Lecture Notes in Electrical Engineering 445

Rongjun Shen
Guangliang Dong
Editors

Proceedings of the 28th Conference of Spacecraft TT&C Technology in China

Openness, Integration and Intelligent Interconnection

清華大學出版社
TSINGHUA UNIVERSITY PRESS

Springer

Lecture Notes in Electrical Engineering

Volume 445

About this Series

"Lecture Notes in Electrical Engineering (LNEE)" is a book series which reports the latest research and developments in Electrical Engineering, namely:

- Communication, Networks, and Information Theory
- Computer Engineering
- Signal, Image, Speech and Information Processing
- Circuits and Systems
- Bioengineering

LNEE publishes authored monographs and contributed volumes which present cutting edge research information as well as new perspectives on classical fields, while maintaining Springer's high standards of academic excellence. Also considered for publication are lecture materials, proceedings, and other related materials of exceptionally high quality and interest. The subject matter should be original and timely, reporting the latest research and developments in all areas of electrical engineering.

The audience for the books in LNEE consists of advanced level students, researchers, and industry professionals working at the forefront of their fields. Much like Springer's other Lecture Notes series, LNEE will be distributed through Springer's print and electronic publishing channels.

More information about this series at http://www.springer.com/series/7818

Rongjun Shen · Guangliang Dong
Editors

Proceedings of the 28th Conference of Spacecraft TT&C Technology in China

Openness, Integration and Intelligent Interconnection

清華大學出版社
TSINGHUA UNIVERSITY PRESS

Springer

Editors
Rongjun Shen
Chinese Academy of Engineering
Beijing
China

Guangliang Dong
Beijing Institute of Tracking and Telecommunications Technology
Beijing
China

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-981-10-4836-4 ISBN 978-981-10-4837-1 (eBook)
DOI 10.1007/978-981-10-4837-1

Jointly published with Tsinghua University Press, Beijing, China

Library of Congress Control Number: 2017940811

The print edition is not for sale in China Mainalnd. Customers from China Mainland please order the print book from: Tsinghua University Press.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

China's aerospace cause will embrace a new round of dynamic development driven by the military-civil integration strategy of the government. In this backdrop, spacecraft TT&C will face both opportunities and challenges. Development and construction of spacecraft TT&C systems must follow an open philosophy and thinking and absorb state-of-the-art ideas and technologies in order to produce an iteratively creative momentum. Based on national demand, integration of the field with other specialties should be improved to seek innovative breakthroughs in a timely manner. At the same time, efforts should be made to extend fields of TT&C services home and abroad by leveraging intelligent linking, intelligent interfacing and intelligent service.

Taking "Openness, Integration and Intelligent Interconnection" as its theme, the 28th China Spacecraft TT&C Conference highlights better activation of the dynamic sources of development of the discipline of spacecraft TT&C and further promotion of creative development of TT&C operation and management integration. The conference will showcase the latest achievements of the country in the field of spacecraft TT&C and explore the future of China's spacecraft TT&C systems.

For the academic conference, totally 310 papers are received from specialists in different fields, and from them 47 papers are selected for publication in the proceedings. With rich contents, clear focus and high academic level, the proceedings has excellent practical and promotional values and we hope it will provide reference and help to officials and scientific and technical personnel at different levels in the field of spacecraft TT&C.

Beijing, China
November 2016

Rongjun Shen

Contents

Part I
Spacecraft TT&C Science and Technology

Chapter 1
Discussion on Networked and Integrated Space-Ground Information System

Jianping Hu, Huizhong Xu, Ting Li, Tian Liu and Hongjun Yang

1.1 Introduction

The space electronic information system includes space-based and ground infrastructure for acquiring, processing, and transmitting information via space platform which is only used as information carrier in order to acquire and apply information.

Our current space information system including spacecraft TT&C and payload control systems is stove-piped, which are developed and deployed independently by various military and civil departments, and using various information acquiring, time reference and information transmission system without uniform technical specifications. There is little effective connection between military and civil resources leading to difficult data sharing and information fusion between them. No integrated space-based information system is formed.

The goal of the proposed space information system is to construct an integrated, open, expandable, interconnected, cross-support, safety, robust, flexible, and reconfigurable space-ground information application system supporting united planning and management. Considering the increasing military and civil application demands and manifold and sophisticated application modes, the system shall be implemented by means of network and integration technologies based on software defined everything (SDX), such as node-and- application-oriented software defined platform (SDP), software defined function (SDF) and so on.

The key to establish a networking and integrated space-ground information system is reconfigurable network topology architecture, definable protocol, and online processing and multifunction adaptability of network nodes, which can provide basis for future resilient construction, function expansion, capability improvement, tailoring by demands, and intelligent situation sensing.

J. Hu · H. Xu · T. Li (✉) · T. Liu · H. Yang
The 10th Research Institute of China Electronics
Technology Group Corporation (CETC), Chengdu 610036, China
e-mail: liting142@126.com

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_1

1.2 Space Information System Development Trends

Space system has developed from single satellite to constellation, and to space network, finally, it will become a networking and integrated space-ground architecture and application [1].

In the early phase of space system development, information acquiring and transmission depended only on a single satellite, whose orbit feature determined that it can only provide unsatisfied coverage, and limited information acquiring and transmission capability. Constellation mode can overcome shortcomings of single satellite mode on coverage area and time. However, in some space information applications, a user may require several kinds of information, such as on navigation and early warning, at the same time, which cannot be met using constellation mode. Space information network may implement connection among various satellites and constellations to provide 24/7 and globe coverage, and integrate resources on space, air, sea, and land to provide full-time and seamless access services. With introduction of integration technologies, space system had developed from a system consisting of single function spacecraft and ground system to a system with multifunction nodes and the SDN architecture. Acting as information carrier, spacecraft and ground infrastructure can join in, maintain, manage, and apply network as server, terminal, node, or transmission line.

Satellite system has developed from single satellite application mode to constellation application mode, with trends to networking mode. Typical constellation system includes GPS navigation satellite system, Iridium satellite system, Space-based Infrared Satellite (SBIRS) system, and so on. GPS system consists of 24 satellites constellation to provide global coverage and real-time location, velocity and timing information. Iridium satellite system consists of 66 satellites constellation in LEO. It is unsuccessful in commercial operation, but it is a technological leap for resolving problem of global mobile satellite communication [2]. SBIRS design utilized a composited constellation combing GEO, HEO, and LEO orbits in order to improve missile finding capability, expand flight midcourse tracking, and implement missile full-range tracking [3].

Onboard processing and crosslink are necessary components of constellation. Although US military TSAT program was cancelled, but its onboard processing and crosslink technologies can used to other applications [4, 5]. Milstar and AEHF communication satellite constellations incorporated crosslink to fast information transmission speed and improve satellite anti-jamming capability. Iridium satellite system had complex and advanced onboard processing capability, and can delivery and exchange information via crosslink without ground station to provide global coverage. GPS constellation began to provide UHF crosslink from Block-IIR satellite to perform inter-satellite radio pseudo range measurement. The ephemeris update information can be acquired frequently to perform onboard real-time obit estimation, keep autonomous navigation for long time, and improve location accuracy and availability in war conditions.

On January 2013, DARPA provided low rate crosslink radio communication platform for F6 satellite project, whose K-band radio allowing any spacecraft joining into F6 swarms. The platform used unique core architecture to provide sustained data links between satellites of swarms and support simultaneous the third part's point-to-point links. Its communication protocol incorporated a data link layer which can integrate the network protocols in higher layer and enable maximum bandwidth assignment via distributed calculation in unique architecture. Although the F6 project was cancelled, this technology can be used to other space networking applications in other smallest project [10].

Constellation system expands the former single satellite application mode and increases the application efficiency. However, the current satellite systems are still stove-piped, which fail to form a uniform and directly interconnected network, leading to little effective interconnection among systems. As a result, application efficiency of information resources in limited space is not fully utilized for data of each system could not be shared and utilized in time.

For the ground section of the space system, U.S. military is prepared to change the state quo where multiple independent ground systems are operated for the on-orbit satellite, and support establishment of a common system. For this purpose, the MOD launched an "Enterprise Ground System (EGS)" program to update the ground section for the military user via a more efficient and more cost effective system and meet challenges confronted with integration and expandability.

In 2015, Lt. Gen. David J. Buck, Commander of 14th Air Force and Joint Functional Component Command for Space have suggested that architecture of the ground section of the U.S. military space system must be safe, flexible and cost effective in an activity held at the Mitchell Institute. Operating multiple separated ground systems will cause stove-piped problem impairing safety, flexibility and economical efficiency. Currently, ground systems of various satellites which are divided by different scope of tasks use special software supplied by contractors, leading to insufficient management from the MOD. In addition, updated and new versions also require data and expertise from the original contractors. It is worth stressing that the government must possess technical baseline and must control interface and standards so as to avoid management ability being limited by those special software. For station network of the space system, the network requires open architecture as well as interface and standards that could be controlled by the user.

U.S. military is developing a common ground operation and control system for satellites. John Hyten, Commander of the Air Force Space Command, suggested that all new satellites for U.S. Air Force must be compatible with the "Multi-Mission Satellite Operation Center (MMSOC)". MMSOC initially established in 2006 is a trial ground system mainly used for demonstration of "Rapid Operationally Responsive Space" program. Characterized by plug and play, MMSOC is easy to modify and update, especially for safety. Leaders of the Air Force deem it as a potential prototype of the common ground architecture for EGS which will operate a number of constellations in the future.

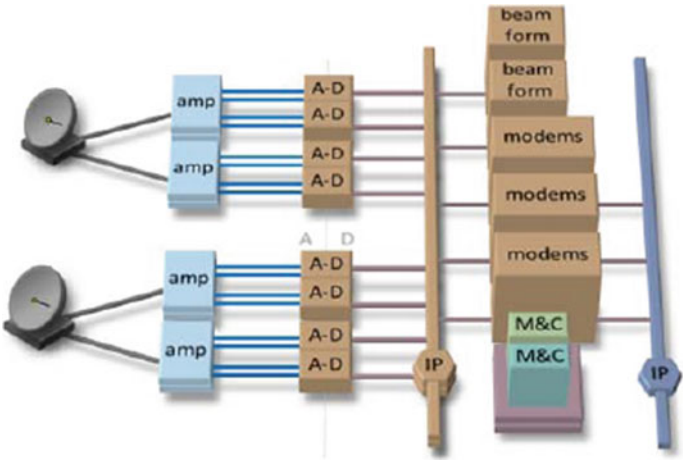


Fig. 1.1 Schematic diagram for SGSS resources pool

U.S. spares no effort to eliminate stove-piped nodes of the ground section. A typical example is the Space Network Ground Section Sustainment (SGSS) for tracking and data relay satellite (TDRS). Having been implementing from 2014 to the end of 2016, main objective of SGSS is to replace all hardwares and software of the space network for constructing a flexible, expandable, upgradable and supportable ground system. SGSS is of "pool" structure as shown in Fig. 1.1. A "resource pool" is formed by standardized and common system equipment. Although there is little special equipment for each TERS, equipment used to perform missions could be selected from the free resources in the pool and after mission, they will be returned to the pool for future use. With all ground terminals sharing the same resources in the pool, this "pool" structure not only decreases demand for equipment but also increases structure flexibility and utilization efficiency of hardwares.

In 1998, JPL initiated the Interplanetary Internet (IPN) [1] program which mainly focused on studying the scheme for end-to-end communication using network outside the Earth and a relevant draft for Internet Engineering Task Force (IETF) protocol was developed.

In 2001, U.S. Goddard Space Flight Center launched Operating Missions Nodes on the Internet (OMNI) which mainly developed the space communication scheme by using ground commercial IP protocol. OMNI took advantage of IP network, data rely satellites (TDRSS) to perform ground tests and airborne flight tests on shuttles, verifying feasibility of using ground IP protocol in space.

NASA plans to build an integrated network architecture as shown in Fig. 1.2 via Space Communications and Navigation (SCaN) program which will integrate existing Near Earth Network (NEN), Space Network (NE) and Deep Space Network(DSN) into one system around 2018. This integrated system will enable integrated management, control, strategy, telemetry, remote control and data

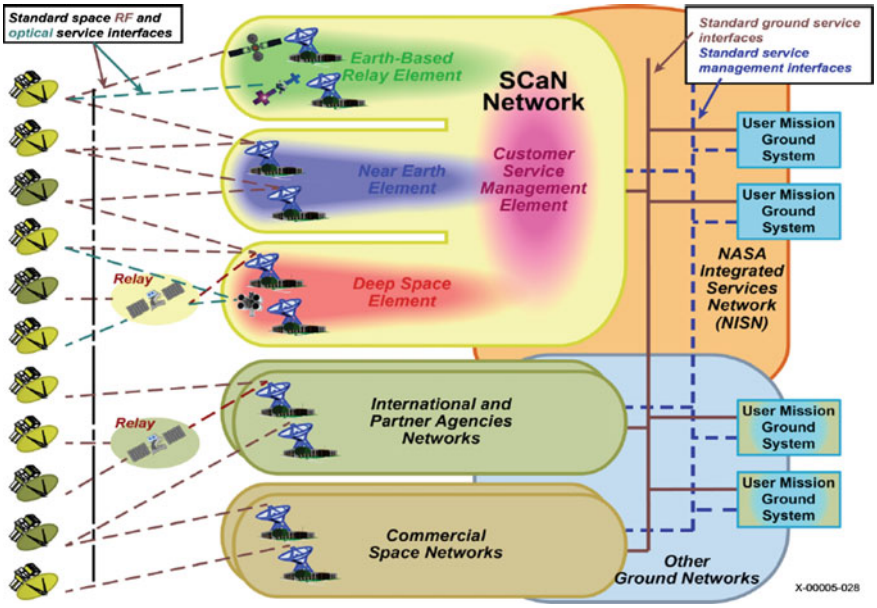


Fig. 1.2 Conceptual schematic for NASA integrated network system architecture around 2018

transmission for various spacecrafts via space/ground/sea-based TT&C navigation and communication resources, so as to gradually achieve the space-ground networking application of space-based information.

This shows that the development trend of the future space information system is to construct a networked, common and flexible space-ground information system, which will not only form ground and space networking, but also enable overall management and integrated application of space-ground resources, route and information, improvement of application efficiency of space-based information as well as more efficient control and management of space-based resources.

However, information application will be limited by the physical border of different platforms within various nodes or a single node caused by networked interconnection only limited among different nodes in the integrated space-ground network. If the satellite payload of space-based nodes still applies the traditional concept, i.e. special hardware/software combinations tailored by function needs leading to a one-to-one correspondence between the "carrier" (hosting hardwares for relevant functions) and the "payload" (functions need to be realized), which means one payload only for one function, information exchange between different functions, or nodes even within a single node will require more space links to meet the demands for information interaction and networking, consequently increasing the network load and requirements for resources processing.

To address this problem, the space-based information system specifically requires payload capacity which may load functions.

According to NASA's SCaN program, a software defined radio technology test was performed for "Communication, navigation and Networking Reconfigurable Testbed" (CoNNeCT) mission at the ISS in 2013. The major test equipment "ScaN Testbed" used software radio platform to enable on-orbit communication capability of software loading and verified the feasibility and maturity of this capability. Function-platform split mutual independence and standardized architecture are prerequisite for integrated application of software loading (Fig. 1.3).

The SpaceCube processing system research program initiated by the Goddard Space Flight Center (GSFC) in 2006 demonstrated a series of system capabilities, including computing capability and reconfigurability of typical space processors. NASA has approved potential of this technology and it used as a main electronic system for test payload of the Relative Navigation Sensor (RNS). SpaceCube V2.0 has basically met the requirements for a mixed, reconfigurable and modular space processing system (Fig. 1.4).

Although integration and multi-function application of the software defined tests performed on the space platform are limited, it is obvious that application of software defined function loading technology on the on-orbit spacecrafts has been verified and approved to some degree.

Some satellite systems of China (including "Beidou" global satellite navigation system [11], etc.) that gradually start to apply onboard processing, crosslink technology and space networking application mode could be independent from the

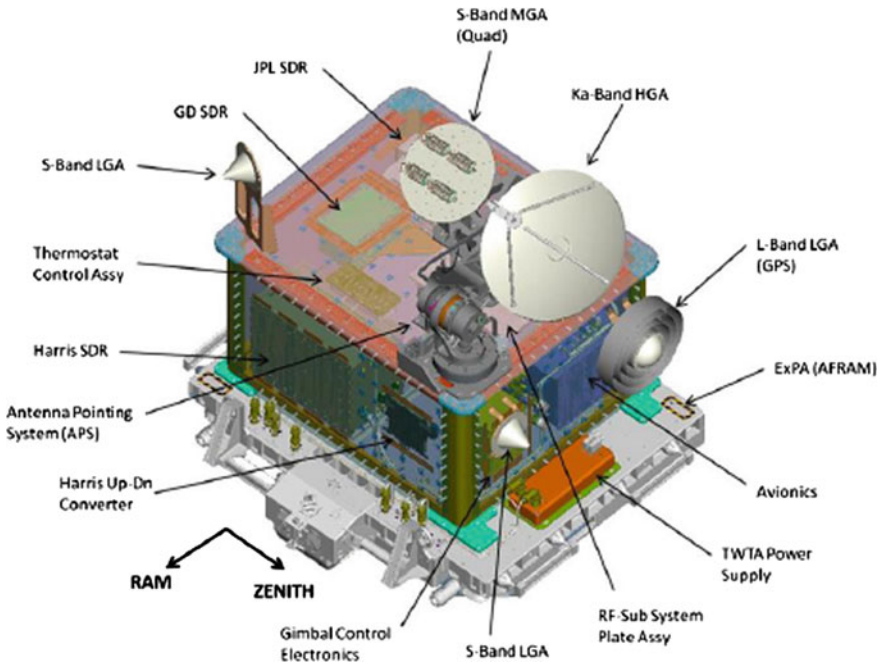


Fig. 1.3 SCaN testbed

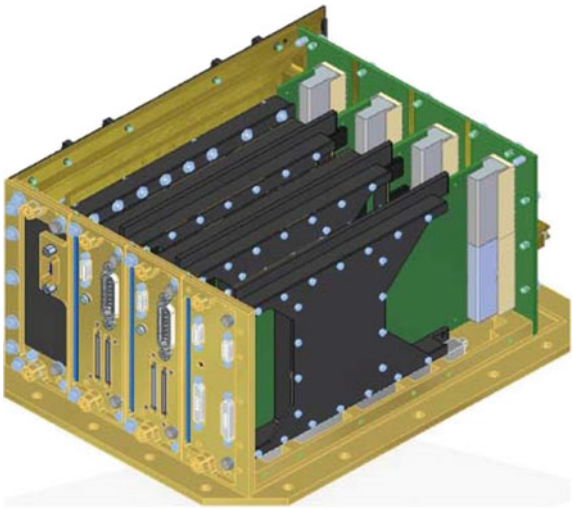


Fig. 1.4 Schematic diagram for function expandable and common processing platform for SpaceCube

ground station network to some degree. However, there is still a certain gap between multi-function processing capacity, interconnection of space heterogeneous networks, space-ground interconnection and those required by networking of space-base information system. A space platform program of China developed a design concept for integrated payload platform, but a breakthrough in series of key technologies is still required so as to meet application needs for integrated payload on middle and small spacecrafts.

According to analysis of relevant documents, research and application of space-based information system networking are now still based on special architecture and protocol, and the network hierarchical structure is still tightly coupled with the network equipment. Although network nodes could enable function conversion via resources reconfiguration and software definition, new protocol couldn't not be deployed and applied during application once a network is constructed. With introduction of software defined node function and software defined network, new technologies for development of space-based information are unfolded.

A common processing platform with multiple functions based on the integrated design concept for software defined platform and software defined function will develop the space-based information system from a network connected by single-function nodes to a flexible network constructed by integrated nodes with configurable hardware and definable function.

As the opportunity for applying networked and integrated technologies in the space-based information system has come, a higher level of space-based information network characterized by software defined everything (SDX) will be the future development trend.

1.3 Suggestions on Networking and Integration of the Space-Based Information System

The space-based information system network [1] consists of backbone networks, access networks, various functional subnetworks and a relevant terminal node of a single user, as shown in Fig. 1.5.

The backbone network consists of space-based backbone networks and ground backbone networks. The dual space-ground backbone networks are linked by multiple space-ground backbone links. The access network consists of access equipment for space-based and ground-based network nodes as well as other special access systems and facilities, functioning as the access link between various functional subnetworks and space/ground-based networks. The functional network consists of user-application-oriented network in the space, air and on the ground, including information acquisition, space-time reference, data transmission networks, etc. As shown in Fig. 1.5, the integrated space-ground information system is a heterogeneous network constructed by multiple platforms and various subnetworks based on open architecture design. In the future, space networks including lunar network, martian network, etc. may be integrated into this system via interplanetary backbone link and access link.

In general construction ideas, in addition to protocol and routing supporting network operations and management, as well as signal, link and other links,

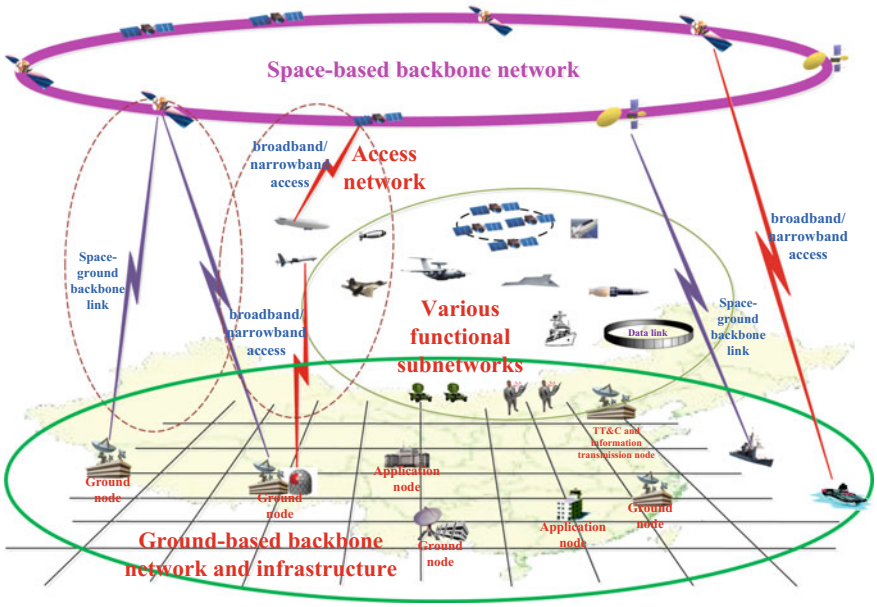


Fig. 1.5 Network architecture for integrated space-ground information system

more attention shall be paid to technologies and equipment formation in the following aspects.

(1) Network architecture

With control functions at network infrastructure layer simplified gradually or split totally in network construction, information forwarding decision are realized by software to introduce and achieve SDN architecture and application formation, so as to make network more suitable and extensible for space application, as well as to improve efficiencies of network operation and resource utilization. Based on SDN idea, control network elements are introduced to space-based and ground backbone networks, access network and other networks (domains), in order to centrally configure, dispatch and manage infrastructures in all kinds of subnetwork (domain) or resources and information transmission function of user nodes. Controllers of all subnetworks (or virtual controllers) are linked to form a general control cluster of space-based information network. One of those control network elements is defined as main controller as required to implement management, which may be set in ground or space-based network. Requirement based definition of control network element, as well as its distributed structure on a multilayer and multi-domain basis, are able to enhance network robustness, survivability and failure configuration capability.

SDN refers to a new three-layer network architecture separating control from forwarding [12, 13]. As shown in Fig. 1.6, the top is an application layer, including various services and applications. It is responsible for customization of network functions, which consists of different application level software. The intermediate, referring to a control layer, deals with network resource management and control, as well as masters the global view of whole network, which is an operating and processing mechanism of network. The bottom, as an infrastructure layer, establishes data paths to realize data forwarding within the network.

After splitting control function of infrastructure, data stream is matched and forwarded according to stream table, through which its performance and efficiency are improved greatly [12, 13]. Also, control layer to be set independently enables flexible resource control and management by network through controllers at different layers, as well as collection and maintenance of network topology, calculation of data forwarding routing, generation and issue of stream table, and control and management of network as needed.

SDN technologies have some unique advantages, such as global view, multi-granularity network control, which bring a new way to solving network security issues [14].

For different applications and network formations, different topology and mission requirements of space-based information system may be achieved by modifying corresponding protocols through software definitions. According to reference [15], current SDN research and application focus on ground network and ground-based wireless network. Comparing with for ground wired and wireless networks, SDN technologies to be applied for space-based information network

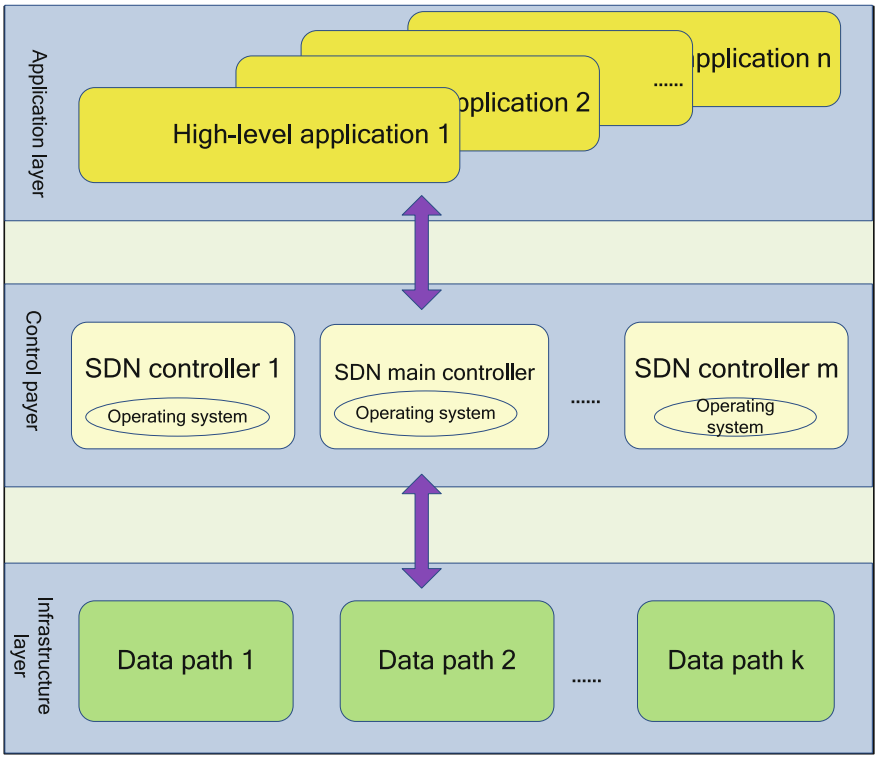


Fig. 1.6 Sketch diagram of SDN architecture

may be faced with a series of technological difficulties in terms of operating system, protocol, software, strategy mechanism and controller.

(2) Integrated multi-function common platform

Based on interconnected space and ground networks, space-based backbone network is constructed by software defined multi-function payloads, to enable its space information processing, multi-platform space network protocol processing, dynamic routing, and SDN control and management capabilities. Supported by SDN architecture, space network protocol and other service systems, space-based node with multi-function payloads as the core enables end-to-end interconnection and integration services of space-based information at space section.

Multi-function payload may be realized by software defined integration technologies. Applicable and efficient integration technologies are able to provide corresponding space nodes with capabilities of totally or partially combining space access and networking, routing and switching and information distribution, efficient calculation and processing, cloud information storage, resource management, autonomous navigation and time reference, and information security and protection. They may be a main infrastructure to construct space nodes of space-based

information system. Realization of this assumption depends on integration levels and efficient processing capabilities.

Although there are some researches about payload technologies of software loading and upgrading based on software radio technologies, function upgrading is limited under certain functional framework and hardware scale conditions, which falls short of the software defined level of overall resource reorganization and function reconfiguration.

The space node capable to obtain, transmit and process information, involves antenna integration, RF channel integration, and integration of signal and information processing. With technological development, antenna and channel may be integrated to achieve integrated configuration of integrated RF and processing. To adapt application requirements of multi-orbit space-based nodes, high requirement are put forward for volume, weight and power assumption of hardware of this part, which needs to weigh capability and scale. Different requirements are met by configuring available integrated RF components and integrated processing modules.

With greatly improved signal processing capability of semiconductor, real-time processing bandwidth has reached hundreds of megabytes. Also, under support of great digital configurable capability provided by FPGA, configurable broadband RF Tx/Rx front-end technologies have been developed rapidly, to enable integration of RF front-end. Advancing digitalization as far as possible, and taking 6 GHz as a dividing point as shown in Fig. 1.7, R/F applications required by different frequency bands and bandwidths may be met through changing RF parts of front-end and array unit combinations. For application requirements of lower than 6 GHz, only applicable components of Rx/Tx channels are selected to constitute RF front-end with Rx/Tx array unit combinations of corresponding band. Digital interfaces are directly connected to back-end integrated processing components to establish a common platform of related scale. Then space payload with corresponding functions and capabilities are formed by uploading related functional software in advance or in orbit. For application requirement of higher than 6 GHz,

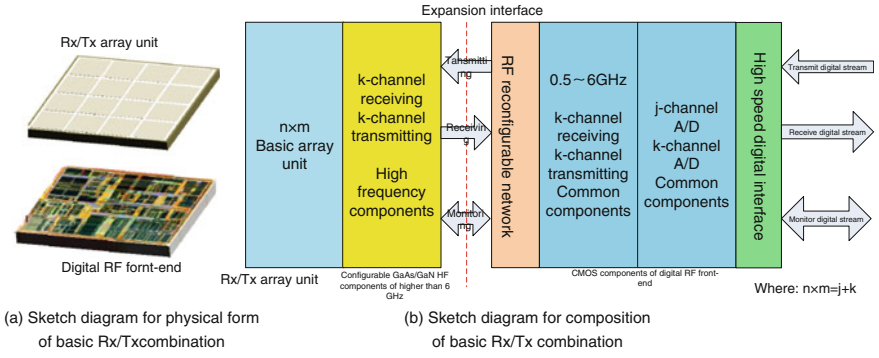


Fig. 1.7 Sketch diagram for basic integrated RF component

it only needs to change array unit combinations of corresponding band and add HF Rx/Tx combinations of corresponding scale.

According to function and mission requirements, different array planes may be arranged at appropriate positions of spacecraft, to achieve its distributed integration application on the space-based platform.

Hierarchical abstraction and virtual design are adopted for integrated processing platform to achieve function-platform split as shown in Fig. 1.8. Functional software modules of mission required system to implement are related to resources of physical platform reasonably and efficiently by software-based deployment function of related logic component layer, which enables combined application of multi-function thread within the range of hardware resource capacity and capability. Achievement of integrated platform is based on construction of open physical hardware and software platforms. On common hardware platform, abstract resource model of system is established by visual modeling technologies, then mapping from logic resources to real physical resources are achieved by blueprint deployment [16].

With reasonable design, basic processing unit is formed by FPGA, DSP and other devices as well as related conversion modules, which is defined as the basic Cube that can construct larger processing platform as shown in Fig. 1.9a. Corresponding space-based processing resource pool is established by open VPX Bus unit block as shown in Fig. 1.9b, to satisfy various applications' demands for signal and information processing resources. For payload function of one Cube only

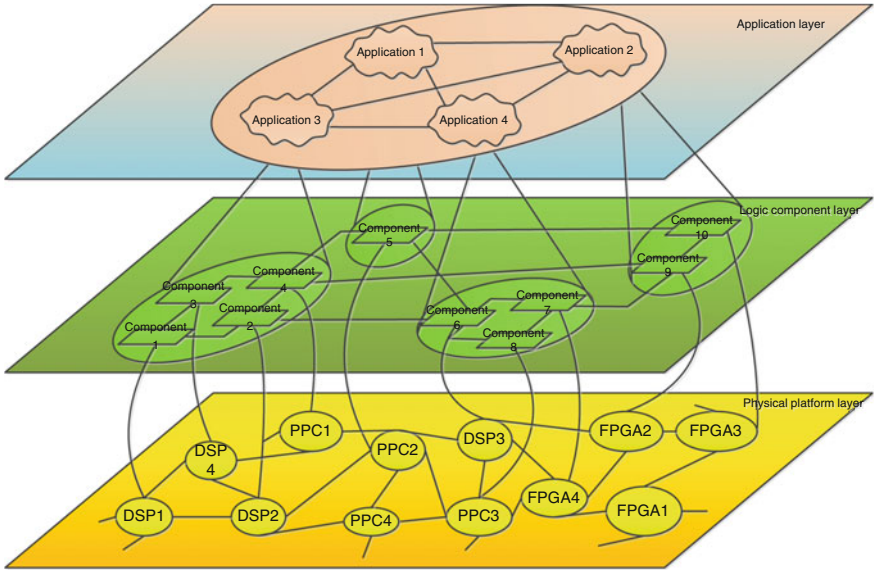
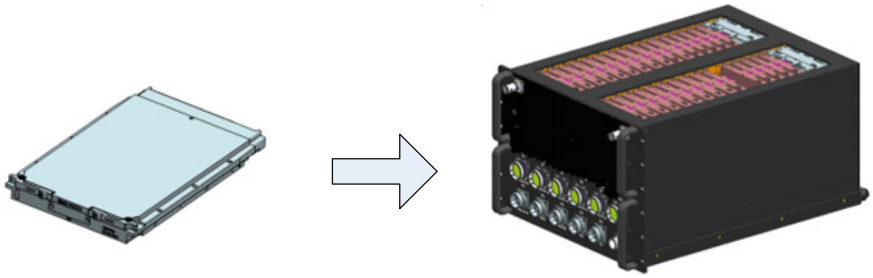


Fig. 1.8 Sketch diagram for hierarchical and abstract model of software defined integrated processing



(a) Sketch diagram for basic Cube

(b) Sketch diagram for unit block of integrated processing resource pool

Fig. 1.9 Sketch diagram for space-based and integrated processing resource pool

needed, it may be realized by directly uploading one Cube on the spacecraft and unit block is unnecessary. The unit block may also be added if resource of one block can not meet application requirements.

Integrated processing unit block of open Bus architecture may be integrated with the spacecraft platform with taking function and information applications as the core, to reduce redundancy requirements for volume and weight as far as possible, therefore, to provide space nodes with optimum performance.

For function-platform split software defined multi-function payloads, their software defined capacities are also provide a technical support and space basic operation platform for integrated space-ground space-based information system to realize SDN architecture.

(3) Multi-function ground infrastructure based on integration and resource pool

As constituent part of the space-based information system, ground sub-network bears kinds of nodes. In the stove-piped construction and development mode, each kind of node is constructed independently and fragmentized. Mutual needs and resource can not be shared timely and efficiently. The application of the networked and integrated technology achieves a high integration of ground node resource and space-based resource in the aspect of information, allowing ground network to measure, remote control, and integrated manage the space node.

The function of ground system to reconfigure the special processing resource in the back-end has been applied in some engineering projects. Resource sharing of the signal processing terminal within multiple system configurations has been realized through massive intermediate frequency switch matrix. And the technologies of reconfiguring resources in front-end also have some relevant researches [17]. The application and development mentality for resource reconfiguration have not yet realized that resources from all domains in ground system can participate in it. The resource that can be re-configured is mainly limited at the special signal processing in terminal layer.

The integration of ground node can adopt the same mentality just as that of the space-based node. But considering the difference of bearing platform, the physical

form of RF integration and integrated process will be fixed facing the ground. And the form of vehicle and ship can be defined according to the needs in capability and dimension.

The front end of radio frequency can adopt a variety of antennas, including parabolic antenna with the same diameter, array antenna, and phased array antenna for all space. But all of those need to realize the digital sampling on radio frequency directly. Just like the RF system in Fig. 1.7, RF chain among ground nodes in different places and heterogeneities can be realized by creating links among broadband transmission, switched network, and integrated processing resource pool in the back end.

Resource processing in integrated processing resource pool also adopts the same function-platform split mentality. The method is that taking advantage of multiple common processing platforms to constitute resource groups, and then resource groups constitute resource clusters, in the end, all resource clusters constitute the ground sub-network processing resource pool. In application, resource reconfiguration and invoking can be implemented in the resource group based on the task needs, which can also be conducted among resource groups in resource clusters. The principle of resource reconfiguration and invoking is to optimize the efficiency of resource utilization under the condition of maximizing node capability. According to the form of broadband transmission and switched network, ground integrated sub-network can have different dimensions and capabilities. As shown in Fig. 1.10, RF resource pool in front end and integrated processing resource pool are connected by the point-to-multipoint broadband optical fiber transmission link and multilayer switches, forming a broadband tree-shaped exchange network. RF resource pool in front end and integrated processing resource pool in back end can realize direct links in point-to-point or point-to-multipoint, achieving a certain capability of function based on task requirements.

As shown in Fig. 1.11, both RF resource pool in front end and integrated processing resource pool are hosted on the gridded broadband digital optical fiber network. Broadband digital switching equipment can form a broadband mesh network of complete exchange integrated resource reconfigurable based on IP. Resources in RF resource pool in front end and integrated processing resource pool in back end can exchange and link at any point, build capability required by task, and implement task reconfiguration and switching in real time. Same type of resources in each resource pool can realize digital IP exchange and link, forming a topical combination to improve processing capability.

In the ground processing resource pool, integrated processing platform can adopt devices like FPGA and DSP and modules like multiplexing/de-multiplexing modules and photoelectric/electro-optical conversion cards to form standard signal and information processing board. Rapid IO high-speed network interconnection will contribute to form a single integrated resource processing platform. And ground sub-network should be equipped with appropriate number of integrated processing platforms based on the network capability. Resource group and resource cluster can be formed in the architecture of concentrated distribution or synchronous distribution.

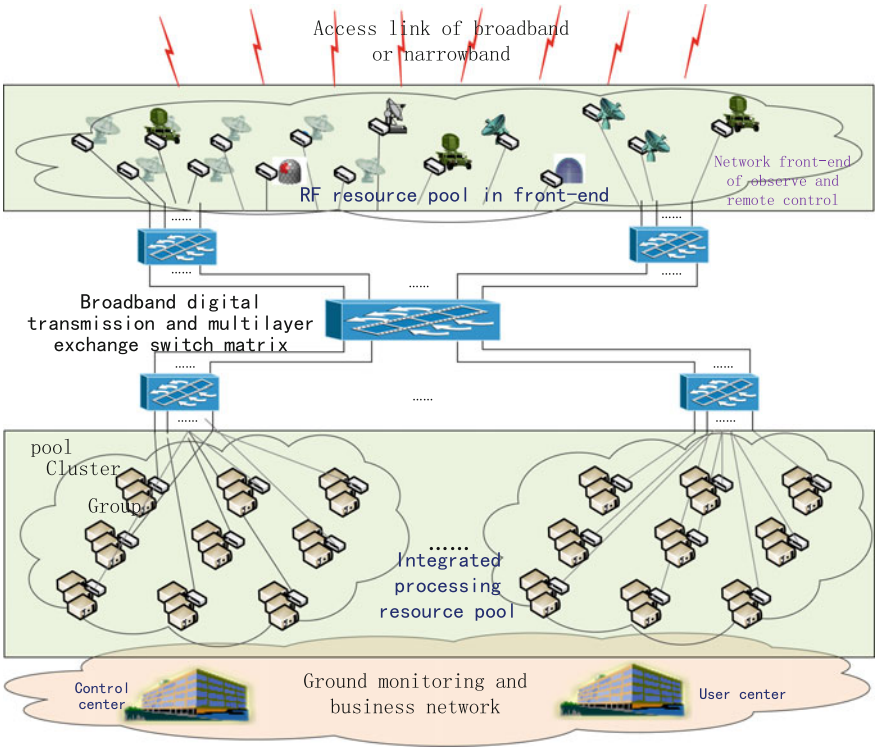


Fig. 1.10 Sketch diagram for ground node resource pool tree-shaped exchange network

With the development of technology, the virtual baseband of common computing platform has been gradually applied into the signal and information processing in various function domains. FPGA + CPU have gradually become the main platform in second generation heterogeneous computing. Real-time processing capability is improved greatly. The possibility has become more and more obvious in adopting virtual integrated processing based on cloud computing and virtual architecture in ground system to realize integrated processing resource.

1.4 Key Technologies Needed to Break Through

Based on integrated common processing platform, combining with the integrated space-ground architecture of SDN technology will contribute to backbone nodes in network possessing full function or multi-function combination capability. Application and control of integrated space-ground networking is the effective technical approach to create integrated space-ground information system.

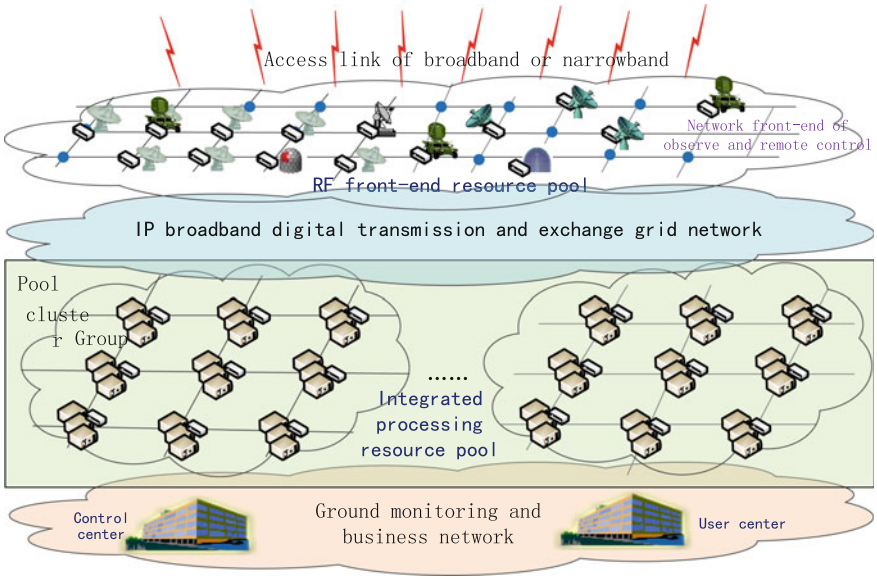


Fig. 1.11 Sketch diagram for ground node resource pool net complete exchange network

However, to realize this goal, a series of technologies need to break through and verify. The main contents are as follows:

(1) Overall technology of space-based information system based on SDN

The introduction of SDN architecture in integrated space-ground information system is closely related to the realization path for technology of each link in the network, which refers to many aspects, including the system architecture design, overall design for node distribution, application requirements, information source, and application pattern. It needs the approach of system engineering investigation to research and demonstrate. The design, application, and perfection of SDN architecture is a scientific, iterative, continuously evolving and progressive process. At the same time, the introduction of function-platform split integrated nodes, its relationship with SDN architecture, and application pattern and so on, all of those technologies need further research and verification.

(2) High integration payload integrated processing basic Cube technology

The aim of the technology is to acquire maximum processing capability at the cost of minimum resource. This will need to combine different pathways, different characteristic overall structures of spacecraft, and application requirements for multifunction, besides, summarize the basic processing capability needed for network nodes in space-based information system. On this basis, the technology will form common basic processing unit module (basic Cube). For this goal, simulate

the common platform performance of system, test and verify the capability and performance of the integrated common hardware platform.

The main difference between space-based multifunction integrated processing platform and integrated electric module of airborne platform is that spacecraft has more severe requirements in the aspects of volume, weight, power dissipation for integrated hardware resources. In addition, spacecraft is more demanding for device performance, processing technology, and algorithm implementation and so on.

(3) Space-based techniques of multi-band multi-function integrated aperture and RF design.

The multi-band multi-function definable design of RF front-end and aperture is closely related to the spacecraft formation. And the design of array units is related to the layout, function and application. So first we should take a full consideration into system functions and planning, and prospectively propose a basic layout and an available scheme which are of common use, then separate common demands from individual demands, finally make a integrated design combined with the spacecraft formation.

The configurable broadband RF T/R front-end technology is combined with radio frequency and programmable technologies, which realize the programmability of radio frequency assembly so as to adapt to various demands of frequency band and broadband application including the high integration multi-band broadband RF front-end technology based on micro system, the multi-band RF chip technology, the broadband RF digital sampling technology, the low delay transmission technology of big data, etc. Among the technologies, we need to research, test and verify these: the interference of channel T/R and the intermodulation isolation under broadband, the isolation technology between array units and channels, the isolation technology between digital and RF, the stability of delay, the controlling of phase noise, the filtering of programmable channel, the high speed digital sampling of broadband, the low delay transmission technology and more.

(4) The intelligent management and control technology of integrated processing resource

We should develop a highly integrated scalable open management and control technology to process resources (set of basic Cubes) based on integration and the dynamic deployment of blueprint. And use this technology to optimize the allocation of all the resources in space-ground processing resource pool oriented toward tasks under the architecture of multi-thread processing and resources management and control which is capable of dynamic loading. It requires multiple functions and the intelligence adaptation of application scenarios.

(5) Space network protocol and the SDN management technology

The core problems of integrated space-ground information system network protocol are to ensure the information is transferred efficiently and stably in an environment of high rate of long delay, error and interruption and the

heterogeneous networks are interconnected with each other [18]. So we should develop and apply the space DTN network technology, the network technology (protocol conversion), the management technology based on the SDN network architecture and the other related technologies, and use the integrated processing resources pool conditions to test the protocol's application effectiveness.

(6) High speed interconnection technology of ground distributed resource pool

The key technology of the ground subnetwork is be interconnected between the RF front-end resource pool and the integrated processing pool by the topological structure of high bandwidth, low delay and low cost. There may be an unexpected delay and delay shaking if the connections among resource pools are of a great number and far away from each other. We should individual the subnetworks and resources into cluster/group according to the future construction scale of the ground subnetwork. And develop a research of the dendriform and reticular exchange or other possible exchange technologies. Thus we could develop an interconnection and exchange architecture that is not only meeting the exchange demands but also functioning well.

(7) The resource pool technology based on real-time cloud computing and virtualization

With the developing of computer technology, it will be the future trend that use the universal computing platform to process resources by signal and information, and the architecture of real time cloud computing and virtualization can help to process the software in real time. The virtualization eliminate the dependency between function service units and the hardware. We should develop a research for the interrupt mechanism in the virtual layer and optimization measures, and minimize the extra cost brought by virtualization to ensure processing in real time.

1.5 Conclusion

The integrated space-ground space-based information system could realize the efficient interconnection of information between space and ground [1, 19, 20]. The architecture of the software defined integrated processing platform and the software defined network based on the integrated RF front-end and Function-platform split software of programmable broadband RF technology is the future trend of the space-based information system which use the software to define all the networked integrated technologies [21]. This article discussed the application suggestions of networking integrated technology for space-ground information system, reviewed the directions of related key technologies and provided the reference thoughts for system constructions.

References

1. Shen R (2006) Some thoughts of Chinese integrated space-ground network system. Eng Sci 8(10):19–30 (in Chinese)
2. Liang M (2010) The inspiration of the rebirth of iridium satellite. Dig Commun World 09: 31–33
3. Wang Q (2012) A new generation of USA missile early warning satellites system and its ability. Nation Def Sci Technol 02:7–12 (in Chinese)
4. Chen X, Chen J (2008) Development of new military communications satellites and future battlefield applications. Aerosp China 7:29–35 (in Chinese)
5. Liu T, Zhang M (2011) The transformational satellite communication system of USA and its development. Satell Netw 5:62–63
6. Yang M, Guo D, Lu L (2012) Study on applications of satellite communications base on global information grid. Wirel Commun Technol 1:37–41 (in Chinese)
7. Cai Y, Wang L, Chen L (2009) New fractionated satellite program of USA and its key technologies. Aerosp Electron Warf 25(4):21–23 (in Chinese)
8. Liu H, Liang W (2010) Development of DARPA's F6 program. Spacecr Eng 9(2):92–98 (in Chinese)
9. Zhang C, Fang Z (2010) Small satellite scenes-responsive space capabilities created by America Department of Defense based on small satellite. Satell Netw 9:46–47 (in Chinese)
10. Fan Z, Zhang Z (2013) Research on New Development of Small Satellites in Foreign Countries. Space International 8:20–29 (in Chinese)
11. Tang J Developing and applying analysis of BeiDou navigation satellites regional system. GNSS World of China (in Chinese)
12. Nunes B, Mendonca M, Nguyen XN et al (2014) A survey of software-defined networking: past, present, and future of programmable networks. IEEE Commun Surv Tutor 16(3): 1617–1634
13. Tennenhouse DL, Wetherall DJ (2007) Towards an active network architecture. ACM SIGCOMM Comput Commun Rev 37(5):81–94
14. Dai B, Wang H, Xu G et al (2014) Opportunities and threats coexist in SDN security. Appl Res Comput 31(8):2254–2262 (in Chinese)
15. Zhang C, Cui Y, Tang H, Wu J (2015) State-of-the-art survey on software-defined networking (SDN). J Softw 26(1):62–81 (in Chinese)
16. Chen Y (2013) Integrated technology of avionics integrated system. National Defense Industry Press, Beijing (in Chinese)
17. Xiao X, Hu J (2016) RF resources reconstruction in TT&C communication system. Telecommun Eng 54(6):707–712 (in Chinese)
18. Zhang J (2011) Space-based mobile communication network. National Defense Industry Press, Beijing (in Chinese)
19. Hu J, Lei L (2010) Technical considerations on an network for air and space vehicles. J Spacecr TT&C Technol 29(5):1–5 (in Chinese)
20. Lei L, Hu J, Zhu Q (2011) Future spacecraft TT&C communication architecture and key technology. Telecommun Eng 51(7):1–6 (in Chinese)
21. Li T, Hu J, Xu H (2016) Application discussion of software defined network in space-based information network. Telecommun Eng 53(3):259–266 (in Chinese)

Chapter 2
Preliminary Discussion on the TTC and Management of Commercial Space in China

Aimin Xu and Guoting Zhang

2.1 Introduction

To the commercial space, there is no unified definition at home and abroad. It is generally accepted that commercial space refers to the use of business, the market model to develop, operate space program, and according to the market rules to carry out space activities [1, 2]. The space programme includes satellite (including spacecraft) services, satellite manufacturing, ground equipment manufacturing, and launching services [3]. The Space activities include commercial rockets manufacturing, commercial satellites manufacturing, commercial launching, commercial space TTC [4]. In recent years, international commercial space improves rapidly, a large number of emerging space enterprises to participate in the development of commercial space, business covers the whole space industry chain, the development of aerospace industry injects new life into the world [5]. Taking The United States for example [6, 7], a number of commercial aerospace companies emerged, including Space Imaging, Orbimage, OneWeb, O3B, which domained in commercial satellites application, and SpaceX, Blue Origin, which engaged in commercial rockets manufacture, commercial satellites manufacture. At present, commercial space in Chinese has just started, which was still in the stage of exploration, a number of commercial aerospace companies emerged as well [8, 9], such as Changchun optical satellite technology LTD., Four-dimensional remote company LTD., which engaged in commercial satellite manufacturing and commercial satellite application, 0–1 space technology LTD., which engaged in commercial rocket manufacturing.

Space TTC mainly completes the rockets, or the satellites tracking, telemetry, remote control and other tasks, which takes the rocket flight state in control, as well as orbit survey and determination of on-orbit satellites, management of running

A. Xu (✉) · G. Zhang
Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China
e-mail: xam23st@sina.com

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_2

state, payload data reception and distribution. Space TTC provides indispensable support for launching and recovery, in-orbit application of satellites [10]. In order to satisfy space engineering of the communication satellites, resource satellites, navigation satellites, ocean satellites, manned space, China has built the ultrashort band, C band and S band space TTC network, which are in appropriate scale, reasonable layout, fully function. To the rise and development of the commercial space of our country in future, space TTC system in China will adjust the service mode to meet the needs of the new task, and provide a more efficient and convenient TTC support service.

Commercial space is the important developing direction of the aerospace industry in the future. This paper analyzes the enlightenment of the foreign commercial space and commercial space TTC to our country in future, and for some of the new commercial aerospace TTC and management questions have carried on the preliminary discussion.

2.2 The Current Developing Situation of Commercial Space and Commercial Aerospace TTC at Home and Abroad

2.2.1 The Current Developing Situation of Commercial Space at Home and Abroad

The United States is the first and the most successful country in the development of Commercial space, who draws up and promulgates a series of relevant laws, regulations and policies, to encourage and support the development of the commercial space, and provide legal protection for commercial space activities. From 1984 to 2015, the United States has promulgated more than 20 laws, regulations and policies, covering space policy, space launch, manned space, satellite communications, satellite navigation, satellite remote sensing, etc. Under the stimulus of national policy, the commercial space industry chain all full blossom, especially in the satellite remote sensing application, satellite communication network and orbital transportation.

Space imaging company, which separated from Lockheed Corporation in October 1994, aimed at making the world's most advanced earth observation technology commercialized, to provide users around the world digital earth information products and services. IKONOS-2 satellite launch success and realize the business operation on September 24, 1999, begun to provide more than 1 and 4 m panchromatic image spectral resolution. This is the world's first commercial operation of high resolution remote sensing satellite, since then opened up a new era of civil high resolution earth observation. The company also has sole global franchise of "Landsat digital image".

Orbit imaging company was founded in November 1993, which is in the domain of design, construction, management and sales of a wide range of products and services related to the space. The company currently has three OrbView satellite, which have sold in the United States the rights of the other satellite systems in remote sensing images, including Canada RADARSAT-1, 2 of synthetic aperture radar platform, and SPIN-2 satellite of Russia. In 2005, orbital imaging company acquired the Space Image's assets [11].

One web company was founded in 2012, is committed to launch small satellites in low earth orbit to form a communication network. 648 small satellites plans to launch, whose altitude is about 200–2000 km, complete the Internet coverage of local area, and provide about 50 Mbps internet speed. OneWeb company plans to launch its first satellite in 2017.

Spire company was founded in 2013 [12], is committed to deploy meteorological satellite network. Spire plans to launch 100 satellites in MEO, by measuring the GPS satellite signal to obtain atmospheric meteorological data, such as temperature, pressure and humidity.

Space exploration technology company (SpaceX) in the United States was founded in 2002 [13], is committed to developing recyclable rocket, and manned spacecraft. Currently SpaceX developes partly reusable falcon 1 and falcon 9 rockets, and the sequence of Dragon spacecraft. In October 2012, the dragon spacecraft was raised to send the goods to the international space station. In March 2015, the falcon 9 rocket will be the first all-electric communications satellite into orbit. In December 2015, falcon 9 rocket goes up and 1st stage recoveries successfully. In April 2016, 1st stage of falcon 9 rocket successfully landed on a maritime unmanned ship.

Blue origin company was founded in 2000 [14], set up by the world's largest online retailer amazon, its main products include engine and suborbital rocket. In November 2015, the orbiter launch New Shepard to an altitude of 100 km, then the rocket intact landing to the predetermined ground position, realized the reusable suborbital vehicle.

Other countries such as Canada, Argentina's aerospace business company has its own satellite launch and operation plan. Canada SkyboxImaging company was founded in 2009, mainly to provide real-time satellite imaging services platform, image accuracy is lower than 1 m. As of March 2014, a total of 24 small satellite SkySat in orbit. Argentina Satellogic company was founded in 2010, mainly to provide real-time earth image data. As of June 2014, three LEO satellites are running in orbit, which will increase to 300 satellites in future. Europe also increased support for the development of commercial space through legislation and policy. In December 2015, the British issued "the national space policy," [15] aimed to become center of Europe commercial space, and occupy a larger share of space market in the world.

At present, the traditional plan mode of "national investment, enterprise development" is predominated in the aerospace industry, the commercial space has just started in China. In recent years, the Chinese government encourages private companies to get into space business, especially when the time the Belt and Road

initiative was put forward since September 2013. The development of aerospace industry especially commercial space industry, brings better service to the national strategy. In October 2014, prime minister Li Keqiang, at a state council executive meeting, proposed to gradually guide the folk capital to participate in the national civil space facilities. In November 2014, the state council announced the 60 document, more specifically to encourage private capital into the spatial domain. In October 2015, the medium and long-term plan for the national civil space infrastructure (2015–2025) is put forward [16], to encourage military and civilian integrated capital into space, base areas are opened by space military-into-civilian as start. Under the encouragement of national policy, lots of innovative enterprises have sprung up which focused on commercial rockets manufacture, or commercial satellites manufacture.

Changchun optical satellite technology LTD., established in December 2014 [17], is funded by Changchun Optical Institute and other shareholders, which focused on manufacturing commercial remote sensing satellites. Company plans to launch 60 video satellites before 2020, and 138 video satellites before 2030.

Four-dimensional remote company was established in September 2015 [18], funded by the aerospace science and technology group, will provide global users with high spatial resolution, high temporal resolution, high spectral observation ability of all-weather of remote sensing for earth observation data services. Company plans to be completed around 2022 by 16, 0.5 m resolution optical satellites, 4 high-end optical satellites, 4 microwave satellites, and some video, hyperspectral satellites.

0–1 space technology LTD., was established in August 2015 [19], is funded by private shareholders as a commercial rocket manufacturing company. The company aims to develop, design, and final assembly of low-cost small vehicles. Its first flight will be carried out in 2018.

2.2.2 *The Current Developing Situation of Commercial Space TTC at Home and Abroad*

At present, space TTC resources, including ground station and control center, have three kinds of subordinate relations [20, 21]:

1. as a national infrastructure, directly by the government and/or military management, such as the United States NASA's tracking and data network (STDN), the U.S. air force satellite control network (AFCSN) acquired, ESA TTC network, the Russian state TTC network.
2. ground stations networking, belongs to an international organization, such as International Maritime Satellite Organization is a global mobile satellite communication of intergovernmental cooperation mechanism, namely the international mobile satellite organization, which domestic habitually call it maritime satellite for short. Maritime satellite organization reformed as an international

business company in 1999, namely the international mobile satellite. Maritime satellite organization has two control center, with main sations in City Road, backup stations in Boreham wood, and eight stations distributed in Italy, China, Canada, New Zealand, the Netherlands, Norway, Russia, the United States. So far, maritime satellite organization manages 11 satellites.

3. a large company has a corresponding ground station and control center, is divided into two kinds, one kind is that the company has its own satellite, established corresponding ground station network, such as China Fengyun, China Satcom. The control center of China Fengyun was built on the National Weather Service, which takes control of four stations in Beijing, Guangzhou, Urumqi and Australia, and is in charge of 5 low-earth orbit and geostationary orbit satellites control. China Satcom operation control center was built in the northwest of Beijing, which takes control of four stations such as Beijing, Hong Kong, Chengdu, and is in charge of 6 geosynchronous orbit satellites. This kind of networks support only professional satellite application, do not support other civilian satellite TTC. Satellite application oriented specific service objects, which belongs to space commercialization. Another kind is that to establish a station network, which provides the service for all satellites, belongs to commercial space TTC. At present, some of the domestic commercial company plans to build its own space TTC network. Some company has certain influence on the international commercial TTC, such as PrioraNet global station network of Sweden Space (Swedish Space Corporation, SSC) and Norway's Satellite service company (Kongsberg Satellite Services, KSAT).

PrioraNet global station network of Swedish Space company(SSC) is composed of the network management center (NMC) [22], SSC's own core station and cooperative stations all over the world with strategic position. The ground stations of SSC are divided into two kinds, one kind is that belongs to Sweden space center, and another kind is cooperation with international partners, who provide C/S, L, X, UHL spectrum TTC services. The NMC is a distributed system, which located in Newport Beach, California, Kiruna, Sweden and Pennsylvania Horsham, which can provide users with a single point of interface in the global earth station. In the whole stages of missions, spacecraft and network operators to provide services and support round the clock. 10 its own core stations, and 6 cooperative stations, distributed in Sweden, Italy, Germany, Australia, Canada, the United States, Chile, India, Mauritius, South Africa, Antarctica, and other countries and regions. PrioraNet network provides reliable solutions for global satellite communication and operation, which provides the services including task management services, mission control services, TT&C services, load data retrieval services. The major customers of PrioraNet global station network include Boeing satellite systems corporation, China aerospace science and technology group LTD., China resources satellite application center, California's space agency, China satellite TTC systems, the geographic eye technology, geographic information from Thailand and space agency, NASA, Taiwan space center, etc. In 2009, the Universal Space Network (USN) bought out by SSC, USN company adopted the mode of commercial

operation to provide satellite operations (including real-time telemetry, tracking, and commanding) services.

Space TTC network of Norway's Kongsberg Satellite Services (KSAT) is composed of the Tromsø Network Operations Centre(TNOC) and four ground stations [23]. TNOC takes control of Tromsø satellite ground stations and TT&C antenna system of Svalbard, and works closely with the owner and the operator of satellite, which is responsible for all operations. Four stations include Svalbard ground station, Tromsø ground station, Grimstad and TrollSat ground stations. This network provides TT&C and data services mainly for polar orbit satellite, which includes telemetry, TTC, TTC support for sounding rockets, launch and early orbit (LEOP) TTC support, Global data dump or on-orbit support. The main client of the network includes: NASA, ESA, NOAA, IPO, JAXA, DigitalGlobe and so on.

2.3 Enlightenment of Commercial Space TTC to Our Country

2.3.1 The General Trend of TTC and Management Development is Diversified

From the perspective of the current foreign developing situation, the coexistence and coordinated development of military/civil space TTC, commercial space TTC is the trend of the times. Military/civil space TTC major services military and civilian satellites, which makes more services for commercial satellites when redundancy. For example, SpaceX's "the falcon 9" tipped the "dragon" spacecraft from military launch pad in Florida, its launching TTC is operated by AFCSN 45th space operation team, "dragon" spacecraft entered orbit outside the international space station a mile run by ferry companies themselves, into the international space station within a mile of the approximation, docking, and out of control was conducted by NASA. Commercial space TTC services mainly for commercial satellites, which can provide service for worldwide space users, either through business model to make its own global ground stations networking, or through business model to cooperate with foreign ground stations by networking operation. Such as SSC's PrioraNet global ground station network, Norway's KSAT provide TTC supports for lots of global commercial satellites.

Both military/civil space TTC and commercial space TTC, the management mode is just the same, composed of mission center and ground stations. Due to the earth synchronous orbit satellite visible to a ground station for 24 h, one single station can be used to manage the satellite. For commercial geosynchronous orbit satellites, commercial companies can accomplish TTC in its own application business simultaneously. Because of the influence of the earth curvature, to meet TTC requirements of the full range in low earth orbit satellites or rockets, multiple ground stations must be distributed in different locations connected in the network

to relay TTC task. The network can be under a TTC system of multiple ground stations, or under different TTC systems. For commercial satellites in LEO or MEO orbit, the investment of a commercial company self-built the network is too high, two more reasonable ways are: 1. to purchase commercial company's TTC service, with global TTC ability; 2. to use of national support network to complete the TTC.

2.3.2 *It is Imperative to Moderate Encourage the Development of Commercial Space TTC*

At present, the main TTC mode of commercial companies abroad [24–28] are shown in Table 2.1.

From present developing situation foreign, there are two main types of commercial space TTC mode: 1. building their own site, such as the United States's Orbimage, One Web, Spire. 2. purchase a TTC service or completely entrust other units to complete the TTC, Such as launching and orbiting TTC of spaceX, launching TTC of Blue Origin. From the situation abroad, orbiting TTC mainly be completed by satellite users or satellite manufacturers. For our country, space industry is open to the commercial capital, inevitablly space TTC must be open to commercial capital. If the developing pattern of commercial space TTC looks abroad, on the one hand, the cost of TTC services will be decreased, the efficiency of TTC services will be improved, on the other hand, it helps to better participate in international competition, or better international cooperation.

From domestic developing situation, the rise of the future commercial space can lead to the development of commercial satellite in explosive growth momentum.

Table 2.1 The main TTC mode of commercial companies abroad

Commercial companies	Main business	TTC mode
Orbimage	Commercial remote sencing satellite application service	Stations built by self, data processing center and more than 10 stations, which can control satellites,and receive remote sensing data
OneWeb	Internet satellite application service	Stations built by self, plans to build lots of Ku band stations
Spire	Meteorologic satellite application service	20 stations built by self, to receive satellite data
SpaceX	Orbital transportation service	Control center built by self, purchase TTC service launching TTC bought by military service, on orbit running controlled by self and by NASA, on orbit running bought by commercial service
Blue origin	Suborbital transportation service	Launching TTC bought by military service, other is unknown

Future commercial space is given priority to with the application of small satellite constellation, small satellite constellation control has the following characteristics: 1. large number of on-orbit space targets, with wide spatial distribution. Satellites inbound or outbound frequently, number of satellites have short interval transit time, requiring a close "all-weather" TTC. 2. a satellite orbit, such as satellite remote sensing application are mainly concentrated in orbit height 500–800 km sun-synchronous orbit, large number of targets inbound or outbound thick and fast, required to provide multiple targets for TTC service simultaneously. Existing TTC resources are close to saturation, which appropriately support only a part of commercial space TTC requirements. It is imperative to moderate encourage the development of commercial space TTC.

2.4 Preliminary Discussion on the TTC and Management of Commercial Space in China

2.4.1 Strengthen Space Legislation

Our country should make space rules at national level, and industry standard level, to regulate commercial behavior and space activities in commercial space and commercial space TTC [29], and to promote the orderly development of commercial space and commercial space TTC. Space activities should be under the legal framework, otherwise may affect subsequent commercial space development. Commercial space needs more space under the rule of law. Legislations should be put out to regulate specifications and requirements of military space, civil space and commercial space activities in our country, in order to ensure the orderly development of aerospace industry (including TTC). Through space legislation, commercial space market access or exit mechanism, fair competition mechanism, insurance and compensation mechanism, safety supervision mechanism, etc. are established, to create an orderly, healthy competition market environment.

2.4.2 Explore the New Commercial Space TTC and Management Model

Commercial company is Encouraged to build mission center, or stations, which protect commercial satellite orbit control, and actively explore the new commercial aerospace TTC management model.

1. support external connection to satellite mission center

External satellite mission center is supported to run with existing satellite TTC center, Satellite TTC center transparently send control instructions to commercial

space TTC ground station by satellite mission center. The health status of commercial satellites can be directly sent to satellite TTC center. The commercial satellite TTC process and interface must be standardized, to support the external network.

2. strengthen international cooperation of commercial space TTC ground station

Commercial space TTC builds their own station at home, makes international cooperation overseas. Domestic and overseas stations can interact with each other.

2.4.3 *Enhance the Management of Commercial Satellite Services*

For all of the satellites, including commercial satellites, strengthen policy management, good services, on the one hand, due to protect national security need to consider, on the other hand due to ensure the safety of commercial satellite and other satellite orbit. On-orbit satellite management services shall include the following aspects:

1. the frequency management services

Satellite frequency become important strategic resources. To top the satellite TTC, digital frequency resources unified planning and design, harmonious and unified management, in order to prevent the on-orbit satellite radio interference between each other.

2. the orbital resource management services [30]

Satellite orbital gradually become scarce strategic resources. For satellite orbital resources unified management, according to the national security, different requirements in terms of hierarchy, such as the national economy needs conflict level according to the distribution of rail.

3. catalogue management services

With the increasing of space target quantity, space target catalog, collision warning is becoming more and more important. All satellites are unified to detect, track and recognise, the trajectory characteristics, geometric characteristics and physical properties of space target is unified for catalogue management, and to provide commercial launch, commercial satellite application collision warning service.

4. launch and recovery services

Satellites enter into space, and return from space to ground, space safety is a matter of territory. The satellite launch and recovery should be unified management services, highlighting the launch and recovery of safety design.

2.5 Conclusion

Commercial space technology development derived from military space, civil space, but with different development model. The arrival of the commercial space puts forward the new requirements, which requires that we should not only by law of space, space concept, and also inject new ideas, new management idea. In this paper, the parts of the commercial space TTC management questions have carried on the preliminary analysis and discussion, the more problems, more in-depth analysis needs to be further discussed in accordance with the commercial space development and progress in China.

References

1. Liu YF (2015) The concept of space commercialization and commercial space. J Satell Netw 2015(8):20–24 (in Chinese)
2. Liu YF (2015) Commercial space forces change of space commercialization. J Satell Netw 2015(8):4–5(in Chinese)
3. Zhang ZH, Bai MS, Shi Y (2015) Development and Enlightenment of foreign commercial space. China Aerosp 11:31–39 (in Chinese)
4. Fan C (2015) Analysis of Foreign aerospace company business model developing trend. J Satell Appl 2015(11):41–43 (in Chinese)
5. Huang ZC (2016) Commercial space is the new power of USA space exploration. J Int Space 2016(5):64–68 (in Chinese)
6. Chen HP, Zhang B, Rong Y (2013) Review of USA space commercial transportation development. China's space 9:32–38
7. Wu Q, Zhang MT (2016) Analysis of USA commercial space development. J Int Space (5):6–11 (in Chinese)
8. Liu YF (2015) The prospect of China's space commercialization and commercial space. J Satell Netw 2015(8):30–33 (in Chinese)
9. Liu YF (2015) Concerned in China's space commercialization and commercial space. J Satell Netw 2015(8):26–28 (in Chinese)
10. Lu XY, Huang LJ, Wang S (2015) Stability analysis of TTC network with dependent properties. J Spacecr TT&C 2:147–152 (in Chinese)
11. WorldView-4 (2016) Satellite sensor (0.31 m). http://www.satimagingcorp.com/satellite-sensors/geoeye-2/.Cited. Accessed 18 Sept 2016 (in Chinese)
12. Spire company will use the miniature satellites for accurate weather forecasts. http://www.wtoutiao.com/a/1508266.html.Cited. Accessed 18 Sept 2016 (in Chinese)
13. SpaceX (2016) https://en.wikipedia.org/wiki/SpaceX.Cited. Accessed 18 Sept 2016 (in Chinese)
14. Blue Origin (2016) https://en.wikipedia.org/wiki/Blue_Origin.Cited. Accessed 18 Sept 2016 (in Chinese)
15. Gong R (2016) NOAA first published "commercial space policy". J Satell Appl (7):48–50 (in Chinese)
16. National civil space infrastructure for medium and long-term development plan (2015–2025). http://www.sdpc.gov.cn/zcfb/zcfbghwb/201510/W020151029394688578326.pdf.Cited. Accessed 18 Sep 2016 (in Chinese)
17. Chang Guang Satellite Technology LTD. (2016) http://www.charmingglobe.com/.Cited. Accessed 18 Sept 2016 (in Chinese)

18. Four-dimensional remote company approved by launching remote sencing satellite on high resolution level of 0.5 meters. http://news.qq.com/a/20151030/034089.htm.Cited. Accessed 18 Sept 2016 (in Chinese)
19. Feelings Entrepreneurship! 0–1 Space Technology LTD. Wants to be SpaceX of China. http://tech.sina.com.cn/i/2016-01-20/doc-ifxnqriy3199852.shtml.Cited. Accessed 18 Sept 2016 (in Chinese)
20. Wei CX (2010) Present situation and prospects of American aerospace TT&C resources. In: National space detecting academic communication' 2010, Xiamen, China, pp 214–218 (in Chinese)
21. Wang G, Wu XY (2010) The structure and development of American TTC system. J Def Technol 31(5):87–91 (in Chinese)
22. Wang Q (2011) PrioraNet global station network deep. In: National space detecting academic communication' 2011, Xi'an, China, pp 119–123 (in Chinese)
23. Guo LH, Huang XW (2012) Overview of satellites TTC network of Norway's satellite services company acquired national conference on space exploration. In: National space detecting academic communication' 2012, Xi'an, China, pp 119–123 (in Chinese)
24. SpaceX mission to space station How it'll work. http://www.space.com/15482-spacex-dragon-space-station-flight.html.Cited. Accessed 18 Sept 2016 (in Chinese)
25. Mission Control Center during SpaceX Rendezvous Operations. http://www.nasa.gov/image-feature/mission-control-center-during-spacex-rendezvous-operations/.Cited. Accessed 18 Sept 2016 (in Chinese)
26. Commercial space strucks with force? Analysis of American commercial space development (Part one). http://www.godeyes.cn/html/2016/07/25/google_earth_16474.html.Cited. Accessed 18 Sept 2016 (in Chinese)
27. Commercial space struck with force? Analysis of American commercial space development (Part two). http://www.godeyes.cn/html/2016/07/25/google_earth_16475.html.Cited. Accessed 18 Sept 2016 (in Chinese)
28. The Legend of SpaceX: the American private sector beyond China's space. http://war.163.com/11/0515/17/7444ULR600014J0G.html.Cited. Accessed 18 Sept 2016 (in Chinese)
29. Zhang ZJ (2015) Thinking about the laws of the Chinese commercial space. China's Space 2015(12):8–12 (in Chinese)
30. Zhang BX, Ju L (2010) Research about development of TTC technology before 2030. J Aircr Meas Control 29(5):11–15

Chapter 3
The Deputy Reflector Control Technology of the Large Deep Space Antenna

Lujian Zhang

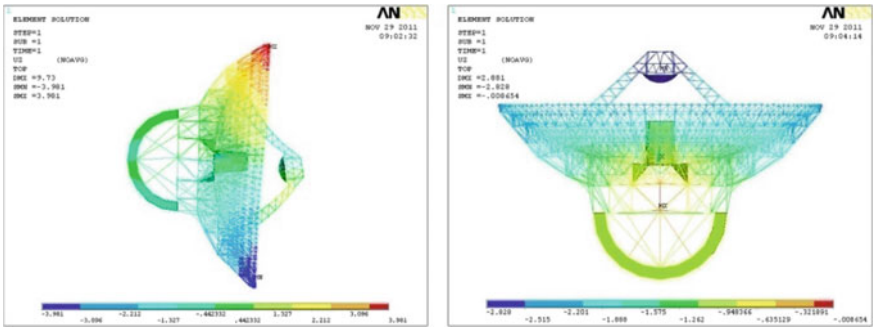
3.1 Introduction

Due to its own gravity [1–3], temperature field gradient and wind power on the antenna surface [4] and so on, the main and deputy reflector and the supporting frame of the deputy reflector of the Deep Space Antenna occur deformation. The deformation of the antenna makes antenna focus shift, thereby reduces the antenna gain even interrupts microwave transmission path [5–8], and ultimately reduces the system performance even does not work. The antenna structure deformations and electrical properties [9] are given can be seen from Fig. 3.1 and Table 3.1, as the main and deputy reflector deformations due to gravity and other factors, resulting in the antenna gain and efficiency decreased, and the larger antenna aperture, the higher antenna operating frequency, the greater loss of the gain and efficiency. So for large-diameter and high-frequency antenna, the deformations of the main and deputy reflector must be effectively addressed.

In order to compensate the performance loss caused by the deformation of the antenna, the method often used the pre-adjustment in the antenna installation or the best parabolic fitting method according to the existing deformation [1, 2]. The pre-adjustment cans not meet the requirement of electrical performance for deep space antenna, so the curve fitting method is used to fit a new parabola according to the deformation of the antenna. The center of the deputy reflector is adjusted to the focus of the new parabolic, so that the electrical performance is guaranteed to the optimal. Compared with the theoretical parabolic shape, with the change of the antenna pitch angle, the shape of the actual parabolic reflector is gradually changed, the position of the focus is gradually changing, which requires the real-time adjustment of the position and attitude of the deputy reflector.

L. Zhang (✉)
The 39th Institute of China Electronic Technology Group Corporation, Xi'an 710065, China
e-mail: nwiee@sina.cn

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_3



(a) El=0deg The structure deformation **(b)** El=90deg The structure deformation

Fig. 3.1 The structure deformation of the large antenna

Table 3.1 The relationship between antenna gain loss and pitch angle for gravity deformation

Band El	L-band	S-band	C-band	X-band	Ku-band	K-band	Ka-band
80°	−0.0023	−0.0050	−0.0307	−0.0604	−0.1785	−0.4675	−1.0439
70°	−0.0018	−0.0040	−0.0248	−0.0489	−0.1446	−0.3786	−0.8455
60°	−0.0016	−0.0035	−0.0215	−0.0424	−0.1253	−0.3282	−0.7328
50°	−0.0015	−0.0033	−0.0205	−0.0404	−0.1194	−0.3125	−0.6979
40°	−0.0016	−0.0035	−0.0214	−0.0421	−0.1245	−0.3260	−0.7279
30°	−0.0018	−0.0039	−0.0237	−0.0467	−0.1380	−0.3614	−0.8070
20°	−0.0020	−0.0044	−0.0270	−0.0532	−0.1571	−0.4114	−0.9188
10°	−0.0023	−0.0050	−0.0308	−0.0607	−0.1793	−0.4695	−1.0485

Since the deputy reflector antenna structure size is relatively large, to ensure reliable installation deputy surface, reduces the weight and gravity deformation of the deputy surface adjustment mechanism, to ensure that the scope of deputy reflector adjustment, real time and accuracy of the antenna, the large Deep space deputy reflector antenna adjustment mechanism with six degree of freedom (6-DOF) parallel mechanical adjustment mode [10] (Fig. 3.2). The 6-DOF parallel mechanisms are composed of the Base platform (lower platform), the mobile platform (upper platform) and the six electric cylinders (linear electric actuators). As the deputy reflector placed on the platform, so the attitude of the platform determines the attitude of deputy reflector. Each motion electric cylinder is controlled by a separate motor. The electric cylinder drives by AC servo motor, and links through Hooke hinge and spherical hinge with upper and lower platform. The movement of the motor is transmitted to the electric cylinder screw, by synchronizing belt. The rotation is converted to an electric cylinder telescopic movement by the screw, so as to realize the control of the 6-DOF.

(a) The main and deputy reflector structure

(b) The 6- DOF parallel mechanisms

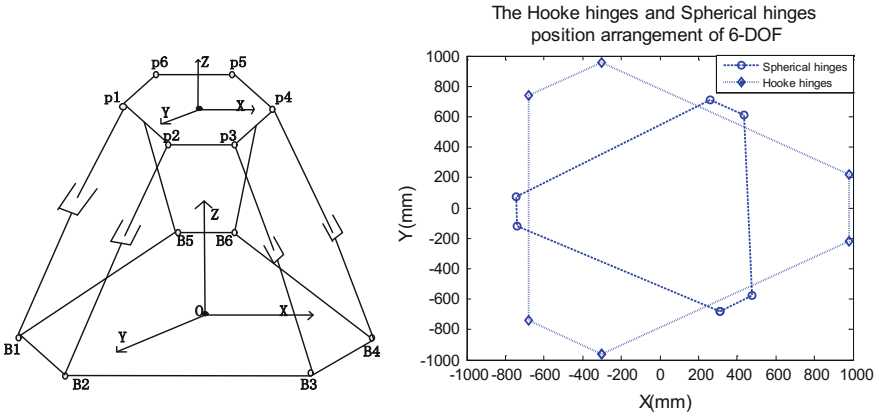
Fig. 3.2 The large antennas and deputy reflector adjustment mechanism

3.2 The Mathematical Model of Deputy Reflector Parallel Mechanism

It is necessary to establish the mathematical model of the 6-DOF parallel mechanism and analyze the kinematics in order to achieve the control of the deputy reflector by 6-DOF parallel mechanism. Kinematic analysis is the theoretical basis for the control of parallel mechanism, and it is also the premise of the dynamic analysis of the rigid body and the analysis of the characteristics of the electric servo system. Through the analysis of the mathematical model of parallel mechanism, the six joints can carry out cooperative movement and realize the multi-dimensional adjustment of the deputy reflector. The mathematical model of 6-DOF parallel mechanism based on joint control is shown in Fig. 3.3, that the deputy reflector is mounted on the moving platform.

3.3 The Coordinate of 6-DOF Parallel Mechanism [11, 12]

The space coordinate system is established according to the parallel mechanism model of Fig. 3.3. The global coordinate system O-XYZ is established with the center of the deputy support and the deputy reflector as the origin, while the local coordinate system p-xyz is established with the center of the deputy reflector as the origin. In the global coordinate system P_{i}, B_{i} (i = 1,…,6) respectively represent Spherical hinge and Hooke hinge Center. In local coordinate system p-xyz, p_{i} represents the position vector of the center point of each Spherical hinge. p represents the position of the center of the deputy reflector in the global coordinate



(a) The configuration of the 6-DOF parallel mechanism

(b) The arrangement points of Hooke hinges and Spherical hinges

Fig. 3.3 The configuration of the 6-DOF parallel mechanism and the arrangement of points

system O-XYZ, respectively, around the antenna deputy platform x, y, z-axis rotation angles. Define the global coordinate system and local coordinate system to meet the right-hand rule.

In order to make the main and deputy reflector motion coherence, the coordinate systems of the main and deputy reflector are needed to be normalized. Define the global coordinate system X axis orientation parallel to the direction of the antenna pitch axis, Y axis orientation parallel to the direction of the yaw axis(antenna overturned), Z direction consistent with the direction of the beam, α the same as the pitching rotation direction, β orthogonal the pitch axis, γ rotating around the electric axis.

3.4 The Position Inverse Solution of 6-DOF Parallel Mechanism [12, 13]

The position inverse solution of the parallel mechanism is that by the deputy reflector positions and attitudes of each posture coordinate to calculate the coordinate of each joint. The position inverse solution is full use of coordinate conversion, the position and attitude vector $\boldsymbol{op}_i$ from the local coordinate system conversion in the global coordinate system. In the fixed coordinate system, the position and attitude parameters of the moving platform, respectively, are the central positions of the moving platform (x, y, **z**) and the rotation angles around the x, y, z. Given the position and attitude in the workspace platform, find the length

of each leg, which is the position inverse solution of Parallel Mechanism. The inverse solution of position can be solved directly by spatial mechanism, and it is closed and unique.

According to the Euler theorem, the matrixes $\boldsymbol{R_x}, \boldsymbol{R_y}$ and $\boldsymbol{R_z}$, respectively, are coordinate rotation matrixes with respect to the angle of α, β and γ.

$$\boldsymbol{R_x} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix}, \boldsymbol{R_y} = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix}, \boldsymbol{R_z} = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3.1}$$

The rotation transformation matrix that the local coordinate system $\mathrm{p} - \mathrm{xyz}$ with respect to the global coordinate system $\mathrm{O} - \mathrm{XYZ}$ is:

$$\begin{aligned} \boldsymbol{R} &= \boldsymbol{R_x}\boldsymbol{R_y}\boldsymbol{R_z} \\ &= \begin{bmatrix} \cos\gamma\cos\beta & \cos\gamma\sin\beta\sin\alpha - \sin\gamma\cos\alpha & \cos\gamma\sin\beta\cos\alpha + \sin\gamma\sin\alpha \\ \sin\gamma\cos\beta & \sin\gamma\sin\beta\sin\alpha + \cos\gamma\cos\alpha & \sin\gamma\sin\beta\cos\alpha - \cos\gamma\sin\alpha \\ -\sin\beta & \cos\beta\sin\alpha & \cos\beta\cos\alpha \end{bmatrix} \end{aligned} \tag{3.2}$$

The relation between $\boldsymbol{P}_i$ and $\boldsymbol{p_i}$:

$$\boldsymbol{P}_i = \boldsymbol{R} \cdot \boldsymbol{p_i} \tag{3.3}$$

In any a branched chain as the research object, the vector of the *i*th leg is in the global coordinate system:

$$\boldsymbol{L_i} = \boldsymbol{p} + \boldsymbol{R} \cdot \boldsymbol{p_i} - \boldsymbol{b}_i \tag{3.4}$$

where, $\boldsymbol{p}$ are the central positions of the deputy reflector in the global coordinate system, and $\boldsymbol{b}_i$ are the positions of the upper platform nodes in the global coordinate system.

So the length of the *i*th leg is:

$$L_i = \sqrt{\boldsymbol{L}_i^T \boldsymbol{L}_i} \tag{3.5}$$

When the central coordinates of the spherical hinge and Hooke hinge are known, the length of the supporting legs can be obtained through the position and attitude of the deputy reflector, and the inverse solution of the 6-DOF parallel mechanism is completed.

3.5 The Position Forward Solution of 6-DOF Parallel Mechanism [14–17]

The position forward solution of 6-DOF parallel mechanism is to solve the position and attitude of the deputy reflector with the known drive legs length L. Due to the complex structure of the parallel mechanism, the position forward solution is more complicated and more difficult. Relative to the inverse position solution [15], Many scholars from the numerical and analytical study, such as genetic algorithm, neural network, Newton Rapson iterative algorithm, but the calculations process are complicated, the computation time are time-consuming, and the positions are not unique and so on. For the deputy reflector control, in order to achieve effective control, the requirements are that the positive solution must be unique, and the solving speed must be very fast. Based on the steepest descent principle, this paper presents an algorithm for the position forward solution of the parallel mechanism. Under the premise of the continuity of the motion of the parallel mechanism, according to the Jacobi variation relation between working space and joint space of parallel mechanism, and the initial point of iteration is taken as an arbitrary point in space, the algorithm can accurately and quickly give the unique position of the parallel mechanism. Due to the deputy reflector movement in its work space, after the control it from any known initial position $\boldsymbol{P}^0$ to reach the current position $\boldsymbol{P}^d$, then length of each leg is $\boldsymbol{L}^d$. In the actual work process according to $\boldsymbol{P}^0$ and $\boldsymbol{L}^d$ to solve the end position $\boldsymbol{P}^d$, which $\boldsymbol{P}^0$ according to the actual calibration to determine, $\boldsymbol{L}^d$ is collected by the linear sensor.

Make $\boldsymbol{q}_i = \boldsymbol{R}\boldsymbol{p}_i$, the type (3.4) both ends are respectively about the time derivative:

$$\dot{\boldsymbol{L}}_i \boldsymbol{u}_i + \boldsymbol{L}_i \dot{\boldsymbol{u}}_i = \dot{\boldsymbol{p}} + \dot{\boldsymbol{q}}_i + \dot{\boldsymbol{b}}_i \tag{3.6}$$

Wherein, $\boldsymbol{u}_i = \boldsymbol{L}_i / \|\boldsymbol{L}_i\|$, the robot kinematics knowledge:

$$\dot{\boldsymbol{q}}_i = \frac{d}{dt}(\boldsymbol{R}\dot{\boldsymbol{p}}_i) = \boldsymbol{R}\dot{\boldsymbol{p}}_i + \boldsymbol{\omega} \times \boldsymbol{q}_i \tag{3.7}$$

Wherein, $\boldsymbol{\omega}$ are the rotational speeds, which are around the three axes X, Y and Z. Type (3.7) to (3.6) and taking into account $\dot{\boldsymbol{b}}_i = \dot{\boldsymbol{p}}_i = \boldsymbol{0}$, then

$$\dot{\boldsymbol{L}}_i \boldsymbol{u}_i + \boldsymbol{L}_i \dot{\boldsymbol{u}}_i = \dot{\boldsymbol{p}} + \boldsymbol{\omega} \times \boldsymbol{q}_i \tag{3.8}$$

The type (3.8) both ends are multiplied by vector $\boldsymbol{u}_i^T$:

$$\dot{\boldsymbol{L}}_i = \boldsymbol{u}_i^{\mathrm{T}} \dot{\boldsymbol{p}} + (\boldsymbol{q}_i \times \boldsymbol{u}_i)^{\mathrm{T}} \boldsymbol{\omega} = \left[\boldsymbol{u}_i^{\mathrm{T}} \ (\boldsymbol{q}_i \times \boldsymbol{u}_i)^{\mathrm{T}}\right] \begin{bmatrix} \dot{\boldsymbol{p}} \\ \boldsymbol{\omega} \end{bmatrix} \tag{3.9}$$

The Jacobi matrix of the adjustment mechanism is [16, 17]:

$$\boldsymbol{J} = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_6 \\ \boldsymbol{q}_1 \times \boldsymbol{u}_1 & \boldsymbol{q}_2 \times \boldsymbol{u}_2 & \cdots & \boldsymbol{q}_6 \times \boldsymbol{u}_6 \end{bmatrix}^{\mathrm{T}} \tag{3.10}$$

Type (3.10) to (3.9):

$$\begin{bmatrix} \dot{\boldsymbol{p}} \\ \boldsymbol{\omega} \end{bmatrix} = \boldsymbol{J}^{-1}\dot{\boldsymbol{L}}_i \tag{3.11}$$

Therefore, the increments of the deputy reflector position and attitude are:

$$\Delta \boldsymbol{p}^k = \boldsymbol{J}^{-1}\dot{\boldsymbol{L}}_i \tag{3.12}$$

The position and attitude increments of the type (3.12) are added to the initial position and pose of the deputy reflector, and the iterative calculation is carried out to meet the accuracy of the algorithm. In this way, the actual position and attitude of the deputy reflector are obtained, and the position forward solution of 6-DOF parallel mechanism is calculated.

3.6 The Electric Control of the Deputy Reflector Mechanism

To ensure the synchronization and the real-time update between multiple axes, the control mechanism system of the deputy reflector adopt full digital multi-axis synchronous motion control network. This mechanical design reduces the cost of manufacturing and services, greatly reducing the development time of the hardware. The topological structure between motion controllers and multiple servo drives, makes it easier to implement the algorithm and the execution mechanism.

6-DOF parallel mechanism is a multi-axis motion control system that composed of six driving branches. Usually the dynamic response characteristics of each axis is different, especially in the high speed profile control will cause significant errors, it is necessary to design a motion controller to solve this problem. The design of the parallel mechanism motion control system adopts two layers of structure: Low order motion controller and high order servo driver. Motion controller is in charge of the motion control command decoding, the relative motion of each position control axis, and the contour control of acceleration and deceleration. The main function of motion controller is to reduce the path error of the whole system motion control. Servo driver is responsible for the position control of servo motor, the main function is to reduce the following error of the servo axis. The upper level motion controller completes the position loop control, and the lower servo driver completes the speed and torque loop control. This combines the advantages of decentralized

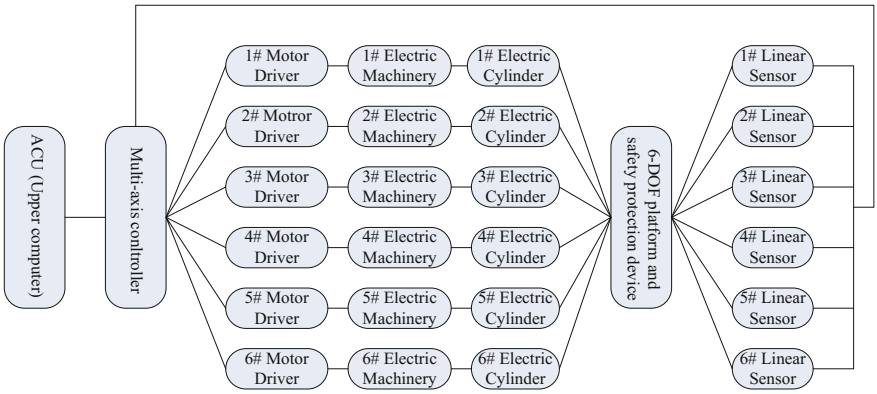


Fig. 3.4 Six-DOF parallel mechanism control system

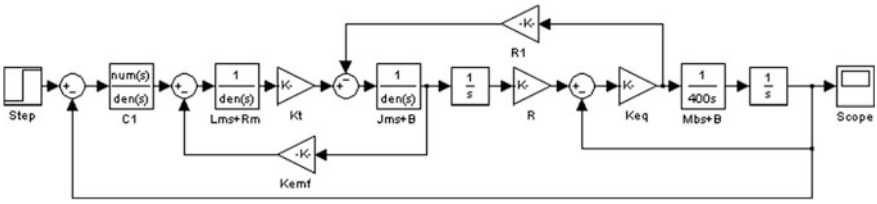


Fig. 3.5 Single leg control loop model of 6-DOF parallel mechanism

control and centralized control, which can meet the needs of each axis, and can make the overall coordination of multiple axes. 6-DOF parallel mechanism control system is shown in Fig. 3.4.

3.7 The Kinematics Analysis of the 6-DOF Deputy Reflector Control Mechanism

6-DOF parallel mechanisms are the MIMO systems which are to coordinate the movement of the six legs to move the auxiliary surface to the corresponding position and attitude. But for each leg, the coordinated movement of the MIMO can be equivalent to the motion of each leg in accordance with certain characteristics. One leg of the servo system is composed of servo drive, motor, encoder, electric cylinder and the position sensor, this design is relatively mature [18]. Figure 3.5 is a single leg control loop model, and Fig. 3.6 is single loop control characteristics of 6-DOF parallel mechanism.

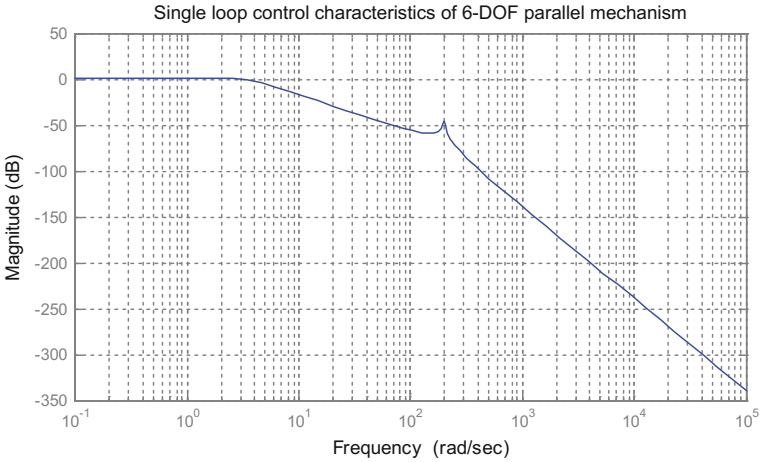


Fig. 3.6 Single loop control characteristics of 6-DOF parallel mechanism

Single loop transfer function of 6-DOF parallel mechanism is:

$$G(s) = \frac{68.4}{1.225 \times 10^{-5}s^5 + 2.743 \times 10^{-3}s^4 + 0.523s^3 + 104.819s^2 + 410.627s} \tag{3.13}$$

6-DOF deputy reflector control system is a complex coupling parallel mechanism. In order to accurately simulate the dynamic and kinematic characteristics of the platform, it is necessary to use ADAMS [19]. In order to generate the mechanism dynamics model, need to build models, configure quality, set constraints, and set the input and output variables. Without loss of generality, the mathematical simulation results of the control system of 6-DOF parallel mechanism shown in Fig. 3.6.

6-DOF deputy reflector control is the MIMO system, whose inputs are the Positions (x, y, z) and attitudes (α, β, γ). Six linear motion parameters are obtained by Inverse solution of positions and attitudes. The antenna can be positioned to the desired positions and attitudes, by controlling the positions of the linear motion of the six paths. The straight line motions of the six branches are mutually restricted, and the forces of each leg are coordinated, in the process of positions and attitudes of deputy reflector adjustment. Figure 3.7 is block diagram of 6-DOF parallel mechanism, and Fig. 3.8 is the relationship between the attitudes and positions of motion and motion of the legs (antenna overturned case).

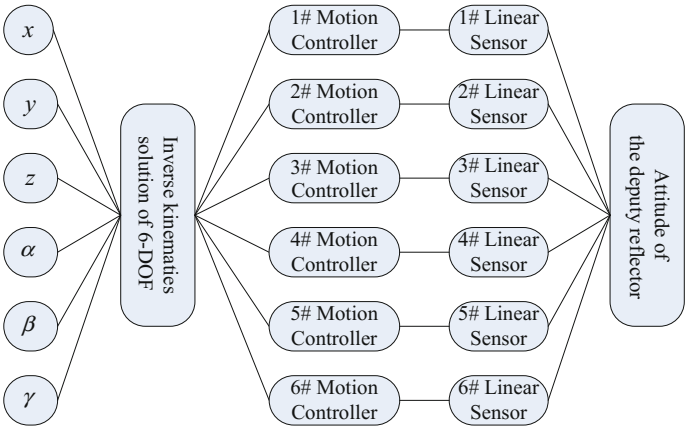


Fig. 3.7 The deputy reflector control of 6-DOF parallel mechanism

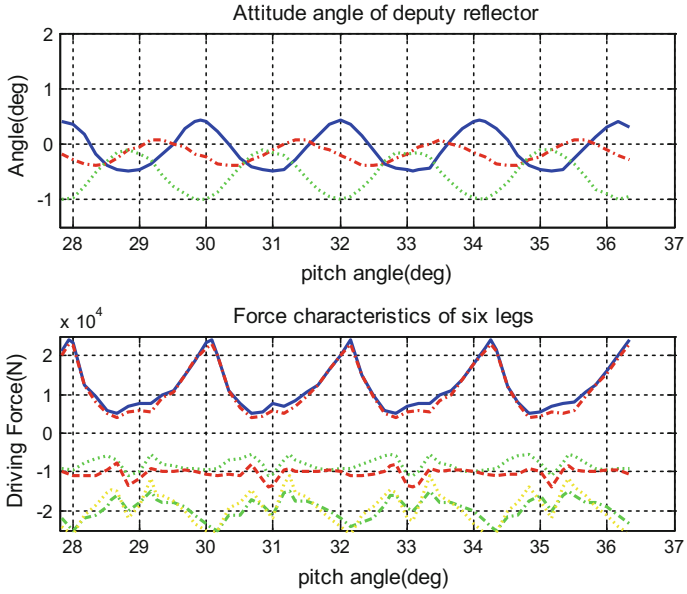


Fig. 3.8 The relationship between the attitude and position of motion and motion of the legs

3.8 The Relationship Between the Deputy Reflector Control and Antenna Control

The position of the deputy reflector surface should be accurately calibrated according to the deformation of the antenna, before the use of large deep space antennas. From the previous analysis, it can be known that the deformation changes with the change of the elevation angle, so it is needed to accurately calibrate the positions of the deputy reflector in different pitch angles. In the calibration process, calibration points the more the better. The positions and attitudes of the deputy reflector are calibrated at different elevation angles, which can be described by the curve fitting and the form function. The analytical expression of the position and attitude of the deputy reflector and the position of the pitch can be obtained by the curve fitting method [20]. Table function method is the antenna pitch angle as the independent variable, which can check the positions and attitudes of the deputy reflector in each pitch angle. According to the positions and attitudes of the adjacent pitch angle, the discrete quantity of the positions and attitudes of the deputy reflector can be obtained by means of linear interpolation, Lagrange interpolation or other polynomial interpolation method. According to the relative between the position and attitude and the pitch angle, the 6-DOF parallel mechanism can control the deputy reflector to the corresponding positions and attitudes to realize the real-time compensation of the antenna's electrical performance.

3.9 Conclusions

Due to the large diameter, high frequency, the decrease of electrical performance caused by the main and deputy reflector deformations of the large deep space detection and control antenna cannot be ignored. The 6-DOF parallel mechanism system control the positions and attitudes of deputy reflector to the calibration positions which are calibrated by the antenna elevation angles, compensation for performance loss due to antenna deformation. Although the 6-DOF parallel mechanism are composed of complex, the solutions are difficult, but compared with the serial mechanism, its bearing capacity is strong, flexible control, has irreplaceable advantages in the attitude and position control, so the 6-DOF parallel mechanism control has good application value. At present, the equipment of the control system of the 6-DOF parallel mechanism has been tested, and the effect is good.

Acknowledge In this paper, we get the guidance and help from Dr. Paul Wang of Xi'an Electronic and Science University!

References

1. Wang W, Duan B, Ma B (2008) Gravity deformation and best rigging angle for surface adjustment of large reflector antennas. Chin J Radio Sci 23(4):643–650
2. Jing H (2007) Deformation and stress analysis for paraboloid antenna reflector. Master's Degree Thesis of National University of Defense Technology, pp 14–16
3. Jianjun Z, Shuixiao A (2000) Determining the gravitational distortion of an antenna under test. Electro-Mech Eng 87(5):27–29
4. Zhang J, Shi H (2008) Time domain wind buffeting response of large-diameter antenna electric performances. Chin J Appl Mech 25(1):84–88
5. Li L, Qinzhuan Z (2014) Analysis of the trend of development of spacecraft TT&C and communication technologies and suggestions. J Spacecr TT&C Technol 33(6):464–468
6. Li Y, Manhong LU (2012) Trends of applications of space-based TT&C system. J Spacecr TT&C Technol 31(4):1–5
7. Zhang B, Ju L (2010) Research on development of aerospace TT&C technology before 2030. J Spacecr TT&C Technol 29(5):11–15
8. Fu L, Zhong W, Qiao H et al (2015) Research on the calculated methods of active control value for antenna panel deformations under gravity. Acta Astron Sin 56(4):378–390
9. Wang C, Duan B (2005) Electrical performance simulation platform and its application in structural design of large reflector antennas. Mach Des Res 21(3):87–89
10. Zhang Z (2009) The research and manufacture of stewart parallel mechanism with 6-DOF, Master's Degree Thesis of Tianjin University, pp 9–24
11. Craig JJ (2006) Introduction to robotics mechanics and control. Chia Machine Press, Beijing, pp 14–28
12. Ding J (2015) Research on methodologies for accuracy analysis and synthesis of six-DOF parallel mechanism. Doctoral Dissertation of Harbin Institute of Technology, pp 13–17
13. Peng X (2015) Dynamic modeling and simulation of the stewart platform as one of 6-DOF parallel mechanism, Master's Degree Thesis of Chongqing University, pp 21–52
14. Huang Z, Kong L (1997) Theory and control of parallel robot mechanism. Chia Machine Press, Beijing, pp 36–65
15. Yang H, Xue K (2016) Iterative method for forward kinematics of the Stewart platform with dynamic search origin. Appl Sci Technol 43(2):54–58
16. Niku SB (2004) Introduction to robotics: analysis, systems, applications. Publ House Electron Ind 34–41:95–98
17. Li L (2008) Research on forward kinematics and control method of 6-DOF parallel platform. Doctoral Dissertation of Harbin Enginee, pp 1–21
18. Liu S, Song J, Li W et al (2006) Motion control of a parallel 6-DOF Platform. Control Theory Appl 25(4):7–10
19. Ma Y (2010) Research on the dynamic analysis and optimal control of parallel six-DOF motion platform, Master's Degree Thesis of Nanjing University of Aeronautics and Astronautics, pp 18–23
20. Mulin H (1994) A best-fit method for deformed surface of antennas. Mod Rader 16(1):75–82

Chapter 4
The Study on the Adjustment Model of Sub-reflector and Engineering Realization Method

Yuhu Duan

4.1 Introduction

The surface accuracy of the antenna reflector is the main mechanical indicator of the antenna, which is closely related with the shortest operating wavelength. The real reflector is different with the theoretical reflector. The error includes the manufacturing error and the deformation caused by self-weight, wind load and temperature load and so on. Therefore, in the process of antenna design, not only the mechanical indicators and radio frequency indicator, but also the mecharonics should be considered [1]. With the increasing of antenna aperture and the operation frequency, the demand for the surface accuracy of the antenna reflector increases. The structural deformation of large reflector antenna is very large. And the gravity deformation is main reason for structural deformation. If the gravity deformation is reduced, the structural stiffness should be increased. So the antenna would become hulking. Especially for the millimeter wave antenna, greater demands are being placed on its surface accuracy. Therefore, many compensation methods for antenna deformation have been made. The method of using array feed to compensate the surface deformation of reflector is introduced in literature [2–8]. The phase of each unit in the feed array is changed and superimposed with the phase of the reflector to form the constant phase surface. However, due to the limit of weight, system complexity and requirements, the method of using feed array to compensate deformation cannot be used for some system. Yahya Rahmat-Samii and Robert A. Hoferer have made researches on the method of shaping the sub-reflector to compensate antenna deformation [9], and developed a fast and effective analysis method, which combines the Fourier Jacobi with physical optics and geometrical optics. The antenna efficiency of a 25 m antenna can be increased from 31 to 72% after the method is used to compensate the deformation in 43 GHz. According to

Y. Duan (✉)
The 39th Institute of China Electronic Technology Group Corporation, Xi'an 710065, China
e-mail: yuhuduan@aliyun.com

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_4

the reaches of William A. Imbriale .et of JPC, The antenna deformation can be compensated through replacing sub-reflector by reflecting array, using a feed or feed array to excite low-profile reflecting array, and adjusting the exciting phase of each unit in array [10, 11]. The active reflector technology and the compensation method of jointly using array feed and transformable plate [12] are widely used in the deformation compensation of large reflector antenna.

According to the analysis of the displacement of each node in the reflector, the translation includes two parts: one is the stiff movement of reflector and focus change of paraboloid; the other is the relative small deformation among the node of the reflector. According to computation results, the first kind of deformation, namely the stiff movement of reflector and focus change of paraboloid dominates, but the second one accounts for a small proportion. Therefore, some mathematical method can be used to compensate gravity deformation. In the engineering practice, a simple compensation method, namely the method of real time adjusting sub-reflector with the elevation angle of the antenna to compensate antenna gravity deformation, has been widely used. The method is based on the best-fit paraboloid technology [13, 14]. A series of new paraboloid can be made for matching with the transformed reflector. The new paraboloid has displacement and rotation compared with the original one. At the same time its focus has small change. The deviation of mean square root between the transformed reflector and one of the new paraboloids is the smallest, and this new one is the best-fit paraboloid. If the feed is placed in the new focus, the surface deviation is only related with the best-fit paraboloid and the error can be much reduced. The structure self-weight deformation is regular, and it is the function of antenna elevation angle. Therefore, the laws of the focus of best-fit paraboloid changing with the elevation angel can be computed. And then the feed and sub-reflector are moved according to the law to ensure the position of the feed or reflector is always in the new focus. So the gravity deformation can be compensated. The deformation of main reflector should be measured to conduct the best fit for main reflector and reduce the influence of gravity deformation on antenna performance through the adjustment of sub-reflector posture. The common measurement methods include the microwave holographic measurement [15, 16] and digital photogrammetry [17–19]. The interpolating calculation of measured data should be conducted to obtain the deformation value of main reflector for all elevation angles, which is the base of compensation method.

4.2 The Influence of the Posture Change of Sub-reflector on the Antenna

4.2.1 The Definition of the Coordinate

In the antenna system coordinate, when the antenna points to the horizontal direction (when the elevation angel is 0°), viewed from the front of the antenna, the

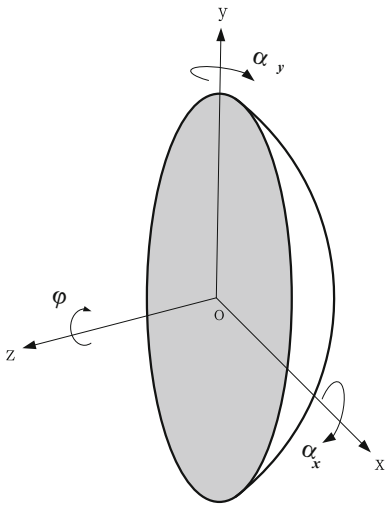


Fig. 4.1 Sub-reflector adjustment coordinates

forward direction is parallel with the antenna elevation axis, the Y axis points to the vertex, and the Z axis is vertical with the antenna aperture. The coordinate for describing sub-reflector is same with the antenna system coordinate. Only in the Z axis direction, some translation happens. The adjustment of sub-reflector uses six DOF platforms, namely the translation in X, Y, and Z direction and the rotation with each axis. The antenna system coordinate is shown in Fig. 4.1

4.2.2 The Influence of the Longitudinal Off-Set Focus on Antenna Gain

When the sub-reflector defocuses alone the antenna axis direction, the radials emitted by radiator is not a plane, but a quadric surface after being reflected by reflector. The reflected radios are no longer parallel with each other and defocusing happens, which can decrease the antenna gain and increase sidelobe. But the maximum wave bam will not excurse and deviate the antenna axis direction. The antenna gain loss and sidelobe increasing is not related with antenna aperture field distribution. Take a large deep space detecting antenna as an example. Commercial reflector antenna analysis software GRASP [20] is used to establish computation model. The antenna is a shaped Cassegrain dual-reflector antenna with uniform aperture field distribution. The diameter of main reflector, the diameter of sub-reflector, focus diameter ratio, and the illuminating angle between feed and sub-reflector are 35, 3.6 m, 0.325 and 8.9° respectively. The feed is a conical corrugated horn. The illuminating taper at the side of sub-reflector is −22.7 dB. The frequency 8.5 GHz is used to calculate the pattern of eleven positions with ± 1λ longitudinal off-set focus, as Fig. 4.2 shown. The results show that the off-set focus

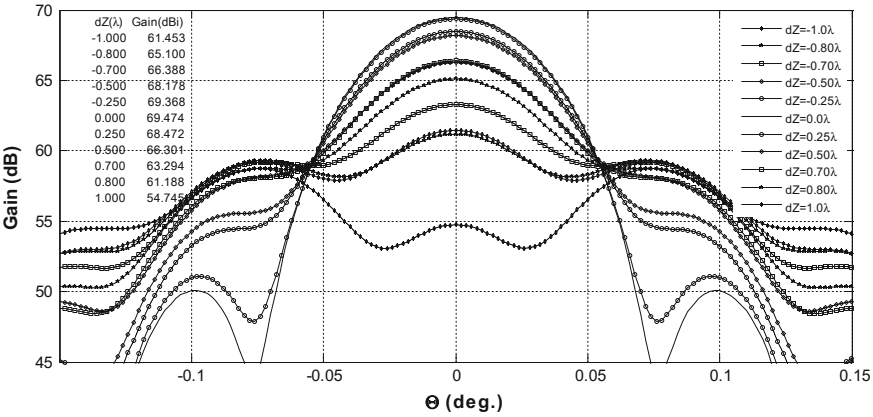


Fig. 4.2 The relationship of antenna gain and longitudinal off-set focus

gain loss in forward direction is larger than that in reverse direction. The gain loss of off-set focus ± 1λ is 14.7 and 8 dB respectively.

4.2.3 The Horizontal Translation of Sub-reflector

The horizontal translation of sub-reflector can reduce the antenna gain, increase the sidelobe level and make it asymmetry, and make the direction of the maximum wave beam deviate the axis direction. Suppose the horizontal translation of the sub-reflector is Δy_s. Its influence on the antenna wave beam pointing can be considered in two steps:

① Suppose the position of main reflector does not change. The position of feed and sub-reflector changes along the y axis. The sub-reflector and feed are in the same straight line of feed, which equivalent to the horizontal offset focus of prime focus reflector. The caused antenna beam deflection angle is:

$$\Delta\theta_1 = \frac{\Delta \mathbf{y}_s}{\mathbf{F_m}} \cdot \boldsymbol{B_F} \tag{4.2.1}$$

② The feed move bake to the original position, and restore the state of only having the horizontal translation of sub-reflector. The beam deflection angle caused by feed translation is:

$$\Delta\theta_2 = \frac{-\Delta y_s}{\mathbf{M} \cdot \mathbf{F_m}} \cdot \boldsymbol{B_F} \tag{4.2.2}$$

In reality, the beam deflection caused by sub-reflector translation is the sum of the above two defection angles, namely:

$$\Delta\theta_{sy} = \Delta\theta_1 + \Delta\theta_2 = \frac{\Delta y_s \cdot (M-1)}{\mathbf{M}\cdot\mathbf{F_m}} \cdot B_F \tag{4.2.3}$$

where F_m it the antenna focal distance, B_F is beam deviation factor, M is the magnification of Cassegrain antenna.

$$\mathbf{B_F} = \frac{1+0.36\left(\frac{D_m}{4F_m}\right)^2}{1+\left(\frac{D_m}{4F_m}\right)^2} \tag{4.2.4}$$

The radiation pattern of sub-reflector alone the Y off-set focus ± 1λ is calculated, as shown in Fig. 4.3a. According to the analysis, The relationship of gain and horizontal off-set focus is $\mathbf{G} = G_0 \cdot e^{a_0(y-y_0)^2}$. The beam peak position has linear relationship with focal offset Δy_s. The slope is $\frac{(M-1)\cdot B_F}{M\cdot F_m}$. Figure 4.3b gives the relation curve of beam peak position and focal offset Δy_s, where, "o" is the simulation result of GRASP software. The full line represents the computation results of formula (4.2.3). They are in good agreement with each other.

4.2.4 The Sub-reflector Rotates About the Vertex

Suppose the prime focus is the axis. The rotation angle of sub-reflector about x axis is α_x. The feed horizontal translation caused by the rotation is $\Delta x_s = (M+1)\cdot Fs \cdot \tan(\alpha_x) \approx (M+1)\cdot Fs \cdot \alpha_x$. The method is same with the analysis method of reflector translation. Substitute Δx_s into (4.2.1) and replace Δy_s. The beam deviation angle caused by sub-reflector rotating α_x about x axis is:

$$\Delta\theta_{\alpha x} = \frac{\alpha_x \cdot (M+1)\cdot F_s}{\mathbf{M}\cdot\mathbf{F}_m} \cdot B_F \tag{4.2.5}$$

where F_s is the focal distance of sub-reflector. 0.2° is the step distance. The radiation pattern after sub-reflector rotating about x axis with ± 1 is calculated by GRASP software, as shown in Fig. 4.4a. In Fig. 4.4b, the formula (4.2.5) is compared with the beam peak position simulated by GRASP. They are in good agreement.

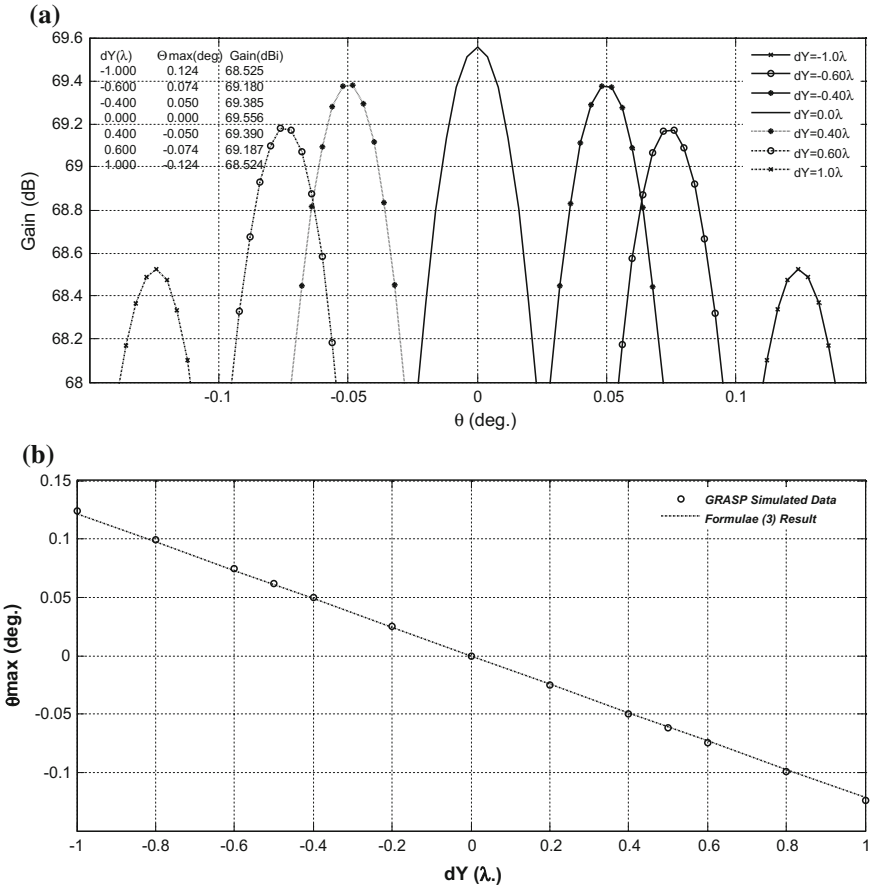


Fig. 4.3 **a** The relationship of antenna beam peak position and sub-reflector horizontal translation. **b** The relationship of Grasp simulation results and beam peak position and sub-reflector horizontal translation calculated according to formula (4.2.3)

4.3 Sub-reflector Posture Adjustment Model

The change of Sub-reflector posture can result in decrease of antenna gain, increase of sidelobe level and pointing deviation from axial direction. Gravity is the main factor affecting antenna sub-reflector posture. Antenna structural distortions induced by gravity load are uncorrelated can be achieved by using algebraic sum method of independent error term. Any independent error term is generated by corresponding distortion. 3 main parts related with antenna pointing are main reflector, sub-reflector and feed, can be denoted with 5 independent error terms respectively: main reflector rotation, main reflector fixed-point translation, sub-reflector rotation, sub-reflector translation and feed displacement. Stimulation and examination show that decrease of gain induced by lateral displacement and

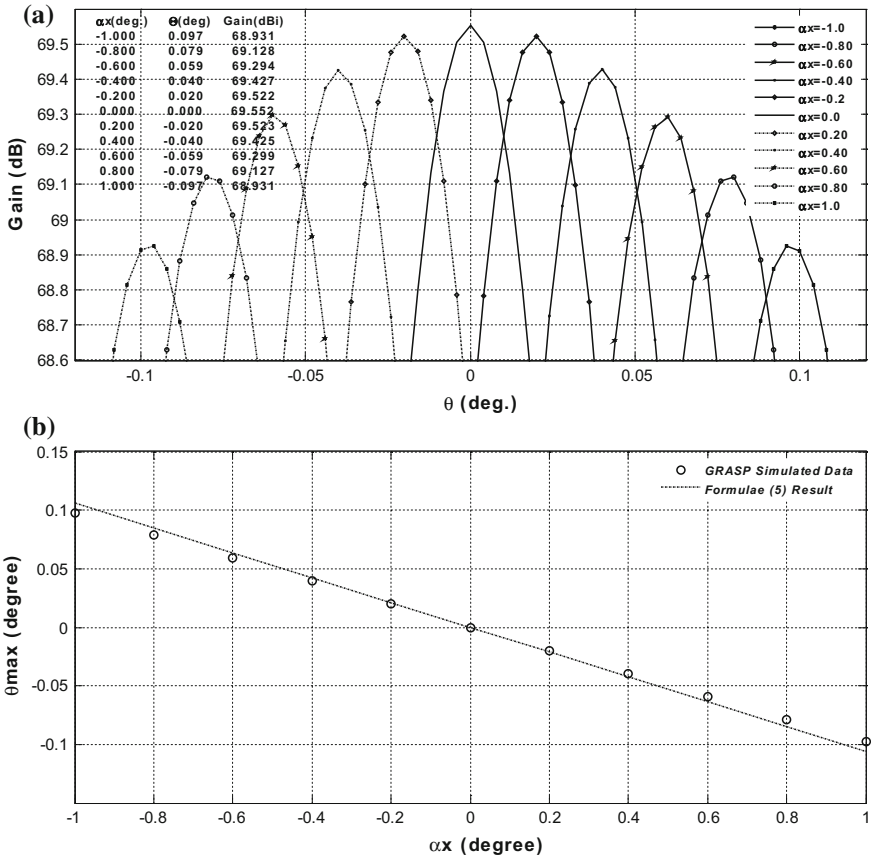


Fig. 4.4 **a** The relationship of rotation angle of sub-reflector about x axis and the beam peak value and pointing deviation. **b** The relationship of Grasp simulation results and beam peak position and sub-reflector rotation angle calculated according to formula (4.2.5)

rotation of sub-reflector is much smaller than that of resulted from Z coordinate displacement. For example, when sub-reflector off-focus in lateral and longitudinal direction is 1λ, gain loss is 1 and 14.6 dB respectively. Therefore, longitudinal adjustment of sub-reflector is more important. Moreover, due to bilateral symmetry of gravity distortion, displacement in X coordinate is very small. Hence, the following formula can be used to denote sub-reflector posture variable induced by gravity distortion for Cassegrain antenna system.

$$\begin{cases} \Delta \mathbf{Y} = \boldsymbol{A}_y + \boldsymbol{B}_y \cdot \mathbf{cos}(\boldsymbol{E}) + \boldsymbol{C}_y \cdot \mathbf{sin}(\boldsymbol{E}) \\ \Delta \boldsymbol{Z} = \boldsymbol{A}_z + \boldsymbol{B}_z \cdot \mathbf{cos}(\boldsymbol{E}) + \boldsymbol{C}_z \cdot \mathbf{sin}(\boldsymbol{E}) \\ \Delta \boldsymbol{\alpha}_{\boldsymbol{x}} = \boldsymbol{A}_{\boldsymbol{\alpha}} + \boldsymbol{B}_{\boldsymbol{\alpha}} \cdot \mathbf{cos}(\boldsymbol{E}) + \boldsymbol{C}_{\boldsymbol{\alpha}} \cdot \mathbf{sin}(\boldsymbol{E}) \end{cases} \tag{4.3.1}$$

where EL is antenna elevation angle. Factors A, B, C can be obtained with measured or stimulated sub-reflector posture data varying with antenna elevation angle by using least square method.

4.4 Project Realization of Sub-reflector Adjustment

4.4.1 Sub-reflector Adjustment Dimension Determination

The analysis described in Sect. 4.2 shows that longitudinal displacement of sub-reflector has more severe effect on antenna performance than lateral displacement and rotation. In practical engineering application, since gravity distortion is symmetrical about YOZ plane, to simplify sub-reflector adjustment system, only perform 3-D adjustment (that is, translation in Y and Z coordinate and rotation about X axis), instead of translation in X coordinate and rotation about Y and Z axis.

Translation in X, Y and Z coordinate and rotation about X and Y axis have been performed to facilitate initial adjustment of sub-reflector for 35 m antenna for deep space observation project. After sub-reflector is adjusted to the optimum position at preadjustment angle, lock translation in X axis and rotation about Y axis, hence simplifying the routine operation. According to theoretically analyzed sub-reflector adjustment value, adjustment range, positioning accuracy and repeatability are listed in Table 4.1.

4.4.2 Project Realization of Sub-reflector Adjustment

In the practical engineering application of antenna for large deep space exploration TT&C station and radio astronomical observation, mathematical model or table looking-up interpolation is often used to perform real-time adjustment for sub-reflector to compensate effect of gravity distortion. Before RF system is installed, measure and adjust main reflector accuracy at preadjustment angle with digital photography to meet accuracy requirement. Afterwards, measure main reflector distortion and sub-reflector posture at other elevation angles, perform best fit for antenna distortion value obtained at other angles to get the optimum position of sub-reflector. At the same time, sub-reflector posture variable varying with

Table 4.1 Sub-reflector adjustment range and repeatability requirement

	X coordinate	Y coordinate	Z coordinate	α_x
Adjustment range	±80 (mm)	±80 (mm)	±100 (mm)	±2.5°
Adjustment speed	2 mm/s	2 mm/s	2 mm/s	0–4°/s
Repeatability	0.2 (mm)	0.2 (mm)	0.2 (mm)	0.002°

antenna elevation angle is also obtained via measurement. Hence, sum of sub-reflector position obtained with best fit method and measured sub-reflector posture variable is final sub-reflector adjustment value. For actually measured elevation angle is limited, substitute relation between final adjustment value and elevation angle into (4.3.1), getting factors A, B and C with least square method.

In a 35 m antenna for deep space observation project, we performed pread-justment for main reflector at 40° of elevation angle, then respectively measured main reflector distortion and sub-reflector posture coordinate at 0°, 15°, 30°, 40°, 60°, and 75° of elevation angles with digital photography. Perform best fit for main reflector distortion value measured at above angles to get sub-reflector adjustment value. Superpose it with measured sub-reflector displacement at corresponding angles induced by sub-reflector support leg distortion, getting final adjustment value of sub-reflector at corresponding angles. Substitute it into mathematical model, use least square best fit to obtain factors in formula (4.3.1):

$$\begin{cases} \boldsymbol{A_y} = -28.78 & \boldsymbol{A_z} = 1.150 & \boldsymbol{A_\alpha} = 0.340 \\ \boldsymbol{B_y} = 34.120 & \boldsymbol{B_z} = -2.95 & \boldsymbol{B_\alpha} = -0.33 \\ \mathbf{C}_y = 4.8900 & \mathbf{C}_z = 1.400 & \mathbf{C}_\alpha = -0.13 \end{cases}$$

Substitute above factors into (6), calculated sub-reflector adjustment value and measured results can be seen in Fig. 4.5a, b and c. Where o represents measured value, dotted line represents calculated value with mathematical model.

4.4.3 Practical Examples

The above section gives factors A, B and C in sub-reflector adjustment mathematical model with 40° of elevation angle as reference. However, it is difficult to find source at 40° of elevation angle to calibrate null of sub-reflector. The following method is often used to recalculate factors in sub-reflector adjustment model at any elevation angles. Take 20° of elevation angle as example, the specific steps are as follows:

Substitute 20° of elevation angle into mathematical model in Sect. 4.4.2, calculate adjustment value $\Delta\mathbf{Y}_{20}$, $\Delta\mathbf{Z}_{20}$ and $\Delta\boldsymbol{\alpha}_{x20}$, which are obtained by subtracting adjustment value at 20° of angle from adjustment value at each angle, obtaining sub-reflector adjustment value at corresponding angles with 20° of angle as reference:

$$\begin{cases} \Delta\mathbf{Y} = (\boldsymbol{A_y} - \Delta\mathbf{Y}_{20}) + \boldsymbol{B_y} \cdot \mathbf{cos}(\boldsymbol{E}) + \boldsymbol{C_y} \cdot \mathbf{sin}(\boldsymbol{E}) \\ \Delta\boldsymbol{Z} = (\boldsymbol{A_z} - \Delta\boldsymbol{Z}_{20}) + \boldsymbol{B_z} \cdot \mathbf{cos}(\boldsymbol{E}) + \boldsymbol{C_z} \cdot \mathbf{sin}(\boldsymbol{E}) \\ \Delta\boldsymbol{\alpha_x} = (\boldsymbol{A_\alpha} - \Delta\boldsymbol{\alpha}_{x20}) + \boldsymbol{B_\alpha} \cdot \mathbf{cos}(\boldsymbol{E}) + \boldsymbol{C_\alpha} \cdot \mathbf{sin}(\boldsymbol{E}) \end{cases} \tag{4.1}$$

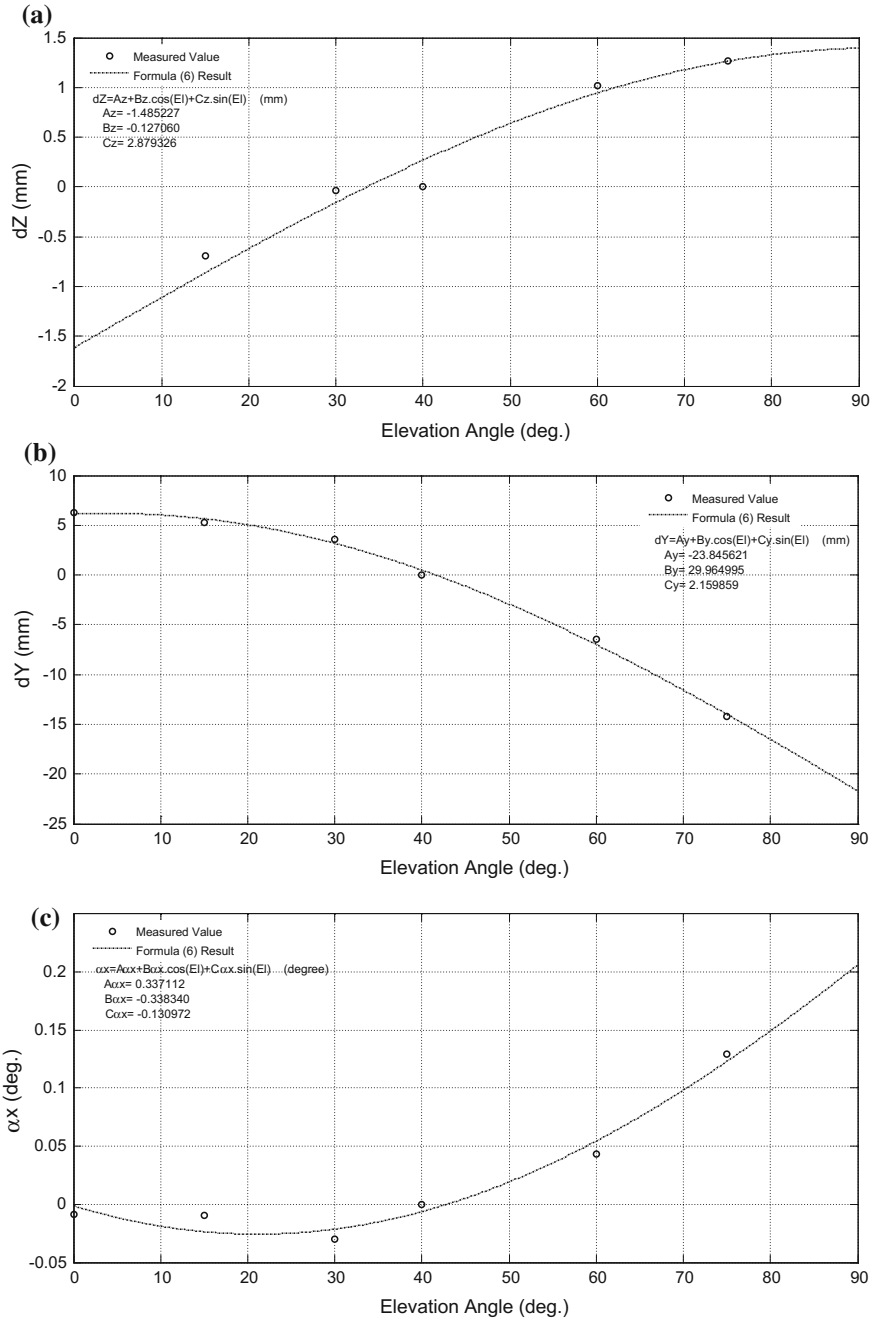


Fig. 4.5 **a** Formula (4.3.1) Calculated and measured sub-reflector longitudinal (Z coordinate) adjustment value versus elevation angles. **b** Formula (4.3.1) calculated and measured sub-reflector lateral (Y coordinate) adjustment value versus elevation angles. **c** Formula (4.3.1) Calculated and measured sub-reflector adjustment value by rotation about X axis versus elevation angles

After reference elevation angle is determined, adjustment value at this angle is known and unchanged. Factor A subtracts adjustment value at 20° of elevation angle, keep other factors unchanged. The newly obtained factors are as follows:

$$\begin{cases} \boldsymbol{A}_y = -23.84 - \Delta \boldsymbol{Y}_{20} = -28.8919 \\ \boldsymbol{B}_y = 29.960 \\ \mathbf{C}_y = 2.1600 \end{cases}$$

$$\begin{cases} \boldsymbol{A}_z = -1.480 - \Delta \boldsymbol{Z}_{20} = -0.8656 \\ \boldsymbol{B}_z = -0.127 \\ \mathbf{C}_z = 2.8800 \end{cases}$$

$$\begin{cases} \boldsymbol{A}_{\alpha x} = 0.337 - \Delta \boldsymbol{\alpha}_{x20} = 0.3624 \\ \boldsymbol{B}_{\alpha x} = -0.338 \\ \mathbf{C}_{\alpha x} = -0.131 \end{cases}$$

Subtract above factors into (4.3.1) to get sub-reflector adjustment model at real time with 20° of angle as reference. Stimulate sub-reflector adjustment value at 20° of angle and distorted main reflector data using GRASP software, through sub-reflector position translation and rotation, the gains of X band and Ka band antennas increase by 0.52 and 1.95 dB respectively. In real project, perform iterative adjustment for sub-reflector with source at 20° of angle to find out the position where the optimum gain and sidelobe level can be obtained. It is also the best position of sub-reflector at this angle. Set it to null position to verify at other angles.

4.5 Conclusions

The effect of gravity load on gain and beam peak pointing of antenna for large deep space TT&C station and radio astronomical exploration is presented here. Actually measured distorted data of antenna main reflector and sub-reflector support structure for 35 m large deep space exploration station are best fit, obtaining the initial data for compensating gravity distortion by adjusting sub-reflector at real time. Sub-reflector adjustment mathematical model is also given. Least square method is used to obtain factor in sub-reflector adjustment mathematical model. Practical application examples are also presented.

References

1. Duan BY (2011) Theory, method and application of electromechanical coupling of electronic equipment. Science Press, Beijing, pp 25–26 (in Chinese)
2. Wang CS, Li JJ (2012) Review of compensation methods of lange reflector antenna with surface deformation. In: Symposium on electromechanical, microwave communication equipment structure technology of China electronics society. Harbin, China, pp 1–7

3. Blank SJ, Imbriale WA (1988) Array feed synthesis for correction of reflector distortion and vernier beamsteering. IEEE Trans Antennas Propag 36(11):1351–1358
4. Cherrette AR, Acosta RJ, Lam PT, Lee S-W (1989) Compensation of reflector antenna surface distortion using an array feed. IEEE Trans Antennas Propag 37(8):966–978
5. Smith WT, Stutzman WL (1992) A pattern synthesis technique for array feeds to improve radiation performance of large distorted reflector antennas. IEEE Trans Antennas Propag 40 (1):57–62
6. Imbriale William A (2002) Large antennas of the deep space network. Jet Propulsion Laboratory, Pasadena, pp 140–149
7. Vilnrotter V, Fort, D. Demonstration and evaluation of the Ka-band array feed compensation system on the 70-meter antenna at DSS 14. The Telecommunications and Mission Operations Progress Report 42-139, pp 1–17. http://tmo.jpl.nasa.gov/tmo/progressreport/42-139/139J.pdf
8. Imbriale WA, Hoppe DJ. Computational methods and theoretical results for the Ka-band array feed compensation system/deformable flat plate experiment at DSS 14 Telecommunications and Mission Operations Progress Report 42-140, pp 1–23. http://tmo.jpl.nasa.gov/tmo/progressreport/42-140/140I.pdf
9. Hoferer RA, Rahmat-Samii Y (2002) Subreflector shaping for antenna distortion compensation: an efficient Fourier-Jacobi expansion with GO/PO analysis. IEEE Trans Antennas Propag 50(12):1676–1687
10. Shenheng Xu, Rahmat-Samii Yahya (2009) Subreflectarrays for reflector surface distortion compensation. IEEE Trans Antennas Propag 57(2):364–372
11. Xu S, Rajagopalan H (2007) A novel reflector surface distortion compensating technique using a sub-reflectarray. In: Proceedings of IEEE antennas and propagation society international symposium, pp 5315–5318
12. Richter P, Franco M, Rochblatt D. Data analysis and results of the Ka-band array feed compensation system deformable flat plate experiment at DSS 14. The Telecommunications and Mission Operations Progress Report 42-139, pp. 1–29. http://tmo.jpl.nasa.gov/tmo/progressreport/42-139/139H.pdf
13. JIANG Zheng and LEI Pei-tian (1991) A best-fit method for the shaped Cassegrain antenna. Syst Eng Electron 13(4):36–42
14. Guo-jun LENG, Wei WANG, Bao-yan DUAN (2011) Subreflector real-time compensation for main reflector deformation of shaped Cassegrain antenna. Syst Eng Electron 33(5):996–1000
15. Kiuchi Hitoshi (2013) A holography receiver design for the ALMA sub millimeter antenna surface measurement. IEEE Trans Antennas Propag 62(10):2763–2772
16. Duan YH (2014) Microwave holographic metrology of the surface accuracy of reflector antenna—simulation method. In: Proceedings of the 27th conference of spacecraft TT&C technology in China, Guangzhou China, pp 103–111
17. Bao-feng WANG, Guang-yun LI (2007) Application of high accuracy digital photogrammetric technology in a 50 meter large antenna. Eng Surv Mapp 16(1):42–46
18. Li H, Yang DH (2010) Research on the pose measurement of a 6-DOF platform using a single camera. Opt Tech 36(3):344–349
19. Fan QH, Fan SH (2010) Application of digital industrial photogrammetric technology to measure the surface accuracy of 13.7 m millimeter-wave radio telescope antenna. ACTA Astron Sin 51(2):210–216
20. GRASP9, Technical Description Editor: Knud Pontoppidan, TICRA, Denmark September 2005

Chapter 5
Experiment and Performance Analysis of Iterative FX Correlation Combining Algorithm for Arraying in Deep Space Network

Youyong Liu, Hujun Geng, Suli Guo and Weijun Yang

5.1 Introduction

The multi-antenna received arraying technique is an effective alternative as the future large-aperture antenna in deep space explorations [1, 2]. DSN studies deeply on the combining method, the combining algorithm and so on. It gives several feasible arraying methods such as FSC, BC, SSC, etc. [2, 3], and then successively puts forward the SIMPLE, SUMPLE and wideband frequency combination as the combining algorithms [4, 5]. Many domestic units conduct thorough research to the arraying correlation technique [6, 7], and have carried out relevant experiments in the CE mission and got good results, whereas their research mainly focus on the full spectrum combining (FSC) method and SUMPLE combining algorithm proposed by DSN. At a Martian distance, however, due to the far transmission, the signal degrade seriously and the received signal of single antenna is weaker, thus the performance of the traditional combining methods (such as FSC etc.) will worsen sharply.

The key of the arraying signal combining is the alignment of the time, frequency and phase difference of the received signal for each antenna. Serving as an open-loop frequency-domain implementation architecture in the VLBI interferometry [8], FX correlator has a high precision in measuring the multi-antenna time, frequency and phase differences and correlation coefficient, which is adapt to various signal types. Based on the "one-to-more" correlating processing idea of FX correlator and SUMPLE algorithm, therefore, an iterative FX correlation combining algorithm in frequency domain is proposed for the combination of the weak

Y. Liu (✉) · H. Geng · S. Guo · W. Yang
The 54th Research Institute of CETC, Shijiazhuang 050081, China
e-mail: yongliu_80@163.com

Y. Liu
CETC Key Laboratory of Space Information Application Technology,
Shijiazhuang 050081, China

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_5

received signals. The performance analysis and the comparison with the FSC arraying method are carried out. Finally, the combining results of the algorithm for the data-transmission signals of a 4-antenna arraying in CE-2 mission are present.

5.2 The Principle of FX Correlation Algorithm

FX correlator is the master method to measure the multi-antenna delay difference in the VLBI interferometry, which aims to get the baseline-projected delay difference coming from the same transmitted wave. The basic processing architecture is shown in Fig. 5.1.

Without loss of generality, based on the signals of two stations located at one baseline, the measuring principle of the multi-antenna delay and phase differences is illuminated in detail. Assuming the modulation of detectors is BPSK, the bandwidth of the IF filter is B_n, the sample rate is f_s, and the reference signal is the received signal of station 1, the received IF signals of two channels can be respectively given by

$$\begin{cases} r_1(t_k) = \sqrt{P_1} \times d(t_k) \times \exp(j \times (2\pi f_{IF} t_k + \varphi_0)) + n_1(t_k) \\ r_2(t_k) = \sqrt{P_2} \times d(t_k - \tau_{g12}(t_k)) \times \exp(j \times (2\pi f_{IF} t_k - 2\pi f_{RF} \tau_{g12}(t_k) + \varphi_{LO12} + \varphi_0)) + n_2(t_k) \end{cases} \tag{5.1}$$

where P_1 and P_2 are the received signal powers, $d(t_k)$ is the modulation data, f_{RF} and f_{IF} are the RF and IF frequencies, $\tau_{g12}(t_k)$ is the multi-antenna real delay difference, φ_{LO12} is the local oscillator initial phase difference between station 1 and station 2, φ_0 is the initial phase of the detector's downlink signal, and $n_i(t_k)$ is the band limited white noise with one-sided spectrum density level N_0 and noise variance $N_0 B_n$.

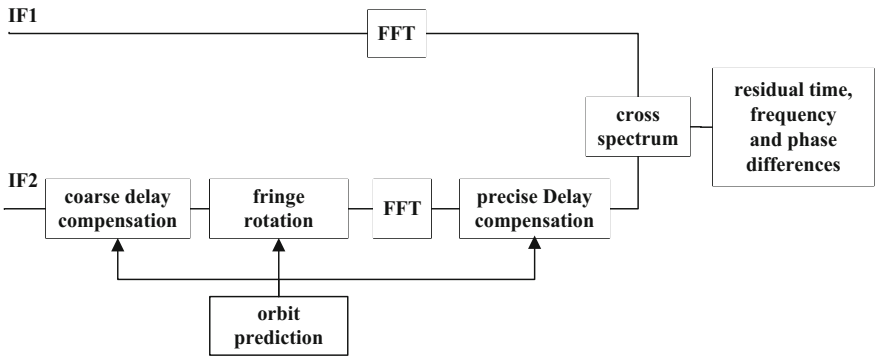


Fig. 5.1 Implementing architecture of FX correlator in VLBI

The orbit prediction based on baseline can be utilized for station 2 to compensate the delay and rotate the fringe. The multi-antenna delay difference of orbit prediction, τ_{gc}, is assumed as

$$\begin{aligned}\tau_{gc}(t_k) &= \tau_{gc}(t_0) + \tau'_{gc}(t_k) * (t_k - t_0) \\ &= \tau_{gc-I}(t_0) + \tau_{gc-F}(t_0) + \tau'_{gc}(t_0) * (t_k - t_0)\end{aligned} \tag{5.2}$$

where $\tau_{gc-I}(t_0)$ and $\tau_{gc-F}(t_0)$ are the corresponding integer bit delay and fraction bit delay of the time t_0 and $\tau'_{gc}(t_0)$ is the corresponding delay rate at the time t_0.

The residual between thee predicted value and real one is assumed as $\Delta\tau = \tau_g - \tau_{gc}$. After compensating the integer bit delay of the IF signal of station 2, we obtain

$$\begin{aligned}r_{2I}(t_k) &= r_2(t_k + \tau_{gc-I}(t_0)) \\ &= \sqrt{P_2} \times d(t_k - (\tau_{g12} - \tau_{gc-I}(t_0))) \times \exp(j \times (2\pi f_{IF} t_k \\ &\quad + 2\pi f_{IF}\tau_{gc-I}(t_0) - 2\pi f_{RF}\tau_{g12} + \varphi_{LO12} + \varphi_0)) + n_2(t_k + \tau_{gc-I}(t_0))\end{aligned} \tag{5.3}$$

The local constructs a frequency signal and rotates the fringe of the received signal to reduce the multi-antenna signal dynamic, which is given by

$$r_{FS}(t_k) = \exp(j[2\pi f_{IF}\tau_{gc-I}(t_0) - 2\pi f_{RF}\tau_{gc}(t_k)]) \tag{5.4}$$

After rotating the fringe, the signal is given by

$$\begin{aligned}r_{2FS}(t_k) &= r_{2I}(t_k) \times r^*_{FS}(t_k) \\ &= \sqrt{P_2} \times d(t_k - (\Delta\tau + \tau_{gc-F}(t_0) + \tau'_{gc}(t_0) * (t_k - t_0))) \\ &\quad \times \exp(2\pi f_{IF} t_k + 2\pi f_{RF}\Delta\tau + \varphi_{LO12} + \varphi_0)\end{aligned} \tag{5.5}$$

Executing the FFT to the reference signal of station 1 and the fringe-rotated signal of station 2, we obtain

$$\begin{cases} R_1(f_k) = \sqrt{P_1} \times D(f_k + f_{IF}) \times \exp(j\varphi_0) + n_1(f_k) \\ R_{2FS}(f_k) = \sqrt{P_2} \times D(f_k + f_{IF}) \times \exp(j(-2\pi f_k(\Delta\tau + \tau_{gc-F}(t_0) \\ \qquad + \tau'_{gc}(t_0) * (t_k - t_0))) + (2\pi f_{RF}\Delta\tau + \varphi_{LO12} + \varphi_0))) + n'_2(f) \end{cases} \tag{5.6}$$

Constructing a phase vector and compensating the fraction bit delay of the frequency signal of station 2, we obtain

$$\begin{aligned}R_{2Fc}(f_k) &= \sqrt{P_2} \times D(f_k + f_{IF}) \times \exp(j(-2\pi f_k \Delta\tau) \\ &\quad \times \exp(j(2\pi f_{RF}\Delta\tau + \varphi_{LO12} + \varphi_0)) + n'_2(f_k)\end{aligned} \tag{5.7}$$

Calculating the multi-antenna cross-spectrum, we obtain

$$\begin{aligned} R_{12}(f_k) &= R_1(f_k) \times R^*_{2Fc}(f_k) \\ &= \sqrt{P_1 P_2} \times |D(f_k + f_{IF})|^2 \times \exp(j(2\pi f_k \Delta\tau - \Delta\varphi_{12})) + n_{12}(f_k) \end{aligned} \tag{5.8}$$

where $\Delta\varphi_{12} = 2\pi f_{RF}\Delta\tau + \varphi_{LO12} + \varphi_0$ and $n_{12}(f_k)$ is the equivalent noise. The cross-spectrum phase is given by

$$\Phi_{12}(f_k) = 2\pi f_k \Delta\tau - \Delta\varphi_{12} + n'(f_k) \tag{5.9}$$

So the residual delay and phase differences are given by

$$\begin{aligned} \Delta\tau &= \frac{d\Phi_{12}(f_k)}{df_k} \\ \Delta\Phi_{12} &= \frac{imag(\sum R_{12}(f_k))}{real(\sum R_{12}(f_k))} \approx \Phi_{12}(f_k)\big|_{f_k=0} \end{aligned} \tag{5.10}$$

Due to the fringe rotation based on orbit prediction could largely eliminate the multi-antenna dynamic, the multi-antenna residual frequency difference is obtained by fulfilling least square fitting of the residual phase difference from the observation time, i.e.,

$$\Delta f = \frac{d\Delta\varphi_{12}(t)}{dt} \tag{5.11}$$

The spectrum of the equivalent noise is given by

$$\begin{aligned} n_{12}(f_k) = &\sqrt{P_1} \times D(f_k + f_{IF}) \times \exp(j\varphi_0) \times n_2'^*(f_k) \\ &+ \sqrt{P_2} \times D(f_k + f_{IF}) \times \exp(j(-2\pi f_k \Delta\tau) \\ &\times \exp(j(\varphi_{12} + \varphi_0)) \times n_1'^*(f_k) + n_1(f_k) \times n_2'^*(f_k) \end{aligned} \tag{5.12}$$

The effective in-band noise power in the time range of T_{NFFT} is given by

$$\sigma_{12}^2 = (P_1 N_{02} B_s + P_2 N_{01} B_s + N_{01} B_s \times N_{02} B_s) \times T_{NFFT} \tag{5.13}$$

where B_s is the effective bandwidth of correlation processing, and $T_{NFFT} = N_{FFT}/fs$. Assuming that the integration time is T seconds, statistical average is replaced by the time averaging operation over T seconds, which could reduce the variance of the equivalent noise by a factor of N, $N = T/T_{NFFT} = (T * f_s)/N_{FFT}$. The in-band SNR at the output of the accumulator is given by

$$\begin{aligned} SNR_{12} &= \frac{|E(R_{12})|^2}{\mathrm{var}(n_{12})} = \frac{NP_1P_2}{P_1N_{02}B_s + P_2N_{01}B_s + N_{01}B_s \times N_{02}B_s} \\ &= \frac{N}{\frac{1}{SNR_1} + \frac{1}{SNR_2} + \frac{1}{SNR_1} \times \frac{1}{SNR_2}} \end{aligned} \tag{5.14}$$

To the deep space weak signal, the bandwidth of the received signal is in the range from a few kHz to MHz. The in-band SNR is commonly negative and easily satisfies $\frac{1}{SNR_1} + \frac{1}{SNR_2} \ll \frac{1}{SNR_1} \times \frac{1}{SNR_2}$. The in-band SNR of the output signal, thus, is given by

$$SNR_{12} \approx N \times SNR_1 \times SNR_2 \tag{5.15}$$

When the in-band combining SNR_{12} satisfies $SNR_{12} > 6\,\text{dB}$, the phase estimation satisfies Gaussian distribution with the precision $\sigma^2_{\varphi_{12}} = \frac{1}{SNR_{12}}$.

5.3 The Iterative FX Correlation Combining Algorithm

Based on the SIMPLE algorithm, SUMPLE algorithm is a method that is used to improve the precision in estimating the phase weighted value of the weak arraying signals by DSN. On the basis of the alignment of the multi-antenna delay difference, the algorithm regards the weighted sum of all other antenna signals except itself as the reference antenna, and then gets the phase difference between each antenna and the reference antenna by the correlating processing of two antennas in time domain. Through the iterative processing of the weighted value in a period of time, the weighted value for each antenna can achieve stable convergence. The essence of the algorithm, therefore, is to improve the SNR of the reference antenna by utilizing the correlation of each antenna signal and summing the weighted values, which could get a better performance than SIMPLE algorithm in lower SNR circumstance.

The iterative FX correlation combining algorithm is a novel frequency-domain combining method based on the "one-to-more" weighted iterative method of FX correlator and SUMPLE algorithm. The processing scheme is shown in Fig. 5.2.

In the algorithm, the correlating processing is done in frequency domain. By the FX correlation of each antenna with the reference antenna composed of the weighted sum of all other antennas, the multi-antenna parameters (such as the residual delay, phase and frequency differences) are estimated on the basis of the signal spectrum characteristics. The algorithm is similar to an implementing architecture of the combination of FSC and match filters in frequency domain, which could get a better combining performance than FSC.

In the iterative FX correlation combining algorithm, the cross-spectrum is given by

$$\begin{aligned} R_{iref}(f_k) &= R_{ref}(f_k) \times R^*_{iFc}(f_k) \\ &= \sqrt{P_{ref}P_i} \times |D(f_k + f_{IF})|^2 \times \exp(j(2\pi f_k \Delta\tau_{iref} - \Delta\varphi_{iref})) + n_{iref}(f_k) \end{aligned} \tag{5.16}$$

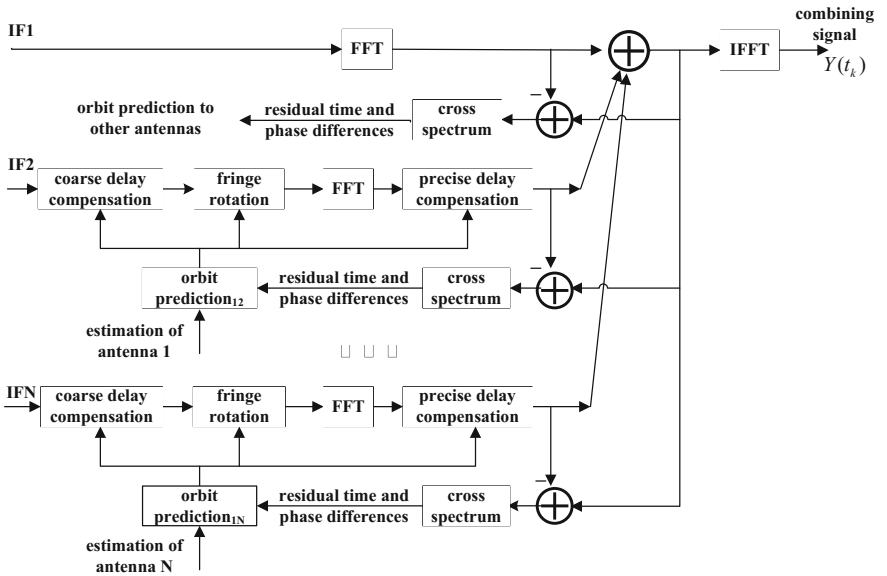


Fig. 5.2 Processing scheme of the iterative FX correlation combining algorithm

Similarly, the in-band SNR at the output of the corresponding accumulator for each antenna is given by

$$SNR_{iref} \approx N \times SNR_{ref} \times SNR_i = N \times (SNR_{SUM} - SNR_i) \times SNR_i \tag{5.17}$$

where SNR_{SUM} is the in-band SNR of the combining signal, and SNR_i is the in-band SNR of the ith antenna signal. The precision in estimating the phase difference satisfies $\sigma^2_{\varphi_{iref}} = \frac{1}{SNR_{iref}}$.

The combining signal of all M antennas is given by

$$\begin{aligned} Y(t_k) &= \sum_{m=1}^{M} \beta_m Y_m(t_k) \\ &= \sum_{m=1}^{M} \beta_m [\sqrt{P_m} e^{j(2\pi f_{IF} t_k + \Delta\varphi_{mref}(t_k))} + n_i(t_k) e^{j(2\pi f_{IF} t_k + \hat{\varphi}_{mref}(t_k))}] \end{aligned} \tag{5.18}$$

Assuming that the weighted value β_i makes the performance of the output SNR best, thus, the weighted value should satisfies $\beta_i = \sqrt{\frac{P_i}{P_1}} \times \frac{N_{01}}{N_{0i}}$. The noise variance of the combining signal is $\sigma_Y^2 = B_s \sum_{m=1}^{M} \beta_i^2 N_{0m} = B_s N_{01} \sum_{m=1}^{M} \gamma_m$, where $\gamma_m = \frac{P_m}{N_{0m}} / \frac{P_1}{N_{01}}$ is

the received signal carrier-to-noise ratio of each antenna to the reference antenna. The total power of the combining signal, thus, is given by

$$\begin{aligned} P_Y &= \overline{\left| [\sum_{m=1}^{M} \beta_m (\sqrt{P_m} e^{j(2\pi f_{IF} t_k + \Delta\varphi_{mref}(t_k))})] \right|^2} = \overline{\left| [\sum_{m=1}^{M} \sum_{n=1}^{M} \beta_m \beta_n (\sqrt{P_m}\sqrt{P_n} e^{j(\Delta\varphi_{mn}(t_k))})] \right|^2} \\ &= \sum_{m=1}^{M} \beta_m^2 P_m + \sum_{m=1}^{M} \sum_{\substack{n=1 \\ n\neq m}}^{M} \beta_m \beta_n (\sqrt{P_m}\sqrt{P_n}\overline{C_{mn}})] \\ &= P_1 [\sum_{m=1}^{M} \gamma_m^2 + \sum_{m=1}^{M} \sum_{\substack{n=1 \\ n\neq m}}^{M} \gamma_m \gamma_n \overline{C_{mn}}] \end{aligned} \tag{5.19}$$

where the correlating phase is $\Delta\varphi_{mn}(t_k) = \Delta\varphi_{mref}(t_k) - \Delta\varphi_{nref}(t_k)$. The signal decay function, $\overline{C_{mn}}$, is given by

$$\overline{C_{mn}} = E(e^{j\Delta\varphi_{mn}(t_k)}) = \begin{cases} e^{-\frac{\sigma^2_{\Delta\varphi_{mref}} + \sigma^2_{\Delta\varphi_{nref}}}{2}} & , m \neq n, \\ 1 & , m = n \end{cases} \tag{5.20}$$

The SNR of the combining signal is given by

$$SNR_{SUM} = \frac{P_Y}{\sigma_Y^2} = \frac{P_1}{N_{01} B_s} \times \left[\frac{\sum_{m=1}^{M} \gamma_m^2 + \sum_{m=1}^{M} \sum_{\substack{n=1 \\ n\neq m}}^{M} \gamma_m \gamma_n \overline{C_{mn}}}{\sum_{m=1}^{M} \gamma_m} \right] \tag{5.21}$$

The combining gain and loss is respectively given by

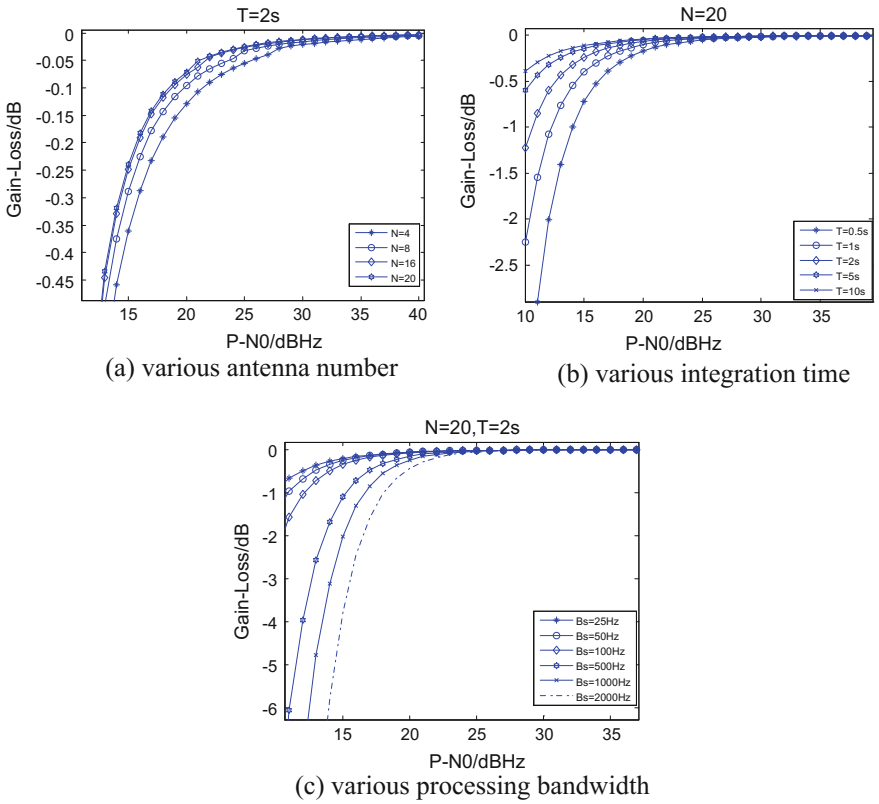
$$\begin{aligned} Gain_{SUM} &= 10 * \log 10 \left(\frac{\sum_{m=1}^{M} \gamma_m^2 + \sum_{m=1}^{M} \sum_{\substack{n=1 \\ n\neq m}}^{M} \gamma_m \gamma_n \overline{C_{mn}}}{\sum_{m=1}^{M} \gamma_m} \right) \\ D_{FX} &= 10 * \log 10 \left(\frac{\sum_{m=1}^{M} \gamma_m^2 + \sum_{m=1}^{M} \sum_{\substack{n=1 \\ n\neq m}}^{M} \gamma_m \gamma_n \overline{C_{mn}}}{\left(\sum_{m=1}^{M} \gamma_m \right)^2} \right) \end{aligned} \tag{5.22}$$

From the results, the combining loss of the iterative FX correlation combining algorithm is a function of the antenna number, the signal-to-noise spectrum density ratio of single antenna, the signal bandwidth and the integration time.

5.4 The Simulation Results and Analysis

5.4.1 The Algorithm Performance Under Various Input Conditions

Under the precondition of aligning delay and frequency differences, the performance curve of the iterative FX correlation combining algorithm versus the antenna number, the integration time and the signal bandwidth, as shown in Fig. 5.3a–c.



(a) various antenna number

(b) various integration time

(c) various processing bandwidth

Fig. 5.3 Performance curve of the algorithm under various conditions.

From the figure above, the antenna number is not a master factor in influencing the performance loss of the iterative FX correlation combining algorithm, whereas the signal-to-noise spectrum density ratio of single antenna, the signal bandwidth and the integration time are the master factors in influencing the combining performance.

5.4.2 *The Performance Comparison of Various Combining Methods*

To illuminate the performance and applicability of the iterative FX correlation combining algorithm in the deep space multiple antenna arraying, the performance comparisons of the algorithm with various arraying methods based on the SUMPLE algorithm (such as FSC, BC, SSC and the combination of BC and CA) are simulated. The simulation conditions are shown in Table 5.1.

The performance curve of various arraying methods is shown in Fig. 5.4.

The figure above shows that: firstly, the combining performance of all arraying methods is similar under a high input SNR; secondly, FSC and SSC require a better input SNR, and their performance would degrade rapidly under a low SNR; thirdly, compared with the FSC, SSC and BC methods, the iterative FX correlation combining algorithm has a better performance under various input SNR, and its performance is similar to the combination of BC and CA.

Table 5.1 Simulation conditions

Antenna number	20 antennas with the same aperture	IF filter bandwidth	200 kHz
Signal type	telemetry signal	Integration time	5 s
SNR of the received signal	15– 40 dBHz	Precision of delay prediction	<50 ns
RF frequency	8.4 GHz	Precision of frequency prediction	<1 kHz
IF frequency	70 MHz	Precision of frequency difference prediction	<0.05 Hz
Subcarrier frequency	65,536 Hz	Length of FFT	1024
Subcarrier data rate	100 bps	Sample rate	56 MHz
Modulation	30°	Length of integrate and dump	64

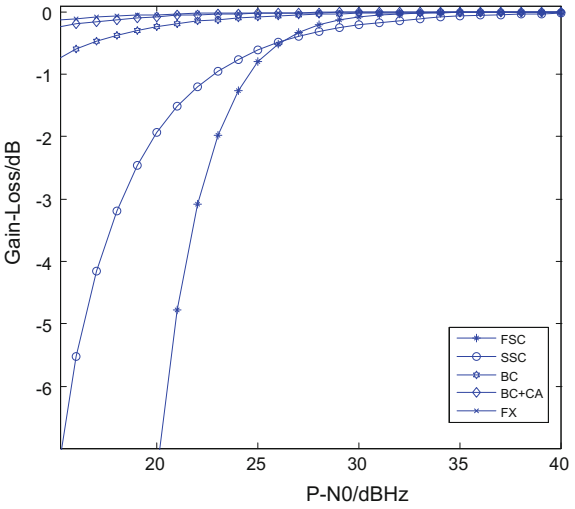


Fig. 5.4 Performance comparison curve of various combining methods

5.5 Conclusions

In allusion to the problem of the efficient combination of the weak signal in Mars arraying, a novel frequency-domain iterative FX correlation combining algorithm is proposed. Through the "one-to-more" correlating processing in frequency domain, the algorithm could precisely obtain the correlating phase between each antenna and the reference antenna in the signal bandwidth, and get the estimation and compensation of the multi-antenna residual delay and frequency differences. The algorithm is similar to a FSC implementing architecture combined with match filters, which could get a better performance than the FSC, BC and SSC combining methods in the weak signals. Compared with the combination method of BC and CA with the similar performance, the iterative FX correlation combining algorithm has a simpler algorithm structure and a lower implementation complexity. The algorithm, therefore, is a good implementation scheme for the Mars antenna arraying.

References

1. Bagri DS (2004) A proposed array system for the deep space network. IPN progress report 42–157, Jet Propulsion Laboratory, Pasadena, pp 1–16
2. Rogstad DH, Mileant A, Pham TT (2005) Antenna arraying techniques in the deep space network. Li HT translating. Tsinghua University Press, Beijing
3. Howard LD (1980) Prototype real-time baseband signal combiner. TDA progress report 42–60, Jet Propulsion Laboratory, Pasadena, pp 145–151

4. Rogstad DH (2005) The SUMPLE algorithm for aligning arrays of receiving radio antennas: coherence achieved with less hardware and lower combining loss. IPN progress report 42–162, Jet Propulsion Laboratory, Pasadena, pp 1–29
5. Navarro R (2012) Frequency domain beamforming for a deep space network downlink array. IEEE
6. Lu MH et al (2013) Analysis of the combining efficiency of full-spectrum combining for antenna arrays. J Spacecr TT&C Technol 32(5):383–388
7. Meng W, Liu Y, Zhang YN (2014) Experiment and analysis of antenna arraying for downlink wideband signals. Radio Eng 44(9):28–31
8. Zheng WM, Shu FC, Zhang D (2008) Application of software correlator to deep space VLBI tracking. J Astronaut 29(1):18–23

Chapter 6
Research on RF Link Management Technology of Telemetry Network System

Xinglai Wang, Kun Lan, Guojiang Xia and Ming Han

6.1 Introduction

It is often needed to measure the parameters of the aircraft in the process of flight test, and the measurement results will be transmitted to ground stations by radio waves. This technology is called telemetry [1]. Telemetry system covers a series of key technologies including the sensor, information collecting, electromagnetic compatibility, wireless communication and so on, and it is an important part of aerospace military system. Ground stations can gain aircraft data accurately and timely through telemetry system and guarantee the experiment safety and efficiency effectively.

In the process of telemetry technology development, the IRIG-106 telemetry standard is the most classical achievement. The standard requires that telemetry system is unidirectional from aircraft to ground station with the information in the form of PCM/FM, in S or L band for point-to-point transmission. The vast majority of telemetry systems use or reference IRIG-106 telemetry standards all over the world [2]. In recent years, with the rapid development of electronics, new materials, computer technology, the craft systems are always changing, but their telemetry system architecture has little change. The traditional IRIG-106 system cannot meet the growing demand of the flight test.

The United States began to develop integrated Network Enhanced Telemetry (iNET) in 2004, which can manage multiple test articles (TAs) and ground stations and realize telemetry network. At present, the iNET standard has been published and improved continuously. Telemetry Network System (TmNS) is the main part of

X. Wang
College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, China

X. Wang · K. Lan · G. Xia · M. Han (✉)
Beijing Institute of Astronautical Systems Engineering, Beijing 100076, China
e-mail: hm890311@163.com

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_6

the iNET. The Radio Frequency (RF) link has not only the original IRIG-106 function, but also realize connection the test articles network and ground stations network [3]. So, TmNS RF network management is extremely important and need to coordinate the allocation of resources to meet the performance requirements of new telemetry network system.

Released in February 2010, «iNET RF Network Element Standard» (V0.7) has comprehensive specifications on the RF network information, interface, security and other issues, and specifically introduces the link management protocol. Link management protocol ensures multiple TAs interoperability in the same range or different ranges and concludes the main specification of wireless link establishment, release, capacity management, forward within TmNS and between TmNS, power control, and so on. The protocol makes the order and behavior of the message transmission clear. Link management protocol is still improved at present.

After simple introduction of TmNS architecture,RF link management technology is elaborated and analyzed in this paper.

