Lecture Notes in Electrical Engineering 87

Jian Lee
Editor

Advanced Electrical and Electronics Engineering

Volume 2

Springer

Lecture Notes in Electrical Engineering

Volume 87

Jian Lee (Ed.)

Advanced Electrical and Electronics Engineering

Volume 2

Springer

Jian Lee
Hubei University of Education,
430000., Wuhan
P.R. China,
Hubei
E-mail: leejan4ever@gmail.com

ISBN 978-3-642-19711-6 e-ISBN 978-3-642-19712-3

DOI 10.1007/978-3-642-19712-3

Lecture Notes in Electrical Engineering ISSN 1876-1100

Library of Congress Control Number: 2011923892

Typeset & Coverdesign: Scientific Publishing Services Pvt. Ltd., Chennai, India.

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Preface

We are delighted to invite you to participate in 2010 International conference on Electrical and Electronics Engineering (ICEEE 2010) in Wuhan, China, and December 4–5, 2010. The objective of ICEEE 2010 is to provide a forum for researchers, educators, engineers, and government officials involved in the general areas of Electrical and Electronics Engineering to disseminate their latest research results and exchange views on the future research directions of these fields.

This year, 2010 International conference on Electrical and Electronics Engineering (ICEEE 2010) invites high-quality recent research results in the areas of Electrical and Electronics Engineering.

The main goal of the Conference is to bring together scientists and engineers who work on Electrical and Electronics Engineering. The ICEEE conference will provide an opportunity for academic and industry professionals to discuss the latest issues and progress in the area of Electrical and Electronics Engineering. Furthermore, we expect that the conference and its publications will be a trigger for further related research and technology improvements in this important subject.

ICEEE 2010 will also include presentations of contributed papers and state-of-the-art lectures by invited keynote speakers. The conference will bring together leading researchers, engineers and scientists in the domain of interest from around the world. We would like to thank the program chairs, organization staff, and the members of the program committees for their hard work. Special thanks go to Springer Publisher.

We hope that ICEEE 2010 will be successful and enjoyable to all participants. We look forward to seeing all of you next year at the ICEEE 2011.

Jian Lee Hubei University of Education, China

ICEEE 2010 Organizing Committee

ICEEE 2011 Keynote Speakers

Chin-Chen Chang, IEEE Fellow	Feng Chia University, Taiwan
Jun Wang	The Chinese University of Hong Kong, Hong Kong

General Chairs

Jian Li	Nanchang University, China
Mark Zhou	International Industrial Electronics Center, Hong Kong

Program Chairs

Dehuai Zeng	Shenzhen University, China
Jian Lee	Hubei University of Education, China

Program Committee

Zhu Min	Nanchang University, China
Peide Liu	Shangdong Economic University, China
Ying Zhang	Wuhan Universtiy, China
Changhui Liu	Huazhong University of Science and Technology, China
Ping He	Liaoning Police Academy, China
Minli Dai	Suzhou University, China
Zhenghong Wu	East China Normal University, China
Qingyuan Zhou	Jiangsu Teachers University of Technology, China
Wei Li	Asia Pacific Human-Computer Interaction Research Center, Hong Kong

Contents

Link Building on Demand in Multi-robots

Yi An Cui and En Ming Xiang

School of Info-Physics Engineering, Central South University, Changsha 410083, China
{csu-iag,enmingxiang}@mail.csu.edu.cn

Abstract. To build effective communication link between two robots while physical link did not exist in multi-robots system, a strategy of active communication link building on demand was proposed. Based on some model assumptions and definitions, so called SSS strategy was presented to build communication link in multi-robot system. And the corresponding algorithms were discussed and compared in simulation. The simulation results indicated that communication link could be built successfully by employing these algorithms and the performances of algorithm could be analyzed quantitatively by using energy consumption and time consumption of robots in link building.

Keywords: link building; multi-robots; simulation; algorithm comparing.

1 Introduction

Communications take a basic role in cooperation of robots with information exchanged [1]. At present, a great deal of research on how to keep connectivity of Ad hoc networks has been done in mobile robotics. But it is still a problem how to maintain correspondence reliably in robots while some key nodes do not work. Many methods of networks building was proposed only under optimal conditions without failure of robots. Howard etc. [2] presented a heuristic strategy to build networks, which is centralized and not robust. Pezeshkian etc. [3] let robots be formed up into a line by following a leader. The failure of robots was considered, but the connectivity of networks was not sure. Ulam etc. [4] discussed 4 different methods to rebuild networks under robot fault in multi-robot system. These methods are limited illuminating but not applied. Vazquez etc. [5] and Anderson etc. [6] constructed and maintained networks based on some hypothesis, such as each robot has at least one neighbor and so on.

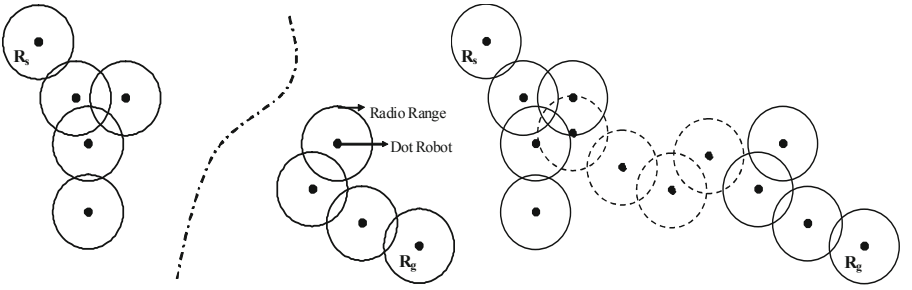
But on some occasion, there is no physical link at all between two robots, any method and strategy mentioned as above is disabled and useless. So an active communication link building method on demand is presented for multi-robot system in unknown environments based on robot's mobility.

2 Problem Description

In cooperative mission robot system, robot R_s and R_g can not move at will for task necessary. But some time they still have to exchange their message. For the reason of

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 1–8.
springerlink.com

limited radio range, they may not communicate with each other directly. In such circumstances, some robots just lying between them act as relays to build communication link. And this is also an ordinary problem studied in Ad hoc networks. But if there are not such relays as shown in figure 1(a), any router discovery protocol would be disable because the link is not here at all. Then it is necessary to move some robots to desired position to act as relay between R_s and R_g. As shown in figure 1(b), 4 robots moved to corresponding place (showed by dotted line) and an active link was build. That is just the problem we will discuss in this paper, so called how to build link between any two robots on demand in multi-robot system.



(a) R_s and R_g can not communicate with each other (b) R_s and R_g communicate with each other by relays

Fig. 1. Active linking in multi-robot system

3 Model and Definition

3.1 Basic Assumptions

For ease of problem presentation, some assumptions about environment and robot were made as follows.

Assumption 1: The environment is an ideal two-dimensional plane without any obstacle.

Assumption 1 extricates us from path planning and obstacle avoiding, that is researched in many literatures about mobile robot.

Assumption 2: All robots are dots and their size is neglected.

Assumption 3: The communication distance of robot is limited to C_l.

Assumption 4: All robots trend group by instinct. The robot without mission would wander in the environment and try to stay nearby other robots.

Assumption 5: The robot has distance and angle sensors as theodolite.

3.2 Connectable Tree and Connectable Frontier

Definition 1: Taking robot R_i as a root node, a tree graph could be constructed by robots that can communicate with R_i directly or indirectly as shown in figure 2. This tree is called connectable tree of R_i , denoted as $T_L(R_i)$.

Definition 2: All leaf nodes in connectable tree are called connectable frontier, and the leaf node with maximum number of ranks is called the most connectable frontier, denoted as $F_R(R_i)$.As shown in figure 2, R_2,R_4,R_6 and R_7 are connectable frontiers of R_i, and R_7 is the most connectable frontier.

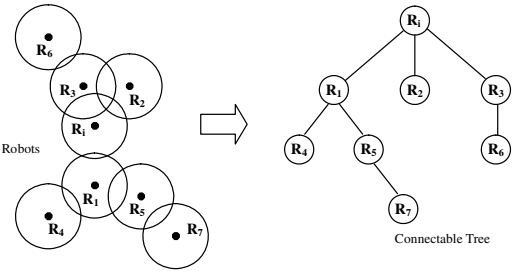


Fig. 2. Connectable tree of R_i

3.3 Space Coordinates and Communication Distance

Definition 3: $p_i(\rho,\theta)$ denotes the polar coordinate of robot R_i in its environment. If a robot is right on the origin (0, 0), it will be called origin robot, denoted as R_0.

Definition 4: $d_p(R_i, R_j)$ denotes the space distance between robot R_i and R_j.

Definition 5: $d_c(R_i)$ denotes the maximum communication distance of R_i, and $d_c(R_i, R_j)$ denotes the maximum communication distance between robot R_i and R_j directly.

Then we can get the equation $d_c(R_i,R_j) = min(d_c(R_i), d_c(R_j))$. And considered of Assumption 3, another equation can be got:

$$d_c(R_i, R_j) = d_c(R_i) = d_c(R_j)=C_l$$

3.4 Communication Link

Definition 6: Beeline link consists of many collinear communication nodes, and forms a maximum connectable communication link. Beeline link can be denoted as $Link\{R_1, R_2, \ldots, R_i\}$.

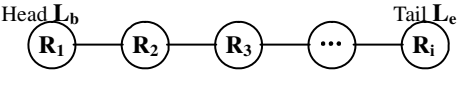


Fig. 3. Beeline link

Definition 7: The two ends of beeline link are called link-end. One is called head-end then the other is tail-end.

Definition 8: Link length means the minimum space distance between the two link-ends on beeline link. It is denoted as d_l.

In figure 3, $d_l = d_p(R_1, R_i)$.

Definition 9: Link-hop, denoted as h_l, equals the number of link nodes subtracted by 1.

In figure 3, $h_l = i-1$.

Definition 10: Increase a new node near the tail-end in a link to be a new tail-end, then a new communication link with $d_l + C_l$ link length is formed. This process is defined as link-increasing, denoted as follows.

$$Link\{\} = Link\{\} + 1$$

Definition 11: Beeline link revolves round it's head-end by an angle θ in a plane while the other parameters are fixed. That is called link-rotating, denoted as follows.

$$Link\{\} = Rev(Link\{\}, \theta)$$

4 Active Communication Link Building

In unknown environments, a strategy called Spiral-Step-Scan (SSS for short) is proposed to build communication link among multi-robot system.

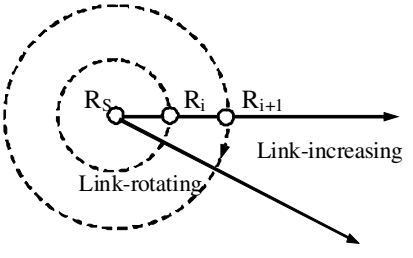


Fig. 4. SSS strategy

The process of SSS strategy includes let R_s be a head-end, combine other robots discovered one by one to form a beeline link, this link run link-increasing and link-rotating by turns. Then the tail-end of link will step in a spiral until R_g is discovered, and the communication link between R_s and R_g will be build at the same time. The detailed algorithm is shown as follows.

Alg. 1. SSS algorithm

Step 1. R_s searchs connectable tree, if R_g is discovered, goto step 10;
Step 2. R_s looks for a free robot R_i in connectable tree;
Step 3. Form a beeline link $Link\{R_s, R_i, \cdots\}$;
Step 4. $h_l \geq \tau$, failure, exit;
Step 5. Set link-rotating counter $Rev\_Counter=1$;
Step 6. $Rev\_Counter \geq 6h_l$, goto Step 9;
Step 7. Tail-end L_e searchs connectable tree, if R_g is discovered, goto step 10;
Step 8. $Link\{R_s, R_i, \cdots\}=Rev(Link\{R_s, R_i, \cdots\},\theta)$; $(\theta=\pi/(3h_l))$
$Rev\_Counter= Rev\_Counter+1$;
Goto Step 6;
Step 9. $Link\{R_s, R_i, \cdots\}= Link\{R_s, R_i, \cdots\}+1$;
Goto Step 4;
Step 10. The link between R_s and R_g is build successfully.

Lacking environmental information, SSS strategy is a blind search algorithm and is inefficient. It may cost much energy and time to build link based on this strategy in multi-robot system. If we take full advantage of the most connectable frontier, the number of link-increasing and link-rotating would be reduced, and the link building would be more efficient. Based on definition 2 and definition 6, there is a corollary as follows.

Corollary 1: If the communication link between Ri and it's most connectable frontier FR(Ri) is a beeline link, then:

$$d_p(R_i, F_R(R_i))=max(d_p(R_i, R_j)), \quad R_j \subset T_L(R_i)$$

In SSS algorithm, change step 2 and Step 3 as follows.

Alg. 2. SSS algorithm based on the most connectable frontier

Step 2. R_s looks for the most connectable frontier robot $F_R(R_i)$ in connectable tree;

Step 3. Let $R_i = F_R(R_s)$, form a beeline link $Link\{R_s,\ldots,R_i,\ldots\}$;

Theoretically, this improved algorithm should work with better performance for link building.

5 Simulation and Analyzing

To verify the validity of these algorithms, some simulations were implemented in multi-robots virtual grid environmental system which was developed by ourselves. 2-D local environment was set to be 88×88 grids, and the visible part were 40×62 grids. 11 dot robots were put in the above environment randomly. Showed as figure 5, the distance between R_s and R_g is 40 grids, denoted as d_p(Rs, Rg)=40, and set C_l=10. All robots are free except R_s and R_g.

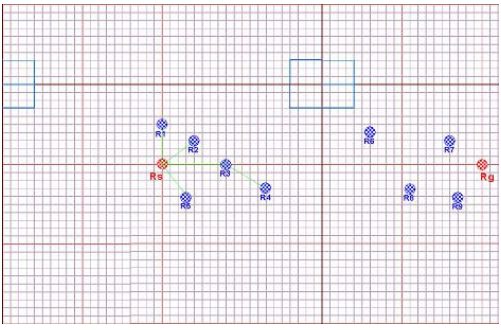


Fig. 5. Multi-robots virtual grid environment

The process of simulation was shown as figure 6. In the figure, (a), (b) and (c) are the main states of simulation with SSS algorithm. The main states of simulation with improved algorithm are shown as (d), (e) and (f).

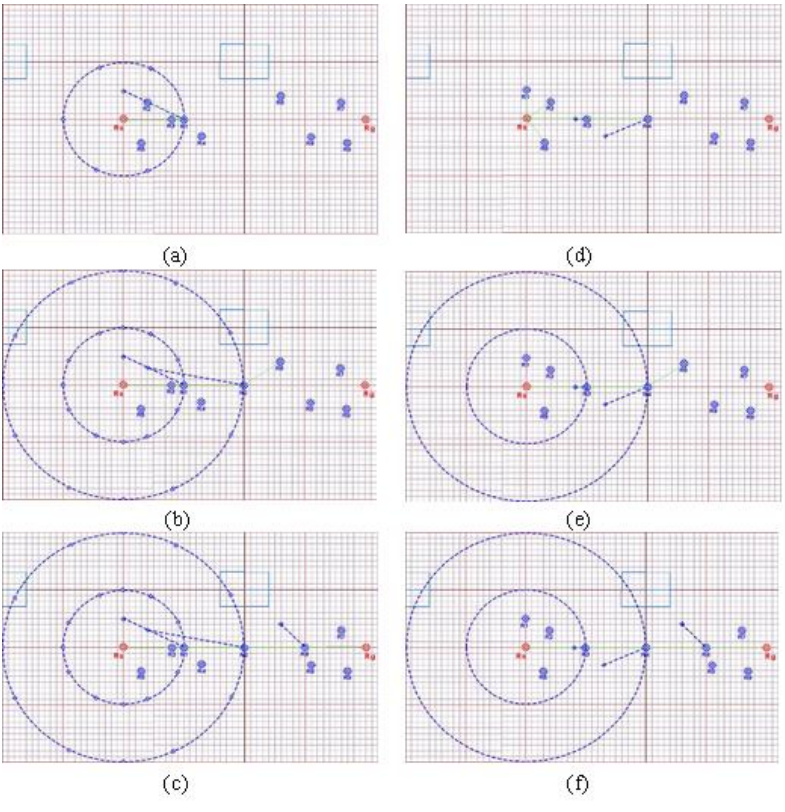


Fig. 6. Process of simulation

Detailed comparison of process with two algorithms was shown in table 1.

Table 1. Comparison of process

SSS Algorithm	Improved SSS Algorithm
Form a beeline link $Link\{R_s,R_1\}$ and link-rotating one revolution, shown as Fig.6 (a).	Look for the most connectable frontier and form a beeline link $Link\{R_s,R_3,R_4\}$, shown as Fig.6 (d).
Link-increasing to be $Link\{R_s,R_1,R_2\}$ and link-rotating one revolution, shown as Fig.6 (b).	$Link\{R_s,R_3,R_4\}$ link-rotating one revolution, shown as Fig.6 (e).
Link-increasing to be $Link\{R_s,R_1,R_2,R_6\}$ and find R_g, shown as Fig.6 (c).	Link-increasing to be $Link\{R_s,R_1,R_2,R_6\}$ and find R_g, shown as Fig.6 (f).

As the simulation results, a communication link between R_s and R_g was build successfully via R_1, R_2 and R_6 by using both two algorithms. But the improved SSS algorithm had more optimal performance than SSS algorithm from two aspects. They were energy consumption and time consumption that could be given by our multi-robots virtual grid environmental system in the form of detailed data. In the system, energy consumption (*TEC* for short) was estimated depend on the total path length of all robots in one mission.And time consumption (*TTC* for short) was the number of clock tempos that was occupied by robots to complete a mission. In the above simulation, compared to SSS algorithm, the improved one reduced 28.8% of the energy consumption and 37% of the time consumption.

6 Conclusion

A concept about active communication link building on demand was proposed for multi-robot system in unknown environments. In order to realize active communication link building on demand, a basic SSS strategy was presented. And the corresponding algorithm even improved algorithm was discussed. The simulation result proved the validity of these strategy and algorithms by building link successfully between two robots on demand. By employed two parameters *TEC* and *TTC*, the improved algorithm was proved more optimal quantitatively. The future work should focus on improving the algorithm further by using some self-learning strategy.

References

1. Yun, X., Yimin, Y.: Overview of autonomous robot soccer system. Robotics 26(5), 475–480 (2004)
2. Howard, A., Matadd, M., Sukhatme, G.: An incremental self-deployment algorithm for mobile sensor networks. Autonomous Robots, Special Issue on Intelligent Embedded Systems 13(2), 113–126 (2002)

3. Pezeshkian, N., Nguyen, H.G., Burmeister, A.: Unmanned Ground Vehicle Radio Relay Deployment System for Non-line-of-sight Operations. In: Proceedings of IASTED International Conference on Robotics and Applications, Wuerzburg, Germany, pp. 29–31 (2007)
4. Ulam, P., Arkin, R.: When good communications go bad: Communications recovery for multi-robot teams. In: Proceedings of 2004 IEEE International Conference on Robotics and Automation, Barcelona, Spain, pp. 3727–3734 (2004)
5. Vazquez, J., Malcolm, C.: Distributed multi-robot exploration maintaining a mobile network. In: Proceedings of Second IEEE International Conference on Intelligent Systems, pp. 113–118 (2004)
6. Anderson, S., Simmons, R., Goldberg, D.: Maintaining line of sight communications networks between planetary rovers. In: Proceedings of the Conference on Intelligent Robots and Systems (IROS), San Diego, pp. 2266–2272 (2003)

Energy Conservation Renovation on 600MW Units Condensate Pump

Yingbai Xie, Jianlin Liu, Zhichao Wang, and Yun Liu

Department of Energy & Power Engineering,
North China Electric Power University, Baoding, China
xieyb@ncepu.edu.cn

Abstract. Condensate pump is employed to pump the water in condenser to deaerator, maintains the water height in deaerator and enables units operate continuously. Currently, system efficiency of 600 MW units is comparatively low due to high design pressure head and large throttling pressure loss. This paper takes NLT500-570x4S condensate pump as example, and analyzes its existing problems based on performances test. Further, aims at solve these problems, some modification works, such as reducing the throttling pressure loss and reforming the through-flow as well as improving the craft of inspection, are performed. The results show that the highest efficiency of the condensate pump was improved with 5.5% under the best design flux.

Keywords: fluid machinery; efficiency; energy conservation renovation; condensation pump; performance test.

1 Introduction

Power plants rated at 600 MW and above are the main development tendency in China. The ratio of power plants rated at 600 MW and above is 48% among all the new founded power generation capacity in year 2006. But from January to September in 2007, the ratio is increasing to 57% [1]. With the emphasis on energy saving in electric power generation industry and the urgent need for electric in China, more and more large-capacity and high efficiency units will be in operation to replace the low capacity, low efficiency and high-polluting units. To guarantee the reliable operation in large capacity units is also important [2-4].

Condensate pump is the main auxiliary equipment of the turbine generator system, the function is to pump the condensation water heated by low pressure heater in condenser water tank to the deaerator, and keep the deaerator water level stable. Currently, condensate pumps of 600 MW units are using fixed-speed pumps. During operation, units load and openings of adjusting valve vary in a wide range. Especially in low load condition, the throttling loss is numerous. In addition, the design is more conservative, and more emphasis on safety, less consideration on energy conservation, these result in the large the design capacity in condensate pump.

Therefore, condensate pump has great energy-saving potential when it operating in the large flow rate tolerance and frequently changed work conditions [5-6].

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 9–13.
springerlink.com

In this paper, energy saving retrofit on condensate pump is processed according on the actual filed condition.

2 Existing Questions

In our plant, each 600MW unit is equipped with two NLT500-570 × 4S type condensate pumps made by Shanghai KSB Ltd., the designed parameters are shown in Table 1.

Table 1. Designed parameters of condensate pumps.

Type	Flux	Head	Rotate Speed	Efficiency
NLT500-570×4S	1427-1630m^3/h	346-328m	1490rpm	83.5%

Matched motor is 2000KW. Active power of a single condensate pump is at about 1800KW when running at rated capacity. Double suction impeller is used in the first stage; the latter three stages adopt single suction impeller. Distribution of pressure lifting is in terms of stage.

During operation, field test indicate that 250 meters head of delivery can meet the requirements, but the actual head of delivery of condensate pump is about 320 meters. Considering the stability operating of unit, 300 meters actual head of delivery is enough. The remaining head of delivery will dissipate in system. This is the main cause to retrofit condensate pump.

Moreover, actual efficiency of condensate pumps is deviated from the designed efficiency. The designed efficiency of condensate pump is 83.5%, but the field test efficiency is about 78%.

3 Performances Testing and Analysis

Performances and pipe resistance tests are carried out for the condensate pumps according to GB/T3216-2005. During test, parameters such as flux, head of delivery, electric power, units load should be steady. Fluctuation of values must be less than 3%. Data for the performances test of condensate pump are shown in Table 2.

Table 2 shows that the proportion of throttling pressure loss in adjusting valve of condensate pump is great. Among the five testing conditions, the head of delivery loss in adjusting valve ΔH accounts for the total head H of delivery at 38.22% to 72.86%. So the first step is to reduce the pressure loss in adjusting valve by replace the throttling adjusting valve with frequency conversion adjusting device.

In addition, the efficiency of condensate pumps is low. The maximum efficiency is 77.31%, and only 60.99% for low load at 298.1MW. There exist a large difference between actual performance and design value.

Table 2. Data for the performance test of condensing pump.

Item	Units	Units Load (MW)				
		298.1	418.7	478.5	538	594.2
Pump Outlet Pressure	MPa	3.596	3.4218	3.321	3.207	2.98
Pump Inlet Pressure	MPa	-0.0805	-0.0813	-0.0824	-0.0827	-0.084
Pressure Before Adjusting Valve	MPa	3.56	3.3955	3.251	3.106	2.86
Pressure After Adjusting Valve	MPa	0.8	1.145	1.3	1.445	1.59
Flux	t/h	761.3	1023.3	1183.7	1325.1	1456.2
Kinetic Pressure Difference with Pump Inlet and Outlet	m	0.137	0.247	0.326	0.416	0.488
Head of Delivery	m	379.06	361.43	351.64	340.23	317.06
Electric Power	kW	1357.3	1518.6	1604.8	1675.2	1739.1
Shaft Power	kW	1289.4	1442.7	1532.4	1599.8	1660.8
Proportion of Throttling Pressure Loss in Adjusting Valve	%	72.86	61.79	55.02	48.07	38.22
Pump Efficiency	%	60.99	69.86	74.01	76.79	77.31

In the performance of testing, we also find that condensate pump has casting defects. There is cellular point on inner cover, and the symmetry of flow passage is bad.

4 Technical Retrofit Measures and Its Realized

According to the results and analysis of performances testing, following technical retrofit measures have been approved.

In order to reduce the throttling pressure drop of adjusting valve, the second stage Impeller is removal, and leaves the position empty. Thus no guide blade casing need to machine, the amount of retrofit work will be reduced. If the second stage Impeller need to re-install, any other technical measures are needed. But the guide blade casing still exists, which will increase the flow resistance, resulting in more than 1 % drop of the efficiency of the entire pump.

Improve on the flow section of the pump. Key local molded line is improved by using the model optimization testing and patented technology to enhance the efficiency of the pump and the ability of preventing cavitations. Detailed measures are:

(1) Optimization of the blade profile of impeller inlet and outlet. The inlet blade profile of the first stage is designed as the fish head shape. Increase the entrance angle β_1 by 1.8^o average and the opening port α_1 by 0.5mm. At the same time, the blade pitch error of the entrance is limited to less than 0.5% to ensure uniformity symmetry of entrance. Increase the angle of blade discharge β_2 by 1.2^o and reduce the thickness of blade outlet from 8.8mm to 3.8mm. The blade pitch error of the outlet is also limited to less than 0.5% to ensure uniformity symmetry of flow section.

(2) Optimization of the flow channel molded line of axial impeller channel. Augment the diameter of the throat D_0 and the wheel D_d in the first stage impeller from φ280mm to φ282.5mm and from φ100mm to φ99mm separately. The eccentricity of D_0 and D_d is eliminated. Outlet width of impeller is optimized from 73mm to 76.5mm, and the local eccentricity of inner flow surface on both sides and the baffle in central is also eliminated.

(3) Optimization of the whole flow channel of impeller. Improve of the glossiness and smoothness in all flow channels, including the transition zone of the adjacent parts, make the irregularity degree less than 0.1mm. This will not only improve the inflow and outflow of the impeller, significantly improve the pump efficiency and capacity for preventing cavitations, enable the high efficiency range of pumps shift to large flow rate, but also enable the flow in the impeller be more uniform and symmetry, enhance the stability of pump operation.

For the other stage impeller, basic principles for optimization are similar to which of the first stage, but parameters are different.

Improve maintenance and assembly process:

(1) Do static equilibrium, high-speed dynamic balance testing and flaw detection inspection for each impeller;

(2) In strict accordance with the manufacturer's process requirements to assemble and to ensure a reasonable seal gap.

After carrying out list retrofit technical measures, performances test of condensate pump is re-done. Referred methods are the same as before. When at the optimum design flow rate $Q=1600m^3/h$, namely capacity is 1588t/h, the head of delivery of condensate pump H=263.5m, efficiency of condensate pump η=83.15%. From the existing performances test curves, the retrofit reaches the expected goal.

5 Conclusions

The designed head of delivery of 600 MW condensate pump is comparatively high which results in larger throttling loss and reduces the system performance. For NLT500-570x4S condensate pump in our plant, energy conservation retrofit is processed based on existing problem analysis and performances test. To solve these problems, optimization improvements, such as reduce the throttling pressure loss in adjusting valve, optimization of the flow section, and improve of maintenance and assembly process are performed. Field test indicates that when operating at optimum designed flow rate condition, maximum efficiency of condensate pump is improved at a order of 5.5%.

All the work done in our plant can provide a reference for energy conservation retrofit on condensate pump of 600MW unit.

References

1. Gu, J., Zhang, J.: Transformation of condensing water pump of 600 MW Units with variable frequency drive technology. North China Electric Power 12, 46–48 (2006)
2. Cheng, W., Xia, G.: Zhou Yin: Optimized governing of condensate pumps. Power Engineering 24, 739–742 (2004)

3. Ju, W., Zhu, L., Yu, X.: Optimization technology application of auxiliary equipment operation mode of steam turbine. Electrical Equipment 8, 16–18 (2007)
4. Das, G., Ray, A.K., Ghosh, S.: Fatigue failure of a boiler feed pump rotor shaft. Engineering Failure Analysis 10, 725–732 (2003)
5. Hua, S.: Energy-saving retrofitting on frequency speed-governing for condensate pump. Northeast Electric Power Technology 26, 50–52 (2005)
6. Fan, K., Mao, J.: Efficiency of Heat-engine Plant Frequency Conversion Control Transformation. Shanghai Electric Power 20, 204–206 (2007)

FPGA-Based PCI Data Acquisition Synchronization Card of Railway Clearance Detection System*

Xining Xu and Zujun Yu

State Key Laboratory of Rail Traffic Control and Safety, School of Mechanical, Electronic and Control Engineering, Beijing Jiaotong University, Beijing 100044, China
{xuxining,zjyu}@bjtu.edu.cn

Abstract. Railway clearance detection system based on rotating laser scanner sensors was designed for the realization of non-contact detection of railway clearance. In the system, in order to acquire and synchronize the data of Laser scanner sensors, gyroscope, position transducers and speed sensor, the FPGA-based PCI data acquisition synchronization card was designed. This paper describes the hardware design for acquisition synchronous card, PCI interface chip configuration, FPGA core programming. The experiment proved that using the data acquisition synchronization card, 2-channels of 500k baud rate serial data, 1-channel of 115200 baud serial data, 4-channel A/D, 2-channel high-speed IO signals can be acquired and synchronized, the time accuracy is 1ms, without missing data. This ensures reliable operation of the whole railway clearance detection system.

Keywords: FPGA, Acquisition, Synchronization.

1 Introduction

As the rapid development of the high-speed railway and urban railway transport, the acceptance of lots of new lines and reconstruction lines needs high speed clearance detection device. The traditional clearance detection system is based mainly on contact mechanical tentacles, which is low degree of automation and difficult to compensate the vibration. With the development of photoelectric technology, the laser scanner sensor based on light propagation principle is widely used in detection system of high-speed railways abroad now. By calculating the time span between the emission and reflection of the laser, the scanner measures the distance between the sensor and the obstacle. When the angle changes with a certain angular velocity, the laser pulse can scan different perspectives on the obstacles to obtain the profile size of the current scanning section.

The system uses two high-speed laser scanners to detect the railway clearance dimension in real-time. The clearance data are stored in database with the railway line kilometer as the index, through data processing to get the current position and the overloaded line clearance gauge size, easy to railway line construction, renovation

* This work is partially supported by the State Key Laboratory of Rail Traffic Control and Safety (Contract No. RCS2009ZT012), Beijing Jiaotong University.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 15–23.
springerlink.com

and maintenance. Comparing the historical clearance data of the same railway line, we can master the variation trend of the line clearance dimension and exclude the hidden danger that affect the railway transportation safety timely, for the safe operation of rail traffic is significant.

2 System Solution

The overall scheme of the clearance detection system is shown in Fig. 1.

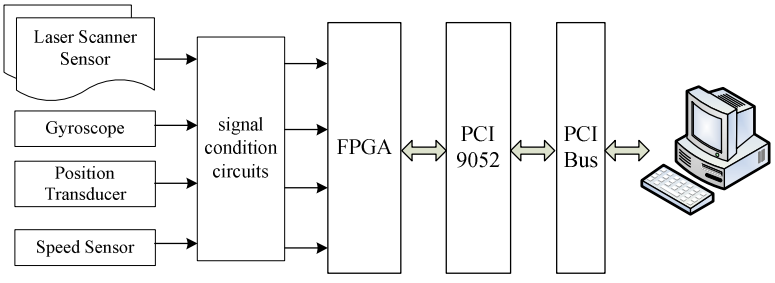


Fig. 1. The overall scheme of the clearance detection system

The LMS200 series high-speed rotation laser scanner sensors made in SICK, a company of Germany, are installed directly on the railroad car or dedicated detection vehicle to measure the clearance dimension of the line. LMS200 can be achieved within the 180º scan, repeated measurement error is less than ±5mm, data transfer rate up to 500kbps, a minimum sampling interval of 0.5º. When two scanners installed on the detection car with an angle of 70º, we can acquire the clearance dimension within a range of 290º.

The laser scanner sensors installed on the detection vehicle will have six degrees of freedom uncertain gesture, which will affect the clearance size measurement accuracy, during the running of the vehicle. System uses one gyroscope and four position transducers to measure the body's movements gesture in real-time, and compensate the clearance data with the rotation matrix method, which will correct the measure error caused by vehicle vibration.

In order to determine the location of exceeding clearance data, the clearance dimension data acquired by laser scanner need to correspondence with the line kilometer, therefore, the system also adds a speed sensor for real-time measurement of vehicle speed and location information.

All the data in the collection must be synchronized, otherwise the dynamic compensation is unable to carry on and it will affect the measure accuracy of clearance dimension.

Currently, the common method of data acquisition is to use various data acquisition card. With this acquisition mode, the sensors' data are read from each acquisition card and the PC time is recorded at the same time, which is used as the time label of the current data pack. Each kind of measure data synchronizes by the time label of the data pack. While because of the internal cache and the time delay of the data

acquisition cards, and the PC time is not precise enough, it is difficult to achieve accurate data synchronization.

In order to achieve high accuracy synchronization in data acquisition, a FPGA-based PCI data acquisition synchronization card was designed. In this card, the high-speed serial port capture, AD acquisition, high-speed IO signal acquisition functions are integrated into one FPGA chip, while a sync pulse of 1KHz is generated in the FPGA. The count of the real-time synchronization pulse is stored in 4Byte registers, and recorded following each packet of data. Using the 4Byte sync pulses count, the acquisition time of each data can be known, its accuracy is 1ms, with these time tag, synchronization can be achieved.

Large amounts of synchronization data which were collected by the FPGA need to be read and stored in PC in high-speed, so the PCI bus interface is used. PCI (Peripheral Component Interconnect) bus data transfer rate is up to 132MB/s, plug and play, and is widely used in various computer systems areas.

3 Hardware Design of the FPGA-Based PCI Data Acquisition Synchronization Card

PCI data acquisition synchronization card hardware circuits include signal conditioning and data acquisition circuits, FPGA chip, PCI interface circuit and the configuration chip.

3.1 Signal Conditioning and Data Acquisition Circuits

1) Laser Scanner Sensors and Gyroscope: The laser scanner sensors and gyroscope output through the RS485 serial port, the communication baud rate of laser scanner sensor is 500kbps and the gyroscope sensor is 115200bps. System first uses the MAX490 as a level converter chip to convert the RS485 signal into a TTL level signal, the signal is isolated through a photo coupler and sent into the IO port of the FPGA.

2) Position Transducer: The output of the position transducer is 4-20mA analog signal, it is converted to 0.48-2.4V voltage signal with a 120Ω precise resistor. Through a capacitor filter and the limited voltage protection circuit, the signal is sent to the AD7888. The TTL signals from the SPI port of AD7888 are sent into FPGA's IO port after isolation by photo couplers. FPGA can read 12-bit AD data by designing the IP core according to the timing of AD7888.

3) Speed Sensor: Speed sensor outputs two pulse signal, the phase difference is 90º, the two pulse signal are sent to the FPGA's IO port after isolation by photo couplers. One signal is used as pulse count and kilometer calculation. According which signal's phase is ahead, we can know whether the vehicle is moving forward or backward.

3.2 Selection of FPGA Chip

FPGA chip is the core of the system, the ALTERA Cyclone series FPGA - EP1C12 is used in the system, it has 12,060 logic cells, 52 M4KRAM, 239616 total RAM bits, 2 PLL, 249 pins of available user IO. It's easy to develop with FPGA, just connect

FPGA's I/O ports to external counterparts, and then design the internal logic of the system, that's ok. You can then implement incremental changes and iterate on an FPGA design within hours instead of weeks. As system requirements often change over time, the cost of making incremental changes to FPGA designs are quite negligible. All of the data acquiring programm, synchronous pulse generator, FIFO, PCI interface timing and other functions are realized in the FPGA internal, FPGA internal functional block diagram shown in Fig. 2.

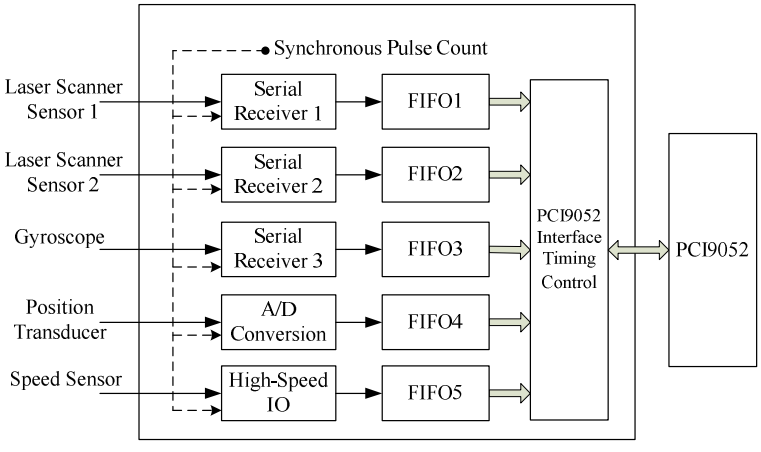


Fig. 2. FPGA internal functional block diagram

3.3 PCI Interface Chip

PCI9052 is a common PCI bus interface chip of U.S. PLX company. Using a dedicated PCI bus interface chip, complex PCI protocol will not be concerned, we only need to develop hardware and drivers, it will reduce the development cycle greatly. The PCI9052 is compliant with PCI r2.1, supporting low cost slave adapters, it allow relatively slow Local Bus designs to achieve 132 MB/s burst transfers on the PCI Bus. Its Local Bus clock runs asynchronously to the PCI clock, allowing the Local Bus to run at an independent rate from the PCI clock. PCI9052 device contains four local chip select signal and five local address spaces, chip select and address space can be configured through the EPPROM or host. The PCI9052 supports 8-, 16-, or 32-bit Local Buses, which may be Non-Multiplexed or Multiplexed mode. PCI bus interface using PCI9052 is shown in Fig. 3.

3.4 Configuration of the Serial EEPROM

PCI bus supports three physical spaces: the memory address space, I/O address space and configuration space. All the PCI devices must provide configuration space. System uses a serial EEPROM - 93CS46 to configure PCI9052. The method of using EEPROM to configure PCI9052 will be introduced.

1) PCI Configuration Registers: Registers of 00H-0FH in EEPROM are used to configure the PCI configuration registers. Only register of 00H needs to be configured, the others use the default settings. Writing configuration data to EEPROM, it should be attentive that low front, high in the post. For example, we write 0x10B5 to registers of 2H and 3H, 2H = 0xB5, 3H = 0x10.
2) Local address space: Registers of 10H, 12H in EEPROM are used to configure the local address space. The hexadecimal data FFFFFFE1 is written to the registers of 10H, 12H in this system, it means that the local address space 0 maps to PCI-IO space, 32-bit PCI addressing mode is used, the local effective address range is 00H to 1FH.
3) Bus Region Descriptors for Local Address Space: Registers of 38H, 3AH in EEPROM are used to configure the bus region descriptors for local address space. The hexadecimal data D0118940 is written to the registers of 38H, 3AH, so 8bit PCI bus mode is used in this system.

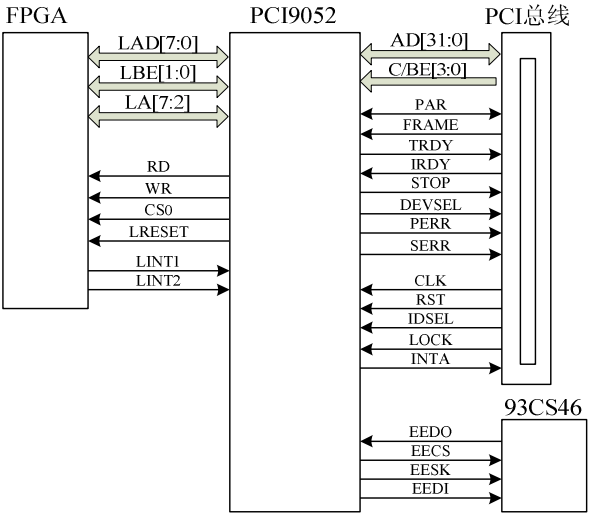


Fig. 3. PCI bus interface

4 FPGA Core Programming

Most of the system functions are implemented in the FPGA, FPGA program mainly completes the sensors' data acquisition, synchronization pulse generation and counting, PCI9052 interface control timing and other functions. The design of receiving data from serial port is more complex, the following will detail the development of the FPGA serial data reception program.

4.1 Design of Serial Data Reception

The data output rate of laser scanner sensor is 500KHz, each packet is 732 bytes, 37.5 packets per second. In order to receive the serial data accuracy, and add synchronization pulse count in the end of each packet of data, the system needs to complete some work as the following:

1) Generate baud rate clock multiplier: Serial data communication is an asynchronous data transfer mode, carrying out according to a certain baud rate, data receiver detects the transfer signal in accordance with the certain baud rate, so as to obtain communications data. During this process, except receiving data bit, it is also the need to detect start bit and stop bits, shift the data bit, add the synchronization pulse count. So we need a clock which is several times of baud rate clock to get more time to complete all the work, the system uses 32 times.
2) Serial to Parallel Converter: Output serial data format of the laser scanner sensor is: 1 start bit, 8 data bits, 1 stop bit, no parity bit. All bits in the serial data stream should be detected, and 8 serial data bits need to be converted to parallel data.
3) Add Sync Pulse Count: After receiving a byte of data, the current byte and the previous byte on a combination compares with the packet's end flag, if meet, after writing the current byte to FIFO, the sync pulse count - 4 bytes are written to the FIFO.

Diagram of serial data reception program is shown on Fig. 4.

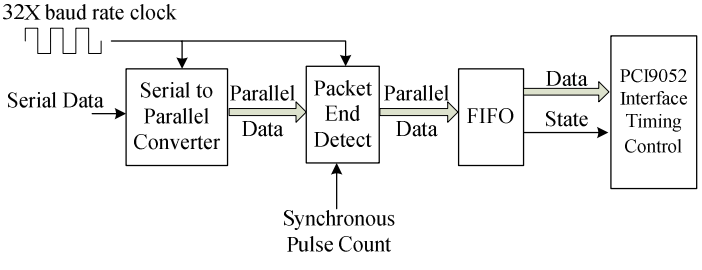


Fig. 4. Diagram of serial data reception program

4.2 State Machine

Serial data receiving process including a total of 7 states, namely IDLE (idle), START (start bit detection), SHIFT (character shift), STOP (stop bit detection), STORE (parallel data storage), DETECT (data Package end flag detect), WRPULSE (write sync pulse count). The data bit count value-COUNTER_BIT, the 32 times the baud rate clock count value-COUNTER32, the RXD input value and data Package end flag detecting result are combined together as a basis for judging the state machine transfers. State transition is shown in Fig. 5.

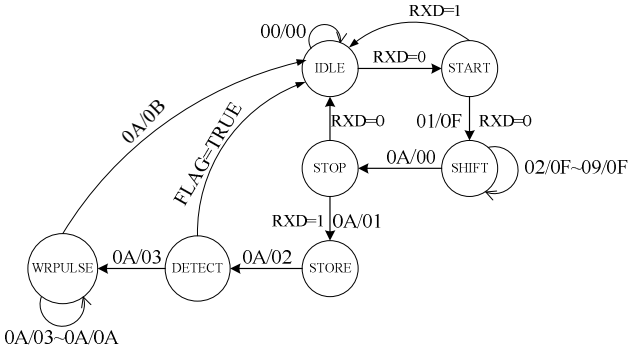


Fig. 5. State transition of serial data reception

5 Experimental Result

After completing the design work, we had a test in the laboratory by connecting the PCI acquisition synchronization card and all the sensors. The data acquisition frequency of laser scanner sensor is 37.5Hz, gyroscope, position transducer, Speed Sensor are all 100Hz. Data samples taken within 100ms, when the packet acquisition is complete, the synchronous pulse count value (in ms) are shown in Table 1.

Table 1. The Synchronous Pulse Count Value

Packet Num	Speed Sensor	Position Transducer	Gyroscope	Laser Scanner Sensor 1	Laser Scanner Sensor 2
1	151815	151815	151812	151819	151821
2	151825	151825	151822	151845	151848
3	151835	151835	151832	151872	151875
4	151845	151845	151842	151899	151901
5	151855	151855	151852		
6	151865	151865	151862		
7	151875	151875	151872		
8	151885	151885	151882		
9	151895	151895	151892		
10	151905	151905	151902		

In this system, the time sampling interval of laser scanner sensor is 26.6ms, the others are 10ms. Among them, the laser scanner sensors and gyroscope are actively sending data, so we cannot control the time of generating data for acquisition synchronization. For this reason, the synchronous pulse count value is added at the time of completing data acquisition to get the moment of data acquisition, the time accuracy is 1ms. By this time label, look for the closest of all data packets, put all this data packets as one group data to process and synthesize, and then to achieve effective data synchronization.

Table 1 shows, speed sensor and position transducer data acquisition can be synchronized by control. The time of Gyroscope's data generation is at random, but the time interval is invariably. The time sampling interval of Laser scanner sensor is

26.6ms, need to find matching data by algorithm. In the system, if the time difference of collection between the Laser scanner sensor and Speed Sensor less than or equal to 5ms, we think them as a group of data. Such as Table 1, the third data packet of Laser scanner sensor and the seventh packet data of others is one group data.

The Railway Clearance detection System with the FPGA-based PCI data acquisition synchronization card is used on the Road-Rail Amphibious Monitoring Vehicle developed at Beijing Jiaotong University. The boundary dimensions of Yizhuang Beijing metro line depot was detected with the vehicle in the acceptance check of the new subway line. After synthesizing the data collecting from the area with ladders, the result shown as Fig. 6. The semi-closed area in the middle of the figure is a border of clearances, on the right-hand is the ladder, the size of the ladder is out of range.

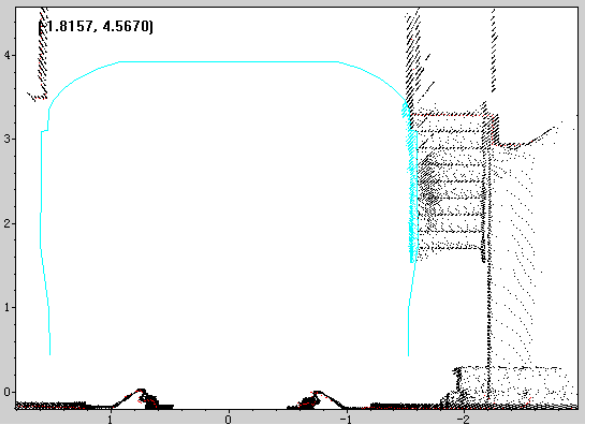


Fig. 6. The boundary dimensions of Yizhuang metro line depot

6 Conclusions

Two channels 500k baud rate serial data, one channel 115200 baud serial data, four channels A/D, two channels high-speed IO signals can be acquired and synchronized, in the FPGA-based PCI data acquisition synchronization card of railway clearance detection system, time synchronization accuracy is 1ms, no packet loss. When the monitoring vehicle runs at the speed of 5Km/h, the interval of each profile clearances is 3.7cm. The line position can be matched with the clearance data by data processing, when there is exceeding clearance, according to the corresponding kilometer mark, we can quickly find the location of overrun points. The FPGA-based PCI data acquisition synchronization card is used on the Road-Rail Amphibious Monitoring Vehicle developed at Beijing Jiaotong University, the clearance of railway and subway line can be acquire in real-time. Now the vehicle has been used in the acceptance test of Beijing new subway line, through the use of this vehicle, we can monitor the new railway line quickly, find the position which out of clearances, facilitate the construction rectification. This system also provides a new method for the urban rail transport monitoring in the future.

References

1. Krobb, C., Krauhausen, M., Priem, R., Katthöfer, H.O., Hülsmann, M.: Measurement of rail profiles using 2D laser sensors, pp. 149–154. VDI Verlag GMBH (2002)
2. Yu, Z., Zhu, L., Zhao, J.: Study on System Scheme for Vehicular Kinematical Measurement of Railway Complete Profile, Engineering Research Center of Structural Reliability and Measurement Technology of Railway Vehicles Ministry of Education. Beijing Jiaotong University, Beijing (2007)
3. Bleicher, F., Puschitz, F.: Laser based measurement system for calibrating machine tools in 6 DOF. In: Annals of Daaam for 2006 & Proceedings of the 17th International Daaam Symposium, pp. 39–40 (2006)
4. SICK. LMS200/211/221/291 Laser Measurement Systems (December 2006)
5. Li, R., Wu, J., Zeng, M.: Metro Clearance Detection Based on Laser Measurement. Urban Rapid Rail Transit 20(5), 70–73 (2007)
6. PLX. PCI 9052 Data Book (September 2009)
7. MacKinnon, D.K., Aitken, V.: A Comparison of Precision and Accuracy in Triangulation Laser Range Scanners. Electrical and Computer Engineering, Canadian Conference on May 2006 200, 832–837 (2006)
8. Altera: Cyclone Device Handbook, vol. 1 (2005)

Short-Circuit Characteristics of Asynchronous Generators and Double-Fed Induction Generators Inter-connected Grid*

Su Chang-Sheng[1,**], Li Feng-Ting[1], and Liu Wei[2,***]

[1] School of Electrical Engineering, Xinjiang University, Urumqi 830047, Xinjiang Uygur Autonomous Region, China
sucs126@sina.com
[2] Zhangjiakou Power System Company, Zhangjiakou 075000, Hebei Province, China
xjlft2009@sina.com

Abstract. In this paper, the electrical transient mode of induction generators is studied, establishing the short-circuit current sequence component of wind generators are when faulty occuring in collect feeder, and analyzed the major effective factors that influence short-circuit characteristics such as wind turbine types, fault type, fault resistance and the number of generators participating ect. Take a 49.5MW wind farm as an example in North China, the dynamic models with asynchronous generators and double-fed induction generators are established based on PSCAD/EMTDC, the fault characteristics of two types wind generators are thoroughly simulation analyzed, the results of simulation are identical with theoretical analysis. The simulation results show that these influence factors should be considered when we design the relay protection of wind farm, it is helpful to enhance the tripped performance of relay protection.

Index Terms: asynchronous generators, double-fed induction generators, short-circuit characteristics, Simulation analysis.

1 Induction

Owing to wind power play an important role in sovling environmental problem and coping with enegy crisis, so the installed wind turbines is rapidly increasing worldwide. with the fast development of wind power, the impact of wind generator on power system become an important reseacher topic[1], however, most study focus on transient stability and power quality, very little attention has been given to analyze the fault characteristics of wind turbine generators and received less achievement.

In the past, the relay protection of wind power system neglect the short-circuit current contrbution of wind farm[2-3]. With the increase of capacity of wind power in

* This work was supported by National Natural Science Foundation of China(NSFC) and Key Project of Xinjiang Uygur Autonomous Region Education Office.

** Su Chang-Sheng is with college of Electrical Engineering, Xinjiang University, Urumqi, 830047, China

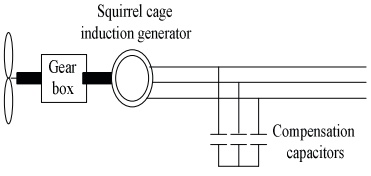
*** Li Feng-Ting is with Xinjiang University, Urumqi, 830047, China

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 25–35.
springerlink.com

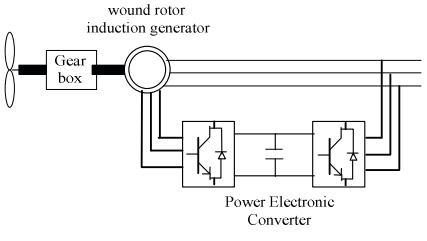
power system every year, short-circuit current of wind turbine contribution is also growing, when the wind power capacity reach to a certain extant, short-circuit current of wind farm will be higher than short-circuit current of grid, and the protection of wind power system must consider the short-circuit current contribution of wind turbine generators, and fault characteristics is foundation of relay protection setting, consequently, research on the short-circuit characteristics of wind farm will be significant.

Wind generator power is different form conventional thermal, hydro power and nuclear generation, which arise different analysis methods regarding the short-circuit characteristics and relay protection of wind farm. Also, reseach regarding wind farm short-circuit characteristics and relay protection are still limited as follows: The trip boundary of distance protection should be setted adaptively that take into account the number of generators, loading level and system frequence[4]. The short-circuit characteristic of asynchronous wind generators was analyzed from fault type and varying wind speed [5-7]. The primary objective of this paper is to simulation analysis the short-circuit characteristics of squirrel cage induction generators and double-fed induction generators, consider the main effctive factors that influence fault characteristics, establishing the dynamic simulation model using PSCAD/EMTDC, the conclusions could applied on the current instantaneous protection. A 49.5MW wind farm in North-China is considered as a simulation example in this reseach.

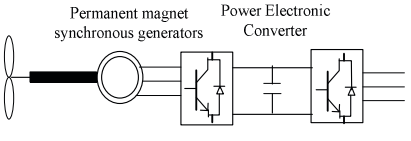
The paper is organsied as follows. The three main types of wind turbine are introduced in section 2, section 3 analyzed the mathematical model of wind turbine generators,the description of the North-China wind farm that used as simulation example in section 4. section 5 highlights the simulation results in different influence factors. Finally, section 6 summarized the conclusion and sppendix that used in this paper.



(a) asynchronous squirrel cage induction generator



(b) double-fed induction generator



(c) permanent magnet synchronous generator

Fig. 1. Three main types of wind turbine

2 Wind Turbine

Nowdays, three main types of wind turbine are commonly being installed in china. The fixed speed wind turbine with squirrel cage asynchronous induction generator directly connect to grid(Fig.1 a), and varible-speed constant frequence wind turbines with wound induction generator and a converter on rotor circuit known as doube-fed induction generators, Fig.1c show the variable speed wind turbine, with the generator connect to the grid through a full power converter in the stator circuit.

3 Mathematical Model

This section describes several dynamic mathematical models of wind turbines and induction generators.

3.1 Wind Tubine

The relation between the wind speed and aerodynamic torque may be described by the following equation[8-9]:

$$T_M = \frac{1}{2}\rho\pi C_{p(\beta,\lambda)} R^3 \frac{V_W{}^2}{\lambda} \tag{1}$$

Where TM is the aerodynamic torque extracted form wind[Nm], ρ the air density[kg/m3], R the wind turbine rotor radius[m], V_W the equivalent wind speed, β the pitch angle of wind turbine[deg], Cp the aerodynamic efficiency, the tip speed ratio λ:

$$\lambda = \frac{\omega_{rot} R}{V_W} \tag{2}$$

here ω_{rot} is the rotational speed of wind turbine [rad/s].

Numerical approximations have been developed to calculate Cp for give values of β and λ, here the following approximation is used:

$$C_p = 0.22\left(\frac{116}{\lambda} - 0.4\beta - 5.0\right)e^{\frac{-12.5}{\lambda}} \tag{3}$$

The dynamic operation of the induction generators is governed by the swing equation given belowp [10]:

$$J\frac{d\omega}{dt} = T_m - T_e \tag{4}$$

Where T_m mechanical torque applied on the rotor of the associated wind turbine(Nm), J the wind turbine mechanical inertia(kg.m^2), ω the rotor speed of generator (rad/s), T_e is generator electro-magnetic torque(Nm).

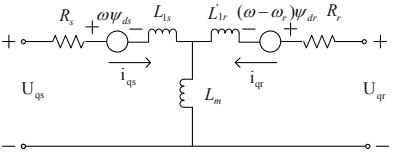
In (4), T_e is electro-magnetic torque developed in the induction generator at the given speed is in proportion with the square of the terminal voltage as follow:

$$T_e = KsU^2 \tag{5}$$

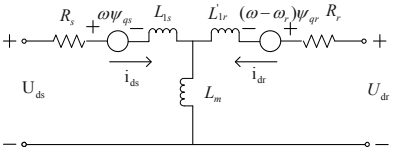
where K is constant value depending on the parameters of the machine and s is the machine slip.

3.2 Induction Generator Model

Fig.1 show the one line diagram of the implemented d-q equivalent circuit of induction generators[11-13] :



(a) q-axis equivalent circuit



(b) d-axis equivalent circuit

Fig. 2. d-q axis equivalent circuit

Detailed explanation of this equivalent circuit was clearly described as follow:

$$\begin{aligned}
U_{ds} &= \frac{d\psi_{ds}}{dt} - \omega_s\psi_{qs} + r_s i_{ds} \\
U_{qs} &= \frac{d\psi_{qs}}{d_t} - \omega_s\psi_{ds} + r_s i_{qs} \\
U_{dr} &= \frac{d\psi_{dr}}{dt} - (\omega_s - \omega_r)\psi_{qs} + r_r i_{dr} \\
U_{qr} &= \frac{d\psi_{qr}}{dt} - (\omega_s - \omega_r)\psi_{ds} + r_r i_{qr}
\end{aligned} \tag{6}$$

$$\begin{aligned}
\psi_{ds} &= (L_{1s} + L_m)i_{ds} + L_m i_{dr} \\
\psi_{qs} &= (L_{1s} + L_m)i_{qs} + L_m i_{qr} \\
\psi_{ds} &= (L_{1r} + L_m)i_{dr} + L_m i_{dr} \\
\psi_{ds} &= (L_{1r} + L_m)i_{qr} + L_m i_{qr}
\end{aligned} \tag{7}$$

where u_{ds} and u_{qs} represent thestator voltage, u_{dr} and u_{qr} the rotor voltage, ψ_{ds} and ψ_{qs} represent the stator flux linkage , ψ_{dr} and ψ_{qr} the rotor flux linkage, i_{ds} and i_{qs}、 the stator

current, idr、 i_{qr} the rotor current, ω_S the synchronous speed, the parameters of the machine L_s, r_s, L_r, r_r, and L_m represent the stator reactance, stator resistance, rotor reactance, rotor resistance, and mutual reactance respectively, s is the rotor slip.

3.3 Influence Factors of Short-Circuit

Fig.3 show the line diagram of the collect circuit for single phase-to-ground fault.

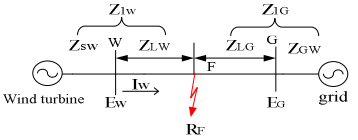


Fig. 3. Line diagram for single phase-to-ground fault

In power system, when occuring the fault, the fault current sequence component can be written as follow:

$$I_1^{(n)} = \frac{E_{1\Sigma}}{Z_{1\Sigma} + Z_\Delta^{(n)}} \tag{8}$$

For the single phase-to-ground fault, the sequence currents component of wind generators can be written:

$$\begin{aligned} &I_{1F}^{(1)} = I_{2F}^{(1)} = I_{3F}^{(1)} \\ &= \frac{E_W - I_W \times Z_{1W}}{Z_{1W} + Z_{2W} + Z_{0W} + Z_{1G} + Z_{2G} + Z_{0G} + 3R_F} \end{aligned} \tag{9}$$

where

$Z_{1W}=Z_{1SW}+Z_{1LW}$,$Z_{1G}=Z_{1SG}+Z_{1LG}$;
$Z_{2W}=Z_{2SW}+Z_{2LW}$,$Z_{2G}=Z_{2SG}+Z_{2LG}$;
$Z_{0W}=Z_{0SW}+Z_{0LW}$,$Z_{0G}=Z_{0SG}+Z_{0LG}$.

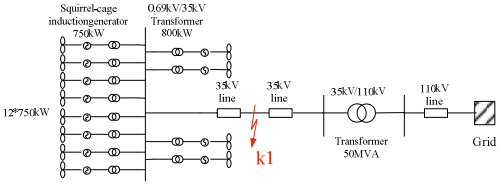
where I_W the prefault current, E_W the prefault voltage of wind generators, Z_{1SW}, Z_{2SW} and Z_{0SW} the positive, negative and zero sequence impedances of wind generators, and Z_{1LW}, Z_{2LW} and Z_{0LW} the positive, negative and zero line sequence impedance from bus W to fault location, Z_{1SG}, Z_{2SG} and Z_{0SG} the positive, negative and zero sequence impedances of system, Z_{1LG}, Z_{2LG} and Z_{0LG} the positive, negative and zero line sequence impedance from bus G to fault location.

When occur two phase fault and three phase fault, short-circuit current coule be described in(8), but the difference among three fault type is the Z_Δ.

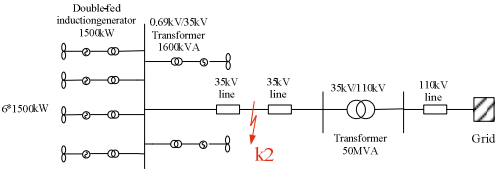
Due to the nature stochastic and intermittent of wind speed, wind turbine always withdrawal and even working in motor sometime, so the number of generator participating will change and the quivalent impedance of generators will change. We can derive the major effective factors that influence short-circuit characteristics from(9) such as wind turbine types, fault type, fault resistor and the number of generator participating in a time.

4 Wind Farm Model

The aimed study is carried out a wind farm in north china as simulation examlpe. The farm was structured with three collect circuits and all capacity is 49.5 MW, the collect circuit 320 contains 20 wind enerators(750kW) providing a total power of 15MW, the wind generators use 800kVA, the collect circuit 322 contains 10 wind generators (1500kW double fed induction generators) provide a total power of 15MW,the wind generators use 1600kVA tranfromers to set-up from the 0.69kV rated voltage to the 35kV collect circuit voltage, wind farm collector line length 5km, the collect circuit connected with wind farm substation to set-up voltage 110kV and connected 16km lines to 110 grid hebei province.



(a) collect circuit 320



(b) collect circuit 322

Fig. 4. Model diagram of the simulation system

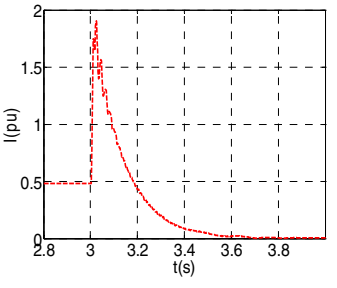
5 Simulation Results

5.1 Wind Turbine Type

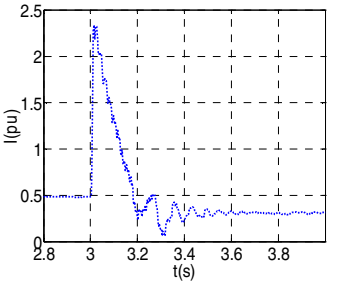
We assume that the active power of squirrel-cage induction generator and double-fed induction generator is similar. Figure.5 Show the simulation results of three phase-to-ground fault occuring in the place of 2.5km distance k1 and k2 away from the collect circuit 320 and 322, In this case, we operated the wind all generators with maximum active power output.

The simulation result show that the circuit current instanours value of squirrel cage induction generator and double-fed induction generator is 4 to 5 times rated current value, because squirrel cage induction generator do not have independent field windings to develop the required electro-magnetic field in the air gap of the

machine, therefore, when the three phase-to-ground happen, the terminal voltage of generator have decreased and the grid can not continue to provide exciting for wind turbine, The asynchronous short circuit current contribution drops from the initial value to zero in a few cycles. however DFIG with wound rotor induction generator, the short-circuit current gradually decay and due to the converter of DFIG, when occuring fault in grid, DFIG could supply continuous short-circuit current. The current instanours protection is infulenced by short-circuit current instanours value, when setting the relay protection of wind farm, we should consider the short-circuit current of different wind turbine, Current protection II is infulenced by trip time, so we should consider attenuation characteristic of wind generators.



(a) three-phase-to-ground fault current of squirrel cage induction generator

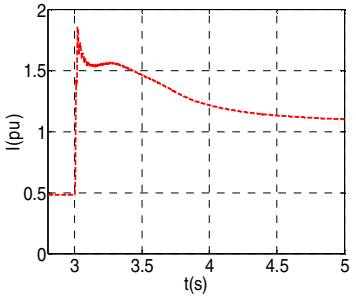


(b) three-phase-to-ground fault current of double-fed induction generator

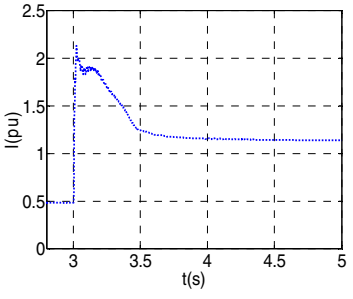
Fig. 5. Short-circuit current of Wind turbines response to ABC-G Fault

5.2 Fault Type

With A-B fault happening occuring in the place of 2.5km distance k1 and k2 away from the collect circuit 320 and 322, In this case, we operated the wind all generators with maximum active power output. The resulting two phase fault short-current current as shown in fig.6. Due to the c-phase without fault,network can continue provide excitation to the wind turbines from c-phase, therefore, asynchronous generators and double-fed induction generator can provide continuous short-circuit current when two-phase fault.



(a) short-circuit current of asynchronous generators



(b) short-circuit current of double-fed induction generators

Fig. 6. Short-circuit current of Wind turbines response to AB Fault

5.3 The Number of Wind Generators

Unlike conventional generation with thermal or hydro,wind farm generators are smaller in size, and a group of them collectively harness bulk power from a large area. The number of wind generators at a time will vary due to withdrawal of generator at high or low wind speed, the equivalent impedeance of the wind farm will change depending on the number of wind generators connected to the collect circuit at a time. Three situation are simulation where the sum capacity (wind generators participating at an instant) is 15MW, and in case2 the sum capacity is 12MW, the capacity is 9MW in case3. The short-circuit current curves of wind generator are described in Fig.7.

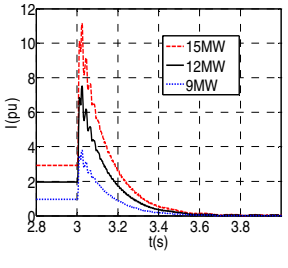


Fig. 7. Short-circuit current in different number of Asynchronous wind generators

The impact on short-circuit current is very obvious as result of different number of wind generators. The simulation results as shown in Fig.7, It is clear that the short-circuit current will increasing when the number of wind generator increasing, because the number of wind generator increasing, the equivalent impedeance of wind generators will descreaing, and according to the formula (8-9), the short-circuit current will increasing, therefore the current protection need to be updated in accordance with the number of wind generators participating in the wind fatm.

5.4 Fault Resistance

Fig.8 show the short-circuit current of wind turbine in different fault resistance. The simulation results show that due to the value of R_F is zero,causes the terminal voltage of generators drop to zero, so the short circuit current contribution drops from the initial value to zero in a few cycles; however,the value of R_F is 4Ω, the terminal voltage of generators drops to 0.2pu, so the wind generators can contribute continuous short-circuit current.

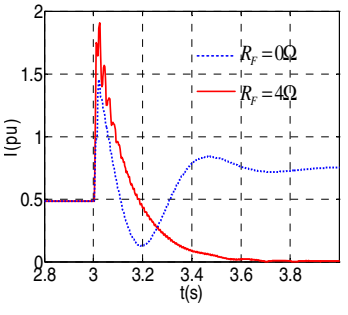


Fig. 8. Currents of wind generators response to differnt fault resistance at position k1

6 Conclusion

The objective of this paper aimed to investigate the short-circuit characteristics of asynchronous generators and double-fed induction generators, and dynamic model was established in PSCAD / EMTDC. the results showed that: when ocurring three-phase-ground fault, the asynchronous short circuit contribution drops from the initial value to zero and DFIG could contribute continuous short-circuit current; two types wind turbines can provide continuous short-circuit current when two-phase fault happening; and the short-circuit current will increasing when the number of wind generator increasing. With the capacity of wind power increasing in power system, we need to consider the short-circuit current contribine of wind farm when we setting the relay protection of wind farm. The simulation results show that the major factors influence the short-circuit characteristic are different wind turbine, fault type and the number of generator operating in a time, we should utilize the adaptive protection method to set the current relay protection using these influence factors.

References

[1] Lei, Y.: Study on Wind Fram Integration into Power System. Automation of Electric Power System 27(8), 84–89 (2003)
[2] Comech, M.P., Montanes, M.A., Garcia-Gracia, M.: Overcurrent Protection Behavior Before Wind Farm Contribution. IEEE, Los Alamitos (2008), 978-1-4244-1633-2
[3] Morren, J., de Haan, S.W.H.: Short-Circuit Current of Wind Turbine With Doubly Fed Induction Generator. Transactions on Energy Conversion 22(1) (March 2007)
[4] Pradhan, A.K., Jóos, G.: Adaptive Distance Relay Setting for Lines Connecting Wind Farms. Transaction Energy Conversion 22(1), 206–313 (2007)
[5] Kawady, T.A., Mansour, N., Taalab, A.E.-M.I.: Performance Evaluation of Conventional Protection Systems for Wind Farms. In: Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, pp. 1–7. IEEE, Los Alamitos (2008)
[6] Jang, S.I., Choi, J.H., Kim, J.W., Choi, D.M., Kim, K.H.: An Adaptive Relaying for the Protection of a Wind Farm Interconnected with Distribution Networks. IEEE, Los Alamitos (2003), 0-7803-81 10-6/03
[7] Haslam, S., Crossley, P., Jenkins, N.: Design and evaluation of a wind farm protection relay. Generation,Transmission and Distribution, IEEE Proceedings 146(1), 38–44 (1999)
[8] Sun, T., Chen, Z., Blaabjerg, F., Fellow: Voltage Recovery of Grid-Connected Wind Turbines After Short –circuit Fault. IEEE, Los Alamitos (2003), 0-7803-7906-3/0
[9] Aghaebrahimi, M.R., Amiri, M., Kamali Moghaddam, M.: A Short Circuit Study of an Induction Generator Wind Farm Considering Wind Speed Changes. In: 40th North American Power Symposium, NAPS 2008, pp. 1–6. IEEE, Los Alamitos (2008)
[10] López, J., Sanchis, P., Roboam, X., Marroyo, L.: Dynamic Behavior of the Doubly Fed Induction Generator During Three-Phase Voltage Dips. IEEE Transactions on Energy Conversion 22(3), 709–717 (2007)
[11] Yao, X., Liang, L., Xing, Z.: Dynamic Characteristic of The Drive Train of DFIG Wind Turbines During Grid Faults. In: Second International Conference on Intelligent Computation Technology and Automation, pp. 503–506 (2009)
[12] Samaan, N., Zavadil, R., Charles Smith, J., Conto, J.: Modeling of Wind Power Plants for Short Circuit Analysis in the Transmission Network. IEEE, Los Alamitos (2008), 978 –1 – 4244 –1904 -3/08
[13] Petru, T., Thiringer, T.: Modeling of Wind Turbines for Power System Studies. IEEE Transactions on Power System 17(4) (November 2002)

Appendix

Table 1. 110kV Line Parameters

	Value
Positive sequence resistance	0.132Ω/km
Positive sequence reactance	0.401Ω/km
Positive sequence susceptance	2.85×10-6s/km
Zero sequence resistance	0.396Ω/km
Zero sequence reactance	1.203Ω/km
Zero sequence susceptance	-

Table 2. 35kV Line Parameters

	Value
Positive sequence resistance	0.132Ω/km
Positive sequence reactance	0.357Ω/km
Positive sequence susceptance	3.21×10-6s/km
Zero sequence resistance	0.396Ω/km
Zero sequence reactance	1.071Ω/km
Zero sequence susceptance	-

Table 3. 110/35 kV Transformer Parameters

	Units	Value
Rated power	MVA	50
Ratio	kV	110/35
Connection	--	YND11
Leakage reactance	Pu	0.1

Table 4. 35/0.69 kV Transformer Parameters

	Units	Value
Rated power	MVA	1.6
Ratio	kV	35/0.69
Connection	--	Dyn11
Leakage reactance	Pu	0.01

Table 5. 750kW squirrel-cage induction generator Parameters

		Value
Stator resistance	Rs	0.0053pu
Stator unsatured leakage reactance	Xs	0.1060pu
Unsatured magnetizing reactance	Xm	4.0209pu
Rotor unsatured reactance	Rr	0.0070pu
Rotor resistance	Xr	0.1256pu

Table 6. 1500kW double-fed induction generator parameters

		Value
Stator resistance	Rs	0.0058pu
Stator unsatured leakage reactance	Xs	0.097pu
Unsatured magnetizing reactance	Xm	5.126pu
Rotor unsatured reactance	Rr	0.0066pu
Rotor resistance	Xr	0.1534pu

Optimization Design of Lifetime Distribution in Power Diode with Fast and Soft Recovery

Tao An, Yi Xing, and Cailin Wang

Xi'an University of Technology, 710048, China

Abstract. The fast and soft recovery is required for power pin diodes supporting the use of IGBT. In this paper, the effects of the local low lifetime and the overall lifetime on device characteristics have been simulated and analyzed by using of software. Based on their effects on the reverse recovery time t_{rr}, the forward voltage drop V_F and the reverse leakage current I_R of power pin diode, those parameters, such as the base location L, lifetime τ^*_P and allover lifetime τ_P In local low lifetime control technology and the overall lifetime τ_P, are studied and the optimal base lifetime distribution parameters are obtained. Those results have practical reference value for research and manufacturing of diodes with fast and soft recovery.

Keywords: local low lifetime, fast and soft recovery, reverse recovery time.

1 Introduction

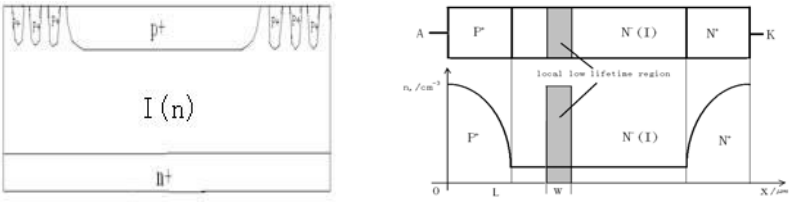
With the Power Electronics towards high-power, high-frequency-based and modular development, the development of the indispensable fast soft-recovery diodes (FRD), matching with the IGBT and power MOSFET high-frequency electronic devices, is urgent and has important practical significance. In recent years, a new lifetime control technology (Local lifetime control technology) has attracts more attention[1]. To form a sheet local low-lifetime region with certain thickness, the high-density recombination centers are introduced on the position with a very small distance to the pn junction. The position is perpendicular to the direction of the pn junction plane. Local low lifetime regions have great impact on performance parameters of FRD and are advantage to realizing the trade-off among of those parameters.

2 PIN Power Diode Basic Structure and Working Principle

2.1 The Basic Structure

The structure of PIN diode is composed of P^+ region, N^+ region and I region. P^+ region and N^+ region are high doping concentration. I region is N-type high-resistivity material and its doping concentration is very low. The basic structural model and the impurities distribution were shown in Figure 1 (a) and Figure 1 (b), Shaded area is the local low lifetime region.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 37–43.
springerlink.com

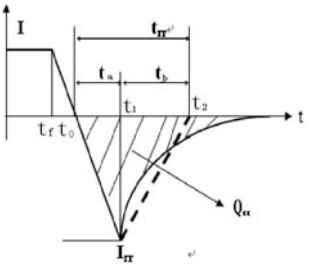


(a) (b)

Fig. 1. The basic structure and impurity distribution of PIN diode

2.2 Theoretical Analysis

When the P-I-N power diode is added forward voltage, its forward voltage drop is low. When the diode is added reverse voltage, its reverse voltage is high[2]. When the voltage added on power diode changes from forward voltage to reverse voltage, the diode can not be cut-off immediately[3][4]. The reason is that many carriers stored in I region can not disappear immediately. It needs period of time that these carriers are extracted or recombined completely and this is the reverse recovery time t_{rr}. In order to protect the device in the main circuit, the diode reverse recovery time must be very short. The reverse recovery characteristics of diodes are shown in Figure 2.



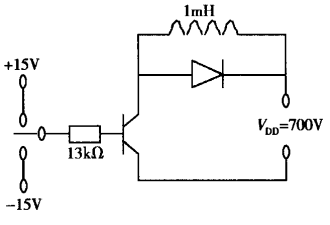


Fig. 2. Reverse recovery characteristics of diodes **Fig. 3.** Test circuit of reverse recovery time

The reverse recovery time of diode is consisted of two parts[5][6]. One is storage time t_a and the other is recombination time t_b. The diode is added reverse voltage V_R at $t = t_f$. Time period from t_0 to t_1 is t_a. Time period from t_1 to t_2 is recombination time t_b. The reverse-recovery softness factor S is described by $S = t_b/t_a$. Namely, the ratio of recombination time and storage time

Based on the study of PIN lifetime control technology of power diode, it can be found that the reverse recovery time of diode can be improved significantly, while its forward voltage drop and reverse leakage current are almost no changed, as shown in Figure 1 (b), by introducing high-density recombination centers in I region.

3 Simulation Analysis

3.1 Model Building

The purpose of this paper is design the diodes supporting the use of IGBT and its reverse voltage is 1200V. Its technical parameters are shown in Table 1. The structure of the device is shown in Figure 1 (a). The specific structural parameters as follows, the total epitaxial wafer thickness H is 150 μm, epitaxial thickness h is 30 μm (thickness of N^+ region), the concentration n_{N^+} is $5\times10^{19}/cm^3$ (concentration of N^+ region), the total area is $9000\times9000\mu m^2$, P^+ area is $8600\times8600\mu m^2$, depth of P^+ is 30 μm, the concentration of P^+ region n_{p^+} is $5\times10^9/\mu m^3$, the thickness of N^- region W is 90 μm, the concentration of N^- region n_{N^-} is $1\times10^{14}/\mu m^3$, the width of Field Limiting Ring are 15 μm, 17 μm and 19 μm, the rings spacing are 50 μm, 48 μm and 46 μm, respectively.

The test circuit of reverse recovery time is shown in Figure 3. Simulated conditions are that forward current I_F equals to 150A, reverse voltage V_R is 600V and reverse current fall rate di/dt is 1650 $A/\mu s$ [6].

Table 1. Main technical parameters

Parameters	Qualification	Units	Notes
Rated Forward Average Current I_F	150	A	
Repeat peak reverse voltage V_R	1200	V	
Reverse leakage current I_R	<27	μA	Reverse Voltage V_R=1200V
Forward voltage drop V_F	<1.6	V	Forward CurrentI_F=150A
Reverse recovery current I_{RRM}	<187	A	di/dt=1650A/ μs
Reverse Recovery Time t_{rr}	<400	ns	

3.2 Results and Discussion

3.2.1 Simulation and Analysis of Location of Low-Lifetime Region

Low-lifetime region width W, lifetime τ_P^* and the overall lifetime of the device τ_P are 10 μm, 1×10^{-8} s and 3×10^{-6} s, respectively. The value of other parameters is given in the previous models. The effects of the low lifetime position L on the recovery time t_{rr} is simulated by changing the position of L from 30 μm to 120 μm (moving low lifetime region from near the P^+ area to the N^+ area, each additional 10 μm). The simulation results are shown in Figure 4.

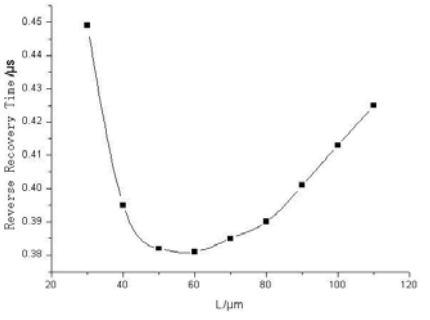


Fig. 4. The effect of the location L on the reverse recovery time trr

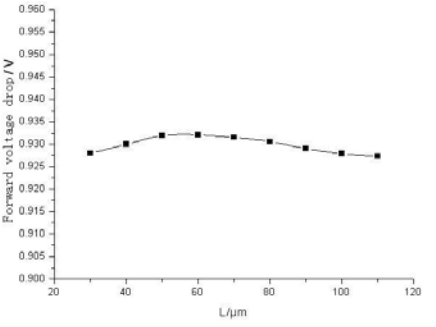


Fig. 5. The effect of the location L on the forward voltage drop V_F

With the increasing of the location L from 30 μm to 60 μm the reverse recovery time t_{rr} decreases from 450ns to 380ns. When the location L is 60 μm (center of I region), the reverse recovery time t_{rr} is minimum. With the increasing of the location L from 60 μm to 120 μm, the reverse recovery time t_{rr} increases from 380ns to 430ns. When the diodes are cut-off, holes and electrons near the P^+ region and N^+ region can be extracted rapidly, while the carriers in the middle of I region can not be extracted but recombined, which needs a long time. Therefore, the reverse recovery time t_{rr} is longer. When the location L is moved from 30 μm to 60 μm, the recombination contribution is more notable and the reverse recovery time decreases. When the location L is far from the center of I region (L>60 μm), the recombination contribution is not remarkable and the reverse recovery time increases. When the location L is in the center of I region (L = 60 μm), large number of holes and electrons without extracting from I region are recombined through low lifetime region and the reverse recovery time shortest.

It can be seen that, from Figure 5, the forward voltage drop V_F is almost not affected by the location L of low-lifetime and the value is only 0.007V. In the local low-lifetime control technology, the thickness of high-density recombination centers is very thin so its impact on forward voltage drop is not notable. In addition, the effect on leakage current is so small that it can be ignored.

3.2.2 Simulation and Analysis of Local Minority Carrier Lifetime

The location of the low lifetime region L is 60 μm and other parameters is unchanged. Minority carrier lifetime τ_P^* in low lifetime region is changed from 1×10^{-5} s to 1×10^{-9} s. The effects of different minority carrier lifetime on recovery time t_{rr} and forward voltage drop have been simulated and analyzed. The results are shown in Figure6, Figure7 and Figure8.

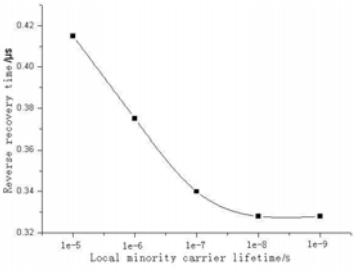


Fig. 6. The effect of the lifetime τ_P^* on the reverse recovery time t_{rr}

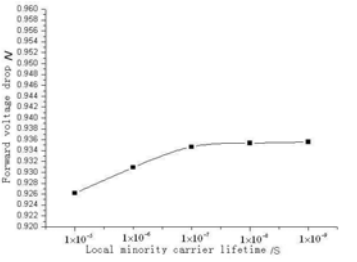


Fig. 7. The effect of the lifetime τ_P^* on the forward voltage drop V_F

Figure 6 shows that the reverse recovery time t_{rr} is reduced from 420ns to 340ns with the lifetime τ_P^* being decreased from 1×10^{-5} s to 1×10^{-7} s. When lifetime is less than 1×10^{-7} s, th of the forward voltage drop V_F is only 0.009V. The effect of lifetime on the forward voltage drop is very little and the local low-lifetime τ_P^* is identified as 1×10^{-8} s.

3.2.3 Simulation and Analysis of the Overall Minority Carrier Lifetime

The location of the low lifetime region L is 60 μm, the lifetime of local low lifetime region τ_P^* is 1×10^{-8} s and other parameters are unchanged. The effects of the overall minority carrier lifetime on the forward voltage drop V_F, the reverse recovery time t_{rr} and the reverse leakage current I_R have been simulated. When the lifetime is decreased from 1×10^{-5} s to 1×10^{-9} s, The forward voltage drop V_F is increased from 0.930V to 0.931V and the change is not obvious as shown in Figure 8. When lifetime τ_P is reduced from 1×10^{-6} s to 1×10^{-8} s, the forward voltage drop V_F is increased from 0.931V to 0.946V and the increase speed is rapidly. When lifetime τ_P is reduced to 1×10^{-8} s, the forward voltage drop V_F is saturated.e reverse recovery time t_{rr} is reduced with a slower speed. When lifetime is reduced to 1×10^{-8} s, t_{rr} is saturated and no longer decreased with the reducing of τ_P^*. When lifetime is less than 1×10^{-8} s, the minority carriers in the center of I region have been able to recombine completely and the lifetime of minority carriers is no longer the major factor determining t_{rr}. Figure7 shows that when the lifetime τ_P^* is changed from 1×10^{-5} s to 1×10^{-9} s, the variation

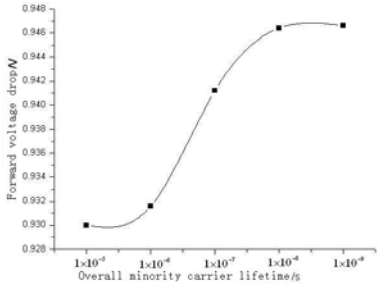


Fig. 8. The effect of the τ_P on the forward voltage drop V_F

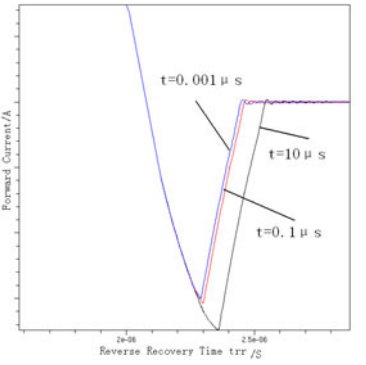


Fig. 9. The effect of the τ_P on the reverse recovery time t_{rr}

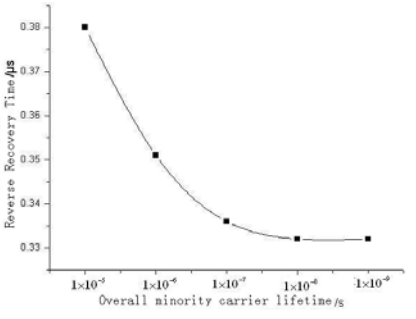


Fig. 10. The effect of the lifetime τ_P on the reverse recovery time t_{rr}

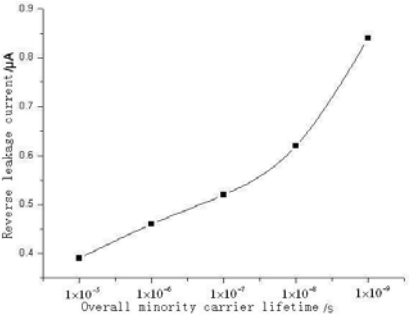


Fig. 11. The effect of the lifetime τ_P on the reverse leakage current I_R

From Figure9, Figure10 and Figure 11, it can be seen that the reverse recovery time t_{rr} is decreased with the decline of minority carrier lifetime τ_P and the decrease of the reverse recovery time is sharp. When the τ_P is declined to 1×10^{-7} s, the reverse recovery time t_{rr} is saturated. The leakage current I_R is increased rapidly with the decline of minority carrier lifetime τ_P. When the lifetime τ_P is declined to 1×10^{-7} s, the leakage current I_R continues increase. Considering the relationship between the forward voltage drop V_F and the leakage current I_R, the overall minority carrier lifetime τ_P^* is identified as 1×10^{-7} s.

In conclusion, based on the comprehensive consideration of the forward voltage drop V_F, the reverse leakage current I_R and the reverse recovery time t_{rr}, the optimal parameters of diode are as follows, (1) the location of low lifetime region L is 60μm,

(2)the minority carrier lifetime of low lifetime region τ_P^* is 10ns, (3) the minority carrier lifetime of allover region τ_P is 100ns. Based on those optimal parameters, the simulation results indicate that the forward voltage drop V_F is 0.935V, the reverse leakage current I_R is 1.1μA and the reverse recovery time t_{rr} is 334ns. Various performance parameters achieve design requirements.

4 Conclusion

In this paper, the effects of low-lifetime region location, minority carrier lifetime and the overall lifetime on forward voltage drop V_F, reverse leakage current I_R and reverse recovery time t_{rr} are simulated comprehensively and systematically. The optimal parameters are obtained based on those simulation results. The results indicate that, the reverse recovery time can be reduced effectively by introducing local low-life region in the base region of $p^+n^-n^+$ fast soft-recovery power diode. The reverse recovery time of diodes is related to the minority carrier lifetime of low-life region of, while the leakage current and forward voltage drop are almost not affected. The reverse recovery time is also related to the position of low-lifetime region. When the low-lifetime region locates in the center of the base region, the reverse recovery time is minimal. Those results provide the instructional methods and conclusions for the optimum design of base region lifetime distribution in fast soft-recovery power diode.

References

1. Vobecky, J., Hazdra, P.: High-Power P-i-N Diode With the Local Lifetime Control Based on the Proximity Gettering of Platinum. IEEE Electron Device Letters 23(7), 392–394 (2002)
2. Baliga, B.J.: Modern power devices, p. 41. Wiley, New York (1987)
3. Napoli, E., Strollo, A.G., Spirito, P.: Numerical analysis of localcontrol for high-speed low-loss P-i-N diode design. IEEE Trans. Power Electron 14(4), 615 (1999)
4. Wolley, E.D., Bevacqua, S.F.: High Speed, Soft Recovery, Epitaxial Diodes for Power Inverter Circuits. In: IEEE Industrial Application Society Meeting Digest, pp. 797–800 (1981)
5. Hazdra, P., Vobecky, J., Brand, K.: Optimum lifetime structure insilicon power diodes by means of various irradiation tech-niques. Nuclear Instruments and Methods in Physics Re-search B 186, 414 (2002)
6. Ma, C.L., Lauritzen, P.O.: A Simple Power Diode Model with Forward and Reverse Recovery. IEEE Transaction on Power Electron, 553–559 (1993)

Research on Double-Base Regions High-Power Fast Recovery Diode

Tao An[1], Yang Li[1], and Qitang Yin[2]

[1] Xi'an University of Technology, 710048, China
[2] Beijing Chunshu Rectifier Co, Ltd 100040, China

Abstract. The paper introduces the $\eta=W_I/X_m=0.25$ mathematical models and adopts mathematical model $V_B=94\rho_n^{0.7}$ to design structure parameters of high power double-base P^+PINN^+ structured fast soft recovery diodes. Platinum doping and electron irradiating technologies are used to mutually control the base minority carrier lifetime and distrbution, the design method is used to optimize the structure parameter of ZKR300A/ 2500V. The design parameters were tested and verified through experiments, which proved the parameters of diodes meet the designed target and achieved the level of similar products in abroad countries. The results prove that the design method and the selected parameters are correct,lifetime control is effective.

Keywords: FRD, P^+PINN^+ structure, soft recovery, minority carrier life.

1 Introduction

At present, new high-power electronic devices such as GTO, IGBT, IEGT, IGCT are widely applied in power electronic field. FRD is indispensable as important "partner" chip must meet application requirements. Because the reverse characteristic of ordinary quick recovery diodes is "hard", it is easy to damage power electronic devices. We adopt P^+PINN^+ in replace of ordinary PN and P^+IN^+, it has large forward current I_{FM},high voltage U_{RRM}, low-state voltage drop V_{FM}, and by controlling the base region of minority carrier's life and impurities distribution accurately,it can decrease reverse recovery charge Q_{rr}, reverse recovery time t_{rr} and soft characteristic S. The research of the device is very important for promoting the development of high-frequency power electronic circuits[1][2][3].

2 Basic Structure and Principles

2.1 The Basic Structure

P^+PINN^+ diode is the combination of a PN junction and two high-low junctions (P^+P and NN^+) which is made of high P^+ and N^+ layer, N and P layer of subsurface on N^- type semiconductor substrate. P^+ region is the anode, N^+ region is the cathode, the base region I is made of light doping N^- and higher doping N region, and N region is the buffer base region or buffer layer. The structure is shown in Figure 1.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 45–51.
springerlink.com

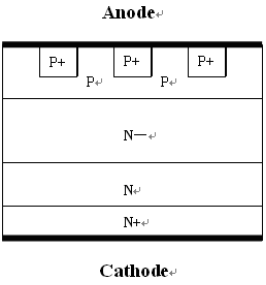
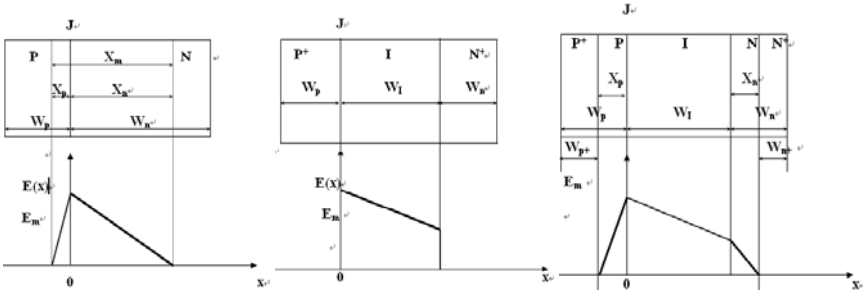


Fig. 1. Double-Base FRD

2.2 Theoretical Analysis

The structures and electric field distributions of P^+PINN^+ fast soft-recovery diode, ordinary rectifier diode of PN and fast rectifier diode of P^+IN^+ are shown in figure 2 (a),(b),(c). P^+PINN^+ owns both the advantages of $P^+IN^+$and PN. When reversing, owing to base region I, the space charge expand fully leading to increasing the breakdown voltage U_{RRM}. Because of N region, there is a restriction on space charge when it expands into N region. It not only meets the breakdown voltage, but also reduces the depth of region I. While the device is forward turned on, the depth of region I is reduced. Since P^+ and N^+ areas inject carriers into the I area, an increase of carrier concentration I area, the role of enhanced conductivity modulation to the low forward voltage drop V_{fp}; because the depth of region I is reduced, the storage of charge is decreased, thus, reverse recovery time t_{rr} is shorter. When the device is turned off, charge in region N are complex and disappeared, which makes the complex time longer, and the recovery characteristic is softened.



(a) (b) (c)

Fig. 2. Diode structure and the electric field distribution

When on-state diode applied a reserve voltage suddenly, a lot of minority carriers are stored in pn junction, it will takes some time to compound those minority carriers until the diode is cut off, which is called the reverse recovery, the time is called reverse recovery time t_{rr}, it contains two processes of storage and recovery. After the time of storage t_1,the minority carriers in charge space are extracted, and reverse current has reached the maximum value I_{rr}.When $t > t_1$, space charge begins to form.

Because the minority carriers are compounded and disappeared, reverse recovery current begins falling to zero gradually, the diode recover to reverse state and can bear high reverse voltage, this time is called recovery time t_2. During t_2, if the reverse recovery current falls quickly, owing to line inductance, there will be a high burr voltage, even a strong shock, which will interfere the circuit. As a good FRD, it has short t_{rr}, lower -di/dt, and a low burr voltage. Generally, the soft coefficient $S= t_2/ t_1$ is used to indicate the softness of reverse recovery characteristic. Bigger the ratio is, better the soft characteristic is. In the same circuit condition, lower the overshoot voltage is, smaller the ratio is, and worse the softness is. The reverse recovery characteristic of diode is shown in figure 3.

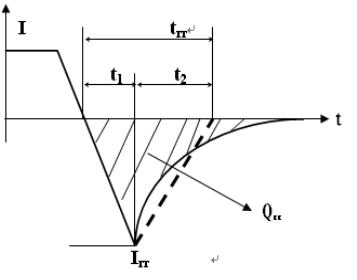


Fig. 3. Diode reverse recovery characteristics

3 Design of Structural Parameters and Technology

This paper adopts high power ZKR round core ZKR300A/2500V as an example, calculating and selecting the structural parameters and process conditions.

3.1 Structural Parameters

1) the main technical parameters of High power ZKR round core

Table 1. Main technical parameters

Parameters	Qualification	Units	Notes
Rate average forward current IF	300/500	A	
Reverse repetitive peak voltage V_{RRM}	>2500	V	
Reverse non-repetitive peak voltage	>2600	V	
Reverse repetitive peak current I_{RRM}	≦50	mA	
Forward on-state voltage V_{FM}	≦2.5	V	
Forward instantaneous voltage V_{fp}	≦50	V	d_i/d_t=500A/μs
Reverse recovery time t_{rr}	≦3	μs	$-d_i/d_t$=100 A/μs
Reverse recovery charge Q_{rr}	≦600	μc	
Reverse recovery softness $S(t_2/ t_1)$	>0.5		

2) Structural parameters

(1) Vertical design

The empirical formula $V_B = 94\ \rho_n^{0.7}$ is used in the design, the utilization factor $\eta = \frac{W_I}{Xm}$ is introduced to optimize design ideas. Both of all make sure the V_B is constant, and make the intrinsic region thinner, thus, the high current characteristic of the high-power fast soft recovery diodes under high voltage and frequency is increased, and switching characteristics are good.

According to the formula of avalanche breakdown voltage[4]:

$$V_B = 94\rho_n^{0.7} . \tag{1}$$

V_B:the avalanche breakdown voltage; ρ_n : the base resistance.

According to the empirical formula of the max width of region and base region resistivity:

$$\eta = \frac{W_I}{Xm} . \tag{2}$$

$$X_m = 4.99\rho_n^{0.85} . \tag{3}$$

$$V_B = V_{B0}/(2\eta \sim \eta^2) . \tag{4}$$

W_I : the thickness of based region; X_m : the space charge region broadening maximum width; η : the ratio of coefficients.

First of all, according to formula (2), η = 0.25 ~ 0.5. while making sure V_B constant, we can adopt the thinner intrinsic region and select the smallest coefficient η, which can increase the switch characteristics and the high-current characteristics of high-power fast soft recovery diode when they are under the high frequency and voltage. From formula (4) and (2), $\rho_n \approx 190 \sim 400\ \Omega\cdot$cm, and according to the formula (3), X_m = 431 ~ 812 μm,so, $W_I = \eta X_m$ = 203 ~ 215 μm ≈ 220 μm; then, adopting W_p = 70 μm, W_n = 60 μm, W_p '= 120 μm, finally, silicon thickness $d = W_p + W_I + W_n + W_p' + \Delta W$ = 500 μm.

In the premise of the lower forward voltage, we make the P^+ and N^+ region thinner, reduce the recombination effects of the P^+ and N^+ storage carriers, and ensure the surface diffusion concentration : $1.0 \sim 10 \times 10^{20}/cm^2$.

(2) Horizontal structure design

The diameter of silicon wafers are selected to determine the width and angle of the angle lap according to the current capacity rating and voltage rating.

According to the empirical formula of the on-state current and on-state current density:

$$I_F = \frac{1}{4}\pi D^2 J_F . \tag{5}$$

J_F : the forward on-state current density, D_e : the chip diameter.

According to the empirical parameters, on-state current density generally range 50 ~ 120A/cm^2, select J_F = 110 A/cm^2. Experience (5)-style get-chip 2.4 cm in diameter. Because of dual-mill grinding angle, rubbing off the width of 3 mm, we adopt the chip diameter ϕ30mm.

Above all, taking the on-state power loss, heat dissipation, the process, cost production and other factors into account, we take the optimization parameter : ρ_n : 200 ~ 400Ω·cm; W_I : 220μm; W_p : 70μm; W_n : 60μm; total thickness of silicon wafers d = 500 μm; wafer diameter ϕ30mm.

3.2 Controlling Theory of Impurity Concentration Distribution

It can improve the reverse recovery characteristics, through controlling the distribution of the impurities. The traditional P^+IN^+ diode's thickness is thick,and life is short, so base storage carriers can diffuse into P^+ and N^+ region and complex in the recovery time. If the base region is not too thick and the minority carrier life is longer, it is good for reverse recovery. If life is short, it will accelerate the reverse recovery, resulting in hard recovery characteristics. So it could add a longer life and lower concentration P and N region, separately, between P^+ region and I, N^+ region and I. In the premise of enabling P^+ and N^+ region thinner,and reduce the base storage carrier combination effects of the P^+ and N^+, it is good for soft recovery characteristics and can ensure the forward voltage drop still lower.

3.3 Minority Carrier Lifetime and Distribution of Reasonable Control

Since changes in the minority carrier lifetime and minority carrier lifetime control method directly affect the fast and soft recovery diode reverse recovery switching frequency characteristics and softness, the minority carrier lifetime control technology is very important. The so-called minority carrier lifetime control methods that use different techniques to increase the base area of the recombination centers, such as increased use of deep level impurities (expanded gold, platinum, etc.) or have some defect levels (electron irradiation) to form a composite center. Thereby increasing the non-equilibrium carrier recombination rate, reducing the base minority carrier lifetime and to the base minority carrier lifetime in different regions in different areas, so as to minimize off-time, increase the recovery soft, increase switching speed.

No matter adopting the expansion of gold, platinum or electron irradiation, which have their own advantages and disadvantages, will reduce the lifetime of the base region minority carrier greatly. High-energy is used at room temperature or low temperature after the end of the chip processing, which is simple, precise control and consistent.The platinum expansion technique is combined with irradiation[5,6]. After forming P^+PINN^+ structure, platinum is expanded in short time, in order to reduce the minority carrier lifetime in the vicinity of PN junction, then electron irradiation is adopted to reduce the minority carrier lifetime of the base region away from the PN junction appropriately to ensure that the reverse recovery time was short and soft. The platinum concentration and the distribution of minority carrier lifetime are shown in Figure 4.

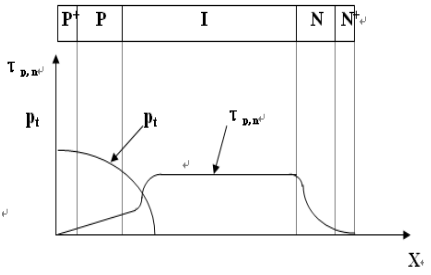


Fig. 4. Pt concentration and Minority carrier lifetime distribution

4 Analysis of Experimental Results

A design and study of process technology are made to the FRD chip of 75A-150A/1200 ~ 1700V, and the trial of the chip. The chip's major testing technical indicators such as shown in Table 2, provides the international well-known enterprise ABB similar product parameters.

Table 2. FRD Performance compared with similar foreign products

Num	Style	Company	Technical indicators					
			V_{RRM} (V)	I_{FAV} (A)	V_{FM} (V)	t_{rr} (μs)	Q_{rr} (μc)	S (t_1/t_2)
1	ZKR75-12	Chunshu	1200	75	≦2.3	0.42	20	
2	5SLX12F1200	ABB	1200	75	≦2.3	0.4	13	
1	ZKR150-17	Chunshu	1200	150	≦2.0	0.6	68	
2	5SLX12K1711	ABB	1200	150	≦2.0	0.66	73	>0.5
1	ZKR300-25	Chunshu	2500	300	≦3.0	3	400	
2	5SDF05D2505	ABB	2500	300	≦2.3	3.6	≦840	
1	ZKR500-25	Chunshu	2500	500	≦3.0	3.5	≦500	
2	5SDF2501	ABB	2500	500	≦1.9	5.7	≦500	

From Table 2 it can be seen that the actual test specific data, models were ZKR75-12, ZKR150-17, ZKR300-25 and ZKR500-25 repetitive peak reverse voltage V_{RRM} for the 1200V, 2500V, rated average forward current I_F for the 75A and 150A. Various types of chips V_{FM}, t_{rr}, Q_{rr} and soft characteristics of the test results meet the design specifications and design requirements. Same chip has been fully achieved the level of internationally famous enterprises ABB parameters similar products. The chip is currently used on self-produced modules. At the same time, products have been gradually applied to the Beijing subway and other devices and have good function. Its capability has reached the level of similar products abroad.

5 Conclusion

In this paper, $V_B = 94\rho_n^{0.7}$ and $\eta = W_I / X_m = 0.25$ for double mathematical model for fast soft recovery FRD use of novel high resistance thin base P^+PINN^+ various structural parameters were optimized; and expanded through the use of chiptwo platinum and electron irradiation lifetime control technology, jointly controlled base region minority carrier lifetime and distribution, which an be designed and manufactured with high current I_{FM}, high pressure U_{RRM}, low-state voltage drop V_{FM}, a small leakage current I_{RRM}, short recovery time t_{rr} and soft characteristics of high-power high-performance index P^+PINN^+ diode, the indicators and performance have reached the level of similar foreign products. This indicates that the design method and the selection of the parameters are correct, reasonable, and realistic, fully can be used in practical design. The P^+PINN^+ structure in this paper for the design and manufacture of diodes provides an important guide and reference for new methods and new technology.

References

1. Zhang, H., Zhang, B.: The current development of fast soft-recovery diode. World Electronic Device 4, 29–33 (2001)
2. Zhang, Q., Zhang, B., Wang, X.: New high-power fast soft recovery diode. Semiconductor Technology 23(1), 30–34 (1998)
3. Zhang, H., Zhang, B.: Overview of high-power fast soft recovery diode. Semiconductor Information 38(3), 1–8 (2001)
4. Faur, M., Maria, F., Chandra, G., et al.: Defect density reduction of n + p and p + n Inp structures fabricated by closed ampoule thermal diffusion. In: Indium Phosphate and Related Materials Fourth International Conference, pp. 322–325 (1992)
5. Jia, Y., Wang, J., Kang, B.: High-voltage Power Diode with Lifetime Control of Localized Platinum. Progress in Solid State Electronics 24(4), 422–426 (2004)
6. Nielsen, P., Blaabjerg, F., Pedersen, J.K.: Space Vector Modulated Matrix Converter with Minimized Number of Switching's and A Feed forward Compensation of Input Voltage Unbalance. In: Proc. of PEDES 1996, vol. (2), pp. 833–839 (1996)

Analysis and Design on Drive Circuit of Integrated Gate Commutated Thyristor

Tao An[1], Xiaowen Zhang[1], Ruliang Zhang[1], and Qitang Yin[2]

[1] Xi'an University of Technology, 710048, China
[2] Beijing Chunshu Rectifier Co. Ltd 100040, China

Abstract. In this paper, a integrated gate drive circuit of reverse-conducting GCT with 1100A/4500V is analyzed and designed. Firstly, the turn-on and turn-off circuit is analyzed and designed according to the requirment of device drive, and using PSPICE to make simulation, moreover, the influence of inductance L_1 and L_2, capacitor C_{off} and switching Q_{off} parameters to the gate drive current were analyzed, as well as the analysis and optimization of these parameters. And then, for FPGA, as the core control chip of the control circuit, using the software QuartusΠ to writing code and completing the function simulation. Finally, get a complete GCT integrated gate drive circuit with the turn-on and turn-off circuit and the logic control circuit, and provide a design method for the development of driving circuit design of integrated GCT.

Keywords: GCT, integrated gate drive circuit, FPGA.

1 Introduction

The integrated gate commutated thyristor (IGCT) is a power switch device which directly integrate gate commutated thyristor (GCT) with gate drive circuit with big current capacity[1], high switching frequency[2], low conduction drop and good EMI characteristics[3]. Now it has begun to be widely used in many high-power fields. Gate drive circuit is a important part of IGCT and has direct influence on performance of GCT[4]. Therefore, the research and development of its high-performance gate drive circuit has broadly practical significance to further expand the application range of IGCT.

Based on the working principle of GCT in the paper, a design to gate drive circuit of reverse-conducting GCT with 1100A/4500V is provided, and simulation analysis and parameter optimization are made while the core control chip with FPGA as control circiut to achieve effective control of GCT.Finally a complete drive circuit of integrated gate commutated thyristor is got, which includse the switch circuit and the logic control circuit.

2 Drive Principle of GCT

Based on the structure and working principle of GCT[5], in order to control the turn-on and turn-off of GCT effectively, gate drive current amplitude and the size of the

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 53–61.
springerlink.com

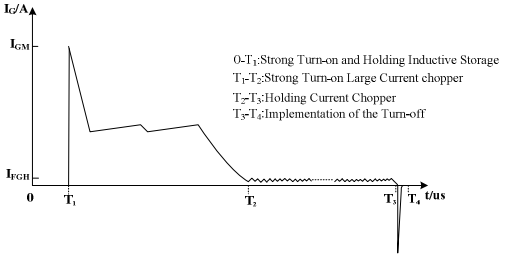


Fig. 1. Gate drive current waveform

rate of current rise di/dt have strict demand, the gate drive current waveform is shown in Fig 1[6][7].

Turn-on Process: When gate drive circuit of GCT has received the opening signal, gate drive circiut make the maxium of gate drive curent I_G, GCT instantly is turned on during the time 0 - T_1, at the time, gate drive circuit remains within the time T_1-T_2 to provides a high gate drive current to GCT until the end of T_2, which GCT is fully turned on; After GCT's turn-on, in order to avoid the GCT automatically turning off for the anode current of GCT is less than the holding current, gate drive circuit continue to provide a gate holding current(I_{FGH}) of the amplitude of 2A to 6A to GCT in the time T_2-T_3 , and the GCT is to be on-state at the time.

Turn-off Process: When gate drive circuit of GCT has received the turn-off signal, the gate drive unit cut off the gate drive current quickly ,and adds reverse current to the gate-cathode ,then GCT is turned off within the time T_3-T_4.

3 Integrated Gate Drive Circiut of GCT

3.1 Integrated Gate Circuit Diagram

Integrated gate drive unit of GCT is made of the fiber optic transceivers, the power circuit, the logic circuit,the turn-on and turn-off circuit, the state display parts and so on, which is shown in Fig 2.

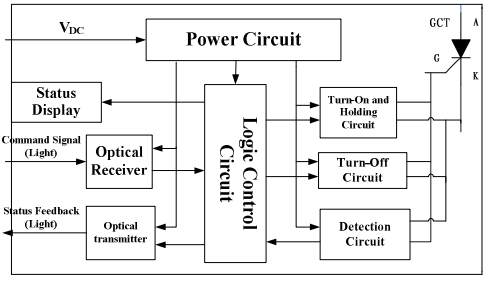


Fig. 2. Integrated gate circuit of GCT diagram

When the fiber receiver has received the light open command signal, the light signal is transformed into electrical signal, and input it to the logic control circuit and judge it, if it meets the open condition,the logic control circuit sends out the open sequence pulses to the turn-on and holding circuit, which produces gate drive current to achieve turn-on of GCT; After GCT's turn-on, the turn-on and holding circuit continue to provide the gate holding current, which make the GCT on-state.The gate state detection circuit detected the turn-on signal of GCT, and transmits it to the logic control circuit, then the logic control circuit transmits feedback signal to the optic transmitter, which is sent to external control system while the state display circuit displays the working state of GCT for the external work staff; If it does not meet the openging condition, the logic control circuit does not response to turn-on demand, but continue to wait for next turn-on command signal, and judge it again until the turn-on command signal is effective. The turn-off of GCT is similar to its turn-on, just in the turn-off process, the operation signal is turn-off signal, and the logic control circuit controls the turn-off circuit.

3.2 The Turn-On and Turn-Off Circuit

3.2.1 Working Principle

The turn-on and turn-off principle diagram is shown in Fig 3. The turn-on and turn-off circuit of integrated gate drive circuit of GCT is made of two buck circuit[8], which mainly complete the turn-on and turn-off process of GCT, the turn-on process includes strong opening large current chopping and holding opening small current chopping.

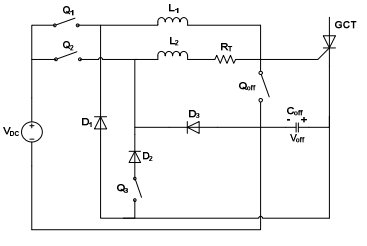


Fig. 3. The turn-on and turn-off principle diagram

Turn-on Process: In strong opening large current chopping process, Q_1 and Q_{off} are on and Q_2 and Q_3 are off, L_1 is charged, Q_1 and Q_{off} are turned off when I_{L1} is the maximum (I_{GM}), L_1 is discharged, and generate strong pulse current into the gate quickly, GCT is turned on instantaneously, I_{L1} will discrease, Q_1 is turned on when I_{L1} discreases to certain value, L_1 is charged in constant voltage(V_{DC}—V_{off}), the gate current gets up slowly to a certain value ,then Q_1 is turned off, L_1 is discharged again, I_{L1} will discrease again. Therefore, by togging on and off, Q_1 will keep the turn-on gate current within specified limits, which GCT is completely turned on; In holding opening small current chopping process, Q_1 and Q_{off} is off， then through sense resistance R_T to judge the size of holding current (I_{FGH}), which is compared with the threshold of gate current that is seted.Therefore, by controling Q_2's on and off to make GCT be

on-state. L_2 is charged in constant voltage (V_{DC}-V_{off}) when Q_2 is on, on the contrary, L_2 is discharged , Q_3 is always on in the process.

Turn-off Process: Q_1, Q_2 and Q_3 is off, Qoff is on, the bias voltage V_{off} is reversed to the gate-cathode of GCT to implement that GCT is turned off by "flow". After GCT's off, Q_{off} will continue to be off to ensure that gate-cathode of GCT is always negative bias and reliable closed.

3.2.2 Analysis and Optimization of Parameters

According to the gate drive curent requirement of on and off of GCT, basic part parameters for basic drive circuit are given through theortical anaysis and calaculation, they are: V_{DC}=20V, L_1=0.2цH, L_2=10цH, C_{off}=1600цF/35V, sense resistor R_T=1Ω, Q_1, Q_2 and Q_3 is a respectively MOSFET with V_{DSS}=100V and R_{DS}=0.055Ω, D_1 is a diode with V_{RRM}=45V and I_{FSM}=150A , D_2 is a diode with V_{RRM}=45V and I_{FSM}=25A, D_3 is a common diode, all above is the basic model parameters for circuit simulation. On this basis, related components parameters (L_1, L_2, Q_{off} and C_{off}) for the influence of the performance of drive circuit are analyzed and optimized.

(1) The Influence on Drive Current from L_1

L_1 mainly determines the maximum aplitude of gate drive current(I_{GM}), under the same parameters, L_1 is given different values, the maximum amplitude of the gate drive curent is changed, as shown in Fig 4.

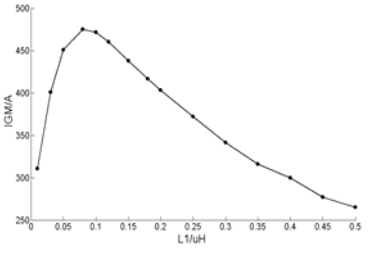


Fig. 4. the change of I_{GM} with the change of L_1

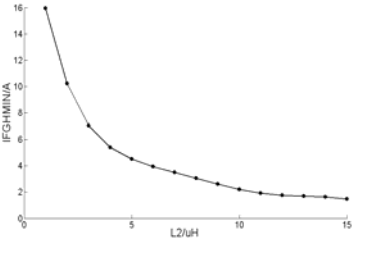


Fig. 5. the change of I_{FGHMIN} with the change of L_2

It can be seen from Fig 4 that I_{GM} is the maximum when L_1 = 0.1цH, all above meet the requrement of gate drive curent amplitude, but in the first gate continuous current discharging, the gate current is zero and GCT is closed, thus L_1 = 0.1цH did not meet the conditions; When L_1 < 0.1цH, the I_{GM} decreases with the decrease of L_1, and the change of gate current is down to zero in the first gate continuous discharging, so all values of L_1 are not desirable during L_1 < 0.1цH.When 0.1цH<L_1<0.2цH, the I_{GM} decreases with the increase of L_1, but in the first gate continuous discharging, gate current can not reach above 50A, so the value of L_1 is not desirable in the range. I_{GM} = 400A When L_1 = 0.2 цH, gate current reaches above 50A after the first gate continuous discharging, which meets the requirement of gate drive current, the value of L_1 is not taken into consideration in L_1 > 0.2цH, thus L_1 = 0.2 цH is for the optimal.

(2) The Influence on Drive Current from L_2

L_2 mainly determine the size of gate holding current, under the same parameters, L_2 is given different values, the maximum (I_{FGHMAX}) and minimum(I_{FGHMIN}) of GCT holding current are also changed, which are shown in Fig 5 and Fig 6.

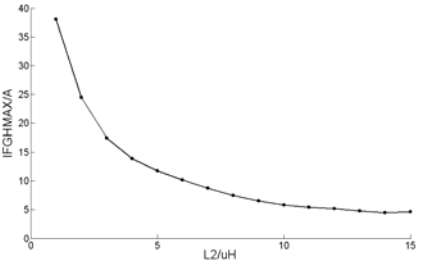


Fig. 6. the change of I_{FGHMAX} with the change of L_2

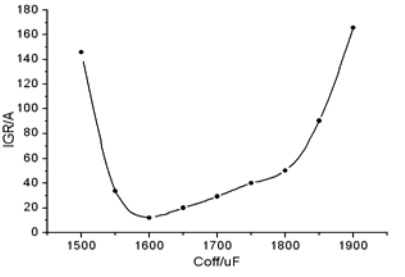


Fig. 7. the change of I_{GR} with the change of C_{off}

According to the requirement of GCT gate holding current($2A \leq I_{FGH} \leq 6A$) , it can be seen from Fig 5 that I_{FGHMIN} will decreases with the increase of L_2, I_{FGHMIN} = 2A when L_2 = 10цH.It can be seen from Fig 6 that I_{FGHMAX} will decreases with the increase of L_2, I_{FGHMAX}=6A when L_2=10цH.Therefore, it can be obtained that it just meet the requirment of GCT holding current $2A \leq I_{FGH} \leq 6A$ when L_2 = 10цH, in that case, GCT is on-state. So L_2=10цH is the best.

(3) The Influence on Drive Current from C_{off}

In the gate turn-off process, C_{off} mainly impact the size of the gate turn-off recoil current (I_{GR}).Under other components parameters are not changed, C_{off} is given different values , the change of the gate turn-off recoil current (I_{GR}) is shown in Fig 7.

According to the need of the gate turn-off of GCT, the gate turn-off recoil current (I_{GR}) is as small as possible. Fig 7 shows that the gate turn-off recoil current is minimum when C_{off}=1600цF, and has the least effect to the turn-off of GCT; the gate turn-on current is distortion when $C_{off} \leq$ 1500цF, so the value of C_{off} is undesirable within this range; When C_{off} >1900цF, the gate recoil current (I_{GR}) will increase and eventually tends to remain constant, thus the value of C_{off} is desirable in the range. C_{off} is electrolytic capacitor because of its high capacity, and can be charged to 20V in the turn-on process of GCT, the voltage of C_{off} across the course of turn-off of GCT is 20V and was added to the gate-cathode of GCT, which makes GCT is turned off instantly.Therefore, electrolytic capacitor C_{off} is 1600цF will be the best.

(4) The Influence on Drive Current from MOSFET

The resistance of Q_{off} directly affects the size of the amplitude of gate drive current (I_{GM}) and the size of I_{GR} during the gate turn-off course of GCT .Therefore, we must take into account a question that how many MOSFET is parallel should be used. Under other components parameters is unchanged, the number of MOSFET is changed, the change of the gate drive current amplitude (I_{GM}) and the gate recoil current (I_{GR}) are shown in Fig 8 and Fig 9.

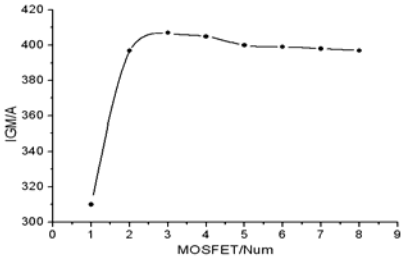


Fig. 8. the change of I_{GM} with the number of MOSFET

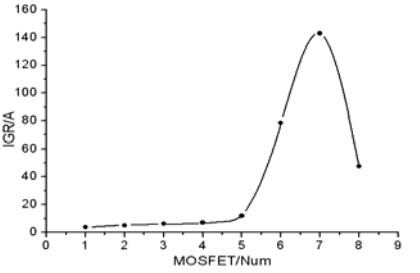


Fig. 9. the change of I_{GR} with the number of MOSFET

It can be seen from the Fig 8 that I_{GM} slowly changes when the number of MOSFET is more than two nearby the 400A, the gate drive current waveform is distorted when the number of MOSFET is between two and four, thus the number of MOSFET is undesirable in this range. When the number of MOSFET is five, I_{GM} reaches the maximum value. At the same time, it can be seen from Fig 9 that the gate turn-off recoil current(I_{GR}) is minimum when the number of MOSFET is five, thus the optimal for Q_{off} is the five parallel MOSFET.

Finally, the turn-on and turn-off of GCT are simulated and analysed and the timing pulse are applied on the Q_1, Q_2, Q_3 and Q_{off}. Fig 10 shows the timing waveform of the switch when GCT is a "turn on-holding- turn off" process. The GCT's gate-cathode structure is equivalent to a diode, therefore, using a diode instead of GCT in simulation, simulation waveforms were shown in Fig 11.

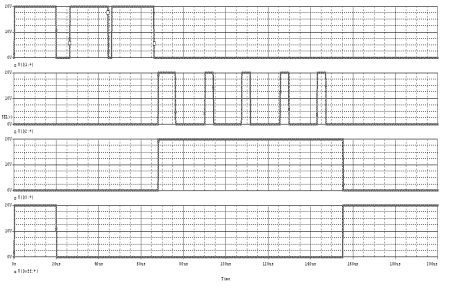


Fig. 10. Switch timing waveform

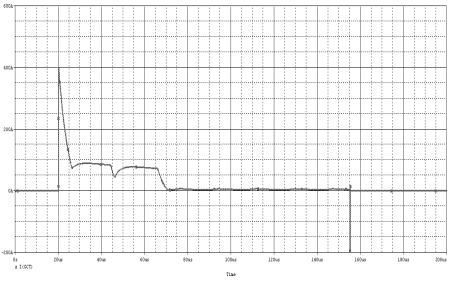


Fig. 11. GCT gate current waveform

It can be seen from Fig 10 that Q_1, Q_2, Q_3 and Q_{off} are four switches of the turn-on and turn-off circuit. High potential represents the closed switch, while low potential represents the turned off switch. GCT is turned on within the 0-66цs; GCT is the state of holding in the 66-155μs; GCT is turned off until 155цs and GCT is turned off within 1μs.

It can be seen from Fig 11 that the maximum gate drive current amplitude I_{GM} = 400A, the gate current rise rate di/dt ≥ 1000A/цs, second trigger current $I_G \geq 50A$, the gate holding current $2A \leq I_{FGH} \leq 6A$ and GCT is turned off within 1μs. These fully meet the gate current needs of turn-on and turn-off of GCT, and also prove that the selection of each component parameters is to be the best.

3.3 Control Circuit

3.3.1 Block Diagram of Control Circuit

The logic control circuits of integrate gate drive circuits of GCT use FPGA as the control core, the control command signal (CS) for the user input is judged by the software programming, and translate them into a series of control timing pulse to control the turn-on and turn-off of Q_1, Q_2, Q_3 and Q_{off} in GCT gate drive circuit, and then realize the turn-on and turn-off of GCT and display its state, the software block diagram is shown in Fig 12.

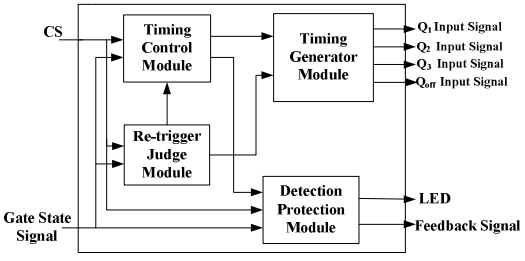


Fig. 12. Block diagram of control circuit

The main roles of timing control module: to determine whether there is the turn-on or the turn-off signals, the minimum gate turn-on and turn-off time whether is met, the detection of minimum gate turn-off interval after the re-triggered. If the signal CS meets the turn-on and turn-off conditions, the turn-on and turn-off signal is generated to control the signal timing generation module and generate the gate switching sequence pulse, finally, GCT is turned on and off. Otherwise, it will filter the received signal CS. The gate re-trigger module will judge the gate status and determine whether the gate status meets the re-trigger conditions, if meet, it will results to a re-trigger. The on and off signal timing generator module mainly generate the gate switching sequence pulse. The detection and protection module are used to detect the working state of GCT and generate the state feedback signal.

3.3.2 Simulation Results and Analysis

According to the requirement of gate drive current of GCT turn-on and turn-off, and through the logic control circuit, the turn-on and turn-off sequence of the switch are controlled. Timing simulation results are shown in Fig 13.

Fig 13 shows a normal "turn on - holding - turn off" process of GCT, the switch timings of Q_1 ($q_{[3]}$), Q_2 ($q_{[2]}$), Q_3 ($q_{[1]}$) and Q_{off} ($q_{[0]}$) are output from the FPGA and reach 20V after amplitude amplification, it can completely control the MOSFET turn-on and turn-off, then produce gate drive current to turn on and turn off GCT effectively. This time, the magnified switch timing are fully consistents with the switch timing in Fig 6. In the process, sf is feedback signal, sf is opposite with cs when GCT is normally turned on and turned off, otherwise, they are same, then it is sent to the control system through external optical transmitter to facilitate the field staff to detect

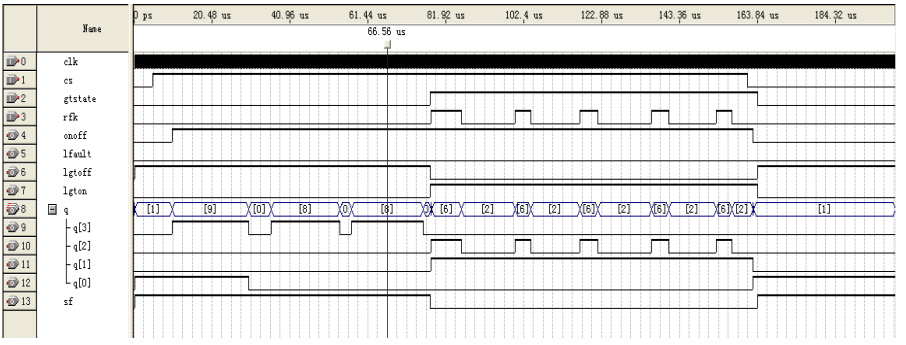


Fig. 13. Switch timing waveform

it. Signal lgton, lgtoff and lfault are respectively the trun-on, turn-off and fault of gate, and through external circuit LED display it to facilitate real time detection of GCT`s working status.

4 Conclusion

The paper is based on the basic structure and working principle of GCT, which presents a design of gate drive circuit of reverse-conducting GCT with 1100A/4500V , mainly make the simulation and parameters optimization of the turn-on and turn-off circuit of GCT and the design of the logic control circuit, finally get V_{DC}=20V, L_1=0.2μH, L_2=10μH, C_{off}=1600μF/35V, current sense resistor R_T=1Ω, Q_1, Q_2 and Q_3 is a respectively MOSFET with V_{DSS}=100V and R_{DS}=0.055Ω, Q_{off} is five parallel MOSFET, D_1 is a diode with V_{RRM}=45V and I_{FSM}=150A, D_2 is a diode with V_{RRM}=45V and I_{FSM}=25A, D_3 is a common diode, and they are the optimum design parameters, under the effective control of the logic control circuit, it achieves the turn-on and turn-off of GCT efectively.

References

1. Tong, Y., Zhang, C.: The influence of Anti-parallel diode to turn-off process of IGCT. The Journal Of Electrical Technology 22(11) (2007)
2. Steimer, P., Apeldoorn, O., Carroll, E.: Nagel A: IGCT Technology Baseline and Furture Opportunities. IEEE/PES 2(2) (2001)
3. Bernet, S., Teichmann, R.: Comparion of High -Power IGBT's Hard-Driver GTO's for High Power Inverters. IEEE, Transactions on Industry Application, 711–718 (1998)
4. Zhang, C.: Rearch On Gate Drive Circuit Of Integrated Gate Commutated Thysistor. Peking Communication University (2007)
5. Klaka, S., Linder, S., Frecker, M.: A Family of Reverse Conducting Gate Commutated Thysitors for Mediu Voltage Drive Applications, PCIM Hong Kong, 1–11 (October 1997)

6. Huang, X., Zhang, X., Xie, L.: Design Of Drive Unit Control Circuit GCT. The High Power Converter Technology (1), 9–13 (2009)
7. Zhuang, L., Jin, X., Wu, X., Xie, L., Huang, X.: Rearch on a Novel Gate Circuit of Integrated Gate Commutated Thysitor. Communication Technology And Electric Traction (2), 16–20 (2008)
8. Qiu, C., Qiu, R.-c.: Research on a Novel Gate Drive Circuit of Integrated Gate Commutated Thyristor

Multi-fuel Combustion Measurement in Utility Boiler for Optimal Operation

Luo Zi-Xue

State Key Laboratory of Coal Combustion, Huazhong University of Science and Technology, Wuhan 430074, P.R. China
luozixue@sina.com

Abstract. This paper presents a combustion measuring method for optimal combustion through adjusting the proportions of coal, blast furnace gas (BFG) and coke oven gas (COG) in a 200 MWe utility boiler. A reconstructed temperature and a flame emissivity algorithm are derived to analyze the mixed combustion in the furnace. The result shows that higher combustion efficiency can be obtained by setting the proper ratio of coal, BFG and COG, namely 0.7:0.2:0.1.

Keywords: multi-fuel combustion; image processing; temperature; flame emissivity; optimal operation.

1 Introduction

Combustion products in pulverized-coal fired boilers include gaseous and solid materials, such as CO2, H2O, char, fly ash, and soot. Many researchers have explored the variation of particle emissivity in pulverized-coal with the volatiles removed [1]. Temperature and flame emissivity measurements have also been carried out as a useful topic [2-7]. Combustion with several kinds of fuel has often been used for the sake of energy saving and emission reduction in a utility boiler [8] in recent years. A 200 MWe power plant boiler was retrofitted for mixed combustion of coal, blast furnace gas (BFG) and coke oven gas (COG).

In this paper, experimental research is performed through adjusting the proportions of coal, BFG and COG. Based on the temperature and a flame emissivity monitoring algorithm, infrared radiation thermometers and a portable image processing technique are applied to analyze the multi-fuel combustion conditions.

2 Measuring Algorithm

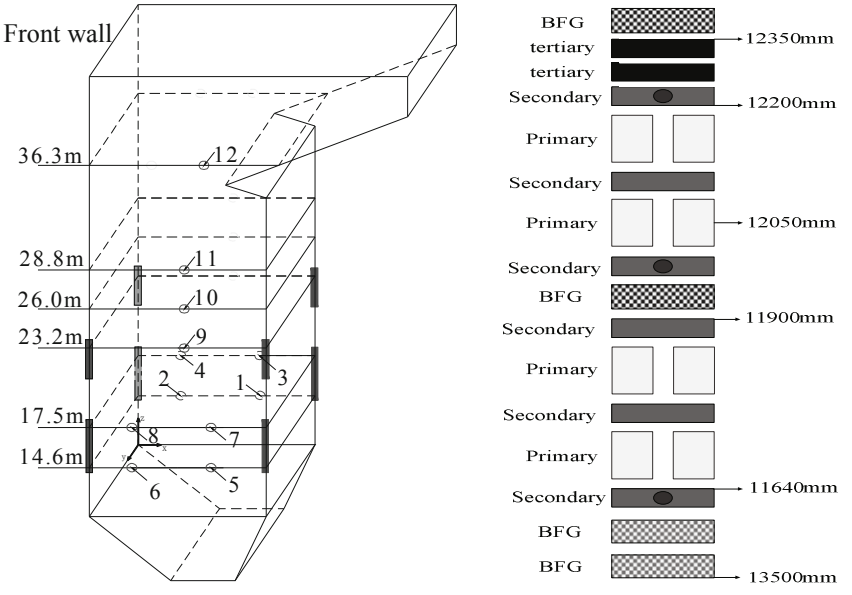
The experimental boiler was designed for burning lean coal. Fig. 1 shows the structural schematic and the location of burners. The 3-D temperature field can be reconstructed from cold ash hopper to flame corner [9]. Fig. 1(a) shows the location of the measuring points in the boiler. The correctness of the measured results was revised by reference to a black-body furnace.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 63–69.
springerlink.com

According to Wien's law, the monochromatic radiation intensity can be expressed as:

$$I(\lambda,T)=\varepsilon_\lambda c_1 e^{-C_2/(\lambda T)}/(\pi\lambda^5) \tag{1}$$

Where T is Kelvin temperature, λ is wavelength, c_1 and c_2 are constants, and ε_λ is the flame emissivity.



(a) Boiler structure and location of measuring points (b) Arrangement of burners

Fig. 1. Schematic views of boiler structure in experimental study

The flame image can be transformed to a digital signal based on image processing technology, which is formed with three primary colors of Red, Green and Blue. Each pixel represents a monochromatic radiant intensity, and it can be illustrated as [10]:

$$\begin{aligned} I(\lambda_r,T)&=k_r R \\ I(\lambda_g,T)&=k_g G \end{aligned} \tag{2}$$

Where, k_r and k_g are calibration coefficients, $I(\lambda_\gamma,T)$ and $I(\lambda_g,T)$ are the monochromatic radiation intensities at the wavelengths λ_γ and λ_g, respectively.

From consideration of equations (1) and (2), the calibration coefficients k_r and k_g are:

$$\begin{aligned} k_r&=c_1\varepsilon_r e^{-C_2/(\lambda_r T)}/(\pi R\lambda_r^5) \\ k_g&=c_1\varepsilon_g e^{-C_2/(\lambda_g T)}/(\pi R\lambda_g^5) \end{aligned} \tag{3}$$

The pulverized-coal combustion flame can be treated as a gray body in visible light, and two arbitrary colors can be extracted from the homogeneous radiation image. Adopting *Red* and *Green* for calculation, the flame temperature can be expressed as follows:

$$\begin{aligned} T &= -c_2\left(\frac{1}{\lambda_r}-\frac{1}{\lambda_g}\right)/\ln\left(\frac{I(\lambda_r,T)\lambda_r^5}{I(\lambda_g,T)\lambda_g^5}\right) \\ &= -c_2\left(\frac{1}{\lambda_r}-\frac{1}{\lambda_g}\right)/\ln\left(\frac{k_r R\lambda_r^5}{k_g G\lambda_g^5}\right) \end{aligned} \tag{4}$$

Therefore, the flame emissivity can be obtained through the expression given below:

$$\varepsilon = \begin{cases} k_r \pi R\lambda_r^5 /(c_1 e^{-c_2/\lambda_r T}) \\ k_g \pi G\lambda_g^5 /(c_1 e^{-c_2/\lambda_g T}) \end{cases} \tag{5}$$

Where, $c_1 = 3.741832\times10^8$ W.μm^4/m^2, $c_2 = 1.4388\times10^4$ μm.K, $\lambda_r = 0.61$μm and $\lambda_g = 0.51$ μm.

3 Results and Discussion

Fig. 2 shows the temperature distribution at several cross sections with different fuel proportions. The highest temperature exists in the region of the burners, and temperature decreases along the height of upper boiler, which is consistent with the practical mode in these four cases.

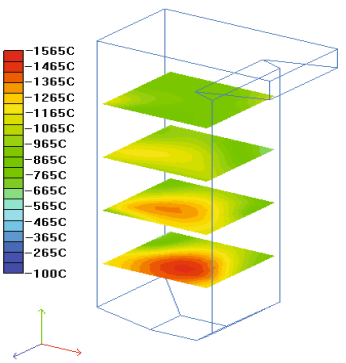
The temperature of the right wall is much higher than that of the left, and the tangential circle of flame center is displaced towards the back-right wall at the first floor(height 16.0 m in boiler). However, the tangential circle at the third floor (height 28.0 m) is displaced towards the front-right wall.

Figs. 2 (a) and (b) show that addition of COG leads to an increase in the highest temperature of 52 K; but the location of the flame center is not influenced. Fig. 2 (c) shows that addition of BFG results in a decrease in temperature of 94 K.

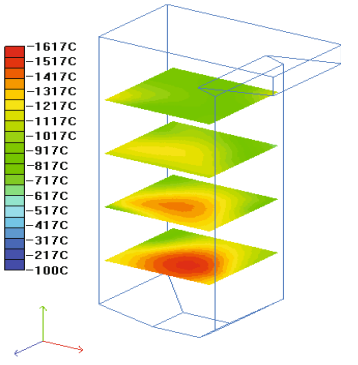
As shown in Fig. 2 (d), the furnace temperature was raised along with a decrease in the height of the flame center by further addition of COG. The percentage distribution of multi-fuel combustion, pulverized-coal, BFG and COG is nearly 70%, 20% and 10% at a load of 170 MWe.

The difference between the temperatures measured by infrared radiation and by 3-D distribution is shown in Fig. 3 (a) and Table 1. The error is less than 5%, and the two methods correctly reflect the tendency of temperature in furnace.

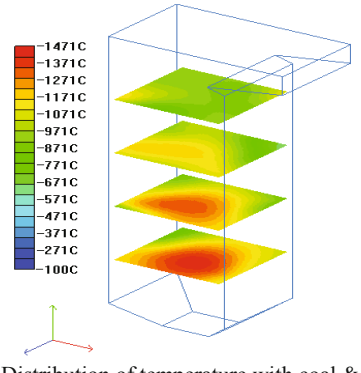
The flame emissivity is considered as the direct reflection of combustion and radiation in the boiler. The developed method is to monitor and judge combustion behavior for varied kinds of fuels by means of flame emissivity.



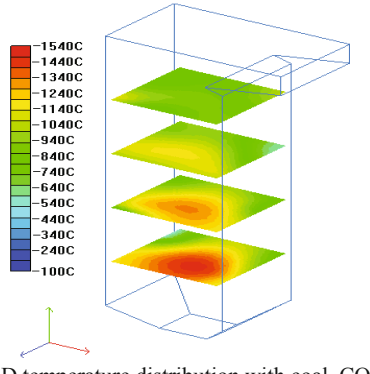
(a) Temperature of single coal combustion



(b) Temperature distribution with coal and COG (10×10^3 m^3/h) mixed combustion

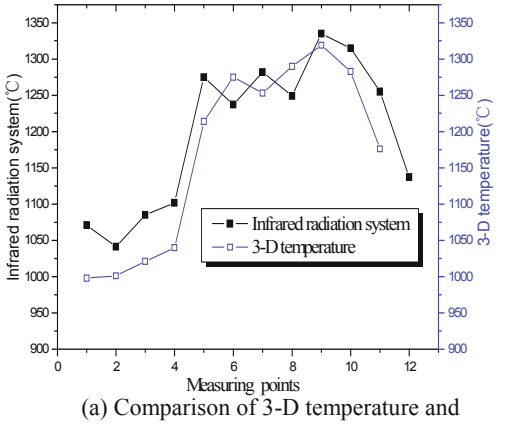


(c) Distribution of temperature with coal & BFG (100×10^3 m^3/h) mixed combustion

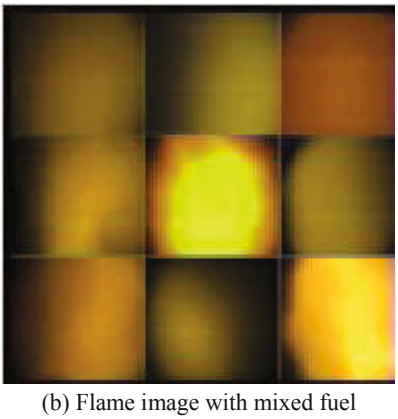


(d) 3D temperature distribution with coal, COG (10×10^3 m^3/h) and BFG (100×10^3 m^3/h) mixed combustion

Fig. 2. Distribution of 3D temperature based on flame image processing with mixed fuel combustion at a load of 170 MWe



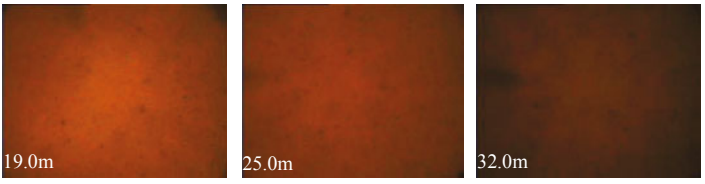
(a) Comparison of 3-D temperature and infrared measurement



(b) Flame image with mixed fuel combustion of coal and COG at 170 MWe

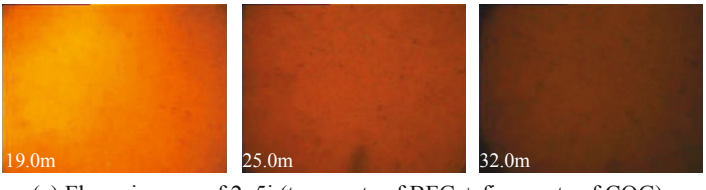
Fig. 3. Flame image and temperature measurement

Fig. 4 shows the flame images at different height with multi-fuel combustion in the furnace. It can be seen that the variation of fuel types has a significant influence on the flame image, and addition of less BFG and addition of more COG can make the flame brighter.



(a) Flame images of 2g0j (two ports of BFG + zero port of COG)

(b) Flame images of 2g2j (two ports of BFG + two ports of COG)



(c) Flame images of 2g5j (two ports of BFG + five ports of COG)

Fig. 4. Flame images at different heights with multi-fuel combustion

Table 1. Temperature comparison between 3D visualization system and infrared measurement

Measuring points	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Infrared measurement (ºC)	1071	1041	1085	1102	1275	1237	1282	1249	1335	1315	1255	1137
3D temp. system(ºC)	998	1001	1021	1040	1214	1275	1253	1290	1319	1283	1176	–

Table 2 shows the flame emissivity with multi-fuel combustion at different heights. The largest flame emissivity existed in the upper region of burners, where the highest temperature accelerated the separation of volatiles and decreased the accumulation of un-burnt particles in the region. Furthermore, addition of COG should introduce a large amount of volatiles to form soot, and produce higher flame emissivity.

However, most fuels had burnt out at 32.0 m, and the flame emissivity changed very little. In general, BOG can reduce the concentration of soot and restrain combustion, and COG may accelerate combustion through increasing the concentration of volatiles. Proper mixture ratio of multi-fuel is one of the important factors for optimal operation in a utility boiler.

Table 2. The variation of emissivity in furnace with mixed fuel combustion

Height in furnace (m)	Emissivity on coal combustion with mixed-fuel of BFG+COG in furnace			
	2g0j *	2g2j	2g4j	2g5j
19.0	0.144	0.146	0.166	0.196
25.0	0.205	0.184	0.197	0.242
32.0	0.134	0.163	0.171	0.161

*: 'g'-BFG, 'j'-COG, coefficients represent the number of operating ports.

4 Conclusions

A suitable method is put forward to analyze the combustion behavior based on flame image digital processing technology in multi-fuel combustion in a power plant boiler. The addition of BOG may lead to a reduction of flame emissivity and decrease the temperature; and increment of COG should alter the procedure. The proportion of mixed fuels; coal, BFG and COG; for optimal combustion is nearly 0.7:0.2:0.1 at a load of 170 MWe in the boiler.

Acknowledgments. This study has been supported by the National Natural Science Foundation of China (Nos. 51076049 and 50721005) and in part supported by the National Hi-Technology Research and Development Program of China (863 Program, No. 2007AA05Z306).

References

1. Bhattacharya, S.P., Wall, T.F., et al.: Development of emittance of coal particles during de-volatilization and burn-off. Fuel 78(5), 511–519 (1999)
2. Qi, J.-A., Leung, C.-W., et al.: Temperature-field measurements of a premixed butane/air circular impinging-flame using reference-beam interferometry. Applied Energy 83, 1307–1316 (2006)
3. Solomon, P.R., Carangelo, R.M., Best, P.E., et al.: The spectral emittance of pulverized coal and char. Proceedings of the Combustion Institute 21(1), 437–446 (1988)
4. Cai, X.-S., Luo, W.-D.: Research of flame temperature and flame emissivity of pulverized coal through spectrum method. Journal of Engineering Thermophysics 21(6), 779–792 (2000) (in Chinese)
5. Bheemul, H.C., Lu, G., Yan, Y.: Three-dimensional visualization and quantitative characterization of gaseous flames. Measurement Science and Technology 13, 1643–1650 (2002)
6. Lu, G., Yan, Y., Riley, G., Bheemul, H.C.: Concurrent measurement of temperature and soot concentration of pulverized coal flames. IEEE Transactions on Instrumentation and Measurement 51(5), 990–995 (2002)
7. Yan, Y., Lu, G., Colechin, M.: Monitoring and characterization of pulverized coal flames using digital image techniques. Fuel 81, 647–656 (2002)

8. Su, S., Xiang, J., Sun, L.-S., et al.: Application of gaseous fuel re-burning for controlling nitric oxide emissions in boilers. Fuel Processing Technology 90, 396–402 (2009)
9. Zhou, H.-C., Lou, C., Cheng, Q., Jiang, Z., et al.: Experimental investigations on visualization of three dimensional temperature distributions in a large-scale pulverized-coal-fired boiler furnace. Proceedings of the Combustion Institute 30, 1699–1706 (2005)
10. Lou, C., Zhou, H.-C., Jiang, Z.-W.: On-line detection and analysis of flame emissivity in coal-fired boilers. Proceedings of the CSEE 26(4), 6–9 (2006) (in Chinese)

Customized Processor Architecture for Model Predictive Control in Magnetic Actuated Small Satellites

Xueliang Bai and Xiaofeng Wu

School of Aerospace, Mechanical & Mechatronic Engineering
University of Sydney, NSW 2006, Australia
{xueliang.bai,xiaofeng.wu}@sydney.edu.au

Abstract. Future spacecraft are envisioned as autonomous, miniature, and intelligent space systems. This paper describes the design and implementation of a model predictive control (MPC) system for satellite attitude control. The MPC algorithm is designed to successfully deal with constraints due to the small control torque given by magnetic torquers and the Earth's magnetic field. Laguerre functions are proposed to simplify the implementation of the MPC controller for on-line computation. A control system processor is designed as a peripheral hard core of the system-on-chip for satellite on-board data handling. Targeting the FPGA technology, this processor runs up to 120 MHz.

Keywords: Control System Processor, FPGA, Model Predictive Control, Satellite.

1 Introduction

The development of what is known as nanosatellites, with typical mass less than 10kg, is quickly transforming the Space development scene as it allows engineering and science researchers to send into orbit various payloads within a short time at low cost. In particular, in the last 10 years many universities like Stanford [1] have carried out university satellite programs, and nanosatellites have been built and successfully launched into space. Such small satellite can be controlled by various actuation methods, including thrusters, reaction wheels, magnetic torquers, or a combination of above. Currently electromagnetic actuator is the most effective approach, and has been adopted for many nanosatellite missions, e.g. the CanX-1 [2], AAUSat [3], Compass One [4] and so on.

The magnetic torquer interacts with the earth's own magnetic field in order to generate a control torque acting on spacecraft. An advantage of magnetic torquers is that they require no fuel. They do require electrical power, but there is no exhaust pollutant and by providing a couple they are not sensitive to movement of the centre of mass. One drawback of this control technique is that the torques which can be applied to the spacecraft for attitude control purposes are constrained to lie in the plane orthogonal to earth magnetic field, and hence the satellite is under-actuated. In the equatorial plane, the magnetic field line always lie horizontally, north-south. Consequently a spacecraft whose orbit lies in this plane cannot use magnetic torquers to counteract

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 71–79.
springerlink.com

the north-south component of their disturbance torque, or to dump this component of momentum. For an inclined orbit, suitable variation of the magnetic field allows controllability in the long term, but presents a significant challenge from a control perspective [5]. The model predictive control is proposed to solve such problems for space missions [6-8].

The MPC algorithm results in very complex matrix operations, which limit its feasibility for small satellites. In real-time control, to execute extremely fast control laws for feedback systems, a control system processor (CSP) was designed [9, 10]. The excellent performance of the CSP is achieved by implementing simple mixed data formation, with 24-bit fixed point data for state variables and 11-bit low precision floating data for coefficients. The CSP takes advantage of single processing element, i.e. multiply and accumulation (MAC) to execute all the arithmetic operation. In [11], another dedicated control system processor was developed based on 1-bit processing [12]. In this paper to effectively implement the MPC, the controller structure is simplified using Laguerre functions [13]. At the same time, a system-on-chip (SoC) is proposed for the satellite on-board data handling system (OBDH). A dedicated model predictive control system processor (MPCSP) is designed as a peripheral hard core of the SoC for attitude control.

The remaining paper is organized as follows. Section 3 proposes a novel MPC approach using Discrete-time Laguerre networks in magnetic attitude control problem. Section 4 introduces the MPCSP design. Section 5 presents the simulation results for attitude control of a student-built nanosatellite. Section 6 concludes.

2 Model Predictive Control

2.1 Discrete-Time MPC Using Laguerre Functions

Considering a linear system described by the discrete-time state space model

$$\begin{aligned} x_m(k+1) &= A_m x_m(k) + B_m u(k) \\ y(k+1) &= C_m x_m(k) \end{aligned} \tag{1}$$

According to Ref. [14], the augmented model is built as follow.

$$\overbrace{\begin{bmatrix} \Delta x_m(k+1) \\ y(k+1) \end{bmatrix}}^{x(k+1)} = \overbrace{\begin{bmatrix} A_m & o_m^T \\ C_m A_m & 1 \end{bmatrix}}^{A} \overbrace{\begin{bmatrix} \Delta x_m(k) \\ y(k) \end{bmatrix}}^{x(k)} + \overbrace{\begin{bmatrix} B_m \\ C_m B_m \end{bmatrix}}^{B} \Delta u(k)$$

$$y(k) = \overbrace{\begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix}}^{C} \begin{bmatrix} \Delta x_m(k) \\ y(k) \end{bmatrix} \tag{2}$$

A set of Laguerre functions can be used to formulate MPC problem. At an arbitrary future sample instant, the difference of the control signal can be expressed as

$$\Delta u(k_i + k) = \begin{bmatrix} L_1(k)^T & 0 & \cdots & 0 \\ 0 & L_2(k)^T & \cdots & 0 \\ \cdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & L_m(k)^T \end{bmatrix} \eta \tag{3}$$

$L_i(k)$ represents the Laguerre network description for the ith control. $L_i(k)$ is given by $L_i(k+1) = A_l L_i(k)$, where matrix A_l is (N×N) and is a function of parameters α and $\beta = (1-\alpha^2)$, and the initial condition is given by $L(0)^T = \sqrt{\beta}[1 - \alpha\ \alpha^2 - \alpha^3 \ldots (-1)^{N-1}\alpha^{N-1}]$, and

$$A_l = \begin{bmatrix} \alpha & 0 & 0 & 0 & 0 \\ \beta & \alpha & 0 & 0 & 0 \\ -\alpha\beta & \beta & \alpha & 0 & 0 \\ \alpha^2\beta & -\alpha\beta & \beta & \alpha & 0 \\ -\alpha^3\beta & \alpha^2\beta & -\alpha\beta & \beta & \alpha \end{bmatrix}$$

The scaling factors α and N can be selected independently for each control signal. η is the parameter vector and η_i comprises N Laguerre coefficients. $\eta = \left[\eta_1^T \eta_2^T \eta_3^T \ldots \eta_m^T\right]$ and $\eta_i = [c_1\ c_2\ c_3 \ldots\ c_N]^T$

The cost function is defined as

$$J == \eta^T \chi \eta + 2\eta^T \Gamma x(k_i) \tag{4}$$

where, $X = \sum_{m=1}^{N_p} \xi(m) Q \xi(m)^T + R_L$, $\Gamma = \xi(m) Q A^m$, $\xi(m)^T = \sum_{i=0}^{m-1} A^{m-i-1} BL(i)^T$, and N_p is the prediction horizon.

It should be noted here that k_i represents the current sampling time and k represents future sampling time.

2.2 Problem Formulation

In this paper, the idea is to design a controller, which guarantees the orthogonality between the geomagnetic field vector b and the magnetic control vector T_c . and make the amplitudes of actuator within the saturations. In this design, the control variable is $T_c = u(k_i) = u(k_i - 1) + \Delta u(k_i)$ and the constraints become

$$b(k_i)^T u(k_i) = 0 \tag{5}$$

$$\begin{bmatrix} -B(b(k_i)) \\ B(b(k_i)) \end{bmatrix} u(k_i) \le \begin{bmatrix} m_{max} \\ m_{min} \end{bmatrix} |b(k_i)|^2 \tag{6}$$

where m is the magnetic dipole. This equation can be expressed in term of U as

$$[\, b(k_i)^T\ 0\ \cdots\ 0]U = 0 \tag{7}$$

$$\begin{bmatrix} B(b(k_i)) & 0 & \cdots & 0 \\ -B(b(k_i)) & 0 & \cdots & 0 \end{bmatrix} U \le \begin{bmatrix} m_{max} \\ m_{min} \end{bmatrix} |b(k)|^2 \tag{8}$$

where, $U = G\,\eta + [u(k_i - 1)^T\ u(k_i - 1)^T \cdots\ u(k_i - 1)^T]^T$,

$$g(k_i + k) = \begin{bmatrix} \sum_{i=0}^{k-1} L_1(i)^T & 0 & \cdots & 0 \\ 0 & \sum_{i=0}^{k-1} L_2(i)^T & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sum_{i=0}^{k-1} L_m(i)^T \end{bmatrix},$$

$G = [\, g(1)^T\ g(2)^T \cdots g\left(N_p\right)^T\ ]^T$.

When the constraints are considered, the optimization procedure is to minimize the cost function J in Eq. 4 and operate quadratic programming, which is able to handle both the equality and inequality constraints to get the optimal solution η_{opt}.

The difference of optimal control signal $\Delta u(k_i)$ is given by

$$\Delta u(k_i) = \begin{bmatrix} L_1(0)^T & 0 & \cdots & 0 \\ 0 & L_2(0)^T & \cdots & 0 \\ \cdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & L_m(0)^T \end{bmatrix} \eta_{opt} \tag{9}$$

And the optimal control signal $u(k_i)$ can be calculated as $u(k_i) = \Delta u(k_i) + u(k_i - 1)$. Finally, the optimal vector of the coils' magnetic dipoles m_{opt} can be calculated by

$$m_{opt}(k_i) = \frac{1}{|b(k_i)|^2} B(b(k_i))' u(k_i) \tag{10}$$

By using Laguerre functions, if the prediction horizon the Number of terms N is set to be 5 for each control signal, only five parameters are needed to capture the current control signal and totally 15 parameters for three control signals. While in traditional MPC, thirty parameters are needed to capture one control signal, and totally 90 parameters are needed for three signals. By reducing the number of parameters in the controller design, Laguerre MPC can dramatically release the computational burden for nanosatellites.

3 Hardware Design

Most MPC applications target process control, in which sample periods are low and the plant is physically large, meaning that processing based upon an industrial computer is adoptable. However, the proposed electro-magnetic control system is targeting very small satellites, in which the controller hardware must be embedded. It may require very high sampling rate for precise attitude control. Also executing the MPC is relatively complex with heavy matrix operations. It is therefore requires some high performance device for control system processing.

3.1 SoC for Satellite On-Board Data Handling

There are two types of satellite on-board data handling systems: central and distributed. The central processing approach has one on-board computer to deal with all the data processing for each subsystem. The distributed processing approach, however, has many on-board computers. Some subsystems may have more than one processor. Nowadays most of the satellites adopt the distributed approach, but for nanosatellites this approach is not efficient due to the limited size and power. Hence, the SoC solution is proposed not only for attitude control but also for data processing for other subsystems.

The SoC can implement the whole digital functionality of the satellite on a single chip. The gate densities achieved in current FPGA devices have enough logic gates/elements to implement different functionalities on the same chip by mixing

self-designed modules with third party ones. In the SoC, it contains dedicated processors for each subsystem. Fig. 1 shows the SoC architecture. It contains a general purpose processor and the dedicated processors. Its structure comprises an AMBA [15] compliant bus that communicates between the general purpose processor and the control system processor. Generally, the MPCSP is independent, but the general purpose processor as the controller of the SoC monitors the control state variables and satellite attitude information from the inertial sensors. These data are a part of telemetry information for satellite house-keeping.

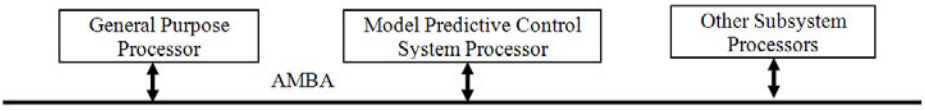


Fig. 1. Soc architecture for satellite OBDH

3.2 MPCSP Design and Implementation

To map the control algorithm to processor architecture, it is divided into tasks or processes. These processes include data input and output (IO), data storage (Memories), timer, instruction fetching and decoding, next instruction address calculation (Program Counter) and arithmetic operations (ALU). This partitioning should allow all the processes to be mapped easily into hardware, minimising the resources required.

The number of concurrent operations can determine the amount and functionality of the hardware structures. For example, the maximum number of simultaneous data transactions that required for arithmetic operations determines the number of ALU ports. Also, communication channels between the ALU, accumulator, memories and IO must be assigned with specific data bus.

The execution of the control algorithm requires the repeated execution of a set of instructions (program). Although the number of instructions in the control loop can be small in the case of implementing a simple controller, the overhead that manipulate the program counter maybe relatively large. We therefore must pay special attention to the architecture of the program counter that implements control loops. Thus, the MPCSP can provide a looping mechanism that introduces a short, or ideally zero, overhead.

The final step is to create a hardware model that supports the operations needed to implement the control algorithm. This hardware model is programmed using the hardware description language. The resulted MPCSP is simulated and verified by running the MPC algorithm with the application-specific instructions. The MPCSP is then synthesized, floor planned and placed & routed. The final netlist can be verified via being downloaded into the FPGA and running the hardware-in-loop simulation.

Fig.2 shows the MPCSP architecture. It adopts a simple mixed data format in 2's complement: the coefficients are in 12-bit floating point format, with 6-bit for mantissa and 6-bit for exponent. The state variables and other data are in 32-bit fixed point format, with 16 integer bits and 16 fractional bits. The memories include program ROM, data ROM and data RAM for control program, coefficients and intermediate states respectively. The sample timer is used to determine the sample

interval for each control loop. The program counter processes one instruction at each clock cycle. It will halt the operation at the end of the control program, and reset to the address where the control loop starts at the rising edge of the sample timer.

The MAC is the only processing element in the previous CSP design. Although this approach reduces the circuitry complexity, it is not efficient in terms of power and speed. Hence, in the MPCSP design, several processing elements are adopted for arithmetic operations as shown in Table 1. Each instruction uses the standard RISC convention of 32-bit fixed-length. All instructions have a single clock cycle execution.

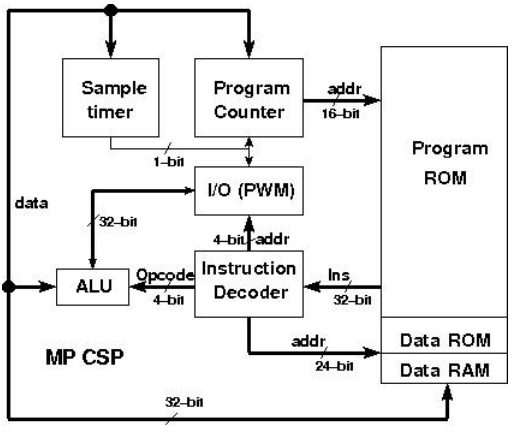


Fig. 2. MPCSP architecture.

Table 1. MPCSP instructions

Opcode	Name	Function Description
0000	HLT	No Operation
0001	RDW	Read data from data ROM
0010	WRW	Write data to data RAM
0011	OUT	Output the control signals
0100	MUL	Multiply
0101	ADD	add
0110	SUB	subtract
0111	INV	invert
1000	SET	Set the sampling frequency
1xx1	WPC	Set the star value for the program counter

The IO block has 12 inputs and 4 PWM outputs. The inputs are connected to the inertial sensors, including accelerometers, gyroscopes, magnetormeters and sun sensors. The outputs are connected to the magnet coils through the power amplifiers. For attitude control, 3 magnet coils are needed to provide three-axis actuation. The data bus and address bus of the MPCSP are connected to the AMBA to allow the general-purpose processor collect house-keeping data. The MPCSP is implemented targeting the Xilinx

Virtex-4 FX100 FPGA technology. The MPCSP occupies less than 8% of the FPGA total area. It runs up to 120MHz. The power consumption is around 183 mW.

4 Implementation and Results

The simulating nanosatellite operates at low Earth orbit altitude of 650km with an orbital inclination of 96° and has inertia matrix J=diag([6.858e-4 6.858e-4 8.164e-4]).

The MPC tuning parameters are listed in Table 2. We can see that in traditional MPC design, the control horizon N_c is normally chosen to be more than 30 to get a sound control performance. While using Laguerre function, N can be selected as 5 to obtain the same performance, which means 6 times less parameters involved in on-line computation.

Table 2. MPC tuning parameters

MPC	Without Laguerre	With Laguerre
Sampling interval	60s	60s
Prediction horizon (N_P)	30	60
Control horizon (N_c)	30	5
Scaling factor for T_{Cx} T_{Cy} T_{Cz}	0.5,0.5,0.5	0.5,0.5,0.5
Number of terms for T_{Cx} T_{Cy} T_{Cz}	5,5,5	5,5,5
Control weighting (R)	0.1,0.1,0.06	0.1,0.1,0.06

The control system is initialized at satellite pointing angles of 1° about each axis and angular rates of 0.0005 rad/s about the roll, yaw and pitch axes.

The first set of simulation was carried out in Matlab. The MPC is then implemented on the MPCSP. A hardware-in-loop simulation platform as shown in Fig. 3 is developed to test the control system processor. The MPCSP is implemented in a Virtex-4 FX100 FPGA board. The satellite attitude dynamics is modelled in C program in the computer. The PWM actuation signals from the FPGA board are sent to the computer using the industrial digital IO card. The feedback signals like angular rate are sent to the FPGA board using the analogue to digital converters.

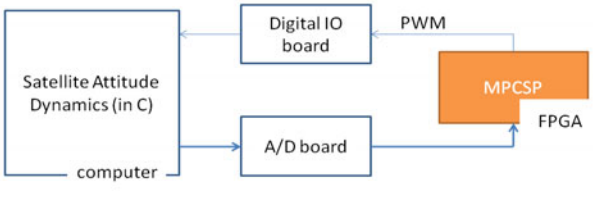


Fig. 3. Hardware-in-loop simulation for MPCSP

Fig.4 shows the simulation results from both the Matlab and the hardware-in-loop simulations. Compared to the Matlab simulation, in the difference is rather small. Hence the MPCSP is feasible for satellite attitude control using magnet torquers.

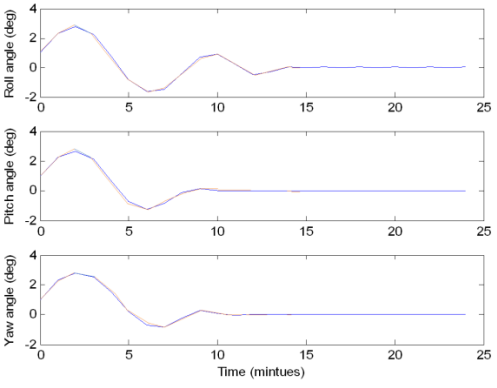


Fig. 4. Satellite attitude angles: comparison between Matlab simulation results and hardware-in-loop simulation results (red).

5 Conclusion

A dedicated control system processor is developed to execute the MPC algorithm. The MPCSP can run upto 120MHz, while consuming 183mW and occupies less than 8% area of a Virtex-4 FX100 FPGA. The MPCSP is used to implement a model predictive attitude control law for a nanosatellite. The hardware-in-loop simulation shows that the MPCSP produces almost the same results as the Matlab simulation.

References

1. CubeSat Kit, http://www.cubesatkit.com/
2. Wells, G.J., Stras, L., Jeans, T.: Canada's Smallest Satellite: The Canadian Andvanced Nanospace eXperiment, http://people.scs.carleton.ca/~barbeau/Picosat/CanX_Paper.pdf
3. AAU CubeSat, http://www.cubesat.auc.dk/
4. Compass One, http://www.cubesat.de/
5. Fortescue, P., Stark, J., Swinerd, G.: Spacecraft Systems Engineering, 3rd edn. Wiley, Chichester
6. Silani, E., Lovera, M.: Magnetic spacecraft attitude control: a survey and some new results. Control Engineering Practice 13, 357–371
7. Chen, W., Wood, M., Fertin, D.: Model Predictive Control of Low Earth Orbiting Spacecraft with Magneto-torquers. In: Proceedings of IEEE International Conference on Control Applications, Munich (2006)

8. Chen, W., Wood, M.: Model predictive control of low earth orbiting spacecraft with magneto-torquers. In: Computer Aided Control System Design, 2006 IEEE International Conference on Control Applications, 2006 IEEE International Symposium on Intelligent Control, October 4-6, pp. 2908–2913. IEEE, Los Alamitos (2006)
9. Goodall, R., Jones, S., Cumplido-Parra, R., Mitchell, F., Bateman, S.: A control system processor architecture for complex LTI controllers. In: Proc. 6th IFAC Workshop AARTC 2000, Palma de Mallorca, Spain, pp. 167–172 (2000)
10. Cumplido-Parra, R.A.: On the design and implementation of a control system processor. Ph.D. thesis, Loughborough University
11. Wu, X., Chouliaras, V., Nunez-Yanez, J., Goodall, R.: A Novel Control System Processor and Its VLSI Implementation. IEEE Trans. on VLSI (2008)
12. Wu, X., Goodall, R.: 1-bit processing for digital control. In: IET Proc. Control Theory & Applications, vol. 152(4) (2005)
13. Stempak, K.: Equiconvergence for Laguerre function series. Studia Mathematica 118(3) (1996)
14. Wang, L.: Model Predictive Control System Design and Implementation Using MATLAB. Springer Press, Heidelberg
15. ARM Limited. AMBA Specification (Rev 2.0) (1999), `http://www.gaisler.com/doc/amba.pdf`

*i*CloudMedia: A Business Model for Cloud-Based Interactive Multimedia Services*

Phooi Yee Lau, Sungkwon Park, Joohan Lee, Joonhee Yoon, and Eunjo Lee

Media Communications Laboratory, Hanyang University, 133-791 Seoul, Republic of Korea
{laupy,sp2996,hitch100,jjdad,leeej}@hanyang.ac.kr

Abstract. The Pay-TV markets are undergoing rapid change. As companies increasingly look for resources efficiencies, cloud computing is seen making waves in the Pay-TV markets. This hugely popular technology will revolutionized how multimedia content is being delivered. Why? We are in dire needs to integrate web, social, mobile and on-demand TV together as interactive content, and deliver them to any television set as a single service. We have seen some companies offering these services and have been rather successful. What they still lack is to provide some form of flexibility in resource(s) provision and content management, especially to subscriber(s). Currently, the resources are allocated based on availabilities and service agreements, while content is managed using on-the-top program guide tightly controlled by the service operator. This paper introduces a new framework, named *i*CloudMedia. The *i*CloudMedia describes a cloud-based framework that allows "push" on-the-fly changes and "pulls" rich personalized multimedia content simultaneously in an interactive resource(s) market. We strongly believe that innovative content branding and configurable resource management through easy-to-manage and flexible user interface is the next logical step for service operators.

Keywords: cloud computing, multimedia content, self-configurable resource management, system architecture.

1 Introduction

Digital content industry is the new driving force for the development of information industry and this has led to the growing importance of research in this area. Some trends shows that the current content distribution networks (CDN) are in a dire need for some sort of improvements and new frameworks need to be designed in order to sustain the viability of interactive multimedia services in a broadband context, in terms of scalability and QoS.

* This work was supported in part by the Brain Korea 21 (BK21) project from the Korean Ministry of Education, the Korean Institute of Science and Technology Information (KISTI) Contract No. N10007 (Research of Video-on Demand Technology for Cable Broadcasting Operator with Cloud Computing) and the Seoul R&BD Program (No. PA090720).

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 81–86.
springerlink.com

Cloud computing, the latest buzzwords in the technology sector, describes a new technology where services, infrastructures and platforms can be delivered over the high-speed network anywhere, anytime and on any device [1-3]. As companies increasingly look for resources efficiencies, this hugely popular technology has revolutionized how multimedia content is being delivered and consumed. Many cable and IPTV operators leap aboard the cloud computing bandwagon from a green computing angle. Why? IPTV operators find it hard to compress HD content below 6 Mb/s, and 3DTV and HDTV cannot run on 12 Mb/s over a standard DSL line. Cable operators, though able to run true HD in a single stream or 3DTV in multiple streams, have to fight against rival IPTV operators to deliver high quality multimedia content as technologies matured.

Historically, these operators has been drive testing vast areas, spending huge effort and money trying to determine which technology would fit what services. However, recently, operators began to focus on how to make more revenues through adding new services, introducing personalization and applying green tech initiatives.

2 Overview

This paper discusses how interactive multimedia content can be delivered in a more efficient and cost effective ways using associated cloud computing technologies such as those presented in [4-7]. We present a new interactive middleware for providing interactive multimedia services on the cloud, to enable subscribers receive and share web TV, social TV, mobile TV and VoD streams from wherever, whenever and however. The iCloudMedia framework is a new framework that delivers and sends "push" on-the-fly changes and "pulls" rich metadata content by both, the operators and the subscribers, through resources availability using an interactive menu in an open market. We strongly believe that innovative content branding and configurable resource management is the next logical step for operators to increase their revenues and for subscriber to obtain optimal viewing experience.

3 *i*CloudMedia

Fig. 1 shows the proposed system architecture of a cloud-based interactive multimedia service. Multimedia content, streamed from any of the virtual media servers, are being delivered according to subscribers' requests based-on resources availability, through *i*CloudMedia.

*i*CloudMedia business model is shown in Fig. 2. A subscriber(s) is able to change his/her reservation and selects any available resource(s) offered in the market through a reservation system which manages incoming service request.

The services available are listed as follows:

- view available resources
- add /remove resources
- move resources to other locations

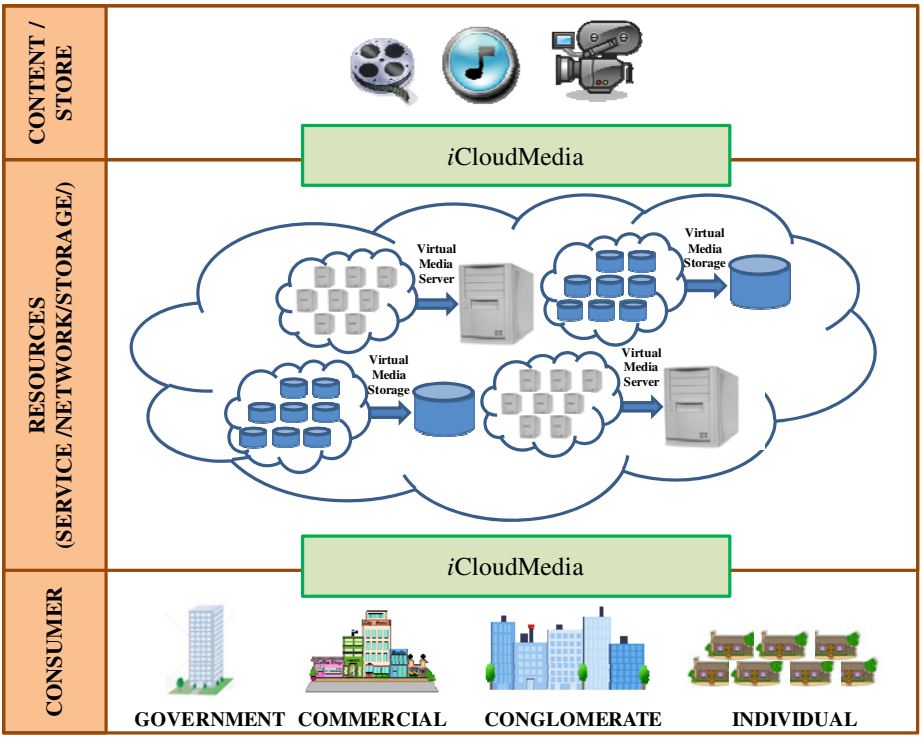


Fig. 1. *i*CloudMedia provides a unified control platform to all multimedia services.

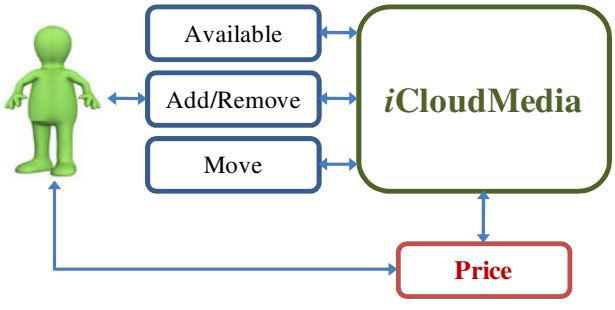


Fig. 2. *i*CloudMedia business model.

Fig.3 describes the how all incoming service requests can be granted using a dynamic management system which allocates available resources efficiently and economically in a coordinated fashion. It ensures that interactive multimedia services are being delivered according to request and continuous high utilization of resources can be achieved, at the same time.

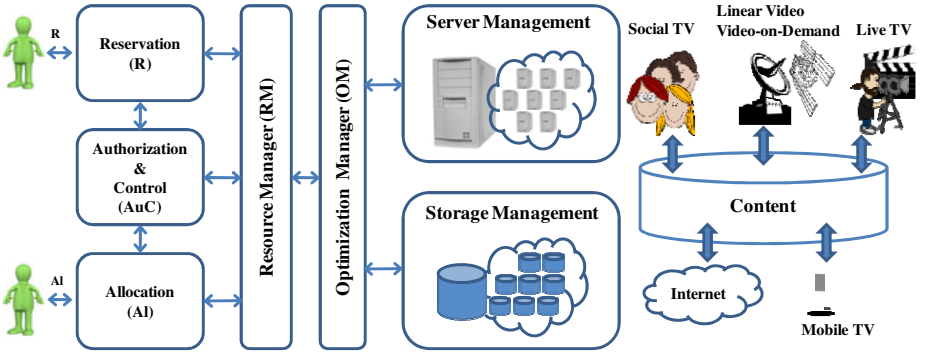


Fig. 3. *i*CloudMedia provides a unified control platform for all multimedia content.

Fig. 4 shows an example of the user interfaces for adding storage, among other resources such as data rate or service plan in an open market using *i*CloudMedia. *i*CloudMedia is flexible and is able to provide the 1) operator the ability to market resources and to introduce new revenue streams through advertisements and promotions 2) subscriber the ability to purchase resources and personalized their content/application(s).

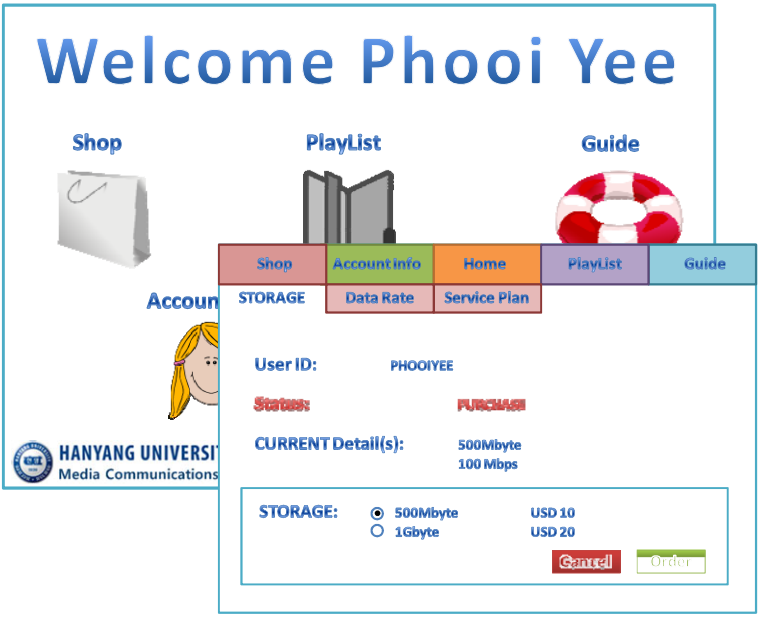


Fig. 4. Sample user interface using *i*CloudMedia for adding resource(s) such as storage on a TV screen – among other resources in the marketplace

We list and categorize the *i*CloudMedia system components as follows:

- *Shop* – consists of resource(s) for sale based on availability offered by service operator
- *Account Info* – consists of account and personal information, managed by the subscriber
- *Home* – consists a list of available services, such as
  1) "push" content such as program guide including seasonal promotions, advertisements and sponsorships,
  2) "pull" content such as Internet, applications, games, telecommunication services, managed by service operator
- *Playlist* – a list of personal content/application(s) stored in the account, managed by the subscriber
- *Guide* – consists of help menu and utilities, managed by the service operator

4 Conclusions

The rapidly evolving digital media delivery has open up new standards and services with the goal of improving digital media experiences for subscribers. Currently, classical media companies are losing their role as the sole providers of home entertainment with clear indication that the current home video framework will not be able to meet the challenges of the scalability, high quality, and real-time distribution of legal video content.

The *i*CloudMedia presented here is more than a mere program list and can be used to reconcile media companies as providers of home entertainment. It built on the concept of *store-share-download-interact* for multimedia contents as well as *retail-need-service-satisfy* for resource(s), all in a same user interface, straight to the television real-time. From the subscriber's point of view, this framework provide an easy, flexible and ubiquitous platform for subscriber to "share" and operator to "retail", as they can now share various multimedia contents – which would become a crucial marketing effort for the survival of Pay-TV services.

The proposed framework is going to revolutionize the digital content industry through green tech initiatives for cloud subscribers. For future work, the framework can be extended to include optimization solutions for interactive resource(s) market to increase the efficiency of virtual media server and virtual storage.

References

1. Amazon.: Amazon elastic compute cloud, Amazon EC2 (2009), `http://aws.amazon.com/ec2/`
2. Ghemawat, S., Gobioff, H., Leung, S.T.: The Google file system. In: 19th ACM Symposium on Operating Systems Principles, New York, pp. 29–43 (2003)
3. Boss, G., Malladi, P., Quan, D., Legregni, L., Hall, H.: Cloud computing. In: IBM White Paper (2007)

4. Pallickara, S., Ekanayake, J., Fox, G.: Granules: A lightweight, streaming runtime for cloud computing with support, for Map-Reduce. In: IEEE International Symposium on Cluster Computing and the Grid Workshops, New Orleans, pp. 1–10 (2009)
5. Sukaridhoto, S., Funabiki, N., Nakanishi, T., Pramadihanto, D.: A comparative study of open source softwares for virtualization with streaming server applications. In: IEEE International Symposium on Consumer Electronics, Kyoto, pp. 577–581 (2009)
6. Van, H.N., Tran, F.D., Menaud, J.-M.: Autonomic virtual resource management for service hosting platforms. In: Workshop on Software Engineering Challenges of Cloud Computing, Vancouver (2009)
7. Lau, P.Y., Park, S., Yoon, J., Lee, J.: Pay-as-you-use on-demand cloud service: an IPTV case. In: 2010 International Conference on Electronics and Information Engineering (ICEIE 2010), Kyoto, Japan, vol. 1, pp. 272–276 (2010)

Study and Practice of an Improving Multi-path Search Algorithm in a City Public Transportation Network

Xiangyang Feng, Xiaofei Zheng, and Cairong Yan

School of Computer Science and Technology
Donghua University
No. 1882,West Yanan Road, Shanghai, China 200051
xiaofei_352@yahoo.com.cn, cryan@dhu.edu.cn

Abstract. Shortest path search is one of key problems for big-scale city public transportation network (CPTN) query system. Based on the current main search algorithms and data models on multi-path search, an improving multi-path search algorithm combined A* and deviation path was proposed in this paper. With our algorithm, not only the optimal path could be provided, but also Kth approaching optimal paths could be displayed. Due to the analysis on algorithm complexity and evaluation experiment, the algorithm efficiency for multi-path search is much better than typical Djikstra algorithm and collection algorithm. Thus, it is more suitable on shortest path search for big-scale real-time CPTN.

Index Terms: shortest path, A* algorithm, Deviation path algorithm, CPTN.

1 Introduction

The current common shortest path search algorithms applied to search the shortest path in a CPTN are normally concerned with these algorithms: improved algorithms based on the traditional Dijkstra algorithm, collection algorithms with a least transfer times through set intersection operations, and AI search algorithms based on genetic algorithms [1-3].

Dijkstra algorithm, which was proposed for single-source problem at 1959, is the most popular algorithm to search the shortest path between two-nodes in a topology network. Its result not only includes the shortest path from the beginning to the destination node, but also includes all the shortest paths to other nodes in the network [1]. Though Dijkstra algorithm has many advantages, such as stability, adaptability of variation on network topology and low requirement on system memory, it is not suitable to obtain the feedback result of all the shortest paths among all node pair of n nodes in a certain time, especially, with the number of node increasing, its computing complexity will be $o(n^3)$. Dijkstra algorithm can be generally applied to search shortest paths in a certain time on a public transportation network in a city with middle and small scope. However, some public transportation system is becoming more and more big such as Beijing, Shanghai and etc. At present, some algorithms such as the collection algorithm [2] and the ant algorithms [3] have been studied to search the shortest path in time with reducing the computational complexity. The collection algorithm is

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 87–96.
springerlink.com

studied to search the shortest path with the least transfer time by set intersection operation among public traffic lines in CPTN, of course, there are many methods to prescribe the order of set intersection operation to obtain the shortest path with the least transfer time between arbitrary given two nodes, its complexity is the problem of $o(q^m)$, with m increasing, in which q is the average number of transfer nodes on a traffic line and m might be the total number of traffic lines in CPTN. Many algorithms like the collection algorithm restrict the value of m so that a result can be fed back in a reasonable time [3]. Many researchers have proposed genetic algorithms for the shortest path based on AI. A typical representative is the ant algorithms [4]. However, in order to reduce search scope and to improve search speed, genetic algorithm removes some little possibility nodes by using feedback mechanism. Though it has some stability degree for a dynamic complex network, the accuracy of studied result is not very reliable because of the possibility losing optimal solution occurred.

In order to solve the problem of search K^{th} shortest path in CPTN with big scope more effectively, a new optimum multi-path search algorithm has been proposed in this paper by combining the A* algorithm with the deviation path algorithm [2-4], it improved the efficiency of multi-path search in CPTN greatly without losing the accuracy of the result.

2 Characteristics and a Topology Structure of Urban Public Transportation Network

A CPTN of a modern city may be composed of rail transit network (CRTN) and public bus routes (PBR), of course, including taxis. A topology structure of PBR is normally more complex than a topology structure of CRTN, because there are mainly these causations:①The number of bus routes is greater than the number of rail transit line; ②Bus routes have different characteristics in run way such as a symmetrical run way which means that the bus station names are the same in run way up and down, and an unsymmetrical run way which means that there are some different bus station names in run way up or down. So different models will be considered and built depending on the characteristics of bus run ways; ③There are some different or the

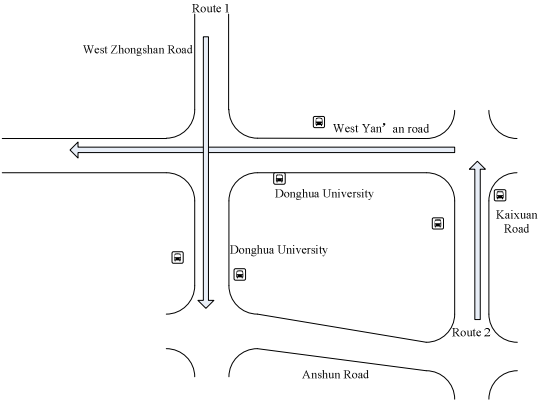


Fig. 1. A part of the actual bus routes' layout

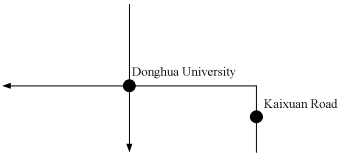


Fig. 2. Geometry layout after merging same stations in Figue1.

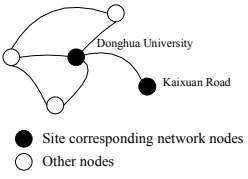


Fig. 3. An abstrct topology structure of Figure 2

same bus station names which are neighbored on a geographical place, and which will bring difficulties to build models of bus routes. It is necessary on PBR to merge these neighbored stations into a virtual unique station code [6] separately.

Figure 1 shows the bus station names neighbored Donghua University in Shanghai that indicates the same geographical place, i.e., Donghua University. Figure 2 shows the two same bus stations in Figure 1 have been merged to a unique bus station. Figure 3 is the abstract topology structure in terms of Figure 2.

3 Modelings of PBR and the Data Structures

PBR can be symbolized as G (V, E, L), in which,V represents the node set of PBR, E represents path section collection of the bus routes (section between two bus station nodes), L represents the collection of bus routes (each bus route includes specific sequence of nodes and sections). The radix of V is m, which means that there are m different nodes in PBR. The radix of L is i, which represents the amount of different bus routes. The up-route and down-route belonging to the same bus route are treated as two independent bus routes. The beginning node and the destination node of ring route are considered as the same node.

Bus route Li can be considered as an orderly collection of n stops:

Li = <S_1, S_2, S_3, S_4 ... Sk | S $\in$ V, k represents the number of stops belonging to route Li.>

Supposed the model as following:

1) Considering up-route and down-route as two different routes.
2) There are no stops which have different names but close to each other should be merged.
3) Regardless of the influence of the departure frequency.
4) Regardless of vehicle types, such as air bus general bus, etc.
5) Regardless of bus route type, travel time is proportional to the number of the passed stops.

In order to carry out algorithm computing, many researchers abstract PBR as adjacency matrix [4]. However, adjacency matrix is suitable for sparse network computing. For major cities such as Beijing and Shanghai, since adjacency matrix formed by PBR must be huge and dense, it should not meet the requirement of real-time computing. On the other hand, with the rapid development of geographic information system (GIS), there has been commercial GIS spatial database currently. Considered the characters of platform independence on spatial data storage, the reliability on data query optimization, GIS spatial database is more appropriate for super big scale public transit data storage. The main table structure is shown as Figure 4.

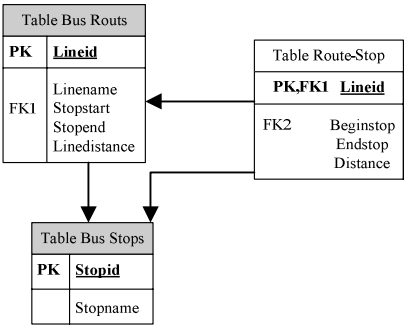


Fig. 4. Database table structure.

Bus Route table is used to store all bus routes, including line ID, line name, start station, terminal station, route mileage, etc. Route-Stop table is used to store information about stop and adjacent stop, including stop ID, adjacent stop number, stop distance and line belongs to. Bus stops table store information of all bus stops, including stop ID, stop name and other attaching information.

4 Search Algorithm Analysis

A * algorithm is a kind of graph search strategy in artificial intelligence. It applies heuristic search function to evaluate the emerging branches during search process, so as to select the most optimal one for searching. Since it effectively reduce search nodes and branches, A* algorithm improves search efficiency significantly. Therefore, A* algorithm is better than traditional Dijkstra algorithm, particularly in big scale network search [4]. Traditional shortest path search algorithm based on deviation path uses classic Dijkstra algorithm to compute the optimal solution firstly, and then seek its deviation path set to get the Kth shortest path collection. As above mentioned, when Dijkstra algorithm is applied on path search against big scale network, its efficiency would decline. Although the optimal solution can be guaranteed, it is obviously that the way of traverse all nodes does not meet the requirement of real-time computing. Thus, A* algorithm tries to improve it by using knowledge-based feedback. The basic definition is as following.

$$f(v)=g(v)+h(v)$$

Among them, f (v) is defined as the heuristic evaluation function of the current node v, g (v) is the actual cost from the initial node to node v in network; h (v) is the estimated cost along the optimal path from node v to the target node. When h (v) equals zero, which means not using any heuristic information, A * algorithm would be equivalent to standard Dijkstra algorithm.

In order to meet diversity of passenger demand actually, not only the shortest path query should satisfy in mass transit network, but also limited Kth shortest path should be provided for passenger's selection. Because the feature of multi-path algorithm requests multiple shortest paths computing, its efficiency would reduce with the search depth increasing. Our proposal firstly obtained the shortest path by a one-time computing using A * algorithm, then also applied A* algorithm to seek deviation path collection. It is proved that the combination of A * algorithm and deviation path algorithm could better meet the search result of Kth shortest path while ensuring the search efficiency. The efficiency comparison of A * algorithm combined with deviation path algorithm, simple Dijkstra algorithm, and simple A * algorithm is shown in table 1. It shows a smaller lose of efficiency can get the accurate search results of the pre-K times shortest path. So the requirement of multi-path selection is well met by A * algorithm combined with the deviation path algorithm.

Deviation path is defined as following, v_1 is the source node, v_n is the destination node, and $v_1v_2v_3v_4$ ... v_n and $v_1v_2'v_3'v_4$ '... v_n are two paths between (v_1, v_n). If $v_1v_2v_3v_4$ ... v_i and $v_1v_2'v_3'v_4'$ ... v_i ' are each corresponding to the same nodes, and get different nodes from v_{i+1}, the latter path is called that the path of $v_1v_2'v_3'v_4$ '... v_n is obtained deviated from the path of $v_1v_2v_3v_4$ ... v_n at v_{i+1}. It is one of deviation paths of $v_1v_2v_3v_4$ ... v_n. The path of $v_1v_2v_3v_4$ ... v_n treated v_{i+1} as deviated node. There exists a shortest one among all of deviation paths that deviated from node v_{i+1}', which consists of $v_1v_2v_3v_4$ ... v_i plus link v_i v_{i+1}' plus the shortest path from v_{i+1}' to v_n. It is defined as the 2th shortest path. And so on could get the Kth shortest path [5]. As shown in Figure 5, the path 1235 is the 2th shortest path of path 145, 12365 is the 3th shortest path that deviated node is 4 and deviating node is 2.

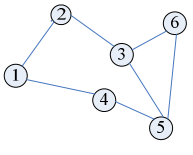


Fig. 5. Deviation Route

5 Algorithm Achievement

1) Firstly, the shortest path from the initial node v_1 to the target node v_n can be got by using A* algorithm. OPEN and CLOSE are two linked lists. OPEN list stores nodes that have been generated but not yet generate successor nodes set from. CLOSE list stores nodes that have generated successor nodes set from, follow these steps:

① Initial OPEN list only contains the original node, initial CLOSE list is empty. Origin's weight is zero, and the other nodes' weight is infinity.

② If OPEN list is empty, it means the search fails. Otherwise, the node with the smallest value f will be selected as the most festive node (named BEST) from OPEN list, and be moved to CLOSE list. If node BEST is the target node, turns to step 3). Otherwise the successor nodes of BEST will be generated based on connected path sections of BEST. For each successor node n, the following procedure should be executed.

Calculate the weight value f to reach node n.

If node n is equal to a node in OPEN list (named m), determines whether it has the lowest cost value. If it is true, replaces the cost value of m with the cost value of n, and points the backward pointer of m to BEST.

If n matches a node in CLOSE list (named m'), compares whether it has the lowest cost value. If it is true, replaces the cost value of m' with the cost value of n, points the backward pointer m' to BEST, and moves m' to OPEN list.

If n is neither in OPEN list nor in CLOSE list, points the backward pointer of n to BEST, moves n to OPEN list, and calculates n's evaluation function f (n) = g (n) + h (n). Repeat ②.

③ Backtracks from node BEST, treats the searching path found as the shortest path.

2) Initializes the deviation path collection G empty. Deals with the shortest path node sequence $v_1v_2v_3v_4$ ... v_n which are outputted by step 1), derives deviation path from deviated nodes v_i ($2 \le i \le n$). Puts all deviation paths into G, and notes deviated node (named devNode) of each deviation path. If G is empty, terminates the search.

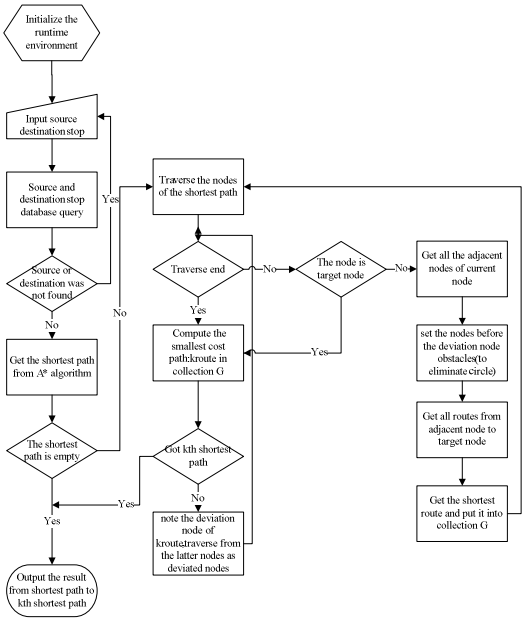


Fig. 6. Algorithm flow chart

3) Sorts the paths in collection G. The shortest path is 2rd shortest path, and derives deviation path from successor nodes of devNode (deviated nodes). Note devNode of every deviation path. Puts the generated deviation path into G, and removes 2th shortest path from G. If G is empty, terminates the search.

4) Supposed that the K-1th shortest path has been obtained. Its deviated node is devNode. The Kth shortest path is the shortest one in G, and derives deviation path from successor nodes of devNode (deviated nodes). Notes devNode of every deviation path. Puts the generated deviation path into G, and removes the Kth shortest path from G.

5) Do not set quit boundary. Repeats step 4). Gets all paths between (v_1, v_n).

The algorithm flow chart was shown in Figure 6.

The core implement program as follows:

```
private GLine DoPlan(IList<StopPoint> route,
StopPoint o, StopPoint d)
{
  foreach (StopPoint p in route)
  {
    if(p.Equals(d)){ break; }
    if(!p.Equals (o))
    {
      //Initialize startNode with p and
    CompassLinesHelper clhelper
      IList<CompassLines> allCompassLines =
clhelper.GetAllCompassLines(startNode);
      foreach (CompassLines cline in
allCompassLines)
      {
       //Get adjacent point pnext and check if it
    is in shortest
        if (pnext in shortest)
        {
         //Get the shortest path from pnext to
destination use AStar
        }
      }
      //Get the shortest path and put into
collection G
    }
  }
  //G is not empty
    foreach (GLine gl in DeviationPlan.G)
    {
      //Tranverse collection G to search the
shortest path
    }
     return kroute;
}
```

6 Algorithm Complexity Analysis

Time complexity of Dijkstra algorithm is O (n^2). In our improving A* algorithm, if the average out degree of stop is b and the search depth from the beginning to the end is d, then the time complexity is O(bd). The Kth shortest path is the shortest path in G (see Section 5). Since each deviation path computing calls A* algorithm one time, the time complexity of the algorithm is limited to O(e × bd) after combined with deviation path. The e is total number of routes in the network. In actual calculation, the number of deviation paths was far less than e [7].

7 Algorithm Efficiency Evaluation

In order to compare the efficiency of three above mentioned algorithms, a database contains part of Shanghai bus routes was established. The execution efficiencies of these three algorithms are listed in Table 1 and Table 2. For comparison, actual algorithm used a variety of data structures, mainly are based on adjacency matrix and adjacency list. When adjacency matrix structure was used to store stops, due to big dimension of the matrix, the algorithm initialization required very big time expense. Therefore, due to the poor maintainability of this algorithm, it was not suitable for big-scale PBR shortest path searching. On the other hand, since the adjacency list read the information of adjacent stops dynamically, it was instantiated only when needed, and the practical efficiency of the algorithm was greatly improved.

Figure 7 showed the efficiency curve based on the algorithm execution time in table 1. Coordinate X axis represented data size (the number of bus routes), Y-axis described algorithm execution time (in millisecond). To use Dijkstra algorithm and A * algorithm in 30, 60, 90, 120, 150 routes respectively, the key points in two curves represented computing time on the shortest path search from Taopu highway to Jiangning road. In accordance with the curves, it was observed directly that the efficiency of A* algorithm was higher than that of Dijkstra algorithm. In the case of the shortest path computing for 30 routes, Dijkstra algorithm needed 774ms, while A * algorithm only needed 12ms. For 60 routes, Dijkstra algorithm needed 1710ms, while A * algorithm only needed 15ms. From the curves, the execution time of Dijkstra

Table 1. The shortest path algorithm efficiency comparison

	Dijkstra	**A***
30 Bus Routes (Stops number:356)	774ms	12ms
60 Bus Routes (Stops number:628)	1710ms	15ms
90 Bus Routes (Stops number:864)	2345ms	15ms
120 Bus Routes (Stops number:991)	4235ms	29ms
150 Bus Routes (Stops number:1118)	5180ms	32ms

emark: Taopu Highway---Jiangning Road(one transfer).

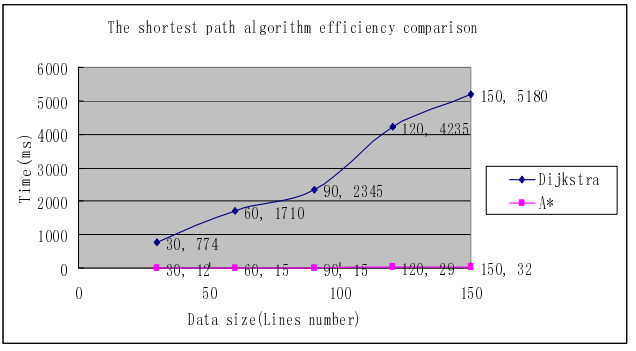


Fig. 7. The shortest path algorithm efficiency curve

algorithm was proportional to grow with bus route number increase, while the efficiency decline of A * algorithm was not obvious. Therefore, A * algorithm was more efficient than Dijkstra algorithm against PBR network search.

Figure 8 described the efficiency curve based on the algorithm execution time in table 2, while the axis meaning was the same as in Figure 7. It mainly reflected the difference against execution time on the shortest path using pure A * algorithm and our A * combined deviation path respectively.

Table 2. Kth shortest path algorithm efficiency comparison (k=3)

	A*	**A* combined with Deviation**
30 Bus Routes (Stops number:356)	1104ms	561ms
60 Bus Routes (Stops number:628)	2181ms	1990ms
90 Bus Routes (Stops number:864)	3224ms	2174ms
120 Bus Routes (Stops number:991)	4475ms	2430ms
150 Bus Routes (Stops number:1118)	5620ms	2884ms

The key points in two curves represented computing time on the 3th shortest path searching from Shanghai Stadium to Jingan Temple using Dijkstra algorithm and A * algorithm in 30, 60, 90, 120, 150 routes respectively. From two curves' trend, the execution time of these two kinds of algorithm both increased rapidly during the data size from 30 to 60 routes, However, when data size was greater than 60 routes, the efficiency of pure A * algorithm decreased proportionately, the efficiency of A * algorithm combined with the deviation path had no significant drop. Thus, A * in combination with the deviation path shows higher efficiency when computing Kth shortest path from source stop to destination stop. It is more suitable for practical application in the path selection.

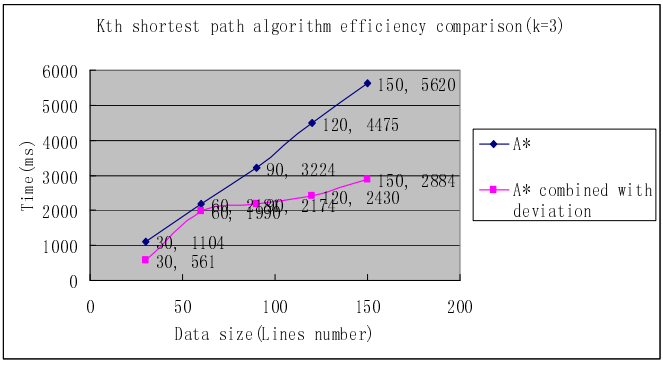


Fig. 8. Kth shortest path algorithm efficiency curve

8 Conclusions

Based on the BPR modeling, this paper proposed a new improved multi-path search algorithm. It combined both A * algorithm and deviation path algorithm. The result not only met the shortest path search of mass transportation network, but also ultimately obtained the Kth shortest path. Through the analysis of different shortest path search algorithms, our algorithm provided various decisions when ensuring the search efficiency. Thus, it is a better query solution on BPR path real-time selection, and has a good practical value. The future work is to extend to adapt more complex queries, such as the requirements of different routes, calculating the different costs, adding mix query capability against rail and bus systems.

