Reza N. Jazar · Liming Dai *Editors*

# Nonlinear Approaches in Engineering Applications

## Advanced Analysis of Vehicle Related Technologies

Springer

# Nonlinear Approaches in Engineering Applications

Reza N. Jazar • Liming Dai
Editors

# Nonlinear Approaches in Engineering Applications

## Advanced Analysis of Vehicle Related Technologies

Springer

*Editors*
Reza N. Jazar
School of Aerospace, Mechanical
and Manufacturing Engineering
RMIT University, Bundoora East Campus
Bundoora, Victoria, Australia

Liming Dai
Department of Industrial Systems
Engineering
University of Regina
Regina, Saskatchewan, Canada

ISBN 978-3-319-27053-1 ISBN 978-3-319-27055-5 (eBook)
DOI 10.1007/978-3-319-27055-5

Library of Congress Control Number: 2016940372

Springer Cham Heidelberg New York Dordrecht London


Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Dedicated to *Xinming* and *Mojgan.*

*You lose only if you give up.*

# Preface

This book follows the same concept of the previous volumes of the series of *Nonlinear Approaches in Engineering Applications*, organized by the editors. The nonlinear analysis techniques have been developed since a few centuries ago when the mathematical modeling of natural dynamical phenomena appeared. Although many of the approaches and techniques have been brought into research and engineering practices, linearization and simplification are still the dominating approaches existing in physics and engineering. Nature is nonlinear in general as the responses of physics and engineering systems are nonlinear. Linearization ends up with simplification, and it loses the original characteristics of the systems. Such simplifications usually lead to inaccuracies, misunderstandings, or even incorrect conclusions. For example, Hooke's law including the generalized Hooke's law is linear, and it composes the foundation of linear elasticity and dominates numerous solutions of physical systems and mechanical designs. However, no material is perfectly linear. Any material used in the real world can actually be a nonlinear system, not only due to its material or structure nonlinearity, but also due to the inhomogeneousness and anisotropy of the materials. The main characteristic of nonlinear systems is that every nonlinear system is nonlinear uniquely, while all linear systems are the same.

Another challenge facing the scientists and engineers in our time is the generation of the solutions and characterization of the nonlinear systems modeled from the physical systems in reality. It would be greatly beneficial in accurately evaluating the behavior of nonlinear systems and revealing the actual nature of the systems, with utilization of the existing mathematical tools and analytical means, if the analytical solutions of nonlinear systems could be pursued. Due to the nonlinearity and complexity of the nonlinear systems, usually, it is very difficult or impossible to derive the analytical and closed-loop solutions for the systems. In solving or simulating the nonlinear systems, one may have to rely on approximate or numerical methods, which may only provide approximate results for the systems while errors are unavoidable during the processes of generating the approximate results. Approximation and inaccuracy are the inescapable shadows following the current research and engineering practices involving nonlinearity or nonlinear systems.

In the role of the editors as well as the chapter contributors of this book, we have tried to present a collection of chapters showing the theoretically and practically sound nonlinear approaches and their engineering applications in various areas, in hoping that this book may provide useful tools and comprehensible examples of solving, modeling, and simulating the nonlinear systems existing in the real world. The carefully selected chapters contained in this book reflect recent advances in nonlinear approaches and their engineering applications. The book intends to feature in particular the fundamental concepts and approaches of nonlinear science and their applications in the fields of engineering and physics. It is anticipated that this book may help to promote the development of nonlinear science and nonlinear dynamics in engineering, as well as to stimulate research and applications of nonlinear science and nonlinear dynamics in physics and engineering practices. It is also expected that the book will further enhance the comprehension of nonlinear science and stimulate interactions among scientists and engineers who are interested in nonlinear science and who find that nonlinearity and complexity of systems play an important role in their respective fields.

With the theme of the book, *Nonlinear Approaches in Engineering Applications: Advanced Analysis of Vehicle Related Technologies*, it covers interdisciplinary studies on theories and methods of nonlinear science and their applications in complex systems such as those in vehicle dynamics, vehicle crashworthiness, vehicle complex steering kinematics, vehicle and train cluster networking, alternative internal combustion engines, alternative fuels, nonlinear vibrations and control, micro-nano vibrating systems, viscoelastic and soft tissue mechanics, and filtering control engineering. Examples include crash beam analysis and optimization in vehicles' application; caster–camber relationship in vehicle steering behavior; a discussion about design of limaçon gas expanders and their role in thermodynamic cycles; networks' approach in train and tram systems; a review and analysis on bio-oil as an alternative fuel for power generating; employing fuzzy sliding mode control in vibration analysis of Euler-Bernoulli beams; investigation on the nanobeam vibrations; analytical approaches in vibration analysis of the systems that obey general Duffing equation; biomechanics investigation including soft tissue, muscles, and viscoelastic materials; solar cells thin film optoelectronics modeling and their nonlinear characteristics; and dynamic model prediction by nonlinear filtering methods.

## Level of the Book

This book aims at engineers, scientists, researchers, and engineering and physics students of graduate levels, together with the interested individuals in engineering, physics, and mathematics. This chapter-book focuses on application of the nonlinear approaches representing a wide spectrum of disciplines of engineering and science. Throughout the book, great emphases are placed on engineering applications, physical meaning of the nonlinear systems, and methodologies of the approaches

in analyzing and solving for the systems. Topics that have been selected are of high interest in engineering and physics. An attempt has been made to expose the engineers and researchers to a broad range of practical topics and approaches. The topics contained in this book are of specific interest to engineers who are seeking expertise in vehicle and automotive-related technologies as well as engines and alternative fuels, mathematical modeling of complex systems, biomechanical engineering approaches to robotics and artificial muscles, nonclassical engineering problems, and modern mathematical treatments of nonlinear equations.

The primary audience of this book is the researchers, graduate students, and engineers in mechanical engineering, engineering mechanics, electrical engineering, civil engineering, aerospace engineering, mathematics, and science disciplines. In particular, the book can be used for training the graduate students as well as senior undergraduate students to enhance their knowledge by taking a graduate or advanced undergraduate course in the areas of nonlinear science, dynamics and vibration of discrete and continuous system, structure dynamics, and engineering applications of nonlinear science. It can also be utilized as a guide to the readers' fulfillment in practices. The covered topics are also of interest to engineers who are seeking to expand their expertise in these areas.

## Organization of the Book

The main structure of the book consists of three parts of automotive applications, vibrations applications, and engineering applications, including 12 chapters. Each of the chapters covers an independent topic along the line of nonlinear approaches and engineering applications of nonlinear science. The main concepts in nonlinear science and engineering applications are explained fully with necessary derivatives in details. Each of the chapters is intended to be organized as essentially self-contained. All necessary concepts, proofs, mathematical background, solutions, methodologies, and references are supplied except for some fundamental knowledge well known in the general fields of engineering and physics. The readers may therefore gain the main concepts of each chapter without frequently having to refer to the concepts of the other chapters and references. Readers may hence start to read one or more chapters of the book for their own interests.

## Method of Presentation

The scope of each chapter is clearly outlined, and the governing equations are derived with an adequate explanation of the procedures. The covered topics are logically and completely presented without unnecessary overemphasis. The topics are presented in a book form rather than in the style of a handbook. Tables, charts, equations, and references are used in abundance. Proofs and derivations

are emphasized in such a way that they can be straightforwardly followed by the readers with fundamental knowledge of engineering science and university physics. The physical model and final results provided in the chapters are accompanied with necessary illustrations and interpretations. Specific information that is required in carrying out the detailed theoretical concepts and modeling processes has been stressed.

## Prerequisites

This book is primarily intended for researchers, engineers, and graduate students, so the assumption is that the readers are familiar with the fundamentals of dynamics, calculus, and differential equations associated with dynamics in engineering and physics, as well as a basic knowledge of linear algebra and numerical methods. The presented topics are given in a way to establish as conceptual framework that enables the readers to pursue further advances in the field. Although the governing equations and modeling methodologies will be derived with adequate explanations of the procedures, it is assumed that the readers have a working knowledge of dynamics, university mathematics, and physics together with theory of linear elasticity.

## Acknowledgments

This book is made available under the close and effective collaborations of all the enthusiastic chapter contributors who have the expertise and experience in various disciplines of nonlinear science and engineering applications. They deserve sincere gratitude for the motivation of creating such a book, encouragement in completing the book, scientific and professional attitude in constructing each of the chapters of the book, and the continuous efforts toward improving the quality of the book. Without the collaboration and consistent efforts of the chapter contributors, the completion of this book would have been impossible. What we have at the end is a book that we have every reason to be proud of.

It has been gratifying to work with the staff of Springer-Verlag through the development of this book. The assistance provided by the staff members have been valuable and efficient. We thank Springer-Verlag for their production of an elegant book.

Bundoora, Victoria, Australia — Reza N. Jazar
Regina, Saskatchewan, Canada — Liming Dai

# Contents

# List of Figures

# Part I
# Automotive Applications

# Chapter 1
# Crashworthiness of Tapered Beams in Automotive Application

**Varshan Beik, M. Fard, and Reza N. Jazar**

**Abstract** In a car crash, the higher level of energy absorption in the frontal structures leads to less transferred energy to the passengers and hence a safer car. S-shaped front rails, also known as S-rails, are one of the main structural elements and energy absorbers in a car body. Energy absorption in the S-rails happens through local buckling. In order to improve the passenger safety in a frontal crash, S-rails design should be optimized to absorb higher level of energy while crushing. In this study, we investigate the crashworthiness impact of tapering S-rails. Two S-rails, one without internal diagonal reinforcement (type-A) and one with this reinforcement (type-B), both are tapered with 20 different tapering ratios ranging from 110 % to 300 % in 10 % increments. All S-rail models are subjected to static and also dynamic loading conditions. The effectiveness of tapering S-rails is assessed through investigating the energy absorption (EA) and specific energy absorption (SEA) variations using finite element method. An equation is developed to verify the numerical results. In this study, we showed the reinforcing and tapering S-rails both could improve the EA and SEA in both static and dynamic loading conditions. Combining reinforcing and tapering the S-rails showed a noticeable improvement in SEA of more than 300 % in static loading condition as well as 275 % SEA increase in dynamic loading condition.

**Keywords** Crashworthiness • Vehicle impact • Vehicle structural • Vehicle safety • Reinforced S-rails • Crash box • Vehicle dynamics

## Nomenclature

$a$ Depth of cross-section
$a_u$ Depth of cross-section in displacement of $u$
$b$ Width of cross-section
$b_u$ Width of cross-section in displacement of $u$

V. Beik (✉) • M. Fard • R.N. Jazar
School of Aerospace, Mechanical, and Manufacturing Engineering, RMIT University, Melbourne, VIC, Australia
e-mail: varshan_bml@yahoo.com; mohamad.fard@rmit.edu.au; reza.jazar@rmit.edu.au



R.N. Jazar, L. Dai (eds.), *Nonlinear Approaches in Engineering Applications*,
DOI 10.1007/978-3-319-27055-5_1

$E$ Young's modulus
$m$ Normalized bending moment
$M$ Moment
$M_0$ Fully plastic bending moment of cross-section
$M_\mathrm{f}$ Fully plastic bending moment
$\overline{P}$ Mean axial load
$R_t$ Tapering ratio
$t$ Wall thickness
$u$ Displacement
$V_0$ Initial velocity
$V_c$ Constant velocity
$W$ Work
$Y$ Yield stress
$\theta$ S-rail bending angle
$\psi$ Angle between bending axis and strong principal bending axis of cross-section
$\sigma_0$ Energy equivalent flow stress
$\delta$ Axial displacement

## 1.1 Introduction

Frontal vehicle collision is the most common type of car cash, resulting in thousands of deaths every year (National Highway Traffic Safety Administration 2012; Australian Road Deaths Database 2012; Roads and Maritime Services 2012). Consumers have become highly aware of the importance of vehicle safety which has made it a primary selling feature now a day. Moreover, the competitive nature in the automobile sector persuades manufacturers to develop safer vehicles. Accordingly, safety is of paramount importance in modern vehicle design. Crashworthiness studies are of the most important safety analysis to be completed. Vehicle crashworthiness, defined as the degree of passenger safety which a vehicle has while involved in a crash, has improved significantly in the past decades. In other words crashworthiness is the ability of a vehicle structure to withstand a crash and yet maintain a sufficient survival space for its occupants. Strict rules and regulations such as FMVSS (Federal Motor Vehicle Safety Standards), Euro NCAP (European New Car Assessment Program), and ANCAP (Australian New Car Assessment Program) set by governments are now controlling car manufacturers. These restrictions and also the highly competitive market have led to innovative advances in the industry. Structural developments, new safety features, and design improvements have all resulted in considerable progress in safety ratings. Extensive research has been carried out recently to improve frontal crash safety. Various simulation methods have been implemented in order to characterize crash conditions. Finite element method has become one of the leading theoretical methods used to predict vehicle safety (Vasudevan et al. 1996; Cheng et al. 2001). The United States National Highway Traffic Safety Administration founded the New Car Assessment Program

(NCAP) in 1979. This program was originally directed at promoting manufacturing of safe vehicles and also encouraging consumers to buy safe cars. Over time, this system has become more user-friendly by implementing safety rating programs. The Transport Research Laboratory of the UK Department for Transport founded the European New Car Assessment Program (Euro NCAP) in 1997 based on the NCAP. Similar programs were then introduced in different countries such as ANCAP in Australia and New Zealand, Latin NCAP in Latin America, and C-NCAP in China (Euro NCAP 2012).

### *1.1.1 Federal Motor Vehicle Safety Standards (National Highway Traffic Safety Administration 2012)*

Federal Motor Vehicle Safety Standards (FMVSS) are the American vehicle manufacturing regulations which identify various safety requirements ranging from design to performance elements. The first FMVSS (No. 209) adopted on 1 March 1967, specifying the requirements for seat belts is still valid. These standards have been updated and made more stringent over the years. In conjunction with the Canada Motor Vehicle Safety Standards (CMVSS) the FMVSS/CMVSS requirements are considerably different from the international UN requirements.

The latest version of these standards coded at 49 C.F.R. 571 is divided into three categories: crash avoidance, crashworthiness, and post-crash standards. The list of standards in each section is listed as below (National Highway Traffic Safety Administration 2012):

**Crash avoidance standards**:

| | |
|---|---|
| Standard No. 101 | Controls and Displays |
| Standard No. 102 | Transmission Shift Lever Sequence, Starter Interlock, and Transmission Braking Effect |
| Standard No. 103 | Windshield Defrosting and Defogging Systems |
| Standard No. 104 | Windshield Wiping and Washing Systems |
| Standard No. 105 | Hydraulic and Electric Brake Systems |
| Standard No. 106 | Brake Hoses |
| Standard No. 107 | [Reserved] |
| Standard No. 108 | Lamps, Reflective Devices, and Associated Equipment |
| Standard No. 109 | New Pneumatic Tires for Passenger Cars |
| Standard No. 110 | Tire Selection and Rims for Passenger Cars |
| Standard No. 111 | Rearview Mirrors |

| | |
|---|---|
| Standard No. 112 | [Reserved] Requirements moved to Standard No. 108—Headlamp Concealment Devices |
| Standard No. 113 | Hood Latch System |
| Standard No. 114 | Theft Protection |
| Standard No. 115 | [Reserved] Requirements moved to Part 565—Vehicle Identification Number |
| Standard No. 116 | Motor Vehicle Brake Fluids |
| Standard No. 117 | Retreaded Pneumatic Tires |
| Standard No. 118 | Power-Operated Window, Partition, and Roof Panel Systems |
| Standard No. 119 | New Pneumatic Tires for Vehicles Other Than Passenger Cars |
| Standard No. 120 | Tire Selection and Rims for Motor Vehicles Other Than Passenger Cars |
| Standard No. 121 | Air Brake Systems |
| Standard No. 122 | Motorcycle Brake Systems |
| Standard No. 125 | Motorcycle Controls and Displays |
| Standard No. 124 | Accelerator Control Systems |
| Standard No. 125 | Warning Devices |
| Standard No. 126 | [Reserved] Requirements moved to Part 575.103 Truck-Camper Loading |
| Standard No. 129 | New Non-Pneumatic Tires for Passenger Cars - |
| | New Temporary Spare Non-Pneumatic Tires for Use on Passenger Cars |
| Standard No. 131 | School Bus Pedestrian Safety Devices |
| Standard No. 135 | Light Vehicle Brake Systems |

## Crashworthiness standards

| | |
|---|---|
| Standard No. 201 | Occupant Protection in Interior Impact |
| Standard No. 202 | Head Restraints |
| Standard No. 203 | Impact Protection for the Driver from the Steering Control System |
| Standard No. 204 | Steering Control Rearward Displacement |
| Standard No. 205 | Glazing Materials |
| Standard No. 206 | Door Locks and Door Retention Components |
| Standard No. 207 | Seating Systems |
| Standard No. 208 | Occupant Crash Protection |
| Standard No. 209 | Seat Belt Assemblies |
| Standard No. 210 | Seat Belt Assembly Anchorages |
| Standard No. 211 | [Reserved] |
| Standard No. 212 | Windshield Mounting |
| Standard No. 213 | Child Restraint Systems |
| Standard No. 214 | Side Impact Protection |
| Standard No. 216 | Roof Crush Resistance |
| Standard No. 217 | Bus Emergency Exits and Window Retention and Release |
| Standard No. 218 | Motorcycle Helmets |

| | |
|---|---|
| Standard No. 219 | Windshield Zone Intrusion |
| Standard No. 220 | School Bus Rollover Protection |
| Standard No. 221 | School Bus Body Joint Strength |
| Standard No. 222 | School Bus Passenger Seating and Crash Protection |
| Standard No. 223 | Rear Impact Guards |
| Standard No. 224 | Rear Impact Protection |

| | |
|---|---|
| Standard No. 301 | Fuel System Integrity |
| Standard No. 302 | Flammability of Interior Materials |
| Standard No. 303 | Fuel System Integrity of Compressed Natural Gas Vehicles |
| Standard No. 304 | Compressed Natural Gas Fuel Container Integrity |
| Standard No. 500 | Low Speed Vehicles |

### *1.1.2 Australian NCAP*

The Australasian New Car Assessment Program (ANCAP) which is conducted by the Australian and New Zealand governments provides detailed information on the protection level of vehicles in frontal and side impact accidents. The Australian NCAP testing method is very similar to the Euro NCAP and includes Frontal Offset, Side Impact, Pole, and Pedestrian tests (Euro NCAP 2012). Star ratings have been used by ANCAP to demonstrate vehicle safety for over 400 vehicles since 1993. The higher the ANCAP star rating, the better the vehicle performance is. The crash tests are carried out by independent specialist laboratories in a rating of 1–5. Test conditions are identical for all vehicles and a standard scientific scoring system is used in this process. To assess the protection level given to passengers dummies are used in all crash tests. Factors such as displacement of the dummies, occupant compartment structural impact, features such as electronic stability control (ESC), anti-lock braking system (ABS), and seat belt reminders are influential in this scoring system. ANCAP 5 Star rating is achieved when the highest recognized standard is obtained in all test categories (Australian NCAP 2012).

Dummy diagrams are used to demonstrate the risk of serious injury and death during the test. Brown and red results (Weak and Poor, respectively) on limb are suggestive of fatalities or a high risk of serious injury rather than an isolated injury to that body part (Fig. 1.1).

#### 1.1.2.1 ANCAP 1 Star

ANCAP 1 star rating indicates a very high risk of serious injury or death. In a vehicle with ANCAP 1 the level of protection provided is minimal and both driver and passenger are exposed to potential death, especially through injury to the head, with slightly better protection for the passenger (Fig. 1.2). The risk of fatalities or life-

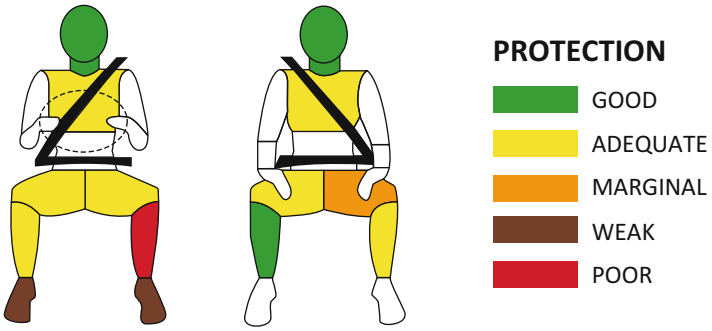



**Fig. 1.1** Readings from dummies in Australian NCAP Frontal Impact (Australian NCAP 2012)


**Fig. 1.2** Common injuries for the driver and passenger in ANCAP 1 Star (Australian NCAP 2012)

threatening injuries in an ANCAP 1 star rated vehicle is two times higher than an ANCAP 5 star rated vehicle (Australian NCAP 2012).

#### 1.1.2.2 ANCAP 2 Star

As demonstrated in Fig. 1.3, driver's head, right leg, and both feet are prone to serious and life-threatening injuries in an ANCAP 2 star rated vehicle (Australian

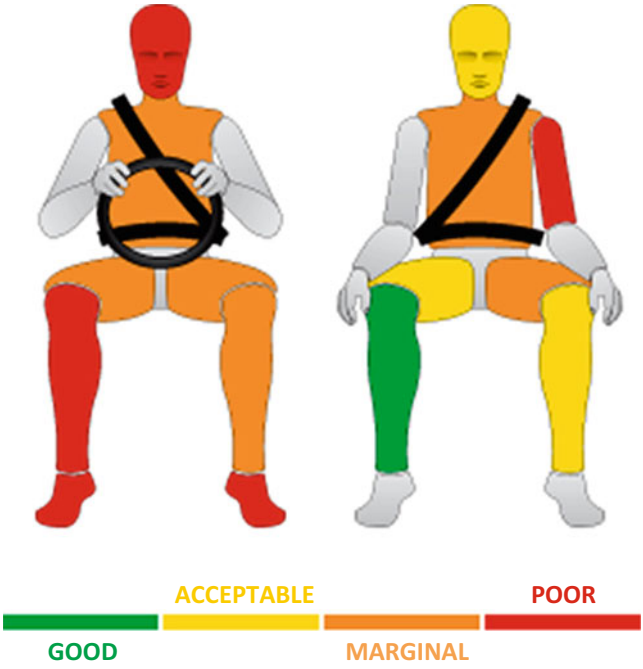



**Fig. 1.3** Common injuries for the driver and passenger in ANCAP 2 Star (Australian NCAP 2012)

NCAP 2012). Also areas in brown (Weak) shown on the driver's body are highly exposed to serious injury or death (Fig. 1.3).

#### 1.1.2.3 ANCAP 3 Star

Orange areas in the driver's chest, head, lower right leg, and upper legs indicate a limited risk of fatality or serious injury (Australian NCAP 2012). The passenger has a good level of protection in the left lower leg (green) and head and moderate protection in the trunk, right lower leg, and left upper leg (yellow) (Fig. 1.4).

#### 1.1.2.4 ANCAP 4 Star

As demonstrated in Fig. 1.5 there is a high level of head and neck protection for both occupants in ANCAP 4. This results in significant reduction of deadly injuries (Australian NCAP 2012). Orange areas demonstrate moderate risk and brown areas (driver left upper leg) show significant risk of damage (Fig. 1.5).

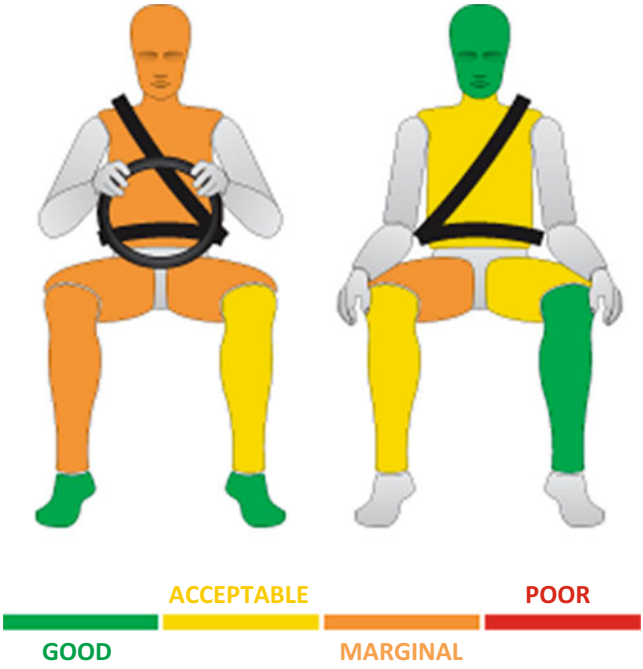



**Fig. 1.4** Common injuries for the driver and passenger in ANCAP 3 Star (Australian NCAP 2012)

#### 1.1.2.5 ANCAP 5 Star

In an ANCAP 5 the legs, body, and head have a high protection in both driver and passenger occupants (Australian NCAP 2012). The chest and lower legs of the passenger have acceptable protection and bear a minimal chance of serious injury (Fig. 1.6).

### *1.1.3 Euro NCAP*

The Euro NCAP testing includes four test conditions: frontal crash, side crash, pole test, and pedestrian crash test.

#### 1.1.3.1 Frontal Impact

As mentioned, frontal crash is the most common type of vehicle accidents. The frontal impact test was primarily designed by the European Enhanced Vehicle-safety Committee. This test is a simulation of the most common and fatal type

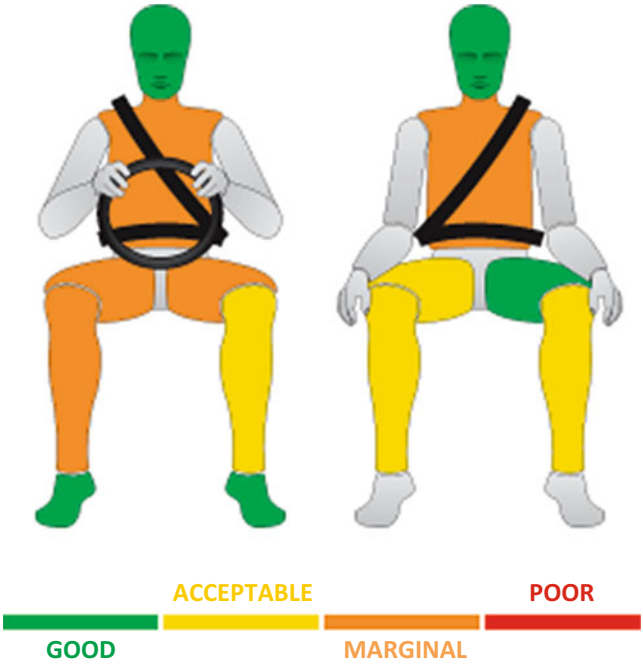



**Fig. 1.5** Common injuries for the driver and passenger in ANCAP 4 Star (Australian NCAP 2012)

of road crash. It challenges the car's ability to endure impact with no passenger compartment intrusion, as contact between the occupants and intruding parts of the passenger compartment primarily leads to serious and fatal injuries. As demonstrated in Fig. 1.7, the car strikes the deformable barrier with a certain speed. Over the years the impact speed has been increased from 56 km/h to 64 km/h (40 mph). This represents a vehicle to vehicle accident with a speed of approximately 55 km/h per each vehicle. The test is designed to be offset with only 40 % of the car being subjected to the impact, as most frontal crashes only involve a proportion of the vehicle front side. The barrier is also designed to be deformable for a better simulation of a car to car crash (Euro NCAP 2012).