6.2 iNET TmNS Architecture

As the core of iNET, TmNS is mainly divided into three parts: the Test Article Segment (TAS), Ground Station Segment (GSS), Radio Access Network Segment (RANS), and also includes two support entities: system management, and metadata (used to describe the Network equipment configuration, etc.). The overall architecture is shown in Fig. 6.1 [4, 5]. TmNS focuses on two networks: the vehicle Network (vNET) and Radio Access Network (RAN). RAN can control one or more radios and radios is the key to the RF communication. The Ground Network (gNET) varies due to the range, and interfaces to the gNET are provided by TmNS.

TAS and GSS all have their own radios and antenna units, which implements the RF interconnection between the vNET and gNET [6]. The RF link has two functions: one is the classic IRIG-106 serial data stream transmission, the other is a new

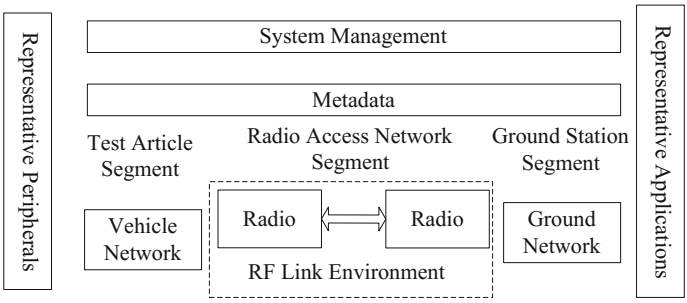


Fig. 6.1 TmNS architecture

bidirectional data link (upward and downward) network in order to realize broadband network between the multiple TAs and multiple the ground stations. Therefore, there will be data in two forms—the PCM data stream and IP packets in TmNS. Compared with traditional telemetry, the function is greatly enhanced [7]. There is a set of special channel access methods in TmNS, which not only guarantee the reliability of the bidirectional link, but also achieve efficient use of spectrum resources. RF Link Manager (LM) is a key element of the process and it is very sensitive to the network connection, transmission queue and network management. LM can be adjusted according to the real-time wireless transmission window and resources to improve TmNS performance.

6.3 RF Link Management

6.3.1 Overview of Link Management

TmNS uses channel access way based on TDMA to complete reliable information interaction between TA and GS, as shown in Fig. 6.2. Time is divided into many Epoches in each distribution frequency point, and the Transmission Opportunity (TxOp) with a certain time interval in Epoch is assigned to each link. In this way, time division multiplexing between uplink and downlink is implemented at the same frequency. Communication equipment used in TmNS network bidirectional transmission is the same as most of the serial data stream telemetry equipment [8].

Link capacity is also called link bandwidth, refers to the amount of data in a unit time that a link can transmit. RF link management system can achieve the best use of a given spectrum through the centralized link manager. LM is the control core of radio TxOp to ensure that the resources can be shared network capacity by TmNS in the range test distributing frequency. LM can allocate the TxOp according to link capacity to achieve the low latency of sudden communication services and reduce the possibility of information loss caused by a buffer overflow.

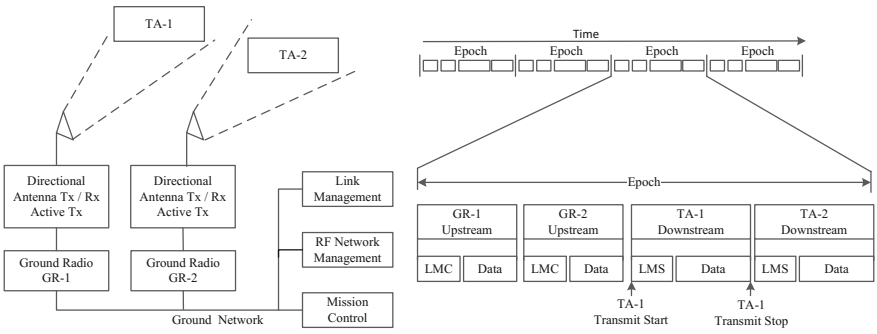


Fig. 6.2 TmNS channel access (including two TAs)

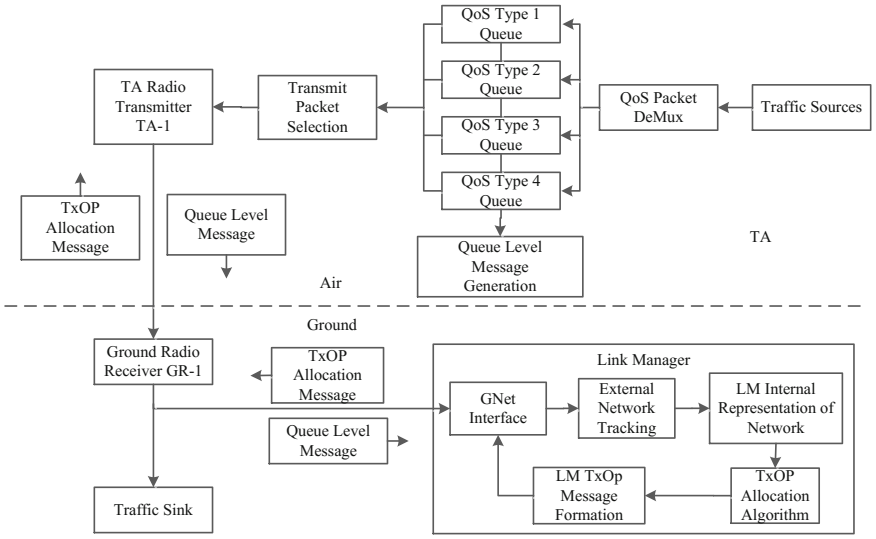


Fig. 6.3 The downstream data and control information transmission process

The ground radios receive the control information that link manager sents to the range ground network. The information provides the uplink transmission start and end time accurately. At the same time, the link manager located in the ground also sends control information to the TA in the air, and the exact time of the downlink transmission of TA is specified to control the data flow of the transmission to the ground equipment. The downstream data transmission process as shown in Fig. 6.3.

The data obtained from TA source is classified and queued according to the Quality of Service (QoS) to transmit. TA radios select packets from the QoS queues that need to transmit and sent them into the transmission windows, which are assigned by the LM. The queue level information in the radio signal is sended to LM in the ground. Queue manager level information to LM and transmission window control information to the TA, this dynamic control loop is the basic process of RF link management.

To achieve the best access to the RF channel, radios and LM use a specific Differentiated Services Code (DSCP) to flag data streams. Radios use the QoS to supervise and rank data flow in the RF channel directly. LM distributes channel capacity through making use of the mark of data flow and each queue level of radios, and based on traffic demand [9]. That is to say, the priorities of data transmission in TmNS are decided by the radios internal queue rules and TDMA network link manager. LM periodically receive queue rating report of each wireless device and compute capacity distribution and adjust the link transmission capacity according to the radio and the network environment to realize QoS data flow control of the whole range.

LM uses low-latency control information and can be free to dynamically adjust the capacity of each link in order to satisfy the unpredictable demands of multiple TAs. When a sudden flow of data to improve the capacity of the link, the LM must make rapid response to new capacity distribution to decrease the increase of time delay and avoid the queue overflow. LM ensures TmNS normal operation of RF environment, but under different network environment, how to improve the efficiency of the distribution of capacity, how to maintain the QoS of data flow effectively, are the practical problems that are necessary to consider.

6.3.2 Dynamic Capacity Allocation Algorithm

In multiple access TmNS, It is necessary to allocate capacity for multiple tasks with different priorities to guarantee QoS so that physical dispersion nodes share RF link. This function is part of the link management and shall be the responsibility of the LM. The capacity allocated by link manager consists generally the minimum capacity based on the running needs (fixed part) and the allocated capacity based on the current demand (dynamic). The minimum capacity sets minimum capacity value allocated and the shortest transfer time of every link. Every task in the network has its own requirements and minimum capacity must meet these requirements. The allocated capacity is related to the steady state level and the actual average demand shall generally be below this value. Various dynamic capacity allocation algorithm are all for the latter.

All kinds of QoS level queues are determined “class weight” according to the DSCP, while the RF link are determined “link weight” according to Mission Service Level Profile (MSLP) priority/weight. LM gets requirements and queue depth of the current each task/QoS level data flow from TA and ground network node, as shown in Fig. 6.4. As TDMA controller, LM uses these inputs to allocate time slots through specifying the RF channel capacity, and generate TxOp information and build uplink and downlink based on packets and task priority.

iNET Management and Operations with the Policy Controls (iMANPOL) project has developed some technologies, which are committed to provide end-to-end QoS for advanced telemetry network. Especially iMANPOL capacity allocation algorithm, it can help LM to deal with severe congestion and existing bad behavior users or data flow. Its fundamental purpose is to protect the “good” link (not at the expense of the other links, not cause system overload operation) and punishment “bad” link (i.e., overload link). The goal of capacity allocation is when traffic is beyond the capacity of radios, it can minimize the delay. If capacity can be adjusted according to the requirements of data flow, data loss caused by queue overflow or timeout can be reduced.

Now Baseline Algorithm and Enhanced Algorithm is dynamic capacity allocation Algorithm with high recognition,respectively B-LM and E-LM Algorithm for short [10].

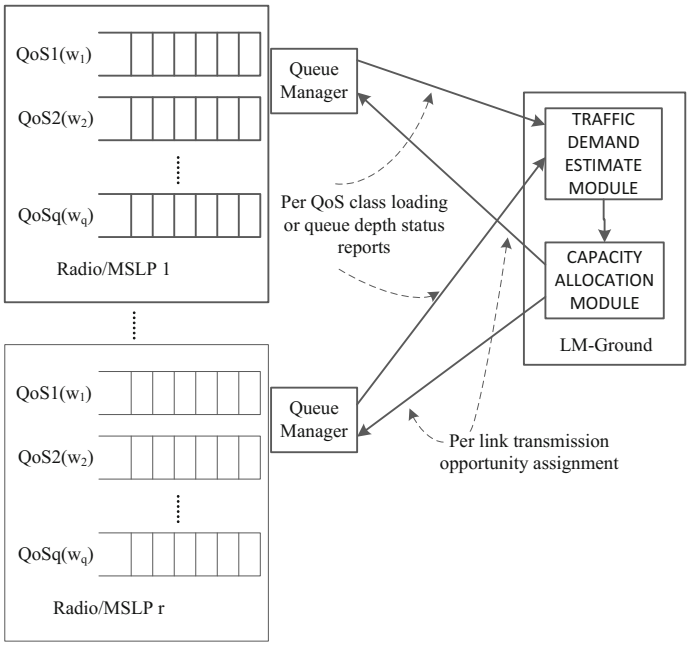


Fig. 6.4 Advanced LM system architecture

(1) B-LM Algorithm

B-LM algorithm can allocate the required capacity based on queue class weight and demand ratio, and demand ratio refers to a proportion that the demand of a given queue accounted for each class data flow demand of all links. After minimum capacity is distributed to each link (guarantee basic communication), the remaining total capacity C_{total} is assigned to each class data flow, According to the demand radio:

$$R_{\text{demand}} = W_{\text{link}} \times W_{\text{class}} \times (D_{\text{class}}/D_{\text{total}}) \tag{6.1}$$

$$C_{\text{allocation}} = C_{\text{total}} \times R_{\text{demand}}, \tag{6.2}$$

where R_{demand} is demand ratio, W_{link} is link weight, W_{class} is class weight, D_{class} is the demand of this class data flow, D_{total} is the total demand of all class data flows, $C_{\text{allocation}}$ is allocated capacity.

B-LM can be applied to most of the data flows and the performance is good.

(2) E-LM Algorithm

E-LM stems from iMANPOL project. There is "guaranteed" part of the total available bandwidth in each class data streams which be referred to as "*Quota*".

Quota is proportional to the class weight and it can be dynamically adjusted according to the proportion of current effective demand capacity.

If applied to the high priority data flows, *Quota* can prevent low priority from capacity shortage problem to ensure the fair. The *Quota* allocated to high priority flows (abbreviated to *Q*) is higher than that of low priority flows, which is proportional to class weight (*W*), namely

$$Q_{\text{high}}/Q_{\text{low}} = W_{\text{high}}/W_{\text{low}} \tag{6.3}$$

If the "bad" flow exceeds *Quota*, it could be downgraded, but it should not cause negative influence on other class data flows. If the traffic of a data flow is below the *Quota*, the balance will be shared with other class data flows. If the traffic of a data flow is higher than the *Quota*, but some other class flows is less than the *Quota*, the capacity requirements can still be met. These principles are applied in the environment where multiple data streams of different classes and multiple links exist, and network performance will be effectively improved.

E-LM algorithm process is shown in Fig. 6.5.

In the first step, the allocated capacity is min (*Quota*, *Demand*), where *Quota* = total capacity × link weight × class weight/all weight, *Demand* for capacity demand. It is important to note that, with the same as B-LM algorithm, the total capacity also here refers to remaining capacity after the minimum capacity

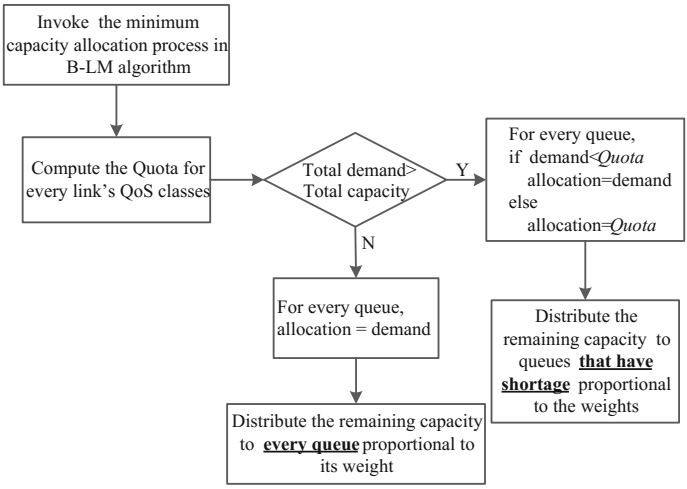


Fig. 6.5 The E-LM algorithm

distribution for each link. *Quota* depends on the link/class weights (fixed) and usable capacity (different), but it does not depend on demand to prevent the greed flow from gaining too much capacity in the overload condition.

The second step, the remaining capacity will be assigned to other queues according to the demand and weight.

(3) Comparison and Analysis between B-LM and E-LM

B-LM algorithm configuration is simple, while E-LM is relatively complex with more parameters. As for performance, if the link capacity load does not exceed the total demand, there is no difference between B-LM and E-LM algorithm generally.

However, B-LM deals with "good" flow and "bad" flow in the same way and it leads to risk. When the queue experience accumulate overmuch, link capacity may be incorrectly assigned to "good" flow and "bad" flow may occupy capacity of the same class flow in other tasks or links, and it will lead to too much delay. In addition, the capacity may not be made full used. If the links loaded different class flows are orthogonal (that is, each class flow has a separate link), it will reduce the overall throughput of the system.

Because *Quota* ensures each communication queue, so no matter what the overload condition is, E-LM can work normally. Only when "bad" data flow capacity is over *Quota*, some disadvantages will appear. E-LM protects the "good" data flow. If the "bad" data flow capacity is less than the *Quota*, their needs may be met. Visible, regardless of the existence of "bad" data flow, E-LM will greatly reduce the delay of good queues.

B-LM algorithm and E-LM algorithm can coexist in LM. LM can use static configuration or dynamic switching according to the data flow. When the additional logic components monitor the serious overload or highly unbalanced traffic patterns, E-LM will be activated; when traffic/mode is back to normal operation system, B-LM will be reactivated.

6.3.3 The Link Manager Software Architecture

LM support Linux, Windows and OPNET platform and usually run as a daemon. At present the mainstream of the LM development is based on Linux, because Linux has the advantage of real-time process and it is more advantageous to gain high performance. Logically speaking, LM software architecture can be divided into four functional areas: LM external interface layer, LM layer, LM operating system (OS) concept layer and LM universal service layer [11]. Representative LM software architecture as shown in Fig. 6.6, this framework can minimize dependence among components and improve the scalability and reliability. It provides a flexible implementation scheme for LM deployed on different platforms.

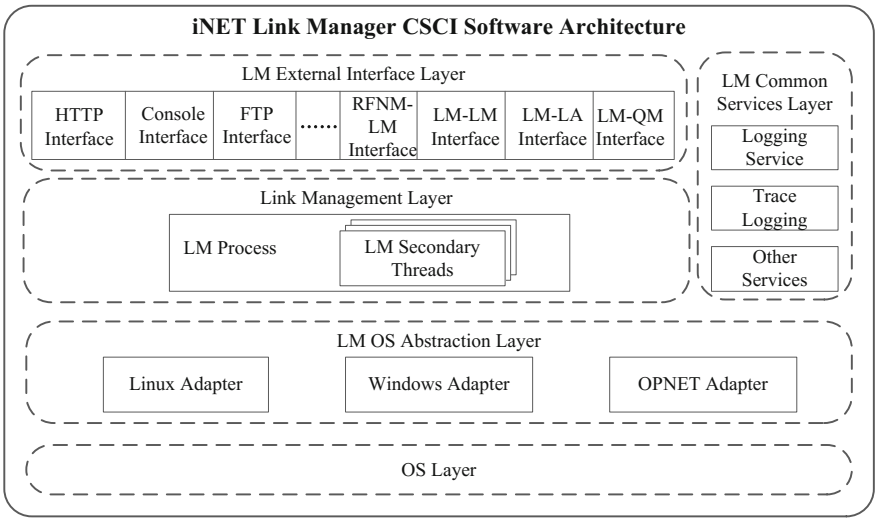


Fig. 6.6 LM software architecture

6.4 Conclusion

As a new network telemetry system, TmNS shares limited spectrum channel and enhances the efficiency of the telemetry task of range test through bidirectional communication link. To implement multiple TAs and multiple GS of the wireless network connection, centralized link manager is the key equipment. Wireless transmission window can be coordinated and the resources can be allocated according by LM according to the network situation. There is important reference value and practical significance to research and development with Chinese characteristics of a new generation of network telemetry system that RF link management principle and process, dynamic capacity allocation algorithm, software architecture are researched and discussed.

References

1. Bai XX, Yang TW, Yuan BN (2010) The telemetry development trend and actions in aviation flight test. Meas Control Technol 29(11):6–9
2. Zhang JM, Yuan BN, Bai XX (2010) The technical framework and application prospect of iNET. Meas Control Technol 29(11):15–17
3. Liu D, Zhang QQ, Li BL (2015) Research on system control and network protocol of TmNS. China Sci Technol 03:93–94
4. Integrated Network Enhanced Telemetry (iNET) RF network element standard, Version 0.7 (2010)

5. Integrated Network Enhanced Telemetry (iNET) Communication link standard, Version 0.7 (2010)
6. Yuan BN, Huo ZH, Bai XX (2010) The new generation network system—TmNS. Meas Control Technol 29(11):18–21
7. Liu D, Huo JH, Guo S (2015) Research on Telemetry network system and transmission network. Comput Meas Control 23(3):1023–1025
8. O'Connell R, Webster L (2011) Telemetry network system (TmNS) link management modeling and simulation. In: Proceedings of the 47th annual international telemetering conference ITC/USA, pp 231–238
9. O'Connell R, Webster L, Kaba J (2012) Telemetry network system (TmNS) RF link management quality of service. In: Proceedings of the 48th annual international telemetering conference ITC/USA, pp 258–266
10. Fecko M, Chang K, Cichocki A et al (2014) Dynamic capacity allocation algorithms for iNET. In: Proceedings of the 50th annual international telemetering conference ITC/USA, pp 321–330
11. O'Connell R, Webster L, Kaba J (2013) Telemetry network system (TmNS) RF link management verification. In: Proceedings of the 49th annual international telemetering conference ITC/USA, pp 398–406

Chapter 7
Analysis of Radar Signals Induced by Symmetric Antenna Interference Region of Transponder

Bo Yang, Lianwen Meng, Yunfeng Liang, Yongsheng Zhang and Hui Zhou

7.1 Introduction

The velocity measurement radar (VMR) [1] can measure aircraft radial velocity using the Doppler frequency shift in Telemetry, Track & Command (TT&C) field. In the paper, a kind of high-precision velocity measurement radar is researched of which major station can launch f_{t1}(xxx5) or f_{t2}(xxx7.9) one of the two kinds of carrier frequency waveform, which both major and slave stations (no transmitter) can receive f_{r1}(xxx8) and f_{r2}(xxx9.6) two kinds of carrier frequency waveform at the same time. To extend radar detection distance, the transponder on the aircraft can be used to improve radar receiver's signal noise radio (SNR). When the aircraft flying smoothly, transponder's antennas located at the surface of the aircraft can receive the signal from radar transmitter and forward signal to radar. Due to the target's rotation and other micro motion including such as vibration, coning motion of the target, the signal discontinuity [2] is generated by antenna gain directivity restrictions of the transponder. The solution to separate the problem has been studied by FENG with multiple antennas [3, 4]. Therefore, symmetrical antennas of the transponder are usually used to solve the discontinuity in engineering practice. However, signal synthesis and signal interference which affect radar received signals obviously will happen when the deviation of radar-target direction and transponder's symmetrical antennas are near 90° [2]. In this paper, we analyzed the effect of the signal interference with velocity measurement radar (VMR) for rotating targets based on the analysis of Automatic Gain Control (AGC) of received signal, Doppler, micro-Doppler, error voltage and so on, especially radar signals in symmetric antenna interference region.

B. Yang (✉) · L. Meng · Y. Liang · Y. Zhang · H. Zhou
Taiyuan Satellite Launch Center, Taiyuan 030031, China
e-mail: ugoder2005@163.com

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_7

7.2 The Oretical Analysis of Returned Signal Induced by Symmetric Antenna Interference Region

7.2.1 Analysis of Transponder's Antenna

A kind of transponder [5] of the VMR can both receive f_{t1} and f_{t2} uplink frequency signal and forward f_{r1} and f_{r2} two kinds of downlink signal; the transponder also can generate two kinds of downlink frequency signal by using its high-stability crystal oscillator when no radar transmitted signal. The transponder has three kinds of work mode, first is duplex measurement in same source, that is, the major station launch signal, the transponder forward, and the major station receive signal; Second is duplex measurement in differ source, that is, the major station launch signal, the transponder forward and the slave stations receive signal; Final mode is simplex measurement, that is, transponder generate downlink signal by itself, radar stations receive signal.

Considering rockets or missiles are usually cylinder-shaped, transponder on the target uses conformal microstrip antennas generally [6], namely the plane or cylindrical microstrip antennas. The gain pattern of the plane antenna can be represented by formula (7.1) according to engineering experience and relative theory [3]; Where r is radius of target aircraft, λ is the wavelength of the radar wave, α is the angle of rotation from antenna located position to the current position in the equatorial plane with a range of [0,360]. Every antenna effective radiation range is limited in an angle $\pm 90^\circ$ from its located position due to the shade effect of cylindrical target because shade of the target can decrease radiation intensity when the transponder antenna is back to radar antenna.

$$F(\alpha) = \frac{\cos\left(\frac{\pi}{2}\sin(\alpha)\right)}{\pi.\cos(\alpha)} \cdot e^{-j\frac{2\pi}{\lambda}r} \tag{7.1}$$

The radiation of the symmetric antennas of the transponder will synthesis under the condition of radar far field, the references of signals synthesis of symmetric antennas are rare in open literature. The synthesis gain pattern of the signal is related to factors such as antenna type, size, material and diameter of target. Considering the radar tracking distance is far greater than the distance between the symmetric antennas in the far field region, the symmetric antennas can be considered as a binary antenna array approximately. Through the simulation of antenna directivity function [7], the pattern of the symmetric antennas can be shown in Fig. 7.1; where the circle angle denotes the angle of space direction and the radius denotes the radiation intensity of the electric field.

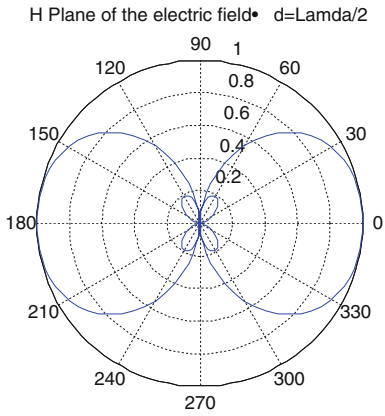


Fig. 7.1 Radiation intensity of the electric field versus space angle

7.2.2 *Analysis of the AGC*

For instance, in the tracking mission of rotating target, the transponder [5] of VMR is located at the fore cabin of the aircraft. One pair of symmetric antennas were located at $\alpha_{m1} = 47°$ and $\alpha_{m3} = 227°$ for a carrier frequency signal (f_{r1}) and the other pair of symmetric antennas were located at $\alpha_{m2} = 137°$ and $\alpha_{m4} = 317°$ for the other carrier frequency signal (f_{r2}). The experiments show that the Relationship between the gain pattern and the angle of β can be shown in Fig. 7.2 [8] when the radar antenna aimed at the antenna of the transponder (α Angle difference is 0°) or in Fig. 7.3 when the radar antenna aimed at the interference region of the transponder (α Angle difference is 90°); where β denotes the angle between the line of longitudinal axis of the target and the line of radar to target. Received signal gain of the transponder antenna was shown using the solid line while the transmitted signal gain was shown using the dotted line. In addition, two typical cases of the gain pattern were listed in Figs. 7.2 and 7.3 to compare with each other to show the difference between the maximum and the minimum. The gain values of the signal are almost above 0 dB with a maximum value of 4 dB when the β angles are in the range of 40°–120° and α angle difference is 0 in Fig. 7.2; while the gain values of the signal are almost below −10 dB with a minimum value of −17 dB when the β angles are in the range of 40°–120° and α angle difference is 90° in Fig. 7.3. The gain difference between Figs. 7.2 and 7.3 is in the range of 10–21 dB.

The distance compensation is necessary to deduct with the formula of $20\lg(\lambda/4\pi R)$ to improve the accuracy of the AGC value of received signal; where λ is the wavelength of radar signal, R is the radar to target range. The AGC with the offset value represents the gain characteristics of the transponder antennas more precisely. Assuming that transponder forward power is 1 W, atmospheric attenuation and polarization attenuation [9] are both 5 dB. Because the tracking sensitivity of VMR is −160 dBW, the received antenna gain is about 38 dB, the range attenuation L is −167 dBW when target from radar at the precision tracking range 1000 km, attach the transponder antenna gain [−17,4], the radar receiver's signal

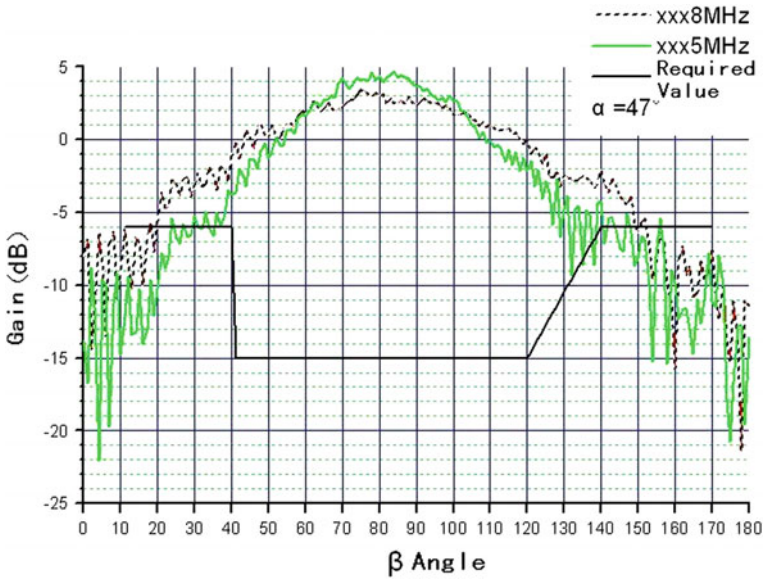


Fig. 7.2 The gain pattern versus β angle [0,180] where α angle is 47°

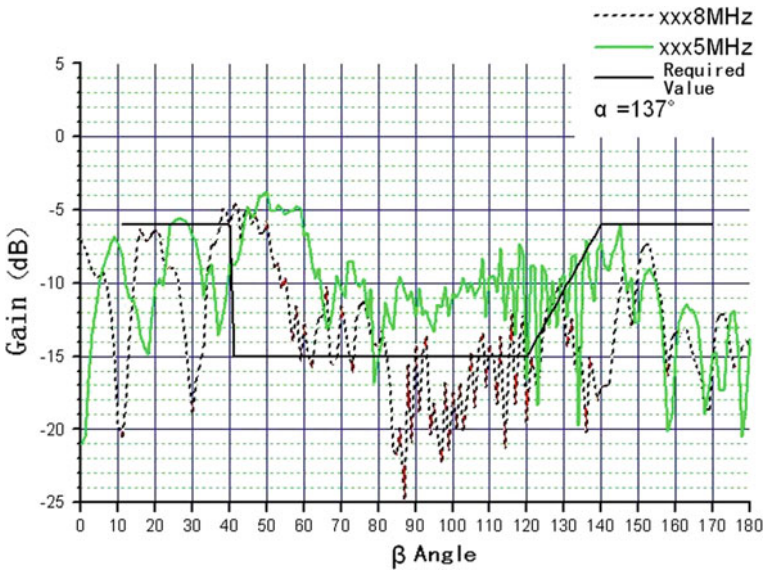


Fig. 7.3 The gain pattern versus β angle [0,180] where α angle is 137°

power is in range [−156,−135], which is higher than the tracking sensitivity, as a result, the interference region of transponder can't lead to signal discontinuity when rotation in precision tracking range.

The fluctuant AGC can affect the error voltage of antenna azimuth (U_A) and the error voltage of elevation (U_E) when the target rotates. The error voltage is determined by the antenna angle error signal which is related to the angle of deviation, difference scope of deviation gain pattern and the amplitude of the sum signal (AGC) [10] considering the synthesis of error voltage (U_Δ) can be calculated by formula (7.2).

$$U_\Delta = \sqrt{U_A^2 + U_E^2} \tag{7.2}$$

7.2.3 *Analysis of Doppler and Micro-Doppler Induced by Symmetric Antenna Interference Region*

Two stations of VMR track the target with one transponder and four transponder antennas including A, B, C and D in Fig. 7.4. The antenna A and C are used for one radar signal (f1) while B and D are used for the other radar signal (f2). The point O is the center of target; the path P_1 D is actual path of the microwave signal of the station P_1 while the path P_1O is necessary path in theory. The relationship of P_1 D and P_1 O can be shown in formula (7.3) as the geometrical shape of target and the radar theory of far field region; where r represents the radius of the cylinder-shaped target, ω represents the rotation speed, θ represents the initial angle of the antenna location, t represents the time.

Assuming that target's rotation is the only consideration without the factors of yaw angle and pitching angle. The phase of the received signal, difference of the phase and difference of Doppler can be represented by formula (7.4); where $\Delta\Phi$ represents the phase difference of the received signals between before and after the moment when the transponder antenna switch from one to the other; Φ_1 and Φ_1' represent the phase and the difference of phase of signal before the switch of the

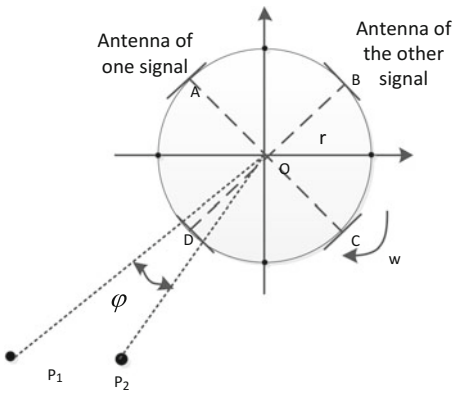


Fig. 7.4 Two radar station P_1, P_2 tracking spinning target

transponder antenna while Φ_2 and Φ_2' represent after the switch of the transponder antenna; $\Delta\Phi'$ denotes the difference of Doppler which is the difference of phase.

$$P_1D = P_1O - r\cos(\theta + \omega t) \tag{7.3}$$

$$\begin{cases} \Delta\Phi = \Phi_1 - \Phi_2 = 0 \\ \Delta\Phi' = \Phi_1' - \Phi_2' = \frac{2}{\lambda}[r\omega \sin(\theta + \omega t) \\ \quad + r\omega \sin(\theta - \omega t)] = \frac{4r\omega}{\lambda} sin(\theta) \end{cases} \tag{7.4}$$

It was shown in Fig. 7.4 that the bistatic Doppler is the difference of the sum of P_1 O and O P_2 and the bistatic micro-Doppler [11] was generated with rotation of target represented in formula (7.5); where f_{DBR} represents the bistatic micro-Doppler, f represents the frequency, c represents the velocity of light, φ represents the angle between the line of station P_1 to target and the line of station P_2 to target.

$$f_{DBR} = \frac{2f}{c}[\omega \times r(t)]\cos(\varphi/2) \tag{7.5}$$

A frequency of VMR f_1 is xxx8 MHz. Assuming that the radius of the target is 0.7 m, the speed of target rotation is 0.4π rad/s; as a result, the difference of Doppler in symmetric antenna interference region is around $4r\,\omega/\lambda$ 63 Hz by formula (7.4) when θ is 90°. In peculiar case for micro-Doppler, the major station plays the roles of P_1 and P_2 at same time, then $\varphi = 0$ the $\cos\left(\frac{\varphi}{2}\right) = 1$, the peak value of the micro-Doppler induced by target rotation is about 34 Hz by formula (7.5). The amplitude of micro-Doppler is closely related to the radius and rotation speed of target and radar frequency, In addition, the bistatic micro-Doppler (slave station) also depends on the triangulation geometry of the transmitter, the target, and the receiver.

7.3 Analysis of Data Gathered by Radar Stations Based on AGC, Error Voltage, Doppler and Micro-Doppler

7.3.1 Data Analysis of AGC and Error Voltage

Focus on the tracking mission mentioned in Sect. 7.2.2, we analysed the data gathered by radar station A, B and C which including time, Doppler, angle of AZ and EL, U_A and U_E and so on with a sampling rate of 20 Hz; where A station is major station with transmitter and receiver and station B and C with receiver only. The AGC value versus the angle of rotation α was shown in Fig. 7.5. The results in Fig. 7.5 show that the AGC values fluctuate from trough to crest with a range of 10–20 dB and a period of 2.5 s around when the target rotates periodically. The crests of the AGC are correspond with the moments when the radar antenna aimed at the transponder antenna; while the troughs of the AGC are correspond with the moments when the radar antenna aimed at the antenna interference of the

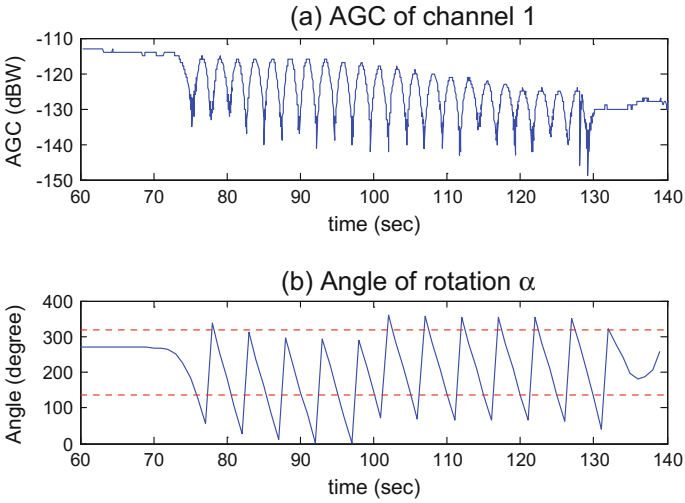


Fig. 7.5 **a** The AGC of channel 1 versus time; **b** the angle of rotation α versus time

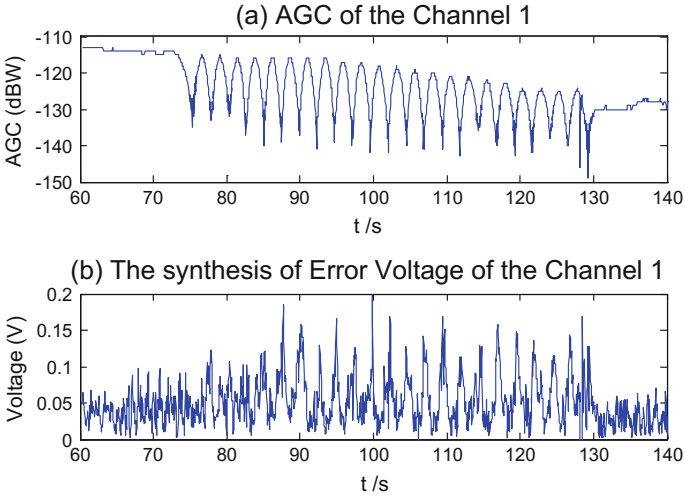


Fig. 7.6 **a** The AGC of channel 1 versus time; **b** the synthesis of error voltage of the channel 1 versus time

transponder. However, the minimum value of AGC is above the receiver sensitivity of radar which is the minimum value of the received signal by receiver.

The integrated error voltage (U_Δ) can be calculated by formula (7.2) using recorded data, The U_Δ versus the AGC value was shown in Fig. 7.6. The results in

Fig. 7.6 show that the value of U_{Δ} rise and fall more obviously in the rotation period of time than other period of time.

7.3.2 Data Analysis of the Doppler and the Micro-Doppler

According to the gathered data, the crest of the difference of the Doppler is more about from 50 to 100 Hz, which is accord with the theoretical value 79 Hz calculated by the formula (7.3) in 7.2.3 section. The micro-Doppler can be obtained from the recorded Doppler using Daubechies 3 wavelet transform [12]; the micro-Doppler versus the angle of rotation is shown in Fig. 7.8 where the micro-Doppler of f1 was shown using the solid line and f2 was shown using the dotted line. The results in Fig. 7.8 show that micro-Doppler generated by target rotation will reach its peak value in antenna interference region with plus and minus two peaks before and after interference region; in addition micro-Doppler period is twice the target rotating period; the micro-Doppler amplitude of the major station were in range of [20, 40] Hz and the micro-Doppler amplitude of slave stations were more smaller by the triangulation geometry of the transmitter, the target, and the receiver (Fig. 7.7).

The auto-correlation [13] is the common method of period detection [14] of the signal. In this case, it is necessary to remove the both edges of the data because the wavelet transform causes the false data at the edges. The auto-correlation of B station versus the data point was shown in Fig. 7.9. The period of micro-Doppler can easily estimate with 50 points which is 2.5 s as the sampling rate is 20 Hz. The theoretical period of the micro-Doppler is half of the period of the rotation because

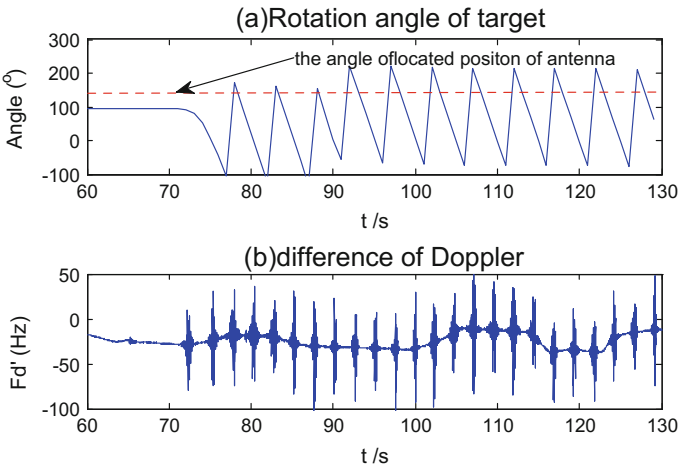


Fig. 7.7 **a** The rotation angle of target versus time **b** the difference of Doppler versus time

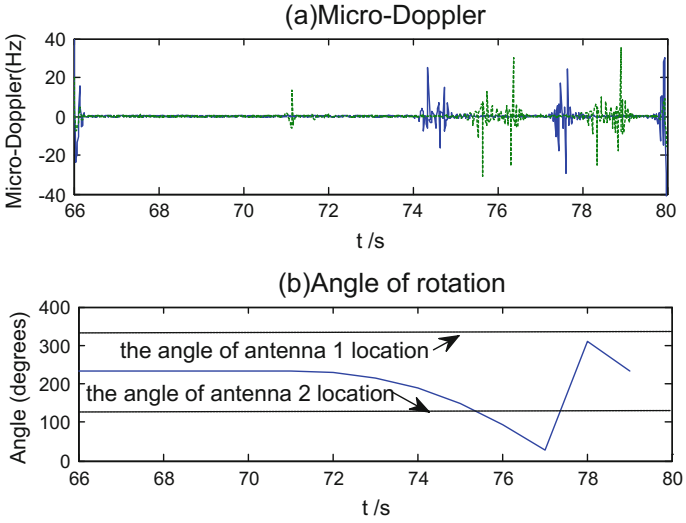


Fig. 7.8 **a** Micro-Doppler of two channel versus time; **b** the angle of rotation versus time

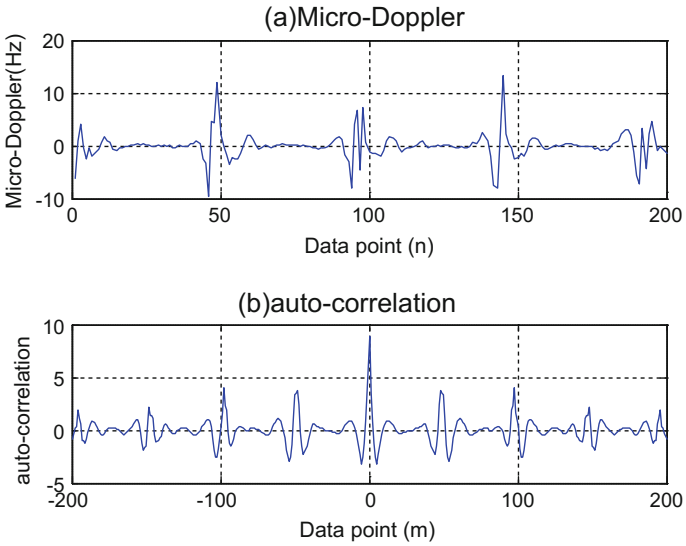


Fig. 7.9 **a** Micro-Doppler in data point **b** the auto-correlation of micro-Doppler in data point

there is twice signal interference in one period of rotation; So, the truly period is accord with a half of theoretical period of rotation which is 5 s in Sect. 7.2.3 as the speed of rotation is 0.4pi rad/s.

7.3.3 The Summary of Results

Transponder symmetric antennas will form the interference region when VMRs track rotating targets; the interference region is around in the middle of symmetric antennas with a range of $\pm 20^{\circ}$ where the amplitude and phase of radar received signal vary greatly.

The AGC of signal will reach its positive peak value when the radar antenna aimed at the antenna of the transponder; while reach its low peak value when the radar antenna aimed at the antenna interference of the transponder. The AGC value changed periodically which amplitude is in the range of 10–20 dB. The value of U_{Δ} rise and fall more obviously in the rotation period of time than other period of time and reach its peak value when in interference region.

A step of Doppler and micro-Doppler are generated by target rotation and interference region. The step of Doppler will reach its peak value when the radar antenna in interference region. The micro-Doppler will reach plus and minus two peaks before and after interference region. Micro-Doppler period is twice the target rotating period; Micro-Doppler amplitude is closely related to the radius and rotation speed of target and radar frequency, In addition, the bistatic micro-Doppler depend on the triangulation geometry of the transmitter, the target, and the receiver.

7.4 Conclusion

Focus on the issue of signal interference of the VMR induced by symmetric antennas of the rotating target, we analyzed the interference effect in theory based on the factors of AGC, Doppler, micro-Doppler and error voltage in this paper; then confirmed theoretical results with processing results of the data gathered by radar stations. The research conclusions are contributed to understand the effect of signal interference with the VMR and the characteristic of the signal induced by rotation and signal interference.

References

1. Jia XQ (2005) Continuous-wave radar data processing. National Defense Industry Press, Beijing
2. Qu WX, Yang WG, Zhang RY (2016) Micro-Doppler signatures analysis of spinning target with double-antenna. Telecommu Eng 56(2):212–217
3. Feng JG, Zhao H, Shi JF (2011) A responder antenna system design for spinning spacecraft. J Astronaut 32(10):2250–2253
4. Feng JG, Qian WP (2009) A rotating aerocraft transponder antenna analysis. The paper collection of 2009 of NCANT
5. Cui LL, Wang W, Ma TH (2015) Analysis and simulation of microstrip antenna mounted curved surface. J Proj Rocket Missiles Guid 35(2):175–178

6. Li J (2011) The development of the high-precision measurement transponder. University of Electronic Science and Technology of China, Chengdu
7. Zeng YX (2009) Analysis of the symmetric diploes antenna based on Matlab. J Quanzhou Norm Univ 27(2):30–33
8. Zhao YD (2011) The specifications of the transponder antenna feed system. China Aerospace Science and Technology Corporation 14th Research Institute
9. Li GJ, Huo DC (2012) Satellite communication link budget computation using STK. Space Electron Technol 1(1):68–72
10. Lan HZ (2012) Dynamic simulation of angle tracking signal. Telecommun Eng 52(7): 1116–1119
11. Chen VC (2011) The micro-Doppler effect in radar. Artech House, Norwood
12. Luo F, Guo ZH, Wang M et al (2013) Experimental researches on the threshold of airborne gravity data denoising based on DB wavelet transform. Geophys Geochem Explor. 37(8):1116–1119
13. Micbael W (2007) Digital signal processing using MATLAB and wavelets. Infinity Science Press, Hingham
14. Pan ZR, Dai R, Zhang N (2015) Improved algorithm for pitch detection. Comput Eng Appl 51(10):223–226

Chapter 8
Deep Learning for Mid-term Forecast of Daily Index of Solar 10.7 cm Radio Flux

Xin Wang

8.1 Introduction

For the low-earth orbit space objects, atmospheric density is the vital factor for the accuracy of orbit determination and prediction. In recent years, with the huge amount of data collected by the space exploration, the modelling of atmospheric density improved rapidly. However, these improvements do not bring about significant increase in the accuracy in the various applications of astrodynamics, such as the orbit prediction problems [1]. The reason is due to the physical mechanism of the atmospheric density, especially the coupling of magnetosphere and thermosphere is not clear yet. And on the other hand, the density of the atmosphere is strongly influenced by the space environment and the uncertainty in the prediction of space environment parameters have great impact on the forecast of density.

Solar 10.7 cm radio flux (F10.7) is a classical parameter for the solar activity proxy as well as a key index for the space environment forecast. As F10.7 index has a very long history of measurement and recording, nearly all empirical atmospheric density models use it to derive the variation of solar activity [2]. So the forecast of F10.7 index attracts many interests in the field space environment prediction especially in the atmospheric density prediction. The forecast of F10.7 has much difficulties because though it has strong relationship with solar cycle and rotation, but the full law is not complete known yet.

X. Wang (✉)
Purple Mountain Observatory, Chinese Academy of Sciences,
Nanjing 210008, China
e-mail: wangxin@pmo.ac.cn

X. Wang
Key Laboratory of Space Object and Debris Observation, PMO,
CAS, Nanjing 210008, China

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_8

Many studies are carried out at home and abroad. Each team developed their own method because of the different goals. In general, the forecast method can be divided into two types. One is to expand F10.7 index to a trigonometric series and another uses the auto-regression of time series analysis to express the F10.7 index as the linear combination of historical values. Autoregressive method is the mainstream in the practice of short- and mid-term forecast nowadays.

Both methods come down to the construction of the function between input values and output predictions, and the parameters of the function are obtained by fitting the historical data. No matter the method, the leading time, input values and fitting span should be determined. For trigonometric series method, time is the only required input. While for auto-regression, the input value is the previous values, so the number of the previous values which is called the order of auto-regression, should also be determined. For different purpose, these parameters differ from each other in references. Reference [3] uses trigonometric series method, the measurements of previous 135 days are fitted to generate the prediction of following 54 days. Reference [4] uses 54 order autoregressive model to give a 27-day prediction, while the fitting span is 750 days. Reference [5] uses 27 order to predict next 3 days and parameters are solved with previous 135 days.

Machine Learning is a rising star in data processing, especially for big data and high dimensional data, and is also more and more widely used in the space environment predictions [6, 7].

There are several studies in the forecast of F10.7 index utilizing machine learning. Reference [8] uses artificial neural network in the prediction of monthly mean F10.7 index and input values considered the previous values as well as the statistical information of last 11-year solar cycle. Reference [5] solves the autoregressive problem with Support Vector Regression (SVR). And Ref. [9] trained a network with one hidden layer of only 2 neutrons with the data of a full solar cycle and realized a static model for 1-day-ahead prediction with previous 11 days.

It is worth noting that, in addition to Ref. [9], all the models are updated in a rolling way, model parameters have to be re-estimated with new measurement, the parameters are not fixed and only effective for the forecast at that time. As we known, the estimation of parameters is much more complicated than the prediction itself, these characteristic causes inconvenient in application.

The autoregressive model is adopted in this study. According the computing ability, a forecast method with deep learning is explored. Following the same idea as Ref. [9], the model is completely fixed after training, and with the new measured value, forecast can be directly done. Beneficial to today's computing power, this paper uses much more number of input units and hidden units, and the data of over 50 years are used for a thorough test of the proposed method.

8.2 Methodology

8.2.1 Model of Forecast

Autoregressive model is the mainstream method in the forecast of F10.7 index, which expresses the forecast value as a linear combination of previous values:

$$y_{t-1+a} = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots + \alpha_p y_{t-p} \tag{8.1}$$

where y_t is the F10.7 value of day t, a is the leading time of the forecast, p is the order of auto-regression. For stationary time series, α_i are constant. While the series of F10.7 index is not stationary, so the previous N ($N > p$) values are used to solve α_i in practice, which cause the estimated α_i is only effective for the next day prediction.

Linear model has many advantages, but it is only a simple approximate to the real model. Considering the complex of solar activity, linear model is replaced by a non-linear model, the output prediction is as:

$$y_{t-1+a} = f(y_{t-1}, y_{t-2}, \ldots, y_{t-p}) \tag{8.2}$$

The model is represented by the function f. Deep learning is very suitable for this kind of problem. The function can be learned from the historical data non-parametrically, in other words without defining certain expression of the function. Then the problem can be solved in minimum assumptions.

As deep learning has a very strong ability to express functional relationship, auto-regressive and mean average model is employed, both the previous values and previous prediction errors are taken as the input, as:

$$y_{t-1+a} = f(y_{t-1}, \ldots, y_{t-p}, \varepsilon_{t-1}, \ldots \varepsilon_{t-p}) \tag{8.3}$$

where ε_t is the a-day-ahead prediction error of the day t. As the model is not update in a rolling way, taking account of previous prediction errors will give auto-repair ability to the model.

The forecast of 27 days is carried out in this study, which meets the requirement of most of the applications and the result is easy to be compared with references.

8.2.2 Model of Learning

Multi-layer perception (MLP) is one of most classical models in deep learning. It was proposed very earlier, but because of the lack of effective training algorithm, it was not widely used for a long time. In recent years, some new training algorithm is invented and this model is widely used in many fields in a very short time [10].

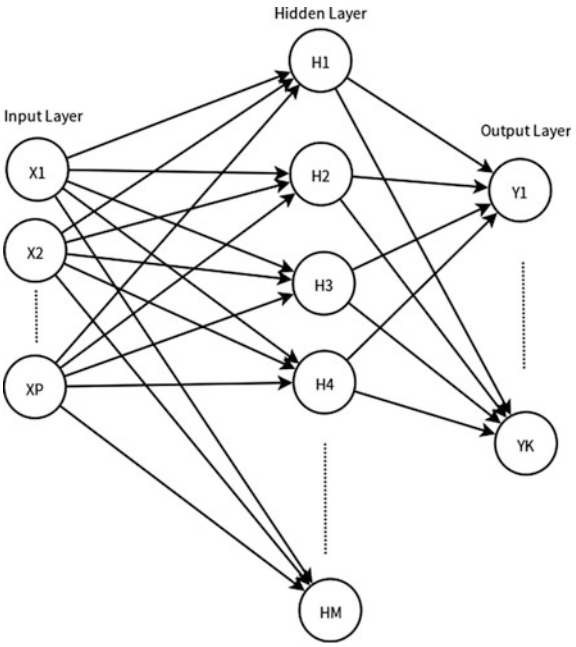


Fig. 8.1 Structure of a neural network

MLP simulates the human brain and the structure is shown in Fig. 8.1. Every circle in the figure stands for a neutron or unit.

A classical 3-layer network is employed in this study, including one input layer, one hidden layer and one output layer. According to universal approximation theorem, a network with one hidden layer can express any continuous function defined in a bounded closed interval with finite number of hidden units. Input layer stands for the input values of the prediction, and output layer is the prediction result. Hidden unit h_i is calculated with Eq. (8.4).

$$h_i = \Phi\left(\sum_{j=1}^{M}\left(w_{ij}^{(1)}x_j\right) + w_{i0}^{(1)}\right), \quad i = 1, 2, \ldots M \tag{8.4}$$

And output unit is computed with Eq. (8.5).

$$Y_i = \Phi'\left(\sum_{j=1}^{M}\left(w_{ij}^{(2)}h_j\right) + w_{i0}^{(2)}\right), \quad i = 1, 2, \ldots K \tag{8.5}$$

where P, M, K are the total number of units for each layer. For the problem discussed here, we have $P = 2p$, $K = a$ and

$$X = (y_{t-1}, \ldots, y_{t-p}, \varepsilon_{t-1,} \ldots \varepsilon_{t-p,}) \tag{8.6}$$

$$Y = (y_t, y_{t+1}, \ldots, y_{t-1+a}) \tag{8.7}$$

In the following numerical experiment, M takes the value of 90. Φ and Φ' are activation functions, which implemented the non-linear representation of the model. Rectified linear units function [11] are chosen for both Φ and Φ'.

$$\Phi(x) = \Phi'(x) = \max(0, x) \tag{8.8}$$

The training is to estimate the parameters w_{ij}. As the prediction with different leading time are independent to each other, the problem of Eqs. (8.6) and (8.7) can be reduced to a one-dimension output sub-problems separately, as $Y = (y_{t-1+i})$, $i = 1,2,\ldots,a$ and $K = 1$.

8.2.3 Evaluation Criteria

To evaluate the performance of the prediction with proposed model, the same criteria defined in Ref. [4] is chosen. The criterion of mean relative error (MRE) is the mean of the daily absolute relative error in the 27-day prediction timespan, defined as Eq. (8.9).

$$E_f = \frac{1}{27} \sum_{t=1}^{27} \frac{|F_t - O_t|}{O_t} \tag{8.9}$$

where F is the forecast value and O is the real observation value. Another criterion is the MRE of different leading time, which is the mean absolute relative error of predictions with same leading time as Eq. (8.10).

$$E_d(a) = \frac{1}{N} \sum_{t=1}^{N} \frac{\left|F_t^a - O_t\right|}{O_t} \tag{8.10}$$

where F_t^a is the predict value with leading time of a day.

8.3 Experiment

From the internet [12], the real measurements of F10.7 index are collected, and Fig. 8.2 gives the full dataset of historical daily F10.7 values since Oct. 1 1957.

The full dataset is divided into 3 sets. The full 55-year data between Jan. 1 1959 and Dec. 31 2013 is as the training dataset. The data before this interval is tagged as

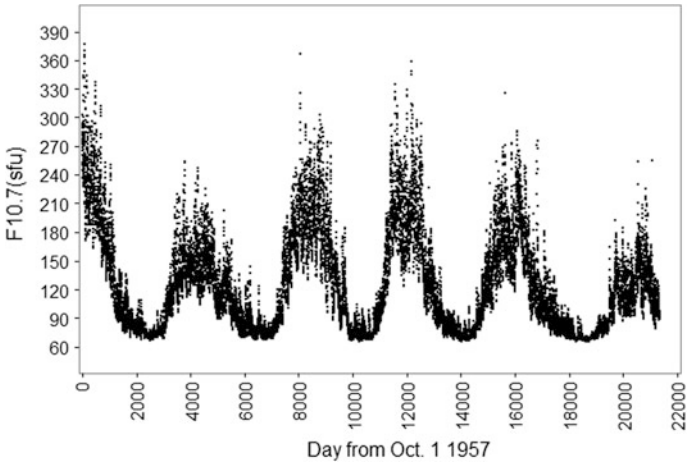


Fig. 8.2 Historical daily F10.7 index since Oct 1, 1957

test dataset 1 and the data after is tagged as test dataset 2. From Fig. 8.2, both of two test datasets are in solar active period and test dataset 1 is located in the solar maximum. So the test datasets can ensure the validation of the model.

Training and prediction computer programs are written in Python language with Keras packages. Keras encapsulates most of the modern models and algorithm for deep neural networks and enables fast experimentation after defining the structure of the network [13].

The networks for leading time of 1–27 days are trained with full training dataset separately. After the training, the networks are fixed. The prediction is done with corresponding fixed networks. Figure 8.3 gives the MRE for each datasets with each leading time. As the training dataset covers 5 full 11-year solar cycle, the model is not easy to be over-fitted. Though the result of training dataset is better than test datasets and test dataset 2 has a better performance, the difference is still in a reasonable range. While considering the difference of solar flux level for the datasets, the training and validation can be regarded in the same error level. Figure 8.3 also clearly shows the same phenomenon as mentioned in some references. With the increase of leading time, the accuracy decreases and the error is saturated after 9 days and the accuracy is much more stationary during solar quiet period.

Figure 8.4 gives the yearly MRE of 27-day prediction since 1958. As year 1958, 2014 and 2015 are fully in test datasets, the result of these years is not divergence and proves that the trained model is not over-fitted. Compared with the result of overlap years in Refs. [3, 4], the performance of proposed method is better, especially during solar maximum.

Table 8.1 shows the yearly MRE of short-term prediction of 1- to 3-day ahead from year 2003 to 2006. The SVR in the table indicates than the results are from

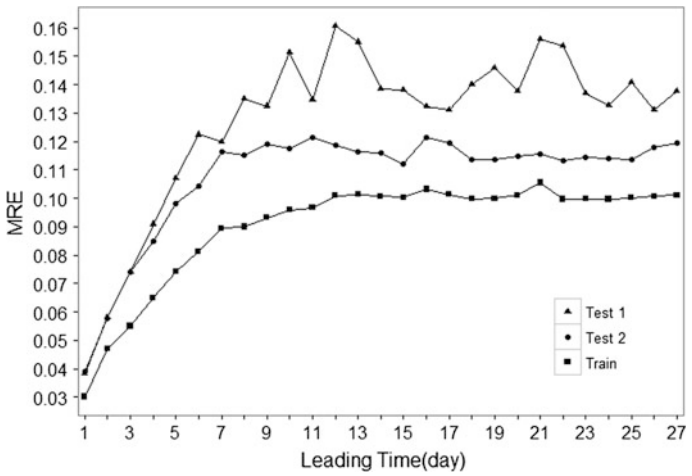


Fig. 8.3 Mean relative error of different leading time

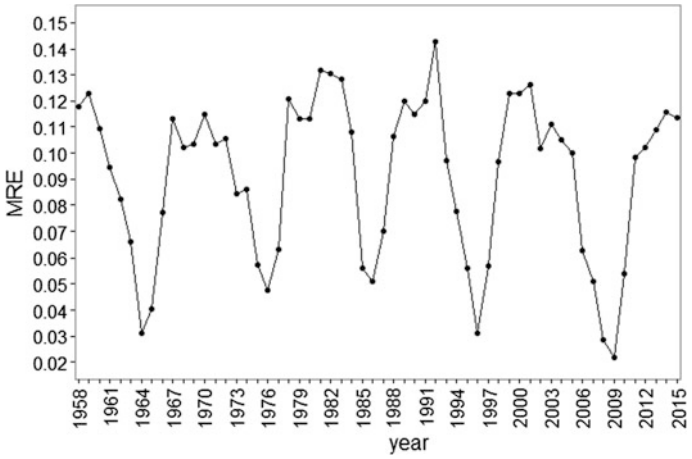


Fig. 8.4 Yearly mean relative error of 27-day prediction

Table 8.1 Yearly mean relative error of short-term prediction

Year	1-day (%)		2-day (%)		3-day (%)	
	SVR	DL	SVR	DL	SVR	DL
2003	5.56	3.46	7.62	5.30	9.32	6.84
2004	3.69	3.43	4.95	5.45	7.28	6.90
2005	4.58	3.41	5.99	5.08	6.83	6.48
2006	2.71	2.46	3.60	3.61	4.48	3.87

reference [5] and obtained with SVR method. DL means the results of this paper with deep learning. Obviously DL results are better in nearly all cases.

In summary, proposed method has a very significant improvement in short- and mid-term forecast of daily F10.7 index. Fixed parameters can represent time series as long as over 55 years, which shows the advantage of deep learning method for similar problems.

8.4 Conclusion and Discussion

With deep learning, the assumptions are not required for the relationship between the input and output value of the prediction. The neural network has a very strong ability in describing the complex solar activity. According to the experiment, the model has significant improvement with convenient implementation with current computer power. Though the parameters are much more than other methods, but the value are fixed, the model is easier to be embedded in other applications.

Research shows than other solar activity proxies, such as the number and area of sun spots have positive sense in the forecast of F10.7 [14]. With deep learning, these values can be directly integrated in the model. The underlying relationship can be learned from the data automatically. The fusion is much easier than with the traditional methods. Further study may be carried out.

Acknowledgements This work is supported by the National Natural Science Foundation of China (11573074)

References

1. Tobiska K (2003) Forecast E10.7 for improved low-earth-orbit satellite operations. J Spacecr Rockets 40(3):405–410
2. Bowman B, Tobiska K, Marcos F, et al (2008) A new empirical thermospheric density model JB2008 using new solar and geomagnetic indices. AIAA/AAS astrodynamics specialist conference, Honolulu, Hawaii, Aug 18–21
3. Wang HB, Xiong JN, Zhao CY (2014) Mid-term forecast method of solar radiation index F10.7. Acta Astron Sin 55(4):302–312
4. Liu S, Zhong QZ, Wen Q et al (2010) Model research of 10.7 cm solar radio flux 27-day forecast(I). Chin J Space Sci 30(1):1–8
5. Huang C, Liu DD, Wang JS (2009) Forecast daily indices of solar activity F10.7 using support vector regression method. Res Astron Astrophys 9(6):694–702
6. Perez D, Wohlberg B, Lovell T et al (2014) Orbit-centered atmospheric density prediction using artificial neural networks. Acta Astronaut 98:9–23
7. Choury A, Bruinsma S, Schaeffer P (2013) Neural networks to predict exosphere temperature corrections. Space Weather 11:592–602
8. Zen W, Huang ZR (1998) Forecast of monthly mean index of solar 10.7 cm radio flux with artificial neural network. In: Proceedings of 14th annual conference of chinese geophysical society, Hangzhou, p. 233

9. Chatterjee TN (2001) On the application of information theory to the optimum state-space reconstruction of the short-term solar radio flux (10.7 cm) and its prediction via a neural network. Mon Not R Astron Soc 323:101–108
10. Bishop C (2007) Pattern recognition and machine learning. Springer, New York, pp 243–308
11. Glorot X, Bordes A, Bengio Y (2010) Deep Sparse Rectifier Neural Networks. J Mach Learn Res 15:315–323
12. Kelso TS (2016) EOP and space weather data. http://www.celestrak.com/SpaceData/sw19571001.txt
13. Chollet F (2016) Keras. https://github.com/fchollet/keras
14. Wen J, Zhong QZ, Liu SQ (2010) Model research of 10.7 cm solar radio flux 27-day forecast. Chin J Space Sci 30(3):198–204

Chapter 9
The Analysis of the Semi-major Axis Changed by the North–South Control of the Beidou GEO Satellites

Quanjun Li, Rui Xue and Dalin Kong

9.1 Introduction

The BeiDou navigation constellation is composed of GEO satellites, IGSO satellites and MEO satellites. And the GEO satellites play an important role in the navigation service [1–6]. In the process of telemetry, track and command (TT&C), the orbit control will make the GEO satellites cannot provide navigation services. Due to the impact of the solar and lunar gravitational perturbations, the north–south control of the GEO satellites is required [7, 8]. In the north–south control of the GEO satellites, the influence of gyro drift, attitude control, orbital plane changes and other reasons changes the semi-major axis [9, 10]. It is hard to accurately estimate the effect of north–south control on the semi-major axis in practice.

It is often need to control semi-major after orbit determination of the north–south control. The extra control affects the continuity of the GEO satellites navigation services.

9.2 The North–South Control Mode of the BeiDou GEO Satellites

The GEO satellites position holding mode is used for attitude and orbit control. There are four types of the position maintenance mode. The differences between these four types are the selection of the gyro angular velocity and the yaw reference, as shown in Table 9.1. The time of the sun in the solar sensor field of view is related to the location of the GEO satellites. And the time is fixed. The time of the

Q. Li (✉) · R. Xue · D. Kong
Xi'an Satellite Control Center, Xi'an 710043, China
e-mail: 13572965391@163.com

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_9

Table 9.1 Definition of the position maintenance mode

Position maintenance mode	Definition
First	No angular velocity
	Measure the yaw angle by solar sensor
Second	no angular velocity
	Measure the yaw angle by gyro
Third	Measure angular velocity by gyro
	Measure the yaw angle by solar sensor
Fourth	Measure angular velocity by gyro
	Measure the yaw angle by gyro

north–south control is mainly related to the initial orbit and the target orbit. The sun may not be in the solar sensor field of view at the time of the nortvouth control. Therefore the fourth position maintenance mode is generally used in the north–south control. In this mode, the three axis angular velocity and the yaw angle are measured by the gyro and the attitude is controlled by the thrusters.

9.3 The Analysis of the Influence of the North–South Control on the Semi-major Axis

According to position maintenance mode of north–south control, the model of various influencing factors on the semi-major axis is established by stripping the factors, such as gyro drift, attitude control and orbital plane change.

9.3.1 The Influence of Gyro Drift on the Semi-major Axis

Gyro drift is the change of the angle with respect to the reference axis of the gyro in the unit time. It is an important indicator of the gyro performance. In practical application, due to the influence of the disturbance torque, the electronic circuit noise, the frequency monitoring error and other factors, the gyro drift is inevitable [11, 12]. With the increase of time, the attitude error caused by drift will gradually increase. The north–south control time of GEO satellite is longer. The long time gyro drift causes the deviation between the measured yaw and the actual yaw. The deviation makes the normal direction of the velocity increment generate the component in the tangential direction. It makes the semi major axis change.

During the north–south control, the yaw angle will be controlled in the vicinity of zero. And the yaw angle measured by the gyro is approximate to zero. If $\dot{\psi}$ refers to the rate of the gyro drift and t_0 refers to the north–south control ignition time,

then the change of the yaw angle caused by the gyro drift can be presented as follow:

$$\psi(t) = -\dot{\psi} \times (t - t_0)$$

The pulse ignition mode is used in the north–south control. In order to facilitate the establishment of the model, a simplified model of the equivalent thrust is adopted. The equivalent thrust is as follow:

$$\bar{F} = Ft_{\text{on}}/t_{\text{p}}$$

F is the actual thrust, t_{on} is the ignition pulse width, t_{p} is the ignition cycle.

The velocity increment in tangential direction component caused by the gyro drift is as shown in formula (9.1).

$$\Delta v_{\text{T1}} = -\frac{1}{m}\int_{t_0}^{t} k_{\text{SN}}\bar{F}\sin(\psi(t))\text{d}t \tag{9.1}$$

m is the mass of the satellite, k_{SN} is the calibration coefficient of the thruster used in the north–south control.

Make $t_0 = 0$, formula (9.1) simplifies to formula (9.2).

$$\Delta v_{\text{T1}} = \frac{1}{m}\frac{k_{\text{SN}}\bar{F}}{\dot{\psi}}(1 - \cos(\dot{\psi}\Delta t)) \tag{9.2}$$

Δt is the control time length.

During the north–south control, $\dot{\psi}\Delta t$ can be considered as a small amount. Formula (9.2) can be changed as follow:

$$\Delta v_{\text{T1}} = \frac{k_{\text{SN}}\bar{F}\dot{\psi}\Delta t^2}{2m} = \frac{k_{\text{SN}}F\dot{\psi}\Delta t^2 t_{\text{on}}}{2mt_{\text{p}}}$$

The semi-major axis variation caused by gyro drift is as follow:

$$\Delta a_1 = \Delta v_{\text{T1}} \times 27.4 = \frac{27.4k_{\text{SN}}F\dot{\psi}\Delta t^2 t_{\text{on}}}{2mt_{\text{p}}}$$

9.3.2 The Influence of Attitude Control on the Semi-major Axis

Attitude control during the north–south control is controlled by the switch modulation. The velocity increment of the tangential direction is mainly caused by the thrust of the tangential direction while controlling the attitude. Without considering the change of the attitude, the velocity increment of the tangential direction is as follow:

$$\Delta v_{T2} = \sum_{i=1}^{N} k_i F \Delta t_i \cos \alpha_i / m$$

n is the number of thrusters for attitude control, k_i is the calibration coefficient, Δt_i is the working time of the thruster for attitude control, α_i is the angle between the thrust direction and the tangential direction.

The semi-major axis variation caused by attitude control is as follow:

$$\Delta a_2 = \Delta v_{T2} \times 27.4 = 27.4F \sum_{i=1}^{N} k_i \Delta t_i \cos \alpha_i / m$$

9.3.3 The Influence of the Orbital Plane Change on the Semi-major Axis

In the fourth position maintenance mode, the yaw angle is measured by the gyro. The yaw angle is relative to the inertial space. The actual yaw angle of the satellite in the north–south control is based on the orbital coordinate system. During the control period, the orbital coordinate system changes with the change of the orbital plane. And it causes the change of the actual yaw angle. It causes the thrust of the north–south control to generate the component in tangential direction. And the semi-major axis has been changed. The thrust direction is always the same in the inertial space, and is perpendicular to the initial orbital plane. With the change of the orbital plane during the control, the thrust direction is no longer perpendicular to the instantaneous orbital plane, as shown in Fig. 5.1. The tangential component is generated. The magnitude of the tangential component is related to the angle between the instantaneous orbital plane and the initial orbital plane (Fig. 9.1).

The angle between the target orbital plane and the initial orbital plane is as follow:

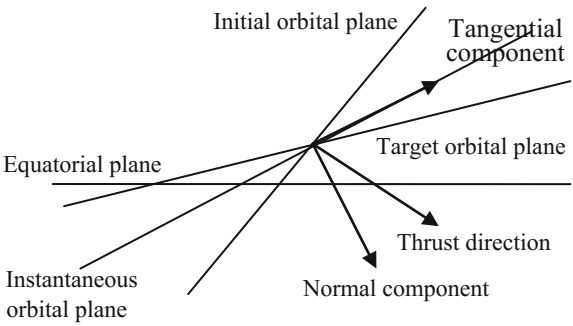


Fig. 9.1 The component of the thrust direction during the north–south control

$$\Delta I = \arccos(-\cos(180 - i_0)\cos i_m + \sin(180 - i_0)\sin i_m \cos(\Omega_0 - \Omega_m))$$

i_0 is the initial inclination, i_m is the target inclination, Ω_0 is the initial ascending node, Ω_m is the target ascending node.

The equivalent thrust is still used to simplify the model. The velocity increment in tangential direction component caused by the change of the orbital plane is as shown in formula (9.3).

$$\Delta v_{T3} = -\frac{1}{m}\int_{t_0}^{t} k_{SN}\bar{F}\sin\Delta I_t dt \tag{9.3}$$

ΔI_t is the angle between the tar instantaneous orbital plane and the initial orbital plane.

Make $t_0 = 0$, formula (9.3) can be changed as follow:

$$\Delta v_{T3} = \frac{k_{SN}\bar{F}\Delta I\Delta t}{2m} = \frac{k_{SN}F\Delta I\Delta t t_{on}}{2mt_p}$$

The semi-major axis variation caused by the orbital plane change is as follow:

$$\Delta a_3 = \Delta v_{T3} \times 27.4 = \frac{27.4k_{SN}F\Delta I\Delta t t_{on}}{2mt_p}$$

9.3.4 The Influence of Other Factors on the Semi-major Axis

In addition to the above reasons, other factors also have an impact on the semi-major axis, such as the installation error of the thrusters, the satellite gravity variation and the performance changes of the thrusters. These factors are difficult to establish models. These factors can be analyzed by stripping the influencing factors. In the previous north–south control of the GEO satellites, the effect of gyro drift, attitude control and orbital plane change on the change of the semi-major axis is eliminated. The remaining change of the semi-major axis is considered to be caused by other factors. The change rate of the semi-major axis caused by other factors in the north–south control is as follow:

$$\Delta\dot{\bar{a}}_i = \Delta\bar{a}_i / \Delta T_i$$

$\Delta\bar{a}_i$ is the actual change of the semi-major axis caused by other factors, ΔT_i is the control time.

The influence factor of other factors on the semi-major axis is as follow:

$$\Delta\dot{\bar{a}} = \frac{1}{n}\sum_{i=1}^{n}\Delta\dot{\bar{a}}_i$$

n is the number of the north–south control.

The semi-major axis variation caused by other factors is as follow:

$$\Delta\bar{a} = \Delta\dot{\bar{a}} \times \Delta t$$

9.4 The Practical Verification

The north–south control situation of three GEO satellites in orbit is as shown in Table 9.2. By using the above model, the theoretical value of the semi-major axis variation caused by various factors affecting is calculated. The parameters of the satellite in the calculation are as follows:

$$F = 10$$
$$t_{\text{on}} = 2.56$$
$$t_{\text{P}} = 5.12$$
$$N = 4$$
$$\cos\alpha_1 = \cos\alpha_2 = 0.96595$$
$$\cos\alpha_3 = \cos\alpha_4 = 0.76552$$
$$k_1 = k_2 = k_3 = k_4 = 0.95$$
$$k_{\text{SN}} = 0.8056$$

Table 9.2 The north–south control situation of GEO satellites in orbit

The satellite	The time of control	The mass of satellite (kg)	The time of ignition (min)
GEO-1	First	1709	83.02933
GEO-1	Second	1709	77.60667
GEO-1	Third	1670	91.22133
GEO-3	First	1651	86.69867
GEO-3	Second	1690	89.79333
GEO-3	Third	1662	97.36533
GEO-4	First	1731	105.0133
GEO-4	Second	1703	51.456

Table 9.3 The coupling of the semi-major axis caused by the gyro drift

The satellite	The time of control	the gyro drift (degree/min)	The coupling (km)
GEO-1	First	−0.0002	−0.09324
GEO-1	Second	0.0007169	0.292009
GEO-1	Third	0.0005273	0.303668
GEO-3	First	−0.00106	−0.55775
GEO-3	Second	−0.001027	−0.56633
GEO-3	Third	−0.001547	−1.0196
GEO-4	First	0.0003333	0.245429
GEO-4	Second	0.0011	0.197654

9.4.1 The Coupling of the Semi-major Axis Caused by Gyro Drift

Gyro drift is calculated according to the change trend of the gyro integration after the final calibration of the north–south control. The coupling of the semi-major axis caused by gyro drift is less than 1 km, as shown in Table 9.3. Its magnitude is mainly related to the effect of gyro calibration.

9.4.2 The Coupling of the Semi-major Axis Caused by Attitude Control

The working time of the thrusters is the actual telemetry parameters of the satellite. Attitude control has a great influence on the coupling of the semi-major axis, as shown in Table 9.4. The impact is not the same each time.

Table 9.4 The coupling of the semi-major axis caused by attitude control

The satellite	The time of control	The working time of the thrusters (s)				The coupling (km)
		1	2	3	4	
GEO-1	First	38.30	0.06	0.19	6.91	4.775
GEO-1	Second	17.00	0.01	3.00	1.72	1.925
GEO-1	Third	0.83	53.24	6.84	0.57	7.188
GEO-3	First	0.095	69.66	0.06	8.74	9.472
GEO-3	Second	36.27	0.06	0.22	8.92	4.315
GEO-3	Third	3.04	6.46	6.08	1.28	0.545
GEO-4	First	1.82	3.77	0.06	11.45	−0.513
GEO-4	Second	0.76	11.73	2.01	1.82	1.396

Table 9.5 The coupling of the semi-major axis caused by the orbital plane change

The satellite	The time of control	The initial inclination (degree)	The initial ascending node (degree)	The target inclination (degree)	The target ascending node (degree)	The coupling (km)
GEO-1	First	1.53	8.48	1.54	351	2.626
GEO-1	Second	1.67	8.63	1.6	353	2.356
GEO-1	Third	1.71	11.77	1.55	355	3.162
GEO-3	First	1.67	12.64	1.55	356	2.931
GEO-3	Second	1.76	35.70	1.4	23	3.072
GEO-3	Third	1.67	33.98	1.32	19	3.531
GEO-4	First	1.23	52.89	0.7	40	4.020
GEO-4	Second	0.92	49.90	0.7	39	0.954

9.4.3 The Coupling of the Semi-major Axis Caused by the Orbital Plane Change

The coupling of the semi-major axis caused by the orbital plane change is as shown in Table 9.5. Because of the long control period of the GEO satellites, the change of the orbital plane is larger, and the coupling of the semi-major axis is larger.

9.4.4 The Coupling of the Semi-major Axis Caused by Other Factors

The coupling of the semi-major axis caused by other factors is as shown in Table 9.6.

Table 9.6 The coupling of the semi-major axis caused by other factors

The satellite	The time of control	Actual coupling (km)	The coupling (km)	The impact factor (km/min)
GEO-1	First	8.6	1.203	0.014487
GEO-1	Second	6	1.124	
GEO-1	Third	11	1.322	
GEO-3	First	13.6	1.256	0.050189
GEO-3	Second	11	4.507	
GEO-3	Third	8.3	4.887	
GEO-4	First	6.9	3.055	0.029088
GEO-4	Second	4	1.497	

Table 9.7 The comparison of estimated value and actual value

The satellite	The time of control	The actual value (km)	The estimated value (km)	The estimation error (km)
GEO-1	First	8.6	8.511	−0.089
GEO-1	Second	6	5.699	−0.301
GEO-1	Third	11	11.976	0.976
GEO-3	First	13.6	13.102	−0.498
GEO-3	Second	11	11.328	0.328
GEO-3	Third	8.3	7.944	−0.356
GEO-4	First	6.9	6.807	−0.093
GEO-4	Second	4	4.046	0.046

9.4.5 Brief Summary

According to the above analysis, the comparison of estimated value and actual value is as shown in Table 9.7. The estimation error of the model is less than 1 km. The estimation error of GEO-1 satellite is less than 500 m except the third time. The estimation error of GEO-3 satellite is less than 400 m. The estimation error of GEO-3 satellite is less than 100 m.

9.5 Conclusions

This paper analyses the influence factors of the north–south control on the semi-major axis. The estimation model of each factor is established. The model is verified by the actual data in orbit. And the results show that the accuracy of the semi-major axis coupling is less than 1 km. The model can be used to compensate the semi-major axis in the north–south control of the BeiDou GEO satellites. The

estimated accuracy of different GEO satellites is not consistent, mainly because the model does not consider the error caused by the semi-major axis control after the north–south control. Other GEO satellites can learn from this method. Due to the difference of the satellite platform, it is needed to re-establish the model.

References

1. Meng B, Yi CHJ, Han CH (2009) Navigation constellation optimization design based on multi objective particle swarm optimization algorithm. J Aviat 30(7):1284–1291
2. Shuai P, Qu GJ, Chen ZHG (2006) Analysis and design of navigation area constellation. J Space Sci 26(4):268–276
3. Du L, Zhang Y, Long B et al (2013) Local optimal design of navigation constellation based on genetic algorithm. J. Navig Position 1(1):34–39
4. Yang YX, Li JL, Wang AB et al (2014) Preliminary assessment of the navigation and positioning performance of BeiDou regional navigation satellite system. Sci China 57:144–152
5. Li HN, Qian SH, Gao WG et al (2014) The constellation stability promotion approaches for compass (beidou) regional navigation and positioning system. Sci Sin-Phys Mech Astron 44 (6):621–629
6. Zhai W, Zhang GZH, Yong SHW (2011) Performance analysis of service volume of COMPASS navigation satellite system based on basic constellation. Gnss World China 36 (4):56–60
7. Peng CHL (2009) Methods research on the orbit control of navigation satellite. National University of Defense Technology, Changsha
8. Huang XT, Zhang WY, He WP (2012) Study of south/north maneuver strategies of geostationary satellite. Aerosp Shanghai 29(5):41–45
9. Wang XJ, Li DL, Qiang LK et al (2005) Orbit-trimming after inclination control. J Spacecr TT&C Technol 24(2):25–27
10. Xi LH, Liu Y, Cui Y (2016) Selection of attitude control strategy for precise orbit spacecraft. J Spacecr TT&C Technol 35(1):48–52 (in Chinese)
11. SHCH Tu (1988) Attitude dynamics and control of satellite orbit (2). China Aerospace Press, Beijing
12. Liu L, Liu Y, Cao JF et al (2014) A realtime on-orbit calibration method for integrated attitude determination by gyros/star sensor. J Spacecr TT&C Technol 33(2):152–157

Chapter 10
Autonomous Orbit Determination of Satellites Around Triangular Libration Points in the Earth–Moon System

Bin Liu, Xiyun Hou, Jingshi Tang and Lin Liu

10.1 Introduction

The Nowadays, due to support from many navigation satellite systems around the Earth (GPS, GLANASS, BeiDou, etc.), performance of satellite navigation close to the Earth has been greatly improved. However, due to limitations of existing navigation satellite constellations, they can not provide powerful support for lunar satellites or deep space probes which are far away from the Earth. An interesting idea is to utilize the collinear libration points (CLPs) of the earth–moon system [1]. Due to uniqueness of these points, AOD between the navigation satellite and the lunar satellites is feasible by utilizing the inter-satellite range data. Recently, this idea is generalized to the distant retrograde orbits around the Moon [2] and the TLPs of the earth–moon system [3]. For the TLPs in the earth–moon system, due to the Sun's perturbation, motions close to them are unstable [4], but the instability is much weaker than that of the collinear libration point [5], so they can be used to locate navigation satellites [6]. Moreover, even with the Sun's perturbation, large stable orbits exist around the TLPs [7]. These orbits are ideal places to locate navigation satellites because theoretically no station-keeping is needed.

Due to unique positions and dynamical properties of the CLPs and TLPs, AOD between the satellites around them and the lunar satellites is possible [1, 3, 8, 9–11]. However, these studies are limited to the simplified circular restricted three-body

B. Liu (✉) · X. Hou · J. Tang · L. Liu
School of Astronomy and Space Science, Nanjing University, Nanjing 210093, China
e-mail: liubin_astronomy@163.cn

B. Liu · X. Hou · J. Tang
Key Laboratory of Modern Astronomy and Astrophysics,
Nanjing University, Nanjing 210093, China

B. Liu · X. Hou · J. Tang · L. Liu
Institute of Astrodynamics and Space Environment, Nanjing University,
Nanjing 210093, China

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_10

problem (CRTBP). The current work generalizes the study to the real earth–moon system where the Sun's perturbation is considered and the orbits of the Moon and the Sun are given by the JPL numerical ephemeris DE 405. In this work, our concern is not the AOD between the lunar satellites and the navigation satellites around the TLPs, but the AOD between the navigation satellites around the TLPs. The main idea to drive us to carry out this study is plain: if the navigation satellites move on stable orbits around the TLPs and can autonomously determine their orbits by utilizing the inter-satellite range data, then the navigation constellation can autonomously execute their missions.

The paper is organized as follows. Section 10.2 is devoted to the force model description on the real earth–moon system. Section 10.3 introduces the DSs around the TLPs and their nearby orbits. Section 10.4 deals with trivial issues of the AOD process in this paper. Section 10.5 shows the numerical results and analysis. Two cases are studied. Section 10.6 concludes the study.

10.2 The Force Model

In our work, two force models will be used. The first model is the Sun–Earth–Moon system, and the second model is the full force model which also incorporates the other seven planets and Pulto in the solar system. The first model is used to obtain the dynamical substitutes, the second model is used as the force model during the AOD of the dynamical substitutes (Fig. 10.1).

10.2.1 The Sun–Earth–Moon System

The motion of a spacecraft, P, of mass m, under the gravitational attraction of three primaries, Earth of mass m_{Earth}, Moon of mass m_{Moon}, Sun of mass m_{Sun}, is considered, and the actual distance satisfy $R, R_2, R_M \ll R_S, \Delta_S, \Delta$.

The orbits of the small body in the Earth-centered inertial fame along with the Sun and the Moon are integrated.

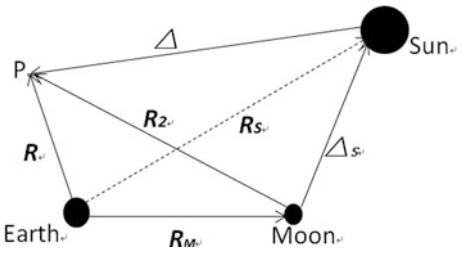


Fig. 10.1 The relative geometry between the Earth, the Moon, the Sun and the small body P in the Earth-centered inertial frame

Equations of motion (EOM) for the small body can be written as

$$\ddot{\mathbf{R}} = -\frac{\mu_E \mathbf{R}}{R^3} - \mu_M \left(\frac{\mathbf{R}_2}{R_2^3} + \frac{\mathbf{R}_M}{R_M^3} \right) - \mu_S \left(\frac{\mathbf{\Delta}}{\Delta^3} + \frac{\mathbf{R}_S}{R_S^3} \right) \tag{10.1}$$

And EOMs for the Moon and the Sun are

$$\ddot{\vec{\mathbf{R}}}_{\mathbf{S}} = -\frac{(\mu_E + \mu_S)\mathbf{R}_S}{R_S^3} - \mu_M \left(\frac{\mathbf{\Delta_S}}{\Delta_S^3} + \frac{\mathbf{R}_M}{R_M^3} \right) \tag{10.2}$$

$$\ddot{\mathbf{R}}_{\mathbf{M}} = -\frac{(\mu_E + \mu_M)\mathbf{R}_M}{R_M^3} + \mu_S \left(\frac{\mathbf{\Delta_S}}{\Delta_S^3} - \frac{\mathbf{R}_S}{R_S^3} \right) \tag{10.3}$$

where $\mu_E = 1$ is the mass parameter of the central body. The model here is used to obtain the dynamical substitutes.

10.2.2 The Full Force Model

The force model here is used when we study the problem of the AOD of the dynamical substitutes around the TLP in the Earth-centered inertial frame corresponding to the epoch J2000.0.

Where M_i is the mass of the third bodies in the solar system and $\mu_i = M_i/M_{Earth}$, $i = 1 \ldots 11, i \neq 3$ is the gravitational parameter of the third body, here, the third bodies considered include the Moon, the Sun, Pluto and the seven planets except the Earth.

And Equations of motion (EOM) for the small body can be written as

$$\mathbf{R} = -\frac{\mu_E \mathbf{R}}{R^3} - \sum_{i=1, i\neq 3}^{11} \mu_i \left(\frac{\mathbf{\Delta}_i}{\Delta_i^3} + \frac{\mathbf{R}_i}{R_i^3} \right) \tag{10.4}$$

where $\mathbf{R}, \mathbf{R_i}$ is the position vector of the small body and the third body respectively,$\mathbf{\Delta}_i$ is the position vector of the small body with respect to the third body.

$$\mathbf{\Delta_i} = \mathbf{R} - \mathbf{R_i} \tag{10.5}$$

Equation (10.1)–(10.4) is formulated with dimensionless time and space units, that is, the total mass of primaries, their distance and their angular velocity are taken as unity. With the new dimensionless units, the value of the gravitational constant G is 1.0.

10.3 Two Special Kinds of Orbits Around the TLPS

There are stable orbits around but far away from the TLPs, For an observer on the Earth, there are only a few days in one month that these orbits are located in the proximity of the libration points.

10.3.1 Dynamical Substitutes Orbit

Two special stable quasi-periodic orbits called dynamical substitutes around each TLP in the earth–moon system are computed in the spatial sun–earth–moon system introduced [6]. The dynamical substitutes have long-term stability characteristics however the periodic or quasi-periodic orbits obtained in the circular restricted three-body problem (CRTBP) model become unstable in the real force model mentioned in (2.2). Beside, the dynamical substitutes are far away from the TLPs compared to the special stable quasi-periodic orbits obtained in the circular restricted three-body problem (CRTBP) mode is close to the TLPs.

In the real force model mentioned in (2.2), starting from the epoch J2000.0.

The four dynamical substitutes is obtained by integrating the initial state of the dynamical substitutes corresponding to the epoch J2000.0 in the Earth-centered J2000.0 equatorial coordinate, lasting for 1000 days, and then transform the state from the Earth-centered inertial frame to the synodic frame.

Figure 10.2 shows the long-term stability characteristics of the dynamical substitutes, and the orbits here can provide navigation orbit for the constellation.

10.3.2 Orbit Around the Dynamical Substitutes

As seen from Fig. 10.2, the dynamical substitutes have along-term stable characteristic with a large stable range of 0.1 to 0.2 of the average distance between the earth and the moon. The distance from the orbit of the translational point of the farther, the stability of the worse, but within a certain distance is stable, orbit around the dynamical substitutes is also very useful in aerospace applications.

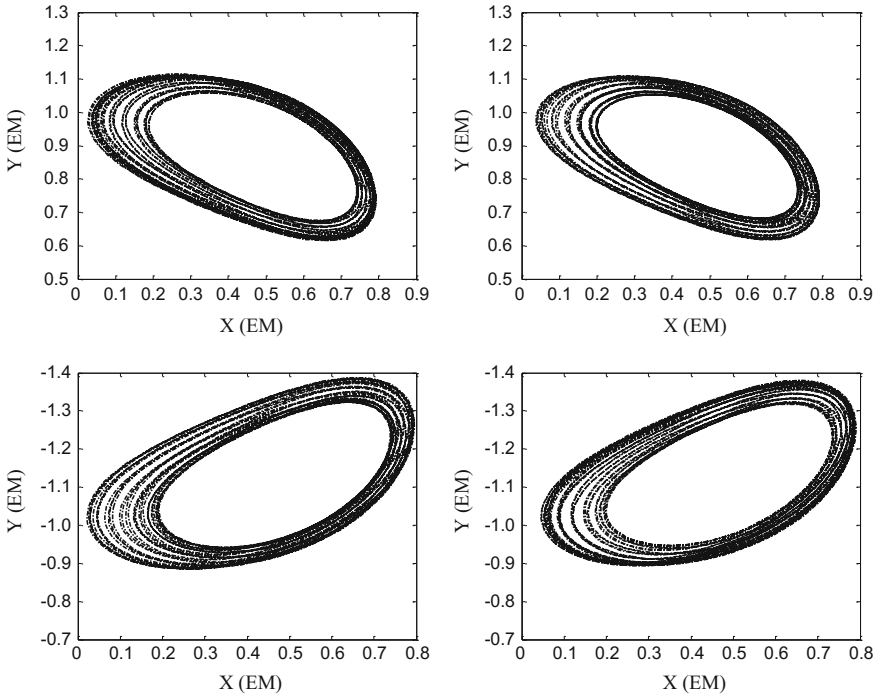


Fig. 10.2 Dynamical substitutes L4–1and L4–2 around TLP L4 (*upper*) and L5–1 and L5–2 around TLP L5 (*lower*) in the Earth-centered J2000.0 equatorial coordinate, starting from the J2000.0 and lasting 1000 days, transform the state from the Earth-centered inertial frame to the synodic frame

10.4 Orbit Determination Issues

10.4.1 OD Process

The batch filter without a priori estimation is used. The criterion for the iteration process to terminate is the change in the residual between the estimated orbit and the observations is smaller than a threshold [12].

The AOD process is carried out in the full force model.

10.4.2 OD Data Generation

Usually, the dynamical substitutes are described in the synodic frame. We transform the state of the dynamical substitutes to the Earth-centered inertial frame considering the fact that it is more convenient to implement the perturbations in the real force model.

To generate the simulated data, we first transform the initial conditions of the dynamical substitutes orbit to the Earth-centered inertial frame and integrate its orbit simultaneously with the other the dynamical substitutes or the orbit around the dynamical substitutes. At each "observation" epoch, the inter-satellite range data are taken as the differences between their positions. In order to simplify our studies, only white-noise errors (with a threshold) are considered to simulate the accuracy of the observation data. In our numerical simulations, three white-noise error thresholds of 1, 10 and 100 m are considered. Generally, the OD accuracy is proportional to that of the observation data. To save space, only the results of the threshold of 1 m are reported in our work. In the following simulations, without specifications, the sample interval is taken to be 60 min. One remark should be made. Different sample data with different intervals may lead to different OD results, but will not change the results qualitatively [8].

The orbit around the dynamical substitutes are Marked as L4–1_around, two sub problems are considered. They are (1) AOD of one dynamical substitutes orbit probe L4–1 and L4–1_around. (2) AOD of two dynamical substitutes orbit probes L4–1 and L4–2 of the same TLP. (3) AOD of two dynamical substitutes orbit probes of the different TLP. To simplify our discussion, the results of the orbit determination for probe (a) (with L4 point as an example) and case (c) are given here. For case (b) we have also made numerical simulations with the same results as cases (a) and (c), and are not given here.

10.5 Numerical Simulation

The AOD problem of two dynamical substitutes is the most important and the most basic one in our studies. Here, the dependence of the OD accuracy on the length of data, the initial point, out of plan component.

10.5.1 L4–1 and L4–1_around

10.5.1.1 Position of L4–1_around

The length of data is taken to be 100 d. Initial point on dynamical substitutes is the orbit point corresponding to the JD = 2,451,595.0.

Tables 10.1 and 10.2 show the results for the L4–1 and L4–1_around respectively, the OD accuracy of the dynamical substitutes is expressed by the radial (R), transverse (T) and normal (N) error.

Table 10.1 The relation between the OD results of the L4–1 satellite and the position of the L4–1_around

σ_R (m)	σ_T (m)	σ_N (m)	σ (m)	σ (m/S)	Z (m)
337.2798	507.8141	15,384.3593	15,396.4328	0.0301	5d3
168.9262	254.2050	7690.3237	7696.3780	0.0150	1d4
17.0207	25.5347	768.3218	768.9344	0.0015	1d5
1.7994	2.6452	75.5501	75.6178	0.0001	1d6
0.4131	0.5807	10.4004	10.4248	0.0000	5d6
0.3980	0.7958	2.7484	2.8889	0.0000	1d7
0.0552	0.1518	0.6406	0.6534	0.0000	5d7

Table 10.2 The relation between the OD results of the L4–1_around satellite and the position of the L4–1_around

σ_R (m)	σ_T (m)	σ_N (m)	σ (m)	σ (m/S)	Z (m)
337.3146	507.9464	15,384.6145	15,396.6929	0.0301	5d3
168.9609	254.3370	7690.5789	7696.6381	0.0150	1d4
17.0556	25.6654	768.5768	769.1944	0.0015	1d5
1.8338	2.7607	75.8026	75.8750	0.0001	1d6
0.4327	0.5860	10.5815	10.6065	0.0000	5d6
0.4846	0.7598	2.8596	2.9982	0.0000	1d7
0.0652	0.1218	0.7406	0.7534	0.0000	5d7

Our studies show that:

(a) The AOD of the L4–1 and L4–1_around based on the inter-satellite range data is feasible.
(b) For dynamical substitutes orbit, the OD accuracy is unsatisfactory for position of L4–1_around is located at the place in Z direction away from the L4–1 less than 10 km, but improves when the distance between the L4–1 and the L4–1_around increase.
(c) For a short distance of between the L4–1 and the L4–1_around, the OD accuracy in R and T direction of the dynamical substitute is the same magnitudes but is one or two magnitudes better than that in N direction, However, the difference improves for a long distance of between the L4–1 and the L4–1_around, the OD accuracy reach the same level of accuracy of the observation data in all direction (R, T, N) when the L4–1_around is located at the place 5×10^4 km away from the L4–1 in Z direction.
(d) When the L4–1_around is selected in the X, Y direction of dynamical substitute L4–1, the result is similar to the Z direction, Here is not repeated again.

10.5.1.2 Length of Data

We set the L4–1_around at 10 km in Z direction away from the L4–1, the initial point of dynamical substitutes orbit corresponding to the epoch 2,451,695.0 (Tables 10.3, 10.4)

Our studies show that:

(a) The AOD of the dynamical substitutes based on the inter-satellite range data is feasible.
(b) For Dynamical Substitutes Orbit, the OD accuracy is unsatisfactory for a short length of data less than30-day, it can be improved when then length of data is increases, until reach the same level of accuracy of the observation data.
(c) For a short length of data, the OD accuracy in R and T direction of the dynamical substitute is the same magnitudes but is one or two magnitudes better than that in N direction, but the difference improves for a long length of data.
(d) When the L4–1_around is selected in the X, Y direction of dynamical substitute L4–1, the result is similar to the Z direction, Here is not repeated again.

Table 10.3 The relation between the OD results of the L4–1 satellite and the length of data

σ_R (m)	σ_T (m)	σ_N (m)	σ (m)	σ (m/S)	Time (day)
16.4144	14.7662	217.0990	218.2188	0.0005	30
4.2277	2.8062	76.9190	77.0862	0.0001	40
2.7667	2.2680	11.8590	12.3869	0.0000	50
0.4218	0.6929	8.1188	8.1592	0.0000	60
0.4293	0.8491	11.3040	11.3440	0.0000	70
0.3563	1.1074	7.4879	7.5777	0.0000	80
0.4255	0.7749	1.2318	1.5162	0.0000	90
0.4086	0.7993	2.9304	3.0648	0.0000	100

Table 10.4 The relation between the OD results of the L4–1_around satellite and the length of data

σ_R (m)	σ_T (m)	σ_N (m)	σ (m)	σ (m/S)	Time (day)
15.0876	18.8442	218.9440	220.2708	0.0006	30
4.4857	3.2209	78.1304	78.3253	0.0001	40
3.0610	2.8357	11.8985	12.6089	0.0000	50
0.4635	0.7738	8.2610	8.3101	0.0000	60
0.5019	1.0130	11.5287	11.5840	0.0000	70
0.3654	1.2100	7.5398	7.6450	0.0000	80
0.5196	0.7926	1.1808	1.5141	0.0000	90
0.4959	0.7641	3.0521	3.1851	0.0000	100

10.5.1.3 Initial Point

In the following, the point from which we start the observation will be called the initial point. It is also the point for which we estimate its state vector in the OD process. However, the geometry between the L4–1 and the L4–1_around is constantly changing in a fixed arc. Different initial points are chosen, so different OD results can be expected.

The figure (a–h) in the Figs. 10.3, 10.4 and 10.5 corresponding to the relation between the OD accuracy and the initial point of dynamical substitute L4–1 when the lengths of data is 30, 40, 50, 60, 70, 80, 90 and 100 days.

We draw the following conclusions:

(a) The OD accuracy with respect to the initial point and for a longer length of data, the vibration becomes smoother and the vibration width also reduces. The reason for this difference lies in the fact that a short length of data can only cover a small portion of dynamical substitutes and the OD accuracy is sensitive to different initial points, but a long length of data can cover nearly the whole dynamical substitutes orbit.

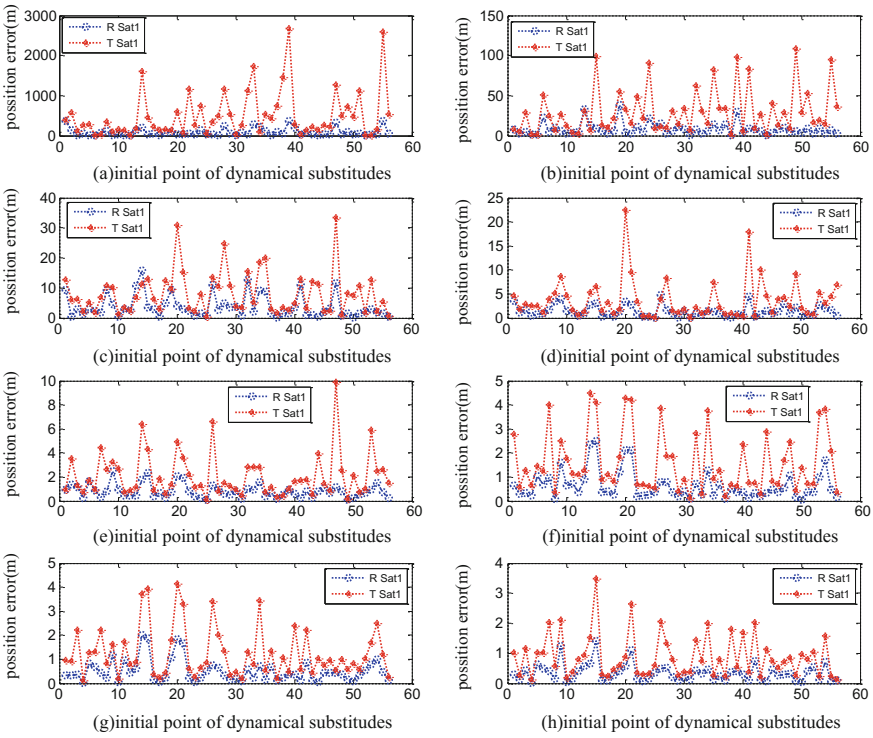


Fig. 10.3 The AOD results of L4–1 in R and T direction corresponding to different initial points of dynamical substitutes and the lengths of the data

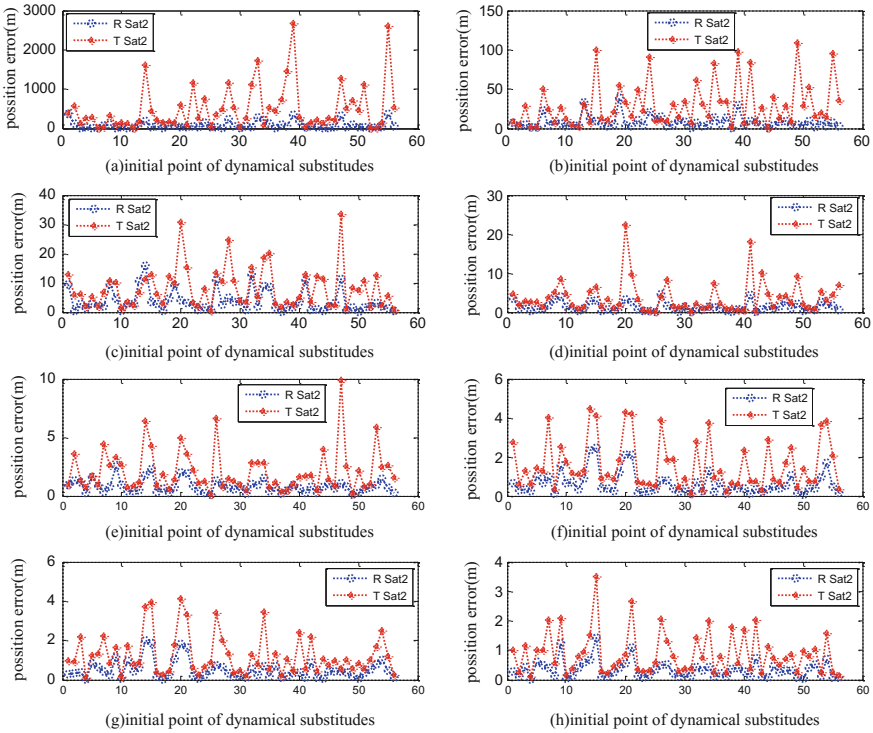


Fig. 10.4 The AOD results of L4–1_around satellite in R and T direction corresponding to different initial points of dynamical substitutes and the lengths of the data

(b) For the same length of data and initial point, the OD accuracy in R and T direction of the dynamical substitutes is the same magnitudes but is one or two magnitudes better than that in N direction. But the difference improves for a long length of data. thus makes the OD accuracy less sensitive to different initial points.
(c) The result of the OD between L4–1and L4–1_around is similar to the Z direction when the L4–1_around is selected in the X, Y direction of dynamical substitute L4–1, the result, here is not repeated again.

10.5.2 L4–1 and L4–2

10.5.2.1 Length of Data

We take dynamical substitutes corresponding to the epoch 2,451,695.0 as the initial point of dynamical substitutes orbit. The out-of-plane amplitude of the dynamical

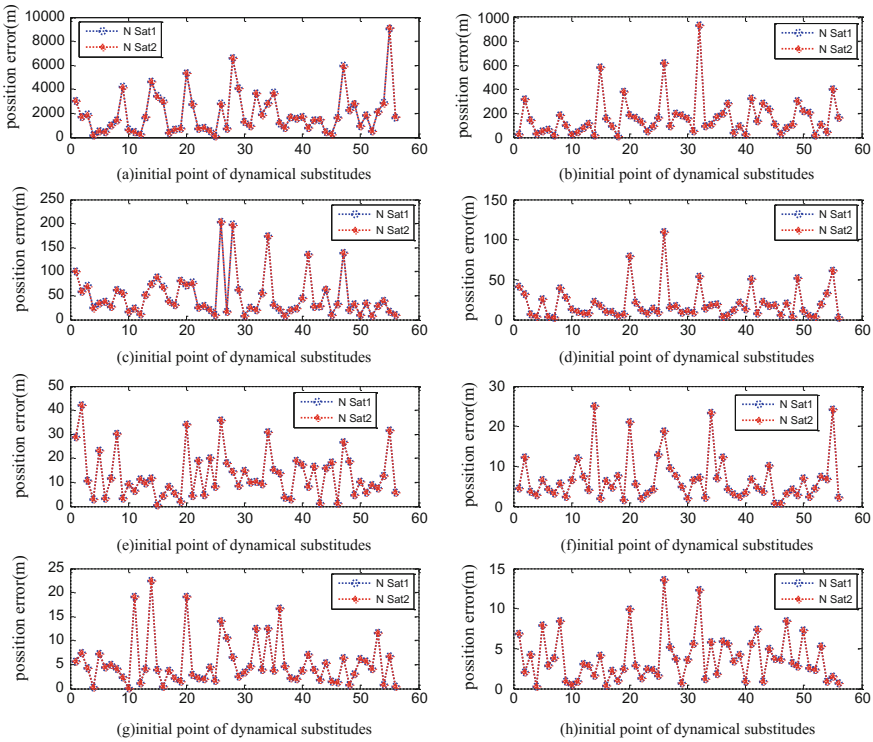


Fig. 10.5 The AOD results of L4–1_around in N direction corresponding to different initial points of dynamical substitutes and the lengths of the data

substitutes is taken to be −123,699.745654622762231 km for L4–1, −56,495.351467153588601 km for L4–2.

Tables 10.5 and 10.6 show the results for the L4–1 and L4–2 respectively.

Our studies show that:

a) The AOD of the dynamical substitutes based on the inter-satellite range data is feasible.

(b) For dynamical substitutes orbit, the OD accuracy is unsatisfactory for a short length of data less than 20-day, but improves when the data length more than 20-day data. However, it can be improved when the length of data is increases, until reach the same level of accuracy of the observation data;

(c) For a short length of data, the OD accuracy in R and T direction of the dynamical substitutes is the same magnitudes but is one or two magnitudes better than that in N direction, but the difference improves for a long length of data.

Table 10.5 The relation between the OD results of the L4–1 and the length of data

σ_R (m)	σ_T (m)	σ_N (m)	σ (m)	σ (m/S)	Time (day)
103.9321	964.7830	30,771.0654	30,786.3618	0.0522	20
67.9650	276.2265	6841.7295	6847.6407	0.0583	25
18.3132	37.0322	3129.5498	3129.8225	0.0136	30
4.3705	11.2337	155.4333	155.9000	0.0003	35
2.0924	4.2881	9.6136	10.7326	0.0001	40
0.5161	1.0887	89.5648	89.5729	0.0003	45
0.6772	3.2448	89.8757	89.9368	0.0003	50
0.5283	2.0321	59.7148	59.7517	0.0002	55
0.1838	0.3074	5.6792	5.6905	0.0000	60
0.1864	0.3795	12.3092	12.3165	0.0000	65
0.1446	0.2924	13.1118	13.1159	0.0000	70
0.1383	0.3832	12.9540	12.9605	0.0000	80
0.0978	0.1774	4.0067	4.0118	0.0000	90
0.0743	0.1215	2.1359	2.1407	0.0000	100
0.0767	0.1227	1.9585	1.9639	0.0000	120

Table 10.6 The relation between the OD results of the L4–2 and the length of data

σ_R (m)	σ_T (m)	σ_N (m)	σ (m)	σ (m/S)	Time (day)
254.7289	1099.5556	83,457.5380	83,465.1697	0.0598	20
6.3853	67.4737	2320.1770	2321.1667	0.0204	25
10.7985	26.8982	343.3903	344.6114	0.0019	30
3.5845	18.0660	81.1445	83.2085	0.0003	35
1.7841	7.4664	79.4377	79.8078	0.0001	40
0.0702	1.4936	67.5923	67.6089	0.0002	45
0.6566	0.7563	20.0142	20.0393	0.0000	50
0.3898	0.3972	22.4746	22.4815	0.0000	55
0.0657	0.5529	9.0585	9.0756	0.0000	60
0.1059	0.3530	4.9496	4.9633	0.0000	65
0.0772	0.3255	4.2745	4.2875	0.0000	70
0.0955	0.0985	8.3633	8.3644	0.0000	80
0.0966	0.1897	3.4676	3.4742	0.0000	90
0.0638	0.1257	1.4235	1.4305	0.0000	100
0.0668	0.1304	1.6415	1.6480	0.0000	120

10.5.2.2 Initial Point

In our numerical simulations, five hundred initial points are taken,the length of the data is Set as 50, 60…100 (day). Figures 10.6 and 10.7 show the OD results for the L4–1 satellite and L4–2 satellite in the direction of T and R respectively with

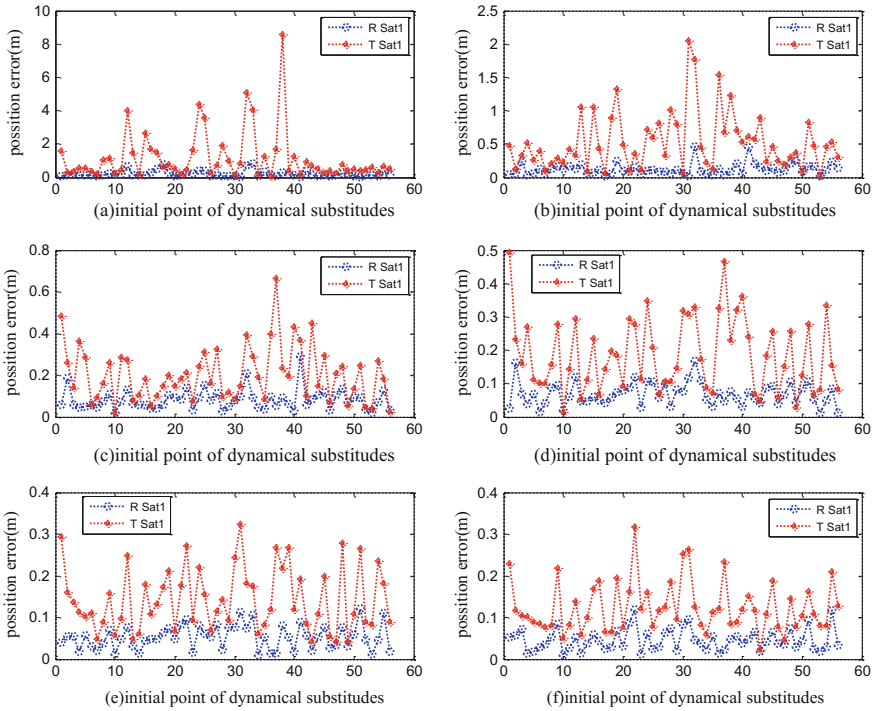


Fig. 10.6 The AOD results of L4–1 in R and T direction corresponding to different initial points of dynamical substitutes and the lengths of the data

respect to different initial points on L4–1 and L4–2 when the length of data is changing, Fig. 10.8 show in the direction of N.

The figure (a–f) in the Figs. 10.6, 10.7 and 10.8 corresponding to the relation between the position and the initial point of dynamical substitutes when the lengths of data is 50, 60, 70, 80, 90 and 100 days.

We draw the following conclusions:

(a) For a short length of data, the accuracy of orbit determination is influenced by the initial point, the fluctuation of the position error is more intense, while for the longer length of data, the orbit accuracy is less affected by the initial point and the fluctuation of the position error is gentler.
(b) When the length of data increases, a longer length of data can increase the effectiveness of measurement data. The sensitivity of the orbit determination to the initial point is reduced and the influence of the initial point selection is reduced
(c) For the same length of data and initial point, the OD accuracy in R and T direction of the dynamical substitutes is the same magnitudes but is one or two magnitudes better than that in N direction. But the difference improves for a

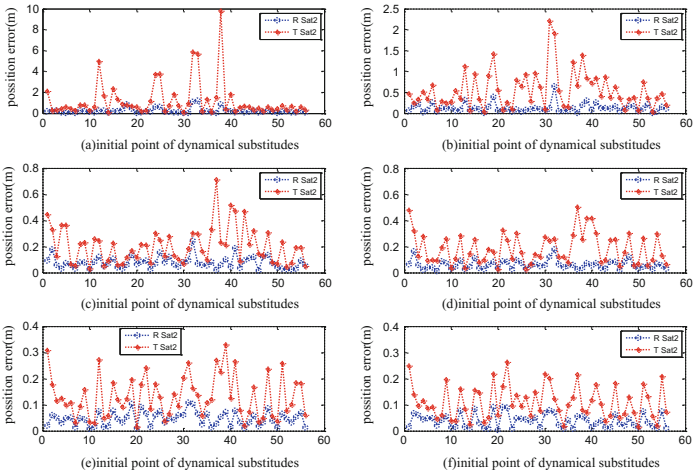


Fig. 10.7 The AOD results of L4–2 in R and T direction corresponding to different initial points of dynamical substitutes and the lengths of the data

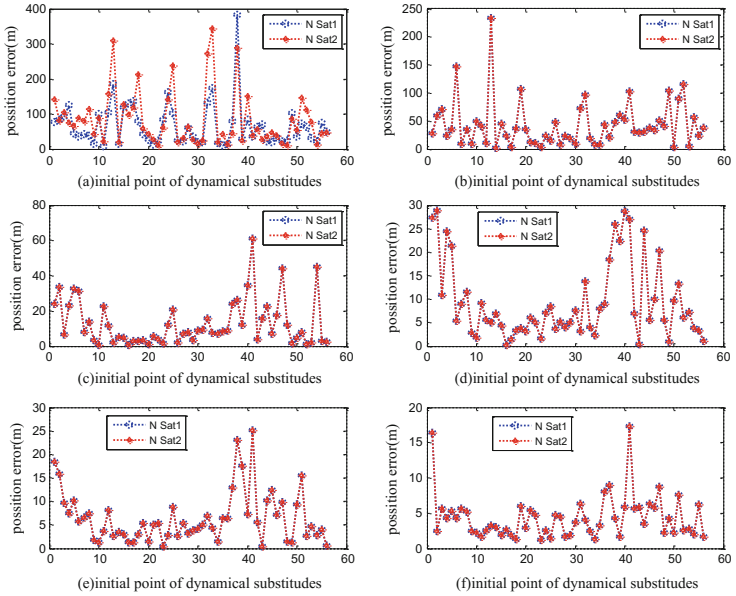


Fig. 10.8 The AOD results of L4–2 satellite in N direction corresponding to different initial points of dynamical substitutes and the lengths of the data

long length of data. and thus makes the OD accuracy less sensitive to different initial points.

(d) The result of the OD between any two of the dynamical substitutes L4–1, L4–2, L5–1, L5–2 is Similar to that of L4–1 and L4–2.

10.5.2.3 Vertical Displacement

Here, we study the effects of the amplitude of the dynamical substitutes on the AOD results.

The length of data is taken to be 40 day. Initial point on dynamical substitutes is the orbit point corresponding to the JD = 2,451,595.0 (Fig. 10.9).

Figure 10.10 shows the relation between the OD accuracy and the out-of-plane amplitude of the dynamical substitutes. Obviously, the OD accuracy of the dynamical substitutes is generally on the level of the observation data (· 1 m) for both L4–1 and L4–2. However, the OD accuracy for the dynamical substitutes shows an obvious tendency to reduce (from several meters to be less than 1 m) dynamical substitutes.

We draw the following conclusions:

(a) The OD accuracy improves when out-of-plane amplitude of the dynamical substitutes increases.

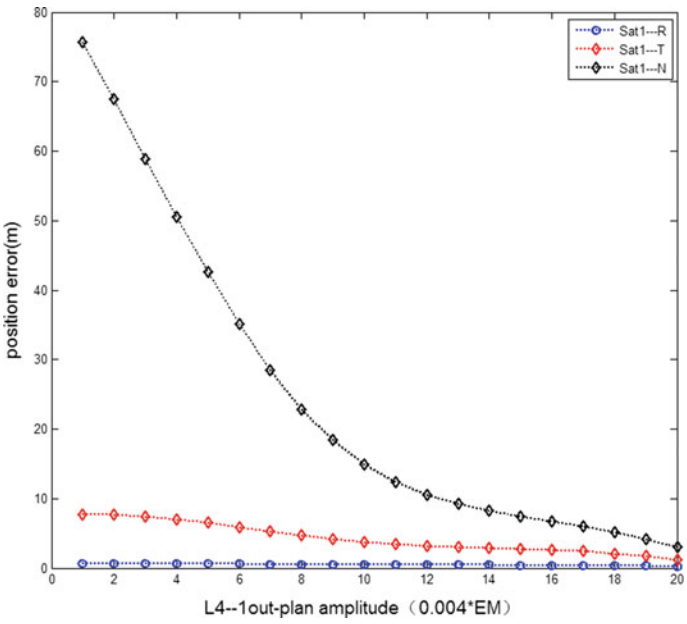


Fig. 10.9 The relations between the OD accuracy of the L4–1 and the out-of-plane amplitude

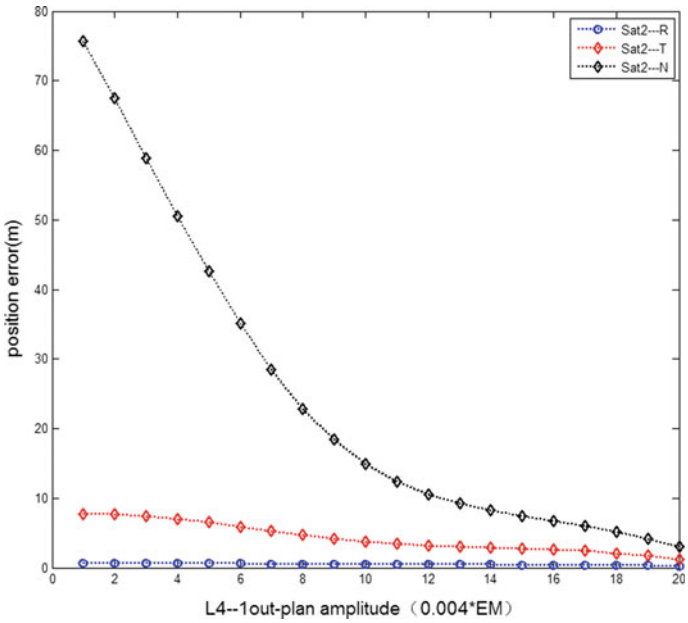


Fig. 10.10 The relations between the OD accuracy of the L4–2and the out-of-plane amplitude

(b) When the amplitude of one of dynamical substitutes is changed in Z direction, the Z direction amplitude needs to be changed a lot to meet the orbit precision requirement, however, the Z direction amplitude of the two dynamical substitutes can be changed at the same time, so that the two directions Z direction change Not much to the case of autonomous orbit determination accuracy requirements.

(c) When changing the Z-direction component of the orbit, the orbit accuracy can be achieved when the sampling arc length is short. Therefore, the Z-direction amplitude can be adjusted and the arc-length parameter combination can be sampled to achieve more accurate orbit determination.

10.6 Conclusions

Two kinds of special orbits are mentioned in this paper, one kind is the four dynamical substitutes around the TLP L4 and L5 in the earth–moon system; the other is the orbit around the dynamical substitutes. the AOD problem between a dynamical substitute and another one or the orbit around the dynamical substitutes is studied, and we have same conclusion

(a) The AOD of the dynamical substitutes based on the inter-satellite range data is feasible. The dependence of the OD accuracy on the length of data, the initial point, out of plan component.
(b) For a short length of data, the OD accuracy in R and T direction of the dynamical substitutes is the same magnitudes but is one or two magnitudes better than that in N direction, but the difference improves for a long length of data.
(c) The OD accuracy with respect to the initial point and for a longer length of data, the vibration becomes smoother and the vibration width also reduces. The reason for this difference lies in the fact that a short length of data can only cover a small portion of dynamical substitutes and the OD accuracy is sensitive to different initial points, but a long length of data can cover nearly the whole dynamical substitutes orbit.
(d) FOR a shorter length of data, the OD accuracy may reach the same level of accuracy of the observation data when Z-direction component of the orbit
(e) The accuracy of orbit determination can be determined by sampling arc, initial point selection and Z direction amplitude parameter, and the orbit determination can be carried out automatically and efficiently.

Acknowledgements This work is supported by national Natural Science Foundation of China (11322330) and national Basic Research Program of China (2013CB834100).

References

1. Hill K, Lo MW, Born GH (2005) Linked, autonomous interplanetary satellite orbit navigation (LiAISON), paper AAS 05-399, AAS/AIAA astrodynamics specialist conference, Lake Tahoe, CA
2. Liu P, Hou XY, Tang JS, Liu L (2014) Application of two special orbits in the orbit determination of lunar satellites. Res Astron Astrophys 14:1307–1328
3. Zhang L, Xu B (2014) A universal light house—candidate architectures of the libration point satellite navigation system. J Navig 67:737–752
4. Hou XY, Liu L (2010) on quasi-periodic motions around the triangular libration points of the real earth–moon system. Celest Mech Dyn Astron 108:301–313
5. Hou XY, Liu L (2011) On quasi-periodic motions around the collinear libration points of the real earth–moon system. Celest Mech Dyn Astron 110:71–98
6. Hou XY, Tang JS, Liu L (2015) Several possible mission orbits around Earth-Moon triangular libration points. In: 30th International Symposium on Space Technology and Science, Kobe-Hyogo, Japan
7. Hou XY, Scheeres DJ, Liu L (2015) Stable motions around triangular libration points in the real Earth-Moon system. Mon Not R Astron Soc 454:4172–4181
8. Du L, Zhang Z, Yu L et al (2013) SST orbit determination of Halo-LMO constellation in CRTBP. Acta Geod Cartogr Sin 42:184–190
9. Farquhar RW (1967) Lunar communication with libration-point satellites. J Spacecr Rocket 4 (10):1383–1384
10. Farquhar RW (1970) The control and use of libration-point satellite. Technical Report NASA TRR–346, NASA Goddard Space Flight Center

11. Hill K (2007) Autonomous navigation in libration point orbits. Ph.D. thesis, Graduate School of the University of Colorado
12. Tapley BD, Schutz BE, Born GH et al (2004) Statistical orbit determination, 200. Elsevier Academic Press, Burlington

Chapter 11
Orbit Maneuver Detection Based on Space-Based Angle Innovation

Lei Liu, Jianfeng Cao, Ye Liu, Songjie Hu and Geshi Tang

11.1 Introduction

Space is becoming crowded as the human activities in the outer space dramatically grow. According to the space target catalog of the United States Space Command, the LEO and GEO space objectives have been more than 10,000 in 1997 [1]. In the crowded outer space, the threats of collision and malicious attacks are more frequent than ever, e.g., the US-Russian satellite collision occurred in 2009. In order to reduce the probability of occurrence of such event, large space targets must be monitored in a real-time or nearly real-time style for awareness of orbit changes, including the sudden change of failure spacecraft because of disintegration or collision. As for the monitoring means, in addition to the traditional ground-based surveillance, the space-based surveillance is highly valued and widely used in recent years due to no limitations of ground light, weather and geographic conditions [2–5]. Therefore, the space-based surveillance probably plays a significant role in the future space surveillance. Furthermore, the space-based optical observations may become the principal means of space-based surveillance for quite a long time in the future [6].

The maneuver detection is generally modeled as an issue of binary hypothesis test [7–11], which is currently solved with the prevalent Chi2 method of sliding window significance test and the corresponding improvement methods. The Chi2 research is focused on the input sequence hypothesis, length selection of sliding

L. Liu (✉) · J. Cao · Y. Liu · S. Hu · G. Tang
Science and Technology on Aerospace Flight Dynamics Laboratory, Beijing 100094, China
e-mail: llbacc@139.com

L. Liu · J. Cao · Y. Liu · S. Hu · G. Tang
Beijing Aerospace Control Center, Beijing 100094, China

L. Liu
College of Science, National University of Defense Technology, Changsha 410073, China

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_11

window, determination of test statistic and so on. Chan et al. proposed the IE (input estimation) method of maneuver detection (Chi2_IE) [10], one of the most typical Chi2 method of sliding window significance test based on the input estimation. Bogler put forward an estimation algorithm of maneuver starting time based on the criterion of maximum likelihood (Chi2_EIE) [11]. With four typical scenes including sharp maneuver, slight maneuver, CT maneuver and random maneuver, Ru et al. used the ROC curve tool to compare the performance of six sliding window detectors (Chi2_MR, Chi2_IE, Gauss_IE, GLRT, MLRT, sliding window CUSUM), and investigated the influence of window length on detection performance [9]. Therefore, on the one hand, the maneuver detection researches are presently focused on the atmosphere targets such as aircraft and missile, while the maneuver detection of space targets using the space-based angle measurements is rarely addressed; on the other hand, some issues are still to be resolved for the maneuver detection of space targets, such as the influences of sliding window length on false alarm probability and detection accuracy, together with enhanced methods of the single detector structure.

Therefore, for the awareness requirement of the space orbit changes in real-time or near real-time, the method of orbit maneuver detection is investigated using the space-based angle innovation in the paper. The real-time motion estimation model for space targets is build using the space-based angle measurement. Finally, several methods of orbit maneuver detection are studied with the selection of the space-based angle innovation as the hypothesis test variable. The preferable methods are determined according to the numeric simulations and comparison.

11.2 Mathematic Model

11.2.1 Space-Based Angle Tracking

Space target tracking using space-based angle measurements is shown Fig. 11.1, in which S is the space-based platform and T is the space target. The geocentric radius vector of S and T is $\boldsymbol{r}$ and $\boldsymbol{R}$, in sequence. The distance between S and T is ρ, while the unit vector from S to T is $\boldsymbol{L}$. So

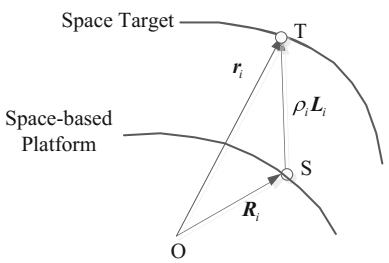


Fig. 11.1 Space target tracking using space-based angle measurements

$$\boldsymbol{r} = \boldsymbol{R} + \rho \boldsymbol{L} \tag{11.1}$$

Equation (11.1) is the measure model of the space-based angle tracking. The measure data can be the right ascension α and declination δ, which gives $\boldsymbol{L}$

$$\boldsymbol{L} = [\cos\alpha\cos\delta \quad \sin\alpha\cos\delta \quad \sin\delta]^T \tag{11.2}$$

The real-time tracking algorithm adopts the Kalman filter and the corresponding improvement method, e.g., EKF [12] and UKF. Following the tracking algorithm, the sate equation and measurement equation are derived to realize the real-time state estimation of space targets.

The state vector $\boldsymbol{X}$ consists of position and velocity of the space target. With the orbit dynamics, the state equation can be derived from the FG series [13]

$$\boldsymbol{X}(\mathrm{k}+1) = \begin{bmatrix} F & G \\ \dot{F} & \dot{G} \end{bmatrix}^{\mathrm{T}} \boldsymbol{X}(\mathrm{k}) + \boldsymbol{W}(\mathrm{k}) \tag{11.3}$$

where $\boldsymbol{W}(k)$ is the state estimation error supposed to be a white noise with a zero mean.

The observable $\boldsymbol{Y}$ directly takes the space-based angle measurement. Consequently, the measurement equation is derived from the combination of Eqs. (11.1) and (11.2). The linearized measurement equation is

$$\boldsymbol{y} = \Delta\boldsymbol{Y} = \boldsymbol{H} \cdot \Delta\boldsymbol{X} + \boldsymbol{V} \tag{11.4}$$

where $\boldsymbol{V}$ is the measurement error supposed to be a white noise with a zero mean. $\boldsymbol{H}$ is the partial derivative matrix of the observable to the state, which is calculated from Eq. (11.2), i.e.,

$$\boldsymbol{H} = \begin{bmatrix} \frac{\partial\alpha}{\partial\boldsymbol{r}} & \boldsymbol{0}_{1\times3} \\ \frac{\partial\delta}{\partial\boldsymbol{r}} & \boldsymbol{0}_{1\times3} \end{bmatrix}, \quad \frac{\partial\alpha}{\partial\boldsymbol{r}} = \begin{bmatrix} -d_y \\ d_x \\ 0 \end{bmatrix} \frac{1}{d_x^2 + d_y^2}, \quad \frac{\partial\delta}{\partial\boldsymbol{r}} = \begin{bmatrix} -d_x d_z \\ -d_y d_z \\ d_x^2 + d_y^2 \end{bmatrix} \frac{1}{d^2\sqrt{d_x^2 + d_y^2}},$$
$$\boldsymbol{d} = \boldsymbol{r} - \boldsymbol{R}.$$

11.2.2 Maneuver Detection Based on Innovation

As the important information in the state estimation, the innovation can accurately reflect the estimation precision and the real-time changes of the estimated state. According to the EKF model, the innovation υ_k of each state estimation for a linear system is

$$\upsilon_k = \boldsymbol{Y}_k - \boldsymbol{H}_k \hat{\boldsymbol{X}}_{k/k-1} \tag{11.5}$$

It can be proved that the innovation sequence $\{\upsilon_k\}$ of the optimal linear filtering is a white noise with a zero mean [12]. If the motion state is changed, the innovation sequence after the change does not turn to be a white noise with a zero mean. Therefore, an orbit maneuver is inferred. There are two specific detection styles, one detecting whether the innovation sequence is a normal distribution and the other detecting whether the distribution parameters of the innovation sequence deviates from the normal situation. The two styles all use the hypothesis test of the probability distribution.

11.2.2.1 Hypothesis Test Method

The hypothesis test methods are selected for the research including the χ^2 test method and KS (Kolmogorov–Smirnov) test method [14, 15].

The χ^2 test method checks the following relationship

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} \sim \chi^2(k - r - 1) \tag{11.6}$$

where p_i is the probability of $A_i (i = 1, 2, \cdots, k)$ that are k mutually exclusive events of the whole sample. n_i/n is the appearance frequency of A_i. r is the number of estimated parameters.

The χ^2 test method has a large probability of committing Type II error when it is used to test the normal distribution. Therefore, the KS test method is frequently used to test the normal distribution. The KS test method checks the sample by comparing the whole distribution with the standard normal distribution [14]. The relationship for the KS method is [15]

$$KS: \quad x^{\mathrm{KS}} = \max(\|F(x) - G(x)\|) \tag{11.7}$$

where $F(x)$ and $G(x)$ are the empirical cumulative distribution function and the standard normal cumulative distribution function, in sequence. $SCDF(x)$ is the empirical cumulative distribution function estimated from the sample. $CDF(x)$ is the normal cumulative distribution function having the same mean and stand deviation with the sample.

11.2.2.2 Maneuver Detection Method

Based on the above hypothesis test methods, four methods of orbit maneuver detection are presented, including the innovation and noise comparison method, distribution parameter detection method, innovation comparison method and

normalized innovation method. The first method is the weighted fusion method of two hypothesis test methods. The last three methods only use a single hypothesis test method. Therefore, in the following section, the four methods are divided into two groups, i.e., the multi-test weighted fusion group and the single test group. For each group, the sample of orbit maneuver detection is taken with a sliding window.

Innovation and Noise Comparison Method

The innovation at the moment of k is supposed to be υ_k. The length of sliding window takes n, which gives the sample $\{\upsilon_i\}$, $i = k-n+1, k-n+2, \cdots, k$. With the significance level taking α, the test result is

$$\nu = \begin{cases} 0, & H_0 : \boldsymbol{\upsilon} \sim N\left(\mu, \sigma^2\right) \\ 1, & H_1 : \boldsymbol{\upsilon} \nsim N\left(\mu, \sigma^2\right) \end{cases} \tag{11.8}$$

The innovation and noise comparison method tests the innovation sample distribution to be the same with that of the measurement noise or not. If the test result shows that they are not the same, the orbit maneuver is inferred. The method consists of the χ^2 method and KS method. Define the integrated distribution test result as

$$v_c = \kappa_1 v(KS) + \kappa_2 v\left(\chi^2\right). \tag{11.9}$$

Distribution Parameter Detection Method

The distribution parameter detection method tests the mean or standard deviation of the innovation sample is the same with that of the measurement noise or not. If the test result shows that they are not the same, the orbit maneuver is inferred.

The measurement noise is a white noise with a mean of μ_0 and a standard deviation σ_0. With the hypothesis precondition that the innovation sample $\{\upsilon_i\}$ has the same but unknown standard deviation with the measurement noise, $\{\upsilon_i\}$ is tested to have the same mean with μ_0 or not, or the same standard deviation with σ_0 or not. There are altogether three methods called μ^1, μ^2 and σ test method, in sequence, with the test relationships and rejection regions as

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1), |z| \geq z_{\alpha/2} \tag{11.10}$$

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim t(n-1), |t| \geq t_{\alpha/2}(n-1) \tag{11.11}$$

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2(n-1), \chi^2 \geq \chi^2_{\alpha/2}(n-1) \text{ or } \chi^2 \leq \chi^2_{1-\alpha/2}(n-1). \quad (11.12)$$

Innovation Comparison Method

Since the distribution characteristic of innovation is approximately the same as the measurement noise, the distribution characteristics of the innovation of each dimension should be approximately the same when the noise characteristics of each measurement are equal. Therefore, the distribution characteristics of the innovation on each dimension can be compared to determine whether the maneuver occurred. As the angle measurements in the research are two-dimensional, the innovation comparison method tests that the innovation samples on two dimensions have the same mean or standard deviation or not. If they are unequal, an orbit maneuver probably occurs.

Thus, the innovation comparison method is similar to the distribution parameter detection method with the approximately same hypothesis test procedures used. With the significance level α and the mean or standard deviation selected for test, there are still three specific ways called $\mu^1_{\alpha\delta}$, $\mu^2_{\alpha\delta}$ and $\sigma_{\alpha\delta}$ test method, in sequence, with the test relationships and rejection regions as

$$z = \frac{\bar{x}_\alpha - \bar{x}_\delta}{\sqrt{(\sigma_\alpha^2 + \sigma_\delta^2)/\mu}} \sim N(0,1), \quad |z| \geq x_{\alpha/2} \quad (11.13)$$

$$t = \frac{\bar{x}_\alpha - \bar{x}_\delta}{\sqrt{(s_\alpha^2 + s_\delta^2)/n}} \sim t(2n-2), \quad |t| \geq t_{\alpha/2}(2n-2) \quad (11.14)$$

$$F = \frac{s_\alpha^2}{s_\delta^2} \sim F(n\text{ - }1, n\text{ - }1), \quad F \geq F_{\alpha/2}(n-1, n-1) \text{ or } F \leq F_{1-\alpha/2}(n-1, n-1). \quad (11.15)$$

Normalized Innovation Method

With the fact that the innovation sequence $\{\upsilon_k\}$ of the optimal linear filtering is a white noise with a zero mean, i.e., $\upsilon_k \sim N(0, \boldsymbol{H}_k \boldsymbol{P}_{k/k-1} \boldsymbol{H}_k^T + \boldsymbol{R}_k)$, the normalized innovation γ_k is defined as

$$\gamma_k = \boldsymbol{\upsilon}_k^T (\boldsymbol{H}_k \boldsymbol{P}_{k/k-1} \boldsymbol{H}_k^T + \boldsymbol{R}_k)^{-1} \boldsymbol{\upsilon}_k \quad (11.16)$$

then γ_k is a χ_2 distribution with a degree of freedom equal to the υ_k dimension m, i.e., $\gamma_k \sim \chi^2(m)$. γ_k can be regarded as the sum of innovation on every dimension, which is tested to detect an orbit maneuver in the normalized innovation method.

When the length of sliding window is n, the sample consists of n independent data. Each data has a χ_2 distribution with the degree of freedom m, i.e., $\gamma_k \sim \chi^2(m)$.

Consequently, the sum γ_n of the normalized innovation of all data has a χ_2 distribution with the degree of freedom km, i.e.,

$$\gamma_n = \sum_{k=1}^{n} \gamma_k \sim \chi^2(km) \tag{11.17}$$

Equation (17) is used to test the distribution characteristic of all sample data with the significance level ρ_2. The orbit maneuver is inferred when $\gamma^n > \chi^2_\alpha(km)$.

11.3 Numerical Simulation

The space-based observation platform has a circular orbit with a height 1600 km and an inclination 54°. The space target has a circular orbit with a height 20,000 km, inclination 30°, right ascension 0° and true anomaly 40°. There exists a 1 h visible arc between the space-based platform and the target. Based on the current performance of space-based optical loads, the observation interval takes 1 s. The measurement error is the white noise with a zero mean and a standard deviation of 5″. The initial position and velocity errors are the white noise with a standard deviation 50 km and 50 cm/s, respectively. The significance level takes 0.01 for all hypothesis tests. The orbit maneuver of the space target is supposed to be 1, 5, 10, 20 and 50 m/s in the simulations. In the simulations, when an orbit maneuver is detected before the actual maneuver moment, the detection is considered to be a false alarm which is invalid. Meanwhile, with consideration of the observation random noise and uncertainty of the maneuver direction, the Monte Carlo method with 100 simulations is used for all hypothesis tests.

11.3.1 Test of Innovation Distribution

Innovation sequence meets the white noise characteristics when the state estimation stabilizes, which is an essential prerequisite for the orbit maneuver detection method based on innovation. With 100 numerical simulations, the maximum, minimum and average values of the measurement innovation are shown in Fig. 11.2.

Figure 11.2 shows the maximum innovation of every estimation is basically 15″, approximately three times of the standard deviation 5″ of the measurement error given in the simulation conditions. Given the significance level 0.01, the distribution of α and δ innovation are tested with the χ^2 method. The test results show that the α and δ innovation are all normal distributions in the 100 simulations. The results reflect that the innovation and observation error have the same distribution characteristics and the state estimation accuracy is high.

The interval estimation method is used to calculate the confidence interval of the mean and standard deviation of α and δ innovation. The results show that the mean range is $0 \pm 0.5''$ and the standard deviation range is $5'' \pm 0.3''$ for the α and δ

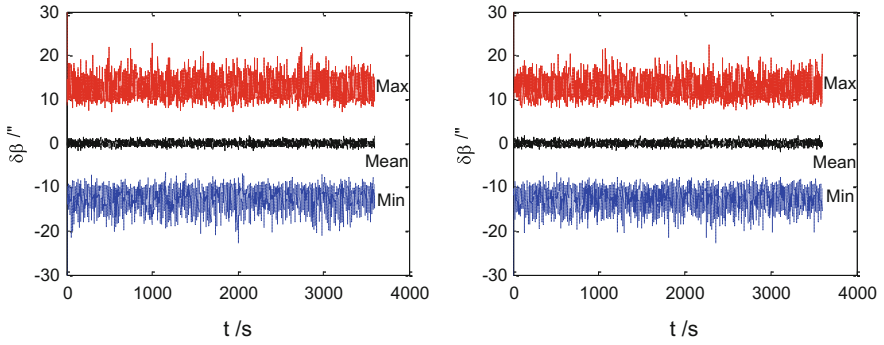


Fig. 11.2 Measurement innovation of an orbit without maneuver

innovation. Therefore, the characteristics of innovation and observation error are the same, which verifies the theoretical analysis of the innovation characteristics.

For the orbit maneuver case, the maneuver of 1 and 10 m/s occurring at the moment of 500 s are taken for example to show the innovation change, shown in Fig. 11.3. Figure 11.3 indicates that the innovation change is not intuitively obvious when the orbit maneuver is 1 m/s, but the innovation is significantly no longer a normal distribution when the maneuver increases to 10 m/s.

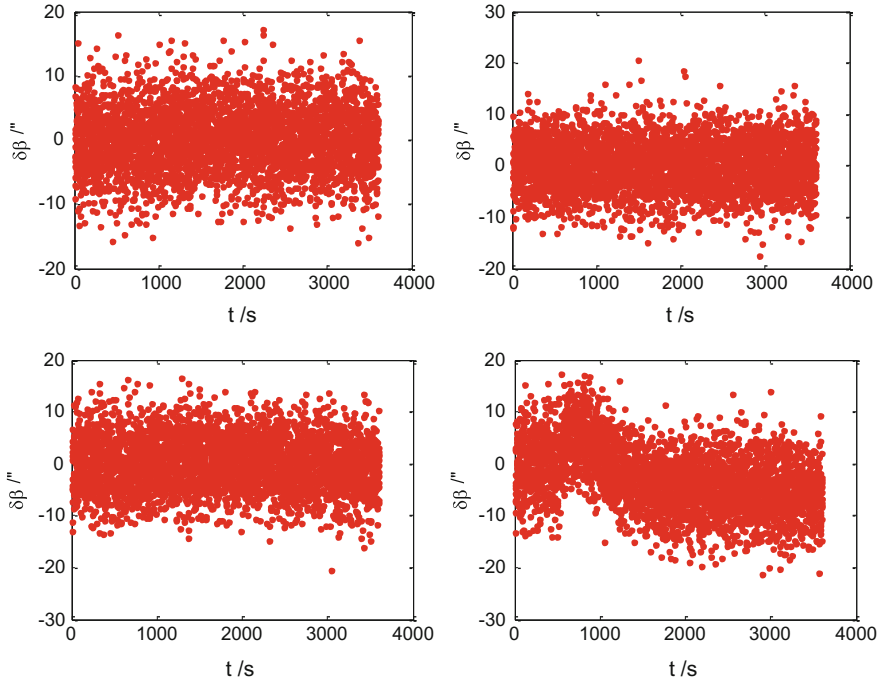


Fig. 11.3 Measurement innovation of orbits with maneuver

11.3.2 Multi-Test Weighted Fusion Group

The multi-test weighted fusion group is the innovation and noise comparison method. With the significance level given, the detection results are decided by the integrated distribution test result and the length of sliding window n. The detection conditions of orbit maneuver are shown in Table 11.1 for the multi-test weighted fusion group.

In Table 11.1, A, B, C and D are the detection conditions with the different combination of the sliding window and integrated distribution test result. Obviously, the results of A and D have the minimum and maximum credibility, in turn. B and C have the results with a medium credibility. However, A and B qualitatively have a higher detection sensitivity and a less amount of calculation than C and D.

With the detection conditions in Table 11.1, the detection result of orbit maneuver for the multi-test weighted fusion group is listed in Table 11.2 as the detection success rate. The orbit maneuver is inferred when the innovation change on any dimension is found.

Table 11.2 shows that D achieves the highest detection success rate, over 85%. B and D have a success rate greater than A and C, which illustrates the accuracy of the χ^2 method and KS method is the same. Furthermore, it is evident that C has a higher success rate than A and the similar case of D and B, which indicates the fact the greater length of the sliding window, the better detection effect, i.e., a larger number of samples is more in line with the overall distribution.

11.3.3 Single Test Group

The single test group includes the distribution parameter detection method, the innovation comparison method and the normalized innovation method. The first

Table 11.1 Detection conditions of orbit maneuver for the multi-test weighted fusion group

Sliding window	$v_c = 1$	$v_c = 2$
$n = 60$	A	B
$n = 120$	C	D

Table 11.2 Detection results of the multi-test weighted fusion group

Δv (m/s)	A (%)	B (%)	C (%)	D (%)
1	10	74	31	89
5	17	70	41	85
10	11	67	37	87
20	11	72	30	94
50	11	72	30	85

and second method include three detection modes that take 120 as the sliding window length from the viewpoint of calculation burden. The normalized innovation method still takes different sliding window lengths, i.e., 30, 60, 120 and 180 s. The detection success rate of all three methods in the single test group is shown in Table 11.3.

Table 11.3 reveals that the success rate of the distribution parameter detection method is greater than 70%. The μ^2 test mode is preferable for a little orbit maneuver while the σ test mode is more suitable for a large orbit maneuver.

For the innovation comparison method, the test modes using mean have a higher success rate than the mode using standard deviation. The former achieve a success rate greater than 80%, while the latter has a success rate lower than 80%. Furthermore, the cases with a sliding window length of 60 and 180 are also simulated. The simulation results show that three test modes have an approximately same success rate higher than 70% when the sliding window length is 60. When the sliding window length is 180, the test mode using mean has a success rate approximate 90%, while the test mode using standard deviation has a success rate among 60 and 90%.

For the normalized innovation method, the sliding window length 120 and 180 have a success rate approximate or higher than 90%, except for the case of 1 m/s orbit maneuver with a success rate or about 70%. Furthermore, the cases with a significance level 0.05 are also simulated. The simulations show that the success rate is lower than 50% for the sliding window length of 30 and 60 but higher than 60% for the window length of 120 and 180. The result indicates a larger significance level produces more false alarms that affect the detection accuracy.

Finally, the detection results of the above four methods are listed in Table 11.4 for a brief comparison.

Table 11.3 Detection results of the single test group

Δv (m/s)	μ^1 (%)	μ^2 (%)	σ (%)	$\mu^1_{\alpha\delta}$ (%)	$\mu^2_{\alpha\delta}$ (%)	$\sigma_{\alpha\delta}$ (%)	30 (%)	60 (%)	120 (%)	180 (%)
1	74	78	74	85	88	71	60	74	71	68
5	74	75	73	86	85	75	72	85	92	98
10	79	78	72	85	83	77	66	78	90	97
20	78	81	76	86	87	76	67	80	86	93
50	73	72	86	81	82	88	65	78	89	94

Table 11.4 Detection results of all methods

No.	Method	PARAMETERS OR MODE	η (%)
1	Innovation and noise comparison method	D: $v = 2$, $n = 120$	85–94
2	Distribution parameter detection method	μ	72–81
3	Innovation comparison method	$\mu_{\alpha\delta}$	82–88
4	Normalized innovation method	n = 180	68–98

11.4 Conclusions

The orbit maneuver detection methods based on space-based angle innovation are investigated in the paper. Four methods of orbit maneuver detection are presented based on four hypothesis test methods, including the innovation and noise comparison method, distribution parameter detection method, innovation comparison method and normalized innovation method. The research shows that the angle innovation contains the information of orbit change and can be used to detect the orbit maneuver. The two specific detection styles that test the innovation distribution or distribution parameters can detect the orbit maneuver with a high success rate. A success rate larger than 80% can be achieved by three methods, i.e., the innovation and noise comparison method, the innovation comparison method and the normalized innovation method. Furthermore, the method of normalized innovation can achieve a success rate larger than 90% for a not too little maneuver. The research is beneficial to the construction of our space situation awareness system.

Acknowledgements The authors are grateful for the support provided for this study by the National Natural Science Foundation of China (Nos. 61573049, 11303001, 11373013, 61304233) and China Postdoctoral Science Foundation (No. 2016M592947).

References

1. United States Air Force Scientific Advisory Board (1997) Space surveillance, asteroids and comets, and space target
2. USSTRATCOM (2010) Space situational awareness (SSA)
3. Smith MS (2006) Military space programs: issues concerning DOD's SBIRS and STSS programs. CRS report for congress
4. Lyon RH (2004) Geosynchronous orbit determination using space surveillance network observations and improved radiative force modeling. Massachusetts Institute of Technology, Cambridge
5. Stokes GH et al (2000) The space-based visible program. In: AIAA space 2000 conference & exposition
6. Liu L (2010) Initial orbit determination of space target based on the space surveillance. Graduate School of National University of Defense Technology
7. Ru JF, Chen HM, Li XR et al (2005) A range based detection technique for tracking a maneuvering target. In: Proceedings of SPIE on signal and data processing of small targets
8. Ru JF, Jilkov VP, Li XR et al (2005) Sequential detection of target maneuvers. In: Proceedings of international conference on information fusion, Philadelphia, PA
9. Ru JF, Bashi A, Li XR (2004) Performance comparison of target maneuver onset detection algorithms. In: Proceedings of SPIE conference on signal and data processing of small targets
10. Chan YT, Hu AGC, Plant JB (1979) A Kalman filter based tracking scheme with input estimation. IEEE Trans AES 15(2):237–244
11. Bogler PL (1987) Tracking a maneuvering target using input estimation. IEEE Trans AES 23(3):298–310

12. Zhang JH, Cai H (1995) Test statistics of flying vehicles. National University of Defense Technology, Changsha
13. Escobal PR (1965) Methods of orbit determination. Wiley, New York
14. Arsaglia G, Tsang W, Wang J (2003) Evaluating Kolmogorov's distribution. J Stat Softw 8(18):1–4
15. Conover WJ (1980) Practical nonparametric statistics. Wiley, Hoboken

Chapter 12
A Simulation Study of Orbit Determination for Lunar Probe via Relay Satellite

Jianfeng Cao, Lei Liu, Ye Liu, Weigang Su and Songjie Hu

12.1 Introduction

Chang'E-4 lunar exploration mission will be performed in 2018. The mission plans to achieve a soft landing on the far-side of the moon. Due to the synchronous rotation of the Moon, it is impossible to build a direct link between an Earth-based ground tracking station and a lunar probe over the far-side. As a result, the only way to build a link between the ground station and the probe on the far-side is to use relay satellites [1, 2]. The point at the Earth–Moon L2 Region is always visible to both the Earth and the far-side of the Moon at the same time. Relay orbits about the Earth–Moon point have great value on the exploration of lunar far-side and the kinetic character is complex. Liu studied the orbit maintenance technology of the relay orbit about the Earth–Moon L2 point, and pointed out that the Lissajous style is suitable for the orbit maintenance [3], which is quite favourable for the implementation of the exploration project. In Chang'E-4 mission, the Lissajous orbit around Earth–Moon L2 point is adopted for relay satellite to support the soft landing on the far-side.

Lunar prospector is the first lunar exploration mission planning to map the far-side gravity field utilizing relay satellite. Unfortunately, the utilization of relay satellite was cancelled due to the reduction of project funding [4]. In September of 2007, the Japanese SELENE mission consisting three separate satellites was launched. It was designed to track the main orbiter on the far-side with four-way Doppler tracking data, which made SELENE the first application using relay satellites in lunar exploration [5, 6]. Gravity Recovery and Interior Laboratory

J. Cao (✉) · L. Liu · Y. Liu · W. Su · S. Hu
Science and Technology on Aerospace Flight Dynamics Laboratory, Beijing 10094, China
e-mail: jfcao@foxmail.com

J. Cao · L. Liu · Y. Liu · W. Su · S. Hu
Beijing Aerospace Control Center, Beijing 100094, China

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_12

adopted a modified version of the inter-satellite tracking system used on the Gravity Recovery and Climate Experiment (GRACE) mission. During the normal mission phase, the orbital height is about 50 km, and GRAIL is about 30 km high above the lunar surface during the extended mission, the Ka band measurement data with high accuracy was obtained [7].

Although relay satellites have been applied in the precedent lunar exploration, the early mission only selected lunar orbit for relay satellite. Chang'E-4 (CE-4) would be the first attempt to adopt orbit around libration point for relay satellite. The earth–moon libration point has a lot of characteristics, such as poor observation geometry of relay satellite and it is difficult to determine its own orbit accurately. The geometric configuration for lunar probe and relay satellite is quite different from that of satellite-to-satellite orbiting the moon. And, the orbital error propagation is unknown. The orbit determination capability using relay satellite is studied based on the Chang'E-4 mission, and it can provide a theoretical support for the pre-research of the mission.

12.2 Basic Measurement Principles

12.2.1 Observation Equation

The relay satellite is a type of tracking satellite that provides four-way ranging or Doppler measurement, and it can also support communications. When the relay satellite is working, the ground station up-links ranging signals, and the relay satellite transponder relays the carrier waves to the spacecraft. Then the transponder on the spacecraft receives the signals and returns to the relay satellite, and down-linked to the ground station [8, 9] (Fig. 12.1).

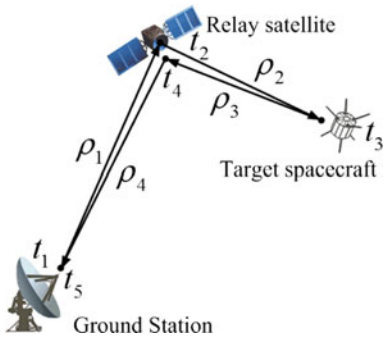


Fig. 12.1 Configuration of four-way measurement

According to the measurement principle, four-way ranging can be written as,

$$\begin{aligned}\rho_{4w} &= \rho_1 + \rho_2 + \rho_3 + \rho_4 + \rho_D + \varepsilon \\ &= |\vec{r}_{\mathrm{RELAY}}(t_2) - \vec{r}_{\mathrm{STA}}(t_1)| + |\vec{r}_{\mathrm{SAT}}(t_3) - \vec{r}_{\mathrm{RELAY}}(t_2)| \\ &\quad + |\vec{r}_{\mathrm{RELAY}}(t_4) - \vec{r}_{\mathrm{SAT}}(t_3)| + |\vec{r}_{\mathrm{STA}}(t_5) - \vec{r}_{\mathrm{RELAY}}(t_4)| + \rho_D + \varepsilon\end{aligned} \tag{12.1}$$

where, $\rho_i, (i = 1, 2, 3, 4)$ represents the geometric distance for each link, ρ_D is the delay introduced by path propagation, ε is measurement noise. $\vec{r}$ is the coordinate of the spacecraft and the relay satellite in geocentric celestial reference system, and the subscript STA refers to the ground station, RELAY refers to the relay satellite and SAT refers to the target spacecraft.

Relay satellite can also build satellite-to-satellite link with a target spacecraft. The basic principle is that the relay satellite transmit signal to the target spacecraft, and the target spacecraft transponder return the signal to the relay satellite. The satellite-to-satellite tracking measures the round trip travel time between the relay satellite and the target spacecraft.

The observation equation of satellite-to-satellite can be written as,

$$\rho = \rho_2 + \rho_3 + \rho_D + c \cdot \tau_{\mathrm{clk}} + \varepsilon \tag{12.2}$$

where, $\rho_i, (i = 2, 3)$ represent the geometric distances for each link, ρ_D is the delay introduced by path propagation, ε is measurement noise, τ_{clk} is clock error, and c is velocity of light.

12.2.2 Measurement Error

The four-way ranging measures the sum of distances of four links, which also includes the tropospheric and ionospheric delay, equipment delay. In general, the tropospheric delay can be calibrated accurately by the water vapor radiometer, and the ionospheric delay can be calculated by ionospheric model, such as international reference ionosphere [10]. The instrumental error from the ground station can be calibrated before each tracking task, and the error may be treated as a constant for each arc tracking. Transponder delay varies depending on the device, but the delay is more stable, and can be calibrated during the ground test phase [11]. In summary, the largest uncertainty of four-way ranging measurement is the path propagation delay, and other delay can be considered as a constant during each tracking path, and can be estimated in orbit determination.

Different from the four-way measurement, the propagation path of the inter-satellite links is relatively "clean". The measurement accuracy of inter-satellite link is affected mainly by satellite clock error, which includes two aspects: (1) error of time stamp; (2) systematic error. Because of the instability of the satellite clock, the time difference should be calibrated. Usually, the error is calibrated when the

clock error is larger than 0.5 ms. The clock error of the relay satellite will also directly introduce the random error, which will enlarge the measurement noise. For the time stamp effect, a calibration before the tracking task will reduce its effect.

12.3 Simulation and Analysis

In reality, visibility is such that the relay satellite can transmit signal to the target spacecraft and receive the transmitted signal. In this section, the visibility is analyzed and the orbit determination is performed by simulation. In the analysis, only the Lissajous orbit is considered for relay satellite. And because the relay satellite on the Lissajous orbit can remain stable for at least two weeks, the effects introduced by orbit maneuver are ignored.

12.3.1 Tracking Performance for Lunar Probes

The most important thing to carry out a tracking task of a lunar probe with relay satellite is the visibility between the relay satellite and the target spacecraft. The long tracking arc can ensure the reliable orbital accuracy. Ignoring the satellite attitude caused by the antenna pointing requirement, the only influence factor is the obstruction of the Moon.

For the lunar probe at a height of 200 km, the full-arc continuous tracking with relay satellite can be achieved if the relay satellite to the Moon connection is perpendicular to the orbital plane. If the connection is parallel to orbital plane, the duration of continuous tracking is the shortest. Even though, a tracking coverage of approximately 64.5% can be obtained. Compared with the ground stations, relay satellites will not be affected by the Earth's rotation, and can point to the Moon for the full time.

12.3.2 Orbit Determination Analysis Based on Four-Way Ranging

Based on current tracking capability, the random error of 3 m for the four-way ranging is considered. The orbit of relay satellite will not be discussed here. According to the orbit analysis of the early Chang'E-5 Test vehicle, it is expected that the accuracy of the orbit around the earth–moon L2 point is better than 100 m. If the long-arc segment can be observed continuously, the orbital accuracy is expected to be further improved. In this section, an orbital error of 10/100 m for relay satellite is considered. The simulation considers three types of error settings:

Table 12.1 Statistics of orbital error based on relay tracking data

Error Setting	Tracking arc (h)	Orbital error statistics (m)			
		x	y	z	Position
A	1	11.798	897.404	367.654	969.8
	2	3.873	109.388	46.293	118.8
	3	1.162	15.210	6.073	16.4
	4	0.909	11.297	4.577	12.2
	5	0.850	7.982	3.198	8.6
	6	0.739	7.622	3.058	8.2
B	1	91.455	4254.021	1690.433	4457.8
	2	51.115	3448.061	1419.852	3729.3
	3	147.084	470.526	215.387	537.9
	4	146.545	546.927	239.671	614.8
	5	140.596	777.989	337.126	859.4
	6	137.844	796.055	341.380	877.1
C	1	14.881	989.064	402.980	1068.1
	2	6.401	361.576	149.270	391.2
	3	14.754	49.426	22.370	56.2
	4	14.682	55.836	24.396	62.6
	5	14.085	78.203	33.862	86.3
	6	13.804	79.966	34.273	88.0

(A) only the measurement random errors; (B) measurement random errors and the orbital errors of the relay satellites with 100 m; and (C) measurement random errors and the orbital errors of the relay satellite with 10 m.

The increase of continuous tracking arc is helpful to improve the precision of orbit determination, especially for short arc condition. The analysis in this section considers tracking arcs of 1–6 h per day (Table 12.1).

If the orbit error of relay satellites is not taken into account, the orbital precision is gradually improved with the increase of the tracking arc. When the tracking arc is increased to 5 h, the orbital accuracy of the lunar probe can be better than 10 m. Once the orbit error of the relay satellite is taken into account, the situation will be changed. When the tracking arc length reaches 5 h, the orbit accuracy tends to be stable, and the increase of tracking arc is not conducive to improving the accuracy of the orbit. In this case, the relay satellite orbital error has become the main error source that affects the orbit solution of lunar probe. Moreover, the orbit error of the relay satellite has a linear effect on the orbit accuracy of the lunar orbiter. The result of the stabilized orbit determination is about one order of magnitude lower than that of the relay satellite itself.

12.3.3 Analysis of Orbit Determination Based on Inter-satellite Tracking Data

If the measurement error of the relay satellite is not taken into account, the simulation result is almost the same as the simulation result of the four-way measurement. This is because the ground station to the satellite link is very similar to the four-way measurement. The geometrical relationship between the relay satellite and the ground station does not change for the orbit calculation.

More detailed simulation analysis needs to consider the influence of the instability of the star clock on the orbit calculation. Ranging measurement error corresponding to the clock error is about 3 m. Long-term stability of satellite clock affects the time stamp and it can be calibrated by the difference between satellites overpasses and ground observation. After calibration, the error can be reduced to 10 μs, which is equivalent to a ranging error of several centimetres and can be ignored for orbit determination. The analysis in this section considers a random error of 10 m with a time stamp error of 10 μs.

The results of the orbit determination using only inter-satellite links are shown in the Table 12.2. Under the ideal conditions (without considering the orbital error of the relay satellites), the orbital accuracy of the lunar probe is improved with the increase of the tracking arc, and the accuracy is stable up to 20 m when the tracking arc is up to 5 h. Considering an orbital error of 100 m for relay satellite, the best orbit solution for lunar probe is about 1.5 km, and the increase of tracking arc will no longer improve the accuracy. If the relay orbit accuracy reaches 10 m, the orbit solution will be greatly improved, and will have an accuracy of 100 meters.

12.3.4 Analysis of the Short-Arc Ground-Based Auxiliary Tracking Data

For the actual mission implementation, the ground station will routinely track the probe each day. In this section, 1 h auxiliary ground station tracking is considered, and the noise of the ground-based two-way ranging is 3 m (Table 12.3).

The orbital accuracy is greatly improved with the assistance of ground station tracking observations. However, the error of the relay satellite is still an important factor that limits the accuracy of the orbit determination. Ignoring the orbital error of relay satellite, an orbital accuracy of better than 100 m can be obtained with 2 h tracking. If the tracking arc reaches 5 h, the orbit accuracy can be better than 10 m. The orbital error of relay satellite is still an important limitation. Although the auxiliary measurement of ground station can increase the orbital accuracy, the best orbit accuracy is several times that of relay satellite. Considering an accuracy of 100 for the relay satellite, the orbit solution can reach 300 m, if the orbital accuracy of relay satellite is 10 m, the orbital accuracy can be better than 100 m.

Table 12.2 Statistics of orbital error based on inter-satellite tracking data

Error setting	Tracking arc (h)	Orbital error statistics (m)			
		x	y	z	Position
A	1	23.1	1758.9	720.6	1900.8
	2	7.5	214.4	90.7	232.8
	3	2.2	29.8	11.9	32.1
	4	1.7	22.1	8.9	23.9
	5	1.6	15.6	6.2	16.8
	6	1.4	14.9	5.9	16.0
B	1	149.9	6976.5	2772.3	7310.7
	2	83.8	5654.8	2328.5	6116.0
	3	241.2	771.6	353.2	882.1
	4	240.3	896.9	393.0	1008.2
	5	230.5	1275.9	552.8	1409.4
	6	226.0	1305.5	559.8	1438.4
C	1	27.6	1839.6	749.5	1986.6
	2	11.9	672.5	277.6	727.6
	3	27.4	91.9	41.6	104.5
	4	27.3	103.8	45.3	116.4
	5	26.1	145.4	62.9	160.5
	6	25.6	148.7	63.7	163.6

Table 12.3 Statistics of orbital error based on inter-satellite tracking data and ground-based tracking data

Error setting	Tracking arc (h)	Orbital error statistics (m)			
		x	y	z	Position
A	1	10.1	100.4	107.6	147.5
	2	1.2	50.3	21.3	54.6
	3	1.3	14.2	4.1	14.8
	4	0.9	10.2	4.5	11.1
	5	0.8	5.9	3.1	6.7
	6	0.7	5.6	3.0	6.3

(continued)

Table 12.3 (continued)

Error setting	Tracking arc (h)	Orbital error statistics (m)			
		x	y	z	Position
B	1	81.4	1143.0	1010.2	1527.6
	2	41.3	918.0	889.2	1278.7
	3	85.4	373.3	194.3	429.4
	4	91.3	243.1	182.5	317.3
	5	100.8	273.4	187.3	346.3
	6	97.7	261.5	141.8	313.1
C	1	13.2	289.0	201.5	352.5
	2	5.4	191.5	101.3	216.7
	3	7.3	38.4	21.5	44.6
	4	7.2	46.2	22.4	51.8
	5	7.3	51.4	28.5	59.2
	6	7.5	53.3	27.3	60.3

12.4 Conclusions

To realize a soft landing on the far-side of the moon, relay satellites will be used in CE-4 mission. The relay satellite enables four-way measurement and inter-satellite link measurements to provide support for the orbit determination. In this paper, the support ability of relay satellites is analyzed, and the following basic conclusions are drawn.

1. The relay satellites located at the earth–moon L2 point can achieve a better tracking of the lunar probe. The tracking coverage can reach at least 64.5%, and in some special phase the relay satellite can track the probe all the time.
2. The orbital accuracy of the lunar probe based on the relay measurement is limited by the orbital accuracy of the relay satellite itself and cannot be compensated by the increase of the tracking arc. When the tracking arc reaches 5 h, the accuracy tends to be stable, and is of an order of magnitude worse than that of relay satellite.
3. The inter-satellite link between the relay satellite and lunar probe can also be used for orbit determination. Under this condition, the stability of the relay satellite clock directly affects the accuracy of the measured data and the accuracy of the time stamp. The orbit error of relay satellites is also the main error source which restricts orbit accuracy of lunar probe. Using inter-satellite tracking data, the orbital accuracy of the lunar probe is much less than that using four-way ranging.
4. The orbit accuracy of the lunar probe will be improved with the auxiliary of ground-based tracking data. Under current condition, an orbital accuracy of better than 100 m can be reached with the assistance of 1 h ground-based tracking data.