References

[1] Wang, L., Cao, S.: Choose a city bus lines travel based on breadth first. In: The Second Intelligent Computing Conference, pp. 3–6 (2008)
[2] Zhou, C., Li, S.: Research of Dijkstra algorithm and A * Algorithm. Software Guide 37(2), 32–36 (2007)
[3] Fu, J., Li, X., Chen, J.: A Practical Algorithm for Finding the Shortest Kth Path Based on Deviation Path. Computer Technology and Development, 1–4 (February 2009)
[4] Wu, M., Su, H., Wang, M.: Research And Implementation for the Algorithm of kth Approaching Shortest Path. Computer Applications and Software, 1–3 (August 2004)
[5] Liang, H., Yuan, X.-q., Liu, R.: Novel model and realization of public transport route inquiring system. Computer Engineering And Applications 43(3), 15–43 (2007)
[6] Mokhtar, A., Azizo gou, M.: Adaptive wavelength routing in all = optical networks. IEEE/ACM Trans. on Networking 7(5), 767–778 (1999)
[7] Bielli, M., Boulmakoul, A., Mouncif, H.: Object modeling and path computation for multimodal travel systems, 10–13 (April 2005)

An Optimal Combination Weights Method Considering Both Subjective and Objective Weight Information in Power Quality Evaluation

Wei Chen[1,2] and Xiaohong Hao[1,2]

[1] College of Electrical and Information Engineering
Lanzhou University of Technology, LZUT
Lanzhou, China
[2] Key Laboratory of Gansu Advanced Control for Industrial Processes
Lanzhou, China
chenlin@lut.cn

Abstract. Power Quality information is not only an important part of AMI (Advanced Metering Infrastructure) in smart grid but also is an important part of technical support system of the electricity market, so the power quality monitoring and evaluation is necessary of smart grid development and operation of power market. based on subjective weighting method (analytic hierarchy process, An expanded least deviations algorithm has been used to combine the multiple expert judgment matrixes to get the subjective weight vector in AHP) and objective weighting methods(variation coefficient method has been used to get the objective weight vector), an optimal combination determining weights method resulting in consistency between subjective and objective evaluation in the least squares sense is present. then bring out a improved fuzzy synthetic evaluation method based on reliability code. A new hierarchy model for power quality evaluation is presented to quantify and evaluate the power quality. Case results clearly show that the proposed method is attractive and effective.

Keywords: smart grid; power quality; analytic hierarchy process; variation coefficient method; expanded least deviations algorithm; optimal combination weights method.

1 Introduction

Power Quality information is not only an important part of AMI (Advanced Metering Infrastructure) in smart grid and is an important part of technical support system of the electricity market, but also is one of the constraints in making tariff policy[1],[2],[3],[4]. Improving the quality of power is also an important part of auxiliary services of electricity market.so the power quality monitoring and evaluation is necessary of smart grid development and operation of power market[5],[6].

The power quality evaluation method is mainly the fuzzy comprehensive evaluation method[7],[8],[9], shepard model[10], projection pursuit[11], entropy principle[12], and the matter-element analysis[13], attribute recognition theory[14], etc.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 97–105.
springerlink.com

The comparative analysis is carried out based on same measured data, which shows there is uncertainty in the evaluation results which relay on the subjective factor influence of evaluation process[15]. Fuzzy Comprehensive evaluation is one of the most widely used methods in the decision-theoretic, but it is usually be influenced significantly by the matrix of fuzzy relation and index vector. For a sequential segmentation category, the principle of the lowest cost, the principle of maximum degree of measure and the principle of maximum degree of membership sometimes can get unreasonable conclusion, even sometimes can get error conclusion, because they conceal the difference of two degree of membership[16]. In this paper, A new hierarchy model for power quality evaluation is presented to quantify and evaluate the power quality, based on subjective and objective weighting methods,this paper presents a combinational evaluation method resulting in consistency between subjective and objective evaluation in the least squares sense, which make comprehensive evaluation result is more reasonable and credibility than single assessment method. then bring out a improved fuzzy synthetic evaluation method based on reliability code. The proposed method can overcome the shortages of the traditional fuzzy synthetic evaluation, Case results clearly show that the proposed method is attractive and effective.

2 Improved Fuzzy Comprehensive Evaluation

2.1 Optimal Algorithm for Combining Subjective and Objective Weight Information

Subjective Weighting Method. In the process of the fuzzy evaluation, determining the proper weight is one of the most important procedures and has direct impact on the results of comprehensive evaluation. The judgment matrix method of the AHP（Analytic Hierarchy Process）is used in this paper. AHP is a method that through the analysis of complex systems and relationship between the factors contained, and then the system will be broken down into different elements, and these elements are incorporated into different levels, so as to form a multi-level analysis model objectively. According to a certain scaling theory, all the elements of each level will be compared so as to get the comparative scales indicating relative importance of the elements and establish the judgment matrix. By calculating the maximum eigenvalue of judgment matrix and the corresponding eigenvector to get the orders of the elements of each level to a certain element from the upper level, and thus the weight vector is determined. For the detailed algorithm about AHP can get in [14][17].

For a number of experts or a group of experts involved in evaluation process, the following expanded least deviation algorithm[18] can be used to combine the multiple expert judgment matrixes to get the final weight vector. The algorithm in one hand can integrate the experts' judgment information to the utmost; in another hand can reduce the dependence on individual expert.

Suppose $\boldsymbol{P}_l = [b_{ij,l}]_{n\times n}$, $l = 1, 2, \cdots\cdots m$ is the judgment matrix group, m is the number of the evaluation experts, $\boldsymbol{P}^* = [\omega_i / \omega_j]_{n\times n}$ is the weight matrix to be solved;

$\boldsymbol{D}=\left\{\boldsymbol{\omega}\in \boldsymbol{R}^n \left| \sum_{i=1}^{n}\omega_i=1,\omega_i\geq 0,i=1,2,\cdots\cdots n\right.\right\}$ is weights vector set,

$\boldsymbol{\Omega}=\left\{\boldsymbol{\lambda}\in \boldsymbol{R}^m \left| \sum_{i=1}^{m}\lambda_l=1,\lambda_l\geq 0,l=1,2,\cdots\cdots m\right.\right\}$ is weighted vector set;

$\boldsymbol{e}=(1,1,\cdots\cdots 1)^{\mathrm{T}}\in \boldsymbol{R}^n$.

Suppose the Disturbance Matrix

$\boldsymbol{E}_l=[\varepsilon_{ij,l}]_{n\times n}=[b_{ij,l}\dfrac{\omega_i}{\omega_j}]_{n\times n}$, $l=1,2,\cdots\cdots m$, Constructing the least squares model as follow:

$$\min f(\boldsymbol{\omega})=\frac{1}{2}\sum_{i,j=1}^{n}\sum_{l=1}^{m}\lambda_l(b_{ij,l}\frac{\omega_i}{\omega_j}+b_{ji,l}\frac{\omega_i}{\omega_j}-2)\qquad \boldsymbol{\omega}\in \boldsymbol{D},\ \boldsymbol{\lambda}\in \boldsymbol{\Omega} \tag{1}$$

This solution to the above optimization problem is given by the Lagrange function:

$$L(\boldsymbol{\omega},k)=h(\boldsymbol{\omega})+k(\sum_{i=1}^{n}\omega_i-1)=\frac{1}{2}\sum_{i,j=1}^{n}\sum_{l=1}^{m}\lambda_l(b_{ij,l}\frac{\omega_i}{\omega_j}+b_{ji,l}\frac{\omega_i}{\omega_j}-2)+k(\sum_{i=1}^{n}\omega_i-1) \tag{2}$$

where $k>0$ is the Lagrange multipliers，then making $\dfrac{\partial L}{\partial \omega_i}=0$,one can obtain

$$\sum_{j=1}^{n}\sum_{l=1}^{m}\lambda_l(b_{ji,l}\frac{\omega_i}{\omega_j}-b_{ij,l}\frac{\omega_i}{\omega_j})+k\omega_i=0 \tag{3}$$

We can obtain $k=0$， then $\boldsymbol{\omega}$ must satisfy:

$$\sum_{j=1}^{n}\sum_{l=1}^{m}\lambda_l(b_{ji,l}\frac{\omega_i}{\omega_j}-b_{ij,l}\frac{\omega_i}{\omega_j})=0 \tag{4}$$

$$\text{Suppose } \varphi_i(\boldsymbol{\omega}(k))=\sum_{j=1}^{n}\sum_{l=1}^{m}\lambda_l(b_{ji,l}\frac{\omega_i(k)}{\omega_j(k)}-b_{ij,l}\frac{\omega_i(k)}{\omega_j(k)}),\quad i=1,2,\cdots\cdots n \tag{5}$$

The steps of using expanded least deviations algorithm for combining index weights are as follows:

Step1: give failure value $\varepsilon>0$，let $k=0$ ，give initial value at random $\boldsymbol{\omega}(0)=(\omega_1(0),\omega_2(0),\cdots,\omega_n(0))$.

Step2:compute $\varphi_i(\boldsymbol{\omega}(k))$, $i=1,2,\cdots\cdots n$,if $|\varphi_i(\boldsymbol{\omega}(k))|$<0 for all i , accept the iteration value, it shows $\boldsymbol{\omega}(k)$ is mineral solution of $h(\boldsymbol{\omega})$, end。or wise , turn to step3。

Step3:suppose $|\varphi_s(\boldsymbol{\omega}(k))|=\max|\varphi_i(\boldsymbol{\omega}(k))|$, then compute

$$g(k)=(\sum_{j\neq s}\sum_{l=1}^{m}\lambda_l(b_{sj,l}\frac{\omega_j(k)}{\omega_s(k)})\Big/\sum_{j\neq s}\sum_{l=1}^{m}\lambda_l(b_{js,l}\frac{\omega_s(k)}{\omega_j(k)}))^{\frac{1}{2}} \tag{6}$$

Making $$\phi_i(k) = \begin{cases} g(k)\omega_s(k), & i = s \\ \omega_i(k) & i \neq s \end{cases} \tag{7}$$

$$\omega_i(k+1) = \phi_i(k) \Big/ \sum_{j=1}^{n} \phi_j(k), \quad i = 1, 2, \cdots\cdots n \tag{8}$$

Step4: let $k = k+1$, back step2.

Objective Weighting Method. Objective weighting method is based on the actual data via mathematical method to get the index weight. generally speaking, they are based on the index variation degree or the relationship to determine the importance of index, objective weighting method due to its absolute objectivity and may violate the index economic significance (or technical significance). at the same time, the samples change may cause weight change,and the weight is unstable. In order to highlight the index of relative variation and variation coefficient of variation, Variation Coefficient Method is be adopted in the paper and its calculation steps:

Step1: For an evaluation matrix $X = (x_{ij})_{m\times n}$, m is evaluated object number, n is the index number. the first thing is to make indexes the same direction, then Normailize and make it standardized matrix $Z = (z_{ij})_{m\times n}$.

$$z_{ij} = x_{ij} \Big/ \sqrt{\sum_{i=1}^{m} x_{ij}^2}, \quad i = 1, 2, \cdots\cdots m, \quad j = 1, 2, \cdots\cdots n \tag{9}$$

Step2: For standardized matrix $Z = (z_{ij})_{m\times n}$, compute mean and variance of samples:

$$\overline{z_j} = \frac{1}{m} \sum_{i=1}^{m} z_{ij}, \quad j = 1, 2, \cdots\cdots n \tag{10}$$

$$s_j = \sqrt{\frac{1}{m-1} \sum_{i=1}^{m} (z_{ij} - \overline{z_j})^2}, \quad j = 1, 2, \cdots\cdots n \tag{11}$$

Step3: According to the mean and variance get variation coefficient:

$$v_j = s_j \Big/ \left|\overline{z_j}\right|, \quad j = 1, 2, \cdots\cdots n \tag{12}$$

Step4: The weight of each index is obtained according to the variation coefficient:

$$\omega_j = s_j \Big/ \sum_{j=1}^{m} v_j, \quad j = 1, 2, \cdots\cdots n \tag{13}$$

Optimal Algorithm for Combining Weight Information. Using subjective weighting method to determine the weight coefficient of index can reflect policymakers inclinations, but evaluation result is very uncertain.using objective weighting method to determine the weight coefficient of index has mathematical theory basis, but it ignores policymakers inclinations. In order to let the decision-making process and result more scientific, a reasonable approach is to combine the weight informations

from the different weighting methods according to certain optimal algorithm.So the results not only can reflect the subjective and objective information, but also can make the unity of the value and information. Based on least-square method,a combinatorial optimization model[19] to determine the index weight is established.

Suppose the index weight vector using the subjective weighting method is $\boldsymbol{u}=[u_1,u_2,\cdots,u_n]^{\mathrm{T}}$, the index weight vector using the objective weighting method is $\boldsymbol{v}=[v_1,v_2,\cdots,v_n]^{\mathrm{T}}$, the final weight vector After optimizing is $\boldsymbol{\omega}=[\omega_1,\omega_2,\cdots,\omega_n]^{\mathrm{T}}$, standardized matrix $Z=(z_{ij})_{m\times n}$, m is evaluated object number, n is the index number.

Constructing the least squares model as follow:

$$\min g(\boldsymbol{\omega})=\sum_{i=1}^{m}\sum_{j=1}^{n}\left\{[(u_j-\omega_j)z_{ij}]^2+[(v_j-\omega_j)z_{ij}]^2\right\}$$

$$\text{s.t.}\quad \sum_{j=1}^{n}\omega_j=1,\quad \omega_j\geqslant 0,\ j=1,2,\cdots\cdots n \tag{14}$$

This optimization problem is given by the Lagrange function:

$$L(\boldsymbol{\omega},\lambda)=g(\boldsymbol{\omega})+\lambda(\sum_{j=1}^{m}\omega_j-1)$$

$$=\sum_{i=1}^{m}\sum_{j=1}^{n}\left\{[(u_j-\omega_j)z_{ij}]^2+[(v_j-\omega_j)z_{ij}]^2\right\}+4\lambda(\sum_{j=1}^{n}\omega_j-1) \tag{15}$$

where $\lambda>0$ is the Lagrange multipliers, then making $\frac{\partial L}{\partial \omega_i}=0$ and $\frac{\partial L}{\partial \lambda}=0$, can obtain

$$-\sum_{i=1}^{m}2(u_j+v_j-\omega_j)z_{ij}^2+4\lambda=0 \tag{16}$$

$$4(\sum_{j=1}^{n}\omega_j-1)=0 \tag{17}$$

set $\boldsymbol{A}=diag\left[\sum_{i=1}^{m}z_{i1}^2,\sum_{i=1}^{m}z_{i2}^2,\cdots\cdots\sum_{i=1}^{m}z_{in}^2\right]$, $\boldsymbol{E}=[1,1,\cdots\cdots 1]^T$,

$\boldsymbol{\omega}=[\omega_1,\omega_2,\cdots,\omega_n]^{\mathrm{T}}$, $\boldsymbol{B}=diag\left[\sum_{i=1}^{m}\frac{1}{2}(u_1+v_1)z_{i1}^2,\sum_{i=1}^{m}\frac{1}{2}(u_2+v_2)z_{i2}^2,\cdots\cdots\sum_{i=1}^{m}\frac{1}{2}(u_n+v_n)z_{in}^2\right]^{\mathrm{T}}$

Using matrix form

$$\begin{bmatrix}\boldsymbol{A} & \boldsymbol{E}\\ \boldsymbol{e}^T & 0\end{bmatrix}\cdot\begin{bmatrix}\boldsymbol{\omega}\\ \lambda\end{bmatrix}=\begin{bmatrix}\boldsymbol{B}\\ 1\end{bmatrix} \tag{18}$$

Solving equations (18), can obtain

$$\boldsymbol{\omega}=\boldsymbol{A}^{-1}\cdot\left[\boldsymbol{B}+\frac{1-\boldsymbol{E}^T\boldsymbol{A}^{-1}\boldsymbol{B}}{\boldsymbol{E}^T\boldsymbol{A}^{-1}\boldsymbol{E}}\boldsymbol{E}\right] \tag{19}$$

Reference [16] showed that for a sequential segmentation category, the principle of the lowest cost, the principle of maximum degree of measure and the principle of maximum degree of membership sometimes can get unreasonable conclusion, even sometimes can get error conclusion, because they conceal the difference of two degree of membership. therefore proposed the principle of reliability code.

If evaluation categories $(y_1, y_2, \cdots\cdots, y_k)$ is a sequential segmentation of the attributes space V, μ_x is the membership, here membership requires to unitary. λ is reliability code, considering the generally range of λ is $0.5<\lambda<1$, here $\lambda = 0.6 \sim 0.7$. If $y_1, y_2, \cdots\cdots, y_k$ meet $y_1 > y_2 > \cdots\cdots > y_K$, and

$$k_0 = \min\{k : \sum_{l=1}^{k} \mu_{x_i}(y_l) \geqslant \lambda,\ 1 \leqslant k \leqslant K\} \tag{20}$$

One can obtain that x_i belongs to category y_{k_0}.

If $y_1, y_2, \cdots\cdots, y_k$ meet $y_1 < y_2 < \cdots\cdots < y_K$, and

$$k_0 = \max\{k : \sum_{l=k}^{K} \mu_{x_i}(y_l) \geqslant \lambda,\ 1 \leqslant k \leqslant K\} \tag{21}$$

One can obtain that x_i belongs to category y_{k_0}.

3 Application in the Comprehensive Evaluation of Power Quality

The elements impacting the quality of power commodities are in large and complex, namely, power quality needs of a number of indicators to measure. How will the sub-indicators reasonable description and organization to reflect the quality of power together is actually a very complex multi-attribute Comprehensive evaluation and decision-making.

The hierarchical model for power quality comprehensive evaluation is established as shown in Figure 1. There are four first-level indicators of power quality; each first-level indicator has some second-level component indicators.

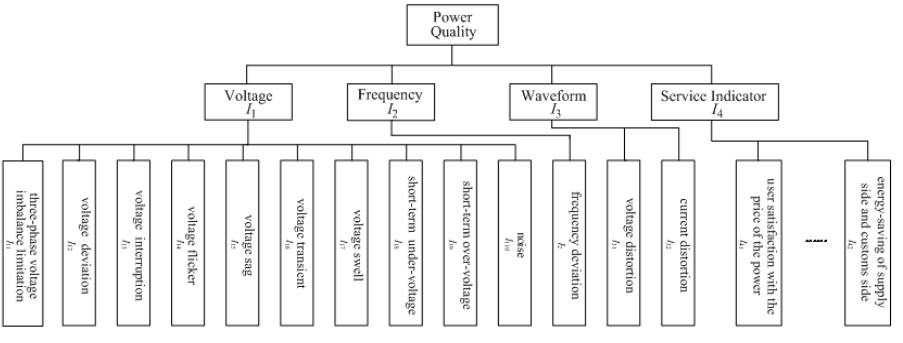


Fig. 1. The hierarchy model for power quality

Conducting Comprehensive evaluation of static indicators figured up with the data from the 95% probability value of the monitoring period and used SARFI (System Average RMS Variation Frequency Index) for transient indicators.

Firstly, With the data provided by Electric company which recorded by monitoring of a PCC(Point of common coupling) in a distribution power network to illustrate the evaluation methods. Detailed data are shown in Table1.

Table 1. Measured data of PCC of distribution network

Sample	I_{11}	I_{12}	I_{13}	I_{14}	I_{15}	I_{16}	I_{17}	I_{18}	I_{19}	I_{110}	I_2	I_3
1	5	2	1.67	0.89	0	4	3	0	0	1.15	0.116	2.47
2	11	2	1.40	0.78	1	1	1	0	0	1.55	0.06	2.39
3	13.7	2	1.11	0.78	0	1	0	0	0	1.20	0.65	2.40
4	7.3	1.1	1.56	0.85	0	2	2	0	0	1.15	0.092	3.50
5	6	2	1.53	0.75	1	3	0	0	0	0.90	0.07	3.19
6	3	1	0.96	0.71	0	1	0	0	0	0.94	0.06	1.61
7	5	2	1.24	0.77	0	2	1	0	0	1.15	0.066	2.74
8	5	2	1.17	0.73	1	1	0	0	0	1.05	0.06	2.59
9	4	2	2.08	0.96	5	8	0	0	0	1.87	0.12	4.28
10	6	2	2.17	0.97	9	10	4	0	0	2.01	0.132	4.34

The weights of 10 second-level indicators of voltage indicators is calculated by AHP Methods, the scores given by four experts used and the expanded optimization algorithm expounded in section Ⅱ.By using the same way we can get the weight vector of the first-level indicator).

In accordance with the above improved fuzzy Comprehensive evaluation method, 10 group data of power quality is used, the Comprehensive evaluation results shows in Table 2, it is consistent with the conclusions provided in references 14.

Table 2. Comprehensive evaluation results of power quality

sample	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Results	m	g	g	m	m	e	g	g	q	q

For Table 1, if using conventional fuzzy Comprehensive evaluation method, about more than 40% (5 group) data will obtain unreasonable even wrong evaluation results. For example, to evaluate sample 8th by proposed method in this paper, one can obtain $k_0 = 2$, and therefore the power quality of this time period belong to C_2 categories, promptly, the Comprehensive evaluation results of the sample is good. According to the conventional method of fuzzy Comprehensive evaluation of the evaluation results should be C_1, in fact we can see that sample belong to C_1 or C_2 are the properties of more or less equal measure, and samples belong to C_1 and C_2 together equivalent to 0.67, accounting for the entire attribute more than half, so that the sample belongs to C_1 attribute category is unreasonable.

4 Conclusion

An improved fuzzy Comprehensive evaluation method is proposed in this paper. Based on subjective and objective weighting methods, this paper presents a combinational evaluation method resulting in consistency between subjective and objective evaluation in the least squares sense, which make comprehensive evaluation result is more reasonable and credibility than single assessment method. then bring out a improved fuzzy synthetic evaluation method based on reliability code. The application in Comprehensive evaluation of the quality of the power commodities illustrates the effectiveness and the feasibility of the proposed method.

References

1. Yu, Y.-x., Luan, W.-p.: Smart Grid. Power System and Clean Energy 1, 7–11 (2009)
2. Hu, X.-h.: Smart Grid-A Development Trend of Future Power Grid. Power System Technology 14, 1–5 (2009)
3. Chen, S.-y., Song, S.-f., Li, L.-x., Shen, J.: Survey on Smart Grid Technology. Power System Technology 8, 1–7 (2009)
4. Hughes, J.: IntelliGrid architecture status report. In: EPRI (2005)
5. Briand, L.C., Emam, E.l., Freimut, B.G.: A comprehensive evaluation of capture-recapture models for estimating software defect content. IEEE Transactions on Software Engineering 6, 518–540 (2000)
6. Yang, J., Li, G.-y., Zhou, M.: Power Quality Service Pricing Method Based on Differential Pricing. Automation of Electric Power Systems 8, 41–45 (2006)
7. Chen, W., Hao, X., Lin, J.: An Improved Fuzzy Comprehensive Evaluation Method Using Expanded Least Deviations Algorithm. In: IEEE International Conference on Machine Learning and Cybernetics, pp. 490–494. IEEE Press, New York (2008)
8. Gao, J.S., Shi, X.D., Wang, L.W.: Study on Machine Equip–ment Selection in Dynamic Alliance Based on Fuzzy Comprehensive Evaluation. In: 6th World Congress on Intelligent Control and Automation, pp. 6787–6791. IEEE Press, New York (2006)
9. Jia, Q.-q., Song, J., Lan, H., et al.: Quality of Electricity Commodity and Its Fuzzy Evaluation. Power System Technology 6, 46–49 (2000)
10. Li, Q.-h., Zhou, L., Liu, H.-y., et al.: Evaluation of Power Quality by Accelerating Genetic Algorithm and Shepard Interpolation. High Voltage Engineering 7, 139–143 (2007)
11. Liu, Y.-y., Xu, Y.-h., Xiao, X.-n.: Analysis of New Method on Power Quality Comprehensive Evaluation for Regional Grid. In: Proceedings of the CSEE, vol. 22, pp. 130–136 (2008)
12. Zhang, B., Wang J: Power quality evaluation based on entropy principles. Electric Power Automation Equipment 10, 35–39 (2009)
13. Huang, J., Zhou, L., Li, Q.-h., et al.: Evaluation of Power Quality Based on the Method of Matter-element. Journal of Chongqing University (Natural Science Edition) 6, 25–30 (2007)
14. Chen, W., Hao, X., Lin, J.: Comprehensive Evaluation System and Method about Electrical Energy Quality Based on Attribute Recognition Theory and AHP. Electrical Engineering 5, 26–30 (2006)
15. Liu, Y., Dai, P., Xu, Y., et al.: Comparison of power quality comprehensive evaluation methods. Electric Power Automation Equipment 6, 25–30 (2007)

16. Cheng, Q.: Attribute Recognition Theoretical Model With Application. Acta Scicentiarum Naturalun Uuniversitis Pekinesis 1, 12–20 (1997)
17. Sstty, T.L.: The Analytic Hierarchy Process. McGraw–Hill Inc., New York (1988)
18. Qin, S.: Comprehensive Evaluation Methods and Application. Electronic Press, Beijing (2003)
19. Mao, D.: A Combinational Evaluation Method Resulting in Consistency between Subjective and Objective Evaluation in the Least Squares Sense. Chinese Journal of Management Science 5, 95–97 (2002)

A Novel Reference Wave Generating Algorithm Based on Sliding Window Iterative DFT and Adaptive Sampling Algorithm for DVR

Wei Chen[1,2], Jie Lin[1,2], and Xiaohong Hao[1,2]

[1] College of Electrical and Information Engineering
Lanzhou University of Technology, LZUT
Lanzhou, China
[2] Key Laboratory of Gansu Advanced Control for Industrial Processes
Lanzhou, China
chenlin@lut.cn

Abstract. Dynamic voltage restorer(DVR)is one of important power quality control equipment. It can compensate voltage sag accurately and quickly, which demands that voltage sags in grid should be accurately and real-time detected and reference voltage wave should be generated. in this paper a sliding window iterative DFT algorithm based on adaptive sampling algorithm is present. magnitude and phase of the Fundamental component of a sample sequence of can be extracted through sliding window iterative DFT algorithm. under the non-synchronized sampling conditons, adaptive sampling algorithm is be used which can adjust automatically the sampling time. So the algorithm can reduce errors caused by the non-synchronous sampling. The simulation results show that the algorithm can effectively detect the reference instruction voltage and has good real-time, tracing and anti-interference performances.

Keywords: power quality; dynamic voltage restorer; voltage sag; Discrete Fourier Transform; sliding-window iterative DFT; adaptive sampling algorithm.

1 Introduction

Power quality has become an important issue over the past several years. One of most important power quality issues in power system is the voltage sags, since voltage sags cause severe effects to end users, general, sag characteristics, e.g., stating and ending points, and depth, are typically determined using an RMS envelope. Dynamic voltage restorer (DVR) is one of important power quality control equipment. The key point of dynamic voltage restorer to compensate voltage sag is quick detection of voltage sag and accurate generation of correct reference voltage wave. At present most widely used method is instantaneous reactive power theory[1], wavelet transform[2] and dq transform[3].These algorithms get more attentions for their real-time. In this paper a sliding window iterative DFT algorithm based on adaptive sampling algorithm is present. Magnitude and phase of the base wave component of a sample sequence of can be extracted through sliding window iterative DFT algorithm. Under the

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 107–113.
springerlink.com

non-synchronized sampling conditons, adaptive sampling algorithm is be used which can adjust the sampling time automatically. So the algorithm can reduce errors caused by the non-synchronous sampling. Simulation results show the feasibility of the algorithm.

2 Sliding Window Iterative DFT Based on Adaptive Sampling Algorithm

In order to examine the quality of power and to get the reference voltage wave for DVR need to sample signal measured. Usually periodic signal sampling requires the sample is strict synchronous with measured signal. If sampling signal or measured signal be interference, the small changes of frequency of signal will cause spectrum leakage and truncation errors, then cause incorrect calculation results. Sliding-window iterative DFT[4],[5],[6] based on adaptive sampling algorithm[7] in present in this paper can effectively solve these problems.

2.1 Sliding Window Iterative Discrete Fourier Transform

For any finite bandwidth periodic signal $x(t)$, Suppose the period T, N is the number of sample, sample periodic $\tau = T/N$,Discrete Fourier Transform expression is:

$$x(k\tau) = A_0 + \sum_{n=1}^{N} A_n \cos(n\omega k\tau) + B_n \sin(n\omega k\tau), \quad k = 0,1,2\cdots, (N\text{-}1) \tag{1}$$

$$A_0 = \frac{1}{N}\sum_{i=0}^{N-1} x(i\tau) \tag{2}$$

$$A_n = \frac{2}{N}\sum_{i=0}^{N-1} x(i\tau)\cos(n\omega i\tau) \tag{3}$$

$$B_n = \frac{2}{N}\sum_{i=0}^{N-1} x(i\tau)\sin(n\omega i\tau) \tag{4}$$

The calculation of expression(1)-(4) need N sample data of whole sampling period, which lead to large amount of calculation and time-consuming, so above expressions are not suitable for instantaneous voltage detection.

Using sliding window iterative thoughts to improve expression (3) and (4):

$$A_n = \frac{2}{N}\sum_{i=N_{\text{new}}}^{N_{\text{new}}-N+1} x(i\tau)\cos(n\omega i\tau) \tag{5}$$

$$B_n = \frac{2}{N}\sum_{i=N_{\text{new}}}^{N_{\text{new}}-N+1} x(i\tau)\sin(n\omega i\tau) \tag{6}$$

N_{new} Noted that the latest sampling point, $x(i\tau)$ indicate sampled data before No. i the sampling point. The latest real-time sampling data involved in the signal detection and

analysis, while the corresponding eliminate the old sampled data, thereby greatly accelerated the rate of sampling data update and improve the system to track changes of load current or voltage. Fundamental signal can be calculated by the following formula:

$$x_1(k\tau) = A_1 \cos(\omega k\tau) + B_1 \sin(\omega k\tau) \tag{7}$$

$$A_1 = \frac{2}{N} \sum_{i=N_{\text{new}}}^{N_{\text{new}}-N+1} x(i\tau)\cos(\omega i\tau) \tag{8}$$

$$B_1 = \frac{2}{N} \sum_{i=N_{\text{new}}}^{N_{\text{new}}-N+1} x(i\tau)\sin(\omega i\tau) \tag{9}$$

When calculating formula (7)-(8), if the system sampling frequency is relatively high, then the Calculated Quantities will also be relatively large. In order to simplify the calculation process software sliding window iterative process [2] can be adopt.

At first a full cycle of N points of sampled data need to be stored n a contiguous space after multiplying with the rotation factor, and then need to set up a data pointer to locate position of the current sampled data, when the completion of a full cycle of N points calculation, the corresponding data pointer should point back to the starting position, begin the next cycle of replacement cycle of data. Thus equation (8) and (9) are simplified as the calculation of a subtraction and a sum. The new calculation result is restored to the old data storage unit, complete the iteration process.

$$\sum_{i=N_{\text{new}}}^{N_{\text{new}}-N+1} x(i\tau)\cos(\omega i\tau) = \sum_{i=N_{cur}-1}^{N_{cur}-N} x(i\tau)\cos(\omega i\tau) -$$

$$x(N_{\text{new}}-N)\tau\cos\left[\omega(N_{\text{new}}-N)\tau\right] + x(N_{\text{new}}\tau)\cos(\omega N_{\text{new}}\tau) \tag{10}$$

$$\sum_{i=N_{\text{new}}}^{N_{\text{new}}-N+1} x(i\tau)\sin(\omega i\tau) = \sum_{i=N_{\text{new}}-1}^{N_{\text{new}}-N} x(i\tau)\sin(\omega i\tau) -$$

$$x(N_{\text{new}}-N)\tau\sin\left[\omega(N_{\text{new}}-N)\tau\right] + x(N_{\text{new}}\tau)\sin(\omega N_{\text{new}}\tau) \tag{11}$$

The whole calculation process needs the whole period summation operations in the initialization stage. Afterwards, when a sample data enters, iterative computation will complete a new value iteration. In this way, the delay caused by calculation process greatly reduced.

This algorithms require the detected signals must be periodic signal, at the same time require digital sampling signals strictly synchronous with the detected signals. For the actual electrical signals, this requirement is too stringent in fact to meet. The actual power system frequency is always changing, although the change is generally slow. According to the characteristics of the electric signal adaptive sampling algorithm can be adopt, this algorithm can automatically adjust the sampling time in order to reduce the synchronization error and increase accuracy.

2.2 Adaptive Sampling Algorithm

$x(t)$ is a continuous periodic signal, Suppose the period T. If the actual sampling period is T_S, Let $x(t)$ go through time window of length LT, get N discrete-time series $x(nT_S)$, N is the number of sample. L is the number of cycles intercepted, Both N and L is integral number. In synchronous sampling, $\frac{LT}{T_S} = N$. When the signal frequency changes while the sampling period is fixed, which called non-synchronized sampling, $\frac{LT}{T_S} \neq N$. Which means according to the original sampling period to sample the sampled data is no longer 'N' in a signal period. So suppose the ideal sampling period is T_{S0}, while the actual sampling period is T_S. ε is the error between T_{S0} and T_S, $T_S - T_{S0} = \varepsilon$.

Suppose $x_0(n)$ an ideal sample sequence:

$$x_0(n) = x_0(nT_{S0}) = x(nT_S - n\varepsilon) \tag{12}$$

In nT_S according to Taylor series, ignoring higher-order items, there are

$$x_0(n) \approx x(nT_S) - x'(nT_S)n\varepsilon \tag{13}$$

And then according to the definition of derivative

$$\begin{aligned} x'(nT_S) &\approx \frac{x(nT_S + N\varepsilon) - x(nT_S)}{N\varepsilon} = \frac{x\left[n(T_{S0} + e) + Ne\right] - x(nT_S)}{N\varepsilon} \\ &= \frac{x\left[nT_{S0} + ne + Ne\right] - x(nT_S)}{N\varepsilon} \end{aligned} \tag{14}$$

where NT_{S0} is the period of sequence $x(n)$, therefore,

$$\begin{aligned} x\left[nT_{S0} + ne + Ne\right] &= x\left[nT_{S0} + ne + Ne + NT_{S0}\right] = x\left[(N+n)(T_{S0} + e)\right] \\ &= x\left[(N+n)T_S\right] = x\left[(N+n)\right] \end{aligned} \tag{15}$$

I $$x'(nT_S) \approx \frac{x(n+N) - x(n)}{Ne} \tag{16}$$

$$x_0(n) \approx x(n) - \frac{x(n+N) - x(n)}{Ne} ne = x(n) + \frac{n}{N}\left[x(n) - x(n+N)\right] \tag{17}$$

Equation 17 is the expression of signal adaptive sampling algorithm expression. Algorithm is simple and easy to project implementation, thus extremely suitable for real-time signal processing applications. Algorithm flowchart shows in Figure 1.

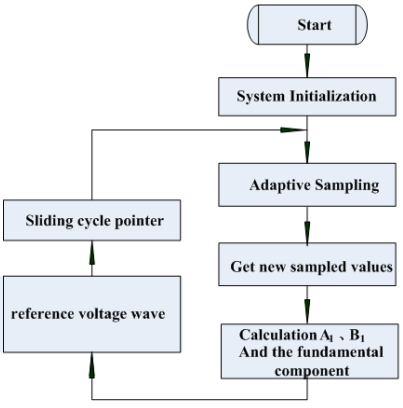


Fig. 1. Flowchart of adaptive sampling DFT algorithm

3 Simulation

Simulation is carried in matlab. In which the signal sampling frequency is set to 6.4k, that each period sample 128 points. When the signal is periodic signal using a fixed sampling period to sample, the simulation results shown in Figure 2. When the signal frequency changes from 50Hz to 51Hz at 0.1s and changes from 51Hz to 50Hz at 0.3s using a fixed sampling period will get incorrect results.

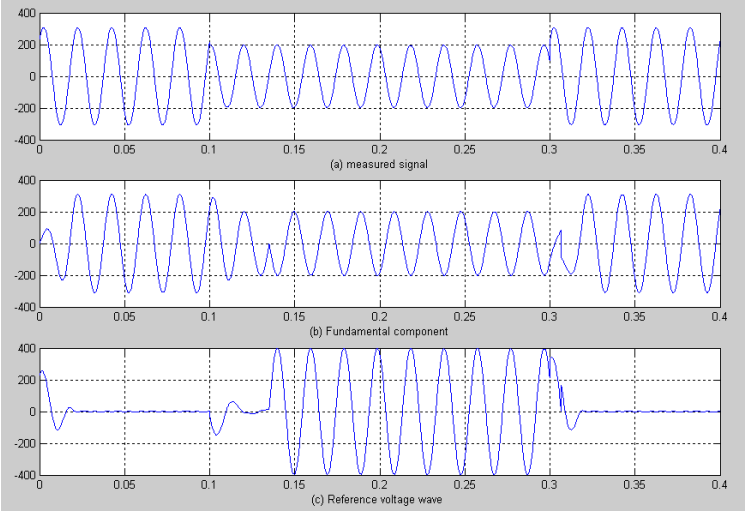


Fig. 2. Simulation results under non-synchronous sample

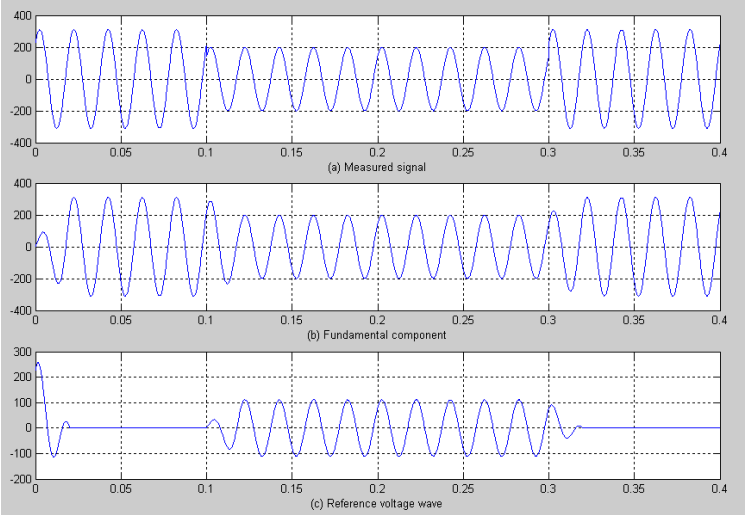


Fig. 3. Simulation results of voltage sag

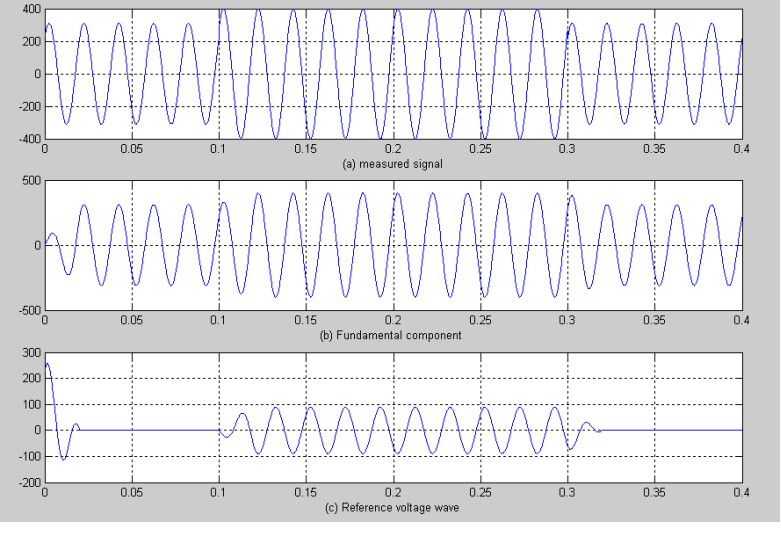


Fig. 4. Simulation results of voltage swell

Adaptive sampling simulation results shows in Figure 3 and 4. The system has a good tracking performance. when the signal is periodic signal, no need to use adaptive sampling algorithm, but when the signal frequency changes, if do not use adaptive sampling algorithm, then the frequency change, the system ran at a fixed sampling interval according to the original sampling, this will get the wrong result, but after adapting the adaptive sampling algorithm, the system can automatically adjust the actual sampling time. This can avoid the problems of DFT in the non-synchronous

sampling. The complexity of the algorithm does not increase in number, namely, the real-time system can still be guaranteed.

4 Conclusion

A new sliding window iterative DFT algorithm based on adaptive sampling algorithm is present in this paper. The algorithm can significantly reduces errors of DFT in the non-synchronous sampling caused through automatically adjusts the sampling time. The algorithm can effectively detect the reference instruction voltage and has good real-time, Simulation results show the feasibility of the algorithm.

References

1. Lee, S.J., Kim, H., Sul, S.-K., et al.: A novel control algorithm for static series compensators by use of PQR instantaneous power theory. IEEE Transactions on Power Electronics 3, 814–827 (2004)
2. Liu, L., Jia, W., Xiao, X., Zhang, Y.: Measurmeng method of voltage sag by wavelet transform and RMS method. Automation of Electric Power Systems 11, 33–36 (2003)
3. Feng, X., Yang, R.: A novel integrated morphology dq transformation detection algorithm for dynamic voltage restorer. In: Proceedings of the CSEE, vol. 11, pp. 193–198 (2004)
4. Luo, A., Shuai, Z., Zhu, W., et al.: Development of hybrid active power filter based on the adaptive fuzzy dividing frequency-control method. IEEE Trans. on Power Delivery 1, 424–432 (2009)
5. Luo, A.: Reactive power compensation and harmonic ontrol technology and equipment. China Power Press, Beijing (2006)
6. Lu, X.-l., Zhou, L.-w., Zhang, S.-h., Cao, C.-k.: Study and Simulation of Harmonic Detection Based on Sliding-window Iterative Algorithm of Discrete Fourier Transform. Journal of System Simulation 14, 3652–3655 (2008)
7. Dai, J., Chen, G.: An Adaptive Sampling Algorithm for Periodic Signal Measurement. Acta Electronica Sinica 8, 11–13 (1998)

An Analysis and Research of the Related Information Aggregation Assessment Based on Sensor Networks

Peng Wu and Yanping Chu

Computing Center, Henan University,
Kaifeng, China
hehe9840@126.com, 33237509@qq.com

Abstract. Sensor networks are battery powered, and they are generally deployed in the place which is hard to access to and environmentally harsh. The replacement of battery is not feasible, so the energy constraint is one of the features of sensor networks. It is essential in sensor networks design to make the best of the limited power of the sensor nodes. This paper is to make some exploration and research about the related information aggregation assessment based on sensor networks.

Keywords: Sensor networks, Data collection, Data redundancy.

1 Introduction

As the rapid development of the embedded system, wireless communication, network and micro-electro-mechanical system, the sensor network with its sensing, computing and wireless communication capabilities and wireless sensor network which is constituted by it aroused great attention. Wireless sensor network is a network made by self-organizing of many sensor nodes with the wireless communication technology. It integrates three great technologies: sensors, micro-electromechanical system and network, aiming to sense, collect and process the information of perception objects within the network coverage and then forward to the observer. It is a system centered in data processing [1]. Wireless sensor network is a special wireless communication network with an enormous number of nodes and cheap cost. So it allows fast deployment without relying on any fixed facilities. And it can collect information on time, accurately and comprehensively in many occasions. Thereby it changes the interact style between human and nature and is considered as one of the most important technologies in the 21st century. Nearly most nodes are battery powered, usually can't supplement energy, so energy-saving is one of the most crucial objectives in wireless sensor design.

2 The Analysis and Exploration of Information Aggregation Assessment Based on Sensor Networks

Sensor is the key component to collect and process data. It can map a physical quantity in the physical world to a quantitative measurement to help people to form

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 115–120.
springerlink.com

quantization cognition of the physical world. The general structure of the sensor includes the sensor module, the information processing module, wireless communication module and power supply modules. Sensor perception module mainly senses and acquires the external information and converts it into digital signals. Information processing module processes and keeps the data collecting by the sensor node itself and other nodes, controls the perception component and the power work mode etc. Wireless communication module communicates with other sensor nodes. In the process, the electricity provides the necessary power for work [2].

Wireless ad hoc network is a multi-hop, peer-to-peer and mobile network made up of dozens, even hundreds of nodes. It adopts the wireless communication, and is dynamic in networking. The aim of it is to transmit the multimedia information flow requiring service quality by means of the dynamic routing and the technology of mobile management. In addition to common characteristics of mobility, self-organization, finiteness of the power to the wireless ad hoc network, wireless sensor network also has distinctive features. For example, the topology of the network is easy to change, resources of nodes are very limited, and the area that can be monitored is wide. Specific features are as follows: The communication capability is limited, and the transmission rate of sensor nodes is low. In most cases, the communication distance is only several dozen or several hundred meters. As the nodes are often working in areas with appalling conditions, they are more exposed to the impact of terrains and landforms like high mountains, buildings, obstacles and of the natural environment such as wind, rain, thunder, lightning, humidity, and flooding, resulting in the unreliability of the communication between sensor nodes, and on the other hand a long-time bug or even damage to the nodes. Computing power is limited, so sensor nodes generally use embedded processors and memories. These nodes have the computing capability to complete certain processing work of information, but due to the limited capability and capacity of embedded processors and memories, the nodes' processing capability therefore is very limited. With limited power energy and a small size, a sensor node usually carries a very limited battery power. Since the number of sensor nodes, who have a wide range of distribution and a complex environment to be deployed, is large, and sensor nodes require low and cheap costs, during the using process, batteries thus can not be recharged or replaced. If the energy of batteries is used up, nodes will lose their effect. Therefore, the constraint of the energy is a serious problem that hinders the application of a wireless sensor network. Sensors consume more electrical energy in transmitting information than making calculations, hence in the working process of the network, we need to save energy, and maximize the life cycle of the network. The size of the network is large, and the topology is very complex. Besides, nodes in a sensor network are intensive. Its large quantity may be hundreds or even tens of millions. In addition, sensor network can be distributed in a wide geographical area, and its range of perception is also very large. Moreover, those three elements in a sensor network, sensor nodes, the perceptual objects, and the observer, are all possible to move, and it is quite often that new nodes will join or the existing nodes may fail. Therefore, along with the constant changes in the network's topology, the path between the n nodes of a sensor and the observer changes at the same time. Sensor network has the nature of self-organization. In the application of a sensor network, sensor nodes usually can not be accurately pre-set, and the relationship between adjacent nodes is also unknown in advance, such as scattering a large

number of sensor nodes by aircraft to the vast primeval forests, or randomly placing them in dangerous areas where people are unable to reach. This requires sensor nodes' capability of self-organizing, meaning that they can automatically configure and manage, and through the topology control mechanism and the network protocol sensors can automatically form a multi-hop wireless network system which forwards the monitoring data. Sensor nodes will lose effectiveness due to the energy depletion or environmental factors, and there are also some nodes who are added to the network in order to make up for the failure nodes, thus making the network topology dynamically change with it Self-organization of the sensor network should be able to adapt to this dynamic change. Data transmission's directionality is also very strong. In a wireless sensor network, data transmission is highly directional. Typically, the inquiry information is transmitted from the observer to the sensor nodes within the network by way of broadcasting or multicast, while information of the detection results clusters with the help of sensor nodes distributed everywhere into the query nodes. Fig. 1 shows the sensor network architecture.

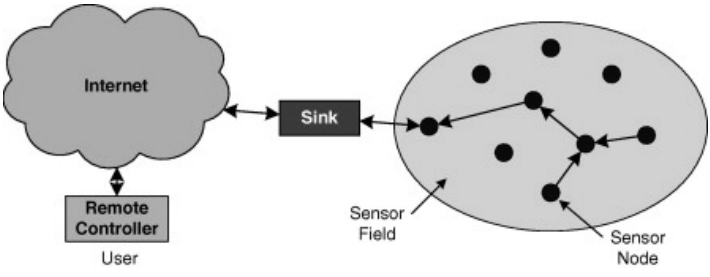


Fig. 1. WSN architecture.

Wireless sensor network is a ubiquitous sensing technology, which allows users to more deeply understand and grasp the world around them. Having very broad application prospects, the wireless sensor network, with a high using value, can be applied to military affairs, environmental monitoring and forecasting, health care, smart home, condition monitoring, space detection, warehouse management and safety monitoring in large areas. Field of military affairs: wireless sensor network is rather widely used in military affairs, such as monitoring the status of the battlefield, acting as a guidance device of smart weapons, detecting and determining attacks of nuclear, biological and chemical weapons, installing sensors in people, equipment, and arms of the military armies for recognition, and throwing sensors to the enemy's positions to detect intelligence, and the like. The U.S. Defense Department and all the military departments attach great importance to the sensor network. They brought forward the C4KSIR plan based on C4ISR, emphasizing the capability of perceiving the intelligence from the battlefield, and of synthesizing and using information. Taking sensor network as an important area of research, military departments set up a series of research projects on military sensor networks. Field of environmental protection: as people become increasingly concerned about the environment, environmental science involves a increasingly larger scope. Wireless sensor network provides convenience for the acquisition of random data for research in the field, for example, monitoring components of oceans, the atmosphere and soil, studying the impact of environmental

changes on crops, determining the rainfall conditions, offering accurate information to prevent floods and drought, real-time monitoring of air pollution, water pollution and soil contamination, and the monitoring of forest fires. Field of medical treatment and health: embedding network sensors in shoes, furniture, household appliances and other devices can help the aged, people who're seriously ill and the disabled with their family life. With the use of sensor networks, necessary information can be delivered efficiently, thus facilitating people's access to nursing, reducing the burden on the nursing staff and improving the quality of nursing. Using sensor networks can also collect people's physiological data for a long time, and accelerate the development process of new drugs, while micro sensors installed in the monitored objects will not bring inconvenience to people's normal life. In short, sensor networks equip the future telemedicine with a convenient and efficient means of technology. Field of space exploration: by means of setting sensors nodes in the objects of the outer space, where humans are inaccessible at present or can not work for a long time, a long period of monitoring of those objects can be achieved. Through the analysis of data returned by these sensor nodes, we may have a better understanding of those objects. Other application fields: self-organization, miniaturization and the perception of the outside world are the three major characteristics of the wireless sensor network, which determine that it is possible for the wireless sensor network to have not a few opportunities in other fields. For instance, sensor network can combine with PDA thus forming a personal server, which is able to obtain the real-time information of the network and filter all the related information by visiting information points so as to provide people with the needed information [3]. Apart from that, in many fields like the rescue during disasters, warehouse management, interactive museums, interactive toys, automatic production lines in factories, wireless sensor network will breed a totally new design and mode of application.

Information aggregation plays a very important role in wireless sensor networks, mainly reflected in saving the energy of the entire network, enhancing the accuracy of the collected data and improving the efficiency of data collecting. Wireless sensor network is constituted of a large number of sensor nodes covering to the monitoring area. In the deployment, sensor nodes need to reach a certain density in order to enhance the robustness of the whole network and the accuracy of the monitoring information. Sometimes the monitoring scopes of several nodes have to overlap. And it will lead to a certain degree of redundancy among the information reported by neighboring nodes [4]. Information aggregation is to deal with the superfluous data within the network by integrating the data and removing the redundant information to minimize the quantity of data conveying under the promise of satisfying the application needs.

There is no unifying method in information aggregation, it needs related computation method depending on different application backgrounds, such as the adaptive weighted aggregation algorithm. Suppose there are n sensors measuring one certain environment object. Firstly, make a data check of $x_i(i=2,\cdots n-1)$, the testing criteria is that the difference between the adjacent numbers should be no more than the given threshold ε, that is to say, $|x_{i+1}-x_i| \le \varepsilon,(i=2,\cdots n-1)$, ε is based on the accuracy degree of the sensor measuring. Suppose the measurement data of n

sensors meets $|x_{i+1}-x_i| \le \varepsilon, (i=2,\cdots n-1)$. Then let's conduct an estimate of the adaptive weighted fusion for these measurements. The basic idea of the adaptive weighted fusion algorithm is: with the optimal condition that the overall mean square error is the minimum, according to measurements provided by various sensors, find the corresponding optimal weighting factor of each sensor in the way of self-adaption so as to make the result of $\overset{=}{x}$ after fusion become optimal. Suppose variances of n sensors respectively are $\sigma_i^2, (i=1,2,\cdots n)$; the truth value needing to be estimated is x, and the measurement value of each sensor is $x_1, x_2, \cdots x_n$, which are independent of each other and are the unbiased estimates of x. The weighting factor of each sensor is $w_i (i=1,2,\cdots n)$, so the truth value of $\overset{=}{x}$ after fusion and the weighting factor satisfy: $\overset{=}{x} = \sum_{i=1}^{n} w_i x_i$, $\sum_{i=1}^{n} w_i = 1$. The overall mean square error is

$$\sigma^2 = E\left[\left(x-\overset{=}{x}\right)^2\right] = E\left[\sum_{i=1}^{n} w_i^2 (x-x_i)^2 + 2\sum_{i=1, j=1, i\neq j}^{n} w_i w_j (x-x_i)(x-x_j)\right].$$

Since $x_1, x_2, \cdots x_n$ are independent of each other, and they are the unbiased estimates of x, thus $E\left[(x-x_i)(x-x_j)\right]=0$, $(i \neq j, i=1,2,\cdots,n; j=1,2,\cdots,n)$.

Therefore, $\sigma^2 = E\left[\sum_{i=1}^{n} w_i^2 (x-x_i)^2\right] = \sum_{i=1}^{n} w_i^2 \sigma_i^2$. Assuming E stands for the set of sensor nodes monitoring events: $E=\{e_1, e_2, \cdots, e_m\}$, X_i stands for the random variables of the monitoring events set in node i, X_i gets its value in the sample space Ω, Ω stands for the set of the possible monitoring events of any nodes in the sensor network, is the power set of E, namely: $\Omega = \{w | w \in \Psi(E)\}$. Suppose E_i stands for the monitoring events set of sensor node i, A_i stands for the monitoring scope of node i, δ stands for the number of the events in per unit area, namely: $|E_i| = \delta A_i$. Assuming K data packets aggregate to one data packet, every data packet received by the sink node is $L_i bit$, changes into $L_f bit$ after aggregation. Then we can calculate the combination entropy of the aggregation entropy, that is to say, the data H_f contained in the mixed data, then: $H_f = H(X_1, X_2, \cdots, X_k) = \left|\bigcup_{i=1}^{k} E_i\right|$. Based on the above assumptions, let's design an assessment of information aggregation as the ratio of joint entropy of aggregation

data in the aggregation nodes and aggregation data bits, namely H_f / L_f . If the ratio of the input and output in the aggregation algorithm is a constant value, then the information aggregation can be defined as $r_f = L_f \Big/ \sum_{i=1}^{k} L_i$. Based on the above defined information aggregation degree, the assessment of information aggregation can be further defined as $H_f / L_f = \left| \sum_{i=1}^{k} E_i \right| \Big/ r_f \sum_{i=1}^{k} L_i$. Consequently, the greater the aggregation data, the smaller the assessment result. The visual effect of information aggregation is to reduce the amount of data transmission, thus save energy. The best polymerization condition is that the intermediate nodes can merge n equal-length input data groups into one output group, its energy-saving efficiency is $(n-1)/n$; in the worst case, the aggregation operation does not reduce the data amount, but reduce lots of operations in consulting and channel contention by decreasing the number of groups. Thus, it decreases the overall overheads of the transmission unit, meanwhile, save the energy [5].

3 Conclusion

There are a great amount of nodes densely distributed in the wireless sensor network, and the network will produce a large number of data. If we do not apply the information aggregation technology, there will be a great deal of excessive information which will consume a lot of energy. Meanwhile, the high probability of conflict will lead to the failure of sending and receiving of information. And the retransmission of information will consume much energy and increase the delay. In a word, the information aggregation technology is essential and necessary.

References

1. Sun, L., Li, J., Chen, Y., Zhu, H.: Wireless Sensor Networks. Tsinghua University Press, Beijing (2005)
2. Yin, A., Wang, B.-w., Hu, X.-y.: Improvement of iHEED Protocol in Wireless Sensor Network Based on Clustering by Node Level. Computer Science (July 2010)
3. Kong, F., Li, C., Ding, Q.: Dynamic key management scheme for wireless sensor networks based on polynomial keys. Journal of Tsinghua University (Science and Technology) (October 2009)
4. Liu, X., Li, J.-b., Ji, S.-l., Guo, L.-j., Hui, L.: A Cluster Head Selecting and Scheduling Strategy in Sensor Networks. Acta Electronica Sinica (August 2010)
5. Younis, O., Fahmy, S.: Heed: a hybrid, energy-efficient, distributed clustering approach for ad-hoc sensor networks. IEEE Transactions on Mobile Computing (4) (2004)

Development of Electronic Compass for Indoor Mobile Robot

Xisheng Li, Donghua Jiang, Xiongying Shu, and Ruiqing Kang

School of Information Engineering, University of Science and Technology Beijing, Xueyuan 30, Haidian, Beijing, P.R. China, 100083
lxs@ustb.edu.cn

Abstract. This paper presents an efficient MEMS gyro aided automatic calibration algorithm for three-axis magnetic compass. This electronic compass module consists of a three-axis magnetic sensor, a two-axis inclinometer and a MEMS gyro. The magnetic electronic compass is used to determine the heading of a indoor mobile robot with respect to the magnetic North. The automatic calibration method requires the mobile robot to make three full 360-degree rotation. In this rotation procedure, magnetic field data, attitude data and angular rate data are recorded. According to magnetic field data and attitude data, raw heading data is calculated. This raw heading data is verified by angular rate data from MEMS gyro. Results of experiment show that the accuracy of calibrated compass is better that 1 deg and MEMS gyro aided automatic calibration algorithm is effective for electronic compass.

Keywords: magnetic electronic compass, MEMS gyro, calibration, mobile robot, heading.

1 Introduction

The Earth's magnetic field intensity is about 20 to 50 A/m and has a component parallel to the Earth's surface that always point toward magnetic north [1-4]. This is the basis for all magnetic compasses. Magnetic compass has been used in navigation for centuries. Today, advances in technology have led to the solid state electronic compass based on MR magnetic sensors and acceleration based tilt sensors. Electronic compasses offer many advantages over conventional "needle" type or gimbaled compasses such as: shock and vibration resistance, electronic compensation for stray field effects, and direct interface to electronic navigation systems.

When a compass is operating in a open area in the absence of any ferrous metals there is no distortion effects on the earth's magnetic field. In reality, though, compasses are mounted in vehicles, aircraft, and platforms that most likely have ferrous materials nearby. The effects of ferrous metals (iron, nickel, steel, cobalt) will distort, or bend, the earth's field which will alter the compass heading. These effects can be thought of as a magnetic field that is added to the earth's field. This will introduce heading measurement error to compass.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 121–128.
springerlink.com

In this paper, an efficient gyro aided compass calibration algorithm is proposed. The magnetic electronic compass is used to determine the heading of a indoor mobile robot with respect to the magnetic North. The automatic calibration method requires the mobile robot to make three full 360-degree rotation. In this rotation procedure, magnetic field data, attitude data and angular rate data are recorded. According to magnetic field data and attitude data, raw heading data is calculated. This raw heading data is verified by angular rate data from MEMS gyro. Results of experiment show that this algorithm can be used to realize heading calibration for magnetic electronic compass.

2 Anistropic Magnetoresistive Sensor and Its Application to Electronic Compass

The anisotropic magnetoresistive (AMR) sensor is one type that lends itself well to the earth's field sensing range. AMR sensors can sense DC static fields as well as the strength and direction of the field. This sensor is made of a nickel-iron (Permalloy) thin film deposited on a silicon wafer and is patterned as a resistive strap. The properties of the AMR thin film cause it to change resistance in the presence of a magnetic field. Typically, four of these resistors are connected in a Wheatstone bridge configuration so that both magnitude and direction of a field along a single axis can be measured. For typical AMR sensors, the bandwidth is in the 1-5 MHz range. The reaction of the magnetoresistive effect is very fast and not limited by coils or oscillating frequencies.

The electrical output of AMR sensor is proportional to the magnetic field strength along its sensitive axis. When a AMR sensor is spun around a horizontal plane starting from magnetic north, the output is a cosine function of the heading angle. A minimum of two sensors that are arranged mutually perpendicular would eliminate the ambiguity in electrical output with respect to heading direction as seen in Fig. 1.

For a two-axis compass without tilt compensation, azimuth or the heading is calculated by the equations given below [1-8]

$$\alpha = \text{arcTan}\left(\frac{Y}{X}\right) \tag{1}$$

The X sensor defines the forward direction and the Y sensor is to the right.

This two-axis compass will perform well as long as it is kept horizontal and is useful in hand held applications. However, operation of the compass while it is not level can result in considerable amount of heading error.

2.1 Electronic Compass with Tilt-Induced-Error Compensation

Most often compasses are not confined to a flat and level. If the compass were tilted, the tilt angles (pitch and roll) and three magnetic field components must be used to calculate heading [1-8]. An inclinometer, or tilt sensor, should be used to determine the roll and pitch angles. The terms roll and pitch are commonly used: ROLL refers to

the rotation around the X-axis, or forward direction, and PITCH refers to the rotation around the Y-axis, or left-right direction.

The general magnetic compass module with tilt-induced-error compensation consists of a three-axis magnetic compass and a two-axis inclinometer to calculate tilt-compensated azimuth information [1-8]. Fig. 1 shows an functional block diagram of a three-axis compass with tilt compensation.

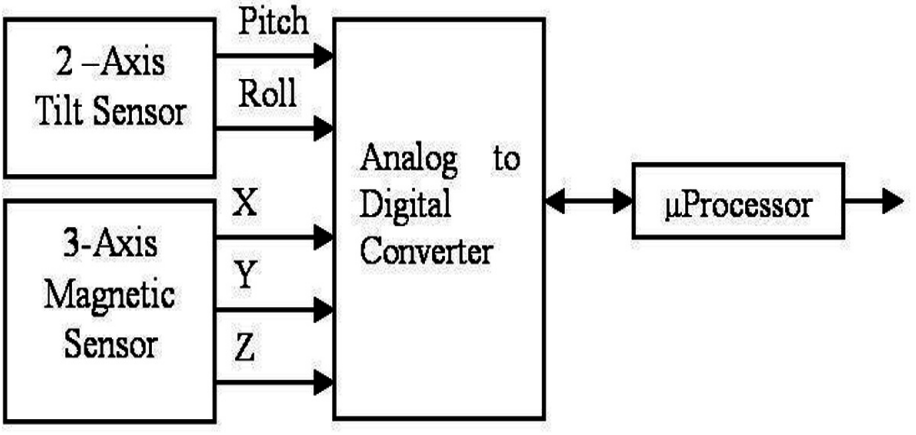


Fig. 1. Functional block diagram of a three-axis compass with tilt compensation

The compass must now rely on all three magnetic axes (X , Y , Z) so that the earth's field can be fully rotated back to a horizontal orientation. In Fig. 2, a compass is shown with roll and pitch tilt angles referenced to the right and forward level directions of the observer or vehicle. The X , Y and Z magnetic field data can be transformed back to the horizontal plane (X_{h} , Y_{h}) by applying the rotational equations shown below:

Roll compensation.
According to the roll angle, roll compensation is done by applying the rotational equations shown below.

$$\begin{cases} X_{\mathrm{rc}} = X \\ Y_{\mathrm{rc}} = Y\cos(\theta) + Z\sin(\theta) \\ Z_{\mathrm{rc}} = -Y\sin(\theta) + Z\cos(\theta) \end{cases} \tag{2}$$

where θ is roll angle.

Horizontal and vertical magnetic components calculation.
After roll compensation, according to the pitch angle, the X , Y and Z magnetic components can be transformed back to the horizontal plane and vertical direction by applying the rotational equations shown below.

$$\begin{cases} X_{\mathrm{h}} = X_{\mathrm{rc}}\cos(\varphi) - Z_{\mathrm{rc}}\sin(\varphi) \\ Y_{\mathrm{h}} = Y_{\mathrm{rc}} \\ Z_{\mathrm{v}} = X_{\mathrm{rc}}\sin(\varphi) + Z_{\mathrm{rc}}\cos(\varphi) \end{cases} \tag{3}$$

where φ is pitch angle. Horizontal and vertical magnetic components are gotten.

Once X_{h} and Y_{h} in the horizontal plane are known, (1) can be used to determine the azimuth.

2.2 Hard Iron and Soft Iron Interference Compensation for Electronic Compass

When a compass is operating in a open area in the absence of any ferrous metals there is no distortion effects on the earth's magnetic field. In reality, though, compasses are mounted in vehicles, aircraft, and platforms that most likely have ferrous materials nearby. The effects of ferrous metals (iron, nickel, steel, cobalt) will distort, or bend, the earth's field which will alter the compass heading.

When the compass is mounted on the mobile robot, the effect of the robot body would distort the Earth's magnetic field. Let the robot rotate in a circle would produce the curves shown in Fig. 2.

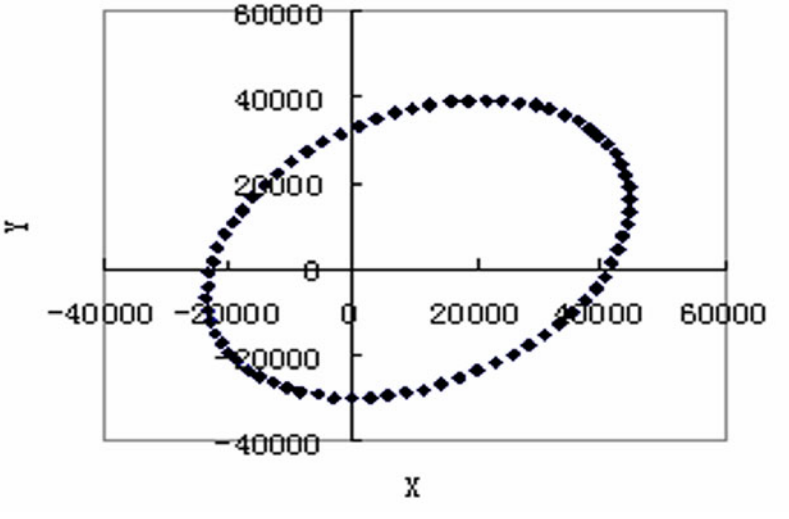


Fig. 2. Hard iron and soft iron interference of electronic compass (Unit: nT)

The X,Y plot is not a circle (slightly ellipsoid) and there is offset from the (0, 0) point. This offset and ellipsoid effect are a result of the fixed distortion of ferrous metals on the earth's magnetic field.

To compensate for the magnetic field distortion, two scale factors for X_{h} and Y_{h} can be determined to change the ellipsoid response to a circle. Offset values X_{off} and Y_{off} can then be calculated to center the circle around the 0,0 origin.

$$\begin{aligned} X_{\mathrm{OFF}} &= \left(\frac{X_{\mathrm{MAX}} - X_{\mathrm{MIN}}}{2} - X_{\mathrm{MAX}} \right) \times \max\left(\frac{Y_{\mathrm{MAX}} - Y_{\mathrm{MIN}}}{X_{\mathrm{MAX}} - X_{\mathrm{MIN}}}, 1 \right) \\ Y_{\mathrm{OFF}} &= \left(\frac{Y_{\mathrm{MAX}} - Y_{\mathrm{MIN}}}{2} - Y_{\mathrm{MAX}} \right) \times \max\left(\frac{X_{\mathrm{MAX}} - X_{\mathrm{MIN}}}{Y_{\mathrm{MAX}} - Y_{\mathrm{MIN}}}, 1 \right) \\ X_{\mathrm{C}} &= \max\left(\frac{Y_{\mathrm{MAX}} - Y_{\mathrm{MIN}}}{X_{\mathrm{MAX}} - X_{\mathrm{MIN}}}, 1 \right) \times X_{\mathrm{h}} + X_{\mathrm{OFF}} \\ Y_{\mathrm{C}} &= \max\left(\frac{X_{\mathrm{MAX}} - X_{\mathrm{MIN}}}{Y_{\mathrm{MAX}} - Y_{\mathrm{MIN}}}, 1 \right) \times Y_{\mathrm{h}} + Y_{\mathrm{OFF}} \end{aligned} \tag{4}$$

where X_{MAX}, Y_{MAX}, X_{MIN} and Y_{MIN} are maximum and minimum values of X_{h} and Y_{h} respectively.

3 MEMS Gyro Aided Automatic Compass Calibration

Traditionally, gyroscopes were mechanical devices that measured the angular rate of rotation. MEMS (microelectromechanicalsystem)-gyroscope technology now provides this function in a variety of packages that enable integration into PCB (printed-circuit-board)-based systems. MEMS gyroscopes employ tiny micromechanical systems on silicon structures, supporting motion-to-electrical transducer functions.

An ideal MEMS gyroscope produces a predictable output when you subject it to a known rate of rotation. It has no noise, perfect linearity, and no offset. Perfection is not typically available on economical manufacturing processes, however, creating the need for a deeper understanding of MEMS-gyroscope errors. A yaw-rate MEMS gyroscope responds to motion about its predefined axis of rotation. The following equation provides a simple linear behavior model for a MEMS gyroscope:

$$\omega_{GYRO} = K\omega_{RATE} + \omega_{BIAS} + \omega_{NOISE} \tag{5}$$

where ω_{GYRO} is output of MEMS-gyroscope, ω_{RATE} is the angular rate of rotation tested, ω_{BIAS} is the offset of MEMS-gyroscope, ω_{NOISE} is the noise of MEMS-gyroscope, K is the scale factor of MEMS-gyroscope.

MEMS angular rate gyroscopes do not measure angular displacement directly but rather the rate of angular motion. However, mathematical integration of angular rate with respect to time produces a relative angular displacement. It can track changes in the heading reliably in the short time. Errors in the mathematical integration accumulate with time due to quantization effects, scale factor and bias changes, and other error sources in the signal output of the gyro.

In order to calibrate the electronic compass, first place mobile robot with heading 0°in a level plane and let it be stationary, mathematical integration of angular rate output of gyro with respect to a fixed time interval produces a quantity proportional to the bias error. According to this quantity, the bias error of gyro can be calculated. The pitch angle (φ) and roll angle (θ) can be calculated according to the output of 2-axis tilt sensor.

Then let mobile robot make three full 360-degree rotation. In this rotation procedure, magnetic field data, attitude data and angular rate data are recorded. Because mobile robot is in level plane, pitch and roll angle are unchanged. According to real time magnetic field data, pitch and roll data in previous step, raw heading data is calculated. On the same time, mathematical integration of angular rate output of gyro with respect to time produces a relative angular displacement. Because mobile robot is of heading 0° at the beginning, this relative angular displacement can be used to induce true value of heading.

$$\alpha_{TRUE} = \begin{cases} \alpha_{GYRO}, & if\ 0 \le \alpha_{GYRO} < 360 \deg \\ \alpha_{GYRO} - N \times 360, & if\ N \times 360 \\ \qquad \le \alpha_{GYRO} < (N+1) \times 360 \deg \end{cases} \tag{6}$$

where N is a integer.

According to true value and raw data of heading angle, electronic compass calibration can be achieved.

4 Experiments

The electronic compass is implemented using Honeywell's Anisotropic Magneto-Resistive (AMR) sensor microcircuits HMC1022/HMC1021S, ADI's accelerometer sensor ADXL203. In order to eliminate the influence of bridge offset, the AMR sensor is flipped periodically at 100 Hz with SET/RESET pulse. The output signal of each sensor is amplified by precise instrumentation amplifier INA118 and then converted into digital quantities. MEMS gyro ADXRS150 is used for automatic calibration of electronic compass. 16-Bit, 6-Channel Simultaneous Sampling Analog-to-Digital Converter ADS8364 is used for analog signal sampling. Microcontroller C8051F120 is used for data processing. The electronic compass is connected to computer through RS-232.

At first, automatic calibration of electronic compass is made. Place the mobile robot in level plane, and let it to make three full 360-degree rotation at speed of 10 deg/s. According to recorded magnetic field data and attitude data, raw heading data is calculated. This raw heading data is verified by the mathematical integration of angular rate data from MEMS gyro. Calibration data is shown in Table 1. And then, let mobile robot rotate one full revolution slowly, raw heading is calculated according to magnetic field data and attitude data. Based on these calibration data, linear interpolation method is chosen to compensate the error of heading. Test result is shown in Table 2. This show that the accuracy of calibrated compass is better that 1 deg and MEMS gyro aided automatic calibration algorithm is effective for electronic compass.

Table 1. Electronic Compass Calibration Data (Unit: degree)

Heading (True value)	Heading (Raw)	Heading (True value)	Heading (Raw)
0.0	9.3	180.0	189.4
10.0	18.5	190.0	198.4
20.0	26.6	200.0	206.7
30.0	34.4	210.0	214.4
40.0	41.7	220.0	221.5
50.0	48.9	230.0	228.8
60.0	56.4	240.0	236.5
70.0	64.1	250.0	244.2
80.0	72.2	260.0	252.3
90.0	81.4	270.0	261.4
100.0	91.5	280.0	271.6
110.0	102.9	290.0	282.8
120.0	115.4	300.0	295.3
130.0	128.9	310.0	308.7
140.0	142.5	320.0	322.7
150.0	155.6	330.0	335.9
160.0	168.5	340.0	348.2
170.0	179.4	350.0	359.4

Table 2. Test Result of Electronic Compass (Unit: degree)

Heading (True value)	Heading Error	Heading (True value)	Heading Error
0.0	0.2	180.0	0.4
15.0	0.4	195.0	-0.4
30.0	-0.3	210.0	-0.3
45.0	0.5	225.0	0.5
60.0	-0.2	240.0	-0.3
75.0	-0.6	255.0	-0.7
90.0	0.1	270.0	0.5
105.0	0.3	285.0	0.4
120.0	-0.2	300.0	-0.3
135.0	0.7	315.0	-0.7
150.0	0.4	330.0	0.2
165.0	-0.5	345.0	0.8

5 Conclusion

In this paper an efficient calibration algorithm for three-axis magnetic compass is presented. The automatic calibration method requires the mobile robot to make three full 360-degree rotation. In this rotation procedure, magnetic field data, attitude data and angular rate data are recorded. According to magnetic field data and attitude data,

raw heading data is calculated. This raw heading data is verified by angular rate data from MEMS gyro. Results of experiment show that the accuracy of calibrated compass is better that 1 deg and MEMS gyro aided automatic calibration algorithm is effective for electronic compass.

Acknowledgments. We would like to thank Prof. Jinhui Lan at University of Science and Technology Beijing for the support she has provided us. This research has been supported by National Natural Science Foundation of China (Project 60574090), National High-tech R&D Program (Project 2007AA12Z320), the Fundamental Research Funds for the Central Universities (Project FRF-TP-09-017B), Program for New Century Excellent Talents in University (Project NCET-09-0216).

References

1. Caruso, M.J.: Application of magnetoresistive sensors in navigation systems, Sensors and Actuators. SAE SP 1220, 15–21 (1997)
2. Caruso, M.J., Bratland, T., Smith, C.H., Schneider, R.: A new perspective on magnetic sensing. Sensors (USA) 15(12), 34–46 (1998)
3. Caruso, M.J., Withanawasam, L.S.: Vehicle detection and compass applications using AMR magnetic sensors. In: Sensors Expo. Proceedings, pp. 477–489 (1999)
4. Caruso, M.J.: Applications of magnetic sensors for low cost compass systems. In: IEEE, Position Location and Navigation Symposium (Cat. No.00CH37062), pp. 177–184 (2000)
5. Platil, A., Kubik, J., Vopalensky, M., Ripka, P.: Precise AMR magnetometer for compass. In: Proceedings of IEEE Sensors 2003 (IEEE Cat. No.03CH37498), vol. 1, pp. 472–476 (2003)
6. Robert, S., Andy, F., Penny, P.: Aspects of heading determination via fusion of inclinometer and magnetometer data. In: ICAR 1997, pp. 739–744 (1997)
7. Basile, G., Pierantoni, M., Pirani, S., Ranaldi, M., Varosi, S.: Attitude compensated electronic compass for aircraft navigation. In: Proceedings of the 1994 Southcon Conference, pp. 126–131 (1994)
8. Li, X., Wang, J., Tang, C., Wang, L., Wang, S.: Development of High Accuracy Electronic Compass. Chinese Journal of Sensors and Actuators 19(6), 2441–2444 (2006)

A Simulation Study of Space Vector PWM Rectifier Based on PSCAD

Dong-dong Li and Lu Wang

Shanghai Key Laboratory of Power Station Automation Technology
Shanghai University of Electric Power
Shanghai, China
powerldd@yahoo.com.cn, cheerwl@126.com

Abstract. The simulation method of three-phase voltage source rectifier(VSR) using the Space Vector pulse width modulation (SVPWM) and feedforward decoupling control strategy is studied in PSCAD. Firstly, the mathematical model of VSR under the rotating coordinates axis was set up; then discussed the realization of SVPWM in PSCAD, and designed a decoupling control system of VSR; lastly, compared the performance of SVPWM and SPWM rectifier, the simulation results proved the dual loop control strategy of VSR is correct, the SVPWM rectifier has the advantages of less Harmonic content ,easier digital realization.

Keywords: three-phase voltage source rectifier; space vector PWM; feedforward decoupling; PSCAD.

1 Introduction

The traditional rectify method usually uses phase-controlled rectify or uncontrollable rectify, they have the shortcomings of slow dynamic response, high harmonic content in power side. Three-phase voltage type pulse-width modulation(PWM) rectifier has the advantages of higher power factor, lower harmonic content ,are mainly used in medium-power transform circuit, act as the direct current(DC) power of inverter circuits or uninterruptible power supply (UPS), it's research focus of green power realization and harmonic contamination elimination in recent years.

This paper first discusses the feedforward decoupling dual close-loop control strategy based on three-phase voltage rectifier, and then elaborates the implement of space vector pulse width modulation (SVPWM) method in the simulation software PSCAD environment, lastly, simulation study of the VSR control based on sinusoidal pulse width modulation (SPWM) and SVPWM modulation is performed, and the corresponding simulation waveform is given.

2 Control Strategy of VSR

2.1 Mathematical Model under Three-Phase Stationary Coordinate System

The voltage-source type PWM rectifier is show in Fig.1.Here, e_x represents the source voltage and i_x represents the input current, x=a, b, c. C is the capacity of the

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 129–137.
springerlink.com

DC-side filter capacitance, R and L denotes the resistance and inductance respectively, i_{load} is the load current of DC side, u_{dc} is the DC bus voltage, R_L is the load.

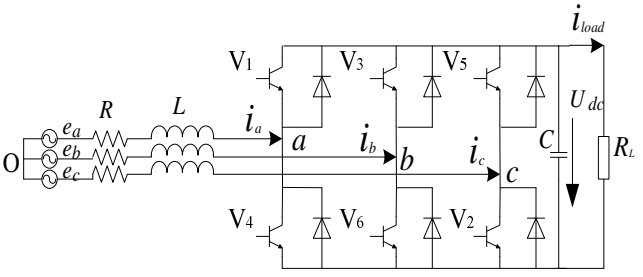


Fig. 1. Main circuit Three-phase voltage-type PWM rectifier

Define switching function of three-phase rectifier bridge as:

$$S_k = \begin{cases} 1 & \text{upper is ON and lower is OFF} \\ 0 & \text{upper is OFF and lower is ON} \end{cases} \quad (k = a,b,c) \tag{1}$$

Considering the displayed state variables on the circuit of Fig.1, neglecting switch delays, on-state semiconductor voltage drops, snubber networks, and applying Kirchhoff laws, Voltage equation of phase A can be obtained[1]:

$$L\frac{di_a}{dt} + Ri_a = e_a - (u_{aN} + u_{NO}) \tag{2}$$

When V_1 is on and V_4 is off, switching function $S_a = 1$, $u_{aN} = u_{dc}$, When V_4 is on and V_1 is off, $S_a = 0$, $u_{aN} = 0$ .for $u_{aN} = u_{dc}S_a$, so (2) can be rewrote as:

$$e_a = L\frac{di_a}{dt} + Ri_a + (u_{dc}S_a + u_{NO}) \tag{3}$$

Similarly the availability of phase B and phase C equation as follows:

$$e_b = L\frac{di_b}{dt} + Ri_b + (u_{dc}S_b + u_{NO}) \tag{4}$$

$$e_c = L\frac{di_c}{dt} + Ri_c + (u_{dc}S_c + u_{NO}) \tag{5}$$

For the three-phase three-wire balanced system, the following equations are established:

$$\begin{cases} e_a + e_b + e_c = 0 \\ i_a + i_b + i_c = 0 \end{cases} \tag{6}$$

derive from(2)- (6):

$$u_{NO} = -\frac{S_a + S_b + S_c}{3}u_{\mathrm{dc}} \tag{7}$$

In addition, using Kirchhoff's current law on DC bus capacitor positive node:

$$C\frac{du_{dc}}{d_t} = i_a S_a + i_b S_b + i_c S_c - i_{load} \tag{8}$$

2.2 Mathematical Model under Synchronous Rotating Coordinate System

Reference coordinate system is shown in Fig.2, and using the Park transform, derive from(2)-(8), we can get mathematical model of VSR under dq coordinate system[2]:

$$\begin{cases} e_d = L\frac{di_d}{dt} + Ri_d - \omega L i_q + u_d \\ e_q = L\frac{di_q}{dt} + Ri_q + \omega L i_d + u_q \\ C\frac{du_{dc}}{dt} = \frac{3}{2}(i_d s_d + i_q s_q) - i_{load} \end{cases} \tag{9}$$

Here e_d, e_q and i_d, i_q denote the input voltages and the input currents in the synchronous reference frame, Where i_d is active current component, i_q is reactive current component, The direction of d-axis orientate according to the direction of voltage vector E ,so $e_d = E, e_q = 0$.

The input of active power and reactive power of grid side are[3]:

$$\begin{cases} P = \frac{3}{2}(e_d i_d + e_q i_q) = \frac{3}{2} e_d i_d \\ Q = \frac{3}{2}(e_q i_d - e_d i_q) = -\frac{3}{2} e_d i_q \end{cases} \tag{10}$$

As can be seen from the above analysis, after coordinate transformation, input active and reactive power of the rectifier have been decoupled, the voltage and current are DC components in the d-q axes, the coordinate transformation provides the conditions for high performance control.

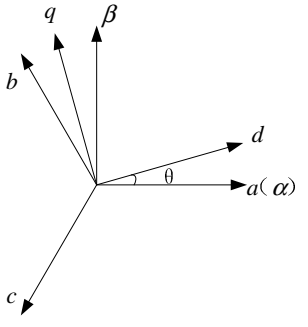


Fig. 2. Reference coordinate systems

2.3 Dual Loop Control Strategy of VSR

Use the feedforward decoupling control strategy, the current regulator with PI regulator, u_d , u_q Equations as follows:

$$\begin{cases} u_d = -(K_{iP} + \dfrac{K_{iI}}{s})(i_d^* - i_d) + \omega L i_q + e_d \\ u_q = -(K_{iP} + \dfrac{K_{iI}}{s})(i_q^* - i_q) - \omega L i_d + e_q \end{cases} \tag{11}$$

When the rectifier operates in unity power factor state, the input of reactive power of grid side is zero, i.e. $i_q = 0$, By controlling the d-axis current to control the active power ,and controlling the q-axis current to control the active power. A control system with dual close-loops of voltage and current was designed based on the mathematical model of three-phase PWM rectifier, the control structure of the system is shown in Fig.3[4].

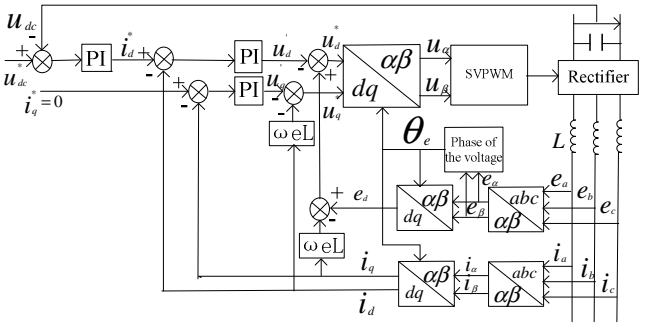


Fig. 3. Dual close-loop control scheme

3 SVPWM Control of Three-Phase VSR

3.1 Theory of SVPWM

The power electronic converter can operate only eight distinct topologies. Six out of these eight topologies produce a nonzero output voltage are known as active vectors (U1- U6) and the remaining two topologies produce zero output voltage are known as zero vectors (U0, U7). The six active vectors numbered I to VI in a stationary reference frame as shown in Fig. 4.

The active vectors U1 to U6 are obtained as follows: [5]

$$\mathrm{U}_k = \frac{2}{3} u_{dc} \bullet e^{j(k-1)(\pi/3)} \tag{12}$$

Here, k =l, 2,…,6.

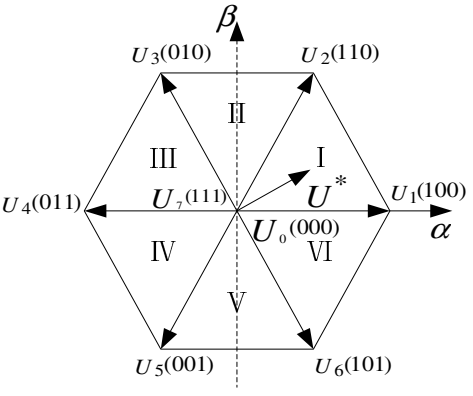


Fig. 4. Representation of the inverter states in the stationary reference frame

Let the desired voltages in a stationary ($\alpha-\beta$ axis) reference frame are U^*. Then, the vector, magnitude and angle of desired voltage can be given as follows:

$$U^* = \left|U^*\right| e^{j\theta^*} \tag{13}$$

$$\theta^* = \tan^{-1}[u_\beta^* / u_\alpha^*]\ ;\ \left|U^*\right| = \sqrt{u_\alpha^{*2} + u_\beta^{*2}} \tag{14}$$

From Fig. 4 finds that, assuming U^* to be lying in sector k, the adjacent active vectors are U_k and U_{k+1}. To achieve the desired stator voltage U^* within the sampling time T_s, the active voltage vectors U_k and U_{k+1} should be activated during the time T_k and T_{k+1}, respectively. Hence, the on-time T_k and T_{k+1} are evaluated by the following equations:

$$U^* T_s = U_k T_k + U_{k+1} T_{k+1} \tag{15}$$

Splitting this vectorial equation into real and imaginary components, from (12)-(15) follows that:

$$\begin{bmatrix} T_k \\ T_{k+1} \end{bmatrix} = \frac{\sqrt{3} T_s \left|U^*\right|}{u_{dc}} \begin{bmatrix} \sin(\frac{k\pi}{3} - \theta^*) \\ \sin(\theta^* - \frac{(k-1)\pi}{3}) \end{bmatrix} \tag{16}$$

The on-time for the zero state vectors is obtained by the following equation:

$$T_0 = T_s - (T_k + T_{k+1}) \tag{17}$$

When $T_k + T_{k+1} > T_s$,the time T_k and T_{k+1} are rescaled as follows:

$$\begin{bmatrix} T_k^{'} \\ T_{k+1}^{'} \end{bmatrix} = \frac{T_s}{T_k + T_{k+1}} \begin{bmatrix} T_k \\ T_{k+1} \end{bmatrix} \tag{18}$$

3.2 The Implementation of SVPWM in PSCAD

Implementation of SVPWM is divided into three steps: [6,7,8]

(1)Determine the sector number of the reference vector Units

Define variables A, B, C, P as: $\begin{cases} A = u_\beta \\ B = \frac{1}{2}(\sqrt{3}u_\alpha - u_\beta) \\ C = \frac{1}{2}(-\sqrt{3}u_\alpha - u_\beta) \\ P = sign(A) + 2sign(B) + 4sign(C) \end{cases}$

Here is a sign function, defined as $sign(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases}$,

Then, the corresponding relations of P and the sector number is shown in Table 1:

Table 1. Correspondence between P and sector NO.

Sec. No.	I	II	III	IV	V	VI
P	3	1	5	4	6	2

(2)Determine the working period of the two adjacent vectors

Define $\begin{cases} X = \sqrt{3}u_\beta T_s / u_{dc} \\ Y = (3u_\alpha + \sqrt{3}u_\beta)T_s / 2u_{dc} \\ Z = (-3u_\alpha + \sqrt{3}u_\beta)T_s / 2u_{dc} \end{cases}$,

When the desired voltage vector is in the six different sectors, the working period of the two adjacent vectors T_1 , T_2 can be got from Table 2:

Table 2. The value of T_1 , T_2

Sec. NO.	I	II	III	IV	V	VI
T_1	Z	Y	-Z	-X	X	-Y
T_2	Y	-X	X	Z	-Y	-Z

If $T_1+T_2>T_s$,we should rescale T_1 , T_2 refer (18),i.e. $T_1^{'}=T_1T_s/(T_1+T_2)$, $T_2^{'}=T_2T_s/(T_1+T_2)$

(3). Determine the on time and off time of power device in the VSR

$$\text{Define}\begin{cases}T_a=(T_s-T_1-T_2)/4\\T_b=T_a+T_1/2\\T_c=T_b+T_2/2\end{cases},$$

Then, the on time of power device V_1, V_3, V_5 in every cycle can be got from Table 3, the Corresponding off time is T_s-T_{cmx}.

Table 3. The value of T_{cmx}

Sec. NO.	I	II	III	IV	V	VI
T_{cm1}	T_a	T_b	T_c	T_c	T_b	T_a
T_{cm2}	T_b	T_a	T_a	T_b	T_c	T_c
T_{cm3}	T_c	T_c	T_b	T_a	T_a	T_b

4 Simulation Study

Using PSCAD simulation software, build a main circuit of three-phase VSR and a dual closed-loop control circuit based on SVPWM and SPWM. The main parameters of the system: The AC source with frequency of 50Hz, the grid-side line voltage is 380V, the line inductance of each phase is $2\,mH$, the AC side resistance is $0.1\,\Omega$, the reference value of DC voltage is 700V, the DC capacitor is $3000\,\mu F$, the load resistor is $30\,\Omega$ and the switching frequency of VSR is $5\,kHz$.

Use SVPWM and SPWM respectively, the DC bus voltage is shown in Fig.5, Fig.6 (the units of voltage and current are KV and KA respectively), we can see that the DC bus voltage is stable at a given value of 700V.

Voltage and current waveforms of phase A are shown in Fig.7,Fig. 8, we can see that the phase angles of phase voltage and phase current are the same, operating at unity power factor. Comparison of Fig. 7 and Fig.8 can be found that the current waveform is sine wave in Fig.7, while the waveform distortion Obviously in Fig.8, according to FFT analysis, as shown in Fig.9, Fig.10,we can see that when using the SVPWM, the harmonic current components content of phase A is zero, while using the SPWM, the harmonic current components content is higher, especially the $6k\pm1$ harmonics $(k=1,2\cdots,n)$.

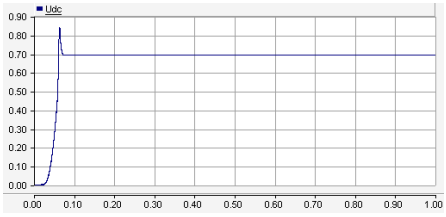


Fig. 5. DC bus voltage using SVPWM

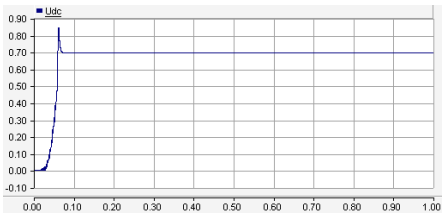


Fig. 6. DC bus voltage using SPWM

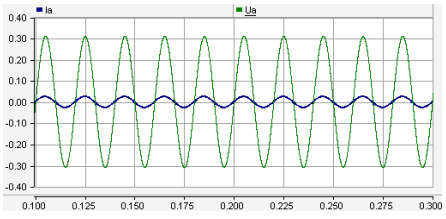


Fig. 7. I_a and U_a using SVPWM

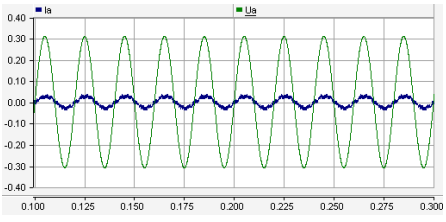


Fig. 8. I_a and U_a using SPWM

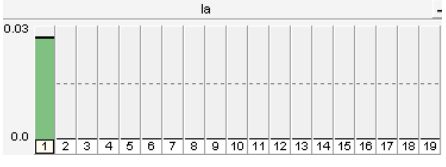


Fig. 9. FFT analysis using SVPWM

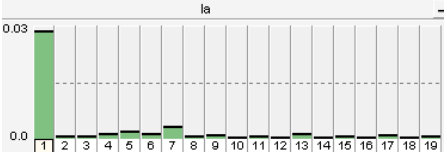


Fig. 10. FFT analysis using SPWM

5 Conclusion

This paper proposed a feedforward decoupling dual close-loop control scheme for the three-phase VSR. Computer simulation results have verified that the controller can offer an excellent steady-state performance, fast dynamic response, as well as the convenient active and reactive power decoupled controls. The SVPWM rectifier has the advantage of less harmonic content ,easier digital realization, which will be widely used in the industry.

Acknowledgments. This work was supported in part by National Natural Science Foundation of China (50807035), Shanghai Rising-Star Program (09QA1402400), Innovation Program of Shanghai Municipal Education Commission (11ZZ169), Shanghai Key Scientific and Technical Project (10DZ1200303) and Leading Academic Discipline Project of Shanghai Municipal Education Commission (J51303).

References

1. Zhang, C., Zhang, X.: PWM rectifier and its control. China Machine Press, Beijing (2003)
2. Xie, M.-x., Tian, M.-x.: Research on PWM rectifier with voltage oriented control strategy. J. Electrical Drive Automation(2009)
3. Li, Q., Li, Y.-d.: Research on Close-loop Control of Three-phase PWM Rectifier. J. Electric Drive (2007)
4. Wang, W.-w., Yin, H.-j.: Parameter Setting for Double Closed-Loop Vector Control of Voltage Source PWM Rectifier. J. Transactions of China Electrotechnical Society (2010)
5. Muyeen, S.M., Mannan, M.A., Ali, M.H.: Simulation technique& Application of Space-Vector PWM method in PSCAD/EMTDC. In: International Conference on Information and Communication Technology, ICICT (2007)
6. Yang, G.-j., Sun, L., Cui, N.-z.: Study on method of the Space Vector PWM. J. Proceedings of the CSEE (2001)
7. Tian, Y.-f., He, J.-a.: Simulation and Analyzation of Space Vector PWM. J. Proceedings of the EPSA (2004), 2004
8. Sun, W.-h., Cheng, S.-m.: A Realization of FPGA-based Space-vector PWM. J. Electric Drive (2000)

Evaluation of Distribution Network Reliability Based on Network Equivalent

Dongdong Li and Yun Wang

Shanghai Key Laboratory of Power Station Automation Technology,
Shanghai University of Electric Power
powerldd@yahoo.com.cn, oliver_yunwang@sina.com

Abstract. The reliability of distribution system of electricity is one of the most important indicators to measure the ability that the power supply system sends electric power to the load continually. Also it is an important technical and economic indicator for the power supply enterprise. In this paper, the theory about the reliability of distribution power system have been studied, a method of network equivalent is introduced. The validity of this method is examined by an example of distribution system of Pudong International Airport. Finally, the article analyses a number of factors which may influence the indicators of reliability.

Keywords: distribution system; reliability evaluation; network equivalent.

1 Introduction

Reliability, one of the important indicators in power system is reflected in the electricity power industry to meet the needs of the national economy. With the development of social production and people's living standard demand for electricity is growing. Reliability now has become equally important power quality standards with the voltage, frequency and so on.

The composition of power system is extremely complex. There are so many components in the system and the structure is precision, variable. Meanwhile, the power grid exposes to the natural environment, it is for these reasons and their own external interference, power accidents have occurred, seriously affects the load to acquire power continuously, making the country suffered great economic losses. Distribution system supplies directly to end users. Frequent failure of its equipment, caused 80% of our power blackout [1].

2 A Method to Evaluate Reliability of Distribution Network Based on Network Equivalent

Distribution system reliability evaluation algorithm can be divided into two categories: analytical method and simulation method. Analytical method use fault enumeration for the selection and estimation of state, establish a more rigorous mathematical

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 139–145.
springerlink.com

model to calculate the reliability index. Simulation carries out the state's options and estimation by sampling the probability distribution of the components to obtain reliability index using the method of reliability.

Distribution network reliability assessment is to traverse all the basic principles of component failure on the reliability of all load points, calculate the load point reliability indexes, and finally integrated all the load point reliability indexes deriving system reliability. Concrete steps generally are:

- Simplify complex network to the formation of a simple radial network
- According to component reliability index, obtain the equivalent load point index, and indicators of the whole system.

2.1 Simplify the Complex Network

In the process of distribution network reliability evaluation, the device can be divided into non-repairable and repairable components. For the specific component in a distribution network, contains normal state and fault state. For the protection components, assuming they all work at 100% reliability. Load is the average load model.

Distribution system usually uses simple radiation network to supply the consumers. In both ends power supply system, the normal form generally will open the switch to cut off two simple radiation networks. This network is characterized by all the components are in series, the branch circuit components and the main trunk line segments are connected in series too.

A number of component feeders form this type of radiation network structure. First of all, divide the network into some layers according to the number of feeder, each feeder and various components connected to the feeder is a layer. Each layer of the network can be equivalent to an equivalent branch line, gradually upward from the end of the equivalent layer, and finally you can translate a number of complex networks with branch feeders to a simple radial network [2][3].

2.2 Network Equivalent and Calculation of Indexes

Assessment of Complex distribution network reliability, including two equivalent processes: upwards and downwards equivalent. In the process of upwards equivalent, the branch feeder line will be replaced by an equivalent node component at a higher level feeder. The impact of higher-level components which is brought by equivalent elements can be represented by the failure rate λ, failure time of power supply U, fault repair time r. In the process of downwards equivalent, the impact of feeder in the low-level which is brought by the feeder in high-level will be replaced by an equivalent component at the beginning of low-level feeder [3][4].

1) The branch feeder with circuit breaker

$$\lambda = (1-P)\sum_{k=1}^{n}\lambda_k^{'}$$

$$r = \frac{\sum_{k=1}^{n} \lambda_k' t_1}{\sum_{k=1}^{n} \lambda_k'} = t_1$$

$$U = \sum_{k=1}^{n} \lambda_k' t_1$$

λ_k' is the failure rate of node k in the branch feeder; P is the reliability of circuit breaker at the beginning of branch feeder; t_1 is the working time of disconnection switch which is couple with the circuit breaker in the branch feeder.

2) The branch feeder without circuit breaker

$$\lambda = \sum_{k=1}^{n} \lambda_k'$$

$$r_e = \frac{u}{\lambda} = \frac{\sum_{k=1}^{n} \lambda_k' r_k}{\sum_{k=1}^{n} \lambda_k'}$$

$$U = \sum_{k=1}^{n} \lambda_k' r_k$$

λ_k' is the failure rate of node k in the branch feeder; r_k is the repair time of feeder which is caused by the failure of node k.

For the system of simple radial feeder, reliable indicators of the load node can be calculated as follows:

$$\lambda_i = \sum_{k \neq i, k=1}^{n} \lambda_k' + \lambda_{il} + \lambda_{it}$$

$$r_i = \frac{U_i}{\lambda_i}$$

$$U_i = \sum_{k \neq i \neq ik, k=1}^{n} \lambda_k' r_{ik} + \lambda_{il} r_{il} + \lambda_{it} r_{it}$$

λ_k' is the failure rate of node k in the branch feeder; r_{ik} is the repair time of node i which is caused by node k. r_{ik} relies on the structure of system and the position relationship between power source and component i.

With the above formulas, in accordance with the foregoing, we can easily get the index, constitute a new simple radial network, and all the load point reliability indexes of the whole system can be obtained [5][6].

3 Example Calculation

The example in this paper bases on a distribution power system of Pudong International Airport.

3.1 Raw Data

This power system contains three inlet wires, they are JiXiang4068, ZhenHang913, and JiXiang4053. Their voltage is 35kv, after passing three main transformers: #1，#2，#3, each of inlet wires divides into two lines, constitute six 10kv buses. Each of these buses has nine feeders to the load. Every subsystem of 35kv can be connected to a ring by circuit breakers, but they will not operate in the ring at normal time. The indexes of main devices are listed as fellows:

Table 1. Component reliability data.

Component reliability data			
Component	Classes	Equipment-forced outages	
		Failure rate （time per year）	Repair time （hour）
circuit breaker	35KV	0.2865	9
	10KV	0.048	4
	0.4KV	0.03	3
transformer	35KV	0.3333	2.5
	10KV	0.1	1.5
bus		0.1	3

3.2 Assumptions

- The generating system and transmission system are all reliable, and the generation system always can meet the requirements of load.

- The circuit breakers are installed at the feeder, the tie lines between buses and transformers contain circuit breakers. All of these circuit breakers can work reliable.
- The high-level substation of JiXiang4068 and JiXiang4053 is JiChang substation. So the feeder of JiChang substation may have three programs:

    a) JiXiang4068 and JiXiang4053 come from the same bus.
    b) JiXiang4068 and JiXiang4053 come from the same feeder which divides into two lines leading to two buses.
    c) JiXiang4068 and JiXiang4053 come from two different feeders; each of these feeders leads to two buses.

Obviously, the reliability of the three programs is different, program c) is the highest, but program a) is the lowest. However, the relationship between the three programs is the same when system operating at non-ring state.

3.3 Networks and Calculation of Indexes

This distribution system is a radial network, upwards equivalent to the 10kv side of transformer #1 which leads to two buses through circuit breakers. Now, the system should be discussed in two conditions.

- Assuming that the network operating at Open-loop, and manually switching the breakers. Each inlet wire and two distribution buses build a small system. On the basis of previous step that the load has been equivalent to the distribution buses, assuming that the bifurcation on the 10kv side of main transformer is a node. Than the load node should be equivalent to this bifurcation.
- When open-loop operation, if there are some fault happen in the small network, relay will operate, the circuit breaker on the key point may open. This condition should be considered. If the bus tie breaker is automatic switching mode, the bus tie line can connect with adjacent small 35kv system automatically. The fault bus can restore operation. This condition looks like a shunt system, now it should be divided into two programs:

    a) If the inlet wires of adjacent buses come from different high-level substation, we can only consider these two buses. This program has higher power supply reliability.
    b) If the inlet wire of adjacent buses come from the same substation, and this inlet wire contains a bifurcation. It is obvious that this kind of program's power supply reliability is lower.

According to the simplification of network, we can calculate the reliability of this distribution system by some common indicators.

4 Calculation Results and Analysis

After the calculation of example system, the result listed as fellow table:

Table 1. Reliability index

Reliability index	No.1	No.2	No.3
SAIFI	1.0158	1.0295	1.1501
SAIDI	4.54575	2.5067	2.8363
CAIDI	4.475	2.4349	2.4661
ASAI	0.99948	0.999714	0.999676
ASUI	0.00052	0.000286	0.000324

From the table above, we can know:

- In the first condition, manually switching mode, its ASAI is 0.99948. And in the other two conditions, they all use automatic device, their ASAI are higher than the first condition. The switching time is usually 2 minutes (automatic switching) or 30 minutes (manual switching). If we ignore the delay of automatic switching, the manual switching should be considered as a non-backup system.
- In the case of automatically switching system, the source of distribution network may influence the reliability. If the mutual backup two inlet wires come from different source, its reliability mustn't lower than the backup line come from different source. From the table, we know that in the second condition, assuming that the ZhenHang913 is the backup line of JiXiang4068, its ASAI is 0.999714. In the third condition, though the final backup is still ZhenHang913, as the non-adjacent buses, JiXiang4068's backup bus is JiXiang4053, they come from the same source. So if ZhenHang operates as backup, it will pass more breakers and buses. The ASAI of whole system is lower.

Now we analyze reliability of system from same source. JiXiang4068 and JiXiang4053 lead from the high-level substation, they may have three programs. Usually, the less devices the more reliability. But in the case of different feeders of the same source, under this adverse factor, if we can divide these two feeders' relationship, the reliability will be higher. C) minimizes the relationship between feeders.

5 Conclusion

This paper introduce a method to evaluate reliability of distribution network based on network equivalent, its key point is the equivalent of complex network. From the example system, we can see that different backup method and different bus arrangement will exert significant influence on the reliability index. This conclusion suggests us to improve system reliability by the following visual angles:

- Power system should use automatic device, they are good for fast action, strengthen system's backup ability, and improving the reliability of overall power grid.
- If it is permitted, use some security network construction such as dual power supply, sectionalized buses. This will help to reduce the probability of power outages, reduce the extent of failure and enhance ability of sustainable supply.

Consider the security, reliability and economy of whole grid, arrangement the system operation reasonable, give full play to the performance, guarantee power quality effectively, and avoid devices overload and other accidents caused by unreasonable arrangement of grid operation [8][9].

Acknowledgments. This work was supported in part by National Natural Science Foundation of China (50807035), Shanghai Rising-Star Program (09QA1402400), Innovation Program of Shanghai Municipal Education Commission (11ZZ169), Shanghai Key Scientific and Technical Project (10DZ1200303) and Leading Academic Discipline Project of Shanghai Municipal Education Commission (J51303).

References

1. Yang, S.: Basic and application of reliability analysis in the power system. Hydraulic and Electric Power Press, Beijing (1986)
2. Zhou, Y., Xiao, C., Wei, X.: The Research Based on Distribution- network Reliability of Network Equivalent Method. Electric Switchgear 46, 20–24 (2008)
3. Wang, C., Xu, Z., Gao, P., Mei, T.: Statistical analysis of fundamental data on component reliability for power systems. East China Electric Power 35, 1–4 (2007)
4. Wan, M., Zhou, Y., Guo, D.: Delaminated Simplified Evaluation of Complex Distribution System Reliability. Electric Switchgear 46, 1–3 (2008)
5. Wan, G., Ren, Z., Tian, X.: Study on model of reliability-network-equivalent of distribution system reliability assesment. Proceedings of the CSEE 23, 48–51 (2003)
6. Wang, J., Zhao, H.: Review of Reliability Evaluation Methods for Distribution System. Guangdong Electric Power 21, 10–13 (2008)
7. Faza, A.Z., Sedigh, S., McMillin, B.M.: The Advanced Electric Power Grid: Complexity Reduction Techniques for Reliability Modeling. In: Harrison, M.D., Sujan, M.-A. (eds.) SAFECOMP 2008. LNCS, vol. 5219, pp. 429–439. Springer, Heidelberg (2008)
8. An, X., Duan, Y.: Reliability Analysis of Distribution Network and Study on Improving Method. Electric Switchgear 46, 48–50 (2008)
9. Li, Y., Guo, H.: Power System Reliability Analysis Techniques. Sichuan Electric Power Technology 31, 14–15 (2008)

Effective Coverage Area Based on Call Dropping Probability of Next Generation Wireless Cellular Networks with Mobile Relay Station

Wei Zheng, Rujing Zhao, and Dongming Su

Beijing University of Posts and Telecommunications,
100876 Beijing, P.R. China
zhengweius@gmail.com, zrj870324@bupt.cn, m_sdm@163.com

Abstract. In this paper, we focus on the effective coverage benefits with the deployment of mobile relay stations in the next generation wireless cellular networks based on call dropping probability discussion. The special scenario is defined that mobile relay stations are applied on carriages of a fast-moving train to improve the quality of communication for user equipments on the train along railway. According to the analysis of effective coverage area performance based on the variation in received signal power over distance in path loss and shadowing with corresponding performance evaluations, it is shown that under the same conditions the effective coverage area with relay stations in the wireless cellular network outperforms that of without relay stations, and the higher the call dropping probability based on user equipment is, the higher BS antenna height is, and the wider area the BS covers.

Keywords: mobile relay station, path loss, shadowing, effective coverage area, call dropping probability.

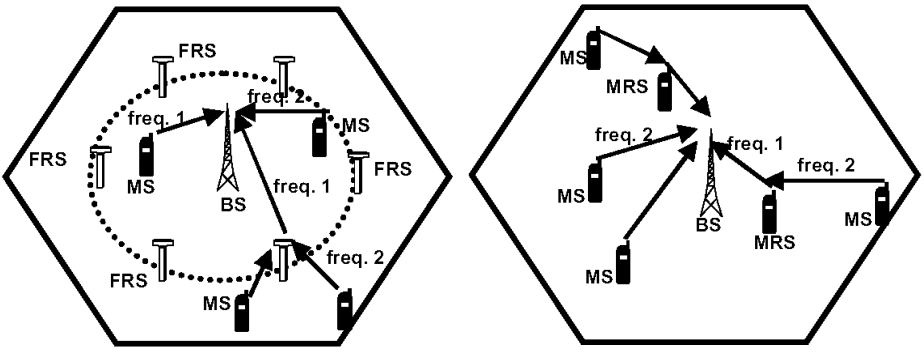
1 Introduction

3GPP proposed LTE-Advanced as the further evolution of LTE. One of the evolution goals is to fulfill and even surpass all of the IMT-Advanced requirements on capacity, data rates, latency, spectrum efficiency and low-cost deployment, etc. In order to reach the requirements for the evolution of LTE, 3GPP proposed some technology components in LTE-Advanced system. One of the potential key technology components is relay technology. By inserting relay stations (RSs), relay technology can effectively increase data rates, extend coverage area, expand system capacity, improve spectrum efficiency, combat fading, and enhance system robustness, etc. In addition, it also deploys the wireless cellular network in a cost efficient manner and thus cuts down OPEX and CAPEX. Relay technology has been studied and considered in the standardization process of 3GPP LTE-Advanced. Relaying is considered for LTE-Advanced as a tool to improve e.g. the coverage of high data rates, group mobility, temporary network deployment, cell-edge throughput and to provide coverage in new areas [1].

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 147–154.
springerlink.com

The so-called relay technology, taking the simple two-hop relay for example, is that a direct bad link from UE(User Equipment) to BS(Base Station) or from BS to UE can be broken down into two shorter better links, a UE-RS link and a RS-BS link, or a BS-RS link and a RS-UE link. In this case, a UE is able to communicate directly with the BS or indirectly uses some relay stations to relay its signals to the BS.

In [2] two different types of relaying network architecture have been investigated. One is proposed to use others users' terminal (Mobile Relay Station, MRS) to relay traffic while the other is proposed to use Fixed Relay Station (FRS) to relay traffic. The two different concepts are shown in Fig.1.Lots of studies have been done on relaying system with either FRS or MRS [3-12]. Compared with FRS, MRS can provide more flexibility to wireless cellular network [13].



(a) FRS (b) MRS

Fig. 1. Cellular networks with relaying.

In this paper, we focus on the effective coverage benefits based on call dropping probability discussion with the deployment of MRSs in the next generation wireless cellular networks. There are mainly two different types of scenarios to deploy MRSs in wireless cellular networks. One is to make MRSs deployed on moving vehicles, such as trains, buses, cars, etc, to cover areas in/on/outside the vehicles. The other is to make the non-active UE (i.e. in idle state) to relay the signals between the active UE to BS, acting as a MRS. We choose the first type of scenario that MRSs are deployed on a fast-moving train as the scenario to be discussed. In the following discussion we assume that MRSs all work in AF (Amplify-and-Forward) relaying way. Considering the variation in received signal power due to path loss and shadowing, we analyze the call dropping probability based on UE (CDPUE) performance and CDPUE-based effective coverage area performance. Comparing with the conventional wireless cellular network without RS, the results show that under certain conditions both the CDPUE performance and CDPUE-based effective coverage performance are improved. Meanwhile, the performance gain under different antenna height is also compared.

In this paper, the system model is specified in section 2. Then the CDPUE-based effective coverage area analysis is discussed in section 3. And the performance

evaluations and discussions are provided in section 4. Finally, a conclusion is drawn in section 5.

2 System Model

The following analysis is based on the specified scene of adopting RSs deployed on a fast-moving train. The scenario of relay-assisted is shown in Fig.2. By the RSs, UEs on the train can communicate with BS. Characteristics for the wireless cellular network along the railway are specified now.

- Low traffic of communication. Thereby, the coverage is the main issue that cut down OPEX and CAPEX.
- Deployment of Macrocell to cover wide area in a cell.
- High-speed mobility of UEs. Due to the mobility of UEs relative to the serving BS, the received signal may have a Doppler shift.

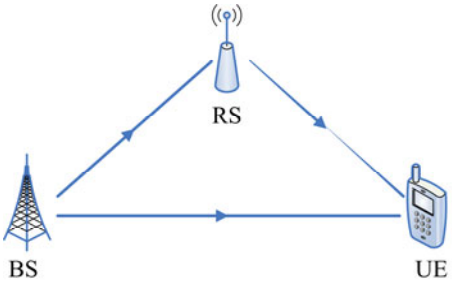


Fig. 2. Relay-assisted scenario in downlink.

According to the summarized characteristics above, MRSs are adopted to improve the effective coverage and communication quality on the train. Two types of RSs have been defined in 3GPP LTE-Advanced, Type-Ⅰ and Type-Ⅱ. Depending on the relaying strategy, a relay may be part of the donor cell and control cells of its own [1]. In this paper, we mainly consider a RS to be part of the wireless cellular network. In this case, smart repeaters, decode-and-forward relays and different types of L2 relays are examples of this type of relaying. The relay station is wirelessly connected to radio-access network via a donor cell. The connection can be inband, in which case the BS-to-RS link shares the same band with direct BS-to-UE links, or outband, in which case the BS-to-RS link does not operate in the same band as direct BS-to-UE links [1]. In our analysis, we consider the inband RSs.

RSs are considered as smart repeaters, and MRSs relay signals using AF (Amplify-and-Forward) relaying. For the inband RSs, since a wireless device cannot simultaneously transmit and receive signals at the same frequency channel, we consider the time division case that RSs receive and transmit in different time..

Consider the downlink relay channel from a BS to a UE via a RS, as shown in Fig.2. It is assumed that Rayleigh fading is between BS and RS and Rice fading is

between RS and UE. In the first time slot, the signal transmitted by a BS is received by a RS, which is:

$$y_1 = \sqrt{P_s} h_1 x_1 + n_1 \tag{1}$$

In equation (1), x_1 represents the signal with the transmission power P_s, and h_1 represents the channel coefficient from BS to RS in Rayleigh distribution, and n_1 represents the corresponding Gaussian noise with zero mean and variance N_0.

Then, during the second time slot the RS amplifies y_1 and retransmits it to the destination UE. The UE receives:

$$y_2 = A\sqrt{P_r} h_2 y_1 + n_2 \tag{2}$$

In equation (2), A represents the amplification factor to scale the transmission power by the RS, P_r represents the power transmitted through the RS, h_2 represents the channel coefficient from RS to UE in Rice distribution, and n_2 represents the corresponding Gaussian noise with zero mean and variance N_0. It is assumed that the RS and the UE receiver chains have identical noise properties.

3 Performance Analysis

The following analysis is to discuss the received signal power variation for the reason of path loss and shadowing over distance.

The COST 231 extension of HATA model is adopted as the pass loss model (in dB), which is [14]:

$$L(d)dB = 46.3 + 33.9\lg f_c - 13.82\lg h_t - a(h_r) + (44.9 - 6.55\lg h_t)\lg d + C_M \tag{3}$$

$$a(h_r) = 3.2(\lg 11.75 h_r)^2 - 4.97 \tag{4}$$

In equation (3), d represents the distance from transmitter to receiver, f_c represents the carrier frequency, h_t/h_r represents the Tx/Rx antenna height, and C_M is a constant factor for suburbs, 0dB.

The log-normal shadowing model is adopted due to shadow fading, which is the most widely used statistical model. In the model the path loss ψ is assumed random changed with a log-normal distribution referred to [14].

$$p(\psi) = \frac{\xi}{\sqrt{2\pi}\sigma_{\psi_{dB}}\psi} \exp\left[-\frac{(10\lg\psi - \mu_{\psi_{dB}})^2}{2\sigma_{\psi_{dB}}^2} \right] \quad \psi > 0 \tag{5}$$

In equation (5), $\xi = 10/\ln 10$, $\mu_{\psi_{dB}}$ refers to the mean of $\psi_{dB} = 10\lg\psi$ in dB and $\sigma_{\psi_{dB}}$ refers to the standard deviation of ψ_{dB} in dB.

Then we can obtain the mean of ψ (the linear average path loss) from (5) due to [14].

$$\mu_\psi = E[\psi] = \exp\left[\frac{\mu_{\psi_{dB}}}{\xi} + \frac{\sigma_{\psi_{dB}}^2}{2\xi^2} \right] \tag{6}$$

The conversion from the linear mean to the log mean (in dB) is derived from (6) due to [14].

$$10\lg \mu_{\psi} = \mu_{\psi_{dB}} + \frac{\sigma^2_{\psi_{dB}}}{2\xi} \tag{7}$$

Based on (5)-(7) and with changes of variables we get the Gaussian distribution of ψ with mean $\mu_{\psi_{dB}}$ and standard deviation $\sigma_{\psi_{dB}}$ [14]:

$$p(\psi_{dB}) = \frac{1}{\sqrt{2\pi}\sigma_{\psi_{dB}}} \exp\left[-\frac{(\psi_{dB} - \mu_{\psi_{dB}})^2}{2\sigma^2_{\psi_{dB}}} \right] \tag{8}$$

Models for path loss and shadowing are typically superimposed to capture power falloff versus distance along with the random attenuation about the path loss from shadowing. According to the combined model the received power in dB is given by:

$$P_r / P_t\,(dB) = -PL - \psi_{dB} \tag{9}$$

In equation (9), P_r represents the received power, P_t represents the transmission power, PL represents the pass loss in (3), and ψ_{dB} represents a Gauss-distributed random variable with mean zero and variance $\sigma^2_{\psi_{dB}}$.

Based on models described above, we discuss call dropping probability based on UE (CDPUE) and CDPUE-based effective coverage area. CDPUE is the call dropping probability in the downlink that the received power of UE falls below a target minimum received power level. In wireless systems there is typically a target minimum received power level P_{min} below which performance becomes unacceptable. The call dropping probability is defined based on UE $CDPUE(P_{min}, d)$ under path loss and shadowing to be the probability that the received power of UE at a given distance d, $P_r(d)$, falls below P_{min}: $CDPUE(P_{min}, d) = p(P_r(d) < P_{min})$. For the combined path loss and shadowing model it can be expressed by the following equation:

$$CDPUE(P_{\min}, d) = 1 - Q((P_{\min} - (P_t - PL)) / \sigma_{\psi_{dB}}) \tag{10}$$

In equation (10), the Q function represents the probability that a Gaussian random variable X with mean zero and variance one is bigger than z [14]:

$$Q(z) = p(X > z) = \int_z^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-y^2/2) dy \tag{11}$$

CDPUE-based effective coverage area refers to the coverage area of the wireless cellular network along the railway when the wireless cellular network along the railway is under the condition of different CDPUE requirement.

4 Results and Discussion

Based on the scenario and analysis described above, the results of performance evaluations are present in this section. It is assumed that there is one MRS on every

carriage to cover the carriage, and each UE in a carriage selects the MRS which covers the carriage as its relay. To evaluate the performance improvement, the assumptions are same both to the conventional cellular network without relaying and our analytical model with relaying. For performance evaluations, the parameters relevant are listed in Table1.

Table 1. Parameters.

Parameter	Value
Carrier Frequency	2GHz
UE Antenna Height	1.5m
BS Antenna Height	30m/100m
RS Antenna Height	5m
Radius of Cell	2000m
UE Transmit Power	23dBm
BS Transmit Power	43dBm
RS Transmit Power	37dBm
Pmin of UE	-97dBm
Shadowing standand deviation	8dB

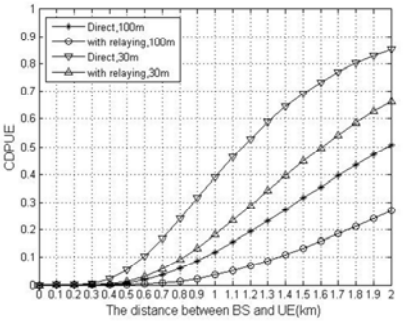


Fig. 3. CDPUE versus the distance between UE and BS comparison.

When a train is run, a UE receives signals from a serving BS in cell with or without relays. Fig. 3 shows the CDPUE on the train in a single cell changing with the distance between UE and BS for the direct path without relays and the relayed path with relays respectively. Furthermore, to evaluate the effect about different antenna heights, the CDPUE changing with different conditions of BS antenna height (when BS antenna height is 30 meters and 100 meters) for both the direct path and the relayed path is also shown in Fig.3. It is observed that the CDPUE performance in the relay-assisted wireless cellular network is better than the conventional wireless cellular network without relays. And, the CDPUE performance under the condition of the BS antenna height 100 meters is better than the condition of 30 meters for both in direct path and relayed path.

Fig. 4 shows the effective coverage area corresponded to CDPUE changing for the direct path without relays and the relayed path with relays respectively. Meanwhile, the curves of the effective coverage area versus the CDPUE under the condition of

different BS antenna heights, both 30 meters and 100 meters, are presented to evaluate the effect of different BS antenna heights on the effective coverage area, for both direct path and relayed path. We can obtain from the curves that under the same conditions the effective coverage area with RSs in the wireless cellular network outperforms that of without RSs. Also, it is observed from the curves that different BS antenna heights has different impact on the effective coverage area, and the wireless cellular network under the condition of 100 meters BS antenna height has better effective coverage area than that of 30 meters BS antenna height. Then we can deduce that under the same condition the higher CDPUE is, the higher BS antenna height is, and the wider area the BS covers.

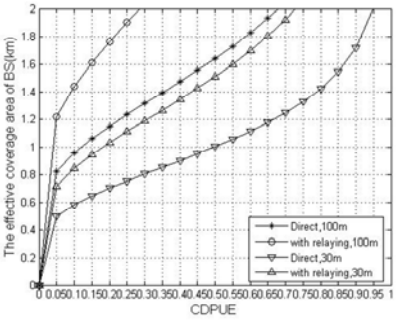


Fig. 4. The effective coverage area versus the CDPUE comparison.

5 Conclusions

In this paper, we focus on the special scenario that MRSs are deployed on carriages of a fast-moving train, and discuss the performances improvement of the CDPUE and the CDPUE-based effective coverage area brought in relay-assisted scenario. By performance evaluation and comparison, it is concluded that the CDPUE performance in the relay-assisted wireless cellular network is better than the conventional wireless cellular network without relays. And under the same conditions the effective coverage area with RSs in the wireless cellular network outperforms that of without RSs. Also, the numerical result shows that the effective coverage area difference comes from different BS antenna heights. According to the discussion above, it is deduced that under the same condition the higher CDPUE is, the higher BS antenna height is, and the wider area the BS covers.

References

1. 3GPP TR 36.814.: Further Advancements for E-UTRA Physical Layer Aspects
2. Pabst, R., Walke, B.: Relay-Based Deployment Concepts for Wireless and Mobile Broadband Radio. IEEE Communications Magazine, 80–89 (2004)
3. Peters, S.W., Panah, A.Y., Truong, K.T., Heath Jr., R.W.: Relay Architectures for 3GPP LTE-Advanced. EURASIP Journal on Wireless Communications and Networking (2009)

4. Nourizadeh, H., Nourizadeh, S., Tafazolli, R.: Performance Evaluation of Cellular Networks with Mobile and Fixed Relay Station. In: IEEE Vehicular Technology Conference, pp. 1–5 (2006)
5. Bakaimis, B.A., Lestable, T.: Connectivity Investigation of Mobile Relays for Next Generation Wireless Systems. In: IEEE Vehicular Technology Conference, pp. 2192–2195 (2005)
6. Beniero, T., Redana, S., Hamalainen, J., Raaf, B.: Effect of Relaying on Coverage in 3GPP LTE-Advanced. In: IEEE Vehicular Technology Conference, pp. 1–5 (2009)
7. Xiao, L., Fuja, T.E., Costello, D.J.: An Analysis of Mobile Relaying for Coverage Extension. In: IEEE International Symposium on Information Theory, pp. 2262–2262 (2008)
8. Li, P., Rong, M., Xue, Y., Wang, L., Schulz, E.: Performance Analysis of Cellular System Enhanced with Two-Hop Fixed Relay Nodes. In: VTC 2007-Spring, pp. 929–933 (2007)
9. Bakaimis, B.A.: Power Control/Allocation Solutions for Mobile Relays for Future Cellular Systems. In: IEEE International Conference on 3G and Beyond, pp. 441–444 (2005)
10. Waisanen, H.A., Shah, D., Dahleh, M.A.: Minimal Delay in Controlled Mobile Relay Networks. In: IEEE Conference on Decision and Control, pp. 1918–1922 (2006)
11. Murad Hossain, M.F., Mammela, A., Chowdhury, H.: Impact of Mobile Relays on Throughput and Delays in Multihop Cellular Network. In: IEEE ICWMC, pp. 304–308 (2008)
12. Lau, V.K.N., Huang, K., Wang, R.: A New Scaling Law on Throughput and Delay Performance of Wireless Mobile Relay Networks over Parallel Fading Channels. In: IEEE International Symposium on Information Theory, pp. 784–788 (2009)
13. Nourizadeh, V.H., Nourizadeh, S., Tafazolli, R.: Performance evaluation of Cellular Networks with Mobile and Fixed Relay Station. In: VTC 2006-Fall, pp. 1–5 (2006)
14. Goldsmith, A.: Wireless Communications. Stanford University, California (2004)

Two-Dimensional Rate Model for Video Coding

Shuai Wan[1], Junhui Hou[1], and Fuzheng Yang[2]

[1] School of Electronic and Information, Northwestern Polytechnical University, Xi'an China, 710072
swan@nwpu.edu.cn, houjunhuihn@163.com
[2] State Key Laboratory of ISN, Xidian Universtity, Xi'an China, 710071
fzhyang@mail.xidian.edu.cn

Abstract. This paper investigates the joint impact of variations of the frame rate and quantization on the coding bit-rate. Through theoretically linking the frame rate and the quantization step to the mean absolute difference (MAD), the MAD is modeled by a 2-D function, based on which a 2-D rate model is proposed. Extensive experimental results have demonstrated the effectiveness of the proposed 2-D rate model. Together with a 2-D distortion model, the proposed rate model can serve as the basis for advanced rate control in video coding and rate-constrained scalable video coding as well as bitstream adaptation.

Keywords: 2-D rate model, rate control, mean absolute difference (MAD), frame rate, quantization step.

1 Introduction

Rate model plays an important role in video coding and transmission. Using an accurate rate model, optimized encoding parameters can be obtained ahead of coding. The relationship between the rate and the distortion for texture coding has been given a considerable amount of attention for video coding, such as in early schemes TM5 where a simple first-order linear rate-quantization (R-Q) model is employed for rate control [1]. The quadratic R-Q models become popular in later schemes, providing better performance than the linear model at the price of higher computational complexity in TMN8 and VM8 [2][3]. Recently, He and Mitra have proposed a ρ-domain model that estimates the rate, where ρ indicates the percentage of zero coefficients after quantization [4]. Many other R-Q models have also been developed along with the development of rate control schemes for video coding [5][6], with the same aim of adjusting the quantization parameter (QP) given a targeted bit-rate.

In low bit-rate applications such as wireless video, however, heavy quantization and temporal resolution reduction (i.e. skipping frames) are usually introduced in video coding to adapt to the limited bandwidth. As a result, quality degradation is inevitable in both the spatial and the temporal domain. In particular, frame dropping causes jitter/jerkiness for human perception. And in existing schemes, frame skipping is always employed when the buffer tends to overflow. This is a rather passive process and the result may impair the overall rate-distortion (RD) performance and causing incoherency in motion considering perceived video quality. Therefore the

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 155–162.
springerlink.com

encoder should determine the frame rate together with the QP for a target bit rate, to optimize the rate-distortion performance. Considering the overall RD performance, the encoder should determine whether to encode the video at a higher frame rate but with significant quantization, or to encode at a lower frame with less quantization.

On the other hand, in scalable video coding (SVC), temporal, spatial and quality scalabilities are implemented to make the video stream scalable. It allows transmission and decoding of partial streams according to varying network conditions. In SVC, a bitstream extractor is needed to extract sub-bitstream. A very simple and inefficient method is to randomly discard bitstream units until the desired bit rate is achieved. For better performance, rate-distortion optimized extraction method can be used. No matter what kind of methods are to be employed, an analytical rate model is useful for a computational solution if without resorting to exhaustive search.

To solve the above problems, a more advanced two-dimensional (2-D) rate model is needed, where the relationship between the bit-rate, the QP and the frame rate is formulated. The integration of temporal resolution with quantization on the perceptual quality has been addressed in [7][8]. However, their joint influence on the rate is still not clear. To the best of the authors' knowledge, the only report on the 2-D rate model is with regard to scalable video coding [9], where the model is empirically developed according to experimental observations.

In this paper, we address how to build a 2-D rate model. In the proposed work, extensive experiments and theoretical analysis are performed in modeling the impact of the quantization step and the frame rate on the bit-rate. Through linking the frame rate and the quantization step to the mean absolute difference (MAD), the coding bit-rate is modeled as a 2-D function. Using the proposed 2-D rate model, the encoding frame rate as well as QP can be determined given a targeted bit-rate for video coding. Such a model is also useful for bitstream extraction in scalable video coding.

The reminder of this paper is organized as follows. The proposed 2-D rate model is described in section 2. To validate the performance of the proposed model, experimental results are reported in section 3. Section 4 closes this paper with concluding remarks.

2 Two-Dimensional Rate Model

After encoding a video sequence, the average bit–rate ($\overline{R}$) of the sequence can be expressed as

$$\overline{R} = f \times \frac{1}{K} \sum_{i=1}^{K} R_i = f \times \overline{R'}, \tag{1}$$

where K is the number of coded frames, f is the original frame rate, and $\overline{R'}$ is the average number of bits per frames. Intuitively, $\overline{R}$ decreases linearly as f decreases, however, changing f will impact statistics of the residual because of predictive coding, and then $\overline{R'}$ will be influenced, which makes this problem complicated. In [10], it has been confirmed that $\overline{R'}$ increase as f decreases according to experimental

results. In the following we will quantitatively investigate the impact of changing f on $\overline{R'}$, based on the quadratic R-Q model. Then the impact of the frame rate on the bit-rate can be evaluated.

2.1 Analysis of Quadratic Rate Model

One of the most widely adopted R-Q models in video coding presents itself in a quadratic polynomial [3]. An improved version of the quadratic R-Q model distinguishes header bits and texture bits, with the MAD introduced as the complexity measure [11]:

$$R_{texture} = MAD \times (a_1 \times q^{-1} + b_1 \times q^{-2}), \tag{2}$$

where $R_{texture}$ denotes the amount of bits to encode the texture information, q is the quantization step, a_1 and b_1 are parameters, and MAD is computed using the motion-compensated residual for the luminance component.

The MAD varies with the following two factors in video coding. (i) Frame rate. Encoding at a lower frame rate is equivalent to down-sampling the original sequence in the temporal domain. Through predictive coding, the MAD after motion estimation would be different for the same video scene at different frame rates. (ii) Quantization step. In the encoding process, reconstructed frames will be used as reference. Therefore, different quantization steps will lead to different distortions, resulting in different MAD values. Armed with the above arguments, the relationship between the MAD and the frame rate as well as the quantization step is built, based on which a new two-dimensional rate model is proposed.

2.2 Modeling MAD and Frame Rate of Original Video

Let $f_n(x, y)$ represent the pixel value at (x, y) in the nth frame in the original video sequence, M and N be the width and height of a frame respectively, and $MAD_{n,m}$ be the MAD value between the mth frame and the nth frame. Denoting the horizontal and the vertical components of the motion vector at (x, y) in the mth frame which references the nth frame as $mvx_{n,m}(x, y)$ and $mvy_{n,m}(x, y)$, respectively. Then

$$MAD_{n,n+1} = \frac{1}{M \times N} \sum_x \sum_y \left| f_{n+1}(x + mvx_{n,n+1}(x, y), y + mvy_{n,n+1}(x, y)) - f_n(x, y) \right| \tag{3}$$

$$MAD_{n+1,n+2} = \frac{1}{M \times N} \sum_x \sum_y \left| f_{n+2}(x + mvx_{n+1,n+2}(x, y), y + mvy_{n+1,n+2}(x, y)) - f_{n+1}(x, y) \right|. \tag{4}$$

Suppose that the motion vectors of successive frames are coherent, which holds for most frames with similar motion and background in the same video scene. Considering the MAD between the nth and the $(n+2)th$ frame

$$\begin{aligned} MAD_{n,n+2} &= \frac{1}{M\times N}\sum_x\sum_y\left|f_{n+2}(x+mvx_{n,n+2}(x,y),y+mvy_{n,n+2}(x,y))-f_n(x,y)\right| \\ &= \frac{1}{M\times N}\sum_x\sum_y\left|f_{n+2}(x+mvx_{n,n+1}(x,y)+mvx_{n+1,n+2}(x,y),y+mvy_{n,n+1}(x,y)+mvy_{n+1,n+2}(x,y))\right. \\ &\quad \left.-f_{n+1}(x+mvx_{n,n+1}(x,y),y+mvy_{n,n+1}(x,y))\ +f_{n+1}(x+mvx_{n,n+1}(x,y),y+mvy_{n,n+1}(x,y))-f_n(x,y)\right| \end{aligned} \tag{5}$$

According to inequality rules, and expending (5), we can obtain

$$\begin{aligned} MAD_{n,n+2} = \alpha(MAD_{n,n+1}+MAD_{n+1,n+2})+\beta(MAD_{n+1,n+2}-MAD_{n,n+1}) \\ +\gamma(MAD_{n,n+1}-MAD_{n+1,n+2}) \end{aligned}, \tag{6}$$

where α, β and $\gamma\in(0,1)$, and $\alpha+\beta+\gamma=1$. Here α, β and γ are probability values, corresponding to each component in (6).

In the same video scene with motion coherence, it can be assumed that

$$MAD_{n,n+1}\approx MAD_{n+1,n+2}=MAD_{org}\,. \tag{7}$$

So (6) can be expressed as

$$MAD_{n,n+2}=2\alpha MAD_{org}\,. \tag{8}$$

Based on the above analysis and extend (8) for any m and n, we have

$$MAD_{n,m}=\begin{cases} MAD_{org} & m-n=1 \\ (m-n-1)\lambda MAD_{org} & m-n\geq 2 \end{cases}, \tag{9}$$

where $\lambda\in(0,1)$. Accordingly, the average MAD of all frames in a sequence can be formulated as a function of the frame rate as follows:

$$MAD(t)=a_2\times\frac{T_{\max}}{t}+b_2\,, \tag{10}$$

where a_2 and b_2 are model parameters. $T_{\max}$ and t in (10) denote the maximum frame rate and the frame rate, respectively.

To verify the (10), experiments have been carried out using the H.264 reference software JM8.6 [12]. Four video sequences in CIF (352×288), i.e., "Mother-daughter", "Container", "Mobile", "News", were encoded under the frame rates of 5fps, 7.5fps, 15fps, 30fps, respectively. Here 30fps is the maximum frame rate. The reference frame is not the reconstructed signal but the corresponding original signal. In all simulations given in this paper, the first IDR frame was intra-coded only, and the remaining frames in a video sequence were inter-coded (P-frame) with all available modes. The number of reference frame was 1. The experimental results are plotted in Figure 1, where a linear relationship between the MAD of the original sequence and the reciprocal of frame rate can be generally observed.

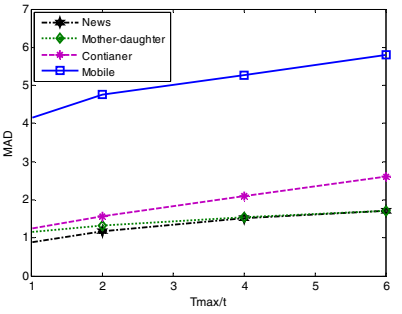


Fig. 1. The relationship between frame rate and MAD of original video

2.3 Modeling MAD and Quantization Step

To see how quantization step influences the MAD, we encoded several test video sequences using the H.264 reference software JM8.6 [12] and measured the actual MAD corresponding to different quantization steps. Specifically, four video sequences , "Mother-daughter", "Container", "Mobile", "News", all in CIF (352×288), were encoded with the quantization parameter (QP) from 15 to 36. In the results reported in this paper, the frame rate are 5fps, 7.5fps, 15fps, 30fps. Using the H.264 mapping between q and QP , i.e., $q = 2^{(QP-4)/6}$, the corresponding quantization step can be computed. The experimental results are plotted in Figure 2. The observations suggest a linear relationship between the MAD and the quantization step at a certain frame rate.

Therefore, at a certain frame rate, the average MAD varies with the quantization step as

$$MAD(q) = a_3 \times q + b_3 , \tag{11}$$

where a_3 and b_3 are parameters, and q is the quantization step.

2.4 The Proposed 2-D Rate Model

According to the above theoretical analysis and experimental observations, combining formula (10) and (11), we propose the following model:

$$MAD(q,t) = a \times q + b \times \frac{T_{\max}}{t} + c , \tag{12}$$

where a , b and c are model parameters, q and t donate the quantization step and the frame rate, respectively, and $T_{\max}$ is the maximum frame rate.

Then substituting (12) into (2), the coding bit-rate can be modeled as

$$\begin{aligned} R(q,t) &= MAD(q,t) \times (X_1 \times q^{-1} + X_2 \times q^{-2}) \\ &= (a \times q + b \times \frac{T_{\max}}{t} + c) \times (X_1 \times q^{-1} + X_2 \times q^{-2}) , \end{aligned} \tag{13}$$

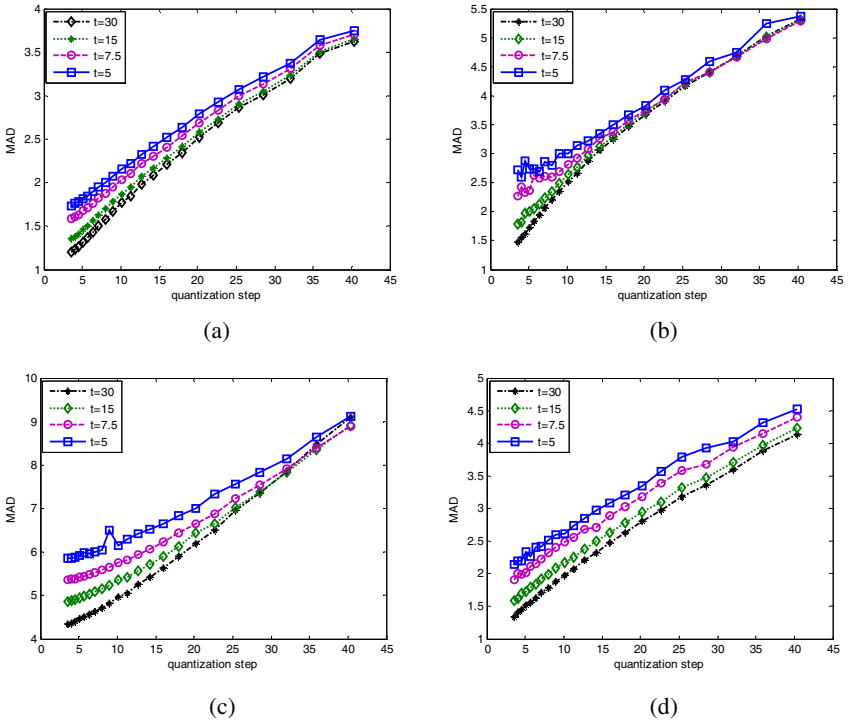


Fig. 2. The relationship between MAD and quantization step(a)"Mother-daughter" (b)"Container" (c) "Mobile" (d) "News

where X_1 and X_2 are model parameters. This rate model is two-dimensional which varies with the frame rate as well as the quantization step.

3 Experimental Results

The performance of the proposed 2-D rate model has been evaluated by using the H.264 video codec of reference software JM8.6 [12]. Standardized test sequences in CIF (352×288) were used. For all considered sequences, the coding video frame rates were 30f/s, 15f/s, 7.5f/s, 5f/s, respectively, and 30f/s was the maximum frame rate. Each sequence was coded with the QP of 15, 20, 25, and 30, respectively. Using the H.264 mapping between q and QP, the corresponding quantization steps were 3.56, 6.35, 11.31, and 20.16. For the sake of conciseness the results reported in this paper include only four test sequences: "Mother-daughter", "Container", "Mobile", "News". The model parameters a, b and c, were obtained by linear regression using the measured and predicted MAD corresponding to all q and t in (12). Parameters X_1 and X_2 were obtained by linear regression using (13). The actual rate data of four test sequences with different combinations of q and t, and the corresponding

estimated rates via the proposed model (13) are illustrated in Figure 3. RM denotes the value of predicted rates. From Figure 3 we observe that the model predictions fit well with the experimental rate points, with an average Pearson correlation coefficient (PCC) of 0.9917 and average relative error (RE) of only 8.38% over four video sequence. Table 1 listed the RE and the PCC between measured and predicted rates.

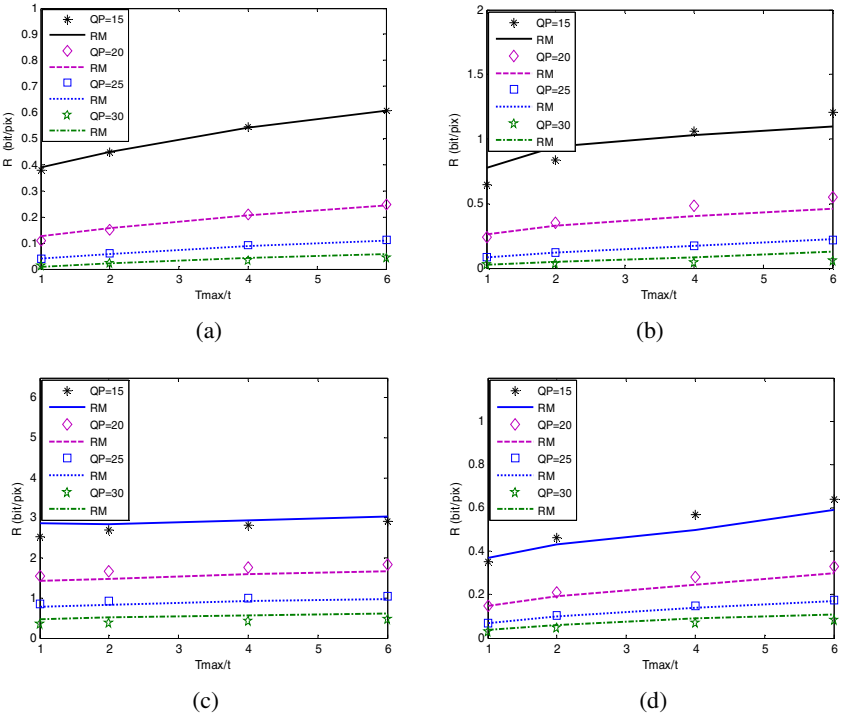


Fig. 3. Measured rate points and predicted rates using the 2-D rate model (13).(a)"Mother-daughter" (b)"Container" (c)"Mobile" (d)"News".

Table 1. PCC and RE between mesured and predicted rates

	News	Container	Mother-daughter	Mobile
PCC	0.9957	0.9855	0.9993	0.9866
RE	0.0923	0.1167	0.0301	0.0961

4 Conclusions

In this paper the impact of the frame rate and the quantization step on the MAD is examined and modeled through theoretical analysis and experimental validation. The corresponding model is brought into the quadratic R-Q model to obtain a new 2-D rate model, in terms of the quantization step and the frame rate. The accuracy of the

proposed 2-D rate model has been verified. Together with a 2-D distortion model, the proposed rate model is not only useful for non-scalable video encoder to determine the encoding frame rate and QP for a target bit-rate, but also useful in scalable video encoding and bitstream extraction.

Acknowledgements

This work was supported by the National Science Foundation of China (60902052, 60902081), and the NPU Foundation for Fundamental Research (JC201038).

References

1. Katto, J., Ohta, M.: Mathematical analysis of MPEG compression capability and its application to rate control. In: Proc. IEEE ICIP, Washington, DC, pp. 555–559 (1995)
2. Corbera, J., Lei, S.: Rate Control for Low-Delay Video Communications. ITU Study Group 16, Video Coding Experts Group, Documents Q15-A-20. Portland (1997)
3. Chiang, T., Zhang, Y.: A new rate control scheme using quadratic rate distortion model. IEEE Trans. on Circuits and Systems for Video Tech. 7(1), 246–250 (1997)
4. He, Z., Mitra, S.K.: A unified rate-distortion analysis frame work for transform coding. IEEE Trans. on Circuits and Systems for Video Tech. 11(12), 1221–1236 (2001)
5. Chen, J., Hang, H.: Source model for transform video coder and its application-part I, II. IEEE Trans. on Circuits and Systems for Video Tech. 7(2), 287–311 (1997)
6. Wei, J., Soong, B., Li, Z.: A new rate-distortion model for video transmission using multiple logarithmic functions. IEEE Signal Processing Letters 11(8), 694–697 (2004)
7. Lai, W., Gu, X.-D., Wang, R.-H., Dai, L.-R., Zhang, H.-J.: Perceptual video streaming by adaptive spatial-temporal scalability. In: Aizawa, K., Nakamura, Y., Satoh, S. (eds.) PCM 2004. LNCS, vol. 3332, pp. 431–438. Springer, Heidelberg (2004)
8. Zhai, G.T., Cai, J.F., Lin, W.S., Yang, X.K., Zhang, W.J.: Three dimensional scalable video adaptation via user-end perceptual quality assessment. IEEE Trans. on Broadcasting 54(3), 719–727 (2008)
9. Wang, Y., Ma, Z., Ou, Y.-F.: Modeling Rate and Perceptual Quality of Scalable Video as Functions of Quantization and Frame Rate and Its Application in Scalable Video Adaptation. In: Proceedings of 17th International Packet Video Workshop, pp. 1–9 (2009)
10. Lee, J.-W.: Bit Allocation For MPEG-4 Video Coding With Spatio-Temporal Tradeoffs. IEEE Trans. on Circuits and Systems for Video Tech. 13(6), 488–502 (2003)
11. Lee, H., Chiang, T., Zhang, Y.: Scalable rate control for MPEG-4 Video. IEEE Trans. on Circuits and Systems for Video Tech. 10(6), 878–894 (2000)
12. H.264/AVC reference software version JM8.6, `http://iphome.hhi.de/suehring/tml/`

Coherent Integration Loss due to Transmit Antenna Scanning Modulation Effect on Passive Bistatic Radar

Caisheng Zhang[1,2], Xiaoming Tang[1], and Jiahui Ding[2]

[1] Research Inst. of Information Fusion, Naval Aeronautical and Astronautical University, Yantai, Shandong, China
[2] Nanjing Research Institute of Electronics Technology, Nanjing, China
caifbi2008@yahoo.com.cn

Abstract. This paper explored the transmit antenna scanning modulation effect on the signal to noise ratio (SNR) loss in a noncooperative illuminator based bistatic radar. The general expression of normalized filed pattern for widely used antenna was analyzed. The direct-path signal and target echo modulated by the transmit antenna, which is mechanically scanned in azimuth, were described. Mathematical representations for direct-path signal and target echo were proposed with respect to the analytical expression of antenna's pattern. The SNR loss of the coherent integration via cross-correlation processing was derived. The relationship between SNR loss and SNRs of direct-path channel and target echo channel was evaluated. The simulation result shows that the SNR loss caused by antenna modulation can be ignored if the average SNR in direct path is greater than 15 dB. However, the SNR loss deteriorates rapidly if the visual angle becomes larger. Some suggestions on choosing the IF bandwidth and data selection were presented.

Keywords: Passive Bistatic Radar; Coherent Integration Loss; Antenna Pattern; Scanning Modulation Effect.

1 Introduction

In recent years, target detection and tracking systems based on illuminators of opportunity have received significant attention and resources, which is so-called passive coherent location system[1-3]. The well known advantages of such systems over conventional monostatic radar include low cost, immunity to ECM threats, etc. Most of the systems exploit radio and television broadcasting signals as noncooperative emitters. Due to the transmitting antennas of these kind of system is omnidirectional, there is no need to consider the effect of modulation caused by the scanning of the transmit antenna [4-6].

However, in the passive bistatic radar system which using radar emitters as sources of opportunity, it is desired to take the effect of modulation into account introduced by mechanical scanning antenna. Even if the system is synchronized in space, the intercepted the direct path wave and the target echo are modulated by different lobes of the transmitting antenna. Then, the gain achieved by coherent integration will be deteriorated. Therefore, the SNR's improvement factor will be limited by the transmitting

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 163–170.
springerlink.com

antenna scanning modulation[7]. [8]discussed the problems for cooperative bistatic radar from the relationship of the probability density of the target echo and its reception angle. The loss caused by radar antenna pattern in different geometric configuration for the cooperative bistatic radar has been illustrated. However, there is little literature focused on the antenna pattern loss in the passive bistatic radar.

This paper concentrates on the property of passive bistatic radar based on noncooperative conventional monostatic radar emitters. The model of transmitting antenna scanning modulation will be proposed in section 2. The coherent integration loss caused by antenna scanning modulation is dealing with in section 3. The analytical expression for SNR loss is derived, and the corresponding simulation analysis is presented as well, and conclusions are made in section 4.

2 Problem Description

A scenario with non-cooperative radar illuminator considered in this paper is illustrated in Fig. 1. Assume that the transmitter is mechanically scanned in azimuth, and the receiver is stationary while the target is moving. Both the transmitter and receiver are focusing on the moving target. The receiving system intercepts the direct-path waveform transmission through a reference antenna when it tunes to the transmitting frequency, and target reflection echoes are intercepted by the target antenna. Time and phase synchronization need to be completed via direct-path signal. Then, surveillance and early-warning may be achieved in the area of interest.

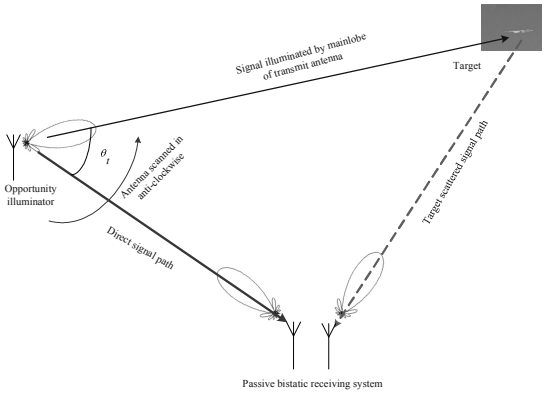


Fig. 1. Schematic of passive bistatic pulsed radar system

When space synchronization between the transmitter and receiver is completed, the antennas' main lobes will cover the target. Then the target echo can be intercepted continuously. While at this time, the direct path reference signal is intercepted via the transmit antenna's sidelobe radiation. It is found that the propagation factor between the direct path channel and the target channel is different during the coherent dwell time. The transmitting visual angle is changing with the movement of target. The geometry of the passive bistatic system is time variant as well. Therefore, it is desired to consider the modulation effect on the output of coherent integration.

3 Normalized Field Lobe Pattern for General Antenna

For most radar antennas, such as parabolic, horn, or array antenna, if the relationship between the antenna aperture size d and the wavelength λ satisfies $d/\lambda > 4 \sim 5$, the normalized field lobe pattern can be approximately expressed as [9]

$$F(\theta) = \frac{\sin(\pi d\theta/\lambda)}{\pi d\theta/\lambda} . \tag{1}$$

where θ is the angle deviation relative to the antenna main lobe, then the relative gain coefficient of different angles is

$$\frac{G(\theta)}{G(0)} = \left[\frac{\sin(\pi d\theta/\lambda)}{\pi d\theta/\lambda}\right]^2 . \tag{2}$$

where $G(0)$ is the maximum gain corresponding to antenna's main lobe, the maximum gain of sidelobes appear at angles when $\sin(\pi d\theta/\lambda) = 1$, thus the relative gain coefficients at the maximum sidelobes are

$$\frac{G(\theta)}{G(0)} = \frac{1}{(\pi d\theta/\lambda)^2} . \tag{3}$$

If $\pi d\theta/\lambda = n\pi (n = 0,1,2,\cdots)$, then $\sin(\pi d\theta/\lambda) = 0$, which are corresponding to the nulls of the antenna's pattern.

4 Coherent Integration Loss due to Antenna's Modulation

4.1 Model for the Received Signals

Let the envelope of the transmit signal be $\tilde{s}_T(t)$, then the direct path signal's envelope can be expressed as $\tilde{s}_D(t) = k_D F_D(t)\tilde{s}_T(t-\tau_D) + \tilde{n}_D(t)$, and the echo signal's envelope will be $\tilde{s}_R(t) = k_R F_R(t)\tilde{s}_T(t-\tau_R) + \tilde{n}_R(t)$, where $\tilde{n}_D(t)$ and $\tilde{n}_R(t)$ are the noisy terms in the direct path channel and the target channel, with zero mean and of $N_0/2$ variance, k_D and k_R are attenuation factors for the direct path channel and the target channel, respectively. $F_D(t)$ and $F_R(t)$ are propagation factors, which are time-varying parameters with the scanning, τ_d and τ_R are time delays of the direct path signal and the target echo path.

Suppose that the pattern propagation factor in the antenna's main beam can be approximately expressed as

$$F_R(t) = \exp\left(-1.39t^2/T^2\right) . \tag{4}$$

Similarly, according to Eq.(3), the pattern propagation factor in the visual angle θ_t can be approximated as

$$F_D(t) = \lambda / \pi d\theta_t \exp\left(-1.39t^2 / T^2\right) . \tag{5}$$

where $-T/2 \le t \le T/2$, herein t corresponding to the antenna's sweeping time in the 3 dB beam width.

Without loss of generality, let the rotating speed of the antenna and the pulse repetition frequency (PRF) of the emitter's signal be constant. And suppose that the receiving system can intercept each pulse accurately during the coherent dwell time. Then coherent integration can be adopted using the target echo pulse train obtained during the space synchronous condition. The integration gain depends upon the target coherent dwell time. In general, the coherent integration time or dwell time target is equal to time of the antenna's 3 dB beam width illuminated on the target.

4.2 The Ideal Output SNR of Cross-Correlation Processing

According to [4-6], passive coherent processing is widely used to detect and estimate the presence of a target. Typically, the received signals are cross correlated over a two-dimensional ambiguity surface to compensate for time delay and Doppler difference of the target echo. If there is no scanning modulation effect in the direct path channel, and the direct path signal is no noise-free, then the expectation of the cross correlation output during the interval $[-T/2, T/2]$ is

$$E_S\left[\tilde{y}(T)\right] = E\left\{\int_{-T/2}^{T/2} k_R k_D F_R(t)\tilde{S}_T(t-\tau_R)\tilde{S}_T^*(t-\tau_D-\tau)\mathrm{d}t\right\} + E\left\{\int_{-T/2}^{T/2} k_D \tilde{n}_R(t)\tilde{S}_T^*(t-\tau_D-\tau)\mathrm{d}t\right\} . \tag{6}$$

where T is the total dwell time. When $\tau = \tau_R - \tau_D$, the output will be

$$E_S\left[\tilde{y}(T)\right] = E\left\{\int_{-T/2}^{T/2} k_R k_D F_R(t)\left|\tilde{S}_T(t-\tau_R)\right|^2 \mathrm{d}t\right\} + E\left\{\int_{-T/2}^{T/2} k_D \tilde{n}_R(t)\tilde{S}_T^*(t-\tau_R)\mathrm{d}t\right\} . \tag{7}$$

The output SNR of Cross-correlation process is defined as

$$SNR = \frac{\left|E_S\left[\tilde{y}(T)\right] - E_{NS}\left[\tilde{y}(T)\right]\right|^2}{\mathrm{var}_S\left[\tilde{y}(T)\right]} . \tag{8}$$

where $E_S\left[\tilde{y}(T)\right]$ denotes the expectation of the echo in the target channel, which contains an actual target, and $\mathrm{var}_S\left[\tilde{y}(T)\right]$ is its variance, $E_{NS}\left[\tilde{y}(T)\right]$ denotes the expectation of the echo in the target channel which contains noise only. If the noise in the receiving system is zero mean, then $E_{NS}\left[\tilde{y}(T)\right] = 0$, and we have

$$E_S[\tilde{y}(T)] = E\left\{\int_{-T/2}^{T/2} k_R k_D F_R(t)\left|\tilde{S}_T(t)\right|^2 dt\right\} . \tag{9}$$

$$\mathrm{Var}_S\left[\tilde{y}(T)\right] = k_D^2 \frac{N_0}{2}\left|\int_{-T/2}^{T/2} \tilde{S}_T^*(t)\,dt\right|^2 . \tag{10}$$

Thus, the peak output SNR in the ambiguity surface will be

$$(SNR)_{\mathrm{idl}} = \frac{2\left|\int_{-T/2}^{T/2} k_R k_D F_R(t)\left|\tilde{S}_T(t)\right|^2 \mathrm{d}t\right|^2}{k_D^2 N_0\left|\int_{-T/2}^{T/2} \tilde{S}_T^*(t)\,\mathrm{d}t\right|^2} . \tag{11}$$

Without loss of generality, let signal amplitude be a constant for simplify the expression, i.e. A, then

$$(SNR)_{\mathrm{idl}} = \frac{2k_R^2 A^2}{N_0 T}\left|\int_{-T/2}^{T/2} \exp\left(-1.39t^2/T^2\right)\mathrm{d}t\right|^2 . \tag{12}$$

Thus

$$(SNR)_{\mathrm{idl}} = \frac{k_R^2 A^2}{N_0 T}\left[\frac{\sqrt{\pi}T}{1.18}\mathrm{erf}(0.59)\right]^2 . \tag{13}$$

where $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-u^2}\mathrm{d}u$.

4.3 SNR Loss Caused by Antenna Scanning Modulation

The actual cross-correlation processing is performed with the received signals that modulated by the antenna's pattern during the coherent dwell time. Therefore, the cross-correlation output should be

$$\tilde{y}(T) = \int_{-T/2}^{T/2}\left[k_R F_R(t)\tilde{s}_T(t-\tau_R) + \tilde{n}_R(t)\right]\left[k_D F_D(t)\tilde{s}_T^*(t-\tau_D-\tau) + \tilde{n}_D^*(t-\tau)\right]\mathrm{d}t . \tag{14}$$

when $\tau = \tau_R - \tau_D$, the peak of correlation output is

$$\begin{aligned}\tilde{y}(T) = &\int_{-T/2}^{T/2} k_R k_D F_R(t) F_D(t)\left|\tilde{s}_T(t-\tau_R)\right|^2 \mathrm{d}t + \int_{-T/2}^{T/2} k_R F_R(t)\tilde{s}_T(t-\tau_R)\tilde{n}_D^*(t-\tau_R-\tau_D)\,\mathrm{d}t \\ &+ \int_{-T/2}^{T/2} k_D F_D(t)\tilde{s}_T^*(t-\tau_R)\tilde{n}_R(t)\,\mathrm{d}t + \int_{-T/2}^{T/2}\tilde{n}_R(t)\tilde{n}_D^*(t-\tau_R-\tau_D)\,\mathrm{d}t .\end{aligned} \tag{15}$$

Obviously, we have $E_{\mathrm{NS}}\left[\tilde{y}(T)\right] = 0$. For the purpose of comparison, we also assume that signal amplitude is A

$$E_S\left[\tilde{y}(T)\right] = k_R k_D A^2 \int_{-T/2}^{T/2} F_D(t) F_R(t)\,\mathrm{d}t . \tag{16}$$

$$\mathrm{Var}_{\mathrm{NS}}\left[\tilde{y}(T)\right] = k_R^2 A^2 \frac{N_0}{2}\int_{-T/2}^{T/2} F_R^2(t)\,\mathrm{d}t + k_D^2 A^2 \frac{N_0}{2}\int_{-T/2}^{T/2} F_D^2(t)\,\mathrm{d}t + \left(\frac{N_0}{2}\right)^2 T . \tag{17}$$

Substitute $F_R(t)$ and $F_D(t)$ into Eq.(17), we have

$$SNR = \frac{k_R^2 k_D^2 A^4 \int_{-T/2}^{T/2} \exp\left(-2\times 1.39t^2/T^2\right) \mathrm{d}t \left(\frac{\pi d\theta}{\lambda}\right)^{-2} \int_{-T/2}^{T/2} \exp\left(-2\times 1.39t^2/T^2\right) \mathrm{d}t}{\frac{N_0}{2} k_R^2 A^2 \int_{-T/2}^{T/2} \exp\left(-2\times 1.39t^2/T^2\right) \mathrm{d}t + k_D^2 A^2 \left(\frac{\pi d\theta}{\lambda}\right)^{-2} \int_{-T/2}^{T/2} \exp\left(-2\times 1.39t^2/T^2\right) \mathrm{d}t + \frac{N_0}{2} T} \tag{18}$$

4.4 Quantity Analysis of SNR Loss

According to the assumptions in Section 4.1, the average SNR direct channel and target channel can be expressed as follow

$$(SNR)_d = k_D^2 A^2 \left(\frac{\pi d\theta_t}{\lambda}\right)^{-2} \left(\frac{N_0}{2} T\right)^{-1} \int_{-T/2}^{T/2} \exp\left(-2\times 1.39t^2/T^2\right) \mathrm{d}t \tag{19}$$

$$(SNR)_r = k_R^2 A^2 \int_{-T/2}^{T/2} \exp\left(-2\times 1.39t^2/T^2\right) \mathrm{d}t \left(\frac{N_0}{2} T\right)^{-1} \tag{20}$$

Then, the SNR loss with respect to the ideal output of match filter will be

$$L = \frac{\int_{-T/2}^{T/2} \exp\left(-2\times 1.39t^2/T^2\right) \mathrm{d}t}{\left|\int_{-T/2}^{T/2} \exp\left(-1.39t^2/T^2\right) \mathrm{d}t\right|^2} \frac{(SNR)_d}{(SNR)_r + (SNR)_d + 1} \approx \frac{(SNR)_d}{(SNR)_r + (SNR)_d + 1} \tag{21}$$

From Eq.(21), it is found that the SNR loss caused by the modulation of antenna depends on the average SNR in the coherent dwell time of the direct path and the target channel.

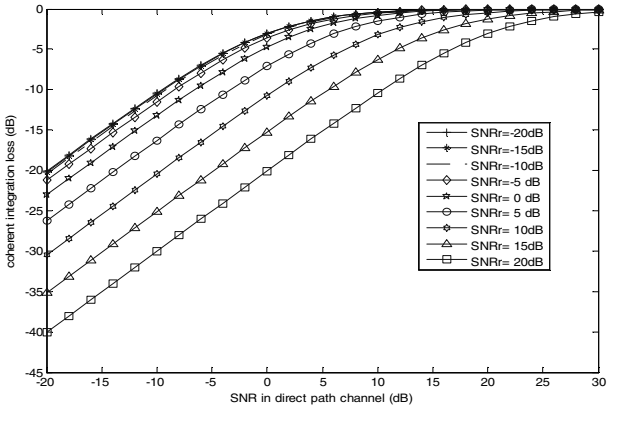


Fig. 2. Coherent integration loss V.S. the average SNR in the direct path

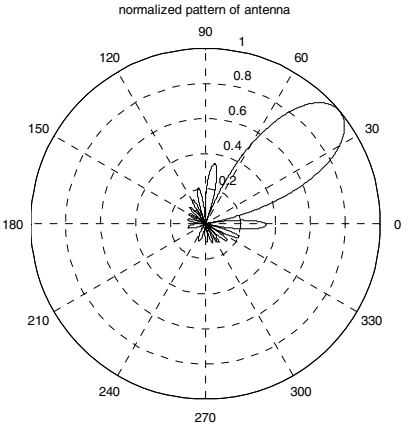


Fig. 3. Normalized field lobe pattern at some instant during the antenna rotating

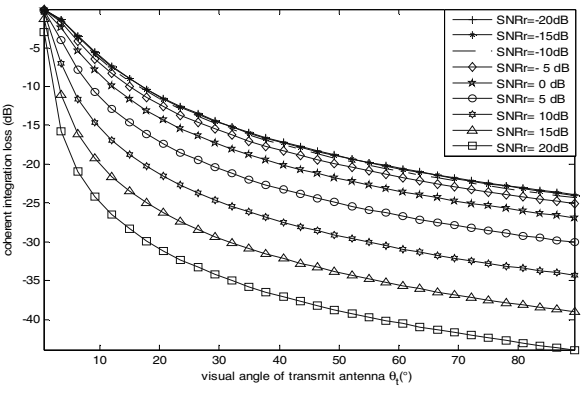


Fig. 4. The curves of the SNR loss factors V.S. the visual angle

Therefore, in order to improve the average SNR in each channel, we should make full use of range cells in every dwell time that contain effective target echo for cross-correlation processing. In general, the mismatch between the receivers' noise bandwidth and signal's the will make the average SNR deteriorate. Thus, it is desired to match the receivers' noise bandwidth with the non-cooperative emitter's signal bandwidth. Generally speaking, the higher is the average SNR in the direct path channel, the smaller is the SNR loss factor, as can be seen in Fig. 2. When the average SNR in the direct path channel is fixed, the SNR loss will increase if the average SNR in the target channel increasing gradually. In practice, the actual SNR in the target channel is much smaller than 0 dB. If the average SNR in direct path is greater than 15 dB, the SNR loss caused by antenna modulation can be ignored approximately.

Fig.3 shows that a normalized field lobe pattern for the aperture $d = 3$ m and $\lambda = 0.03$ m . Let the SNR of the direct path received via its main lobe illumination

be 20 dB. If the direct path signal received from different sidelobes is adopted as reference signal of cross correlation processing, then the SNR loss is illustrated in Fig. 4. According to the curves of SNR loss V.S. θ_t , the SNR losses deteriorate rapidly as the visual angle becomes larger gradually. In fact, the average SNR in the direct path declines sharply as the level of the sidelobes are decreasing quickly. Therefore, only the signals intercepted from relatively higher sidelobes can fulfill the target detection.

5 Conclusion

The modulation phenomenon caused by the rotating of transmit antenna was illustrated in detail. The general expression of normalized filed pattern for widely used antenna was introduced for the purpose of quantity analysis. The direct-path signal and target echo modulated by the transmit antenna, were described in detail. Mathematical representations for direct-path signal and target echo were proposed with respect to the analytical expression of antenna's pattern. Then, the SNR loss of the coherent integration via cross-correlation processing was derived. The relationship between SNR loss and SNRs of direct-path channel and target echo channel was evaluated by simulation. According to the simulation result, it is found that the SNR loss caused by antenna modulation can be ignored if the average SNR in direct path is greater than 15 dB. However, the SNR losses deteriorate rapidly if the visual angle becomes larger gradually. In order to increase the average SNR, it is also desired to match the IF receivers' noise bandwidth with the non-cooperative emitter's signal bandwidth.

References

1. Wang, X., Kuang, Y., Chen, Z.: Surveillance radar technology, pp. 366–369. Publishing House of Electronic Industry, Beijing (2008) (in Chinese)
2. Yang, Z., Zhang, Y., Luo, Y.: Bistatic/Multistatic Radar System, pp. 209–214. Publishing House of National Defense Industry, Beijing (1998) (in Chinese)
3. Song, J., He, Y., Cai, F., et al.: Overview of Passive Radar Technology Based on Non-cooperative radar Illuminator. Systems Engineering and Electronics 31(9), 2151–2156 (2009) (in Chinese)
4. Howland, P.E., Maksimiuk, D., Reitsma, G.: FM Radio Based Bistatic Radar. IEE Radar, Sonar and Navigation 152(3), 107–115 (2005)
5. Griffiths, H.D., Baker, C.J.: Passive Coherent Location Radar Systems Part 1: Performance Prediction. IEE Radar, Sonar and Navigation 152(3), 153–159 (2005)
6. Kulpa, K.S., Czekała, Z.: Masking Effect and its Removal in PCL Radar. IEE Radar, Sonar and Navigation 152(3), 174–178 (2005)
7. Thomas, D.D.: Synchronization of noncooperative bistatic radar receivers. Syracuse University, N.Y (1999)
8. Wang, Y., Xu, S.: Antenna Pattern Loss Analysis of Bistatic Radar. Systems Engineering and Electronics 25(10), 1219–1222 ,1273 (2003) (in Chinese)
9. Zhang, Y.-S., Ningning, T., Guoqing, Z., et al.: Radar & Electronic Warfare Theory, pp. 28–34. Publishing House of Electronic Industry, Beijing (2004) (in Chinese)

Research on Compressive Sensing Based GPR Data Acquisition

Huimin Yu

College of Physics and Information Science, Hunan Normal University,
410081, Changsha, China
huiminyu@yahoo.cn

Abstract. Results in theory of compressive sensing enable the reconstruction of sparse signals from a small set of non-adaptive linear measurements by solving a convex optimization problem. Considering the sparse structure of actual target space in Ground Penetrating Radar (GPR) application, a data acquisition method based on random aperture compressive sensing (RACS) is studied in this paper, which requires the GPR transceiver to record only a minimum amount of samples through incoherent measurement at each aperture point, and only to measure a small number of random apertures in x-y plane of interested target space. Results indicate that the method allow much fewer sampling data.

Keywords: compressive sensing; GPR; random aperture; data acquisition.

1 Introduction

The efficient utilization of GPR system for above civil and military application areas nondestructive speculative searches not only lies on the performance of GPR hardware system, but also rests on the efficiency of GPR subsurface data acquisition, imaging and feature detection method [1][2][3][4]. Familiar GPR method such as Range Migration (RM) Method [5], Reverse Time Migration (RTM) Method [6] and time-domain Standard Back Projection (SBP) Method [7] [8]require fine spatial sampling and Nyquist rate time sampling of the received signals, or a high aperture density measurement, and then they perform matched filtering with the impulse response of the data acquisition process. They don't use any prior knowledge about the target space, such as the spatial sparsity of targets.

Compressive Sensing involves taking a relatively small number of non-traditional samples in the form of randomized projections that are capable of capturing the most salient information in original signal [9] [10]. In this paper we study data acquisition method for impulse Ground Penetrating Radars which based on random aperture compressive sensing by exploiting structure sparseness of the target space. First we give an introduction of compressive sensing theory. Then we present random aperture compressive sensing method and verify it's validity.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 171–177.
springerlink.com

2 Compressive Sensing

Consider a real-valued, one-dimensional, discrete-time signal x, which can be viewed as an N×1 column vector in R^N with elements x[n], n=1,2,…,N. Any signal in R^N can be represented in terms of a basis of N×1 vectors $\{\psi_i\}_{i=1}^N$. Using the N×N basis matrix $\psi=[\psi_1][\psi_2]\ldots[\psi_N]$ with the vectors $\{\psi_i\}$ as columns, a signal x can be expressed as [11] [12]

$$\mathbf{x}=\sum_{i=1}^{N} s_i\psi_i \quad \text{or} \quad \mathbf{x}=\Psi\mathbf{s} \tag{1}$$

where s is the N×1 column vector of weighting coefficients $s_i=\langle \mathbf{x},\Psi_i\rangle=\Psi_i{}^T\mathbf{x}$ and $\cdot^T$ denotes transposition.

In common data acquisition systems, the full N-sample signal x is acquired, and the complete set of transform coefficients s is computed via $\mathbf{s}=\Psi^T\mathbf{x}$, then the K largest coefficients are located and the (N-K) smallest coefficients are discarded. Compressive sensing acquires a compressed signal representation without going through the intermediate stage of acquiring N samples [13] [14].

Consider a general linear measurement process that computes M<N inner products between x and a collection of vectors $\{\phi_j\}_{j=1}^M$ as in $y_j=\langle \mathbf{x},\Phi_j\rangle$. Arrange the measurements y_j in an M×1 vector y and the measurement vectors ϕ_j^T as rows in an M×N matrix Φ. Then, by substituting Ψ from (1), y can be written as

$$\mathbf{y}=\Phi\mathbf{x}=\Phi\Psi\mathbf{s} \tag{2}$$

Compressive sensing theory indicate that if the matrix $\Phi\Psi$ has the Restricted Isometry Property (RIP) [15] [16] then it is possible to recover s exactly from O(Klog(N)) measurements by solving the following l_1–norm minimization problem

$$\min\|\hat{\mathbf{s}}\|_1 \quad s.t. \quad \mathbf{y}=\Phi\Psi\hat{\mathbf{s}} \tag{3}$$

Compressive sensing now has found a variety of interesting applications [17] [18]. This paper explains the application of Compressive sensing to data acquisition in GPR.

3 Random Aperture Compressive Sensing Method

For the i^{th} aperture point, the discrete data GPR sampled and received can be written as [19] [20]

$$\mathbf{R}_i=\left[R_i(t_0),R_i(t_0+\frac{1}{F_s})\cdots,R_i(t_0+\frac{N_s-1}{F_s})\right]^T \tag{4}$$

Where F_s is the sampling frequency, t_0 is the initial time of received i^{th} aperture point signal, and N_s is the number of temporal samples. Usually, F_s is large in order to

afford adequate sampling data for existing algorithm, such as RM, RTM, SBP algorithm to implement high-resolution subsurface imaging, and all aperture points in x-y plane of target space need to be measured. Based on compressive sensing, a signal sampling and recovery theory, random aperture compressive sensing algorithm (RACS) record only a minimum amount of samples through incoherent measurement at each aperture point, and chosen to measure only a small number of random apertures in x-y plane of interested target space, RACS utilize the acquired small amount of random aperture and incoherent measurement data to reconstruct the original target space, the random aperture compressive sensing data acquisition process can be presented as

$$\mathbf{D}_i = \Phi \mathbf{R}_i = \Phi \Psi_i \mathbf{b} \tag{5}$$

Where Φ is an $M \times N_s$ measurement matrix and $M \square N_s$, $\mathbf{D}_i$ is the $M \times 1$ compressive sensing acquisition data at i[th] aperture point.

By solving an l_1–norm minimization problem

$$\min \left\| \hat{\mathbf{b}} \right\|_1 \quad s.t. \quad \mathbf{D}_i = \Phi \Psi_i \hat{\mathbf{b}} \tag{6}$$

The target space indicator vector at i[th] aperture point can be reconstructed [21], and the original target space can be obtained by add all indicator vectors of chosen random aperture point.

4 Simulation Results

In simulation we set $N_s = 512$ and sampling interval to be $10ps$, width of Gaussian pulse $w = 80ps$. Virtual target space is shown in fig.1. The discrete samples of all aperture points in x-y plane are obtained according to formula (4). Instead of measuring all the space-time domain response at each aperture position, RACS form 20 inner product measurements at 32 randomly chosen aperture points making 640 measurements in total. The inner products is the product of time-domain response with Φ, which of size 20×512 with all entries drawn independently form $N(0,1/\sqrt{512})$.

The sparsity pattern vector for virtual space has length 2560 and there are 640 measurements which result in an underdetermined equation, $\mathbf{D} = \Phi \mathbf{R} = \Phi \Psi \mathbf{b}$. Least squares method provide a possible solution $\mathbf{b} = (\Phi\Psi)^T (\Phi\Psi(\Phi\Psi)^T)^{-1} \mathbf{D}$, the target space image for this is shown in fig.2 for compare. SBP using all of 512×256 space-time data, and do matched filtering of the measured data with the impulse response of the data acquisition process for each spatial location, and then add up the amplitude value of received signal at all aperture points with same round-trip delay value to obtain target space indicator vector and reconstruct target space, which result is shown in fig.3.

RACS utilizes 20 measurements of each aperture point, and measure 32 random aperture points to recover target space indicator vector, for the numerical solution of

(6) a convex optimization package called l_1-magic [22] is used, the target space image by RACS is shown in fig. 4.

It can be seen that the actual target positions are found correctly by RACS method, and the obtained image is much sparser compared to the SBP result in fig.3, even though the SBP result is obtained using all of the space-time data, since RACS exploits sparseness characteristic of the target space. The convex optimization result has less side lobe in the image since the problem (6) forces a sparse solution through the l1-norm minimization.

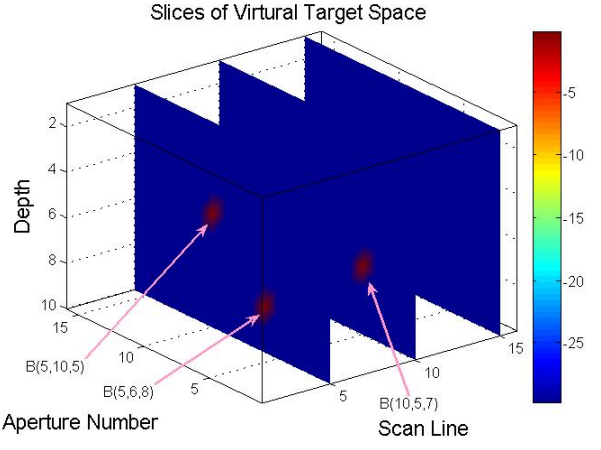


Fig. 1. Slices of Virtual Target Space.

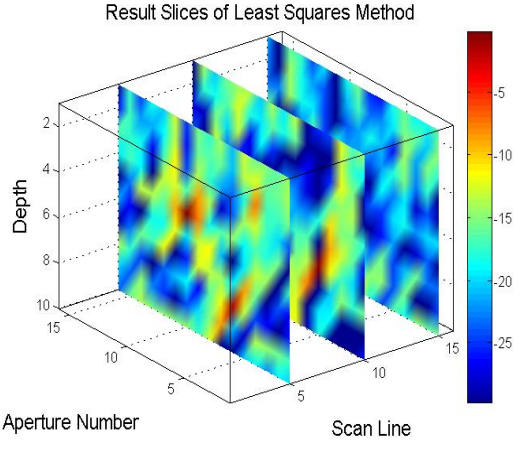


Fig. 2. Result Slices of Least Squares Methods.

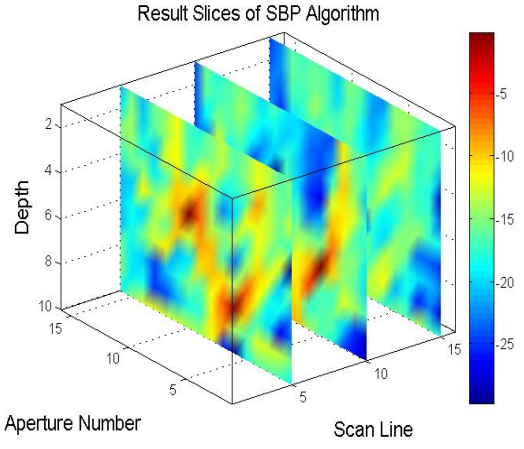


Fig. 3. Result Slices of SBP Algorithm.

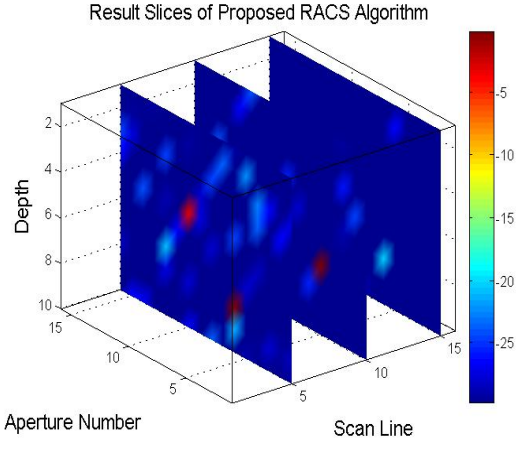


Fig. 4. Result Slices of Proposed RACS Algorithms.

5 Conclusion

None existing subsurface data acquisition method uses the spatial sparsity prior knowledge about the target space, and they need the GPR transceiver to work at high sampling frequency, high measurement aperture density in order to provide there enough data to practice subsurface imaging. In this paper we studied a data acquisition method for impulse GPR based on random aperture compressive sensing by exploiting sparseness in the target space, RACS method record only a minimum number of samples through incoherent measurement at each aperture point, and chosen to measure only a small number of random apertures in x-y plane of interested target space,

which proved to be much more robust to noise and fewer measurements needed when compared to existing subsurface data acquisition procedure.

Acknowledgments. The work was partially supported by the Natural Science Foundation of Hunan Province of China under Grant No.10JJ6102.

References

1. Daniels, D.J.: Surface-penetrating radar. Electronics and Communication Engineering Journal 8, 165–182 (1996)
2. Fang, G.Y.: The research activities of ultrawide-band (UWB) radar in China. In: IEEE International Conference on Ultra-Wideband, Singapore, September 24-26, pp. 43–45 (2007)
3. Daniels, D.J.: Introduction to subsurface radar. IEEE Proceeding 133(4), 278–326 (1999)
4. Daniels, D.J.: A study of surface penetrating radar and its application. Microwave Journal 37(2), 68–82 (1999)
5. Xu, X., Miller, E.L.: Optimization of migration method to locate buried object in lossy medium. In: IEEE International Geoscience and Remote Sensing Symposium, Toronto, Canada, June 24-28, pp. 337–339 (2002)
6. Leuschen, C.J., Plumb, R.G.: A matched-filter-based reverse-time migration algorithm for ground-penetrating radar data. IEEE Trans. on Geoscience and Remote Sensing 39(5), 929–936 (2001)
7. Carin, L., Geng, N., McClure, M., et al.: Ultra-wide-band synthetic-aperture radar for mine-field detection. IEEE Antennas and Propagation Magazine 41(1), 18–33 (1999)
8. Hayakawa, H., Nadamoto, A., Uesaka, S.: 3D radar imaging of buried objects using arbitrary scaning GPR. In: Eighth International Conference on Ground Penetrating Radar, Queensland, Australia, May 23-26, pp. 273–276 (2000)
9. Donoho, D.L.: Compressed sensing. IEEE Trans. on Information Theory 52(4), 1289–1306 (2006)
10. Candes, E.J., Romberg, J., Tao, T.: Stable signal recovery from incomplete and inaccurate measurements. Comm. on Pure and Applied Math. 59(8), 1207–1223 (2006)
11. Candes, E.J.: Compressive sampling. In: International Congress of Mathematicians, Madrid, Spain, vol. 3, pp. 1433–1452 (2006)
12. Candes, E.J., Romberg, J., Tao, T.: Sparsity and incoherence in compressive sampling. Inverse Problems 23, 969–985 (2006)
13. Haupt, J., Nowak, R.: Signal reconstruction from noisy random projections. IEEE Trans. on Information Theory 52(9), 4036–4048 (2006)
14. Baraniuk, R.: Compressive sensing. IEEE Signal Processing Magazine 24(4), 118–121 (2007)
15. Candes, E.J., Tao, T.: Decoding by linear programming. IEEE Trans. on Information Theory 51, 4203–4215 (2005)
16. Baraniuk, R., Davenport, M., DeVore, R., Wakin, M.: A simple proof of the restricted isometry property for random matrices. Constructive Approximation 28(3), 365–372 (2008)
17. Wakin, M.B., Laska, J.N., Duarte, M.F., et al.: An Architecture for Compressive Imaging. In: IEEE International Conference on Image Processing, Atlanta, GA, USA, October 8-11, pp. 1273–1276 (2006)

18. Lustig, M., Donoho, D.L., Pauly, J.M.: The application of compressed sensing for rapid MR imaging. Magnetic Resonance in Medicine 58(6), 1182–1195 (2007)
19. Baraniuk, R., Steeghs, P.: Compressive radar imaging. In: IEEE Radar Conference, Waltham, Massachusetts, USA, pp. 128–133 (April 2007)
20. Gurbuz, A.C., McClellan, J.H., Waymond, R.S.: Compressive sensing for GPR imaging. In: Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, November 4-7, pp. 2223–2227 (2007)
21. Candes, E.J., Rudelson, M., Tao, T., Vershynin, R.: Error correction via linear programming. In: The 46th Annual IEEE Symposium on Foundations of Computer Science, Pittsburgh, PA, USA, October 23-25, pp. 295–308 (2005)
22. l1-magic, `http://www.acm.caltech.edu/l1magic/`

A Projection Access Scheme Based on Prime Number Increment for Cone-Beam Iterative Reconstruction

Huihua Kong and Jinxiao Pan

School of science, North University of China, Taiyuan
Shanxi, 030051, China
fengerkong@nuc.edu.cn

Abstract. The convergence speed of algebraic reconstruction technique (ART) depends heavily on the order in which the projections are considered. In this study, a projection access scheme based on prime number increment is proposed, which is applicable to uniform projection sampling in any angle range. We compared the results reconstructed from the proposed method with the results reconstructed from the conventional sequential method, the prime number decomposition method and random ordering method, for cone-beam X-ray computed tomography reconstruction and for the case of circular acquisition. The results indicate that using the proposed method can accelerate the convergence of ART and produces more accurate images with fewer artifacts.

Keywords: CT reconstruction, ART, projection order, prime number increment.

1 Introduction

Cone-beam x-ray computed tomography (CT) is one of the most important non-invasive imaging techniques. In the x-ray CT reconstruction, a volumetric image of object is reconstructed from the projection data. There are two major categories of CT image reconstruction: analytic and iterative methods. Iterative methods such as algebraic reconstruction technique (ART) [1] and expectation maximization (EM)[2-3] are superior to analytic methods such as the FDK [4] and the Katsevich algorithms [5] in handling incomplete and noisy projection data.

A relatively high demand for computational time is the main drawback to use iterative methods. Several approaches have been developed to accelerate the computation of iterative methods. One method is using the optimized ordering schemes for ART. Several methods have been proposed and evaluated so far [6–9].

The aim of this study is to design a "good" permutation of the ordering of the projection view's index using the simple method. This paper is organized as follows. A brief review of ART and the proposed projection access system based on prime

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 179–185.
springerlink.com

number increment are given in Section 2. Section 3 describes the experiments, quantitative evaluation metrics, and the results. The conclusion is given in section 4.

2 Methods

2.1 Algebraic Reconstruction Techniques

The image reconstruction problem in CT can be modeled by the following equation:

$$\boldsymbol{AX} = \boldsymbol{Y} \tag{1}$$

where $\boldsymbol{Y} = (y_1, y_2, \cdots, y_I)^{\mathrm{T}}$ is the projection data and I is the total number of projection rays, $\boldsymbol{X} = (x_1, x_2, \cdots x_J)^T$ is an unknown image and J is the total number of the voxel in the image, and $\boldsymbol{A} = (a_{ij})_{I \times J}$ is the projection matrix and a_{ij} is the length of projection ray *i* through voxel *j*. The problem is to reconstruct the $\boldsymbol{X}$ from the $\boldsymbol{Y}$.

The ART algorithm provides an efficient iterative way to solve the problem. It can be written as:

$$x_j^{(k+1)} = x_j^{(k)} + \lambda_k \cdot \frac{y_i - \sum_{j=1}^{J} a_{ij} x_j^{(k)}}{\sum_{j=1}^{J} a_{ij}^2} \cdot a_{ij} . \tag{2}$$

where $i = k \bmod(I) + 1$ and λ_k is the relaxation parameter. This method was originally discovered by Kaczmarz in [10].

2.2 A projection Access System Based on Prime Number Increment

Denoting M as the total number of projections and the projections $P_i, 0 \le i \le M-1$, are assumed to be equally spaced by an angle $\varphi = \frac{2\pi}{M}$ in the interval [0,2π). In the following, we will give the permutation τ of the ordering of the M projections $\{1, 2, \cdots, M\}$ using prime number increment:

$$\begin{aligned} &\tau(i) = (\tau(i-1) + P) \bmod(M) \qquad 1 \le i \le M-1 \\ &\tau(0) = 0 \end{aligned} . \tag{3}$$

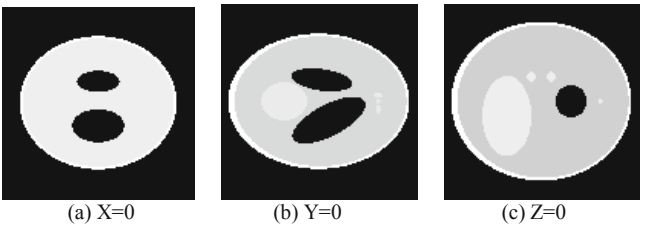
where $P(< M)$ is the prime number and the P is not divided exactly by the M.

Note also that in the special case when M is a prime number, $P(< M)$ can be any positive integer. The sampling pattern is non-periodic because M is not divided exactly by the P. This implies that all measured projections index is processed during the iteration.

3 Evaluations

3.1 Phantom and Simulation Setup

The 3D 128×128×128 voxels Shepp-Logan model is selected as the test model. The three central slices are shown in Fig. 1.



(a) X=0 (b) Y=0 (c) Z=0

Fig. 1. The central slices of 3D Shepp-Logan model.

We use a circular orbit to acquire projection views $[P_0, P_1, \cdots, P_{M-1}]$. System dimensions were as follows: source-to-detector distance equal to 1,000 pixels, source-to-centre-of-rotation distance equal to 200 pixels. The detector consisted of 160 × 160 detection channels. To eliminate the effects of another variable in the comparison process, we used a fixed value of $\lambda_k = 0.2$ throughout the reconstruction procedure.

To verify the validity of the proposed method, projection data comprise 360 projection views distributed uniformly in angle [0, 2π], and ART using Prime Number Increment (PNI) is compared with ART using Sequential Access (SAS), Prime Number Decomposition (PND) and Random Access (RAS).

3.2 Assessment of Image Quality

Both qualitative and quantitative evaluation methods are used in this work. Qualitative evaluation involves visual comparison of different reconstruction methods. Quantitative evaluation involves calculation of the normalized mean square error (NRMSE) and the correlation coefficient (CC).

The NRMSE is defined as

$$\mathrm{NRMSE} = \left[\frac{\sum_{i=1}^{J} (x_i - \hat{x}_i)^2}{\sum_{i=1}^{J} (x_i - \bar{x})^2} \right]^{\frac{1}{2}}, \tag{4}$$

where x_i refers to the i th voxel value in the true image , $\hat{x}_i$ represents the value of voxel i in the reconstructed image, $\bar{x}$ represent image average value in the true images.

The CC is defined as

$$\mathrm{CC} = \frac{\sum_{i=1}^{J} (x_i - \bar{x})(\hat{x}_i - \tilde{x})}{\left[\sum_{i=1}^{J} (x_i - \bar{x})^2 \sum_{i=1}^{J} (\hat{x}_i - \tilde{x})^2 \right]^{\frac{1}{2}}}. \tag{5}$$

where $\tilde{x}$ represent image average value in the reconstructed images. The CC measures the extent to which two images are similar to each other and it takes the highest value of unity if the two are exactly the same.

3.3 Results

Fig. 2 gives the plot of NRMSE as a function of the prime number after 1 iteration when the 360 projection views are distributed uniformly in angle [0, 2π). In fig. 2, the upper and the nether beeline denote the NRMSE reconstructed from PND (NRMSE=0.1996) and RAS (NRMSE=0.1980) after 1 iteration. From fig. 2, we can see the NRMSE reconstructed from PNI after 1 iteration is less than the NRMSE reconstructed from RAS, when P=11, 13, 17, 29, 31, 37, 41, 53, 71, 73, 101, 109, 113, 139, 157, 167, 199, 223, 227, 239, 241, 251, 307, 331, 347, 349.

In the following, we compared the PNI (P=157) with SAS, PND and RAS. Table1 and table 2 list the NRMSE and the CC reconstructed from PNI (P=157), PND and RAS after 5 iterations. From table 1 and table 2, we can see the optimized projection ordering systems (PND, RAS and PNI) are better than conventional sequential method (SAS) and the PNI is slightly better than PND and RAS in convergence speed.

Reconstruction results for three central axial slices after three updates are shown in Fig. 3. The graph clearly shows differences in intensity between the images generated by SAS and the optimized projection ordering systems. However, the differences in intensity between the images generated by PND, RAS and IWDS are almost invisible.

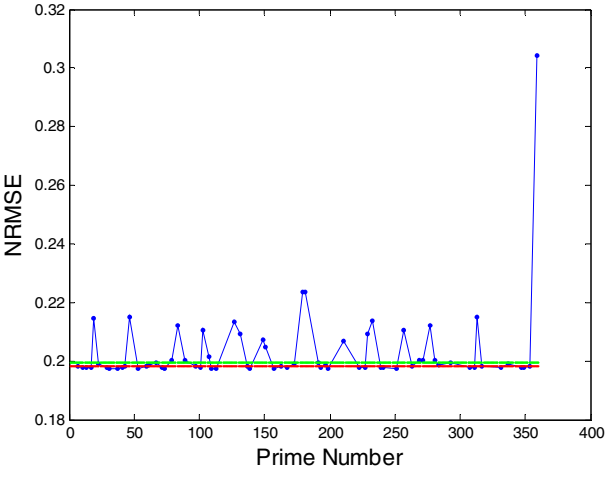


Fig. 2. Plots of the NRMSE versus the prime number after 1 iteration.

Table 1. The NRMSE reconstructed from projection access systems(SAS,PND,RAS,PNI) after 5 iteration

Iterations	**SAS**	**PND**	**RAS**	**PNI (P=157)**
1	0.304499	0.199631	0.198078	0.197382
2	0.164394	0.120987	0.120266	0.119865
3	0.109118	0.086962	0.086848	0.086393
4	0.081984	0.068216	0.068468	0.067936
5	0.066087	0.056454	0.056923	0.056333

Table 2. The CC reconstructed from projection access systems(SAS,PND,RAS,PNI) after 5 iteration

Iterations	**SAS**	**PND**	**RAS**	**PNI (P=157)**
1	0.948793	0.975845	0.976484	0.976613
2	0.980074	0.988581	0.98889	0.988959
3	0.988645	0.992707	0.992862	0.993014
4	0.992525	0.994706	0.994688	0.994819
5	0.994619	0.995840	0.995788	0.995933

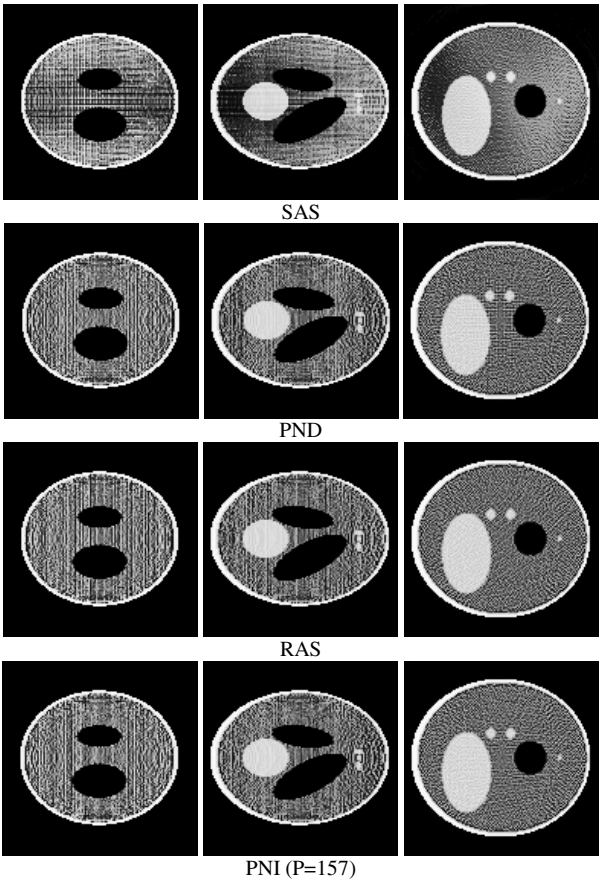


Fig. 3. Reconstruction results at the central location of three different axes after three iterations using different projection access systems.

4 Conclusion

In this paper, we introduced a new and simple projection access system PNI to accelerate the ART algorithm and improve the quality of the reconstructed image. This method is still adapted to the case that the projection views are distributed in limited angle. We analyzed the validity of the proposed method for cone-beam circular orbit CT. The PNI method is superior to SAS and slightly better than PND and RAS in convergence speed and reconstructed image quality. Results have been obtained for ART only, but it is anticipated that other iterative reconstruction algorithms like OS-type methods [11-15] will behave similarly.

Mueller et al. also evaluated a method with constant angular increment in [9]. However, their choices of 66.0^{o}, 69.75^{o}, 73.8^{o} may be result in repeat projection sampling and differ from the method of this paper.

Acknowledgments. This study was partly supported by National Natural Science Foundation of China (60772102, 60876077, 61071193) and Shanxi Natural Science Foundation (2010011002-1).

References

1. Gordon, R., Bender, R., Herman, G.T.: Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography. J. Theo. Biol. 29, 471–482 (1970)
2. Shepp, L.A., Vardi, Y.: Maximum likelihood reconstruction in positron emission tomography. IEEE Trans. Med. Imag. 1, 113–122 (1982)
3. Lange, K., Carson, R.: EM reconstruction algorithms for emission and transmission tomography. J. Comput. Assist. Tomogr. 8, 306–316 (1984)
4. Feldkamp, L.A., Davis, L.C., Kress, J.W.: Practical cone-beam algorithm. J. Opt. Soc. Am. A 1, 612–619 (1984)
5. Katsevich, A.: Theoretically exact filtered backprojection-type inversion algorithm for Spiral CT. SIAM Journal on Applied Mathematic 62, 2012–2026 (2002)
6. van Dijke, M.C.: Iterative methods in image reconstruction, Ph.D. Dissertation. Rijksuniversiteit Utrecht, The Netherlands (1992)
7. Herman, G.T., Meyer, L.B.: Algebraic reconstruction technique can be made computationally efficient. IEEE Trans. Med. Imag. 12(3), 600–609 (1993)
8. Guan, H., Gordon, R.: A projection access order for speedy convergence of ART (algebraic reconstruction technique): a multilevel scheme for computed tomography. Phys. Med. Biol. 39, 2005–2022 (1994)
9. Mueller, K., Yagel, R., Cornhill, F.: The weighted distance scheme: A Globally Optimizing Projection Ordering Method for ART. IEEE Trans. Med. Imag. 16(2), 223–230 (1997)
10. Kaczmarz, S.: Angenäherte Auflösung von Systemen linearer Gleichungen. Bull. Acad. Polon. Sci. Lett. A., 335–357 (1937)
11. Hudson, H.M., Larkin, R.S.: Accelerated image reconstruction using ordered subsets of projection data. IEEE Trans. Med. Imag. 13(4), 601–609 (1994)
12. Kamphuis, C., Beekman, F.J.: Accelerated iterative transmission CT reconstruction using an ordered subsets convex algorithm. IEEE Trans. Med. Imag. 17(6), 1101–1105 (1998)
13. Erdoğan, H., Fessler, J.A.: Ordered subsets algorithms for transmission tomography. Phys. Med. Biol. 44, 2835–2851 (1999)
14. Beekman, F.J., Kamphuis, C.: Ordered subset reconstruction for x-ray CT. Phys. Med. Biol. 46, 1835–1844 (2001)
15. Kole, J.S., Beekman, F.J.: Evaluation of the ordered subset convex algorithm for cone-beam CT. Phys. Med. Biol. 50, 613–623 (2005)

The Simulation of the Noise Eliminated Based on Artificial Neural Network

Huang Yu Shui and Liu Chong

College of Information Engineering, Nanchang University, Nanchang, Jiangxi, China 330031
huangyushui@ncu.edu.cn, Liuchong_12@163.com

Abstract. In order to solve the communication system among the noise interference problem, we use a new method, the neural network and adaptive fuzzy system to model the system. With a certain signal source and a random noise to the training, the simulation effectively deal with the noise reaching the purpose of no noise.