The main purpose of all safety measures is to prevent and minimize the occupants' injuries and the car's interior impact. Generally safety measures have been directed at two major aspects, restraining occupants and minimizing the impact force.

Providing restraint for the occupants is one of the basic elements in damage control in vehicle safety studies. Different methods have been implemented to reach this goal. Steering wheel mounted airbags and seatbelts are two important restraints in modern vehicle design (Euro NCAP 2012).


**Fig. 1.6** Common injuries for the driver and passenger in ANCAP 5 Star (Australian NCAP 2012)

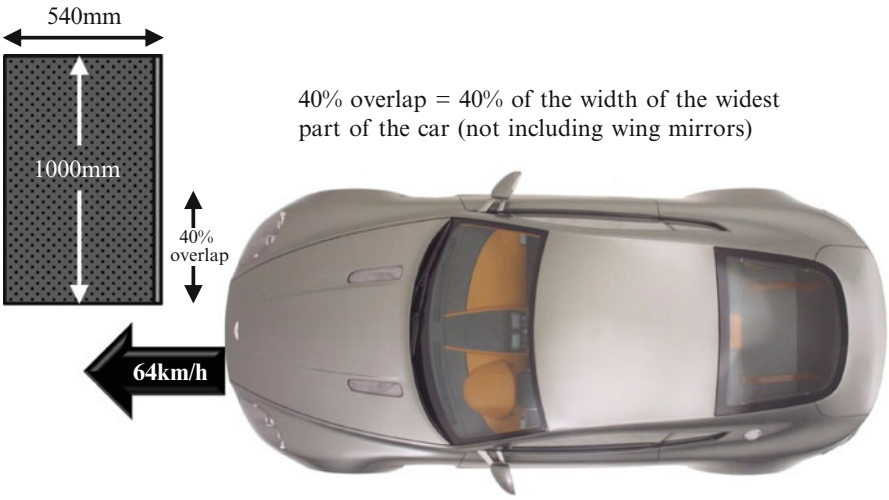



**Fig. 1.7** Euro NCAP frontal impact

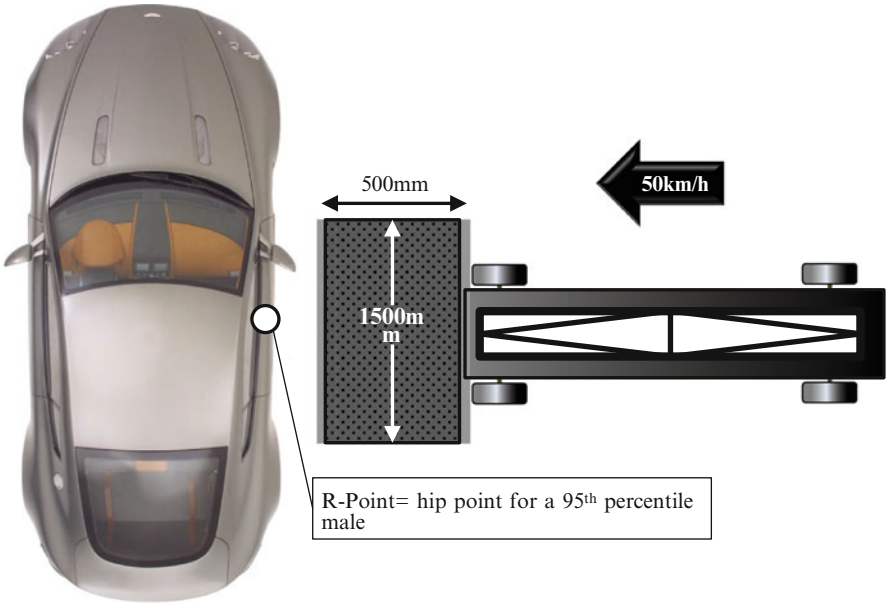



**Fig. 1.8** Car to car side impact in Euro NCAP

### 1.1.3.2 Side Impact and Pole Test

One of the major types of collision is car side impact. This can be simulated by collision of a mobile deformable barrier into the driver's door at 50 km/h (Euro NCAP 2012). A dummy is placed in the driver's seat and injury is evaluated (Fig. 1.8).

In the pole side impact test, the side of the car is directed at 29 km/h (18 mph) into a rigid pole. The pole is quite thin, so there is a pure penetration into the side of the car (Fig. 1.9). This test was not mandatory before 2009 and manufacturers were given the option of performing this test to demonstrate efficacy of fitted side air bags. Head protecting air bags have been demonstrated to significantly reduce fatal head injuries. Thus as of 2009 this test has become an Euro NCAP requirement (Euro NCAP 2012).

### 1.1.3.3 Pedestrian Impact

As part of the Euro NCAP, a series of tests are done to assess pedestrian injury in vehicle accidents. Impact conditions replicate motor vehicle accidents with the vehicle traveling at a speed of 40 km/h (25 mph) (Euro NCAP 2012). Due to complexity of predicting pedestrian response upon impact using a full dummy, individual component tests including leg form (Fig. 1.10a), upper leg form (Fig. 1.10b), and head form (Fig. 1.10c) tests are implemented.

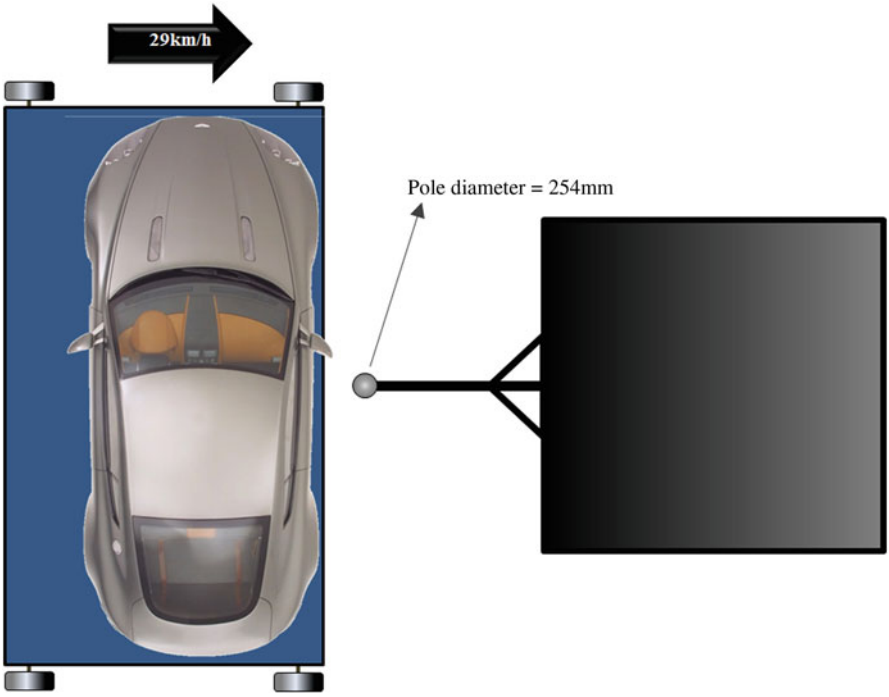



**Fig. 1.9** Pole side impact in Euro NCAP


**Fig. 1.10** Subsystem test for Pedestrian Protection in Euro NCAP (Euro NCAP 2012) (**a**) Leg form (**b**) Upper leg form (**c**) Head form

### *1.1.4 Structural Requirement*

Higher energy absorption and less energy transfer to the passengers are two fundamental elements in vehicle structural design. Vehicle structural design consists of chassis and body design. Structural energy absorbers should be placed in these parts to minimize damage. Longitudinal rails are one of the main elements which act as energy absorbers. Alterations in their design would result in major changes

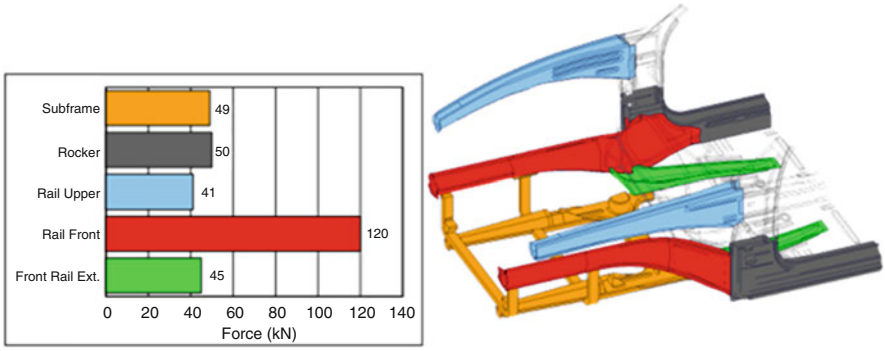



**Fig. 1.11** Front rails are the Main Energy Transfer Structure (A. I. a. S. Institute 1998)

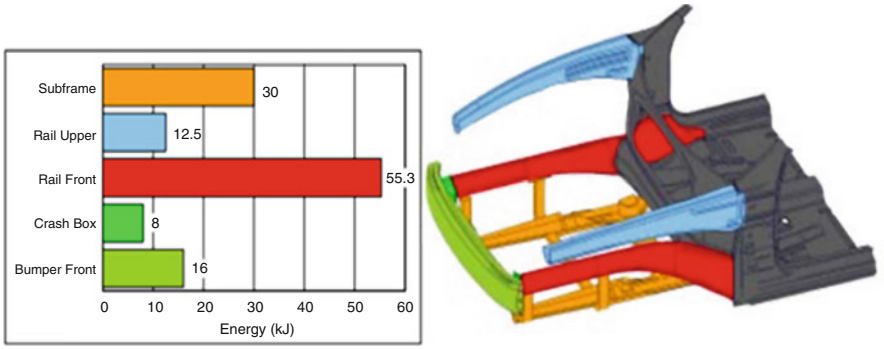



**Fig. 1.12** Front rails are the Main Energy Absorption Structure (A. I. a. S. Institute 1998)

in their functionality. The structures in the front end of the vehicle are involved in impact energy absorption and energy transfer to the occupiers depending on their structural design (A. I. a. S. Institute 1998).

As demonstrated in Figs. 1.11 and 1.12, the front rail (S-Rail) is considered the main energy load path. As previously mentioned S-Rail design and its effective parameters are indispensable factors in vehicle safety.

Energy absorption in the chassis happens through local buckling. To understand this process completely it is essential to be familiar with the parameters influencing local buckling. Plastic deformity should be designed to happen as far as possible from the passenger cabin and closest to the impact point. It is optimal for this deformity to happen on the top of the chassis which is considered the farthest point from the cabin. To reach this goal, soft metals are used in desirable buckling sites.

Energy transfer to the occupants is also an important parameter in side S-Rail design. Acceleration changes which result from energy transfer to the cabin could eventually lead to significant trauma to the passengers.

The frontal structures of a vehicle are described as soft, rigid, and hard (Fig. 1.13). Soft points or urban vulnerable points are able to absorb energy at

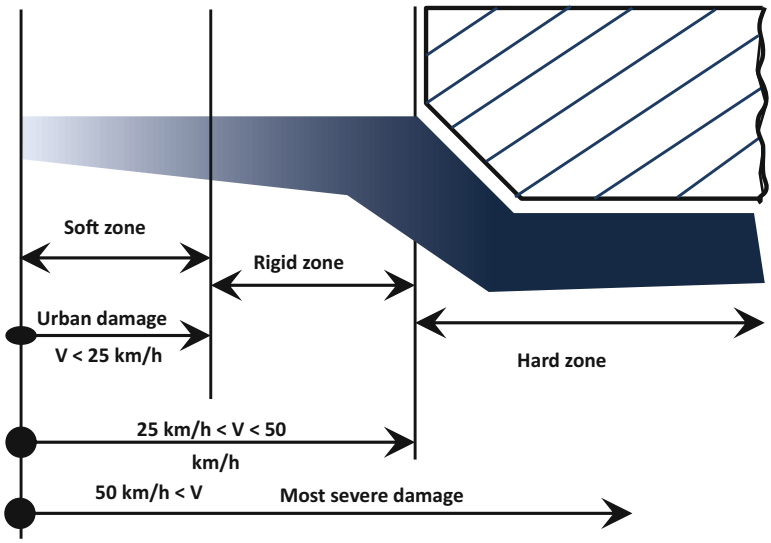



**Fig. 1.13** Car body front end structural zones

less than 15 km/h. The rigid points are resistant to impacts resulted at 25–50 km/h. Finally hard points are designed for maximum energy absorption in speeds higher than 50 km/h (Dieter 2000).

Structures such as the engine are considered as "rigid." In the case of a high velocity crash the engine might become displaced towards the passenger cabin. Ultimate design of the frontal part of the car should aim to maximize energy absorption and also to direct the engine beneath the cabin to reduce damage.

## 1.2 Background

Improvement of side rail structures has been of great interest to researchers in the past decades. Different strategies have been implemented to increase the level of absorbed energy in rails and consequently to decrease the energy transfer through rails. The main studies in this field include but are not limited to cross-section shapes, dimensions, and thickness. Also alterations in materials used in rails have been the subject of many studies. Using fillers such as foam, installing internal stiffeners, and also application of ribs in the S-rails have also been discussed by a number of authors. The available investigations can be categorized as constant section S-shaped frames, constant cross-section straight rails, and tapered straight rails studies. The major researches in the aforementioned categories are reviewed below.

### *1.2.1 Constant Cross-Section S-Shaped Frames*

As explained in the previous chapter, due to buckling in S-rails, the level of energy absorption in this category is not high. Among the studies which have been done to improve the energy absorption of S-rails, the following are of high importance due to their findings. Kim and Wierzbicki (2001a) studied the crash behavior of three dimensional thin-walled structures under combined loading. The authors proposed three S-frame cross-section types (Fig. 1.14) and two methods of internal reinforcement (diagonally positioned sheet metal stiffener and ultra-light metallic foam filler) aiming to decrease weight and maximize energy absorption. They concluded that the foam filled section and also diaphragm type 2 have the highest reinforcing effect. They also demonstrated that the diagonal diaphragms show higher degrees of reinforcement comparing to the parallel ones. In terms

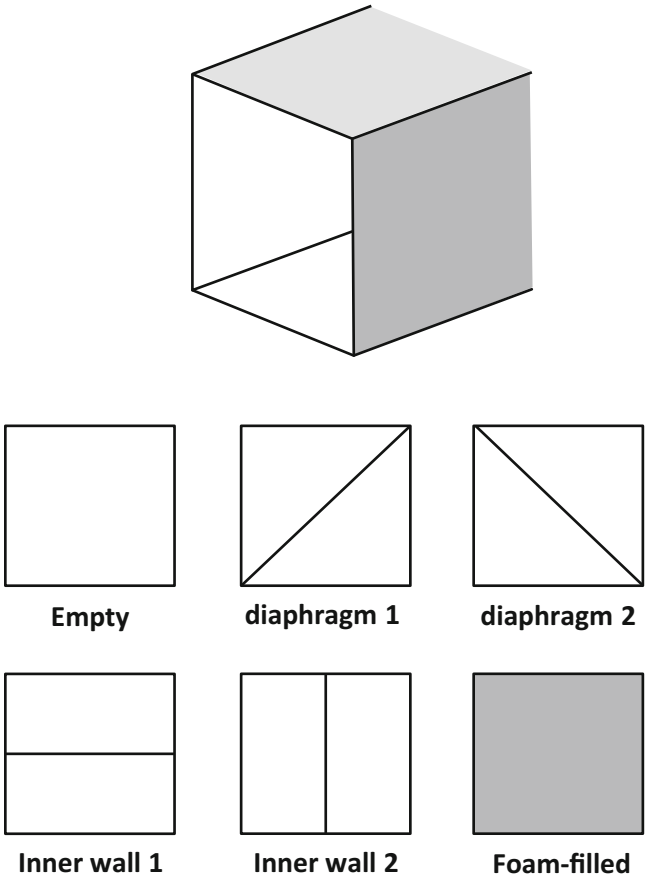



**Fig. 1.14** S-Rail cross-section types studied by Kim et al.

of collapse, type 1 showed lateral collapse while type 2 did not. This collapse pattern demonstrated the significant importance of the diagonal flange orientation plane of the S-rail. The foam filled model and diaphragm type 2 showed 300 % and 200 % more energy absorption than the original empty models, respectively. Specific energy absorption (SEA) was highest in the foam filled models followed by the diaphragm type 2.

Kim and Wierzbicki (2001a) showed that the ultra-light metallic foam filling technique is the most efficient reinforcement. The author also compared three different foam filling densities. He concluded that stronger foam resulted in more SEA, but this will increase the weight as well. So he proposed a partial foam filling technique (Fig. 1.15) to optimize the model. While the SEA for this model was significantly lower than the fully foam filled model, it was still higher than the empty structure.

As diaphragm type 2 showed the best strengthening effect and also prevented lateral collapse, yielding 100 % increase in SEA. The author proposed a new design for this structure to increase weight efficacy. Removing the front and rear ends of the structure, axial deformation would be induced in the front and end members (Fig. 1.16).This model resulted in 10 % decrease in the weight while increasing SEA by 7 % comparing to the original diaphragm 2 model (Kim and Wierzbicki 2001a).

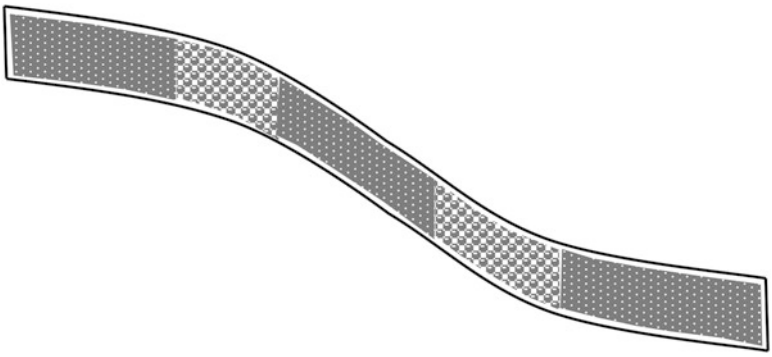


**Fig. 1.15** Partially foam filled S-rail (*shaded areas*)

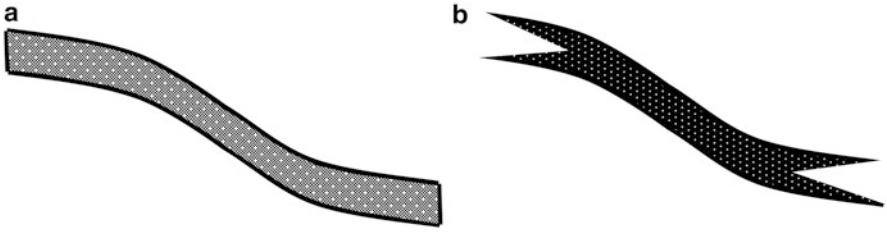



**Fig. 1.16** Diaphragm type two (**a**) original and (**b**) new design with removed front and rear end

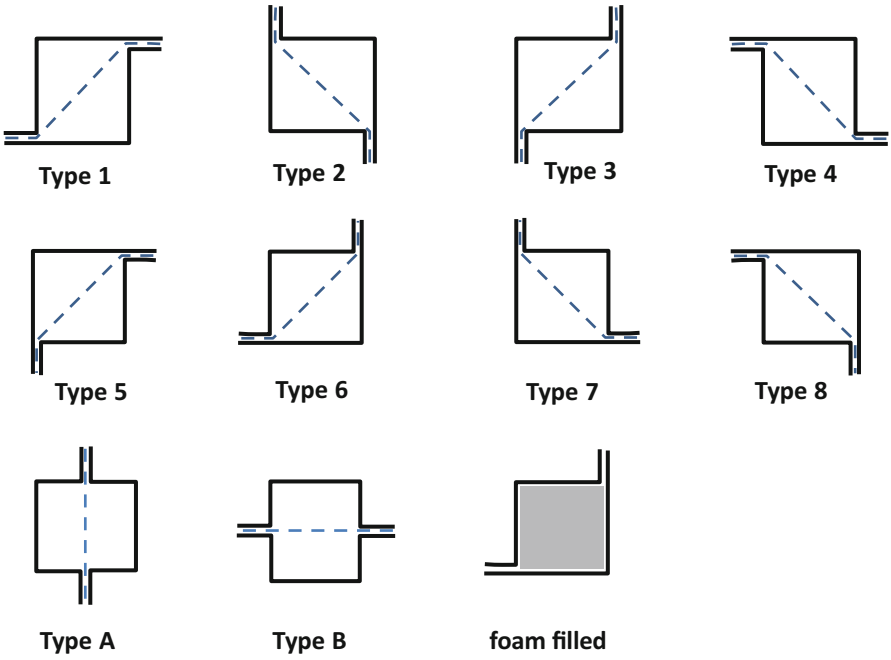



**Fig. 1.17** Studied hat-type cross-sections by Kim and Wierzbicki

Kim and Wierzbicki (2000, 2001a, b) carried out numerical and analytical studies on deep biaxial bending collapse of thin-walled beams (Kim and Wierzbicki 2000). Furthermore they studied the cross-sectional shape of hat-type cross-sections of S frames (Kim and Wierzbicki 2001b). They studied three section profiles. As shown in Fig. 1.17, type 1–4 had side web flanges which were parallel. Type 5–8 flanges were located at a right angle. Type A and B were described as "typical double-hat profiles." They studied these structures with and without internal stiffeners. Kim and Wierzbicki concluded that in the case of using no internal stiffeners, Type A and B show a higher force response than others. While using internal stiffeners the double-cell profile with diagonal inner reinforcement had the maximum energy absorption. Eventually they generated a new design of S frames with internal stiffeners considering that the cross-section Type 6 showed the best energy absorbing capacity.

As illustrated in Fig. 1.18 the front and rear ends of the stiffeners were removed in a modified model. This resulted in the induction of ongoing axial deformation in the front and rear ends. The middle part was preserved to avoid lateral sectional collapse.

Kim and Wierzbicki studied the crushing behavior of thin-walled rectangular cross-section S-rails subjected to two loading conditions (static and dynamic loading) later in 2004 (Kim and Wierzbicki 2004). In this research, they generated simplified calculations to explain the crushing behavior. Also they identified

**Fig. 1.18** Type 6 with new design of internal reinforcement and trigger

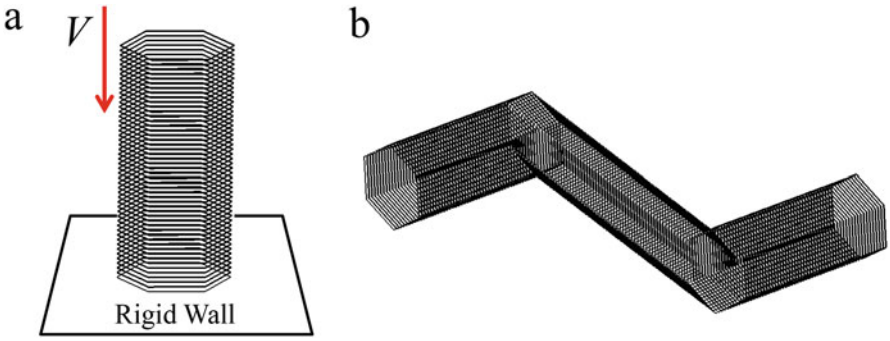



**Fig. 1.19** Straight octagonal (**a**) and curved hexagonal (**b**) columns studied by Liu

deformation modes and validated their results with the finite element output. They proposed that their calculations could be used in early phases of a vehicle body design.

Han and Yamazaki (2003) investigated the crushing pattern of S-rails with square shape sections with and without reinforcements (Han and Yamazaki 2003). They demonstrated that S shaped tubes without reinforcement buckle at the curvatures with no other deformation. In the case of having reinforcements, buckling happens in the curves and also middle parts. The authors demonstrated that using reinforcements and also shape perturbation increases energy absorption by 27 %.

Crashworthiness design of multi-corner thin-walled columns was studied in 2008 by Yucheng (2008). He studied thin-walled columns with straight octagonal columns and S-shaped hexagonal columns (Fig. 1.19). The design objective was to achieve the maximum SEA with the maximum crushing force as the design constraint. Factors such as cross-sectional width and wall thickness were selected as variables in this study. It was observed that in the straight multi-corner columns, increase in side length will reduce SEA and increases the peak crushing force. On the other hand, in the curved multi-corner columns, this was exactly the opposite.

Improvement of SUV S-frame was studied by Wu et al. in 2010 (Jun et al. 2010). Several rectangular shaped reinforcements were welded to the internal surface of S frame. Three designs (with 5, 6, and 7 reinforcements) were proposed and subjected to crash conditions (Fig. 1.20).

Increasing the number of reinforcements demonstrated more stiffness of the S frame and led to more energy absorption. Considering factors such as deformation

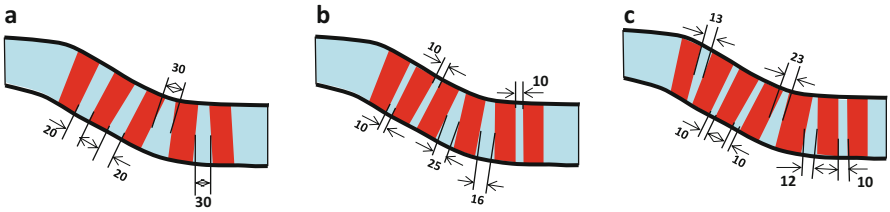



**Fig. 1.20** Five (**a**), six (**b**), and seven (**c**) reinforcement parts studied by Wu et al.

of the S beam, energy absorption of the longitudinal beam, crash acceleration, and passenger injury, the model with six internal reinforcements (Fig. 1.20b) was considered to be the optimal design.

Wu and Xin (2009) studied cross-section width, height, and the bending angle of the S-frame through finite element analyses to optimize energy absorption (Wu and Xin 2009). They demonstrated that reduction in width and curvature angle and increase in height results in an increase in SEA.

Crash behavior of a three dimensional S-shape structure was investigated in 2007 by Zhang and Saigal (2007). They demonstrated that although aluminium foam fillers are very good energy absorbers they do not increase SEA as they increase the structure weight significantly.

### *1.2.2 Constant Cross-Section Straight Rails*

Straight rails have been studied widely in the past decades. Many studies targeted variations in cross-sections, thickness, material, etc., to investigate the energy absorption characteristic of straight rails.

Chen and Wierzbicki studied the axial crushing of hollow multi-cell columns with and without foam fillings analytically and numerically (Chen and Wierzbicki 2001). They generated three different column designs (Fig. 1.21) and studied their crushing behavior with and without foam filling. The foam-filled columns showed higher SEA values. The gain in SEA was about 40 % for double-cell and triple-cell columns which was about 15 % more than what it was for single cell ones. All sections had similar SEA.

Shujuan et al. conducted a multi-objective optimization of multi-cell sections for the crashworthiness design to reduce peak crushing force and maximize SEA (Hou et al. 2008). Studying four different cross-section shapes they applied longitudinal impact to the columns (Fig. 1.22).

They used single objective, Pareto frontiers, and geometrical average multi-objective optimization methods. It was noted that increasing the number of cells results in a higher SEA. However no regularity was seen in the peak crushing force.