References

1. Namiki N, Takahiro I, Koji M et al (2009) Farside gravity field of the moon from Four-way doppler measurements of Selene (kaguya). Science 323:900–905
2. Zuber MT, Smith DE, Lemoine FG et al (1994) The shape and internal structure of the moon from the Clementine mission. Science 266:1839–1843
3. Liu L, Cao JF, Hu SJ et al (2015) Maintenance of relay orbit about the Earth–Moon collinear libration points. J Deep Space Explor 2:318–324
4. Konopliv AS, Binder AB, Hood LL et al (1998) Improved gravity field of the moon from lunar prospector. Science 281:1476–1480
5. Vighnesam NV, Sonney A, Subramanian B, et al (2009). Precise orbit computation of India's first lunar mission Chandrayaan-1 using accelerometer and tracking data during early phase. In: 21st international symposium on space flight dynamics, Toulouse, pp 1–16
6. Ikeda H, Ogawa M, Inoue T, et al (2009). Selene(Kaguya) orbit determination results and lunar gravity field estimation by using 4-way Doppler measurements. In: 21st international symposium on space flight dynamics. Toulouse, pp 1–11
7. Zuber MT, Smith DE, Watkins MM et al (2012) Gravity field of the moon from the gravity recovery and interior laboratory (Grail) mission. Science 339:668–671
8. Zhai Z (2016) Development of next generation data relay satellite systems. J Spacecr TT&C Technol 35:89–97
9. Shi XB, Li BJ, Wang K et al (2016) Development of three generations of tracking and data relay satellite system. J Spacecr TT&C Technol 30:1–8
10. Huang L, Wang H, Fan M (2012) Application of three-way measurement in deep space exploration. J Spacecr TT&C Technol 31:6–10
11. Chen XJ, Zhao J, Wang XJ et al (2016) Validating effects of ionospheric delay correction model. J Spacecr TT&C Technol 35:181–187

Chapter 13
A Multi-dimensional Genetic Algorithm for Spacecraft TT&C Resources Unified Scheduling

Jian Bai, Huili Gao, Xiaosong Gu and Huiying Yang

13.1 Introduction

Genetic algorithm (GA) was promoted by Prof. Holland [1] of the University of Michigan from the end of 1960s to the beginning of 1970s. After developed by many scholars, the GA has become the most successful intelligent optimal algorithm in the application. It simulates the reproduction law that the adapted one survives in the nature environment. The GA constructs a fitness function according to the object of the problem, performs evaluating, calculating, selecting on a population containing multiple chromosomes that each chromosome stands for a solution of the problem. After many reproductions, it takes the solution with best fitness value as the most optimal solution for the problem.

The spacecraft tracking, telemetry and command (TT&C) resources scheduling (STRS) problem aims to satisfy the requirements of TT&C services from versatile satellites, space probes, spaceships, and space stations, and schedules the resources rationally. Along with the aerospace industry development, the amounts and types of spacecraft are increasing more and more. They have been applied widely in many different scopes, such as communication, navigation, resource detection, meteorological observation, environmental observation, military affairs and etc. How to use the limited TT&C resources to support their TT&C events, namely solving the STRS problem, becomes a research hotspot for the industry.

STRS is a combination optimal problem with multi-constraints and high conflicts, its calculation is complex and costly, and now so far does not have significantly effective algorithm. Usually, it adopts algorithms such as linear programming [2, 3], heuristic algorithm [4–6], neighbourhood search [7], ant colony [8], GA [9–14], hybrid algorithm [15–17], etc. [18–20]. Many research results showed that the GA is the most effective one for the STRS relatively. In current researches,

J. Bai (✉) · H. Gao · X. Gu · H. Yang
Xi'an Satellite Control Center, Xi'an 710043, China
e-mail: baijianxscc@163.com

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_13

common GA adopts penalty function to decrease the fitness of the solution which includes conflicts. Under the circumstances of the resources shortage in the TT&C scheduling, the conflicts happen seriously and the solution gotten by penalty method could whether or not be sure with conflict. It is necessary to solve the conflict during the transform from the searching space to the result space, whereas the transformed result may not be the best optimal one. On the other side, the current researches are developed mainly on the scheduling of spacecraft in the low earth orbit (LEO), there is few researches explored on the high earth orbit (HEO) spacecraft.

This paper designed a GA based on gene repair method. At the same time, a multi-dimensional gene encoding method is specialized for the situation that it is not all the visible pass that requires TT&C support for the mission such as HEO spacecraft. This is much practical with the domestic status that the space TT&C network is developed and utilized integrated.

13.2 Hybrid Genetic Algorithm with Multi-dimensional Encoding

13.2.1 Two Methods of One-Dimensional Encoding

For the STRS problem of LEO, a visible arc has only two statuses: to be tracked or not to be tracked. So it is suitable to adopt binary encoding. When we sort all the visible passes according to the time of acquisition of signal (AOS), we get a queue in sequence: $V = (v_1, v_2, \ldots, v_n)$. We set the corresponding chromosome as $C = (g_1, g_2, \ldots, g_n)$, in which g_i is the i-th gene where 1 means that the visible pass i will be tracked, and 0 means that the visible pass i will not be tracked.

Assume there are m genes conflicts with gene g_k, which means that the pass g_k can't be tracked with the other passes at the same time. The corresponding mission weight of gene g_i is ω_i, where $i = 1, \ldots, m$. Thus the penalty function for g_k is as following:

$$\mathrm{P_l}(k) = \sum_{i=1}^{m} \omega_i \tag{13.1}$$

The penalty for the whole chromosome is as:

$$\mathrm{P_l} = \sum_{k=1}^{n} \mathrm{P_l}(k) \tag{13.2}$$

For the STRS of HEO, the basic requirement is to select a sector with satisfied duration to track in a long visible pass, so it is suitable to adopt real decimal encoding. In the real number encoding situation, it is difficult to design a repair

method for decoding. So the penalty method is a good option for fitness function. When the conflict happens, the penalty value is determined by the ratio conflicted duration to the required tracking length. If the required length could be compressed, it could be decided based on randomization.

Assume that the length of the gene g_k is L_k, and there are m genes conflicts with g_k. The i-th conflicted gene is g_i and its length is L_i. The L_k overlapped length with L_i is O_{ki}. Thus the penalty function for g_k is as following:

$$\mathrm{P_h}(k) = \sum_{i=1}^{m} \omega_i \cdot \frac{O_{ki}}{L_i} \tag{13.3}$$

where ω_i represents the weight of g_i.

The penalty for the whole chromosome is as:

$$\mathrm{P_h} = \sum_{k=1}^{n} \mathrm{P_h}(k) \tag{13.4}$$

13.2.2 Multi-dimensional Encoding

To unify these two methods of encoding described above, simplify the calculation and standardize the GA operator, a multi-dimensional encoding is designed which can unite the different requirements from both LEO and HEO.

13.2.2.1 Encoding Rules

A chromosome C corresponding to a solution for the problem is made up by L genes, where L is the number of visible passes to be scheduled in the STRS problem. Every gene g_i, $i = 1, \ldots, L$, includes 3 code (C_i^1, C_i^2, C_i^3), which indicates whether the pass will be tracked or not, the beginning of the tracking and the length of the tracking.

C^1 is the tracking flag. Its value is in the set of [0,1] which indicate whether the pass will be tracked, where 1 is *Yes*, and 0 is *No*.

C^2 is the beginning time of the tracking, namely t_b. Its value is in the scope of $[T_\mathrm{b}, T_\mathrm{e} - L_\mathrm{min}]$. T_b is the beginning of the pass and T_e is the end of the pass. L_min is the minimum length that the tracking length requirement could be compressed.

C^3 is the length of the tracking. Its value is in the scope of $[L_{\min}, L_{\max}]$. $L_{\min}$ is the minimum length that the tracking length requirement could be compressed and $L_{\max}$ is the required tracking length.

When the value of tracking flag (code C^1) is 0, the code C^2 and code C^3 does not effect in decoding, as they will be meaningless.

For the LEO scheduling, only the C^1 is changed. The C^2 and C^3 is fixed real decimal. It means that the tracking can't be compressed or moved. Thus the beginning of the tracking is the beginning of the visible pass.

13.2.2.2 Encoding

(1) Encoding

To search and index data conveniently, the one-dimensional gene code (C_i^1, C_i^2, C_i^3) is expanded to a two-dimensional matrix, which includes the necessary parameters for decoding.

Every gene is represented with a matrix as following:

$$g_i = \begin{pmatrix} c_i^{11} & c_i^{12} & c_i^{13} \\ c_i^{21} & c_i^{22} & c_i^{23} \\ c_i^{31} & c_i^{32} & c_i^{33} \end{pmatrix} \tag{13.5}$$

where, the first line of the matrix represents the code of resource selection:

c_i^{11} Indicates whether the pass will be tracked. 1 is *Yes* and 0 is *No*
c_i^{12} Indicates whether it permits the tracking beginning moving. 1 is *Yes* and 0 is *No*
c_i^{13} Indicates whether it permits the tracking length compressed. 1 is *Yes* and 0 is *No*.

The second line of the matrix represents the code of the tracking beginning time:

c_i^{21} Indicates the tracking beginning time. Its value is a real number. Its integer part is a simplified Julian calendar date number, and its fraction part is the time ratio of day
c_i^{22} Indicates the possible tracking beginning time. Its representation is same to c_i^{21}
c_i^{23} Indicates the possible tracking end time. Its representation is same to c_i^{21}.

The third line of the matrix represents the length of the tracking:

c_i^{31} Indicates the scheduled tracking length. Its integer part is the number of days, and its fraction part is the time ratio of day
c_i^{32} Indicates the minimum length of the required tracking. Its representation is same to c_i^{31}
c_i^{33} Indicates the maximum length of the required tracking. Its representation is same to c_i^{31}.

(2) Decoding

Every TT&C facility lines up in the chromosome, and the visible passes of each facility line up in the chromosome according to their beginning times.

In the decoding process, the three codes (C_i^1, C_i^2, C_i^3) of the gene g_i are translated to whether the visible pass is tracked, the beginning time of the tracking and the length of the tracking.

13.2.3 Fitness Function

In our research we use the multi-object integrated evaluation function promoted by BAI [21] as the fitness function, which is as following:

$$f = 0.630\,2f^{(1)} + 0.051\,7f^{(2)} + 0.213\,8f^{(3)} + 0.102\,5f^{(4)} \tag{13.6}$$

The definitions of each sub-function could be found in the reference document [21], where $f^{(1)}$ indicates the task fulfillment, $f^{(2)}$ the resources plentifulness, $f^{(3)}$ the term conformity and $f^{(4)}$ the duration satisfaction.

13.2.4 Constraints Process

Because of the encoding method, although the chromosome generated in the process may be valid in code, its corresponding solution exceeds the feasible scope and becomes invalid. In the STRS problem, when the corresponding solution contains tracking conflicts, which means that it breaks the tracking capability constraints of the facility, the solution would be invalid. In order not to contaminate the population, the invalid solution needs to be processed.

In general, the constraints processes of GA include following methods:

- Rejection
- Repair
- Penalty
- Special encoding and reproducing strategy

Rejection: This strategy rejects all the infeasible solutions produced in the evolution. It is the simplest and most inefficient method. When a feasible solution is not easy to achieve, even the initial population will be difficult to be built.

Repair: this strategy fixes the infeasible solution and transforms it to a feasible solution. It is easy to construct a repair process for the combination optimal problem, but it may decrease the versatility of the population. The repair methods

include partially mapped crossover, sequential crossover, cyclic crossover, transposition mutation, shift mutation, and etc.

Penalty: this strategy is a commonest process method for constraint satisfaction problem (CSP). It transforms the constrained problem to an unconstrained problem via setting penalty into the infeasible solution. Any violation of the constraints should add the penalty item to the objective function. By using penalty function, the fitness of the infeasible solution decreases and the possibility of its reproduction decreases too. But it require the appropriate penalty function, an inappropriate penalty function will cover up the optimal of the objective function.

In this paper special repair strategy is designed to fix the chromosome with conflicts. The strategy is as following:

1. Random correction

   (a) Put the passes which conflicts to each other into the same set, so to make multiple isolated conflict sets.
   (b) Select a set, draw a pass randomly from the set by its possibility, and process the pass according to its characteristic.
   (c) If this pass tracking is compressible, search if its free part could satisfy the scopes required by C^2 and C^3. If the answer is *Yes*, correct the gene to the feasible code value which could satisfy the requirement, then go to step (e).
   (d) If this pass tracking is uncompressible and unmovable, or it is compressible or movable, but it is not possible to find a free slot to insert it into the facility working queue even it is compressed to the minimum length, set this pass to *Not* be tracked, then go to step (e).
   (e) Check this set, remove the passes without conflicts.
   (f) If this set is empty now, go to step (b) to process another set.
   (g) When all the sets are already processed, the correction is finished and a feasible solution is produced. Go to the next process of calculating the fitness.

2. Heuristic correction

Considering the random correction strategy would not promote the fitness of the chromosome in the best way, a new step is added into the above strategy after the step (b):

(b)+ If the section with the drawn pass already has enough tracks to satisfy the task requirement, set the drawn pass to *Not* be tracked, then go to step (e); otherwise go to step (c).

13.3 Algorithms Comparison

GA has been taken as a better algorithm for the STRS commonly. Parish [9] presented a heuristic hybrid GA. It queued the task requirements, adopted permutation encoding method, and optimized the requirements sequence by GA. After the queue determined, heuristic algorithm is used according to the sequence to select the best scheme in the usable resources in turns, so that a scheduling result is built. When all of the requirements are processed, the result is evaluated with evaluation function to get its fitness value. The number of unsatisfied task is used as the evaluation function in [9]. To compare with the algorithm promoted in this paper, the formula (6) is taken as fitness for the GA. The modified algorithm is tagged as PGA.

To optimize the TT&C scheduling for LEO spacecraft, Wu [12] designed a GA using fitness function with penalty strategy. We referred the penalty method of his research, and expanded the encoding to multi-dimensional. The fitness function is as following:

$$f'(P) = \frac{f(P)}{N_c(P)} \tag{13.7}$$

where $N_c(P)$ represents the conflicts number of the scheme P.

If there is any gene fault in the chromosome, the conflict should be corrected before the output. This modified algorithm is tagged as pmGA.

13.3.1 Simulation Descript

To compare the effectiveness of different GAs, a simulation ia designed as following. Multiple combinations of different spacecrafts and facilities are selected from the database of in-orbit spacecrafts randomly. The spacecrafts requirements are same to their real work content. The number of spacecrafts and facilities of each combination is decided randomly also. The drawn data is shown in Table 13.1.

Three algorithms are performed on the same data sets of Table 13.1 separately and the scheduling results are compared.

The GA using multi-dimensional encoding and repair strategy designed in this paper is tagged as hmGA. The second algorithm tagged as pmGA is using penalty strategy for gene conflicts and the others are same to hmGA. The third algorithm tagged as PGA is the heuristic hybrid GA using requirements permutation encoding.

The population of each algorithm is 100. The regeneration number is 100. The crossover possibility is 0.6 and the mutation possibility is 0.002.

Table 13.1 Test data sets

Set	Spacecraft/number	Facility/number	Pass/number
1	8	3	76
2	13	3	130
3	19	2	154
4	12	5	185
5	13	5	189
6	9	6	190
7	11	5	199
8	10	6	212
9	15	5	262
10	19	4	268
11	14	5	302
12	14	6	384

Table 13.2 Time cost of the algorithms

Set	hmGA	pmGA	PGA
1	1006.8580	966.2508	266.7033
2	1771.6820	1769.3874	369.8194
3	2845.7818	3041.8636	1025.0877
4	2037.7986	2021.8077	597.2369
5	2596.5050	2684.4705	867.9778
6	1743.0325	1739.4050	570.6584
7	2005.4621	2080.9163	592.5237
8	1968.0996	1952.2419	608.2113
9	3021.4783	3062.2001	949.8157
10	3587.0092	3887.7756	1770.0946
11	2921.7663	2114.2176	657.7892
12	2430.0314	2435.0056	880.9915

13.3.2 Simulation Result

Three algorithms are performed on each data set and repeated 10 times. The 10 results of each set by each algorithm are evaluated and the evaluations are averaged in order to eliminate the bias raised by the data random nature.

The simulation results are shown in Tables 13.2, 13.3 and 13.4.

Table 13.3 Result evaluation of the algorithms

set	hmGA	pmGA	PGA
1	0.8760	0.8506	0.8520
2	0.8606	0.8544	0.8280
3	0.8076	0.7178	0.8296
4	0.9194	0.9030	0.8600
5	0.9240	0.8348	0.8570
6	0.8726	0.8652	0.8640
7	0.9040	0.8926	0.8750
8	0.9074	0.8770	0.8720
9	0.9006	0.8314	0.8690
10	0.8590	0.7996	0.8436
11	0.9220	0.9106	0.8730
12	0.9068	0.8954	0.8718

Table 13.4 Task fulfillment of the algorithms

Set	hmGA	pmGA	PGA
1	1.0000	0.9708	1.0000
2	1.0000	0.9952	1.0000
3	0.9678	0.8752	1.0000
4	1.0000	0.9912	1.0000
5	1.0000	0.9464	1.0000
6	1.0000	0.9942	1.0000
7	1.0000	0.9916	1.0000
8	1.0000	0.9810	1.0000
9	0.9910	0.9240	0.9910
10	0.9880	0.9396	1.0000
11	1.0000	1.0000	1.0000
12	1.0000	1.0000	1.0000

13.3.3 Simulation Result

Generally, the hmGA has best evaluation value, the PGA better, and the pmGA has the lowest value. On task fulfillment value, the hmGA and the PGA are at similar level, the both are better than pmGA.

In most situations, the results evaluation of the hmGA are better than PGA, which indicated that the hmGA is better in searching solution. But under the condition with significantly insufficient resources (e.g. 2 facilities support 19 spacecrafts), the result of PGA is better than hmGA. It indicates that in this condition the heuristic method has more guidance capability.

The time cost of hmGA and pmGA are similar, but the result of hmGA are much better than pmGA. It indicates that the repair strategy is more efficient in promoting scheduling result than the penalty strategy. The time cost of hmGA are around 3

times comparing with that of the PGA, which suggests that the calculation time of hmGA costs mainly on the processing of encoding and decoding.

13.4 Conclusion

This multi-dimensional encoding GA could describe and solve the STRS problem. It transforms the scheduling result evaluation method into effective chromosome fitness function, so that it transforms the resources allocation problem into the optimal solution searching problem in the solution space with chromosome. The hmGA promoted by this paper could solve the general scheduling problem including LEO and HEO, and maybe also applying to other kind of spacecraft such as deep space probe. It ensures the validation of the solution via special gene repair strategy, thus it can be applied to STRS better than the others.

References

1. Holland JM (1975) Adaptation in natural and artificial systems. MIT Press, Cambridge, MA
2. Gooley TD (1993) Automating the satellite range scheduling process. Air Force Institute of Technology, Wright-Patterson AFB, Ohio
3. Brrowbridge SE (1999) Optimal allocation of satellite network resources. Virginia Polytechnic Institute and State University, Blacksburg
4. Wolfe WJ, Sorensen SE (2000) Three scheduling algorithms applied to the earth observing systems domain. Manag Sci 46(1):148–166
5. Jin G, Wu XY, Gao WB (2004) Ground station resource scheduling optimization model and its heuristic algorithm. Syst Eng Electron 26(12):1839–1841
6. Yang P, Yang F, Wu B et al (2007) Heuristic algorithm and conflict-based backjumping algorithm for satellite TT&C scheduling. J Astronaut 28(6):1609–1613
7. Zufferey N, Vasquez M (2015) A generalized consistent neighborhood search for satellite range scheduling problems. RAIRO Oper Res 49(1):99–121
8. Zhang N, Ke LJ, Feng ZR (2009) A new model for satellite TT&C resource scheduling and its solution algorithm. J Astronaut 30(5):2140–2145
9. Parish DA (1994) A genetic algorithm approach to automating satellite range scheduling. Air Force Institute of Technology, Wright-Patterson AFB, Ohio
10. Barbulescu L, Watson JP, Whitley LD et al (2003) Scheduling space-ground communications for the air force satellite control network. J Sched 7(1):7–34
11. Barbulescu L, Howe A, Whitley D (2006) AFSCN scheduling: how the problem and solution have evolved. Math Comput Model 43(9–10):1023–1037
12. Wu B, Li YX, Huang YX (2006) Optimal scheduling of TT&C network resources based on genetic algorithm. J Astronaut 27(6):1132–1136,1167
13. Li YQ, Wang RX, Xu MQ et al (2012) An improved genetic algorithm for a class of multi-resource range scheduling problem. J Astronaut 33(1):85–90
14. Dong JQ (2013) TT&C network resource assignment model and simulation based on genetic algorithm. J Comput Appl 33(7):2074–2077
15. Wang HB, Xu MQ, Wang RX et al (2012) Solving space and ground TT&C resources integrated scheduling problem with ant colony optimization-simulated annealing algorithm. J Astronaut 33(11):1636–1645

16. Zhang TJ, Li JS, Li J et al (2016) Space-ground integrated scheduling based on the hybrid ant colony optimization. Syst Eng Electron 38(7):1555–1562
17. Ling XD, Wu XY, Liu Q (2009) Study of GATS algorithm for multi-satellite TT&C scheduling problem. J Astronaut 30(5):2133–2139
18. Chen DX, Xu R, Cui PY (2014) A temporal topological sort processing method for spacecraft resources constraints. J Astronaut 35(6):669–676
19. Du SY, Hu J, Jia LJ et al (2012) Multi-satellite TT&C scheduling method based on multi-living agent theory. J Spacecr TT&C Technol 31(5):11–14
20. Zufferey N, Amstutz P, Giaccari P (2008) Graph colouring approaches for a satellite range scheduling problem. J Sched 11(4):263–277
21. Bai J, Feng ZR, Li J, et al (2010) Design of evaluation method for resources scheduling in TT&C network management system. In: 61st international astronautical congress. Prague, Czech Republic 12:9962–9967

Chapter 14
Research on Health State Evaluation Method of Ground-Based TT&C Network

Tao Wu, Huili Gao, Junchao Chen, Yindi Wang and Huiying Yang

14.1 Introduction

The ground-based TT&C network [1, 2] is a global distribution network for aircraft services, including three interconnecting modules, which are the control center module, the TT&C node module and the communication link module, as shown in Fig. 14.1. With the rapid development of the space industry, the sharp increase in the number of on orbit spacecraft and the TT&C equipment, and the continuous improvement of technical complexity higher requirements are put forward for the use and management of the ground-based TT&C network. The management personnel of the ground-based TT&C network must grasp the health state of the ground-based TT&C network, especially when its health state become deteriorated, they should be able to timely corresponding measures, to ensure that the ground-based TT& C network can operate in a reliable and stable and continuous way. Therefore, it is necessary to evaluate the health state of the ground-based TT&C network.

Since the experts in the U.S.A, the United Kingdom, etc. proposed the concept of Prognostics and Health Management (PHM) [3–6] in 1990s, the health state evaluation as a primary function of the PHM technology has been widely used in the fields of bridge, electric power, ships, air-crafts, etc. [7–20] after years of study and practice. However, there has been fewer studies on the large open system with complementary system node function like the ground-based TT&C network. In the paper, we established a health assessment mode based on the characteristics of the ground-based TT &C network.

T. Wu (✉) · H. Gao · J. Chen · Y. Wang · H. Yang
Xi'an Satellite Control Center, Xi'an 710043, China
e-mail: itunk@126.com

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_14

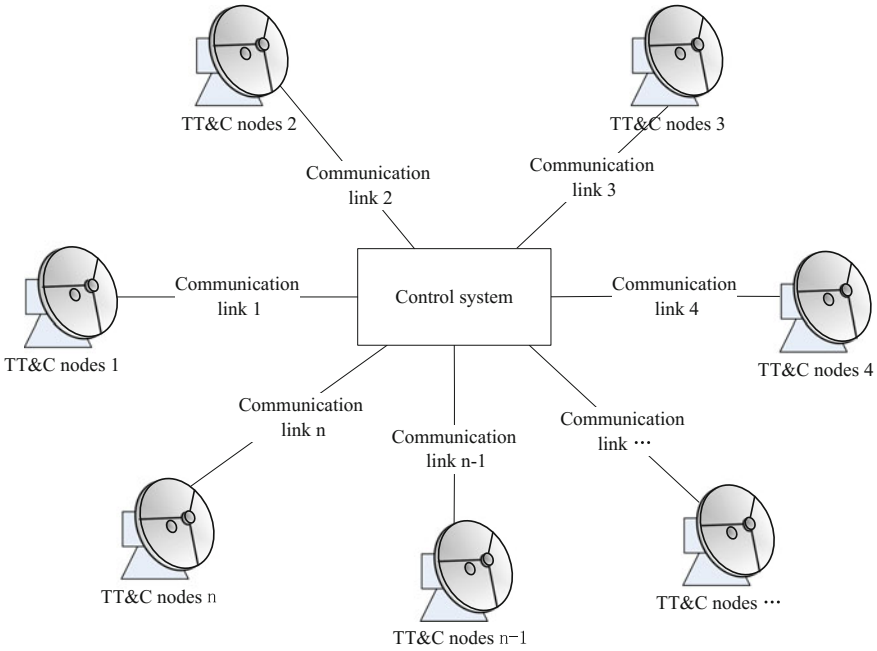


Fig. 14.1 Schematic diagram of the ground-based TT&C network

14.2 Health Evaluation Model for the Ground-Based TT&C Network

14.2.1 Overall Evaluation Design

In this paper, we classified the health state levels of the ground-based TT&C control network into four levels: Excellent, Good, Medium and Poor, proposed to use the health indexes of the ground-based TT&C network (N) to reflect the health level of the ground-based TT&C network (it is specified that when $N < 50\%$, it belongs to "Poor", when $50\% \leq N < 70\%$, it belongs to "Medium"; when $70\% \leq N < 90\%$, it belongs to "Good"; when $90\% \leq N$, it belongs to "Excellent") and determined the health level of the ground-based TT&C network using the quantitative calculated health index.

The ground-based TT&C network in China is a star network composed of a large number of TT &C Nodes connecting to the control center by corresponding communication links. The health state of the whole TT&C network is determined by the health state of the control center, the control nodes and the communication links. Considering the control center cannot be replaced, and the functions of each control node are complementary and have certain differences, each TT&C node and

its communication link jointly determine the health level of the TT&C node of this communication link, so we designed Formula (14.1) to calculate the health index of the ground-based TT&C network.

$$N = C \times \left(\sum_{i=1}^{n} p_i \times D_i \times L_i \right) \tag{14.1}$$

where, N is the health index of the ground-based TT&C network; C is the health index of the control center; D_i is the health index of the ith TT&C node; L_i is the health index of the link between the ith TT& node and the control center; for N *is within* [0,1], C *is within* [0,1], D_i *is within* [0,1] and L_i *is within* [0,1], "0" represent the system is completely damaged and can not be used, "1" represents that the system is entirely healthy, without any fault; "n" represents the number of the TT&C nodes; "p_i" represents the weight coefficient of the ith TT&C node, and the sum of the weights of all the nodes shall be 1.

To carry out the overall health evaluation of the ground-based TT&C network, we needed to study the influence factors of the health state of the three modules (the control center module, TT&C nodes module and communication links module), and calculated the value of "C" (the health index of the control center), "D" (the health index of the TT&C node), "L"(the health index of the communication link) and "p" (the weight of the TT&C node) respectively using an appropriate method, then the overall health index of the ground-based TT&C network can be obtained, as shown in Fig. 14.2.

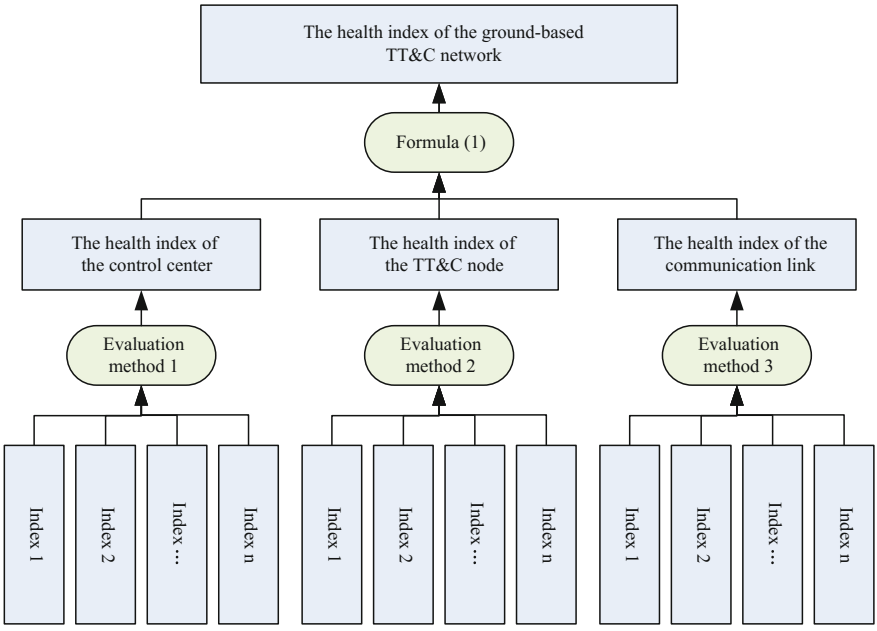


Fig. 14.2 Health evaluation plan chart for the ground-based TT&C network

14.2.2 Module Evaluation Design

A lot of similar studies had been carried out for the three Modules *C*, *D* and L and the typical studies included the bridge system health state evaluation [11, 12] and the aircraft system health state evaluation [17–20]. In this paper, we determined the health index through calculating the special indexes influencing the health state of the module, as shown in formula (14.2).

$$H = \sum_{i=1}^{m} q_i \times I_i \tag{14.2}$$

where, H is the health index for single module, such as the control center module, the TT&C nodes module, and the communication link module, I_i is the ith index to evaluate H and I_i *is within* [0,1]; m is the number of indexes to evaluate H; q_i is the weight coefficient of the ith index, and the sum of the weights of all the nodes shall be 1. Therefore, evaluated the health state of the various modules of the ground-based TT&C network, namely analyzing of the indexes affecting the health state of each module and calculating it according to Formula (14.2).

In addition, according to the idea of setting the health state standard value in reference documentation [15], this paper set the standard value of evaluation indexes for each module, as shown in Fig. 14.3. The I_i index was evaluated by the difference between the measured value and the standard value. When the measured index value was equal to the standard value, I_i was equal to 1; when the measured index does not meet the standard value, evaluated I_i according to the difference, making I_i *within* [0,1); in order to achieve dimensionless processing on the evaluation indexes, integrated the real-time data, periodic data and other data and achieved a comprehensive evaluation of the health state of each module.

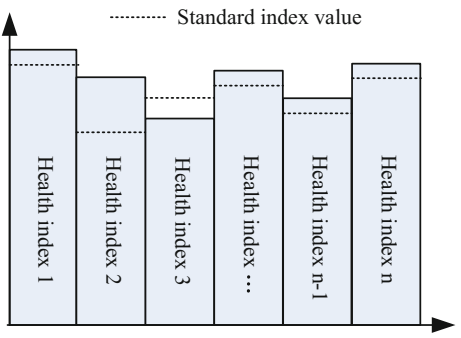


Fig. 14.3 Standard index value histogram for each module

14.3 Implementation of Ground-Based TT&C Network Health Condition Evaluation

Based on the module described in Sect. 14.2, the implementation of ground-based TT&C network health condition evaluation requires the following 4 steps:

1. Establishment of an indicator system for each evaluation module.
2. Determination of the weights of various kinds and standard values of the indicators in the indicator systems involved in the evaluation program.
3. Collection of various kinds of measured indicator data in the evaluation cycle.
4. Calculation of the health indicator of the ground-based TT&C network and evaluation of the health level of the ground-based TT&C network.

14.3.1 Indicator System Establishment

14.3.1.1 Control Center Indicators

A control center is a complex information management center that can process multi-spacecraft tasks parallelly and complete the life-cycle management of spacecraft task preparation, implementation and long-term in-orbit movement. There are numerous factors influencing the health of the control center. In this paper, the control center is evaluated as an overall information system mainly in terms of the load, stability and reliability of the control center, with the following evaluation indicators designed:

1. Task resource load. It reflects the control center's ability to support TT&C tasks, task resource load = the quantity of spacecrafts being supported by the control center/the maximum quantity of spacecrafts supported by the control center.
2. Equipment resource load. It reflects the control center's ability to support TT&C equipment, equipment resource load = the quantity of equipment being supported by the control center/the maximum quantity of equipment supported by the control center.
3. Storage resource load. It reflects the control center's ability to store information, storage resource load = the used storage capacity of the control center/the maximum storage capacity of the control center.
4. Real-time traffic load. It reflects the control center's ability to transmit information, real-time traffic load = actual data traffic of the control center/the maximum data traffic supported by the control center.
5. Mean time between failures (MTBF). It reflects the control center's stability, MTBF = working hours of the control center/number of failures of the control center within working hours.
6. Mean time to repair (MTTR). It reflects the control center's reliability, MTTR = total time required for repairs of the control center/number of failures.

14.3.1.2 TT&C Node Indicators

TT&C nodes refer to a single set of TT&C equipment and are basic units that provide TT&C support for spacecrafts. In this paper, from the angle of the TT&C network's administrator, key attention is paid to the equipment statuses of TT&C nodes. The TT&C equipment statuses fall into three statuses, i.e. normal use of equipment (including normal TT&C, emergency TT&C, ground station-satellite compatibility, task joint debugging), equipment unavailability (including equipment transition, equipment transformation, equipment maintenance, equipment fault, adverse weather) and equipment idleness. Therefore, the following evaluation indicators are designed in this paper.

1. Equipment availability. It reflects the degree to which TT&C equipment is available for TT&C tasks in the evaluation cycle, equipment availability = actual available working hours/evaluation cycle. The actual working hours refer to the actual available working hours of equipment in the evaluation cycle, excluding equipment maintenance, transition, transformation and faults.
2. Equipment plan completion rate. It refers to the degree to which TT&C equipment completes a TT&C plan in the evaluation cycle, equipment plan completion rate = actual number of tracked laps/arranged number of laps in the TT&C plan. The arranged number of laps in TT&C plans involve the arranged number of laps in conventional TT&C plans and the arranged number of laps in emergency TT&C plans. The actual number of tracked laps refers to the total number of laps actually tracked by equipment, also including the number of laps involved in emergency tracking.
3. Equipment failure index. It refers to the degree to which TT&C equipment suffers hardware and software failures, equipment failure index = number of equipment hardware and software failures/equipment working hours, in which the unit time of equipment working hours is generally 1,000 h.
4. Environmental impact index. It reflects the degree to which TT&C equipment affects TT&C tasks due to environment-related reasons (such as typhoons and thunders) in the evaluation cycle, environmental impact index = number of environment affecting TT&C tasks/equipment working hours, in which the unit time of equipment working hours is generally 1000 h.
5. Equipment work load. It reflects the use of TT&C equipment in the evaluation cycle, equipment work load = the actual equipment tracking work hours/the theoretical equipment tracking work hours.

14.3.1.3 Communication Link Indicator System

TT&C nodes refer to a single set of TT&C equipment and communication links are data channels connecting TT&C nodes and control center. Currently, all communication links for ground-based TT&C networks feature biplane communication and mutual backup, contributing to the communication link health index

$L_i = (L_{ia} + L_{ib})/2$, where L_{ia} is the ith communication link and a is the plane health index; L_{ib} is the ith communication link and b is the plane health index, with the calculation mode still being Formula (14.2). The indicators for evaluating the health state of communication links are mature, mainly as follows:

1. Data transfer rate, which is used for evaluating the actual data transfer rate.
2. Data transfer band width, which is used for evaluating the channel capacity, and describing the maximum data transfer rate of a channel, the limit of data transfer capacity of the channel.
3. Data transfer delay, which is used for evaluating data transfer delay, transfer delay = channel length/data transfer rate in the channel.
4. Data transfer bit error rate, which is used for evaluating the reliability of data transfer, bit error rate = number of bits involving errors/number of bits of transferred data.

14.3.2 Determination of Weights

Weight refers to the relevant importance degree of indicator items in the evaluation system, and such methods as the Delphi method, analytic hierarchy process (AHP), fuzzy comprehensive evaluation method, variation coefficient method and entropy method [21–23] are commonly used for determining the weights of indicator items.

14.3.2.1 Weights of Module Indicators

AHP is used in this paper to determine the weights of indicator items, with the specific method as follows:

1. Judgment matrix construction. Based on Formula (14.2), the judgment matrix $\boldsymbol{P}$ is constructed, in which u_{ij} refers to the value of relative importance of Indicator Item I_i to I_j, as shown in formula (14.3).

$$P = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & & \vdots \\ u_{n1} & u_{n2} & \cdots & u_{nn} \end{bmatrix} \tag{14.3}$$

2. Calculation importance ranking. Based on the judgment matrix, the characteristic vector of its maximum characteristic root $\lambda_{\max}$ corresponds to is obtained, with the equation as follows:

Table 14.1 Weights of indicators

Module	Indicator 1	Indicator 2	Indicator 3	Indicator 4	Indicator 5	Indicator 6
Control center	0.154	0.151	0.056	0.198	0.230	0.211
TT&C nodes	0.214	0.266	0.165	0.158	0.197	–
Communication link	0.211	0.199	0.269	0.321	–	–
Difference factor	0.237	0.699	0.064	–	–	–

$$Pw = \lambda_{\max} w \tag{14.4}$$

Upon normalization, the obtained characteristic vector w is the weight of indicator items.

3. Consistency check. Whether the obtained weight assignment is rational requires the consistency check of the judgment matrix, with the formula used in the check as follows:

$$CR = CI/RI = \frac{(\lambda_{\max} - n)/(n-1)}{RI} \tag{14.5}$$

where, RI is the average random consistency indicator of the judgment matrix, which can obtained by checking the table. In the case of $CR < 0.1$, it is thought $\boldsymbol{P}$ has satisfactory consistency, otherwise the elements in $\boldsymbol{P}$ call for adjustment to make it have satisfactory consistency.

In this paper, based on the indicator system, 10 experts have been invited to construct a control center judgment matrix, a TT&C node judgment matrix and a communication link judgment matrix respectively. After data processing and testing, the weight of each module indicator is obtained, as shown in Table 14.1. Due to the limited space in this paper, the data operation process is omitted here.

14.3.2.2 Weights of TT&C Nodes

The functions of the TT&C nodes in the TT&C network are not only similar but also different to a certain degree, with the differences mainly shown in three aspects, i.e. position (S_i), feature (T_i) and years of use (Y_i), where the position of TT&C nodes determines the visible duration for different spacecrafts, the features of TT&C equipment (such as supported frequency band, supported TT&C system, multi-objective design) determine the spacecrafts that can be served, the years of use of TT&C equipment influences the performance of such equipment. The setting of TT&C node weights is intended to correct the differences between TT&C nodes. Different from common weight determination, a weighted mean idea is used in this paper, and Formula (14.6) is used for calculation.

$$g_i = (k1Si + k2Ti + k3Yi)/3 \tag{14.6}$$

where, k_1, k_2, k_3 are the weights of TT&C node difference factors, with the determination method stated in Sect. 14.3.2.1, the result as shown in Table 14.1;g_i is the calculated weight value of the ith TT&C node and after normalization treatment, the weight value p_i of each TT&C node can be obtained.

14.3.3 Determination of Standard Values

The standard values of evaluation indicator items are the indicator criteria that the system should meet and the basic requirements for the sound operation of the TT&C network. Some indicators involved in this paper lack similar definite requirements, which are given based on the authors' experience.

1. It is required that the task resource load shouldn't be greater than 80%.
2. It is required that the equipment resource load shouldn't be greater than 80%.
3. It is required that the storage resource load shouldn't be greater than 50%.
4. It is required that the real-time traffic load shouldn't be greater than 50%.
5. It is required that MTBF should be more than 4000 h.
6. It is required that MTTR should be less than 3 h.
7. It is required that the equipment availability should be greater than 90%.
8. It is required that the equipment plan completion rate should be greater than 95%.
9. It is required that the equipment failure index rate should be smaller than 3‰.
10. It is required that the environmental impact index should be smaller than 3‰.
11. It is required that the equipment working load should be smaller than 80%.
12. It is required that the data transfer rate shouldn't be smaller than 64 kbps.
13. It is required that the data transfer band width shouldn't be smaller than 128 kbps.
14. It is required that the data transfer delay shouldn't be longer than 1 s.
15. It is required that the data transfer bit error rate should be smaller than 10^{-6}.

14.3.4 Data Collection and Evaluation

At the data collection and evaluation stage, it is necessary to determine the scope of data collection first, then followed by the collection of data about indicators of various kinds based on the constructed evaluation system, with evaluation result obtained through calculation based on the formula provided by the model and the evaluation report given finally.

14.3.4.1 Evaluation Scopes

The evaluation scopes in this paper are as follows:

1. Time scope: the whole year 2015 is selected as the scope of this evaluation.
2. TT&C data scope: This evaluation is aimed at the spacecrafts in S Frequency Band.
3. TT&C node scope: This evaluation is aimed at 34 TT&C nodes that can be used in S Frequency Band.

14.3.4.2 Indicator Calculation

Based on the differences in the evaluation indicator items, real-time calculation and regular calculation modes are adopted respectively to obtain these data evaluation indicators. Individual indicator items involves a case where a real-time indicator value (obtained through real-time calculation) and a mean indicator value (obtained through regular calculation) coexist. It is stipulated in this paper that in the case of the real-time indicator value not exceeding the allowed limit, the mean indicator values should be used; in the case of the real-time indicator value exceeding the allowed limit, the real-time indicator value should be used, with the calculation of various kinds of data as follows:

1. Control center health index: $C = q_1I_1 + q_2I_2 + q_3I_3 + q_4I_4 + q_5I_5 + q_6I_6 = 94.83\%$.
2. TT&C node health index: after normalization of the measured data of each TT&C node based on the indicator standard value, Formula (14.2) is used to calculate the health index of each TT&C node, such as $D_{14} = q_1I_1 + q_2I_2 + q_3I_3 + q_4I_4 + q_5I_5 = 95.97\%$, with the calculation results of the 34 sets of TT&C equipment as shown in Table 14.2.
3. Communication link health index: the equipment health index is referred to, such as $L_{ia} = q_1I_1 + q_2I_2 + q_3I_3 + q_4I_4 = 100\%$.

14.3.4.3 Health Evaluation

Based on Formula (14.1) in the evaluation model, coupled with the actual situation of the TT&C network involved in this evaluation, the ground-based TT&C network health index can be obtained through calculation based on Formula (14.7).

$$N = C \times \left(\sum_{i=1}^{34} \frac{g_i}{\sum_{j=1}^{34} g_j} \times D_i \times \frac{L_{ia} + L_{ib}}{2} \right) \tag{14.7}$$

Table 14.2 Equipment health index calculation

Equipment serial no.	Measured value of equipment availability	Measured value of equipment completion rate	Measured value of failure index/(‰)	Measured value of environment index/(‰)	Measured value of equipment load	Equipment health index/(%)
1	0.899	0.984	2.740	1.090	0.636	99.97
2	0.650	0.993	2.450	1.840	0.709	91.28
3	0.868	0.947	2.790	3.090	0.709	98.63
4	0.633	0.946	2.670	2.890	0.790	90.57
5	0.860	0.987	1.440	1.150	0.436	98.60
6	0.868	0.994	0.620	0.000	1.000	69.18
7	0.866	0.963	2.610	6.270	0.685	93.01
8	0.949	0.999	0.300	0.000	0.702	100.00
9	0.735	0.971	1.910	0.000	0.607	94.24
10	0.920	1.000	0.000	0.000	0.659	100.00
11	0.879	1.000	0.000	0.000	0.482	99.27
12	0.982	1.000	0.000	0.000	1.000	70.30
13	0.909	0.976	0.750	3.940	0.782	98.18
14	0.806	0.979	1.130	3.390	0.694	95.97
15	0.522	1.000	0.000	0.000	0.512	86.81
16	0.950	0.998	0.320	0.000	0.966	75.35
17	0.668	0.993	2.620	1.570	0.478	91.91
18	0.847	0.995	1.310	1.310	0.482	98.15
19	0.973	1.000	0.000	0.000	0.688	100.00
20	0.973	0.998	0.000	0.000	0.852	92.23
21	0.979	1.000	0.000	0.000	1.000	70.30
22	0.979	1.000	0.000	0.000	1.000	70.30
23	0.967	1.000	0.000	0.000	1.000	70.30
24	0.112	0.974	2.760	0.000	0.731	72.51
25	0.868	0.988	2.930	2.510	0.619	98.88
26	0.691	0.970	1.670	3.330	0.665	92.07
27	0.567	0.975	1.640	1.640	0.654	88.38
28	1.000	1.000	0.000	0.000	1.000	70.30
29	1.000	1.000	0.000	0.000	1.000	70.30
30	1.000	1.000	0.000	0.000	1.000	70.30
31	1.000	1.000	0.000	0.000	1.000	70.30
32	1.000	1.000	0.000	0.000	1.000	70.30
33	1.000	1.000	0.000	0.000	0.800	100.00
34	0.934	0.997	2.290	0.000	0.604	100.00

The TT&C network health index, 88.32%, is obtained based on calculation by substituting the control center health index, and health index and weight value of 34 TT&C nodes and the communication link index into Formula (14.7), indicating good health level of the ground-based TT&C network and the overall good health condition of the ground-based TT&C network.

At the same time, it is found that individual indicators exceed the allowed limit, which reflects certain hidden health risks involved in the ground-based TT&C network, with specific information as follows:

1. In Table 14.2, the measured values of the environment indexes of Equipment 3, 7, 13, 14 and 26 exceed the allowed limit, reflecting these sets of equipment are significantly influenced by the environment.
2. In Equipment 6, 12, 16, 21, 22, 23, 28, 29, 30, 31 and 32 (11 sets of equipment in all) in Table 14.2, the measured equipment load values of 10 sets of equipment are 100% (indicating full-load operation) and measured equipment load value the remaining 1 set of equipment is 96.6%, showing too heavy load of these sets of equipment, involving hidden health risks.
3. In Table 14.2, the measured equipment availability value of Equipment 24 is merely 11.2%, so this equipment went through nearly 300-day transformation in 2015, severely influencing its availability.

14.4 Conclusions

The result shows that the foundation TT&C evaluation model presented in this paper features strong operability, easy engineering implementation, intuitive reflection of the ground-based TT&C network health index and health level and effective solution to the actual difficulties involved in the overall foundation TT&C management. At the same time, based on the information about indicators exceeding the allowed limits, hidden health risks of various kinds can be found early so that measures can be taken in a timely manner to eliminate or minimize or delay such risks, which can provide scientific basis for TT&C resource maintenance decisions.

References

1. Zhao YF (2002) Technical development of satellite TT&C network. J Spacecr TT&C Technol 21(3):1–4
2. Zhai ZA, Wu B (2002) TT&C network of development conception. J Spacecr TT&C Technol 19(3):7–12
3. Peng Y, Liu DT, Peng XY (2010) A review: prognostics and health management. J Electron Meas Instrum 1:1–9

4. Guo YM, Cai XB, Zhang BZ et al (2008) Review of prognostics and health management technology. Comput Meas Control 16(9):1213–1216
5. Sun B, Kang R, Xie JS (2007) Research and application of the prognostic and health management system. Syst Eng Electron 29(10):1762–1767
6. Jing B, Tang W, Huang YF et al (2014) Summary of PHM system standards. J Electron Meas Instrum 28(12):1301–1307
7. Wu B (2009) Research on methods and application of health assessment. Comput Meas Control 17(12):2345–2362
8. Wang L, Lv WM, Li W et al (2013) Status and perspective of health condition assessment for complex system. Comput Meas Control 4:002
9. Yao YF, Wu YF, Feng YG et al (2012) Health condition assessment of equipment. Mod Def Technol 40(5):156–161
10. Yuan WC, Cui F (1999) Current research and development of structural health monitoring and condition assessment for bridges. J Tongji Univ 27(2):184–188
11. Guo T, Li AQ, Li ZX et al (2004) Progress in condition assessment methods for long span bridges. J Southeast Univ 34(5):699–704
12. Yue Q, Zhu LM (2006) Assessment methods of bridge health conditions in bridge health monitoring system for Donghai bridge. Bridge Constr 2006(s2):166–170
13. Zhang JP, Liu GX, Yuan Q et al (2004) Design and realization of evaluation of health condition system for transformation equipment. Mod Electric Power 21(4):45–49
14. Wang YY, Chen BJ (2014) An integrated life estimation model of power transformer based on hierarchical architecture and health index. Power Syst Technol 38(10):2845–2850
15. Lv JW, Yu P, Wei J et al (2011) Assessment method of health conditions of naval vessel system. J Naval Univ Eng 23(3):72–76
16. Yu P, Lv JW, Liu ZH (2010) Research on the health state evaluation and application of warship equipment. China Shiprep 6:47–50
17. Huang JY, Liu X, Bo WH et al (2014) Research on health condition assessment of civil aircraft. Comput Meas Control 22(010):3256–3258
18. Li Y, Chen H, Zhang QJ et al (2011) Assessment method of health condition for UAV systems. Syst Eng Electron 33(3):562–567
19. Cao HL, Huang LT, Kang LP (2015) To study on health assessment of aero-engine based on AHP and GRA. Math Pract Theory 2:122–129
20. Cui JG, Lin ZL, Lv R et al (2014) Comprehensive assessment method of aircraft health status based on fuzzy gray clustering and combination weighting. Acta Aeronaut 35(3):764–772
21. He C, Li M, Li TT et al (2016) Comparison and analysis of the four methods of determining weights in multi-objective comprehensive evaluation. J Hubei Univ 2:172–178
22. Deng X, Li JM, Zeng HJ et al (2012) Research on computation methods of AHP wight vector and its applications. Math Pract Theory 42(7):93–100
23. Chang JE, Jiang TL (2007) Research on the weight of coefficient through analytic hierarchy process. J WUT 29(1):153–156

Chapter 15
Space-Ground TT&C Resources Integrated Scheduling Based on the Hybrid Ant Colony Optimization

Zexi Li, Jing Li and Wenting Mu

15.1 Introductions

With China's regional navigation system began to service, China's GNSS is developing rapidly. Thus, the number of satellites in orbit is growing with great speed and the disadvantages of ground-based TT&C network become impossible to ignore. Inter-satellite links (ISLs) in Global navigation constellation have the ability of complete coverage and high manoeuvrability, which can provide effective support for making up for the deficiencies of the ground-based resources and emergency uses. How to tackle the multi-satellite scheduling problem with considering the increasing constraints and the integrated space-ground TT&C resources becomes difficult [1].

TTCRISP is a kind of large combinatorial optimization problem belonging to the aerospace field, which has been proved to be a NP-complete problem [2]. It specializes how to assign integrated resources (i.e., time windows) to tasks properly over a time period, which plays a vital role in the management of satellite network. A lot of works have already been done on satellite scheduling problems. In general, we may divide the methods used to tackle this kind of problems into three groups:

- deterministic methods based on programming model,
- general heuristic methods based on rules and constraints,
- metaheuristics.

Conventional deterministic methods can rarely find an optimal solution to a given NP-hard problem in reasonable time, and general heuristic methods do not guarantee finding a solution with good quality. So, many researchers have been engaged in metaheuristics, mainly include ant colony optimization (ACO) and genetic algorithm (GA), to solve it by taking the efficiency and performance into account. GA was first

Z. Li (✉) · J. Li · W. Mu
State Key Laboratory of Astronautic Dynamics, Xi'an 710043, China
e-mail: zexi57@163.com

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_15

proposed by Parish et al. [3]. for the satellite TT&C resources scheduling problem to obtain a 24-h scheduling plan of Air Force Satellite Control Network (AFSCN). In order to maximize the number of assigned tasks, Soma et al. [4] designed a GA based software, named IMPACT, for a broad range of satellite scheduling problem instances. Facing a large scale multi-satellite scheduling problem, Li et al. [5] presented the improved GA by designing the proper operators of crossover and mutation, and the numerical computational results showed its good performance. Xing et al. [6] analyzed the features of the multi-satellite observation scheduling problem and proposed a knowledge based ACO method which improved the efficiency of ACO dramatically. Aiming to minimize the working burden as optimization objective, Zhang et al. [7] built a complex independent set model based on visible arcs and working periods, and then an ACO approach combined with a guidance solution based pheromone updating strategy was proposed to solve it. The experimental results demonstrated that the novel approach was competitive. Chen et al. [8] founded two types of solution construction graph based model for satellite data transmission scheduling problem and proposed an ACO method to deal with it, the results verify the feasibility and effectiveness of the algorithms.

In recent years, ACO has been demonstrated in solving hard combinatorial optimization problems effectively. ACO represents an optimization problem by using a construction graph and letting amount of artificial ants walk on the graph. But the basic ACO also has several disadvantages such as trapped into local optimal easily and its poor efficiency [9, 10]. Due to this, the hybrid algorithms which combined ACO with other optimization mechanisms were developed. In the field of combining ACO with GA [11–21], Abbattista [18] first proposed an ACO-GA hybrid algorithm for TSP and achieved competitive results on instance Oliver30TSP and Eilon50TSP. White [19], Pourya [20] and Nemati [21] proposed another kind of ACO-GA hybrid algorithms for a variety of optimization problems in both discrete and continuous domains. Specially, White [19] used GA to optimize the parameters of ACO. Pourya, Nemati ran ACO and GA in parallel in each iteration, and the good solutions were selected as the initial population into the next iteration. As far as we know, there isn't any research on using ACO-GA hybrid algorithms in the field of satellite range scheduling, so two types of ACO-GA algorithms with different fusion mechanisms, that is, static sequential fusion mechanism and dynamic sequential fusion mechanism respectively, were presented.

15.2 Space-Ground TT&C Resources Integrated Scheduling System

In this section, A schematic representation of the process of TTCRISP is give in Fig. 15.1 based on the structure of the conventional TT&C resources scheduling system [22]. The grey parts are needed to be modified due to the newly increased space-based TT&C resources.

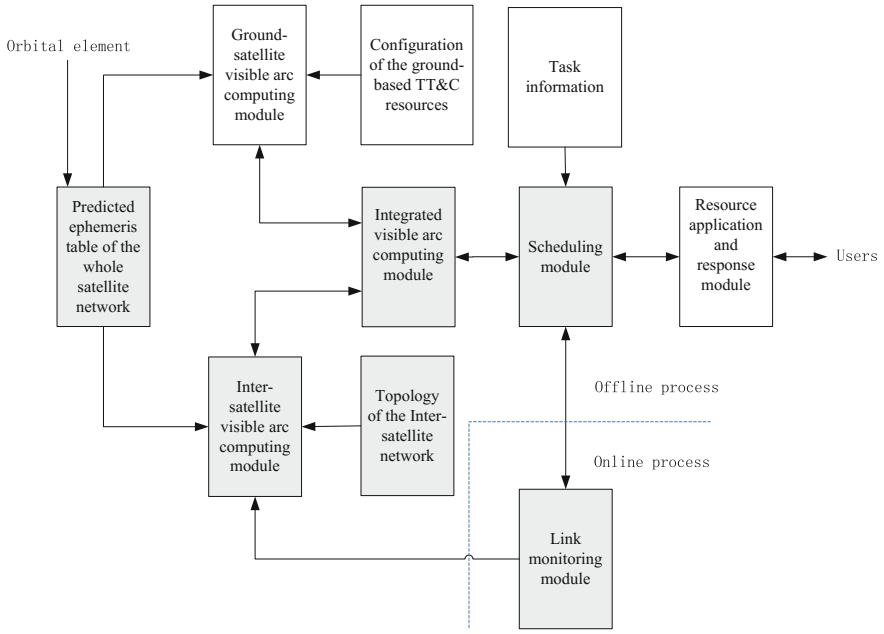


Fig. 15.1 Space-ground TT&C resources integrated scheduling system

Scheduling module is the core of the system. It coordinates the communication between satellites and ground stations so as to keep the orbiting satellites working effectively. The scheduling module responses to serve requests of the satellites with considering the constraints, space-ground integrated resources and task information together, resulting an optimal scheduling plan at last.

Integrated visible arcs are composed of the inter-satellite and ground-satellite visible arcs, which are generated according to the visibility between entities. Inter-satellite visibility depends on both the topology and the predicted ephemeris table of satellite network.

In contrast to the other modules work in an offline manner, link monitoring module works online. It is response for monitoring the state of ISLs. When an ISL breaks, the monitoring module sends a feedback message to the scheduling module, and then the scheduling module relocates the resources to ensure that the follow-up tasks can be served.

15.3 Space-Ground TT&C Resources Integrated Scheduling Model

For the sake of simplify, we make the following assumptions in terms of practical conditions:

- both satellite failure and facility failure are not in consideration,
- dysfunction of satellites and facilities are not in consideration,
- each visible arc cannot be preempted once it have been assigned to the corresponding task,
- each satellite involves only one antenna for establishing ground-satellite link, and every antenna used for establishing ISLs has the same configuration,
- the time delay of multi-hop ISLs is always unacceptable, so we assume that a ground station cannot serve a satellite via more than one relay satellite.

15.3.1 Preliminaries

The notations for elements of TTCRISP in the context are as following:

(1) A set of satellites $S = \{s_1, s_2, \ldots, s_{|S|}\}$, $\forall s_i \in S$, $s_i = (sNo_i, sO_i, sN_i)$, where $|S|$ means the number of satellites needed to TT&C support in the scheduling period, sNo_i is the identifier of satellites, sO_i is the sign of the orbital characteristics, namely, $sO_i \in sO$ = {1,2,3,4} = { early orbit phase, orbiting segment, transfer orbit segment, return segment}, sN_i is the minimal numbers that a satellite needs to be served in the scheduling period.

(2) A set of TT&C equipments $E = \{e_1, e_2, \ldots, e_{|E|}\}$, where $|E|$ means the number of TT&C equipments in the scheduling period.

(3) For each $(s_i, e_j) \in S \times E$, the visibility period can be expressed as a set $A(s_i, e_j) = \{a_1, \ldots, a_{|A(s_i,e_j)|}\}$ of disjoint visible arcs which are made up of direct visible arcs (i.e., corresponding to ground-based resources) and indirect visible arcs (i.e., corresponding to space-based resources). Each visible arc a_i is characterized by the following parameters:

- aS_i, the (unique) satellite establishing communication in a_i;
- aE_i, the (unique) ground station establishing communication in a_i;
- aJ_i, the service which can be processed with aS_i;
- $[aS_i^s, aW_i^e]$, the time window of a_i;
- $relay\_sat_i$, the relay satellite. If it is the same as aS_i, aS_i is a direct visible arc (i.e., aE_i applies service to aS_i directly), otherwise an indirect visible arc (i.e., indirectly service via $relay\_sat_i$ is executed.);
- aPr_i, the profit of a_i.

(4) A set of tasks $J = \{J_1, J_2, \ldots, J_{|J|}\}$ have to be scheduled for the requirement of satellite control operations. Each task J_i specifies a minimal communication duration $tlast_i$ and can only be processed in a given time window $JWin_i$. Each task J_i belongs to a unique satellite JS_i. Once a task is assigned with a legal visible arc a_j, the corresponding profit aPr_i will be got. Moreover, the interval of any two successive services of each satellite must be between the minimum and the maximum interval time, denoted as $t_{\max}, t_{\min}$, respectively.

(5) The decision variable of the TTCRISP is $V = \{x_i, t_j^s, t_j^e\}, i = 1, \ldots, |A|$, where $|A|$ is the number of the visible arcs in the scheduling period; x_i is a binary decision variable, $x_i = 1$ if and only if the visible arc is assigned to a corresponding task J_j; t_j^s, t_j^e denote the actual start time and end time of executing task J_j, respectively.

15.3.2 Constraints

During the process of determining whether a task is scheduled, a set of domain constraints must be satisfied. In general, they are temporal constraints, resource constraints, ground facility exclusivity constraints, quality of service constraints and mission requirements, which are formalized as follows [23–27]:

(1) Temporal constraints

Temporal constraints is a pair of fixed metric times related to a visible arc, only during which period the visible arc is available and scheduled. For example, if a visible arc a_i is assigned to a proper task J_j, the succeeding constraints must be satisfied:

$$C_1 = \{x | \forall a_i \in A, \exists J_j \in J, aS_i = JS_j, aW_i^s \le t_j^s \le t_j^e \le aW_i^e, if\ x_i = 1\} \quad (15.1)$$

(2) Resource constraints

Compared with ISLs (i.e., space-based resources), ground-satellite links (i.e., ground-based resources) have higher reliability, stability and accuracy when transmitting data. Due to this, the space-based resources are only used in the orbiting segment during which segment the satellite is stable. In summary, the resource constraints can be expressed as follows:

$$C_2 = \{x | \forall a_i \in A, sO(aS_i) = 2, if\ (x_i = 1) \wedge (aS_i \neq relay\_sat_i)\} \quad (15.2)$$

(3) Ground facility exclusivity constraints

Each ground facility can serve only one satellite at a time, so any two time-overlapping visible arcs which belong to the same ground station conflict with each other. The ground facility exclusivity constraints can be expressed as follows:

$$C_3 = \{x|\forall a_i, a_j \in A, x_i + x_j \leq 1, if(aE_i = aE_j) \wedge ([aW_i^s, aW_i^e] \cap [aW_j^s, aW_j^e] \neq \varphi)\} \tag{15.3}$$

(4) Mission requirements

The ground stations must serve the orbiting satellites several times a day to receive information from or transmit commands to the antennas in order to keep satellites working effectively. Each service may satisfy the shortest duration, maximal and minimal interval time constraints. The tasks that the satellite s_k required are denoted as $sJ_k = \{J_i|\forall J_i \in J, JS_i = s_k\}$.

Service number constraints is as follows:

$$C_4^1 = \{x|\forall s_k \in S, |sJ_k| \geq sN_k\} \tag{15.4}$$

Serve time constraints is as follows:

$$\begin{aligned} C_4^2 = \{x|\forall J_m \in sJ_k, \exists a_i \in A, t_{\min} \leq t_n^s - t_m^e \leq t_{\max}, t_m^e - t_m^s \geq tlast_m, \\ t_n^e - t_n^s \geq tlast_n, if(x_i = 1) \wedge (aS_i = JS_m) \wedge \\ (n = \min\{n'|\exists J_{n'} \in sJ_k, \exists a_j \in A, \\ x_j = 1, aS_j = JS_{n'}, n' - m > 0\})\} \end{aligned} \tag{15.5}$$

(5) Quality of service constraints

In order to improve the accuracy of orbit determination, the number of ground stations which provide TT&C services must be as many as possible; the distance between ground station and satellite must be less than 49,000 km. Every kind of resource utilization strategies must subject to management. Quality of service constraints is marked as C_5.

(6) Task execution constraints

Each task can be executed at most once. Task execution constraints can be expressed as follows:

$$C_6 = \{x|\forall a_i, a_j \in A, aJ_i \neq aJ_j, if(x_i = 1) \wedge (x_j = 1)\} \tag{15.6}$$

15.3.3 Scheduling Model

The TTCRISP aims to find a feasible schedule plan Ω in which the visible arcs are assigned to as many preferential tasks as possible such that the total received profits F are maximized. Hence, the TTCRISP can be stated as follows:

$$\max \quad F(X) = \sum_{|A|} x_i \cdot aPr_i \tag{15.7}$$

$$s.t. \quad X = (x_1, x_2, \ldots, x_{|A|}); \tag{15.8}$$

$$x_i \in \{0, 1\}, \forall a_i \in A; \tag{15.9}$$

$$X \in C_1 \cap C_2 \cap C_3 \cap C_4^1 \cap C_4^2 \cap C_5 \cap C_6 \tag{15.10}$$

Formulation (15.7) states the objective function. Constraints (15.8) and (15.9) impose the restriction on decision variables. Constraints (15.10) are intersection of the domain constraints.

15.4 GA-ACO Hybrid Algorithm

One of the issues that probabilistic optimization algorithms might face in solving global highly non-convex optimization problem is premature convergence. One of the causes of premature convergence is the lack of diversity, the other reason is related to their lack of advanced search capability around the global solution. Several studies have shown that incorporating some knowledge about the search space can improve the search capability of evolutionary algorithms (EAs) significantly. In order to eliminate the shortcomings and the insufficient robustness of the global search ability of the ACO algorithm, an hybrid algorithm which combines ACO with GA is proposed in this paper for two reasons:

(1) It costs so much time for basic ACO to initialize and update pheromone. In the new method, part of the updating work is handled by GA in advance. This causes GA generally affects the searching process since it can accelerate the low optimization efficiency and quickly find the near-optimal solution.
(2) The performance of hybrid algorithms which combined with a variety of optimization mechanism is better than any of the original algorithms which only use a single optimization mechanism.

15.4.1 Algorithm Framework and Hybrid Strategy

The flowchart of the hybridization plan is illustrated in Fig. 15.2. Due to the advantages of ACO, GA is performed to generate feasible solutions for initialization and updating pheromone until the distribution of pheromone is up to a standard, followed by the further optimization and getting a near-optimal solution by ACO since it has the characteristics of positive feedback, distributed computation and constructive greediness for example.

The optimization performance comparison of GA and ACO is shown in Fig. 15.3. The lack of pheromones in stage $[t_0, t_a]$, makes searching speed slow. Only with pheromone accumulation, can accelerate solving process. While GA could find acceptable solution after short period, the overall precision of solution is relatively low. If invited at the static point t_b, ACO would slow down the searching speed due to the lack of pheromone; if invited at the static point t_c, GA would slow down the searching speed due to its late phase slow speed disadvantage. Therefore, the key to optimal solution is to find the dynamic optimal combining point t_a of two algorithms. The steps to determine the dynamic combining point are as below:

(1) Set the lower and upper bound of GA iterations (such as t_b and t_c);
(2) Count offspring population evolution rate in the iterative process of GA, and set minimum offspring evolution rate accordingly;
(3) In the specified range of the number of iterations, if offspring population evolution rate is less than the minimum evolution rate for N generations, we think the efficiency of GA is lower. Then terminate process of genetic algorithm and enter the stage of ant colony algorithm.

15.4.2 Pheromone Initialization Process Based on the Genetic Algorithm

To use genetic algorithm, the three key problems are tackled as follows:

15.4.2.1 Coding and Population Initialization

Chromosome coding is the foundation of GA. A design of genetic coding requires completeness, integrity and non-redundancy. The paper takes chromosome binary code as decision variable and encodes all the visible arcs as binary code.

Individuals in initial population are generated randomly. But the search performance is closely related to initial population diversity, so initial population is generated from excessive random population and then sampling by certain distribution function.

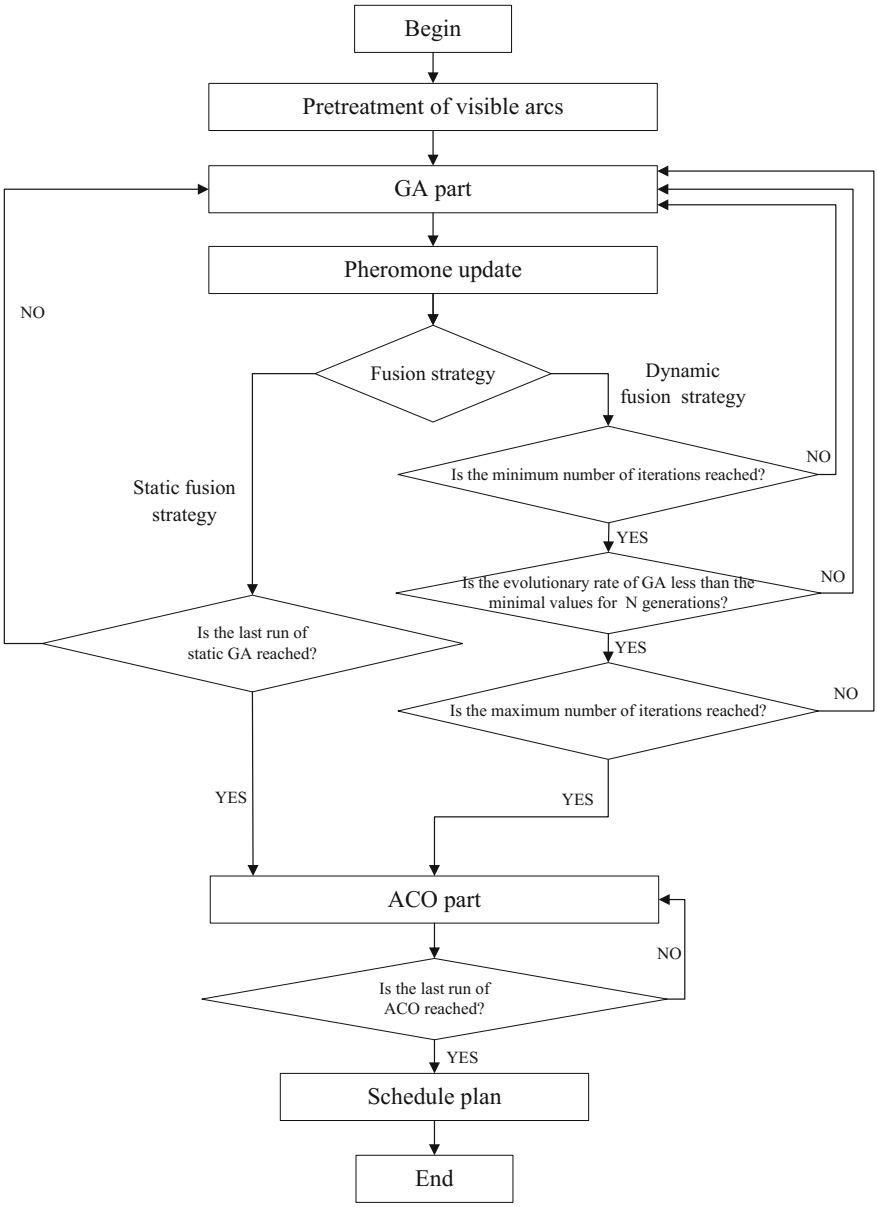


Fig. 15.2 The flowchart of GA-ACO

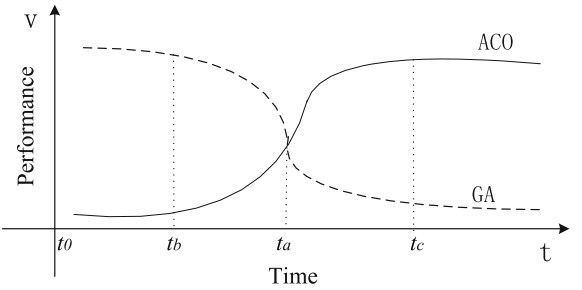


Fig. 15.3 Comparison of solving speed-time

15.4.2.2 Genetic Component Design

In consideration of coding method and this specific issue of space-ground TT&C resources integrated scheduling, this article applies single-point crossover operator and multi-point random mutation operator. Fit value functions bridges optimization problem and genetic algorithm. The quality of corresponding feasible solution is evaluated by calculating the fit value. Design of fit value function should be geared to the needs of specific issues. In this paper, with reference to the optimization target given by space-ground TT&C resource integrated scheduling model, takes the earning of task scheduling as the fit value function of GA. GA simulates the population evolution phenomenon of "survival of the fittest" by selecting the operator. It has effect of reducing the population diversity monotonously and guarantees the global convergence of GA. With considering the balance between the performance and the efficiency, this article designs a hybrid selective factor based on region partition to make the population diversity. All the chromosomes are partitioned and sorted by its fitness at first, then a hybrid selective factor combined with a roulette strategy and a elite presentation strategy is used in each partition. The selection phase is as follows:

(1) Compute the individual fit value of generation k, $f(X_i^{mid}(k))i = 1, 2, \ldots, 2 * N$(N is population quantity) and sort by descending order;
(2) Divide the sorted individual fit value into m regions and sort these m regions in descending order of their individual fit value.
(3) In every set of these m regions, according to the proportion of fit value, apply roulette wheel selection method to pick $\frac{N}{m}$ individuals independently and randomly to compose a new population $X^{new}(k)$. To ensure that optimal solution of old population is preserved, elite preservation strategy applies.

In the roulette strategy, the probability of selecting a individual is proportional to its fitness. Specially, calculate the fitness of each individual and generate a random number $\xi \in [0, 1]$ at first. Then, if the following formula is satisfied, the individual x_i will be chosen.

$$\frac{\sum_{j=1}^{i-1} f_j}{\sum_{j=1}^{N} f_j} < \xi \leq \frac{\sum_{j=1}^{i} f_j}{\sum_{j=1}^{N} f_j} \tag{15.11}$$

where f_i is the fitness of the individual x_i.

15.4.2.3 Decoding

Decoding refers to allocating visible arc with gene value true to corresponding task in a fixed order to generate scheduling plan.

15.4.3 The Solving Process of ACO

Provided the initial solution set from GA, ACO continues the solving process. The graph based space-ground integrated scheduling ACO consists of three key steps:

15.4.3.1 Construction Graph

This paper adopted the idea of Gutjahr, he presented a construction graph based ACO framework which takes $C = <V, E>$ for searching space of ACO, and function Φ for mapping the route of ants in the directed graph to the set of feasible solution.

The rule of constructing the visible arc timing graph is as follows: The visible arcs are abstracted as nodes and each node corresponds to a Boolean decision variable which represents whether the visible arc is chosen. When two nodes have no TT&C conflict (i.e. requiring services from the same ground equipment), connect them with directed line which points from early starting arc to late starting arc. The rule ensures the traversing of ants. Each visible arc subset in $ArcS = \{as_1, as_2, \ldots, as_{|ArcS|}\}$ transforms into a sub graph.

15.4.3.2 Solution Construction

It takes two steps to construct a solution. First, $|ArcS|$ groups of ants walk on the $|ArcS|$ directed graphs independently. Then link the subset together to generate scheduling plan. In the process of traversing, ants choose next node according to the pseudo-random state transition rule expressed in formula (15.12). When ant takes step from node a_i to a_j in step k, probability can be calculated by formula (15.12):

$$P_{ij}(k) = \begin{cases} \frac{[\tau_{ij}(k)]^{\alpha} \cdot [\eta_{ij}(k)]^{\beta}}{\sum_{u \in J(i)} [\tau_{iu}(k)]^{\alpha} [\eta_{iu}(k)]^{\beta}}, j \in J(i) \\ 0, otherwise \end{cases} \tag{15.12}$$

where τ_{ij} is pheromone trail; η_{ij} is heuristic information; α and β are the weight of pheromone trail and heuristic information; $J(i)$ is the set of nodes which can be reached directly from a_i and is not visited by ants.

From the experience of pragmatic situation, scheduling prefers to concentrate TT&C tasks into a relatively short time span in each divided span. For efficiency, that is, to pursue the max parallel degree, so the heuristic rule designs as:

$$\eta_{ij} = 1 / \left| aW_i^s - aW_j^s \right| \tag{15.13}$$

Formula (15.13) denotes that ant would prefer node which is nearest to the node that has TT&C service already.

15.4.3.3 Pheromone Trail Update

The distribution of pheromone will finally determine the quality of solution. This paper designs a local update strategy,which is based on elitist ants, for pheromone to lead the ant colony to search and find the optimal solution. Set the i iterations of paths as $\{T_{\text{lg}}^1(i), T_{\text{lg}}^2(i), \ldots, T_{\text{lg}}^m(i)\}$, and arrange the corresponding objective function values according to the order from ideal to non-acceptable. Then the updating mechanism of local update strategy is as below:

$$\tau_{ij} = (1 - \rho) \cdot \tau_{ij} + \sum_{k=1}^{|em|} \Delta\tau_{ij}^k \tag{15.14}$$

$$\Delta\tau_{ij}^k = \begin{cases} (C_{\text{lg}}^k)^{-1}, & (i,j) \text{ is on path } T_{\text{lg}}^k \\ 0, & otherwise \end{cases}$$

where $|em|$ is the number of elitist ants, ρ is pheromone evaporation rate.

15.5 Simulation Example and Result Analysis

15.5.1 Scene Design

Compass GNSS consists of 30 satellites: 24 MEO satellites which form a Walker24/3/1 constellation, 3 Geosynchronous Earth Orbit (GEO) satellites, and 3 Inclined Geosynchronous Satellite Orbit (IGSO) satellites, the layout of

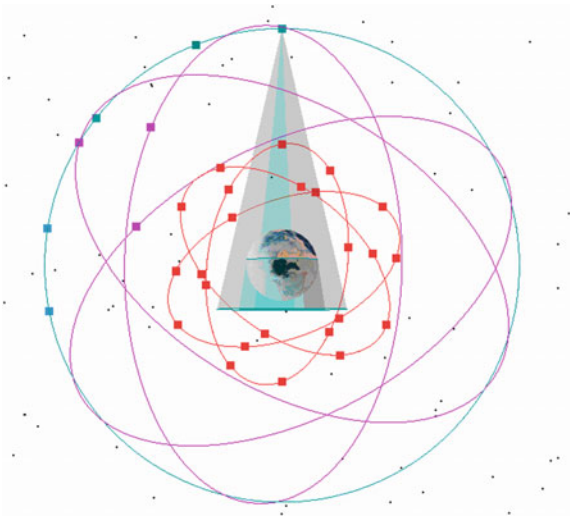


Fig. 15.4 Layout of compass GNSS

Table 15.1 The profit of resources

The resource type	Profit
Satellite-ground visible arc	1.00
Inter-orbit plane inter-satellite visible arc	0.95
Intra-orbit plane inter-satellite visible arc	0.90

Compass GNSS is shown in Fig. 15.4, red, violet, and blue squares represent the MEO satellites, IGSO satellites and GEO satellites respectively.

We consider two instances designed for the orbit determination services scheduling of the MEO satellites in Compass GNSS. The basic information is given in Table 15.2. To meet the accuracy requirement of orbit determination, a MEO satellite must be served 4 times which are evenly distributed throughout a day. For this reason, a scheduling period is always partitioned into four equal time slots and the visible arcs are often divided into four clusters. Without loss of generality, a long visible arc of a MEO satellite is divided into short arcs by the standard of the minimum duration of all services in order to make well use of these visible arcs. Moreover, all short visible arcs are divided into 4 clusters corresponding 4 time slots. In summary, within each time slot, every satellite needs to be served with not less than 30 min duration. The profit of visible arcs are shown in Table 15.1.

Table 15.2 Scheduling scene configuration

Problem instance scale	# satellites	Scheduling period	TT&C resources	# visible arcs	# ground stations	Design purpose
Small	24	2012/7/1–2	Ground-based TT&C resources	1773	4	It is used to verify the feasibility and effectiveness of the algorithm
Large	24	2012/7/1–2	Space-ground TT&C resources	8865	4	It is used to verify the robustness of the algorithm

Table 15.3 Parameter settings of GA-ACO

Symbol	Meaning	Value
N_{GA}	Maximum number of iterations of GA	$N_{GA} = \{50, 100, 150, 200, 250, 300\}$
N_{ACO}	Maximum number of iterations of ACO	$N_{ACO} = 300$
N_{chrom}	Population size of GA	$N_{chrom} = 50$
N_{ants}	Population size of ACO	$N_{ants} = 20$
GA_last	Maximum stagnant generations	$GA\_last = 30$
P_C	Crossover probability	$P_C = 0.65$
P_m	Mutation probability	$P_m = 0.08$
ρ	Pheromone evaporation ratio	$\rho = 0.3$
α	Weight of pheromone trail	$\alpha = 1$
β	Weight of heuristic information	$\beta = 1/3$

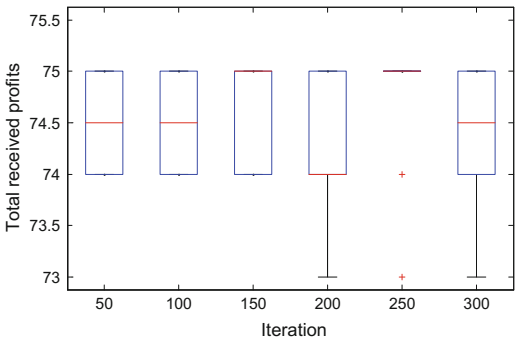
15.5.2 Parameter Settings

The parameter settings of the hybrid GA-ACO are shown in Table 15.3.

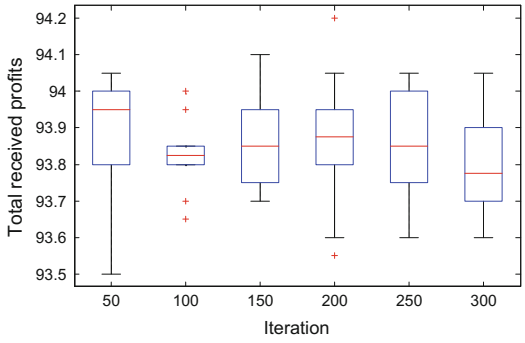
15.5.3 Simulation Result

To find the best combining point of GA-ACO, it is necessary to determine the proper iteration range of GA. Therefore, firstly, we study the effect of iteration number of GA on optimization performance of hybrid algorithm. To find out the best combining point, only the parameter N_{GA} is varied while others are fixed. For each instances, our algorithm was executed 10 times independently.

The effect of iteration number of GA in the hybrid algorithms is shown in Fig. 15.5. As seen from Fig. 15.5, for the instance of ground-based TT&C resources integrated scheduling, the median and the maximum value are the largest and the dispersion degree of the scheduling results is small when $N_{GA} = 150$. Due to this, the lower and the upper bound of the GA iterations are set as 50 and 200,



(a) Instance of ground-based TT&C resources scheduling



(b) Instance of space-ground TT&C resources integrated scheduling

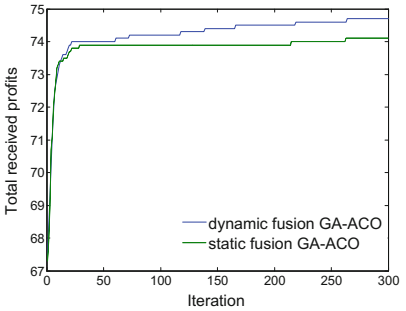
Fig. 15.5 Computational results obtained when different number of iterations are tested

respectively. For the other instance, the median and the maximum value are both the best when $N_{GA} = 150, 200, 250$, so the lower and the upper bound of the GA iterations of space-ground instance are set as 150 and 250, respectively. The GA iterations in static fusion GA-ACO for the two instances are set as 150.

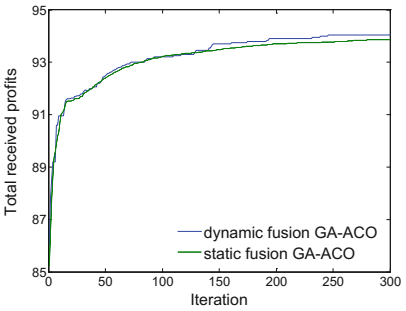
We implemented four algorithms to check the performance of the hybrid ones. The four algorithms are GA, ACO, static fusion GA-ACO, dynamic fusion GA-ACO, respectively. Figure 15.6a, b shows the comparison results of the two fusion strategies obtained for two instances. We can observe that the dynamic fusion GA-ACO was able to find all best results of average solution values for both the ground-based and the space-ground instances. The best average result for the ground-based instance is 74.7, while for space-ground instance is 94.05.

Figure 15.6c, d presents the comparison results obtained for dynamic fusion GA-ACO and GA for two instances. With respect to the average value, dynamic fusion GA-ACO is better than GA on all instances. We can observe that although the performance of GA-ACO was not as good as GA in the early optimization stage, but GA was trapped into local optimal with the increase of iterations.

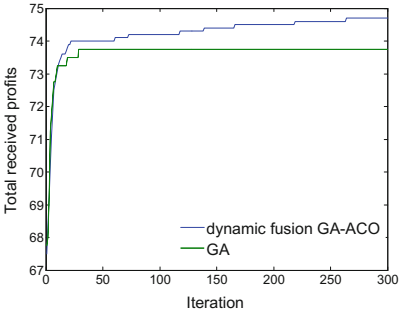
As shown in Fig. 15.6e, f, the performance of dynamic fusion GA-ACO is competitive compared with ACO.



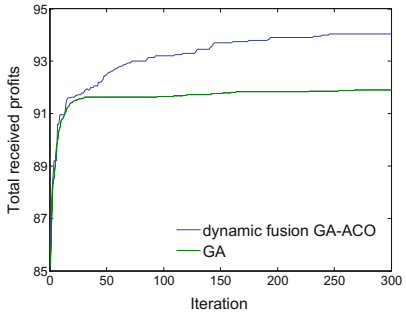
(a) Comparison between GA-ACO using static fusion and dynamic fusion strategy on ground-based instance



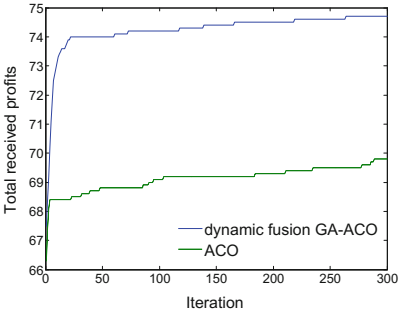
(b) Comparison between GA-ACO using static fusion and dynamic fusion strategy on ground-space instance



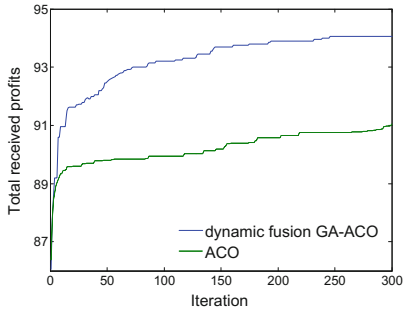
(c) Comparison between GA and GA-ACO using dynamic fusion strategy on ground-based instance



(d) Comparison between GA and GA-ACO using dynamic fusion strategy on ground-space



(e) Comparison between ACO and GA-ACO using dynamic fusion strategy on ground-based instance



(f) Comparison between ACO and GA-ACO using dynamic fusion strategy on ground-space instance

Fig. 15.6 Optimizing performance comparison of dynamic fusion of GA-ACO, static fusion of GA-ACO, ACO and GA

In summary, dynamic fusion GA-ACO can combine the advantages of GA and ACO effectively, the performance and the efficiency of the hybrid algorithm is better than any of the algorithm with single optimization mechanism.

15.6 Conclusions

Motivated by a real-life application in China Satellite Control Center, this paper considers a space-ground TT&C resources integrated scheduling problem. Two hybrid algorithms which combine ACO with GA were proposed to tackle this combinatorial optimization problem. The computational results obtained by four proposed heuristics show that all the algorithms were able to get solutions with good quality in reasonable time for the TTCRISP and the dynamic fusion GA-ACO outperformed the others in this paper.

TT&C resource scheduling problem in dynamic uncertain environment and with multiple objectives are the further research directions, and combining ACO with other optimization mechanisms will be continuously explored.

Acknowledgements Thanks are due to the State Key Laboratory of Astronautic Dynamics for their supporting this work and International GNSS monitoring and assessment system (iGMAS) for providing the observation data (www.igmas.org). This work is supported by National 973 Program of China (NO.613237201506), National 863 Program of China (NO.2014AA7013035) and the Open Research Fund of the State Key Laboratory of Astronautic Dynamics under Grant 2013GKJ11-ADL.

References

1. Yu ZJ (2006) Status quo and development of spaceflight TT&C systems. Eng Sci 8(10): 42–46
2. Chen F (2010) Research on genetic algorithm for multi-satellite TT&C scheduling problem. National University of Defense Technology, Changsha
3. Parish DA (1994) A genetic algorithm approach to automating satellite range scheduling. Air Force Inst of Tech, Wright-Patterson AFB OH, School of Engineering
4. Soma P, Venkateswarlu S, Santhalakshmi S, et al (2004) Multi-satellite scheduling using genetic algorithms. ISTRAC/ISRO, SpaceOps
5. Li YQ, Wang RX, Xu MQ (2012) An improved genetic algorithm for a class of multi-resource range scheduling problem. J Astronaut 33(1):85–90
6. Xing LN, Rohlfshagen P, Chen Y et al (2010) An evolutionary approach to the multidepot capacitated arc routing problem. IEEE Trans Evol Comput 14(3):356–374
7. Zhang N, Feng ZR, Ke LJ (2011) Guidance-solution based ant colony optimization for satellite control resource scheduling problem. Appl Intell 35(3):436–444
8. Chen XG, Wu X (2009) Ant colony algorithm for satellite data transmission scheduling problem. J Syst Eng 4:011
9. Dorigo M, Maniezzo V, Colorni V (1996) Ant system: optimization by a colony of cooperating agents. IEEE Trans Syst Man Cybern B 26(1):29–41
10. Dorigo M (1992) Optimization, learning and natural algorithms. Politecnico diMilano, Milan
11. Wu YM, Xu CF (2010) Genetic ant colony algorithm for job-shop scheduling problem. Appl Res Comput 27(9):3247–3250
12. Wang L, Zheng DZ (2002) A modified genetic algorithm for job shop scheduling. Adv Manuf Technol 20(1):72–76
13. Irina C, Elias K (2012) Hybrid ant colony-genetic algorithm (GAAPI) for global continuous optimization. IEEE Trans Syst Man Cybern B 42(1):234–245

14. Silva CA, Faria JM, Abrantes P (2005) Concrete delivery using a combination of GA and ACO. In: Proceedings of the 44th IEEE conference on decision and control, and the European control conference, pp. 129–134
15. Acan A (2002) GAACO: a GA + ACO hybrid for faster and better search capability. In: Proceedings of 3rd international conference on ant algorithm, pp. 205–212
16. Liu B, Meng P (2008) Hybrid algorithm combining ant colony algorithm with genetic algorithm for continuous domain. In: Proceedings of the 9th international conference for young computer scientists, pp. 1064–1072
17. Zhang DZ, Du LN (2011) Hybrid ant colony optimization based on genetic algorithm for container loading problem. In: Proceedings of the international conference of soft computing and pattern recognition, pp. 890–900
18. Abbattista F, Abbattista N, Caponetti L (1995) An evolutionary and cooperative agents model for optimization. In: Proceedings of the evolutionary computation, pp. 215–224
19. White T, Pagurek B, Oppacher F (1998) ASGA: improving the ant system by integration with genetic algorithms. In: Proceedings of the 3rd annual conference genetic programming, pp. 1069–1075
20. Pourya H, Mahrokh GS (2006) Hybrid ant colony optimization, genetic algorithm, and simulated annealing for image contrast enhancement. In: Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, pp. 755–761
21. Nemati S, Basiri ME, Ghasem-Aghaee N et al (2009) A novel ACO-GA hybrid algorithm for feature selection in protein function prediction. J Expert Syst Appl 36(10):12086–12094
22. Chen F, Wu XY (2010) Space and ground TT&C resource integrated scheduling model. J Astronaut 31(5):1405–1412
23. Barbulescu L, Watson JP, Whitley LD et al (2004) Scheduling space-ground communications for the air force satellite control network. J Sched 7(1):7–34
24. Lee ZJ, Su SF, Liu KH (2008) Genetic algorithm with ant colony optimization for multiple sequence alignment. J Appl Soft Comput 8(1):55–78
25. Li YQ, Wang RX, Xu MQ (2012) An improved genetic algorithm for a class of multi-resource range scheduling problem. J Astronaut 33(1):85–90
26. Gooley TD (1993) Automating the satellite range scheduling process. Air Force Institute of Technology, Ohio
27. Zhang N, Ke LJ, Feng ZR (2009) A new model for satellite TT&C resource scheduling and solution algorithm. J Astronaut 30(5):2141–2145

Chapter 16
Design and Realization of the Three Layers Telemetry Data Transfer Software Frame

Guanghui Ren, Xiangyu He, Shuangcheng Gao and Xin Zhang

16.1 Introduction

Data Transfer Equipment (DTE) is the data interface of telemetry equipments and the center data disposal terminal [1]. In recent years, With the promotion of Consultative Committee for Space Data Systems (CCSDS) standard, telemetry data formats of spacecraft change continually and high density and normalization tasks make DTE need to reduce workload of software maintenance and shorten preparation time as far as possible [2–4]. As the hub of telemetry system data, DTE gradually is developing towards network, modularization, abstraction and generalization. Therefore, many experts in the related areas do a lot of research: Haijiang et al. [5] analyze feasibility of IP multicast technology used in the telemetry data transfer deeply; André Weiskopf, Frank Weichert describes a system for data acquisition and remote maintenance via wireless communication [6]; Wang Gang, Guo Libing, Li Ru and Beijing telemetry technology institute develop respective DTE software using plug-in or IceGrid technology [7–9]; Conceptions of "data source", "data channel" and "selection parameter table" are proposed by Bian Changwan [10] who designs telemetry data and disposal method abstractly; Zhang Caiyue, Chen Liang and other people [11] expound the idea of abstraction design in the data transmission system based on database and implement it.

Based on above achievements, a new and highly abstract "three layers data" transfer software frame is designed in this paper. Data received from outside are all abstracted as "data pool"; methods, processes and results of data disposal are abstracted as "data table"; interfaces of data towards outside and data storage are abstracted as "data channel". Through "data pool"-"data table"-"data channel" three layers data which are encapsulated independently and work collectively, the DTE software has higher generality, flexibility and adaptability. Meanwhile, the

G. Ren (✉) · X. He · S. Gao · X. Zhang
PLA Unit 63819, Yibin 644000, China
e-mail: gh.ren@qq.com

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_16

telemetry DTE software has been implemented and applied in the homebred Kylin operation system based on this frame. Looking from the application effect, the new DTE software shortens the time of task software maintenance and improves the visual monitoring ability of real-time data which produces obvious benefits.

16.2 Functions and Features of DTE

The main functions of DTE are: in the process of preparing a task and the process of tracking and measuring spacecraft, all telemetry data are real-time disposed, selected, transmitted and saved and other data are received, disposed and sent as the task configuration.

DTE sub-system needs to transfer information between the inside system and the outside disposal display terminal. They have several features:

- A wide variety of receiving data: there are above ten kinds of data in the inside system and the outside disposal display terminal and some new kinds of data may be added along with task functions expanded.
- Numerous methods of data disposal: a mass of data need to be disposed before sending; disposal methods are various and change as the different of species of receiving data.
- Manifold sending data: the number of base band telemetry flow, the disposal display number of sending towards outside, the frame frequency and the format of framing change with the needs of different tasks.

From the above features, it is known that the highly abstract and low coupling software frame must be designed if generality, flexibility and adaptation of DTE are improved. On the basis of meeting current demands, the software should be abstract and decoupling as much as possible to meet possible changes and demands in the future.

16.3 The Original Design Frame of Telemetry Transfer Software

16.3.1 The General Frame of Telemetry Transfer Software

Most DTE software are written by object-oriented C++ or C#. A general method of telemetry transfer data disposal process is described in Fig. 16.1 and it is a natural train of thought with reception, disposal and sending data. For example: data received from network are classified according to the data format and framed according to different data types, disposed data are sent to destinations through the

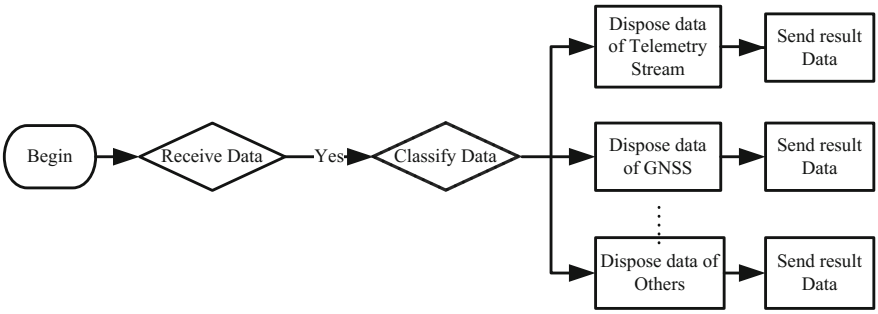


Fig. 16.1 Disposal flow diagram of DTE using general method

network. Although this kind of software is simple, efficient and has a strong understandability, there are many problems in aspects of flexibility and adaptation.

- Poor adaptability and insufficient abstraction when the software is designed

The flexibility and adaptation of software depend on abstract degree. Although the software is written using the object-oriented languages, the abstract degree of data and the data disposal are not enough [5]. Common problems in the data disposal process are not extracted which is a disposal of process in essence. Once there are large demand changes occurring (such as the number of base band changes, destination of sending data increases, format and standard of transmitting information change and the disposal method changes), designers need to amend the software.

- There are heavy workload and high difficulty when maintaining software

Because of poor adaptation, when facing on ordinary demands change, the software could not work using visual configuration methods and code must be modified. When facing on big demands change, the software frame and disposal process need to be modified which lead to heavy workload and high standard to maintainer. It is difficult to establish task state quickly in the situation of high density task.

16.3.2 The Frame of Telemetry Transfer Software Based on "Data Source"

Conceptions of "data source", "communication channel" "parameter disposal method" and "disposal parameter table" are proposed by Bian Changwan who designs telemetry data and disposal method abstractly on the basis of summarizing many kinds of DTE software. A new thought of generalization and modularization of data is proposed. The frame of transfer software is described in Fig. 16.2 based on this thought.

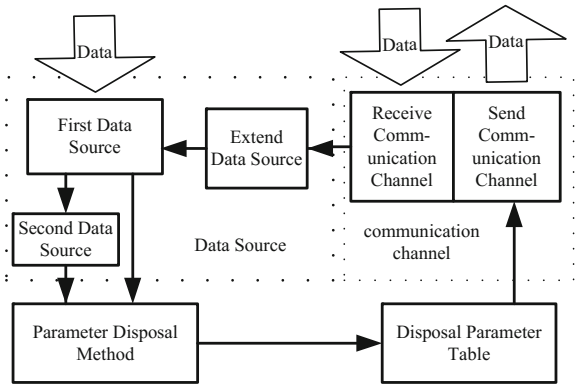


Fig. 16.2 Software frame of DTE based on "data source"

"Data source" in the frame is a logic conception which indicates any data inputted into the module of software disposal transmission. All "data source" are considered as a (n, m) planar data block. The new data block which has been disposed can be used by other software modules. "Communication channel" is an approach through which data are transmitted and receiving and sending any data are through "communication channel". "Parameter disposal method" is an abstraction of different data disposal methods. "Disposal parameter table" includes "parameter disposal method" and "communication channel" which takes different kinds of parameter disposal results to frame as configurations and sends framing data through "communication channel".

This frame is an abstraction of telemetry data and disposal methods. The adaptation of software is improved obviously, the workload of software maintenance is reduced and the efficiency of maintenance is improved. But there are some shortcomings:

- Part of data are abstracted irrationally and some couplings exist. "Data source" and "communication channel" overlap in the function. In order to reduce codes repetition, the conception of "extended data source" is proposed in the "data source" which leads to couplings of software structures and goes against the modularization of software.
- The conception of "data source" is divided too small which leads to complicated structures and difficult codes. In order to solve the problem of single data source existing several kinds of data, two layers data are used in the frame to guarantee the oneness of data in the data source. Although complexity of structures and difficulty of codes are improved, efficiency is improved limitedly.
- Part of data are abstracted halfway and the design thought that data is center isn't carried out completely. Data sending and saving are based on communication channels in the frame which leads to some inconveniences such as data types difficult to identify in the storage and can't control data directly when sending data.

16.4 Software Frame Design

16.4.1 Software Frame Design of "Three Layers Data"

Aiming at problems of the current design frame, task demands and future development demands, the current software frame of "data source", "communication channel", "parameter disposal method" and "disposal parameter table" is analyzed, reconstructed, improved and optimized. The conception of three layer data of "data pool", "data table" and "data channel" is formed. A new highly abstract transfer software frame is proposed as shown in Fig. 16.3.

The design thought is: the data pool receives DTE external data and pre-disposes data through different methods. Then uniform format data are formed. Pre-disposed data and external data in the transfer system are disposed as target data through different data disposal methods. These target data after combination form the "data table" which accomplishes different kinds of framing work of sending data. Finally, data channels and data tables establish corresponding relations and data tables are sent to target addresses through the grouping method.

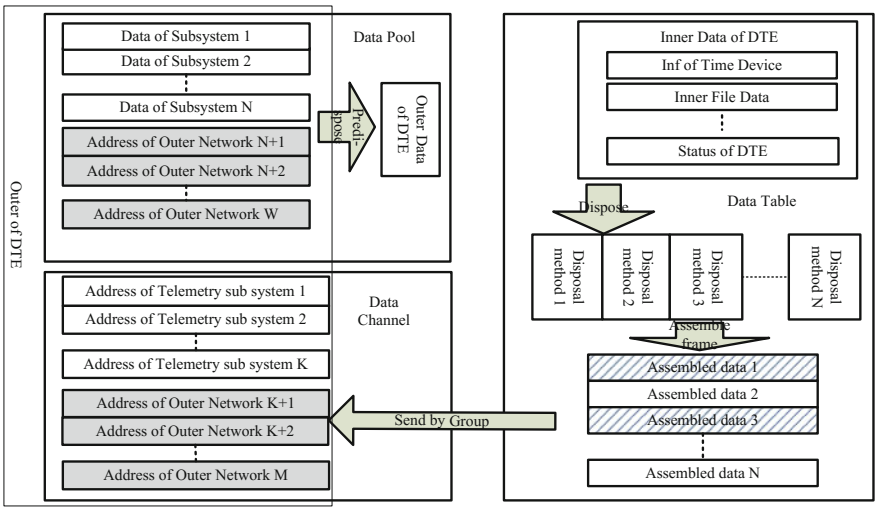


Fig. 16.3 Design diagram of three layers data software frame

16.4.2 Detailed Disposal Flow of "Three Layers Data" Frame

The detailed disposal flow design of "three layers data" frame is shown in Fig. 16.4.

- According to current task demands, data pool objects K, R and M are created. When objects K and M receive corresponding data, the related parameter disposal objects K_1, K_2……K_n and M_1, M_2……M_n are disposed. Objects K_n and M_n have the function of control data tables sending;

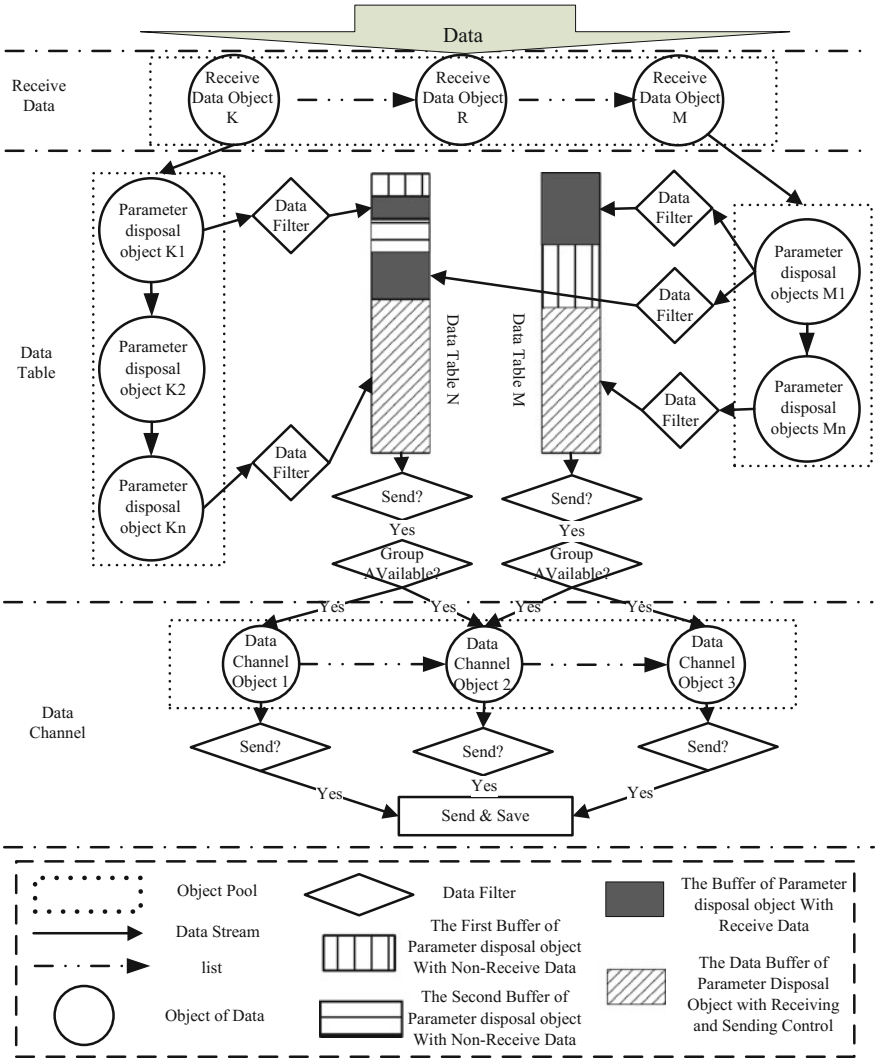


Fig. 16.4 Specific dispose flow diagram based on "three layers data" frame

- These data are filtered and then disposed, the results are laid in the corresponding data buffer of selection parameter table.
- After parameter disposal objects with the function of control data tables sending are disposed, parameter disposal objects which are not related with non-data pool are disposed and form “data table”.
- When data tables can be sent and groups are opened, data can be sent through corresponding data sending objects.

In addition, for timing sending data, after selection data tables gather data, data are sent through data channels using timing intermit.

16.4.3 Advantages of “Three Layers Data” Frame

16.4.3.1 Reasonable Data Abstraction, Reduce the Coupling of Different Links

As shown in Fig. 16.5, external data are abstracted, concluded and form “data pool”. The communication channel forms “data channel”.

In the new frame, on the one hand, objects of “data pool” and “data channel” can be created independently which don’t overlap in the function of receiving data. On the other hand, the key “data table” can flexibly dispose data in “data pool” and send results to the external system through “communication channel”. The coupling

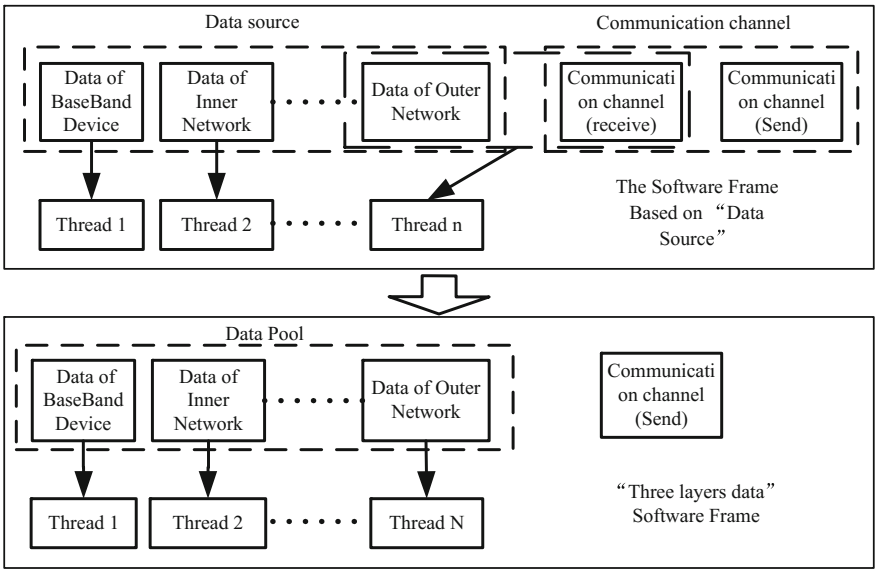


Fig. 16.5 Concept diagram of “data pool”

of conception in three is avoided which is very beneficial to design modular software, reduce coupling in modules and improve the adaptation, maintenance and expansibility.

16.4.3.2 Efficiently Filter and Judge Data, Simplify the Frame to Improve Efficiency

In the software frame based on "data source", "sub-data source" is added in order to distinguish different kinds of data in the single data source. Every "data source" creates a thread of receiving data. Data sources are judged by "sub-data source" in this thread which doesn't improve the software operation efficiency. Data in "data pool" are not partitioned directly, but using the regular expression judge data whether or not need to dispose which simplifies the software frame and improves readability of code and disposal efficiency of data.

16.4.3.3 Flexible Data Grouping, Improve Visual Operation Ability

Because there are many kinds of data tables and data channels in the process of executing the task, in order to increase operability, sending data is not only controlled by two methods of controlling data tables and data channels, but also group to data tables in the new frame. As shown in Fig. 16.6.

Every data table binds a "sending group", each sending group is named as "Persist group", "Forward group", "Fixed group", "Measure group", "Simulation group", "Backup group" and so on.

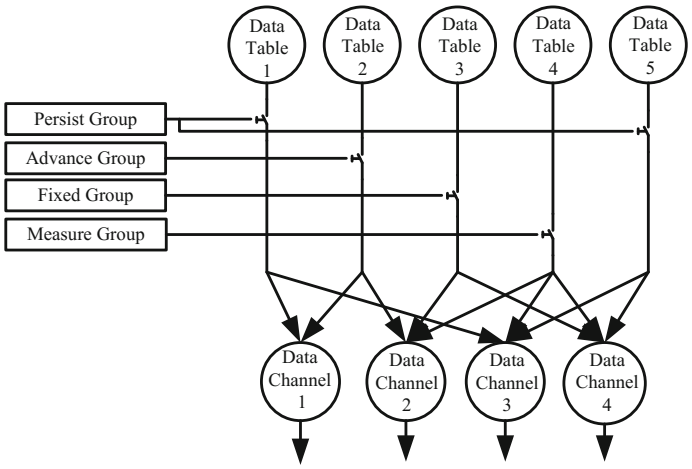


Fig. 16.6 Diagram of sending group controlling data table

In the normal state, each data table is in different sending groups. When the group switch is opened, data tables are sent immediately after disposed data. According to joint debugging procedure in different phases, open different groups and complete joint debugging requirements. In the emergency circumstance, special requirements are completed by controlling every "data table" and "data channel" which increases flexibility of sending data and improving operability of users.

16.5 The New Frame Software is Completed in the Kylin Operation System

16.5.1 The Realization of DTE

The telemetry transfer software based on the new frame in the Kylin operation system using C++ program language on QT flat is completed. The timeboard supported Kylin operation system is used. DTE software in the new frame is divided to seven modules: Task configuration and management module. Visual task state configurations and other management works are completed based on eXtensible Markup Language (XML) and regular expression; Time disposal module, which collection highest to 200 Hz are completed using the system signal waiting mechanism; Task disposal and management module, which is in charge of creating objects of data at all levels and completing the task process; Data pool and management module, which is in charge of receiving and pre-disposing external data and managing receiving data; "Data table" and management module, which is in charge of managing data disposal methods in the data pool and sending data; Data channels and management module, which is in charge of sending and saving "data table" and managing sending data channels; Display modules, which is in charge of displaying information of software configurations and tasks.

16.5.2 The Key Technology and Characteristics

16.5.2.1 Module Design Without Telemetry, Improve Generalization of Transfer Software

This software is based on a highly abstract frame and all classes are out of the natural attribute define of telemetry equipments. "Meta-object" system is used to complete dynamically creating different kinds of objects. The generalization in the process of data disposal and transfer using space measuring equipments is achieved. For other measuring equipments, this software only need to use "meta-object" system and increase corresponding data disposal methods.

In view of current tasks demands, many kinds of data disposal methods are designed, such as:

- Data selection class: telemetry selection, common selection, base selection, GNSS selection, etc.;
- Data framing class: Packet Data Exchange Protocol (PDXP), calculating number of days, count, telemetry state, servo state, etc.;
- Data conversion class: Global Navigation Satellite System (GNSS) Data, instantaneous orbit, computer word, bite control parameter, telemetry time, time transformation, etc.;
- Data simulation class: fixed value filling, simulation file etc.

16.5.2.2 The Process of Establishing the Task State like Building Blocks, Improve the Efficiency of Maintenance

Using QT with flexible interface program technology, highly visual configuration pages are made. Operators can assemble step by step to basal modules and configurate parameters by visual and building blocks methods. Then complicated task states are established. States of similar tasks are established just need to modify part of parameters of original task configuration, the whole process of establishing task state no longer need to modify codes. It improves the efficiency of software maintenance and reduces the time of preparing tasks. Meanwhile, because of the design of high abstraction, low coupling and building blocks, this software can meet new demands which may appears in future. As shown in Fig. 16.7, corresponding "building blocks" are dismantled, added or modified according to contents of software maintenance.

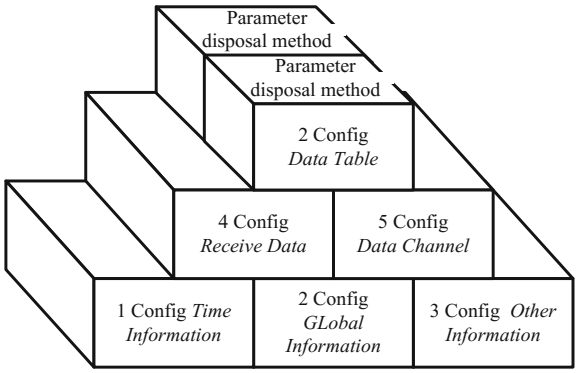


Fig. 16.7 Diagram of task state establishing process

16.5.2.3 Data Display of Jigsaw, Focus on Key Data

In view of problems of various kinds, large amount of data and inconvenient monitor of telemetry data, MIME (Multipurpose Internet Mail Extensions) technology is used which can drag, assemble, fold and delete display information of different kinds of data expediently, quickly and visually. The most key data and the most care states can be placed on the main window. Meanwhile, key data of different kinds of data flows are judged real-time. The system would give an alarm to abnormal states through indicator lights and widget colors which can improve ability of operators monitoring real-time data and help operators carry out emergency disposal.

16.5.2.4 The United Structure of Data Storage, Accelerate Data Analysis After Tasks

Receiving and sending data in this software are saving according to objects which are created, the name of data storage is same as the name of creating objects. In addition, in order to accelerate the speed of disposing data after tasks, the format of data storage is defined. In the frame head of data, the state of this task and the type, structure, takeoff absolute time of this data are defined. Time code, data length and end symbol are added in each frame data. It is convenient to divide data and analyze key information of losing frames, frequency and sending time.

16.5.2.5 Long-Distance Order Interface, Adapt Future Automatic Operation Demands

In order to adapt automatic operation demands, a suit of order format of transfer software is designed. By judging the type of receiving data and according to order type of receiving data control switch of data group and data storage, many operation items are completed. It lays the foundation for DTE long-distance, equipment one-click and automatic operation.

16.6 Conclusion

It indicates that the transfer software based on this frame has high generalization, flexibility and adaptability which shortens the time of the software maintenance, reduces the difficulty of the software maintenance and improves the visual monitor ability of real-time data. It can meet the demand of DTE software in future tasks and has the value of generalization and application in other space measurement equipment DTE sub-systems.

References

1. Yuncai L (2000) Telemetry, tracking and command system, vol 1. National Industry Publishing House, Beijing
2. Yien C (2002) Telemetry data process. National Industry Publishing House, Beijing
3. Kim J, Mo HS, Jung W, Lee H (1999) CCSDS packet data processing of KOMPSAT MCE system. Asia-pacific conference on communications & fourth optoelectronics & communications conference, vol 9, no. 2, pp 133–156
4. Zhang Y (2008) CCSDS data software design for ground receive system. J Telem Track Command 29(3):1–5
5. Haijiang Z, Changmin Y (2011) Analysis of IP multicast technology application in telemetry data transmission. J Telem Track Command 32(04):70–72
6. Weiskopf A, Weichert F, Fränzel N, Schneider M (2014) Flexible embedded telemetry system for agriculture and aquaculture. Computer and computing technologies in agriculture VII. Springer, Berlin, 552–560
7. Wang G, Zhao Q (2013) A satellite telemetry data processing system based on ice middleware. Sci Lit China 10:67–70
8. Guo LB, Rao AS, Xi HM, Mao WH, Zhang XH (2011) A Distributed socket telemetry data processing system architectural design based on ICE. J Telem Track Command 32(4):15–18
9. Li R, Liu GS (2004) Application of the technology of plug-in in range telemerty data transmission software design. Comput Appl Softw 21(12):37–39
10. Bian CW (2013) YQ-2505 S wave band immovable telemetry station principles and its application index. Xichang Satellite Lanch Center Publishing House, XiChang
11. Zhang CY, Chen L (2013) Telemetry data transmission system design based on database. Radio Commun Technol 39(3):46–49

Chapter 17
Telemetry Communication in Complex Attitude Conditions Based on Space-Time Coding

Hongpeng Zhu, Jun Cai, Zhiqiang Li and Zhongwu Xiang

17.1 Introduction

For aircraft under complicated attitude conditions, when the aircraft's position is located between the sending and receiving antennas, the signal would attenuate greatly and even been blocked. To solve this problem, there are generally two kinds of method: one is to use steerable antenna with tracking-pointing mechanism for the aircraft, but the antenna technology is too complicated, and the antenna tracking-pointing range is limited. Another one is to use a pair of antenna on both sides of the body, which ensures at least one antenna is visible to earth station. Under certain attitude conditions, both two antennas are visible, but the distances between the sending and receiving antennas are different. This leads to the phase difference of two signals from different antennas, which manifests as the phenomenon of destructive interference. While these two signals use different frequencies, the above problem can be solved effectively but signal bandwidth is doubled, and earth stations need to be equipped with a pair of receiving antennas. Frequency resource is becoming more precious today and the solution requiring double bandwidth is inappropriate.

With space-time coding [1] of these two signals, the reliability of telemetry communication under complex attitude can be improved without extra bandwidth. Space-time coding for aeronautical telemetry were studied by Rice [2, 3], but the decoder is based on hard decision. With the development of channel coding technology, turbo codes [4, 5], LDPC codes [6–8], Polar codes [9, 10] and other advanced coding technologies have been widely used in telemetry communication. These advanced coding are all based on soft decoding. If space-time codes are hard-decoded, subsequent channel coding cannot give full play to their advantages.

H. Zhu (✉) · J. Cai · Z. Li · Z. Xiang
PLA University of Science and Technology, Nanjing 210000, China
e-mail: hongpengzhu@126.com

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_17

For ensuring the reliability of telemetry communication, the adoption of erasure code technique was proposed in paper [11], while situations under complex attitude conditions was not considered. Concatenated LDPC and space-time coding in Rayleigh fading channel has been studied by Fang [12, 13], but they placed their intersection construction and analysis of LDPC codes and didn't illustrate how to perform soft space-time decoding. There are also a number of other researches on Concatenated LDPC and space-time coding [14–17], but they are not for the telemetry communication under complex attitude conditions.

In this paper, space-time coding for telemetry communication under telemetry communication are studied. We mainly focus on how to perform soft space-time decoding based on channel estimation. The rest of this paper is organized as follows. Section 2 introduces the system model with concatenated LPDC space-time codes and a 2×1 antenna system. Section 3 elaborates the channel estimation algorithm and provides the parameters for space-time decoding. Section 4 proposes an algorithm of space-time soft decoding. Section 5 gives the bit error performance of telemetry communication based on the space-time coding. Finally Sect. 6 concludes the paper.

17.2 System Model

The telemetry communication system includes one aircraft and an earth station, as shown in Fig. 17.1. A pair of transmitting antennas is mounted on the aircraft. The earth station is equipped with only one receiving antenna.

Assume $\boldsymbol{s}$ to be the transmitted information, and $\boldsymbol{c}$ is the corresponding LDPC codes. With space-time coding, the codes denoted as $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are generated. After BPSK modulation, the two signals would be emitted by different antennas respectively.

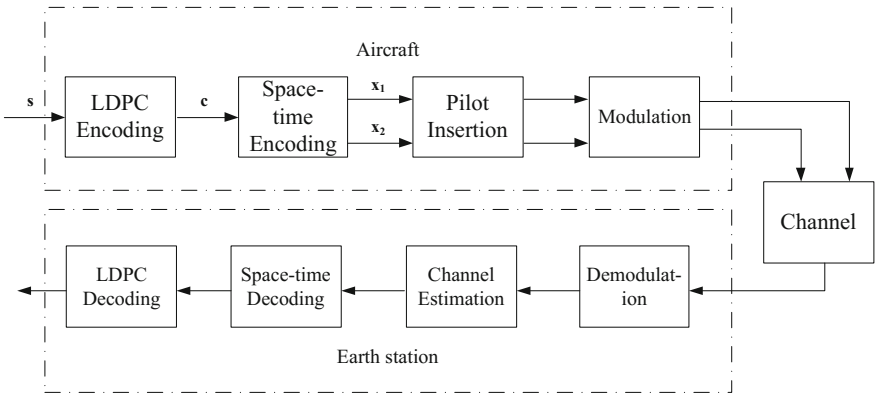


Fig. 17.1 Adams model of stable-platform

Space-time coding is performed as follows [18–20].Break the LDPC codes into several groups, and each group contains two symbols, as expressed below.

$$c_0, c_1|c_2, c_3|c_4, c_5|c_6, c_7| \cdots |c_{2k}, c_{2k+1}| \cdots \tag{17.1}$$

The space-time coding generates two vectors denoted as x_1 and x_2. $x_{1,i}$ and $x_{2,i}$ are the elements of corresponding vector. The encoding rule is expressed as (17.2) and (17.3).

$$x_{1,2k} = c_{2k}, x_{1,2k+1} = 1 - c_{2k+1} \tag{17.2}$$

$$x_{2,2k} = c_{2k+1}, x_{2,2k+1} = c_{2k} \tag{17.3}$$

The space-time codes are divided into blocks, with L_c symbols in each block, where L_c is integer times of 2. As the receiver need channel information for space-time decoding, pilot symbols of length L_p should be inserted before each code block for channel estimation. The pilot symbols for two signals are independent of each other.

When the mixed signals from two transmitting antennas arrive at the receiver, synchronization should be performed to find out the pilot information. Then, we can do the pilot-aided channel estimation to assist space-time decoding. Space-time decoder is based on soft-in-soft-out (SISO) decoding algorithm and outputs soft likelihood information to LDPC decoder. In this paper, we mainly focus on channel estimation and space-time soft decoding.

17.3 Channel Estimation Algorithm

Pilot sequence in each block can be used for channel estimation by iterative algorithm.

Assume baseband signals transmitted from two antennas to be $s_0(t)$ and $s_1(t)$. Parameters of the corresponding 2 channels are $h_0 = |h_0|e^{j\theta_0}$ and $h_1 = |h_1|e^{j\theta_1}$. Since two transmitting antennas are installed on the same aircraft which is distant from earth station, frequency offset of two signals can be considered approximately the same, denoted as ω_0. As the distances between two transmit antennas and receiving antenna are different, there exists difference of propagation delay. Suppose the propagation delay of the first signal is τ_0, and the second is τ_1. When the difference $\Delta\tau = \tau_1 - \tau_0$ is much smaller than the symbol period T_s, we consider $\tau_1 \approx \tau_0$.

The baseband form of received signal is as follows.

$$r(t) = [h_0 s_0(t - \tau_0) + h_1 s_1(t - \tau_1)]e^{j\omega_0 t} + n(t) \tag{17.4}$$

where $n(t)$ is the complex noise.

The objective of channel estimation is to estimate the parameters h_0, h_1, ω_0, τ_0 and τ_1.The iterative estimation algorithm can be described as follows:

(a) Initialization: $\hat{h}_0 = 1$, $\hat{h}_1 = 1$, $\hat{\tau}_0 = 0$, $\hat{\tau}_1 = 0$.
(b) Estimating frequency offset according to $\hat{h}_0$, $\hat{h}_1$, $\hat{\tau}_0$ and $\hat{\tau}_1$.
(c) Compute $\hat{h}_0$, $\hat{h}_1$, $\hat{\tau}_0$ and $\hat{\tau}_1$ based on $\hat{\omega}_0$.
(d) Go to step (b) for next iteration.

The maximum likelihood algorithm for frequency offset estimation is as follow:

$$\hat{\omega}_0 = \text{argmax}_{\omega_0}\left\{\left|\sum_{n=0}^{N_p-1} r(nT_s)\left[h_0^* s_0^*(nT_s-\tau_0)+h_1^* s_1^*(nT_s-\tau_1)\right]e^{-j\omega_0 nT_s}\right|^2\right\} \tag{17.5}$$

where T_s is the sampling period, and N_p is the number of sample points in the pilot segment. Suppose the up sampling rate is 4, that is $T_s = T_d/4$, where T_d is the symbol period, and $N_p = 4L_p$, where L_p is the number of symbols in the pilot segment. Study shows that an appropriate increase of sampling rate can improve the accuracy of the frequency estimation.ω_0 depends on the estimate h_0, h_1 and τ, which we use their estimated value $\hat{h}_0$, $\hat{h}_1$ and $\hat{\tau}$ instead respectively. The above estimation algorithm has no closed form solution. It can be solved by traversing search.

Frequency offset can be compensated for the pilot segment based on frequency estimation. The compensated signal can be written as

$$r_c(t) = [h_0 s_0(t-\tau_0)+h_1 s_1(t-\tau_1)]e^{j\Delta\omega tt}+n'(t) \tag{17.6}$$

where $\Delta\omega$ is the residual frequency offset.

The discrete form of $r_c(t)$ can be expressed as:

$$r_c(kT) = [h_0 s_0(kT-\tau_0)+h_1 s_1(kT-\tau_1)]e^{j\Delta\omega kT}+n'(kT) \tag{17.7}$$

Define vectors $\mathbf{r_c}$, $\mathbf{w}$, $\mathbf{s}_0(\tau_0)$, $\mathbf{s}_1(\tau_1)$ and $\boldsymbol{\omega}$ as

$$\mathbf{r_c} = \begin{bmatrix} r_c(0) \\ r_c(T) \\ \vdots \\ r_c\big((L_p-1)T\big) \end{bmatrix} \tag{17.8}$$

$$\mathbf{s}_0(\tau_0) = \begin{bmatrix} s_0(-\tau_0) \\ s_0(T-\tau_0) \\ \vdots \\ s_0\big((L_p-1)T-\tau_0\big) \end{bmatrix} \tag{17.9}$$

$$\mathbf{s}_1(\tau_1) = \begin{bmatrix} s_1(-\tau_1) \\ s_1(T - \tau_1) \\ \vdots \\ s_1\big((L_p - 1)T - \tau_1\big) \end{bmatrix} \tag{17.10}$$

$$\boldsymbol{\omega} = \begin{bmatrix} 1 \\ e^{j\Delta\omega T} \\ \vdots \\ e^{j\Delta\omega(L_p-1)T} \end{bmatrix} \tag{17.11}$$

$\mathbf{r_c}$ can be written as:

$$\mathbf{r_c} = (h_0\mathbf{s}_0(\tau_0) + h_1\mathbf{s}_1(\tau_1)).*\boldsymbol{\omega} + \mathbf{w} \tag{17.12}$$

Express $\mathbf{s}_0(\tau_0)$ and $\mathbf{s}_1(\tau_1)$ with a matrix

$$\mathbf{S}(\boldsymbol{\tau}) = [\mathbf{s}_0(\tau_0) \quad \mathbf{s}_1(\tau_1)] \tag{17.13}$$

and h_0 and h_1 with a vector

$$\mathbf{h} = \begin{bmatrix} h_0 \\ h_1 \end{bmatrix} \tag{17.14}$$

Thus $\mathbf{r_c}$ can be rewritten as vector

$$\mathbf{r_c} = (\mathbf{S}(\boldsymbol{\tau})\mathbf{h}).*\boldsymbol{\omega} + \mathbf{w} \tag{17.15}$$

The maximum likelihood estimation for channel parameter $\boldsymbol{\tau}$ is as follows:

$$\hat{\boldsymbol{\tau}} = \mathrm{argmin}_{\boldsymbol{\tau}}\left\{\left|\left[\mathbf{I} - \mathbf{S}(\boldsymbol{\tau})\big(\mathbf{S}^{\mathbf{H}}(\boldsymbol{\tau})\mathbf{S}(\boldsymbol{\tau})\big)^{-1}\mathbf{S}^{\mathbf{H}}(\boldsymbol{\tau})\right]\mathbf{r_c}\right|^2\right\} \tag{17.16}$$

The solution of $\hat{\tau}$ is the one that makes $\left\{\left|\left[\mathbf{I} - \mathbf{S}(\boldsymbol{\tau})\big(\mathbf{S}^{\mathbf{H}}(\boldsymbol{\tau})\mathbf{S}(\boldsymbol{\tau})\big)^{-1}\mathbf{S}^{\mathbf{H}}(\boldsymbol{\tau})\right]\mathbf{r}\right|^2\right\}$ minimum,so we need to calculate the expression for all the possible values of $\boldsymbol{\tau}$. Since $\boldsymbol{\tau}$ has infinite possibilities, it is impossible to traverse all cases. We need to design the resolution of $\boldsymbol{\tau}$ and quantify accordingly. Then we can calculate corresponding results for each $\boldsymbol{\tau}$ and find the minimum value. The corresponding $\boldsymbol{\tau}$ for the minimum is the estimated value $\hat{\boldsymbol{\tau}}$.

The estimation of channel parameter $\mathbf{h}$ is as follows.

$$\hat{\mathbf{h}} = \left(\mathbf{S}^{\mathbf{H}}(\hat{\tau})\mathbf{S}(\hat{\tau})\right)^{-1}\mathbf{S}^{\mathbf{H}}(\hat{\tau})\mathbf{r} \tag{17.17}$$

Substituting (17.16) into above equation, we get

$$\hat{\mathbf{h}} = \left(\mathbf{S}^{\mathbf{H}}(\hat{\tau})\mathbf{S}(\hat{\tau})\right)^{-1}\mathbf{S}^{\mathbf{H}}(\hat{\tau})((\mathbf{S}(\tau)\mathbf{h}).*\boldsymbol{\omega}+\mathbf{w}) \tag{17.18}$$

Assume the impact of noise components is not considered, we get

$$\hat{\mathbf{h}} = \left(\mathbf{S}^{\mathbf{H}}(\hat{\tau})\mathbf{S}(\hat{\tau})\right)^{-1}\mathbf{S}^{\mathbf{H}}(\hat{\tau})(\mathbf{S}(\tau).*\boldsymbol{\omega})\mathbf{h} \tag{17.19}$$

From the above equation, we can see that the residual frequency offset will cause error of channel parameter estimation. But as long as the residual frequency offset is small enough, the impact can be negligible.

17.4 Space-Time Soft Decoding Algorithm

Suppose the information to be sent is x_1 and x_2. After space-time coding, two signals$[x_1, 1-x_2]$ and $[x_1,x_2]$ are generated.

Define **s** to be the transmitted signal for each code block.

$$\mathbf{s} = \begin{bmatrix} \mathrm{x}_1 & 1-\mathrm{x}_2 \\ \mathrm{x}_2 & \mathrm{x}_1 \end{bmatrix} \tag{17.20}$$

The channel parameter is denoted as **h**.

$$\mathbf{h} = [\,\mathrm{h}_1 \quad \mathrm{h}_2\,] \tag{17.21}$$

Received signal **r** can be written as:

$$\mathbf{r} = \mathbf{hs} + \mathbf{n} \tag{17.22}$$

where **n** is the noise, $\mathbf{n} = [\,\mathrm{n}_1 \quad \mathrm{n}_2\,]$ and $\mathbf{r} = [\,\mathrm{r}_1 \quad \mathrm{r}_2\,]$.

Space-time decoding outputs likelihood information.

$$\mathrm{L_D}(\mathrm{x_k}|\mathbf{r}) = \ln\frac{\mathrm{Prob}\{\mathrm{x_k}=0|\mathbf{r}\}}{\mathrm{Prob}\{\mathrm{x_k}=1|\mathbf{r}\}}, \quad \mathbf{k} = 1,2 \tag{17.23}$$

According to the received signal **r**, the likelihood information of each transmitted bit is determined, and the information is sent to the LDPC decoder.

$L_D(x_k|\mathbf{r})$ can be expressed as

$$L_D(x_k|r) = \ln \frac{\sum_{\mathbf{s}\in\mathbf{s}^+} p(\mathbf{r}|\mathbf{s})\text{Prob}\{\mathbf{s}\}}{\sum_{\mathbf{s}\in\mathbf{s}^-} p(\mathbf{r}|\mathbf{s})\text{Prob}\{\mathbf{s}\}} \tag{17.24}$$

where s^+ represents all sets of $\mathbf{s}$ corresponding to $x_k = 0$, and $\mathbf{s}^-$ corresponding to $x_k = 1$. Each code block contains two bits. Take the first bit of code for example. $\mathbf{s}^+$ contains{0,0}and{0,1},and $\mathbf{s}^-$ contains{1,0} and {1,1}.

The above formula can be converted to

$$L_D(x_0|\mathbf{r}) = \ln \frac{\sum_{\mathbf{s}\in\mathbf{s}^+} p(\mathbf{r}|\mathbf{s})(\text{Prob}\{x_0 = 0\}\text{Prob}\{x_1\})}{\sum_{\mathbf{s}\in\mathbf{s}^-} p(\mathbf{r}|\mathbf{s})(\text{Prob}\{x_0 = 1\}\text{Prob}\{x_1\})} \tag{17.25}$$

$$L_D(x_0|\mathbf{r}) = \ln \frac{\text{Prob}\{x_0 = 0\}}{\text{Prob}\{x_0 = 1\}} + \ln \frac{\sum_{\mathbf{s}\in\mathbf{s}^+} p(\mathbf{r}|\mathbf{s})(\text{Prob}\{x_1\})}{\sum_{\mathbf{s}\in\mathbf{s}^-} p(\mathbf{r}|\mathbf{s})(\text{Prob}\{x_1\})} \tag{17.26}$$

Suppose 0 and 1 have equal probability. The above equation can be further simplified to

$$L_D(x_0|\mathbf{r}) = \ln \frac{\sum_{\mathbf{s}\in\mathbf{s}^+} p(\mathbf{r}|\mathbf{s})}{\sum_{\mathbf{s}\in\mathbf{s}^-} p(\mathbf{r}|\mathbf{s})} \tag{17.27}$$

where

$$p(\mathbf{r}|\mathbf{s}) = \frac{1}{\left(4\pi(\text{SNR})^{-1}\right)^2} e^{\frac{\text{Tr}\left((\mathbf{r}-\mathbf{hs})^H(\mathbf{r}-\mathbf{hs})\right)}{2(\text{SNR})^{-1}}} \tag{17.28}$$

$L_D(x_0|\mathbf{r})$ can be simplified as

$$L_D(x_0|\mathbf{r}) = \frac{r_1 h_1^* + r_2^* h_2}{h_1^2 + h_2^2} \tag{17.29}$$

$$L_D(x_1|\mathbf{r}) = \frac{r_1 h_2^* - r_2^* h_1}{h_1^2 + h_2^2} \tag{17.30}$$

(17.29) and (17.30) can be regarded as the log-likelihood ratio(LLR) output of soft space-time decoder.

17.5 Performance Simulation

In this section, the performance based on space-time coding under complex attitude conditions is simulated. The system uses two transmitting antennas and one receiving antenna. Assume that data is encoded with LDPC codes of length 512 and rate 1/2. The threshold of E_b/n_0 corresponding to BER < 10^{-5} for this LDPC codes is 3 dB. Each sending frame is comprised of five coding frames. The length of sending frame is 2560.

Perform space-time coding to frame to generate 2 signals. A pilot segment of 256 symbols is added in front of each frame. It is notable that 2 signals must have uncorrelated pilots. Each signal is BPSK modulated. Assume the symbol rate is 512 K baud. The channel conditions are simulated by setting the signals' channel parameters $\mathbf{h}_0$ and $\mathbf{h}_1$. Doppler effect is simulated by adding a frequency offset $\Delta\omega$. The receiver uses the above channel estimation algorithm and space-time decoding method. It outputs LLR information to LDPC decoder. Bit error rate performance is calculated by comparing LDPC decoder output with source data. The direct transmission without space-time coding is also simulated and its performance is compared with space-time coding. In the simulation, we assume 2 signals have the same transmitting power. For each signal its' signal to noise ratio E_b/n_0 is 3 dB. When the phase differences of two received signals changes from 90° to 180°, the bit error rate performance is shown in Fig. 17.2.

In Fig. 17.2, the curve labeled with square represents BER performance without space-time coding, while the curve labeled with star is based on space-time coding.

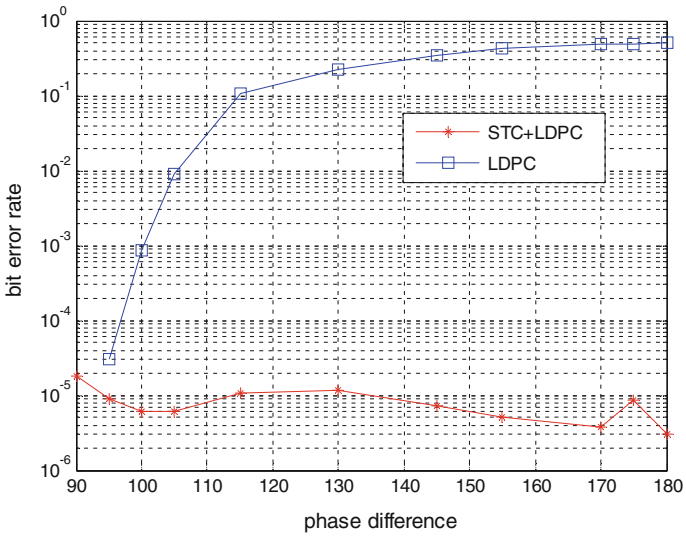


Fig. 17.2 Comparison of bit error rate performance with and without space time coding under complex attitude conditions

As shown in the figure, if data is transmitted directly, these 2 signals appear to cancel with each other while the phase difference changes between 90° and 180°. The more closer to 180°, the signal cancellation phenomenon is more obvious, thus the BER performance is much worse.

When the difference reaches 180°, the BER is close to 0.5. However, with space-time coding, BER performance is not sensitive to the phase difference any longer. The error performance almost remains unchanged while the phase difference changes from 90° to 180°. The BER is less than 10^{-5} when $E_b/n_0 = 3dB$. This simulation result justifies that space-time coding can improve the reliability of telemetry communication under complex attitude conditions.

17.6 Conclusion

In this paper, telemetry communication under complicated attitude conditions is studied. We mainly focus on how to improve the reliability of telemetry communication by adopting a dual antenna system and space-time coding. Based on dual-antenna system, the high-gain LDPC coding and space-time coding, a telemetry communication system model is established. A pilot-aided method for channel estimation is stated. Then a soft space-time decoding algorithm based on the channel parameter is proposed. The space-time decoder can output soft information to subsequent LDPC decoder. The simulation result shows that the method with space-time coding is not sensitive to the phase difference of 2 signals from 2 transmitting antennas. Its reliability is much better than transmitting signal directly without space-time coding. The result is meaningful for solving the reliability problem on telemetry communication under complicated attitude conditions.

Acknowledgements This work is supported by National Natural Science Foundation of China (No. 61571464).

References

1. Alamouti M (1998) A simple transmit diversity technique for wireless communications. IEEE J Sel Areas Commun 16(8):1451–1458
2. Rice M (2001) Space-time coding for aeronautical telemetry: part I—system description. In: Proceedings of the international telemetering conference, Las Vegas, USA
3. Temple K, Rice M (2011) Space-time coding for aeronautical telemetry: part II—experimental results. In: Proceedings of the international telemetering conference, Las Vegas, NV
4. Berrou C, Glavieux A, Thitmajshima P (1993) Near Shannon limit error-correcting coding and decoding: turbo-codes. In: IEEE conference proceedings, Geneva, pp 1076–1080
5. Valenti MC (2000) Inserting turbo code technology into the DVB satellite broadcasting system. In: IEEE military communications conference MOLCOM, pp 650–654

6. Winters JH (1984) Optimum combining in digital mobile radio with co-channel interference. IEEE Trans Veh Technol 33(4):144–155
7. Tanner RM (1981) A recursive approach to low complexity codes. IEEE Trans Inf Theory 27 (5):533–547
8. Gallager RG (1963) Low-density parity-check codes. M.I.T. Press, Cambridge, MA
9. Arikan E (2009) Channel polarization: a method for constructing capacity achieving codes for symmetric binary-input memoryless channels. IEEE Trans Inf Theory 55(7):3051–3073
10. TalI VardyA (2015) List decoding of polar codes. IEEE Trans Inf Theory 61(5):2213–2226
11. Jinrong Z, Ling W, Shasha Y (2013) Erasure correction codes for frame loss recovery in telemetry links [J]. J Spacecr TT&C Technol 32(6):501–506 (in Chinese)
12. Fang Y et al (2015) Asymptotic performance analysis of protograph LDPC–coded STBC system in fading channels. IEEE Trans Aerosp Electron Syst 43(1):262–272
13. Fang Y, Han GJ, Chen PP et al (2015) Protograph LDPC codes for STBC Rayleigh fading channels. In: 15th international symposium on communications and information technologies (Iscit), pp 93–96
14. Yan YR, Wei QF, Li Y, Wang L (2009) Interleave-division-multiple space-time system using LDPC codes. J Chongqing Univ Posts Telecommun 21(1):57–60
15. Zhang Y, Yang FF (2015) Detection algorithm for LDPC combined with vertical layered space time code. J Nanjing Univ Posts Telecommun 35(3):47–51
16. Ghayoula E, Taieb MH, Chouinard JY et al (2015) Improving MIMO systems performances by concatenating LDPC decoder to the STBC and MRC receivers. In: 2015 world symposium on computer networks and information security (Wscnis)
17. Gupta P, Gupta B. LLR (2014) Based analysis of LDPC codes concatenated with orthogonal STBC coded MIMO Systems2014 international conference on signal propagation and computer technology (Icspct), pp 680–684
18. Jensen M, Rice M, Anderson A (2007) Aeronautical telemetry using multiple-antenna transmitters. IEEE Trans Aerosp Electron Syst 43(1):262–272
19. Jensen M, Rice M, Nelson T, Anderson A (2004) Orthogonal dual-antenna transmit diversityfor SOQPSK in aeronautical telemetry channels. In: Proceedings of the international telemetering conference, San Diego, CA
20. Nelson T, Rice M, and Jensen M (2005) Experimental results with space-time coding using FQPSK. In: Proceedings of the international telemetering conference, Las Vegas, NV

Chapter 18
Development of the Lunar-Earth and Deep Space TT&C System with Several Key Techniques

Haifeng Yang, Lin Chai, Ouxin Lu, Jianping Hu, Maoge Xu and Hui Yan

18.1 Introduction

Deep space telemetry, track and command (TT&C) system, which refer to the detection of the moon and further celestial bodies or space, is an important symbol of both national strength and the development of science and technology. Developing deep space TT&C can effectively drive the progress of space science, technology and application. Meanwhile, it can promote a variety of significant scientific researches, such as the influence of celestial activity, the origin of the solar system, and the exploration of extra-terrestrial life [1, 2].

Under the circumstances of lunar-earth and even further deep space TT&C system, it is difficult to improve the performance of traditional radio frequency TT&C to satisfy the current need. At the same time, a series of new assignment demands new technology and higher-performance. On the basis of analysis of the current research status of the lunar-earth, deep space TT&C system at home and abroad, and this paper summarizes the major difficulties and challenges to establish the deep space TT&C system of our country. The recommendations of future developments of the lunar-earth, deep space TT&C system have been proposed, which can provide references for the development of the lunar-earth, deep space TT&C system.

H. Yang (✉) · L. Chai · O. Lu · J. Hu · M. Xu · H. Yan
The 10th Research Institute of China Electronics Technology Group Corporation (CETC), Chengdu 610036, China
e-mail: haifeng_ioe@163.com

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_18

18.2 Lunar-Earth and Deep Space TT&C Task Requirements

18.2.1 Overall Demand for Lunar-Earth and Even Further Deep Space TT&C System

In November 2000, Chinese government published "China's space" white paper, formally announced to launch the advanced study of moon-oriented deep space exploration. In recent, the State Council of the People's Republic of China published "The Outline of National Medium and Long Term Science and Technology Development Program", in which deep space exploration is proposed as an independent division. This explicitly implies the urgent demand of deep space exploration.

At the same time of finalizing the plan of china lunar exploration project, according to the general objective of deep space exploration, 9 exploration tasks will be scheduled during the next 20 years. The aim of these tasks is to improve the relative technology and expand the detection range.

The whole assignment is divided into three stages, and the time node of each stage is 2020, 2025 and 2030. The TT&C tasks of each stage are as follows [3, 4]:

In the first stage: to expand the TT&C range from 1.5 million km to 400 million km, and to improve both measurement accuracy and data rate, thus to support the first Mars, Sun and Asteroid exploration task.

In the second stage: to expand the TT&C range from 400 million km to 1 billion km, and to improve the measurement accuracy by one order of magnitude, thus to support the first Venus exploration task, and second Mars, Sun and Asteroid exploration task.

In the third stage, the goal is to complete the task of orbiting around the Jupiter, and the task of returning the Mars sample. Further improve the measurement accuracy, and improve the transmitted data rate to dozens of Mbps between 1 billion km. the deep space TT&C technology of our country is expected to reach the international advanced level.

18.2.2 The Requirement of Deep Space TT&C System to Ka Band and Optical Frequency Band

Based on all the long term planning and requirement listed above, improving the frequency band is an effective approach to improve both the transmitted data rate and measurement accuracy. Thus, Ka-band deep space TT&C and laser communication technology is a powerful means for lunar-earth high-speed communication and ultra-long-range TT&C. The detailed technical requirements are as follows [5–9]:

Take deep space exploration for instance, some foreign countries, especially the Unites States, has completed the deep space exploration of Mars, Venus, Jupiter,

Saturn, Uranus, Neptune and other planets and their moons. Meanwhile, the United State has constructed a complete deep space TT&C system, which can fully utilize the existing resources to support the implement of Ka-band deep space TT&C and laser communication system.

China has been developing a great number of high-resolution earth observation projects, including both the space science and aviation science projects. In the future, there will be more related projects. What all these projects have in common is the focus on proprietary intellectual property rights and real-time data transmission. At present, limited communication bandwidth restricts the real-time transmission of data and its application. As pointed out by "The Outline of National Medium and Long Term Science and Technology Development Program", high-resolution data mainly depend on foreign agencies, while China has not formed a complete earth observation system yet. However, since TT&C system is an essential part of the earth observation system, the application of Ka-band deep space TT&C and laser communication technology will be a strong support and complement for earth observation system.

In the application of satellite & detector navigation and relay satellite communication, the integrated laser communication and ranging system can provide cm-level, even mm-level ranging accuracy, and a transmitted data rate of tens of Gbps. At the same time, it has a notable anti-interference ability. So it is regarded as a very promising development trend.

Moreover, our country has launched a variety of large aperture ground-based telescope station project. Meanwhile, deep space antenna array project is under research and preliminary implementation, which effectively promote the development of related industries. As a part of deep space TT&C system, Ka-band deep space TT&C system and laser communication system can fully utilize the existing resources to support their own implementation.

In the field of international advanced technology, quantum communication system shares a similar structure with integrated laser communication and ranging system. Thus, the integrated laser communication and ranging system can be transformed to implement quantum communication. So far, there is no report on successful quantum communication between the satellite and ground, while the integrated laser communication and ranging system can provide a physical platform for the experiments of quantum communication, which can effectively improve the utilization of experiment platform to obtain both the technical and economic benefit.

At the present, the TT&C system of our deep space exploration project mainly relies on the radio frequency of S/X band. Considering the practical experience of foreign deep space TT&C and laser communication, we need to construct some TT&C station which based on Ka band or, even laser. At the same time, we need to initiate the researches on integrated high-speed communication, ultra-long distance communication, key technology of lunar-earth laser communication, as well as principle prototype of small-size Ka-band TT&C equipment.

While analyzing and researching the technical requirement of the integrated laser communication and ranging system and its the key technology such as ATP,

according to our national conditions, we need to take a full advantage of existing radio-frequency and optical TT&C station in order to satisfy the increasing demand of space TT&C. It can be predicted that, in the next 20 or 30 years, the space mission of our country will enter a brand new period in which the distance between the spacecraft and the earth will be hundreds of millions of kilometers or even further. Simultaneously, the requirement of data transmission will constantly increase with the progress of detection payload. The application of Ka/laser TT&C technology will satisfy not only the need for ranging and data transmission, but also the need of data relay, thus maximize the benefits. Therefore, Ka/laser TT&C technology is a highly-integrated and multi-functional technology which is in line with the future trend.

18.3 An Overview of the Current Development Situation

18.3.1 The General Trend

Due to the tremendous path loss in deep space communication, improving the intensity of received signal is an eternal technical problem. In order to solve the problem, several measures have been adopted by space powers, such as increasing the carrier frequency to optical band, using laser communication technology, even integrated laser communication and ranging technology. In the aspect of radio frequency, main measures are, enlarging the antenna aperture, increasing the carrier frequency and reducing the noise temperature of receiving system. As for the uplink channel, it is also crucial to improve the EIRP of antenna. In order to achieve this goal, the relevant measures are, enlarging the antenna aperture, increasing the carrier frequency and the capacity of transmitter and feed system, and utilizing the technology of antenna array and low-temperature receiver.

Based on our existing condition, it is necessary to supplement the current TT&C station with Ka uplink channel and transform its Ka downlink channel. Also, the deep space TT&C antenna in the background of ultra-long distance tasks (like the moon, Mars, and Jupiter exploration missions) need be illustrated. The Ka band large aperture antenna channel technology, antenna array technology and the advanced laser communication ATP technology is the necessary technical means.

In the face of ultra-long distance, weak signal, and longtime delay, deep space TT&C has always been struggling to obtain high measurement accuracy and reliable high-speed data rate. In lunar-earth TT&C system, measurement and communication are usually conducted on Ka band radio frequency, while laser communication works as a complementary. As to the ultra-distant detection, take Jupiter detector for instance, TT&C is completed via deep space antenna array technology. In this case, radio frequency and laser link works interactively to take full advantage of existing resources, and to improve the efficiency of ATP, as well as the reliability and availability of the link channel. The overall composition

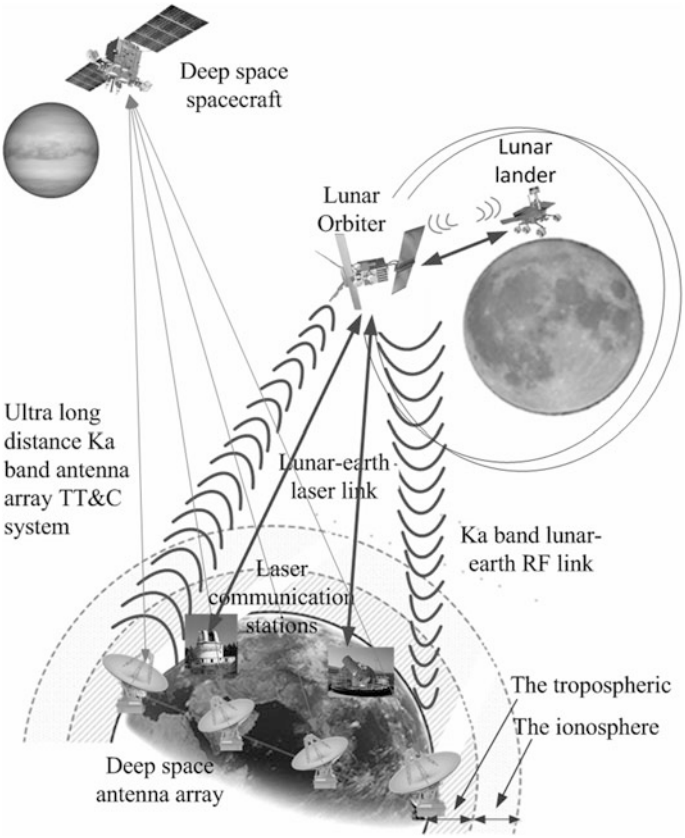


Fig. 18.1 The overall composition diagram of the lunar-earth and deep space TT&C system

diagram of the lunar-earth and deep space TT&C system, as shown in Fig. 18.1, blue line represents Ka link channel, while the red one represents laser link.

In order to match the key technical requirements of one-billion-kilometer ultra-distant deep space TT&C, it is not enough to rely merely on existing TT&C pattern, the fundamentally new technology or system is essential.

18.3.2 Development of Deep Space TT&C System in Radio Frequency Band

The research and construction of the deep space TT&C network have been vigorously developed in the world. At present, only the United States, the European Space Agency and Russia have completed the construction of deep space network. From the perspective of technical capability and coverage, the deep space network

of the United States (referred to as deep space network DSN) is the highest level in the world. The Jet Propulsion Laboratory (JPL) of NASA is in charge of constructing and managing DSN, as well as the implementation of tracking, navigation, measurement and control, communication tasks for deep space exploration spacecraft. As the largest and most sensitive scientific remote communication system in the world, the control center of DSN is located at the JPL headquarter in California Pasadena, and the whole system is composed of 3 deep space facilities in a longitude interval of 120°, which are located in the U.S. state of California, Madrid and Canberra stone Australia respectively.

The deep space TT&C network of Russia is composed of 3 deep space ground stations, 2 command centers and 2 ballistic centers. Its deep space station is also equipped with 32-m and 70-m antennas. Japan, Germany and India are also constructing their own deep space TT&C stations, in which the facilities are basically the models of DSN of the United States. At present, the developing trends of TT&C deep space station in foreign countries are as follows:

Construction of reflector antenna with large aperture, normally the antenna with 35-m, 64-m or even 70-m aperture. The aperture of antenna is reaching the manufacturing limit.

The signal frequency band is switching from S/X to Ka. The down link signal of deep space network is extremely weak, to achieve longer transmission distance, Ka band attracts more and more attention.

BWG antenna is adopted, and antenna feeds are moved to the ground to reduce the feeder loss. Also, LNA and HPA are easier to install.

Low- noise technology is adopted to reduce the noise level of ground system, and maximize the G/T value. At present, low noise level is guaranteed by LNA ultra-low- temperature refrigeration technology.

China has completed the construction of 35-m deep space TT&C station in Kashi, 66-m deep space TT&C station in Kiamusze. And the deep space TT&C station in South America, which is under construction, will be completed by 2016. Till then, China will accomplish the construction of its deep space network, and possess the capability of global TT&C for deep space spacecraft.

18.3.3 Development of Deep Space TT&C System in Optical Frequency Band

Inter-satellite laser communication and networking technologies have not been widely promoted and used in civil space area. From the perspective of current development of foreign countries, the experiments of laser communication link, and network test, especially for the civil space, has not been reported. However, from the previous trials, the development of optical communication technology is generally dependent on the mode of civil military integration. Here are some typical laser communication systems in foreign countries [10–22]:

Since 2000, nearly 20 foreign space laser communication experiments have been conducted, most of which were led by the United States and Europe.

In 2000, 17 low-earth-orbit and ground laser communication link set-up tests were conducted between the communication terminal of the United States space technology research satellite No. 2 (STRV-2) and ground-based terminal, all failed.

In 2001, the United States geostationary orbit lightweight technology test satellite (GeoLITE) conducted the earth orbit to ground (GEO-Ground) laser communication test. The result has not been reported yet.

In 2008, the world's first two-way inter-satellite coherent laser communication link was successfully set up between NFIRE satellite of the United States and the German TerraSAR-X microwave imaging satellite, the link distance was 4900 km, and the data transmission rate was 5.625 Gbps.

In 2010, bi-directional coherent laser communication tests were conducted between NFIRE satellite of the United States and the mobile optical ground station located in Maui in Hawaii of Tenerife of Spain respectively. The link distance was 500 km; the data transmission rate was 5.625 Gbps. However, Test system is not stable because of the high error rate.

Based on the establishment of complete the space TT&C system, NASA is more and more aware of the value of laser communication technology, which can not only improve the communication rate, but also greatly reduce power requirements compared with radio frequency communication. In 2008, NASA started to develop the lunar atmosphere and dust environment detector (LADEE). Since the satellite had the ability to carry extra loads, NASA took the opportunity to conduct the laser communication experiment (LLCD). In September 2013, LADEE were launched, and soon enough the laser communication terminal equipped on it created the history, realizing the data transmission between the moon and the earth, the distance of which is about four hundred thousand kilometers. The downlink date rate historically reached 622 Mbps, and the uplink data rate was up to 20 Mbps. The bidirectional delay accuracy was 200 picoseconds, and ranging accuracy reached 1 cm.

The satellite-based terminal of LLCD project is known as LLST. As shown in Figs. 18.2 and 18.3, it consists of optical module, modulation module and control module. A reflecting telescope with 100 mm aperture is adopted as T/R antenna of LLST. Downlink beacon light divergence angle is designed to be 15 μrad. The telescope is installed on a 2-axis cardan joint through a Magnetic fluid inertial reference unit (MIRU, which originated from technology of MLCD project). The angle measuring device in the MIRU provides the angle error measurement for the system, and restrains the interference though a group of voice coils motor. The receiving and transmitting optical fiber is fused together at the end of the optical path through an optical fiber coupler, and controlled by the piezoelectric ceramic to achieve fine tracking and advanced targeting. In the ATP system of LLST, there is only one detector, which is a four-quadrant detector. The transmitting and receiving antennas of ground-based terminal LLGT adopts T/R arrays which are composed of four telescopes. The aperture of the receiving telescope is 400 mm, and the transmitting aperture is 150 mm.

Fig. 18.2 Schematic of the lunar laser communication scenario (*left*) and the locations of the three ground station (*right*)

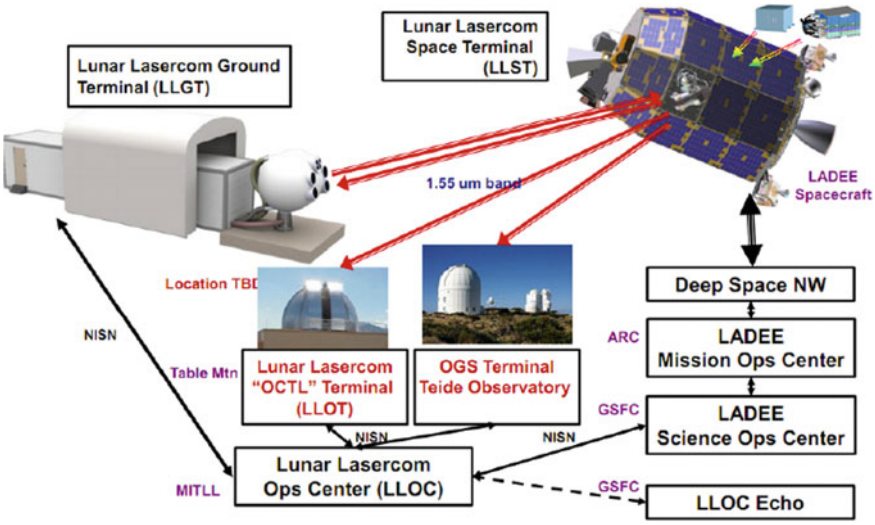


Fig. 18.3 Schematic of the lunar-ground space laser communication link

On June 5, 2014, NASA launched a laser optical communication sciences load (OPALS) experiment, using OOK modulation, the equivalent transmission rate of which reached 50 Mbps. The experiment accomplished the data transmission between the international space station and ground, verified the communication performance between the international space station and ground under the influences of joint jitter, atmospheric disturbance, background noise.

On the basis of LLCD, The United States deployed a laser communication relay demonstration project (LCRD), which is expected to conduct the high-speed laser communication test between the earth orbit and the ground (GEO-earth) in 2017. In this experiment, the link distance is 40,000 km, data rate is 2.88 Gbps, using DPSK modulation.

In 2001, an optical communication link was established between the European ARTEMIS geostationary orbiting data relay satellite and SPOT-4 orbit of earth observation satellite, successfully completing the inter-satellite laser communication link test between the relay satellite and the LEO satellite for the first time. The link distance is 45,000 km, data transmission rate are 50 Mbps.

In 2012, Russia performed a 400 km space station to ground laser communication experiment, with a bit-rate of 125 Mbps.

In 2013, the Europe data relay satellite Alphasat, with a coherent laser transmitter, has orbited at longitude 25° east, and a space-to-earth laser acquisition and tracking experiments have been conducted with Alphasat. The design path length for the link is 45000 km, and the download bit-rate is 1.8 Gps.

ESA has launch two data relay satellite EDRS-A and EDRS-C in 2015 and 2016. EDRS-A orbited at longitude 9° east, and EDRS-C orbited at longitude 31° east. These satellites are delivering data relay service for Europe countries, America, and Japan, eta. It is estimated that, after 2017, ESA will accomplish the networking of data relay satellites. The bit-rate of the backbone networks can reach 1.8 Gbps.

A lot of relevant theories of laser technology and studies for partial system have been carried out in China. However, few projects of laser technology were conducted. In view of the trends, new opportunities for development of laser technology are coming.

The relevant study for free-space laser communication was started from the Ninth Five-Year plan. During the Tenth Five-Year plan, the second generation data relay satellite has been studied. Systematic designing, simulation verification, terminal designing and prototype developing were accomplished. During the Eleventh Five-Year plan, space link Acquisition, Tracking and Pointing (ATP) system demonstration experiments were conducted, and PAT system was integrated.

In 2008, BITTT (Beijing Institute of Tracking and Telecommunications Technology, BITTT) of the General Armaments Department put forward the new concept, namely, “unified laser TT&C system”. In 2014, BITTT conducted a demonstration experiment of high-speed laser communication and precision ranging based on coherent laser communication systems. The bit-rate is Gbps level and the measuring accuracy is the millimeter level.

In 2012, aiming at the advanced world level, a 45,000 km LEO-GEO-earth high-speed laser communication system with the bit-rate of 5 Gbps has been built. Four kinds or two pairs of prototypes, which are the user equipment terminal, the space-to-space laser terminal, the space-to-earth laser terminal and the earth terminal were developed. Studies of high-speed coherent laser communication technology, unified laser ranging and controlling technology, space-based high-speed optical switching networks technology, space-to-earth laser communication adaptive optics compensation technology, and large field high sensitivity high speed laser detection technology have been conducted.

The Harbin Institute of Technology (HIT) has accomplished the confirmatory experiment of space-to-earth laser communication. The cutting edge research of quantum communication conducted by the Institute of Optics and Electronics, of

Chinese Academy of Sciences (IOE), and the University of Science and Technology of China (USTC) is world-renowned. The physical link of the quantum communication is also based on the laser communication link. The experiment for quantum communication was based on a 50 km horizontal path laser communication link experiment. In this experiment, the 700 mm aperture telescope's quantum entanglement receiver module was changed to a quantum key receiver module, and a compound 2 axes closed-loop tracking scheme was employed. A 96 km quantum encryption key distribution between the telescope and a satellite payloads prototype has been achieved in this quantum communication experiment. This achievement was published on "Nature Photonics". This is the world's first successful comprehensive confirmatory experiment for space-to-earth quantum encryption key distribution, the schematic of experimental set-up, as shown in Fig. 18.4.

Compared with the foreign countries, the development for free-space laser communication in China started late. The major institution developing free-space laser communication technology includes the Institute of Optics and Electronics, of Chinese Academy of Sciences (IOE), the Changchun Institute of Optics, Fine Mechanics and Physics, of Chinese Academy of Sciences (CIOMP), Shanghai Institute of Optics and Fine Mechanics, of Chinese Academy of Sciences (SIMO), the Harbin Institute of Technology (HIT), the Changchun University of Science and Technology (CUST), the Wuhan University (WHU), and the Xian branch academy

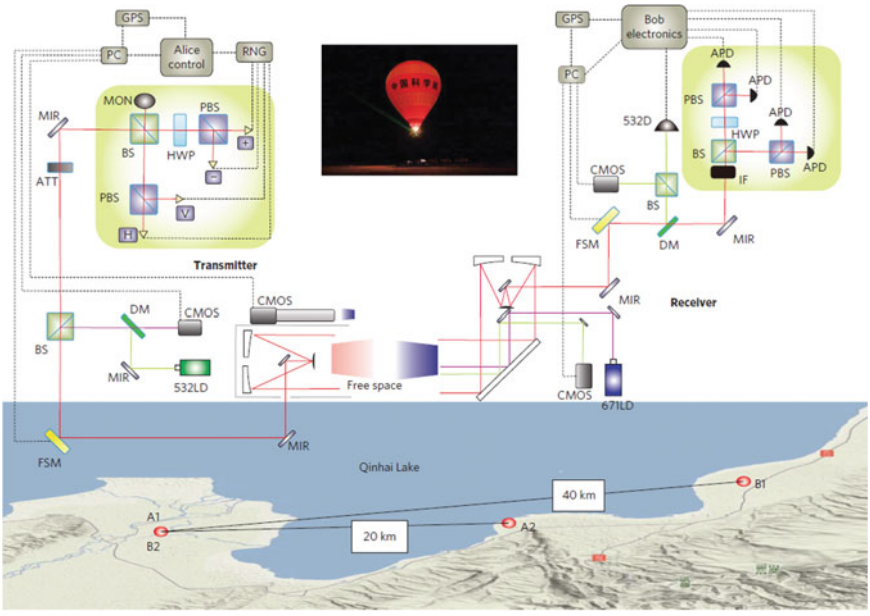


Fig. 18.4 Schematic of China's full-scale experimental verifications towards ground–satellite quantum key distribution

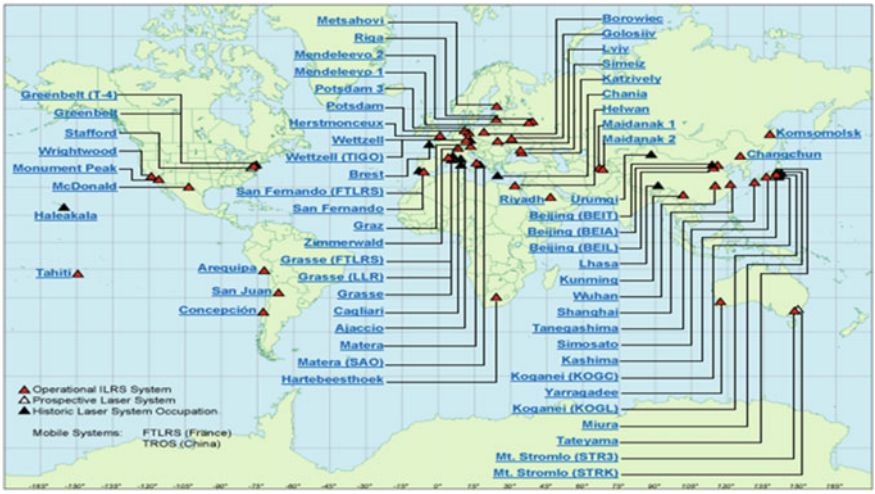


Fig. 18.5 Distribution of satellite ranging stations

Table 18.1 Technology achievements of space laser communication and ranging

Parameters	Terra-SAR	LLCD	OPALS	Alphasat
Link type	LEO–LEO	lunar-earth	LEO	GEO–GEO
Link distance(km)	1000–5100	400,000	700	45,000
Emitting wavelength (nm)	1064	1550	1550	1064
Divergence angle (mrad)	–	–	1–1.6	–
Peak data rate	5.625 Gbps	622 Mbps	30–50 Mbps	1.8 Gbps
Antenna size (mm)	125	100	–	135
Weight (kg)	32	30	–	50
Power (W)	–	140	–	160
Bit error rate	10^{-9}	–	10^{-4}	10^{-8}
Ranging accuracy (cm)	–	1	–	–

of the China Academy of Space Technology (CASC), etc. However, Institutions developing the unified laser communication and ranging technology are few.