Keywords: artificial neural network, MATLAB, model.

1 Introduction

As the development of the artificial neural network, it has got much better applied effect in many fields. The model which is made in nonlinear system is the more important applied direction in discern and control. The artificial neural network provides a strong tool for the nonlinear system which bases on its nonlinear approach ability and its study ability. At the same time the fuzzy inference can get the good fuzzy model which bases the given data adjusting the parameter[1]. This paper bases on the knowledge express ability and the neural network study ability in the fuzzy system and builds the ANFIS(Adaptive neuron-fuzzy system)[2], and to build a model of the nonlinear eliminates the noise system and to get the purpose of eliminating the noise.

2 Neural Network PID Controller

Neural network with learning ability and approach any nonlinear mapping ability, and thus resolve the uncertainty in the control of complex systems with very large applications. Domestic and foreign scholars in recent years, neural network and traditional technology, applied to nonlinear system control aspects of a number of useful attempt, achieved some encouraging results. How to proceed from the classical PID control theory with the PID weights in the form proposed by the network structure to form complex neural network PID controller, both of neural networks in order to avoid possible local minimum, the establishment of a hybrid neural direct adaptive control structure and the corresponding learning algorithms, and achieved a good control effect.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 187–191.
springerlink.com

2.1 Robust Adaptive Neural Network Controller

Unknown system for the model proposes a complex control structure - the parallel self-learning neural network Robust adaptive control structure, it can use the learning ability of neural networks and nonlinear mapping capabilities to address the traditional online model of adaptive control online identification and controller design problem for uncertain nonlinear systems in order to achieve high precision output tracking control; run the monitor through the introduction, neural network control method is usually overcome the existing problem of poor real time; use of a robust feedback controller to ensure the neural network model for studying the stability of the initial closed-loop system. Two methods are proposed by a conventional feedback controller to guarantee system stability. Therefore, when the neural network controller is designed with larger degree of freedom, it is better than the pure control of intelligent control performance.

2.2 Study of Neural Network Optimal Controller

Optimal control of linear and nonlinear neural network learning methods combined, a new complex nonlinear optimal controller. This control method with artificial neural networks in parallel, adaptive, self-learning ability to apply the present optimal control, control system of compensation as part of completion of more accurate modeling and stability control, the control system has more advanced intelligence, is a very effective combination of methods.

3 The System Model

The noise eliminated may be look as a reflection from the noise space to the no noise space. Generally, the reflection is a complicated nonlinear function, the fuzzy system may be realize the complicated nonlinear reflection in this neural network, and makes it adapt the surrounding change and make it have the most robust. It can be used to the complicated analogy nonlinear system. It may be make up a completed noise eliminated system. The noise eliminated is to get rid of the noise from the advanced tackle and only restore the goal signal, using the neural network fuzzy system to complete the network train model[3]. After the network stabilized, it may be complete the noise eliminated. The number M which the network may be remembered is determined by the neural element. The model of system is in Fig 1.

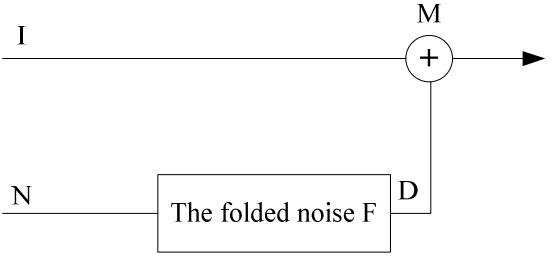


Fig. 1. The folded noise measure process

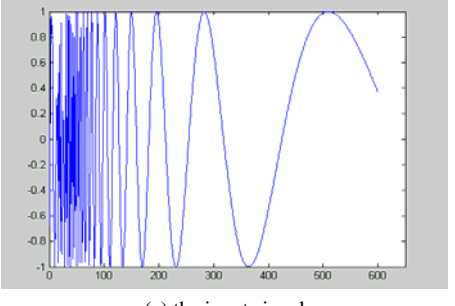
In Fig1, I is the signal which is going to be measured, N is the noise source, D is the noise signal which is folded the measure signal M. D and N satisfied the nonlinear reflection f. As follows:

$$D(k) = f(N(k), N(k-1), \ldots) \tag{1}$$

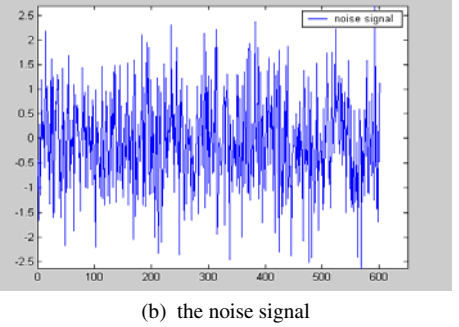
$$M(k) = I(k) + D(k) = I(k) + f(N(k), N(k-1), \ldots) \tag{2}$$

4 System Simulation

For example: the input and noise signal are X and n1 respectively. Their curves are in Fig. 2 (a) and (b).



(a) the input signal



(b) the noise signal

Fig. 2. The input and noise signal

If the folded measure noise signal n2(k) is as the Fig 3

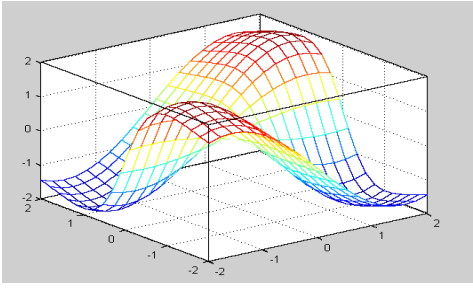


Fig. 3. The folded noise

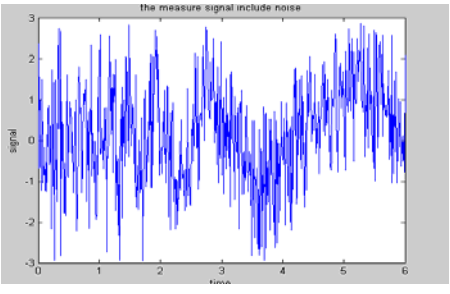


Fig. 4. The measured folded signal

The measure signal M of the folded noise n2(k) satisfies the following relation:

$$M(k) = X(k) + n2(k) \tag{3}$$

The measure signal change curve as the following Fig 4.

Choosing the two input and output neural vague system to simulate the nonlinear character and gains the input –output data (n1(k),n1(k-1), m(k)), using the ANFIS function to train the neural network fuzzy system. In training, every input variable subordinates the function number chooses 2, and the total fuzzy rule number chooses 4, As the following, use the MATLAB sentence to carry out the training process.

To eliminate the noise of the measure signal, if m1 is the measure signal which the system has eliminated the noise: the measure signal curve as Fig 5.

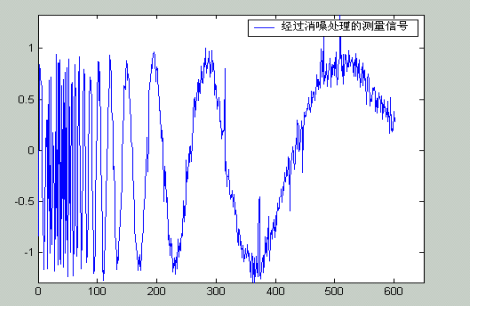


Fig. 5. Measurement of the signal after noise cancellation M

By comparing the input to output, the signal source passes the ANFIS model, the output signal is the same as the input I, and it effectively eliminates the noise and gains the purpose of the noise eliminated.

5 The Conclusion

Given above in the context of non-Gaussian noise neural network adaptive fuzzy systems nonlinear processing, for the noisy source, by the neural network model of fuzzy system consisting of system and noise elimination in real time performance can meet the requirement. The nonlinear adaptive fuzzy neural network intelligent control strategy and the traditional control structure combining intelligent composite controller, whichever is in control of the advantages and features, has become a hot topic in the field of control, but also part of the solution types control of complex systems an important tool in both theory and practical application is very important.

References

1. Long, F., Yanda, L.: Neural network signal tackle. Electronic industry publishing press (October 1993)
2. Xiaoli, W., Zhehui, L.: MATLAB supplementary fuzzy system design. Xi'an electronic technology university publishing press (August 2002)
3. Shuang, C.: Neural network, fuzzy system and applied in control. Chinese technology university publishing press (May 2001)
4. Abdelnour, G.M.: Design of a fuzzy controller using input and output mapping factors. IEEE Trans. on Systems, Man and Cybemetics 27(5), 884–889 (1997)
5. Ying, H.: The Takagi-sugeno fuzzy controllers using the simplified linear control rules are nonlinear variable gain controllers. Automatica 34(2), 157–167 (1998)
6. Zhang, E., Shi, S., Weng, Z.: Comparative study of fuzzy control and PID control methods. Journal of Shanghai Jiaotong University 33(4), 501–503 (1999) (in Chinese)

Simulation Research on μ Synthesis Robust Control for Driving of Hybrid-Power Electric Vehicle*

Jianbo Cao[1], E. Shiju[1], Tianfeng Zhao[2], Xilin Zhu[1], and Hongkui Jiang[1]

[1] Department of Mechanical Engineering, Zhejiang Normal University, 321004 Jinhua, China
[2] Information Center, Zhejiang Normal University, 321004 Jinhua, China
{caojianbo2008,jhkass}@163.com, {esx_2001,ztf,zhuxl}@zjnu.cn

Abstract. To deal with two major problems of electric vehicle (EV): the short driving range and the short life of batteries, a hybrid power system was designed and applied to the EV. It was composed of an ultracapacitor with high-specific power and long life, four lead-acid batteries with high-specific energy, and a bi-directional DC/DC converter. To improve the stability and reliability of the system, based on researching driving process and μ synthesis robust control, the driving mathematical model of the system was established, and the driving μ synthesis robust controller for the system was designed. The simulation experimental results show that the hybrid power could not only increase the starting current, but also reduce the batteries' discharge current and lengthen the life of batteries. Additionally, the μ synthesis robust controller is superior to PID controller at response speed, steady-state tracking error and resisting disturbance.

Keywords: Electric vehicle, Ultracapacitor, Hybrid power, Driving, Robust control.

1 Introduction

Due to the dual pressure from environmental pollution and energy crisis, it has become a general trend of the development of electric vehicles (EVs). EVs have improved their performance and made suitable for commercial and domestic use during the last decades. Nevertheless, they still have not achieved ranges as good as gas-powered conventional vehicles. This problem, due to the low specific energy contained in most electric batteries compared to that of gasoline, restricts EVs fast developing. It is very significant for the increase in range that the technology of energy-regenerative baking is applied to EVs [1]. However, batteries have a poor ability to recover energy from a regenerative braking, and a scarce power capacity for a fast acceleration. For this reason, EVs may use an auxiliary energy system able to receive regeneration fast and give power during peak periods. The ultracapacitor with high-specific power looks as the most appropriate choice [2]. In this paper, a hybrid power system, in which batteries are in use as "main source" and ultracapacitor as "auxiliary

* The work was supported by the Foundation of Education Department of Zhejiang Province, China (Y200909210), and the Doctor Fundation (ZC304010013).

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 193–200.
springerlink.com

source", was designed and applied to the EV. It is essential the development of an energy-management system to control the power flow between both sources. In order to improve the energy-regenerative efficiency, the optimal-control strategy is very important and should be researched. μ synthesis robust controller has strong robustness. In this paper, based on establishing the mathematical model of the driving process, a μ synthesis robust controller for the system was developed and tested successfully. The hybrid power system controlled by μ synthesis robust controller can recover more energy and lengthen batteries' life.

2 Driving Process and Its Mathematical Model of Hybrid-Power EV

In this paper, the EV employed a brushless DC motor (BLDCM), which has been widely applied in the field of EVs, due to a series of advantages: simple structure, reliable performance, high efficiency, large starting torque, etc [3]. The main circuit topology of the EV control system designed in this paper is shown in Fig. 1.

During the start-up process, the batteries and the ultracapacitor drive the motor together. When EV runs at normal speed, batteries drive the motor alone, and meanwhile recharge some energy to the ultracapacitor to prepare for accelerating or climbing. The principle of ultracapacitor driving motor is the same as batteries', so we take batteries as an example to analyze the driving process and establish the mathematical model [4-7].

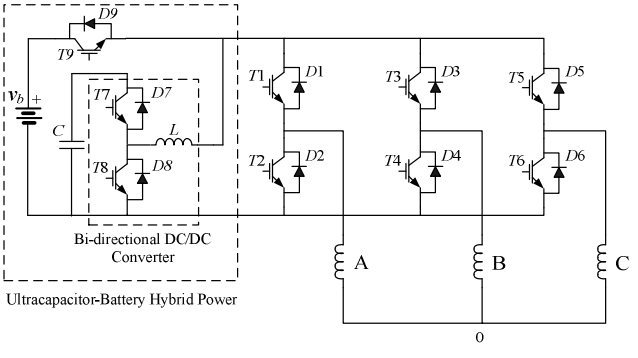


Fig. 1. Hybrid-power EV's main circuit topology.

In this paper, the three-phase BLDCM works according to six states, that is, T1 to T6 of the inverter work in turn according to the six states of Hall sequence, and in any state, two-phase work principle is the same. The BLDCM could be seen as the DC motor under the condition of two phases conducting. In this, we take phases A and B as an example to establish the driving equivalent circuit. During the driving process, the batteries drive the BLDCM by buck converter, at this time, T1 is PWM, T4, T9 are on, and other MOSFETs are off, the current flow direction and equivalent circuit topology as buck converter are shown in Fig. 2.

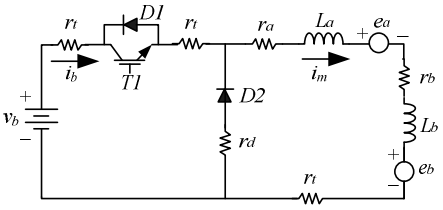


Fig. 2. Equivalent circuit when batteries alone.

The driving mathematical model can be established as follows (take Fig. 2 as an example):

While T1 is in conducting state and cut-off state, the state equations [8] are respectively described as:

T1_on ($0 \le t \le d \cdot T$)

$$(L_a + L_b)\frac{di_m}{dt} = v_b - e_a - e_b - i_m \cdot (3r_t + r_a + r_b + r_{bat}) . \tag{1}$$

T1_off ($d \cdot T < t \le T$)

$$(L_a + L_b)\frac{di_m}{dt} = -e_a - e_b - i_m \cdot (r_a + r_b + r_d + r_t) , \tag{2}$$

where d is the duty cycle of T1_PWM, T is the operation period, r_{bat} is the ESR of the batteries.

Because the motor's three phases are similar, in permitted scope, we have: $L_a = L_b = L_m$, $e_a = e_b = e_m$, $r_a = r_b = r_m$, then (1) and (2) can be simplified as:

$$2L_m \frac{di_m}{dt} = v_b - 2e_m - i_m \cdot (3r_t + 2r_m + r_{bat}) , \tag{3}$$

$$2L_m \frac{di_m}{dt} = -2e_m - i_m \cdot (r_d + 2r_m + r_t) . \tag{4}$$

According to the deflecting couple equilibrium, we have:

$$J\frac{d\omega}{dt} = K_t \cdot i_m - T_L , \tag{5}$$

$$e_m = K_e \omega , \tag{6}$$

where J is the moment of inertia, ω is the rotational speed of motor, K_t is the torque coefficient, T_L is the load torque, and K_e is the BEMF coefficient of armature winding.

Suppose state variables are $\boldsymbol{x}=\begin{bmatrix} i_m & \omega \end{bmatrix}^T$, output is $y=i_m$, after average processing, then:

$$\begin{cases} \dot{\boldsymbol{x}}=\begin{bmatrix} -\dfrac{2r_m+r_t+(2r_t+r_{bat})\cdot d+r_d\cdot(1-d)}{2L_m} & -\dfrac{K_e}{L_m} \\ \dfrac{K_t}{J} & 0 \end{bmatrix}\boldsymbol{x}+\begin{bmatrix} \dfrac{v_b}{2L_m}d \\ -\dfrac{T_L}{J} \end{bmatrix} \\ y=\begin{bmatrix} 1 & 0 \end{bmatrix}\boldsymbol{x} \end{cases}. \tag{7}$$

After handling by perturbation method, and separation of steady-state variables and instant variables, we can get the linear small-signal mathematical model of driving:

$$\begin{cases} \dot{\boldsymbol{x}}=\begin{bmatrix} -\dfrac{2r_m+r_t+(2r_t+r_{bat})\cdot D+r_d\cdot(1-D)}{2L_m} & -\dfrac{K_e}{L_m} \\ \dfrac{K_t}{J} & 0 \end{bmatrix}\boldsymbol{x}+\begin{bmatrix} \dfrac{r_d-2r_t-r_{bat}}{2L_m} & 0 \\ 0 & 0 \end{bmatrix}\cdot\boldsymbol{X}\cdot d+\begin{bmatrix} \dfrac{v_b}{2L_m} \\ 0 \end{bmatrix}\cdot d \\ y=\begin{bmatrix} 1 & 0 \end{bmatrix}\boldsymbol{x} \end{cases}, \tag{8}$$

where D and $\boldsymbol{X}$ are respectively the circuit steady values of d and $\boldsymbol{x}$.

When the batteries and the ultracapacitor drive the motor together, the current flow direction and equivalent circuit topology are shown in Fig. 3.

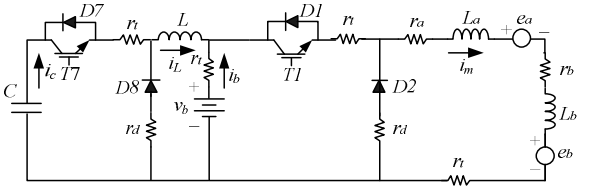


Fig. 3. Equivalent circuit when batteries and ultracapacitor together.

3 Driving μ Synthesis Robust Controller Design for Hybrid-Power EV

3.1 Equivalent Transformation of Control Problem

The internal link structure of EV driving control system was shown in Fig. 4. The figure reflects the connection relations of nominal model G_0, feedback structure K_d, uncertainty model, and performance index. In the modelling process, all uncertainty belong to the normalization transfer matrix $\boldsymbol{\Delta}$, which describes the difference between nominal model G_0 and the actual model G. Moreover, $\boldsymbol{\Delta}$ is stable, and its norm boundary condition is $\|\boldsymbol{\Delta}\|_\infty < 1$.

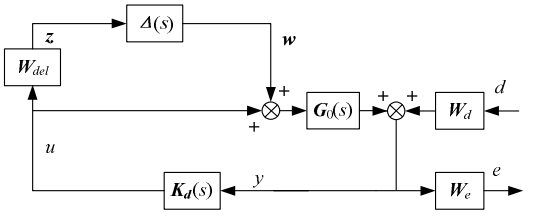


Fig. 4. The EV driving control system with the uncertainty (z - weighted control, p – perturbation, u - control variable, y - measurement input, d – disturbance, e - performance output).

μ control of the system can be described as follows: for all the stable perturbation $\boldsymbol{\Delta}$, and $\|\boldsymbol{\Delta}\|_\infty < 1$, finding a stable controller Kd that enabled the closed-loop system still to maintain stable in the perturbation situation. the weighting sensitivity transfer matrix is $\boldsymbol{S}(\boldsymbol{\Delta}) = \boldsymbol{W}_e(\boldsymbol{I} + \boldsymbol{G}_0(\boldsymbol{I} + \boldsymbol{\Delta}\boldsymbol{W}_{del})\boldsymbol{K}_d)^{-1}\boldsymbol{W}_d$ and has $\|\boldsymbol{S}(\boldsymbol{\Delta})\|_\infty < 1$. The control problem in Fig. 4 can be equivalently transformed into a typical μ design problem, and the transformation process is shown in Fig. 5.

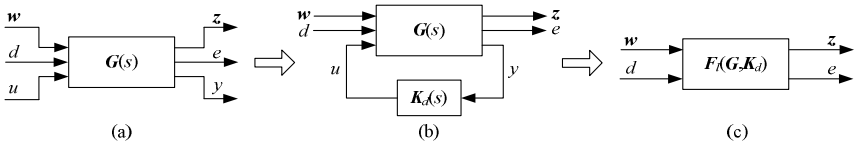


Fig. 5. Equivalent transformation of control problem. (a) Equivalent connection diagram of open-loop control system, (b) Equivalent connection diagram of closed-loop control system after adding K_d, (c) Closed-loop linear fractional transformation.

3.2 The Selection of Weighting Function

The choice of weighting function is the key to design the system controller, and directly determines the performance of the system. Through the rational choice of weighting function, it can be guaranteed that the system has strong robust stability, good capacity of command signal tracking, anti-interference and noise suppression. Some literature works gave a few weighting function selection methods for some specific system [9][10], but due to the different characteristics of each system, it is difficult to give a general design method of weighting function. The main difficulty of weighting function selection lies in how the system closed-loop performance index was expressed in the weight function with as far as possible simple structure. The weighting function mainly includes performance index weighting function, model uncertainty weighting function, and input weighting function.

According to the actual situation of the EV in the paper, after several adjustments, the frequency weighting functions of EV driving controller were taken as follows:

Performance index weighting function is

$$W_e(s) = \frac{0.3s + 100}{s + 0.005} \cdot \tag{9}$$

Model uncertainty weighting function is

$$W_{del}(s) = \frac{3(s+10)}{s+150}. \tag{10}$$

Input weighting function is

$$W_d(s) = \frac{s+10}{150}. \tag{11}$$

3.3 μ Synthesis Robust Controller Solving Process

According to the driving mathematical model described above, and the weighting functions from (9) to (11), μ synthesis robust controller can be optimally solved by the D-K algorithm in μ-Tools of Matlab Robust Control Toolbox [11]. As a result, the structure singular value μ=0.768<1, and closed-loop system achieved robust performance requirements. After handling by truncated equilibrium model, the μ synthesis robust controller reduced to 4 bands as follows:

$$K_d(s) = \frac{0.0312s^3 + 5.2186s^2 + 117.6569s + 19.3253}{s^4 + 256.68s^3 + 202.5034s^2 + 35.215s + 0.67}. \tag{12}$$

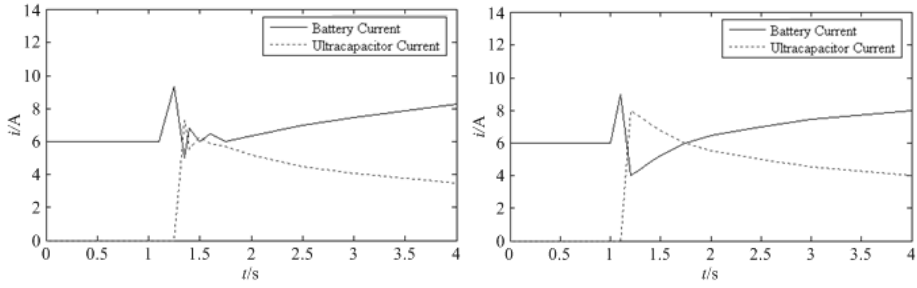
In the practical application, (12) must be discretized to realize digital programming. In this paper, the control period was T_s=2 ms. The discrete controller is gotten as follows by the double linear transformation method:

$$K_d(z) = \frac{(2.907z^4 - 4.946z^3 - 0.8304z^2 + 4.946z - 2.076)\cdot 10^{-5}}{z^4 - 3.591z^3 + 4.773z^2 - 2.774z + 0.5916}. \tag{13}$$

4 Simulation Results and Analyses

The controller of hybrid-power EV designed in this paper was tested by MATLAB/Simulink. The hybrid power system uses a 3.125 farad ultracapacitor and four batteries with a nominal voltage of 48 volts. The driving simulation comparisons were respectively done with the PID controller and the μ synthesis robust controller. The simulation results are shown as Fig. 6 and Fig. 7.

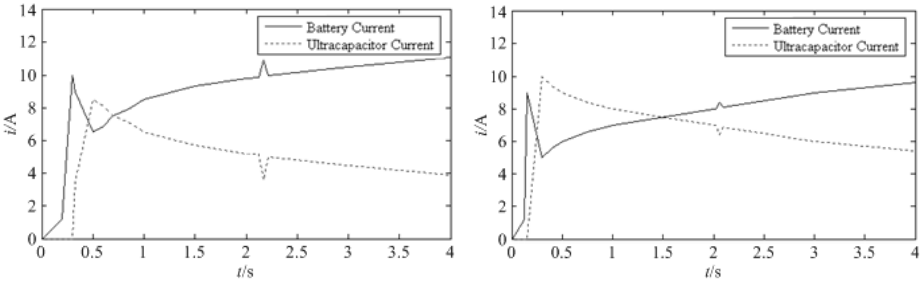
Fig. 6 shows the response of battery current and ultracapacitor current when the load of EV was suddenly increased. When the driving current of EV was not big, the ultracapacitor stopped discharging. If the load was suddenly increased, the driving current of EV would be also increased. When the driving current exceeded 9 amperes, the ultracapacitor began to discharge, and worked together with the batteries. As shown in Fig. 6, the ultracapacitor could reduce the batteries's discharging current. Additionally, the μ synthesis robust controller is better than the PID controller at response speed and steady-state tracking error.



(a) PID controller (b) μ synthesis robust controller

Fig. 6. Current response when the load was suddenly increased.

In order to compare the robustness of two controllers, the disturbance signal with 2-ampere amplitude and 50-ms pulse width was imposed on the system at 2.5 seconds, and the simulation results are shown in Fig. 7. As we see, the μ synthesis robust controller is superior to PID controller at response speed and resisting disturbance.



(a) PID controller (b) μ synthesis robust controller

Fig. 7. Resisting-disturbance comparison.

5 Conclusions

Based on constructing the control system of the hybrid-power EV, a μ synthesis robust controller was designed and applied to the EV. The following conclusions can be gotten from the above simulations:

The ultracapacitor-battery hybrid power system with the μ synthesis robust controller can enhance the instantaneous performance of EV when driving, avoid batteries being charged and discharging by big electric current, reduce the batteries' charge times, and lengthen the batteries' life. In the driving processes, comparing with traditional PID controller, the μ synthesis robust controller has better robustness, and can improve the stability and reliability of the system.

References

1. Cao, B.G., Zhang, C.W., Bai, Z.F., Li, J.C.: Technology Progress and Trends of Electric Vehicles. Journal of Xi'an Jiaotong University 38, 1–5 (2004)
2. Sun, F.Q., Zou, J.Z., Yao, X.D., Wang, X.Y.: An Ultracapacitor Controller Based on DSP for Electric Vehicle. Mechatronics, 58–61 (2005)
3. Liao, L.Y., Wang, Z.F., Deng, Y.D., Hu, X.L.: Simulation of a Permanent Magnet Brushless DC Motor for Electric Vehicle Using the Bond Graph Method. Journal of WUT 25, 192–195 (2003)
4. Jorge, M., Micah, E.O., Juan, W.D.: Energy-Management System for a Hybrid Electric Vehicle, Using Ultracapacitors and Neural Networks. IEEE Transactions on Industrial Electronics 53, 614–623 (2006)
5. Juan, W.D., Micah, E.O.: Ultracapacitors + DC-DC Converters in Regenerative Braking System. IEEE AESS Systems Magazine, 16–21 (2002)
6. Wang, X.M., Wang, L.: Motor Control Based on DSP. Beihang University Press, Beijing (2004)
7. Bai, Z.F.: Research and Development on Ultracapacitor-Battery Composite Power System and Its H_∞ Robust Control of Electric Vehicle. Xi'an (2006)
8. Zhang, C.W., Bai, Z.F., Cao, B.G., Li, J.C.: Study on Regenerative Braking of Electric Vehicle. In: 4th International Power Electronics and Motion Control Conference, pp. 836–839 (2004)
9. Pascoli, B.C., Soares, N.M.F., Antonio, A.F.S.: Robust Speed Control of an Induction Motor: an H_∞ Control Theory Approach with Field Orientation and μ-Analysis. IEEE Trans. on Power Electronics 15, 908–915 (2000)
10. The MathWorks Inc.: Robust Control Toolbox User's Manual. Version 2 (2005)
11. The MathWorks Inc.: μ-Analysis and Synthesis Toolbox User's Manual (2005)

Influencing Factors on Power Generation Mode of Electroactive Polymer*

Jianbo Cao[1], E. Shiju[1], Tianfeng Zhao[2], Xilin Zhu[1], Chunfu Gao[1], and Anfeng Hui[1]

[1] Department of Mechanical Engineering, Zhejiang Normal University, 321004 Jinhua, China
[2] Information Center, Zhejiang Normal University, 321004 Jinhua, China
{caojianbo2008,huianfeng.123}@163.com,
{esx_2001,ztf,zhuxl,cfgao2007}@zjnu.cn

Abstract. In order to promote the application of electroactive polymer, the generator mode theory fundamentals of electroactive polymer (take acrylic elastomer as an example) were discussed and studied. The power generation mode of electroactive acrylic elastomer (EAE) was studied by experiments based on designing a power generation measurement system for EAE. The power generation process of EAE was tested under different influencing factors. The experimental results show that power generated when EAE relaxed. The bias voltage and pre-strain level had a great impact on the power generation of EAE. The voltage of power generation increased with the augmentation of the bias voltage and pre-strain level.

Keywords: Electroactive polymer, Power Generation mode, Influencing factors, Bias voltage.

1 Introduction

Electroactive polymer (EAP) is a new type of polymer which responds to external electrical stimulation. Electroactive acrylic elastomer (EAE) is a particular type of EAPs and it has best demonstrated exceptional performance. This type of field-activated EAPs has shown tremendous promise for applications as actuators. Their advantages in converting mechanical to electrical energy in a generator mode are less well known [1]. EAE is formed by infiltering or coating the acrylic elastomer matrix with electrode material in the upper and lower surface based on Maxwell effect [2]. EAE could be viewed as a compliant capacitor, that is, a sandwich structure that the thin passive elastomer is sandwiched between two compliant electrodes. Compared with other types of EAPs, EAE could produce bigger strain, and have better flexibility, lower density, and lower cost, etc [3][4]. This material could be used as the driver material, the bionic muscles, and the related components used to construct micro electromechanical system (MEMS), such as micro-motors, micro-pumps, micro-valves, etc, moreover, it could be used in power generation, especially suitable for distributed power generation [5-7]. Therefore, EAE will have broad application prospects.

* The work was supported by the National Natural Science Foundation of China (50777028).

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 201–207.
springerlink.com

EAE is capable of converting energy in the form of electric charge and voltage into different form of mechanical force and vice versa. In this paper, EAE was realised with a pre-stretched circular film coated with compliant electrodes of graphite powder and conductive adhesive. The mechanism and influencing factors of EAE power generation were studied by theoretical analysis and experiments. The research on the EAE was expanded. The chief purpose of this study is to provide the basis for the design and use of the EAE power generation.

2 Power Conversion Process of EAE

In this paper, we employed EAE as EAP. Acrylic elastomers have demonstrated an estimated 0.4J/g specific energy density, compared to around 0.1J/g for advanced single crystal ceramics and around 0.04J/g for peak electromagnetic. Much higher energy densities, over 1J/g, are predicted. High conversion efficiency is predicted, theoretically up to 80-90%. In addition to superior performance, EAEs have two features that distinguish them from other energy-conversion materials: they are made from low cost materials that can be easily fabricated and they are compliant.

There are many advantages [1] of EAE power, such as, lighter – low density, high performance, multifunctional polymers; cheaper – inexpensive materials, fewer parts, no precision machining; quieter – high energy density and compliance of polymers allows quiet primarily sub-acoustic operation with few moving parts; softer – rubbery materials are impedance matched to large motions (e.g. human motion, engines); versatile – polymers are scale-invariant; systems can be made in variety of form factors (conformal, elongated, etc.).

EAEs could convert electrical energy to mechanical work and vice versa, as shown in Fig. 1.

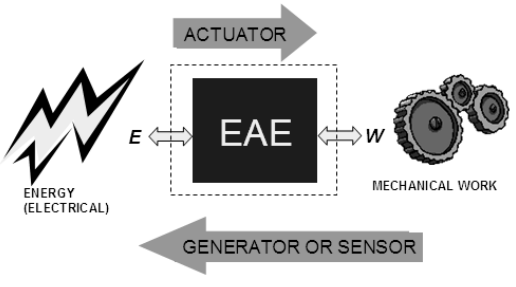


Fig. 1. Power conversion process of EAE.

EAEs are a type of EAP that uses an electric field across a rubbery dielectric with compliant electrodes. The basic functional element is shown in Fig. 2. Incompressible polymer gets thinner and stretch in area when a voltage is turned on. Incompressible polymer gets thicker and contracts in area when a voltage is turned off.

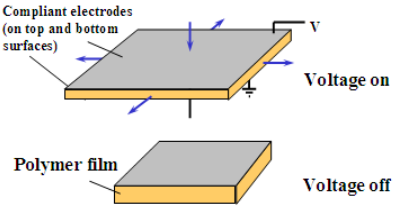


Fig. 2. Basic functional element.

Power generation principle of EAE is shown in Fig. 3. It is like a variable capacitor generator. Energy generated as nearly incompressible elastomer layers increase in area and decrease in thickness when stretched, and power generated when EAE relaxed.

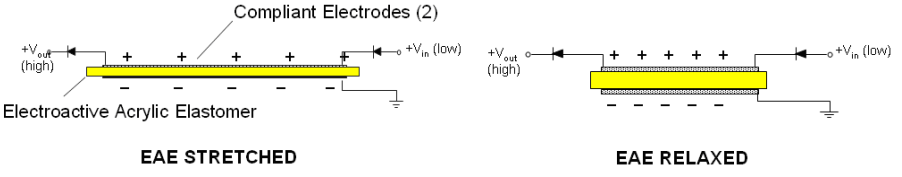


Fig. 3. Power generation principle diagram of EAE.

3 Power Generation Measurement System Design

Fig. 4 shows the power generation measurement system. The bias power supplies the voltage V_{in} for EAE through a diode. Based on the bias voltage, the power generation was achieved by the EAE film contraction. The increased voltage in power generation process can be detected by a high-voltage probe (1000:1) and could be seen through an oscilloscope.

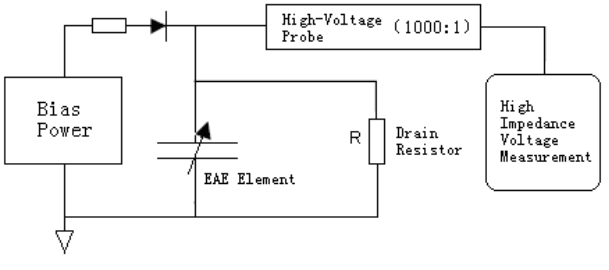


Fig. 4. Measurement circuit.

4 Theoretical Analysis Model

4.1 Strain Energy Function of Super-Elastic Materials

In order to simplify the calculation in the theoretical analysis of acrylic elastomer, it could be assumed to be super-elastic material (Poisson's ratio is 0.5) which could not be compressed, and possess uniformity. The mechanical property of super-elastic material is generally depicted by the general strain energy function W that may be expressed in various forms [8]. According to the analysis of simulation results in [9], Yeoh strain energy form was adopted in this paper. The strain energy form depends on I_1, namely the first constant of the so-called left Cauchy-Green deformation tensor. The equation of the strain energy form is expressed as follows:

$$W = C_{10}(I_1 - 3) + C_{20}(I_1 - 3)^2 + C_{30}(I_1 - 3)^3, \tag{1}$$

where, C_{10}, C_{20}, C_{30} are the material parameters, and respectively $C_{10} = 0.063$ MPa, $C_{20} = -8.88 \times 10^{-4}$ MPa, $C_{30} = 16.7 \times 10^{-6}$ MPa. I_1 was calculated in (2) through the eigenvalue of deformation gradient tensor, namely strain rate λ_i ($i = 1, 2, 3$).

$$I_1 = \lambda_1^2 + \lambda_2^2 + \lambda_3^2. \tag{2}$$

For incompressible materials, Cauchy principal stress (the force per unit area of deformation structure) could be determined by λ_i:

$$t_i = \lambda_i \frac{\partial W}{\partial \lambda_i} - p^t, \tag{3}$$

where, p^t is the super-elastic force, and related with dynamic boundary conditions.

4.2 Power Generation Principal Analyses

According to [1], EAEs could be simplified as variable capacitors. The capacitance can be expressed as

$$C = \varepsilon \varepsilon_0 A / d, \tag{4}$$

where ε is the relative dielectric constant, ε_0 is the permittivity of free space ($8.85 \times 10^{-12} F/m$), A is the total area of the polymer film, and d is the thickness. Both A and d depend on the strain. Because they are elastorners, the volume B of the EAE stays the same during stretching, that is,

$$B = A \cdot d. \tag{5}$$

From (4) and (5), have

$$C = \varepsilon \varepsilon_0 A^2 / B. \tag{6}$$

The voltage V of EAE can be expressed as

$$V = \frac{Q}{C} = \frac{QB}{\varepsilon\varepsilon_0 A^2}, \tag{7}$$

where Q is the charge.

Based on the theory of variable capacitor, the electrical energy generated by EAE can be deduced as follows:

Suppose C_s and C_c are the capacitances of the EAE capacitor in the stretched and contracted states, respectively. In the stretched state, a bias voltage $\boldsymbol{V_{in}}$ is applied to the EAE, and the same voltage $\boldsymbol{V_{in}}$ is on the EAE capacitor in the contracted state after some amount of charge has drained through the drain resistor R. So have the stored electrical energy in the EAE capacitor in the stretched and contracted states respectively [1][10] as follows:

$$E_s = \frac{1}{2} V_{in} C_s^{\,2} = \frac{1}{2} V_{in} Q_s, \tag{8}$$

$$E_c = \frac{1}{2} V_{in} C_c^{\,2} = \frac{1}{2} V_{in} Q_c. \tag{9}$$

Suppose ΔQ is the charge that goes through the drain resistor, Q_s and Q_c are the stretched and contracted charges on the film, then have

$$\Delta Q = Q_s - Q_c = [V(t)/R]dt. \tag{10}$$

Suppose E_R is the energy dissipated through the drain resistor, $E_R = [V(t)^2/R]dt$, and the electrical energy E_g generated by the EAE can be deduced as follows:

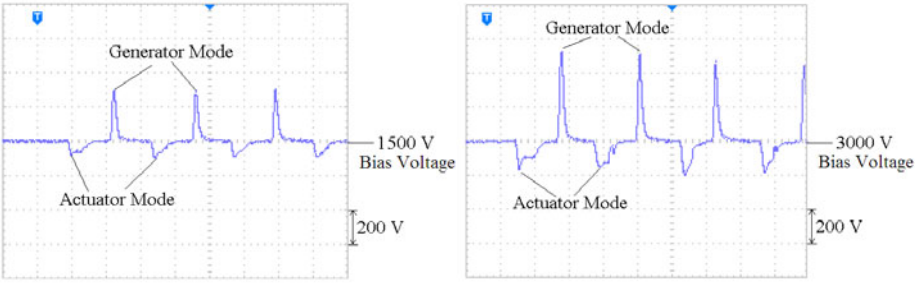
$$E_g = E_s - E_c - E_R = \frac{1}{2} V_{in} \Delta Q - [V(t)^2/R]dt = \frac{1}{2} V_{in} [V(t)/R]dt - [V(t)^2/R]dt \cdot \tag{11}$$

5 Experiments and Analyses

The power generation process of EAE was tested by the experimental equipment and environment which comprises bias power, EAE element, high-voltage probe, and oscilloscope. The experimental results were recorded by the oscilloscope.

The actuator and generator cycle process of EAE under different bias voltage were compared and shown in Fig. 5. In the experiments, the pre-strain was 300%, and the bias voltage was 1500 volts and 3000 volts respectively. When the EAE element was stretched and compressed, it worked in actuator mode and generator mode respectively. As shown in Fig. 5, EAE element can achieve a continuous power generation

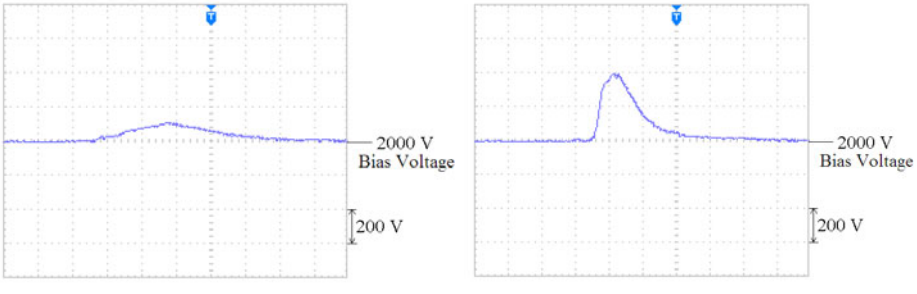
in the stretched and compressed cycle process. Additionally, through contrast, it was found that with the bias voltage increasing, the amplitude of increased voltage and the cycle power generation level also increased.



(a) Bias voltage 1500 V (b) Bias voltage 3000 V

Fig. 5. Actuator and generator cycle process.

The pre-strain influence on the power generation level was tested and shown in Fig. 6. In the experiments, the bias voltage was 2000 volts, and the pre-strain was 100% and 200% respectively. When pre-strain was 100%, the peak voltage reached about 2100 volts, increased about 100 volts. When pre-strain was 200%, the peak voltage reached about 2200 volts, increased about 200 volts. Through contrast, the result showed that the pre-strain level had a great influence on power generation, and with the pre-strain level increasing, the power generation level also increased.



(a) Pre-strain 100% (b) Pre-strain 200%

Fig. 6. Power generation waveform vs. different pre-strain.

6 Conclusions

Based on theoretical analysis and experimental measurement for EAE, the following conclusions could be gotten:

When EAE relaxed, EAE became thicker and contracted in area, as a result, power generation was achieved. The bias voltage and pre-strain level had a great impact on the power generation of EAE. With the bias voltage and pre-strain level increasing, the power generation level also increased. In order to achieve the best result, the values of the two factors could be maximized under the bearing capacity of the system. The experimental results accorded with analytical model and lay the foundation for generator and sensors of EAE.

References

1. Pelrine, R., Kornbluh, R., Eckerle, J., Jeuck, P., Oh, S., Pei, Q., Stanford, S.: Dielectric Elastomers: Generator Mode Fundamentals and Applications. In: Proceedings of SPIE, vol. 4329, pp. 148–156 (2001)
2. Bar-Cohen, Y.: Electroactive Polymer (EAP) Actuators as Artificial Muscles—Reality, Potential and Challlenges. SPIE Press, USA (2001)
3. Perline, R., Kohrnbluh, R., Pei, Q.: High-Speed Electrically Actuated Elastomers with Over 100% Strain. Science 287, 836–839 (2000)
4. Pelrine, R.E., Kornbluh, R.D., Joseph, J.P.: Analysis of the Electrostriction of Polymer Dielectrics with Compliant Electrodes as a Means of Actuation. Sens. Actuators 64, 77–85 (1998)
5. Heydt, R., Kohrnbluh, R., Perline, R.: Design and Performance of an Electrostrictive-Polymer-Film Acoustic Actuator. Jou. of Sound Vib. 215, 297–311 (1998)
6. Goulbourne, N.C., Mockensturm, E.M., Frecker, M.I.: Electro-elastomers: Large Deformation Analysis of Silicone Membranes. Int. Jou. Solids Struct. 44, 2609–2626 (2007)
7. Jung, K., Kim, K.J., Choi, H.R.: A Self-Sensing Dielectric Elastomer Actuation. Sensors and Actuators A: Physical 143, 343–351 (2008)
8. Yeoh, O.H.: Some Forms of the Strain Energy Function for Rubber. Rubber Chem. Technol. 66, 754–771 (1993)
9. Wissler, M., Mazza, E.: Modeling for a Pre-Strained Circular Actuator Made of Dielectric Elastomers. Sensors and Actuators 120, 184–192 (2005)
10. Liu, Y., Ren, K., Hofmann, H.F., Zhang, Q.M.: Electrostrictive Polymer for Mechanical Energy Harvesting. In: Proceedings of SPIE, vol. 5358, pp. 17–28 (2004)

An Adiabatic Register File Using Gate-Length Biasing Technique with Dual-Threshold CMOS

Yangbo Wu, Jiaguo Zhu, and Jianping Hu

Faculty of Information Science and Technology, Ningbo University
315211 Ningbo City, China
nbuhjp@yahoo.cn

Abstract. Gate-length biasing is an active leakage reduction technique that utilizes the short-channel effect by marginally increasing the gate-length of MOS devices to reduce their leakage current with a small delay penalty. DTCMOS (Dual-threshold CMOS) has been proven as an effective way to reduce sub-threshold leakage consumption in active mode. This paper proposes a gate-length biasing technique for clocked adiabatic logic (CAL) circuits using dual-threshold technique to reduce sub-threshold leakage dissipations. An improved CAL register file using DTCMOS gate-length biasing is addressed in this paper. All circuits are verified with HSPICE using a NCSU 45nm technology. The simulations show that the improved CAL register file with gate-length biasing and DTCMOS techniques can attain large energy savings.

Keywords: Nanometer circuits, Gate-length biasing, Dual-threshold technique, Clocked adiabatic logic, Leakage reduction, Register file.

1 Introduction

The power dissipation is a key factor of limiting circuit performances and costs. In previous studies of low-power integrated circuits, the dynamic power consumption is the most concerns, and leakage dissipation is often neglected [1]. However, with the feature size of integrated circuits continues to reduce, the leakage dissipation caused by leakage currents catches up with the dynamic power consumption gradually and the standby power consumption is becoming an important factor in low-power design, which attracts extensive attentions [2].

A number of approaches have been proposed to reduce static leakage power when the system is in standby mode, such as VTCMOS or MTCMOS, input vector control, etc [3, 4]. Very few approaches, such as stacking transistor techniques [5, 6], dual threshold CMOS [7], and P-type CMOS design technology [8], have also been proposed mainly to reduce runtime leakages.

For dual-threshold logic circuit, a higher threshold voltage can be assigned to some transistors on non-critical paths so as to reduce leakage current, while the performance is maintained due to the low threshold transistors in the critical paths. Therefore, no additional transistors are required, and both high performance and low power can be achieved simultaneously.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 209–215.
springerlink.com

Gate-length biasing is an active leakage reduction technique that utilizes the short-channel effect by marginally increasing the gate-length of MOS devices to reduce their leakage current with a small delay penalty. Similar to the dual-threshold technology, the gate-length biasing technology use longer gate-lengths transistor in non-critical paths to reduce leakage currents.

Compared with conventional CMOS circuits, adiabatic circuits obtain low power dissipations, because they utilize AC power supplies to recycle the energy of node capacitances. In the adiabatic circuits, the dynamic power dissipation can be reduced effectively. This paper presents a gate-length biasing technique for clocked adiabatic logic (CAL) circuits using dual-threshold technique to reduce sub-threshold leakage dissipations. An improved CAL register file using DTCMOS gate-length biasing is also addressed in this paper. All circuits are verified with HSPICE using a NCSU 45nm technology.

2 Gate-Length Biasing Technology for Improved CAL Circuits

The basic CAL gate, buffer / inverter, is shown in Fig. 1 (a) [9]. It is a dual-rail logic with true and complementary NMOS functional blocks (N1, N2) and cross-coupled PMOS latch (P1, P2). A sinusoidal power clock (*clk*) supplies the CAL circuits. *CX* is the auxiliary clock. The clamp transistors (N3 and N4) make the un-driven output node grounded. When the *clk* ramps down towards zero, the energy stored on the capacitance is recovered.

In basic CAL gate, the clamp transistors (N3 and N4) are on non-critical paths, since they only used for making the un-driven output node grounded. Therefore, the clamp transistors (N3 and N4) can use high-V_t transistors. The CAL buffer using gate-length biasing technology with dual-threshold techniques is shown in Fig. 1 (b). The cross-coupled PMOS latch (P1, P2) uses longer gate-length transistors further to reduce sub-threshold leakage currents.

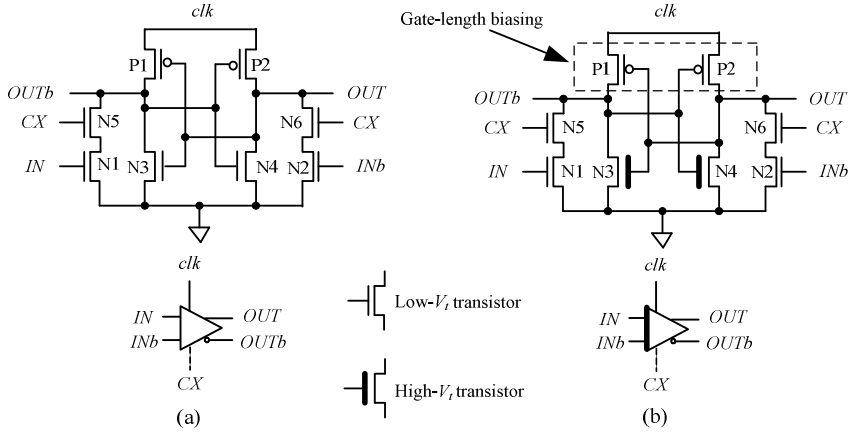


Fig. 1. (a) Basic CAL buffer, and (b) CAL buffer using gate-length biasing technology with dual-threshold CMOS.

In the improved CAL circuits, the charge of the auxiliary clock lines can be well recycled the power-clock *clk* because of using the sinusoidal clock signal, thus the energy loss of the improved CAL circuits is smaller than the conventional CAL circuits, as shown in Fig. 2 [9].

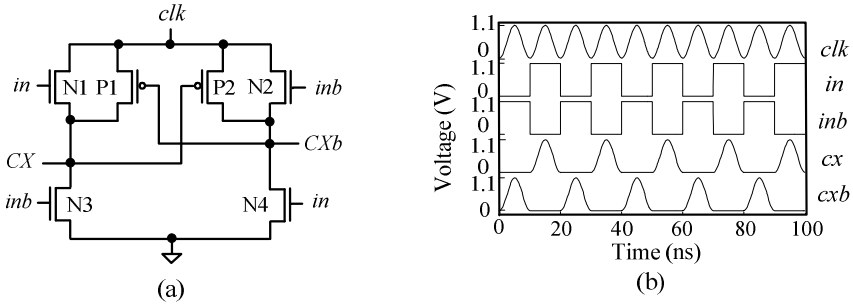


Fig. 2. Auxiliary clock generator and its simulation waveforms.

The power loss component can be analyzed from the output waveforms of CAL shown in Fig. 3. The energy loss occurs when the nodes of the CAL buffer are charged or discharged. Based on the power dissipation models of adiabatic circuits, an estimation technique for the active leakage dissipations of CAL circuits has been proposed in [10].

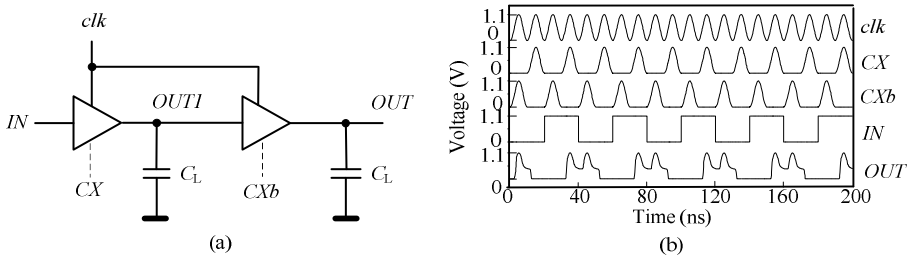


Fig. 3. CAL buffer chain and its simulation waveforms.

The energy dissipation per cycle of CAL circuits can be represented as

$$E_{total} = E_{adiabatic} + E_{non-adiabatic} + E_{leak}\,, \tag{1}$$

where $E_{adiabatic}$ is the full-adiabatic energy dissipation, $E_{non-adiabatic}$ is the non-adiabatic energy dissipation and E_{leak} is the average leakage energy dissipation per cycle of CAL circuits, respectively.

The full-adiabatic energy loss per cycle of CAL circuits can be represented as

$$E_{adiabatic} \approx (\pi^2 R_p C_L / 2T) C_L V_{DD}^2\,, \tag{2}$$

where C_L is the additional load capacitance of the CAL buffer, which is introduced for the purpose of estimating leakage dissipations, T is the period of the power-clock, V_{DD} is peak-peak voltage of the power clock, and R_p is the turn-on resistance of the transistors (P1 and P2). As shown in (1), the full-adiabatic dissipations of the CAL circuits with gate-length biasing is slightly larger than the basic CAL ones without using the gate-length biasing technique, since the gate-length of PMOS devices is marginally increased.

As energy recovery through the PMOS transistors, the *OUT* or *OUTb* of the CAL buffer falls to V_{DD} - V_{tp}, where V_{tp} is the threshold voltage of PMOS transistors, thus resulting in non-adiabatic energy dissipation, which is given by

$$E_{non-adiabatic} = C_L V_{tp}^2, \tag{3}$$

where V_{tp} is the threshold voltage of PMOS transistors. As shown in (2), the non-adiabatic energy dissipation of the DTCMOS CAL circuits is the same as the basic CAL ones without using the DTCMOS technique.

The average leakage energy dissipation per cycle of CAL circuits can be represented as

$$E_{leakage} = (V_{DD} I_{leak} / 2)T, \tag{4}$$

where I_{Leak} is the average leakage current per cycle of CAL circuits. In DTCMOS CAL with gate-length biasing, a higher threshold voltage is assigned to the clamp transistors (N3 and N4), and the cross-coupled PMOS latch (P1, P2) use longer gate-length transistors. Therefore, the leakage currents of the DTCMOS CAL circuits with gate-length biasing are reduced compared with the basic CAL ones without using DTCMOS technique and gate-length biasing.

The total energy dissipation per cycle of CAL circuits can also be represented as

$$E_{\text{total}} = \frac{k_1 C_{\text{L}}^{\ 2}}{T} + k_2 C_L + k_3 T, \tag{5}$$

where k_1 is $\pi^2 R_p V_{dd}^2 / 2$, k_2 is V_{tp}^2, and k_3 is $\frac{1}{2} V_{DD} I_{\text{Leak}}$, respectively. The leakage power dissipation of CAL circuits can be estimated by measuring total energy dissipations (E_{total1}, E_{total2} and E_{total3}) in three different capacitances (C_L, $2C_L$ and $3C_L$) of power-clocks [10]. The leakage power dissipation of CAL circuits can be given by

$$E_{leak} = E_{total3} - 3E_{total2} + 3E_{total1} \tag{6}$$

3 Improved DTCMOS CAL Register with Gate-Length Biasing

The improved CAL 32×32 register file using the dual-threshold technology and gate-length biasing technology is shown in Fig. 4, and its structure is the same as [11, 12], which consists of a storage-cell array, address decoders, read/write word-line drivers,

sense amplifiers, read bit-line and write data-line drivers, and a auxiliary clock generator that supplies auxiliary clock *CX* and *CXb* of the whole circuits. In the improved CAL 32×32 register files, the dual-threshold technology and gate-length biasing technology have been used to reduce sub-threshold currents.

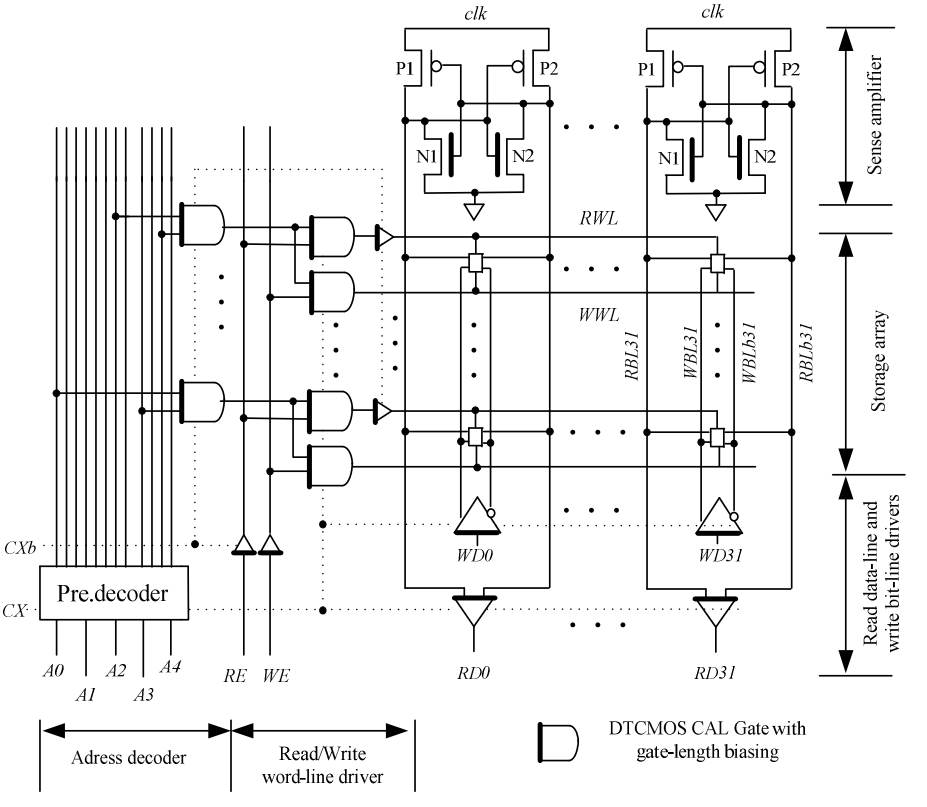


Fig. 4. The structure of improved DTCMOS CAL register using gate-length biasing.

For comparison, three register files based on the basic improved CAL circuits (CAL register), improved CAL circuits using DTCMOS technology (DT-CAL register), and improved CAL circuits using gate-length biasing and DTCMOS technologies (GLB-DT-CAL register) are simulated with HSPICE at a NCSU 45nm CMOS technology. In the GLB-DT-CAL register, the gate length of the transistors using gate-length biasing technology is 8% longer than the original transistors.

Table 1 show the total energy dissipation per cycle of the three register files. Fig. 5 show the leakage energy dissipation per cycle calculated according to Eq. 6. At 20MHz, the GLB-DT-CAL register file can save the leakage power dissipation of 25.2% and 10.4% compared with DT-CAL register and CAL register, respectively.

Table 1. Total energy dissipation per cycle of the 32×32 register files based on the basic improved CAL circuits (CAL register), improved CAL circuits using DTCMOS technology (DT-CAL register), and improved CAL circuits using gate-length biasing and DTCMOS technologies (GLB-DT-CAL register). The peak-to-peak voltage of the power-clock *pc* is taken as 1.1V.

Operation frequency (MHz)	10			20			40		
Additional load capacitance (fF)	10	20	30	10	20	30	10	20	30
CAL register	791fj	544fj	430fj	860fj	656fj	566fj	962fj	797fj	731fj
DT-CAL register	638fj	478fj	397fj	747fj	600fj	538fj	857fj	750fj	713fj
GLB-DT-CAL register	522fj	448fj	384fj	654fj	570fj	531fj	817fj	711fj	700fj

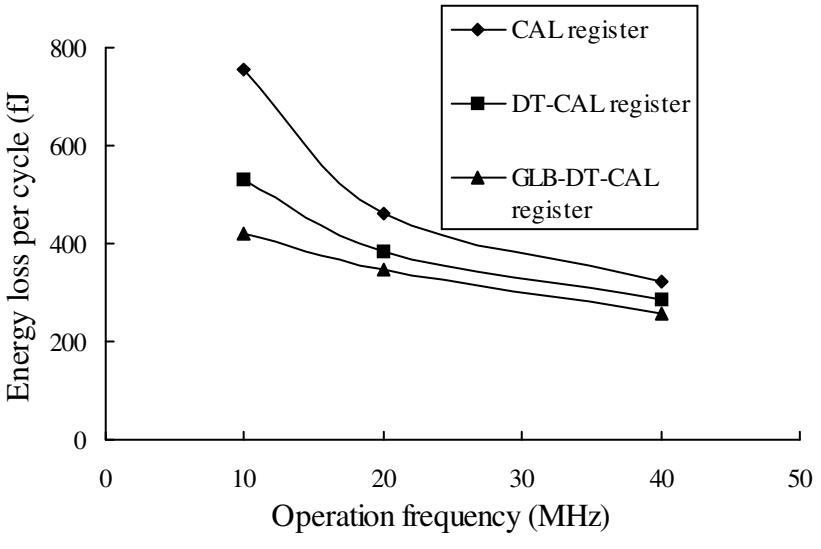


Fig. 5. Leakage energy dissipation of the 32×32 register files based on the basic improved CAL circuits (CAL register), improved CAL circuits using DTCMOS technology (DT-CAL register), and improved CAL circuits using gate-length biasing and DTCMOS technologies (GLB-DT-CAL register). The peak-to-peak voltage of the power-clock *pc* is 1.1V.

4 Conclusion

This paper has presented a gate-length biasing technique for clocked adiabatic logic (CAL) circuits using dual-threshold technique to reduce sub-threshold leakage dissipations. An improved CAL register file using DTCMOS gate-length biasing is addressed in this paper. All circuits are verified with HSPICE using a NCSU 45nm technology. The simulations show that the register file with gate-length biasing technology can attain large energy savings. The register based on DTCMOS CAL circuits using gate-length biasing can save leakage power dissipation of 25.2% compared with the basic CAL register.

Acknowledgments. Project is supported by National Natural Science Foundation of China (No. 61071049), Scientific Research Fund of Zhejiang Provincial Education Department (No. Z200908632), and Ningbo Natural Science Foundation (No. 2009A610066).

References

1. Rabaey, J.M., Pedram, M.: Low Power Design Methodologies, pp. 65–95. Kluwer Academic Publishers, Boston (1996)
2. Kim, N.S., Austin, T., et al.: Leakage Current: Moore's Law Meets Static Power. Computer 36(12), 68–75 (2003)
3. Fallah, F., Pedram, M.: Standby and Active Leakage Current Control and Minimization in CMOS VLSI Circuits. IEICE Trans. on Electronics E88-C(4), 509–519 (2005)
4. Kim, K.K., Kim, Y.B., Choi, M., Park, N.: Leakage Minimization Technique for Nanoscale CMOS VLSI. IEEE. Design and Test of Computers 24(4), 322–330 (2007)
5. Gupta, P., Kahng, A.B., Sharma, P., Sylvester, D.: Selective Gate-Length Biasing for Cost-Effective Runtime Leakage Control. In: Proc. ACM/IEEE Design Automation Conference, pp. 327–330 (2004)
6. Sirisantana, N., Wei, L., Roy, K.: High-Performance Low-Power CMOS Circuits Using Multiple Channel Length and Multiple Oxide Thickness. In: Proc. ICCD, pp. 227–232 (2000)
7. Wei, L., Chen, Z., Johnson, M., Roy, K.: Design and Optimization of Low Voltage High Performance Dual Threshold CMOS Circuits. In: Proc. ACM/IEEE DAC, pp. 489–492 (1998)
8. Hamzaoglu, F., Stan, M.R.: Circuit-Level Techniques to Control Gate Leakage for Sub-100 nm CMOS. In: Proc. Int. Symp. Low Power Electronics and Design, pp. 60–63 (2002)
9. Luo, C.N.: and Hu, J. P.: Single-Phase Adiabatic Flip-Flops and Sequential Circuits Using Improved CAL Circuits. In: Proc. IEEE ASICON 2007, pp. 126–129 (2007)
10. Hu, J.P., Zhu, J.G.: An Improved CAL Register File with Dual-Threshold Technique for Leakage Reduction, Advanced Materials Research (2010)
11. Moon, Y., Jeong, D.K.: A 32 x 32-b adiabatic register file with supply clock generator. Proc. IEEE Journal of Solid-State Circuits 33(5), 696–701 (1998)
12. Hu, J.P., Liu, B.B., Hu, X.Y., Zhang, S.: A Test Chip for CPAL Register File Fabricated in Chartered 0.35(μm CMOS Process. In: Proc. IEEE 51st Midwest Symposium on Circuits and Systems, pp. 434–437 (2008)

Near-Threshold Computing of ECRL Circuits for Ultralow-Power Applications

Yangbo Wu, Jindan Chen, and Jianping Hu

Faculty of Information Science and Technology, Ningbo University
315211 Ningbo City, China
nbuhjp@yahoo.cn

Abstract. Ultralow-power design has been a main challenge in modern VLSI circuits. This paper explores the near-threshold computing of adiabatic circuits. The characteristic of energy dissipations of the ECRL (efficient charge recovery logic) circuits is analyzed. It is found that its energy consumption is dependent linearly on operating voltage. In the near-threshold region, the ECRL circuits obtain considerable energy savings with a little performance penalty. An 8-bit ECRL Kogge-Stone adder is realized and simulated using HSPICE at a 45nm process with the NCSU PTM model. Simulations show that the power consumption of the near-threshold ECRL Kogge-Stone adder is reduced about 50% compared with the super-threshold one for clock rates ranging from 50MHz to 1.0 GHz.

Keywords: Low power, Near-threshold, Efficient charge recovery logic, Kogge-Stone adder.

1 Introduction

With the rapid progress in semiconductor technology, the density and operation speed of CMOS chips have been increasing, so that power consumption has become a critical concern in VLSI circuits. The classical circuit techniques to reduce power dissipation include transistor sizing and interconnect optimization, gated clock, multiple supply voltages, and dynamic controlling of supply voltage [1]. However, with the emergence of wireless sensors and biomedical applications that require ultralow energy dissipations, the classical circuit techniques are not sufficient to reduce the power consumption.

The power dissipation of a circuit can be significantly reduced by operating at a low supply voltage, since it has a square dependence on power supply voltage. It is shown that the minimum-energy point occurs in the sub-threshold region of MOS transistors for most logic families. Sub-threshold logic has shown significant improvement in the term of power consumption, in contrast to operation in strong inversion. Therefore, the sub-threshold digital logic design has obtained quit some attention over past the decade [2-5]. However, the performance of sub-threshold circuits is much lower than super-threshold ones due to the exponential relationship between delay and supply voltage. Scaling supply voltage to sub-threshold region only suits for ultra-low operation frequencies. Moreover, the robustness of

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 217–224.
springerlink.com

sub-threshold logic circuits must carefully be considered, since their operation relies only on leakage currents that are exponentially dependent on V_T and are therefore more sensitive to process variation than traditional super-threshold designs [6].

Recently, the near-threshold computing is presented. The supply voltage of near-threshold circuits is slightly above the threshold voltage of the transistors. This region retains much of the energy savings of sub-threshold operation with more favorable performance and variability characteristics [7].

The design ideas mentioned above aim at power optimization for the conventional CMOS logic circuits. In fact, the adiabatic circuit that utilizes AC power supplies to recycle the charge of node capacitances is a particularly attractive approach to reduce power dissipation [8]. However, the previously proposed adiabatic logic families focus mainly on nominal voltage circuits. As the operation frequency of adiabatic circuits is determined by AC power supplies, for a given operation frequency, the energy dissipation of adiabatic circuits can also be reduced by operating at a low supply voltage with a little performance penalty. This paper explores the near-threshold computing of the ECRL (efficient charge recovery logic) circuits. An 8-bit ECRL Kogge-Stone adder is realized and simulated using HSPICE at 45nm process with NCSU PTM model. In the near-threshold region, the ECRL circuits obtain considerable energy savings with a little performance penalty.

2 Review of ECRL Circuits

The basic structure of the ECRL buffer (inverter) and trapezoidal power clocks are shown in Fig. 1 [8]. The buffer consists of cross-coupled PMOS loads (P1 and P2) and NMOS pull-down transistors (N1 and N2) that are used for evaluation.

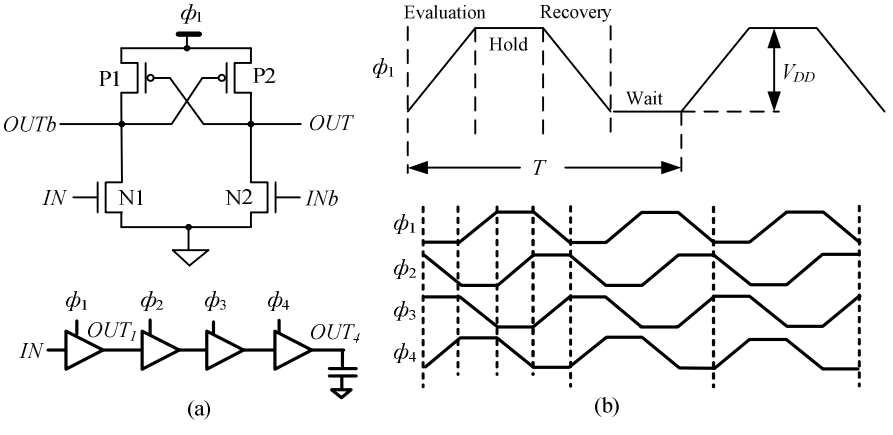


Fig. 1. ECRL buffer. (a) Schematic and buffers chain, and (b) Its four-phase power clocks.

The operation of the ECRL buffer can be divided four processes: evaluation, hold, recovery, and wait phases. For convenience, we assume *IN* is high and *INb* is low at the beginning of a cycle. As the supply clock ϕ_1 rises from zero to V_{DD}, *OUTb*

remains at a ground level, because *IN* signal turns on N1, while *OUT* follows ϕ_1 through P2. When ϕ_1 reaches V_{DD}, the outputs hold valid logic level. These values are maintained during the hold phase and used as inputs for evaluation of the next stage. After the hold phase, ϕ_1 falls down to a ground level, *OUT* node returns its energy to ϕ_1 so that the delivered charge is recovered. Thus, the clock ϕ_1 acts as both a clock and power supply. The wait phase is inserted for clock symmetry. In this phase, valid inputs are prepared in the previous stage. The ECRL circuits use four-phase power clocks to recover the charge delivered by the power clocks. Each clock is followed by the next clock with a 90° phase lag. When the previous stage is at the hold phase, the next stage must evaluate the logic value in the evaluation phase.

The general schematic of ECRL is shown in Fig. 2 (a), and it consists of PMOS loads and NMOS pull-down networks. The pull-down networks (PDN1and PDN2) are complementary, and implement the required logic function. The schematics of ECRL AND/NAND, OR/NOR, XOR/XNOR, and AND-OR/NAND-OR gates are also shown in Fig. 2.

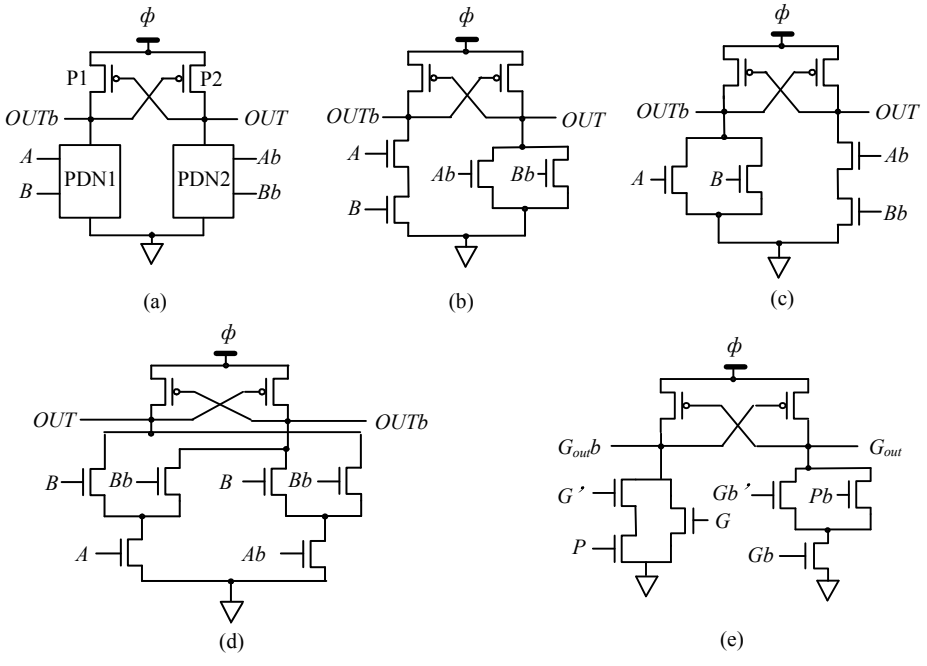


Fig. 2. ECRL buffer. (a) General Schematic, (b) AND/NAND gate, (c) OR/NOR gate, (d) XOR/XNOR gate, and (e) AND-OR/NAND-OR gate.

3 Near-Threshold Computing of ECRL Circuits

Voltage scaling represents a proper way for reducing the power dissipations in static CMOS and the adiabatic circuits. In this section we analyze the energy consumption characteristics of the ECRL circuits firstly, and then explore the relationship of supply voltage, energy consumption and performance of the basic ECRL gate.

3.1 Energy Consumption Analysis of ECRL Circuits

There are three main energy dissipations in adiabatic circuits, named as adiabatic energy dissipation, non-adiabatic energy dissipation and leakage power dissipation. For convenience, we will assume a symmetrical trapezoidal waveform in the following considerations, and thus the four phases (evaluation, hold, recovery, wait) are of equal duration, which is a quarter of a period $T=1/f$.

The energy loss occurs when the nodes of the ECRL circuits are charged or discharged, which is regarded as the main part in the total energy dissipation. The full-adiabatic energy dissipation per cycle of ECRL circuits can be represented as

$$E_{\text{adiabatic}} = 8fRC_L^2V_{\text{DD}}^2, \tag{1}$$

where C_{L} is the load capacitance of the ECRL circuits, f is the frequency of the power-clock, and R is the turn-on resistance of the PMOS transistors. The energy dissipation per cycle of the static CMOS circuits is

$$E_{stiatic} = C_{\text{L}}V_{\text{DD}}^2 . \tag{2}$$

Based on the Eq. (1) and Eq. (2), the cross-over frequency f_{c} can be derived as

$$f_{\text{c}} = 1/8RC_L . \tag{3}$$

The cross-over frequency f_c will rise with decreasing capacitance C_L and R. This means that adiabatic computing would more save energy for a given frequency with decreasing capacitance C_L and R.

During the evaluation and recovery phases, as the power clock falls below the $|V_{\text{tp}}|$, the PMOS transistor is turned off, so that the path between the power clock and the output node is disconnected, thus resulting in non-adiabatic energy dissipation. The amount of the energy loss is given by

$$E_{\text{non-adiabatic}} = C_{\text{L}} | V_{\text{tp}} |^2, \tag{4}$$

where $|V_{\text{tp}}|$ is the threshold voltage of PMOS transistors. This energy dissipation does not depend on the frequency, and it represents the lower bound energy of ECRL circuits.

The leakage power dissipation of ECRL circuits can be expressed as

$$p_{\text{s}}(t) = i_{\text{leak}}(t)v_{\text{DD}}(t) . \tag{5}$$

For convenience, we use the average leakage current I_{Leak} to replace $i_{leak}(t)$. The average leakage energy dissipation per cycle of ECRL circuits is

$$E_{\text{leak}} = \int_0^T p_{\text{s}}(t)dt = \int_0^T i_{\text{leak}}(t)v_{\text{DD}}(t)dt \approx \int_0^T I_{\text{leak}}v_{DD}(t)dt = \frac{1}{2}V_{\text{DD}}I_{\text{leak}}T , \tag{6}$$

where T is the period of the power-clock, V_{DD} is the peak-peak value of the trapezoidal power clock. The total energy dissipation per cycle of the ECRL buffer can be expressed as

$$E_{\text{total}} = E_{\text{adiabatic}} + E_{\text{non-adiabatic}} + E_{\text{leak}} = 8fRC_L^2V_{\text{DD}}^2 + C_{\text{L}}|V_{\text{tp}}|^2 + \frac{1}{2}V_{DD}I_{\text{leak}}T\,. \quad (7)$$

According to Eq. (7), the total energy dissipation can be reduced, when the operation supply voltage is lowered.

When a MOS transistor operates in the linear region, the on-resistance R is

$$R = 1/(\mu_n C_{ox}\frac{W}{L}(V_{GS} - V_T))\,. \quad (8)$$

R is inversely proportional to the overdrive voltage. Therefore, the total energy losses scale approximately linearly with the supply voltage compared to the proportionality to the square of the supply voltage for static CMOS.

3.2 Near-Threshold Computing of ECRL Circuits

To estimate the power dissipation of the ECRL circuits in different supply voltages, the basic ECRL gates are simulated using HSPICE at a 45nm CMOS process using the NCSU PTM BISM4 model. The threshold voltage of PMOS and NMOS transistors is -0.423V and 0.471V, respectively. The peak-to-peak voltage of the sinusoidal power clocks varies from 1V to 0.3V. Fig. 4 shows the energy loss curve of the ECRL XOR gate in 50MHz, 100MHz, 500MHz, and 1GHz. The size of the MOS transistors using in the XOR gate is W/L=40λ/2λ and λ=25nm.

The operating regions of ECRL circuits can be divided three parts: Subthreshold Region, Near-threshold Region and Super-threshold Region. From the energy curve shown in Fig. 4, we can observe that the energy dissipation of the ECRL XOR gate depend linearly on the supply voltage. For a certain operating frequency, there exists a minimum supply voltage, at which the circuits can operate correctly. Though adiabatic circuits can operate at the subthreshold region, and obtain considerable energy reduction, the maximum operation frequency is lower than the other two regions.

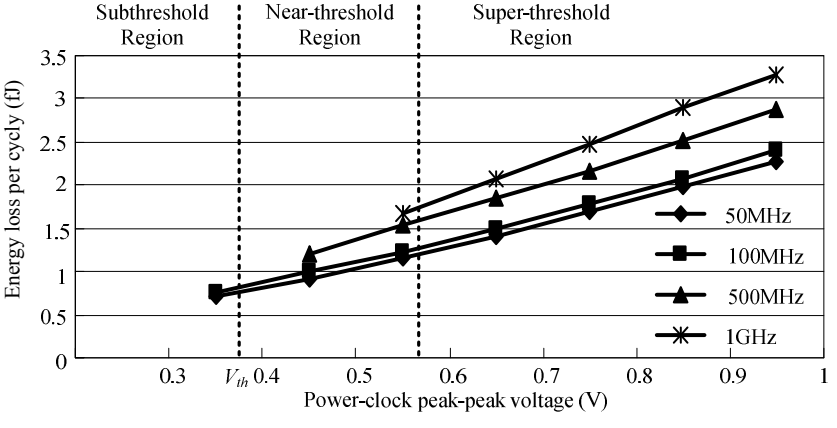


Fig. 4. Energy consumption of ECRL XOR gates in different supply voltage operating regions and frequencies.

When the ECRL XOR gate operating at 0.4V peak-peak voltage of the power-clock, the maximum frequency only reaches 100MHz. Moreover, subthreshold circuits are extremely sensitive to variations in supply, temperature and processing factors.

In near-threshold region, the operation frequency is wider than the subthreshold one. For the peak-peak voltage at 0.6V, the maximum operation frequency of the ECRL XOR gate is up to 1GHz. And it can save about 50% of energy dissipation compare to nominal operation voltage (1.0V). In the near-threshold region, the MOS transistors usually are in strong or moderation inversion, the robustness of near-threshold circuits are improved greatly than that of subthreshold circuits. These benefits are much more attractive for a wide variety of new applications.

4 Near-Threshold ECRL Kogge-Stone Adder

The Kogge-Stone adder is a carry-lookahead adder with parallel prefix form, and is widely used in the industry for high performance arithmetic circuits [9]. In this section an 8-bit Kogge-Stone adder is realized using ECRL gates to verified near-threshold computing ideas. The schematic diagram of the 8-bit ECRL Kogge-Stone adder is shown in Fig. 5. It consists of propagate and generate signal generation circuits, carry-lookahead tree using dot operation and sum generation circuits.

The propagate and generate functions have been defined as

$$P_n = A_n + B_n \text{ and } G_n = A_n \cdot B_n . \tag{9}$$

The carry-lookahead tree using dot operation generates the complete set of carry bits. The dot operator is defined by

$$(G, P) \bullet (G', P') = (G + PG', PP') . \tag{10}$$

However, the designed adder does not have an input carry, C_{in}, since it is to be used as a standalone adder, and is not a part of a bigger adder. Therefore, C_{in} of the adder is assumed to be 0, and then the sum generation function can be expressed as

$$S_n = P_n \oplus C_{n-1} = P_n \oplus G_{n-1} . \tag{11}$$

In Fig. 5, the triangles are buffers, which are used for maintaining a pipeline. These buffers propagate the correct logic values for addition like latch in the general pipeline structure. To obtain the first result of the addition, a few cycles of latency are needed. The latency is only 1.25 cycles for the 8-bit adder.

The 8-bit Kogge-Stone adder is simulated using HSPICE at the 45nm CMOS process using the NCSU PTM model. The basic gates using in the adder are shown in Fig. 1 and Fig. 2. The device size of PMOS transistors of all the gates is taken with $W/L = 40\lambda/2\lambda$, and $\lambda = 45$nm. The size of all NMOS transistors is taken with $40\lambda/2\lambda$, except for the AND gate's N3 and N4, OR gate's N1and N2, AND-OR gate's N1 with $W/L = 20\lambda/2\lambda$. The energy consumption of the adder operating at different supply voltages is evaluated. Fig. 7 shows the energy loss curve of the adder operating at 0.6V and 1.0V peak-peak voltage of sinusoidal power-clocks.

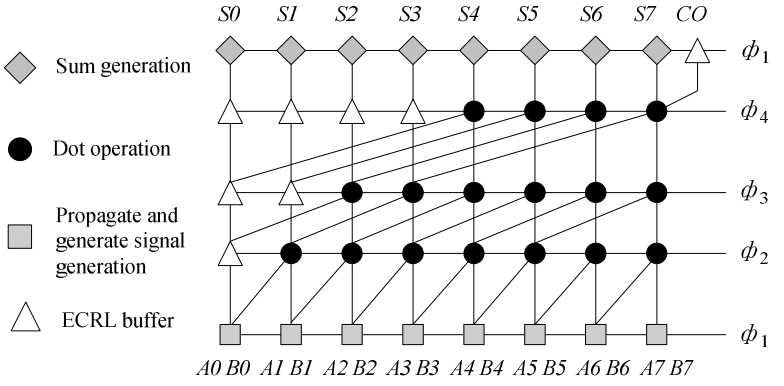


Fig. 5. Schematic diagram for 8-bit ECRL Kogge-Stone adder.

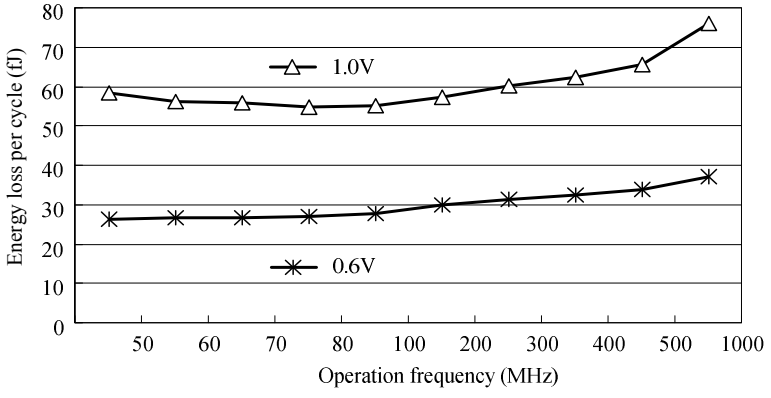


Fig. 6. Energy consumption of 8-bit Kogge-stone adder.

The simulation results show that the proposed near-threshold ECRL 8-bit Kogge-Stone adder has considerable power savings. Compared to the conventional ECRL 8-bit Kogge-Stone adder operating at the nominal voltage (1.0V), the near-threshold ECRL 8-bit Kogge-stone one operating at the peak-peak voltage of 0.6V attains about 50% energy savings for clock rates ranging from 50MHz to 1.0 GHz.

5 Conclusion

In this paper, we have proposed a near-threshold computing scheme of ECRL circuits. We analyzed the characteristic of the power consumption of ECRL circuits, and found that power dissipation scales down linearly with the supply voltage. For the ECRL circuits, the near-threshold operation region is the optimal one compare to the other two regions. The simulation results showed that the near-threshold ECRL adder consumed about 50% of the dissipated energy of the conventional ECRL one operating at

normal source voltage. So the near-threshold computing is an attractive method for ultra-low-power applications.

Acknowledgments. Project is supported by National Natural Science Foundation of China (No. 61071049), Zhejiang Science and Technology Project of China (No. 2010C31G2070017), Scientific Research Fund of Zhejiang Provincial Education Department (No. Z200908632), and Ningbo Natural Science Foundation (No. 2009A610066).

References

1. Paul, B.C., Agarwal, A., Roy, K.: Low-Power Design Techniques for Scaled Technologies. Integration, the VLSI Journal 39(2), 64–89 (2006)
2. Soeleman, H., Roy, K.: Ultra-low Power Digital Sub-threshold Logic Circuits. In: Proc. Int. Symp. Low Power Electron. Design, pp. 94–96 (1999)
3. Calhoun, B.H., Chandrakasan, A.P.: Characterizing and Modeling Minimum Energy Operation for Subthreshold Circuits. In: Proc. Int. Symp. Low Power Electron. Design, pp. 90–95 (2004)
4. Hanson, S., Seok, M., Sylvester, D., Blaauw, D.: Nanometer Device Scaling in Subthreshold Logic and SRAM. IEEE Trans. Electron Devices 55(1), 175–185 (2008)
5. Kim, T.-H., Liu, J., Keane, J., Kim, C.H.: Circuit Techniques for Ultra-low Power Subthreshold SRAMs. In: Proc. IEEE Int. Symp. Circuits Syst., pp. 2574–2577 (2008)
6. Verma, N., Kwong, J., Chandrakasan, A.P.: Nanometer MOSFET Variation in Minimum Energy Subthreshold Circuits. IEEE Trans. Electron Devices 55(1), 163–173 (2008)
7. Dreslinski, R.G., Wieckowski, M., Blaauw, D., Sylvester, D., Mudge, T.: Near-Threshold Computing: Reclaiming Moore's Law Through Energy Efficient Integrated Circuits. Proceedings of the IEEE 98(2), 253–266 (2010)
8. Moon, Y., Jeong, D.K.: An Efficient Charge Recovery Logic Circuit. IEEE Journal of Solid-State Circuits 31(4), 514–522 (1996)
9. Kogge, P.M., Stone, H.S.: A Parallel Algorithm for the Efficient Solution of a General Class of Recurrence Equations. IEEE Transactions on Computers C-22(8), 786–793 (1973)

A Dual-Threshold CMOS Technique of P-Type CAL Circuits for Leakage Reduction

Weiqiang Zhang, Li Su, and Yangbo Wu

Faculty of Information Science and Technology, Ningbo University
315211 Ningbo City, China
nbuhjp@yahoo.cn

Abstract. Dual-threshold CMOS (DTCMOS) has been proven as an effective way to reduce sub-threshold leakage currents both in the active and standby modes, while the high performance of the circuits is still maintained. P-type logic circuits can reduce the gate leakage dissipations significantly in nanometer CMOS processes. This paper presents a dual-threshold CMOS scheme for p-type clocked adiabatic logic (CAL) circuits to reduce leakage power dissipations. An ISCAS 74182 benchmark circuit from the ISCAS 74X-series is verified using DTCMOS p-type adiabatic circuits. All circuits are simulated using the 65nm CMOS process with gate oxide materials by HSPICE simulation tool. The results show that ISCAS 74182 benchmark circuit based on P-type adiabatic circuits with the dual-threshold voltage technique can achieve large energy savings, since both sub-threshold and gate leakage consumptions are reduced effetely.

Keywords: Logic circuits, DTCMOS, P-type logic, Clocked adiabatic logic, Leakage reduction.

1 Introduction

As technology scaling trends continue in future generations, leakage currents are not neglectable any more. Even today, leakage current consumption has become the major contributor to power consumption as technology drops below the 65nm feature size. Leakage currents are mainly from the following three sources: sub-threshold leakage current due to very low threshold voltage (V_{th}), gate leakage current due to very thin gate oxide (T_{ox}), and band-to-band tunneling leakage current due to heavily-doped halo [1]. The subthreshold leakage currents are exponentially dependent on device threshold voltage, while the gate leakage currents increase exponentially with reducing gate oxide thickness [2, 3].

The DTCMOS (Dual-threshold CMOS) has been proven as an effective way to reduce subthreshold leakage currents both in the active mode and the standby mode. The DTCMOS strategy involves using low threshold voltage (V_{th}) transistors for the gates on the critical paths and high threshold voltage (V_{th}) transistors for the gates in the non-critical paths [4, 5]. High-V_{th} devices can be used to reduce leakage currents while low-V_{th} devices can be used to maintain high performances [6]. The DTCMOS is a very

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 225–232.
springerlink.com

attractive technique for reducing subthreshold leakage currents with no extra area by simply using high V_{th} transistors.

P-type logic circuits that consist mostly of PMOS transistors have been proposed to reduce gate leakage dissipations [3]. The technique is based on the fact that the gate leakage through SiO_2 for the PMOS transistors is an order of magnitude lower than for the NMOS ones [3]. The reason for this difference is that electron tunneling from conduction band (ECB) is the dominant component of gate leakages for the NMOS, whereas it is the hole tunneling from valence band (HVB) for the PMOS. As the barrier height for HVB (4.5ev) is significantly larger than for ECB (3.1ev), PMOS transistors have much lower gate leakages than NMOS [3, 7].

Adiabatic logic is a promising low power circuit to reduce the energy consumption in digital by using a constant current to efficiently charge a capacitor. Therefore, a clocked power supply (power clock) is used, that consists of four states. Only during one of the four states the whole supply voltage V_{DD} drops across the gate. Hence a reduction of the leakage currents is implemented explicitly by the power clock in adiabatic logic circuits [7].

This paper explores P-type CAL (Clocked Adiabatic Logic) circuits [8] with dual-threshold voltage technique for reducing leakage currents. An ISCAS 74182 benchmark circuit from the ISCAS 74X-series set [9] is verified using the DTCMOS P-Type CAL circuits. Both sub-threshold and gate leakage consumptions are reduced effetely.

2 Dual-Threshold Technology for p-Type CAL

N-type logic circuits, such as CPL (complementary pass-transistor logic) and DVSL (differential cascode voltage switch logic) circuits, have been addressed extensively in the past years, which use more NMOS transistors than PMOS ones, since NMOS transistors has better conduction, smaller area and node capacitance than PMOS ones. As MOS transistors get smaller and the gate oxide gets thinner, PMOS transistors have lower gate leakage current than NMOS ones. Based on this fact, the circuits that consist mostly of PMOS transistors, which are named as P-type logic circuits, can reduce gate leakage power dissipations compared with N-type logic circuits [8].

Basic P-type CAL buffer/inverter has been reported in [8], as shown in Fig. 1 (a). Cascaded P-type CAL gates are driven by a single-phase power clock (*clk*), as shown in Fig. 1(b). The structure and operation of the P-type CAL circuit are complementary to the N-type CAL one [10, 11]. It is realized by using NMOS loads and PMOS pull-up transistors. Its simulation waveforms are shown in Fig. 1 (c).

The DTCMOS has been widely explored in conventional CMOS circuits. It uses low threshold voltage (V_{th}) transistors in the critical paths and high threshold voltage (V_{th}) transistors in the non-critical paths. Similarly, the DTCMOS can be also used for the adiabatic circuits [12]. A dual-threshold scheme for the P-Type CAL buffer is illustrated in Fig. 2. The clamp transistors (P3 and P4) use high V_{TH} transistors to reduce leakage power dissipations, while the other transistors of the P-type CAL circuits use the low V_{TH} transistors to retain circuit performances.

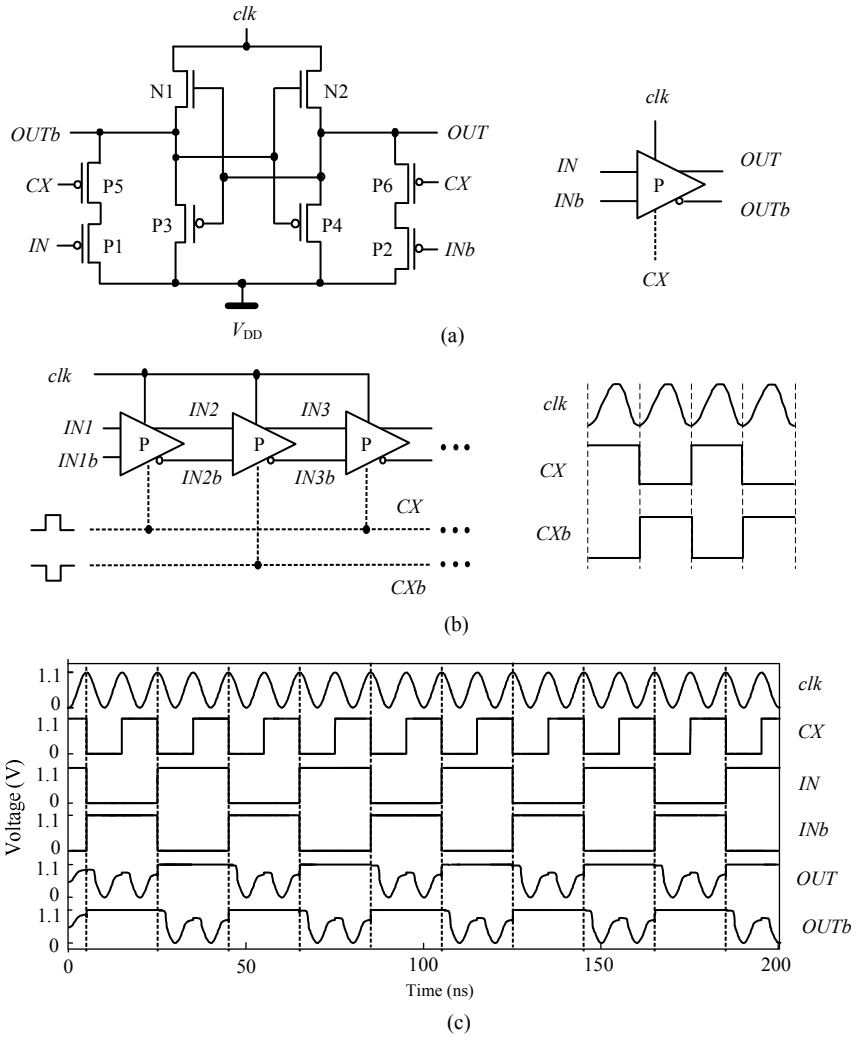


Fig. 1. Basic P-Type CAL buffer. (a) schematic and symbol, (b) buffer chain, and (c) simulation waveforms.

Based on the power dissipation models of adiabatic circuits, active leakage dissipations can be estimated by testing total leakage dissipations using SPICE simulations. The estimation technique for basic P-CAL circuits has been reported in [8], which can also be used for estimating active leakage dissipations of the DTCMOS p-type CAL circuits.

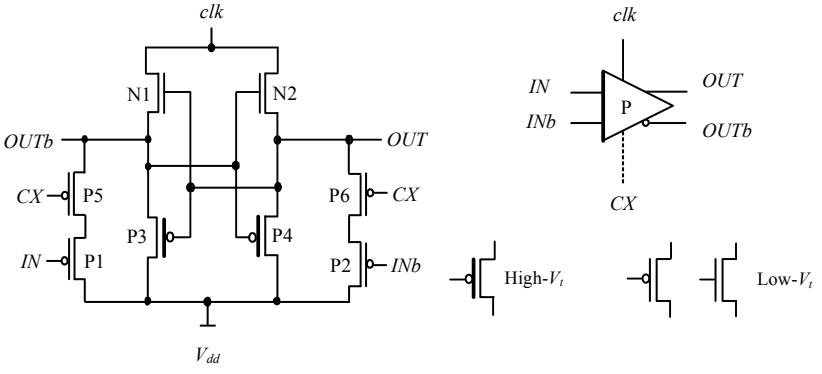


Fig. 2. P-Type CAL buffer with dual-threshold technology.

The energy dissipation per cycle of CAL circuits includes adiabatic energy dissipation ($E_{adiabatic}$), non-adiabatic energy loss ($E_{non\text{-}adiabatic}$) and leakage dissipation (E_{Leak}). The total energy dissipation per cycle of the CAL buffer can be represented as

$$\begin{aligned} E_{total} &= E_{adiabatic} + E_{non-adiabatic} + E_{leak} \\ &\approx (\pi^2 R_N C_L / 2T) C_L V_{DD}^2 + C_L V_{TN}^2 / 2 + (V_{DD} I_{leak} / 2) T \end{aligned}, \tag{1}$$

where C_L is additional load capacitance of the CAL buffer, which is introduced for the purpose of estimating leakage dissipations, T is period of the power-clock, R_N is turn-on resistance of the transistors (N1 and N2), V_{TN} is threshold voltage of NMOS transistors, and I_{leak} is average leakage current per cycle of P-type CAL circuits.

According to Eq. (1), the total energy dissipation per cycle of CAL circuits can also be represented as

$$E_{\text{total}} = \frac{k_1 C_{\text{L}}^{\ 2}}{T} + k_2 C_L + k_3 T \,, \tag{2}$$

where k_1 is $\pi^2 R_{\text{N}} V_{\text{DD}}^2 / 2$, k_2 is $V_{\text{TN}}^2 / 2$, and k_3 is $V_{DD} I_{\text{Leak}} / 2$. The leakage power dissipation of P-type CAL circuits can be estimated by measuring total energy dissipations (E_{total1}, E_{total2} and E_{total3}) in three different capacitances (C_{L}, $2C_{\text{L}}$ and $3C_{\text{L}}$) of power-clocks. The leakage power dissipation of CAL circuits can be given by

$$E_{leak} = E_{total3} - 3E_{total2} + 3E_{total1} \,, \tag{3}$$

3 Combinational Logic Circuits Using DTCMOS P-Type CAL

An ISCAS 74182 benchmark circuit from the ISCAS 74X-series is shown in Fig. 3 [9]. The 74182 benchmark circuit based on P-type CAL is verified using the dual-threshold voltage technique and single-threshold voltage. All the circuits use P-type CAL circuits, which need a single-phase power clock (*clk*) and an auxiliary clock

generator that supplies auxiliary clock *CX* and *CXb* of the whole circuits. The auxiliary clock *CX* and *CXb* control the cascade circuit alternately.

We can use dual-threshold technology in the P-Type CAL ISCAS 74182 benchmark circuit, and their gates are realized by using required PMOS logic blocks to replace the input PMOS transistors (P1 and P2) of the P-type CAL buffer shown in Fig. 2. The simulated waveforms of the ISCAS 74182 benchmark circuit based on P-Type CAL are shown in Fig. 4. Fig. 5 and Fig. 6 show the total and the leakage energy dissipations per cycle of ISCAS 74182 benchmark circuits based on P-Type CAL with the dual-threshold circuits and single-threshold circuits at the 65nm CMOS technology, respectively. Compared with P-Type CAL circuits with the single-threshold, both total and leakage energy dissipations of the dual-threshold P-Type CAL circuits can be reduced effectively.

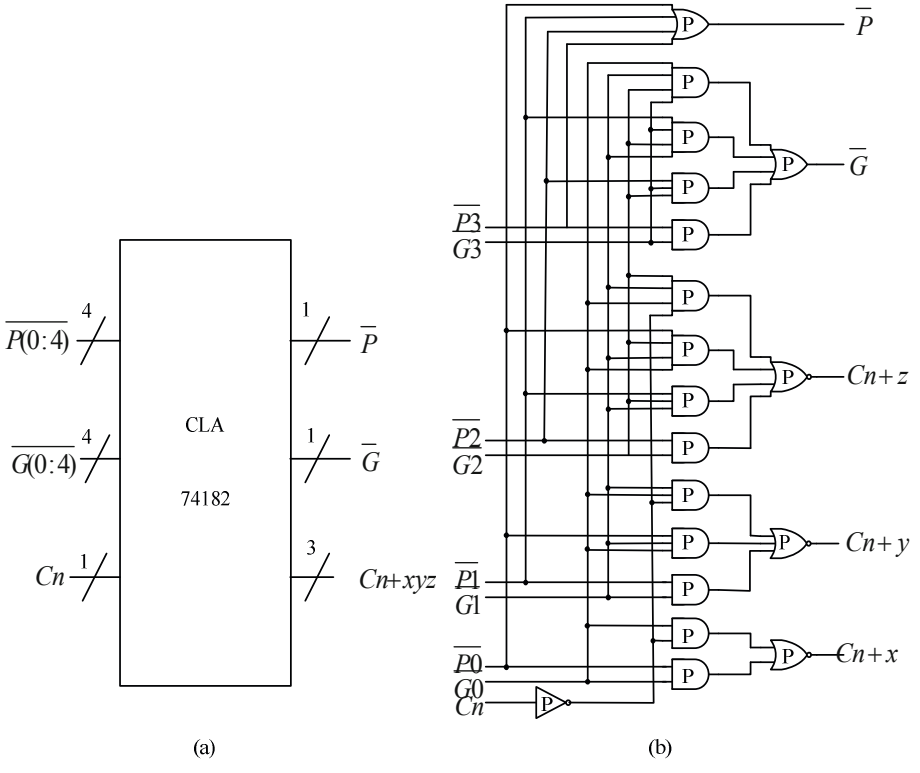


Fig. 3. The ISCAS 74182 benchmark circuit. (a) symbol, and (b) gate-level schematic.

Fig. 7 shows the leakage power consumption saving rate. Compared with the single-threshold implementation, the leakage power consumption saving of the ISCAS 74182 benchmark circuit based on the dual-threshold P-Type CAL circuits is about 48% at 100MHz.

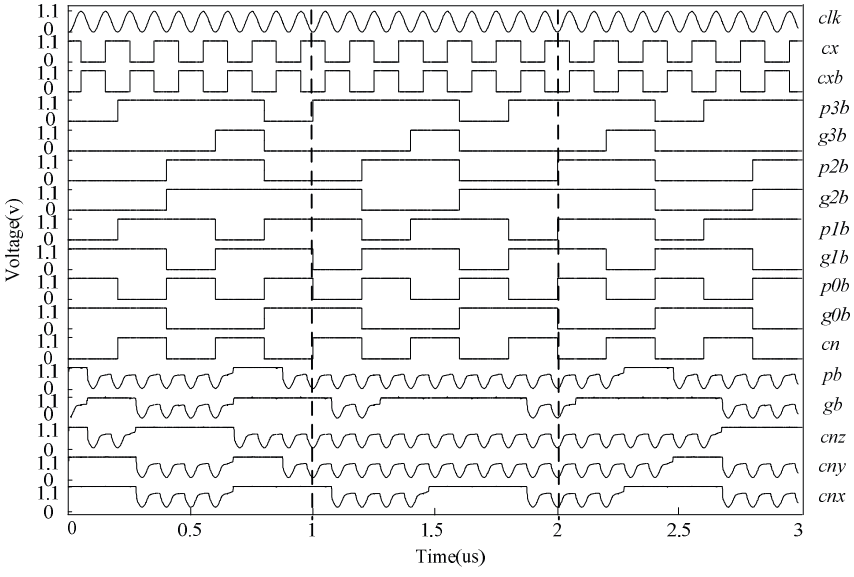


Fig. 4. The simulation waveforms of the ISCAS74182 benchmark circuit based on P-Type CAL.

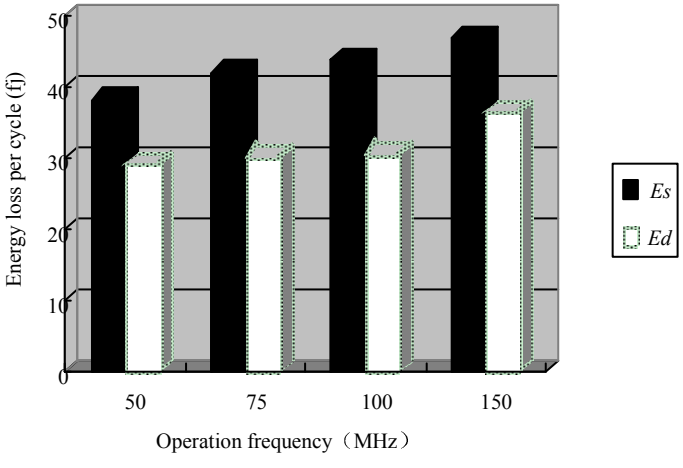


Fig. 5. *Es* and *Ed* are the total energy dissipations of the ISACS 74182 benchmark circuits based on single-threshold and dual-threshold P-Type CAL circuits at the 65nm CMOS technology, respectively.The peak-to-peak voltage of the power-clock *pc* is 1.1V.

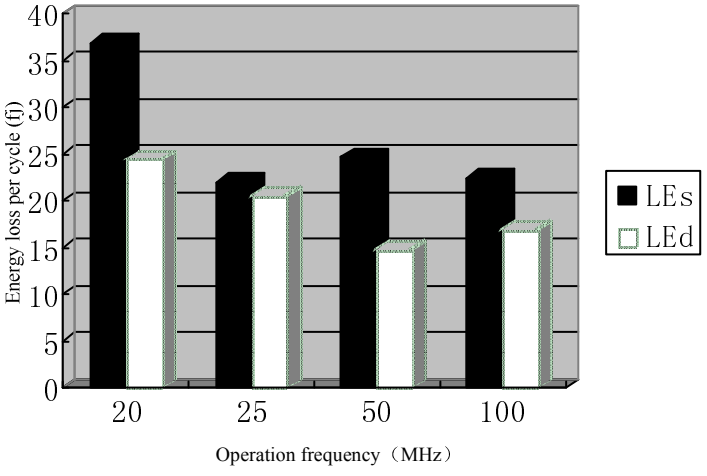


Fig. 6. *LEs* and *LEd* are the leakage energy dissipations of the ISACS 74182 benchmark circuits based on single-threshold and dual-threshold P-Type CAL circuits at the 65nm CMOS technology, respectively.The peak-to-peak voltage of the power-clock *pc* is 1.1V.

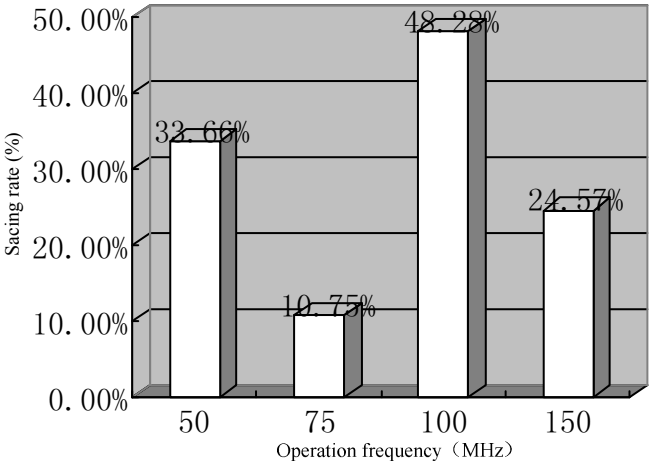


Fig. 7. Leakage dissipation saving rate of ISACS 74182 circuits based on dual-threshold P-Type CAL circuits compared with single-threshold P-Type CAL ones at the 65nm CMOS technology.

4 Conclusion

This paper proposes a dual-threshold technology for P-type clocked adiabatic logic (CAL) circuits to reduce leakage power. The ISCAS 74182 benchmark circuits based on the P-Type CAL circuits are used to testify this technology. Compared with the single-threshold P-Type CAL, the total and leakage energy dissipations of ISCAS 74182 benchmark circuit based on the dual-threshold P-Type CAL circuits is significantly reduced, since both sub-threshold and gate leakage consumption are reduced effetely.

Acknowledgments. Project is supported by National Natural Science Foundation of China (No. 61071049), Scientific Research Fund of Zhejiang Provincial Education Department (No. Z200908632), and Ningbo Natural Science Foundation (No. 2009A610066), and supported by the Scientific Research Foundation of Graduate School of Ningbo University (No. PY20090011).

References

1. Roy, K., Mukhopadhyay, S., et al.: Leakage Current Mechanisms and Leakage Reduction Technique in Deep-submicrometer CMOS Circuits. Proceeding of IEEE 91(2), 305–327 (2003)
2. Kim, N.S., Austin, T., et al.: Leakage Current: Moore's Law Meets Static Power. Computer 36(12), 68–75 (2003)
3. Hamzaoglu, F., Stan, M.R.: Circuit-Level Techniques to Control Gate Leakage for Sub-100 nm CMOS. In: Proc. Int. Symp. Low Power Electronics and Design, pp. 60–63 (2002)
4. Fallah, F., Pedram, M.: Standby and Active Leakage Current Control and Minimization in CMOS VLSI Circuits. IEICE Trans. on Electronics E88–C(4), 509–519 (2005)
5. Wei, L., Chen, Z., Johnson, M., Roy, K.: Design and Optimization of Low Voltage High Performance Dual Threshold CMOS Circuits. In: Proc. Annual ACM IEEE Design Automation Conference, pp. 489–492 (1998)
6. Kao, J.T., Chandrakasan, A.P.: Dual-Threshold Voltage Techniques for Low-Power Digital Circuits. IEEE Journal of Solid State Circuits 35(7), 1009–1018 (2000)
7. Teichmann, P., et al.: Gate Leakage Reduction by Clocked Power Supply of Adiabatic Logic Circuits. Advances in Radio Science (3), 281–285 (2005)
8. Hu, J.P., Ye, L.F., Su, L.: A New P-Type Clocked Adiabatic Logic for Nanometer CMOS Processes with Gate Oxide Materials. Applied Mechanics and Materials 29-32, 1930–1935 (2010)
9. http://www.eecs.umich.edu/~jhayes/iscas/
10. Maksimovic, D., Oklobdzija, V.G., Nikolic, B., Current, K.W.: Clocked CMOS Adiabatic Logic with Integrated Single-Phase Power-Clock Supply. IEEE Trans. VLSI Syst. 8(4), 460–463 (2000)
11. Luo, C.N., Hu, J. P.: Single-Phase Adiabatic Flip-Flops and Sequential Circuits Using Improved CAL Circuits. In: Proc. IEEE ASICON 2007, pp. 126–129 (2007)
12. Hu, J.P., Yu, X.Y., Sheng, X.L.: An Adiabatic Register Based on Dual Threshold Leakage Reduction Technique. Advanced Materials Research 121-122, 148–153 (2010)

An Improved Phase Shift Modulation Method for Cascade Active Power Filter with Low Switching Frequency*

Congzhe Gao[1], Xinjian Jiang[1], Yongdong Li[1], and Junling Chen[2]

[1] Department of Electrical Engineering, Tsinghua University
Beijing, China
gcz07@mails.thu.eu.cn
[2] Institute of Electrical Engineering, Chinese Academy of Science
Beijing, China
chenjunling@mail.iee.ac.cn

Abstract. The cascade H-bridge inverters are widely utilized in high voltage and large capacity occasions. Furthermore, it is suitable for the medium voltage level active power filter (APF) field. Considering the high speed response is of great importance for APF application, the large capacity APF generally applies low level switching frequency which will affect the response speed. So it is very important to design appropriate modulation method for large capacity APF. Phase shift PWM (PS-PWM) method can be regarded as a good solution for inverter with low level switching frequency. Further study on PS-PWM is made in this paper and an improved PS-PWM method is presented. With the novel modulation method, high response speed of APF can be guaranteed with relative low switching frequency. The algorism is simple to be implemented. A three-stage cascade APF (CS-APF) simulation model is built and simulated, and the results show that the novel modulation method is an effective and usable solution to promote the response of CS-APF.

Keywords: CS-APF, modulation, PS-PWM, time delay.

1 Introduction

Harmonics exist in the network and EMI brought by harmonics is becoming serious with an increasing implementation of nonlinear power-electronic devices. Nowadays, passive filter (PF) is the main filter device used in the network, especially in the high-voltage and medium-voltage level. Although significant progress has been made in active power filter (APF) technology which has been used in some low-voltage (e.g. 380V) and low-capacity levels, only few are put into practical operation. Constrained by voltage level of power-electronic devices, less APFs exist for successful medium-voltage applications (e.g. 10kV), where series and parallel of several single APFs are usually utilized to enlarge the APFs' capacity and voltage [1, 2]. This undoubtedly

* This work is funded by 863Plan(2007AA05Z227).

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 233–242.
springerlink.com

requires expensive and large volume transformers. Another solution is to select hybrid topology, where the PF devices stress most of the voltage from the grid [2, 3].

In cascade H-bridge topology, H-bridge units are in series to achieve high voltage levels which makes it possible to connection the inverter directly to the grid without transformers. At present, static synchronous compensator (STATCOM) with cascade topology have already been applied for commercial application [4, 5]. Furthermore, this topology can be directly used in APF field and the relative studies have been made by some researchers [6, 7]. Cascade APF is used to enlarge the capacity and voltage. However, switching frequency of large-capacity power-electronic devices is usually kept at low levels. However, the APF has to respond fast enough in order to attain a better harmonic compensation performance. High-frequency devices such as IGBT are selected in this application. In practical application, switching frequency of large-capacity IGBT (e.g. 1200A/1700V) is only several kHz which is up to the cooling rate. Meanwhile, frequency of the switching devices is required as lower as possible in order to guarantee the stability and efficiency of the whole system.

Therefore, PWM modulation strategy with fast response and low switching frequency is of great importance to cascade APF. Low-switching-frequency cascade APF is studied in the paper [8]; A hybrid modulation method is presented in paper [9]. Low switching frequency is realized in two-level cascade topology but the equivalent switching frequency is half of the carrier in phase-shifting PWM method. Thus, this method is appropriate only in the case of high switching frequency; SVPWM is studied in paper [10, 11]. However, there are many redundant vectors with an increase of level numbers, which results in complexity of the design.

This paper presents an improved PS-PWM method, which keeps high equivalent switching frequency and fast response speed with real low switching frequency. This paper designs a three-stage cascade APF simulation system, and introduces the control principles. And the new PWM method is verified by the simulation at the end of the paper.

2 Structure of Cascade APF

Medium-voltage cascade APF topology is shown in Figure1. Without transformers the cascade APF directly connects to the grid. H-bridge cascade can be either delta connection or star connection. The star connection is selected here. N is the cells number in per phase leg of cascade APF, and it is determined by the grid voltage level, together with the H-bridge DC-link voltage level. Capacitor on the DC-link side is float and a parallel discharge circuit is added.

2.1 Control of the Cascade APF

As shown in Fig2 to Fig4, the control system consists of internal current loop and external voltage loop. The harmonic detection method is based on the instantaneous reactive power theory (IRP) as shown in Fig2.

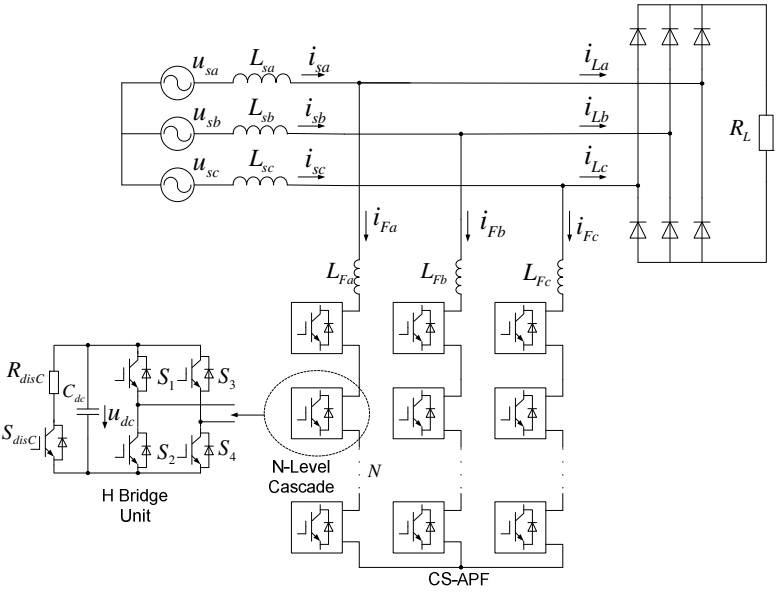


Fig. 1. Topology of CS-APF

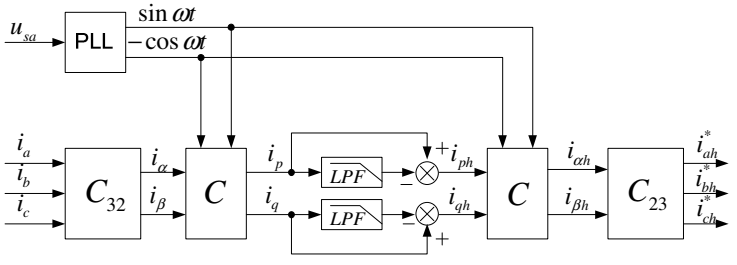


Fig. 2. Harmonic detection method

$$\mathbf{C}_{32} = \sqrt{\frac{2}{3}}\begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} \sin \omega t & -\cos \omega t \\ -\cos \omega t & -\sin \omega t \end{bmatrix} \tag{1}$$

The current control loop consists of load current feed-forward control and grid current feedback control. Fig3 shows the structure of current loop (The three phase system is replaced by single phase for simple). A proportional controller is adopted to achieve the current tracking control. To get better performance and suppress the effects of the line voltage, a grid voltage feed-forward compensation regulator is adopted, then the proportional controller equals as follows:

$$P' = \frac{P}{u_{dc} - |u_s|} \tag{2}$$

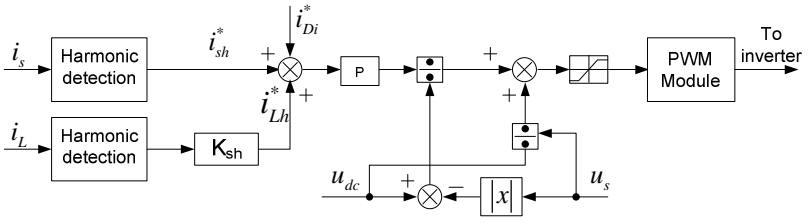


Fig. 3. Current control scheme

For the DC-link voltage loop, a PI controller is designed to keep the total DC-link voltage of one phase leg equal to others. However, unbalanced voltage problem may exist in CS-APF, and a control method is introduced in papers [6, 7]. Besides the reference current generated by the PI controller maybe not suitable for star connection, so a discharge circuit is added to the DC-link capacitor. When the voltage is higher than the setting level u_{high}^*, this circuit can discharge the capacitor. And it is important that the discharging voltage level has to be higher than the DC-link working level. A proportional controller is used to control the discharging PWM duty.

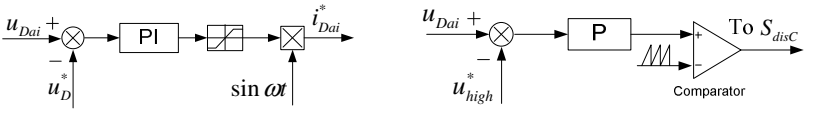


Fig. 4. DC-link voltage control scheme

2.2 New PWM Method for Cascade APF

For CS-APF with PS-PWM modulation method, the phase of each cell carrier wave should be shift by $\theta = \pi / n$, n is the cascade cell number. Besides the left leg and right leg modulation wave phase is opposite (Fig5). Then the equivalent switching frequency of the output voltage is $2n$ times of the switching frequency, and the harmonic component is very low, as shown in Fig5:

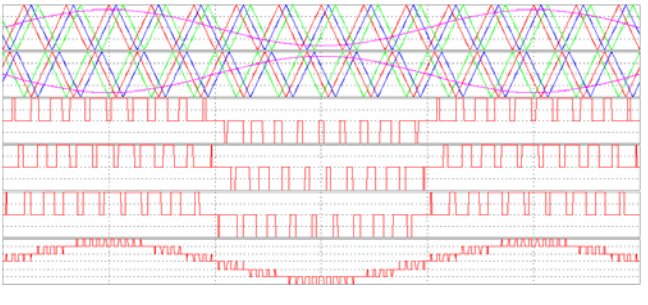


Fig. 5. PS-PWM principle

Focus on the time delay. As for common two level voltage inverter with SPWM, time delay brought by modulation is: Δt=0.5T_s.

T_S is the sampling frequency (equal to switching frequency). The PS-PWM method will also bring about time delay as a result of PWM duty sampling. This delay will reduce the current loop phase margin. If the gain of current loop is not reduced, the current loop may resonate. When the switching frequency is high, carrier modulation will produce less delay component compared to the total delay in the system, and the modulation delay has no significant effect. The carrier phase-shifting is to improve the equivalent switching frequency and the other way round, to reduce the actual switching frequency. When the actual switching frequency is reduced, measures must be taken into consideration to guarantee the fast response of the inverter.

Response speed can be improved by increasing the sampling frequency, as is shown in Fig6.

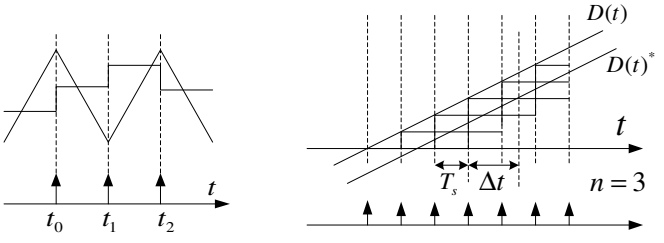


Fig. 6. Repeated sampling for PS-PWM

Update the duty value in the carrier peak point or valley point of each cell won't affect the actual switching frequency and will improve the response speed of inverter simultaneously. Set the sampling cycle to T_s, and then (cascade cell number is n) the switching cycle is $2nT_s$, the equivalent duty cycle after sampling is$D(t)^*$. According to the V-S (voltage and second balance characters) theory:

$$\begin{aligned} D(t)^* &\approx \frac{1}{n}[D(t-\frac{T_s}{2})+D(t-\frac{3T_s}{2})+D(t-\frac{5T_s}{2})\cdots] \\ &= \frac{1}{n}\sum_{k=1}^{n}\{D[t-\frac{(2k-1)T_s}{2}]\} \end{aligned} \tag{3}$$

If the duty cycle is linear during one switch cycle as is shown in Fig6(b), then we can get the approximate formula as follows:

$$\begin{aligned} D(t)^* &\approx D(t-\frac{nT_s}{2}) \\ &= D(t-\frac{T_{switch}}{4}) \end{aligned} \tag{4}$$

According to the duty cycle formula, with sampling frequency up to 2 times of switching frequency, PS-PWM modulation delay is quarter of switching cycle, and it will not decrease as the n increase. So with this modulation method and taking time

delay into account, the equivalent switching frequency is not equal to the real $2n$ times switching frequency.

The time delay always exits with digital control and carrier wave modulation. Low level switching frequency will bring in long time delay. Even if the above modulation strategy was adopted, the time delay of the modulation could be considerable, especially in the occasion of very low level switching frequency.

As Fig7 shows, with PS-PWM strategy, when one stage PWM duty cycle gets updated, for example D_3 at the time t_0, the D_1 keeps constant between time t_0 to t_1. It shows that when one cell's duty cycle is updated, other cells' duty cycles will be kept constant for a predictable time, so these constant duty cycles can be compensated during the time. Then the response speed of CS-APF can be increased, in another word, the modulation time delay can be reduced.

The compensation ΔD is derived according to the theory of V-S(volt and second balance characters). Taking three-stage cascade modulation for example:

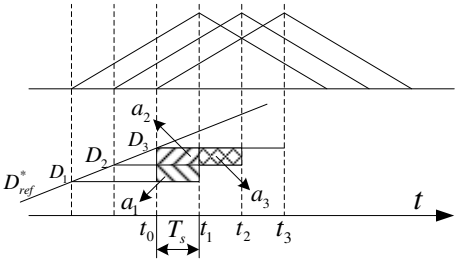


Fig. 7. Duty compensation for PS-PWM

As Fig7 shows, suppose the reference duty value D^*_{ref} is linear during one switching cycle T_{switch}. D_3 will get updated at time t_0; D_1 will be kept constant from t_0 to t_1, and D_2 will be kept constant from t_0 to t_2. Taking the other two stages' duty cycle (D_1, D_2) errors into account, the compensation for each stage can be set as follows:

$$\Delta D_{32} = \frac{2}{3}(D_3 - D_2)\text{'} \quad \Delta D_{31} = \frac{1}{3}(D_3 - D_1) \tag{5}$$

Then define the new duty cycle value $D_3^{'}$:

$$D_3^{'} = D_3 + \Delta D_{32} + \Delta D_{31} = 2D_3 - \frac{2D_2 + D_1}{3} \tag{6}$$

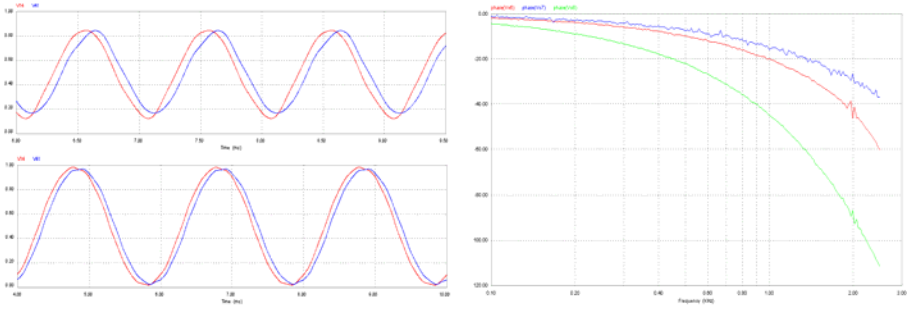
And the other two duty cycles can be derived as follows:

$$\begin{cases} D_3^{'} = 2D_3 - \dfrac{2D_2 + D_1}{3} \\ D_2^{'} = 2D_2 - \dfrac{2D_1 + D_3}{3} \\ D_1^{'} = 2D_1 - \dfrac{2D_3 + D_2}{3} \end{cases} \tag{7}$$

Time delay introduced by carrier modulation will be significantly reduced after compensation. The improved modulation method and the former one are compared through simulation in the following.

The bottom two curves of Fig8(a) show the result of tracking the 500Hz-sinusoidal wave with the two above modulation methods. For clear expression, both signals (output voltage waveforms) shown in the Figure are filtered by filter with same cut-off frequency (1kHz). The red one uses the improved modulation method, and the other uses the usual PS-PWM. It's obvious that the improved method can shorten the time delay a lot, especially in the case of tracking 1kHz-sinusoidal wave (Fig8(a)top ones).

Fig8(b) shows the phase characteristic of three kinds of modulation method after AC sweep. The upper curve is simulated with 12kHz switching frequency and 12kHz sampling frequency, the middle one with the improved PS-PWM method and the lower one with the former PS-PWM method.



(a) (b)

Fig. 8. Comparison of PS-PWM and new PS-PWM

At last, we can derive the common duty compensation formula as follows:

$$D_k^{'} = D_k + \sum_{i=1}^{k-1}[\frac{n+i-k}{n}(D_k - D_i)] + \sum_{i=k+1}^{n}[\frac{n-i}{n}(D_k - D_i)] \tag{8}$$

In the above formula, n is the total cascade stage number, D_k is the k stage PWM duty cycle value, and $D_k^{'}$ is the new value.

3 Simulation Study

A three-stage CS-APF simulation model has been accomplished in the environment of PSIM, and the simulation parameters are as follows:

Network voltage: 3kV; Network equivalent inductance: 0.02mH; Load: three phase rectifier, resistance: 20Ωat start, reduced to 5.72Ω at 0.2 second. Cascade stage number:3, DC-link voltage of each cell: 1200V; DC-link capacitor: 5000uF; Inductance of CS-APF: 0.6mH; Switching frequency:2kHz; Sampling frequency: 12kHz; Discharge resistance: 100Ω; Discharge switching frequency: 1kHz.

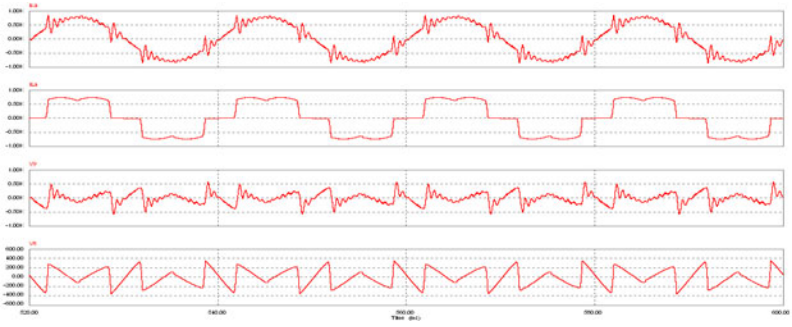


Fig. 9. General PS-PWM method

Fig9 shows static response of the simulation model with general PS-PWM strategy. From top to the bottom, Fig9 shows grid current, load current, CS-APF current and the reference current for the current closed-loop control. It's obvious that resonance exists in the CS-APF current because time-delay during the control and modulation. Only by reducing the gain in the current control loop with the results of lower response, the resonance could be restrained. However, if the modulation strategy is replaced by the novel method, the result will be improved (Fig 10). With the novel PWM strategy, the resonance has been restrained, and it also increases the response speed with lower peak of leakage harmonics. The current ripple is small enough for this application and is equivalent to the one with switching frequency of 12kHz.

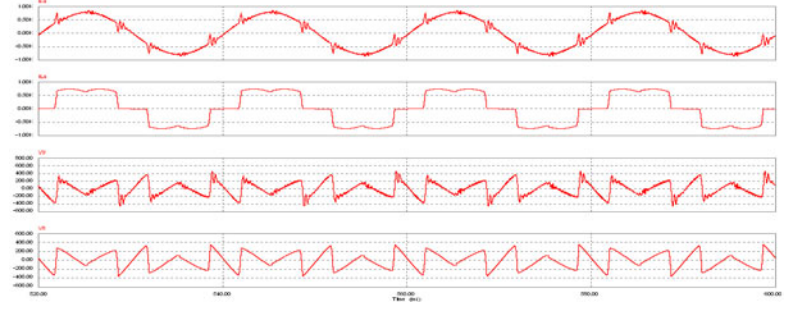


Fig. 10. Novle PS-PWM method

Fig11 shows the dynamic simulation results of the CS-APF with novel PS-PWM method. During the process of the step increase of the load, the CS-APF can also response immediately, and the harmonics on the grid current is kept at low level. For the real power is not separated completely in the short process due to the LPF in the harmonic detection algorithm, the DC-link voltages go lower. But the DC-link capacitors get charged soon and the voltages are kept at the reference level, which proves the effectiveness of the DC-link-voltage control method.

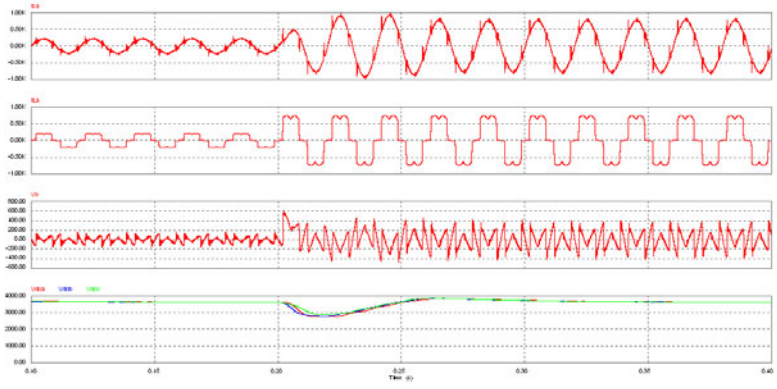


Fig. 11. Dynamic process of CS-APF

4 Conclusion

The PS-PWM method can improve the equivalent switching frequency for cascade inverter, but the time delay brought by PS-PWM is still considerable. The time delay almost keeps constant, when the stage number increases. This paper presents a novel modulation method for cascade active power filter. With the advantages of high response speed and relative high equivalent switch frequency, this modulation method fits for cascade inverter application. At last a simulation model of three-stage CS-APF is implemented, and the simulation results proved the above method is effective.

References

1. Bhim, S., Kamal, A.H., Ambrish, C.: A Review of Active Filters for Power Quality Improvement. IEEE Transactions on Industry Electronics 46(5), 960–971 (1999)
2. Hirofumi, A.: New Trends in Active Filters for Power Condition. IEEE Transactions on Industry Application 32(6), 1312–1322 (1996)
3. Luo, A., Shuai, Z.K., Zhu, W.J., et al.: Combined System for Harmonic Suppression and Reactive Power Compensation. IEEE Transactions on Industrial Electronics 56(2), 418–428 (2009)
4. Song, Q., Liu, W.H.: Control of a Cascade STATCOM With Star Configuration Under Unbalanced Conditions. IEEE Transactions on Power Electronics 24(1), 45–58 (2009)
5. Zang, C.Y., Pei, Z.J., He, J.J., et al.: Comparison and Analysis on Common Modulation. In: Strategies for the Cascaded Multilevel STATCOM PEDS, pp. 1439–1442 (2009)
6. Wang, G., Li, Y.D., You, X.J.: A Novel Control Algorithm for Cascade Shunt Active Power Filter. In: 35th Annual IEEE Power Electronics Specialists Conference, pp. 771–775 (2004)
7. Chen, J.L., Yin, Z.Z., Wang, P., et al.: Capacitor Voltage Balancing Control of Cascaded Multilevel Inverter for High-power Active Power Filters. In: DRPT 2008, pp. 1683–1687 (2008)

8. Zhou, G., Wu, B., Xu, D.: Direct Power Control of A MultiIevel Inverter Based Active Power Filter. In: IEEE ICIT, pp. 498–503 (2004)
9. Chen, Y.G., Wang, P., Li, Y.H., et al.: A Novel Hybrid Modulation Method for Cascaded H-bridge Active Power Filter. In: EPE-PEMC, pp. 1981–1986 (2008)
10. Wang, L.Q., Wu, W.Y.: Shunt Active Power Filter with Sample Time Staggered Space Vector Modulation Based Cascade Multilevel Converters. In: IPEMC, pp. 1–4 (2006)
11. Ahmed, M.M., Stephen, J.F., Andrew, J.C., et al.: Three-Phase, Three-Wire, Five-Level Cascaded Shunt Active Filter for Power Conditioning, Using Two Different Space Vector Modulation Techniques. IEEE Transactions on Power Delivery 22(4), 2349–2361 (2007)

An Integration Method for Edge Detection

Su-Qun Cao[1], Wei-Min Chen[2], and Hong Zhang[3]

[1] Faculty of Mechanical Engineering
Huaiyin Institute of Technology
Huai'an 223003, P.R. China
caosuqun@126.com
[2] Computer Technology & Artistic Design Department
Jiangsu Vocational and Technical College of Finance & Economics
Huai'an 223003, P.R. China
jshacwm@163.com
[3] Department of Personnel
Huaiyin Institute of Technology
Huai'an 223003, P.R. China
zh0822@sina.com

Abstract. Edge detection plays an important role in many fields. Many edge detection operators such as Roberts edge operator, Sobel edge operator and Prewitt edge operator have been proposed to solve edge detection problems. Each of them has its own characteristic. But it is difficult to select a suitable operator according to a specific problem. In this paper, a novel method for edge detection is presented. Firstly, traditional edge detection operators are used in a gray image to obtain the results which are viewed as the feature vectors of the image data. Then kernel-based fuzzy c-means clustering algorithm is applied in these feature vectors to detect out the edge points adaptively. Experimental results show the proposed method's efficiency.

Keywords: Edge detection operator, Fuzzy clustering, Roberts operator, Sobel operator, Prewitt operator.

1 Introduction

Edge detection plays an important role in many fields such as image analysis, computer vision and automated driving et al. [1] Edge points are pixels at which abrupt gray-level changes occur because of changes in surface orientation, depth or physical properties of materials. The aim of edge detection is providing a meaningful description of identifying and locating sharp discontinuities in an image.

There are many edge detection operators available. Commonly used operators for edge detection are mainly [2]: Roberts edge operator, Sobel edge operator and Prewitt edge operator. Each of them is designed to be sensitive to certain types of edges. In addition, some people put forward new edge detection algorithms. Tuba Sirin

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 243–249.
springerlink.com

et al. presents an edge detection method which is based on the use of the clustering algorithms and a gray scale edge detector [3]. Wenshuo Gao et al. [4] proposes a method which combines Sobel edge detection operator and soft-threshold wavelet denoising to do edge detection on images which include white Gaussian noises. All these methods mentioned above are used only single edge detection operator.

In this paper, an integration method of edge detection operator with kernel-based fuzzy c-means clustering algorithm (KFCM) is presented. The rest of this paper is organized as in the following; commonly used edge detection operators are discussed in the next section. In section 3 KFCM used in the proposed method is discussed. Section 4 describes the proposed algorithm. The comparative experimental results are discussed in Section 5. Conclusions are given in the last section.

2 Edge Detection Operators

2.1 Roberts Edge Operator

The Roberts operator is very quick to compute 2-D spatial gradient measurement on an image. It highlights regions of high spatial frequency which often correspond to edges. In its most common usage, the input to the operator is a grayscale image, as is the output. Pixel values at each point in the output represent the estimated absolute magnitude of the spatial gradient of the input image at that point. Only four input pixels need to be examined to determine the value of each output pixel, and only subtractions and additions are used in the calculation. In addition there are no parameters to set. Its main disadvantages are that since it uses such a small kernel, it is very sensitive to noise.

2.2 Sobel Edge Operator

The Sobel operator is a discrete differentiation operator, computing an approximation of the gradient of the image intensity function to find the edge. If one takes the derivative of the intensity value across the image and find points where the derivative is maximum, the edge could be found. The Sobel operator is based on convolving the image with a small, separable, and integer valued filter in horizontal and vertical direction and is therefore relatively inexpensive in terms of computations. The result of the Sobel operator is either the corresponding gradient vector or the norm of this vector which is relatively crude, in particular for high frequency variations in the image.

2.3 Prewitt Edge Operator

The Prewitt edge detector is an appropriate way to estimate the magnitude and orientation of an edge. It calculates the maximum response of a set of convolution

kernels to find the local edge orientation for each pixel. The set of kernels is limited to eight possible orientations. However experience shows that the Prewitt edge detector has a major drawback of being very sensitive to noise. In addition, the size of the kernel filter and coefficients are fixed and sometimes it cannot be adapted to a given image.

Thus, it is necessary to provide a robust edge detection algorithm to help distinguish valid image contents from visual artifacts introduced by noise.

3 Kernel-Based Fuzzy c-Means Clustering Algorithm (KFCM)

The fuzzy c-means (FCM) algorithm was introduced by J.C.Bezdek [5]. Given $X = \{x_1, x_2, ..., x_n\}$ where x_k in R^s, the idea of FCM is using the weights that minimize the total weighted mean-square error:

$$J_m(U,v) = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^m \left\| x_k - v_i \right\|^2 \tag{1}$$

Here c is the number of clusters, n is the number of data points, and u_{ik} is the membership of x_k in class i takes value in the interval [0,1] such that

$$\sum_{i=1}^{c} u_{ik} = 1 \tag{2}$$

for all k.

Dao-Qiang Zhang et al. [6] introduced kernel methods to FCM and proposed kernel-based fuzzy c-means clustering (KFCM) algorithm which is a robust clustering approach.

As we know, a pattern in the original input data space X can be mapped into the higher dimensional feature space F through the nonlinear mapping function Φ.

$$\Phi : X = (x_1, x_2, ..., x_n) \to \Phi(X) = (\Phi(x_1), ..., \Phi(x_N)) \tag{3}$$

The objective function of KFCM is constructed as follows:

$$J_m(U,v) = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^m \left\| \Phi(x_k) - \Phi(v_i) \right\|^2 \tag{4}$$

Scalar product calculation in input space is transformed into kernel function calculation by nonlinear mapping

$$\left\langle x_i \cdot x_j \right\rangle \to \left\langle \Phi(x_i) \cdot \Phi(x_j) \right\rangle = K(x_i, x_j) = K_{ij} \tag{5}$$

Thus, we have

$$\left\| \Phi(x_k) - \Phi(v_i) \right\|^2 = K(x_k, x_k) + K(v_i, v_i) - 2K(x_k, v_i) \tag{6}$$

To minimize J_m under the constraint of Eq.(2), the alternate iterative equations can be obtained as follows:

$$u_{ik} = \frac{(1/(K(x_k, x_k) + K(v_i, v_i) - 2K(x_k, v_i)))^{1/(m-1)}}{\sum_{j=1}^{c}(1/(K(x_k, x_k) + K(v_j, v_j) - 2K(x_k, v_j)))^{1/(m-1)}} \tag{7}$$

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^m \tilde{K}(x_k, v_i) x_k}{\sum_{k=1}^{n} u_{ik}^m \tilde{K}(x_k, v_i)} \tag{8}$$

Here we use Gaussian kernel. It can be defined as follows:

$$\tilde{K}(x_k, v_i) = \exp(-\frac{\|x_k - v_i\|^2}{2\sigma^2}) \tag{9}$$

KFCM algorithm can be summarized in the following steps:

Step 1: Fix c and select parameters, $t_{\max}$, $m > 1$ and $\varepsilon > 0$ for some positive constant.

Step 2: Initialize the memberships u_{ik}^0.

Step 3: For $t = 1, 2, \ldots, t_{\max}$, do:

(a) update all v_i with Eq.(8);

(b) update all memberships u_{ik}^t with Eq.(7);

(c) if $\max_{i,k} \left| u_{ik}^t - u_{ik}^{t-1} \right| \le \varepsilon$, stop;

End.

4 Proposed Edge Detection Method

A two-stage edge detection operator integration method is applied on the images. Novelty in our study is the use of KFCM to combine three edge detection operators. The stages of our algorithm as follow:

Stage 1: Gray scale edge detection by using Roberts, Sobel and Prewitt operators.
Stage 2: Clustering of the results of the edge detection operator by using KFCM.

Firstly, a pixel point in a gray image is regarded as a data sample, and its gray values which are processed by Robert operator, Sobel operator and Prewitt operator make up of the feature vectors of this data sample. In this way a data set with three-dimensional features can be obtained. Then the KFCM clustering algorithm is used in this data set, the edge points can be detected out adaptively.

5 Experimental Results

Edge detections of traditional edge detection operators and the proposed method are performed on the gear image. Fig. 1 is the gear image and Fig. 2 is edge detection results of Fig. 1. Fig. 3 is the gear image with Gaussian noises and Fig. 4 is edge detection results of Fig. 3. As is shown in the experimental results, the proposed method yields the best results.

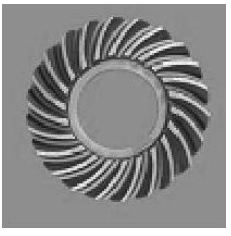


Fig. 1. Gear image

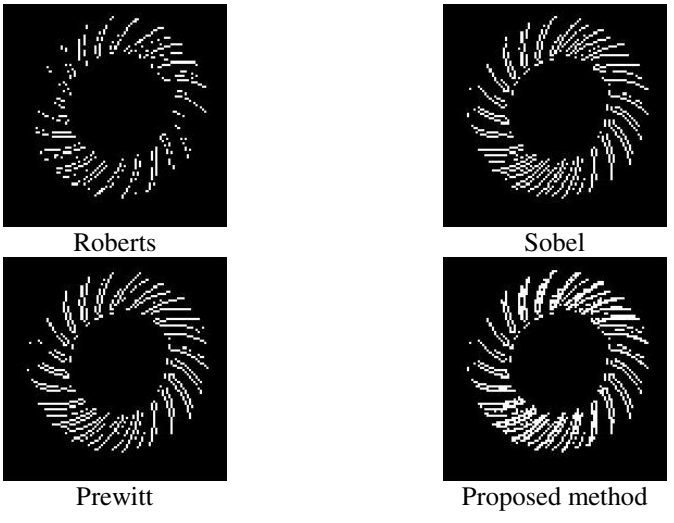


Fig. 2. Edge detection results of gear image

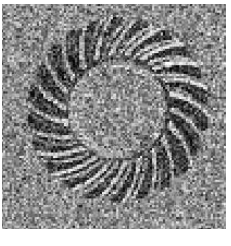


Fig. 3. Noised gear image

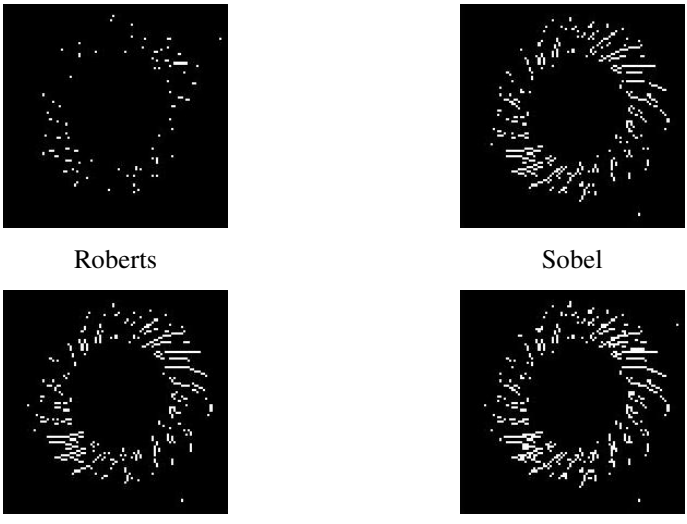


Fig. 4. Edge detection results of noised gear image

6 Conclusions

A novel method for edge detection is presented. It uses kernel-based fuzzy c-means clustering algorithm to combine Roberts, Sobel and Prewitt edge detection operators to detect out the edge points adaptively. The resulting edge images show that the performance of the proposed integration method is superior to the single edge detection operator. In the future, we will also explore our previously proposed clustering algorithm [7-8] to integrate the edge detector operators.

Acknowledgements. This work is supported by National Natural Science Foundation of China (Grant No.60773206/F020508), Natural Science Major Basic Research Program of Jiangsu Colleges and Universities, China (09KJA460001, 10KJA460004), the International Technology Cooperation Project of Huaian City, China (HG004), Youth Foundation of Huaiyin Institute of Technology, China (HGQN0701) and Huaian Science and Technique Program (HAG2010039, HAG2010062).

References

1. Application of Edge Detection in Automated Driving Systems, http://homepages.cae.wisc.edu/~ece533/project/f06/kratzke_rpt.doc
2. Algorithms for Edge Detection, http://www.ee.sunysb.edu/~cvl/ese558/s2005/Reports/Srikanth%20Rangarajan/submission.doc

3. Sirin, T., Saglam, M.I., Erer, I., et al.: Edge Detection in Images Using Clustering Algorithms. In: 4th WSEAS International Conference on Telecommunications and Informatics, Prague, Czech Republic (2005)
4. Gao, W.-S., Zhang, X.-G., Yang, L., et al.: An improved Sobel edge detection. In: 3rd IEEE International Conference on Computer Science and Information Technology, pp. 67–71. IEEE Press, Chengdu (2010)
5. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York (1981)
6. Zhang, D.-Q., Chen, S.-C.: A Novel Kernelized Fuzzy C-means Algorithm with Application in Medical Image Segmentation. Artificial Intelligence in Medicine 32, 37–50 (2004)
7. Cao, S.-Q., Wang, S.-T., Bo, Y.-F., et al.: A Novel Semi-fuzzy Clustering Algorithm with Application in Image Texture Segmentation. In: International Symposium on Distributed Computing and Applications for Business, Engineering and Sciences, Dalian, China, pp. 214–218 (2008)
8. Cao, S.-Q., Wang, S.-T., Chen, X.-F., et al.: Fuzzy Fisher Criterion based Semi-fuzzy Clustering Algorithm. Journal of Electronics & Information Technology 30, 2162–2165 (2008)

A New Approach for Fingerprint Matching Based on Singular Points Detection

Rohollah Moosavi Tayebi[1], Samaneh Mazaheri[2], and Bahram Sadeghi Bigham[2]

[1] Islamic Azad University, Shahr-e-Qods Branch,Tehran, Iran
Moosavi_tayebi@shahryariau.ac.ir
[2] Institute for Advanced Studies in Basic Science (IASBS)
Department of Computer Science & IT, RoboCG Lab, Zanjan, Iran
{S.mazaheri,b_sadeghi_b}@iasbs.ac.ir

Abstract. In this paper, we introduce a new approach, using combined feature for fingerprint matching. Our proposed approach is the idea that this feature combines the information of every two singular points such as core and delta, of the image and takes the ridges structure between them into account. The combined feature is invariant with respect to the global transformations such as rotation and transformation. The proposed recognition is performed in three steps: singular points extraction, combined features extraction and matching. Experimental results on FVC2006 show efficiency and accuracy of the proposed method.

Keywords: Image processing, fingerprint matching, singular point extraction, core and delta detection.

1 Introduction

The fingerprint -a biological feature of humankind- has many particular properties such as uniqueness, stableness, and inseparability from the host. It has been used for personal verification for more than one hundred years, and is the most widely used biological recognition technique today. Among the many current biometric technologies, fingerprint matching is the oldest and the most popular method widely used in different commercial and security applications. More than 100 features are defined for representation of a fingerprint. Singular points such as minutiae, Core and Delta are shown in Figure 1.

Among the features, ridge ending and ridge bifurcation, shown in Figure 1, are the most commonly used features which are named minutiae points. Several minutiae-based approaches have been proposed to match fingerprints [1], [2], [3], [4]. Some of them are based on singular points [5] and some others are independent of singular points [1], [2]. There are two major types of features that are used in fingerprint matching: local and global features. Local features such as minutiae contain the

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 251–260.
springerlink.com

information in local area and invariant with respect to global transformation such as rotation and transformation. On the other hand, global features, such as number, type and position of singularities, spatial relationship and geometrical attributes of ridge lines, size and shape of the fingertips are characterized by the attributes that capture the global spatial relationships of a fingerprint [6]. It has been showed that the geometric deformation can be more easily controlled than global deformation [7]. In [1], [2], a secondary feature has been introduced using relative distance, radial angle and minutia orientation. Jiang and Yau [8] have used this relative information along with ridge count and minutiae type.

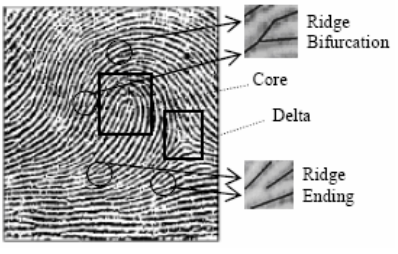


Fig. 1. Singular Points (Core and Delta) and Minutiae (ridge ending and ridge bifurcation).

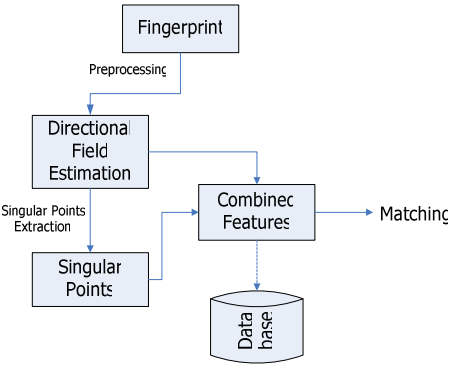


Fig. 2. The proposed fingerprint matching system.

In this paper, we introduce a combined feature for fingerprint matching. The combined features are derived from singular points features. We use the minutiae-based representation of a fingerprint. The combined features are extracted from image obtained by preprocessing steps. The feature, introduced in Section 3, is defined for every two points, delta and core. Figure 2 shows the proposed matching system. Three main steps of our proposed method are: 1) Singular points extraction 2) Combined feature extraction 3) Matching

In this paper, it is assumed that a fingerprint with no distortion is available. The combined features are extracted from the clear fingerprint. In Section 2, we extract singular points. The new combined feature is described in section 3. Matching is described in section 4. Experimental results on FVC2006 are presented in Section 5. The paper is concluded in Section 6.

2 Singular Points

In this section, we use an accurate method to singular points detection [9]. In [9], a method is presented for using directional fields of fingerprints images and orientation in extraction of singular points. The main results are repeated here.

2.1 Directional Field Estimation

The directional field describes the coarse structure, or basic shape, of a fingerprint [9]. The directional field is defined as the local orientation of the ridge-valley structures. This is for instance used for classification of fingerprints. In [10], a method is presented for the estimation of a high-resolution directional field. The main results are repeated here. The method is based on the gradient vector $[G_x(x,y)\ G_y(x,y)]^T$ of the grayscale image $I(x,y)$, which is defined by:

$$\begin{pmatrix} G_x(x,y) \\ G_y(x,y) \end{pmatrix} = \nabla I(x,y) = \begin{pmatrix} \frac{\delta I(x,y)}{\delta x} \\ \frac{\delta I(x,y)}{\delta y} \end{pmatrix} \tag{1}$$

The directional field is, in principle, perpendicular to the gradients. However, the gradients are orientations at pixel scale, while the directional field describes the orientation of the ridge-valley structures. This requires a much coarser scale in which local fluctuations do not have influence. Therefore, the directional field can be derived from the gradients by performing some averaging operation on the gradients, involving pixels in some neighborhood [11].

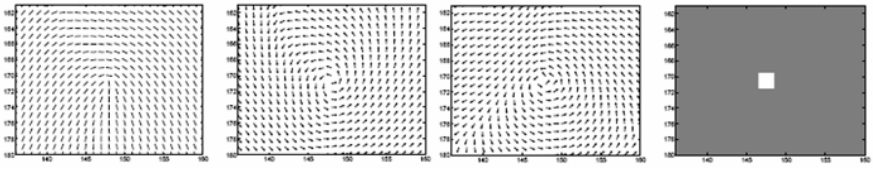
Gradients cannot simply be averaged in some local neighborhood, since opposite gradient vectors will then cancel each other, although they indicate the same ridge-valley orientation. A solution to this problem is to double the angles of the gradient vectors before averaging. Then, opposite gradient vectors will point in the same direction and therefore will reinforce each other, while perpendicular gradients will cancel. After averaging, the gradient vectors have to be converted back to their single-angle representation. The main ridge-valley orientation is perpendicular to the direction of the average gradient vector. This method was proposed by [12] and was adopted in some way for the estimation of the directional field of fingerprints by various researchers.

In the version of algorithm used in this paper, not only the angle of the gradients is doubled, but also the length of the gradient vectors is squared, as if the gradient vectors are considered as complex numbers that are squared. This has the effect that strong orientations have a higher vote in the average orientation than weaker orientations and this approach results in the cleanest expressions.

Another difference of this method is that we do not estimate the average directional field for a number of blocks in the image. Instead, the directional field is estimated for each pixel in the image using a Gaussian window W for averaging. Using $G_{xx} = \sum_w G^2_x$, $G_{yy} = \sum_w G^2_y$ and $G_{xy} = \sum_w G_x G_y$. The average gradient direction θ, with $-\frac{1}{2}\pi < \theta \le \frac{1}{2}\pi$ is given by: $\theta = \frac{1}{2}\angle(G_{xx} - G_{yy}, 2G_{xy})$ (2)

2.2 Singular Points Extraction

Singular points are the discontinuities in the directional field. They are clearly visible in a fingerprint image, as can be seen from Figure 1, which shows examples of the two different types of singular points that are called core and delta. Singular points can for instance be used for the registration (alignment) of two fingerprints of the same finger. Many different algorithms for singular point extraction are known from literature. The method that is used here is based on the Poincare index [13]. It can be explained using Figure 3(a). Following a counter-clockwise closed contour in the directional field around a core results in a cumulative change of π in the orientation, while carrying out this procedure around a delta results in $-\pi$. On the other hand, when applied to a location that does not contain a singular point, the cumulative orientation change will be zero.



(a) Directional field (b) Squared Directional field (c) Gradient (d) Rotation

Fig. 3. Processing Steps in the Extraction of a Core.

The method is capable of detecting singular points that are located only a few pixels apart. In the rest of this section, all calculations are made for the case of a core. It is left to the reader to adapt them to the case of a delta. First, the squared directional field is taken. This eliminates the transition of π which is encountered in the directional field between the orientations $\theta = \frac{1}{2}\pi$ and $\theta = -\frac{1}{2}\pi$. As a result, the Poincare index is doubled. The orientation of the squared directional field is depicted in Figure 3(b) for the area around the core example. Instead of summing the changes in orientation, it is possible to sum the gradients of the squared orientation as well. The gradient vector J can be efficiently pre-calculated for the entire image by:

$$\begin{bmatrix} J_x(x,y) \\ J_y(x,y) \end{bmatrix} = \nabla 2\theta(x,y) = \begin{bmatrix} \frac{\delta 2\theta(x,y)}{\delta x} \\ \frac{\delta 2\theta(x,y)}{\delta y} \end{bmatrix} \tag{3}$$

In the calculation of the discrete version of this gradient, both components of J should be calculated 'modulo 2π', such that they are always between $-\pi$ and π. This makes the transition from $2\theta = -\pi$ to $2\theta = \pi$ continuous or, in other words, the orientation

is considered to be cyclic. The gradient vectors of the squared orientation around the core are shown in Figure 3(c). The next step is the application of Green's Theorem, which states that instead of calculating a closed line-integral over a vector field, the surface integral over the rotation of this vector field can be calculated. This theorem is applied to the summation of the gradients of the squared orientation over the contour:

$$\sum_{\partial A} J_x + J_y = \sum_{A} rot[J_x J_y]^T = \sum_{A} \left(\frac{\delta J_y}{\delta J_x} - \frac{\delta J_x}{\delta J_y} \right) \tag{4}$$

Where A is the area and ∂A is the contour. Since the directional field is assumed to be smooth, A can be taken as small as a square of 1×1 pixel. This results in a very efficient method for computation of the Poincare index. Application of the proposed method will indeed lead to the desired singular point locations. Unlike all other singular point extraction methods, a core results in a Poincare index of 2π a delta in -2π while the index for the rest of the image is exactly equal to 0. This is illustrated in Figure 3(d) for the core example and the area around it.

2.3 Orientation of Singular Points

The last stage of singular points detection is the estimation of the orientations φ of the extracted singular points. The method that is described here makes use of the squared gradient vectors in the neighborhood of a singular point, both for the image to be analyzed and for the reference singular point. The averaged squared gradients of the core, repeated in Figure 4(a), ideally look like the reference model in Figure 4(b). For a core at $(x, y) = (0, 0)$, this reference model are given by:

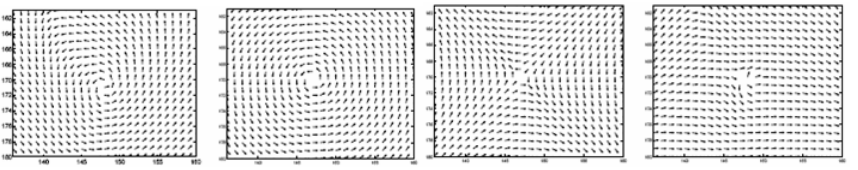
$$Core_{Ref} = \frac{(y, -x)}{\sqrt{x^2 + y^2}} \tag{5}$$

The model of a core that has rotated over an angle φ is given by a reference model with all its components multiplied by $e^{j\varphi}$. $Core_{\varphi} = Core_{Ref} \cdot e^{j\varphi}$ (6)

This property is used for the estimation of the orientation of the core. The orientation of the core, with respect to the reference model, is found by taking the inner product of the estimated squared gradient data and the complex conjugated of the reference model, depicted in Figure 4(c). Then, it is divided by the number matrix elements N, and the angle is taken.

$$\hat{\varphi} = \angle \frac{1}{N} \sum_{x,y} Core^*_{Ref}(x, y) Core_{Obs}(x, y) \tag{7}$$

For a delta, a related reference model can be made. Because of the symmetries in a delta, its relative orientation with respect to the reference model is given by one third of the angle resulting from Equation 7, which was used for a core.



(a) Squared Directional field (b) Reference Core (c) Conjugated (d) Orientation

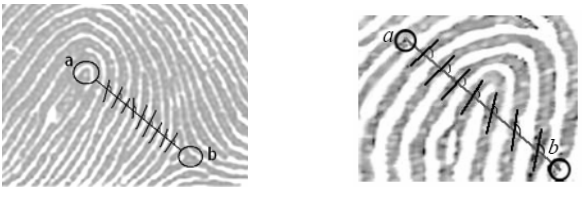
Fig. 4. Processing Steps in the Calculation of the Orientation of a Core.

3 Combined Features

The combined features are derived from singular points and localized information between them. The combined feature is defined for every two singular points in the skeletonized image. We use ridge count between them along with relative orientation of each middle ridge and the line connects the two singular points. Ridge count is an abstract measurement of the distances between any two points in a fingerprint image [14]. Let a and b be two points, core and delta in a fingerprint; then the ridge count between a and b is the number of ridges intersected by segment ab. Figure 5 illustrates the combined feature. For two points, core and delta $m_i\,(x_i, y_i, \theta_i, type_i)$ and $m_j\,(x_j, y_j, \theta_j, type_j)$ we contrast a vector sf_{ij} as follows:

$$sf_{ij} = [x_i, y_i, x_j, y_j, rc_{ij}, \theta_{ij}^1, \theta_{ij}^2, ..., \theta_{ij}^{rc}] \tag{8}$$

Where (x_i, y_i) and (x_j, y_j) are location coordinates of m_i and m_j rc_{ij} is the ridge count between m_i and m_j θ_{ij}^k is the kth relative orientation of middle ridges and the line which connects m_i and m_j θ_{ij}^k is the clockwise rotation of $m_1 m_2$ to middle ridge. The ridge count and the relative orientation is calculated while scan the skeletonized image from m_i to m_j .



(a) (b)

Fig. 5. (a) The combined feature on singular points, Core and Delta (b) the combined feature on two minutiae.

Since different impressions of a fingerprint cause a variety of orientation, we considered some regions of orientations as follows [15]:

$$R_1 = 0 \le \theta_i < \frac{\pi}{4} \quad (9) \quad , \quad R_2 = \frac{\pi}{4} \le \theta_i < \frac{\pi}{2} \quad (10)$$

$$R_3 = \frac{\pi}{2} \le \theta_i < \frac{3\pi}{4} \quad (11) \quad , \quad R_4 = \frac{3\pi}{4} \le \theta_i < \pi \quad (12)$$

In order to increase the stability of the combined feature a tolerance area is considered. R_1, R_2, R_3 and R_4 with constant tolerance area are shown in Figure 6.

We consider a constant interval for tolerance area.

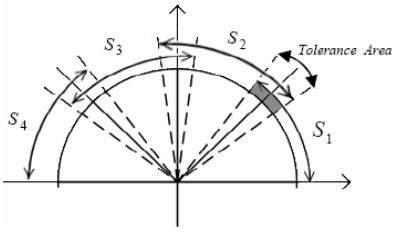


Fig. 6. The Orientation Regions.

$$S_1 : 0 \le \theta_i < \frac{\pi}{4} + \gamma \quad (13) \quad , \quad S_2 : \frac{\pi}{4} - \gamma \le \theta_i < \frac{\pi}{2} + \gamma \quad (14)$$

$$S_3 : \frac{\pi}{2} - \gamma \le \theta_i < \frac{3\pi}{4} + \gamma \quad (15) \quad , \quad S_4 : \frac{3\pi}{4} - \gamma \le \theta_i < \pi \quad (16)$$

A set of combined features, Q, is stored in template database for each fingerprint. For an image with n singular points the template size is $\frac{n(n-1)}{2}$.

$$Q = \{q_k = sf_{ij} \mid m_i \ne m_j, \quad sf_{ij} = < x_i, y_i, x_j, y_j, rc_{ij}, \theta_{ij}^1, ...\theta_{ij}^{rc} >\}$$

$$k = 1, 2, ... \frac{n(n-1)}{2}, i = 1, 2, ...n, j = 1, 2, ..., n \quad (17)$$

4 Matching

The larger *rc* the combined feature has, the more distinctive it is. The combined feature with the largest *rc* is selected as a reference point in input fingerprint image and the best matched combined feature with the same *rc* is selected as a reference point in template fingerprint images. The input fingerprint aligns with the template fingerprint according to this reference points. The similarity score is a two element

vector $< Sc_1, Sc_2 >$ that Sc_1 corresponds to the number of matched combined features and Sc_2 is the sum of Euclidean distance between every two matched combined features in input image and template images.

Let I be the input fingerprint image, a set of candidate fingerprint template, MF, is provided according to Sc_1.

$$MF = \{Q_i \mid Sc_1(Q_i, I) > \alpha\} \tag{18}$$

$$Sc_1(Q, I) = \frac{\#of \quad matched \quad combined \quad features(Q, I)}{\#of \quad total \quad combined \quad features \quad of \quad input \quad image} \tag{19}$$

$$Sc_2 = \sum_{\substack{i,j=1 \\ q_i(5)=q_j(5)}}^{\frac{n(n-1)}{2}} \sqrt{\left|q_i - q_j\right|^2} \tag{20}$$

Where $q(5)$ is the fifth element of the combined feature corresponds to $rc.q_i(5) = q_j(5)$ in formula (16) means that Sc_2 is calculated for the combined features of the input image and template image with the same rc.

The template that has the largest value of Sc_1 is selected as the matched template. For two templates with the same Sc_1, the second element, Sc_2 is used to make the final decision. The template that has the smallest Sc_2 is selected as a result.

5 Experimental Results

In the experience, $\gamma = \frac{\pi}{36}$ and $\alpha = 0.9$ have been used (See (14)-(17) and (19)). The Normalized Matching Score are shown in Figure 7. The experiments reported in this paper have been conducted on the public domain collection of fingerprint images, DB3 in FVC2006. It comprises 800 fingerprint images of size 300×300 pixels captured at a resolution of 500 dpi, from 100 fingers (eight impressions per finger). We conducted a set of experiments on DB3 meant to compare our algorithm with the approach proposed in [16], which matches fingerprints based on both the local and global structures of minutiae. For convenience we label the method in [16] as II. The algorithm II uses the local minutia structures for registration and computes matching score based on the similarity level of corresponding local structures. Matching experiments have been performed on DB3 using algorithms I and II. We have performed the two algorithms on DB3 and obtained the results expressed in terms of ROC curves, as shown in Figure 8. The results demonstrate that our method is more effective than algorithmII.

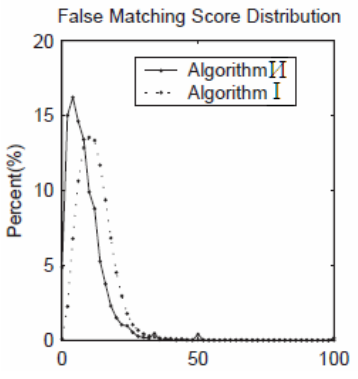


Fig. 7. Normalized Matching Score.

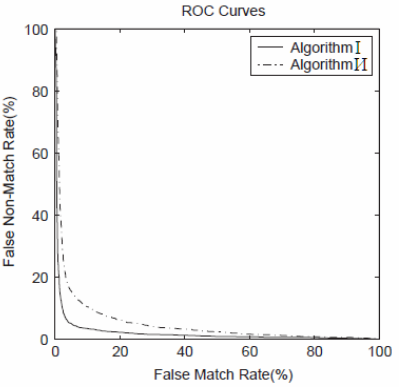


Fig. 8. ROC-Curves on DB3 obtained with our matching method (solid line) and the algorithm И (dash-dot line).

6 Conclusion

In this paper, we proposed a new localized feature based on singular points detection. With using these singular points of the fingerprint we apply our approach named combined feature. This feature combines the information of two singular points and the ridges structure between them. The combined features have the following advantages: 1) These are generated from singular points, so this method can be adopted for existing applications. 2) These are invariant with respect to translation and rotation. 3) Since reference point detection includes a search for the largest rc, the alignment of two fingerprints is simple and fast. 4) The experimental results show that the proposed method is efficient and suitable for fingerprint matching.

Acknowledgments. The authors are highly grateful to Dr.Mansoor Davoodi Monfared, Hamzeh Khazaei and Amin Gheibi for their helpful comments and suggestions, which help us very much.

References

1. Jea, T., Govindaraju, V.: A Minutia-based partial fingerprint recognition system. Pattern Recognition 38, 1672–1684 (2005)
2. Jea, T., Govindaraju, V.: Partial Fingerprint Recognition Based on Localized Features and Matching. In: Proceedings of Biometric Consortium Conference, Crystal City, VA (2005)
3. Jea, T., Chavan, V.K., Govindaraju, V., Schneider, J.K.: Security and matching of partial fingerprint recognition systems. In: Proceeding of SPIE, vol. 5404, pp. 39–50 (2004)
4. Ko, K.H., Maekawa, T., Patrikalakis, N.M.: Algorithms for optimal partial matching of free-form objects with scaling effects. Graphical Models 67, 120–148 (2005)
5. Chan, K.C., Moon, Y.S., Cheng, P.S.: Fast Fingerprint Using Subregions of Fingerprint Images. IEEE Transactions on Circuits and Systems for Video Technology 14(1) (2004)

6. Wegstein, J.H.: An automated fingerprint identification system", U.S. Government Publication, U.S. Department of Commerce, National Bureau of Standards, Washington DC (1982)
7. Jain, A.K., Hong, L., Pankanti, S., Bolle, R.: An identity authentication system using fingerprints. Proceeding of the IEEE 85(9), 1365–1388 (1997)
8. Luo, X., Tian, X.J., Wu, Y.: A minutia matching algorithm in fingerprint verification. In: Proceedings of the International Conference on Pattern Recognition, pp. 833–836 (2000)
9. Bazen, A.M., Gerez, S.H.: Extraction of Singular Points from Directional Fields of Fingerprints. In: Mobile Communications in Perspective, Annual CTIT Workshop, Enschede, The Netherlands (Febraury 2001)
10. Bazen, A.M., Gerez, S.H.: Directional field computation for fingerprints based on the principal component analysis of local gradients. In: Proceedings of ProRISC2000, 11th Annual Workshop on Circuits and Systems, The Netherlands (November 2000)
11. Lindeberg, T.: Scale-Space Theory in Computer Vision. Kluwer Academic Publishers, Boston (1994)
12. Kass, M., Witkin, A.: Analyzing oriented patterns. Computer Vision, Graphics, and Image Processing 37(3), 362–385 (1987)
13. Kawagoe, M., Tojo, A.: Fingerprint pattern classification. Pattern Recognition 17(3), 295–303 (1984)
14. Maltoni, D., Maio, D., Jain, A.K., Prabhakar, S.: Handbook of Fingerprint Recognition. Springer, New York (2003)
15. Abdoos, M., Mozayani, N.: A combined feature for partial fingerprint recognition. In: Proceedings of the 2007 International Conference on Computational Intelligence and Security (2007)
16. Jiang, X., Yau, W.Y.: Fingerprint minutiae matching based on the local and global structures. In: Proceedings of the 15th International Conference on Pattern Recognition, vol. 2, pp. 1038–1041 (2000)

On the Enhancement of TCP Protocol

Seifedine Kadry

Faculty of General Education
American University of the Middle East
220 DASMAN, 15453 Kuwait
Seifedine.kadry@aum.edu.kw, skadry@gmail.com

Abstract. The Transmission Control Protocol (TCP) is the most popular transport layer protocol for the internet. Congestion Control is used to increase the congestion window size if there is additional bandwidth on the network, and decrease the congestion window size when there is congestion. This paper uses a classic TCP which we called Robust TCP with an accurate algorithm of congestion detection in order to improve the performance of TCP. Our TCP Robust only reacts when it receives an ECN (Explicit Congestion Notification) mark. The evaluation result shows a good performance in the terms of drop ratio and throughput.

Keywords: Congestion Control, TCP, ECN, Implicit Congestion Notification.

1 Introduction

TCP is a connection-oriented, end-to-end reliable protocol designed to fit into a layered hierarchy of protocols which support multi-network applications. Congestion events in communication networks leads to packet losses, and it's well known that these losses occur in burst.TCP congestion control involves two tasks:

1. Detect congestion
2. Limit Transmission rate

To achieve good performance and obtain a Robust TCP, it is necessary and important to control network congestion, by limiting the sending rate and regulating the size of congestion window (Cwnd) after the detection of congestion.TCP congestion control operates in a closed loop that infers network conditions and reacts accordingly by means of losses. A negative return is due to a loss of a segment which can be translated by decreasing the flow from the source through a reduction in the size of window control. TCP considers loss of a segment as a congestion in the network, the detection of this loss can be done in several ways: Timeout, Three Duplicate ACKs (Fast retransmit) and by receiving a partial ACK.

The state is:

- If Packet Loss or congestion event =>TCP decreases Cwnd.
- All is well and no congestion in the network, i.e., TCP increases Cwnd.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 261–270.
springerlink.com

At all cases, loss indication should be done with accuracy because it may lead to false indications like: *Spurious retransmission.*

Spurious timeout occurs when a non lost packet is retransmitted due to a sudden RTT (Round Trip Time) increase (hand over, high delay, variability, rerouting . .) which implies to an expiration of the retransmission timer set with a previous and thus outdated RTT value. This effect is known to be the root cause of spurious retransmission. The function of the congestion control is an essential element to the stability of the internet. Indeed, TCP congestion control reduces the flow when it detects a loss in the network. Therefore, it is important to be accurate in the loss detection to improve the performance of TCP. A congestion event (or loss event) corresponds to one or several losses (or in the context of ECN: at least one acknowledgment path with an ECN-echo) occurring in one TCP window during one current RTT period, it means that a congestion event begins when the first loss occurs and finishes one RTT later. In this paper, we propose a congestion detection algorithm that is realized independently of the TCP code. To improve the TCP by reducing the Cwnd, we aim to illustrate the feasibility of the concept by demonstrating that we can both obtain similar performances and also improve the accuracy of the detection outside the TCP stack. We implement the Implicit Congestion Notification (ICN) algorithm to better understand and investigate the problem of congestion events estimation. This paper is organized as follows: section 2 presents related works, section 3 shows the architecture of the congestion detection, section 4 presents the detailed discussion for the Robust TCP with ICN congestion detection algorithm, and section 5 presents an evaluation of the TCP Robust using simulations. Finally, section 6 concludes this article and presents some perspectives.

2 Related Works

Over the past few years, several solutions have been proposed to improve the performance of TCP. In [5] proposed TCP-DCR modifications to TCP's congestion control mechanism to make it more robust to non-congestion events, this is implemented by using the delay "tau" based on a timer. Our mechanism is different; it relies on the accurate congestion detection algorithm (ICN) and uses the timestamp option to detect spurious timeout which can more improve the reliability of the algorithm and leads to a real Robust TCP. In **Forward RTO-Recovery (F-RTO)**: the F-RTO algorithm of the TCP sender monitors the incoming acknowledgments to determine whether the timeout was spurious. TCP suffers from the inaccuracy of the congestion detection in the other TCP agents, for this reason we design an accurate mechanism of congestion detection (ICN) that interacts with TCP robust. Our study must prove the functionality of our TCP with ICN is better than other versions of TCP. For this point we have to show that the mechanism of congestion detection for some TCP variants (New-Reno, Sack) doesn't detect well when there is congestion and doesn't work better than TCP Robust with ICN. In [5], the idea or the solution proposed for the detection of congestion is the delay of the time to infer congestion by T, and this value should be large to recover from non-congestion event, and should be small to avoid expensive RTO. Our approach is different by using a classic TCP that responds only to an accurate algorithm of congestion detection.

3 Stand-Alone TCP Congestion Events Algorithms

In this section we present the architecture of decorrelating congestion Detection from the Transport Layer (figure 3).The main goal of this architecture is to simplify the task of kernel developers as well as improve TCP performances. This scheme opens the door to another way to react to congestion by enabling ECN emulation at end-host. In this case ICN emulates ECN marking to imply a congestion window reduction.

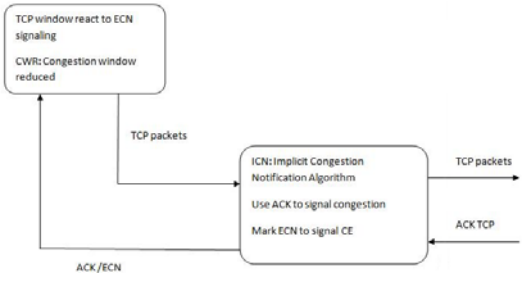


Fig. 3. Decorrelating Congestion Detecting from the Transport Layer

4 Enhancement of TCP Algorithms

Our proposed algorithm which we called Robust TCP is to make the congestion detection reliable and to distinguish the causes of losses in order to improve the flow control. The main idea is to determine CE (i.e. the congestion detection) which impact on the TCP flow performance by monitoring the TCP flow itself. The principle is to obtain a detection system at the edge of a network or at the sender side which analyses the TCP behavior through the observation of both data packets and acknowledgments paths. So, the scenario is to make a new version of TCP (Robust TCP) without detection of congestion. Robust TCP doesn't reacts (reducing of Cwnd) whenever it doesn't receive a notification ECN. Robust TCP must interact with ICN algorithm through ECN. Once we have congestion indication and the congestion event is validated, in this case it must notify the TCP we are exploring the functionality of Robust TCP and the ICN algorithm with the interaction between each other. Robust TCP maintains all the functions of TCP Reno (slow start, Congestion avoidance, Fast retransmit and Fast recovery) and modified by adding error control and limited transmit (like in New-Reno TCP) to avoid unnecessary timeouts.

Robust TCP is a classic version of TCP but very sensitive to packet loss. It contains the major congestion control phases:

1. Slow start
    - When ACK received: cwnd++ which means for every ACK received, the sender sends two more segments.

- Exponential increase in the window (Every RTT: cwnd = 2*cwnd)
- Threshold (sstrhesh) controls the change to congestion avoidance.

2. Congestion avoidance (increase Window size).
   - When ACK received: cwnd+ = 1/cwnd.
   - Linear increment of cwnd (every RTT: cwnd++) slow start is exists until cwnd is smaller or equal to ssthresh. Later congestion avoidance takes over.
3. Fast retransmit (Detection of congestion).
   - TCP generates duplicate ACK when out-of-order segments are received. In this case Fast retransmit uses "duplicate ACK" to trigger retransmission packets, so the sender does not wait until timeout for retransmission, sender retransmits the missing packet after receiving 3 DUPACK.
4. Fast recovery.
   - TCP retransmits the missing packet that was signaled by three duplicate ACKs and waits for an acknowledgment of the entire transmit window before returning to congestion avoidance. If there is no acknowledgment, TCP Robust experiences a timeout and enters the Slow-start state.

TCP recovers much faster from fast retransmit than from timeout. When congestion window is small, the sender may not receive enough dupacks to trigger fast retransmit and has to wait for timer to expire but under Limited transmit; sender will transmit a new segment after receiving 1 or 2 DUPACKs if allowed by receivers advertised window to generate more dupacks. Robust TCP is poor in performance without detection of congestion and worse than other TCP like TCP New-Reno and Sack. It reacts only on the receiving of ECN notification. Once it doesn't receive a notification that means there is no congestion control on TCP and the window keep increasing, but in case of receiving ECN that will indicate the occurrence of congestion indication notified by ICN, than Robust TCP reacts by limiting its sending rate and takes the full meaning of its name.

4.1 ICN with Timestamp

ICN (implicit congestion notification) is an algorithm for congestion detection implemented outside the TCP stack to analyze TCP flow and to better understand the problem of congestion events and then to conclude if the congestion occurs in the network or no and it is also more accurate in congestion detection than TCP. The main goal of ICN is to determine the losses (i.e. the congestion detection) which impact on the TCP flow performance by observing the flow itself which mean by looking at the losses occurring over an RTT period given. ICN is a generic algorithm that doesn't depend on the TCP version used which implements a congestion control where a negative feedback means a loss. It is important to note that ICN doesn't manage the error control which remains under the responsibility of TCP. Starting from the observation of the data segments and the acknowledgments, we identify each TCP connection with a state machine. This state machine indentifies the control congestion

phase and classifies retransmission as spurious or not.TCP congestion control reacts following binary notification feedbacks allowing assessing whether the network is congested or not. ICN algorithm consists of two states:

1. **Normal state:** which characterizes TCP connection without losses, in this state no congestion occurs and the sender receive the ACK normally.
2. **Congestion state:** This state starts from the loss of the first window data segment. When a loss occurs ICN enters in this state and waits to the congestion event to be validated to notify Robust TCP about this loss. When the top of the window is acknowledged, ICN enters in the normal state.

To improve the performance of the congestion detection algorithm and especially against spurious timeout we added the timestamp option, in order once the congestion happens ICN enter in this state and append a timestamp to let the sender to compute the RTT estimate based on returned timestamp in ACK. Time stamps used in this state to measure the round trip time (RTT) of a given TCP segment and including retransmitted segment, this option also can help to eliminate the retransmission ambiguity (due to false indication) and identifies when retransmission is spurious or not. Spurious Timeout are inevitable and not rare in data networks, for this reason and once the congestion event occurs, ICN enter in the congestion event state, timestamp is added for each data segment. Timestamp can be considered as an acknowledging mechanism in the time domain. In the figure (4) shown below we will present the flowchart of TCP Robust with ICN mechanism:

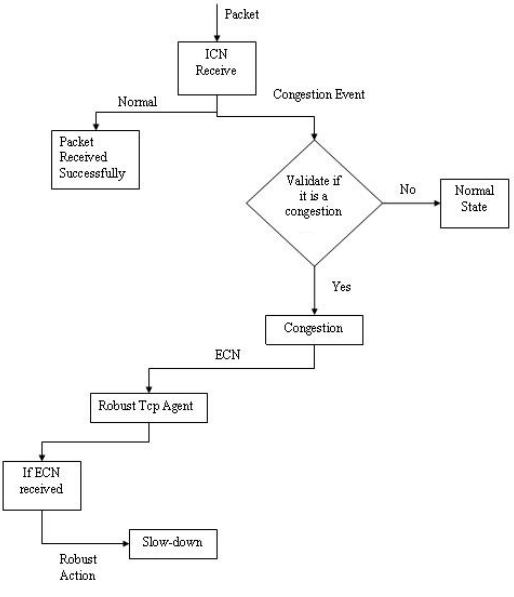


Fig. 4. Robust TCP with ICN detection algorithm

4.2 Robust TCP and ICN interaction

ICN is an accurate congestion detection algorithm where after detecting a loss event in the congestion state, the congestion event (CE) must be validated. The validation of CE should lead to a congestion indication which is the principle responsible to inform the Robust TCP about the congestion. The confirmation method due to a congestion indication is ECN (Explicit congestion notification), which is the main fag in the ACK to notify the loss to the source TCP. Once the source is signaled by ECN notification it reacts by reducing its window (Cwnd) and this time Robust TCP takes the full meaning of its name. After reducing its window, we can notice very well the decreasing of the number of dropped packets (d) in the network due to using of ICN congestion detector and our TCP becomes better in performance than others like TCP New-Reno and Sack.

5 Validations and Evaluations

In this section we evaluate the performance of Robust TCP with ICN algorithm. The main idea is to build an algorithm of congestion detection outside the TCP stack that is responsible to detect the loss and notify it to Robust TCP. The architecture of our tools is shown in the figure (3), which is mainly composed from the following components:

1. Network topology.
2. Traffic model.
3. Performance evaluation metrics.

After the simulation is done, a set of result statistics and graphs are generated.

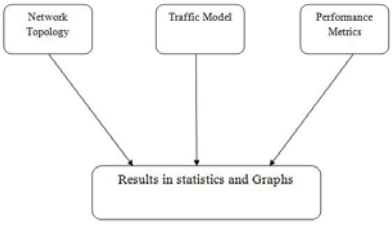


Fig. 5. Architecture of our tools

5.1 Network Topology

To study our TCP and ICN behavior we built our Network and application model shown in figure 6, in which source nodes and sink nodes connect to router 1 or router 2. The bandwidth between the two routers is much lower than the other links, which causes the link between the routers to be a bottleneck. (Traffic can be either unidirectional or bidirectional).

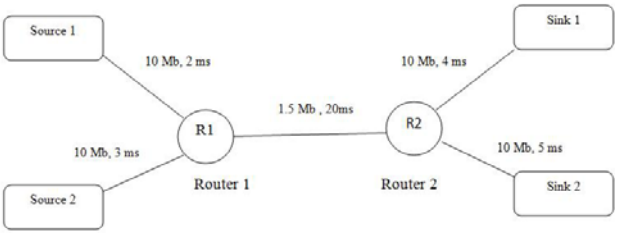


Fig. 6. Network topology

5.2 Traffic Model

The tool attempts to apply the typical traffic settings. In our application include the FTP traffic that uses infinite, non-stop file transmission, which begins at a random time and runs on the top of TCP. Implementation details and a comparative analysis of TCP Tahoe, Reno, New-Reno, SACK and Vegas choices of TCP variant are decided by users.

5.3 Performance Evaluations Metrics

The metrics used in our simulations are Throughput and Drop ratio. Throughput is the total elapsed flow since the beginning of simulation time. Throughput may also includes retransmitted traffic (repeated packets).Drop ratio is the total rate of packet loss during the simulation time. To obtain network statistics, we measure also the drop ratio metric that result in the failure of the receiver to decode the packet and simulation time is 100 seconds. Robust TCP is poor in performance as a standalone TCP but after adding the ICN it becomes much better (see figure 7) and accurate than TCP New-Reno as show in the figure (8). To evaluate our scenario, we compare TCP Robust with other TCP variants (TCP New-Reno) by using different metrics that will show us clearly the improvement of our TCP version compared to others. (Figure 8 and 9).

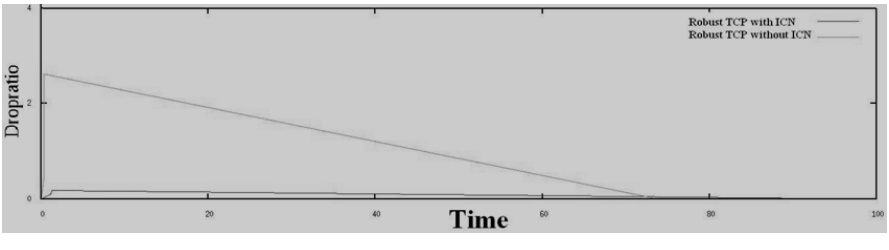


Fig. 7. Comparison between TCP Robust before and after adding ICN algorithm.

The main difference between Robust and New-Reno TCP occurs in the reaction of each protocol. In the TCP New-Reno the reaction will be whenever an error or congestion occurs on the network by slowdown the transmission without being accurate if there is a congestion or not. In addition of that the main problem of New-Reno TCP

that it suffers from the fact that it takes one RTT to detect each packet loss. When the ACK for the first retransmitted segment is received only then we can deduce which other segment was lost. This problem of inaccuracy in TCP New-Reno is solved by the ICN algorithm that the ICN receive the packet and check the presence of congestion by using the normal and congestion phase and by adding the timestamp option which can be make sure of the presence of congestion or no. The deduction of congestion in TCP Robust is different from New-Reno, it will be deduced after signaling ECN from ICN to TCP robust, and then the TCP reacts by decreasing the transmission. This accuracy in detection of congestion can be up to 24 % as difference between the two protocols (Figure 8) before reaction of each one and starting slowdown retransmission. Due the fast reaction of TCP robust, the transmission of TCP become less than in TCP New-Reno which means that the throughput in the TCP robust must be less than in New-Reno, this is clear and deduced in the figure 8.

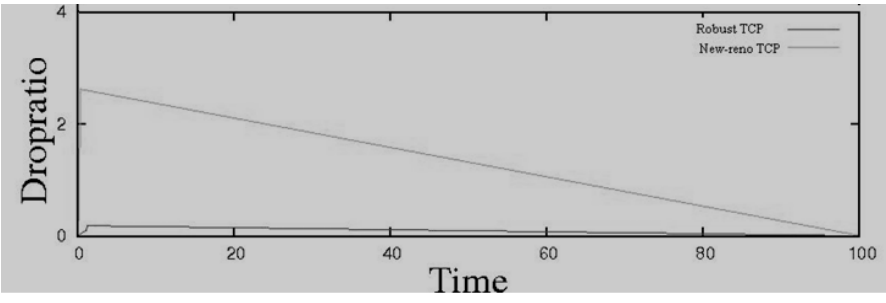


Fig. 8. Comparison between TCP Robust and TCP New-Reno

In figure (9) represents that the drop ratio is less in Robust than in New-Reno due that TCP reacts only when receiving ECN which make its reaction faster.

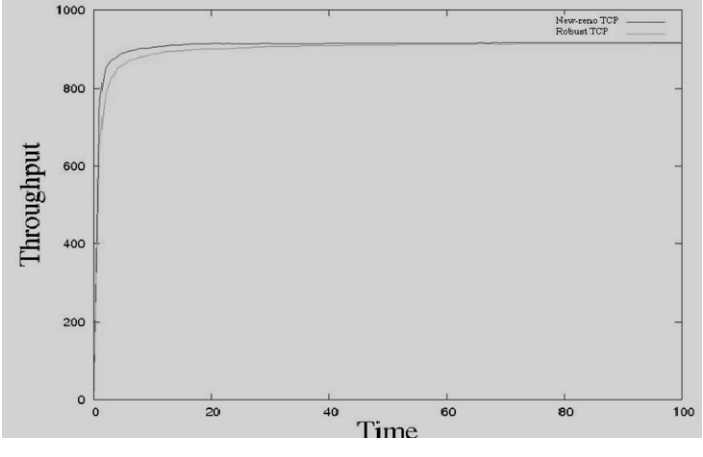


Fig. 9. Comparison between TCP Robust and TCP New-Reno

In figure (9) Robust TCP algorithm reaction is faster than the Reaction of New-Reno, thus Throughput in New-Reno is higher than when using Robust TCP. Congestion detection used by ICN algorithm is more accurate when using the timestamp option for detecting a spurious timeout which improve more the performance of TCP.

The main difference between spurious timeout algorithms relies on the method how to detect spurious timeout by solving the retransmission ambiguity in many circumstances. After clarifying this ambiguity TCP can tell whether the data is there is spurious timeout has happened or not. DSACK, F-RTO and Robust TCP can see the problem of spurious timeout in different aspects. DSACK, an extension of TCP SACK, works it out in the sequence space. It requires the TCP receiver explicitly acknowledging duplicate segments with duplicate SACK options. F-RTO algorithm is used for detecting spurious retransmission timeouts with TCP. It is a TCP sender-only algorithm that does not require any TCP options to operate-RTO delays the decision of loss recovery and waits further two ACK. If the first arrived ACK forwards the sender's transmitting window, TCP concludes a spurious timeout and resume transmitting new data. Our approach is different than other TCP by using an algorithm of congestion detection outside the TCP code, where it can detect congestion and spurious timeout by using the timestamp option at the occurrence of loss or congestion event. The main advantage of ICN with timestamp algorithm is that it can work with spurious timeouts and the others loss events by detecting the congestion in the network immediately and then directly will be notified to Robust TCP in order that TCP after this action will reduce its window, which can improve very well the performance of our TCP.

6 Conclusion and Perspective

This paper has proposed a new algorithm, which is implemented as a standalone component and not inside a TCP stack. This algorithm that interacts with a classic version of TCP is able to detect congestion and notify directly the loss to the Robust TCP through the congestion notification (ECN) in order to reduce its window which leads to a Robust TCP compared to other variants like New-Reno and SACK TCP. In our work we demonstrate that congestion event detection can be realized independently of the TCP code in sake of better detecting congestion occurring in the network. Following this work and the results obtained so far, we are currently planning to develop more the detection of congestion by using the delay-based in the congestion detection algorithm (ICN) and the effect of fast reaction of TCP robust in the Network.

References

1. sarolahti, P., kojo, M.: Forward rto-recovery (f-tro): An algorithm for detecting spurious retransmission timeouts with tcp and the stream control transmission protocol (sctp).rfc 4138, IETF(August 2005)
2. Anelli, P., Harivelo, F., Lochin, E.: On TCP congestion events detection
3. Ludwig, R., Katz, R.H.: The Eifel algorithm: making TCP robust against spurious retransmissions. SIGCOMM Comput. Common. Rev. 30(1), 30–36 (2000)

4. Ramakrishnan, K., Floyd, S., Black, D.: The addition to explicit congestion notification (ECN) to ip. Request for comments 3168, IETF (September 2001)
5. Bhandarkar, S., Reddy, A.L.N.: Networking, May TCP-Dcr: Making TCP Robust to Non-Congestion Events (2004)
6. Floyd, S.: ICSI, RFC 3649 - High Speed TCP for Large Congestion Windows (December 2003)
7. A Comparative Analysis of TCP Tahoe, Reno, New- Reno, SACK and Vegas
8. RFC 793 Transmission Control Protocol (September 1981)
9. Floyd, S.: RFC 3649 High Speed TCP for Large Congestion Window (December 2003)

Transformer Fault Diagnosis Method Based on Improved ANFIS

Hongsheng Su

School of Automation and Electrical Engineering, Lanzhou Jiaotong University
Lanzhou 730070, P.R. China
shsen@163.com

Abstract. In view of some traditional defects, say, incomplete codes, and imperfect criterions on critical values, of IEC 3-ratio-code law in transformer faults diagnosis, a novel transformer faults diagnosis method is proposed based on adaptive neuro-fuzzy inference system (ANFIS) in the paper. The ridge-type distribution functions serve as the fuzzy membership functions in input layer of ANFIS, and Fletcher-Reeves (FR) conjugation gradient algorithm with good global convergence properties acts as the learning algorithm of ANFIS, the learning quality of ANIFS is therefore improved, dramatically. Eventually, the test results in transformer faults diagnosis show the validity of the method.

Keywords: Fletcher-Reeves conjugate gradient method, ridge-type distribution function, ANFIS, transformer, fault diagnosis.

1 Introduction

Transformers are key pieces in power supply and distribution systems, whose normal operation is a basic guarantee for power supply reliability. During the occurrence of the faults the insulating material of the transformers are degraded, resulting in the generation of gases. The type, the amount and the proportion of these gases depend on the type of degraded material, of the responsible phenomenon for the degradation and the levels of energy involved in the action. And so, it is possible to characterize the type and the severity of the transformer faults through the analysis of the gases composition that find dissolved in the insulating oil. Presently, the broadly applied criteria for the fault diagnosis in transformers starting from the analysis of the dissolved gases analysis (DGA) in the oil is IEC 3-ratio-code diagnosis law[1]. Later, the law is further improved to suit the code-combination under condition that several faults are coexistent, which is named as improved IEC 3-ratio-code law[2]. However, the method is inadaptable to the condition that the volume capacities of the gases lower than the threshold value. In addition, IEC codes are incomplete and the criteria for critical values are imperfect[3]. In present, the methods to solve the problems mostly include fuzzy synthesis judgment[4], evolutionary fuzzy logic[5], fuzzy neural network[6], neuro-fuzzy hybrid system[7], Kohonen networks[8], adaptive neuro-fuzzy inference

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 271–280.
springerlink.com

system (ANFIS)[9], and et al. In [4], fuzzy logic is applied to solve the "bottle neck" difficulty in fuzzy boundary knowledge acquisition. It effectively improves the disposing abilities of fuzzy information, but fuzzy membership functions and the shape parameters are subjectively decided by man beforehand. In [5] the traditional IEC 3-ratio-code criteria are applied to establish the initial architecture of the diagnostic system, including the diagnostic rules and membership functions of the fuzzy subsets. After this first step, a genetic algorithm is applied to simultaneously adjust the diagnostic rules and the shape parameters of fuzzy membership function in order to obtain the best performance for the group of samples given. However, the process is no self-adaptive. When the distribution of the samples space is complicated and comprises noise, it is difficult to acquire the fuzzy rules. In [6] fuzzy neural network is applied to implement fault diagnosis so as to improve the learning ability of fuzzy information, but the fuzzy membership functions, and the shape parameters related to it, are still specified by man, subjectively. In [7] a neuro-fuzzy hybrid system that combines a ANN with a fuzzy evolutionary expert system, with a purpose of allying the advantages of high learning and high capacity of non linear map of ANN's with the explicit knowledge represented by the rules of a fuzzy expert system is presented to improve the knowledge denotation and inference learning abilities of fuzzy expert system. In [8] a Kohonen neural network approach is presented. The application of the technique is due to the high capacity patterns classification and low computational cost. In [9] an adaptive neuro-fuzzy inference system(ANFIS), that not only incorporates with the advantages of fuzzy logic and ANN but also overcomes the flaws of the two, is presented to improve the knowledge processing ability and adaptability. But the distribution of the samples space is complicated and comprises noise, the learning of ANFIS is a difficult task, i.e., it is unsuitable to ANFIS learning. Due to that many improved algorithm is proposed. But these methods mostly implement revision and selection for input-output data, and don't resolve the problem, radically. Hence, to improve the diagnostic efficiency and intelligence of IEC 3-ratio-code further, and mine faults information sufficiently, in this paper, the fuzzy membership functions and algorithm of ANFIS are improved so as to construct a high-efficient expert system for transformer faults diagnosis, it boosts up the robustness and adaptability of the diagnosis system, dramatically.

2 ANFIS

2.1 ANFIS Architecture

ANFIS is an integration of neural network and fuzzy inference system(FIS) as in [10]. Sugeno-type FIS is comparatively simple and makes for mathematics analysis, it is for that adopted in ANFIS model. ANFIS provides a modeling method based on data, and easily incorporates with other optimizing self-adaptive methods to form a fuzzy modeling tool with optimizing and self-adaptive ability. Fig.1 shows the ANFIS architecture for transformer fault diagnosis based on IEC 3-ratio-code.

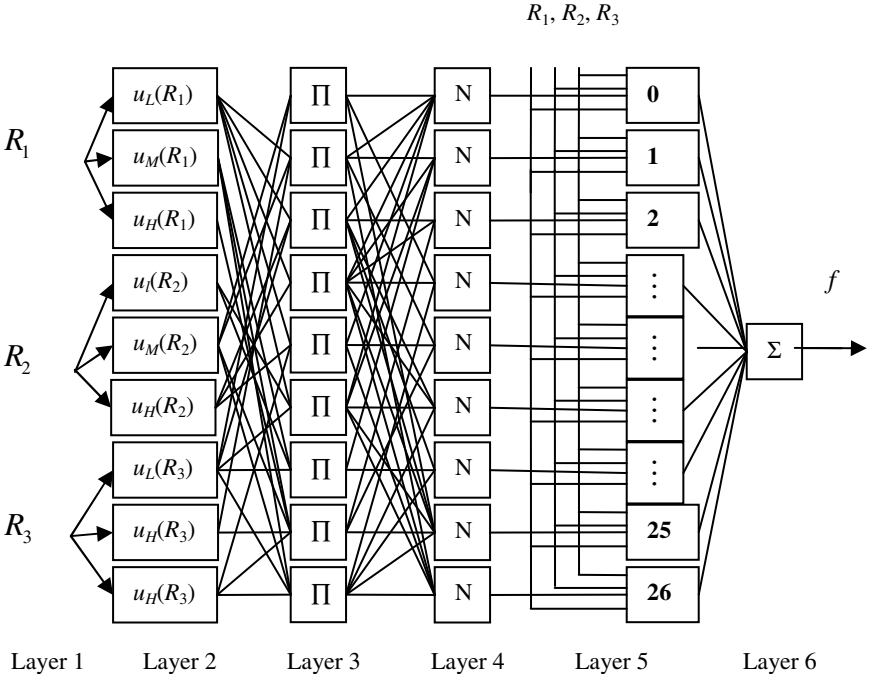


Fig. 1. ANFIS structure

The architecture of ANFIS may be divided as 6 layers. In Fig.1 the 1st layer means input layer where $\boldsymbol{X}$={R_1=C_2H_2/C_2H_4, R_2=CH_4/H_2, R_3=C_2H_2/C_2H_6}.The 2nd layer is fuzzy segmentation layer where there are 9 nodes, and each node applies ridge-type distribution function, whose shape parameters may be adjusted automatically through the learning of the networks. The third layer is rule layer where there are 27 nodes fully covering universe space, each neuron in this layer is corresponding to single Sugeno-type fuzzy rule. The rule neuron accepts its inputs from fuzzy segmentation neurons and computes the rule strength expressed by it. The fourth layer is normalized layer where each neuron accepts all inputs coming from rule layer, and works out active strength for specified rule. The fifth layer is reverse fuzzy segmentation layer where each neuron connects itself to each normalized neuron, and simultaneously, accepts the initial inputs R_1, R_2, and R_3. The sixth layer is a summation neuron where all the outputs in reverse fuzzy segmentation layer are incorporated as the final output y expressed by

$$y = \sum_{i=1}^{n} R_i^{(6)} = \sum_{i=1}^{n} \overline{\mu}_i [k_{i0} + k_{i1} R_1 + k_{i2} R_2 + k_{i3} R_3] \cdot \tag{1}$$

where k_{i0}, k_{i1}, k_{i2}, k_{i3} is the linear parameters of Takagi-Sugeno-Kang(TSK) equation.