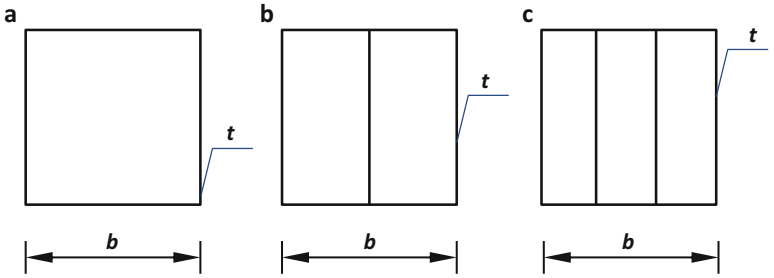



**Fig. 1.21** Single-cell (**a**), double-cell (**b**), and triple-cell (**c**) cross-sections studied by Chen and Wierzbicki

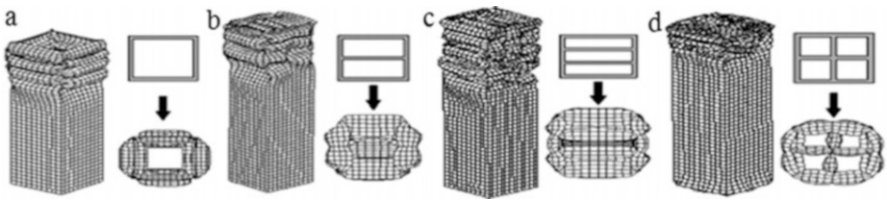



**Fig. 1.22** Deformation of columns in (**a**) single-cell, (**b**) double-cell, (**c**) triple-cell, and (**d**) quadruple-cell (Hou et al. 2008)

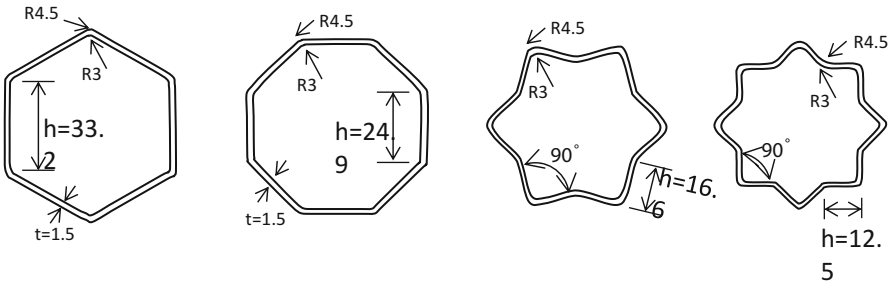



**Fig. 1.23** Multi-sided models studied by Fan et al.

Thin-walled tubes with different cross-sectional shapes were studied by Fan et al. (2013). Four shapes including hexagon, octagon, 12-sided, and 16-sided star were studied (Fig. 1.23). The results showed that increase in inward corners increased energy absorption. The 12-sided star shaped tube had the highest SEA. Increase in the number of corners up to 12, improved SEA while the 16-sided star shape design demonstrated a very poor SEA.

Song et al. studied the axial crushing of thin-walled structures with origami patterns (Song et al. 2012). In this study square, hexagonal, and octagonal cross-sections with origami patterns were analyzed. Patterned tubes demonstrated much lower initial peaks and therefore were more suitable energy absorbers.

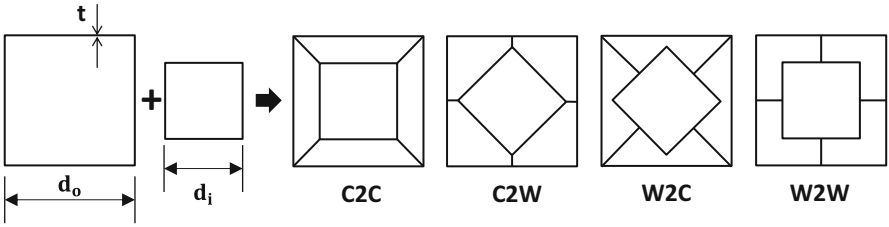



**Fig. 1.24** Multi-cell, multi-corner tube models studied by Najafi et al.

In a research done by Najafi and Rais-Rohani (2011), a static nonlinear finite element simulation study was done on the energy absorption characteristics of axially crushed thin-walled aluminium beams with different multi-cell, multi-corner structures (Fig. 1.24). The results indicated that multi-cell tubes had less crush force fluctuation. A complex deformation pattern in multi-cell, multi-corner configurations was noted. Configuration W2W and C2C had maximum and minimum SEA, respectively. The mean crush force for model C2W was very similar to W2C, but as C2W was significantly lighter than W2C, SEA and effective crush distance in C2W were higher than W2C.

A number of studies have been focused on using fillers. Santosa and Wierzbicki studied crash behavior of box columns filled with aluminium honeycomb or foam in 1998 (Santosa and Wierzbicki 1998). They showed that aluminium honeycomb filling is more weight efficient than aluminium foam filling. Aluminium foam filler demonstrated better crash behavior in combined loading. Also aluminium honeycomb had a higher strength to weight ratio compared to aluminium foam.

Also in another study by the aforementioned authors crush behavior of a closed cell aluminium foam tube was modelled (Santosa and Wierbicki 1998). Results were obtained from analytic and numerical methods which correlated. The authors studied foam-filled sections via experimental tests and numerical methods. Nonlinear dynamic finite element analysis was used to simulate test conditions. They carried out a thorough study of crush behavior of foam filled sections under axial loading. Sections were modelled as illustrated in Fig. 1.25. The crush behavior of these sections was also compared with empty columns (Santosa and Wierbicki 1998).

Experimental and numerical methods have been used by Sigit et al. to study the deformity pattern in both empty and foam filled columns in 2000 (Santosa et al. 2000). They demonstrated that for a foam filled tube there is a linear dependency between increasing the mean crushing force with the compressive resistance of the foam and cross-sectional area of the tube (Fig. 1.26).

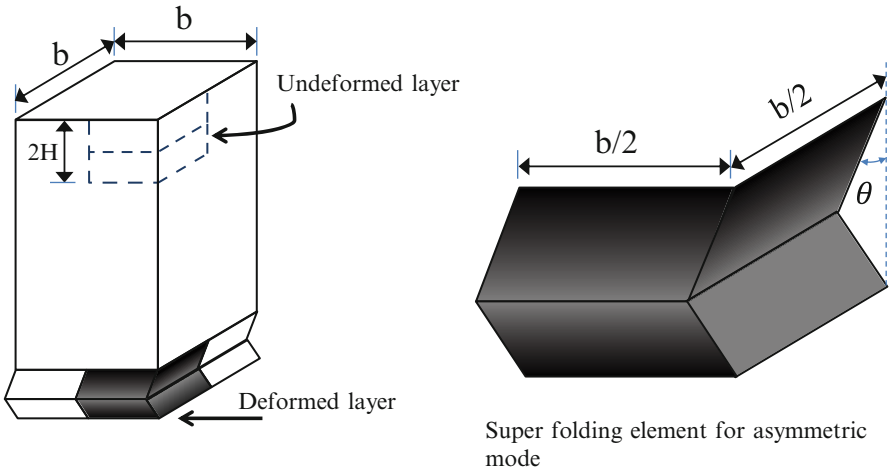



**Fig. 1.25** Geometric modelling of a foam-filled column

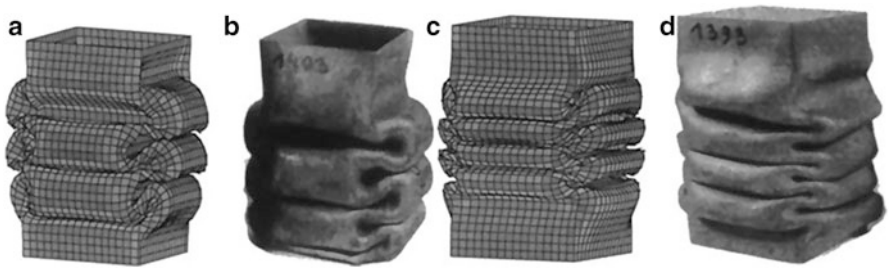



**Fig. 1.26** Simulation (**a**, **c**) and experimental (**b**, **d**) results of deformation pattern in empty (**a**, **b**) and foam filled (**c**, **d**) columns (Santosa et al. 2000)

### *1.2.3 Tapered Straight Rails*

Although never applied on 3D S-shape structures a series of structural stability studies have been done on thin wall tubes with variable cross-sections along the length. Reid and Reddy (1986) initially studied static and dynamic crushing behavior of tapered metal beams with rectangular cross-sections with an analysis based on the Abramowicz and Wierzbicki model (Wierzbicki and Abramowicz 1983). Single and double tapered tube behaviors were compared with tubes with uniform cross-sections. Figure 1.27 demonstrates the generated tubes. Tapered beams demonstrate more stability compared to straight tubes while subjected to axial loading. Approximately 20 % more crushing load was needed to achieve the same crushing impact in tapered tubes. Tapered tubes were also subjected to oblique impact. These structures demonstrated the same efficiency subjecting oblique impact as they were under axial loading. Reid and Reddy proposed that these

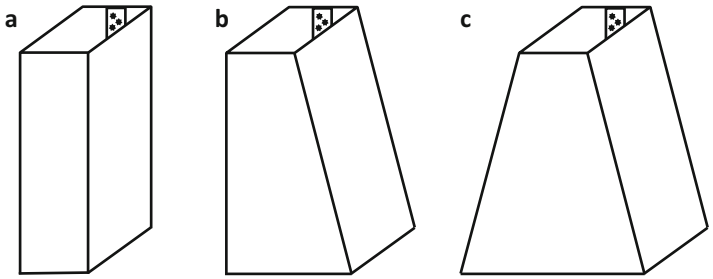



**Fig. 1.27** Straight (**a**), single-tapered (**b**), and double-tapered (**c**) tubes studied by Reid and Reddy

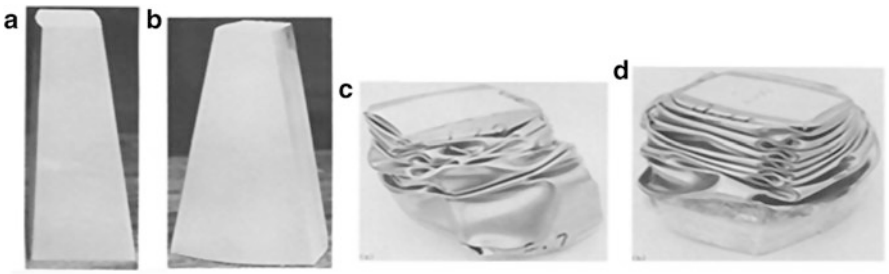



**Fig. 1.28** Single (**a**, **c**) and double (**b**, **d**) tapered tubes before and after compression (Reid et al. 1986)

structures show to be more efficient because they demonstrated an ideal constant mean load–deflection response subjecting dynamic loading. Their efficiency under axial pressure makes these structures suitable to use in the automotive industry as cars are subjected to both axial and oblique impacts in the event of crash.

Reid et al. (1986) studied axial crushing (dynamic and static loading) of foam filled tapered sheet metal tubes (Reid et al. 1986). The authors generated single and double tapered tubes and filled them with foam. The behavior of the structures before and after axial compression is illustrated in Fig. 1.28. Using foam fillers reduced load fluctuations and also increased energy absorption.

Karamanlidist and Jasti (1987) analyzed the crushing behavior of tapered beams through finite element (FE) modelling. They demonstrated that the model is an accurate and efficient mode of crush analysis in non-prismatic tapered beams (Karamanlidis and Jasti 1987).

Nagel and Thambiratnam (2004) conducted a mathematical investigation on the force response and absorbed energy of tapered beams via the FE method (Nagel and Thambiratnam 2004). They compared the energy absorption of straight and tapered thin-walled beams subjecting both static and dynamic axial loading. Variables were wall thickness, taper angle, impact mass, and impact velocity. The authors concluded that the energy absorption response of tapered rectangular tubes can be controlled dominantly by wall thickness variation. Also they showed that straight tube energy absorption response is not too sensitive to wall thickness variations.

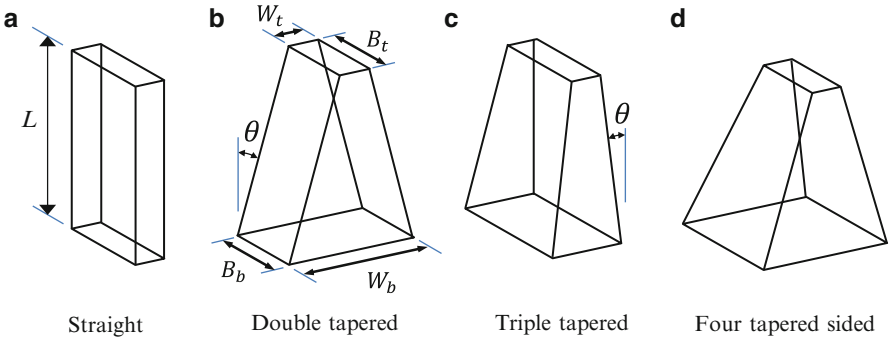



**Fig. 1.29** Straight and tapered tubes studied by Nagel and Thambiratnam

The authors compared the two structures in low and high impact velocities. In impact velocity of less than 30 m/s both structures went through progressive dynamic buckling with sequential folding forming at the impact end. When applying impact with velocity of 90 m/s buckling was formed at both ends of the straight tube. Although in the tapered tube, a wrinkle formed at the fixed end rather than buckling. To have buckling in the other end of the tapered beam, higher impact velocity was required. Also it was demonstrated that the larger the width of the tapered tube at the fixed end, the more the energy absorption.

Nagel and Thambiratnam also compared energy absorption of tapered and straight tubes with variable wall thickness, tapering angle, and tapering sides. Increase in the tapering angle increased energy absorption (Nagel and Thambiratnam 2005). They demonstrated that the energy absorption in tapered tubes was mostly influenced by wall thickness rather than tapering angle or taper sides. The triple-tapered tube showed the highest energy absorption capacity among others (Fig. 1.29).

The energy absorption and peak transferred load of tapered rectangular beams was studied in 2007. It was demonstrated that tapering angle and the length of the section have great influence on energy absorption (Shariatpanahi et al. 2008).

Qi et al. (2012) studied the crashworthiness of straight and tapered tubes with single-cell and multi-cell cross-sections. Multi-cell tapered tubes had the best crashworthiness under oblique loading (Fig. 1.30d).

### *1.2.4 Finite Element Modelling*

The finite element method which was introduced initially in aeronautical engineering is a numerical method for finding solutions to engineering problems. This method is based on sub-dividing a complex physical system into multiple simpler systems which are called finite elements. There are two types of FE analysis: linear and nonlinear. Linear analysis usually involves small displacements. Usually

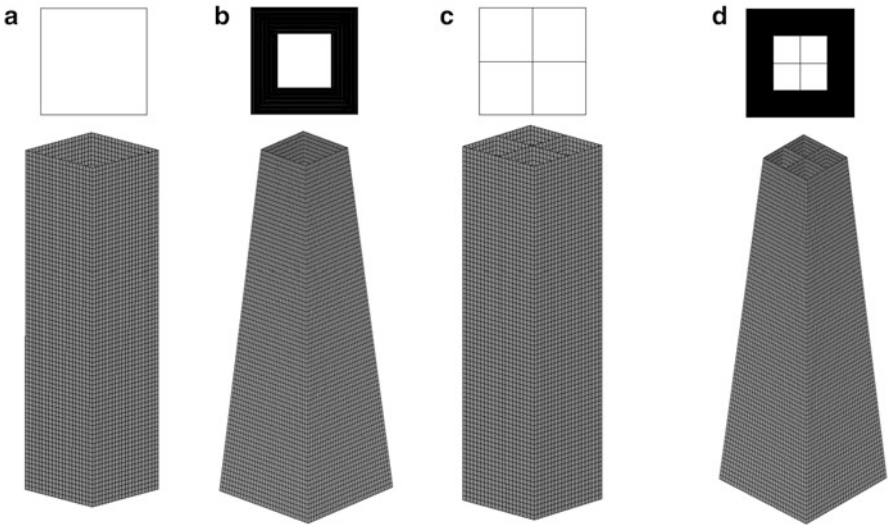



**Fig. 1.30** FE models of straight single-cell tube (**a**), tapered single-cell tube (**b**), straight multi-cell tube (**c**), and tapered multi-cell tube (**d**)

linear analysis is most suitable where the material has a linearly elastic behavior (Reddy 2006). Nonlinear FEM consists of material and geometric nonlinearity while geometric nonlinearity material may behave in a linear-elastic manner. Depending on the nature of the problem, linear FEM can be quite good as nonlinear FEM. Due to the nonlinearity and the complexity of the crash mechanism, nonlinear FEM is more suitable to be used.

### *1.2.5 The Aim of the Present Work*

Although many studies have been directed towards straight rail and S-shaped rail crushing behavior, no study has investigated the effect of tapering on S-rails crashworthiness. This study aims to investigate the impact of tapering on crashworthiness of S-rails as the main load path and energy absorbers in frontal vehicle crash. Two types of S-rails, one with and one without internal diagonal reinforcement are tapered with 20 different tapering ratios. Finite element analysis is performed to simulate crash conditions in a total of 42 models. All models are subjected to static and dynamic loading conditions. The FE results are validated through analytical analysis. The effectiveness of tapering S-rails is assessed through investigating the energy absorption and SEA variations in different models.

## 1.3 Analytical Analysis of Crush Mechanism

### *1.3.1 Energy Absorption*

To gain a thorough understanding of crush mechanisms one should be familiar with energy absorbing mechanisms. Impact energy dissipates in plastic deformation of metallic structures in frontal crash. Energy absorbers are designed to optimally absorb all the impact energy within them, thus increasing the safety of the passengers. Other than the absorbers themselves, factors such as transmission rates, magnitude of trauma, and deformation patterns are effective in energy conversion (Johnson and Reid 1978). Structures such as steel drums (Carney and Pothen 1988), circular tubes (Alexander 1960), tubular rings (Reid et al. 1984), square tubes (Langseth and Hopperstand 1996), and corrugated tubes (Singace and El-Sobky 1997) have been studied as deformable energy absorbers. Multiple studies are done on these elements in different settings such as filled with foam (Reddy and Wall 1988), liquid (Nahas 1993), and wood shavings (Reddy and Al-Hassani 1993). Axial and lateral crushing of tubes, tube inversion, and tube nosing have been previously discussed (Fan et al. 2013; Kotełko et al. 2009; Harte et al. 2000). Each of the aforementioned systems has its own special features. Collapse mechanisms in the case of crash are very complicated to describe and a good understanding can only be achieved through the study of energy absorber performance in respond to impulsive loads. Axially loaded structures undergo deformation via geometrical changes in various forms such as concertina (axisymmetric) and the diamond (non-axisymmetric) modes of collapse. There are many different types of energy absorbers described in literature such as thin-walled tubes, frusta, multi-corner columns, struts, sandwich plates, honeycomb cells. Structural components of the vehicle body can also act as energy absorbers. As explained previously, side rails absorb the largest amount of impact energy in frontal crash. Thus the focus of this thesis will be on methods for strengthening S-frames.

### *1.3.2 Crush Mechanism*

In compression cylindrical structures undergo a minimum of four different mechanisms of deformation: (1) tube inversion, (2) progressive crushing, (3) axisymmetric buckling, and (4) diamond-shape buckling as demonstrated in Fig. 1.31 (Al-Hassani et al. 1972; Abramowicz 1983).

In the case of frontal collision collapse of metallic toolbars occurs. Cylindrical shell collapse subjecting axial load was initially described by Alexander (1960). In 1983 Abramowicz completed the collapse model that Alexander had previously proposed (Abramowicz 1983). Wierzbicki and Abramowicz (1983) were the first to propose an analytic solution for axial crushing of thin-walled prismatic box columns. This model followed the same trend as the proposed model for cylindrical

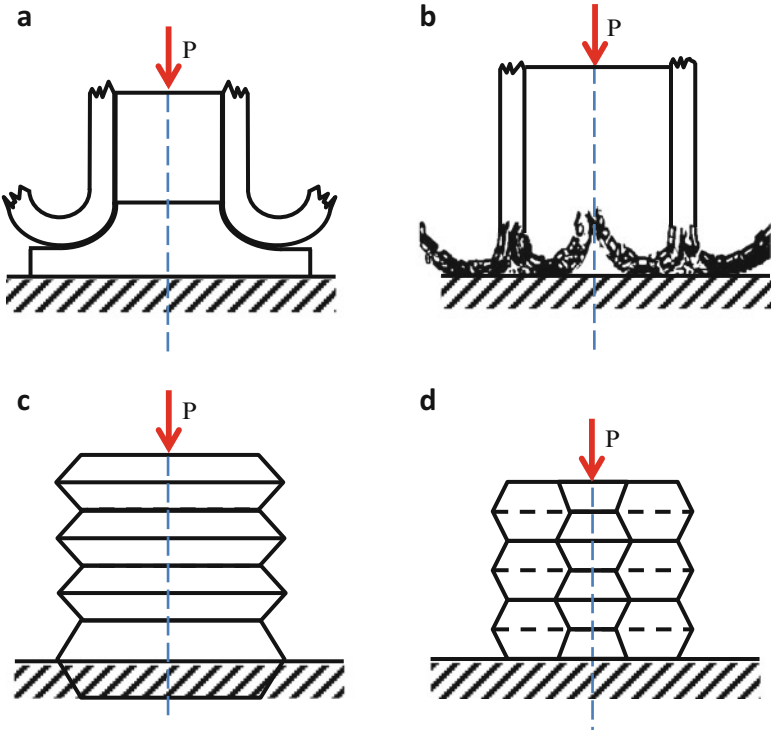



**Fig. 1.31** Axial modes of collapse for cylinders in (**a**) tube inversion, (**b**) progressive crush, (**c**) axisymmetric buckle, (**d**) diamond buckling

structures. They subsequently extended this model to multi-corner prismatic membranes (Abramowicz and Wierzbicki 1989). A model for calculation of local distribution of stress at a cross-section of the tube under axial crushing was also proposed by Wierzbicki et al. (1994). A crush analysis of thin-walled tubes was done by Abramowicz (1981). Kecman studied deep bending collapse in closed sections through a kinematic approach to plasticity and introduced a proper folding mechanism via plastic hinges. Actually this was the first experimental study on deep bending collapse in prismatic membranes (Kecman 1983). The aforementioned models are explained briefly below.

#### 1.3.2.1 The Alexander Model

The Alexander model which explains cylindrical shell collapse under axial loading was initially described to design more efficient energy absorbers in unintentional reducing of elements in nuclear reactor fuel channels. Alexander (1960) assumed the collapse as a "concertina" form with straight sided complexities resulting in a symmetrical crush. He also eliminated elastic strains from the study and considered the shell as "plastic-rigid" material (Fig. 1.32).

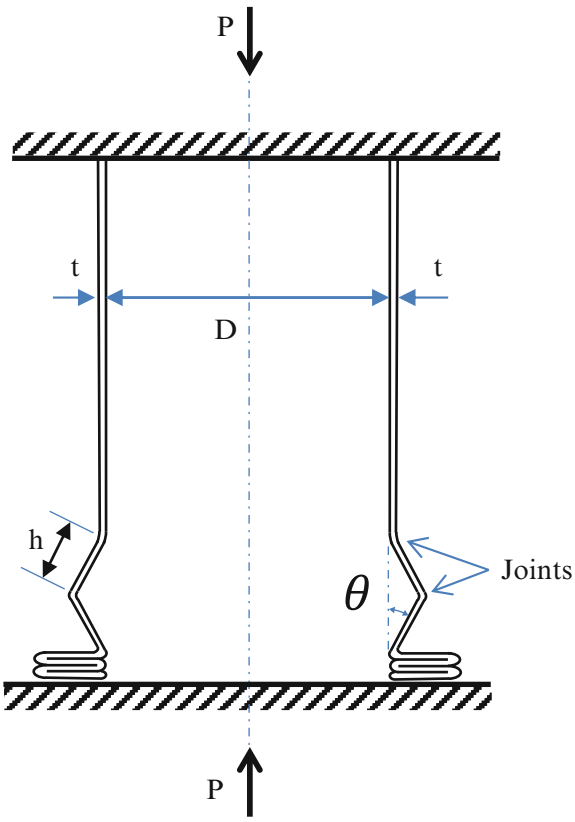



**Fig. 1.32** Collapse mode in the Alexander model

The amount of work done to increase angle $\theta$ by $d\theta$ at the three joints is calculated as:

$$dW_1 = 4\,M\,d\theta\,\pi\;\;(D + h\;\sin\theta) \tag{1.1}$$

where $M$ is considered as collapse moment per unit circumferential length of the joint.

Assuming the material follows the Von Mises criterion of yielding $M$ would be

$$M = \left(\frac{2}{\sqrt{3}}\right)\left(\frac{Yt^2}{4}\right) \tag{1.2}$$

where $Y$ is the yield stress in uniaxial tension or compression. So,

$$dW_1 = \frac{2\pi}{\sqrt{3}}\,Y\,t^2 d\theta\,(D + h\;\sin\;\theta) \tag{1.3}$$

As Fig. 1.32 is demonstrating the convolutions are formed towards the exterior. The mean strain during the incremental changes in the joints is calculated as:

$$\frac{\pi\,[D + h\ \sin(\theta + d\theta)] — \pi\,[D + h\ \sin\theta]}{\pi\,[D + h\ \sin\theta]} = \frac{h\,d\theta\ \cos\ \theta}{D + h\ \sin\theta} \tag{1.4}$$

in which $Y$ is the yield stress of the metal hence the work to stretch the metal would be:

$$dW_1 = \frac{Yhd\theta\ \cos\ \theta}{D + h\ \sin\ \theta}\pi\ \ (D + h\ \sin\ \theta)\,.2ht = 2\pi Yh^2td\theta\ \cos\ \theta \tag{1.5}$$

So the total work to collapse the shell for 90 degrees incremental change will be:

$$W = \int (dW_1 + dW_2) = \int_0^{\frac{1}{2\pi}} \left[\frac{2\pi}{\sqrt{3}}Yt^2\,(D + h\ \sin\ \theta) + 2\pi\ Y\,h^2t\ \cos\ \theta\right]d\theta \tag{1.6}$$

This shall be equal to the mean axial load $\overline{P}$ multiplied by the total displacement $2\,h$ (the metal thickness is neglected).

$$\frac{\overline{P}}{Y} = \frac{\pi t^2}{\sqrt{3}}\left(\frac{\pi\ D}{2h} + 1\right) + \pi ht \tag{1.7}$$

$h$ is then calculated, by minimizing the equation gives

$$h = \sqrt{\left(\frac{\pi}{2\sqrt{3}}\right)}\sqrt{Dt} = 0.953\ \sqrt{Dt} \tag{1.8}$$

Comparing to the $h$ calculated as by Timoshenko and Gere (1961):

$$h = 1.213\sqrt{Dt} \tag{1.9}$$

Substituting $h \cong k\ \sqrt{Dt}$ in Eq. (1.9),

$$\frac{\overline{P}}{Y} = \left(\frac{\pi^2}{2\sqrt{3}k} + \pi k\right)t^{1.5}\sqrt{D} + \frac{\pi t^2}{\sqrt{3}} \tag{1.10}$$

It is then assumed that the convolutions will be shaped internally and not externally. Repeating the analysis for this case results in:

$$\frac{\overline{P}}{Y} = \left(\frac{\pi^2}{2\sqrt{3}k} + \pi k\right)t^{1.5}\sqrt{D} - \frac{\pi t^2}{\sqrt{3}} \tag{1.11}$$

Knowing the fact that the analysis is approximate and considering that the actual deformation mode is between the two cases, it looks ok to assume the mean value between Eqs. (1.10) and (1.11).

$$\frac{\overline{P}}{Y} = \left(\frac{\pi^2}{2\sqrt{3k}} + \pi k\right) t^{1.5}\sqrt{D} \tag{1.12}$$

By replacing the $k = 0.953$ from Eq. (1.8) gives:

$$\frac{\overline{P}}{Y} = 5.99 t^{1.5}\sqrt{D} \tag{1.13}$$

and replacing $k = 1.213$ from Eq. (1.9)

$$\frac{\overline{P}}{Y} = 6.16 t^{1.5}\sqrt{D} \tag{1.14}$$

Then the final solution will be

$$\overline{P} = KY t^{1.5}\sqrt{D} \tag{1.15}$$

where $K \cong 6.08$, $Y =$ yield strength of the material, $t =$ thickness of the cylinder, and $D =$ mean diameter.