The SLR stations in active use are shown in Fig. 18.5. The red triangles in Fig. 18.5 represent the station are currently working (44 stations, from ILS). The main results of space laser communication and ranging technology in recent years are shown in Table 18.1.

18.3.4 Summary of Development Trend in China and International

The Domestic deep space exploration starts relatively late and the overall development level are lagging behind that of the United States, Russia and other space powers. So far, the United States has completed almost all of the planetary exploration and other deep space activities within the solar system. The deep space exploration activities in our country are mainly highlighted on the moon and we have completed the phase targets of the round-moon flying and landing on the moon. The following engineering plan is being carried out too.

As the prober will fly deep into the distant space, the deep space TT&C task has confronted a series of problems, such as long-time period, long distance, long signal delay and weak signal received and so on. Deep space TT&C stations must receive signals from tens of billions of kilometers away in the distant deep space. Because of the signal loss in the path, it will be very weak when arriving at earth. It is well known that the power level of the microwave signal decays with the square of transmission distance. When the transmission distance is increased by N times, the power level will loss $20\log N$ dB. Therefore, relative to the earth synchronous orbit (GEO) distance which is 36,000 km, 2 million km distance defined by the deep space is 55.56 times of it, so the path loss is increased by about 35 dB; The distance of the moon's orbit to earth is about 0.348 million kilometers which is 10.67 times than it, so the path loss is about 20 dB; while the hot spot for deep space exploration-Mars which is the nearest to earth is about 59 million kilometers to earth and the farthest distance is about 400 million kilometers being 1655.56 times to 11111.11 times to the earth synchronous orbit, so the minimum path loss will be increased by about 65–81 dB. The distant between Jupiter in the deep space and the earth is about 1 billion km 27777.78 times to the earth synchronous orbit, so the path loss is even increased by about 89 dB.

For deep space exploration, the moon-earth deep space TT&C network for communication construction has the following gaps:

(1) The measurement accuracy is still unable to meet the higher requirements of scientific exploration

In terms of the angle measurement, although the ΔDOR interferometry is introduced and the measurement bandwidth reached 40 MHz in X frequency, the measurement accuracy of DOR is only 4 ns. Based on the measurement baseline of 4307 km from Kiamusze to Kashi, the position measurement error of the Mars probe is expected to be 173.4 km, which requires building large-scale antenna array or remote antenna array to support the demand for further deep space mission.

(2) The data transmission capability cannot meet the higher demands for data transmission rate

In terms of the data transmission, as deep space station in China with the largest diameter of 66 m, its data transmission capacity at 400 million km distance is only dozens of kbps.

At present, the foundation of deep space station is mainly based on the TT&C mode of radio frequency control and measurement model of large parabolic of radio frequency, it does not have the ability to carry out group array, deep space relay, and the integration of the space and the earth, laser measurement, control and communication and other new TT&C abilities.

(3) The key components/the autonomous control of components are relatively weak

According to the deep space network development plan of the United States from 2010 to 2030, the characteristics of NASA deep space missions are undergoing significant changes, so the deep space network must be adjusted to adapt to these changes. It is estimated that the future of the task will require the downlink transmission rate and increase its number to increase by 1 to 2 magnitude and the uplink data rates may also increase by 10 to 100 times. In order to meet the needs of rapid task increasing of NAS, the jet propulsion laboratory has developed a series of development plans and the key is highlighted on optimizing the structure, reduce the cost of operation and maintenance on the condition that the budget is being cut and improve the service ability. One of the essential parts of the development plan is to make the deep space network implement the Ka band reconstruction plan to improve the data transmission rate. When the development of the deep space network is carried out toward the higher frequency and higher performance, upgrading the key devices, realize the device independent shell and optimize the cost efficiency of the station is a key factor. Therefore, the high-power receiver, low-temperature receiver will also develop toward the direction of lower noise. Moreover, the development of superconducting nanowire single photon detector in optical communication has also become a key device in deep space exploration [23].

In terms of the high-power receiver, with the constant increasing of the output power of the transmitter, more and more difficult has been brought for the design of the core component of the transmitter-speed regulating tube. At the same time, the demand for the frequency quality of output signal and channel of transmission power is becoming higher and higher, that is to say, the power capacity will become a technical problem. It is expected in "13th Five-Year" period that the new requirements for the working frequency and the power level of the transmitter are put forward. In the Ka waveband, its emission power has been increased by nearly three times compared to the mature technology previously. Therefore, the device's power density is higher and it is more difficult for R&D; at the same time, as the device in the system will be closely connected with the antenna system, transmitter system, on the one hand, the device must be in smaller size so that the system integration is possible, on the other hand, it is required to increase the power density of the system to a large extent. If the power capacity is not enough, it is extremely easy to cause the spark discharge. In serious situation, it may cause damage to the

transmitter. Therefore, in view of the above difficulties, careful planning must be conducted with organized personnel to ensure the success of the device design and pass power verification in one time [24].

In terms of the cryogenic receiver technology in Ka waveband and millimeter wave, the domestic research starts late and due to the policies of restriction for domestic technology industry and other reasons, there is a large gap between the domestic and the foreign technologies in this regard. The technology gap is mainly reflected in the aspect that the core device of the domestic transmitter-low temperature amplifier has adopted the device of INP technology (there is embargo for this device in China) and the low temperature receiver with higher frequency uses SIS node technology (currently the experimental design is being made in China and foreign films are being implemented). The noise temperature of cooled receiver of JPL in Ka waveband is generally less than 20 K and the low temperature amplifier module is about 8.5 K.

In terms of the single photon detector of superconducting nanowire, the recovery time (20 ns) of unit detector developed domestically is relatively long, which cannot meet the requirements of modulation and detection signal of high-speed optical communication; moreover, the detector's photosensitive surface is small-sized (10 m) and it adopts the fiber coupling mode of single fiber with the core diameter of the fiber being only 9 mu. For larger imaging spot, the efficiency of light energy coupling is low and it is difficult to effectively receive the weak light signal under the condition of ultra-long distance.

(4) The laser communication is of high precision and its fast acquisition method is single. The orbit experience for satellite terminal design and the key technologies is less, so it is difficult to meet the actual demand and application of deep space communication

It should be noted that that in view of the development of laser communication technology at home and abroad, the development direction is to improve the system reliability, based on which, the distance and rate of transmission is increased. At the same time, it should be noted that foreign advanced technology has been supported for a long time and verified in the field of deep space measurement, control and communication. It can be predicted that in the future we will carry out a number of tasks for deep space laser measurement, control and communication.

In moon-earth laser linkage, the distance between the two terminals is 400,000 km or longer with large beam propagation delay and large relative motion of rate. Therefore, the strict requirements are put forward for the technology of rapid establishment and stabilization of linkage. In the stage of linkage establishment, the terminal uses the track, posture and other measurement data (such as microwave guidance) to calibrate the targeted coordinate system to reduce the uncertain angle of the acquisition. It also uses the bidirectional beam scanning acquisition strategy and non-beacon acquisition technique to carry out bidirectional, continuous and the synchronized scanning of laser to improve the acquisition ability of inter-satellite linkage in the delay network.

Based on the above key technologies, the R&D of the real-time coherent receiver should be carried out and the key technologies are verified and improved in the system. For the transmission terminal which requires a higher transmission power and signal to noise ratio, the production of wide band laser light source with narrow line, quadrature modulation and amplifier with high power and low noise should be further studied. The receiver requires high sensitivity and large dynamic range. The main research contents include the large dynamic range with function of automatic gain and control, low noise optical preamplifier, frequency mixing of relevant light and photoelectric conversion with high speed and response, high sampling rate and high resolution analog to digital conversion and FPGA circuit design of high speed and large scale.

At the same time, it should be noted that in order to aim at the high precision measurement and high data transmission. Each country has selected the high-performance telescope to carry out advanced ground test and star-earth verification experiment. Because of the short wavelength and small emission angle of the laser, we are required to consider more problems than that of the microwave communication system in the construction of a free-space optical communication system. In order to reduce the bit error rate caused by the alignment error and to improve the communication rate and efficiency, the tracking error of the optical ground station is required to keep a fraction of laser divergence angle. The divergence angle of the signal light on the star is usually within ten to tens of micro arc magnitude due to the limitation of the load volume, weight and power consumption of the stars. It requires the tracking precision of laser communication, optical ground station of ranging measurement integrated system and ATP system to achieve a few radians. This tracking accuracy is not possible for the system which uses a single telescope frame to realize the tracking. Therefore, each country has upgraded the original optical ground station to make it equipped with compound axis tracking system with acquisition, tracking and targeting functions.

Moreover, from the perspective of laser communication and the planning of ranging optical ground station, each country has proposed that the performance of the telescope should be improved by adding adaptive optics system, including improving image quality, tracking accuracy and coupling efficiency. The research of adaptive optics compensation mainly focuses on two aspects. The first aspect is the measurement of the phase distortion and the phase reconstruction is usually performed by measuring the wavefront slope. The second aspect is to implement wavefront phase correction by using a deformable mirror. ESA optical ground stations and JPL optical ground stations have carried out the corresponding adaptive optics correction experiment in order to suppress the influence of atmospheric turbulence and improve the link margin. JPL experiments show that the tracking telescope with adaptive optics system can enhance more than 5 dB in terms of the OOK laser linkage of 100 Mbps and the linkage margin.

Although China started late in this area, but after the investment and development being implemented, the majority of the key technologies have been successful or close to some breakthrough. We have also laid some certain foundation, so by drawing on lessons from foreign experience, the leapfrog development can be expected.

18.4 Related New Concept and Technology

18.4.1 IROC Technology

NASA has put forward the IROC project (Integrated Radio and Optical Communications, IROC) in its new version of 2015 technology roadmap (NASA Technology Roadmaps TA 5: Communications, Navigation, and Orbital Debris Tracking and Characterization Systems). Moreover, NASA GRC have researched the principle prototype and engineering prototype in recent years. Recently, it has also carried out a number of environmental adaptability experiments so as to use the prototype in future deep space exploration (Mars Exploration) with its weight being 4 kg and the power consumption being about 30–50 W [25].

The system also contains a radio-frequency network antenna which targets celestial bodies and whose diameter is 3 m and an optical antenna whose diameter is 30 cm. Ka waveband and laser communication will share integrated modulation and demodulation device defined by a software to break the bottleneck limitation of the non-beacon tracking technique based on data fusion, dual-mode antenna technology based RF/optical common aperture, resource optimization being not sufficient, targeting being difficulty, the transmission rate being low and weak robustness of the corresponding satellite-carrier equipment. The integrated RF and optical aperture, as shown in Fig. 18.6.

The technical parameters of IROC are summarized as Table 18.2.

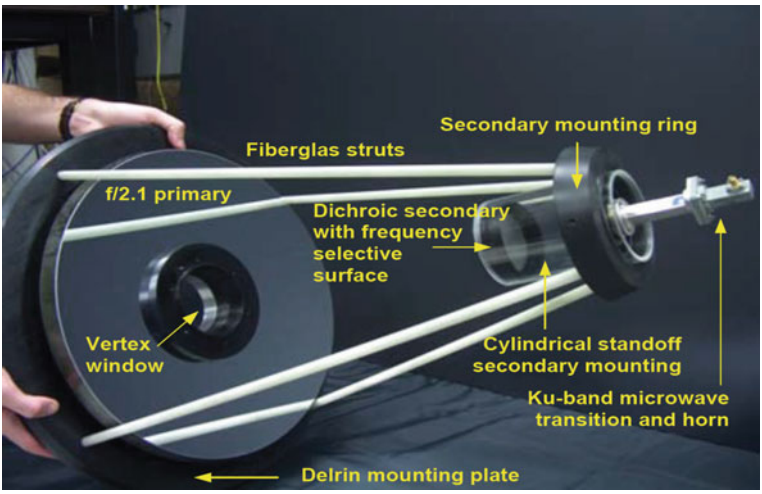


Fig. 18.6 Integrated RF and optical aperture. Schematic view is on the top and mock-up is on the bottom

Table 18.2 The technical parameters of IROC

Index	Optical	Raidio
1	Peak data rate: 267 Mbps	Peak data rate: 85 Mbps
2	Link Distance: 0.55 AU	Link Distance: 0.55 AU
3	Modulation: SCPPM-16	Modulation: GMSK
4	Code rate: 2/3	Coding: LDPC or Turbo
5	Guard time: 4	Code rate: 1/2
6	Slot clock: 0.5 ns	TWTA: 75 W
7	Antenna size: 0.3米	Antenna size: 3 m
8	Pulse repetition frequency: 10–100 MHz	Bandwidth: 200 MHz
9	Peak power: 700 W	Frequency: 32 GHz
10	Average power: 13 W	–
11	Wavelength: 1550 nm	–

18.4.2 Stable Phase Transmission Technology of Microwave Photonic RF Signals

At present, the TT&C stations in China are small-sized and the distance between the equipment room of time-frequency signal and equipment is not more than 100 m. If the distance between them is close, the transmission medium is mainly the coaxial cable, but the stability of time-frequency signal transmission is poor. When comprehensively carrying out the deep space exploration mission, it is inevitable to expand the TT&C station and the time-frequency signal should be transmitted to other new equipment and computer room at long distance. The existing time-frequency signal transmission system cannot meet the demand, which is a bottleneck for the location of the antenna array, room construction, so it is difficult to implement the ground station expansion task for deep space exploration. At the same time, considering its future application, in terms of the downlink antenna array, each antenna is manifested as paralleled and distributed linage, so there is higher demand for the relative stability and space stability. The array system requires all antenna signal phases are aligned, so the antenna front-end, distribution link and signal processing center must have very stable and efficient signal processing capabilities. Additionally, ESA and NASA have carried out a lot of experiments on the technology in deep space network and project verification. Japan, France and other developed countries have also exerted much effort in a timely manner; however, China has not yet fully grasped the technology, so the establishment of a new type of system for radio frequency signal transmission and compensation is imminent.

Since the end of 1970s, some foreign countries have carried out the research on the time frequency transfer by using the optical fiber with some important results being obtained [26]. To sum up, the research on the time frequency signal of the optical fiber transmission can be divided into two major aspects: the first category is no-compensation optical transmission of radio frequency signals and the second

category is the high-precision RF signal light. The first category basically does not consider the changes of the environment, machinery, transmission and the stability degree of no-compensation transmission frequency is about 2×10^{-13}. The second category takes some compensation and control measures, the daily stability degree of transmission are close to 10^{-17} magnitude. The uncertainty of the transmission delay can be controlled at the sub-picosecond level. To divide them from the perspective of methods, there are three categories which are respectively the stable phase transmission of direct light signal, the stable phase transmission of radio frequency signal light and the stable phase transmission by using the optical frequency comb [27–31].

In the national "fifteen five-year plan" pre-research project, the microwave photonic technology is used as a pre-research project with some research funds being invested. However, the research is mainly carried out on the basis of the application of radar while in the field of measurement, control and communication field, there is few report. The research or testing is not much in the aspect of the system level and the engineering application level, so there is a huge gap between China and the foreign countries. Therefore, we must carry out relevant research in this regard so as to support for the upcoming deep space exploration and lay the foundation for the application of antenna array and other fields.

18.5 Conclusions

This paper makes a brief summary of moon-earth, the demand and task of long-distance communication of moon-earth. It also introduces the development trend of moon-earth, the long-distance measurement, control and communication of moon-earth and some key technologies at home and abroad and summarizes the current technology gap between China and the foreign countries. By combining with the relevant technology situation in our country, it is suggested that the next step should be carried out based on the following four points:

(1) The overall design of study on high speed communication scheme of space-earth integration and ultra-long distance communication

The research is emphasized on high-speed TT&C and laser communication system of 400,000 mile Ka waveband. On such basis, it also conducts research on the deep space TT&C system with the effective distance being up to 6.47 AU. It completes the demand analysis, performance demonstration; sorts out the following month-earth deep space communication architecture with ultra-long distance. On such basis, by combining with China's national conditions, it carries out the overall technology research and makes breakthrough in key technologies for the deep space antenna array, self-controlled devices and so on.

(2) Acquisition and tracking of narrow beacon for moon-earth laser communication and high-sensitivity receiving technology

In moon-earth laser linkage, the distance between the two terminals is up to 400,000 km or more and the beam propagation delay is large, there are relative motions with large rate. Strict requirements are put forward for the rapid establishment and stabilization of linkage. The highlight of the research is high precision system of tracking, capture, targeting (ATP) in laser communication system of 400,000 miles system and high-sensitivity receiver technology. Moreover, it completes the demand analysis, the performance index demonstration, sorts out the follow-up overall plan and the structure design of ground, the satellite carrier equipment, making breakthrough of the key technologies.

(3) Equipment design and principle prototype research on light and small-sized Ka TT&C equipment

The main breakthrough to be made is lighted on the key technologies of light and small-sizes equipment, signal processing and space environmental adaptability of deep space transponder. According to the engineering conditions, the prototype machine research is carried out and creates conditions for the development of engineering products.

(4) Research on key devices and related new concepts of optical communication

We can draw lessons from the LLCD project of America and consider the current situation of related technology in China. On such basis, the single photon detector system of superconducting nanowire with high quantum efficiency, fast response speed, large receiving area and high coupling efficiency of optical fiber suitable for moon-earth laser signal detection. And it fully absorbs the integrated microwave/optical TT&C technology which is the latest technology developed by NASA for deep space exploration so as to further reduce the weight of the equipment on the satellite, power consumption and improve its redundancy, reliability of deep-space task.

Acknowledgements This project is supported by the innovation fund of The 10th Research Institute of China Electronics Technology Group Corporation (CETC) (Grant No. H15017).

References

1. Wu W, Liu W, Tang Y et al (2013) Development of deep space exploration and several key technique. Space Int 420(12):45–51
2. Lei L, Zhu Q (2014) Analysis of the trend of development of spacecraft TT&C and communication technologies and suggestions. J Spacecr TT&C Technol 33(6):463–468
3. Ye PJ, Yang MF, Peng J et al (2015) Review and prospect of atmospheric entry and earth re-entry technology of China deep space exploration. Sci Sin Technol 420(12):45–51
4. Ye P, Peng J (2006) Deep space exploration and its prospect in China. Eng Sci 8(10):13–18

5. Li H, Zhou H, Hao W et al (2013) Development of radio interferometry and its prospect in deep space navigation. J Spacecr TT&C Technol 32(06):470–478
6. Li J, Zhang J, Liu L et al (2012) VLBI technique applied in deep space exploration. Spacecr Eng 21(02):62–67
7. Tang S (2012) Radiometric measuring techniques for deep space navigation. National Defense Industry Press, Beijing, pp 87–96
8. Wu W, Yu D (2014) Development of deep space exploration and its future key technologies. J Deep Space Explor 1(1):5–17
9. Jiang H, An Y, Zhang Y et al (2015) Analysis of the status quo, development trend and key technologies of space laser communication. J Spacecr TT&C Technol 34(3):207–217
10. Grotzinger JP, Crisp J, Vasavada AR et al (2012) Mars science laboratory mission and science investigation. Space Sci Rev 170(1–4):5–56
11. Qi B, Chen H, Ren G et al (2013) ATP technology for 100-kilometer quantum entanglement distribution experiment. Optics Precis Eng 21(6):1628–1634
12. Wang J, Yang B, Liao S et al (2013) Direct and full-scale experimental verifications towards ground satellite quantum key distribution. Nat Photon 7:387–393
13. Thom S, Kenneth F, Raymond G et al (2008) A lunar communication satellite network architecture employing internet protocol, laser communication technologies and small satellites. In: Proceedings of 26th international communications satellite systems conference
14. Hemmati H (2001) Overview of laser communications research at JPL. In: Proceedings of SPIE 4273, the search for extraterrestrial intelligence (SETI) in the optical spectrum III
15. Burnside JW, Conrad SD, Pillsbury AD et al (2011) Design of an inertially stabilized telescope for the LLCD. In: Proceedings of SPIE 7923, free-space laser communication technologies XXIII
16. Hemmati H, Wright M, Sanii B, et al (2002) Multi-gigabit data-rate optical communications depicting LEO-to-GEO and GEO-to-Ground links. In: Proceedings of SPIE, vol. 4635, pp. 95–301
17. Fieldsa R, Lundea C, et al (2009) NFIRE-to-Terra SAR-X laser communication results: satellite pointing, disturbances, and other attributes consistent with successful performance. In: Proceedings of SPIE 7330, sensors and systems for space applications III
18. Zhang L, Du Z, Chen M et al (2016) Progress and trends of development of laser TT&C technology. J Spacecr TT&C Technol 35(1):010–020
19. Edwards BL, Robinson B, Biswas A, Hamkins J (2015) An overview of NASA's latest efforts in optical communications. In: 2015 IEEE international conference on space optical systems and applications
20. Roberts WT, Petrie HL, Pickles AJ, et al (2004) Feasibility of utilizing the 200-inch Hale Telescope as a deep-space optical receiver. In: SPIE international symposium on optical science and technology international society for optics and photonics
21. Boroson DM, Bondurant RS, Murphy DV (2004) A novel laser communications receiver array architecture. In: Proceedings of SPIE, free-space laser communications technologies XVI
22. Chapoton C, White J (2011) Deep-space optical communications. In: Communications satellite systems conference, international communications satellite systems conferences
23. Li H, Chen S, Lixing You et al (2016) Superconducting nanowire single photon detector at 532 nm and demonstration in satellite laser ranging. Opt Express 24(4):3535–3542
24. Wang Z, Liu M (2014) Ka-band cryogenic receiver technology. J Spacecr TT&C Technol 33 (3):225–230
25. Zeleznikar D, Nappier J, Downey J (2014) Ka-band link study and analysis for a mars hybrid RF/optical software defined radio. In: 32nd AIAA international communications satellite systems conference, SPACE conferences and exposition
26. Pan S, Zhu D, Liu S (2015) Satellite payloads pay off. IEEE Microw Mag 16(8):61–73
27. Wang B, Gao C, Chen WL et al (2011) Precise and continuous time and frequency synchronization at the 5×10^{-19} accuracy level. Sci Rep 2(2):556

28. Jiang W, Tan Q, Qin W et al (2015) A linearization analog photonic link with high third-order intermodulation distortion suppression based on dual-parallel Mach-Zehnder modulator. IEEE Photonics J 7(3):1–8
29. Maat DHP, Broekema PC, Schoonderbeek GW, Weiss J (2013) Multi-Tb/s data transport system for the square kilometre array radio telescope. In: 2013 IEEE Photonics Society summer topical meeting series
30. Cliché J-F, Shillue B (2006) Precision timing control for radio astronomy maintaining femtosecond synchronization in the atacama large millimeter array. IEEE Control Syst Mag 22(02):19–26
31. Olivier L, Haboucha Adil, Kéfélian Fabien et al (2014) Cascaded multiplexed optical link on a telecommunication network for frequency dissemination. Opt Express 18(16):16849–16857

Chapter 19
An Improved MFSK Signal Detection Algorithm for Mars Probe Entry, Descent, Landing Phase

Tiansheng Zhang, Xiaolin Zhang, Zan Li and Junhai Bao

19.1 Introduction

The EDL (entry, descent, landing) phase of the Mars probe starts from the beginning of entering into the Martian atmosphere to the landing. During the EDL phase, it is very important to maintain communications from the spacecraft to the earth. This is especially true in case of a mission failure, when the diagnostic data would be very important. In 1999, the United States launched the "Mars Polar Lander" (MPL), however, there is no communication system designed for EDL process due to funding constraints. After the task failed, it is difficult to find the prime reasons [1]. Since then, two mars rovers (MERs) were launched by the United States in 2003, the Phoenix in 2007, and the Mars Science Laboratory (MSL) in 2012, and all of them used a special form of MFSK modulation to send messages from the spacecraft to the earth. Before Entry and until the lander is separated from the backshell, communication is by a direct-to-Earth (DTE) X-band (8.4 GHz) link. After separation the backshell antenna can no longer be used, and communication is achieved using two links: a main UHF relay link to the Mars Odyssey or Mars Global Surveyor spacecraft, and a backup DTE link using the rover antenna. The DTE link uses the MFSK modulation to transmit one out of 256 possible data tones every 10 s.

To detect MFSK signal correctly, three main procedures are included, which are consisted of carrier acquisition, carrier tracking and data tone calculation. The carrier acquisition process is very important, because that the subsequent procedures are depended on the results of the carrier acquisition. The carrier acquisition

T. Zhang (✉) · X. Zhang · J. Bao
Beihang University, Beijing 100191, China
e-mail: zhang_tian_sheng@163.com

Z. Li
Beijing Institute of Tracking and Telecommunication Technology,
Beijing 100094, China

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_19

algorithms for processing signal with extremely low signal noise ratio and high dynamics mainly include: ① Maximum likelihood (ML) criterion estimation algorithm: the traditional ML algorithm is a nonlinear optimization problem based on multi-dimension search, which is difficult to be applied in practice. A parameter estimation algorithm based on the sum of weighted phase and 1-order or 2-order phase differences is proposed in the paper [2], which simplified the traditional ML algorithm. A discrete high-order phase function method is proposed in the paper [3], which can be used for the signal estimation with higher-order Doppler dynamics, but it is applicable only when signal noise ratio is not too low. ② Maximum likelihood algorithm with FFT: The time-domain matching-average periodogram algorithm (TDMAP) is proposed in the paper [4] and paper [5]. This algorithm reduces the accuracy requirement of the Doppler rate matching as well as the computation complexity, which is suitable for carrier acquisition. However, when the dynamic range is expanded, the matching branch will increase correspondingly, which will cause the computation complexity to be enlarged largely, and make it more difficult to do real-time carrier acquisition. Aiming at the problem that the computation complexity of the traditional time-domain matching algorithm is high, an improved algorithm is proposed in the paper [6] and paper [7], which decreased the computation complexity without performance loss by padding extra zeros to the signal. By analyzing the ML algorithm, an algorithm named carrier recovery enhancement for maximum-likelihood Doppler shift estimation (CRE-ML) is proposed which takes into account the power in the data tones to enhance carrier recovery. The algorithm reduces the signal detection threshold by 3 dB, but the complexity of the data processing is increased by about 2000 times.

In this paper, we studied the algorithm principle in paper [8] and paper [4], and analyzed the relationship between the four search dimensions and carrier recovery enhancement in CRE-ML algorithm. We applied the principle of 4-dimensions search into the TDMAP and proposed the time-domain matching-average periodogram algorithm based on 4-dimensions searching (4D-TDMAP). The computation complexity of 4D-TDMAP is only about 0.26% of that of CRE-ML algorithm with the same signal detection performance. By adding two search dimensions which are the data tone and the phase of data tone, we make full use of the power in data tones as well as in carrier. The signal detection threshold using 4D-TDMAO decreased by about 3 dB compared with the traditional TDMAP. Referring to the dynamics during EDL of Mars Science Laboratory, we verify the performance of the 4D-TDMAP (computation complexity, signal detection threshold) by doing experiments under the condition of high dynamics and low signal noise ratio. The experiment results are in agreement with the theoretical analysis.

19.2 The Analysis of MFSK During EDL

The DTE link uses the MFSK modulation to transmit one out of 256 possible data tones every 10 s. The transmitted EDL signal is represented mathematically as follows [4, 5]:

$$s(t) = \sqrt{2P_{\mathrm{T}}}\cos\left(2\pi \cdot f_{\mathrm{c}}^{0} t + \Delta \cdot \mathrm{Sqr}\left(2\pi \int_{-\infty}^{t} \mathrm{d}\tau \cdot f_{\mathrm{d}}(\tau)\right)\right) \tag{19.1}$$

where P_{T} is the transmitted power, f_{c}^{0} is the transmitted carrier frequency, f_{d} is the data tone frequency (1 of 256 possible tones transmitted every 10 s), Δ is the modulation index, $\mathrm{Sqr}(\cdot)$ is the hard-limited sine function.

$$\mathrm{Sqr}(x) = \begin{cases} 1, & 0 < x \le \pi \\ -1, & \pi < x \le 2\pi \end{cases} \tag{19.2}$$

$$\mathrm{Sqr}(x) = \mathrm{Sqr}(x + 2\pi) \tag{19.3}$$

$$f_{\mathrm{d}}(t) = f_{\mathrm{dinit}} + n_{\mathrm{d}}(t) \cdot f_{\mathrm{interval}} \quad , \quad n_{\mathrm{d}}(t) \in [1, 256] \tag{19.4}$$

where f_{dinit} is the initial data tone frequency, f_{interval} denotes the frequency interval between two data tones, $n_{\mathrm{d}}(t)$ is the number of transmitted data tone.

The received EDL signal is represented mathematically as follows:

$$r(t) = \sqrt{2P_{\mathrm{R}}(t)}\cos\left(2\pi \int_{t_0}^{t} \mathrm{d}\tau \cdot f_{\mathrm{c}}(\tau) + \Delta \cdot \mathrm{Sqr}\left(2\pi \int_{t_0}^{t} f_{\mathrm{d}}(\tau) d\tau\right)\right) + v(t) \tag{19.5}$$

where $v(t) = v_{\mathrm{I}}(t)\cos(2\pi f_{\mathrm{c}}^{0} t) - v_{\mathrm{Q}}(t)\sin(2\pi f_{\mathrm{c}}^{0} t)$ is equal to $\mathrm{Re}\left\{\tilde{v}(t)\mathrm{e}^{j2\pi f_{\mathrm{c}}^{0} t}\right\}$, $f_{\mathrm{c}}(t) = f_{\mathrm{c}}^{0} + \Delta f_{\mathrm{c}}(t)$. The received signal $r(t)$ is a bandpass signal with the center frequency of f_{c}^{0}. Its mathematical expression is:

$$\begin{aligned} \tilde{r}(t) = \sqrt{2P_{\mathrm{R}}(t)}\exp\Bigg(&-\mathrm{j}2\pi f_{\mathrm{c}}^{0} t_0 + \mathrm{j}2\pi \times \int_{t_0}^{t} \triangle f_{\mathrm{c}}(\tau)\mathrm{d}\tau \\ &+ \mathrm{j}\Delta \cdot \mathrm{Sqr}\left(2\pi \int_{t_0}^{t} f_{\mathrm{d}}(\tau) d\tau\right)\Bigg) + \tilde{v}(t) \end{aligned} \tag{19.6}$$

where $\tilde{r}(t)$ is the lowpass equivalent of the received signal $r(t)$. $\tilde{v}(t)$ is the lowpass equivalent of the received signal $v(t)$. The variance of the complex $\tilde{v}(t)$ is $2\sigma_{v}^{2}$.

19.3 The Analysis of CRE-ML

We can express $\tilde{r}(t)$ in Eq. (19.6) in discrete-time by sampling at the rate $F_s = 1/T_s$, and use $\tilde{r}(nT_s)$, $\tilde{v}(nT_s)$ to refer to $\tilde{r}(n)$, $\tilde{v}(n)$ repectively, then we can get [8, 9]:

$$\tilde{r}(n) = A(n)\exp\{\mathrm{j}\phi(n)\} + \tilde{v}(n) \tag{19.7}$$

$$\phi(n) \triangleq 2\pi \int_{t_0}^{nT_s} \Delta f_c(\tau)d\tau + \Delta \cdot \mathrm{Sqr}\left[2\pi \int_{t_0}^{nT_s} f_d(\tau)d\tau\right].$$

We assume that data tone frequency is constant ($f_d(t) = f_d$), so we can get:

$$\mathrm{Sqr}\left[2\pi \int_{t_0}^{nT_s} f_d(\tau)d\tau\right] = \mathrm{Sqr}(2\pi f_d n/F_s + \varphi) \tag{19.8}$$

where $\varphi = 2\pi\left[\int_{t_0}^{mT_s} f_d(\tau)d\tau - f_d m T_s\right]$ is an unknown phase.

We assume: ① The noise $\tilde{\mathbf{v}}(n)$ is complex zero-mean circular Gauss with variance is $2\sigma_v^2$; ② The data is analyzed in segments of $N \times M$ samples. Typical values are $N = 10000$, $M = 10$, and $F_s = 100$kHz, therefore, an analysis segment duration of T = 1 s. During each segment, we assumed that the Doppler shift f_0 and the Doppler rate $\dot{f}$ are constant, and can model them as unknown parameters. ③ The complex amplitude $A(n)$ can be modeled as piecewise constant deterministic parameter. Assuming that $A(n)$ is a constant in intervals of duration N samples, and the amplitude corresponding to the mth segment is A_m . In order to ensure an average signal power of $2P_R$, the $\frac{1}{M}\sum_{m=0}^{M-1}|A_m|^2 = 2P_R$ is needed.

Under this assumption, the probability density function of the received signal is:

$$\begin{aligned} f_{\tilde{\mathbf{r}}}(\tilde{r}(n)) &= f_{\tilde{\mathbf{v}}}(\tilde{r}(n) - A(n)\mathrm{e}^{\mathrm{j}\phi(n)}) \\ &= \frac{1}{(2\pi\sigma_v^2)^{N\times M}} \times \exp\left\{-\frac{1}{2\sigma_v^2}\sum_{m=0}^{M-1}\sum_{n=mN}^{nN+N-1}\left|\tilde{r}(n) - A_m\mathrm{e}^{\mathrm{j}\phi(n)}\right|^2\right\} \end{aligned} \tag{19.9}$$

By taking the natural logarithm, we arrive at the following log-likelihood function:

$$-\frac{1}{2\sigma_v^2}\sum_{m=0}^{M-1}\sum_{n=mN}^{nN+N-1}\left|\tilde{r}(n) - A_m\mathrm{e}^{\mathrm{j}\phi(n)}\right|^2 - \log(2\pi\sigma_v^2)^{N\times M} \tag{19.10}$$

The maximum likelihood solution of Let the problem (19.10) is:

$$\max_{f_0,\dot{f},f_d,\varphi}\sum_{m=0}^{M-1}\left|\sum_{n=mN}^{mN+N-1}\tilde{r}(n)\mathrm{e}^{-\mathrm{j}2\pi(\frac{f_0 n}{F_s}+\frac{\dot{f}n^2}{2F_s^2})-\Delta\cdot\mathrm{Sqr}(\frac{2\pi f_d n}{F_s}+\varphi)}\right|^2 \tag{19.11}$$

We will refer to (19.11) as the CRE-ML method. CRE-ML is more complex because we have to search over the phase φ and the data tone frequency f_d besides Doppler frequency f_0 and Doppler rate $\dot{f}$, which will increase the computation complexity about 2000 times and also will decrease the threshold by about 3 dB.

In this paper, to solve the high computational complexity problem of CRE-ML, we analyze the relationship between four searching dimensions and carrier recovery enhancement, and propose the 4D-TDMAP algorithm which maintains the performance of CRE-ML and reduces the computation complexity.

19.4 The Analysis of 4D-TDMAP

From the Eq. (19.5), we can see that the $f_c(t)$ is the time-varying carrier frequency, which can be seen as the sum of initial carrier frequency f_c^0 and the unknown Doppler frequency shift $\Delta f_c(t)$. The Taylor expansion of $\Delta f_c(t)$ is [10]:

$$\Delta f_c(t) = f_0 + \dot{f}\cdot t + \ddot{f}\cdot\frac{t^2}{2}+\cdots \tag{19.12}$$

where f_0 is Doppler frequency offset, $\dot{f}$ is the first derivative of Doppler frequency, $\ddot{f}$ is the second derivative of Doppler frequency.

$$\begin{aligned} x_{l,q,w}(t) = r(t)\cdot\exp\left\{-2\pi\mathrm{j}\left(\hat{f}_c^l(t)\cdot\tfrac{t^2}{2}\right)-\mathrm{j}\left(\Delta\cdot\mathrm{Sqr}\left(2\pi f_d^q t+\varphi_w\right)\right)\right\} \\ 0\le t\le T;1\le l\le N_r;1\le q\le 256;1\le w\le 8 \end{aligned} \tag{19.13}$$

where $\hat{f}_c^l(t)$ is the l-th one of N_r hypothesized carrier frequency rate. T is the time span of the segment. f_d^q is one of the 256 data tone frequencies which are searched in formula (19.4). φ_w is unknown data tone phase in formula (19.8). Usually we take 45° as step and search 8 phases in total. Each segment $x_{l,q,w}(t)$ is further divided into M continuous sub-segments. The duration of each sub-segment is $\Delta T = T/M$ and the data length is $N_{FFT} = F_s\cdot\Delta T$. The magnitude-squared FFTs for each l, q, w th rate, data tone and data tone phase are averaged together to form the periodogram estimates $P_{l,q,w}(f_k)$, $1\le l\le N_r$, $1\le q\le 256$, $1\le w\le 8$; $f_k\equiv(k/\Delta T)([-N_{FFT}/2]+1\le k\le[-N_{FFT}/2])$. At last, according to the branch and position of the maximum value of $P_{l,q,w}(f_k)$, we can get the estimated Doppler frequency, Doppler rate of change and data tone. 4D-TDMAP Algorithm principle is shown in Fig. 19.1.

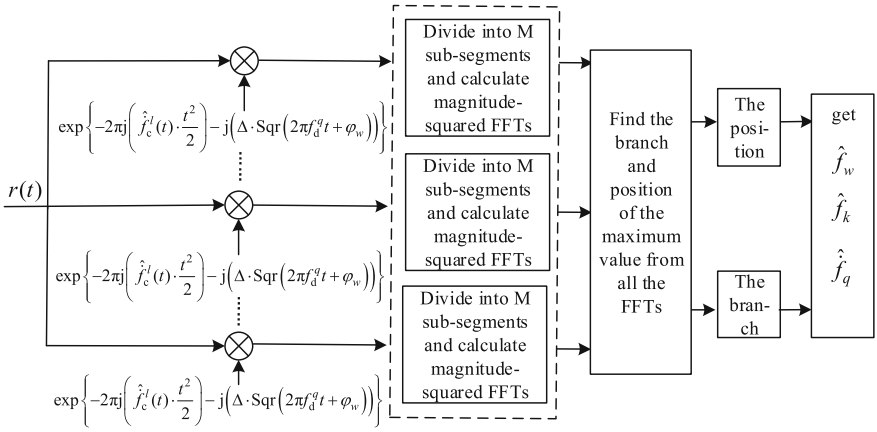


Fig. 19.1 Algorithm of time domain matching based on four dimensional search algorithm

$$\left.\begin{matrix}\hat{f}_k \\ \hat{f}_l \\ \hat{f}_q \\ \hat{\varphi}_w\end{matrix}\right\} = \arg\max_{l,q,w,f_k}\left\{P_{l,q,w}(f_k)\right\} \tag{19.14}$$

19.5 Simulation and Verification

19.5.1 Simulation Parameter Setting

According to the paper [12, 13], the range of Doppler frequency shift of the DTE X-band (8.4 GHz) link during EDL is about 90 kHz. Positive Doppler frequency is about 50 kHz. The maximum of Doppler rate is about 700 to 1200 Hz/s. The range of the second derivative of Doppler frequency is about −25 to 40 Hz/s^2. We assume that one of 256 different data tones are transmitted every 10 s. The parameters of received MFSK signal are shown in Table 19.1

Doppler frequency, Doppler rate, the second derivative of Doppler frequency and other parameters all refer to the dynamics of MER and MSL in paper [4, 5]. Assume that the initial Doppler frequency is 20 kHz and initial Doppler rate is −250 Hz/s. Generate received MFSK signal transmitted from spacecraft during EDL according to the formula (19.5). Simulated Doppler dynamics profiles and data tones during EDL are shown in Fig. 19.2.

The high dynamics between 150 s and 300 s is the result of intense atmospheric friction when the Mars probe enters the Martian atmosphere. The spike around 500 s in Fig. 19.2b, c is a violent shake when the parachute is deployed. Detect the received MFSK signal using CRE-ML in paper [8], TDMAP in paper [4, 5] and

Table 19.1 Parameter setting of received signal in EDL process

Sampling rate	100 kHz
Doppler frequency	0–50 kHz
Doppler frequency rate	−1000 to 1000 Hz/s
Second derivative of Doppler frequency	−30 to 30 Hz/s^2
Modulation index Δ	48°
Data tone	1–256
Initial data tone frequency	2000 Hz
Frequency interval of data tone	80 Hz

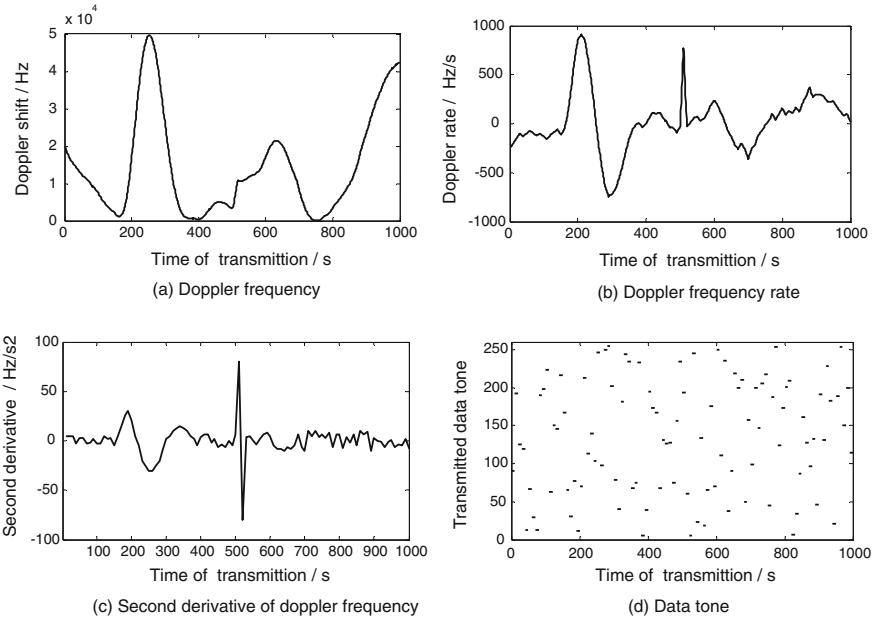


Fig. 19.2 Simulated Doppler dynamics profiles and data tones during EDL

4D-TDMAP in this paper respectively, and compare the performance of the three algorithms.

19.5.2 The Simulation Results

Do simulation using the baseband signal. The sampling rate is F_s = 100 kHz, $N_{FFT} = 20000$. The number of sub-segments is $M = 5$. Doppler frequency resolution is 5 Hz. The results of MFSK signal detection under the same dynamics

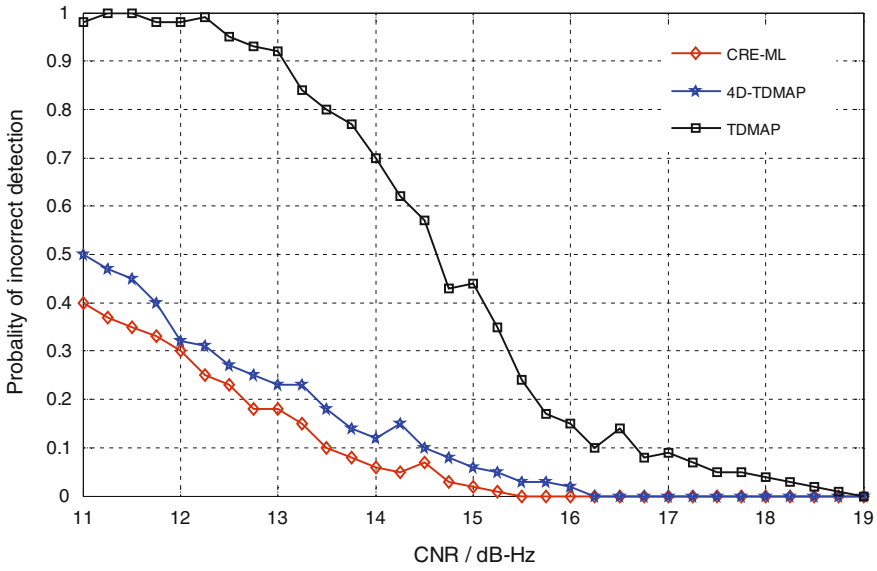


Fig. 19.3 The results of MFSK signal detection under the same dynamics using CRE-ML, TDMAP and 4D-TDMAP

using CRE-ML, TDMAP and 4D-TDMAP are shown in Fig. 19.3. We can come to the conclusion that two algorithms of CRE-ML and 4D-TDMAP have the similar threshold which is reduced by approximately 3 dB respect to TDMAP algorithm.

We divide the Doppler dynamics into four levels according to the value of Doppler frequency rate during EDL. The ranges of absolute value of Doppler rate for the four levels are 0–100, 100–300, 300–500 and 500–1000 Hz/s respectively. The results of MFSK signal detection under the four-level dynamics using 4D-TDMAP and TDMAP are shown in Fig. 19.4a–d. We can come to the conclusions as follows. Firstly, The threshold of MFSK signal detection under higher-level Doppler dynamics is greater than that under lower-level Doppler dynamics using 4D-TDMAP and TDMAP respectively. Secondly, the threshold of MFSK signal detection under four-level Doppler dynamics using TDMAP is larger than that under the same Doppler dynamics using 4D-TDMAP by about 3 dB.

From Figs. 19.3 and 19.4, we can get that the experiment results for 4D-TDMAP are consistent with the theoretical analysis.

19.5.3 The Analysis of Computation Complexity for CRE-ML and 4D-TDMAP

The sampling rate is $F_s = 100\text{kHz}$, and the duration of each segment is 1 s. Each segment will be divided into M sub-segments, thus the length of each sub-segment

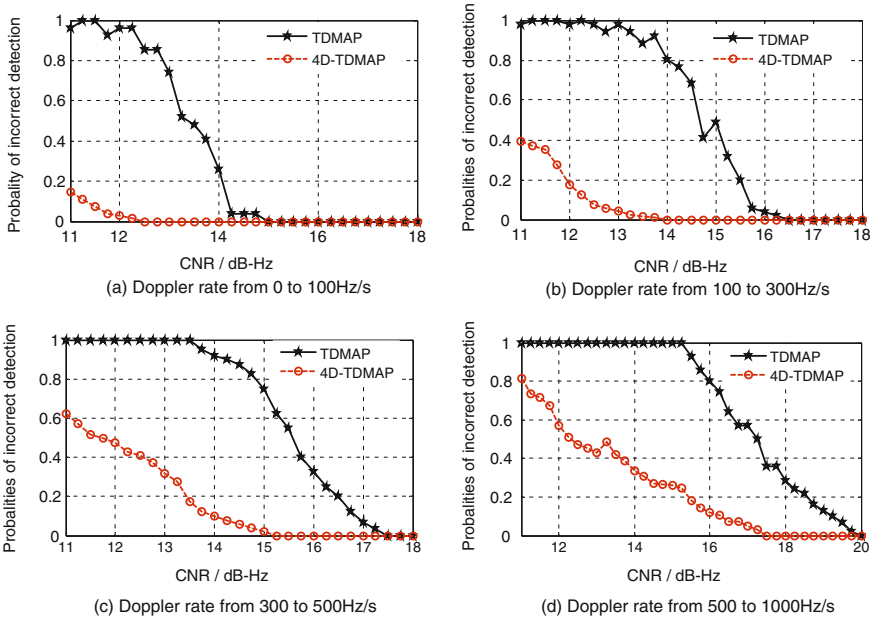


Fig. 19.4 The results of MFSK signal detection under the four-level dynamics using 4D-TDMAP and TDMAP

Table 19.2 The number of complex additions and complex multiplications for CRE-ML 4D-TDMAP

Algorithms	Number of complex multiplications	Number of complex addition
CRE-ML	$M \cdot N \cdot N_{f_0} \cdot N_{\dot{f}} \cdot N_{f_\mathrm{d}} \cdot N_\varphi$	$(M \cdot N - 1) \cdot N_{f_0} \cdot N_{\dot{f}} \cdot N_{f_\mathrm{d}} \cdot N_\varphi$
4D-TDMAP	$M \cdot [N] \cdot \log_2^{[N]} \cdot N_{\dot{f}} \cdot N_{f_\mathrm{d}} \cdot N_\varphi$	$\left(M \cdot [N] \cdot (\log_2^{[N]} + 1) + (N - 1)\right) \cdot N_{\dot{f}} \cdot N_{f_\mathrm{d}} \cdot N_\varphi$

is $N = F_s/M$. N_{f_0} , $N_{\dot{f}}$, N_{f_d} , N_φ represents Doppler frequency, Doppler rate, number of total data tone and data tone phase. f_{res} , $\dot{f}_{\mathrm{res}}$ represents Doppler frequency resolution and Doppler rate resolution. According to Eqs. (19.11) and (19.13), the number of complex additions and complex multiplication for CRE-ML and 4D-TDMAP is shown in Table 19.2.

In the table, [N] represents the minimum 2^n Greater than N. Let $f_{\mathrm{res}} = 5\mathrm{Hz}$, $\dot{f}_{\mathrm{res}} = 10\mathrm{Hz/s}$, $N = 20000$, $N_{f_0} = F_\mathrm{s}/2 \cdot f_{\mathrm{res}} = 10000, N_{\dot{f}} = 57, N_{f_\mathrm{d}} = 256, N_\varphi = 8$. We can get that the numbers of complex multiplications for CRE-ML and 4D-TDMAP are 11.6736 trillion times and 28.689 billion times. The numbers of complex additions for CRE-ML and 4D-TDMAP are approximately 11.6736 trillion times 30.835 billion times. So we can get the conclusion that the computation complexity of 4D-TDMAP decreased to 0.26% of CRE-ML.

19.6 Conclusion

The algorithm named 4D-TDMAP is proposed in this paper which searches over the phase φ and the data tone frequency f_{d} besides Doppler frequency f_0 and Doppler rate $\dot{f}$. This algorithm can make full use of the power of data tone as well as the carrier, which results in the threshold of MFSK signal detection using 4D-TDMAP decreased by 3 dB compared with TDMAP. The computation complexity of 4D-TDMA decreased to 0.26% of CRE-ML with almost no performance degradation. In spite of this, it is difficult to detect MFSK signal in real time, and 4D-TDMAP is still needed to be improved.

References

1. Brian CS, Melissa S, Peter I (2014) Telecommunications performance during entry, descent, and landing of the mars science laboratory. J Spacecr Rockets 51(4):1237–1250
2. Li Y, Fu H, Kam PY (2009) Improved, approximate, time-domain ML estimators of chirp signal parameters and their performance analysis. IEEE Trans Signal Process 57(4):1260–1272
3. Farquharson M, O'Shea P, Ledwich G (2005) A computationally efficient technique for estimating the parameters of polynomial-phase signals from noisy observations. IEEE Trans Signal Process 53(8):3337–3342
4. Satorius E, Estabrook P, Wilson J et al (2003) Direct-to-earth communications and signal processing for mars exploration rover entry, descent, and landing. Interplanet Netw Prog 153:42–153
5. Soriano M, Finley S, Fort D et al (2013) Direct-to-earth communications with mars science laboratory during entry, descent, and landing. In: IEEE aerospace conference, Big Sky, MT: AERO, pp. 1–14
6. Duan RF, Liu RK, Zhou Y et al (2013) A Low-complexity coarse carrier acquisition algorithm for signals with extremely low signal noise ratio and high dynamics. Acta Aeronaut Astronaut Sin 34(3):662–669
7. Gan H, Zhang XL, Ma YH et al (2015) An improved frequency acquisition algorithm for high dynamic and weak signal in deep space. J Telem Track Command 36(4):6–11
8. Cattivelli FS, Estabrook P, Satorius EH et al (2008) Carrier recovery enhancement for maximum-likelihood doppler shift estimation in mars exploration missions. IEEE J Sel Topics Signal Process 2(5):658–669
9. Pan X (2010) ALE-FFT algorithms for weak signal acquisition. In: IEEE international symposium on intelligent signal processing and communication systems (ISPACS 2010), pp. 1–4
10. Vilnrotter VA, Hinedi S, Kumar R (1989) Frequency estimation techniques for high dynanic trajcetories. IEEE Trans Aerosp Electron Syst 25(4):559–577
11. Liu S, Huang XF, Qiao XJ (2015) Telecommunication system scheme for mars probe during EDL. Spacecr Eng 24(4):94–101
12. Kornfeld RP, Prakash R, Devereaux AS et al (2014) Verification and validation of the mars science laboratory/curiosity rover entry, descent, and landing system. J Spacecr Rockets 51(4):1251–1269
13. Soriano M, Finlcy S, Jongcling A et al (2012) Spacecraft to-earth communications for Juno and Mars Science Laboratory critical events. In: 2012 IEEE aerospace conference. New York, IEEE

Chapter 20
Robust Fault Detection for a Spacecraft with Lipschitz Nonlinear System

An Liu, Zhibin Wu and Dong Han

20.1 Introduction

Observer is widely used in fault detection, isolation and reconstruction of dynamic systems [1]. With the development of control theory, the control methods such as sliding mode and adaptive control have been successfully applied to the fault diagnosis method based on the observer [2]. Adaptive method can effectively overcome the linear parameter uncertainty in the system, and sliding mode control can be effectively used to deal with the external disturbance of the system. Therefore, the observer based on the sliding mode is more suitable for the fault diagnosis of the system with external disturbance.

For the linear time invariant system, sliding mode observer for fault detection is proposed by Edwards et al. [3], they reconstruct actuator and sensor faults based on sensor fault, however it is only suitable for slow time-varying sensor fault detection. Alwi et al. build the augmented system based on sliding mode, and further extended the slow time-varying fault detection to reconfigurable general time variable sensor fault detection and isolation methods [4]. However, the literatures [3, 4] are not taking into account a variety of inevitable external interference in the actual application system. In view of the nonlinear system with external disturbance, various Lipschitz nonlinear systems are discussed in the existing literatures. For a class of Lipschitz nonlinear systems, the literature [5, 6] are proposed respectively fault reconstruction methods based on the sliding mode observer for actuator. Yan et al. obtain two subsystems by using the equivalent state transformation, one of the subsystems including external disturbances and actuator fault, and the other

A. Liu (✉) · Z. Wu
Key Laboratory of Diagnosis and Maintenance of Spacecraft in Orbit,
Xi'an 710043, China
e-mail: 15191681263@163.com

D. Han
Beijing Information Science and Technology University, Beijing 100192, China

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_20

subsystems decoupling from external disturbances and actuator fault; then, the sliding mode observer is designed for the subsystem includes external disturbances and actuator faults, and the actuator fault is reconstructed [5, 6]. However, a strong mathematical hypothesis is needed in the state estimation to reconstruct the fault mode in this method, which limits the use of proposed method in literatures [7, 8].

The results of fault detection for Lipschitz nonlinear systems with unknown disturbances are not seen based on the sliding mode observer. Although proposed method in [5, 6] can be directly used for fault detection, it is not necessary to reconstruct assumptions. In this paper, a fault detection method based on sliding mode observer is proposed for a class of Lipschitz nonlinear systems with unknown input disturbances. Due to the difficulty of fault detection is less than fault reconstruction, the proposed method has relaxed the mathematical hypothesis in [5, 6] which is necessary for the state estimation.

In this paper, we use the idea of equivalent decomposition to decompose the original system into two subsystems, in which the first subsystem is included external disturbance and actuator fault similar to the [5, 6], while the other subsystem does not contain external interference and contains the actuator fault which is different with [5, 6]. According to the two subsystems, the sliding mode observers are respectively designed to track the state of the subsystem asymptotically when the actuator fails to occur. Finally, we apply the proposed method to a spacecraft actuator fault detection based on a simplified simulation model under condition of high attack angle.

20.2 Problem Description

A class of Lipschitz nonlinear systems with external disturbances is considered.

The linear discrete system is considered as follows.

$$\begin{cases} \dot{x} = Ax + \Phi(x, u) + E\eta + Df_a \\ y = Cx \end{cases} \tag{20.1}$$

In which, $x \in \mathbf{R}^n$ is system state vectors, $y \in \mathbf{R}^p$ is observable output, $u \in \mathbf{R}^m$ is the known input, $\eta \in \mathbf{R}^r$ is the external disturbance of the system, $f_a \in \mathbf{R}^q$ is the actuator failure. System matrix $A \in \mathbf{R}^{n\times n}$, $E \in \mathbf{R}^{n\times r}$, $D \in \mathbf{R}^{n\times q}$, and $C \in \mathbf{R}^{p\times n}$ are constant matrix. Matrix E and D are for full column rank, C is for full row rank, and $r \le p \le n$.

In order to facilitate the subsequent analysis, the system (20.1) does the following assumptions [9, 10].

Hypothesis 1 The invariant zeros of $\{A, C, E\}$ are located in the left half plane of the complex plane.

Hypothesis 2 $\text{rank}(E) = \text{rank}(CE)$

Hypothesis 3 Nonlinear term $\Phi(x,u)$ is to satisfy the Lipschitz condition, that is, $\|\Phi(x,u)-\Phi(\hat{x},u)\| \le L_\Phi \|x-\hat{x}\|$, L_Φ is a Lipschitz constant, $\|\cdot\|$ is 2- norm of vector.

Hypothesis 4 External disturbance η is norm bounded, that is, the existence of $b_1 > 0$ makes $\|\eta\| \le b_1$ set up.

For systems subject to external disturbances, assumptions 1 and 2 are necessary conditions for the design of the observer. Without loss of generality, as shown in the literature [3, 6], it is also assumed that the output matrix of the system (20.1) has the following form $C = [I_p \quad 0]$, there, I_p is a unit matrix with p order. For other forms matrix C with full row rank can be turned into this special form through a simple equivalent transformation.

Here the lemma is as following.

Lemma 1 (Gronwall-Bellman inequation) [11]

If a continuous function $f:[a \quad b] \to \mathbf{R}$ *is not negative, and a continuous function* $g:[a \quad b] \to \mathbf{R}$ *satisfies* $g(t) \le c + \int_a^t f(s)g(s)ds$, then $g(t) \le ce^{\int_a^t f(s)ds}$, *there*, c *is a constant*.

Lemma 2 [12] *Based on hypothesis 1 and 2, in view of the specific positive matrix* $Q \in \mathbf{R}^{n\times n}$, *there is a matrix* $L \in \mathbf{R}^{n\times p}$, $F \in \mathbf{R}^{r\times p}$ *and positive matrix* $P \in \mathbf{R}^{n\times n}$, *to make the following equality holds*.

$$(A-LC)^T P + P(A-LC) = -Q \tag{20.2}$$

$$E^T P = FC \tag{20.3}$$

20.3 Design of Sliding Mode Observer

In this section, a sliding mode observer is proposed to design for the system (20.1). This observer is robust to external disturbance of the system, when the system has no fault, it can guarantee the asymptotic convergence of the state observation error; and the system fails, it can ensure the sensitivity of the system fault.

20.3.1 System Equivalent Decomposition and Sliding Mode Observer Design

System (20.1) can be rewritten as following.

$$\begin{cases} \dot{x}_1 = A_{11}x_1 + A_{12}x_2 + \Phi_1(x,u) + E_1\eta + D_1 f_a \\ \dot{x}_2 = A_{21}x_1 + A_{22}x_2 + \Phi_2(x,u) + E_2\eta + D_2 f_a \\ y = x_1 \end{cases} \tag{20.4}$$

In which,$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $x_1 \in \mathbf{R}^p$, $\Phi_1(x,u) \in \mathbf{R}^p$ and $\Phi_2(x,u) \in \mathbf{R}^{n-p}$, were respectively the pre p and post $n-p$ components of $\Phi(x,u)$.

System matrix blocks as following.

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = A \quad \begin{bmatrix} E_1 \\ E_2 \end{bmatrix} = E$$

The matrix P and Q in the Lyapunov Eq. (20.2) is blocked out corresponding partition. That is

$$P = \begin{bmatrix} P_{11} & P_{12} \\ P_{12}^T & P_{22} \end{bmatrix} Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12}^T & Q_{22} \end{bmatrix}$$

From Eq. (20.3), may obtain

$$\begin{bmatrix} P_{11} & P_{12} \\ P_{12}^T & P_{22} \end{bmatrix} E = C^T F^T = \begin{bmatrix} I_p \\ 0 \end{bmatrix} F^T$$

Thus,

$$\begin{bmatrix} P_{12}^T & P_{22} \end{bmatrix} E = 0 \tag{20.5}$$

Both sides of the equation and take the left by P_{22}^{-1}, note as $K = P_{22}^{-1} P_{12}^T$, can obtain,

$$\begin{bmatrix} K & I_{n-p} \end{bmatrix} E = 0 \tag{20.6}$$

Thus, the Eq. (20.2) can be written as follows block form

$$\begin{bmatrix} * & * \\ * & (A_{22} + KA_{12})^T P_{22} + P_{22}(A_{22} + KA_{12}) \end{bmatrix} = - \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12}^T & Q_{22} \end{bmatrix}$$

Obviously,

$$(A_{22} + KA_{12})^T P_{22} + P_{22}(A_{22} + KA_{12}) = -Q_{22} \tag{20.7}$$

Because $Q > 0$, then $Q_{22} > 0$. And because (20.7) is the Lyapunov equation, $A_{22} + KA_{12}$ is stable.

Then introduced state transformation $\bar{x} = Tx$, in which, $T = \begin{bmatrix} I_p & 0 \\ K & I_{n-p} \end{bmatrix}$, $\bar{x}_1 \in \mathbf{R}^p$ and $\bar{x}_2 \in \mathbf{R}^{n-p}$. Then (20.4) can be written as follows

$$\begin{cases} \dot{\bar{x}}_1 = (A_{11} - A_{12}K)\bar{x}_1 + A_{12}\bar{x}_2 + \Phi_1(T^{-1}\bar{x}, u) + E_1\eta + D_1 f_a \\ \dot{\bar{x}}_2 = (A_{22} + KA_{12})\bar{x}_2 + [K(A_{11} - A_{12}K) + A_{21} - A_{22}K]y + K\Phi_1(T^{-1}\bar{x}, u) \\ \qquad + \Phi_2(T^{-1}\bar{x}, u) + (KD_1 + D_2)f_a \end{cases} \tag{20.8}$$

And $y = \bar{x}_1$.

In which, $\bar{x} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix}$, $\bar{x}_1 \in \mathbf{R}^p$.

At this point, the original system is decomposed into two subsystems. As can be seen from (20.8), the first subsystem also contains external disturbance and actuator faults, and the second subsystem contains the actuator failure without external interference. The equivalent system (20.8), respectively, can design the following formula (20.9) that contained two observers.

$$\begin{cases} \dot{\hat{\bar{x}}}_1 = (A_{11} - A_{12}K)\hat{\bar{x}}_1 + A_{12}\hat{\bar{x}}_2 + \Phi_1\left(T^{-1}\hat{\bar{x}}, u\right) + v \\ \dot{\hat{\bar{x}}}_2 = (A_{22} + KA_{12})\hat{\bar{x}}_2 + [K(A_{11} - A_{12}K) + A_{21} - A_{22}K]y + K\Phi_1\left(T^{-1}\hat{\bar{x}}, u\right) + \Phi_2\left(T^{-1}\hat{\bar{x}}, u\right) \\ \hat{y} = \hat{\bar{x}}_1 \end{cases} \tag{20.9}$$

In which $\hat{\bar{x}} = \begin{bmatrix} y \\ \bar{x}_2 \end{bmatrix}$, input terms v that are equivalent output are defined as

$$v = \left(\|A_{11} - A_{12}K\|\left\|e_y\right\| + \rho\|E_1\|\right)\frac{e_y}{\left\|e_y\right\|} \tag{20.10}$$

where $e_y = y - \hat{y}$, ρ is to be determined for sliding mode gain.

In order to alleviate the chattering phenomenon, the following definitions are usually used in practice.

$$v = \left(\|A_{11} - A_{12}K\|\left\|e_y\right\| + \rho\|E_1\|\right)\frac{e_y}{\left\|e_y\right\| + \varepsilon}$$

There, ε is a small positive constant.

Compared with the classical sliding mode observer given in [3], the sliding mode control (20.10) is more complicated. This is due to the more complex Lipschitz nonlinear system. Compared with the sliding mode observer in the literature [5, 6], the proposed observer structure on (20.10) in this paper is relatively simple. In order to reconstruct the fault in [5, 6], some special treatment means had to been add. In addition, it also assumes that the interference η is Lipschitz to time t in the literature [5, 6], which is a very strong restriction condition. External disturbances, however, are random; the type of interference actually is not restricted. In this paper, a sliding mode observer is used for fault monitoring on the assumption of interference η cancellation, and the application range of the system is enlarged.

20.3.2 Analysis of the Convergence and Existence of Sliding Mode

Considering the sliding mode observer in (20.9), with the different assumptions in the [5, 6], the design of sliding mode observer is also different. So the existence and convergence of the observer are reanalysed as following.

Let $e_2 = \bar{x}_2 - \hat{\bar{x}}_2$, the dynamic error equations based on (20.8) and (20.9) are as follows.

$$\begin{cases} \dot{e}_y = (A_{11} - A_{12}K)e_y + A_{12}e_2 + \Phi_1(T^{-1}\bar{x}, u) - \Phi_1\left(T^{-1}\hat{\bar{x}}, u\right) + E_1\eta + D_1 f_a - v \\ \dot{e}_2 = (A_{22} + KA_{12})e_2 + K\left[\Phi_1(T^{-1}\bar{x}, u) - \Phi_1\left(T^{-1}\hat{\bar{x}}, u\right)\right] + \Phi_2(T^{-1}\bar{x}, u) - \Phi_2\left(T^{-1}\hat{\bar{x}}, u\right)\bar{\theta}_r \\ \qquad + (KD_1 + D_2)f_a \end{cases} \tag{20.11}$$

By the structure of the matrix T, there are

$$\left\|T^{-1}\bar{x} - T^{-1}\hat{\bar{x}}\right\| = \left\|T^{-1}\left(\bar{x} - \hat{\bar{x}}\right)\right\| = \left\|T^{-1}\begin{bmatrix} 0 \\ e_2 \end{bmatrix}\right\| = \|e_2\| \tag{20.12}$$

Theorem 1 *Considering the dynamic error system* (20.11), *based on the assumption* 1–4 *and when* $f_a = 0$, *if*

$$L_\Phi \leq \frac{k_1}{k_2\left\|\begin{bmatrix} K & I_{n-p} \end{bmatrix}\right\|} \tag{20.13}$$

Then e_2 is bounded, in which k_1 and k_2 are both positive constants.

Proof When $f_a = 0$, the second equations of (20.11) can be written as

$$\dot{e}_2 = (A_{22} + KA_{12})e_2 + \begin{bmatrix} K & I_{n-p} \end{bmatrix}\left[\Phi\left(T^{-1}\bar{x}, u\right) - \Phi\left(T^{-1}\hat{\bar{x}}, u\right)\right] \tag{20.14}$$

From (20.14), there is

$$e_2(t) = e^{(A_{22}+KA_{12})t}e_2(0) + \int_0^t e^{(A_{22}+KA_{12})(t-\tau)}\begin{bmatrix} K & I_{n-p} \end{bmatrix}\left[\Phi\left(T^{-1}\bar{x}(\tau), u(\tau)\right) - \Phi\left(T^{-1}\hat{\bar{x}}(\tau), u(\tau)\right)\right]\mathrm{d}\tau$$

which, $\hat{\bar{x}}(\tau) = \begin{bmatrix} y(\tau) \\ \hat{\bar{x}}_2(\tau) \end{bmatrix}$.

Because $A_{22} + KA_{12}$ is stable, according to the [13], at the same time consider (20.12), there are $k_1 > 0$ and $k_2 > 0$, when,

$$\begin{aligned}\|e_2\| &\le k_2 e^{-k_1 t}\|e_2(0)\| + \int_0^t k_2 e^{-k_1(t-\tau)} L_\Phi \left\|\begin{bmatrix} K & I_{n-p}\end{bmatrix}\right\| \|e_2(\tau)\| \mathrm{d}\tau \\ &= k_2 e^{-k_1 t}\|e_2(0)\| + k_2 e^{-k_1 t} L_\Phi \left\|\begin{bmatrix} K & I_{n-p}\end{bmatrix}\right\| \int_0^t e^{k_1 \tau}\|e_2(\tau)\| \mathrm{d}\tau\end{aligned} \tag{20.15}$$

On both sides of the formula (20.15) multiplied by $e^{k_1 t}$ at the same time, there is,

$$e^{k_1 t}\|e_2(t)\| \le k_2\|e_2(0)\| + k_2 L_\Phi \left\|\begin{bmatrix} K & I_{n-p}\end{bmatrix}\right\| \int_0^t e^{k_1 \tau}\|e_2(\tau)\| \mathrm{d}\tau \tag{20.16}$$

By lemma 1, as can be obtained

$$\|e_2(t)\| \le k_2\|e_2(0)\| e^{\left(k_2 L_\Phi \left\|\begin{bmatrix} K & I_{n-p}\end{bmatrix}\right\| - k_1\right)t} \tag{20.17}$$

Thus theorem 1 is proved. □

Theorem 1 shows that the dynamic error Eq. (20.11) is bounded when $f_a = 0$. Then $b_2 > 0$, the equality $\|e_2\| \le b_2$ holds.

In view of (20.11), taking into account the sliding surface

$$S = \left\{(e_y, e_2)\middle| e_y = 0\right\} \tag{20.18}$$

The conditions of satisfying the motion stability and reachability of sliding mode are discussed as follows.

Theorem 2 *Based on the assumption that* 1–4 *and* $f_a = 0$, *if*

$$L_\Phi < \frac{\lambda_{\min}(Q)}{2\lambda_{\max}(P)} \tag{20.19}$$

$$\rho > b_1 + \frac{b_2(\|A_{12}\| + L_\Phi)}{\|E_1\|} \tag{20.20}$$

Hold at the same time, then the dynamic error Eq. (20.11) will reach the sliding mode surface (20.18) in finite time and keep on the sliding surface, and the sliding mode motion is asymptotically stable. There, $\lambda_{\min}(\cdot)$ represents minimum eigenvalue of a matrix, $\lambda_{\max}(\cdot)$ represents maximum eigenvalue of a matrix.

Proof Consider the Lyapunov candidate function $V = e_2^T P_{22} e_2$, by (20.14) and the definition of K can be obtained that

$$\begin{aligned}\dot{V} &= 2e_2^T P_{22}(A_{22}+KA_{12})e_2 + 2e_2^T P_{22}[K \quad I_{n-p}]\left[\Phi\left(T^{-1}\bar{x},u\right)-\Phi\left(T^{-1}\hat{\bar{x}},u\right)\right] \\ &= e_2^T\left[P_{22}(A_{22}+KA_{12})+(A_{22}+KA_{12})^T P_{22}\right]e_2 + 2e_2^T\begin{bmatrix}P_{12}^T & P_{22}\end{bmatrix}\left[\Phi\left(T^{-1}\bar{x},u\right)-\Phi\left(T^{-1}\hat{\bar{x}},u\right)\right]\end{aligned}$$

By (20.7) and (20.12) there

$$\begin{aligned}\dot{V} &= -e_2^T Q_{22}e_2 + 2\|e_2\|\left\|\begin{bmatrix}P_{12}^T & P_{22}\end{bmatrix}\right\|\left\|\Phi\left(T^{-1}\bar{x},u\right)-\Phi\left(T^{-1}\hat{\bar{x}},u\right)\right\| \\ &\le -\lambda_{\min}(Q_{22})\|e_2\|^2 + 2L_\Phi\left\|\begin{bmatrix}P_{12}^T & P_{22}\end{bmatrix}\right\|\|e_2\|^2 \\ &\le -\left[\lambda_{\min}(Q)-2L_\Phi\lambda_{\max}(P)\right]\|e_2\|^2\end{aligned}$$

When (20.19) holds, there $\dot{V}<0$, the dynamic error Eq. (20.11) is asymptotically stable in the sliding mode motion on the sliding surface (20.18).

From the first Eq. (20.11) can be obtained

$$e_y^T\dot{e}_y = e_y^T\left[(A_{11}-A_{12}K)e_y + A_{12}e_2 + \Phi_1\left(T^{-1}\bar{x},u\right)-\Phi_1\left(T^{-1}\hat{\bar{x}},u\right)+E_1\eta - v\right] \tag{20.21}$$

Let the first equations in (20.9) and (20.10) into the formula (20.21), it can be obtained as follows.

$$\begin{aligned}e_y^T\dot{e}_y &= e_y^T(A_{11}-A_{12}K)e_y + e_y^T A_{12}e_2 + e_y^T\left[\Phi_1\left(T^{-1}\bar{x},u\right)-\Phi_1\left(T^{-1}\hat{\bar{x}},u\right)\right] \\ &\quad + e_y^T\left[E_1\eta - \left(\|A_{11}-A_{12}K\|\|e_y\|+\rho E_1\right)\frac{e_y}{\|e_y\|}\right]\end{aligned}$$

It further can be gotten the following inequality by assuming 4

$$\begin{aligned}e_y^T\dot{e}_y &\le \left[\left(\|A_{12}\|+L_{\Phi_1}\right)\|e_2\| + \|E_1\|(\|\eta\|-\rho)\right]\|e_y\| \\ &\le \left[b_2(\|A_{12}\|+L_\Phi)+\|E_1\|(b_1-\rho)\right]\|e_y\|\end{aligned}$$

According to the literature [14], when the (20.20) holds, the dynamic error Eq. (20.11) satisfies the reachability condition, which is it will reach the sliding surface (20.18) in the finite time and keep on the sliding surface.

Remark 1 To make the hypothesis 3 in the literature [6] hold, the following conditions must be met.

$$\operatorname{rank}([E \quad D]) = \operatorname{rank}(C[E \quad D])$$

This condition is much stronger than the assumption 2 in this paper. It can be seen that the method proposed in this paper is also applicable to dealing with external disturbances in general sense. In this paper, in addition to the requirement that the interference η is norm bounded, it is not required that its time t is Lipschitz. Obviously, general external interference can meet such conditions.

20.3.3 Application of Sliding Mode Observer in Robust Fault Detection

Theorem 2 shows that output error signal e_y is robust from external disturbance η. In order to detect faults, let e_y the residual signal, only need to ensure that the residual e_y is sensitive to the actuator failure. From (20.8) it can be seen, as long as it holds as follows.

$$D_1 \neq 0, \quad KD_1 + D_2 \neq 0 \tag{20.22}$$

The residual e_y will be affected by actuator failures. In fact, the condition (20.22) is not difficult to meet.

In practice, the following judgment logic is used to determine whether or not the system has an actuator failure:

$$\begin{cases} \|e_y\| < Th, \text{when actuators are without failure} \\ \|e_y\| \geq Th, \text{when one of actuator failure occurred at least} \end{cases}$$

There Th is the threshold that can be set according to the specific application.

20.4 Simulation and Verification

In this section, a simplified model of a spacecraft in return with large angle of attack is carried out for the fault diagnosis simulation [15]. The model represents the following matrix form [6], in which, trim values for the Maher number: 0.8, height: 5000 feet. System matrix is as follows.

$$A = \begin{bmatrix} -0.0367 & -0.0318 & 0.0831 & -0.0008 \\ 0 & -0.0716 & -1.4850 & 0.9848 \\ 0 & -0.2797 & -5.6725 & -1.0253 \\ 0 & 0 & 0 & 1.0000 \end{bmatrix}, D = \begin{bmatrix} 0.0120 & -0.0071 \\ -0.3058 & -0.0223 \\ -22.4293 & 7.8777 \\ 0 & 0 \end{bmatrix},$$

$$E = \begin{bmatrix} 0 \\ 2.0275 \\ 10 \\ 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$\Phi(x,u) = \begin{bmatrix} 0.0120u_1 - 0.0071u_2 \\ -0.3058u_1 - 0.0223u_2 + \frac{F_e}{M}(\sin x_3)/(1+x_2) \\ -22.4293u_1 + 7.8777u_2 \\ 0 \end{bmatrix}$$

There, parameters F_e/M represent engine thrust per unit spacecraft mass. There are 4 states in the system, respectively: x_1-standard deviation of airspeed

$(v - v_0)/v_0$, in which v as the airspeed (unit: m/s), v_0 is for the desired airspeed, x_2—attack angle (unit: deg), x_3—pitch angular rate (unit: deg/s), x_4—pitch angle (unit: deg). The system has 2 inputs, respectively: u_1—deflection angle of symmetrical horizontal stabilizer (unit: deg), u_2—deflection angle of symmetric front wing (unit: deg). It has 3 outputs, respectively: y_1—standard deviation of airspeed, y_2—attack angle (unit: deg), y_3—pitch angular rate (unit: deg/s). External disturbance of the system $\|\eta\| = 0.001 * \|y\| \sin^2 x_4$, which caused into model error by air drag and lift term.

In the following simulation, a linear state feedback controller is used to configure the closed-loop poles $\{-2.5, -4.5, -4, -5\}$. If selected

$$L = \begin{bmatrix} 1.9682 & 0.0609 & 0 \\ -0.0716 & 0.0150 & 0.9848 \\ -0.2797 & -5.6725 & 0.9747 \\ -27.2784 & 0.9086 & 0.9544 \end{bmatrix}$$

For

$$Q = \begin{bmatrix} -1.0000 & -27.2968 & 0.6058 & -0.0228 \\ -27.2968 & -746.1127 & 16.5354 & -0.6229 \\ 0.6058 & 16.5354 & -1.3669 & 0.0138 \\ -0.0228 & -0.6229 & 0.0138 & -1.0005 \end{bmatrix}$$

Then the Lyapunov Eq. (20.2) has a positive definite solution

$$P = \begin{bmatrix} 372.5563 & -8.2621 & 0.3113 & 13.6392 \\ -8.2621 & 0.5168 & -0.0069 & -0.3029 \\ 0.3113 & -0.0069 & 0.2503 & 0.0114 \\ 13.6392 & -0.3029 & 0.0114 & 0.5000 \end{bmatrix}$$

There

$$F = [0 \quad 0.6758 \quad 2.5]$$

Figure 20.1 shows the state estimation when the fault occurs, and Fig. 20.2 shows the residual information when the actuator fails. There, fault signal is

$$f_{a1} = \begin{cases} 0.25, & 60 \le t \le 80 \\ 0, & \text{others} \end{cases}$$
$$f_{a2} = \begin{cases} 0.25, & 60 \le t \le 80 \\ 0, & \text{others} \end{cases}$$

The Figs. 20.1 and 20.2 shows that when there is no fault in the system, the observer can track the state of the system asymptotically. When the system fails, the residual error can be more sensitive to give alarm information. At the same time,

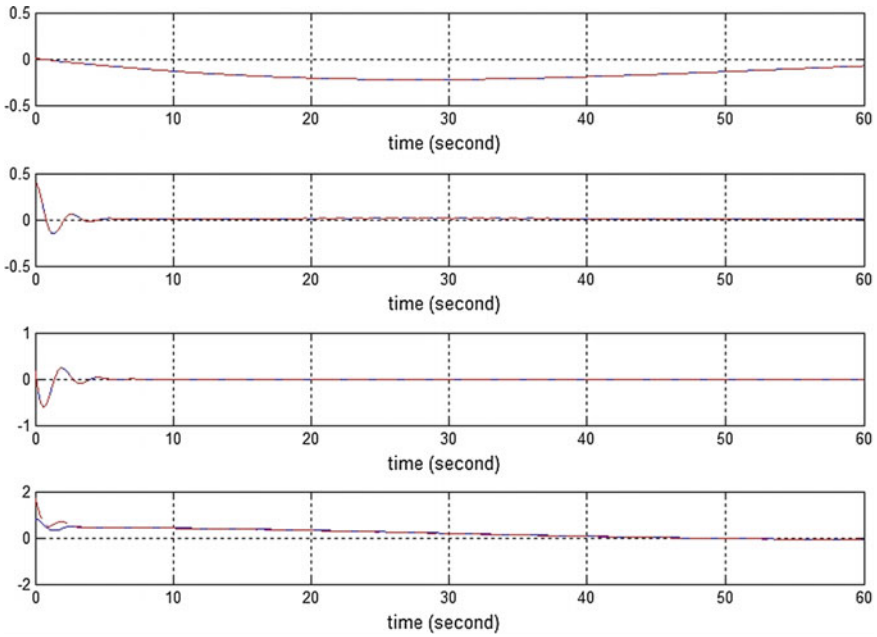


Fig. 20.1 System state and observation value without fault

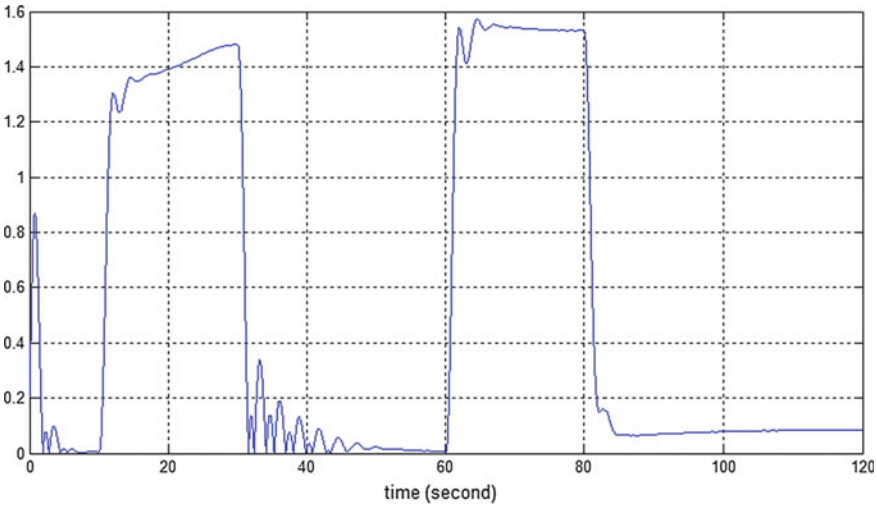


Fig. 20.2 The value of residual generation failure

the residual can also roughly determine when the actuator failure occurs, when the fault will disappear. It is seen that the effect of the fault detection and isolation method is satisfactory in this paper.

20.5 Conclusions

In this paper, for a class of Lipschitz nonlinear systems with external disturbances, a fault detection method based on sliding mode observer is proposed. This sliding mode observer is robust to external disturbance. When the actuator is free of faults, the state of the system can be asymptotically tracked. When actuator failure occurs, the robust residual error can send alarm signal. The proposed method is applied to a spacecraft reentry model with high angle of attack, and the simulation results show the effectiveness of the proposed method.

References

1. Tan CP, Edwards C (2002) Sliding mode observers for detection and reconstruction of sensors faults. Automatica 38(5):1815–1821
2. Zhang X (2011) Sensor bias fault detection and isolation in a class of nonlinear uncertain systems using adaptive estimation. IEEE Trans Autom Control 56(5):1220–1226
3. Edwards C, Spurgeon SK, Patton RJ (2000) Sliding mode observers for fault detection and isolation. Automatica 36(4):541–553
4. Alwi H, Edwards C (2008) Fault detection and fault-tolerant control of civil aircraft using a sliding-mode-based scheme. IEEE Transactions on Control System Technology 16(3):451–499
5. Yan XG, Edwards C (2007) Non-linear robust fault reconstruction and estimation using a sliding mode observer. Automatica 43:1605–1614
6. Yan XG, Edwards C (2008) Robust sliding mode observer based actuator fault detection and isolation for a class of nonlinear systems. Int J Syst Sci 39(4):349–359
7. Lee D, Park YJ, Park Y (2012) Robust H∞ sliding mode descriptor observer for fault and output disturbance estimation of uncertain systems. IEEE Trans Autom Control 57(11):2928–2934
8. Veluvolu KC, Soh YC (2011) Fault reconstruction and state estimation with sliding mode observers for Lipschitz non-linear systems. IET Control Theory Appl 5(11):1255–1263
9. Mao ZH, Jiang B, Shi P (2010) Fault-tolerant control for a class of nonlinear sampled-data systems via an Euler approximate observer. Automatica 46(11):1852–1859
10. Edwards C (2004) A comparison of sliding mode and un-known input observers for fault reconstruction. In: 43rd Conference on Decision and Control. IEEE
11. Khalil HK (2002) Non-linear systems. Prentice-Hall, New Jersey
12. Corless M, Tu J (1998) State and input estimation for a class of uncertain systems. Automatica 34(6):757–764
13. Yang J, Li S, Yu X (2013) Sliding modes control for systems with mismatched uncertaintied via a disturbance observer. IEEE Trans Industrial Electronica 60(1):160–169
14. Rahme S, Meskin N (2015) Adaptive sliding mode observer for sensor fault diagnosis of an industrial gas turbine. Cont Eng Pr 38(3):57–74
15 Fadjar RT, Bambang RT, Trio A, etc. (2015) Nonlinear dynamic modeling of a fixed-wing unmanned aerial vehicle: a case study of wulung. J Mechatron Electri Power and Veh Tech 6(1):19–30

Chapter 21
Human Motion Capture Similarity Control for Space Teleoperation

Zhong Shi, Xuexiang Huang and Tianjian Hu

21.1 Introduction

With the rapid development of space technology and the continuous deepening of space exploration activities, the tasks of space operations including in-space maintenance, planetary surface exploration and space scientific experiment are gradually rising [1]. As one of the most important ways of executing tasks of space operations, space teleoperation combines the intelligence of human beings and the adaptability of space robots, to reduce the risk and cost of space operations and improve the abilities of adapting to the space environment, which has been a new hotspot in the research of space technology [2].

At present, the operator is mainly used to control the end pose of space robot, and many domestic and foreign scholars have studied the Cartesian space feedback control method: The scattering operator is proposed based on the passive theory in [3]; On this basis, the wave variable method is proposed in [4]; Based on the H∞ optimal control theory, the two step adaptive controller for space teleoperation is designed in [5]; A modified force position hybrid control method is proposed in [6] for the ORU assembly processes of space teleoperation; The stability, transparency and traceability of the bilateral PD control method are studied in [7] based on the absolute stability theory; A fuzzy adaptive terminal sliding mode controller for the teleoperation system is designed in [8]. The Cartesian space feedback control for space teleoperation has advantages in overcoming the adverse effect of space environment disturbances, and ensuring the stability and precision of the space teleoperation. The terminal sliding mode controller is widely used in the complex

Z. Shi (✉) · X. Huang · T. Hu
Beijing Institute of Tracking and Telecommunications Technology,
Beijing 100094, China
e-mail: clock1991529@163.com

T. Hu
School of Aerospace Engineering, Tsinghua University, Beijing 100084, China

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_21

robot control system due to its control characteristics of rapidity, robustness and anti-interference [9, 10], so it is suitable for the space teleoperation system with high real-time and accurate requirements. However, only using the Cartesian space control method, the operator can not directly control the joint space of the robot, and the flexibility of operation is limited, which will reduce the operator's execution ability for the complex teleoperation tasks in unknown environment.

Motion capture technology has been used in introducing the human's motion information into the teleoperation control loop, which is an important way to improve operational flexibility. [11, 12] use the position tracker based on the magnetic field and the motion capture system based on the vision, respectively, to capture the movement of the operator's head, hands and chest for the use of controlling the movements of the robot astronaut; The humanoid robots are flexibly controlled based on the optical motion capture equipment and the inertial motion capture device in [13, 14], respectively. However, most of space robots are non-humanoid robots, whose linkage length and joint configuration are not strictly the same as the body structure, so it is difficult to use the motion capture technology in space teleoperation system.

Aiming at the deficiency of the traditional space teleoperation control method, an ATSMS control method is proposed in this paper based on motion capture technology. Firstly, an ATSM controller is designed to ensure the accurate control of the human arm to the end position of the robot; Secondly, a similarity control method is designed to control the robot's joint configuration to tend to the human arm joint configuration, and achieve the direct control of the joint space; Meanwhile, the space teleoperation time-delay system is transformed into a non-time-delay system by the output predictor; Then, the accuracy of the end position control and the similarity between the human arm and the robot joint space are proved theoretically; Finally, the effectiveness of the ATSMS control method is analyzed and verified by numerical simulation and ground experiments.

21.2 Adaptive Terminal Sliding Mode Similarity Control Method

The nonlinear dynamic model for a free-flying space robot with n-degrees of freedom can be described as

$$\boldsymbol{\tau}_{\mathrm{r}} + \boldsymbol{\tau}_{\mathrm{d}} = \boldsymbol{M}(\boldsymbol{q}_{\mathrm{r}})\ddot{\boldsymbol{q}}_{\mathrm{r}} + \boldsymbol{C}(\boldsymbol{q}_{\mathrm{r}}, \dot{\boldsymbol{q}}_{\mathrm{r}}) \tag{21.1}$$

where, $\boldsymbol{q}_{\mathrm{r}}$, $\dot{\boldsymbol{q}}_{\mathrm{r}}$, $\ddot{\boldsymbol{q}}_{\mathrm{r}} \in \mathbf{R}^n$ are the joint angle, velocity and acceleration, respectively; $\boldsymbol{M}(\boldsymbol{q}_{\mathrm{r}}) \in \mathbf{R}^{n\times n}$ is the inertia matrix; $\boldsymbol{C}(\boldsymbol{q}_{\mathrm{r}}, \dot{\boldsymbol{q}}_{\mathrm{r}}) \in \mathbf{R}^n$ is the coefficient matrix of centrifugal, Coriolis and velocity damping; $\boldsymbol{\tau}_{\mathrm{r}} \in \mathbf{R}^n$ is the control torque; and $\boldsymbol{\tau}_{\mathrm{d}} \in \mathbf{R}^n$ is the disturb torque.

For space teleoperation, the accurate control of the end position is the premise of the fine operation; In addition, in order to improve the flexibility of space

teleoperation, it is necessary to control the robot joint space directly. Thus, it is needed to consider the coupling relationship between the joint space and the Cartesian space of space robot, which can be expressed as

$$\dot{\boldsymbol{p}}_{\mathrm{r}} = \boldsymbol{J}(\boldsymbol{q}_{\mathrm{r}})\dot{\boldsymbol{q}}_{\mathrm{r}} \tag{21.2}$$

in which, $\dot{\boldsymbol{p}}_{\mathrm{r}} \in \mathbf{R}^m$ represents the end velocity of the robot; $\boldsymbol{J}(\boldsymbol{q}_{\mathrm{r}}) \in \mathbf{R}^{m\times n}$ represents the Jacobian matrix, in general $m < n$ and $\boldsymbol{J}$ is underdetermined.

By using the underdetermined characteristic of Jacobian matrix, it can be effectively dealt with that the problem of controlling the joint space and Cartesian space of the robot simultaneously [15, 16]. By differentiating formula (21.2) with respect to the time, the differential relationship for accelerations can be obtained as

$$\ddot{\boldsymbol{p}}_{\mathrm{r}} = \boldsymbol{J}\ddot{\boldsymbol{q}}_{\mathrm{r}} + \dot{\boldsymbol{J}}\dot{q}_{\mathrm{r}} \tag{21.3}$$

By inversing formula (21.3), the joint acceleration of the robot can be expressed as

$$\ddot{\boldsymbol{q}}_{\mathrm{r}} = \boldsymbol{J}^{+}\left(\ddot{\boldsymbol{p}}_{\mathrm{r}} - \dot{\boldsymbol{J}}\dot{q}_{\mathrm{r}}\right) + \ddot{\boldsymbol{q}}_{\mathrm{N}} \tag{21.4}$$

in which, $\boldsymbol{J}^{+} \in R^{n\times m}$ is the Moore-Penrose inverse of Jacobian matrix, and $\boldsymbol{J}^{+} = \boldsymbol{J}^{\mathrm{T}}(\boldsymbol{J}\boldsymbol{J}^{\mathrm{T}})^{-1}$ generally; $\ddot{\boldsymbol{q}}_{\mathrm{N}} \in \mathbf{R}^n$ represents the component of the $\ddot{\boldsymbol{q}}_{\mathrm{r}}$ in the null space of Jacobian matrix $\boldsymbol{J}$, i.e., $\boldsymbol{J}\ddot{\boldsymbol{q}}_{\mathrm{N}} = 0$.

Substituting formula (21.4) into formula (21.1), the following result can be obtained

$$\boldsymbol{\tau}_{\mathrm{r}} + \boldsymbol{\tau}_{\mathrm{d}} = \boldsymbol{M}(\boldsymbol{q}_{\mathrm{r}})\left\{\boldsymbol{J}^{+}\left(\ddot{\boldsymbol{p}}_{\mathrm{r}} - \dot{\boldsymbol{J}}\dot{q}_{\mathrm{r}}\right) + \ddot{\boldsymbol{q}}_{\mathrm{N}}\right\} + \boldsymbol{C}(\boldsymbol{q}_{\mathrm{r}}, \dot{\boldsymbol{q}}_{\mathrm{r}}) \tag{21.5}$$

By using the computed torque method, the control torque can be calculated as

$$\boldsymbol{\tau}_{\mathrm{r}} = \hat{\boldsymbol{M}}(\boldsymbol{q}_{\mathrm{r}})\left\{\boldsymbol{J}^{+}\left(\ddot{\boldsymbol{p}}_{\mathrm{h}} + \boldsymbol{u} - \dot{\boldsymbol{J}}\dot{q}_{\mathrm{r}}\right) + \boldsymbol{\phi}_{\mathrm{N}}\right\} + \hat{\boldsymbol{C}}(\boldsymbol{q}_{\mathrm{r}}, \dot{\boldsymbol{q}}_{\mathrm{r}}) \tag{21.6}$$

in which, $\hat{\boldsymbol{M}}(\boldsymbol{q}_{\mathrm{r}}) \in \mathbf{R}^{n\times n}$ and $\hat{\boldsymbol{C}}(\boldsymbol{q}_{\mathrm{r}}, \dot{\boldsymbol{q}}_{\mathrm{r}}) \in \mathbf{R}^n$ represent the online identification values of $\boldsymbol{M}(\boldsymbol{q}_{\mathrm{r}})$ and $\boldsymbol{C}(\boldsymbol{q}_{\mathrm{r}}, \dot{\boldsymbol{q}}_{\mathrm{r}})$, respectively; $\ddot{\boldsymbol{p}}_{\mathrm{h}} \in \mathbf{R}^m$ is the end acceleration of human arm, which can be measured by the motion capture device as the input of space teleoperation; $\boldsymbol{u} \in \mathbf{R}^m$ represents the feedback control law; and $\boldsymbol{\phi}_{\mathrm{N}} \in \mathbf{R}^n$ represents the arbitrary vector in the null space of Jacobian matrix $\boldsymbol{J}$, i.e., $\boldsymbol{J}\boldsymbol{\phi}_{\mathrm{N}} = 0$.

Thus, an ATSMS control system is designed in this paper, and the control system mainly includes three components of a output predictor, an ATSM controller and a similarity controller, as shown in Fig. 21.1.

21.2.1 Output Predictor

Time delay is an important factor affecting the stability of space teleoperation system, and the output predictor based on tracking differentiator [17] can transform the space teleoperation with time delay into an non-time-delay system. The discrete form expression for output predictor is

$$\begin{cases} \boldsymbol{f}_y = \text{fhan}(\boldsymbol{y}_2(k) - \boldsymbol{y}_1(k), \boldsymbol{y}_3(k), r, h_0) \\ \boldsymbol{y}_2(k+1) = \boldsymbol{y}_2(k) + h\boldsymbol{y}_3(k) \\ \boldsymbol{y}_3(k+1) = \boldsymbol{y}_3(k) + h\boldsymbol{f}_y \\ \boldsymbol{y}_0(k) = \boldsymbol{y}_2(k) + \alpha_y \boldsymbol{y}_3(k) \end{cases} \tag{21.7}$$

in which, $\boldsymbol{y}_1 = [y_{1i}]_{n\times 1} \in \mathbf{R}^n$ represents the real joint angle of space robot of the time-delay system; $\boldsymbol{y}_2 = [y_{2i}]_{n\times 1} \in \mathbf{R}^n$ and $\boldsymbol{y}_3 = [y_{3i}]_{n\times 1} \in \mathbf{R}^n$ represent the extended outputs of the output predictor; $\boldsymbol{y}_0 \in \mathbf{R}^n$ represents the non-time-delay joint angle produced by the output predictor; α_y represents the differential weight coefficient of $\boldsymbol{y}_0$; $\text{fhan}(\boldsymbol{y}_2 - \boldsymbol{y}_1, \boldsymbol{y}_3, r, h_0) = [\text{fhan}(y_{2i}\text{-}y_{1i}, y_{3i}, r, h_0)]_{n\times 1} \in \mathbf{R}^n$ is the optimal control synthesis function of discrete-time system, which can be expressed as

$$\begin{cases} d = rh_0^2, \quad a_0 = h_0 y_{3i}, \quad y = y_{2i} - y_{1i} + a_0 \\ a_1 = \sqrt{d(d + 8|y|)}, \quad a_2 = a_0 + sign(y)(a_1 - d)/2 \\ fsg(y, d) = (sign(y + d) - sign(y - d))/2 \\ a = (a_0 + y)fsg(y, d) + a_2(1 - fsg(y, d)) \\ fhan = -r\left(\frac{a}{d}\right)fsg(a, d) - r \cdot sign(a)(1 - fsg(a, d)) \end{cases} \tag{21.8}$$

where h_0 and r are the adjusting parameters of function $\text{fhan}(\boldsymbol{y}_2 - \boldsymbol{y}_1, \boldsymbol{y}_3, r, h_0)$.

21.2.2 Adaptive Terminal Sliding Mode Controller

Considering the uncertain disturbances including the environmental interference, measurement noise and modeling error and the real-time requirements of space teleoperation, the feedback control law needs to be robust, fast and anti-interference, thus, an ATSM controller is designed to generate the feedback control law and the sliding mode surface is designed as

$$\boldsymbol{S} = \boldsymbol{e}_p + \frac{1}{k}\text{sig}^{1/\alpha}(\dot{\boldsymbol{e}}_p) \tag{21.9}$$

in which, represents the control error of end position; represents the control error of end velocity; and $\mathrm{sig}^{1/\alpha}(\dot{\boldsymbol{e}}_p) = \left[\left|\dot{e}_{p1}\right|^{1/\alpha}\mathrm{sign}(\dot{e}_{p1}),\ \left|\dot{e}_{p2}\right|^{1/\alpha}\mathrm{sign}(\dot{e}_{p2}),\ \ldots,\ \left|\dot{e}_{pm}\right|^{1/\alpha}\mathrm{sign}(\dot{e}_{pm})\right]^{\mathrm{T}}$, $0 < \alpha < 1$, $k > 0$.

Chattering is an inherent phenomenon of sliding mode control, and to eliminate the chattering, the sliding mode reaching law is designed as:

$$\dot{\boldsymbol{S}} = -\beta_1 \boldsymbol{S} - \beta_2 \mathrm{sig}^{\eta}(\boldsymbol{S}) \tag{21.10}$$

in which, β_1, $\beta_2 > 0$; and the range of η is limited as (0.5, 1), in order to avoid the chattering of sliding mode as far as possible while ensuring the convergence of the system.

In order to overcome the uncertain disturbances of the space teleoperation system, the adaptive terminal sliding mode control law is designed as

$$\boldsymbol{u} = k\alpha \mathrm{sig}^{2-1/\alpha}(\dot{\boldsymbol{e}}_p) + \beta_1 \boldsymbol{S} + \beta_2 \mathrm{sig}^{\eta}(\boldsymbol{S}) + \hat{\boldsymbol{\psi}} \tag{21.11}$$

in which, $\hat{\boldsymbol{\psi}}$ represents the adaptive estimation of the uncertain disturbances, and the adaptive law can be calculated as

$$\dot{\hat{\boldsymbol{\psi}}} = \lambda \boldsymbol{S} \tag{21.12}$$

in which, $\lambda > 0$ represents the adaptation gain.

21.2.3 Similarity Controller

In order to realize the direct control of the joint space, it is necessary to design a similarity control method to make the robot joint space tend to the operator's arm, i.e., the joint velocity of robot meets the following relationship:

$$\begin{cases} \varepsilon_{qi} \geq \dot{q}_{ri} > 0, & e_{qi} > \varepsilon_e \\ \dot{q}_{ri} = 0, & \left|e_{qi}\right| \leq \varepsilon_e \\ -\varepsilon_{qi} \leq \dot{q}_{ri} < 0, & e_{qi} < -\varepsilon_e \end{cases} \tag{21.13}$$

in which, $\dot{q}_{ri}$ and e_{qi} represent the joint velocity and similarity error of the i-th ($i = 1, 2, \ldots, n$) joint, respectively (Fig. 21.1); $\varepsilon_{qi} > 0$ represents the upper bound of the i-th joint velocity; $\varepsilon_e > 0$ defines the size of the joint similarity region, and the robot joint space is thought to be similar to the human arm when $|e_{qi}| \leq \varepsilon_e$.

Considering the continuity and smoothness of the joint velocity of the space robot, the clamping velocity is designed as the following:

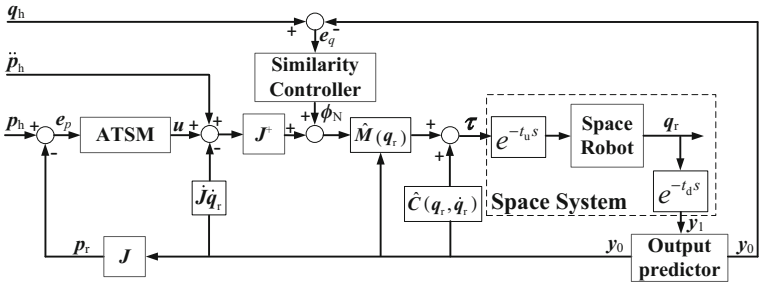


Fig. 21.1 ATSMS control system for space teleoperation

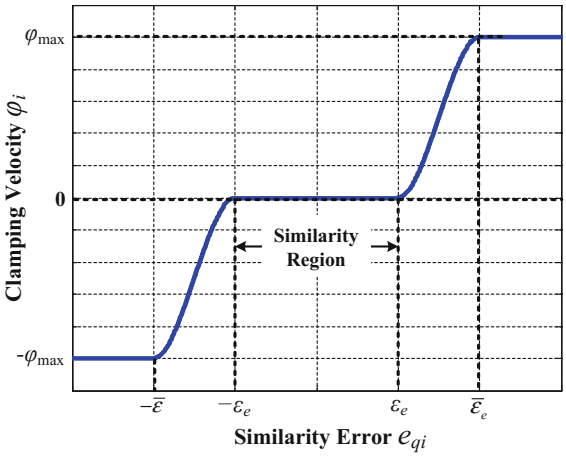


Fig. 21.2 Relationship curve between the clamping velocity and similarity error for i-th joint

$$\varphi_i = \begin{cases} \dfrac{\left(e_{qi}-\varepsilon_e\right)^2\left(2e_{qi}+\varepsilon_e-3\bar{\varepsilon}_e\right)}{\left(\varepsilon_e-\bar{\varepsilon}_e\right)^3}\varphi_{\max} & \bar{\varepsilon}_e > e_{qi} > \varepsilon_e \\ \dfrac{\left(e_{qi}+\varepsilon_e\right)^2\left(2e_{qi}-\varepsilon_e+3\bar{\varepsilon}_e\right)}{\left(\varepsilon_e-\bar{\varepsilon}_e\right)^3}\varphi_{\max} & -\bar{\varepsilon}_e < e_{qi} < -\varepsilon_e \\ 0 & \left|e_{qi}\right| \le \varepsilon_e \\ -\varphi_{\max} & e_{qi} \le -\bar{\varepsilon}_e \\ \varphi_{\max} & e_{qi} \ge \bar{\varepsilon}_e \end{cases} \tag{21.14}$$

in which, $\bar{\varepsilon}_e$ represents the threshold of similarity error, and $\varphi_{\max}$ represents the maximum clamping velocity. The relationship curve between the clamping velocity and similarity error is depicted as Fig. 21.2. When $|e_{qi}| < \varepsilon_e$, the calmping velocity is set as 0; and when the similarity error of the i-th joint is out of the joint similarity region, the i-th joint will be pushed toward the joint similarity region in the null space of Jacobian matrix $\boldsymbol{J}$ by the clamping velocity φ_i. Meanwhile, φ_i is continuously differentiable everywhere so that there is no jump to joint velocities and joint torques become continuous.

In order to avoid interrupting the accurate control of end position, the clamping velocity $\boldsymbol{\Phi}$ is projected on the null space of Jacobian matrix by designing $\boldsymbol{\phi}_N$ as

$$\boldsymbol{\phi}_{\mathrm{N}} = (\boldsymbol{I} - \boldsymbol{J}^{+}\boldsymbol{J})\left(\dot{\boldsymbol{\Phi}} + \boldsymbol{K}\underline{\boldsymbol{e}}_{\mathrm{N}}\right) - \left(\boldsymbol{J}^{+}\dot{\boldsymbol{J}}\boldsymbol{J}^{+} + \dot{\boldsymbol{J}}^{+}\right)\boldsymbol{J}(\boldsymbol{\Phi} - \dot{\boldsymbol{q}}_{\mathrm{r}}) \tag{21.15}$$

in which $\boldsymbol{K} \in \mathbf{R}^{n}$ is a positive-definite matrix; $\dot{\boldsymbol{e}}_{\mathrm{N}} \in \mathbf{R}^{n}$ are the errors between the clamping velocity and joint velocity in the null space of Jacobian matrix, which is defined as

$$\dot{\boldsymbol{e}}_{\mathrm{N}} = (\boldsymbol{I} - \boldsymbol{J}^{+}\boldsymbol{J})(\boldsymbol{\Phi} - \dot{\boldsymbol{q}}_{\mathrm{r}}) \tag{21.16}$$

21.3 Performance Analysis of Control System

21.3.1 Stability Analysis

The convergence speed of adaptive estimation is increased as the adaptation gain λ increases, but the excessive adaptation gain λ will destroy the stability of the control system designed in this paper. The following theorem gives a further explanation for the determination of λ.

Theorem 1 *The control system depicted in Fig. 21.1 is asymptotically stable and the end position control error tends to zero, if the adaptation gain λ in the adaptive law (21.12) takes the following value*

$$\lambda = \min_{i=1,2,\cdots,m}\left\{\frac{1}{\alpha k}\left|\dot{e}_{pi}\right|^{1/\alpha-1}\right\} \tag{21.17}$$

in which $\dot{e}_{pi}$ is the i-th component of the end velocity control error $\dot{\boldsymbol{e}}_p$.

Proof Due to the effect of the output predictor, only consider the non-time-delay systems here. Substitute formula (21.5) into formula (21.6), and the following equality can be obtained

$$\boldsymbol{J}^{+}\left(\ddot{\boldsymbol{e}}_p + \boldsymbol{u}\right) + \boldsymbol{\phi}_{\mathrm{N}} - \ddot{\boldsymbol{q}}_{\mathrm{N}} = \hat{\boldsymbol{M}}^{-1}\left[(\boldsymbol{M} - \hat{\boldsymbol{M}})\ddot{\boldsymbol{q}}_{\mathrm{r}} + \left(\boldsymbol{C} - \hat{\boldsymbol{C}}\right) - \tau_{\mathrm{d}}\right] \tag{21.18}$$

Due to $\boldsymbol{J}\ddot{\boldsymbol{q}}_N = \boldsymbol{0}_{m\times 1}$, and

$$\begin{aligned}\boldsymbol{J}\boldsymbol{\phi}_{\mathrm{N}} &= \boldsymbol{J}(\boldsymbol{I} - \boldsymbol{J}^{+}\boldsymbol{J})\left(\dot{\boldsymbol{\Phi}} + \boldsymbol{K}\underline{\boldsymbol{e}}_{\mathrm{N}}\right) - \boldsymbol{J}\left(\boldsymbol{J}^{+}\dot{\boldsymbol{J}}\boldsymbol{J}^{+} + \dot{\boldsymbol{J}}^{+}\right)\boldsymbol{J}(\boldsymbol{\Phi} - \dot{\boldsymbol{q}}_{\mathrm{r}}) \\ &= -\left(\dot{\boldsymbol{J}}\boldsymbol{J}^{+} + \boldsymbol{J}\dot{\boldsymbol{J}}^{+}\right)\boldsymbol{J}(\boldsymbol{\Phi} - \dot{\boldsymbol{q}}_{\mathrm{r}}) = -\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{J}\boldsymbol{J}^{+})\boldsymbol{J}(\boldsymbol{\Phi} - \dot{\boldsymbol{q}}_{\mathrm{r}}) \\ &= -\left(\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{I}\right)\boldsymbol{J}(\boldsymbol{\Phi} - \dot{\boldsymbol{q}}_{\mathrm{r}}) = \boldsymbol{0}_{m\times 1}\end{aligned} \tag{21.19}$$

by multiplying $\boldsymbol{J}$ on the left of formula (21.18), the system error equation can be obtained as

$$\ddot{\boldsymbol{e}}_p + \boldsymbol{u} = \boldsymbol{J}\hat{\boldsymbol{M}}^{-1}\left[(\boldsymbol{M} - \hat{\boldsymbol{M}})\ddot{\boldsymbol{q}}_{\mathrm{r}} + (\boldsymbol{C} - \hat{\boldsymbol{C}}) + \boldsymbol{D}\right] \triangleq \boldsymbol{\psi} \tag{21.20}$$

in which $\boldsymbol{\psi}$ represents the uncertain disturbances of the control system, including the unmolded dynamics and external disturbances.

The following Lyapunov function is defined to verified the stability

$$V = \frac{1}{2}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{S} + \frac{1}{2}\tilde{\boldsymbol{\psi}}^{\mathrm{T}}\tilde{\boldsymbol{\psi}} \tag{21.21}$$

in which $\tilde{\boldsymbol{\psi}}$ are the estimation errors of the uncertain disturbances, i.e.,

$$\tilde{\boldsymbol{\psi}} = \boldsymbol{\psi} - \hat{\psi} \tag{21.22}$$

Differentiating $\boldsymbol{V}$ with respect to the time along the system trajectory, the following equality can be obtained

$$\dot{V} = \boldsymbol{S}^{\mathrm{T}}\dot{\boldsymbol{S}} + \tilde{\boldsymbol{\psi}}^{\mathrm{T}}\dot{\tilde{\boldsymbol{\psi}}} \tag{21.23}$$

Substituting formulas (21.9), (21.11), (21.20) and (21.22) into formula (21.23) and after simplification, the following equality can be obtained

$$\begin{aligned}\dot{V} = {} & \boldsymbol{S}^{\mathrm{T}}\frac{1}{\alpha k}\mathrm{diag}\left(\left|\underline{\boldsymbol{e}}_p\right|^{1/\alpha-1}\right)(-\beta_1\boldsymbol{S} - \beta_2\mathrm{sig}^{\eta}(\boldsymbol{S})) \\ & + \boldsymbol{S}^{\mathrm{T}}\left(\lambda\boldsymbol{I}_m - \frac{1}{\alpha k}\mathrm{diag}\left(\left|\underline{\boldsymbol{e}}_p\right|^{1/\alpha-1}\right)\right)\tilde{\psi} - \tilde{\psi}^{\mathrm{T}}\psi\end{aligned} \tag{21.24}$$

Since $\lambda = \min\limits_{i=1,2,\cdots,m}\left\{\frac{1}{\alpha k}\left|\dot{e}_{pi}\right|^{1/\alpha-1}\right\}$, the following inequality can be obtained

$$\dot{V} \leq -\lambda\beta_1\|\boldsymbol{S}\| - \lambda\beta_2\|\boldsymbol{S}\|^{\eta+1} + \left|\tilde{\boldsymbol{\psi}}^{\mathrm{T}}\dot{\boldsymbol{\psi}}\right| \tag{21.25}$$

By adjusting the parameters β_1 and β_2, we can obtain $\dot{V} < 0$. Thus, the control system is asymptotically stable, and the sliding mode surface $\boldsymbol{S}$ and the estimation errors of the uncertain disturbances $\tilde{\boldsymbol{\psi}}$ converge to zero. Then, we can obtain

$$\dot{\boldsymbol{e}}_p = -k\mathrm{sig}^{\alpha}(\boldsymbol{e}_p) \tag{21.26}$$

Hence, the end position control error $\boldsymbol{e}_p$ will converge to zero, and the convergence time of the feedback control system is

$$t_c = \frac{1}{k^{\alpha}(1-\alpha)} \left| \boldsymbol{e}_p(t_s) \right|^{1-\alpha} \tag{21.27}$$

in which t_s is the time of system states reaching the sliding mode surface. □

Theorem 1 indicates that: the ATSMS method can effectively estimate the modeling error and disturbance torque of the system, so that the system has good robustness and anti-disturbance; the ATSMS method can guarantee the asymptotic stability of the system, and the end position control error tends to 0; the clamping velocity is projected on the null space of the Jacobian matrix, which has no disturbances to the stability of the system.

21.3.2 Joint Similarity Verification

The clamping velocity can make the space robot joint position tend to the human arm on the null space of the Jacobi matrix to realize the direct control of the robot joint space. In order to verify the effectiveness of similarity control method, firstly, the Jacobian matrix is decomposed by the singular value decomposition, which can be expressed as

$$\boldsymbol{J} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{\mathrm{T}} \tag{21.28}$$

in which, $\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m] \in \mathbf{R}^{m\times m}$, $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n] \in \mathbf{R}^{n\times n}$, and $\boldsymbol{D} = \mathrm{diag}(\sigma_1,\ldots, \sigma_s,0,\ldots,0) \in \mathbf{R}^{m\times n}$ with $s \leq m$ as the rank of $\boldsymbol{J}$ and $\sigma_1 \geq \sigma_2 \geq \cdots \geq {}_s > 0$ as the singular values of $\boldsymbol{J}$. The joint velocity of the robot can be calculated as

$$\dot{\boldsymbol{q}}_{\mathrm{r}} = \sum_{j=1}^{s} \frac{1}{\sigma_j} \boldsymbol{u}_j^{\mathrm{T}} \dot{\boldsymbol{p}}_{\mathrm{r}} \boldsymbol{v}_j + \dot{\boldsymbol{q}}_{\mathrm{r}}^{\mathrm{N}} \tag{21.29}$$

in which $\dot{\boldsymbol{q}}_{\mathrm{r}}^{\mathrm{N}}$ represents the joint velocity in the null space of the Jacobian matrix. The ATSM controller can ensure the accurate control of the end position of the space robot, so $\dot{\boldsymbol{p}}_{\mathrm{h}}$ can be used as a substitute for $\dot{\boldsymbol{p}}_{\mathrm{r}}$ to solve the robot joint angular velocity.

To simplify the verification, assume that only the last joint locates outside the joint similarity region, i.e. $e_{qn} > \varepsilon_e$; while other joints locate inside the joint similarity region, i.e. $|e_{qi}| \leq \varepsilon_e$ for $i = 1, 2, \ldots, n-1$.

Theorem 2 *When* $e_{qn} < -\varepsilon_e$ *and* $|e_{qi}| \leq \varepsilon_e$ $(i = 1, 2, \ldots, n-1)$, *the velocity of the n-th joint will satisfy*

$$\dot{q}_{\mathrm{r}n} > 0 \tag{21.30}$$

which means a joint motion towards the similarity region of n-th joint, if the maximum value of clamping velocity satisfies

$$\varphi_{\max} > \frac{\left|\sum_{j=1}^{s} \frac{1}{\sigma_j} \boldsymbol{u}_j^{\mathrm{T}} \dot{\boldsymbol{p}}_{\mathrm{h}} \boldsymbol{v}_j(n)\right| (\varepsilon_e - \bar{\varepsilon}_e)^3}{\left(1 - \sum_{j=1}^{s} \left[\boldsymbol{v}_j(n)\right]^2\right) (e_{qn} - \varepsilon_e)^2 (2e_{qn} + \varepsilon_e - 3\bar{\varepsilon}_e)} \tag{21.31}$$

Proof Differentiating formula (21.16) with respect to the time, we can obtain

$$\begin{aligned} \ddot{\boldsymbol{e}}_{\mathrm{N}} &= (\boldsymbol{I} - \boldsymbol{J}^+ \boldsymbol{J})(\dot{\boldsymbol{\Phi}} - \ddot{\boldsymbol{q}}_{\mathrm{r}}) - (\dot{\boldsymbol{J}}^+ \boldsymbol{J} + \boldsymbol{J}^+ \dot{\boldsymbol{J}})(\boldsymbol{\Phi} - \dot{\boldsymbol{q}}_{\mathrm{r}}) \\ &= (\boldsymbol{I} - \boldsymbol{J}^+ \boldsymbol{J})(\dot{\boldsymbol{\Phi}} - \ddot{\boldsymbol{q}}_{\mathrm{r}}) - \boldsymbol{J}^+ \dot{\boldsymbol{J}} \dot{\boldsymbol{e}}_{\mathrm{N}} - (\boldsymbol{J}^+ \dot{\boldsymbol{J}} \boldsymbol{J}^+ + \dot{\boldsymbol{J}}^+) \boldsymbol{J}(\boldsymbol{\Phi} - \dot{\boldsymbol{q}}_{\mathrm{r}}) \end{aligned} \tag{21.32}$$

Derived from formula (21.6), the joint acceleration of the robot can be obtained as

$$\ddot{\boldsymbol{q}}_{\mathrm{r}} = \boldsymbol{J}^+ (\ddot{\boldsymbol{p}}_{\mathrm{h}} + \boldsymbol{u} - \dot{\boldsymbol{J}} \dot{q}_{\mathrm{r}}) + \boldsymbol{\phi}_{\mathrm{N}} \tag{21.33}$$

Substitute formulas (21.14) and (21.33) into formula (21.32), yields

$$\ddot{\boldsymbol{e}}_{\mathrm{N}} = -(\boldsymbol{I} - \boldsymbol{J}^+ \boldsymbol{J}) \boldsymbol{K} \underline{\boldsymbol{e}}_{\mathrm{N}} - \boldsymbol{J}^+ \dot{\boldsymbol{J}} \dot{\boldsymbol{e}}_{\mathrm{N}} \tag{21.34}$$

Define the Lyapunov function as

$$V = \frac{1}{2} \dot{\boldsymbol{e}}_{\mathrm{N}}^{\mathrm{T}} \dot{\boldsymbol{e}}_{\mathrm{N}} \tag{21.35}$$

Differentiating Lyapunov function V with respect to the time and substituting formula (21.34) into it, yields

$$\dot{V} = \dot{\boldsymbol{e}}_{\mathrm{N}}^{\mathrm{T}} \ddot{\boldsymbol{e}}_{\mathrm{N}} = -\dot{\boldsymbol{e}}_{\mathrm{N}}^{\mathrm{T}} (\boldsymbol{I} - \boldsymbol{J}^+ \boldsymbol{J}) \boldsymbol{K} \underline{\boldsymbol{e}}_{\mathrm{N}} - \dot{\boldsymbol{e}}_{\mathrm{N}}^{\mathrm{T}} \boldsymbol{J}^+ \dot{\boldsymbol{J}} \dot{\boldsymbol{e}}_{\mathrm{N}} \tag{21.36}$$

Substituting formula (21.16) into formula (21.36), yields

$$\dot{V} = -\dot{\boldsymbol{e}}_{\mathrm{N}}^{\mathrm{T}} \boldsymbol{K} \dot{\boldsymbol{e}}_{\mathrm{N}} \tag{21.37}$$

Since $\boldsymbol{K}$ is a positive-definite matrix, then $\dot{V} < 0$, satisfied the Lyapunov criteria. Hence, the errors $\dot{\boldsymbol{e}}_{\mathrm{N}}$ converge to zero vector, i.e., $(\boldsymbol{I}\text{-}\boldsymbol{J}^+\boldsymbol{J})\boldsymbol{\Phi}$ tends to be equal to $\dot{\boldsymbol{q}}_{\mathrm{r}}^{\mathrm{N}}$, and the robot joint velocity can be expressed as

$$\dot{\boldsymbol{q}}_{\mathrm{r}} = \sum_{j=1}^{s} \frac{1}{\sigma_j} \boldsymbol{u}_j^{\mathrm{T}} \dot{\boldsymbol{p}}_{\mathrm{h}} \boldsymbol{v}_j + (\boldsymbol{I} - \boldsymbol{J}^+ \boldsymbol{J}) \boldsymbol{\Phi} \tag{21.38}$$

Since $e_{qn} > \varepsilon_e$, and $|e_{qi}| < \varepsilon_e$ for $i = 1, 2, \ldots, n - 1$, the clamping velocity satisfies

$$\varphi_n = \begin{cases} \frac{\left(e_{qn}-\varepsilon_e\right)^2\left(2e_{qn}+\varepsilon_e-3\bar{\varepsilon}_e\right)}{\left(\varepsilon_e-\bar{\varepsilon}_e\right)^3}\varphi_{\max} & \bar{\varepsilon}_e > e_{qn} > \varepsilon_e \\ \varphi_{\max} & e_{qn} \geq \bar{\varepsilon}_e \end{cases} \tag{21.39}$$

and

$$\varphi_i = 0 \quad (i = 1, 2, \cdots, n-1) \tag{21.40}$$

Then, the n-th joint velocity can be calculated as

$$\dot{q}_{\mathrm{r}n} = \sum_{j=1}^{s} \frac{1}{\sigma_j} \boldsymbol{u}_j^{\mathrm{T}} \dot{\boldsymbol{p}}_{\mathrm{h}} \boldsymbol{v}_j(n) + \left(1 - \sum_{j=1}^{s} \left[\boldsymbol{v}_j(n)\right]^2\right) \varphi_n \tag{21.41}$$

in which $\boldsymbol{v}_j(n)$ is the n-th element of $\boldsymbol{v}_j$, and

$$\sum_{j=1}^{s} \left[v_j(n)\right]^2 < \sum_{j=1}^{n} \left[v_j(n)\right]^2 = 1 \tag{21.42}$$

Since

$$\varphi_n \geq \frac{\left(e_{qn}-\varepsilon_e\right)^2\left(2e_{qn}+\varepsilon_e-3\bar{\varepsilon}_e\right)}{\left(\varepsilon_e-\bar{\varepsilon}_e\right)^3}\varphi_{\max} > \frac{\left|\sum_{j=1}^{s} \frac{1}{\sigma_j} \boldsymbol{u}_j^{\mathrm{T}} \dot{\boldsymbol{p}}_{\mathrm{h}} \boldsymbol{v}_j(n)\right|}{1 - \sum_{j=1}^{s} \left[\boldsymbol{v}_j(n)\right]^2} \tag{21.43}$$

the n-th joint velocity satisfies

$$\dot{q}_{\mathrm{r}n} > -\left|\sum_{j=1}^{s} \frac{1}{\sigma_j} \boldsymbol{u}_j^{\mathrm{T}} \dot{\boldsymbol{p}}_{\mathrm{h}} \boldsymbol{v}_j(n)\right| + \left|\sum_{j=1}^{s} \frac{1}{\sigma_j} \boldsymbol{u}_j^{\mathrm{T}} \dot{\boldsymbol{p}}_{\mathrm{h}} \boldsymbol{v}_j(n)\right| = 0 \tag{21.44}$$

21.4 Experimental Verification

21.4.1 Numerical Simulation

As a validation of the proposed method, a planar 3 DOF linkage non-humanoid robot controlled by human motion is simulated numerically. The robot demo is depicted in Fig. 21.3 and nomenclatures of the demo are shown in Table 21.1.

During the numerical simulation, the operator is simulated to execute an planar arc motion, which can be expressed as $q_{\mathrm{h}1} = q_{\mathrm{h}2} = q_{\mathrm{h}3} = (45\sin(0.2\pi t - \pi/2) + 45)$ deg. The sampling time is set as $\Delta t = 0.001$ s, and the up and down time delay are

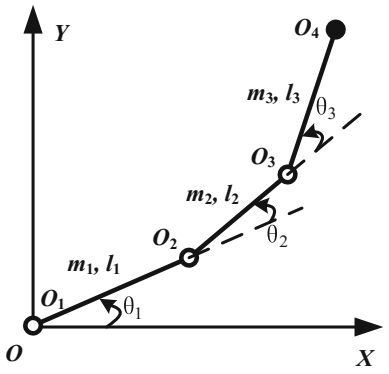


Fig. 21.3 Planar 3 DOF linkage non-humanoid robot

Table 21.1 Nomenclatures of the robot demo

Nomenclatures(unit)	Specification
XOY	World coordinate
O_1, O_2, O_3, O_4	Robot's joints(O_1, O_2, O_3) End-effector(O_4)
θ_1, θ_2, θ_3	Joint rotation angles
m_1, m_2, m_3	Linkage mass(kg) $m_1 = 30$, $m_2 = 20$, $m_3 = 10$
l_1, l_2, l_3	Linkage length(m) $l_1 = 0.3$, $l_2 = 0.2$, $l_3 = 0.1$

set as $t_u = t_d = 1.1$ s. The initial states of the robot are selected as $\boldsymbol{q}_r = [0\ 0\ 0]^T$, $\dot{\boldsymbol{q}}_r = [0\ 0\ 0]^T$. The external disturbances considered are $\boldsymbol{\tau}_d = [2\sin(t) + 0.5\sin(200\pi t),\ \cos(t) + 0.5\sin(200\pi t),\ \cos(2t) + 0.5\sin(200\pi t)]^T$, in which the second terms exhibit the effect of high frequency measurement noises. And the unmodeled error is considered as a 30% estimation error of the linkage inertias. In addition, the control system adopts PD, ATSM and proposed ATSMS methods, respectively, in which, the parameters of PD controller [18] are set as $k_p = 26$ and $k_v = 9$; the parameters of ATSM controller are set as $k = 0.1$, $\alpha = 0.6$, $\beta_1 = 20$, $\beta_2 = 20$ and $\eta = 0.7$; the parameters of ATSMS controller are set as: $\boldsymbol{K}$ is a *3*-dimensional identity matrix, $r = 1$, $h_0 = 0.01$, $\alpha_y = 0.5$, and the other parameters are the same as the ATSM controller.

The end position control errors between the human arm and robot are depicted in Fig. 21.4, which show that the end position control errors of ATSM and ATSMS method are about 0.01 m while that of PD method are about 0.05 m. In addition, the end position control errors of ATSM and ATSMS control method are quite similar, which indicates that the joint similarity control has no disturbances to the end position control.

The joint similarity errors between the human arm and robot are depicted in Fig. 21.5, and Fig. 21.6 shows the configuration changes of the human arm and robot. It is clear that the joint space of the robot controlled by the ATSMS method is

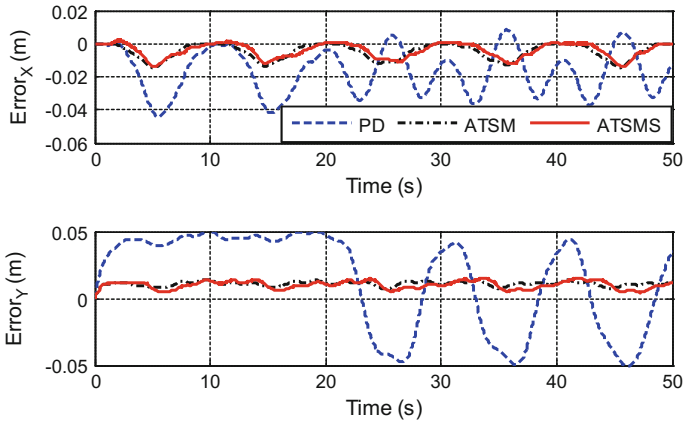


Fig. 21.4 The end position control errors between the human arm and robot

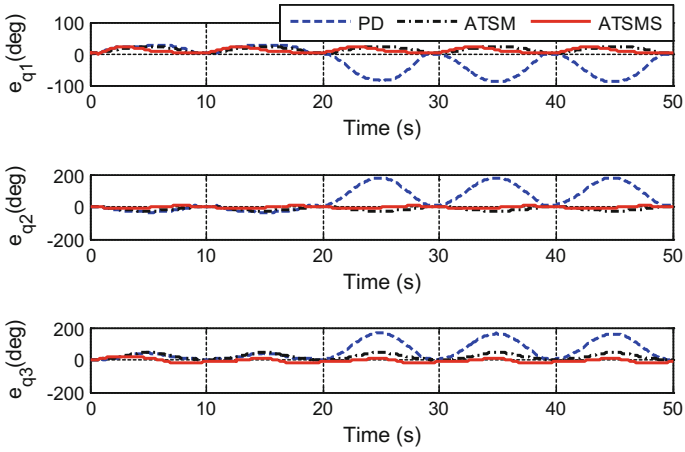


Fig. 21.5 The joint similarity errors between the human arm and robot

more similar to the human arm compared to the PD and ATSM method, and the joint similarity errors of ATSMS method are smaller than 10 deg.

In order to further compare the control effects of the three methods, the control accuracy of end position is defined as

$$TA = 1 - \frac{1}{T}\int_0^T \sqrt{\frac{\boldsymbol{e}_p^{\mathrm{T}}(t)\boldsymbol{e}_p(t)}{\boldsymbol{p}_h^{\mathrm{T}}(t)\boldsymbol{p}_h(t)}}\mathrm{d}t \tag{21.45}$$

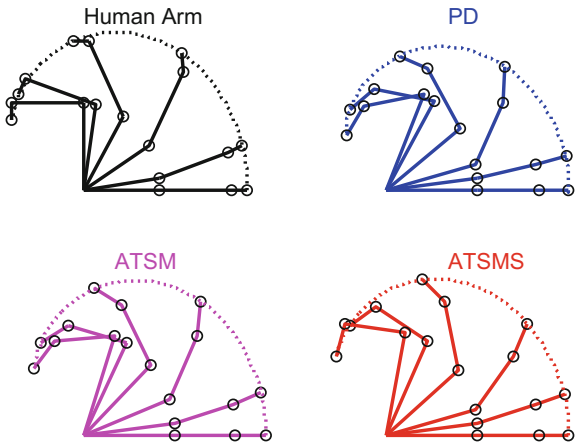


Fig. 21.6 The configuration changes of the human arm and robot

and from the view of joint trajectory consistency, the similarity function and average similarity for joint space are defined as

$$S(t) = \left[1 + \sum_{i=1}^{n} \left(\frac{q_{\mathrm{h}i} - q_{ri}}{q_{ri\max} - q_{ri\min}}\right)^2\right]^{-1} \tag{21.46}$$

and

$$\bar{S} = \frac{1}{T}\int_0^T S(t)\mathrm{d}t \tag{21.47}$$

respectively, in which the running time $T = 50$ s, and $q_{ri\min}$ and $q_{ri\max}$ are the minimal and maximal value of joint angles for i-th joint, respectively, which are set as $\boldsymbol{q}_{r\min} = [0, -90 \text{ deg}, -90 \text{ deg}]^{\mathrm{T}}$ and $\boldsymbol{q}_{r\max} = [180 \text{ deg}, 90 \text{ deg}, 90 \text{ deg}]^{\mathrm{T}}$.

The curves of similarity function for joint space are depicted in Fig. 21.7. It is obvious that with the processing of the numerical simulation, the joint space of the robot controlled by PD and ATSM method gradually deviates from the joint space of human arm while ATSMS can ensure the robot's joint space tends to the human arm all the time. The comparison results of end position control accuracy and average similarity are shown in Table 21.2, which demonstrates the better performance of ATSMS method in accurate controlling of end position and similarity of joint space between the robot and human arm compared to the PD and ATSM control method.

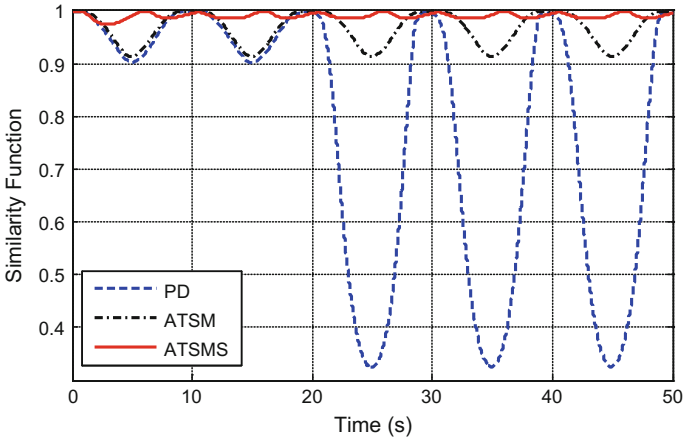


Fig. 21.7 The similarity function for joint space

Table 21.2 Comparison results of control accuracy and average similarity

Parameters	PD (%)	ATSM (%)	ATSMS (%)
TA	91.12	97.43	97.58
$\overline{S}$	78.04	96.70	99.06

21.4.2 Ground Experiments

The ATSMS control method can realize the simultaneous control of the end position and joint space of space robot, therefore, in the complex and unknown space environment, the operator can control the robot to carry out some complex operations flexibly, e.g., obstacle avoidance. A teleoperation ground experimental system based on human motion capture has been established, as shown in Fig. 21.8, where the motion capture device is an inertial motion capture device based on MEMS sensor and the controlled object is an industrial robot with 6 degrees of freedom. The target of the teleoperation task is that the operator controls the robot to avoid an unstructured obstacle based on the visual information and the end-effector of the robot to arrive at the desired position. PD and ATSMS control method are used in the teleoperation system, respectively.

The final state of the robot is shown in Fig. 21.9. It is obvious that, the PD control method can only control the end position of the robot, and when the operator finds the robot will hit the obstacle through the visual information, due to its inability to control the joint space of the robot, it is difficult to control the robot to avoid obstacles in order to control the end-effector reaches the designated position; while the ATSMS control method can control the joint space of the robot directly to avoid the obstacle without disturbances to the accurate control of end position, thus, the operator can complete the teleoperation task.

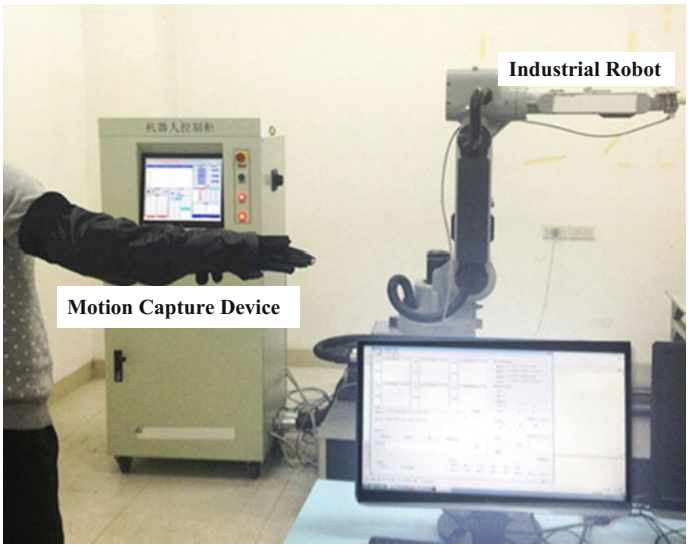


Fig. 21.8 A teleoperation ground experimental system based on human motion capture

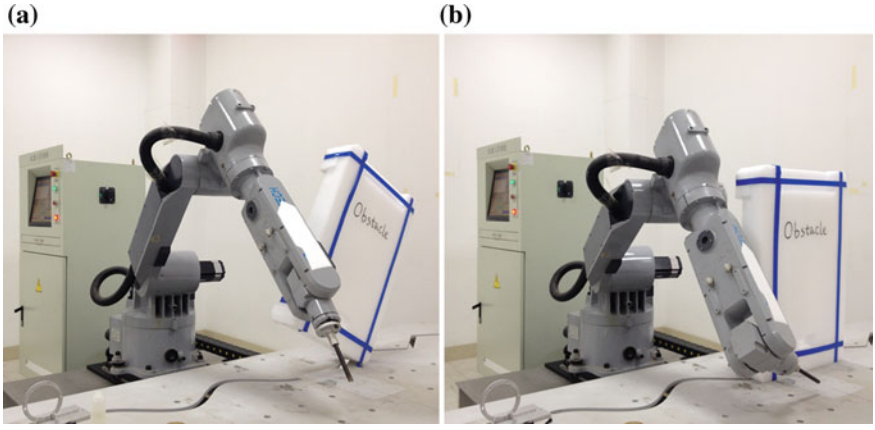


Fig. 21.9 The final state of the robot controlled by two method: **a** PD control method; **b** ATSMS control method

21.5 Conclusions

An ATSMS control method is proposed in this paper for the lack of flexibility of the traditional space teleoperation control method. An ATSM controller is used to guarante the accurate control of space robot end position due to its characteristics of robustness and anti-interference; and a similar control method is used by designing the clamping velocity to realize the direct control of space robot joint space and

improve the flexibility of operation. The results of numerical simulation showed that, compared with PD and ATSM control method, ATSMS control method can realize the higher control accurate of the robot's end position (about 97.58%) and higher joint similarity (the average similarity is up to 99.06%) between the human arm and robot. Finally, the ground experiment shows that by using the ATSMS control method, the operator can accurately control the robot end effector reach the designated position, and can control the robot joint space to avoid obstacles.

References

1. Wang Y, Xie Y, Zhou JL (2010) A research survey on teleoperation of space robot through time delay. J Astronaut 31(2):299–306
2. Huang PF, Liu ZX (2015) Space teleoperation technology. National Defense Industry, Arlington
3. Anderson RJ, Spong MW (1989) Bilateral control of teleoperators with time delay. IEEE Trans Autom Control 34(5):494–501
4. Niemeyer G, Slotine JJE (1991) Stable adaptive teleoperation. IEEE J Ocean Eng 16(1):152–162
5. Ali S, Shahin S (2009) Adaptive/robust control for time-delay teleoperation. IEEE Trans Robot 25(1):196–205
6. Xu WF, Zhou RX, Meng DS (2013) A hybrid force/position control method of space robot performing on-orbit ORU replacement. J Astronaut 34(10):1353–1361
7. Deng QW, Wei Q, Li ZX (2006) Numerical analysis of thermal environment of lunar surface. J Astronaut 27(2):196–200
8. Yang Y, Hua C, Guan X (2014) Adaptive fuzzy finite-time coordination control for networked nonlinear bilateral teleoperation system. IEEE Trans Fuzzy Syst 22(3):631–641
9. Yu S, Yu X, Shirinzadeh B, Man Z (2005) Continuous finite-time control for robotic manipulators with terminal sliding mode. Automatica 41(11):1957–1964
10. Neila MBR, Tarak D (2011) Adaptive terminal sliding mode control for rigid robotic manipulators. Int J Autom Comput 8(2):215–220
11. Bluethmann W, Ambrose R, Diftler M, Askew S et al (2003) Robonaut: A robot designed to work with humans in space. Auton Robots 14(2–3):179–197
12. Diftler MA, Mehling JS, Abdallah ME, Radford NA, et al. (2011) Robonaut 2-the first humanoid robot in space. 2011 IEEE International Conference on Robotics and Automation (ICRA)
13. Ke WD, Peng ZP, Cai ZS, Piao SH et al (2014) Study of trajectory tracking control for humanoid robot based on similarity locomotion. Acta Automatica Sinica 40(11):2404–2413
14. Koenemann J, Burget F, Bennewitz M (2014) Real-time imitation of human whole-body motions by humanoids. 2014 IEEE International Conference on Robotics and Automation (ICRA)
15. Hsu P, Mauser J, Sastry S (1989) Dynamic control of redundant manipulators. J Robotic Syst 6(2):133–148
16. Flacco F, De LA (2014) Discrete-time velocity control of redundant robots with acceleration/torque optimization proper-ties. 2014 IEEE International Conference on Robotics and Automation (ICRA)
17. Hu T, Huang X, Tan Q, Huang J (2012) Error feedback controller for autonomous space teleoperation. Procedia Eng 29(1):1142–1149
18. Silva GJ, Datta A, Bhattacharyya SP (2007) PID controllers for time-delay systems. Springer Science & Business Media, Berlin

Chapter 22
Analysis and Design of the Stabilization Loop for Ship-Borne Antenna Servo System

Jianhui Jia and Shuyang Zhao

22.1 Introduction

When flying target is tracked by ship-borne antenna,the tracking accuracy is reduced because of ship's shaking,and the target is even loss in bad sea conditions [1]. At this time, the stabilization loop that uses gyroscope as the feedback element is the commonly used [2], block diagram of stabilization loop was shown in Fig. 22.1[3], for AE type antenna, the gyroscope is installed on the elevation arm to obtain the rates of azimuth axis and elevation axis. In order to complete the stabilization loop design, the corresponding regulator is needed. In addition to the regulator and gyroscope, the speed loop is the important part of stabilization loop, and the bandwidth and the dynamic characteristics restrict the behavior of the stabilization loop [4] and make the regulator of the stabilization loop difficult to design. This paper is to provide a design method of the regulator and verify the correctness of the method in the actual system.

22.2 The Design of the Stabilization Loop Regulator

The design of regulator of servo system includes form selection and parameter debugging, and the form of regulator is generally determined in the early stage of design, and the parameters are obtained through being repeatedly debugged in the actual system.

In the servo loop, such as position loop, speed loop and current loop, PI regulator is usually used. Because of the feedback element of the stabilization loop is located on the load axis, it is especially sensitive to the noise, at the same time, the

J. Jia (✉) · S. Zhao
Beijing Research Institute of Telemetry, Beijing 100094, China
e-mail: 466403592@qq.com

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_22

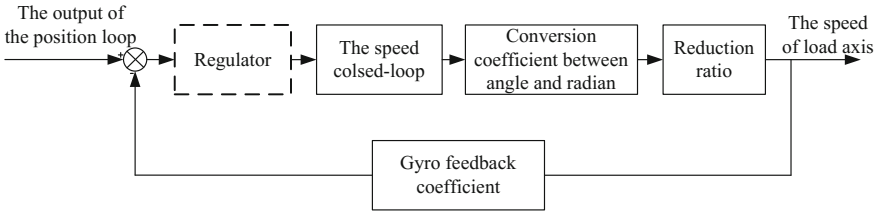


Fig. 22.1 Block diagram of stabilization loop

bandwidth of the stabilization loop is lower than the bandwidth of the speed loop, and that makes it difficult or impossible to adjust the parameters for the PI regulator.

In servo system, the bandwidth of servo loop usually meets the following relationship [5].

$$\omega_b = (1.5 \sim 2)\omega_c \tag{22.1}$$

ω_b is the closed-loop bandwidth and ω_c is the open-loop break frequency in the servo loop. In addition, the bandwidth of the inner loop is generally higher than that of the outer loop [6]. Because of speed loop is the inner loop of the stabilization loop, the break frequency ω_c of the stabilization loop is less than the bandwidth ω_b of the speed loop on the basis of the above relationship.

The Bode diagram of the speed closed-loop is generally consistent with the form shown in Fig. 22.2a, the purpose of the regulator of the stabilization loop is to transform it, so that the stabilization loop has good dynamic and steady state characteristics. The integrator is added to the regulator to improve the type of the stabilization loop. At this time, the form of the bode diagram that was transformed is consistent with the form that is shown in Fig. 22.2b. In order to satisfy the requirement of amplitude margin, the proportional gain is required, and the form of bode diagram as shown in Fig. 22.2c. The break frequency ω_c of the stabilization loop is more than the bandwidth of speed loop right now, and that does not accord with the result of the previous analysis, to meet the requirement of the break bandwidth of stabilization loop is lower than that of the speed loop, lag compensation is needed by using the feature of attenuating the amplitude of high frequency signal [7]. So far, the bode diagram is shown in Fig. 22.2d, and achieves the desired expectation.

According to the above analysis, the equation for the regulator of the stabilization loop is shown below.

$$G_{jz}(s) = \frac{K(\tau_1 s + 1)}{s(\tau_2 s + 1)} \tag{22.2}$$

K is proportional gain, τ_1 and τ_2 are the coefficients of lag compensation, and τ_1 is less than τ_2.

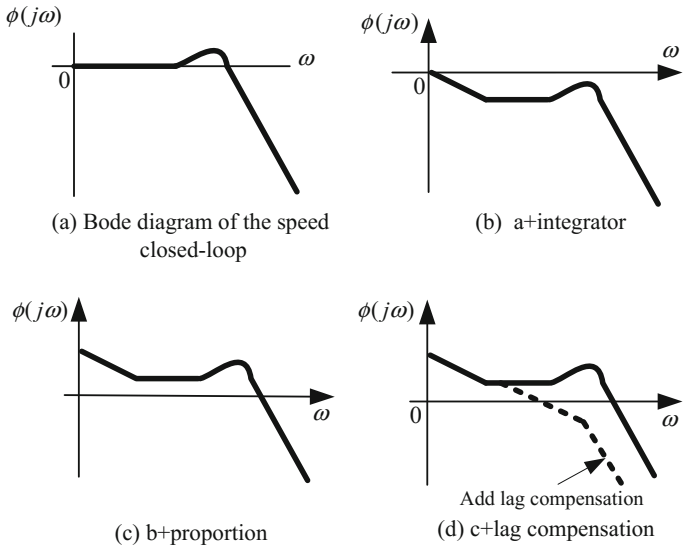


Fig. 22.2 The design flow diagram of the stabilization loop regulator

In the practical system, the speed loop bandwidth is relatively low, and that affects the width of the intermediate frequency signal of stabilization loop, furthermore, the dynamic characteristics is impacted [7]. In order to improve that, taking the regulator mentioned above and the speed loop as an object, then to adjust it by adding the PID regulator that is often used and we have rich debugging experience

22.3 The Realization of the Regulator

22.3.1 Specific Form of Regulator

According to the above analysis, the regulator designed in this paper is composed of PID part and integrator + lag compensation.

The form of PID algorithm that is realizated by computer is shown in formula (22.3).

$$u(k) = K_p \times e(k) + K_I \times \sum_{i=1}^{k} e(i) + K_d \times [e(k) - e(k-1)] \tag{22.3}$$

e is error that output of position loop minus feedback of gyroscope, u is the output of PID algorithm, K_p, K_I, K_d is the proportional integral differential coefficient.

The integrator + lag compensation corresponding formula (22.2) requires discretization before computer programming, and the difference equation is shown in formula (22.4) [8].

$$y(k) = a_1 \times y(k-1) - a_2 \times y(k-2) + b_1 \times K \times u(k-1) - b_2 \times K \times u(k-2) \tag{22.4}$$

u is the output of PID algorithm, K is the proportional gain that is equal to K of formula (22.2) and a_1, a_2, b_1, b_2 are the coefficients of differential term.

22.3.2 Parameter Design

Parameters of the regulator discussed in the servo system are usually obtained in final debugging process, while the initial values of the parameters need to be determined at the beginning of design. In antenna servo system, the speed loop is generally designed in the driver, whose bandwidth is fixed and not easy to be changed. In this article, Parameters of the regulator of stabilization loop are evaluated based on this.

Assuming that ω_n is the bandwidth of speed loop, ω_{gc} is the break frequency of stabilization loop, D is a technical index that reflects the ability of the servo system to isolate the ship's shaking and was put forward in the early design stage, and the unit is dB, Kg is the open loop gain of the stabilization loop. To guarantee the stability margin of the stabilization loop, the following relationships are generally required according to the experience that are accumulated in engineering practice.

$$\omega_{gc} = \frac{\omega_n}{3}, D = 20 \lg \frac{1}{1 + Kg} \tag{22.6}$$

The $\tau 1$ and $\tau 2$ in formula (22.2) are designed according to the following equation.

$$\tau 1 = 3/\omega_{gc}, \tau 2 = 3\tau 1 \tag{22.7}$$

Combined with Fig. (22.1) and Kg in formula (22.6), K in formula (22.2) can be calculated. Formula (22.2) can be discretized into the form of the formula (22.4) after completion of the above parameters, future more,a_1, a_2, b_1, b_2 can be determined in debugging process.

The parameter design of PID algorithm is no longer described in this paper due to being widely discussed in engineering applications.

22.4 The Experiment and Analysis

22.4.1 Step Test of Servo Loop

In the servo system, the response speed of the loop is closely related to the bandwidth [4], and the higher the bandwidth, the faster the response. Because the stabilization loop bandwidth is between the speed loop and the position loop, and the position loop bandwidth is less than the speed loop bandwidth, then the response speed of the stabilization loop is slower than that of the speed loop and it is faster than response speed of the position loop. The rise time of the step response of the loop reflects the response speed in servo system. We take this as the basis for the stabilization loop debugging.

Based on the above analysis, we have carried out experimental tests in our system. The position loop step response curve is shown in Fig. 22.3a, the speed loop step response curve is shown in Fig. 22.3b, and the stabilization loop step

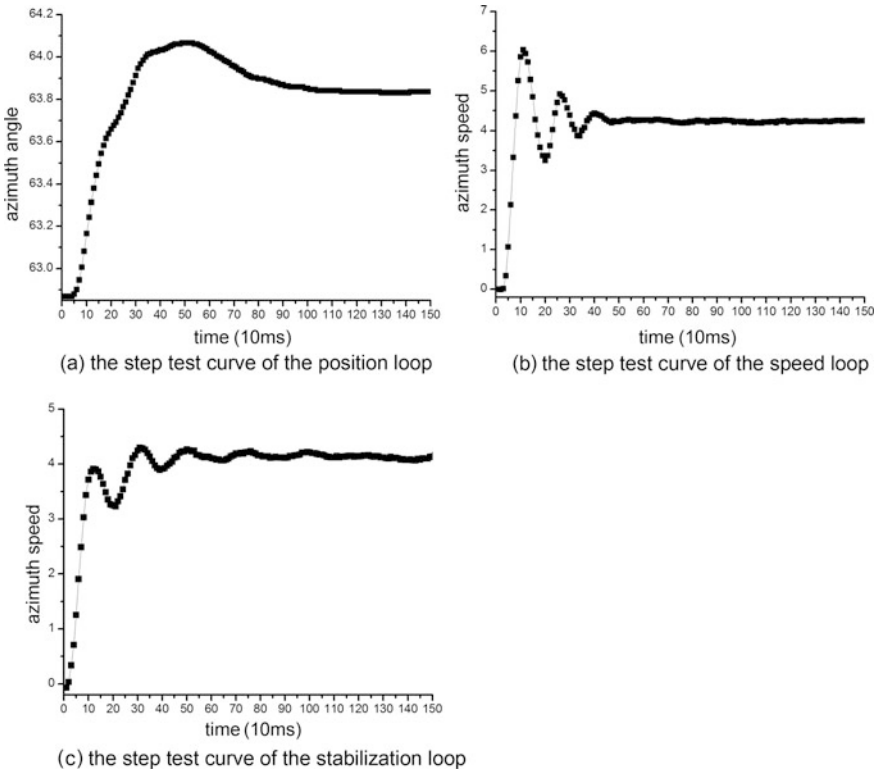


Fig. 22.3 The step test curves of multiple loops

response curve is shown in Fig. 22.3c. It can get that the rise time for the position loop is about 300 ms, and it is about 100 ms for the speed loop, and it is about 200 ms for stabilization loop. The experimental measurements are in good agreement with the above analysis result.

22.4.2 Test of Isolating the Ship-Swaying

After the stabilization loop design is completed, the ability of the servo system to isolate the ship's shaking is tested under two different conditions using and not using the stabilizing loop by using sinusoidal signal that amplitude is 8° and cycle is 10 s to simulate the shaking of the ship [9, 10]. The test result is shown in Fig. 22.4.

The test result is that the tracking error of azimuth axis is reduced from 0.1° to 0.05° and the tracking error of elevation axis is reduced from 0.24° to 0.49° after adding stabilization loop. It shows that it can improve tracking accuracy and isolate the disturbance that is caused by the shaking of the ship by adding stabilization loop, at the same time. It also shows that the design method of the regulator that is proposed in this paper is correct.

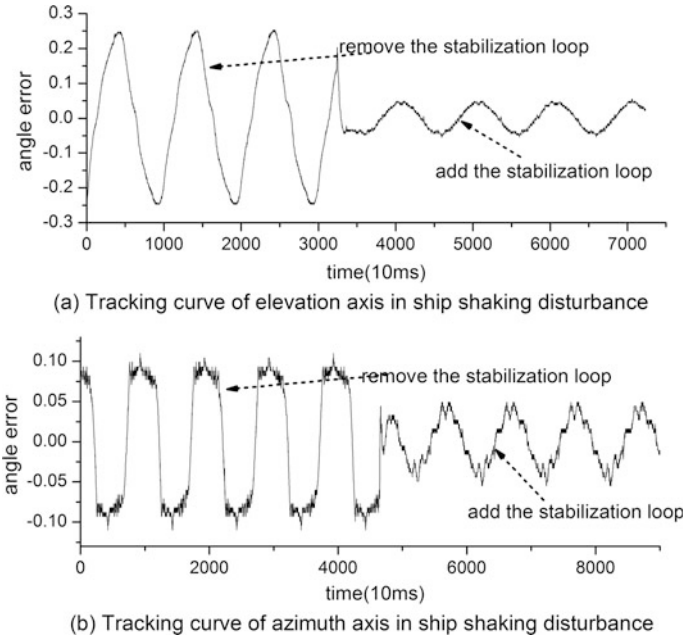


Fig. 22.4 Tracking angle error of antenna servo system in ship shaking disturbance

22.4.3 The Application of the Stabilization Loop

To make the dynamic characteristic of the stabilizing loop satisfied, it is often necessary to reduce the bandwidth of the position loop when the bandwidth of the speed loop is not high, the previous parameters of the position are usually no longer applicable, and they need to be re-adjusted. At the point, the response speed will drop.

After all design is completed, the step test of the position loop is carried out under two conditions using and not using the stabilizing loop. The test result is shown in Fig. 22.5.

From this figure, you can see that the rise time of the position loop is slow when the stabilization loop is used. If the flying target is highly mobile, the slow response of the position loop also affects the tracking accuracy. So it is necessary to weigh whether to use stabilization loop according to the feature of flying target and sea state, while it is mainly rely on human experience at present.

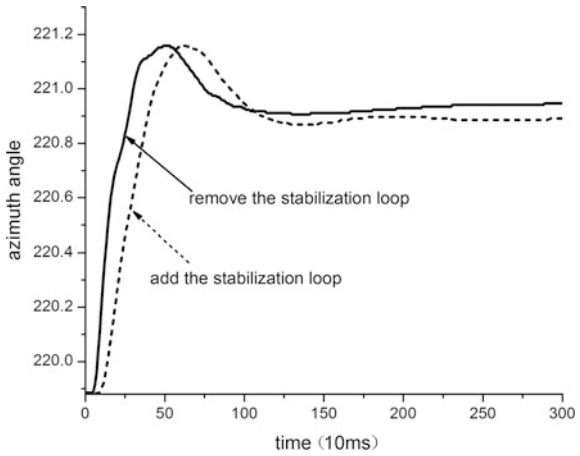


Fig. 22.5 The position loop step test curves before and after adding the stabilization loop

22.5 Conclusion

The experiment result demonstrates that it can isolate the disturbance caused by the shaking of the ship by adding stabilization loop in the ship-borne antenna servo system, and shows that the design method of the stabilization loop regulator that is presented in this paper is effective and has a certain engineering significance. In addition, selection of the stabilization loop is mainly rely on human experience at present, it is need to be further studied to give clear working conditions.

References

1. Qu YX (2009) Stability technique of TT&C equipments on the space tracking ship. National Defence Industry, Beijing
2. Wang XR, Tao M, Geng WJ (2013) Application of the kalman filter in the rate gyroscopes of ship-borne servo system. AMSMT 2013 International conference on advances in materials science and manufacturing technology. Transtech Publications, Fujian, pp 765–767
3. Pan GF, Zhou J (2007) Modeling and simulation of ship-borne radar servo system. Electron Eng 33(6):7–14
4. Shi JD (2004) Research on servo bandwidth and tracking performance. Electron Eng 30(9):4–5, 16
5. Dong RX (2003) Design for airborne radar serve bandwidth. Modern electronic technology. Communication 9:83–84
6. Chen BS (2003) Electricity pull automation control system. China machine press, Beijing
7. Hu SS (2007) Automatic control theory. Science press, Beijing
8. Luo LX, Wang XB, Zhou YH (2011) The digital of servo loop controller. Electron Des Eng 19(6):72–74
9. Tao B (2009) Simulation test method of isolating ship rocking interference in radar servo system. Inf Res 35(3):20–22
10. Qu YX, Zhou JB, Xie Y (2006) Research on automated test of DISRI of ship-borne TT&C equipment. J Spacecr TT&C Technol 25(4):62–65

Part II
Object Exploration and Identification

Chapter 23
Target Recognition of Radar HRRP Using the Envelope Reconstruction

Pengfei Zhang, Li Chan, Hongxi Zhou and Xiaguang Yu

23.1 Introduction

With the developing of wideband radar and signal processing technology, it can easily obtain the high resolution range profile (HRRP). It contains important target structure signatures, such as target size, scatterer distribution, etc. at that aspect. Therefore, target recognition of radar HRRP has received intensive attention for a real-time system from the radar automatic target recognition (RATR) community [1–12]. Li and Yang [1] directly used HRRP as feature vectors and developed matching scores as discriminate rule for target recognition. Feature transformation is an efficient method for feature extraction. Zyweck and Bogner [2] extracted magnitude by performing Fourier transform (FT) of HRRP to recognize the commercial aircraft. and Kim et al. [3] presented some time-shift invariant features of HRRP. Du et al. [4] utilizes the frequency domain feature with various higher-order spectra. In order to overcome the limitation of the highly sensitivity to target-aspect angle variation from radar, the consecutive range profile samples have to be divided into several aspect-frames [5, 6]. And then statistical models are developed to parameterize HRRP in a frame. Shi et al. [7] studied the two-phase model on local factor analysis. Gaussian mixture distribution [8] is used to model the short-term relation of consecutive HRRPs. Wang et al. [9] studied the T-mixture model for radar HRRP target recognition. Since HRRP is typical high-dimensional distributed signal, the traditional linear classify method often fail to gain good target recognition. Liu et al. [10] proposed support vector machine (SVM) for HRRP-based RATR. Lin et al. [11] applied support vector data description (SVDD) to the target rejection. Fu and Yang [12] presented a kernel trick via multiclass synthetical discriminant analysis for radar HRRP recognition. All the development trends

P. Zhang (✉) · L. Chan · H. Zhou · X. Yu
Xi'an Satellite Control Centre, Xi'an 710043, China
e-mail: zh_pengfei@sina.com

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_23

shows that with well feature extraction of target, how to design a effective classifier to measure test HRRP sample is also very important to HRRP target recognition.

Moments and various functions of moments have been extensively utilized as pattern features to achieve invariant recognition of 2-D image patterns [13, 14]. Kim [3] introduced the central moments into HRRP-based recognition and presented some time-shift invariant features of HRRP. However, since the monomial function of central moments is not orthogonal, the correlation of moment weighting kernels directly results in information redundancy. To deal with these redundancy contained in the central moments, the principal component analysis (PCA) is applied to reduce and achieve more dimensionality reduction. Since the transformations of feature vectors in PCA are based on the training data samples only and are applied to both the training and test feature vectors. When the training dataset is small, the transformations often can't effectively represent the intrinsic information of the target. It becomes necessary to review those orthogonal moments which are based on the theory of orthogonal polynomials. Different types of orthogonal moments [15, 16] defined in a rectangular region have been investigated as the unique image shape features. As one of the important orthogonal moments [17, 18], the Legendre moments have near zero information redundancy and show potential usefulness in the field of HRRP analysis.

In this paper, we propose a new algorithm of translation invariant feature extraction based on Legendre moments. With preprocess of aspect-frames separated, amplitude ℓ_1 normalized and range values mapped, the calculated Legendre moments feature vectors are dependent not on the amplitude or position of HRRP but only on its shape. Meanwhile, spectrum analysis (SA) was adopted to reconstruct the real HRRP and extract principal information in order to further improve recognition performance of HRRP obtained by inverse fast fourier transform (IFFT) in low SNR environment. The recognition experiments based on measured data are simulated to verify the effectiveness of the proposed algorithm.

The rest of this paper is organized as follows. In Sect. 23.2, we analyze the real HRRP features. In Sect. 23.3, the detail target recognition scheme is described. In Sect. 23.4, Several experiments based on measured data are performed to evaluate the recognition performances of the proposed algorithm. Finally, conclusions are made in Sect. 23.5.

23.2 Analysis of the Real HRRP Features

The moving toward range cells (MTRC) and noise effect are the most important phenomena causing the target-aspect sensitivity of HRRP [4–6]. Therefore, how to overcome this problem and establish a robust template database is a very important issue, which will be discussed in this section.

23.2.1 Signal Model of HRRP

Even for high resolution radar, the targets size is still much larger than the wavelength of the radar, the electromagnetism characteristics of targets can be described by the scattering center target model [4]. According to this model, a HRRP is the vectorial summations of the returns from the target's scattering centers projected onto the radar line-of-sight (LOS) as show in Fig. 23.1 [7], which represents the scattering centers distribution of a target along the radial distance.

In real applications radar, we can only collect limited discrete sample frequency points N within a finite bandwidth of frequency B. Therefore, the returned echo in frequency domain is described as formula (23.1).

$$X(k) = \sum_{i=1}^{K} \sigma_i \exp[-j4\pi kBR_i/Nc] = \sum_{i=1}^{K} \sigma_i \exp[-j4\pi kR_i/c \cdot \triangle f] \tag{23.1}$$

where the frequency Δf and the distance R_i constitute a pair fourier transform (FT) variables. For each sample echo in a aspect angle, the K scatterers distribute in N range cell. Suppose there are K_n scatterers in mth sample echo in the nth range cell then echo signal in a range cell can be written as

$$x_n(m) = \sum_{i=1}^{K_n} \sigma_{ni} \exp[-j(4\pi \triangle r_{ni}(m)/\lambda + \phi_{ni0})] \triangleq \sum_{i=1}^{K_n} \sigma_{ni} \exp[-j\phi_{ni}(m)] \tag{23.2}$$

where σ_{ni} represent the strength of the ith scatterer in the nth range cell, $\triangle r_{ni}(m)$ denotes the radial displacement of the ith scatterer in the nth range cell ϕ_{ni0} represent the initial phase.

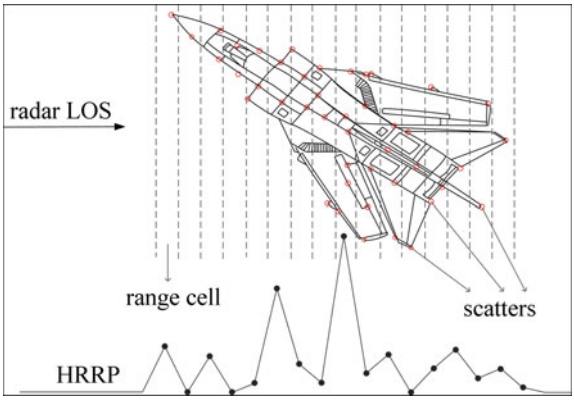


Fig. 23.1 Illustration of target HRRP [7]

23.2.2 Properties of Real HRRP

As the above derivation, the mth sample real HRRP can be defined as

$$x(m) = [|x_1(m)|, \ldots, |x_n(m)|, \ldots, |x_N(m)|]^T = \left[\left|\sum_{i=1}^{K_1} \sigma_{1i} \exp[-j\phi_{1i}(m)]\right|, \ldots, \left|\sum_{i=1}^{K_{N-1}} \sigma_{ni} \exp[-j\phi_{ni}(m)]\right| \ldots, \left|\sum_{i=1}^{K_N} \sigma_{Ni} \exp[-j\phi_{Ni}(m)]\right|\right]^T \tag{23.3}$$

One property as stated in [5] is that it is very sensitive to time-shift, target-aspect and amplitude-scale variation and needs to deal with this three sensitivities. Another property is that in a subset of HRRPs with the limitation of the target-aspect angle avoiding the occurring of MTRC, HRRP has a statistical invariant.