2.2 ANFIS Learning Algorithm

Traditional ANFIS adopts the blend learning algorithm of BP and Least Mean Square (LMS), where BP algorithm is applied to decide the shape parameters of fuzzy function and LMS is used to confirm the linear parameters of TSK equations. The algorithm of ANFIS comprises two processes in each iterative cycle, that is, forward propagation of information and backward propagation of error. In forward propagation, the training set of input patterns presents in ANFIS, the outputs of the neurons are calculated layer by layer, where the parameters of the latter item of the rule is represented using LMS. In Sugeno-type FIS, the output y is expressed by the linear function. Hence, give out the shape parameters of fuzzy membership functions and P training sets of the input-output patterns, then P equations are formed below.

$$\begin{cases} y_d(1) = \overline{\mu}_1(1) f_1(1) + \overline{\mu}_2(1) f_2(1) + \cdots + \overline{\mu}_n(1) f_n(1) \\ y_d(2) = \overline{\mu}_1(2) f_1(2) + \overline{\mu}_2(2) f_2(2) + \cdots + \overline{\mu}_n(2) f_n(2) \\ \qquad\qquad \vdots \\ y_d(p) = \overline{\mu}_1(p) f_1(p) + \overline{\mu}_2(p) f_2(p) + \cdots + \overline{\mu}_n(p) f_n(p) \\ \qquad\qquad \vdots \\ y_d(P) = \overline{\mu}_1(P) f_1(P) + \overline{\mu}_2(P) f_2(P) + \cdots + \overline{\mu}_n(P) f_n(P) \end{cases} \tag{2}$$

where

$$f_n(a) = k_{n0} + k_{n1} x_1(a) + k_{n2} x_2(a) + \cdots + k_{nm} x_m(a). \tag{3}$$

In the above equations, m is the number of the neurons in input layer, and n is the number of the neurons in rule layer. Assume that the inputs are represented by $x_1(p)$,⋯, $x_m(p)$, and $y_d(p)$ serves as the desired output of ANFIS, we then have

$$y_d = Ak. \tag{4}$$

where y_d is P×1-dimensional desired output vector expressed by

$$\boldsymbol{y}_d = [y_d(1), y_d(2), \ldots, y_d(P)]^{\mathrm{T}}. \tag{5}$$

and $\boldsymbol{A}$ is P×n(1+m)-dimensional matrix expressed by

$$A = \begin{bmatrix} \overline{\mu}_1(1) & \overline{\mu}_1(1)x_1(1) & \cdots & \overline{\mu}_1(1)x_m(1) & \cdots & \overline{\mu}_n(1) & \overline{\mu}_n(1)x_1(1) & \cdots \overline{\mu}_n(1)x_m(1) \\ \overline{\mu}_1(2) & \overline{\mu}_1(2)x_1(2) & \cdots & \overline{\mu}_1(2)x_m(2) & \cdots & \overline{\mu}_n(2) & \overline{\mu}_n(2)x_1(2) & \cdots \overline{\mu}_n(2)x_m(2) \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots & \cdots \quad \vdots \\ \overline{\mu}_1(p) & \overline{\mu}_1(p)x_1(p) & \cdots & \overline{\mu}_1(p)x_m(p) & \cdots & \overline{\mu}_n(p) & \overline{\mu}_n(p)x_1(p) & \cdots \overline{\mu}_n(p)x_m(p) \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots & \cdots \quad \vdots \\ \overline{\mu}_1(P) & \overline{\mu}_1(P)x_1(P) & \cdots & \overline{\mu}_1(P)x_m(P) & \cdots & \overline{\mu}_n(P) & \overline{\mu}_n(P)x_1(P) & \cdots \overline{\mu}_n(P)x_m(P) \end{bmatrix} \tag{6}$$

and $\boldsymbol{k}$ is the n(1+m) ×1-dimensional vector expressed by

$$k = \begin{bmatrix} k_{10} & k_{11} & k_{12} & \cdots & k_{1m} & k_{20} & k_{21} & k_{22} & \cdots & k_{2m} & k_{n0} & k_{n1} & k_{n2} & \cdots & k_{nm} \end{bmatrix}^T \tag{7}$$

Usually, the number of the input-input pattern P is larger than $n(1+m)$, and the exact solution of equation (4) is possibly inexistent. However, we may find the LMS estimation k^* of k that can make $\| Ak - yd \|^2$ least. This can be realized using the following fake-reversion technique.

$$k^* = (A^T A)^{-1} A^T y_d . \tag{8}$$

During learning of the network, the calculating error may be adjusted according to the practical outputs and the desired outputs of the system. The mean square error (MSE) of the network is expressed by

$$H = \frac{1}{2}\left(y_{ej} - y_j \right)^2 . \tag{9}$$

where y_{ej} and y_j are the desired value and the output value of the j^{th} node in output layer. During learning, centre d_{ij}, width σ_{ij}, weigh w_{ij} are adjusted by

$$d_{ij}(n+1) = d_{ij}(n) - \eta \frac{\partial H}{\partial d_{ij}} = d_{ij}(n) - \eta\lambda . \tag{10}$$

$$\sigma_{ij}(n+1) = \sigma_{ij}(n) - \eta \frac{\partial H}{\partial d_{ij}} = \sigma_{ij}(n) - \eta\lambda \underline{w}_j f_j (w_i - y_{ej})[x_i - d_{ij}(n)] / \sigma_{ij} . \tag{11}$$

$$w_{ij}(n+1) = w_{ij}(n) - \eta\mu_{A_j}\mu_{B_j}(y_{ej} - y_j) / \sum_j w_j \tag{12}$$

where n is the iterative times, η is learning factor, $0<\eta<1$, and

$$\lambda = [(y_{ej} - y_j) / \sum_j w_j]\mu_{A_j}\mu_{B_j}\{2[x_i - d_{ij}(n)] / \sigma_{ij}^2(n)\} . \tag{13}$$

2.3 Innovation of ANFIS Learning Algorithm

In the traditional ANFIS, the adjustment of network parameters are implemented through BP learning algorithm or the combination of BP and LMS algorithms. When applying BP algorithm, the shape and linear parameters of ANFIS are adjusted at the same time. When applying combination algorithm, the shape parameters are adjusted through BP algorithm and the linear parameters are adjusted through LMS algorithm. Clearly, the two apply BP algorithm. However, BP algorithm has some defects, say, low training speed and easily falling into local extremum. Due to that many improved methods are presented. They can be generally classified as the two types. One is to apply the heuristic technique through analysis on performance function of BP algorithm, for instance, additive momentum law, self-adaptive learning rate law, and resilient back-propagation (RBP), and et al. The other is to speed BP algorithm through standard numerical value optimization, such as conjugation gradient method, quasi Newton method, and Levenberg-Marquart(LM) law and so on. Through a number of

analysis and comparison on diverse algorithms in terms of structure, complicity, and training accuracy, Fletcher-Reeves(FR) conjugation gradient method is considered to be conveniently realized and achieve better performance[11], hence it can be adopted in this paper. Standard BP algorithm adjusts the weights along the most rapid direction of loss function descending, that is, negative gradient direction. Its single step algorithm is expressed as follows.

$$W_{n+1} = W_n + a_n G_n = W_n - a_n \nabla f(W_n). \tag{14}$$

where W_n is the current weigh vector, $\nabla f(W_n)$ is the current gradient, and a_n is the learning rate.

In conjugation gradient method, the first step of the algorithm is explored along negative gradient direction too, i.e., $D_0=G_0=-\nabla f(W_n)$, and then, linearly scout is performed to ensure an optimal motion distance in the current searching direction.

$$D_n = G_n + \beta_n D_{n-1}. \tag{15}$$

where D_n is the scout direction of conjugation gradient algorithm. In diverse algorithms the calculation of β_n is different. In FR algorithm, β_n is computed by

$$\beta_n = \frac{G_n^T G_n}{G_{n-1}^T G_{n-1}}. \tag{16}$$

During searching, the direction still required to be corrected, and the algorithm will not be ensured to be convergent otherwise. The corrected method is described as follows. When the training time is integral times the weigh number, let $\beta_n=0$. At the same time, in every step of the training, if $\nabla f(W_n)^T D_n \geq 0$, then $D_n= -\nabla f(W_n)$. This ensures that the searching is always implemented along error descending.

3 Transformer Fault Diagnosis Model Based on ANFIS

In fuzzy region partition of IEC 3-ratio-code, the boundary points 0.1, 1, and 3 are educed based on a lot of trials and analysis on transformer fault examples. Clearly, they are statistic and of decentralization. The statistic method inevitably abnegates some minor factors, and the precision may be sacrificed. However, fuzzy membership function may describe the decentralization phenomenon. When fuzzy mathematics is applied to deal with code problems, the key is to construct the fuzzy membership functions of the codes. According to knowledge and experience of expert system(ES), descending ridge-type distribution function is selected to describe the fuzziness of code regions. Thus the fuzzy membership functions of the inputs $R_1=C_2H_2/C_2H_4$, $R_2=CH_4/H_2$, and $R_3=C_2H_2/C_2H_6$ on code 0, 1, and 2 can be constructed beolw.

1) Fuzzy membership function of R_1

$$u_0(x) = \begin{cases} 1 & x \leq 0.08 \\ 0.5 - 0.5\sin[25\pi(x-0.1)] & 0.08 < x \leq 0.12 \\ 0 & x > 0.12 \end{cases}. \tag{17}$$

$$u_1(x)=\begin{cases}0 & x<0.08\\ 0.5+0.5\sin\left[25\pi(x-0.1)\right] & 0.08<x\le 0.12\\ 1 & 0.12<x\le 2.9\\ 0.5-0.5\sin\left[5\pi(x-3)\right] & 2.9<x\le 3.1\\ 0 & x>3.1\end{cases}. \tag{18}$$

$$u_2(x)=\begin{cases}0 & x\le 2.9\\ 0.5+0.5\sin\left[5\pi(x-3)\right] & 2.9<x\le 3.1\\ 1 & x>3.1\end{cases}. \tag{19}$$

2) Fuzzy membership function of R_2

$$v_0(x)=\begin{cases}0 & x<0.08\\ 0.5+0.5\sin[25\pi(x-0.1)] & 0.08<x\le 0.12\\ 1 & 0.12<x\le 0.9\\ 0.5-0.5\sin[5\pi(x-1.0)] & 0.9<x\le 1.1\\ 0 & x>1.1\end{cases}. \tag{20}$$

$$v_1(x)=\begin{cases}1 & x\le 0.08\\ 0.5-0.5\sin[25\pi(x-0.1)] & 0.08<x\le 0.12\\ 0 & x>0.12\end{cases}. \tag{21}$$

$$v_2(x)=\begin{cases}0 & x\le 0.9\\ 0.5+0.5\sin[5\pi(x-1.0)] & 0.9<x\le 1.1\\ 1 & x>1.1\end{cases}. \tag{22}$$

3) Fuzzy membership function of R_3

$$w_0(x)=\begin{cases}1 & x\le 0.9\\ 0.5-0.5\sin[5\pi(x-1.0)] & 0.9<x\le 1.1\\ 0 & x>1.1\end{cases}. \tag{23}$$

$$w_1(x)=\begin{cases}0 & x<0.9\\ 0.5+0.5\sin[5\pi(x-1.0)] & 0.9<x\le 1.1\\ 1 & 1.1<x\le 2.9\\ 0.5-0.5\sin[5\pi(x-3.0)] & 2.9<x\le 3.1\\ 0 & x>3.1\end{cases}. \tag{24}$$

$$w_2(x) = \begin{cases} 0 & x \le 2.9 \\ 0.5 + 0.5\sin[5\pi(x-3.0)] & 2.9 < x \le 3.1 \cdot \\ 1 & x > 3.1 \end{cases} \tag{25}$$

4 Example

270 samples are randomly drawn out from the DGA diagnostic knowledge base of the no.1 main transformer in one substation to train ANFIS. These samples are averagely classified as 9 groups corresponding to 9 fault-types of IEC code table. Below we take the fifth fault in IEC code table, i.e., low temperature overheating, as an example to illustrate modeling procedure of ANFIS. Let 30 samples present low temperature overheating fault, and the corresponding outputs be 1, the rest 240 samples be other type's fault-type and the outputs be 0. Finally, some data are randomly drawn out from the rest data in data base as checking data to perform model affirmation. Fig.2 shows the change condition of the step length during training process.

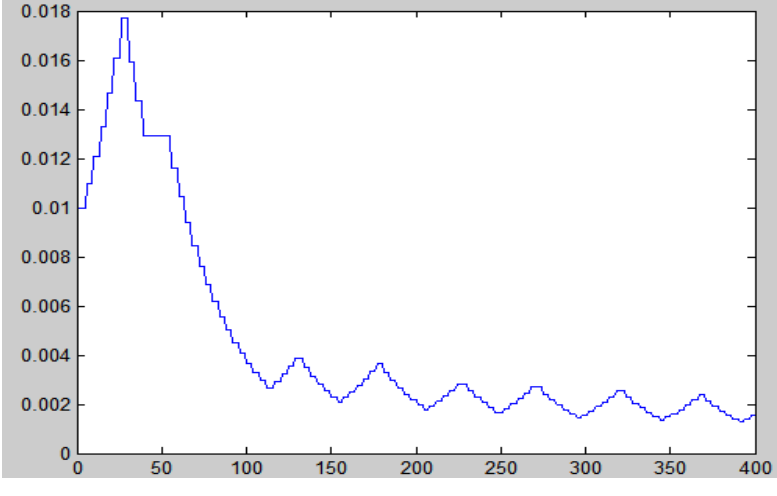


Fig. 2. Curve of training step-length change

Fig. 3 shows the error graph during learning.

Seen from Fig.3, applying the descending ridge-type membership function with its parameters given, the system may rapidly converge, the computing cost is, due to that, saved to very extent.

To test the validity of the model, some data are randomly selected to serve as checking data that they don't participate in training. Fig.4 is the fitting results of the checking data of ANFIS.

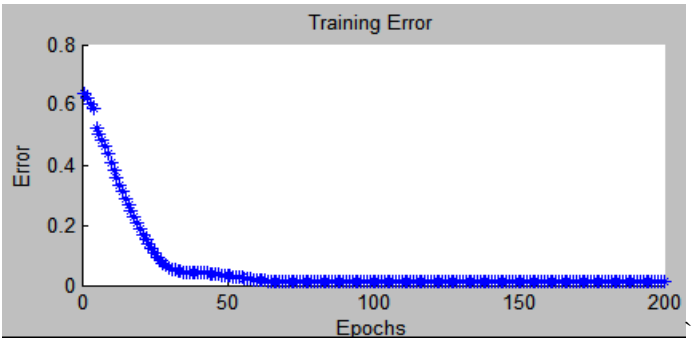


Fig. 3. Training error curve

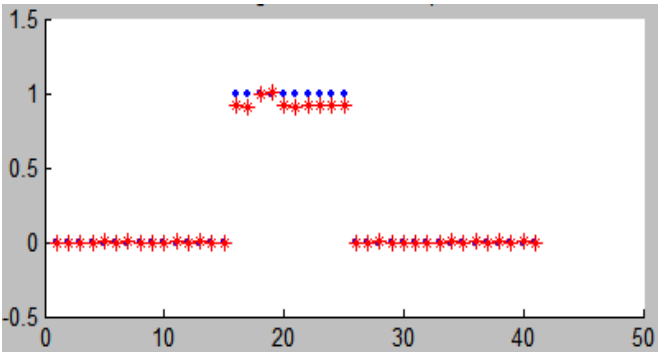


Fig. 4. Fitting test data

Below a comparison is made between the improved ANFIS and traditional ANFIS learning algorithm.

Table 1. Transformer faults indentification results.

Input data	Num. of samples tested			% Accurate	
	Correct diagnosis	Improved ANFIS	Traditional ANFIS	Improved ANFIS	Traditional ANFIS
Fault-type 0 sample	30	28	19	93	63
Fault-type 1 sample	30	27	16	90	53
Fault-type 2 sample	30	28	20	93	66
Fault-type 3 sample	30	27	23	90	76
Fault-type 4 sample	30	29	22	96	73
Fault-type 5 sample	30	28	22	93	73
Fault-type 6 sample	30	27	20	90	66
Fault-type 7 sample	30	27	23	90	76
Fault-type 8 sample	30	28	21	93	70

Seen from Table1, compared with traditional ANFIS with BP algorithm based, the improved ANFIS algorithm shows higher accuracy and excellent property.

5 Conclusions

ANFIS can automatically acquire fuzzy rules through training samples and overcome the defects of fuzzy inference system(FS) that excessively relies on knowledge and experience of experts. ANFIS applies FR conjugation gradient algorithm to effectively overcome the flaws of BP algorithm and speed the convergent speed. The spot applications prove that mentioned ANFIS in this paper applied in transformer faults diagnosis is feasible. ANFIS is a self-adaptive system and can rapid track input-output data samples to adjust own parameters, it can be so used to investigate on-line faults diagnosis technique for the transformers, further.

References

1. Dong, M., Zhao, W., Yan, Z.: An Improved Method for Ttransformer Fault Diagnosis Based on DGA. High Voltage Engineering 4, 6–8 (2002)
2. Cao, D.K.: Dissolved Gases Analysis and Faults Checking in Transformer Oil. China Electric Power Publishing House, Beijing (2005)
3. Xu, K., Zhou, J., Ru, Q.: Development and Prospect of Ttransformer Oil Dissolved Gas On-line Monitoring Technology. High Voltage Engineering 8, 30–32 (2005)
4. Su, H., Wang, X., Chen, H.: An ANFIS Based Fuzzy Synthesis Judgment for Transformer Fault Diagnosis. In: Yu, W., He, H., Zhang, N. (eds.) ISNN 2009. LNCS, vol. 5552, pp. 172–181. Springer, Heidelberg (2009)
5. Huang, C., Yang, H.T., Huang, C.L.: Developing a New Transformer Fault Diagnosis System Through Evolutionary Fuzzy Logic. IEEE Transactions on Power Delivery 12, 761–776 (1997)
6. Su, H., Li, Q.: Fuzzy neural classifier for transformer fault diagnosis based on EM learning. In: Huang, D.-S., Li, K., Irwin, G.W. (eds.) ICIC 2006. LNCS (LNAI), vol. 4114, pp. 222–229. Springer, Heidelberg (2006)
7. Wang, Z., Griffin, P.J., Liu, Y.: Neural Net and Expert System Diagnose Transformer Faults. IEEE Computer Applications in Power 13, 50–55 (2000)
8. De, A., Chatterjee, N.: Recognition of Impulse Fault Patterns in Transformers Using Kohonen's Self-organizing Feature Mmap. IEEE Trans. on Power Delivery 17, 489–494 (2002)
9. Zhao, F., Su, H.: Transformer Fault Diagnosis Method Based on ANFIS. Hi-TECH 4, 69–70+77 (2008)
10. Jhy-Shing Jang, R.: Anfis: Adaptive–Network-Based Fuzzy Inference System. IEEE Transactions on Systems, Man and Cybernetics 23, 665–685 (1993)
11. Hong, L., Mo, L., Wei, Z.: A Modified Conjugate Gradient Mmethod with Global Convergence Property. Journal of Mathematical Research & Exposition 29, 65–75 (2009)

Causes and Solutions of Electrical Circuit Fire of the Residential Basements in Beijing City, China

Ziyue Xu[1],Yuanfang Lu[2], and Ning Han[1]

[1] School of Engineering, Beijing Forestry University, 100083 Beijing, China
[2] School of Architecture and Urban Planning, Beijing University of Civil Engineering and Architecture, 100044 Beijing, China
{abyssliz,luyuanfang}@gmail.com, hn217@bjfu.edu.cn

Abstract. In this paper, a large number of studies which are about the typical residential basements have been done in Haidian District, Beijing City, China. Under the national regulations and the power of relevant authorities, we still regard the security risks of an electric circuit fire, in the first place, as a result of wrongly installment of electrical circuits. In this article, these risks have been carefully analyzed as well as the corresponding solutions being proposed. At the same time, we hope that an increasing attention will be paid to the residential basement and those users.

Keywords: Beijing City, residential basements, electrical circuit, circuits fire, electric design.

1 Introduction

In Beijing, the basement as a rental lease, the residence of the phenomenon is widespread. Some of the basements are constructed for residence and they are used directly as a part of the building, the other residential basements are lately converted from underground storerooms or underground parking lots. The basements provide domiciles for low-income workers in catering, logistics, cleaning and other fields in Beijing, nonlocal students and job seekers living in Beijing for a short time.

2 Current Situation

Because laws and regulations are not sound and the management system is not smooth[1], living in the residential basement deals with a variety of security risks, such as entrances, corridors and other public spaces blocked with all kinds of debris which leads to no access to disperse the crowd when an emergency occurs, and no access control systems to avoid the theft events[2]. After conducting researches of the rental residential basements in Haidian District, Beijing City, it is obvious that the design and construction problems of electrical circuit contribute most to the common electric fire.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 281–286.
springerlink.com

2.1 Inhabitants

Those people who live in basements are the city's low-income workers, as a consequence, most of them are lack of necessary knowledge and have little command of living safety, especially the use of electrical equipments.

2.2 Living Environment

The residential basements can be divided into two categories, semi-basements or those entirely built underground. Furthermore, the basements constructed fully under the ground also consist of two kinds, negative one or two. They have no chance to acquire the sunshine. Due to the features of sunshine and air flowing, the lower of a basement level, the darker and more humid the house will be, so the less rental is charged. The low-incomers' choosing living in basements results from the valid fact that they are comparable cheap to rent in big cities like Beijing. According to the investigation of the residential basements near the Fourth Ring Road in Haidian District, the monthly rental for each semi-basement room (10 square meters residential area on average) varies from RMB500 to 1200, those entirely beneath the surface are charged from RMB200 to 500 per month (10 square meters residential area on average), to the lowest, the monthly rental of a negative two basement is RMB150.

Because of the low price, characters of the residences and the poor knowledge of the tenements in safe basement living which have already drawn the public attention, cause the cheat on labor and materials and short of maintaining so as to get both the living quality and residential environment into stuck initially.

The weather in Beijing is hot and rainy in summer, cold and dry in winter with a short period of spring and autumn. Take the statistics in 2007 for example, it is an average of 14 centigrade annually, negative 7 to 4 in January and 25 to 26 centigrade in July. The extreme low record is negative 27.4 centigrade as well as the highest is beyond 42 centigrade. 180 to 200 days are without frost all year round with a shorter period of western mountain regions. In addition, the average amount of raining is 483.9mm, whose rainfall is one of the most ones in northern China.[3]As a result, humidity is one of the regional weather characters. Worse still, humid environments do harm to the safe use of electrical circuits. Besides, those basements with common shared toilets, laundries and bathrooms lead to a large wet environment, which calls for our consideration in electrical design and construction, as well as the living safety of residences.

3 Formation of Electrical Circuit Fire of the Residential Basement

Electrical circuits consist of lines and cables, and electrical lines are long pieces with lots of branches. They have many opportunities to contact with combustible material and wire cable insulation materials are mostly combustible organic materials. When the cable or local overheating or the insulation reaches a certain temperature, it will burn and cause fire accidents result from the rapid spread of flames, but also produce large amounts of toxic gases such as hydrogen chloride and carbon monoxide which will cause greater harm to the human body. [4]

In Western Europe, the electricity consumption per capita is tens of times greater than the one in China. But in Western Europe, electrical fire only accounts for a few percent of the total fire, which is in sharp contrast when compared with that of China which is reaching more than 50%. [5]

Also, according to the fire statistics of the whole nation in the first half of 2009 provided by the Fire Department Ministry of Public Security, among the direct causes of fire accidents, the wrong electrical installation caused fire accidents up to a total of 19,852 accounting for 26.8% of the total. Second, careless use of fire caused 16,119 fire accidents making up 21.8%. In addition to the two main reasons, the fire playing caused 8,138 fires accounting for 11%, 5,727 fires caused by smoking accounting for 7.7%, advertence of operation caused 3,171 fires accounting for 4.3%. Also, spontaneous combustion leaded 1,418 fires accounting for 1.9% and lightning, static electricity and other causes caused 12,385 fires accounting for 16.7%. Meanwhile, the fires whose causes are unknown and under investigation are 7,134, accounting for 9.6%. [6]

After in-depth analysis of some typical residential basements and its tenants, we found that the main causes of the electrical fire are human factors. For example: on-site operator is not in strict accordance with the rules of construction, some staffs are even not equipped with the appropriate qualifications; the regulators pay far less attention than then should and electrical facilities are lack of necessary inspection or maintenance management; the poor electrical knowledge and safety awareness lead the tenants randomly change the protecting devices to meet the requirements of some electrical facilities, causing a large burden of electrical wiring which makes the devices operate irregularly.

4 Causes of Electrical Circuit Fire of the Residential Basement

Constraints such as cost control, when designing or constructing the residential basement, the electrical circuits are often one-sidedly for the sake of conserving materials and other resources, especially for the basements for living purpose adapted from warehouses or parking lots. Since the absence of heaters in such basements, most tenants will use high-power heating equipments in winter, which will result in frequent tripping of electrical circuits. More seriously, the long-term overload of these circuits and the electrical insulation will lead to electrical accidents. According to statistics, China's electric fire holds the first place because of the following reasons. [7]

4.1 Low Standard of Electrical Circuit Design and Small Circuit Capacity

When designing the electrical circuits, mode selection is usually relevant to the lower limit of the circuit capacity, without a great foresight. The new national standard, "Residential Design Code" (GB50096-1999) put into effect in 1999 has published some requirements as follows. Electrical lines should be consistent with fire-resistance requirements of laying wire. Copper wire should be used, and each line into the family home should not be less than $10mm^2$ and the branch circuit cross-section should not be less than $2.5mm^2$. The power outlet sockets of kitchens and bathrooms

should be an independent circuit. Every housing unit should be equipped with a power circuit breaker, and should not touch the phase line and the neutral line switching equipment at the same time. [8] But it must be noted that these standards are still the lowest ones in residential electrical design requirements.

Due to the specialties of the development and construction of the residential basements, their electrical circuits are not only much different from the residential units above ground, but also not in the same with those of public institutions. Such basements do not have full-time professional electrical maintenance personnel, and tenants do not understand electrical knowledge and have no ability to maintain electrical safety. Therefore, when some problems occur, fire accidents are prone to merge. Coupled with the design requires electrical safety, functionality and adaption to the development, electrical circuit design of residential basements needs more long-term plans.

4.2 Negligence of Installation and Maintenance of Electrical Circuits

When laying the electrical lines with poor planning or in order to save lines, some wire connectors are in the tubes, and some wires go out of the protective casing 1m to 2m then expose to other objects, and simple wire does not twist as insulated. These situations will sooner or later cause a fire disaster. Insulated wires are laid directly using ordinary non-flame-retardant plastic casings. As soon as there is a spark, the fire may easily spread quickly and then becomes a fire disaster. [9]

In the researching process, particularly in toilets and bathrooms and other damp places found that there is no insulation for the insulated to handle the situation. What is worse, when part of the electrical power system is running, the connector is directly exposed which is not only make it easier to pull the private wire, but also prone to electric shocks threaten people's safety or electric fire caused by a short circuit.

4.3 Poor Operating Condition of Electrical Circuits

The basement is very damp, combined with insufficient ventilation set reasonably. Electric leakages are very likely to occur. And because of lacking in knowledge of electric fire, tenants often accumulate flammable debris such as useless boxes and waste newspaper in the channel and other public spaces. On one hand, it makes it easier for the fire disaster to take place when an electric leakage occurs. On the other hand, due to the accumulation of such debris blocked the channels making it difficult to escape when a fire emergency comes into being, it is not hard to lead a more serious personal and property losses.

5 Solutions of Electrical Circuit Fire of the Residential Basement

5.1 Design Reasonably and Focus on Residential Electrical Fire

With economic development, people's power demand increases, making the design of electrical circuits guaranteed in safety and reliability at least 20 to 30 years. In terms of wiring should be noted: it is much likely to set a fire when wiring with aluminum than doing with copper. The placing of lines should be the mean of concealed way

which is combined with both beauty and safety. Increasing the number of sockets to remove all kinds of security risks and set a reasonable number of residential branch circuits can ensure the quality of work. [10] Meanwhile, the high-power equipments, especially the heating equipments used for long periods in the residential basement when winter comes, should be set a single loop for each. Last but not least, setting a fire alarm device is also very important.

5.2 Standard Installation and Adaption of the Electrical Circuits

Creating a safe environment are positive measures to reduce electrical fires. Standard installation and alternation is the first step to cut off a fire from the source which greatly avoids the line aging, overloading and other hazards. This requires the electrical design, construction and management to work together. And the professional ability of the personnel should be strengthened with training and practicing.

5.3 Regular Testing of the Existing Electrical Circuits

Fire agencies should conduct regular testing for the electrical circuits. It is also necessary to establish regular electrical safety testing system. Regular testing should be managed by professionals to detect random access, damaging and aging phenomena which can effectively prevent the occurrence of fire accidents.

5.4 Improvement of the Relevant Regulations and Management

Although the system of residential design has been basically established, the laws and regulations of living in the residential basement are still not authoritative and aimed enough. Secondly, for this special residential environment, it lacks in effective management of organizations, and even can be regarded as a management blind spot.

5.5 Popularization of the Electric Knowledge

Electrical fires of residence are caused mainly by the weakness in electrical fire knowledge. All of the levels in government and community should provide useful accesses to knowledge of electrical circuit at the very beginning. The tenants need regular and long-term education of the proper use of electricity, such as posting a variety of easy-to-learn knowledge of fire illustrated, and organizing professionals to explain electric knowledge to the tenants in their tenements in order to promote safe use of electricity.

5.6 Active Use of New Materials and New Technology

In new construction projects, it is better to utilize more new materials and modern technology, such as the use of flame-retardant or fire resistant wires, cables, pipes with fire blocking materials, electrical leakages and overload protection with circuit breakers and so on. [11]

6 Conclusion

The residential basements provide a shelter for more than 10 million residents in Beijing. Along with the development of Beijing, the status of the residential basements will play an increasingly critical role for population from other places, and therefore such basements require more staffs in relevant fields like electrical engineers to pay attention. What is more, the construction and renovation of electrical engineering of the residential basements calls for improvement of reliability and security urgently, as well as the supplement of the load for the long-term with reserving margin.

References

1. Liu, Y.: Obstacles and Strategies of Energy Saving for the Existing Building in Wuhan City(in Chinese). A dissertation of Huazhong University of Science and Technology (2007)
2. Yu, P.: Immunity Rules of Residing safely(in Chinese). Illustrate Real Estate by Pictures, 26–27
3. Climate and Environment of Beijing(in Chinese), http://baike.baidu.com/view/2621.htm#3
4. Yang, Y.: Technology of Preventing Electrical Circuit Fires of Civil Engineering(in Chinese). Guangdong Architecture Civil Engineering 4, 15–17 (2007)
5. Shi, W.: The Formation of Electric Fires and the Relevant Prevention(in Chinese). Technology Forum, 77
6. The National Fire Statistics of the First Half of, 2009 Released by Fire Department Ministry of Public Security(in Chinese) (2009), http://www.gov.cn/gzdt/2009-07/03/content_1356580.htm
7. Dong, L.: On Electric Wiring for Residence Projects. House Science(in Chinese), 59–62 (2007)
8. Design Code for Residential Buildings, http://baike.baidu.com/view/2683768.htm
9. Shan, L.: Prevention of Electric Fires of Civil Construction(in Chinese). Science and Technology Consulting Herald 16, 83 (2007)
10. Zhu, J.: Design for Fire Protection of Electrical Circuits of Residential buildings(in Chinese). China Science and Technology Information 7, 112–124 (2006)
11. Beijing Municipal People's Air Defense Engineering and General Management of the Safe Use of the Basement for Trial Implementation (2006) (in Chinese)
12. Luo, J.: Analysis and Prevention Strategies of Electric Circuit Fires in Residential Buildings(in Chinese). Engineering, 655–656

Minimization of Torque Ripple in Switched Reluctance Motor Drive – A Review

S. Jebarani Evangeline and S. Suresh Kumar

EEE Department, Karunya University (Karunya Institute of Technology and Sciences)
Coimbatore 641 114
jeba_eva@karunya.edu

Abstract. In Switched Reluctance Motor, the basic control issue is production of ripple free torque. The presence of ripple in the torque leads to production of undesirable noise and undesirable vibration during the operation. The static characteristics along with the magnetization pattern of the individual phases dictate the amount of torque ripple during operation. Concept of Electronic Control approaches for dynamic torque ripples are extensively reviewed and presented.

Keywords: Switched Reluctance Motor, Torque Ripple Minimization, Electronic Control, Torque Sharing Function, Current Profiling.

1 Introduction

Switched Reluctance Motor (SRM) has become popular in Industrial and domestic application due to its high torque to inertia ratio, high efficiency, low cost, variable speed operation and good dynamic response. The primary disadvantage of SRM is torque ripple in the generated torque, leading to acoustic noise and mechanical vibration, which must be kept within the permissible limits. The torque ripple is not tolerable in Direct Drive applications. There are primarily two approaches for reducing the torque ripple; one method is to improve the magnetic design of the motor, while the other is to use sophisticated electronic control techniques. In magnetic design, the reduction in torque pulsations is obtained by changing the stator and rotor pole structures with some penalty on the motor torque. The electronic approach is based on optimizing the control parameters, which include the supply voltage, turn-on and turn-off angles, and current level.

Firstly, the cause of torque ripple in SRM is mainly due to the switching of phase currents into its windings and the highly nonlinear nature of phase inductance variation when the rotor rotates. When the successive phase windings are excited in sequence to produce continuous rotation, the total torque is the sum of the torque generated due to the currents in the outgoing phase and the current in the incoming phase, which are controlled independently. Torque pulsations are encountered at these instants of current switching from one active phase to the other. From the following Equation (1), it is observed that if improvements in torque profile are to be achieved, the excitation current i and /or the phase inductance L (θ, i) are to be modified.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 287–294.
springerlink.com

$$T(\theta, i) = \frac{1}{2} i^2 \frac{\partial L(\theta, i)}{\partial \theta} \quad (1)$$

Therefore, the torque depends upon the torque – current – angle characteristics of the machine. Secondly, the absence of reference frame transformation for SRM leads to dependency of rotor angle for torque production.

Torque ripple is defined as as the difference between the maximum and minimum instantaneous torque expressed as a percentage of the average torque during steady state operation. Mathematically, Percentage Torque ripple is expressed as given in equation (2)

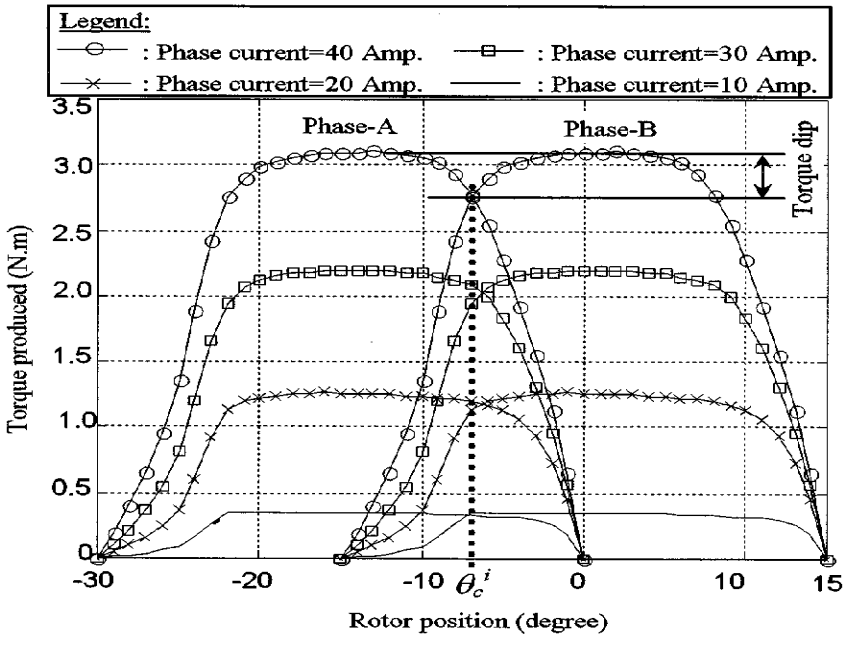


Fig. 1. Torque – current – angle characteristics of SRM

$$Torque\ Ripple\ (\%) = \frac{T_{Inst}(\max) - T_{Inst}(\min)}{T_{ave}} \quad (2)$$

Where T_{Inst} is the instantaneous torque produced during every switching and T_{ave} is the average torque value. The torque ripple is evaluated from the torque dips in the static characteristics. Torque dip occurs at the overlap region during which the two overlapping phases produce equal torque at equal levels of current. When the incoming phase becomes deficient in supplying the required torque, torque dip occurs. The period of overlap is related to the no of strokes per revolution or step angle ε.

$$Step\ Angle\ \varepsilon = \frac{2\pi}{N_r N_{Ph} N_{rep}} \tag{3}$$

Where N_r is the number of rotor poles, N_{ph} is the number of phases and N_{rep} represents the number of pole pairs per phase. A low value of step angle will increase the overlap angle and also influences the frequency of control per revolution. Higher number of strokes would lessen the torque ripple with high value of rotor poles Nr which would decrease the rotor saliency, increase VA rating of the inverter, switching frequency and copper loss [1]. A lower torque ripple would improve the average torque produced. The number of strokes per revolution could be increased by larger N_r with the penalty in the saliency ratio. In [3], the author had discussed about the choice of number of rotor poles. Increasing N_r would increase the voltage ampere rating of the controller used. The core losses will also increase due to higher frequency switching. Torque ripple occur mainly in the overlap region as the torque production shifts from one phase to another. The value of overlap angle is given by the formula.

$$Overlap\ angle = min(\beta_r, \beta_s) - \frac{2\pi}{N_r N_s} \tag{4}$$

As the amount of overlap increases with N_r and N_s, one would try to maximize these values. Maximising the value of N_r and N_s will lead to complexity in the converter design and would limit the operating speed of the machine. It also depends upon the stator pole arc β_s and rotor pole arc β_r.

This paper presents a detailed survey of various electronic strategies implemented for Torque Ripple Minimization.

2 Electronic Control Approach

2.1 Torque Sharing Function

One of the efficient methods for minimization of dynamic torque ripples is to track the current / torque produced in every individual phase by a Torque Sharing Function (TSF). The TSF are function of turn on angle T_{on}, overlap angle θ_{ov}, turn off angle T_{off} and motor speed ω.

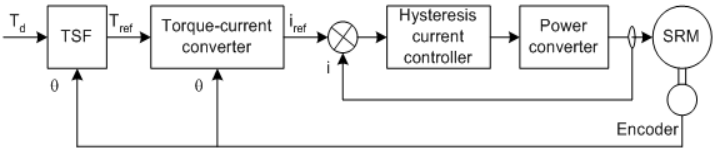


Fig. 2. Torque Control using TSF

The torque/flux/current controllers tracks the expected value of torque/flux/current based on TSF. With TSF, the SRM drive operates either on hysteresis or PWM control. In order to maintain the desired instantaneous torque, a high bandwidth current regulator is need. The TSF defined could be linear or nonlinear. In the paper [3], a rising and falling exponential TSF, referred as *m* function relating to rotor position θ

has been used. The reference phase voltage is obtained from rotating frame transformation technique under constant speed. The limitations in this technique are the assumption of magnetic linearity, constant speed operation and converter requirement for obtaining the spiky applied voltage. In the paper [4], Schramm et.al proposed a linear TSF in which the torque varies linearly during the commutation and the value of current in both incoming and outgoing phases are equal at the central commutation point. A high torque to current was obtained from this TSF. The nonlinear TSF like sinusoidal TSF and Cubic TSF were developed to obtain the smooth control and highest torque to current ratios [5, 6]. Xue et.al has carried out a critical evaluation of all four TSFs for torque ripple minimization with rate of change of flux and rate of change of current to be the evaluation criteria [7]. He further proposed the Torque Ripple Factor (TRF), a performance index to evaluate the effectiveness of TSF for ripple reduction. Genetic Algorithm is used to optimize the all the four TSF for minimum TRF and concluded that the cubic TSF yields less in both TRF and computation time. Thus, it can be inferred that nature of function and its interdependency on rotor position decides the smoother and ripple less torque production.

2.2 Current Profile Strategy

From equation (1), it can be understood that the value of torque is dependent on the phase currents and rotor position and it is evident to control the phase current profile through wave shaping technique for reducing the torque pulsation. In the paper [8], a self learning technique in which an accurate value of reference current required for obtaining the desired torque for various rotor positions could be obtained by physical measurements and stored as a loop up table. The total time taken for computation and conduction of test is very high. In the paper [9], Lovatt et.al has described for the first time a method for finding the optimum current waveforms for a given power and speed for an SRM with limited supply voltage and limited peak current, using computer search techniques. A prior knowledge about the static characteristics of the motor is required in LDT approach while determining the reference current [10, 11]. Reference [12] showed that the motor torque ripples could be effectively cancelled by injecting the harmonics of amplitude and spatial phase into the reference current at appropriate frequency. In the paper [13], the coefficient of added amplitude and spatial phases were found plotting surface of ripple variance. The surface was optimized for minimal value by Gradient Descent method in online process. It was found effective in reducing steady-state torque ripple, given a feedback signal proportional to torque ripple, over the full speed range. In reference [14], the author designed an effective PI controller for tracking the pre-optimized current references. A digital solution (Microcontroller or Digital Signal Processor) for hysteresis control eliminates the use of current feedback filters and limits the bandwidth of the current control loop.

As the developed in SRM is proportional to the square of the phase current, Nihat et.al presented a method for modulation of phase currents i_{ref} [15]. Such modulation leads to decrease in torque dip. Due to hysteresis current control, the system experiences subsonic noise which could be reduced by reducing the controller bandwidth and increasing the switching frequency. The above modulation strategy is implemented with Sliding Mode Control (SMC) for the speed control of the drive [16]. The

SMC used, adjust the i_{ref} and torque dips are eliminated with the modulation strategy. The low frequency vibration in torque has been eliminated by SMC.

In the paper [17], the author proposed a nonlinear internal model control in which the current error is passed through LPF and is added to compensate for plant model mismatches. Sahoo et.al proposed an iterative learning controller (ILC) for solving tracking control problems in nonlinear systems with limited plant knowledge, where the task is performed repetitively. In the above proposed controller, a feed forward control scheme together with a proportional current controller was employed to improve the tracking performance [18]. The ILC stores the control input and the plant output error at each operation cycle, and the control input was updated according to a learning law. The learning law ensures that the error is reduced from cycle to cycle until the desired level of accuracy was achieved. In the paper [19], the author proposed a novel technique Fuzzy Iterative Technique (FIT) modulates the reference current waveform iteratively by the use of the multiplying factor determined by fuzzy systems using torque error and rotor position as the two inputs. The FIT computes the incremental change (correction term) for the current profile at each iteration.

In the paper [20], the current profiling is carried out from flux linkage profiling or magnetization characteristics of motor using two dimensional B-spline neural networks (BSNN). The system proposed with BSNN does not necessitate the use of high bandwidth current controller and a torque sensor for initial training of the network. Henriques et al. [21] suggested a new method for shaping the motor phase currents to minimize the torque ripple using a neuro-fuzzy compensator. In this method, a compensating signal was added to the output of the proportional integral (PI) controller in the current regulated speed control loop. In [22], the authors developed a fuzzy system for maintaining the speed constant by controlling the current waveform. The reference current was then modulated by subtracting the output of the fuzzy system from the sum of phase currents computed at the previous sampling period. In [23], the author formulated a multi-objective optimal design problem for generating the reference current using GA. The results show that the four design parameters can be automatically selected by GA and much smoother current waveforms are generated when comparing with conventional TSF design using heuristic knowledge. In paper [24], authors proposed the concept of context based emotional learning for computing the reference phase current for maintaining the speed constant and inherently reducing the torque ripple.

2.3 Commutation Strategy

One of the reasons for torque ripple in SRM is rapid switching of phase currents into its windings, the method of switching the currents are performed either at optimal turn on (T_{on}) or turn off (T_{off}) angle. In paper [25], Iqbal et.al, devised a hybrid controller in which the concept of balanced commutation algorithm is infused in TSF with the control of Central commutation angle θ_c, which is varied between two limits as function of speed. In [26], the magnetization duration depends on turn off (T_{off}) angle, an offline calculation for optimum torque control at both chopping and full wave mode can be determined with an analytical model. In [27], the author proposed a compensation technique, by varying the value of turn off angle as a function of current and speed, using fuzzy logic control proved it to be effective in reducing the ripple. It was an off

line process and cannot be implemented for real time system due to difficulty in dynamic measurements. In paper [28], Fisch et.al presented the Pareto optimal optimization technique using GA for maximizing the value of turn on and turn off angles. The optimized firing angles were applied to an inverse model of SRM.

2.4 Multiphase Operation

The torque ripple produced in a unipolar operation is more. The total torque is the sum of sequential torque pulses produced by each phase alone which leads to existence of quite large torque ripple. Multiphase excitation of Switched reluctance motor with a special accent on three phase operation would certainly reduce the torque ripple [29]. Chris et.al analyzed the distribution of magnetic forces with multiphase excitation. Bipolar Excitation has produced more number of Short Flux Path Excitation (SFPE) which improves the average torque produced and reduces the core loss, vibrations and acoustic noise in the motor [30]. The Bipolar excitations to the phase winding improve the current profile and reduce the total harmonic distortion (THD) by 26.77% [31]. In a two phase excitation model discussed in [32], the torque ripple minimization is carried out using Fibonacci and exhaustive search methods. It was observed that Fibonacci and exhaustive search yielded closer results and ripple reduction is nearly 50% compared to the traditional constant current scheme. The Fibonacci search requires less storage space and it is preferred.

3 Conclusion

In this work, a review of the ripple reduction strategies in SRM is presented. It is observed from the study that the torque control should be on instantaneous basis for dynamic torque ripple reduction. The multiphase operation accounting mutual inductance and extension of Random Pulse Width Modulation (RPWM) in SRM are in progress.

References

1. Anwar, M.N., Hussain, I., Radun, A.V.: Comprehensive Design Methodology for Switched Reluctance Machines. IEEE Transactions on Industrial Applications 37(6) (November 2001)
2. Lawrenson, P.J., Stephenson, J.M., Blenkinsop, P.T., Corda, J., Fulton, N.N.: Variable-speed switched reluctance motors. Proc. Inst. Elect. Eng., pt. B 127(4), 253–265 (1980)
3. Illic-Spong, M., Miller, T.J.E., MacMinn, S.R., Thorp J.S.: Instantaneous Torque Control of Electric Motor Drives. IEEE Transaction of Power Electronics 2, 55–61 (1987)
4. Husain, I., Eshani, M.: Torque Ripple Minimization in Switched Reluctance Motor Drives by PWM Current Control. IEEE Transaction on Power Electronics 11(1) (January 1996)
5. Schramm, D.S., Williams, B.S., Green, T.C.: Torque Ripple Reduction of Switched Reluctance Motors by Phase Current Optimal Profiling. In: Proceeding of IEEE PESC 1992, Toledo, Spain, Vol. 2, pp. 857–860 (1992)
6. Sahoo, S.K., Panda, S.K., Xu, J.-X.: Indirect Torque Control of Switched Reluctance Motors using Iterative Learning Control. IEEE Transaction on Power Electronics 20(1) (January 2005)

7. Xue, X.D., Cheng, K.W.E., Ho, S.L.: Optimization and Evaluation of Torque-Sharing Functions for Torque Ripple Minimization in Switched Reluctance Motor Drives. IEEE Transactions on Power Electronics 24(9) (September 2009)
8. Kavanagh, R.C., Murphy, J.M.D., Egan, M.G.: Torque Ripple Minimization in Switched Reluctance Drives using Self Learning Techniques. In: Proceedings of IEEE IECON 1991, pp. 289–294 (1991)
9. Lovatt, H.C., Stephenson, J.M.: Computer Optimised Current Waveforms for Switched Reluctance Motors. IEE Proc. - Electr. Power Appl. 141(2), 45–51 (1994)
10. Ilic-Spong, M., Marino, R., Peresda, S., Taylor, D.G.: Feedback Linearizing Control of Switched Reluctance Motors. IEEE Transactions on Automatic Control 32, 371–379 (1987)
11. Stephenson, J.M., Hughes, A., Mann, R.: Torque ripple minimisation in a switched reluctance motor by optimum harmonic current injection. IEE Proceedings, Electrical. Power Application 148(4) (July 2001)
12. Wallace, R.S., Taylor, D.G.: A Balanced Commutator for Switched Reluctance Motors to reduce Torque Ripple. IEEE Transaction on Power Electronics 7(4) (October 1992)
13. Stephenson, J.M., Hughes, A., Mann, R.: Online Torque -Ripple Minimisation in a Switched Reluctance Motor over a Wide Speed Range. IEE Proceeding, Electrical Power Application 149(4) (July 2002)
14. Chapman, P.L., Sudhoff, S.D.: Design and precise realization of optimized current waveforms for an 8/6 switched reluctance drive. IEEE Trans. Power Electron. 17, 76–83 (2002)
15. Inanc, N.: Phase Current Modulation of Switched Reluctance Motor to Minimize Torque Ripple. Electric Power Systems Research 61, 51–55 (2002)
16. Inanc, N., Ozbulur, V.: Torque Ripple Minimization of a Switched Reluctance Motor by using Continuous Sliding Mode Control Technique. Electric Power Systems Research 66(3), 241–251 (2003)
17. Baoming, G., Xiangheng, W., Pengsheng, S., Jingping, J.: Nonlinear Internal-Model Control for Switched Reluctance Drives. IEEE Transaction on Power Electronics 17, 379–388 (2002)
18. Sahoo, S.K., Panda, S.K., Xu, J.X.: Iterative Learning-Based High-Performance Current Controller For Switched Reluctance Motors. IEEE Transaction on Energy Conversion 19(3), 491–498 (2004)
19. Gobbi, R., Ramar, K., Sahoo, N.C.: Fuzzy Iterative Technique for Torque Ripple Minimization in Switched Reluctance Motors. Electric Power Components and Systems 37(9), 982–1004 (2009)
20. Lin, Z., Reay, D.S.: Torque Ripple Reduction in Switched Reluctance Motor Drives using B-Spline Neural Networks. IEEE Transactions on Industry Applications 42(6) (November 2006)
21. Luis, O.A.P., Henriques, L.G.B., Rolim, W.I.S.: Torque Ripple Minimization in a Switched Reluctance Drive by Neuro - Fuzzy Compensation. IEEE Transaction on Magnetics 36(5) (September 2000)
22. Derdiyok, A., Inanc, N., Ozbulur, V., Ozoglu, Y.: Optimal phase current profiling of SRM by fuzzy logic controller to minimize torque ripple. In: Proceedings of the 1997 IEEE International Symposium on Intelligent Control, Istanbul, Turkey, vol. 1, pp. 77–82 (1997)
23. Xu, J.-X., Panda, S.K., Zheng, Q.: Multi-objective Optimization of Current Waveforms for Switched Reluctance Motors by Genetic Algorithm. In: Proceedings of the 2002 Congress on Evolutionary Computation, vol. 2, pp. 12–17 (2002)

24. Levachkine, S., Torres, M., Moreno, M., Quintero, R.: Knowledge-Based Method to Recognize Objects in Geo-Images. In: Negoita, M.G., Howlett, R.J., Jain, L.C. (eds.) KES 2004. LNCS (LNAI), vol. 3215, pp. 718–725. Springer, Heidelberg (2004)
25. Husain, I.: Minimization of Torque Ripple in SRM Drives. IEEE Transactions on Industrial Electronics 49(1) (February 2002)
26. Orthmann, R., Schoner, H.P.: Turn-Off Angle Control of Switched Reluctance motors for Optimum Torque Output. The European Power Electronics Association (1993)
27. Rodrigues, M., Costa Branco, P.J., Suemitsu, W.: Fuzzy Logic Torque Ripple Reduction by Turn-Off Angle Compensation for Switched Reluctance Motors. IEEE Transaction on Industrial Electronics 48(3) (June 2001)
28. Fisch, J.H., Yun Li Kjaer, P.C., Gribble, J.J., Miller, T.J.E.: Pareto-Optimal Firing Angles for Switched Reluctance Motor Control. In: Second International Conference on Genetic Algorithms in Engineering Systems: Innovations and Applications, GALESIA 1997 (1997)
29. Pillay, P., Liu, Y., Cai, W., Sebastian, T.: Multiphase operation of Switched Reluctance Motor Drives. In: Industry Applications Conference, Thirty-Second IAS Annual Meeting, IAS apos; 1997, Conference Record of the 1997, vol. 1(5-9), pp. 310–317 (October 1997)
30. Alkhazendar, M.A., Kader, F.M.A., Hawwas, Y.A., Kalas, A.E.: Performance analysis and realization of Bipolar Current Waveforms for a Switched Reluctance Motor. In: 12th International Middle-East Power System Conference, MEPCON 2008, vol. 12/15, pp. 387–391 (March 2008)
31. Erdington, C.S., Krishnamurthy, M., Fahimi, B.: Bipolar Switched Reluctance Machines: A Novel Solution for Automotive Applications. IEEE Transactions on Vehicular Technology 54(3) (May 2005)
32. Bhiwapurkar, N., Brekken, T.K.A., Mohan, N.: Torque Ripple Optimization of Switched Reluctance Motor Using Two-phase Model and Optimization Search Techniques. In: 37th IEEE Power Electronics Specialists Conference, PESC, pp. 1–6 (June 2006)

A New Scheme for Preprocessing Image Sequence in Wearable Vision System

Lu Han[1], Fangzheng Xue[1], Zushu Li[1], and Dongyi Chen[2]

[1] College of Automation, Chongqing University, Chongqing 400044, China
[2] School of Automation Engineering, University of Electronic Science and Technology of China Chengdu 610054, China
{hanlu,xuefangzheng}@cqu.edu.cn, zushuli@vip.sina.com, chenxigt@hotmail.com

Abstract. Cameras' miniaturization enable us feasible to capture photos and video at anywhere and anytime, even we wear them on human body and look the world in egocentric point of view. However, image quality degradation is caused by various factors in wearable environment, preprocess image acquired in wearable vision system is an unavoidable step. In this paper, we propose a new scheme to preprocess image sequence according to different quality of image frames in contiguous sequence. Firstly we detect image quality using hybrid methods, then classify them into some categories and given corresponding pre-process policy. We present experiment on a dataset collected within a cluttered environment, a prototype implementation validates its validity.

Keyword: wearable vision, preprocess image, blur detection, motion blur.

1 Introduction

Two technology's development drive the emerging wearable vision[1],[2],[3]. One is the cameras' miniaturization, so we are feasible to capture photos and video at anywhere and anytime. The other is development of wearable computing, so we easily wear the cameras on person's body and looking out at the world.

Capture high quality in wearable environment is a hard and challenge task, because the quality of image via wearable camera is relate to the motion of body, camera and human body's movement easier cause image quality degradation. There are many image quality problems are caused by the constantly and arbitrarily moving wearable cameras. Some sample images with degradation were acquired by head mounted cameras is showed in Fig.1.

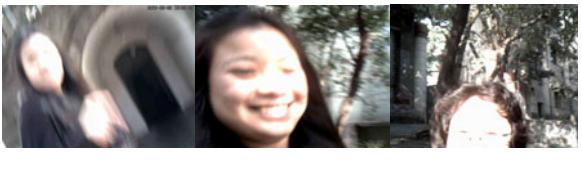


Fig. 1. Acquired image quality problems via wearable cameras: image motion blur, magnifying scaled face with out of focus; excessive light.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 295–301.
springerlink.com

The decrease of quality image even let the further image analysis doesn't work, object detection or recognition all need good quality image. Though the wearable camera could be consciously controlled by the user in reality application, for instance, key parts of the body well-positioned may help shoot good object and may help simplify vision problem. But it is notice that for continuous video sequences, most standard technique of object detection is simply to treat each frame as a still image, sometimes an object found in frame n+1 is the same object found in frame n, if the object doesn't move or detect wrong object, that means the system will keep on detecting and consuming vast resources all the while. So in resource-limited wearable system, we should consume fewer resources as effectively as possible to extract more useful information from the best quality image in frame sequence. Thus the key question is how to fast detect best quality image in frame sequence as showed in Fig.2.

Fig. 2. Four continuous video frames captured from wearable camera, the question is which one is the best quality we should prior process?

In this paper, we propose a new schema for pre-processing image in wearable vision system, the purpose is help resource-limited vision system to select the best quality image and repair lower quality image. Considering motion blur is the common problem in wearable computing, so we only pay close attention to motion blur in our first step work.

The organization of the remainder of this paper is as follows: the next section describes the schema design and pre-process strategy, then implementation of crucial module is presented in following section, mainly introduces relate works of blur detection and the special solution method for blur identification in wearable system, simulation section presents an example use case using the proposed schema and method. Finally, the conclusion and outlook provides summary and directions for ongoing and future work.

2 Schema Design

In this section, we designed series requirement policies for image preprocess in wearable vision system, the objective of which is to consider all preprocess requirement and select the best quality image for preprocessing different image according to different quality. The whole requirement policies are showed as follows:

1) When image is captured by wearable vision system, do image graying and blur detection, If image is no blur detected then enter the queen of image process, otherwise calculate the extend of blur and automatically sorted by blur extend.

2) According to application requirements and computational power, system will preset threshold of blurred images, the threshold will filter blurred image into different process policy: direct handling, handling after repair and cannot be repaired.
3) Little blur image could directly handling in the pending queue and was given priority 2, repairable image should do image inpainting and was given priority 3 in pending queue. Unrepairable image should adopt discarding, saving or sending to server policy according to application requirement.
4) All the images on the queue are in order of priority and waiting for the exposure and light balance of pre-processing methods, if in the previous the same scene of frame has been get recognized results, the remaining images in the queue as needed to save or discard.

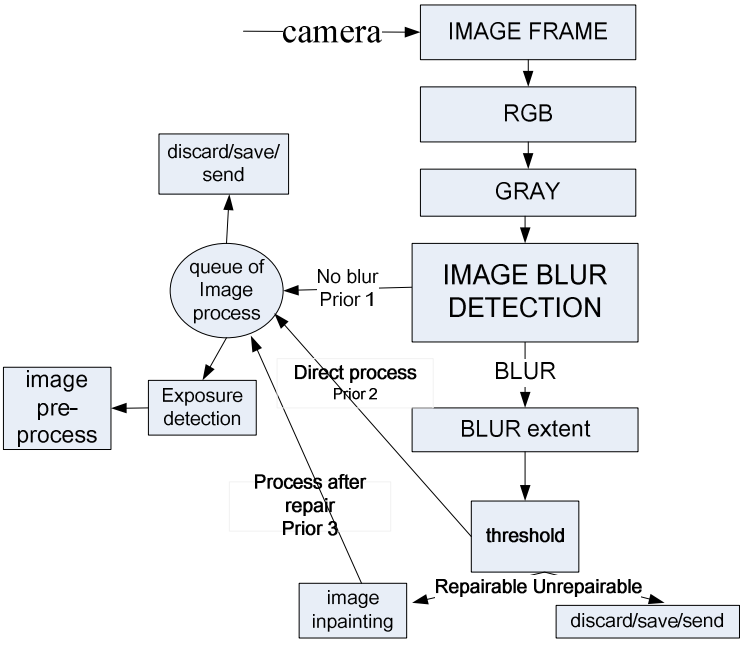


Fig. 3. The framework with process module of image quality selection policy.

According to above preprocess strategy, we have designed the pre-processing framework (see Fig.3) of selecting the best quality image. Through this framework we can capture the image and divide them into four categories: high-quality image without blur, good quality image with little blur, repairable low quality image and unrepairable bad quality image. In wearable vision system, motion blur is the most common degradation, so the key step of preprocess execute image blur detection.

3 Image Blur Detection

3.1 Previous Work

Motion blur is a common image degradation phenomenon in digital cameras, so many technologies have been proposed to enhance the quality of image, for instance, stabilized lens and shift-CCD/CMOS are popular and expensive hardware solution, and other software approaches for blur detection are proposed in recently years. Tone et al[4] proposed a scheme that makes the use of the ability of Harr wavelet transform in different types of edges. Reeves S et al. [5] simplified the identification problem by parameterizing the PSF and described a parameterized blur estimation technique using generalized cross validation, but computation complexity is high. Rooms et al. [6] assumed the PSF can be modeled with a single parameter, and they used a Gaussian function to estimate the single-parameter PSF from a single observation of a degraded image. Fergus et al. [7] adopted a variational Bayesian approach to estimate the PSF of an image. Chong et al [8] proposed a method based on the analysis of extrema values in an image. Hsu et al [9] propose a blur detector based on support vector machines to estimate the blur extent of an image. Ko et al [10] detect the blurred image by constructing the Bayes disciminant function[11] with the statistics of the gradients of the input image.

3.2 Proposed Method

In real application of wearable vision, wearable camera's motion relate to body's movement, so usually more than one degradation function affects the image blur. we hope system could fast detect the quality of image frame, so in this section, we propose a low-cost and simple approach using the statistics of image gradients to detect blur and classify image sequence into different prior label class[9][10].

In this paper, we calculate Magnitude of Gradient using first order derivative of Gaussian with 2-D Gaussian kernel. When Sigma=1.0, outputs Fig.4, which showed the comparison of good quality image, horizon motion and vertical motion blur image in original RGB image, gradient magnitude of the corresponding image and corresponding histogram image. Comparing Fig.4-A2, B2 and C2 and Fig.5., we intuitively notice that once the image is blurred, its amount of gradients is decreased and the mean and standard deviation are also decreased. Fig.3-A3, B3 and C3 shows the histogram of gradients image, we notice that the gradient magnitude distribution of a blurred image is almost empty on the large values and there only exists small values. This suggests that the statistics of the image gradients can be used for detecting the blurred images.

In Fig.5, we present the distribution of comparison of Sigma=1.0 and Sigma=0.2. Obviously, blurred images have smaller the mean and standard deviations than the sharp in this distribution, and smaller sigma distribution could be easier classified. With these statistics, we use the Bayes linear discriminant function [10][11] to detect images for two class: blurred and no blur images.

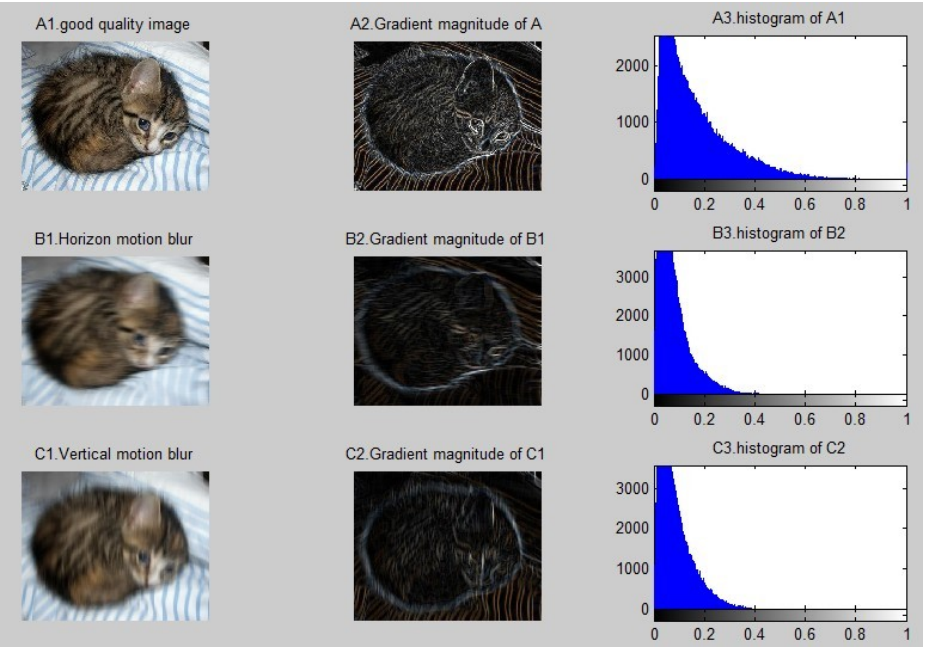


Fig. 4. Comparison of good quality, horizon motion blur and vertical motion blur images in corresponding gradient magnitude and histogram image.

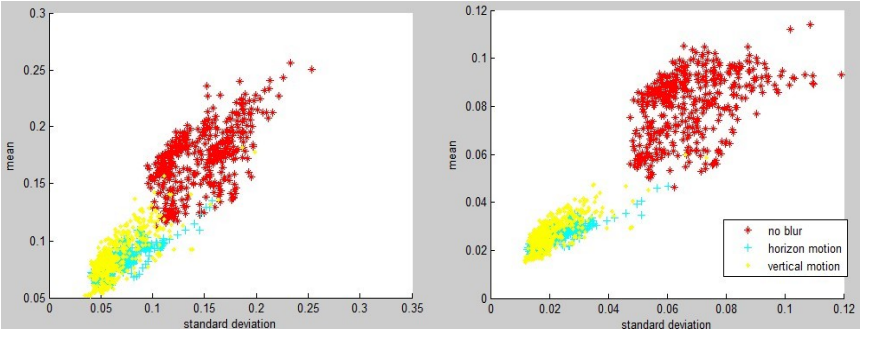


Fig. 5. Distribution of the statistics of the gradient magnitude: no blur image, horizon motion image and vertical motion image. (left) sigma=1.0 (right) sigma=0.2.

4 Simulate Experiment

As mentioned in section 2.1, we use 100 blurred images and 100 sharp images captured by wearable vision system as the training data, see Fig.6. From this training data, we

can calculate the mean vector and the covariance matrix of the input vector x=[stdev mean]^T for the blur and no blur classes, and apply[11] to detect the blurred images. We prepare 125 test image sequence frames (see Fig.7) to examine the proposed method and the result of accuracy rate is about 80.8%.

Fig. 5. The training data with (1st row)different direction blur images and (2th row) no motion blur images.

Fig. 6. The test image sequence with different scene were captured in wearable vision system.

5 Conclusion

In this paper, we propose a new schema for pre-processing image in wearable vision system for the goal of selecting the best quality image and repairing lower quality image. Considering motion blur is the common problem in wearable computing, so we realize the motion blur detection in our first step work. We have provided experimental results showing the proposed scheme is efficient. A future extension of our work will be to optimize the blur detection classifier and stabilize the motion blur images.

Acknowledgment. This research was supported by the National Natural Science Youth Foundation of China (No. 60905053).

References

1. http://www.seattle.intel-research.net/egovision09/
2. Mann, S.: Wearable Computing: A first step towards personal imaging. IEEE Computer 30(2), 25–32 (1997)
3. Mayol, W.W., Ben, J., Murray, D.W.: On the choice and placement of wearable vision sensors. IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans 39, 414–425 (2009)
4. Tong, H., Li, M., Zhang, H., Zhang, C.: Blur detection for digital images using wavelet transform. In: Proceedings of IEEE 2004 International Conference on Multimedia and Expo., pp. 17–20 (2004)
5. Reeves, S.J., Mersereau, R.M.: Blur identification by the method of generalized cross-validation. IEEE Transactions on Image Processing 1(3), 301–311 (1992)
6. Rooms, F., Philips, W., Portilla, J.: Parametric PSF estimation via sparseness maximization in the wavelet domain. In: Proceedings of SPIE Wavelet Applications in Industrial Processing II, pp. 26–33 (2004)
7. Fergus, R., Singh, B., Hertzmann, A., Roweis, S.T., Freeman, W.T.: Removing camera shake from a single photograph. ACM Transactions on Graphics 25, 787–794 (2006)
8. Chong, R.M., Tanaka, T.: Image Extrema Analysis and Blur Detection with Identification, In: IEEE International Conference on Signal Image Technology and Internet Based Systems, pp. 320–326 (2008)
9. Hsu, P., Chen, B.-Y.: Blurred Image Detection and Classification. In: Satoh, S., Nack, F., Etoh, M. (eds.) MMM 2008. LNCS, vol. 4903, pp. 277–286. Springer, Heidelberg (2008)
10. Ko, J., Kim, C.: Low cost blur image detection and estimation for mobile devices. In: Proceedings of the 11th International Conference on Advanced Communication Technology, Gangwon-Do, South Korea, vol. 3, pp. 1605–1610 (2009)
11. Duda, R., Hart, P.E., Stock, D.G.: Pattern Recognition, pp. 29–45. Wiley Interscience, Hoboken

Wrapper Roll Control System of Hot Strip Mill

Haiying Liu[1] and Long He[2]

[1] College of Information, Hebei Polytechnic University
063009, Tangshan, Hebei, China
pearllhy@sinar.com
[2] Automatic Company, Tangshan Iron and Steel Co,.L td.
063000, Tangshan, Hebei, China
Helios315@sohu.com

Abstract. Wrapper roll is one of the most important components in rolling coiler. And it has great relation to the quality of the final products. Based on the programming and optimization experience, a hot strip coiler with new coiling process, which operates with position control and pressure control is introduced. Control method and operational mode are given and auto jump control is described specially. The configuration of the system will be more compact. It acts slightly, obviously reduces dynamic loads and noises, and improves the surface quality of strip. the wrapper roll control system is introduced in detail.

Keywords: hot strip, wrapper roll, AJC control.

1 Introduction

With the large demand of hot strip in the market, high quality of the hot strip becomes more important. Down Coiler area is the finished product area of the hot strip area and has great influence on the quality of band steel. Wrapper roll is the most important parts of the whole system. The high precision of position and force control of wrapper roll control system enhances the quality of band steel wrapping.

2 Control Method

There are two control methods in wrapper roll control. One is position control which is called CPC (constant position control), and the other is pressing force control, which is called CPR (constant pressure rolling).

2.1 Position Control

Wrapper roll (WR) position is controlled by comparing WR gap setting with the actual gap. The actual gap is detected by rotary type Manescale which is installed at pivot of wrapper arm. The WR gap setting is converted to rotation angle of WR arm

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 303–309.
springerlink.com

in the DCPC unit. DCPC unit calculates the deviation between WR arm rotation angle setting and WR arm actual rotation angle. The deviation is multiplied by the control gain. And then, DCPC unit outputs as servo valve opening amount command. The illustrated scheme is as followed.

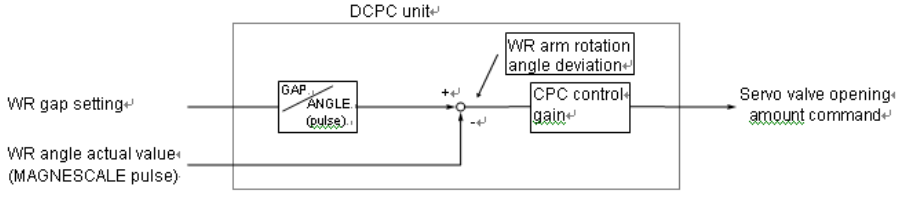


Fig. 1. Position control

2.2 Pressing Force Control

Pressing force control is illustrated in figure 2. WR pressing force is controlled by comparing WR force setting with the actual force. The actual force is detected by pressure transducers which are installed at cylinder head side and rod side. The WR force setting is converted to cylinder pressure considering WR position and WR arm weight. in the DCPC unit. DCPC unit calculates the deviation between WR pressure setting and actual pressure. The deviation is multiplied by the control gain. And then, DCPC unit outputs as servo valve opening amount command.

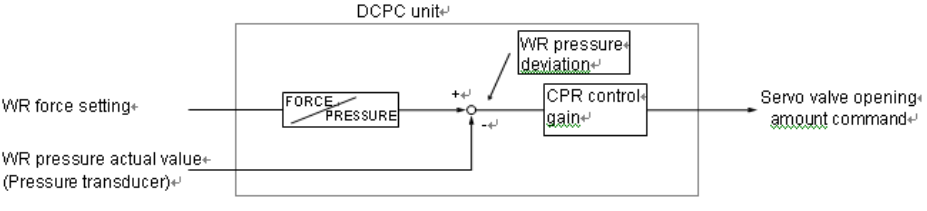


Fig. 2. Pressing force control

3 WR Control System

3.1 AJC Control System

AJC (automatic jumping control) control system consists of PLC portion and D-CPC unit portion. PLC outputs the reference and the command to D-CPC unit and D-CPC unit consists of position control loop (CPC) or pressing force control loop (CPR).

3.2 Position Control Loop

Wrapper roll gap position control system compares the feedback with the reference in the unit of wrapper roll rotation angle converted from wrapper roll gap. Gap reference

from PLC gap reference is made in the DCPC unit during AJC etc. is converted to the reference of wrapper roll arm rotation angle. The other hand feedback signal is obtained by the D-CPC unit by counting pulses from rotary MAGNESCALE installed at the wrapper roll arm. The between reference and feedback is multiplied by the difference control gain "G". and then this multiplied value is output to servo valve via servo amplifier as servo valve opening amount command.

The above mentioned position control and pressing force control mentioned below are done every 1msec in the D-CPC unit.

The above mentioned is fundamental position control function, in addition to this there are speed compensation function and null compensation function as auxiliary function. Speed compensation function is the function of make the response faster and to prevent vibration of wrapper roll by detecting rotation speed of wrapper roll arm and outputting some open or close amount command as compensation to servo valve.

Null compensation function is the function to improve the wrapper roll gap setting error caused by null point shift of the servo valve. The difference between the reference and the feedback are input null compensation circuit and from this circuit the some open or close amount command as compensation is output to servo valve.

3.3 Pressing Force Control Loop

When pressing force control is done, pressing force reference and pressure of head side and rod side of the wrapper roll cylinder are converted to the pressing pressure of head side. The difference between the reference and the feedback is multiplied by the control gain "G" and then this multiplied value is output to servo valve via servo amplifier as servo valve opening amount command.

Straight pressing force direction of the wrapper roll cylinder is different from the direction of wrapper roll toward the mandrel. In other word the direction changes by wrapper roll gap. Therefore wrapper roll pressing force reference conversion to cylinder pressure reference (the pressing pressure of head side) is done considering wrapper roll gap in the D-CPC unit.

Pressure transducers installed at head side and rod side of wrapper roll cylinder are used as feedback signal. However rod side pressure signal is converted to head side pressure considering the difference of cylinder area of head side and rod side. And the cylinder pressure feedback is obtained as the pressure converted to head side pressure.

4 Jumping Control

4.1 AJC Control

Automatic jumping control is carried out to minimize the top mark of the strip. The system is illustrated in figure 3. The top mark occurs when the strip head end hits against the wrapper roll. Therefore control concept of automatic jumping control (AJC) is as follows.

The wrapper roll jumps (start position control based on step reference) before strip head end reaches to the wrapper roll. After the strip head end passes through the wrapper roll position, wrapper roll goes down towards the mandrel and starts pressing force control (CPR). The jump timing and the pressing force control timing are made by strip head end tracking function.

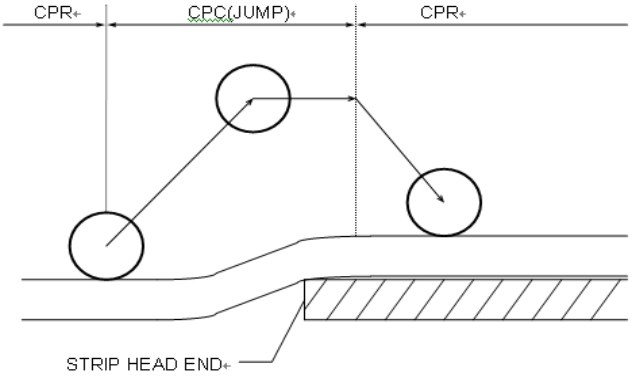


Fig. 3. AJC control

4.2 AJC Control Process

When the strip head end breaks both laser sensor beams (in normal operation, both OS and DS laser are used for detecting the strip head end), the strip head end tracking for AJC is started to generate wrapper roll jump and press timing signal for each wrapper roll.

The pulse signal from PLG attached to the bottom pinch roll is used for this tracking. Wrapper rolls start jumping and pressing in accordance with the timing signal generated as above. One example of the wrapper roll operation during the strip coiling with AJC (JUMP) is given. When the strip head end breaks both laser sensor beams (in normal operation, both OS and DS laser are used for detecting the strip head end), the strip head end tracking for AJC is started to generate wrapper roll jump and press timing signal for each wrapper roll, same as AJC selected. (The strip head end tracking is performed, and wrapper roll jump and Press timing signals are generated even though AJC selection is off.)

4.3 Jump Timing and CPR Timing Compensation

For prevention of the WR to hit against the strip head end at AJC coiling, wrapper roll must reach at preset jump position before the strip head end reaches hit point which is calculated in SH.93. The Jump timing and CPR timing are calculated based on the following formulas.

$$L_{jti}=V_{PR}\bullet T_{ji}+\gamma_3=V_{PR}(T_{jbi}+\Delta T_{ji})+\gamma_3 \ \ [mm] \tag{1}$$

$$L_{cti}=\gamma_2 \ \ [mm] \tag{2}$$

Relation between jump amount and time to jump is as the following formulas.

$$T_{ji}=T_{jbi}+\Delta T_{ji}=T_{jbi}(1+\alpha_{jti}) \ \ [sec] \tag{3}$$

Where T_{jbi} is time to reach preset jump position under the condition, that serve valve is not in full open, α_{jti} is coefficient of jump timing compensation at aturation of serve valve, ΔT_{ji} is time compensation, V_{PR} is bottom pinch roll speed at HMD

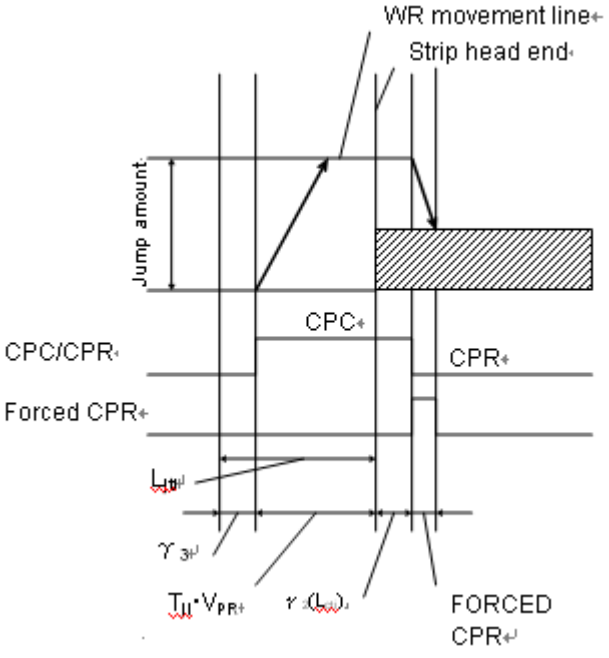


Fig. 4. Jump timing and CPR timing compensation

on timing, γ_3 is jump start compensation length(preset in AJC panel) and γ_2 is CPR start compensation length(preset in AJC panel).

WR calibration is conducted to match the actual gap between WR and MD and the WR gap actual value memorized in the AJC control panel.

WR calibration operation is required in the following cases: After starting up the AJC control panel (power on) including the replacement of the control unit and the position sensor; After starting up the PC in the AJC control panel (running the PC from its stop status); After replacing WR; After replacing MD and MD segment.

WR gap is controlled using WR arm rotation angle as feedback signal. Therefore, converting WR gap to WR rotation angle or WR rotation angle to WR gap are necessary for WR position control. The formulas for conversion are as follows.

WR gap to WR rotation angle (pulse) conversion is as followed.

$$\theta = \cos^{-1}\left\{\frac{L1^2 + L2^2 - \left(\frac{Dm + Dw}{2} + G\right)^2}{2 \times L1 \times L2}\right\} - \alpha \tag{4}$$

$$\text{Magnescalenpulse} = \frac{\theta}{360} \times 100000 \tag{5}$$

WR rotation angle (pulse) to WR gap conversion is as followed.

$$\theta = \frac{\text{Magnescalepulse}}{100000} \times 360 \tag{6}$$

$$G = \sqrt{L1^2 + L2^2 - 2 \times L1 \times L2 \times \cos(\theta + \alpha)} - \frac{Dm + Dw}{2} \tag{7}$$

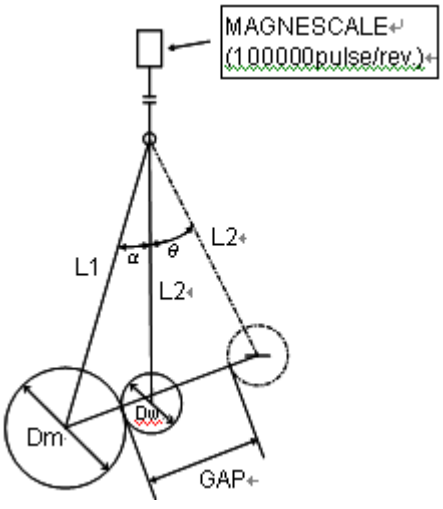


Fig. 5. WR gap to WR rotation angle conversion

Where Dm is Mandrel diameter, Dw is WR diameter, G is WR gap, θis WR arm rotation angle from WR gap "0" and αis WR arm rotation angle (WR gap "0") based on position of center of mandrel

5 Conclusion

The hot strip is drived by cylinder that has fast response speed and high control precision. It help the wrapper roll operate effectively and enhance the quality of band steel. The control system has been one of the most important functions of the whole control system.

References

1. Zhang, D., Zhou, F.: TDC-based wrapping roll automatic jump control system for a 1500mm hot strip mill coiler. Journal of University of Science and Technology Beijin. (6) (2007)

2. Li, Y-l., Liu, X.: Design and Research on AJC Hydraulic Servo Controlling System of Hot-rolling Coiler. Journal of Chongqing Institute of Technology (Natural Science Edition) (5) (2007)
3. Zhang, L., An, R.-c.: Introduction of Critical Equipment of Tangsteel 1810mm Semi-endless Rolling Line. Steel Rolling (01) (2007)
4. Wang, Y.-q., Lin, B.-j.: Dynamic characteristics analysis of automatic pressure control system in wrapping roller hydraulic jump system. Heavy Machinery (06) (2006)
5. Zhao, B.-l., Wang, H.-b., Ma, Z.: Simulation and Test Research of Electro-hydraulic Servo System in Coiler. Mechanical Engineering & Automation (06) (2005)
6. Lv, X.-y.: The AJC Jump Time Calculation and Control Analysis of Coiler Wrapper Roll. Shanxi Metallurgy (01) (2005)

Analysis and Amend Online Certificate Service Protocols

Xiaojun Deng[1], Gelan Yang[2], Zhiqiang Wen[1], Tao Zhang[1], and Bing Deng[1]

[1] Hunan University of Technology, College of Computer and Communication, Zhuzhou, China
[2] Hunan City University, Department of Computer Science, Yiyang, China
little_army@139.com

Abstract. Digital certificate is the key element that implements trust and trust authentication in e-government and e-commerce, however CA may be canceled and revoked in advance by accident in practical applications. Therefore it is critical for the users of this certificate to obtain the latest certificate status as soon as possible, it is also critical for realizing credibility in PKI system. It summarizes the advantages and disadvantages in PKI system, and discusses the difficulties in practical application by analyzing the OCSP protocol. Some corresponding methods are proposed to solve the above problems.

Keywords: PKI, digital certificate, certificate revoke, OCSP.

1 Introduction

As e-commerce and e-government continued popularity to ensure transmission of digital information security, in addition to adopting stronger encryption algorithm, the need to establish trust and mechanism of trust authentication, that the parties in participating e-commerce must have an identifier that can be verified, which is a digital certificate. Digital certificate is unique, and its public key associates with the entity itself. The authentication of certificates solves the security issues in online trading and settlement, Including the establishment of e-commerce trust relationship between the subjects, namely the establishment of safety certification authority(CA) and selection of secure protocol (such as SET, SSL)[1]. CA issues Public key certificate to user which has expiration data. However, in many cases the validity of the certificate may be invalid before the expiration data. For example: certificate subject name changing, the relationship of CA and the subject of certificate changing (such as an employee and his organization end the employment relationship), the private key of the leaked, destroyed or lost and so on. In these cases, the certificate must be canceled and revoked, Then the user of using this certificate should as soon as possible to obtain the latest certificate status which is critical for realizing credibility in PKI system.

2 The Theory of OCSP

Nowadays, the revocation of PKI system is widely implemented by CRL. A general description of CRL is in [2]. This system works as the issuance of certificate does, which means the information without trust of correspondence and service system

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 311–317.
springerlink.com

needn't encryption. This method is easily implemented and is widely adopted because of its low requirements of system resources. But, the shortage of this method is obvious. Because CRL is issued at regular intervals, if one certificate is abolished, the message of revocation has to be apprized in the coming cycle. Meanwhile, because of the existence of all information of revocations, CRL will become very huge with more and more revocation information. With the increase of the size of CRL, the verification period will be longer, if the latest CRL is continuously downloaded, it will suffer great network losses and result in CA server inefficiency and network congestion, which keeps clients from getting CRL information. On account of the periodicity of information issued by CRL, time delay between certificate being revoked and clients getting message occurs. The creditability of PKI is confronted with great challenges because of the time delay. Some special discussion on improving methods is presented in [3] and [4]. OCSP can get the information whether the certificate has been revoked in time. As the supplement of CRL, the application can test "revoked" state on account of OCSP. OCSP allows the application to determine the "revoked" state of a certificate. Contrasting to CRL, OCSP can meet many requirements and provide quicker "revoked" message and other states information. A general description of OCSP is in [5].The two main advantages of OCSP are: efficiently reducing the network load and providing timely state information of a certificate for clients. Chart 1 describes the interaction between clients and OCSP Responder.

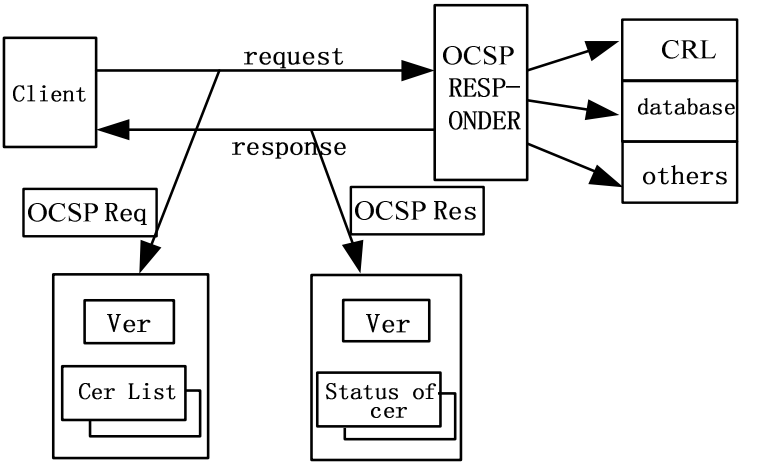


Fig. 1. Interactive process of OCSP

The protocol of OCSP is an easy request/response protocol, which provides on-line revocation by means of the trusted third party. An OCSP client sends a request for state detection to an OCSP responder and keeps the certificate in receiving order until receiving a response from Responder. A request of OCSP consists of a protocol version number, type of service request and one or more certificate identifiers. A certificate identifier consists of identifying names of the certificate authority organization, the hash values of shared key of the certificate authority organization and serial

number of the certificate and additional expansion. Response consists of certificate identifiers and certificate states, namely, normal, revoke or unknown. If a certificate is in a revoked state, the exact time and the reason for revoking are required to describe.

```
OCSPResponse::=SEQUENCE {
        responseStatus          OCSPResponseStatus,
        responseBytes           [0] EXPLICIT ResponseBytes OPTIONAL }
    OCSPResponseStatus ::= ENUMERATED {
        successful              (0),    —Response has valid confirmations
        malformedRequest        (1),    —Illegal confirmation request
        internalError           (2),    —Internal error in issuer
        tryLater                (3),    —Try again later
                                        —(4) is not used
        sigRequired             (5),    —Must sign the request
        unauthorized            (6),    —Request unauthorized
    }
```

3 OCSP Analysis

OCSP can provide timely and latest information for certificate state, but it doesn't mean the response from OCSP is zero-delay contrasting to the current state of certificate. OCSP protocol does not give a clear definition on the background database for collecting the revoked information. As Chart 1 shows, the background of OCSP still adopts CRL and other similar means for collecting the revoked certificate information. The real-time performance of OCSP Responder is determined by the time-delay of gathering the information. Therefore, we cannot simply think OCSP responder can update information automatically and provide timely service.

OCSP response must be guaranteed by digital signature from a trusted party and it is not tampered during transmitting process. The third party signature may be a certificate authorized by CA or an entity admitted by CA and clients must acquire the certificate copy of the public key. In order that a proper responder is conveniently achieved by clients, authoritative information access extending field of X.509 shared-key certificate indicates the address of OCSP Responder, which means the localities of OCSP Responder are attached to certificate.

Because of the adoption of request-responder, OCSP has no need to distribution of CRL and it can eliminate the distribution limitations of CRL; Because every responder from OCSP is smaller than CRL, OCSP responds to information model that relies on C/S model, which supports more clients; OCSP adopts HTTP, LDAP and so on, which travels the protocols of TCP/IP, so the configuration and realization of OCSP can be conveniently achieved; Because the requester makes information requests for a particular certificate not an invalid certificate list, OCSP provides more effective solutions; In OCSP the time of sending request and the responding of showing certificate valid can gain the resisted denial to trades history.

In spite of the mentioned advantages, OCSP in nature relies on CRL, which results in some disadvantages. OCSP Responder gets the revoked information of certificate through the periodical CRL; If plenty of requests are submitted to the Responder in a

definite interval, OCSP Responder will easily be attacked by DoS. And the Responder need to sign for responses, which consumes the vast majority of CPU time; because the Responder does not sign for returned false notifications, attackers would fake returned false notifications to attack. The protocol indicates certificate state information comes from Responder, but it does not definitely indicate how to get certificate state information. As for clients of OCSP Responder, the best approach to getting the information is that CA directly transmits it to Responder. According to the relation between CA and OCSP Responder, certificate authority (CA) organization can forward revoked information of information and then offers it to clients immediately. However, in practice most Responders gain revoked information of certificate through periodical CRL.

4 Solutions to the Problems

On the basis of the above analyses if a reply that meets a demand of real-time state needs to be realized, that is to say providing a CRL that has not been stored, it is from a real-time certificate storage. Allow Responder offers high performance extended information without CRL. In abstract terms, the responder is providing an implementation of an authenticated dictionary that responds to membership queries from relying parties. A conventional OCSP responder answers the question "Is x excluded from D?", while an improved OCSP responder answers the question "Is x present in D?"[6]. When returning a response, the responder only indicates that request certificate that is present in its valid certificates, a definite verified list that does not verify certificate in any way and operates directly on CA certificates storage.

4.1 Some Improving Methods

OCSP defines a complex certificate identifier as a part of certificate. Some parts are transformed by Hash and some are not, and even need to regard datum of other certificates as a part of identifier, which makes it very difficult verifying a unique certificate. The main goal is providing an identifier that is easily and widely adopted by all certificates, and thereby neglects their model and coding. Meanwhile, the new certificate identifier can be compatible with traditional OCSP certificate identifier. Therefore, the definition of new identifier can be adopted by new protocol as follows to extend existent OCSP identifiers:

```
NEWIdentifier ::= SEQUENCE {
        certHash        OtherHash,
        legacyID        IssuerAndSerialNumber OPTIONAL   }
```

SHA-1 arithmetic is adopted by certHash to hash certificate, which applies fingerprint system or other similar systems. And this operational and correct verification is easily achieved. Here Othterhash adopts sha1Hash ::= OCTET STRING SIZE(20) directly or other methods in [7]. LegacyID provides feasible methods that is based on traditional CRL. Its complete definition makes reference to [8]. This identifier can easily develop from X.509 and it can be widely applied in CMS and S/MIME. Regardless of

whether this identifier applies a traditional method or in an uncertain state, this identifier can be adopted. And this method can be run in constrained environments.

Because of its CRL-based origins, OCSP can only return a negative response. For example, when fed a freshly-issued certificate and asked "Is this a valid certificate", it can't say "Yes" (a CRL can only answer "revoked"), and when fed an Excel spreadsheet it can't say "No" (the spreadsheet won't be present on any CRL). This problem interacts badly with the one mentioned above, so there is no way to confirm what the actual problem is. The second major design goal of OCSP then is to provide a clear, unambiguous response to any query, either "This certificate is definitely valid right now", "This certificate is definitely not valid right now", or "The object you have queried doesn't exist"(The standard OCSP cannot perform the operations mentioned above.). The new one can apply basic and extended responses. The definition of basic response is as follows:

```
NewResponseBasic ::= SEQUENCE {
        certHash          OtherHash,
        status            BOOLEAN,
        extensions        Extensions OPTIONAL     }
```

A returned value 'true' indicates that the certificate is valid right now. A returned value 'false' indicates that the certificate is not valid right now. This is a clear, unambiguous response that is useful for the clients who want to know definitely whether they can safely use it or not. Relying parties who require further information should use the extended response. The definition is as follows:

```
RESPONSEINFO ::= CLASS {
      &status          CertStatus UNIQUE,
      &StatusInfo      OPTIONAL
      } WITH SYNTAX  { &status [WITH DETAILS IN &StatusInfo] }
  NewResponseExtended ::= SEQUENCE {
      certHash            OtherHash,
      status              RESPONSEINFO.&status({ CertStatus }),
      statusInfo          RESPONSEINFO.&StatusInfo({ CertStatus }{ @status }),
      extensions          Extensions OPTIONAL     }
   ResponseTypes RESPONSEINFO ::= {
      { statusOK                                } |
      { statusRevoked       WITH DETAILS IN RevocationInfo } |
      { statusSuperseded WITH DETAILS IN SupersededInfo } |
      { statusUnknown                           },
         }
     CertStatus ::= ENUMERATED {
      statusOK (0),
      statusRevoked (1),
      statusSuperseded (2),
      statusUnknown (3),
      ...
      }
```

The consistency of statusOK value with the basic response shows certificate valid. If the certificate has been revoked or rendered invalid in some form, the responder will return a "revoked" response. The definition of RevocationInfo is as follows:

```
RevocationInfo ::= SEQUENCE {
    revocationTime        RelativeTimeInfo OPTIONAL,
    revocationReason      CRLReason OPTIONAL     }
```

RevocationTime indicates the time at which the revocation or invalidation took place, if available. RevocationReason provides the reason why the certificate was revoked or rendered invalid, if available.

OCSP uses timestamps for all responses, assuming that the relying party and responder somehow have perfectly synchronized clocks. This is almost never the case, with systems having been encountered with clocks that are as much as decades out of sync. The new method does not rely on synchronized clocks for its operation.

This method eliminates the need that the responder and relying party are in need of precisely synchronized clocks. The relying party may use the absolute revocation time if they have a mechanism for precise clock synchronization with the responder, or the difference between the two times to determine how far in the past relative to its own clock the revocation took place. The definition is as follows:

```
RelativeTimeInfo ::= SEQUENCE {
    localTime             GeneralizedTime,
    timeValue             GeneralizedTime     }
```

5 Conclusion

This paper solved the problems of the excessive complexity of certificate identifier of OCSP, unclearness and inaccurateness of response, an overreliance on responder and the disadvantages of synchronized clocks. And the new protocol can be applied better in some resource-constrained environments because of these improvements. The adoption and operation of these improvements can also solve the existent problems of a real-time certification querying mechanism, which will be useful for the development and adoption of PKI. We only accept references written using the latin alphabet. If the title of the book you are referring to is in Russian or Chinese, then please write (in Russian) or (in Chinese) at the end of the transcript or translation of the title.

Acknowledgment. This work was supported by Scientific Research Fund of Hunan Provincial Education Department under Grant (No. 10C0622, No. 10C0498).

References

1. Deng, X.: Analysis and research mechanism of certificate revocation. Computer Engineering and Design 28(7), 1538–1540 (2007)
2. Housley, R., Polk, W., Solo, D.: Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile[S]. RFC 3280 (2002)

3. Liu, X., Jiang, X.: Study of a certificate status information distributi on approach based on Over-Issued CRL mechanism. Journal of Zhejiang University (Engineering Science) 36(2), 224–226 (2002)
4. Zheng, Z., Yu, Z.: An improved CRL schema based on segementation. Computer Application and Software 26(12), 271–273 (2009)
5. Myers, M., Ankney, R.: X.509 Internet Public Key Infrastructure Online Certificate Status Protocol – OCSP. RFC 2560 (1999)
6. Gutmann, P.: X.509 Internet Public Key Infrastructure Real-time Certificate Status Facility for OCSP. RFC draft (2002)
7. Pinkas, D., Pinkas, D., Pope, N.: Electronic Signature Formats for long term electronic signatures. RFC 3126 (2001)
8. Housley, R.: Cryptographic Message Syntax (CMS). RFC 3369 (2002)

Contourlet-2.3 Retrieval Algorithm Using Absolute Mean Energy and Kurtosis Features

Xinwu Chen and Liwei Liu

College of Physics and Electronic Engineering
Xinyang Normal University
Xinyang, China
chenxinwu@126.com

Abstract. To improve the retrieval rate of contourlet texture image retrieval system, a contourlet-2.3 transform based texture image retrieval system was proposed. In the system, the contourlet transform was contourlet-2.3, a new version of the original contourlet, sub-bands absolute mean energy and kurtosis in each contourlet-2.3 sub-band were cascaded to form feature vectors, and the similarity metric was Canberra distance. Experimental results on 109 brodatz texture images show that using the features cascaded by absolute mean energy and kurtosis can lead to a higher retrieval rate than the combination of standard deviation and absolute mean energy which is most commonly used today under same dimension of feature vectors. Contourlet-2.3 transform based image retrieval system is superior to those of the original contourlet transform, non-subsampled contourlet system under the same system structure with same dimension of feature vectors, retrieval time and memory needed.

Keywords: content based image retrieval; contourlet-2.3 transform; texture image; retrieval rate; contourlet transform; non-subsampled contourlet transform.

1 Introduction

With the fast development of internet and all kinds of imaging system technology, image resources are expanding more quickly than ever, the classic retrieval approaches using keywords can not describe the visual characters in the query image such as color, texture and contour. To overcome the difficulties of keyword retrieval systems, a new type of retrieval system called content-based image retrieval (CBIR) system was proposed[1],[2]. In the CBIR system, before retrieval work, every image in the image database which will be retrieved should be represented with a feature vector, all the vectors should be placed together to form a feature vector database, that is, a feature vector is used to represent a real image and is linked to the corresponding "true image". When a query image is input, the retrieval system will extract its features to form a feature vector which is used to compare the similarity between the query vector and each vector in the database, the N number most similarity vector will be chosen as the retrieval result. The most important technology of the CBIR system

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 319–326.
springerlink.com

is feature matching which includes three aspects: some certain transform (e.g. wavelet transform), feature extraction and similarity measure. During the past ten years, wavelet transform has played an important role in the system due to its good characters of multi-scale and local time-frequency [3], [4], [5]. Yet, some disadvantages of wavelet transform including shift sensitivity and the lack of directionality limits its abilities in texture representation. To overcome the deficiencies of wavelet transform, researchers have developed many improved approaches, such as: ridgelet, curvelet, beamlet, contourlet, bandelet, etc. In the family of "X-let", contourlet transform (CT) [6] is more acceptable greatly because of its easier implementation and strong ability in direction information representation. Since the transform was proposed in 2002 by Do, several modified versions have been proposed and form a new family including non-subsampled contourlet transform (NSCT), semi-subsampled contourlet transform [7], and a sharper frequency localization contourlet version [8], etc. Non-subsampled contourlet transform which was proposed by Cunha in 2005 has higher shift insensitivity level than the original contourlet transform but has higher redundancy as described by (1), where S denotes the scale number of the transform. The high redundancy makes the transform much more time consuming and much larger memory needed.

$$Re = 1 + \sum_{s=1}^{S} 2^s . \tag{1}$$

To overcome the limitation of high redundancy, Cunha presented a compromise transform which was a cascade of non-subsampled Laplacian pyramid and critical subsampled directional filter banks, and made the redundancy fall to S+1, Here we call the transform Contourlet-S, and CTS as abbreviation. To further reduce the redundancy of the transform, and aliasing, Lu and Do proposed a modified version of the original contourlet transform including three different variants and their redundancy ratio are approximately 2.3, 1.6 and 1.3, respectively. For convenience, we call them contourlet-2.3, contourlet-1.6 and contourlet-1.3. In [8], Lu and Do announced that contourlet-2.3 performs better than the other two versions in image de-noising application. In this paper, we will use contourlet-2.3 to implement a new retrieval algorithm.

Ever since the contourlet transform was proposed, many literatures reported the application approaches in many different areas including CBIR systems [9], [10]. But the original based retrieval system has very limited retrieval rate due to the drawbacks as mentioned before.

On the other hand, all the literatures as we know use absolute mean energy and standard deviation of each sub-band coefficients as features, which will be shown has low retrieval rates. Here, in this work, we will use absolute mean energy and kurtosis as features.

The remaining parts of this paper are organized as follows: key techniques of contourlet-2.3 texture image retrieval algorithm will be covered in section 2, experimental method and results will be shown in section 3 and in the section 4, the last section, we will conclude the whole paper and give some future work directions.

2 Key Techniques of Contourlet-2.3 Texture Retrieval Algorithm

The key technologies of contourlet-2.3 texture image retrieval system include contourlet-2.3 transform, feature vectors construction and distance measure. So we will introduce them separately in this section.

2.1 Contourlet-2.3 Transform

We will review the contourlet-2.3 transform in this section to explain why we choose it to implement our retrieval system.

The original contourlet transform is constructed as a combination of Laplacian Pyramid (LP) and directional filter banks (DFB), where the LP iteratively decomposes a 2-D image into low pass and high pass sub-bands, and the DFB are applied to the high pass sub-bands to further decompose the frequency spectrum. Using ideal filters, the contourlet transform will decompose the 2-D frequency spectrum into trapezoid-shaped regions. Due to this cascade structure, multiscale and multi-directional decomposition stages in the contourlet transform are independent of each other. One can decompose each scale into any arbitrary power of two's number of directions, and different scales can be decomposed into different numbers of directions. This feature makes contourlet a unique transform that can achieve a high level of flexibility in decomposition while being close to critically sampled (up to 33% redundancy, which comes from the LP structure).

Experiments have shown that the original contourlet transform has poor time-frequency localization character which leads to severe artifacts and aliasing exist in the recovered images. Lu and Do proposed a new design based on the basic structure of the original contourlet transform to enhance the localization of the contourlet basis.

In their design, a new LP structure which is composed of $Di(\omega)$ and $Li(\omega)(i=0,1,2\ldots)$ replaced the LP structure in the original CT. The DFB in the new design uses the same one as the original transform. Depending on the different subsampled matrix in LP-like structure, the redundancy of the new contourlet transform is different. The choices for the subsampled matrix (d, d) maybe (1,1), (1.5,1.5) or (2,2), and the redundancy of the new version will be about 2.3, 1.6 and 1.3 respectively.

It is notable that the new LP-like structure considered the practical reasons which lead to the deficient time-frequency localization, the new transform can avoid most aliasing and artifacts problem which widely exists in the first CT de-noising method. We communicate with Mr. Yue Lu through E-mail and appreciate to get their CTSD toolbox. Experimental results show that the new transform has better localization basis than that of the original version. According to different redundancies the retrieval results are different, roughly speaking, higher redundancy results in higher retrieval rate. In this paper, we use contourlet-2.3 to implement our retrieval system.

2.2 Construction of Feature Vectors

Many methods have been used to construct feature vectors including energy combined with standard deviation, generalized Gaussian model and co-occurrence matrix,

here we use the absolute mean energy value and kurtosis of each contourlet domain directional sub-band.

For a sub-band in contourlet domain, we can use (2) to calculate its absolute mean energy, and the kurtosis used here is defined as (3), where s and k denote the index of scale and direction, M,N stand for the row and column number of the sub-band coefficients, W is the coefficient of row M and column N in sub-band indexed by s and k, μ andσrepresent mean and standard deviation, respectively.

$$E(s,k)=\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}|W_{s,k}(m,n)| \ . \tag{2}$$

$$Kur(s,k)=\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}\frac{\left(W_{s,k}(m,n)-\mu(s,k)\right)^4}{\sigma(s,k)^4} \ . \tag{3}$$

Each feature vector is constructed by cascading the energy value and kurtosis of each contourlet domain directional sub-band. For every image in the database which will be retrieved, a certain feature vector can be obtained and then is put into the feature vector database as the signature of the corresponding image for retrieval.

2.3 The Determination of Similarity Measure

The similarity measure is used to calculate the distance between different feature vectors. Up to now, at least there are 10 different types of distance measure, they are: Manhattan (L1), Weighted-Mean–Variance (WMV), Euclidean (L2), Chebychev (L), Mahalanobis, Canberra, Bray-Curtis, Squared Chord, Squared Chi-Squared and Kull-back Leibler. Kokare compared the nine measures except Kull-back distance (KLD) and declared that Canberra and Bray-Curtis are superior to others [11], and we compared Canberra and Kull-back distance, the result is that Canberra is more suitable in such kind of situation. So in this paper, we directly choose Canberra distance as distance measure. The Canberra distance is defined as (4), where $d(\mathbf{x},\mathbf{y})$ means the distance between vector $\mathbf{x},\mathbf{y}$, D denotes the dimension of the feature vectors, x_i, y_i are the i-th components of $\mathbf{x}$ and $\mathbf{y}$, respectively.

$$d(\mathbf{x},\mathbf{y})=\sum_{i=1}^{D}\frac{|\mathbf{x}_i-\mathbf{y}_i|}{|\mathbf{x}_i|+|\mathbf{y}_i|} \ . \tag{4}$$

3 Experiment and Results

In this section, we will introduce the implementation approach of the contourlet-2.3 texture image retrieval system, and evaluate the retrieval rate of the algorithm. Furthermore, we will study the factors which influence the retrieval rate and how to improve the retrieval rate of the system.

3.1 Experimental Objects

The experimental objects are the 109 texture images come from Brodatz album [12]. For each 640×640 pixels image, we cut them into non-overlapped 16 sub-images and each one is 160×160 pixels size, then we can obtain an image database with 109×16=1744 sub-images. The 16 sub-images come from the same original image can be viewed as the same category.

3.2 Experimental Approach

The experimental approach can be divided into 4 steps:

Step 1: For each sub-image in the database, we used contourlet-2.3 to transform it into contourlet-2.3 domain. In contourlet-2.3 domain, for each image, we calculated the absolute mean energy and kurtosis of each directional sub-band using (2) and (3), respectively. Then we cascaded them together as the feature vector of that image. Choosing the decomposition parameter as [4 3 3] means that the numbers of directional sub-bands are 16, 8, 8, adding the low frequency sub-band, the number of sub-bands are 33, each sub-band needs two parameters to describe, so, for every sub-image in the database, the dimension of feature vector is 66. Using the same method for every sub-image, we can extract 1744 feature vectors altogether. All the feature vectors were put together into feature vector database.

The following steps used to evaluate the performance of the retrieval system.

Step 2: Select the first sub-image in the database, using (4), calculate the Canberra distance between its feature vector and every one in the feature vector database. Then find the N=16 nearest images as the retrieval result. Examine how many images belong to the corresponding group, and divided the value by 16 to get the retrieval rate;

Step 3: For next image feature vector in the 1744 sub-image vector database, using the same method as in step 2, calculate the average retrieval rate R, and repeat the procedure until all the feature vectors have been processed.

Step 4: For $N \in \{16, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$, repeat step 2 and 3, calculate the average retrieval rate for each N.

Step 2 to Step 4 can be described by formula (5) as follows, where q=1744, R(p) denotes the average retrieval rate for each $p \in \{16, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$, hence 10 retrieval results can be acquired. S(p, i) is the number of images belong to the correct group when the i-th image used as query image.

$$R(p) = \frac{1}{q}\sum_{i=1}^{q} R(p,i) = \frac{1}{q}\sum_{i=1}^{q} \frac{S(p,i)}{16} \quad . \tag{5}$$

3.3 Experimental Results

Using the above approach, we can get the average retrieval rate of contourlet-2.3 texture image retrieval system as shown in table 1.

It should be noted that CO stands for contourlet-2.3 with the combination feature of absolute mean energy and standard deviation which is used widely in wavelet and contourlet-like situations as in (6) while CN means contourlet-2.3 with the combination feature of absolute mean energy and kurtosis which was proposed in this paper. The row named "av" means the average retrieval rate of the corresponding coloum.

In table 1, we compared the retrieval rates and some other systems including the original contourlet transform (CT), and non-subsampled contourlet transform (NSCT, here use NT for short) under the same structure and decomposition parameters. In the experiment, for CT condition, we used "pkva" and "9-7" bi-orthogonal wavelet as DFB and LP filters, respectively. For NSCT condition, the "maxflat" and "dmaxflat7" were used as LP and DFB filters, respectively. For contourlet-2.3 (CT23) transform "pkva" were used for DFB filters.

$$\sigma(s,k)=[\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}|W_{s,k}(m,n)-\mu_{s,k}|^2]^{1/2} \quad . \tag{6}$$

From table 1 we can see that no matter what decomposition parameters were selected, contourlet-2.3 retrieval system always has higher average retrieval rate that CT system, especially under small N values. Comparing with NSCT, CT23 has some little privilege. It should be noted that, CT23 has significant lower redundancy than NSCT, hence needs much less time in building the feature vector database, which is a welcome character at present time due to too fast expanding image database.

Table 1. Comparison of different texture image retrieval algorithms (%)

	[4 3 3]				[3 2 2]				[3 3 2 2]			
N	**CT**	**NT**	**CO**	**CN**	**CT**	**NT**	**CO**	**CN**	**CT**	**NT**	**CO**	**CN**
16	68.3	71.4	67.3	71.2	70.0	71.0	68.2	70.8	70.6	72.0	69.1	72.5
20	74.2	76.7	72.5	76.2	74.9	75.9	73.5	76.1	76.1	77.4	74.5	78.2
30	80.5	81.3	78.1	81.9	80.1	80.8	78.7	81.3	81.6	81.9	79.5	83.1
40	83.5	84.1	81.2	85.0	83.1	83.6	81.3	84.2	84.1	84.2	82.2	85.6
50	85.7	86.2	83.0	87.1	85.1	85.6	83.1	86.3	86.0	85.8	84.1	87.5
60	87.0	87.6	84.4	88.5	86.8	87.2	84.5	87.8	87.5	87.1	85.8	88.9
70	88.1	88.7	85.7	89.7	88.0	88.3	85.7	88.9	88.7	88.1	87.1	89.9
80	89.1	89.7	87.0	90.6	89.0	89.2	87.0	89.8	89.7	89.0	88.4	90.7
90	90.0	90.4	87.9	91.4	89.8	90.0	88.0	90.5	90.6	89.8	89.3	91.5
100	90.6	91.0	88.8	92.1	90.6	90.7	89.0	91.1	91.3	90.5	90.2	92.2
av	83.7	84.7	81.6	85.4	83.7	84.2	81.9	84.7	84.6	84.6	83.0	86.0

Table 1 show that decomposition parameters including the scale number and directional sub-band number have great influence on the retrieval rates for all the four retrieval systems. With the increasing of scale number, the retrieval rate tends rising, and with the increasing of directional number, the average retrieval rates tends falling.

For the former, the increasing of scale number means to describe the same image from more different distances, but the scale number should be limited so that the approximation image is no smaller than 10×10 pixels for robust reason. For the latter phenomenon, the reason rises from the texture images of this database are not rich enough texture information.

We can also find that the features what we use can make significant effect to retrieval rates. Even using the original contourlet transform with absolute mean energy and kurtosis features can perform better than contourlet-2.3 with absolute mean energy and standard deviation features. Of course, contourlet-2.3 with absolute mean energy and kurtosis features is superior to CT and NSCT situations.

4 Conclusion

A contourlet-2.3 based texture image retrieval system was proposed in this paper which utilized the CT23 combined with the Canberra distance and the features including absolute mean energy and kurtosis of each sub-band coefficients. The new retrieval system has higher retrieval rate than absolute mean energy and standard deviation under same structure and same dimension of feature vectors. The new algorithm has higher retrieval rate than the original contourlet transform and nonsubsampled contourlet transform under same structure.

We will further improve the retrieval rate through the means including new transforms, features and similarity measure.

Acknowledgments. This paper is supported by the High Level Science Research Starting Fund of Xinyang Normal University and Project of The Education Department of Henan Province (2010B120009).

References

1. Smeulders, A., Worring, M., Santini, S.: Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Recognit. Machine Intell. 22(12), 1349–1380 (2000)
2. Minh, N.D., Vetterli, M.: Wavelet-based texture retrieval using Generalized Gaussian density and kullback-leibler distance. IEEE Transactions on Image Processing 11(2), 146–158 (2002)
3. Laine, A., Fan, J.: Texture classification by wavelet packet signatures. IEEE Trans. Pattern Recognit. Machine Intell. 15, 1186–1191 (1993)
4. Chang, T., Kuo, C.: Texture analysis and classification with tree-structure wavelet transform. IEEE Trans. on Image Processing 2, 429–441 (1993)
5. Smith, J.R., Chang, S.F.: Transform features for texture classification and discrimination in large image databases. In: Proceedings of IEEE Int. Conf. on Image Processing, Texas, pp. 407–411 (1994)
6. Do, M.N., Vetterli, M.: Contourlets: a directional multiresolution image representation. In: International Conference on Image Processing, New York, pp. 357–360 (September 2002)
7. Cunha, D., Zhou, J., Do, M.N.: The nonsubsampled contourlet transform: theory, design, and applications. IEEE Transactions on Image Processing 15, 3089–3101 (2006)

8. Lu, Y., Do, M.N.: A new contourlet transform with sharp frequency localization. In: Proceeding of IEEE International Conference on Image Processing, Atlanta, pp. 8–11 (October 2006)
9. Qimin, C., Guangxi, Z.: Contourlet spectral histogram for texture retrieval of remotely sensed imagery. In: Proceeding of SPIE on Remote Sensing and GIS Data Processing and Other Applications, Yichang, pp. 74981R–74981R-6 (2009)
10. Arun, K.S., Hema, P.M.: Content Based Medical Image Retrieval by Combining Rotation Invariant Contourlet Features and Fourier Descriptors. International Journal of Recent Trends in Engineering 2, 35–39 (2009)
11. Kokare, M., Chatterji, B.N., Biswas, P.K.: Comparison of similarity metrics for texture image retrieval. In: IEEE TENCON Conference, Bangalore, pp. 571–575 (2003)
12. Trygve, R.:Brodatz texture images, `http://www.ux.uis.no/~tranden/brodatz.html`

Predict Demonstration Method and Its Application on Robot Soccer Obstacle Avoidance

Peng Huang, Lijin Guo, Ying Li, and Wuxi Shi

Tianjin Polytechnic University, Tianjin, China
huangpeng@tjpu.edu.cn

Abstract. This paper presents a new predict demonstration approach for robot soccer path planning under complex and uncertain environment. By predicting and analyzing the future position and attitude of concerned object, the position and attitude of the object is controlled by demonstration algorithm. The proposed method is successfully used in the obstacle avoidance and is realized on the MiroSot 3vs3 simulating platform. Experiment results show that this algorithm has good real-time ability and adaptability to environment.

Keywords: obstacle avoidance, predict control, path planning, robot soccer.

1 Introduction

Robot soccer game is full of intense competition where accurate collision-free path planning is one of the most important challenges. During the realization of path planning, both the dynamic behaviors and tendency of the robots and obstacles should be considered. The changing state of robots will affect the state of obstacles, which will accordingly affect the state of robots. As a result, it is quite difficult for path planning since both robots and obstacles have dynamic characters. Recently, there are lots of methods for robot path planning such as artificial field algorithm, genetic algorithm, neural network, adaptive control6, etc[1-3]. It seems that none is adaptive to all circumstance as each has its advantages and disadvantages. Furthermore, in a robot soccer game, successful obstacle avoidance and rapidness are both of great importance. This paper presents an approach based on predict demonstration method by predicting the uncertainty of the size, direction of the movement of the obstacle [4] and then using demonstration method to controlled the object. The proposed algorithm is successfully applied in the obstacle avoidance by the MiroSot 3vs3 simulating platform. Results show that this algorithm has good real-time ability and adaptability to environment.

2 Demonstration Algorithm

The game field is divided into several zones as shown in fig.1. Each zone is corresponding to a value which is defined as aggressive coefficient (AC). The larger value of

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 327–330.
springerlink.com

AC means more successful attacking probability at that very point. If we divide all behaviors of the robot at any point into one zone, then each robot will have an available zone. So how to choose an optimal point becomes to the key of this problem. Obviously, it is not efficient to predict, estimate and calculate each point concerned and the speed limitation of each wheel makes the problem worse. For example, the robot usually has the following behaviors: (1) Directions: front and back; (2) Speed: including 4 levels as left or right. So the speed combination of points that the robot can arrive is 2*4*4=32, and the available distance of each point is different. Therefore we choose the optimal speed combination of the left and right wheel from the 32 kinds of speed according to the dissimilar circumstances in the game field every period.

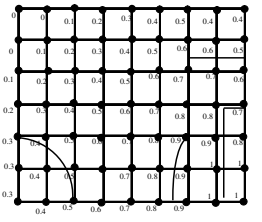


Fig. 1. Demonstration algorithm sketch

According to the evaluation of each point, we judge the weight value of the point and select the point which has the highest weight value as the objective of the movement. Available zone and evaluation factors assure the reachability of the concerned point so we need not to discuss more conditions. Assume t=80ms as a computer command period:

(1) Concerning available zone:

- orientations of each robot of our side, which influence the speed selection;
- coordinates of all robots;

(2) Concerning objective:

- with obstacles or not, including the robots of the opponent and our own;
- on the sideline or not, namely outside or not
- the position of the ball relative to the robot;
- controllability of the ball, namely whether the attacking robot is approaching the ball;
- the defending effect on the opponent member;
- the importance coefficient of local zone;
- attitude after movement, namely whether toward the goal or the next action to take;
- the amount of robot in each zone;

From above we choose the max value. The assignment of weight values can be measured by the importance of factors at each point, and the weight values of those points that the robot can't reach are defined zero. As long as a certain point is chosen, the corresponding speeds of the left and right wheel are ready to use.

3 Predict Demonstration Algorithm

The circumstance of robot soccer game is complex, uncertain and changing rapidly, however the robot members' action is much limited. So how to reasonably plan the restricted acts and obtain the efficiency is important. This following part discusses a predictive approach to decide the robot actions by the prediction of the effects based on demonstration algorithm. This model has the characteristic of high robustness, high efficiency and high antagonism.

The main Agent's decisions are determined by the camera above the field which captures all information about the robots and the ball in the entire course of the game. The concerned information can be regarded as static in some relative time, while the trajectory of the robots and the ball is dynamic and continuous. So we must predict the future positions of the robots and the ball in order to succeed in obtaining the optimal path. As fig.2. shows, where *A1, A2*, and *A3* means respectively positions of the target at the former time*(t-T)*, current time*(t)*, next time *(t+T)*, when $t \to t+T$ (T is the sample period of computer), the motive distance of the target is $\overline{A_2A_3}$. Since the T value is very small, there is $\overline{A_2A_3} = \overline{A_1A_2}$ approximately. So we can get:

$$\overline{OA_3} = \overline{OA_2} + \overline{A_2A_3} = \overline{OA_2} + \overline{A_1A_2} \tag{1}$$

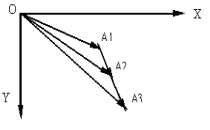


Fig. 2. Inertia prediction sketch

Inertia prediction algorithm is introduced as follows by a robot pursuing the ball example (as shown in fig.3.). Suppose that the current speed of the ball (size and direction given) and the position of robot R are known; *Bt* is the current position of the ball; *Bt-T* is the former position of the ball; *nT* is the predictive time. Where *n* is a constant obtained from experiment, *T* is sample period. After calculating we can get:

(1) The position of the ball after *nT* time, namely the coordinate of point *Bt+nT*;

(2) Assuming *R* is in regular speed, calculate $\left|\overline{RB_{t+nT}}\right|$, the speed V' of *R* can be obtained from *Bt* to *Bt+nT* as $V' = \left|\overline{RB_{t+nT}}\right| \Big/ nT$ (2)

Obviously, the smaller *n* is, the smaller *nT* is, and the predict approach is more accurate. However, there is no inertia to the ball if *n* is too small, and there is an increase to step length which may cause overshoot and larger error if *n* is too big. So it

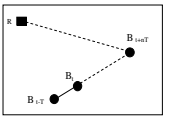


Fig. 3. The predict method diagram

is very important to adjust appropriately the value of *n.* In addition, whatever *n* is, all the values obtained from predict method have an error margin and need continuous improvement. The prediction and demonstration approach are combined as predict demonstration algorithm (PDA).

4 Simulation and Results

The mirosot3vs3 experiment plat is provided by the robot soccer competition BBS for the hardware system simulating and decision training. Users can program their own decision-making system based on this platform. The develop environment is Windows2000 VC++6.0. The PDA has been applied in robot soccer obstacle avoidance. One case is considered about the simulating experiment: one robot of the own team, one robot of the opponent team and the ball in the match field. The simulating result is shown in Fig.4. As the obstacle position (here is the opponent robot which in regular speed) are known at the current time and the former time, the position of the obstacle after *nt* time can be predicted.

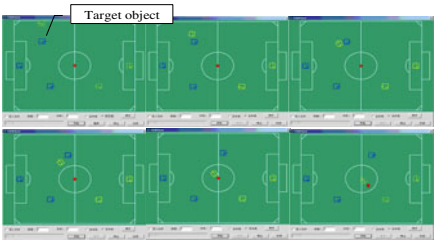


Fig. 4. Simulation result of robot obstacle avoidance with PDA

5 Conclusions

With PDA, the computer predicts and analysis the position of the target in future time and uses demonstration approach in path planning. Experiments and simulating results show that the robot path designed by DDA is more preferable and rapid. Using the proposed approach to plan the robot track is equal to change the static circumstance into dynamic one, so it has good real-time ability and adaptability to environment. Experiments show that this algorithm can be applied in penalty kick and holding actions.

References

1. Fan, C., Chen, W.: A Neural Networks-based Approach to Safe Path Planning of Mobile Robot in Unknown Environment. Acta Automatica Sinica 30, 816–823 (2004)
2. Fan, X., Li, S.: Dynamic Obstacle-avoiding Path Planning for Robots Based on a New Artificial Potential Field Function. Control Theory & Applications 22, 703–707 (2005)
3. Zhong, B., Zhang, Q., Yang, Y.: Anti-collision path planning for soccer robot using modified potential field method. Control Theory & Applications 20, 623–626 (2003)
4. Xi, Y., Zhang, C.: Rolling path planning of robots in a kind of dynamic uncertain environment. Acta Automatica Sinica 25, 161–175 (2002)

Study on the Gain of SNR for Programmable Amplifiers

Zheying Li[1] IEEE Member, Limei Xiu[1], Jia Liu[1], Caixia Lv[1],
Qianru Li[2], and Li Luo[3]

[1] Institute of Microelectronic Application Technology, BUU, Beijing China
[2] College of Astronomy, HIT, Harbin, China
[3] Electronics and Communication School, BJTU, Beijing, China

Abstract. SNR gain for programmable amplifiers is analyzed in this paper. The results show that programmable amplifier with amplifier-array (PACAA) has higher SNR gain compared with programmable amplifier with single Op amp. SNR gain of PACAA is over 20dB and direct proportional to the number of amplifiers connected in serial. The input signal of the amplifier is divided into two parts according to the frequency, one of which is signal if its frequency is below a certain value and another is noise if its frequency is over the certain value. To the same design requirements, PACAA not only has good SNR gain but also reduces demands for the 0dB bandwidths of Op amps. However, the integrated circuits with PACAA will increase the area of the dice. It is needed to consider how to obtain higher SNR gain and minimize the chip size.

Keywords: programmable amplifier, SNR, gain.

1 Introduction

Variable gain amplifier can be divided into two types in engineering technology. One is automatic gain control amplifier (AGC) [1] [2], and the other is programmable gain amplifier (PGA) [3] [4] [5]. In signal detection and measurement systems, PGA is generally used as the front-end in measurement or data acquisition circuits for the gain control of analog input signals to meet the requirements of data acquisition and measurement [4].

There are two ways to achieve PGA. One is programmable amplifier constructed with an Op amp [4], and the other is programmable amplifier constructed with amplifier-array (PACAA). The gains of two kinds PGA are both programmable, but their technical features and design methods are relatively different. The different are embodied in the parameters of 0dB bandwidth requirement for Op amp and the SNR gain corresponding to the same amplification factor. SNR gains of programmable amplifier with single Op amp and PACAA are discussed in this paper. Results of the study show that PACAA has higher SNR gain, which is benefit for reducing the output noise of the amplifier.

In this paper, frequency characteristics of the Programmable amplifier with single Op amp and PACAA are discussed in part 2; in part 3 and part 4, SNR gains of both kinds of programmable amplifiers are analyzed and discussed, respectively.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 331–338.
springerlink.com

2 Programmable Amplifiers and Amplifier Array

2.1 Frequency Characteristics of Non-inverting Amplifier Circuit

The frequency response of the Op amp is directly related to the connection states of the circuit. Fig. 1 shows the closed loop feedback circuit of the Op amp. It is known that the Op Amp circuit can run steadily under the negative feedback state condition. Fig. 1 is a non-inverting amplifier with negative feedback structure.

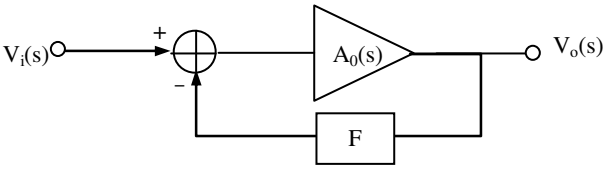


Fig. 1. Closed loop negative feedback structure of Op Amp

In Fig. 1 $A_0(s) = a_0/(1+s/p_1)$ is the open loop gain of the Op Amp, a_0 is the low frequency amplification factor of open loop of the Op Amp, p_1 is the main pole of the Op Amp, and F is the negative feedback function. The transfer function of Fig.1 is

$$H(s) = V_o(s)/V_{in}(s) = A_0(s)/(1+A_0(s)F) \tag{1}$$

Under the condition of the deeply negative feedback, $A_L = a_0/(1+a_0F) \approx 1/F$, and let $p_2 = p_1(1+a_0/A_L)$, therefore

$$H_n(s) = A_L/(1+s/p_2) \tag{2}$$

To non-inverting amplifier there is $F = 1/(1+R_f/R)$, therefore,

$$H_n(s) = A_L \Big/ \left(1+s \Big/ p_1\left(1+a_0/\left(1+\frac{R_f}{R}\right)\right)\right) = A_L/(1+s/p_2) \tag{3}$$

where $p_2 = p_1\left(1+a_0/(1+R_f/R)\right) \approx p_1a_0/(1+R_f/R)$, so

$$p_2 = p_1a_0/(1+R_f/R) = p_1a_0/A_L \tag{4}$$

2.2 Programmable Amplifier with Single Op Amp

A basic architecture of programmable non-inverting amplifier is shown as Fig.2. In Fig.2, A_L is formed by connecting different feedback resistors.

A non-inverting amplifier can be described with equation (2) and the amplitude-frequency characteristics are shown in Fig. 3, where 20dB thick dashed line

represents $A_L = 10$ and thick solid line represents $A_L = 100$, $A_0 = 20\lg a_0$, f_m and f_{p1} are input signal bandwidths and the frequency corresponding to the main pole of the amplifier, respectively. For the Op amp, Af is a constant. Therefore the unit-gain bandwidth is $f_0 = a_0 f_{p1}$ (i.e. 0dB bandwidth). For example, let maximum frequency of the signal is f_m =20MHz, and the amplifications of non-inverting amplifier with single Op amp are 10 and 100, then the Op amp's unit-gain bandwidth is $20\times10^6\times100 = 2\text{GHz}$, i.e. the 0dB bandwidth of the Op amp has to be $f_0 \geq 2\text{GHz}$.

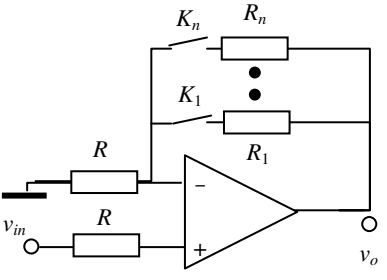


Fig. 2. Non-inverting PGA with single Op amp

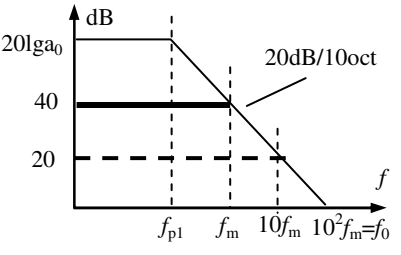


Fig. 3. Bode plot of Fig.2

From Fig. 3, suppose the low frequency amplifications of PGA are A_{La} and A_{Lb}, and $A_{La} < A_{Lb}$, the 3dB bandwidths are from f_a to f_b, then

$$f_a = \left(A_{Lb}/A_{La}\right) f_b \tag{5}$$

From equations (2) and (5), it can be known that there must be $f_m = f_b < f_a$ for the amplification A_{La} because the same Op amp is used. Equation (5) shows that if the programmable amplifier is designed with single Op amp, it must be taken full account of the requirements for the unit-gain bandwidth of the Op amp during it work with maximum voltage gain.

2.3 Gain Programmable Amplifier Array

Fig. 4 is an example of a programmable amplifier constructed with amplifier-array. It can be seen from Fig. 4 that the voltage gain of the amplifier depends on the connection of amplifiers. When the switches K1, K3 and K5 are closed, the voltage amplification factor is 20, while K1, K2 and K4 are closed, the amplification is 100.

Among PACAA, the amplifications of the amplifiers are fixed, so their 3dB bandwidths are fixed. Since each level of magnification do not have to be large, therefore, the requirements for each Op amp's unit-gain bandwidth will be significantly reduced. Suppose the circuit with amplification factor 100 is realized by two amplifiers

with the magnification of 10 in series, and the maximum frequency of the input signal is f_m=20MHz, then $20\times10^6\times10$=200MHz. This shows that the unit-gain bandwidth of the Op amp should be $f_0\geq$200MHz. This means that the requirement for the Op amp's unit-gain bandwidth is reduced 10 times compared with Fig. 2,.

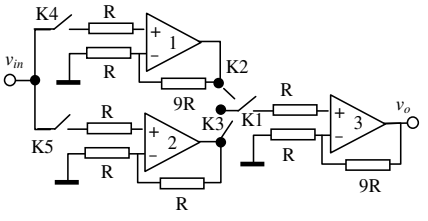
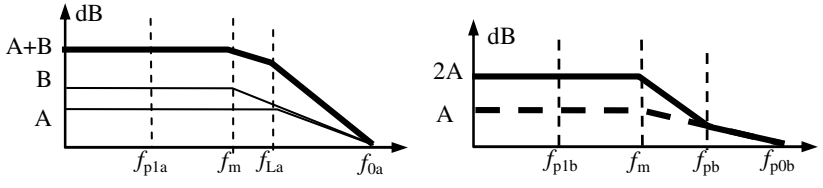


Fig. 4. Amplifier-array

Let A_{La} and A_{Lb} are low frequency gains of the two series amplifiers, $A_{La}A_{Lb}=A_L$ and $A_{La} < A_{Lb}$. Meanwhile, assuming two amplifiers have the same unit-gain bandwidths (0dB bandwidth) f_0, i.e. $f_{p1a}=f_{p1b}$. 3dB bandwidth of the amplifier with largest gain is $f_{Lb}=f_m$ and $f_m<f_{La}<f_0$. The transfer function of the amplifier in Fig. 4 is

$$H_i(s)=\frac{A_{La}A_{Lb}}{(1+s/p_{2a})(1+s/p_{2b})} \tag{6}$$



(a) Different gaia (b) The same gain

Fig. 5. Bode plot of the PACAA

The Bode plot of equation (6) is shown in Fig. 5. The thick dotted line in Fig. 5(b) represents two amplifiers with the same magnification. It can be seen from Fig. 5, due to the unit-gain bandwidth of the amplifier is effectively reduced, the programmable amplifier has a better inhibitory effect to the out-band noise mixed in the signal

3 SNR Gain Analysis of Programmable Amplifier Constructed with Single Op Amp

Assuming the input signal is $s(t)=s_1(t)+n(t)$, where the effective bandwidth of $s_1(t)$ is f_m, $n(t)$ is the out-band noise mixed in the signal whose minimum frequency is greater than f_m. The SNR gain of an amplifier is defined as

$$D = 20\lg\left|SNR_o / SNR_{in}\right| \tag{7}$$

where SNR of the input signal is $SNR_{in} = S_1(s)/N_{in}(s)$ and SNR of the output signal is $SNR_o = S_{1o}(s)/N_o(s)$. For the PGA shown as Fig. 3, the Laplace Transform of the output signal $Y_n(s)$ is

$$Y_n(s) = \frac{A_L}{1+s/p_2}[S_1(s)+N(s)] = A_L S_1(s) + \frac{A_L}{1+s/p_2}N(s) \tag{8}$$

In Fig. 3, 3dB bandwidth is f_L corresponding to A_L, and $f_L \geq f_m$ can be met at any magnification. Let input noise is $N(s)=N_{m0}(s)+N_{op}(s)$, where, $N_{m0}(s)$ is the noise between $f_m < f_L$ and $N_{op}(s)$ is the noise between $f_L < f_o$. Therefore the output noise can be divided into two parts:

1) For the noise between $f_m \leq f_L$, because it is in the pass band of the amplifier (Fig. 3), the output noise of this part is

$$N_{o1}(s) = A_L N_{m0}(s) \tag{9}$$

2) For the noise between $f_L < f_0$, because it is outside the pass band of the amplifier, the output noise of this part is

$$N_{o2}(s) = \frac{A_L}{1+s/p_2}N_{op}(s) \tag{10}$$

Then, the output noise of the amplifier is $N_o(s) = N_{o1}(s) + N_{o2}(s)$. According to equation (7), the SNR of the output signal is

$$SNR_o = \left(\frac{1+\dfrac{s}{p_2}}{1+\dfrac{sN_{m0}(s)}{p_2 N_{in}(s)}}\right)\frac{S_1(s)}{N_{in}(s)} \tag{11}$$

Above all, SNR gain of the programmable amplifier with single Op amp is

$$D = 20\lg\left|1+\frac{s}{p_2}\right| - 20\lg\left|1+\frac{sN_{m0}(s)}{p_2 N_{in}(s)}\right| \tag{12}$$

Considering $p_2 = p_1 a_0 / A_L$, set $s = j\omega = j2\pi f$, $p_1 = 2\pi f_{p1}$ and substitutes them into equation (13), equation (14) is received.

$$D = 20\lg\left|1+\frac{jfA_L}{f_{p1}a_0}\right| - 20\lg\left|1+\frac{jfA_L N_{m0}(f)}{f_{p1}a_0 N_{in}(f)}\right| \tag{13}$$

The graph is shown in Fig. 6.

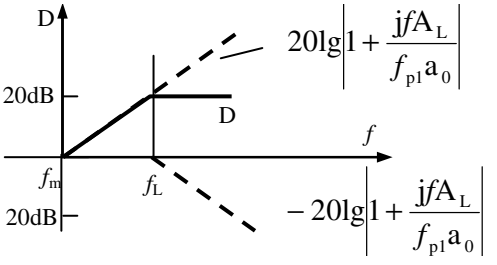


Fig. 6. SNR gain of programmable amplifier with single Op amp

Equation (13) shows that for programmable amplifier with single Op amp:

1) If the noise within the pass band of the amplifier $N_{m0}(s)$ is equal to 0, i.e. $f_m = f_L$, the SNR gain is:

$$D = 20\lg\left|1+\frac{jfA_L}{f_{p1}a_0}\right| \tag{14}$$

Equation (14) shows that SNR of the output is larger than that of the input.

2) If all the noise is in the pass band of the amplifier,

$$D = 20\lg\left|1+\frac{jfA_L}{f_{p1}a_0}\right| - 20\lg\left|1+\frac{jfA_L}{f_{p1}a_0}\right| = 0 \tag{15}$$

It shows that the SNR is not improved.

3) If neither of $N_{m0}(s)$ and $N_{op}(s)$ is equal to 0, the SNR gain is smaller than 20. There is little improvement in low frequency noise but more obvious improvement in high frequency noise. This shows that higher SNR gain can be obtained in the case of higher gain in low-frequency while a smaller SNR gain can be obtained in the case of smaller low-frequency gain. Both cases significantly inhibit the high-frequency noise but the low-frequency noise suppression is not obvious.

4 SNR Gain of the PACAA

Consider the PACAA in Fig. (4), it is composed by two amplifiers in series (where the amplifier 3 is always included in any combination). Using equation (5), equation (16) is received.

$$Y_n(s) = \frac{A_{La}A_{Lb}}{(1+s/p_{2a})(1+s/p_{2b})}[S_1(s)+N(s)] \tag{16}$$

In order to compare with programmable amplifier with single Op amp, let $A_{La}A_{Lb} = A_L$ and $A_{La} < A_{Lb}$, meanwhile, both of the Op amps have the same unit-gain bandwidth (0dB bandwidth), $f_m < f_{La} < f_0$ and $f_{pa} = f_{pb}$.

Assuming $f_{Lb} = f_m$, the effective signal in the output is

$$S_n(s) = A_L S_1(s) \tag{17}$$

The noise in the output is

$$N_o = \frac{A_{La}A_{Lb}}{(1+s/p_{2a})(1+s/p_{2b})} N_{in}(s) \tag{18}$$

Similarly, assuming

$$N_o = \frac{A_{La}A_{Lb}}{(1+s/p_{2a})(1+s/p_{2b})}[N_{o1}(s)+N_{o2}(s)] \tag{19}$$

Considering $f_m = f_{Lb} < f_{La} < f_0$, it is received like this:

1) For the low gain amplifier which is in the input port, considering there is noise in the pass band, from equation (9) and (10), the output noise of first amplifier is

$$N_{oa}(s) = N_{oa1}(s) + N_{oa2}(s) = A_{La}\left(\frac{1+sN_{m0}(s)/(p_{2a}N_{in}(s))}{1+s/p_{2a}}\right)N_{in}(s) \tag{20}$$

2) For the high gain Op amp, there is no noise in the pass band,

$$N_{ob} = \frac{A_{Lb}}{1+s/p_{2b}} N_{oa}(s) \tag{21}$$

Substitute equation (20) into (21), the noise in the output is

$$N_o(s) = N_{ob}(s) = \frac{A_{La}A_{Lb}}{1+s/p_{2b}}\left(\frac{1+sN_{m0}(s)/(p_{2a}/N_{in}(s))}{1+s/p_{2a}}\right)N_{in}(s) \tag{22}$$

With complex frequency form, SNR gain is

$$D = 20\lg\left|1+\frac{jfA_{La}}{f_{1ab}a_{0ab}}\right| + 20\lg\left|1+\frac{jfA_{Lb}}{f_{1ab}a_{0ab}}\right| - 20\lg\left|1+\frac{jfA_{La}N_{m0}(f)}{f_{1ab}a_{0ab}N_{in}(f)}\right| \tag{23}$$

Compared with equation (14), there are two positive 20dB in equation (23). The Bode plot is shown in Fig. 7. It can be seen that SNR gain of PACAA is 20dB larger than that of programmable amplifier with single Op amp, and higher SNR is received.

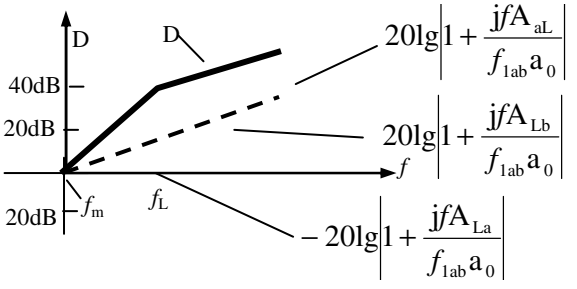


Fig. 7. SNR gain of PACAA

5 Conclusion

SNR gain is one of the basic requirements for signal conditioning circuits in signal measurement systems. This paper points out that PACAA has higher SNR gain compared with programmable amplifier constructed with an Op amp. In normal conditions, when other design requirements are the same, PACAA reduces requirements for unit-gain bandwidths of Op amps besides the better SNR gain. Therefore, it makes the amplifier design easier. However, compared with programmable amplifier constructed with an Op amp, PACAA will increase the area of analog circuits in integrated circuits design. Therefore, it is required careful analysis to obtain the lowest possible chip area under conditions of high SNR in relevant integrated circuits design.

Acknowledgment. This work is sponsored by PHR (IHLB) 20090513 of Beijing Education committee and NSFC609976024.

References

1. Alam, S.K., DeGroat, J., Roblin, P.: A 5 GHz CMOS Variable Gain Low Noise Amplifier for wireless LAN Applications. In: Proce. of IEEE 49th MWSCAS 2006, pp. 289–293 (2006)
2. Leung, L.H.Y., Buchwald, A.: 10-MHz 60-dB dynamic-range 6-dB variable gain amplifier. In: Proce. of IEEE ISCAS 1997, vol. 1, pp. 173–176 (1997)
3. Huang, C.-C., Hung, S.-H., Chung, J.-F., Van, L.-D., Lin, C.-T.: Front-end amplifier of low-noise and tunable BW/gain for portable biomedical signal acquisition. In: Proce. of IEEE ISCAS 2008, pp. 2717–2720 (2008)
4. Compagne, E., Martel, G., Senn, P.: A -100 dB THD, 120 dB SNR programmable gain amplifier in a 3.3 V, 0.5 μm CMOS process. In: Proce. of IEEE 21st ESSCIRC 1995, pp. 38–41 (1995)
5. De Lima, J.A.: A low-voltage programmable-gain current-mode amplifier. In: Proce. of IEEE 9th International Conference on Electronics, Circuits and Systems, vol. 1, pp. 33–36 (2002)

Testbench Design for a Mixed-Signal SoC

Chengcheng Wang[1], Caixia Lv[2], and Zheying Li[2]

[1] School of Electronic Information Engineering, Beijing Jiaotong University, Beijing China
wangcheng850318@hotmail.com
[2] Institude of Microelectronic Application Technology, BUU, Beijing China
zheying@buu.edu.cn

Abstract. In this paper, the system's verification testbench is established according to the functional requirements of a mixed-signal SoC system. The testbench tests the system processor's control on the analog signal input circuit's magnification and ADC's sampling frequency. The simulation results with Modelsim indicate that the SoC System has achieved the expected functional requirements. And the testbench validates the accuracy of the SoC design well, reducing the risk of tape-out.

Keywords: mixed-signal SoC, system function, testbench.

1 Introduction

Owing to the high performance, high integration and high complexity, the cost of the chip's design and taping-out is also high. Therefore, the system's function must be verified to ensure the correct function of SoC chip system before taping-out or placing and routing. In the whole process of the SoC chip design, the proportion of simulation and verification is increasing, and the logical inaccuracy is the main reason to cause the failure of SoC chip design and taping-out. Therefore, using the advanced design and simulation methods is the key to the success of SoC chip design. This can not only reduce the risk of SoC design taping-out and the cost, but also greatly reduce the SoC chip's development cycle [1].

Currently, SoC verification methods can be divided into four main categories: verification method based on simulation, static verification method, formalization verification and physical verification. Static verification method is to verify the timing information of the specific circuit; physical verification method is applied to the physical layer of chip design. Therefore, SoC system verification methods are mainly verification method based on simulation and formalization verification [2]. And which is commonly used is software and hardware co-verification method based testbench.

In this paper, a software simulation testbench for a mixed-signal SoC system is established, which is used to carry out system's simulation verification to ensure the realization of its functional requirements. In part II, the structure and functional requirements of the mixed-signal SoC system are described. Part III provides the

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 339–346.
springerlink.com

simulation testbench of the mixed-signal SoC system, and explains the specific method for testing its function. And the results of simulation are given in part IV.

2 A Mixed-Signal SoC System

2.1 System Structure

The system which is needed to establish testbench is a complex circuit mixed with digital-analog circuits, and is composed of some modules such as a set of analog signal input circuits, ADC, processor system, and so on. The structure is shown in Fig.1.

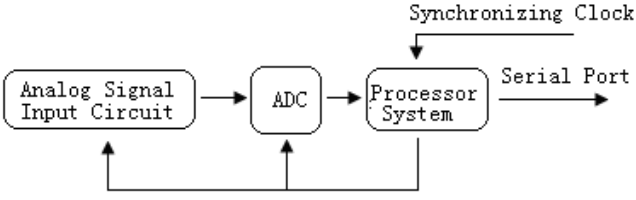


Fig. 1. Structure of a mixed-signal SoC system

As Fig.1 shown above, the system's basic function is to complete the acquisition of the analog signal and control the amplifier and ADC by processor system. Specifically, the analog signal into the system first passes an analog signal input circuit with a magnification, then goes into ADC converting analog signal into digital signal. The digital signal output from ADC is stored into the processor system. At the same time, the processor system can control the selection of the analog input circuit magnification, and the sampling clock of ADC.

2.2 Functional Requirements

The functional requirements of main modules are as follows:

Analog Signal Input Circuit: amplify the input analog signal. The circuit includes a set of amplifiers with four different magnifications which are controlled by four switches to choose the different amplified paths. Users can choose to set an amplifier with a magnification in four according to the actual need.
ADC: achieve to convert analog signal into digital signal, and the conversion rate is controlled by the processor system.
Processor System: realize the control of the analog signal input circuit's magnification and the control of ADC's sampling frequency, receive or transmit data through the serial port.
Serial Port: achieve data communication with the outside.
Synchronous Clock: provide the control signal of the mixed-signal SoC system.

3 Testbench Design

This mixed-signal SoC system is a complex circuit with digital-analog circuits, and each module is a function IP core in the simulation. After completing the simulation of each IP core, it has to establish the system's simulation testbench. The testbench is mainly to test the processor system's control on the analog signal input circuit's magnification and ADC's sampling frequency, detecting whether the system has realized the functional requirements.

3.1 Structure for Test

Each module in the system has to generate the corresponding IP core, and the digital signal generator is used to replace the analog signal input circuit and ADC. The whole structure used for test is shown in Fig.2, and the description of each module's function is shown in Table 1.

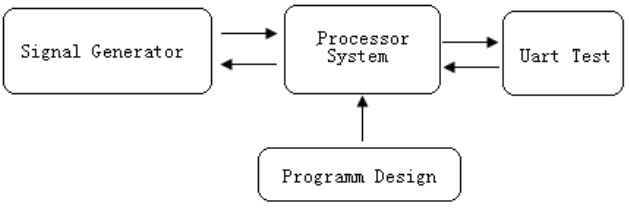


Fig. 2. Structure for test

Table 1. Description of each module's function in Fig.2

Name	Function
Signal Generator	Provide digital signal to Processor System
Processor System	Store digital signal, and control Signal Generator which is used to replace amplifier and ADC
Uart test	Transmit and receive data With Processor System to achieve the serial communication
Program Design	Achieve the functional verification of the system through programs, mainly including the acquisition of the input signal and the control of Signal Generator which replaces amplifier and ADC.

3.2 Testbench

The established testbench is mainly to test the processor system's control on the analog signal input circuit's magnification and ADC's sampling frequency. The structure and the associated control signals are shown below in Fig.3 and Table 2.

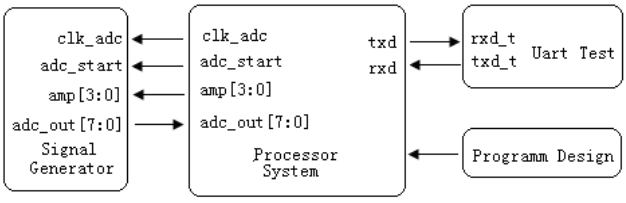


Fig. 3. Structure of system testbench

Table 2. Description of each module's function in Fig.3

Name	Description
clk_adc	The sampling clock of ADC
adc_srart	The start signal of ADC
Amp[3:0]	Choose the path with different magnifications
adc_out[7:0]	The output of ADC
Txd	Transmitting data of system serial port
Rxd	Receiving data of system serial port
txd_t	Transmitting data of testing module serial port
rxd_t	Receiving data of testing module serial port

The testbench is used to test the analog signal's magnification and the ADC's sampling frequency controlled by the processor system. And this control is achieved by the controlling registers.

In the test, the processor system can control the signal's magnification. The corresponding controlling register goes through the 2-4 decoders, and then exports the control signals. The signals control the relay switch to select the amplifier. The four configurations of decoder are corresponding to the four different signal amplifiers. In Table 3.3, when the reg[1:0] = 00, the output amp[3:0] = 0001, which means to select the first signal amplification path. The truth table is shown in Table 3.

Table 3. The truth table

reg[1]	Reg[0]	amp[3]	amp[2]	amp[1]	amp[0]
0	0	0	0	0	1
0	1	0	0	1	0
1	0	0	1	0	0
1	1	1	0	0	0

The control register in processor system can be configured directly by the assembler, and the program is as follow. In program, control register named REG_AMP is configured with 0X01(00000001). So reg[1:0] = 01, and when it is enabled, the corresponding output is amp[3:0] = 0010, namely 2, which means to choose the second signal amplification path.

```
REG_AMP=0X01;
REG_AMP_EN=0X01;
```

The controlling register for ADC controls to generate the clock and the start signal. When 1 bit in the control register namely reg[0] is 0, ADC's sampling clock is the same as the system clock; otherwise the clock offered to the ADC is only half of the system clock. The control register can be configured directly by the assembler too, and the program is as follow. In program, control register named REG_ADC is configured with 0X01(00000001). So reg[0]=01, and when it is enabled, the corresponding output clk_adc is only half of the system clock.

```
REG_ADC=0X01
REG_ADC_EN=0X01;
```

Of course, the control registers can also be configured through the serial port. About serial communication, it requires the same baud rate between system and the uart test module. TMOD = 0x20 means both using timer T1 which works with mode 2; TH1 = 0xe6, TL1 = 0xe6, setting the same initial value of timer T1; TR1 = 1 means to start timing. When transmitting, SCON = 0x40 sets the serial port with mode 1 to send data. And when receiving, SCON = 0x50 sets the serial port with mode 1 to allow reception. So that it can achieve the serial communication.

4 Test Results

The simulation based on testbench, is to simulate the function of the mixed-signal SoC system using assembler through Modelsim. The acquired data of the system is stored in processor system. The result is shown in Fig.4.

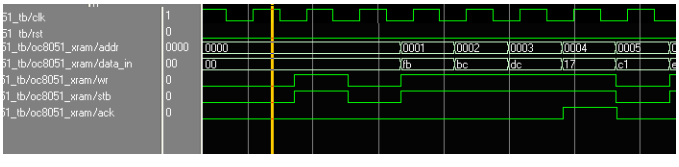


Fig. 4. Test result (1)

According to the design requirements of the address space, the controlling registers are written directly by instructions. The configured result is exported by processor's P0 port, and the result is shown in Fig.5. The register is configured with fcg = 11, acg = 21, and when fcg_en and acg_en are enabled with 1, fcg_o = 11, acg_o = 21, and export through P0 port.

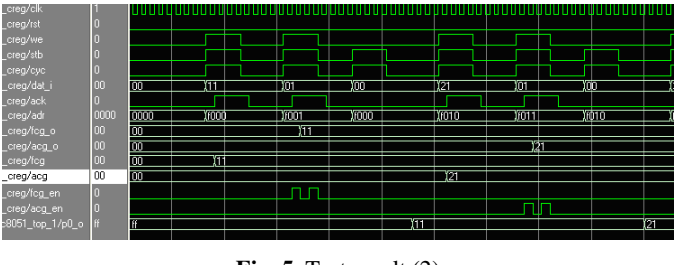


Fig. 5. Test result (2)

It can also be configured through the serial port, as shown in Fig.6. SCON = 0x50 means it allows to receive data. When sbuf=01, fcg_o=01. It makes amp [3:0] = 0010, namely 2, which means the system chooses the second amplifier to amplify the signal.

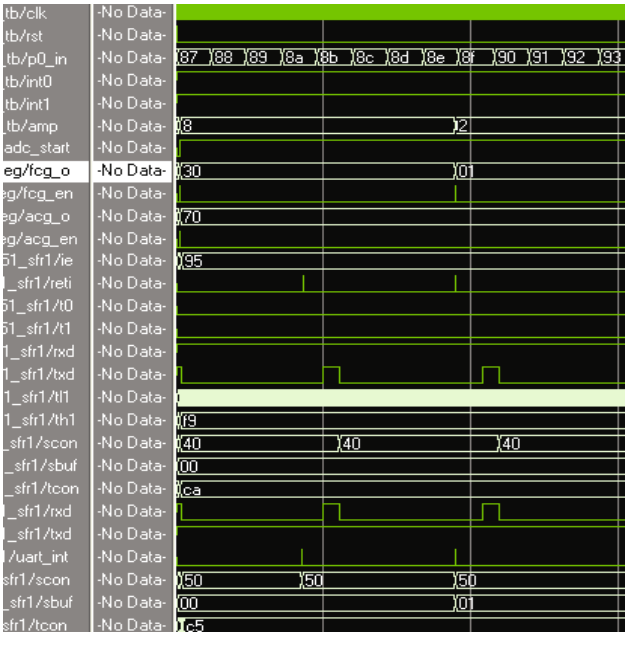


Fig. 6. Test result (3)

Through directly configuring the register to control the ADC's sampling frequency, it is the same as the system clock or only half of the system clock. The result is shown in Fig.7. When acg_o = 00, clk_adc is the same as the system clock 'clk'; when acg_o = 01, clk_adc is only half the system clock 'clk'.

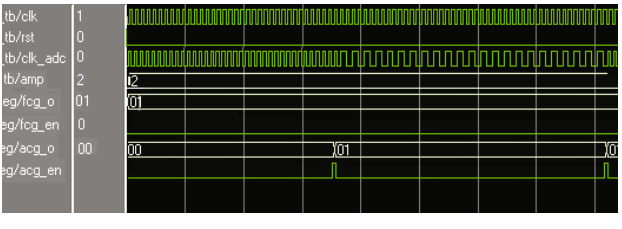


Fig. 7. Test result (4)

The test result of serial communication is shown in Fig.8, and when the baud rate is set the same, SBUF register of the receiving module can receive the data of transmitting module. SCON = 0x40 means it allows to transmit data, and SCON = 0x50 means it allows to receive data. The data is 01,02……, and exported by P0 port.

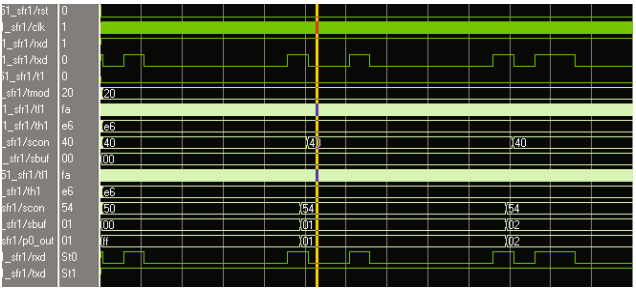


Fig. 8. Serial communication

Through the above test results, the testbench design of the mixed-signal SoC system has completed the acquisition of analog signals, and has realized the processor system's control of amplifier and ADC.

5 Conclusion

This paper provides a simulation testbench for a mixed-signal SoC system, which is to verify the system's function by the software and hardware co-verification method. The results validates that the mixed-signal SoC system has realized its functional requirements, such as the processor system's control of amplifier's magnification and ADC's sampling frequency. The establishment of the mixed signal SoC testbench, not only reduces the risk of SoC design and the cost, but also greatly reduces the development cycle of SoC chip. With the development of verification methodology among the design of SoC and other chips in the future, more scientific and advanced verification techniques will appear [9].

Acknowledgments. This work is sponsored by PHR (IHLB) 20090513 of Beijing Education committee.

References

1. Yu, Z.-l., Jiang, K., Li, F.: The Design of SoC Verification platform based FPGA. Information Technology and Informatization 5 (2008)
2. Ren, X.-j., Feng, C.-c., Zhao, Z.-y., Li, Y., Wan, J.-h.: The Study of SoC System Verification Method. China Computer Federation
3. Feng, Y.-t., Fu, Y.-z., Zhao, F.: The Research and Implementation of Efficient FPGA Verification Technology in Large-scale SoC Design. Application of Electronic Technique 2, (2006)
4. Zhao, G., Hou, L.-g., Liu, Y., Zhu, X.-d., Wu, W.-c.: The Hardware and Software Co-Verification study based on SoC design. Microelectronics & Computer 23(6) (2006)
5. He, W., Zhang, D.-l., Zhou, M.: The Research of SoC Verification Environment Building Methods. China Integrated Circuit 15(8), 22–26 (2006)

6. Duan, Q.-y., Huang, S.-t., Jin, R.-l., Zhang, S.-s., Miao, S.: Study and Implementation on Method of SoC Automatic Verification. Journal of Jilin University 28(3) (2010)
7. Bruce, A., Goodenough, J.: Re-useable Hardware/Software Co-verification of IP-Blocks. In: Proceedings of 14th Annual IEEE International ASIC/SoC Conference, pp. 413-417 (2001)
8. Lu, S., Yu, L.-l.: SoC-Oriented Verification Strategy. Microelectronics 32(4), 265–268 (2002)
9. Shi, G.-z., Shi, Q.-l., Yu, Z.-g., Yu, Z.-g.: The Study on Software and Hardware Co-verification Platform of SoC Design based on FPGA

Implementation of Rapid Prototype Verification for Block-Based SoC*

Xianju Guo[1], Caixia Lv[2], Zheying Li[3], and Haitao Xu[4]

[1] School of Electronics Engineering, Beijing Jiaotong University,
100044, Beijing, China
gxjsky386@gmail.com
[2] Information School, Beijing Union University,
100101, Beijing, China
caixialv04@hotmail.com
[3] Institute of Micro-electronic Application Tech., Beijing Union University,
100101, Beijing, China
zheying@buu.edu.cn
[4] School of Electronics Engineering, Beijing Jiaotong University,
100044, Beijing, China
08120043@bjtu.edu.cn

Abstract. Based on Altera EP2C70, a block-based method is proposed to optimize the Mixed-SoC design in this paper. A rapid verification platform for SoC prototyping system is studied and implemented. The Hardware/Software parallel design and co-verification are realized to greatly reduce the chip time to market. The verification platform establishes an integrated data stream from the PC to the targeted system, and solves the problems of the accurate, predictable and real-time sending, transmission, acquisition of the SoC verification system.

Keywords: Mixed-SoC, FPGA, Block-based, Co-Verification, Data Acqusition System.

1 Introduction

With the development of large scale integrated circuits, integrated circuits has entered SoC (System-on-Chip) era. The difficulty of verification is increasing rapidly due to the constant expansion of SoC. Currently, the success rate of tape-out for the first time is only about 35%, which is largely due to the verification. SoC verification consumes 60 to 80% [4] of the whole design time, which has great negative impact on the return on investment and time to market.

With the ever-increasing complexity of SoC and the increasingly urgent time to market, more attentions and concern are paid to the block-based design SoC prototype verification in the SoC design and verification. Rapid system prototyping, co-verification of hardware prototype and software prototype, has become a common verification method in the early stage of SoC design process. This powerful

* This work is sponsored by PHR(IHLB) 20090513 of Beijing Education committee.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 347–354.
springerlink.com

FPGA-based hardware verification platform can quickly achieve Hardware modules in SoC design, which greatly reduce the chip time to market.

In this paper, the data acquisition system is designed hierarchically using module-based design method of Altera QuartusII. Each module is optimized efficiently only once. This method significantly reduces design iteration time, reserve performance characteristics unprecedented and greatly improves work efficiency. On this basis, study the design of the hardware verification platform based on ALTERA EP2C70 FPGA and the implementation of 8051 SoC rapid prototyping system, and then complete the Parallel design and co- verification of hardware and software.

2 Mixed-SoC Design Optimization

The main features of Mixed-SoC are IP-reuse and HW/SW co-design. General design flow is shown in Fig. 2.1[5]. This paper presents a block-based HW/SW co-design method to optimize the Mixed-SoC design flow. This method significantly shortens the total development time and realizes highly efficient team design. Designers can complete the high-density FPGA design iterations 4 to 5 times per day with this method while they can finish the high-density FPGA design iterations 1 to 2 times per day with the traditional design method [3].

First, divide the entire design into software components and hardware components according to the design specifications. Then, divide the function parts of the software and hardware into blocks. Every block is designed, implemented and imported into the top engineering individually. Some of the functional blocks commonly used in the system are modular designed, debugged, packed and prepare for call. The general principles of module division are as fellows. The contact inside the module should be close. The function of each block is independent. The connection between the blocks should be as simple as possible.

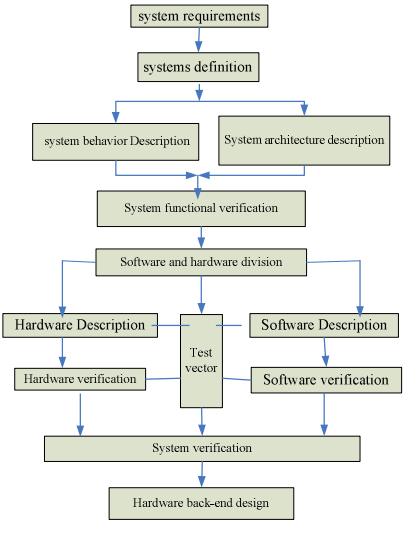


Fig. 1. SoC HW/SW co-design process

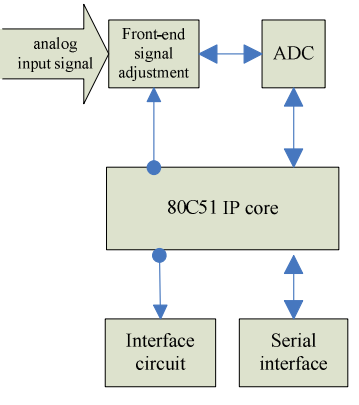


Fig. 2. Modular data acquisition system

In this paper, the data acquisition system is a Mixed-SoC system. Its modular structure is shown in Fig. 2. The front-end signal adjustment circuit and ADC conversion are realized with analog circuit. The interface circuit and serial interface are realized with the digital circuit. The 80C51 IP core as the core control unit control the analog input, signal adjustment, A/D conversion, digital data transfer and storage, serial port communication.The design and implementation of 80C51 soft core are completed in the software design phase. The modular structure of 80C51 is shown in Fig. 3[1].

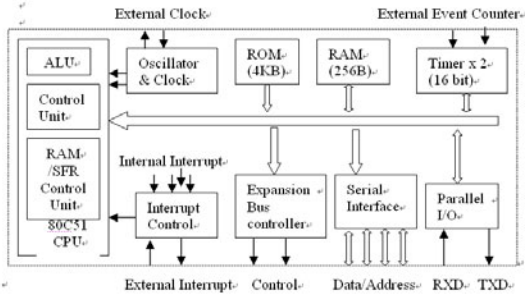


Fig. 3. Modular 80C51 IP soft core

3 SoC Hardware Verification Platform Design

The synthesis result of 80C51 SoC shows that 3121 logic elements are used. We choose EP2C70 considering the expansibility and cost-effective of the SoC verification platform, which is shown in Fig. 4.

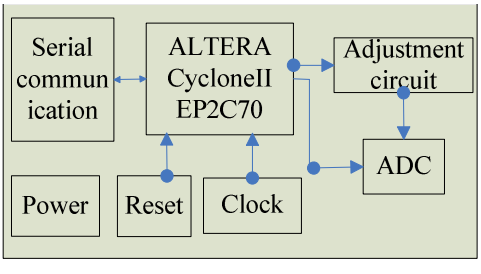


Fig. 4. SoC hardware verification platform

3.1 Hardware Module

The hardware circuit of the SoC verification platform includes the power circuit, reset circuit, clock circuit, a serial interface circuit, the front-end signal adjustment circuit and the ADC conversion.

3.1.1 SoC Verification Circuit Requires Multiple Powers

The main power is 5v, 3.3v and ±1.8V, and the 3.3v and ±1.8V are conversed from 5v by the integrated voltage regulator LM1117. The current load capacity of the 5V power is no less than 5A, and the current load capacity of the ±1.8V power is no less than 800mA.

3.1.2 System Clock

The 50MHZ system clock, generated by the active crystal, is linked to CLK1 of EP2C70.24MHZ block, generated by the internal PLL, is supplied to the 80C51 soft-core.

3.1.3 Reset Circuit

System reset signal becomes effective automatically after remains high more than two machine cycles. This paper uses RC reset circuit with a gate to improve the stability and performance of the reset circuit. The reset circuit is showed in Fig. 5[7].

3.1.4 Serial Communication

Level conversion between RS232 and TTL is required to realize the serial communication between the system chip and PC. The conversion circuit is shown in Fig. 6.

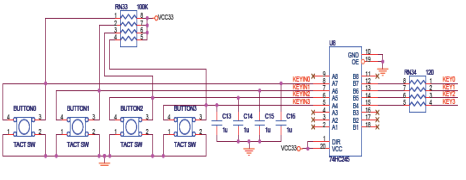


Fig. 5. Reset circuit

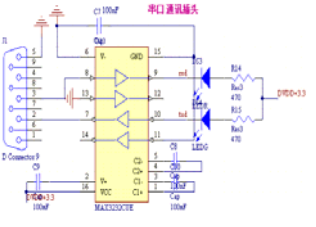


Fig. 6. Serial communication

4 The Transplant From SoC to FPGA

SoC rapid Prototype system Verification is different from other verification method. Any of the other verification methods is part of SoC verification process, while the SoC rapid Prototype system Verification is an entire process. SoC is based on standard cell library, while FPGA is based on themacro-cell block provided by the manufacturers.

Since SoC and FPGA are different in physical structure and performance, the RTL code should be modified first [2]. Then the mapping tools optimize the RTL code based on constraints and map the basic unit of the selected FPGA device to the net list. If the timing satisfies the constraints, the download can be realized using configuration file [6]. If not, we can confirm the critical path according to the timing report to optimize the timing.

Since the synthesized 80C51 IP soft core does not include ROM and RAM, we can use the ROM module provided by QuartusII 7.0 to realize the function. The HEX file for test is embedded by "Mega Wizard Plug-In Manager" provided by QuartusII7.0.

The specific verification process is as follows:

l) Create a new project，add the SoC project files including 80C51 IP soft core and then synthesize the RTL code calling the ROM, RAM and PLL provided by Quartus Ⅱ7.0. ROM modules need to be modified according to actual situation.

The synthesis maps of RTL are shown in Fig.7 and 8. The resource consumption reports after the synthesis based on FPGA is shown in Fig. 9.

2) Write the test file into ROM, and simulate with ModelSim SE6.2b.

3) Complete the pin configuration referencing FPGA pin reference manual, and compiled a downloadable executable file (Top.sof). The resource consumption reports after the download is shown in Fig. 10.

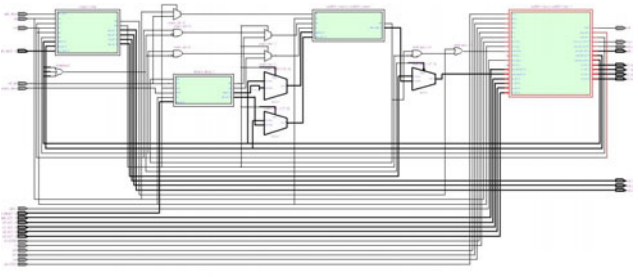


Fig. 7. Synthesis map of 80C51 core and the periphery

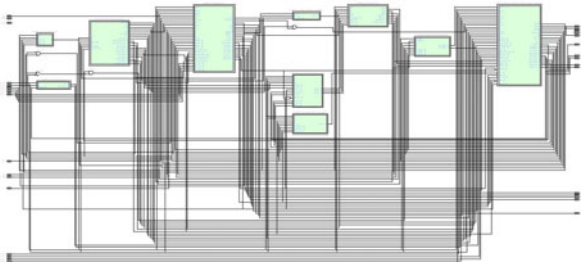


Fig. 8. Synthesis map of 80C51 IP core

Analysis & Synthesis Status	Successful - Sun May 16 15:32:18 2010
Quartus II Version	7.2 Build 151 09/26/2007 SJ Full Version
Revision Name	top
Top-level Entity Name	top
Family	Cyclone II
Total logic elements	3,121
Total combinational functions	3,121
Dedicated logic registers	1,798
Total registers	1798
Total pins	102
Total virtual pins	60
Total memory bits	553,728
Embedded Multiplier 9-bit elements	1
Total PLLs	1

Fig. 9. Resource consumption reports after the synthesis

Flow Status	Successful - Sun May 16 15:35:36 2010
Quartus II Version	7.2 Build 151 09/26/2007 SJ Full Version
Revision Name	top
Top-level Entity Name	top
Family	Cyclone II
Device	EP2C70F896C6
Timing Models	Final
Met timing requirements	No
Total logic elements	3,823 / 68,416 (6 %)
Total combinational functions	3,126 / 68,416 (5 %)
Dedicated logic registers	1,798 / 68,416 (3 %)
Total registers	1798
Total pins	102 / 622 (16 %)
Total virtual pins	60
Total memory bits	553,728 / 1,152,000 (48 %)
Embedded Multiplier 9-bit elements	1 / 300 (< 1 %)
Total PLLs	1 / 4 (25 %)

Fig. 10. Resource consumption reports after the routing

5 Co-verification of the SoC Prototype Verification Platform

The next task is to carry on the Co-verification of the SoC prototype verification platform. HW/SW co-verification environment is shown in Fig. 11.

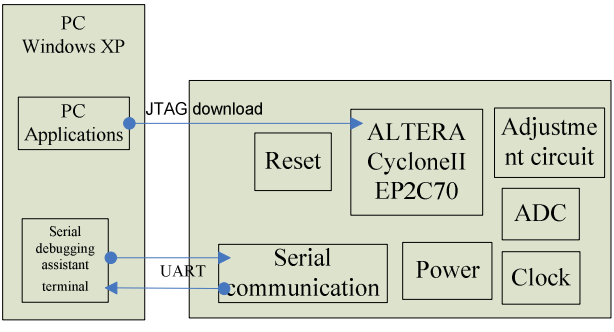


Fig. 11. HW/SW co-verification environment

Verification and debugging process is as follows:

1) Control the verification platform under the control instructions issued by PC Applications. 80C51 control the periphery by configuring registers and output processed signal in parallel and serial way. The input analog signal is shown in fig. 12.

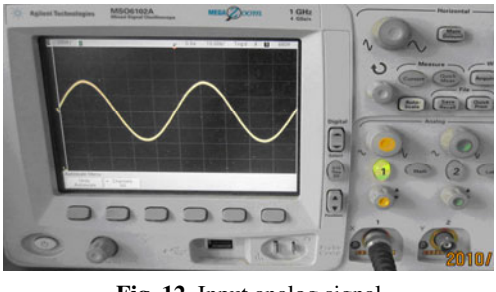


Fig. 12. Input analog signal

2) Observe the run results

a) Observe the feedback information of the hardware verification platform with serial debugging assistant terminal. The processed input analog is transferred to PC, which is shown in fig. 13. The magnification is set two. The result is consistent with expected. It shows that the SoC logic function and serial communication is right.

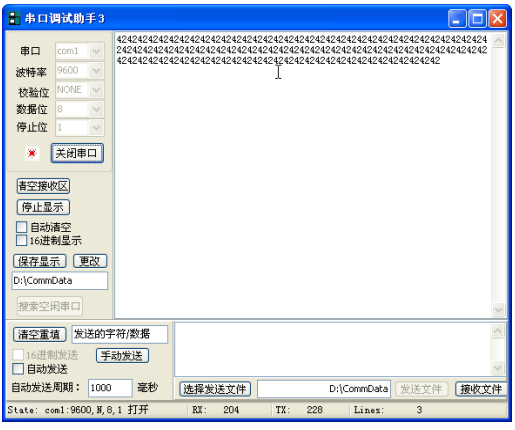


Fig. 13. The serial output data

b) Observe the data from the FPGA internal parallel port by the embedded Signal TapII provided by the QuartusII 7.0. The fig. 14 shows that the parallel data is consistent with the expected and serial data.

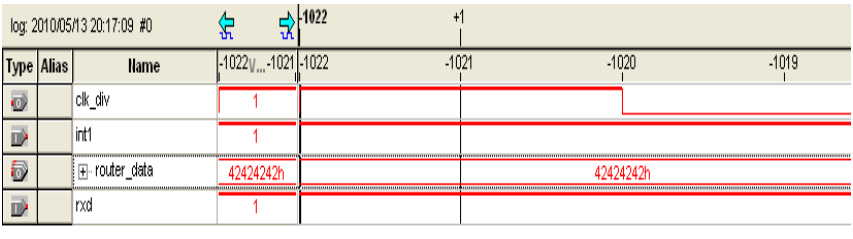


Fig. 14. Parallel output data

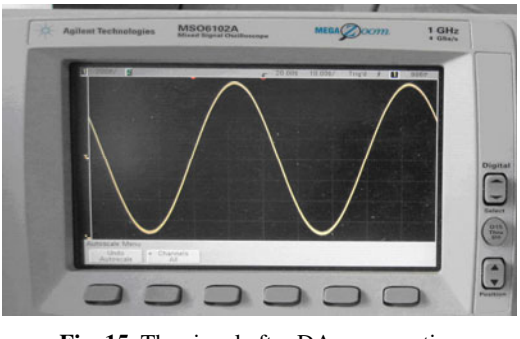


Fig. 15. The signal after DA conversation

c) The parallel digital data is transferred into analog signal by DA conversation, which is shown in fig 15. It shows that the signal is amplified by two as expected.

6 Conclusion

A block-based HW/SW co-design optimization method is proposed in this paper based on Altera EP2C70 FPGA prototype system to optimize the Mixed-SoC design flow. Then, a real-time hardware verification platform is developed to monitor and analysis the behavior and realize the close and flexible coupling of hardware and software. Finally, an independent verification platform is constituted to carry on the field monitoring. Code generated by the HW/SW Co-design in real-time is run on the platform. Control of the entire verification platform is realized and an optimize verification result is achieved.

Acknowledgments. This work is sponsored by PHR (IHLB) 20090513 of Beijing Education committee.

References

1. Wang, J., Li, Z.: Research on the test and application of 80C51 IP soft core. Master's degree thesis, Microelectronics, Beijing Jiaotong University (2007) (in Chinese)
2. Coffman, K.: Real word FPGA design with verilog. Prentice Hall PTR, Englewood Cliffs (2000) (in Chinese)
3. Luo, R.: Research and Implementation of SoC Prototype Verification Based on FPGA. Master's degree thesis, Microelectronics, Beijing University of Technology (2007) (in Chinese)
4. Zhang, Z., Fu, L.: Development Review of SoC Design and verification techodoly. Information Technology Bulletin 3(4), 17–26 (2005) (in Chinese)
5. Li, J., Wang, C., Zhou, H., Zou, Y.: Verification and Simulation of SoC Based on 8051 Core. Microcontroller and Embedded System 9 (2003) (in Chinese)
6. Altera corporation. ASIC to FPGA design methodology and guildlines. Ver.1.0 (July 2003)
7. Jia, Z.: System reset circuit design of MCU. Information Technology (2007)

Design of a H.264 Main Profile Video Decoding Coprocessor*

Limei Xiu[1], Zheying Li[1] IEEE Member, and Weijie Xiu[2]

[1] Institude of Microelectronic Application Tech., Beijing Union University, Beijing, China
[2] College of Mechnical Electronical and Engineering, North China University of Technology, Beijing, China

Abstract. A video decoding coprocessor used as H.264 main profile is presented. In the coprocessor, some decoder accelerating modules are involved, such as inter predictor, intra predictor, IDCT and weighted prediction. When the frequency of the system clock is 133MHz, the average time for this coprocessor to decode 720x480's P frame is 9.2ms. After being synthesized by Design Compiler, 90k gates are totally needed, including 1Kbyte RAM. This coprocessor can be easily integrated in ARM-based SoC and some other processor- based SoC after modification.

Keywords: SoC, H.264/AVC, coprocessor.

1 Introduction

H.264/AVC is the newest, state-of-the-art, video compression standard [1]. Compared with other standards, H.264 has higher compression rate because of many new adding tools, such as flexible inter prediction, CABAC, integer transform and so on. Also much more resources are needed for H.264 decoding, and it will take much more time for general embedded CPU to decode large size H.264 frame. Therefore, in embedded systems or SoCs, decoding coprocessor is needed to accelerate the decoding process in order to meet the requirements of video communications, HDTV and so on.

This paper describes a coprocessor used as H.264 main profile decoder. In the coprocessor many decode accelerate modules are include, such as inter predictor, intra predictor, IDCT and weighted prediction. AHB slave interface is added in the coprocessor to configure the register files used to control calculation progress and transfer decoder parameters. DMA interfaces are used to fetch reference frames and transfer the final results to external ram, and also to fetch the calculation parameters when the coprocessor is worked on parameter buffer mode.

In the simulation at 133MHz clock, the average time for this coprocessor to decode 720x480's P frame is 9.2ms. After synthesis used Design Compiler, 90k gates are needed, include 1Kbyte ram. So this coprocessor can be easily integrated in ARM based SoC. The AMBA slave interface can be modified to other bus interface so the coprocessor can be used in other bus system.

* This work is sponsored by PHR (IHLB) 20090513 of Beijing Education committee.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 355–360.
springerlink.com

2 System Overview

2.1 Hardware/Software Partition

There are two major categories of hardware-software partitioning methodology, ASIPs (Application Specific Integrated Processors) and processor-coprocessor systems [2]. ASIP will modify the main processor core instructions to achieve application-specific function, so the compliers, libraries, operating system functions, and simulation and debugging environment must also be modified. But in processor-coprocessor mode, application specific coprocessor is added and is controlled through register files configuration by the main processor. So, little modification is needed to software development environment.

Most of the SoC designs are based on ARM or other fashion processors. To modify these processors' instructions will cost much for software development and other application work. And because this design is mainly used in AMBA bus system, so we design this H.264 decoder as a coprocessor which has an AHB slave interface to the main processor for decode control and parameters transfer.

In the coprocessor, main calculation modules, which cost most calculation time and less dependent on the software parsing process of NAL, including inter predictor, intra predictor, IDCT, weighted predictor, are designed. The calculation is based on macro block, which includes a 16x16 Luma block and two 8x8 Chroma blocks. Software configures the control and calculation parameter of one macro block, start the decoder, and wait for the end of decoding process, then configure another macro block parameters.

2.2 Decoder Parameter Configure

Register files are used to store macro block parameters. There are two methods to configure the register files. One is configure through AHB slave bus at the beginning of every macro block decoding. The other is software write parameters of many macro blocks to memory, and coprocessor fetches the parameters through DMA to register files.

The first method does not need much memory, but software must wait coprocessor decoding end. Relatively the second method needs much memory, but software needn't wait for coprocessor, so the software can work more independently.

2.3 System Architecture

The final coprocessor architecture is showed as Fig.1.

AHB Slave and DMA interfaces are used to configure register files in control part. IDCT, Intra and inter predictor and other calculation modules are included in calculation part. There are three DMA interfaces in calculation part: one is used for fetching IDCT source residual data, the other is for fetching reference frame for inter predict, and the third is for the final decoded macro block output.

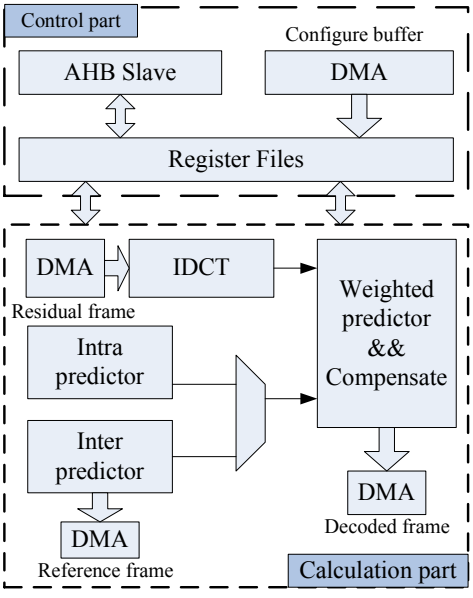


Fig. 1. H.264 coprocessor architectrue

3 Inter Predictor

The inter prediction in H.264 has been made more flexible. This increased flexibility is one of the important changes which improve the level of compression achievable with H.264 [3].

In inter prediction; a number of different block sizes can be used: 4x4, 4x8, 8x4, 8x8, 8x16, 16x8, and 16x16. The reference frame number can be up to 16, and each 8x8 block can have respective reference frame. So one B frame macro block, which include four 8x8 Luma blocks can have 8 reference frames at most. The interpolation position can be half pixel or quarter pixel. First a 6-tap filter for obtaining sample values at half pixel and afterwards a 2-tap filter for calculating sample values at quarter pixel positions [4].

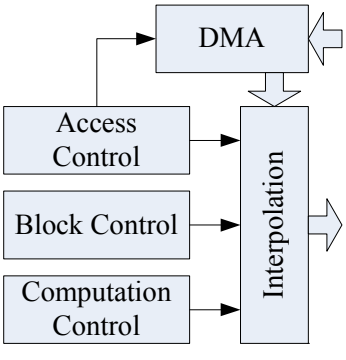


Fig. 2. Inter predictor architectrue

The architecture of inter predictor is shown as Fig.2. Access Control unit controls access to DMA for reference macro block, block control unit is used for different block partition control, and computation control unit take charge of row or column interpolation data input to interpolation unit. Interpolation unit has 6-tap and 2-tap filters and 4 transpose rams to store the temporary data.

4 Intra Predictor

Intra prediction uses row or column of macro block to predict the decoding macro block. A number of intra prediction modes are provide in H.264: nine modes for 4x4 Luma block, and four modes for 16x16 Luma block and Chroma block [5].

Because intra prediction needs the neighbor row or column of the neighbor block in the same frame, these neighbor data must be write as back to register files as calculation parameters before computation begin.

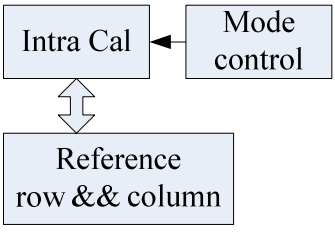


Fig. 3. Intra predictor architecture

5 IDCT and Result Compensation

IDCT in H.264 is simplified compare to MPEG4 and other standard. 4x4 integer transform is used and only shift and add operation is needed [6]. The 2D IDCT is implemented as two 1D IDCT operation. IDCT module architecture is shown as Fig.4.

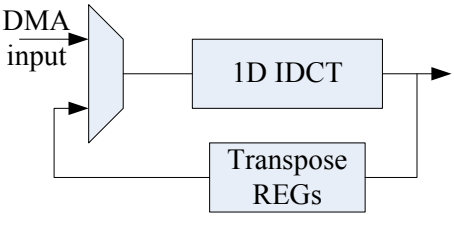


Fig. 4. IDCT architectrue

Result compensation adds the IDCT results and intra or inter prediction results, gains the final decoded results. Weighted prediction function is added in this part. When weighted prediction enable, weighted coefficients configured in Register Files are fetched and used for weighted compensation. Fig.5 shows the Result compensation module architecture.

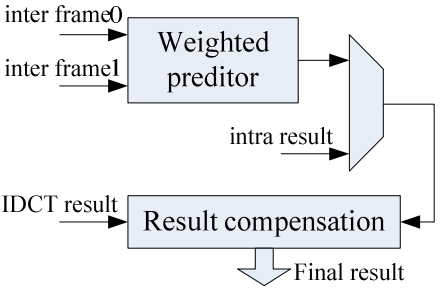


Fig. 5. Result compensation architectrue

6 Conclusion

A verification platform based on System-Verilog is developed, including ARM VMM module, external DDR SRAM, internal data ram. The test clock is set to 133MHz, and the average decoding time for different size video is noted as the follow Table 1.

Finally in TSMC 90nm, synthesis shows the critical path is 6ns, and 90k gate is needed.

Table 1. H.264 Coprocessor simulaion proformance

Image	Size	Decoding Time
Forman	176x144	0.9ms
	320x240	2.8ms
	352x288	4.2ms
	720x480	9.2ms

Acknowledgments. This work is sponsored by PHR (IHLB) 20090513 of Beijing Education committee.

References

1. de Queiroz, R.L., Ortis, R.S., Zaghetto, A., Fonseca, T.A.: Fringe benefits of the H.264/AVC. In: 2006 International Telecommunications Symposium, September 3-6, pp. 166–170 (2006)
2. Kun, Y., Chun, Z., Guoze, D., Jiangxiang, X., Zhihua, W.: A Hardware-Software Co-design for H.264/AVG Decoder. In: IEEE Asian Solid-State Circuits Conference, ASSCC 2006, pp. 119–122 (November 2006)
3. Hu, Y., Simpson, A., McAdoo, K., Cush, J.: A high definition H.264/AVC hardware video decoder core for multimedia SoC's. In: 2004 IEEE International Symposium on Consumer Electronics, September 1-3, pp. 385–389 (2004)

4. Yang, S., Zhenyu, L., Goto, S., Ikenaga, T.: A VLSI architecture for motion compensation interpolation in H.264/AVC. In: 6th International Conference on ASIC 2005, ASICON 2005, October 24-27, vol. 1, pp. 279–282 (2005)
5. Staehler, W.T., Berriel, E.A., Susin, A.A., Bampi, S.: Architecture of an HDTV Intraframe Predictor for a H.264 Decoder. In: 2006 IFIP International Conference on Very Large Scale Integration, October 16-18, pp. 228–233 (2006)
6. Stabernack, B., Wels, K.-I., Hubert, H.: A Video Coprocessor for Mobile Multi Media Signal Processing. In: IEEE International Symposium on Consumer Electronics, ISCE 2007, June 20-23, pp. 1–6 (2007)

An Authentication System Using Finger Vein Features*

Haitao Xu[1], Zheying Li[2], and Xianju Guo[3]

[1] School of Electronics Engineering, Beijing Jiaotong University,
100044, Beijing, China
haitaoxu84@hotmail.com
[2] Institute of Micro-electronic Application Tech., Beijing Union University,
100101, Beijing, China
zheying@buu.edu.cn
[3] School of Electronics Engineering, Beijing Jiaotong University,
100044, Beijing, China
gxjsky386@gmail.com

Abstract. Compared with the fingerprint, finger vein is blood vessel network under finger skin, which is unique for each individual and hard to be forged. In this paper, an authentication system based on Nios Ⅱ soft-core CUP is designed, which uses finger vein features as identity information. To overcome the problem of high time-consuming of software algorithms brought by the use of Nios Ⅱ, a method of hardware acceleration is introduced. Through time analysis of various parts of the image processing algorithm, high time-consuming algorithms are hardware accelerated and the requirement of real-time system is met.

Keywords: finger vein, Nios Ⅱ, FPGA, image processing, hardware acceleration.

1 Introduction

How to accurately identify the individual's identity information, and protect their information security, is one of the problems need to be solved in today's information society. Finger vein recognition technology is an emerging biometric identification technology. Compared with the fingerprint, finger vein is a blood vessel network under finger skin, which is unique for each individual and hard to be forged. According to statistics, only 8 per 1,000 people have similar hand vein distribution [1]. Finger vein image capture is easy, and takes up small storage space. So, as an important feature on authentication system, the biometric finger vein network also has the stability, high accuracy, low cost, and non-invasive characteristics.

The system is based on Nios Ⅱ soft-core processors, which realize the information of finger vein characteristic processing systems. Compared with the PC implementation, the system is low cost, low power consumption, high flexibility, reconfigurability and easy to realize. The system in this paper uses human finger vein as the identity

* This work is sponsored by PHR (IHLB) 20090513 of Beijing Education committee.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 361–368.
springerlink.com

information. First, capture the user's finger vein of information in real-time. Then do the finger vein image processing. Finally compare the result of image processing and original personal information in the database to make the information authentication. So, customers need to get their finger vein information collected and stored in the Finger Vein Information Database which is established firstly. As finger vein is less influenced by the growth of human body, it can be ensured that the database will be available in no less than 5 years.

Finger vein information processing system is a real-time image acquisition and processing system, which need to run and finish within the given time. Through time analysis of various parts of the image processing algorithm, high time-consuming algorithms are hardware accelerated and the requirement of real-time system is met.

The rest of this paper is arranged as follows: Section 2 gives a brief introduction of the hardware architecture of the system and functional description of each module. Section 3 describes the image processing software algorithms, and analyzes the time-consuming and reasons of each parts. Section 4 introduces the hardware acceleration to the main time-consuming algorithm. Some conclusions of this paper are given in section 5.

2 System Hardware Architecture

Using Nios Ⅱ soft-core as the platform of image processing and success in the FPGA transplant, the system includes two parts: Hardware image acquisition and software image data processing. The hardware system is divided into outside image acquisition devices and FPGA embedded platform. The acquisition device connected to the system through the analog video port on the development platform. Nios soft-core is used for running the feature extraction algorithms and the authentication process. By designing the appropriate custom peripherals to accelerate algorithms and ensure the system's real-time. Connecting to the database through UART, users' characteristics and their original information in database were compared. Meanwhile, AES algorithm is added in the system to ensure the confidentiality of users' characteristics information. The system architecture is shown in Figure 1.

2.1 Finger Vein Image Acquisition Device

With a wavelength of 700nm-1000nm infrared light irradiation fingers, most organizations will be penetrated. As infrared light will be absorbed by hemoglobin in the finger vein, the transmission infrared light from the vein network portion is weak, and none vein network areas one is strong. With the help of image sensor, the finger vein network topology is formed. The acquisition system is shown in reference [3].

2.2 IP Design

Beside Nios Ⅱ, there are other IP cores designed in FPGA, including: Image acquisition control and decoding module, image hardware-accelerated processing module, SDRAM interface control module, VGA display control module, FIFO, on-chip shared RAM and custom peripheral module, which are shown in Figure 1. There are

two independent data channels after video decoding. One is through the SDRAM controller and the VGA controller which directly output to VGA monitor, the other is through the image hardware-accelerated processing module and FIFO, then stored in the On-chip RAM which is shared by Nios Ⅱ.

(1). Image acquisition control and decoding module: This module is used for controlling video decoding chip ADV7180 via I2C bus for the analog signal capture.

(2). Image hardware-accelerated processing module: Since only the gray scale is used in the image date processing, in order to streamline the storage space and speed up image processing, the module is used for capturing a frame image from the NTSC standard data flow and extracting the Y component. This module is controlled by Nios Ⅱ, and provides the data address defined by the user.

(3). FIFO and on-chip shared RAM: As the clock of the video capture is 28MHz and the on-chip shared RAM is 100MHz, so FIFO is used in the cross-clock domain part of system. The on-chip shared RAM is used for storage the image data shared by FIFO and Nios Ⅱ.

(4). SDRAM control module and VGA display control module: After decoded by video decoding chip, the image signal can go through this channel and be output to the VGA. Each frame of NTSC video data is storage in the SDRAM dynamically, and then converted into RGB signals by VGA display control module, finally, displayed in the VGA.

(5). Custom peripheral module: as the clock of Nios Ⅱ is 100MHz, the real-time requirement of the system can not be met by using software to run the image data processing. After analyzing time-consuming of each step of the image processing algorithm, high time-consuming algorithms are hardware accelerated to reduce their run-time. So the real-time requirement of system is achieved.

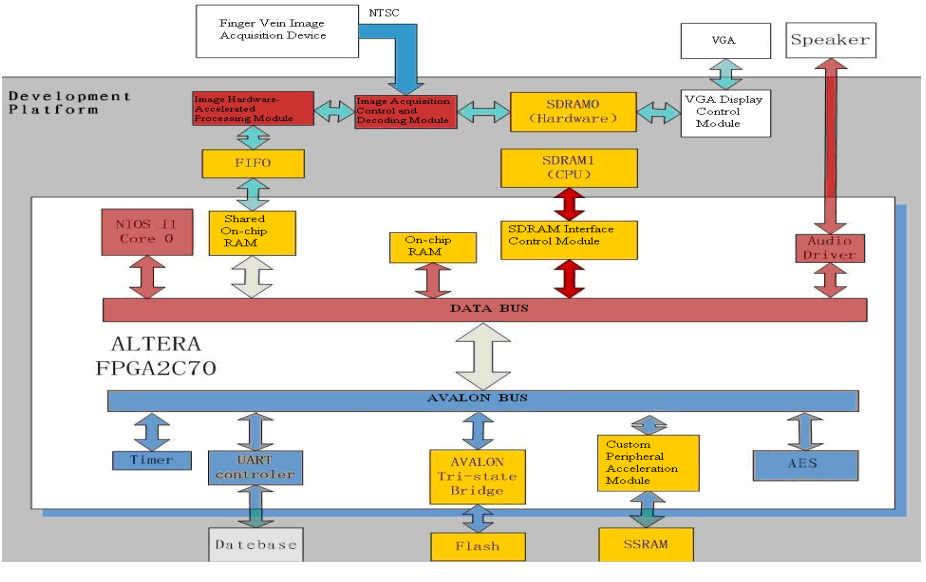


Fig. 1. Finger vein authentication system structure

3 Image Data Processing Algorithm

Image data processing is run on the Nios Ⅱ, including three parts, which are preprocessing, venous contour extraction and authentication. The data processing flow is shown in Figure 2.

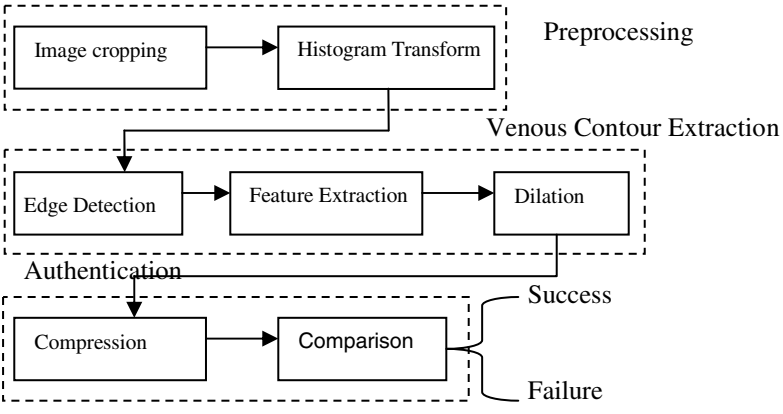
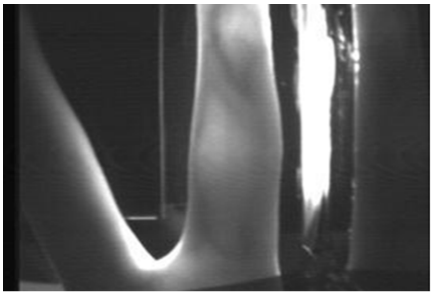
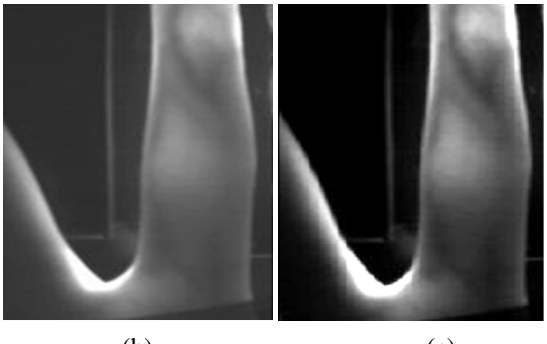


Fig. 2. Data processing flow



(a)



(b) (c)

Fig. 3. The result of image preprocessing: (a) Original image. (b) Image after copping. (c) Image after histogram transforms.

3.1 Image Preprocessing

Due to the limited of the image quality, the image preprocessing is required, including image cropping, and histogram transform. Image cropping is used for deleting unnecessary background information, which is realized by hardware. Original picture size is 480*330*8bits, and changed to 210*330 after cropping. Histogram transform can enhance the gray-scale image to strengthen its finger vein information. After cropping, the image size is 210*330. So, it can be estimated that the histogram transform algorithm cycles about 140 thousand times. Figure 3 shows the results of image preprocessing.

3.2 Venous Contour Extraction

Venous contour extraction includes edge detection, feature extraction, and dilation. The application of edge detection algorithm is Canny operator [6]. The area of finger vein can be defined by the closed region of finger contour. Reference [6] shows that advanced functions, such as trigonometric functions, power function and exponential function, domain the edge detection algorithm. After the area of finger vein is defined, vein topological features can be extracted by the feature extraction algorithm. The steps of the feature extraction algorithm are shown as follows:

1) Set outside of the image to zero to eliminate the feature points generated from the image border.

2) According to the region of fingers vein create a same size memory matrix Z, and initialize each value to zero.

3) The general finger vein width in image is 26 pixels, so set the vein width w = 26.

4) Set the count value c = 0. Successively select points within the given region. Compare the selected points to their four sides, of which the size is 26*26. If the value of both the top and bottom sides or both left and right sides is less than the centre, then c = c + 1. That means each point should be compared to it's around 104 times

5) After comparison of each point, if c>8 and its value (gray scale) is less than 170, then set the corresponding value in matrix Z to 255.

6) Skeletonize the extraction features.

To reduce the storage space, the binarization of vein features is used, i.e. 1 and 0 stand for whether or not this position is the region of finger vein. The statistics shows that the average closed region of the finger contour is 120*330, so there are about 2 million cycles.

Dilation is used for expanding the finger vein to eliminate the impact made by small deformation of finger. The main part of the algorithm is cycle, about 40 thousand times. Figure 4 shows the results of extraction algorithm in different fingers.