#### 1.3.2.2 The Abramowicz Model

Abramowicz studied the effective crushing distance of a thin wall structure while taking the hardening property into account. He demonstrated that the work needed for a thin wall cylindrical structure to collapse under axial pressure with an effective crushing distance of $\sigma_e$ will be equal to $P_m \delta_e$. Studying crush in a non-elastic column with the length of $2\lambda$ the following results were obtained (Abramowicz 1983):

$$X_m = 0.28\, L \tag{1.16}$$

He demonstrated that in complete collapse Fig. 1.33 would be closer to reality than Fig. 1.32 which Alexander proposed. According to Fig. 1.16, the effective crushing distance will be:

$$\delta_e = 2\lambda - 2X_m - H \tag{1.17}$$

As Alexander demonstrated:

$$\lambda = \left(\frac{\pi RH}{\sqrt{3}}\right) 1/2 \tag{1.18}$$

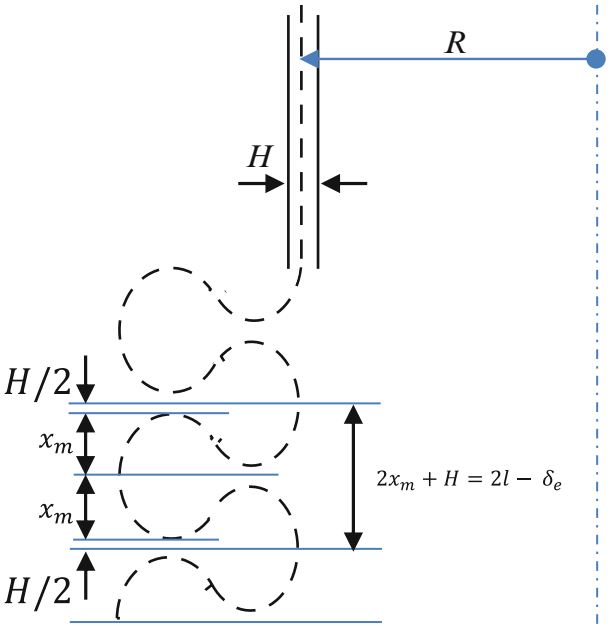



**Fig. 1.33** Collapse prediction in Abramowicz Model

Substituting Eqs. (1.15) and (1.17) in to Eq. (1.16):

$$\delta_e = 1072L - H$$

$$\frac{\delta_e}{2L} = 0.86 - \frac{H}{2L} \tag{1.19}$$

That would be:

$$\frac{\delta_e}{2L} = 0.86 - 0.37\left(\frac{H}{R}\right)^{1/2} \tag{1.20}$$

which results in a more accurate equation for crushing force:

$$P_m = 2(\pi H)^{3/2} R^{1/2} Q_0 / \left\{3^{1/2}\left[0.86 - 0.37\left(\frac{H}{R}\right)^{1/2}\right]\right\} \tag{1.21}$$

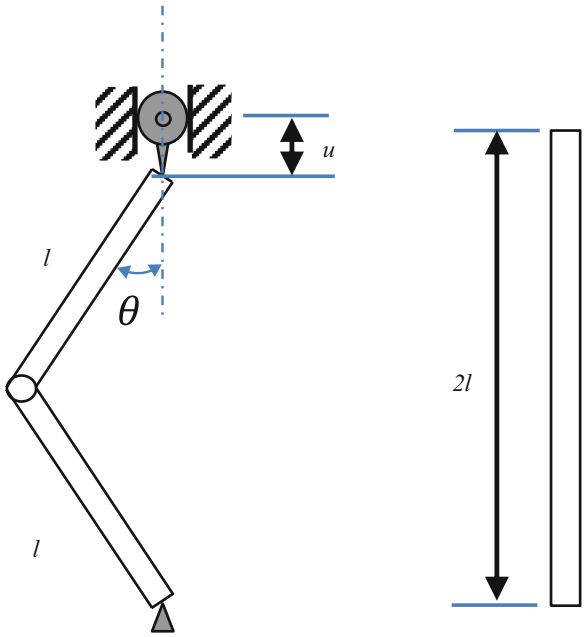



**Fig. 1.34** Pin–pin supported model

### 1.3.2.3 Kim and Wierzbicki Model for S-Shaped Frame

Kim and Wierzbicki, being pioneers in rail crush analysis, have developed the fundamentals of mathematical analysis of the S-rail (Kim and Wierzbicki 2004). The authors studied a three dimensional S-shaped tube with rectangular cross-sections subjected to crash load resulting in biaxial bend and combination of compression and bending. Force response $P$ was calculated in three different conditions: Constant bending moment, decaying function of bending moment, and combined load of bending and compression.

Initially a simplified model compromising of a pin–pin support prismatic column with a plastic hinge in the middle was developed by Kim and Wierzbicki (Fig. 1.34).

In a constant bending moment, $u$ was defined based on the model geometry and rotation angles

$$u = 2l - 2l\ \cos\ \theta \cong 2l\ \frac{\theta^2}{2} = l\theta^2 \tag{1.22}$$

hence,

$$\theta \cong \sqrt{\frac{u}{l}} \tag{1.23}$$

The authors then went on to define the force response $P$. The constant resisting moment was considered equal the fully plastic bending moment $M_0$ of the cross-section and hence $P$ was defined as:

$$P = \frac{M_0}{\sqrt{ul}} \tag{1.24}$$

To calculate $P$, decaying function of bending moment $M_0$ was substituted with the actual decaying function of bending moment response (Santosa et al. 2000),

$$P = \frac{11.03\,\sigma_0 b^{4/3} t^{5/3}\left(0.576 + l^{1/4}/\,2u^{1/4}\right)}{\sqrt{lu}} \tag{1.25}$$

Equations (1.24) and (1.25) did not consider the interaction between the compression and bending. $P$ was calculated in combined load of bending and compression conditions, using the normalized interaction curve of bending and compression in deep plastic deformation range which was defined earlier (Al-Hassani et al. 1972)

$$P = \frac{M_0 C_0}{\sqrt[4]{l}} \times \frac{1}{u^{3/4}} \tag{1.26}$$

where,

$$C_0 = 3.4342\left(\frac{t}{b}\right)^{5/6} \tag{1.27}$$

$t$ being the cross-section thickness and $b$ the cross-section height.

The aforementioned calculations apply to a simplified model. To define crash behavior in a three dimensional S-shaped frame, a two dimensional S-shaped model (Planar model) was also studied. As demonstrated in Fig. 1.35 the frame consisted of three plastic hinges.

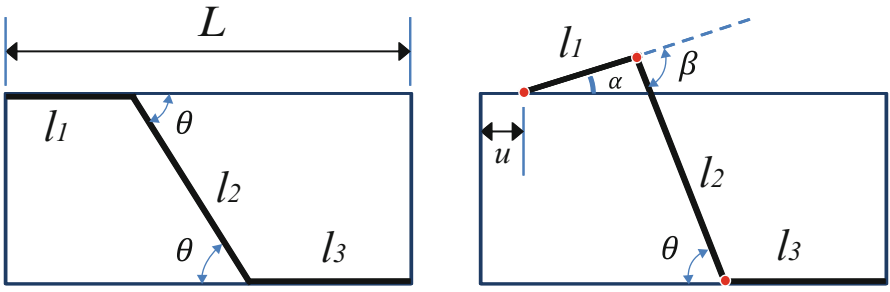



**Fig. 1.35** Planar S-shaped frame

To calculate the peak force more accurately, the beam was divided into three segments and the equilibrium for force and moment for each segment was defined. A global equilibrium was then expressed and peak crushing force was calculated as:

$$P_{\text{peak}} = \min_{i=1,2,a,b,c,d} \left\{ \frac{N_0 M_0 N_i}{\alpha_i \left( M_0 N_i + N_0 M_i \right)} \right\} \tag{1.28}$$

Further on, force response was extracted as a function of $u$:

$$P = \frac{M_0 C_0}{\sqrt{Au+B}} \left( \begin{array}{l} \frac{\sqrt{l_1 (l_1+l_2)}}{\left(-l_1 l_2 \ \sin\ \theta + l_2 \sqrt{Au+B}\right)^{1/2}} + \frac{(l_1+l_2)\sqrt{l_1}}{l_2 \left(\sqrt{Au+B} - l_1 \ \sin\ \theta\right)^{1/2}} \\ + \frac{l_1 \sqrt{l_1+l_2}}{l_2 \left(-l_1 \ \sin\ \theta + \sqrt{Au+B}\right)^{1/2}} \end{array} \right) \tag{1.29}$$

where

$$A = \frac{2 l_1 \left(l_1 + l_2\right)}{l_2} \tag{1.30}$$

$$B = l_1^2 \sin^2\theta - \frac{2 l_1 \left(l_1 + l_2\right)}{l_2} \left( L - l_1 - l_2 + \frac{l_1}{l_2} \sin^2\theta - l_3 \right) \tag{1.31}$$

$$M_0 = \frac{3}{2} b^2 t \sigma_0 \tag{1.32}$$

and $C_0$ can be calculated from Eq. (1.27).

To be able to define the bending moment in three dimensional structures it is essential to determine the bending axis. Figure 1.36 demonstrates the orientation angle of bending axis $\psi$ and its relation with the cross-section.

The fully plastic bending moment $M_{\text{f}}$ was calculated as:

$$M_{\text{f}} = \sigma_0 t \ \cos\ \psi \left( \frac{a^2}{2} + \frac{b^2}{2} \tan^2\psi + ab \right) \quad \left\{ 0 \le \psi < 90^\circ - \tan^{-1}\left(\frac{b}{a}\right) \right\} \tag{1.33}$$

$$M_{\text{f}} = \sigma_0 t \ \sin\ \psi \left( \frac{a^2}{2} \cot^2\psi + \frac{b^2}{2} + ab \right) \quad \left\{ 90^\circ - \tan^{-1}\left(\frac{b}{a}\right) \le \psi \le 90^\circ \right\} \tag{1.34}$$

When the orientation angle of minimum $M_{\text{f}}$ is calculated as:

$$\frac{dM_{\text{f}}}{d\psi} = 0 \tag{1.35}$$

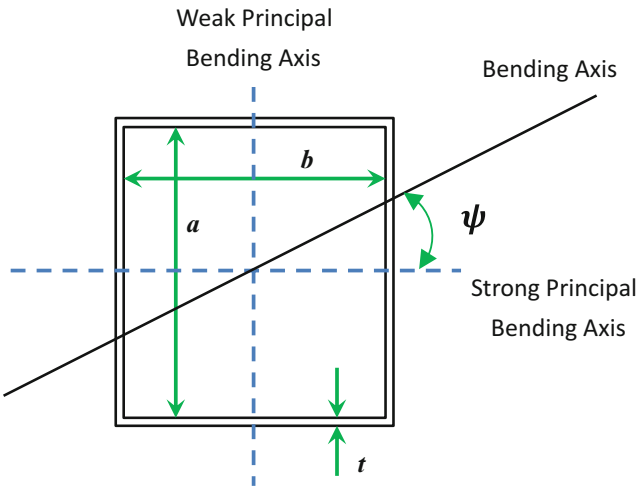
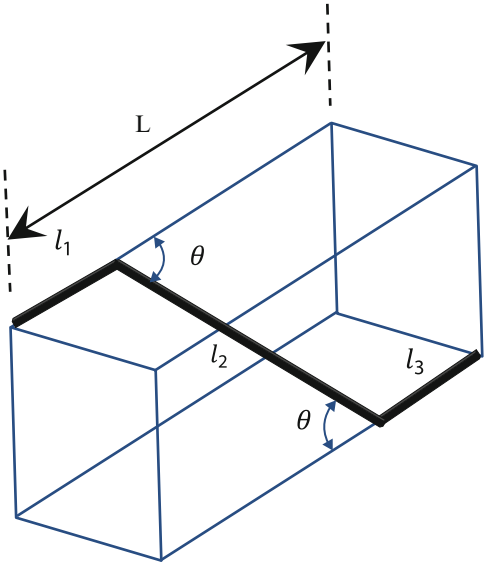



**Fig. 1.36** Orientation angle of bending axis in the frame section

**Fig. 1.37** Three dimensional S-shaped frame




Bending axis $\psi$ can be calculated from:

$$\sin\psi = \sqrt{\frac{(a/b)^2}{1+2\,(a/b)-(a/b)^2}}. \tag{1.36}$$

Finally, for an S-frame with a cross-section aspect ratio less than 1.366, the crushing response of a three dimensional S-shaped frame (Fig. 1.37) can be calculated by replacing $M_0$ in Eq. (1.29) with Min ($M_f$ ($\psi$)) by Eqs. (1.33), (1.34):

$$P = \frac{M_f C_0}{\sqrt{Au + B}} \left( \begin{array}{l} \frac{\sqrt{l_1(l_1+l_2)}}{\left(-l_1 l_2 \sin\theta + l_2\sqrt{Au+B}\right)^{1/2}} + \frac{(l_1+l_2)\sqrt{l_1}}{l_2\left(\sqrt{Au+B} - l_1 \sin\theta\right)^{1/2}} \\ + \frac{l_1\sqrt{l_1+l_2}}{l_2\left(-l_1 \sin\theta + \sqrt{Au+B}\right)^{1/2}} \end{array} \right) \tag{1.37}$$

where $C_0$, $A$, $B$, $M_f$ can be calculated from Eqs. (1.27), (1.30), (1.31), and (1.32).

Using Eq. (1.37), the response force $P$ of an S-frame for a displacement of $u$ can be predicted.

## 1.4 Formulation and Finite Element Modelling

The structure of three dimensional S-rails is illustrated in Fig. 1.37. Due to the horizontal and vertical eccentricities, when an external force is applied, the S-rails are subject to complex bending and combined load of compression and bending. In this chapter the finite element method is used to study the compound crushing mode. In the next chapter, numerical analysis is provided to validate the finite element results.

To formulate the problem the following steps are followed:

- Designing a novel square section S-rail with and without internal lateral reinforcement
- Creating the models geometry using CATIA V5
- Meshing the created geometries with HyperMesh
- Developing the finite element model in LS-PrePost
- Models solving by LS-DYNA solver

To create the finite element models, the two base model S-rails (one with and one without internal reinforcement) are created using CATIA 5. Forty tapered S-rails are developed from the two base models, all with the same central profile line and with 10 % tapering increments. All models are then imported to HyperMesh for meshing. Nonlinear finite element code LS-PrePost is used to develop the finite element models and to define all required parameters including material properties, element features, boundary, and loading conditions, etc. The numerical solutions are extracted using LSDYNA solver. The simulation results are then viewed via LS-PrePost, HyperView, and HyperGraph software.

### *1.4.1 Geometry Modelling*

The studied structures in this research are thin-walled S-shaped rails with square cross-sections. Two base models are defined: one with a normal square cross-section (Type A) and the other with a square cross-section including a diagonal lateral reinforcement (Type B) as illustrated in Fig. 1.38.

The base models are 1000 mm (X length) S-rails with two straight ends of 350 mm and bending angle of $\theta = 43.3°$ (Fig. 1.38). The cross-section aspect ratio

(width/thickness) is 50 (100/2 = 50) in both types A and B where width ($a$) is 100 mm and thickness ($t$) is 2.0 mm.

In order to investigate the effect of different tapering ratios on crushing behavior of S-Rails, a series of models with different tapering ratios are developed. Each type A and type B S-rails are tapered with 20 different tapering ratios ($R_t = av/a$) from 1.1 to 3.0 with 0.1 increments (Fig. 1.39).

The profile center line is identical for all models. In total, 42 different geometries are developed and coded as shown below in Table 1.1.

### *1.4.2 Finite Element Modelling*

The generated models are meshed using HyperMesh (Altair HyperWorks) software. The whole structure is modelled using 4-node *Quads only* shell element. If any other mesh type is automatically created during meshing, all must be manually updated

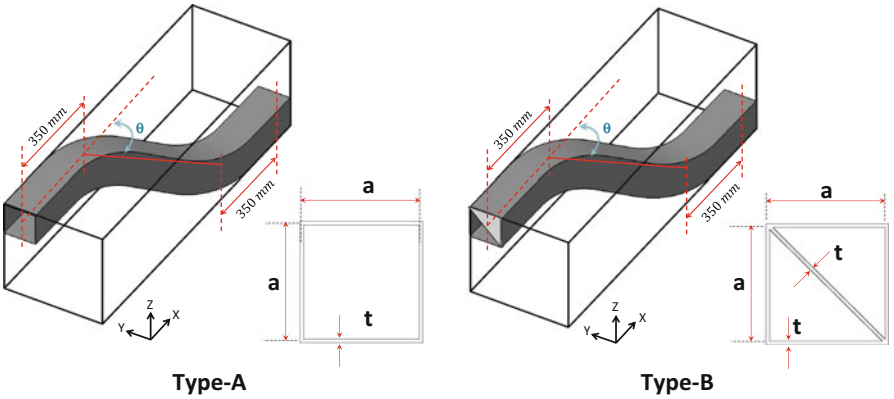



**Fig. 1.38** Base models S-rails type A (without reinforcement) and type B (with reinforcement)

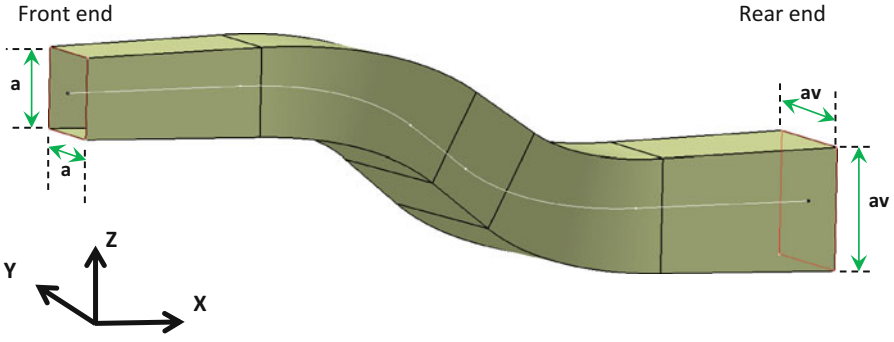



**Fig. 1.39** Tapered S-rails with tapering ratios from $av/a = 1.1$ to $av/a = 3.0$

**Table 1.1** Type A and Type B S-rail models coding (Fig. 1.39)

| Type A (without reinforcement) | | | Type B (with reinforcement) | | |
|---|---|---|---|---|---|
| No. | $R_t = av/a$ | Name | No. | $R_t = av/a$ | Name |
| 1 | 1.0 | A | 1 | 1.0 | B |
| 2 | 1.1 | A11 | 2 | 1.1 | B11 |
| 3 | 1.2 | A12 | 3 | 1.2 | B12 |
| 4 | 1.3 | A13 | 4 | 1.3 | B13 |
| 5 | 1.4 | A14 | 5 | 1.4 | B14 |
| 6 | 1.5 | A15 | 6 | 1.5 | B15 |
| 7 | 1.6 | A16 | 7 | 1.6 | B16 |
| 8 | 1.7 | A17 | 8 | 1.7 | B17 |
| 9 | 1.8 | A18 | 9 | 1.8 | B18 |
| 10 | 1.9 | A19 | 10 | 1.9 | B19 |
| 11 | 2.0 | A20 | 11 | 2.0 | B20 |
| 12 | 2.1 | A21 | 12 | 2.1 | B21 |
| 13 | 2.2 | A22 | 13 | 2.2 | B22 |
| 14 | 2.3 | A23 | 14 | 2.3 | B23 |
| 15 | 2.4 | A24 | 15 | 2.4 | B24 |
| 16 | 2.5 | A25 | 16 | 2.5 | B25 |
| 17 | 2.6 | A26 | 17 | 2.6 | B26 |
| 18 | 2.7 | A27 | 18 | 2.7 | B27 |
| 19 | 2.8 | A28 | 19 | 2.8 | B28 |
| 20 | 2.9 | A29 | 20 | 2.9 | B29 |
| 21 | 3.0 | A30 | 21 | 3.0 | B30 |

and changed to quad type to ensure the integrity of meshing. Mesh size 4 is selected to increase the accuracy of simulation. By doing so, there will be a total number of approximately 20,000 nodes on each model.

LS-PrePost is used to complete the modelling after meshing. The material used is mild steel defined as *MAT_PIECEWISE_LINEAR_PLASTICITY_24 with Young's module of 205 GPa, and mass density of 7.8e-6 kg/mm$^3$ and Poisson's ratio of 0.3. A section of 2.0 mm thick shell with material type 16 (Fully integrated shell element) was selected for ELFORM (*SECTION_SHELL). Hourglass control type 8 with coefficient of 0.1 was defined to eliminate instabilities of the material (*HOURGLASS). The true strain-true stress curve of the used material is applied to the model (Fig. 1.40).

To analyze the crushing behavior of the S-Rails between two rigid bodies, the rear end of the S-Rails is fully fixed in all six degrees of freedom. The transitional degree of freedom along the length is set free for the front end of the S-rails.

S-rails are stationary and a moving rigid wall is utilized to crush the rails by impacting the front end of the S-rails. In order to study both static and dynamic crushing behavior of the S-Rails, a 1000 kg rigid wall is crushed into the models in two different conditions: once with an initial velocity $V_0$ of 10 m/s which represents

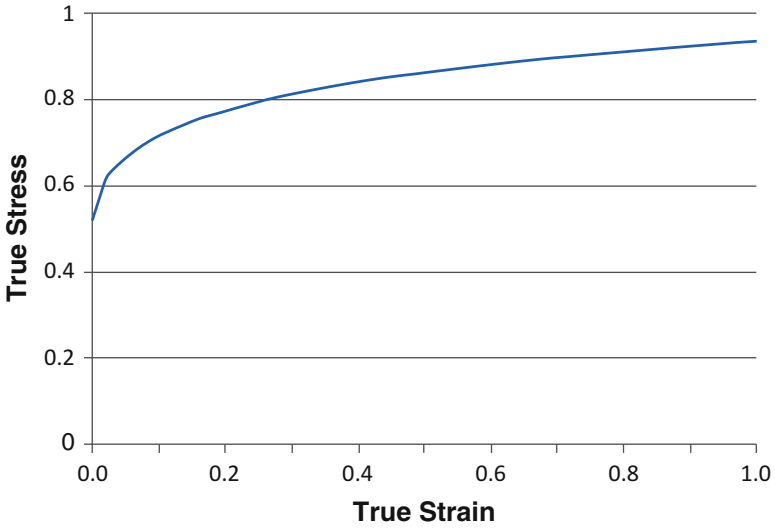



**Fig. 1.40** The true strain—true stress of the used material

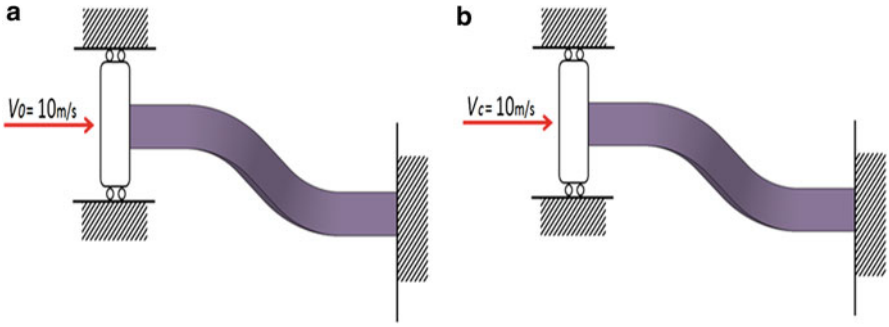



**Fig. 1.41** Boundary and loading in (**a**) dynamic and (**b**) static conditions

the dynamic loading condition (Fig. 1.41a) and once with a constant velocity $V_c$ of 10 m/s to simulate the static loading condition (Fig. 1.41b). All 42 models were subjected to both static and dynamic crushing condition resulting in 84 model-load conditions.

Dynamic crushing is simulated by a rigid wall (*RIGIDWALL_PLANAR_MOVING_FORCES) with an initial velocity of 10 m/s while the static simulation is performed using a manually defined rigid wall (*MAT_RIGID) with a constant velocity as shown in Fig. 1.42 (*DEFINE_CURVE).

The actual boundary condition in the simulation is demonstrated below in Fig. 1.43.

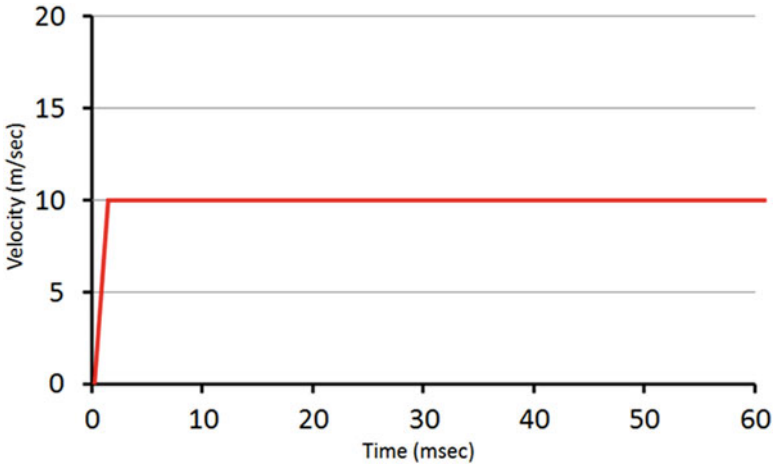



**Fig. 1.42** Velocity–time curve applied to the rigid wall in static loading condition

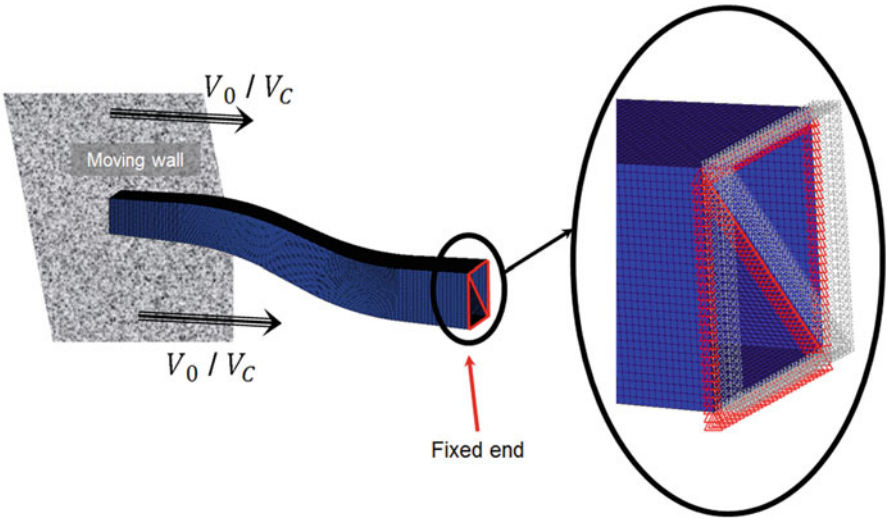



**Fig. 1.43** Boundary condition definitions in finite element simulation

Models are set to run for 60 ms (*CONTROL_TERMINATION) with 0.5 ms time intervals (*DATABASE_OPTION). Automatic single surface contact is also defined with both static and dynamic coefficient of friction of 0.08 (*CONTACT_AUTOMATIC_SINGLE SURFACE). All models are then solved using LS-DYNA 3.2 solver. After defining the above-mentioned elements in LSDYNA, the code is generated for each model.