Define the power of $x_n(m)$ in formula (23.3) as

$$|x_n(m)|^2 = x_n(m)x_n^*(m) = \sum_{i=1}^{K_n} \sigma_{ni}^2 + 2\sum_{i=1}^{K_n}\sum_{k=1}^{i-1} \sigma_i \sigma_k \xi_{nik}(m) \tag{23.4}$$

The first term at the right side of formula (23.4) is referred as the scatterer auto term (SAT), while the second term at the right side of formula (23.4) is referred as the scatterer cross term (SCT). Without scatterers' MTRC, the SAT can be considered as a stable variable, and the SCT tends to be a random variable with zero mean [4, 5]. Therefore, we can obtain the expectation of the power of $x_n(m)$

$$E\left\{|x_n(m)|_x^2\right\} = E\left\{\sum_{i=1}^{K_n} \sigma_{ni}^2 + 2\sum_{i=1}^{K_n}\sum_{k=1}^{i-1} \sigma_i \sigma_k \xi_{nik}(m)\right\} = \sum_{i=1}^{K_n} \sigma_{ni}^2 \tag{23.5}$$

The result in formula (23.5) means that the average HRRP in a HRRP aspect-frame is invariant therefore can be used to reduce the target-aspect sensitivity.

Considering the actual measured data affected by noise, the signal-to-noise ratios (SNR) is defined as [5]

$$\mathrm{SNR} = 10 \times \log_{10}\left(\overline{P}_x / P_n\right) = 10 \times \log_{10}\left(\sum_{i=1}^{N} P_x(i) \Big/ N \cdot P_n\right) \tag{23.6}$$

where $\{P_x(i)\}$ denote the power of the signal in per range cell,$\overline{P}_x$ denotes the average signal power, $P_n = w_n^2$ denotes the power of noise, N denotes the number of range cells.

As with formula (23.3), the detailed SNR represents of a real HRRP can be given as

$$SNR = 10 \times \log_{10}\left(1/N\sum_{i=1}^{N}|x_i(m)|^2/P_n\right) \approx 10 \times \log_{10}\left(1/N\sum_{i=1}^{K}\sigma_i^2/w_n^2\right) \tag{23.7}$$

In a HRRP aspect-frame, we can also some equalities of the average HRRP

$$\begin{cases} P_x^{Ave}(i) = |\bar{x}_i(m)|^2 = \left(1/M\sum\limits_{m=1}^{M}|x_1(m)|^2\right) \\ \overline{P}_x^{Ave} = 1/N\sum\limits_{i=1}^{N}P_x(i) = 1/MN\sum\limits_{m=1}^{M}\sum\limits_{i=1}^{N}|x_i(m)|^2 \\ E\left\{\overline{P}_x^{Ave}\right\} = E\left\{1/MN\sum\limits_{m=1}^{M}\sum\limits_{i=1}^{N}|x_i(m)|^2\right\} = 1/N\sum\limits_{i=1}^{K}\sigma_i^2 \end{cases} \tag{23.8}$$

Suppose the noise level keep stable w_n^2, then the SNR of average HRRP in a HRRP aspect-frame can be given as follow

$$SNR = 10 \times \log_{10}\left(\overline{P}_x^{Ave}\Big/P_n\right) \approx 10 \times \log_{10}\left(1/N\sum_{i=1}^{K}\sigma_i^2/w_n^2\right) \tag{23.9}$$

Compare formula (23.7) with formula (23.9), we can easily get a consultation that the SNR of average HRRP in a HRRP aspect-frame is equal to the original SNR of any HRRP in that aspect-frame. That means the average HRRP contribute to reduce the SCT, but do nothing improving the SNR.

23.2.3 Reconstruction of the Real HRRP with SA

It was noted that the HRRP obtained by IFFT in formula (23.1) often result in a limited range resolution and ignore noise effect. The rewritten format from formula (23.1) with measured noise is given as

$$X(k) = \sum_{i=1}^{K}\sigma_i \exp[-j4\pi R_i k/c \cdot \triangle f] + \omega(k) \tag{23.10}$$

In the above, conventional IFFT bring measured noise into HRRP. A great number of different algorithm have been proposed, such as the MUSIC [19], Prony [20], Relax [21],to provides super-resolved HRRP, but studies about their noise effect relieve and application to target recognition are rare. Since our HRRP used for recognition is real, we have adopted a method via spectrum analysis (SA) to suppress noise and reconstruct the real HRRP.

The real HRRP in formula (23.3) can be rewritten as

$$x(m) = [\sigma_1(m), \ldots, \sigma_n(m), \ldots, \sigma_N(m)]^T \tag{23.11}$$

where

$$\sigma_n(m) = \left| \sum_{i=1}^{K_{N-1}} \sigma_{ni} \exp\left[-j\phi_{ni}(m)\right] \right|$$

Considering noise effect, we operate formula (23.11) with FFT, the spectrum form can be written as

$$X(k,m) = \sum_{i=1}^{N} \sigma_i(m) \exp\left[-j2\pi k r_i\right] + \omega(k) \tag{23.12}$$

It is noted that there are no difference in formats between formula (23.11) with formula (23.3) and formula (23.12) with formula (23.10), but has completely different meaning. In formula (23.11) $\sigma_n(m)$ no longer represent the strength of a scatterer in a practical situation, but can be seen as one and only one equivalent scatterer in a range cell and also changes with sample instant m. In formula (23.12) r_i no longer represent the distance between radar and the scatterer and just a parametric variable related to the sequence of range cell. N represents the number of range cell. For analysis conveniently, formula (23.12) can be written as matrix form

$$X = Ea + \omega \tag{23.13}$$

In the above, the problem is considered as the estimation of the parameters $\{\sigma_i, r_i\}$ from the frequency data $X(k,m)$ for the noise model given in formula (23.12). This problem can be solved by minimizing the following nonlinear least-squares (NLS) criterion [21].

$$\{\sigma_i, r_i\} = \arg\min_{\{\sigma_i, r_i\}} \|X - Ea\|^2 \tag{23.14}$$

Actually, we can also apply Relax algorithm to (13)form the complex HRRP and obtained all the parameters of K scatterers and then reconstruct the real HRRP, which is equivalent to formula (23.12). In narrow radar, a first-order approximation of RCS for a complex target can be obtained by summing the RCS components given as

$$\sigma \cong \left| \sum_{i=1}^{M} \sigma_i \exp\left[-j4\pi r_i/\lambda\right] \right| \tag{23.15}$$

Similarly, a range cell can be seen as a micro narrow radar for the K_n target scatterers in that range cell as show in formula (23.11). Moreover, take the average HRRP into consideration

$$\bar{x}(m) = [\bar{\sigma}_1(m), \ldots, \bar{\sigma}_n(m), \ldots, \bar{\sigma}_N(m)]^T \tag{23.16}$$

where-

where $\bar{\sigma}_n(m)$ is the equivalent scatterer obtained by SA form real HRRP and $\left(\sum_{i=1}^{K_n} \sigma_{ni}^2\right)^{1/2}$ is the average of all scatterers obtained by Relax from complex HRRP. Above means to reconstruct the average HRRP, SA also can obtain the same noise mitigate as Relax. SA can be seen as a first-order approximation of Relax algorithm [21].

23.3 Target Recognition Scheme

Target recognition scheme consists of training (or learning) and testing (or recognition) phases on data and also of feature extraction and classification phases on procedure. Figure 23.2 shows a typical flow chart of HRRP-based recognition. In the train phase, the training samples must be divided into several aspect-frames as

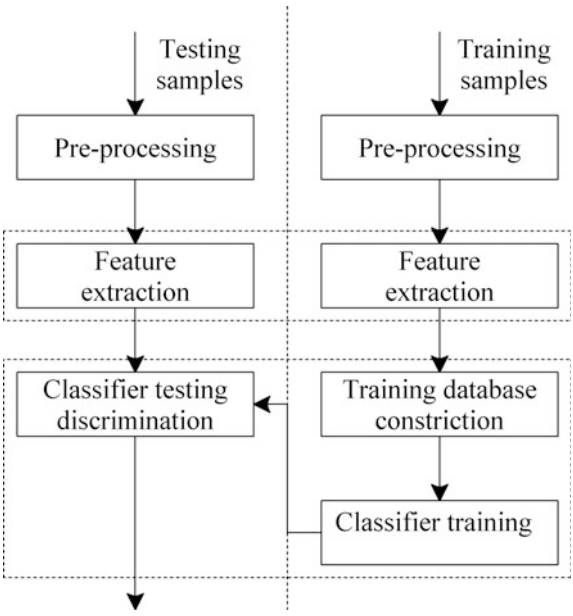


Fig. 23.2 Typical flow chart of radar HRRP recognition

noted before to obtain templates. Then in order to extract features of templates and test samples, a preprocessing is required to be applied to the data.

23.3.1 Preprocessing

Firstly, since it needs only to store the waveforms of HRRP as a function of aspect to construct the database for target recognition, we and then normalize each HRRP with amplitude ℓ_1 normalization.

$$\tilde{x}(m) = [\tilde{x}_1(m), \ldots, \tilde{x}_n(m), \ldots, \tilde{x}_N(m)]^T \tag{23.17}$$

With the above ℓ_1 normalization, the desired amplitude-scale invariance has been achieved and moreover the normalized HRRP can be seen as one-dimensional (1-D) probability density function.

Secondly, the center of coordinate region $[1, 2, 3, \ldots, N]$ is reset to

$$n_0 = \sum_{n=1}^{N} n\tilde{x}_n(m) \tag{23.18}$$

where n_0 is the 1th order origin moments, which can be seen as the mean of HRRP associated with time-shift as shown in Fig. 23.3.

Thirdly, the new coordinate region is mapped into [−1, 1] by choosing the new resolution as 1/*N*.

$$\eta(n) = (n - n_0)/N\ ,\ n = 1, 2, 3, \ldots, N \tag{23.19}$$

The final region can meet to the orthonormal interval of Legendre function.

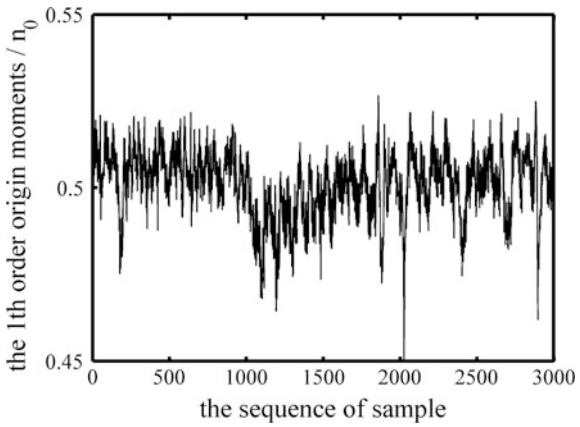


Fig. 23.3 Illustration of the 1th order origin moments n_0 drift

23.3.2 Feature Extraction Based on Legendre Moments

The mth order Legendre moments [15, 16] L_m of a function $f(x)$ are defined as

$$L_m = \int_{\zeta} p_m(x) f(x)\, dx \tag{23.20}$$

where $P_m(x)$ is the mth Legendre polynomial in the coordinate region ζ

After the preprocessing of HRRP above, its kth order Legendre moments can be computed by

$$L_k = \sum_{n=1}^{N} p_k(\eta(n))\tilde{x}_n(m) \tag{23.21}$$

The Legendre moments and the Central moments are related by [14]

$$L_k = (2k+1)/2 \sum_{k=1}^{k} a_{kj}\mu_j \tag{23.22}$$

where

$$\begin{cases} a_{kj} = \frac{2k+1}{2} \sum_{j=0}^{k} \binom{k}{j} \binom{(k+j-1)/2}{k} \\ \mu_j = \int_{\zeta} (x-\eta_x)^j f(x)\, dx \\ \eta_x = \int_{\zeta} x f(x)\, dx \end{cases}$$

Using the Legendre moments in formula (23.21), since $L_1 \equiv 0$, we choose the order form two to generate the feature vector

$$v = [L_2, L_3, \ldots, L_m]^T \tag{23.23}$$

where m denotes the maximum order of Legendre moments. There is no absolutely right way of choosing optimum m. Considering the relationship between Legendre moments and Central moments and also According to analysis in Sect. 1.2.3 and experiments in, we choose $m = 20$, which is sufficient for classifying different targets, to be discussed later.

To design a classifier, we need to construct a training database containing different aspect angles and targets. Suppose there are N_c targets and N_a aspect-frame corresponding to $Q = N_c \times N_a$ average HRRPs, we can obtain the feature vector L and the training database V as follows:

$$\begin{cases} V = [v_1, v_2, \ldots, v_Q] \\ v_j = [L_2, L_3, \ldots, L_m]^T,\ j = 1, 2, \ldots, Q \end{cases} \tag{23.24}$$

23.3.3 Classification Algorithms Using SVM

With the final the training database V, how to improve the correct recognition rate is an important problem to be considering in designing a favorable classifier. Using the structure risk minimization criterion and the kernel trick, SVM can map features from the original feature space to a high dimensional space feature space, only a (typically small) subset of the training data, which is commonly called "support vectors" [22] need to be used. Therefore, SVM is a promising candidate for HRRP classification. Unfortunately, SVM are binary classifiers [23] and can't generalize naturally to the multi-class classification case. There are two approaches have been suggested to perform multi-class classification. One is one-versus-all, Another is one-versus-one. In this paper, we have only 4 different targets and focus on studying the performance of target recognition based on Legendre moments descriptors, there are not much significant difference between the two approaches and choose one-versus-one here.

23.4 Experiment Results

23.4.1 Recognition Results for Measured Data

To demonstrate the classification performance of proposed algorithm, several experiments are performed on continuously measured data of four different targets recorded by a X-band ISAR experimental system. The system transmits 800-MHz linear modulated chirp signal with 80-μs pulse width. The measured HRRP is a 1600-dimensional vector, for the length of IFFT is 1600 points. The measured data of each target are divided into several subsets. The maximum angle difference between any two adjacent sets is about 6 degrees. The training data set and test data set are chosen from different data subsets respectively. For each target, 3072 HRRPs are chosen equidistantly from about 40,000 HRRPs for testing, and 72 average HRRPs are used for training, each of which is averaged by 128 HRRPs. The adjustable parameter σ in one-versus-one SVM classifier is set as 1. The correct recognition rate can be estimated [3].

$$P_c = \frac{\text{Number of correct classifications}}{\text{Number of test samples}} \tag{23.25}$$

Table 23.1 Confusion matrix of Legendre moments descriptors

Actual class	Classification			
	Target-1	Target-2	Target-2	Target-4
Target-1	2831	42	24	175
Target-2	186	2144	693	49
Target-3	29	593	2424	26
Target-4	423	14	16	2619

The selected parameters are: maximum order m = 20, adjustable parameter $\sigma = 1$ (overall correct recognition rate $p_c = 81.5\%$)

Table 23.2 Confusion matrix of Central moments descriptors

Actual class	Classification			
	Target-1	Target-2	Target-2	Target-4
Target-1	2085	293	309	385
Target-2	277	1519	1076	200
Target-3	283	842	1745	202
Target-4	1918	154	351	649

The selected parameters are: maximum order m = 20, new feature dimension = 7, adjustable parameter $\sigma = 1$ (overall correct recognition rate $p_c = 48.8\%$)

Table 23.3 Confusion matrix of Legendre moments descriptors

Actual class	Classification			
	Target-1	Target-2	Target-2	Target-4
Target-1	2659	21	32	360
Target-2	11	2173	852	36
Target-3	38	931	2006	97
Target-4	812	15	23	2222

The selected parameters are: maximum order m = 8, adjustable parameter $\sigma = 1$ (overall correct recognition rate $p_c = 73.7\%$)

We choose maximum order m = 20 and new feature dimension of Central moments is equal to 7. Another maximum order m = 8 for Legendre moments is chosen for the same dimension as Central moments. The experimental results are shown in Tables (23.1, 23.2) and (23.3). Note that target-4 based on Central moments is unable to recognize for its correct recognition rate P_c is lower than 30%, and using Legendre moments descriptors it can achieve high correct recognition rate $P_c = 85\%$ in Tables (23.1) and $P_c = 72\%$ in Table (23.3), while the corresponding overall correct recognition rate is 81.5, 48.8 and 73.7% respectively.

23.4.2 Recognition Performances of Legendre Moments Versus Central Moments with Different Orders

Firstly, it is necessary review the moments' noise sensitivity of HRRP. Let $f(x)$ be the measured of a HRRP, a reasonable autocorrelation function for $f(x)$ is

$$K_{ff}(x,u) = K_{ff}(0)\exp\left(-\lambda|x-u|\right) \tag{23.26}$$

where $K_{ff}(0) = E\left\{[f(x)]^2\right\}$ is the average energy of $f(x)$, λ is positive constants to be chosen for representing the amounts of detail in HRRP. We can estimate the approximate value of λ. Experimental evidence indicates that λ is often larger than 20, take Fig. 23.4 as instance. And also, the mapped coordinate region ζ is a subset of the interval $[-1, 1]$, which means the value $f(x)$ vanish outside ζ. These two respects effect will be analyzed in detail bellow.

For conveniently, we define the Legendre moments and Central moments as following general form:

$$\Phi_m = \int_{\zeta} \phi_m(x) f(x)\, dx \tag{23.27}$$

where $\phi_m(x)$ represent the moments kernel function and Φ_m is the mth order moments of $f(x)$.

The noise $n(x)$ is assumed zero mean and white with autocorrelation function

$$K_{nn}(x,u) = \sigma_n^2 \delta(x-u) \tag{23.28}$$

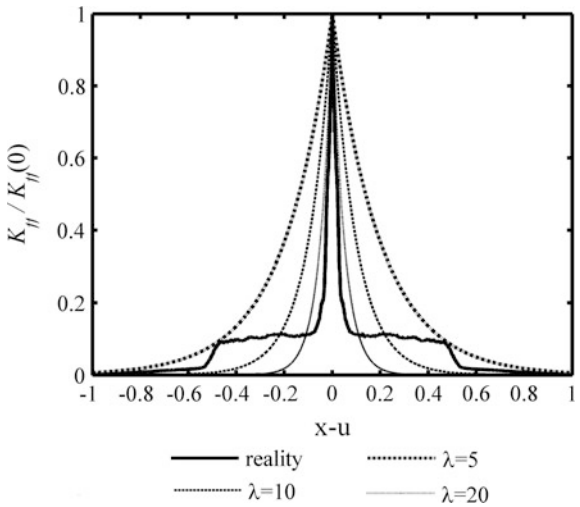


Fig. 23.4 Comparison measured HRRP autocorrelation function with various parameters empirical function

where σ_n^2 is spectral density. Similarly the moments of noise is defined

$$N_m = \int_{\zeta} \phi_m(x) n(x) dx \tag{23.29}$$

Then we define the normalized SNR (NSNR) of mth order moments as

$$\overline{\mathrm{SNR}}_m = \mathrm{SNR}_m \frac{\sigma_n^2}{K_{ff}(0)} = \frac{E\{\Phi_m^2\}}{E\{N_m^2\}} \cdot \frac{\sigma_n^2}{K_{ff}(0)} \tag{23.30}$$

The explicit NSNR for Legendre moments and Central moments are evaluate as in formula (23.31).

$$\begin{cases} \mathrm{SNR}_m^L = \frac{2m+1}{2\sigma_n^2} \iint p_m(x) p_m(u) K_{ff}(x,u) dx du \\ \mathrm{SNR}_m^C = \frac{2m+1}{2\sigma_n^2} \iint (xu)^m K_{ff}(x,u) dx du \end{cases} \tag{23.31}$$

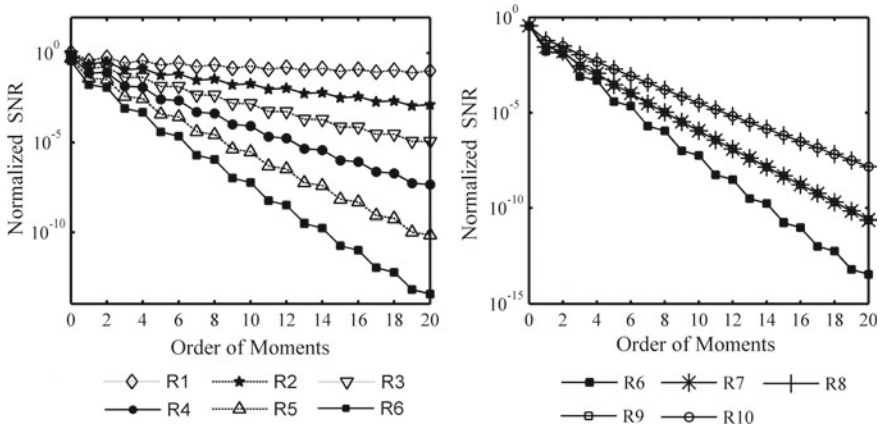
Considering the mapped coordinate region firstly, the ideal center in formula (23.18) is zero and the ideal mapped coordinate region in formula (23.19) is [−0.5 0.5], and often drift as shown in Fig. 23.3 such as [−0.6 0.4] or [−0.1 0.9]. Experimental evidence indicates that the drifts large than 0.3 like [−0.3 0.7] or [−0.7 0.3] appear sparingly. So we assume [−0.7, 0.7] is the drift bound.

For illustration purposes, we choose λ = 1 and different coordinate region listed in Table 23.4. The NSNR defined in formula (23.30) are plotted in Figs. 23.5 and 23.6.

Both figures show the general trend that noise sensitivity increase in higher moments. Figures 23.5 a and 23.6a show that with effective coordinate region decreasing, higher Central moments are extremely more sensitive, while higher Legendre moments can maintain stable. Figures 23.5b and 23.6b show that with coordinate region drifting around the ideal mapped interval [−0.5 0.5], higher Central moments can be mitigated but still more sensitive than the drift bound correspond to the absolute maximum for interval, while higher Legendre moments

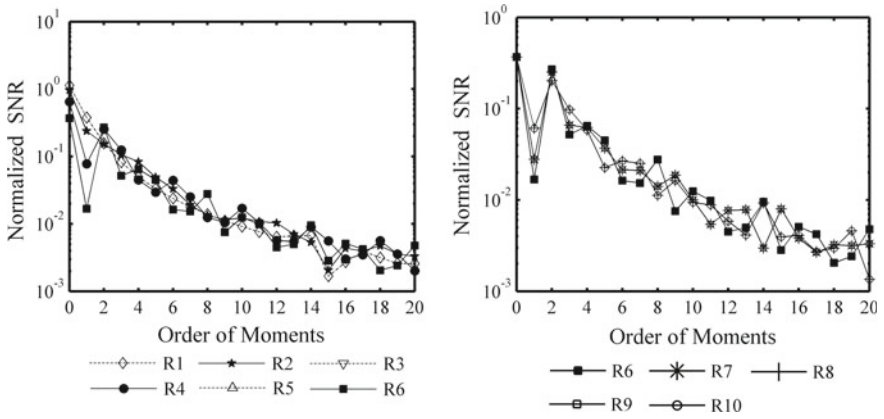
Table 23.4 Interval set

Symbol	Interval
R1	[−1.0,1.0]
R2	[−0.9,0.9]
R3	[−0.8,0.8]
R4	[−0.7,0.7]
R5	[−0.6,0.6]
R6	[−0.5,0.5]
R7	[−0.6,0.4]
R8	[−0.7,0.3]
R9	[−0.4,0.6]
R10	[−0.3,0.7]



(a) Illustration of decrease coordinate region (b) Illustration of coordinate region

Fig. 23.5 Illustration of different coordinate region for the Central moments ($\lambda = 1$)



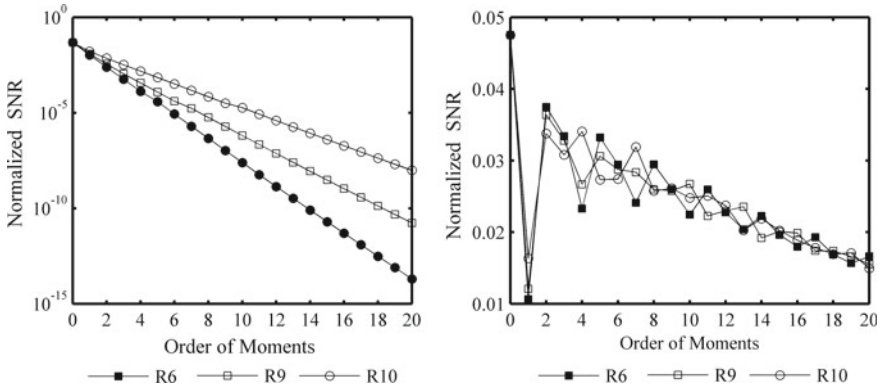
(a) Illustration of decrease coordinate region (b) Illustration of drift coordinate region

Fig. 23.6 Illustration of different coordinate region for the Central moments ($\lambda = 1$)

can maintain stable. In addition, the NSNR are same for mirror coordinate region since NSNR in (38) depend only on the absolute values, which can be seen in Figs. 23.5b and 23.6b.

Further $\lambda = 20$ is chosen, which is more precise to approximate the actual, is computed as show in Fig. 23.7. One measured HRRP is also computed as show in Fig. 23.8. All figures show the same conclusions as discussed above.

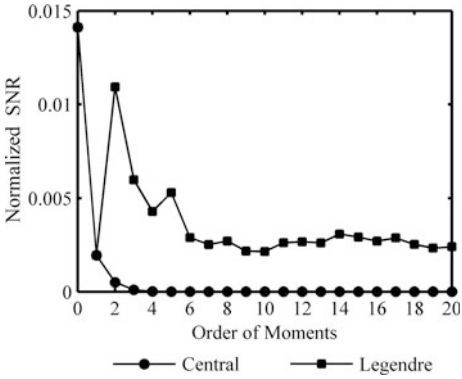
To demonstrate conclusion above, we choose different maximum order from 2 to 20 for Legendre moments named as case 1, and different maximum order from 2 to 20 with the same new feature dimension for Central moments named as case 1.



(a) Illustration of decrease coordinate region (b) Illustration of drift interval

Fig. 23.7 Illustration of different coordinate region for the Legendre moments (λ = 20)

Fig. 23.8 Illustration of different coordinate region for the Central moments and the Legendre moments of measured HRRP



Case 3 is defined as maximum order m = 20 with different new feature dimension from 1 to 19 of transformation matrix for Central moments. We run twenty Monte-Carlo simulations and the average resultants are show in Fig. 23.9.

Case 1 shows that the general trend of P_c for Legendre moments increase with the increase of the maximum order but doesn't change much more when the maximum order beyond 10 sometimes however even increase slightly. It is obvious that the higher order moments is more sensitive to noise and also considering the compute complex and real time required, it is enough to make a choice between 10 and 15. Case 2 shows that with low order moments Central moments can obtain the similar recognition performance as Legendre moments. However with the increase of the maximum order, the affection of information redundancy of the Central appears. PCA is applied to obtain transformation matrix, eigenvector with low eigenvalue can't provide robust features therefore cause the decrease of P_c.

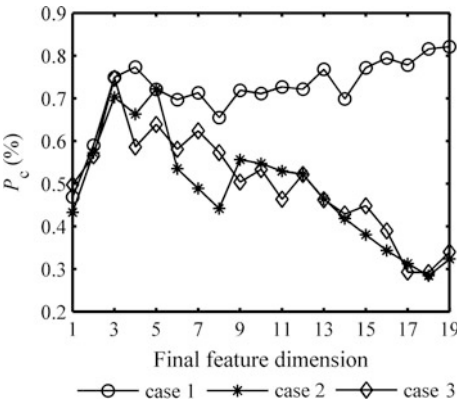


Fig. 23.9 P_c of the Legendre moments versus Central moments

Case 3 shows that with the increase of new feature dimension of transformation matrix, more eigenvectors with low eigenvalue are chosen. With the new feature dimension beyond 7, P_c decreases strictly as the same trend in case 2.

Compare case 1 with case 2 and case 3, we can see that Legendre moments based on orthogonal basis circumvent the problem of information redundancy and obtain better recognition results than Central moments. We can choose enough order moments to construct feature vector if the SNR of high order moments is acceptable.

23.4.3 Comparison Between SA and IFFT with Various SNR

Although the SNR in the measured data by real radar is large, it is necessary to performance the effect of noises since high Legendre moments are noise sensitivity. Actually in our experimental data above, some HRRP's SNR is found not satisfy the index requirements affected by complex environment. For aerial targets, the noises in the inphase and quadrature echoes of targets can be assumed to be Gaussian white noises. Then we add this kind additional noise to our measured data to evaluate the classification performance of the proposed algorithm in a noisy environment. Note that the actual noise level contained in the additional data may be somewhat higher than the desired noise level since the measurement noise level was not considered in the process of desired noise addition.

The desired SNR of HRRP with desired noise addition is defined as

$$\mathrm{SNR} = 10 \times \log_{10}\left(\sum_{i=1}^{N} P_x(i)/N \cdot P_A\right) \tag{23.32}$$

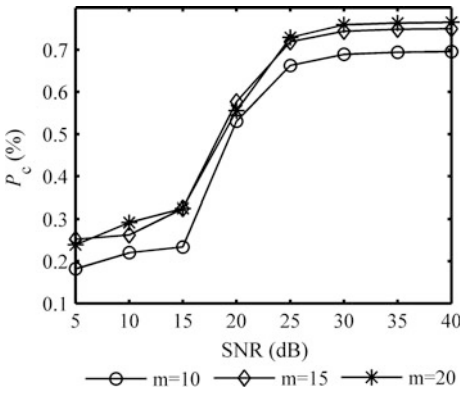


Fig. 23.10 P_c of the Legendre moments with various SNR by IFFT

where $\{P_x(i)\}$ denotes the power of each measured range cell, P_A denotes the power of additive noise, N denotes the number of HRRP range cells. The desired SNR is added from 5 to 40 dB with a 5-dB step in our experiments. We performed twenty independent Gaussian white noises each SNR and obtained twenty's average resultant, which provide more reliable results.

The average recognition rates P_c the Legendre moments with various SNR are shown in Fig. 23.10. With SNR = 5 dB, P_c of different maximum order is less than 30%. P_c increases with the increase of SNR. With SNR = 40 dB, P_c of different maximum order is higher than 70%. Given the same SNR, with the maximum order increase more information offered and the average P_c increase thereby, but don't change much further as with Sect. 1.4.2 shows.

To improve the poor recognition in low SNR, we adopt SA proposed in Sect. 1.2.3. Figure 23.11 shows a measured real HRRP obtained by IFFT and reconstructed by SA. We can see that reconstructed HRRP can effectively mitigate noise affect. The same measured data are used and the experimental results are shown in Table 23.5. The corresponding overall correct recognition rate P_c is 65.73% lower than that used by IFFT directly in Table 23.1 and shows no any

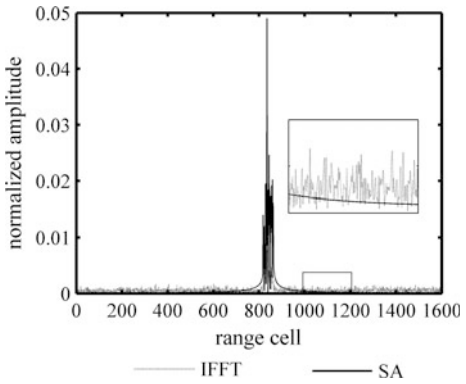


Fig. 23.11 Normalized HRRP obtained by IFFT and reconstructed by SA

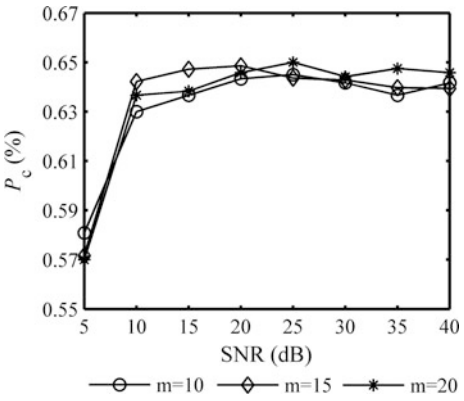


Fig. 23.12 P_c of the Legendre moments with various SNR using SA

Table 23.5 Confusion matrix of Legendre moments descriptors

Actual class	Classification			
	Target-1	Target-2	Target-3	Target-4
Target-1	2276	594	29	173
Target-2	292	1528	866	386
Target-3	30	152	2208	682
Target-4	121	523	363	2065

The selected parameters are: maximum order m = 20, adjustable parameter σ = 1,assume scatters K = 30 (overall correct recognition rate p_c = 65.73%)

particular advantage. Since for assumed K scatterers, SA could extract the strongest scatterers and ignored various weaker scatterers with their information. But when we add additional noise to measured data, weaker scatterers polluted by the serious noise can't provide any useful information contrarily increase noise sensitivity of higher moments. Meanwhile, reconstruct the real HRRP by SA, the average recognition rates P_c versus the SNR are shown in Fig. 23.12. Even with SNR = 5 dB, the average P_c of different maximum order is more than 60% double than that obtained by IFFT. Comparing Figs. 23.11 and 23.12, it can be seen that the SA could extract the strongest scatterers effectively even in low SNR therefore much more robust to noise than IFFT, and has significant advantages in lower SNR environment.

23.5 Conclusions

HRRP carry rich information about structural features of targets. A new target recognition approach based on Legendre moments has been proposed and analyzed. The proposed approach utilizes the average HRRP and Legendre moments feature

in addition to the appropriate preprocessing. The SVM classifier recognizes target using the proposed feature vectors. The experiment results based on the measured data show that the proposed approach has significant advantages than that of Central moments. In view of maximum order and noise sensitivity, it needs consider the real-time and computation time to select the suitable orders. Moreover, HRRP through the conventional IFFT for the measured echo signal data can't obtain much more robust result in lower SNR environment. Therefore we adopted another method via spectrum analysis to reconstruct the real HRRP and make Legendre moments more robust in a noisy environment. Other orthogonal moments extensively used in image analysis will be investigated in the future work.

References

1. Li HJ, Yang SH (1993) Using range profiles as feature vectors to identify aerospace objects. IEEE Trans Antennas Propag 41(3):261–268
2. Zyweck A, Bogner RE (1996) Radar target classification of commercial aircraft. IEEE Trans Aerosp Electron Syst 32(2):598–606
3. Kim KT, Seo DK, Kim HT (2002) Efficient radar target recognition using the MUSIC algorithm and invariant features. IEEE Trans Antennas Propag 50(3):325–337
4. Du L, Liu H, Bao Z, Xing M (2005) Radar HRRP target recognition based on higher order spectra. IEEE Trans Signal Process 53(7):2359–2368
5. Du L, Liu H, Bao Z, Zhang J (2006) A two-distribution compounded statistical model for radar HRRP target recognition. IEEE Trans Signal Process 54(6):2226–2238
6. Ajorloo A, Hadavi M, Bastani MH, Nayebi MM (2014) Radar HRRP modeling using dynamic system for radar target recognition. Radioengineering 23(1):121
7. Shi L, Wang P, Liu H, Xu L, Bao Z (2011) Radar HRRP statistical recognition with local factor analysis by automatic Bayesian Ying-Yang harmony learning. IEEE Trans Signal Process 59(2):610–617
8. Ajorloo A, Hadavi M, Nayebi MM, Bastani MH (2013) Statistical modeling of consecutive range profiles for radar target recognition. In: 2013 14th International Radar Symposium (IRS) IEEE
9. Wang C, Xie J (2013) The T-mixture model approach for radar HRRP target recognition. Int J Comput Electr Eng 5(5):500
10. Liu H, Bao Z. (2004). Radar HRR profiles recognition based on SVM with power-transformed-correlation kernel. In: International Symposium on Neural Networks. Springer, Heidelberg.
11. Li Q, Li B, Yang Z. (2011) Plane HRRP rejection based on SVDD technology. In: Synthetic aperture radar (APSAR), 2011 3rd International Asia-Pacific Conference on. IEEE
12. Fu JS, Yang WL (2012) KFD-based multiclass synthetical discriminant analysis for radar HRRP recognition. J Electromagn Waves and Appl 26(2–3):169–178
13. Hu MK (1962) Visual pattern recognition by moment invariants. IRE Trans Inf Theory 8 (2):179–187
14. Teh CH, Chin RT (1988) On image analysis by the methods of moments. IEEE Trans Pattern Anal Mach Intell 10(4):496–513
15. Kumar TS, Babu KA (2011) Recognition of degraded images by Legendre moment invariants. Digital Image Process 3(15):939–944

16. Hosaini S J, Alirezaee S, Ahmadi M, Makki S V. A. D. (2013). Comparison of the Legendre, Zernike and Pseudo-Zernike moments for feature extraction in iris recognition. In: Computational Intelligence and Communication Networks (CICN), 2013 5th International Conference on. IEEE
17. Singh K, Gupta I, Gupta S (2013) Classification of bamboo species by Fourier and Legendre Moment. Int J Adv Sci Technol 50:61–70
18. Dai X, Zhang H, Liu T, Shu H, Luo L (2014) Legendre moment invariants to blur and affine transformation and their use in image recognition. Pattern Anal Appl 17(2):311–326
19. Schmidt R (1986) Multiple emitter location and signal parameter estimation. IEEE Trans Antennas Propag 34(3):276–280
20. Carriere R, Moses RL (1992) High resolution radar target modeling using a modified Prony estimator. IEEE Trans Antennas Propag 40(1):13–18
21. Li J, Stoica P (1996) Efficient mixed-spectrum estimation with applications to target feature extraction. IEEE Trans Signal Process 44(2):281–295
22. Burges CJ (1998) A tutorial on support vector machines for pattern recognition. Data Min Knowl Disc 2(2):121–167
23. Kittler J, Roli F (2010) Multiple classifier systems. Springer, Heidelberg

Chapter 24
An Improved Adaptive SRCKF Algorithm for Non-cooperative Target Orbit Determination

Guangde Xu, Zhongqiu Gou and Bainan Zhang

24.1 Introduction

Tracking and orbit determination of space objects have two methods, based on Earth and space. Compared with the cooperative tracking mode, the satellite-to-satellite passive tracking system can obtain angles and frequencies by means of optical or radioed measurements. Space based orbit determination using bearings only measurements overcomes the limitation of ground station and is not affected by the weather and national boundary. There has been literature that validated the observability of orbit determination of space objects using bearings only measurements [1, 2]. Research on passive tracking using bearings-only measurements has great significance in space surveillance systems and space situation awareness [3].

The accuracy of orbit determination of non-cooperative targets using bearings only measurement mainly relies on the accuracy of the dynamics model, the measurement error and the performance of the filter algorithm. As for the dynamics model, present study generally used the spherical earth orbit dynamics model of J2 perturbation, ignoring the atmospheric drag perturbation, the sun gravitational and light pressure perturbation, which are seen as system error. In 1997, Crassidis and Markley presented a real-time nonlinear filter named nonlinear predictive filter (NPF) [4], which has good ability to cope with model errors and has attracted wide attention. Space based orbit determination is a typical nonlinear estimation problem. The most classical nonlinear estimation method is the extended Kalman filter (EKF), but high running complexity induced by the computation of the Jacobian matrix limit its application [5]. Julier and Uhlmann proposed a new algorithm named unscented Kalman filter (UKF) [6], which does not need to calculate the Jacobian matrix. When the system state dimensionality is relatively high, the

G. Xu (✉) · Z. Gou · B. Zhang
China Academy of Space Technology, Beijing 100094, China
e-mail: xuguangde.2007@163.com

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_24

performance of UKF will obviously decrease. A kind of Monte-Carlo method, Particle filter (PF) [7] greatly improves the estimation precision of the nonlinear filter, but leads to tremendous computational complexity. Recently, the square-root cubature Kalman filter (SCKF) was proposed by Arasaratnam and Haykin [8], which uses the spherical cubature rule and radial rule to optimize the sigma points and weights and clearly improves the estimation precision and stability. In the paper I try to combine the NPF with the SCKF to solve the orbit determination problem of non-cooperative target.

24.2 Problem Formulation

24.2.1 Orbit Dynamics Model

In this study, the satellite was modeled as point mass with a state vector $\boldsymbol{x}$ composed of the satellite's position $\boldsymbol{r}$, velocity $\boldsymbol{v}$ in J2000.0 inertial coordinate. In this study we have considered, taking in account the chaser orbital configuration, an orbital dynamic with only two spherical harmonics [9], namely

$$\dot{\boldsymbol{x}} = \begin{cases} \dot{v}_x \\ \dot{v}_y \\ \dot{v}_z \\ -\frac{\mu x}{r^3}\left[1 + \frac{3}{2}J_2\left(\frac{R_e}{r}\right)^2\left(1 - \frac{5z^2}{r^2}\right)\right] \\ -\frac{\mu y}{r^3}\left[1 + \frac{3}{2}J_2\left(\frac{R_e}{r}\right)^2\left(1 - \frac{5z^2}{r^2}\right)\right] \\ -\frac{\mu z}{r^3}\left[1 + \frac{3}{2}J_2\left(\frac{R_e}{r}\right)^2\left(3 - \frac{5z^2}{r^2}\right)\right] \end{cases} \tag{24.1}$$

where μ is the Keplerian constant, $\boldsymbol{r}$ is the length of the position vector of the target from the center of the Earth, $\boldsymbol{R}_\mathrm{e}$ is the Earth radius, $\boldsymbol{J}_2 = 1.08263 \times 10^{-3}$.

Equation (24.1) can be written as $\dot{\boldsymbol{x}} = F(\boldsymbol{x})$, if the state vector at t_k is $\boldsymbol{x}_\mathrm{k}$, then the state vector at $t_{\mathrm{k}+1}$ is

$$\boldsymbol{x}_{k+1} = \int_{t_k}^{t_{k+1}} F(\boldsymbol{x}_k)\, dt \tag{24.2}$$

The above equation is written as

$$\boldsymbol{x}_{k+1} = f(x_k) \tag{24.3}$$

24.2.2 Measurement Model

The measurement values are azimuth α and β pitching angle between the observer satellite and the target [10]

$$\begin{aligned} \alpha &= \arctan\left(\frac{\rho_y}{\rho_y}\right) + v_\alpha \\ \beta &= \arctan\left(\frac{\rho_z}{\sqrt{\rho_x^2 + \rho_y^2}}\right) + v_\beta \end{aligned} \tag{24.4}$$

where v_α and v_β are modeled as zero-mean white Gaussian noise with variances of σ_α and σ_β, respectively. $\boldsymbol{\rho} = \left[\rho_x\ \rho_y\ \rho_z\right]^{\mathrm{T}}$ is the relative position vector from the observer to the target.

Supposing that the orbit of the observer satellite can be obtained using GPS or other method, named as $\boldsymbol{x}_c$, then the relationship between $\boldsymbol{\rho}$ and state vector $\boldsymbol{x}$ is described as $\boldsymbol{\rho} = \boldsymbol{R}_{oi}(\boldsymbol{x} - \boldsymbol{x}_c)$, where $\boldsymbol{R}_{oi}$ is the transformation matrix from the inertial coordinate to the orbit coordinate of the observer satellite.

Define the measurement vector as $\boldsymbol{z} = [\alpha \quad \beta]^{\mathrm{T}}$, then the measurement equation is described as follows

$$\boldsymbol{z} = h(\boldsymbol{x}) + \boldsymbol{v} \tag{24.5}$$

where $\boldsymbol{v} = \left[v_\alpha\ v_\beta\right]^{\mathrm{T}}$ is the measurement noise, its covariance matrix $\boldsymbol{R}$ is $\begin{bmatrix} \sigma_\alpha^2 & 0 \\ 0 & \sigma_\beta^2 \end{bmatrix}$.

24.3 Filter Algorithm

24.3.1 Nonlinear Predictive Filter

In the nonlinear predictive filter, it is assumed that the state and output estimates are given by a preliminary model and a to-be-determined model error vector, given by [11]

$$\begin{aligned} \dot{\hat{\boldsymbol{x}}}(t) &= f(\hat{\boldsymbol{x}}(t), t) + \boldsymbol{G}(t)\,\boldsymbol{d}(t) \\ \hat{\boldsymbol{z}}(t) &= h(\hat{\boldsymbol{x}}(t), t) + \boldsymbol{v}(t) \end{aligned} \tag{24.6}$$

where $\hat{\boldsymbol{x}}(t) \in \mathrm{R}^n$ is the state estimation vector, $\boldsymbol{d}(t) \in \mathrm{R}^l$ is the model error vector, $\boldsymbol{G}(t) \in \mathrm{R}^n$ is the model-error distribution matrix, $\hat{\boldsymbol{z}}(t) \in \mathrm{R}^m$ is the estimated output vector.

A Taylor series expansion of the output estimate is given by

$$\hat{\boldsymbol{z}}(t+\Delta t) = \hat{\boldsymbol{z}}(t) + \boldsymbol{Z}(\hat{\boldsymbol{x}}(t), \Delta t) + \boldsymbol{\Lambda}(\Delta t)\, \boldsymbol{S}(\hat{\boldsymbol{x}}(t))\, \boldsymbol{d}(t) \tag{24.7}$$

where $\Delta t = t_{k+1} - t_k$ is the sampling interval.

Define $\hat{\boldsymbol{z}}(t)$ as $(\hat{z}_1(t) \cdots \hat{z}_m(t))^{\mathrm{T}}$, then

$$\begin{cases} \dot{\hat{\boldsymbol{z}}}_i(t) = \frac{\partial h_i(\hat{\boldsymbol{x}}(t),t)}{\partial \hat{\boldsymbol{x}}(t)} \dot{\hat{\boldsymbol{x}}}(t) \\ \ddot{\hat{\boldsymbol{z}}}_i(t) = \frac{\partial}{\partial \hat{\boldsymbol{x}}(t)} \left[\frac{\partial h_i(\hat{\boldsymbol{x}}(t),t)}{\partial \hat{\boldsymbol{x}}(t)} \dot{\hat{\boldsymbol{x}}}(t) \right] \dot{\hat{\boldsymbol{x}}}(t) \end{cases} \tag{24.8}$$

where i = 1,2,…m, $\boldsymbol{p}_i$ is the lowest order of the derivative of $h_i(\hat{\boldsymbol{x}}(t), t)$ in which any component of the model error $\boldsymbol{d}(t)$ first appears.

The ith element of $\boldsymbol{Z}(\hat{\boldsymbol{x}}(t), \Delta t)$ is given by

$$\boldsymbol{Z}\,(\hat{\boldsymbol{x}}(t), \Delta t) = \sum_{k=1}^{p_i} \frac{(\Delta t)^k}{k!} \boldsymbol{L}_f^k(h_i), \quad i = 1, 2, \ldots, m \tag{24.9}$$

$\boldsymbol{L}_f^k(h_i)$ is a k^{th} Lie derivative, defined by

$$\begin{cases} \boldsymbol{L}_f^k(h_i) = h_i & k = 0 \\ \boldsymbol{L}_f^k(h_i) = \frac{\partial \boldsymbol{L}_f^{k-1}(h_i)}{\partial \hat{\boldsymbol{x}}} f & k \geq 1 \end{cases} \tag{24.10}$$

$\boldsymbol{\Lambda}(\Delta t) \in \mathrm{R}^{\mathrm{m}\times 1}$ is a diagonal matrix with elements given by

$$\lambda_{ii} = \frac{(\Delta t)^{p_i}}{p_i!}, \quad i = 1, 2, \ldots, m \tag{24.11}$$

$\boldsymbol{S}\,(\hat{\boldsymbol{x}}(t)) \in \mathrm{R}^{\mathrm{m}\times \mathrm{m}}$ is a matrix with each ith row given by

$$s_i = (\boldsymbol{L}_{g1}(\boldsymbol{L}_f^{p_i-1}(h_i)) \cdots \boldsymbol{L}_{gl}(\boldsymbol{L}_f^{p_i-1}(h_i))) \quad i = 1, 2, \ldots, m \tag{24.12}$$

A cost functional consisting of the weighted sum square of the measurement-minus-estimate residuals plus the weighted sum square of the model correction term is minimized, given by

$$\boldsymbol{J}[\boldsymbol{d}(t)] = \frac{1}{2}[\boldsymbol{z}(t+\Delta t) - \hat{\boldsymbol{z}}(t+\Delta t)]^{\mathrm{T}} \boldsymbol{R}^{-1} \cdot [\boldsymbol{z}(t+\Delta t) - \hat{\boldsymbol{z}}(t+\Delta t)] + \frac{1}{2}\boldsymbol{d}^{\mathrm{T}}(t)\, \boldsymbol{W}\, \boldsymbol{d}\,(t) \tag{24.13}$$

Minimizing Eq. (24.15) with respect to $\boldsymbol{d}$(t) leads to the following model error solution

$$\hat{\boldsymbol{d}}(t) = -\left\{[\boldsymbol{\Lambda}(\Delta t)\boldsymbol{S}(\hat{\boldsymbol{x}})]^{\mathrm{T}}\boldsymbol{R}^{-1}[\boldsymbol{\Lambda}(\Delta t)\boldsymbol{S}(\hat{\boldsymbol{x}}(t))] + \boldsymbol{W}\right\}^{-1}\cdot[\boldsymbol{\Lambda}(\Delta t)\boldsymbol{S}(\hat{\boldsymbol{x}})]^{\mathrm{T}} \boldsymbol{R}^{-1}[\boldsymbol{Z}(\hat{\boldsymbol{x}}, \Delta t) + \hat{\boldsymbol{z}}(t) - \hat{\boldsymbol{z}}(t + \Delta t)] \tag{24.14}$$

24.3.2 Square-Root Cubature Kalman filtering

Consider a nonlinear continuous–discrete system with additive process and measurement noises [12]

$$\boldsymbol{x}_k = f(\boldsymbol{x}_{k-1}, \boldsymbol{u}_{k-1}) + \boldsymbol{w}_{k-1} \tag{24.15}$$

$$\boldsymbol{z}_k = h(\boldsymbol{x}_k, \boldsymbol{u}_k) + \boldsymbol{v}_k \tag{24.16}$$

where $\boldsymbol{x}_{\mathrm{k}}$ and $\boldsymbol{z}_{\mathrm{k}}$ are the state vector and measurement vector of the system. $\boldsymbol{w}_{\mathrm{k}}$ and $\boldsymbol{v}_{\mathrm{k}}$ are the process and measurement noise. $f(\cdot)$ and $h(\cdot)$ are the state and measurement function. $\boldsymbol{w}_{\mathrm{k}}$ and $\boldsymbol{v}_{\mathrm{k}}$ are uncorrelated white Gaussian noise with the constant statistical properties, $\boldsymbol{w}_{\mathrm{k}}$ satisfies $\mathrm{N}(0, \boldsymbol{Q}_k)$, $\boldsymbol{v}_{\mathrm{k}}$ satisfies $\mathrm{N}(0, \boldsymbol{R}_k)$.

In the standard CKF algorithm, the error covariance matrix may lose symmetry and positive definiteness, causing unstable or even divergence behavior. To circumvent this problem, square-root cubature Kalman filter (SCKF) is introduced to propagate the square root A of error covariance P, and the square root can also preserve the symmetry and positive definiteness of the covariance matrix for improving numerical stability of CKF. The specific SCKF algorithm process based on the nonlinear system Eqs. (24.15) and (24.16) is given as follows.

(1) ***Initialization***

The initial value $\boldsymbol{S}_{k|k}$ is the square-root factor of the error covariance matrix

$$\boldsymbol{S}_{k|k} = \left[chol(\boldsymbol{P}_{k|k})\right]^{\mathrm{T}} \tag{24.17}$$

(2) ***Time update***

Evaluate the cubature points

$$\boldsymbol{X}_{i,k|k} = \boldsymbol{S}_{k|k}\xi_i + \hat{\boldsymbol{x}}_{k|k} \tag{24.18}$$

Evaluate the propagated cubature points

$$\boldsymbol{X}^{*}_{i,k+1|k} = f_k(\boldsymbol{X}_{i,k|k}),\ i = 1, 2, \cdots m \tag{24.19}$$

where $\boldsymbol{X}_{i,k|k}$, $\boldsymbol{X}^{*}_{i,k+1|k}$ are the cubature points, m = 2n, $\boldsymbol{S}_{k|k}$ is the triangle square-root matrix of the error covariance matrix.
Evaluate the predicted cubature points

$$\hat{x}_{k+1|k} = \frac{1}{m}\sum_{i=1}^{m} \boldsymbol{X}^{*}_{i,k+1|k} \tag{24.20}$$

Calculate the triangular square-root matrix of the predicted error covariance matrix $\boldsymbol{P}_{k|k}$.

$$\boldsymbol{S}_{k+1|k} = Tria\left([\boldsymbol{X}^{*}_{i,k+1|k}, S_{\boldsymbol{Q},k}]\right) \tag{24.21}$$

where $\boldsymbol{Q}_k = \boldsymbol{S}_{\boldsymbol{Q},k}\boldsymbol{S}^{\mathrm{T}}_{\boldsymbol{Q},k}$, $\boldsymbol{S}_{\boldsymbol{Q},k}$ is the square root factor of $\boldsymbol{Q}_k$, and the centered matrix

$$\boldsymbol{X}^{*}_{k+1|k} = \frac{1}{\sqrt{m}}\left[\boldsymbol{X}^{*}_{1,k+1|k} - \hat{\boldsymbol{x}}_{k+1|k}, \boldsymbol{X}^{*}_{2,k+1|k} - \hat{\boldsymbol{x}}_{k+1|k}, \cdots \boldsymbol{X}^{*}_{\mathrm{m},k+1|k} - \hat{\boldsymbol{x}}_{k+1|k}\right] \tag{24.22}$$

(3) ***Measurement update***
Recalculate the cubature points

$$\boldsymbol{X}_{i,k+1|k} = \boldsymbol{S}_{k+1|k}\xi_i + \hat{\boldsymbol{x}}_{k+1|k} \tag{24.23}$$

Evaluate the measured cubature points

$$\boldsymbol{Y}_{i,k+1|k} = h_{k+1|k}(\boldsymbol{X}_{i,k+1|k}) \tag{24.24}$$

Evaluate the predicted measurement vector

$$\hat{\boldsymbol{z}}_{k+1|k} = \frac{1}{m}\sum_{i=1}^{m} \boldsymbol{Y}_{i,k+1|k} \tag{24.25}$$

Estimate the square-root of the innovation covariance matrix

$$\boldsymbol{S}_{zz,k+1|k} = Tria\left([\boldsymbol{Y}_{k+1|k}, \boldsymbol{S}_{\boldsymbol{R},k+1}]\right) \tag{24.26}$$

where $\boldsymbol{S}_{\boldsymbol{R},k+1}$ denotes a square-root factor of $\boldsymbol{R}_{k+1}$ and the weighted, centered matrix

$$\boldsymbol{Y}_{k+1|k} = \frac{1}{\sqrt{m}}[\boldsymbol{Y}_{1,k+1|k} - \hat{\boldsymbol{z}}_{k+1|k}, \boldsymbol{Y}_{2,k+1|k} - \hat{\boldsymbol{z}}_{k+1|k}, \cdots \boldsymbol{Y}_{m,k+1|k} - \hat{\boldsymbol{z}}_{k+1|k}] \tag{24.27}$$

Estimate the innovation covariance matrix

$$\boldsymbol{P}_{zz,k+1|k} = \boldsymbol{S}_{zz,k+1|k}\boldsymbol{S}^{\mathrm{T}}_{zz,k+1|k} \tag{24.28}$$

$$\boldsymbol{P}_{z,k+1|k} = \boldsymbol{P}_{zz,k+1|k} - \boldsymbol{R}_{k+1} \tag{24.29}$$

Estimate the cross-covariance matrix

$$\boldsymbol{P}_{xz,k+1|k} = \boldsymbol{X}_{k+1|k}\boldsymbol{Y}^{\mathrm{T}}_{k+1|k} \tag{24.30}$$

where the weighted, centered matrix

$$\boldsymbol{X}_{k+1|k} = \frac{1}{\sqrt{m}}[\boldsymbol{X}_{1,k+1|k} - \hat{\boldsymbol{x}}_{k+1|k}, \boldsymbol{X}_{2,k+1|k} - \hat{\boldsymbol{x}}_{k+1|k}, \cdots \boldsymbol{X}_{m,k+1|k} - \hat{\boldsymbol{x}}_{k+1|k}] \tag{24.31}$$

Estimate the Kalman gain

$$\boldsymbol{K}_{k+1} = (\boldsymbol{P}_{xz,k+1|k}/\boldsymbol{S}^{\mathrm{T}}_{zz,k+1|k})/\boldsymbol{S}_{zz,k+1|k} \tag{24.32}$$

Estimate the updated state as

$$\hat{\boldsymbol{x}}_{k+1|k+1} = \hat{\boldsymbol{x}}_{k+1|k} + \boldsymbol{K}_{k+1}(\boldsymbol{z}_{k+1} - \frac{1}{m}\sum_{i=1}^{m} h_{k+1|k}(\boldsymbol{X}_{i,k+1|k})) \tag{24.33}$$

Estimate the square root factor of the corresponding error covariance

$$\boldsymbol{S}_{k+1|k+1} = \mathrm{Tria}\left([\boldsymbol{X}_{k+1|k} - \boldsymbol{K}_{k+1}\boldsymbol{Y}_{k+1|k}, \boldsymbol{K}_{k+1}\boldsymbol{S}_{\boldsymbol{R},k+1}]\right) \tag{24.34}$$

24.3.3 Adaptive Square-Root Cubature Kalman Filter

Supposing that the state estimation at time of k is $\hat{\boldsymbol{x}}_{k/k}$, $\boldsymbol{S}_{k|k}$, the measurement value at time of k+1 is $\boldsymbol{z}_{k+1}$. The ASRCKF algorithm process is given as follows

(1) Calculate the predicted output at time of k

$$\hat{\boldsymbol{z}}_{k/k} = h\left(\hat{\boldsymbol{x}}_{k/k}\right) \tag{24.35}$$

(2) Using Eqs. (24.8)–(24.12) to calculate the parameters of predictive filter at time of k $\boldsymbol{Z}(\hat{\boldsymbol{x}}(t), \Delta t)$、$\boldsymbol{\Lambda}(\Delta t)$、$\boldsymbol{S}(\hat{\boldsymbol{x}}_{k/k})$

(3) Using Eq. (24.14) to calculate the model error $\hat{\boldsymbol{d}}_k$

(4) Based on the nonlinear discrete system described by Eq. (24.35), use $\hat{\boldsymbol{d}}_k$ to modify Eq. (24.19)

$$\boldsymbol{X}^{*}_{i,k+1|k} = f_k(\boldsymbol{X}_{i,k|k}) + \boldsymbol{G}_k\boldsymbol{d}_k, i = 1, 2, \cdots m \tag{24.36}$$

Using Eqs. (24.18), (24.32), and Eqs. (24.20)–(24.34) to get the estimation $\hat{\boldsymbol{x}}_{k+1|k+1}$, $\boldsymbol{S}_{k+1|k+1}$.

24.3.4 Simulation and Results

The observer satellite is at the geosynchronous earth orbit (GEO) and the target satellite is at low earth orbit (LEO), the initial orbit elements are:

Satellite	a (km)	e	i (°)	Ω (°)	w (°)	f (°)
Observer	42,000	0.1	120	30	45	0
Target	7171	0.1	30.0	75	60	29.784

Using STK to generate the standard emperies data of the observer satellite and the target satellite. The simulation time is 6000 s, the sampling time is 1 s. The Monte-carlo simulation times are 100. The average root mean square error (RMSE) [2] is used to evaluate the result of orbit determination.

(1) *Case 1*

The measurement errors are set as follow. For the angle measurement, the random error is 5″ and the system error is 5″. For the position measurement of the observer satellite, the random error is 100 m and the system error is 50 m. The initial errors are [300 300 300] km, [0.1 0.1 0.1] km/s, the results of EKF, UKF and ASRCKF are shown in Figs. 24.1, 24.2 and 24.3.

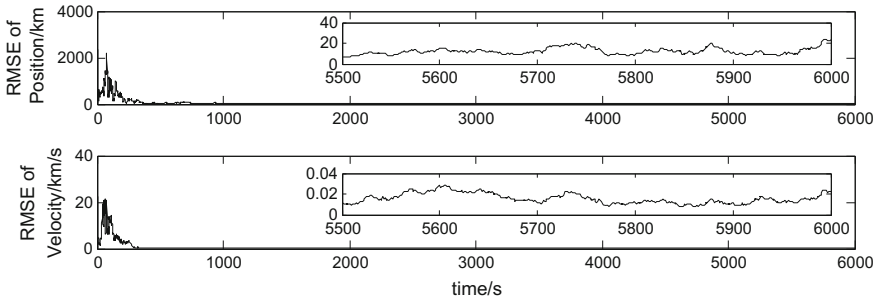


Fig. 24.1 Tracking errors of EKF

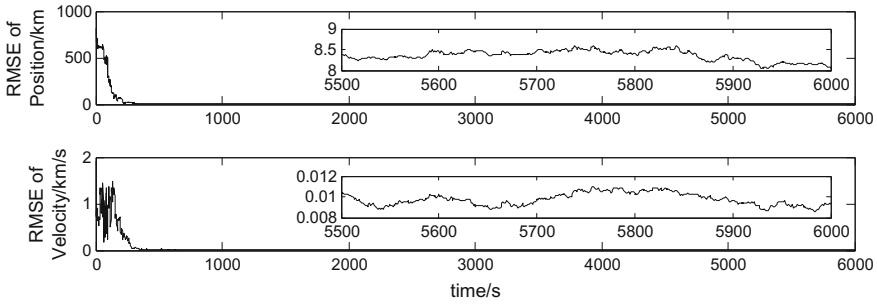


Fig. 24.2 Tracking errors of UKF

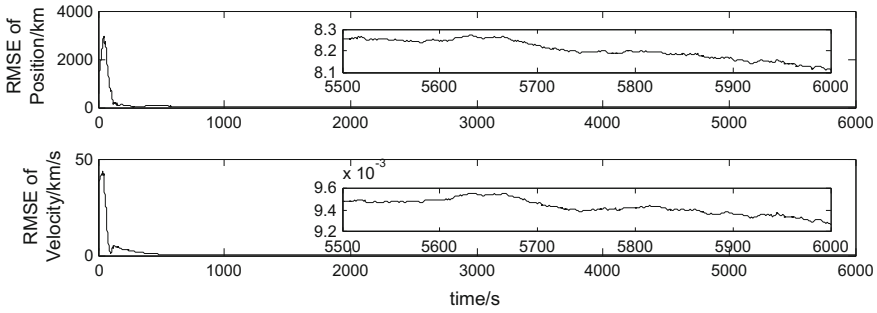


Fig. 24.3 Tracking errors of ASRCKF

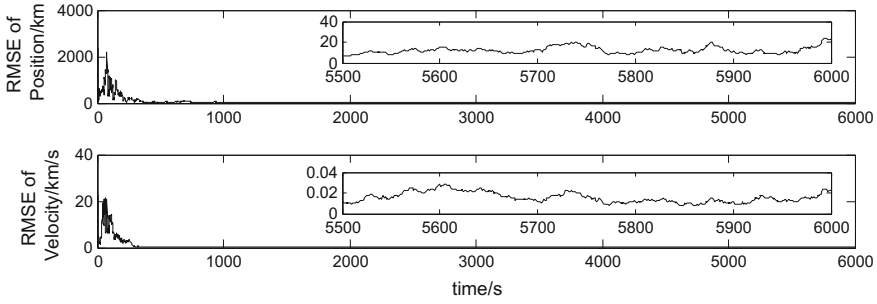


Fig. 24.4 Tracking errors of EKF

It can be seen from the simulation results that the ASRCKF algorithm has faster convergence ability and higher tracking precision than EKF and UKF.

(2) *Case 2*

Compared with case 1, the measurement errors are increased to 50″ and 200 m, the simulation results are shown in Figs. 24.4, 24.5 and 24.6.

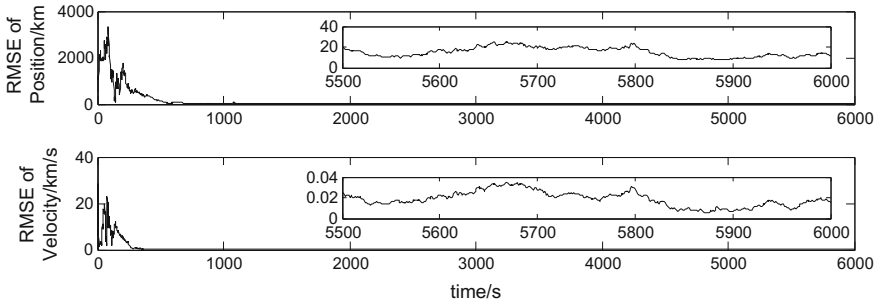


Fig. 24.5 Tracking errors of UKF

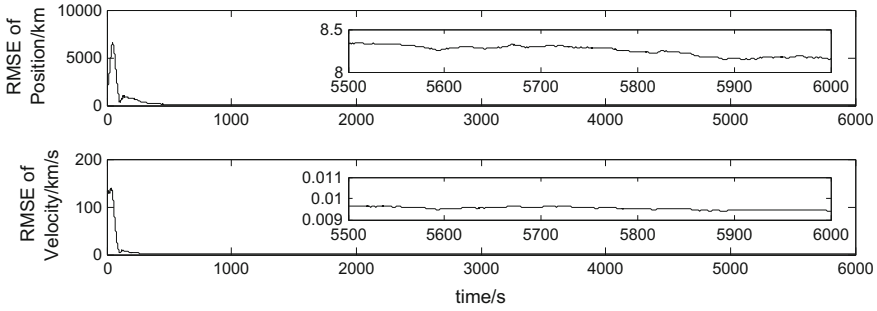


Fig. 24.6 Tracking errors of ASRCKF

It can be seen that the steady estimation error of the EKF algorithm and UKF increase clearly, while the tracking errors of ASRCKF keep at 8.2 km and 10 m/s.

By increasing the initial error and measurement error, the above simulation results fully demonstrated the robustness and stability of the ASRCKF algorithm.

24.4 Conclusions

To solve the problem of inaccuracy of dynamic model in the orbit determination of non-cooperative targets, the paper tries to combine nonlinear predictive filter with square-root cubature kalman filter and propose an improved adaptive filter algorithm, the simulation results show its superiority over traditional filter algorithm.

References

1. Qiang LI, Shi LH, Wang HX, Guo FC (2009) Observability of satellite-to-satellite three-dimensional passive tracking using bearings-only measurements. J Astronaut 30(5): 1957–1966
2. Shun-Hua WU, Qin X, Wan JW (2009) Novel satellite-to-satellite passive orbit determination and tracking method with bearing-only measurements. J Syst Simul 21(12):3693–3701
3. Sharma J, Stokes GH, Von Braun C (2002) Toward operational space-based space surveillance. Lincoln Lab J 13(2):309–334
4. Crassidis JL, Markley FL (2015) Predictive filtering for nonlinear systems. J Guid Control Dynam 20(3):566–572
5. Daum F (2005) Nonlinear filters: beyond the Kalman filter. IEEE Aerosp Electron Syst Mag 20(8):57–69
6. Julier SJ, Uhlmann JK (2004) Unscented filtering and nonlinear estimation. Proc IEEE 92(3): 401–422
7. Speekenbrink M (2016) A tutorial on particle filters. J Math Psychol 73:140–152
8. Arasaratnam I, Haykin S (2009) Cubature kalman filters. IEEE Trans n Autom Control 54(6): 1254–1269
9. Armellin R, Lizia PD, Zanetti R (2016) Dealing with uncertainties in angles-only initial orbit determination. Celest Mech Dynam Astron 125(4):435–450
10. Kaufman E, Lovell TA, Lee T (2016) Nonlinear observability for relative orbit determination with angles-only measurements. J Astronaut Sci 63:1–21
11. Zhi-Jun LI, Hou LQ (2014) An improved NPF-SRCKF based algorithm for spacecraft orbit determination. J Astronaut 35(7):811–817
12. Zhang L, Yang H, Lu H, Zhang S, Cai H, Qian S (2015) Cubature kalman filtering for relative spacecraft attitude and position estimation. Acta Astronaut 105(1):254–264

Chapter 25
Influence Analysis of the High-Energy Electrons on Geosynchronous Orbit Satellite

Zhenghe Wang, Baosheng Sun and Shengpeng Liu

25.1 Introduction

Geosynchronous orbit is located above the earth's equator about 36,000 km (from the center of the earth is about 6.6 Re), where hundreds of all kinds of applications satellites, which include the communications, meteorology, navigation and other professional satellites are concentrated. With the wide application in the field of satellite, the number of satellite continues to increase. The application of new spaceborne high performance microelectronic devices increases the risk of fault caused by space environment with complexity and changes, which is a serious threat to the safety of the satellite in the orbit [1]. The statistical results from USA's National Oceanic and Atmospheric Administration (NOAA) about 200 abnormal events on GOES satellites, shows that what leads to these abnormal events are Single Event Upset (SEU) and electrostatic discharge caused by abnormal environment [2–5].

Although the fault on orbit has been effectively corrected, which hasn't influence the service life of the satellite, it is of great significance to analyse the correlation between the fault and the space environment, which can improve satellite design level and the reliability of satellite mission, reduce the risk of long-term management as well as personnel pressure and so on, also is an important part of space situation awareness.

Z. Wang (✉) · B. Sun · S. Liu
Beijing Space Information Relay and Transmission Technology Research Center, Beijing 100094, China
e-mail: 13671039594@163.com

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_25

25.2 Mechanism

In the solar system, the sun sent out high-speed charged particles, mainly including protons and electrons, which are captured by the earth's magnetic field, and format charged particle aggregation region in a high altitude area from the surface of the earth. At the beginning of the last century, Norway space physicist F.C.M. Stormer observed and analyzed the high latitude auroral, and proved the theory that there is a charged particle capture region around the earth (most of the region is found in the earth's radiation zone).

At the end of 1950s, American scientist Van Allen analyzed deeply the data from the Explorer satellites, verified the F.C.M. Stormer theory with the measured data, called the charged particles aggregation region, which formed by these particles captured by the earth's magnetic field, the earth's radiation belt and ascertained it's structure and scope. The earth's radiation belt, whose outer edge gets a range of about 10 earth radii from the center of the earth, has ringed distribution and crescent cross section.

The earth's radiation belt contains inner and outer radiation belts. The inner radiation belt, whose core height is about 1–2 Re, limited to about 40° magnetic latitude, has good stability. The outer radiation belt, whose core height is about 3–4 Re within the magnetic latitude about 60°, has weak stability, and can expand more than 100 times under some special circumstances, which means that includes all kinds of GEO communication satellites and scientific satellites which are exposed to harmful radiation.

Constrained by solar activity, long-term geomagnetism variation, geomagnetic storms and solar cosmic events, the changes in spatial distribution and intensity of earth's radiation belts are complex.

Ideally, with stable solar activity, without geomagnetic storms and other disturbances, the earth's radiation belts remained stable. Actually, the sun is always active, the solar flares, sunspots, geomagnetic storms and other disturbances are frequently active and drastically changed, that is difficult for the earth's radiation belts to keep stable. For the convenience of the study, the grade of activity intensity of the solar flares, sunspots and geomagnetic storms is prioritized. The earth's radiation belts is considered approximately stable in the lower level and the shorter interval. The GEO satellite is in the outer radiation belt, which is a relatively harsh area in space, especially the high energy electrons and high energy protons in it. The high energy electrons with high energy of 2–10 MeV, can penetrate the GEO satellite shell, deposited in the electronic devices or circuits of GEO satellite. Local strong electric field is formed due to high energy electrons accumulation over time, which causes deep dielectric charging inside the satellite. When this charging exceeds certain threshold, there will be internal charging and discharging, resulting in signal voltage level flip, which causes satellite fault and abnormality [6–8]. Therefore, in the study of the space environment of the GEO satellite, it is mainly concerned about the high energy electrons flux anomaly.

25.3 Space Environment and Satellite Fault Analysis

25.3.1 Space Environment Situation

Statistical results show that the fault of the JPL's GOES on orbit is caused by the abnormal space environment with great probability. To explain the complexity of the space environment near the GEO, this paper refers to space environment data in 2015 from SEPC and NSMC [9, 10], focuses on three space environment events, which include geomagnetic storms, high energy electron flux anomalies and energetic proton flux anomalies, and systematically analyzed. By analyzing and refining, the space environment anomaly in 2015 can be seen in Table 25.1.

Tab 25.1 GEO satellite anomaly in 2015

Month	Geomagnetic storm	High energy electron flux anomaly	Energetic proton flux anomaly
1	4, moderate 7, strong	2, exceed threshold	Not reach hazardous level
2	18–24, moderate	4–5, exceed threshold	Not reach hazardous level
3	17, extreme 18–20, moderate 22, moderate	19–28, exceed threshold	Not reach hazardous level
4	11, moderate 15–17, moderate	Not exceed threshold	Not reach hazardous level
5	11, moderate 13–14, moderate	14–18, exceed threshold	Not reach hazardous level
6	8–9, moderate 22–23, extreme 25, moderate	10–13, exceed threshold 24, exceed threshold 26–30, exceed threshold	18, reach hazardous level 22–24, exceed hazardous level 26–27, reach hazardous level
7	5, moderate 13, moderate 23, moderate	1–3, near threshold 15–19, near threshold	Not reach hazardous level
8	15–17, strong 23, moderate 26–29, moderate	8, not exceed (exceed threshold in short time) 11, exceed threshold 17–18, exceed threshold 20–22, exceed threshold 31, exceed threshold	Not reach hazardous level
9	8–10, moderate 11, strong 12, moderate 20, moderate	1, exceed threshold 12–17, exceed threshold	Not reach hazardous level
10	7, moderate 8, strong	8–15, exceed threshold 17, exceed threshold	29, reach hazardous level
11	7, moderate 10, moderate	4–17, exceed threshold	Not reach hazardous level
12	20–21, moderate 31, moderate	7–8, exceed threshold 10–14, exceed threshold 23–25, exceed threshold 27, exceed threshold 30, exceed threshold	Not reach hazardous level

There is some information concluded from Table 25.1:

(1) Moderate geomagnetic storm occurred throughout the year (inclusive) more than 29 times, and it occurred in every month. It occurred 4 times in September.
(2) Daily integration of high energy electron flux achieved or exceeded threshold more than 22 times. it occurred in each month except in April, and 5 times in December at most.
(3) The number of high energy proton flux to the solar proton (burst) event level is 4 times in the whole year, and occurred 3 times in June.

It can be seen that the space environment of the GEO is very bad, whose influence on the GEO satellite is objective and can not be ignored, which needs high attention.

25.3.2 Analysis

In 2015, one GEO satellite occurs many faults on orbit. Analyzing the time of the satellite fault and the space environment variation, it can be found that:

(1) Relationship between satellite fault and geomagnetic storm
Among more than 10 times of satellite fault in 2015, the occurrence of geomagnetic storm and the faults overlaps 2 times which is about 14% of the whole faults, and other 86% doesn't overlap. There is no time relationship between geomagnetic storm and satellite faults. It suggests that satellite faults has no necessary connection with geomagnetic storm.
(2) Relationship between satellite fault and high energy electron flux anomaly
Among more than 10 satellite faults in 2015, there are 11 time overlapping with the high energy electron flux anomaly, the ratio of total faults is about 79%, and the ratio of no overlapping is about 21%. This shows that the high energy electron flux near GEO is higher than that of the warning threshold value, the probability of the satellite fault is larger, and the correlation is about 80%.
(3) Relationship between satellite fault and high energy proton flux anomaly
Among more than 10 satellite faults in 2015, only one in June overlaps in time with high energy proton flux, but there is also high energy electron flux anomaly. All the other satellite faults and high-energy proton flux anomaly do not overlap in time, which means there is no obvious connection between them.

25.3.3 Summary

(1) the relation between high energy electron flux anomaly and satellite fault is strong, and the time overlapping is about 80%;
(2) the relation between geomagnetic storms and satellite faults is weak, which means that the guidance, that geomagnetic storm forecasts satellite fault, is weak.
(3) the high energy proton flux anomaly and satellite fault are independent with weak connection.

25.4 Research on High Energy Electron Forecasting Technology

Due to the complexity of the space environment, the high energy electrons which cause the fault of GEO satellite is the objective existence. To avoid the occurrence of similar satellite fault, the fundamental strategy is to improve the design level of satellite, and optimize satellite systems and stand-alone products, such as the circuit resistance–capacitance filtration, cable shielding, which can improve the anti-interference performance of satellite. When the satellite is operating on orbit, in order to reduce the risk of long-term management personnel pressure, improve the reliability of space mission and enhance the ability of space situational awareness, more effective strategy is to strengthen the forecasting of energetic electron flux anomaly.

25.4.1 Theoretical Analysis

The GOES series satellite of United States is located in the West Zone, including GOES10, 11, 12, which are located in the West zone four, west zone nine and West Zone five. Chinese FY-2D GEO satellite is located in the eastern time zone six. The local time difference between FY-2D satellite and GOES10, 11 and 12 is 10, 15 and 11 h. When the FY-2D satellite local time is 20 h (evening), GOES10, 11 and 12 local time is 10 h (morning), 5 h (morning), 9 h (morning) respectively. Wang Xinyue [11] compares and analyses the observations of FY-2D satellite and GOES satellite with space particle, which comes to a conclusion that different satellite observation data of the same area of high-energy electrons has good correlation. High energy electron data from FY-2D satellite in local time of 20 h and high energy electron data from GOES10 satellite in 10 h is relevant, which is to say that the high energy electron data from FY-2D satellite can be used as a forecast for GOES10 satellite after 10 h. Also, the high energy electron data measured by the

GOES10 satellite can be used as the prediction value for satellite FY-2D after 14 h. When the electron data can be real-time access to multiple GEO satellites, according to the above characteristics, using multi-source data redundancy information, establishes and improves the prediction model of GEO electron of 24 h or longer, and releases the dynamic forecast information in roll.

25.4.2 Research on Conventional Forecasting

At present, the NSMC can realize real-time posting of GEO energetic electron flux information through the network every 30 min, broadcast a UTC time of the energetic electron flux values, by using space environment data acquisition at home and abroad. It can be seen in Fig. 25.1.

Figure 25.1 is a screenshot from NSMC website, the lower right corner of the red box shows the contents of the GEO orbital energy electron flux of the current 5 min value and daily integration value. For the user, you can browse the web page regularly, record the UTC time corresponding to the high energy electron flux value, and concern whether the high energy electron flux value exceeds the warning threshold value. When the energetic electron flux exceeds the warning threshold, and there is no other geomagnetic disturbances occurrence (the earth's radiation

Fig. 25.1 Screenshot from SEPC website

belts can keep stable within a short period of time) the user can conduct the time conversion between UTC and local time based on fixed point position of the GEO satellite, also can pre-warn fault on GEO satellite on orbit, or make fault prevention of GEO satellite, and enhance the ability to space environment situation awareness.

25.4.3 Research on Special Forecasting

For special users, the need is to know the space environment situation of the GEO satellite in real time, offline browsing can not meet the requirements.

In order to solve this problem, this paper puts forward the following ideas: setup a data link which connects the special user and forecast center, to achieve real-time sharing of space environment data. Special users use multi-source data redundancy information real-time acquisition, establish and improve the automatic forecasting and warning mechanism in GEO space environment for 24 h or longer, as an important implementation of long-term management of satellite.

This procedure is: the space environment satellite measured high energy electron data, data downloaded to the ground station via wireless channel in real time, then transmitted to the satellite ground station through the ground center, the ground center transmits the data which is received from the satellite ground station and other domestic and international space environment data after comprehensive treatment to the special user in real time [12]. Its flow chart is shown in Fig. 25.2.

The specific processes are as follows:

(1) when developing the GEO satellite, it is appropriate to consider particle detector as payload on the GEO satellite;
(2) the space environment data acquired by each satellite is transmitted in real time through the satellite-ground wireless channel;

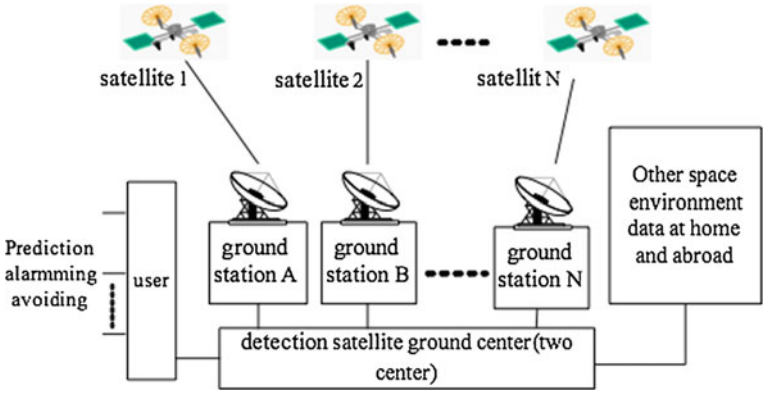


Fig. 25.2 Chart of GEO environment data prediction

(3) the ground station receives satellite data, adds the reception timing and other information, stores and forwards received real-time data to the satellite ground receiving station at the center; the ground calibration and satellite measured environment data of the time difference need to provide as a prior calibration;
(4) detection satellite ground station receives data from the ground center, while receives other domestic and foreign space environment data. Through comprehensive data processing, the data will contain the space environment data of high energy electrons and forwards to special users in real time;
(5) establish real-time routing between special users and the ground station of the detection satellite;
(6) special users receive data in real time, use the prediction model to deal with the core calculation, forecast the information about the GEO satellite interfered by high energy electron, and establish automatic forecast and warning mechanism.

25.5 Conclusion

Based on the space environment data and fault data of one GEO satellite, this paper analyses the reasons of the satellite fault. Based on the mechanism analysis, this paper analyzes the influence of high energy electron on the GEO satellite. In order to enhance the accuracy of fault prediction and ability of space situational awareness, reduce risk of the long-term management of satellite personnel pressure and improve space mission reliability, the need to strengthen the work for high energy electronic space environment real-time forecast, proposes detection satellites, ground stations, satellite ground center and related special users the establishment of real time routing between them, realize real-time sharing of space environmental information including high energy electron data and long-term data accumulation, establish and perfect the system of automatic forecasting warning satellites in GEO space environment for 24 h or longer.

References

1. Bedingfield KL, Leach RD, Alexander MB (1996) Spacecraft system failures and anomalies attributed to the natural space environment. National Aeronautics and Space Administration, Marshall Space Llight Center, Huntville
2. Y-x Xue, S-s Yang, Ba D (2012) Analysis of spacecraft system 1 failures and anomalies attributed to the natural space radiation environment. Vac Cryog 18(2):63–70 (in Chinese)
3. Paniccia M (2014) Cluster Space Weather Anomalies [OL][2014-12-15]. http://lasp.colorado.edu/media/education/reu/2007/docs/slides/paniccia-slides.pdf
4. Love DP, Toomb DS, Wilkinson DC et al (2000) Penetrating electron fluctuations associated with GEO spacecraft anomalies. Plasma Sci 28(6):2075–2084
5. Harboe-Sorensen R (2013) 40 years of radiation single event effects at the European Space Agency, ESTEC. IEEE Trans Nucl Sci 60(3):1816–1823

6. Tian T, Yaoping WU, Zheng C (2015) Analysis of the Chinese GEO satellite anomaly on 9 March 2012. Chin J Space Sci 35(6):687 (in Chinese)
7. Jianguo H, Jianwei H (2010) Analysis of a typical internal charging induced spacecraft anomaly. Acta Phys Sin 59(4):2907–2913 (in Chinese)
8. Ce G, Bingsen X, Zhaoxiang L (2013) Approach for predicting the energetic electron flux in geosynchronous earth orbit. Chin J Space Sci 33(4):418–426
9. www.space-weather.gov.cn
10. http://blog.sepc.ac.cn
11. Xinyue W, Chunqin W (2008) Comparison first results of space particle detector of FY-2D with GOES. Chin J Geophys 51(3):611–617 (in Chinese)
12. Nanyin Xia (2002) Spaceflight TT&C system. National Defence Industry Press, Beijing

Chapter 26
Image Fusion Method Based on Sparse and Redundant Representation

Jianglin Shi, Changhai Liu, Rong Xu and Tao Men

26.1 Introduction

As the constant development and application of sensor technology, image fusion has now become an essential research branch in image processing. The multi-frame image, which contains complex information, of the same target or scene can be obtained by one or more sensors. Image fusion is a process of examining the characteristic feature in the original image and compound the detailed information into one frame image. By the fusion of images, it benefits humans and machine perception a lot that providing information about target and scene which is more understandable. The main application field includes digital image, medical image, remotely-sensed image and machine visual [1–5].

According to the difference of image fusion layer, it can be classified as: pixel level fusion, characteristic level fusion and decisive fusion. Pixel fusion concentrates mainly to improve the image quality in order to the convenient the continued image operation. Characteristic level image fusion is targeted at accurately obtaining the objective features from the original image, which approve certain evidence for the classification and examination of objective. The decisive one is to improve the accuracy of classification of decision. Due to the variation in different described methods, the image fusion can be classified roughly as: space based and transform based image fusion. Image fusion based on spatial domain is directly fused pixel intensity of source image into intensity value [6–9]. Image fusion based

J. Shi (✉) · C. Liu · R. Xu · T. Men
State Key Laboratory of Astronautic Dynamics, Xi'an 710043, China
e-mail: shijianglin89@163.com

J. Shi · C. Liu · R. Xu · T. Men
Xi'an Satellite Control Center, Xi'an 710043, China

J. Shi
School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_26

on transform domain firstly transforms the source image into a certain frequency or time domain, and then uses the fusion rules to transform the image.

Assuming $F(\cdot)$ represents "fusion operation", then space range fusion methods can be summarized as follows:

$$I_F = F(I_1, I_2, \ldots, I_K). \tag{26.1}$$

The simplest fusion methods based on space range is to calculate the average of the original image pixel, while this methods always has several side effect like causing the decrease of contrast. If the original is not exactly suitable, the fake images may appear using methods based on single pixel like the spacial gradient.

A much more prevalent method in researching image fusion in recent years is multi scale transform methods. Commonly used multi scale transform methods includes Laplace Pyramid fusion method (LAP), Discrete Wavelet transform fusion method (DWT), Curvelet transform fusion method (CVT) and Non-subsampled Contourlet transform fusion method(NSCT) [10–13]. The methods based on transform domain can be summarized as follows:

$$I_F = T^{-1}(F(T(I_1), T(I_1), \ldots, T(I_k))) \tag{26.2}$$

Where $T(\cdot)$ is multi scale transform, $F(\cdot)$ is fusion rules.

In the field of image fusion, pyramid decomposition is the first multi-scale transformation method, at the start it decompose the image into a series of pyramid images in different resolution, and then fuse in different pyramid layers using various fusion rules. At last, the fusion image can be received after the inverse transformation of fused pyramid layers. The fusion methods based on wavelet transform take similar schemes, while the multiple resolution decomposition method is limited, because most of their collection procedure is not linear shift invariance. However, discrete wavelet transform based on linear shift invariant performance producing over-complete signals shows that, it's suitable for image fusion [14–16]. In recent years, many advanced geometric multi-scale transforms like curvelet, ridgelet and contourlet, etc. begin to be researched, and have certain improvement in performance [17–19]. However, since the fusion image obtained based on the transform domain method is generated globally, a single coefficient change on the transform domain will result in a change in all pixel values in the spatial domain. Thus, several unsatisfactory artifacts may be generated in the fusion process.

Obviously, an efficient and complete extraction of the implicit information from the original image will make the fused image more accurate. Unlike the transform domain approach, the sparse representation uses the overcomplete (redundant) dictionary containing the signal atomic prototypes to describe the signal by sparsely linear combinations of atoms [20–24]. The two major characteristics of sparse representations is redundant and sparse. Over-completeness means the atom number in the dictionary is greater than the signal dimensionality, and sparsity means the index in counter signal is sparse, that is, less non-zero element. Benefit from

these two things, sparse representation has been applied successfully in many occasions, including: compressing, denoising, feature extraction, and classification, etc. This article applies the sparse and redundant representation in image fusion field, and the fusion quality depends on the accuracy of the local features representation, therefore in the article, "sliding window" technology is adopted to acquire local features, and at the same time shift-invariance is maintained.

The methods used in this article is the partition of the original image firstly, then separately express the image block using over-completeness dictionary, and the larger counter index means including more characteristic information. As a result, in the third step, adopting the fusion rule, choosing the bigger norm, to the coefficient of the fusion to obtain the fusion coefficient. At last sparse reconstruct combined with over-completeness to obtain the fusion image. In sparse representation, redundant dictionary is established through two ways: the first one is to analysis dictionary, whose theory model is the primary function of all kinds of transformation: DCT, Gabor, and ridgelet transformation etc. The second one is the learning dictionary, which is obtained from the image samples through iterative training, the common methods includes MOD and K-SVD, through comparative analysis, this article prefer the K-SVD method to establish the dictionary.

This article is organized by the following ways: the second part is to briefly introduce the basic theory of sparse representation. The third part introduces the image fusion algorithm based on sparse and redundant representation. The experiment result and discussion is presented in the fourth part and the fifth part illustrate the summary of the algorithm and the further development.

26.2 Sparse and Redundant Representation Theory

The sparse representation theory is based on the assumption that a signal can be represented or approximated as a linear combination of "minority" atoms in the dictionary. For a given signal $\Gamma \subset R^n$, sparse representation theory suggests the existence of a dictionary $D \in R^{n \times T}$, which contains T prototype signals that are referred to atoms. For any signal $x \in \Gamma$, there exists a linear combination of atoms from D which approximates it well. That is,

$$x \approx Ds \tag{26.3}$$

Where $s \in R^T$ is sparse coefficients. It usually assumes that $T > n$, indicating that the dictionary D is redundant. Looking for the smallest possible number of nonzero components of s contains solving the following optimization problem:

$$\min_s \|s\|_0 \text{ subject to } \|Ds - x\|_2^2 \leq \varepsilon \tag{26.4}$$

Where $\|s\|_0$ denotes the l_0 normal of coefficients s, that is the number of nonzero components in s. $\varepsilon \geq 0$ denotes tolerance error. Formula (26.4) is called sparse representation model. (Fig. 26.1)

Sparse representation model solving is a critical matter in sparse representation, however, the optimization of norm l_0 is a combination question, it's difficult to solve it directly. Therefore, researchers has come up with a large amount of the sparse representation approximation algorithm, the commonest one is the MP based on greedy policy. This algorithm firstly calculate residual approximation and the inner product of the atom to determine the correlation, then according to the correlation, select the most correlated residual approximation atom in the dictionary each time, and approximate the signal by selected atom, finally calculate the approximation residual and perform the next iteration process, until it satisfies the terminal condition. Although the algorithm performance of MP is limited, it's the initial prototype in greedy algorithm research development. OMP is a most representative advanced MP algorithm, which introduces the least square problem to solve the approximation signal in iteration process, the specific process of algorithm is shown as follows. In OMP algorithm, both the selection of atom number and approximation residual can be terms of the termination of algorithm (Fig. 26.2).

Algorithm 1: OMP
Input: Given initial sparse coefficient $s^0 = 0$, Residual $r^0 = x$, set $\Gamma^0 = \emptyset$, counter $k = 1$.
Step 2: Select most relevant atoms to the residuals and update the set $i^k = \arg\max_i \lvert\langle r^k, d_i\rangle\rvert, \Gamma^k = \Gamma^{k-1} \cup \{i^k\}$
Step 3: Calculate the approximation sparse coefficient $s^k = \arg\min_s \|x - D_{\Gamma^k} s\|_2^2$
Step 4: Update residual $r^k = x - Ds^k$
Step 5: Update counter $k = k + 1$, repeat steps 2–4 until the termination condition is satisfied
Output: Sparse coefficient s^k

Dictionary construction is another core problem in sparse representation theory. Dictionaries are not only related to the validity of images, but also affect the performance of sparse representation algorithms. Researchers have proposed a

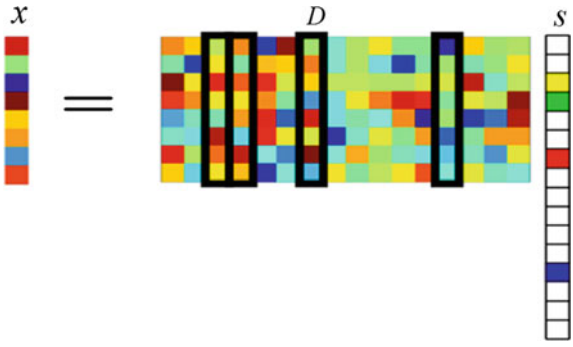


Fig. 26.1 Sparse representation of signal

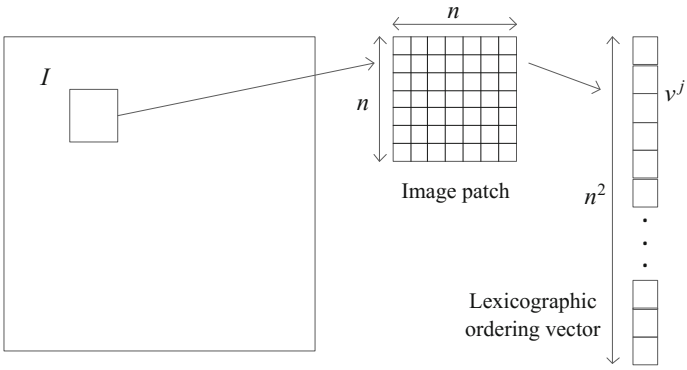


Fig. 26.2 The original image is decomposed into image blocks and arranged in a vector column

number of dictionary construction methods, which can be divided into two categories: (1) based on mathematical model construction (DCT dictionary). (2) based on sample learning method construction (MOD and K-SVD dictionary). This paper will compare the performance of the three dictionaries, and choose K-SVD dictionary structure for sparse representation of the image.

26.3 Image Fusion Method

26.3.1 Sparse Representation of Image Fusion

Because sparse representation theory is the global processing of the whole image, and image fusion depends on the local information of the source image, so sparse representation can't be directly used for image fusion. In this paper, a sliding window technique is adopted to divide the image with size $M \times N$ into image blocks with size $n \times n$ by one pixel from the upper left corner to the lower right corner, and these small image blocks are arranged in a column direction. So as to obtain the column vector matrix V_k composed of image blocks, it can be expressed as

$$V_k = DS_k = [d_1, d_2, \ldots, d_T] \begin{bmatrix} s_{k_1}(1) & s_{k_2}(1) & \cdots & s_{k_J}(1) \\ s_{k_1}(2) & s_{k_2}(2) & \cdots & s_{k_J}(2) \\ \vdots & \vdots & \ddots & \vdots \\ s_{k_1}(T) & s_{k_2}(T) & \cdots & s_{k_J}(T) \end{bmatrix} \tag{26.5}$$

Where $k = 1, 2, \ldots, K$ is the number of frames of the original images. The dimension of the dictionary D is $(n * n) \times T$. The dimension of the sparse coefficient matrix S_k is $T \times J$. The dimension of the column vector of the image blocks V_K is $(n * n) \times J$, and the number of the image block is $J = (M - n + 1) \times (N - n + 1)$.

26.3.2 Fusion Strategy

The flow of the fusion algorithm is shown in Fig. 26.3, The OMP algorithm is used to solve the sparse model, and the sparse coefficient matrix S_k is obtained, The OMP algorithm uses the approximation residual as the iteration termination condition. The 1-norm of the k-th source image is

$$A_{k_j} = \left\| s_{k_j} \right\|_1 \tag{26.6}$$

According to the rule of fusion with a large number of 1-norm to generate a matrix of the fusion coefficients S_F, and the j-th columns are obtained as follows

$$s_{F_j} = s_{k_j^*}, \qquad k_j^* = \arg\max_{k_j} \left(A_{k_j} \right) \tag{26.7}$$

Then the vector matrix of the fused image is

$$V_F = DS_F \tag{26.8}$$

Finally, the fusion image is reconstructed according to the inverse sliding window technique shown in Fig. 26.2. The algorithm steps are shown in the following table.

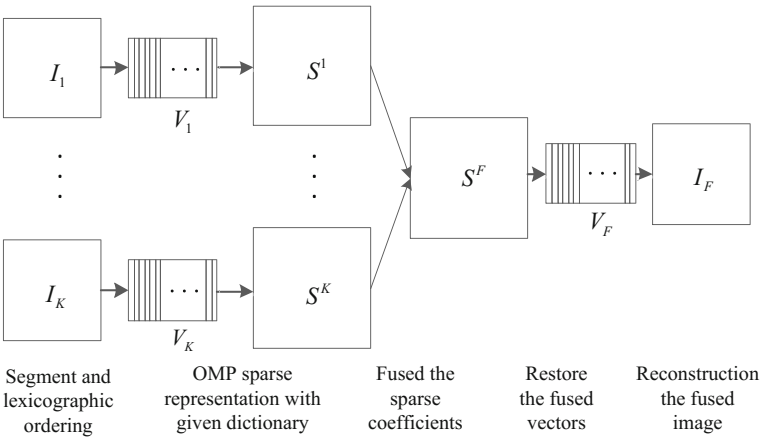


Fig. 26.3 Flow chart of image fusion based on sparse representation

Algorithm 2: Image Fusion Algorithm Based on Sparse Representation

Input: Source images $I_k(k = 1, \ldots, K)$, redundant dictionary D;
Step 2: Divide I_k into image blocks with size of $n * n$ by sliding window technique, and arranged it in a column vector $v_{k_j}(j = 1, 2, \ldots, J)$, all column vectors corresponding to I_k make up the matrix V_K;
Step 3: Decompose V_K sparsely in D to obtain sparse coefficient matrix S_k ,then $V_k = DS_k$;
Step 4: Suppose the elements of the j-th column in S_k is s_{k_j}, and the characteristic index of S_k is calculated by l_1 norm, select the sparse coefficients by maximization selection rule, that is $s_{F_j} = s_{k_j^*}$, $k_j^* = \arg\max_{k_j}(A_{k_j})$;
Step 5: All of the s_{F_j} formed the sparse coefficient matrix S_F of fused image, then fused image matrix $V_F = DS_F$;
Output: And the output fused image I_F is obtained from V_F;

26.4 Experimental Result

26.4.1 Dictionary Training

Redundancy dictionary is the basis of sparse representation method, its advantages and disadvantages directly determine the sparseness of the coefficient matrix. Sparse representation of the image fusion method is based on the sparsity of the coefficient matrix, so the performance of the redundant dictionary directly affect the performance of image fusion. In this section, we compare the performance of DCT, K-SVD and MOD dictionary respectively, and draw some meaningful conclusions.

In general, a dictionary is well-performed if it is complete and redundant and can sparse reconstruct any signal belonging to its defined space. We define the reconstruction error $E = \|x - DA\|_2^2$ of the signal x, and assume that signal has been accurately reconstructed when $E \leq \varepsilon$. Then, we observe the non-zero element number of coefficient vector A to measure the performance of the dictionary D. Where, ε is a threshold of the reconstruction error and assume $\varepsilon = 21$ for 8-bit grayscale image block with size of 8×8 (Figs. 26.4, 26.5).

Figure 26.6 shows the learning dictionary MOD and K-SVD training based on OCNR5 image. It can be seen that the MOD dictionary atom is similar to the K-SVD dictionary atom, which indicates that the performance of the two methods is close.

Figure 26.7 shows the number of non-zero elements in the coefficient vector for the OCNR5 image blocks when using three kinds of dictionaries. It can be seen that the number of non-zero elements in the corresponding coefficient vector is generally more than 15.65 when DCT dictionary is adopted, and the number of non-zero elements in the corresponding coefficient vector when using MOD dictionary and K-SVD dictionary relatively less, the former average of about 11.86, the latter an average of about 11.99. This shows that the performance of the training dictionary

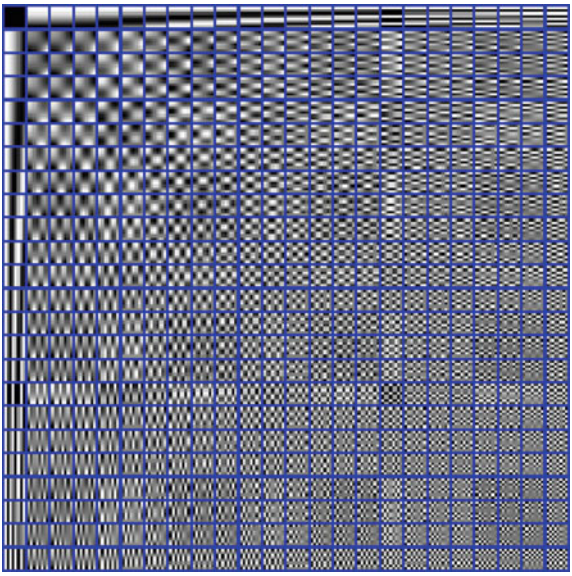


Fig. 26.4 DCT dictionary atomic

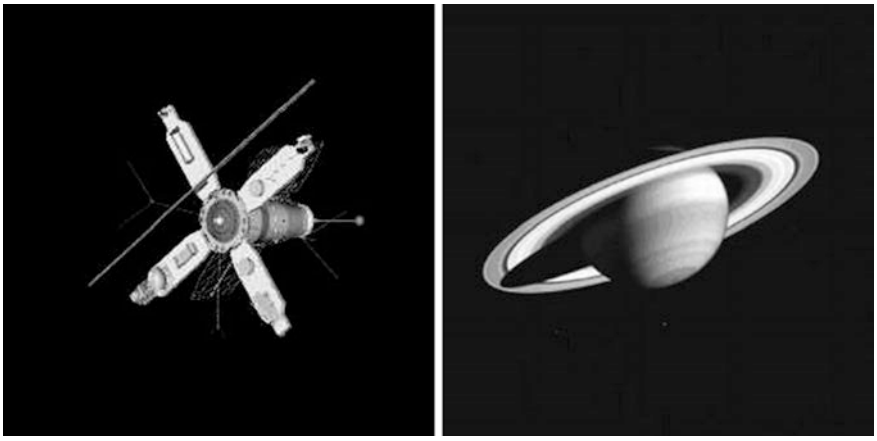


Fig. 26.5 Image data *left* is OCNR5 and *right* is Saturn

is higher than the analytical dictionary for the actual image and the performance of MOD dictionary and K-SVD dictionary is very close. (Fig. 26.8)

We construct a redundant dictionary corresponding to the Saturn image in a similar way as before and then perform a sparse representation of the Saturn image block on the corresponding dictionary. Figure 26.9 shows the number of non-zero elements in the coefficient vector for the Saturn image blocks when using three

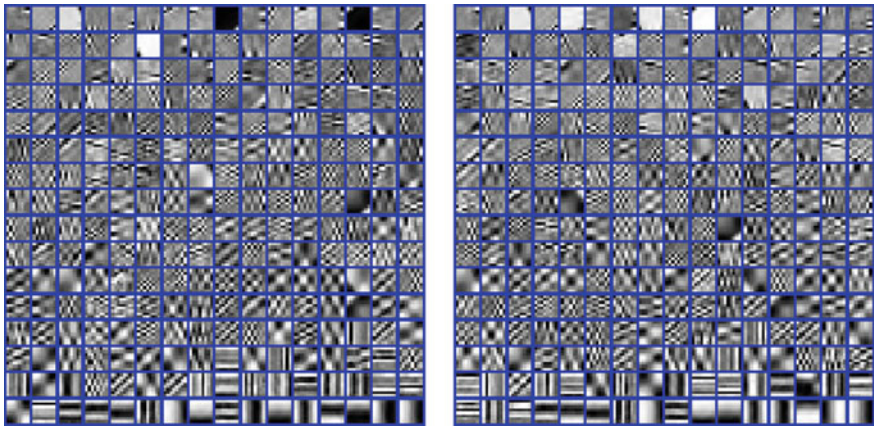


Fig. 26.6 OCNR5 Single-frame image learning dictionary atomic (*left* is MOD and *right* is K-SVD)

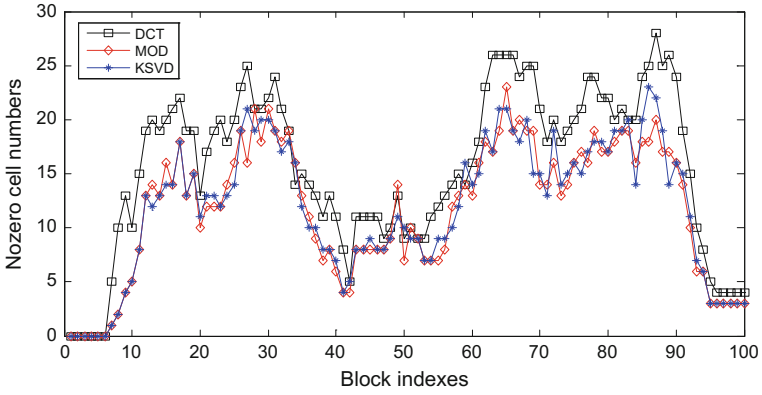


Fig. 26.7 The number of non-zero elements in sparse coefficients for OCR5 image blocks

kinds of dictionaries. The result is similar to the earlier when the OCNR5 image was processed. That is to say, when the image blocks are sparsely represented by the DCT dictionary, the number of non-zero elements in the corresponding coefficient vectors is about 7.49, and when the MOD and K-SVD dictionary are adopted, the number of non-zero elements in the coefficient vector is relatively small, the former average of about 5.63, the latter an average of about 5.72.

So we can conclusion that the performance of MOD and K-SVD dictionary is close and they are both better than DCT dictionary. Finally we select the K-SVD dictionary in this paper because it is simpler and more efficient than MOD.

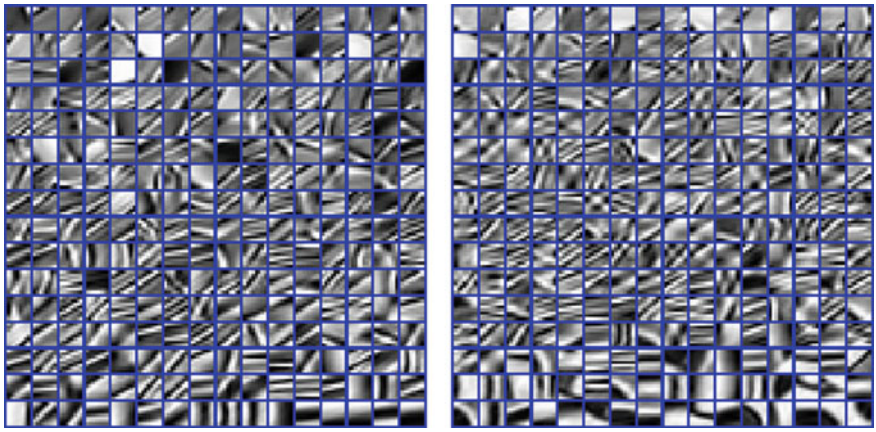


Fig. 26.8 Staturn single-frame image learning dictionary atomic (*left* is MOD and *right* is K-SVD

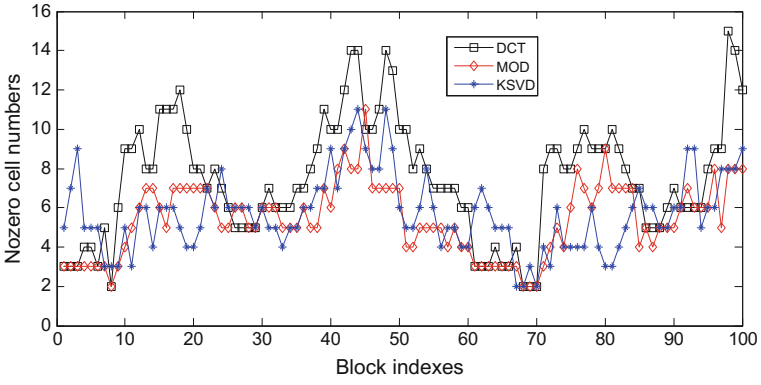


Fig. 26.9 The number of non-zero elements in sparse coefficients for Saturn image blocks

26.4.2 Parameter Determination

In this paper, the fusion data are shown as Fig. 26.10. They are multi focus images, medicine image include CT and MRI and satellite visible and near infrared images. The proposed algorithm involves two parameters that need tuning: (1) the size of image blocks n, (2) approximation residual ε. First, we fix the approximation residuals to study the performance of algorithm and the time consumed under different size of the image blocks. Then the image block size is fixed and the approximation residual is adjusted to study the algorithm performance and time consumed under different residuals. We adopted the nine performance evaluation indicators [25] for fusion performance evaluation, they are $Q_{MI}, Q_{TE}, Q_{NICE}, Q_G, Q_{SF}, Q_P, Q_S, Q_Y, Q_{CB}$.

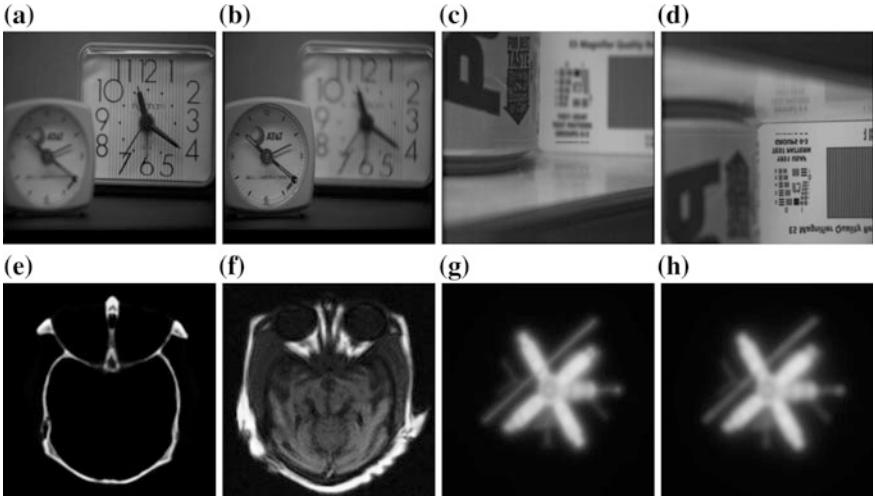


Fig. 26.10 Fusion image data **a**, **b** multi focus image 1; **c**, **d** multi focus image 2; **e**, **f** medicine image; **g**, **h** satellite image

From Fig. 26.11a–d, we can see the influence of different image block size and approximation residual on the time and performance of the fusion algorithm. Obviously, we can conclude that the image fusion performance increases with the increase of the image block and the decrease of the approximation residual, but the time-consuming is also increasing. But with the rapid increase in the level of computer hardware now, computing time-consuming will no longer be a problem. Through the above experiment, we set the image block size $n = 8 \times 8$, approximation residual $\varepsilon = 1$. (Figs. 26.12, 26.13)

26.4.3 Algorithm Comparison

If we compare the approved algorithm in this article with the five existed different fusion algorithms (LAP, PCA, DWT, CVT and NSCT), whose perimeters are measured according to the constant given in the document. The fusion image result is illustrated as the followings(Due to the space limitation, only show the image fusion result from multi-focus images and medical image), the performance evaluation index proposed by using literature 25 and eight different subjective performance evaluation using three kinds in terms of varied fusion result is shown s chart 1. Among this, the maximum in eight evaluation performance is expressed in bold and black. From the chart, better performance is obtained by the method proposed in this article, rather than the other five fusion methods.

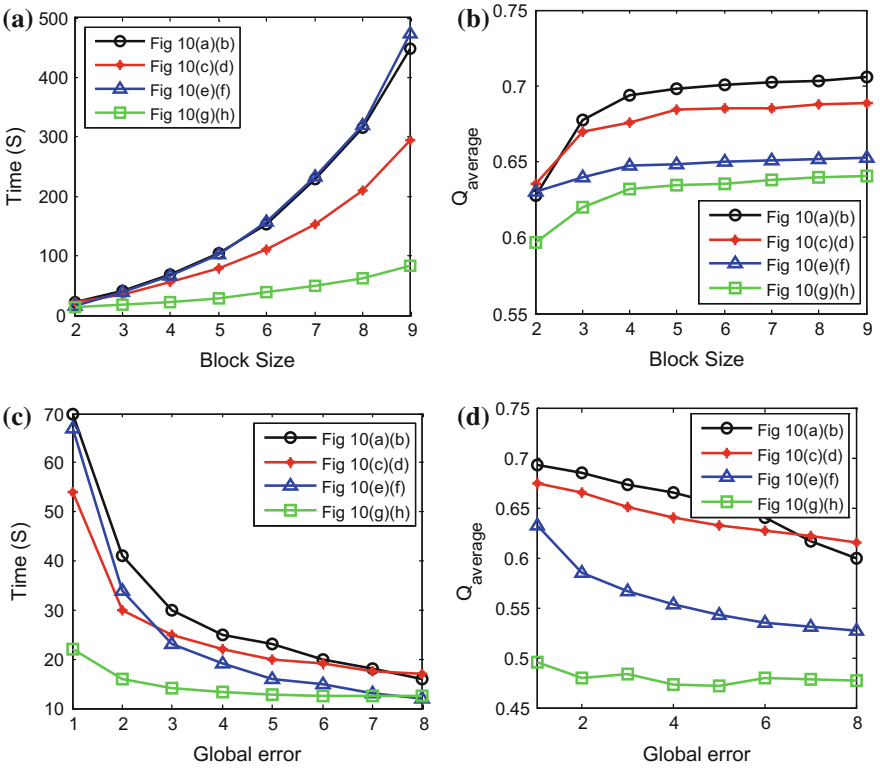
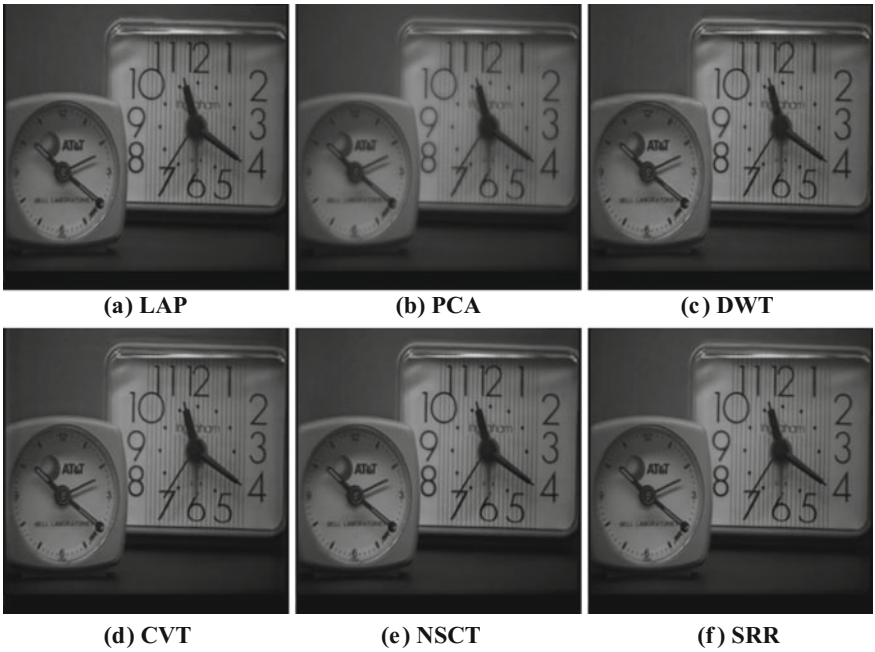


Fig. 26.11 Time—consuming and performance of image fusion algorithm under different image block sizes and approximation residuals

Source image	Evaluation index	LAP	PCA	DWT	CVT	NSCT	SRR
Fig. 26.10a, b	Q_MI	1.003	1.021	0.911	0.947	0.932	1.163
	Q_TE	0.465	0.469	0.467	0.474	0.462	0.442
	Q_NCIE	0.83	0.83	0.826	0.828	0.827	0.838
	Q_G	0.635	0.622	0.592	0.579	0.618	0.695
	Q_P	0.863	0.829	0.822	0.851	0.879	0.882
	Q_S	0.937	0.905	0.933	0.935	0.9359	0.938
	Q_Y	0.899	0.893	0.87	0.858	0.893	0.959
	Q_CB	0.739	0.681	0.713	0.698	0.746	0.765
Fig. 26.10c, d	Q_MI	1.025	1.06	0.924	1	1.037	1.178
	Q_TE	0.492	0.45	0.475	0.491	0.494	0.453
	Q_NCIE	0.831	0.833	0.827	0.83	0.832	0.839
	Q_G	0.623	0.624	0.578	0.57	0.622	0.675
	Q_P	0.844	0.775	0.786	0.818	0.842	0.845
	Q_S	0.944	0.887	0.94	0.942	0.945	0.944
	Q_Y	0.907	0.889	0.872	0.891	0.92	0.948

(continued)



(a) LAP (b) PCA (c) DWT

(d) CVT (e) NSCT (f) SRR

Fig. 26.12 Fusion results of different algorithm with multifocus images

(continued)

Source image	Evaluation index	LAP	PCA	DWT	CVT	NSCT	SRR
	Q_CB			0.68		0.723	0.726
Fig. 26.10e, f	Q_MI	0.456	0.535	0.359	0.383	0.327	0.907
	Q_TE	0.51	0.566	0.506	0.545	0.42	0.591
	Q_NCIE	0.806	0.729	0.804	0.805	0.803	0.821
	Q_G	0.646	0.845	0.571	0.449	0.301	0.877
	Q_P	0.539	0.544	0.458	0.45	0.294	0.58
	Q_S	0.805	0.534	0.725	0.717	0.631	0.851
	Q_Y	0.673	0.672	0.644	0.596	0.443	0.946
	Q_CB	0.411	0.335	0.382	0.402	0.455	0.47
Fig. 26.10g, h	Q_MI	0.916	1.127	0.925	0.964	0.984	1.26
	Q_TE	0.554	0.563	0.613	0.634	0.49	0.645
	Q_NCIE	0.809	0.813	0.809	0.81	0.812	0.813
	Q_G	0.602	0.594	0.517	0.505	0.439	0.662
	Q_P	0.188	0.197	0.148	0.16	0.192	0.378
	Q_S	0.824	0.825	0.83	0.821	0.813	0.834
	Q_Y	0.812	0.816	0.772	0.782	0.896	0.982
	Q_CB	0.255	0.218	0.223	0.204	0.24	0.297

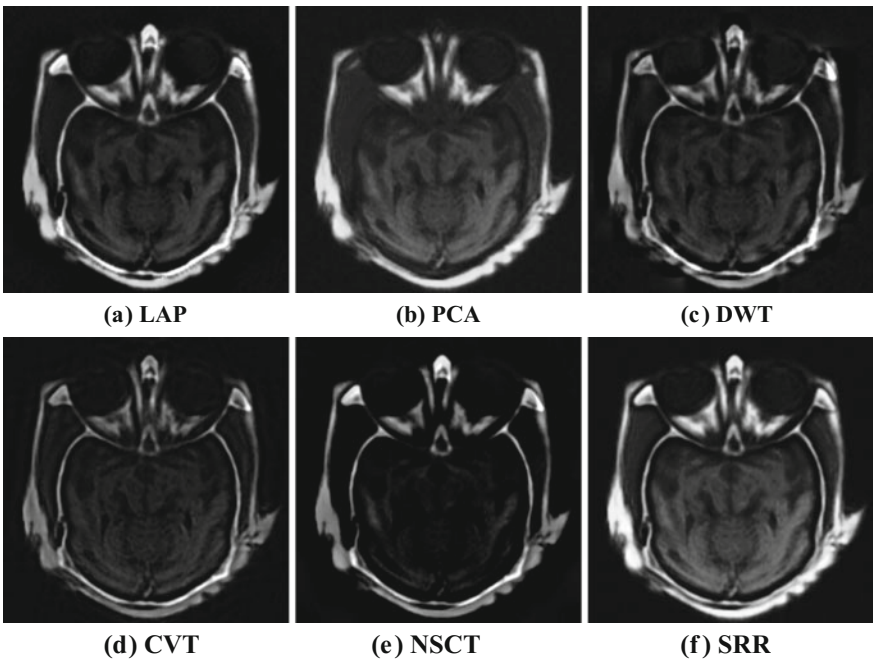


Fig. 26.13 Fusion results of different algorithm with CT and MRI images

Based on the figures above, PCA methods can result in blur problem in fused image, and from the result in the fusion of the satellite image, DWT, CVT and NSCT fusion image may produce reconstruction artifacts, which is due to the downsampling process. Although LAP fusion method has a reasonable fusion result, but to some extent, the contrast in the fusion image decreases. Taking the subjective visual evaluation together, the sparse representation fusion method proposed in this article has the best performance.

26.5 Conclusion

This paper studies the image fusion methods based on sparse and redundant representation theory. Meanwhile it works on the subject mainly in these two aspects: firstly, apply the sparse representation in the image fusion field, secondly, ascertain the dictionary training methods and parameterization approach in sparse representation theory through experiment. By experimental analysis of comparison between several varied fusion data types, the method in this article is more feasible and effective than LAP, PCA, DWT, CVT and NSCT both in subjective and

objective ways. Further steps in space objective image, a specific data type and application field, demand intensive studies on the combination of image denoising and fusion, which benefit much better visual effects.

References

1. Zhang Z, Blum RS (1999) A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application. Proc IEEE 87(8): 1315–1326
2. Liu Z, Forsyth DS, Safizadeh M-S (2008) A data-fusion scheme for quantitative image analysis by using locally weighted regression and Dempster-Shafer theory. IEEE Trans Instrum Meas 57(11):2554–2560
3. Goshtasby AA, Nikolov S (2007) Image fusion: advances in the state ofthe art. Inf Fusion 8 (2):114–118
4. Petrovic VS, Xydeas CS (2004) Gradient-based multiresolution image fusion. IEEE Trans Image Process 13(2):228–237
5. Pajares G, Cruz J (2004) A wavelet-based image fusion tutorial. Pattern Recognit 37(9): 1855–1872
6. Eltoukhy HA, Kavusi S (2003) A computationally efficient algorithm for multi-focus image reconstruction. Imaging 50(17):332–341
7. Li ST, Kwok JT, Wang YN (2001) Combination of images with diverse focuses using the spatial frequency. Inf Fusion 2(3):169–176
8. Lin PL, Huang PY (2008) Fusion methods based on dynamic segmented morphological wavelet or cut and paste for multi focus images. Signal Process 88(6):1511–1527
9. Li ST, Yang B (2008) Multifocus image fusion using region segmentation and spatial frequency. Image Vis Comput 26(7):971–979
10. Burt PT, Andelson EH (1983) The Laplacian pyramid as a compact image code. IEEE Trans Commun 31(4):532–540
11. Li H, Manjunath B, Mitra S (1995) Multisensor image fusion using the wavelet transform. Graph Models Image Process 57(3):235–245
12. Nencini F, Garzelli A, Baronti S, Alparone L (2007) Remote sensing image fusion using the curvelet transform. Inf Fusion 8(2):143–156
13. Song HH, Yu SY, Song L, Yang XK (2007) Fusion of multispectral and panchromatic satellite images based on contourlet transform and local average gradient. Opt Eng 46(2): 020502-1–020502-3
14. Rockinger O (1997) Image sequence fusion using a shift-invariant wavelet transform. Proc Int Conf Image Process 3:288–291
15. Li ST, Kwok JT, Wang YN (2002) Discrete wavelet frame transform method to merge Landsat TM and SPOT panchromatic images. Inf Fusion 3(l):17–23
16. Beaulieu M, Foucher S (1989) Multi-spectral image resolution refinement using stationary wavelet transform. Proc IEEE Int Geosci Remote Sens Symp 6:4032–4034
17. Cunha LD, Zhou JP (2006) The nonsubsampled contourlet transform: theory, design, and applications. IEEE Trans Image Process 15(10):3089–3101
18. Yang B, Li ST, Sun FM (2007) Image fusion using nonsubsampled contourlet transform. In: Proceeding of IEEE 4th International Conference on Image and Graphics, pp 719–724
19. Tessens L, Ledda A, Pizurica A, Philips W (2007) Extending the depth of field in microscopy through curvelet-based frequency-adaptive image fusion. In: Proceeding of IEEE International Conference of Acoustics, Speech, and Signal Processing, pp 1861–1864
20. Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse coding for natural images. Nature 381(6583):607–609

21. Davis G, Mallat S, Avellaneda M (1997) Adaptive greedy approximations. Constr Approx 13(1):57–98
22. Aharon M, Elad M, Bruckstein A (2006) K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans Signal Process 54(11):4311–4322
23. Gorodnitsky IF, Rao BD (1997) Sparse signal reconstruction from limited data using FOCUSS: a re-weighted norm minimization algorithm. IEEE Trans Signal Process 45(3): 600–616
24. Elad M, Aharon M (2006) Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans Image Process 15(2):3736–3745
25. Liu Z, Blasch E, Xue Z, Zhao J, Laganiere R, Wu W (2012) Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: a comparative study. IEEE Trans Pattern Anal Mach Intell 34(1):94–109

Chapter 27
An Improved Test Method to Study the pBRDF of the Rough Surface of Targets

Qing Liu, Yonghong Zhan, Di Yang, Yaping Wang and Change Zeng

27.1 Introduction

What we call the properties of the target are the attributes of the observed target showed on the detection equipment. And there are different kinds of characteristics according to different detection mechanisms, such as the optical characteristic of target and the radar characteristic of target [1]. In the optic area, we try to improve the abilities of target detection and recognition using the principle that surfaces of different targets will perform different properties of polarimetric bidirectional reflectance distribution. The polarimetric bidirectional reflectance distribution function, which is called pBRDF for short, can describe the characteristics of the intensities and polarization of the light reflected from the target surface in the upper hemisphere space. Since that with pBRDF, we could study the overall relations between the characteristics of reflectance light and the surface of target conveniently, pBRDF is one of the fundamental characteristic parameters to study the optical characteristic of target [2]. Usually, we get the pBRDF data of surfaces of different targets through experiments, and now it is an effective method to study the law of the polarimetric characteristics of light reflected from the different rough surfaces, which is the focus of this paper [3].

As consulted by other papers, a traditional method to measure the pBRDF data should use four different polarization states of incident light, which would be transformed into other polarization state by the Mueller matrix of the surface of the targets, then calculate all of the parameters of pBRDF matrix according to the algebra relations between them [4]. This kind of method need 16 times experiments to measure the pBRDF data of single-point on the surface of target, which is such complex and inefficiently, not to mention that the error is large. Considering this problem, in this paper, we introduce a new measure method, which just needs only

Q. Liu (✉) · Y. Zhan · D. Yang · Y. Wang · C. Zeng
Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China
e-mail: Qing_liu_0812@126.com

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_27

4 times experiments to get the all parameters of the pBRDF matrix of single-point. The verification verifies the correctness of the method, and apparently its efficiency.

27.2 The Experimental Theory and Method

27.2.1 pBRDF Measuring Method

According to the definition of the Stokes parameters, we can get the polarimetric result of the reflectance light by setting a polarization analyzer and some phase retardations before the detection. But after all is said and done, the results of the optical detection are intensities of the light, such as $I(0^\circ, 0^\circ)$, $I(45^\circ, 0^\circ)$, $I(90^\circ, 0^\circ)$, $I(135^\circ, 0^\circ)$, which means the intensities of different analyzer angles and phase retardation angles [5]. Then we can represent the Stokes parameters of the incidence of reflectance in formulation like:

$$\begin{cases} S_0' = I(0^\circ, 0^\circ) + I(90^\circ, 0^\circ) \\ S_1' = I(0^\circ, 0^\circ) - I(90^\circ, 0^\circ) \\ S_2' = I(45^\circ, 0^\circ) - I(135^\circ, 0^\circ) \\ S_3' = I(45^\circ, 90^\circ) - I(135^\circ, 90^\circ) \end{cases} \tag{27.1}$$

When measure the S_3' we used a ¼ phase retardation. After the measurement, we get the normalized Stokes parameters by normalization processing of the results, as follows.

$$S = \frac{1}{S_0'} [\, S_0' \quad S_1' \quad S_2' \quad S_3' \,]^T = [\, S_0 \quad S_1 \quad S_2 \quad S_3 \,]^T \tag{27.2}$$

After that, we begin to calculate the pBRDF matrix of the rough surface. At this point, there are three different methods to realize up to now, defining method, comparing method and comparative monadic method [6]. Defining method means to calculate pBRDF matrix according to its definition by the division operation between the Stokes of the incident and scattered light. But this method imposes restrictions on that the illuminant, the samples to detect, the detection and the experiment system must be calibrated strictly. However, it is very difficult to guarantee the accuracy of the calibration system and to avoid the interference noise of the experiment, which leads to large errors of the results, so this method is seldom used.

Comparing method needs to make a standard reflector, always taking advantage of the material of the tetrafluoroethylene (TFE). Measure the Stokes parameters of the scattering light reflected from the sample and the standard reflector under the

same experimental condition, then make a division to calculate the pBRDF matrix of the sample. This method can avoid the effect of the interference noise, but needs double times experiments than other methods, which makes the work too huge. In order to decrease the workload of the experiments, we try to measure the BRDF of the standard reflector at some fixed incident angle in advance, then the pBRDF of the sample to detect, make a division of them to invert the parameters of the PG model. This method we create is called comparative monadic method. Making the f_{00} of the pBRDF matrix as an example, its expression is as follow.

$$f_{00}(\theta_i, \varphi_i, \theta_r, \varphi_r) = \frac{L((\theta_i, \varphi_i, \theta_r, \varphi_r))}{E(\theta_i, \varphi_i)} \tag{27.3}$$

Where, $L(\theta_i, \varphi_i; \theta_r, \varphi_r)$ is the radiance of the scattering light, $E(\theta_i, \varphi_i)$ is the irradiance of the incidence, which means flux per unit area onto a surface and can be expressed as

$$\begin{aligned} E(\theta_i, \varphi_i) &= \frac{d\Phi(\theta_i, \varphi_i)}{dA_\perp} = \frac{\cos(\theta_i d\Phi(\theta_i, \varphi_i))}{dA} \\ &= \frac{\Phi}{4\pi r^2}\cos(\theta_i) = E(0^\circ, \varphi_i)\cos(\theta_i) \end{aligned} \tag{27.4}$$

Then the f_{00} of the Eq. (27.3) can be expressed as

$$f_{00}(\theta_i, \varphi_i, \theta_r, \varphi_r) = \frac{L(\theta_i, \varphi_i, \theta_r, \varphi_r)}{E(\theta_i, \varphi_i)} = \frac{L(\theta_i, \varphi_i, \theta_r, \varphi_r)}{E(0^\circ, \varphi_i)\cos(\theta_i)} \tag{27.5}$$

If we just make a measurement at 30° incident zenith angle about the standard reflector, then

$$f_{00b}(30^\circ, \varphi_i; \theta_r, \varphi_r) = \frac{L(30^\circ, \varphi_i; \theta_r, \varphi_r)}{E_b(0^\circ, \varphi_i) \cdot \cos 30^\circ} \tag{27.6}$$

At the same time, to standard reflector it has another expression as

$$f_{00b}(30^\circ, \varphi_i; \theta_r, \varphi_r) = \frac{\rho}{\pi} \tag{27.7}$$

Because of $E_s(0^\circ, \varphi_i; \lambda) = E_b(0^\circ, \varphi_i; \lambda)$, then the expression of the f_{00} of the sample to detect is

$$
\begin{aligned}
f_{00}(\theta_i, \varphi_i; \theta_r, \varphi_r) &= \frac{L(\theta_i, \varphi_i; \theta_r, \varphi_r)}{E_b(0^\circ, \varphi_i) \cdot \cos\theta_i} \\
&= \frac{L(\theta_i, \varphi_i; \theta_r, \varphi_r)}{[L(30^\circ, \varphi_i; \theta_r, \varphi_r)/[f_{00\,b}(30^\circ, \varphi_i; \theta_r, \varphi_r) \cdot \cos 30^\circ]] \cdot \cos\theta_i} \\
&= f_{00\,b}(30^\circ, \varphi_i; \theta_r, \varphi_r) \cdot \frac{L(\theta_i, \varphi_i; \theta_r, \varphi_r)}{L_b(\theta_i, \varphi_i; \theta_r, \varphi_r)} \cdot \frac{\cos 30^\circ}{\cos\theta_i} \\
&= \frac{\rho}{\pi} \cdot \frac{L(\theta_i, \varphi_i; \theta_r, \varphi_r)}{L_b(\theta_i, \varphi_i; \theta_r, \varphi_r)} \cdot \frac{\cos 30^\circ}{\cos\theta_i}
\end{aligned}
\tag{27.8}
$$

Thus it can be seen that, if we adopt the comparative monadic method to acquire the pBRDF matrix of the sample, we just need to measure the reflectance of the standard reflector one time, for example, 30° of the zenith angle, to get all parameters of pBRDF matrix of the sample to detect at any angle using Eq. (27.8). Obviously, this method we create can not only avoid the interference noise, but also increase the efficiency of the experiments.

27.2.2 The Angle Measuring Method

If we know the incidence angle, observation angle and relative azimuth angle, we can calculate all angles parameters in the model using the equations as follows [7].

$$\theta = \arccos\left(\frac{\cos\theta_i + \cos\theta_r}{2\cos\beta}\right) \tag{27.9}$$

$$\cos(2\beta) = \cos(\theta_i)\cos(\theta_r) + \sin(\theta_i)\sin(\theta_r)\cos(\Delta\varphi) \tag{27.10}$$

Incidence angle measuring method: make sure that the center of a circle of the arc-shaped supporter must be at the center of the reference platform when the illuminant supporter is installed, then the angle scale on the supporter is the incident zenith angle. Make the plane that the supporter is in as the reference plane, then the incident azimuth angle is 0°. Use the light intersection method to determine the height of the reference platform and the center of a circle of the arc-shaped supporter. In special, setting the laser at the 0° zenith angle of the supporter, the laser beam is at the height of the reference platform; setting the laser at the 90° zenith angle, the joint of the laser beam and the platform is the center of a circle of the arc-shaped supporter. Move the center of the platform to the joint in order to make sure the veracity of the angle measurement.

Observation angle measuring method: there are angle scales on the swivel bearing. Make sure that the observation zenith angle is 0 when the detection is above the center of the platform, then the reading on the scale of the detection stent

is the observation zenith angle. When the detection stent is 90° relative to the are-shaped supporter plane, the angle scale of the platform is 0, then spin the platform to change the observation azimuth angle. When experimenting, (1) fix the observation azimuth angle, spin the detection stent by 5°, (2) change the observation angle by 15° one time, then repeat step (1).

Relative azimuth angle measuring method: set the plane that the supporter is in as the x, y plane of the coordinate system, then the incidence azimuth angle is 0, and the observation azimuth angle is the relative azimuth angle that the incidence to the reflected.

27.2.3 pBRDF Matrix Calculating Method

As stated earlier, it is able to get all parameters of the pBRDF matrix of single point on the rough surface through measurements if different states of the incident light are accessible. Traditional methods need 16 times measurements to get all parameters, so it is essential to improve. It is proven that, the Mueller matrix of the isotropic material possesses symmetrical structure as follow [8].

$$\begin{cases} \vec{L_{r0}} = \begin{bmatrix} L_0 \\ L_1 \\ L_2 \\ L_3 \end{bmatrix} = \begin{bmatrix} f_{00} & f_{01} & 0 & 0 \\ f_{01} & f_{00} & 0 & 0 \\ 0 & 0 & f_{22} & f_{23} \\ 0 & 0 & -f_{23} & f_{22} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} f_{00} \\ f_{01} \\ 0 \\ 0 \end{bmatrix} \\ \vec{L_{r1}} = \begin{bmatrix} L_0 \\ L_1 \\ L_2 \\ L_3 \end{bmatrix} = \begin{bmatrix} f_{00} & f_{01} & 0 & 0 \\ f_{01} & f_{00} & 0 & 0 \\ 0 & 0 & f_{22} & f_{23} \\ 0 & 0 & -f_{23} & f_{22} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} f_{00}+f_{01} \\ f_{01}+f_{00} \\ 0 \\ 0 \end{bmatrix} \\ \vec{L_{r2}} = \begin{bmatrix} L_0 \\ L_1 \\ L_2 \\ L_3 \end{bmatrix} = \begin{bmatrix} f_{00} & f_{01} & 0 & 0 \\ f_{01} & f_{00} & 0 & 0 \\ 0 & 0 & f_{22} & f_{23} \\ 0 & 0 & -f_{23} & f_{22} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} f_{00} \\ f_{01} \\ f_{22} \\ f_{23} \end{bmatrix} \\ \vec{L_{r3}} = \begin{bmatrix} L_0 \\ L_1 \\ L_2 \\ L_3 \end{bmatrix} = \begin{bmatrix} f_{00} & f_{01} & 0 & 0 \\ f_{01} & f_{00} & 0 & 0 \\ 0 & 0 & f_{22} & f_{23} \\ 0 & 0 & -f_{23} & f_{22} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} f_{00} \\ f_{01} \\ -f_{23} \\ f_{22} \end{bmatrix} \end{cases} \tag{27.11}$$

Through analysis of the Eq. (27.11), we know that if set the polarization state of the incident light as 45°/135° linear polarization or circular polarization, all the parameters of the pBRDF matrix can be measured by least amount of measuring frequency. If the incident light is natural light, we can deduce other parameters of

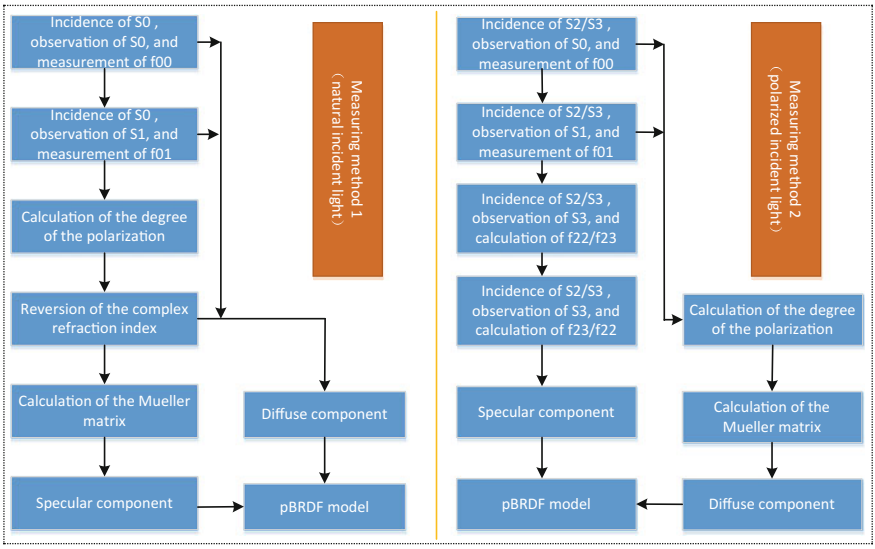


Fig. 27.1 The schematic diagram of the measuring and data processing of the pBRDF matrix

the pBRDF model by inversing the complex refractive index of the sample, then work out the pBRDF matrix. In theory, it just needs two points on the surface to be measured when we want to construct the model, however it needs to measure some more points in consideration of the inversing precision of the parameters. Considering that, we design two different measuring methods, as showed in Fig. 27.1.

There are two methods described in Fig. 27.1. One adopts the natural incident light, measures several points to get the data of f_{00} and f_{01} under the same incident conditions. After that, substitute the data of f_{00} and f_{01} into the pBRDF model to inverse the complex refractive index and other factors, then calculate all parameters of the pBRDF matrix. All of this is tested and verified in my other papers, the reader can refer to the reference documents [9, 10]. The other one adopts the polarized incident light, then all of the parameters of the pBRDF matrix can be measured according to Eq. (27.9). This method can test and verify the correction of the factors inversed, so it is a potential method.

In order to decrease the numbers of measuring cycles, all measurements can be done at ten different azimuth and zenith angles without measuring all points on hemisphere space. Since what stated before is only applicated to single point at single condition, we need to move forward a single step to get all pBRDF models at different conditions of incidence and observation. Then it is well-reasoned to study the polarimetric characteristics of the rough surfaces.

27.2.4 The Equipment of Experiment

The polarization scattering characteristics test system consists of test turntable, data acquisition processor, wide spectrum illuminant, fiber optical spectrometer, liquid crystal polarization camera, optical elements and so on. The test turntable contains an illuminant supporter, the reference platform and the control computer, while the data acquisition processor contains a system of PXI acquisition unit, and the whole system is equipped with two computer workstation. In one workstation which we name it as workstation A runs the software to control the test turntable, in the other workstation which we name it as workstation B runs the software to control the data acquisition processor, whose specific functions are as follows: workstation A is connected to workstation B by two serial ports, one serial port exports pulse signal whether the turntable is in place, the other exports the location signal of the reference platform. In experiments, test turntable runs to the position according to the instructions of the workstation A, when it is done, the workstation A will export a pulse signal to workstation B through RS422, then workstation B triggers the detection to collect the data and receive the location signal at the same time. Then workstation B exports the experiment data into the database.

The illuminant supporter is fixe on the wall and is divided into ten sections by 10°. There is a mounting hole on every section in order to mount the illuminant at fixed position, as showed in Fig. 27.2.

Choose a tungsten halogen lamp as the natural illuminant, and four laser devices as the polarized illuminant. Their fundamental information is as follows. The natural

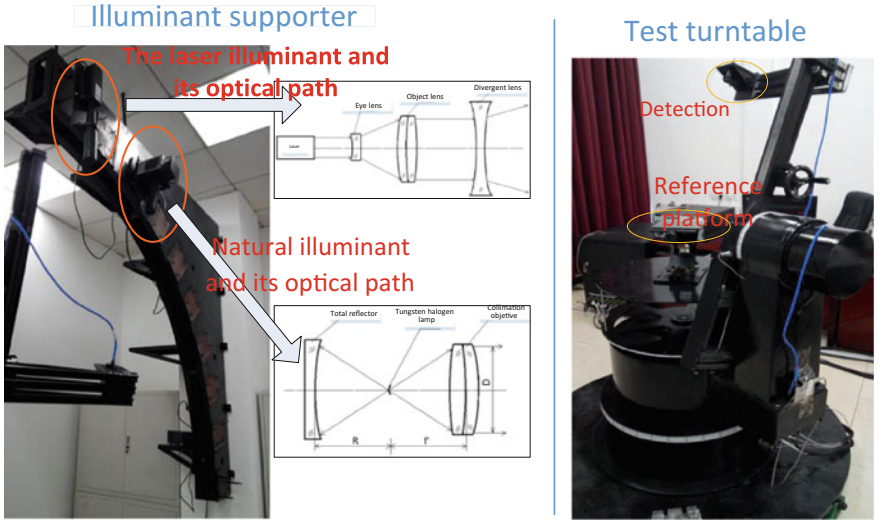


Fig. 27.2 The photograph of the illuminant and test system to measure the pBRDF data of the samples

illuminant's power is no less than 250 W and its power stability exceeds 97%. When illuminating by collimation, the diameter of the natural illuminant is between 30 and 300 mm more or less. The laser illuminants have four kind of wavelength: 457, 532, 671, 1064 nm, and among them, the power of 457 nm laser is no less than 1 W while the others' is above 1.5 W. In order to mount laser easily, we design a universal interface, provide a piece of standard attenuation lens that the magnifications are 5, 10 and 20 dB. The weight of all illuminant devices is less than 5 kg.

The test turntable is a two axis turntable which can support the liquid crystal polarization camera, fiber optical spectrometer and other detection devices to measure the characteristics of the scattering light on different angles of zenith and azimuth. The turntable consists of a reference platform, a detection cantilever supporter and the control workstation, showed in Fig. 27.2.

The test turntable adopts vertical structure, and its azimuth rotating shaft of the axis is perpendicular to the horizontal plane, while its pitch axis runs in the plane perpendicular to horizontal plane too. The detection we used in the experiments is mounted on the pitch axis by the cantilever supporter, and its measuring range is $\pm 180°$ in azimuth angle and $\pm 90°$ in zenith angle. The angle-position precision is better than $\pm 0.02°$. Both of minimum smooth angular velocities of azimuth and zenith are better than 0.005°/s. The maximum angular velocity is more than 15°/s, and the maximum angular acceleration is more than $15°/s^2$. The sample to detect is set on the center of the platform, and its size is bigger than Æ500 × 300 mm. The distance from detector to the surface of the sample can be changed, and its adjustable extent is between 300 and 1000 mm.

The control computer workstation mainly controls the movement of the test turntable. It has more than three channels of the serial port expansion card. Meanwhile, the PXI data acquisition system can also control the turntable, and receive the pulse signal of the movement in place. When experiment, test turntable can work according to the command file of the azimuth and zenith angles pre-decided, as well as according to the real-time command.

The liquid crystal polarization camera is produced by Bossa Nova Company of USA named SALSA which can measure total polarization information correctly in R, G, B wave band. Its precision of the degree of linear polarization is more or less 1.2%, the precision of the polarization angle is 0.4°, the precision of the degree of circular polarization is less than 1.0%

27.3 The Experiments and the Results

27.3.1 Information About the Standard Reflector and the Samples to Detect

The standard reflector we adopt in our paper is made up of the tetrafluoroethylene (TFE). Its surface is quite smooth and uniformity without any obvious blemishes

and imperfections. Measure the scattering characteristics of the standard reflector under the natural illuminant, and calculate the criterion data of the measurements according to the calibration results of the standard reflector, which serves as the benchmark data of the after experiments and data processing.

The sample to detect in our paper is green coating on the surface of the aluminum plate. Because of that the thickness of the coating is 0.1 mm and the visible light cannot penetrate it, the scattering characteristic of the surface is regarded as the characteristics of the green coating.

27.3.2 The Experimental Results of the Standard Reflector

The standard reflector is relative ideal Lambert reflector, so its scattering light is even distribution all over the twoπ hemisphere space, besides the light is more or less unpolarized. According to this features, the data measured can be the benchmark data as well as can be the validation for the correction of the test system. Specifically speaking, if the analysis of the data of the measuring standard reflector demonstrates to be more or less lambert reflectance, the test system is reliable, otherwise the test system cannot meet the precision requirements, and so it needs to redesign the test system.

A tungsten halogen lamp is illuminated at 20° of the zenith angle, and detection adopts the SALSA camera with a green wavelength filter before its lens working at 180° of the relative azimuth angle. The results of the measurement are showed in Fig. 27.3 and Table 27.1. Figure 27.3 shows the result of the measuring at 180° of the relative azimuth angle and the specular reflection direction which is 20° of

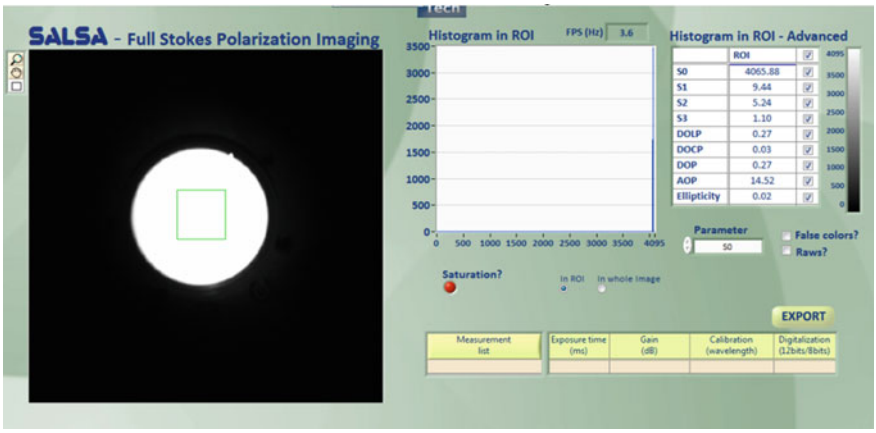


Fig. 27.3 The measurement result of standard reflector at the specular direction when incident zenith angle is 20°

Table 27.1 Measurement results of the standard reflector

Zenith angle		−40	−35	−15	−10	0	10	20	30
Parameters	S_0	4065.79	4065.85	4065.87	4065.69	4065.1	4065.54	4065.53	4063.89
	S_1	9.44	9.49	9.51	9.69	10.13	10.36	10.56	10.97
	S_2	5.24	5.24	5.24	5.24	5.38	6.38	7.86	8.03
	S_3	1.1	1.1	1.1	1.1	1.12	0.98	2.43	2.77
	DoLP	0.0026	0.0027	0.0027	0.0027	0.0028	0.0030	0.0032	0.0033
	DoCP	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002	0.0006	0.0007
	DoP	0.0027	0.0027	0.0027	0.0027	0.0028	0.0030	0.0033	0.0034
	AoP	14.52	14.45	14.53	14.20	13.98	15.81	18.33	18.10

observation zenith angle in this paper, while Table (27.1) shows measurement results at other observation zenith angles.

From the figure and table we can come to some conclusion. Firstly, the scattering energy of the platform is so small relative to the scattering energy of the standard reflector that it can be ignored. Secondly, the measurement results of the S_0 of the standard reflector are very stable, more or less fluctuating around 4065 mV, but they are decreasing little by little while the observation zenith angle grows, this testifies that the standard reflector is an relative ideal Lambert reflector. Thirdly, the polarization components such as S_1,S_2, S_2, DoP etc. can be ignored when observation zenith angle is small, specifically, they are stable at 9.44, 5.24, 1.10 mV, 0.0027 respectively. However, when the angle is big, the polarization components fluctuate severely, which is because of the noise reflected from the wall and other things. All of the results of the experiments demonstrate the reliability of the test system.

27.3.3 The Experimental Results of the Green Coating

The general process of the research about the pBRDF of the surface of the fundamental material is that: measure the scattering data of the samples at some positions, then analysis the data to calculate other factors of the rest positions to construct the pBRDF model. In this paper, we choose the green coating as the representative. Same to the experiments of the standard reflector, the results of the measurements showed in Fig. 27.4 and Table 27.2.

According to the analysis of the data we know that: Firstly, the maximum reflected energy appears at the specular direction, and decreases when leaving the

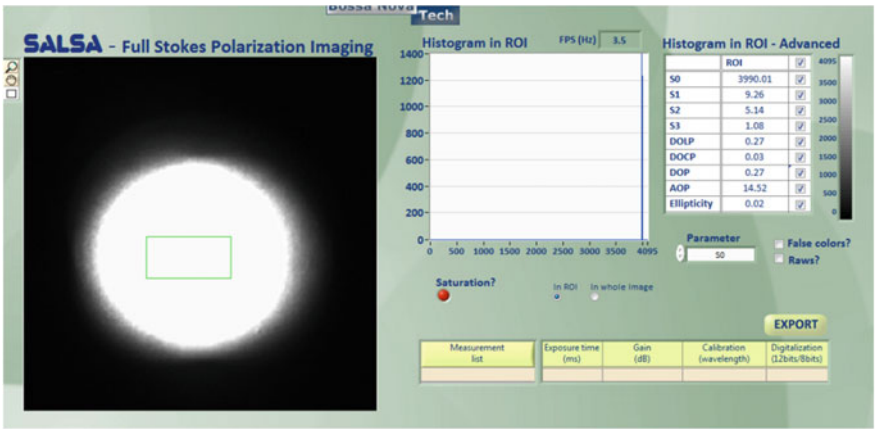


Fig. 27.4 The measurement result of green coating at the specular direction when incident zenith angle is 20°

Table 27.2 Measurement results of the standard reflector

Zenith angle		−15	−10	−5	5	35	40	45	50
Parameters	S_0	799.26	834.59	1056.22	3192.84	3162.57	1744.38	1158.91	884.43
	S_1	33.05	39.42	60.61	368.08	1147.39	592.74	331.08	189.75
	S_2	22.22	28.61	40.93	58.54	51.41	66.54	48.88	57.45
	S_3	1.16	0.3	0.9	186.32	42.2	26.61	4.74	1.56
	DoLP	0.0498	0.0584	0.0692	0.1167	0.3632	0.3419	0.2888	0.2242
	DoCP	0.0014	0.0004	0.0008	0.0584	0.0133	0.0153	0.0041	0.0018
	DoP	0.0498	0.0584	0.0692	0.1305	0.3634	0.3423	0.2888	0.2242
	AoP	16.96	17.98	17.02	4.52	1.28	3.2	4.2	8.42

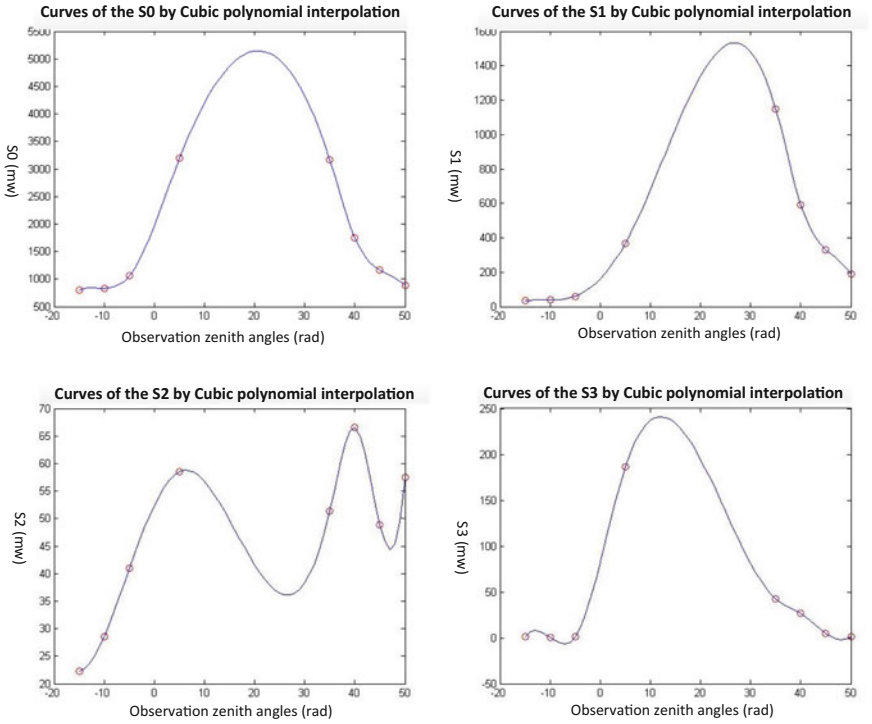


Fig. 27.5 The measured data and interpolation curves of the S_0, S_1, S_2 and S_3 components of the green coating

specular direction. Secondly, the backscatter of the green coating is not obvious, and there is no mirror reflection peak. Thirdly, the polarization components of the scattering light from the surface of the green coating are changing with the change of the observation angles, and it changes violently, what accounts for the envisage that there is an optimal observation angle of the polarization detection.

Make an interpolation operation to the data in the Table 27.2 to get data of all the zenith angles, the results showed in Figs. 27.5 and 27.6.

From Figs. 27.5 and 27.6 we know that, the scattering intensity S_0 of the green coating is of normal distribution and the biggest S_0 appears at the specular direction, this illustrates the micro-facet theory that the direction of the micro-facet is about Gaussian distribution. Meanwhile, polarization components S_1 is of normal distribution too, but the peak appears at the bigger zenith angle, this illustrates that polarization detection can gain more information than gray-scale imaging because of the noncoincidence of their peaks. In addition, the S_2 component fluctuates violently, and S_3 is more or less normal distribution, but the data is so small compared to the others, so it can be ignored. The curves of the index of the polarization rise with the angles before 40°, while decrease with the angles after 40°, and arrive the peak at 40°. The experiment results have a bit of differences

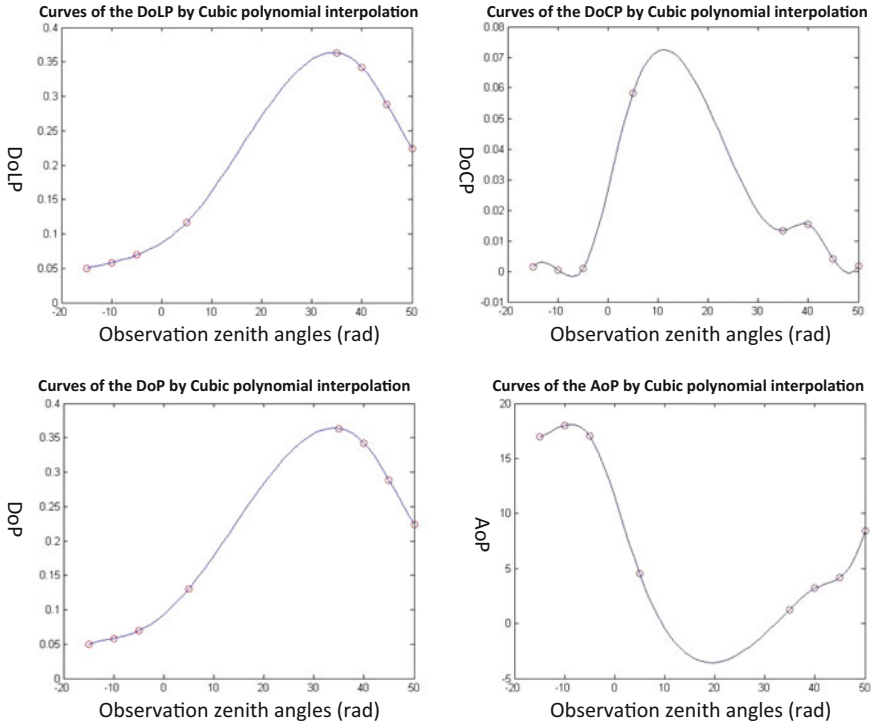


Fig. 27.6 The measured data and interpolation curves of the DoLP, DoCP, DoP and Aop components of the green coating

compared to the simulation curves, this is because of that measurement at big observation angles is effected by bigger noises from the wall and other surroundings.

However, the experiments illustrate the correction of the measuring method we design to measure the pBRDF data in high efficiency.

27.4 Conclusions

This paper described the theory and method to measure the polarimetric characteristics of the rough surfaces in detail, as well as the experimental equipment, experimental procession and the experiment results. It introduced a comparative monadic method to measure the Stokes parameters, and the angle calculation formula of the test system. It introduced a new test method to get the pBRDF matrix through analysis the symmetry structure of the pBRDF matrix. It introduced the composition of the pBRDF test system we designed, as well as the sample to detect and the standard reflector. After experiments, we obtained the pBRDF data of the

standard reflector and the green coating, and analyzed the data that verified the feasibility of our new method. Specifically, the standard reflector is more or less Lambert reflector and can be regarded as the benchmark of the test. In addition, the reflectance characteristics of the green coating illustrates that polarization detection has better detection efficiency than gray-scale imaging detection at some angles. In our subsequent work, we will do more experiments of the samples by the method introduced in this paper, and analyze the pBRDF data to study the law of the polarization characteristics of the scattering light from different fundamental materials, then improve the pBRDF model.

References

1. Cao YH, Wu ZHS, Zhang HL (2008) Research on visible light scattering of spatial targets based on spectral BRDF of target samples. Acta Photonica Sin 37(11):2269–2273
2. Priest RG, Germer TA (2000) Polarimetric BRDF in the microfacet model: theory and measurements. In: Proceeding of 2000 meeting of the military sensing symposia specialty group on passive sensor, pp 169–181
3. Ward GJ (1992) Measuring and modeling anisctropic reflection. Comput Graph 26(2): 265–272
4. Zhao YQ, Pan Q, Cheng YM (2011) Imaging spectropolarimetric remote sensing and application. National Defence Industrial Press, Beijing
5. Lu WG, Wu YM, Gao LM (2013) Rapid measurement of spatial azimuth by using polarized light. Opt Precis Eng 21(3):539–545
6. Yang TH, Zhao YQ, Pan Q (2008) A new measurement method of polarimetric spectral bidirectional reflectance distribution function. Acta Photonica Sin 37(12):2520–2524
7. Wang X, Zou XF, Jin WQ (2011) Study of polarization properties of radiation reflected by roughness objects. Trans Beijing Inst Technol 31(11):1327–1331
8. Priest RG, Meier SR (2002) Polarimetric microfacet scattering theory with applications to absorptive and reflective surface. Opt Eng 41(5):988–993
9. Liu Q, Zhan YH, Yang D (2015) Parameters of the polarimetric bidirectional reflectance distribution function of rough surfaces and parameter inversion. J Spacecr TT&C Technol 34(5):482–487
10. Zhan YH, Liu Q, Yang D (2015) Inversion of complex refractive index for rough-surface objects. Opt Precis Eng 23(8):2178–2184

Chapter 28
Study on the Geometric Super Resolution by Code Division Multiplexing Technology

Di Yang, Xinyue Liu, Change Zeng and Yonghong Zhan

28.1 Introduction

The spectrum, gray, and space resolution should affect the performance of the optic-electric system [1]. The space resolution is the key specification for non-spectrum imaging system and non-energy calibration imaging system. For remote sense optic-electric imaging system, the instantaneous field of view was small, so increasing the optical aperture should realize the high space resolution and acquire the enough power to meeting the operating requirements easily [2, 3]. However, the size of the detective pixels may be about 10–30 μm, the limited sample frequency should limit to increase the space resolution [4].

To resolve the problem, Jonathan Solomon in Tel Aviv college of Israel provided code division multiplexing method (CDMA) [5, 6]. The light from the object should be encoded in frequency, and then be decoded in frequency after imaged on the CCD. The method means that the origin image had been sampled twice, one is in the frequency field, and another is in space field. Focus on the CDMA method applied in practice, the simulation, analysis, and study should be provided on the paper [7–9].

D. Yang (✉) · C. Zeng · Y. Zhan
Beijing Institute of Tracking and Telecommunications Technology,
Beijing 100094, China
e-mail: guangxuesheji@sina.com.cn

X. Liu
China Academy of Sciences Institute of Optoelectronics, Beijing 100094, China

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_28

28.2 Under-Sample Effect

Suppose that the focus length for some optical system is 300 mm, the entrance aperture is 150 mm, and the image range is 10 km, the resolution on the object space is 0.04 m for lens. However, if the pixel size is 10 μm, the acquired space resolution should be only 0.33 m. About 10 times difference should exist between the lens' resolution and the system's resolution. For the optical system approaching the diffraction limit, the aperture might be larger, and the non-zero pixel should be affecting the image performance seriously. The under-sample should limit the space resolution for the optical system [10].

The effect could be analyzed in frequency field. For an ideal optical system in the incoherent light illumination, optical transfer function (OTF) based on amplitude is shown as formula (28.1). Where, ρ_0 and f_0 are radius and cutoff frequency. For a circular aperture, as the formula shown, the optical transfer function is multiplex of the two triangle functions in x and y coordinates.

$$\begin{aligned} G(\rho) &= \frac{2}{\pi}\left[\cos^{-1}\left(\frac{\rho}{2\rho_0}\right) - \left(\frac{\rho}{2\rho_0}\right)\sqrt{1-\left(\frac{\rho}{2\rho_0}\right)^2}\right] \\ &= G(f_X, f_Y) = \Lambda\left(\frac{f_X}{2f_0}\right)\Lambda\left(\frac{f_Y}{2f_0}\right) \end{aligned} \tag{28.1}$$

The detector could be expressed by comb function. The g(x,y) is the inverse Fourier transformation of $G(f_x,f_y)$. The sample course of the detector could be expressed by formula (28.2). In the frequency field, the multiplex in space filed could be transformed as convolution of comb function and $G(f_x,f_y)$, which is shown as in the right hand of the formula (28.2).

$$\begin{aligned} &g_s(x,y) = comb\left(\tfrac{x}{X}\right)comb\left(\tfrac{y}{Y}\right)g(x,y) \\ &\Rightarrow G_s(f_X,f_Y) = F\left\{comb\left(\tfrac{x}{X}\right)comb\left(\tfrac{y}{Y}\right)\right\} * G(f_X,f_Y) \end{aligned} \tag{28.2}$$

The image system could be approximate line system. So the two dimension sample could be multiplex by two samples in dimensions. Only one dimension was considered for now, which is shown as Fig. 28.1. The right and left direction arrow

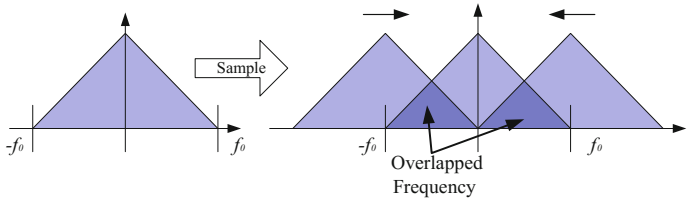


Fig. 28.1 OTF and duplicated OTF in the frequency field

in the figure means the course that the frequency gets closer caused by sample interval distance larger, at last, the two frequencies get overlapped.

The interval between the pixels and the interval between frequencies are inverse proportion. For ideal optical system, the sample frequency needs to be higher than the lens cutoff frequency, which means that the sample frequency need to meet the requirement of Whittaker–Shannon sample principle, so the overlap could not be existed. If the under-sample is existed and the triangle function gets closer enough, the frequency should be existed, which should effect the performance of the optical system.

28.3 Encoded and Decode Method

The basic method for encode and decode as follow: the virtual panel which is Fourier transfer panel of object should be existed near by the aperture, if the Fourier transfer panel could be encoded in frequency field, after the origin image have been imaged in the CCD, we could decoded for restoring the origin image, which is shown as Fig. 28.2.

The light of the object transferred by the lens before sampled should form a real image. Based on the object-image conjugate relationship, the image could be recognized as an object, and then on the fore-focus panel a frequency image should be existed, shown as formula (28.3).

$$U_f = c \cdot \int_{-\infty}^{+\infty} t_0(x_0, y_0) \cdot \exp\left[-j\frac{2\pi}{\lambda f}\left(x_0 x_f + y_0 y_f\right)\right] \mathrm{d}x_0 \mathrm{d}y_0 \tag{28.3}$$

where, d_0 is equal to f, and t_0 is transmissivity. We can see that if the object is in the infinite range, on the fore-focus panel an object frequency should be existed. Which means the lens is a Fourier transfer. If a code panel is in the frequency panel, the relationship between interval distance of the codes and of the pixels should be shown as formula (28.4).