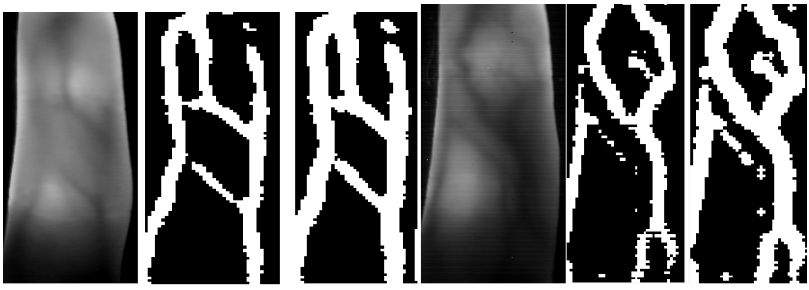


Fig. 4. Original, vein extraction and dilation images of two different fingers

3.3 Authentication

To further reduce vein information storage space, the vein feature is 3*3 compressed after binarization. I.e. vein feature matrix is divided into several 3*3 sub-matrixes, if the number of 1 in each sub-matrix is greater than 4, then the sub-matrix is replaced by 1, otherwise use 0 instead of. The new matrix storage space can be shortened to 1 / 9, and, as a result, reduce the running time of authentication algorithm. Number of cycles of the algorithm is about 4 million times.

There are three methods of comparison to achieve authentication. The one is frequency comparison. Via DCT, the vein image is transformed into the frequency domain. An advantage of this approach is high recognition success rate, but the drawback is the high time-consuming as complexity of the algorithm. The second is recognition algorithm using phase only correlation [5]. When the vein information is rich, the algorithm accuracy is high, but in the case of the information-poor veins, becomes very low, with strong instability. The third comparison method is point by point comparison algorithm, i.e. compare the vein features point by point in a specific region. Advantage of the algorithm is simple and easy to implement, the disadvantage is low recognition success rate due to its high constraints. In this system, as the use of Canny operator and dilation algorithm, the region of the finger vein can be defined automatically, and the small deformation would be offset, which will greatly reduce the comparison constraints. So the third method is used in the system, and the recognition success rate can reach a high level. After the compression algorithm the vein information matrix size is 40*110, so the number of algorithm cycles is about 5 thousand times.

4 Algorithm Timing Analysis and Hardware Acceleration

Figure 3.1 shows that the image processing algorithm includes image cropping, histogram transform, edge detection using Canny operator, vein feature extraction, dilation, compression, and point by point comparison. Since the system is user-oriented operation, the system running time is set to no more than 2 seconds to meet the individual's tolerance.

Because the image cropping done by the hardware, under the 100MHz of operating frequency, the time-consuming is about 2ms, and can be negligible. The high time-consuming algorithms are: 1) Edge detection algorithm (Canny operator), 6.8s. 2) Feature extraction algorithm, 4.1s. Running time of each part of the algorithm detailed

in table 1 below. Thus, the total running-time in software can not meet the system's real-time requirements. The major time-consuming algorithms, edge detection algorithm and feature extraction algorithm, need to be hardware accelerated.

Table 1. Software running time of each part of the algorithm

	Software running time:(CPU 100MHz)	Algorithm characteristics:	Remarks:
Histogram Transform	0.3s	Cycle	About 140 thousand
Edge Detection	6.8s	Advanced computing	Trigonometric function, power function
Feature Extraction	4.1s	Cycle	About 2 million
Dilation	79ms	Cycle	About 40 thousand
Compression	81ms	Cycle	About 40 thousand
Comparison	11ms	Cycle	About 5 thousand
Total Time:	11.4s		

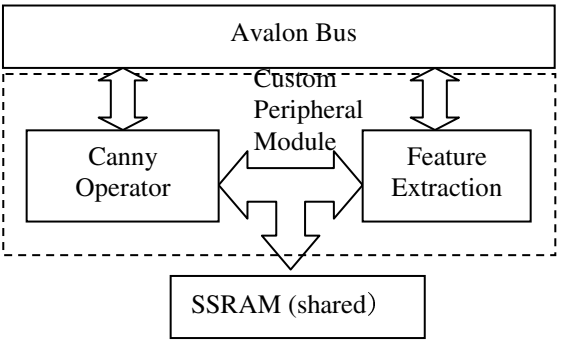


Fig. 5. Custom peripheral system structure

The hardware structure of edge detection algorithm (Canny operator) is shown in reference [7]. The image date stored in SDRAM1 is transferred through Avalon and processed by hardware. The result of edge detection is stored in SSRAM. Finally, a "done" signal will be sent to feature extraction module when the processing is over. Hardware vein feature extraction module is running in following four steps:

(1). Define the region of finger vein according to the edge data stored in the SSRAM.

(2). Process the defined region of image stored in SDRAM1, i.e. compare the selected point to four sides of rectangular around, and store the comparison result to SSRAM.

(3). If all points in the selected region are processed, read the finger vein result in SSRAM and store it to SDRAM1 through Avalon bus.

(4). Send the conversion completion signal to the CPU and stop operation.

The structure of custom peripheral system is shown in fig.5. After hardware-accelerated, the total running time is reduced to 0.6s, to meet the system real-time requirement. The details are shown in table 2.

Table 2. Final running time after hardware-accelerated

	Running time：	Realization：
Image cropping	2ms	Hardware
Histogram Transform	300ms	Software
Edge Detection	1.8ms	Hardware
Feature Extraction	20ms	Hardware
Dilation	79ms	Software
Compression	81ms	Software
Comparison	11ms	Software
Total Time:	<0.6s	

5 Conclusion

The system uses FPGA-based IP design and Nios Ⅱ soft core processor. Nios Ⅱ is used for software scheduling and controlling. While, some high time-consuming functions, such as some transform algorithms and video data stream processing, are implemented in hardware via HDL. Though the analysis of various parts of image processing software algorithms, high time-consuming ones are hardware accelerated and called by Nios Ⅱ to meet the time requirements for data processing.

References

1. Yu, D.: The Identification Technology of Living Things Feature. China Criminal Police (1), 56 (2008) (in Chinese)
2. Kalomiros, J.A., Lygouras, J.: A host/co-processor FPGA-based Architecture for Fast Image Processing. In: Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, pp. 373–378 (2007)
3. Mulyono, D., Jinn, H.S.: A study of finger vein biometric for personal identification. In: Biometrics and Security Technologies (ISBAST 2008), pp. 1–8 (June 2008)
4. Zhang, X., Li, Y., Wang, J., Chen, Y.: Design of High-Speed Image Processing System Based On FPGA. In: The Ninth International Conference on Electronic Measurement & Instruments (ICEMI 2009), pp. 65–69 (2009)
5. Mahri, N., Sundi, S.A., Rosdi, B.A.: Finger Vein Recognition Algorithm Using Phase Only Correlation. In: Emerging Techniques and Challenges for Hand-Based Biometrics (ETCHB 2010), pp. 1–6 (2010)
6. Wang, B., Fan, S.: An improved CANNY edge detection algorithm. In: Computer Science and Engineering (WCSE 2009), pp. 497–500 (2009)
7. He, W., Yuan, K.: An improved Canny edge detector and its realization on FPGA. In: Intelligent Control and Automation, WCICA 2008, pp. 6561–6564 (2008)

Excitation Parameter Identification Based on the Adaptive Inertia Weight Particle Swarm Optimization

Wei-hua Zha[1,2], Yue Yuan[1], and Tao Zhang[2]

[1] College of Electrical Engineering, Hohai University, Nanjing, 210013
[2] Nuclear Power Qinshan Joint Venture Co., Ltd, Zhejiang, Haiyan, 314300

Abstract. Particle Swarm Optimization of excitation system parameters identification in the study, A new adaptive inertia weight particle swarm optimization of excitation parameters used in the identification. Test proved that the new algorithm more powerful search capabilities to overcome a swarm of elementary particles easily fall into the shortcomings of local optimum.

Keywords: Parameter Identification, PSO, Adaptive, Inertia weight, Excitation.

1 Introduction

System identification is through the observation of a system or a process the relationship between input and output, describe the system or process to determine the input and output relations describing the system or process to determine the dynamic characteristics of the mathematical model. Measurement system in accordance with the awareness level of treatment, the system under test can be divided into the black box system, gray box systems, white-box system. Excitation system is a gray box system, in accordance with the physical mechanism of a mathematical model, and then find the parameters of system identification. In this paper, time domain identification method, which first identified the measurement system to treat non-parametric characteristics, time domain response obtained, and then the dynamic fitting technique, to strike from the dynamic characteristic curve model parameters.

Normal operation of power system operation or accident, synchronous generator excitation system plays an important role. It has a control voltage to control the distribution of reactive power to improve the stability of synchronous generators operating in parallel, improving the ability of power system stability[1,2]. Most power system simulation software comes with some of the standard model of excitation system for the use of simulation analysis, however, the original model of the actual system is often not the software in the standard model. Therefore, the parameter identification method must be the actual model of the original model simulation software into the standard model of the library. In other words, the simulation software to determine the

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 369–374.
springerlink.com

model parameters, so that the actual excitation system model has the same or similar characteristics, for calculations.

The particle swarm optimization (PSO) algorithm was first introduced by Kennedy and Eberhart in 1995[3]. As simple and easy to implement and robust features, has now been applied to many fields. However, the basic PSO algorithm to balance the global search and local search on the deficiencies. PSO updates the particle degree of inertia weight in the formula of the system global and local search plays an important role. In this paper, an adaptive inertia weight strategy, according to the objective function value of each particle, adaptive selection of inertia weight. The objective function value of current particle of all particles smaller than the average target value, to reduce the flight speed of particles; the current objective function value is greater than the particles the average of all the particles, the particles to accelerate the speed, set it as the maximum. Adaptive inertia weight particle swarm algorithm can meet their needs for each particle, which greatly improved the diversity of the population of particles to improve the PSO's search capabilities. The basic PSO algorithm with the known excitation system parameter identification comparative study to prove the validity of the algorithm and advanced.

2 Mechanism of Excitation System Parameters Identification

Excitation system model and simulation software in the original model structure and parameters are known, the parameter identification task is to determine the parameters of the standard model, so that the original excitation system model and the standard model of the input and output to maintain a certain error range. Identification principle is shown in Figure 1

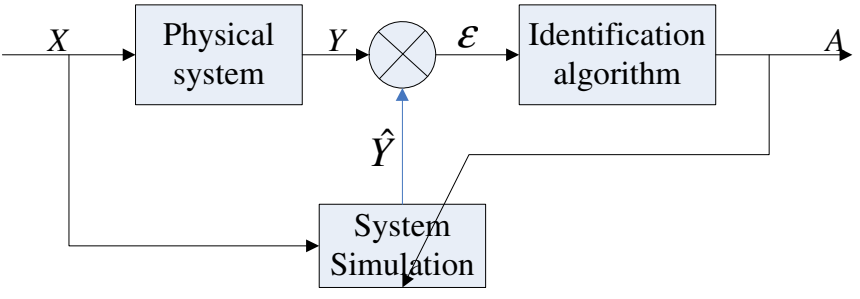


Fig. 1. Schematic diagram of parameter identification

Parameter identification process is, in the same excitation signal r under the influence of the physical model of excitation system produces output $\hat{y}(t)$, The standard model of excitation system produces an output signal $y(t)$, both error is ε, after the

standard identification algorithm continuously adjust and optimize the model parameters until the error is ε until the minimum[5].

3 Standard PSO Algorithm[3]

The basic idea of PSO: random in the solution space to initialize a random population, which contains a number of particles, these particles in the solution space seeking the position of representative of solution to the problem, particles in the solution space to obtain the best according to their current location and the entire population in the best position to determine the current flight path, step by step approach the optimal region. Let $\mathbf{z}_i = (z_{i1}, z_{i2}, \cdots, z_{id}, \cdots, z_{iD})$ as the first i particles (i = 1,2, ..., m) of the D-dimensional position vector, Let $\mathbf{v}_i = (v_{i1}, v_{i2}, \cdots, v_{id}, \cdots, v_{iD})$ for the particle i of the flight speed, that is, particles moving distance; Let $\mathbf{p}_i = (p_{i1}, p_{i2}, \cdots, p_{id}, \cdots, p_{iD})$ to search for the particles so far the best location; $\mathbf{p}_g = (p_{g1}, p_{g2}, \cdots, p_{gd}, \cdots, p_{gD})$ for the whole particle swarm's best position so far to search. In each iteration, according to the following formula particle velocity and position updates:

$$v_{id}{}^{k+1} = w v_{id}{}^{k} + c_1 r_1 (p_{id} - z_{id}{}^{k}) + c_2 r_2 (p_{gd} - z_{id}{}^{k}) \tag{1}$$

$$z_{id}{}^{k} = z_{id}{}^{k} + v_{id}{}^{k+1} \tag{2}$$

k is the number of iterations, r1 and r2 [0,1] random number between. r1 and r2 is called learning factors or accelerating factor, which makes particle self-summary and excellent individuals to groups the ability to learn to the best advantage to their own history and the history within the group close to the most advantages. The first is the inertia weight, ability to play an optimal balance of local and global optimal capacity; the second is the "cognitive" part, on behalf of the particles on their own learning; second is "social" part, on behalf of collaboration among the particles.

4 Adaptive Inertia Weight PSO

When a particle in a local optimum, it may always be the best time evaluation, due to the interference of local optimum, resulting in a premature local optimum particles. According to the objective function value of each particle, adaptive selection of inertia weight. The current objective function value is less than the particle, the average target value of all the particles, to reduce the flight speed of particles; the current objective function value is greater than the particle, the average target value of all the particles, the particles to increase the maximum flight speed. Equation (1) the selection of inertia weight:

$$w_i = \begin{cases} w_{\min} + \dfrac{(w_{\max} - w_{\min})(E_i - E_{avg})}{E_{avg} - E_{\min}}, & E_i \le E_{avg} \\ w_{\max}, & E_i > E_{avg} \end{cases} \tag{3}$$

One, $w_{\min}$ and $w_{\max}$ represent the w of the minimum and maximum values, f for the particle current objective function value, $E_{\min}$ and E_{avg} are the minimum and average of all target particles. Objective evaluation function is the mean square error:

$$\varepsilon = \sum_{i=1}^{T} (y_i - \hat{y}_i)^2 \tag{4}$$

One, T as the number of sampling points, $\hat{y}_i$ as the best target.

5 Example

The following is based on adaptive inertia weight particle swarm optimization of excitation system parameters identification of specific steps:

Step1: In the Matlab software, excitation system were established in the original model and the standard model, the equation (4) as the parameter identification of the objective function;

Step2: the standard model to determine the parameters to be identified;

Step3: Termination conditions and given the maximum number of iterations;

Step4: For each particle is initialized;

Step5: Enter the excitation signal, the simulation calculation, based on the original model of excitation system, the standard model of particle output error calculated fitness.

Step6:Particle swarm optimization using parameter identification

Step7:When the objective function value less than or equal eps, the end loop($eps = 2.2204 \times 10^{-16}$)

Model (Figure 2), the parameter value: $k_d = 5.0513, k_p = 100.718, k_i = 2.8965$

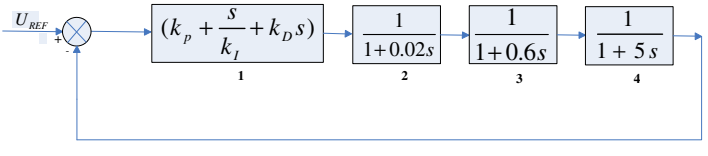


Fig. 2. Excitation System Model

1:PID;2: Regulator;3: Exciter;4: Generator

Figures 3 and 4 are SPSO trend curve fitting with WSPSO

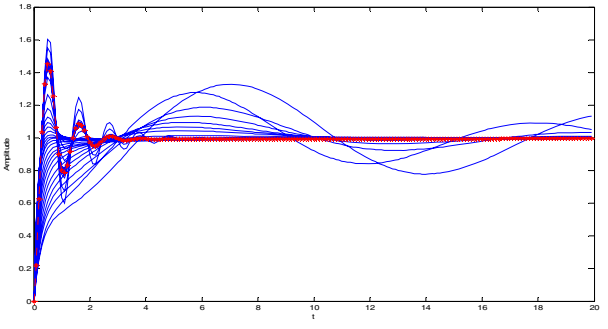


Fig. 3. PSO algorithm based on the standard output curve fitting trend

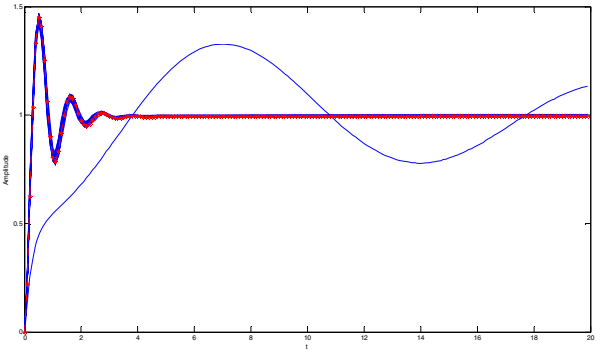


Fig. 4. Adaptive inertia weight PSO algorithm output curve fitting trend

Red response curve for the actual model output; Blue Wave as the standard system output. Figure 3 and Figure 4, we can draw contrast, the basic optimization ability of PSO algorithm is significantly higher than adaptive inertia weight PSO algorithm optimization poor. Identification results shown in Table 1:

	err	**Algebra**	**time**	$\mathbf{k_D}$	$\mathbf{k_p}$	$\mathbf{k_I}$
SPSO	eps	200	55	5.0513	100.718	**2.8965**
WSPSO	eps	100	42	5.0513	100.718	**2.8965**
Real value				5.0513	100.718	**2.8965**
Limit				6	200	**5**
Lower limit				**1**	**1**	**1**

6 Conclusion

According to the objective function, the dynamic adjustment of flight speed particle inertia weight PSO can be well balanced regulation of local search and global search

capability. Avoid premature local optimum particle swarm. Basic PSO with the excitation system parameters identification comparison proved adaptive inertia weight particle swarm algorithm, accuracy and advanced.

References

1. Kundar, P.: Power System Stability and Control. McGraw-Hill, NewYork (1994)
2. Huang, C., Wang, M.: A particle swarm optimization to identifying the ARMAX model short-term load forecasting. IEEE Trans. on Power Systems 2(20), 1126–1133
3. Eberhart, R.C., Kennenedy, J.: Particle swarm optimization. In: Proceedings of IEEE International Conference on Neural Networks, Perth, Australia, November 27-December 1, pp. 1942–1948 (1995)
4. Angeline, P.J.: Evolutionary optimization versus particle swarm optimization and philosophy and performance difference. In: Proceedings of 7th Annual Conference on Evolutionary Programming, San Diego, CA, USA, March 25-27, pp. 601–610 (1998)
5. Feng, H.: Hybrid stages particle swarm optimization learning fuzzy modeling systems design. Tamkang Journal of Science and Engineering 2(9), 167–176
6. Ping, J.: Nonlinear Parameter Identification of Power System. Hohai University Press, Nanjing (1999)
7. Eberhart, R., Kennedy, J.: A new optimizer using partilce swarm theory. In: Proc. 6th Int Symp. Micro Machine and Human Science, pp. 39–45 (1995)

Biographies

Wei-hua Zha (1978-) was born in Anhui province, on July 20,1978. Ph.D,to major power system planning and designing.Email:zwhcsm@163.com
Yue Yuan(1966-) Male, Ph.D., Professor, research interests include optimal operation of power systems, power system stability analysis and control, power economic and technical and operations management
Tao Zhang(1968-) Male, professorial senior engineers, operations research and management for nuclear power

The Page Scheduling Schema Exploiting Periodic Access Pattern

Fang Hou[1,2], Yue-long Zhao[1], and Yu Wang[1]

[1] School of Computer Science and Engineering, South China University of Technology, Guangzhou China, 510641
[2] Department of Computer Science and Technology, Guangdong University of Finance Guangzhou China, 510521
hfhoufang@126.com

Abstract. The Temporal locality and spatial locality have been repeatedly studied and widely exploited in the computer storage hierarchy. They are crucial principles of cache and buffer. But current spatial locality researches have been confined to a small time scale such as several instruction cycles. The periodic page accesses pattern can be discovered by expanding the time scale to seconds. To reduce the rate of cache and buffer miss is the primary goal of most cache and buffer management algorithms. When the miss is inevitable, the practical solution is to reduce the miss cost. A buffer cache management algorithm (Periodicity and Miss Latency Algorithm, PLC) has been proposed based on the periodic page accesses pattern and page miss cost evaluation. By keeping the pages that cost a lot of instruction cycles to swap in or out in cache/buffer as long as possible, the PLC algorithm has been proved practical and efficient.

Keywords: page scheduling, management algorithm, periodicity, latency cost.

1 Introduction

To balance the conflict of performance and cost, the hierarchical architecture has been introduced into the storage subsystem. This hierarchical architecture consists of three levels of cache in CPU, RAM, and External Storages. The caches and buffers between two levels make it possible to get the approximate performance of higher level at the approximate cost of lower level.

Since the caches are built by the CPU-similar techniques, their prices are high and capacities are restricted. A lot of algorithms have been proposed to maximize the efficiencies of the cache and buffer. The most common strategies are as followings:

Optimal (OPT): This is the theoretic best strategy to decide which exist data or instruction in the cache or buffer will be replaced by the pre-fetched one. This strategy will swap out the data or instructions whose next use will occur in the farthest future. Since it need to know the information about future, it can only been deployed for algorithms comparisons.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 375–382.
springerlink.com

Least Frequently Used (LFU): This algorithm swaps out the page accessed least frequently. The access times are recorded in an access counter.

Random (RDM): It just randomly selects out the page to be replaced by new one.

Least Recently Used (LRU): This algorithm takes the least recently used content as the farthest future used. It is an acceptable simulation of optimal algorithm.

Fist In First Out (FIFO): It just simply moves out the head content of the cache or buffer queue.

Clock (CLK): The CLK is a combination of LRU and FIFO in practice. Because the LRU needs several bits to record its time stamp, it is a great consumption in the capacity restricted cache or buffer. The CLK algorithm uses a ring to organize space of the cache or buffer. A pointer is designed to clockwise check pages of the ring. The first page encountered with used-bit of value 0 will be replaced. The other pages' used-bits will be switched from 1 to 0.There are a lot of CLK algorithm variations have been adopted by Unix/Linux operating systems to manage their caches and buffers.

The temporal and spatial localities [1] are the fundamental principles of these cache and buffer management algorithms. This locality theory tells us that most of instruction requests and data requests occur very closely in time scale or/and spatial scale. If we pre-fetch the instructions and data that have been stored near to the instruction and data being processed, the seek time and reading latency will be reduced while the real fetch is required.

2 The Page Access Behaivior Analysis

2.1 The Latency Cost of Page Miss

To reduce the rate of cache and buffer miss is the primary goal of most cache and buffer management algorithms. When the miss is inevitable, the practical solution is to reduce the miss cost.

A specific page that be requested by an active process exists in cache of CPU, working set in RAM, cache in Hard Disk's controller circuit or 8 sectors of a disk's plate. Hence the request will be satisfied by cache hit or swap-in with latency cost. Table 1 shows how the storage system works when a page access is requested.

Table 1. Cache hit and page swap-in

	Access CPU's Cache	Access RAM	Access HD's Cache	Cache Hit	Swap-in To Higher Hierarchy
Page in CPU's Cache	Hit	-	-	1 time	0 time
Page in Working Set	Miss	Hit	-	1 time	1 time
Page in HD's Cache	Miss	Miss	Hit	1 time	2 times
Page in Sectors	Miss	Miss	Miss	0	3 times

The access latency of the top 3 hierarchies such as CPU's cache, working set in RAM and HD's cache is determined by electronic pulses of semi-conductor. It takes several nanoseconds or microseconds. While the access latency of the sector access

on a disk plate depends on mechanical movement. The average track seek time latency and rotary latency of a Seagate's ST3750528AS are 8.5 milliseconds and 4.17 milliseconds [2]. The latency cost takes 12.67ms which is thousands of times higher than the former.

2.2 The Periodic Page Access Pattern

Current spatial locality researches have been confined to a small time scale such as several instruction cycles. The periodic page accesses pattern can be discovered by expanding the time scale to seconds.

DiskMon [3] is a system event tracing tool provided by Microsoft to monitor disk access activities in Windows environment. It runned in an active web page server to record any disk access event in our laboratory. A department-level business assessment system has been deployed in this server. The hardware and software inventory of this server are shown in Table 2:

Table 2. Inventory of Server

Components	Specifications
CPU	Intel Core E6300
RAM	DD2-667 4GB
Hard Disk	Seagate ST80P15K scis 300GB
Operating System	Windows Server 2003 SP1
Internet Information Services	MS IIS 6.0
Active Page Services	ASP.net 2.0
Database	SQL2005
Development language	C#2005

The DiskMon's monitoring lasted 900 seconds. Table 3 illustrates the distribution of disk sector accesses. The percentage of single access to a specific sector is less than 28%.

Table 3. Distribution of Disk Accesses

Range of Sector Accesses	Quantity of Sectors in the Range	Quantity of Accesses in the Range	CDF of Access Times
1time	1335	1335	27.3%
2~5times	310	913	46.0%
6~10times	88	683	60.0%
11~15times	33	423	68.7%
16~20times	15	265	74.1%
21~30times	4	100	76.2%
31~100times	6	305	82.4%
More than 100	2	858	100%

The layout of disk accesses is shown in Figure 1. The horizon axis means time, while the vertical axis stands for the sequence number of disk sector.

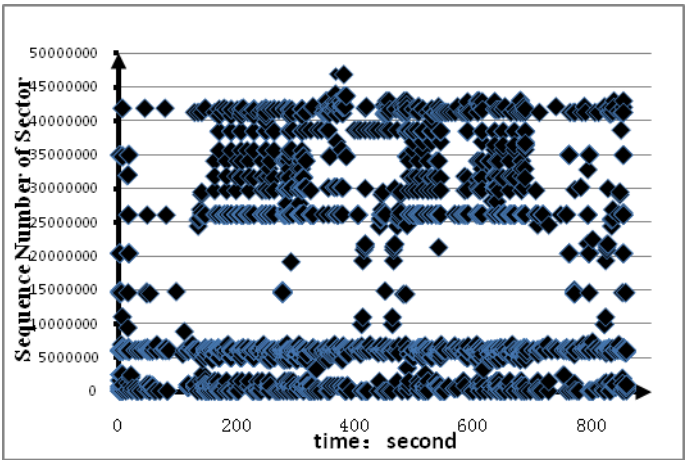


Fig. 1. Sector Accesses Layout

The DiskMon monitors disk access activeties by the unit of sector. Universally, a sector covers 512 Bytes, and a pge covers 4 kilobytes both in Windows and Linux environment. Each page consists of 8 sectores. Consequently, 2 disk accesses to the same sector means repetitive visit to a same page. Even 2 disk accesses to adjacent sectors most probably means visit to a same page. It makes acceptable to assess the periodical page accesses by monitoring sector accesses.

The periodic page access pattern can be defined from the figure and table above:

1) Page accesses concentrate on in a small scope of the sectors.
2) These partial sectors are accessed at intervals.

3 The New Page Scheduling Schema

As we have discussed in 2.1, the performance of storage subsystem can be improved by reducing the miss rate. Many cache management algorithms try to keep the pages that most probably be accessed in near future. The methodology of these algorithms is to maximize the possibility of cache hit [4][5].

On the other hand, it is obviously that the cost of cache miss is expensive, especially when the miss happens in the HD's cache. Accordingly, we propose a new cache management schema with the methodology of minimizing the cost of cache miss based on the periodic page access pattern. We name it a PLC (Periodicity and Latency Cost) scheme.

Traditionally, the physical memory of a computer can be shared by several running applications. Each applications occupies an active working set consist of an accessed area (AA) and a prefetching area (PA). The PLC takes an additional area for each cache working set. This area is reserved (RA) for the pages which have a high latency cost of being swapped into cache. The structure of PLC is shown in figure 2.

If there is a page with a swap-in latency cost higher than average cost (12.67ms in our case), and the following page request was satisfied after an above-average latency

again. This means the latter page request is an isolated and random access. This page will be moved into the reserved cache area where is not checked by LRU or CLK algorithm. Figure 2 shows the structure of PLC shema.

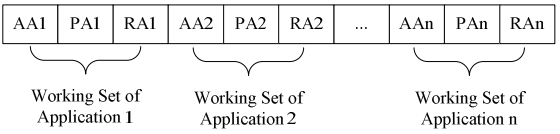


Fig. 2. Structure of PLC

Suppose there are two applications running in ram. Application 1 is with a reference string of A, B, C, D, and 1,2,3,4. For example, the A, B, C and D are neighboring pages in an executable file located in a user directory, while the 1, 2, 3 and 4 are pages of dynamic library files in a system file fold. Application 2 is with a similar page request combination of α, β, γ , δ and 1, 2, 3, 4. Hence A, B, C, D andα, β, γ , δ are sequential page accesses separately, and 1, 2, 3, 4 are random page accesses.

The system sends out a reference string of A, B, C, D, 1, 2, 3, 4, A, B, C, D, α, β, γ , δ, 1, 2, 3, 4. Figure 3 shows how PLC algorithm works.

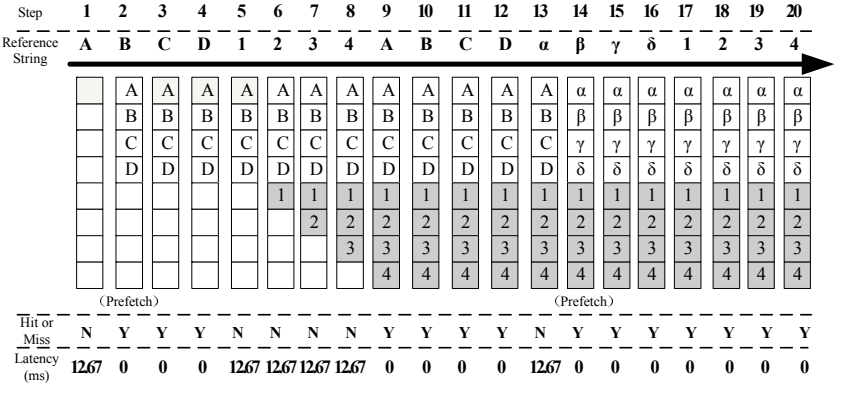


Fig. 3. Example of PLC

Step1~4: Because A, B, C and D are sequential pages, only the access of page A leads a cache miss. Then, the prefetch mechanism works. All four pages will be read into cache at one time. The time consumption is 12.67ms.

Step 5~8: Each access for page 1, 2, 3 and 4 will result a cache miss. The cost of time is 50.68ms. All these four pages are to be reserved because of their high cost. The shadowed grids mean to be reserved.

Step 9~12: Another access for page A, B, C and D. Because of cache hit, the cost of time is approximate 0.

Step 13~16: Visiting pageα, β, γ and δ will result cache swapped out some old pages to get space. The LRU/CLK algorithm will swap out the page 1,2,3 and 4,because these pages are less recently used than page A,B,C and D. Instead, the PLC will reserve page 1,2,3 and 4 for their high latency cost.

Step 17~20: Another access for page 1, 2, 3 and 4. Due to PLC algorithm, they can be visit in cache. The time consumption is about 0. On the contrary, the LRU/CLK has already swapped them out. Each access will result in a cache miss. These 4 misses will cost 50.68ms totally.

Figure 2 shows PLC encounters 6 misses that cost 76.02, while LRU/CLK will takes 126.7ms.

4 The Implementation of PLC Schema

The PLC has been deployed in a Linux2.6.32 system. The cache and buffer page frame is managed by the function of refill_inactive_zone() in the source file of mm/vmscan.c. The function was modified by the following algorithm of this schema:

```
/*Function will be invoked when a page frame is
requested*/
void* PLC(page_frame_request p){
cost=0;
while(p.next!=null){
if p is in PLC_stack{
      p.use_bit = 1;
      return p.pBPF;}/*return the page in cache*/
else if(p.use_bit==0&&p.res_bit==0){
      cost = Jiffiles; /*Global variable of Linux,
records the clock tick of system*/
      replacestack(p);/*read page from disk*/
      cost = Jiffiles - cost;/*the Latency Cost*/
      p.usebit = 1;
      if (cost>threshold)/*threshold equals average disk
latency */
           p.res_bit = 1;/*reserved page*/
      else
          p.res_bit = 0;
      return p.pBPF;}
p=p.next;}
}
```

Primarily, the PLC schema was tested in desktop applications. The hardware of the test machine is with a CPU of Intel P4 3.2G, 2GB DDR2-667 RAM and a 320GB HDD of ST3320620AS. The diff is a Linux tool to compare text files, and iostat is a tool to monitoring the I/O stat of the Linux system. We use the iostat to count activities and throughput of the system while running the command of diff. Table 4 shows the differences between original LRU/CLK and PLC schema. The actual quantity of I/O activities has been reduced because of higher latency cost operations has been avoided.

Table 4. Test Results From iostat

	rrqm/s	wrqm/s	rsec/s	avgqu-sz	await
PLC	96	23	95123	4.35	12.5ms
LRU/CLK	102	25	96321	4.94	13.78ms
Improvement	5.88%	8.00%	1.24%	11.94%	9.29%

Then, a Linux file system benchmark tool-PostMark [6] developed by NetApp was adopted to test the random access performance. The PostMark initially set up a file pool, then it create, delete, open, close, open, read and write file(s) randomly.

The configurations of the test are as followings:

```
set size 1000 50000 #amount of files varies from 1000 to 5000
set location /usr/cfs      #directory of file pool
set transactions 5000 #file transaction amount
set number 5    #max concurrent transactions
run result.txt    #file for test results
```

The results of the test illustrated that the whole test lasted 9 seconds, 7455 files were created successively, 2553 read and 2447 append(write) transactions occurred. Figure 4 shows PLC improves the read speed and write speed by 4.11% and 4.91% respectively.

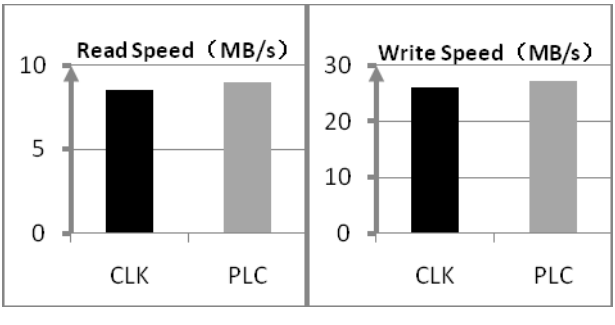


Fig. 4. Test Results from PostMark

5 Conclusions

Spatial locality manifests periodicity if we expand the time scale from several instruction cycles to several seconds. Since pages will be accessed repeatedly, reproducibility

of latency cost need to be considered carefully in cache and buffer management algorithms design and implementation. The PLC schema intends to minimize the high latency cost operations of the storage systems. Our experiments illustrates that under the given circumstance, this schema is practical and effective.

References

1. Denning, P.J.: The Locality Principle. Communication of the ACM 48(7), 19–23 (2005)
2. Seagate: Barracuda 7200.12. SAS Product Manual, http://www.seagate.com
3. Microsoft: DiskMon for Windows v2.01, http://technet.microsoft.com
4. Gill, B.S., Bathen, L.A.D.: Optimal Multistream Sequential Prefeching in a Shared Cache. ACM Transaction on Storage 3(3), 34–59 (2007)
5. Fide, S., Jenks, S.: Proactive Use of Shared L3 Caches to Enhance Cache Communications in Multi-Core Processors. IEEE Computer Architecture Letters 7(2), 57–60 (2008)
6. Yi, J.J., Lilja, D.J.: Simulation of Computer Architectures: Simulators, Benchmarks, Methodologies, and Recommendations. IEEE Transactions on Computers 55(3), 268–280 (2006)

Design of Simple Home Service Robot Based on Infineon Microcontroller

Jie Jiang and Yanqiang Wang

School of Information Engineering,
Inner Mongolia University of Science and Technology,
Bao Tou, Inner Mongolia, 014010, China

Abstract. Aiming at the status that commercial robot structure is complex, expensive and unable to popularize, this paper designes a simple home service robot system based on Infineon 16-bit single chip microcontroller. It introducs the mechanical structure, and mainly recommends hardware and software to elaborate the design and realization of the control system. The robot adopts double controllers, with two model cars as its feet, each of which communicates with each other through wireless communication module to walk simultaneously. It uses Camera to automatically identify target objects, and emploies ultrasonic sensor to measure the distance between robot and target objects. It adopts E-compass to fix position, uses phonetic module to realize human-computer interaction and grabs target objects with arms. Experimental results show that the robot realizes the expected functions, with its features of simple structure, low cost, flexibility of control and certain application value.

Keywords: Infineon; Microcontroller; Home service robot; Electronic compass; Wireless communication.

1 Introduction

Along with the rapid development of intelligent robot technology, the application of intelligent robot is expanding constantly. It is not only applied in industrial and agricultural production, but also even used in home service industry. Home service robot, which merges the core technologies of robot vision, phonetics, intelligent human-machine interaction, network technology, sensor detection, control technology and system integration together, will replace human to finish all kinds of housework, including clean sanitation, items, handling, family entertainment, patient monitors, could greatly improve the quality of people's lives. Meanwhile, China is facing more and more serious problem of aging population, so robot will have great application prospects in many families with older or disable people. However, in China, the research of humanoid robot starts later, there is still a big disparity between our country and the robot powers such as Japan and Europe countries. Although current high-end

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 383–391.
springerlink.com

home service robots abroad have a great number of functions, their structures are most complex, and the price is high and are unable to popularize temporarily[1].

Aiming at the characteristics that home service robot should have, this paper designs a home service robot system with the advantages of simple and practical mechanical structure, low cost and flexibility of control. It uses two cost-effective Infineon site microcontroller as the core controllers and aims at system reliability and stability, real-time ability and optimization of overall performance.

2 Summary of the Mechanical Structure and Function of Robot

2.1 The Mechanical Structure of Robot

The robot is consisted of five parts of head, arms, waist, legs and feet. It has 17 degrees of freedom. Head has two degrees of freedom, which is used to realize pitching and horizontal rotation. It is equipped with a camera, which is used to realize target recognition and allocation and to provide target information to the master controller;Ultrasonic sensorh is used to detect the distance between robot and target, and then report it back to the master controller to control the speed. Each arm has five degrees of freedom and can extend and bend to grab the target accurately. Waist, which is equipped with a E-compass to detect the pitch and azimuth of robot, has one degree of freedom to realize horizontal rotation of the upper body of robot. Knee has two degrees of freedom to realize crouch. Feet are two Freescale intelligent vehicles, each of which is consisted of a car body, a steering servo and a DC motor. Each of the feet is equipped with a microcontroller, which is used to realize the information collection and processing and control algorithm; Wireless communication module is used to exchange information between two controllers. The mechanical structure of robot is shown in Fig. 1.

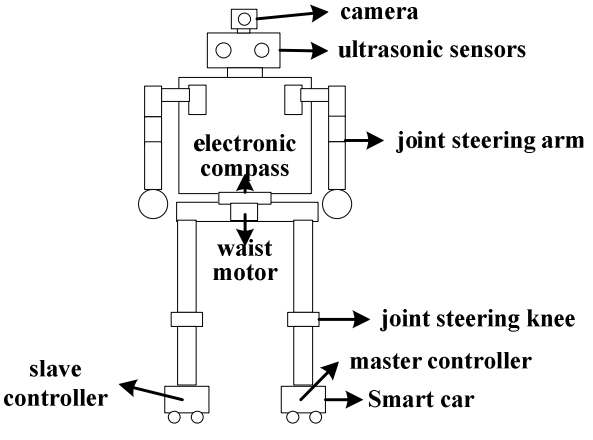


Fig. 1. Robot mechanical structure

2.2 Summary of the Function of Robot

The robot can realize human-computer interaction, target tracking, path recognition and object grabbing. The specific work process is that, turn on the power switch firstly, then robot will wait to receive a certain voice command. Once start command is received, the motor of waist will rotate and the camera of head will search the certain target. After got the certain target, the angle and location from E-compass will be obtained. Then the motor of Waist will rotate back according to the angle. After adjusting its body, the steering servo on the feet will act according to the angle from the camera and the E-compass and the DC motor will act to track the certain target. Meanwhile, the distance from ultrasonic sensor is got to control the speed of robot. When robot is close to the target, it will stop moving and the arm will act to grab the target. Then robot will return back to the starting point and put down the target.

3 The Overall Design of Control System

The control system is a hierarchy control system based on two controllers, one of which is master controller and the other is slave one. In this hierarchy control architecture, the controllers have a clear division of work and it exchanges information between the two controllers through wireless communication, which can relieve the master controller's pressure generated by the frequent interruption of multi-interrupt sources and can enhance the robot's response ability and stability of operation.

The control system of home service robot adopts modular design. The master control system includes master-control module, perception module, display module, wireless communication module, power conversion module and driver module. The slave control system is consisted of distance measuring module, driver module, power module. The whole block diagram of the system is shown in Fig 2.

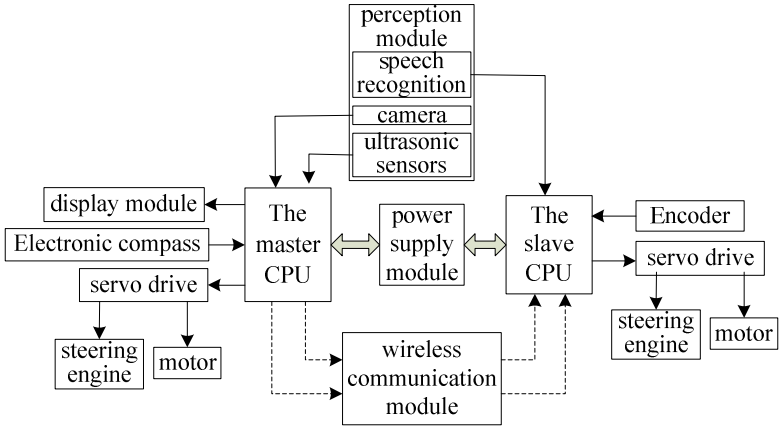


Fig. 2. Control system structure

4 The Hardware of Control System Design

The hardware of control system also adopts modular design, and the detailed design modules are shown as follows:

4.1 The CPU Control Module

According to the fact that the control system should have good expansibility and rich resources, this system adopts the latest Infineon XE164FM - 72F80L 16-bit microcontroller which is produced by Germany Siemens Infineon Technologies. The microcontroller has advantages of small volume,high performance-price ratio, high integration, easy system extension and strong ability of data and can carry out digital signal processing. It contains 16 A/D converters with 10-bits precision, 16 channel universal capture/comparison, can produce 16 unit PWM waves, and possess strong ability to process the interrupts generated by 96 interrupt sources and 16 interrupt priorities[2].

4.2 Perception Module

It mainly includes the visual system, the hardware circuit of ultrasonic sensor, phonetic recognition. The visual system uses CCD camera, whose pixel is 320×240. The CCD camera, whose output signal is PAL, outputs 50 fields of pictures per second(divided into strange and dipole field). In order to collect camera video signal effectively, it is needed to capture sync signal and vertical sync signal, otherwise microcontroller will not receive the video signal correctly. The video sync separator chip, LM1881,which is produced by America's national semiconductor company, is used to separate the video signal. Fig. 3 is the circuit principle diagram of signal separation of video.

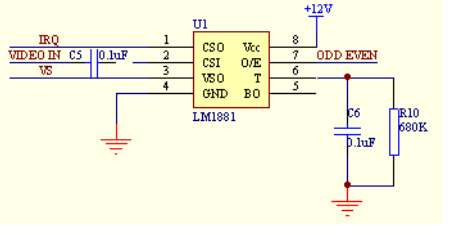


Fig. 3. LM1881 video signal circuit separation

Phonetic recognition module mainly completes certain people's voice recognition and speech synthesis output. The phonetic recognition chips, AP7003, is a new type, low cost, phonetic recognition, application-specific integrated circuit with microphone amplifier, an A/D converter, phonetic processor and the I/O controller. It can identify 12 groups of words after pretreatment, each of which spends 1.5 seconds, and is highly programmable and easy to use. Voice signal is input to AP7003 by microphones and then to the robot.

4.3 Power Conversion Module

The system uses four 7.2 v, 2000mAh nimh batteries. Because each part requires different voltage, different power switching circuit are designed to provide the voltage of +12v, +5v, 6v and +3.3v separately.

4.4 Motor Driver Module

It uses PWM to adjust the motor speed. The motor drive chip is MC33886. It is a special motor drive chip manufactured by Motorola company, and the maximum output current is 5A. Two parallel chips could improve the drive ability, and the biggest drive current can reach 10A. The circuit diagram is shown in Fig 4.

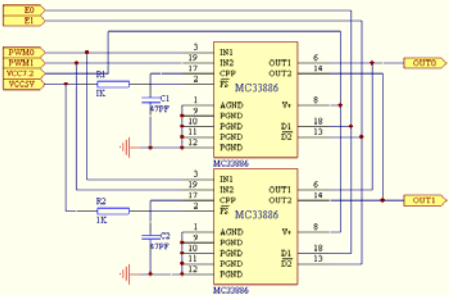


Fig. 4. Motor driving circuit

4.5 Wireless Communication Module

NRF24L01 which is a single chip wireless transceiver integrated chip is adopted. It employs the GFSK modulation with strong anti-jamming capability, with the advantage of stable and reliable working frequency, and its peripheral devices are less, power consumption is low and it is suitable for handheld and portable product design. It is mainly used in robot communication between master and slave controller. In order to make the two feet coordinated and synchronized, the master controller uses wireless module to exchange information, which includes the visual information measured by camera and distance information measured by ultrasonic sensor, to control the direction and speed of the two feet.

4.6 Electronic Compass

It is put in the waist and mainly used for measuring the heading angle and pitch to navigate and keep balance. It is mainly consist of control unit (MSP430), three axis acceleration sensors (SCA3000), a temperature sensor (TMP100) and magnetic sensors (PNI11096)[3]. The temperature sensor is responsible to compensate for measurements of acceleration sensor and magnetic sensor. PNI is a drive chip, which converts the weak current measured by PNI11096 into the digital quantity that can be identified by MSP430.

5 The Software Design of Control System

The whole structure of control system is shown in Fig. 5. The software of control system includes two main parts: software of master control system and slave control system.

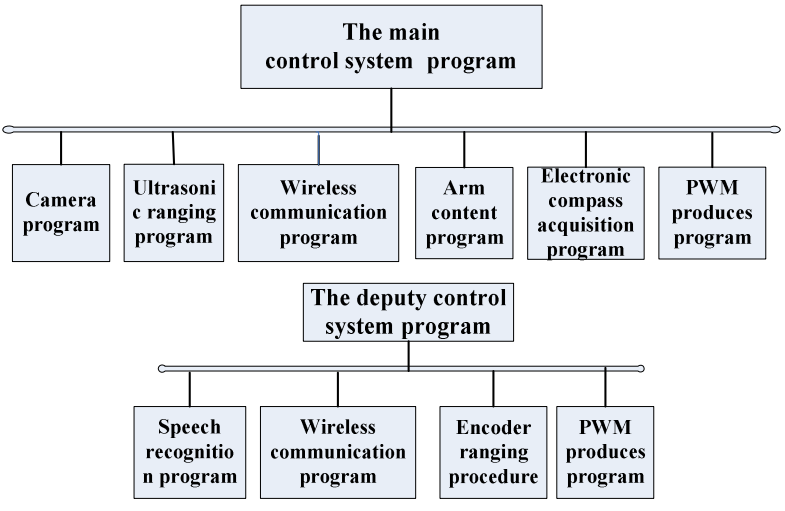


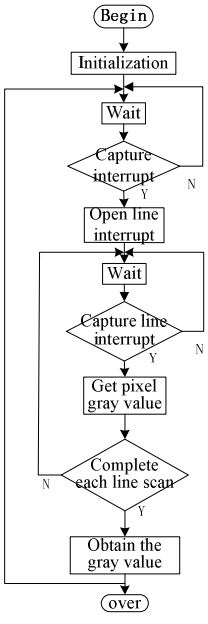
Fig. 5. Overall structure of control system software

The software of master control system is consist of main program, subprograms of camera information acquisition, ultrasonic distance measuring, wireless communication, grabbing object of arm, E-compass information acquisition and PWM generation. The main program includes system initialization, control flow and realization of typical control algorithm. The subprogram of camera information acquisition accomplishes the acquisition of camera information. It gets the gray level of every frame picture through capturing the link interruption and field interruption and then gets the object's location in the camera's world coordinate system. The software flow pattern is shown in Fig 6.

The subprogram of ultrasonic distance measuring completes the distance measuring between robot and the target. The controller firstly generates a trigger pulse of 10us to trigger the ultrasonic module to transmit a series of ultrasonic. Meanwhile the timer and the capture interruption is opened. The ultrasonic will return back when it comes across the obstacle and will be received by the receiver of ultrasonic module[4][5]. Then the receiver will generate a falling edge pulse. After the controller captures this falling edge, it will close timer and t is got. The distance will be calculated according to the formula:

$$S = 340 * t / 2 \tag{1}$$

The software flow pattern is shown in Fig 7.



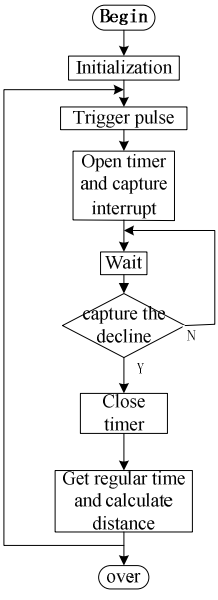


Fig. 6. Camera acquisition flowchart **Fig. 7.** Ultrasonic ranging flowchart

The subprogram of wireless communication completes the data exchange between master controller and slave controller. It includes the initialization of NRF24L01 chip, configuration of address and data length and transceiver sub-function. The subprogram of serial communication realizes the data delivering between the controller and E-compass. The subprogram of PWM generation generates PWM of different period by using timer to drive the servos of arm and motors of head and feet.

The subprogram of camera information acquisition, subprogram of ultrasonic distance measuring and the feet of robot realize target tracking together. Target tracking means that robot moves forward or backward and turn left or right by following a certain object. In the view of camera, a plane coordinate system is built. In the coordinate system, the original point is on the left bottom and both X-axis and Y-axis are divide into 100. The target is black and the controller store the gray level of object into a two-dimensional array. The data in the two-dimensional array is detected to get the coordinate values of the edge of target. The coordinate values of upper and lower edge are added together and then is divided by 2. The coordinate values of left and right edge are also added together and then is divided by 2. And the coordinate values (P_1, P_2) of the center of target in the camera world coordinate system are got. To compare (P_1, P_2) with the center $(50, 50)$ of camera world coordinate system, if P_2 is bigger than 50, the camera will deflect downward; otherwise, the camera will deflect upward. This can ensure that the camera can always capture the target. If P_1 is bigger

than 50, that the target is on the left of robot and robot will turn left; otherwise, it will turn right. This ensure that robot can always track the target[6].

The subprogram of grabbing object of arm controls the arm to grab the target object according to the position got by camera and ultrasonic module.

The software of slave control system includes main program, wireless communication subprogram, voice recognition subprogram and PWM generation subprogram. The master control system and the slave control system coordinate together to realize the basic function of robot.

The whole software flow pattern of master-slave control system is shown in Fig 8.

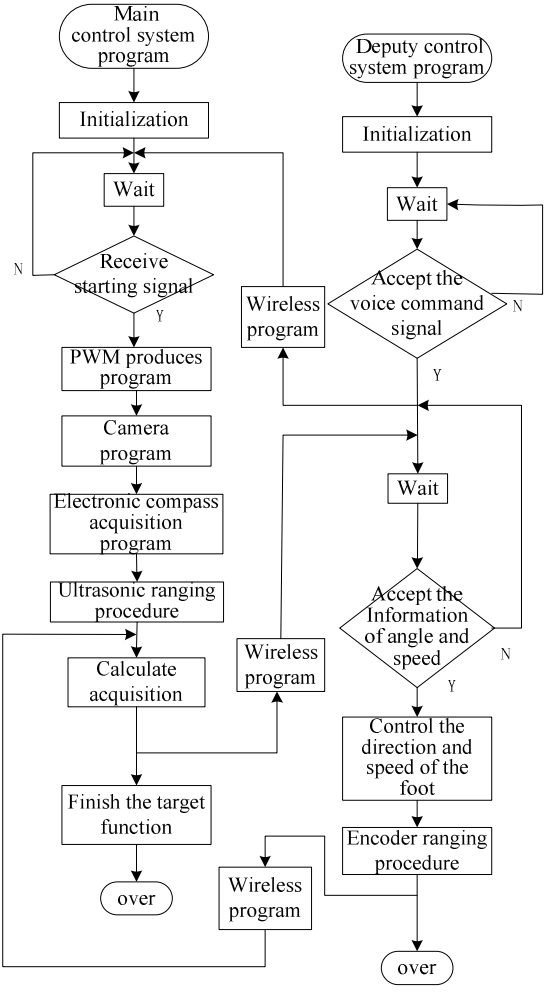


Fig. 8. Overall software flowchart

6 Conclusions

By combining with the design function of robot, this paper establishes a hardware platform that uses two micro-controllers as the core. The basic function of the home service robot is realized by the cooperation of every module. This robot has the advantage of low cost, modularization, easy expansibility, easy portability and high reliability. More functions can be realized through external sensors and flexible software programming to really satisfy the design requirements of home service robot.

References

1. Tian, G., Duan, F., Arai, T.: Modular Design for Home Service Robot System Based on Hierachical Colored Petri Nets. In: Proceeding of the 9th International Conference on Intelligent Autonomous System IAS-9, Tokyo, Japan, Marth 7-9, pp. 542–549. IOS Press, Amsterdam (2006)
2. Infineon Technologies AG.XE164xM 16-Bit Single-Chip Real Time Single Controller User's Manual Version 2.0. Germany: Infineon Technologies AG (2009)
3. Chen, H., Li, W., Li, B.: Design of small mobile robot based on multiprpocessor. Micro Computer Information(in chinese) 24(10), 214–216 (2008)
4. Ambastha, M., Busquets, D.: Evolving a multi-agent system for landmark-based robot navigation. International Journal of Intelligent System 20(5), 523–539 (2005)
5. Maja, J.M.: Real-time obstacle avoidance algorithm for visual navigation. In: The 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems, vol. 2, pp. 925–930 (October/November 2000)
6. Yang, G.s.: Research on corridor navigation based on sensor fusion of infrared and ultrasonic for mobile robot. Journal of Electronic Messurement and Instrument 2, 1749–1754 (2004)

Computer Simulation of the Leontief Model in ADA

Xiumei Wu

School of Computer and Information, Shanghai Second Polytechnic University,
200000 Shanghai, China
xmwu88@163.com

Abstract. ADA is a high-performance language for scientific and technical computing. Different program packages can be used in corresponding areas, such as math, control theory, economics. We describe the design of the Leontief input-output model toolbox which can simulate the trajectories of input and output of economic systems and automatically control the run of economic systems. By means of mathematic methods, the Leontief input-output systems are directly treated without converted to general systems. A sufficient condition under which the Leontief input-output models are stable is investigated. Based on this, the control algorithm is derived. Finally, the code of the input-output control program is provided.

Keywords: ADA, Design, Input-output Model.

1 Introduction

ADA is a high-performance language for scientific and technical computing. At first, the scalar noninteractive languages, such as C or Fortran, are used to compute the scientific and technical problems. But these languages can hardly deal with the matrix and vector equations. ADA's basic data element is an array which does not require dimensioning. So ADA can easily solve many computing problems, especially those with matrix and vector formulations. Matrix is a mathematic tool which is utilized in many areas, such as economics [1].

In the past years much attention has been paid for the simulation and control of economics. Many economic models were founded to simulate the run of economic systems. Most of those models were matrix and vector equations. Then economists tried to solve these equations and hoped to control these models. At first, economists completed the study of the static economic systems' matrix equations. However, most economic problems are dynamic. Then, the economists have to study the dynamic economic systems' matrix equations. They often investigate the current state of an economic system and ask how various policies can be used to move the system from its present status to a future more desirable state when they deal with those dynamic systems [2]. A number of fundamental notions and methods based on the theory of economical cybernetics have been extended. When a more detailed description of the production side of an economic organization is desired with the development of macroeconomics, this leads to a so-called input-output analysis (i.e. input-output

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 393–399.
springerlink.com

economics). Harvard Professor Leontief, who opened the door to the input-output economics, put forward the Leontief input-output model in 1949 [3]. The classic Leontief input-output models are static systems with linear matrix formulations. Then in the region of input-output economics, many models were established to describe the real economics [4]. The general linear models of input-output economics were well investigated. However, many systems are singular linear models, which are much harder to be solved than general linear models. In the area of cybernetics there is a rapid progress in the control of singular systems because such systems can describe many real systems such as economic systems [5]. Singular models in economics are generally converted into general linear models by means of the selection of state vector, control vector and output vector. We hope to research the structure of the singular Leontief input-output model's matrix formulation and find out a control algorithm of this kind of model. Based on this algorithm, the simulation and automatic control of the Leontief input-output model will be completed [6]. Furthermore, the classic Leontief Input-output Model does not consider the economic architecture's change. The coefficient matrix of classic Leontief Input-output Model, which reflects the quantity of investment in each sector, is static and can not correctly describe the variety of economic system when the architecture of economy has changed. So, the form of coefficient matrices of Input-output Model can be improved to describe the variety of economic system's architecture.

Different program packages can be used in corresponding areas, such as math, control theory, economics. A new input-output model toolbox need be designed for the research of the Leontief input-output model. In this paper, ADA is used to design such toolbox.

2 Simulation of Leontief Model

Prior to the design of the input-output model package, we must confirm the function of this toolbox. Firstly, this toolbox should be able to solve the equations of the Leontief input-output model and simulate the run of the Leontief input-output model. Secondly, the toolbox could automatically search a stable solution of the Leontief input-output model when we make an economic plan.

Now we will firstly investigate the equations of the Leontief input-output model and simulate the trajectories of input and output. The classic economic dynamic input-output model is described by

$$x(k) = Ax(k) + B[x(k+1) - x(k)] + \bar{Y}(k). \tag{1}$$

The vector $x(k) = [x_1(\mathrm{k}) \ldots x_n(k)]^T \in \mathbb{R}^n$ is the total output vector and $x_i(k)$ is the total output from sector i. $\bar{Y}(k) = [y_1(\mathrm{k}) \ldots y_n(k)]^T$ is the final net product vector and $y_i(k)$ denotes the final net product of sector i. The matrix $A = [a_{ij}]$ is the direct consumption coefficient matrix, $B = [b_{ij}]$ is the capital coefficient matrix. In fact, Matrix A and B do not maintain unchanged in the real economic life. Generally, the average profit margins vary from sector to sector: a higher one for the sector whose demand exceeds supply and a relatively lower one for the sector oversupplied.

Consequently, an enterprise with a low profit margin would invest capital in a high-profit sector and thus the economic structure is changed, which means the change of the coefficient of matrix A and B. Manifested in the input-output model, matrix A and B keep changing every year and the value of matrix A and B is determined by supply and demand relations. The allocation of investment is described by capital coefficient matrix B. If the investment in sector j is raised, the corresponding coefficient of the jth row of matrix B will be increased; if the investment in sector j is reduced, the corresponding column coefficient of matrix B will be decreased. Assuming that Bc(k) is an elementary matrix, when the investment in sector i is raised, element on Row i / Column i is constant Pi which is greater than 1, and when the investment in sector i is reduced, element on Row i / Column i is constant Pi which is less than 1. Then the change of capital coefficient matrix could be described via the equations:

$$B(k) = B \cdot Bc(k) \text{ and } Pi = 1 + (P(i) - Paverage). \tag{2}$$

Matrix B is the initial capital coefficient matrix; Constant P(i) is the profit margin of sector i; Paverage denotes the average profit margin. Then the above dynamic input-output model can be rewritten as

$$B(k)x(k+1) = (I - A + B(k))x(k) - \bar{Y}(k) \tag{3}$$

where $rankB(k) = r < n$.

The equation (3) is a difference equation. Matrices A and B can be determined by economic policy makers. Then the next question to be solved is the value of $\bar{Y}(k)$. In fact, $\bar{Y}(k)$ can be considered as the control vector of economic discrete-time singular dynamic input-output model because we can affect the quantity of final net product by controlling the scale of investment. Then $\bar{Y}(k)$ can be treated as control vector which can be specified by governments or companies. Thus the equation (3) can be solved.

Remark 1. In this paper, $\mathbb{R}^n$ is the n-dimensional Euclidean space; $\mathbb{R}^{m\times n}$ denotes the $m\times n$ real matrices space; I is the $n\times n$ identity matrix; A^T denotes the matrix transposition; "*" is used as the term that is introduced by symmetry; when matrices X, Y are symmetric, $X < Y$ means that matrix $Y - X$ is positive-definite.

Remark 2. By using of the function "dsolve()", we can get the solution of equation (3). Of course, we can also design the new function to solve the equation (3). The following code is a one of such a function.

```
procedure read_abc is
   type big is new real;
   d,y11,y21:real;
   b11,b12,b21,b22,a11,a12,a21,a22:big;
   i: integer
   for i in 1..100 loop
        b11*dx+b12*dy=a11*x(1)+a12*x(2)-y11;
        b21*dx+b22*dy=a21*x(1)+a22*x(2)-y21;
   end loop;
end
```

We can get the solution of the equation (3) by chase method and recursive invocation of the above procedure.

In this section, we describe the design of simulation of the Leontief input-output model. When we can know the values of matrices A, B and $\overline{Y}(k)$, the solution of Leontief input-output model can be easily derived. Figure 1 shows one example of Leontief input-output model.

3 Stable Solution of Model

In real society, there are many problems concerning the control of the economic systems' scale. One important control problem is about the quantity of final net product. Generally, if the final net product in the first year is d_0, it is hoped to increase to a higher level several years later. However, the quantity of the final net product could not increase forever because resources on the earth are scarce. Scarcity means that society has limited resources and can not allocate all the resources to one sector. So it is hoped that the quantity of the final net product remains at a quite stable level. So, we need to research the stability of the input-output models. In economics, capital coefficient matrix B is not always invertible, because the product of some sectors can not be treated as capital product and applied to invest. Then the system (3) is a singular input-output model. The singular systems are hard to handle. Fortunately, the stability of singular system (3) is equal to the stability of the following system:

$$B(k)x(k+1) = (I - A + B(k))x(k) \tag{4}$$

where $B(k) = B \cdot Bc(k)$. So the research of the stable solution of singular systems (3) can be converted into the research of system (4). To investigate the stability of system (4), the following definition of stability is needed:

Definition 1: The discrete-time singular equation:

$$B \cdot Bc(k)x(k+1) = (I - A + B \cdot Bc(k))x(k) \tag{5}$$

(1) System (5) will be called regular if $\det(sB_k - B_k - I + A)$ is not identically zero where $B_k = B \cdot Bc(k)$.

(2) System (5) is causal if $\deg(\det(sB_k - B_k - I + A)) = rankB_k$ where $B_k = B \cdot Bc(k)$.

(3) System (5) is asymptotically stable if any root of $\det(sB_k - B_k - I + A) = 0$ lies in the interior of the unit disk with center at the origin where $B_k = B \cdot Bc(k)$.

(4) System (5) is called to be stable if it is regular, causal and asymptotically stable.

In general, it can be assumed that the quantity of final net product rely on the quantity of total output product. In other words, we assume that the quantity of final net product can be described by $\overline{Y}(k) = gx(k)$. Then the system (5) turns into the form of

$$B \cdot Bc(k)x(k+1) = (I - A' + B \cdot Bc(k))x(k) \tag{6}$$

where $A' = A + g$. Next, we need to research the stability of system (6) and the value of g. This problem can be solved with the following algorithm.

Theorem 1: The singular dynamic input-output system (6) is stable if there exist matrices $P > 0$, Q and S_k with proper dimensions such that the relations

$$\Sigma + \Sigma^T < 0 \tag{7}$$

and $$B_k^T S_k = 0 \tag{8}$$

can establish for each k where

$$B_k = B \cdot Bc(k),$$

$$\Sigma = \frac{1}{2}P + \frac{1}{2}A'^T PA' - PA' + PB_k - A'^T PB_k + QS_k{}^T - QS_k{}^T A' + QS_k{}^T B_k . \tag{9}$$

Proof of Theorem 1: For system (6), there exist two nonsingular matrices M_k and N_k such that

$$B_k = M_k \begin{bmatrix} I_1 & 0 \\ 0 & 0 \end{bmatrix} N_k \tag{10}$$

where $I_1 \in \mathbb{R}^{rankB_k \times rankB_k}$ is an identity matrix.

Then, pre-multiplying $(I - A' + B_k)$ by $M_k{}^{-1}$ and post-multiplying $(I - A' + B_k)$ by $N_k{}^{-1}$ one can get a new matrix:

$$M_k{}^{-1}(I - A' + B_k)N_k{}^{-1} = \begin{bmatrix} T_{1k} & T_{2k} \\ T_{3k} & T_{4k} \end{bmatrix}. \tag{11}$$

It can be written as: $$A' = I - M_k \begin{bmatrix} T_{1k} - I_1 & T_{2k} \\ T_{3k} & T_{4k} \end{bmatrix} N_k .$$

Obviously, S can be:

$$S_k = M_k{}^{-T} \begin{bmatrix} 0 \\ I_3 \end{bmatrix}. \tag{12}$$

Then we can write

$$P = M_k{}^{-T} \begin{bmatrix} P_{1k} & P_{2k} \\ P_{2k}^T & P_{3k} \end{bmatrix} M_k{}^{-1},\ Q = N_k{}^T \begin{bmatrix} Q_{1k} \\ Q_{2k} \end{bmatrix}. \tag{13}$$

where the block is compatible with that of (10).

Together with (9)-(13), (7) is equivalent to:

$$\begin{aligned} &\Sigma + \Sigma^T = P - PA' - A'^T P + A'^T PA' + PB_k - A'^T PB_k + B_k{}^T P \\ &-B_k{}^T PA' + QS^T - QS^T A' + QS^T B_k + SQ^T - A'^T SQ^T + B_k{}^T SQ^T \\ &= N_k{}^T \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} N_k < 0 \end{aligned} \tag{14}$$

where

$$Z_{22} = T_{2k}^T P_{1k} T_{2k} + T_{4k}^T P_{2k} T_{2k} + T_{2k}^T P_{2k} T_{4k} + T_{4k}^T P_{3k} T_{4k} + Q_{2k} T_{4k} + T_{4k}^T Q_{2k}^T .$$

From formula (14), we can get $Z_{22} < 0$. Utilizing the reduction to absurdity, we can easily prove that the matrix T_4 is nonsingular. Now, define: $T_k = I - A' + B_k$.

Then, for system (6) it is easy to see:

$$\det(sB_k - T_k) = \det(M_k \begin{bmatrix} sI_1 - T_{1k} & -T_{2k} \\ -T_{3k} & -T_{4k} \end{bmatrix} N_k) = \det(M_k N_k) \det(\begin{bmatrix} sI_1 - T_{1k} & -T_{2k} \\ -T_{3k} & -T_{4k} \end{bmatrix}) .$$

Obviously, there exist two nonsingular matrices L and R such that

$$L \begin{bmatrix} sI_1 - T_{1k} & -T_{2k} \\ -T_{3k} & -T_{4k} \end{bmatrix} R = \begin{bmatrix} sI_1 - T_{1k} & -T_{2k}^{'} \\ -T_{3k}^{'} & -I_2 \end{bmatrix} .$$

So, it can be easily seen that $\det(\begin{bmatrix} sI_1 - T_{1k} & -T_{2k}^{'} \\ -T_{3k}^{'} & -I_2 \end{bmatrix})$ is not identically zero and the degree is $rankB_k$. For M, N, L, and R are nonsingular, $\det(sB_k - T_k)$ is not identically zero and $\deg(\det(sB_k - T_k)) = rankB_k$. So this discrete-time singular input-output system (6) is regular and causal.

Next, we will prove that system (6) is asymptotically stable. Since system (6) is regular and causal as proved above, according to [8] can find two nonsingular matrices $\tilde{M}$ and $\tilde{N}$ such that

$$B_k = \tilde{M}_k \begin{bmatrix} I_1 & 0 \\ 0 & 0 \end{bmatrix} \tilde{N}_k , \; I - A' + B_k = \tilde{M}_k \begin{bmatrix} \tilde{T}_k & 0 \\ 0 & I_2 \end{bmatrix} \tilde{N}_k . \tag{15}$$

Then, S can be chosen as

$$S = \tilde{M}_k^{-T} \begin{bmatrix} 0 \\ I_3 \end{bmatrix} . \tag{16}$$

Define

$$P = \tilde{M}_k^{-T} \begin{bmatrix} \tilde{P}_{1k} & \tilde{P}_{2k} \\ \tilde{P}_{2k}^T & \tilde{P}_{3k} \end{bmatrix} \tilde{M}_k^{-1} \quad Q = \tilde{N}_k^T \begin{bmatrix} Q_{1k} \\ Q_{2k} \end{bmatrix} . \tag{17}$$

Substituting (15)-(17) into (7), we can acquire

$$\tilde{N}_k^T \begin{bmatrix} \tilde{T}_k \tilde{P}_{1k} \tilde{T}_k - \tilde{P}_{1k} & \tilde{T}_k \tilde{P}_{2k} + \tilde{Q}_{1k} \\ \tilde{P}_{2k}^T \tilde{T}_k + \tilde{Q}_{1k}^T & \tilde{P}_{3k} + \tilde{Q}_{2k} + \tilde{Q}_{2k}^T \end{bmatrix} \tilde{N}_k < 0 .$$

Using the Schur complement formula, we have $\tilde{T}_k \tilde{P}_{1k} \tilde{T}_k - \tilde{P}_{1k} < 0$. From [9, 19], we know that every root of $\det(sB_k - T_k) = 0$ lies in the interior of the unit disk with center at the origin. So system (6) is asymptotically stable. Thus system (6) is regular,

causal and asymptotically stable which means system (6) is stable. This is the end of the proof.

Remark 3. Theorem 1 offers a sufficient condition for the discrete-time singular dynamic input-output system to be stable. Based on this, we can get the stable solution of the Leontief input-output model (3). The following algorithm is such a result.

4 Conclusion

The design of the Leontief input-output toolbox in ADA has been provided. This toolbox provides the simulation function of the Leontief input-output model and can find out the stable solution. One important aspect of our design is the extension of classic Leontief Input-output Model. The capital coefficient matrix is designed as function of time k. A new algorithm is designed to get the stable solution of singular input-output model. Finally the Leontief input-output model toolbox is completed.

References

1. Brandimarte, P.: Numerical methods in finance and economics. Wiley Interscience, Hoboken (2006)
2. Kendrick, D.: Stochastic Control for Economic Models. McGraw-Hill, New York (1981)
3. Leontief, W.: Input-Output Economics. Oxford University Press, New York (1966)
4. Dai, L.: Singular Control Systems. Springer, Berlin (1989)
5. Xu, S., Lam, J.: Robust stability for uncertain discrete singular systems with state delay. Asian J. Control 5, 399–405 (2003)
6. Yue, D., Han, Q.: Delay-dependent exponential stability of stochastic systems with time-varying delay, nonlinearity, and markovian switching. IEEE Trans. Automat. Control 50, 217–222 (2005)
7. Wang, Z., Qiao, H., Burnham, K.J.: On stabilization of bilinear uncertain time-delay stochastic systems with markovian jumping parameters. IEEE Trans. Automat. Control 47, 640–646 (2002)
8. Xu, S., Lam, J.: Robust stability and stabilization of discrete singular systems: an equivalent characterization. Asian J. Control 5, 399–405 (2003)
9. Wu, Z., Zhou, W.: Delay-dependent robust H infinite control for uncertain singular time-delay systems. IET Control Theory Appl. 1(5), 1234–1241 (2007)

Direct Radial Suspension Force Control Algorithm of Bearingless Synchronous Reluctance Motor

Weichao Hu, Yanchao Li, Huangqiu Zhu, and Zebin Yang

School of Electrical and Information Engineering, Jiangsu University,
212013 Zhenjiang, China
lyghwc@126.com

Abstract. The bearingless synchronous reluctance motor (BSynRM) is a multivariable, nonlinear and strong-coupled system. To solve the difficult problem of precise decoupling in electromagnet torque and radial suspension force control, the direct suspension force control (DSFC) method is presented with the reference of direct torque control based on space vector pulse width modulation (SVM), and the direct suspension force control (DSFC) algorithm is deduced. The direct suspension force control (DSFC) system, which a double closed-loop control system of the rotor radial displacement and radial suspension force is constructed to control the radial suspension force directly, is designed and simulated. The simulation results show the stable suspension of the rotor and good performance of the proposed algorithm that realizes the decoupling control of electromagnet torque and radial suspension force.

Keywords: bearingless synchronous reluctance motor (BSynRM); direct suspension force control (DSFC); space vector pulse width modulation (SVM); decoupling control.

1 Introduction

A bearingless synchronous reluctance motor (BSynRM) has many special advantages and characteristics. Compared with the bearingless permanent magnet synchronous motor, the BSynRM without rotor windings and permanent magnetic material can be used for high-speed applications including operation at high temperature. The BSynRM in comparison with bearingless induction motor is low cost and control algorithm of the BSynRM can be simplified because it is not necessary to compute the slip. Therefore, the BSynRM has a wide range of applications, for instance high-speed machine tool, turbo molecular pump, turbo compressor, flywheel energy storage etc[1-5].

Radial suspension force of the BSynRM is generated by the magnetic field interaction between the torque windings and radial suspension force windings, which determines between the torque and radial suspension force existing strong coupling[6-7]. To ensure the BSynRM stable operation, decoupling control of electromagnet torque and radial suspension force is necessary and realizes the dependent control of torque and radial suspension force. At present, radial suspension force control algorithm of the BSynRM mainly uses the rotor magnetic orientated method. However, this control

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 401–408.
springerlink.com

method exists some weaknesses as follow: (1) this algorithm is difficult to achieve the theoretical results due to the complexity of vector coordinates transformation; (2) an open-loop control of radial suspension force affects the accuracy and dynamic response performance of radial suspension force control; (3) this control method adopts current following inverter, which has high switch frequency and the utilization rate of inverter capacity is relatively low.

With the reference of direct torque control based on space vector pulse width modulation (SVM) [8-10], hence the direct suspension force control (DSFC) method is presented and the control algorithm is deduced. Moreover, the DSFC system for the BSynRM is designed and simulated using Matlab software. The simulation results have shown that the rotor can be suspended stability, torque and radial suspension forces can be controlled independently and control system has strong robustness.

2 Mathematical Model of the BSynRM

The rotor saliency effect results in the magneto-resistance torque. The torque of the bearingless synchronous reluctance motor can be written as follow.

$$T_e = \frac{3}{2} p_M \left(\psi_{s1\alpha} i_{s1\beta} - \psi_{s1\beta} i_{s1\alpha} \right) \tag{1}$$

where $i_{s1\alpha}$, $i_{s1\beta}$ are the stator α-axis and β-axis currents of the stator torque windings respectively; $\psi_{s1\alpha}$, $\psi_{s1\beta}$ are the stator α-axis and β-axis air gap flux linkages of the stator torque windings respectively; p_M is pole pairs of stator torque windings.

The mathematical model of rotor radial suspension force is built on the as assumption that: (1) Assuming the radial displacement of the rotor is small enough to air gap length δ_o; (2) Assuming the initial value of the space position angle and the initial position angle of the rotor are zero; (3)Neglecting the leakage flux of suspension force windings. The radial suspension force acting on the rotor α-axis and β-axis can be written as follows.

$$\begin{cases} F_\alpha = \int_0^{2\pi} \dfrac{lr}{2\mu_0} B^2(\varphi) \cos\varphi \mathrm{d}\varphi \\ F_\beta = \int_0^{2\pi} \dfrac{lr}{2\mu_0} B^2(\varphi) \sin\varphi \mathrm{d}\varphi \end{cases} \tag{2}$$

where l is the core stack length, r denotes the rotor radius of the salient poles, μ_0 is permeability of vacuum, φ is the space position angle, B is motor air gas flux density including torque windings and suspension force windings.

In order to simplify the radial suspension model, so the equation (2) integral divides into four integrals over the rotor salient poles[11]. The mathematic model of radial suspension force acting on the rotor can be written as follows.

$$\begin{cases} F_\alpha = k_{M1} \psi_{m1} \psi_{s2} \cos(\lambda - \mu) + k_{M2} \psi_{m1} \psi_{s2} \cos(\lambda + \mu) \\ F_\beta = k_{M1} \psi_{m1} \psi_{s2} \sin(\lambda - \mu) + k_{M2} \psi_{m1} \psi_{s2} \sin(\lambda + \mu) \end{cases} \tag{3}$$

where $k_{M1} = \frac{\pi}{12N_1N_2lr\mu_0}$, $k_{M2} = \frac{\sqrt{3}}{8N_1N_2lr\mu_0}$, ψ_{m1} is the amplitude of torque windings air-gap flux leakage, ψ_{s2} is stator flux linkage amplitude of suspension force windings, N_1, N_2 are the number of turns which are belong to the torque windings and suspension force windings respectively, λ, μ are the initial phase angle of air gas flux leakage.

3 DSFC of the BSynRM

The direct suspension force control (DSFC) method adopts double-loop control, in which the inner closed-loop is radial suspension force loop and the outer closed-loop is rotor radial displacement loop. According to the increment between given value of the radical suspension force and the feedback value of the radical suspension force, incremental flux leakage of suspension force windings is generated by the DSFC algorithm, and then uses SVM principle generates the corresponding space voltage vector to drive the voltage source inverter to achieve the suspension force windings and radial force flux leakage direct control.

3.1 Basic Principle of DSFC

According to the equation (3), the radical suspension force can be divided two parts. Fig. 1 shows the vector diagram for DSFC. λ-μ is the angle between radial suspension force vector $\boldsymbol{F}_1$and A-phase windings axis, λ+μ is the angle between radial suspension force vector $\boldsymbol{F}_2$and A-phase windings axis. So, the $k_{M1}\psi_{m1}\psi_{s2}$, $k_{M2}\psi_{m1}\psi_{s2}$ are the amplitude of radial suspension force vector $\boldsymbol{F}_1$ and $\boldsymbol{F}_2$ respectively. The equation (3) can be transformed as follows.

$$\begin{cases} F_\alpha = F_1\cos(\lambda-\mu) + F_2\cos(\lambda+\mu) \\ F_\beta = F_1\sin(\lambda-\mu) + F_2\sin(\lambda+\mu) \end{cases} \tag{4}$$

Radial suspension force depends on the amplitude and direction of suspension force windings flux leakage and torque windings air-gap flux leakage. A stable radial force of the rotor is required by the control of suspension force windings flux leakage and torque windings air-gap flux. The motor electrical time constant is much smaller than mechanical time constant, assuming the location of the rotor is stationary within a very little time period Δt, hence the small-signal model can be used to analyze the mathematic relationship about the radial suspension force, flux leakage of suspension force windings and the torque.

(1) In steady state, the BSynRM operates at rated speed or rated torque. Within time Δt, the amplitude and phase of torque windings air-gap flux linkage are constant, and the torque angle δ is constant too. So suspension force can be controlled by regulating the amplitude ψ_{s2} and phase λ of suspension force windings flux linkage derived from equation (3).

(2) In transient state, torque of the BSynRM is variable or motor speed changes. Keeping amplitude ψ_{s1} constant, hence torque control is implemented by regulating the torque angle δ shown from Fig. 1. Assuming torque increases ΔT, which will result

in the decrease of amplitude ψ_{m1} and increase of angleμ, and the rotor radial suspension force amplitude and phase angle will vary at the same time derived from equation (3). To compensate radial suspension force, it is necessary to increase the amplitude and phase of suspension force windings flux linkage. For reducing the electromagnetic torque, the situation is vice versa.

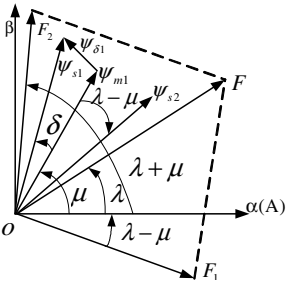


Fig. 1. Vector diagram of direct suspension force control

3.2 DSFC Algorithm of the BSynRM

Constant torque operation of the BSynRM can be regarded as a special state of the variable torque operation, so control algorithm of DSFC in variable torque operation state only is deduced considering the universality of control algorithm. Fig. 2 shows vector diagram of rotor radial suspension force and suspension force windings flux linkage.

From Fig. 2, it can be seen vector diagram of two variables during the period from t to t+1, which include rotor radial suspension force and suspension force windings flux linkage. At t moment, suspension force is $\boldsymbol{F}(t)$. Using small signal model, assuming torque increase ΔT_e during the period from t to t+1, which makes the amplitude $\psi_{m1}(t)$ reduce and phase angel μ change, derived from equation (3). Supposing the changed values are $\psi'_{m1}(t)$ and μ' respectively; suspension force vector $\boldsymbol{F_1}(t)$ becomes $\boldsymbol{F_1}'(t)$, phase angel λ-μ becomes λ-μ', suspension force vector $\boldsymbol{F_2}(t)$ becomes $\boldsymbol{F_2}'(t)$, phase angel λ+μ becomes λ+μ'.

According to geometry knowledge, the triangle $\Delta\boldsymbol{F_1}(t+1)O\boldsymbol{F_1}'(t)$,$\Delta\boldsymbol{\psi_{s2}}(t+1)O\boldsymbol{\psi_{s2}}(t)$ and $\Delta\boldsymbol{F_2}(t+1)O\boldsymbol{F_2}'(t)$ are three mutual similar triangles in Fig. 2. So, the angle between vector $\boldsymbol{\psi_{s2}}(t)$ and vector $\boldsymbol{F_1}'(t)$ can be written as follow.

$$\lambda-(\lambda-\mu')=\mu' \tag{5}$$

At the same times, the angle between vector $\boldsymbol{\psi_{s2}}(t)$ and vector $\boldsymbol{F_2}'(t)$ can be written as follow.

$$\lambda-(\lambda+\mu')=-\mu' \tag{6}$$

Assuming k_{F1}=$k_{M1}\psi'_{m1}(t)$, k_{F2}=$k_{M2}\psi'_{m1}(t)$, the triangle $\Delta\boldsymbol{\psi_{s2}}(t+1)O\boldsymbol{\psi_{s2}}(t)$ is equivalent to triangle $\Delta\boldsymbol{F_1}(t+1)O\boldsymbol{F_1}'(t)$, which takes O as the endpoint with the counterclockwise rotation angle μ', and its sides length are reduced to $1/k_{F1}$ times as well as is

equivalent to triangle $\Delta \boldsymbol{F}_2(t+1)O\boldsymbol{F}_2'(t)$, which takes O as the endpoint with the counterclockwise rotation angle $-\mu'$, and its sides length are reduced to $1/k_{F2}$ times. So vector $\Delta \boldsymbol{\psi}_{s2}(t)$ is equivalent to vector $\Delta \boldsymbol{F}_1(t)$, which takes O as the endpoint with the counterclockwise rotation angle μ' and its sides length are reduced to $1/k_{F1}$ as well as is equivalent to vector $\Delta \boldsymbol{F}_2(t)$, which takes O as the endpoint with the counterclockwise rotation angle $-\mu'$ and its amplitude is reduced to $1/k_{F2}$ times.

Rotating coordinates transformation can be used to calculate the mathematic relationship about $\Delta \boldsymbol{F}_1$, $\Delta \boldsymbol{F}_2$and $\Delta \boldsymbol{\psi}_{s2}$. Fig. 3 shows the coordinates relationship increment of suspension force $\Delta \boldsymbol{F}_1$, $\Delta \boldsymbol{F}_2$ and increment of flux linkage $\Delta \boldsymbol{\psi}_{s2}$. So, $\Delta F_{1\alpha}$ and $\Delta F_{1\beta}$ are determined from αβ components of suspension force windings flux linkage $\Delta \psi_{s2\alpha}$ and $\Delta \psi_{s2\beta}$ as follows.

$$\begin{pmatrix} \Delta F_{1\alpha} \\ \Delta F_{1\beta} \end{pmatrix} = k_{F1} \begin{pmatrix} \cos\mu' & \sin\mu' \\ -\sin\mu' & \cos\mu' \end{pmatrix} \begin{pmatrix} \Delta \psi_{s2\alpha} \\ \Delta \psi_{s2\beta} \end{pmatrix} \tag{7}$$

Accordingly, $\Delta F_{2\alpha}$ and $\Delta F_{2\beta}$ are determined from αβ components of suspension force windings flux linkage $\Delta \psi_{s2\alpha}$ and $\Delta \psi_{s2\beta}$ as follows.

$$\begin{pmatrix} \Delta F_{2\alpha} \\ \Delta F_{2\beta} \end{pmatrix} = k_{F2} \begin{pmatrix} \cos\mu' & -\sin\mu' \\ \sin\mu' & \cos\mu' \end{pmatrix} \begin{pmatrix} \Delta \psi_{s2\alpha} \\ \Delta \psi_{s2\beta} \end{pmatrix} \tag{8}$$

The error of radial suspension force ΔF includes ΔF_1 the error of radial suspension force F_1 and ΔF_2 the error of radial suspension force F_2. So, the ΔF from αβ components can be written as follows.

$$\begin{pmatrix} \Delta F_{\alpha} \\ \Delta F_{\beta} \end{pmatrix} = \begin{pmatrix} (k_{F1}+k_{F2})\cos\mu' & (k_{F1}-k_{F2})\sin\mu' \\ -(k_{F1}-k_{F2})\sin\mu' & (k_{F1}+k_{F2})\cos\mu' \end{pmatrix} \begin{pmatrix} \Delta \psi_{s2\alpha} \\ \Delta \psi_{s2\beta} \end{pmatrix} \tag{9}$$

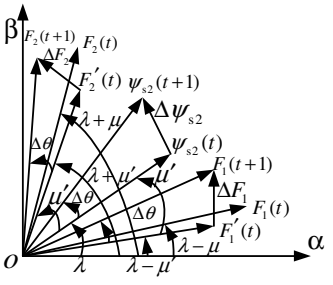


Fig. 2. Vector diagram of suspension force and flux linkage

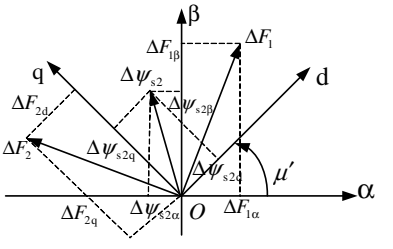


Fig. 3. Connected vector diagram of increment of suspension force ΔF_1, ΔF_2 and increment of flux linkage $\Delta \psi_{s2}$

3.3 DSFC System Design

Fig. 4 shows the block diagram of DSFC system. It can be seen that torque windings flux linkage calculator can real time calculate torque windings air-gap flux linkage

amplitude ψ_{m1} and phaseμ. The increments of rotor radial displacements are regulated by PID controllers to produce the command values of radical suspension force F_α^*, F_β^*. Suspension force windings flux linkage estimator calculates the amplitude ψ_{s2} and phaseλ, which can work out the feedback value of radical suspension force F_α, F_β by combining with the amplitude ψ_{m1} and phaseμ. Then the increment of suspension force windings flux linkage $\Delta\psi_{s2\alpha}$ and $\Delta\psi_{s2\beta}$ can be deduced by equation (9), according to the increments between radical suspension force F_α, F_β and its command values F_α^*, F_β^*. Finally the appropriate voltage space vectors are selected for suspension force windings, and voltage source inverter switching signals are generated by SVM.

4 Simulation Results

Control system of the BSynRM including torque control subsystem and radical suspension force control subsystem is constructed. Control strategy of torque control subsystem adopts the DTC based on SVM. DSFC algorithm is used to direct radical suspension force control subsystem. In order to verify the DSFC algorithm, the schematics shown in Fig. 4 can be implemented and simulated in Matlab6.5-Simulink environment.In this system, variable step size selects ode23t, and simulation time selects 0s to 0.1s. Simulation main parameters are given as follow.

Motor parameters: rotor quality m is 1kg, rotational inertia J is 0.002kg·m^2, air-gap δ_0 is 0.25mm, and auxiliary machinery bearings δ_1 is 0.20mm. Torque windings: rated voltage is 240V, rated current is 4.8A, and pole pairs p_M is 2, stator resistance R_s is 0.25Ω, direct axis inductance L_d is 35mH, cross axis inductance L_q is 4.2mH. Suspension force windings: pole pairs p_B is 1.

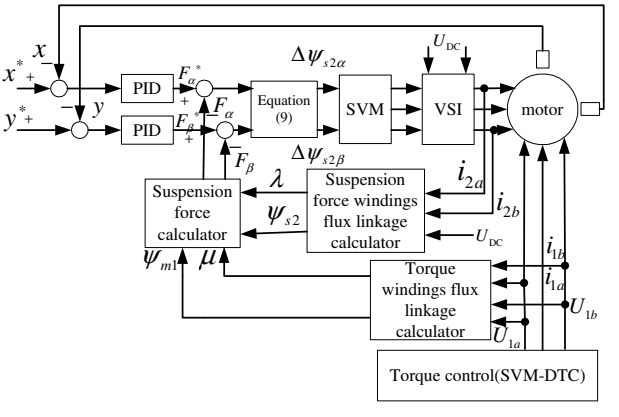
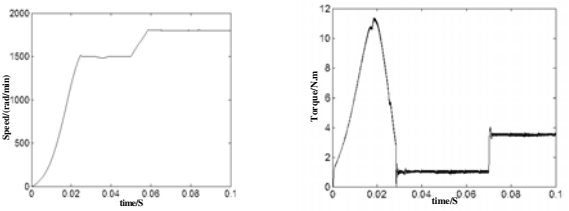


Fig. 4. Block diagram of direct suspension force control system

The motor load torque has been set to 1N·m, speed is 1500r/min at starting, then added to 1800r/min at t=0.03s, the Fig. 5 (a) shows the speed output characteristics, speed overshoot is under 0.4%, and the fluctuating error of speed in steady state is less than 10r/min. The speed has been set to 1500r/min, motor load torque is 1N·m at

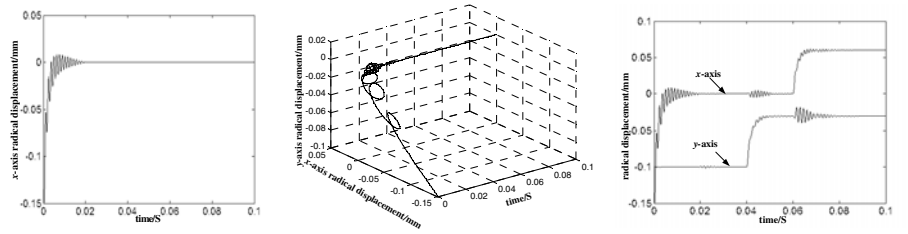
starting, then added to 3.5N·m at 0.07s. Fig. 5 (b) shows the performance of the torque, the pulsating movement of torque is less than 10%. Motor torque control system has good speed performance shown in Fig.5 (a)-(b).



(a) Speed output characteristics (b) Torque performance

Fig. 5. Simulation results of the torque control system

The starting position *x*-axis offset is-0.15mm, *y*-axis offset-0.10mm; Fig. 6 (a) shows the radial displacement in *x* axis; Fig. 6 (b) shows the track for the mass center of the suspended rotor. Although at t=0.07s step change of load torque from 0 to 3.5N.m, the radial displacement cannot be interfered by the load torque, indicating DSFC control algorithms realize electromagnetic torque and radial suspension force decoupling. Fig. 6(c) shows the dynamic response curve of *x*, *y*-axis radial displacement, with step change of *x*-axis from -0.15mm to 0.06mm at t=0.06s and *y*-axis from -0.1mm to -0.04mm at t=0.04s. It can be seen that the control system has good decoupling performance, torque and rotor radial suspension force can be controlled independently shown in Fig. 6 (a)-(c).



(a) *x* axis radial displacement (b) Rotor suspend (c) radial displacement response curve

Fig. 6. Simulation results of the control system

5 Conclusions

Based on the relation of flux linkage and suspension force of the BSynRM, the direct suspension force control (DSFC) method is proposed. The ideas of DTC are applied to a BSynRM system to control suspension force directly. The control algorithms of DSFC are deduced, and the simulation models of control system are established in Matlab-Simulink. The simulation results have shown that the method is valid, and the

control system has a good static and dynamic performance. The rotor can be suspended stability, torque and radial suspension forces can be controlled independently.

Acknowledgements. This work is sponsored by the National Natural Science Foundation of China (60974053), the Natural Science Foundation of Jiangsu Province (BK2009204), and the Research Fund for the Doctoral Program of Higher Education of China (20093227110002).

References

1. Chiba, A., Fukao, T., Rahman, M.A.: Radial Force and Speed Detection for Improved Magnetic Suspension in Bearingless Motors. IEEE Trans. on Ind. Appl. 42(2), 414–422 (2006)
2. Hertel, L., Hofmann, W.: Design and Test Results of High Speed Bearingless Reluctance Motors. In: Proceedings of PCIM, Nuremberg, German, pp. 143–147 (1999)
3. Chiba, A., Hanazawa, M., Fukao, T., Rahman, M.A.: Effects of Magnetic Saturation on Radial Forces of Bearingless Synchronous Reluctance Motors. IEEE Trans. on IAS 32(2), 354–362 (1996)
4. Michioka, C., Sakamoto, T., Chikawa, O.I., Chiba, A., Fukao, T.: A Decoupling Control Method of Reluctance-type Bearingless Motors Considering Magnetic Saturation. IEEE Trans on IAS 32(5), 1204–1210 (1996)
5. Hertel, L., Hofmann, W.: Magnetic Couplings in A Bearingless Reluctance Machine. In: International Conf. on Electrical Machines., vol. (3), pp. 1776–1780 (2000)
6. Ooshima, M., Chiba, A., Rahman, M.A.: An Improved Control Method of Buried-type IPM Bearingless Motors Considering Magnetic Saturation and Magnetic Pull Variation. IEEE Trans. Energy Convers. 19(3), 569–575 (2004)
7. Cao, J.R., Chen, Q.S., Yu, L.: Input-output Linearizing Control for Levitated Rotating Motors. IEEE Ind. Appl. Society, 1650–1654 (2002)
8. Zhou, L.Q., Zhu, J.H., Gu, C.L.: Research on Direct Torque Control of ALA Rotor Synchronous Reluctance Machine. In: Proceedings of the CSEE, vol. 24(7), pp. 155–158 (2004) (in Chinese)
9. Wang, Y., Deng, Z.Q., Wang, X.L.: Direct Torque Control of Bearingless Induction Motor. In: Proceedings of the CSEE, vol. 28(21), pp. 80–84 (2008) (in Chinese)
10. Sun, D., He, Y.K.: Space Vector Modulated Based Constant Switching Frequency Direct Torque Control for Permanent Magnet Synchronous Motor. In: Proceedings of the CSEE, vol. 25(12), pp. 112–116 (2005) (in Chinese)
11. Chiba, A., Chida, K., Fukao, T.: Principles and Characteristics of A Bearingless Synchronous Reluctance Motors with Windings of Magnetic Bearing. In: Proc. of IPEC, Tokyo, pp. 354–362 (1990)

Application of Wireless Sensor Network for Monitoring Water Temperature Difference in Blast Furnace Cooling System*

Yuefeng Liu and Jie Jiang

Information Engineering Institute, Inner Mongolia University of Science & Technology, Baotou, Inner Mongolia, China

Abstract. The wirless sensor network is a new kind of technique for getting or processing information. It's a new try for WSN with application to the blast furnace cooling system. This paper initiate the accomplishment of sensor nodes. And describe the construction and design of monitoring system. The system can monitor the water temperature difference of a blast furnace cooling system in real-time, and provide the blast furnace with basic data for its safety production.

Keywords: Wireless sensor network; blast furnace; cooling system; water temperature difference.

1 Introduction

At present, many large or middle blast furnace in china universally used the direct labor method monitoring water temperature at regular time. It's very dangerous to the workers because the blast furnace will leak gas everywhere at last. Sometimes the furnace jar had not been heated evenly during the blast furnace metallurgy, and the local edge gas stream bloat excessively so that water temperature increased rapidly. If this phenomenon can't be found in time, it will bring about the fatal accident such as furnace jar or furnace body may be burned to penetrate or explode and so on[1][2].

The shortcomings of direct labor monitoring are more workload, not in time and unsafe. Paper [1] and paper [2] proposed that using the DCS (Distributed Control System) to solve this problem, and effectively bridge the gap of direct labor method. But the sensor signal is transmitted by long pull wires, these wires will be damaged inevitably in the high temperature, so the maintain cost will be very high. Furthermore, temperature measuring points were distributed broadly and multitudinously, this made the wire layout very elaborate and installation cost also very high. Therefore, this paper mainly discussed how to accomplish the application of WSN(Wireless Sensor Network) for monitoring water temperature difference in blast furnace cooling system.

* Fund Project: Inner Mongolia Natural Science Foundation (20080404MS0703).
About the author: Yuefeng Liu (1977 -), Inner Mongolia Baotou people, Inner Mongolia University of Technology lecturer, Master.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 409–414.
springerlink.com

2 Wireless Monitoring Technic for Cooling System Water Temperature Difference

The temperature sensor uses bridge mode Pt1 000 thermistor of 1/3DIN grade, piping collection design used the standard pipe reed, installation and maintenance is very convenient [2]. With our own developed sensor nodes collect the temperature of every points, then every node send the data as particular rule to its own cluster head in the cluster. The cluster head collect all the data of sensor nodes in its own cluster, and then send these data to the relay node. Finally the relay nodes send the data to sink node in the operating room. See the structure in Figure 1.

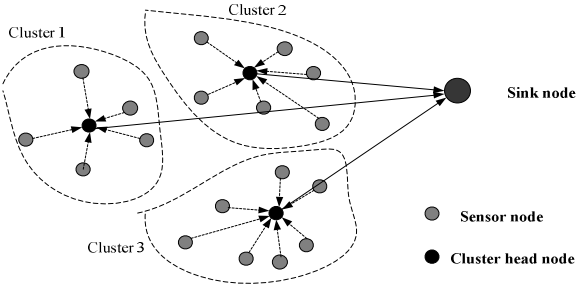


Fig. 1. System Structure of Monitoring Temperature

WSN can not only decrease the costs of installation and maintenance, but also can share some function of multiple path control unit under the system of DCS. 16 path of sensor signal are collected in one control unit before, it is changed now as the mode that every sensor node acquire the signal each other so that the data is more accurate and collected timely. According to each sensor node is a small processor, it has enough ability to compute and process. In this view, compared with DCS, WSN is a distribution system with strong fault tolerance.

2.1 Application Requirement Analysis

1) Water temperature measuring precision is ±0.05℃;
2) Temperature resolution ratio is 0.01℃;
3) There have been distributed more than 300 temperature measuring points in the area of 200 meter long, 20 meter width and 30 meter high around the blast furnace;
4) The distance between sensor nodes and operating room is 250 meter the most far and 50 meter the nearest;
5) Sensor node can work continuously for five months with four batteries.

2.2 System Sensing Reliance

System sensing result is based on the water temperature difference between the inner water and outer water in the cooling wall and the data of thermal stream strength, and finally is calculated by the administration computer.

Thermal stream strength:

$$J = (V \times \Delta t \times C \times 1000) / S \tag{1}$$

In the formula above, J is thermal stream strength; V is current velocity; Δt ($\Delta t = t\_out - t\_in$) is water temperature difference, t_out 、 t_in is outer water temperature and inner water temperature; C is water specific heat; S is cooling wall area [1].

2.3 Water Temperature Sensing Operational Principle

Constant-current source provide the electric current to bridge mode temperature sensor, and then invert the temperature variation of medium touched by sensor to the corresponding output voltage variation. Then the analog signals are sent to the A/D converter, amplified, converted and digital filtered. The measuring temperature data is sent to administration computer through WSN after linearization processed by single chip microcomputer.

2.4 Water Temperature Sensing Error Analysis

The temperature measuring error is very important to system. It is the key that how to reduce and calibrate error. The error is related to the grade of Pt thermistor and the nonlinearity of bridge circuit. The measuring bridge circuit is the signal convert link in the system, whose error is more important in the error of whole measuring system. Even using the Pt1000 sensor whose grade is A and allowed error is ±0.15℃, the error is also beyond the demand of system sensing precision. The Pt thermistor error becomes the main part of measuring error. So the system chooses the special Pt1000 which its grade is 1/3DIN, in order to reduce the sensor error. Besides, the system software also correct the nonlinearity of bridge circuit. According to "Industry Pt Thermistor Separated Table", using the standard resistance box calibrate temperature parameter in several section, and make it linear. Temperature measuring system send the temperature data after it is linear processed in several section and with multiple junction to the administration computer which shows these data. This intelligent function can further improve the measuring precision.

3 Sensor Nodes Design

In this paragraph we design the hardware device of sensor motes, and emphasis the realization that how to sense water temperature. The sensor node structure is illustrated by the Figure 2[3].

The key of sensor node design is that the sensor is very hard to reach the demand of water temperature sensing precision. In china some blast furnace water temperature difference is demanded between 0.3℃ and 0.4℃. It is proved in practice that the system could be normal applied in blast furnace only water temperature sensing precision reach to 0.05℃[1].

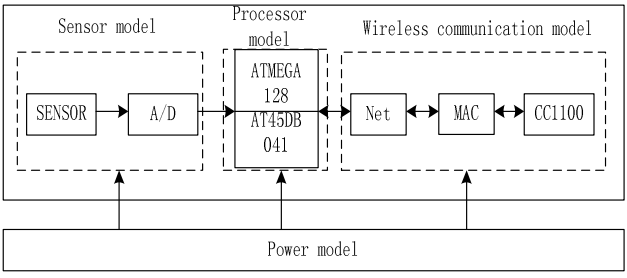


Fig. 2. Wireless Sensor Mote Structure

3.1 Wireless Communication Model

Radio chip chooses Chipcon's CC1100. CC1100 is a low cost true single chip UHF transceiver designed for very low power wireless applications. The circuit is mainly intended for the ISM (Industrial, Scientific and Medical) and SRD (Short Range Device) frequency bands at 315, 433, 868 and 915 MHz. Programmable data rate is up to 500 kbps. It can improve the transmission capability by using inner antecedent error correction, thus can be applied in the wireless ugly surroundings [4].

3.2 Processor Model

CPU chooses the ATMEAL's ATMEGA128. This microprocessor has abundant on-chip source and interface.

3.3 Sensor Model

The sensor is in charge of the function sensing inner water temperature or outer water temperature of measuring point. The sensor modular circuit is illustrated by Figure 3. The temperature signal acquired by sensor is converted through stable bridge circuit, amplified by the differential input circuit which consists of interface chip MAX492, transmitted to ADC1 interface of ATMEGA128, so thus the final temperature data is collected. The reason why analog signal do differential input is that it has very high common-mode rejection ratio for even times harmonic wave, it can improve the circuit function. There are also other advantages with differential input. First, it has high common-mode rejection ratio for noise signal cause by power supply and ground; Second, it has strong restraint function for in-phase signal cause by original vibration feedback.

The sensor of this system uses the bridge mode Pt1000 thermistor whose grade is 1/3DIN. Pt thermistor is a temperature sensor used very widely. Because of the nonlinearity of Pt thermistor, it must be corrected and calibrated in application. So we design a linearity calibration algorithm against the nonlinearity between temperature and electrical potential. Calibration uses segmentation linearity algorithm (segment every ten degree), improve the precision of measuring temperature; We also design a filter algorithm about the interfering signal. First continuously collect 16 instantaneously temperature data, then sort these data from low to high, for expel the interference data as

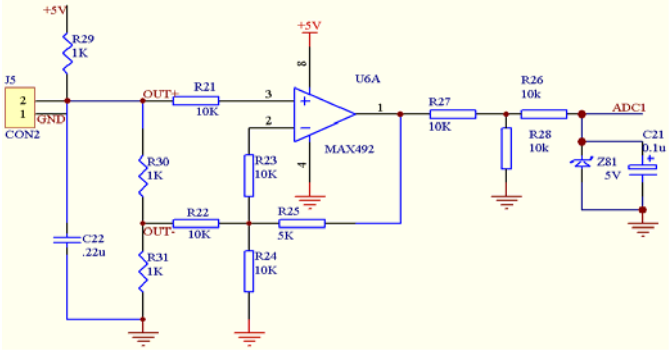


Fig. 3. Sensor circuit

possible, we respectively detract the biggest 3 number and the smallest 3 number, then figure out the average of the middle 10 number. It is practically proved that this filter algorithm can guarantee 2 numbers follow the decimal point keeping stable.

4 System Structure and Design

Blast furnace monitoring system for water temperature difference in cooling system consists of sink node, relay nodes, cluster heads and sensor nodes. See the structure in Figure 1. There are more than 300 sensor nodes, each cluster contain about 30 sensor nodes, sensor nodes communicate with sink node by cluster head and relay nodes. Every sensor node is assigned the cluster head number and index. This can assure the unique identity. Sensor nodes cyclical send the data to cluster head with TDMA mode, then the cluster head send data to sink node through the relay nodes, different cluster work under different TDMA mode. This can avoid channel collision. The following declaration is about the function of sink nodes, cluster head and sensor nodes.

The function of sink node is:

1)collecting the cluster head data;
2)maintaining the cluster heads time synchronization;
3)decreeing (user command, administrator command) to cluster heads;
4)sending data to administration computer;
5)accepting the order of administration computer.

According to the received data, the client judges whether the temperature is normal, whether the battery voltage down to the critical point, and whether need to alarm. The administration computer also has the function of data backup and data query.

The function of cluster head is:

1)accepting and confirming the sensor nodes data;
2)maintaining the time synchronization of sensor nodes in its own cluster;
3)sending data to sink nodes.

The sensor nodes are located in every part of blast furnace. Its main function is to acquire the temperature data at regular time and send the data to cluster head on the basis of regulation time-slot.

5 Conclusion

The costs of installation and maintenance are very low, the precision of measuring temperature is very high, and the expansibility of sensor nodes function is very strong. It can fully attain to the demand of production technology, and it has practical sense in guidance of operating blast furnace, assuring the production safety and prolonging the furnace life and so on.

References

1. Cui, D., Yang, Y., Jiang, J.: An on-line supervision system for water temperature difference and head flux intensity in blast furnace cooling system. Metallurgical Industry Automation (1), 55–57 (2004)
2. An, S., Jiang, J., Wang, J., Cui, D.: The on-line measurement of heat charge in a blast furnace cooling system. Industrial Instrumentation & Automation, 44–47 (2001)
3. King, M., Zhu, B.: Gaming strategies. In: Tang, S., King, M. (eds.) Path Planning to the West, vol. II, pp. 158–176. Jiaoda Press, Xian (1998); Sun, L., Li, J., Chen, Y., Zhu, H.: Wireless Sensors Network. TsingHua University Press (2005)
4. Lu, G., Krishnamachari, B., Raghavendra, C.: An adaptive energy-efficient and low-latency MAC for data gathering in wireless sensor networks. In: Proc 18th Int'l Parallel and Distributed Processing Symp. (IPDPS 2004), April 26-30, pp. 224–230 (2004)
5. Akyildiz, I., Su, W., Sankarasubramainiam, Y., Cayirci, E.: A survey on sensor networks. IEEE Commun. Mag. 40(8), 102–114 (2002)
6. Karaoguz, J.: High-rate Wireless Personal Area Networks. IEEE Commun. 39(12), 96–102 (2001)
7. Estrin, D., Girod, L., Pottie, G., Srivastava, M.: Instrumenting the world with wireless sensor networks. In: Proc. IEEE Intl. Conf. Acoustics, Speech, and Signal Processing, pp. 2033–2036 (2001)

Digital Control Technology of Bearingless Permanent Magnet Synchronous Motor

Jinfu Wei, Jing Yu, Qiuliang Cheng, Huangqiu Zhu, and Zebin Yang

School of Electrical and Information Engineering, Jiangsu University,
212013 Zhenjiang, China
weijinfu1986@163.com

Abstract. The digital control technology of a bearingless permanent magnet synchronous motor (BPMSM) includes torque control technology and radial suspension force control technology. On the basis of explaining the producing principle of radial suspension forces, electromagnetic torque equation and radial suspension force equations of the BPMSM are given. Based on the functional block diagram of the BPMSM control system, optimized software structure of the control system is designed, realizing methods of its various functional blocks and management mechanisms of two kinds of critical resources in TMS320LF2407A DSP are put forward. The experiment results show that, adopting this software structure and the realizing methods and management mechanisms, the functions of the BPMSM torque control, suspension control and human-computer interaction can be implemented well. The software structure and realizing methods and management mechanisms have important reference value for developing digital control of bearingless motors.

Keywords: bearingless motor; permanent magnet synchronous motor; torque; suspension; software structure; critical resource.

1 Introduction

A control system of bearingless permanent magnet synchronous motor (BPMSM) includes torque control subsystem and radial suspension forces control subsystem. The electromagnetic torque is produced by torque windings. The speed signal is gained by using rotary encoder or adopting sensorless operation technology, and then motor speed can be accurately controlled [1-2]. The radial suspension forces are produced by suspension force windings. Through adding control current of suspension force windings, the generated magnetic fields are superposed to the magnetic fields of torque windings, unbalanced magnetic fields are produced and then radial suspension forces are generated. When the difference between the pole pairs of torque windings and suspension force windings is 1, single radial suspension force in one direction can be generated [1-3]. The rotor radial displacements can be measured by virtue of capacitive displacement sensors or eddy current sensors [4-5]. The rotor can be suspended at the balance position by the closed loop control of rotor radial displacements [3]. Adopting digital signal processor (DSP), the rapidity requirement of response time can be satisfied for controlling rotor radial displacements [6].

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 415–422.
springerlink.com

In the paper, based on setting up electromagnetic torque equation and complete radial suspension force equations of the BPMSM, optimized software structure of the control system is designed, realizing methods of its various functional blocks and management mechanisms of two kinds of critical resources in TMS320LF2407A DSP are put forward., software program of the control system is then developed. The experiment results demonstrate that the functions of the BPMSM torque control, suspension control and human-computer interaction can be implemented well.

2 Production Principle of the Radial Suspension Forces

In order to produce controllable radial suspension forces, the pole pairs relationship between torque windings P_M and suspension force windings P_B should be $P_B=P_M\pm1$ [7]. Fig. 1 shows the production principle of the radial suspension forces. Additional 2-pole suspension force windings N_x and N_y are wound in the stator slots together with conventional 4-pole torque windings N_a and N_b. The radial suspension forces can be produced by the unbalanced magnetic flux density in the airgap caused by the interactions between 4-pole excitation magnetic fluxes Φ_p of permanent magnets and the magnetic fluxes generated by 2-pole suspension force winding currents i_x and i_y. For example, if the positive suspension force winding current i_x exits in N_x winding as shown in Fig. 1, the 2-pole magnetic fluxes Φ_x are generated. Therefore, the magnetic flux density increases in the airgap 1 while decreases in the airgap 3. Radial suspension force F is generated in the positive direction of x-axis. If the suspension force winding current is negative, a radial suspension force can be produced in the negative direction of x-axis. If currents exist in the N_y winding, y-axis direction force can be produced.

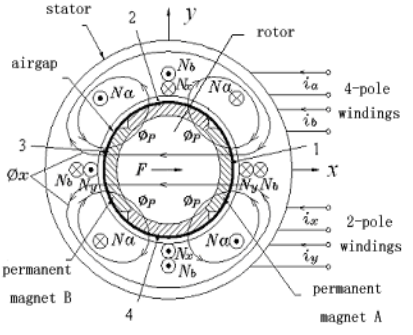


Fig. 1. Production principle of the radial suspension forces

It can be seen that the paths of 2-pole fluxes Φ_x pass through the permanent magnets A and B. It is well known that the permeability of permanent magnets is approximately equal to permeability in the air. Therefore, the thick permanent magnets result in large 2-pole MMF (Magnetomotive Force) requirements to produce radial suspension force. Consequently, thin permanent magnets are preferred to generate radial suspension force efficiently. However, thick permanent magnets have advantages to achieve

reasonable magnetic flux density for motor performance as well as to avoid demagnetization. So there exists an optimum thickness of permanent magnet to produce radial suspension forces most efficiently. The relationships between the radial suspension forces and permanent magnet thickness are described in the next section.

3 Mathematics Model of the BPMSM

In this paper, a two-phase machine model is used for simplicity, though a three-phase is practical. The current and magnetic flux linkage of the BPMSM are shown in a synchronous rotating reference coordinates. The relationship among the radial suspension forces and the currents of suspension force windings can be expressed as

$$\begin{cases} F_{ix} = (K_{\mathrm{M}} \pm K_{\mathrm{L}}) \cdot (i_{2d} \cdot \psi_{1d} + i_{2q} \cdot \psi_{1q}) \\ F_{iy} = (K_{\mathrm{L}} \pm K_{\mathrm{M}}) \cdot (i_{2q} \cdot \psi_{1d} - i_{2d} \cdot \psi_{1q}) \end{cases} \tag{1}$$

where, F_{ix}, F_{iy} are the radial suspension forces which are made up of Maxwell forces and Lorentz forces. K_{M} is Maxwell forces constant. K_{L} is Lorentz forces constant. i_{2d}, i_{2q} are current components of suspension force windings. ψ_{1d}, ψ_{1q} are the airgap magnetic flux linkages components of torque windings.

In addition, according to the theory of Electromagnetic Field, when the rotor is out of the center, another radial suspension force will be generated. This effect is known as the magnetic tensile force in the electromagnetic field of electrical motor. The generated Maxwell forces F_{sx}, F_{sy} are in proportion to the displacement. These inherent forces can be written as

$$\begin{cases} F_{sx} = k_s x \\ F_{sy} = k_s y \end{cases} \tag{2}$$

where, $k_s = k \cdot \dfrac{\pi r l B^2}{\mu_0 \delta}$, k_s is the force-displacement coefficient. μ_0 is the vacuum permeability. δ is the airgap length. k is the attenuation factor, $k \approx 0.3$.

So the radial suspension forces F_x and F_y in x- and y- direction can be expressed as

$$\begin{cases} F_x = F_{ix} + F_{sx} \\ F_y = F_{iy} + F_{sy} \end{cases} \tag{3}$$

Substituting expression (1), (2) into (3), so expression (3) can be written as

$$\begin{cases} F_x = (K_{\mathrm{M}} \pm K_{\mathrm{L}}) \cdot (i_{2d} \cdot \psi_{1d} + i_{2q} \cdot \psi_{1q}) + k_s \cdot x \\ F_y = (K_{\mathrm{L}} \pm K_{\mathrm{M}}) \cdot (i_{2q} \cdot \psi_{1d} - i_{2d} \cdot \psi_{1q}) + k_s \cdot y \end{cases} \tag{4}$$

When $P_{\mathrm{B}} = P_{\mathrm{M}} + 1$, expression (4) can be written as

$$\begin{cases} F_x = (K_{\mathrm{M}} + K_{\mathrm{L}}) \cdot (i_{2d} \cdot \psi_{1d} + i_{2q} \cdot \psi_{1q}) + k_s \cdot x \\ F_y = (K_{\mathrm{L}} + K_{\mathrm{M}}) \cdot (i_{2q} \cdot \psi_{1d} - i_{2d} \cdot \psi_{1q}) + k_s \cdot y \end{cases} \tag{5}$$

When P_B=P_M -1, expression (4) can be written as

$$\begin{cases} F_x = (K_M - K_L)\cdot(i_{2d}\cdot\psi_{1d} + i_{2q}\cdot\psi_{1q}) + k_s\cdot x \\ F_y = (K_L - K_M)\cdot(i_{2q}\cdot\psi_{1d} - i_{2d}\cdot\psi_{1q}) + k_s\cdot y \end{cases} \tag{6}$$

The stator magnetic flux linkage equation is as follows

$$\begin{cases} \psi_{1d} = L_d i_{1d} + \psi_r \\ \psi_{1q} = L_q i_{1q} \end{cases} \tag{7}$$

where, ψ_{1d} and ψ_{1q} are all airgap magnetic flux linkages. ψ_r is rotor magnetic flux linkage. L_d and L_q are the self-inductances of torque windings in the 2-phase rotating coordinates, respectively.

4 Control System Structure of the BPMSM

Functional block diagram of the BPMSM control system software is shown in Fig. 2 (when P_B=P_M+1). In this diagram, the control method, that is, adjusting voltage by enhancing magnetic flux is adopted in speed loop, x_{ref} and y_{ref} are the given rotor radial displacements in the *x*-axis and *y*-axis. ω^* is the command value of angular velocity. θ_1 denotes the initial phase angle between A-phase axis of torque windings and *x*-axis, θ_2 denotes the initial phase angle between A-phase axis of suspension force windings and *x*-axis.

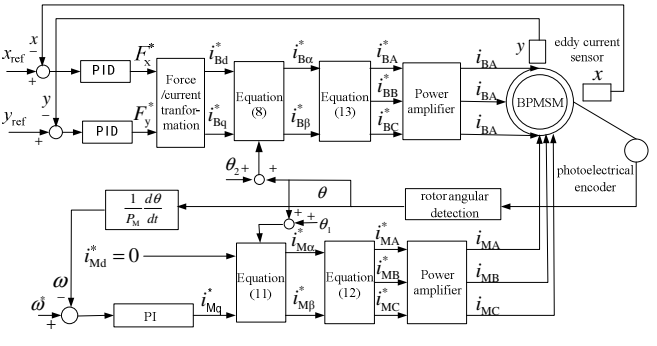


Fig. 2. Functional block diagram of the BPMSM control system software

The equation of PARK inverse transformation of the currents of suspension force windings can be written as

$$\begin{bmatrix} i_{B\alpha} \\ i_{B\beta} \end{bmatrix} = \begin{bmatrix} \cos(P_M\omega t + \Delta\theta_1 + \Delta\theta_2) & -\sin(P_M\omega t + \Delta\theta_1 + \Delta\theta_2) \\ \sin(P_M\omega t + \Delta\theta_1 + \Delta\theta_2) & \cos(P_M\omega t + \Delta\theta_1 + \Delta\theta_2) \end{bmatrix} \cdot \begin{bmatrix} i_{Bd} \\ i_{Bq} \end{bmatrix} \tag{8}$$

where, ω is the rotor's actual angular velocity, t is the time, $\Delta\theta_1$ is the space mechanical angle among torque windings, suspension force windings and *x*-axis, $\Delta\theta_2$ is the advanced electrical angle of torque windings airgap magnetic field. When P_B=P_M±1,

$$\Delta\theta_1 = P_B(\theta_2 - \theta_1) \pm \theta_1 = P_B\theta_2 - (P_B \mp 1)\theta_1 = P_B\theta_2 - P_M\theta_1 \tag{9}$$

When A-phase axis of torque windings and A-phase axis of suspension force windings superpose, namely, $\theta_1=\theta_2=\theta$, then

$$\Delta\theta_1 = P_B\theta_2 - P_M\theta_1 = P_B\theta - P_M\theta = (P_B - P_M)\theta = \pm\theta \tag{10}$$

The control system hardware of the BPMSM mostly includes four parts, that is, DSP controller, power amplifier, BPMSM interface circuit board and computer system. The functional block diagram of the control system hardware of the BPMSM is shown in Fig. 3. In this diagram, i^+_M and i^+_B are the modulated given currents of torque windings and suspension force windings, i^*_M and i^*_B are the demodulated given currents of torque windings and suspension force windings. i_M and i_B are the practical currents of torque windings and suspension force windings.

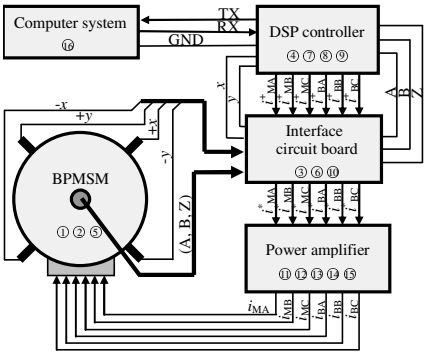


Fig. 3. Functional block diagram of the BPMSM control system hardware

In the control system, the functional modules are as follows: ①BPMSM; ②rotary encoder; ③adjusting circuit of rotary encoder's signals; ④quadrature encoding unit; ⑤rotor radial displacement sensor; ⑥adjusting circuit of rotor radial displacement sensor's signals; ⑦A/D convertor; ⑧processing, computing and controlling program; ⑨PWM signals generator; ⑩demodulating circuit of signals; ⑪generating circuit of power switch control signals;⑫ commutator, filter and inverter unit; ⑬malfunction control logical circuit; ⑭current sensor; ⑮adjusting circuit of current sensor's signals; ⑯human-computer interaction interface.

5 Control System Software of the BPMSM

The equations of PARK inverse transformation and CLARK inverse transformation of torque windings' currents adopted in the BPMSM control system in the paper are as follows

$$\begin{bmatrix} i_{M\alpha} \\ i_{M\beta} \end{bmatrix} = \begin{bmatrix} \cos(P_M \omega t) & -\sin(P_M \omega t) \\ \sin(P_M \omega t) & \cos(P_M \omega t) \end{bmatrix} \cdot \begin{bmatrix} i_{Md} \\ i_{Mq} \end{bmatrix} \tag{11}$$

$$\begin{bmatrix} i_{MA} \\ i_{MB} \\ i_{MC} \end{bmatrix} = \sqrt{\frac{2}{3}} \begin{bmatrix} 1 & 0 \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{1}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} \cdot \begin{bmatrix} i_{M\alpha} \\ i_{M\beta} \end{bmatrix} \tag{12}$$

The equation of CLARK inverse transformation of suspension force windings' currents can be written as

$$\begin{bmatrix} i_{BA} \\ i_{BB} \\ i_{BC} \end{bmatrix} = \sqrt{\frac{2}{3}} \begin{bmatrix} 1 & 0 \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{1}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} \cdot \begin{bmatrix} i_{B\alpha} \\ i_{B\beta} \end{bmatrix} \tag{13}$$

Fig. 4 shows the flowchart of this control system program, from the Fig. 4, the control program of motor synchronous rotation includes: calling subprogram of calculating sine value and cosine value of feedback rotor position angle, gaining torque windings' current value calculated by speed loop PI(Proportion Integration), calling PARK inverse transform subprogram, calling CLARK inverse transform subprogram and

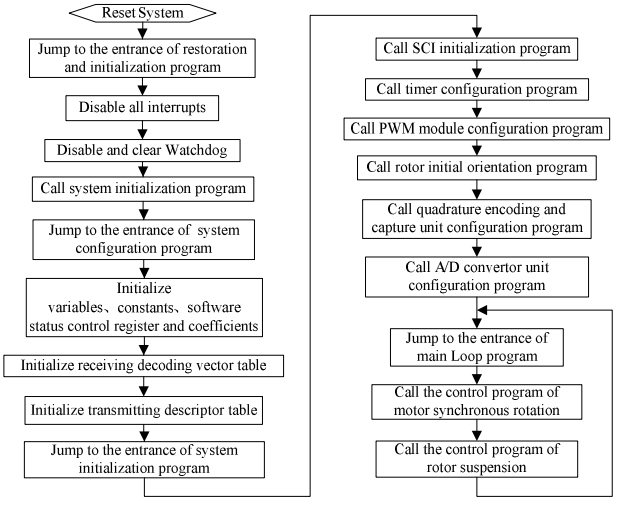


Fig. 4. Flowchart of the BPMSM control system program

calling sine wave PWM driving subprogram of torque windings. In this program, the sine value and cosine value of feedback rotor position angle are calculated through the above-mentioned method of look-up table. The control program of rotor suspension includes: getting corresponding sine value and cosine value based on the sum of feedback rotor position angle and $\Delta\theta_1$、$\Delta\theta_2$(By default this program runs without motor load, $\Delta\theta_2$=0) through the look-up table method(sharing the same sine table with the control program of motor synchronous rotation), gaining suspension force windings' current value calculated by rotor radial displacement loop PID(Proportion Integration Differentiation), calling PARK inverse transform subprogram, calling CLARK inverse transform subprogram and calling sine wave PWM driving subprogram of suspension force windings.

Critical resources which have been used in this program include two forms, one is that can be saved and then restored, the other is that can not be restored. The former involves accumulator ACC, auxiliary register AR0~AR7, status register ST0 and ST1, and this kind of critical resource is managed through locale information protection mechanism. The latter involves TREG register and PREG register of multiplier, and it is managed through controlling interrupt mode. Namely, before main program uses these resources all the related interrupts are disabled, and after using them, the corresponding interrupts are enabled again. This management mechanism of controlling interrupts makes program execution cause a period of delay. But the delay can be reduced to a minimum through designing program structure reasonably and enabling corresponding interrupts opportunely. The experiment results show that this management mechanism is viable, and the delay time can be controlled within 10 machine cycle under the condition of assuring program execution in the maximum efficiency.

6 Experiment Analysis

When the BPMSM operates steadily, the waveforms of given and feedback C-phase current of torque windings are shown in Fig. 5 (gained by the output voltage signal of eddy current sensor). Fig. 6 is the rotor's displacement wave in *x*-direction in the BPMSM (through detecting output voltage signal of eddy current sensor's interface circuit). From Fig. 6, it is obvious that rotor can be limited to the central position. Because of the factors of mechanism imbalance, and so on, rotor is recurrently divorced from the central position. However, it doesn't affect the validity and feasibility of suspension control.

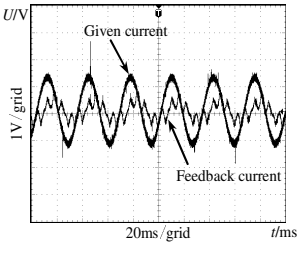


Fig. 5. Waveforms of given and feedback C-phase current in torque windings.

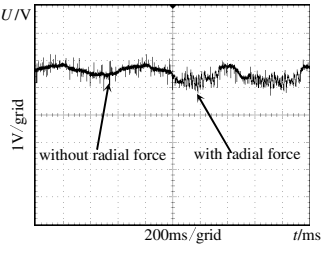


Fig. 6. Waveform of rotor displacement in *x*-direction.

7 Conclusion

This paper presents electromagnetic torque equation, radial suspension force equations and functional block diagram of control system of a BPMSM. On these bases, optimized software structure of the control system is designed, and realizing methods of various control parts and management mechanisms of two kinds of critical resources in TMS320LF2407A DSP are researched. The experiment results show that, this software structure and the realizing methods and management mechanisms can satisfy all functions of the BPMSM control and lay a foundation for developing digital control of bearingless motors.

Acknowledgements. This work is sponsored by the National Natural Science Foundation of China (60974053), the Natural Science Foundation of Jiangsu Province (BK2009204), and the Research Fund for the Doctoral Program of Higher Education of China (20093227110002).

References

1. Zhu, H.Q., Boesch, P.N., Jiang, J.W.: Design and Realization of a Digital Control System for a 4 kW Bearingless Permanent Magnet Slice Motor. In: The Eighth International Conference on Electrical Machines and Systems, Nanjing, vol. (3), pp. 1807–1810 (2005)
2. Nian, H., He, Y.K., Qin, F., Liu, Y.: Sensorless Operation of PM Type Bearingless Motor. In: Proceedings of the CSEE, vol. 24(11), pp. 101–105 (2004) (in Chinese)
3. Khoo, W.K.S.: Bridge Configured Winding for Poly-phase Self-bearing Machines. IEEE Trans. on Magn. 41(4), 1289–1295 (2005)
4. Salazar, A.O., Dunford, W., Stephan, R.: A Magnetic Bearing Dystem Using Capacitive Sensors for Position Measurement. IEEE Trans. on Magn. 26(5), 2541–2543 (1990)
5. Li, X.A., Wang, F.X.: Non-touch Sensors in Bearingless Magnetic Suspension Motor System. Small & Special Electrical Machines (3), 36–39 (2005) (in Chinese)
6. Ferreira, J.M.S., Paiva, D.J.A., Salazar, A.O.: DSP Utilization in Radial Positioning Control of Bearingless Machine. In: 2003 IEEE International Symposium on Industrial Electronics, vol. (1), pp. 312–317 (2003)
7. Bichsel, J.: The Bearingless Electrical Machine. In: Proc. Int. Symp. Magn. Suspension. Technol., pp. 561-573. NASA Langley Res. Center, Hampton(1991)

A New Ball Bearing Fault Diagnosis Method Based on EMD and SVM

Du Wei and Liu Quan

School of Information Engineering, Wuhan University of Technology Wuhan, China
duwei216@sina.com, quanliu@whut.edu.cn

Abstract. This paper presents a new method which combines empirical mode decomposition (EMD) and support vector machine (SVM) together for bearing fault diagnosis in low speed-high load rotary machine. EMD is a novel self-adaptive method which is based on partial characters of the signal. Vibration signal measured from a defective rolling bearing is decomposed into a number of intrinsic mode functions (IMFs), with each IMF corresponding to a specific range of frequency components contained within the vibration signal. Then calculate the energy entropy mean of each IMF and normalization motor speed(RPM) to construct feature vector to train SVM classifiers. The results of application in simulation signal and practical bearing fault signal both show its efficiency.

Keywords: EMD, IMF, SVM, feature vector, fault diagnosis, ball bearing.

1 Introduction

Bearing is one of the most important components in machine. Once it has any sort of faults, it would lead to serious economic lose. The traditional fault diagnosis techniques mainly contain demodulation and envelope-based method. But both have drawbacks. The first one is that bearing defect characters always perform as nonlinear and non-stationary. Secondly compared with the complex noise the bearing defect characters are very weak[1-2]. They are usually submerged in the noise signals and hard to be picked out. Finally, by using envelope-based method, it suffers from the drawback of having to determine a proper filtering width, in order to obtain consistent results under varying operating conditions. Therefore, it is quite difficult to extract the defect characters just by using traditional techniques.

Recently, a new signal analysis method, namely empirical mode decomposition (EMD) developed by Huang et al., has been based on the local characteristic time scale of the signal and can decompose the complicated signal into a number of intrinsic mode functions[3]. By analyzing each resulting IMF component that involves the local characteristic of the signal, the characteristic information of the original signal could be extracted more accurately and effectively. In addition, the frequency components involved in each IMF not only relates to sampling frequency but also changes with the signal itself; therefore, EMD is a self-adaptive signal processing method that can be applied to nonlinear and non-stationary process perfectly.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 423–428.
springerlink.com

In this paper, EMD is applied to the ball bearing fault diagnosis. The original acceleration vibration signal is decomposed by EMD and some IMF components are obtained, then the concept of energy entropy mean is introduced, which defined by calculating the mean value of the vibration signal entropies of a bearing with a fault in different various speeds a loads. Then we use the IMF energy entropy mean as an input of SVM classifiers to identify the bearing faults.

The theories of EMD and SVM are introduced in Section 2. Section 3 proposes the method combined EMD and SVM to bearing fault diagnosis. Section 4 is an experimental verification of the proposed method, as it is engaged in a bearing fault diagnosis. The conclusions are given in Section 5.

2 EMD Algorithm and SVM

2.1 EMD Algorithm

The EMD method is a fully data driven approach. Since the decomposition of the EMD is based on the local characteristics time scale of the data, it is applicable to nonlinear and non-stationary processes. The EMD decomposes into a sum of IMFs. An IMF is a function that must satisfy the following two conditions [4].

i. The number of extrema must either be equal to, or at most differ by one from the number of zero crossings.
ii. The mean values of both the envelope defined by the local maxima, and the envelope defined by the local minima, are zero at any point in the data.

The sifting process is defined by the following steps:

Given a signal x(t), the effective algorithm of EMD can be summarized as follows:
1) Determine all the local maxima, and local minima from the signal x(t)
2) Perform interpolation so that an upper envelope, $x_{up}(t)$, and a lower envelop, $x_{low}(t)$, can be formed by all of the local maxima, and local minima respectively.
3) Obtain the mean, $m(t)$, of the upper, and lower envelopes using

$$m(t) = \frac{x_{up}(t) + x_{low}(t)}{2} \tag{1}$$

4) Extract the detail $h_i(t) = x(t) - m(t)$, the index i is the number of the relevant IMF. If $h_i(t)$ could satisfy the two IMF conditions (that is, conditions i, and ii as defined before), then it is the valid IMF. Otherwise, it is not a valid IMF; iterate the above steps until a new IMF, which satisfies the two conditions, is found.
5) Define the residue $r_n(t) = x(t) - h_i(t)$, the index n is the number of the relevant IMF As $r_n(t)$ still contains much information from the lowest frequency; replace $x(t)$ with $r_n(t)$, and repeat the sifting process until the amplitude of the residue is

lower than a predetermined threshold, or it contains the lowest frequency component of the signal $x(t)$ .The original signal can be precisely reconstructed using

$$x(t) = \sum_{i=1}^{n} h_i(t) + r_n(t) \tag{2}$$

2.2 SVM

SVM is developed from the optimal separation plane under linearly separable condition. SVM uses a hypothesis space of linear functions in a high dimensional feature space to estimate decision surfaces directly rather than modeling a probability distribution across training data.[5-6] It uses support vector kernel to map the data from input space to a high dimensional feature space which facilitates the problem to be processed in linear form. SVM always finds a global minimum because it usually tries to minimize a bound on the structural risk, rather than the empirical risk. The structural risks, defined as a structure derived from the inner class of the function in the nested subset, find the subset of the function that minimizes the bound on the actual risk. SVM achieves this goal by minimizing the following Lagrangian formulation:

$$L = \frac{1}{2} \| \omega \|^2 - \sum_{i=1}^{l} \alpha_i y_i (x_i \bullet \omega + b) + \sum_{i=1}^{l} \alpha_i \tag{3}$$

Where α_i is positive Lagrange multiplier. SVM uses some kernels to map the data from the input space to a high dimensional feature space which facilitates the problem to be processed in linear form. In this paper linear, radial basis function (RBF), quadratic and polynomial kernels have been used.

3 Fault Diagnosis Combined EMD and SVM

While the ball bearing with different faults is operating, the corresponding resonance frequency components are produced in the vibration signals, and here the energy of fault vibration signal changes with the frequency distribution. To illustrate this change case as mentioned above, the energy entropy mean concept is proposed here.

If n IMFs and a residue $r_n(t)$ are obtained by using the EMD method to decompose the ball bearing vibration signal x(t) where the energy of the n IMFs is E1, E2,...,En, respectively; then, due to the orthogonality of the EMD decomposition, the sum of the energy of the n IMFs should be equal to the total energy of the original signal when the residue $r_n(t)$ is ignored. As the IMFs $h_i(t)$ include different frequency components, E={ E1, E2,...,En }, forms an energy distribution in the frequency domain of ball bearing vibration signals, and the corresponding EMD energy entropy is

$$H_{EN} = -\sum_{i=1}^{n} p_i \log p_i \tag{4}$$

Where $p_i = Ei / E$ is the percent of the energy of $h_i(t)$ in the whole signal energy, and $E = \sum_{i=1}^{n} E_i$ .Then energy entropy mean concept is as the mean of energy entropy of vibration signals collect from a faulty ball bearing in various speeds and loads. Therefore, IMFs' energy entropy mean could be used as fault feature vectors. After a fault feature vector has been extracted, SVM could be chosen as classifier to identify the work condition and fault pattern of bearing.

The EMD and SVM method for fault diagnosis can be summarized as follows:

1) Sample N times at a certain sample frequency fs under the condition that the bearing is normal, ball fault, inner-race fault, out-race fault respectively. And the 4N signals are taken as samples that are divided into two subsets, the training samples and testing samples.
2) Each sample signal is decomposed by EMD method and calculated energy entropy mean for each IMFs to construct one initial feature vector.
3) Each sample's motor speed(RPM) is normalized to construct another initial feature vector.

Design SVM classifier. The energy entropy mean and normalization RPM of the initial feature vector of the training samples are used as the fault feature vectors to be input to the SVM classifiers and the classifiers are trained.

4 Experimental Verification

In order to verify the efficiency of our method, we adopt fault test data from the Case Western Reserve University Bearing Data Center. Experiments were conducted using a 2 hp Reliance Electric motor, and acceleration data was measured at locations near to and remote from the motor bearings. Motor bearings were seeded with faults using electro-discharge machining (EDM). Faults ranging from 0.007 inches in diameter to 0.040 inches in diameter were introduced separately at the inner raceway, rolling element (i.e. ball) and outer raceway. Faulted bearings were reinstalled into the test motor and vibration data was recorded for motor loads of 0 to 3 horsepower (motor speeds of 1797 to 1720 RPM).

SVM classifiers are needed to design if four classes of bearing conditions are to be identified, like normal, with ball fault, with inner-race fault and with out-race fault. First of all, for SVM1, define the condition with normal as y =+1 and the other condition as y = −1, thus the normal condition could be separated from other condition by SVM1. Then define the condition with ball fault as y =+1 and the other condition as y = −1 for SVM2, thus the ball fault could be separated from other condition by SVM2.Then define the condition with inner-race fault as y =+1 and the other condition as y = −1 for SVM3, thus the inner-race fault could be separated from other condition by SVM3. Since we have known there are only three conditions to be identified, the remaining should be out-race fault. The identification approach is the same as above, that is, extract 11 samples as training ones at random (3 samples with normal condition, 4 samples with inner-race fault and 4 samples with out-race fault). The

Table 1. The identification results of proposed method for roller bearing

Testing Samples	Fault feature vector		SVM1 Distance 1	SVM2 Distance 2	SVM3 Distance 3	Identification Results
	Energy Entropy	Normalization RPM				
Normal	0.52845	0.5(1797)	0.7120 (+1)			Normal
Normal	0.48377	-0.5(1772)	0.5469 (+1)			Normal
Ball	0.43559	0.5(1797)	-0.8321 (-1)	0.6492 (+1)		Ball
Ball	0.40264	-0.5(1772)	-0.9762 (-1)	0.7941 (+1)		Ball
Inner-race	0.49733	0.5(1797)	-0.5211 (-1)	-0.8947 (-1)	0.9213 (+1)	Inner-race
Inner-race	0.45937	-0.5(1772)	-0.498(-1)	-0.5876 (-1)	1.0224 (+1)	Inner-race
Out-race	0.55867	0.5(1797)	-0.9561 (-1)	-0.6412 (-1)	-1.2246 (-1)	Out-race
Out-race	0.51346	-0.5(1772)	-0.7476 (-1)	-0.8454 (-1)	-0.8298 (-1)	Out-race

part identification results are shown in Table 1 from which we can see that two SVM classifiers can identify the working conditions and fault patterns of bearing accurately.

5 Conclusions

In this paper, original vibration signals based on EMD are preprocessed, IMF energy entropy mean are calculated. The energy entropy mean are extracted as training and testing samples of SVM. The final result shows that the proposed method of fault diagnosis for bearing based on EMD and SVM algorithms if effective.

Acknowledgments. This work was supported by State Program of National Natural Science of China(Grant No.50935005 and Grant No.50775167).

References

1. Yan, R., Gao, R.X.: Rotary Machine Health Diagnosis Based on Empirical Mode Decomposition. Journal of Vibration and Acoustics 130, 021007-1–021007-12 (2008)
2. Holm-Hansen, B.T., Gao, R.X.: Vibration Analysis of a Sensor Integrated Ball Bearing. ASME J. Vibr. Acoust. 122(5), 384–392 (2000)
3. Huang, N., Shen, Z., Long, S., Wu, M., Shih, H., Zheng, Q., Yen, N., Tung, C., Liu, H.: The empirical mode decomposition and the Hilbert spectrumfor nonlinear and non-stationary time series analysis. Proceeding of the Royal Society of London Series A-Mathematical Physical and Engineering Sciences 454, 903–995 (1998)

4. Yang, W.: Interpretation of mechanical signals using an improved Hilbert–Huang transform. Mechanical Systems and Signal Processing 22, 1061–1071 (2008)
5. Huang, P., Pan, Z., Qi, X., Lei, J.: Bearing Fault Diagnosis Based on EMD and PSD. In: Proceedings of the 8th World Congress on Intelligent Control and Automation, Jinan, China, July 6-9 (2010)
6. Chenga, J., Yua, D., Tangb, J., Yanga, Y.: Application of SVM and SVD technique based on EMD to the fault diagnosis of the rotating machinery. Shock and Vibration 16, 89–98 (2009)

Self-tuning Centralized Fusion Wiener Filter with Applied to Signal Processing

Jinfang Liu[1,2] and Zili Deng[1]

[1] Department of Automation, Heilongjiang University, Harbin 150080, China
[2] Department of Computer and Information Engineering,
Harbin Deqiang College of Commerce, Harbin 150025, China
dzl@hlju.edu.cn

Abstract. For the multisensor systems with unknown model parameters and noise variances, based on the system identification algorithm and correlation method, the estimators of model parameters and noise variances can be obtained. Based on the information matrix, a self-tuning centralized fusion Wiener filter is presented by substituting the estimators into the corresponding optimal filter. Using the dynamic error system analysis (DESA) method, it is proved that the self-tuning centralized fusion Wiener filter has asymptotic global optimality, i.e. it converges to the optimal centralized fusion Wiener filter. A simulation example applied to signal processing shows its effectiveness.

Keywords: Self-tuning, centralized fusion, Wiener filter, information matrix, convergence.

1 Introduction

No doubt that the multisensor information fusion has become one of the most popular fields for its widely application [1]. The conventional information fusion methods based on Kalman filtering include the centralized and distributed fusion methods [2]. The former gives the globally optimal state estimation by combining all local measurement data, with the disadvantage of requiring a larger computational burden. The latter gives the globally optimal and suboptimal state estimations by combining or weighting local state estimators [3], respectively. It can facilitate fault detection and isolation more conveniently, and reduce the computational burden. The centralized fusion method based on information matrix presented in [4] can give globally optimal estimation and avoid calculating the high-dimension inverse matrix.

In this paper, for the multisensor systems with unknown model parameters and noise variances, based on the information matrix approach, a self-tuning centralized fusion Wiener filter is presented. It makes up the shortcoming that the existing literatures of information filter are seldom and just for the system with unknown noise statistic [4,5]. It is strictly proved that the self-tuning fuser presented in this paper has asymptotical global optimality by the DESA method.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 429–436.
springerlink.com

2 Problem Formulation

Consider the linear discrete time-invariant stochastic system with L-sensor

$$x(t+1) = \Phi(\theta)x(t) + \Gamma w(t) \ . \tag{1}$$

$$y_i(t) = H_i x(t) + v_i(t) \ , i = 1,2,\cdots,L \ . \tag{2}$$

where t is the discrete time, $x(t) \in R^n$, $y_i(t) \in R^{m_i}$, $w(t) \in R^r$ and $v_i(t) \in R^{m_i}$ are the state, measurement, input noise and measurement noise, respectively. $\Phi(\theta)$, Γ and H_i are constant matrices with compatible dimensions. The vector $\theta \in R^s$ denotes the unknown constant parameters in the state transition matrix Φ, and each unknown element of Φ is a continuous function with respect to θ. When θ is known, we denote $\Phi(\theta) = \Phi$.

Assumption 1. (Φ, H_i) is a completely controllable pair, (Φ, Γ) is a completely observable pair, and Φ is a nonsingular matrix.

Assumption 2. $w(t) \in R^r$ and $v_i(t) \in R^{m_i}$ are uncorrelated white noises with zero means and variances $Q > 0$ and $R_i > 0$.

Assumption 3. The matrices H and Γ are known, vector θ and the noise variances Q and R_i are unknown.

Assumption 4. The measurement data $y_i(t)$ are bounded, i.e. a realization of the stochastic process $y_i(t)$ is bounded, $i = 1,2,\cdots,L$.

The problem is to find the self-tuning centralize fusion Wiener filter $\hat{x}^s(t \mid t)$ of the state $x(t)$ based on the information matrix.

3 Globally Optimal Centralized Fusion Wiener Filter

From (2), the centralized fusion measurement equation is given by

$$y(t) = Hx(t) + v(t) \ . \tag{3}$$

where

$$y(t) = \left[y_1^{\mathrm{T}}(t) \quad \cdots \quad y_L^{\mathrm{T}}(t)\right]^{\mathrm{T}}, H = \left[H_1^{\mathrm{T}} \quad \cdots \quad H_L^{\mathrm{T}}\right]^{\mathrm{T}}, v(t) = \left[v_1^{\mathrm{T}}(t) \quad \cdots \quad v_L^{\mathrm{T}}(t)\right]^{\mathrm{T}} \ . \tag{4}$$

and the fused measurement noise $v(t)$ has the variance as

$$R = \mathrm{diag}\left[R_1 \quad \cdots \quad R_L\right] \ . \tag{5}$$

When θ, Q and R_i are known, from (1) and (3), the globally optimal centralized fusion Kalman filter is given by

$$\hat{x}(t\,|\,t) = \Psi(t)\hat{x}(t-1\,|\,t-1) + K(t)y(t) \ . \tag{6}$$

$$\Psi(t) = P(t\,|\,t)\Sigma^{-1}(t\,|\,t-1)\Phi \ , \ K(t) = P(t\,|\,t)H^{\mathrm{T}}R^{-1} \ . \tag{7}$$

$$P^{-1}(t\,|\,t) = \Sigma^{-1}(t\,|\,t-1) + H^{\mathrm{T}}R^{-1}H \ , \ \Sigma(t+1\,|\,t) = \Phi P(t\,|\,t)\Phi^{\mathrm{T}} + \Gamma Q\Gamma^{\mathrm{T}} \ . \tag{8}$$

where inverse matrices $P^{-1}(t\,|\,t)$ and $\Sigma^{-1}(t\,|\,t-1)$ are called information matrix.

Theorem 1. For the multisensor system (1) and (2) with Assumption 1 and 2, the globally optimal centralized fusion Wiener filter is given by

$$\psi(q^{-1})\hat{x}(t\,|\,t) = M(q^{-1})\sum_{i=1}^{L} H_i^T R_i^{-1} y_i(t) \ . \tag{9}$$

where

$$\psi(q^{-1}) = \det(I_n - q^{-1}\Psi(t)) \ , M(q^{-1}) = \mathrm{adj}(I_n - q^{-1}\Psi(t))P(t\,|\,t) \ . \tag{10}$$

$\Psi(t)$, $P^{-1}(t\,|\,t)$, $\Sigma(t+1\,|\,t)$ are given by (7) and (8), and $P^{-1}(t\,|\,t)$ can also be given by a new form $P^{-1}(t\,|\,t) = \Sigma^{-1}(t\,|\,t-1) + \sum_{i=1}^{L} H_i^{\mathrm{T}} R_i^{-1} H_i$, which can reduce the computational burden because compared with the standard Kalman filter, it can avoid the computation of high-dimension inverse matrix.

Proof. From (3) - (7), $\hat{x}(t\,|\,t) = \Psi(t)\hat{x}(t-1\,|\,t-1) + P(t\,|\,t)\sum_{i=1}^{L} H_i^{\mathrm{T}} R_i^{-1} y_i(t)$ is obtained, which can be transformed into $(I_n - q^{-1}\Psi(t))\hat{x}(t\,|\,t) = P(t\,|\,t)\sum_{i=1}^{L} H_i^T R_i^{-1} y_i(t)$. It has been proved that $\Psi(t)$ is asymptotically stable [6], so $(I_n - q^{-1}\Psi(t))^{-1} = \mathrm{adj}(I_n - q^{-1}\Psi(t)) / \det(I_n - q^{-1}\Psi(t))$. Thus (9) and (10) hold. From (8) and applying (4) and (5), the new form of $P^{-1}(t\,|\,t)$ can be obtained. This completes the proof.

4 Self-tuning Centralized Fusion Wiener Filter

When the model parameters and the noise variances are unknown, the self-tuning centralized fusion Wiener filter can be realized by the following steps:

Step 1. Applying the system identification algorithm [7] (for example, RIV algorithm or RELS algorithm), the fused estimator $\hat{\theta}(t)$ of θ is obtained by taking the average of all local estimators.

Step 2. Applying the estimator $\hat{\theta}(t)$ and the correlation function method [8], the fused estimators $\hat{Q}(t), \hat{R}_i(t)$ of the noise variances Q, R_i are obtained in the similar way.

Step 3. Substituting all the fused estimators into Theorem 1 yields the self-tuning centralized fusion Wiener filter as

$$\hat{\psi}(q^{-1})\hat{x}^s(t\,|\,t) = \hat{M}(q^{-1})\sum_{i=1}^{L} H_i^T \hat{R}_i^{-1}(t)\, y_i(t) \ . \tag{11}$$

$$\hat{\psi}(q^{-1}) = \det(I_n - q^{-1}\hat{\Psi}(t)) \ , \hat{M}(q^{-1}) = adj(I_n - q^{-1}\hat{\Psi}(t))\hat{P}(t\,|\,t) \ . \tag{12}$$

$$\hat{\Psi}(t) = \hat{P}(t\,|\,t)\hat{\Sigma}^{-1}(t\,|\,t-1)\Phi(\hat{\theta}(t)) \ . \tag{13}$$

$$\hat{P}^{-1}(t\,|\,t) = \hat{\Sigma}^{-1}(t\,|\,t-1) + \sum_{i=1}^{L} H_i^{\mathrm{T}} \hat{R}_i^{-1}(t) H_i \ . \tag{14}$$

$$\hat{\Sigma}(t+1\,|\,t) = \Phi(\hat{\theta}(t))\hat{P}(t\,|\,t)\Phi^{\mathrm{T}}(\hat{\theta}(t)) + \Gamma\hat{Q}(t)\Gamma^{\mathrm{T}} \ . \tag{15}$$

The above three steps are repeated at each time t.

5 The Convergence Analysis

Assumption 5. The parameter estimator $\hat{\theta}(t)$ and the noise variance estimators $\hat{Q}(t)$ and $\hat{R}_i(t)$ are consistent, i.e.

$$\hat{\theta}(t) \to \theta, \Phi(\hat{\theta}(t)) \to \Phi(\theta), \hat{Q}(t) \to Q, \hat{R}_i(t) \to R_i \text{, as } t \to \infty \text{, w.p.1} \ . \tag{16}$$

where the notation "w.p.1" denotes "with probability one" [7].

Theorem 2. For the multisensor system (1) and (2) with Assumption 1-5, the self-tuning centralized fusion Wiener filter (11) converges to the optimal centralized fusion Wiener filter (9) in a realization, i.e.

$$[\hat{x}^s(t\,|\,t) - \hat{x}(t\,|\,t)] \to 0, \text{ as } t \to \infty, \text{ i.a.r.} \ . \tag{17}$$

Proof. When Assumption 1-5 are satisfied, for the globally optimal centralized fusion Kalman filter (6) - (8), by applying the dynamic variance error system analysis (DVESA) method [9], it has been proved [4] that $[\hat{\Sigma}(t+1\,|\,t) - \Sigma] \to 0$, $[\hat{P}(t\,|\,t) - P] \to 0$, $[\hat{K}(t) - K] \to 0$, $[\hat{\Psi}(t) - \Psi] \to 0$, as $t \to \infty$, i.a.r., where Σ , P , K and Ψ are the corresponding values in the stable Kalman filter and Ψ is a stable matrix, so $\hat{K}(t), \hat{\Psi}(t)$ are bounded. Then from (10) and (12), it holds that $\hat{\psi}(q^{-1}) \to \psi(q^{-1}), \hat{M}(q^{-1}) \to M(q^{-1})$, as $t \to \infty$, i.a.r..

Setting $\hat{\psi}(q^{-1}) = \psi(q^{-1}) + \Delta\hat{\psi}(q^{-1})$, $\hat{M}(q^{-1}) = M(q^{-1}) + \Delta\hat{M}(q^{-1})$ then we have $\Delta\hat{\psi}(q^{-1}) \to 0$, $\Delta\hat{M}(q^{-1}) \to 0$, as $t \to \infty$, i.a.r.. Defining $\delta(t) = [\hat{x}^s(t \mid t) - \hat{x}(t \mid t)]$, subtracting (9) from (11) yields the dynamic error system

$$\psi(q^{-1})\delta(t) = u(t) \ . \tag{18}$$

$$u(t) = M(q^{-1})[\sum_{i=1}^{L} H_i^T(\hat{R}_i^{-1}(t) - R_i^{-1})y_i(t)] + \Delta\hat{M}(q^{-1})\sum_{i=1}^{L} H_i^T\hat{R}_i^{-1}(t)y_i(t) - \Delta\hat{\psi}(q^{-1})\hat{x}^s(t \mid t) \ . \tag{19}$$

Assumption 4 and 5 yield that $\sum_{i=1}^{L} H_i^T\hat{R}_i^{-1}(t)y_i(t)$ is bounded. Then from (11), it can yield that $\hat{x}^s(t \mid t)$ is bounded [10]. Noting that $\psi(q^{-1})$ is a stable polynomial, which can be obtained from (10) and the stability of Ψ . And it yields $(\hat{R}_i^{-1}(t) - R_i^{-1}) \to 0$ from Assumption 5, i.e. $M(q^{-1})[\sum_{i=1}^{L} H_i^T(\hat{R}_i^{-1}(t) - R_i^{-1})y_i(t)] \to 0$. So it can be obtained that $u(t) \to 0$, as $t \to \infty$, i.a.r.. Thus from (18), it [10] yields $\delta(t) \to 0$, as $t \to \infty$, i.a.r., i.e. (17) holds. This completes the proof.

6 Application to Signal Processing

Consider the multisensor single channel Autoregressive (AR) signal with L-sensor

$$A(q^{-1})s(t) = w(t-1) \ . \tag{20}$$

$$y_i(t) = s(t) + v_i(t), i = 1,2,\cdots,L \ . \tag{21}$$

where $s(t)$ is the AR signal to be estimated, $y_i(t)$ is the measurements of the ith sensor, $w(t)$ and $v_i(t)$ are independent white noises with zero mean and variances Q and R_i, respectively. $A(q^{-1})$ is a stable polynomial of q^{-1} with the following form: $A(q^{-1}) = 1 + a_1q^{-1} + \cdots a_nq^{-n}$.

The AR signal model denoted by (20) and (21) can be transformed into the state space model denoted by (1) and (2), where Φ, Γ, H_i are given by

$$\Phi(\theta) = \begin{bmatrix} -a_1 & & & \\ -a_2 & & I_{(n-1)} & \\ \vdots & & & \\ -a_n & 0 & \cdots & 0 \end{bmatrix}, \Gamma = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, H_i = H = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} . \tag{22}$$

where Φ contains the unknown parameter vector $\theta = [a_1, a_2, \cdots, a_n]^T$, I_n denotes the $n \times n$ unite matrix. Thus, when $a_1, a_2, \cdots, a_n$, Q and R_i are unknown, the problem of

obtaining the self-tuning fusion Wiener signal filter $\hat{s}^s(t\,|\,t)$ can be converted into the one of obtaining $\hat{x}^s(t\,|\,t)$ because of the relationship

$$s(t) = Hx(t) \ . \tag{23}$$

Substituting (20) into (21) yields the least squares structure $y_i(t) = \varphi_i^T(t)\theta + r_i(t)$, with definition $\varphi_i(t) = [-y_i(t-1), \cdots, -y_i(t-n)]^T$ and $r_i(t) = w(t-1) + v_i(t) + a_1 v_i(t-1) + \cdots + a_n v_i(t-n)$. Based on the ith sensor, we can obtain the local RIV [7,11] estimates $\hat{\theta}_i(t)$ of θ, and it has been proved that the RIV estimate of the AR parameters is strongly consistent. Then the fused estimate $\hat{\theta}(t) = \frac{1}{L}\sum_{i=1}^{L}\hat{\theta}_i(t)$ is also strongly consistent, i.e. $\hat{\theta}(t) \to \theta$, as $t \to \infty$, w.p.1.

From (20), (21) and the definition of $r_i(t)$, we have

$$r_i(t) = A(q^{-1})y_i(t) = w(t-1) + A(q^{-1})v_i(t) \ . \tag{24}$$

Define the correlation function of $r_i(t)$ as $R_{rij}(k) = E[r_i(t)r_j(t-k)]$. Computing the correlation function of (24) yields

$$R_{rij}(0) = Q + \sum_{\alpha=0}^{n} a_\alpha^2 R_i \delta_{ij} \ , \ R_{rij}(k) = \sum_{\alpha=k}^{n} a_\alpha a_{\alpha-k} R_i \delta_{ij} \ , i, j = 1, \cdots, L \ , k = 1, \cdots, n \ . \tag{25}$$

where $\delta_{tt} = 1, \delta_{ik} = 0$. Then the estimator of $r_i(t)$ can be obtained by $\hat{r}_i(t) = \hat{A}(q^{-1})y_i(t)$. Further, the estimator $\hat{R}_{rij}^t(k)$ of the sampled correlation function is defined as $\hat{R}_{rij}^t(k) = \frac{1}{t}\sum_{\alpha=1}^{t}\hat{r}_i(\alpha)\hat{r}_j(\alpha-k)$, and it has been proved [8] that $\hat{R}_{rij}^t(k) \to R_{rij}(k)$, as $t \to \infty$, i.a.r.. Substituting the estimates $\hat{R}_{rij}^t(k)$ and $\hat{A}(q^{-1})$ into (25), the estimator of R_i can be obtained by

$$\hat{R}_{ik}(t) = \frac{\hat{R}_{rii}^t(k)}{\sum_{\alpha=k}^{n}\hat{a}_\alpha(t)\hat{a}_{\alpha-k}(t)}, \ \hat{R}_i(t) = \frac{1}{n}\sum_{k=1}^{n}\hat{R}_{ik}(t) \ , i = 1, \cdots, L \ , k = 1, \cdots, n \ . \tag{26}$$

where $\hat{R}_{ik}(t)$ and $\hat{R}_i(t)$ are the local and fused estimators of R_i, respectively. Similarly, the local and fused estimators of Q are obtained by

$$\hat{Q}_{ij}(t) = \hat{R}_{rij}^t(0) - \sum_{\alpha=0}^{n}\hat{a}_\alpha^2(t)\hat{R}_i(t)\delta_{ij}, \ \hat{Q}(t) = \frac{1}{L^2}\sum_{i,j=1}^{L}\hat{Q}_{ij}(t) \ , i, j = 1, \cdots, L \ . \tag{27}$$

Applying the method in [4], it can be easily proved that $\hat{R}_i(t) \to R_i, \hat{Q}(t) \to Q$, as $t \to \infty$, w.p.1. So Assumption 5 holds.

Substituting the estimators into the optimal centralized fusion Wiener filter, the self-tuning centralized fusion Wiener filter $\hat{x}^s(t\,|\,t)$ is obtained. Hence from (23), the self-tuning fusion Wiener signal filter can be given by

$$\hat{s}^s(t\,|\,t) = H\hat{x}^s(t\,|\,t) \ . \tag{28}$$

7 Simulation Example

Consider the single channel Autoregressive (AR) signal with L-sensor

$$(1 + a_1 q^{-1} + a_2 q^{-2}) s(t) = w(t-1) \ . \tag{29}$$

$$y_i(t) = s(t) + v_i(t), i = 1,2,3 \ . \tag{30}$$

where $s(t)$ is the AR signal to be estimated, $y_i(t)$ is the measurements of the ith sensor, $w(t)$ and $v_i(t)$ are independent white noises with zero mean and variances Q and R_i, respectively. In simulation we take $a_1 = 0.4, a_2 = -0.45, Q = 1$, $R_1 = 0.1, R_2 = 0.3, R_2 = 0.5$.

When the model parameters and noise variances are unknown, the fused estimators of the model parameters and noise variances are shown in Fig.1-Fig.3, where the curves and straight lines denote the estimates and real values, respectively. The error curves between the self-tuning and optimal fused Wiener signal filters are presented in Fig.4, which show the convergence of the self-tuning fusion Wiener signal filter.

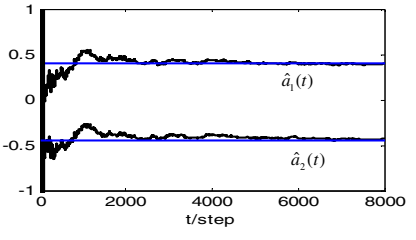


Fig. 1. Fused estimators of $a_j, j = 1,2$

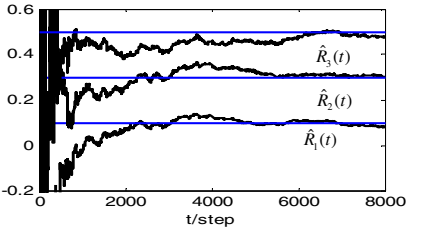


Fig. 2. Fused estimators of $R_i, i = 1,2,3$

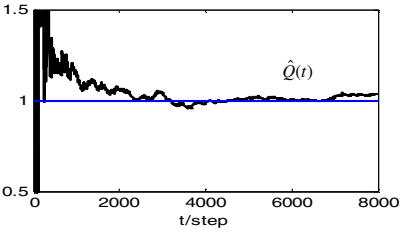


Fig. 3. Fused estimator of Q

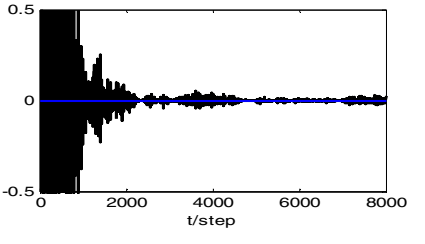


Fig. 4. Error curves $e(t) = \hat{s}^s(t\,|\,t) - \hat{s}(t\,|\,t)$

8 Conclusion

For the multisensor system with unknown model parameters and noise variances, the information fusion estimators are obtained by the system identification algorithm and the correlation method. Based on the fused estimators and information matrices, a new self-tuning fusion Wiener filter is presented. Further, it is rigorously proved that the self-tuning centralized fusion Wiener filter converges to the optimal centralized fusion Wiener filter in a realization by the DESA method, so it has asymptotical optimality.

Acknowledgment. The authors thank to the support from National Natural Science Foundation of China (60874063), Automatic Control Key Laboratory of Heilongjiang University, Support Program for Young Professionals in Regular Higher Education Institutions of Heilongjiang Province (1155G13), and Science and Technology Research Foundation of Heilongjiang Education Department (11553101).

References

1. Bar, S.Y., Li, X.R.: Multitarget-Multisensor Tracking: Principles and Techniques. YBS publishing, Storrs (1995)
2. Li, X.R., Zhu, Y.M., Wang, J., Han, C.Z.: Optimal linear estimation fusion-Part 1: Unified fusion rules. IEEE Trans. Information Theory 49(9), 2192–2208 (2001)
3. Zhu, Y.M., You, Z.H., Zhang, J., et al.: The optimality for the distributed Kalman filtering fusion with feedback. Automatica 37, 1487–1493 (2001)
4. Tao, G.L., Guan, X.H., Deng, Z.L.: Self-tuning Distributed Fusion Kalman Filter with Asymptotic Global Optimality. In: 29th Chinese Control Conference, pp. 1268–1272 (2010)
5. Guan, X.H., Deng, Z.L., Shi, Y.: Self-tuning centralized fusion information filter. Science Technology and Engineering 10(2), 1671–1815 (2010)
6. Kanen, E.W, Su, J.K: Introduction to Optimal Estimate. Springer-verlag Lindon Linited, Heidelberg (1999)
7. Deng, Z.L.: Multisensor Information Fusion Filtering Theory with Applications. Harbin Institute of Technology Press, Harbin (2007)
8. Deng, Z.L., Gao, Y., Li, C.B., Hao, G.: Self-tuning decoupled information fusion Wiener state component filters and their convergence. Automatica 44, 685–695 (2008)
9. Gu, L., Sun, X.J., Deng, Z.L.: The convergence analysis of the self-tuning Riccati equation. In: 2009 Chinese Control and Decision Conference, Guilin, pp.1165–1170 (2009)
10. Gao, Y., Jia, W.J., Sun, X.J., Deng, Z.L.: Self-tuning multisensor weighted measurement fusion Kalman filter. IEEE Trans. Aerospace and Electronic System 45(1), 179–191 (2009)
11. Ljung, L.: System Identification. Theory for the User, 2nd edn. Tsinghua University Press, Beijing (1999)

Covariance Intersection Fusion Kalman Filter

Peng Zhang[1,2], Wenjuan Qi[1], and Zili Deng[1]

[1] Department Of Automation, Heilongjiang University, Harbin 150080, P.R. China
[2] Harbin Deqiang College of Commerce, Harbin 150025, P.R. China
dzl@hlju.edu.cn

Abstract. By the covariance intersection (CI) fusion method, the covariance intersection fusion steady-state Kalman filter is presented for two-sensor system with unknown cross-covariances between local filter errors. It is proved that its accuracy is higher than that of each local filtering, and is lower than that of the optimal fuser with known cross-covariances. A Monte-Carlo simulation result shows that its accuracy is approximates to that of the optimal fuser.

Keywords: Information fusion Kalman filter, steady-state Kalman filter, covariance intersection fusion, unknown cross-covariances, covariance ellipse.

1 Introduction

Multisensor information fusion Kalman filtering has been applied to many fields, such as guidance, defence, robotics, tracking, signal processing, GPS positioning. To compute the optimal distributed weighted fusion Kalman filter [1,2] requires that the cross-covariances among local Kalman filtering errors are known exactly. However, in many practical applications, the cross-covariances are unknown or computation of the cross-covariances is very complex, or to find the cross-covariances is very difficult. In order to overcome these drawbacks, a covariance intersection (CI) fusion method was presented [3,4], which can handle the fusion problem with unknown cross-covariances. This paper, using the CI fusion method, a CI fusion steady-state Kalman filter is presented for two-sensor system with unknown cross-covariances. It is rigorously proved that the accuracy of CI fuser is higher than that of each local filter, and is lower than that of the optimal fuser with known cross-covariances, and the CI fuser is consistent, i.e. the actual variance for CI fuser has the theoretical upper bound obtained from the CI fuser.

2 The CI Kalman Fuser

Consider two-sensor system

$$x(t+1) = \Phi x(t) + \Gamma w(t) \tag{1}$$

$$y_i(t) = H_i x(t) + v_i(t),\ i = 1,2 \tag{2}$$

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 437–442.
springerlink.com

where $x(t) \in R^n$ is the state, $y_i(t) \in R^{m_i}$ is the measurement, $w(t) \in R^r$ and $v_i(t) \in R^{m_i}$ are uncorrelated white noises with zero mean and variances Q_w and Q_{vi}, respectively.

The local steady-state Kalman filters are given by [5]

$$\hat{x}_i(t \mid t) = \Psi_{fi} \hat{x}_i(t-1 \mid t-1) + K_{fi} y_i(t), \ i = 1,2 \tag{3}$$

$$\Psi_{fi} = [I_n - K_{fi} H_i] \Phi \tag{4}$$

$$K_{fi} = \Sigma_i H_i^{\mathrm{T}} [H_i \Sigma_i H_i^{\mathrm{T}} + Q_{vi}]^{-1} \tag{5}$$

where the symbol T denotes the transpose, Σ_i satisfies the Riccati equation and the steady-state filtering error variance P_i is given by

$$P_i = [I_n - K_{fi} H_i] \Sigma_i, \ i = 1,2 \tag{6}$$

the cross-covariance P_{12} between the local filtering errors satisfies the Lyapunov equation

$$P_{12} = \Psi_{f1} P_{12} \Psi_{f2}^{\mathrm{T}} + \Delta_{12} \tag{7}$$

$$\Delta_{12} = [I_n - K_{f1} H_1] \Gamma Q_w \Gamma^{\mathrm{T}} [I_n - K_{f2} H_2]^{\mathrm{T}} \tag{8}$$

when P_i and P_{12} are known, the optimal fuser weighted by matrices is given [1]

$$\hat{x}_0(t \mid t) = \Omega_1 \hat{x}_1(t \mid t) + \Omega_2 \hat{x}_2(t \mid t) \tag{9}$$

with the optimal weighted matrices [1,2]

$$\Omega_1 = (P_2 - P_{21})(P_1 + P_2 - P_{12} - P_{21})^{-1} \tag{10}$$

$$\Omega_2 = (P_1 - P_{12})(P_1 + P_2 - P_{12} - P_{21})^{-1} \tag{11}$$

with $P_{12} = P_{21}^T$. The fused error variance matrix P_0 is given by

$$P_0 = P_1 - (P_1 - P_{12})(P_1 + P_2 - P_{12} - P_{21})^{-1} (P_1 - P_{12})^{\mathrm{T}} \tag{12}$$

When P_1 and P_2 are known, but the cross-covariance P_{12} is unknown, the CI Kalman fuser without P_{12} is given by [3,4]

$$\hat{x}_{CI}(t \mid t) = P_{CI} [\omega P_1^{-1} \hat{x}_1(t \mid t) + (1-\omega) P_2^{-1} \hat{x}_2(t \mid t)] \tag{13}$$

$$P_{CI} = [\omega P_1^{-1} + (1-\omega) P_2^{-1}]^{-1} \tag{14}$$

where $\omega \in [0,1]$, and minimizes the performance index

$$\min_{\omega} \mathrm{tr} P_{CI} = \min_{\omega \in [0,1]} \mathrm{tr}\{[\omega P_1^{-1} + (1-\omega) P_2^{-1}]^{-1}\} \tag{15}$$

where the notation tr denotes the trace of matrix. The optimal weighting coefficient ω can fast be obtained by the gold section method or Fibonacci method [6].

Theorem 1. The local and fused Kalman filters have the accuracy relation

$$\mathrm{tr}\overline{P}_{CI} \le \mathrm{tr} P_{CI} \le \mathrm{tr} P_i, \quad i = 1,2 \tag{16}$$

where the theoretical variance P_{CI} is determined by (15), $\overline{P}_{CI}$ is the actual cross-covariance [3,4] for CI fuser $\hat{x}_{CI}(t \mid t)$ given by (13) without P_{12}.

Proof. From (15) we have that taking $\omega = 0$ yields $\mathrm{tr} P_{CI} = \mathrm{tr} P_2$, taking $\omega = 1$ yields $\mathrm{tr} P_{CI} = \mathrm{tr} P_1$. Hence from (15) we have $\mathrm{tr} P_{CI} \le \mathrm{tr} P_i$ for $\omega \in [0,1]$, $i = 1,2$. The relation $\overline{P}_{CI} \le P_{CI}$ was proved in [3,4], which yields $\mathrm{tr}\overline{P}_{CI} \le \mathrm{tr} P_{CI}$. This shows consistency of CI fuser.

Remark 1. Theorem 1 shows that the actual accuracy of CI fuser is higher than that of each local filter.

In order to give a powerful geometric interpretation with respect to accuracy relations of local and fused filters, the covariance ellipse for a covariance matrix P is defined as the locus of points $\{x : x^T P^{-1} x = c\}$ where c is a constant. In the sequel, $c = 1$ will be assumed without loss of generality. The following facts were proved: $P_1 \le P_2$ is equivalent to that the covariance ellipse for P_1 is enclosed in the covariance ellipse for P_2. The ellipse for fused covariance P_0 with known P_1, P_2 and P_{12} lies within the intersection of ellipses for P_1 and P_2. The ellipse for CI fused covariance P_{CI} with unknown P_{12} encloses the intersection region of ellipses for P_1 and P_2, and passes through the four points of intersection of ellipses for P_1 and P_2. These facts are shown in Fig.1.

Theorem 2. The local and fused Kalman filters have the accuracy relations

$$P_0 < P_i, \ \mathrm{tr} P_0 \le \mathrm{tr} P_i, \ i = 1,2 \tag{17}$$

$$P_0 \le P_{CI}, \ \mathrm{tr} P_0 \le \mathrm{tr} P_{CI} \tag{18}$$

Proof. The analytic proof was presented in [1]. The geometric proof is show in Fig.1. Since the ellipse for P_0 is enclosed in the intersection of ellipses for P_1 and P_2, then it is enclosed in the ellipse for P_1, which yields $P_0 \le P_1$, and it is also enclosed in the ellipse for P_2, which yields $P_0 < P_2$. Hence (17) holds. Since the ellipse for P_0 is enclosed in the ellipse for P_{CI}, which yields (18). The proof is completed.

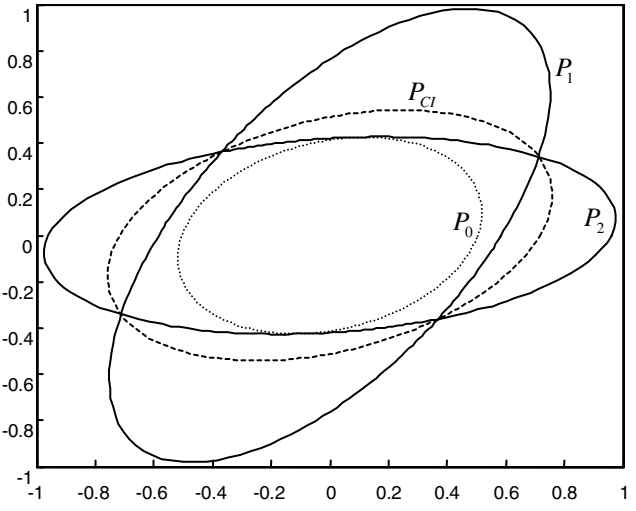


Fig. 1. The accuracy comparison of P_1, P_2, P_0 and P_{CI} by covariance ellipses

Remark 2. Theorem 2 shows that the accuracy of the optimal fuser is higher than that of local filters or CI fuser.

3 Simulation Example

Consider the two-sensor tracking system (1) and (2) with

$$\Phi = \begin{bmatrix} 1 & T_0 \\ 0 & 1 \end{bmatrix}, \ \Gamma = \begin{bmatrix} 0.5T_0^2 \\ T_0 \end{bmatrix} \tag{19}$$

$$H_1 = \begin{bmatrix} 1 & 0 \end{bmatrix}, \ H_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{20}$$

where $x(t) = \begin{bmatrix} x_1(t) & x_2(t) \end{bmatrix}^{\mathrm{T}}$ is the state, $x_1(t)$ and $x_2(t)$ are the position and velocity of target, T_0 is the sampled period, $w(t)$, $v_1(t)$ and $v_2(t)$ are white Gauss noises with zero mean and variances σ_w^2, σ_{v1}^2, Q_{v2}, respectively. In simulation we take $T_0 = 0.5$, $\sigma_w^2 = 2$, $\sigma_{v1}^2 = 1$, $Q_{v2} = diag(16, 0.25)$, $t = 1, \cdots, 300$.

In order to verify the above theoretical results for accuracy relation, 200 Monte-Carlo runs are performed for $t = 1, \cdots, 300$, the MSE curves of local and fused Kalman filters are computed, where the MSE value at time t is defined as the sampled average for $\mathrm{tr}P_i = \mathrm{trE}[(\hat{x}_i(t \mid t) - x(t))(\hat{x}_i(t \mid t) - x(t))^{\mathrm{T}}]$, (E denotes the expectation), i.e.

$$\mathrm{MSE}_i(t)=\frac{1}{N}\sum_{j=1}^{N}(\hat{x}_i^{(j)}(t\,|\,t)-x^{(j)}(t))^{\mathrm{T}}(\hat{x}_i^{(j)}(t\,|\,t)-x^{(j)}(t)),\ i=0,1,2,CI\ . \tag{21}$$

where $N=1,\cdots,200$, $t=1,\cdots,300$, $j=1,2$, $\hat{x}_i^{(j)}(t\,|\,t)$ or $x^{(j)}(t)$ denote the jth realization of $\hat{x}_i(t\,|\,t)$ or $x(t)$, respectively. According to ergodicity

$$\mathrm{MSE}_i(t)\to \mathrm{tr}P_i,\ as\ N\to\infty, t\to\infty, i=0,1,2,CI \tag{22}$$

The simulation results are shown in Fig.2.

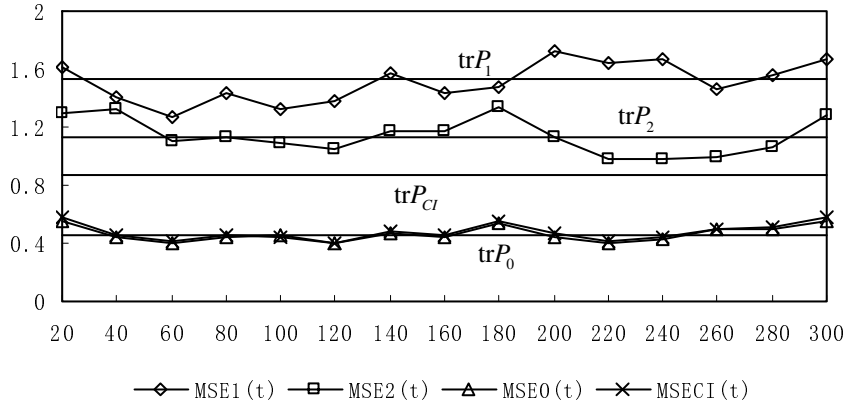


Fig. 2. The comparison of $\mathrm{MSE}_i(t)$ and $\mathrm{tr}P_i$, $i=0,1,2,CI$.

where the straight lines denote $\mathrm{tr}P_i$, $i=0,1,2,CI$, the curves denote the corresponding $\mathrm{MSE}_i(t)$. We see that the values of $\mathrm{MSE}_i(t)$, $(i=0,1,2)$ are close to the corresponding $\mathrm{tr}P_i$, which verifies the ergodicity (22). We also see that $\mathrm{MSE}_{CI}(t)<\mathrm{tr}P_{CI}$. According the ergodicity, $\mathrm{MSE}_{CI}(t)\to\mathrm{tr}\bar{P}_{CI}$, hence $\mathrm{tr}\bar{P}_{CI}<\mathrm{tr}P_{CI}$. This verifies P_{CI} is a upper bound of $\bar{P}_{CI}$. Finally, we see the CI fuser has good performance, whose accuracy is approximates to the accuracy of the optimal filter, because $\mathrm{MSE}_{CI}(t)$ is approximates to $\mathrm{MSE}_0(t)$.

4 Conclusion

A CI fusion steady-state Kalman filter has been presented for systems with unknown cross-covariances, which has advantage that the computation of cross-covariances can be avoided. It is proved that its accuracy is higher than that of each local filter, and is lower than that of the optimal fuser with known cross-covariances. Simulation results show that its accuracy is approximates to the accuracy of the optimal fuser.

Acknowledgments. This work is supported by the Natural Science Foundation of China under grant NSFC-60874063, and Foundation of Automatic Control Key Laboratory of Heilongjiang University.

References

1. Sun, S.L., Deng, Z.L.: Multi-sensor optimal information fusion Kalman filter. Automatica 40, 1017–1023 (2004)
2. Shalon, Y.B., Campo, L.: The effect of the common process noise on the two-sensor fused-track covariance. IEEE. Trans. Aerospace and Electronic Systems 22, 803–805 (1986)
3. Liggins, M.E., Hall, D.L., Llinas, J.: Handbook of Multisensor Data Fusion Theory and Practice, 2nd edn. CRC Press, Boca Raton (2009)
4. Julier, S.J., Uhlman, J.K.: Non-divergent estimation algorithm in the presence of unknown correlations. In: Proceedings of the IEEE American Control Conference, Albuquerque, NM, USA, pp. 2369–2373 (1997)
5. Kailath, T., Sayed, A.H., Hassibi, B.: Linear Estimation. Prentice-Hall, Upper Saddle River (2000)
6. Yuan, Y.X., Sun, W.Y.: Optimization Theory and Methods. Science Press, Beijing (2003)

Self-tuning Decoupled Fusion Kalman Smoother for Signal Processing

Guili Tao[1,2] and Zili Deng[1]

[1] Department of Automation, Heilongjiang University, 150080 Harbin, China
[2] Computer and Information Engineering College, Heilongjiang Institute of Science and Technology, 150027 Harbin, China
dzl@hlju.edu.cn, taoguili_5605@163.com

Abstract. For the multisensor systems with unknown model parameters and noise variances, based on the system identification method and correlation method, the online information fusion estimators of model parameters and noise variances can be obtained. Substituting them into the optimal fused Kalman smoother weighted by scalars for components, a self-tuning fusion Kalman smoother weighted by scalars for components is presented.The proposed self-tuning Kalman smoother converges to the time-varying optimal fusion Kalman smoother in a realization, so that it has asymptotic optimality. It can be applied to self-tuning signal processing. A simulation example shows its effectiveness.

Keywords: multisensor information fusion,decoupled information, identificatication, self-tuning Kalman smoother.

1 Introduction

With the high-accuracy requirement on target tracking and signal estimation in many high-technology fields, the multisensor information fusion has received great attention in recent year[1].The optimal fusion Kalman smoother by [2] requires the model parameters and noise statistics are known. This restricts their practical applications. And the filtering for the systems with unknown model parameters and/or noise variances is called self-tuning filtering [3]. Several self-tuning weighted fusion filters were presented only for systems with unknown noise variances [4-7]. Their drawbacks are that only the noise variances are assumed to be unknown, while the model parameters are assumed to be known.

In this paper, using the classical Kalman filtering method, the self-tuning information fusion Kalman smoothers weighted by scalars for components is presented for the multisensor systems with unknown model parameters and noise variances. The proposed self-tuning information fusion Kalman smoother converges to the optimal information fusion Kalman smoother in a realization, so it has asymptotic optimality. In [8], a self-tuning information fusion Kalman smoother is presented by using the modern time series analysis method. Therefore, compared with [8], the on-line identification of the ARMA innovation model is avoided.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 443–450.
springerlink.com

2 Optimal Fusion Kalman Smoother Weighted by Scalars for Components

Consider the multisensor linear discrete-time stochastic system

$$x(t+1) = \Phi(\theta)x(t) + \Gamma w(t) \tag{1}$$

$$y_i(t) = H_i x(t) + v_i(t), \quad i = 1,2,\cdots,L \tag{2}$$

where t is the discrete time, $x(t) \in R^n$, $y_i(t) \in R^{m_i}$, $w(t) \in R^r$ and $v_i(t) \in R^{m_i}$ are the state, measurement output, the input noise and measurement noise of the ith sensor subsystem, respectively. The transition matrix Φ contains a unknown parameter vector $\theta \in R^s$, and each element of Φ is a continuous function with respect to θ. When θ is known, we denote $\Phi(\theta) = \Phi$.

Assumption 1. $w(t)$ and $v_i(t)$ are the uncorrelated white noises with zero means and variances $Q > 0$ and $R_i > 0$, i.e.

$$\mathrm{E}\left\{\begin{bmatrix} w(t) \\ v_i(t) \end{bmatrix} [w^{\mathrm{T}}(k) \quad v_j^{\mathrm{T}}(k)]\right\} = \begin{bmatrix} Q & 0 \\ 0 & R_i\delta_{ij} \end{bmatrix} \delta_{tk} \tag{3}$$

where δ_{ij} is the Kronecker delta function, i.e. $\delta_{tt} = 1$, $\delta_{tk} = 0 (t \neq k)$, and the superscript T denotes the transpose.

Assumption 2. When θ is unknown, each unknown element of Φ is a continuous function with respect to θ. When θ is known, Φ is non-singular, (Φ, H_i) is a completely observable pair, (Φ, Γ) is a completely controllable pair.

Assumption 3. The matrices H_i and Γ are known, but the parameter vector θ, and the noise variances Q and $R_i (i = 1,\cdots,L)$ are completely or partially unknown.

Assumption 4. The measurement data $y_i(t)$ (a realization of measurement stochastic process $y_i(t)$) are bounded for t, $i = 1,2,\cdots,L$.

Denoting the state $x(t)$ in the component form as

$$x(t) = [x_1(t), \quad \cdots, \quad x_n(t)]^{\mathrm{T}} \tag{4}$$

The problem is to find the self-tuning information fusion Kalman smoother weighted by scalars.

Lemma 1 [9]. For the multi-sensor system (1) and (2) with known model parameters and noise variances, the ith sensor subsystem has the local optimal Kalman predictor $\hat{x}_i(t+1|t)$ of $x(t)$ as

$$\hat{x}_i(t+1|t) = \Psi_{Pi}(t)\hat{x}_i(t|t-1) + K_{Pi}(t)y_i(t) \tag{5}$$

$$\Psi_{Pi}(t) = \Phi - K_{Pi}(t)H_i \tag{6}$$

$$K_{Pi}(t) = \Phi\Sigma_i(t \mid t-1)H_i^{\mathrm{T}}(H_i\Sigma_i(t \mid t-1)H_i^{\mathrm{T}} + R_i)^{-1} \tag{7}$$

where the prediction error variance matrix satisfy the optimal Riccati equations

$$\begin{aligned}\Sigma_i(t+1 \mid t) = \Phi[\Sigma_i(t \mid t-1) - \Sigma_i(t \mid t-1)H_i^{\mathrm{T}}(H_i\Sigma_i(t \mid t-1)H_i^{\mathrm{T}} + R_i)^{-1} \\ \times H_i\Sigma_i(t \mid t-1)]\Phi^{\mathrm{T}} + \Gamma Q\Gamma^{\mathrm{T}}\end{aligned} \tag{8}$$

and the local prediction cross-covariances satisfy the Lyapunov equation

$$\Sigma_{ij}(t+1 \mid t) = \Psi_{Pi}(t)\Sigma_{ij}(t \mid t-1)\Psi_{Pj}^{\mathrm{T}}(t) + \Gamma Q\Gamma^{\mathrm{T}},\ i \neq j,\ i, j = 1,2,\cdots,L \tag{9}$$

Lemma 2 [9]. For the multi-sensor system (1) and (2) with assumptions 1 and 2, the ith sensor subsystem has the local optimal time-varying Kalman smoother as

$$\hat{x}_i(t-N \mid t) = \hat{x}_i(t-N \mid t-N-1) + \sum_{j=0}^{N} K_i(t-N \mid t-N+j)\varepsilon_i(t-N+j), i = 1,\cdots,L \tag{10}$$

where the optimal Kalman predictor $\hat{x}_i(t-N \mid t-N-1)$ can be computed by (5), and we can obtain

$$\varepsilon_i(t) = y_i(t) - H_i\hat{x}_i(t \mid t-1) \tag{11}$$

$$\begin{aligned}K_i(t \mid t+j) = \Sigma_i(t \mid t-1)\{\prod_{k=0}^{j-1}\Psi_{pi}^{\mathrm{T}}(t+k)\}H_i^{\mathrm{T}}(H_i\Sigma_i(t+j \mid t+j-1)H_i^{\mathrm{T}} + R_i)^{-1} \\ K_i(t \mid t) = \Sigma_i(t \mid t-1)H_i^{\mathrm{T}}(H_i\Sigma_i(t \mid t-1)H_i^{\mathrm{T}} + R_i)^{-1}\end{aligned} \tag{12}$$

The error variance matrices and covariance matrices of local optimal Kalman smoother are given as

$$\begin{aligned}P_i(t-N \mid t) = \Sigma_i(t-N \mid t-N-1) - \sum_{j=0}^{N} K_i(t-N \mid t-N+j) \\ \times(H_i\Sigma_i(t-N+j \mid t-N-1+j)H_i^{\mathrm{T}} + R_i)K_i^{\mathrm{T}}(t-N \mid t-N+j)\end{aligned} \tag{13}$$

$$\begin{aligned}P_{ij}(t-N \mid t) = \Psi_{iN}(t-N)\Sigma_{ij}(t-N \mid t-N-1)\Psi_{jN}^{\mathrm{T}}(t-N) + \sum_{\rho=0}^{N} K_{i\rho}^{w}(t-N)QK_{j\rho}^{w\mathrm{T}}(t-N) \\ i \neq j,\ i, j = 1,2,\cdots,L\end{aligned} \tag{14}$$

with the definition $P_{ii}(t-N \mid t) = P_i(t-N \mid t)$.

$$\begin{aligned}\Psi_{iN}(t-N) = I_n - \sum_{k=0}^{N} K_i(t-N \mid t-N+k)H_i\Psi_{pi}(t-N+k, t-N),\ K_{iN}^{w}(t-N) = 0 \\ K_{i\rho}^{w}(t-N) = -\sum_{k=\rho+1}^{N} K_i(t-N \mid t-N+k)H_i\Psi_{pi}(t-N+k, t-N+\rho+1)\Gamma, \\ \rho = 0,\cdots,N-1\end{aligned} \tag{15}$$

Lemma 3 [9]. For the multi-sensor systems (1) and (2) with assumptions 1 and 2, we have the optimal information fusion Kalman smoother weighted by scalars for components

$$\begin{aligned}\hat{x}_0(t-N\,|\,t) &= \sum_{j=1}^{L} \Omega_j(t)\hat{x}_j(t-N\,|\,t) \\ \Omega_j(t) &= \mathrm{diag}(\omega_{j1}(t),\cdots,\omega_{jn}(t))\end{aligned} \tag{16}$$

where the optimal scalar weighting coefficient vectors $\Omega_i(t)$ are given by

$$\omega_i(t) = [\omega_{1i}(t), \quad \cdots, \quad \omega_{Li}(t)] = [e^{\mathrm{T}}(P^{ii}(t-N\,|\,t))^{-1}e]^{-1}e^{\mathrm{T}}(P^{ii}(t-N\,|\,t))^{-1}, i=1,\cdots,n \tag{17}$$

where $e^{\mathrm{T}} = [1,\cdots,1]$, and the $L\times L$ matrix $P_{kj}^{ii}(t-N\,|\,t)$ is defined as

$$P^{ii}(t-N\,|\,t) = (P_{kj}^{ii}(t-N\,|\,t)), \ k, j = 1,\cdots,L \tag{18}$$

whose (k,j) element $P_{kj}^{ii}(t-N\,|\,t)$ are the (i,i) diagonal elements of $P_{kj}(t-N\,|\,t)$.

3 Self-tuning Fusion Kalman Smoother Weighted by Scalars for Components

When model parameters and noise variances are unknown, substituting their estimators into the optimal fusion Kalman smoother will yield the self-tuning fusion Kalman smoother. It consists of the following steps:

Step 1. Applying the system identification algorithm[10](for example, the recursive instrumental variable (RIV) algorithm, the recursive extended least square (RELS) algorithm), and the correlation method [11], the information fusion estimators $\hat{\theta}(t)$, $\hat{Q}(t)$ and $\hat{R}_i(t)$ are that for the model parameters θ and noise variances Q, $R_i(i=1,\cdots,L)$ can be obtained [12].

Step 2. In Lemma 1, Φ, Q and $R_i(i=1,\cdots,L)$ are replaced by $\Phi(\hat{\theta}(t))$, $\hat{Q}(t)$ and $\hat{R}_i(t)$, repectively. So the estimates $\hat{\Psi}_{pi}(t)$ and $\hat{K}_{pi}(t)$ can be obtained. And the estimates $\hat{\Sigma}_i(t\,|\,t-1)$ satisfy the self-tuning Riccati equations

$$\begin{aligned}\hat{\Sigma}_i(t+1\,|\,t) = \Phi(\hat{\theta}(t))[\hat{\Sigma}_i(t\,|\,t-1) - \hat{\Sigma}_i(t\,|\,t-1)H_i^{\mathrm{T}}(H_i\hat{\Sigma}_i(t\,|\,t-1)H_i^{\mathrm{T}} + \\ \hat{R}_i(t))^{-1}H_i\hat{\Sigma}_i(t\,|\,t-1)]\Phi^{\mathrm{T}}(\hat{\theta}(t)) + \Gamma\hat{Q}(t)\Gamma^{\mathrm{T}}\end{aligned} \tag{19}$$

and the prediction cross-covariance matrices satisfy the self-tuning Lyapunov equations

$$\hat{\Sigma}_{ij}(t+1\,|\,t) = \hat{\Psi}_{Pi}(t)\hat{\Sigma}_{ij}(t\,|\,t-1)\hat{\Psi}_{Pj}^{\mathrm{T}}(t) + \Gamma\hat{Q}(t)\Gamma^{\mathrm{T}} \ i\neq j \ , i, j = 1,2,\cdots,L \tag{20}$$

Substituting the estimates into (5), self-tuning local Kalman predictor can be given as

$$\hat{x}_i^s(t+1|t) = \hat{\Psi}_{Pi}(t)\hat{x}_i^s(t|t-1) + \hat{K}_{Pi}(t)y_i(t) \tag{21}$$

Substituting the estimates into (10)-(15), yields the self-tuning local Kalman smoother

$$\hat{x}_i^s(t-N|t) = \hat{x}_i^s(t-N|t-N-1) + \sum_{j=0}^{N} \hat{K}_i(t-N|t-N+j)\hat{\varepsilon}_i(t-N+j), i = 1,\cdots,L \tag{22}$$

And, the estimates $\hat{\varepsilon}_i(t), \hat{K}_i(t-N|t-N+j), \hat{P}_i(t-N|t), \hat{P}_{ij}(t-N|t)$ can be obtained from Lemma 2.

Step 3. Applying (18) and (19), the estimates $\hat{P}(t-N|t)$ and $\hat{\Omega}_i(t)$ are obtained, and the self-tuning fused Kalman smoother is given as

$$\hat{x}_0^s(t-N|t) = \sum_{j=1}^{L} \hat{\Omega}_j(t)\hat{x}_j^s(t-N|t) \tag{23}$$

The above three steps are repeated at each time t.

Assumption 5. The parameters and noise variance estimators $\hat{\theta}(t)$, $\hat{Q}(t)$ and $\hat{R}_i(t)(i=1,\cdots,L)$ are consistent, i.e.

$$\hat{\theta}(t) \to \theta, \hat{Q}(t) \to Q, \hat{R}_i(t) \to R_i \text{ , as } t \to \infty, \text{ i.a.r} \tag{24}$$

where the notation "i.a.r" denotes "in a realization" .

Applying to the dynamic variance error system analysis (DVESA) method [13, 14], the solution $\hat{\Sigma}_{ij}(t+1|t)$ of the self-tuning Lyapunov equation (20) converges to the solution $\Sigma_{ij}(t+1|t)$ of the time-varying optimal Lyapunov equation (9), i.e.

$$[\hat{\Sigma}_{ij}(t+1|t) - \Sigma_{ij}(t+1|t)] \to 0 \text{ , as } t \to \infty \text{ , i.a.r} \tag{25}$$

Then, it can be proved that local self-tuning Kalman predictor $\hat{x}_i^s(t+1|t)$ converges to the local optimal Kalman predictor $\hat{x}_i(t+1|t)$ by the the dynamic error system analysis (DESA) method [4],i.e.

$$[\hat{x}_i^s(t+1|t) - \hat{x}_i(t+1|t)] \to 0, \text{ as } t \to \infty \text{ , i.a.r.} \tag{26}$$

From (25) and (26), it can be obtained easily that the local self-tuning Kalman smoother $\hat{x}_i^s(t-N|t)$ converges to the local optimal Kalman smoother $\hat{x}_i(t-N|t)$, and the self-tuning fused Kalman smoother $\hat{x}_0^s(t-N|t)$ converges to the optimal fused Kalman smoother $\hat{x}_0(t-N|t)$ in a realization, i.e.

$$\hat{x}_i^s(t-N|t) - \hat{x}_i(t-N|t) \to 0. \text{ as } t \to \infty \text{ , i.a.r.} \tag{27}$$

$$\hat{x}_0(t-N|t) \to \hat{x}_0(t-N|t), \text{ as } t \to \infty \text{ , i.a.r.} \tag{28}$$

4 Simulation Example

Consider the multisensor single channel Autoregressive (AR) signal with white measurement noise

$$A(q^{-1})s(t) = w(t-1) \tag{29}$$

$$y_i(t) = s(t) + v_i(t), i = 1,2,3 \tag{30}$$

$$A(q^{-1}) = 1 + a_1 q^{-1} + a_2 q^{-2} + a_3 q^{-3} \tag{31}$$

where $y_i(t) \in R$ is the measurement of the ith sensor, $v_i(t)$, $w(t)$ are independent Gaussian white noises with zero mean and variance $\sigma_{vi}^2, \sigma_w^2$, respectively. Assume a_1, a_2, a_3 , σ_w^2 and σ_{vi}^2 are unknown. The aim is to obtain the self-tuning fused Kalman signal smoother $\hat{s}_0^s(t-1 \mid t)$.

The system (29) and (30) have the state space model

$$x(t+1) = \Phi(\theta)x(t) + \Gamma w(t) \tag{32}$$

$$y_i(t) = Hx(t) + v_i(t), \quad i = 1,2,\cdots,L \tag{33}$$

$$s(t) = Hx(t) \tag{34}$$

with the companion form

$$\Phi = \begin{bmatrix} -a_1 & & \\ -a_2 & I_{(n-1)} & \\ \vdots & & \\ -a_n & 0 \quad \cdots \quad 0 & \end{bmatrix}, \Gamma^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} \; H = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} \tag{35}$$

where Φ contains the unknown parameter vector $\theta = [a_1, a_2, \cdots, a_n]^{\mathrm{T}}$. Using the above method, the optimal and self-tuning fused Kalman signal smoothers $\hat{s}_0(t-1 \mid t)$ and $\hat{s}_0^s(t-1 \mid t)$ are given by

$$\hat{s}_0(t-1 \mid t) = H\hat{x}_0(t-1 \mid t), \; \hat{s}_0^s(t-1 \mid t) = H\hat{x}_0^s(t-1 \mid t) \tag{36}$$

where $\hat{x}_0(t-1 \mid t)$ and $\hat{x}_0^s(t-1 \mid t)$ are the optimal and self-tuning fused Kalman smoothers for the system (29) and (30), respectively.

In simulation we take that

$$\begin{gathered} A(q^{-1}) = (1+0.6q^{-1})(1+0.7q^{-1})(1-0.8q^{-1}) = 1+0.5q^{-1}-0.62q^{-1}-0.336q^{-3} \\ \sigma_w^2 = 1.5 \quad \sigma_{v1}^2 = 0.2 \quad \sigma_{v2}^2 = 0.4 \quad \sigma_{v3}^2 = 0.6 \end{gathered} \tag{37}$$

Applying the RIV algorithm, the estimate $\hat{a}_j$ of a_j $(j=1,2,3)$ can be obtained. Based on the identification of model parameters, and based on the correlation method, the on-line estimators of the noise variances are obtained. The convergence of estimators of AR parameters and noise variances is shown in Fig.1-Fig.3, Fig.1-Fig.3 show the consistence of estimators of the model parameters and noise variances, where the straight curves denote the true values, the solid curves denote the estimators. The error curve between the self-tuning and optimal fused Kalman signal smoothers is presented in Fig.4, where we see that the self-tuning fuser converges to the optimal fuser, i.e. its error converges to zero. Therefore, the self-tuning fused Kalman signal smoother has asymptotic optimality.