## 1.5 Analytical Analysis and FE Result Validation

Finite element results can either be validated using experimental or analytical methods. Due to the complexity and the number of models studied in this project, analytical validation is considered as the more time and cost efficient method.

### *1.5.1 Analytical Analysis*

As explained in Sect. 1.3.2.3, Kim and Wierzbicki, developed the force-displacement Eq. (1.37) for a three dimensional S-shaped frame with constant cross-section under static loading condition. This equation was validated by the authors (Kim and Wierzbicki 2004) and can be used for other similar designs. As the tapered models pose different geometrical properties, the aforementioned equations must be modified before using for S-rails. $C_0$, $M_f$, $\sin\psi$ parameters in Eqs. (1.27), (1.34), (1.36) are not constant in a tapered S-rail due to different cross-section width and length in each displacement of $u$. To reproduce the equations for a tapered S-rail, the model is simplified as illustrated in Fig. 1.44.

From Fig. 1.44 the tapering ratio $R_t$ can be expressed as

$$R_t = \frac{a_v}{a} = \frac{a + 2\,a_t}{a} \tag{1.38}$$

So

$$a_t = {}^{1}\!/_{2}\;\, a.\,(R_t - 1) \tag{1.39}$$

$$\tan\ \beta = \frac{a_t}{L_f} \tag{1.40}$$

where

$L_f$ is the length of the center line profile.

By replacing $a_t$ from Eq. (1.39) into Eq. (1.40)

$$\tan\ \beta = \frac{{}^{1}\!/_{2}\;\, a.\,(R_t - 1)}{L_f} \tag{1.41}$$

From Fig. 1.44 $\tan\beta$ can also be considered as

$$\tan\ \beta = \frac{a_{u0}}{u} \tag{1.42}$$

where $u$ is the displacement.

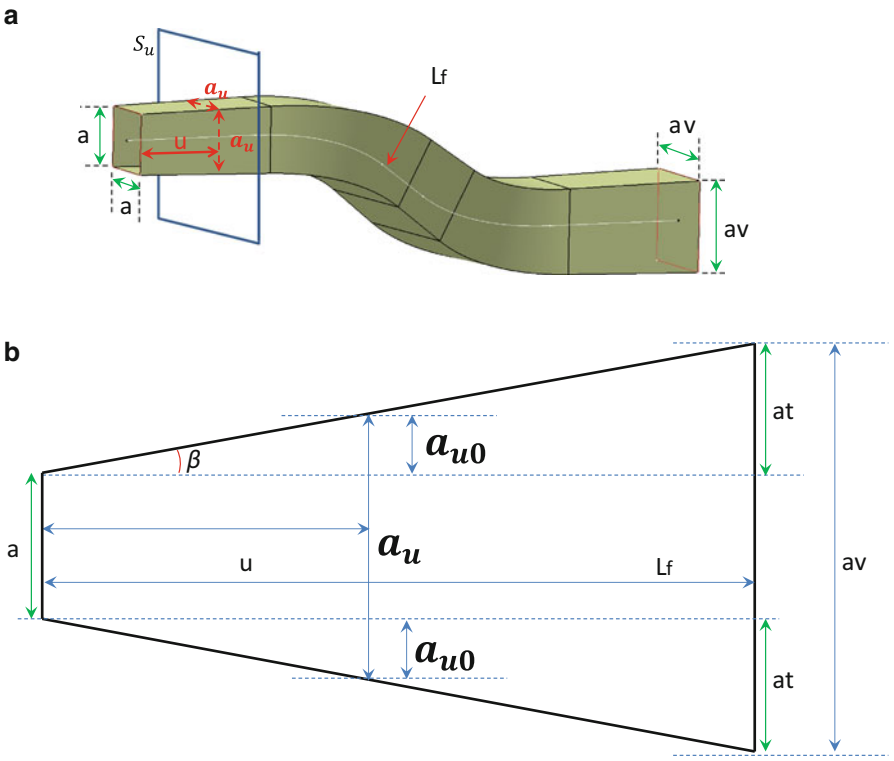



**Fig. 1.44** Tapered S-rail (**a**) and simplified side view of it (**b**)

From Eqs. (1.41) and (1.42)

$$\frac{{}^{1}\!/_{2}\ a.(R_t - 1)}{L_f} = \frac{a_{u0}}{u} \tag{1.43}$$

As demonstrated in Fig. 1.44b

$$a_u = 2a_{u0} + a \tag{1.44}$$

By replacing $a_{u0}$ from Eq. (1.44) into Eq. (1.43):

$$a_u = a\left(1 + \frac{u\,(R_t - 1)}{L_f}\right) \tag{1.45}$$

The cross-section width for each section with a displacement of $u$ can be calculated from Eq. (1.45). The same method can be used to define the equation for length of the cross-section in a rectangular cross-section (Fig. 1.45):

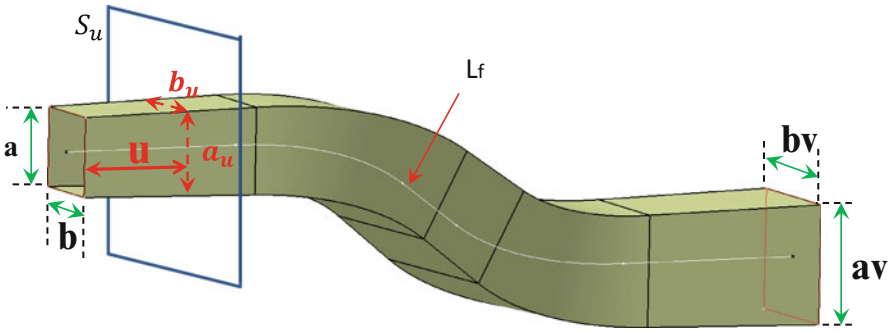



**Fig. 1.45** $a_u$ & $b_u$ of displacement $u$ in an S-rail

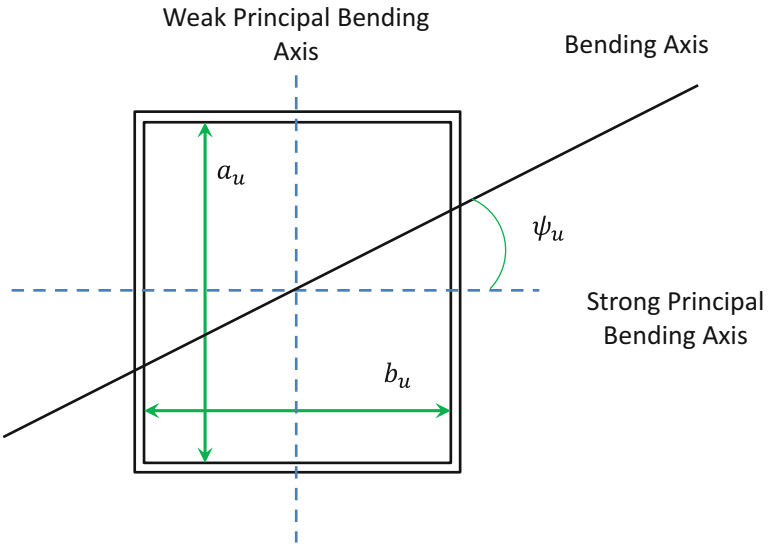



**Fig. 1.46** Orientation angle of bending axis in the S-rail section

$$b_u = b\left(1 + \frac{u\,(R_t - 1)}{L_\mathrm{f}}\right) \tag{1.45}$$

The bending axis angle $\psi_u$ (Fig. 1.46) and the moment ($M_{f,u}$) for each section with a displacement of $u$ is calculated by inserting $a_u$ and $b_u$ in Eqs. (1.36) and (1.34), respectively:

$$\sin\psi_u = \sqrt{\frac{(a_u/b_u)^2}{1 + 2\,(a_u/b_u) - (a_u/b_u)^2}} \tag{1.46}$$

Replacing Eq. (1.46) into Eq. (1.34) gives:

$$M_{f,u} = \sigma_0 t \sin\psi_u \left( \frac{a_u^2}{2}\cot^2\psi + \frac{b_u^2}{2} + a_u b_u \right) \tag{1.47}$$

where $M_{f,u}$ is the moment in displacement of $u$ for each section in a tapered S-rail. Having $a_u$, $b_u$, $\psi_u$, and $M_{f,u}$ we can modify Eq. (1.37) for a tapered S-rail:

$$P_u = \frac{M_{f,u} C_0}{\sqrt{Au+B}} \left( \begin{array}{l} \frac{\sqrt{l_1(l_1+l_2)}}{\left(-l_1 l_2 \sin\theta + l_2\sqrt{Au+B}\right)^{1/2}} + \frac{(l_1+l_2)\sqrt{l_1}}{l_2\left(\sqrt{Au+B} - l_1\sin\theta\right)^{1/2}} \\ + \frac{l_1\sqrt{l_1+l_2}}{l_2\left(-l_1\sin\theta + \sqrt{Au+B}\right)^{1/2}} \end{array} \right) \tag{1.48}$$

where $P_u$ is the force response of a S-rail with a displacement of $u$ in a static loading condition.

In Eq. (1.48) $\psi_u$, $M_{f,u}$, $C_0$, $A$, $B$ can be derived from Eqs. (1.46), (1.47), (1.27), (1.30), and (1.31), respectively. However, a more refined analytical solution considering all the dynamic effects such as strain rate effects and inertia effects can be developed for more accurate prediction of crash response of structural components.

### *1.5.2 Finite Element Results Validation*

Now that the crushing force of the S-rail can be predicted by Eq. (1.48), the finite element results and analytical prediction can be compared for static loading condition. Force-displacement curve from Eq. (1.48) and finite element results are plotted and compared for different tapering ratios. As illustrated in Fig. 1.47 for tapering ratio of $R_t = 1.1$ the correlation between analytical prediction and finite element results is almost perfect and this validates the FE results.

As another example of the correlation, Fig. 1.48 illustrates the correlation between analytical prediction and finite element results for tapering ratio of $R_t = 1.8$. The two results are in good correlation in this case as well.

Now that the results are validated, the energy absorption figures in finite element results are reviewed and discussed in the following chapter.

## 1.6 Results and Discussion

The study is carried out under both static and dynamic loading conditions. The static results are reviewed and discussed initially with the dynamic loading results and discussion following. LS-PrePost 3.2 and Altair HyperWorks 11.0 are used to view

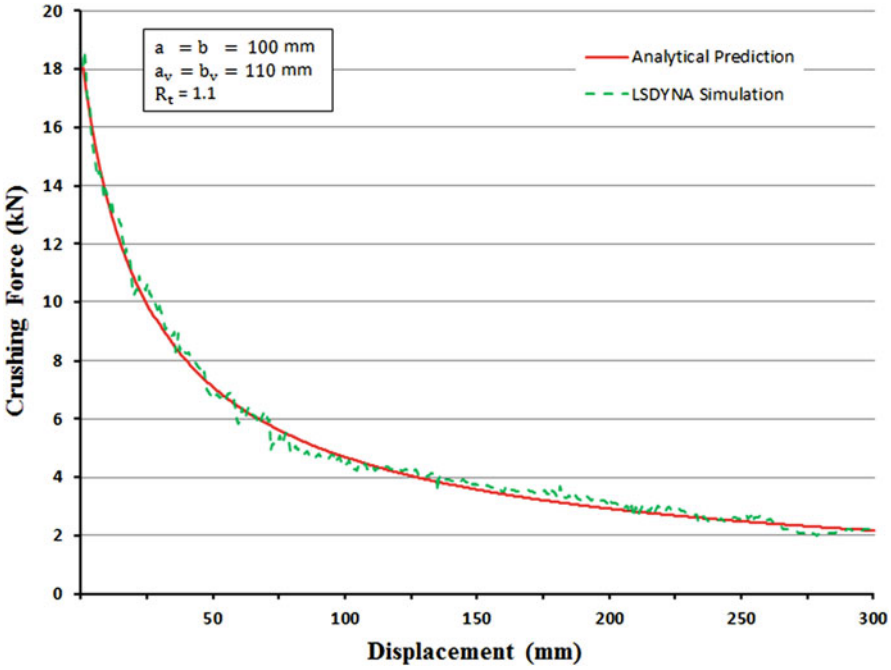



**Fig. 1.47** Displacement-crushing force comparison between analytical prediction and FE results for $R_t = 1.1$ for the first 300 mm displacement

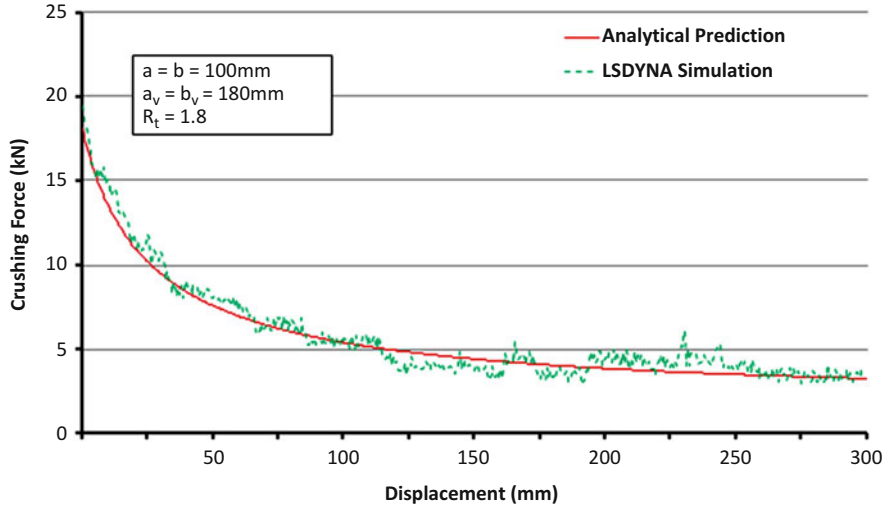



**Fig. 1.48** Displacement-crushing force comparison between analytical prediction and FE results for $R_t = 1.8$ for the first 300 mm displacement

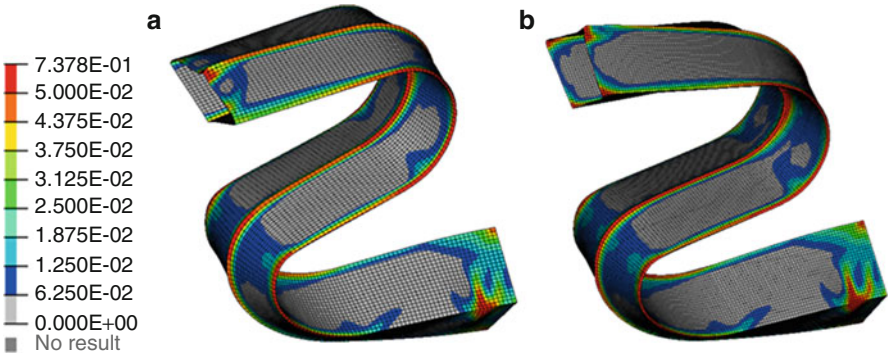



**Fig. 1.49** Model A0 deformation simulation subjecting (**a**) static and (**b**) dynamic loading

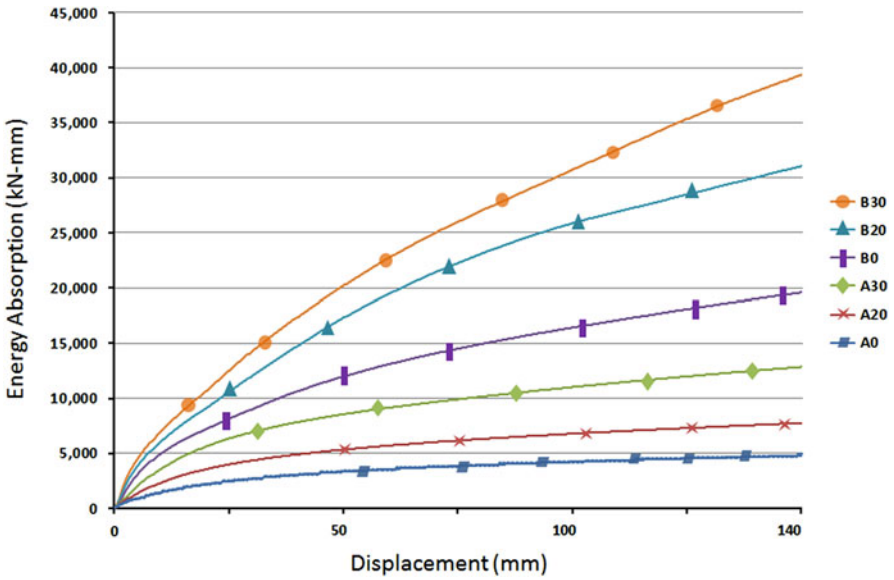



**Fig. 1.50** Energy absorption—displacement curves for models A0, A20, A30, B0, B20, and B30 subjecting dynamic loading

the FE results. Figure 1.49 illustrated the deformed base model A0 S-rails resulting from (a) static and (b) dynamic loading conditions.

The results demonstrate that global bending is the most prominent process during crash. This conclusion aligns totally with the findings in other studies done on S-rails. As illustrated in Fig. 1.50, the results demonstrated a direct correlation between deformation and energy absorption. Under dynamic loading, an ascending trend in absorbed energy is observed when tapering ratio increases from A0 to A30 and also from B0 to B30. The trend of calculated energy absorption agrees with

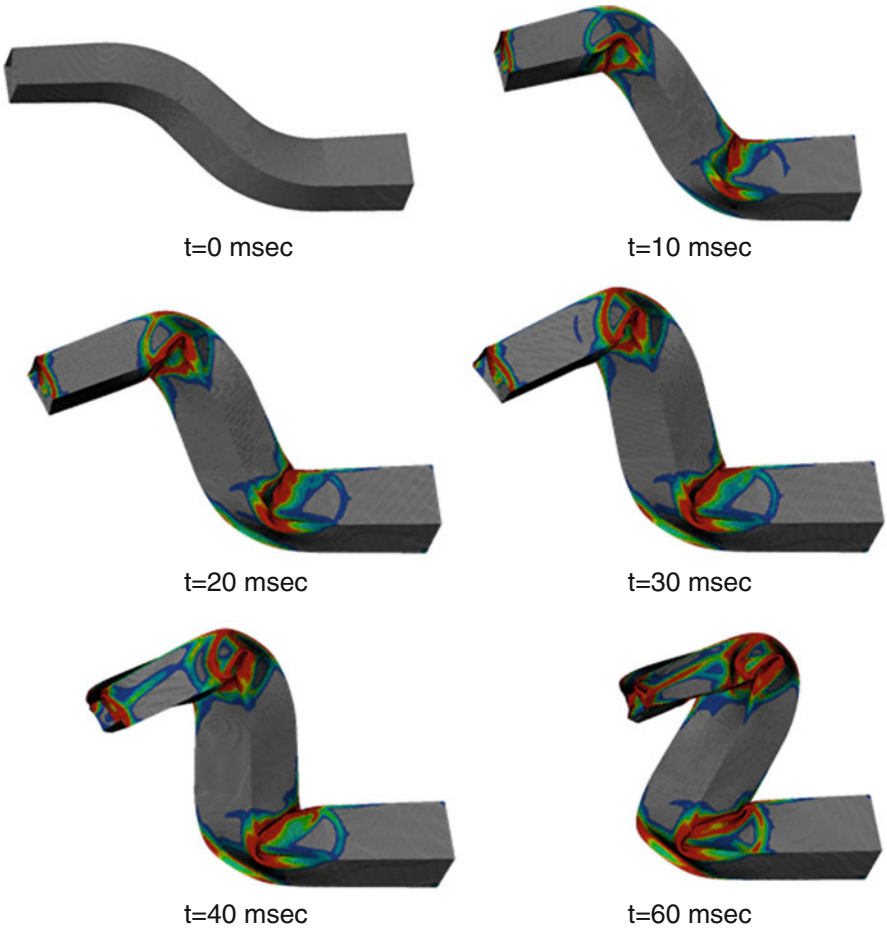



**Fig. 1.51** Deformed B14 S-rail subjecting dynamic static loading in time $t = 0$, 10, 20, 30, 40, 60 ms

the results achieved by Kim and Wierzbicki (2004). Reinforced S-rail B0 shows a significant increase in absorbed energy comparing to normal base model A0 confirming Zhang C. and Saigal A. findings (Reid et al. 1984).

Figure 1.51 illustrates model B14 deformation in time = 0, 10, 20, 30, 40, 60 ms. As illustrated, the S-frame deformation is mainly in the form of bending around the hinge points as assumed in Kim and Wierzbicki investigations (Kim and Wierzbicki 2004)

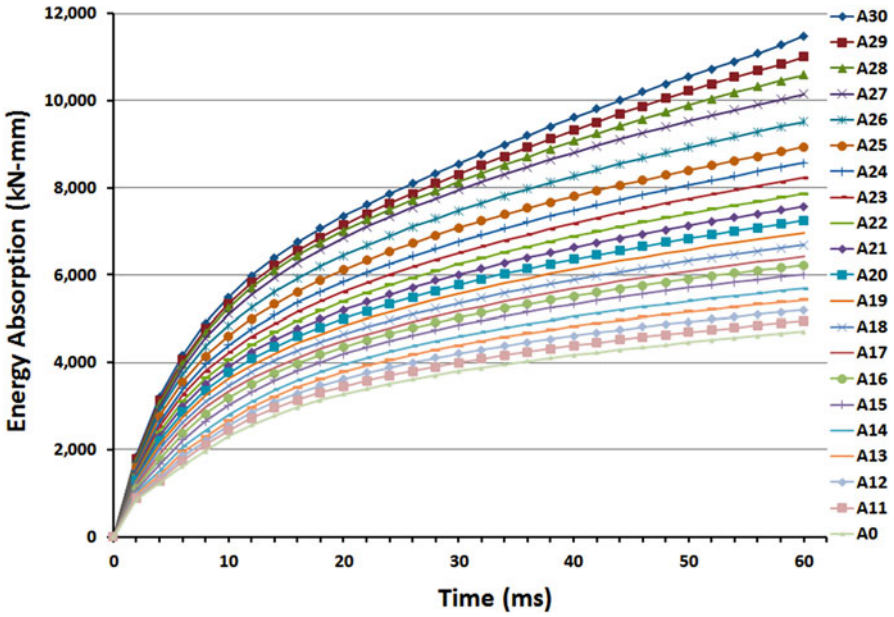



**Fig. 1.52** Time-energy absorption trends for type A S-rails subjecting static load

## *1.6.1 Static Loading*

### 1.6.1.1 Type A Models Subjecting Static Load

All 21 type A models are subjected to static loading. FE results show that type "A" S-rails subjected to static loading demonstrate more energy absorption as the tapering ratio increases. Figure 1.52 demonstrates the energy absorption trends in type A models. A noticeable increase in energy absorption from 4703 kN-mm in model A0 to a maximum of 11,483 kN-mm in A30 can be seen due to tapering in the studied models. This is a 144 % improvement in energy absorption. The energy absorption rate is higher in the first 10 ms almost in all models. This trend continues with a moderate and steady rate till time = 60 ms.

### 1.6.1.2 Type B Models Subjecting Static Load

In type "B" S-rails subjected to static loading, increasing the tapering ratio generally results in an increase in the level of energy absorption but a uniform trend such as the one observed in type A models is not documented. Figure 1.53 shows the energy absorption trends in type B models subjected to static loading. As illustrated, there is a significant increase in energy absorption rates, increasing from 24,100 kN-mm in model B0 to 52,443 kN-mm in model B30. This is a 118 % improvement in

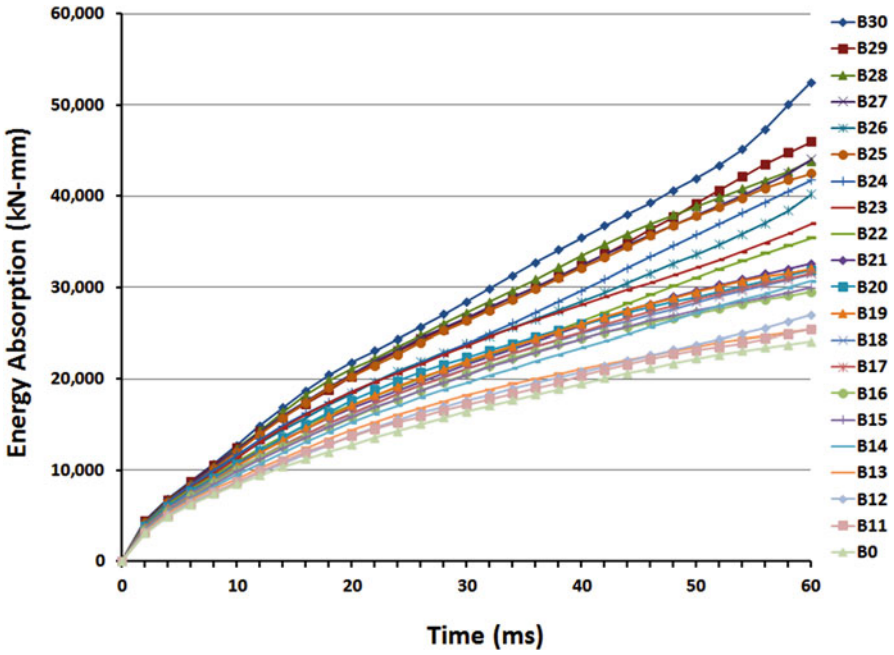



**Fig. 1.53** Time-energy absorption trends for type B S-rails subjecting static load

energy absorption from B0 to B30 model. As demonstrated in Fig. 1.53 the energy absorption trends are different in type B models and some models follow a different trend after 45 ms.

### 1.6.1.3 Specific Energy Absorption in Static Loading Condition

The weight factor is eliminated from the energy absorption figures to avoid bias. As illustrated in Fig. 1.54, the energy absorption increases with an almost constant slope and a linear behavior by increasing the tapering ratio from A0 to A30. Meanwhile the SEA trend does not follow a constant pattern. SEA remains almost constant from A0 to A20 then shows a slight increase up to A26. The ascending trend continues with a minimally faster slope from A26 to A30. An increase of 22 % in SEA from 694.3 kN-mm/kg in A0 to 843.9 kN in A30-mm/kg is observed in type A models subjected to static loading.

While there is an absolute increase in the level of energy absorption in tapered type A models, type B models show a different pattern when subjected to static loading (Fig. 1.55). While the maximum SEA is achieved in model B14 (tapering ratio $R_t = 1.4$) the total trend is variable with multiple increase and decline phases. As described the peak in SEA is achieved in model B14 with 6 % increase (Fig. 1.54).