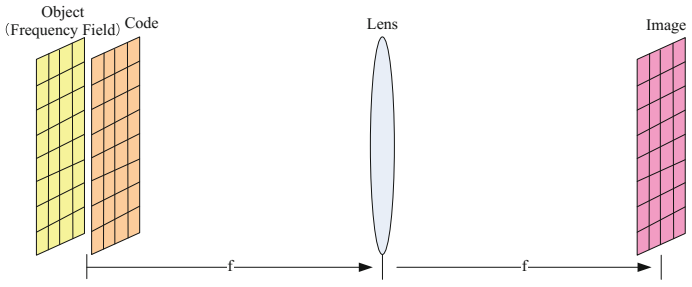


Fig. 28.2 Basic scheme of code

$$t \cdot s = \lambda \cdot f \tag{28.4}$$

To design the code, the detector sample frequency should be considered. If we wanted to increase the resolution two times, the new sample unit needed insert the center between two pixels, and the encode should be the most effective. The code must be orthogonal for decode the original image. The code could be 01 code. For simple 01 code, the transmissivity is 25% (50% of one dimension). In order to increase the transmissivity, some special code needed to be design. For simulating easily, in the paper only 01 code was considered. As we known the Fourier transfer is line transfer. The code panel in the fore-focus point could be expressed as follow.

$$\begin{aligned} S(x) &= \mathrm{rect}\left(\frac{x}{x_s}\right) \otimes \mathrm{comb}\left(\frac{x}{x_i}\right) \\ C(y) &= \mathrm{rect}\left(\frac{y}{y_t}\right) \otimes \mathrm{comb}\left(\frac{y}{2y_t}\right) \end{aligned} \tag{28.5}$$

where, S(x) and C(y) are stand for sample panel and code panel. x_i is sample interval distance, x_s is pixels size, and x_s is smaller than x_i, the ratio of x_s and x_i are duty ratio, y_t is code unit size. The relationship of x and y could meet the requirement of formula (28.4).

Supposed that an object was imaged on the focus panel, the inverse Fourier transfer should be existed near the aperture. If we set the code panel in the frequency panel, the image frequency and code panel should be multiplied. This means that the code panel only allows the special light pass. The sampled image could be expressed as formula (28.6).

$$\begin{aligned} I_s = F\left[F^{-1}(I_o) \cdot \mathrm{rect}\left(\frac{y}{y_t}\right) \otimes \mathrm{comb}\left(\frac{y}{2y_t}\right)\right] \\ \cdot \mathrm{rect}\left(\frac{x}{x_s}\right) \otimes \mathrm{comb}\left(\frac{x}{x_i}\right) \end{aligned} \tag{28.6}$$

From the formula (28.6) we can see that the code panel is like a space sample. The detector not only sampled the original image, but also sampled the coded frequency image. From the formula (28.6) and (28.4), we can deduce the formula (28.7).

$$\begin{aligned} I_s = \left(2|y_t|^2\right) \cdot I_o \cdot \mathrm{rect}\left(\tfrac{x}{x_s}\right) \otimes \mathrm{comb}\left(\tfrac{x}{x_i}\right) \\ \otimes \underline{[\sin \mathrm{c}(y_t x) \cdot \mathrm{comb}(2y_t x)]} \end{aligned} \tag{28.7}$$

From the formula (28.7), we can see that the image sampled by the CCD is a normal image convoluted with the frequency of the code. The normal image had been sampled by the CCD and the code frequency. After decoded, the original image could be retrieved. The retrieved image method is using the code to multiply the CCD detected image frequency, which is shown as formula (28.8).

$$I_s = F\left[F^{-1}(I_o)\cdot \mathrm{rect}\left(\frac{y}{y_t}\right)\otimes \mathrm{comb}\left(\frac{y}{2y_t}\right)\right]\cdot \left\{\begin{array}{l}\left[\mathrm{rect}\left(\frac{x}{x_s}\right)\otimes \mathrm{comb}\left(\frac{x}{x_i}\right)\right]\cdot \\ \left[\mathrm{rect}\left(\frac{y}{y_t}\right)\otimes \mathrm{comb}\left(\frac{y}{2y_t}\right)\right]\end{array}\right\} \tag{28.8}$$

Based on the formula (28.4), the multi-sample could be combined, which is shown as follow.

$$I_s = F\left[F^{-1}(I_o)\cdot \mathrm{rect}\left(\frac{y}{y_t}\right)\otimes \mathrm{comb}\left(\frac{y}{2y_t}\right)\right] \tag{28.9}$$

where, the frequency had not lost and the original image had been retrieved because the code panel existed.

28.4 Simulation

The design of the experiment is shown as Fig. (28.3). The experiment set is a 4f system. On the fore-focus panel is a high resolution film illuminated by a laser source. In the Fourier panel which is in the center of the system an aperture and a code was set. The aperture capability is preventing the light out of the view of the field enter the image optical path which shall cause the aliased. The code is 01 code, and in the 4 quadrant the 1000, 0100, 0010, and 0001 are existed. 0 means pass, and 0 means block. The code size is 40 40 μm, the aperture is 10 mm, the optical focus is 400 mm and the pixel size is 10 μm.

The illuminator could be incoherent light. The Fourier transfer includes space transfer and phrase transfer, and the phrase could not affect the power image, because the OTF equals the CTF's autocorrelation normalization function. When using the 01 code, the OTF is the quadratic of the CTF. The coherent light was considered in the paper.

From the Fig. 28.4a is the original image, which film's resolution is 256 × 256, and the pixel size is 2.5 μm. Because the aperture is 10 mm, the image should be

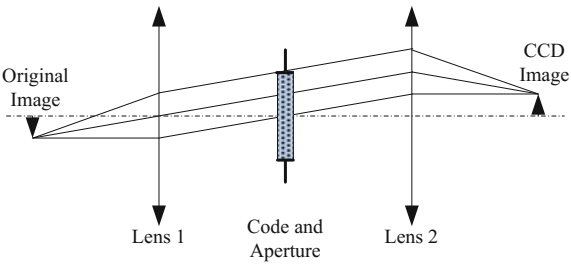


Fig. 28.3 The optical route of the experiment

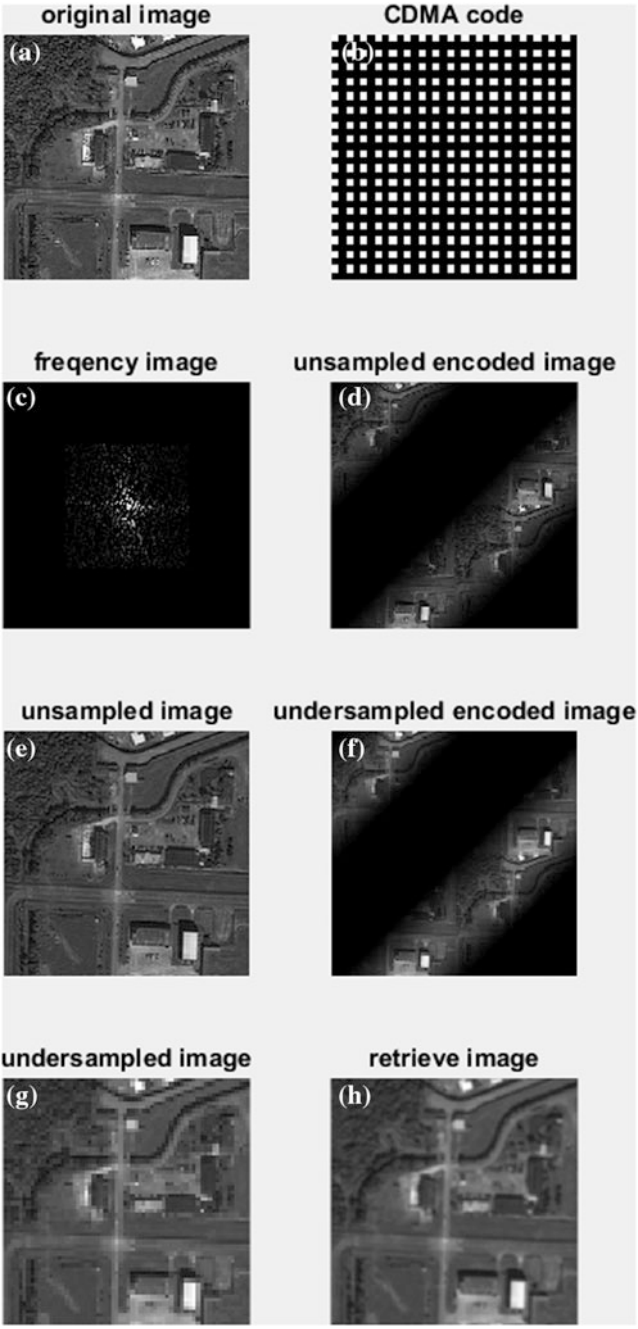


Fig. 28.4 Simulation of the CDMA

transfer into frequency field, and only center frequency could pass the aperture, which is shown as b. The filed image transferred in space field is shown as c. The figure d is sampled image by the CCD and the figure e is the CDMA code, and the code is the 01 code. In the frequency, the code is multiply with the frequency image, which is shown as figure f. This image is transferred in space field, and the bright dark stripe is caused by the code. The strip was sampled by the CCD, which is under-sampled and is shown as figure g. Noticed that the imaged area should be lager, the 64×64 pixels should enlarge to 128×128 pixels, the pixel size is same. At last, the under-sampled should be transferred into the frequency filed, and multiplied with the 01 code, then, the image still should be transferred into the space field. The retrieved image is shown as figure h.

Comparing the figure d which is directed sampled by the CCD with the figure h which is used the encode and decode CDMA imaging, we can see that the resolution increased, and we can get more detail information from the picture. For macro view, the CDMA technology is using the larger scale pixel to repair the shortage of the under-sampled image. For technology, the CDMA is similar as grating which can separate the frequency apart and encode the frequency to prevent the image from the alias cause by the close frequency.

28.5 Discussions

Focus on the effect of the under-sample problem of the optical system which should cause the decrease of the image resolution, a CDMA method is provided for refrain the image alias. The course of the image is simulated, and the feasibility is validated in the paper. No moveable unit is in the system, and is convenience for modifying the existed system to improve the resolution. The method is suit for applying to the situation in where the gray resolution is not important.

Several applications could be discussed. (1) If the code panel is not in the Fourier panel, how to realize the CDMA is effectively. Should we do some experiment to measure the distant between the code panel and the Fourier panel for retrieve the original image perfectly. (2) The only the 01 code was validated, the other type code, higher resolution, and higher transmissivity should be validated. (3) The CDMA provide a new method for design the zoom lens without the movable unit. (4) The 4f system should have image 2 times. How to clear the alias of the atmosphere using the CDMA code in the frequency is an interesting field. (5) The spectral with should affect the CDMA image. Coherent illuminator is different from the incoherent. So, the special method should be used in the incoherent light illuminating.

References

1. Jonathan S, Zeev Z, David M (2005) Geometric superresolution by code division multiplexing. Appl Opt 44(1):32–40
2. Lukosz W (1966) Optical systems with resolving powers exceeding the classical limit. J Opt Soc Am 56:1463–1472
3. Gartner W, Lohmann AW (1963) An experiment going beyond Abbe's limit of diffraction. Z Phys 174:18–21
4. Kartashev AI (1960) Optical systems with enhanced resolving power. Opt Spectrosc 9: 204–206
5. Riesenberg R, Seifert T, Berka A, Dillner U (1999) Opt micromechanical super resolution detector system. In: Merkle F (ed) Design and engineering of optical systems II, Proc SPIE (3737), pp 367–383
6. Peleg S, Keren D, Schweitzer L (1987) Improving image resolution using sub pixel motion. Phys Rev Lett 5:223–226
7. Keren D, Peleg S, Brada R (1988) Image sequence enhancement using sub-pixel displacements. Comput Vis Pattern Recognit 88:742–746
8. Bertero M, De Mol C (1997) Superresolution by data inversion. In: Wolf E (ed) Progress in optics. North-Holland, Amsterdam
9. Xian HY, Fu ZHZH, Wan Q (2011) Block-based adaptive super resolution algorithm. Opto-Electron Eng 38(9):119–123
10. Zhao XY, Su Y, Wang JM (2010) Kind of reconstruction method of CCD aliasing image in frequency domain. Appl Res Comput 27(1):368–373

Chapter 29
Processing the Reflectance Data of Rough Surface for Inversing the Index of Refraction

Yonghong Zhan, Di Yang, Qing Liu, Change Zeng and Yaping Wang

29.1 Introduction

Polarization embody the vibrate direction and phase error of the electric-field vector, and embody transverse wave characteristics [1, 2]. When the light reflected by the rough surface, the power and transmit direction should change, and the polarization characteristic should change too [3]. Using the polarization detection could acquire the rich information except the power and spectrum; and detection capability of the optical imaging system should be improved. Polarization image is a useful tool for remote sense [4].

Polarization characteristic of target was related with geometry characteristic of the target and polarization characteristic of target material. The material polarization characteristic is related with statistic micro-face and complex refraction index of the material surface. The statistic micro-face should effect the transmit direction of the light; and complex refraction index should effect the polarization status of the light [5, 6]. The complex index of the refraction is the core parameters for generating the polarization characters. So, studying the characteristic of target should study and measure the complex refraction index of the target material.

The method to acquiring the complex index of refraction of the material included transmission measurement method and reflection measurement method [7–9]. The transmission method need to measure the refraction angle for inversing the index, so the method was used for translucent material; the classic reflection method need to measure the Brewster angle or phase error, and require the material is mirror face. Vimal and Milo brought out a method that measuring the pBRDF data. The core processing method is acquiring the polarization degree according to the phase angle for inversing the complex index of the refraction. Because the total reflection light

Y. Zhan (✉) · D. Yang · Q. Liu · C. Zeng · Y. Wang
Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China
e-mail: zhanyonghong@sina.com

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_29

was assumed mirror reflection light, and the Lambert light was ignored, the inversed index should not be accurate.

Focus on the above problem, the Vimal–Milo method was studied on, inversing method of complex index of the refraction based on "relative polarization value & phase angle" was brought out, the inversing method model was established, and contrast validation was performed in the experiment. The experiment data indicate that the advanced method could avoid the effect of the Lambert reflection light for inversing the index, and the accuracy was improved, especially for rough surface material.

29.2 Problem Describe

29.2.1 P-G Model

Vimal–Milo inversing method was based on the measurement of material pBRDF, and pBRDF was brought out based on the scalar BRDF. The pBRDF could describe the power and polarization power distribution according to the semi-sphere space above the rough surface. The pBRDF was defined was the ratio of illuminance on the surface and luminance off the surface, the formula is shown as follow [10, 11] (Fig. 29.1).

$$\vec{F}_r(\theta_i, \theta_r, \Delta\varphi, \lambda) = \mathrm{d}\vec{L}_r(\theta_i, \theta_r, \Delta\varphi)/\mathrm{d}\vec{E}_i(\theta_i, \theta_r) \tag{29.1}$$

where, (θ_i, φ_i) incident light, (θ_r, φ_r) is sight line, λ is wavelength, F_r is pBRDF, $\mathrm{d}L_r$ is luminance of the sight line, $\mathrm{d}E_i$ is illuminance of the incident light. For the incident could be expressed as 4×1 Stokes vector, the F_r is a 4×4 pBRDF matrix.

Priest and Germer brought out a classic pBRDF model which was based on Torrance-Sparrow scale model in 2002. The P-G model includes two parts: the mirror reflected light and non-mirror reflected light. The mirror reflect model is a physical model, and the non-mirror model is an experience model. The mirror reflect model is shown as follow.

$$f_{jl}(\theta_i, \varphi_i, \theta_r, \varphi_r, \lambda) = f_{\mathrm{spec}} + f_d = \frac{m_{jl}(\beta, n, \kappa) f_{SO} f_P}{4 \cos\theta_{\mathrm{i}} \cos\theta_{\mathrm{r}}} + f_d \tag{29.2}$$

where, subscript j and l are stand for row and column, and the value is 0–3. f_{spec} is mirror reflected component. f_d is non-mirror reflected component which is appeared in f_{00} caused by rough surface scatter. f_{SO} and $\mathrm{f_P}$ are shadow and obscure function and probability of micro-face normal function those is derived by the micro-face assume. The P-G model's Mueller matrix is shown as follow.

$$M_{mueller} = \begin{bmatrix} m_{00} & m_{01} & 0 & 0 \\ m_{01} & m_{00} & 0 & 0 \\ 0 & 0 & m_{22} & -m_{23} \\ 0 & 0 & m_{23} & m_{22} \end{bmatrix} \tag{29.3}$$

29.2.2 Vimal–Milo Method

Based on the P-G model, when the incident light is nature light, the reflected light from the surface is expressed as the formula.

$$\begin{bmatrix} S_0^r \\ S_1^r \\ S_2^r \\ S_3^r \end{bmatrix} = \vec{F}_r \times \begin{bmatrix} S_0^i \\ 0 \\ 0 \\ 0 \end{bmatrix} = S_0^i \begin{bmatrix} f_{00} \\ f_{01} \\ 0 \\ 0 \end{bmatrix} \tag{29.4}$$

where, $S^i{}_0$ is illuminance of depolarized incident light; is luminance of reflected light, the subscript i is 0–3. Using the define of Stokes vector and formula (29.2, 29.4), we could get the formula of polarization degree which is shown as follow.

$$P(n,k,\beta) = \frac{f_{01}}{f_{00}} = \frac{\xi m_{01}}{\xi m_{00} + f_d} = \frac{S_1^r}{S_0^r} \tag{29.5}$$

where,

$$\xi = \frac{f_S o f_p}{4 \cos\theta_{\mathrm{i}} \cos\theta_{\mathrm{r}}} \tag{29.6}$$

The mirror reflected light was considered by Vimal–Milo method, then, the scatter component f_d was ignored. So, the depolarization degree is expressed as follow.

$$P(n,k,\beta) = \frac{S_1^r}{S_0^r} = \frac{m_{01}}{m_{00}} \tag{29.7}$$

where, $S^r{}_0$, $S^r{}_1$ is measured value, β is a function of incident angle and reflectance angle. In the formula (29.7) have two uncertain parameters, if the polarization degree value could be tested in two different angle, the n,k should be inversed by Vimal–Milo method.

According to formula (29.5), (29.6), and (29.7), when and only when $\xi m_{01} \gg f_d$, in another word, the non-mirror reflected light is very low compared with mirror reflected light, the complex index of refraction of the surface could be inversed by

Vimal–Milo method. In fact, the real material surface is rough surface, the appliance of the Vimal–Milo method was restricted in very narrow field.

29.3 Inversing the Complex Index of Refraction

29.3.1 Basic Theory

Based on the Vimal–Milo method, non-mirror reflect component f_d in $S^r{}_0$ is the reason that the formula might false. Avoiding the effect of the scatter light for inversing the index complex of refraction, two methods could be considered: solving the and avoiding the f_d. Because f_d is different according different angle, solving the f_d is difficult. So, we wish avoiding the f_d or *Sr 0*.

When, the Flux of the light is expressed, the measured value of the reflected light could be expressed as follow.

$$I = I_{spec} + I_d \tag{29.8}$$

Conform to stokes vector (Only S_0, and S_1).

$$S_0 = (I_{spec} + I_d)_{(0^\circ,0^\circ)} + (I_{spec} + I_d)_{(90^\circ,0^\circ)} \tag{29.9}$$

$$S_1 = (I_{spec} + I_d)_{(0^\circ,0^\circ)} - (I_{spec} + I_d)_{(90^\circ,0^\circ)} \tag{29.10}$$

Based on the P-G model, non-mirror reflected light which is depolarized is caused by the scatter of the rough surface.

$$I_d\left(0^\circ, 0^\circ\right) = I_d\left(90^\circ, 0^\circ\right) \tag{29.11}$$

Put above formula to the formula (29.10).

$$S_l = (I_{spec})_{(0^\circ,0^\circ)} - (I_{spec})_{(90^\circ,0^\circ)} \tag{29.12}$$

S_1 is related to the mirror reflected light, using the relationship between f_{01} and S_1 to inverse the complex index of refraction. As the function (2) shown, the scatter S_1 is according to the f_{01} in the pBRDF matrix. And the f_{01} is shown as.

$$f_{01} = \frac{m_{01} f_{SO} f_p}{4\cos\theta_\mathrm{i}\cos\theta_\mathrm{r}} = \frac{S_1^r}{S_0^i} \tag{29.13}$$

Comparing the function of solving the polarization degree, and assume the parameter y.

$$y(n,k,\beta) = \frac{f_{01}}{\bar{f}_{01}} = \frac{S_1^r/S_0^i}{\bar{S}_1^r/S_0^i} = \frac{S_1^r}{\bar{S}_1^r} \tag{29.14}$$

where, f_{01}, S^r_1 was chosen for reference parameters which could be any measurement value theoretically, and *y(n,k,β)* replaces the polarization degree *P(n,k,β)* for inversing the complex index of refraction.

29.3.2 Inversing Function

f_{01} is the line polarization component (except non-mirror component). For Fresnel reflect law, the line polarization component is the flux error of s-wave and p-wave.

$$f_{01} \propto m_{01} = \frac{1}{2}\left|R_s - R_p\right| \tag{29.15}$$

The s-wave and p-wave reflectance R_s and R_p is the function of complex index of refraction and incident angle, which is shown as follow.

$$R_s = \frac{(A - \cos\beta)^2 + B^2}{(A + \cos\beta)^2 + B^2} \tag{29.16}$$

$$R_p = R_s \cdot \left[\frac{(A - \sin\beta\,\tan\beta)^2 + B^2}{(A + \sin\beta\,\tan\beta)^2 + B^2}\right] \tag{29.17}$$

where, A, B are process parameter, shown as follows.

$$A = \sqrt{\frac{\sqrt{C} + D}{2}} \tag{29.18}$$

$$B = \sqrt{\frac{\sqrt{C} - D}{2}} \tag{29.19}$$

$$C = 4n^2k^2 + D^2 \tag{29.20}$$

$$D = n^2 - k^2 - \sin^2(\beta) \tag{29.21}$$

where, β is the angle between incident light and normal of micro-face, shown as follow.

$$\cos(2\beta) = \cos(\theta_i)\cos(\theta_r) + \sin(\theta_i)\sin(\theta_r)\cos(\Delta\varphi) \tag{29.22}$$

Shadow and obscure function f_{SO} and normal probability function of micro-face f_P are the function of geometry characteristic parametersτ, Ω, σ, B_n, shown as follows.

$$f_{SO}(\theta_N, \beta, \tau, \Omega) = \frac{\Omega + \theta_N e^{-2\beta/\tau}}{\Omega + \theta_N} \tag{29.23}$$

$$f_p(\theta_N, B, \sigma) = \frac{B_n e^{-\frac{\tan^2(\theta_N)}{2\sigma^2}}}{2\pi\sigma^2\cos^3(\theta_N)} \tag{29.24}$$

where, θ_N is the angle between normal of micro-face and means normal of surface.

$$\cos(\theta_N) = \frac{\cos(\theta_i) + \cos(\theta_r)}{2\cos(\beta)} \tag{29.25}$$

Put the function (29.16–29.25) to function (29.2),

$$f_{01} = \frac{(A - \cos\beta)^2 + B^2}{(A + \cos\beta)^2 + B^2} \cdot \frac{A\sin\beta\tan\beta}{(A + \sin\beta\tan\beta)^2 + B^2} \cdot \frac{\Omega + \theta_N e^{-2\beta/\tau}}{\Omega + \theta_N} \cdot \frac{B_n e^{-\frac{\tan^2(\theta_N)}{2\sigma^2}}}{8\pi\sigma^2\cos^3(\theta_N)} \cdot \frac{1}{\cos\theta_i\cos\theta_r} \tag{29.26}$$

Put the function (29.26) to (29.13)

$$y(n,k) = \frac{\frac{(A-\cos\beta)^2 + B^2}{(A+\cos\beta)^2 + B^2} \cdot \frac{A\sin\beta\tan\beta}{(A+\sin\beta\tan\beta)^2 + B^2} \cdot \frac{\Omega + \theta_N e^{-2\beta/\tau}}{\Omega + \theta_N} \cdot \frac{e^{-\frac{\tan^2(\theta_N)}{2\sigma^2}}}{\cos^3(\theta_N)} \cdot \frac{\cos\bar{\theta}_i\cos\bar{\theta}_r}{\cos\theta_i\cos\theta_r}}{\frac{(\bar{A}-\cos\bar{\beta})^2 + \bar{B}^2}{(\bar{A}+\cos\bar{\beta})^2 + \bar{B}^2} \cdot \frac{\bar{A}\sin\bar{\beta}\tan\bar{\beta}}{(\bar{A}+\sin\bar{\beta}\tan\bar{\beta})^2 + \bar{B}^2} \cdot \frac{\Omega + \bar{\theta}_N e^{-2\bar{\beta}/\tau}}{\Omega + \bar{\theta}_N} \cdot \frac{e^{-\frac{\tan^2(\theta_N)}{2\sigma^2}}}{\cos^3(\bar{\theta}_N)}} \tag{29.27}$$

where, n, k are unknown value, the others are known value and process value. When the reference parameters θ_i, $\bar{\theta}_r$ were confirmed, and measure the value y according to charged angle of incident or reflect, the complex index of refraction should be acquired by function (29.27).

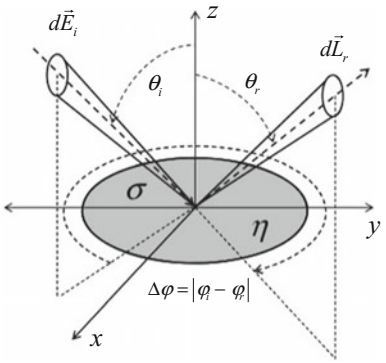


Fig. 29.1 Geometrical illustration of pBRDF

29.3.3 Algorithm Design

Considerate that the incident face and reflect face are the same plane, the angle between the incident and reflect light, when the incident angle is fixed and reflect angle is change at the test, we would get different y value according to the reflect angle.

$$y_j(n,k,\theta_{rj}) = \frac{f_{01j}}{\bar{f}_{01}} = \frac{S^r_{1j}}{\bar{S}^r_1} \tag{29.28}$$

where, $j = 0,1,2,\ldots n$. f_{01j} are the measurement value f_{01} at the angle of θ_{rj}. When the test data groups are more than 3 groups, function (29.28) are over determined non-line equation. For using the least square method, the evaluation function should be given as follow.

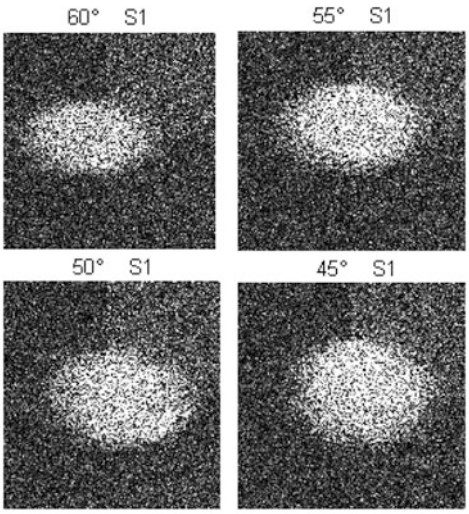
$$z(\mathbf{x}) = \sum_{j}^{R} \left(y_j(n_i,k_i,\theta_{rj}) - \hat{y}_j(\theta_{rj})\right)^2 \tag{29.29}$$

where, $x = [n,k]^T$ are the solved value, $y_j(n_i,k_i,\theta_{rj})$ are calculated value, and $\hat{y}(\theta_{rj})$ are measurement data.

29.4 Experiments and Results

29.4.1 Experiments and Data

Experiment object are green paint and aluminum plane. For improving the test accuracy standard Lambert body plant was using in the test. The light source is tungsten bromine lamp, the stability is 97%; the detector is SALSA polarization



Degree	40°	45°	50°	55°	60°	65°	70°
Green Paint	37.7	47.8	61.2	78.6	102.1	135.0	184.0
Al. Plane	82.2	104.1	132.1	168.6	217.7	285.8	387.1

Fig. 29.2 Some imaging S_1 data of green paint's surface

Table 29.1 Estimating results of the green paint and Aluminum

	Material	Reflect angle	n	k	z_{min}
Proposal method	Green paint	40°–70°	1.37	0.34	0.0055
	Al. plane		1.25	6.59	0.0343
Vimal-Milo method	Green paint	40°–70°	1.47	0.47	0.1362
	Al. plane		5.02	7.14	1.5791

camera, degree of line polarization is 1.2%, angle of polarization is 0.4°, and accuracy of the polarization degree is 1.0%; and position precise of the rotating platform mounting the light is ±0.02°. The light source is mounted on the shelf, and incident angle is 20°, the camera mounted on the rotating platform rotate between 40° and 70°, the step interval range is 5°, the gain of the camera is 22 dB. The camera image and the test data is shown as Fig. 29.2.

29.4.2 Inversing the Index

For green paint the 45° test data and for aluminum plane 55° are selected for reference value. Put the S1 data to the function (29.29), and use the least square algorithm to inverse index. The inversed result included the Vimal–Milo method are shown as Table 29.1.

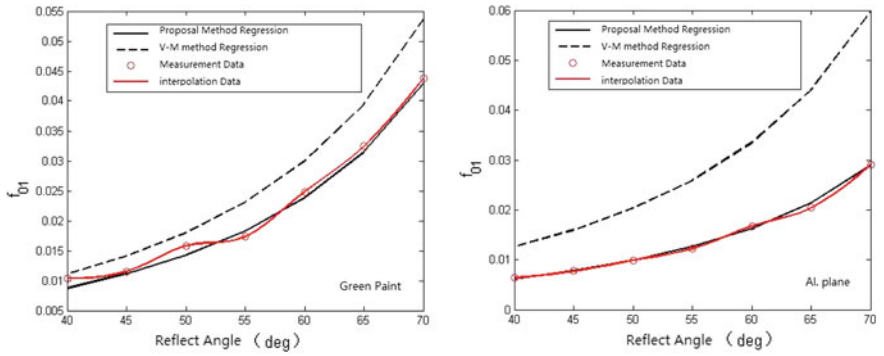


Fig. 29.3 Green paint and Aluminum plane f01 curves of simulation by model and fitted with experimental data with the change of angles

The complex index of refraction of aluminum is 1.24 + i6.60, the inversed result of proposal is very close to the real value, and while the inversed resulted of Vimal–Milo method is far from the real value. The evaluation factor z_{min} is a parameter which could be used to evaluation the coincidence between the inversed value and experiment value. Comparing the Vimal–Milo method, the proposal method could acquire high coincidence.

29.4.3 Regression Validation

Put the inversed result by proposal and Vimal–Milo method to function (29.26) for regressing the valid. The regression analysis is shown as Fig. 29.3. The left figure is green paint result, and the right one is aluminum plane result. The circle point is measurement data, the red line is interpolation data, the dash line is regression data used Vimal–Milo inversed result, and the black line is regression data used proposal inversed result.

The root-mean-square deviations of the proposal for green paint and aluminum plane are 0.0022 and 0.0015, while the root-mean-square deviations of the Vimal–Milo are 0.0148 and 0.0452. Comparing the root-mean-square deviations, we can see, the proposal method have low error to the measurement data, which means the proposal method is a fine method. Comparing the regression data between the green paint and aluminum, we can see, for rough surface, the two inversed methods have fine accuracy for smooth surface.

29.5 Conclusion

The complex index of refraction is an important parameter for studying the target polarization characteristics. The classic method of inversing the complex index of refraction is not meeting the accuracy requirement for rough surface. An advanced inversed method is suggested which based on "relationship between the relative polarization and phase angle". The non-mirror component could not be draw into the inversed processing of the proposal method, so it could be used not only for smooth surface but also rough surface.

References

1. Lu WG, WU YM, GAO LM et al (2013) Rapid measurement of spatial azimuth by using polarized light. Opt Precis Eng 21(3):539–545
2. Kerekes J, Hart C, Gartley M et al (2010) Characterization of material reflectance variation through measurements and simulation. In: Proceedings of SPIE 7695(0P), pp 1–11
3. Wang X, Zou XF, JIN WQ (2011) Study of polarization properties of radiation reflected by roughness objects. Trans Beijing Inst Technol 31(11):1327–1331
4. Zhang Y, Zhao HJ, XING H et al (2009) Polarization aberrations of transmitting rotationally symmetric optical systems. Opt Precis Eng 22(5):757–762
5. Sawyer MA, Hyde MW (2013) Material characterization using passive multispectral polarimetric imagery. In: Proceedings of SPIE 8873(0Y), pp 1–15
6. Li-xiang MA, Fan-ming LI, Ji-yong NIU et al (2013) Polarization model based on complex refractive index and its applications. Laser Infrared 43(10):1338–1341
7. Li Q, Yang RF, ZHAI Y et al (2013) Nondestructive measurement of refractive index of capillary in non-paraxial conditions. Opt Precis Eng 21(3):616–623
8. Thilak V, Creusere CD, Voelz DG (2006) Estimating the complex index of refraction and view angle of object using multiple polarization measurements. IEEE (1-4244-0785-0/06), pp 1067–1071
9. Thilak V, Voelz DG, Creusere CD (2007) Polarization-based index of refraction and reflection angle estimation for remote sensing applications. Appl Opt 46(30):7527–7536
10. Liu Q, Zhan YH, Yang D et al (2014) Methods of modeling of polarized BRDF for rough surfaces. J Spacecr TT&C Technol 33(4):354–359
11. Zhan YH, LIU Q, YANG D (2015) Inversion of complex refractive index for rough-surface objects. Opt Precis Eng 23(8):2178–2184

Chapter 30
Polarization Optical Image Processing Used in the Target Detection and Identification

Change Zeng, Qing Liu, Di Yang, Yonghong Zhan and Yaping Wang

30.1 Introduction

People see images everywhere and every day, and people get much more information from the images rather than other medium like words, sounds etc. The images transmit some massages to people directly, and make people understand easily. For example, we can quickly figure out where the targets are or how to get the destination in the picture without any trouble, while it is difficult to distinguish these information from words, especially for those who never be trained. On account of these advantages, the images are becoming one of the most important media people use to communicate in their everyday lives, and almost all of information people get is showed as a picture. However, there are so many noises around and disturbing the transmission of the images, especially those obtained in the faraway place to detect something useful. So it is important to make the images clearly that people can tell the differences between things they are interested in and the other background. Since then, there are so many ways used in this area of study, however, they are all methods about how to deal with the images already acquired, and their results cannot be satisfied with needs of the users all the time [1].

All of the images are the intensity characteristics of light, but the light has more than one characteristic, there are also wavelength, phase and polarization, which can be used to show different characters of the targets. For example, the wavelength can describe not only the motion characteristics, but also the temperature, component etc. of the targets with other parameters of the light, like amplitude. What's more, the polarization of the light can show so much information about the target, such as its roughness and its complex index of refraction [2, 3], as showed in Fig. 30.1.

C. Zeng · Q. Liu (✉) · D. Yang · Y. Zhan · Y. Wang
Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China
e-mail: Qing_liu_0812@126.com

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_30

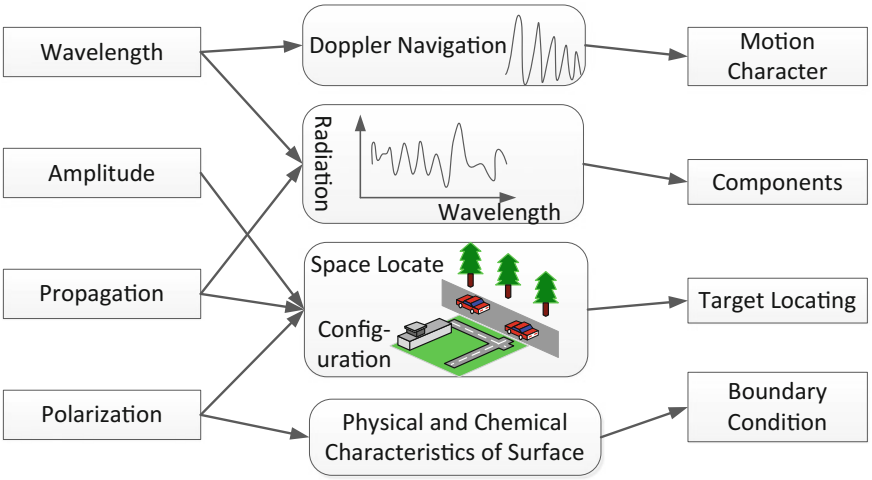


Fig. 30.1 The different characteristics of light figure out the targets' characteristics

On the other hand, study has verified that all the things in the world, no matter where they are, on the ground or flying in the air, or hiding under the water, will change the polarization characteristics of the light reflected from their surface. In the other words, light reflected from different surfaces will show different polarization forms, and the differences will become significance between the nature background and material man made. So we can make use of this discovery to make optical image processing, then it will be easy for us to figure out the interesting target from complex background. In this way, we call the optical image processing as polarization front-processing, and it will show people clear pictures they want [4].

30.2 Theories of Polarization Front-Processing

To begin, we will emphasize the optical theories. All of the inventions we employ in our daily lives are invented based on the basis optical theories. And without exception, polarization front-processing of optical image is based on the optical theories indeed. In this chapter we will introduce some of the principles and rationales we refer simplify, but we try to make you understand our ideas directly.

The first principle we need to talk is the Fresnel reflectance and refraction equation. Imagine that, a light incident on a flat surface, as showed in Fig. 30.2, what performance it will show? Is its polarization or phase changed? Using our eye only, we cannot distinguish that, but we know there something changes actually. Though we can see nothing of these changes, the Fresnel reflectance equation can describe all the changes, even those details [5, 6].

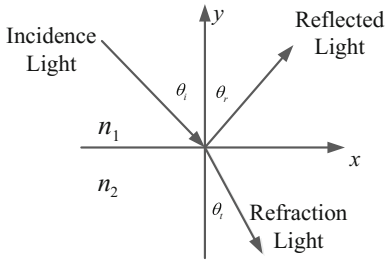


Fig. 30.2 The schematic of the laws of the Fresnel reflectance and refraction

The Fresnel reflectance and refraction equations have forms as Eqs. (30.1)–(30.4).

$$r_p = \frac{n_2 \cos\theta_i - n_1 \cos\theta_t}{n_2 \cos\theta_i + n_1 \cos\theta_t} = \frac{\tan(\theta_i - \theta_t)}{\tan(\theta_i + \theta_t)} \tag{30.1}$$

$$r_s = \frac{n_1 \cos\theta_i - n_2 \cos\theta_t}{n_1 \cos\theta_i + n_2 \cos\theta_t} = \frac{\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)} \tag{30.2}$$

$$t_p = \frac{2n_1 \cos\theta_i}{n_2 \cos\theta_i + n_1 \cos\theta_t} \tag{30.3}$$

$$t_s = \frac{2n_1 \cos\theta_i}{n_1 \cos\theta_i + n_2 \cos\theta_t} \tag{30.4}$$

The index s, p represents s-wave and p-wave, r and t represent the amplitude index of reflectance and refraction separately. n_1 is the index of refraction of the air, and can be set as 1, while n_2 is the index of refraction of the medium, and usually has a complex form, which has an expression as $n_2 = n + ik$. θ_i and θ_r represent the incident angle and the refraction angle. These four equations show us how the light reflected from the surface of the medium or the light refracted into the medium changes its characteristics, particularly its condition of polarization. Generally speaking, the polarization of a light can be described by the degree of the polarization (DoP) and angle of the polarization (AoP). To make use of the Eqs. (30.1)–(30.4), the DoP of the reflectance has an expression as Eq. (30.5) [7, 8].

$$DoP_r = \frac{r_s^2 - r_p^2}{r_s^2 + r_p^2} \tag{30.5}$$

If n_2 is in the complex form, the amplitude index of reflectance and refraction will be calculated by Eqs. (30.6) and (30.7).

$$R_s = \frac{(A - \cos\theta_i)^2 + B^2}{(A + \cos\theta_i)^2 + B^2} \tag{30.6}$$

$$R_p = R_s \cdot \left[\frac{(A - \sin\theta_i \tan\theta_i)^2 + B^2}{(A + \sin\theta_i \tan\theta_i)^2 + B^2}\right] \tag{30.7}$$

Then we can rewrite DoP as Eq. (30.8).

$$DoP = \frac{2A\sin^2(\theta_i)\cos^2(\theta_i)}{A^2\cos^2(\theta_i) + \sin^4(\theta_i) + B^2\cos^2(\theta_i)} \tag{30.8}$$

where A and B are all auxiliary variables, as well as C and D in the Eq. (30.9).

$$\begin{aligned} &A = \sqrt{\frac{\sqrt{C}+D}{2}}; \quad B = \sqrt{\frac{\sqrt{C}-D}{2}} \\ &C = 4n^2k^2 + D^2; \\ &D = n^2 - k^2 - \sin^2(\theta_i). \end{aligned} \tag{30.9}$$

Use these equations, maybe we cannot solve problems of optical image processing, but with them we will make it out that the way to solve them. Optical image processing is a complex and big topic [9, 10]. In this paper we try to use polarization theories introduced before into this area in order to dig up a new road to read the optical images. Following, we will make some simulations and experiments to explain our thoughts.

30.3 Simulations of the Processing

Taking the samples of green paint and aluminum as the examples, their complex indexes of refraction are 1.4 + i0.4 and 1.24 + i6.60, their surfaces are flat like a mirror (though it is hard to realize). Their simulation curves show in Figs. 30.3 and 30.4.

From the figures, we see that, when a light reflected from a no matter the green paint or the aluminum, the index of reflectance, including the index of s-wave, p-wave and the whole wave, increases with the increasing incident angle, while the degree of polarization increases at first until it reaches the top, and then it decreases to zero. However, there are still differences between the green paint and aluminum. Obviously, the index of reflectance of green paint much less than that of aluminum when the incident angle is small, while the degree of polarization shows an opposite phenomenon, as DoP of green paint is much bigger than that of aluminum. The

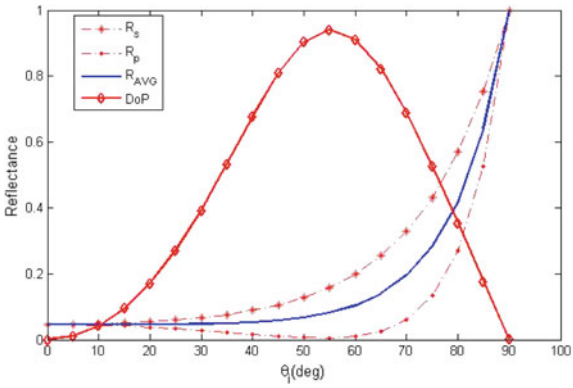


Fig. 30.3 The simulation gauss-curves for green paint surface

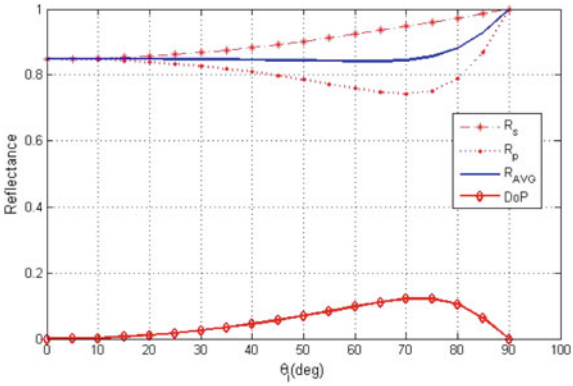


Fig. 30.4 The simulation gauss-curves for Aluminum flat surface

biggest points of DoP are around 0.6 for green paint but less than 0.2 for aluminum. All of these things demonstrate that characteristics of reflectance of different materials, just like green paint and aluminum, are different greatly. Generally, green paint can represent the background or the camouflage, and aluminum represents the targets manmade. If we cannot figure out the target from the background from the intensity images, we can make the images polarized and this way will enlarge the differences between the targets and background which will make it easy to read.

30.4 Experiments Outside and Analysis

We also do some experiments to test and verify our optical image processing methods. We choose some typical samples such as aluminum, iron, wood and a model of a vehicle, and put them on the wild grass. We use a visible light camera with a polarization element in front of the camera to take photos in order to analyze

the polarization characteristics' benefits for target detection. At the same time, we must construct a coaxial system to receive the light from the same direction which is showed as Fig. 30.5.

When we do this experiment, it is a sunshine day. Set the sample on the ground with a constant bias, and the camera system face east and back west, around 7 m away from position of the sample. In this way, the normal direction of the plane of the sample is 30° to the horizontal direction, and the view angle is 60°. Another thing is that the azimuth angle is 180° because of the same plane that the camera and the sun are. We make a measurement once in an hour. The following Fig. 30.6 shows what we get at 9 am and 12 am, that is the original picture, the intensity picture in another words.

In Fig. 30.6, left is the image we get at 9 am, and right is 12 am. At 9 am, the sun cannot incident on the sample directly, so the picture is dark. At 12 am, the sun in the middle of the sky and can incident on the sample directly, so the picture is bright. But there is something fuzzy, and boundary confused. Then from this picture we usually get useful information little. So we make it polarized, that is to say we

Fig. 30.5 The photograph of the experiment's system outside

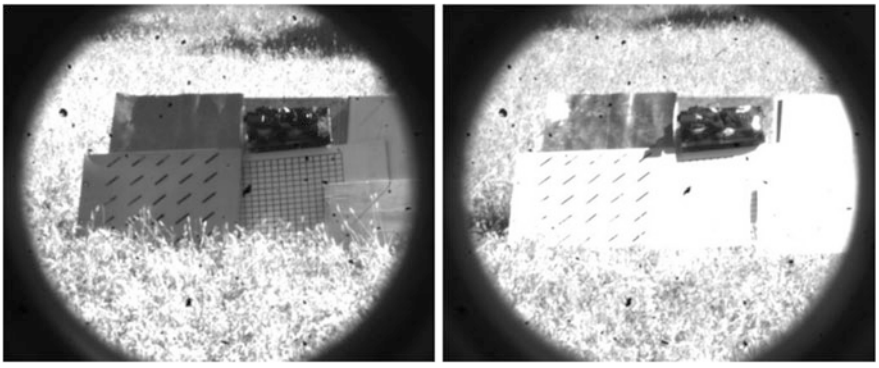


Fig. 30.6 The original image get at 9 am and 12 am

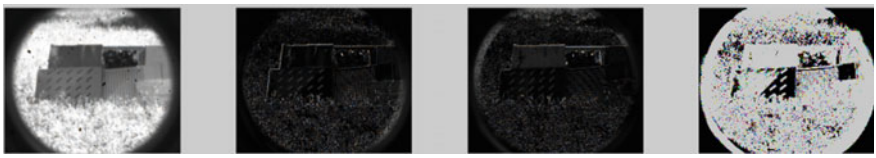


Fig. 30.7 Results of the 9 am image processing. First one is the original picture, second and third are the polarization intensity pictures, and the last is DoP picture

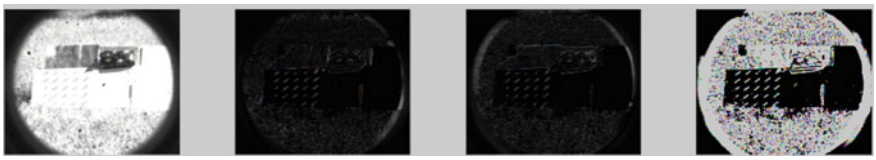


Fig. 30.8 Results of the 12 am image processing. First one is the original picture, second and third are the polarization intensity pictures, and the last is DoP picture

take optical image processing with polarization theory. Figures 30.7 and 30.8 are the results of the processing.

Comparing these pictures, we can see that, the pictures after polarization are usually darker than the original ones, but the boundary of each thing in the same picture is clear, that is the main important point to distinguish the target. Especially in the DoP pictures, things show inverse characteristics compared to the original pictures. For example, the dark things in the original pictures change into bright things in the DoP pictures, while bright change into dark. Usually, targets are hidden in dark place in order to avoid the detection, if we use intensity image, it is hard to discover the target, but now, with polarization optical image processing, we can find out anything in dark easily. That is our thoughts' value in target detection and identification using optical image processing.

30.5 Conclusion

In this paper, we introduce the thoughts that using polarization in optical image processing. First of all, we analyze the use of the polarization of light in target detection and identification with image processing theoretically. Then, we make some simulations of behaviors the light reflected from target materials and background materials, and compare their differences. Finally, we make some outdoor experiments to verify our thoughts and the results achieve that indeed. Later on, we will continue this study, to find out the way to use the polarization method in optical image processing practically.

References

1. Scott Tyo J et al (2006) Review of passive imaging polarimetry for remote sensing application. Appl Opt 45(22):543–5469
2. Shell JR, Dr Salvaggio C, Dr Schott JR (2004) A novel BRDF measurement technique with spatial resolution-dependent spectral variance. IEEE Int Geosci Remote Sens Symp 7:4754–4757
3. Hapke BW (1981) Bidirectional reflectance spectroscopy theory. J Geophys Res 86:3039–3054
4. Zhao YQ, Zhang HC, Pan Q (2003) Object detection through polarization characteristics. J Quant Electron 20(6):685–688
5. Xu SX (2011) Research on the scattered light polarization characteristics analysis for space target detection. Master thesis of Nanjing University of Science & Technology
6. Priest RG, Germer TA (2000) Polarimetric BRDF in the microfacet model: theory and measurements. In: Proceedings of 2000 meeting of the military sensing symposia specialty group on passive sensor, pp 169–181
7. Priest RG, Meier SR (2002) Polarimetric micro-facet scattering theory with applications to absorptive and reflective surface. Opt Eng 41(5):988–993
8. Liu Q, Zhan YH et al (2015) Parameters of polarimetric bidirectional reflectance distribution function of rough surfaces and parameter inversion. J Spacecr TT&C Technol 34(5):481–487
9. Yang D, Zhan YH, Liu Q, Zeng CHE (2014) Estimating the complex index of refraction based on the rough surface polarization degree. In: 1st international conference on photonic and optical engineering in conjunction with annual Western China photonics conference, Xi'an, China
10. Zhan YH, Liu Q, Yang D (2015) Inversion of complex refractive index for rough-surface objects. Opt Precis Eng 23(8):2178–2184

Chapter 31
Analysis of Drift Adjustment by Space Optical Camera Platform

Sanhai Ren, Xiang Fan, Fan Zhang and Zengli Su

31.1 Introduction

In space optical camera, TDI-CCD is predominantly used as the imaging sensor to improve the camera's sensitivity and SNR by multistage time-delay integral. Because of the special operating mode of TDI-CCD, the moving direction and velocity of the CCD array must be identical to that of the target image along the orientation of CCD array accurately. But, there is always an angle between the ground track of the satellite platform and the orientation of the TDI-CCD array due to the earth rotation, which is called the drift angle. The drift angle will cause image motion during the CCD integration, and destroy the motion synchronization between the light-yield electronic charges and the image in the focal plane, resulting the deterioration of the image quality.

There are two main ways in drift adjustment for space optical camera. One way of the camera drift angle compensation is realized by yaw control of the satellite, and the other way is by the camera compensation mechanism. The second one is the commonly used method, which focus on the effect on imaging quality and the compensation method of the drift angle [1–5], or the design and control of drift adjustment mechanism [6, 7], or influencing parameters analysis of estimated errors of space camera's drift angle [8–10]. Whereas, the research on the first method is little, and just stay on the theory analysis [1], no experimental results have been reported.

The drift adjustment by space platform should not be ignored, though it is just the backup means. Taking the universal scroll imaging as example, the cause and the effect on imaging quality of drift angle are analyzed, also the influence of target latitude and scrolling angle on drift angle are analyzed by simulation. In the end, the theory analysis is validated by the flight test.

S. Ren (✉) · X. Fan · F. Zhang · Z. Su
Beijing Institute of Tracking and Telecommunication Technology, Beijing 100094, China
e-mail: sanhairen@163.com

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_31

31.2 The Mechanism and Calculation Method of Drift Angle

31.2.1 The Mechanism of Drift Angle

The geometric relation in scroll imaging is showed in Fig. 31.1. Suppose that the scroll angel is a_s, the ground track is S'_{p_1} when the space optical camera's position is at S_{p_1}, then the ground imaging position will be at S'_{T_1}. When the space optical camera's position is at S_{p_2} after time interval of Δt, the ground track will be S'_{T_2}, ignoring the earth rotation. But, if the earth rotation can't be ignored, though the ground track still is S'_{T_2}, the planning image position will be at S''_{T_2}. Then the drift angle a_d appears, which is between $S'_{T\_1}S'_{T\_2}$ and $S'_{T\_1}S''_{T\_2}$.

31.2.1.1 The Calculation Method of Drift Angle

Based on the analysis above, the velocity of the target includes two parts, considering the earth rotation. One is v_{T1}, which is along the course, and the other is v_{T2}, which is across the course, as shown in Fig. 31.2.

The drift angle a_d can be calculated as follow:

$$a_d = \arctan(v_{T2}/v_{T1}) \tag{31.1}$$

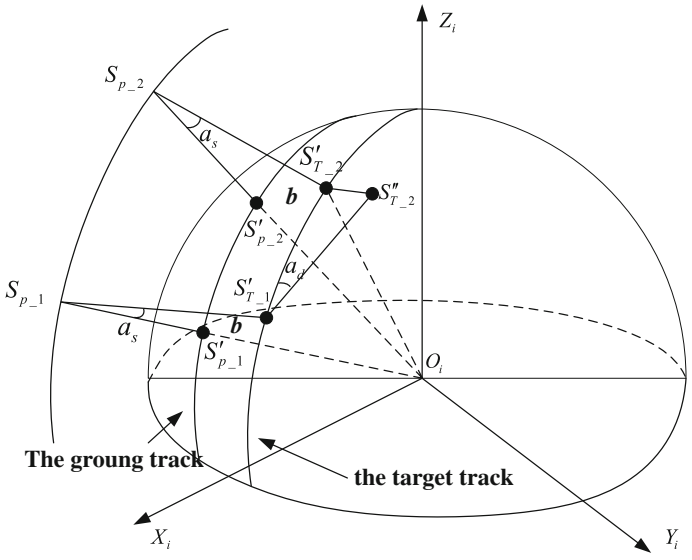


Fig. 31.1 The drift angle in azimuth scroll imaging

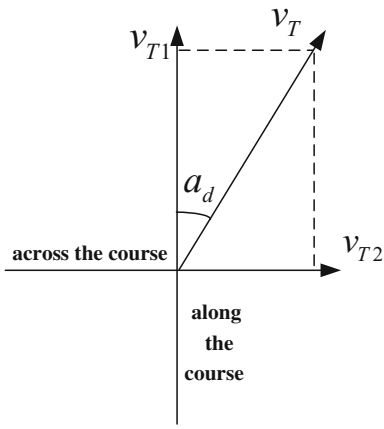


Fig. 31.2 The calculation of drift angle

Thereby, the calculation of drift angle depends on two velocities, which can be calculated as follow [1].

$$V_{T1} = \Omega_s(R_e + h)\cos\ b - \omega_e(R_e + h)\cos\delta_T \cos\ i \tag{31.2}$$

$$V_{T2} = \omega_e(R_e + h)\cos\delta_T \sin\ i \tag{31.3}$$

where $\Omega_s = 360/(T_S \times 60)$ is the angular velocity of the ground track relative to the earth's center, which is equal to the angular velocity along the satellite moving course. T_S is the orbital period. δ_T and h is the latitude and altitude of the ground target T respectively. ω_e, R_e is the spin velocity and radius of the earth. i is the orbit inclination. b is the arc length of $\angle S'_{P\_2}O_iS'_{T\_2}$ based on the earth radius. In the triangle $\Delta S_{P\_2}O_iS'_{T\_2}$, $|O_iS'_{T\_2}| = R_e$, $|O_iS_{P\_2}| = R_s$, then the following relationship can be got by the law of sines.

$$\frac{R_e}{\sin\ a_s} = \frac{R_s}{\sin\angle O_iS'_{T\_2}S_{P\_2}} \tag{31.4}$$

Then, we can get $\angle O_iS'_{T\_2}S_{P\_2}$.

$$\angle O_iS'_{T\_2}S_{P\_2} = \pi - \arcsin\left(\frac{R_s}{R_e}\sin\ a_s\right) \tag{31.5}$$

Thereby, the b can be calculated.

$$b = \angle S'_{P\_2}O_iS'_{T\_2} = \pi - \angle O_iS'_{T\_2}S_{P\_2} - a_s = \arcsin\left(\frac{R_s}{R_e}\sin\ a_s\right) - a_s \tag{31.6}$$

where a_s is the scroll angle, and R_s is the geocentric distance.

Take the formula 31.2–31.4 into formula 31.1, we can get the drift angle.

$$a_d = \arctan\left(\frac{\omega_e \cos\ \delta_T \sin\ i}{\Omega_s \cos\ b - \omega_e \cos\ \delta_T \cos\ i}\right) \tag{31.7}$$

When the scroll angle a_s is zero, the arc length b is zero too, and $\delta_T = \delta_p$, then the formula 31.7 becomes the calculation of drift angle in ground track imaging.

31.3 The Analysis of the Effect of Drift Angle on Imaging Quality

TDI-CCD adopts multiple integral exposure to the same target by multi-level photosensitive cell in turn, and adds the weak signal obtained by every photosensitive cell to a relatively strong signal. Because a time internal exists when different level photosensitive cell images the same target, so the drift angle will bring an image motion to the different photosensitive cell. Then, the image quality will be deteriorated when the signal addition is performed for multi-level photosensitive cell.

From Fig. 31.2 we can get that if there is no drift angle, the target velocity will along the course, and the pixel distance in one integral time is d_{pix}, which equals to the size of integral region. If the drift angle exists, the pixel distance d_{pix} becomes $d_{pix} \cdot \cos a_d$ in one integral time T. So, the image motion along the course can be calculated as follow.

$$\Delta d_1 = d_{pix} - d_{pix} \cdot \cos a_d = d_{pix}(1 - \cos a_d) \tag{31.8}$$

In the same way, we can get the image motion cross the course.

$$\Delta d_2 = d_{pix} \cdot \sin a_d \tag{31.9}$$

A residual drift angle Δa_d will exist in the drift adjustment by platform because of the satellite attitude, especially in yaw, which brings a residual drift adjustment, as follows.

$$\xi_{\Delta d_1} = \frac{\partial \Delta d_1}{\partial a_d} \Delta a_d = d_{pix} \cdot \Delta a_d \cdot \sin a_d \tag{31.10}$$

$$\xi_{\Delta d_2} = \frac{\partial \Delta d_2}{\partial a_d} \Delta a_d = d_{pix} \cdot \Delta a_d \cdot \cos a_d \tag{31.11}$$

Based on the image pixel size, the residual drift adjustment in two directions can be expressed as $\Delta a_d \cdot \sin a_d \cdot S_{CCD}$ and $\Delta a_d \cdot \cos a_d \cdot S_{CCD}$, for the camera whose integral level is S_{CCD}. For $\Delta a_d \ll 1$,$a_d \ll 1$, the effect on image equality of residual drift adjustment cross the course is bigger than that along the course.

31.4 The Flight Test Result Analysis for Platform Drifting Adjustment

31.4.1 The Theoretical Calculation Results of Drift Angle and Analysis

Take a certain satellite as example, its orbit is sun-synchronous, with an orbital period $T_S = 90$ min, orbit inclination $i = 97°$, orbit altitude $H = 270$ km. The theoretical calculation results is shown in Fig. 31.3, including the curve of drift angle changing with respect to target latitude, the curve of drift angle changing with respect to scroll angle, and the curve of the effect on image quality across the course and along the course.

From formula 31.5 we get the changing curve of drift angle when the satellite orbits the earth on a whole round, with the scroll angle being zero. The curve is shown in Fig. 31.3a. From the figure we get that the drift angle reaches the maximum in equator, which is 3.523°, and decreases with latitude increasing. The curve of drift angle changing with respect to scroll angle is shown in Fig. 31.3d–f, from which we get that the maximal variation of drift angle is just 0.003°, when the scroll angle changes from −45° to 45°. The change trend of drift angle is similar at different latitude. So we can get the conclusion that there is no relationship between drift angle change and scroll angle.

The image motion across the course and along the course in one integral period is shown in Fig. 31.3b, c, which is obtained from formulas 31.8 and 31.9, with the CCD integral level being 1. From the figure we know that the effect on image quality across the course is worse than that along the course, and the effect at low latitude is bigger than high latitude.

Based on the theoretical analysis and simulation results, the drift angle is merely related to orbit parameters and ground latitude. The drift angle reaches maximum in equator, at the same time, the effect on image quality is also the biggest, and that across the course is bigger than that along the course.

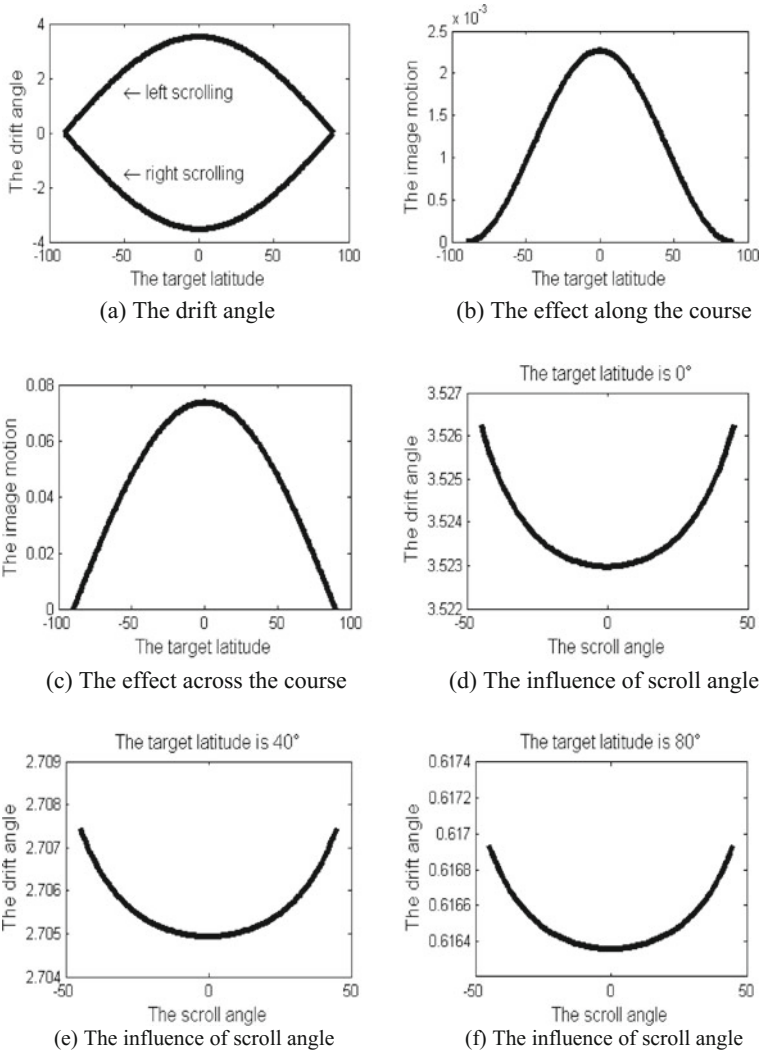
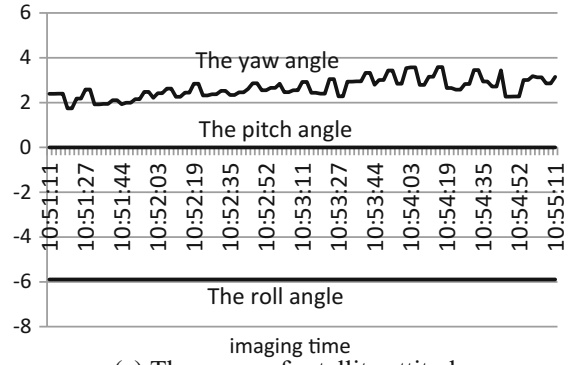


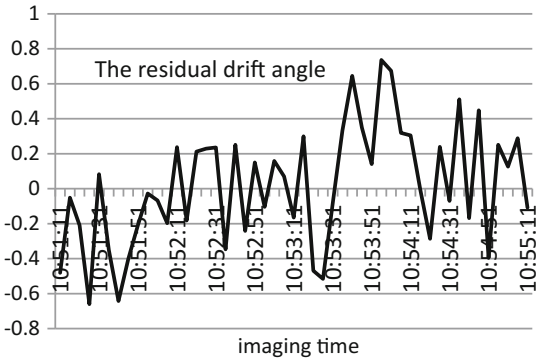
Fig. 31.3 The theoretical analysis results of drift angle

31.4.2 The Result Analysis on Flight Test

In order to validate the effect of drift adjustment by platform, ten flight tests are implemented by a certain satellite on July 18–23, 2014. Taking one flight test as example, the imaging area is Harbin of China, whose coordinate is 45.7400°N and 126.6300°E. The scroll angle is −5.893°, and the solar altitude angle is 63.44°.



(a) The curve of satellite attitude



(b) The residual drift angle

Fig. 31.4 The test results of drift adjustment by platform

The CCD integral level is 16, and imaging time is 10:51:10 am to 10:55:10 am. The result of the flight test is shown in Fig. 31.4, with the flight direction from north to south.

The curves of satellite attitude are shown in Fig. 31.4a, from the top of which are yaw, pitch and roll respectively. From the figure, we know the pitch angle and roll angle are in accordance with theoretical value, where the roll angle is the scroll angle. The drift adjustment by platform amends the drift angle by adjustment the yaw angle. The drift angle goes bigger with time increasing for the satellite flight from north to south, so the curve of yaw angle is smooth upward. From Fig. 31.4a, the curve of yaw angle is dithering, which will bring the residual drift angle, as shown in Fig. 31.4b. The residual drift angle will produce image motion. For the TDI-CCD camera with integral level being 16, the image motion is 0.003 pixel along the course, and 0.2 pixel across the course, calculated from formulas 31.10 and 31.11, which will be worse if the integral level increases.

A piece of image of the flight test is shown in Fig. 31.5, from which we get that the target texture is clear and the boundary is visible, with no image blur. But, the image motion caused by residual drift angle is not acceptable for the high-resolution

Fig. 31.5 The partial image of Harbin

camera, especially when the image is used in target localization or target recognition.

The imaging latitude of the ten fight tests includes −33.857° to 60.794°, and the scroll angle changes from −45° to 30°. From the results we can get that the trend of drift angle changing with scroll angle and target latitude is in accordance with theoretical analysis and simulation, and the yaw angle dithered in every test.

31.4.3 The Analysis of Residual Drift Adjustment by Platform

Based on the principle of drift adjustment by camera and platform, we find that the platform response mechanism is main causation of residual drift angle. For the drift adjustment by camera, the calculated drift angle is sent to the adjustment mechanics directly, and the camera imaging continues after drift adjustment. Whereas, the drift adjustment by platform is different. The calculated drift angle is sent to the satellite platform, and the platform responds to it. Then, the platform adjusts the yaw angle to compensate it. The slow response of platform will cause overshoot. The camera uses a big angle, which might be bigger or smaller than the real value, to do coarse

adjustment firstly, then fine adjustment, inducing the yaw angle vibrates around the real value. In addition, the satellite attitude disturbs when the platform adjusts drift, and the disturbances is collected and sent to platform, which brings a dither to satellite attitude angles.

31.5 Conclusions

The drift adjustment by satellite platform is analyzed in the paper. The calculation method of drift angle and the effect on image quality across the course and along the course is introduced. Moreover, the influence of scroll angle and target latitude on drift angle is also simulated. From the flight test we find that the yaw angle vibrates around the real value, which brings a residual drift angle, deteriorating the image quality. The reason of residual drift adjustment by platform is analyzed, and some improvement are offered to the platform development departments, which are applied to the second generation satellite.

References

1. Yuan XK (2006) Calculation and compensation for the deviant angle of satellite borne TDI-CCD push scan camera. Aerosp Shanghai 06:10–13
2. Fan C, Li YC, Yi HW (2006) Analysis of drift angle on space camera. Infrared Laser Eng 35:216–220
3. Wang JQ, Yu P, Yang CX et al (2005) Space optical remote sensor image motion velocity vector computational modeling, error budget and synthesis. Chin Opt Lett 3(7):414–417
4. Fan C, Liang YT, Li W et al (2008) Study on influence of drift angle on the space camera. Electron Opt Control 15(11):76–79
5. Gu S, Xu Z (2014) The drift angle adjusting method of a new type space camera. Infrared Laser Eng S1:209–213
6. Bao H, Yang LW, Jiang XN et al (2005) Design of draft adjusting mechanism for space optical camera. Opto 39(6):22–28
7. Yang H, Li Z, Xu H et al (2014) Kinematic formula fitting of drift adjusting mechanism for space camera. Opt Precis Eng 12:3342–3347
8. Li WX, Xu SY, Yan DZ (2011) Influencing parameters of estimate errors of space camera′s drift angle. Infrared Laser Eng 40(8):1530–1536
9. Zhong WC (2009) The effect analysis of spacecraft orbit and attitude parameters on optical imaging. Master's thesis of the Harbin institute of technology
10. Wang D, Yan DZ, Wu WP et al (2014) Error analysis and ground verification of high resolution space camera real-time drift angle adjustment. Chin J Sci Instrum 35(9):1955–1962

Chapter 32
The Study on Retrieval Technique of Significant Wave Height Using Airborne GNSS-R

Fei Xu, Xiechang Sun, Xinning Liu and Ruidong Li

32.1 Introduction

Economically developed coastal areas account for 13.6% of Chinese territory and play a leading role in the national economic development. Therefore, the effect of support and protection of marine disaster prevention and mitigation on coastal economic and social development is further demonstrated. The sea parameters is an important basis for the tsunami/typhoon warning, meteorological and weather forecast and climate change. Significant wave height, sea surface wind field and sea surface height are typical parameters of sea state, current space-based detection method of these parameters are mostly only for subastral point observations and thus their spatial coverage is limited. Active detection mode is often used, but it has problems on high-cost, heavy weight and high power consumption, so methods with active mode can not maintain good performance in long-term continuous work [1–4].

A multi-source and multi-destination GNSS-R system for marine environment remote sensing, which is independently developed, can receive reflected GNSS signal of sea surface and retrieve significant wave height and other parameters of sea state. The system has the characteristics of abundant signal source, wide coverage, low power consumption and so on, so small satellite constellation networking observation using this system can be easily realized [5–7]. Figure 32.1 depicts technical principle of multi-source and multi-destination GNSS-R technique for maritime remote sensing, shows detection structure formed by multiple

F. Xu (✉) · X. Sun · X. Liu · R. Li
Space Star Technology Co., Ltd, 100086 Beijing, China
e-mail: promisefy@126.com

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_32

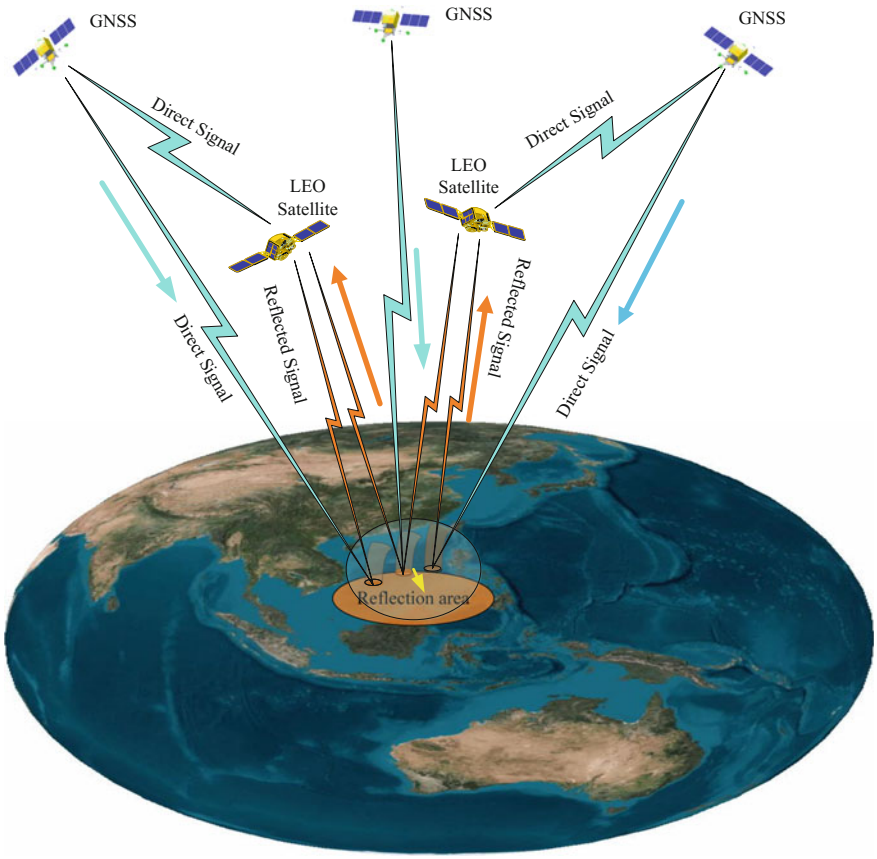


Fig. 32.1 Schematic diagram of technical principle of multi-source and multi-destination GNSS-R technique for maritime remote sensing

navigation satellites and multiple LEO (low earth orbit) satellites equipped with receiver. Through the receiver on the LEO satellite, GNSS-R signal of observation area is received and processed, high-precision retrieval of significant wave height with high spatial and temporal resolution is implemented [6–8]. Three airborne flight tests using the system were organized in specific area of Bohai Bay, Tianjin. GNSS and GNSS-R signal data collected in the flight tests were used for inversion calculation of significant wave height, the calculated results are compared with the theoretical calculation, and the feasibility of the principle is verified [9, 10].

32.2 Significant Wave Height Retrieval Method

32.2.1 *Collection and Storage Technology of Massive GNSS-R Data*

The experimental data used in this paper is collected by our independently developed marine environment remote sensing system. The receiving antenna of the reflected signal used in the system is a GNSS-R multi-beam digital array antenna, which is designed with 31 elements. The reflected signal received by each element of the antenna is amplified, down converted, and sampling. The sampling rate is 56.8 MHz, and the data width is 8 bit. In order to save completely the raw data of reflected signal received by 31 elements, storage rate must be higher than 1.8 GB/s, and there should be sufficient storage space, such that the data can uninterruptedly stored and data files can be formed throughout the flight test process. The system realizes the function of file management and data transfer. The design of the interface for the high-speed and large-scale data acquisition card that meet above requirements is shown in Fig. 32.2.

The data acquisition card performs real-time acquisition and packaged storage of large amount of data which received by the system. The data is transmitted through ZD high speed connector and backplane to 15 SSD (mSATA) disks in storage module. Each time the data is collected to form a data file, marked with the

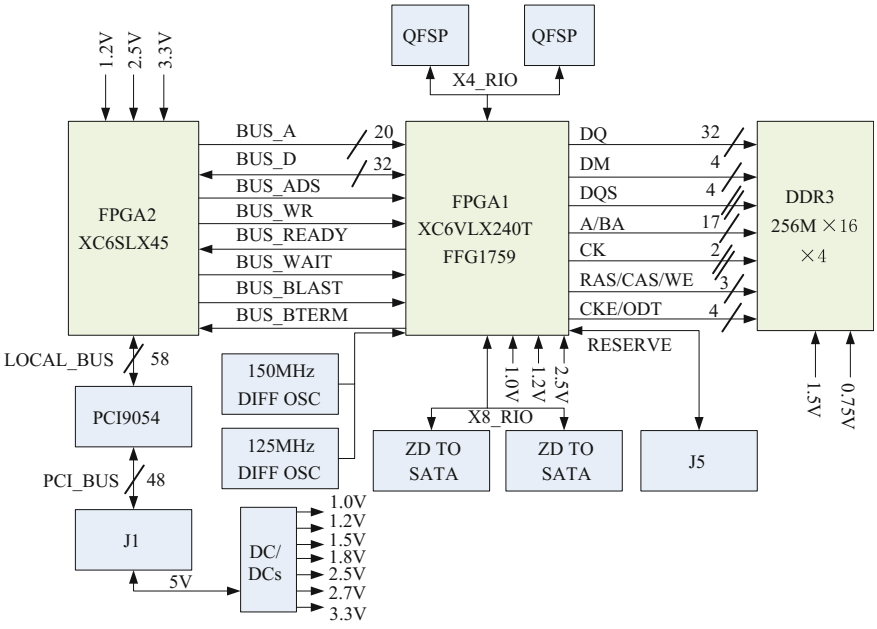


Fig. 32.2 Interface for high-speed and large-scale data acquisition card

collection time, file size and other information. After the acquisition, the memory bank can be taken out and connected to suitable ground equipment for data transfer and storage. A single memory module consists of 15 SSD (mSATA) disks, each disk capacity is 480 GB, the total capacity is more than 7 TB; Supported acquisition rate of a storage module can reach more than 2.4 GB/s, which can meet real-time data acquisition and storage requirements of the system.

32.2.2 *Significant Wave Height Retrieval Method Using GNSS-R*

In this paper, a lot of GNSS-R raw data obtained in three airborne flight tests is processed and reflected signal of sea surface is successfully captured. The reflected signal carries the characteristics of sea surface, which can be described by waveform, polarization, amplitude, phase, frequency and other features. Therefore, the physical characteristics of the reflecting surface can be obtained by receiving and high-precision processing of the reflected signal. Specific approach is to generate DDM imagery from GNSS-R data, then calculate significant wave height using DCF function. Figure 32.3a is a schematic diagram of significant wave height measurement, Fig. 32.3b shows the flow chart of significant wave height retrieval.

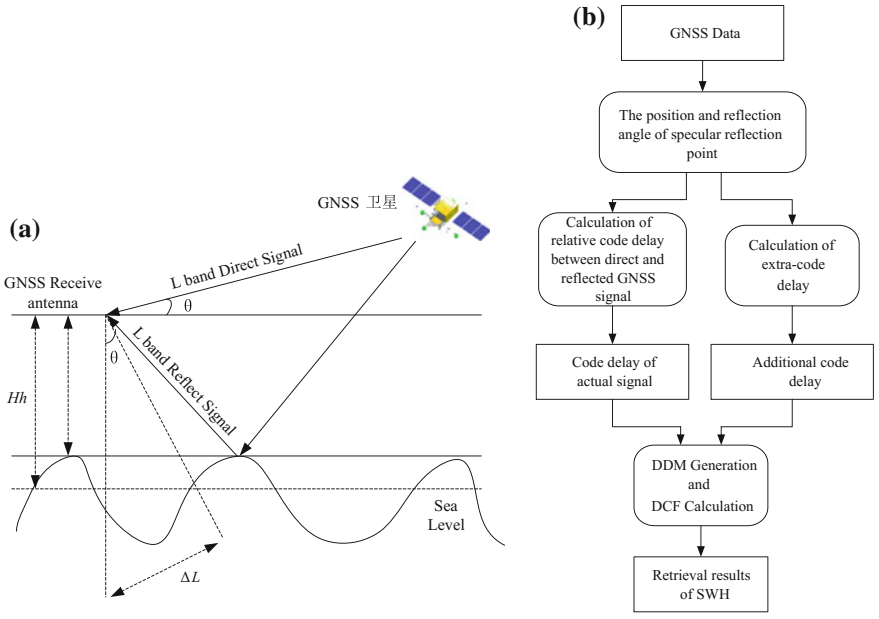


Fig. 32.3 Process of SWH measurement and retrieval. **a** Schematic diagram of SWH measurement. **b** Flow chart of SWH retrieval

As shown in Fig. 32.3b, the specific retrieval process is as follows:

1. Measured data, including L1 carrier data and pseudo code data, is read and correlated waveform of the direct wave and reflected wave is calculated. GPS satellites' position and speed information is obtained.
2. According to observation time or the position and velocity information of the receivers and the satellites, the position and the reflection angle of the specular reflection point are calculated.
3. Single frequency correction of additional delay caused by ionosphere is performed using local carrier and code data.
4. Correlation operation of corrected data is done and three-dimensional DDM imagery is generated using coherent and non-coherent integration method.
5. Significant wave height is calculated as DCF waveform width based on the DDM imagery.

32.3 Experimental Data Processing

32.3.1 DDM Imagery Generation

Figure 32.4a is equipment connection diagram of the system used in three flight tests in Bohai bay, Fig. 32.4b is standby state diagram of the installed equipments. Figure 32.5a shows the actual distribution of GPS satellites in one of operation areas, Fig. 32.5b depicts actual flight path in the operation area of one of flight tests, and it can be seen that there are 8 GPS satellites are visible in the operating area at the same time.

Dextrorotary direct signal stored by data acquisition card and 31 channel laevorotatory reflected signal are processed, dextrorotary direct signal is configuration processed with direct channel correlator. Positioning and specular reflection point calculation is completed through signal capturing and tracking by code loop. By controlling chip step size, correlation processing is performed between direct signal with different Doppler frequency shift and reflected signal from reflection channel, and three-dimensional DDM imagery is generated using coherent and non-coherent integration method and output accumulation method. DDM imagery generation principle is shown in Fig. 32.6.

In the multi-source and multi-destination GNSS-R system for marine environment remote sensing, signal receive platforms are apparently different with signal transmit platforms, the system is a typical multi-based (dual based) remote sensing detection system, the physical quantities in echo model for imagery generation are shown in Fig. 32.7.

In the model, t represents time, $P_T(t)$ represents trajectory of phase center of transmit antenna, $P_R(t)$ represents trajectory of phase center of receiving antenna, $\boldsymbol{r}$ represents the position vector of the scattering point P to the origin of the

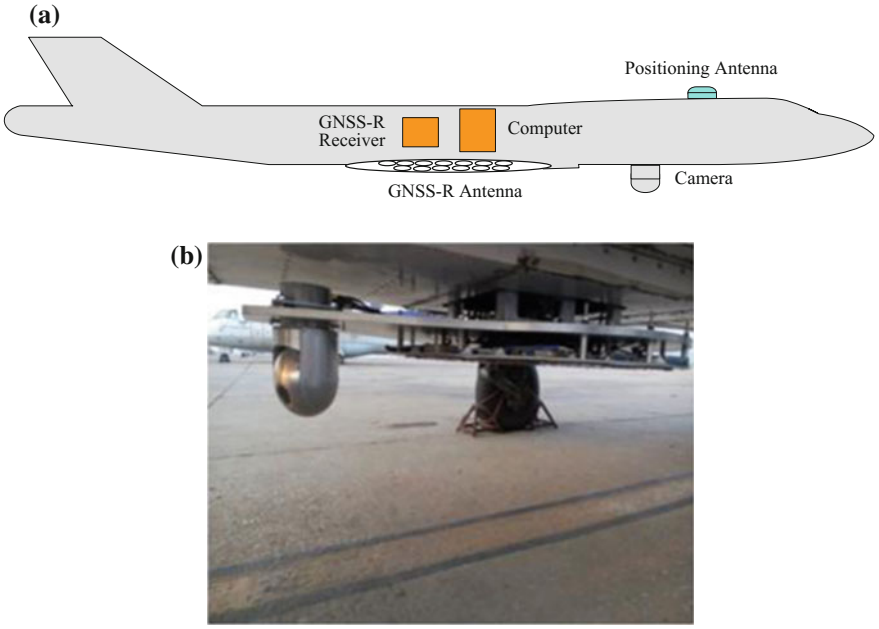


Fig. 32.4 Equipment connection and to be working. **a** Equipment connection. **b** Equipment to be working

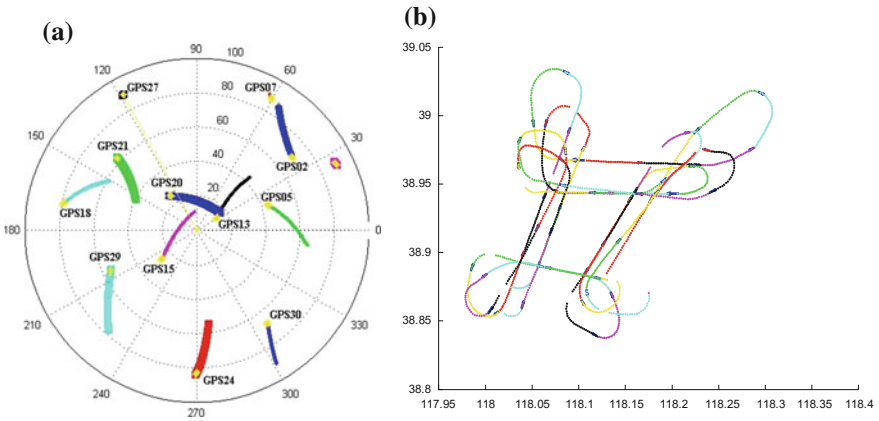


Fig. 32.5 Visible satellites in operation area and flight path. **a** Visible GPS satellites. **b** Actual flight path

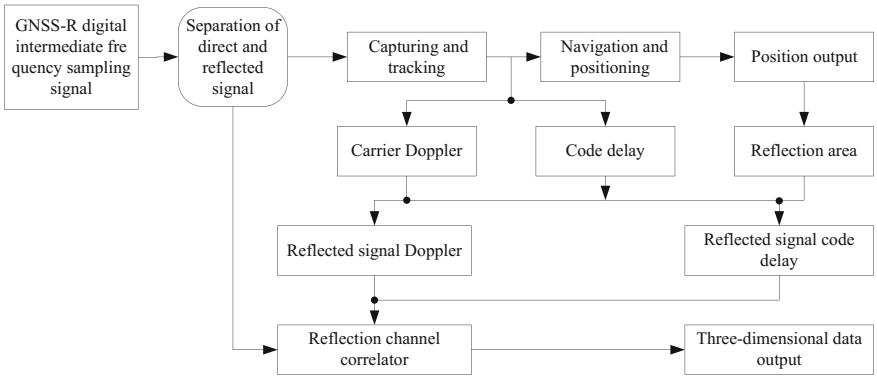


Fig. 32.6 DDM imagery generation principle

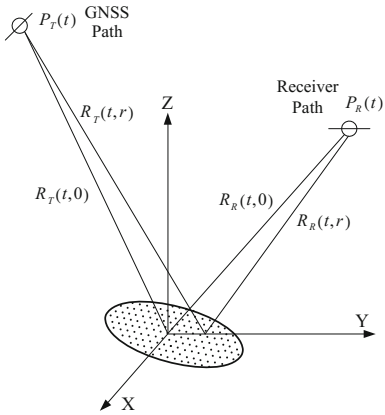


Fig. 32.7 Echo model for imagery generation

coordinate, $R_T(t,\boldsymbol{r})$ represents distance from the transmitting antenna to the scattering point P, $R_R(t,\boldsymbol{r})$ represents the distance from the receiving antenna to the scattering point P.

For a given base band signal $f(t)$, two-dimensional echo signal $d_t(t,\boldsymbol{r})$ whose carrier frequency is removed of scattering point can be described by

$$d_t(t;\boldsymbol{r}) = w_T(\boldsymbol{r}) \cdot w_R(\boldsymbol{r}) \cdot \sigma \cdot \exp\left(j \cdot K_0 \cdot R(t,\boldsymbol{r})\right) \cdot f\left(t - \frac{R(t;\boldsymbol{r})}{c}\right) \tag{32.1}$$

in which, σ represents radar cross section (RCS) of scattering point P, $w_T(\boldsymbol{r})$ and $w_R(\boldsymbol{r})$ represents amplitudes of antenna pattern of receiving antenna and transmitting antenna respectively in the direction of target area, K_0 represents wave number, c is the speed of light, $R(t,\boldsymbol{r})$ represents slant range of the scattering point P, which is the sum of two distances, one of which is from the scattering point P to GNSS satellite, the other is from the scattering point P to the receiving platform:

$$R(t;\boldsymbol{r}) = R_T(t;\boldsymbol{r}) + R_R(t;\boldsymbol{r}) \tag{32.2}$$

Two-dimensional echo signal $d_t(t,\boldsymbol{r})$ of scattering point P is compressed using matched filter along range direction, range compressed single-point two-dimensional echo signal $d_{II}(r_d,t,\boldsymbol{r})$ can be obtained. And then to a observation area Ω, range compressed echo $d_{II}(r_d,t)$ is the integral of single-point two-dimensional echo $d_{II}(r_d,t,\boldsymbol{r})$ in the area. In order for easy analysis, discrete approximation of the observed area Ω is performed and approximate sum can be obtained:

$$d_{II}(r_d,t) \approx \sum_{P\in\Omega} \sigma \cdot d_{II}(r_d,t;\bar{r}) \tag{32.3}$$

Acquisition of three-dimensional DDM imagery can be regarded as distribution estimation of the scattering coefficient of each scattering point utilizing difference between echoes of scattering points in the space. The time delay and Doppler frequency shift of direct signal is set as coordinate origin, the time delay (τ_0) and Doppler frequency shift (f_0) at the specular reflection point is set as the reference point of the delay–doppler window. Three-dimensional DDM imagery was generated by two dimensional correlation of the reflected signals, which is defined by:

$$\begin{aligned} &DDM_k(\tau,f) \\ &= \sum_{n=(k-1)T_i f_s}^{kT_i f_s} u_R(nT_s+\tau+\tau_0)\cdot a(nT_s+\tau_0)\cdot \exp[2\pi j(f_R+f_0)(nT_s+\tau_0)] \end{aligned} \tag{32.4}$$

In a flight test GNSS-R data is received and analyzed, the system receives the sea surface reflection signal from No. 29 GPS satellite and reflected by the specular reflection point. Three-dimensional DDM imagery which is generated using the above method is shown in Fig. 32.8a–c.

32.3.2 SWH Calculation Using DCF

Delay–doppler two-dimensional correlation of reflected signal is needed for generating three-dimensional DDM imagery. The presence of significant wave height make sea surface roughness change, which cause diffuse reflection of electromagnetic wave. Different peak values of delay correlation result is superimposed on the peak value of the specular reflection point's correlation result, the maximum of reflection correlation peak values correspondingly shifted, the shape of the correlation peak also changed, which means the slope of echo power varies with the distribution of significant wave height. Thus significant wave height and other parameters of sea surface can be retrieved from further operation of correlation function of DDM imagery using DCF.

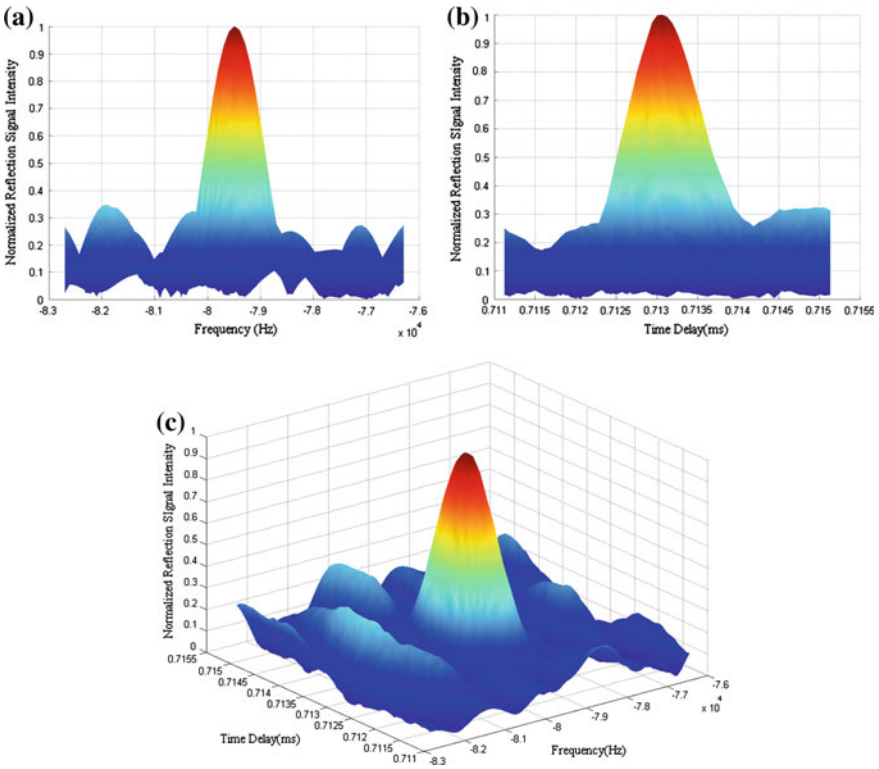


Fig. 32.8 Three-dimensional DDM imagery generated using reflected signal data from No. 29 GPS satellite. **a** Two-dimensional doppler imagery. **b** Two-dimensional time delay imagery. **c** Three-dimensional DDM imagery

In the Sect. 32.3.1, Two-dimensional correlation function of reflected signal is

$$R_P(\tau_m) = \sum_k a_k \exp(i(\varphi_m - \varphi_k))\Lambda(\tau_m - \tau_k) \tag{32.5}$$

in which τ_m represents delay of reflected signal. Assume that reflection elements are non-coherent, their cross-correlation value is 0, then

$$\begin{aligned}\left\langle R_P R_P^* \right\rangle(\tau_m) &= \sum_k \left(\frac{P_t G_t}{4\pi D_k^2}\right)\left(\frac{\sigma_{0k} A_k}{4\pi d_k^2}\right)\left(\frac{\lambda^2 G_r}{4\pi}\right)\Lambda(\tau_m - \tau_k) \\ &= \left(\frac{P_t G_t}{4\pi D_k^2}\right)\left(\frac{\sigma_{0k} A_k}{4\pi d_k^2}\right)\left(\frac{\lambda^2 G_r}{4\pi}\right)\sigma_0 \iint \Lambda^2(\tau_m - \tau_k(x,y))dxdy\end{aligned} \tag{32.6}$$

in which x, y represent the horizontal and vertical coordinates of sea level respectively. The relationship between reflected signal correlation function and the reflecting surface can be obtained by the above formula, which is described by the DCF as following:

$$\Lambda_R^2 = \left(\frac{P_t G_t}{4\pi D_k^2}\right)\left(\frac{\sigma_{0k} A_k}{4\pi d_k^2}\right)\left(\frac{\lambda^2 G_r}{4\pi}\right) \int_{-\infty}^{+\infty} f(z)\Lambda^2\left(\tau_m - \tau(x,y,0) + 2\frac{z}{c}\sin(\varepsilon)\right)dz \tag{32.7}$$

Sea surface is considered as a rough surface, the above formula contains the information of roughness and tilt effect information of sea surface. When σ_0 is very small, Λ_R is close to the ideal situation. With the increase of σ_0, the peak value of Λ_R decreases and the correlation time increases. The sea surface tilt effect makes ΛR to be asymmetrical, which is caused by electromagnetic deviation.

The test results and results from buoy and other auxiliary means are compared and analyzed. Through the analysis, the wind speed of sea surface can be obtained by calculation of the height of the DCF, the average height of sea surface can be obtained by calculation of the code delay of DCF peak value, significant wave height can be obtained by calculation of the DCF peak width.

32.4 Conclusions

The system independently developed is used in airborne flight test to collect raw data of reflected signal of sea surface, and then three-dimensional DDM imagery is generated, SWH is retrieved based on DDM map using DCF function. Feasibility and reliability of the retrieval method is qualitatively analyzed, which can support quantitative calculation and verification of SWH precision.

The multi-source and multi-destination GNSS-R system for ocean remote sensing has real-time performance, and also has high spatial and temporal resolution due to the use of multi beam digital array antenna (beam scanning). The system can realize the spatial resolution of 25 km by processing reflected signal data of a single satellite. Compared with radar altimeter's 50 km spatial resolution, the spatial resolution is increased to two times, at the same time measuring efficiency is improved by more than 20 times. The technology can be one of marine environment parameter detection methods with good performance, and serves for national marine strategy and the promoted application of Beidou satellite navigation and positioning system.

References

1. D'Addio S, Martin-Neira M (2014) GNSS-R altimeter based on doppler multi-looking. IEEE J Sel Top Appl Earth Obs Remote Sens 7(5):1452–1460
2. Martin-Neira M, Caparrini M, Font-Rossello J (2001) The PARIS concept: an experimental demonstration of sea surface altimetry using GPS reflected signals. IEEE Trans Geosci Remote Sens 39(1):142–150
3. Xuesong ZHU, Songhua YAN, Xunxie ZHANG et al (2012) Sea wave height retrieving system based on GNSS—R carrier phase. Chin J Radio Sci 27(6):1070–1074
4. Schiavulli Domenico, Nunziata Ferdinando (2014) Reconstruction of the normalized radar cross section field from GNSS-R delay-doppler map. IEEE J Sel Top Appl Earth Obs Remote Sens 7(5):1573–1583
5. Zhang X, Zhang D, Hu X et al (2004) Remote sensing of the global oceanic state using GPS-reflected signals. GNSS World China 29(5):2–9
6. Rius A, Oliveras S (2012) Altimetry with GNSS-R interferometry: first proof of concept experiment. Orig Article GPS Solut 16:231–241
7. Clarizia MP (2012) Investigating the effect of ocean waves on GNSS-R microwave remote sensing measurements. University of Southampton
8. Shao L, Zhang X, Wang X et al (2008) Sea surface wave height retrieve using GNSS-R signals. Geomat Inf Sci Wuhan Univ 33(5):475–478
9. Gleason S, Hodgart S, Sun Y et al (2005) Detection and processing of bistatically reflected GPS signals from low earth orbit for the purpose of ocean remote sensing. IEEE Geosci Remote Sens 43(6):1229–1241
10. Rius A, Cardellach E, Martı´n-Neira M (2010) Altimetric analysis of the sea-surface GPS-reflected signals. IEEE Trans Geosci Remote Sens 48(4):2119–2127

Chapter 33
The Calibration Method of Channel Consistency of Distributed Digital Phased Array

Na Wang, Xinshi Hu, Tian Yuan and Wentao Zhou

33.1 Introduction

As the rapid development of the array signal theory and digital integrated circuit technology, adaptive digital beam forming technology has been applied in the phased array system [1, 2]. The traditional phased array system, which uses the attenuator and phase shifter to control beam, has been replaced by the adaptive digital beam forming that weights baseband signal with digital technology and controls the gain and shape of antenna pattern. For distributed digital phased array, in order to achieve the precise synthesis of the beam, it is required that gain and time delay of radio frequency transceiver modules must be the same. So, the synchronization between each array element has been a key problem.

The digital phased array characteristics of radio frequency transceiver modules are not the same and impossible to remain the same. Performance differences between radio frequency transceiver chips, temperature changes, etc., can lead to various gain and time delay of radio frequency transceiver modules, causing the distortion map in the direction of the beam. The clocks and the synchronous signals in distributed digital phased array need to pass long transmission lines. As the wiring paths of the transmission lines are different, the errors of the signal delay on the transmission lines directly lead to the phase errors of radio frequency transceiver modules. Therefore, for distributed digital phased array, because of the phase error between channels usually beyond the compensation scope of equalizer, some traditional calibration methods of phased array channel, including time domain equalization algorithm and frequency domain equalization algorithm [3, 4], have not been effectively implemented. Based on the demand of engineering practice, the paper proposes a new solution. Firstly, the time delay estimation algorithm based on

N. Wang (✉) · X. Hu · T. Yuan · W. Zhou
The 10th Research Institute of China Electronics Technology Group Corporation (CETC), Chengdu 610036, China
e-mail: wangna_cetc@163.com

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_33

temporal coherence is used to align the large time delay among channels, within the scope of compensation. Then it is achieved consistency of magnitude and phase among channels according to use the frequency domain equalization algorithm based on the least squares fitting. It has been carried on the detailed simulation analysis using MATLAB platform.

33.2 The Time Delay Estimation Algorithm Based on Temporal Coherence

The time delay estimation algorithm based on temporal coherence uses a known linear frequency modulation signal, which is conjugated with the emission signal, to multiply the output signal of T/R component. After multiplying, the data contains information of time delay [5], and then it is used to correct the difference of large time delay between each channel.

Assume the sending complex linear frequency modulation signal is shown in formula (33.1) [6].

$$s(t) = e^{j(2\pi f_0 t + \pi\mu t^2)} \tag{33.1}$$

f_0 is starting frequency, μ is modulation slope. Ignoring the channel to the influence of the signal magnitude, the channel output signal is as follows:

$$y(t) = e^{j\left[2\pi f_0(t-\tau) + \pi\mu(t-\tau)^2\right]} \tag{33.2}$$

τ is channel delay.

Then, the channel output signal is multiplied by the conjugated sending signal, as shown in formula (33.3).

$$\begin{aligned} m(t) &= s^*(t) \times y(t) = e^{-j(2\pi f_0 t + \pi\mu t^2)} \times e^{j[2\pi f_0(t-\tau) + \pi\mu(t-\tau)^2]} \\ &= e^{j[-2\pi f_0 t - \pi\mu t^2 + 2\pi f_0(t-\tau) + \pi\mu(t-\tau)^2]} = e^{j(-2\pi\mu\tau t - 2\pi f_0\tau + \pi\mu\tau^2)} \end{aligned} \tag{33.3}$$

Above result contains the information needed for the delay. The time delay estimated is shown in formula (33.4).

$$\hat{\tau} = \frac{\hat{w}}{-2\pi\mu} \tag{33.4}$$

$\hat{w}$ is the angular frequency estimated using fast Fourier transform.

33.3 The Frequency Domain Equalization Algorithm Based on the Least Squares Fitting

Time delay estimation algorithm based on temporal coherence between each channel achieves the large time delay estimation, beneficially to control the fluctuation within the scope of the equalizer to compensate. The next is to use equalizer to calibrate magnitude and phase consistency of each channel. The frequency domain equalization algorithm based on the least squares fitting is adopted in this paper [7].

The principle of frequency domain equalization is as shown in Fig. 33.1. Assume $C_i(w)$ is the frequency response of mismatch channel, $H_i(w)$ is the frequency response of expected equalizer and $B_i(w)$ is the frequency response of equilibrium. The relations are as follows [8, 9]:

$$\begin{aligned} B_i(w) &= C_i(w)H_i(w), \quad i = 1, 2, \ldots, N \\ H_i(w) &= \frac{C_{\text{ref}}(w)}{C_i(w)} H_{\text{ref}}(w) = \frac{B_{\text{ref}}(w)}{C_i(w)}, \quad i = 1, 2, \ldots, N \end{aligned} \tag{33.5}$$

$$\begin{aligned} D &= (L-1)T/2 \\ H_{\text{ref}}(w) &= e^{-jwD} \end{aligned}$$

$C_{\text{ref}}(w)$ is the frequency response of reference channel, $H_{\text{ref}}(w)$ is the frequency response of linear phase all-pass network which is used to ensure that the time delay of reference channel is the same as the time delay of equalized channel, D is the value of the same time delay, L is the length of the filter.

The actual frequency response of the channel $C_i(w)$ can be achieved according to work points into the same linear frequency modulation signal by power division network. Then $H_i(w)$ can be calculated through formula (33.5).

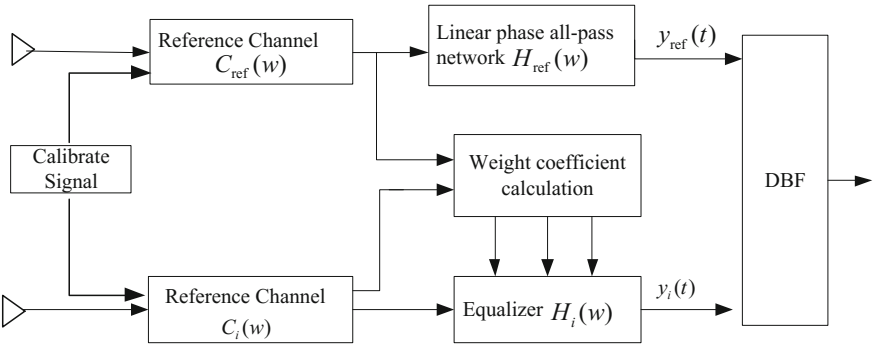


Fig. 33.1 The diagram of equalization in frequency domain

FIR filter is used to fit the expect equalizer. The frequency response of filter is as shown in formula (33.6).

$$\begin{aligned} E_i(w) &= \sum_{l=0}^{L-1} h_i(l)\mathrm{e}^{-jwlT} = \boldsymbol{a}^{\mathrm{T}}(w)\boldsymbol{h} \\ \boldsymbol{a}(w) &= [1, \mathrm{e}^{-jwT}, \ldots, \mathrm{e}^{-jw(N-1)T}]^{\mathrm{T}} \\ \boldsymbol{h} &= [h_i(0), h_i(1), \ldots, h_i(L-1)]^{\mathrm{T}} \end{aligned} \tag{33.6}$$

$\boldsymbol{a}(w)$ is the phase shift vector, T is the unit time delay of FIR filter, L is the filter order, $\boldsymbol{h}$ is the filter weight vector.

Assume the frequency response of expected equalizer $H_i(w)$ is calculated according to the M points of discrete frequencies. The frequency response of equalization filter is shown in formula (33.7).

$$H_i(m) = \frac{C_{\mathrm{ref}}(m)}{C_i(m)} H_{\mathrm{ref}}(m), \quad m = 0, 1, \ldots, M-1; \quad i = 1, 2, \ldots, N \tag{33.7}$$

The value of $\frac{C_{\mathrm{ref}}(m)}{C_i(m)}$ can be achieved by requiring the ratio between discrete Fourier transform of reference channel signal and discrete Fourier transform of equalized channel signal, according to use the same calibration signals. The frequency response of FIR filter $E_i(m)$ is discrete as follows:

$$\begin{aligned} E_i(m) &= \sum_{k=0}^{L-1} h_i(l)\mathrm{e}^{-j\frac{2\pi ml}{M}} = \boldsymbol{a}^{\mathrm{T}}(m)\boldsymbol{h}_i, \quad m = 0, 1, \ldots, M-1; \quad i = 1, 2, \ldots, N \\ \boldsymbol{a}(w) &= \left[1, \exp\left(-j\frac{2\pi m}{M}\right), \ldots, \exp\left(-j\frac{2\pi m(L-1)}{M}\right)\right]^{\mathrm{T}}, \quad m = 0, 1, \ldots, M-1 \end{aligned} \tag{33.8}$$

The least squares fitting is used to approximate $E_i(m)$ to $H_i(m)$, as shown in formula (33.9).

$$\begin{aligned} e_i(0) &= H_i(0) - \boldsymbol{a}^{\mathrm{T}}(0)\boldsymbol{h}_i \\ e_i(1) &= H_i(1) - \boldsymbol{a}^{\mathrm{T}}(1)\boldsymbol{h}_i \\ &\ldots\ldots \\ e_i(M-1) &= H_i(M-1) - \boldsymbol{a}^{\mathrm{T}}(M-1)\boldsymbol{h}_i \end{aligned} \tag{33.9}$$

Namely, it is as follows:

$$\begin{aligned} \boldsymbol{e}_i &= \boldsymbol{H}_i - \boldsymbol{A}\boldsymbol{h}_i \\ \boldsymbol{e}_i &= [e_i(0), e_i(1), \ldots, e_i(M-1)]^{\mathrm{T}} \\ \boldsymbol{H}_i &= [H_i(0), H_i(1), \ldots, H_i(M-1)]^{\mathrm{T}} \\ \boldsymbol{A} &= \begin{bmatrix} a_{0,0} & a_{0,1} & \cdots & a_{0,N-1} \\ a_{1,0} & a_{1,1} & \cdots & a_{1,N-1} \\ \cdots & \cdots & \cdots & a_{0,N-1} \\ a_{M-1,0} & a_{M-1,1} & \cdots & a_{M-1,N-1} \end{bmatrix} \\ a_{m,n} &= \exp\left[-j\frac{2\pi(m-1)(n-1)}{M}\right] \end{aligned} \tag{33.10}$$

The weight vector $\boldsymbol{h}_i$ of equalization filter can be optimized as follows:

$$\min_{\boldsymbol{h}_i} \sum_{m=0}^{M-1} |e_i(m)|^2 = \min_{\boldsymbol{h}_i} \sum_{m=0}^{M-1} \left|H_i(m) - \boldsymbol{a}^{\mathrm{T}}(m)\boldsymbol{h}_i\right|^2 = \min_{\boldsymbol{h}_i}(\|\boldsymbol{H}_i - \boldsymbol{A}\boldsymbol{h}_i\|^2) \tag{33.11}$$

It is a classical issue of least squares optimization to solve $\boldsymbol{h}_i$, as shown in formula (33.12).

$$\boldsymbol{h}_i = (\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{H}_i \tag{33.12}$$

In order to obtain better equalization effect, it is usually to modify the basic algorithm. One approach is to equalize the required frequency band, and another approach is to use weighted least squares fitting method, such as the reference channel magnitude response used to be the weighted matrix of the diagonal elements [10]. The paper has adopted the first approach.

In order to achieve the calibration of the channel consistency, it is commonly used to select a channel as a reference channel. In addition to the reference channel, the equalizer is joined in other each channel. Through the calculation of equilibrium coefficient of various equalizer, it is achieved to make the frequency response function of each channel consistent with the reference channel.

33.4 The Simulation and Analysis of Algorithm

33.4.1 The Simulation Model

The frequency response of FIR filter is used to simulate the receiving channel. As shown in Fig. 33.2, T is the sampling interval, and $b_i(i = 0, 1, 2, \ldots, M)$ is the ideal weights which are the channel impulse response without distortion.

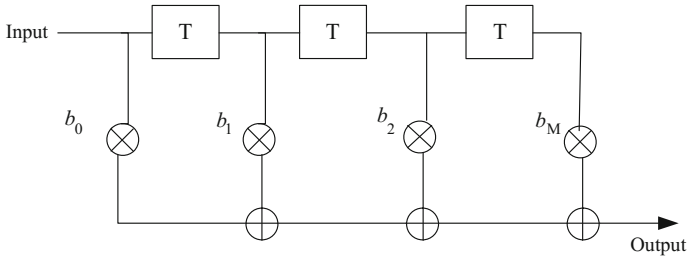


Fig. 33.2 The model of receiving channel simulated by FIR filter

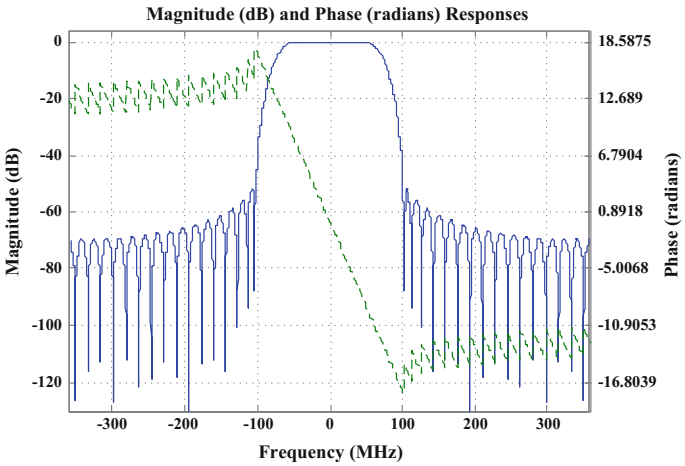


Fig. 33.3 The magnitude and phase response of reference channel

The frequency response of ideal filter is as follows:

$$H_0(w) = \sum_{i=0}^{M} b_i \exp(-jwiT) \qquad (33.13)$$
$$w = 2\pi f$$

w is the angular frequency.

The ideal 40 order FIR filter is used to simulate the reference channel. The frequency responses are shown in Figs. 33.3 and 33.4. The balanced band is −50 to +50 MHz. The magnitude and group delay in pass band are 0 dB and 27.78 ns respectively.

The twenty order complex coefficient FIR filter is used to simulate the mismatch channel. The frequency responses are shown in Figs. 33.5 and 33.6. Because the balanced band is −50 to +50 MHz, the fluctuations of magnitude and group delay in pass band and are −1 to +0.7 dB and 8.4–15.9 ns respectively.

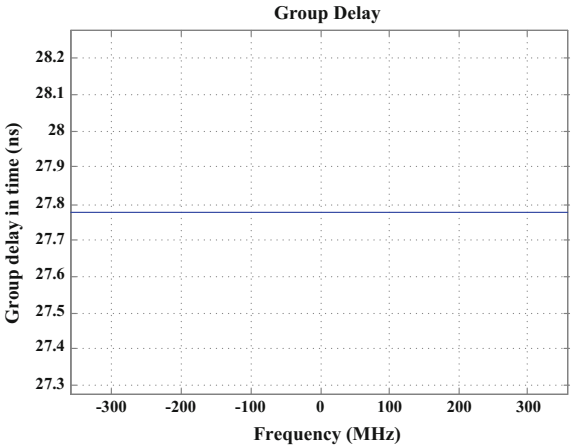


Fig. 33.4 The group delay of reference channel

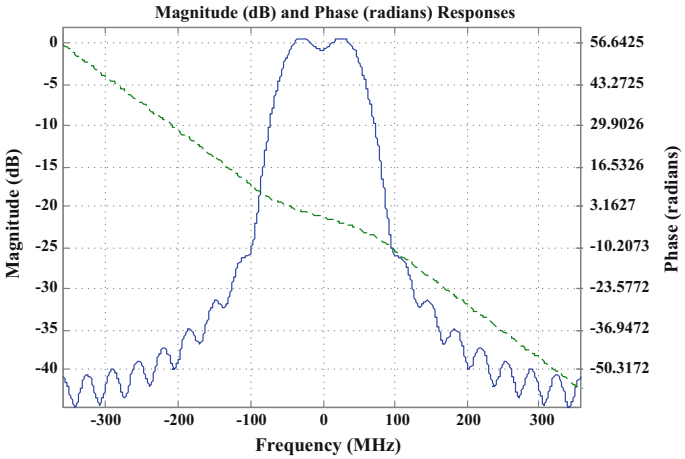


Fig. 33.5 The magnitude and phase response of mismatch channel

33.4.2 The Simulation Process

1. The SIMULINK platform is used to build the reference channel and the mismatch channel. The same linear frequency modulation signal, as the input signal to reference channel and mismatch channel, is conjugated and multiplied by the output signal of each channel.

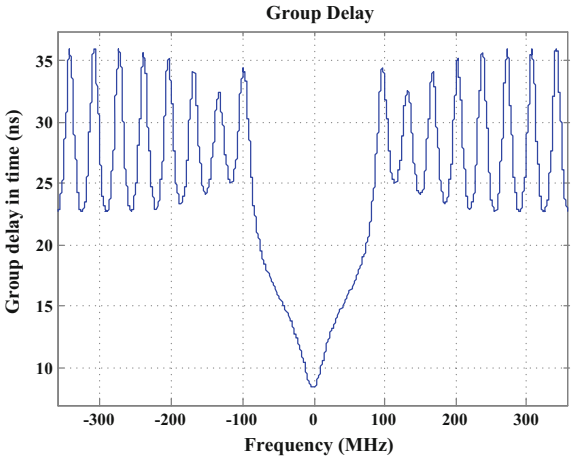


Fig. 33.6 The group delay of mismatch channel

2. The multiplied data is performed FFT, then the frequency $\hat{w}$ can be estimated by extracting the peak point. According to utilize formula (33.4), the channel delay estimation can be achieved and benefit to correct the large time delay difference between channels.
3. After correcting the large time delay difference between channels, the same linear frequency modulation signals enter each channel at the same time. Read the output sample data of reference channel and the mismatch channel, and mark the results as $h_{\text{ref}}(n)$ and $h_i(n)$ respectively.
4. $h_{\text{ref}}(n)$ and $h_i(n)$ are performed FFT with corresponding points. Mark the results as $Y_{\text{ref}}(m)$ and $Y_i(m)$ respectively.
5. Achieve the expected equalizer frequency response $H_i(m)$ among balanced frequency range by calculating the ratio between $Y_{\text{ref}}(m)$ and $Y_i(m)$.
6. After the previous step, the FIR filter coefficient can be obtained by using the least squares fitting method as showed in formula (33.12).
7. Analyze and compare the equilibrium properties of the filter.

33.4.3 The Results of Simulation and Analysis

The simulation parameters used are as follows. The 40 order ideal FIR filter is used to simulate the reference channel, as shown in Fig. 33.3. The 20 order complex coefficient FIR filter is used to simulate the mismatch channel, as shown in Fig. 33.5 (Table 33.1).

Table 33.1 The table of configuration parameters in balanced band

Channel	Fluctuation of magnitude (dB)	Time delay (ns)	Additional time delay (ns)	Total of time delay (ns)
Reference channel	0	27. 8	6250	6277.8
Mismatch channel	−1 to +0.7	8.4–15.9	0	8.4–15.9

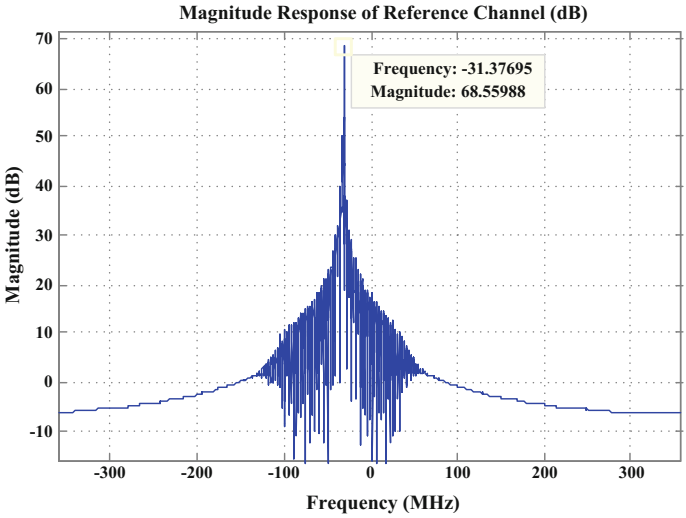


Fig. 33.7 The FFT results of reference channel data

The balanced band is −50 to +50 MHz. The equalizer is chosen to be 40 order. The scanning width of LFM signal is −50 to +50 MHz. The time width T is 10 μs. The sample frequency is 720 MHz. Signal-noise ratio is 20 dB. The reference channel is exerted 6250 ns (4500 samples) time delay in addition.

33.4.3.1 The Estimation of Time Delay

The same linear frequency modulation signal, as the input signal to reference channel and mismatch channel, is conjugated and multiplied by the output signal of each channel. The results after FFT are shown in Figs. 33.7 and 33.8. It is concluded that the peaks of FFT results correspond to the frequency of 31.37695 and 0.08789063 MHz respectively. Further, according to the formula (33.4), the time delay estimations of reference channel and mismatch channel are 6275.4 and 17.6 ns, respectively. Then, the difference of time delay between reference channel and mismatch channel is 6257.8 ns. Namely, the time delay of mismatch channel

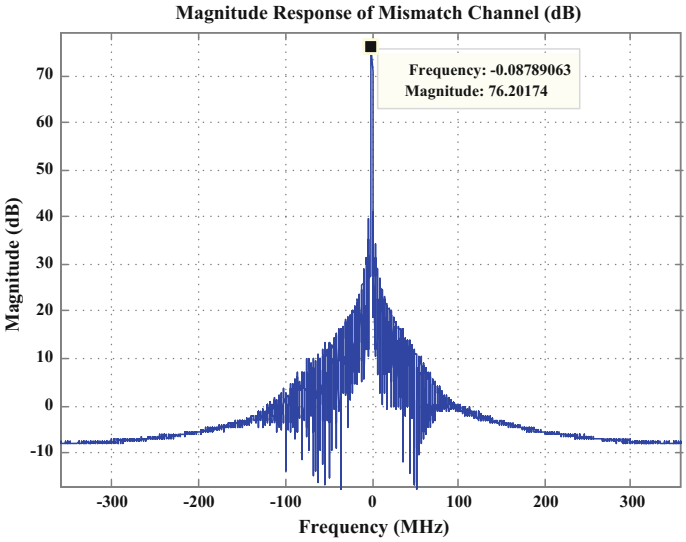


Fig. 33.8 The FFT results of mismatch channel data

should be get 6257.8 ns compensation. As a result of the reference channel total delay of 6277.8 ns and mismatch channel in pass band (−50 to +50 MHz) group delay within range of 8.4–15.9 ns, the rest of the time delay difference is 4.1–11.6 ns, after the mismatch channel is compensated 6257.8 ns time delay. Based on the theory, the maximum time delay of the equalizer to compensate is N/f_s. f_s is the sampling clock of system, and N is the order of equalizer. Under the simulation environment, the 40 order equalizer can compensate the biggest 55.6 ns time delay. So, the equalizer can completely correct remaining time delay difference, achieving phase and magnitude consistency at the same time.

33.4.3.2 The Result of Frequency Domain Equalization

As known from the analysis of previous section, the rest of the time delay difference between reference channel and mismatch channel is 4.1–11.6 ns after large time delay compensation. Next, the frequency domain equalization algorithm based on the least squares fitting is used to calibrate the rest of the time delay difference. The results of magnitude and phase differences in the pass band (−50 to +50 MHz) after equalization are shown in Figs. 33.9 and 33.10.

Figures 33.9 and 33.10 shows that the magnitude and phase differences between reference channel and mismatch channel in the pass band (−50 to +50 MHz) are rescaled in the range of 0.015 dB and 0.1° respectively after the calibration of equalizer. The results indicate that the consistency between channels is well calibrated by the use of the time delay estimation algorithm based on temporal

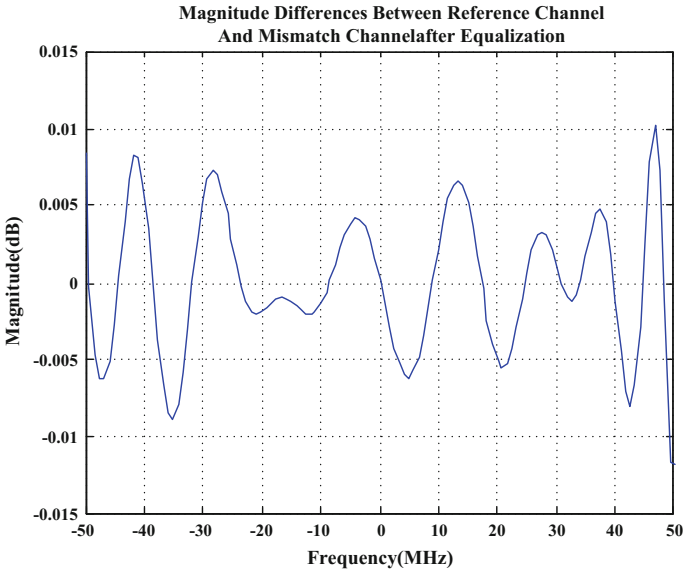


Fig. 33.9 The results of magnitude differences between reference channel and mismatch channel after equalization

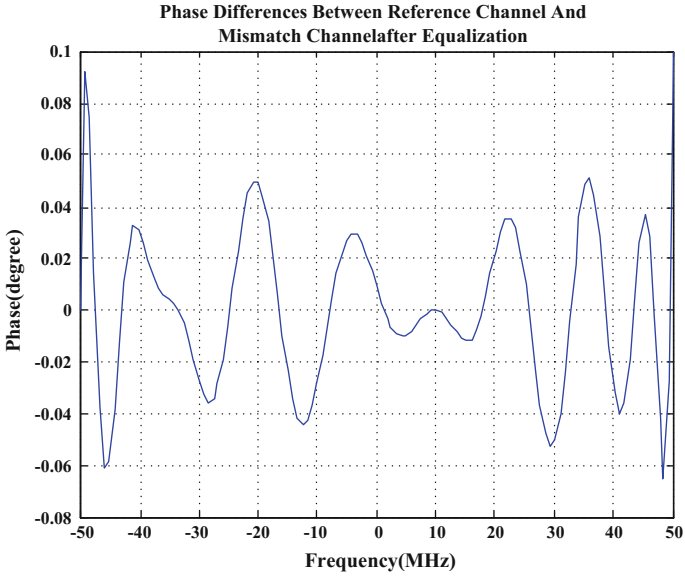


Fig. 33.10 The results of phase differences between reference channel and mismatch channel after equalization

coherence to correct the difference of large delay among channels and the method of frequency domain equalization to calibrate the rest of different delay.

33.5 Conclusions

Based on the request of channel consistency of distributed digital phased array, the paper use the time delay estimation algorithm to correct the large time delay differences among channels. As a result, the phase differences are controlled in the scope of equalizer compensation. Then, the frequency domain equalization algorithm based on the least squares fitting is used to complete the calibration of magnitude and phase consistency. The results show that the proposed method can effectively solve the issue of phase inconsistency caused by magnitude fluctuations and large time delay differences among channels in distributed digital phased array system. The calibration method can meet the demand of practical engineering project.

References

1. Mailloux RJ (2005) Phased array antenna, 2nd edn. Artech House, Norwood, pp 353–377
2. Han YC, Zhao XL, Zhao GQ (2009) Development and actuality of adaptive beamforming technology. Aerosp Electron Warf 5:44–47
3. Aalfs DD, Holder EJ (2000) Impact of wideband channel-to-channel mismatch on adaptive array. In: Proceedings of IEEE SAMSP, Cambridge, pp 459–463
4. Wang F, Fu YG, Meng B, Chen LM (2006) Performance analysis and improvement of the equalization algorithm based on Fourier transform for radar channel. Acta Electron Sin 34 (9):1677–1680
5. Moddemeijer R (1991) On the determination of the position of extrema of sampled correlators. IEEE Trans Signal Process 39(1):216–219
6. Deng B, Wang X, Tao R, Liu XZ (2012) Performance analysis of time delay estimation for linear frequency-modulated pulse based on fractional Fourier transform. Acta Armament 33 (6):764–768
7. Wang DC (2010) Wideband phased array radar. National Defend Industry Press, Beijing
8. Chen G (2008) Channel equalization technology research in wideband digital array radar. Master Thesis, University of Electronic Science and Technology of China, Chengdu
9. Zhang Y, Bao QL, Yang J, Chen ZP (2010) Design and implementation of channel equalization method for wideband digital array radar. Signal Process 26(3):453–457
10. Chen G, Li HY, He ZS (2008) An improved algorithm of channel equalization for wideband digital array. Radar Sci Technol 6(6):015

Chapter 34
Multi-hit Method for Weak Signal Detection of the Diffuse Reflection Laser Ranging in Daylight

Peng Zhao, Yan Zhang, Kunpeng Wang and Chenglin Wang

34.1 Introduction

Diffuse reflection laser ranging technology is one of the feasible ways to achieve high precision in the measurement for space target, especially for the space debris. However, since the space debris and most spacecraft are non-cooperative target without the corner reflector, the echo signal diffused form the target will be attenuate to the level of single photon at the front head of the detector. Geiger-mode avalanche photodiode (GM-APD) possessing high sensitivity and gain are widely used in diffuse reflection laser ranging system [1]. However, GM-APD is still difficult to distinguish between noise and the weak echo signal effectively because of its binary response. At the same time, the existing laser ranging system will shut down the detector after the echo signal is detected in a gate time (that is, the single-hit). As a result, the noise among the gate will reduce the detection probability of the echo signal. Therefore, the weak echo signal will not be detected under the condition of the strong background noise in daylight.

The multi-hit method implements the detection of many firing times in one gate through reasonable design of the ranging system, which is well investigated to improve the detection probability of partially covered targets in ladar [2–4]. The diffuse reflection laser ranging in daylight and the detection for the partially covered target in ladar are facing the similar problem, namely the low detection probability of weak echo signal caused by the strong noise. Therefore, the multi-hit method is one possible method to improve the weak echo signal detection of the diffuse reflection laser ranging in daylight, requiring further theoretical analysis. The

P. Zhao (✉) · Y. Zhang · K. Wang · C. Wang
Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China
e-mail: roczhao1991@163.com

P. Zhao · Y. Zhang · K. Wang · C. Wang
Key Laboratory of Space Object Measurement, Beijing 100094, China

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_34

detection probability of GM-APD is the key to analyze the performance of the multi-hit method. Johnson et al. [2] proposed recursive method under discrete time to solve the detection probability of the GM-APD. Gatt et al. [5] on the basis of Johnson, developed a recursive method under continuous time to calculate more accurate signal detection probability of the GM-APD. Recursive method is a universal and accurate method. However, when the dead time is small compared with the gate time, the complexity of the recursive method will increase sharply. Zhao et al. [6] proposed LTR method to solve the detection probability. The LTR method is fast and accurate performance prediction method, but can only solve the detection probability under the condition of the biggest firing time.

In order to analyze the performance of the multi-hit method in improving the weak echo signal detection of the diffuse reflection laser ranging in daylight theoretically, this paper introduced the improved LTR method, which can be used to solve the detection probability under the condition of different firing times. Then, taking the typical parameters of the diffuse reflection laser ranging system, we analyze the performance of the multi-hit method in improving the weak echo signal detection of the diffuse reflection laser ranging in daylight theoretically based on the detection probability of the echo signal under different firing times. Furthermore, the cost of the echo signal is also analyzed while the effective detection ratio is defined to evaluate the extraction cost of the echo signal. Finally, the simulation system is given to prove the validity of the theoretical analysis above.

34.2 Solving of the GM-APD Detection Probability

Each firing of the GM-APD depends on the generation of primary electrons and the avalanche followed in sequence [7]. Primary electrons (PEs) are generated mainly from the echo light and the background light by imperfect quantum efficiency as well as the dark noise inside the GM-APD. Then, the avalanche initiated by the PEs produces a detectable output waveform independent of the number of PEs. Once the avalanche occurs, a period of time denoted the dead time (T_{dead}) follows in which the GM-APD detector will be invalid for re-firing. Define the average rate of echo light and background light as $R_{\text{echo}}(t)$ and R_{back}. Define the average rate of dark noise as R_{dark}. Then, the total rate function of mean input flux, $\psi(t)$, is shown as

$$\psi(t) = R_{\text{echo}}(t) + R_{\text{back}} + R_{\text{dark}}. \tag{34.1}$$

In order to calculate under discrete time, the gating time (T_{gate}) is discretized into N bins with the same time period T_{bin}. The dead time is discretized into d bins where d equals to ceil ($T_{\text{dead}}/T_{\text{bin}}$) and the function ceil() is the rounding up function. Then, the mean number of PEs generated from arbitrary input at the ith bin, $\overline{K}_i$, is given by

$$\overline{K}_i = \int_{(i-1)T_{\text{bin}}}^{iT_{\text{bin}}} \psi(t)\mathrm{d}t. \tag{34.2}$$

The number of PEs, K_i, is generally assumed to obey Poisson distribution [7]. Then, the probability that more than one PE is generated at the ith bin is given by [5]

$$P_{\text{PE}}(i) = \Pr(K_i \geq 1) = 1 - \exp(-\overline{K}_i). \tag{34.3}$$

Only if the end of the gating time (T_{gate}) is not reached and the dead time is over, the GM-APD detector can be fired again. Therefore, the dead time and the gating time together decide the maximum firing time of the GM-APD, which is shown as

$$M_{\max} = \left\lceil \frac{T_{\text{gate}}}{T_{\text{dead}}} \right\rceil. \tag{34.4}$$

We might as well set certain firing time, M to investigate the performance of GM-APD of different firing time as long as M is no more than $M_{\max}$. The probability of avalanche for M firing time at the ith time interval is a function of M, which is defined as $P_{\text{ava}}(i, M)$. When the multi-hit GAPD detector responses to the input flux, three states could occur at each bin, the GM-APD succeeds to avalanche, the GM-APD fails to avalanche and the GM-APD is blocked during the dead time. Denote these three states as '1', '0' and 'X' respectively. When the GM-APD is armed at the ith bin, possible states can be listed by looking back upon the previous bins. Three look-back-upon rules should be followed in a recurrent way, (a) the initial state is '1' at the ith bin, (b) the prior states of '0' or '1' are '0'and 'X', (c) the prior states of 'X' are '1' or 'X', depending on the location of current states 'X' in the sequence of $\underbrace{\text{X} \rightarrow \text{X} \cdots \text{X} \rightarrow 1}_{d}$. The structure obtained during the look-back-upon process after i times recurrence is defined as look-back-upon tree (LT) of size i, LT_i for short. An example of LT_5 where GAPD is armed at the fifth bin and $d = 2$ is shown in Fig. 34.1.

As a result, $P_{\text{ava}}(i, M)$ can be calculated by summing up all the probabilities of valid connections in LT_i. Those state connections with the state 'X' at the end are invalid because the state 'X' cannot happen at the beginning of the gated time. Meanwhile, those state connections with state '1' and state '0' at the end are valid connections. Define $P^1_{\text{ava}}(i, M)$ and $P^0_{\text{ava}}(i, M)$ as the probability of valid connection with state '1', '0' at the end respectively for the M firing time, while the number of state '1' is actually the firing time in each state connection. Then, $P_{\text{ava}}(i, M)$ is given by

$$P_{\text{ava}}(i, M) = P^1_{\text{ava}}(i, M) + P^0_{\text{ava}}(i, M), \tag{34.5}$$

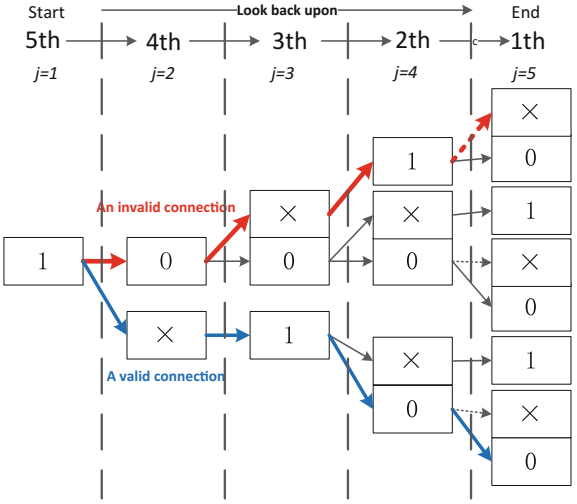


Fig. 34.1 Look-back-up tree where $i = 5$ and $d = 2$

where $P^1_{\text{ava}}(i, M)$ and $P^0_{\text{ava}}(i, M)$ will be solved in the recurrence deduction of LT_i. When LT_j is deduced from LT_{j-1}, the possible state of $i–j + 1$ bin is going to be listed where j is a temporary variable less than i.

According to the look-back-upon rules, state '0' at the end of LT_j comes from state '0' and state '1' at the end of LT_{j-1}. In addition, state '0' at the end of LT_j means no PE will be generated at the $i–j + 1$th bin, besides state '0' will not increase the firing time in the state connection. Therefore, $P^0_{\text{ava}}(j, M)$ equals the product of the probability that no PE is generated at the $i–j + 1$th bin and the sum of $P^0_{\text{ava}}(j-1, M)$ and $P^1_{\text{ava}}(j-1, M)$. When $j = 1$, $P^0_{\text{ava}}(j, M)$ equals to 0 since the initial state is '1' at the ith bin. In the discussion above, M can take the value from 1 to $M_{\max}$. In conclusion, the recurrent solving equations for $P^0_{\text{ava}}(i, M)$ is shown as

$$\begin{cases} P^0_{\text{ava}}(j, M) = 0 & j = 1, \quad 1 \le M \le M_{\max} \\ P^0_{\text{ava}}(j, M) = (1 - P_{\text{PE}}(i-j+1))(P^0_{\text{ava}}(j-1, M) + P^1_{\text{ava}}(j-1, M)) & 1 < j \le i, \quad 1 \le M \le M_{\max} \end{cases} \tag{34.6}$$

According to the look-back-upon rules, state '1' at the end of LT_j only comes from state 'X' at the end of LT_{j-1} while state 'X' at the end of LT_j comes from state '0', '1' and 'X' at the end of LT_{j-1}. However, because the sequence of $\underbrace{\text{X} \rightarrow \text{X} \cdots \text{X} \rightarrow 1}_{d}$ is always bound together, the situation that new 'X' comes from old 'X' only happens within the sequence while the first 'X' of the sequence comes from state '0' and '1' at the end of LT_{j-d}. In addition, state '1' at the end of LT_j means PEs will be generated at the $i–j + 1$th bin and state '1' will increase the firing time by one in the state connection. Therefore, $P^1_{\text{ava}}(j, M)$ equals the product of the probability that more than one PE is generated at the $i–j + 1$th bin and the sum of

$P^0_{\text{ava}}(j-d,M)$ and $P^1_{\text{ava}}(j-d,M)$. Since the initial state is '1' at the ith bin, $P^1_{\text{ava}}(1,1)$ equals to $P_{\text{PE}}(i)$ while $P^1_{\text{ava}}(1,M), \quad 1<M\le M_{\max}$ and $P^1_{\text{ava}}(j,1), \quad 1<j\le i$ equals to 0. Because there will be no more '1' within the first d bins, $P^1_{\text{ava}}(j,M)$ equals to zero when $1<j\le d$ no matter what the firing time M is. In conclusion, the recurrent solving equations for $P^1_{\text{ava}}(i,M)$ is shown as

$$\begin{cases} P^1_{\text{ava}}(j,M)=P_{\text{PE}}(i) & j=1, \quad M=1 \\ P^1_{\text{ava}}(j,M)=0 & j=1, \quad 1<M\le M_{\max} \\ P^1_{\text{ava}}(j,M)=0 & 1<j\le i, \quad M=1 \\ P^1_{\text{ava}}(j,M)=0 & 1<j\le d, \quad 1<M\le M_{\max} \\ P^1_{\text{ava}}(j,M)=P_{\text{PE}}(i-j+1)(P^0_{\text{ava}}(j-d,M-1)+P^1_{\text{ava}}(j-d,M-1) & d<j\le i, \quad 1<M\le M_{\max} \end{cases} \tag{34.7}$$

Through Eqs. (34.6) and (34.7), $P^1_{\text{ava}}(i,M)$ and $P^0_{\text{ava}}(i,M)$ are obtained. Finally, $P_{\text{ava}}(i,M)$ can be obtained according to Eq. (34.5). Define M_{lim} as the trigger limit and define $P_r(t,M_{\text{lim}})$ as the probability of avalanche when M is less than M_{lim}. $P_r(t,M_{\text{lim}})$ is shown as

$$P_r(t,M_{\text{lim}})=\sum_{M=1}^{M_{\text{lim}}} P_{ava}(i,M)\,(i-1)T_{\text{bin}}\le t\le iT_{\text{bin}}, \quad 1\le M_{\text{lim}}\le M_{\max}. \tag{34.8}$$

34.3 Theoretical Analysis of the Multi-hit Method

34.3.1 Typical Parameters of Diffuse Reflection Laser Ranging

According to the typical parameters of diffuse reflection laser ranging and estimation formula of the echo photon number for laser ranging [8], the average echo photon number per pulse, N_s, is equal to 0.85. When the Q-switched pulsed laser at 532 nm is used, the normalized pulse distribution formula is shown as [9]

$$f(t)=\frac{1}{\tau}\frac{t}{\tau}\exp\left(-(t/\tau)\right), \tag{34.9}$$

where $\tau = T_p/3.5$ and $T_p = 10$ ns. Due to the existence of the ranging error, the firing time of the echo signal is difficult to determine. Thus, for the convenience of theoretical analysis, we might as well assume that the ranging error is a known constant and the echo signal arrives at the middle of the gate. Consequently, $R_{\text{echo}}(t)$ is shown as

$$R_{\text{echo}}(t)=\begin{cases} N_s f\left(t-\frac{T_{gate}}{2}\right) & \frac{T_{gate}}{2}\le t\le T_{gate} \\ 0 & 0\le t<\frac{T_{gate}}{2} \end{cases}. \tag{34.10}$$

Assuming T_{gate} = 1 μs, typical values of R_{back} under the condition of cloudless night and cloudless daylight can be taken as [10, 11]

$$R_{\text{back}}=\begin{cases} 150\,\text{Hz} & \text{cloudless night} \\ 3\times 10^6\,\text{MHz} & \text{cloudless daylight} \end{cases}. \tag{34.11}$$

As for the average dark noise counting of the GM-APD, the diffuse reflection laser ranging system generally takes 4–10 kHz [12, 13]. In this paper, R_{dark} is assumed to be 10 kHz. In addition, the dead time of common type silicon detector is generally between 30 and 200 ns. In this paper, we assume that the dead time of the GM-APD is 100 ns. At the same time, in order to obtain accurate theoretical calculation results, N takes 1000 in the performance prediction method of GM-APD [6].

34.3.2 Analysis of Echo Signal Detection Probability

Substituting Eq. (34.1) with Eqs. (34.10), (34.11) and R_{dark}, $\psi(t)$ is obtained under the cloudless night and the cloudless daylight observation. Then, we can solve $P_r(t, M_{\text{lim}})$. Ignoring the shape effect of the target, Integrate $P_r(t, M_{\text{lim}})$ within the signal pulse width, T_p, we can further obtain the echo signal detection probability, defined as $P_s(M_{\text{lim}})$. $P_s(M_{\text{lim}})$ is shown as,

$$P_s(M_{\text{lim}}) = \int_{T_{gate}/2}^{T_{gate}/2+T_p} P_r(t, M_{\text{lim}})\mathrm{d}t. \tag{34.12}$$

Therefore, according to Eq. (34.12), $P_s(M_{\text{lim}})$ under the cloudless night and the cloudless daylight observation condition can be calculated as shown in Fig. 34.2.

As shown in Fig. 34.2, under the cloudless night observation condition, the echo signal detection probability is basically unchanged with the changing of M_{lim}. While under the cloudless daylight observation condition, the echo signal detection probability increases with the increasing of M_{lim}, and is saturated when M_{lim} is more than 4. Although the echo signal detection probability in daylight is lower than the echo signal detection probability at night, the increase of M_{lim} can improve the detection probability of the diffuse reflection laser ranging by about 4 times. The above results show that the multi-hit method is a feasible way to detect the weak echo signal of the diffuse reflection laser ranging in daylight.

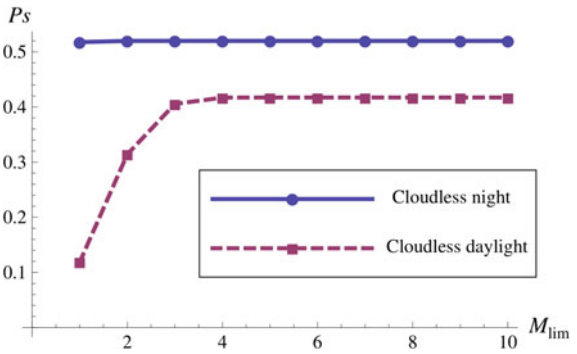


Fig. 34.2 $P_s(M_{\text{lim}})$ for the cloudless night and the cloudless daylight

34.3.3 *Effective Detection Ratio*

Research shows that the increase of the noise will increase the extract cost of the echo signal [14]. At the same time, the multi-hit method can increase the detection probability of noise during the gate [15]. Therefore, in addition to studying the detection probability of weak echo signal in daylight using multi-hit method, the effect of multi-hit method on the extraction of weak echo signal should be also considered. Define $P_t(M_{\text{lim}})$ as the total detection probability and define $R_{ef}(M_{\text{lim}})$ as the effective detection ratio, then $R_{ef}(M_{\text{lim}})$ is shown as

$$R_{ef}(M_{\text{lim}}) = \frac{P_s(M_{\text{lim}})}{P_t(M_{\text{lim}})}, \tag{34.13}$$

where $P_t(M_{\text{lim}})$ is the integral of $P_r(t, M_{\text{lim}})$ over the entire gate time, T_{gate}, which is shown as

$$P_t(M_{\text{lim}}) = \int_0^{T_{gate}} P_s(t, M_{\text{lim}})\, dt, \tag{34.14}$$

In this paper, the effective detection ratio is used as a measure of the cost of te echo signal extraction. The higher the effective detection ratio, the higher the proportion of the echo signal detection probability, the lower the cost of the echo signal extraction, and vice versa. According to Eq. (34.13), the effective detection ratio under the cloudless daylight observation condition can be calculated, which is shown in Fig. 34.3.

As we can see from Fig. 34.3, under the cloudless night observation condition, the effective detection ratio is basically unchanged with the changing of M_{lim}. While under the cloudless daylight observation condition, the effective detection ratio gradually increased to the maximum value with the increasing of $M_{\text{lim}}(M_{\text{lim}} = 3)$. With the further increase of M_{lim}, the effective detection ratio is gradually reduced to a constant value. At the same time, the optimization of the firing time for

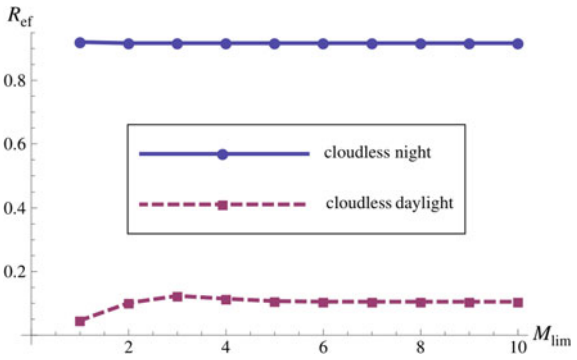


Fig. 34.3 The effective detection ratio for the cloudless night and the cloudless daylight

GM-APD, that is, selecting M_{lim} corresponding to the maximum effective detection ratio, can reduce the extraction cost of the echo signal under the cloudless daylight observation condition. However, even the maximum value of the effective detection ratio under the cloudless daylight observation condition is only about 1/5 of the effective detection ratio under the cloudless night observation condition. This shows that the echo signal extraction cost of cloudless daylight observation is difficult than the echo signal extraction cost of cloudless night observation. Meanwhile, since the gap between them is still within a reasonable range, the multi-hit method works for the echo signal extraction under the cloudless daylight observation condition basically.

34.4 Simulation Verification

For the performance analysis of the multi-hit method, defuse reflection laser ranging simulation system is designed based on typical parameters of the diffuse reflection laser ranging experiment of Yunnan observation. the simulation system mainly consists of three parts: the target distance simulation, echo signal simulation and GM-APD detection simulation.

34.4.1 Target Distance Simulation

In the theoretical analysis, it is assumed that the target distance is constant, the echo arrival time is fixed and is set in the central part of the gate. In the actual system, the distance of the detection target will change with time. Therefore, our simulation system presents the distance changing feature based on the 25,400 space target during an observation of 160 s. When the laser pulse frequency of the simulation system is 10 Hz, the target simulation system will accomplish laser pulse ranging for 1600 times. Selecting the corresponding measured TLE orbit data of the 25,400

space target within the same time length, we can obtain the distance data of each laser pulse ranging in the simulation system. At the same time, it is necessary to consider the influence of the ranging error. For the sake of simple discussion, we assume that the error characteristic is linear. Then ranging error is added to the target distance data where the error rate is 3 m/s. Besides, the gate time of Yunnan observation is approximately ten microseconds. As a result, $T_{\text{gate}} = 10$ μs in the simulation system.

34.4.2 Echo Signal Simulation

In Sect. 34.3.1, according to the typical parameters of diffuse reflection laser ranging and estimation formula of the echo photon number for laser ranging, the average echo photon number per pulse has been given ($N_s = 0.85$), but in the actual system, the presence of telescope tracking problem, the influence of laser energy fluctuation, the atmospheric turbulence and other factors needs to be considered in the simulation system.

At the beginning of the diffuse reflection laser ranging, the telescope is adjusted to track the target. The echo signal cannot be received until the tracking and aiming are correct. Therefore, the echo ranging data cannot be received at the initial stage. In the simulation, the initial stage of the tracking and aiming is assumed to last for 6 s. The laser pulse energy fluctuation will lead to the fluctuation of the average number of the echo signal. And the average number of the echo signal is proportional to the energy of the laser pulse. If the laser echo pulse energy obeys Gauss distribution, the average number of echo signal should also obey the Gauss distribution. As a result, the average number of the echo signal is assumed to obey Gauss distribution with the mean of 0.85 and the standard deviation of 0.2 in the simulation. In addition, the echo signal may have the overall lack in a short period of time. In the simulation, it is assumed that the probability of the overall lack is 1% and its duration is 1 s.

34.4.3 GM-APD Detection Simulation

First of all, the GM-APD detection simulation needs to determine the GM-APD key parameters: the dark noise and the dead time. The noise of the GM-APD used by Yunnan observation for diffuse reflection laser ranging has a high dark noise GM-APD (about 60 kHz [16]). In the simulation, R_{dark} is assumed to be 60 kHz. As for the value of the dead time, related literature of the existing test in Yunnan observation has not discussed the dead time. The reason is that the dead time almost has no effect on the laser ranging system. Consequently, this paper failed to determine the dead time of the GM-APD used by the Yunnan observation. Here, we take the dead time of GM-APD as 100 ns. It should be noted that the A033-ET

timer [17] with the dead time of 50 ns should be used on the echo pulse timing, in order to avoid the situation that the dead time of timer is greater than the dead time of the detector.

In the echo signal detection of the GM-APD for one pulse, this paper uses Monte Carlo method to simulate the detection process of the GM-APD. The gate time is divided into N bins and the average number of PE in each bin is calculated by Eq. (34.2). Further, according to Eq. (34.3), the probability of the PE generation in each bin is calculated. Then, random numbers uniform distributed from 0 to 1 are generated at each bin. If the random number is greater than the probability of the PE generation in the current bin, avalanche will occur at this bin. Otherwise, continue to check whether the random number is greater than the probability of the PE generation in the next bin until all the bins are checked. Once the avalanche occurred, the triggered time of the avalanche event and the current firing number within the gate are recorded. Then, move backward $d - 1$ bins as the dead time interval and start over the checking process from the first bin after the end of the dead time interval.

After completing the echo signal detection of the GM-APD for one pulse, repeat this detection process for the next pulse until the observation time meet the time requirement.

Based on the recorded trigger time, the number of the received echo signal within the observation period is obtained by summing up all the firing number during the echo pulse time within the gate under certain M_{lim}. The number of the received echo signal and noise within the observation period is obtained by summing up all the firing number within the gate under certain M_{lim}. Then, the simulated $P_s(M_{\text{lim}})$ is obtained when the number of the received echo signal within the observation period is divided by the total laser pulse. The simulated $P_t(M_{\text{lim}})$ is obtained when the number of the received echo signal and noise within the observation period is divided by the total laser pulse. Moreover, $R_{ef}(M_{\text{lim}})$ is obtained by substituting Eq. (34.13) with $P_s(M_{\text{lim}})$ and $P_t(M_{\text{lim}})$.

34.4.4 Simulation Results

10,000 simulation experiments of the diffuse reflection laser ranging system are repeated. The average value of the simulated echo signal detection probability is shown in Fig. 34.4a. The feature of the echo signal detection probability changing with M_{lim} is consistent with the theoretical analysis. The average value of the simulated effective detection ratio is shown in Fig. 34.4b. The feature of the effective detection ratio changing with M_{lim} is consistent with the theoretical analysis.

As shown in Fig. 34.4, the detection probability of the echo signal is almost zero under the single-hit situation. With the increase of the number of the firing times, the detection probability of the echo signal can still reach the same level as the

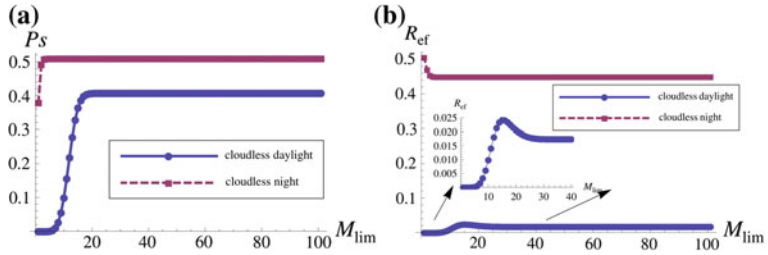


Fig. 34.4 Average results of 10,000 simulation experiments: **a** simulated echo signal detection probability, **b** simulated effective detection ratio

detection probability of the echo signal in the theoretical analysis (about 0.4). This fully shows the outstanding performance of the multi-hit method in improving the detection probability of the diffuse reflection laser ranging. At the same time, the effective detection ratio is almost zero under the single-hit situation. While the optimization of the firing time for GM-APD can greatly improve the effective detection ratio (the maximum of the effective detection ratio is 0.025). This shows that the multi-hit method can greatly reduce the cost of the signal extraction under the cloudless daylight observation condition. Therefore, the multi-hit method is a feasible way to realize the detection of the weak echo signal of the diffuse reflection laser ranging in daylight.

34.5 Conclusions

This paper carried out the theory analysis of the multi-hit method in solving the weak signal detection of the diffuse reflection laser ranging in daylight. First of all, the LTR method based on the look-back-upon tree for the performance prediction of the GM-APD is introduced. Then, taking the typical parameters of the diffuse reflection laser ranging system, the performance of the multi-hit method in improving the detection probability of the echo signal is analyzed. Furthermore, the effective detection ratio is defined as the standard to evaluate the extraction cost of the echo signal. The performance of the multi-hit method in improving the effective detection ratio is analyzed. The analysis results show that under the cloudless daylight observation condition, the multi-hit method can improve the detection probability of the echo signal greatly. At the same time, through the optimization of the firing time, the multi-hit method can also improve the effective detection ratio and reduce the extraction cost of the echo signal under the cloudless daylight observation condition.

Finally, the simulation system of diffuse reflection laser ranging is presented, which considers the distance error, the target tracking time, the laser energy fluctuation, the atmospheric turbulence, and the Poisson detection process. The simulation results of 10,000 repeated experiment show that the changing feature of the

detection probability of the echo signal and the effective detection ratio is consistent with the theoretical analysis. The multi-hit method improve the weak signal detection probability and effective detection ratio significantly, where the detection probability of the echo signal can still reach the same level as the detection probability of the echo signal in the theoretical analysis. Therefore, the multi-hit method is a feasible way to realize the detection of the weak echo signal of the diffuse reflection laser ranging in daylight.

References

1. Courde C, Samain E, Laas-Bourez M et al (2012) Laser ranging on space debris with the MeO station. In: Proceeding of the international technical laser workshop, OCA, France
2. Johnson S, Gatt P, Nichols T (2003) Analysis of Geiger-mode APD laser radars. Proc SPIE 5086:359–368
3. Degnan J, Machan R, Leventhal E et al (2008) Inflight performance of a second generation, photon counting, 3D imaging lidar. Proc SPIE 6950:695007
4. Oh MS, Kong HJ, Kim TH et al (2010) Time-of-flight analysis of three-dimensional imaging laser radar using a Geiger-mode avalanche photodiode. Jpn J Appl Phys 49(2R):026601
5. Gatt P, Johnson S, Nichols T (2009) Geiger-mode avalanche photodiode ladar receiver performance characteristics and detection statistics. Appl Opt 48:3261–3276
6. Zhao P, Zhang Y, Hua YM et al (2015) Look-back-upon tree recurrence method for Geiger-mode avalanche photodiode performance prediction. Opt Lett 40(16):3822–3825
7. Fouche DG (2003) Detection and false-alarm probabilities for laser radars that use Geiger-mode detectors. Appl Opt 42(27):5388–5398
8. Li Y, Li Z, Fu H et al (2011) Experimentation of diffuse reflection laser ranging of space debris. Chin J Lasers 38(9):0908001 (in Chinese)
9. Wang F, Zhao Y, Zhang Y et al (2010) Range accuracy limitation of pulse ranging systems based on Geiger mode single-photon detectors. Appl Opt 49(29):5561–5566
10. Yang F, Chen W (1998) Design and experimental results of the satellite laser ranging system in daylight. Sci China A 28(11):1048–1056 (in Chinese)
11. Zhang Z, Wu L, Zhang Y et al (2013) Method to improve the signal-to-noise ratio of photon-counting chirped amplitude modulation ladar. Appl Opt 52(2):274–279
12. Stipcevic M, Skenderovic H, Gracin D (2010) Characterization of a novel avalanche photodiode for single photon detection in VIS-NIR range. Opt. Exp. 18 (16), 1748–1759
13. LASER COMPONENTS. Silicon Geiger mode avalanche photodiode. http://www.lasercomponents.com/de-en/product/silicon-apds-for-photon-counting. Accessed 15 Oct 2015
14. Zhao P, Zhang Y, Zhang X et al (2015) Echo online extraction for diffuse reflection laser ranging. Acta Opt Sin 10:280–288 (in Chinese)
15. Zhao P, Zhang Y, Qian WP et al (2015) Investigation of Geiger-mode detector in multi-hit model for laser ranging. Sci China Technol Sci 58:16
16. Chan HM, Ming ZX, Guo JC (2003) The detectors of satellite laser ranging in yunnan observatory. Astron Res Technol 3:62–68 (in Chinese)
17. Li Z, Li Y, Fu H et al (2012) Design and realization of a 10 Hz diffuse-reflection laser ranging control system. Astron Res Technol 3:010 (in Chinese)

Part III
Information Science and Technology

Chapter 35
A Real-Time Classification Algorithm for Multi-Velocity Measuring Data

Xiaohu Liang, Hua Zhao and Jiagui Huang

35.1 Introduction

Multi-velocity measuring system has two measuring modes: the response mode and the beacon mode [1]. Under the response mode, the multi-velocity measuring system is configured to one master station and some vice station. The master station sends the uplink signal, which is received by a transponder. Then the transponder forwards the downlink signal (in order to avoid interference, the uplink and downlink signals have different frequencies). the master and vice stations all receive the downlink signal, measuring the Doppler frequency shift, which can recover the change rate of the whole range from the master station to the target and from the target to the receiving station [2]. This kind of measuring data is called the response data. While under the beacon mode, the measuring system has no any master station. The transponder sends the downlink signal directly. All the stations receive the downlink signal and measure the Doppler frequency shift, which can recover the change rate of the range from the target to the receiving station [3]. This kind of measuring data is called the beacon data. Because the measuring accuracy of the response data is higher than the beacon data, the response measuring mode is the main working pattern of the multi-velocity measuring system.

But in practical tasks, even if the multi-velocity measuring system uses the response mode to measure target, the received measuring data may contain the response data, the beacon data and the abnormal data. The latter two kinds are called non-response data. In response mode, the causes for generating the beacon data include: the master station can't track the target in the starting period, the master station suddenly fault, the target tail flame degrades the uplink signal

X. Liang (✉) · H. Zhao · J. Huang
Beijing Institute of Tracking and Telecommunications Technology, 100094 Beijing, China
e-mail: liangxiaohu_2006@163.com

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_35

seriously, and there exists no master station in the relay section between two master stations, etc. [4]. In these cases, no uplink signal is sent or the transponder doesn't receive the uplink signal, then the measuring data of all stations turns into the beacon data. In order to obtain the correct target trajectory and improve the precision, the two kinds of data must be classified correctly, because they have different measuring equations, different wave refraction correction methods and different trajectory calculating methods [5, 6]. The data classification error will lead to real-time trajectory calculating error, with the phenomenon of trajectory jump in large sections, which has bad effects on many aspects such as the flight status display, flight security control, real-time guidance, trajectory impact forecast and so on [7].

For the problem of real-time classification of the multi-velocity measuring system data, the existing algorithm used in the test range takes the beacon data as abnormal data to reject them. But it can only handle the situation that the non-response data is discrete isolated distribution, which is unable to process the large segment of the beacon data. And the two master stations relay situation is also not for special consideration. It usually determines the real master station in accordance the designed switch time of the two master stations in the pre-populated scheme. When the actual switch time is not identical to the designed time, it will cause the trajectory calculation error in the relay paragraph. This paper analyses two typical kinds of experimental data, and proposes a new classification algorithm, which is validated with the real measured data and achieves good classification result. On this basis, the smooth and continuous real-time trajectory can be calculated in the whole measuring period.

35.2 Data Classification Problem Under the Response Mode

This paper studies the real-time data classification problem for the multi-velocity measuring system configured to the response mode. Because under the beacon mode, there will be no response data. Then the problem of data classification is not encountered. According to the number of the master station in practical tasks, the response mode can be divided into two situations: one has only one master station sending the uplink signal, the other has two master stations relay sending the uplink signal. The effective data of the measuring station in the former situation is called effective data 1 for short, as shown in Fig. 35.1. The effective data of the measuring station in the latter situation is called effective data 2 for short, as shown in Fig. 35.2.

For effective data 1, data classification refers to classify the effective data further into the response data, the beacon data and the abnormal data. For effective data 2, data classification refers the same as the former before and after the relay paragraph. But during the relay paragraph, we not only need to determine whether the data is

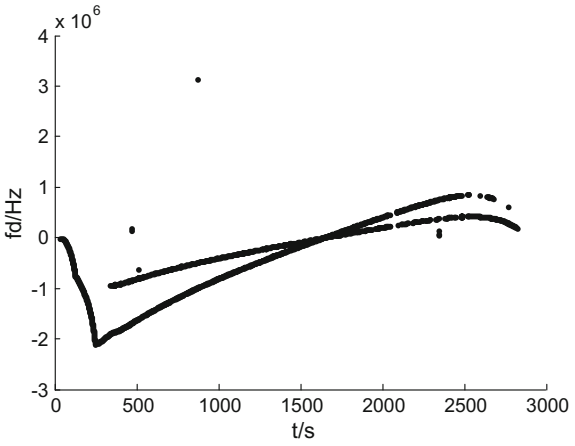


Fig. 35.1 Effective data 1 with one master station

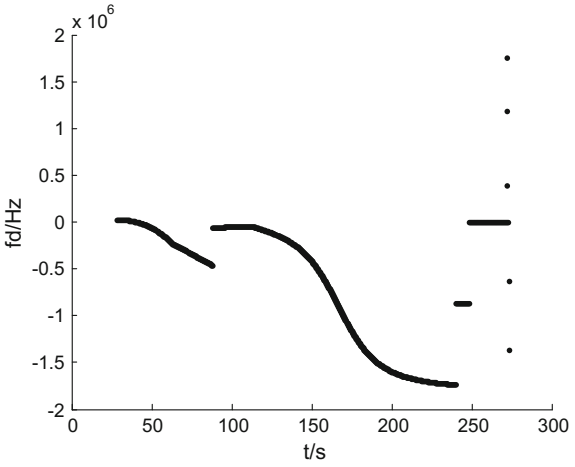


Fig. 35.2 Effective data 2 with two master stations relay

the response data, but also need to determine its relevant master station. If the master station is mismatched, the real-time trajectory will be wrong.

35.3 Analysis of the Existing Method

The existing method is rationality test to classify the measuring data into the response data and the abnormal data. Its basic idea is: under the response mode, the majority of effective data should be the response data, and only little is non-response data. So the data passing the rationality test is regarded as the

response data, while the data not passing the test is regarded as the abnormal data to remove away.

The rationality test uses a five-point linear extrapolation prediction method. Firstly, the fourth-order difference test is used to find a reasonable five-point data sequence, which is considered as the effective response data sequence. Then, the five-point data sequence is used to calculate the prediction data in the next measuring time using linear extrapolation method. When the deviation between the real measuring data and the prediction is lower than the designed threshold, the real measuring data is determined as a response data. While the deviation exceeds the threshold, the real data is determined as an abnormal data and is replaced by the prediction data to calculate the target trajectory. If five consecutive real data points are determined as abnormal data, a new reasonable data sequence need to be found out again using the fourth-order difference test.

In the case of two master stations relay measurement, the existing method just matches the master station according to the pre-populated scheme. Assume the designed relay time segment in the scheme is $[t_1, t_2]$, that means the master station 1 sending the uplink signal before the time t_1 and the master station 2 sending the uplink signal after the time t_2. During the segment $[t_1, t_2]$, no master station sends the uplink signal. For the rational real data passing the rationality test, if the measuring moment is before time t_1, the data is determined as the response data corresponding to the master station 1; if the measuring moment is after time t_2, the data is determined as the response data corresponding to the master station 2; if the measuring time is during $[t_1, t_2]$, the data is determined as the beacon data.

Use the above mentioned method to perform the rationality test for effective data 1 and effective data 2 respectively. The rational data passed the test is shown in Figs. 35.3 and 35.4. All these data are determined as the response data. Adopt the UKF (Unscented Kalman Filter) algorithm to calculate the target trajectory using the classification result of all measuring stations. The height curves of the target trajectory are shown in Figs. 35.5 and 35.6 respectively. From these figures, we can

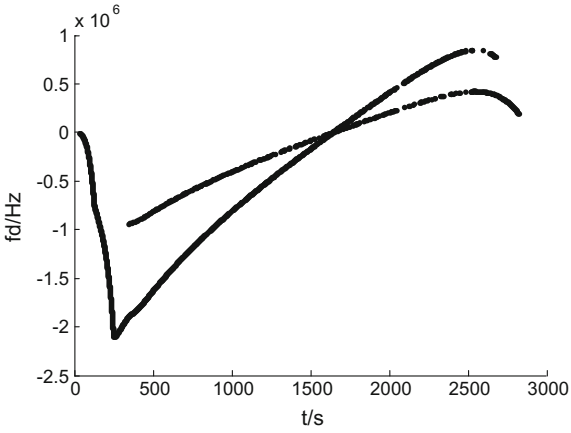


Fig. 35.3 The response data of the effective data 1 classified by the existing method

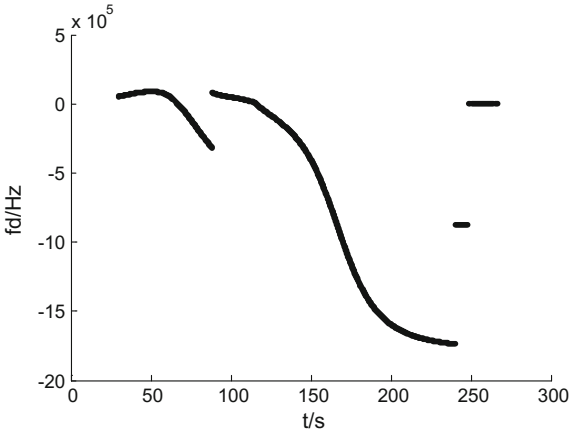


Fig. 35.4 The response data of the effective data 2 classified by the existing method

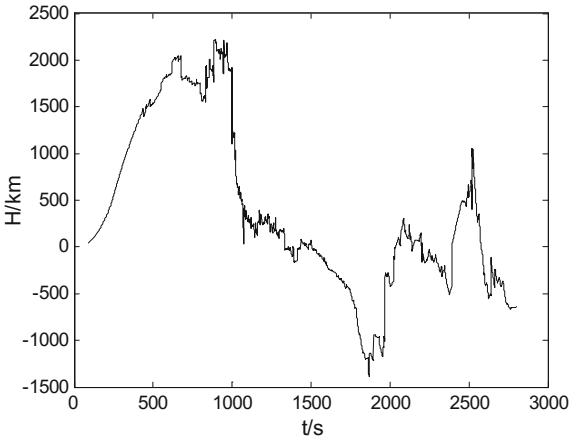


Fig. 35.5 The height curve of target trajectory calculated using the effective data 1 classification result with the existing method

see that the existing method leads serious errors in the trajectory calculation. The calculated trajectory presents sharp ups and downs, even divergence.