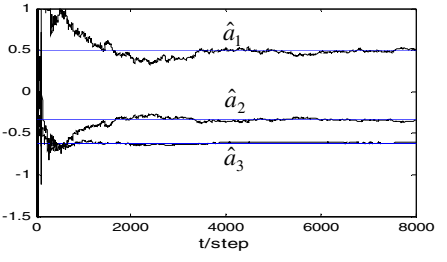


Fig. 1. The convergence of estimates of a_i

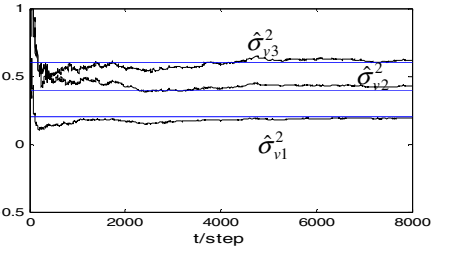


Fig. 2. The convergence of estimates of σ^2_{vi}

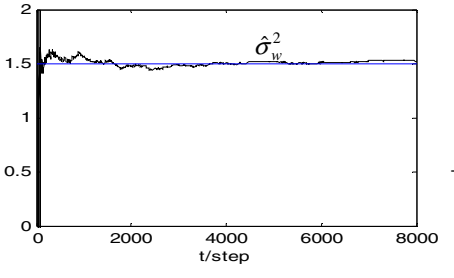


Fig. 3. The convergence of $\hat{\sigma}^2_w$

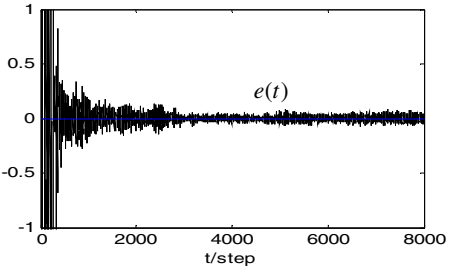


Fig. 4. The error curves $e(t)=\hat{s}^s_0(t-1|t)-\hat{s}_0(t-1|t)$

5 Conclusions

For the multisensor system with unknown model parameters and noise variances, a self-tuning information fusion Kalman smoother weighted by scalars for components has been presented by the classical Kalman filter method. By the system identification algorithm and the correlation function method, the estimators of the model parameters and noise variances are obtained. The proposed self-tuning fused Kalman smoother converges to the optimal fused Kalman smoother in a realization, so that it has the asymptotic optimality.

Acknowledgements. This work is supported by National Natural Science Foundation of China under NSFC-60874063, Automatic Control Key Laboratory of Heilongjiang University, and Science and Technology Research Foundation of Heilongjiang Education Department under Grant 11553101.

References

1. Bar-shalom, Y., Li, X.R.: Multitarget-Multisensor Tracking:Principles and Techniques. YBS public, Storrs (1995)
2. Sun, S.L.: Distributed optimal component fusion weighted by scalars for fixed-lag Kalman smoother. Automatica 41, 2153–2159 (2005)
3. Deng, Z.L.: Self-tuning filtering Theory with applications. Harbin Institute of Technology Press, Harbin (2003)
4. Deng, Z.L., Gao, Y., Li, C.B., Hao, G.: Self-tuning decoupled information fusion Wiener state component filters and their convergence. Automatic 44, 685–695 (2008)
5. Deng, Z.L., Li, C.B.: Self-tuning information fusion Kalman predictor weighted by diagonal matrices and its convergence analysis. Acta Automatic Sinica 33(2), 156–163 (2007)
6. Sun, S.L.: Optimal and self-tuning information fusion Kalman multi-step predictor. IEEE Trans. Aerospace and Electronic Systems 43(2), 418–427 (2007)
7. Sun, X.J., Zhang, P., Deng, Z.L.: Self-tuning decoupled fusion Kalman filter based on Riccati equation. Frontiers of Electronic and Electronic Engineering in Chinese 3(4), 459–464 (2008)
8. Deng, Z.L., Li, C.B.: Self-tuning information fusion Kalman smoother. Control Theory and Application 24(2), 236–248 (2007)
9. Deng, Z.L.: Information Fusion Filter Theory with Applications. Harbin Institute of Technology Press, Harbin (2007)
10. Ljung, L.: System Identification: Theory for the User, 2nd edn. Tsinghua University Press, Prentice Hall, Beijing, China (1999)
11. Gao, Y., Wang, W.L., Deng, Z.L.: Information fusion estimation of noise statistics for multisensor systems. In: The 21th Chinese Control and Decision Conference, Guilin, China, pp. 1127–1131 (2009)
12. Gao, Y., Xu, H.Q., Deng, Z.L., et al.: An information fusion identification method for model parameters and noise statistics of multisensor systems. Science Technology and Engineering 9(17), 4896–4900 (2009)
13. Gu, L., Sun, X.J., Deng, Z.L.: The convergence analysis of the self-tuning Riccati equation. In: The 21th Chinese Control and Decision Conference, Guilin, China, pp. 1165–1170.
14. Ran, C.J., Tao, G.L., Liu, J.F., Deng, Z.L.: Self-tuning decoupled fusion Kalman predictor and its convergence analysis. IEEE Sensors Journal 9(12), 2024–2032 (2009)

The Formation of Virtual Backbone Network in MANETs Based on Cognitive Radio

Feng Wen-Jiang, Wu Di, Jiang Wei-Heng, and Jia Nian-Long

College of communications Engineering, Chongqing University, Chongqing, China
fengwj@cqu.edu.cn, tonywoo1983@yahoo.com.cn

Abstract. In this paper, we propose a new algorithm adopting the spectrum sensing function of cognitive nodes, selections of cluster-heads and gateways can be made on the basis of dynamic information. Taking the cluster-head as anchor point, by choosing gateway nodes that are less affected by primary user in quest of connected dominating set, a virtual backbone network can be established. Based on the comparison of weights, through the sensing of the location of the primary user, the spatial distribution area of the cognitive nodes are divided and assigned different weights, thus realizing partition processing in the network. On one hand, it maintains relative stability of the sub-clusters in motion; on the other hand, it reduces the probability of interference links. In the end, and simulation results confirm the verification of the algorithm.

Keywords: Cognitive radio; Virtual backbone network; Topology; Clustering.

1 Introduction

Mobile ad hoc wireless networks (MANET) is a wireless communication systems which deploying in small domain without fixed infrastructure support. Traditionally, the network architecture are classified as the planar structure and the hierarchy structure. Clustering is considered to the core technology to construct the management and transmission layer for the hierarchical networks. Although the tentative plan is absorbing, it's the famous NP-hard problem in Graph Theory[1]. Nowadays, the virtual backbone networks construction methods can be divided into two groups, one is the centralized type, and another one is the distributed type. Zhao[2] design a distributed approximation algorithm to calculate the connected dominating set by choosing the one-hop gateways in the dense ad hoc networks. HAN[3] proposed a zone-based distributed algorithm for CDS formation. IAN[4] puts forward the system architecture of the wireless networks being integrated cognitive radio technologies. The study about topology management is few now, in accordance with the new features of cognitive radio ad hoc network.

The MANET networks integrated with cognitive radio(CR) support the capacity of temporarily borrowing the Primer User(PU)'s authorized spectrums as the additional communication bandwidth while the PU is off. But the system is demanded to switch to another available band. The spectrum handoff will result in some delay, because the CR user need to configure the parameters of the transmitter such as the operating

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 451–458.
springerlink.com

frequency and the modulation mode, to hold the current connections, and even to trigger the process of the routing refresh. In this paper, a novel method is proposed to building the topology with lower PU interferences and relative stability in motion.

2 Related Convention and Mathematical Model

Dominating Set. In graph theory, a dominating set for a graph G is a subset D of V such that every vertex not in D is joined to at least one member of D by some edge. If every proper subset of dominating set D is not the dominating set, D is said to be minimum dominating set.

Independent Set. An independent set is a set of vertices in a graph, no two of which are adjacent. A maximal independent set is an independent set such that adding any other vertex to the set forces the set to contain an edge. A maximum independent set is a largest independent set for a given graph G and its size is denoted $\alpha(G)$.

Unit Disk Graph. In the simple undirected graph G , two vertices u and v are connected by an edge if and only if the distance between u and v is less than the definite value. The graph G is said to be Unit Disk Graph(UDG).

Unit Disk Graph is used in computer science to model the topology of ad-hoc wireless communication networks, among which, vertex set represents the nodes and edge set represents the links. And the area within which a signal from one node can be received by another node is modeled as the circle.All above, the formation of virtual backbone network can be abstracted to the calculating approximated solution to the connected dominating set in the unit disk graph.

3 Formation of Virtual Backbone Network

3.1 Calculation of the Weight

Taking the cluster-head as anchor point, by choosing gateway nodes that are less affected by primary user in quest of connected dominating set, a virtual backbone network can be established. The cognitive user gain the capacity, based on the PU location technology of cognitive radio[5], to estimate that PU in the shadow of it's omni-directional antenna. Furthermore, the node judge whether it should cause the potential interference. Based on this, the vertex set V is divided into Vi , the set of nodes locate in the potential interference zone and Vn, the complementary set of Vi, V= Vi +Vn. For the nodes locate in the potential interference zone, select the nodes which one potentially bring less interference to play the role of cluster head and the gateway. For the nodes locate out of the potential interference zone, select the nodes which one taking on more stability to play the role of cluster head and gateway.For the sake of providing a easy solution based on open standards for the practice of weight comparison of the nodes in or out the potential interference zones, it is necessary that make the weights meet the totally ordering relation. We define the mathematic expression of the weight calculation as follows. In the cluster head electing and clustering stage:

$$w_C = \begin{cases} f_{node}(v_x) & v_x \in V_n \\ \text{interf}_n(v_x) & v_x \in V_i \end{cases} . \quad (1)$$

In the gateway choosing stage:

$$w_G = \begin{cases} f_{gw}(g) & \forall v \in \{g\}, v \in V_n \\ \text{interf}_{gw}(g) & \exists v \in \{g\}, v \in V_i \end{cases} . \quad (2)$$

Where vch is the cluster head, g is the gateway. fnode and the fgw are the weight calculation formulas, the function interfn and the function interfgw are used to estimate the interference degree of the node and gateway with PU, respectively.The function f_{node} designed in this paper refers to the cluster availability, which is determined by the stability of every links between the current node and its neighbors in MANET.

$$f_{node} = \sum_{i=0}^{k-1} I_i . \quad (3)$$

Links Availability I_i is defined as follows:

$$I_i = \begin{cases} \left(1 - \frac{R - S_{m,n}}{\Delta t \times 2 \times v_{\max}}\right) \times \left(1 - \frac{\Delta S_{m,n}}{2 \times v_{\max} \times \Delta \tau}\right) & \Delta S_{m,n} > 0 \\ 1 & \Delta S_{m,n} \le 0 \end{cases} . \quad (4)$$

Where the Sm,n is the distance between the node m and n; ΔSm,n is the distance increment, ΔSm,n<0 denote the pair nodes keep the co-rotating motion, while the ΔSm,n >0 means them move away from each other; vmax is the maximum speed; Δt is the adjacent time interval between the changes of speed, supposing that speed and direction change only at certain point of time, and remain unchanged during a fixed period of time. The movement in time interval Δt can be predicted by the movement in Δτ, which can be estimated by the signal intensity measured in the points T and T+Δ τ . I_i is the Links Availability of current link, which is determined by both the prediction results of relative motion and the actual distance of the source and destination. Links Availability $I_i \in [0,1]$, the bigger the I_i is, the more stable the link is. The function *interf*$_n$ in (1) is defined as follows, indicating the interference degree.

$$\text{interf}_n = 1 / dist(v_{x,} v_{PU}) . \quad (5)$$

In terms of the Gateway, the functions fgw and interfgw in (2) is designed as follows, where the V2hop is the 2 hops gateway set and the V3hop is the 3 hops gateway set, the min refers to the minimum value, the dist indicate the distance between the current pair of nodes.

$$f_{gw} = \begin{cases} dist(g_x, v_{chs}) + dist(g_x, v_{chd}) & g_x \in V_{2hop} \\ \min(dist(g_i, v_{chs}) + dist(g_j, v_{chs})) + \min(dist(g_i, v_{chd}) + dist(g_j, v_{chd})) & g_x \in V_{3hop} \end{cases} \quad (6)$$

$$f_{gw} = \begin{cases} 1/dist(g_{x,}v_{PU}) & g_x \in V_{2hop} \\ 1/dist(g_{i,}v_{PU}) + 1/dist(g_{j,}v_{PU}) & g_i, g_j \in V_{3hop} \end{cases}. \quad (7)$$

3.2 Algorithm Design

The formation of virtual backbone network is the method based the mode that constructing the connected dominating set in the basis of the maximum independent domination. The method is divided into two phases: first to solve the maximum independent domination, second to insert few nodes to connect the independent points, and to make the independent domination is connected.

The clustering algorithm using staining patterns can be seen as the solution to the maximum independent domination. The nodes in the network are divided into three states, the cluster head, the cluster member, and the undecided state. Solution maximal independent set of algorithm flow is as follows.

Algorithm 1. Solution maximal independent set

1: Cluster heads extract the gateway status from the local topology information;

2: The node under the primary user location information to determine if they located a potential interference area, and calculates the weight of its own for cluster head according to cluster stability or the interference degree, then broadcast the weight to all its neighbors;

3: After get all neighbors weight, the node compare them with its own, if the current node hold the greatest one (In our design, the greater the weight is, the higher the priority is.), it upgrades to the cluster head state from the undecided state, and request the neighbors to be the cluster members in the cluster;

4: For the node at undecided state and get requests, it chooses to join the cluster dominated by the cluster head with highest priority. Then, it broadcasts the state information;

5: Repeat 3 and 4, till none of the nodes are marked as the initial state.

Connected dominating set construction algorithm flow is as follows.

Algorithm 2. Connected dominating set construction

1: In the initial moment, all nodes are in the undecided state, and periodically send control messages with the basic information to discover the local topology situation;

2: Cluster heads sort the candidate gateways in accordance with the principle that the gateway located out the potential interference area is more prior than the one in and the 2 hop gateway is more prior than the 3 hop gateway. For the gateways in same type, compare the weight described above;

3: Cluster heads chooses the one with highest priority to mark as the gateway. The process is over, till every cluster head have and only have a gateway to the adjoining cluster head.

3.3 Proof of Correctness of Algorithm

Lemma 1. Eventually the algorithm terminates.

Proof. In the beginning, any node in the network is in the undecided state. The Γ_i refers to the nodes in the undecided state set. Since every node can determine its cluster, the Γ_i set will eventually become empty. Thus, the clustering phase will terminate. After clustering phase, it turns into the gateway selection phase. The cluster head with the lower ID of adjacent clusters should choose the optimal candidate gateway to mark as the gateway and notify the other of the result. The total number of nodes in the network is given, so the sum of cluster heads is limited. There will be one and only gateway between every adjacent clusters pair. The Γ_j refers to the set of the adjacent clusters pair which have not chosen the gateway. Γ_j set will be empty. Thus, the gateway selection phase will terminate. The proposed algorithm will converge.

Lemma 2. The algorithm obtains a connected dominating set under the connected graph condition

Proof. Set I is a maximum independent set, which denoting the cluster head set. I is a dominating set. Any vertex $v \in V\text{-}I$, v is adjacent to one vertex u, $u \in I$. Assume that G is a simple connected graph, any vertex $v \in I$, there is a vertex u, they are reachable for each other within 3 hops. In the proposed gateway selecting, every adjacent cluster pair will be inserted a gateway. So, if the vertex v and $u \in I$, they are reachable for each other within 3 hops, there will be a gateway to connect them. The induced graph $G(V',E')$ of $G(V,E)$ is connected, set $V'=I+S_g$. It means that the reachability will not change after the algorithm execution. The set $S_{cds}= I+S_g$ is a connected dominating set.

4 Performance Simulation and Results Analysis

In our simulation, N mobile nodes were randomly and uniformly distributed in a plane area of X×Y, the nodes could move in all possible directions with displacement varying uniformly within[0,Vmax], and the radius of antenna coverage is R. The simulation parameters have been listed in Table 1. The running time is 200s.

Table 1. Simulation Parameter

Parameter	Meaning	Value
N	Number of nodes	50
$X \times Y$	simulation area	100×100(m)
V_{max}	Max speed of the node motion	2-18 (m/s)
R	radius of antenna coverage	10-60(m)
Number of PU_	Number of Primer user	1-5

The number of potential interfering links. If one or both nodes the current link connecting is in the potential interference area, the link is considered to be cause a potential interference to the PU. A potential Interference link means that the spectrum has to hand off as soon as the primary user is working.

Re-affiliations. it refers to the sum of the nodes whose logical relationship to its neighbor change caused by the relative motion per unit time.

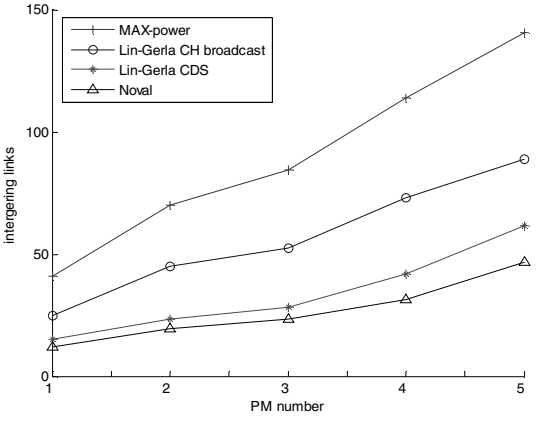


Fig. 1. Number of Interfering links number VS number of PU

Fig. 1 shows the variation of the number of interfering links with respect to the amount of PU in the different topology formation algorithms. From top to bottom they are: the max power topology, Lin-Gerla algorithm[6] with redundant gateway, Lin-Gerla algorithm with choosing the only gateway according to the lowest id principle[7], and the novel algorithm. As we know, the lower the amount of links is, the lower the number of interfering links is. The amount of interfering links of max power topology is the highest; the amount of the topology of Lin-Gerla algorithm with redundant gateway is lower than the max power topology because the links between the nods in the same cluster is wiped off. The cluster head election and gateway selection mechanism, which tends to select the nodes with high stability and far away from PU position as the member in virtual backbone network, of proposed algorithm, ensures the number of interfering links is the smaller than the Lin-Gerla algorithm.

Fig. 2 shows the variation of Re-affiliations with respect to the max speed of motion. The novel algorithm, realizing partition processing in the network, can produce a stable cluster out the potential interference area, so that the re-affiliations per unit time are lower than Lin-Gerla.

In the Fig. 3, the triangle indicates the PU, the circle indicates the cluster head, the cross-sing indicates the gateway, and the dot is the common node. The figure shows the novel algorithm maintain the connectivity of the generated topology, the structure is more sparse, and the nodes less play the role of cluster head and the gateway, in the vicinity of the PUs of cognitive. As the structure of the virtual backbone network is mesh-like, it can support the network layer to select the right path.

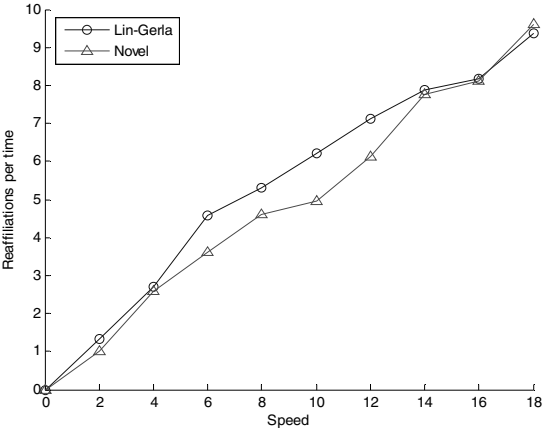


Fig. 2. Re-affiliations per unit time VS speed

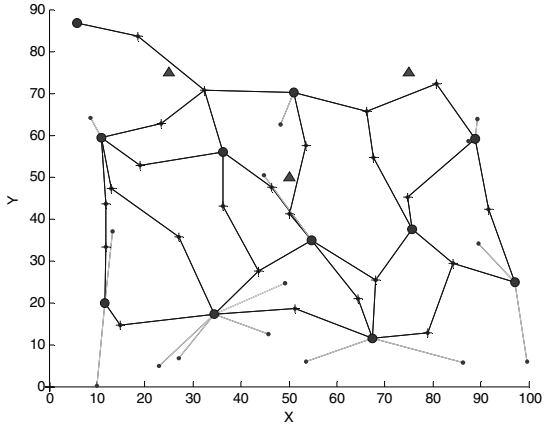


Fig. 3. Topology graph of the novel algorithm

5 Conclusion

This paper focuses on topology management based on clustering in cognitive MANET, combines PU location and spectrum sensing of cognitive radio and MENAT network topology management. Subject to cognitive MANET, by spectrum sensing and cognitive node positioning, a virtual backbone construction method is proposed, which is based on the space and time information of the primary user's working status. Virtual backbone construction includes cluster head election and gateway selection. Cluster heads are taken as anchor points, the connected dominating sets are obtained by selecting gateway nodes that are less affected by the primary user, and then the virtual backbone can be established. The results show that it can maintains

the stability while reducing the interference of the PU. We simulated and verified the proposed algorithm, and analyzed the relative results.

Acknowledgments. This work is supported by the National Natural Science Foundation of China (60872038), and the Natural Science Foundation Project of CQ CSTC (CSTC,2009BA2064).

References

1. Karaki, J.A., Kamal, A.E.: Efficient virtual-backbone routing in mobile ad hoc networks. Computer Networks 52(2), 327–350 (2008)
2. Zhao, C.X., Wang, G.X.: Gateway Selection Strategy in Dense Manet. Chinese Journal of Computers 28(2), 195–200 (2005)
3. Han, B.O.: Zone-based virtual backbone formation in wireless ad hoc networks. Ad Hoc Networks 7, 183–200 (2009)
4. Akyildiz, I.F., Lee, W.Y., Chowdhury, K.R.: Cognitive radio ad hoc networks. Ad Hoc Networks 7(5), 810–836 (2009)
5. Ma, Z.Y., Chen, W., Cao, Z.G.: Analysis on Detection Probability Based Primary User Localization Algorithm in Cognitive Radio Networks. Journal of Beijing University of Posts and Telecommunications 32(2), 14–19 (2009)
6. Lin, R., Gerla, M.: Adaptive Clustering for Mobile. Wireless Networks 15(7), 1265–1275 (1997)
7. Stojmenovic, I., Seddigh, M., Zunic, J.: Dominating sets and neighbor elimination-based broadcasting algorithms in wireless networks. IEEE Transactions on Parallel and Distributed Systems 13(1), 14–25 (2002)

Development of the Nerve-Central Listen System Based on Component

Xu Dahua

College of Engineering, Nanjing Agricultural University,
Nanjing 210031, China
xudahua@njau.edu.cn

Abstract. Researching on component, a nerve-central listen system model was designed based on the HIS (Hospital Information System). Data collection, transmission and retrieval in the listen system was described. By comparison of video coding technology, selected a suitable coding technology for this system, and discussed the yawp removal technology to video data stream in the system.

Keywords: Component; Topology architecture; video collection; yawp.

1 Introduction

Nerve-central listen system is a set of software and hardware systems, which is used to achieve the patient's physiological parameters for real-time monitoring and centralized management. Through the application of the system, patients' life and physical parameters can be real-time viewed and took timely first-aid on critically ill patients, greatly reduce the intensity of nurses, improve rapid response capability on heavy patient condition monitoring. So the application prospect is very wide. The system has been put into use in many hospitals' care units and worked very well.

Component is a combination of a group of packages, and completes one or more functions on behalf of specific services, but also provides users with multiple interfaces. At different levels, the components can combine the underlying logic into a larger particle size of high levels of new components, or even directly package into a system that allows reuse of modules from the code-level, object level, schema level to the system level are possible, so that the same software as the hardware can assemble custom made submissive dream a reality. Using component-based system model can greatly reduce the system development cycle and improve code reuse and extensibility.

The nerve-central listen system, as a technology project and business application project, has a wide theory and clinical application. The project will ultimately achieve in use in several hospitals. As different requirements required by hospitals, in order to enhance the system's flexibility and scalability, using a component-based development methodology can enhance the system scalability and creativity, and improve the theoretical research value and business value of the system.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 459–464.
springerlink.com

2 Component Technology

Component technology promoted the content of object encapsulates, it focused on the harmonious relations of complex components and stressed the existence of physical form in the environment. It has four basic attributes:

(1) Component is independently configurable elements, the component must be self-tolerance.

(2) Component stresses separation of the environment and other components, so component is the strictly packaged. The external has no chance or no necessary to know the internal implementation details.

(3) Components can be complex used in an appropriate environment. The components need to provide a clear interface specification that can interact with the environment.

(4) Component should not be continued (i.e. component has no individual-specific properties), it should not be distincted with its own copy, and only one copy at most in any environment. Easy to see, the components follow the package characteristics of the object, but not limite to an object, which can be encapsulated within one or more classes, the prototype object or process, the structure is flexible. Component highlights the characteristics of tolerance and self-tolerance, that is the necessary features as a part of software product line.

Viewed from the abstract level, object-oriented technology has reached a class-level reuse (code reuse), It was packaged in units of class. Such reuse granularity was too light to solve the heterogeneous interoperability and efficient reuse. Component would be referred to a higher abstractive level. Component is a combination of a group of packages, and completes one or more functions on behalf of specific services, but also provides users with multiple interfaces. The components hide the concrete implementation, only provide services by interfaces. Thus, at different levels, the components can combined the underlying logic into a larger particle size of high levels of new components, or even directly packaged into a system that allows reuse of modules from the code-level, object level, schema level to the system level are possible, so that the same software as the hardware can assemble custom made submissive dream a reality.

3 System Architecture

The system consists of the bedside machine for distributed monitoring and the central machine for centralized management. The central machine and the bedside machine are connected by the corresponding data lines to transfer data in two-way. The central listen system software partly includes the central machine's control software and the bedside machine's function software. in addition to the physical hardware connection to central listen system, but also the coordination of software. The central machine's control software and the bedside machine's function software each has a communication software module responsible for the connection to ensure two-way data transmission. Using RS485 to transmiss data is simple, easy and reliable, so we use serial port for data communication in central listen system.

The central machine and bedside machine were connected with RS485 cable through their respective configuration of serial communication ports to achieve a physical hardware connection. Using RS485 bus, as data transfer link, achieves half-duplex asynchronous communication. On the one hand, RS485 standard, as multi-point, differential data transmission of electrical specifications have become one of the most widely used in standard communication interface. This communication interface allowed simple multi-point, two-way communication on twisted pair, each terminal just hanging in the bus through an interface, this could achieve the real multi-point bus architecture. Its yawp rejection capability, and data transfer rate, cable length and reliability were unmatched by other standards. On the other hand, RS485 standard only made regulations upon interface electrical characteristics, rather than connector, cable or agreement, users could create their own communication protocol on the basis.

As the system is complex and reliability demanding, so we use component-based intelligent control technology, shown as in Figure 1 the central nervous control system topology architecture, effectively improve reliability and stability of the system.

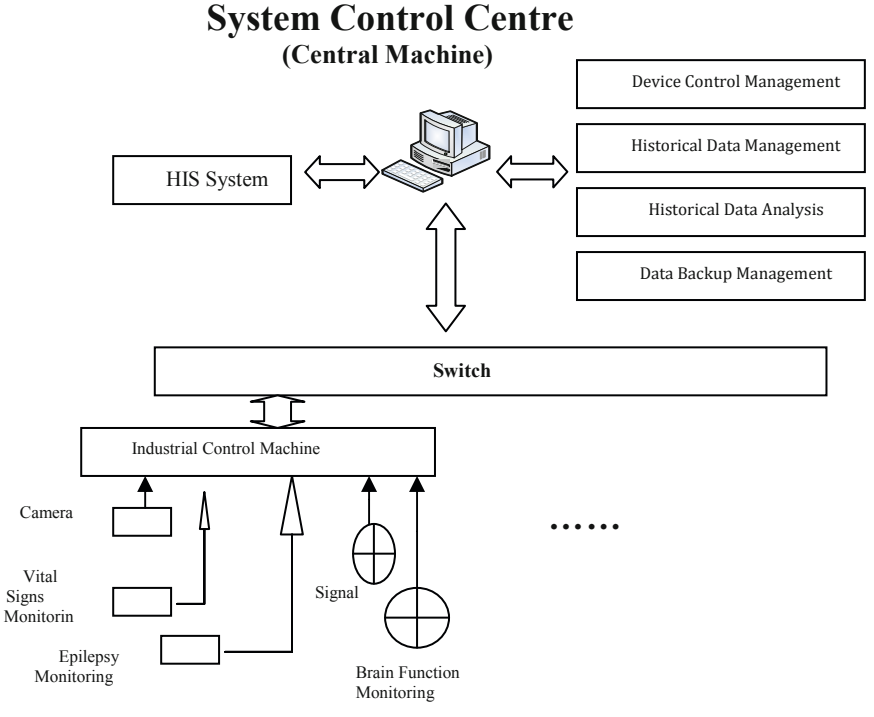


Fig. 1. System Topology Architecture

4 Bedside Machine Design

Using industrial single board computers, communications components completed the calling provided by the MOXA serial port application development libraries, the

embedded process of bedside machine was designed with WINCE C language. The central processing components of bedside machine used interrupts to receive control commands sent by the central machine, and forwarded the command to the data acquisition components. When the central processing components organized data into a predetermined data packet. And then sent to the central machine through the serial port. When the system started to initialize the serial port (designated port communication parameters and set the interrupt service routines), and then waited for the disruption, control commands were received and processed by the interruption service components.

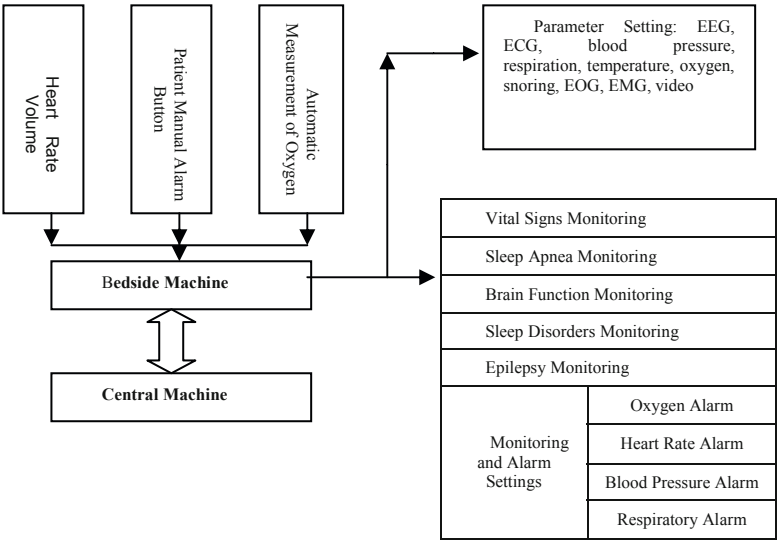


Fig. 2. Schematic of work of bedside machine

In this system, many of the data involved. In all data to be collected and transferred, video data transmission was more complex. The key coding techniques were MPEG-4 and H.264 compression technology, as H.264 encoding made the video compression decompression at less than the rate of 28.8Kbps ,so it could real-time mobile video capture, compression, decompression and playback. H.264 encode used unique P frames and B frames adaptive compression technology to further enhance the compression algorithm of the compression ratio, with IP multicast capabilities, greatly reduced the real-time transmission network bandwidth. When multiple control points work simultaneously, which caused transmission bit stream of data at the same time. The average bandwidth less than 250kbps, greatly reduced the input for user's network resource. The video streams of the system used H.264 and AVI data file was stored in the system, all the data files saved in binary form, not only guaranteed the data transfer speed, but also effectively reduced the amount of data acquisition systems. System video processing core code was as follows:

```
AVIFileInit();     // AVIFILE Library of Initialized system
AVIFILE  af ;    // File pointer
HRESULT hr;
Hr=AVIFileOpen(&af,LPCTSTR(filename),OF_CREATE,NULL);
AVISTREAMINFO sthd;    // Video data stream structure
AVISTREAM  as;         // Video data stream interface
SetRect(&sthd.reFrame,0,0,biout.biWidth,biout.biHeight);
```

The video stream transmitted to the central machine through data lines, the central machine stored and analyzed the collection of data. The transmission process of video streams completed by a secondary thread and the secondary thread obeyed the RTP protocol, combined with asynchronous transfer mode and multi-buffer the video data stream real-time transmission, and thus to be a good solution to the local broadcast and network transmission time difference.

The hospital environment was complex, the acquired video images existed yawp. In order not to affect system monitoring performance, we have developed an "Edge Enhancement Algorithm Components" in this system. Through this component to enhance images, the specific implementation was divided into the following two steps:

(1) By morphological filtering the images with the Y component of YUV system, first needs to convert RGB image to YUV to obtain the image intensity values Y component. Then mathematical morphological filtered the Y component.

(2) After morphological filtering on the Y component for median filtering and mathematical morphology filtering for Y component, it could eliminate a portion of the yawp and make the details of the images coming together. Then using the median filter method could eliminate the remaining noise in the image, make image smooth and loss of image detail was not serious. In order to preserve the details, use 3×3 two-dimensional sliding window when making median filter.

5 Central Machine

The development of the central machine is the core of the whole system, more functions. According to parameters and the functions realized, we first used UML to model and gradually refining the system's functions, made the system features modular, component-based, and finally got the system fully functional diagram and software component framework, and developed a component library management system. In the development process, reserved an interface for the existing HIS, so the system has good usability. When doctors, nurses used the system to access patient-related information, it could have direct access to the default modules, without the need go to find information such as patient records, this could save valuable time for first-aid and treatment for the patient and greatly improve the efficiency of the treatments.

After designing topology architecture, based on the needs of the system, we designed a number of common components and a component library management system. This was available to project team members to use, not only improves efficiency, but also ensured uniform standards for each functional module control interface. More

importantly, it was convenient for the system development process. As demand changed, upgraded the system conveniently. We could achieve: single functional, common, simple interface, clear, highly independent. The component library management system was completed for the following functions: find components, component expansion, component integration, component removed, component changes.

"Patient Management Component" first entered the patient's basic information, selected the monitoring parameters, set the physiological parameter. Because of the reserved the HIS interface, we could enter the patient's hospital number to transfer the patient's basic information directly from the HIS system, and save the patient's physical signs monitoring parameters into the HIS system.

After setting physiological monitoring parameters, the real-time monitoring component started to work, where using multi-way monitoring and real time monitoring technique. Interface included digital display monitor, video display, parameter setting areas. Digital display and video image display area were floating windows, position could be adjusted and moved, without affecting the waveform display. Through setting the filter parameters, you could set the system state to monitor or record. On the monitoring screen, it could display from 4, up to 16 bedside machines' all parameters. Each bedside machine could make video stream freeze, zoom and other operations, the system would real-time alarm for exceptional data, and display a specific exception information or the patient's bed number.

6 Summary

The system was developed with technical innovations. The main innovative technology has the following two point: 1, Using the unique streaming technology, achieved real-time monitoring of unlimited number of remote sites and can view all the monitoring data; 2, Using standard networking technologies to support IP unicast, multicast functionality, It may form point to point monitoring system, and also form a central monitoring system for each point; Setting a number of monitoring stations on the network to meet the needs of multiple users, and also formed multi-class monitoring system in WEB. With component design, system functions can be smoothly upgrade, as well as the interface with other medical systems, such as HIS, CIS and other system interfaces. The development and the successful commercial operation of nerve-central listen system, verified the correctness and feasibility of software development based on component.

References

1. Barbier, F., et al.: Business Component-based software engineering. Kluwer Academic Publishers, Boston Hardbound (2002)
2. Barroso, L.A., Dean, J., Hlzle, U.: Web Search for a Planet:The Google Cluster Architecture (Febraury 2003)

Research of the Strip Shape Control System on Cold Strip Mill

Guimei Cui and Yuxing Zhao

School of Information Engineering, Inner Mongolia University of Science and Technology, 014010 Baotou, China

Abstract. Strip shape is one of important quality indicators in cold strip steel production. The strip shape control system of six roller CVC rolling mill is studied in this paper. According to the measured data, the BP neural network is established, which identifies nonlinear part of the strip shape control system. Research results help a deeper study on the strip shape control system, which has important significance on obtaining good strip shape control.

Keywords: flatness; actuator efficiency; BP neural networks.

1 Introduction

The strip shape control is one of two important control quality indicators in the modern rolling production. Strip shape control directly influence end-product quality and market sales of the strip. Therefore, the development of strip shape control research has far-reaching significance to society and economic. At present, there are few studies on the strip shape control system through consulting a lot of related references. Therefore, this article analyses and researches the strip shape control system of six roller CVC rolling mill.

2 Strip Shape Measurement

According to the definition, the strip shape can be described as the flatness in direction of rolling. The strip shape metering equipment of six roller CVC rolling mills adopts Siemens's BFI flatness measuring roll. There are mainly 5 kinds of actuators: the back-up roll depresses inclined, the intermediate roll bending, the working roll bending, the intermediate roll CVC shifting, and the working roll multi-area cooling system.

The measure principle of strip shape is that different strip elongation results in different tension of discrete points in the horizontal. It uses measuring roll to measure tension and indirectly measured strip flatness. Flatness measuring signal flow chart is as follow in figure 1.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 465–470.
springerlink.com

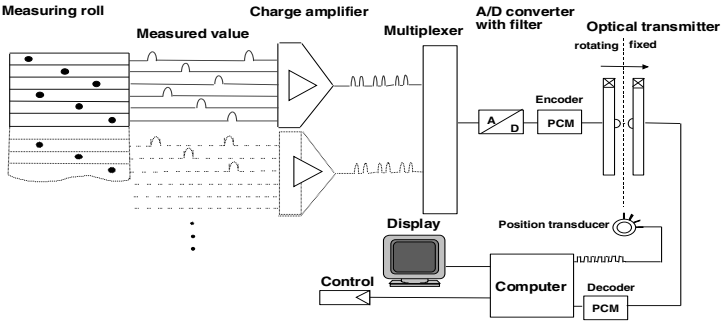


Fig. 1. Flatness measuring signal flow chart

3 Strip Shape Control

The strip shape control system mainly includes two parts: learning efficiency of actuators flatness and actuator movement controller, control structure as shown in Figure 2.

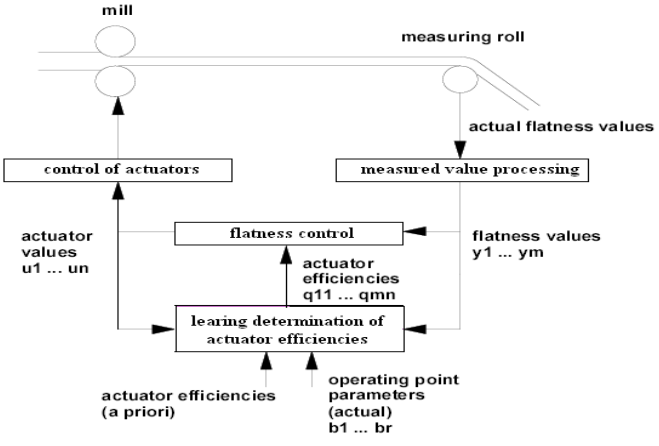


Fig. 2. Flatness control structure

The detailed organization diagram of flatness neural network control as shown in Figure 3,and NNI is the system identifier, Y is flatness actual value; Y_{sr} is flatness setting value; Both Y and Y_{sr} are vectors containing 52 elements; $\Delta\sigma$ is the current flatness deviation; $\Delta\sigma_i$ is a flatness deviation which is determined by recently activated controls; N_σ is flatness deviation which can be improved by the neural network identifier(NNI) of actuator; *P-P* is the prior actuator efficiency; F is the current rolling force; W is the current strip width; *LP* is the actual position deviation of actuator; g and p are efficiency which is in the direction of strap width including flatness deviation weighting factor and actuators; neural computing controller (NCC) is the controller

which is consist of the mathematical method. It sends out the optimal value to the actuator; E_1 is the actual improvement flatness deviation; E is the deviation between output value of network identification and the actual expected value, which is used in training the network.

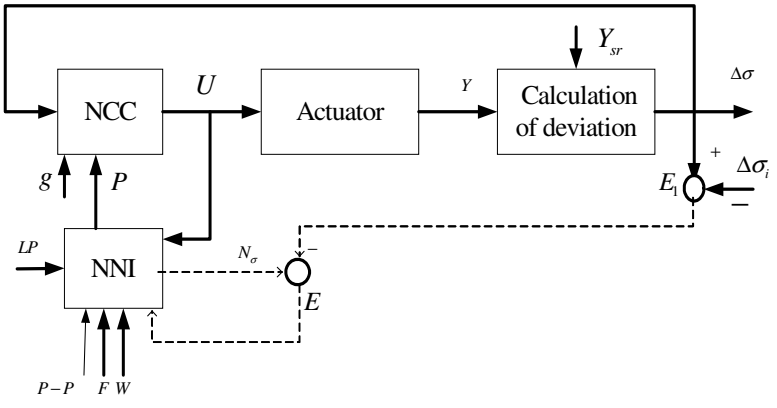


Fig. 3. Flatness neural network control detailed diagram

3.1 NNI Design

3.1.1 Determination of Input and Output

According to Figure 3, the neural network identifier can be achieved by using the BP neural network. According to the impact factor of NNI, the network inputs can be set to 4: the current rolling force F; strap width W, optimal setting of actuator value U, and changes in the actual location of the executing agency LP. According to the function of NNI and the type of settings of secondary computer, the strip shape flatness deviation is used as the network output, which calculated by the direction of strap width of 52 points, so the network outputs are 52 Real data.

3.1.2 Determination of Hidden Layer and Nodal

According to network input and output node number, and the training time of the network, the network is defined as the single hidden layer network structure. The node number is decided using the trial-and-error method, and finally determine the hidden node number are 16.The structure of the neural network identifier which is established as shown in Figure 4.

3.1.3 Network Training

Index function of the training network is:

$$E = \frac{1}{2p}\sum_{p}\sum_{k}(t_{pk} - O_{pk}) \tag{1}$$

Where, p represents the training standard sample number; t_{pk} is network actual output; O_{pk} is network expectation output.

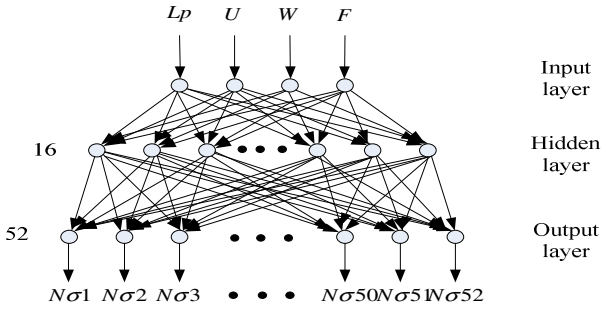


Fig. 4. Neural network identification structure

Use the gradient descent method to cause the minimum of the equation (1), namely using the generalized regular algorithm. The weight revision is:

$$\Delta W_{ji}(n+1) = \eta \cdot \frac{\partial E}{\partial W_{ji}} + \alpha \cdot \Delta W_{ji}(n) \tag{2}$$

Where, $\Delta W_{ji}(n+1)$ is this time weight allowance; $\Delta W_{ji}(n)$ is previous time weight allowance; a is inertia factor, value between [0,1]; η is study factor, value between [0,1].

3.2 Calculation of the Optimum Value of Actuator

NCC is the model, which is used to calculate the optimal settings of the actuator, the specific method is: when minimize the quadratic variance in the formula (3), corresponding U is the optimal settings of the actuator.

$$F = \sum_{i=1}^{n} [g_i(\Delta\sigma_i - p_i(x_i) * U)]^2 \tag{3}$$

Where, x_i is strap coordinate in width direction; n is the node number; g_i is the deviation weight of flatness in position x_i; p_i is the efficiency of actuator in position x_i; $\Delta\sigma_i$ is the flatness deviation in position x_i; U is the optimum value of actuator which is going to be calculated.

NCC is divided into five analysis modules in this project, corresponding to the five neural networks, which are used to calculate the efficiency of actuator. When using these five modules to calculate the latest setting value of the actuator separately, the order should be noted. The order is: the back-up roll depresses inclined, the intermediate roll bending, the working roll bending, the intermediate roll CVC shifting, the working roll multi-area cooling system. Since the priority is different, so the output of the five actuators is different when the flatness deviation of input is different. For the back-up roll depresses inclined, the flatness deviation of input is the deviation between flatness observed value and the two-stage system flatness setting value; As for the other analysis modules which are used to compute the setting value of the other four actuator, the input is the deviation, removing the higher authority actuator corrected value from the smoothness deviation of first-level input (namely multiply efficiency by the most

superior setting value of this actuator). Then sending this differential value into the current actuator analysis functional model, the given quota of the current actuator will be obtained.

In addition, the working roll multi-area cooling system is different from the other actuators after obtained the setting value of actuator. It directly sends into the traditional PI regulator to adjust, while the output value of the other four flatness actuator need to be amplified and limited, then send to the implementing agencies.

3.3 Strip Shape Control System Simulation

The program is edited in MATLAB7.0 Environment in order to simulate the neural network model and mathematical model, and use the measured data to carry on off-line training. The predicted value of each actuator controlling amount is quite close with the display result of the live WinCC picture. Therefore, this model is correct. The WinCC configuration screen shows in figure 5.

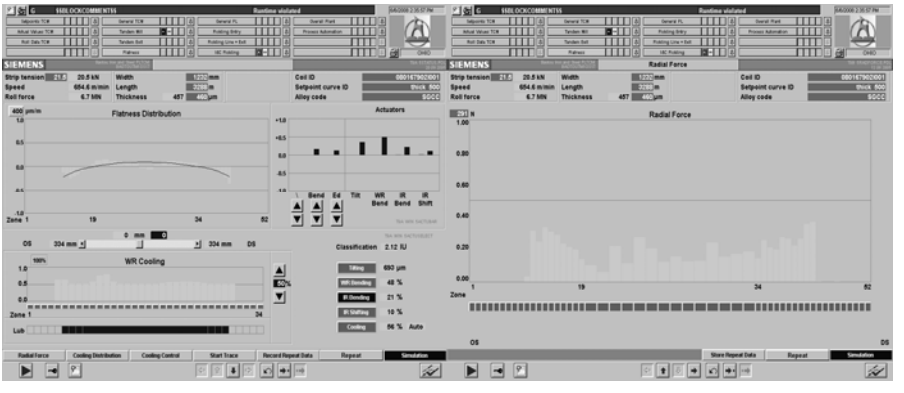


Fig. 5. WinCC configuration screen

4 Conclusion

Using strip shape control system based on the neural network can greatly improve the control precision of strip shape in the rolling scene. Practice proved that the model can well control the flatness and the quality can completely achieve the anticipated goal, when the flatness of the coming strip meets the requirement.

References

1. SIEMENS: New Shape measurement and control system achieve greatly improved performance
2. Peng, Y., Liu, H., Du, R.: A neural network-based shape control system for cold rolling operations. Journal of Materials Processing Technology 202(1-3), 54–60 (2008)

3. Larkiola, J., Myllykoski, P., Korhonen, A.S., Cser, L.: The role of neural networks in the optimisation of rolling processes. Journal of Materials Processing Technology 80-81, 16–23 (1998)
4. Hsu, C.-M.: A New Flatness Pattern Recognition Model Based on Cerebellar Model Articulation Controllers Network. International Journal of Iron and Steel Research 15(5), 32–36 (2008)
5. Jia, C.-y., Shan, X.-y., Liu, H.-m., Niu, Z.-p.: Fuzzy Neural Model for Flatness Pattern Recognition. Journal of Iron and Steel Research, International 15(6), 33–38 (2008)

Research on the Optimal Control of Tube Billet Temperature for Rotary Reheating Furnace

Guimei Cui and Guibao Ding

School of Information Engineering, Inner Mongolia University of Science and Technology, 014010 Baotou, China

Abstract. At present, furnace temperature settings are achieved by the operators in the rotary reheating furnace, and control accuracy is no good. The temperature prediction model of a tube billet surface at the exit based on RBF neural network and the modified particle swarm optimization are put forward to search the optimal steady-state temperatures in this paper. The simulation results indicate that it meets heating process indicators and improve the heating quality and precision of the tube billet. The modified particle swarm optimization solves the local optima problem resulting from population degradation, and improves the search precision.

Keywords: temperature control of rotary reheating furnace; RBF neural network; modified particle swarm optimization.

1 Introduction

In rotary reheating furnace, the temperature of the tube billet should be elevated up to intended temperature before its exit to the mill in order not to damage the rolling mill. Its heating quality, energy consumption and control level directly affect the quality, production and cost. At present, it mainly sets the temperature of each control zone by operators with experience in order to control the entire furnace process. But it is hard to meet billet heating process indicators and minimize energy consumption. Temperature optimal setting is a typical problem of optimal decision. CHAI et al. [1] established a temperature optimal settings model. It made use of the looking up table, a feedback model and the difference between temperature of the billet at the furnace exit and target temperature to calculate a modifier of temperature settings. Kim et al. [2] adopted the principal component analysis to classify the input data. A temperature settings model was established, which consisted of multiple experts neural networks and a door network. However, there are still some problems in these optimization methods. It is difficult to establish an accurate model because of creating a number of assumptions. Many of key factors are not available and so on. In industrial production, it limits the application of temperature optimal settings policy.

The standard particle swarm optimization (PSO) algorithm has been used to search the optimal steady-state temperatures. It is easy to fall in local optima. Therefore, there have been some modified PSO algorithms in recent years, such as intelligent PSO [3],

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 471–477.
springerlink.com

fuzzy PSO [4] and other methods. It ensures the diversity of population, and improves the overall search ability of PSO. In the following pages temperature prediction model and modified PSO algorithm will be proposed.

2 Configuration of Rotary Furnace Optimal Control System

Configuration of rotary furnace optimal control system is adopted in figure 1. Where, g, d, v represent the tube billet type, diameter and rolling speed; charging temperature is room temperature; $T(\boldsymbol{T}_{sc})$ is the temperature predicted value of the tube billet at the exit; T^* is the objective value; E is the difference between the temperature predicted and objective value; $T_F(t)$ is a function of the temperature settings vector $\boldsymbol{T}_{sc}$ in the distribution of time; $\boldsymbol{T}_f$ is a vector consisting of temperature feedback value of each control zone. Temperature controller adopts double-cross limiting PID controlling method.

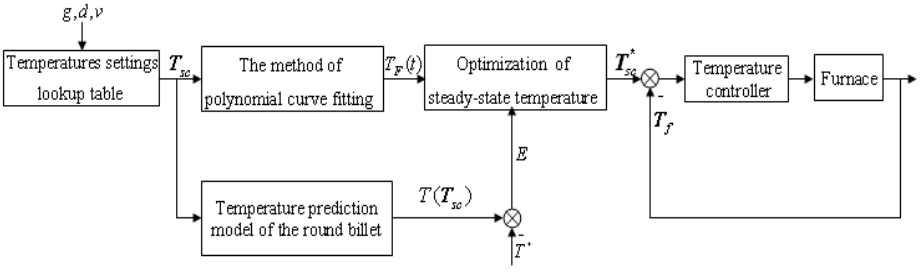


Fig. 1. Configuration of rotary furnace optimal control system

The modified PSO algorithm is used to search the optimal steady-state temperatures settings vector $\boldsymbol{T}_{sc}^*$. According to g, d, v and temperatures settings lookup table with experience, it uses linear interpolation method to obtain the initial temperature settings vector $\boldsymbol{T}_{sc}$, which is used as the particles of initial swarm. It uses the minimum value of the objective function constituted by E and $T_F(t)$ as a standard of searching the optimal steady-state temperatures.

3 Temperature Prediction Model

Heat transfer mechanism needs to create many assumptions, and many key factors are not available. So it is difficult to establish an accurate mathematical model [5], [6]. Therefore, this paper adopts radial basic function (RBF) [7] neural network to establish a temperature prediction model. RBF neural network is three layers forward neural network, which is composed of the input layer, hidden layer and output layer. It has features such as simple structure, short training time as well as nothing to do with the initial weights. Structure of RBF neural network is shown in Figure 2:

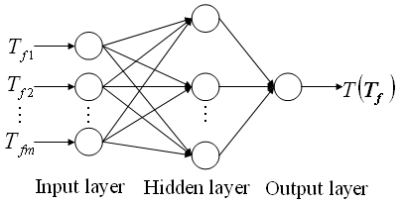


Fig. 2. Structure of RBF neural network

Gaussian function of hidden layer is:

$$u_i = \exp\left[-\frac{\left\|\boldsymbol{T}_f - \boldsymbol{c}_i\right\|}{2\sigma_i^2}\right], i = 1,2,...,q$$

Where, u_i and σ_i are the output and standardized constant of the ith hidden node; $\boldsymbol{T}_f$ is an input sample vector; q are the number of hidden layer nodes; c_i is the center column vector of Gaussian function of the ith hidden node.

The output of RBF neural network is:

$$T(\boldsymbol{T}_f) = \sum_{i=1}^{q} \omega_i u_i - \theta$$

Where, $T(\boldsymbol{T}_f)$ represents output of output layer; ω_i is the weight coefficient from hidden layer to output layer; θ represents threshold of hidden layer.

RBF neural network model can accurately reflect the heat transfer process of the billet. Temperature of the tube billet surface $T(\boldsymbol{T}_f)$ at the exit is used as the only one output. Each zone temperature T_{fm} is the input of the network. Where, $m \in 1,2,\ldots,7$.

4 Optimization of Steady-State Temperature Setting

Each zone steady-state temperature setting is the key to ensure the heating quality of the billet. Billet temperature should reach a specified temperature at the exit, meet the requirements of rolling process, and minimize energy consumption. Thus, the optimal objective function is:

$$J(\boldsymbol{T}_{sc}) = \frac{1}{2} P\left[T(\boldsymbol{T}_{sc}) - T^*\right]^2 + \frac{1}{2} Q \int_0^{t_a} {T_F}^2(t)dt$$

When $\boldsymbol{T}_{sc}$ minimizes optimal objective function, the $\boldsymbol{T}_{sc}$ is equal to $\boldsymbol{T}_{sc}$*. According to practical experiences and process analysis, $T_F(t)$ can be described as a quadratic function of $\boldsymbol{T}_{sc}$ in the distribution of billet heating time. It can be obtained by the method of polynomial curve fitting. Where, P, Q are weighing coefficients; t_a is the heating time of billet.

4.1 Standard PSO Algorithm

The PSO [8] algorithm was first proposed by Eberhart and Kennedy in 1995, inspired by the natural flocking and swarming behavior of birds and insects. Each individual in particle swarm, called a "particle", represents a potential solution which moves its position in search space and updates its velocity according to its own flying experience and that of its neighbor's, aiming for a better position for itself at the next move.

It is assumed below that the swarm consists of m particles. Thus, $i \in 1,\ldots,m$. During each of iterations, every particle in the swarm is updated using equations (1) and (2). Two pseudorandom sequences $\xi \in$ Rand [0,1] and $\eta \in$ Rand [0,1] are used to affect the stochastic nature of the algorithm. For all dimensions$\in 1,\ldots, D$, let x_{iD}, p_{iD}, v_{iD} and p_{gD} be the current position, current personal best position, velocity of the Dth dimension of the ith particle, and global best position, respectively. The velocity update step is according to Eq. (1) as follow:

$$v_{iD}^{k+1} = \omega v_{iD}^{k} + c_1 \xi (p_{iD}^{k} - x_{iD}^{k}) + c_2 \eta (p_{gD}^{k} - x_{iD}^{k}) \tag{1}$$

The new velocity is then added to the current position of the particle to obtain its next position as follow:

$$x_{iD}^{k+1} = x_{iD}^{k} + v_{iD}^{k+1} \tag{2}$$

Where, the acceleration constants $c1$ and $c2$ control how far a particle will move in a single iteration. The inertia weight ω is used to control the convergence behavior of the PSO. In general, the inertia weight ω is set according to the following equation:

$$\omega = \omega_{\max} - \frac{\omega_{\max} - \omega_{\min}}{iter_{\max}} \times k$$

Temperature settings of seven control zones constitute the position of the particle. Velocity of the particle is used to search the optimal settings vector. The reciprocal of optimal objective function plus 0.1 is used as the fitness function. Standard PSO algorithm has a quick convergence rate in the early stages, but it is easy to fall in local optima in the later stages. In this paper, modified PSO algorithm will solve this problem.

4.2 Modified PSO Algorithm of Velocity and Position Mutation

Standard PSO algorithm has a lot of defects. Thus, a modified PSO algorithm is proposed in this paper, and mutation [9] strategy is introduced into it.

The variable object of the modified PSO algorithm is current velocity and position of the particle swarm. The variation can happen during each of iterations and each dimension of the particle, and break up centralized particles. It increases the diversity of population, and avoids the phenomenon of premature convergence. The modified PSO algorithm is as follows. First, initialize the position vector and associated velocity according to the temperatures settings lookup table, and evaluate the fitness of each particle via the fitness function. Velocity and position vector are reinitialized in each dimension on certain mutation rate. Second, compare the particle's fitness evaluation with the particle's previous best solution and population's global best in order to search

population's best solution, which is the optimal steady-state temperatures settings vector. Then, change velocities and positions by using Eq. (1) and (2). Finally, output population's global best solution and corresponding fitness value at the end of the iterations.

5 Verification and Discussion

In a seamless steel tube plant, rotary reheating furnace can be divided into seven control zones: a preheating zone, five heating zones, and a soaking zone. The charging temperature are 20°C, the temperature of the tube billet at the exit should be between 1,200°C and 1,290°C. It requires that the surface temperature error of the predicted values and measured values should be less than 30°C.

5.1 Verification of the RBF Neural Network Model

The steel grade is 20CrMo steel. It selects 50 groups of data as a network training samples, and the other 25 groups of data are used to verify the model. Figure 3 shows that temperature prediction error of the tube billet surface at the exit is less than 30°C, and it meets the requirements of the heating process.

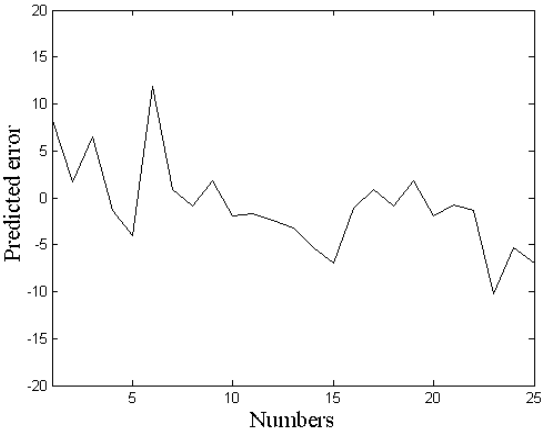


Fig. 3. Predicted error of RBF neural network

5.2 Verification and Discussion of Modified PSO Algorithm

Preferences of modified PSO algorithm is: population sizes are 30; iteration times are 100; $c1$=$c2$=2; mutation rate is 0.5. The target value is 1230°C. For the modified PSO algorithm, the bias is only the 16.197% of manual value, which is much less than the manual operation as shown in Table 1. Obviously, it can meet the tube billet heating process indicators and improve the heating quality.

Table 1. Surface temperature estimation results of a round billet at the exit

Methods	**The difference between the target and the predicted value**
Manual operation	8.15
Modified PSO algorithm	1.32

The modified PSO and the PSO have the same parameters settings. Figure 4 shows that the improvement of the PSO has no significant change after the 13th generation, while, in modified POS algorithm, it can be improved to 29 generations. It can be seen that, the modified PSO can solve the local optima problem caused by population degradation of the standard PSO algorithm, and improve the search precision.

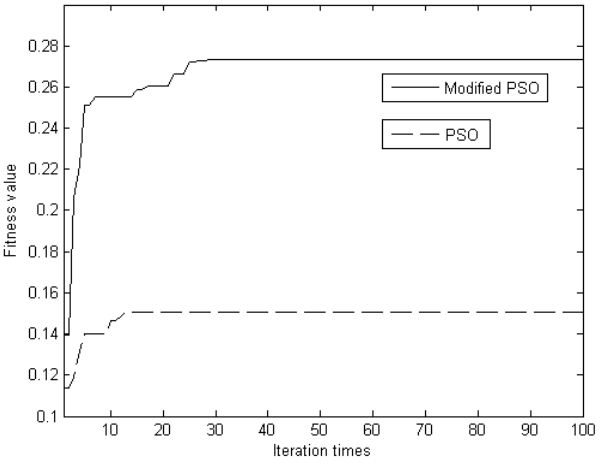


Fig. 4. Evolutionary curves of modified PSO and PSO

6 Conclusion

The RBF neural network model is established in this article, and the modified PSO algorithm is used to search $\boldsymbol{T_{sc}}$*.The simulation indicated that the prediction error satisfies the request of the heating process, and this method is superior to manual operation and improves the heating quality. The modified PSO algorithm breaks up centralized particles through mutating velocity and position, and improves the overall search ability and search precision.

References

1. Chai, T., Wang, Z., Zhang, L.: Optimal Setting model of reheat furnace temperature. J. Acta Automatica Sinica 26(4), 537–541 (2000) (in Chinese)
2. Kim, Y.I., Moon, K.C., Kang, B.S., et al.: Application of neural network to the supervisory control of a reheating furnace in the steel industry. J. Control Engineering Practice (6), 1009–1014 (1998)
3. Ciuprina, G., Ioan, D., Munteanu, I.: Use of intelligent-Particle Swarm Optimization in Electromagnetics. J. IEEE Trans. on Magnetics 38(2), 1037–1040 (2002)
4. Shi, Y., Eberhart, R.C.: Fuzzy Adaptive Particle Swarm Optimization. In: Proc. Congress on Evolutionary Computation, pp. 101–106 (2001)
5. Feng, S., Wen, Z., Su, F., et al.: Study on three-dimensional heat transfer mathematic model for tube rounds in ring shaped furnace. J. Metallurgical Industry Automation 31(5), 26–29 (2007) (in Chinese)
6. Shang, G., Jiao, J., Zhang, S., et al.: Development and application of 3D temperature field mathematic model system for slab reheating furnace. J. Metallurgical Industry Automation 33(2), 38–42 (2009) (in Chinese)
7. Li, G.: Neural-Fuzzy Control Theory and Application. Publishing House of Electronics Industry, Beijing (2009) (in Chinese)
8. Ji, Z., Liao, H., Wu, Q.: Particle Swarm Algorithm and Its Application. Publishing House of Science, Beijing (2009)
9. Li, N., Sun, D., Cen, Y.: Particle Swarm Optimization with Mutation Operator. J. Computer Engineering and Applications 17, 12–14, 15 (2004) (in Chinese)

Research of Traffic Evacuation Methods of Unexpected Events

Fu Wang[1,2], Jie Li[2], and Yong-hui Shi[3]

[1] School of Civil Engineering and Architecture, Wuhan Polytechnic University, Wuhan 430023, China
[2] School of Civil Engineering & Mechanics, Huazhong University of Science & Technology, Wuhan 430074, China
[3] Department of Scientific Research, Wuhan Traffic Management Bureau, Wuhan 430000, China
wangfu@whpu.edu.cn

Abstract. To rescue and evacuate people of unexpected events in city, it is very necessary to organize traffic system. On the base of analyzing traffic characteristics of unexpected events, traffic evacuation methods are submitted, such as combination of one-way traffic and special use way, no turn -left, temporary traffic channeling, one-way green wave, and estimation methods also are set. Estimation result illustrates that these traffic evacuation methods of unexpected events can improve network capacity and promote rescue and evacuation.

Keywords: traffic engineering, unexpected events, traffic evacuation, one-way traffic, traffic channeling, traffic estimation.

1 Introduction

Emergency in city causes the drastic increase of traffic volume in local area within a short time, resulting in intensifying contradictions between traffic supply and demand, the formation of traffic jams, causing traffic chaos and creating difficulties for rescue and evacuation, having a great impact on normal traffic flow. However, the impact caused by this incident is temporary, therefore, generally it can not be solved simply relying on increasing road capacity (road building)[1,2], because this solution not only works very hard but also is uneconomical. Throughout the active respond to emergency effectively and successfully at home and abroad, it can be discovered that the key of relief the contradiction between supply and demand after the incident and ensuring a successful response to emergencies is to formulate and implement the traffic evacuation plan of emergency systemically, scientifically and rationally[3].

Traffic flow of emergency in city has focused on the characteristics of one-way, and it can easily form traffic shock waves which can result in concentrated outbreak of traffic demand in the local region, causing traffic congestion and traffic chaos.

After analyzing and studying the traffic characteristics of unexpected events and various traffic evacuation methods, the author submits the traffic evacuation methods of unexpected events, such as combination of one-way traffic and special use way,

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 479–486.
springerlink.com

changing the way of turn-left in intersection, temporary traffic channeling, one-way green wave, and evaluation methods of traffic evacuation programs of unexpected events also are set.

2 Traffic Evacuation Methods of Unexpected Events

2.1 Combination of One-Way Traffic and Special Use Lane

After the occurrence of emergencies in city, in order to evacuate the largest crowd in the shortest possible time, one-way traffic is the ideal method of traffic evacuation. Evacuation people of emergency with the characteristic of one-way evacuation, which coincides with the one-way traffic, so one-way traffic in the direction of the evacuation capacity than the two-way traffic could double[4] to speed up evacuation.

But during the evacuation, we must make sure that a variety of special vehicles (ambulance, fire engines, etc.) can be fast and convenient access to event area. Therefore, special use lane must be opened up in the evacuation road, exclusively for dealing with emergency services vehicles, such as engineering emergency, fire rescue, ambulance or the implementation of urgent official matters by police. Any community vehicle is prohibited from entering, or staying in the lane for various reasons. In the period of public emergency evacuation, in the emergency special use lane, evacuation of the provisions only allow the rescue vehicle traffic and other vehicles can not enter the special lanes.

Based on the above analysis, after the unexpected incident, the method of traffic evacuation that can ensure emergency vehicles to enter the incident area to perform a rescue while can evacuate the crowd as soon as possible is the combination of one-way traffic and special use lane, shown in Fig.1.

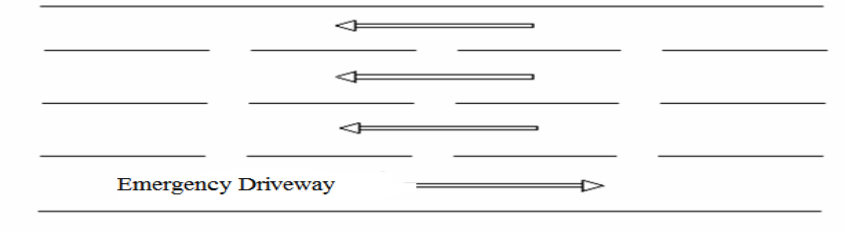


Fig. 1. Combination of one-way traffic and special use lane

2.2 Change the Way of Left Turn in Intersection

Vehicles entering and leaving the intersection, due to the different moving direction, the conflict between vehicles is not the same way, resulting in confluence points, diverging points and crossing points[5]. In the three types of conflict points, the points of crossing are produced by straight vehicles, turning left vehicles, turning left and going straight, turning left vehicles have the greatest impact on safety of traffic interference and running, followed by the confluence point, again, is the diverging point.

Intersections without traffic control, there are all kinds of conflict points. Their quantity is increasing significantly with the increase of lanes of roads that are crossed and the crossing points grow fastest among them. The crossing points are produced most by left-turn vehicles. Left-turn traffic at intersections is a major factor in the formation of conflict points and the main reasons of the formation of traffic conflict. Therefore, the rational evacuation of left-turn traffic is a effective method to ensure traffic safety and to improve intersection capacity[6].

After the occurrence of emergencies in city, in order to evacuate the largest crowd in the shortest possible time, the most effective way of evacuation methods of intersection is to prohibit turning left and it is feasible in the incident area. Because the evacuation of people with one-way evacuation will leave the incident area as soon as possible along the direction of evacuation, they need a very small chance of turning in the incident area. Prohibition of left-turn implemented, the reduction in the number of conflict points at intersection will reduce the level of traffic conflict so that the capacity of intersection is greatly improved, especially the straight direction.

The management of prohibition of left-turn implemented, left-turn lanes or straight-left lanes are changed directly to straight lanes. Because the headway of straight traffic is different from that of left-turn vehicles, according to the formula for calculating capacity, its capacity will be significantly changed. For non-signalized crossing, left-turn traffic has no impact on the straight vehicles so that the capacity will increase significantly. For signalized intersections, the implementation of left-turn ban will reduce the number of intersections of the phase, so that a cycle of straight traffic can be assigned to more passing time and there is a significant increase in traffic capacity. In the sudden incident, people expect to leave the danger area away as soon as possible for their fear psychology, while left-turn traffic is the culprit of congested intersection and the production of delay. Therefore, in the event of an emergency, measures taken to prohibit left-turn can effectively improve the intersection capacity and reduce intersection delays.

2.3 Temporary Traffic Channeling

View of the temporary characteristics of traffic evacuation management measures of unexpected events, on the basis of using existing channeling facilities, facilities of temporary traffic channeling should become the main ones, such as barrier, separator, divided blocks, marking, temporary traffic signs, etc.

For serious consequences after the emergency, it will be greatly inconvenient for the crowd in chaos, if there have been no temporary traffic channeling. For example, road traffic accident occurred, when the accident causes traffic disruption, then you can organize the traffic by means of temporary channels. Isolate the scope of the accident with barrier to ensure the on-site for sampling and to judge cause of the accident and the responsibility. And set up temporary signs at the intersections which are before and after the accident area to inform past drivers of the road break, please bypass to avoid congestion on the road. Showed by Fig. 2.

The traffic can be orderly by using of temporary traffic channeling mode and evacuation will achieve more with less.

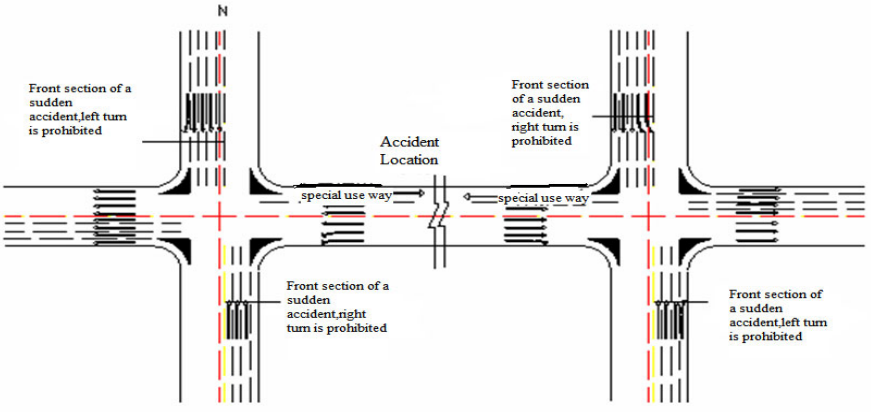


Fig. 2. Sketch of temporary traffic channeling

2.4 One-Way Green Wave Control

The main road in the urban road traffic network is the aorta which ensures the normal operation of urban socio-economic activities. A main road traffic running condition will directly affect the road traffic condition that it covers the large areas around. When signal control of intersection along the main road in the city done, in order to maintain the continuity of the main road traffic to the extent possible, the intersection along the main road should be managed group by group, based on road and environmental conditions, namely, urban trunk road section traffic signal base is implemented system control characterized by coordinated form[7,8].

On the collector-distributor roads that public emergencies in the city mainly occur, traffic signal coordination control is a effective measure to improve the road capacity within the event. After the incident, traffic flow on the road with very obvious characteristic of uneven direction, presenting a one-way flow. According to this feature, the "Green Wave Control" can implement "one-way Green Wave Control" different from the conventional "two-way Green Wave Control".

After the occurrence of emergency in the city, consideration at that time is how to evacuate the crowd quickly in the shortest possible time. When organizing traffic by using one-way green wave, there is no need to take road speed into account and just the distance between adjacent intersection and intersection signal cycle are needed to consider to build trunk one-way Green Wave Control. Therefore, it is more simple than the two-way Green Wave Control, and more quickly and efficiently, and it can evacuate the largest crowd in the shortest possible time.

3 Evaluation of Traffic Evacuation Methods of Unexpected Events

It can be seen from the above analysis that measures applied to public emergency in the city are mainly one-way traffic, prohibition of left turn, temporary traffic channeling and one-way green wave control, of which one-way traffic is the more core content.

Measures of prohibition of left turn and temporary traffic channeling are both aimed at the implementation of one-way traffic and one-way Green Wave is on the basis of one-way traffic.

3.1 Intersection Capacity of Traffic Evacuation of Unexpected Events

As we all know, the greatest impact on traffic in the intersection is the conflict point which caused by turn left and straight car. After the one-way traffic the number of conflict points significantly reduced, as shown in Fig.3, which undoubtedly will improve the intersection capacity.

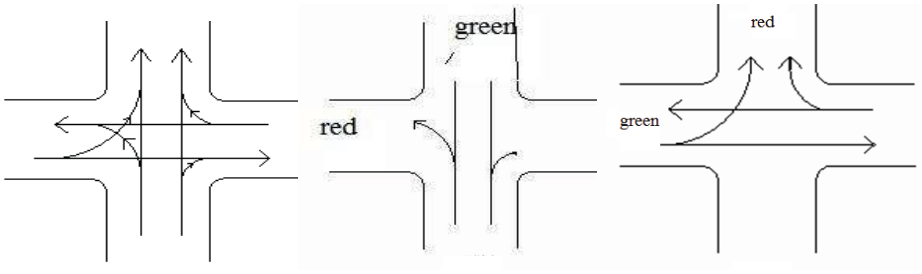


Fig. 3. Traffic conflict of one-way traffic

Following the intersection capacity on the one-way traffic signal controlled was calculated by using the stop line method according to the "urban road design standard".

Design capacity of straight drive was calculated by the following formula:

$$N_s = 3600 \times \varphi_s [(t_g - t_1) / t_{is} + 1] / t_c \quad (1)$$

Where N_s —The design capacity of one straight lane,

t_c —Signal period,

t_g —The green time in one signal period,

t_1 —The time when the first car start and through the stop the line after the green light on, the value can be used 2.35,

t_{is} —The average interval of straight or right vehicles through the stop line,

φ_s —The reduction factor of capacity of straight drive traffic, the value can be used 0.9.

In the case of Figure 3 (b), the design capacity of imported road A is $N_A = N_{sl} + N_{sr}$, because of $N_{sl} = N_{sr} = N_s$,

So $N_A = 2N_s$,

Where N_{s1} —The design capacity of one straight and left lane,

N_{sr} —The design capacity of one right lane.

In the case of Figure 3 (c), the design capacity of imported road B is $N_B = N_{sr} + N_s$. Because of taking the one-way traffic, the capacity of imported road C is 0 and the design capacity of imported road D is $N_D = N_{sl} = N_s(1-\beta_1/2)$,

Where β_1 is the percentage of the left turn car in the right and left lane before taking the one-way traffic control.

The design capacity of one-way traffic intersection is calculated by the formula (2):

$$N_{oneway} = N_A + N_B + N_C + N_D = 2N_s + N_s + 0 + N_s(1-\beta_1/2) = N_s(4-\beta_1/2) \quad (2)$$

But $N_{slr} = N_{sl} + N_s(1-\beta_1/2)$

Where N_{s1r} —The design capacity of one straight, left and right lane;

So the capacity of a normal cross intersection of signal control is calculated as formula (3):

$$N_{normal} = N_A + N_B + N_C + N_D = 4N_{slr} + N_s(4-2\beta_1) \quad (3)$$

After the intersection is changed into the one-way traffic, the improved percentage of the capacity of intersection is expressed as formula (4) :

$$\lambda = \frac{N_{oneway} - N_{normal}}{N_{normal}} \times 100\% = \frac{3\beta_1}{N_s - 4\beta_1} \times 100\% \quad (4)$$

According to the above formula (4), it can be seen that the greater the proportion of the left turn cars in intersection, the greater the rate of increase of the intersection capacity after the intersection changed into a one-way traffic.

3.2 Section Capacity of Traffic Evacuation of Unexpected Events

After using the evacuation of one-way traffic, the design capacity of road section is:

The possible capacity of a driveway is $N_p \frac{3600}{t_i}$,where t_i is the headway.

The design capacity of driveway affected by the intersection is:

$$N_m = \alpha_c \bullet \alpha_d \bullet N_p \quad (5)$$

Where: α_c —Road classification factor of capacity of driveway;

α_d —Influence factor of intersection.

$$\text{Where } \alpha_d = \frac{\frac{l}{v}}{\frac{l}{v} + \frac{v}{2a} + \frac{v}{2b} + t_r}$$

Where: l—The distance between the intersection;
v—design speed of section;
a—The average acceleration of start:
b—The average deceleration of braking;
t_r—The waiting time r of vehicles between intersection.

It is not difficult to see from the formula (5) , in the case of certain other conditions, two-way traffic changed into one-way traffic, due to the improvement of intersection capacity, the waiting time of vehicles at intersection is reduced, resulting in the improvement of road section capacity. In addition, because two-way traffic is changed into one-way traffic, one-way traffic direction on the number of lanes increase and the road section capacity will be significantly enhanced also.

4 Conclusion

Environment and energy disasters and emergencies brought by densely populated cities, frequently economic activities and all types of dense buildings are threatening the production and life of urban residents continually, even life. More and more frequent emergencies are testing the emergency response capacity of modern urban managers increasingly. The traffic in response to unexpected events plays a key role. As sudden changes in traffic flow caused by the emergency, the conventional traffic evacuation is very difficult to adapt to changes in traffic flow. There is need to research traffic evacuation methods for emergencies according to changes in traffic flow characteristics. This article is based of unexpected events traffic flow with characteristics of one-way evacuation, proposing traffic evacuation methods that fit the characteristics of one-way evacuation ,such as one-way traffic, special road, organization of intersection turning, one-way Green Wave and so on. It is hoped that it can provide a reference for the traffic management department and lay the foundation for the development of traffic emergency plans.

References

1. Chen, Q., Wang, F.: Spacial and temporal influences of emergency on traffic flow in large-scale activity. Journal of Traffic and Transportation Engineering 3(9), 81–85 (2009)
2. Zhai, Z.-m.: Road traffic organization optimization. China Communication Press, Beijing (2004)
3. NCHRP. Transportation Planning and management for Special Events[R], A synthesis of highway practice, Transportation research Board, Washington D.C., 2003:32-41
4. Xiao, A.: Setup conditions and methods of one-way traffic. TRAFFIC & TRANSPORTATION.2005.3:33

5. Jin, L.: The Improvement of the Traffic Organization of Mixed Traffic Flow on Signalized Intersection.Chang'an University master paper, 2006
6. Ge, H-q.: Study on prohibition of left-turn of urban road. Chang'an University master paper (2004)
7. Lu, K., Xu, J-m.: Offset Model for Arterial Road Coordinate Control and Its Optimization Method China. Journal of Highway and Transport 1(21), 83–88 (2008)
8. Wang, W-p.: Study of Optimal Method for the Traffic Control in Urban Plane Intersection. Shandong University of Science and Technology master paper (2004)
9. Li, R.-b.: Research of capacity of one-way traffic. China Municipal Engineering 2, 1–3 (1998)

Medical Image Registration Based on Improved PSO Algorithm

Lin-Tao Zheng, Gui-Ping Qian, and Lan-Fen Lin*

Department of Computer Science and Engineering
Zhejiang University
Hangzhou, P.R. China
{zhenglintao,llf}@zju.edu.cn,
qianguiping@163.com

Abstract. Image registration is playing an increasingly important role in modern medical image processing. It refers to the process of overlaying two or more images of the same scene taken at different times, from different viewpoints, and/or by different sensors. During the last decades, a large number of image registration methods have been described in the literature. Unfortunately, there is no one method that works very well for all applications. In this paper, we propose a new approach using improved Particle swarm optimization for medical image registration. The algorithm has been successfully used for medical image registration. The feasibility of the proposed method is demonstrated and compared with the standard PSO based image registration technique. The experimental results show that the propose method better results than the standard PSO method.

Keywords: image registration; improved particle swarm optimization; mutual information.

1 Introduction

Medical imaging is about establishing shape, structure, size, and spatial relationships of anatomical structures within the patient, together with spatial information about function and any pathology or other abnormality. Establishing the correspondence of spatial information in medical images and equivalent structures in the body is fundamental to medical image interpretation and analysis. In many medical clinical applications, images of similar or differing modalities often need to be aligned as a preprocessing step for many planning, navigation, data-fusion and visualization tasks. This alignment process is known as image registration.

Since the mid 1980s medical image registration has evolved from being perceived as a rather minor precursor to some medical imaging applications to a significant subdiscipline in itself. Entire sessions are devoted to the topic in major medical imaging conferences, and workshops have been held on the subject. Image registration has

* Corresponding author.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 487–494.
springerlink.com

also become one of the more successful areas of image processing, with fully automated algorithms available in a number of applications [1, 2, 11].

Medical image registration has a wide range of potential applications. These include[1]:

1) Combining information from multiple imaging modalities, for example, when relating functional information from nuclear medicine images to anatomy delineated in high-resolution MR images.

2) Monitoring changes in size, shape, or image intensity over time intervals that might range from a few seconds in dynamic perfusion studies to several months or even years in the study of neuronal loss in dementia.

3) Relating preoperative images and surgical plans to the physical reality of the patient in the operating room during image-guided surgery or in the treatment suite during radiotherapy.

4) Relating an individual's anatomy to a standardized atlas.

Image registration is the process of overlaying two or more images of the same scene taken at different times, from different viewpoints, and/or by different sensors [3]. In recent years a number of image registration approaches have been devised. These techniques can be divided into feature-based techniques and intensity-based techniques. Feature-based techniques require some preprocessing, prior to registration, to extract relevant information, like anatomical land-marks, edges or shapes. In contrary to feature-based techniques, intensity-based measures get by without prior preprocessing. Thus images can be registered right after image acquisition. Intensity-based measures use the full raw image information for image alignment. Here we adopt the latter approach.

The focus of the current paper is the search strategy (optimization) for maximizing the similarity metric for registering images. This paper focuses on a new improved PSO technique. To the authors' knowledge, this technique is relatively recent and is better know for optimization but has not been applied to medical image registration previously. In the following sections, we will introduce this method in more details, and then apply it to medical image registration. Theoretical analysis and experiments show that this method is effective and accurate to register medical images.

This reminder of this paper is organized as follows. We first introduce the improved particle optimization algorithm used as optimization method in section 2. Then we introduce the mutual information used as registration criterion of two images in section 3.Our experiments and discussion is in section 5.

2 Particle Swarm Optimization

Particle Swarm Optimization (PSO) is a recently proposed algorithm by James Kennedy and R. C. Eberhart in 1995, motivated by social behavior of organisms such as bird flocking and fish schooling[2]. PSO algorithm is not only a tool for optimization, but also a tool for representing sociocognition of human and artificial agents, based on principles of social psychology. The particle swarm concept originated as a simulation of simplified social system. The original intent was to graphically simulate

the choreography of bird of a bird block or fish school. However, it was found that particle swarm model can be used as an optimizer. During the past few years, PSO has been successfully applied to multidimensional optimization problems, artificial neural network training, and multiobjective optimization problems.

Particle Swarm Optimization optimizes an objective function by undertaking a population-based search. The population consists of potential solutions, named particles, which are metaphor of birds in flocks. These particles are randomly initialized and freely fly across the multi dimensional search space. During flight, each particle updates its own velocity and position based on the best experience of its own and the entire population.

2.1 Original Particle Swarm Optimization

Mathematical notation of original PSO is defined as follow:

An individual particle i is composed of three vectors: its position in the D-dimensional search space $\vec{x}_i = (x_{i1}, x_{i2}, ..., x_{iD})$, the best position that it has individually found $\vec{p}_i = (p_{i1}, p_{i2}, ..., p_{iD})$, and its velocity $\vec{v}_i = (v_{i1}, v_{i2}, ..., v_{iD})$ .Particles were originally initialized in a uniform random manner throughout the search space; velocity is also randomly initialized.

These particles then move throughout the search space by a fairly simple set of update equations. The algorithm updates the entire swarm at each time step by updating the velocity and position of each particle in every dimension by the following rules[4]:

$$v_{id} = v_{id} + c_1 * rand() * (p_{id} - x_{id}) + c_2 * Rand() * (p_{gd} - x_{id}) \quad (1)$$

$$x_{id} = x_{id} + v_{id} \quad (2)$$

where in the original equations c_1 and c_2 are a constant with the value of 2.0, rand and Rand are independent random numbers uniquely generated at every update for each individual dimension d = 1 to D, and $\vec{p}_g$ is the best position found by any neighbor of the particle.

2.2 Improved Particle Swarm Optimization

Particle swarm optimization for image registration was introduced in [5]. The basic PSO algorithm can be enhanced in a variety of ways.

Although the original PSO algorithm has ability of exploiting the global maximum, it can not guarantee to achieve the maximum but often falls into local maxima. To overcome this shortcoming, In this paper, we use the following equations to instead equation (1) and (2)[6]:

$$v_{id} = w * v_{id} + c_1 * rand() * (p_{id} - x_{id}) + c_2 * Rand() * (p_{gd} - x_{id}) \tag{3}$$

$$x_{id} = x_{id} + v_{id} \tag{4}$$

$$w = \mu + \sigma * N(0,1)$$

$$\mu = \mu_{\min} + (\mu_{\max} - \mu_{\min}) * rand(0,1)$$

In the above equations, $N(0,1)$ is a random numbers drawn from the standard normal distribution.

```
For each particle
    Initialize particle
For each particle
    Calculate fitness value
    If the fitness value is better than the best fitness
    value (pbest) in history
    set current value as the new pbest
End
Choose the particle with the best fitness value of all
the particles as the gbest
For each particle
    Calculate particle velocity according to Eq.1.3
    Update particle position according to Eq.1.4
End
Continue while maximum iterations or minimum error
criteria is not attained
```

Fig. 1. The Pseudo code of the improved PSO procedure

3 Mutual Information

Mutual information(MI) is a basic concept originating from information theory, measuring the statistical dependence between two random variables or the amount of information that one variable contains about the other. Mutual information is an information theoretic topic that has quickly become one of the most popular techniques available for use in image registration[9].

MI originated many decades ago from works based on entropy and its roots stem from information and communication theory. However, it was first proposed as a registration measure in medical image registration in 1995, independently by Viola

and Wells[7] and by Collignon[8]. The mutual information I of two images X and Y is defines as

$$I(X,Y) = H(X) + H(Y) - H(X,Y)$$

where $H(X)$ and $H(Y)$ are individual entropies and $H(X,Y)$ is the joint entropy. For two images, the mutual information is computed from the joint probability distribution of the images' intensity or gray-values. When two images are aligned, the joint probability distribution is "peaky" resulting in a high mutual information value. The greater the value of MI, the better the match between the two images, so image registration becomes a typical maximization problem. The definitions of these entropies are

$$H(X) = -\sum_{x} P_X(x) \log P_X(x)$$

$$H(Y) = -\sum_{y} P_Y(y) \log P_Y(y)$$

$$H(X,Y) = -\sum_{x,y} P_{X,Y}(x,y) \log P_{X,Y}(x,y)$$

where $P_X(x)$ and $P_Y(y)$ are the marginal probability mass functions and $P_{X,Y}(x,y)$ is the joint probability mass functions.

The success of MI registration lies in its simplicity as it is considered to be quite a general similarity measure. It makes very few assumptions regarding the relationship that exists between different images. Assumptions regarding linear correlation or even functional correlation are not made.

4 Image Registration Model

Fig. 2 shows the framework of our proposed method in which the optimizer is based on the Improved Particle Swarm Optimization(Eq. 3 and Eq. 4).

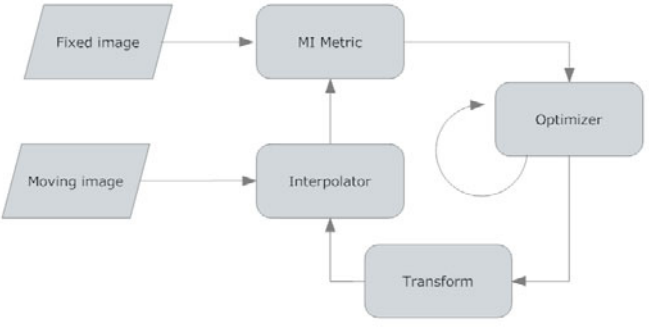
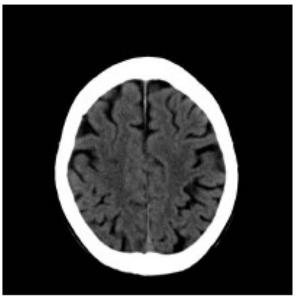


Fig. 2. Framework of our proposed method

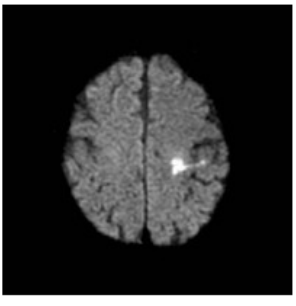
5 Results

In order to evaluate the proposed PSO-based image registration, we compared the proposed method with original pso method-based image registration.

In this section, we perform several registration experiments with medical image data to evaluate the performance of the proposed technique. Moreover, we also perform conventional PSO for comparison.



(a) target image

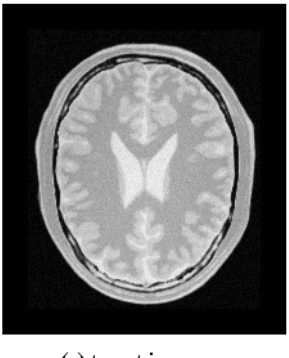


(b) floating image

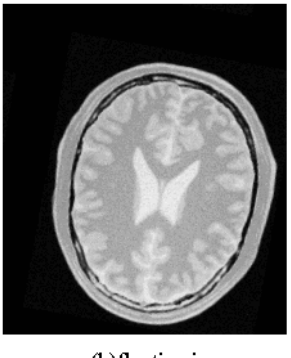
Brain CT data and MR data are used as the reference image and the test image, respectively.

Table 1. Registration results of experiment 1

	T_x	T_y	θ
Ground truth	7	0	2
Standard PSO	9	1	0
Our algorithm	7	0	2



(a) target image



(b)floating image

Table 2. Registration results of experiment 2

	T_x	T_y	θ
Ground truth	13	17	10
Standard PSO	20	14.5	6.4
Our algorithm	13.1	15.7	10.1

6 Conclusion

In this paper we have presented a new method for the registration of medical images which is based on the combination of mutual information and a new improved PSO technique. Results show that the proposed technique is better than the original PSO technique. According to the experiment, we can conclude that the method proposed is more effective than the traditional PSO method.

Acknowledgement

The paper was supported by Zhejiang Provincial Natural Science Foundation of China (Grant No. Y1100018).

References

1. Hajnal, J.V., Hawkes, D.J., Hill, D.L.G.: Medical image registration. CRC, Boca Raton (2001)
2. Brown, L.: A survey of image registration technique. ACM Comput. Surv. 24, 325–376 (1992)
3. Zitová, B., Flusser, J.: Image registration methods: A survey. Image Vis. Comput. 21, 977–1000 (2003)
4. Eberhart, R.C., Kennedy, J.: A new optimizer using particle swarm theory. In: Proceedings of the Sixth International Symposium on Micro Machine and Human Science, Nagoya, Japan, pp. 39–43 (1995)
5. Talbi, H., Batouche, M.: Particle swam optimization for image registration. Information and Communication Technologies: From The to Applications, 397–398 (2004)
6. Chun, G., Zheng, W-l.: Master MATLAB for optimization computation, PUBLISHING HOUSE OF ELECTRONICS INDUSTRY (2009)
7. Viola, P., Wells III, W.M.: Alignment by maximization of mutual information. In: Grimson, E., Shafer, S., Blake, A., Sugihara, K. (eds.) International Conference on Computer Vision, pp. 16–23. IEEE Computer Society Press, Los Alamitos (1995)
8. Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., Marchal, G.: Automated multimodality image registration based on information theory. In: Bizais, Y., Barillot, C., Di Paola, R. (eds.) Information Processing in Medical Imaging, pp. 263–274. Kluwer Academic Publishers, Dordrecht (1995)

9. Luan, H., Qi, F., Xue, Z., Chen, L., Shen, D.: Multimodality image registration by maximization of quantitative-qualitative measure of mutual information. Pattern Recognition 41, 285–298 (2008)
10. Raghavendra, R., Dorizzi, B., Rao, A., Hemantha Kumar, G.: PSO versus AdaBoost for feature selection in multimodal biometrics. In: Proceedings of Third IEEE International Conference on Biometrics: Theory, Application and System, BTAS 2009, Washington, DC, USA, pp. 1–4 (2009)
11. Shams, R., Sadeghi, P., Kennedy, R., Hartley, R.: A Survey of Medical Image Registration on Multicore and the GPU. Signal Processing Magazine 27(2), 50–60 (2010)

The Application of Sliding Mode Control in High Precision Low Speed Servo System

Liu Hui-bo[1,2], Zhou Hai-jing[2], and Feng Zheng[3,4]

[1] School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China
[2] Information Engineering School, UST Inner Mongolia, Baotou 014010, China
[3] National Astronomical Observatories/Nanjing Institute of Astronomical Optics & Technology, Chinese Academy of Sciences, Nanjing 210042, China
[4] Graduate University of Chinese Academy of Sciences, Beijing 100049, China

Abstract. In this paper, the principle and characteristics of friction drive is analyzed. The mathematical model of the LAMOST telescope mount drive control system is established and the fuzzy sliding mode control theory is put forward to contrast with traditional PID control and ordinary sliding mode control, which is realized by Matlab/simulink simulation. The simulation results show that the fuzzy sliding mode control has good performance of speed tracking.

Keywords: servo system; sliding mode; fuzzy sliding mode.

1 Introduction

Friction is a common physical phenomenon, which exists in all sports. In servo system, friction that can cause some adverse impacts of limit cycle oscillation and low speed creeping of system is a major factor affecting the low-speed performance. The impacts can be reduced by changing the mechanical properties or using high-precision controller, considering the control algorithm is a better choice to inhibit the impact. Sliding mode control, which has strong insensitivity to the model error and parameter changes of and external disturbance, has strong robustness. The simulation on Matlab/simulink proves the superiority of fuzzy sliding mode control.

2 The Structure of LAMOST Mount Drive Servo System

LAMOST mount drive servo system is a two-axis servo system which adopts friction drive and uses permanent magnet DC torque motor to drive. The control model established according to the laws of kinematic is shown in Figure 1.

In Figure 1, u represents the control input signal; L and R respectively represents the total inductance of armature circuit and the total resistance; K_t represents the motor torque coefficient; T_m represents motor shaft torque; T_{eq} represents the equivalent torque which converted on the motor shaft; J_{eq} represents the equivalent moment of inertia

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 495–502.
springerlink.com

which converted on the motor shaft, N represents the friction drive reduction ratio; K_e represents the inverse EMF constant; S represents the Laplace operator; $\dot{\theta}$ represents the speed; θ represents the tracking position.

The state space equation of the telescope mount drive control system can be obtained as follows:

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= \frac{1}{J_{eq}\mathrm{N}}(x_3 - T_{eq}) \\ \dot{x}_3 &= \frac{1}{L}\left(K_t u - K_t K_e \mathrm{N} x_2 - R x_3\right) \end{aligned} \tag{1}$$

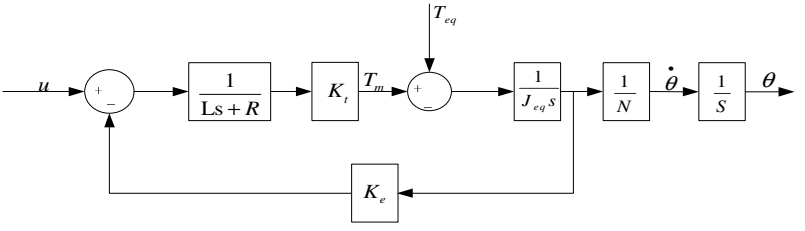


Fig. 1. The model of LAMOST mount drive servo system

3 The Model of Friction

The Stribeck friction model which is shown in Figure 2 is used in this paper.

When $|\dot{\theta}(t)| < \alpha$, the static friction is

$$\mathrm{F}(t) = \begin{cases} F_m & F(t) > F_m \\ F(t) & -F_m < F < F_m \\ -F_m & F_t < -F_m \end{cases} \tag{2}$$

When $|\dot{\theta}(t)| > \alpha$, the dynamic friction is

$$\mathrm{F}_t = \left[F_c + (F_m - F_c) e^{-\alpha |\dot{\theta}(t)|_1} \right] \mathrm{sgn}(\dot{\theta}(t)) + k_v \dot{\theta} \tag{3}$$

where, F_t represents the driving force; F_m represents the maximum friction; F_c represents the coulomb friction; K_V represents the ratio coefficient of viscous friction torque; $\dot{\theta}(t)$ represents the rotation velocity; α、α_1 represent the small and positive constants.

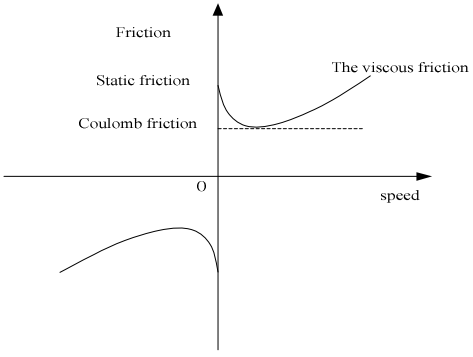


Fig. 2. Friction-Speed steady state curve(Stribeck curve)

4 Design of the General Sliding Mode Controller

4.1 Design of Controller

θ represents the position signal of the azimuth axis. The drive motor which used in the simulation model is the permanent magnet DC torque motor manufactured by Precision Motor Factory of Chengdu. To ignore the armature inductance, the position state equation of the drive control servo system can be described as follows:

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & -\frac{K_e K_t}{JR} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{K_t}{JNR} \end{bmatrix} u(t) - \begin{bmatrix} 0 \\ \frac{1}{JN} \end{bmatrix} F_f(t) \tag{4}$$

Where, $x_1(t) = \theta(t)$ represents the rotation angle; $x_2(t) = \dot{\theta}(t)$ represents the rotation velocity.

To design the switching function as $s = ce + \dot{e}$ and adopt the exponential reaching law as $\dot{s} = -\varepsilon \operatorname{sgn}(s) - ks$, where $\varepsilon > 0, k > 0$, the follows can be got:

$$\dot{s} = c\dot{e} + \ddot{r} - (-\frac{K_e K_t}{JR}\dot{x} + \frac{K_t}{JNR}u - \frac{F_f}{JN}) \tag{5}$$

Follows can be got based on the above fomula:

$$u = \frac{JNR}{K_t}(c\dot{e} + \ddot{r} + \varepsilon \operatorname{sgn}(s) + ks + \frac{K_e K_t}{JR}\dot{x} + \frac{F_f}{JN}) \tag{6}$$

4.2 Analysis of Algorithm Convergence

To adopt the Lyapunov algorithm convergence and take the Lyapunov function as follows:

$$\mathrm{V}(x)=\frac{1}{2}s^2 \tag{7}$$

Follows can be got according to $s=ce+\dot{e}$ and $\dot{s}=-\varepsilon\,\mathrm{sgn}(s)-ks$, we can get:

$$\dot{\mathrm{V}}(x)=s\dot{s}=(ce+\dot{e})[-\varepsilon\,\mathrm{sgn}(s)-ks]=-(ce+\dot{e})[\varepsilon\,\mathrm{sgn}(s)+ks] \tag{8}$$

Where, $\varepsilon>0,k>0$, and $\dot{\mathrm{V}}(x)<0$ when $s<0$, $\mathrm{sgn}(s)<0$; when $s>0,\mathrm{sgn}(s)>0$ and $\dot{\mathrm{V}}(x)<0$. The system is asymptotically stable according to Lyapunov stability of the algorithm, and $s\dot{s}<0$ is permanent establishment, and then the reaching condition of sliding mode is satisfied. System dynamics will arrive at the sliding surface within a limited time and stay on it, so the controller designed is feasible.

5 Design of Fuzzy Sliding Mode Controller

Fuzzy sliding mode controller is usually composed of the equivalent control u_{eq} and the switching control u_{sw}, and the form of the control volume is $u=u_{eq}+u_{sw}$. To adopt a two-dimensional fuzzy controller, and suppose the fuzzy controller input is S and ds. The switching sliding mode control u_{sw} can be obtained according to the output of fuzzy controller.

(1)Fuzzy

Suppose the domain of S、ES and U is $\{-3,-2,-1,0,1,2,3\}$,and the corresponding subset of fuzzy language is $\{\mathrm{NB},\ \mathrm{NM},\ \mathrm{NS},\ \mathrm{ZO},\ \mathrm{PS},\ \mathrm{PM},\ \mathrm{PB}\}$.

(2)Fuzzy rules and fuzzy reasoning

To design the fuzzy control rules based on the existence and the reaching condition of sliding mode, which are shown in table 1. The fuzzy rule is: If S is A and ES is B, then U is C.

(3)De-fuzzy method

Using the center of gravity method, defuzzification formula according to the center of gravity method is shown as follows:

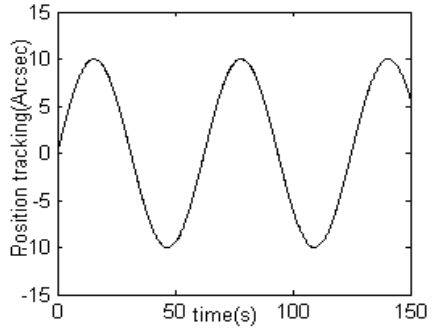
$$\mathrm{U}_0=\frac{\sum_{\mathrm{U}}\mu(\mathrm{U})*\mathrm{U}}{\sum_{\mathrm{U}}\mu(\mathrm{U})} \tag{9}$$

Table 1. Fuzzy rules

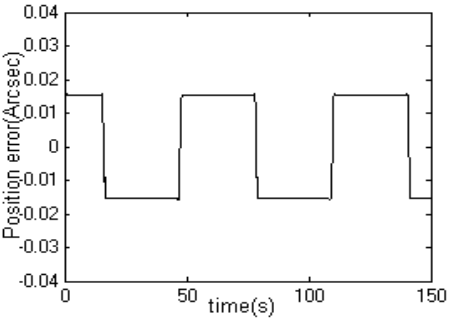
s ds	NB	NM	NS	ZO	PS	PM	PB
PB	ZO	PS	PM	PB	PB	PB	PB
PM	NS	ZO	PS	PM	PB	PB	PB
PS	NM	NS	ZO	PS	PM	PB	PB
ZO	NB	NM	NS	ZO	PS	PM	PB
NS	NB	NB	NM	NS	ZO	PS	PM
NM	NB	NB	NB	NM	NS	ZO	PS
NB	NB	NB	NB	NB	NM	NS	ZO

6 Simulation

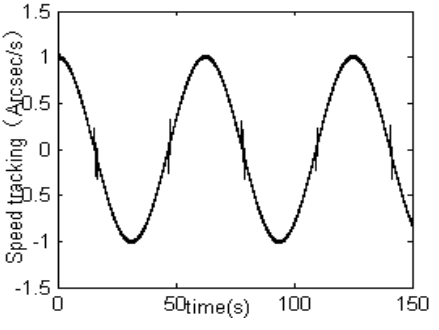
To create a simulation model of the system in Matlab / Simulink, in which the position tracking signal is *10"sin(0.1t)*, and the corresponding speed signal is *1"cos(0.1t)*. The simulation results are shown in Fig.3、Fig.4 and Fig.5



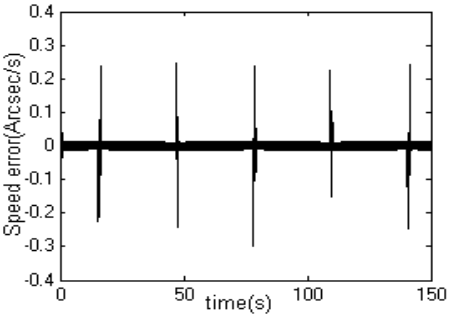
a(1)Position tracking curve



a(2)Position error curve

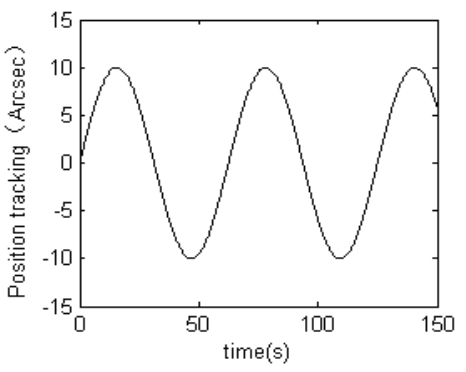


b(1)Speed tracking curve

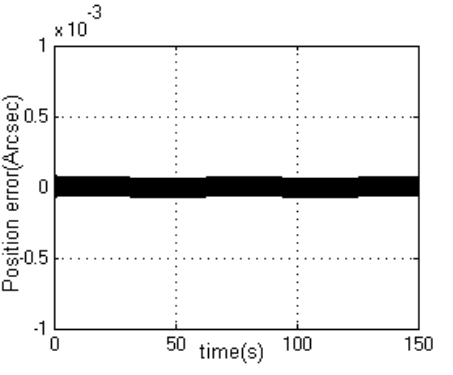


b(2)Speed error curve

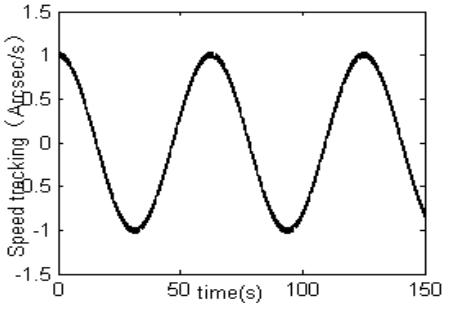
Fig. 3. PID control tracking curve



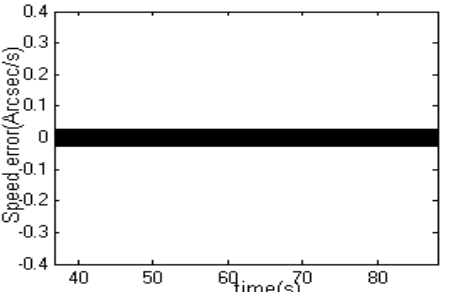
c(1)Position tracking curve



c(2)Position error curve



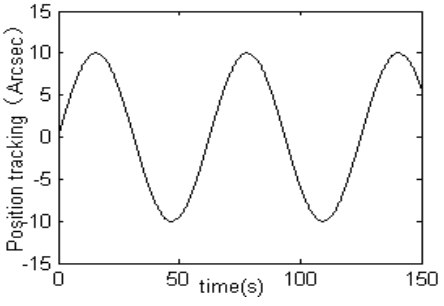
d(1)Speed tracking curve



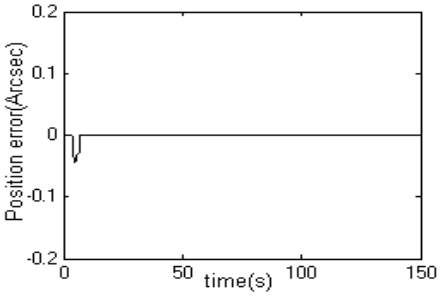
d(2)Speed error curve

Fig. 4. Sliding mode control tracking curve

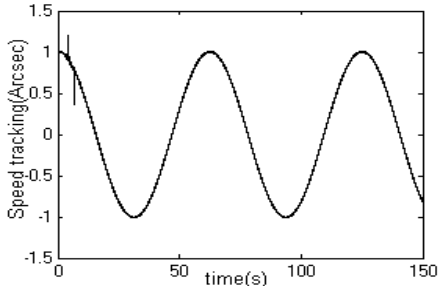
Fig.3 shows that, when adopting the PID control the position error is between *+0.02″* and *-0.02″*, the speed error is between *+0.3″/s* and *-0.3″/s*, and the speed signal chatters when crossing the zero. The simulation results show that the conventional PID control is ineffective on the low speed system which contains more nonlinear interference. Fig.4 shows that, the tracking accuracy of the position and the speed are more highly improved than conventional PID control when using the sliding mode control ,and however, the chattering phenomenon is produced. Fig.5 shows that, both position tracking and speed tracking can achieve the system performance requirements, and the chattering in the general sliding mode control is weakened. Meanwhile the system tracking accuracy is improved significantly.



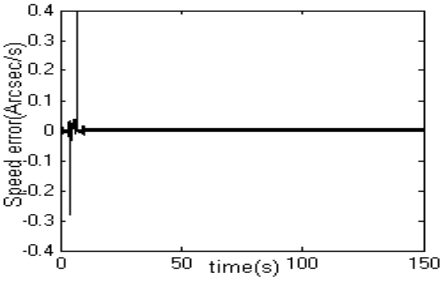
e(1)Position tracking curve



e(2)Position error curve



f(1)Speed tracking curve



f(2)Speed error curve

Fig. 5. Fuzzy sliding mode control tracking curve

7 Conclusion

The control of high-performance servo system is very difficult. This paper takes the LAMOST telescope mount drive system, which has the control difficulties of large inertia, low speed and multi-disturbance, as the research object. The conventional PID control is ineffective on the low speed system which contains more nonlinear interference. The simulation results show that the sliding mode control can achieve the performance requirements of the drive system, and can obtain a satisfactory low-speed tracking performance, and however chattering exists. Fuzzy sliding mode control, which designs the fuzzy rules based on reducing chattering, effectively reduces the chattering of sliding mode control, and improves low-speed tracking performance.

References

1. Xu, X.: Control system of LAMOST telescope. J. Aota Astrophysica Sinica 20, 43–52 (Supplementary 2000) (in Chinese)
2. Liu, J.: MATLAB Simulation for Sliding Mode Control. Tsinghua University Press, BeiJing (2005)

3. Liu, J.: MATLAB Simulation of Advanced PID Control(2). Electronic Industry Press, Bei-Jing (2004)
4. Zhang, L.: Principle and Design of Servo System. Beijing Institute of Technology Press, Beijing (2006)
5. XU X.: Technical requirements of LAMOST:20030909. Nanjing: Nanjing Institute of Astronomical & Technolog（LAMOST-TS-ED-L5-002）:1-18. (in Chinese)
6. Mark, J.S., Peter, M.T.: Azimuth/elevation Servo Performance of the W.N. Keck Telescope. In: Proc.of SPIE,2002, vol. 2199, pp. 126–141. SPIE, USA (2002)

One Improvement Control Method of Maximum Power Point Tracking

Zheng-ming Li, Xiao-hui Xia, and Yan-yan Yan

College of Electronic and Information Engineering
Jiangsu University
Zhenjiang, China
lzming@ujs.edu.cn, xiaxh_cool@126.com

Abstract. Synthesizing the advantages and disadvantages of each kind of maximum power point tracking (MPPT) control method, one improvement control method is proposed. This method uses the constant voltage method to find the approximate location of maximum power point rapidly in the earlier period, and uses hysteresis comparison method to increase the precision in the later period. Compared with the traditional MPPT control method, this method tracking speed is quicker, and the tracking precision is higher, and can realize the smooth start under each kind of changeful environment.

Keywords: solar cell; MPPT; control method.

1 Introduction

Driven by increasing costs and decreasing reserves of fossil-fuels, as well as by global environmental concerns, renewable energy is becoming a significant fraction of the total energy generation. [1] The solar energy already obtains more and more attention of various countries as the most valuable renewable energy source in the world. But the external environment is changeful (for example temperature, sunshine). The solar cell needs maximum power point tracking to use solar energy by the maximum efficiency. At present, the methods of maximum power point tracking (MPPT) are many, such as the constant voltage method, the perturbation and observation method, the incremental conductance method, the fuzzy control method, hysteresis comparison method [1-5] and so on, but the different methods have the different advantages and disadvantages in the actual application. In order to use the merit of each method fully, one improvement starting characteristic MPPT method is proposed in the foundation of the analysis to several common MPPT methods.The tracking speed and the precision could raise obviously by this method.

2 The Equivalent Model and Characteristic

2.1 Solar Cell Equivalent Model

The equivalent circuit model of solar cell unit is shown in Fig. 1. And, I_{ph} is the photoproduction electric current,which is in direct proportion to the area and the

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 503–510.
springerlink.com

incident light intensity of solar cell. I_d is the characteristic that the solar cell unit is similar to the ordinary diode in the non-illumination's situation because of the internal semiconductor PN ties. R_s is the series resistance. R_{sh} is the shunt resistance. Generally R_s is very small, but R_{sh} is very big.

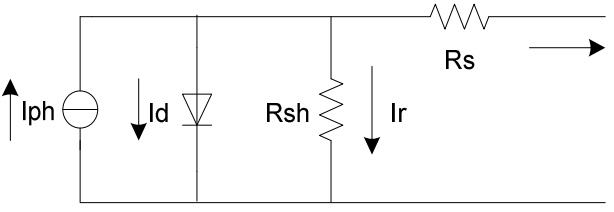


Fig. 1. Solar cell equivalent model

The relations of the solar cell's operating current and the voltage are as follows:

$$I = \frac{R_{sh}}{R_s + R_{sh}}[I_{ph} - \frac{U_{pv}}{R_{sh}} - I_D] \tag{1}$$

Where

$$I_D = I_0(e^{qU_{pv}/AKT} - 1) \tag{2}$$

With combination (1) and (2), obtain

$$I = I_{ph} - I_0[e^{qU_{pv}/AKT} - 1] - \frac{U_{pv} + IR_s}{R_{sh}} \tag{3}$$

I_0-------diode saturation current, I_{ph} ------- Photoproduction electric current, I ------- solar cell operating current, q------- electron charge $1.6\times10^{-19}C$, U_{pv} -------solar cell operating voltage, K-------Boltzmann constant $1.38\times10^{-23}J/K$,T------- absolute temperature A------- diode characteristic factor

In ideal situation, R_s=0, $R_{sh}\to\infty$, A =1.

$$I = I_{sc} - I_0[e^{qU_{pv}/KT} - 1] \tag{4}$$

When I=0, it can obain the open-circuit voltage of solar cell.

$$V_{oc} = \frac{kT}{q}\ln(\frac{I_{sc}}{I_0} + 1) \tag{5}$$

2.2 Solar Cell Characteristic

The open-circuit voltage and the electric current of solar cell change with the strength of illumination and the cell array temperature. Under certain strength of illumination, the I-V characteristic curve and the P-V curve of solar cell are shown in Fig. 2 and Fig. 3.

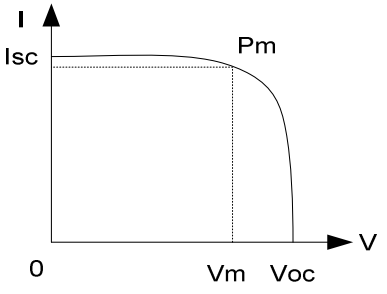


Fig. 2. I-V characteristic curve of solar cell

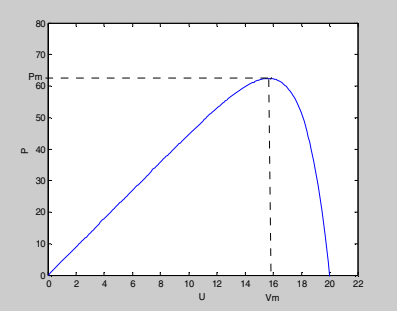


Fig. 3. P-V characteristic curve of solar cell

The position marked in the figure is the maximum power point. From the figure, it shows that solar cell has the obvious misalignment.

Fig. 4 and Fig. 5 are the I-V and P-V characteristic curve of solar cell under the same temperature and different sunshine (S) separately. From the figure, it shows that the output power of solar cell array increases with the increase of luminous intensity(S1<S2<S3).

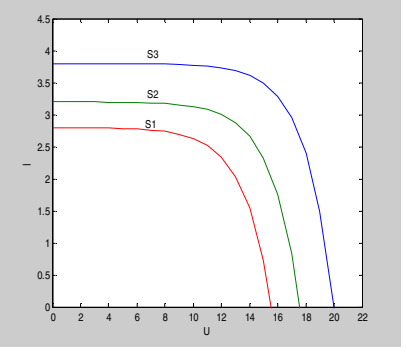


Fig. 4. I-V characteristic curve under different sunshine

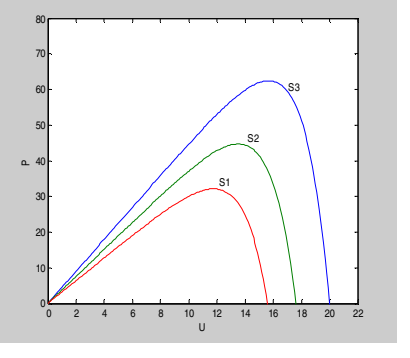


Fig. 5. P-V characteristic curve under different sunshine

Fig. 6 and Fig. 7 are the I-V and P-V characteristic curve of solar cell under the same sunshine and different temperature (T) separately. From the figure, it shows that the output power of solar cell array decreases with the increase of temperature (T1<T2<T3).

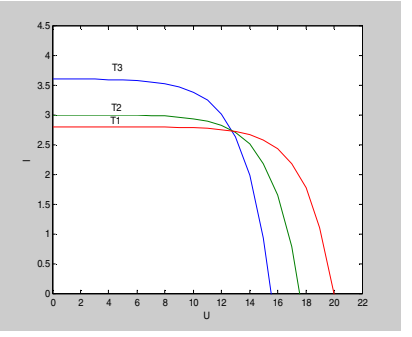


Fig. 6. I-V characteristic curve under different temperature

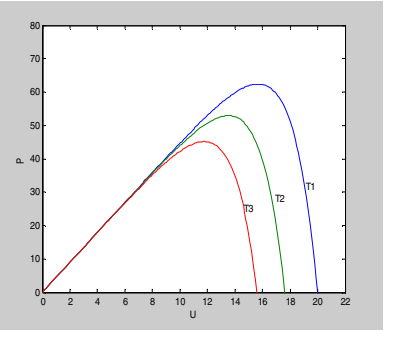


Fig. 7. P-V characteristic curve under different temperature

3 The MPPT Algorithm Comparison

3.1 The Perturbation and Observation Method

The perturbation and observation method (also called mountain climbing method) is one kind of control algorithm which is used most widely at present. Generally it controls the output voltage of the solar cell board through the control of the DC-DC switch's dutyfactor. First record the first time dutyfactor D_1, and write down the output voltage V_1 and the electric current I_1 of the solar cell at this time, and calculate the powe P_1. Next, let $D_2 = D_1 + \Delta D$ and measure the voltage V_2 and the electric current I_2 at the second time, then calculate the power P_2. D is decided to increase or reduce through the comparison of P_1 and P_2, accordingly, the maximum power point is found. Selecting the different size ΔD, the tracking speed and the precision are different. When ΔD is bigger, the tracking speed is quicker and the precision is lower. On the contrary, When ΔD is smaller, the tracking speed is slower and the precision is higher.

The perturbation and observation method has the merit of simple principle, quick tracking speed and high tracking accuracy, however, it also has the inherent shortcoming. It has the oscillatory occurrences and create the nonessential power loss for this nearby the maximum power point. When the external environment changes fiercely, it has the possibility to produce voltage collapse phenomenon [6].

3.2 The Incremental Conductance Method

The incremental conductance method decides the direction of the perturbation mainly through the judgment of positive and negative of $\frac{dP}{dV}$, and it is actually one kind of distortion of the perturbation and observation method. From the PV curve of solar cell

board in Figure 3, it shows that: When $\frac{dP}{dV}$ >0, the curve is at the left side of the maximum power point, and needs to increase the dutyfactor to increase the output voltage; When $\frac{dP}{dV}$ <0, the curve is at the right side of the maximum power point, and needs to decrease the dutyfactor to decrease the output voltage; When $\frac{dP}{dV}$ =0,it is at the maximum power point.

The incremental conductance method eliminates oscillatory occurrences of the perturbation and observation method completely, but it still has the voltage collapse phenomenon. Because of its complex algorithm, high measuring accuracy request and big cost, it is very difficult to be used widely.

3.3 The Constant Voltage Method

The principle of the constant voltage method is that the voltage of solar cell has the approximate proportional relationship with the open-circuit voltage at the maximum power point. Moreover this scale factor is invariant nearly when the sunshine and temperature outside changes. The working voltage can be adjusted through the measure of the open-circuit voltage of the battery board. Thus the maximum power point could be found.

$$V_{MAX} = M_V \times V_O \tag{6}$$

thereinto, M_V is scale factor.

This control method is simple and the facility cost is low, however, the tracking accuracy is low and the error is big.

3.4 The Constant Electric Current Method

The principle of the constant electric current method is that the operating current of the solar cell has the approximate proportional relationship with the short-circuit current at the maximum power point.This scale-up factor,the same as the constant voltage method, is invariant nearly when the sunshine and temperature outside changes. The operating current can be adjusted through the measure of the short-circuit current of the battery board.Thus the maximum power point could be found.

$$I_{MAX} = M_I \times I_{sc} \tag{7}$$

thereinto, M_I is scale factor.

The tracking accuracy of this kind of control method is low. Because it needs to let the solar cell board short circuit for measuring the short-circuit current, there is great affects to the life of the cell board. It holds the inferiority compared with the constant voltage method.

3.5 Hysteresis Comparison Method

Hysteresis comparison method is proposed to overcome the oscillatory occurrences and the voltage collapse phenomenon of the perturbation and observation method at the maximum power point. It does not make any movement when the external environment changes fiercely, and only carries on the tracking when the environment is stable. The concrete principle is as follows:

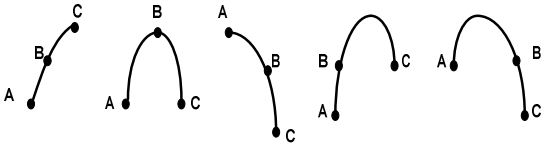


Fig. 8. Each kind of situation nearby the maximum power point

It needs to measure the duty factor D_b at a time , the voltage V_b and the electric current I_b nearby the maximum power point and calculate the power P_b ; Next let $D_a = D_b - \Delta D$ and measure V_a , I_a at this time and calculate the power P_a ; Next let $D_c = D_b + \Delta D$ and measure V_c , I_c ,and calculate the power P_c . Last set a comparison mark Tag.

If $P_c >= P_b$, Tag1=1; or Tag1=-1;If $P_b > P_a$, Tag2=1; or Tag2=-1;

calculate Tag=Tag1+Tag2;

If Tag=2, let $D_b = D_c$, and make the operating point move to C;

If Tag=2, let $D_b = D_a$, and make the operating point move to A;

If Tag=0, The operating point is motionless and the maximum power point is found. When external environment changes fiercely, Tag=0, it still regards it as the maximum power point and does not make any movement.

4 MPPT Improvement

It is known that hysteresis comparison method is able to eliminate the oscillatory occurrences and the voltage collapse phenomenon from the analysis of several kind of control algorithm, but it requests the three spots A, B, and C are nearby the maximum power point. The structure of the constant voltage method is simple, the cost is low and could find the maximum power point fast, but the tracking accuracy is insufficient. One better algorithm is synthesizing the merit of these two methods. In the earlier period it uses the constant voltage method to find the maximum power point rapidly and in the later period uses hysteresis comparison method to enhance control precision. Such system has the most superior starting characteristic; the toggle speed

is quick and steady; the power increases monotonously,; it will not emerge the fluctuation phenomenon.

In order to adapt the changement of the external environment fast, the timer setting function is used to skip to the constant voltage method periodically to find the maximum power point. The flow diagram is as Fig. 9.

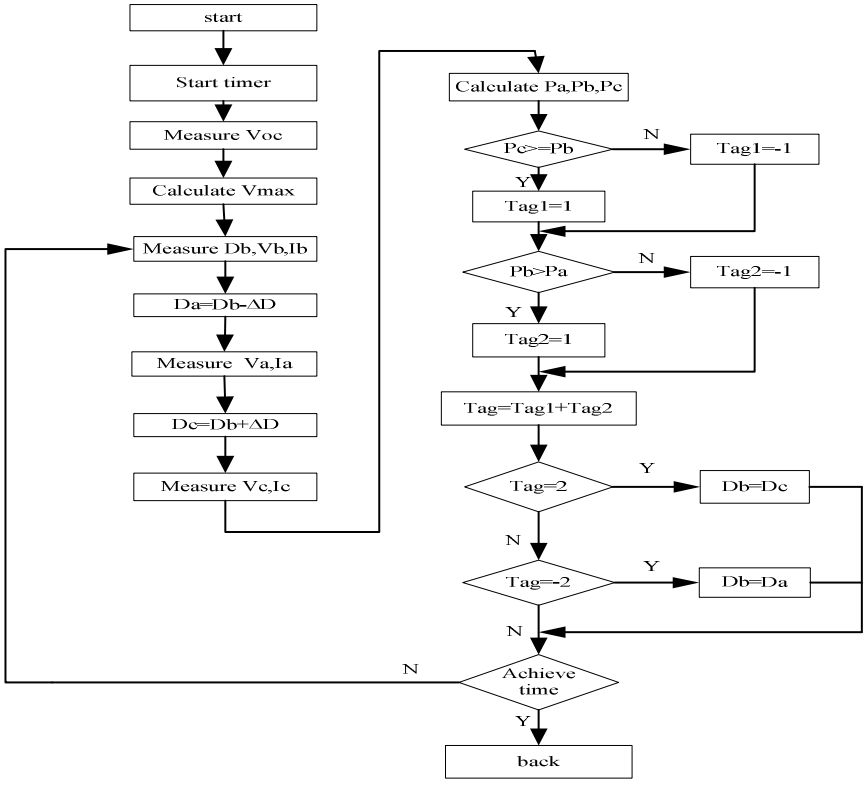


Fig. 9. Flow diagram of improvement control algorithm

5 Conclusion

This article carrys on the research to the maximum power point tracking of the photovoltaic system using the improvement control algorithm. The improvement control algorithm could find the maximum power point fast when the external environment changes fiercely. It has the merit of simple control policy, smooth startup process, quick tracking speed and high tracking accuracy. It eliminates oscillatory occurrences and the voltage collapse phenomenon of the perturbation and observation method completely, and has the great superiority compared with other algorithms.

References

1. Park, S.-S.: Department of Electrical Engineering, Changwon National University, Korea, Republic of; Jindal, A.K., Gole, A.M., Park, M., Yu.: In-Keun Source: Canadian Conference on Electrical and Computer Engineering, CCECE 2009, pp. 720–724 (2009)
2. Wei, Z.: Solar photovoltaic system's research. Hefei Industrial university doctorate paper, 25-44 (2003)
3. Femia, N., Petrone, G., Spagnuolo, G., Vitelli, M.: Optimization of Perturb and Observe Maximum Power Point Tracking Method. IEEE Transactions on Power Electronics 20(4), 963–973 (2005)
4. Mutoh, N., Matuo, T., Okada, K., Sakai, M.: Predicion-Data-Based Maximum Power Point Tracking Method for Photovoltaic Power Generation Systems. In: IEEE 33rd Annual Power Electronics Specialists Conference.PESC 2002, June 23-27, vol. 3, pp. 1489–1494 (2002)
5. Won, C.-Y., Kim, D.-H., Kim, S.-C., Kim, H.-S.: A New Maximum Power Point Tracker of Photovoltaic Arrays Using Fuzzy Controller. In: 25th Annual IEEE Power Electronics Specialists Conference. PESC 1994 Record, June 20-25, vol. 1, pp. 396–403 (1994)
6. Fei, G.: The MPPT controller's research and the design of Solar photovoltaic system. Zhejiang University master's degree paper, 26-27 (2007)

A Quality Control Method Based on Data Mining Technology

Li Jun[1,2], Kan Shulin[1], and Liu pengyu[1,2]

[1] School of Mechatronics Engineering and Automation, Shanghai University, 200072, Shanghai, China
[2] Taizhou University, 318000, ZheJiang Taizhou, China
lijunlh@shu.edu.cn

Abstract. For enhances the control method efficiency in the quality control, the causes the production process to be in the control area as far as possible, enhances product the qualified rate. This article proposed quality control method based on data mining, and proposed the main body cooperates the data mining, introduced data mining's theory, the excavation process and the excavation method, finally through the experiment, confirmed compared based on the data mining quality control method with traditional the statistical quality control to have the archery target enhancement, to a great extent raised the efficiency which the enterprise produced.

Keywords: data mining, quality control, statistical quality control, main body assists, experiment analyzes.

1 Introduction

The quality control is to achieve the work technology which the quality requirement adopts and the activity. That is, the quality control is for through the surveillance quality forming process, eliminates on the quality link all stages to cause unqualified or not the satisfactory effect factor, achieves the quality requirement, the gain economic efficiency. Along with data warehouse and data mining technology starting, the people use from the massive quality historical data withdraw the knowledge to be able the better control online production, compared the based on data mining's method and original use the statistical quality control to have the archery target distinction, to a great extent raised the efficiency which the enterprise produced. The conventional routes already could not adapt under the new situation quality control, becomes the restriction enterprise development the bottleneck, therefore under the new manufacture pattern inquired about that the new quality control method, and realizes the higher efficiency quality control to have the very vital significance.

2 Data Mining Theory

2.1 Data Mining Concept

The data mining [1] is from massive, incomplete, has the noise fuzzily, in the stochastic practical application data, extraction concealment, the people do not know

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 511–517.
springerlink.com

beforehand, but is also the latent useful information and the knowledge process. Said simply is withdraws or "the excavation" from the mass data the knowledge.

2.2 Data Mining Knowledge's Process

The data mining process generally needs to experience; Data preparation, data mining, result expression and explanation. Data mining primary mission [2] May divide into two kinds: Description and forecast. The descriptive data mining portray data's general characteristic, the forecasting data mining is carries on the inference in the current mission, carries on the predict that like Figure 1 has demonstrated several kind of basic data mining duty:

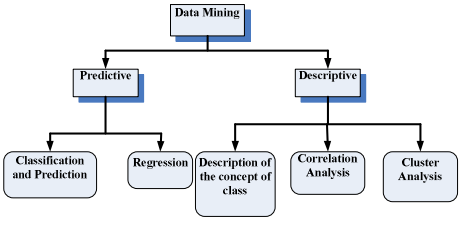


Fig. 1. The duty of data mining

Its use's main method has the classification and the predict and the return, concept/kind of description, the connection analysis, the cluster analysis.

3 Quality Control Technology Research

The quality control has experienced three stages approximately; first, the performance test stage, second, statistical quality control stage, third, Total Quality Management stage, corresponds to the above three quality control stage, the quality control technology also develops from the most universal statistical quality control to based on the intelligence information processing control method, The quality control most main goal is enhances the control method the efficiency, the objective is enhances the control method the efficiency, causes the production process to be in as far as possible the control area, enhances product the qualified rate. What this article main research is the control technology, but non-quality control entire flow. At present mainly concentrates to its technology's research in two kinds of technologies.

3.1 Statistical Quality Control Technology

Statistical quality control (SQC) is the present application scope broadest quality control method; its research is also mature. It uses the product the qualitative index to follow normal distribution this kind of rule to carry on the control. The control method is uses the control chart [3] general control chart principle as shown in Figure 2.

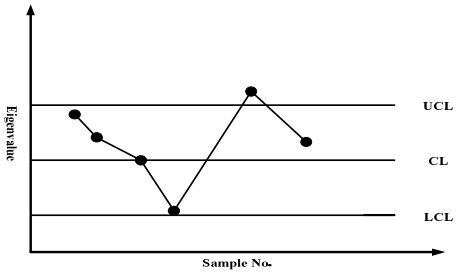


Fig. 2. The control of the figure

The statistical quality control engineering research's primary coverage counts the technology, how to raise the statistical rate of accuracy to draw up the closer actual production process the control chart, uses the main method is the mathematical statistic method.

3.2 Quality Control Technology Based on Data Mining

The data mining is refers to from the data discovers usefully, hideaway pattern process. Advance which produces along with the enterprise, accumulates the massive historical data, based on data mining technology knowledge discovery characteristic, the people then apply this kind of technology in the quality control, unearths from the historical data has the value quality control information to use in the following production control, this is based on the data mining quality control technology. This kind of technology's core is the data mining technology. Has the essential difference based on data mining's quality control technology and the statistical quality control, it does not need to establish the control rule beforehand, the control rule withdraws directly from the quality historical data, avoided establishing not the flexibility beforehand. The data mining is one kind of relatively mature technology, at present the people mainly concentrate to this kind of controlling force method research in the data mining in the quality control application. In view of controlled member's characteristic, selects the appropriate excavation method, realizes the concrete quality control.

4 Data Mining Method Based on the Main Body Assists

The data isomerism which needs the domain expert in view of the tradition data mining to participate under the limitation which as well as the new manufacture pattern possibly exists the difficulty which brings for quality control, this article introduces the main body to solve this kind of problem.

4.1 Definition of Main Body

To solve in the knowledge integration the applied technology multiplicity, questions and so on unity and standard, the main body development has brought the hope for the question solution. Main body's called the entity; it carries on the decomposition

to objective world's thing, discovers its basic constituent, then research objective things abstract essence. In recent years, the main body concept by more and more applications in the computer knowledge 1 two regulation domain, uses to the objective world existence reality carries on the systematization to describe, convenient knowledge entrusting with heavy responsibility with alternately.

4.2 Main Bodies Assist Data Mining

The main body may assist the data mining from two aspects advance; first, based on excavation method main body [4]; second, based on excavation object main body [5]. based on main body data mining process as shown in Figure 3.

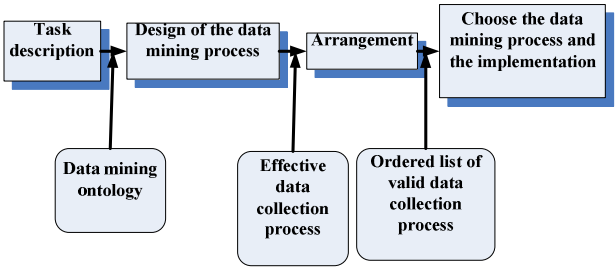


Fig. 3. Main bodies assist data mining process

First, the data mining worker through obtains the data with the user exchange, the ultimate objective must unearth as well as anything is the related information which the user is most urgently needed knew. Second step, the restraint which conceals according to the user input, the main body which the data characteristic and already existed obtains all effective DM process set (each effective DM process is may implement carries out plan). in this in step including How to choose the suitable data pretreatment, how to choose the appropriate data mining algorithm and to carry on optimized, the visualization model operation to the excavation result. Then, the demand order which is urgently needed according to the user obtains to the second step institute carries out the plan set to carry on the arrangement, forms one to be possible to carry out the plan detailed list, Thus, the user may choose appropriately in this plan detailed list carries out the plan. Last step, chooses in the detailed list the plan and carries out.

5 Data Mining in Quality Control of Assistance Based on Ontology

Production activities, quality control is a very important part, only by improving the efficiency of quality control of products, the efficiency of enterprises in order to fundamentally improve. Now widely used for enterprise statistical quality control (SQC) methods of abuse, combined with most of the current quality control based on data mining methods are still used after the extraction efficiency of control lag v. disease,

presents a more efficient quality control. And the general production process for the establishment of a quality control model.

5.1 Quality Modeling

Quality modeling to determine the substance of both control parameters, in general, the general production process of products is divided into three parts, one is the input part, that is, the production line of raw materials, quality control of the middle part, the last part of the output of the product, it is shown in Figure 4, the quality of the purpose of modeling is to determine how to set determine the QCM, but also how to set control parameters only makes the end product is qualified. This model is based on the quality of data mining, basic idea is to first select the object modeling of historical data, after preprocessing, and then extract the information contained and quality of knowledge related to the formation of control rules.

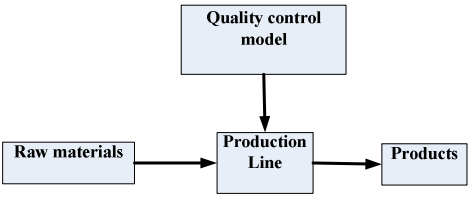


Fig. 4. The quality of modeling diagrams

5.2 Turning the Quality of Flow Based on Data Mining Thread

Quality modeling method is using the first quality control of turning thread model, and through the model test. CNC lathe with CJK0635A object, CJKO635A turning screw is used in quality control of statistical quality control method, the parameters of the distribution of turning thread statistics, the parameters input to the Department in the production of CNC machine tools.

It is using the quality modeling method, the union general turning thread's production line, carries on the knowledge excavation based on data mining's way to the turning historical data, is as follows based on this kind of control thought's control chart:

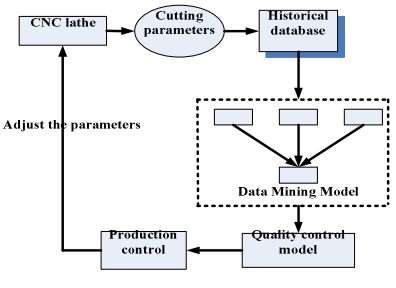


Fig. 5. It is turning thread quality control system based on data mining

It uses the above control method, glass's processing parameter each time stores the historical database, applies the rough collection in the historical database's foundation to carry on the extraction to the control rule, The controlled variable feedback which obtains for lathe's control system carries on the reset to achieve again uses the new parameter to carry on the control the function.

6 Experiment Analyses

6.1 Data Selections

Union this article proposed the theory method, the tentative data elects, from to corresponds in the CJK0635A numerical control lathe's quality database data. The experiment take Microsoft Windows the Server2003 operating system as the platform, the database picks Microsoft Sqlserver 2000, the main body edition tool selects Prot6g63.1.l. The data inducts selects tool DTS, first establishes the ODBC data driven for the data file, the data object is the DBC database file, but in system, so long as has installed the VFP software, has Visual FoxPro Tables the ODBC data pool, found windows the data pool supervisor, disposed the table of contents which the Visual FoxPro Tables table of contents was at to the data file.

6.2 Data Mining Using Rough Sets

In accordance with the method of rough set reduction of two-dimensional table, decision rules expressed using the IF-THEN expression, the second column of Table 2. Training results in Table 1, the third and fourth columns.

Table 1. Decision Rules

No.	Rules	Detection record	Percentage
1	IF(Fl=Al6061)AND(F3=.2)THEN(D=l)	102	16
2	IF(Fl=A1 6061)AND(FS=.017)THEN(D=l)	91	15
3	IF(Fl=AI7075)AND(FS=600)THEN(D=l	125	20
4	IF(FI=Al 7075)AND(FS=.005)THEN(D=l)	75	12
5	IF(Fl=Fe 4140)AND(F2=1200)AND(F8=300) THEN(D=O)	7	17
6	IF(FI=Al 6061)AND(F8=600)THEN(D=O)	8	20

6.3 Assessment

Select from the historical data 333 (l9 a defective) of data as test data, the results shown in Table 2. It can be seen. And objective decision making model the pass rate has been very similar, reflecting the good quality control results.

Table 2. Models the Classification Capability of the Test Record

Test set	Test Project					
333 data (314 qualified, 19 defects)	Decision Rules	1	2	3	4	5
	Related to feature set	F1,F3	F1,F5	F1,F8	F1,F5	F1,F2, F8
	Percentage	16%	15%	20%	12%	17%
	Pass rate model	94.9%	100%	70.1%	88.9%	94.6%
	Pass rate	95.4%	100%	71%	89.3%	95.3%

This paper selects the control and the present model of statistical comparison of control results in Table 3.The model established by this qualified products and substandard products on the classification accuracy rate is significantly higher than in SQC, and the fluctuation range of smaller than the SQC.

Table 3. Quality Control Methods and Compare the Effect of SQC-based Data Mining

Control	Test data	Training data	Features re-centralization	Accuracy	Error rate
SQC	8/3333	no	no	82.12±4.4777	17.88±4.47
This model	3/(667+333)	yes	no	95.05±0.8333	4.96±0.83

7 Summaries

In this paper, quality control, the introduction of the data mining technology, and proposed data mining based on ontology, describes the process of its implementation, Validated by experiment based on data mining for the quality control compared to the original use of statistical quality control has a qualitative difference, it has greatly improved production efficiency.

References

1. Han, J., Kamber, M.: Data Mining Concepts and Techniques.Fan Ming, X. Meng Such as translation, Machinery Industry Press. 8, (2001)
2. Fayyad, U.M., Piatetsky-shapiro, G., Smyth, P.: The KDD Process for extracting useful Knowledge from volumes data. Communication of the ACM 39, 27–35 (2006)
3. LI, X.-c.: Quality Management. Beijing University of Posts and Telecommunications Press (2003)
4. Xing, P.-p., Shi, P.-f., Zhao, Y.: Ontology-based data mining method. Computer Engineering 5(27), 15–16 (2001)
5. Zou, L-p., Wang, L-z., Yao, S-w.: Ontology of data mining method. Computer Science 32 (2005)

Two Switched-Current Memory Cells for High-Speed Digital Communication System

Li Cheng, Ning Yang, Ming Yan, and Xiaodi Zhou

Institute of Electricity and Information, Jiangsu University, Zhenjiang 212013, Jiangsu Province, P.R. China
lcheng@ujs.edu.cn, yangning7410@163.com, yanming321s@126.com

Abstract. Two new switched-current (SI) memory cells suitable for high-speed digital communication circuits are introduced. A kind of novel BiNMOS technology is introduced in the first switched-current memory cell, and a high-speed fully differential BiCMOS structure is introduced in the second switched-current memory cell. Besides, measures of deducing the delay time formulae, optimum seeking the Parameters, and reducing power consumption have been carried out. All the results of hand calculations, SPICE simulations and measurement indicate that with a bipolar supply of 2.0 V~4.0 V, the proposed cells achieve a good comprehensive properties index-delay-power product (DP). The DP of the proposed cells achieves 12.9 pJ, which is about 16.5 pJ lower than the DP of CMOS SI memory cell AD585. Applications of the two new cells in a high-speed digital communication circuits are demonstrated.

Keywords: switched-current memory cells; BiCMOS technology; delay-power product; high-speed; digital communication circuits.

1 Introduction

Switched current (SI) memory cells (known as current copier cells [1]) initially conceived to overcome the inherent matching limitations of continuous time current mirrors [2, 3] are the simplest analog memory building blocks that can be realized in purely digital CMOS technologies. They became basic building blocks for many useful circuits such as dynamic current mirrors, accurate current dividers, A/D and D/A converters [4, 5]. So there is considerable interest in the research of SI memory cells. A low-power low-mismatch low-glitch class AB first-generation switched-current memory cell is presented in [6], and a low-voltage, low-power, low switching error, class-AB SI memory cell is proposed in [7]. The circuit decomposes the input signal into two components by a low-voltage class-AB current splitter and subsequently processes the individual signals by two low switching error class-A memory cells. But all the designs above adopt only CMOS technology. The CMOS devices always keep on alternatively under the static state. So the drive current only exists in the switching of movements. Thus, the switching speed of the current is very low. In order to meet the demand of high switching speed in information systems of digital communication, two kinds of novel high-switching speed SI memory cells are proposed. Combine the advantages of BJT with CMOS, a BiCMOS technology to improve the switching

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 519–524.
springerlink.com

speed and power consumption is adopted, because the switching speed of BJT devices is much higher than the CMOS devices but the supply voltage and the power consumption are very larger than the CMOS devices.

2 Designs of BiNMOS/BiCMOS SI Memory Cells

2.1 BiNMOS SI Memory Cells

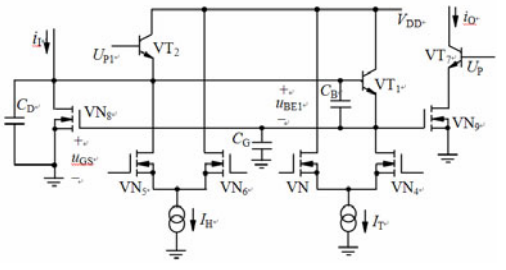


Fig. 1. BiNMOS switched-current memory cells

The complete circuit schematic of the proposed BiNMOS SI memory cell is shown in Fig.1. In which VT_1 and VT_2 are switching devices, VN_3, VN_4 and VN_5, VN_6 form two CMOS differential pairs respectively. If the circuit is in the sampling mode, VT_1, VN_4 and VN_6 will turn on, and the current i_I will charge the CG through VT_1. This time the grid-to-source voltage u_{GS} of VN_8 rise. When value of u_{GS} beyond the threshold voltage of VN_9, VN_9 turns on. This time the output current i_O flows across VT_7 and VN_9, and $i_O \approx K_i I$ (where K is the proportional coefficient, and $1.4 \le K \le 2.5$). Thus i_I is slightly amplified. If the circuit is in the holding mode, VT_2, VN_3 and VN_5 will turn on, and the potential of VT_1's base u_{B1} is lower than the value of $u_{GS} + u_{BE1}$. So u_{B1} is lower than the potential of the emitter. Then the VT_1 turns off and holds on the i_O.

2.2 Differential BiCMOS SI Memory Cells

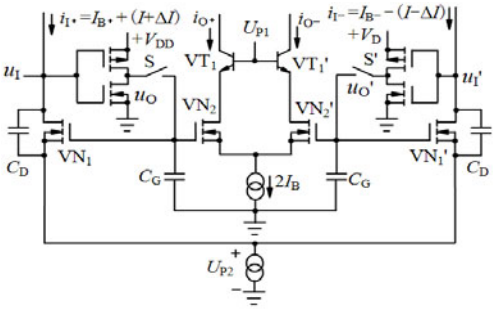


Fig. 2. Differential BiCMOS switched-current memory

The complete circuit schematic of the proposed BiCMOS SI memory cell is shown in Fig.2. In which the differential pair is powered by the current source I_{B+} and I_{B-}. As the differential pair has a fast switching speed and the currents can be added up easily, so the dependence of the differential pair's output current on input current is weakened, and the feedthrough of the holding mode decreases. When the circuit is in tracking state, u_{BE} of all the BJTs are equal, and the collector current of all the BJTs are also equal. In Fig.2 the differential currents i_{I+} and i_{I-} are the functions of input voltage u_I and u_I' respectively. When the circuit is in the holding mode, the drain currents of VN_1' and VN_2 are added up, and the drain currents of VN_2' and VN_1 are

added up. Then i_{I+} and i_{I-} are independent of u_I and u_I', while $i_{O+}\approx Ki_{I+}$, $i_{O-}\approx Ki_{I-}$ (where K is the proportional coefficient, and $1.4\leq K\leq 2.5$).

3 Measures for Speed Increasing and Voltage Decreasing

3.1 The Deduction of the Time Delay Formula

The formula of average time delay tPD is deduced by analysing the circuit schematic of BiCMOS SI memory cell in Fig.1. After being analysed, the formula (1) and (2) can be written to express the time delay tPLH and tPHL in the pull-up and pull-down process of SI memory cell respectively:

$$t_{PLH}\approx C_D U_{BE1}/I_{PLH1}+[C_B C_G/(C_B+C_G)]V_{DD}/(2I_{PLH2})\tilde{}. \quad (1)$$

$$t_{PHL}\approx C_D U_{BE1}/I_{PHL1}+[C_B C_G/(C_B+C_G)]V_{DD}/(2I_{PHL2})\tilde{}. \quad (2)$$

Where C_B and U_{BE1} are average stray capacitance of the base-to-ground and voltage drop of the base-to-emitter in VT_1 respectively, C_G and C_D are average value of grid capacitance and drain capacitance of NMOS devices respectively. V_{DD} is the power supply. I_{PLH1} and I_{PLH2} are average charging currents in the pull-up process. I_{PHL1} and I_{PHL2} are average discharging currents in the pull-down process. It can be seen from the formula (1) and (2) that the speed of the SI memory cells is determined by charging and discharging of the electrode capacitance in pull-up and pull-down processes. Because the connection mode of BiCMOS circuit's electrode capacitor in Fig.2 is similar to the connection mode of BiNMOS circuit's electrode capacitor in Fig.1, the delay time formula deduced from the Fig.2 is similar to the formula (1) and (2).

Besides, it can be seen from the formula (1) and (2) that V_{DD} is a part of the numerator in the second term, so the lower V_{DD} can make the time delay smaller.

3.2 Contractive Analysis in the Time Delay of the Two Memory Cells

The SI memory cell in Fig.1 is also the object of research so as to simplify the analysis. When the switch devices VT_1 and VN_8 are in the turn-off state, then the current conducting channel will does not exist and the total equivalent capacitance will be determined by C_B. When the VT_1 and VN_8 are in the transition state, the inversion layer formed in VN_8 act as the channel between drain and source. So the substrate is shielded by the channel at the terminal of grid, and $C_G\approx 0$. When the VT_1 and VN_8 are in the turn-on state, the channel of VN_9 is pinched off and the capacitance between grid and source is almost zero. The capacitance C_G between grid and substrate is also about zero. Because the value of C_G in the three states above is about zero and C_G is in series with C_B, the equivalent capacitance is also about zero and the formula can be simplified as follows:

$$t_{PLH}\approx C_D U_{BE1}/I_{PLH1}\tilde{}. \quad (3)$$

$$t_{PHL}\approx C_D U_{BE1}/I_{PHL1}\tilde{}. \quad (4)$$

So the time delay t_{PLH} and t_{PHL} in the pull-up and pull-down process are determined by C_D (because of U_{BE1}=0.6 V). Compared with the CMOS SI memory source AD585 (in sample-hold mode), in AD585, the circuit where C_G is in series with C_B does not exist, but there are several C_G connected in parallel. On the contrary, a topology of the circuit in which C_G is in series with C_B is included in the proposed SI memory cells. So the time delay of proposed SI memory cells is very small according to the estimating formula above. Then the estimating formula of the average time delay t_{PD} is that:

$$t_{PD}=(t_{PLH}+t_{PHL})/\tilde{2}. \tag{5}$$

3.3 Optimizing of the Parameters of the Devices

Table 1. Parameters of the BiCMOS SI memory cell

Technology/Transistors	Parameters
Technology	0.35 μm BiCMOS
Fabrication	Double well CMOS+NPN poly-silicon emitter transistor
BJT	A_E=0.85×4.0 μm2 f_T=15 GHz R_E=15 Ω R_C=20 Ω R_B=53 Ω β=20~30
MOS device	0.42 μm（NMOS） 0.50 μm（PMOS） U_{TH} (according to VDD) Tox=100 nm

According to the formula (3) and (4), the optimal parameters in BiCMOS SI memory cells are shown in Table 1. The purpose of optimizing the parameters of the devices is to guarantee the low time delay and small power consumption of the proposed BiCMOS SI memory cells.

3.4 Preparation of the Devices for Speed Improving

BJT devices: The key of the emitter technology is to maintain the enough high current gain and reduce the RE;The extended pressure results in increasing the R_E under the high forward current density, so the reliability of the emitter should be taken into account; The technology to reduce the capacitance C_{CB} of J_c include adopting double-well polycrystalline silicon emitter structure, reducing the junction area of extrinsic J_e and reducing the doping concentration of the n-type S_i below the extrinsic J_e.

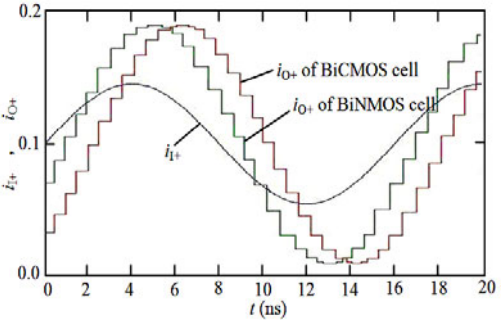


Fig. 3. The Input-output transfer characteristics of the tow kinds of BiCMOS switch SI memorie cells: BiNMOS memory of Fig.1 (K=1.9, sample mode)； Differential BiCMOS memory of Fig.2 (K=1.9, sample mode)

CMOS devices: Si surface silicified treatment technology is adopted to reduce the sheet resistance and enhance the ohm contact; In order to control the short-channel effect, an injection technology which adds the halo on nonuniform (shrinking) channel is adopted; The shallow trench isolation process with minimal space consumption is used to isolate the CMOS devices; Reducing R_{DS}, C_D, t_{PLH} and t_{PHL} at the same time is carried out according to the formula (3) and (4).

4 Experimental Results and Discussions

4.1 Date of the Simulation and Hardware Experiments

Simulation and hardware circuit experiments for the proposed BiNMOS/BiCMOS SI memory cells are implemented respectively, and the hardware is implemented in Jiangsu electronics key laboratory. The tools for simulation and hardware circuit experiments are PSpice8.0 and Agilent electronic test instrument respectively. The simulation waveform which reflects the relationship between the input and output is shown in Fig.3. It can be seen easily from Fig.3 that the time delay of output signal relative to input signal is about 2-3 ns, and the waveform of output signal always follows the input signal without any distortion. Table 2 compares the proposed BiNMOS/BiCMOS SI memory cells with previous CMOS SI memory cell (taking the AD585 as the representative.). The test condition is that all the SI memory cells work in a sample-hold mode. The result of hardware experiment shows that the both proposed IS memory cells can work in the power supply V_{DD}＝3.1 V, the time delay t_{PD} has reduced to 2.0 ns-3.1 ns and the power consumption P_D has reduced to less than 4.7 mW.

Table 2. Simulations and experiments results of CMOS/BiNMOS/BiCMOS memories（f＝60 MHz，VDD＝3.1 V）

The different experiments memories	Av. power consumption PD at 100 MHz /mW		Av. propagation delay t_{PD} / ns		Delay-power product DP / pJ		Drive current i_O / mA
	Sim.	*Exp.*	*Sim.*	*Exp.*	*Sim.*	*Exp.*	
CMOS memory of AD585	1.1	1.2	26.7	27.3	29.4	32.8	0.1
BiNMOS memory of Fig.1	4.0	4.1	2.0	2.3	8.0	9.4	1.4
BiCMOS memory of Fig.2	4.6	4.7	2.8	3.1	12.9	14.6	1.5

Sim.: Simulation; Exp.: Experiment.

5 Relationship among Delay-Power Product *DP*, Power Supply V_{DD} and Time Delay t_{PD}

From the analysis in [8] it is easily to notice that the delay-power product DP reflect the comprehensive performance of circuit. In order to compare the performance of the CMOS, BiNMOS and BiCMOS SI memory cells, the relation curves between DP and

V_{DD} of the three memory cells have been carried out. The result in Fig.5 shows that under the same condition of V_{DD}=2.0 V~4.0 V, the DPs of the BiNMOS and BiCMOS SI memory cells are deeper than the CMOS SI memory converter. When the V_{DD} is about 3.04 V, DP of the three SI memory cells achieve minimum value. But when V_{DD} is less than 2.0 V or more than 4.0 V, the DP increases gradually, and the advantage goes down. It can be seen clearly that when 2.0 V$<V_{DD}<$4.0 V, DP of this BiNMOS/BiCMOS SI memory cells are about 21.4 nJ and 16.5 nJ deeper than CMOS SI memory cell. All the results show that the designed SI memory cells have some advantages both in speed and in the comprehensive properties index DP. So they belong to high speed and high performance SI memory cells.

6 Conclusion

Two BiCMOS SI memory cells were introduced and analyzed. The advantages in high speed and low power supply of the BiNMOS/BiCMOS SI memory cells are shown in Pspice 8.0 and hardware experiments. The reasons are that: The parametric optimizations of the components have been worked out; the complementary advantage between single-polar and bipolar devices is worked out through adopting the BiCMOS technology. The experiments show that the proposed SI memory cells can work under the power supply of 2.0 V$\leq V_{DD}\leq$4.0 V, the time delay tPD of it is 24.2 ns smaller than similar CMOS SI memory cell AD585, and the power consumption PD is only 3.5 mW larger than that of AD585. So the DP of the proposed SI memory cells is about half of AD585's and the drive current come from the output of it is about 14 times larger than that of AD585.

References

1. Daubert, S.J., Vallancourt, D., Tsividis, Y.P.: Current copier cells. Electron. Lett. 24, 1560–1562 (1988)
2. Groeneveld, W., Schouwenaars, H., Termeer, H.: Self calibration technique for monolithic high-resolution D/A converters. In: Dig. Tech. Pap. IEEE Int. Solid State Circuits Conf., vol. 32, pp. 22–23 (1989)
3. Hughes, J.B., Bird, N.C., Macbeth, I.C.: Switched currents-A new technique for analog sampled-data signal processing. In: Proc. IEEE Int. Symp. Circuits Syst., vol. 3, pp. 1584–1587 (1989)
4. Sawigun, C., Serdijn, W.A.: 0.75V micro-power SI memory cell with feedthrough error reduction. Electron. Lett. 44, 561–562 (2008)
5. Wang, C.Y., Ahmad, M.O., Swamy, M.N.S.: Design and Implementation of a Switched-current Memory Cell for Low-power and Weak-current Operations. IEEE J. Solid State Circuits 36, 304–307 (2001)
6. San-Um, W., Srisuchinwong, B., Tantaratana, S.: A low-power low-mismatch low-glitch class ab first-generation switched-current memory cell and its applications. In: IEEE Reg. 10 Annu. Int. Conf. Proc. TENCON, pp. 258–261 (2004)
7. Sawigun, C.W., Serdijn, A.: Low-voltage, low-power, low switching error, class-AB switched current memory cell. Electron. Lett. 44, 706–708 (2008)
8. Cheng, L., Wang, Z.Y., Zhang, B.: 3 Kinds of Low-voltage, High-speed and Low-comsumption BiCMOS Tristate Logic Gates. Guti Dianzixue Yanjiu Yu Jinzhan/Research and Progress of Solid State Electronics 26, 166–170 (2006)

A Novel High-Performance BiCMOS F/V Converter for Low-Frequency Signal Processing Systems

Li Cheng, Ming Yan, Ning Yang, and Ling Ding

Institute of Electricity and Information, Jiangsu University, Zhenjiang 212013, Jiangsu Province, P.R. China
lcheng@ujs.edu.cn, yanming321s@126.com, yangning7410@163.com

Abstract. This paper presents a novel high-linearity and low-consumption BiCMOS frequency-to-voltage (F/V) converter which is composed of three operational amplifiers (OAs) A1, A2 and A3, in which A1 and A2 are common-source CMOS OAs, but A3 is a low-pass filter (LPF) which adopts the BiCMOS technology. The component parameters of the entire converter are optimized and some measures such as raising speed and reducing power-consumption are taken. Experimental results show that the pulse signal frequency f2 input into the LPF of the converter is equal to the input signal frequency fi, and the output average voltage (Uo) of the converter varies directly with fi. This converter is tunable from 4 Hz to 10 kHz, and its delay-power product (DP) is about 1.09 nJ, and its conversion linearity (LC) is only 1.7× 10-2. With these characteristics, the presented converter is very suitable for low-frequency signal processing systems.

Keywords: BiCMOS analog integrated circuits; frequency-to-voltage converter; high-performance; operational amplifiers.

1 Introduction

A F/V converter is a device that generates an output voltage proportional to the frequency of an input signal. It can convert a signal from alternating frequency to alternating voltage linearly. This device is very useful and has many applications in power system control, in processing of very-low-frequency signals, and in many fields of instrumentation. Because of the strong anti-jamming capability of frequency signal, it is also often used for long-distance transmission. It can be modulated in the RF signal for wireless transmission. Usually, the F/V converters are designed to comply with several requirements such as high speed operation, low-output ac ripples, good linearity and wide frequency range [1~3]. Besides, the application and the principle of a series of integrated F/V converter chip such as LM131, LM2907 and LM2917 were presented in [4~6]. Another new F/V converter which was used to generate coded signals for AC/DC servomechanisms was designed in [7]. Another kind of F/V converters can produce a voltage that is proportional to the frequency of a sinusoidal wave form input signal as proposed in [8]. Besides there are three kinds of high-performance F/V converters presented in [9~11]. In order to meet the performance requirements of intelligent multi-parameter high precision measurement and the high-performance fiber-optic

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 525–531.
springerlink.com

transmission system, a new F/V converter which has low-consumption and high-linearity based on BiCMOS technology is proposed, and the simulation and hardware circuit experiment are carried out.

2 BiCMOS F/V Converter

The complete circuit schematic of the BiCMOS F/V converter composed of A_1, A_2 and A_3 is shown in Fig.1. The voltage comparator composed of A1 convert input signals into square wave signal with the same frequency, and then transmit the narrow pulse with positive pulsatile capability to A_2 through the capacitor C_1 and the diode VD_3. Because the potential u_N in the inverting input is negative before triggering; the potential in the output of A_2 is positive; the VN_1, VN_2 is turning on and the potential u_2 is negative. When the voltage $+V_{DD}$ in the non-inverting input trigger A_2 and shift it to low level swiftly, VN_1 is turned-off and u_2 jumps to high level which is equal to the stable voltage U_Z of the silicon stabilivolt VS. At the same time u_N jumps to the high level U_H correspondingly; then the VN_2 is turned-off and $+V_{DD}$ charges the capacitor C through the resistance R. So the voltage u_P in the non-inverting input of A_2 rises following the exponential law. After the time of T_W, when the voltage u_P increases to U_H, the voltage of A_2 reverse swiftly again to reset. Then the process of the monostability is over. Thus the drain voltage which is the voltage pulses with the width T_W and the amplitude U_Z increases with the rise of the input frequency f_i. According to the principle of superposition, when VN_1 turns off, the potential in the inverting input of A_2 is following:

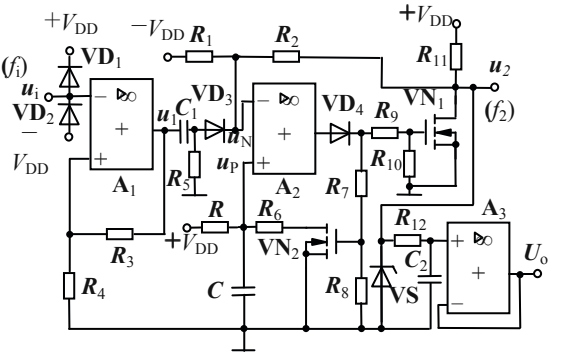


Fig. 1. BiCMOS F/V converter

$$U_H = R_1 U_Z/(R_1+R_2) - R_2 V_{DD}/(R_1+R_2). \quad (1)$$

When the charging interval is up to T_W, $u_P(T_W)=U_H$. Therefore, after calculation, the time charging to T_W of the RC circuit is reached:

$$T_W = RC\ln\{[R/(R+R_6)]\times[(R_1+R_2)V_{DD}]\div[(R_1+R_2)V_{DD}-(R_1U_Z-R_2V_{DD})]\}. \quad (2)$$

In Fig.1, the low-pass filter (LPF) formed by A_3 is also can be considered as voltage follower. So the average voltage in the output of the whole converter is follows:

$$U_o = T_W U_Z f_i. \quad (3)$$

The formula(3)shows that the U_o of the converter is proportional to the input signal frequency f_i, and the ratio of coefficient is $T_W U_Z$.

3 Measures for Speeding Promoted, Power Consumption Reduced and Linearity Improved

3.1 The Improvement of the Circuit

In Fig.1 VD_1 and VD_2 in the input of the F/V converter protect the circuit by limiting the voltage. The R_3 and R_4 accelerate the jumping of the voltage by inducing the positive feedback in the comparator A_1. Because of the backlash, the capacity of resisting disturbance of the comparator A_1 is stronger than other comparator. A3 can be considered as the voltage follower. So the series voltage negative backward is induced to increase input resistance, reduce output resistance, broad band, restrain interference and noise.

3.2 Optimizing for the Parameters of Elements

The supply voltage $\pm V_{DD}=\pm 6$ V; the accuracy class is ±0.001; and the stabilized voltage of VS is 4 V. The silicon reference diode with lower temperature coefficient should be selected. The principles of selecting the parameter of *RC*-components are higher speed, higher linearity, and lower power consumption. After the selective preference, the metal film resistor which belong to the E24 nominal value series with error ±5% and tantalum electrolytic capacitor which have a mall leakage current are worked as the *RC*-components of the converter in Fig.1. The parameters of every *RC*-component are that $R_1=R_4=10$ kΩ, $R_0=R_3=R_5=R_7=R_8=R_9=R_{10}=R_{11}=R_{12}=1$ kΩ, $R_6=0.1$ kΩ, $R_2=75$ Ω, $C_1=C_2=0.1$ μF.

3.3 Optimizing for the Preparation of the Semiconductor Devices

CMOS devices: ① Electrode is made by adopting an impure and self-aligned ion implantation. ② In order to reduce the sheet resistance and strengthen the point-pressure ohmic contact, siliconizing is using to the surface of the devices. ③ Non-uniform channel added with halo injection is adopted. ④ Shallow trench isolation process with minimal space occupation is used to insulate the MOS devices.

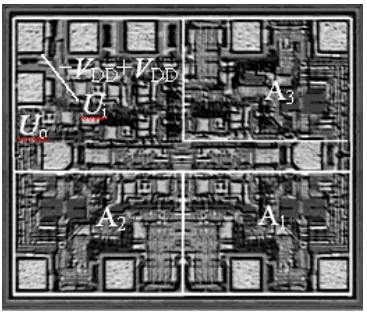


Fig. 2. Microphotograph of the BiCMOS F/V converter

BJT: ① Polysilicon emitter bipolar BJT pair transistor with ultra shallow junction is made. The depth of E junction is scaled down. The width of the base and the base transit time of carriers are reduced. ② The key to the technology of emitter is that an adequate size of the current gain should be maintained and RE should be reduced as small as possible. ③ Considering the reliability of the emitter contact, the contact area between the base and emitter is made small.

Layout: The standard 0.18 μm BiCMOS technology which provide six layers of metal, diversified BJT and MOS models, poly resistors and MIM capacitors is adopted in the

whole chip. Fig.2 is the microphotograph of the BiCMOS F/V converter. The voltage comparator composed by A_1 and the monostable trigger composed by A_2 are in the below two-line boxes of the layout. However the LPF composed by A_2 is in the upward side of the layout. Excluding the weld spot, the whole circuit occupys an active area of 160 μm×120 μm.

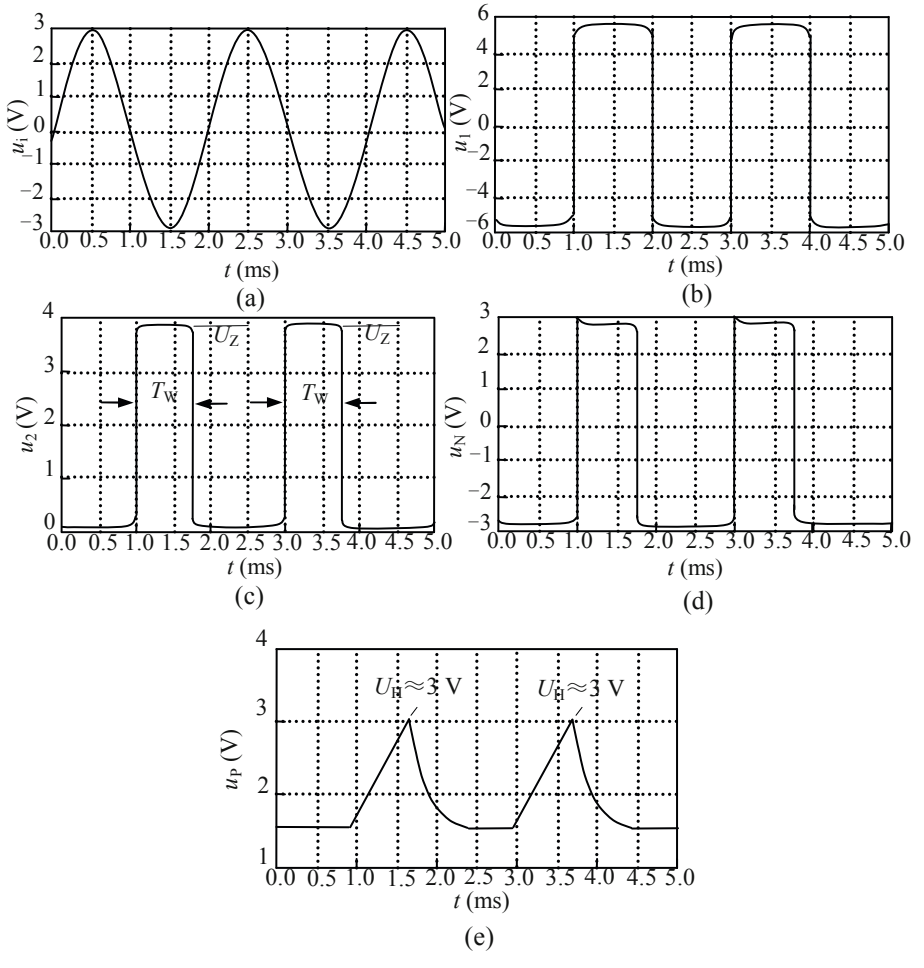


Fig. 3. Waveform of key points in the voltage test. (a) Input voltage of F/V Converter, (b) Output voltage of comparer A_2, (c) Input voltage of LPF, (d) Voltage in the anti-phase input of A_2, (e) Voltage in the in-phase input of A_2

4 Discussions and Analysis of Experimental Results

4.1 Voltage Waveform of Points in F/V Converter

The proposed BiCMOS F/V converter has been implemented in a standard TSMC 0.18 μm BiCMOS technology. The tools for simulation and hardware circuit experiment are PSpice8.0 and Agilent electronic test instrument respectively. The digital oscilloscope is the Tektronix TDS5034B and the signal generator is E4438C made by Agilent Company. At the beginning of the experiment, alternating sinusoidal voltage ui=3sin(1000πt) V is input. Then the key points of the circuit are measured, and the waveforms are obtained. Fig. 3 shows that when the whole F/V converter input a cycle of sinusoidal signal, the monostable trigger A_2 will output a positive pulse with pulse width (T_W≈0.7 ms). This work moves in cycles. Then the drain of VN_2, which is also the input of LPF, will obtain a sequence of positive voltage pulse u_2. The number of the pulses in the pulse sequence is equal to the number of the circles of sinusoidal voltage u_i. That means the input frequency f_2 of the pulse sequence pushed into the LPF equaling to the frequency fi of the input signal is 500 Hz. According to Fig.3(c) and Fig.3(e), it can bee seen easily that U_Z≈3.90 V, and U_H≈3.02 V.These data are consistent with the analysis in the section 2, which shows that u_2 in the circuit has a good sensitivity and accuracy. So like the U_o, the voltage u_2 also can be used to connect load.

4.2 Relationship between Frequency and Voltage

Fig.4 shows that the measured the average U_o of three output voltage versus the input signal frequency f_i under the condition that the temperatures are 60℃, 27 ℃ and 5℃ respectively. To make it easier to read, f_i in the Fig.4 is expressed by abscissa and the decibels of the voltage ratio U_o/U_R is expressed by ordinate under the condition that U_R=1 V, which means the operation 20lg(U_o/U_R) is made on the ordinate. The measured three curves show that the converter can work normally in the frequency range from 4 Hz to 10 kHz. U_o following with f_i grows linearly. However, the three linear curves show that the converter can work normally in a wide temperature range (5℃~60℃) and eliminate the impact on the linearity of the converter from temperature variation.

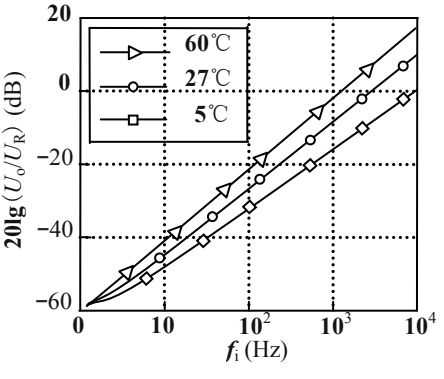


Fig. 4. Relationship between U_o and f_i

4.3 Date of the Hardware Circuit Experiment

Tab.1 shows the experimental date from the three times testing of the BiCMOS F/V converter. So it is easy to see that though the arithmetic mean value of the time delay t_{PD}

Table 1. Main date of this BiCMOS F/V converter

No. of tests	First	Second	Third	AV
3 dB(kHz)	≈10	≈11	≈9	10
Linearity $L_D(\times 10^{-2})$	≈1.8	≈1.7	≈1.6	1.7
P_D (mW)	23.6	23.9	24.5	24.0
t_{PD}(ns)	45.3	45.1	45.7	≈45.4

is 45.4 ns, the power consumption P_D is only 24.0 mW. Then the aggregative performance indicator we call it the arithmetic mean value of delay-power product (*DP*) is that:

$$(23.6\times45.3+23.9\times45.1+24.5\times45.7)\div3\approx1.09\ \text{nJ}. \tag{4}$$

The 3 dB bandwidth means the frequency is about 10 kHz. In order to see clearly, the curves reflecting the relationship between *DP* and supply voltage V_{DD} of the bipolar, CMOS F/V converter in [8, 9] and this BiCMOS F/V converter are plotted by experiment in Fig.5. The experimental result can be achieved only on the condition that the voltage range is from 2 V~8 V. All the results show us that the designed BiCMOS F/V converter has advantages in the aggregative performance indicator *DP*.

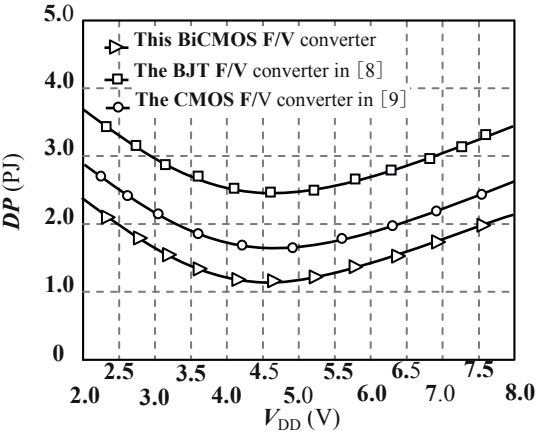


Fig. 5. Relationship between *DP* and V_{DD} of the three F/V converters

5 Summary

A kind of high-performance F/V converter is presented. Both simulation and hardware circuit experiment show that this converter has obvious advantages both in power consumption and in speed. The reason is that ① BiCMOS technology is applied in this paper. The CMOS devices form the main part of the converter. The two BJTs are only used in output stage which is composed of LPF operational amplifier, and they turn on alternately as pull-up and pull-down devices respectively. ② The parametric optimization of the components have been worked out. The positive pulse sequence u_2 with the width T_W and magnitude U_Z is generated by BiCMOS F/V converter, and the value of its output voltage U_Z varies directly with the input signal frequency f_i. The pulse sequence u_2 can connect directly with the interface of computer. This way can take the place of the ADC and occupy less hardware resources. In addition, there is a hysteresis loop in voltage comparator. Because of these reasons, the designed converter has a strong anti-interference ability.

References

1. Stitt, R.M., Burt, R.: Get F-V converters to settle fast without ripple. Electronics Design 39, 73–79 (1991)
2. Laopoulos, T.L., Kqbakas, C.A.: High Performance Frequency-to-Voltage Converter. International Journal of Electronics 66, 303–307 (1990)
3. Hee, S.D., Heyong, C.G.: A Simple High-Performance Analog Frequency-to-Voltage Converter. IEEE Transactions on Industrial Electronics 34, 295–298 (1987)
4. Fang, Y.J., Cheng, J.H.: Detection technologies and systems, pp. 154–160. China Electric Power PRESS (2006)
5. Qi, Y.l., Lu, Y.F., Liu, M.: Integrated F/V converter LM2917 and its application. International Electronic Elements 20, 3–5 (2000)
6. Zhang, Y.G., Yang, G.: Integrated F/V converter LM2005 and its application. International Electronic Elements 15, 71–72 (2005)
7. Zhang, L.X., Yang, Y., Wang, H.J.: F/V circuit of encoder signals for AC and DC servo system. Applied Science and Technology 32, 1–2 (2005)
8. Wanlop, S., Yongyut, C., Suree, B.: An analog sinusoidal frequency-to-voltage converter. IEEE Transactions on Instrumentation and Measurement 40, 925–929 (1991)
9. Papazoglou, C., Siskos, S., Laopoulos, C.A.: Karybakas: A current-mode, CMOS F/V converter for low-frequency applications. In: Proceedings of the IEEE International Symposium on Applications Industrial Electronics, vol. 2, pp. 476–479 (1995)
10. Djemouai, A., Sawan, M., Slamani, M.: New circuit techniques based on a high performance frequency-to-voltage converter. In: Proceedings of The 6th IEEE International Conference on Electronics, Circuits and Systems, vol. 1, pp. 13–16 (1999)
11. Djemouai, A., Sawan, M., Slamani, M.: High performance integrated CMOS frequency-to-voltage converter. In: Proceedings of the Tenth International Conference on Microelectronics, vol. 12, pp. 63–66 (1998)

Novel 0.15 μm BiCMOS A/D Optoelectronic Converter with Schmitt Trigger Circuit

Li Cheng, Decheng Lou, Ming Yan, and Ning Yang

Institute of Electricity and Information, Jiangsu University, Zhenjiang 212013, Jiangsu Province, P.R. China
lcheng@ujs.edu.cn, dclou@126.com, yanming321s@126.com, yangning7410@163.com

Abstract. A novel 0.15 μm BiCMOS A/D optoelectronic converter (OC) was designed, which is composed of optoelectronic receive devices and a Schmitt A/D OC. To improve the gain linearity, accuracy and stability of this proposed A/D converter, negative feedback and bandgap reference were introduced in this circuit, respectively. Additionally, component parameters were optimized, which can enhance the performance of the whole system. The layout of operational amplifier (OA) was designed and its chip area was 0.42 mm×0.32 mm. The simulation and hardware circuit experiments were also given. Test results show that the photocurrent increases with exponentially regularity followed by increasing of the laser power. The converter can use 3.3V power supply, -3 dB bandwidth is about 58.4 kHz and the power consumption is about 69 mW, so the proposed converter can meet the requirement of low-frequency displacement optoelectronic control systems (OCS).

Keywords: BiCMOS; Optoelectronic receive device; Bandgap reference; Schmitt A/D converter circuit.

1 Introduction

With the advantage of the fast arithmetic speed, high reliability and the powerful function for data processing, storage and control, PC is widely used in optoelectronic measurement and control system, especially in the industrial process control and production lines. As computer can only receive the digital signal "0" or "1", while the luminous intensity of the OCS reflects variation of the physical quantity (displacement, rotate speed and pressure etc.), and optoelectronic devices (ODs) convert physical quantity changes to analog signal, therefore, a interface circuit—A/D converter is needed between the ODs and the computer. For example, when producing armor plate, a steel company want to make the plate stack neatly, it uses photoelectric control system based on plate edge location. When the measured plate swung to the left or right edge, the output of photoelectric converter is "1", otherwise is "0". If the plate specification is different, it also needs to adjust the displacement in the length direction. So this case can not only reflect the variation of the displacement and its direction, but also require the A/D conversion for photoelectric signal to control the stacked position of plate accurately. Obviously, the luminous intensity and stability of photoelectronic

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 533–539.
springerlink.com

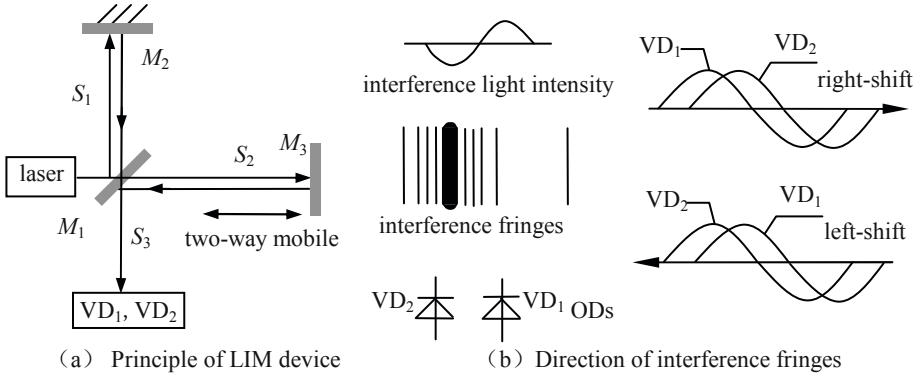
devices directly affect the measurement error. Additionally, the bright-dark of the workshop brightness will also affect the detection of plate edge. So it requires to manufacture a high stability A/D OC.

According to the published literature, it has not been reported about the design of high stability, low power A/D OC[1~6]. With the characteristics of BiCMOS technology—low power consumption and high integration, a novel 0.15 μm BiCMOS Schmitt trigger circuit A/D OC is present.

2 Design of Schmitt A/D OC

2.1 Laser Single-Channel Interference Measurement Systems

In Fig.1(a), firstly, He-Ne laser with frequency stabilization circuit emits beam, when the incidence beam reaches to the viewfinder M_1 (splitter), it is divided into two beams, one is the reference beam (S_1), the other is the measurement beam (S_2). These two beams are reflected back to M_1 through total reflection mirror M_2, M_3 respectively, and it produces interference beam (S_3) in the location of M_1, then the interference fringes are received by ODs VD_1, VD_2, at last pulse analog signal is formed, so the displacement information of M_3 can be carried in this analog signal. Fig.1(a) shows the principle of laser interference measurement (LIM) device: M_2 is fixed and M_3 is installed on the movable measuring head, according to the linear relationship between the movement amount of M_3 and the length of detected object, the length of object can be measured; The optical length of reference beam is not equal to the measurement beam, when the optical length difference Δ is integer multiple of the wavelength λ, i.e., when $\Delta = \pm K\lambda$ (K=0, 1, 2,…), the two beams are in phase, now the light intensity is maximum, when M_1 emits bright spots, ODs can receive this bright spot signal; when $\Delta = \pm(K+1/2)\lambda$, the phase difference of this two beams is π, and the light intensity is zero, when M_1 emits dark spots, without light signal incidence, the output signal of optoelectronic receiver is zero; When the movable mirror M_3 moves along the optical axis



（a） Principle of LIM device （b）Direction of interference fringes

Fig. 1. Single-channel LIM systems

of measurement beam, There will be alternating bright-dark interference fringes in the location of M_1, so the variation of luminous intensity I_V is

$$I_V = I_{OV} + I_{OV} K_1 \cos(2\pi \Delta/\lambda). \quad (1)$$

I_{OV} is the average light intensity, K_1 is the interference fringe contrast. From above formula, when Δ changes λ, the bright-dark of interference fringe will change a cycle. If interference fringe changes n cycles, then $\Delta = n \times \lambda$, interference measurement device as Fig.1 shows, Δ is twice as long as the displacement L of M_3, and the measured displacement:

$$L = n\lambda/2. \quad (2)$$

so if the change cycle of interference fringe is calculated, then the move length of measurement head will be known. Fig.1 (b) schematically shows the interference fringe direction. The bright-dark of interference fringes interchanges with the movement of M_3, it is equivalent that interference fringe is moving. So the output signal of VD_1, VD_2 is approximately sine wave; while the two narrow gaps make phase difference of output sine wave of VD_1, VD_2 reach π/2. If the interference fringes moves to the right, then the phase of output sine wave of VD_1 is ahead of VD_2 π/2; if interference fringes moves to the left, then the phase relationship is contrary.

2.2 Schmitt A/D OC Circuit

As is shown in Fig.2, VD_1 (VD_2) is followed by A/D OC circuit, VD_1 (VD_2) is connected to source output device which is composed of VT, R_P, R_G and R_0. The latter outputs analog voltage signal u_I which is linearly related to the luminous intensity of ODs; u_I is added to the inverting input of OA A. Bandgap reference composed of R_1, R_2, R_3, Q_1, Q_2 and A_1 outputs reference voltage U_R, which is added to the non-inverting input of A through R_4; A, R_4, R_5, R_6, R_7 and two-way regulator VD_Z etc. constitute a Schmitt trigger; positive feedback was introduced through R_4, R_6, the purpose is to accelerate hopping rate of the output digital voltage u_O, and it makes the voltage transmission characteristics have hysteresis, in order to improve the anti-jamming capability of the circuit; VD_Z can restrict the amplitude, it makes positive and negative amplitude of u_O within $\pm U_Z$, after derivation, the threshold voltage of this trigger is:

$$U_{TH} = (\pm R_4 U_Z + R_6 U_R)/(R_4 + R_6). \quad (3)$$

obviously, this trigger has two threshold voltage which are variable.

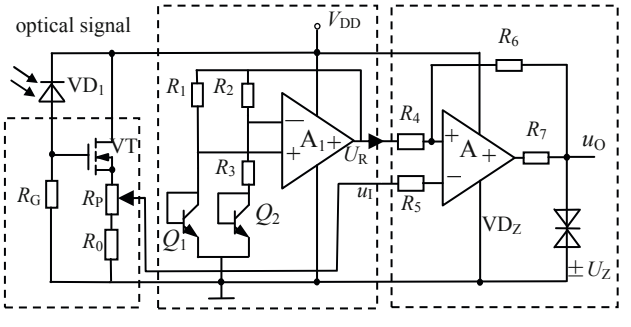


Fig. 2. Variable threshold binarization processing circuit

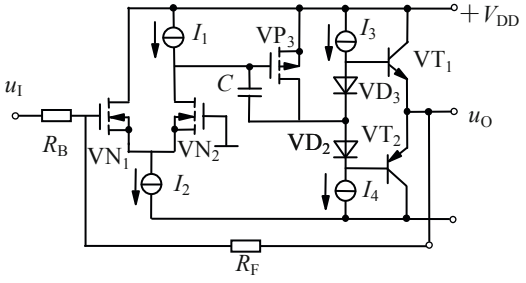


Fig. 3. Internal circuit of the BiCMOS OA A

2.3 Key Design for BiCMOS OA

Fig.3 gives internal circuit of the BiCMOS OA A. It is a differential BiCMOS OA[7]. Source-coupled differential input stage was composed of VN_1 and VN_2; common-source voltage amplification was composed of VP_3, and C is a vibration capacitance, VT_1 and VT_2 are complementary push-pull driver stage. Feedback network is constituted of external resistance R_B and R_F, high input impedance and broadband properties are achieved through differential input stage using insulated gate MOS devices. Push-pull output not only reduces the output impedance, and increases bandwidth[8]. Voltage parallel negative feedback was introduced in R_F. All MOS devices in OA can use the selected BiCMOS parameter from References[9] in Table 1.

3 Design for BiCMOS OA Chip

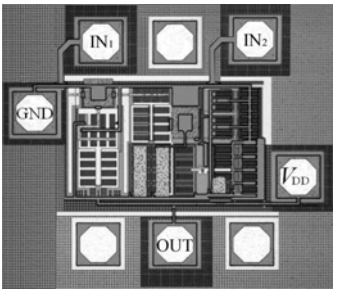


Fig. 4. Design for layout of OA

OC circuit can be achieved by 0.18 μm BiCMOS process using the standard of TSMC company. The preparation method is: OA A is made into a chip singly, or OA A and source-output device are made to co-produce a ASIC. For the source output devices is influenced great by ODs, the design for layout of OA A is finished firstly in the Cadence environment, then the parasitic *RLC* parameter extraction and post-simulation experiment is carried out, The results show that gain-bandwidth product of circuit declines for influenced by parasitic parameter of circuit, so it needs to re-adjust *RC* parameters and optimize layout-wiring. For example, the function block is placed in a reasonable position, so the connection line between each other can be the shortest; the output line is separated from feedback connection to reduce high-frequency interference between them, double-layer metal wiring is used and so on. The chip area of A was 0.42mm×0.32mm in Fig.4. Similarly, the layout of single chip ASIC can be designed and the size is a bit larger than OA A.

4 Simulation and Hardware Circuit Experiments

4.1 Design for BiCMOS OA Chip

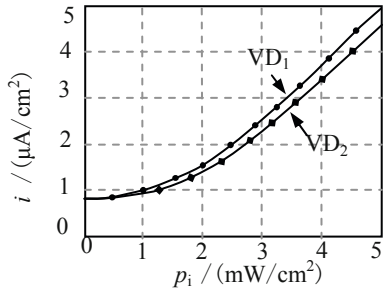
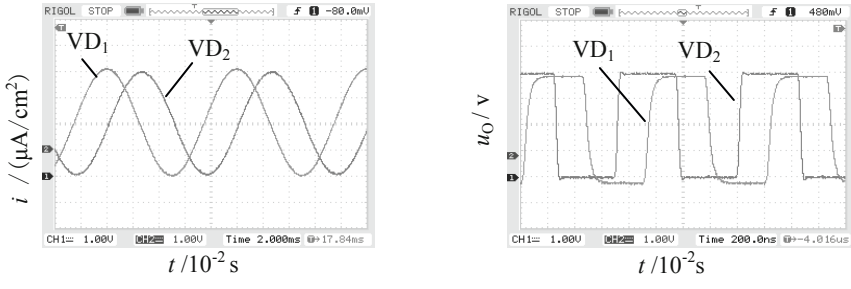


Fig. 5. Curve relationship between current density i and incident optical power p_i

Fig.5 shows the curve relationship between current density i and incident optical power p_i from 1 mW/cm^2 to 5 mW/cm^2, using 570 nm incident light wavelength. The change trajectory of photocurrent i of VD_1 and VD_2 is almost same from the Figure, They are in accordance with exponential increase with the increase of p_i, which shows that once VD_1 or VD_2 exceeds the threshold voltage, even if absorbing only a bit optical power, photocurrent will also increase sharply.



(a) the input voltage u_I (b) the output voltage u_O

Fig. 6. Measured waveforms of the u_I and u_O

4.2 Input and Output Waveforms of Schmitt A/D OC Circuit

Firstly, the input and output voltage test is carried out for OC circuit using the above method, and the supply voltage ±VDD is ±3.3V in test, the external resistor of source output device and schmitt trigger are: R_G=R_6=R_7=4 kΩ, R_0=100 Ω, R_5=R_P=2 kΩ, R_4=1.5 kΩ, the RC parameter of OA are: C=10 pF, R_F=10 kΩ; two sets circuit as Figure 2 shows are used to test the results, the quasi-sinusoidal signal frequency exported from optoelectronic receiver takes 100 Hz, the reference voltage of OC: U_R=0 V, VD_Z is connected with two silicon voltage-regulators 2CW103, and the voltage is restricted at ±2.6 V. MATLAB soft is used to get the waveform of input and output voltage, as is shown in Fig.6. It can be seen clearly that the simulation waveforms and theoretical results are consistent, and the output waveform of ODs is proved to be sine

wave; After OC, the final output voltage is rectangular wave pulse signal, therefore, the functions for optoelectronic signal A/D converter is achieved. Then, it begins to test frequency response feature, circuit and device parameters are ibid, selecting 26 frequencies in the frequency range 0.1~100 kHz are measured, drawing out amplitude-frequency characteristic curve of the measured and simulation results by MATLAB soft, as Fig.7 shows. From the Figure, it can be seen that the results are consistent basically.

4.3 Main Parameters of BiCMOS OC

Table 1 shows the other parameters of BiCMOS OC which are measured by experiment, it proves that the new proposed converter has extremely low power consumption(P_D≈69 mW) , -3 dB bandwidth is about 58.4 kHz, the gain linearity *NF* is 8.3×10^{-5}, the other parameters (e.g. input impedance, CMRR, leakage current) reach a good level too.

Table 1. Measured parameters of the BiCMOS A/D Photoelectric Converter

-3 dB (kHz)	L_D (NF)	Z_I(GΩ)	CMRR	I_D (μA)	±V_{DD}(V)	P_D (Mw)
58.4	8.3×10^{-5}	5.1	0.96×10^{8}	0.21	3.3	69

5 Summary

This proposed BiCMOS Schmitt OC module is very suitable for low-frequency displacement OCS. Its advantages include: ①With the advantages of BiCMOS devices—low power consumption and high integration, it can be designed into OC chip with low power consumption (69 mW) and small device size (0.15 μm) to meet the trend of nano semiconductor devices; ②Under the situation of approximate same parameters, the two selected optical devices can reduce the impact of dark current, and also can reduce conversion error caused by temperature change; ③As the lower level of ±VDD, bandwidth is about 58.4 kHz, it is suitable for low frequency and low pressure. ④ The change of circuit parameters, especially the fluctuation of ±VDD and light intensity change in workshop can affect the stability of signal detection, so negative feedback is introduced in OA A. Additionally, unipolar devices, isolation effect of the source follower and strong anti-interference ability of the Schmitt Trigger are adopted, so the problem about the stability of OC can be solved.

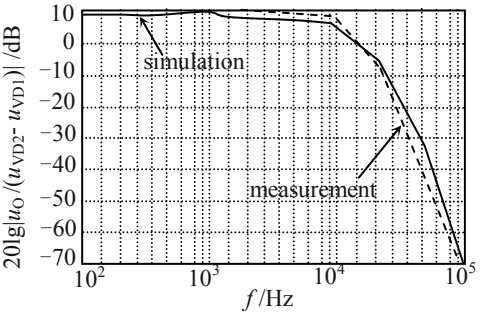


Fig. 7. Amplitude-frequency characteristics of OA A

References

1. Fattaruso, J.W., Sheahan, B.: A 3V 4.25 Gb/s laser driver with 0.4 V output voltage compliance. IEEE Journal of Solid-State Circuits 45, 1930–1937 (2006)
2. Dragoner, V.V.: The photoelectric displacement converte. Proceedings of SPIE 5822, 98–108 (2004)
3. Gorbunov, N.A., Flamant, G.: Qualitative model of a plasma optoelectronic converter. Technical Physics 54, 68–77 (2009)
4. Witte, J.F., Makinwaka, K.A.A., Huijsing, J.H.: A CMOS chopper offset-stabilized op-amp. IEEE JSSC 42, 1529–1535 (2007)
5. Hu, Y., Loge, A., Dulinski, W., Loge, W.: A Low noise, low power BiCMOS preamplifier-shaper for microstrip detectors using DMILL process and irradiation measurements. Nuclear Instruments and Methods in Physics Research, Spectrometers, Detectors and Associated Equipment 423, 272–281 (1999)
6. Thomas, L.F., David, B. , et al.: Fundamentals of Analog Circuits. Prentice Hall/Pearson Press (2004)
7. Plevridis, S., Birbas, A., Triantis, D.: Monolithic amplifier with A GC and differential output for 622 Mbit/s optical communications. In: Proceeding of IEE Circuit Devices and System, vol. 146, pp. 55–58 (2003)
8. Dejhan, K., et al.: A full-swing BiCMOS schmitt trigger circuit design. In: International Symposium on IC Technology. Systems and Applications, vol. 146, pp. 266–269 (1997)
9. Cheng, L., Xu, Z.-C., et al.: Design of a BiCMOS optically coupled isolation amplifier. Bandaoti Guangdian/Semiconductor Optoelectronics 28, 613–616 (2007)

A Kind of Differential BiCMOS Optical Converter with Three Op Amp and Triple Loop Control

Li Cheng, Chao Chen, Ming Yan, and Ning Yang

Institute of Electricity and Information, Jiangsu University, Zhenjiang 212013, Jiangsu Province, P.R. China
lcheng@ujs.edu.cn, chenchaoaichenchao@yahoo.cn, yanming321s@126.com, yangning7410@163.com

Abstract. A kind of differential optical converter was designed, the whole conversion circuit used three op amp amplifier circuits, the voltage series negative feedback was introduced in the first stage op amp A1, A2 which were used CMOS op amp. The voltage parallel negative feedback was introduced in the second stage op amp A3 which was taken into parameter symmetry differential input subtraction device. A3 was designed into BiCMOS op amp, so the whole circuit made up three op amp and triple loop control system. The territory size on the operational amplifier was designed and the chip area was 0.44 mm×0.31 mm, and the simulation and hardware circuit experiments were carried out. These experiments results show that the converter can use 3.0 V~5.0 V power supply, the gain linearity is up to 5.7×10-5, common-mode rejection ratio is 0.9×109. With these characteristics the designed converter is very suitable for high performance optical fiber communication systems.

Keywords: BiCMOS and CMOS devices; differential optical converters; closed-loop systems; ASIC design.

1 Introduction

With the rapid development of computer networks, the universal application of multimedia communication and large-scale construction of information highway, high-performance fiber-optic communications and other optical transform system are needed urgently. The computers have the advantages of high speed operation, high reliability, data processing, storage and control advantages, so it has been widely used in optical measurement and control systems, especially in industrial process control and production pipeline. Simulation of electron converter as an application specific integrated circuit (ASIC), when the measured non-electrical signals (such as optical signal, medium thickness, displacement, etc.) load in the optical signal, it often uses the way of luminous flux to transmit to optoelectronic devices, optoelectronic device receives optical signals, while enabling photoelectric conversion circuit voltage output analog signals. The output voltage u_O is the function of the measured signals Q, while u_O=f(Q). Obviously u_O is not only related with the Q and the carrier flux, but also related with the electron converter's parameters change. Therefore, design of a high-performance photovoltaic system and use of

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 541–546.
springerlink.com

convert information transformation ASIC module are both particularly important. According to inspection of domestic and foreign literature, design of high-speed, high accuracy, photoelectric conversion of low-power ASIC reports[1~2] has not been proposed. Therefore, it is necessary to use BiCMOS technology to design a differential electron converter and the operational amplifier chip to meet the need for optical engineering.

2 Design of a Differential BiCMOS Electron Converter

2.1 Dual Beam and Dual Optoelectronic Devices' Design

Figure 1 shows the dual beam and dual differential optical detection system device. Light source send the beam, and the beam through the mirror into road reference system and measuring optical system. The up output flux is received by optoelectronic device VD_1 and the down flux is received by optoelectronic device VD_2. The optoelectronic devices VD_1 and VD_2 are covered and packaged together in the tank. The VD_1 and VD_2's characteristic parameters should be consistent as possible[3].

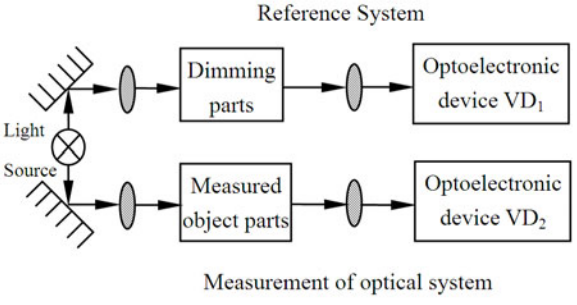


Fig. 1. Dual beam and Dual optoelectronic devices

2.2 Differential Photoelectric Conversion Circuit

Optoelectronic devices followed by a differential conversion circuit can be seen in Figure 2. VD_1 and VD_2 are accessed to the op amp circuit in the form of figure 2. Output voltage of optoelectronic devices VD_1 is u_{VD1}, output voltage of the measurement system VD_2 is u_{VD2}, Op amp A_1 and A_2 are connected to a series of negative voltage feedback circuit, to increase the input impedance differential circuit, and to drift the temperature offset and to improve the circuit's common mode rejection ratio. Amplifier A_3 is connected to form a symmetrical subtraction device parameters, R_6 is in the introduction of a negative voltage feedback, so that the introduction of the three op amp are in the depth of negative feedback, which greatly improves the performance of many electron converter. Three op amp circuit voltage gain is:

$$u_O/(u_{VD2}-u_{VD1}) = (R_6/R_5) \times [1+(2R_4/R_3)] \quad (1)$$

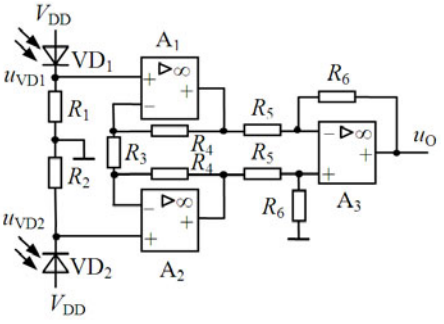
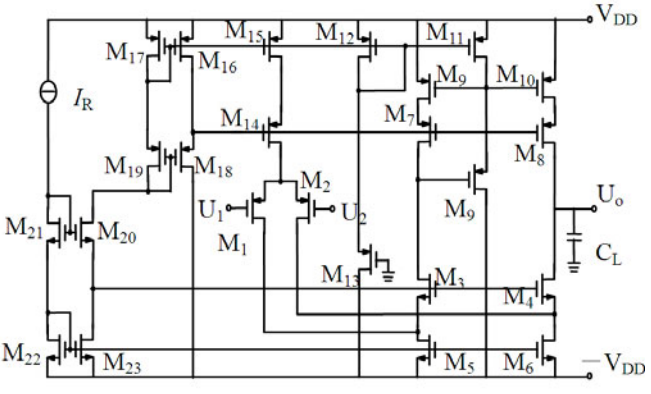
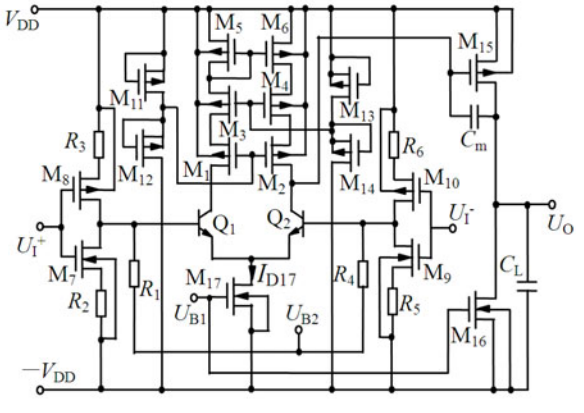


Fig. 2. Differential photoelectric conversion circuit



(a) A_1, A_2: CMOS op amp



(b) A_3: BiCMOS op amp

Fig. 3. Differential conversion circuit op amp BiCMOS F/V converter

when $R_3=R_4=R_5=R_6$, the output voltage can be written in more concise form:

$$u_O=3(u_{VD2}-u_{VD1}) \tag{2}$$

In fig.2, A_1 and A_2 are designed to CMOS operational amplifier, A_3 is designed to BiCMOS operational amplifier, and the entire optical converter is becoming a differential BiCMOS circuit[4~5]. Therefore, all the single and bipolar devices in the circuit can use the preferred device parameters in table 1 of the work[6].