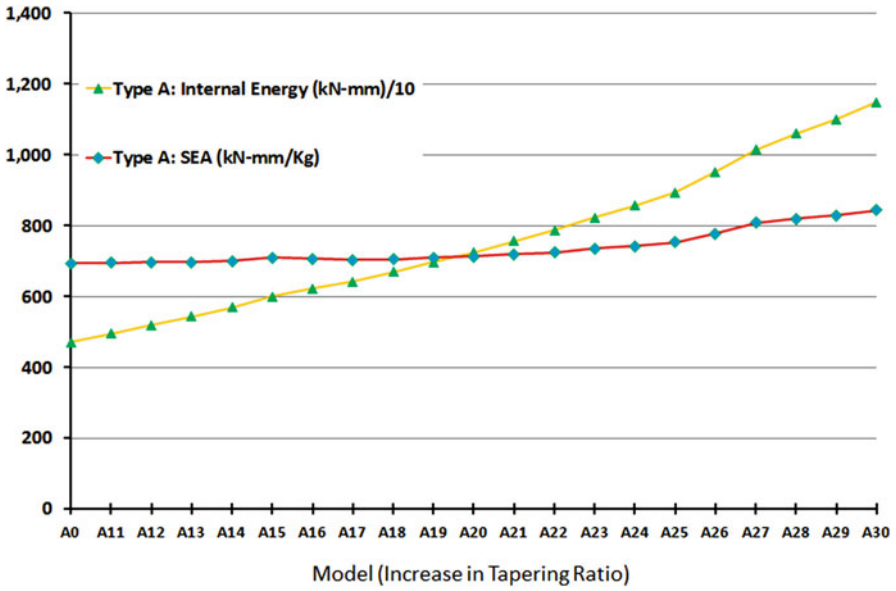



**Fig. 1.54** EA and SEA trends in type A models under static load

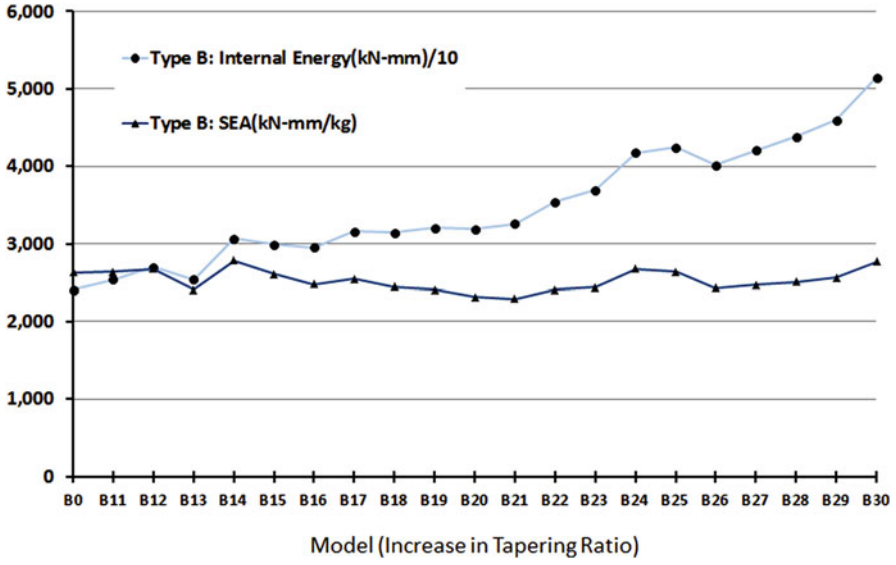



**Fig. 1.55** EA and SEA trends in type A models under static load

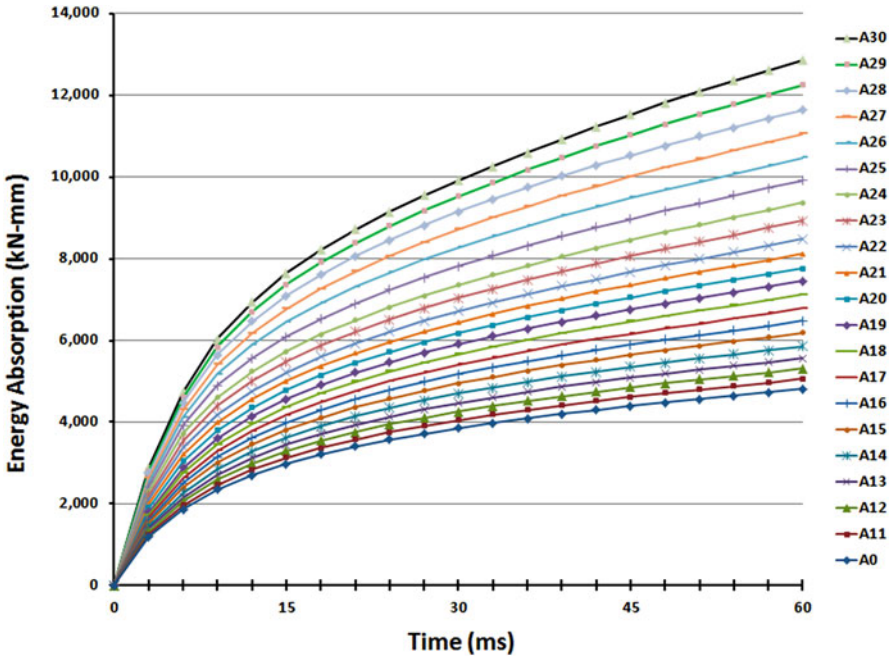



**Fig. 1.56** Time-energy absorption trends for type A S-rails subjecting dynamic load

## *1.6.2 Dynamic Loading*

### 1.6.2.1 Type A Models Subjecting Dynamic Load

Similar to static loading conditions, an increase in energy absorption is observed in dynamic loading. Figure 1.56 illustrates the energy absorption trend in type A models under dynamic loading. Accordingly, a large increase in energy absorption is observed from 4811.7 kN-mm in model A0 to 12,865 kN-mm in model A30 in the studied models. This is a 167 % improvement in energy absorption from A0 to A30 models. The energy absorption increase rate is observed to be higher in the first 10–12 ms. The ascending trend then continues with a moderate pace and no significant changes in the slope for all models

### 1.6.2.2 Type B Models Subjecting Dynamic Load

The energy absorption level increases in type "B" S-rails subjected to dynamic loading by increasing the tapering ratio. Figure 1.57 shows the energy absorption trends in type B models subjected to dynamic loading. As illustrated, there is a noticeable increase in energy absorption from 19,591 kN-mm in model B0 to 39,375 kN-mm in model B30 resulting in an improvement of 101 % in energy

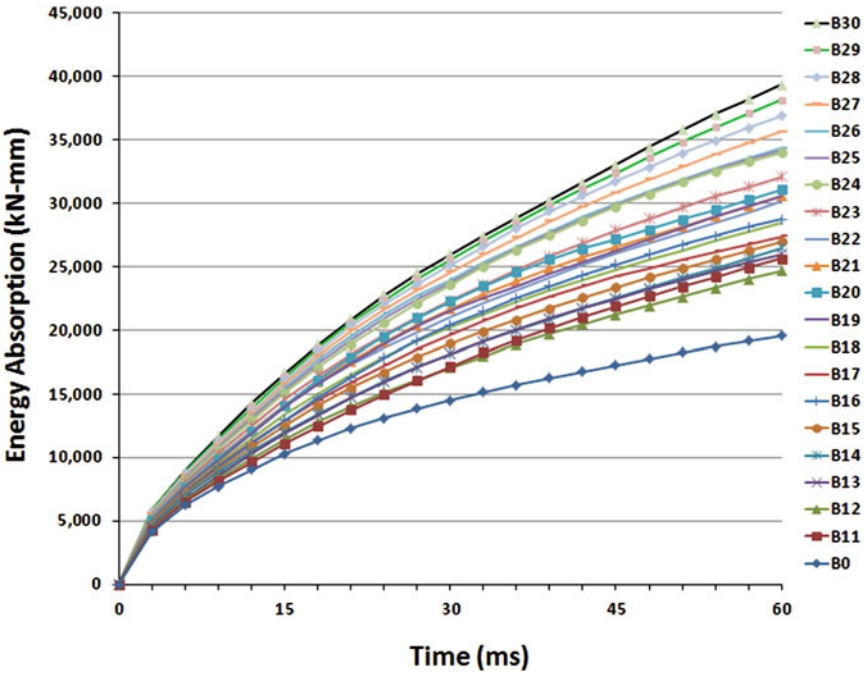



**Fig. 1.57** Time-energy absorption trends for type B S-rails subjecting dynamic load

absorption. The energy absorption increasing trend follows a faster pace in the first few milliseconds and then continues with a slightly more moderate slope which also increases as the tapering ratio is increased

### 1.6.2.3 Specific Energy Absorption in Dynamic Load Condition

As illustrated in Fig. 1.58, the increase in tapering ratio from A0 to A30 results in an increase in the energy absorption with an almost constant slope and a linear behavior. On the other hand, SEA does not follow a similar pattern. SEA remains almost constant from A0 to A22 then an increasing trend is documented with a slow pace up to A30. In total, an increase of 33 % in SEA from 710.32 kN-mm/kg to 945.47 kN-mm/kg is observed in the studied type A models in dynamic loading condition.

As illustrated in Fig. 1.59, despite an increase in the level of energy absorption in type B models subjected to dynamic load, the SEA trend does not follow a constant increasing trend. Maximum SEA is achieved in model B11 with 25 % increase from 2136.66 kN-mm/kg in B0. SEA follows a variable trend with multiple ascends and descends after wards.

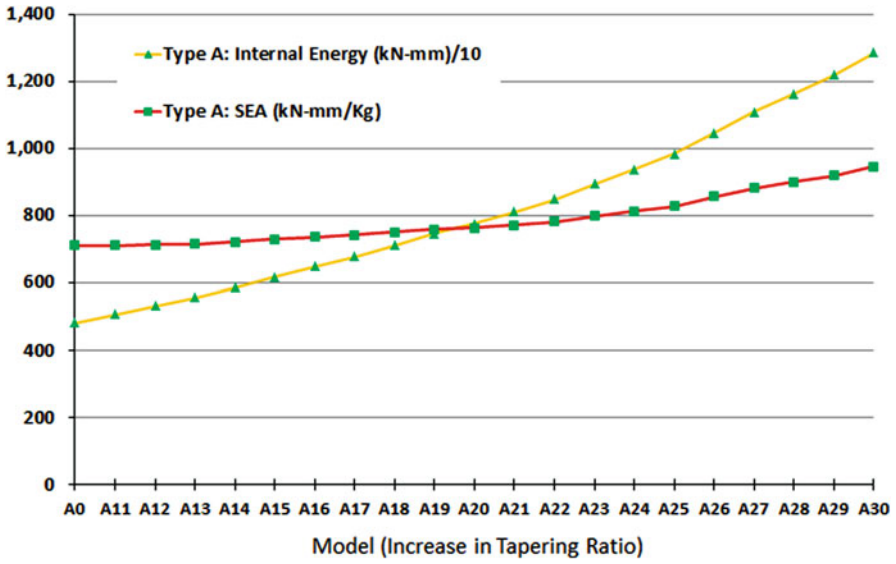



**Fig. 1.58** EA and SEA trends in type A models under dynamic load

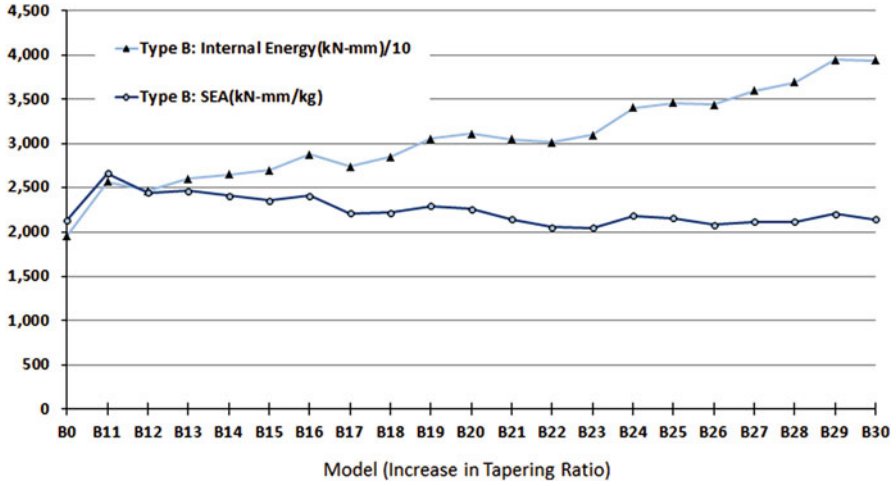



**Fig. 1.59** EA and SEA trends in type B models under dynamic load

### *1.6.3 The Effect of Reinforcing S-Rails*

Our results indicate that in dynamic loading conditions, reinforcing the base model A0 results in an increase of 307 % in the energy absorption from 4811.7 kN-mm in A0 to 19,591 kN-mm in B0. The improvement in SEA is about 201 % from

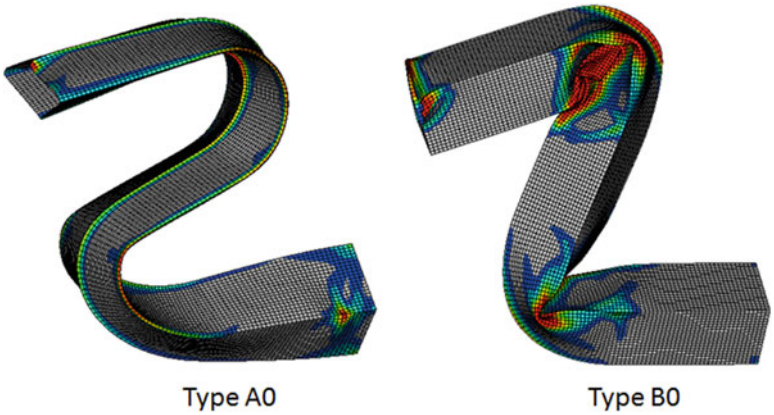



**Fig. 1.60** Type A0 and B0 deformation in time $t = 60$ ms

710.32 kN-mm/kg in model A0 to 2136.66 kN-mm/kg in model B0 when internal reinforcement is applied.

In static loading condition, a 412 % increase in energy absorption is observed by reinforcing the base model A0 (energy absorption of 4703.1 kN-mm in A0 to 24,100 kN-mm in B0). This is equivalent to 279 % increase in SEA. As illustrated in Fig. 1.60 for time $t = 60$ ms, type B0 demonstrates a very different deformation in comparing to A0.

The increase in energy absorption levels after applying lateral internal reinforcement are in line with previous studies and confirm the findings by Zhang and Saigal (2007)

### *1.6.4 EA and SEA in Dynamic Loading Versus Static Loading*

As illustrated in Fig. 1.61, finite element analysis demonstrates that in all type A models the energy absorption level is slightly higher in dynamic loading condition compared to static loading condition. This result is in agreement with Kim and Wierzbicki findings (Kim and Wierzbicki 2004).

The same behavior is also observed in SEA (Fig. 1.62). Higher level of SEA can be seen in dynamic load compared to static load.

Contrary to type A, in type B models the energy absorption levels are usually higher in static loading compared to dynamic load conditions (Fig. 1.63).

Likewise, the SEA levels have the same behavior: higher SEA in static load compared to dynamic loading condition as demonstrated below in Fig. 1.64.

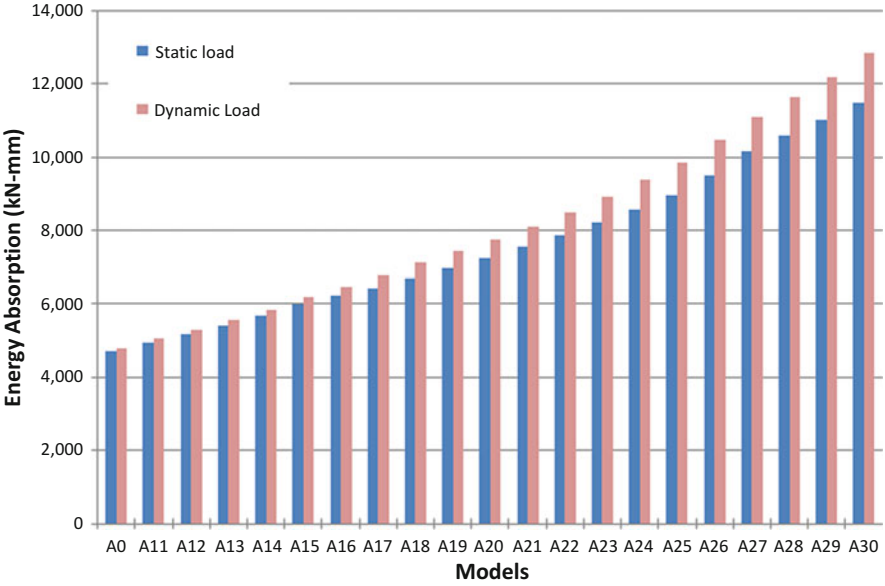



**Fig. 1.61** Energy absorption of type A models in static and dynamic load

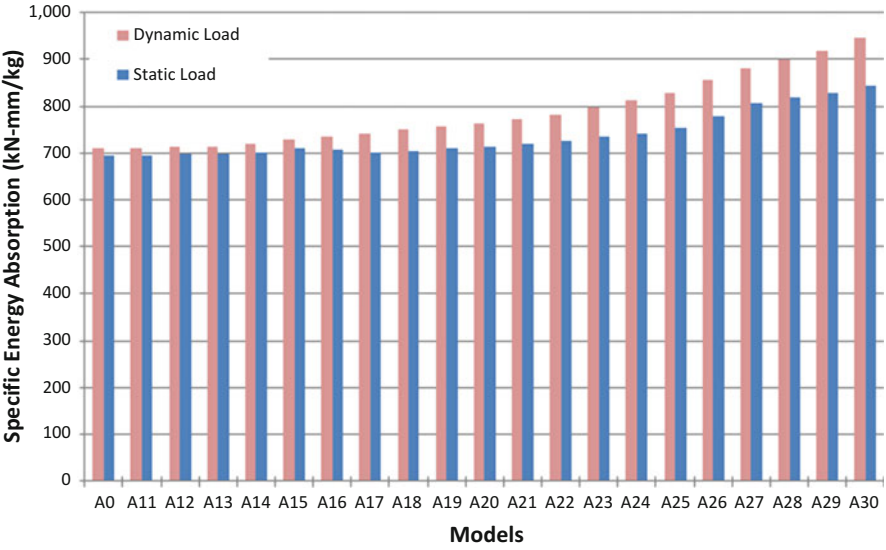



**Fig. 1.62** Specific energy absorption of type A models in static and dynamic load

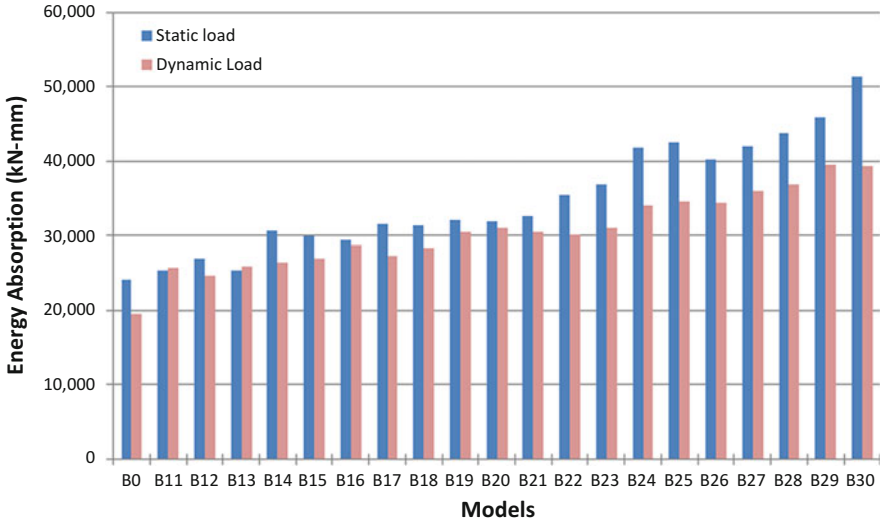



**Fig. 1.63** Energy absorption of type B models in static and dynamic load

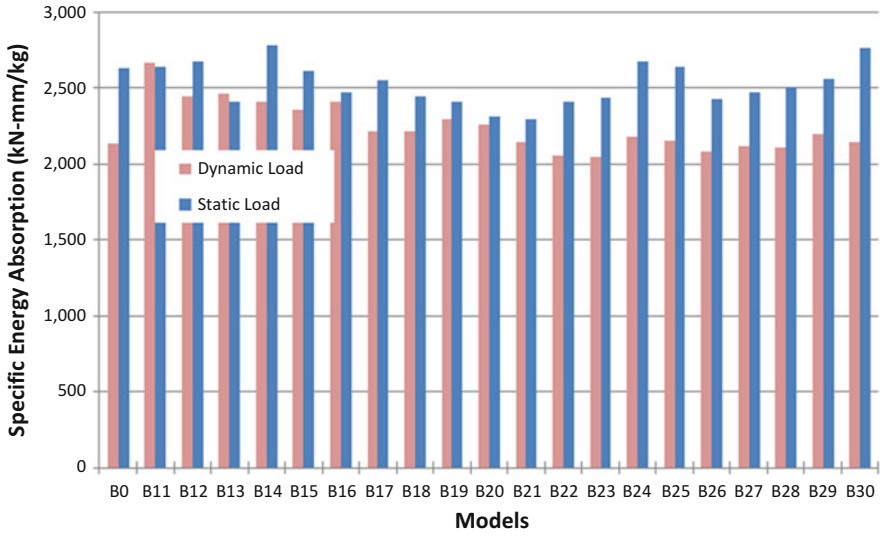



**Fig. 1.64** Specific energy absorption of type B models in static and dynamic load

### *1.6.5 Models with the Highest SEA*

Tapering the base model S-rail A0 by 40 % together with the addition of a diagonal internal reinforcement, improves the SEA from 694.29 kN-mm/kg in model A0 to 2785.44 kN-mm/kg in model B14 in static loading condition. This is the optimal

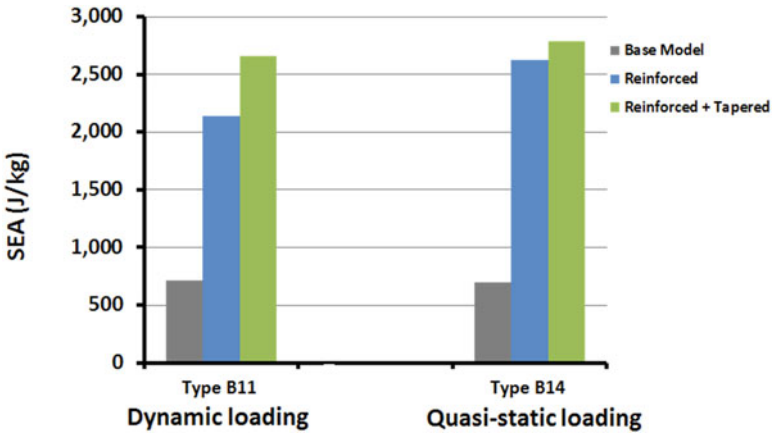



**Fig. 1.65** Specific energy absorption for optimum design models B11 (dynamic load) and B14 (static load)

design in this loading condition with a SEA improvement of 301 % compared to the base model. In the studied models, tapering ratio of $R_t = 1.4$ can be considered as the optimum design point in static loading condition. In dynamic loading condition the SEA significantly increases by 275 % from 710.32 kN-mm/kg in model A0 to 2664.69 kN-mm/kg in model B11. In other words, in the models subjected to dynamic loading condition the optimum tapering ratio is $R_t = 1.1$. These results are illustrated below in Fig. 1.65.

## 1.7 Conclusion and Recommendations

Safety is a key design priority in the automotive industry. In a car crash, the higher level of energy absorption in the frontal structures leads to less transferred energy to the passengers and hence a safer car. S-shaped front rails, also known as S-rails, are one of the main structural elements and energy absorbers in a car body. Energy absorption in the S-rails mainly happens through local buckling. In order to improve the passenger safety in a frontal crash, S-rails design should be optimized to absorb higher level of energy while crushing. Though many studies have targeted S-rails and investigated many different factors in design such as cross-section forms, thickness, material, etc., none have focused on tapered S-rails.

In this study, a model of steel S-rail with a length of one meter and a thickness of 2.0 mm was subjected to 20 different tapering ratios (from 110 % to 300 %) with 10 % increments which resulted in generation of 20 models (type-A). The same base model was reinforced by adding a lateral internal reinforcement and again subjected to 20 tapering ratios like the first S-rail (type-B). In total 42 models generated. Two different loading conditions were applied to crush the S-rails: once

a rigid wall with a constant velocity of 10 m/s used to crush the S-rails (static load), and in another test, a rigid wall with an initial velocity of 10 m/s crushed the S-rails (dynamic load). Finite element models were developed, solved, and analyzed using Altair HyperWorks and LS-DYNA software. The effectiveness of tapering S-rails was assessed through investigating the SEA variations in different models. To analytically validate the results, a modified mathematical equation has been introduced. The modified equation is based on Kim and Wierzbicki (Kim and Wierzbicki 2004) response force equation and is in a good agreement with the numerical results. In this study we showed that in static loading conditions, tapering type-A models showed 144 % increase in EA and 22 % improvement in SEA. Also, tapering type-B models showed 118 and 6 % improvement in EA and SEA, respectively. In this loading condition reinforcing type-A0 model to B0 model demonstrated 412 % increase in EA and 279 % in SEA. In dynamic loading condition, an increase of 167 % EA was observed in type-A models as well as 33 % increase in SEA. Type-B models subjected to dynamic loading showed 101 % increase in EA and a maximum of 25 % improvement in SEA. In this loading condition, reinforcing type-A0 model to B0 model also increased the EA by 307 % and SEA by 201 %. Models B14 and B11 demonstrated the Maximum SEA by 275 and 301 % improvement in static and dynamic loading, respectively. In the other words, considering the physical and geometrical constraints in car body design, as well as the optimization goal in auto industries, the optimal design to maximize the energy absorption with minimum mass has been determined.

## References

A. I. a. S. Institute, Ultra Light Steel Auto Body, *1.0.2,* pp. 19–24 (1998).

Abramowicz, W. (1981). Simplified crushing analysis of thin-walled columns and beams. *Engineering Translations, 29*(1), 5–26.

Abramowicz, W. (1983). The effective crushing distance in axially compressed thin walled metal columns. *International Journal of Impact Engineering, 1*, 309–317.

Abramowicz, W., & Wierzbicki, T. (1989). Axial crushing of multicorner sheet metal columns. *Journal of Applied Mechanics, 56*, 113–120.

Alexander, A. (1960). An approximate analysis of the collapse of thin cylindrical shells under axial loading. *Quarterly Journal of Mechanics and Applied Mathematics, 13*, 10–15.

Al-Hassani, S., Johnson, W., & Lowe, W. (1972). Characteristics of inversion tube under axial loading. *Journal of Mechanical Engineering Science, 14*, 370–381.

Australian Road Deaths Database, Bureau of Infrastructure, Transport and Regional Economics, [Online]. Available: www.bitre.gov.au/statistics/safety/fatal_road_crash_database.aspx. Accessed July 2012.