35.4 The New Classification Algorithm

Through the analysis of the existing method, there are two main defects leading that the method cannot classify the effective measuring data correctly. First, the method can't recognize the continuous beacon data. Second, in the case of two master stations relay measurement, when the actual switch time doesn't conform to that designed switch time in the pre-populated scheme, the master station for the

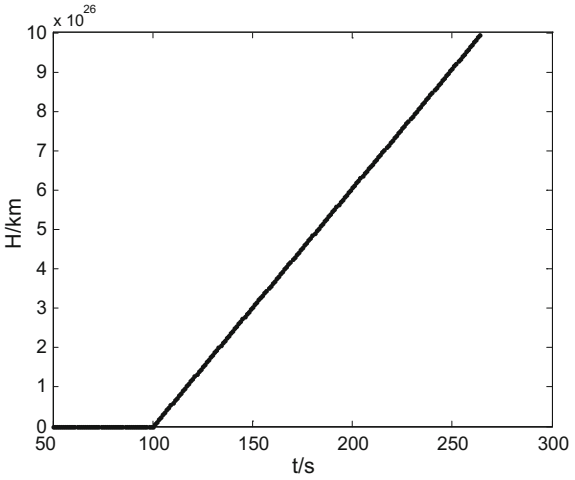


Fig. 35.6 The height curve of target trajectory calculated using the effective data 2 classification result with the existing method

response data in the relay section cannot be matched correctly. To solve these two problems, we must find an appropriate classification reference.

35.4.1 Classification Reference

The available classification references in the real-time data processing are the theoretical trajectory, the calculated trajectory at the historical moment and the real measuring data at the historical moment. Considering that classifying the data in the pre-processing phase is independent of calculating the trajectory and can avoid to generate coupling errors, the new classification algorithm chooses the theoretical trajectory and the real measuring data at the historical moment as the classification reference.

In the case of flight test successful, the actual flight trajectory is consistent with theoretical trajectory. The deviation between them increases with the flight distance. Even in the case of flight test unsuccessful, the two kinds of trajectory are still consistent at the early stage.Therefore, the theoretical trajectory is much the same as the actual trajectory in the early phase. So we can use the calculated theoretical response and beacon measurement according to the theoretical trajectory as the classification references.

In the selected coordinate system, assume that the position coordinate of the master station is $[x_m, y_m, z_m]$, the position coordinate of the received station is $[x_r, y_r, z_r]$, the theoretical trajectory coordinate at the moment t_i is $[x_i, y_i, z_i, \dot{x}_i, \dot{y}_i, \dot{z}_i]$, the downlink signal frequency is f_D, the velocity of light in the air is c. Then the theoretical beacon measuring data f_b(the Doppler frequency shift) is calculated as shown in formula (35.1) [8]

$$\begin{cases} f_b = \frac{v_r}{c - v_r} f_D \\ v_r = \frac{(x_i - x_r)}{R_{ri}} \dot{x}_i + \frac{(y_i - y_r)}{R_{ri}} \dot{y}_i + \frac{(z_i - z_r)}{R_{ri}} \dot{z}_i \\ R_{ri} = \sqrt{(x_i - x_r)^2 + (y_i - y_r)^2 + (z_i - z_r)^2} \end{cases} \tag{35.1}$$

And the theoretical response measurement data f_r(the Doppler frequency shift) is calculated as shown in formula (35.2) [9]

$$\begin{cases} f_r = \frac{v_m + v_r}{c - v_r} f_D \\ v_m = \frac{(x_i - x_m)}{R_{mi}} \dot{x}_i + \frac{(y_i - y_m)}{R_{mi}} \dot{y}_i + \frac{(z_i - z_m)}{R_{mi}} \dot{z}_i \\ R_{mi} = \sqrt{(x_i - x_m)^2 + (y_i - y_m)^2 + (z_i - z_m)^2} \end{cases} \tag{35.2}$$

The v_r calculation in formula (35.2) is identical with that in formula (35.1).

The effective data 1 and effective data 2 contrast with the theoretical measurement calculated according to the theoretical trajectory are show in Figs. 35.7 and 35.8 respectively. In Fig. 35.8, the theoretical response data 1 means the response data corresponding to the master station 1, and the theoretical response data 2 corresponding to the master station 2.

There are two questions to be solved using the above reference. One of them lies in the late flight phase, when the deviation between the reference and the real data is quite large. The suitable threshold to carry out the rationality test is difficult to be pre-populated. The other lies in the phase when the response reference and the beacon reference cross each other. In these special phases, classifying the real data according to a single moment distance between the real data and the references can also lead mistakes.

From Figs. 35.7 and 35.8, we can find that using a consecutive data sequence instead of a single data to perform data classification is much easier. Because the

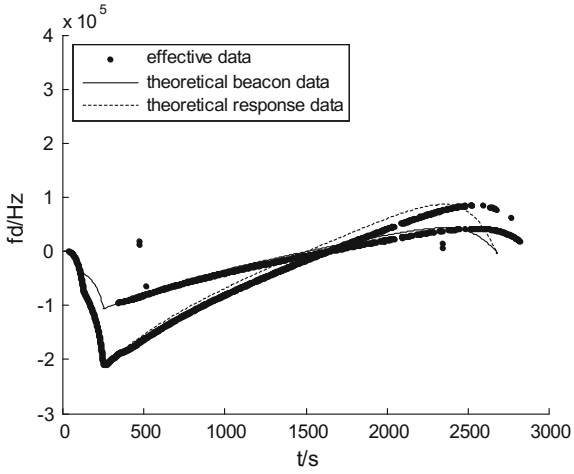


Fig. 35.7 The effective data 1 contrast with the theoretical references

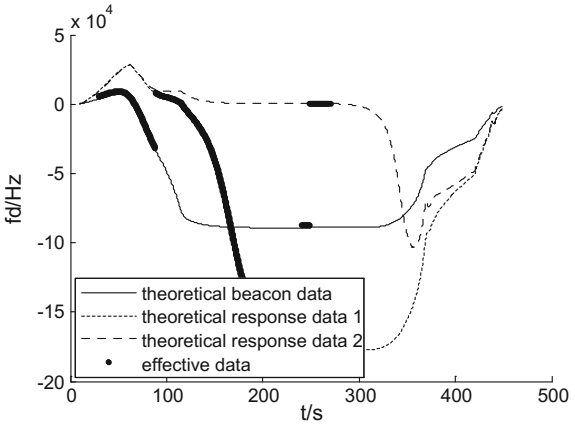


Fig. 35.8 The effective data 2 contrast with the theoretical references

measuring data has strong correlation in time. In the late phase for long range flight, although the deviation between the theoretical reference and the real data is quite large, the curve trend is consistent. So based on the time correlation, the measuring data at the historical moments can be used as an another new auxiliary classification reference to improve the classification performance.

35.4.2 Algorithm Flowchart

The basic idea of the new classification algorithm proposed in this paper is: in the initial stage of data classification, because the deviation between the actual trajectory and the theoretical trajectory is very small, the calculated theoretical measurement according to the theoretical trajectory can be used as the classification reference to classify the real data passing the rationality test in the nearest principle. When a certain points of the classified data have been accumulated, the classified response data and beacon data at the historical moments can be used as a new classification reference. When the deviations between the real data and the classification references are all larger than the designed threshold, the real data is determined as the abnormal data. The specific algorithm flowchart is as follows:

(1) Find an initial five-point reasonable data sequence using the fourth-order difference test method. Mark a five-point sequence as $f_{j-4}, f_{j-3}, f_{j-2}, f_{j-1}, f_j$. Its fourth-order difference are as shown in formula (35.3)

$$\Delta^4 f_j = f_{j-4} - 4f_{j-3} + 6f_{j-2} - 4f_{j-1} + f_j \tag{35.3}$$

Assume the given threshold is M_1. If $|\Delta^4 f_j| \le M_1$, the sequence is determined as a reasonable data sequence; otherwise, let $j = j + 1$, and go on the fourth-order difference test until a reasonable data sequence is found.

(2) Calculate the theoretical references using the theoretical trajectory at the measuring moment t_j corresponding to f_j with the formula (35.1) and (35.2). Mark the calculated theoretical beacon data as f_{bj} and the calculated theoretical response data as f_{rj}. If $|f_j - f_{rj}| \le |f_j - f_{bj}|$, the real data f_j is classified into the response data. Otherwise, the real data f_j is classified into the beacon data.

(3) For the any subsequent sampling moment t_k, mark the corresponding measuring data as f_k, the number of historical data has been classified into the beacon data category as i, the number of historical data has been classified into the response data category as j, the given data accumulation length as N. Then calculate the two kinds of classification reference as the following:

(3a) When $i < N$, use the theoretical beacon data $\hat{f}_{bk}$ as the beacon classification reference in the formula (35.1); when $i \ge N$, assume the latest N historical beacon data sequence is $f_{i-N+1}, f_{i-N+2}, \ldots, f_i$, the relevant measuring moment sequence is $t_{i-N+1}, t_{i-N+2}, \ldots, t_i$, fitting the sequence into a straight line with the parameters calculated in the formula (35.4)

$$\begin{cases} \hat{a} = \dfrac{\sum\limits_{j=N-1}^{0} t_{i-j} \sum\limits_{j=N-1}^{0} f_{i-j} - N \sum\limits_{j=N-1}^{0} t_{i-j} f_{i-j}}{\left(\sum\limits_{j=N-1}^{0} t_{i-j}\right)^2 - N \sum\limits_{j=N-1}^{0} t_{i-j}^2} \\ \hat{b} = \dfrac{\sum\limits_{j=N-1}^{0} f_{i-j} - \hat{a} \sum\limits_{j=N-1}^{0} t_{i-j}}{N} \end{cases} \tag{35.4}$$

The prediction for the current measuring moment is gained in the formula (35.5)

$$\hat{f}_{bk} = \hat{a} t_k + \hat{b} \tag{35.5}$$

Use the prediction data $\hat{f}_{bk}$ as the beacon classification reference.

(3b) When $j < N$, use the theoretical response data $\hat{f}_{rk}$ as the response classification reference in the formula (35.2); when $j \ge N$, assume the latest N historical response data sequence is $f_{j-N+1}, f_{j-N+2}, \ldots, f_j$, the relevant measuring moment sequence is $t_{j-N+1}, t_{j-N+2}, \ldots, t_j$, fitting the sequence into a straight

line with the parameters calculated in the formula (35.4). Use the prediction data $\hat{f}_{rk}$ gained in the formula (35.5) as the response classification reference.

(4) Assume the given threshold is f_{th}, classify the real data at the current measuring moment f_k in the nearest principle: if $|f_k - \hat{f}_{bk}| > f_{th}$ and $|f_k - \hat{f}_{rk}| > f_{th}$, f_k is determined as an abnormal data; if $|f_k - \hat{f}_{bk}| < |f_k - \hat{f}_{rk}|$, f_k is determined as a beacon data; if $|f_k - \hat{f}_{rk}| < |f_k - \hat{f}_{bk}|$, f_k is determined as a response data.

(5) For the situation of two master stations relay work, the algorithm flowchart is much the same as the steps (1)–(4). The only difference is that three kinds of classification references are needed: the beacon reference ($\hat{f}_{bk}$), the response reference corresponding to the master station 1 ($\hat{f}_{rk1}$) and the response reference corresponding to the master station 2 ($\hat{f}_{rk2}$). If $|f_k - \hat{f}_{bk}| > f_{th}$ and $|f_k - \hat{f}_{rk1}| > f_{th}$ and $|f_k - \hat{f}_{rk2}| > f_{th}$, f_k is determined as an abnormal data. Otherwise, if $|f_k - \hat{f}_{bk}|$ is minimal, f_k is determined as a beacon data; if $|f_k - \hat{f}_{rk1}|$ is minimal, f_k is determined as a response data corresponding to the master station 1; if $|f_k - \hat{f}_{rk2}|$ is minimal, f_k is determined as a response data corresponding to the master station 2.

35.5 Real Data Validation

The new classification algorithm is validated using the effective data 1 and effective data 2. The classification results are shown in Figs. 35.9 and 35.10 respectively. The real-time trajectory height curves calculated with the classified data are shown in Figs. 35.11 and 35.12 respectively. Visible from these figures, the new algorithm

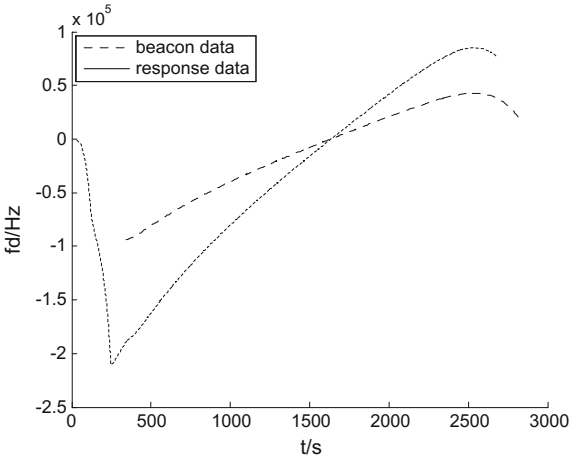


Fig. 35.9 The classification result of effective data 1 using the new algorithm

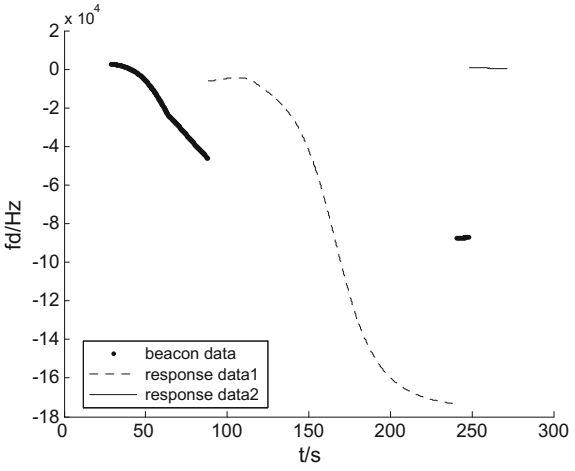


Fig. 35.10 The classification result of effective data 2 using the new algorithm

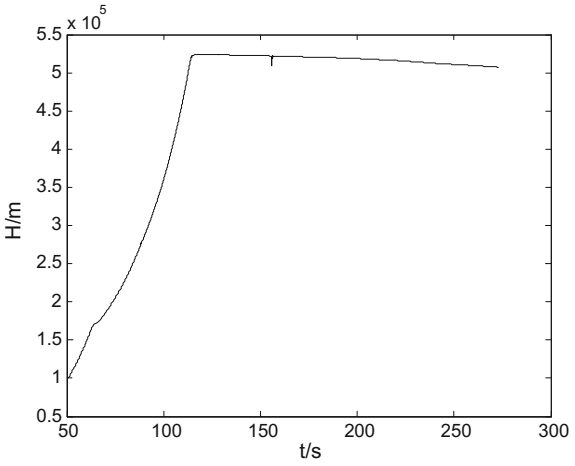


Fig. 35.11 The height curve of target trajectory calculated using the effective data 1 classification result with the new algorithm

has good classification effectiveness and can classify the real measuring data into the response data, the beacon data and the abnormal data correctly. On this basis, the real-time trajectory calculated from the classified data is smooth and continuous. For the flight task corresponding to the effective data 1, the time length of the correct real-time trajectory increases to nearly 3000 s from 180 s, which significantly improve the utilization of the measuring data gained by the multi-velocity measuring system. For the flight task corresponding to the effective data 2, the problem of the target trajectory jumping largely in the relay segment is solved, resulting a smooth and continuous real-time trajectory in the whole measuring time duration.

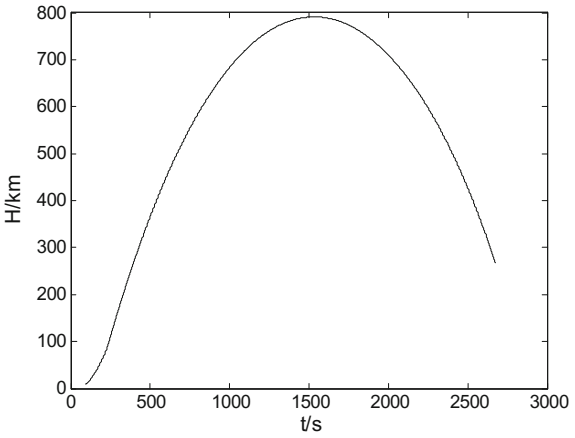


Fig. 35.12 The height curve of target trajectory calculated using the effective data 2 classification result with the new algorithm

35.6 Conclusion

This paper aims at the real-time data classification problem for the multi-velocity measuring system configured to the response measurement mode. After analyzing the defects of the existing method, a new classification algorithm is proposed. The new classification algorithm involves two different kinds of classification references: the theoretical trajectory calculated references at the beginning and the classified historical measuring data after a certain accumulation of data number. Based on the suitable classification references, the real measuring data could be classified into three categories: the response data, the beacon data and the abnormal data. Using two typical kinds of measuring data corresponding to one master station pattern and two master stations relay pattern to validate the performance and flexibility of the new algorithm. The results show that the algorithm can classify the three kinds of data correctly and suit the different situations. On this basis, the real-time trajectory calculated is smooth and continuous. Therefore, the new classification algorithm will play an important role in the practical data processing of the multi-velocity measuring system to improve the trajectory precision [10].

References

1. Guo JH, Wu ZR, Huang XD (2002) Trajectory measurement system research on multi-velocity radar. J Spacecr TT&C Technol 21(3):5–11
2. Zhong SH, Hu SB, Wang ZJ (2015) The beacon measuring data processing of continuous wave Doppler radar. Telecommun Eng 2015(11):1244–1248
3. Shang LF, Xing KY, Gan YY (2012) The real-time data processing for maneuvering target tracking under the beacon mode. Control Theory Appl 29(1):59–63
4. Li CF, Liu LJ, Zhu SY (2015) The dual-channel data fusion method of Doppler radar in the manned space mission. J Spacecr TT&C Technol 34(6):505–511

5. Cui SH, Hu SL, Song WX (2012) The optimal trajectory estimation method and application of multi-velocity measuring system. J Proj Rockets Missiles Guid 32(4):215–218
6. Shang LF, Shao YP, Zhou JG (2009) The real-time trajectory calculating method and application in multi-velocity measuring system. Missile Test Technol 003:63–66
7. Cui SH, Song WH, Min Wang (2011) The data fusion algorithm research and application in multi-velocity measuring system. J Proj Rockets Missiles Guid 31(5):161–164
8. Li XG, Jia XQ, Shi ZR (2006) The realization and demonstration of the velocity measuring method applying the beacon mode. J Spacecr TT&C Technol 25(2):37–43
9. Liu ZW, Pan JP, Zhang XC (2013). The new data recovery method of the response data in multi-velocity measuring system. J Telem Track Command (3)
10. Liang XH, Zhu WX, Zhang Y (2012) The real-time estimation and calibration algorithm of systematic error in multi-velocity system. J Spacecr TT&C Technol 05:49–53

Chapter 36
New Algorithm for Guidance Instrument Error Separation

Hua Zhao, Jiagui Huang, Xiaohu Liang and Yuming Hua

36.1 Introduction

Hit precision is most important to missile. There are so many factors which have effect on the hit precision, but the guidance instrument error is the main factor. Some scholars pointed out that 70–80% of error of hit precision is caused by the guidance instrument error [1–7]. It is obvious that the guidance instrument error is determined by guidance instrument, However the quality of guidance instrument is affected by many technology field such as metallurgy manufacture mechanism Electronic Industries and so on, There are so many challenges in improving the quality only by hardware with the more request in hit precision. So many scholars focus on how to separate and compensate guidance instrument error by mathematical method.

Now guidance instrument error separation is based on the model $\Delta \boldsymbol{W} = \boldsymbol{SC}$. In this model, the environmental matrix $\boldsymbol{S}$ is usually ill-conditioned. In order to obtain the accuracy result, the most important thing is how to decrease the ill-conditioned of the matrix. In the past years, many scholars have researched some error separation method [8–17]. One kind of methods emphasizes the application of prior information, such as Bayes method, ridge estimation, hyper-ellipsoid model, PB method. The other kind of methods is non-prior information methods, such as PCA (Principal Component Analysis), modified PCA, characteristic roots estimate and so on. The accuracy of the first kind of methods depends on the accuracy of the Prior Information. The higher the accuracy of the prior information is, the higher the accuracy of the result of the error separation is. The other kind of methods doesn't

H. Zhao (✉) · J. Huang · X. Liang · Y. Hua
Beijing Institute of Tracking and Telecommunication Technology, 100094 Beijing, China
e-mail: zhaohcq@163.com

J. Huang · X. Liang · Y. Hua
Key Laboratory of Space Object Measurement, 100094 Beijing, China

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_36

depend on prior information, they firstly analyze the characteristic of Matrix $\boldsymbol{S}^{\mathrm{T}}\boldsymbol{S}$, then choose some principal component with some rules, finally reconstruct the environmental matrix and calculate the guidance instrument error. By this way the ill-condition of the matrix is decreased, the higher accuracy will be obtained. But how to choose the principal component is a big problem.

In the recent years, some scholars find that solving the linear ill-conditioned algebraic equation is equivalence to solving a stiff dynamic system [18]. According to this idea, a new method is given to separate the guidance instrument error.

36.2 Background

The guidance instrument error model is as follows:

$$\begin{cases} \Delta \boldsymbol{W} = \boldsymbol{S}\boldsymbol{C} + \boldsymbol{\eta} \\ E\boldsymbol{\eta} = 0, D\boldsymbol{\eta} = \Sigma \end{cases} \tag{36.1}$$

Where

$$\Delta \boldsymbol{W} = \begin{bmatrix} \Delta \boldsymbol{W}_1 \\ \vdots \\ \Delta \boldsymbol{W}_N \end{bmatrix}, \boldsymbol{S} = \begin{bmatrix} \boldsymbol{S}_1 \\ \vdots \\ \boldsymbol{S}_N \end{bmatrix} \tag{36.2}$$

$\Delta \boldsymbol{W}_i, i = 1, \ldots, N$ is the *ith* difference vector between telemetering apparent velocity and exterior measuring apparent velocity in inertial frame, they are 1×3 vector. $\boldsymbol{S}_i, i = 1, \ldots, N$ is the *ith* Environmental Matrix,they are $3 \times n$ matrix. $\boldsymbol{C}$ is $n \times 1$ vector, which is the guidance instrument error will be estimated. N is data number, n is the number of the parameters and $N > n$. In order to get $\boldsymbol{C}$, we need to solve the Eq. (1).

It is obvious that the Least square solution of the equation above is $\boldsymbol{C}_{LS} = (\boldsymbol{S}^{\mathrm{T}}\boldsymbol{S})^{-1}\boldsymbol{S}^{\mathrm{T}}\Delta \boldsymbol{W}$, but the matrix $\boldsymbol{S}^{\mathrm{T}}\boldsymbol{S}$ is seriously ill-conditioned, in other words, its columns vector have linear correlation, the matrix is almost singularity, the condition number of the matrix is very large. The condition number is given as follows:

$$Cond(\boldsymbol{B}) = \|\boldsymbol{B}\| * \left\|\boldsymbol{B}^{-1}\right\| \tag{3}$$

where $\|\bullet\|$ stands for the matrix norm. On this condition, small error of $\Delta \boldsymbol{W}$ will lead to a big error of result.

36.3 New Algorithm Based on the Dynamic System

Consider the following equation, where $\tilde{A}$ is symmetry and positive definite.

$$\tilde{A}\tilde{x} = \tilde{b} \tag{36.4}$$

When $\tilde{A}$ is ill-conditioned, the following iterative method will be used to solve this problem:

$$\tilde{x}_{n+1} = \tilde{x}_n + M_n(\tilde{b} - \tilde{A}x_n), n = 0, 1, 2, \ldots \tag{36.5}$$

where $\{M_n\}$ is $m \times m$ non-singular matrix sequence (36.5) will be rewritten as follows:

$$\begin{cases} x_{n+1} = x_n + z_n \\ z_n = M_n(b - Ax_n), \quad n = 0, 1, 2, \ldots \end{cases} \tag{36.6}$$

From equation above, this iterative improvement of solution can be viewed as explicit Euler method with step size $h = 1$ for solving the system of ordinary differential equations:

$$\begin{cases} \dfrac{dz}{dt} = M(t)(b - Az) \\ z(0) = x_0 \in R^m \end{cases} \tag{36.7}$$

where $M(t)$ is $m \times m$, and $\det(M(t)) \neq 0$, $\bar{z} = A^{-1}b$ is a unique stationary point of system. In other words,any linear iterative method is equivalence to some dynamics, so we can convert the problem of solving the linear equations to the problem of solving the corresponding dynamics system.

So Letting $A = S^{\mathrm{T}}S$, $x = C$, $b = S^{\mathrm{T}}\Delta W$ (1) can be rewritten as

$$Ax = b \tag{36.8}$$

where A is badly ill-conditioned and symmetry positive definite.

Some scholars find that the iteration method (5) is not effective when the matrix A is badly ill-conditioned [18]. In order to obtain more suitable dynamic system, new iterative will be used.

Consider the following iterative improvement of solution with parameter $u \geq 0$.

$$(uI + A)x_{n+1} = ux_n + b \tag{36.9}$$

It is easy to see that the equations can be rewritten as

$$\begin{cases} (u\boldsymbol{I}+\boldsymbol{A})\boldsymbol{y}_n = \boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}_n \\ \boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \boldsymbol{y}_n \end{cases} \tag{36.10}$$

The corresponding dynamic system is

$$\begin{cases} \dfrac{d\boldsymbol{x}}{dt} = (u\boldsymbol{I}+\boldsymbol{A})^{-1}(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{z}) \\ \boldsymbol{x}(0) = \boldsymbol{x}_0 \end{cases} \tag{36.11}$$

where $\boldsymbol{x}_0 \in R^m$.

Now the characteristics of the linear Eqs. (36.8) the dynamic system (36.11) and the iteration method (36.10) will be discussed.

Theorem 36.1 *The solution* $\boldsymbol{x}^*$ *of linear system* (8) *is a unique globally asymptotically stable equilibrium point of the dynamic system.*

Proof let

$$f(\boldsymbol{x}) = \boldsymbol{b}-\boldsymbol{A}\boldsymbol{x} \tag{36.12}$$

$$v(\boldsymbol{x}) = f^{\mathrm{T}}(\boldsymbol{x})f(\boldsymbol{x})/2 \tag{36.13}$$

Then

1. if $\boldsymbol{x} \neq \boldsymbol{x}^*$ and $v(\boldsymbol{x}) > 0$, then $v(\boldsymbol{x}^*) = 0$.
2. if $\boldsymbol{x} \neq \boldsymbol{x}^*$, because of $u > 0, \boldsymbol{A}, \boldsymbol{A}^{-1}, (\boldsymbol{I}+u\boldsymbol{A}^{-1})$ 及 $(\boldsymbol{I}+u\boldsymbol{A}^{-1})^{-1}$ are all positive definite, then

$$\begin{aligned} v'(\boldsymbol{x}) &= \frac{dv(\boldsymbol{x})}{dt} \\ &= f^{\mathrm{T}}(\boldsymbol{x})(-\boldsymbol{A})\frac{d\boldsymbol{x}}{dt} \\ &= -f^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{A}(u\boldsymbol{I}+\boldsymbol{A})^{-1}f(\boldsymbol{x}) \\ &= -f^{\mathrm{T}}(\boldsymbol{x})[(u\boldsymbol{I}+\boldsymbol{A})\boldsymbol{A}^{-1}]^{-1}f(\boldsymbol{x}) \\ &= -f^{\mathrm{T}}(\boldsymbol{x})(\boldsymbol{I}+u\boldsymbol{A}^{-1})^{-1}f(\boldsymbol{x}) < 0 \end{aligned} \tag{36.14}$$

So $v(\boldsymbol{z})$ is a strict Lyapunov function of dynamic system,Thus the unique solution of linear system is a unique globally asymptotically stable equilibrium point of dynamic systems.

Corollary 1 *Suppose that the solution of system is expressed in the form*

$$\boldsymbol{x} = \boldsymbol{x}(t, \boldsymbol{x}_0)$$

Then the following result will be obtained.

$$\lim_{t \to +\infty} \boldsymbol{x}(t, \boldsymbol{x}_0) = \boldsymbol{A}^{-1}\boldsymbol{b} \tag{36.15}$$

Proof It if follows immediately from Theorem 36.1 and the definition of the well-known Liapunov's asymptotical stability.

From above, the new iteration method can be converted to a dynamic system, and the solution of the dynamic is the solution of the algebra equation. However, it is difficult to get the analytic solution of the dynamic system. Fortunately, numerical integrations for stiff dynamic system will be employed. Consequently, the effect of solving dynamic system by numerical integrations will be analyzed. For simplicity and convenience, explicit Euler method with step size $h = 1$ is applied to system (36.11), the following formula will be obtained:

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + (u\boldsymbol{I} + \boldsymbol{A})^{-1}(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_n), n = 0, 1, 2, \ \ldots \tag{36.16}$$

In this case we have the recursion

$$\begin{cases} \boldsymbol{x}_{n+1} = (u\boldsymbol{I} + \boldsymbol{A})^{-1}(u\boldsymbol{x}_n + \boldsymbol{b}), n = 0, 1, 2, \ldots \\ \quad \boldsymbol{x}_0 = 0 \end{cases} \tag{36.17}$$

This can be regarded as a stationary iterative method with the iterative matrix

$$\boldsymbol{M} = u(u\boldsymbol{I} + \boldsymbol{A})^{-1} \tag{36.18}$$

Because $\boldsymbol{A}$ is positive definite normal matrix and $u > 0$, $u\boldsymbol{I} + \boldsymbol{A}$,$(u\boldsymbol{I} + \boldsymbol{A})^{-1}$ and $u(u\boldsymbol{I} + \boldsymbol{A})^{-1}$ are all positive normal matrices.

Theorem 36.2 *Assume that* $\sigma_i (i = 1, 2, \ldots, m)$ *are eigenvalues of positive definite normal matrix A in Linear system and*

$$0 < \sigma_1 \leq \sigma_2 \leq \cdots \leq \sigma_m \tag{36.19}$$

Then the spectral radius of the iterative method (17) *is*

$$\rho(\boldsymbol{M}) = \rho[u(u\boldsymbol{I} + \boldsymbol{A})^{-1}] = \frac{u}{u + \sigma_1} \tag{36.20}$$

And for any $u > 0$ the asymptotic rate of convergence is

$$R(\boldsymbol{M}) = -\ln \rho(\boldsymbol{M}) = -\ln(\frac{u}{u+\sigma_1}) \tag{36.21}$$

Proof Because $\sigma_i (i = 1, 2, \ldots, m)$ are the eigenvalues of $\boldsymbol{A}$ from (36.19) and u > 0, we have

$$0 < \sigma_1 + u \leq \sigma_2 + u \leq \cdots \leq \sigma_m + u \tag{36.22}$$

are the eigenvalues of $u\boldsymbol{I} + \boldsymbol{A}$. Therefore, $\frac{1}{u+\sigma_i}, i = 1, 2, \ldots, m$ are the eigenvalues of $(u\boldsymbol{I} + \boldsymbol{A})^{-1}$ and

$$\frac{1}{u+\sigma_m} \leq \frac{1}{u+\sigma_{m-1}} \leq \cdots \leq \frac{1}{u+\sigma_1} \tag{36.23}$$

And, $\frac{u}{u+\sigma_i}, i = 1, 2, \cdots, m$ are the eigenvalues of $u(u\boldsymbol{I} + \boldsymbol{A})^{-1}$ and

$$\frac{u}{u+\sigma_m} \leq \frac{u}{u+\sigma_{m-1}} \leq \cdots \leq \frac{u}{u+\sigma_1} \tag{36.24}$$

Thus

$$\rho(\boldsymbol{M}) = \rho(u(u\boldsymbol{I} + \boldsymbol{A})^{-1}) = \max_{1 \leq i \leq m} \frac{u}{u+\sigma_i} = \frac{u}{u+\sigma_1} \tag{36.25}$$

And asymptotic rate of convergence of (36.17) is as follows

$$R(\boldsymbol{M}) = -\ln \rho(\boldsymbol{M}) = -\ln(\frac{u}{u+\sigma_1}) \tag{36.26}$$

Corollary 2 *The iterative method (36.16) of solution is convergent unconditionally for any $u > 0$.*

Proof For any $u > 0$, $\rho(\boldsymbol{M}) < 1$, because of Eq. (36.25) in Theorem 36.2,the conclusion is obvious.

Theorem 36.3 *Suppose $0 < \sigma_1 \leq \sigma_2 \leq \cdots \leq \sigma_m$ are eigenvalues of the positive definite normal matrix $\boldsymbol{A}$ in (36.8). Then for any $u > 0$ we always have*

$$\kappa(u\boldsymbol{I} + \boldsymbol{A}) < \kappa(\boldsymbol{A}) \tag{36.27}$$

where, $\kappa(\bullet)$ express the spectral number.

Proof The spectral condition number of $\boldsymbol{A}$ and $u\boldsymbol{I}+\boldsymbol{A}$ are

$$\kappa(\boldsymbol{A}) = \frac{\sigma_m}{\sigma_1} \tag{36.28}$$

$$\kappa(u\boldsymbol{I}+\boldsymbol{A}) = \frac{u+\sigma_m}{u+\sigma_1} \tag{36.29}$$

since $u > 0, \sigma_m > \sigma_1 > 0, \varphi'(x) = \frac{\sigma_1-\sigma_m}{(x+\sigma_1)^2}, x \geq 0$ so

$$\varphi(x) = \frac{x+\sigma_m}{x+\sigma_1}, x \geq 0 \tag{36.30}$$

we can get $\frac{u+\sigma_m}{u+\sigma_1} < \frac{\sigma_m}{\sigma_1}$, 即 $\kappa(u\boldsymbol{I}+\boldsymbol{A}) < \kappa(\boldsymbol{A})$.

So from the theorem above, we can know that for any $u > 0$, the spectral number of the new iterative improvement (36.17) is less than the spectral condition of the original problem.

Obviously, the ill-conditioned linear algebra equation can be converted to a stiff dynamic system. But the explicit method is not suitable to solve the stiff dynamic system, so we consider one-step second derivative extended backward differentiation methods of order four as follows:

$$\alpha_0 \boldsymbol{y}^n + \alpha_1 \boldsymbol{y}^{n+1} = h\beta_1 f^{n+1} + h\beta_2 f^{n+2} + h^2\gamma_1 g^{n+1} + h^2\gamma_2 g^{n+2} \tag{36.31}$$

where $g = f' = -\boldsymbol{A}f$, the coefficients α_i, β_i, γ_i are the parameters.

$$\begin{cases} \alpha_0 = -1 \\ \alpha_1 = \dfrac{1}{12} \\ \beta_1 = -\dfrac{1}{2} \\ \beta_2 = \dfrac{3}{2} \\ \gamma_1 = -\dfrac{17}{12} \\ \gamma_2 = -\dfrac{17}{12} \end{cases} \tag{36.32}$$

Since method (36.31) is implicit, f^{n+2} and g^{n+2} are unknown, so we need apply explicit method to calculate them in advance. Here the classical Runge-Kutta method of four steps and four order is used.

Table 36.1 Real value of guidance instrument error

Name of the error	Value of the error	Name of the error	Value of the error	Name of the error	Value of the error
C1	4.9043×10^{-4}	C10	4.8489×10^{-7}	C19	9.6962×10^{-5}
C2	4.9043×10^{-4}	C11	4.8489×10^{-7}	C20	2.1818×10^{-4}
C3	4.9043×10^{-4}	C12	4.8489×10^{-7}	C21	9.6950×10^{-5}
C4	3.9996×10^{-5}	C13	7.5459×10^{-8}	C22	2.9659×10^{-6}
C5	7.0002×10^{-5}	C14	8.6561×10^{-8}	C23	2.9659×10^{-6}
C6	9.9776×10^{-5}	C15	7.4157×10^{-8}	C24	2.9659×10^{-6}
C7	5.1096×10^{-7}	C16	7.5228×10^{-9}	C25	2.9659×10^{-6}
C8	5.0982×10^{-7}	C17	7.4712×10^{-9}	C26	2.9659×10^{-6}
C9	1.0055×10^{-5}	C18	7.4597×10^{-9}	C27	2.9659×10^{-6}

36.4 Numerical Illustrations

36.4.1 Conditions of Numerical Illustration

In order to verify the effective of the methods, numerical experiment will be carried out. The simulation conditions are shown below.

1. Guidance instrument error model parameter

The Guidance instrument error model used in this paper has 27 parameters,the value of these parameter is shown in Table 36.1.

2. The error of difference vector of between telemetering apparent velocity and exterior measuring apparent velocity

There are three kinds of errors used in the numerical experiment, they are 0.01, 0.02 and 0.03 m/s respectively.

3. Numerical methods

Numerical methods mentioned in the Sect. 36.3 are used to solve the problem.

36.4.2 Result of Numerical Illustration

In order to verify the effective of the method presented in this paper, the PB method is used to compare with it. The numerical Illustration shown as Figs. 36.1, 36.2, 36.3.

From the picture above, it is easy to see that PB method can separate a majority of the guidance instrument error, but the 1th, 2th, 3th, 6th and 20th guidance instrument errors have a big deviation with the theoretical value. And with the growth of apparent velocity error, the differences between these five guidance

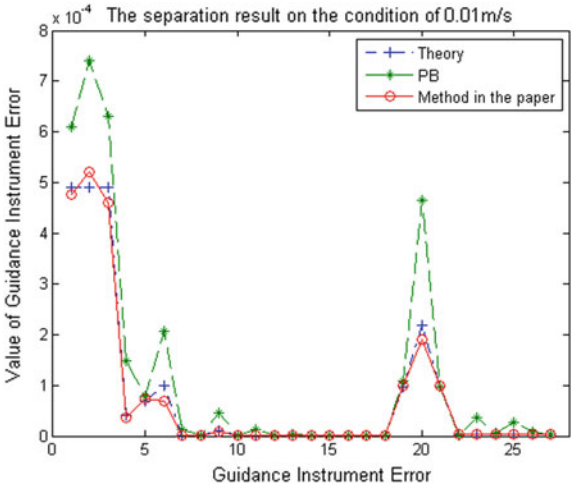


Fig. 36.1 The separation result on the condition of error 0.01 m/s

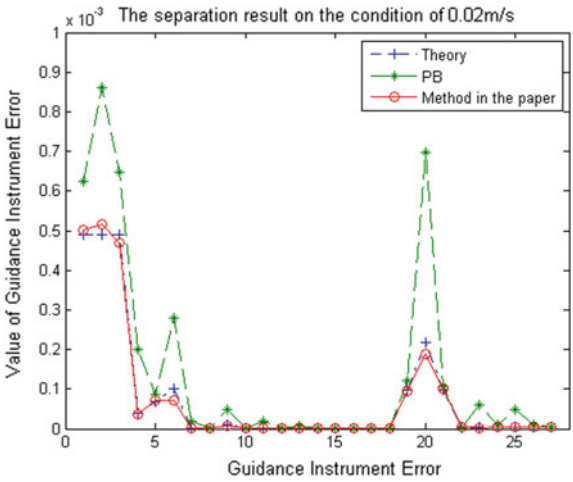


Fig. 36.2 The separation result on the condition of error 0.02 m/s

instrument errors and the theoretical value is becoming more and more big. This leads to that the other 22 errors are also becoming bigger. Compared with the PB method, the separation result by the method proposed in this paper have good consistency with the theoretical value almost in all guidance instrument error. Even if the apparent velocity error becomes bigger, the error of separation result is also acceptable. So the method in this paper decreases the ill-conditioned of the environmental matrix, it is more robust and the separation result is more accuracy.

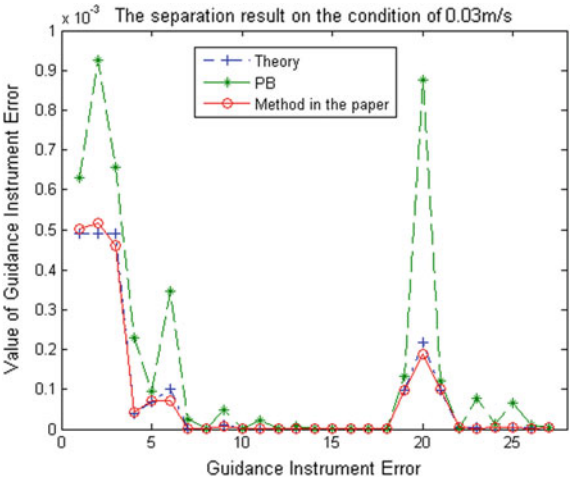


Fig. 36.3 The separation result on the condition of error 0.03 m/s

36.5 Conclusions

Hit precision is a very important index of missile. It is a useful way to improve the hit precision of the missile by separating the guidance instrument error and compensating it. In this article, a method base on solving the stiff dynamitic system is presented. The numerical result shows that the method is more effective than the traditional method.

References

1. Yong Xian, Yi Zhang (1999) Analysis on effect of a missile's guidance instrument error on fire accuracy[J]. Flight, Dyn 17(2):82–86
2. Jing Yao, Xiaojun Duan, Haiyin Zhou (2005) Modeling and parameters estimation of marine guidance instrumentation systematic error[J]. J Ballist 17(1):33–39
3. Jian Li, Shifeng Zhang (2008) An analysis of the effect of guidance instrument error on fall point accuracy and accuracy estimate[J]. J Proj Rockets, Missil Guid 28(1):67–71
4. Yuzhen Xie, Zhenxing Li (2013) Influence factors of guidance instrumentation system error linear model estimation[J]. Aerosp Control 31(5):46–49
5. Yuzhen X (2012) The impact of origin error on separating guidance instrument systematic error. Digit Commun[J] 4:48–50
6. Zhengdong Hu, Huabo Yang, Shifeng Zhang et al (2006) Research on separation of initial error and guidance instrumentation error in submarine launched missile[J]. J Project Rockets, Missil Guid 26(3):46–49
7. Lin Yuan (2013) Separation of systematic errors of guidance instruments of moving base spacecraft in geocentric coordinate system[J]. J Spacecr TT&C Technol 32(1):80–83
8. Yuzhen Xie (2013) The research of combining ridge and principal components estimate in separating guidance instrument systematic error[J]. J Project Rocket Missil Guid 33(3):189–191
9. Wenjun Wang, Xiaojun Duan, Jubo Zhu (2013) Adaptive parameter regularization model for guidance instrumental error separation and its Application[J]. J Ballist 25(3):28–33

10. Lin Yuan, Zhaogang Wang, Chengzhu Han (2012) The model optimization method for the guidance instrument systematic error separation[J]. Digit Commun 2:51–53
11. Jie He, Xianlin Huang, Hongbo Jia (2006) Real-time and segmented error compensation of guidance instrument for reentry aircraft[J]. J Chin Inert Technol 14(2):29–32
12. Yuan Y, Wang M (2006) Research on compensation methods of guidance instrument error based on analysis of environmental function[J]. J Project Rockets[J] Missil Guid 26(3): 279–282
13. Jie He, Xianlin Huang, Qinghua Zhu (2007) Application of improved latent regression to SINS guidance instrument error estimation[J]. J Chin Inert Technol 15(3):282–285
14. Hu Xiaoxiang Hu, Changhua Zheng Jianfei et al (2007) On evaluating method of guidance instrument error[J]. Control Eng China 14(S2):14–16
15. Huabo Yang, Shifeng Zhang, Hong Cai (2007) Separation and conversion for guidance instrumentation systematic error using support vector machines regression[J]. J Syst Simul 19 (10):2177–2180
16. Dekun Xu, Huabo Yang, Shifeng Zhang et al (2007) A genetic principal components for conversion of guidance instrumentation system error[J]. Aerosp Control 25(6):22–26
17. Weijun Song (2008) Influence factor analysis of converted accuracy for guidance instrumentation systematic error[J]. Aerosp Control 26(4):13–15
18. Wu X, Shao R, Zhu Y (2002) New iterative improvement of a solution for an ill-conditioned system of linear equation based on a linear dynamic system[J]. Comput Math Appl 44: 1109–1116

Chapter 37
Ionosphere's Effect on the Demodulation BER Performance of a DS/FH-BPSK Signal

Liyi He, Xiao Chen, Jinhai Sun and Junfu Chen

37.1 Introduction

The ionosphere is the atmosphere of the earth that range from 50 to 1000 km, which is ionized due to sun radiation and the cosmic rays. The transmission speed of radio waves is affected by the ionized electrons and the refraction, reflection and scattering and rotating of the polarization of the radio waves occur. To the radio wave with a higher frequency than ultra-high frequency (UHF) the ionosphere dispersion effect can be modeled as phase advance and group delay, which deteriorates the BER performance of BPSK signals.

Hybrid direct sequence and frequency hopping (DS/FH) spread spectrum is a new satellite telemetry signal scheme with a much wider signal band than that of traditional DSSS signal. BPSK is a commonly-used modulation in satellite telemetry and communication system and its BER of information transmission is up to the mathematical distribution of the recovered signal phase. Due to a higher spread-spectrum ratio, the signal channel dispersion effect is severer and worsens the BER performance.

The sources of phase estimation error include the noise and phase scintillation caused by the channel. Some [1] investigated the QPSK BER under a white Gaussian noise and a phase reference source obeying Tikhonov distribution, derived an approximate expression of the BER and performed data simulation to verify the accuracy of the expression. Corvaja and Pupolin [2] sorted the phase noise out as white phase noise, white frequency noise and flicker, analyzed the effect of phase noise spectrum on the BER of a DPSK receiver. Taggart and Kumar [3] found that

L. He · X. Chen (✉) · J. Sun · J. Chen
Rocket Force University of Engineering, Xi'an 710038, China
e-mail: ty_chenxiao@163.com

X. Chen
PLA Unit 96813, Huangshan 245400, China

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_37

the square rooted raised cosine filter had little impact on the PDF of the receiver's phase noise, pointed out that phase noise causes the decision regions to overlap and increases BER. Nie et al. [4] considered the impact of timing and carrier phase recovering error on the demodulation BER, where the phase noise is mainly caused by the clock source and the receiving circuits. Zhu [5] concluded that the phase noise results in the rotation of QPSK constellation and worsens the BER of receiver and performed BER simulations of QPSK signal.

In a DSSS telemetry system, the signal band is narrow and the dispersion effect is not obvious [6], however, to the DS/FH signal with a wider band, the non-linearity of ionosphere introduces phase noise and degrades the BER severely. Based on the works mentioned above, this paper discusses the ionosphere dispersion's effects on the BER of a DS/FH BPSK signal. Firstly, the BPSK BER of pilot signal with a fixed FH pattern under ionosphere dispersion is discussed. Secondly, the BPSK BER of long-term signal with a varying FH pattern is analyzed.

37.2 Signal Model

The hybrid DS/FH spread spectrum signal discussed here is a phase-coherent and fast-frequency-hopping BPSK signal. In this telemetry system, the signal is consisted of a pilot signal and a long-term signal. The pilot signal adopts a block hopping FH pattern, in which the cycle of FH pattern equals the time span of one BPSK modulation symbol. Let the FH pattern and pseudo random code change chronically and the cycle length be N_{hop} hops, so the hybrid DS/FH telemetry signal in one cycle can be written as

$$s(t) = \sqrt{2S}\sum_{i=1}^{N_{\text{hop}}} D(t)\text{Re}\left[c(t)e^{j2\pi f_i t}\right] \tag{37.1}$$

Let S be the average power of the transmitted signal, f_i be the hopping radio frequency, T_{h} be the length of one hopping frequency, $N_{\text{hop}}T_{\text{h}}$ be one frequency hopping cycle, $c(t)$ be the DS code with a chip rate f_{c}, Re[·] represent the real part of a complex and $D(t)$ represent BPSK data, the hybrid DS/FH signal can be shown graphically in Fig. 37.1.

To ensure that the phases in each T_{h} hold the same initial value, the FH pattern is designed as below

$$\forall i = 1, 2, 3, \ldots, N_{\text{hop}}, \quad f_i = \frac{k}{T_h} + f_{RF}, \quad k \in \{1, 2, \ldots, K\}, \tag{37.2}$$

where k is the hopping code, f_{RF} is the frequency of the radio local oscillator, the hopping bandwidth is $B_{\text{hop}} = K/T_{\text{h}}$, K is the total number of FH points, N_{hop} is the number of FH points chosen in the pilot signal, and $N_{\text{hop}} \ll K$. To the pilot signal, the DS code has a cycle of M chips with chip length T_c and M satisfies

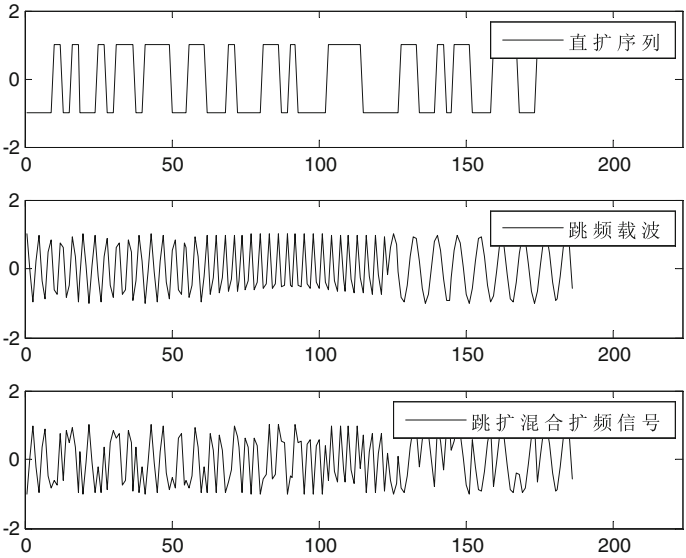


Fig. 37.1 Time domain waveform of a hybrid DS/FH signal

$$MT_c = N_{\text{hop}} T_h. \tag{37.3}$$

The long-term DS/FH signal also uses both a chaotic sequence as the DS code and another chaotic sequence as the FH code. The difference is that in multiple BPSK symbols of pilot signal there exist one cycle of both the FH code and the DS code, whereas the BPSK symbols of long-term signal contains different FH schemes. This difference leads to the difference of BER of these two signals, which is discussed in the later part of this paper.

In order to investigate the demodulation performance of DS/FH BPSK signal, first we study the convergence of code tracking. According to [7], under ionosphere dispersion, the ionosphere gives the correlation peak a time delay through the total electron content (TEC) as

$$\tau_{IONO} = \frac{\lambda_{TEC} TEC}{2\pi f_{RF,Cent}^2}. \tag{37.4}$$

$f_{\text{RF,cent}}$ represents the mean of the FH frequencies in one single BPSK symbol. Assume the phase discrimination speed equals the BPSK symbol rate, the output of the phase discriminating function in the carrier loop can be written as

$$\Theta_{\text{r}} = \tan^{-1}\left(\frac{Q_P}{I_P}\right) = \arg\left[R_{IONO}(\tau) e^{jData \cdot \pi}\right], \tag{37.5}$$

where $I_P = \text{real}(R_{IONO}(\tau)e^{jData\cdot\pi})$, $Q_P = \text{imag}(R_{IONO}(\tau)e^{jData\cdot\pi})$, and arg[·] gives the argument of a complex and *Data* is the information bit, either "0" or "1". Using e_{TEC} to denote the value of TEC, the output of carrier phase discriminator in each symbol can be represented as

$$\Theta_{\text{r}} = \arg\left[\sum_{i=0}^{N_{hop}-1} e^{2\pi f_i\left(\frac{\lambda_{TEC}e_{TEC}}{2\pi f_i^2}-\tau_{IONO}\right)} + n_i\right], \tag{37.6}$$

where n_i is the correlation of a FH wave and white noise in the i-th hop, here we have

$$n_i = \int_{(i-1)T_h+\tau_{IONO}}^{iT_h+\tau_{IONO}} e^{2\pi f_i(t-\tau_{IONO})}\cdot n_{cmplx}\,\mathrm{d}t. \tag{37.7}$$

Then we investigate the BER of BPSK through studying Eq. (37.6). Firstly, we study the convergence condition of the phase discrimination output. According to Eqs. (37.4) and (37.6), we find the convergence is related to the FH pattern. When the f_is in consecutive symbols are identical, or, they conform to a uniform distribution, the τ_{IONO}s and Θ_rs in consecutive symbols stay the same, otherwise the τ_{IONO}s and Θ_rs suffer scintillations when the symbols elapse, which affects the demodulation performance. Thus there are two situations regarding the convergence of the output of phase discrimination:

1. To the pilot signal with identical FH patterns in different symbols, the τ_{IONO}s and Θ_rs remain the same during the tracking and demodulating process.
2. To the long-term signal with varying FH patterns, the τ_{IONO}s and Θ_rs scintillate in consecutive hops and the BER deteriorates.

Secondly, we study the TOA estimation error's effect on the BER. In receiver-designing practices, a tracking loop with a narrow noise-band can guarantee that the TOA precision is within 1 ns and the TOA error's effect on Θ_r is ignored.

37.3 BPSK Demodulation Performance of Pilot Signal under Ionosphere Dispersion

37.3.1 BER Derivation

Since the mean of code and carrier discriminating outputs are constant, the BPSK demodulator of the pilot signal is modeled as a phase decider, which calculates the phase of the received vector $\mathbf{r}$ and choses the maximum-likely signal vector $\mathbf{s}_{\text{m}}$. The phase decider is described below. The sending vector of BPSK signal is

$$\mathbf{s_m} = \left[\sqrt{E_S}\cos\frac{2\pi}{M}(m-1), \sqrt{E_S}\sin\frac{2\pi}{M}(m-1)\right], \tag{37.8}$$

where m=1 or m=0. The receiving vector is **r**and the phase of the receiving vector is

$$\Theta_r = \tan^{-1}\frac{r_2}{r_1} = \tan^{-1}\frac{Q_P}{I_P}. \tag{37.9}$$

According to [8], the received phase can be expressed as

$$\Theta_{\mathrm{r}} = \Theta_{\mathrm{AWGN,IONO}} + \Theta_{\mathrm{Scint}}, \tag{37.10}$$

where $\Theta_{\mathrm{AWGN,IONO}}$ is the phase under white noise and Θ_{Scint} is the phase caused by time scintillation, these two phases are independently distributed variables. To pilot signal Θ_{Scint} is zero, so

$$\Theta_{\mathrm{r}} = \Theta_{\mathrm{AWGN,IONO}}. \tag{37.11}$$

Define that $V = \sqrt{r_1^2 + r_2^2}$, then the joint PDF of V and Θ_r is

$$p_{V,\Theta_r}(V,\Theta_r) = \frac{V}{2\pi\sigma_r^2}\exp\left(-\frac{V^2 + E_{s,\mathrm{IONO}} - 2V\sqrt{E_{s,\mathrm{IONO}}}\cos\Theta_r}{2\sigma_r^2}\right). \tag{37.12}$$

Calculating the integral of $p_{V,\Theta r}(\mathrm{V},\Theta_r)$ over $\mathrm{V}\in(0,\infty)$, we obtain $p_{\theta r}(\Theta_r)$ as below

$$p_{\theta_r}(\Theta_r) = \frac{1}{2\pi}e^{-2\gamma_s\sin^2\Theta_r}\int_0^\infty Ve^{-(V-\sqrt{2\gamma_s}\cos\Theta_r)^2/2}\mathrm{d}V. \tag{37.13}$$

Again, calculate the integral over the decision region and then obtain the BER as

$$P_{\mathrm{e,IONO}} = 1 - \int_{-\pi/2+\overline{\theta_{\mathrm{dem}}}}^{\pi/2+\overline{\theta_{\mathrm{dem}}}} p_{\theta_r}(\Theta_r)\mathrm{d}\Theta_r, \tag{37.14}$$

where

$$\overline{\theta_{dem}} = \arg\left[\sum_{i=0}^{N_{hop}-1} e^{2\pi f_i\left(\frac{\lambda_{TEC}\cdot TEC}{2\pi f_i^2} - \tau_{IONO}\right)}\right]. \tag{15}$$

Because the integral in (13) is complicated, knowing that BPSK is equal to PAM, we have

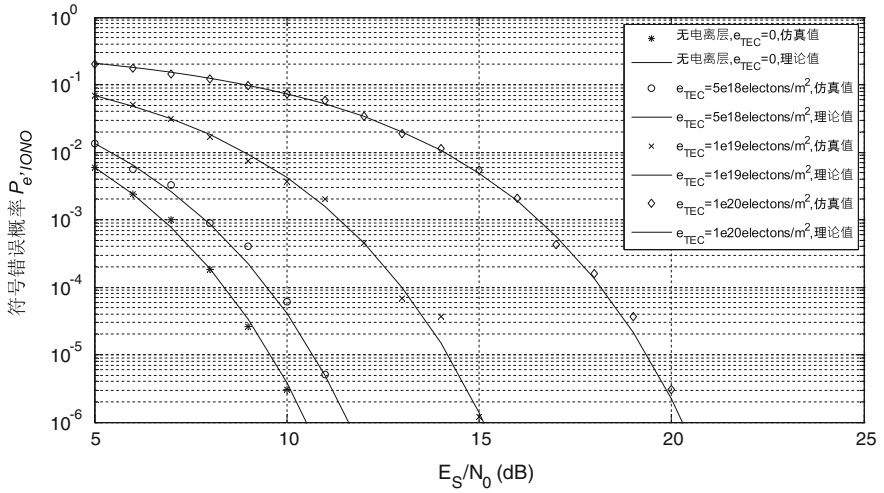


Fig. 37.2 BER of pilot signal $P_{\mathrm{e,IONO}}$ under different E_s/N_0 and e_{TEC}

$$P_{\mathrm{e,IONO}} = Q\left(\sqrt{\frac{2E_{s,\mathrm{IONO}}}{N_0}}\right), \tag{37.16}$$

where

$$\gamma_s = E_S/N_0, E_{s,\mathrm{IONO}} = L_{TEC}E_s. \tag{37.17}$$

L_{TEC} is the signal to noise ratio caused by ionosphere.

37.3.2 Simulation

By using Matlab the BER of BPSK demodulation is simulated. e_{TEC} is set to be 0.5×10^{18}, 10^{19} and 10^{20}, the signal-to-noise ratio (SNR) ranges from 5 to 25 dB, and the decision region is $(-\pi/2, \pi/2)$. The simulation result BER (denoted as P_{e}) is shown in Fig. 37.2 and the BER formula in Eq. (37.16) is verified.

37.4 BPSK Demodulation Performance of Long-Term Signal under Ionosphere Dispersion

37.4.1 PDF of Ionosphere Delay Scintillation

The FH pattern of long-term signal changes with the symbols, causes the $f_{\mathrm{RF,cent}}$ to change slightly in different symbols and the BPSK demodulation performance is

affected. In this section the ionosphere delay scintillation $\tau_{\mathrm{IONO,Scint}}$ is investigated and its impact on Θ_{Scint} is analyzed.

Assume the mean of FH pattern is $f_{\mathrm{RF,cent}}(p)$, p denotes the number of demodulation symbol and $p = 1,2,3, \ldots, \infty$. Based on the design principle of FH code balance [9], the $f_{\mathrm{RF,cent}}$ obeys uniform distribution over (f_a, f_b), that is, $f_{\mathrm{RF,cent}} \sim U(f_a,f_b)$, the FH bandwidth is $B_{\mathrm{Scint}} = (f_b - f_a)$, so the PDF of $f_{\mathrm{RF,cent}}$ is written as

$$p_f(f_{RF,cent}) = \frac{1}{f_b - f_a}, \quad f_{\mathrm{RF,cent}} \in (f_a,f_b). \tag{37.18}$$

In order to obtain the PDF of ionosphere delay scintillation $p_\tau(\tau_{\mathrm{IONO,Scint}})$ caused by ionosphere e_{TEC}, we use the following theorem.

Theorem 37.1 [10, 11] *To an N-dimensional random variable* $\boldsymbol{X} = (X_1, X_2, \ldots, X_N)$ *with the joint PDF* $p_{\boldsymbol{x}}(x_1, x_2, \ldots, x_N)$ *if the following THREE conditions concerning a N-variate function* $g_i\,(x_1, x_2, \ldots, x_N)$, $(i = 1, 2, \ldots, N)$ *are satisfied,*

(i) *There exists the only reverse function* $x_i = g_i^{-1}(y_1, y_2, \ldots, y_N)$ of $g_i\,(x_1, x_2, \ldots, x_N)$;

(ii) *Both* $g_i\,(x_1, x_2, \ldots, x_N)$ *and* $g_i^{-1}(y_1, y_2, \ldots, y_N)$ *are continuous functions*;

(iii) *There exist continuous partial derivatives* $\frac{\partial g_i^{-1}}{y_j}, (i,j = 1,2,\ldots,N)$, *together forming the Jacobi determinant, as* $\mathbf{J} = \frac{\partial(g_1^{-1},g_2^{-1},\ldots,g_N^{-1})}{\partial(y_1,y_2,\ldots,y_N)}$.

The PDF of $\boldsymbol{Y}=(Y_1, Y_2, \ldots, Y_N)$ *can be calculated as* $p_y\,(y_1, y_2, \ldots, y_N) = p_{\boldsymbol{x}}\,[g_1^{-1}(y_1, y_2, \ldots, y_N), g_2^{-1}\,(y_1, y_2, \ldots, y_N), \ldots, g_N^{-1}(y_1, y_2, \ldots, y_N)]\cdot|\mathbf{J}|$.

Since $\tau_{IONO} = g(f_{RF,cent})$ is a continuous function and its only continuous reverse function is

$$f_{RF,cent} = g^{-1}(\tau_{IONO,\mathrm{Scint}}) = \sqrt{\frac{\lambda_{TEC} e_{\mathrm{TEC}}}{2\pi\tau_{IONO,\mathrm{Scint}}}}, \tag{37.19}$$

and the derivative

$$\frac{df_{RF,cent}}{d\tau_{\mathrm{IONO,Scint}}} = \sqrt{\frac{\lambda_{TEC} e_{\mathrm{TEC}}}{2\pi}}(\tau_{IONO,\mathrm{Scint}})^{-\frac{3}{2}}, \tag{37.20}$$

is continuous, we apply Theorem 37.1 and obtain the PDF of $\tau_{\mathrm{IONO,Scint}}$ as

$$p_\tau(\tau_{IONO,\mathrm{Scint}}) = p_f(g^{-1}(\tau_{IONO,\mathrm{Scint}})) \cdot \frac{\mathrm{d}g^{-1}(\tau_{IONO,\mathrm{Scint}})}{\mathrm{d}\tau_{IONO,\mathrm{Scint}}}. \tag{37.21}$$

After some simple math we have

$$p_\tau(\tau_{IONO,\text{Scint}}) = \frac{1}{f_b - f_a}\sqrt{\frac{\lambda_{TEC} e_{\text{TEC}}}{8\pi\tau^3_{IONO,\text{Scint}}}}, \quad \tau_{IONO,\text{Scint}} \in \left(\frac{\lambda_{TEC} e_{\text{TEC}}}{2\pi f_b^2}, \frac{\lambda_{TEC} e_{\text{TEC}}}{2\pi f_a^2}\right), \tag{37.22}$$

and the ionosphere delay of the center frequency is

$$\tau_{IONO,cent} = 2\lambda_{TEC} e_{\text{TEC}} \Big/ \pi(f_a + f_b)^2. \tag{37.23}$$

According to Eq. (37.22), the $\tau_{\text{IONO,Scint}}$ under different e_{TEC}s is drawn below in Fig. 37.3 with f_a = 2.420 GHz, f_b = 2.425 GHz. Obviously when the mean of FH pattern conforms to uniform distribution, the corresponding $\tau_{\text{IONO,Scint}}$ does not obey uniform distribution.

37.4.2 BER of BPSK Under FH Pattern Scintillation

37.4.2.1 PDF of Θ_r Under FH Pattern Scintillation

Knowing the PDF of $\tau_{\text{IONO,Scint}}$, we analyze the PDF of Θ_r and obtain the BER of BPSK $P_{\text{e,IONO,Scint}}$ as below.

$$P_{\text{e,IONO,Scint}} = 1 - \int_{-\pi/2+\overline{\theta_{\text{dem}}}}^{\pi/2+\overline{\theta_{\text{dem}}}} p_{\theta_r}(\Theta_r)\mathrm{d}\Theta_r. \tag{37.24}$$

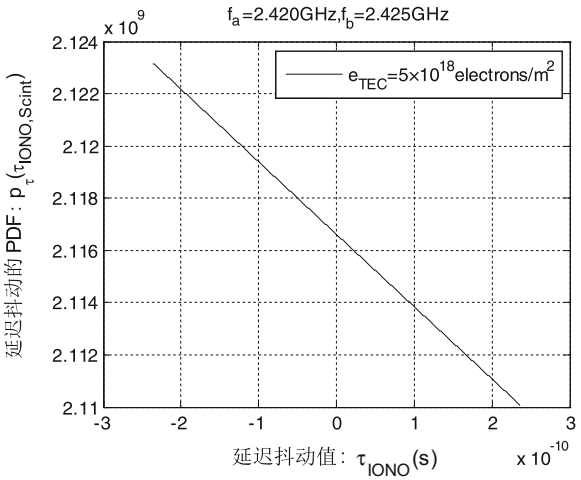
According to (37.10) we have $\Theta_{\text{r}} = \Theta_{\text{scint}} + \Theta_{\text{AWGN,IONO}}$, referring to Eq. (37.13) we already know the PDF of $\Theta_{\text{AWGN,IONO}}$, then the remaining unknown factor is the PDF of Θ_{Scint}.

According to Eq. (37.6), when $|\tau_{\text{IONO}}| < 1/(6B_{\text{hop}})$ is true, Θ_{Scint} can be approximated as the function of $\tau_{\text{IONO,Scint}}$

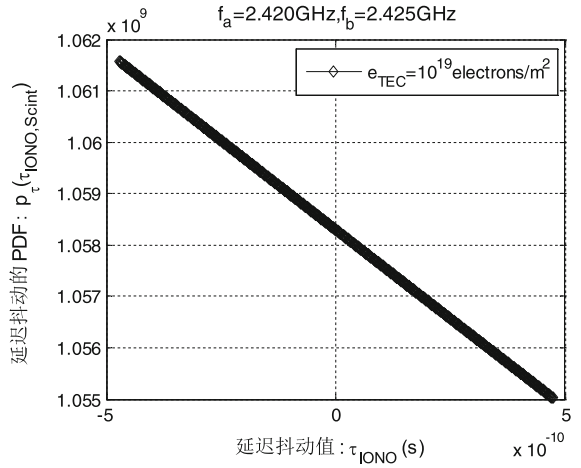
$$\Theta_{\text{Scint}}(\tau_{\text{IONO,Scint}}) = 2\pi\tau_{\text{IONO,Scint}} B_{hop}. \tag{37.25}$$

Since

$$\tau_{\text{IONO,Scint}} = h^{-1}(\Theta_{\text{scint}}) = \frac{\Theta_{\text{Scint}}}{2\pi B_{\text{hop}}}, \tag{37.26}$$



(a) $e_{\mathrm{TEC}}=5\times10^{18}\mathrm{electrons/m^2}$



(b) $e_{\mathrm{TEC}}=10^{19}\mathrm{electrons/m^2}$

Fig. 37.3 PDF of ionosphere delay scintillation $\tau_{\mathrm{IONO,Scint}}$ under different e_{TEC}

and

$$\frac{dh^{-1}(\Theta_{\mathrm{Scint}})}{d\Theta_{\mathrm{Scint}}}=\frac{1}{2\pi B_{\mathrm{hop}}}, \tag{37.27}$$

according to Theorem 37.1, we have

$$p_{\theta\,\mathrm{Scint}}(\Theta_{\mathrm{Scint}})=p_{\tau}(h^{-1}(\Theta_{\mathrm{Scint}}))\cdot\frac{dh^{-1}(\Theta_{\mathrm{Scint}})}{d\Theta_{\mathrm{Scint}}}. \tag{37.28}$$

Plug Eq. (37.21) into Eq. (37.28), the PDF of Θ_{Scint} can be derived as

$$p_{\theta\,\text{Scint}}(\Theta_{\text{Scint}}) = \frac{1}{f_b - f_a}\sqrt{\frac{\lambda_{TEC} e_{\text{TEC}}}{8\pi(\frac{\Theta_{\text{Scint}}}{2\pi B_{hop}})^3}} \cdot \frac{1}{2\pi B_{\text{hop}}} = \frac{1}{2(f_b - f_a)}\sqrt{\frac{B_{hop}\lambda_{TEC} e_{\text{TEC}}}{\Theta^3_{\text{Scint}}}} \tag{37.29}$$

where $\Theta_{\text{scint}} \in \left(\frac{B_{hop}\lambda_{TEC} e_{\text{TEC}}}{f_b^2}, \frac{B_{hop}\lambda_{TEC} e_{\text{TEC}}}{f_a^2}\right)$.

Now the PDF of Θ_{r}, the sum of two random variables, denoted as $p_{\theta r}(\Theta_{\text{r}})$, can be derived. Let $x = \Theta_{\text{Scint}}$, $y = \Theta_{\text{AWGN,IONO}}$ and $z = \Theta_{\text{r}}$. Since x and y are independent random variables, the joint PDF of x and y is

$$f(x, y) = f(x)f(y). \tag{37.30}$$

To derive f(z), firstly we derive $F(z)$. Since $z = x + y$, we have

$$\begin{aligned} F(z) &= \iint\limits_{x+y\le z} f(x, y)dxdy \\ &= \int_{-\infty}^{z} p_{\theta\,\text{AWGN,IONO}}(\Theta_{\text{AWGN,IONO}})d\Theta_{\text{AWGN,IONO}} \int_{\theta_b}^{\theta_a} p_{\theta\,\text{Scint}}(\Theta_{\text{Scint}})d\Theta_{\text{Scint}} \\ &= -2A\left(\sqrt{\theta_a} - \sqrt{\theta_b}\right)\int_{-\infty}^{z} p_{\theta\,\text{AWGN,IONO}}(\Theta_{\text{AWGN,IONO}})d\Theta_{\text{AWGN,IONO}}, \end{aligned}$$

and

$$f_z(z) = \frac{dF(z)}{dz} = -2A\left(\sqrt{\theta_a} - \sqrt{\theta_b}\right)p_z(z). \tag{37.31}$$

Plug $z = \Theta_{\text{r}}$ into Eq. (37.31) then we obtain

$$p_{\theta_r}(\Theta_r) = 2A\left(\sqrt{\theta_b} - \sqrt{\theta_a}\right)\frac{1}{2\pi}e^{-2\gamma_s \sin^2\Theta_r}\int_0^{\infty} Ve^{-(V-\sqrt{2\gamma_s}\cos\Theta_r)^2/2}\text{d}V, \tag{37.32}$$

where

$$A = \frac{1}{2(f_b - f_a)}\sqrt{B_{hop}\lambda_{TEC} e_{\text{TEC}}}, \quad \theta_b = \frac{B_{hop}\lambda_{TEC} e_{\text{TEC}}}{f_b^2}, \quad \theta_a = \frac{B_{hop}\lambda_{TEC} e_{\text{TEC}}}{f_a^2}.$$

Transform Eq. (37.16) and we get

$$\int_{-\pi/2+\overline{\theta_{\text{dem}}}}^{\pi/2+\overline{\theta_{\text{dem}}}} \frac{1}{2\pi}e^{-2\gamma_s \sin^2\Theta_r}\int_0^{\infty} Ve^{-(V-\sqrt{2\gamma_s}\cos\Theta_r)^2/2}\text{d}V = 1 - Q\left(\sqrt{\frac{2E_{\text{S,IONO}}}{N_0}}\right). \tag{37.33}$$

Substitute Eq. (37.33) into Eq. (37.32), the BER under FH pattern scintillation is derived as

$$\begin{aligned} P_{\mathrm{e,IONO,Scint}} &= 1 - \int_{-\pi/2+\overline{\theta_{\mathrm{dem}}}}^{\pi/2+\overline{\theta_{\mathrm{dem}}}} p_{\theta_{\mathrm{r}}}(\Theta_r)\mathrm{d}\Theta_r \\ &= 1 - 2A\left(\sqrt{\theta_b} - \sqrt{\theta_a}\right)\left(1 - Q\left(\sqrt{\frac{2E_{\mathrm{S,IONO}}}{N_0}}\right)\right). \end{aligned} \tag{37.34}$$

37.4.2.2 BER Simulations

In this section, the BPSK demodulation performance under different FH pattern scintillation ranges B_{Scint}s is simulated. The simulation settings are listed in Table 37.1.

Firstly, according to Eq. (37.32), under the condition of different ionosphere dispersion levels, different FH pattern scintillation bandwidths and E_s/N_0 = 15 dB, the $p_{\theta r}(\Theta_r - \overline{\theta_{dem}})$ under different B_{Scint}s is shown in Fig. 37.4. The origin of horizontal coordinate in Fig. 37.4 assumes that the center of phase scintillation is zero, that means $\Theta_r = \overline{\theta_{dem}}$=0. Comparing the three curves in Fig. 37.4 with B_{Scint} = 1 MHz, B_{Scint} = 10 MHz and B_{Scint} = 100 MHz, we find that the variance of Θ_r increases when B_{Scint} increases. Comparing (a) and (b), we find that the variance of Θ_r increases when e_{TEC} increases.

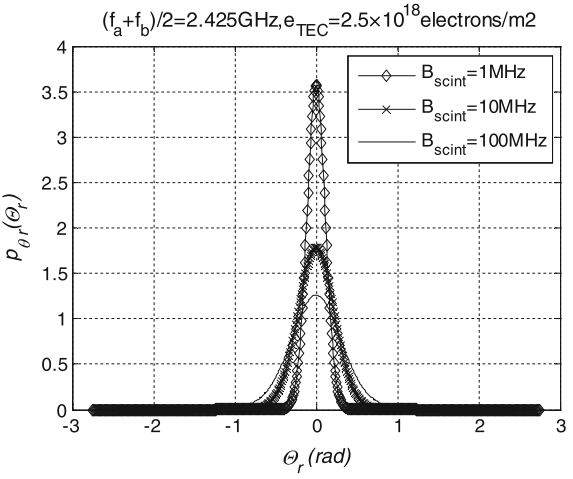
Then the demodulation performance of BPSK of long-term signal is simulated. Assume the transmitted BPSK phase to be θ_s = 0, the decision criterion is as follows

If $\Theta_r \in (-\pi/2 + \overline{\theta_{dem}}, \pi/2 + \overline{\theta_{dem}})$, then the decision result is θ_s = 0, and vice versa.

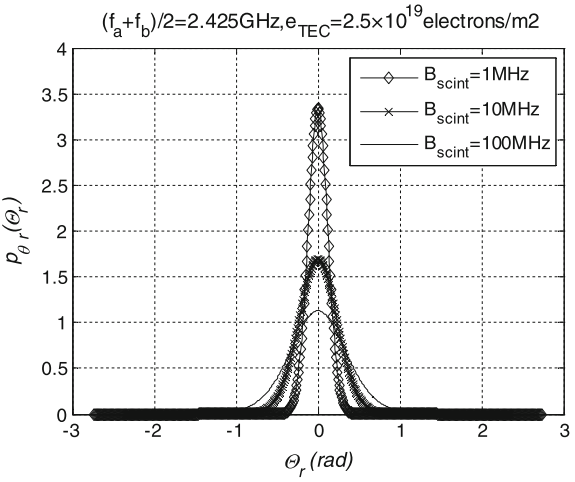
Through simulation the error probability $P_{\mathrm{e,IONO}}$ changes while e_{TEC} and B_{Scint} change. The demodulation process repeats for 10^7 times during the BER Monte Carlo simulation and the results are shown in Fig. 37.5. The solid line represents the BPSK demodulation BER when e_{TEC} equals to 0 and the FH pattern scintillates,

Table 37.1 Parameter settings of long-term signal for BER simulation

Parameter	Settings
Type of DS code	Chaotic
Type of FH code	Chaotic
DS code rate	1 Mcps
FH rate	50,000 hop/s
Radio frequency	S band
FH bandwidth	150 MHz
Modulation keying	BPSK



(a) $e_{\mathrm{TEC}}=2.5\times10^{18}\mathrm{electrons/m^2}$



(b) $e_{\mathrm{TEC}}=2.5\times10^{19}\mathrm{electrons/m^2}$

Fig. 37.4 Θ_r under different e_{TEC} and B_{Scint}

which shows that when there is no ionosphere the BER is not affected by the FH pattern scintillation. The solid line with "○" represents the BPSK demodulation BER when e_{TEC} equals to 2.5×10^{18} electrons/m^2 and the FH pattern scintillation range B_{Scint} equals to 1MHz, and so on, for the rest of the curves. These curves show that when e_{TEC} and B_{Scint} increase, the BER raises, and the BER results closely fits the result in Eq. (37.32).

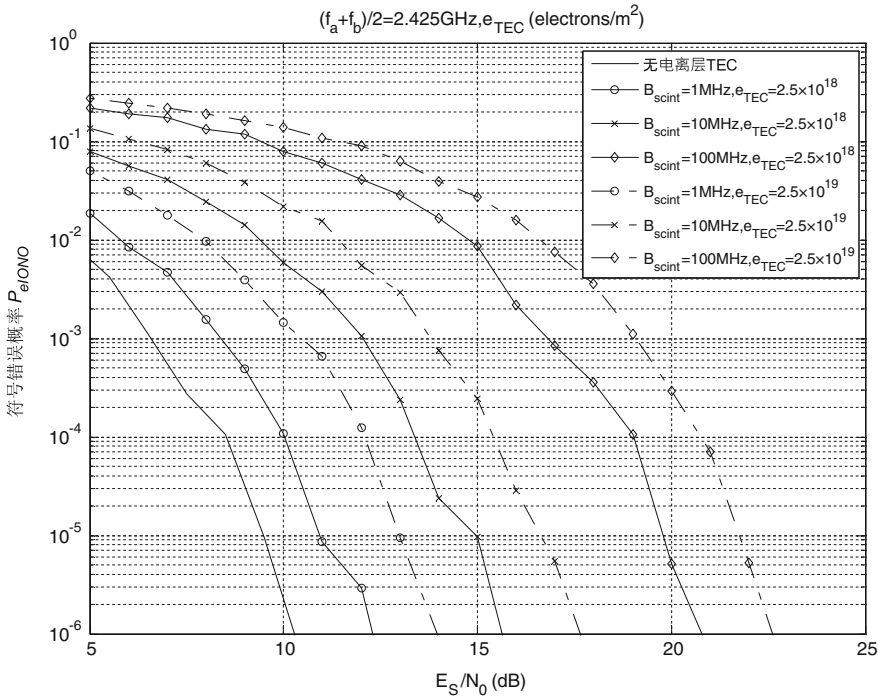


Fig. 37.5 BER of long-term signal ($P_{e,IONO}$) under different E_s/N_0 and e_{TEC}

37.5 Conclusion

The BER performance of DS/FH-BPSK signal deteriorates when ionosphere dispersion and FH pattern scintillation exists. Through investigating the mathematical laws within the signal phase in ionosphere transmission condition, this paper gives a closed-form BER solution for BPSK modulation. Simulations under different conditions are run to verify the analytical expression.

Acknowledgements This work is supported by PLA Science Research Foundation (No. EP2015032200B21087).

References

1. Some YK, Kam PY (1995) Bit-error probability of QPSK with noisy phase reference. IEEE Proc Commun 142(5):292–296
2. Corvaja R, Pupolin S (2002) Effects of phase noise spectral shape on the performance of DPSK systems for wireless applications. Eur Trans Telecommun 13(3):203–210

3. Taggart D, Kumar R (2011) Impact of phase noise on the performance of the QPSK modulated signal. In: Aerospace conference, 2011. IEEE, pp 1–10
4. Nie W, Shuai T, Liu Y, Li G (2012) Analysis of influence of phase noise on demodulation BER of satellite signals. In: Electronic Proceedings of third annual china satellite navigation conference—SO_4 atomic clock technology and time/frequency system
5. Zhu L (2015) Influence of phase noise on QPSK demodulation performance. Radio Eng 45 (10):38–40
6. Le Nir V, Scheers B (2011) Performance of delay and add direct sequence spread spectrum modulation scheme with fast frequency hopping in frequency selective Rayleigh channels. In: Military communications conference, 2011-MILCOM, pp 622–627
7. Yeh KC, Liu C (1982) Radio wave scintillations in the ionosphere. Proc IEEE 70(4):324–360
8. Hamaguchi K, Shoji Y, Ogawa H (2001) BER performance of coherent-QPSK transmissions affected by phase noise from frequency converters. In: Vehicular technology conference, 1988, IEEE 38th, pp 2281–2284
9. Cai S, Liu G, Tang H (2011) Statistical detection of sequential performance of long period frequency-hopping code. Telecommun Eng 51(5):28–32
10. Zhaolong Su (2004) Teaching materials of stochastic process. Press of PLA University of Science and Technology, Nanjing
11. Wang Fanbin (2013) Promotion of variable transformation method of multidimensional random variablefunction. J Jinggangshan Univ: Nat Sci 3:10–12

Chapter 38
De-noising Method Research on RF Signal by Combining Wavelet Transform and SVD

Junyao Li, Yongbin Li, Xiaoqiang Wang and Peijie Zhang

38.1 Introduction

As the space distance transmission characteristics, RF signal are used widely in the fields of aerospace measurement and control, radio communication, radio and television and so on. RF signal would be influenced by noise and interference inevitably from atmosphere, channel noise and equipment devices when it spreads in space. Furthermore affecting signal communications and leading to received signal distortion and transmission errors. Therefore, the noise signal preprocessing should be done before signal analysis, in order to restore the useful signal.

There are many current researches on signal noise elimination and inhibition, such as traditional filter de-noising method, wavelet threshold method, adaptive filtering method and neural network method. These methods all have their own advantages, disadvantages and application conditions. But there are few research about RF signal systemic noise reduction. Traditional low-pass filter and band-pass filter are usually used in projects to filter the no-aliasing noise, but it can only remove the out-band noise, the belt noise are not included. Few researches used wavelet threshold method to remove the noise in RF signal [1–3], but there is a premise that the useful signal must be of low frequency and noise must be of high frequency, which has some limitations.

Traditional filter de-noising method, wavelet threshold method and SVD (Singular value decomposition) method were studied comparatively in RF signal. Then Method of combining wavelet threshold and SVD was put forward to reduce the noise. Simulation results and extraction effects of useful RF signal were analyzed by different methods. Finally, effectiveness and superiority of noise suppression were verified by de-noising performance metrics.

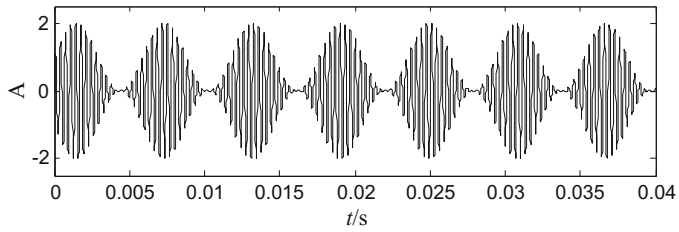
J. Li (✉) · Y. Li · X. Wang · P. Zhang
Xi'an Satellite Control Center, Xi'an 710043, China
e-mail: rachelyao_li@163.com

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_38

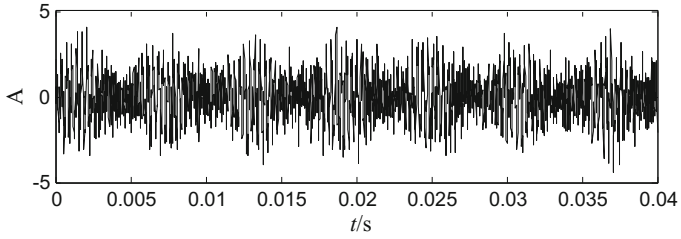
38.2 Analysis of RF Signal with Noise

According to the nature of noise, it can be divided into single frequency noise, impulse noise, and fluctuation noise. The single frequency noise mainly exist in the communication power supply and self-excited oscillation feedback system. And the spectrum is concentrated, which can prevent by taking some appropriate measures. The characteristics of impulse noise is in the form of pulse, and have short duration and strong intensity, which usually happened in industrial electric spark, thunder and lightning. Fluctuation noise are introduced by device noise inside the channel, thermal noise and spatial noise, it's a kind of irregular random noise. It can be seen that fluctuation noise is the main factor affecting the signal transmission performance. When the system is in the low signal noise ratio (SNR), the accuracy and certainty of frequency domain characteristics analysis will be affected. Especially when the frequency resolution is low, more influence will be brought in for fence effect, spectrum leakage and so on.

The simulation results of pure signal and noised-signal are showed respectively in Fig. 38.1a, b. The simulation parameters are as follows: carrier frequency f_c = 10.7 MHz, sample frequency f_s = 362 MHz, sampling points N = 4001, signal noise ratio ξ_{SNRin} = 0 dB.



(a) Pure RF signal



(b) Noised RF signal

Fig. 38.1 Time domain waveform of RF signal

38.3 RF Signal de-Noising and Simulation Analysis

38.3.1 Traditional Filter De-noising

Generally speaking, the RF signal has a band pass frequency domain feature. The early signal processing is usually happened in the device receiving end by analog low-pass filter to reduce white noise and color noise in higher frequency band. Then outside the scope of the required band signal will be further removed before decoding and processing by digital band pass filter. But filters can't reduce all noise around the expected frequency band, and noise in the frequency band are still retained. It can be seen that filter de-noising method can't meet requirements to systems of low signal noise ratio and high signal precision.

38.3.2 Wavelet Threshold De-noising

The principle of wavelet threshold de-noising [4–8] is that resolving noised signal to multi-scale through wavelet transformation. In order to separate the useful signal apart useless signal. Then eliminate noise element in every scales and retain the signal component.

The method steps is shown in Fig. 38.2. Firstly, an appropriate wavelet basis function should be chosen to get every wavelet coefficients after signal wavelet decomposing. Then, choosing appropriate threshold value λ and threshold function to deal with those decomposition coefficients. Discarding coefficients below λ and remaining coefficients greater than λ. Finally, realizing discrete wavelet inverse transformation of every wavelet coefficients, and getting the reconstruction RF de-noising signal.

38.3.3 SVD De-noising

Singular value decomposition (SVD) [9–11] is an important decomposition algorithm of linear algebra, and has an important application status in signal processing and numerical analysis. SVD method is similar with wavelet transformation, they all make use of the energy divisibility of signal and noise. Through the matrix transformation realizing the noised signal transformed into a new matrix space.

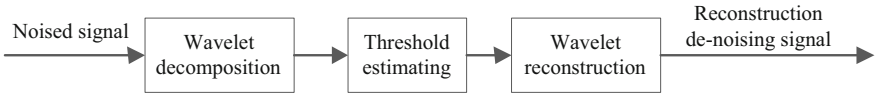


Fig. 38.2 Steps of wavelet threshold de-noising

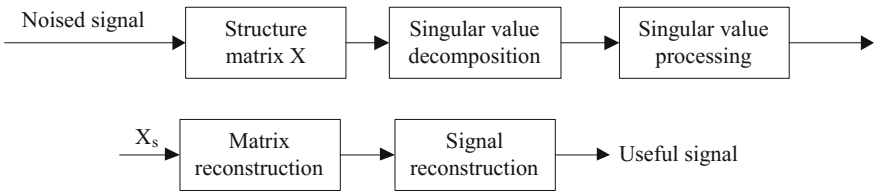


Fig. 38.3 Principle of SVD de-noising

Then turning the singular value which reflecting the noise component into zero. Ultimately achieving the goal of noise suppression. The principle of SVD method is shown in Fig. 38.3.

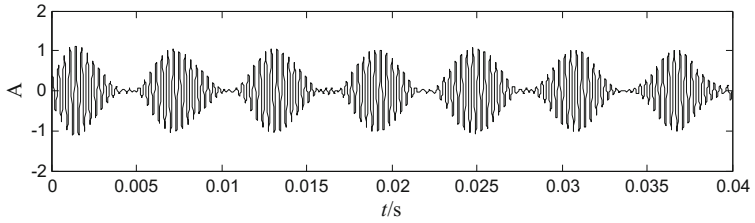
Assuming that noised RF signal $x(t) = s(n) + v(n)$, among them $s(n)$ is useful signal, $v(n)$ is noise, $n = 1, 2\ldots N$ (N is sampling points). Using Hankel matrix to structure a $m \times n$ matrix X. The expression is shown in formula (38.1), among it $N = m + n-1$.

$$X = \begin{bmatrix} x(1) & x(2) & \cdots & x(n) \\ x(2) & x(3) & \cdots & x(n+1) \\ \vdots & \vdots & \vdots & \vdots \\ x(m) & x(m+1) & \cdots & x(N) \end{bmatrix} \tag{38.1}$$

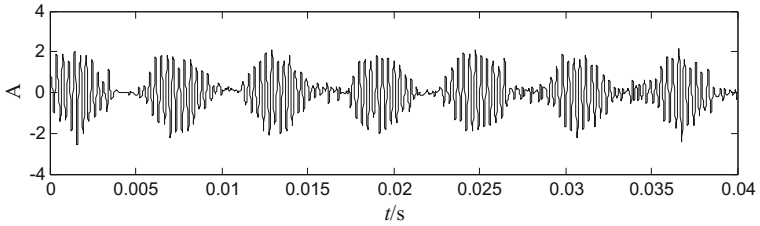
We can get $X = USV^{\mathrm{T}}$ by singular value decomposing to matrix X, among them: U is $m \times m$ unitary matrix, V is $n \times n$ unitary matrix. $S = \begin{bmatrix} \sum & 0 \\ 0 & 0 \end{bmatrix}$, $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2 \ldots \sigma_k \ \ldots \ \sigma_r)$, $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_k \geq \cdots \geq \sigma_r > 0$ are singular value of matrix X, they reflected energy concentration of useful signal and noise. The reconstructed signal after de-noising got received by remaining the singular value before k with greater energy and turning the rest (r-k) smaller singular value which did not meet the requirements of the threshold value into zero.

38.4 Analysis of Simulation Result

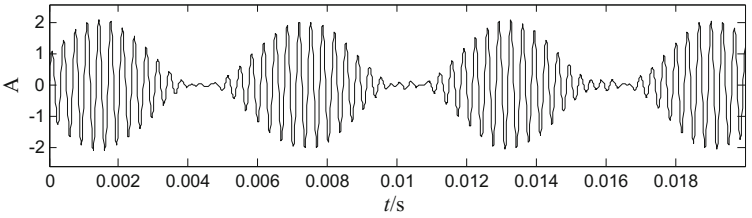
Simulation results were gained and shown in Fig. 38.4 by using those methods above to dispose the noised signal in Fig. 38.1b. Among them, the noise suppression results showed in Fig. 38.4a used analog band-pass filter by Fourier transform filtering method. Figure 38.4b showed the result by using wavelet threshold de-noising, choosing sym8 wavelet function and taking 7 layers decomposition, heuristic soft heursure was chosen at the same time. Figure 38.4c showed the result by using SVD method, the matrix dimension is 1000×1001, and the rank $k = 6$.



(a) Fourier transform de-noising



(b) Wavelet threshold de-noising



(c) SVD de-noising

Fig. 38.4 Time domain waveform of de-noising RF signal

Table 38.1 Noise suppression performance metrics by different algorithm (ξ_{SNRin} = 0 dB)

performance index	Fourier transform method	Wavelet threshold method	SVD method
ξ_{SNRo} (dB)	6.1495	10.6322	22.3088
ξ_{MSE}	0.1860	0.0663	0.0049

Signal de-noising performance metrics were shown in Table 38.1, among them, ξ_{SNRo} is signal noise ratio of de-noising RF signal. The greater value it is, the better noise suppression effect will have. ξ_{MSE} is mean square error of de-nosing signal, while the smaller value means the better effect. We got some conclusions as follows according to the above simulation results:

1. Fourier transform de-noising method using FFT to realize signal band-pass filtering, so a better waveform characteristic of original signal was retained. But when the noise was reduced, the useful signal energy was attenuated at the same time.

Table 38.2 De-noising signal ξ_{SNRo} (dB) by SVD

Hankel matrix dimension	$\xi_{SNRin} = -10$ dB	$\xi_{SNRin} = -5$ dB	$\xi_{SNRin} = 0$ dB	$\xi_{SNRin} = 5$ dB
1000×1001	12.6565	17.8026	21.8849	25.9105
500×501	9.8582	13.4066	19.1512	23.6707
100×101	9.7898	13.5978	17.1229	19.2792
100×1001	6.8660	8.2483	10.7223	10.5023
100×501	7.1940	10.0337	11.6473	11.7797

2. The wavelet threshold de-noising was effected by wavelet decomposition characteristics and selection of wavelet basis function, which would cause the reconstructed signal distortion.
3. SVD de-nosing retained better original signal time domain features, and performance of the reconstructed signal is better than Fourier transform method and wavelet threshold method. So it can be used in noise RF signal reduction.

SVD de-noising method needs to consider the decomposition algorithm effect of matrix dimension. Hankel matrix of different dimension were structured by 1000×1001, 500×501, 100×101, 100×1001, 100×501. Taking 100 times simulation calculation for RF signal with different input SNR ($\xi_{SNRin} = -10, -5, 0, 5$ dB), then calculated the mean value. Table 38.2 showed the result of de-nosing signal SNR. It can be seen from the statistical result that the larger ξ_{SNRin} was, the better noise reduction effect had. And the effectiveness of de-noising signal was increasing when structure matrix closed to square matrix.

38.5 Combining Wavelet Threshold Method and SVD

The simulation analysis found that although SVD algorithm had these advantages of simple principle, simple algorithm and better effect of de-noising, but it needed longer operation time when choosing a larger matrix dimension. Combing wavelet threshold method and SVD method can solve this problem effectively. Firstly, choosing sym8 wavelet function for 2 layer decomposition to noised signal, removing the higher frequency noise component in the signal. Then using Hankel matrix to construct a new matrix of 1000×1001 dimension. Finally, choosing the input RF signal $\xi_{SNRin} = 0$ dB to emulate 100 times and get the average value. Table 38.3 showed the noise suppression performance metrics of these two methods. It can be seen that the de-noising efficiency was improving availably by combining these two methods.

Table 38.3 Noise suppression performance metrics

De-noising method	Operation time(s)	ξ_{SNRo} (dB)	ξ_{MSE}
SVD	59.2	21.8849	0.0054
Wavelet transform combined with SVD	24.3	22.5974	0.0046

38.6 Conclusions

1. Traditional band-pass filter can only filter out noise outside the band, which can't remove noise in mixed frequency band. And Fourier transform affect de-noising effect for limiting by frequency resolution.
2. Wavelet threshold method assumes that the useful signal had lower frequency component and the noise had higher frequency component. And it needs to choose a appropriate wavelet function to reduce noise in different scales. But the energy weight of useful signal component and noise component in different scale space is different. The useful signal would be filter when each dimension used the same threshold value law, and leading to the distortion and non-ideal de-noising effect of reconstruction signal.
3. SVD method separated noise from the useful signal in different vector space by using matrix decomposition. Correlation between the matrix vector strengthened the extraction of useful signal, which made the singular value reflected useful signal energy is much larger than that reflected noise energy. The method avoided effectively about the filter mistake of useful signal causing by threshold choice, and the RF signal noise suppression performance was improving. But this method cut off the length of original signal affected by matrix dimension, and the problem will be solve subsequently.
4. Combining the wavelet threshold method and SVD method took advantage of those two methods effectively, and noised RF signal inhibition was realized much more efficiently.

References

1. Sun F, He YG, Xiao YQ (2010) The RFID radio frequency signal wavelet de-noising method. J Comput Simul 27(11):332–335
2. Sun XY (2011) Research on RFID radio frequency signal of wavelet de-noising method. Taiyuan University of Technology, Taiyuan
3. Tian Y (2015) Radio signal de-noising study based on wavelet packet threshold method. Monit Test, pp 57–59
4. Li HD, Li Q (2009) Threshold method based on wavelet de-noising algorithm research. Comput Technol Dev 19(7):56–585
5. Jin BL, Li H, Zhao NJ (2011) A new wavelet threshold de-noising algorithm. J Proj Rockets, Missiles Guid 31(1):167–169 (176)
6. Wu CJ, Gu YH, Xu XL (2015) New wavelet threshold method in the application of vibration signal de-noising. Mech Eng Autom 4:1–2

7. Liu XF, Xu HZ, Gong L (2009) Improvement on the wavelet packet transformation in telemetry rapid variation data noise reduction. J Spacecr TT&C Technol 28(3):57–60
8. Guo LH, Zhu YC, Yin WL (2011) Improvement on the wavelet packet transformation in telemetry rapid variation data noise reduction. J Spacecr TT&C Technol 28(3):57–60
9. Qian ZW, Cheng L, Li YH (2011) Noise reduction method based on singular value decomposition. J Vib Meas Diagn 31(4):459–463
10. Hu WH, Shu H, Luan YG (2010) Power quality signals' de-noising method based on singular value decomposition (SVD). Power Syst Prot Control 38(2):30–33
11. Gao CW, Sun YP (2014) Simulation and implementation of de-noising method for fast changing parameter data based on Hankel matrix and SVD. J Telem Track Command 35 (1):60–64

Chapter 39
The Multi-Objective Routing Optimization Algorithm for Hybrid SDN

Suolin Gu, Lijuan Luo, Zhekun Zhao and Xiaofang Li

39.1 Introduction

Hybrid SDN is a network architecture of the coexistence of traditional IP network nodes and SDN nodes. The core idea of SDN is the separation of control plane and forwarding plane, which can randomly distribute traffic for outbound port at the forwarding nodes to optimize network performance. However, due to various restrictions, SDN cannot be fully deployed to replace the traditional network in the short run. As a result, hybrid SDN will persist as a transition state over a long period of time [1]. In order to make full use of advantages of SDN nodes to optimize network performance, traffic engineering of hybrid SDN becomes a current research focus.

This paper explores the traffic engineering [2] of SDN/OSPF hybrid network (SONet). In SONet, a routing node is composed of a SDN-Forwarding Element (SDN-FE) controlled by the SDN controller (SDN-C) and a traditional router (OSPF-Forwarding Element, OSPF-FE) running OSPF [4], as shown in Fig. 39.1.

Interconnection between SDN-FE and OSPF-FE can be done through Route Flow or other mechanisms. The objective of traffic engineering is to optimize the utilization of network resources and improve network performance. Routing optimization is one of the key strategies to realize the goal of traffic engineering.

In a traditional OSPF network, the routing path of network traffic is fixed once the connection relations and link weight of the network topology are fixed. Load balancing of the OSPF network is achieved through ECMP mechanism. The routing optimization capacity of OSPF is restricted since ECMP can only equally distribute traffic on equivalent paths. In SONet, SDN-FE can split traffic converged to the

S. Gu · L. Luo · Z. Zhao · X. Li (✉)
Jiuquan Satellite Launch Center, Jiuquan 732750, China
e-mail: lxf000@mail.ustc.eud.cn

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_39

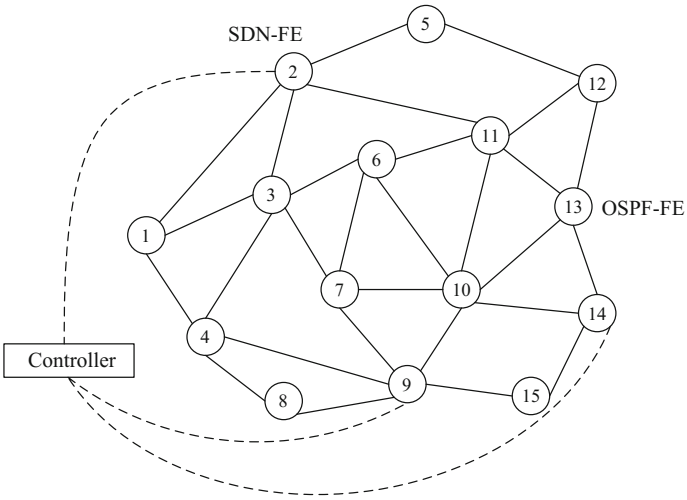


Fig. 39.1 SONet topology

node at the outbound port in any proportion, which gives SONet stronger routing optimization capacity than a traditional OSPF network.

For SONet, routing optimization can be achieved in two stages: (1) Balance outbound traffic of OSPF-FE by adjusting weight setting of the whole network; (2) Further balance the load of the whole network by splitting traffic converged at SDN-FE under fixed network weight setting.

Much work has been done in optimization of OSPF-FE link weight. Since it is a NP-complete problem, many inspiring algorithms were proposed. Paper [3] studied methods to reduce network load through random splits of outbound traffic by SDN-FE when link weight is 1. Paper [4] first comprehensively considered the two stages, and put forward the routing optimization algorithm in simultaneous changes of network weight setting and traffic splitting ratio.

Papers [3, 4] made pioneering work for routing optimization of SONet with respective algorithms of FLEX and SOTE. However, they share a drawback: the objective of routing optimization is to minimize maximum link utilization of a network, so optimization is available only for links with maximum utilization and other links will not be considered when such optimization cannot be proceeded. Meanwhile, failing to take restrictions in application of SDN-FE into consideration when seeking network load balancing increases burden of SDN-FE and even exceeds its service limit, bringing too much delay into the network.

In order to overcome the above disadvantage,this paper proposes a multi-objective routing optimization algorithm to optimize link utilization and transmission delay of SONet under the restriction of SDN-FE.

39.2 Mathematical Model of Multi-Objective Routing Optimization

We use $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$ to describe the network topology of SONet. G represents a weighted directed graph. $\boldsymbol{V}$ stands for the set of all the routing nodes in SONet, and $\boldsymbol{CV} \in \boldsymbol{V}, \boldsymbol{DV} = \boldsymbol{V} \backslash \boldsymbol{CV}$, of which $\boldsymbol{CV}$is the set of SDN-FEs and $\boldsymbol{DV}$ is the set of OSPF-FEs.$\boldsymbol{Out}(v), v \in \boldsymbol{V}$ means the set of all the links passed by traffic flowing out of node v. E is the set of links. For $c(e), e \in \boldsymbol{E}$ stands for capacity of link e, $l(e)$ means load of link e, $f_e^{(s,t)}$ means amount of the traffic between node s and node t borne by link e. In particular, when $x_e^{(s,t)} = f_e^{(s,t)}, e \in \boldsymbol{Out}(v), v \in \boldsymbol{CV}$, meaning traffic split by SDN-FE node v.$N(x_e^{(s,t)})$ is used to count the number of $x_e^{(s,t)}$.$\boldsymbol{S}(v), v \in \boldsymbol{CV}$ means the maximum capacity of SDN-FE v to process traffic splitting. *TM* stands for traffic matrix, and $\boldsymbol{TM}_{(i,j)}$ means the traffic from routing node i to routing node j.

39.2.1 Optimization Objective

The optimization objective of this paper is to achieve minimum network congestion of SONet while bringing in least network delay.

Minimum network congestion means optimum load balancing. The objective of current load balancing algorithms of SONet is to minimize link utilization, the advantage of which is simple and easy to get the optimum solution. However, it does not take network traffic balancing into consideration from a comprehensive perspective. Paper [5] put forward a new optimization model called the minimum link cost sum model, which overcame the shortcoming of the minimum link utilization model. The model introduces the cost function $\boldsymbol{\Phi_e}(\boldsymbol{l}(\boldsymbol{e})/\boldsymbol{c}(\boldsymbol{e}))$ demonstrated in formula (39.1) as the function for link utilization [6], to punish heavy-load links emerging in the optimization process.

$$\boldsymbol{\Phi_e} = \begin{cases} \boldsymbol{l}(\boldsymbol{e}) & \boldsymbol{0} \le \boldsymbol{l}(\boldsymbol{e})/\boldsymbol{c}(\boldsymbol{e}) \le \boldsymbol{1/3} \\ \boldsymbol{3l}(\boldsymbol{e}) - \frac{\boldsymbol{2}}{\boldsymbol{3}}\boldsymbol{c}(\boldsymbol{e}) & \boldsymbol{1/3} \le \boldsymbol{l}(\boldsymbol{e})/\boldsymbol{c}(\boldsymbol{e}) \le \boldsymbol{2/3} \\ \boldsymbol{10l}(\boldsymbol{e}) - \frac{\boldsymbol{16}}{\boldsymbol{3}}\boldsymbol{c}(\boldsymbol{e}) & \boldsymbol{2/3} \le \boldsymbol{l}(\boldsymbol{e})/\boldsymbol{c}(\boldsymbol{e}) \le \boldsymbol{9/10} \\ \boldsymbol{70l}(\boldsymbol{e}) - \frac{\boldsymbol{178}}{\boldsymbol{3}}\boldsymbol{c}(\boldsymbol{e}) & \boldsymbol{9/10} \le \boldsymbol{l}(\boldsymbol{e})/\boldsymbol{c}(\boldsymbol{e}) \le \boldsymbol{1} \end{cases} \tag{39.1}$$

$\boldsymbol{\Phi_e}(\boldsymbol{l}(\boldsymbol{e})/\boldsymbol{c}(\boldsymbol{e}))$ is defined as piecewise linear increasing convex function, indicating traffic pays less cost when passing through links with low utilization, and higher cost on the contrary. The optimization objective of the model is $\boldsymbol{\min\Phi} = \sum_{\boldsymbol{e}\in\boldsymbol{E}} \boldsymbol{\Phi_e}$, i.e. comprehensive consideration of utilization of all links, to prevent optimization suspension caused by bottleneck links. Given the advantage of the minimum link cost sum model, this paper will take it as the objective of network load balancing in SONet routing optimization.

Another optimization objective of this paper is minimum network delay. SDN-FE can split traffic in outbound links in any proportion, which is conducive to network load balancing. However, such function of SDN-FE depends on SDN-C. SDN-FE must communicate with SDN-C for each split traffic on SDN-FE, and SDN-C will generate a corresponding configuration directive and send it to SDN-FE. Then, SDN-FE must install the directive before splitting traffic. We can see that traffic splitting will bring in some delay into the network. For large amount of split traffic, such delay cost can be neglected compared with benefits in load balancing. However, it will be another matter for small split traffic. When 10 KB traffic takes 1 μs to pass through a 10 Gpbs link, the transmission delay will be at least doubled even if traffic splitting only needs 1 μs. Therefore, the amount of split traffic must be controlled. Since introduced network delay is related to the amount of split traffic, the delay cost function $\boldsymbol{\Gamma}_{x_e^{(v,t)}} = \boldsymbol{\Gamma}(x_e^{(v,t)}/c(e))$ in formula (39.2) will punish the split traffic that is too little.

$$\boldsymbol{\Gamma}_{x_e^{(s,t)}} = \begin{cases} 100 & 0 \le x_e^{(s,t)}/c(e) \le 1/10^3 \\ -\frac{2\times 10^3}{c(e)} x_e^{(s,t)} + 102 & 1/10^3 \le x_e^{(s,t)}/c(e) \le 1/10^2 \\ -\frac{8\times 10^2}{c(e)} x_e^{(s,t)} + 90 & 1/10^2 \le x_e^{(s,t)}/c(e) \le 1/10 \\ -\frac{1\times 10^2}{9c(e)} x_e^{(s,t)} + \frac{100}{9} & 1/10 \le x_e^{(s,t)}/c(e) \le 1 \end{cases} \tag{39.2}$$

The total cost of introduced network delay is $\Gamma = \sum_{v\in CN} \sum_{e\in out(v)} \Gamma_{x_e^{(v,t)}}$, and the optimization objective is $\min \Gamma$.

The overall optimization objective is defined as $\min \Theta = \{\Phi, \Gamma\}$. Since it is a multi-objective optimization problem, the weight method can be used to turn it into single-objective optimization. $\delta \in [0, 1]$ is set as the weight factor to evaluate relative priorities of load balancing and network delay in optimization, then $\min \Theta = \alpha \Phi + (1 - \alpha)\Gamma$. The following is constraints to the optimization objective, and analysis of solution complexity.

39.2.2 Mathematical Model

The following mathematical model of routing optimization can be constructed from analysis in Sect. 39.2.1.

$$\min \Theta = \alpha \Phi + (1 - \alpha)\Gamma$$

subject to

$$\sum_{i:(i,j)\in E} f_{(i,j)}^{(s,t)} - \sum_{k:(j,k)\in E} f_{(j,k)}^{(s,t)} = \begin{cases} -TM_{(s,t)} & if\ i = s \\ TM_{(s,t)} & if\ j = tj, s, t \in V \\ 0 & otherwise \end{cases} \tag{39.3}$$

$$l(e) = \sum_{(s,t)\in V\times V} f_e^{(s,t)} \le c(e) \tag{39.4}$$

$$\mathbf{N}\left(x_e^{(s,t)}\right) = \begin{cases} 1 & if \quad x_e^{(s,t)} > 0 \\ 0 & if \quad x_e^{(s,t)} = 0 \end{cases} e \in Out(v); v \in CN; s,t \in V \tag{39.5}$$

$$\sum_{e\in Out(v)} \sum_{(s,t)\in V\times V} N\left(x_e^{(s,t)}\right) \le S(v) v \in CN \tag{39.6}$$

$$\Phi_e = \begin{cases} l(e) & 0 \le l(e)/c(e) \le 1/3 \\ 3l(e) - \frac{2}{3}c(e) & 1/3 \le l(e)/c(e) \le 2/3 \\ 10l(e) - \frac{16}{3}c(e) & 2/3 \le l(e)/c(e) \le 9/10 \\ 70l(e) - \frac{178}{3}c(e) & 9/10 \le l(e)/c(e) \le 1 \end{cases} \tag{39.7}$$

$$\Gamma_{x_e^{(s,t)}} = \begin{cases} 100 & 0 \le x_e^{(s,t)}/c(e) \le 1/10^3 \\ -\frac{2\times 10^3}{c(e)} x_e^{(s,t)} + 102 & 1/10^3 \le x_e^{(s,t)}/c(e) \le 1/10^2 \\ -\frac{8\times 10^2}{c(e)} x_e^{(s,t)} + 90 & 1/10^2 \le x_e^{(s,t)}/c(e) \le 1/10 \\ -\frac{1\times 10^2}{9c(e)} x_e^{(s,t)} + \frac{100}{9} & 1/10 \le x_e^{(s,t)}/c(e) \le 1 \end{cases} \tag{39.8}$$

$$x_e^{(s,t)} = f_e^{(s,t)} e \in Out(v); v \in CN \tag{39.9}$$

$$f_e^{(s,t)} \ge 0 e \in E; s,t \in V \tag{39.10}$$

Formula (39.3) is the constraint to traffic amount to ensure conservation of outbound and inbound traffic at the node. Formula (39.4) is link load, which shall not exceed link capacity. Formula (39.5) counts the number of traffic splitting, and formula (39.6) ensures traffic splitting by the node will not exceed its processing capacity. Formula (39.7) is the constraint to network congestion cost, which is obtained by transferring the piecewise function in formula (39.1). Formula (39.8) is the constraint to delay cost, which is obtained by transferring the piecewise function in formula (39.2). Formulas (39.9), (39.10) show that split traffic is a special type of traffic, and all traffic is non-negative.

39.2.3 Analysis of Solution Complexity of the Model

Since formula (39.5) is a nonlinear function, the above optimization is a nonlinear planning problem. However, it is easy to find that the root cause of non-linearization is: the number of split traffic by node s of SDN in outbound links is restricted by the service capacity of node s. The number of split traffic by node s of SDN in outbound links is

$$\boldsymbol{P}(\boldsymbol{v}) = \sum_{\boldsymbol{e} \in \boldsymbol{Out}(\boldsymbol{v})} \sum_{(\boldsymbol{s},\boldsymbol{t}) \in \boldsymbol{V} \times \boldsymbol{V}} \boldsymbol{N}(\boldsymbol{x}_{\boldsymbol{e}}^{(\boldsymbol{s},\boldsymbol{t})}) = |\boldsymbol{Out}(\boldsymbol{v})| \times (|\boldsymbol{V}| - \mathbf{1}), \boldsymbol{v} \in \boldsymbol{SV}$$

In ideal conditions, i.e. $\boldsymbol{S}(v) \geq \boldsymbol{P}(v), v \in \mathbf{SV}$, the optimization is translated into a linear problem, whose optimum solution can be obtained in polynomial time [7].

In general conditions, i.e. $\mathbf{S}(v) < \mathbf{P}(v)$, $\boldsymbol{S}(v)$ possible split flows shall be picked up from $\boldsymbol{P}(v)$ first, and then outbound traffic of SDN nodes will be split into $\mathbf{S}(v)$. In this way, the problem becomes seeking the optimum one from $\prod_{\boldsymbol{v} \in \boldsymbol{SV}} \boldsymbol{C}_{\boldsymbol{P}(\boldsymbol{v})}^{\boldsymbol{S}(\boldsymbol{v})}$ solutions of a linear planning problem. From the Stirling formula of combinatorics, we can see that the complexity of search space is $\mathbf{O}\left(\left(2^{(|\boldsymbol{Out}(v)| \times (|\boldsymbol{V}| - \mathbf{1}))}\right)^{|\boldsymbol{SV}|} \right)$. It is clear that method of exhaustion cannot obtain the optimum solution in polynomial time. As a heuristic searching algorithm, the genetic algorithm can effectively solve this searching problem. However, when solving the problem of this paper with the typical genetic algorithm, some individuals no longer belong to the original species after crossover and mutation, which is called parent variants in this paper. Therefore, a Reject Variant Adaptive Genetic Algorithm (RV-AGA) is proposed to search the optimum solution. Based on this, the paper finally puts forward a routing optimization model based on link cost sum and a corresponding Minimal Cost Sum (MCS) routing algorithm to optimize link utilization and transmission delay of SONet under the restriction of SDN-FE.

39.3 Algorithm Implementation

This section introduces implementation of the MCS routing algorithm and the RV-VGA algorithm.

39.3.1 MCS Algorithm

The overall process of the MCS algorithm is as follows and the detailed design of the RV-AGA algorithm is explained in Sect. 39.3.2.

Initial conditions:

(1) Set of OSPF link weights: $\boldsymbol{W}$
(2) Set of SDN nodes: $\boldsymbol{SV}$
(3) Traffic matrix: $\boldsymbol{TM}$
(4) Link capacity matrix: $\boldsymbol{C}$
(5) Network topology: $\boldsymbol{G = (V,A)}$

Implementation:

Step 1: Conduct topological sorting of all the nodes of $\boldsymbol{G}$ to obtain *Seq(**V**)*.

Step 2: Repeat the following steps until the directed acyclic graph DAG*(v)* of all the vertexes (v) of $\boldsymbol{G}$ is constructed.

```
DAG= DijkstraShortestPath (W,v);
foreach link(i,j)∈outgoing_links(SDNi) do
    DAG(i,j) =1;
    if CheckLoop (DAG)==1 then DAG(i,j)=0
    end
end
```

DAG stands for all the possible paths from other nodes of the topology to node v, and DAG(i,j) is the link between node i and node j of DAG. DijkstraShortestPath ($\boldsymbol{W}$,v) calculates the shortest path from other node s to node v, and CheckLoop(DAG) is used to ensure DAG is acyclic.

Step 3: Repeat the following steps, until all the traffic routing conditions of DAG (v) are determined to obtain the constraint expression set *expr_set* in ideal conditions.

```
foreach d∈Seq(V) do
    RouteFlow(d)
end
```

The implementation process of RouteFlow(d) is: if the node is a SDN node, split traffic will be represented by x_i, and be added to the next hop. Using the hash function to map traffic can split the traffic in any proportion. If the node is a traditional IP network node, traffic will be equally distributed among all equivalent shortest paths, and be added to the next hop.

Step 4: Obtain the constraint expression set *expr_set* in ideal conditions.

Step 5: Add restriction of node service capacity to *expr_set* in line with actual conditions to obtain generative rules of the actual constraint expression. Use the RV-AGA algorithm to conduct heuristic search in corresponding solution space of generative rules and find and output the satisfactory solution of routing optimization.

39.3.2 RV-AGA Algorithm

In order to solve the problem in this paper, the RV-AGA algorithm modified the initial population, crossover operator and mutation operator on the basis of standard adaptive genetic algorithm [8]. See topological structure in Fig. 39.2 for details. In Fig. 39.2, subgraph (a) is a network topology composed of 3 nodes and node 1 is a SDN node. The number of each edge in the topology represents OSPF link weight.

1. Selection of the initial population

The paper intends to optimize load balancing of network links, so flows of divisible traffic can be allocated in proportion according to traffic amount. Take topology in Fig. 39.2 as an example. Assume the initial population is 2, $TM_{12} = 100$, and $TM_{13} = 50$, i.e.$(x_1 + x_2) = 100$,and $(x_3 + x_4) = 50$. Since $(x_1 + x_2) : (x_3 + x_4) = 2{:}1$, 2 flows shall be picked from split flows x_1 and x_2, and 1 flow shall be picked from split flows x_3 and x_4.

2. Crossover operator

This paper uses the one-point crossover operator. Assume the codes of two parents are $s_1 = 01|101$ and $s_2 = 10|011$. "|" stands for a cross point, and the offspring after crossover are $s_1' = 01|011$ and $s_2' = 10|101$, as shown in Fig. 39.3.

However, Sect. 4.3.1 reveals that the number of 1 in individual code strings is definite. Therefore, the crossover operator in this paper must exclude the situation when the number of 1 in individual code strings of the offspring does not conform

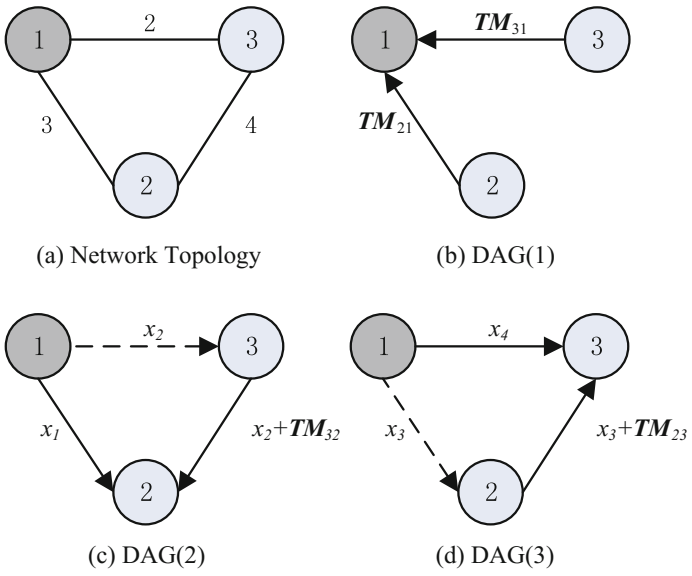


Fig. 39.2 SONet topological structure with 3 nodes

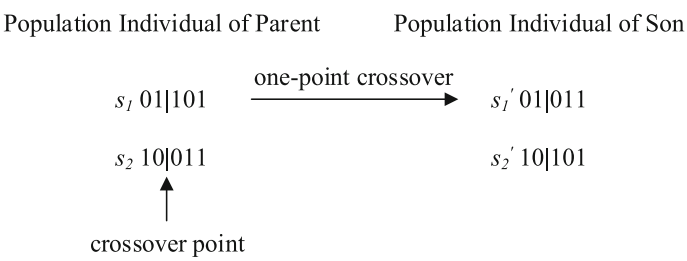


Fig. 39.3 Operation of crossover operators under constraints

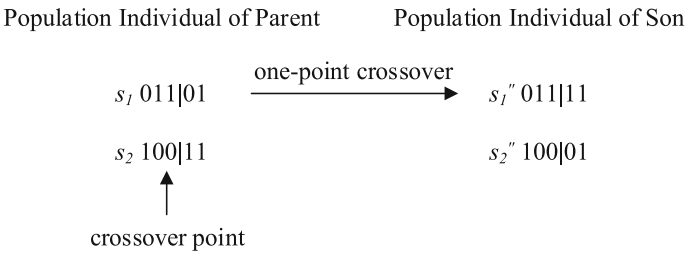


Fig. 39.4 Operation of crossover operators that violates constraints

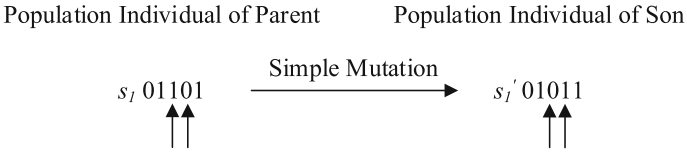


Fig. 39.5 Operation of mutation operators under constraints

to the constraints after crossover. As shown in Fig. 39.4, assume the number of 1 in individual code strings is 3, the parents are still s_1 and s_2, and location of the cross point is changed, then the offspring will be s_1'' and s_2''. The number of 1 in s_1'' and s_2'' is 4 and 2 respectively, failing to conform to constraints.

3. Mutation operators

Similar to crossover operators, the paper must evaluate the number of 1 in new individuals after mutation to judge whether constraints are followed.

Figure 39.5 shows the mutation scenario under constraints, and Fig. 39.6 shows the mutation scenario that do not conform to constraints Generally, when mutation simultaneously occurs in an even number of genes and the numbers of 0 and 1 are the same, the offspring conform to constraints after mutation.

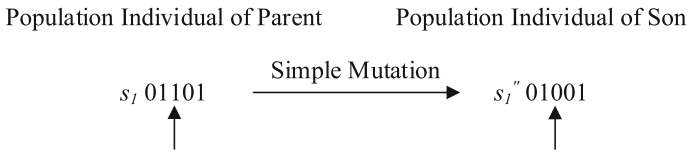


Fig. 39.6 6 Operation of mutation operators that violates constraints

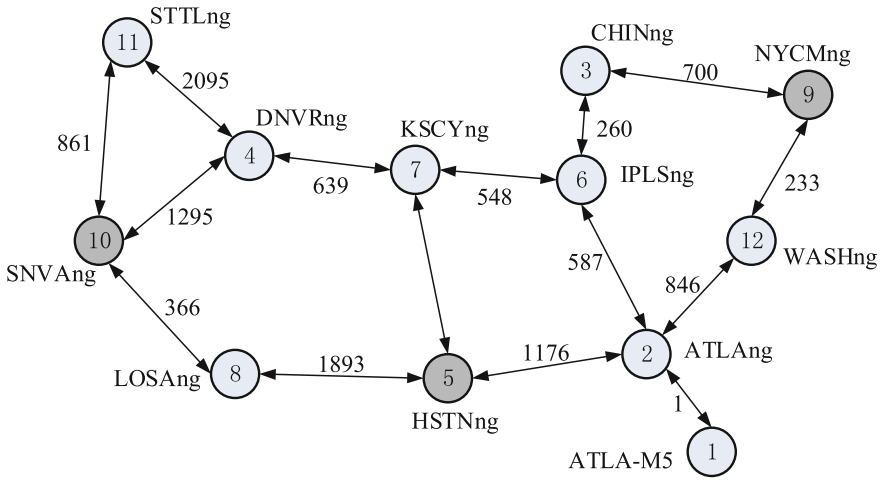


Fig. 39.7 Hybrid SDN topology for simulation

39.4 Simulation and Analysis

In order to verify effectiveness of the algorithm proposed in this paper, use Abilene network data collected by Zhang from 2004.3.1 to 2004.3.7 [9], as shown in Fig. 39.7, Assume the grey nodes represent SDN nodes, and others are traditional IP network nodes running the OSPF protocol [10].

39.4.1 Algorithms and Parameters Setting

The core of the MCS algorithm is the RV-AGA algorithm, and the setting of related parameters is: Maximum number of generations $T = 50$, initial population $M = 100$, and initial crossover probability $p_{c1} = 0.8$ and $p_{c2} = 0.7$; initial mutation probability $p_{m1} = 0.01$ and $p_{m2} = 0.006$. In ideal conditions, the maximum number of split flows supported by a SDN node is $S(v_5) = 29$, $S(v_9) = 15$, and $S(v_{10}) = 24$; in actual conditions, the maximum number of split flows supported by a SDN node is $P(v_5) = 25$, $P(v_9) = 13$, and $P(v_{10}) = 20$.

39.4.2 RV-AGA Performance Analysis

Performance changes in the iterative process are tracked to analyze performance of RV-AGA. Figure 39.8 demonstrates the changing trend of the corresponding objective function value of the optimum individual in the iterative process when the traffic data is the traffic matrix at 00:05 on March 1, 2004.

We can see that RV-AGA can rapidly limit the objective function value around the optimum solution in the search space. After about 30 iterations, change of the average value of the population stabilizes at [397,739.2968,397,921.8187], and the optimum objective function value stabilizes at [397,509.6868,397,522.0831], which are quite close to the theoretical optimum solution: 397,498.0000.

39.4.3 Effectiveness Analysis of MCS Load Balancing

In order to compare the effectiveness of the MCS algorithm and the SOTE algorithm in network balancing, use the two algorithms to optimize the same traffic matrix, and take maximum link utilization (MLU) of the network after optimization as the indicator for load balancing. The network traffic data comes from 288 traffic matrices in 24 h of March 3, 2004, and the experiment result is shown in Fig. 39.9. In light network load (6–12 h), the two algorithms have little difference in optimization performance with MLU improving by about 1.10% at most. However,

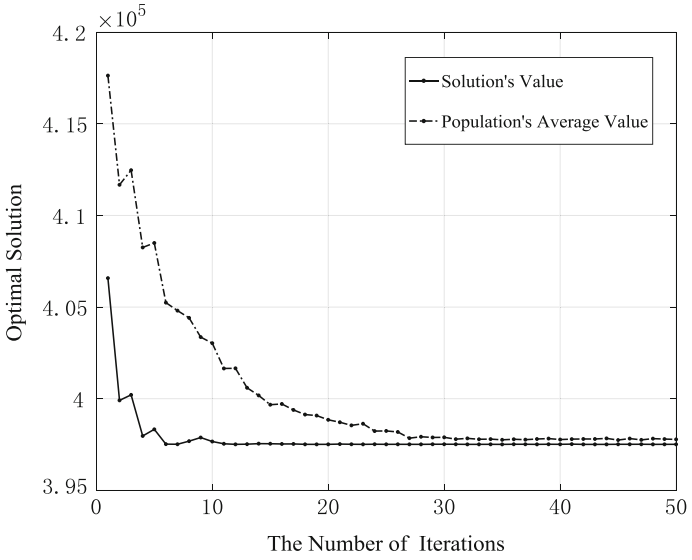


Fig. 39.8 Tracking of RV-AGA performance

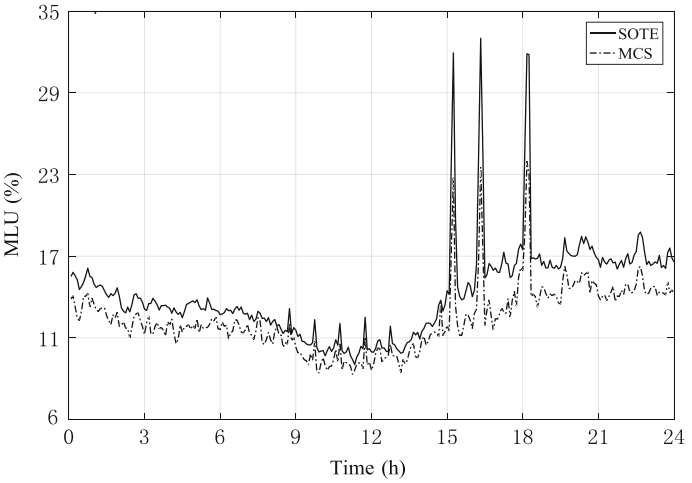


Fig. 39.9 Comparison of MCS and SOTE in load balancing

when network balancing becomes heavier (such as 15–24 h), MCS demonstrates much better optimization performance than SOTE, especially at traffic peaks at 15:15, 16:50, and 18:05 with MLU down by 9.10, 9.47, and 8.82% respectively.

39.5 Conclusion

This paper puts forward a routing optimization model and a corresponding optimization algorithm based on link cost sum, the objective of which is minimum link cost sum from the perspective of the overall network. The key point is that the model includes actual processing capacity of SDN nodes in hybrid SDN into constraints to realize optimization of network links under actual restrictions, overcoming limitations of current models. The experiment result shows that the algorithm proposed by this paper has better network load balancing than the current routing algorithms of hybrid SDN.

References

1. Lin P, Hart J, Krishnaswamy U et al (2013) Seamless Interworking of SDN and IP[J]. ACM Sigcomm Comput Commun Rev 43(4):475–476
2. Caria M, Jukan A, Hoffmann M (2013) A performance study of network migration to SDN-enabled Traffic Engineering[C]// Global Communications Conference. IEEE, pp 1391–1396

3. Agarwal S, Kodialam M, Lakshman TV (2013) Traffic engineering in software defined networks[C]//INFOCOM. In: 2013 Proceedings IEEE. IEEE, pp 2211–2219
4. Guo Y, Wang Z, Yin X, et al (2014) Traffic Engineering in SDN/OSPF Hybrid Network[C]// Network Protocols (ICNP), 2014 IEEE 22nd International Conference on. IEEE, pp 563–568
5. Fortz B, Thorup M (2000) Internet traffic engineering by optimizing OSPF weights[C]// INFOCOM 2000. In: Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE. IEEE, 2: 519–528
6. Bharti S, Pattanaik KK (2014) Dynamic distributed flow scheduling for effective link utilization in data center networks[J]. J High Speed Netw 20(1):155–158
7. Yang M, Zheng KQ, Lei YJ Method of Uncertain Reasoning Based on IFBN[J]. Fire Control & Command Control, 2015
8. Guo YH, Liu FF, Zhang N et al (2014) Joint optimization combining genetic algorithm and numerical algorithm based on MATLAB[J]. Adv Mater Res 1065–1069:3442–3445
9. Zhang Y (2009) 6 months of Abilene traffic matrices http://www.cs.utexas.edu/ ~ yzhang/
10. Jayakumar M, Rekha NRS, Bharathi B (2015) A comparative study on RIP and OSPF protocols[C]// International Conference on Innovations in Information, Embedded and Communication Systems. IEEE

Chapter 40
Parameterized Unified Modulation Model Design for Satellite Communications

Yabo Yuan, Bo Wang and Bin Wu

40.1 Introduction

Nowadays, to meet the globalisation and full-coverage demand in wireless communications, satellite networks are supposed to be built into the so-called 'backbone in the air', which provide access and information transmission services to other satellites and ground users in next-generation digital communications [1, 2]. At present, the main modulation types for satellite and ground users include BPSK, QPSK, M-PSK, M-QAM, and so on [3]. Furthermore, continuous phase modulation (CPM) with a constant envelope and continuous phase [4] has been applied to satellite communications to overcome the non-linear problem associated with the high-power amplifiers (HPAs) in satellites [5]. Moreover, in the extremely high frequency (EHF) broadband satellite field [6], many potential modulation schemes with novel chip waveforms have been proposed to mitigate inter-channel interference [6, 7]. Thus, the satellite network must overcome the challenges of multiple users, variable access, and highly diversified modulation schemes. To efficiently satisfy the communication needs of different users, a 'one-stop shop' which can provide nearly universal real-time handover and multiple modulation services is especially necessary and urgent.

Recently, there have been quite a few studies on the unified modulation technology of existing communication systems. Kuei-Chiang Lai et al. proposed a CPM/PSK unified modulator for the Global System for Mobile Communication (GSM) [8], and Hao Lin et al. proposed a unified structure for multi-carrier modulation in power-line communications [9]. These two technologies obtain unification by combining the common modules of different modulators, which is easy to design and realise but has insufficient compatibility and flexibility. Moreover, the structural efficiency of the modulators is quite low. Francois D. Cote et al. proposed

Y. Yuan (✉) · B. Wang · B. Wu
Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China
e-mail: yybsky0523@163.com

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_40

a unifying framework which can generate nearly all signals employed by the Global Navigation Satellite Systems (GNSS) including GPS and Galileo by appropriately positioning and scaling a sequence of Dirac delta functions [10]. The structural efficiency of this modulation framework is higher; however, the application of the model is limited to GNSS signals. Marcus L. Roberts et al. developed a general spectrally modulated, spectrally encoded signal framework for 4G communication systems [11, 12], which can generate various signals through frequency-domain control. However, the modulation mapping rules, which are not consistent for different modulation schemes, are not considered in this model. In conclusion, there still exist quite a few limitations for the application of existing unified modulation technology which should be further studied, especially with regard to improving the modulation structure integration, enhancing the signal expression flexibility, and designing consistent mapping rules.

In this study, a novel parameterised sub-component decomposition and state-mapping-based unified modulation model (SDSM-UMM) is first developed, which has the features of structural efficiency, signal expression flexibility, and mapping rule consistency. The proposed SDSM-UMM model not only cover popular existing modulation schemes but also describe potential and even completely new modulation schemes, which represent a substantial contribution to next-generation satellite communications.

40.2 Problem Analysis

In digital communication systems, a modulation process is equivalent to mapping a data sequence into analogue waveforms which match the channel characteristics. It mainly consists of two parts: a modulated signal expression and waveform mapping.

The modulated signal expression contains three aspects. The first one is the expression of chip waveforms. A chip waveform varies with the modulation. Despite common ones such as rectangular, half-sine, and raised-cosine pulses, some novel designed waveforms are widely used in different application scenarios. The second aspect is the expression of a waveform transformation. A chip waveform is transformed into a modulated waveform which carries information through a waveform transformation. The waveform transformation includes the transformation of the amplitude, phase, frequency, position, and so on. Various chip waveforms and their transformation methods require high signal expression flexibility and adaptability. The last aspect is the modulated signal frameworks, which divide modulation schemes into linear ones and non-linear ones. If the modulated signal is a linear superposition of modulated waveforms at each time slot, the modulation is a linear modulation; otherwise, it is a non-linear modulation. Significant differences in linear and non-linear modulated signal frameworks cause difficulties in the integration of a modulator structure, where the non-linearity greatly increases the system complexity.

Another main part of modulation is waveform mapping, which means to map the incoming information data into an element of the modulated waveform set. If the modulation mapping process is only determined by the current information data, the modulation is said to be memoryless. Otherwise, the modulation has memory if the previous data sequence influences the mapping process. At present, the modulator mapping structures of different modulation schemes are designed on the basis of the diversity of mapping rules, which incurs serious resource and time consumption during modulation handover.

40.3 Unified Modulation Model

According to the analysis in Sect. 40.2, we can see that it is necessary to develop a flexible signal expression method and a corresponding consistent state mapping algorithm in order to realise multiple modulation schemes with a unified system model. This section presents the unified modulation model using sub-component decomposition and the state mapping approach.

40.3.1 Unified Signal Expression Method

In this subsection, we develop a sub-component-decomposition-based linear signal expression method. First, we express the non-linear modulated signal as a sum of time-limited modulated waveforms. Then, each modulated waveform is decomposed into a set of sub-components whose coefficients are alterable. Therefore, the modulated signals of different modulation schemes can be linearly generated through parameter settings.

40.3.1.1 Modulated Signal Linearisation

Suppose the bit period is T_b, the sampling period is T_s, and the number of samples is N_s ($N_s = T_b/T_s$) per bit period, where N_s is an integer. We denote the number of samples per symbol period for M-ary modulation as N_M, $N_M = N_s{*}\log_2 M$. For M-ary linear modulation, the modulated signal $s[n]$ can be expressed as

$$s[n] = \sum_{m=-\infty}^{\infty} g^{(m)}[n - mMN_s] \tag{40.1}$$

where $g^{(m)}[n], n \in [0, (L+1)MN_s)$ is the modulated waveform mapped at the mth symbol period, and $L(L \in \mathbb{N})$ is the memory span parameter. For memoryless modulation, $L = 0$, and for memory modulation, $L > 0$.

Non-linear modulation mainly includes various CPM schemes [4]. According to Laurent in [13], binary CPM can be expressed as a finite number of time-limited amplitude-modulated pulses (AMP), and further research shows that *M*-CPM shares similar properties [14, 15]. In this study, we take binary CPM as an example to introduce signal linearisation and modulation unification for non-linear modulation, and the method can be further expanded to *M*-CPM. First, we denote the modulation index as h, the phase shift function as $q(t)$, the phase shift function as $q[n]$, the bipolar data sequence as $\{d_n\}$. Parameter $B \in \{0, 1, \ldots, 2^{L-1}-1\}$ can be decomposed into L bit binary number, where $B[i]$ denotes the ith bit. For example, for $L = 2$, $B = 2$, $B[0] = 0$, $B[1] = 1$. Then, we express the CPM signal with (40.1) by AMP decomposition [16], where the modulated waveform is

$$g^{(m)}[n] = \sum_{B=0}^{2^{L-1}-1} e^{j\pi h D_{B,m}} C_B(nT_s - mT_b) \tag{40.2}$$

where $n = 0, 1, \ldots, (L+1)N_s - 1$.

$$D_{B,m} = \sum_{v=-\infty}^{m} d_v - \sum_{i=1}^{L-1} B[i] d_{m-i} \tag{40.3}$$

$$\begin{aligned} C_B(t) &= \frac{\sin q'(t)}{\sin h\pi} \prod_{i=1}^{L-1} \frac{\sin q'(t + (i + B[i]L)T_b)}{\sin h\pi} \\ q'(t) &= \begin{cases} q(t), & 0 \le t < LT_b \\ h\pi - q(t - T_b), & LT_b \le t \le 2LT_b \end{cases} \end{aligned} \tag{40.4}$$

40.3.1.2 Signal Model

We decompose the *M*-ary modulated signal into a set of sub-components, and the unified modulated signal expression can be formulated as follows:

$$s[n] = \sum_{m=-\infty}^{\infty} \sum_{k=0}^{K-1} w_k a_k^{(m)}[n - mMN_s] \psi_k[n - mMN_s] \tag{40.5}$$

where $\boldsymbol{\psi}_k (k = 0, 1, \ldots, K-1)$ indicates a sub-component in the sub-component set which contains K elements; $\boldsymbol{a}_k^{(m)}$ is a modulated waveform mapping coefficient vector of which the superscript m indicates the mth symbol period, and the subscript k corresponds to sub-component $\boldsymbol{\psi}_k$; and $w_k (k = 0, 1, \ldots, K-1)$ is a sub-component ON–OFF indicator. The details of the main parameters in (40.3) are presented in the following parts of this subsection.

(1) Sub-component

$\boldsymbol{\psi}_k=[\psi_k[0],\psi_k[1],\ldots,\psi_k[(L+1)MN_s-1]]^{\mathrm{T}}$ denotes the kth sub-component, where $\psi_k[n]$ is the value of $\boldsymbol{\psi}_k$ at the moment nT_s, and $\psi_k[n]$ is equal to zero when n is out of range. Each sub-component is constructed with a cascade of $L+1$ normalised base vectors $\boldsymbol{\varphi}_k=[\varphi_k[0],\varphi_k[1],\ldots,\varphi_k[MN_s-1]]^{\mathrm{T}}$, $\boldsymbol{\psi}_k=[\boldsymbol{\varphi}_k^{\mathrm{T}},\ldots,\boldsymbol{\varphi}_k^{\mathrm{T}}]^{\mathrm{T}}/\sqrt{L+1}$. The base vectors are orthogonal; thus, the combined base matrix $\boldsymbol{\Phi}=[\boldsymbol{\varphi}_0,\boldsymbol{\varphi}_1,\ldots,\boldsymbol{\varphi}_{K-1}]$ satisfies

$$\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}=\boldsymbol{I} \tag{40.6}$$

which guarantees the intra- and inter- symbol orthogonality of the sub-components, i.e.

$$\sum_{n=-\infty}^{\infty}\psi_i^*[n]\psi_j[n+mMN_{\mathrm{s}}]=\begin{cases}1-\frac{|m|}{L+1} & i=j,|m|\le L\\ 0 & i\ne j\ or\ |m|>L\end{cases}. \tag{40.7}$$

The choice of the base matrix $\boldsymbol{\Phi}$ has a direct effect on the complexity and accuracy of the modulated waveform expression; that is, when the characteristics of the base vectors are consistent with the modulated waveform characteristics, the modulated waveforms can be precisely denoted by a low-dimensional base matrix. In this study, we use prolate spheroidal wave functions (PSWFs), which are widely used in ultra-wideband communications, to form the base vectors. The PSWFs are energy concentrated in both the time and frequency domains; moreover, they are fully programmable with regard to the size and spectrum shape [17], which make the PSWFs qualified in the representation of manifold band-limited signals. Because the closed form of the PSWF is difficult to obtain, a discretisation algorithm was employed by Parr to obtain a numerical solution [18]:

$$\lambda_k\varphi_k[n]=\sum_{m=0}^{T/T_{\mathrm{s}}}\varphi_k[m]\frac{\sin 2\pi B[[n-m]T_{\mathrm{s}}]}{\pi[n-m]T_{\mathrm{s}}} \tag{40.8}$$

where B is the energy-concentrated bandwidth of the PSWFs; and λ_k is the energy concentration factor, and $T=(L+1)MN_sT_s$. The base vectors obtained from (40.8) meet the requirements of (40.6).

(2) Sub-component ON–OFF indicator

We define $w=[w_0,w_1,\ldots,w_{K-1}]$ as the sub-component ON–OFF indicator vector, where $w_k\in\{0,1\}$ is used to control the mode of the kth sub-component with $w_k=1$ indicating the ON mode and $w_k=0$ indicating the OFF mode. Its main effects are reflected in two aspects. One is to choose effective sub-components. Because multiple modulation schemes share a corporate sub-component set in our model, it is quite reasonable that a particular modulated signal would correspond to part of the sub-components which have similar characteristics, such as the odd-even

property and spectrum range, with the signal. By setting the ON–OFF indicators of these sub-components to one and the rest to zero, the signal expression can be simplified. The other effect is to improve the signal-to-noise ratio (SNR) by shutting down sub-components whose projected SNR is lower than a given threshold.

(3) Modulated waveform mapping coefficient

$A^{(m)} \triangleq [\boldsymbol{a}_0^{(m)}, \boldsymbol{a}_1^{(m)}, \ldots, \boldsymbol{a}_{K-1}^{(m)}]$ is the modulated waveform mapping coefficient matrix, and the modulation mapping relation is equal to the corresponding relation between $A^{(m)}$ and the data sequence. $\boldsymbol{a}_k^{(m)} = \left[a_k^{(m)}[0], \ldots, a_k^{(m)}[(L+1)MN_\mathrm{s} - 1]\right]^\mathrm{T}$ is the kth projection coefficient vector of the modulated waveform. For memoryless modulation, $a_k^{(m)}[n]$ only relates to m and k, while for modulation with memory, $a_k^{(m)}[n]$ also relates to n. Combine (40.1) with (40.5), $a_k^{(m)}[n]$ is obtained as

$$a_k^{(m)}[n] = a_k^{(m,l)}, n \in [lMN_\mathrm{s}, (l+1)MN_\mathrm{s}) \tag{40.9}$$

where $a_k^{(m,l)} = \sqrt{L+1} \sum_{v=0}^{MN_\mathrm{s}-1} g^{(m)}[v + lMN_\mathrm{s}] \varphi_k^*[v], l = 0, \ldots, L.$

40.3.2 Unified Modulation Mapping Algorithm

As mentioned in the previous section, modulation mapping is equivalent to the corresponding relation between $A^{(m)}$ and the data sequence, and it varies with the modulation type. By defining and designing the mapping function, a novel mapping state set which processes a universal one-to-one mapping rule with the data sequence is constructed in this section; thus, a consistent modulation mapping algorithm is obtained. For M-ary modulation, by denoting the data sequence as $I_m \in \{0, 1, \ldots, 2^M - 1\}$, the definitions of the mapping function and mapping state set are presented as follows.

Definition 1 We define the mapping function as $f(\boldsymbol{x}^{(m)})$, which satisfies the following conditions: (1) the independent variable $\boldsymbol{x}^{(m)}$ of the mapping function $f(\boldsymbol{x}^{(m)})$ consists of data sequence items at or before the mth symbol period, and it meets the following requirement: by traversing its own definition domain, the corresponding $A^{(m)}$ can attain all its possible values. (2) The value of $f(\boldsymbol{x}^{(m)})$ and its independent variable $\boldsymbol{x}^{(m)}$ have a one-to-one correspondence with each other. (3) The range of $f(\boldsymbol{x}^{(m)})$ is $\{0, 1, \ldots, M_0 - 1\}$, where M_0 is equal to the minimum state number of the independent variable $\boldsymbol{x}^{(m)}$ which satisfies condition 1). (4) $f(\boldsymbol{x}^{(m)})$ can be obtained with the values of I_m and $f(\boldsymbol{x}^{(m-1)})$.

Definition 2 We define the mapping state at the mth symbol period as $\mathrm{state}^{(m)}$ on the basis of the mapping function, where

$$\text{state}^{(m)} \triangleq [f(\boldsymbol{x}^{(m)}), A^{(m)}] \tag{40.10}$$

Furthermore, we define the value set of $\text{state}^{(m)}$ as the mapping state set $\{\text{state}_i, i = 0, 1, \ldots, M_0 - 1\}$,

$$\text{state}_i \triangleq [i, A_i] \tag{40.11}$$

where $A_i = [\boldsymbol{a}_{i0}, \boldsymbol{a}_{i1}, \ldots, \boldsymbol{a}_{iK-1}]$. It can be proved that $\{f(\boldsymbol{x}^{(m)})\}$ is a time-homogeneous Markov chain. Therefore, we indicate the corresponding mapping waveform where $f(\boldsymbol{x}^{(m)}) = i$ as $g_i[n]$. According to (40.9) and (40.11),

$$a_{ik} = \left[a_{ik}^0 \cdot \mathbf{1}_{1\times MN_s}, \ldots, a_{ik}^L \cdot \mathbf{1}_{1\times MN_s}\right]^{\mathrm{T}} \tag{40.12}$$

where $a_{ik}^l = \sqrt{L+1}\sum_{v=0}^{MN_s-1} \varphi_k^*[v] g_i[v + lMN_s]$. The definition of the mapping state set ensures that it is a set of certainty, mutual dissimilarity, and non-sequentiality.

Mapping function characteristics (1)–(3) in Definition 1 guarantee that the values of $\text{state}^{(m)}$ determined by data sequence have a one-to-one correspondence with the elements in the mapping state set $\{\text{state}_i\}$, and the mapping rule is

$$\text{state}^{(m)} = \text{state}_{f(\boldsymbol{x}^{(m)})}. \tag{40.13}$$

From the above analysis, we can obtain the following unified modulation mapping algorithm.

First, we design the mapping function $f(\boldsymbol{x}^{(m)})$ which satisfies the conditions in Definition 1. For M-ary memoryless modulation, $f(\boldsymbol{x}^{(m)}) = I_m$, where the independent variable $\boldsymbol{x}^{(m)} = I_m$ and the state set size $M_0 = 2^M$. For memory modulation, e.g. binary CPM whose modulation index $h = c/p$ (c and p are coprime), the mapping function is

$$f(\boldsymbol{x}^{(m)}) = \left(\sum_{i=-\infty}^{m-L} d_i\right) \bmod p_0 + p_0 \sum_{l=1}^{L} 2^{l-2}(d_{m+1-l} + 1)$$

where $p_0 = p$ when m is even and $p_0 = 2p$ when m is odd. The independent variable $\boldsymbol{x}^{(m)} = \left[d_m, d_{m-1}, \ldots, d_{m-L+1}, \sum_{i=-\infty}^{m-L} d_i\right]$, and the state set size $M_0 = p_0 2^L$. Then, the mapping waveform $g_i[n]$ is determined by the mapping function. For memoryless modulation $g_i[n] = g[n|I^{(m)} = i]$, and for memory modulation, such as binary CPM, according to (40.2)–(40.4),

$$g_i[n] = e^{j\pi h\left(i - p_0\left\lfloor \frac{i}{p_0}\right\rfloor\right)} \sum_{B=0}^{2^{L-1}-1} e^{j\pi h\left((2*\left\lfloor \frac{i}{p_0}\right\rfloor(0) - 1 + \sum_{w=1}^{L-1}[1-B(w)](2*\left\lfloor \frac{i}{p_0}\right\rfloor(w) - 1)\right)} C_B(nT_s).$$

Next, the mapping waveform is substituted into (40.11) and (40.12), and the mapping state set is determined. Finally, we can obtain the current state state$^{(m)}$ based on the mapping rule formulated in (40.13), and the modulated waveform mapping coefficient vector $A^{(m)}$ is obtained.

40.3.3 Unified Modulation Model Structure

According to the sub-component-decomposition-based signal expression in Sect. 40.3.1 and the corresponding consistent mapping-function-based modulation mapping algorithm in Sect. 40.3.2, a novel unified modulation model, SDSM-UMM, is obtained, as shown in Fig. 40.1. This unified modulator structure consists of three parts: the modulation control parameter (MCP) part, state mapping (SM) part, and signal representation (SR) part. The MCP part is used for system assignment, and it chooses the desired system status including the modulation type, modulation parameters (chip waveform, modulation order, memory span parameter, modulation index, etc.), transmission rate, and so on. The main parameters consist of the mapping function $f(\boldsymbol{x}^{(m)})$, mapping state set $\{\text{state}_i\}$, sub-component parameters $P = [K, L, M, N_s, T_s, B]$, and sub-component ON–OFF indicator vector $\boldsymbol{w}$. The SM part uses the incoming data I_m to calculate the value f of the mapping function, which determines the current state from the stored mapping state set $\{\text{state}_i\}$ according to the mapping rule in (40.13), and then outputs the modulated waveform mapping coefficient matrix A_f. The SR part stores the sub-component set, and the sub-component ON–OFF indicator vector $\boldsymbol{w}$ is used to control each sub-component. By substituting A_f and $\boldsymbol{w}$ into (40.5), the SDSM-UMM signal $s[n]$ is generated.

The SDSM-UMM model developed in this study also has some extended functionality, such as the realisation of the DS/FH spread spectrum by introducing a pseudo-code or frequency-hopping sequence to the modulated waveform mapping coefficient matrix, and the realisation of OFDM by changing the sub-components in the parallel structure of the SE part to exponential functions.

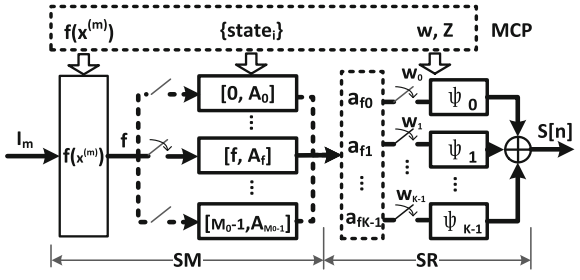


Fig. 40.1 SDSM-UMM model structure

40.4 Simulation and Analysis

In this section, the applicability and feasibility of the proposed SDSM-UMM model is verified by (1) representatively modelling several typical modulated signals and calculating the normalised mean-squared error (NMSE) of the simulation results and (2) verifying the analytical ACF and PSD expressions with numerical simulations and examining whether they are consistent with the classical results. We denote the modulated signal generated through practically implementing the SDSM-UMM model as the SDSM-UMM signal and the signal generated through conventional modulation methods as the conventional signal. The theoretical and numerical results demonstrate that the SDSM-UMM signals are almost identical to the conventional signals, which reinforces the practical utility of the proposed SDSM-UMM model.

40.4.1 Results and Analysis for the NMSE

We perform Monte Carlo simulations using the PSWF-based SDSM-UMM model, and the main simulation parameters are listed in Table 40.1.

The SDSM-UMM and conventional signals for memoryless linear modulation using BPSK, QPSK, 8PSK, and 16QAM and memory nonlinear modulation using MSK ($L = 1$, $h = 0.5$) and GMSK ($L = 4$, $h = 0.5$, BT = 0.3) are simulated using the Monte Carlo method with 10,000 bits of randomly generated binary data. By changing the size K of the sub-component set, variation in the NMSE of the SDSM-UMM and conventional signals versus K is obtained, as shown in Fig. 40.2. It can be seen that the NMSE rapidly decreases as K increases. This is because the sub-component set constructed by the first K PSWF functions which have the maximum energy concentration constitutes a K-dimensional signal space, and the SDSM-UMM signal is equivalent to the projection of the conventional signal onto the K-dimensional signal space. For a higher number of signal space dimensions, the SDSM-UMM signal is more consistent with the conventional signal; thus, the NMSE is smaller. Furthermore, the accuracy of the signal expression as well as the expression complexity are enhanced as K increases; therefore, a trade-off is required.

More specifically, it is obvious from Fig. 40.2 that the NMSE curves rapidly decrease to less than 0.03 at the point $K = 7$ and gradually decrease after that. Thus,

Table 40.1 Simulation Parameters

T_b, bit period	1 s
B, energy-concentrated bandwidth of the PSWF	1 MHz
N_s, number of samples per bit period	100
$\boldsymbol{w}$, sub-component ON–OFF indicator vector	**1**
t, simulation time	10000 T_b

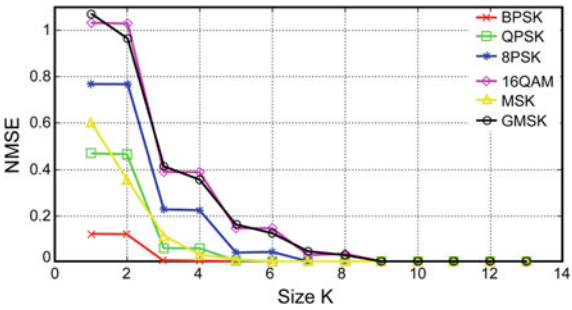


Fig. 40.2 NMSE of SDSM-UMM and conventional signals versus the sub-component set size K

we can choose the first seven PSWF functions to construct the sub-component set of the modulator. In addition, the degradation in the NMSE at the point $K = k$ indicates the contribution of the kth sub-component to the signal expression; therefore, it can be used to indicate the selection of w_k. For example, in Fig. 40.2, the NMSE curves of memoryless signals which are even symmetric in the projective interval hover at the even points of K because of the odd symmetry of the even-order PSWF functions. Thus, we can fix $w_{2k} = 0$ for memoryless signals to simplify the signal expression without lowering the expression accuracy.

40.4.2 Results and Analysis for the ACF and PSD

In this subsection, we examine the analytical statistical expressions of two typical modulation signals, BPSK and MSK. According to the previous subsection, we choose the sub-component set size $K = 7$ and fix $\boldsymbol{w} = [1, 0, 1, 0, 1, 0, 0]$ for BPSK and $\boldsymbol{w} = [1, 1, 1, 1, 1, 1, 1]$ for MSK, i.e. three and seven sub-components are used to express the BPSK and MSK signals, respectively. The other parameters are consistent with those in Table 40.1.

We design numerical simulations using the Monte Carlo method. The empirical ACF [4] is obtained from the circular finite-time autocorrelation of 10,000 bits of randomly generated binary data; then, it is normalised with respect to the maximum magnitude. The PSD [4] is obtained from the 10,000-point discrete Fourier transform of the normalised empirical ACF. The simulation results of the empirical ACF and PSD are compared with the analytical results of the conventional signals. The normalised results are shown in Fig. 40.3. As can be seen, the simulation results of the two second-order statistics of the SMSE-UMM signal are basically identical to those of the conventional signal, which demonstrates that the proposed SDSM-UMM model can provide a fixed applicability and feasibility with seven sub-components. Moreover, for simple modulated signals, such as a BPSK signal, the model can be simplified by adaptively assigning the sub-component ON–OFF indicator vector.

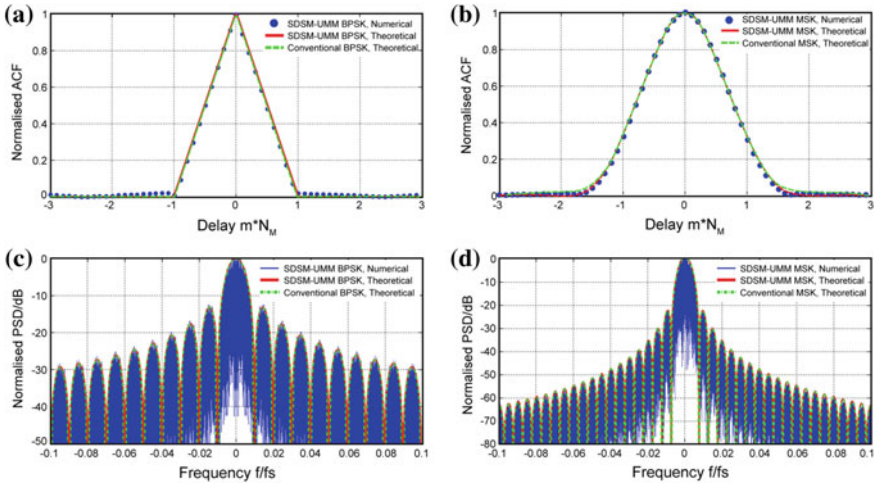


Fig. 40.3 Numerical and theoretical results for the second-order statistics. **a** Normalised ACF of BPSK, **b** normalised ACF of MSK, **c** normalised PSD of BPSK, **d** normalised PSD of MSK

40.5 Conclusion

In this paper, we first developed a novel parameterised sub-component decomposition and state-mapping-based unified modulation model. First, we presented a unified modulated signal expression method which can linearly express an arbitrary modulated signal through sub-component decomposition. Then, we used the waveform mapping coefficient vector obtained through the signal expression to structure a time-homogeneous Markovian mapping state set. By designing mapping functions, we developed a general one-to-one mapping rule between the mapping state set and the data sequence. Finally, Monte Carlo simulations were performed for the SDSM-UMM signals. The simulation results of the NMSE and second-order statistics suggested that our model can generate both linear and non-linear, memoryless and memory modulated signals with seven sub-components, of which the characteristics were virtually identical to the conventional ones. With the features of a high structural efficiency, strong signal expression flexibility, and modulation mapping-rule consistency, the proposed SDSM-UMM model can provide feasible ideas for solving the problem of multiple users in satellite communications. In addition, there are some extended functions of the SDSM-UMM model, such as the spread spectrum and OFDM, which will be studied in future work.

References

1. Ibnkahla M, Rahman QM, Sulyman AI et al (2004) High-speed satellite mobile communications: technologies and challenges. Proc IEEE 92(2):312–339
2. Chan S, Division M (2005) Architectures for a space-based information network with shared on-orbit processing. Massachusetts Institute of Technology. Boston, pp 1–65
3. Gérard M, and Bousquet M (2011). Satellite communications systems: systems, techniques and technology. Wiley, New york
4. Proakis JG (2011). Digital communications, 5th edn. McGraw-Hill Higher Education. New york
5. Remlein P (2014) Energy efficient continuous phase modulation signals for satellite intelligent transportation systems. IET Circ Dev Syst 8(5):406–411
6. Cianca E, Rossi T, Yahalom A et al (2011) EHF for satellite communications: the new broadband frontier. Proc IEEE 99(11):1858–1881
7. Sacchi C, Rossi T, Ruggieri M et al (2011) Efficient waveform design for high-bit-rate W-band satellite transmissions. IEEE Trans Aerosp Electron Syst 47(2):974–995
8. Lai KC, O'shea HD (2011) Unified modulator for continuous phase modulation and phase-shift keying. U.S. Patent 8,009,761
9. Lin H, Siohan P (2010) Modulation flexibility in PLC: a unified MCM transceiver design and implementation. IEEE Trans Circ Syst I Regul Pap 57(10):2762–2775
10. Cote FD, Psaromiligkos IN, Gross WJ (2011) GNSS modulation: a unified statistical description. IEEE Trans Aerosp Electron Syst 47(3):1814–1836
11. Roberts ML, Temple MA, Raines RA et al (2007) Communication waveform design using an adaptive spectrally modulated, spectrally encoded (SMSE) framework. IEEE J Select Top Signal Process 1(1):203–213
12. Roberts M, Temple MA, Oxley ME et al (2006) A general analytic framework for spectrally modulated, spectrally encoded signals. In: IEEE International Conference on Waveform Diversity and Design, pp 1–5
13. Laurent P (1986) Exact and approximate construction of digital phase modulations by superposition of amplitude modulated pulses (AMP). IEEE Trans Commun 34(2):150–160
14. Mengali U, Morelli M (1995) Decomposition of M-ary CPM signals into PAM waveforms. IEEE Trans Inf Theory 41(5):1265–1275
15. Perrins E, Rice M (2005) PAM decomposition of M-ary multi-h CPM. IEEE Trans Commun 53(12):2065–2075
16. Huang K (2005) Supplementary proof for" Exact and approximate construction of digital phase modulations by superposition of AMP" by P. A. Laurent. IEEE Trans Commun 53 (2):234–237
17. Sacchi C et al (2008) Utilization of UWB transmission techniques for broadband satellite connections operating in W-band. In: IEEE GLOBECOM Workshops, pp 1–5
18. Parr B, Cho B, Wallace K (2003) A novel ultra-wideband pulse design algorithm. IEEE Commun Lett 7(5):219–221

Chapter 41
Study on the Influencing Factors of Frequency Locked Loop Based on Stochastic Resonance

Weitong Zhang, Zhiqiang Li, Huan Chen and Shengchao Shi

41.1 Introduction

Carrier tracking plays an extremely essential role in digital communication system. As an important part of carrier tracking, the frequency locked loop can realize continuous and stable tracking when the signal changes dynamically. However, the FLL has noise disadvantages [1]. In order to improve the performance of FLL, a lot of efforts have been made. In order to get better dynamic performance, a method has been put forward with broad loop bandwidth, which results an poor loop tracking performance [2, 3]. By using the technologies of wild-value elimination and bandwidth switching, the FLL gets better frequency tracking performance with an unusually complex and cumbersome structure [4]. Based on the weakening effect of parameter-induced stochastic resonance [5–9], an FLL based on SR has been designed and its tracking performance has greatly improved [10].

The ingenious combination with FLL and SR is an adventurous innovation. So far, only few works on this field have appeared. Because of the no-clear influence factors of FLL based on SR, a study with signal to noise ratio, sampling frequency and system parameters has been carried out in this paper. This work composes five sections: in next section, basic principle of FLL based on SR is simply interpreted and simulated. In Sects. 41.3–41.5, the studies are brought out serially, and the conclusion is given out finally.

W. Zhang (✉) · Z. Li · S. Shi
Department of Space-Based Information System, ICE, PLAUST, Nanjing 210007, China
e-mail: 15605143455@163.com

W. Zhang
PLA Unit 66008, Tianjin 300171, China

H. Chen
Luoyang Electronic Equipment Test Center, Luoyang 471000, China

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_41

41.2 The Basic Principle of Frequency Locked Loop Based on Stochastic Resonance

41.2.1 The Structure of Frequency Locked Loop Based on Stochastic Resonance

The structure of frequency locked loop based on stochastic resonance consists of orthogonally demodulated module and frequency discriminator and loop filter and numerically controlled oscillator (NCO) and SR processor, just as shown in Fig. 41.1.

The SR processor, which is designed by the nonlinear effect of SR, will weaken the noise influence. The complete workflow of the improved FLL is as follows:

1. Adjusting the NCO and the FLL will start to work after the signal acquisition is completed.
2. The orthogonal demodulated signal is processed by the SR processor.
3. After processing of the frequency discriminator and loop filter, the NCO moves close to the signal frequency.
4. Repeat step 1–3 until the local frequency is gotten.