2.3 The Design of Operational Amplifiers A1, A2, A3

The internal circuit of CMOS A_1 and A_2 are showed in Fig. 3 (a). It is a two-stage large circuit: The first level is common-source amplifier, and the second stage is common-gate amplifier[7~9]. The internal circuit of CMOS A_3 is shown in Fig.3 (b).

3 Design of op amp Chip

The photo electric conversion circuits in fig.2 can be implemented by TSMC company's standard using 0.25 μm BiCMOS technology. The specific method is that each of $A_1(A_2)$and A_3 can separately be made into chip, or three chip circuits are made into a single ASIC. Because of the effect of optical device the level chip $A_1(A_2)$is larger than A_3. Therefore, in the Cadence environment firstly complete $A_1(A_2)$map design、and obtain parasitic *RLC* parameter and the later simulation experiment. Simulation results show that because of the influence of circuit parasitic parameter, the circuit-bandwidth is declined, therefore by reelecting compensation capacitance value and optimizing plat fitter can attain the design demands. Fig.4 shows plat area of chip $A_1(A_2)$, and die size is 0.44 mm×0.31 mm. The same can design chip A_3 map, and die chip is the same as $A_1(A_2)$.

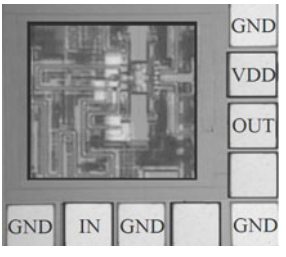
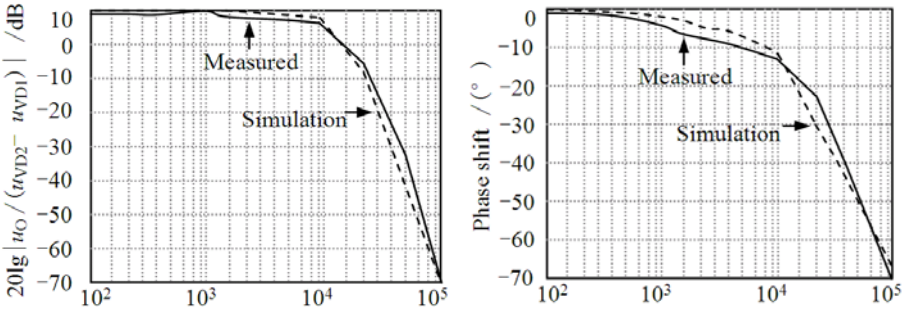


Fig. 4. Layout design of op amp

4 The Experimental Results for Discussion and Analysis

4.1 The Circuit of the Transmission Properties on the Voltage and Frequency of the Curve

Firstly adopting this method to test the voltage of transport properties, and three discharge of the external resistors: $R_3=R_4=R_5=R_6=100$ kΩ, and input resistance $R_1=R_2=10$ kΩ; and amp impedance let parameters: $R_0=R_F=100$ kΩ, $C_1=C_2=10$ pF; When testing input signal frequency sine is 1 kHz. Using MATLAB software to draw and simulate for obtaining voltage-transfer characteristic curve, and together with the hardware circuit test results of voltage transmission characteristics curve drawing in figure 5. From the curve, and the simulation result obviously is consistent with type(2) expressing the operation result in characteristics of drawing, and illiterately the amplifier closed-loop gain of voltage is truly 3; But actual transmission



(a) Amplitude-frequencycharacteristics (b) Phase frequency

Fig. 6. The frequency response of photoelectric conversion circuit

characteristic curve and the ideal simulation curve exists error, and this is mainly induced by external disturbance signal interference and parasitic capacitance charging.

Then the frequency responding characteristic is tested. Circuit and device parameters are the same as above, and 14 frequency points in 50~100 kHz frequency range are measured, while drawing simulation results and the amplitude frequency characteristics curve by software MATLAB, as shown in figure 6. By the graph shows, practical testing and simulation results are consistent. When the voltage circuit amplitude dropped to 0.707 times of low frequency, and phase delay is about 45º, the moment sinusoidal input frequency is more than 60 kHz, and illiterately three simulated transmittal circuits of bandwidth are more than 60 kHz.

4.2 The Main Parameters of Photoelectric Conversion Circuit

The main parameters of the experimental measured BiCMOS differential photoelectric conversion circuit are listed on Table 1, which prove the good performance of the converter design, its speed is faster (t_{PD} is 15.4 ns), -3 DB bandwidth is about 77.3 kHz, gain linearity is 5.7×10^{-5}, and other performance indicators (such as input impedance, common mode rejection ratio, leakage current) was also improved.

Table 1. The main design of parameters differential photoelectric conversion circuit

−3dB /Hz	Amplifier gain linearity NF	Input impedance Z_I /MΩ	CMRR K_{CMRR}	Leakage current I_D /μA	Supply voltage $\pm V_{DD}$ /V	Delay t_{PD} /ns
77.3	5.7×10^{-5}	85.9	0.9×10^{9}	0.23	3.0~5.0	15.4

5 Summary

Differential BiCMOS optical converter is designed in this paper, and op-amp chip size is determined, the advantage of converter gain, bandwidth, power and speed is illustrated by experiments. The converter can be used for optical communications and other optical signal detection system, because it has the following advantages: (1) Using the characteristics of BiCMOS technology "strength of gain - bandwidth product and power - delay product of single and bipolar device" to achieve high-speed low-power electron converter indicators; (2) By selected two photoelectric device in the close of characteristic parameters situations, the dark current influence can be reduced, and conversion errors can be reduced when the temperature changes. (3) Optical path and reference optical path use the same light source, light source drift and volatility will not bring great influence on conversion results; (4) Changes of circuit parameters, especially $\pm V_{DD}$ fluctuation will affect the stability of the measured signal. In addition, Luminous flux is related with light, optical system and the performance of mechanical structures. Therefore, the operational amplifier introduces the deep negative feedback to solve the problem of design simulation of electron converter.

References

1. Emre, A., Kenn, C., Pietro, A.: 45% power saving in 0.25 μm BiCMOS 10 Gb/s 50 Ω-terminated packaged active-load laser driver. IEEE International Solid-State Circuits Conference (ISSCC), 552–553 (2007)
2. John, W.F., Benjamin, S.: A 3V 4.25 Gb/s laser driver with 0.4 V output voltage compliance. IEEE Journal of Solid-State Circuits 8, 1930–1937 (2006)
3. Laopoulos, T.L., Kqbakas, C.A.: High Performance Frequency-to-Voltage Converter. International Journal of Electronics 66, 303–307 (1990)
4. Ferreira, L.H.C., Pimenta, T.C.: Extraction of MOS parameters from BSIM3V3 model using minimum square method for quick manual design. IEEE Circuits, Devices and Systems 48, 153–158 (2006)
5. Cheng, L., Wang, Z.Y., Wang, Y.: Digital Electronic Technology, 2nd edn. China Machine Press (CMP), Beijing (2009)
6. Cheng, L., Zhi, W.J., Fan, H.H.: 0.25 μm BiCMOS Differential Photoelectric Converter. Semiconductor Optoelectronics 152, 16–19 (2008)
7. Cheng, L., Wang, Z.Y., Zhu, J.: Low-voltage, high-speed and large-drive-current BiCMOS analog switch cells. Research & Progress of SSE Solid State, 597–601 (2008)
8. Dejhan, K., et al.: A full-swing BiCMOS schmitt trigger circuit design. In: International Symposium on IC Technology. J. Systems and Applications, vol. 146, pp. 266–269 (1997)
9. Cheng, L., Xu, Z.-C., et al.: Design of a BiCMOS optically coupled isolation amplifier. J. Bandaoti Guangdian/Semiconductor Optoelectronics 28, 613–616 (2007)

Analysis of Transparent Coating Technology on the Surface Texture of Ash Veneer Based on Wavelet Component Parameters

Fenghu Wang[1], Bing Liu[1,*], Anying Jiao[2], Xiaodong Zhu[1], Jianping Sun[1], and Yongfeng Li[2,3]

[1] Material Science and Engineering College, Northeast Forestry University, 150040 Harbin, China
fenghuwa@hotmail.com, {ice8515,jiaoanying008,sjp_jpcn}@163.com, pse4646@126.com
[2] Forestry School, Northeast Forestry University, 150040 Harbin, China
[3] College of Chemistry and Chemical Engineering, Shanghai University of Engineering Science 201620 Shanghai, China
dr_lyf@163.com

Abstract. In order to rationally optimize the clear lacquer practice in modern furniture production application, manchurian ash rotary cut veneers was used as the substrate. The experiment selected the biorthogonal wavelet to decompose wood picture using multi-level technique, and extracted low-frequency subgraph (LL) and high-frequency subgraph (HL, LH) on wood texture after clear lacquer and conduct multi-scale spectral analysis. The quantitative comparison of the texture parameters were studied and acquired transformation rule of the wood surface texture features via after coating process through nitrocellulose varnish (glossy, matt), alkyd varnish (glossy, matt) and polyurethane varnish (glossy, matt). The results showed that : (1) Feature vector obtained by wavelet, examine the value of sub-image and standard deviation of the energy reflected in the texture characteristics, can effectively reflect the wood grain under several clear lacquer variation rule, characteristic and directional; (2) Texture enhancement significantly different, the overall effect performance is alkyd varnish (glossy) > nitrocellulose varnish (glossy) > polyurethane varnish (glossy) > Alkyd varnish (matt) > polyurethane varnish (matt) > nitrocellulose varnish (matt). Studies suggest that, clear lacquer is conducive to enhancing the substrate texture features, the effect is significant and positive for changing the visual effect.

Keywords: Wavelet, Texture analysis, Digital image processing, Quantitation, Multiscale representation, Feature parameters, Clear lacquer.

1 Introduction

Clear lacquer is effective to preserve the natural wood grain and color, is to extend the service life of wood and enhance the natural wood pattern and color, and increases the

* Corresponding author.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 547–554.
springerlink.com

value of timber products effective technical measures, is the senior production of wood products, furniture and commonly used treatment process. Because of its composition ratio and the finishing processes of the change will directly affect the timber, after finishing the visual effects, thus affecting the value of the goods and the construction of the level of the visual effect of environmental studies, so its control technology has been developed for the paint manufacturers and furniture manufacturing industries of the importance [1].

There have been some scholars in China before and after finishing physical changes in wood vision a quantitative measurement (Jian Li, 1998, Yi-xing Liu, 1995, Xin-fang Duan,1998, Hai-peng Yu, 2005, however, after wavelet feature vectors obtained to study the energy value and sub-image as reflected in the standard deviation of the texture features, to comparative analysis of several transparent coating on the surface of the same visual effect veneer, not been reported so far[2-3].

The method of drawing on the basis of previous studies (Hai-peng Yu, 2005), lacquer Nitrocellulose as a common market in general subjects, nitrocellulose varnish (glossy, matt), alkyd varnish (glossy, matt), polyurethane varnish (glossy, matt), which are commonly used for wood products painting in China for the experimental subjects, reference furniture factory standard finish on the wood finishing process, wavelet method (filter length of 8, the decomposition scale is 2) multi-scale texture on the wood spectrum analysis, comparative study and summarizes the impact of their rules, is related to methods used in the furniture industry and decoration industry, improve the scientific nature of the paint color to guide the improvement of finishing processes, better learn to play the quality of wood play a role in the visual environment.

2 Materials and Methods

2.1 Materials Preparation

Substrate: Manchurian Ash rotary cut veneers, size 100mm × 100mm × 10mm, air-dry at room temperature, moisture content 8.26%;

Paint types: nitrocellulose varnish (glossy, matt), alkyd varnish (glossy, matt), polyurethane varnish (glossy, matt), Hua Run Paint Co., Ltd., Shunde, Guangdong;

Process of clear lacquer: primer paint 2 times, coat finish 2 times. Shoot digital image, the image of the sample surface and the parts of the image is consistent with the material, if screenshots are necessary, please make sure that you are happy with the print quality before you send the files.

2.2 Experimental Procedures

Application scanner of EPSON1670, image sampling accuracy is set to 512 × 512 pixels, gray levels of 256, saved as a BMP image format, procedures submitted to the texture. Presented a good image will be saved texture procedures, eigenvectors programming using Matlab wavelet sub-image from the vector concentrated extract texture feature vectors needed for the analysis.

Texture characteristic frequency decomposition: An image (two-dimensional signal) obtained by the decomposition of 4 sub-plans: (1) subgraph (LL); (2) subgraph

(LH); (3) subgraph (HL); (4) subgraph (HH), in Fig. 1. This is the nature of the subgraph: a^{LL} in the horizontal and vertical directions have low-pass characteristics, subgraph (LL) concentrated original image of mainly low frequency components, from the visual point of view, approximate profile of the image information; d^{LH} in the horizontal direction for the low-pass characteristics, high-frequency characteristics in the Vertical direction, from the visual point of view, subgraph (LH) well preserved the original image level of the boundary line of the boundary points; d^{HL} in the vertical direction for the low-pass characteristics, high-frequency characteristics in the horizontal direction, from the visual point of view, subgraph (HL) well preserved the original image of the vertical boundary line boundary points; d^{HH} in the horizontal and vertical directions have high-frequency characteristics, from the visual point of view, subgraph (HH) retain only a few scattered boundary points of the original image[4-5].

These subgraph fully reflect the image at different scales, different frequency texture features in different directions, for the image analysis and classification provide a good foundation.

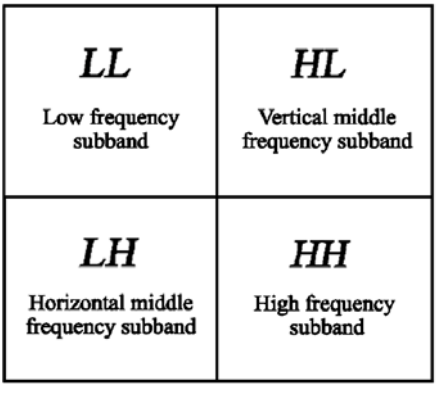


Fig. 1. Two-dimensional duple discrete wavelet decomposition

Texture feature extraction. Matlab programming using wavelet decomposition subimage from the vector extraction parameters required for texture analysis.

Wavelet energy distribution: For an N × N size image, and its energy distribution is defined as:

$$Ef = \sum_{m=1}^{N} \sum_{n=1}^{N} \frac{f^2(m,n)}{N^2} \tag{1}$$

The multi-scale wavelet decomposition, the original image detail sub-images LL, LH and HL k-order wavelet energy distribution is defined as:

$$ELH^{(k)} = \sum_{m=(N/2^k)+1}^{N/2^{k-1}} \sum_{n=1}^{N/2^k} \frac{[LH^{(k)}(m,n)]^2}{(N/2^k)^2} \tag{2}$$

$$EHL^{(k)} = \sum_{n=1}^{N/2^k} \sum_{m=(N/2^k)+1}^{N/2^{k-1}} \frac{[HL^{(k)}(m,n)]^2}{(N/2^k)^2} \tag{3}$$

$$EHH^{(k)} = \sum_{m=(N/2^k)+1}^{N/2^{k-1}} \sum_{n=(N/2^k)+1}^{N/2^{k-1}} \frac{[HH^{(k)}(m,n)]^2}{(N/2^k)^2} \tag{4}$$

The proportion of energy and the directional characteristics of a special detail sub-image to all details sub-image of the same scale, is defined as ratio of wavelet energy distribution.

$$EPLH^{(k)} = \frac{ELH^{(k)}}{ELH^{(k)} + EHL^{(k)} + EHH^{(k)}} \tag{5}$$

$$EPHL^{(k)} = \frac{EHL^{(k)}}{ELH^{(k)} + EHL^{(k)} + EHH^{(k)}} \tag{6}$$

$$EPHH^{(k)} = \frac{EHH^{(k)}}{ELH^{(k)} + EHL^{(k)} + EHH^{(k)}} \tag{7}$$

Thus, $EPLH^{(k)}$ reflected the original image in the horizontal direction energy distribution and specific gravity; $EPHL^{(k)}$ reflected the original image in the vertical direction energy distribution and specific gravity; $EPHH^{(k)}$ reflect the original image in the diagonal direction energy distribution and specific gravity (Hai-peng Yu, 2005).

3 Results and Discussion

3.1 Visual Comparison of Samples Before and After Clear Lacquer

For the clear lacquer of wood samples before and after visual observation of the visual image can be seen: Compared with the substrate, after finishing in the visual sense of color more vivid, light saturation, texture becomes more obvious and clear. Clear lacquer help enhance transparency texture coating, the effect of changing the visual effect is significant and positive [6].

3.2 Texture Characteristic Analysis Before and After Clear Lacquer Finishing Changes

Because the image by the wavelet first and second decomposition, the texture information with the gradual decomposition of the increase in the number of amplification,

decomposition of the second layer, fully reflects the texture information, standard deviation reaches its maximum, the most obvious difference in texture reflect, therefore, choose the decomposition scale of 2.

Sub-images LL analysis of texture parameters. Tab. 1, Fig. 2 were calculated and analyzed by the orthogonal design assistant 2 obtained wavelet decomposition on the sub-images LL visual analysis of the energy value chart table, variance analysis diagram.

Table 1. Subgraph visual analysis of the energy value table before and after clearing lacquer in WaVenLL 2 wavelet decomposition scale.

Decomposition scale at 2				
Sub-image types	Treatment	Before	After	Difference
WavEnLL	Alkyd varnish (glossy)	28604.39	13991.9	-14612.49
	Alkyd varnish (matt)	25885.07	20020.96	-5864.11
	Polyurethane varnish (glossy)	35604.61	27770.91	-7833.7
	Polyurethane varnish (matt)	30857.27	31220.37	363.1
	Nitrocellulose varnish (glossy)	42657.76	30760.81	-11896.95
	Nitrocellulose varnish (matt)	41834.29	43202.73	1368.44

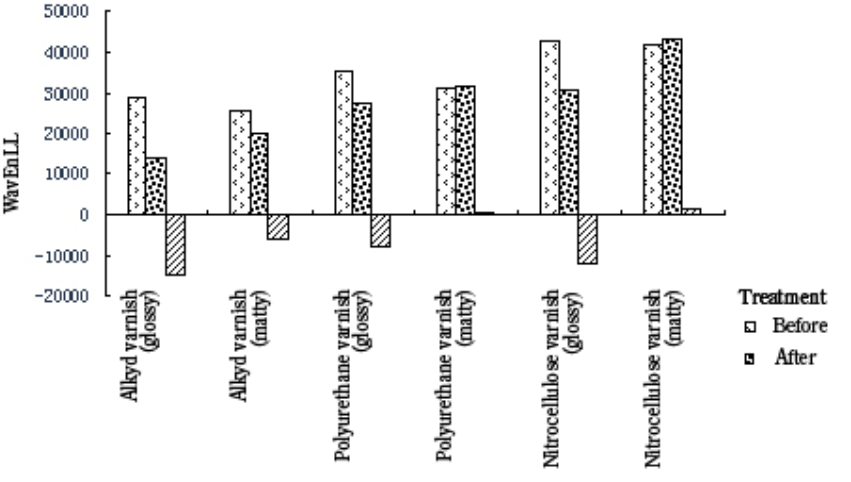


Fig. 2. Subgraph visual analysis of the energy variance chart before and after clearing lacquer in WaVenLL at 2 wavelet decomposition scale

Tab. 1 and Fig. 2 shows, subgraph (LL) in the low frequency components, after polyurethane varnish (matt) and nitrocellulose varnish (matt), energy values were increased, difference was 363.1,1368.44,the others approach is to make the substrate energy value decreased, reduce the rate of alkyd varnish (glossy) > nitrocellulose varnish (glossy) > polyurethane varnish (glossy) > alkyd varnish (matt); Clear lacquer (glossy) and alkyd varnish (matt), texture enhancement, the degree of enhancement is

directly proportional to reduce the magnitude of the energy value; Nitrocellulose varnish (matt) and polyurethane varnish (matt), texture enhancement.

Sub-images LH analysis of texture parameters. Tab. 2, Fig. 3 were calculated and analyzed by the orthogonal design assistant 2 obtained wavelet decomposition on the sub-images LH visual analysis of the energy value chart table, variance analysis diagram.

Table 2. Subgraph visual analysis of the energy value table before and after clearing lacquer in WaVenLH 2 wavelet decomposition scale

Decomposition scale at 2				
Sub-image types	Treatment	Before	After	Difference
WavEnLH	Alkyd varnish (glossy)	5.119006	7.598602	2.479596
	Alkyd varnish (matt)	7.00006	3.978562	-3.021498
	Polyurethane varnish (glossy)	4.008253	6.366469	2.358216
	Polyurethane varnish (matt)	7.83156	2.835516	-4.996044
	Nitrocellulose varnish (glossy)	3.978271	7.387462	3.409191
	Nitrocellulose varnish (matt)	5.73135	3.253840	-2.47751

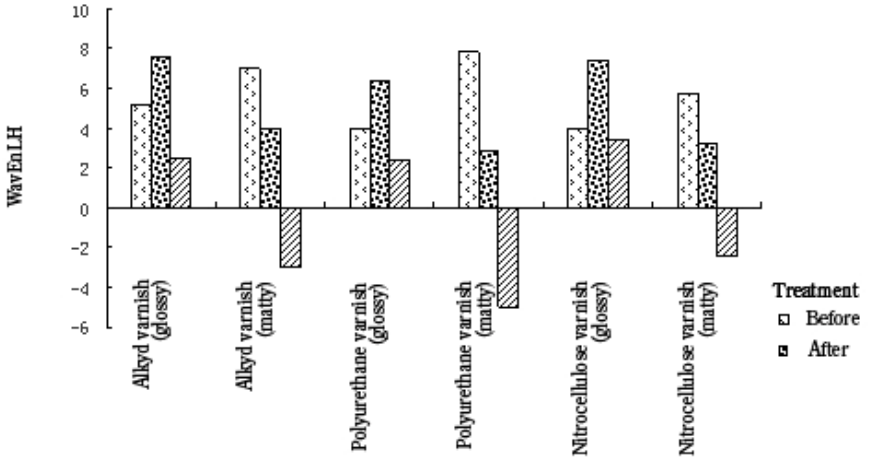


Fig. 3. Subgraph visual analysis of the energy variance chart before and after clearing lacquer in WaVenLH 2 wavelet decomposition scale

Tab. 2 and Fig. 3 shows, subgraph (LH) in the high frequency components, after alkyd varnish (glossy) and polyurethane varnish (glossy), energy values were increased, Reduce the rate of Nitrocellulose varnish (glossy) > alkyd varnish (glossy) > polyurethane varnish (glossy), the others approach is to make the substrate energy value decreased. Because with the texture from coarseness to fineness, from weakness to strength, subgraph (LH) is strengthened gradually, so we can see, clear lacquer (Glossy) treatment, substrate in the horizontal direction (horizontal grain direction) texture enhancement, enhancing the extent and magnitude of increase is proportional

to the energy value; Clear lacquer (matt) treatment, substrate texture weakening in the horizontal direction, because the energy value is reduced by a big margin, so the greater the degree of weakening.

Sub-images HL analysis of texture parameters. Tab. 3, Fig. 4 were calculated and analyzed by the orthogonal design assistant 2 obtained wavelet decomposition on the sub-images HL visual analysis of the energy value chart table, variance analysis diagram.

Table 3. Subgraph visual analysis of the energy value table before and after clearing lacquer in WaVenHL 2 wavelet decomposition scale

Decomposition scale at 2				
Sub-image types	Treatment	Before	After	Difference
WavEnHL	Alkyd varnish (glossy)	56.02797	26.82398	-29.20399
	Alkyd varnish (matt)	46.40508	34.02359	-12.38149
	Polyurethane varnish (glossy)	28.59873	35.98652	7.38779
	Polyurethane varnish (matt)	37.44300	19.2408	-18.2022
	Nitrocellulose varnish (glossy)	25.61983	29.34062	3.72079
	Nitrocellulose varnish (matt)	27.48571	20.31202	-7.17369

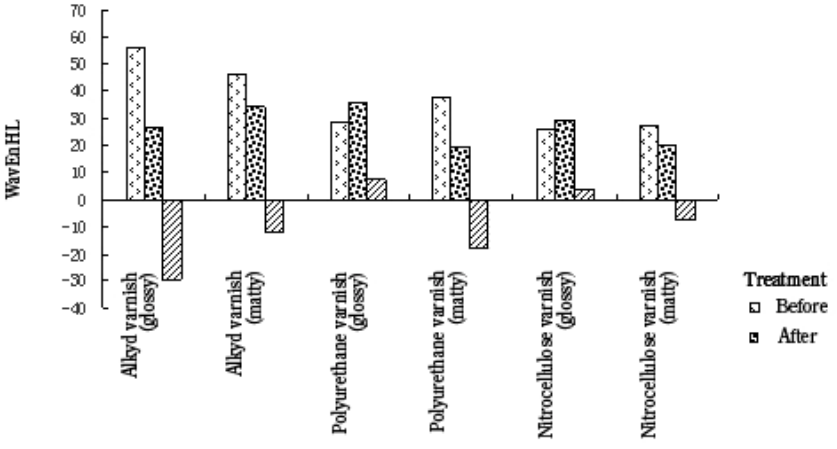


Fig. 4. Subgraph visual analysis of the energy variance chart before and after clearing lacquer in WaVenHL 2 wavelet decomposition scale

Tab. 3 and Fig. 4 shows, subgraph (HL) in the high frequency components, after polyurethane varnish (glossy) and Nitrocellulose varnish (glossy), energy values were increased, difference was 7.38779, 3.72079, the others approach is to make the substrate energy value decreased. Because with the texture from coarseness to fineness, from weakness to strength, subgraph (HL) is strengthened gradually, so we can see, nitrocellulose varnish (glossy), polyurethane varnish (glossy) treatment, substrate in the vertical direction (along the texture direction) texture enhancement, enhancing the extent and magnitude of increase is proportional to the energy value; Alkyd varnish,

polyurethane varnish (matt) and nitrocellulose varnish (matt) treatment, substrate texture weakening in the vertical direction, because the energy value is reduced by a big margin, so the greater the degree of weakening.

4 Conclusions

Wavelet analysis, multi-scale texture of the wood spectrum analysis, transparent coating can effectively direct quantitative visual effects on the wood surface influence.

The results showed: Clear lacquer (glossy) and alkyd varnish (matt), texture enhancement, the degree of enhancement is directly proportional to reduce the magnitude of the energy value; nitrocellulose varnish (matt) and polyurethane varnish (matt), texture enhancement. Clear lacquer (glossy) treatment, substrate in the horizontal direction (horizontal grain direction) texture enhancement, enhancing the extent and magnitude of increase is proportional to the energy value, clear lacquer (matt) treatment, substrate texture weakening in the horizontal direction; Nitrocellulose varnish (glossy), polyurethane varnish (glossy) treatment, substrate in the vertical direction (along the texture direction) texture enhancement, enhancing the extent and magnitude of increase is proportional to the energy value, alkyd varnish, polyurethane varnish (matt) and nitrocellulose varnish (matt) treatment, substrate texture weakening in the vertical direction.

References

1. Van de Wouwer, G., Scheunders, P., Van Dyck, D.: Statistical texture characterization from discrete wavelet representations. IEEE Transactions on Image Processing 8, 592–598 (1999)
2. Yu, H., Liu, Y., Liu, Z.: Auto detection of wood texture orientation by Radon transform. Journal of Forestry Research 16, 1–4 (2005)
3. Yu, H., Cao, J., Luo, W., Liu, Y.: Image retrieval of wood species by color, texture, and spatial information. In: 2009 IEEE International Conference on Information and Automation, ICIA 2009, pp. 1116–1119. IEEE Computer Society, Los Alamitos (2009)
4. Tsai, D.-M., Chiang, C.-H.: Automatic band selection for wavelet reconstruction in the application of defect detection. Image and Vision Computing 21, 413–431 (2003)
5. Yu, G., Kamarthi, S.V.: A cluster-based wavelet feature extraction method and its application. Engineering Applications of Artificial Intelligence 23, 196–202 (2010)
6. Bing, L., Fenghu, W., Da, T., Anying, J., Yongfeng, L.: Comparative Analysis of the Wood Surface Material Chromaticity and Gloss Changes before and after Hot Wax and Transparent Coating Process. In: 2010 Second International Conference on Future Computer and Communicationm, FCC 2010, vol. 4, pp. 638–641. Institute of Electrical and Electronics Engineers, Inc. (2010)

Research on Architectures and Key Techniques of Gansu Science and Technology Documentation Sharing Platform

Quan Wang[1], Jiyu Bao[2], and Weixin Shen[3]

[1] Institute of Science & Technology Information of Gansu, Lanzhou 730000, P.R. China
wang-quan-li@163.com
[2] School of Information and Communication Engineering,Inha University, Incheon 402-751, Republic of Korea
[3] Shenzhen Yako Automation Technology Co., Ltd, Shenzhen 518057, P.R. China

Abstract. Gansu Science & Technology Documentation Sharing Platform consists of five systems: full-text retrieval and web-publishing system, heterogeneous digit resources unitive search system, original text delivery system, user management and accounting system, statistical analysis system. The application and technique architectures of the platform were elaborated in this paper, and the major key technologies on the platform were also expounded, which include unitive search system, web2.0, web services and data security. The platform having been running shows that it integrated 173 resource databases, implemented "one stop" services, improved document resource integration degree., improved service quality, level of management and market competitiveness capacity of documentary information organizations, reduced the repetitive investment of document resource and the development of duplicate of databases resource which have the same content.

Keywords: document sharing, architecture, unitive search, original text delivery, web service.

1 Introduction

July 2004, the general office of state council forwarded "2004-2010 National Science and Technology Infrastructure Construction Program", and proposed the target of building science and technology infrastructure which is resource-rich, layout reasonable, technologically advanced, fully functional, operational efficiency. Currently, from national to provincial and municipal, they builded science & technology documentation sharing platform at all levels so as to provide better service for the scientific innovation. Since 2005, Gansu province relied on Institute of Science & Technology Information of Gansu as the main undertaker for building Gansu Science & Technology Documentation Sharing Platform (http://www.gsstd.cn), to the present, nearly 2,000 individuals and 104 group users registered, in the construction process, building models and key technology of the platform are very worthy of study and reflection, which sum up, can serve the future development of the role of inspiration and reference.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 555–561.
springerlink.com

References [1] proposed four modes of current science & technology documentation sharing platform and analized in detail: resource-oriented mode, integrating service mode, technological application mode and comprehensive mode. Gansu science & technology documentation sharing platform is comprehensive, joint catalog, reference, the original text delivery and other work shows the construction model of integrated service, on the other hand, resource integration reflects technological application. At present, the platform integrated literature resources in seven major collection of literature units, a total of 173 resource databases. Document types include journal papers, standards, patents, dissertations, union catalog, report of the meeting, agency products, local characteristic resources, network of development research center of the state council family library literature, through the unitive portal platform website to offer free secondary document search services for community, according to the user's need to provide the appropriate primary document paid services.

2 Platform's Architecture

Software architecture directly restricts success or failure of software development[2]. The design and implementation of software development of science & technology documentation sharing platform which was acceptable and robust has important significance to its construction. Based on previous researchers' R&D harvest, and according to requirement of construction and maintenance, its software architecture was constructed from different aspects of application and technology.

2.1 Application Architecture

This platform is composed of five systems, as shown in fig. 1.

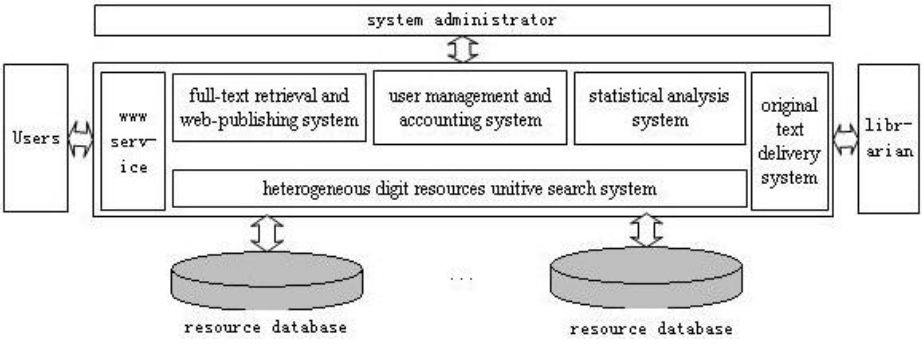


Fig. 1. Application Architecture

1) www service
WWW service including resource directory and notice information. It supplied entry of system such as online registration, personal information query, accounting query and document delivery query. The entry of www servie supplied entry of registered

users and nonregistered users. Nonregistered users cannot download document and, but can search title and abstract.

2) full-text retrieval and web-publishing system
As construction unit of documentation, information which is massive and repeatedly used was collected. However, in these data that cannot be transformed into field such as text, image, audio, video, compound document, it also accounts for certain proportion. They cannot be efficiently handled by traditional RDBMS. And their real value was substantially reduced. Meanwhile, as structured information which could be efficiently handled by traditional RDBMS, It has defects such as processing speed slow, none-uniformity, insufficeient that indexed them. It could not meet information swift growth needs.

The full text retrieval takes the text data as the main processing object, provides the advanced inquiry method according to the material content but not the external characteristic realizes. This platform uses full text retrieval system which was widely applied in the domestic books intelligence system——TRIP(http://trip.istic.ac.cn/html/tripchn/docs/trip.htm).

3) heterogeneous digit resources unitive search system
The unification retrieval system refers to the user submit retrieval request through sole and user-friendly interface which could access many web databases and search engine at a time, gain more accurate and orderly retrieval result. It shows high precision and retrieval efficiency in higher recall.

User may retrieve various resources database on one retrieval and two retrievals. The field include title, author, full text, keyword, category number and so on. The inquiry condition was saved which could directly use in the later inquiry. The retrieval result display by paging. User can collect the retrieval result and browse digest information. If it has online full text in retrieval result, registration user may directly download it which could be automatically accounted by system. If it has offline full text, user could gain full text by original text delivery.

4) original text delivery system
It process request of original text delivery which user submit from search result. Before processing request, it authenticate users' identity and the balance of account. If passed, request of original text delivery would be transmited to corresponding collection unit. It adopts advanced service pattern which is end user-oriented. Readers submit request of original text delivery on line by myself and gain full text in email when register in this system. The entire process does not need any third party involvement. The way of gaining literature is convenient, quick, conforms to reader's information acquisition custom under especially informationization swift development environment.

5) user management and accounting system
User management include several service such as user registeration, user management, send messages, send email, batch processing on user period of validity, batch processing on user prestore fee and user imformation online query. Among them, sends messages or email to the registered user is advantageous to publicize, then two batch processing functions greatly raised the literature service efficiency. The user management also provides the user authenticate [3] and the jurisdiction inspection, it

assists to complete the user authentication of the literature WWW service, the unification retrieval and the original text delivery system.

Accounting system completes account of full text downloading and the original text request, it may recharge and refund money for the user.

6) statistical analysis system

According to the received document retrieval services, full download and the original ordering information, we count the use of various types of documents resources, users geographical and age distribution to provide reliable data for the decision analysis.

Statistical analysis of original text delivery system provides "Costs Statistics ", "User Statistics", "Sended Request Statistics", "received request statistics" and other modules, notably the three statistical feature of the design can be described as originality. "User Statistics" provides the results from the user education, job title, type, total number of accounts, using or not using the number of multiple dimensions of analysis, screening: the "send or receive requests statistics" module, not only count each document collection unit of sended or received requests number, but also provide details of the treatment results——satisfied, not satisfied and the specific reasons of not satisfied, the service hours, tolls. This is easy to understand document service units of the various documents for the processing of collections, the museum services of external circumstances, from mining problem, then the document collection for each unit of the library resources development, training and readers to communicate with the outside hall and other work provide the basis for the launching.

2.2 Technical Architecture

This platform based on php and Web Service to view, business logic and data layers of a reasonable division of the overall framework of the system more optimized. Technical architecture was divided into the portal layer (service channel level), application support layer, application layer, information resources layer, as shown in Fig. 2.

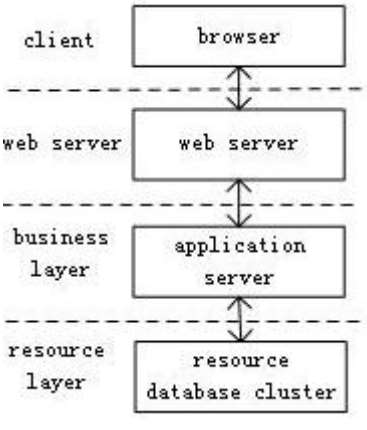


Fig. 2. Technical Architecture

System architecture with four layers: client (request information), business (process the request) and data (to operate) which are physically isolated. Client layer consists of the browser and rich client, the browser interface for the display system, the rich client to handle ajax requests and XML data. Service layer consists of Web servers and application servers of literature gateway, Web server handle display logic, application servers of literature gateway deal with document retrieval business logic, business logic in the middle layer, do not care about what type of client request data, but also with the back-end system to maintain relative independence, is conducive to system expansion. Four- layer structure has better portability, you can work across different platforms, allowing users to request load balancing between multiple servers. Security is more easy to implement, because the business logic has been isolated with the customer.

Four- layer communication between the following: The browser through Ajax asynchronous call to send user requests to the Web server, Web server and the literature gateway were used between the SOAP protocol, the gateway application server through HTTP, Z39.50, ODBC or JDBC with the literature resources such as databases.

3 Key Techniques

3.1 Unitive Search

Unitive search [4] also known as cross-database searching, unitive search system must work around two of the most significant features such as the heterogeneous and distributed computing, as long as the shield of heterogeneous database resources, realistic solutions will be able to propose with rational use of distributed processing technology to achieve a unitive document resources inquiries.

From the technical point of view, there are two kinds of unitive search model: joint search and integration search (http://www.chnlib.com/zylwj/shuzitsg/200605/221.html).Joint search process commonly used simulation Web access [5], the search criteria which was entered in unified search interface will be automatically saved and sended to multiple digital resources system, the digital resources system start their search system to search and show search results in the same interface.

In method of implementation of web-based simulation, the core technology is web information extraction. This paper proposes a new method which can extract the useful information from the different document sites automatically——web information extraction based on sub-tree breadth [6].The method is that view title number of per page of scientific and technical literature web site and store the number in the database, and then use the HTML Tidy(http://tidy.sourceforge.net) to clean up these pages into XML documents, and then generate DOM tree, computing breadth of someone sub-tree in DOM tree, by judging the breadth of a sub-tree is equal title number of per page of pre-stored in the database to establish the key information block[7-8], and then extract information from key information block. Experiments show that the method can guarantee a high accuracy in terms of recall and precision [9].

3.2 Web2.0

To improve user experience, the system uses Ajax, Tag and other Web2.0 technology to implement personalized service. Personalized service refers to users can obtain information or services by their own purposes and needs. For example: Users can collect their own commonly used search document type and commonly used literature database by tag for later use. In user's document classification, a free label was used, allowing users to freely define. In resource database display, ajax was used, the user clicks the label to select resources database list of data output, in the whole operation, user does not need to refresh the entire page. In addition, my database, my favorites, my search history is the performance of personalized service.

Although the main advantage of PHP is superior processing character speed and reliability, through combination of Apache 2.0 makes the unitive search system has good stability and performance [10], but it does not support multi-threading, and unitive search system needs to search multiple databases at the same time, if an ordinary single-threaded program, processing speed will slow that people can not tolerate. Multi-task programming techniques was improved by ajax to enhance the program efficiency and avoid a "suspended animation"state of program interface.

3.3 Web Service

The distributed store of digital resources is one of its main features, the traditional unitive search system was designed for a specific portal, it was lack of sharing resources and accounting between sharing units. System needs to establish Web service program for title search, abstracts obtain and full-text download by asynchronous call procedures, results transfer by XML file between users and documents node group, it achieved grid resource sharing, resource exchange visits and billing. For users, document search services mainly refers to it receives user requests and then starts the searching machine, and a XML file of search results is generated by services. The main services include resource list service, browsing titles service, browsing abstracts service and full-text download service.

3.4 Data Security

This platform provide services by network, in this process, how to protect data security, mainly depend on following several technology: (1)strengthen right setting, legitimate users can be accessed by password; or through the IP address settings, users in IP segment can be accessed. (2)restrict data traffic of accessing databases to prevent malicious downloads and databases collapse. (3)using encryption and digital signature technology during network transmission, to prevent theft and destruction.

4 Conclusion

Key technology and the model of this platform has been successfully applied in addition in Gansu province, but also promote to China petroleum, Qinghai province(http://www.textqh.com), Ningxia province(http://www.nxkjwx.com.cn) and other provincial-level scientific and technical document sharing platform. The

practice shows that the platform improve their integration degree. This research offers a feasible approach for the realization of "one stop" services. The application of this platform greatly promoted the process of information sharing of documentary information organizations, improved service quality, level of management and market competitiveness capacity of documentary information organizations, reduced the repetitive investment of document resource and the development of duplicate of databases resource which have the same content.

Along with the development of science & technology documentation sharing platform, research and development of heterogeneous digital resources unitive search will enter a deeper level, function will get further rich, but unitive search based on web service was created to solve non-standard data interface of vendors, and access and retrieval interface which was based on standard and norms of resource database is the solution to the current "information island" phenomenon of the effective ways and means.Our country should early strengthen standards and norms and compulsory promotion to complete our country digital libraries as soon as possible.

References

1. Hu, M.: An Inquiry into Mode of Constructing Sharing Platforms of Scientific and Technical Documents in China and Thinking of Sustainable Development. Digital Library Forum (7), 67–70 (2008) (in Chinese)
2. Bass, L., Clements, P., Kazman, R.: Software Architecture in Practice Second Edition. Addison-Wesley, Reading (2003)
3. Xiao, W., Yang, S.: Design and Implementation of Unified Identity Authentication System Based on LDAP. Computer Science 35(5), 298–301 (2008) (in Chinese)
4. Huang, D.: A Review of Cross Searching Technique for Heterogeneous Databases. Library and Information Service (6), 94–97 (2003) (in Chinese)
5. Cao, F., Shi, S.: Design and Implementation of Heterogeneous Digital Documentation Unification Retrieval System Based on Web Simulation Process. Journal of the China Society for Scientific and Technical Information 25(5), 575–579 (2006) (in Chinese)
6. Wang, Q., Shi, S.: Web Information Extraction Based on Sub-tree Breadth. Computer Engineering 35(3), 89–90, 93 (2009) (in Chinese)
7. Deng, C., Yu, S., Wen, J., et al.: Block-based Web Search. In: Proc.of the 27th Annual lnternational ACM SIGIR Conference, pp. 456–463. ACM Press, Sheffield (2004)
8. Deng, C., He, X., Wen, J., et al.: Block-level Link Analysis. In: Proc.of the 27th Annual International ACM SIGIR Conference, pp. 440–447. ACM Press, Sheffield (2004)
9. Gaizauskas, R., Wilks, Y.: Information Extraction:Beyond Document Retrieval. Computational Linguistics and Chinese Language Processing 3(2), 17–60 (1998)
10. Wang, Q., Shi, S.: Design and implementation of unitive search system based on PHP. Journal of Lanzhou University of Technology 34(1), 91–94 (2008) (in Chinese)

Effective Design for CRFF Sigma-Delta Modulators Using Inverters Based on 0.13μm CMOS

Hongyi Li, Yuan Wang, Song Jia, and Xing Zhang

Key Laboratory of Microelectronic Devices and Circuits,
Institute of Microelectronics, Peking University, Beijing, China
{lhyyym,wangyuan}@pku.edu.cn

Abstract. In this paper, an efficient method to relax timing requirements of CRFF sigma-delta modulators has been proposed. A system optimization to circuit level design was finished. Class-C inverter was used to realize half delay integrators of the proposed structure. A 4th-order 1-bit CRFF topology was implemented in smic 0.13μm CMOS technology. With 31.25MHz sampling frequency and 64x oversampling ratio, 13.2-bit resolution has been reached. The whole circuit consumes 472.63-μW power from a single 0.6V supply voltage. Thus, a low-voltage low-power medium-bandwidth high-accuracy sigma-delta modulator has been obtained.

Keywords: sigma-delta modulator, switch-capacitor (SC) integrator, class-C inverter, low-distortion topology, low-power.

1 Introduction

Sigma-delta data conversion technique is commonly used in many fields such as wired and wireless communication systems, consumer electronics, radar systems and medical applications because of its high-accuracy and wide-band. In order to satisfy the ever-increasing demands for portable devices, the development of sigma-delta modulators (SDM) whose power in macro watt rang has become a hot research topic recently [1][2][3].

Although complex structures, such as cascade and multi-bit modulators, could be used to realize extra high resolution and speed, high power consumption is inevasible. Of course, continuous-time (CT) implements seem to show lower power consumption than switch-capacitor (SC) ones, however, high sensitivity to clock jitter and large coefficient warp restrict their design. Thus, simple SC architecture, i.e. single-loop single-bit SC modulator, is still the main method to realize low-power high-performance sigma-delta modulators. In addition, the low-voltage low-power performance of feed-forward topologies could be superior to the feedback ones, since they have reduced internal signal swings by cancelling the input signal at the input of loop filters.

Therefore, single-loop single-bit low-distortion cascade of integrators with distributed feed-forward form (CIFF) and cascade of resonators with distributed feed-forward form (CRFF) [4] could be effectively used for low-power applications. For high-speed modulators realized by SC integrators, there are advantages to have a

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 563–570.
springerlink.com

delay associated with each integrator, since it reduces the speed requirements of the operational amplifiers (op-amps) used. As a result, CIFF such as [5] is used more than CRFF which needs a delay-free integrator in each resonator to insure the poles stay on the unit circuit [4]. However, as a kind of effective topology, CRFF also including other sigma-delta topologies which contain delay-free integrators ought not to be limited to theory. This paper suggests a feasible method to realize the CRFF structures.

Our design goal is to finish high-resolution (Effective-Number-of-Bits ENOB>13-bit), medium-bandwidth (input signal bandwidth fb>200 kHz) and low power (Power<1mW) sigma-delta modulators, so that high figure-of-merit (FOM) performance can be reached.

This paper is organized in the following manner. Section 2 provides the system level considerations of the modulators. Section 3 describes circuit implement of the SC circuits. Section 4 shows the experiment results of the designed sigma-delta modulators.

2 System Considerations

2.1 Proposed Structures

Take a 4th-order 1-bit CRFF SDM as shown in Fig.1 for example, and Fig.2 is the proposed structure. V and U denote the output and input of overall modulator, respectively. The key is to use half delay integrators instead of the delay-free ones so that

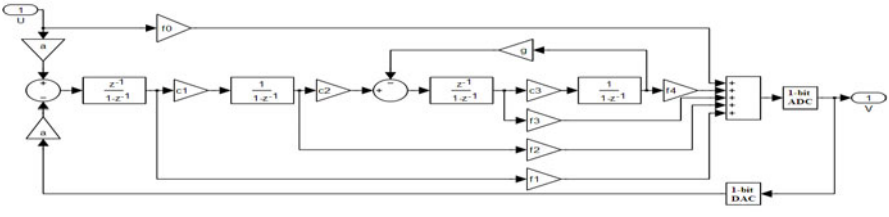


Fig. 1. Block diagram of the traditional 4th-order 1-bit CRFF SDM.

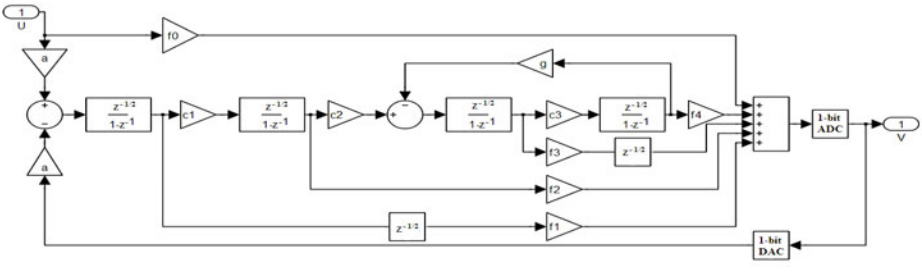


Fig. 2. Block diagram of the proposed 4th-order 1-bit CRFF SDM.

rigorous timing requirements could be relaxed. For this purpose, the method of equivalence transformations has been used to change all the integrators to half delay ones, and meanwhile, half delay elements have been introduced in the first and third integrator feed-forward paths. As a result, the signal (STF) and noise transfer function (NTF) of the CRFF would not be changed, and the timing could be relaxed as well.

2.2 Coefficient Selection

In order to satisfy our design goal, through behavior simulation and design margin consideration, a 4th-order single-loop 1-bit CRFF form was chosen as the proposed topology, since they balance high resolution, high speed and low circuit complexity effectively.

Using the 'delsig' toolbox available on [6], an inverse Chebyshev NTF which has local feedback path has been derived in the discrete-time domain and the coefficients of the sigma-delta topology could also be produced. The order of the NTF has been chosen to 4, oversampling ratio (OSR) is 64 and the sampling frequency (fs) is 31.25MHz. The maximum out-of-band gain of NTF Hinf = 1.5 has been chosen to balance resolution and loop stability. Dynamic-range scaling has also been performed to restrict integrators' output swings to known and practical values.

After simulation and scaling, the final scaled values used are shown in Table 1. Here, for the sake of using unit capacitors to ensure the capacitor ratios to be insensitive to fringing effects and processing-induced length and width variations in real circuit implement, rational approximations have been introduced as well [4]. With the coefficients, ideal CRFF could exhibit 16-bit resolution in system simulation using Matlab & Simulink.

Table 1. Modulator's Coefficients

Integrator coefficients	Resonator coefficients	Feedback coefficients	Feed-forward coefficients
a=0.4005 → 2/5	g=0.0083 → 1/120	a=0.4005 → 2/5	f0=1
c1=0.3929 → 2/5			f1=1.388 → 25/18
c2=0.2799 → 7/25			f2=1.5804 → 79/50
c3=0.2076 → 1/5			f3=1.1718 → 34/29
			f4=0.6274 → 19/30

2.3 Non-idealities Simulation

For the sake of finding out the requisite design specifications of building blocks, the proposed 4th-order 1-bit CRFF sigma-delta modulator first has been simulated using MATLAB environment with the method of [7]. Here, a -4.1245dB full-scale 64.85 kHz sine wave has been used as input signal. After the behavior level simulation, the required amplifier DC-gain which determines the accuracy of charge transfer and

GBW that determines its operation speed for the first integrator is 35dB and 40MHz respectively, and its sampling capacitor is 4pF for 86.9dB SNDR, as seen in Fig.3.

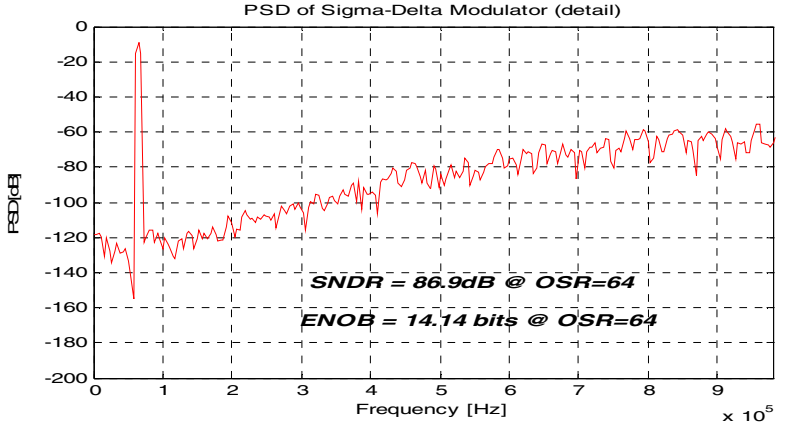


Fig. 3. Output power spectrum with 8192 samples of real CRFF modulator modeled in Matlab

3 Circuit Implementation

The major building blocks of the modulator include SC integrators, 1-bit quantizer, 1-bit feedback DAC, switches, clock bootstrap and clock generator.

3.1 Integrator Design

In the proposed modulator structure, all the integrators are implemented with half delay forms. This could be done by switched-RC integrators proposed in [8]. However, [3] shows us a more power-efficient method to realize half delay integrators. It suggests us to use inverters to substitute general op-amps. In fact, the op-amp is the most vital component for low-power sigma-delta modulator, since it dominates the performance and consumes most of power of the overall system. Thus, to find out a kind of high performance and low cost op-amp satisfying the requirement of sigma-delta modulator has become an important work for low-voltage low-power design, and using inverter as op-amp gives us a new way to this job.

As shown in Fig. 4, the inverter-based half delay SC integrator using an auto-zeroing technique to cancel the offset and form a virtual ground for the integrator. The node connecting the bottom plate of sampling capacitor Cs could be considered as a virtual ground, and the charge should be transfer from Cs to integrator capacitor Ci as the conventional SC circuits. The capacitor Cc has be used to sample and hold the offset voltage causing by the one input terminal inverter. A common-mode feedback (CMFB) circuit has been used in this kind of pseudo-differential SC integrator, and gain of the CMFB is defined as the ratio of C_M/Ci.

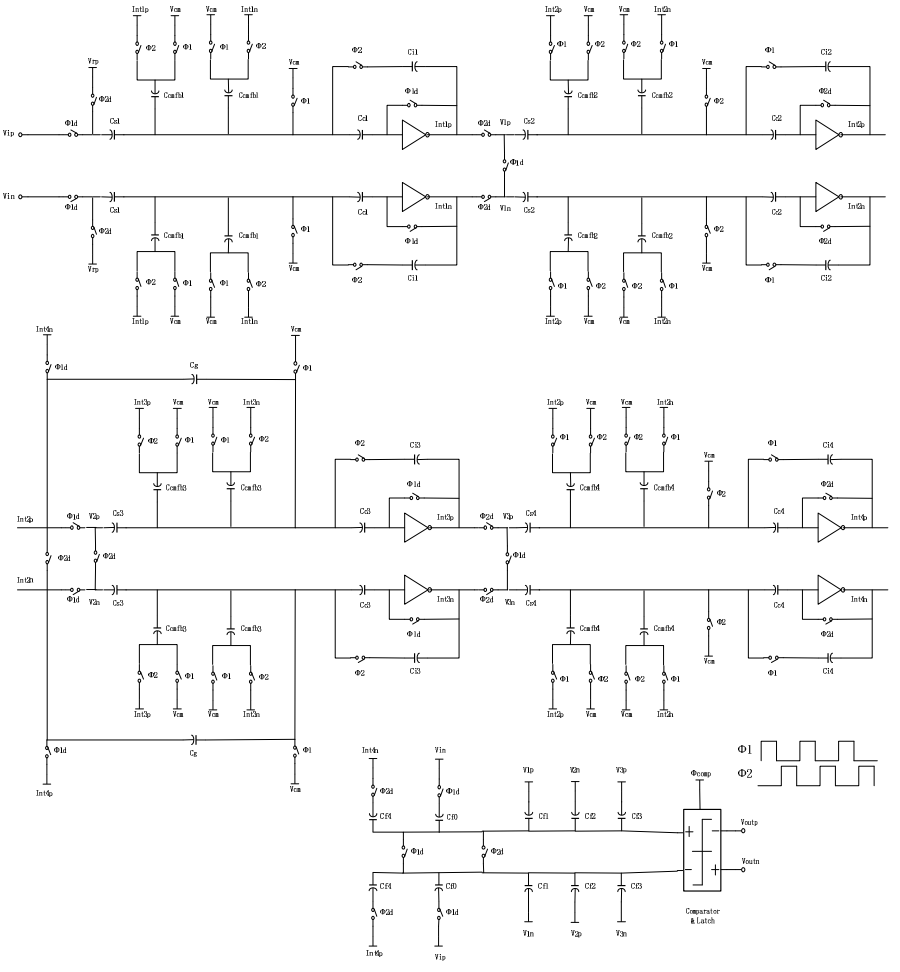


Fig. 4. Schematic of the proposed complete CRFF modulator.

Table 2. Inverters' Performance

DC gain	GBW	Slew rate (SR)	Power
41.13dB	63.543MHz	51.35V/μs	8.425μW

As for the inverter, a simple push-pull inverter is more suitable than the cascode one in a low power supply voltage Vdd, since a simple inverter could still obtain both high DC-gain and wide GBW with very low Vdd. Based on the PMOS and NMOS transistors' threshold voltage of our technology, 0.6V power supply voltage should be chosen to insure class-C operation which is the most power efficiency and high performance for an inverter [3]. Table 2 shows the basic performance parameters with 10pF load capacitor of the inverter used in the first integrator. Considering the non-idealities of other building blocks in the modulator that have not been modeled in the

above Matlab simulation, higher performance inverter has been used in the circuit level design.

3.2 Other Building Blocks

The modulators are controlled by two phase, non-overlapping clocks: one for the sampling phase and the other for the integrating phase. In order to reduce the effects of charge injection, delayed clocks (Φ1d and Φ2d) are needed. Thus, an on-chip clock generator was designed to generate the above clocks.

Since the Vdd is 0.6V, the input and output common-mode voltage is set to 0.3V. And then, the feedback reference voltages Vref+ and Vref- are chosen as the power supply rails Vdd and ground.

A traditional clocked comparator which consumes dynamic power only, followed by latch has been used as the 1-bit quantizer. The 1-bit feedback DAC is implemented by four transmission gate switches which are controlled by the output voltage of overall modulator.

All switches are implemented by transmission gates and driven by a pair of invert clock signals. A simple bootstrap circuit [9] is used to raise the voltage level of the clock in order for turning on the switches completely. In addition, sharing sampling capacitors and redundant switches techniques are introduced in the circuits to save power and chip area.

The complete schematic of the overall CRFF modulator is displayed in Fig. 4.

4 Simulation Results

This circuit has been realized using Cadence environment under smic 0.13μm CMOS technology. Using MATLAB, FFT with hanning window has been used to calculate signal to noise and distortion ratio (SNDR) under 8192 samples.

The output swings of each integrator in the CRFF sigma-delta modulator are shown in Fig. 5 (left). Here, a 64.85 kHz, -2dB to full scale (dBFS) sinusoidal input signal has been used. As shown, besides the output swing of the first integrator which is a ± 0.5V waveform, the rest ones present relative small output levels ranging from ± 0.2V to ± 0.3V voltage, all of that exhibit the low-distortion property.

Fig. 5 (right) plots the SNDR versus input amplitude. As seen from that, with 64x OSR, the proposed modulator could finish a large dynamic range (DR). The characteristics of the overall 4th-order 1-bit CRFF SC sigma-delta modulator are summarized in Table 3. The FOM for measuring performance is defined as [3]:

$$\mathrm{FOM} = \frac{\mathrm{Power}}{2 \cdot 2^{\mathrm{ENOB}} \cdot \mathrm{Bandwidth}} \tag{1}$$

In Table 4, it is a comparison of the performance for several recent low-voltage low-power sigma-delta modulators. It is obvious that the proposed structure has high performance among recent designs. In addition, as shown, in general, low-power designs are used as ten kHz bandwidth range, while our design could be used as hundred kHz applications.

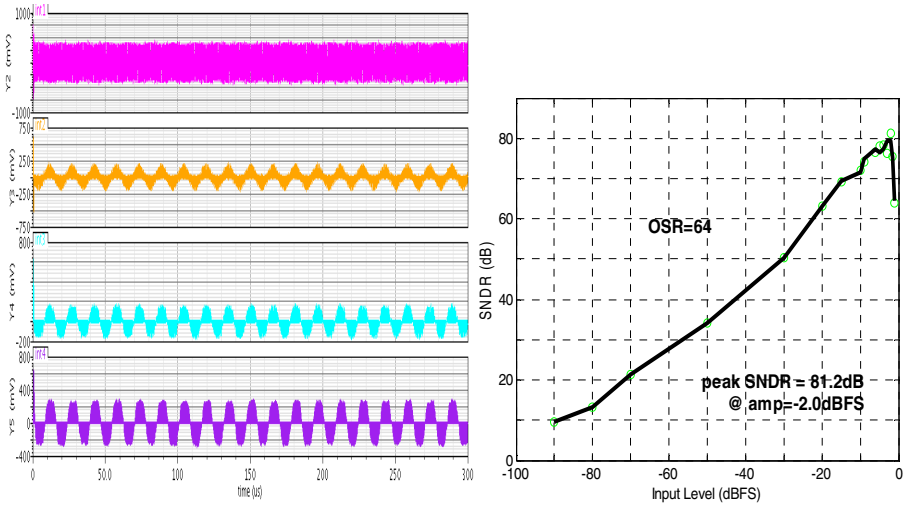


Fig. 5. Integrator output swings (left) and SNDR versus input amplitude of the CRFF (right) with 64.85 kHz, -2dBFS input signal.

Table 3. Modulator's Performance Summary

Supply Voltage	Technology	fs	OSR	fb	Power	Peak SNDR	ENOB	DR	FOM
0.6V	smic 0.13μm CMOS	31.25MHz	64	245kHz	427.63μW	81.2dB	13.2bits	95dB	0.103 pJ/conversion-step

Table 4. Performance Comparison

Paper	Vdd [V]	CMOS [μm]	BW [kHz]	SNDR [dB]	DR [dB]	Power [μW]	FOM [pJ/conv]
Roh [1]	0.9	0.13	20	73	83	60	0.412
Par [2]	0.7	0.18	25	95	100	870	0.378
Cha [3]	0.7	0.18	20	81	85	36	0.098
Su [5]	1	0.18	16	63.4	N/A	18.1	0.468
Ahn [8]	0.6	0.35	20	81	82	1000	2.731
Su [9]	1	0.18	20	61.96	66	42	1.025
This work	0.6	0.13	245	81.2	95	472.63	0.103

5 Conclusion

A method to overcome tight timing requirements of CRFF sigma-delta modulators have been proposed in this paper. Class-C inverters have been used as the amplifiers in half delay integrators. Then, a 4th-order 1-bit CRFF modulator has been implemented under 0.13μm CMOS technology with only 0.6V supply voltage. Compared

to the other low-power structures, this design achieves high FOM performance. It indicates that both power and performance are well optimized.

Acknowledgments. Project supported by the National Science Fund for Distinguished Young Scholars of China (Grant No. 60925015).

References

1. Roh, J., Byun, S., Choi, Y., Roh, H., Kim, Y.G., Kwon, J.K.: A 0.9-V 60-μW 1-Bit Fourth-Order Delta-Sigma Modulator With 83-dB Dynamic Range. J. Solid-State Circuits 43(2), 361–370 (2008)
2. Park, H., Nam, K., Su, D.K., Vleugels, K., Wooley, B.A.: A 0.7-V 870uW Digital-Audio CMOS Sigma-Delta Modulator. J. Solid State Circuits 44(4), 1078–1088 (2009)
3. Chae, Y., Han, G.: Low voltage, low power, inverter-based switched-capacitor delta-sigma modulator. J. Solid-State Circuits 44(2), 458–472 (2009)
4. Schreier, R., Temes, G.C.: Understanding Delta-Sigma Data Converters. Wiley/IEEE Press, New York (2004)
5. Su, P.H., Chiueh, H.: The design of low-power CIFF structure second-order sigma-delta modulator. In: 52nd IEEE International Midwest Symposium on Circuits and Systems, MWSCAS 2009 (2009)
6. Schreier, R.: Delta Sigma Toolbox 2009 (2009), `http://www.mathworks.com/matlabcentral`
7. Malcovati, P., Brigati, S., Francesconi, F., Maloberti, F., Cusinato, P., Baschirotto, A.: Behavioral Modeling of Switched-Capacitor Sigma-Delta Modulators. IEEE Transactions on Circuits and Systems I - Fundamental Theory and Applications 50(3), 352–364 (2003)
8. Ahn, G.H., Chang, D.Y., Brown, M.E., Ozaki, N., Youra, H., Yamamura, K., Hamashita, K., Takasuka, K., Temes, G.C., Moon, U.K.: A 0.6-V 82-dB Delta-Sigma Audio ADC Using Switched-RC Integrators. J. Solid State Circuits 40(12) (2005)
9. Su, C., Lin, P.C., Lu, H.W.: An Inverter Based 2-MHz 42-μW $\Delta\Sigma$ ADC with 20-KHz Bandwidth and 66dB Dynamic Range. In: IEEE Asian Solid-State Circuits Conference, ASSCC 2006 (2006)

Hardware Design of Sensor Nodes in the Nilaparvata Lugens Monitoring System Based on the Internet of Things*

Ken Cai[1,2], Xiaoming Wu[1], Xiaoying Liang[3], and Keqiang Wang[2]

[1] School of Bioscience and Bioengineering,
South China University of Technology, Guangzhou 510641, China
[2] Information College,
Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China
[3] Guangdong Women's Polytechnic College, Guangzhou 511450, China
caiken218@gmail.com

Abstract. With the rapid developing of electronics, communication and embedded system technologies, some specific applications of the Internet of Things have made a success in the military and civilian aspects. The application of the Internet of Things in Nilaparvata lugens and environmental online monitoring is an effective solution for present wired sensor monitoring system, which has much more disadvantages, such as complex wiring, susceptible to interferences, poor reliability, poor real-time capability, difficult maintenance, heavy repair work load and so on. In this paper, a novel wireless sensor node in the Internet of Things is proposed to achieve the functions of automatic data collection, transmission and the real-time promulgation. The wireless sensor nodes include a temperature sensors, a Humidity Sensors, a module of data processing based on ARM, a wireless communication module, etc. The whole structure of system is given and the key technologies of system design are introduced. In additional, the design and its realizing method are described in the hardware aspect. The system has the advantage of simple structure, low-budget, low power consumption, etc. This makes the system applicable to Nilaparvata lugens and environmental online monitoring and has a strong practical value.

Keywords: Nilaparvata lugens, the Internet of Things, farmland information, sensor nodes, hardware.

1 Introduction

The Internet of Things is an emerging information network and is described as a self-configuring wireless network of sensors whose purpose would be to interconnect all things [1]. The concept is attributed to the former Auto-ID Center, founded in 1999, based at the time at the Massachusetts Institute of Technology (MIT) [2]. The Internet

* This research was supported by GuangDong Provincial Science and Technology Planning Project of China under grant 2010B020315028.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 571–578.
springerlink.com

of Things is neither science fiction nor industry hype, but is based on solid technological advances and visions of network ubiquity. Its appearance enables the whole world to be intelligently perceived and more mutually connected. Any objects will be able to exchange information and actively process information according to predefined schemes. In this vision of the future, is it easy to imagine things that are able to transport themselves: e.g. sensors in an electronic jacket can collect information about changes in external temperature and the parameters of the jacket can be adjusted accordingly. There will be fully automated supply networks, autonomous warehouses.

Nilaparvata lugens is one of the major pests of rice; not only in China but in most rice growing countries in South and Southeast Asia. Since rice crops are continuously cultivated in tropical Asia, both nymphs and adults of Nilaparvata lugens damage rice plants through extensive feeding on them. Nilaparvata lugens also transmits viruses such as rice ragged stunt (RRSV) and rice grassy stunt (RGSV) [3], [4]. From 2005 to 2006, more than 485,000 ha of rice production area in southern Vietnam were severely affected by Nilaparvata lugens. It resulted in the loss of 828,000 tons of rice valued at US$120 million [5]. Thus, once the monitoring and prevention work is relaxed, it is very likely to cause serious damages in a large scale. Given the above cases, the technology of the Internet of Things is used to monitor Nilaparvata lugens and environmental information. It provides an important basis to study the rules of the Nilaparvata lugens disaster and the key factors leading to the disaster.

2 Wireless Sensor Network Architecture

2.1 Wireless Sensor Network Structure

The wireless farmland information acquisition system took distributed structure which is showed in Fig.1.

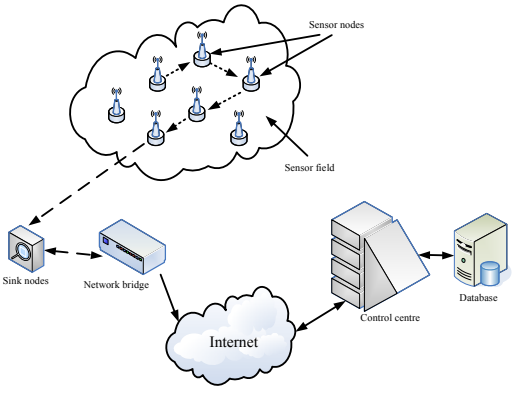


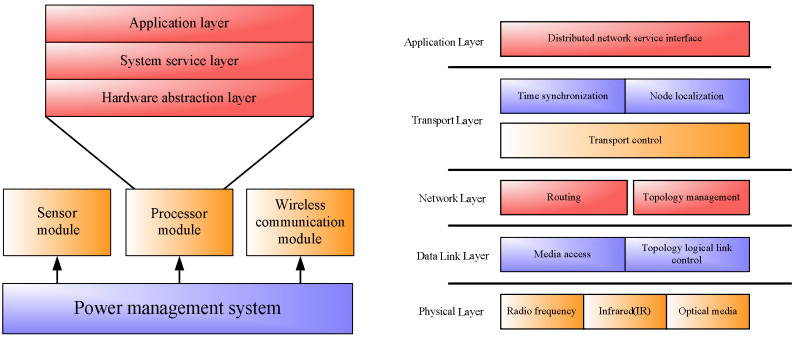
Fig. 1. Sensor network architecture.

The system consists of wireless sensor nodes, sink nodes, wireless bridge, and control centre. The measure data transmit from the sensor nodes to the sink nodes using the ZigBee communication network. The data transmit from the sink nodes to the

server by network bridge. The sink nodes in this system act as gateway, it is the protocol conversion used to transform a data package in ZigBee protocol to TCP/IP protocol before transmitting and a data package in TCP/IP protocol to ZigBee protocol.

2.2 Structure of Wireless Sensor Nodes

In wireless sensor network, there are a lot of sensor nodes, which not only have hardware part, but also have software part. The hardware part consists of four hardware components of sensor modules, processor modules, wireless communication modules and power supply modules, while the software part is made up of the hardware abstraction layer, the service layer and the application layer. The entire structure is shown in Figure 2(a).



(a) (b)

Fig. 2. (a) Composition of wireless sensor nodes. (b) Protocol architecture of wireless sensor network.

At the hardware layer, the sensor module is responsible for collecting information in sensor field. The processor module is responsible for coordinating the work of various parts of nodes. The wireless communication module is responsible for wireless communication with other sensor nodes. The power technical module is responsible for providing the power required for sensor nodes.

At the software layer, the hardware abstraction layer realizes abstraction on the hardware platform (modules of power supply, data acquisition, data processing and wireless communication), which hides the details of the specific platform's hardware interface. The service layer includes communication services, sensor services, power management services and real-time kernel. The application layer is defined by users based on the needs of specific use, which uses the interface offered by the service layer to easily design top software.

2.3 Wireless Sensor Network Protocol Stack

Network architecture is a collection of protocol layering and network protocols in a network, which defines and describes the functions which a network and its

components should complete. The same as other networks, wireless sensor network also has its own protocol stack. Figure 2(b) shows the protocol architecture of wireless sensor network, which is different from a traditional computer network [6], [7]. The application of different software at the application layer can achieve different application purposes of sensor network. The transport layer provides functions of error control and flow control. The network layer is responsible for routing the data offered by the transport layer to information collection nodes. The data link layer is mainly responsible for nodes' access so as to reduce nodes' transport conflicts. The physical layer is responsible for bit-stream transport.

3 Hardware Design of Sensor Nodes

3.1 The Overall Structure of Hardware

The Internet of Things is a wireless sensor network where many sensor nodes are arranged in certain areas in a random way. Sensor nodes are the most important and basic components of the Internet of Things. Different sensor nodes have different function: e.g. humidity sensor nodes acquire humidity, image sensor nodes capture images, and temperature sensor nodes acquire temperature, etc. No matter what type of sensor nodes, in addition to the differences in sensor modules, the composition of other parts is almost the same. A sensor node is actually an embedded system which is consists of the microprocessor, memory, sensors, network interface, camera interface and power supply. The Figure 3 shows the structure.

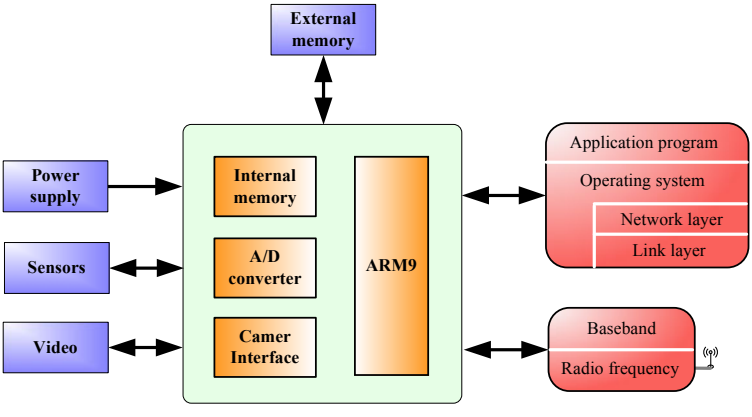


Fig. 3. The overall hardware design of sensor nodes in the Internet of Things.

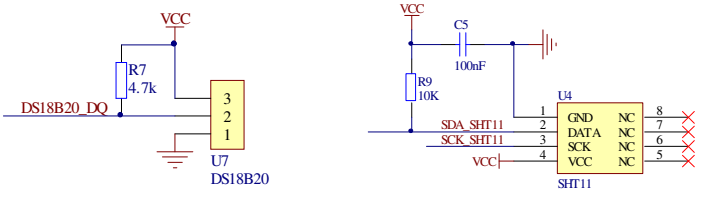
3.2 The Processor Module

The processor module is the core of sensor nodes. It completes the system resource management, task scheduling, information exchange, command analyzing,

multi-channel sampling, and other tasks. The selection of processors is essential to the design of sensor nodes. In this project, the processor module selects NXP Semiconductors's LPC2478 powered by the ARM7TDMI-S core, which is designed for a wide range of applications that require advanced communications. The LPC2478 microcontroller incorporates an LCD controller, a 10/100 Ethernet Media Access Controller (MAC), a USB full-speed Device/Host/OTG Controller with 4 kB of endpoint RAM, four UARTs, two Controller Area Network (CAN) channels, an SPI interface, two Synchronous Serial Ports (SSP), three I2C interfaces, an I2S interface, 32-bit timers, a 10-bit ADC, a 10-bit DAC, two PWM units, and up to 160 fast GPIO lines. All of these features make the LPC2478 particularly suitable for the hardware's requirements on a processor in this paper.

3.3 Temperature Sensors

The temperature sensors used in this project are high precision digital sensors DS18B20 designed by Dallas Company. The core functionality of the DS18B20 is its direct-to-digital temperature sensor and the resolution of the temperature sensor is user-configurable to 9, 10, 11, or 12 bits. The default resolution at power-up is 12-bit. The DS18B20 uses Maxim's exclusive 1-Wire bus protocol that uses single signal lines to transmit not only the clock but also data. In additional, each DS18B20 has a unique 64-bit serial code. This feature allows multiple DS18B20s to function on the same 1-Wire bus. Thus, it is simple to use one microprocessor to control many DS18B20s distributed over a large area. Its circuit diagram is shown in Figure 4(a).



(a) (b)

Fig. 4. (a) The circuit diagram of temperature sensors. (b) The circuit diagram of humidity sensors.

3.4 Humidity Sensors

The humidity sensor uses SHT11 produced by the Swiss Sensirion, which is Sensirion's family of surface mountable relative humidity and temperature sensors. It widely applied in HVAC, automobiles, consumer electronics, automatic control and other fields and the applied CMOSens technology guarantees excellent reliability and long term stability. The SHT11 has many characteristics, such as programmable

regulation of measurement resolution (8/12/14 bit data), a tiny foot print, 2-wire serial interface, low power consumption, etc. Its circuit diagram is shown in Figure 4(b).

3.5 The Wireless Communication Module

The wireless communication module uses Texas Instruments's CC2430 which can meet the need of high performance and low power in 2.4 GHz IEEE 802.15.4 band based on ZigBee technology. The CC2430 is a true System-on-Chip (SoC) solution, which combines a high-performance 2.4GHz DSSS (Direct Sequence Spread Spectrum) RF transceiver core and an industrial-grade compact 8051 controller. The enhanced 8051 MCU has 128KB programmable flash memory, 8KB RAM and many other powerful features. The CC2430 is highly suited for systems where ultra low power consumption is required. Its working current loss is 27 mA. In receiving and transmitting mode, the current consumption is respectively less than 27 mA or 25 mA. Its circuit diagram is shown in Figure 5.

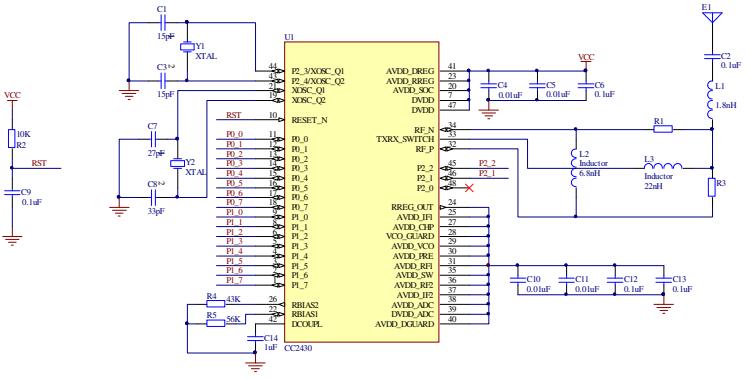


Fig. 5. The circuit diagram of wireless communication module

3.6 The CMOS Camera Module

In this project, the OV6620 CMOS Image sensors is used to capture image, which are single-chip video/imaging camera devices designed to provide a high level of functionality in a single, small-footprint package. The OV6620 incorporates sensor includes a 356 x 292 resolution image array capable of operating up to 60 frames per second image capture, an analog signal processor, dual 8-bit Analog-to-Digital converters, analog video multiplexer, digital data formatter and video port, SCCB interface and registers. Thereinto, the video output can be programmed to provide video output in 4-bit/8-bit/16-bit digital formats. The SCCB interface provides access to all of the device's programmable internal registers including exposure control, gamma, gain, white balance, color matrix, windowing, and more. Its circuit diagram is shown in Figure 6.

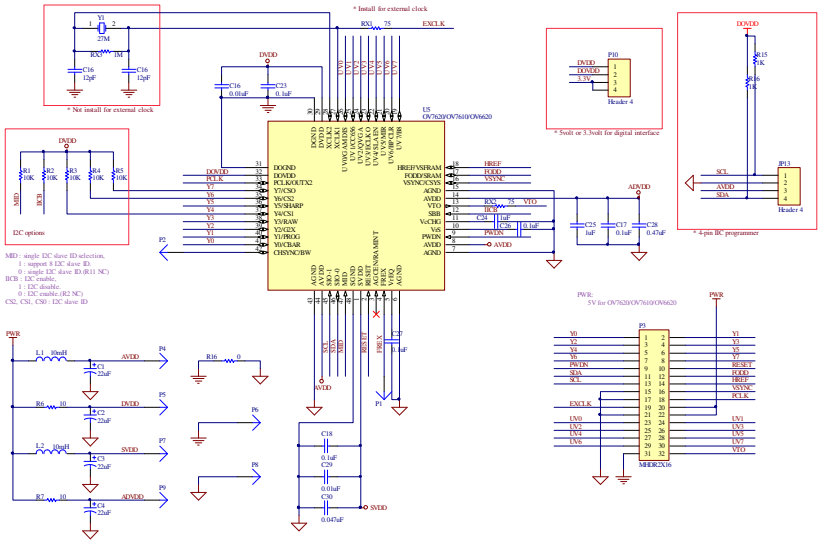


Fig. 6. Schematic circuit diagram of CMOS camera module

3.7 The Ethernet Interface

This system uses ENC28J60 network chip as the Ethernet interface. The ENC28J60 chip produced by Microchip is a highly integrated and fast Ethernet controller with an industry standard Serial Peripheral Interface (SPI). It is designed to interface directly with the SPI port available on many microcontrollers. The ENC28J60 is IEEE 802.3 compatible Ethernet controller and integrates MAC and 10BASE-T PHY. It also provides an internal DMA module for fast data movement and hardware assisted checksum calculation for various network protocols. Communication with the host controller is implemented via an interrupt pin and the SPI interface with clock speeds up to 20 MHz. Its circuit diagram is shown in Figure 7.

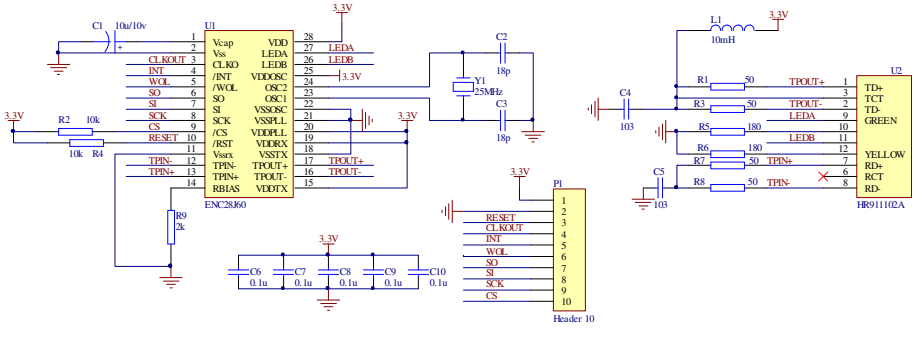


Fig. 7. Schematic circuit diagram of Ethernet module

4 Conclusion

It is the hot field of recent international research for the Internet of Things. In the other hand, recent advances in wireless communication, semiconductor, embedded system and sensor technology have made it possible to build low cost, low power consumption, high performance wireless sensor nodes. In this paper, the Internet of Things is used in the Nilaparvata lugens monitoring and warning system. The sensor nodes in the Nilaparvata lugens monitoring system based on the Internet of Things is also designed, which uses high performance arm processor as control core to realize the data collection, storage and wireless transmission, improving the system speed and transfer efficiency. This paper has introduced its overall structure, and described in detail the hardware design and implementation of the various components of sensor nodes. The future work is mainly reflected in the selection of CPU with lower costs, smaller size, and lower power consumption to replace LPC2478. Through the use of small-size components, the size of nodes can be further reduced.

References

1. Conner, M.: Sensors empower the "Internet of Things", pp. 32–38 (2010)
2. APPENDIX F: THE INTERNET OF THINGS (BACKGROUND). Disruptive Technologies: Global Trends 2025. SRI Consulting Business Intelligence (2010)
3. Hibino, H.: Insect-borne viruses in rice. In: Harris, K.F. (ed.) Advances in Disease Vector Research, vol. 6, pp. 209–241. Springer, New York (1989)
4. Hibino, H.: Biology and epidemiology of rice viruses. Annu. Rev. Phytopathol. 34, 249–274 (1996)
5. Du, P.V., Cabunagan, R.C., Cabauatan, P.Q., Choi, H.S., Choi, I.R., Chien, H.V., Huan, N.: Yellowing syndrome of rice: etiology, current status, and future challenges. Omonrice 15, 94–101 (2007)
6. Sabbineni, H.: Location-aided flooding: an energy-efficient data dissemination protocol for wireless-sensor networks. IEEE Transactions on Computers 54(1), 36–46 (2005)
7. Shih, E., Cho, S., Ickes, N., et al.: Physical layer driven protocol and algorithm design for dnergy-efficient wireless sensor networks. In: Proccessings of the ACM MobiCom 2001, pp. 272–286. ACM Press, Rome (2001)
8. Shih, C.H., Shieh, C.K., Hwang, W.S.: An integrated rate control scheme for TCP-friendly MPEG-4 video transmission. Signal Processing: Image Communication 23(2), 101–115 (2008)
9. DAVICOM Semiconductor Inc. DM9000A Ethernet Controller with General Processor Interface Data Sheet (2005)

Finding All Maximum Non-crossing Subsets of Nets Using Dynamic Programming

Xinguo Deng[1,2], Jiarui Chen[1,2], and Shu Zhang[2]

[1] Center for Discrete Mathematics and Theoretical Computer Science, Fuzhou University, Fuzhou, China
[2] College of Software, Fuzhou University, Fuzhou, China
xgdeng@fzu.edu.cn

Abstract. Current algorithms of the dynamic programming or longest increasing subsequence with the time complexity of $O(n^2)$ or O(nlogn) can only find one maximum non-crossing subset of nets even if there is more than one with the same length. In order to traverse all possible ones in the VLSI (Very Large Integration Circuits) wire routing, a dynamic programming algorithm is adopted and modified to calculate the length of maximum non-crossing subsets. For this purpose, an adjacent list is created for the traverse and a recursive function is used to output all maximum non-crossing subsets of nets. The effectiveness of the algorithm with a time complexity of $O(n^2)$ is illustrated through the theoretical analysis and experimental results of corresponding C++ program.

Keywords: maximal non-crossing subsets; dynamic programming; adjacent list; recursive function; circuit wiring.

1 Introduction

A circuit consists of a set of modules and a set of nets. Each net specifies a subset of points, called terminals, on the boundary of the modules. The layout problem is to interconnect the modules as specified by the nets in terms of different technological design rules. Due to the complexity of the problem, VLSI layout design is typically decomposed into three phases: placement, global routing, and detailed routing. In the placement phase, circuit modules are geometrically positioned on a layout surface (chip). In the global routing phase, the routing region is partitioned into simple subregions, each called an elementary region, and global assignment of the wiring paths is determined for each net. In the detailed routing phase, detailed wirings of the individual routing regions are given. [1] The crossing distribution problem occurs before the detailed routing. It is observed that nets crossing each other are more difficult to route than those nets that do not cross. The layout of crossing nets must be realized in more than two layers, thus requiring a larger number of vias. [2]

Documents [3, 4] and [5] studied the problem of maximum non-crossing subset (MNS) of nets using dynamic programming [6, 7] and longest common subsequence [8] respectively. They have the complexity of $O(n^2)$ and O(nlogn) respectively. However, these algorithms can only find one maximum non-crossing subset of nets even if there is more than one with the same length. In order to traverse all possible ones in the

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 579–586.
springerlink.com

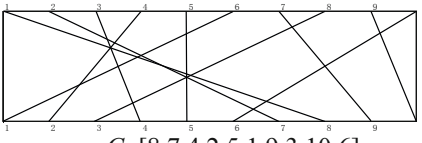
VLSI (Very Large Integration Circuits) wire routing, a modified dynamic programming algorithm is brought forward in this paper. The effectiveness of this algorithm whose time complexity is $O(n^2)$ is illuminated by the theoretical analysis and experimental results of corresponding C++ program.

The remainder of the paper is organized as follows. Section 2 introduces the problem of maximum non-crossing subsets of nets in the circuit wiring. Section 3 illustrates the core thoughts of the algorithm using dynamic programming and analyzes the time complexity of the algorithm. Section 4 is the C++ program implementation corresponding to the algorithm. Section 5 provides the experimental results of the C++ program corresponding to the algorithm. The last section concludes and gives the directions for future work in this field.

2 Maximum Non-crossing Subsets of Nets

A routing channel has n pins on the top side and n pins with a permutation C on the bottom side. Pin i on the top side of the channel is to be connected to pin C_i on the bottom side, $1 \leqslant i \leqslant n$. The pair (i, C_i) is called a **net**. In total, we have n nets that are to be connected or routed. Suppose that we have two or more routing layers, of which one is a *preferred layer*. For example, in the preferred layer it may be possible to use much thinner wires, or the resistance in the preferred layer may be considerably less than in other layers. Our task is to route as many nets as possible in the preferred layer. The remaining nets will be routed, at least partially, in the other layers. Since two nets can be routed in the same layer if they do not cross, our task is equivalent to finding a maximum non-crossing subset (MNS) of the nets. Such a subset has the property that no two nets of the subset cross. Since net (i, C_i) is completely specified by i, we may refer to this net as net i.[3]

Consider the example in figure 1, the nets（1,8）and（2,7）(or equivalently, the nets 1 and 2）cross and so cannot be routed in the same layer. The nets (1,8), (7,9), and (9,10) do not cross and so can be routed in the same layer. These three nets do not constitute a MNS as there is a larger subset of non-crossing nets. There are two MNS of the routing instance given in figure 1. They are four nets {(4,2),(5,5),(7,9),(9,10)} and {(3,4),(5,5),(7,9),(9,10)}.



C=[8,7,4,2,5,1,9,3,10,6]

Fig. 1. A wiring instance.

3 Algorithm of Dynamic Programming

The algorithm outlined below finds MNS with dynamic programming efficiently, using only arrays and an adjacent list [10]. It processes the sequence elements in order, maintaining the MNS found so far.

Array element *dp[i]* represents the length of LIS that ends on the position *i*. The state transition equation is $dp[i]=max\{ dp[j]+1 \mid 1 \leq j < i,\ bottom[j]<bottom[i] \}$. Here position *j* is the possible predecessor of position *i* in LIS.

Adjacent list with virtual source node S and target node T can be created based on the array *dp[ ]*. Variable *length* is the length of the longest increasing subsequence. Elements with *dp[i]=length and dp[i]=1* are linked to the source node S and target node T respectively. Predecessor elements are linked to their successor element.

Depth-first search (DFS) retraces the adjacent list from the source node S to the target node T and then outputs MNS sequentially.

The algorithm, then, proceeds as follows.

```
input circuit wiring;
for i = 1, 2, ... n
    {for j = 1, 2, ... i-1
        find possible predecessor j of i in LIS
        such that wire[j].bottom< wire[i].bottom;
     calculate length dp[i] of every wire in LIS; }
create an adjacent list of all LIS based on dp[ ];
recursively output all MNS using DFS.
```

The time complexity of the algorithm is equivalent to that of calculating the length of LIS, and it is $O(n^2)$.

4 C++ Implementation

4.1 Design

The methodology of top-down modular is adopted to design the program. Therefore, we considered the structure of the C++ program with three big modules: 1. inputting the circuit wiring, 2. the central module to find the MNS using dynamic programming, 3. outputting the result.

The second module computes the length of the MNS and finds the MNS using the dynamic programming. This module is further divided into three sub modules: “LIS” module, “ConstructAdjacentList” module and “addEdge” module. The last module outputs all MNS implied in the adjacent list. This module is further divided into two sub modules: “OutputMNS” module and “DFS” module. A seventh module “Welcome” that displays the function of the program is also desirable. While this module is not directly related to the problem at hand, the use of such a modular enhances the user-friendliness of the program.

4.2 Program Plan

In last section, we have already pointed out the need for seven program modules. A root (or main) module invokes five modules in the following sequence: “Welcome” module, “InputCircuitWiring” module, “LIS” module, “ConstructAdjacentList” module and “OutputMNS” module. “ConstructAdjacentList” and “OutputMNS” modules invoke “addEdge” and “DFS” modules respectively.

A C++ program is designed by following the modular structure in figure 2. Each program module is coded as a function. The root module is coded as the function "main"; "Welcome", "InputCircuitWiring", "LIS", "ConstructAdjacentList", "addEdge", "OutputMNS", and "DFS" modules are implemented through different functions.

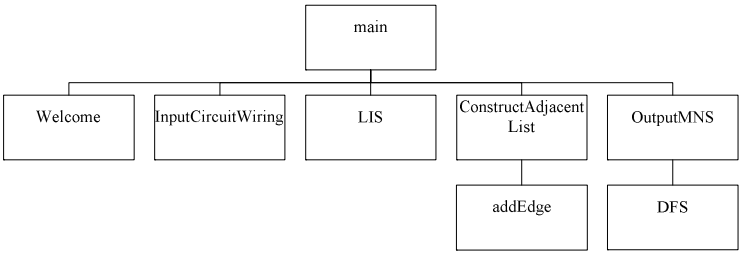


Fig. 2. Modular structure.

4.3 Program Development

Function "Welcome" explains the function of the whole C++ program. Function "InputCircuitWiring" informs the user that the input is expected as a permutation of pins on the bottom of the channel corresponding to the pins on the top. The size of the group of the pins is determined first, so the total number of pins is needed before an input begins. In our program, the input process is implemented by importing the input data from a text file called "input.txt". The result is outputted to the text file "output.txt".

Struct *wire* consists of top side *top* and bottom side *bottom*. Bottom side is actually the wire permutation (i.e., C_i corresponding to i). The following figure 3 details the functions of "LIS", "addEdge" and "constructAdjacentList". The first and last functions are invoked by main function. The function "addEdge" is invoked by function "constructAdjacentList". The function "addEdge" helps to construct adjacent list.

The function "LIS()" calculates the length and implicitly helps to construct adjacent list of LIS for the later output using dynamic programming. Integer length represents the length of LIS. Array element *dp[i]* represents the length of LIS that ends on the position *i*. The state transition equation is $dp[i]=max\{\ dp[j]+1 \mid 1 \leqq j < i,\ bottom[j]<bottom[i]\ \}$. There are double loops in the function "LIS". The inner loop finds the possible predecessor of *dp[i]*. The outer loop calculates the length of current wire in the LIS. When the whole loop terminates, the length of LIS (i.e., MNS) is achieved.

The function "addEdge (int u, int v)" adds an edge *(u, v)*. The function "ConstructAdjacentList" constructs adjacent list for the later output via function "DFS". At first, two added nodes, source node S and target node T are created. Then, the adjacent list is initialized. Next, there are loops of two layers. In the outer loop, the last and first elements of LIS are linked to the source node S and target node T respectively. In the inner loop, predecessor elements are linked to their successor element. Array of length *dp[ ]* and array of struct wire *wire[ ]* are used in the condition of judgment.

```
int LIS(int n)  //find LIS using dynamic programming, O(N^2)
{   int length=0,i,j;
    //memset(dp,-1,sizeof(dp));
    for(i=1 ;i<=n ;i++)
    {   int tmp=0;
        for(j=1 ;j<i ;j++)
        {   //find possible predecessor of dp[i]
            if(wire[j].bottom<wire[i].bottom && dp[j]>tmp)
                tmp=dp[j];
        }
        dp[i]=tmp+1;
        if(dp[i]>length)    //find maximal length of LIS
            length=dp[i];
    }
    return length;
}
inline void addEdge(int u,int v)    //add an edge (u,v)
    {g[u].push_back(v);}
void ConstructAdjacentList(int n,int len)   //ConstructAdjacentList
{   S=0,T=n+1;
    int i,j;
    for(i=0 ;i<=n+1 ;i++)
        {g[i].clear();}
    for(i=1 ;i<=n ;i++)
    {   if(dp[i]==len)  //The last element of LIS is linked to the node S.
            addEdge(S,i);
        if(dp[i]==1)    //The first element of LIS is linked to the node T.
            addEdge(i,T);
        for(j=1 ;j<i ;j++)
            {if(dp[i]==dp[j]+1 && wire[i].bottom>wire[j].bottom)
                addEdge(i,j);
            }
    }
}
```

Fig. 3. LIS, addEdge and ConstructAdjacentList functions

In figure 4, function "OutputMNS" initializes variable *pathNum* with zero first, then invokes the recursive function "DFS(S, 0, len, pathNum)". Function "DFS(int u, int dep, int len, int& pathNum)" retraces from the source node S to the target node T using DFS, and then outputs MNS sequentially. Here the source node S is the successor of the last element and the target node T is the predecessor of the first element in the LIS.

The details of other functions "Welcome", "InputCircuitWiring" and "main" are omitted here. The effects of these functions will be illustrated in the next section.

```
void DFS(int u,int dep,int len,int& pathNum)    //DFS outputs all MNS
{   int i;
    if(dep==len+1)
    {   printf("The path %d of MNS:\n",++pathNum);
        printf("top\tbottom\n");
        for(i=dep-2 ;i>=0 ;i--)
            printf("%d\t%d\n",wire[path[i]].top,wire[path[i]].bottom);
    }
    for(i=0 ;i<g[u].size() ;i++)
    {   path[dep]=g[u][i];
        DFS(g[u][i],dep+1,len,pathNum);
    }
}
void OutputMNS(int n,int len)
{   int pathNum=0;
    printf("The length of MNS is %d.\n",len);
    printf("The following is all maximum non-crossing subsets of nets.\n");
    DFS(S,0,len,pathNum);
}
```

Fig. 4. DFS and OutputMNS functions

5 Experimental Results

5.1 Example Illustration

TABLE I shows the dynamic changes of some variables and arrays in the process of program running for the example in figure 1. Variable *i* represents the time of loop. It is the position of pins on the top side, too. Array element *wire[i].bottom* records the position of pins on the bottom side of the current wire. Array element *dp[i]* represents the length of LIS that ends on the position *i*.

Table 1. Dynamic changes of some variables and arrays

i	1	2	3	4	5	6	7	8	9	10
wire[i].bottom	8	7	4	2	5	1	9	3	10	6
dp[i]	1	1	1	1	2	1	3	2	4	3

According to the array *dp [ ]*, the following adjacent list can be constructed.

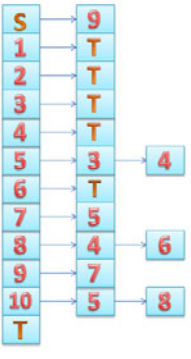


Fig. 5. Output of all maximum non-crossing subsets

In order to elaborate more clearly, the above figure 5 is converted to its equivalent figure 6. Two paths marked with red color can be found starting from the source node S and ending at the end node T. They are S → 10 → 9 → 5 → 4 → T and S → 10 → 9 → 5 → 2 → T. All maximum non-crossing subsets are achieved after the previous sequences are reversed. They are 2 → 5 → 9 → 10 and 4→ 5 → 9 → 10.

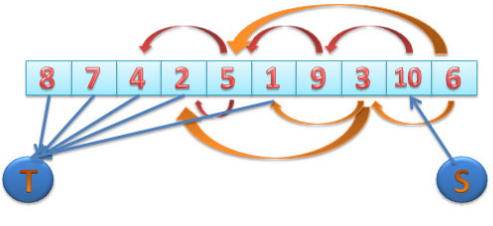


Fig. 6. Output of all maximum non-crossing subsets

5.2 Final Output

In figure 7 our program outputs length and all maximum non-crossing subsets of nets for the example in figure 1. The length of MNS is 4. Set {(3,4),(5,5),(7,9),(9,10)} is a MNS of nets {(1,8),(2,7),(3,4),(4.2),(5,5),(6,1),(7,9),(8,3),(9,10),(10,6)}. And {(4,2),(5,5),(7,9),(9,10)} is another MNS of nets in figure 1.

```
//////////////////////////////////////////////////////////////////
  This program finds all maximum non-crossing
  subsets of nets using dynamic programming.
//////////////////////////////////////////////////////////////////
Input the size of wires:
Input the array of pins on the bottom side:
The length of MNS is 4.
The following is all maximum non-crossing subsets of nets.
The path 1 of MNS:
top   bottom
3     4
5     5
7     9
9     10
The path 2 of MNS:
top   bottom
4     2
5     5
7     9
9     10
```

Fig. 7. Output of all maximum non-crossing subsets

6 Conclusion

In this paper, a modified dynamic programming algorithm with the complexity $O(n^2)$ to output all maximum non-crossing subsets of nets and to calculate its length in the circuit wiring was presented. Furthermore, the data structure of adjacent list is applied in the algorithm to improve the efficiency.

The complexity of this algorithm is analyzed theoretically. In contrast to this, the previous algorithms of using either the dynamic programming or longest increasing subsequence can only find one maximum non-crossing subset of nets even if there is more than one with the same length. Further, a C++ program is developed to test this enhanced algorithm, and the satisfactory experimental results of the C++ program are presented.

Though some achievements about maximum non-crossing subsets of nets are made in this paper, more efficient algorithm of minimizing wiring congestion is our next research task.

Acknowledgements

The work was supported by the Natural Science Foundation of Fujian Province under Grant 2009J05142, the Natural Science Foundation of Fuzhou University under Grant 0220826788 and Technology Development Fund of Fuzhou University under Grant 2008-xy-16.

References

1. Wolf, W.: Modern VLSI Design: System-on-Chip Design, 3rd edn., pp. 510–522. Pearson Education, Inc., London (2003)
2. Song, X., Wang, Y.: On the crossing distribution problem. ACM Transaction on Design Automation of Electronic Systems 4(1), 39–51 (1999)
3. Sahni, S.: Data Structures, Algorithms, and Applications in C++, 2nd edn., pp. 735–740. McGraw-Hill, Inc., New York (2004)
4. Wang, X.: Algorithm Design and Analysis, 3rd edn., pp. 74–76. Publishing House of Electronics Industry (2007)
5. Deng, X., Zhong, R.: Finding Maximum Noncrossing Subset of Nets Using Longest Increasing Subsequence. In: 2nd IEEE International Conference on Multimedia and Computational Intelligence, pp. 697–700. IEEE Press, Wuhan (2010)
6. Skiena, S.S.: The Algorithm Design Manul, 2nd edn., pp. 273–315. Springer, Inc., Heidelberg (2008)
7. Pandey, H.M.: Design Analysis and Algorithms, 1st edn., pp. 258–286. University Science Press (2008)
8. Iliopoulos, C.S., Sohel Rahman, M.: A New Efficient Algorithm for Computing the Longest Common Subsequence. Theory of Computing Systems 45(2), 355–371 (2009)
9. Krusche, P., Tiskin, A.: Parallel Dynamic programmings in Scalable Time and Memory. In: Wyrzykowski, R., Dongarra, J., Karczewski, K., Wasniewski, J. (eds.) PPAM 2009. LNCS, vol. 6067, pp. 176–185. Springer, Heidelberg (2010)
10. Neapolitan, R., Naimipour, K.: Foundations of Algorithms, 4th edn., pp. 277–328. Jones and Bartlett Publishers, LLC (2009)

Non-fragile $\mathcal{H}_\infty$ Control with Pole Constraints for a Class of Nonlinear Sample-Data System

Shigang Wang*

School of of Mechanical and Electrical Engineering, Heilongjiang University, Harbin, 150080, China
wsg@hlju.edu.cn

Abstract. This paper deals with non-fragile $\mathcal{H}_\infty$ control with pole constraints for a class of nonlinear sample-data system. Firstly, the continuous control plant of sampled-data system is transformed into an uncertain discrete system with bounded nonlinearities model. Then, the pole constraints theory and the linear matrix inequality approach are incorporated to design a non-fragile $\mathcal{H}_\infty$ controller against possible perturbations, which results in the closed-loop system being $\mathcal{D}$-stable and the system's performance index being less than a prescribed scalar. Simultaneously, the existence condition and the design approach of non-fragile controller are derived. Finally, simulation examples are presented to illustrate the feasibility of the proposed control algorithm.

Keywords: Sampled-data system, Nonlinear system, Non-fragile control, Pole constraints, Linear matrix inequality.

1 Introduction

Sampled-data control theory has been well-developed in the last two decades[1], and established methodology is available for analysis and synthesis, which is embodied by the following two aspects.

On one hand, such control problem can usually be equal to designing proper controller such that the closed-loop system is asymptotically stable and its performance index is less than some prescribed scalar. However, most of these results are based on the accurate state feedback controllers[2,3,4,5]. In fact, due to the existence of the parameter drift, accuracy problem and other factors. the parameters of the controller are possible to accrue gain variations. And relatively small perturbation of the controller parameters might destabilize the closed-loop system, even lead to the performance degradation. This is known as the non-fragile control problem[6]. So far this Problem has been widely investigated by many researchers[7,8,9,10].

* This work was supported by the Natural Science Foundation of Heilongjiang Province China (F200918), Youth Foundation of Heilongjiang University and Open Laboratory Foundation of Heilongjiang University.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 587–594.
springerlink.com

On the other hand, it is desirable to construct sampled-data system to achieve better transient performance. To this aim, a more practical way is to put the closed-loop poles in a specified region[11,12,13].

In this paper, we will design the non-fragile $\mathcal{H}_\infty$ control with pole constraints for a class of nonlinear sample-data system. At the same time, the existence condition and design approach of non-fragile controller with pole constraints are presented.

2 Preliminaries and Problem Statement

Let the control plant of sampled-data system be given by

$$\begin{cases} \dot{x}(t) = (A_0 + \Delta A_0)x(t) + (B_0 + \Delta B_0)u(t) + f(x,u,t) + B_1\omega(t) \\ z(t) = C_1 x(t) + H_2\omega(t) \end{cases} \tag{1}$$

where $x(t) \in \mathbf{R}^n$ is the state, and $u(t) \in \mathbf{R}^m$ is the control input, $z(t) \in \mathbf{R}^l$ is measured output, $\omega(t) \in \mathbf{R}^p$ is the external disturbance input that belongs to $\mathbf{L}_2[0,\infty]$, A_0, B_0, B_1, C_1, H_2 are the constant matrices of appropriate dimensions. $f(x,u,t)$ is the uncertain nonlinear function vector, and satisfies the Lipschitz condition with $f(0,0,0) = 0$.

Assumption 1. The continuous plant is time-driven with a constant sampling period $h(h > 0)$.

Discretizing system (1) in one period, we can obtain the discrete state equation of the plant of sampled-data system

$$\begin{cases} x(k+1) = (G_0 + \Delta G_0)x(k) + (H_0 + \Delta H_0)u(k) + \bar{f}(x,u,k) + H_1\omega(k) \\ z(k) = C_1 x(k) + H_2\omega(k) \end{cases} \tag{2}$$

where

$$G_0 = e^{A_0 h}\ ,\ H_0 = \int_0^h e^{A_0(h-\omega)}d\omega B_0,\ ,\ H_1 = \int_0^h e^{A_0(h-\omega)}d\omega B_1$$
$$\bar{f}(x,u,k) = \int_0^h e^{Aw}d\omega f(x,u,t)$$

$\Delta G_0, \Delta H_0$ are uncertain matrix, and satisfy the following form

$$[\Delta G_0 \quad \Delta H_0] = MF(k)[E_0 \quad E_1] \tag{3}$$

In the sequel, we assume that the nonlinear uncertainty $\bar{f}(x,u,k)$ satisfies

$$\bar{f}^{\mathrm{T}}(x,u,k)\bar{f}(x,u,k) \le x^{\mathrm{T}}(k)Q_{11}^{\mathrm{T}}Q_{11}x(k) \tag{4}$$

where Q_{11} are known positive definite matrix.

The objective of this paper is to design a non-fragile state feedback controller

$$u(k) = (\mathcal{K} + \Delta\mathcal{K})x(k) \tag{5}$$

where, $\mathcal{K}$ is the nominal controller gain, and $\Delta\mathcal{K}$ represents the gain perturbations. In general, there exist the following two types of perturbations:

Type 1: the additive form

$$\Delta\mathcal{K} = MF(k)E_2 \tag{6}$$

Type 2: the multiplicative form

$$\Delta\mathcal{K} = MF(k)E_3\mathcal{K} \tag{7}$$

where M, E_2 and E_3 are known constant matrices, and $F(k)$ is uncertain parameter matrix with satisfying $F^{\mathrm{T}}(k)F(k) \leq I$.

Remark 1. The controller gain perturbations can result from the actuator degradations, as well as from the requirements for re-adjustment of controller gains during the controller implement stage[14]. These perturbations in the controller gains are modeled here as uncertain gains that are dependent on the uncertain parameters. The model of additive uncertainties [9] and multiplicative [10] are used to describe the controller gain variations.

Definition 1. A real matrix A is $\mathcal{D}$-stable, i.e., has all its eigenvalues in the LMI region D,which can be denoted as $\sigma(A) \subset \mathcal{D}$, $\sigma(A)$ is spectral set of matrix A.

Lemma 1. [15] Let $A \in \mathbf{R}^{n\times n}$be a given matrix. The eigenvalues of A belong to $D(-\alpha, r)$ if and only if there exists a symmetric matrix $P \in \mathbf{R}^{n\times n}$ such that

$$\begin{bmatrix} -P^{-1} & r^{-1}(A+\alpha I) \\ r^{-1}(A+\alpha I)^{\mathrm{T}} & -P \end{bmatrix} < 0$$

Lemma 2. [16] For given matrices $Q = Q^{\mathrm{T}}$, H, and E, with appropriate dimensions

$$Q + HF(t)E + E^{\mathrm{T}}F^{\mathrm{T}}(t)H^{\mathrm{T}} < 0$$

holds for all $F(k)$ satisfying $F^{\mathrm{T}}(k)F(k) \leq I$, if and only if there exists $\varepsilon > 0$

$$Q + \varepsilon HH^{\mathrm{T}} + \varepsilon^{-1}E^{\mathrm{T}}E < 0$$

Lemma 3. [17] Let (G, H, C, D) be a minimal state space realization of system $T(s)$, let γ be a positive scalar. The system is stable and the inequality

$$\|T(s)\|_\infty \leq \gamma$$

is true, if and only if there exists a positive definite matrix P such that the following linear matrix inequality (LMI) holds:

$$\begin{bmatrix} -P + G^{\mathrm{T}}PG & G^{\mathrm{T}}PH & C^{\mathrm{T}} \\ H^{\mathrm{T}}PG & -\gamma^2 I + H^{\mathrm{T}}PH & D^{\mathrm{T}} \\ C & D & -I \end{bmatrix} < 0$$

3 Main Result

Theorem 1. Consider system(2), there exist scalars $\varepsilon_1 > 0, \varepsilon_2 > 0$ and matrices $\mathcal{K}$, $\Delta\mathcal{K}$, $X_1 = X_1^{\mathrm{T}} > 0$, and Y_1 such that for all admissible uncertainties the following matrix inequality hold:

$$\begin{bmatrix} -X_1 & * & * & * & * & * & * & * \\ 0 & \gamma^2 I & * & * & * & * & * & * \\ \varphi_{31} & -X_1 & * & * & * & * & * & * \\ C_1X_1 & H_2 & 0 & I & * & * & * & * \\ 0 & 0 & \varepsilon_1 M^{\mathrm{T}} & 0 & -\varepsilon_1 I & * & * & * \\ \varphi_{61} & 0 & 0 & 0 & 0 & -\varepsilon_1 I & * & * \\ 0 & 0 & \varepsilon_2 M^{\mathrm{T}} H_0^{\mathrm{T}} & 0 & 0 & 0 & -\varepsilon_2 I & * \\ E_2X_1 & 0 & 0 & 0 & 0 & 0 & 0 & -\varepsilon_2 I \end{bmatrix} < 0 \tag{8}$$

$$\begin{bmatrix} -rX_1 & * & * & * & * & * \\ \varphi_{21} & -rX_1 & * & * & * & * \\ 0 & \varepsilon_3 M^{\mathrm{T}} & -\varepsilon_3 I & * & * & * \\ \varphi_{41} & 0 & 0 & -\varepsilon_3 I & * & * \\ 0 & \varepsilon_4 M^{\mathrm{T}} H_0^{\mathrm{T}} & 0 & 0 & -\varepsilon_4 I & * \\ E_2X_1 & 0 & 0 & 0 & 0 & -\varepsilon_4 I \end{bmatrix} < 0 \tag{9}$$

where

$$\varphi_{21} = \varphi_{31} = G_0X_1 + Q_{11}X_1 + H_0Y_1$$
$$\varphi_{41} = \varphi_{61} = E_0X_1 + E_1Y_1 + E_2X_1$$

then $u(k) = (\mathcal{K} + \Delta\mathcal{K})x(k)$ is a additive non-fragile H_∞ control law with pole constraints of system (2) and $\mathcal{K} = Y_1X_1$.

Proof. system (2) can be transformed into the following form

$$\begin{cases} x(k+1) = Gx(k) + H_1\omega(k) \\ z(k) = C_1x(k) + H_2\omega(k) \end{cases} \tag{10}$$

where

$$G = G_0 + Q_{11} + \Delta G_0 + H_0\mathcal{K} + H_0\Delta\mathcal{K} + \Delta H_0\mathcal{K} + \Delta H_0\Delta\mathcal{K} \tag{11}$$

Owing to Lemma 3, one obtains

$$\begin{bmatrix} -P & * & * & * \\ 0 & \gamma^2 I & * & * \\ G & H_1 & -P^{-1} & * \\ C_1 & H_2 & 0 & I \end{bmatrix} < 0 \tag{12}$$

Substituting (9) into the left side of (10) yields

$$\mathcal{W} = \mathcal{W}_1 + \begin{bmatrix} 0 & 0\,0\,0 \\ 0 & 0\,0\,0 \\ \varphi_1 & 0\,0\,0 \\ 0 & 0\,0\,0 \end{bmatrix} < 0 \tag{13}$$

where

$$\mathcal{W}_1 = \begin{bmatrix} -P & * & * & * \\ 0 & \gamma^2 I & * & * \\ G_0 + Q_{22} + H_0 K & H_1 & -P^{-1} & * \\ C_1 & H_2 & 0 & I \end{bmatrix}$$

$$\varphi_1 = \Delta G_0 + H_0 \Delta \mathcal{K} + \Delta H_0 \mathcal{K} + \Delta H_0 \Delta \mathcal{K}$$

From (12), and together with (3) and (6) yields

$$\mathcal{W} = \mathcal{W}_1 + \varepsilon_1 \theta_1 \theta_1^{\mathrm{T}} + \varepsilon_1^{-1} \theta_2^{\mathrm{T}} \theta_2 + \varepsilon_2 \theta_3 \theta_3^{\mathrm{T}} + \varepsilon_2^{-1} \theta_4^{\mathrm{T}} \theta_4 \tag{14}$$

where

$$\theta_1 = \begin{bmatrix} 0 & 0 & M^{\mathrm{T}} & 0 \end{bmatrix}^{\mathrm{T}}, \qquad \theta_2 = \begin{bmatrix} E_0 + E_0 K + E_2 K & 0 & 0 & 0 \end{bmatrix}$$
$$\theta_3 = \begin{bmatrix} 0 & 0 & M^{\mathrm{T}} H_0^{\mathrm{T}} & 0 \end{bmatrix}^{\mathrm{T}}, \quad \theta_4 = \begin{bmatrix} E_2 & 0 & 0 & 0 \end{bmatrix}$$

From Lemma 2, (12) is equivalent to

$$\begin{bmatrix} \mathcal{W}_1 & * & * & * & * \\ \varepsilon_1 \theta_1^{\mathrm{T}} & -\varepsilon_1 I & * & * & * \\ \theta_2 & 0 & -\varepsilon_1 I & * & * \\ \varepsilon_2 \theta_3^{\mathrm{T}} & 0 & 0 & -\varepsilon_2 I & * \\ \theta_4 & 0 & 0 & 0 & -\varepsilon_2 I \end{bmatrix} < 0 \tag{15}$$

Pre- and Post-multiplying (13) by diag$\{P^{-1},\ I,\ I,\ I,\ I,\ I,\ I, I\}$ and its transpose respectively, one gets that (13) is equivalent to (7).

On the other hand, the closed-loop system (8) must be $\mathcal{D}$-stable, from Lemma 1, one obtains

$$\begin{bmatrix} -\gamma P & G^{\mathrm{T}} \\ G & -\gamma P^{-1} \end{bmatrix} < 0 \tag{16}$$

From Lemma 2 and Lemma 3, and together with (3) and (5) yields

$$\begin{bmatrix} \mathcal{W}_2 & * & * & * & * \\ \varepsilon_1 \theta_5^{\mathrm{T}} & -\varepsilon_3 I & * & * & * \\ \theta_6 & 0 & -\varepsilon_3 I & * & * \\ \varepsilon_2 \theta_7^{\mathrm{T}} & 0 & 0 & -\varepsilon_4 I & * \\ \theta_8 & 0 & 0 & 0 & -\varepsilon_4 I \end{bmatrix} < 0 \tag{17}$$

where

$$\mathcal{W}_2 = \begin{bmatrix} -rP & * \\ G_0 + Q_{11} + H_0 \mathcal{K} & -rP^{-1} \end{bmatrix}$$

$$\theta_5 = \begin{bmatrix} 0 & M^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}, \qquad \theta_6 = \begin{bmatrix} E_0 + E_1 K + E_2 & 0 \end{bmatrix}$$
$$\theta_7 = \begin{bmatrix} 0 & M^{\mathrm{T}} H_0^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}, \ \theta_8 = \begin{bmatrix} E_2 & 0 \end{bmatrix}$$

Pre- and Post-multiplying (13) by diag $\{P^{-1},\ I,\ I,\ I,\ I,\ I\}$ and its transpose respectively, one gets that (16) is equivalent to (8). Thus, this completes the proof.

Theorem 2. Consider system(2), if there exist scalars $\varepsilon_1 > 0, \varepsilon_2 > 0$ and matrices $\mathcal{K}$, $\Delta\mathcal{K}$, $X_2 = X_2^{\mathrm{T}} > 0$, and Y_2 such that for all admissible uncertainties the following matrix inequality hold:

$$\begin{bmatrix} -X_2 & * & * & * & * & * & * & * \\ 0 & \gamma^2 I & * & * & * & * & * & * \\ \psi_{31} & H_1 & -X_2 & * & * & * & * & * \\ C_1 X_2 & H_2 & 0 & I & * & * & * & * \\ 0 & 0 & \varepsilon_5 M^{\mathrm{T}} & 0 & \varepsilon_5 I & * & * & * \\ \psi_{61} & 0 & 0 & 0 & 0 & \varepsilon_5 I & * & * \\ 0 & 0 & \varepsilon_6 (H_0 M)^{\mathrm{T}} & 0 & 0 & 0 & \varepsilon_6 I & * \\ E_3 Y & 0 & 0 & 0 & 0 & 0 & 0 & \varepsilon_6 I \end{bmatrix} < 0 \tag{18}$$

$$\begin{bmatrix} -\gamma I & * & * & * & * & * \\ \psi_{21} & \gamma X_2 & * & * & * & * \\ 0 & \varepsilon_7 M^{\mathrm{T}} & -\varepsilon_7 I & * & * & * \\ \psi_{41} & 0 & 0 & -\varepsilon_7 I & * & * \\ 0 & \varepsilon_8 M^{\mathrm{T}} H_0^{\mathrm{T}} & 0 & 0 & -\varepsilon_8 I & * \\ E_3 Y_2 & 0 & 0 & 0 & 0 & -\varepsilon_8 I \end{bmatrix} < 0 \tag{19}$$

where

$$\psi_{21} = \psi_{31} = G_0 X_2 + Q_{11} X_2 + H_0 Y_2$$
$$\psi_{41} = \psi_{61} = E_0 X_2 + E_1 Y_2 + E_2 X_2$$

then $u(k) = (\mathcal{K} + \Delta\mathcal{K})x(k)$ is a multiplicative non-fragile H_∞ control law with pole constraints of system (2) and $\mathcal{K} = Y_2 X_2$.

Proof. Its proof is same with theorem1, so omitted.

4 Numerical Example

Consider the control plant of the sampled-data system with

$$A_0 = \begin{bmatrix} -1 & 4.5 \\ 0 & -1.5 \end{bmatrix},\ B_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix},\ B_1 = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix},\ C_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}^{\mathrm{T}},\ H_1 = H_2 = 0.1$$

$$M = \begin{bmatrix} 2 & 3 \\ -1 & 0 \end{bmatrix},\ E_0 = \begin{bmatrix} 0 & 0.1 \\ -0.2 & 0.3 \end{bmatrix},\ E_1 = \begin{bmatrix} 0 & -0.2 \\ 0.5 & 0.1 \end{bmatrix},\ E_2 = \begin{bmatrix} 0 \\ -0.2 \end{bmatrix},\ E_3 = \begin{bmatrix} 1 \\ 0.1 \end{bmatrix}$$

The disk $\mathcal{D}$(-1.0,1.8) is given, according to Theorem1, solve the corresponding non-fragile with pole constraints problem via the solver feasp in LMI toolboxes, it yields

$$X_1 = \begin{bmatrix} 0.4612 & -0.7026 \\ -0.7026 & 0.7322 \end{bmatrix},\ Y_1 = \begin{bmatrix} 0.2435\ 0.2001 \end{bmatrix}$$

$$\mathcal{K} = \begin{bmatrix} -2.0450\ -1.6890 \end{bmatrix}$$

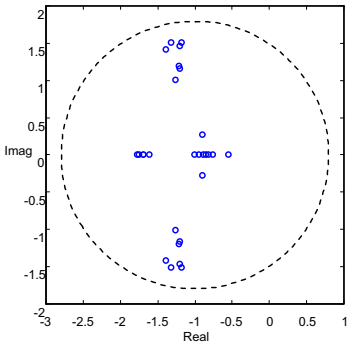


Fig. 1. The eigenvalues distribution under additive variation

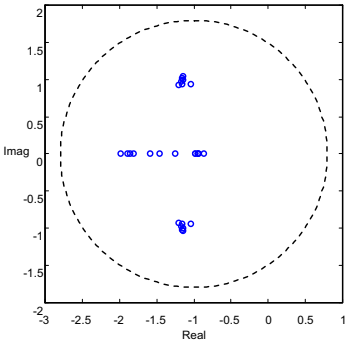


Fig. 2. The eigenvalues distribution under multiplicative variation

5 Conclusion

We have addressed the problem of non-fragile $\mathcal{H}_\infty$ control with pole constraints for a class of nonlinear sample-data system, and non-fragile controller is derived by solving a set of LMIs, which guarantees that the closed-loop system is $\mathcal{D}$-stable and system's performance index is less than a prescribed scalar. Finally, simulation example are presented to illustrate the feasibility of the proposed control algorithm.

References

1. Chen, T., Francis, B.: Optimal Sampled-Data Control System. Spring, New York (1995)
2. Junfeng, W., Qiang, W., Shanben, C.: Robust stability for sampled-data systems. Control Theory & Applications 18(8), 99–102 (2001)
3. Junfeng, W., Qiang, W., Shanben, C.: Robust control of a class of sampled-data systems with structured uncertainty. Control and Decision 17(5), 681–685 (2002)
4. Fuchun, L., Yu, Y., Fenghua, H.: Robust H_∞ controller design for sampled-data systems with parametric uncertainties. Journal of Systems Engineering and Electronics 20(2), 371–378 (2009)
5. Guangxiong, W., Yanwen, L., Zhen, H.: H_∞ mixed sensitivity design for sampled-data systems. Control Theory & Applications 23(3), 351–354 (2006)
6. Keel, L.H., Bhattacharyya, S.P.: Robust, fragile, or optimal. IEEE Trans. Automatic Control 42(8), 2678–2683 (1997)
7. Wang, S. (ed.): Resilient control of a class of state-delay sampling system, Nanjing, China, September 21–23, vol. 2 (2009)
8. Shigang, W., Zhiqiang, H., Yingsong, L. (eds.): Non-fragile H-infinity Control for a Class of Nonlinear Sampled-data System, Changsha, China, September 26–28, vol. 2 (2010)
9. Yang, G.L., Wang, J.L., Lin, C.: H_∞ control for linear Systems with additive gain variations. International Journal of Control 73(16), 1500–1506 (2000)

10. Yang, G.L., Wang, J.L.: Nonfragile control for linear Systems with multiplicative controller gain variations. Automatica 7(5), 727–737 (2001)
11. Li, Y., Guoding, C., Maying, Y.: Robust regional pole assignment of uncertain systems via output feedback controllers. Control Theory & Applications 19(2), 244–246 (2002)
12. Li, Y., Haiqing, Y.: Robust output feedback H_∞ control for linear discrete-time systems with closed-loop disk pole constraints. Control Theory & Applications 20(5), 773–775 (2003)
13. Yuanqing, X., Jingqing, H.: Robust H_2 control with regional stability constraints. Acta Automatica Sinica 30(3), 321–329 (2004)
14. Yue, D., Lam, J.: Non-fragile guaranteed cost control for uncertain descriptor systems with time-varing state and input delay sytem. Optimal Control Applications and Methods 26(2), 85–105 (2005)
15. Garcia, G., Bernussou, J.: Pole assignment for uncertain systems in a specific disk by state feedback. IEEE Transactions on Automatic Control 40(1), 184–190 (1995)
16. Barmish, B.: Necessary and sufficient conditions for quadratic stabilizability of an uncertain system. Journal of Optimization Theory and Applications 46(4), 399–408 (1985)
17. Junfeng, W., Shanwu, L., Libo, H.: Robust H_∞ state feedback control of uncertain data-sampled system with pole constrain. Journal of Nanjing University of Aeronautice & Astronautics 38(7), 1–4 (2006)

A Dual-Band Tapered Slot Omni-directional Antenna with an Orthogonal Polygon Parasitic Element

Ying-song Li, Xiao-dong Yang, Chengyuan Liu, and Lin-lin Feng

College of Information and Communication Engineering,
Harbin Engineering University, Harbin, Heilongjiang, 150001, P.R. China
{liyingsong,yangxiaodong,liuchengyuan,jiangtao}@hrbeu.edu.cn

Abstract. A dual-band tapered slot omni-directional antenna with an orthogonal polygon parasitic element which covers GSM/DCS/PCS/CDMA2000/WCDMA /TD-SCDMA/WLAN is proposed experimentally and numerically. The proposed antenna comprises of two shunted tapered slot antennas. The tapered slot antenna is an asymmetric structure which is made up of a straight line and an exponential line. A polygon parasitic element which improves the radiation pattern is orthogonal with the radiator. The measured 14-dB impedance bandwidth of the antenna is 29.5% and 54.6% at the lower band and the higher band, respectively. And the measured 10-dB impedance bandwidth is 129.2% (0.69GHz-3.21GHz) centered at 1.95GHz, which is about 46 times that of the corresponding monopolar wire-patch antenna. The antenna is successfully simulated, designed, and measured, showing dual-band impedance bandwidth, stable gain and good omni-directional radiation patterns.

Keywords: Tapered slot antenna, wideband antenna, omni-directional antenna, indoor base station antenna, dual-band antenna.

1 Introduction

With the development of the third generation mobile communication, more and more buildings and malls require the indoor base station antenna mounted on the ceilings and districts. In fact, many monopole antennas and sleeve antennas had been used widely, while the omni-directional radiation pattern with the vertical polarization was obtained [1-3]. However, the proposed antennas have large dimensions, narrow bandwidth and complex structures.

In order to obtain a low-profile and wideband antenna which has omni-directional radiation patterns, a monopolar wire-patch antenna has been reported [1]. For this antenna, the wire monopole is top-loaded with a square patch. Two shorting wires are used to connect the patch to the ground plane. Under such configuration, the antenna height can be reduced to at the center operating frequency, as the antenna is operated at a resonance under the fundamental cavity mode. Another monopolar antenna was investigated in [3], which had wide bandwidth and monopole-like radiation pattern. But the antenna also has large size, which is too large to mount on the ceilings. In [4], a wide-band monopolar wire-patch antenna with L-probe fed is proposed for indoor

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 595–602.
springerlink.com

base station applications. However, the impedance bandwidth of these antennas are narrow. In [5-6], two sleeve antennas are investigated, the experimental results show that impedance bandwidths of about 50% are achieved. So, it can't cover GSM/DCS/PCS/ CDMA2000/WCDMA/TD-SCDMA/WLAN operation bands.

In this paper, a dual-band tapered slot omni-directional antenna with an orthogonal polygon parasitic element is presented experimentally and numerically. By properly selecting the straight line, the exponential line, the ground plane, the height of the radiator, and the orthogonal parasitic element of the slot antenna, a dual band antenna with wide impedance bandwidth, small size, good radiation characteristics suitable for the GSM, CDMA, CDMA2000, WCDMA, TD-SCDMA, DCS, WLNA (2.4GHz-2.483GHz) base station applications could be obtained. Details of the antenna design and both the numerical and experimental results are presented and discussed.

2 Antenna Design

The basic geometry of the proposed antenna consists of two tapered slot antennas which are formed by combining the straight line and the exponential line, orthogonal parasitic element, the ground plane. The tapered slot was supported by FR4 with relative permittivity 4.4 and thickness 1.6mm. The structure of the antenna is shown in Fig.1. The straight line and the exponential line can be described by equation (1) and equation (2), respectively.

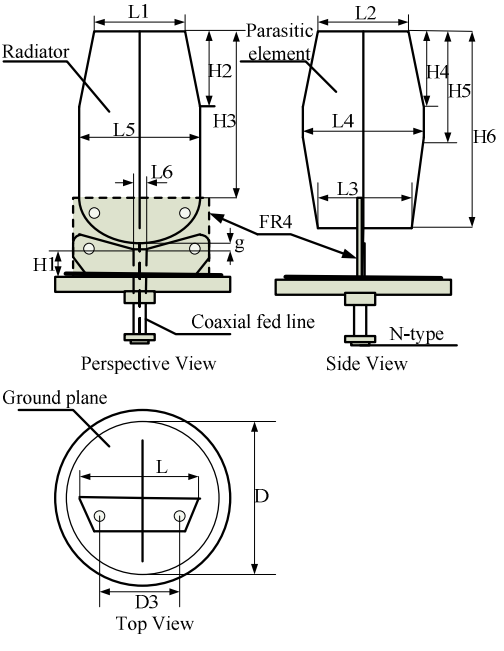


Fig. 1. Geometry of the proposed antenna

$$z = k(x - A) - B \tag{1}$$

$$z = Ce^{p(x-D)} + E \tag{2}$$

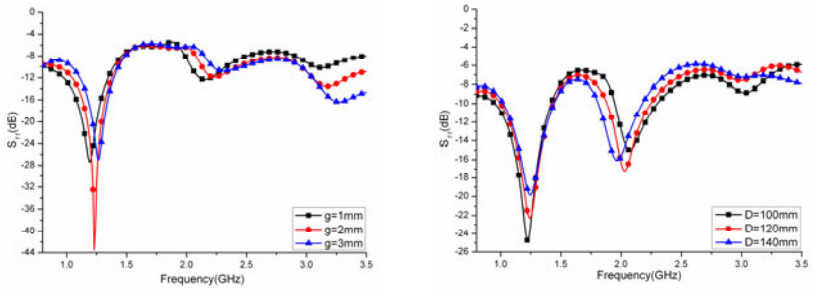
Where, x is the horizontal ordinate and z is the height of the antenna. In this design, k, A, B, C, p, D, E are constant. And E is zero. From the Fig.1, the proposed antenna with an orthogonal polygon parasitic element is fixed on a ground with diameter D like a monopole antenna. And the two tapered slot antennas which are modified Vivaldi antennas are shunted to obtain good omni-directional radiation patterns. The orthogonal polygon parasitic element utilized in the paper is similar to the radiator, only the exponential line was replaced by a tapered structure. The parasitic element is also orthogonal with the radiator to improve the radiation characteristics effectively [7]. However, very different from the conventional tapered slot antenna, the proposed antenna combined two tapered slot antenna which is shunted to form an omni-directional antenna. The asymmetrical structure also gives a good omni-directional radiation pattern which can well meet the wireless and mobile communication applications. For achieving the dual-band and wide-band operation, the dimensional parameters of the proposed antenna, were all first iteratively approached from the High Frequency Structure Simulator (HFSS) and then adjusted from experiment. Finally, the optimal antenna dimensions were obtained and shown in Table 1.

Table 1. Dimension of the proposed antenna(in mm)

Paramters	Value
L	56
L1	45
L2	43
L3	43
L4	56
L5	56
L6	9
g	1.8
H1	10
H2	44
H3	85
H4	32
H5	48
H6	94
D	120
D3	39

3 Parameters Study

As the proposed idea above, the key parameters were obtained by using HFSS. Due to the diameter of the ground D and the distance of the slot g play an important role in the impedance of the antenna. So, the two parameters are selected in the parametric study. In order to obtain accurate influence of the two parameters on its impedance



(a) (b)

Fig. 2. Effect on the reflection coefficient of the key parameters vs. frequency

bandwidth, only one parameter is changed at each time while other parameters are kept constant which is listed in table 1.

The effects of the two key parameters mentioned above on the return loss vs. frequency are plotted in Fig.2 (a) and (b), respectively. From the Fig.2 (a), it can be seen that the resonant frequency of the both band is moving to the higher frequency with the distance between the straight line and the exponential line g increased, whereas the resonant frequency will reduce by decreasing the distance g. The distance plays an important role in improving the impedance bandwidth and the resonant frequency. Fig. 2(b) gives that the effects of the diameter of the ground D. The resonant frequency at the lower band has little change, but the resonant frequency at the higher band changes rapidly. In order to broaden the impedance bandwidth of the antenna, the radiator and the orthogonal parasitic element are also cut as a tapered structure. The proposed antennas with orthogonal polygon parasitic element and without orthogonal polygon parasitic element are realized by HFSS. The effect on the impedance bandwidth of the orthogonal polygon parasitic element is shown in Fig.3. From Fig.3, the orthogonal parasitic element has little effect on the impedance. But the orthogonal parasitic element can improve the omni-directional radiation pattern which is realized in [7]. So, the antenna can be optimized by adjusting the distance g, the ground plane D and the tapered slot antenna structure.

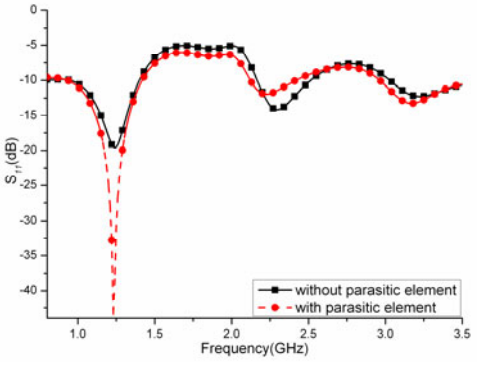
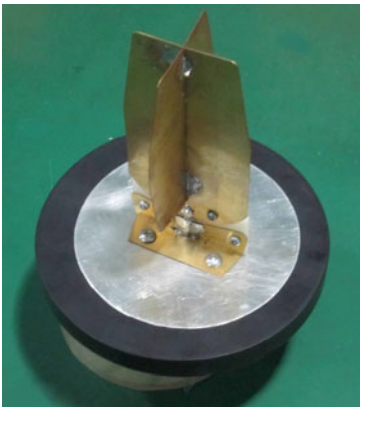


Fig. 3. reflection coefficient of the proposed antenna with and without parasitic

4 Results and Discussions

The proposed antenna is optimized by HFSS and the optimized antenna is manufactured. The prototype of the proposed dual-band tapered slot antenna with an orthogonal parasitic element is shown in Fig.4. To evaluate the performance of the optimized antenna, the proposed antenna was tested. The detailed dimensions are listed in table 1. The size of proposed antenna is reduced by 33.3%, which is smaller than the previous structure in [1-3] and more suitable for the indoor base station application installed on ceilings and suburban area. In this section, the simulated and measured return loss, Gain and the measured radiation pattern of the proposed antenna are presented and discussed. The simulations are given by using HFSS, and the measurement is achieved by using HP8757D network analyzer and an anechoic chamber. Fig.5 shows the simulated and measured return losses of the antenna. It can be seen from the Fig.5, the measured result is well agree with the simulated one with an acceptable discrepancy. The differences between the simulated and measured values may be due to the errors of the manufactured antenna and the N-type connector to the feeding probe, which is included in the measurements but not taken into account in the calculated results. For return loss less than -14dB, the measured dual-band impedance width are about 29.5% (0.75GHz-1.01GHz) centered at 0.88GHz and 54.6% (1.57GHz- 2.75GHz) centered at 2.16GHz. And the measured 10-dB impedance bandwidth is 129.2% (0.69GHz-3.21GHz) centered at 1.93GHz, which is about 45 times that of the corresponding monopolar wire-patch antenna [2]. The measured result also meets the GSM, CDMA, CDMA2000, WCDMA, TD-SCDMA, DCS, WLAN (2.4GHz-2.483GHz) applications.



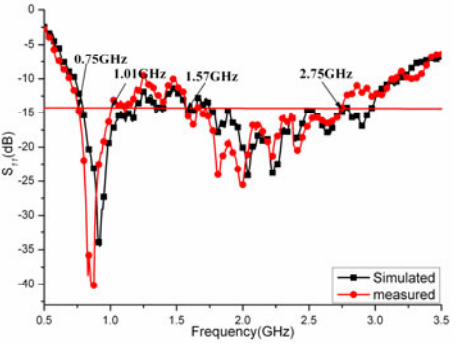


Fig. 4. Prototype of the proposed antenna; **Fig. 5.** Reflection coefficient of the antenna

Fig.6 gives the radiation pattern of the proposed antenna at 0.824GHz, 0.96GHz, 1.71GHz, 2.17GHz, 2.4GHz, respectively. It can be seen from the Fig. 6, the antenna has good omni-directional radiation pattern among the impedance bandwidth in the H-plane. In the E-plane, the radiation pattern of the antenna has monopole-like, which is bidirectional radiation pattern. But, the radiation pattern at the higher band has little change which may caused by the feed cable placed in the near field of the antenna and

the asymmetric tapered slot structure. Fig.7 shows the simulated and measured gains at the major point. In the lower band, the gain is more than 2dBi which is similar to the monopole antenna. In the higher band, the gain is more than 5dBi, which is raised by 2dBi compared to the antenna in [1-3]. It can be well fulfill the indoor base station applications.

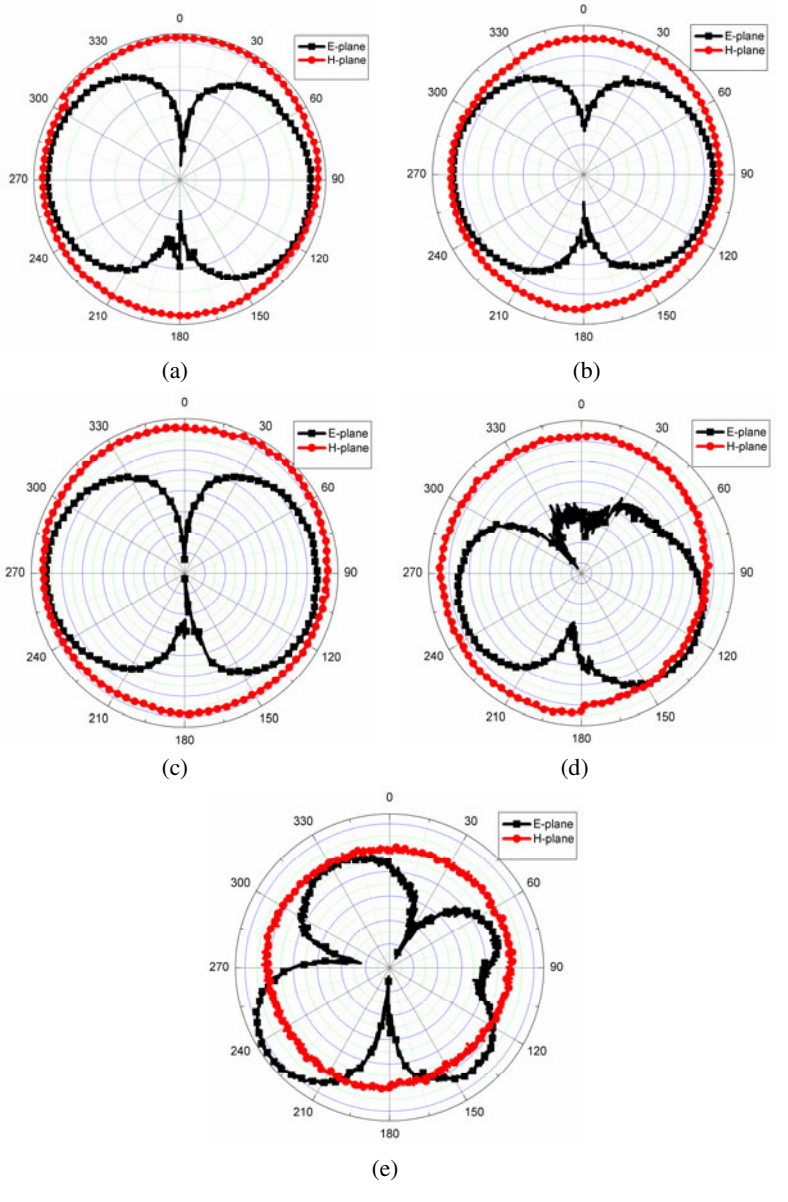


Fig. 6. The radiation patterns of the antenna.
(a)0.824GHz; (b)0.96GHz; (c) 1.71GHz; (d)2.17GHz; (e)2.4GHz

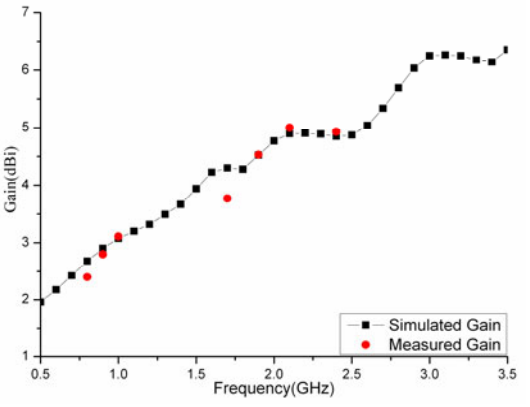


Fig. 7. Gain of the antenna

5 Conclusions

In the paper, a dual-band tapered slot omni-directional antenna consisted by two shunted tapered slot antenna with an orthogonal parasitic element is realized experimentally and numerically. The size of the antenna is reduced about 33.3% and the gain is 2dBi higher than the previous similar antenna. From the experimental results, return loss less than 14-dB, the measured dual-band impedance bandwidth is about 29.5% at lower band and 54.6% at higher band. And the measured 10-dB impedance bandwidth is 129.2% at the center frequency. Owing to the results, the antenna can be applied for indoor base station communication applications.

Acknowledgments

The authors are grateful to Mr. Wang Jixiang for his kindly help and valuable suggestion. The work is also partially supported by Nature Science Fund of China (No.60902014), China. This paper is funded by the International Exchange Program of Harbin Engineering University for Innovation-oriented Talents Cultivation.

References

1. Delaveaud, C., Leveque, P., Jecko, B.: New kind of microstrip antenna:The monopolar wire-patch antenna. Electron. Lett. 30(1), 1–2 (1994)
2. Row, J.S., Yeh, S.H., Wong, K.L.: A wide-band monopolar plate-patch antenna. IEEE Trans. Antennas Propag. 50(9), 1328–1330 (2002)
3. Lau, K.L., Li, P., Luk, K.-M.: A Monopolar Patch Antenna With Very Wide Impedance Bandwidth. IEEE Trans. Antennas Propag. 53(2), 655–661 (2005)
4. Lau, K.L., Luk, K.M.: A wide-band monopolar wire-patch antenna for indoor base station applications. IEEE Antennas Wirel. Propag. Lett. 4, 155–157 (2005)

5. Thomas, K.G., Lenin, N., Sreenivasan, M.: A monopolar patch antenna with very wide impedance bandwidth. IEEE Trans. Antennas Propag. 54, 1034–1037 (2006)
6. Zhang, Z.Y., Gong, F., Gong, S.X., Zuo, S.L., Lu, Q.Y.: Sleeve monopole antenna for DVB-H applications. Electron. Lett. 46, 112–113 (2010)
7. Ammann, M.J., Cordoba, R.S., Uzelac, M., Evans, J.A., Schwarzbacher, A.T.: On pattern stability of the crossed planar monopole. Microw. Opt. Technol. Lett. 40, 294–296 (2004)

Local Approximation in Manifold Learning

Huixia Jin[1], Jun Tang[2], and Gelan Yang[3]

[1] Department of Physics and Telecom Engineering, Hunan City University, Yiyang, China
[2] Department of Information Engineering, Hunan Urban Construction College, Xiangtan, China
[3] Department of Computer Science, Hunan City University, Yiyang, China
Jinhuixia1980@163.com

Abstract. Both manifold learning and semi-supervised learning have been widely investigated in the past few years. Some of manifold learning algorithms, which are based on the idea of local approximation, can be used to control the way of transmitting information between point clouds. We combine local approximation with the idea of preserving projections and weighted integration, and give a set of solutions to semi-supervised regression and manifold alignment. Finally, we validated the effectiveness of the presented schemes in the experiments.

Keywords: semi-supervised learning, manifold learning, classification, manifold alignment.

1 Introduction

Many high-dimensional data in real-world applications can be modeled as data points lying close to a low-dimensional nonlinear manifold. Manifold learning algorithms aim at recovering the embedded low-dimensional manifold to study the primary property of the high-dimensional data. Most of them are based on preserving some metric info of the sample space. For example, Isometric Mapping(ISOMAP) [1] holds global geodesic distance; Diffusion Maps [2] preserves a kind of global metric defined by gauss kernel function; Hessian Eigenmaps Locally Linear Embedding(HLLE) [3] maintains that the low-dimensional data representation is locally isometric. Some algorithms are based on the idea of local approximation, such as Locally Linear Embedding(LLE)[4], Laplacian Eigenmaps (LE)[5] and Local Tangent Space Alignment(LTSA)[6]. Some others are modified from local approximation algorithms, and preserve projections in the transformation of high-dimensional data to low-dimensional data. For example, Linearity Preserving Projection(LPP)[7], Neighborhood Preserving Embedding(NPE) [8], and Lineal Local Tangent Space Alignment (LLTSA)[9] preserve a global linear projection; Locally Linear Coordination(LLC) [10~14] preserves local linear projections between every local dimension reduction coordinates and final coordinates. We combine local approximation with the idea of preserving projections and weighted integration, and give a set of solutions to semi-supervised regression and manifold alignment. Here, classification is thought of as a special case of regression.

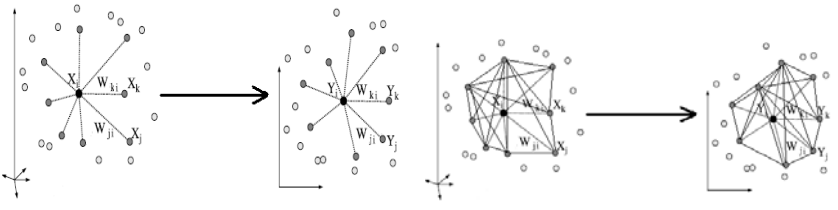
J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 603–610.
springerlink.com

The rest of this paper is organized as follows: Section 2 introduces the local approximation. Following in section 3 applications of Local approximation is developed for local approximation algorithms. We give an analysis and experiments of SLTSA and others on classification in Section 4. Finally, the conclusions are given in Section 5.

2 Local Approximation

Suppose we have a $D \times n$ matrix X consisting of n datavectors x_i with dimensionality D, we need to transform it into a new dataset Y with dimensionality d, while preserving the main property of X.

Local approximation is extracting some local approximation property from X to organize Y. There are two types of approximation, approximation of a point, and approximation of a block, and the former is a special case of the other one.



(a) (b)

Fig. 1. Local approximation (a) Local approximation for point (b) Local approximation for block

Let Γ_i be a vector of indices of points in the $(k-1)$-neighbor of x_i, $\overline{\Gamma}_i = \begin{bmatrix} i \\ \Gamma_i \end{bmatrix}$. I is the identify matrix, e is a vector of all 1's, $J = (I - \frac{1}{k} ee^T)$ is a mean removal operator, $\overline{X}_{\overline{\Gamma}_i} = X_{\overline{\Gamma}_i} J$ is the local coordinate of the high-dimensional data and its counterpart is $\overline{Y}_{\overline{\Gamma}_i} = Y_{\overline{\Gamma}_i} J$ .Given coefficient $w_{\Gamma_i,i}$, the approximation to a point is $\overline{Y}_{\Gamma_i} w_{\Gamma_i,i} \to \overline{Y}_i$.

2.1 Point Approximation

Let Γ_i be a vector of indices of points in the $(k-1)$-neighbor of x_i, and $\overline{\Gamma}_i = \begin{bmatrix} i \\ \Gamma_i \end{bmatrix}$ is a vector including i and Γ_i. S_i is a 0-1 selection matrix satisfying $XS_i = X_{\overline{\Gamma}_i}$, similarly in the low-dimensional space, we have $\overline{Y}S_i = Y_{\overline{\Gamma}_i}$. Let e be the vector of all 1's, I_k be the identify matrix with rank k. Then, $J = (I - \frac{1}{k} ee^T)$ is a mean removal operator.

We can get the local coordinate of the high-dimensional data $\overline{X}_{\overline{\Gamma}_i} = X_{\overline{\Gamma}_i} J$, and its counterpart in low-dimension $\overline{Y}_{\overline{\Gamma}_i} = Y_{\overline{\Gamma}_i} J$. The approximation of a point is defined as $\overline{Y}_{\Gamma_i} w_i \to \overline{Y}_i$, where w_i is the local approximation vector, which is extracted from X.

The point approximation error of y_i is defined as

$$err_{pi} = \left\| \overline{Y}_i - \overline{Y}_{\Gamma_i} w_i \right\|^2 = \left\| YS_i J \begin{bmatrix} 1 \\ -w_i \end{bmatrix} \right\|^2 \tag{1}$$

by summing the approximation error of each point, the total approximation error of points is

$$err_p = \sum_{i=1}^{n} err_{pi} = \sum_{i=1}^{n} \left\| YS_i J \begin{bmatrix} 1 \\ -w_i \end{bmatrix} \right\|^2 = \left\| YB_p \right\|_F^2 \tag{2}$$

where B_p is a sparse matrix satisfying $B_P(\overline{\Gamma}_i \quad i) = J \begin{bmatrix} 1 \\ -w_i \end{bmatrix}$ and $\|\bullet\|_F$ stands for the Frobenius norm of a matrix.

2.2 Block Approximation

Similar to the definition of point approximation, the block approximation is defined as $\overline{Y}_{\overline{\Gamma}_i} W_i \to \overline{Y}_{\overline{\Gamma}_i}$, where W_i is the local approximation matrix around y_i.

The block approximation error around y_i is defined as

$$err_{bi} = \left\| \overline{Y}_{\overline{\Gamma}_i} - \overline{Y}_{\overline{\Gamma}_i} W_i \right\|_F^2 = \left\| YS_i J \left(I_k - W_i \right) \right\|_F^2 \tag{3}$$

The summation of the approximation error of all local blocks is

$$err_b = \sum_{i=1}^{n} err_{bi} = \sum_{i=1}^{n} \left\| YS_i J \left(I_k - W_i \right) \right\|_F^2 = \left\| YS_b B_b \right\|_F^2 \tag{4}$$

where $S_b = \left[S_1, \cdots, S_n \right]$, $B_b = diag\left\{ J\left(I_k - W_1 \right), \cdots, J\left(I_k - W_n \right) \right\}$.

3 Applications of Local Approximation

Learning with the label info can be regarded[8] as the problem of approximating a multivariate function from labeled data points. The function can be real valued as in regression or binary valued as in classification. Learning with the label info can also be regarded as a special case of dimension reduction that maps all the data points in the label space. The label error of y_i is defined as $err_{li} = s_i \left\| y_i - f_i \right\|^2$, where $F = \left[f_1, \cdots, f_n \right]$ is the label value, s_i is the flag to identify the labeled points satisfying

$s_i = \begin{cases} 1 & \cdots \quad i \in L \\ 0 & \cdots \quad i \notin L \end{cases}$, and L is the collection of index of labeled points. By weighted combining the point approximation error and label error, we can get

$$Err_p = \sum_{i=1}^{n} \left((1-a_i)^2 err_{pi} + a_i{}^2 err_{li} \right) = \left\| YB_p (I_n - A) \right\|_F^2 + \left\| (Y-F)A \right\|_F^2 \tag{5}$$

optimal

$$Y^* = FAA^T \left(M_p + AA^T \right)^{-1} \tag{6}$$

Here, $M_p = B_p (I_n - A)(I_n - A)^T B_p^T$, $a_i = (\frac{l}{n}(1-a^0) + a^0) s_i$ is the weight coefficient at y_i, l is the number of labeled points, a^0 is the minimal weight coefficient set by user, and $A = diag(a_i)$. We set the weight coefficient on the following intuition that: if the proportion of the labeled points is very small, we have to reduce our dependence on the knowledge only retained from the labeled points; if all the points are labeled, we must totally discard the geometry info of the point clouds, for at that moment the label info is more reliable. As a result, the coefficient a_i has to be adjusted with the proportion of labeled points.

Similarly with the point approximation, the total error defined for block approximation is

$$Err_b = \sum_{i=1}^{n} \left((1-a_i)^2 err_{bi} + a_i{}^2 err_{li} \right) = \left\| YS_b B_b (I_K - A_b) \right\|_F^2 + \left\| (Y-F)A \right\|_F^2 \tag{7}$$

optimal

$$Y^* = FAA^T \left(M_b + AA^T \right)^{-1} \tag{8}$$

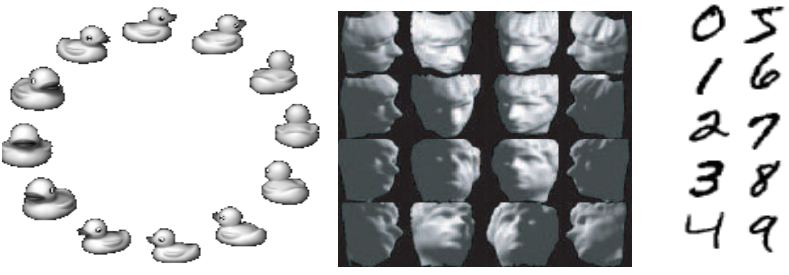
Here, $M_b = S_b B_b (I_k - A_b)(I_k - A_b)^T B_b^T S_b^T$, $K = k \times n$, $A_b = diag\{a_1 I_k, \cdots, a_n I_k\}$ is a sparse weight matrix. We take $y^0 = \frac{1}{n-l} \sum_{i \notin L} y_i$ as the decision threshold for classification.

4 Experiment and Discussion

In this section, we investigate the performance of our proposed Local approximation (SLP)method for face representation and recognition. The system performance is compared with the Eigenface method Neighborhood preserving embedding (NPE), the Linear local tangent space alignment (LLTSA), and the Laplacianface method (LPP), three of the most popular linear methods in face recognition. We use the same graph structures in the Laplacianface and Local approximation (LP)method, which is built based on the label information.

Three experiments were conducted to evaluate the performance of the SLP algorithms. The nearest neighbor algorithm was used to evaluate the recognition

technique. Each dataset was truncated into two subsets, one as a training set and the other as a learning set. The classification rates are means of all tests on each dataset. It should be noted that, since the focus in this paper is on feature representation, all of our experiments use a very simple classifier, i.e. nearest-neighbor classifier. In this experiment, the Duck dataset are randomly partitioned into two subsets, namely. Samples are shown in Fig. 1, which shows our experiment datasets. Duck dataset is extracted from COIL20. We use duck and face datasets and MNIST dataset for classification.



(a) (b) (c)

Fig. 2. Datasets for experiment (a) DUCK dataset (b) FACE dataset (c) MNIST dataset

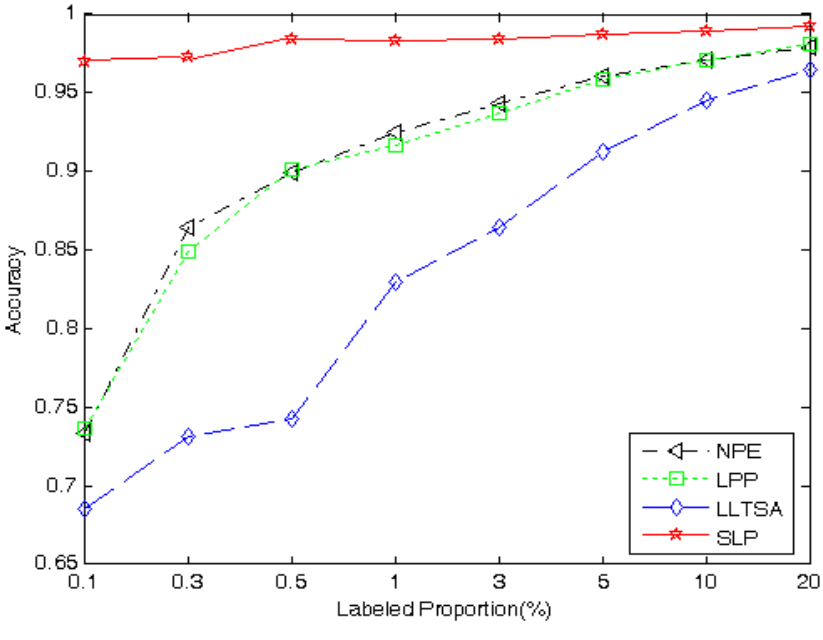


Fig. 3. DUCK recognition accuracy

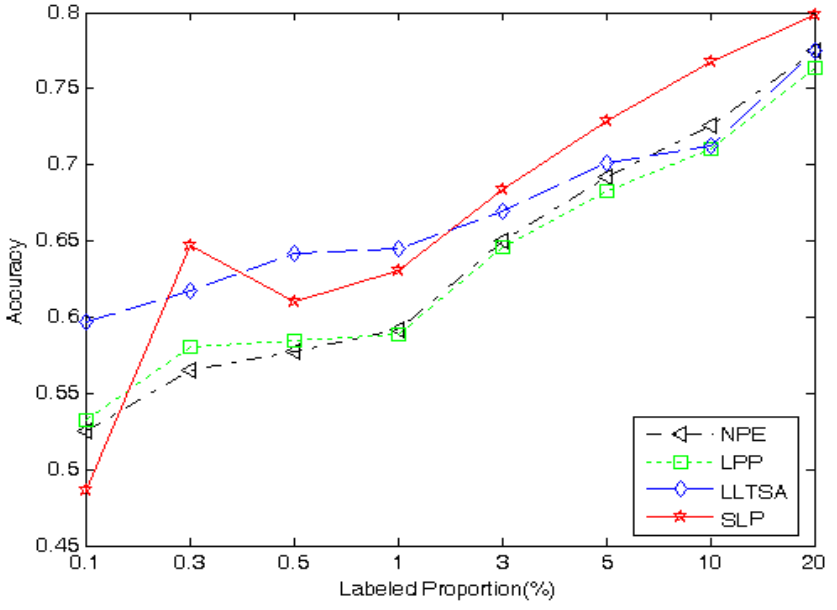


Fig. 4. Facial recognition accuracy

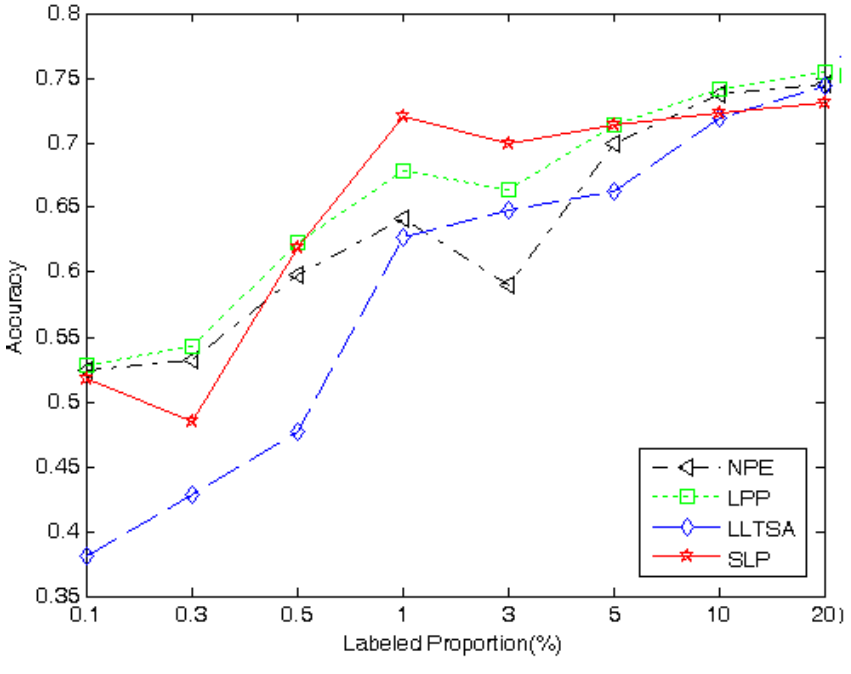


Fig. 5. MNIST recognition accuracy

In the experiments, SLP (Semi-supervised local coordinate) is compared with other three inductive methods: LPP, NPE and LLTSA. Here, LPP is performed in semi-supervised manner. NPE and LLTSA are semi-supervised extended with prior information. The final results on all data sets are separately listed in figure 3, 4 and 5, and the best performance is shown in bold. The result in figure 3 shows that SLC outperforms other methods on DUCK in all cases. figure 4 demonstrates the superiority of SLC and LPP on FACE. In figure 5, the classification accuracy reveals the benefits of both SLC and LLTSA. From these results, we can see that, SLP is a relatively robust classification algorithm, and achieves good performance on all the data sets.

5 Conclusion

An local approximation (SLC) algorithm is presented in this paper. It combines local approximation with the idea of preserving projections and weighted integration, and give a set of solutions to semi-supervised regression and manifold alignment for classification. This algorithm presented greatly reduces the number of labeled data the classifier system needs in order to achieve satisfactory performance. Experiments on standard database show that our algorithm performs better than current classifier combination rules when considering both labeling cost and classification accuracy..

Acknowledgment

This work was supported by Scientific Research Fund of Hunan Provincial Education Department under Grant (No.10C0498). Project supported by the Research Foundation of Science & Technology office of Hunan Province under Grant(No. 2010FJ4213).

References

1. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by Locally Linear Embedding. Science 290(5500), 2323–2326 (2000)
2. Belkin, M., Niyogi, P.: Laplacian Eigenmaps and spectral techniques for embedding and clustering. Neural Computation 15(6), 1373–1396 (2003)
3. Zhang, Z.Y., Zha, H.Y.: Principal Manifolds and Nonlinear Dimension Reduction via Local Tangent Space Alignment. CSE-02-019, Technical Report, CSE, Penn State Univ. (2002)
4. He, X., Niyogi, P.: Locality preserving projections. In: Advances in Neural Information Processing Systems, vol. 16, p. 37. The MIT Press, Cambridge (2006)
5. He, X., Cai, D., Yan, S., Zhang, H.J.: Neighborhood preserving embedding. In: Proc. Int. Conf. Computer Vision (2005)
6. Zhang, T., Yang, J., Zhao, D., Ge, X.: Linear local tangent space alignment and application to face recognition. Neurocomputing 70, 1547–1553 (2007)
7. Teh, Y.W., Roweis, S.T.: Automatic alignment of hidden representations. In: Advances in Neural Information Processing Systems, vol. 15, pp. 841–848. The MIT Press, Cambridge (2002)

8. Tenenbaum, J.B., Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science 290(5500), 2319–2323 (2000)
9. Lafon, S., Lee, A.B.: Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(9), 1393–1403 (2006)
10. Donoho, D.L., Grimes, C.: Hessian Eigenmaps: new locally linear embedding techniques for high-dimensional data. Technical Report TR-2003-08, Department of Statistics, Stanford University (2003)
11. Hoerl, A.E., Kennard, R.W.: Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 12(1), 55–67 (1970)
12. Pothen, A., Fan, C.J.: Computing the Block Triangular Form of a Sparse Matrix. ACM Transactions on Mathematical Software 16(4), 303–324 (1990)
13. Gelan, Y., Xue, X., Li, T.: Manifold alignment via local block coordinate. In: 2nd International Workshop on Knowledge Discovery and Data Mining, WKKD 2009, pp. 697–700 (2009)
14. Gelan, Y., Xue, X., Huixia, J.: Semi-supervised regression via local block coordinate. In: 2nd International Congress on International Congress on Image and Signal Processing, pp. 232–236 (2009)

Semi-supervised Local Tangent Space Alignment

Liyuan Jia[1], Li Huang[2], and Lei Li[3]

[1] Department of Computer Science, Hunan City University, Yiyang, China
[2] Department of science and technology, Hunan City University, Yiyang, China
[3] Department of computer and information technology, Henan Xinyang Normal College, Xinyang, China
jia_4211003@126.com

Abstract. Conventional semi-supervised learning leverages the unlabeled data by intensive exploring the pairwise relation among the data points. However, it is well known that such relation cannot capture the complex interaction in many real-world applications. To address this problem, we proposed in this paper a new approach to effectively modeling the labeled and unlabeled data by local tangent space alignment, which is superior due to the properties of invariant to shift and scale. We apply the local tangent space alignment to semi-supervised learning tasks including semi-supervised classification. The experiments compared with state-of-the-art semi-supervised learning methods demonstrated the effectiveness of the proposed approach.

Keywords: semi-supervised learning, manifold learning, local tangent space alignment.

1 Introduction

Recently, technology for dimension reduction has attracted much attention in pattern recognition, usually raw data taken with capturing devices are multidimensional and therefore are not very suitable for accurate classification. To obtain compact representations of raw data, some techniques about dimension reduction have come forth. From the geometrical point of view, dimension reduction can be considered as discovering a low-dimensional embedding of high-dimensional data assumed to lie on a manifold. The key of dimension reduction is to preserve the underlying local geometrical information of raw high-dimensional data while reducing insignificant dimensions. However, if the original data lie on a nonlinear manifold in nature, traditional dimension reduction methods such as Principal Component Analysis (PCA) will fail to well preserve its geometrical information in a low-dimensional space while unfolding the nonlinear manifold. That is, in the case of nonlinear manifolds, PCA often maps close points in the original space into distant points in the embedded space. In the recent years, a number of techniques have been proposed to perform nonlinear mappings, such as MDS [1], locally linear embedding (LLE) [2],

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 611–618.
springerlink.com

Laplacian Eigenmaps (LE) [3], manifold charting [4], Hessian-based locally linear embedding (HLLE) [3], Modified locally linear embedding (MLLE) [6~9], Laplacian Eigenmap (LE) [6] and ISOMAP [5]. One basic idea of local methods is to regard a small neighborhood of the manifold as a linear one and find a local geometry around each data point, and then use the collected local geometric information to nonlinearly map the manifold to a lower dimensional space. All of these are problematic in application in some way: firstly, multi-dimensional scaling and neural networks are hard to train and time-consuming. Mixtures of localized linear models require the user to set a number of parameters, which are highly specific to each data set and determine how well the model fits the data. Secondly, these nonlinear methods aim to preserve local structures in small neighborhoods and successfully derive the intrinsic features of nonlinear manifolds. Recently, Zhang and Zha [10] proposed a fine method: local tangent space alignment (LTSA). The research of nonlinear dimensionality reduction is based on that many high-dimensional data is confirmed by several hidden variables, such as the effect of face image sampling is determined by brightness, the distance between person and camera, head pose, facial muscles and so on. From the perspective of cognitive psychology, Psychologists think that cognitive processes are based on topological continuity. Thus nonlinear method is feasible. This paper introduces the principle of locally linear embedding, applies it to face recognition, and analyses the recognition rate. Comprehensive comparisons and extensive experiments show that the approach achieves much higher recognition rates than a few competing methods.

Face recognition and fingerprint identification technology have been applied widely. Because of different light and expression, every person has many different face images. However, gaining effective and reasonable low dimension face image is much more difficult from high dimension face by keeping whole face information. It is a imperative solved problem. Consequently, potential framework of high dimensional data should be discovered through studying the low dimensional character embedding the high dimensional space, then we can recognize face efficient.

The rest of the paper is organized as follows. We outline the basic steps of LTSA and illustrate its failure modes using two examples in Section 2. The SLTSA will be proposed in Section 3. We give an analysis and experiments of SLTSA and others on classification in Section 4. Finally, the conclusions are given in Section 5.

2 Local Tangent Space Alignment

We first outline the basic steps of LTSA. The basic idea of LTSA is to construct local linear approximations of the manifold in the form of a collection of overlapping approximate tangent spaces at each sample point, and then align those tangent spaces to obtain a global parametrization of the manifold. Details and derivation of the algorithm can be found in [10]. Given a data set $X_i = [x_1,\ldots,x_2]$ with $x_i \in R^m$, sampled

(possibly with noise) from a d-dimensional manifold ($d << m$), $x_i = f(\tau_i) + \varepsilon_i$, where $f : \Omega \subset R^d \to R^m$, Ω is an open connected subset, and ε_i represents noise. LTSA assumes that d is known and proceeds in the following steps.

(1) Local neighborhood construction. For each x_i , $i = 1,\ldots,N$, determine a set $X_i = \lfloor x_{i_1},\ldots,x_{i_k} \rfloor$ of its neighbors (k nearest neighbors, for example).

(2) Local linear fitting. Compute the optimal rank-d approximation to the centered matrix $(X_i - \bar{x}_i e^T)$, where $\bar{x}_i = \frac{1}{k}\sum_{j=1}^{k} x_{i_j}$, and e is a k-dimensional row vector of all 1's. By the SVD of $X_i - \bar{x}_i e^T$, we can obtain the orthonormal basis Q_i for the d-dimensional tangent space of the manifold at x_i, and the orthogonal projection of each x_{i_j} in its neighborhood to the computed tangent space $\theta_j^{(i)} = Q_i^T (x_{i_j} - \bar{x}_i)$.

(3) Local coordinates alignment. Align the N local projection $\Theta_i = \left[\theta_1^{(i)},\ldots,\theta_k^{(i)}\right]$, $i = 1,\ldots,N$, to obtain the global coordinates Denote $\tau_1,\ldots,\tau_N$, and $T_i = \lfloor \tau_{i_1},\ldots,\tau_{i_k} \rfloor$ which consists of a subset of the columns of T with the index set $\{i_1,\ldots,i_k\}$ determined by the neighbors of each x_i. Let $E_i = T_i - c_i e^T - L_i\Theta_i$ be the local reconstruction error matrix, where $c_i = \frac{1}{k} T_i e$ and $L_i = T_i(I - \frac{1}{k} ee^T)\Theta_i^+ = T_i\Theta_i^+$, where Θ_i^+ is the Moor-Penrose generalized inverse of and e is a vector of all ones. Then the alignment of LTSA is achieved by minimizing the following global reconstruction error:

$$E(T) = \sum_i \|E_i\|^2 \equiv \sum_i \min_{c_i, L_i} \left\|T_i - c_i e^T - L_i\Theta_i\right\|^2 = \|TSW\|^2 . \tag{1}$$

Where $S = [S_1,\ldots,S_N]$ and $W = diag(W_1,\ldots,W_N)$, with

$$W_i = (I - \frac{1}{k} ee^T)(I - \Theta_i^+\Theta^+) \tag{2}$$

To uniquely determine T, we will impose the constraints $TT^T = I_d$, it turns out that the vector e of all ones is an eigenvector of B corresponding to a zero eigenvalue,

$$B = SWW^T S^T \tag{3}$$

Therefore, the optimal T is given by the d eigenvectors of the matrix B, corresponding to the 2nd to $d+1$ st smallest eigenvalues of B

From the basic steps of LTSA, one can see that recovering the real local tangent space is the key issue that relates to whether LTSA can discover the true manifold structure faithfully. In the presence of noise, however, the recovered tangent space by standard SVD technique will deviate from the real one due to its sensitivity to noise, then it will further influence embedding result of LTSA.

When computing local alignment error, LTSA gives each point the same weight in equation (1). Obviously, in noise case, we should make distinction between clean points and noise points, and this can be implemented by specifying different weights to them. In addition, there is no need to minimize the local align error sum of all neighborhoods in equation (3). On the one hand, in no noise case localtangent space coordinates derived from each neighborhood can character local geometric well, they are heavily redundant. On the other hand, in noise case, forcedly aligning the local coordinates of neighborhoods that are not well approximated will result in fatal error in the final embedding. Therefore, one should discard the neighborhoods whose local coordinates can't character the local geometric well due to dominant effect of noise, and select neighborhoods that are approximated well by the local tangent space coordinates, then minimize the alignment error sum of these selected neighborhoods in equation (3).

3 Semi Supervised Local Tangent Space Alignment

LTSA offers us a method to determine the approximation matrix W_i for the block mode. Here θ_i is the mapping of $\overline{X}_{\overline{\Gamma}_i}$ in the local tangent space, θ_i^+ is the Moor-Penrose generalized inverse of θ_i, and $W_i = \theta_i^+ \theta_i$ acts like a correlation matrix of the points around x_i.

Learning with the label value can be regarded[8] as the problem of approximating a multivariate function from labeled data points. The function can be real valued as in regression or binary valued as in classification. Learning with the label value can also be regarded as a special case of dimension reduction that maps all the data points into the label value space. The label error of y_i is defined as $err_{li} = s_i \| y_i - f_i \|^2$, where s_i is the flag to identify the labeled points satisfying $s_i = \begin{cases} 1 & \cdots \quad i \in L \\ 0 & \cdots \quad i \notin L \end{cases}$, L is the collection of indices of labeled points, and $F = [f_1, \cdots, f_n]$ is the given label value. The loss function $Err_p(Y)$ defined on weighted combination of point approximation error and label error and its optimal solution Y^* are shown in (3).

$$Err_p = \sum_{i=1}^{n} \left((1-a_i)^2 err_{pi} + a_i{}^2 err_{li} \right) = \left\| Y B_p (I_n - A) \right\|_F^2 + \left\| (Y - F) A \right\|_F^2 \tag{4}$$

Then

$$Y^* = FAA^T \left(M_p + AA^T \right)^{-1} \tag{5}$$

where $M_p = B_p (I_n - A)(I_n - A)^T B_p{}^T$, $a_i = (\frac{l}{n}(1-a^0) + a^0) s_i$ is the weight coefficient at y_i, l is the number of labeled points, a^0 is the minimal weight coefficient set by user,

and $A = diag(a_i)$. Here, the setting of the weight coefficient a_i is based on the following two assumptions that: if the proportion of the labeled points decreases, we have to reduce our dependence on the knowledge only retained from the labeled points; if all the points are labeled, we must totally discard the geometric knowledge of the point clouds, for the label information is more reliable and the geometric knowledge is completely useless at that moment.

Similarly, the total error defined for block approximation and its optimal solution are

$$Err_b = \sum_{i=1}^{n} \left((1-a_i)^2 err_{bi} + a_i{}^2 err_{li} \right) = \left\| YS_b B_b \left(I_K - A_b \right) \right\|_F^2 + \left\| (Y-F)A \right\|_F^2 \tag{6}$$

Optimal

$$Y^* = FAA^T \left(M_b + AA^T \right)^{-1} \tag{7}$$

Here,

$$M_b = S_b B_b \left(I_K - A_b \right) \left(I_K - A_b \right)^T B_b{}^T S_b{}^T \tag{8}$$

$K = k \times n$, $A_b = diag\{a_1 I_k, \cdots, a_n I_k\}$is a sparse weight matrix. We take $y^0 = \frac{1}{n-l} \sum_{i \notin L} y_i$ as the decision threshold for classification.

4 Experiment

Two experiments were conducted to evaluate the performance of the SLTSA algorithms. The nearest neighbor algorithm was used to evaluate the recognition technique. Each dataset was truncated into two subsets, one as a training set and the other as a learning set. The classification rates are means of all tests on each dataset. It should be noted that, since the focus in this paper is on feature representation, all of our experiments use a very simple classifier, i.e. nearest-neighbor classifier.

In this experiment, the ORL data sets are randomly partitioned into two subsets, namely, 200 training images and 200 test images, with no overlap. Samples are shown in Fig. 1.

Fig. 1. Some original ORL faces

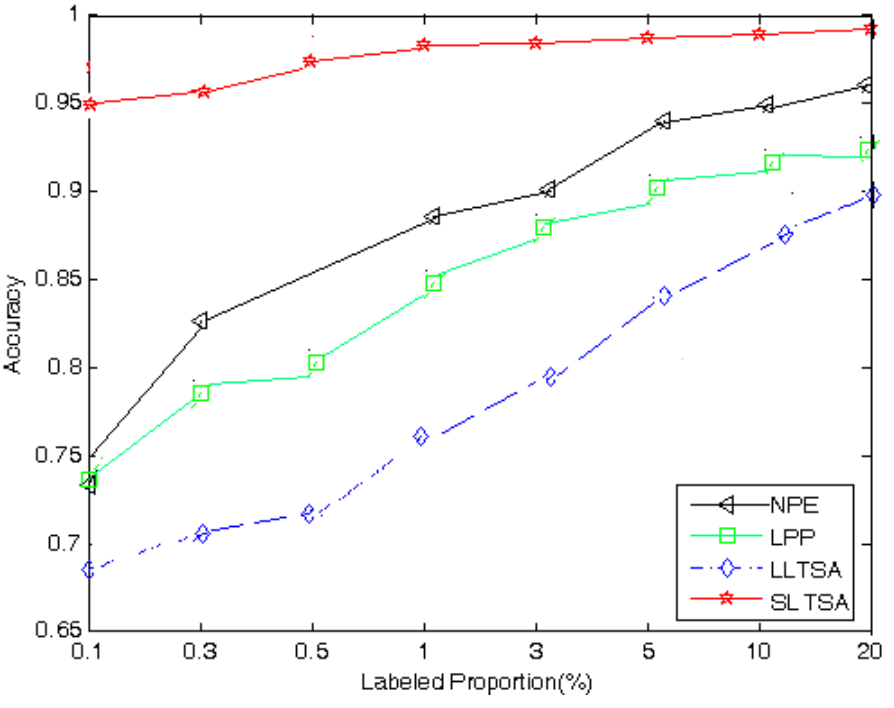


Fig. 2. Facial recognition accuracy

In this experiment, the training sets are randomly selected from the training set with 1000 images of each digit, while the test sets are all images of digits in the testing set.

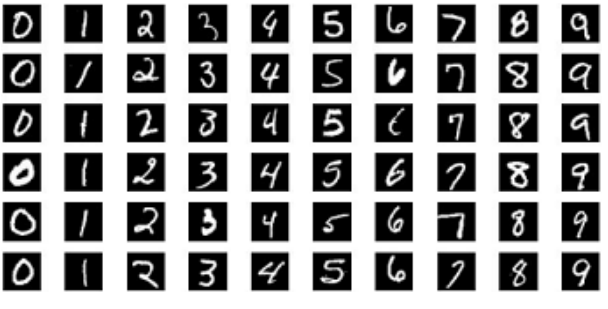


Fig. 3. Digit samples in MNIST data set.

The handwriting digits vary based on the habits of each person. The digits have been size-normalized and centered in 28×28 pixels gray-scale images. In this experiment each algorithm is performed on the data set ten times. The mean accuracies are shown in Table 1. Here we also choose K-NN as the classifier, and K=15. The feature space is set to

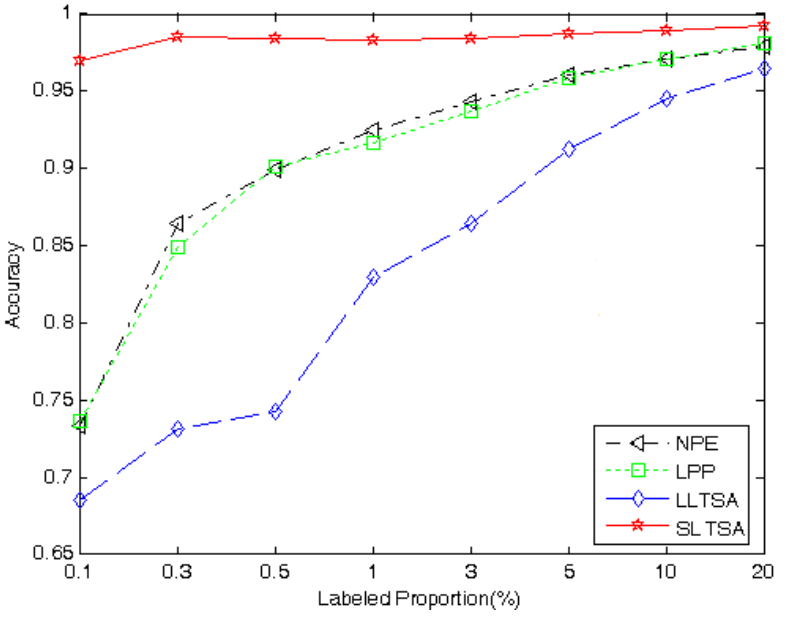


Fig. 4. MNIST recognition accuracy

However, our method still has some shortcomings. The method just gives an idea for semi-supervised manifold learning. First it introduces an additional parameter; second it ignores the statistical feature, which has been used in Charting a Manifold. From the theoretical analysis we can obviously find the effective parameters to noisy learning; thus the next research will focus on using probability to reduce the noise affection according to the effect of these parameters. Nowadays Jenkins, et al proposed an average method which can greatly reduce the incidence of face recognition errors. This method stimulate us to research a new manifold learning method to find texture manifold and structure manifold, which can increase the robustness applied in pattern recognition of manifold learning methods against noise disturbance.

5 Conclusion

In this paper, we proposed a novel semi-supervised method based on Local Tangent Space alignment. Different from some manifold learning algorithms that preserve the global or local metric knowledge of the high-dimensional data, local approximation is invariant to shift and scale, also, It utilizes the unlabeled data by intensive exploring the pairwise relation among the data points, and offers us a useful approach to model the labeled and unlabeled data in semi-supervised learning tasks. There are some problems to be solved for future work: 1.Choice of coefficients: find new approximation coefficients to precisely describe the relationship between point clouds; 2.find new way of computing inter-data to better maintain the global or local

property of high-dimensional data; 3.Nonlinear function: most of functions in this paper are linear, that are represented by matrix. We can replace them by nonlinear function to achieve better performance. We could perfect the algorithm and find better method to improve the efficiency in the future work.

Acknowledgment

This work was supported by the Research Foundation of Science & Technology office of Hunan Province under Grant(No. 2010FJ4213); Project supported by Scientific Research Fund of Hunan Provincial Education Department under Grant (No. 10C0498).

References

1. Zhu, X.: Semi-supervised learning literature survey. Technical Report 1530, Computer Sciences, University of Wisconsin-Madison (2005)
2. Zhang, T., Yang, J., Zhao, D., Ge, X.: Linear local tangent space alignment and application to face recognition. Neurocomputing 70, 1547–1553 (2007)
3. Comon, P.: Independent component analysis. A new concept. Signal Process. 36(3), 287–314 (1994)
4. Belkin, M., Niyogi, P.: Laplacian eigenmaps and spectral techniques for embedding and clustering. In: Proceedings of Advances in Neural Information Processing System, Vancouver, Canada, vol. 14 (December 2001)
5. Saul, L., Roweis, S.: Think globally, fit locally: unsupervised learnin of nonlinear manifolds. J. Mach. Learning Res. 4, 119–155 (2003)
6. Tenenbaum, J., de Silva, V., Langford, J.: A global geometric framework for nonlinear dimensionality reduction. Science 290, 2319–2323 (2000)
7. Zhang, Z., Zha, H.: Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. SIAM J. Sci. Comput. 26(1), 313–338 (2004)
8. Roweis, S.T., Saul, L.K.: Nonlinear Dimensionality Reduction by Locally Linear Embedding. Science, 2323–2326 (2000)
9. Gelan, Y., Xue, X., Li, T.: Manifold alignment via local block coordinate. In: 2nd International Workshop on Knowledge Discovery and Data Mining, WKKD 2009, pp. 697–700 (2009)
10. Gelan, Y., Xue, X., Huixia, J.: Semi-supervised regression via local block coordinate. In: 2nd International Congress on International Congress on Image and Signal Processing, pp. 232–236 (2009)

The Characteristics of Underwater Plasma Discharge Channel and Its Discharge Circuit*

Kaizhuo Lei**, Ning Li, Hai Huang, Jianguo Huang, and Jiankang Qu

College of Marine Engineering,
Northwestern Polytechnical University, Xi'an 710072, China
Lsr@nwpu.edu.cn

Abstract. In the paper, different types of energy generated by the underwater plasma discharging are analyzed and calculated. Together with data of the discharge voltage, discharge current, shock wave signal and the bubble wave, the method for calculating the resistance of the channel is deduced, based on the conversation of energy. With the method, a time-varying resistance model R(t) about the discharge channel is built up (Eq11) and some simulations were done. Moreover, the discharge circuit loop equation is deduced based on the proposed time-varying resistance model. By using the finite difference method, the discharge voltage and current are computed (Eq14-15). The simulation results are in a good agreement with the experimental result (Fig.4). The result in this paper provides guidance for the research of the discharging channel characteristics, the design of its circuit, the controlling of the energy conversion and the enhancement of the efficiency of the electricity to sound.

Keywords: underwater plasma acoustic source, discharge channel and circuit characteristics, simulation and experiment, time-varying resistance model.

1 Introduction

The impulse sound wave generated by the underwater plasma acoustic source, in virtue of underwater high-voltage discharge[1], has many advantages, such as high instantaneous emitting power, narrow wave width, wide frequency bandwidth, fast reaction and easy to focus or control [2]. Since these characteristics, this technology has been used widely in industry, medical treatment and so on, like lithotrities[3], removing the dirty in pipeline[4], food sanitation and sewage treatment[5], etc. In present years, this kind of source also have been applied in ocean geology reconnoitering[7], underwater targets detection[8], wideband acoustic interference[9] and sound power weapon[10].

Although the acoustic source based on the underwater plasma discharge has the said advantages and wide application, the transient and randomicity of the discharge process make the physical phenomenon very complex and the mechanism of electric-sound and

* National Natural Science Foundation of China (10974154);
National Innovation Project of China for Undergraduates (091069935);
NWPU Sci-tech Innovation Foundation (2007KJ01021).
** Corresponding author.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 619–626.
springerlink.com

the characteristics of the discharge channel not explicit enough. Therefore, exact calculating the resistance of the discharge channel will be helpful to the reaches for the mastery of the discharge circuit features. It also can give a guidance for enhancing the discharge efficiency and decreasing the energy dissipate in the discharge channel[11].

In this paper, a model about the resistance of the channel was built based on the plasma channel energy balance equation. With the help of the model and the finite difference method, the discharge current and discharge voltage during the underwater high-voltage pulse discharge were simulated. Simulation results and experimental results were in good agreement.

2 Conversation of Energy Equation

If water is stressed with a high voltage pulse having a rise time of tens of nanoseconds, stream develop and multi conductive channel forms between the electrodes. The intense Joule heating of the plasma in the channel results in the temperature in the channel up to 10^4K[12]-[15]. At the same time, the shock wave is generated together with radiation. Then bubble will be formed between the electrodes and emits bubble wave because of its fast expanding and collapsing. When plasma discharging underwater, many types of energy are produced concluding[16]: the internal energy inside the channel, the radiation energy, the shockwave and the babble energy. The energy assumed by the light and electromagnetic radiation can be ignored. Thus, the relationship among the different energy is

$$W_{dis} = W_{in} + W_{ra} + W_s + W_b \tag{1}$$

Where W_{in} is the internal energy inside the channel, W_{ra} the radiation energy, W_s the shock wave energy, W_b the bubble wave energy, W_{dis} the energy transfused into the channel and can be expressed by Eq2

$$W_{dis}(t) = \int_0^t i^2(t) R(t) dt \tag{2}$$

Here $R(t)$ is the resistance of the channel, $i(t)$ is the discharge current and t is the duration of the channel. Associating with the Eq1 and Eq2, it can be seen that, while the different energy at the right hand of the Eq1 and the discharge current are known, the resistance of the channel can be got.

3 Analyze and Computation of the Energy in Discharge Channel

3.1 The Internal Energy

The temperatures in the plasma channel between charging period of the two electrodes can reach in excess of 10000 K. At this temperature, water vapor no longer behaves as an ideal gas with constant specific heat; which rather dissociates into hydrogen and oxygen. Ionization of hydrogen and oxygen can begin at temperatures of this order as

well. For easy calculation, assuming the gas, including water molecules and different particle generated by the dissociation and ionization of the water molecules, are at a balance state[17]. Based on Landau-Lifshitz law, the particle or ion inside the charging channel can react in the following 5 forms (Eq3).

$$\begin{aligned}
&\alpha_1 : 2H_2O \leftrightarrow 2H_2 + O_2 && [+4.96eV] \\
&\alpha_2 : H_2 \leftrightarrow 2H && [+4.48eV] \\
&\alpha_3 : O_2 \leftrightarrow 2O && [+5.11eV] \\
&\alpha_4 : H \leftrightarrow H^+ + e^- && [+13.6eV] \\
&\alpha_5 : O \leftrightarrow O^+ + e^- && [+13.6eV]
\end{aligned} \tag{3}$$

The datum inside the bracket is the minimum energy for the corresponding reaction. α_i is the reacting rate of the corresponding reaction, with $0 \le \alpha_i \le 1$, where 0 depicts non reaction, 1 depicts reaction finish. Therefore the energy added inside the channel is

$$W_{in} = \sum_i E_i \tag{4}$$

Where E_i is the energy added during each reaction in Eq.3, can be calculated by Eq5.

$$E_i = \frac{1}{2} N_j \left(c_{vj} kT + \varepsilon_{0j} \right) \tag{5}$$

Here, $1 \le j \le 8$, $1 \le i \le 5$, c_{vj} is the freedom of the molecule or the atom, k is Boltzmann constant, ε_{0j} is the ground state energy of the each atom, T is the temperature in the channel, N_j is the number of different particle. If the initial number of the water molecule is N, after dissociation and ionization, $N_j = n_j N$ .Where n_j is the function of α_i which is gotten from Equation (3).

3.2 The Radiation Energy

Thermal radiation, which is strongly dependent on the temperature of the system, can be an important contribution to the energy partition of the underwater spark. In a naive model of the gas globe as a perfect blackbody radiator, thermal radiation represents a power loss proportional to the fourth power of the temperature as described by the Stefan–Boltzmann law,

$$W_{ra} = 4\pi R^2 S \cdot t \tag{6}$$

Where, R is the radius of the discharge channel. S is the electromagnetic flux normal to the surface of the channel. The flux at the surface of the discharge channel is the integration over photon frequencies

$$S = \int_0^\infty S_v dv \tag{7}$$

S_v is the flux at special photon frequency v.

3.3 The Shock Wave Energy and Bubble Wave Energy

When underwater plasma discharging, both shock wave and bubble wave are generated. At time domain, the shock wave generated while the plasma channel formed is before the bubble. Although the shock wave and bubble wave are two kinds of energy, they can be calculated with the same formula[18]

$$W_{s,b} = \frac{4\pi \cdot r^2}{\rho C} \int_0^{t_1} P(t)^2 dt \tag{8}$$

Where, r is the distance between the sensor and the source, ρ is the density of the water, C is the sound speed, t_1 is the signal's width, $P(t)$ is the signal.

4 The Model and Computation of the Resistance in the Discharging Channel

4.1 Simulation of the Resistance in the Discharging Channel

With the experimental data of the shock wave, bubble wave and the discharge current, the model is simulated. The parameters used in the simulation are as follows: temperature of the discharge channel is 30000K; the temperature inside the bubble is 4000K. α_i in equation (3) are α_1 =0.9, α_2 =0.3, α_3 =0.3, α_4 =0.01, α_5 =0.01 respectively.

For easy simulation, the derivative of the Eq2 about t is

$$i(t)^2 R(t) = \frac{dW_{dis}}{dt} \tag{9}$$

Where

$$\frac{dW_{dis}}{dt} = \frac{d}{dt}\left(W_{in} + W_{ra} + W_s + W_b\right) \tag{10}$$

Combining Eq3~10, the resistance of the discharge channel is deduced and the curves are drawn in Fig.1. From which, it can be seen that, before the formation of the discharge channel, the two electrodes can be considered as open circuit. Because of the discharge, the resistance of the channel changes from a few Ohms at the beginning of the discharge to only a few milliohms at the metaphase. At the end of the discharge, the resistance will increase.

In Fig.2, resistances inside the discharging channel are the result from reference, curve 1, [19], curve 2 [20]. Curve 1 was gotten with method of calculation of the discharge channel conductance. Curve 2 was a experimental result. Comparing Figures 1 and 2, a good agreement between the results of the references with that of the simulation of this paper will be found.

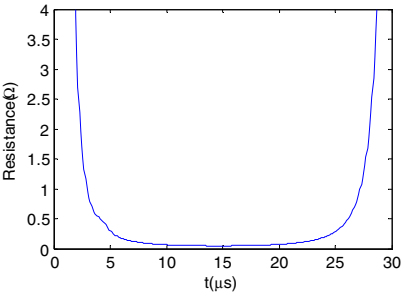


Fig. 1. Resistance waveform during the discharge

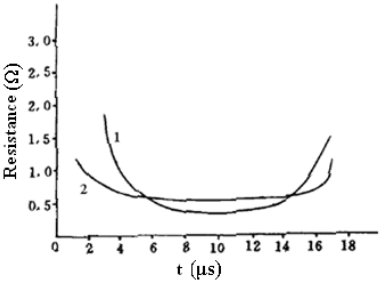


Fig. 2. Resistance curve in the discharge channel (Ref. 19-20)

4.2 Time-Varying Model of the Discharge Channel

In fact, the resistance during the discharging can be from kΩ, at the begin of discharging, to mΩ, at the end of the discharging. The equivalent resistance *R(t)* can be denoted with the exponential function

$$R(t) = A \cdot e^{-\left(\frac{t+a}{b}\right)^2} + R_0 \tag{11}$$

Where R_0 is the equivalent resistance of the electrodes after the channel formation. For the case of tap water (resistance rate3.42 Ωm), the distance and the top radius between the electrodes 2mm and 0.5mm respectively, R_0-7.5 kΩ *A, a, b* can be got from the Fig.2 and the experiments.

5 The Simulation and Experiments of the Discharge Circuit

5.1 The Model of the Discharge Circuit and Its Simulation

The equivalent circuit diagram of the discharge circuit is shown in fig.3, where, K is the switch and L is the inherent inductance.

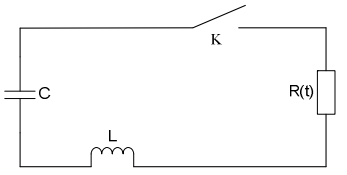


Fig. 3. Equivalent circuit diagram of the discharge circuit

Eq.12 is the loop equation of the discharge circuit.

$$L\frac{di(t)}{dt}+R(t)i(t)+\frac{1}{C}\int i(t)dt=0 \tag{12}$$

The derivative of Eq12 about t is

$$L\frac{d^2i(t)}{dt^2}+R(t)\frac{di(t)}{dt}+\left(\frac{dR(t)}{dt}+\frac{1}{C}\right)i(t)=0 \tag{13}$$

The Eq13 is solved with the finite difference method and the discharge current is expressed in Eq14:

$$i(n+1)=\frac{i(n)\left(2L+2R(n)\Delta t-\Delta tR(n+1)-\frac{\Delta t^2}{C}\right)}{L+R(n)\Delta t}-\frac{i(n-1)L}{L+R(n)\Delta t} \tag{14}$$

The initial conditions of the Eq14 are $i(0) = 0$ and $i(1) = 10$ μA. The discharge voltage is

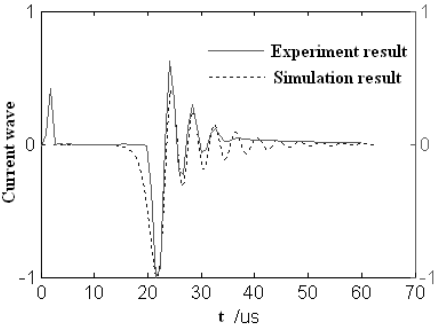
$$u(n)=i(n)\cdot R(n) \tag{15}$$

Based on Eq14, Eq15 with some certain initial conditions (according to the setup and the discharge), the normalized discharge current and voltage waveforms are shown in Fig.4. The reason of the oscillatory of the discharge current and voltage waveforms is: after the plasma channel formed, the equivalent resistance of the electrodes is at the level of mΩ, at this condition, the type of the discharge will convert to underdamped oscillations decay mode, just like the Fig.4.

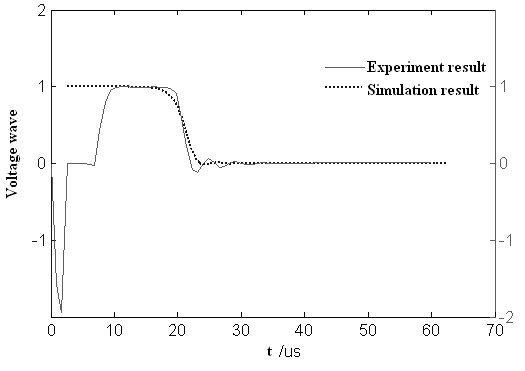
5.2 Comparison between the Experiment and Simulation

Experiment are done inside a 20×7×8m water pool with noise elimination. The charging capacity is 100μF. The charging voltage is 6kv.The charging electrodes is a coaxial cable, which in the depth of 2m. The shock wave and the bubble wave are collected with PCB pressure sensor, resolution 6.95MPa/V. The discharging current is measured with a Rogowski coil with 50kA/V; high-voltage detector is TEK6105A.

In order to test the validity of the method to calculate the discharge voltage and current which is based on the exponential model of the resistance, some simulation have been done at the same condition with the experiment. Both simulation result and experimental result are plotted in Fig.5. From which we can find a good agreement between the results.



(a) Discharge current



(b) Discharge voltage

Fig. 4. Comparison of the simulation and the experiment results

6 Conclusion

(1) The main energy generated during the underwater discharge are systematically analyzed and computed. Based on the measurement of the discharge current, the resistance of the discharge channel has been calculated due to the conservation of energy. From the comparing the simulation results to other references, a good agreement are found and the validity of the method is confirmed.

(2) The models of time-varying resistance and the discharge circuit are built up and some simulations have been done. The simulation results have a good agreement with the discharge voltage and current measured.

(3) A model for describing the discharging circuit is set up based on the proposed time-varying channel resistance. The results simulated with the model and the finite differential method are in good agreement with that measured.

(4) The simulation and experiment indicate that the main electric energy is converted to the internal energy when discharging, which heats the water and make the dissociation and ionization happen. Only 4%-8% energy converts to sound wave. In

order to enhance this efficiency, maybe some salt can be put into the water or a fine wire is set between the electrodes, which makes the formation of the discharge channel become easily.