Australian NCAP, [Online]. Available: http://www.ancap.com.au. Accessed August 2012.

Carney, J., & Pothen, S. (1988). Energy dissipation in braced cylindrical shells. *International Journal of Mechanical Science, 30*(3/4), 203–2016.

Chen, W., & Wierzbicki, T. (2001). Relative merits of single-cell, multi-cell and foam-filled thin-walled structures in energy absorption. *Thin-Walled Structures, 39*(4), 287–306.

Cheng, Z., Thacker, J., Pilkey, W., Hollowell, W., Reagan, S., & Sieveka, E. (2001). Experiences in reverse-engineering of a finite element automobile crash model. *Finite Elements in Analysis and Design, 37*(11), 843–860.

Dieter, A. (2000). *The passenger car body, design, deformation characteristics, accident repair*. Warrendale, PA: SAE International.

Euro NCAP, [Online]. Available: http://www.euroncap.com/. Accessed August 2012.

Fan, Z., Lu, G., & Liu, K. (2013). Quasi-static axial compression of thin-walled tubes with different cross-sectional shapes. *Engineering Structures, 55*, 80–89.

Han, J., & Yamazaki, K. (2003). Crashworthiness optimisation of S-shape square tubes. *International Journal of Vehicle Design, 31*(1), 72–85.

Harte, A.-M., Fleck, N. A., & Ashby, M. F. (2000). Energy absorption of foam-filled circular tubes with braided composite walls. *European Journal of Mechanics/A Solids, 19*(1), 31–50.

Hou, S., Li, Q., Long, S., Yang, X., & Li, W. (2008). Multiobjective optimization of multi-cell sections for the crashworthiness design. *International Journal of Impact Engineering, 35*(11), 1355–1367.

Johnson, W., & Reid, S. (1978). Metallic energy dissipating systems. *Applied Mechanics, 31*(3), 277–288.

Jun, W., Libo, C., Tianzhi, C., Chenchen, H., Binghui, J., & Huan, C. (2010). Structure improvement of the S beam based on the safety of SUV. *Applied Mechanics and Materials, 34–35*, 675–680.

Karamanlidis, D., & Jasti, R. (1987). Geometrically nonlinear finite element analysis of tapered beams. *Computers & Structures, 25*(6), 825–830.

Kecman, D. (1983). Bending collapse of rectangular and square section tubes. *International Journal of Mechanical Sciences, 25*, 623–636.

Kim, H.-S., & Wierzbicki, T. (2000). Numerical and analytical study on deep biaxial bending collapse of thin-walled beams. *International Journal of Mechanical Sciences, 42*(10), 1947–1970.

Kim, H., & Wierzbicki, T. (2001a). Effect of the cross-sectional shape on crash behaviour of a three dimensional space frame. *International Journal of Vehicle Design, 25*(4), 295–316.

Kim, H.-S., & Wierzbicki, T. (2001b). Effect of the cross-sectional shape of hat-type cross-sections on crash resistance of an "S"-frame. *Thin-Walled Structures, 39*(7), 535–554.

Kim, H.-S., & Wierzbicki, T. (2004). Closed-form solution for crushing response of three-dimensional thin-walled "S" frames with rectangular cross-sections. *International Journal of Impact Engineering, 30*(1), 87–112.

Kotełko, M., Sebastian, L., & Radosław, J. (2009). Dynamic crushing tests of thin walled members under compression. *Materials Engineering, 16*(1), 14–19.

Langseth, M., & Hopperstand, O. (1996). Static and dynamic axial crushing of square thin-walled aluminium extrusions. *International Journal of Impact Engineering, 18*(7/8), 949–968.

Roads and Maritime Services, Australia, NSW, [Online]. Available: http://www.rms.nsw.gov.au/licensing/tests/hazardperceptiontest/novicedriversareatrisk/5mostcommoncrashtypes.html. Accessed July 2012.

Nagel, G., & Thambiratnam, D. (2004). A numerical study on the impact response and energy absorption of tapered thin-walled tubes. *International Journal of Mechanical Sciences, 46*(2), 201–216.

Nagel, G., & Thambiratnam, D. (2005). Computer simulation and energy absorption of tapered thin-walled rectangular tubes. *Thin-Walled Structures, 43*(8), 1225–1242.

Nahas, M. (1993). Impact energy dissipation characteristics of thin-walled cylinders. *Thin-Walled Structures, 15*(2), 81–93.

Najafi, A., & Rais-Rohani, M. (2011). Mechanics of axial plastic collapse in multi-cell, multi-corner crush tubes. *Computers and Structures, 49*(1), 1–12.

National Highway Traffic Safety Administration, [Online]. Available: http://www.nhtsa.gov/. Accessed July 2012.

Qi, C., Yang, S., & Dong, F. (2012). Crushing analysis and multiobjective crashworthiness optimization of tapered square tubes under oblique impact loading. *Thin-Walled Structures, 59*, 103–119.

Reddy, J. (2006). *An introduction to the finite element method*. New York, NY: McGraw-Hill Higher Education.

Reddy, T., & Al-Hassani, S. (1993). Axial crushing of wood-filled square metal tubes. *International Journal of Mechanical Science, 35*(3/4), 231–246.

Reddy, T., & Wall, R. (1988). Axial compression of foam-filled thin-walled circular tubes. *International Journal of Impact Engineering, 7*, 151–160.

Reid, S., Austin, C., & Smith, R. (1984). Tubular rings as impact energy absorber. *Structural Impact and Crashworthiness*, 2, 555–563.

Reid, S., & Reddy, T. (1986). Static and dynamic crushing of tapered sheet metal tubes. *International Journal of Mechanical Science, 28*, 623–637.

Reid, S., Reddy, T., & Gray, M. (1986). Static and dynamic axial crushing of foam filled sheet metal tubes. *International Journal of Mechanical Science, 28*, 295–322.

Santosa, S., & Wierbicki, T. (1998). On the modeling of crush behavior of a closed-cell aluminum foam structure. *Journal of the Mechanics and Physics of Solids, 46*(4), 645–669.

Santosa, S., & Wierzbicki, T. (1998). Crash behaviour of box columns filled with aluminum honeycomb or foam. *Computers and Structures, 68*(4), 343–367.

Santosa, S., Wierzbicki, T., Hansen, A., & Langseth, M. (2000). Experimental and numerical studies of foam-filled sections. *International Journal of Impact Engineering, 24*(5), 509–534.

Shariatpanahi, M., Masoumi, A., & Ataei, A. (2008). Optimum design of partially tapered rectangular thin-walled tubes in axial crushing. *Journal of Engineering Manufacture, 222*, 285–291.

Singace, A., & El-Sobky, H. (1997). Behaviour of axially crushed corrugated tubes. *International Journal of Mechanical Science, 39*(3), 249–268.

Song, J., Chen, Y., & Lu, G. (2012). Axial crushing of thin-walled structures with origami patterns. *Thin-Walled Structures, 54*, 65–71.

Timoshenko, S., & Gere, J. (1961). *Theory of elastic stability* (2nd ed.). New York: McGraw-Hill.

Vasudevan, S., Okada, H., & Atluri, S. (1996). Development of a new frame finite element for crash analysis, using a mixed variational principle and rotations as independent variables. *Finite Elements in Analysis and Design, 23*(2), 155–171.

Wierzbicki, T., & Abramowicz, W. (1983). On the crashing mechanics of thin walled structures. *Journal of Applied Mechanics, 50*, 727–734.

Wierzbicki, T., Recke, L., Abramowicz, W., Gholami, T., & Huang, J. (1994). Stress profiles in thin-walled prismatic columns subjected to crush loading-II. Bending. *Computers and Structures, 51*(6), 625–641.

Wu, H., & Xin, Y. (2009). Optimal design of the S-rail for crashworthiness analysis. *International Joint Conference on Computational Sciences and Optimization, 2*, 735–738.

Yucheng, L. (2008). Crashworthiness design of multi-corner thin-walled columns. *Thin-Walled Structures, 46*(12), 1329–1337.

Zhang, C., & Saigal, A. (2007). Crash behavior of a 3D S-shape space frame structure. *Journal of Materials Processing Technology, 191*(1), 256–259.

# Chapter 2
# Caster–Camber Relationship in Vehicles

**Dai Q. Vo, Hormoz Marzbani, Mohammad Fard, and Reza N. Jazar**

**Abstract** A variation in the camber of an automotive wheel is desired to compensate a side-slip force reduction owing to transversely normal load shift when the car is cornering. The camber of a steered wheel can be varied by adjusting caster or/and lean angles which are the representations of steering axis orientation. Thus, adjustable camber can be created by variable caster or/and lean angle. Choosing which parameter among the two angles to be the main source of camber generation is a key starting point and dependent on its different effects. Here, homogeneous transformation is employed to establish camber as a function of caster, lean angle, and steering angle in the general case. A comparison between caster and lean angle based on different criteria is made. The comparison shows that a variable caster is more effective than a variable lean angle in generating a smart camber.

**Keywords** Variable camber • Controllable camber • Variable caster • Adjustable caster • KPI angle • Lateral force • Camber force • Steering tire vehicle dynamics

## 2.1 Introduction and Background

In automobile, cornering force is mainly contributed by two components: side-slip force and camber thrust, as shown in Fig. 2.1.

In cars and trucks, side-slip force is the main source of the cornering force; and, as illustrated in Fig. 2.2, it increases with normal tire force in the form of a saturating curve. Therefore, due to a normal load transfer between the left and the right wheels of the vehicle when cornering, the side-slip forces change. This leads to a reduction

D.Q. Vo (✉) • H. Marzbani • M. Fard • R.N. Jazar
School of Aerospace, Mechanical and Manufacturing Engineering, RMIT University, Melbourne, VIC, Australia
e-mail: dai.voquoc@rmit.edu.au; hormoz.marzbani@rmit.edu.au; mohamad.fard@rmit.edu.au; reza.jazar@rmit.edu.au



R.N. Jazar, L. Dai (eds.), *Nonlinear Approaches in Engineering Applications*,
DOI 10.1007/978-3-319-27055-5_2

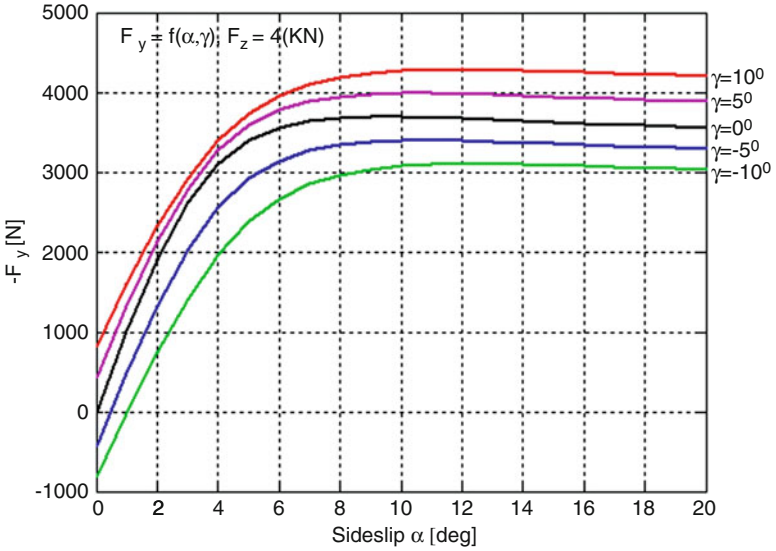



**Fig. 2.1** Lateral force as a function of side-slip angle and camber angle

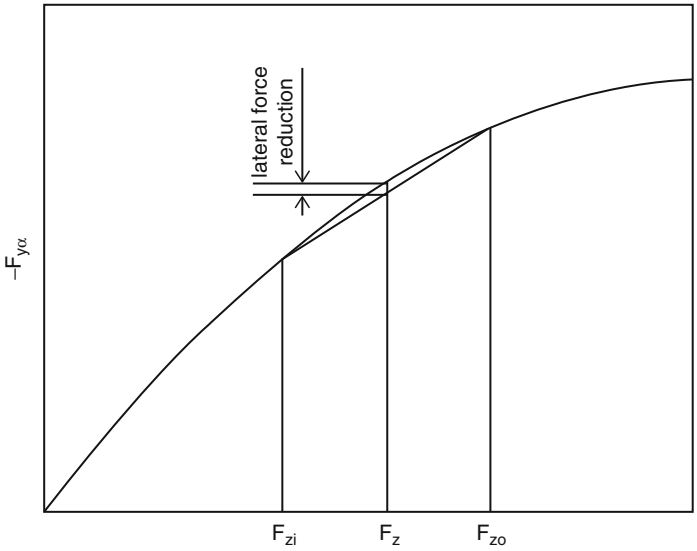



**Fig. 2.2** The reduction of lateral force of a car when turning

of the total lateral force of the vehicle (Jazar 2014; William et al. 2002; Reimpell et al. 2000; Wong 2001). In order to keep the vehicle on road, variations in camber are required to compensate the changes in the side-slip forces. For a steerable tire, camber can be varied by changing the orientation of the steering kingpin axis which is represented by caster and lean angle (Jazar et al. 2012).

The homogeneous transformation is the one that maps coordinates in one frame to another frame rotating about an arbitrary axis. In this paper, the homogeneous transformation is employed to develop the kinematics of the tire. Different simplified designs are also presented. As camber is the key parameter of the topic the camber function of a steered tire is also expressed in general case. Based on the function, a comparison between caster and lean angle is made for choosing the most effective one to control camber.

## 2.2 Homogeneous Transformation

The homogeneous transformation can be employed to transfer coordinates between two coordinate frames in which one rotates about an arbitrary axis with respect to the other. Here, we introduce a basic understanding of the transformation.

### *2.2.1 Rigid Body Motion*

To begin with, we consider a simple situation in which a rigid body $B$ rotating about a point $O$ fixed to ground as shown in Fig. 2.3. Two coordinate systems are defined: the first one called local coordinate frame $B(Oxyz)$ is attached to the body; the other

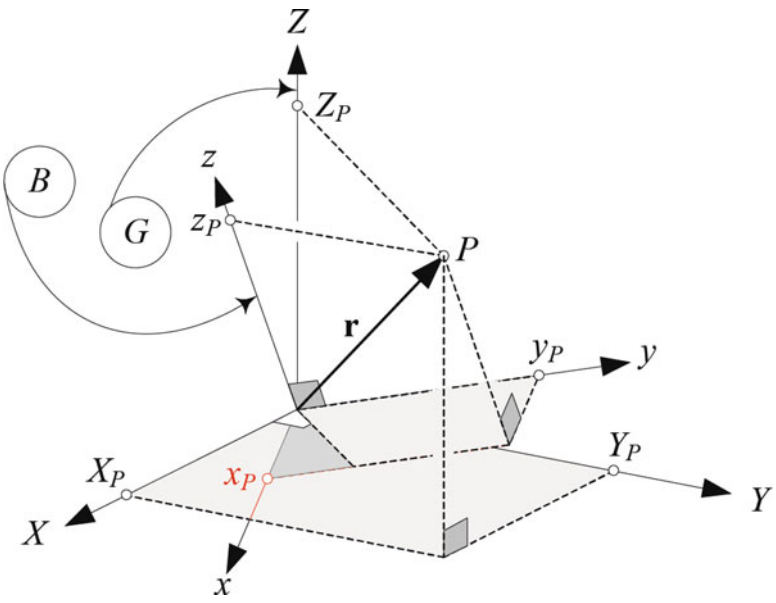



**Fig. 2.3** The local coordinate frame $B(Oxyz)$ rotates about a fixed point $O$ with respect to the global coordinate frame $G(OXYZ)$

is global coordinate frame $G(OXYZ)$ which is motionless with the ground. The two frames, which have the common origin $O$, are employed to express a vector $\mathbf{r}$. There is always a rotation matrix ${}^{G}R_{B}$ to transform the coordinates of the vector $\mathbf{r}$ from the reference frame $B(Oxyz)$ to the other frame $G(OXYZ)$ (Jazar 2010).

$$ {}^{G}\mathbf{r} = {}^{G}R_{B}\,{}^{B}\mathbf{r} \tag{2.1} $$

where ${}^{B}\mathbf{r}$ and ${}^{G}\mathbf{r}$ are position vectors of a point $P$ expressed in the body coordinate frame and the global coordinate frame, respectively.

$$ {}^{B}\mathbf{r} = \begin{bmatrix} x_P \\ y_P \\ z_P \end{bmatrix};\ {}^{G}\mathbf{r} = \begin{bmatrix} X_P \\ Y_P \\ X_P \end{bmatrix};\ {}^{G}R_{B} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \tag{2.2} $$

More specifically, when the body rotates an angle $\delta$ about the axis denoted by its direction unit vector $\hat{u}$ and passing through the origin $O$, as illustrated in Fig. 2.4, the rotation matrix becomes *Rodriguez matrix* $R_{\hat{u},\delta}$ as shown in (2.3). This motion is also called axis-angle rotation.

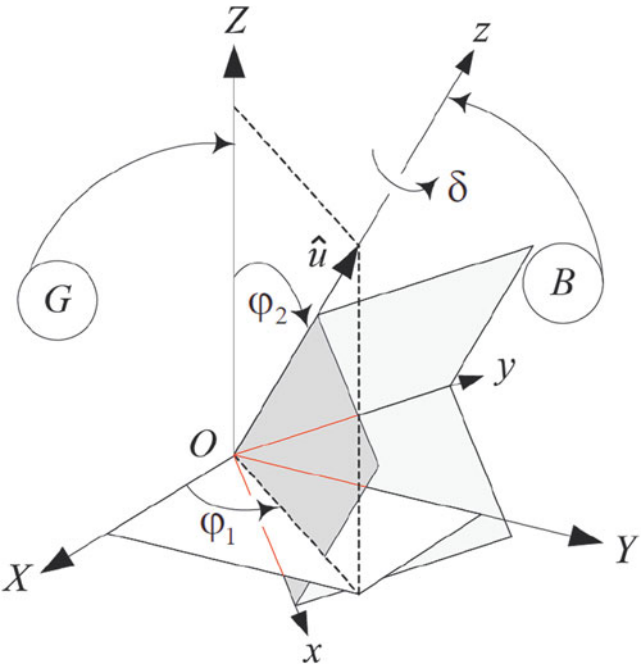



**Fig. 2.4** Axis-angle rotation

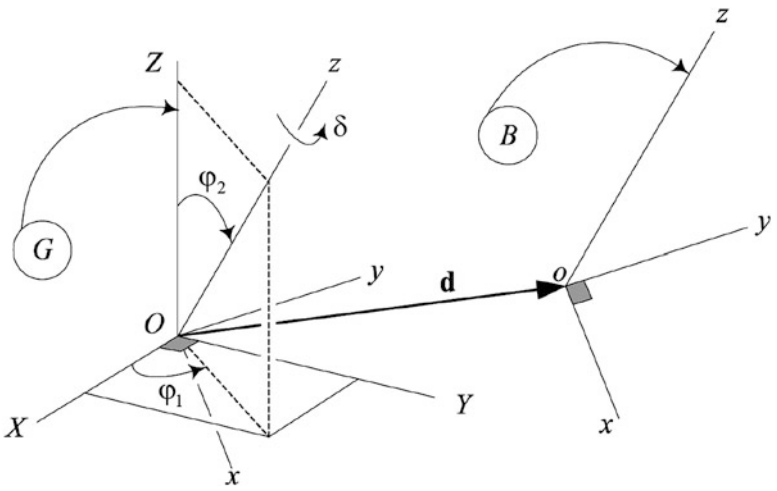



**Fig. 2.5** The local coordinate frame$B(Oxyz)$ rotates and translates in the global coordinate frame $G(OXYZ)$

$$
{}^{G}R_{B} = \begin{bmatrix} u_1^2\text{vers}\,\delta + \cos\delta & u_1u_2\text{vers}\,\delta - u_3\sin\delta & u_1u_3\text{vers}\,\delta + u_2\sin\delta \\ u_1u_2\text{vers}\,\delta + u_3\sin\delta & u_2^2\text{vers}\,\delta + \cos\delta & u_2u_3\text{vers}\,\delta - u_1\sin\delta \\ u_1u_3\text{vers}\,\delta - u_2\sin\delta & u_2u_3\text{vers}\,\delta + u_1\sin\delta & u_3^2\text{vers}\,\delta + \cos\delta \end{bmatrix} = R_{\hat{u},\delta} \tag{2.3}
$$

where $\text{vers}\,\delta = 1 - \cos\delta$, and $\hat{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$

Now the situation in Fig. 2.3 is made more complicated by giving the origin o of the body coordinate frame $B(oxyz)$ a free move in the global coordinate frame $G(OXYZ)$. In other words, the rigid body can rotate in the global frame, while its point $o$ can translate with respect to the origin $O$ of the frame $G$ as shown in Fig. 2.5.

Let us denote the position of the moving point $o$ relative to the fixed point $O$ as vector ${}^{G}\mathbf{d}$. Accordingly, the coordinates of a point $P$ in the local coordinate system can be matched with that in the global coordinate system by the following equation:

$$
{}^{G}\mathbf{r} = {}^{G}R_{B}\,{}^{B}\mathbf{r} + {}^{G}\,\mathbf{d} \tag{2.4}
$$

or

$$
\begin{bmatrix} X_P \\ Y_P \\ Z_P \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x_P \\ y_P \\ z_P \end{bmatrix} + \begin{bmatrix} X_o \\ Y_o \\ Z_o \end{bmatrix} = \begin{bmatrix} r_{11}x_P + r_{12}y_P + r_{13}z_P + X_o \\ r_{21}x_P + r_{22}y_P + r_{23}z_P + Y_o \\ r_{31}x_P + r_{32}y_P + r_{33}z_P + Z_o \end{bmatrix} \tag{2.5}
$$

${}^{G}\mathbf{d}$ is called the displacement or translation of $B$ with respect to $G$. ${}^{G}R_{B}$ is the rotation matrix to map ${}^{B}\mathbf{r}$ to ${}^{G}\mathbf{r}$ when ${}^{G}\mathbf{d} = 0$. Such a motion is called *rigid motion* (Jazar

2010). This is a combination of a rotation and a translation. In other words, the location of a rigid body can be described by the position of the origin $o$ and the orientation of the body frame in the global frame.

### 2.2.2 Homogeneous Transformation

If we add an element to the position vector and extend the rotation matrix as below:

$$ {}^{G}\mathbf{r}_P = \begin{bmatrix} X_P \\ Y_P \\ Z_P \end{bmatrix} \longrightarrow \begin{bmatrix} X_P \\ Y_P \\ Z_P \\ 1 \end{bmatrix}; \; {}^{B}\mathbf{r}_P = \begin{bmatrix} x_P \\ y_P \\ z_P \end{bmatrix} \longrightarrow \begin{bmatrix} x_P \\ y_P \\ z_P \\ 1 \end{bmatrix}; \; {}^{G}\mathbf{d} = \begin{bmatrix} X_o \\ Y_o \\ Z_o \end{bmatrix} \longrightarrow \begin{bmatrix} X_o \\ Y_o \\ Z_o \\ 1 \end{bmatrix} \tag{2.6} $$

$$ {}^{G}R_B = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \longrightarrow {}^{G}T_B = \begin{bmatrix} r_{11} & r_{12} & r_{13} & X_o \\ r_{21} & r_{22} & r_{23} & Y_o \\ r_{31} & r_{32} & r_{33} & Z_o \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} {}^{G}R_B & {}^{G}\mathbf{d} \\ 0 & 1 \end{bmatrix} \tag{2.7} $$

The following equation will be established:

$$ \begin{bmatrix} X_P \\ Y_P \\ Z_P \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & X_o \\ r_{21} & r_{22} & r_{23} & Y_o \\ r_{31} & r_{32} & r_{33} & Z_o \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_P \\ y_P \\ z_P \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11}x_P + r_{12}y_P + r_{13}z_P + X_o \\ r_{21}x_P + r_{22}y_P + r_{23}z_P + Y_o \\ r_{31}x_P + r_{32}y_P + r_{33}z_P + Z_o \\ 1 \end{bmatrix} \tag{2.8} $$

By comparing (2.5) with (2.8) we can come up with the conclusion that introducing such a new $4 \times 4$ matrix and the appended element to the position vector enables the rigid motion to be expressed by only a single matrix transformation as (2.8) or (2.9)

$$ {}^{G}\mathbf{r} = {}^{G}T_B {}^{B}\mathbf{r} \tag{2.9} $$

The new $4 \times 4$ matrix ${}^{G}T_B$ as shown in (2.7) is called *homogeneous transformation matrix*; and, the new expression ${}^{G}\mathbf{r}_P$ as a $4 \times 1$ vector in (2.6) are called *homogeneous coordinates* of point P, expressed in the global frame $G$.

It is interesting that the homogeneous transformation matrix can be decomposed into two other matrices, one as following:

$$ {}^{G}T_B = \begin{bmatrix} r_{11} & r_{12} & r_{13} & X_o \\ r_{21} & r_{22} & r_{23} & Y_o \\ r_{31} & r_{32} & r_{33} & Z_o \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & X_o \\ 0 & 1 & 0 & Y_o \\ 0 & 0 & 1 & Z_o \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & 0 \\ r_{21} & r_{22} & r_{23} & 0 \\ r_{31} & r_{32} & r_{33} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = {}^{G}D_B {}^{G}R_B \tag{2.10} $$

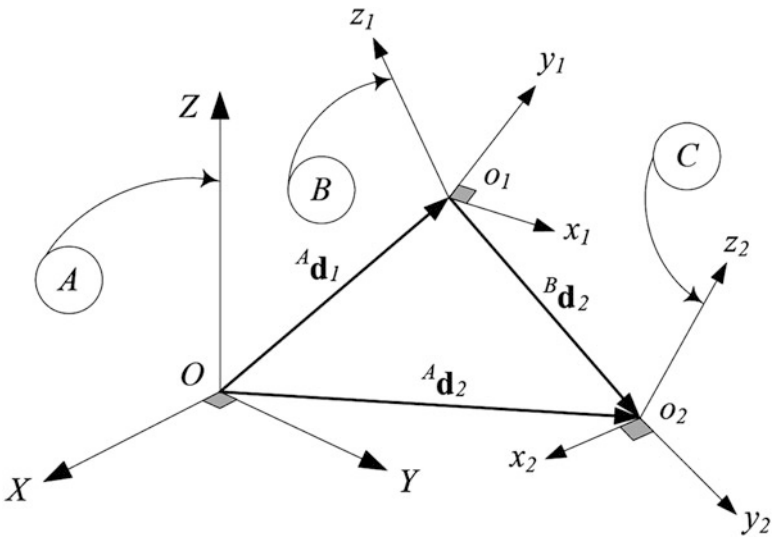



**Fig. 2.6** Compounding two homogeneous transformations

As mathematically expressed, a transformation can be achieved by a pure rotation first, followed by a pure translation.

Two homogeneous transformations can also be compounded with each other as one homogeneous transformation. Consider a situation in which there are reference frames: $A$, $B$, and $C$, as shown in Fig. 2.6.