41.2.2 Simulation of Frequency Locked Loop Based on Stochastic Resonance

In this part, a computer simulation platform based on MATLAB is carried out in order to check the performance of the improved FLL. Generally speaking, traditional FLL could get stable and precise tracking with the signal whose SNR is $E_b/N_0 \geq 5$ dB. The signal is assumed as sinusoidal periodic signal buried with

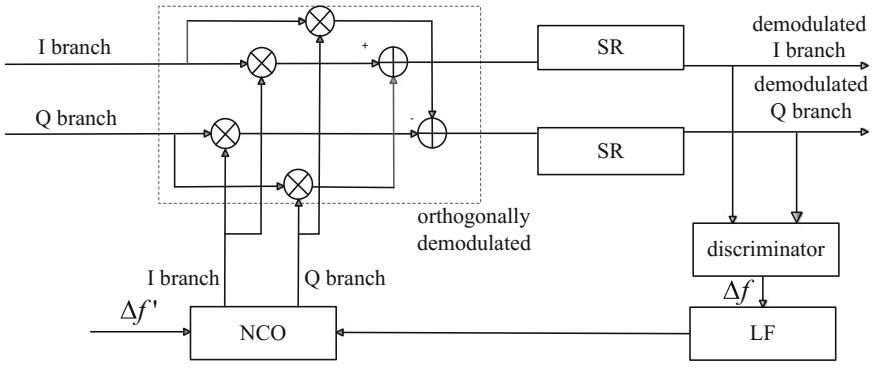


Fig. 41.1 The structure of frequency locked loop based on stochastic resonance

additive Gaussian white noise, whose frequency is no more than 20 Hz, and the signal to noise ratio is E_b/N_0 = 4 dB. A small loop bandwidth and a short coherent integration time is adopted in this paper. The random change of signal and the influence of Doppler will not be considered. All in all, the simulation platform is set as follows:

Signal to noise ratio: $E_b/N_0 = 4\,\text{dB}$
Sampling frequency: $f_s = N \times f_c$
Coherent integration time: $T = 5\,\text{ms}$
Loop bandwidth: $B_L = 2\,\text{Hz}$
Sampling frequency multiple: $N = 500$

A mass of researches indicate that the parameter a is less effective to the system performance. Set a = 1, and stimulate the nonlinear system to reach stochastic resonance state by adjusting the parameter b. We firstly set b = 1,000,000.

Figure 41.2 shows the signal is affected by the noise and its waveform jitters severely. The content in Fig. 41.3 is the signal processed by SR processor with the parameter a = 1 and b = 1,000,000.

The frequency tracking performance of traditional FLL and improved FLL is shown in the following Figs. 41.4 and 41.5.

Figure 41.4 shows the traditional FLL cannot realize stable and accurate tracking with the SNR is E_b/N_0 = 4 dB. From Fig. 41.5, we can see that the improved FLL can realize stable tracking with the signal processed by the SR processor and the frequency offset can reach 0.5 Hz. This result verifies the significant effectiveness that the SR processor brings to FLL. A great number of simulation indicate that different signal to noise ratio, sampling frequency and system parameters would influence the tracking performance of the improved FLL. Then, some researches will be carried out about it.

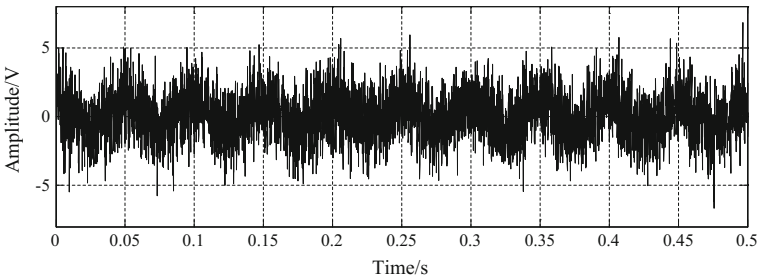


Fig. 41.2 The waveform of signal whose SNR is E_b/N_0 = 4 dB

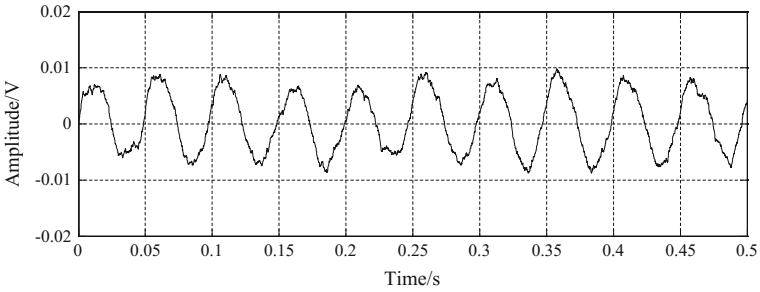


Fig. 41.3 The waveform of processed signal

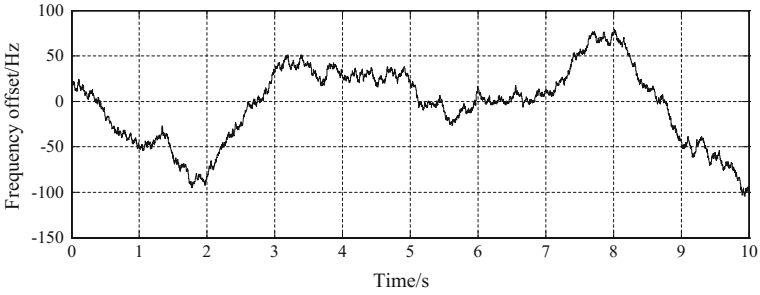


Fig. 41.4 The frequency tracking performance of traditional FLL

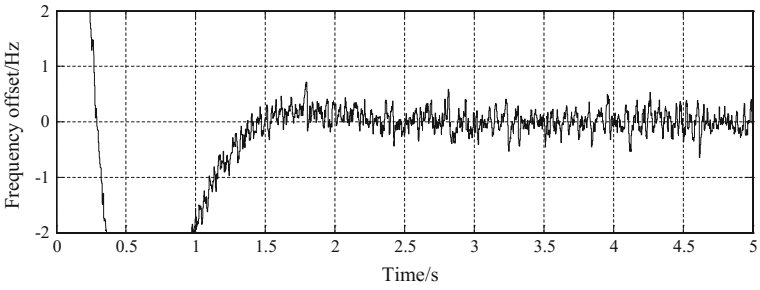


Fig. 41.5 The frequency tracking performance of improved FLL

41.3 The Influence of Signal to Noise Ratio on the Frequency Locked Loop Based on Stochastic Resonance

Obviously, the noise has a negative effect to the FLL. Smaller the noise is, better the performance of FLL gets. Part 1.2 shows that traditional FLL could get continuous and stable tracking with the SNR is $E_b/N_0 \geq 5$ dB while the improved FLL could with the SNR is $E_b/N_0 = 4$ dB. In order to explore the limit SNR that the improved FLL could bear, following researches have been pursued.

Based on the simulation platform, the SNR value is changed gradually. Set $E_b/N_0 = 3$ dB and $E_b/N_0 = 1$ dB respectively, and the processed signal is shown in the following Figs. 41.6 and 41.7.

Compared with Fig. 41.3, the cycle of waveforms in Figs. 41.6 and 41.7 is unchanged with the SNR declining. At the same time, the waveform changes from smoothly to roughly, and the quantity of the burrs begins to increase.

The frequency tracking performance is shown in the following Figs. 41.8 and 41.9.

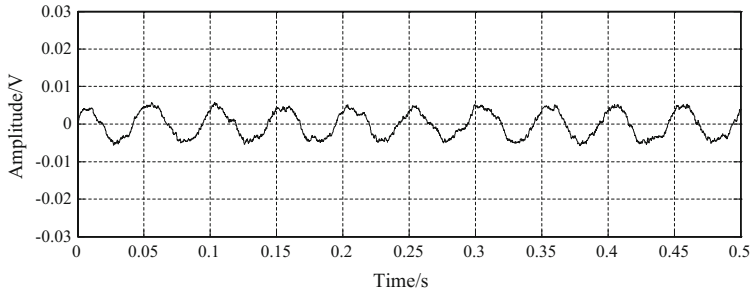


Fig. 41.6 The waveform of processed signal whose SNR is $E_b/N_0 = 3$ dB

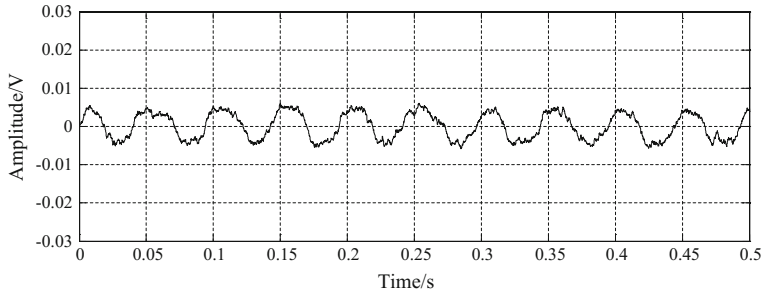


Fig. 41.7 The waveform of processed signal whose SNR is $E_b/N_0 = 1$ dB

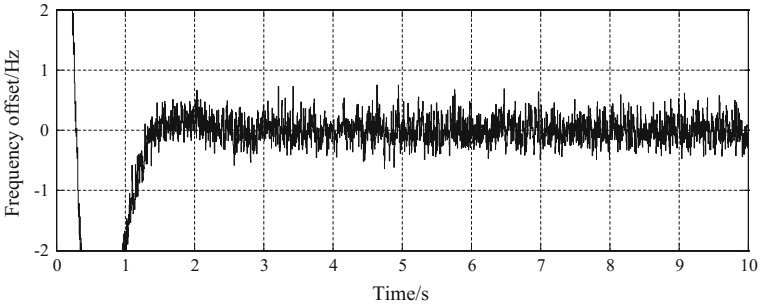


Fig. 41.8 The frequency tracking performance of improved FLL with E_b/N_0 = 3 dB

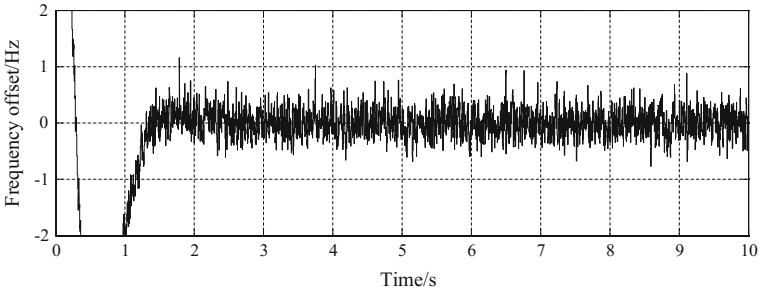


Fig. 41.9 The frequency tracking performance of improved FLL with E_b/N_0 = 1 dB

When the SNR is E_b/N_0 = 3 dB, the frequency offset of improved FLL is about 0.6 Hz, and the average frequency offset is 0.11 Hz. When the SNR is E_b/N_0 = 1 dB, the frequency offset of improved FLL is about 0.9 Hz, and the average frequency offset is 0.15 Hz. With decreasing of SNR, the frequency tracking performance of improved FLL becomes worse. It indicates that the additional noise has a negative effect on both the stochastic resonance and the frequency tracking performance of FLL.

At the same time, with more appropriate system parameters being set, the frequency offset of the improved FLL will change slightly, just as shown in Figs. 41.10 and 41.11.

A mass of simulations show that the adjustment of system parameters can not promote the improved FLL realizing continuous and stable tracking if the SNR is E_b/N_0 <1 dB. So the lowest SNR is E_b/N_0 = 1 dB that the improved FLL can bear.

All in all, the main influence of SNR on the improved FLL is:

1. The FLL based on SR could bear the SNR is E_b/N_0 = 1 dB, whose performance has been improved about 4 dB than traditional FLL.

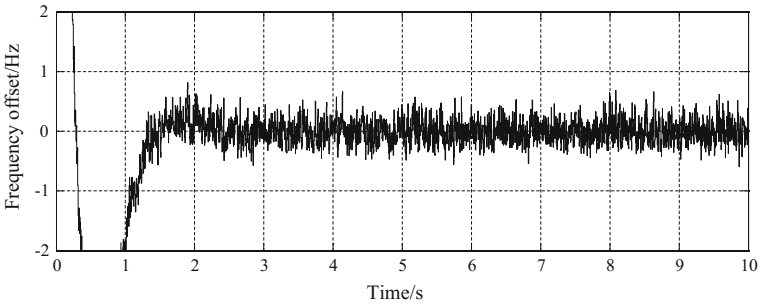


Fig. 41.10 The frequency tracking performance of improved FLL with E_b/N_0 = 3 dB

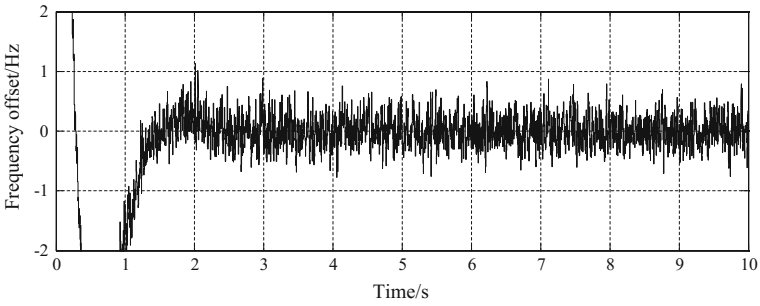


Fig. 41.11 The frequency tracking performance of improved FLL with E_b/N_0 = 1 dB

2. Under the condition of low SNR, the system parameters are the key factors leading to stochastic resonance. The noise, which plays a negative role, could be ignored.
3. When the SNR changes in a reasonable range, the system parameters can be adjusted to compensate for the negative effects of the noise, and the frequency tracking performance are kept unchanged.

41.4 The Influence of Sampling Frequency on the Frequency Locked Loop Based on Stochastic Resonance

Sampling frequency is an important factor which is related to SR. Higher the sampling frequency is, smaller the calculated step is, and better the resonance is.

Based on the simulation platform, the sampling frequency is changed gradually Set sampling frequency N = 100 and N = 1000, respectively, and the processed signal is shown in the following Figs. 41.12 and 41.13.

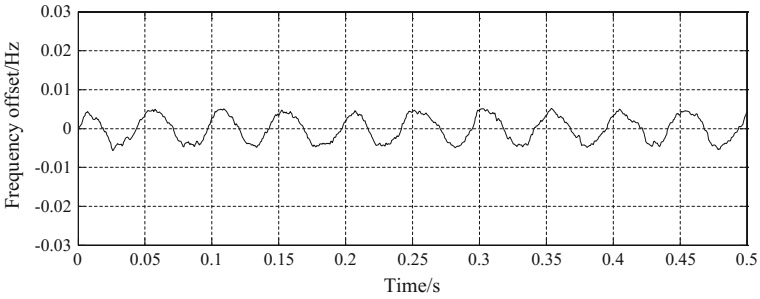


Fig. 41.12 The waveform of processed signal, $N = 100$

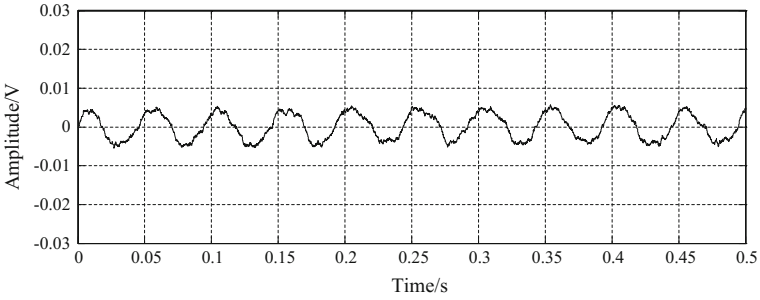


Fig. 41.13 The waveform of processed signal, $N = 1000$

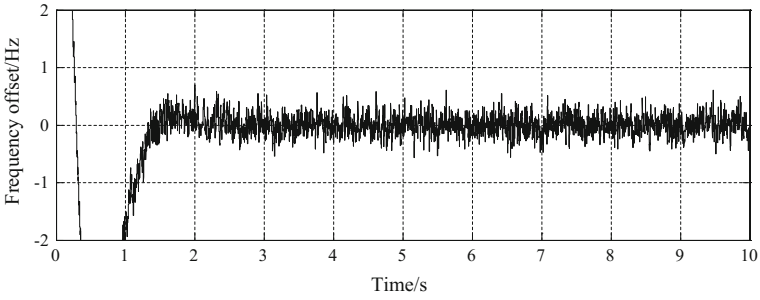


Fig. 41.14 The performance of improved FLL with $E_b/N_0 = 4$ dB, $N = 100$

Obviously, the increase of sampling frequency has little effect on the processing of the signal. The frequency tracking performance is shown in the following Figs. 41.14 and 41.15. When sampling multiple $N = 100$, the frequency offset is about 0.4 Hz, and the average frequency offset is about 0.112 Hz. When sampling multiple $N = 1000$, the frequency offset is about 0.4 Hz, and the average frequency offset is about 0.111 Hz.

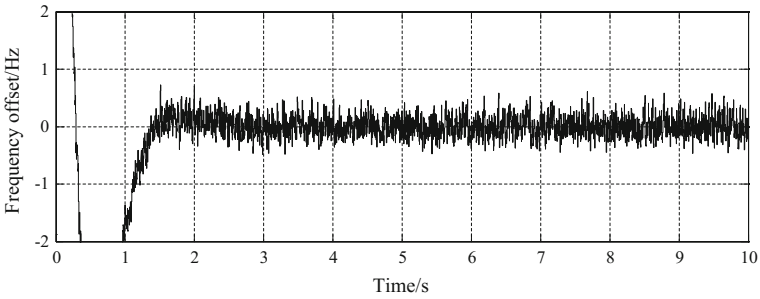


Fig. 41.15 The performance of improved FLL with E_b/N_0 = 4 dB, N = 1000

Table 41.1 The average frequency offset changes with different sampling multiple

Sampling multiple N	Average tracking frequency offset/Hz	Sampling multiple N	Average tracking frequency offset/Hz
100	0.112	600	0.113
200	0.111	700	0.115
300	0.112	800	0.112
400	0.115	900	0.111
500	0.114	1000	0.114

A mass of simulation shows that the improved FLL will realize continuous and stable tracking if the sampling multiple $N \geq 50$, and the frequency offset could reach 0.5 Hz. Table 41.1 shows the average frequency offset changes with sampling multiple. Clearly, there is no significant relationship between them.

All in all, the main effect of sampling frequency on the improved FLL is:

1. The sampling frequency must be set 50 times higher than the frequency of signal.
2. When the SNR (E_b/N_0) is kept constant, the noise intensity superimposed on the signal will change with the sampling multiple. High sampling frequency will make the bandwidth of the noise become larger, and the noise intensity will become larger. With the increasing of sampling frequency, the tracking performance of the improved FLL changes little. This is mainly because there is a relationship between the noise intensity superimposed on the signal and the sampling frequency, which is "this eliminate, that rise". When small sampling frequency is captured, the noise intensity superimposed on the signal is light, and the resonance is normal. By contrast, the noise intensity superimposed on the signal is large, and the intensive calculation guaranteed the resonance is normal similarly.
3. When the SNR (E_b/N_0) is kept constant, the sampling frequency has no limit, as long as it satisfies the law of Nyquist. With the decreasing of sampling frequency, the randomness of noise enhances, which carries out a negative effect on the improved FLL.

41.5 The Influence of System Parameters on the Frequency Locked Loop Based on Stochastic Resonance

The coordination of SR processor and signal and noise is the crux of stochastic resonance. System parameters plays an important role in SR processor. Based on the simulation platform, different system parameters are selected to explore the effect it brings out to the improved FLL.

Set $b = 10^6$ and $b = 10^9$, respectively. The processed signal is shown in Figs. 41.16 and 41.17. Compared with Fig. 41.2, the periodicity of the waveform can be identified clearly, which indicates the stochastic resonance has occurred. The waveform in Fig. 41.16 is relatively smooth and more similar to the triangular wave, whose amplitude is about 0.009 V. The waveform in Fig. 41.17 is more similar to the sinusoid wave, whose amplitude is about 0.001 V.

The frequency tracking performance of improved FLL under different system parameters is shown in Figs. 41.18 and 41.19. Obviously, different system parameters lead to different tracking results. When $b = 10^6$, the frequency offset can reach 0.5 Hz. When $b = 10^9$, the frequency offset can reach 0.8 Hz. If the parameter $b < 10^6$ or $b > 10^9$, it is very difficult to realize continuous and stable tracking for the improved FLL.

All in all, the main influence of system parameters can be concluded as:

1. System parameters that can stimulate stochastic resonance is not a particular value, but present as an alternative numerical interval.
2. The amplitude of the signal processed by stochastic resonance becomes smaller.
3. Differentsystem parameters leads to different resonance results. When the system parameters is small, the waveform is more likely triangular. By contrast, the waveform is more likely sinusoidal. Under the conditions of improved FLL, a small system parameter could obtain a better tracking accuracy.

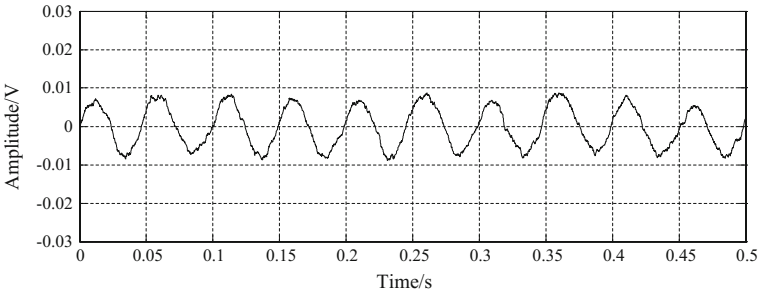


Fig. 41.16 The waveform of processed signal, $a = 1$, $b = 10^6$

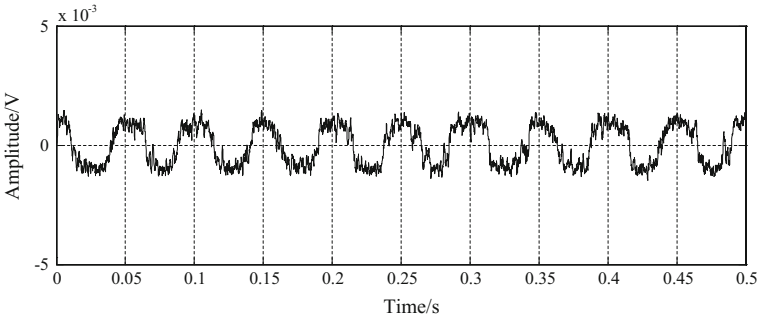


Fig. 41.17 The waveform of processed signal, $a = 1$, $b = 10^9$

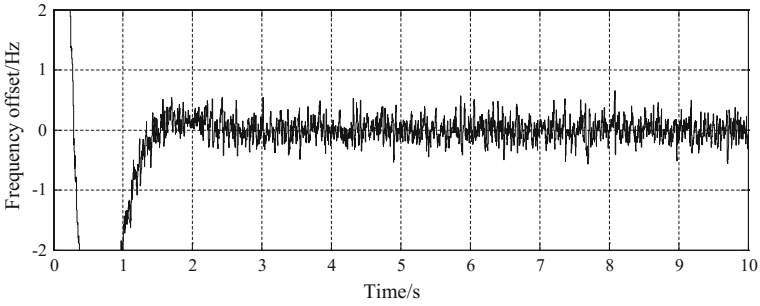


Fig. 41.18 The performance of improved FLL with $a = 1$, $b = 10^6$

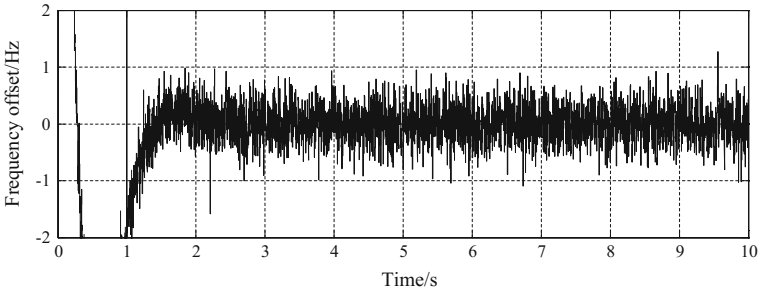


Fig. 41.19 The performance of improved FLL with $a = 1$, $b = 10^9$

41.6 Conclusions

The influence factors of the frequency locked loop based on stochastic resonance are explored in this paper. Simulation results show that signal to noise ratio and sampling frequency and system parameter are the three key factors that must be

deliberated before designing and using the improved FLL. The tracking performance gets poor with the SNR declining and the limit of SNR is $E_b/N_0 = 1$ dB. The sampling frequency must be set 50 times higher than the signal frequency, and the change of sampling frequency has little effect on the performance improvement. System parameters are the key factors that influence the FLL based on stochastic resonance, and it presents as an alternative numerical interval.

Acknowledgements This work is supported by the National Natural Science Foundation of China (No. 61571464).

References

1. Tian Y, Wu CQ, Niu XZ (2009) Research of carrier synchronization for low SNR and high dynamic signals. Electron Meas Technol 31(5):40–43
2. Cheng NP, Ren YF, Lu QF (2003) Study of carrier tracking for high-dynamic spread spectrum signals. Acta Electron Sin 31(12A):2147–2150
3. Zhang BC, Zhang QS (2003) research on the key issue about high dynamic receiver. Acta Electron Sin 31(12):1844–1846
4. Han MF, Wang YQ, Wu SL, Cui W (2009) A study on carrier tracking FLL for extremely-low-CNR highly-dynamic signals. Mod Radar 31(1):76–80
5. Benzi R, Parisi G, Sutera A (1983) A theory of stochastic resonance in climate change. SIAM J Appl Math 43(3):565–578
6. Duan FB (2002) The application of parameter-turning stochastic resonance in digital signal transmission. Doctoral Thesis of Zhejiang University
7. Zhu WN (2014) Study on parameters optimization of stochastic resonance and its application. Master Thesis of China Jiliang University
8. Zhang LB (2010) To detect high-frequency signals with stochastic resonance by improving sampling frequency. J Xiangfan Univ 31(2):39–41
9. Zhang G, Hu T, Zhang TQ (2015) Large parameters weak periodic signal detection based on stochastic resonance. Sci Technol Eng 15(35):189–191
10. Zhang WT, Li ZQ, Lou DK, Wang CF, Feng CC (2013) Improved carrier synchronization FLL based on stochastic resonance. J Milit Commun Technol 34(2):23–29

Chapter 42
Frequency Stabilization of an Optoelectronic Oscillator Based on Phase-Locked-Loop

Rongrong Fu, Yanhong Zhu and Xiaofeng Jin

42.1 Introduction

High performance microwave oscillators are the foundation of modern microwave applications, which is critical for the sensitivity of communication, radar, sensing and measurement systems. Modern communication features large bandwidth and high frequency, which impose high spectral purity and stability of microwave oscillators. The optoelectronic oscillator (OEO) is proposed by Yao and Maleki at 1994 [1], which is capable of high spectral purity microwave signal generation even at frequencies up to 100 GHz. The OEO utilizes optical fibers or other microwave photonic filters with high quality (Q) as a low loss energy storage component to achieve ultra-low phase noise output at such high frequencies. However, most free-running OEOs suffer from inferior frequency stability since optical fibers or high Q components are highly sensitive to the temperature and vibration.

A few exploration has been carried out to obtain frequency stabilization of the OEO in recent years. Injection locking (IL) proves to be viable for phase noise reduction of the oscillators [2, 3], while it requires an external microwave reference with high stability and low phase noise. Self-injection-locking (SIL) is a simpler scheme to suppress the phase noise of the OEO [4, 5], within which part of the oscillation is exported into a long delay line before injected back into the oscillating loop, where the overall phase noise reduction is proportional to the delay time. In spite of this, the long-term frequency stability of the OEO cannot be ensured since the drift of the environmental sensitive elements still exists [6]. Phase-locked loop (PLL) is recently applied to OEO to enhance the long-term stability of the OEO and reduce its close-in phase noise [7–9]. However, the equivalent phase noise of the reference after frequency multiplication in the PLL is of no superiority from

R. Fu (✉) · Y. Zhu · X. Jin
Department of Information Science and Electronic Engineering,
Zhejiang University, Hangzhou 310027, China
e-mail: jinxf00@zju.edu.cn

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_42

frequency offsets of several kilohertz or tens of kilohertz. Therefore, the frequency range of phase noise suppression is limited since the loop bandwidth of the PLL is generally chosen at this crossing offset.

In this paper, frequency stabilization of an X-band OEO using the technique of PLL and self-injection-locking (SIL) is proposed. The employment of SIL suppresses the phase noise of OEO within frequency offsets from 10 Hz to 10 kHz. The long term frequency stabilization is implemented by phase-locking the SIL OEO to a 100 MHz oven-controlled crystal oscillator (OCXO) via a fractional PLL. The employment of fractional PLL enables easier locking process and tunable output with high frequency resolution.

42.2 Principle

42.2.1 OEO System Setup

The schematic of the OEO system configuration is displayed in Fig. 42.1. The laser outputs a 1550 nm light of 20 mW power with a relative intensity noise (RIN) of −150 dB/Hz. After electro-optical modulation in a quadrature biased dual output Mach-Zehnder modulator (DOMZM), the light is split into two portions, where 50% of the optical power is distributed to the oscillating loop and the other 50% to the self injection loop. The oscillating loop of the OEO contains a fiber of 120 m, corresponding to a free spectral range (FSR) of 1.67 MHz. The self-injection loop is with a 10 km fiber. The optical signal is then converted into microwave signal by the photodiode (PD). In the electrical feedback of the oscillating loop 60 dB of gain are inserted while in self-injection link 35 dB of gain is employed to keep the open-loop gain below unity so that no oscillation can be sustained. The combined signal is filtered by a band pass filter (BPF) with 3 dB bandwidth of 10 MHz and center frequency of 9.95 GHz to select a proper oscillating mode. The following electrical phase shifter (EPS) provides a 2π phase shift at bias voltage of 2 V, utilized to tune the frequency of the SIL OEO through the feedback of the PLL.

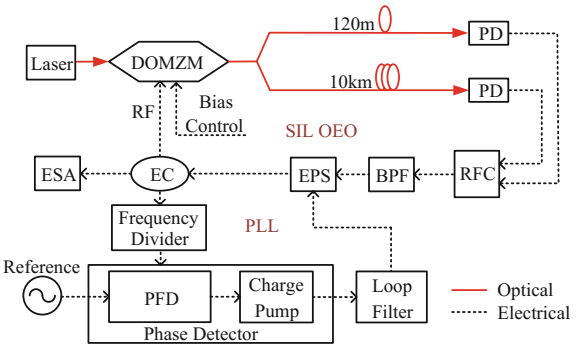


Fig. 42.1 Schematic of the SIL and PLL combined OEO

The SIL OEO as a whole operates as a voltage-controlled-oscillator (VCO), which is frequency divided to 100 MHz then phase detected with a 100 MHz OCXO. The phase error signal is low-pass filtered and fed back to the EPS to balance out the phase drift of the SIL OEO and suppress the close-in phase noise within the loop bandwidth.

42.2.2 The Mechanism of OEO Frequency Stabilization by PLL

The PLL's close loop transfer function $H(s)$ and error transfer function $H_e(s)$ can be expressed by [10]

$$H(s) = \frac{A_d \cdot A_F(s) \cdot k_v/N}{s + A_d \cdot A_F(s) \cdot k_v/N}, \tag{42.1}$$

$$H_e(s) = \frac{s}{s + A_d \cdot A_F(s) \cdot k_v/N}, \tag{42.2}$$

respectively, where A_d is the gain of the phase detector, $A_F(s)$ is the transfer function of the loop filter, k_v is the gain of VCO and N is the frequency divide value of the VCO. It can be inferred from Eqs. (42.1) and (42.2) that the PLL is high-pass to the phase argument of the VCO and low-pass to the reference phase argument. The phase noise of the 100 MHz OCXO reference will be about 40 dB higher under the equivalent frequency multiplication in the PLL since an N-times frequency multiplication corresponds to a phase noise degradation of $20\log_{10}N$ dB. The phase noise of the PLL's output at any given offset frequency f within the loop bandwidth is determined by that of the reference at the given offset frequency f. Therefore, the bandwidth of the loop filter should be at the crossing offset frequency where the phase noise of the OEO and the reference frequency-multiplied by N are equal to avoid far-out performance deterioration.

42.3 Results

The single side band (SSB) phase noise of the free-running OEO and the 10 km SIL OEO is discussed in Fig. 42.2, measured by the phase noise measurement module in an R&S FSW67 Electrical Spectrum Analyzer (ESA). A reliable and repeatable phase noise measurement of the free-running single loop OEO within 100 Hz offset is unavailable through the ESA due to the poor instantaneous stability of the OEO. When the SIL loop is closed, the phase noise up to 10 kHz offset is suppressed for over 20 dB, which accords with the theoretical relation between the

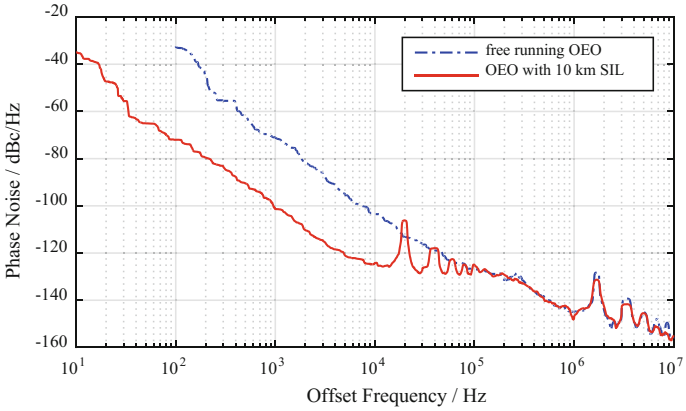


Fig. 42.2 SSB phase noise spectrum of the free-running OEO and the SIL OEO

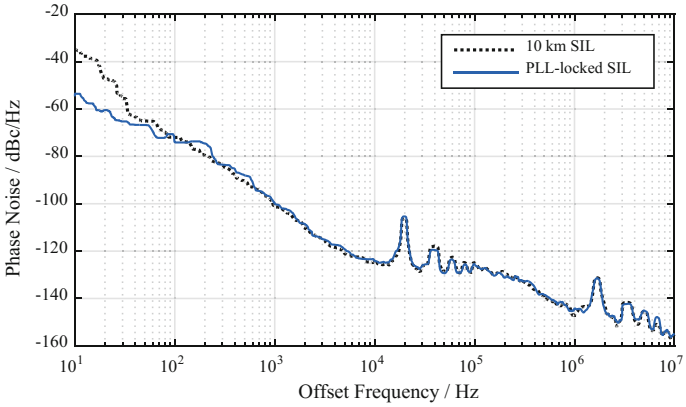


Fig. 42.3 SSB phase noise spectrum of the SIL OEO and PLL-locked SIL OEO

phase noise reduction and the delay time of SIL. However, spurious modes are induced by the SIL, which can be mutually suppressed by employing dual loop SIL [11, 12].

The output of the SIL OEO is fractional frequency divided to 100 MHz then phase detected with the 100 MHz OCXO. The phase error signal is low-pass filtered and fed back to the voltage control port to lock the PLL. The phase noise of the SIL OEO and the PLL-locked SIL OEO is displayed in Fig. 42.3. A close-loop bandwidth of 200 Hz is selected, where further phase noise suppression is seen within the loop bandwidth and beyond that the phase noise remains the same with that of the SIL OEO. The obtained phase noise of the PLL-locked SIL OEO is −55, −74 and −124 dBc/Hz at 10 Hz, 100 Hz and 10 kHz offset. It is decreased by 42 dB at 100 Hz offset and 21 dB at 10 kHz offset compared to the free-running single loop OEO with loop length of 120 m.

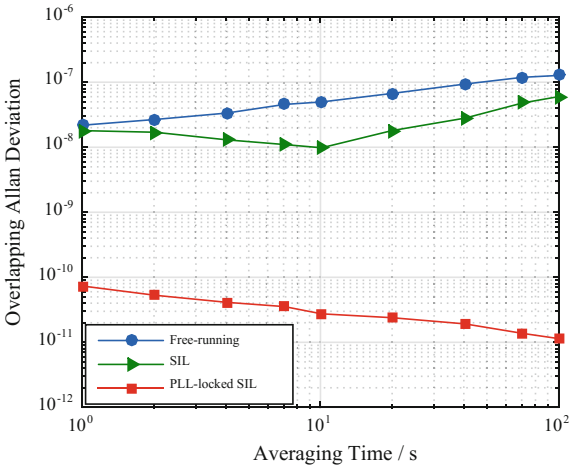


Fig. 42.4 Frequency overlapping ADEV of the free-running OEO, SIL OEO and the PLL-locked SIL OEO

A comparison of frequency overlapping Allan deviation (ADEV) of the PLL-locked SIL, the SIL and the free-running OEO is presented in Fig. 42.4. The frequency data is collected by the microwave frequency counter in the ESA. The SIL OEO represents slightly improved frequency stability than the free-running OEO, though both of them still show frequency drift with the increasing of average time. The SIL cannot eliminate the environment induced disturbance essentially in a lack of a highly stable reference. While the overlapping ADEV of the frequency of PLL-locked SIL OEO behaves convergence, reaching 1.14×10^{-11} at 100 s, which is more than 3 orders of magnitude better than that of the SIL OEO and the free-running OEO.

42.4 Conclusion

We demonstrate frequency stabilization in an OEO through the combination of fractional PLL and SIL. The influence of loop transfer characteristic of the PLL to the phase noise of the PLL's output is analyzed. A 9.95 GHz microwave output is obtained with phase noise of −55 dBc/Hz at 10 Hz and −124 dBc/Hz at 10 kHz offset and the overlapping ADEV is 1.14×10^{-11} at average time of 100 s. The employment of fractional PLL puts more flexibility on the frequency of both the reference and the microwave output's frequency, which enables the stabilization and practical utilization of the OEO with the help of the highly performance OCXO.

References

1. Yao XS, Maleki L (1994) High frequency optical subcarrier generator. Electron Lett 30(18):1525–1526
2. Zhang X, Zhou X, Daryoush AS (1992) A theoretical and experimental study of the noise behavior of subharmonically injection locked local oscillators. IEEE Trans Microw Theory Tech 40(5):895–902
3. Zhenghua Z, Chun Y, Zhewei C et al (2016) An ultra-low phase noise and highly stable optoelectronic oscillator utilizing IL-PLL. IEEE Photonics Technol Lett 28(4):516–519
4. Chang HC (2003) Phase noise in self-injection-locked oscillators-theory and experiment. IEEE Trans Microw Theory Tech 51(9):1994–1999
5. Wang TP, Tsai ZM, Sun KJ et al (2007) Phase-noise reduction of-band push–push oscillator with second-harmonic self-injection techniques. IEEE Trans Microw Theory Tech 55(1):66–77
6. Zhang L, Poddar AK, Rohde UL et al (2013) Self-injection locked phase-locked loop OEO In: IEEE MTT-S international microwave and RF conference. pp 1–4
7. Zhang Y, Hou D, Zhao J (2014) Long-term frequency stabilization of an optoelectronic oscillator using phase-locked loop. J Lightwave Technol 32(13):2408–2414
8. Jia S, Yu J, Wang J et al (2015) A novel optoelectronic oscillator based on wavelength multiplexing. Photonics Technol Lett 27(2):213–216
9. Bluestone A, Spencer DT, Srinivasan S et al (2015) An ultra-low phase-noise 20-GHz PLL utilizing an optoelectronic voltage-controlled oscillator. IEEE Trans Microw Theory Tech 63(3):1046–1052
10. Ronald E (2003) Phase-lock loops: design, simulation and application, 5th edn. McGraw Hill, New York, USA
11. Zhu Y, Jin X, Jin X et al (2016) A novel scheme of microwave generation based on heterodyne phase locking of an OEO. Photonics Technol Lett 28(22):2637
12. Zhang L, Poddar AK, Rohde UL et al (2013) Analytical and experimental evaluation of SSB phase noise reduction in self-injection locked oscillators using optical delay loops. Photonics J 5(6):6602217

Chapter 43
The Techniques of Network Coding Applied in the Physical-Layer of the Wireless Communication Systems: A Survey

Xiaoting Wang, Qiang Mei and Xu Yao

43.1 Introduction

Network coding [1, 2] is an innovative multicast transmission technique which allows a node to encode its received data before passing it on. Different from the traditional storage-and-forward way, the intermediate routing nodes encode the data and then forward. Through the encoding scheme the data rate of each node can achieve the max-flow bound. The original intention of the network coding technique is to improve the system transmission data rate or the network throughput. But furthermore, the advantages extend to intensify the network robust ability, promote the system security and save the energy consumption.

This survey focuses on discussing the techniques of network coding applied in the physical-layer of the wireless communication systems for two domains: the PNC and the joint design of the network and the channel coding. Firstly, the main research contents and directions of PNC are concluded on the basis of analyzing the PNC mechanism and studying the current situation and the developing trend. Then, different designs of the network and channel coding are categorized and summed up.

43.2 Network Coding Applications

Figure 43.1, which is commonly known as the butterfly network, depicts a communication network where two data bits b_1 and b_2 are multicast from the source node S to both the nodes R_1 and R_2. The vertices correspond to terminals and the edges correspond to channels. The data rate of each channel is assumed to 1 bit per time unit. According to the max-flow min-cut theorem of the network information

X. Wang (✉) · Q. Mei · X. Yao
Beijing Institute of Tracking and Telecommunications Technologies, Beijing 100094, China
e-mail: wxt.xiti@gmail.com

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_43

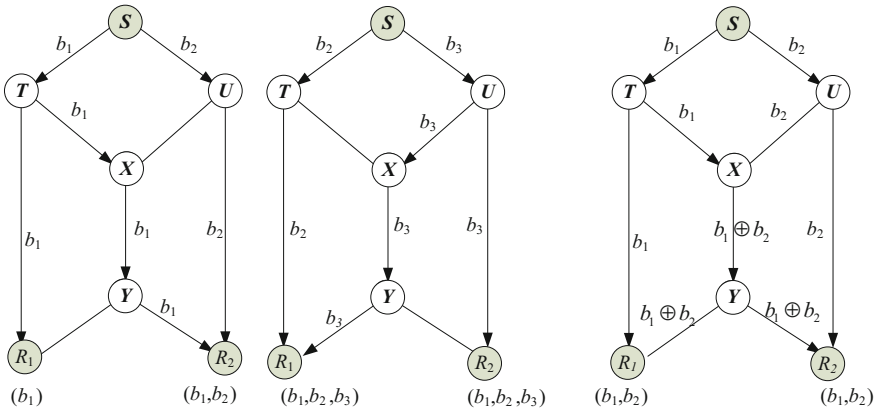


Fig. 43.1 Network coding in the butterfly network

theory, the maximum information rate we can send from S to R_1 and R_2 is equal to the min-cut value, that is, 2 bits per time unit. In the traditional network transmission, as the Fig. 43.1a depicted, S sends b_1, b_2 and b_3 (3 bits) to R_1 and R_2 in 2 time units. This achieves a transmission efficiency equals 1.5. The channel XY, which has unit capacity, becomes the bottleneck of the butterfly network where the intermediate node X has to transmit one bit per time unit. While using network coding, as the Fig. 43.1b depicted, the node X derives from the received bits b_1 and b_2 the exclusive-OR bit $b_1 \oplus b_2$ and then sends it. Through the replicated transmission of node Y, the nodes R_1 receives b_1 and $b_1 \oplus b_2$, from which the bit b_2 can be decoded. Similarly, the node R_2 decodes the bit b_1 from the received bits b_2 and $b_1 \oplus b_2$. As we can see, it achieves transmission efficiency equals 2. Actually, by applying the network coding, the capacity of the network reaches the maximum flow limit while it cannot be accomplished where the intermediate nodes perform just bit replication.

Network coding is an important breakthrough in the information transmission domain. It stems from the wired network and now focuses more on the wireless network. The wireless channel is characteristic as time varying and the downlink transmission is a multicast model, both of which make the network coding technique be prone to combine with the wireless communication. As the Fig. 43.2 shows, the application techniques of network coding in the wireless communication systems can be divided roughly depending on the scenario, the layer and the goal of the applications.

The objectives of the network coding applications are different depending on the scenario and structure of the network. Generally, the wireless communication network can be divided into the single-hop and the multiple-hop networks. In the traditional cellular mobile network and the cognitive radio network which belong to

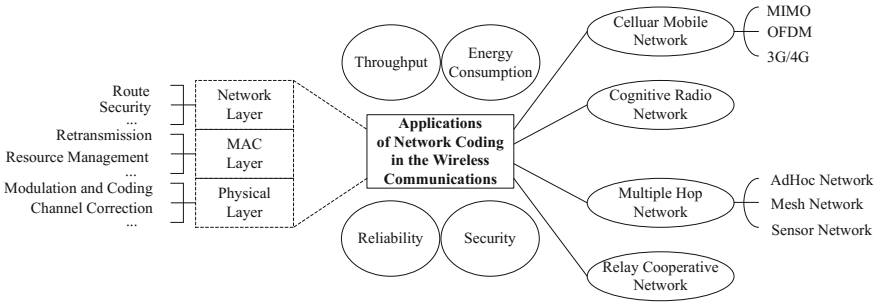


Fig. 43.2 The diagram of applications of network coding in the wireless communication systems

the single hop networks, the objectives of the network coding applications are to improve the system throughput, enhance the transmission reliability and strengthen the system security. While in the wireless Ad Hoc network, the wireless mesh network and the wireless sensor network which belong to the multiple-hop networks, the objectives of the network coding applications are highlighted on reducing energy consumption of the nodes and enhancing the transmission reliability. The relay cooperative network which also belongs to the multiple-hop networks recently become the research hotspot of the network coding application in wireless communication systems. But it is usually combined with the cellular mobile network and the scenarios of which are different from the mesh network. So the application objectives are also throughput, reliability and security.

There are diverse development directions for the network coding applications in each layer of the wireless communication network. In the physical layer, the network coding technique is combined with various transmission techniques such as modulation and coding, channel error correction, multiple input multiple output (MIMO) technique and etc. Specially, the physical-layer network coding (PNC) is an important network coding application technique in the physical layer, based on which lots of technique combinations spring up. In the media access control (MAC) layer, the network coding is more likely to combine with the retransmission, the scheduling, the resource management and etc. In the network layer, in which the network coding presents firstly, the combination is with routing techniques naturally and on the other fact the secure network coding is a newly developing research content.

43.3 Physical-Layer Network Coding (PNC)

43.3.1 The PNC Scheme

PNC technique was firstly proposed by Zhang S. and etc. in 2006 [3]. The main idea of PNC is directly superposing the electromagnetic (EM) wave signals in the physical layer of a relay node through a designed proper

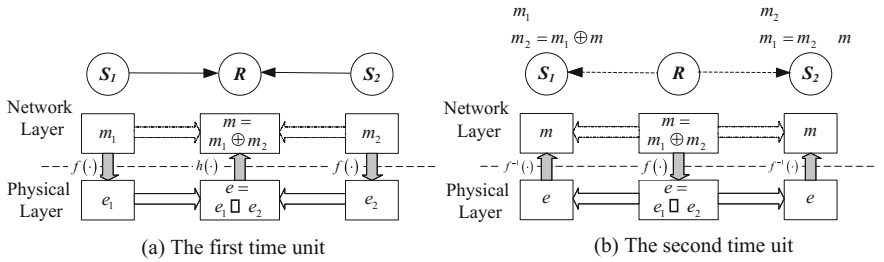


Fig. 43.3 The PNC scheme

modulation-and-demodulation mechanism. The superposition of EM wave signals can be mapped to digital bit streams in GF(2^n), which seem equally as the XOR operation in traditional network coding.

Taking QPSK modulation for example, it can be considered as two BPSK data streams: an in-phase stream and a quadrature-phase stream. We can handle with the two streams in the same way. As depicted in the Fig. 43.3, a three node two-way transmission linear network scenario is considered, m_i denotes the bit data sent by two source nodes S_1 and S_2. After the corresponding modulation, the value domain of the electromagnetic wave signals is $e_i \in \{-1, 1\}$. The modulation mapping function is $f(\cdot) : e_i = 2m_i - 1$ for QPSK modulation and the EM wave signal is transmitted through the physical layer of each source node. Assuming symbol-level and carrier phase synchronization, the physical layer of the relay node receives the superposition of the EM wave from two source nodes and the amplitude value domain is $(e_1 + e_2) \in \{-2, 0, 2\}$. A many-to-one mapping function $h(\cdot)$ can be designed in order to map the superposition EM wave signals in the physical node to the bit data value in the network layer of the relay node. Making the values of the function equal to the XOR values of the bit data if applying network coding directly, thus the scheme actually realizes the network coding in the physical layer.

For QPSK modulation, if $m_1 \neq m_2$, thus $m = m_1 \oplus m_2 = -1$, and if $m_1 = m_2$, thus $m = m_1 \oplus m_2 = 1$. As a result, the mapping function $h(\cdot)$ can be designed as follows,

$$m = \begin{cases} -1, e_i = 0 \\ 1, e_i = -2, 2 \end{cases} \tag{43.1}$$

The relay node modulates and transmits the mapped bit data, then the nodes S_1 and S_2 receive and demodulate thus both can get coded bit data $m = m_1 \oplus m_2$. According to the signal m_1 and m_2 already known, through the XOR operation, i.e. $m_2 = m_1 \oplus m$ and $m_1 = m_2 \oplus m$, S_1 and S_2 ultimately get mutual information.

Through processing the signal superposition in the physical layer directly, a two-way relay transmission only needs 2 time units using PNC. Compared with 4 time units in the traditional relay system and 3 time units in the traditional network coding system, PNC can yield a 2-time and 1.5-time throughput improvement respectively.

43.3.2 The Research Directions

The PNC combines the coding in the network with the physical layer techniques, such as the modulation mapping and the antenna selection. In utilizing of the wireless channel characteristics, it is more likely to improve the system throughput and enhance the system robustness. The paper [4] analyzes the problems that have to be considered in applying the network coding in the real communication system, including the noise, the synchronization, the channel fading and etc., which are of course the key research contents of the PNC. Only these problems are solved that could the PNC theory step forward to application [5]. Besides, the theoretical researches are still the hot topics as the guide of this innovative conception in the transmission techniques. The research contents are mostly focused on the following aspects.

1. The coding and decoding design

As the PNC scheme shows, the coding and decoding process of PNC is actually the mapping process. For certain modulation mode and given the mapping function $f(\cdot)$, a mapping function $h(\cdot)$ has to be found in order to demodulate the superposed physical signal to get the network coding form of the bit data. The coding and decoding design of PNC lies in the design of the mapping function pair $(f \cdot h)$. The merits of the functions decide the performance of the PNC scheme. The PNC schemes can be categorized as the physical network coding in finite field (PNCF) and in infinite field (PNCI) according to the value domain of the network coding function. Firstly, the PNCF was analyzed assumed in AWGN channel and QPSK modulation. Thereafter more complex modulation modes and channel conditions such as FSK, MPSK and QAM [6–8] in the Rayleigh channel, Nakagami channel and etc. [9, 10] are considered. The representative technique of PNCI is analog network coding (ANC) [11] which outputs analog signal through function mapping. Instead of requiring strict synchronization, it accepts the signal collision interference and makes use of the asynchronization to improve the system throughput. But compared with PNCF, PNCI also amplifies the additive noise in the relay node hence the system performance will deteriorate if the uplink channel conditions grow bad.

2. The synchronization problem

The PNC is presented based on the assumption of fully transmission synchronization at first. However, there exists symbol asynchronization, phase asychronization and etc. and certainly will impact the on the PNC performance. Many researches follow up to analyze how much the effect will be and propose various PNC schemes to cope with the asychronization [12–15]. The study of paper [12] and [13] show that the performance loss of the carrier phase shift, the carrier frequency shift and the symbol shift is <3, <0.6 and <2.2 dB respectively. The paper [16] presents the simulation results in QPSK modulation mode which show that the loss will arrive at 6–7 dB at 10^{-2} bit error rate, in the condition that the phase difference equals $\pi/4$ while the symbols are fully synchronized. But the

paper [4] believes that the loss is only about 1 dB if considering the symbol shift. It is on account of the diversity and the transmission certainty effect because of which the symbol shift on the contrary compensates on the loss produced by the phase shift. It demonstrates that in actual system the transmission performance may not deteriorate so much owing to the asychronization as in the limiting case. In spite, it needs further step research that how much of symbol shift could decrease the loss of the performance. Besides, the channel estimation in the PNC is also a research direction [17, 18].

3. The combination with other transmission techniques

PNC realizes the networking coding technique in the physical layer, thus naturally could combine with the transmission techniques of the physical layer to exploit the function of the network coding. The joint design of PNC and channel coding is an aspect more researches focus on [19–21]. Besides, PNC is combined with MIMO-OFDM. The asynchronized PNC can be constructed utilizing the advantages of OFDM and multiple antennas [22, 23]. It is worth to mention that PNC could combine with secure transmission techniques. By making use of the physical layer characteristics of the wireless channel (such as the fading, the noise and etc.), the quality of the main channel need to be guaranteed better than that of the wiretap channel so that the eavesdropper is unable to acquire any useful information. It integrates the physical layer characteristics, the secure transmission and the network coding effectively and becomes an innovative research direction [24, 25].

4. The PNC capacity

The information theory study of PNC is aiming at deducing the channel capacity of the two-way relay channel, i.e. the maximal information data rate of the two source nodes in the AWGN channel condition. The capacity of the TWRC is seemed as a bound and the capacity of different PNC scheme is compared with that, such as the PNC capacity with channel coding, multiple antennas [26, 27], or with ANC [11, 27]. Specially, because the PNC technique is used for security transmission as well, there comes up the theory researches from the angle of improving the security capacity [28, 29].

5. The application scenario extensions

The two-way relay channel is the main application scenario in the PNC researches. Moreover, the scenarios with multiple users, multiple relays and multiple hops are becoming important contents of PNC researches [30, 31].

43.4 Joint Design of the Network Coding and the Channel Coding

In the wireless network, the most popular scenario of the network coding application is the relay system owing to the mechanism of the network coding. Uniting the relay node taking decode and forward mode with the network coding, obviously could increase the diversity gain and the capacity. And on the other hand, the combination of network coding and channel coding is considered. The redundant information of the relay node which passes on to the destination node can give better error protection and effectively fight against the influence of fading and noise in the channel and hence improve the system reliability.

There are two kinds of methods to combine the network coding and the channel coding. One is the simple combination where each technique is independent. The paper [32] proves that when the channel is discrete memory less and independent in the single source direction network, the separation theorem of the network coding and the channel coding is to be true. The network coding helps reaching the maximal flow bound and the channel coding guarantees the transmission reliability of every link. Considering the complexity of the system and the convenience of the application, the two techniques could be designed separately. As the Fig. 43.4 shows, the flow is as follows:

1. In the source nodes, broadcast the information S_1 and S_2 to the relay node and the base station through channel coding and modulation;
2. In the relay node, get the estimations $\bar{S}_1$ and $\bar{S}_2$ of the source data through demodulation and channel decoding, XOR the estimated data to realize network coding, process the channel coding and modulation again, send them to the base station;
3. In the base station, demodulate and channel decoding the three way data from the two source nodes and the relay node, process the network decoding later, get the final estimations of the data $\hat{S}_1$ and $\hat{S}_2$.

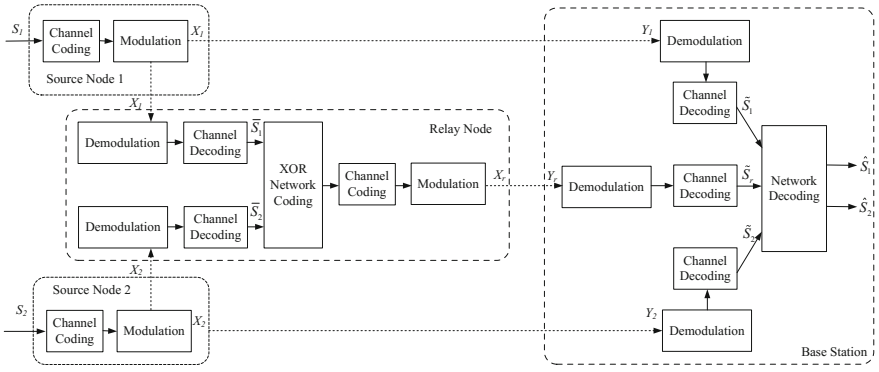


Fig. 43.4 The simple combination of the network coding and the channel coding

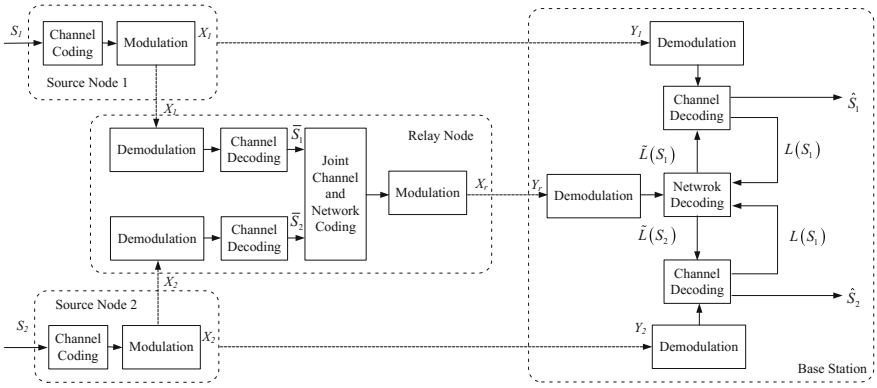


Fig. 43.5 The joint design of the network coding and the channel coding

The other combination method of the network coding and the channel coding is joint design codec. The paper [33] points out that adopting joint coding could get better transmission reliability than separate coding in certain multi-user channel. As the Fig. 43.5 shows, the flow is as follows:

1. In the source nodes, broadcast the information S_1 and S_2 to the relay node and the base station through channel coding and modulation;
2. In the relay node, get the estimations $\bar{S}_1$ and $\bar{S}_2$ of the source data through demodulation and channel decoding, process the channel coding and modulation according to the designed joint network-channel coding scheme which takes the channel coding features into account, send them to the base station;
3. In the base station, receive the data from the two source nodes and the relay node. As the data sent by the relay node is after the joint network-channel coding, there exists parity information of the two source data to assist decoding and the soft information $\tilde{L}(S_1)$ and $\tilde{L}(S_2)$ can be got. The soft information will go into a new round channel decoding process and after several iterations the two channel decoders give hard decisions, thus the final estimations of the data $\hat{S}_1$ and $\hat{S}_2$ will be obtained.

As the researches of the network coding going deeper, the combination of which with the channel coding generally are from the joint design thought. Depending on different angle of the design, the joint designs can be divided into the following categories.

1. Based on the type of the channel coding

The Turbo code and the LDPC code have outstanding error control performance that could nearly approach the Shannon limit bound. The researches focus on techniques based on these two types and so as the network coding. For the Turbo code, the relay node unites the Turbo code and XOR network coding and sends the parity information. The base station makes the joint iterative decoding. For the

LDPC code, the coding principles are usually exploited such as the product code, the Tanner graph and etc. Joint coding and decoding schemes can be realized considering different generator matrix design of the source node information to institute of XOR network coding and optimizing the corresponding LDPC code [34–37].

2. Modify the mode of handling the soft information

The hard decision error of the relay node will lead to whole system performances decrease, so we can use the soft information to design the codec algorithm from the angle of enhancing the signal strength. The relay node does not do the hard decision and instead transmits the soft information after the network coding [38]. If the system complexity is considered, the relay node could still adopt hard decision while adding certain compensation in the soft information of the base station to help the channel and network decoding [39].

3. Utilizing of PNC

The PNC is able to realize network coding through the signal superposition and basing on different modulation and mapping modes. It makes the bit operation on the network layer come true on the physical layer, so the channel coding as an important physical layer technique is apt to jointly design with it. The relay node does not need to do coding and decoding thus can simplify the system complexity greatly and raise the transmission efficiency at the same time [19–21].

4. Considering various system scenarios

For the joint design, the system scenario of two source and single relay is gradually turned into that of multiple relays and multiple user accesses.

Compared with the simple combination design in connecting separated part, the joint network and channel coding starts from the channel coding principles and utilizes the soft information process. Taking the advantages of the PNC, it can acquire improvement on the aspects of the transmission efficiency and the system reliability. Moreover, the joint design shows large superiority on error rate performance especially in good signal to noise conditions.

43.5 Conclusion

The survey studies the application techniques of the network coding in the physical layer of the wireless communication systems. It aims at the hot topics: PNC and the joint design of the network and channel coding, in which the technique application schemes are discussed and the developing directions are presented on the basis of analyzing the technique principles. The researches of the network coding technique are spreading from the physical layer, the MAC layer and the network layer. All kinds of advanced techniques can be combined with it. And the consistent goal is promoting the transition from the theory study to the practical use.

References

1. Ahlswede R, Cai N, Li SY et al (2000) Network information flow. IEEE Trans Inf Theory 46 (4):1204–1206
2. Li SY, Yeung RW, Cai N (2003) Linear network coding. IEEE Trans Inf Theory 49(2):371–381
3. Zhang S, Liew SC, Lam PP (2006) Hot topic: physical-layer network coding. In: Proc. of ACM MOBICOM'06, Los Angeles, USA, pp 358–365
4. Zhou QF, Zhang S et al (2014) Wireless network coding. Posts & Telecom Press, Beijing
5. Zhao MF, Zhou YW, Yuan Q, Yang YX (2011) Research survey on physical layer network coding. J Comput Appl 31(8):2015–2020
6. Wang J (2009) Study on the theory of network coding and its application. Doctoral thesis of the Xidian University
7. Ruohan C, Tiejun lv, Feichi L (2011) Symbol-based physical-layer network coding with MPSK Modulation. In: Proc. Of IEEE GLOBECOM'11, Houston, USA, 263(4):1–5
8. Liu B (2013) The modulation techniques of the physical-layer network coding. Electron Technol Quality Eng 2013(6):9–11
9. MinChul J, Min K (2010) Error performance analysis of bpsk modulation in physical layer network-coded bidirectional relay networks. IEEE Trans Commun 58(10):2770–2775
10. Li B (2013) Research on Physical-layer Networkcoding and the Performance in Asymmetric Two-way Relay Communications. Doctoral thesis of the Harbin Institute of Technology
11. Katti S, Gollakota S, Katabi D (2007) Embracing wireless interference: analog network coding. In: Proc. of ACM SIGCOMM'07, Kyoto, Japan, 9(4):397–408
12. Zhang S, Liew SC, LAM PP (2006) On the synchronization of physical-layer network coding. in Proc. IEEE ITW'06, Punta del Este, Uruguay, pp 404–408
13. Zhang S, Liew SC, Wang H (2009) Synchronization analysis in physical layer network coding. In: Computer Science http://arxiv.org/abs/1001.0069
14. Zygmunt JH, Wang D, Fu S et al (2009) Channel coding design to support asynchronous physical layer network coding. In: Proc. IEEE GLOBECOM'09, Hawaii, USA, pp 1–6
15. Lu L, Liew SC (2012) Asynchronous physical-layer network coding. IEEE Trans Wirel Commun 11(2):819–831
16. Hao Y, Goeckel D, Ding Z et al (2007) Achievable rates for network coding on the exchange channel. In: Proc. of IEEE MILCOM'07, Orlando, USA pp 1–7
17. Jiang B, Gao G, Gao X (2010) Channel estimation and training design for two-way relay networks with power allocation. IEEE Trans Wirel Commun 57(10):2022–2032
18. Sjodin T, Gacacin G, Adachi F (2010) Two-slot channel estimation for analog network coding based on OFDM in a frequency-selective fading channel. In: Proc. IEEE VTC' Ottawa, Canada, pp 1–5
19. Wilson MP, Narayanan K, Pfister HD et al (2010) Joint physical layer coding and decoding for bidirectional relaying. IEEE Trans Inf Theory 56(11):5641–5654
20. Sui D (2010) Joint decoding scheme of physical layer network and channel coding. Master thesis of Beijing University of Posts and Telecommunications
21. To D, Choi J (2010) Convolutional codes in two-way relay networks with physical-layer network coding. IEEE Trans Wirel Commun 9(9):2724–2729
22. Zhang S, Liew SC (2010) Physical layer network coding with multiple antennas. In Proc. IEEE WCNC'10, Sydney, Australia, pp 1–6
23. Tanoli SAK, Khan I, Rajatheva N, Atachi F (2010) Advances in relay networks: performance and capacity analysis of space-time analog network coding. EURASIP J Wirel Commun Netw 2010(1):1–10
24. Dong L, Han Z, Petropulu A et al (2010) Improving wireless physical layer security via cooperating relays. IEEE Trans Signal Process 58(3):1875–1888
25. Yuan Q (2012) Research of wireless cooperative relay technology based on network coding. Doctor thesis of Beijing University of Posts and Telecommunications

26. Zhang S, Liew SC, Wang H et al (2009) Capacity of two-way relay channel. Berlin: Springer Press, pp 219–231
27. Zhang R, Liang Y, Chai CC (2009) Optimal beamforming for two-way multi-antenna relay channel with analogue network coding. IEEE Sel Areas Commun 27(5):699–712
28. Lai L, El-Gamal H (2007) Cooperative secrecy: the relay-eavesdropper channel. In: Proc. IEEE ISIT'07, Nice, France, pp 931–935
29. Tekin E, Yener A (2008) The general gaussian multiple access and two-way wiretap channels: achievable rates and cooperative jamming. IEEE Trans Inf Theory 54(6):2735–2751
30. Gao F, Cui T, Jiang B, et al (2009) On communication protocol and beamforming design for amplify-and-forward n-way relay networks. In: Proc. IEEE International Workshop CAMSAP, Aruba, pp 109–112
31. Amah AUT, Klein A (2010) Beamforming-based physical-layer network coding for nonregenerative multi-way relaying. EURASIP J Wirel Commun Netw 2010(1):1–12
32. Song L, Yeung RW, Cai N (2006) A separation theorem for single-source network coding. IEEE Trans Inf Theory 52(5):1861–1871
33. Slepian D, Wolf JK (1973) A coding theorem for multiple access channels with correlated sources. Bell Syst Techn J 52(7):1037–1076
34. Li ZY (2012) Research on key technology for joint channel and network coding. Doctor thesis of Beijing University of Posts and Telecommunications
35. Ji PY, Xie H, Zhao YL (2012) Joint design of network-channel coding on relay node. Comput Eng Appl 48(34):102–106
36. Gu Y (2012) Research on the network coding in the wireless two-way relay Systems. Master thesis of Fudan University
37. Razaghi P, Yu W (2007) Bilayer low-density parity check codes for decode-and-forward in relay channels. IEEE Trans Inf Theory 53(10):3723–3739
38. Song GH (2009) Joint design of the network coding and the channel coding in wireless relay networks. Master thesis of Xidian University
39. Wu K (2013) Joint design of network coding and channel coding in wireless relay network. Master thesis of Nanjing University of Posts and Telecommunications

Chapter 44
Research on the Improvement of LTP Protocol in Space DTN Network Based on Network Coding

Peng Wan, Shijie Song, Zhongjie Hua and Shengli Zhang

44.1 Introduction

The future space missions ask for more and more data transmission bandwidth along with the increase of space application services, and the spacecraft would be of powerful processing and storage capabilities with the development of the space technologies on board. It is obviously that the communication society would go into the times of the space-ground integrated networks. In the recent researches on the space Delay Tolerant Networks (DTN), in order to conquer the worse channel conditions, ACK [1], NACK [2], SACK [3] and other similar mechanisms in transport layer have been generally used to improve the reliability of data transmission tasks. However, because of the long time delay of the ACK channel and the large diversity of the random data loss rate, utilizing the feedback and retransmission methods would severely impair the transmission performance of the whole system. Reference [4] suggests the Deep Space-Transport Protocol (DS-TP), which performs well in the network throughput, but its simple retransmission of the sent messages might decrease the working efficiency of this method. Based on the DS-TP protocol, some researchers promoted the coded transport protocol [5], which just utilize the channel coding to produce and send the redundancy messages, in order to minimize the number of the feedback and retransmission cycles, but the chosen channel coding method of this protocol is so complicated and inflexible as to limit its application in space missions.

This article firstly introduces the characteristics of the space DTN networks and its communication protocol stacks, especially the designs and procedures of the

P. Wan (✉) · S. Song · Z. Hua
Beijing Institute of Tracking and Telecommunications Technology,
Beijing 100094, China
e-mail: wanpeng@bittt.cn

S. Zhang
School of Information Engineering, Shenzhen University, Shenzhen 518060, China

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_44

Licklider Transmission Protocol (LTP). Then, with the detail analysis of the shortcomings of LTP protocol used in the space DTN networks, we suggest an improvement of LTP protocol based on Network Coding (NC-LTP), and make efforts in the theoretical works and simulation of such new approach. The theoretical and simulation results show that, compared with the traditional LTP protocol, NC-LTP performs well in the different situations, such as time-delay and packets loss rate, which is suitable to be used in the space DTN networks with the characteristics of high dynamic, high bit-error rate and band-width limited.

44.2 Space DTN Networks

Space DTN networks [6] is a new information network model suitable to the space communications situations, which derived from one of the NASA sponsored project conducted by the united research group of NASA Jet Propulsion Laboratory (JPL) and the Maryland University. The effectiveness of the space DTN networks have been validated by several kinds of the space experiments, such as the Deep Impact Network experiment (employing the EPOXI space cruise), the JAXA jointly performed space link demonstrations with NASA (where the JAXA's GEO relay satellite called Data Relay Test Satellite has been used), the Space Data Routers European Project, and the pilot operation of a DTN implementation on the International Space Station (ISS).

Recently, DTN has been accepted by many international organizations such as Internet Engineering Task Force (IETF) and Consultative Committee for Space Data Systems (CCSDS). In order to adapt well to the extreme space communication environment, DTN mainly introduces two concepts: (1) insert a new middleware called Bundle Protocol (BP) [7] between the application layer and transport layer, in order to provide hop-by-hop acknowledgement; (2) supplement a new protocol called Licklider Transmission Protocol (LTP) [8] in the transport layer, in order to minimize the number of communication handshake between the end users.

Figure 44.1 shows the typical space DTN network communication protocol stacks, which is made of seven layers from the application layer to the physical layer.

1. Application Layer

As the top layer of the space DTN network protocol stacks, the application layer includes some application services in space, such as CCSDS File Delivery Protocol (CFDP) [9] and other test applications. CFDP is not the core part of the DTN protocols, yet it could provide some useful functions at the delay tolerant mode based on the BP, such as the sectoring, sending, receiving and reassigning of a file, so CFDP could be used as the application layer in this stacks.

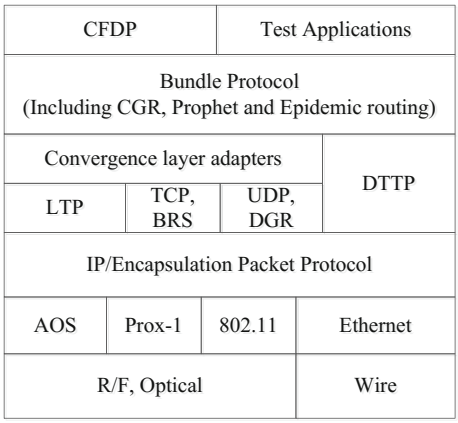


Fig. 44.1 Space DTN network communication protocol stacks

2. Bundle Protocol layer

In the second layer, BP works as the critical protocol for the whole DTN protocol stacks. If there is no complete end-to-end routes in the DTN networks, BP would provide the data forwarding service with the characteristic of delay tolerant, as well as the dynamic routing. The widely studied routing algorithms include Contac Graph Routing (CGR) [10], Probabilistic Routing Protocol using History of Encounters and Transitivity (PRoPHET) [11], Epidemic [12], Spray and Wait [13], and so on.

3. Convergence and Transport Layer

The third and fourth layer of DTN protocol stacks are respectively the Convergence and Transport Layer. The Convergence Layer provides the interface and translation service between the Bundle Layer and the Transport Layer, so the most of its function should be determined by the underneath layer. Recently, the typical Convergence Layer protocols include LTP, BRS, TCPTL, and so on, generally derived from the traditional TCP and UDP protocols. In the actual application, in order to improve the facility and implement, we always utilize the standard and universe convergence layer interface, allowing the different service functions to be transferred to the next transport layer. The typical Transport Layer protocols in space DTN networks include TCP, UDP, DTTP, LTP, and so on, follows as:

- TCP [14] and UDP [14] function as the counterparts in terrestrial Internet.
- DTTP [15] designs for the deep space communication scene, as a Delay-Tolerant transport protocol, which could perform well and reliably in the challenged network through the method pointing to packets' transmissions, and could also get the required network bandwidth based on the rate-concerned transport behaviors.
- Similar to UDP protocol, LTP is a point to point communication protocol, which could be not care about the issues such as routing and congestion control.

Furthermore, LTP tries to solve the problems such as delay and disturb in the point to point communication situation, especially the long time delay transmission difficulty.

4. Network, Data Link and Physical Layer

The left three lower layer in the DTN protocol stacks are Network, Data Link and Physical Layer, which are not the core parts in this system. The DTN nodes could choose different protocols in such three layers based on the different communication situations, and such lower protocols would not introduce impacts or other requirements for the upper layers.

Concretely, we could choose not only the terrestrial IP protocol, but also the Encapsulation Service [16] from the CCSDS standards, which is originally designed for the space networks. With wired interface in Physical layer, we could choose Ethernet 802.3 [14] as the Data Link layer; however, with the wireless interface in Physical layer, such as R/F or optical communication condition, we could choose either the terrestrial 802.11 [17], or the CCSDS AOS [18] or Proximity-1 [19] protocols.

44.3 Analysis of LTP

LTP is named for the momentary for the American computing scientist Joseph Carl Robnett Licklider, who was one of the most famous experts in the area of computer science and normal calculation. Originally, the purpose to design a new LTP protocol focused on the issues on the interplanetary long-haul transmission, acting as a point- to-point communication method in the deep space wireless links. LTP usually works on the up layer of the Data Link Layer, however, it could also be used in the circumstance such as terrestrial mobile internet, VPN without band-limit, or software developing, even cooperated with the UDP protocol. Because of the point-to-point trait, LTP might not care about the issues such as routing and congestion control.

LTP is generally considered as the standard convergence layer protocol of the bundle protocol, widely used in various kinds of networks. Some important characteristics of LTP describe as follows:

- No data loss on the disturb-tolerant links
- Aim on the least payload or the low capacity on asymmetry links
- Expand on the segments (e.g., secure coding)
- Cancel the transmission or reception procedure of some blocks
- Accelerated retransmission: utilizing multiple check points mechanism for each block
- Partial reliability: utilizing check point and retransmission mechanism for the first N bytes of each block.

Concretely, LTP could not only provide the reliable transmission for the essential data (e.g., the header of a file), but also provide the unreliable transmission for the non-essential data (e.g., the pixel of an image). According to the potential long time delay circumstance, in order to avoid the lower efficiency of the wireless link utilization, LTP supports the message transfer without negotiation. The time counter and communication procedure table should cooperate with each other, meanwhile, in case the wireless links break down anytime, the time counter should be stopped. Additionally, LTP should be noticed with the availability, round trip time, and communication procedure status of the Data Link Layer. Furthermore, LTP could avoid the long RTT communication service by the one-way conversation mechanism.

Figure 44.2 shows the typical LTP communication procedures.

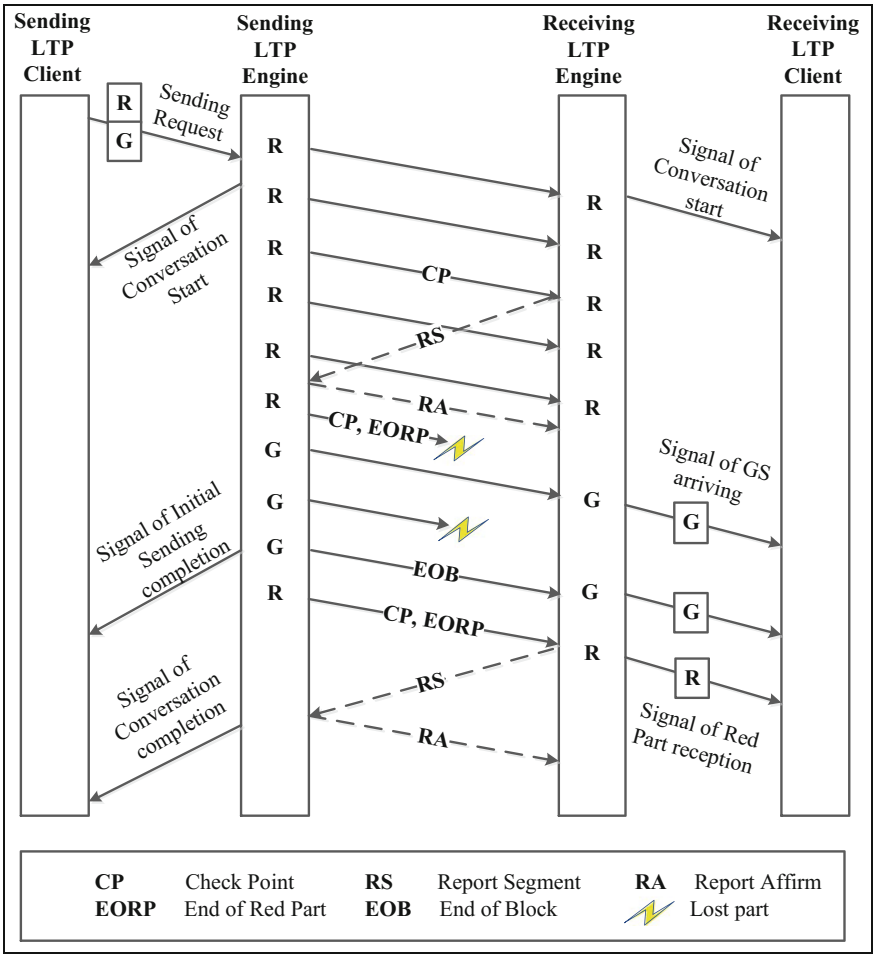


Fig. 44.2 Typical LTP communication procedures

At the source node, LTP usually saves the data segments in its protocol engine, which would check whether there were some available wireless links accessible to the sink node: if exists, then the LTP engine would start the transmission; if not, then the LTP engine would not delete or discard such data segments until it receives the Link State Cue information from either the lower Data Link layer or Management Information database (MIB). If the Red-parts are transferred, it would produce the mark of End of Red Part (EORP) for the last segment, meanwhile, it would start the time counter for this transmission task. If there is no Green-parts at the end of this transmission procedure, it would produce the mark of End of Block (EOB); otherwise, it would transfer the Green-parts and produce the mark of End of Block (EOB) for the last segment. Once the sink node receives the End of Red Part (EORP), then it would reply a report data segment to the source node, finishing the data transmission session. After the transmission of data segments are finished, both the Red-parts and Green-parts, the source node would start a time counter. Once received the report data segment, the source node would cancel such time counter, rearrange the acknowledgements and tell the application layer about the successful transmission of all the Red-parts. Once the sink node receives the End of Block (EOB), then the whole transmission task is finished.

There is one special characteristics of LTP protocol, which divided all the blocks to be transferred into 2 parts as Red Parts and Green Parts. The Red Parts require the acknowledgement and retransmission mechanisms, which means that the red part data block must be transferred in a reliable way, and they should maintain available before being acknowledged. Take the header as an example, if there exist some important information for the decoding consequently, then the loss of such header would bring harmful impact for the processing of the subsequent data blocks. The Green Parts do not require reliable transmission, which could be dropped just after the source sending without any acknowledgements. The data length of either Red parts or Green parts could be zero respectively, namely for any data block, it could include all of the Red parts or Green parts. If the length of Green parts is zero, and all of the data are Red parts, then LTP would provide the transmission function similar to TCP; otherwise, it works similar to UDP. LTP receives the acknowledgement with selectivity. In order to guarantee the reception capacity for any data segments, LTP should operate normally all the time. With the LTP protocol, the most important data segments which have not been correctly received by the sink node and should be transferred again could be saved by the LTP engine, accounting for the discontinuous links problem in the space DTN networks.

44.4 Improvement of LTP Based on Network Coding

44.4.1 Design Methodology

In this paper, we suggest an improvement of LTP based on Network Coding [20], NC-LTP, with low complexity and suitable to the space DTN networks, which could effectively enhance the throughput and decrease the number of retransmission.

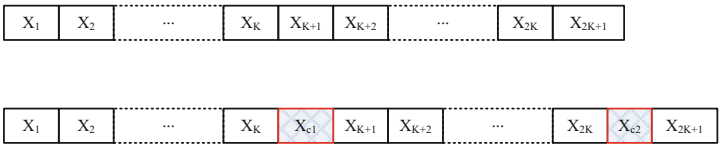
In the NC-LTP, we could produce and send each coded packet every K blocks, which is encoded by the XOR calculation of the former K blocks. Even if anyone of such $K + 1$ blocks was lost, the sink could still recover original K blocks from the remaining received blocks. However, if the number of lost blocks are no less than 2, then we need to retransmission the additional blocks. With the low complexity and the automatic adjustment for the value of K and the number of redundancy messages, so we could effectively improve the performance for the space DTN networks.

The characteristics of NC-LTP lies on the ideas as follows: After all the Red-parts' segments have been sent, then the source DTN node would encode all the sent segments into a new Network Coded segment, which acts as the EOPR to be sent to the sink DTN node. Such new segment could recover all the former segments if anyone were lost, which could obviously decrease the whole communication time delay. LTP divides all the segments into 2 parts, the Red-parts must be transferred with acknowledgements, while the Red-parts could be transferred with the reliable methods. Although both the Red-parts and the Green-parts could operate on the NC-LTP protocol stacks, we prefer the application of NC-LTP only on the Red-parts. There would be both normal segments and control segments in the transmission procedure of the Red-parts, however, NC-LTP would only encode the normal segments, even if the control segments might include data to be transferred.

44.4.2 Design of NC-LTP

1. Operation at NC-LTP source node

The NC-LTP source node encode the data blocks through XOR method into a new coded data segment. The content of the data segment include: Service ID, Segment Offset, Segment Length, CP sequence, RS sequence, and the user data. For the Network coding data blocks, only the Segment Offset, Segment Length and user data need to be encoded. For the SDNV data blocks, NC-LTP would encode all the data but MSB through XOR. Assume the data segments to be sent marked as X_1,



Notes: the upper line shows the original data segments without coding; the bottom line shows the data segments after coding, with the new coded data segment *Xci* to be inserted.

Fig. 44.3 The encoding and sending procedure at the source node

X_2, …, X_K, X_{K+1}, …, X_{2K}, X_{2K+1}, …, then we could encode every K segments into X_i as follows:

$$\begin{aligned} X_{ci} &= X_{iK+1} \oplus X_{iK+2} \oplus \cdots \oplus X_{iK+K} \\ i &= 0, 1, \ldots \end{aligned} \tag{44.1}$$

In Eq. (44.1), the symbol $\oplus$ means the XOR process for two data segments as the neighbors to each other, and the calculation should be operated from left to right. The newly encoded data segment should be inserted after the proper data segment, and the other data segments need not to be changed. The encoding and sending procedure at the source node could be seen as Fig. 44.3.

2. Operation at NC-LTP sink node

After receiving the Network coded EORP segment, the sink node should firstly calculate the feasibility about whether all the data segments could be correctly received:

- If available, then the sink node will reply with the RS acknowledge segments for the correct transmission status, and wait for the RA data packets.
- Else if not available, then the sink node will utilize the Network coded EORP segment to decode all the segments, in order to get the data segment which had been lost.
- Else, the sink node will reply with the RS acknowledge segments for the successful received segments, and wait for the retransmission.

44.4.3 Analysis of LTP

Assume that, t_0: the transmission time of each data segment; t_1: the propagation time of signal; p: the data loss rate of each data segment.

1. Total transmission time delay based on traditional LTP

$$\begin{aligned} T_{LTP} &= (Kt_0 + 2t_1) + (Kpt_0 + 2t_1) + (Kp^2t_0 + 2t_1) + \cdots + (Kp^{n-1}t_0 + 2t_1) \\ &= 2nt_1 + Kt_0\frac{1-p^n}{1-p} \end{aligned} \tag{44.2}$$

2. Total transmission time delay based on NC-LTP

$$\begin{aligned} T_{NC-LTP} &= ((K+1)t_0 + 2t_1) + \{[(K+1)p-1]t_0 + 2t_1\} + \{[(K+1)p-1]pt_0 \\ &\quad + 2t_1\} + \cdots + \{[(K+1)p-1]p^{m-2}t_0 + 2t_1\} \\ &= 2mt_1 + (K+1)t_0 + [(K+1)p-1]t_0\frac{1-p^{m-1}}{1-p} \end{aligned} \tag{44.3}$$

3. Theoretical analysis

The theoretical analysis could be seen as Figs. 44.4 and 44.5, which show that NC-LTP always performs better than LTP, and the advantage of NC-LTP increases with the increase of error probability.

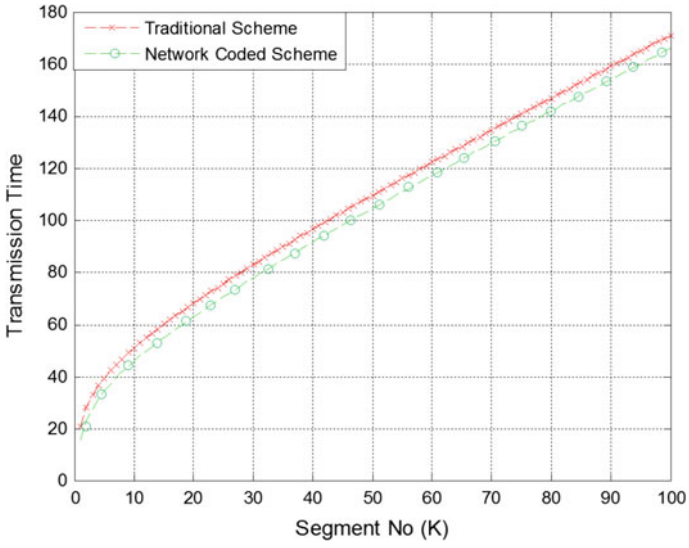


Fig. 44.4 Transmission time difference with different loss rate, segments loss rate $p = 0.1$

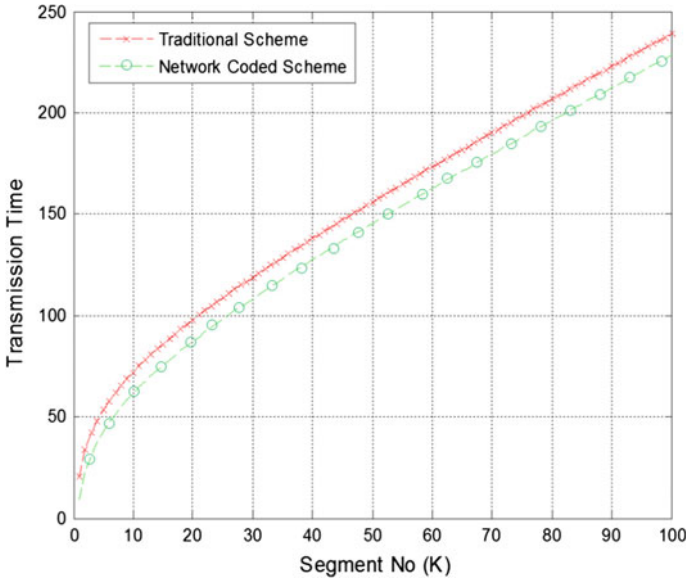


Fig. 44.5 Transmission time difference with different loss rate, segments loss rate $p = 0.3$

44.5 Simulation

44.5.1 Simulation Scene

Simulation scene of the deep space DTN networks is shown as Fig. 44.6.

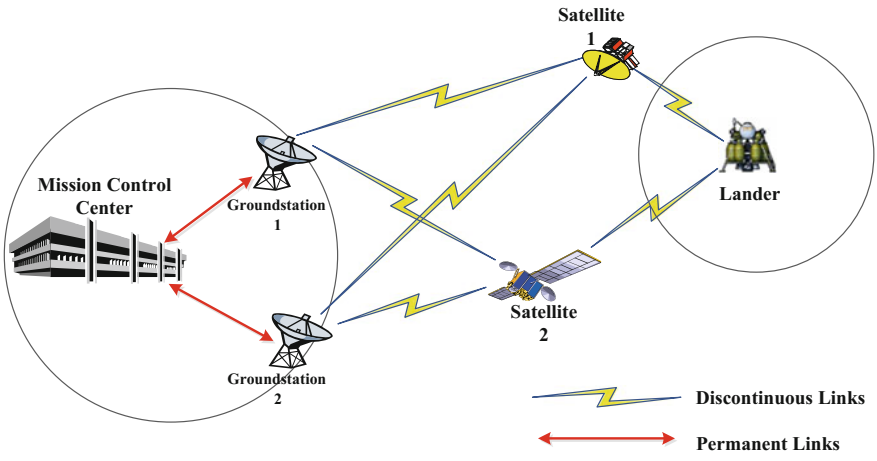


Fig. 44.6 Simulation scene of deep space DTN networks

The protocol stacks between the nodes in the networks include application layer (e.g., telecommand and telemetry), Bundle Protocol, transport layer (e.g., Licklider Transmission Protocol, LTP), network layer (e.g., IP, Encapsulation Service), data link layer (CCSDS AOS) and physical layer (RF links). Theoretically, the DTN satellites, both Satellite 1 and Satellite 2, should locate at the equilibrium points in the Earth and Mars system, such as the Lagrange points which refers to some stable points where one relatively small object can remain stationary under the gravitational effect of two bigger objects.

44.5.1.1 Simulation Results

1. Comparison between LTP and NC-LTP under different error probability

Figure 44.7 shows the comparison between LTP and NC-LTP under different error probability, which means that NC-LTP always performs better than LTP, and the advantage of NC-LTP increases with the increase of error probability. With the further analysis, we could conduct that when the data loss rate increases for the reason of wireless channel error rate or network congestion, the loss probability of data segments would be increased, and the advantage of Network Coding would be much more significant.

2. Comparison between the LTP and NC-LTP under different propagation time

Figure 44.8 shows the comparison between LTP and NC-LTP under different propagation time, which means that NC-LTP always performs better than LTP, and

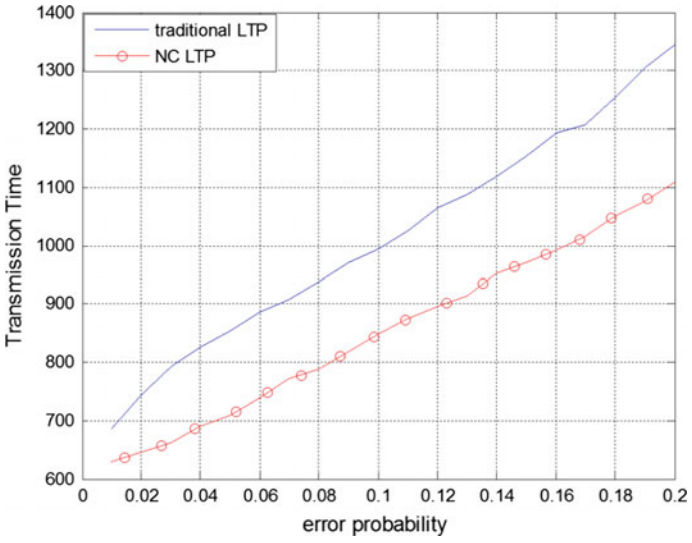


Fig. 44.7 Comparison between LTP and NC-LTP under different error probability

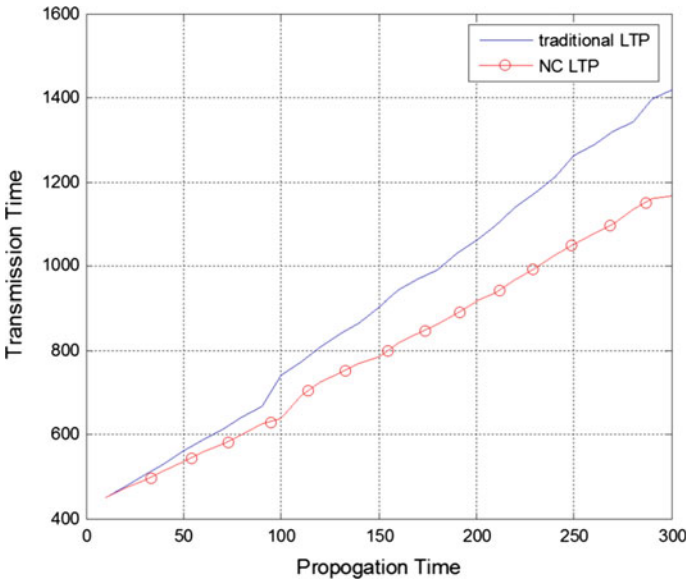


Fig. 44.8 Comparison between LTP and NC-LTP under different error probability

the advantage of NC-LTP increases with the increase of propagation time. With the further analysis, we could conduct that the advantages of NC-LTP is due to the fixed number of additional transferred encoded segments, and the propagation time it saves is proportional to the time delay.

44.6 Conclusions

This paper makes great effort on the study of existing LTP protocol in the space DTN networks, and suggest the improvement of LTP protocol based on Network Coding (NC-LTP). In order to verify the NC-LTP's performance, we have done some research works to prove its availability with the DTN networks simulation test-bed. The analysis and simulation results shows that, compared with the traditional LTP, our new NC-LTP performs well on under the different error probability and propagation time, which proves to be suitable for the highly dynamic space DTN networks.

References

1. Akan OB, Fang J, Akyildiz IF (2002) Performance of TCP protocols in deep space communication networks. IEEE Lett Commun 6:478–480
2. Akyildiz IF, Giacomo M, Sergio P (2001) TCP-Peach: a new congestion control scheme for satellite IP networks. IEEE/ACM Trans Netw 9:307–321
3. Mascolo S, Casetti C, Gerla M et al (2001) TCP westwood: Bandwidth estimation for enhanced transport over wireless links. In: Proceedings of the 7th annual international conference on Mobile computing and networking. ACM, pp 287–297
4. Psaras I, Papastergiou G, Tsaoussidis V et al (2008) DS-TP: Deep-space transport protocol. In: Aerospace Conference, 2008 IEEE, pp 1–13
5. Cola T, Ernst H, Marchese M (2007) Performance analysis of CCSDS File Delivery Protocol and erasure coding techniques in deep space environments. Elsevier Comput Netw 51(14):4032–4049
6. Fall K (2003) A delay-tolerant network architecture for challenged internets. In: Proceedings of the 2003 conference on applications, technologies, architectures, and protocols for computer communications. ACM, pp 27–34
7. CCSDS Bundle Protocol Protocol Specification (2015) Recommendation for space data system standards, CCSDS 734.2-B-1. Blue Book, Washington D. C.
8. Licklider Transmission Protocol (LTP) for CCSDS (2015) Recommendation for space data system standards, CCSDS 734.1-B-1. Blue Book, Washington D. C.
9. CCSDS File Delivery Protocol (2007) Recommendation for space data system standards, CCSDS 727.0-B-4. Blue Book, Washington D. C.
10. Birrane E, Burleigh S, Kasch N (2012) Analysis of the contact graph routing algorithm: bounding interplanetary paths. Acta Astronaut 75(1):108–119
11. Lindgren A, Doria A, Schelén O (2003) Probabilistic routing in intermittently connected networks. ACM SIGMOBILE Mobile Comput Commun Rev 7(3):19–20
12. Mundur P, Seligman M, Lee G (2008) Epidemic routing with immunity in delay tolerant networks[A]. In: Proceedings of IEEE MILCOM, pp 1–7
13. Spyropoulos T, Psounis K, Raghavendra CS (2005) Spray and wait: an efficient routing scheme for intermittently connected mobile networks. In: Proceeding of ACM SIGCOMM workshop on delay-tolerant Networking, pp 252–259
14. Stevens WR (1994) TCP/IP illustrated, vol 1. Addison-Wesley Reading, Upper Saddle River, pp 15–268
15. Samaras CV, Tsaoussidis V, Peccia N (2008) DTTP: a delay-tolerant transport protocol for space internetworks, 2nd ERCIM Workshop on eMobility. Tampere, Finland, pp 3–14
16. Encapsulation Service (2009) Recommendation for space data system standards, CCSDS 133.1-B-2. Blue Book, Washington D. C.
17. Matthew S Gast (2005) IEEE 802.11 wireless networks: the definitive guide, 2nd edn. O' Reily Media, Inc, 1005 Gravenstein Highway North, Sebastopol, CA 95472, The United States of America
18. AOS Space Data Link Protocol (2015) Recommendation for space data system standards, CCSDS 732.0-B-3. Blue Book, Washington D. C.
19. Proximity-1 Space Data Link Protocol (2006) Recommendation for space data system standards, CCSDS 211.0-B-4. Blue Book, Washington D. C.
20. Li SYR, Yeung RW, Cai N (2003) Linear network coding. IEEE Trans Inf Theory 49(2): 371–381

Chapter 45
Research on Security Protection of the Communication Network for Space TT&C Based on TCP/IP Protocol Vulnerabilities

Shuai Yuan, Peng Liu and En Zhao

45.1 Introduction

The communication network for Space TT&C is composed of a centre and a number of stations distributed around the world. The task of the communication network for Space TT&C is to guarantee the communication and transmission of data, image and scheduling information between the centre and the stations, and to realize the tracking, measurement and control of the spacecraft. Since the new generation of the communication network for Space TT&C was built, it has been safe and stable for three years. From the operation situation, the network is stable and reliable, and basically meets the communication needs of space TT&C system. With the development of China's space industry, the number of satellites has increased year by year, the scale of the communication network for Space TT&C is expanding, and the task is increasing day by day, the information transmission quality and safety requirements are also getting higher and higher.

At present, network attack and defense technology is developing rapidly, network attack measures emerge frequently and develop rapidly. The vulnerability of the TCP/IP protocol cluster in the network has become the first choice for hackers to network intrusion. TCP/IP protocol security has become a top issue in the current network security. How to effectively protect the network attacks based on TCP/IP protocol cluster vulnerabilities is a pressing matter of the moment.

In this paper, based on the analysis of the basic principle of the TCP/IP layered protocol and the protocol vulnerabilities, the security problems are analyzed. With all kinds of security problems which have been analyzed, the effective defense measures and protection methods are given. Finally, the design scheme of a set of configuration and maintenance management system for the communication network

S. Yuan (✉) · P. Liu · E. Zhao
Xi'an Satellite Control Center, Xi'an 710043, China
e-mail: 3984924@qq.com

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_45

for Space TT&C is discussed. The effective management of the safe and stable operation of the communication network for space TT&C is realized.

45.2 Vulnerability Analysis of TCP/IP Protocol and Common Attack Methods

TCP/IP protocol is a kind of network communication protocol, which regulates the data exchange format and transmission mode of communication devices on the network. TCP/IP protocol is the basis of the Internet protocol. At the beginning of design, the protocol does not take into account the current network has so many threats. So for the protocol vulnerabilities, there are many kinds of attacks [1].

45.2.1 TCP/IP Hierarchy

As shown in Table 45.1, TCP/IP system adopts a hierarchical structure, and the lower layer provides service to the upper layer. It can be seen from the table that the TCP/IP protocol defines the network layer, transport layer and application layer, and does not provide the physical layer and data link layer services. But TCP/IP supports all of the standard physical and data link protocols [2].

Because the application layer is oriented towards the end user, the service is different. So this issue is not discussed in this article. Although TCP/IP protocol does not provide the data link layer services, it is very important in our practical application, so this paper mainly discusses the data link layer, network layer and transport layer.

45.2.2 Security Vulnerability Analysis of Data Link Layer

Switch is a network device that works on the data link layer. Based on the results of the matching search for the destination MAC address and the MAC address table of its own store, the data packet is forwarded at the specified port. MAC address

Table 45.1 TCP protocol layer

Application Layer	Telnet	FTP	HTTP	SMTP	DNS
Transport layer	TCP		UDP		
Network layer	IP				
	ARP	RARP	ICMP	IGMP	IP
Network interface layer (data link layer and network layer)	Ethernet, 802.3, 802.5, FDDI, etc. (TCP/IP supports all of the standard physical and data link protocols)				

translation table is the basis of the switch, which has two kinds of generation methods. One is the static configuration, by the network administrator to manually enter the MAC address and port mapping records; the other is to use the switch address learning mechanism to automatically generate MAC address conversion table, which is the default switch address table generation method.

Switch address learning mechanism is as follows. When the switch just started to work, the MAC address table is not completely formed; it will forward the data in all ports. When a host sends IP packets through the port, it records the corresponding host MAC address and port information. When the MAC address learning process is complete, the MAC address table records the MAC address and port mapping information of the entire host. So the switch only forwards the received packets to the specific port. Each mapping record corresponds to a timer, when the time is up, the record will be deleted. When the communication data appears again, the switch will relearn the corresponding MAC address [3].

By using this automatic address learning mechanism of the switch, network attacks can be carried out by sending IP data packets with pseudo MAC address. Next, we take the typical MAC-PORT attack as an example to illustrate, as shown in Fig. 45.1:

Host 1 sent spoofed IP packets, destroyed the MAC address conversion table of the switch. The MAC address of the port 2 of the switch is changed, so that the data packets of host 2 which were sent to the external host cannot be sent through the switch, but misguided to host 1, causing the host 2 cannot contact with the external network.

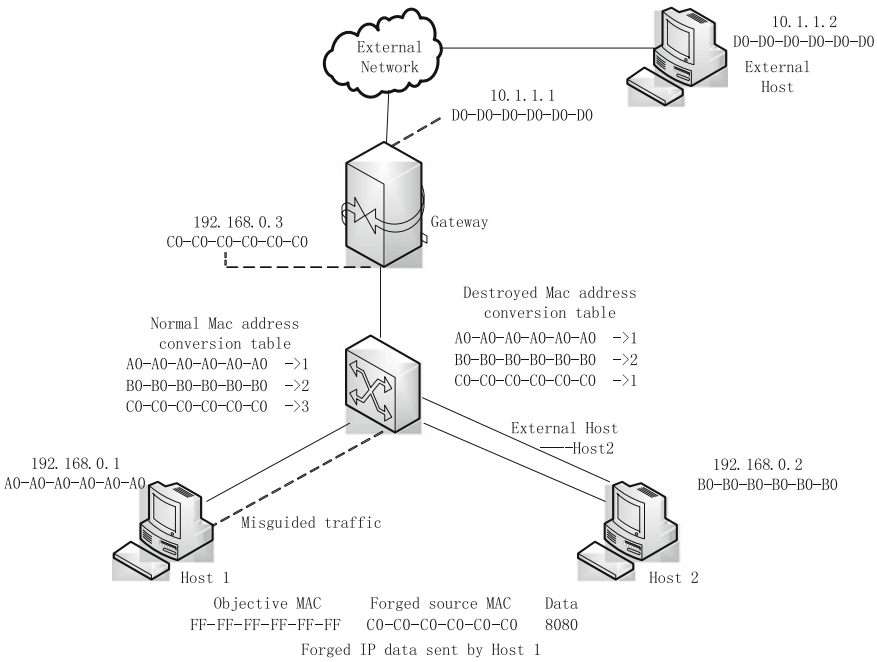


Fig. 45.1 Schematic diagram of MAC-PORT attack

45.2.3 Security Vulnerability Analysis and Protection of Network Layer Protocol

In the TCP/IP protocol, the network layer consists of 5 protocols, among which the most important one is the IP protocol, whose main responsibility is to deliver the IP data packets from the source host to the destination host in the whole network. In the process of transferring IP data packets, IP protocol requires ARP to determine the address of the next hop router and ICMP to deal with the abnormal conditions such as errors.

Limited by physical characteristics, the maximum length of a single IP data which can be transmitted by each type of network has a certain limit, which produces the problem of the division and recombination of IP data. This will produce an opportunity of fragment attack. The intruder forges the data message, and sends the abnormal data with overlapping migration into the target host. When the data is divided into the target host and reorganized in the stack, an error will occur while restructuring, causing the collapse of the protocol stack [4].

The network layer attacks can be divided into three types:

1. Abuse Header: Including malicious constructed, corrupted or illegally modified data packet of the network layer header, such as IP data packet with fake source address or containing false offset value.
2. Network Stack Vulnerabilities: The data packet contains components which are specially designed that enable the target host code which specifically process the network layer information to reject the service itself.
3. Bandwidth Saturation: A large amount of garbage data is used to consume the bandwidth of the target network in order to prevent legal communication.

45.2.4 Security Analysis and Protection of Transport Layer Protocol

The transport layer has two different protocols, the transmission control protocol (TCP) and the user data protocol (UDP). Both UDP and TCP use the port and the process of application layer for communication. Final target representation which UDP protocol used is to use the UDP port. Each port is an ultimate goal. In each of the hosts of the TCP/IP protocol stack, the UDP protocol has a port set, each of which is marked with an integer. Application program must apply for one or more of its own port number before the network communication, and all packets sent to the port will be sent to the application program. TCP protocol provides a connection-oriented reliable transport service, and uses the connection to indicate the final processing of data [5]. A connection is represented by two addresses and ports, such as: (192.168.8.21, 20)–(192.168.22.29, 29).

The most common attack on the transport layer is the port scan, which is the first step for a hacker to attack the network. By scanning the network host, the hacker can judge the operating system of the target host from the port data and the port number, and can grasp the structure of the network combined with other information, finally, implements the network attack.

SYN FLOOD attack utilizes the TCP protocol defects, by sending a large number of forged TCP connection requests, makes the target resource depleted. In the process of network operation, TCP half connection situation often appears. If half of the connection is much more, that is the phenomenon of SYN FLOOD attack.

45.3 The Network Security Deployment of the Communication Network for Space TT&C

As the main communication network of satellite launching experiments and controlling in orbit, communication network for space TT&C mainly has the following characteristics:

1. From the network structure, the topology of the communication network for space TT&C is relatively fixed. Static routing is used outside of the field, and dynamic routing is used in the field.
2. From the host security, terminal IP and the configuration are controllable, and the equipment personnel are relatively fixed.
3. From the use of the network, the size and direction of the data stream are controllable, and the demand for transmission quality is high.

The communication network for space TT&C configure the firewall between the intranet and extranet site, and strictly control the data exchange of Intranet and Internet. Through the construction of Unified security management, anti-virus and virus warning, host input and output control, security audit, network protection, information transmission encryption and other type. of systems at the sub stations and center, in order to meet the basic information security requirements of the communication network for space TT&C. Information security management node is located in the center, which has deployed a more comprehensive security management platform, including management server of antivirus, virus early warning, host control, security audit and other systems.

From the deployment situation, existing safety facilities have more comprehensive protection for the internal control of information flow, virus defense, information transmission security, but not involved the malicious active attack from inside. As an independent network, its most important network threats mainly come from the lack of understanding of the TCP/IP protocol and the improper configuration of the network.

45.4 Network Security Problems and Maintenance Measures of the Communication Network for Space TT&C

Through the above analysis of the TCP/IP layered protocol vulnerabilities and common attacks, we conduct a longitudinal analysis of security issues according to the data link layer, network layer, transmission layer of the order, and give the corresponding protective measures as follows.

45.4.1 Security Analysis and Protection of Data Link Layer of the Communication Network for Space TT&C

The communication network for space TT&C has a plurality of DTE terminals, and the data is sent through the gateway to the center by connecting the bottom layer switch. At present, in the bottom layer switch of the communication network for space TT&C, we can access the network by setting the internal IP address. If the MAC-PORT attack occurred in a new access host or a virtual machine, the communication between the intranet and extranet will be blocked, causing data cannot be sent to the center [6].

In response to this situation, the following measures can be taken: Bind the MAC address to the port of the bottom layer switch, and configure the MAC address table of the bottom layer switch in static way, which not only can prevent any host access, but also can effectively avoid the pseudo MAC-PORT attacks [7]. The disadvantage of this kind of response is that the time of handling the fault is prolonged when the switch port is in trouble. Not only need to replace the switch port, but also need to reconfigure the MAC address table. In addition, we can also use a variety of methods such as strengthen network management, strict examination and approval of network terminal, fixed IP address, regularly check the network interface configuration, etc. to effectively deal with the data link layer security problems.

45.4.2 Security Analysis and Protection of Network Layer of the Communication Network for Space TT&C

According to the security deployment of the communication network for space TT&C, we can see that through the filtering operation of the firewall, router, switch and other network equipment, we can effectively prevent some attack with illegal IP address, And we can also reject the attacker to send packets to the attack target by configuring black hole routing and other ways. But there are some loopholes in the

prevention of internal attack as follows: On the one hand, the forged data packets from the inside cannot be filtered; on the other hand, the traffic attacks from the internal host are unable to actively find and intercept [8].

The internal attack problem can be solved from the following three aspects:

1. Strictly control the configuration of the firewall and router. Strictly control the source address, the destination address and the traffic threshold of the past data. Ensure that the IP data package in the non-normal range cannot be passed.
2. Implement the secure transmission of IP packets by using the protection services based on encryption, the security protocol and the dynamic key management provided by IPSec protocol, and effectively avoid all kinds of attacks caused by the fake IP fragmentation.
3. Using software to monitor the data of the bottom layer switch, once the traffic out of range is discovered, we must carry out interception or manual intervention, and control the flow rate to the minimum.

45.4.3 Security Analysis and Protection of Transport Layer of the Communication Network for Space TT&C

Both the application program of the host and the port number of development in the communication network for space TT&C are controllable. So scanning the host and checking the port state, are effective methods to ensure the security of the transport layer [9]. Nmap is the most commonly used tool for port scanning, which provides a wide range of port scanning techniques.

In the communication network for space TT&C, voice and image is transmitted by using TCP protocol. Semi connection phenomenon often occurs, which leading to the three times handshake cannot be completed, the connection cannot be established, and the communication cannot be carried out. TCP three-way handshake is shown in Fig. 45.2.

The cause of this phenomenon occurs when the TCP third handshake. If there is a crash or a drop occurred after a user sent the SYN message to the server, then the server will not receive the ACK message of the client after a SYN + ACK response

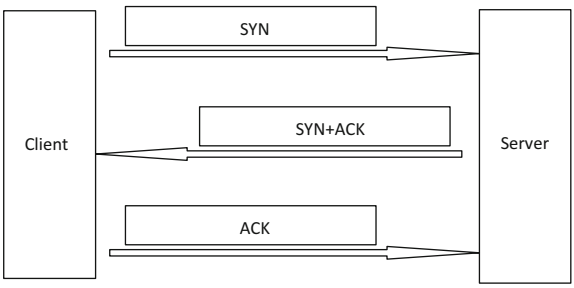


Fig. 45.2 The diagram of TCP three-way handshake

message is sent. At this point, the server will usually try to retry the connection, after waiting for a period of time, discard this unfinished connection. This time is called SYN Timeout. In general, this time takes minutes as orders of magnitude, and the length is about 30 to 60 s. If the server only maintains one or a few half connections, there is no problem. But a large number of semi connections can lead to stack overflow and system crash. This kind of problem is inevitable, but we should try to reduce the probability of occurrence of the problem.

From the defense point of view, we can use the following methods to solve the problem:

The first method is to reduce the Timeout SYN time. The harm of the semi connection depends on the semi connection number of the server. Therefore, by reducing the time between the server receiving the SYN message and discarding the connection, we can effectively reduce the burden on the server.

The second method is to set the SYN Cookie. Assign a Cookie to each IP address that is connected to the server. If the server received a continuous IP repeated SYN messages in a short time, we can identify that the server is attacked by this IP address, and drop the message from this IP directly.

For transport layer the most commonly used method is the network scanning. Using tools to scan the target host, found the loophole of the host, and repair it in time to ensure security [10].

45.4.4 The Design Scheme of the Configuration and Maintenance Management System for the Communication Network for Space TT&C

In summary, the safe operation of the communication network for space TT&C, not only needs the security technology to protect, but also needs the strict management and control. This section has designed a set of the configuration and maintenance management system for the communication network for Space TT&C, which can manage and maintain the communication equipment. The system function module diagram is shown in Fig. 45.3:

We collect, review and store the configuration information of all communication nodes of the communication network for space TT&C by the establishment of equipment configuration database, as a basis for regular maintenance and inspection. The information database includes the configuration information of firewall, router, security machine, switch, and the software of terminal host, open ports, etc. [11]. The server checks the configuration information periodically, and the information is classified and put into database. The network fault database classifies the faults in the network, in order to query for reference when dealing with faults in the future. Traffic monitoring software is deployed in the bottom network to monitor the traffic in LAN. Authority management is used to classify the users, and ensure the security of information by opening the permissions at different levels.

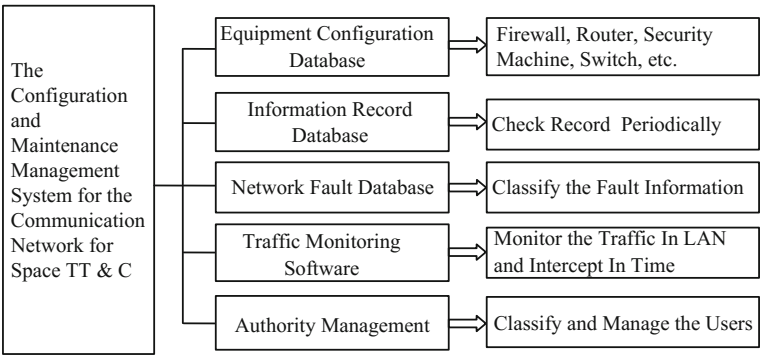


Fig. 45.3 Function module diagram of the configuration and maintenance management system for the communication network for space TT&C

45.5 Conclusions

In this rapidly changing and developing world, network security vulnerabilities are everywhere. Even if the old security loopholes were repaired, the new security vulnerabilities will continue to emerge. The network security system must adopt multiple technologies, and establish a highly security protection mechanism of joint defense, to ensure the network security. At the same time, in addition to the technical aspects of the factors, the network security also includes a large number of management issues. Therefore, the information security of the communication network for space TT&C is not only to use a variety of effective security measures to improve the level of security technology constantly, but also to improve the network management level by establishing and perfecting the network management standards. Finally, the information security of the communication network for space TT&C can be ensured.

References

1. Kumar A, Karthikeyan S (2011) Security model for TCP/IP protocol suite. J Adv Inform Technol 22(1):87–91
2. Gary W, Stevens WR (2000) TCP/IP illustrated, Volume 1: the protocols. Addison Wesley, New Jersey
3. Xu GT, Duan YB, Qin YH (2014) Network Security Foundation. Tsinghua University Press, Beijing
4. Premchand, B. Ambhore, BB Meshram, Wankhade KAD (1999) Protection of Public Network Communication by using Internet Protocol Security. Int J Comput Sci Eng 2(3):493–499
5. Zhuge JW (2001) Network attack and defense technology. Electronic Industry Press, Beijing
6. Huang FW, Xu JR, Chen JS (2002) Research on security technology of computer Network. Appl Res Comput 05:46–48

7. Kota S, Goyal M, Goyal R, Jain R (2001) Multimedia satellite networks and TCP/IP traffic transport. Int J Comput Appl 2001(232):18–21
8. Xu LG, Chi JY (2015) Vulnerabilities and protection of the computer protocol security. China New Telecommun 16:5–6
9. Zhang YR, Zhao ZC (2004) Research on scanning technology of computer network. Comput Eng Appl 02:173–176
10. Zeng ZF, Yang YX (2000) The development and research of network security. Comput Eng Appl 10:1–3
11. Michel Riguidel (2000) Creating a New Security for Tomorrow's Communication Networks and Information Systems. Annales Des Telecommunications 2000(557):401–418

Chapter 46
Timeliness Analysis and Countermeasure of Remote Control of Equipment Monitoring and Control System

Jianglai Xu, Lei Wang and Hui Zhang

46.1 Introduction

Satellite ground station is an important part and hub of the application system of the satellite. In recent years, with the continuous rise of satellite data communication and relay transmission applications, show the communication satellite coverage rate is high, a large amount of data transmission and real-time advantages, played a prominent role in order to improve the application efficiency of space equipment.

In the early construction of the satellite ground station, we pay more attention to reliability and stability of remote monitoring system, but the control timeliness of equipment is not perfect in depth. The main performance is the satellite ground station equipment configuration time is long, timeliness is not high, in the task of preparing the occupation using is the large scale, which restricts the application of efficiency satellite system. At present, the system applications continue to expand, user demand is increasing, the use of increasingly strong, the requirement of remote control for the remote control of the ground station equipment is more and more high. Therefore, it is necessary to analyze the related factors that affect the control efficiency in the field of equipment monitoring, in the future system construction to be applied, and continuously improve the efficiency of equipment control.

J. Xu (✉) · L. Wang · H. Zhang
Beijing Spatial Information Relay Transmission Technology Research Center, Beijing 100094, China
e-mail: mybird1234@sina.com

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_46

46.2 Present Situation Analysis

Equipment control timeliness is a measure of the length of time that the device is required to be configured, the equipment is scheduled to complete the requirements of the time refers to the monitoring system of ground station to send the task script command, through the monitoring system decomposition and analysis, is sent to the hardware, the hardware required to complete the configuration and scheduled a successful response by the monitoring system layer analysis back to the task script command to end OK.

The shorter the time length is, the higher the timeliness of the monitoring system, and vice versa. The relationship of the factors affecting the timeliness of the monitoring system is shown in Fig. 46.1 it can be seen from Fig. 46.1, the timeliness of the monitoring system in order to get high, the device must be scheduled to complete the requirements of the T time as short as possible, while the T = t1 + t2 + t3 + t4 + t5 + t6 + t7 + t8, therefore, every aspect of the time will directly affect the timeliness of the system.

In the early stage of the satellite ground station monitoring system, the control method of the configuration items is script commands, macro commands, block commands, process commands, and single commands, Script commands are composed of macro commands, process commands, and single commands [1]. In this collection, According to a certain process logic, a single command combined into a process command and block commands, single command and block command consists of macro command, in accordance with the module parameter classification combined parameters of macro command, in accordance with the classification of link configuration combination configuration macro command, so from the point of view of a single set of scripts. The single command is an element, the process command, command block, macro command is set respectively. The relationships of various control commands are shown in Fig. 46.2.

Several commands are transferred between the system of internal control of each configuration item, in order to ensure the correctness of control information, monitoring system designed the control response feedback mechanism,

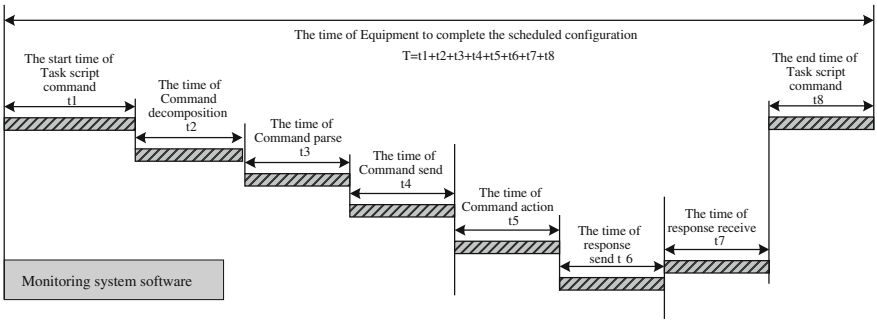


Fig. 46.1 The various factors affecting the timeliness of the monitoring system

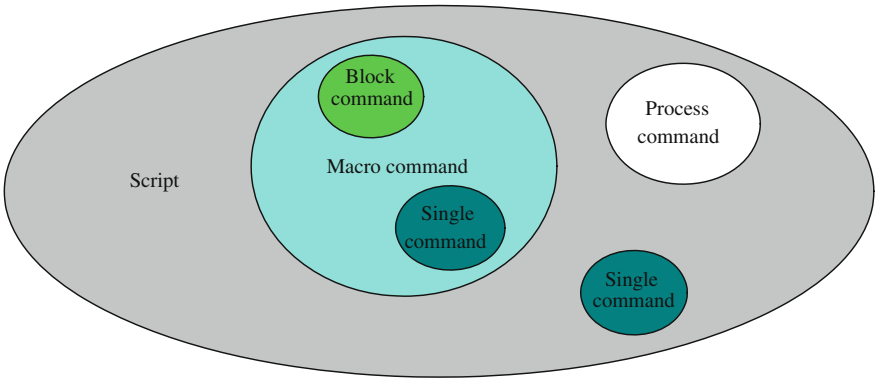


Fig. 46.2 Schematic diagram of the relationship of multiple control commands

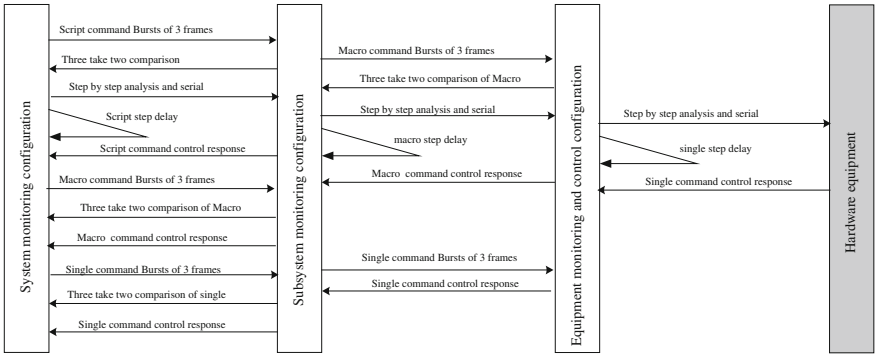


Fig. 46.3 Interaction mechanism between software configuration items in a typical monitoring system

for example, three take two judgments, parameter comparison, timeout mechanism design, but if the mechanism is not perfect situation, then will bring a certain control problem. Take system monitoring, subsystem monitoring, equipment monitoring three configuration items as an example, Fig. 46.3 shows a typical monitoring system software configuration items of interaction mechanism [2].

Without taking into account the inherent control of equipment hardware control time consuming. Only in the satellite ground station monitoring system level considerations, summed up, mainly due to the following factors:

1. device control script step delay

When the equipment control script performs step serial in editing, every step has increased to varying degrees of equipment state logic between step and step, also increased the response delay, resulting in equipment configuration parameters when more control, low efficiency.

2. the serial delay between single command and command block

In order to prevent the catching of the command frame, the margin of the reliable control device is set aside, and the time delay between the command frame is artificially added, Script contains more macro and single command, more cumulative effect of delay.

3. a variety of response mechanisms between the preparation of the redundant time consuming

There are three to take two to judge, control response, parameter comparison, timeout mechanism. These interaction mechanisms are integrated with each other, and the redundant design is redundant, which increases the time consuming of equipment control.

4. the parameters of the host and backup machine sending is repeated

Agreed ground station monitoring system in the early stage of the interface protocol, only online machine equipment and monitoring system in response to the communication, not online and sub machine monitoring system does not respond to communication, because of the preparation of host and backup machine can not be automatically configured for synchronous control, the same content, need to control the host and backup preparation machine, repeat the main parameters of the device were caused by equipment configuration parameters when more control, low efficiency.

5. device serial control

From the point of view of the early design of the monitoring system, equipment control command is different in serial issued, at the same time, the same equipment of different equipment unit (board) control command is issued in serial, each device controlled by will accumulate, every single element control equipment will be used accumulate, when the cumulative effect of the continuous accumulation, greatly increasing the equipment configuration parameters with time [3].

6. device macro commands to resolve duplicate

The agreed monitoring system in the early stage of interface structure, macro equipment is fixed, regardless of the structure parameters of macro control command is changed, system monitoring analysis will all commands in macro structure in turn issued again, this equipment configuration parameters when more control, low efficiency.

46.3 Strategies and Methods

46.3.1 Device Control Script Refinement

To classify the script, the script is divided into task script and emergency automatic script. The two have in common is the integration of macro command and single

command, the difference is to cancel the mission script logic (logic to automation scripts), integrated macro command and single command is issued all single parallel, serial no step. Automated scripts have a logical judgment, the integration of the macro commands and single command are all based on step serial. Most of the time because the system is working properly, often do not need to start the emergency automatic script (corresponding to the operation time can not accumulate), through the classification process, the task can be prepared in the monitoring system to further reduce the timeliness.

46.3.2 Delay Redundancy Between Commands

Early monitoring system for the prevention and control of the order to catch up, there may be caused by equipment control is not in place. But from the practical application effect, in the single command frame to add the sleep delay, resulting in the script contains the macro, the single command more, the cumulative effect of delay.

In fact, in the design of the interface, if you have already designed a single command, block commands to control the response to upload and timeout limit, sleep design is redundant design, it should be abolished. Table 46.1 shows the cancellation of the sleep design of the equipment to control the contribution, you can see the abolition of the design of sleep, the device after the script control to shorten the 2–3 m, the effect is obvious.

46.3.3 The Response Control Mechanism of the Host Equipment

The response mechanism of the host machine is adopted to solve the problem that the main machine parameters of the host equipment response and the backup machine are not repeated. The up-monitor will direct the control command to the standby unit broadcast issued, the host equipment and backup equipment synchronous configuration parameters, return control response to monitor, the host and backup machine with the maximum value, namely the equipment configuration parameters with time.

Table 46.1 Comparison of the use of a certain type of task script in a satellite ground station

Task script ID	Command frame addition sleep (1000)	Command frame cancel sleep (1000)
A_ task	4 m 33 s	1 m 26 s
B_ task	3 m 27 s	1 m 27 s
C_ task	3 m 43 s	1 m 19 s

Table 46.2 The device identification field and the host and backup of the relationship field

ID of use	ID of backup	Description of backup relations
1 + 2	0001	High-speed terminal 1/Master, High-speed terminal 2/Backup
1 + 3	0010	High-speed terminal 1/Master, High-speed terminal 3/Backup
1 + 4	0011	High-speed terminal 1/Master, High-speed terminal 4/Backup
2 + 1	0100	High-speed terminal 2/Master, High-speed terminal 1/Backup
2 + 3	0101	High-speed terminal 2/Master, High-speed terminal 3/Backup
2 + 4	0110	High-speed terminal 2/Master, High-speed terminal 4/Backup
3 + 1	0111	High-speed terminal 3/Master, High-speed terminal 1/Backup
3 + 2	1000	High-speed terminal 3/Master, High-speed terminal 2/Backup
3 + 4	1001	High-speed terminal 3/Master, High-speed terminal 4/Backup
4 + 1	1010	High-speed terminal 4/Master, High-speed terminal 1/Backup
4 + 2	1011	High-speed terminal 4/Master, High-speed terminal 2/Backup
4 + 3	1100	High-speed terminal 4/Master, High-speed terminal 3/Backup

But some equipment design is not a simple 1:1 backup, they do not have a host and backup binding relationship, their backup relationship is not sure, in the control command frame, we use the device identification field and the main preparation relation field to solve the problem that the device returns to the upper monitor.

For any 2 sets of equipment, in order to identify their main preparation, only need to configure the design of backup identification field, it is P_n^2, the host and backup machine receives the control command frame, parsing out the backup identification field, the corresponding backup device is synchronized with the device parameter configuration, and the control response is returned to the upper control unit, the maximum value of the host and backup machine, that is, the time when the parameters of the device configuration.

In the high-speed terminal 1, high-speed terminal 2, high-speed terminal 3, high-speed terminal 4, their backup relationship is not sure, we can identify in the control command frame, configuration backup identification field, the corresponding device according to the identity of the corresponding settings, ready to return the response. As shown in Table 46.2.

46.3.4 Control Response Mechanism Between Commands

Between the equipment and the up-monitor, we use the method of multicast communication, in order to prevent UDP bag loss may occur, for a control command, we use three take two mechanism + control response, or the use of control response + monitoring of the mechanism [4].

Three take two control mechanism + response mechanism is: as far as a control command is concerned, controller bursts three frame of control command, after receiving the control, as long as the three frames received 2 frames of the same, to the control side to return two sentenced to three response, otherwise not responding.

The prosecution was sentenced to two return three response, according to the control command, control in place, return control command response.

The control response mechanism + the monitoring comparison mechanism is: the controller sends out a frame of control commands, and after being received by the controller, the control command action is controlled, and the control command response is returned.

When the number of control commands is more, the time consuming of the 2 mechanisms is obviously different, as shown in Fig. 46.4, which corresponds to the left side of the three take two mechanism + control response mechanism, the right side of the control response mechanism + the monitoring comparison mechanism.

Three sentences two mechanism + control response mechanism, as shown in the Fig. 46.4, a control command from the control side to the control response returned to the control side, the characterization of the completion of this control, a total of 5 steps:

The control command frame 3 bursts, consuming 3 ms;
Three take two, time consuming 1 ms;
Three take two response, time consuming 1 ms;
The control associated with the control object, denoted by Tc;
The control response, time consuming 1 ms.

The completion of a single control command requires time consuming 3 + 1 +1 + Tc + 1 = 6 ms + Tc. But Tc is related to the hardware of the control object, and it has nothing to do with the monitoring system. Therefore, in the case of only considering the efficiency of the monitoring system, two take to three + control response mechanism is 6 ms. Assuming the 100 control commands, the time required to take 100 * 6 = 600 ms.

Using the control response + monitoring comparison mechanism, as shown in the Fig. 46.4, a control command from the control side to the control response returned to the control side, the characterization of the completion of this control, a total of 4 steps:

Control command 1 frames to send, time consuming 1 ms;
Control, and the control object, with Tc said;
Control response, time consuming 1 ms;
The monitoring on time 1 ms.

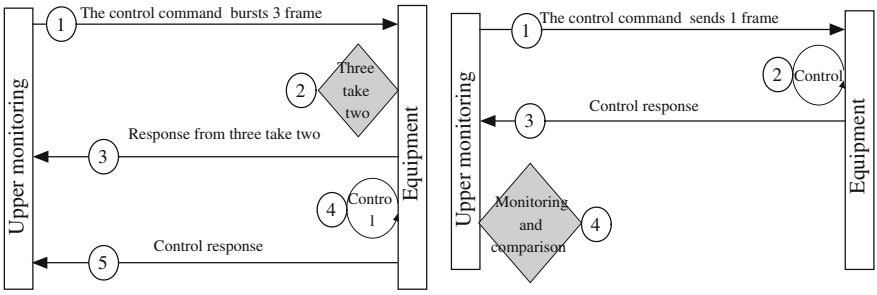


Fig. 46.4 Comparison of two control response mechanisms

In this case, the completion of a single control command requires time consuming 1 + Tc + 1+1 = 3 ms + Tc. But Tc is related to the hardware of the control object, and it has nothing to do with the monitoring system. Therefore, in the case of only considering the efficiency of the monitoring system, the control response + monitoring comparison mechanism is 3 ms. Assuming the 100 control commands, the time required to take 100*3 = 600 ms.

Compared to three take two + control response mechanism, the control response time is reduced by 1 times, which improves the efficiency of equipment control.

46.3.5 Device Parallel + Serial Control Mechanism

Parallel serial hybrid control equipment, there are 2 different situations, one is the different categories of equipment + serial parallel hybrid control, the other is a different unit of the same device (card) + serial parallel hybrid control. For each part or unit control is completely independent of the situation, should use the parallel control mechanism for each part or unit control is not completely independent of the situation, should use the serial control mechanism, if each part of the system or the control unit is completely independent of the situation, there are not completely independent of the situation, should use the parallel hybrid control mechanism + serial, which can greatly reduce the time to control the entire system equipment. From a system point of view, we assume that a system has a number of parts or components, which can be independently controlled by the N_1, N_2……N_m parts or elements, the corresponding control time were t_1, t_2……t_m, which is not an independent control of the parts or units by N_i, N_{i+1}……N_{i+k} parts or elements, the corresponding control time were t_i, t_{i+1}……t_{i+k}, if the serial control mode, the system control of the total time is $t_1 + t_2 +$……$+ t_m + t_i + t_{i+1} +$……$+ t_{i+k}$, if the parallel + serial hybrid control mode, the system control of the total time is Max (t_1, t_2……t_m) + $t_i + t_{i+1} +$……$+ t_{i+k}$.

A satellite ground station as an example, the station has to achieve equipment, equipment, power amplifier fed up converter, switch matrix, field equipment, mixer, low-speed and high-speed baseband baseband sets, using equipment serial control of the total time to reach 544 s. Because these devices are controlled independently, so if the equipment can control the process of parallel control station monitoring system on the ground in accordance with the types of equipment and set the dynamic open control system equipment control thread corresponding to the total time for a single set of control time maximum, namely Max (2, 4, 1.5, 1.5, 1, 32, 30) = 32 s. It is 16 times lower than the total time of the device serial control, which greatly improves the efficiency of the equipment control. As shown in Table 46.3.

In addition, can be seen from Table 46.3, a single set of equipment control is the most time-consuming baseband terminal equipment, if the set of M7 is relatively large, the cumulative control is the most time-consuming, to reduce the control of this part of the time consuming relates to different equipment unit the same equipment (board) of the parallel control problem. In the low-speed baseband

Table 46.3 Comparison of serial and parallel control of independent device

N	Device name	Amount	Single set of control time (s)	Accumulation of control time (s)	Explain	Serial control time (s)	Parallel control time (s)
1	Antenna servo	m1	2	2*m1	Independent	544	32
2	Power amplifier	m2	4	4*m2	Independent		
3	Up converter	m3	1	1*m3	Independent		
4	Switch matrix	m4	1.5	1.5*m4	Independent		
5	Field amplifier	m5	1.5	1.5*m5	Independent		
6	Down converter	m6	1	1*m6	Independent		
7	Medium and low-speed baseband	m7	32	32*m7	Independent		
8	High-speed baseband	m8	30	30*m8	Independent		

terminal as an example, generally by the machine, prior to the test receiver, modulator, analog source, back to the receiver composed of equipment, including machine, prior to the test receiver, modulator, analog control source of these devices can be done independently of each other, but returned to the receiving machine control needs to be carried out in the whole control is completed, at the same time returns to the receiver by back to the internal processing board, receiving FPGA1, receiving FPGA2, receiving FPGA3, receiving DSP, FPGA, DSP etc. the capture, which return to the motherboard, FPGA3, DSP, receiving, receiving, can be independently configured to capture FPGA.

Because these devices are controlled independently, and some are non independent control, analysis for simplicity, assume that unit set amount is 1, namely K1 = K3 = K4 = H1 = H2 = H3 = H4 = H5 = H6 = H7 = 1, hybrid control mechanism adopts parallel equipment + serial, time-consuming for the Max (1, 2, 1, 2, 1, 6, 3, 2) +6+6 + 2=20 s, and the use of equipment serial control of the total time to reach 32 s, such as shown in Table 46.4.

46.3.6 Macro Command Parsing Advance Optimization Mechanism

The macro command is a set of commands that are arranged according to a certain structure, complete the predetermined control, and have the equipment link configuration macro command and the device parameter configuration macro command in the equipment monitoring system [5].

Table 46.4 Comparison of serial and parallel control for independent and non independent devices

Device	N	Unit		Amount	Single set of control time (s)	Accumulation of control time (s)	Explain	Serial control time (s)	Serial/parallel control time (s)
Medium and low -speed baseband	1	Total device		K1	1/K1	1	Independent	32	20
	2	Forward modulator		K2	2/K2	2	Independent		
	3	Test receiver		K3	1/K3	1	Independent		
	4	Simulation source		K4	2/K4	2	Independent		
	6	Backward receiver	Main	H1	1/H1	1	Independent		
	7		FPGA1	H2	6/H2	6	Non independent		
	8		FPGA2	H3	6/H3	6	Non independent		
	9		FPGA3	H4	6/H4	6	Independent		
	10		DSP	H5	3/H5	3	Independent		
	11		Capture FPGA	H6	2/H6	2	Independent		
	12		Capture DSP	H7	2/H7	2	Non independent		

Monitoring system before the macro command issued, the first optimization command set, that is, according to the existing monitoring content and macro content to do the difference set, get the smallest set, and then send. This can remove the redundant control of the macro control of the device to achieve the purpose of improving the timeliness.

A device parameter configuration macro command as shown in Table 46.5.

If the current device corresponding to the macro n parameters, only K parameters and the equipment does not match the real state, need to be set, and other n-k equipment status and the actual state of the equipment parameters are consistent, no change, in this case, if we do not advance macro to resolve optimization, monitor still according to the parameters of macro structure will sort n devices under the control of parameters in turn again, but in fact there are n-k equipment control parameters is redundant, so it takes time for equipment configuration.

If the macro to resolve pre optimization mechanism to monitor the n parameters are compared with the current equipment monitoring, detection equipment parameter change of k parameter re optimizing composition of macro set then issued, the optimization of macro set must be a subset of a set of original macro command, thus reducing the equipment configuration parameters when the time. As shown in Fig. 46.5, the left side is the case that the macro analysis is not performed in advance, and the right side is the case that the macro analysis is performed in advance.

For example: a macro command intensive task parameters corresponding to 200 parameters, according to the optimization way to perform 2 min (redundant execution, that has been in place of the control parameters should be reset once per set once, need to return the corresponding control response, these will affect the implementation of the monitoring efficiency). If the switch to macro configuration analysis pre optimization mechanism, monitoring system according to the existing monitoring contents and macro content difference set, get the minimum set, that is through the operation mechanism of the 200 parameters may have 190 parameters is already setting, only 10 parameters are not set to the value, then just the 10 is not in place of the parameter settings, so as to improve the timeliness of monitoring. Due to the different parameters of each task, the minimum set length is dynamically changed, which can only be implemented by dynamically changing the macro structure in the current monitoring multicast programming.

Table 46.5 Frame structure of macro command

Field class	Macro ID	Macro length	Parameter 1	Parameter 2	Parameter 3	Parameter 4	……	Parameter n
Assignment	14303451123	233	01	10	1E	2A	……	AD

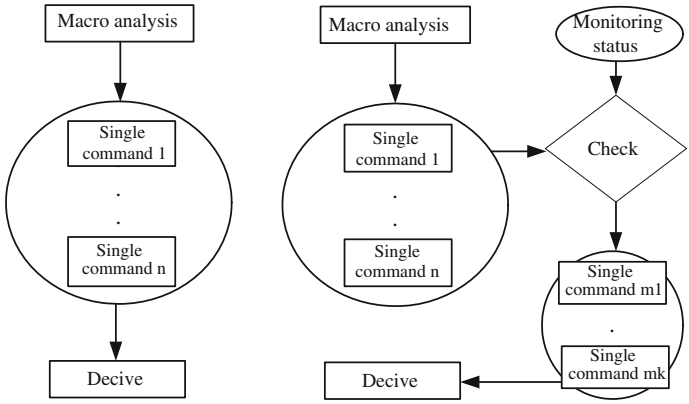


Fig. 46.5 Comparison of pre optimization of macro analysis

46.4 Conclusions

At present, all kinds of communication satellite applications expand continuously, Carry out all kinds of test tasks to enhance the ability of imminent requirements, satellite ground station equipment requirements more and more high efficiency parameter configuration. Especially with the further construction of satellite ground application system further, satellite ground station equipment set to further increase the amount, in practical work, should rethink the design of satellite ground station equipment configuration ideas and methods, to adapt to all kinds of test tasks, fast switching tasks, improve the timeliness of satellite ground station remote control equipment monitoring system, it is a useful attempt.

References

1. Zhiqin Z (2001) Aerospace measurement and control equipment software, automation. Beijing, Expo (5):56
2. Pengfei Y (1998) SQL server database application development technology, the people's. Posts and Telecommunications Publishing House, Beijing, p 689
3. Xinyan L (2000) ASP application classic. China Railway Publishing House, Beijing, p 347
4. Xiushan Y (2003) Software test automation. Electronic Industry Publishing House, Beijing, p 498
5. Nian D (2007) The software performance testing process and case analysis. Tsinghua University Publishing House, Beijing, p 279

Chapter 47
Designed on Operation and Management System for Aerospace TT&C Station

Qi Dang, Weiping Li, Dong Guo, Shuncheng Ning and Xiaopeng Wang

47.1 Introduction

Since the middle of 60 s in twentieth century, from the satellite observation network after more than 50 years of development and perfection, aerospace measurement and control system in our country has experienced from nonexistent to existent, from weak to strong development process, and gradually is formed to a rational layout, work coordination, strong adaptability of Tracking Telemetry and Control (TT&C) network [1]. The space measurement and control network is usually composed of TT&C center and TT&C station (ship) [2]. The station is a basic part of TT&C network, the equipment in the station includes tracking and measuring system, telemetry system, unified time system, communication system, meteorological system, service support system and so on.

For a long time, because of the vertical management of the professional equipment, making the development of four professional major including TT&C, communication, meteorological, service support in the station management domain is very uneven, formed to the development pattern of making the TT&C and communication domain as the centre of gravity, and the meteorological and service major is a relatively low degree of concern. As aerospace TT&C missions are becoming more and more intensive, the technology of equipment automation operation is applied gradually [3, 4], and the number of duty staff in TT&C station is relatively reducing [5], the service equipment is becoming increasingly important as the infrastructure, if these equipments occur fault during the measuring and

Q. Dang (✉) · W. Li · D. Guo · S. Ning · X. Wang
Xi'an Satellite Control Center, Xi'an 710043, China
e-mail: dangqifuping@163.com

W. Li
Key Laboratory for Fault Diagnosis and Maintenance of Spacecraft in Orbit, Xi'an 710043, China

R. Shen and G. Dong (eds.), *Proceedings of the 28th Conference of Spacecraft TT&C Technology in China*, Lecture Notes in Electrical Engineering 445,
DOI 10.1007/978-981-10-4837-1_47

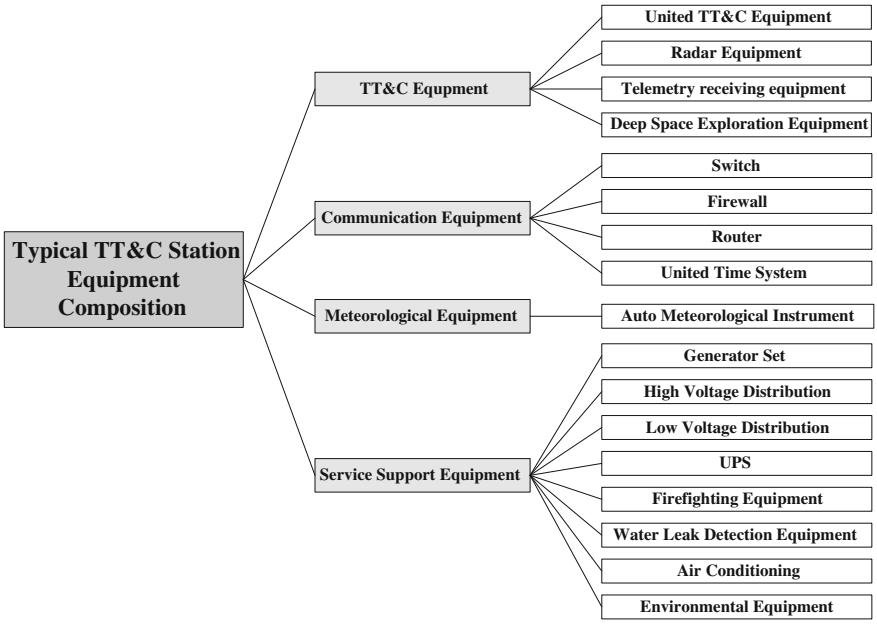


Fig. 47.1 The composition diagram of typical TT&C station equipment

tracking of spacecraft, the influence on operation security of spacecraft will be immeasurable.

This paper takes the typical TT&C station as the basis, starting from the main equipment and the information composition in the TT&C station, designed the station operation management network, and the function of equipment status acquisition, processing, information fusion and integrated display from the perspective of TT&C and comprehensive management, and realized the interconnection and interflow between different professional domain equipment in TT&C station, through the designing and applying of the TT&C duty process, laid the foundation for the follow-up intelligent development.

47.2 The Composition of TT&C Station Equipment

The space TT&C station in our country mostly is the comprehensive TT&C station [1], generally includes various types of different TT&C equipment, such as unified TT&C equipment, radar equipment, telemetry receiving equipment and deep space exploration equipment. Communication equipment mainly includes router, firewall,

access switch, aggregation switch, and unified timing system and so on [6]. Meteorological equipment is generally auto meteorological instrument. The service support equipment in station mainly includes generator set, high and low voltage power distribution, UPS(Uninterrupted Power Supply), firefight equipment, water leak detection equipment and air conditioning and so on [7–10]. Considering these service equipments, the geographical distribution is wide, the information capacity is relatively low, and the management means is backward (Fig. 47.1).

47.3 The Information Composition of TT&C Station

Station operation management information includes equipment basic information, mission related information and equipment status information.

The equipment basic information includes equipment name, type, TT&C frequency band, TT&C system, and so on. The mission information includes mission name, tracking plan, the orbit prediction and so on. The equipment status information includes the real-time running state and comprehensive status from various kinds of equipments including TT&C, communication, meteorological and service equipment, the specific content seen in Table 47.1.

Table 47.1 Equipment status information list

Professional domain	Type	Information content
TT&C	TT&C Data	Including the range, velocity, angle and telemetry source information
	Data check result	The data check result of range, velocity, angle information, and the check statistical information
	Equipment monitoring information	Real-time running state information of TT&C equipment, including automatic operation status, locking state, etc.
	Computer	The resource occupation information of TT&C computers
Communication		Communication link state information, communication link structure topology and fault alarm information
Meteorological		Temperature, humidity, pressure and other meteorological data and the status information of meteorological instrument

(continued)

Table 47.1 (continued)

Professional domain	Type	Information content
Service	Generator set	The running status such as output voltage, current, frequency (speed) and water temperature, oil level, oil pressure and the state such as the on-off state of fuel valve, etc.
	UPS	Input and output voltage, current, frequency, power, battery voltage, backup time, temperature and the state of rectifier, inverter, battery, bypass, load and other parts [7, 8]
	DC power supply	Battery voltage, battery temperature,the state of input electricity and battery voltage
	High and low voltage power distributors	The main circuit and sub-circuit voltage, current, frequency, active power, power factor, reactive power, apparent power, active power degree, reactive power degree and other parameters of Level 1 and Level 2 AC power distributors
	Smoke detector	The output state of detector
	Water leak detection equipment	On-off state
	Environment monitoring equipment	Temperature and humidity environment parameters of the key room

47.4 Design of TT&C Station Operation Management System

47.4.1 Design of Station Operation Management System

Station operation management system uses a layered architecture, and is divided into management layer and managed layer [11], is shown in Fig. 47.2.

Management Layer realizes the comprehensive operation management and the professional operation management functions. The professional operation management for TT&C, communication, meteorological, service and other professional domain business, is responsible for collecting the managed objects and status information and reporting the information to the station comprehensive operation management system, realizes the management functions which are closely related to business affairs. The comprehensive operation management for global situation realizes the global transaction management function which needs multi business coordinate such as comprehensive status management, integrated status monitoring, voice alarm, data record and duty process management and so on.

Managed Layer is composed of managed objects including TT&C, communication, meteorological, service and so on. The managed objects are responsible

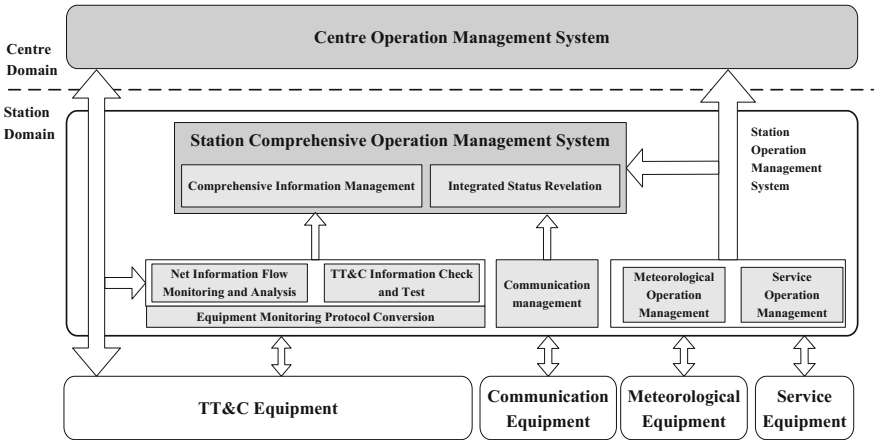


Fig. 47.2 The architecture diagram of TT&C station comprehensive management system

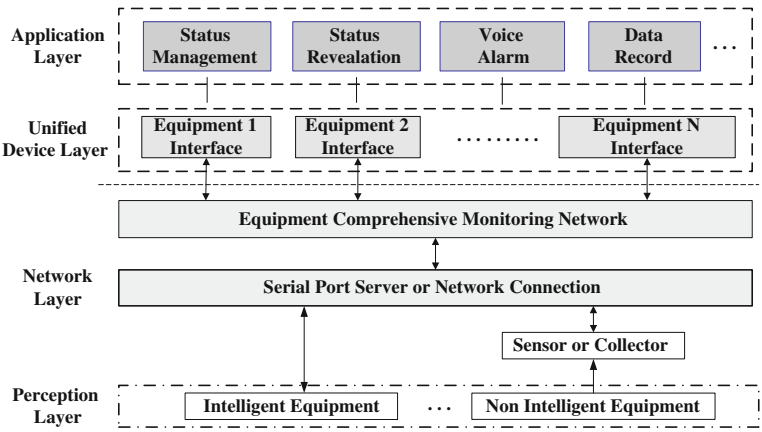


Fig. 47.3 The hierarchical structure diagram of TT&C station comprehensive management system

for reporting the status of equipment, receiving the configuration command or control command from a higher operation management layer.

The structure model is refined, which can be divided into four hierarchical layers: perception layer, network layer, unified device layer and application layer [12], as shown in Fig. 47.3.

Perception layer is composed of the field unit, which completes the on-line data collection, including the collection of non intelligent equipment and the collection of intelligent equipment.

Network layer based on the UDP/IP network structure, using cable transmission, is to meet the needs of different network channel transmission.

Unified device layer is to complete the normalization of different application protocol, establish the unified device model and realize the unity of the equipment monitoring interface.

Application layer is mainly for the user and the software module of automation process, to realize the functions such as status management, integrated status revelation, fault alarm, data record, and data proxy transmitting to the superior management domain, etc.

47.4.2 Design of Functional Structure

As functional structure considering, the station operation management system includes 5 function parts which are station comprehensive management, station TT&C management, communication management, meteorological management, and service management. The function structure diagram is shown in Fig. 47.4.

1. Comprehensive information management function. Responsible for the equipment static information management in TT&C station management domain, to realize the function of basic information storage and maintenance, equipment working situation statistics, station duty process management, and the service interface of accessing the database.
2. Integrated status revelation function. To realize the display of TT&C network running situation, achieve sound and light alarm when fault occurs, and transmit the integrated status information of TT&C, meteorological, service equipment, and the communication equipment topology and equipment status to the centre real-time information display system by the method of active push.

<table>
<tr><td rowspan="2">Comprehensive Operation Management</td><td colspan="7">Integrated Status Revelation</td></tr>
<tr><td colspan="7">Comprehensive Information Management</td></tr>
<tr><td rowspan="3">Professional Operation Management</td><td colspan="4">TT&C Operation Management</td><td>Communication Operation Management</td><td>Meteorology Operation Management</td><td>Service Operation Management</td></tr>
<tr><td colspan="2">TT&C Information Check and Test Software</td><td colspan="2">TT&C Equipment Fault Diagnose Software</td><td rowspan="2">Net Information Flow Monitoring and Analysis</td><td rowspan="2"></td><td>Station Meteorology Operation Management</td><td>Station Service Operation Management</td></tr>
<tr><td colspan="5">TT&C Equipment Monitoring Protocol Conversion</td><td>Sub Region Meteorology Operation Management</td><td>Sub Region Service Operation Management</td></tr>
<tr><td>Information Acquisition</td><td>Radar Equipment Monitoring Protocol Conversion</td><td>United TT&C Equipment Monitoring Protocol Conversion</td><td>Telemetry Receiving Equipment Monitoring Protocol Conversion</td><td>Optical Equipment Monitoring Protocol Conversion</td><td>Three-level network Information Acquisition Interface</td><td>Meteorological Information Acquisition</td><td>Service Information Acquisition</td></tr>
</table>

Fig. 47.4 The function structure diagram of TT&C station comprehensive management system

3. TT&C information check and test function. To acquire all kinds of equipment internal and external network TT&C data (telemetry, tracking) and complete the comprehensive analysis, and provide a visual display of information to the user on duty or mission command personnel through the B/S (Browser/Server) mode.
4. Information flow monitoring and analysis function. The Information flow monitoring and analysis software is to complete the compliance testing of application layer information flow in designated network environment, and complete the statistics and reporting of information flow. The software can be deployed in any station network node, and complete the network information flow monitoring and analysis [13].
5. TT&C equipment fault diagnose function. By acquiring and processing the monitoring data reported from monitoring point and working process, to check the equipment fault, realize the fault alarm, give the diagnostic result by fault reasoning based on the existing expert knowledge and fault tree, and provide supplementary information for equipment fault diagnosis and troubleshooting [14].
6. TT&C equipment monitoring information protocol conversion function. To complete the centralized conversion between the non-standardized monitoring protocol which is currently used by TT&C equipment and the standard monitoring protocol, complete the monitoring information acquisition and processing, and realize the interface between the integrated status revelation system.
7. Service operation management function. Responsible for service support equipment resource management, information management, and the centralized management and monitoring of generator set, power supply, high and low voltage distributor, air conditioning, environment detection equipment, and water leak equipment and so on, realize the managed object parameter setting, state acquisition and forwarding in service support domain, and realize the fault early warning according to the working parameters of the service equipment [15].
8. Meteorological operation management function. To complete the management of meteorological terminal instrument, and the status acquisition and monitoring of TT&C station meteorological equipment, report the meteorological equipment status information of local station domain to the centre meteorological operation management software, or in response to the query from the centre meteorological operation management software.

47.4.3 Design of Management Network

The professional equipment management network in TT&C station domain using TCP/IP network, through three level switches, collected the equipment running state information and TT&C data from each field area and equipment room to the centre monitoring room, and realized centralized processing and display, the connection network diagram is shown in Fig. 47.5.

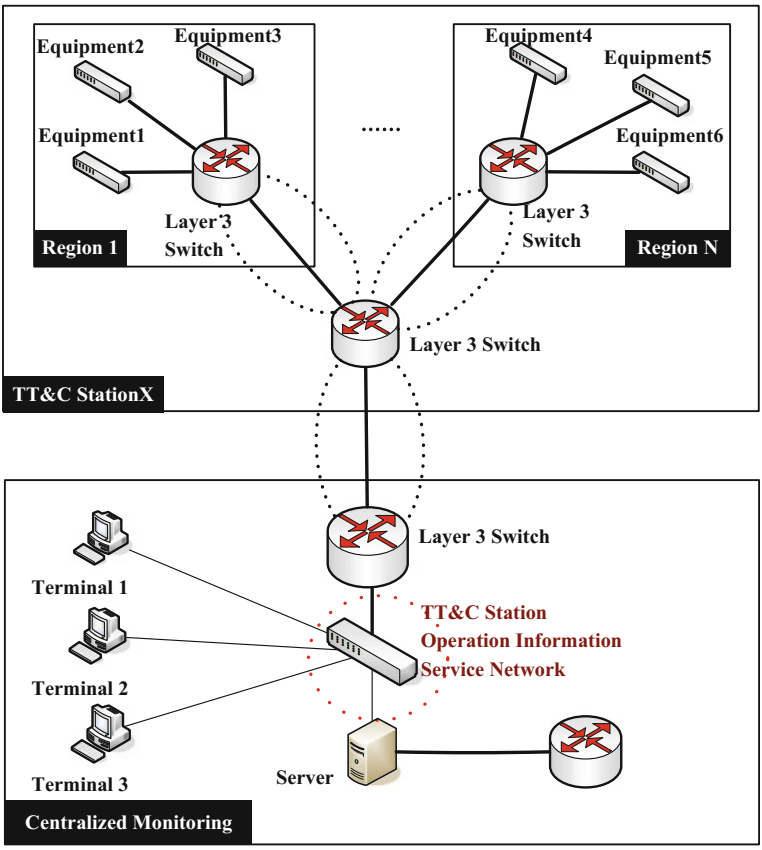


Fig. 47.5 TT&C station management connection network diagram

For the TT&C station which has multi sub-stations or regions, the service equipments are deployed in different regions, so in each region, establish the service operation management system of the region, if exists meteorological equipment independently, also establish corresponding meteorological operation management system.

47.4.4 Design of TT&C Equipment Status Information Acquisition

For the TT&C equipment monitoring information, using the method of centralized acquisition, the related work is carried out from three aspects, which are parameter acquisition, parameter normalization and parameter classification [16]. The TT&C equipment in station has the following characteristics:

1. Each TT&C equipment has a monitoring system;
2. Equipment monitoring system software uses C/S architecture [17];
3. The system composition of the TT&C equipment is basically the same, with the same or similar sub system.

Due to the status parameters of each TT&C equipment sub-system converging to monitoring server through the monitoring network, the equipment status acquisition has been realized, so all the equipment parameters and status information of a TT&C equipment can all be acquired from monitoring system theoretically. Also due to the composition of TT&C equipment subsystem is almost the same, it is possible to make the parameter to be modular. Therefore, when realizing the TT&C equipment information acquisition, using the method of centralized agent, with the monitoring server as the breakthrough point, developed and designed the TT&C equipment monitoring protocol conversion software based on monitoring client to realize the receiving, parsing and standardizing of equipment status information, and standardized encapsulate and send [16, 18].

Because there are many kinds of sub-system in the TT&C system, and the difference between different equipment parameters is large, even the parameters of the same equipment type have some differences. The difference of parameters has a serious impact on the data processing and using, so realized the parameters normalization, including the normalization of the parameter type, the parameter information and the meaning of the parameters.

47.4.5 Design of Service Equipment Status Information Acquisition

The status acquisition of service equipment used multi-level flat structure, composed of tree tandem level by level including the station service operation management, region service operation management, and monitoring unit, and so on, and can be used in multi-level tandem, hierarchical management. The model structure diagram of the system is shown in Fig. 47.6.

Using “equipment - order - node” model, for different access equipments using different communication protocols, designed the consistent data interface and normalized equipment model [15], including the unified object of equipment, order and data node, realized the information interaction with the equipment.

47.4.6 Design of Information Fusion and Display

Information acquisition has realized the information convergence of different system and different major to the server, and achieved the data fusion physically. In information processing and application level, through building the given logic and

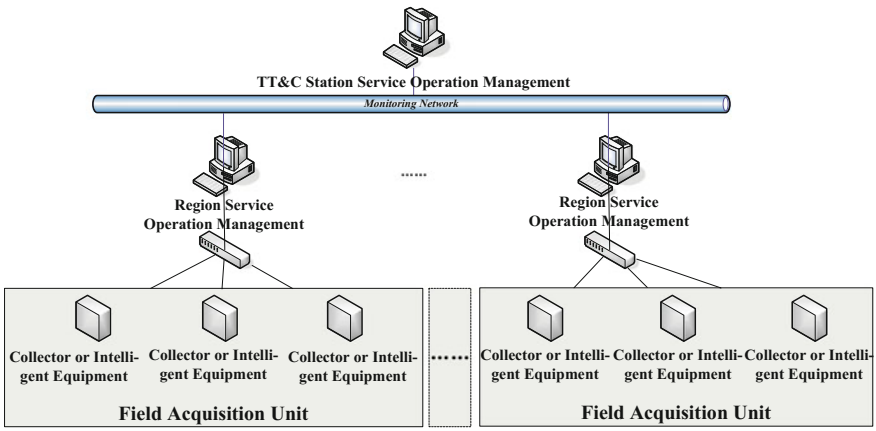


Fig. 47.6 The model structure diagram of service operation management system information acquisition

Source	Time	Level	Content
TICTS	2015-5-16 20 03:1	Severe	××× Plain 2 link break

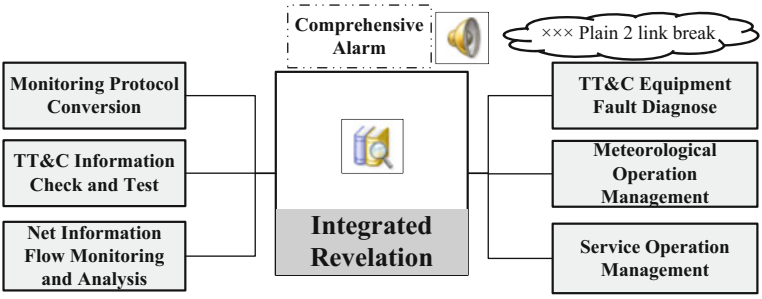


Fig. 47.7 The model diagram of information fusion and display

analyzing the relationship between the data, complete data fusion and integrated data using, the function includes the following three aspects, as shown in Fig. 47.7.

1. The fusion and analysis of equipment status and TT&C data

In the spacecraft tracking process, through the appropriate division of the TT&C process, complete the abnormal tracking data processing which generated in the initial period of tracking and the last period of outbound tracking. For example, when the equipment has not yet to establish stable communication link with the satellite, the generated data is not reliable, at this time, the system combined with the equipment locking status to determine whether to send tracking and telemetry data to the center.

2. The centralized fault alarm of multi domain and multi information

In the station operation management system, each system can realize the voice alarm according to the internal exceptions and errors of its own, but the exception information is isolated before the information fusion, and probably will alarm at the same time. So the system realized the integration and fusion of all kinds of alarm information, and set the corresponding priority and processing logic, which makes the warning process is based on the preset alarm logic.

3. Fault linkage and processing

Realized the fault linkage on the basis of multi equipment key parameters monitoring and the physical centralization of multi equipment operating platform, and also through the software scheduling, determine the problem and countermeasures combined with alarm source of integrated revelation system and alarm content, and frontward present the software or software interface need to deal with to make the staff on duty to deal with conveniently.

47.5 Conclusion

Based on the management requirements of TT&C station, designed the station TT&C equipment monitoring protocol conversion software, station service operation management software, station meteorological operation management software, station TT&C information check and test software, comprehensive information management software and integrated status revelation software, realized the integration status acquisition, data processing, information display and fault alarm functions of four professional domain equipment including TT&C, communication, meteorological, service support in TT&C station management domain, has the following advantages:

1. Realized the interconnection and interflow of information. From the perspective of centralized management, realized the status acquisition, integration, interconnection and interflow of four professional domain equipment including TT&C, communication, meteorological, service support, accumulated first-hand information resources for large data processing of TT&C network.
2. Realized the hierarchical management of information. From the perspective of system integration, realized a centralized alarm service system, collected the log and alarm information generated from all kinds of functions such as mission process, net information flow checking, fault diagnose and professional equipment management and so on, and implemented hierarchical management.

After the system put into operation, promoted the fusion of duty post, reduced the skill requirements of duty post, which is conducive to form a professional comprehensive duty team through the proper training, and is conducive to release the technical backbone, increase the human resource investment in scientific research and innovation.

References

1. Zhijian Y (2006) Status Quo and development of spaceflight TT&C systems. Eng Sci 8 (10):42–46 (in Chinese)
2. Fuming H, Henian H (2004) Spacecraft flight control and emulation. National Defence Industry Press, Beijing (in Chinese)
3. Yufeng M, Dong L, Haiqi Zh (2014) Research on automatically running technology to TT&C station. Gansu Sci Technol 30(19):43–46 (in Chinese)
4. Qi D, Dong G, Rujun G et al (2016) Research on the automatic flow remote customizing technology of TT&C equipment. 46(1):16–19, 64 (in Chinese)
5. Jinglong G (2011) Analysis on unattended technology for space TT&C station. Radio Eng China 41(12):38–40, 58 (in Chinese)
6. Xiaojun Zh, Li H, Keyun S (2016) Typical failure of IP network equipment and response measures. J Spacecr TT&C Technol 35(1):070–075 (in Chinese)
7. Wen Zh, Changjin L (2012) UPS battery remote on-line monitoring management system in television transmitting room. Comput Meas Control 20(11):2929–2931 (in Chinese)
8. Yao Y, Wenzhou G (2015) Intelligent room power and environment construction. China Digit Cable TV 02:199–201 (in Chinese)
9. Zhiqiang W (2008) Surveillance system design in power supply environment of computer room. Video Eng 32(08):74–75 (in Chinese)
10. Jing W (2015) The application of power supply and environment monitoring system in the passenger dedicated railway line's information equipment room. Inf Commun 146(02): 177–178 (in Chinese)
11. Qi D, Jun L, Weiping L (2014) Research of spaceflight TT&C network management mode. J Astrodyn 4(2):89–93 (in Chinese)
12. Simon B (2015) Software architecture for developers. Posts and Telecom Press, Deng Gang, Beijing (in Chinese)
13. Shengjun Q, Yan Ch, Min Y (2011) System modeling based on information flow structure analysis. J Dali Marit Univ 37(2):97–100 (in Chinese)
14. Ying D, Chao S, Fei W et al (2015) Research on general fault diagnosis system of aerospace TT&C network equipment. J Astrodyn 5(4):89–93 (in Chinese)
15. Guo D, Dongxing W, Qi D et al (2014) Design and realization of configuration software in TT&C network service monitor and control system. J Astrodyn 4(2):89–93 (in Chinese)
16. Weiping L, Shuo T, Daming B (2012) Design and implementation of an integrated monitor system for TT&C equipment. J Spacecr TT&C Technol 31(6):52–56 (in Chinese)
17. Qi D, Dongxing W, Daming B et al (2015) Web-based realization of integrated monitoring system to TT&C equipments. In: Proceedings of the 27th conference of spacecraft TT&C technology in China, Lecture notes in electrical engineering 323, Tsinghua University Press, Beijing, Springer, Berlin, pp 503–511
18. Qi D, Weiping L, Daming B et al (2015) Construct a centralized monitoring system of TT&C equipment with MLA theory. J Spacecr TT&C Technol 34(5):459–468