References

1. Yutkin (USSR), The Electrohydraulic Effect. Jiashan Yu. Trans. Science Publish House, Beijing (1962)
2. Li, N., Chen, J., Huang, J., et al.: Sounding mechanisms and characteristics of various underwater sound sources. Applied Acoustics 29(4), 1–8 (2009) (in Chinese)
3. Fennigkoh, L., Traner, M., Reimer, D.: Characteristics and performance of lithotripter electrodes. Clinical Eng. 15(1), 23–28 (1990)
4. Mazzola, M.S., Grothaus, M.G., Walch, M., et al.: New electrical control methods to prevent power plant fouling. In: Proc.10th IEEE Int. Pulsed Power Conf., pp. 34–39. IEEE, Albuquerque (1995)
5. Yi, C., Wu, C., Xu, W., et al.: Study on sterilization by high voltage pulsed electric fields shock wave. High Voltage Engineering 33(2), 109–111 (2007) (in Chinese)
6. Zhang, Y., Zheng, J., Chen, H.: Theoretical study on high voltage pulse discharge technology in waste water treatment. High Voltage Engineering 33(2), 136–140 (in Chinese)
7. Wang, R.: Progress in detecting the geological formations and sediment properties by sound. Technical Acoustics 21(1-2), 96–98 (2002) (in Chinese)
8. Yan, P., Sun, Y.H., Zhou, Y.X., et al.: Underwater sound source based on pulsed corona discharge. In: CEIDP 2004, pp. 596–598. IEEE, Boulder (2004)
9. Pan, X., Huang, J., Chen, J.: Study on Underwater Wideband Acoustic Interference Source Based on Electro hydraulic Effect. Audio Engineering 32(5), 66–68 (2008) (in Chinese)
10. Lei, K., Huang, J., Zhang, Q., et al.: Overview of Directed Energy Weapon System. Torpedo Technology 18(3), 161–166 (2010) (in Chinese)
11. Timoshkin, I.V., Fourcre, R.A., Given, M.J., et al.: Analysis of the Transient Process in Underwater Spark Discharges. In: Conference Record of the 2006 27th International Power Modulator Symposium, pp. 310–313 (2006)
12. Lu, X., Pan, Y., Zhang, H.: The electrical and acoustical characteristics of pulsed discharge in water. ACTA Physica Sinic. 51(7), 1549–1553 (2002) (in Chinese)
13. Sultanov, M.A.: Hydrodynamic processes in a high-power pulsed discharge. Sov. Phys., Thys. 21(7), 815–821 (1976)
14. Edward, A.M.: Experimental Investigation of a High-Energy Density, High-Pressure Arc Plasma. J. Apply. Phys. 31(2), 255–267 (1960)
15. Lu, X., Pan, Y., Liu, K., et al.: Spark model of pulsed discharge in water. JAP 91(1), 25–31 (2002)
16. Lu, X., Pan, Y.: Theory of Plasma Character in a Pulsed Discharge in Water. Journal Of Basic Science And Engineering 8(3), 310–316 (2000) (in Chinese)
17. Roberts, R.M., Cook, J.A., Rogers, R.L.: The energy partition of underwater sparks. J. Acoust. Soc. Am. 99(6), 3465–3475 (1996)
18. Cole: Underwater explosions. Princeton University Press, Princeton (1948)
19. Li, P., Jing, F.: Calculations of shock wave and plasma parameters of the discharge in liquid. Journal of Zhejiang University (Natural Science) 28(1), 27–35 (1994) (in Chinese)
20. Tsuchiya, H., et al.: Generation and Propagation of Pressure Wave by Spark Discharge in Liquid. Aun. CIRP 31(1), 107–110 (1988)

The Evaluation Index System of the Industrial Clusters' Core Competence and Fuzzy Synthetic Evaluation*

Yumin Liu[1] and Zhenhua Hu[2]

[1] Business School, Central South University, College of Business, Hunan Normal University
Changsha, China
lym.corn@163.com
[2] Business School, Central South University,
Changsha, China
Zhenhhu@163.com

Abstract. The article illustrated that there are four factors work on the core competence of industrial clusters. Then the writer designed a model of evaluation index system, and used fuzzy synthetic evaluation to analyze the core competence of industrial clusters quantitatively and to identify the problems. So we can put forward corresponding countermeasures and measures to enable the rapid development of industrial clusters, at the same time promoting the regional economy and the country's macroeconomic development.

Keywords: industrial clusters, core competence, fuzzy synthetic evaluation.

1 Introduction

Industrial clusters are playing a more and more significant role in regional economic development, both academic and government departments pay high attention to the phenomenon of industrial clusters. Many countries and regions have developed industrial clusters research programs to identify and implement cluster development strategies. Through the interactive cooperation and exchange, enterprises of cluster can bring economies of scope and economies of scale into play, and can also have a strong spillover effect, so as to promote economic development of a particular region and the whole nation. Industrial clusters have become a worldwide economic phenomenon. Many scholars and experts carried out researches and put forward many theories. Such as the external economic theory of Marshall (1920) [1] and Krugman(1991) [2];the Regional aggregation economic theory of Weber (1909), Hoover (1948) and Barton; the regional production complexes and growth pole theory of Korosovski and Perroux; the diamond model /new competition economic theory of Porter (1998); regional economic dynamics/new industrial space theory of Scott, Storper [3], Harrison and

* This research work was supported by the National Natural Sciences Foundation of China (No. 70921001) "Research on Uncertain Decision Theory and Application under Complex Environment".

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 627–633.
springerlink.com

Waler. And the transaction cost theory of Williamson and the economic network of Harrison (Harrison) in recent years. From the 1990s, Chinese scholars have joined the research ranks, and advanced the theory of innovation spaces (Wang Jici, 1994), the new industrial district scale structure, contacts, embeddedness (Li Xiaojian, 1997), the regional economy, the informal system (Qiu Baoxin, 1999), business networks and entrepreneurial networks (Guangdong scholars), social network (Taiwan scholars) and innovation networks (Gai Wenqi, 2002). These ideas and theories explained and demonstrated the law of the existence, developing and changing of industrial clusters from different levels and different points of view. However, how did the core competence of industrial clusters form? How to evaluate the core competence of the industrial cluster qualitatively and quantitatively? The series issues of how to enhance the core competence of industrial clusters of China still need further study and explanation.

2 The Core Competence of Industrial Clusters

C. K. Prahalad and G. Hamel (1990) published an article "The Core Competence of the Corporation" in the Harvard Business Review, raised the original notion of the core competence of the corporation for the first time. They considered that core competence is "the accumulation of knowledge of the organization, particularly the knowledge of how to coordinate different production skills and how to combine a variety of skills organically [5]. In their opinion, due to the difference of the core competence, the corporations in the same industry have different performances. The core competence is the key dominance for the corporation to survive and develop in the fierce market.

As one of the important backbones to promote regional and national economies, industrial clusters should also strengthen its core competence continually in the background of production and competition globalization. The core competence of industrial clusters formed in long-term development of the clusters, so that clusters maintain long-term stable competitive advantage, and obtain the most basic competence for long-term sustainable development. This competence builds on the foundation of the clusters' core resources, embodied by the comprehensive competitiveness beyond of other competitors in production, culture, innovation, marketing, and so on.

The core competence of industrial clusters should satisfy the following three aspects: Firstly, embeddedness, the clusters' core competence is accumulated by unique methods during the practice in the long term, which rooted in the corporate group and which is difficult to mimic or create for the outside competitors; secondly, scalability, the core competence of industrial clusters is not limited to a particular enterprise or product, but content in a variety of business areas within the cluster; thirdly, periodical, the clusters' core competence is a performance that the cluster gains a dominant position in the market competition in a certain period.

So, while many capabilities have contributed to the generation of competitive advantage, but only the core competence can create a sustainable dynamic competitive advantage. Industrial clusters have low-cost flexible manufacturing system, excellent competitiveness of collaborative culture, dynamic mechanism of continuous innovation and marketing competitiveness.

3 The Elements and Evaluation Index System of Industrial Clusters' Core Competence

3.1 Low-Cost Flexible Manufacturing Competitiveness

In product manufacturing, because the industrial enterprises within the cluster are geographically close to each other, so that they can find a variety of upstream and downstream industry resources in the region nearby, and can form vertical integration of corporate chain within the industrial cluster rapidly through the professional division of labor and collaboration. Saving transaction costs, reducing the search range and information search cost and shortening the matching path, the close geographical location and frequent interaction help enterprises to rapidly launch new products and quickly capture the market and form industrial scale in a short time. Concerning the competition of modern enterprises, the advantages of time and speed becomes more and more obviously. Enterprises and supporting cooperative ones that can produce new products to meet market demand fast, often can take high monopoly profits because of its first-mover advantage in competition.

The author use following index to evaluate low-cost flexible manufacturing competitiveness: the rate of adoption of advanced equipment; average labor productivity; average capital input-output ratio; human Resources Recruitment average annual cost; ability to adapt to market changes.

3.2 The Competitiveness of Collaborative Culture

Collaborative culture is the deep-level factor during the formation of the clusters' core competence, which is difficult for competitors to imitate.

On the one hand, The formation of cluster's collaborative culture, stems from the basis of cluster formation——"trust and cooperation".[6] The long-term cooperative relationships between enterprises within the cluster do not need to maintain completely by the contract, but carry out through commitment and trust. Established in the corporate culture, the divisions of labor and cooperation relations are very stable and able to generate synergies, which cannot be obtained by the individual enterprise outside the cluster. So the enterprises within the cluster have a unique competitive advantage when faced with external competitors.

On the other hand, the formation of cluster's collaborative culture stems from the cluster talents mechanism. Under the pressure of competition, cluster enterprises attach importance to human resources development, especially the high-tech human resources development, and the mechanism to attract and stimulate talent, and form their own unique products. Thus, the products produced by the cluster corporations are of the same type, but vary in quality, style or otherwise. The clusters convert the competition among clusters' enterprises to a synergistic competition by achieving product differentiation.

These indicators are used to evaluate the competitiveness of the cluster collaborative culture: degree of division of labor; the average compliance rate of cluster enterprises; cluster cohesion; communication degree; insight and innovation in entrepreneurs.

3.3 Continuous Innovation Competitiveness

The enterprises, universities, research institutions within the cluster locate closing to each other, which is conducive to the rapid diffusion and sharing of technology. We know the most basic model of technology diffusion is neighbor-proliferation: from the center to the edges, decaying with the distance. And the geographic proximity of the cluster enterprises is particularly conducive to enterprise employees carry out a range of formal and informal exchanges and promote the dissemination of tacit knowledge, and access to knowledge spillovers[4]. Geographical concentration will also help stimulating the companies to win the race peer effects. Within the regional cluster, comparisons between enterprises continue to promote the constant improvement of business management and accelerate technological innovation. The presence of the high intensity and constant comparison, coupled with easy access to innovation resources and the pursuit of the leading aspire of value, making the innovation continually producing in cluster region and quickly imitated by other companies within the cluster, the value of innovation and rapid implementation and testing, thus the innovation cycles continue to be shortened and the innovation cycle is accelerating.

These part indicators list as follows: the rate of adoption of advanced technology; the ratio of new products; R&D investment rate; Rate of technology developers; R&D input-output ratio.

3.4 Marketing Competitiveness

Marketing competitiveness of industrial clusters mainly assumes as a region brand and marketing innovation system [7]. On the one hand, it is easy to establish for the industrial cluster to built a region brand. Compared to a single corporate brand, region brand (such as West Lake Longjing tea, Italian fashion) is a more visual image, has a broader, continuing brand effect. The more product brand name within the cluster, the bigger cluster force is, and the more likely to attract a variety of resources. It is conducive to the cluster companies gain market share, and create a competitive cluster. On the other hand, it is easy for the cluster enterprises to establish marketing innovation system. Cluster enterprises can establish a marketing innovation system, which built market development as the core and customer's demand as the center. In this system, the marketing innovation training is strengthened, the quality of personnel is improved, and the marketing innovation is promoted, thereby strengthening the cluster market competitiveness.

4 Fuzzy Comprehensive Evaluation Model of Industrial Clusters' Core Competence

Many indicators of industrial clusters' core competence cannot be quantified with the specific value, can only be qualitative assessment of its extent, so the author used two-lay fuzzy comprehensive evaluation method to evaluate [8] [9]. Based on fuzzy

comprehensive evaluation and the core competence of the cluster evaluation index system, the author established the following evaluation model.

4.1 Single Factor Evaluation

4.1.1 Establish Factor Set

We call factors that influence the core competence of industrial clusters as a factor set. It is a common collection, use U to stand for:

$$U = \{U_1, U_2, U_3, U_4\}$$

Of which:

U_1 Indicate that the industry cluster competitiveness of low-cost flexible manufacturing

U_2 Indicate that the industry cluster competitiveness of the cultural synergy

U_3 Indicate that the industry cluster competitiveness of continuous innovation

U_4 Said the marketing industry cluster competitiveness

Considering U_1, U_2, U_3, U_4, each have 5 influence factors, so

$$U_i = \{A_{i1}, A_{i2}, A_{i3}, A_{i4}, A_{i5}\} \quad (i=1, 2, 3, 4)$$

4.1.2 Establish Weight Set

In general, the importance of each factor is different. Therefore, to reflect the degree of the importance of each factor, each factor should be assigned a certain weight, and U establish a corresponding set of weights A:

$$A = \{a_1, a_2, a_3, \cdots\cdots, a_n\}$$

And meet $\sum_{i=1}^{n} a_i = 1$

Suppose we use the AHP method or random survey by the weight of U is: A_i (i = 1,2,3,4).

$$A_i = \{a_{i1}, a_{i2}, a_{i3}, a_{i4}\}$$

4.1.3 Establishment Evaluation Set

This evaluation set to evaluate a variety of evaluation objects may make the composition of the total set of evaluation results. With V stand for:

$$V = \{V_1, V_2, V_3, \cdots\cdots, V_m\}$$

The core competence of industrial clusters evaluation, the evaluation set can be set:

$$V = \{v_1, v_2, v_3, v_4, v_5\}$$

Which v_1 = very strong; v_2 = strong; v_3 = moderate; v_4 = weak; v_5 = very weak.

Obtained by methods such as Delphi, the evaluation results matrix of U_1 is $C_{ij}^{(1)}$

$$C_{ij}^{(1)} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C15 \\ C21 & C22 & C23 & C24 & C25 \\ C31 & C32 & C33 & C34 & C35 \\ C41 & C42 & C43 & C44 & C45 \\ C51 & C52 & C53 & C54 & C55 \end{bmatrix}$$

(i＝1, 2, 3, 4, 5;j＝1, 2, 3, 4, 5)

Order: $r_{ij}^{(1)} = \dfrac{C_{ij}}{\sum_{j=1}^{5} C_{ij}}$

Construction U_1 single factor comparison matrix R_1

$$R_1 = \begin{bmatrix} r_{11} & r_{12} & r_{13} & r_{14} & r_{15} \\ r_{21} & r_{22} & r_{23} & r_{24} & r_{25} \\ r_{31} & r_{32} & r_{33} & r_{34} & r_{35} \\ r_{41} & r_{42} & r_{43} & r44 & r_{45} \\ r_{51} & r_{52} & r_{53} & r_{54} & r_{55} \end{bmatrix}$$

Similarly, comparison matrix can be constructed R_2 , R_3, R_4 ,

Order U_1 evaluation vector B_i there are:

$$B_i = A_i o R_i = \{b_{i1}, b_{i2}, b_{i3}, b_{i4}, b_{i5}\}$$

Therefore, it results:

$$B_1 = A_1 o B_1$$

$$B_2 = A_2 o B_2$$

$$B_3 = A_3 o B_3$$

$$B_4 = A_4 o B_4$$

4.2 Two-Lay Judges

For the first-lay index set $U = \{U_{1,}, U_2, U_3, U_4\}$, the weight assigned to $A_i = \{a_{i1,}, a_{i2}, a_{i3}, a_{i4}\}$, Make two-lay evaluation matrix as:

$$R = \begin{bmatrix} B_1 \\ B_2 \\ B_3 \\ B_4 \end{bmatrix}$$

For the two-lay evaluation, have $B = A \circ R = \{b_1, b_2, b_3, b_4, b_5\}$

According to the principle of maximum membership, if b_1 the most, it shows that low-cost flexible manufacturing competitiveness, cluster collaborative cultural competition, continuous innovation, marketing competitiveness are strong, the core competitiveness is also strong; if b_5 is the maximum, then its core competence is weak.

5 Conclusion

The strength of industrial clusters' core competence is important to industrial clusters development in the long-term, and also important to China's entire economic development process. According to the evaluation of the cluster's core competence, implementing targeted strategies and taking corresponding measures to promote industrial clusters developing rapidly and healthily has been a priority. Using the fuzzy comprehensive evaluation method for evaluating indicator system model can comprehensively evaluate the cluster's core competence and help to identify the problems. We can put forward corresponding countermeasures and measures to enable the rapid development of industrial clusters, at the same time promoting the regional economy and the country's macroeconomic development.

References

[1] Krugman: Development, Geography and Economic Theory. The MIT Press, Cambridge (1995)
[2] Porter, M.E.: Clusters and New Economics of Competition. Harvard Business Review (11) (1998)
[3] Scott, A.J., Stoper, M.: Regions, Globalization, Development. Regional Studies 37 (2003)
[4] Wang, J.C.: Innovative space: Enterprise Clusters and Regional Development. Peking University Press, Beijing (2001)
[5] Prahalad, C.K.: The Core Competence of the Corporation. Harvard Business Review (3) (1990)
[6] Zheng, S., Zhou, L.: Knowledge of industrial cluster coordination mechanism. World Regional Studies (3) (2004)
[7] Wei, S., Zhao, Y.: Explore the competitive advantage of industrial clusters. Financial and Economic Issues (5) (2002)
[8] Li, Y., Xu, X.: The mathematic methods of business decision-making. Economic Science Press (2003)
[9] Jiang, Y.: Fuzzy Math Tutorial. National Defence Industry Press (2003)

Research on Integration of Heterogeneous Wireless Access Communication Networks

Yu Liu

[1] School of Resources and Environment Engineering,
Chongqing University, Chongqing, China
[2] School of Information Engineering
Chongqing Vocational Institute of Engineering, Chongqing, China
dfiafsuqf2@163.com

Abstract. In recent years, the wireless communication networks shows a high-speed, broadband and general development trend, including the rapid development of wireless technology represented by WLAN, WiFi, WiMAX, WSN, Mesh, 3G and B3G. As 3G is a present heat technology, the research on how to achieve the integration of 3G networks with WLAN and WiMAX networks is full of practical value. This paper just discusses and researches about the issue, and proposes appropriate solutions.

Keywords: heterogeneous networks, Integration, WCDMA ,WLAN.

1 Introduction

Recent years, WLAN, Adhoc networks, WSN, WiFi, WiMAX and many types of wireless networks technology were generally researched, simultaneously, the mobile communication technology has experienced rapid development from 1G (AMPS, TACS) to 2G (GSM, CDMAOne) until the present heat 3G (WCDMA, CDMA2000, TDSCDMA) just in a short decade. So far, the global wireless communication system turns out the trend of broadband mobile and mobile broadband. Under the present heterogeneous networks environment, there are types of wireless technology exists and which are relatively independent, lacking of an effective coordinate system. This may cause problems of system disturbance, frequency resource scarcity, networks seamless handoff and etc unsolved. So it is an urgent issue how to realize the effective integration of these heterogeneous networks, especially the 3G networks with present wireless technology [1].

2 The Integration of WLAN and 3G

WLAN is the wireless LAN standards made by IEEE. At present, serial standards 802.11 is widely used, frequency in the 2.4 GHz range, which has an advantage of cheap price, flexible networks, high-speed wireless data access supporting, unrestricted band and etc. The expanding of WLAN from cable LAN into wireless world, is aiming

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 635–642.
springerlink.com

at supplying data communication under the wireless environment, but not voice capacity. The 3G networks are built on cellular infrastructure, which is most suitable in supporting data service under mobile environment. Cellular infrastructure backs up signal handoff between different cellular networks, thereby provide customers with mobility covering the whole networks.

3G whose logos is to supply the public with carrier-class businesses, set up it's aim of at providing multimedia business at beginning, which contains voice and data, global roaming and so on. So the user can enjoy nice networks safety as well as different QoS. As know, WLAN has obvious advantages in indoor wireless data business, but it's weaknesses of low coverage, bad mobility and nonsupport of voice are also prominent; otherwise, 3G can support wireless voice and data business, has a good coverage and mobility, but low bandwidth is its big weakness. If the 3G operator combines with WLAN supplier, integrates 3G networks with WLAN networks, then the wireless users can enjoy high-speed data transmission as well as seamless roaming between different networks without any restriction [2, 3].

Table 1. Contrast parameters table of WLAN with 3G technology

Parameter	WLAN	3G
Working frequency	2.4GHZ working frequency	3.11GHZ permissive frequency
Broadband	11-108Mbps	2Mbps
Coverage	50-150m	global roaming
Service orientation	high-speed data	voice and data integration
Equipment	Data/PC center	telecom operations center
Main technology	FH and DSSS	CDMA

As in table 1 has listed the contrast parameters of WLAN and 3G technology, and the following is to show you two modes for integration.

2.1 The Tightly-Coupled Integration Mode of 3G with WLAN

As showed in figure 1, the WLAN networks apply the same way with 3G base stations to connect with 3G core networks, which takes full advantages of mobility, safety and quality service existed in 3G core networks. But to connect with 3G core networks directly, the present WLAN production of tightly-coupled integration need to be remodified and remade, and the equipments support UMTS and CDMA are also different. In this mode, WLAN brings 3G main transmission networks business traffic and signaling, connects to core networks' AAA server through AC, and be authenticated by HLR (Home Location Register). And the data channels of WLAN directly connected with PDN through AC. The service charge information of using WLAN are gathered by AC, and reported to 3GPP charge system through AAA server. So, making a little modification to the standards and architecture of 3GPP networks, can realize uniform authentication and charging. In this mode, apply mobile IP technology can also support roaming operations between two different systems.

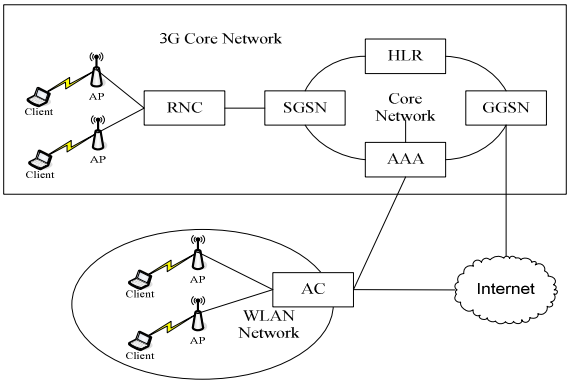


Fig. 1. The tightly-coupled integration architecture of 3G with WLAN

2.2 The Loosely-Coupled Integration Mode of 3G with WLAN

As showed in figure 2, WLAN connects with mobile networks as a supplementary form, only shares the usage of AAA with 3G, so avoid the hybrid of data flows caused by two different access technology at the 3G core networks node. This combination guarantees the totally separate of two wireless networks, make them complete independence. Sharing AAA infrastructure, to make the 3G operators can use accordant user authentication under the 3G and WLAN hybrid networks circumstances, and help the operators gathering the using information of WLAN, then form a uniform customer bills including the using information of 3G and WLAN.

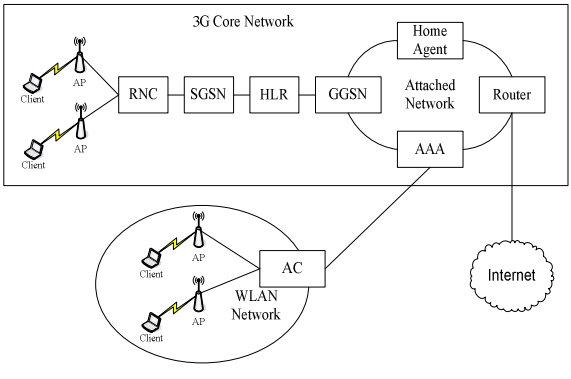


Fig. 2. The loosely-coupled integration architecture of 3G with WLAN

This integration mode has flowing characteristics: the data of two systems are independent from each other. 3GPP maintains the present standards unchanged, and the users enjoy services as before. The WLAN systems do not need to make any changes, and the users can get one piece of bill from the same operator. So there's no need of roaming operation between WLAN and 3G, it's the simplest integration method. In

this mode, using mobile IP technology can also support the roaming operation within two different systems. The concrete application: realize function FA (Foreign Agent) at the connecting node of AC and GGSN (Gateway GPRS Support Node), and set up a server as HA (Home Agent) in additional networks. So applying the mobile IP operating process, and users can roam between two systems by Wi-Fi mobile converged devices. But in this mode, WLAN can't directly visit resources and businesses in 3GPP networks, as the present mobile IP is not perfect whose performance can hardly meet the carrier-class requirements.

3 The Integration of WiMAX with 3G

WiMAX (Worldwide Interoperability for Microwave Access), is a broadband wireless access method with mobility, has higher speed and wider coverage. It can supply users with a 75MB transmission speed, 50Km coverage at farthest, support 120km/h high-speed vehicular movement, and a better hierarchical QoS guarantee in wireless multimedia businesses. It is mainly used in wireless communication within MAN (Metropolitan Area Networks).

Table 2. Contrast parameters table of WiMAX with 3G technology

parameters	WiMAX	3G
Working frequency	2.6GHZ frequency	3.11GHZ frequency
Acceptance range	10km	unrestricted
Transmission speed	120Mbps	2Mbps
Mobility	120km/h	250km/h
Service	Broadband low-speed data	Voice and data
Main technology	OFMD/MIMO/OFDMA	CDMA/TDMA

As showed in table2, WiMAX mobile communication system positions at packet data transmission service, its peak data transmission speed can arrive at 75Mbit/s, much higher than 3G system. And it is mostly used for fixed, portable or low-speed users access, don't back up seamless roaming under high-speed movement. But 3G mobile communication system has advantages of supporting high-speed roaming as well as a whole network covered communication service. If combine the two systems of WiMAX with 3G when networking, make WiMAX emphasis on achieving broadband mobile and meet the high-speed service requirements in some hot area, while 3G emphasis on achieving high-quality voice communication and seamless roaming needs of mobile communication, then it can realize advantages complementary, save investment cost, and arrive a win-win situation as well as meet the users needs of smart card [4, 5].

3G is established in mobile broadband, while WiMAX based on broadband mobile, the both have a nice complementarity in technology and service, so it is feasible to unite networks of 3G with WiMAX, even a preponderant hybrid network system. According to the level of their interdependences, the integration architecture of WiMAX with 3G can be divided into two modes: tightly-coupled integration architecture and loosely-coupled integration architecture.

3.1 The Tightly-Coupled Integration Mode of 3G with WLAN

As showed in figure 3, in the tightly-coupled integration architecture, the WiMAX system gateway connects with GGSN of 3G system through tightly-coupled interface, the IF shield the property of WiMAX to SGSN, make SGSN take WiMAX system as a separate base station. WiMAX uses the authentication, charging of 3G networks, its upper runs 3G relative protocol, needs some transformation to the protocol stack and adds corresponding interfaces to 3G protocol stacks. In the tightly-coupled integration architecture, the access networks of WiMAX use the same way as 3G to combine with the core network, and directly connect to SGSN node of 3G networks, both of the access network of WiMAX and 3G are equal. The WiMAX gateway conceals the details of WiMAX network technology from 3G networks, deploys all protocols (authentication, mobility management and etc) required by 3G wireless access networks.

This tightly-coupled integration mode will make a big attack to the core network of 3G, because the SGSN network components of 3G core network have to been redesigned to support the ascending service whose service load is far from its property. Among the tightly-coupled integration architecture, there also brings up some new requirements on users' equipments, the mobile termination of the integration network can run the IEEE802.16 and 3G protocol stack at the same time.

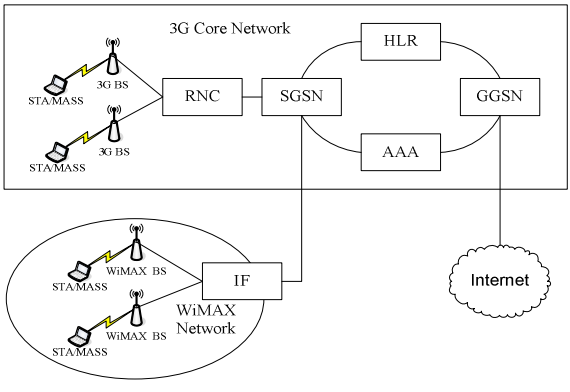


Fig. 3. The tightly-coupled integration architecture of 3G with WiMAX

3.2 The Loosely-Coupled Integration Mode of 3G with WiMAX

As showed in figure4, in the loosely-coupled integration architecture, the WiMAX system gateway connects with GGSN of 3G system through loosely-coupled interface to achieve interconnection with each other. WiMAX system support the authentication based on SIM, then can share users' data, authentication and charging function with 3G system. In the tightly-coupled integration architecture, the data of WiMAX network directly connects with public networks out of 3G core network. To 3G core network, there is no directly interface with WiMAX, their access networks are parallel. To make the users from WiMAX access network can also enjoy the service provided by packet switching area of 3G core network, there need to construct a data tunnel among 3G core network and WiMAX network, through which can transmit the data from 3G core

network to WiMAX access network. WiMAX access network and 3G access network can use different authentication, charging, mobile management mechanisms and protocols. But, to realize seamless handoff, different mechanisms and protocols must coordinate each other.

The loosely-coupled integration mode utmost maintains the independence of WiMAX and 3G network, makes less changes to present networks, is more open, more advantages and competitive. But in the loosely-coupled integration architecture, as there is no direct tunnel between the two networks, the data and signal are all switched through internet, and the data received by mobile node have to pass HLR (Home Location Register). So the transmission is inefficiency and handoff delay is long, not easy to provide highly real-time service, such as video streaming transmission.

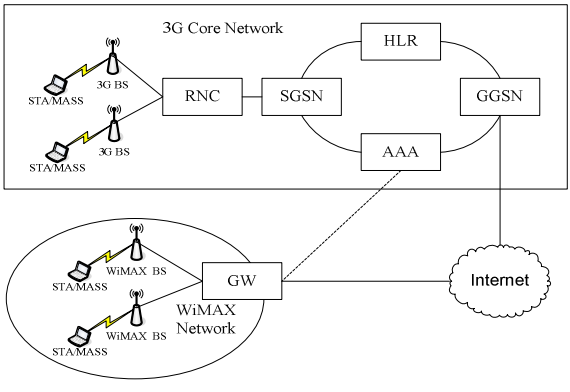


Fig. 4. The loosely-coupled integration architecture of 3G with WiMAX

4 The Integration of WLAN, WiMAX and 3G Networks

As showed in figure 5, connecting with the two network architecture of WiMAX and WLAN Mesh, realize the wireless broadband access of Mesh gateway node with WiMAX the bone back network through the broadband wireless access system. The MAP in WLAN Mesh network can connect with the user terminal directly, also be used as AP access interface of traditional WLAN and combine multi-wireless local area networks together by the way of wireless Mesh. On the one hand, it can achieve intercommunication among the traditional wireless local area networks; on the other hand, it can also realize the wireless broadband access. So, WLAN can satisfy some hot areas with higher speed data transmission, and WiMAX connects different hot areas together, then it can achieve a wider coverage of high-speed data access. Otherwise, 3G network positions at mobile subscribers' voice communication and low-speed data wireless communication over the whole networks, it arrives an advantage complementary of three technologies integrates multi access ways, and forms a hierarchical broadband wireless access network, and provides more efficient service to different needs.

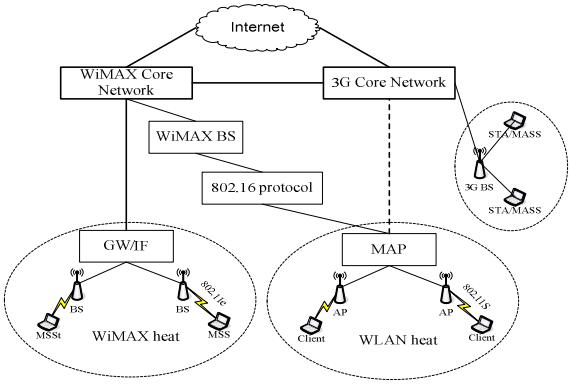


Fig. 5. The integration architecture of WLAN, WiMAX and 3G networks

5 Handoff between 3GPP System and Non-3GPP System

HMIPv6 (Hierarchical Mobile IPv6), is the enhanced version of IPv6. It was proposed for reducing signal traffic and speeding mobile connection. HMIPv6was proposed by Internet Engineering Task Force, mainly aimed at providing calculation function for Internet, and solving seamless handoff problems within heterogeneous networks. As showed in figure 6, HMIPv6 brings a new node- MAP (Mobile Anchor Point) as a local entity to manage the mobile handoff. Using mobile IPv6, the mobile node will send location update information to its relational nodes when each time it updates the location. UE directly access to Internet through common wireless access ways. To support the mobility, it needs to be updated in local bindings and home bindings, and also the updated the registration to HSS through AR. UE (users' equipments) access the 3GPP target system, including authentication, address allocation and wireless resources construction. The authentication is from 3GPP system to UE, and UE after updating local bindings through target system, the hierarchical data can then pass the new transmission link.

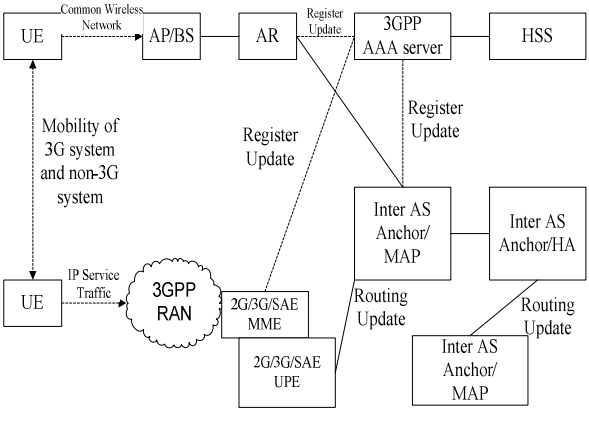


Fig. 6. Mobility proposal based on HMIPv6 system

6 Conclusion

The integration of all kinds of wireless networks is an inevitable trend with the development of wireless technology, has important realistic meaning and research value. This paper discusses on the heat technology-integration of 3G with WLAN, proposes some typical solutions, and gives out a handoff method between 3G system and non-3G system. But as the integration of heterogeneous wireless networks is complicated system engineering, referring to system architecture, protocol stack, seamless handoff, seamless management and security, so to realize the real seamless integration between wireless networks is still facing many challenges and having much work to be done.

References

[1] Li, J.: Theory and Technology Research on integration of heterogeneous wireless networks. Electronic Industry Press, Beijing City (2009)
[2] Tang, Z.-j.: Solutions and Application of The Integration between WLAN and 3G Networks. Jiangsu Communication 6, 60–63 (2008)
[3] Xiao, L.: The Integration of WLAN with 3G wireless data networks. Telecommuni Cations Information 6, 33–35 (2006)
[4] Yuan, L.: Research on combined networking in 3G and WiMAX. Mechanical and Electrical Engineering and Application 43(3), 153–156 (2007)
[5] Zhang, J.-l.: Study of Hybrid Networks Between WiMAX and 3G technologies. Software Guide 6, 83–84 (2008)

Research on the Radiation Field of Mono-radiation Wide-Band Leaky Coaxial Cable Using for 3G Mobile Communication

Guojian Yang and Zhuo Wang

Mechanical and Electrical Engineering College,
Northeast Forestry University, 150040 Harbin, China
{fll1981314,yguojian}@163.com

Abstract. To solve the problem of shortage of low frequencies along with the development of mobile communication, the research on the leaky communication system using for the mobile communication in blind and semi-blind zone was towards to high frequency and wide frequency area. Combining with the 3G and 2G mobile communication frequency band, the method of expanding the frequency range of single mode radiation leaky coaxial cable was studied, and the radiation field of leaky coaxial cable whose frequency range is expanded was simulated, and the relationship between frequency range expanding and radiation field was analyzed, the conclusion set a basis for the application of leaky coaxial cable.

Keywords: leaky coaxial cable, wide-band, mono-radiation, radiation field.

1 Introduction

Leaky coaxial cable is also called sequential antenna, according to the electromagnetic theory, it has some periodic or non-periodic slots on its outer conductor[1]. All the slots along the cable are the radiation sources of electromagnetic wave, and when the signal is transmitting in the cable, part of the energy is radiated out of the cable from the slots. The leaky electromagnetic signal can be received by the receiver along the cable; or signal from the mobile transmitter get into the leaky coaxial cable, and this communication method can make up the disadvantage of the existing of blind zone in the traditional communication, so that the omni-bearing duplex communication between leaky coaxial cable and outer space can be realized easily[2][3]. Leaky coaxial cable is the combination of wire and wireless communication, at the same time is the integration of microwave and electromagnetic field and electric wave transmission theory, and it can be widely used in the communication in subways, tunnels, caves, mines, high ways, ships and other blind or semi-blind zone. Especially along with the coming of 3G, the requirement of wide-band communication is getting higher and higher[4], it is very necessary to expand the frequency band

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 643–650.
springerlink.com

of leaky coaxial cable, and the research on the radiation field of wide-band leaky coaxial cable is turned to be an important issue today[5][6]. In this paper, the method of expanding the frequency band of leaky coaxial cable is analyzed, and the simulated results show that in the mono-radiation zone, the wide-band leaky coaxial cable can work well in the 3G communication.

2 Mono-Radiation Theory

If the propagation coefficient of an uniform infinitely long magnetic current J_m in z direction is β, the electric field of the point $p(z_0, r)$ that out of this magnetic current is [1]:

$$E = -j\omega\mu_0 \nabla \times \prod_m \tag{1}$$

Here, ω is angular frequency, $\prod_m$ is magnetic hertz vector, and $\prod_m = (1/j4\pi\omega\mu_0)\int_{-\infty}^{+\infty} J_m\left(e^{-jk_0R}/R\right)dz$, k_0 is wave number in free space, μ_0 is the permeability, and $R = \sqrt{r^2 + (z - z_0)^2}$ is the distance between view point P and magnetic dipole as shown in Fig.1:

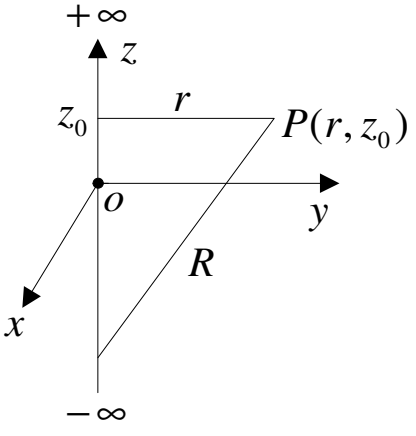


Fig. 1. Magnetic current source

If the magnetic J_m transmits in z direction, there is only z component of $\prod_m$ [1]:

$$\prod_{mz} = \frac{1}{j4\pi\omega\mu_0}\int_{-\infty}^{+\infty} J_m \frac{e^{-j(\beta z + k_0\sqrt{r^2 + (z - z_0)^2})}}{\sqrt{r^2 + (z - z_0)^2}} dz \tag{2}$$

Let $\beta_\varphi^{\ 2} = k_0^{\ 2} - \beta^2$, $z - z_0 = r\sinh t$, then:

$$\prod_{mz} = \frac{J_m e^{-j\beta z}}{j4\pi\omega\mu_0}\int_{-\infty}^{+\infty} e^{-j\beta_\varphi r\cosh t}dt = -\frac{J_m}{4\omega\mu_0}H_0^{(2)}(\beta_\varphi r)e^{-j\beta z} \tag{3}$$

Here $H_0^{(2)}(x)$ is zero-order Hankel function of second kind, from (3) and (1), we can obtain:

$$E_\varphi(r,z) = j\frac{J_m\beta_\varphi}{4}H_1^{(2)}(\beta_\varphi r)e^{-j\beta z} \tag{4}$$

Where $H_1^{(2)}(x)$ is one-order Hankel function of second kind.

The magnetic current J_m of leaky coaxial cable which period of vertical slots is P is also a periodic function with P, and the Fourier series of it is:

$$J_m(z) = \sum_{n=-\infty}^{+\infty} J_{mn}e^{-j\frac{2\pi n}{p}z} \tag{5}$$

Here J_{mn} is the amplitude of J_m, n is integral number, then the field in φ direction of leaky coaxial cable can be written as:

$$E_\varphi(r,z) = \frac{j}{4}\sum_{n=-\infty}^{+\infty} J_{mn}\beta_\varphi H_1^{(2)}(\beta_\varphi r)e^{-j\beta_n z} \tag{6}$$

and $\beta = k_0\sqrt{\varepsilon_r}$, $\beta_n = \beta + 2\pi n/p$, β_φ is the propagation constant of electromagnetic wave in radial direction, β_n is the propagation constant of *n*th spatial harmonic wave in z direction.

It is known from $\beta_\varphi = \sqrt{k_0^{\ 2} - \beta_n^{\ 2}}$ that if $\beta_\varphi \le 0$, the field falls in radial direction, and there is no radiation, so the surface wave exists around leaky coaxial cable; only if $\beta_\varphi > 0$, the electromagnetic wave can be radiated in radial direction, and that is the requirement of radiation:

$$k_0^{\ 2} - \beta_n^{\ 2} > 0 \tag{7}$$

Now we can obtain the following expression by using $k_0 = 2\pi f/c$ and $\beta = k_0\sqrt{\varepsilon_r}$:

$$-nf_1 < f < -nf_2 \tag{8}$$

where $f_1 = c/P(\sqrt{\varepsilon_r}+1)$, $f_2 = c/P(\sqrt{\varepsilon_r}-1)$, c is the velocity of light in free space, ε_r is the relative permittivity of the dielectric material in the cable, and n must be integral number only.

It is obviously that there are infinitely many spatial harmonic components around leaky coaxial cable with periodic slots, and they are called modes because they are similar to the transmission modes in waveguide[7]. Only if $n \le -1$, the radiation exists, and (8) is the frequency range which can product radiation wave[8].

The diagram for radiation frequency band of different harmonics is illustrated in Fig. 2:

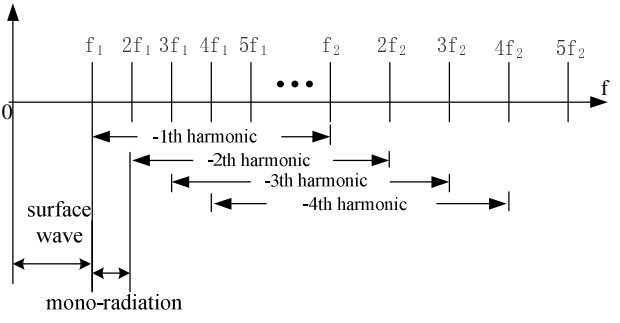


Fig. 2. Spatial harmonic modes

According to the interference of leaky wireless wave from slots, the frequency band can be divided into 3 areas[9]:

1. The frequency band in which β_φ is imaginary number, that is in $f < f_1$. In this area, there is only surface wave around leaky coaxial cable and the field of them can be disturbed very easily, the transmission distance is short, so it is not suitable to be used in the mobile communication.

2. The frequency band in which β_φ is real number and there is only mono-radiation, and that is in $f_1 < f < 2f_1$. In this area, there is only -1th mode radiation wave around, and this mode is needed in the mobile communication.

3. The frequency band in which β_φ is real number and there is multi-modes radiation, when the frequency increases to $2f_1$, -2th mode radiation wave appears, and along with the increasing of the frequency, -3th, -4th, etc. modes begin to radiate, but different modes have their different transmission constants, the result of it is the standing wave, which will product strong interference between every modes, and the fluctuation of field make the quality of mobile communication fall, so it is necessary to suppress the high order spatial harmonic component, that is to make the frequency between f_1 and $2f_1$, realize mono-radiation, and the frequency band is f_1.

3 Wide-Band Leaky Coaxial Cable

To expand the frequency band of leaky coaxial cable, all the high order spatial harmonic in the -1th radiation band must be suppressed, and the more the modes are suppressed, the wider the frequency band is. There are three methods to achieve it:

1. Add some new slots on the outer conductor, and their size and shape are the same as the original ones. The according high order harmonic can be suppressed by adjusting the distance between the new and old slots.

2. Adjust the inclined angle of slots while keep the arrangement and length of original slots invariant to suppress the high order harmonics.

3. Adjust the length of the slots while keep the arrangement and inclined angle of original slots invariant can also suppress the high order harmonics.

The first method is often chosen because the latter two methods are hard to achieve.

The following is leaky coaxial cable with zigzag slots which is usually used in our country:

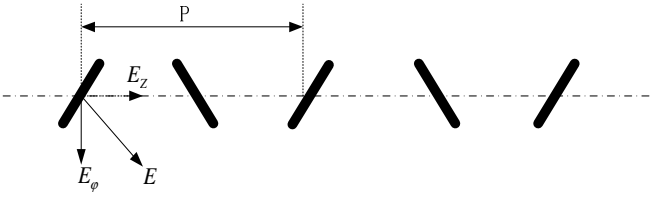


Fig. 3. Leaky coaxial cable with zigzag slots

The field around it is[10]:

$$E_\varphi(r,\varphi,z)=\sum_{n=-\infty}^{\infty} Z_n\left(1-e^{jn\pi}\right)R(\eta_n,r,\varphi)e^{-j\beta_{zn}^{z}} \tag{9}$$

Here Z_n is periodic function in z direction, $R(\eta,r,\varphi)$ is related with r, φ and radial transmission constant η. When n is even, $1-e^{jn\pi}$ equals to zero, nth harmonic can be suppressed, so there is odd order modes of leaky coaxial cable with zigzag slots, and the frequency band is between f_1 and $3f_1$ [11].

The following figure is leaky coaxial cable with double zigzag slots in one period:

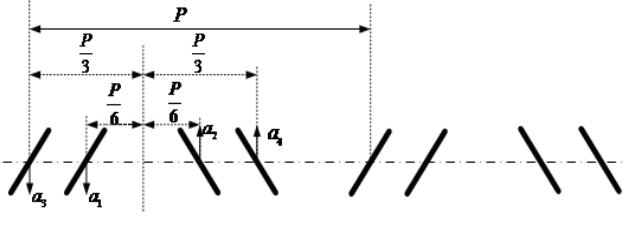


Fig. 4. Leaky coaxial cable with double zigzag slots

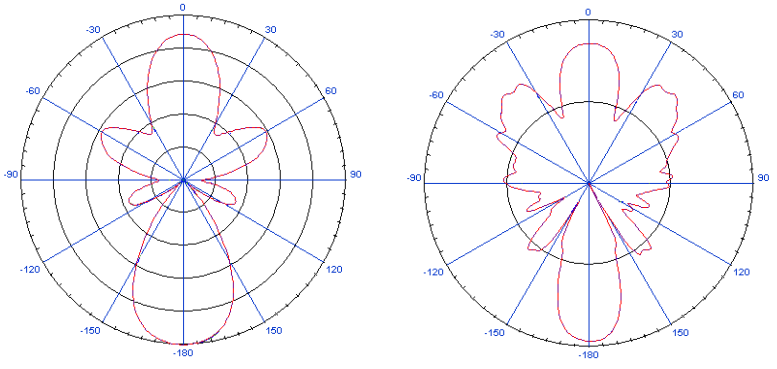
The same as above, the frequency band of this kind of cable is between f_1 and $5f_1$, and it is four times of original band. So we can know that the expanding multiple of frequency band is two times of the pairs of zigzag slots in one period of leaky coaxial cable, that is if there are m pairs of zigzag slots in a period, the frequency band can get 2m to increase after expanding.

According to the International Telecommunication Union standardization, the frequency band for 3G communication is 1885MHz to 2025MHz, and in general case,

the slot period of leaky coaxial cable is 0.22m, $\varepsilon_r = 1.25$, then we can obtain that the mono-radiation frequency bands of leaky coaxial cable with periodic slots, zigzag slots and double zigzag slots are 643.83MHz—1287.66MHz, 643.83MHz—1931.49MHz and 643.83MHz—3219.15MHz respectively. So the last kind of leaky coaxial cable is suitable for using in 3G mobile communication in blind or semi-blind zones.

4 Simulation Results

To check the performance of the leaky coaxial cable after frequency band expanding, the simulated radiation fields analyzed by HFSS are as follows. The cable's parameters are: the length of it is 300m; the radiuses of inner and outer conductor are 9mm and 22.8mm; the incline angle of slot is $\pi/4$; $\varepsilon_r = 1.25$, and the mono-radiation frequency band is 643.83MHz—1287.66MHz.

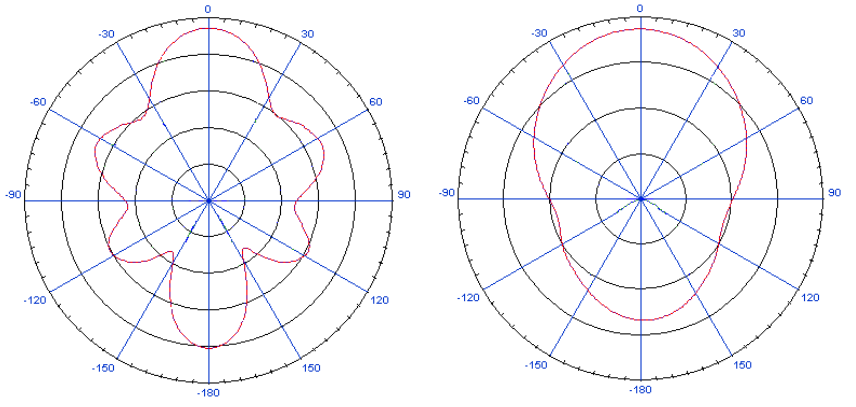


(a) Frequency is 1885MHz (b) Frequency is 2025MHz

Fig. 5. Radiation field of leaky coaxial cable with zigzag slots

These two figures are the radiation fields of leaky coaxial cable with zigzag slots at the lowest and highest frequency of 3G. Figure (a) shows that the leaky coaxial cable is in the mono-radiation status when the frequency is 1885MHz, and the field of it is smooth, so it can work well in the mobile communication. When the frequency increases to 2025MHz, the leaky coaxial cable gets out of mono-radiation, the field showed in figure (b) is not stable, and more side lobes which can disperse the energy of leaky coaxial cable or import some unnecessary interference emerge, so if the cable works under this frequency, the quality of mobile communication will be affected, and the communication distance will be cut short too.

The following two figures are the radiation field of leaky coaxial cable with double zigzag slots in one period at the frequency of 1885MHz and 2025MHz, and its frequency band has been expanded to 643.83MHz—3219.15MHz.



(a) Frequency is 1885MHz (b) Frequency is 2025MHz

Fig. 6. Radiation field of leaky coaxial cable with double zigzag slots

We can know from Fig.6 that leaky coaxial cable is in the mono-radiation status no matter the frequency is 1885MHz or 2025MHz, the main lobe remains in the original position, and the field is stable, so it can be used in 3G mobile communication.

5 Conclusion

The simulation results show that change the slot structure of leaky coaxial cable in one period can expand the frequency band and application range, and turn the cable be useful in 3G communication. The radiation character and main lobe of leaky coaxial cable does not change after band expanding, and the side lobe energy decreases at the same time. So it is effective of adding zigzag slots in one period to expand the frequency band of leaky coaxial cable until to the frequency range of 3G mobile communication, but if there are enough many slots in one period, because of the close distance, the electromagnetic wave from different slots will disturb with each other, so the limitation of the adding slots and the radiation field of leaky coaxial cable with more zigzag slots will be studied in the future.

References

1. Zhang, X., Zhao, Z.X.: Research on the Radiation Characteristics of Leaky Coaxial Cable and Spiral Antenna in the Blind Zone. In: 4th International Conference on Communications, Circuits and Systems, Guilin, China (2006)
2. Heddebaut, M.: Leaky Waveguide for Train-to-Wayside Communication-Based Train Control. IEEE Transactions of Vehicular Technology 58, 1068–1076 (2009)
3. Hamid, M.R., Hall, P.S., Gardner, P., Ghanem, F.: Switched WLAN-wideband Tapered Slot Antenna. Electronics Letters 46, 23–24 (2010)

4. Ohira, M., Miura, A., Ueba, M.: 60-GHz Wideband Substrate-Integrated-Waveguide Slot Array Using Closely Spaced Elements for Planar Multisector Antenna. IEEE Transactions on Antennas and Propagation 58, 993–998 (2010)
5. Dai, H., Liu, Y., Li, Y., Wang, X., Xiao, S.: Far Field Measurement for Spatial Polarization Characteristics of Antenna. Journal of Applied Sciences 27, 606–611 (2009)
6. Brockett, T., Rahmat-samii, Y.: A novel portable bipolar near-field measurement system for millimeterwave antennas: construction, development, and verification. IEEE Magazine on Antennas and Propagation 50, 121–130 (2008)
7. Kim, S.-T., Yun, G.-H., Park, H.-K.: Numerical Analysis of the Propagation Characteristics of Multiangle Multislot Coaxial Cable Using Moment Method. IEEE Transactions on Microwave Theory and Techniques 46, 269–278 (1998)
8. Zhou, Y., Zhang, X., Yang, X., Guo, L.: Research on the expansion technology of the single-moderadiation frequency band of leaky coaxial cable. Applied Science and Technology 33, 4–6 (2006)
9. Sakabe, I.: VHF-UHF Super-wideband Leaky Coaxial Cable. Optical Fiber & Electric Cable 6, 41–45 (1995)
10. Zhang, X. Yang, X.-d., Guo, L.-l., Zhang, S.: Research of Leaky Coaxial Cable Using for Mobile Radio. In: Proceeding of 1st Workshop on Multidisciplinary Researches for Human Life and Human Support, Harbin, China (2001)
11. Wang, J., Jian, S.: Analysis of the radiation mode of the leaky coaxial cable and the restraint of the high-order mode. Journal of China Institute of Communications 21, 17–22 (2000)

Pre-design and Analysis of Flow Field of Spacer-Free Nozzle of Aluminum Roll-Casting

Ying Zhou[1], Pei Feng[2], Huanlong Hu[2], Pengjian Li[2], and Cijun Shuai[2,*]

[1] School of Material Science and Engineering, Central South University, Changsha, 410083, Hunan Province, P.R. China
zhouying@mail.csu.edu.cn
[2] School of Mechanical and Electrical Engineering, Central South University, Changsha, 410083, Hunan Province, P.R. China
Fengpei851688@163.com, huhuanlong@126.com, 632404633@qq.com, shuai@mail.csu.edu.cn

Abstract. To optimize the design of flow field of nozzle in aluminum roll-casting, a coupled fluid-thermal finite element analysis using ANSYS software, was performed to explore the distributing of velocity and temperature of melt aluminum in spacer-free nozzle by MATLAB. Curve of velocity was gently sloping but curve of temperature was steep at outlet of spacer-free nozzle. It was explored that geometrical sizes of nozzle should be pre-designed to get better curves of velocity and temperature before placing spacers into nozzle.

Keywords: roll cast; spacer-free nozzle; flow field; coupled fluid-thermal; finite element analysis.

1 Introduction

Modern twin-roll casting is appealing urgently for discovery and research of optimization of processing. Normal roll casting doesn't request stricter design for flow field and thermal field, so spacers inside nozzle are used to hold flow's passage and separate the flow. There are many kinds of spacers used in nozzles, and at the normal speed of roll casting, they are not sensitive to the uneven spreading of velocity and temperature of flow field of melt aluminum, especially with lower speed and thicker sheet of roll casting [1, 2].

But for high speed and thin sheet of roll cast, those kinds of nozzles with spacers are not good enough. As industrial experiments cost too much, simulation of flow field and thermal field of melt aluminum in nozzle is used before on-site experiments. Usually comparing with a lot of results of different designs, and combining with experience of engineering, then the most reasonable optimized plans are chosen for doing industrial experiments.

From former work, conclusion is drawn that one-spacer nozzle has potential capacity to be designed for higher speed and thinner sheet of roll cast. Before optimized designing one-spacer nozzle, the shape of nozzle without spacer should be pre-design

* Corresponding author.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 651–655.
springerlink.com

first, which will be discussed in this paper. Section 2 gives the detailed description of the problem; Section 3 explains characters of this model of flow field, and describes the coordinating and boundary conditions used in the analysis; Section 4 presents and analyzes the simulation results; Section 5 is conclusion.

2 Description

Spacer-free nozzle's main geometry sizes of flow fluid include thickness and width of entrance and outlet, and the slope of side edges from entrance to outlet. Due to the limitation to producing process, the width of either entrance or outlet is fixed: if the width of sheet is decided, then the width of outlet of nozzle is determined. Melted aluminum flows from entrance of front-box, crossing the interface of front-box and nozzle, passes nozzle, and finally outflows at the outlet of nozzle, as shown in Fig. 1. Then the fluid of melt metal with certain spreading of velocity and temperature will arrive at cast-rolling-area between the gap of two rollers. This model of front-box and nozzle is prepared for the future simulation model of cast rolling, which will include multiphysics fields considering casting rollers, melt cave area, and the cast-rolling area altogether.

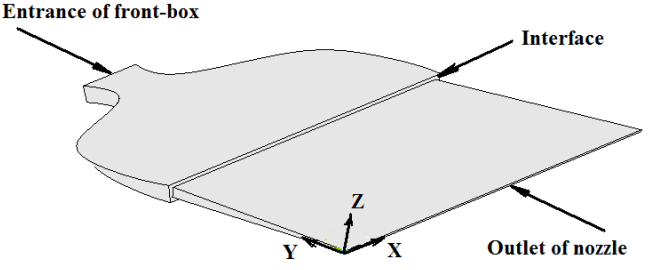


Fig. 1. Schematic of flow field of front-box & spacer-free nozzle

Model of flow field of both front-box and nozzle is built to take the liquids in both the front-box and the nozzle into consideration all together. As an isolated model of the nozzle field will bring inconsistency to the model at the interface of entrance, a merged field of both front-box and nozzle will avoid errors caused by separated modeling. Geometrical sizes are shown in Table 1.

Table 1. Geometrical sizes of nozzle

Name of size	Width of entrance	Width of outlet	Thickness of entrance	Thickness of outlet	Depth of nozzle	Speed of roll casting sheet
Along axis	*x*	*x*	*z*	*z*	*y*	*y*
Units	m	m	m	m	m	m/s
Value	0.780	0.890	0.016	0.007	0.370	0.0167

3 Analysis of Modeling

3.1 Mesh Type

In module of FLOTRAN of ANSYS software, there are two different meshing methods: free mesh and mapped mesh, the former is strongly recommended to module of FLOTRAN, as the latter suitable to module of Structure. In the analysis of fluid FEA, mapped mesh method is helpful to reduce the calculating error, avoid divergence, and obtain more accurate results. Even if the models are complex, the mapped mesh method should be applied as much as possible. Fig. 2 is mapped meshing model of flow field of front-box and nozzle.

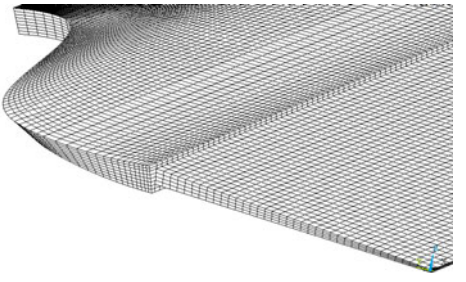


Fig. 2. Mapped meshing model of flow field

3.2 Boundary Condition

3.2.1 Fluid Field Boundary Conditions

At the entrance of the front-box the boundary condition is set to the Entrance Condition, $v_x=0$, $v_y= v_0$, $v_z=0$, where v_x,v_y,v_z: variable component of fluid velocity vector whose units is [m/s]. At the exit of the nozzle it is set to the Free Exit Conditions, $P=0$, where P: Pressure of the fluid whose units is [N/m^2]. At the bottom of the front-box and at nozzle where fluids are constrained they are set to the Fixed Boundary Conditions, $v_x =0$, $v_y =0$, $v_z =0$. At the top of the front-box the flowing speed along the Z-direction is set to zero, $v_x =0$, $v_y =0$.

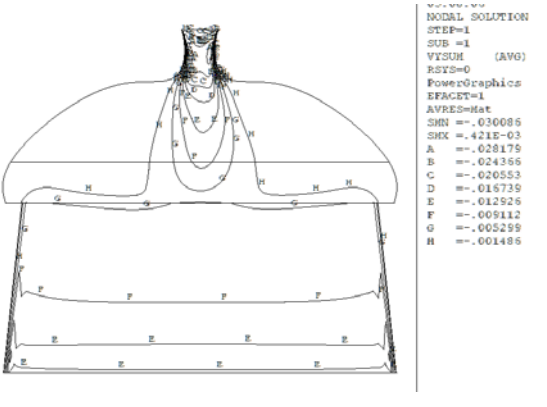


Fig. 3. Isosurface of contour display of v_y

3.2.2 Thermal Boundary Conditions

At the entrance of the front-box the boundary condition is set to temperature: $t=t_0$, where t: temperature of nodes, its units is K. At the bottom of the front-box the heat flux to the boundaries between melted aluminum and front-box is applied: $h=h_1$. At the top

of the front-box the heat flux to the boundaries between melted aluminum and air is set: $h=h_2$. At internal nozzle surfaces where fluids are constrained corresponding heat flux to the boundaries between melted aluminum and nozzle are set: $h=h_3$, where h represents: heat flux to the boundaries whose units is [W/m^2]. [3]

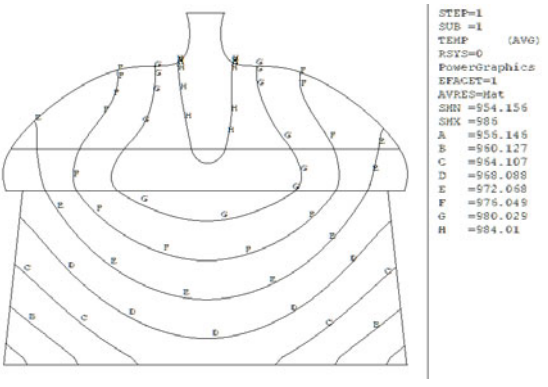


Fig. 4. Isosurface of contour display of temperature

4 Result of Simulation

Postprocessing by ANSYS[4]: ANSYS's default result display of flow fields in front-box and nozzle is in vector display, but not as clear as contour display. Plotting of analysis result of y-dimensional velocity vector (*vy*) with 8 equational-division contours, on the mid-plane(*z=0*), which is shown as in fig. 3. Fig. 4 is the isosurface of temperature in contour display on the mid-plane (*z=0*). Mid-plane is where the maximum of *vy* and temperature locate.

Stricter analysis of result by software MATLAB: Make comprehensive graphics based on the nodes' two degrees of freedom: velocity and temperature of midline and outlet surface getting from result

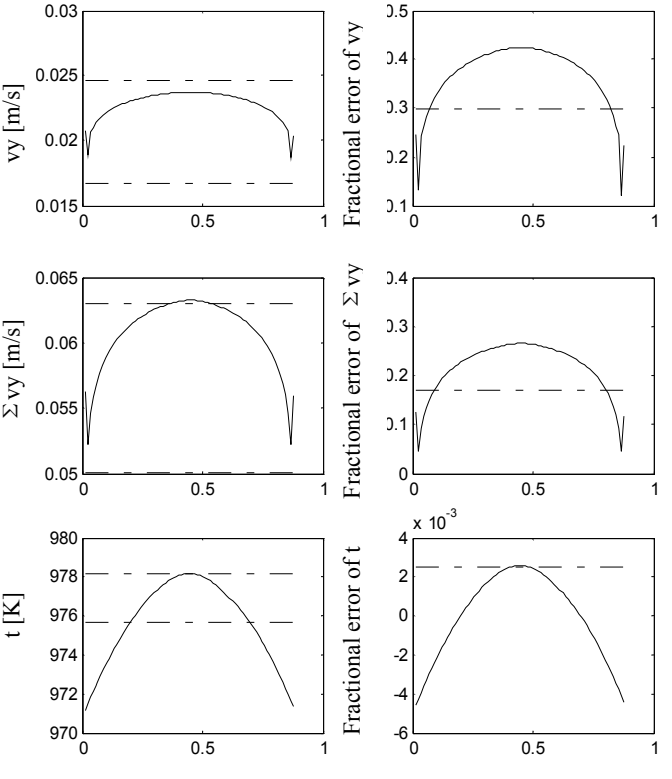


Fig. 5. Comprehensive graphic of *vy* & *t* at 0.0167 m/s

files of ANSYS, by using MATLAB software, as shown in Fig. 5. Postprocessing of results by MATLAB figured out absolute and fractional errors of velocity and temperature of melt aluminum at outlet of nozzle, which could judge how suitable the flow field of nozzle will be for producing with higher casting speed and thinner thickness of aluminum sheet.

5 Conclusion

Nozzle without spacer is disscussed for optimizing flow fluid of nozzle, which will be set spacers to obtain even spreading of velocity (*vy*) and temperature (*t*). From analysis of spacer-free nozzle, it is concluded that *vy* is smooth and flat, as *t* is not at outlet of nozzle, which means that spacer-free nozzle is good for *vy* but bad for *t*. Temperature at outlet shows a convex curve, and maximum of *vy* and *t* are at the midpoint of width, so spacers at mid part have to be added to reduce the detrimental effect. Further more, it is supposed that optimazing the geometrical sizes of nozzle without spacer, is helpful to get better shape of nozzle which will serve explanate curve of both *vy* and *t* before placing spacers in nozzle[5].

Acknowledgments

This thesis is sponsored by the following funds: (1) The Natural Science Foundation of China (50875267 and 50605063); (2) Program for Doctoral Dissertation of P. R. China (201032) (NCET-10-0792); (3) Program for New Century Excellent Talents in University (NCET-10-0792); (4) Postdoctoral Science Foundation of Central South University.

References

1. Amit, S., Yogeshwar, S.: Modeling of fluid flow and heat transfer in twin-roll casting of aluminum alloys. Materials Transactions 43, 206–213 (2002)
2. Haga, T.: A high speed twin toll caster for aluminum alloy strip. Journal of Materials Processing Technology 113, 291–295 (2001)
3. Kang, C.G., Kim, Y.D.: Model experiments for the determination of the heat-transfer coefficient and transition thermal analysis in the direct rolling process. Journal of Materials Processing Technology 84, 210–224 (1998)
4. Forum of Simulation We Use (in Chinese), `http://www.simwe.com`
5. Zhou, Y.: Physical-geometrical Coupling Rule and Design of Fluid Field of Aluminum Melt of Continuous Roll Casting, Ph.D. thesis. College of Mechanical and Electrical Engineering, Central South University, Changsha, P. R. China (2007) (in Chinese)

Optimize Structure Dimension of Spacer-Free Nozzle of Aluminum Roll Casting Using Orthogonal Experiment Design

Ying Zhou[1], Pei Feng[2], Pengjian Li[2], Huanlong Hu[2], and Cijun Shuai[2,*]

[1] School of Material Science and Engineering, Central South University, Changsha, 410083, Hunan Province, P.R. China
zhouying@mail.csu.edu.cn
[2] School of Mechanical and Electrical Engineering, Central South University, Changsha, 410083, Hunan Province, P.R. China
Fengpei851688@163.com, 632404633@qq.com, huhuanlong@126.com, shuai@mail.csu.edu.cn

Abstract. Optimization of nozzle was crucial process to improve quality of production of roll casting, and nozzle without spacer should be pre-designed before placing spacer in it. Using orthogonal experiment design, three main structure dimensions and the speed of roll casting sheet were chosen as 4 factors, to make orthogonal array of 3 levels for simulation of nozzle. It was concluded that an optimal combination of factors and levels could be provided by orthogonal experiment, which might achieve optimum of good distributing of velocity and temperature of flow fluid in nozzle. The optimized result would be used as the original condition of integrated design of nozzle and spacers.

Keywords: orthogonal experiment design; spacer-free nozzle; roll cast; coupled fluid-thermal; finite element analysis.

1 Introduction

Accurate analysis and design of modern twin-roll casting is appealed for producing aluminum roll casting sheet with high speed and thin sheet [1, 2]. During many technological parameters, the distributing of velocity and temperature of flow fluid at outlet of nozzle used in roll casting process, is a key to get high quality sheet. As there are some main geometrical parameters of nozzle to consider, such as sizes of nozzle, shape and location of spacers and so on, so optimization methods should be introduced to improve geometry parameters.

The orthogonal experiment design is chosen here as optimization algorithm for spacer-free nozzles' geometry. It can offer excellent factor design, and find an optimal combination of factor levels that may achieve optimum [3,4].

* Corresponding author.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 657–661.
springerlink.com

Section 2 gives schematic of main geometry parameters of spacer-free nozzle which will act as factors of orthogonal experiment; Section 3 explains orthogonal arrays used in this experimental design; Section 4 presents and analyzes simulation results of orthogonal experiments; Section 5 is conclusion.

2 Description

For batch simulation, structure of model should be simplified to cut down the time needed for modeling and analysis. Structure of nozzle without spacer is all built by planes, and it is already very simple, so only front-box can be changed for easier drawing. Without changing shape of curves on horizon, bottom of front-box is from curved surface to planes, as shown in Fig.1.

It is supposed that sizes of front-box are fixed, so are width (0.78 m) of entrance and depth (0.37 m) of nozzle. Then three other geometrical parameters of nozzle, as shown in Fig. 2: width of outlet (A), thickness of entrance (B), and thickness of outlet (C) can be optimized. Besides, speed of roll casting sheet is worth considering too [5]. Values range of these four parameters is shown in Table 1.

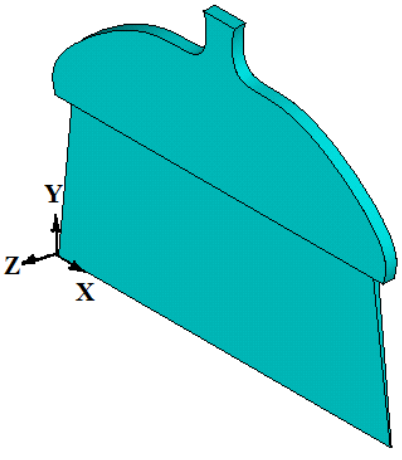


Fig. 1. Schematic of flow field of front-box & spacer-free nozzle

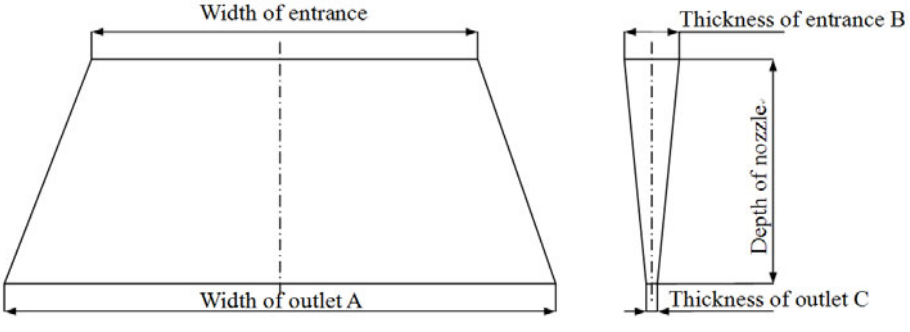


Fig. 2. Front view and side elevation of spacer-free nozzle

Table 1. Range of optimization parameters

	Width of outlet A	Thickness of entrance B	Thickness of outlet C	Speed of roll casting sheet
Units	m	m	m	m/s
Upper limit	0.920	0.019	0.007	0.167
Lower limit	0.860	0.013	0.003	0.0167

3 Orthogonal Experiment Design

3.1 Choose Orthogonal Array for Optimization Parameters

Orthogonal arrays are used in this experimental design in order to explore an optimal combination of geometry values of spacer-free nozzle, shown as factor levels that may achieve optimum.[] Four factors chosen here are width of outlet, thickness of entrance, thickness of outlet, and speed of roll casting sheet. Orthogonal array is shown in Table 2.

Table 2. Orthogonal array of 4 factors and 3 levels of each factor

Number of experiment	Width of outlet A	Thickness of entrance B	Thickness of outlet C	Speed of roll casting sheet
Units	m	m	m	m/s
1	0.860	0.013	0.003	0.0167
2	0.860	0.016	0.005	0.0835
3	0.860	0.019	0.007	0.167
4	0.890	0.013	0.005	0.167
5	0.890	0.016	0.007	0.0167
6	0.890	0.019	0.003	0.0835
7	0.920	0.013	0.007	0.0835
8	0.920	0.016	0.003	0.167
9	0.920	0.019	0.005	0.0167

3.2 Finite Element Analysis of Each Experiment

Proceeding of coupled fluid-thermal analysis with ANSYS software is seen in reference [6,7].

3.3 Evaluation System

To judge how suitable the flow field of nozzle will be for producing with higher casting speed and thinner thickness of aluminum sheet, stricter analysis of result by software MATLAB is utilized to make comprehensive graphics of flow fluid's velocity and

temperature on midline of nozzle's outlet surface: two degrees of freedom of nodes in finite element analysis. Fig. 3 is an example of comprehensive graphics from experiment No. 4.

Based on industrial experience, on midline (y=0, z=0) of outlet of nozzle, the best distributing of velocity is a flat line which parallel to outlet; and that of temperature is also a flat line, but parts of two sides of midline will be a little higher temperature than mid part. So the shape and error of each curve of velocity and temperature of flow fluid of outlet are decisive of evaluation.

4 Result

According to 9 comprehensive graphics of results, first of all, compare every 3 results of the same speed. For example, at speed of 0.167 m/s, there are experiments of No. 3, 4, 8, apparently it is judged by above method, that No. 8 is better than other two, and No. 4 is better than No. 3. Secondly, compare results of other same factor with same level, to get other order of superiority. But finally it is realized as an original and rough judgment, which cannot supply accurate analysis between results which are of different levels and factors.

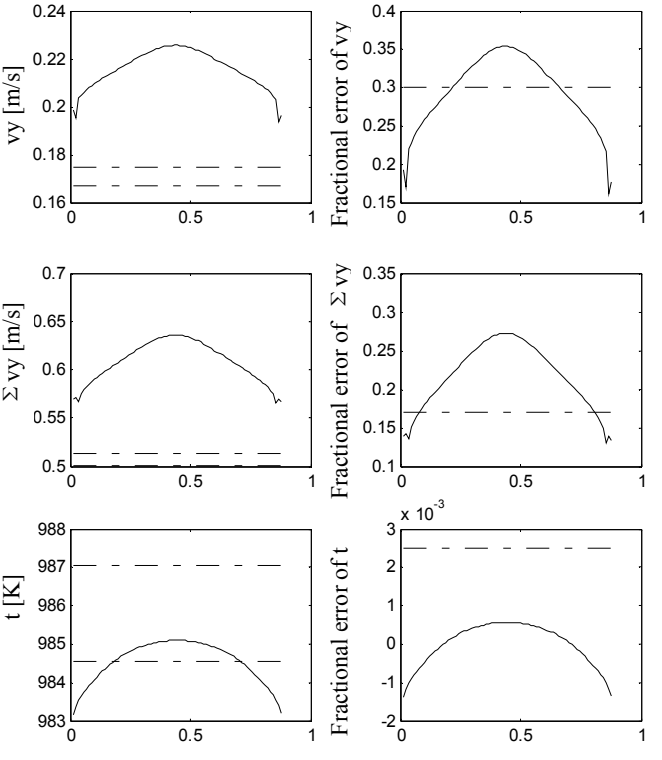


Fig. 3. Comprehensive graphic of *vy* & *t* at 0.0167 m/s

5 Conclusion

Orthogonal experiment design is used to make orthogonal array of 3 levels of 4 factors for simulation of nozzle, while three main structure dimensions and the speed of roll casting sheet are chosen as 4 factors. An optimal combination of factors and levels could be provided to achieve optimum of good distributing of velocity and temperature

of flow fluid in nozzle. Experiment No. 8 is one of better results, but to provide appropriate designs for investigating the main effects in order of priority, quantitative analysis of simulation result is needed and will be discussed later. Once the optimized result is decided, it can be used as the original condition of integrated design of nozzle and spacers.

Acknowledgments

This thesis is sponsored by the following funds: (1) The Natural Science Foundation of China (50875267 and 50605063); (2) Program for Doctoral Dissertation of P. R. China (201032) (NCET-10-0792); (3) Program for New Century Excellent Talents in University (NCET-10-0792); (4) Postdoctoral Science Foundation of Central South University.

References

1. Amit, S., Yogeshwar, S.: Modeling of fluid flow and heat transfer in twin-roll casting of aluminum alloys. Materials Transactions 43, 206–213 (2002)
2. Kang, C.G., Kim, Y.D.: Model experiments for the determination of the heat-transfer coefficient and transition thermal analysis in the direct rolling process. Journal of Materials Processing Technology 84, 210–224 (1998)
3. Li, Y., Hu, C.: Experiment design and data processing. Chemical Industry Press, Beijing (2005) (in Chinese)
4. Zhou, J.: Statistics Method for Quality Control, Chinese statistics publication, People's Republic of China (1999)
5. Zhou, Y.: Effect of velocity of Fluid Field of 3-spacer Nozzle in Roll-casting Models Using Coupled Fluid-thermal Finite Element Analysis. Applied Mechanics and Materials 29-32, 1481–1487 (2010)
6. Forum of Simulation We Use (in Chinese), http://www.simwe.com
7. Zhou, Y.: Physical-geometrical Coupling Rule and Design of Fluid Field of Aluminum Melt of Continuous Roll Casting, Ph.D. thesis. College of Mechanical and Electrical Engineering, Central South University, Changsha, P. R. China (2007) (in Chinese)

Quantitative Analysis of Orthogonal Experiment on Simulation of Fluid Field in Spacer-Free Nozzle in Aluminum Roll-Casting

Ying Zhou[1], Pei Feng[2], Huanlong Hu[2], Pengjian Li[2], and Cijun Shuai[2,*]

[1] School of Material Science and Engineering, Central South University, Changsha, 410083, Hunan Province, P.R. China
zhouying@mail.csu.edu.cn
[2] School of Mechanical and Electrical Engineering, Central South University, Changsha, 410083, Hunan Province, P.R. China
Fengpei851688@163.com, huhuanlong@126.com, 632404633@qq.com, shuai@mail.csu.edu.cn

Abstract. Conventional analysis of orthogonal experimental results was applied to optimize structure of spacer-free nozzle in aluminum roll-casting. Weighting factors of velocity and temperature were determined in analysis with extreme deviation. It was concluded that good distributing of flow fluid's velocity and temperature, would come with bigger width of outlet, modest thickness of entrance, and smaller thickness of outlet. The main effects in order of priority were investigated and advices on improving performance were given for industrial experiments.

Keywords: orthogonal experiment design; extreme deviation; spacer-free nozzle; roll cast; finite element analysis.

1 Introduction

To achieve high quality production of twin-roll casting with higher speed and thinner sheet [1, 2], the optimization design of flow fluid of nozzle should be attached great importance to. Some geometrical parameters of structure are decided by special producing conditions, such as width of aluminum sheet, but other geometrical parameters can vary within a certain range, which are suitable to design levels of orthogonal experiment design. At the beginning of optimization design of nozzle, geometrical sizes of space-free nozzle should be considered to optimize at first, then spacers added. The orthogonal experiment can find an optimal combination of levels of factors [3, 4] for finite element analysis of flow fluid of spacer-free nozzle. As discussed before [5], three geometry parameters of structure of nozzle and speed of roll-casting sheet are chosen as factors of orthogonal design.

Section 2 serves preliminary analysis of experimental result; Section 3 presents weighting factors for analyzing simulation results of orthogonal experiments; Section 4

* Corresponding author.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 663–667.
springerlink.com

is result of evaluating by extreme deviation; Section 5 draws conclusion and gives suggestion for improvement of orthogonal experiment.

2 Preliminary Analysis of Experimental Result

Comparing 3 cases with each same speed, sort these cases by order of superiority. For example, at speed of 0.0835m/s, there are three groups of No. 2, 6 and 7. It can be observed from Comprehensive graphic of *vy* & *t* that Group 6 (as shown in Fig. 1) is better than other two groups, and group 2 (as shown in Fig. 2) is better than group 7. But if analysis with more than one factor, it is hard to judge which one is better, and how much it is better than others. Quantitative analysis of orthogonal experiment has to be applied.

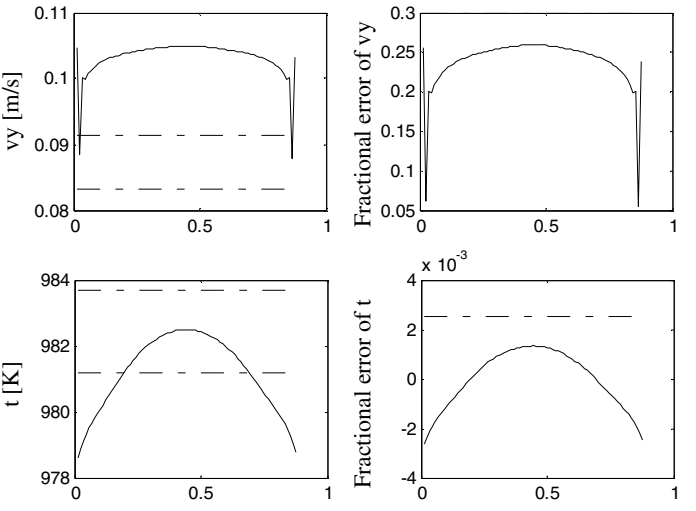


Fig. 1. Comprehensive graphic of *vy* & *t* of group No. 6

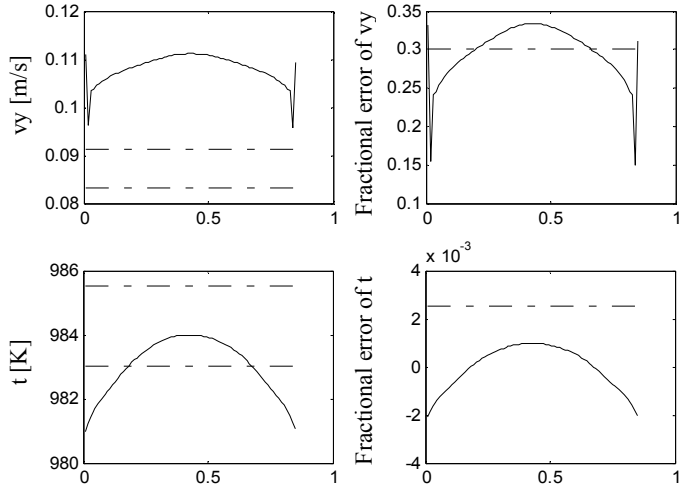


Fig. 2. Comprehensive graphic of *vy* & *t* of group No. 2

3 Weighting Factor for Quantitative Analysis

Measurement of each group's superiority is to adjust every possible weighting factor to

fit for the order of visual analysis by observer or analyst. Usually more than one weighting factor will be settled. From coupled fluid-thermal finite element analysis, there are two degrees of freedom are involved: velocity and temperature of flow fluid, which suggests that two weighting factors of velocity and temperature will be considered.

3.1 Weighting Factor of Velocity (*W*)

As 3 different kinds of speed of roll-casting sheet used in design, even a same model will show different distribution of velocity and temperature, so weighting factor of velocity (W) must be designed as a synthesized factor, which means W is the last one to be multiplied by.

3.2 Weighting Factor of Temperature (*Wt*)

With the same simulation model, the higher speed of roll-casting sheet increases, the more even distribution of temperature is. So weighting factor of temperature (*Wt*) only applies to item of temperature. By lots of comparing and calculating, weighting factors are chosen in Table. 1:

Table 1. Values of weighting factors

Velocity of roll-casting [m/s]	Weighting Factor of velocity: *W*	Weighting Factor of temperature: *Wt*
0.0167	1.0	1.4
0.0835	1.1	1.2
0.167	1.2	1.0

4 Result of Evaluating

Conventional analysis [6] by extreme deviation is as below in Table. 2:

ΔK: extreme deviation, is equal to maximum minus minimum of one row.

Values of *K1~K3* corresponding to Level 1~3 are calculated as below:

K_1^A=25+24.64+19.2=68.84

K_2^A=24+24.4+27.72=76.12

K_3^A=22.88+30+24.5=77.38

K_1^B=25+24+22.88=71.88

K_2^B=24.64+24.4+30=79.04

K_3^B=19.2+27.72+24.5=71.42

K_1^C=25+27.72+30=82.72

K_2^C=24.64+24+24.5=73.14

K_3^C=19.2+24.4+22.88=66.48

Table 2. Visual analysis of experimental result of simulation of nozzle

No.	Width of outlet	Thickness of entrance	Thickness of outlet	Speed of roll casting	Weighting factor of velocity	Grade on *vy*	Grade with weighting factor	Grade on *vy*t*	Final grade
					Evaluation system: $S=W*(Sv+Wt*St+Sq)$				
	A	*B*	*C*	*D*	*W*	*Sv*	*Wt*St*	*Sq*	*S*
1	860	13	3	0.0167	1.0	9	1.4*5	9	25
2	860	16	5	0.0835	1.1	7	1.2*7	7	24.64
3	860	19	7	0.167	1.2	4	1.0*8	4	19.2
4	890	13	5	0.167	1.2	6	1.0*8	6	24
5	890	16	7	0.0167	1.0	8	1.4*8	8	24.4
6	890	19	3	0.0835	1.1	9	1.2*6	9	27.72
7	920	13	7	0.0835	1.1	5	1.2*9	5	22.88
8	920	16	3	0.167	1.2	8	1.0*9	8	30
9	920	19	5	0.0167	1.0	9	1.4*5	8.5	24.5
K1	68.84	71.88	82.72						
K2	76.12	79.04	73.14						
K3	77.38	71.42	66.48						
ΔK	8.54	7.62	16.24						

The grade of effect on experiment result calculated by extreme deviation of each factor, results that sequence from high effect to lower effect, is in order of factor *C*, *A* and *B*. Optimization result in the orthogonal experiment is from group No. 8.

5 Conclusion

By the result of evaluation, it is concluded that thickness of outlet (*C*) is the key factor of all, width of outlet (*A*) is less important factor, and minor factor is thickness of entrance (*B*). To obtain good distributing of flow fluid's velocity and temperature, bigger width of outlet, modest thickness of entrance, and smaller thickness of outlet will be helpful.

Advices from industrial experiments:

Width of outlet (*A*) determined by machine of roll-casting and production, it cannot be adjusted casually, so it should be treated as an unchangeable parameter, and instead width of entrance can be designed as a factor.

In this experiment, speed of roll-casting sheet has not been treated as an independent parameter, but as a factor. Now it is explored that next design will be project of 3 factors with 3 levels, without speed of sheet, it will be more accurate and reasonable.

Acknowledgments

This thesis is sponsored by the following funds: (1) The Natural Science Foundation of China (50875267 and 50605063); (2) Program for Doctoral Dissertation of P. R. China (201032) (NCET-10-0792); (3) Program for New Century Excellent Talents in University (NCET-10-0792); (4) Postdoctoral Science Foundation of Central South University.

References

1. Amit, S., Yogeshwar, S.: Modeling of fluid flow and heat transfer in twin-roll casting of aluminum alloys. Materials Transactions 43, 206–213 (2002)
2. Kang, C.G., Kim, Y.D.: Model experiments for the determination of the heat-transfer coefficient and transition thermal analysis in the direct rolling process. Journal of Materials Processing Technology 84, 210–224 (1998)
3. Hillmann, M., Kubli, W.: Optimization of sheet metal forming processes using simulation programs. In: Proceedings the of NUMISHEET 1999, Besancon, France, pp. 287–292 (1999)
4. Zhou, J.: Statistics Method for Quality Control, Chinese statistics publication, People's Republic of China (1999)
5. Zhou, Y.: Physical-geometrical Coupling Rule and Design of Fluid Field of Aluminum Melt of Continuous Roll Casting, Ph.D. thesis. School of Mechanical and Electrical Engineering, Central South University, Changsha, P. R. China (2007) (in Chinese)
6. Li, Y., Hu, C.: Experiment design and data processing, pp. 83–98. Chemical Industry Press, Beijing (2005) (in Chinese)

An Optimization Scheme of Single-Spacer Nozzle of Aluminum Roll Casting Using Coupled Fluid-Thermal Finite Element Analysis

Ying Zhou[1], Pei Feng[2], Pengjian Li[2], Huanlong Hu[2], and Cijun Shuai[2,*]

[1] School of Material Science and Engineering, Central South University, Changsha, 410083, Hunan Province, P.R. China
zhouying@mail.csu.edu.cn
[2] School of Mechanical and Electrical Engineering, Central South University, Changsha, 410083, Hunan Province, P.R. China
Fengpei851688@163.com, 632404633@qq.com, huhuanlong@126.com, shuai@mail.csu.edu.cn

Abstract. To optimize shape of single-spacer nozzle, width of single spacer was chosen to be the only optimized parameter to discover its influence on distribution of velocity and temperature of nozzle's flow fluid. According to results of coupled fluid-thermal finite element analysis with four different widths of single spacer in nozzle, it was discovered that the greater width was, the more uneven of distribution of velocity and temperature of flow fluid at outlet of nozzle would be.

Keywords: optimization design; single-spacer nozzle; roll cast; coupled fluid-thermal; finite element analysis.

1 Introduction

Single-spacer nozzle used in aluminum roll casting is one of structures which had potential to obtain good flow fluid which matches the cooling capacity of casting rollers [1]. Theoretically an integrated scheme of optimization of single-spacer nozzle should be followed as these steps: 1. analysis spacer-free nozzle's geometrical parameters by orthogonal experiment; 2. determine shape and variables of single spacer, and analysis single-spacer nozzle; 3. analysis of combining variable of single spacer with geometrical parameters of nozzle [2, 3].

The optimization scheme has been designed for industry experiments, limited by the factory's technical conditions, many structure parameters of nozzle have been fixed, and it's why only one variable of single-spacer nozzle discussed in this paper.

* Corresponding author.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 669–674.
springerlink.com

Section 2 gives optimization scheme of single-spacer nozzle; section 3 displays boundary conditions applied to finite element analysis of nozzle; section 4 shows result of analysis; section 5 indicates conclusion.

2 Optimization Scheme of Single-Spacer Nozzle

Under limited conditions of industrial actual experiments, there are many fixed parameters of nozzle, and a few parameters of spacer can be designed as optimization parameters this time.

2.1 Modeling with Geometrical Parameters of Spacer-Free Nozzle

Spacer-free nozzle's main geometrical parameters are shown in Fig. 1 and Table. 1. First, since front-box is appointed, then size of nozzle's entrance is followed fixedly too. Furthermore, nozzles are custom-made, and thicknesses of nozzle's entrance and outlet are hard to adjust, so they have to be changeless. At last, geometrical parameter of nozzle is width of outlet, it is determined as constant. Finally all sizes of nozzle are fixed.

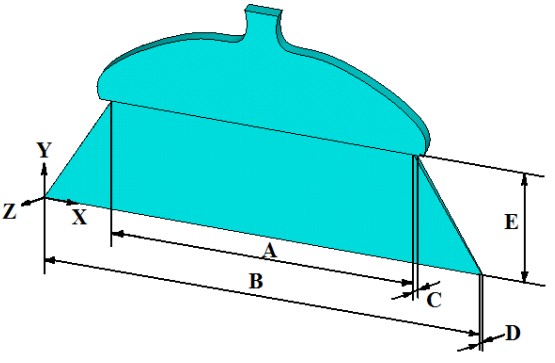


Fig. 1. Geometrical parameters of flow field of spacer-free nozzle with front-box

Table 1. Parameters of spacer-free nozzle for 3-d modeling

Name	Width of entrance *A*	Width of outlet *B*	Thickness of entrance *C*	Thickness of outlet *D*	Depth of nozzle *E*	Speed of roll-casting sheet
Units	m	m	m	m	m	m/s
Value	0.89	1.300	0.017	0.007	0.37	0.0835

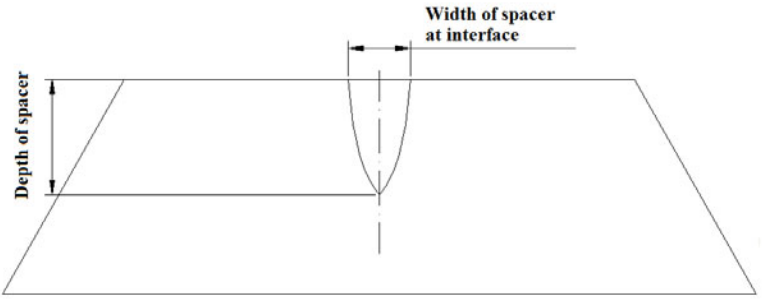


Fig. 2. Geometrical parameters of single spacer in nozzle

2.2 Variable of Single-Spacer Nozzle

Shape of the single spacer is designed as droplike curve created by spline curves which use fitting curves to link several key points. Coordinates of key points can be calculated by program with position, depth of spacer, and width of spacer at interface of front-box and nozzle (as shown in Fig. 2). Depth of spacer in this plan is 0.2 m selected by former results of both simulation and real experiments. Width of spacer at interface is the only variable in this optimization scheme, and values are listed in Table 2:

Table 2. Values' list of width of single spacer

Group No.	1	2	3	4
Width of spacer [m]	0.1	0.15	0.2	0.3

3 Finite Element Analysis of Nozzle

Mesh type and boundary conditions of coupled fluid-thermal FEA are shown in reference [3~5].

4 Result

Simulation of four groups brings about results in Fig. 3. Workplane of z=0 is the most convincing place where contour display of vy (y - axis component of velocity) & t (temperature) shows most details. [6]

Comprehensive graphics are drawn by MATLAB in Fig. 4, which come from nodes' two degrees of freedom: vy & t on midline (y=0, z=0).

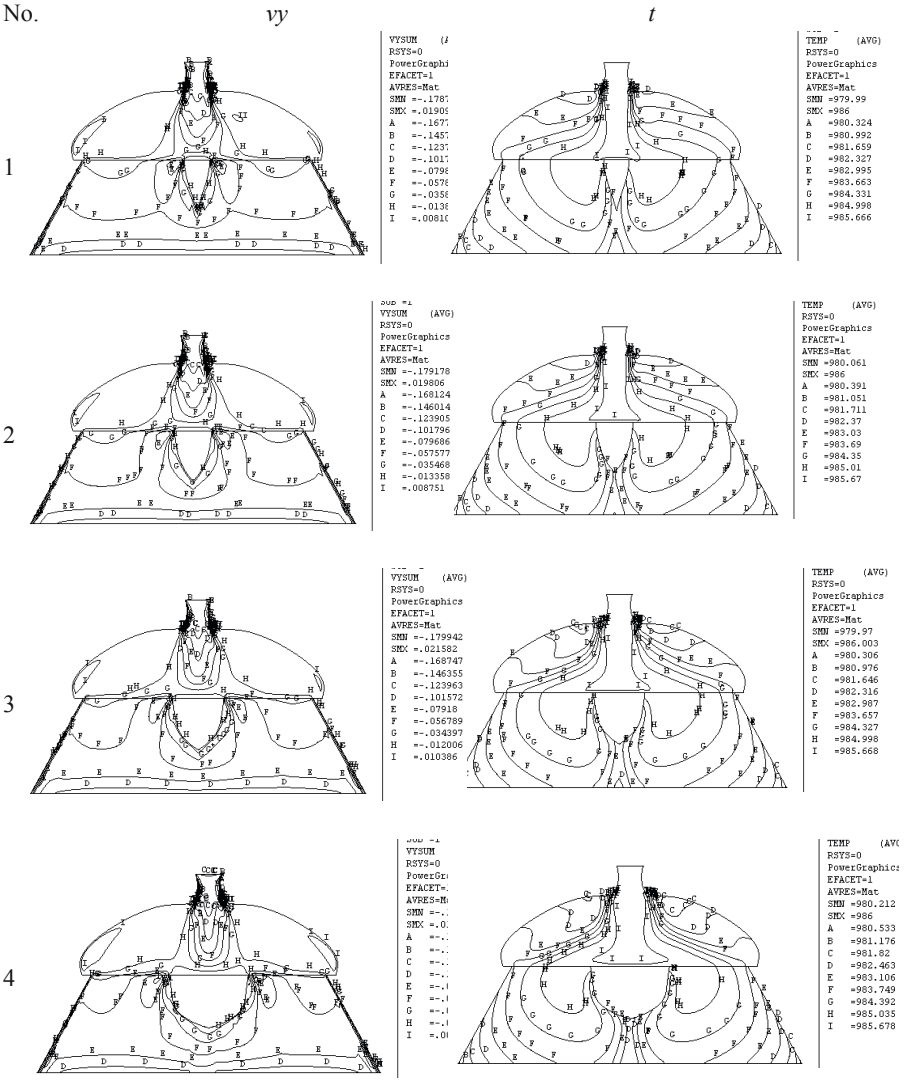


Fig. 3. Isosurface of contour display of *vy* & *t* on workplane (*z*=0) of four groups

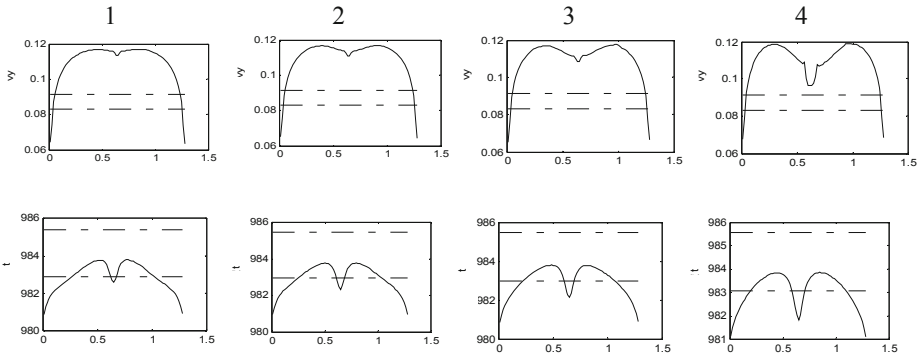


Fig. 4. Comprehensive graphics of *vy* & *t* on midline (*y*=0, *z*=0) of four groups

From Fig. 3 and Fig. 4, it shows that the greater width of single spacer is, the more and sharply sunken curves of *vy* & *t* around mid part of nozzle's outlet are. Especially when width is equal to 0.3 m, the sunken has rapid swing than others.

5 Conclusion

Though in this case, a lot of factors could not be set as optimized parameters, and only width of single spacer on interface of nozzle and front-box acts as variable, but it still can be discovered that smaller width of spacer will produce smoother distribution of velocity and temperature at nozzle's outlet. Width of spacer will also influence width of entrance besides spacer, here comes another explanation of why 0.3-m-width spacer has such sudden change of curves of *vy* & *t*, it is because area of entrance is much smaller than those of other groups, as volume of flow (which is approximately equal to area plus velocity) is constant, so flow's velocity through entrance must be bigger than others, and partially rising of velocity increases uneven of distribution of *vy* & *t*. Analysis of single-spacer nozzle can be reference for further optimization of 3C-spacer nozzle.

Acknowledgments

This thesis is sponsored by the following funds: (1) The Natural Science Foundation of China (50875267 and 50605063); (2) Program for Doctoral Dissertation of P. R. China (201032) (NCET-10-0792); (3) Program for New Century Excellent Talents in University (NCET-10-0792); (4) Postdoctoral Science Foundation of Central South University.

References

1. Zhou, Y., Huang, M.-h., Mao, D.-h., Liang, T.: 3-D Coupled Fluid-thermal Finite Element Analysis of 3C-style Nozzle's Fluid Field of Al Roll-casting. Materials Science Forum 546-549, 741–744 (2007)
2. Li, Y., Hu, C.: Experiment design and data processing, pp. 83–98. Chemical Industry Press, Beijing (2005) (in Chinese)
3. Zhou, Y.: Physical-geometrical Coupling Rule and Design of Fluid Field of Aluminum Melt of Continuous Roll Casting, Ph.D. thesis. Central South University, Changsha, P. R. China (2007) (in Chinese)
4. Kang, C.G., Kim, Y.D.: Model experiments for the determination of the heat-transfer coefficient and transition thermal analysis in the direct rolling process. Journal of Materials Processing Technology 84, 210–224 (1998)
5. Amit, S., Yogeshwar, S.: Modeling of fluid flow and heat transfer in twin-roll casting of aluminum alloys. Materials Transactions 43, 206–213 (2002)
6. Forum of Simulation We Use (in Chinese), `http://www.simwe.com`

Association Rules Algorithm in Bank Risk Assessment

Guorong Xiao

Department of Computer Science and Technology
GuangDong University of Finance Guangzhou, China
newducky@126.com

Abstract. Domestic financial enterprise data management applications generally brings together the vast amounts of data, but can not find the relationship and business rules exists in the data to do risk prediction assessment. Therefore, the domestic financial companies need to accelerate the pace of information technology in regions of integration of customer resources, business analysis and investment decisions. This paper analyzes the risk assessment approach of banks, mainly focuses on the analysis of association rules data mining in bank risk assessment, and discusses the working principle of improved association rules algorithm genetic algorithm in commercial bank risk assessment. We described the methods and processes of system application. We select the matrix form, only scan the database once, and use the method of selecting assumption frequent items and numbers, find the frequent item sets through high end item sets, minimize the number of candidate data sets, greatly improve the efficiency of the algorithm.

Keywords: bank; risk assessment; association rules; algorithm.

1 Introduction

With the extensive application of database and computer networks, demands on banking sector increased continuously. In the face of three key factors, change, competition and customer, which impact and determine the development of bank, if there is no information technology, it will become increasingly difficult for banks to understand, grasp and respond to change and increasingly difficult to integrate resources, planning restructuring to cope with competition, also difficult to achieve their business process reengineering and strategic tasks of intelligent marketing and management decisions. Comprehensive information management is the real core competitiveness of bank. Therefore, domestic commercial banks all consider the development of entering informatization as an important strategic move, established a relatively perfect system of financial informatization. Since the end of 2006, China opens the full domestic financial markets to WTO. Foreign banks are allowed to engage in RMB totally, the competition between domestic and foreign banks become more intensive. In this global financial crisis, although China's banking sector was not seriously impacted itself, the market turbulence remains intense, and China's banking sector also needs to seize the opportunity, make great efforts to promote the process of informatization of domestic banking sector.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 675–681.
springerlink.com

For credit card, here exist the issues of malicious overdraft and fraud which bring a great deal of risk to bank. Therefore, card issuers should take effective measures to prevent risks in advance, to quantify the credit rating to applicants' qualifications and credit, and then assist decision-makers to decide whether to give the credit card to the applicant. Usually bank judge the credit of applications through statistical techniques and experience, however, with rapid increase of credit card users and the trading volume, Experience alone is not enough to effectively make the right judgments, therefore, we need to introduce intelligent information processing technology to provide decision support to decision-makers. In this paper, we focus on the association rules of intelligent algorithm, discussed and researched upon risk assessment of banks.

2 Bank Risk Assessment Methods

Bank risk refers to the possibility of encountering economic loss during operating by various factors, or just the possibility for banks to meet assets and income loss. According to the cause of risk, the risks include credit risk, market risk, interest rate risk and legal risk, credit risk is the current key risk facing banking sector and also the main research of this paper. Since the 30's of 20th century, bank credit risk assessment method has mainly gone through 3 stages, judge according to experience, statistical analysis and artificial intelligence.

Mainly includes the following methods.

- Expert Judgment, in the initial phase of the credit rating, as the historical data information of trading partners is not enough; the level of trading partners' credit is entirely based on subjective experience of credit experts. This method is not efficient, costs high, and often have inconsistent conclusions.
- Scoring method, Banks and credit rating companies based on a pre-designed set of standardized indicators system, to rate each indicators of risk status about trading partners and customers, then average the rate according to the importance, and make the totally score the main judgment to customer risk rating. This method requires risk management experts to set indicator and importance according to their experience. The scoring of each indicator also need experts to use their experience and feelings, therefore, the level and experience of experts has a great impact on the effectiveness of ratings.
- Model approach, the method of risk rating system is based on trading partners or customers' historical database. Built the probability statistical model on historical data, including the discriminate analysis model, probability of default measurement model and loss given default rate measurement model. This method has the advantage of high efficiency; low cost, high accuracy measurement of default risk factors, inadequacies is difficult to directly enter the model of qualitative indicators, making it difficult to reflect the qualitative indicators of information.

3 Data Mining

With the rapid development of database technology and the widely application of management systems, people accumulated more and more data. Many important

information are hidden behind data, people want to have higher level of analysis to make better use the data.

The current database system can efficiently implement data entry, modification, statistics, query and other functions, but can not find the relationships and rules that exist among data, thus can not predict the future trends according to the current data. As currently lack of means to detect the knowledge behind the data mining, this led to the "data explosion but lack of knowledge" phenomenon. Therefore, intelligently and automatically valuable knowledge and information research among large amounts of data, know as data mining, is of great practical significance and wide application prospects.

From the technical point of view, Data mining is the process of the extraction of implicit, and unknown but potentially useful information and knowledge among a lot of, incomplete, noisy, fuzzy, random, real data. From the perspective of business applications, data mining is a new business information processing technology. Its main feature is a commercial database or data warehouse large amounts of data extraction, transformation, analysis and pattern processing, to extract the key of knowledge supporting business decisions, from a database or data warehouse model to automatically find related business.

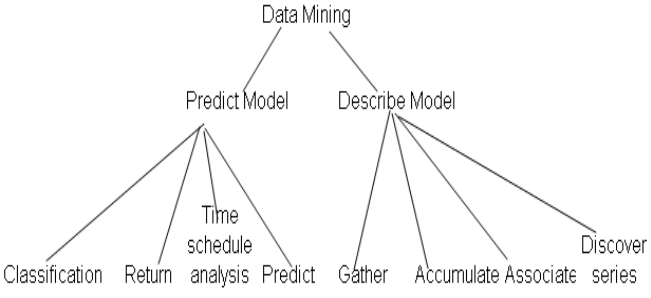


Fig. 1. Classification of data mining.

In short, data mining is actually a class of in-depth data analysis method. Data analysis itself has many years of history, only that the data collection and analysis in the past targeted at scientific research. But data mining is to do exploration and analysis on a large number of enterprise data in accordance with corporate business objectives. It's the effective ways of reveal the hidden, unknown or the verify regularity, and further the model.

The predictive model will predict the value of the data from results known from different data. It can accomplish data mining tasks, including classification. Map data to the predefined group or class. Return (the data item is mapped to a real predictor variable). Time series analysis, the behavior of determines the series according to the distance and structure. Forecast data value based on the historical event sequence diagram. Predict (based on past and current data to predict the future state of the data). Descriptive model identify the data in patterns or relationships. Different from the predictive models, descriptive models provide the method of data nature analysis, rather than predict the new nature. It usually consists of Gather (unsupervised learning or

partition), Accumulate (to be accompanied by a brief description of the data mapped to the subset), association rules (reveal the relationship between the data) and the series discovery (identifying data related to time between the sequence modes).

4 Association Rules in Bank Risk Assessment

At present, the association rule mining technology has been widely used in western financial industries and enterprises; it can successfully forecast demand for bank customers. Once got this information, banks can improve their marketing. The current banks are developing new methods of exploring new way of communication with customer every day. All banks bundled product information that may be interesting to customer in their ATM machines. To benefit customer who use their ATM and what to understand the products. If the database shows that a high credit customer change the address, the client is likely to buy a new bigger house, so there may need a higher credit limit, the higher end of the new credit card, or need a housing improvement loans, these products can be mailed to the customer through the credit card bill. When customers call to consult, the database can effectively assist telephone sales representative. Sales representative's computer screen can show the characteristics of clients, and what products the customer would be interested in.

4.1 Association Rules

Agrawal first proposed in 1993, equivalent to mining a database of customer transactions association rules between sets of items, and designed a basic algorithm, its core is based on the frequency of the recursive method of set theory, which is based on the frequency set of two-stage method of thinking , the design of association rules is decomposed into two sub-problems: ① found that the frequency set. This sub-issue is most important, the most expensive, therefore, focused on various algorithms to improve the efficiency of frequent item set discovery. ② According to the obtained frequent item sets to generate strong association rules. The algorithm uses the following two basic properties. The nature of a subset of any frequency band must be set. The nature of any non-frequent item sets 2 superset of a non-frequent item sets.

Apriori Algorithm is one of the classic data mining algorithms of association rules. One important step of Apriori is pruning, which matches every subset of the candidates with the frequent sets of previous layer, and then those infrequent sets will be removed according to one character of Apriori. This operation becomes the fact of time-cost of Apriori. A new algorithm named NPA (No Pruning Apriori) referred in this article is based on Apriori and, by means of modifying the JOIN operation; the pruning operation has been canceled. Such improvement enhances the speed of the algorithm and it brings practice application value in a certain degree. Apriori specific algorithm is as follows:

(1) L1 = {large 1-itemsets};
(2) for (k = 2; Lk-1 ≠; k + +) do begin
(3) Ck = apriori_gen (Lk-1);
(4) for all transactions t ∈ D do begin
(5) Ct = subset (Ck, t);

(6) for all candidates c ∈ Ct do
(7) C.count + +;
(8) End;
(9) Lk = {c ∈ Ck | c.count ≥ minsup}
(10) End;
(11) Answer = UkLk;

The basic idea of this algorithm is: first find all frequent sets, such as frequent item sets occur at least a predefined minimum support the same. Generated by the frequency of collection and strong association rules, these rules must satisfy minimum support and minimum confidence. Then use Step 1 to find the desired frequency set of rules generated, resulting in only a collection of items containing all the rules, in which the right side of each rule is only one, here is the rules used in the definition. Once these rules are generated, then only those greater than the minimum confidence given by the user was only to stay the rules. In order to generate all frequency sets, using the recursive method.

For many applications, the dispersion of the data distribution, it is difficult in the most detail level data find strong association rules. Although the rules drawn on a higher level may be general information, however, it is common for a user's information, but not necessarily so for another user. Therefore, data mining should provide a dig at multiple levels of functionality. Multi-level association rules mining are generally two ways: one is the single-level association rule mining algorithms directly applied to multi-level; the other is applied at different levels of different support threshold and confidence threshold. Existing multi-level association rule mining algorithm is mainly Improved Association Rules Algorithm.

4.2 Improved Association Rules in Bank Risk Assessment

Set each item I as an example, each transaction is a line in database D, together constructed incidence matrix M, M is equal to m * n, n is the total number of I, m for the database D contains the total number of Project Services. In the association matrix that contains the items that each firm, also contains a similar transaction.

Algorithm ideas, for the sum of each row of the matrix associated, the number of items calculated Affairs. Calculated to support the number of items removed is less than the number of sets to support the remaining items to our collection to find.

Input, transactional date T, the minimum support minsup count digital. Out put, the maximum frequency item set L.

- 1. C [n] = 0; / / C [n] n the maximum number of items
- 2. For each ti is Tdo {
- 3. I = | ti |
- 4. C [i] = C [i] +1
- 5.}
- 6. For i = n to 1 {
- 7. If (c [i] is greater than minsup) then {
- 8. K = i
- 9. Break

- 10.}
- 11.}
- 12. For i = k to 1 do {
- 13. Ck = {select k - item sets}
- 14. For each Ci is Ck do {
- 15. Lk = {Ci count Ci is greater than minsup}
- 16.}
- 17. If Lk is not equal to return Lk
- 18.}

From the point of example calculation, the use of matrix, only scan the database once, reducing the operation, and break the conventional approach using the first hypothesis, set out to find items from the high set, to minimize the number of data sets to improve work efficiency.

4.3 Improved Association Rules in Bank Risk Assessment

Select a month credit card records and customer information as credit card data, raw data, set the item sets C1, C2 ... Ck represent clients in T1, T2 ... Tk business consumer behavior.

First, scan personal credit card data, to obtain a matrix set, select the previous data sets.

Second, calculate the number of frequent item sets, the maximum was 7. Based on business experience, book the minimum support is 50.

According to Apriori algorithm, calculate the various supports. Assume that the most frequent item sets of 6, to obtain {C1, C2, C5, C8, C12, C13} support for the 67, {C1, C2, C5, C9, C12, C14} support for the 23, {C1, C2, C4, C10, C12, C13} support 82, which said that {C1, C2, C5, C8, C12, C13} and {C1, C2, C4, C10, C12, C13} for the maximum frequent item sets. Compared to the classic Apriori algorithm, the improved method only scans the database once, and improves work efficiency. Based on this data, you can engage in business promotion with the proposed business alliances {T1, T2, T5, T8, T12, T13} and {T1, T2, T4, T10, T12, T13}, and expand consumer groups, the maximum boost consumer spending, meaning a positive win-win business.

5 Conclusion

Domestic financial enterprise data management applications generally brings together the vast amounts of data, but can not find the relationship and business rules exists in the data to do risk prediction assessment. Therefore, the domestic financial companies need to accelerate the pace of information technology in regions of integration of customer resources, business analysis and investment decisions. This paper analyzes the risk assessment approach of banks, mainly focuses on the analysis of association rules data mining in bank risk assessment, and discusses the working principle of improved association rules algorithm genetic algorithm in commercial bank risk assessment.

We described the methods and processes of system application. We select the matrix form, only scan the database once, and use the method of selecting assumption

frequent items and numbers, find the frequent item sets through high end item sets, minimize the number of candidate data sets, greatly improve the efficiency of the algorithm.

References

[1] Fayyad, U.M., Piatetsky_Shapiro, G., Smyth, P., Uthurusamy: Advances in Knowledge Discovery and Data Mining (1996)
[2] Houtsma, M., Swami, A.: Set-oriented mining of associateion rules. IBM Almaden Research Center (1993)
[3] Keyun, H., Yu, C., Chunyi, S.: International Discovering Association Rules (1999)
[4] Agrawal, R., Yu, P.S.: Online generation of association rules. In: Proceedings of the 14th International Conference on Data Engineering, Orlando, Florida, USA (1993)
[5] Jiawei, H.: Data mining concept
[6] Btin, S.: Dynamic item counting and implication rules for market basket data (1997)
[7] Savasere, A.: An efficent algorithm for mining accosiation rules (1995)

A Wide Bandwidth Microstrip Array Antenna with Inverted L-Shaped Slots

Peng Chen[1,2], Xiao-dong Yang[1], Ying-song Li[1], and Yu Bai[1]

[1] College of Information and Communications Engineering,
Harbin Engineering University,
Harbin, Heilongjiang, 150001, P.R. China
[2] College of Information Engineering,
Jimei University, Xiamen, Fujian, 361021, P.R. China
{PengChen,XiaodongYang,Ying-songLi,YuBai}@hrbeu.edu.cn

Abstract. A wide bandwidth microstrip array antenna is proposed numerically and experimentally. By cutting an inverted slot in the patch and using the stacked structures, the impedance bandwidth of the presented antenna is broadened. In addition, the distance of elements of the array is properly settled and the coupling between the elements has been reduced effectively and the dimension of the array antenna is reduced by about 50%. Experimental and numerical result shows that the proposed array antenna, with compact size, has an impedance bandwidth range from 5.01GHz to 6.09GHz for voltage standing-wave ratio less than 2, which can meet the requirement of WLAN, WiMAX and C-band communication applications.

Keywords: microstrip array antenna; wideband antenna; L-shaped slot.

1 Introduction

With the increasing applications of the wideband wireless technologies, the WLAN, WiMAX and SAR systems have been in the spotlight worldwide. Various antennas with wide bandwidth have been proposed recent years. For the available designs, the printed monopole antennas reported in [1-5] have narrow bandwidth and lower gains. However, most of them were addressed to the needs of WLAN applications, and personal communication systems [5], very limited compact antenna designs have included the 5-6GHz for WLAN, WiMAX and SAR applications. Recently, a lot of wideband antennas have been investigated in [6-9]. The proposed technologies increased the impedance bandwidth of the antennas. However, the dimensions of the antennas are large and bandwidth is still narrow which limited its applications.

For these reasons, the bevel technology and parasitic elements method [10-11] are given to improve the bandwidth of the antennas. However, these antennas have larger size and lower gain. Therefore, the proposed antennas can't meet the demand of the wideband applications. In this paper, a wideband array antenna is investigated numerically and experimentally. In order to enhance the bandwidth of the array antenna, the stacked antenna is employed and an inverted L-shaped. The L-shaped slots are cut in

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 683–688.
springerlink.com

the different stacked patches. The distance of elements of the array is properly settled and the electromagnetic coupling between the elements is also reduced. What is more, the dimension of the array antenna is reduced by about 50%. Experimental and numerical result shows that the proposed array antenna, with compact size, has an impedance bandwidth range from 5.01GHz to 6.09GHz for voltage standing-wave ratio less than 2, which can meet the requirement of WLAN, WiMAX and SAR communication applications. The gains, the simulated and measured standing-wave ratio, radiation patterns are also given and discussed herein.

2 Antenna Design

Fig.1 illustrates the geometry and the configuration of the proposed array element o the proposed wide band antenna for WLAN, WiMAX and SAR applications. Three layers are used to construct and support the investigated wide band antenna. The antenna element consists of three substrates, two patches and two foams. The two resonance frequencies are so near that the bandwidth is enhanced. The sub1 used in the paper has a dielectric constant $\varepsilon = 2.2$ and the thickness is 1.5mm. The sub2 and the sub3 used in the paper have a dielectric constant $\varepsilon = 2.9$ and the thickness is 1.5mm. The foam between the patch1 and patch2 is Rohacell 71HF which has a dielectric constant $\varepsilon = 1.07$ with thickness is 5.3mm. In order to obtain a directional radiation pattern, a reflector is employed in the array element. The patch1 and patch2 are simple rectangular patches with L-shaped slot and inverted L-shaped slot. The ground hidden in the middle of the sub2 and sub3 can improve the radiation patterns effectively. A probe is implemented to feed the array element. The probe via the sub3 and sub2, connect to the patch2, and the patch1 is excited by the coupling of the patch1. The details parameters of the array element are listed in the table 1.

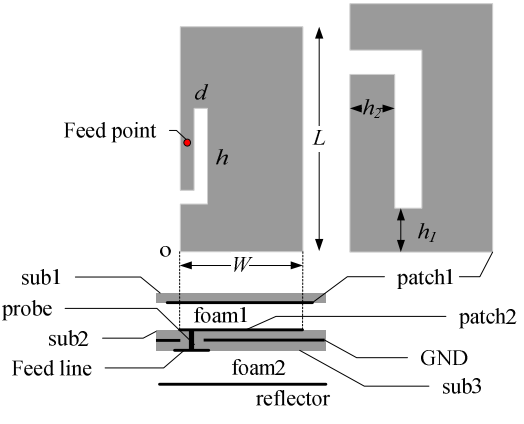


Fig. 1. Configuration of array element

Table 1. Structure size of array element

	W	*L*	*d*	*h*	h_1	h_2
Patch1	11	18.6	2	12.5	3.3	3.6
Patch2	8.5	16	0.9	6.6	3.4	1

The feed line and probe transmitted the energy to the main radiation structure. The patch2 is excited and the patch1 is also coupled. With the matching line, the return loss of the array antenna is improved very well. In this design, the patch2 coupled the energy to the patch1. By using the L-slot, the current path is changed and the path is also prolonged, which leaded to the increased bandwidth. The resonant frequency can be easy controlled by adjusting the dimension of the patch2 and patch1 and the L-shaped slot cut in the patches. In the design, the patch1 is the parasitic patch, the radiation patch (patch2) excites the parasitic element. Therefore, two resonance frequencies produced. The two resonance frequencies are so near that a wide band antenna is formed.

3 Results and Discussions

Based on the above discussions and analysis, an array antenna is designed, manufactured and tested. In order to meet the SAR applications, a 2 × 2 array is analyzed by using High Frequency Structure Simulator(HFSS), and the fabricated array is measured by using Anristu 37347D vector network analyzer and the radiation patterns are obtained in the Chamber. The photo-type of the array is shown in Fig. 2.

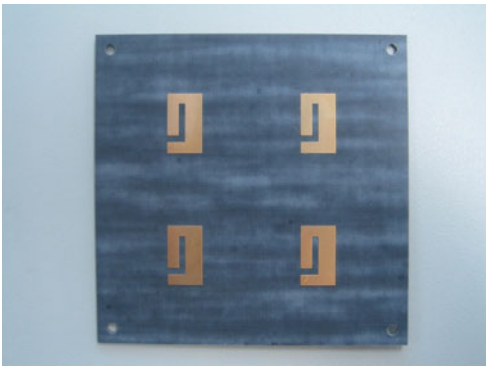


Fig. 2. Geometry of the array antenna

In the Fig.2, array spacing is about $0.75\lambda_0$. The simulated and measured VSWR (VSWR<2) is shown in Fig.3. It can be seen from the Fig.3, the simulated and measured results meet very well. The simulated impedance bandwidth is range from 4.9GHz to 5.8GHz. And the measured bandwidth covers 5.01GHz to 6.09GHz. The differences between the simulated and measured values may be due to the errors of

the manufactured antenna and the SMA connector to CPW-fed transition, which is included in the measurements but not taken into account in the calculated results.

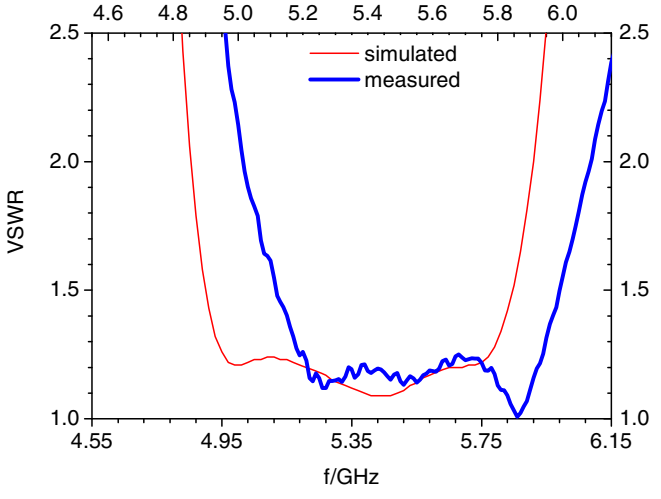
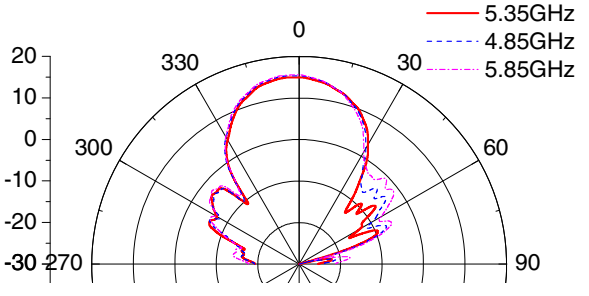
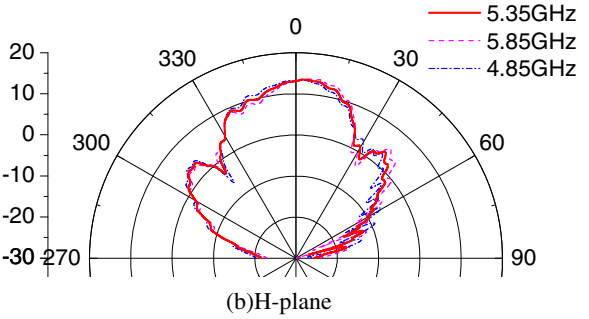


Fig. 3. Measured return loss of the array



(a) E-plane



(b)H-plane

Fig. 4. Radiation patterns of the array antenna

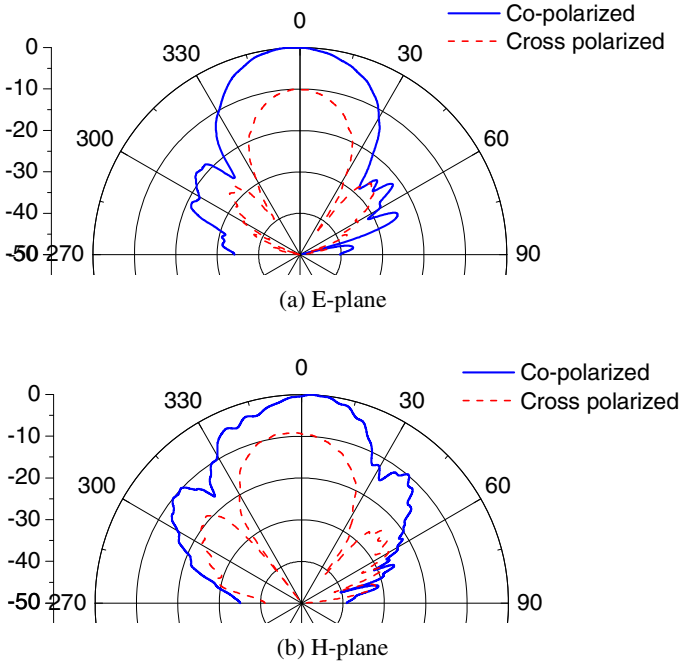


Fig. 5. Measured pattern at 5.35GHz

The radiation patterns of the array antenna at 4.85GHz, 5.35GHz and 5.85GHz are measured and shown in Fig.4. And the cross polarized is also tested and shown in Fig.5. The radiation pattern is in the broadside direction. And the cross polarized level is above 12dB. The E-plane and the H-plane at varying frequencies have the similar radiation patterns which meet the requirement of SAR applications. The radiation patterns in the other frequencies have the consistent radiation characteristics, which are not given herein. The maximum gain of the array is 14.76dBi.

4 Conclusion

A novel probe-fed stacked array antenna with L-shaped slot is realized in the paper. The array antenna is constructed by using multi-substrate and two patches. The proposed array shows wide bandwidth and consistent radiation patterns in the operation band. The simulated and measured results show that the array has higher gain and smaller size which is suitable for SAR applications.

Acknowledgments. This paper is funded by the International Exchange Program of Harbin Engineering University for Innovation-oriented Talents Cultivation.

References

1. Sabban, A.: A new broadband stacked two-layer microstrip antenna. In: Antennas and Propagation Society International Symposium. IEEE, London (1983)
2. Gao, S., Li, L., Leong, M., Yeo, T.: Dual-polarized slot-coupled planar antenna with wide bandwidth. IEEE Transactions Antennas Propagat. 51, 441–448 (2003)
3. Chakrabarty, S.B., Klefenz, F., Dreher, A.: Dual polarized wide-band stacked microstrip antenna with aperture coupling for SAR applications. In: Antennas and Propagation Society International Symposium. IEEE, New York (2000)
4. Gao, S., Sambell, A.: Dual-polarized broad-band microstrip antennas fed by proximity coupling. IEEE Transactions Antennas Propagat. 53, 526–530 (2005)
5. Deshmukh, A., Kumar, G.: Compact broadband U-slot-loaded rectangular microstrip antennas. Microwave and Optical Technology Letters 46, 556–559 (2005)
6. Jung, J.K., Choi, W.Y., Choi, J.: A small wideband microstrip-fed monopole antenna. IEEE Microwave and Wireless Components Letters 15, 703–705 (2005)
7. Kwak, S.I., Chang, K., Yoon, Y.J.: Small spiral antenna for wideband capsule endoscope system. Electronics Letters 42, 1328–1329 (2006)
8. Ammann, M.J.: Wideband antenna for mobile wireless terminals. Microwave and Optical Technology Letters 26, 360–362 (2000)
9. Lau, K.L., Luk, K.M., Lee, K.F.: Wideband U-slot microstrip patch antenna array. In: IEE Proceedings - Microwaves, Antennas and Propagation, vol. 148, pp. 41–44 (2001)
10. Ammann, M.J., Chen, Z.N.: A wide-band shorted planar monopole with bevel. IEEE Trans. Antennas Propag. 51, 901–903 (2003)
11. Chung, K., Yun, T., Choi, J.: Wideband CPW-fed monopole antenna with parasitic elements and slots. Electron. Lett. 40, 1038–1040 (2004)

Sonar Image Fusion Denoising Method Based on Multiple Morphological Wavelet Packets

Hong Shi[1], Chunhui Zhao[1], Zhengyan Shen[2], and Guangyu Liu[2]

[1] Information and Communication Engineering College, Harbin Engineering University, 150001 Harbin, China
[2] Underwater Acoustics Engineering College, Harbin Engineering University, 150001 Harbin, China
{ariel_0528,heu_lgy,etc_lgy}@yahoo.cn

Abstract. Aiming at the information loss, serious noisy and low resolution of sonar image, a sonar image fusion denoising method based on multiple morphological wavelet packets is proposed. Firstly, gave a morphological midpoint wavelet under the perfect reconstruction condition; Secondly, defined morphological wavelet packet to construct morphological Haar wavelet packet, morphological median wavelet packet and morphological midpoint wavelet packet for the noisy image decomposition and threshold processing; Finally, did fusion processing of low-frequency and high-frequency components separately according to certain fusion rules and got the final output image through the wavelet packet inverse transform. The simulation experiment result shows that the proposed method is more adapted to sonar image denoising than the single morphological wavelet denoising method.

Keywords: Sonar image denoising; Image fusion; Morphological wavelet; Wavelet packet.

1 Introduction

For single acoustics imaging instrument is usually unable to obtain satisfactory image, and the underwater environment is extremely complex, most of the obtained sonar images have serious noise pollution and low resolution, which undoubtedly brings great difficulties to the late detection, identification, tracking, etc. As an important branch of inter-discipline which relates to Data Fusion and Image processing, image Fusion aims at getting a more accurate, comprehensive and reliable image description through fusion rules by doing the information extraction, automatic analysis and optimal synthesis for multiple image which is the same object obtained in different situations (different observation time, different viewing angles, different sensors, etc). Image fusion has many outstanding advantages: on the one hand, it improves image resolution to increase the information credibility, and on the other hand, it reduces the requirements on a single image quality[1]. For fused image describes the integrated features of the object, it contains more abundant amount of information than the original image, therefore, from another point, image fusion can be consider as image enhancing. In addition, image fusion is not merely a synthesis of original images, but

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 689–696.
springerlink.com

a optimal synthesis that makes the synthesis image contain maximum useful information and discard useless information, so it is also do image denoising simultaneously [2]. This paper considered the application of image fusion technology in the sonar image denoising field, and gave a sonar image fusion denoising method based on multiple morphological wavelet packets. The experimental simulation result validates the feasibility and effectiveness of the proposed method.

2 Image Fusion Based on Wavelet Packet

Image fusion has three kinds: pixel-level image fusion, feature-level image fusion and decision-level image fusion [3]. The popular pixel-level image fusion methods are weighted average method, tower transform method, wavelet transform method and recent super wavelet transform (ridgelet transform[4], curvelet transform[5], contourlet transform[6], directionlet transform[7], etc). Image fusion based on wavelet packet transform is shown in Fig.1, and the steps are: firstly, do wavelet packet decomposition for two registration images separately; secondly, make data fusion of different frequency sub-bands on each decomposition level using different fusion rules; finally, take wavelet packet inverter transform, and the reconstruction image namely is the fusion result image.

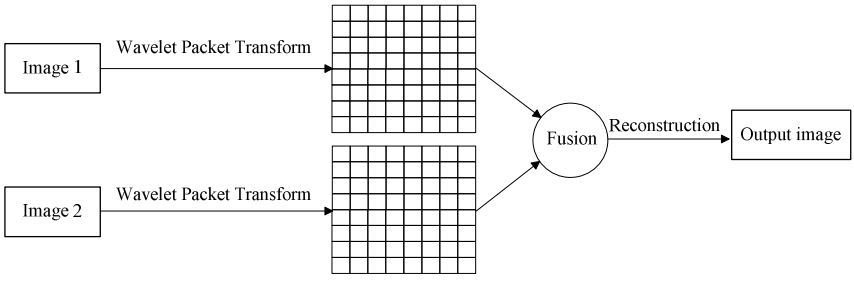


Fig. 1. Image fusion based on Wavelet packet transform (three levels).

3 Sonar Image Fusion Denoising Method Based on Multiple Morphological Wavelet Packets

3.1 Morphological Wavelet

Both the nonlinear property of morphological filter and the multi-resolution characteristic of wavelet transformation are taken account of in the morphological v wavelet algorithm, so it has higher research value than linear wavelet. It is proposed by Heijmans and Goutsias, and they also gave the perfect reconstruction condition when constructing morphological wavelet [8]. The perfect reconstruction condition for the uncoupled Morphological Wavelet is as follows:

$$\varphi_j^{\downarrow}\varphi_j^{\uparrow}(x) \dotplus \omega_j^{\downarrow}\omega_j^{\uparrow}(x) = x, \text{ if } x \in \Phi_j . \tag{1}$$

$$\varphi_j^{\uparrow}(\varphi_j^{\downarrow}(x) \dotplus \omega_j^{\downarrow}(y)) = x, \text{ if } x \in \Phi_{j+1},\ y \in W_{j+1} . \tag{2}$$

$$\omega_j^{\uparrow}((\varphi_j^{\downarrow} \dotplus \omega_j^{\downarrow}(y))) = y, \text{ if } x \in \Phi_{j+1},\ y \in W_{j+1}. \tag{3}$$

where Φ_j is the signal space, $\varphi^{\uparrow} : \Phi_j \to \Phi_{j+1}$ and $\varphi^{\downarrow} : \Phi_{j+1} \to \Phi_j$ is the signal analysis and synthesis operators respectively; W_j is the detailed space, $\omega^{\uparrow} : W_j \to W_{j+1}$ and $\omega^{\downarrow} : W_{j+1} \to W_j$ is the detailed analysis and synthesis operators respectively.

A number of new morphological wavelet can be constructed in this framework, such as the existed morphological Haar wavelet (MHW) in [8] and morphological median wavelet (MMedW) in [9]. In this paper, we will construct a morphological midpoint wavelet (MMW).

Firstly, make the midpoint filter as the low-pass filter of morphological wavelet, namely the signal analysis operator:

$$\begin{aligned}\varphi^{\uparrow}(x)(n) = 0.5\times[&x(2n) \wedge x(2n^+) \wedge x(2n_+) \wedge x(2n_+^+) + \\ &x(2n) \vee x(2n^+) \vee x(2n_+) \vee x(2n_+^+)].\end{aligned} \tag{4}$$

where the signal " $\wedge$ " and " $\vee$ " means the minimum and maximum operation respectively.

Secondly, define the detailed analysis operators as the difference of the two pixels in horizontal, vertical, diagonal directions as follows:

$$\omega^{\uparrow}(x)(n) = (\omega_v(x)(n), \omega_h(x)(n), \omega_d(x)(n)), \tag{5}$$

$$y_v(n) = \omega_v(x)(n) = x(2n) - x(2n^+), \tag{6}$$

$$y_h(n) = \omega_h(x)(n) = x(2n) - x(2n_+), \tag{7}$$

$$y_d(n) = \omega_d(x)(n) = x(2n) - x(2n_+^+). \tag{8}$$

Thirdly, according to the perfect reconstruction condition (1)-(3), synthesis operators can be obtained as follows:

$$\varphi^{\downarrow}(x)(2n) = \varphi^{\downarrow}(x)(2n^+) = \varphi^{\downarrow}(x)(2n_+) = \varphi^{\downarrow}(x)(2n_+^+) = x(n), \tag{9}$$

$$\begin{aligned}\omega^{\downarrow}(y)(2n) = 0.5\times[&y_v(n) \vee y_h(n) \vee y_d(n) \vee 0 + \\ &y_v(n) \wedge y_h(n) \wedge y_d(n) \wedge 0] = D,\end{aligned} \tag{10}$$

$$\omega^{\downarrow}(y)(2n^+) = D - y_v(n), \tag{11}$$

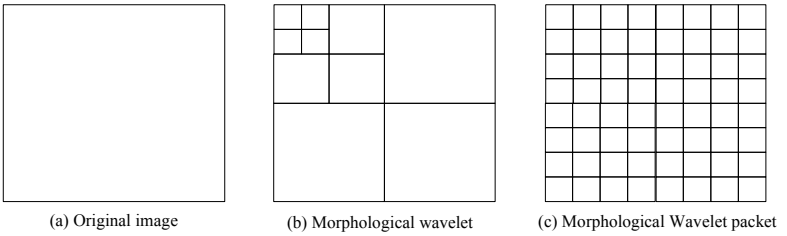
$$\omega^{\downarrow}(y)(2n_{+}) = D - y_{h}(n), \tag{12}$$

$$\omega^{\downarrow}(y)(2n_{+}^{+}) = D - y_{d}(n). \tag{13}$$

Then morphological midpoint wavelet based on the perfect reconstruction condition is constructed.

3.2 Morphological Wavelet Packet

In 1992, Coifman, Meyer and Wickerhauser gave the concept of wavelet packet [10], while in this paper, we give the concept of morphological wavelet packet. Morphological wavelet transform contains signal analysis operator and detail analysis operator, which respectively corresponding to low-pass filter and high-pass filter. The image is decomposed into 4 sub-bands by each level of morphological wavelet transform: a low-frequency sub-band (signal component) and three different directions of the high frequency sub-band (detail component). The low-frequency sub-band will continue to do multiresolution decomposition at the next level, but the high-frequency sub-bands are no longer decomposed. Similar to wavelet packet transform, in morphological wavelet packet transform, both the low-frequency sub-band and the high frequency sub-band of each layer after decomposition will do the multiresolution decomposition at the next level. Morphological wavelet decomposition and morphological wavelet packet decomposition is shown in Fig.2, by morphological wavelet packet, image can be decomposed into 4 sub-bands at the first level, 16 sub-bands at the second level, 64 sub-bands at the third level, and so forth, 4^m sub-bands at the m level.



(a) Original image (b) Morphological wavelet decomposition (c) Morphological Wavelet packet decomposition

Fig. 2. Morphological wavelet and morphological wavelet packet decomposition (three levels).

Image after morphological wavelet fusion, the low-frequency information can be well represented, while the high frequency information is lost a lot, which is not satisfactory fusion result for the image contains large amount detail information [11]. So, in this paper, we consider the more extensive and accurate image fusion method based on morphological wavelet packet. Morphological wavelet packet has the advantages

of both morphological wavelet and wavelet packet, such as nonlinearity of morphological filters, multi-resolution of wavelet and comprehensiveness of wavelet packet. According to the existing morphological Haar wavelet, morphological median wavelet and morphological midpoint wavelet proposed in the paper, morphological Haar wavelet packet, morphological median wavelet packet and morphological midpoint wavelet packet can be constructed respectively.

3.3 Sonar Image Fusion Denoising System

Make the above three wavelet packets for sonar image fusion denoising system, the flow chart is shown in Fig.3, and the specific steps are as follows:

(1) Do morphological Haar wavelet packet, morphological median wavelet packet and morphological midpoint wavelet packet decomposition respectively for the noisy image to obtain multi-resolution high-frequency and low-frequency sub-bands;

(2) Do threshold processing for morphological wavelet packet coefficients of high-frequency sub-bands as image denoising method based on wavelet transform;

(3) According to appropriate fusion rules, do data fusion for morphological wavelet packet coefficients of high-frequency sub-bands and low-frequency sub-bands respectively, then the new morphological wavelet packet coefficients of high frequency sub-band and low frequency sub-band can be obtained;

(4) Do morphological wavelet packet reconstruction, and right now the output image is the result image using fusion denoising method based on multiply morphological wavelet packets.

Among them, the fusion rule has a great impact on image fusion result. Because low-frequency component after image decomposition represents the image approximate part, this paper chooses mean algorithm as the fusion rule, make the average value of low-frequency coefficients as the fusion coefficient; for high-frequency component, it reflects the image detail part, so we selects the maximum (MS) fusion rules, which choosing the modulus maxima of corresponding pixels as the fusion coefficients.

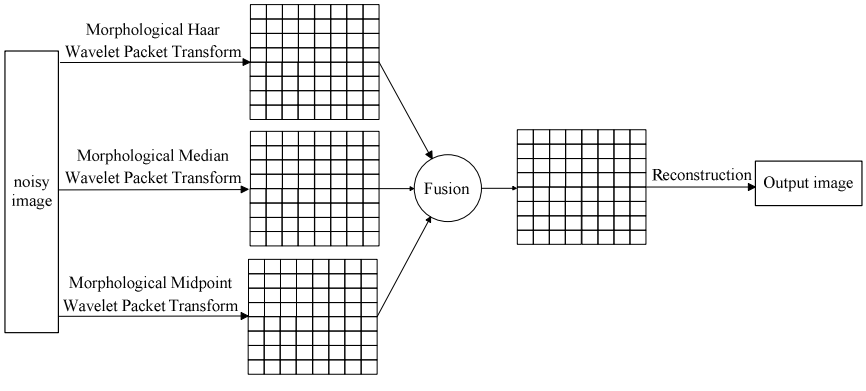


Fig. 3. Image fusion denoising method using three different morphological wavelet packets (three levels).

The low-pass filter of morphological Haar wavelet packet is the maximum or minimum filter, according to the characteristics that maximum and minimum filters can remove the "pepper" noise and "salt" noise, morphological Haar wavelet packet can be inferred that the denoising performance for the "pepper" noise and " salt "noise is more pronounced; by analogy, morphological median wavelet packet, using median filter as low-pass filter, is more effective for the single-stage or bipolar impulse noise remove; morphological midpoint wavelet packet, constructed by midpoint filter in this paper, has better performance on Gaussian and uniform random noise denoising system. The fusing denoising method based on the three morphological wavelet packets above combines the characteristic of them, so it has all their advantages and is more applicable to image denoising of the image with complex noise. For the noise sources of sonar image is extremely complex, such as instrument itself noise, underwater environment and so on, denoising use a single morphological wavelet packet may cannot get a satisfactory sonar image. So, in this paper, we combine three morphological wavelet packets together to obtain a more comprehensive and detailed denoising result.

4 Simulation Experiment

In the simulation experiment, we selected a 256×256 sonar image as original image (OI), and captained noisy image (NI) by adding Gaussian noise whose noise standard deviation is 0.1. Then we did the compare simulation experiment of the fusion denoising method in this paper, the single morphological Haar wavelet (MHW) denoising method in literature [8], the single morphological median wavelet (MMedW) denoising method in literature [9] and the single morphological midpoint wavelet (MMW) denoising method in our paper to verify the feasibility and effectiveness of the proposed method. For morphological wavelet packet decomposition, although more decomposition levels can bring more extensive details to the fusion image, but with the number of decomposition level increases, the number of sub-band will increase exponentially at the same time, which results in large computation, so we set the decomposition level to 3 in the experiment. The simulation result is shown in Fig.4.

Noise ratio (SNR), peak signal to noise ratio (PSNR) and mean square error (MSE), as the current performance indicators, were taken in the following to concretely evaluate the various denoising methods above. For image denoising, the higher SNR and PSNR, the lower MSE, means the better effect of the system. Table 1 shows the statistical properties of different methods in our experiment.

It can be seen in Fig.4 and Table 1, comparing with single morphological wavelet denoising method, our method has better denoising effect. Table 1 shows that the indicators of fusing denoising method are all improved, this is because that the fusion denoising method makes a more comprehensive denoising by combine the characteristics of the various morphological wavelet. In addition, it also can be seen from Fig.4 that the proposed method not only have better performance in denoising, but also reserve the detail of the image well, the reason is that taking morphological wavelet packet instead of morphological wavelet can makes the decomposition more detailed and comprehensive, so the fusion image retains the edge and other detail information. To sum up, the proposed method is more suitable for sonar image denoising.

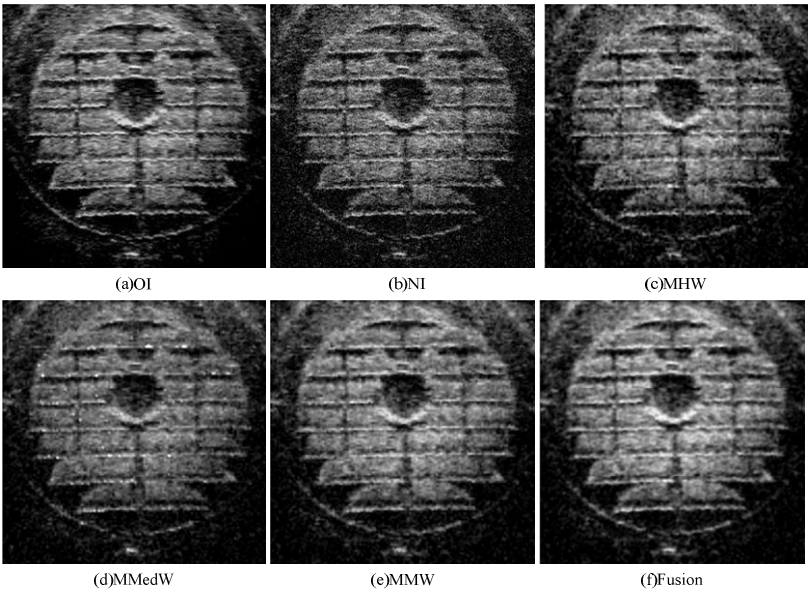


Fig. 4. The simulation results of sonar images denoising using different methods.

Table 1. The statistical properties of sonar images denoising using different method(dB).

	NI	MHW	MMeW	MMW	Fusion
SNR	5.6036	7.1778	8.1896	9.7436	9.9864
PSNR	19.9734	21.5476	22.5594	24.1134	24.8316
MSE	0.0101	0.0070	0.0055	0.0039	0.0030

5 Conclusion

In this paper, we combined nonlinearity of morphological filters, multi-resolution of wavelet and comprehensiveness of wavelet packet together to give the definition of morphological wavelet packet, and constructed a sonar image fusion denoising system based on multiple morphological wavelet packets. The simulation experimental result shows that the proposed method has better performance than the single morphological wavelet denoising method, and the edge preserving effect is also superior, so it is more adapted to the denoising of sonar image with serious noise pollution.

Acknowledgement

The project is sponsored by Natural Science Foundation of China (No. 50909025 and No. 61077079). This paper is funded by the International Exchange Program of Harbin Engineering University for Innovation-oriented Talents Cultivation.

References

1. Pajares, G., de la Cruz, J.M.: A Wavelet-based Image Fusion Tutorial. Pattern Recognition 37, 1855–1872 (2004)
2. Wenjie, L.: The Study on De-noising Method by Image Fusion Based on Wavelet Packet Analysis. Journal of Hebei Polytechnic University (Natural Science Edition) 32, 42–44 (2010)
3. Cofer, R.H., Kozaitis, S.P.: Pixel-Registered Image Fusion. Pro. SPIE 2484, 650–656 (1995)
4. Liu K., Guo L, Chang W.: Algorithm of Image Fusion based on Finite Ridgelet Transform. In: Proceedings of SPIE-The International Society for Optical Engineering. MIPPR: Automatic Target Recognition and Image Analysis; and Multispectral Image Acquisition, Vol. 6786 (2007)
5. Li, D., Duan, Z.: New Method Based on Curvelet Transform for Image Denoising. In: International Conference on MT&MA, Hangsha, pp. 760–763 (2010)
6. Yang, S.Y., Wang, M., Jiao, L.C., Wua, R.X., Wang, Z.X.: Image Fusion based on a new Contourlet Packet. Information Fusion 11, 78–84 (2010)
7. Xiang, L., Yueshun, H., Xuan, Z.: Edge-enhancement EPMA image fusion based on Directionlet transform. In: The Third International Symposium on Intelligent Information Technology Application, Nanchang, pp. 180–183 (2009)
8. Heijmans, H.J.A.M., Goutsias, J.: Nonlinear Multiresolution Signal Decomposition Schemes- Part II: Morphological Wavelets. IEEE Tans. Image Processing 9, 1897–1913 (2000)
9. Qingyun, D., Yinglin, Y.: A line Feature Extraction Method for Online Palmprint Images Based on Morphological Median Wavelet. Chinese Journal of Computers 26, 234–239 (2003)
10. Coifman, R.R., Meyer, Y., Wickerhauser, M.V.: Size Properties of Walvelet Packets. In: Ruskai, M.B., et al. (eds.) Wavelet and Their Applications, pp. 453–470. Jones and Bartlett (1992)
11. Ting, W., Jie, L., Ming, Z.: Application of Image Fusion Technology based on the Wavelet Packet Transform. Journal of Tongji Univesity(Natural science) 34, 1137–1141 (2006)

Research and Design of Fiber Grating Temperature Sensing System Based on Interferometric Demodulation Technique

Hong Zhang

School of Computer and Communication Engineering
Weifang University
Weifang 261061, China
zhanghongwf@sina.com

Abstract. Aimed at the deficiency that non-equilibrium M-Z interferometric demodulation method only applies to dynamic demodulation systems and according to the characteristic that there is a linear relationship or near-linear relationship between wavelength shift and demodulation parameters (light intensity or transmission rate, etc) caused by the change of measured field physical quantity, this paper designs a kind of fiber Bragg grating linear demodulation sensor which demodulates Bragg wavelength by linear demodulation method and it realizes the measurement of quasi-static parameters. Apply this design scheme to the design of fiber grating temperature sensing system and experiments show that the system finishes research requirements according to schedule.

Keywords: fiber Bragg grating; fiber grating sensor; M-Z interferometer; wavelength demodulation.

1 Introduction

Fiber grating sensor is a kind of optical fiber sensor and the sensing process based on fiber grating is to get sensing information by external physical parameters modulating fiber Bragg wavelength and therefore, it is a kind of wavelength modulation fiber sensor. Because there is a natural compatibility between fiber grating and fiber, quasi-distributed sensing can be realized easily and the sensing signals of fiber grating itself are wavelength modulation, measuring signals are unaffected by light source fluctuation, fiber bending loss, light source power fluctuation and system loss, the use of fiber grating in the sensing field has caused wide attention and great interest among related scholars all over the world [1].

When the temperature or stress on a fiber grating sensor changes, Bragg reflection wavelength shifts and the quantity to be measured can be judged according to the shift, and therefore people put forward many demodulation methods, such as filtering method, interferometry and adjustable light source scanning method, etc. Compared with other demodulation techniques, interferometer demodulation method has extremely high detection sensitivity and many foreign researchers have studied the method [2, 3]. Combined with the working principles and characteristics of

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 697–703.
springerlink.com

non-equilibrium M-Z fiber interferometer demodulation method, this paper demodulates Bragg wavelength by linear demodulation method, designs a kind of fiber grating linear demodulation device, establishes a fiber grating temperature sensor experimental system and makes an experimental result analysis of the system.

2 Signal Demodulation Technique of Fiber Grating Sensor

2.1 Requirements for Accuracy

The theoretical analysis and research of fiber Bragg grating show that the temperature and strain sensitivity of fiber Bragg grating are very small. When the wavelength of grating is 1500nm, typical temperature and strain sensitivity are 0.011nm/^{o}C and 0.012nm/με respectively. To reach the measurement accuracies of 1^{o}C and 10με, the measurement accuracy of shifting center wavelength by $\triangle\lambda$ should be superior to the magnitude of 0.01nm. And therefore, the detection accuracy of $\triangle\lambda$ directly limits the detection accuracy of the whole system. The detection technique of $\triangle\lambda$ becomes one of key techniques in fiber grating sensing.

2.2 Interferometer Demodulation Method

At present, there are three kinds of different interference measurement structures adopted in fiber grating sensors, namely Michelson structure, Sagnac structure and Mach-Zehnde structure, among which Mach-Zehnde（M-Z）interferometer can realize wide-bandwidth and high-resolution demodulation ability and its cost is low, so this paper carries out scheme design by M-Z demodulation method.

M-Z interferometer demodulation technique mainly has three different implementation methods, namely non-equilibrium M-Z interferometric demodulation method, external modulation M-Z interferometric demodulation method and interferometric demodulation method based on 3×3 coupler [4].

Non-equilibrium M-Z fiber interferometer demodulation method has advantages of wide bandwidth and high resolution, etc, but the influence of random phase shift decides that this scheme only applies to dynamic demodulation systems [5], unsuitable for quasi-static detection. For this reason, this paper puts forward an improved scheme which can be used for quasi-static detection and the improved scheme is to add a reference grating to the non-equilibrium M-Z interferometer, which is equivalent to adding a modulating frequency whose size is w to the received signal. The reflection signals of sensing grating and reference grating are processed by a phase meter after passing a band-pass filter, which can eliminate the interference of random phase difference and make it apply to the linear measurement of quasi-static strain.

2.3 Analysis of Linear Demodulation Principle Based on M-Z Interferometer

The fiber Bragg grating sensor designed in this paper demodulates Bragg wavelength by linear demodulation method which is a kind of wavelength shift demodulation technique put forward for solving field practicality. The starting point of adopting this scheme is based on that there is a linear relationship or near-linear relationship between wavelength shift and demodulation parameters (light intensity or transmission

rate, etc) caused by the change of measured field physical quantity [6] and the scheme is suited to the measurement of quasi-static parameters. M-Z interferometer is composed of two 3dB fused-taper couplers, shown as Fig 1.

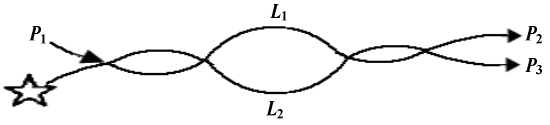


Fig. 1. Schematic Diagram of M-Z Interferometer

When satisfying the interference condition $L_1 - L_2 \leq \lambda_2 / \triangle\lambda$, M-Z interferometer has a comb filtering characteristic, at the rising or falling edge of output spectrum, the power ratio of two output arms changes very quickly with wavelength change, a filter with steep edges can be got by adjusting the length difference of the two arms of M-Z interferometer, high-accuracy wavelength shift detection can be carried out by using the characteristic, it is the linear demodulation principle based on M-Z interferometer[7].

3 Design of Sensing Demodulation Experimental System

3.1 Composition of Demodulation System

The demodulation system designed in this scheme is mainly composed of the following parts, namely light source, fiber coupler, fiber Bragg grating, spectrometer, M-Z interferometer, photodetector, conversion circuit, operation processing circuit and A/D converter.

3.2 Composition of System Hardware

Optica path

Selection of light source. The characteristics of light source decide if a fiber system can reach expected indexes. In this topic, the light-emitting device which is used as light source should meet the following conditions:

(1) Small volume, the light-emitting area should be matched with the size of fiber core diameter and there should be a high coupling efficiency between light source and fiber.

(2) Emission wavelengths should be suitable for two low-loss wave bands of fiber, namely short wavelengths should be between 0.8 and 0.9μm and long wavelengths should be between 1.2 and 1.6μm.

(3) It can carry out light intensity modulation directly and its connection with the modulator should be very convenient.

The light-emitting devices which are often used for fiber sensing are semiconductor laser (LD) and light emitting diode (LED). This scheme adopts a light emitting diode as the light source of the system [8].

Selection of coupler. The fiber coupler adopted in this topic is a kind of all-fiber directional coupler and its main characteristics are as follows:

(1) Its body is optical fiber, excluding other optical elements.
(2) It realizes light coupling function by the coupling effect of transmission mode in fiber.
(3) The direction of optical signal transmission is fixed.

Bragg grating. Sensing grating plays an important role in fiber grating sensing and we mainly consider the following aspects:

(1) Working waveband of sensing grating.
(2) Reflection bandwidth, reflectivity and length of sensing grating.
(3) Side mode suppression.
(4) Wavelength interval and buffer area of sensing grating (for sensor network).

The Bragg grating used in this system has a center wavelength of 1543.357nm and a reflection bandwidth of 0.3nm, its reflectivity is greater than 90%, its length is 10nm and its side mode suppression ratio should be higher than 15dB.

Spectrometer. The spectrometer used in this experiment is AQ6317C spectrum analyzer which is used to monitor the reflection wavelength of fiber Bragg grating and the shift value of reflection wavelength with temperature change. In the experiment, wavelength resolution Res = 0.5 nm and spectrometer scan number AVG = 10.

Interferometer

Making of interferometer. All-fiber M-Z interferometer is made by welding two 3dB couplers on two fibers continuously and the length difference between its two arms is about 0.5mm.

Noise of interferometer. The noise of interferometer is the main noise source this system considers and it directly influences measurement results.

$$\zeta = \frac{P_2 - P_3}{P_2 + P_3} = \cos\varphi = \cos(\frac{2\pi nd}{\lambda}) \tag{1}$$

It can be known from formula (1) that $\Delta\zeta = -\sin\varphi \cdot \Delta\varphi$, the phase change of interferometer will lead to a significant change of system output. When the arm length difference d and fiber core refractive index n of interferometer change with the change of external environment, the phase output of interferometer can be expressed as:

$$\Delta\varphi = \frac{2\pi}{\lambda} n\Delta d + \frac{2\pi}{\lambda} d\Delta n - \frac{2\pi}{\lambda^2} dn\Delta\lambda \tag{2}$$

In formula (2), $\triangle d$ is the change of arm length difference caused by environmental factors; $\triangle n$ is the change of refractive index of two arms of interferometer caused by environmental factors. It can be seen from (2) that $\triangle d$, $\triangle$n and $\triangle\lambda$ can all modulate phase to give rise to interference noise. In this way, noise can be divided into

stress-strain effect noise, thermal expansion and cold contraction effect noise and thermodynamic intrinsic phase noise. Environmental mechanical vibration and sound field are main factors causing stress-strain effect. The modulation of phase by temperature change mainly considers the effects of $\triangle d$ and $\triangle n$.

Circuit realization. The circuit is the most important part as well as the core of the demodulation system. The circuit system design integrates multiple functions as a whole, such as photoelectric detector, regulated power supply, SCM and serial port, etc, and the circuit can magnify signals about 1,000 times.

The light outputted from two arms of the interferometer enters a switch after passing the photoelectric detector, the switch can control the reception of signals of two arms or one arm and then control the magnification of signals by an amplification circuit controller, at last, the light enters A/D converter, these control circuits are all controlled by a SCM. In addition, an upper serial port must be designed in the circuit to transmit signals to the computer for processing.

4 Experimental Result Analysis

4.1 Temperature Sensing Experiment

Before carrying out a characteristic experiment of the sensing system, we should debug every part in light path and circuit first. The center wavelength of the light source used in this experiment is about 1551nm and its bandwidth is about 40nm. At 24°C, the reflection wavelength of the used fiber grating is 1543.322nm, its reflection bandwidth is about 0.3nm and it has good side mode suppression effects.

In the experiment, put a fiber grating device in water and reflect the wavelength variation of fiber grating by changing water temperature. The wavelength shift of fiber grating is measured by AQ6317B spectrometer OSA; spectrum scan number AVG = 10; the 3dB coupler not only couples the light emitted by the wideband light source into fiber grating but also couples the light reflected by the fiber grating into OSA for detection; in the experiment, the matching liquid should be added to eliminate the effects of another path of reflected light. The relation curve between the wavelength variation of fiber grating encapsulating device and temperature can be got, shown as Fig 2. The change of water temperature is between 24°C and 91°C.

It can be seen from Fig 2 that as temperature sensing, fiber Bragg grating has a good linearity. However, its temperature sensitivity is very low and encapsulating the fiber grating with an aluminum cap can increase its temperature sensitivity by three orders of magnitude. The fiber grating has a great shortcoming: it is very fragile. It will fracture unless you give your whole attention to it.

4.2 Sensing Experimental Result Analysis

There are the following problems and deficiencies in the experimental process and the analysis of experimental results.

(1) Inaccurate water temperature control.
(2) The detection accuracy of Bragg wavelength is limited.
(3) Effects of stress on temperature sensing system.

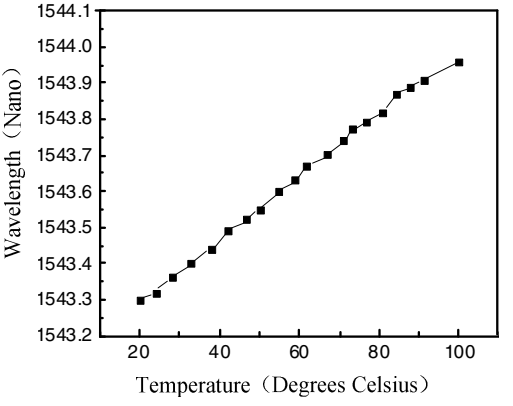


Fig. 2. Schematic Change of Fiber Grating Wavelength with Temperature

4.3 Temperature Demodulation Experiment and Error Analysis

The sensing demodulation experiment is the crux of this scheme. The splitting ratio of the two 2×2 couplers constituting the interferometer often can not reach 50: 50 strictly, the maximum values of optical power got at two output arms are unequal, which will bring the experiment an error. It's OK if the consistency of maximum values of electric signals got at the two ends of circuit input is ensured, so we can realize it by adjusting the multiples of photoelectric conversion circuit and amplification circuit.

In the experiment, we seal one M-Z interferometer in a closed plastic box first and then examine if the M-Z interferometer is under interference conditions: in the experiment, the interferometer is in the linear region when the wavelength shift of fiber grating is between 1544nm and 1546nm; and then drive the light source to operate, change the temperature of external environment by changing the temperature on the sensing grating and measure the output of operation processing circuit, if the temperature detection range of temperature sensor is within 100°C, the temperature sensitivity of fiber grating is 0.01nm/°C and the variation range of wavelength is about 1nm.

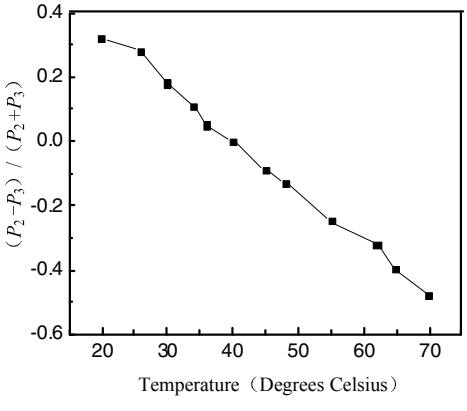


Fig. 3. Test Results of Temperature Sensor

Fig 3 is the ratio of output power difference to output power sum of two arms of the interferometer drawn according to measurement data, namely the relation curve between the cosine of phase difference and temperature variation. It can be seen from Fig 4 that there is a good linearity between temperature and $(P_2-P_3)/(P_2+P_3)$, the resolvable $(P_2-P_3)/(P_2+P_3)$ is 0.0025, then the wavelength detecting device can detect 0.018°C, if the temperature sensitivity of fiber grating is 10pm/°C, the minimum wavelength that the system can detect is 0.18pm, reaching the magnitude of pm.

5 Conclusion

On the basis of understanding the current development status of fiber grating demodulation techniques at home and abroad, this paper carries out the research on fiber grating demodulation techniques. It analyzes the linear sensing demodulation principles based on M-Z interferometer, designs a new kind of fiber grating linear demodulation device and establishes a fiber grating temperature sensor experimental system. The experimental system finishes research requirements according to schedule, but there are still some shortcomings which need to be further studied and improved in the future.

References

1. Kersey, A.D., Davis, M.D., Patrick, H.J., et al.: Fiber grating sensors. IEEE Journal of Lightwave Technology 15, 1442–1463 (1997)
2. Ezbiri, A., Kanellopoulos, S.E., Handerek, V.A.: High-resolution instrumentation system for fiber-Bragg grating aerospace sensors. Opt. Commun. 150, 43–48 (1998)
3. Rao, Y.J., Jackson, D.A., Zhang, L., et al.: Dual-cavity interferometric wavelength-shift detection for in-fiber Bragg grating sensors. Opt. Lett. 21(19), 1556–1558 (1996)
4. Keiser, G.: Optical Fiber Communications Third Edition, vol. (1), pp. 308–310. Publishing House of Electronics Industry (2002)
5. Ball, G.A., Morey, W.W., Cheo, P.K.: Fiber Laser Source/Analyzer for Bragg Grating Sensor Array Interrogation. Journal of Lightwave Techology 12(4), 700–703 (1994)
6. Culshaw, B., Dakin, J.: Optical Fiber Sensors, pp. 45–46. Artech House, Boston (1996)
7. Hocker, G.B.: Fiber-optic sensing of pressure and temperature. Appl. Opt. 18(9), 1445–1448 (1979)
8. Hill, K.O.: Bragg gratings fabricated in monomode photosen-sitive optical fiber by UV exposure through a phase mask. Applied Physics Letter 62, 1035–1037 (1993)

Analysis on Authentication Secrecy of Non-repudiation Protocols

Hong Zhang

School of Computer and Communication Engineering
Weifang University
Weifang 261061, China
zhanghongwf@sina.com

Abstract. This paper discusses the secrecy authentication goals of non-repudiation protocols and verifies the deficiencies of A（0） protocol in authentication secrecy by an attack method; and then, it modifies the confirmation mode of A（0） protocol after its message format and session key are established and puts forward NA（0） protocol; and finally, it makes a formal analysis of NA（0） protocol by SVO logic and verifies the authentication meeting subject identity of NA（0） protocol and the secrecy of its session key.

Keywords: non-repudiation protocol; authentication secrecy; formal analysis; SVO logic.

1 Introduction

Non-repudiation protocols are designed to prevent dishonest people denying that they have ever participated in a certain affair and refusing to undertake the corresponding responsibility. Non-repudiation protocols have two goals, one is to confirm non-repudiation of the sender and the other is to confirm non-repudiation of the receiver. A good and secure non-repudiation protocol is the necessary condition of completing e-commerce transactions [1，2]. This paper mainly discusses the authentication secrecy of non-repudiation protocols and puts forward the design and formal analysis methods of authentication secrecy.

2 Authentication Secrecy

In a computer network and distribution system, one subject usually needs to confirm the identity of the other subject when carrying out resource access or communication. Sometimes keys or other kinds of secrets need to be distributed among subjects, the authentication protocol is used to describe how to confirm identities and distribute secrets among subjects and it is usually composed of a series of exchange messages among subjects. The authentication may concern two subjects or more; it may be a one-way authentication or a mutual authentication and it may use a symmetric key system or an asymmetric key system. The same with authentication protocols, non-repudiation protocols also need to confirm identities and distribute keys among

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 705–711.
springerlink.com

subjects and their requirements for authentication secrecy are the same, so the authentication secrecy of non-repudiation protocols can be discussed by design and formal analysis of authentication protocols.

3 Design of Authentication Secrecy

The authentication protocol is the basis of network security and even an authentication protocol which is established on a perfect cryptographic system may still have various kinds of security vulnerabilities. The difficulties in design and analysis of authentication protocol consist in the subtlety of security target itself, the complexity of protocol operation environment, the complexity of attacker model and the high concurrency of authentication protocol itself. This paper will discuss authentication secrecy by the design and formal analysis [3] of A(0) protocol and its improved protocol next.

3.1 A (0) protocol

A(0) protocol is a kind of key agreement protocol that Matsumoto, Takashima and Imai got by modifying the key exchange protocol of Diffie-Hellman. It has advantages of lightening the burden on the authentication center and limiting its authority. As the premise of the protocol, a public big prime number P and the primitive element a on finite field $GF(P)$ must be selected first. Before the protocol starts, the two communication parties A and B select a random integer $\bar{x}$ and $\bar{y}$ respectively and send $\overline{R_a} = a^{\bar{x}} (\mathrm{mod}\, P)$ (for A) and $\overline{R_b} = a^{\bar{y}} (\mathrm{mod}\, P)$ (for B) got by calculation to the authentication center T to obtain their respective public agreement key certificates. The certificate is the result that the authentication center T signs the identity of any subject C and its public agreement key $\overline{R_c}$. And then, A and B select a random integer x and y respectively, A calculates $R_a = a^x (\mathrm{mod}\, P)$ and B calculates $R_b = a^y (\mathrm{mod}\, P)$. The obtained R_a and R_b are called the temporary public agreement key of A and B respectively. On this basis, A(0) protocol can be executed. The concrete A(0) protocol is as follows:

(1) $A \rightarrow B$: A, $\overline{R_a}$, $\left\{A, \overline{R_a}\right\}_{K_i^{-1}}$, R_a

(2) $B \rightarrow A$: B, $\overline{R_b}$, $\left\{B, \overline{R_b}\right\}_{K_i^{-1}}$, R_b

Here, $\left\{A, \overline{R}_a\right\}_{K_i^{-1}}$ is the public agreement key certificate of subject A issued by the authentication center T. After B receives message (1), it confirms the identity of A by verifying the signature of T and then calculates $K_{ab} = (\overline{R_a})^y \cdot (R_a)^{\bar{y}} = a^{\bar{x}\bar{y}} \cdot a^{x\bar{y}}$ as the shared session key with A, similarly, A can also verify the identity of B and obtain the shared session key with B: $K_{ab} = (\overline{R_b})^x \cdot (R_b)^{\bar{x}} = a^{x\bar{y}} \cdot a^{\bar{x}\bar{y}}$. At last, A and B establish the session key between them K_{ab} by executing A(0) protocol.

3.2 Improved protocol of A（0）protocol-NA（0）

Many kinds of methods for attacking A(0) protocol have been found so far. The most common attack method is as follows: the attacker *P* carries out normal communication and initiates the first round of protocol execution ahead of *A* [4]:

(1) $P \rightarrow A$: P, $\overline{R_P}$, $\left\{P, \overline{R_P}\right\}_{K_t^{-1}}$, R_P

(2) $A \rightarrow P$: A, $\overline{R_a}$, $\left\{A, \overline{R_a}\right\}_{K_t^{-1}}$, R_a

And then, *P* retransmits the message that *A* sends to him to *B* to initiate the second round of protocol execution and intercepts the message that *B* sends to *A*.

(1) $P(A) \rightarrow B$: A, $\overline{R_a}$, $\left\{A, \overline{R_a}\right\}_{K_t^{-1}}$, R_a

(2) $B \rightarrow P(A)$: B, $\overline{R_b}$, $\left\{B, \overline{R_b}\right\}_{K_t^{-1}}$, R_b

It can be seen from the above attack method that the attackers mainly use the identity of A(0) protocol message format and the characteristic that the subjects of the protocol can not differentiate the initiating party of the protocol from the response party of the protocol to initiate effective attacks. To ensure the security of the protocol, modify the message format of the protocol first to make the protocol differentiate the initiating party from the response party and then carry out handshake confirmation after the session key is established to make both parties of the protocol confirm that the other party has had an agreement session key. For the defects of A(0) protocol, we improve it and get NA(0) protocol as follows:

(1) $A \rightarrow B$: N_a, A, $\overline{R_a}$, $\left\{A, \overline{R_a}\right\}_{K_t^{-1}}$, R_a

(2) $B \rightarrow A$: B, $\overline{R_b}$, $\left\{B, \overline{R_b}\right\}_{K_t^{-1}}$, R_b, $\left\{N_a, B, N_b\right\}_{K_{ab}}$

(3) $A \rightarrow B$: $\left\{N_a, A, N_b\right\}_{K_{ab}}$

Here, $N_a = H($ *Date Time A* $)$ is the random number *P* generated by *A* for marking the operation of the protocol and N_a links the mutual information in the protocol together. *Date Time A* marks the time subject *A* spends in initiating the protocol. $N_b = H($ *Date Time B* $)$ is the random number generated by *B* for marking the freshness of agreement key of the protocol and *Date Time B* marks the time subject *B* spends in accepting the protocol. *H* is a strong one-way collisionless function. Different *Date Time* have different $H($ *Date Time* $)$ random numbers. The protocol uses *Date Time* to have the ability of resisting and refusing service attacks. If attackers want to let the subjects of the protocol waste a lot of time on useless wait and unable to provide services for honest subjects, the subjects can cancel the execution of long wait protocols according to their own *Date Time* to make attacks fail.

3.3 Analysis of NA（0）protocol

The execution premise of NA(0) protocol is very simple, it only needs a certificate issuing center. After the participation parties of the protocol obtain their respective

certificates, they can execute NA(0) protocol with the person whom they want to communicate with to obtain a temporary key shared by both parties and then they can use the temporary key to communicate. NA(0) protocol has made good adjustments to message format to make every step of message contain the step information of the protocol and avoid type flaw attacks. NA(0) protocol is simple and efficient, the establishment of its key only needs three steps in all and it doesn't have redundant information. NA(0) protocol has the ability of resisting and refusing service attacks by including communication initiation and acceptance time in protocol information. In addition, it can be known by formal analysis of NA(0) protocol that it has the ability of resisting replay and forge attacks.

4 Formal Analysis of Authentication Secrecy

The design and analysis of authentication secrecy are a very difficult task. Even if we only discuss the most basic authentication protocol and it only has two or three participation subjects and three or five exchange messages, to design a correct authentication protocol which meets authentication goals and has no redundancy is also very difficult. And therefore, a kind of proper formal analysis tool is urgently needed to make a rigorous formal analysis of authentication secrecy in the protocol and examine if the protocol reaches authentication secrecy and there are security flaws and redundancies in the protocol.

The most direct and simplest security protocol analysis method is modal logic method based on knowledge and belief inference [5]. They are composed of some propositions and inference rules, propositions represent the knowledge or beliefs of the subject for messages and new knowledge and beliefs can be deduced from known knowledge and beliefs by using inference rules. In this kind of methods, the most famous method is logic of BAN, including BAN logic, GNY logic, AT logic, VO logic and SVO logic. SVO logic absorbs the advantages of BAN logic, GNY logic, AT logic and VO logic and integrates them in one logic system. In the aspect of formal semantics, SVO logic redefines some concepts as distinguished from AT logic, thereby canceling some restrictions in AT logic system.

4.1 SVO Logic

The marks that SVO logic uses are similar to BAN logic and there are 12 special symbols in all. The formal analysis of a security protocol by SVO logic can be divided into three steps: the initialization assumption set Ω of the protocol is given first and then the goal set Γ that the protocol may or should reach is given; and finally, prove if the conclusion $\Omega|\text{-}\Gamma$ is tenable in SVO logic, if it is tenable, it shows that the protocol reaches the expected design goal and the design of the protocol is successful.

Using SVO logic can analyze not only all kinds of authentication protocols but also the secrecy of non-repudiation protocols which find an increasingly extensive application in the electronic commerce successfully. SVO logic obeys two inference rules and ten axioms [6]. The two inference rules are as follows:

MP rule: ψ can be deduced from φ and $\varphi \supset \psi$;

Nec rule: $\vdash P \mid \equiv \varphi$ can be deduced from $\vdash \varphi$.

Here, φ and ψ are formulas, P represents a subject and $\vdash \varphi$ represents that φ is a formula which can be deduced from an axiom.

The ten axioms are trust axiom, message source axiom, key agreement axiom, acceptance axiom, message possession axiom, message comprehension axiom, jurisdiction axiom, message freshness axiom, temporary value verification axiom and "good" shared key symmetry axiom. Only those axioms which are related to authentication are listed here.

4.2 Formal Analysis of NA (0) protocol

To verify the security of the modified protocol, now use SVO logic to make a formal analysis. Analyze the initiating party A and receiving party B of the protocol as follows:

For subject *A*. The initialization assumption set about subject A[7]:

P_1: $A \mid \equiv \mathrm{P} K_\sigma(T, K_t)$,

P_2: $A \mid \equiv A \ni (\overline{R_a}, R_a, \overline{x}, x)$,

P_3: $A \mid \equiv \mathrm{SV}(\{B, \overline{R_b}\}_{K_t^{-1}}, K_i, (B, \overline{R_b}))$,

P_4: $A \mid \equiv \mathrm{EV}((N_a, B, N_b), K_{ab}, \{N_a, B, N_b\}_{K_{ab}})$,

P_5: $A \mid \equiv \mathrm{P} K_\delta(A, (\overline{R_a}, R_a))$,

P_6: $A \mid \equiv \#(R_a)$,

P_7: $A \mid \equiv ((T \mid \sim PK_\delta(B, \overline{R_b}) \wedge A \lhd (\ (B, \overline{R_b}, \{B, \overline{R_b}\}_{K_t^{-1}})\ , *_b\ ) \wedge$

$EV\ (\ (N_a,\ B,\ N_b)\ ,\ K_{ab}, \{N_a,\ B,\ N_b\}_{K_{ab}})\ ) \supset \mathrm{P} K_\delta(B,\ (\overline{R_b}, *_b)\ )\ )$,

P_8: $A \lhd (B, \overline{R_b}, \{B, \overline{R_b}\}_{K_i^{-1}},\ R_b, \{N_a,\ B,\ N_b\}_{K_{ab}})$,

P_9: $A \mid \equiv A \lhd (B, \overline{R_b}, \{B, \overline{R_b}\}_{K_i^{-1}},\ R_b, \{N_a,\ B,\ N_b\}_{K_{ab}})$,

P_{10}: $A \mid \equiv (T \mid \sim (B, \overline{R_b}) \supset T \mid \sim \mathrm{P} K_\delta(B, \overline{R_b}))$,

P_{11}: $A \mid \equiv (B \mid \sim (N_a,\ B,\ N_b) \supset B \mid \equiv B \xleftarrow{K_{ab}-} A \wedge B \mid$
$\sim (B \ni K_{ab}) \wedge B \mid \equiv \#(K_{ab}))$,

$P_1 \ldots P_7$ reflect the initial beliefs of subject A, P_8 receives messages, P_9 comprehends messages and P_{10} and P_{11} interpret messages. In addition, there is a supplement to SVO logic, namely introducing $EV(X, K, Y)$ to express that the result of encrypting X by an encryption key K is Y.

Protocol goals[8]:

G_1: $A \equiv \xleftarrow{K_{ab}+} B$

G_2: $A \mid\equiv \#(K_{ab})$

Use rules and axioms to infer.

The results got by all steps are written out first and then the rules, axioms, formulas and initialization assumptions used for deducing the results are given.

(1) $A \mid\equiv (A \lhd \{B, \overline{R_b}\}_{K_i^{-1}})$ can be got by P_8, acceptance axiom and Nec rule.

(2) $A \mid\equiv (T \mid\sim (B, \overline{R_b})$ can be got by formula (1), P_1, P_3 and message source axiom.

(3) $A \mid\equiv T \mid\sim \mathrm{P}K_\delta(B, \overline{R_b})$ can be got by formula (2), trust axiom and MP rule.

(4) $A \mid\equiv \mathrm{P}K_\delta(B, (\overline{R_b}, *_b))$ can be got by formula (3), P_9, P_4, trust axiom, P_7 and MP rule.

(5) $A \mid\equiv A \xleftarrow{K_{ab}} B$ can be got by formula (4), P_5, key agreement axiom, trust axiom and MP rule. K_{ab} is shown as formula (6).

(6) $K_{ab} = F_0(\overline{R_a}, R_a, \overline{R_b}, *_b) = (\overline{R_b})^x \cdot (R_b)^{\bar{x}} = (\overline{R_a})^y \cdot (R_a)^{\bar{y}} = g^{\bar{x}\bar{y}+x\bar{y}} (\mathrm{mod}\, P)$

(7) $A \mid\equiv \#(K_{ab})$ can be got by formula (6), P_6 and message freshness axiom. The goal G_2 is achieved.

(8) $A \mid\equiv A \ni (\overline{R_b}, *_b)$ can be got by P_9, message possession axiom, trust axiom and MP rule.

(9) $A \mid\equiv A \ni K_{ab}$ can be got by formula (8), P_2, formula (6) and message possession axiom.

(10) $A \mid\equiv A \xleftarrow{K_{AB}-} B$ can be got by formula (5), formula (9), the definition of $A \xleftarrow{K_{AB}-} B$, trust axiom and MP rule.

(11) $A \mid\equiv (A \lhd \{N_a, B, N_b\}_{K_{ab}})$ can be got by P_8, acceptance axiom and Nec rule.

(12) $A \mid\equiv B \mid\sim (N_a,\ B,\ N_b)$ can be got by formula (5), formula (1) and message source axiom.

(13) $A \mid\equiv B \mid\sim (B \ni K_{ab})$ can be got by formula (12), P_{11}, trust axiom and MP rule.

(14) $A \mid\equiv B \mid\approx (B \ni K_{ab})$ can be got by formula (7), message freshness axiom, formula (13), temporary value verification axiom, trust axiom and MP rule.

(15) $A \mid\equiv A \xleftarrow{K_{ab}+} B$ can be got by formula (10), formula (13) and the definition of $A \xleftarrow{K_{ab}+} B$. The goal G_1 is achieved

It can be known from formula (7) and formula (15) that protocol goals G_1 and G_2 have been achieved.

For subject *B*. The combination of the first message and the second message that subject *B* receives corresponds to the message that subject *A* receives. It can be got by symmetry that the protocol can achieve the following goals[9]:

G_3: $B \equiv B \xleftrightarrow{K_{ab}+} A$

G_4: $B \equiv \#(K_{ab})$

It can be known from Ⅰ and Ⅱ that the modified protocol meets security goals.

The conclusion shows that after the modified A(0) protocol is executed successfully, both subject *A* and subject *B* believe that K_{ab} is the session key of freshness owned jointly by them and unknown to the others. So it is said that the improved A(0) protocol completes a definite key authentication, thereby reaching ideal authentication goals.

5 Conclusion

This paper discusses the design and formal analysis methods of authentication secrecy in non-repudiation protocols, puts forward an improved NA(0) protocol and verifies its authentication and secrecy. It is to be noted that the existing formal analysis methods are still far from perfect, it's because they can only find the shortcomings of protocols but can not ensure trouble-free protocols after analysis necessarily secure and non-attacking. And therefore, the existing formal analysis methods remain to be further deepened and perfected.

References

1. Even, S., Goldreich, O., Lempel, A.: A Randomized Protocol for Signing Contracts. Commun. ACM 28(6), 637–647 (1985)
2. Schneider, S.: Formal Analysis of a Non-repudiation Protocol. In: IEEE Computer Secutity Foundations Workshop, pp. 54–65. IEEE Computer Society Press, Los Alamitos (1998)
3. Abadi, M., Rogaway, P.: Reconcilling two views of cryptography (the computational soundness of formal encryption). In: IFIP TCS, pp. 3–22 (2000)
4. Ezhilchelvan, P.D., Shrivastava, S.K.: A Family of Trusted Third Party Based Fair-exchange Protocols. IEEE Transactions on Dependable and Secure Computing 2, 273–286 (2005)
5. Zhou, J., Gollmann, D.: A Fair Non-repudiation Protocol. In: Symposium on Security and Privacy, pp. 55–61. IEEE Computer Society Press, Oakland (1996)
6. Kremer, S., Markowitch, O.: Fair multi-party non-repudiation protocols. Internation Journal of Information Security 1(4), 223–235 (2003)
7. Khan, A.S., Mukund, M., Suresh, S.P.: Generic Verification of Security Protocols. In: Godefroid, P. (ed.) SPIN 2005. LNCS, vol. 3639, pp. 221–235. Springer, Heidelberg (2005)
8. Dolev, D., Yao, A.: On the Security of Public Key Protocols. IEEE Transactions on Information Theroy 29(2), 198–208 (1983)
9. Damgård, I.B.: Practical and Provably Secure Release of a Secret and Exchange of Signatures. In: Helleseth, T. (ed.) EUROCRYPT 1993. LNCS, vol. 765, pp. 200–217. Springer, Heidelberg (1994)

Research on Control Strategy of Intelligent Monitoring System for Greenhouse Environment

Hong Zhang

School of Computer and Communication Engineering
Weifang University
Weifang 261061, China
zhanghongwf@sina.com

Abstract. Based on the characteristics of nonlinearity, large delay, strong coupling, etc, the fuzzy control method was selected as the control strategy for the system. According to the analysis and research about fuzzy control theory and greenhouse environment, temperature-humidity deviation and deviation variation rate of the system were selected as I/O variables and would be fuzzified, and a fuzzy controller was designed for the greenhouse environment monitoring system. After that, a dynamic greenhouse environment model was constructed and simulations were carried out. The analytic results got a good agreement with the experiment data, which proved the accuracy of the model and the feasibility of the control strategy. In the end, the outcome of fuzzy control was compared with that of PID control. It was confirmed that the fuzzy control with a smaller overhoot and shorter adjusting time is superior to the PID control.

Keywords: greenhouse environment; fuzzy control; dynamic greenhouse model; PID control.

1 Introduction

The greenhouse, a place that can create the best conditions for plant growth and avoid the effects of external seasonal variations and bad weather, is an important component of modern agriculture[1]. With the economic development, technological advances and importance attached to energy saving, the traditional greenhouse production technology has been unable to meet the needs of agricultural development, which is mainly reflected in the management and control strategy of greenhouse environment. This paper mainly studied the control strategy. The fuzzy control method was used as the system control strategy, the fuzzy controller for greenhouse environment monitoring system was designed, and the dynamic greenhouse model[2] was obtained. The results of simulation experiment on the system carried out by Matlab/Simulink verified the accuracy of the model and the feasibility of control method. Meanwhile, the comparison with PID control was made.

J. Lee (Ed.): Advanced Electrical and Electronics Engineering, LNEE 87, pp. 713–720.
springerlink.com

2 Design of System Controller

The greenhouse system is a large system, which has the characteristics of strong mutual coupling and time-varying between nonlinearity, time-varying, delays, uncertainty, multiple objectives, and control parameters. Therefore, an accurate mathematical model is difficult to be established for the greenhouse environment monitoring system, and traditional control method adopted can not effectively control the greenhouse environment.

The fuzzy control does not require an accurate mathematical model established the controlled object, has better response time, system stability and robustness, and is ideal for greenhouse control system. Therefore, this paper focuses on researching the control function of fuzzy control method on the environment state.

2.1 Simplification of Fuzzy Controller

The fuzzy control theory is a nonlinear control [3] based on fuzzy sets theory and fuzzy linguistic variables and fuzzy reasoning. Currently, the fuzzy control theory has become widely used in single-input single-output (SISO) system, but for complex control system, the multiple-input multiple-output (MIMO) variable system with strong coupling is often encountered. For the fuzzy controller, the control rule increases exponentially with the increase of input, but too many control rules will make fuzzy controller become too complex and difficult to control. In the case of multiple variables, the structure of fuzzy controller shall be simplified to reduce fuzzy control rules. For the MIMO fuzzy controller, its rule has the following form:

$$R = \left\{ R^1_{\mathrm{MIMO}}, R^2_{\mathrm{MIMO}}, \ldots, R^n_{\mathrm{MIMO}} \right\} \tag{1}$$

Where: R^i_{MIMO}: if x is A_i, and ... and y is B_i, then z_1 is C_{i1}, ..., z_q is C_{iq}.

The antecedent (input and preconditions) of R^i_{MIMO} is a fuzzy set in the direct product space $X \times ... \times Y$, the seccedent (conclusion) is the combination of q control actions which are mutually independent. Therefore, i-rule can be expressed as the following fuzzy implication formula, i.e..

$$R^i_{\mathrm{MIMO}} : (A_i \times \ldots \times B_i) \rightarrow (C_{i1} + \ldots + C_{iq}) \tag{2}$$

2.2 Design of Fuzzy Controller

The fuzzy controller is the core of the fuzzy control system. The design of fuzzy controller usually includes the following items: (1) determine the input and output variables of fuzzy controller (i.e., determine the control volume); (2) develop fuzzy control rules; (3) fuzzy quantifying is made on the control variables; (4) selection of the universe of discourse and determination of quantization factor, scale factor and other parameters.

The main factors affecting the greenhouse environment are temperature, humidity, light intensity, CO_2 concentration, etc. The temperature and humidity have the most obvious effects on the greenhouse, so the system made specific research on them. The

temperature and humidity control system in the greenhouse environment is a multiple-input multiple-output (MIMO) fuzzy control system, and four inputs and seven outputs need to be considered[4]. According to the principle that the MIMO system can be simplified into several multi-input single-output (MISO) systems, a multiple of MISO fuzzy controllers can be used to replace the MIMO fuzzy controller, to reduce fuzzy control rules and simplify the design of the controller.

The control of greenhouse temperature and humidity has relative independence, so the system uses temperature and humidity fuzzy controllers to solve the issue of too large control rules caused by multiple inputs. The following analyzes the process of building the fuzzy controller of the greenhouse system based on the example of temperature control.

Fuzzification of input and output. In temperature control, the system selects the temperature deviation e (t) and temperature variation rate ec (t) as input variables, and the corresponding fuzzy sets are E_T and EC_T. The temperature deviation within setting value ± 1°C is fuzzy control area, i.e., the basic domain of discourse is [-1, 1]. The value beyond the domain is treated as the boundary value, the quantification domain is [-6, 6], so the quantization factor Ket = 6; the temperature variation rate EC_T = dE_T/dt reflects the variation trend of temperature deviation in on-site state. The basic domain of discourse is [-1, 1], the quantification domain is [- 6, 6], so the quantification factor is $Kect$ = 6.

The fuzzy sets E_T and EC_T of both the temperature deviation and temperature variation rate are expressed by seven fuzzy states, namely PB (positive big), PM (positive middle), PS (positive small), ZO (zero), NS (negative small), NM (negative small), and NB (negative big). The shape of antecedent membership function of fuzzy rules has small impacts on control performance, but the size of breadth has larger impacts on performance; the breadth of seccedent membership function has small impacts on control performance, so the system used triangular membership function to reduce the calculated amount of the system.

The temperature output control includes heating and cooling, the output fuzzy sets fuzzy U_T is expressed by 6 fuzzy states, namely PB, PS, ZO (moderate), NS (light cooling), NM (moderate cooling) and NB (heavy cooling). The heating is expressed by two fuzzy states. PB means the rapid heating of stove, and PS means slow heating. The cooling of greenhouse system is composed by three states: substantial cooling, moderate cooling and slight cooling, and the output membership function uses triangle distribution.

Establishment of control rule. The basic idea of establishing fuzzy control rules is that when the error is large or relatively large, it focuses on selecting controlled quantity to eliminate the error as soon as possible; while when the error is small, the selection of controlled quantity shall avoid overshoot, with the stability of the system as the main premise. According to the above principles, combined with the corresponding control rules, the error can be eliminated by judging the deviation E_T and deviation rate EC_T during the temperature adjustment. The winter control rules can be described by fuzzy condition statement as follows.

IF E=NB AND EC=PB THEN U=PB;
IF E=NB AND EC=PM THEN U=PB;
IF E=NB AND EC=PS THEN U=PB;
IF E=NB AND EC=ZO THEN U=PB;
IF E=NB AND EC=NS THEN U=PB;
IF E=NB AND EC=NM THEN U=PB;
IF E=NB AND EC=NB THEN U=PB;
……

49 winter fuzzy control rules can be obtained respectively by analogy. Similarly, the summer fuzzy control rules can also be obtained.

Clarification of fuzzy variables. There are many clarification methods for the fuzzy variables, and the most commonly used methods are the maximum membership degree method, median judgment law and the weighted average method[5]. The system uses the weighted average method for the clarification of variables. In general, the decision of weight coefficient is related to the system response. Therefore, the appropriate weighting coefficient can be selected according to the system design requirements or experience.

3 System Simulation

The establishment of simulation on the field monitoring system of greenhouse environment is based on the environmental dynamic model, and is the validation on the control effects and feasibility of control rules of the designed fuzzy controller.

3.1 Dynamic Greenhouse Model

The greenhouse system is generally divided into five components: soil layer, crop layer, heating layer, indoor air layer and greenhouse covering layer. By learning from the theoretical results of modeling environment and climate at home and abroad, and considering the control function of actuating mechanism, a specific temperature dynamic model [6][7] of greenhouse environmental climate is obtained shown as follows:

$$v\rho C_P \frac{\mathrm{d}T}{\mathrm{d}t} = Q_\gamma + Q_{\text{heater}} + Q_{\text{vent}} + Q_c + Q_l + Q_{\text{soil}} + Q_{\text{leaf}} - Q_{\text{cool}} - Q_{\text{tran}} - Q_P \tag{3}$$

Where, v is the volume of greenhouse (m^3), ρ is the air density (kg/m^3), C_P is the heat content in the air ($Jkg^{-1}K^{-1}$), and Q is energy.

This paper conducts studies and system simulation on the above model based on the greenhouse environment under winter weather conditions. When the outdoor temperature in winter is low, the skylight is basically closed, without considering venting heat exchange, while ignoring the minor impacts of blade surface heat transfer, photosynthesis, and transpiration on greenhouse[8]. The equation (3) can be simplified as:

$$v\rho C_P \frac{\mathrm{d}T}{\mathrm{d}t} = Q_\gamma + Q_{\text{heater}} + Q_c + Q_l \tag{4}$$

Where: Q_γ - solar radiation energy, Q_{heater} - heating energy, Q_c - energy for external thermal conduction, Q_l - long-wave radiation energy.

Table 1 shows the parameter table of greenhouse environment in the model.

Table 1. Parameter Table

Parameter	Symbol	Value
Greenhouse volume	v	1048.32 m^3
Surface area of greenhouse covering materials	A_s	305.68 m^2
Light transmittance of glass	τ	0.89
Heat transfer coefficient of covering materials	K_g	$2.97\triangle T^{0.33}$Wm-^{2}K^{-1}
Air emissivity	ε_1	0.90
Glass emissivity	ε_2	0.90
Air density	ρ	1.2 kgm^{-3}
Heat content in the air	C_P	1006 Jkg^{-1}K^{-1}
Stefan-Boltzman Constant	σ	5.67×10^{-8}wm^{-2}k^{-4}

3.2 System Simulation

The system conducted simulation experiment on the built model by using software Matlab/Simulink to verify the control effect of fuzzy control strategy and the feasibility of control rules. According to the dynamic model of the greenhouse environment, the block diagram of establishing Simulink simulation of the indoor temperature control system was shown in Figure 1. The system input was composed by indoor temperature deviation *E* and temperature variation rate *EC*; the system setting temperature was added into simulation for the selection of fuzzy control rules; *E* and *EC* output control quantity to regulate greenhouse heating system through fuzzy control, and output the control quantity to the workspace of Matlab.

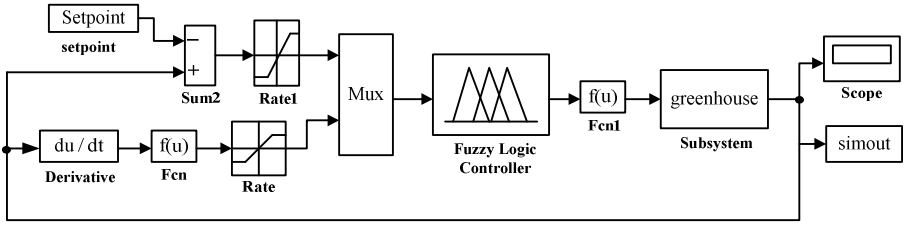


Fig. 1. Simulation Block Diagram of Winter Simulink

Since the outdoor temperature is low in winter, the wet curtain fan is generally not opened for cooling. According to growth requirements of crops, the setting value of control system temperature is 22°C during the day and 10°C during the night. The simulation experiment selected actual climate data on December 23, 2007, and introduced them into the greenhouse model in the form of Excel sheet as the basis for the computer simulation of greenhouse environment. Figure 2 shows outdoor climate

conditions of a day, including outdoor temperature, humidity and light intensity. The equation of the model is integrated by Runge - Kutta algorithm with fixed time and step. Figure 3 is the comparison chart of computer simulation temperature and actual temperature. The figure shows that in the temperature simulation curve, the lowest controlled indoor is about 10°C, and the maximum temperature is less than 26°C. Meanwhile, the optimum temperature for the growth of general crops is between 20°C ~ 30°C, so the above fuzzy control method used for indoor temperature control can meet the needs of crop growth. From the comparison of the simulation curve and the actual curve, the actual curve had a good agreement with the simulation curve, which verified the accuracy of the dynamic greenhouse model and the feasibility of fuzzy control strategy.

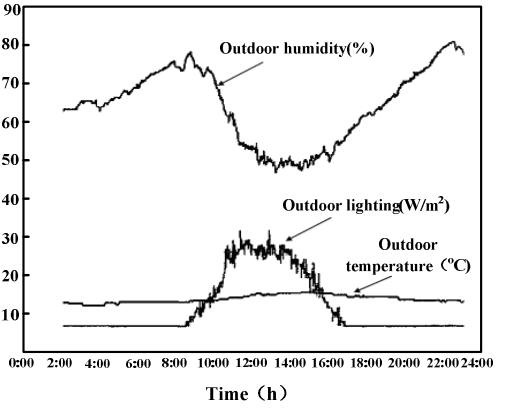


Fig. 2. Outdoor Climate Condition

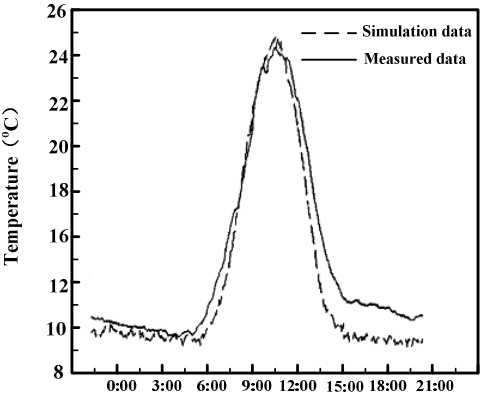


Fig. 3. Comparison Chart of Fuzzy Control

Finally, the control effects of both fuzzy control and PID control are compared. The mathematical model of temperature parameters of greenhouse environmental control can be approximated as first order pure delay plus disturbance model. The transfer function of first-order pure delay system is assumed as $G(s) = \frac{e^{-2s}}{5s+1}$ by using unit step input, and the software Matlab/Simulink is used for simulation. Figure 4 shows the comparison chart of control effects of the two controls. It can be seen from the figure that from the comparison of simulation curve obtained by using fuzzy control method and that obtained by using PID control, the former had smaller overshoot, shorter adjustment time, and better control effects than those of conventional PID control, verifying the rationality of fuzzy control being used as the system control strategy.

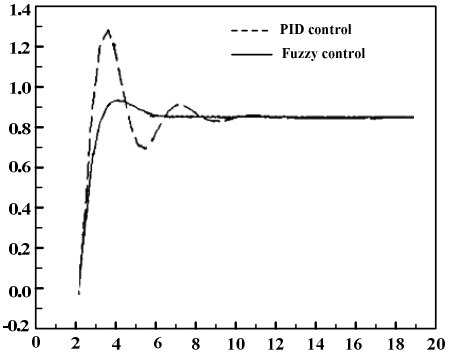


Fig. 4. Comparison of PID Control and Fuzzy Control

4 Conclusion

This paper studied the control strategy of monitoring system for greenhouse environment, and focused on introduction of the design of temperature fuzzy controller, including the structure of fuzzy controller, selection of input and output membership function, and the establishment of fuzzy control rules. The fuzzy control algorithm was simulated, and the feasibility of the control method and the accuracy of the model were validated. In the future, with the continuous development of fuzzy control technology, the improved fuzzy control technology can be used for the further optimized research of control strategy of greenhouse control system to achieve better control purposes.

References

1. Yusof, R., Omatu, S.: Application of Generalized Predictive Control To Temperature Control Process. In: IECON 1991, pp. 1935–1940. IEEE, Los Alamitos (1991)
2. van Straten, G.: Acceptance of optimal operation and control methods for green house cultivation. Annual Reviews in Control 23, 83–90 (1999)

3. van Starten, G.: Acceptance of Optimal Operation and Control Methods for Greenhouse Cultivation. Annual Reviews in Control 23, 83–90 (1999)
4. Blasco, X., Martinez, M., Herrero, J.M.: Model-based predictive control of greenhouse climate for reducing energy and water consumption. Computers and Electronics in Agricultures (55), 49–70 (2007)
5. Eker, I., Torun, Y.: Fuzzy logic control to be conventional method. Energy Conversion and Monagement 47(4), 377–394 (2006)
6. Castafieda-Miranda Jr., R., Ventura-Ramos, E., del Rocfo Peniche-Vera, R., Herrera-Ruiz, G.: Fuzzy Greenhouse Climate Control System based on a Field Programmable Gate Array. Biosystems Engineering 94(2), 165–177 (2006)
7. Stipanicev, D., Marasovic, J.: Networked Embedded Greenhouse Monitoring and Control. In: Istanbul Turkey: Conference on Control Application, pp. 1350–1355 (2003)
8. Coelho, J.P., de Moura Oliveira, P.B.: Greenhouse air temperature predictive control using the particle swarm optimization algorithm. Computers and Electronics in Agricultures 49, 330–344 (2005)

Author Index