If the matrices to transform coordinates from the frame $B$ to the frame $A$, and from the frame $C$ to the frame $B$, respectively, are:

$$ {}^A T_B = \begin{bmatrix} {}^A R_B & {}^A\mathbf{d}_1 \\ 0 & 1 \end{bmatrix} \tag{2.11} $$

$$ {}^B T_C = \begin{bmatrix} {}^B R_C & {}^B\mathbf{d}_2 \\ 0 & 1 \end{bmatrix} \tag{2.12} $$

The transformation matrix from the frame $C$ to the frame $A$ will be:

$$ \begin{aligned} {}^A T_C = {}^A T_B {}^B T_C &= \begin{bmatrix} {}^A R_B & {}^A\mathbf{d}_1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} {}^B R_C & {}^B\mathbf{d}_2 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} {}^A R_B {}^B R_C & {}^A R_B {}^B\mathbf{d}_2 + {}^A\mathbf{d}_1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} {}^A R_C & {}^A\mathbf{d}_2 \\ 0 & 1 \end{bmatrix} \end{aligned} \tag{2.13} $$

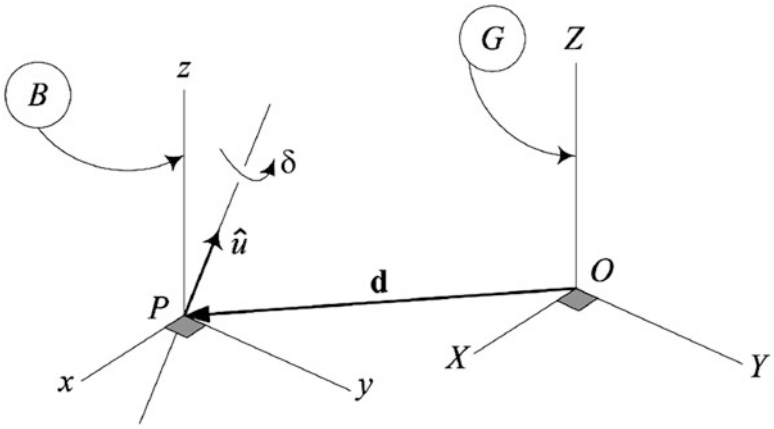



**Fig. 2.7** Rotation about an axis not going through origin

### *2.2.3 A Rotation about an Axis Not Going Through Origin*

The homogeneous transformation matrix can represent rotations about an axis going through a point different from the origin of the frames. Figure 2.7 shows an axis with its direction unit vector $\hat{u}$ passing through point $P$ that denoted by position vector $\mathbf{d}$. In order to express the transformation, a local frame $B$ is set at $P$ that is parallel to the $G$ frame. By doing so, the rotation can be expressed as a translation along $-\mathbf{d}$, followed by a rotation about $\hat{u}$ and a translation $\mathbf{d}$ (Jazar 2010).

$$ {}^{G}T_{B} = D_{\hat{d},d} R_{\hat{u},\delta} D_{\hat{d},-d} = \begin{bmatrix} I & \mathbf{d} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} R_{\hat{u},\delta} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} I & -\mathbf{d} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} R_{\hat{u},\delta} & \mathbf{d} - R_{\hat{u}}\mathbf{d} \\ 0 & 1 \end{bmatrix} \tag{2.14} $$

## 2.3 Steering Motion

To express steering motion of the tire as the homogeneous transformation, different coordinate systems are utilized. The steering axis is also determined.

### *2.3.1 Coordinate Frames*

In this sections, the four following types of coordinate system are defined.

Body coordinate frame $B(Cxyz)$ is the frame attached to the vehicle at the mass center C, as depicted in Fig. 2.8. The $x$-axis is longitudinal, going through $C$, and

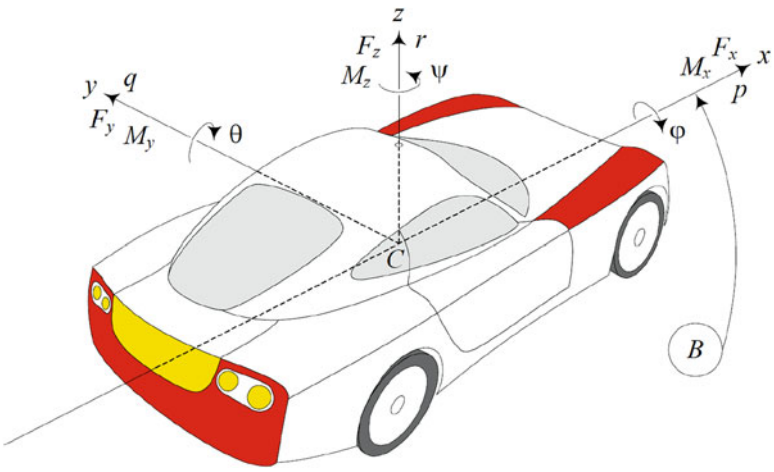



**Fig. 2.8** Body coordinate frame $B(Cxyz)$

directed forward. The $y$-axis goes laterally to the left of the drive's viewpoint. The $z$-axis is perpendicular to flat horizontal ground and upward. Therefore, it makes the coordinate system a right-hand triad. It is also called B-frame.

To define coordinate systems at wheels with mathematic consistency the wheels are numbered as follow: the front left wheel is wheel number 1; the front right wheel is number 2 ; the rear right wheel is number 3; and wheel number 4 is for the rear left wheel. Without loss of generality, we also assume that when the wheel is not steering the wheel is in straight position so it has zero initial camber and toe. Having identified by the numbers and accepting the assumption, the following coordinate frames, as shown in Fig. 2.9, are defined.

Wheel coordinate frame $W_i(x_{wi}, y_{wi}, z_{wi})$, or $W_i$-frame is the frame that its origin attached to the center of wheel number $i$. When the wheel is not steering, the $x_{wi}$-axis is parallel to the ground plane and directed forward; the $z_{wi}$ is vertical to the ground and upward. The $y_{wi}$ is always the spin axis and makes the coordinate frame a right-hand triad. $(x_{wi}W_iz_{wi})$ plane is coincident with the $i$th tire plane which made by centrally narrowing the $i$th wheel into a flat disk. The $W_i$ frame follows every motions of the wheel except spin motion transmitted from the drive line.

When the wheel number $i$ is not steering, a wheel-body coordinate system $C_i(x_{ci}, y_{ci}, z_{ci})$ that initially coincident with the $W_i$-frame is attached to the car body. The wheel-body coordinate frame is motionless with the car body and so it does not follows any motion of the wheel when steering.

We also define a tire coordinate frame $T_i(x_{ti}, y_{ti}, x_{ti})$ attached to the $i$th tireprint center. The $x_{ti}$ axis is the intersection of the flat ground and the corresponding tire plane. The $z_{ti}$ axis is always vertical to the ground and upward; and $y_{ti}$ makes the system a right-hand triad. This frame solely follows steering motion of the wheel. The orientations and locations of the three coordinate frames will not be affected by the spin motion of the wheel.

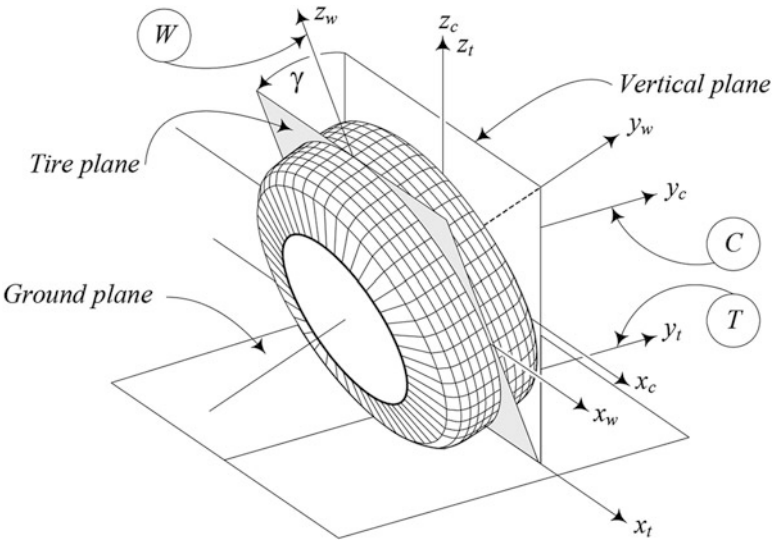



**Fig. 2.9** Wheel coordinate frame $W$, wheel-body coordinate frame $C$, and tire coordinate frame $T$

### *2.3.2 Steering and Homogeneous Transformation*

In this investigation, we assume that the steering axis is a fixed line with respect to car body. Thus, steering a tire (there is no difference between tire and wheel hereafter) about the pivot is considered as relative motions of the tire to the car body; and it is convenient to express the steering axis in the wheel-body coordinate frame.

Owing to different purposes the steering axis is generally not vertical to the ground. Its orientation is represented by caster angle $\phi$ and kingpin inclination angle $\theta$ (also called steering axis inclination angle - *SAI* angle, or lean angle), as illustrated in Fig. 2.10. In order to find the direction unit vector of it, the steering pivot is considered as the intersection line of two planes: caster plane $\pi_c$ and *KPI* plane $\pi_I$. The caster plane is a plane that has an angle $\theta$ with $(z_c, x_c)$ plane; the *KPI* plane has an angle $\phi$ with $(y_c, z_c)$ plane; and both of them contain steering axis. Thus, the two planes can be denoted by their normal unit vectors in the wheel-body coordinate frame, respectively:

$$\hat{n}_{\pi c} = \begin{bmatrix} 0 \\ \cos\theta \\ \sin\theta \end{bmatrix}; \quad \hat{n}_{\pi I} = \begin{bmatrix} -\cos\phi \\ 0 \\ \sin\phi \end{bmatrix} \tag{2.15}$$

Therefore, the unit direction vector of the steering axis is:

$$\hat{u} = \frac{\hat{n}_{\pi c} \times \hat{n}_{\pi I}}{|\hat{n}_{\pi c} \times \hat{n}_{\pi I}|} \tag{2.16}$$

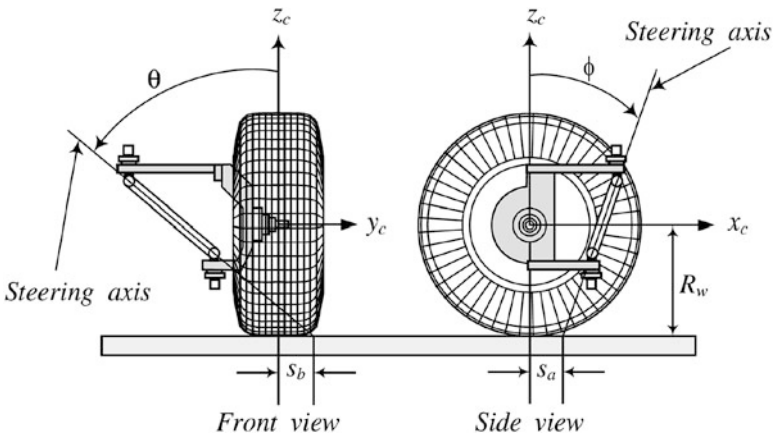



**Fig. 2.10** The orientation and location of the kingpin axis for wheel number 1

or

$$ {}^{C}\hat{u} = \frac{1}{\sqrt{\cos^2\phi + \cos^2\theta \sin^2\phi}} \begin{bmatrix} \cos\theta \sin\phi \\ -\sin\theta \cos\phi \\ \cos\theta \cos\phi \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \tag{2.17} $$

To define the axis an arbitrary point on it is required. We choose this point as the intersection between the axis and the ground: $P(s_a, s_b, -R_w)$.

With the previously defined coordinate frames and steering axis, a steering motion can now be mathematically equivalent to a rotation $\delta$ of the $W$ frame about the steering axis denoted by its direction unit vector $\hat{u}$ and the point $P$ fixed in the $C$ frame. Within a practical range of caster and lean angle, i.e., $\theta, \phi \in (-90°; 90°)$, the steering angle is positive when the wheel steers to the left. Since the homogeneous transformation is applicable to mapping coordinates in one frame onto another that rotates about an axis not going through the origin of the two initially coincident frames, it is employed to transfer coordinates between the two frames. Mapping coordinates in the wheel coordinate frame onto the wheel-body coordinate frame is governed by the following transformation:

$$ {}^{C}r = {}^{C}T_{W}\,{}^{W}r \tag{2.18} $$

where ${}^{C}r$ and ${}^{W}r$ are homogeneous representations of a position vector in the $C$ frame and the $W$ frame, respectively.

$$ {}^{C}r = \begin{bmatrix} {}^{W}x_P \\ {}^{W}y_P \\ {}^{W}z_P \\ 1 \end{bmatrix}; \quad {}^{W}r = \begin{bmatrix} {}^{C}x_P \\ {}^{C}y_P \\ {}^{C}z_P \\ 1 \end{bmatrix} \tag{2.19} $$

and ${}^{C}T_{w}$ is a $4 \times 4$ homogeneous matrix:

$$ {}^{C}T_{w} = \begin{bmatrix} R_{\hat{u},\delta} & d_P - R_{\hat{u},\delta} d_P \\ 0 & 1 \end{bmatrix} \tag{2.20} $$

where $d_P$ is position vector of the point $P$ expressed in the $C$ frame.

$$ {}^{C}d_P = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} = \begin{bmatrix} s_a \\ s_b \\ -R_w \end{bmatrix} \tag{2.21} $$

## 2.4 Kinematics of a Steered Tire

The transformation enables us to find coordinates of any point of the tire expressed in the wheel-body coordinate frame. Kinematic parameters of the steerable tire are extracted. Simplified designs are also presented.

### *2.4.1 The Camber Angle*

Camber is determined as the angle $\gamma$ that the tire plane has rotated about the $+x_t$ axis from the vertical position. Camber is positive when the wheel leans to the right as viewed from the rear regardless left or right wheels. The easiest way to find camber using the transformation is through calculating the angle $\rho$, as depicted in Fig. 2.11, between the normal vectors of the tire plane and the ground plane as the following equation:

$$ \gamma = \frac{\pi}{2} - \rho \tag{2.22} $$

If the unit vectors in the direction of $x_c$, $y_c$, and $z_c$ of the $C$ frame are denoted by $\hat{I}$, $\hat{J}$, and $\hat{K}$; and, the unit vectors in the direction of $x_w$, $y_w$, and $z_w$ of the $W$ frame are symbolized by $\hat{i}$, $\hat{j}$, and $\hat{k}$. Then $\rho$ will be:

$$ \rho = a \cos \frac{{}^{C}\hat{j} \cdot {}^{C}\hat{K}}{\left| {}^{C}\hat{j} {}^{C}\hat{K} \right|} \tag{2.23} $$

where

$$ {}^{C}\hat{K} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \tag{2.24} $$

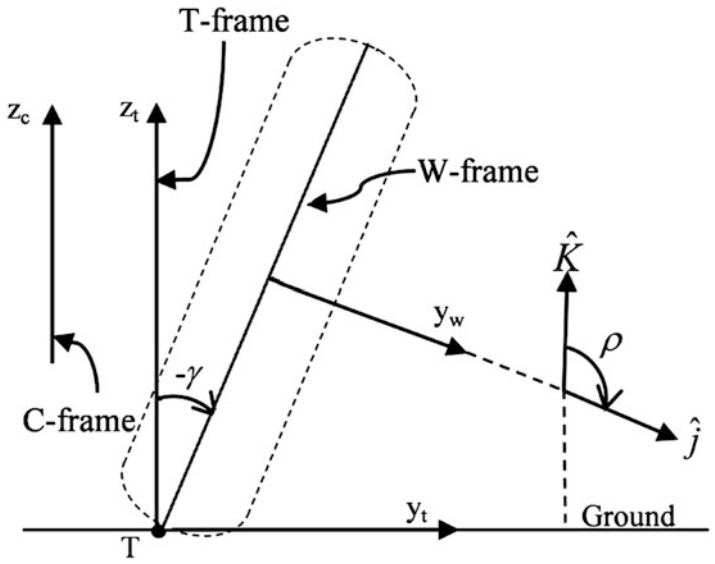



**Fig. 2.11** The front view of the steered wheel with camber sign convention

$$ {}^{c}\hat{j} = {}^{c}T_{W}\,{}^{W}\hat{j} = \begin{bmatrix} R_{\hat{u},\delta} & d_P - R_{\hat{u},\delta} d_P \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} u_1 u_2 (1 - \cos\delta) - u_3 \sin\delta \\ u_2^2 (1 - \cos\delta) + \cos\delta \\ u_2 u_3 (1 - \cos\delta) + u_1 \sin\delta \\ 0 \end{bmatrix} \tag{2.25} $$

By substituting (2.24) and (2.25) in (2.23) and then in (2.22), we have:

$$ \gamma = \frac{\pi}{2} - a\cos[u_2 u_3 (1 - \cos\delta) + u_1 \sin\delta] \tag{2.26} $$

Substitute (2.17) in (2.26), the camber of a steered wheel will be:

$$ \gamma = \frac{\pi}{2} - a\cos\left[ \frac{\cos\theta \sin\phi}{\sqrt{\cos^2\phi + \cos^2\theta \sin^2\phi}} \sin\delta - \frac{\cos^2\phi \sin\theta \cos\theta}{\cos^2\phi + \cos^2\theta \sin^2\phi} \mathrm{vers}\,\delta \right] \tag{2.27} $$

Using the camber function we can verify that the wheel is cambered by steering around a tilted kingpin axis, as presented in Section 2.5.1. Furthermore, apart from steering angle this camber only depends on the orientation of the steering axis ($\phi$ and $\theta$); it is independent of the location of the axis ($s_a$, $s_b$). In other words, in order to have a particular value of camber, the only thing we need to vary/control is steering axis orientation.

### 2.4.2 The Wheel Center

The homogeneous coordinates of the wheel center in the $W$ frame are:

$$ {}^{W}r_{W} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \tag{2.28} $$

They are transformed into the homogeneous coordinates expressed in the $C$ frame as:

$$ {}^{C}r_{W} = {}^{C}T_{w}\,{}^{W}r_{W} = \begin{bmatrix} R_{\hat{u},\delta} & d_P - R_{\hat{u},\delta}d_P \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} d_P - R_{\hat{u},\delta}d_P \\ 1 \end{bmatrix} \tag{2.29} $$

Substitute (2.3) and (2.21) in (2.29), we have:

$$ {}^{C}r_{W} = \begin{bmatrix} d_1(1-u_1^2)\text{vers}\,\delta - u_1(d_2u_2 + d_3u_3)\text{vers}\,\delta + (d_2u_3 - d_3u_2)\sin\delta \\ d_2(1-u_2^2)\text{vers}\,\delta - u_2(d_3u_3 + d_1u_1)\text{vers}\,\delta + (d_3u_1 - d_1u_3)\sin\delta \\ d_3(1-u_3^2)\text{vers}\,\delta - u_3(d_1u_1 + d_2u_2)\text{vers}\,\delta + (d_1u_2 - d_2u_1)\sin\delta \end{bmatrix} \tag{2.30} $$

If we substitute (2.17) in (2.30), then the coordinates of the steered wheel center in the wheel-body frame will be:

$$ {}^{C}x_{W} = \left[ s_a + \frac{1}{4}\frac{(2R_w\sin 2\phi - 4s_a\sin^2\phi)\cos^2\theta + s_b\sin 2\theta\sin 2\phi}{\cos^2\phi + \cos^2\theta\sin^2\phi} \right]\text{vers}\,\delta + \frac{s_b\cos\theta - R_w\sin\theta}{\sqrt{\cos^2\phi + \cos^2\theta\sin^2\phi}}\cos\phi\sin\delta \tag{2.31} $$

$$ {}^{C}y_{W} = \left[ s_b - \frac{1}{4}\frac{(2R_w\sin 2\theta + 4s_a\sin^2\theta)\cos^2\phi - s_a\sin 2\theta\sin 2\phi}{\cos^2\phi + \cos^2\theta\sin^2\phi} \right]\text{vers}\,\delta - \frac{s_a\cos\phi + R_w\sin\phi}{\sqrt{\cos^2\phi + \cos^2\theta\sin^2\phi}}\cos\theta\sin\delta \tag{2.32} $$

$$ {}^{C}z_{W} = \left[ -R_w + \frac{1}{2}\frac{(2R_w\cos^2\theta + s_b\sin 2\theta)\cos^2\phi - s_a\cos^2\theta\sin 2\phi}{\cos^2\phi + \cos^2\theta\sin^2\phi} \right]\text{vers}\,\delta - \frac{s_a\cos\phi\sin\theta + s_b\cos\theta\sin\phi}{\sqrt{\cos^2\phi + \cos^2\theta\sin^2\phi}}\sin\delta \tag{2.33} $$

${}^{C}x_W$, ${}^{C}y_W$, and ${}^{C}z_W$ indicate how much the wheel center longitudinally, laterally, and vertically displaces from the point $C$ fixed to the car body, respectively. The ${}^{C}z_W$ is required to determine the car body's vertical displacement as will be discussed in Sect. 2.4.3.

### 2.4.3 The Car Body Displacement

Having camber angle $\gamma$ and ${}^{C}z_W$ is enough to define the ${}^{C}z_T$-coordinate of tire-print center as the following equation:

$$\frac{{}^{C}z_W - {}^{C}z_T}{R_w} = \cos\gamma \Leftrightarrow {}^{C}z_T = {}^{C}z_W - R_w \cos\gamma \tag{2.34}$$

By assuming the ground is flat and absolutely rigid; and hence, the center of tireprint is always on the ground; the origin of $C$-frame on car body will experience a displacement of $H$ as the following formula:

$$H = -{}^{C}z_T - R_w = R_w(\cos\gamma - 1) - {}^{C}z_W \tag{2.35}$$

Substituting (2.33) and (2.27) in (2.35) yields:

$$\begin{aligned} H = {} & R_w \sin\left[a\cos\left(\frac{\cos\theta\sin\phi}{\sqrt{\cos^2\phi + \cos^2\theta\sin^2\phi}}\sin\delta - \frac{\cos^2\phi\sin\theta\cos\theta}{\cos^2\phi + \cos^2\theta\sin^2\phi}\mathrm{vers}\,\delta\right)\right] \\ & - R_w\cos\delta - \frac{1}{2}\frac{(2R_w\cos^2\theta + s_b\sin 2\theta)\cos^2\phi - s_a\cos^2\theta\sin 2\phi}{\cos^2\phi + \cos^2\theta\sin^2\phi}\mathrm{vers}\,\delta \\ & + \frac{s_a\cos\phi\sin\theta + s_b\cos\theta\sin\phi}{\sqrt{\cos^2\phi + \cos^2\theta\sin^2\phi}}\sin\delta \end{aligned} \tag{2.36}$$

It can be seen from the formula that, $H$ is a function of the orientation, location of the steering axis as well as tire radius.

### 2.4.4 Simplified Designs

The kinematics of the steered wheel for some simplified cases can be derived from the general one. Below are some simplified situations that may be applied to the real vehicle for different situations.

#### 2.4.4.1 Ideal Wheel

Ideal wheel is the wheel that has the steering axis coincident with $z_w$. Such the wheel has $\phi = 0$, $\theta = 0$, $s_a = 0$, and $s_b = 0$, and experiences the minimum tire wear and steering effort.

From (2.31) to (2.33), the coordinates of the wheel's center of the case turn into:

$$x_w = 0; y_w = 0; z_w = 0 \tag{2.37}$$

The car body displacement remains zero:

$$H = 0 \tag{2.38}$$

while (2.27) shows that, camber for this configuration is always zero.

#### 2.4.4.2 Shopping Cart Wheel

In order to provide shopping cart wheel with self-steering ability when pulled, the steering pivot of the wheel is not coincident with $z_w$ although it is usually perpendicular to the ground. The wheel has $\phi = 0$, $\theta = 0$, $s_a \neq 0$, and $s_b \neq 0$. Substituting the above orientation and location of the steering kingpin, the coordinates can be simplified as below:

$$x_w = s_a(1 - \cos\delta) + s_b \sin\delta \tag{2.39}$$

$$y_w = s_a(1 - \cos\delta) - s_b \sin\delta \tag{2.40}$$

$$z_w = 0 \tag{2.41}$$

The car body displacement and camber are:

$$H = 0; \gamma = 0 \tag{2.42}$$

#### 2.4.4.3 $s_a = 0, s_b = 0, \phi \neq 0, \theta \neq 0$

The coordinates of the wheel's center are:

$$^{C}x_W = R_w \left[ \frac{\sin\phi \cos\phi \cos^2\theta}{\cos^2\phi + \cos^2\theta \sin^2\phi} \operatorname{vers}\delta + \frac{\sin\theta \cos\phi}{\sqrt{\cos^2\phi + \cos^2\theta \sin^2\phi}} \sin\delta \right] \tag{2.43}$$

$$^{c}y_W = -R_w \left[ \frac{\sin\theta \cos\theta \cos^2\phi}{\cos^2\phi + \cos^2\theta \sin^2\phi} \text{vers}\,\delta - \frac{\sin\phi \cos\theta}{\sqrt{\cos^2\phi + \cos^2\theta \sin^2\phi}} \sin\delta \right] \tag{2.44}$$

$$^{c}z_W = -R_w \left( \frac{\cos^2\phi \cos^2\theta}{\cos^2\phi + \cos^2\theta \sin^2\phi} - 1 \right) \text{vers}\,\delta \tag{2.45}$$

Car body displacement is:

$$H = R_w \sin\left[ a\cos\left( \frac{\cos\theta \sin\phi}{\sqrt{\cos^2\phi + \cos^2\theta \sin^2\phi}} \sin\delta - \frac{\cos^2\phi \sin\theta \cos\theta}{\cos^2\phi + \cos^2\theta \sin^2\phi} \text{vers}\,\delta \right) \right]$$
$$-R_w \cos\delta - \frac{R_w \cos^2\theta \cos^2\phi}{\cos^2\phi + \cos^2\theta \sin^2\phi} \text{vers}\,\delta \tag{2.46}$$

Camber of the steered wheel is the same as the general case as it is only dependent of caster and lean angles.

#### 2.4.4.4 Variable Caster and Lean Angles

Consider a practical configuration, as shown in Fig. 2.12, where a fixed point $I$ on the steering axis with the coordinates $(0, s_{b0}, 0)$, is kept motionless with respect to

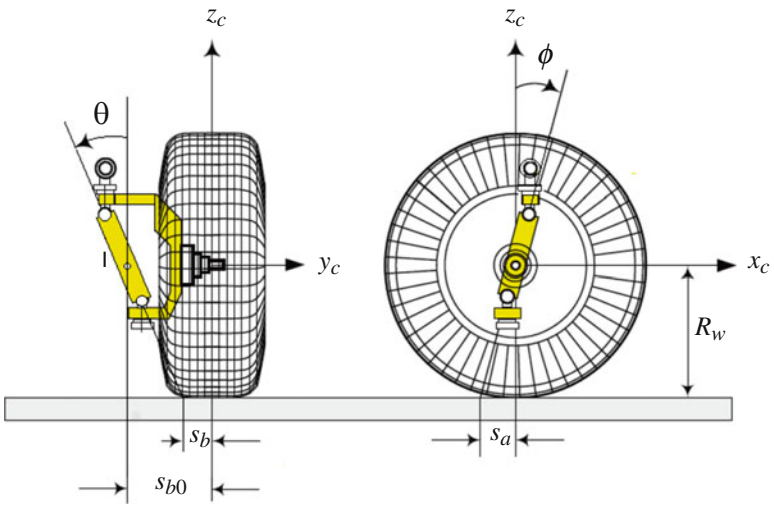



**Fig. 2.12** A practical variable caster and variable lean angle

the wheel-body frame; and, the steering axis rotates about the point to have different orientations. Then, $s_a$ and $s_b$ are functions of caster and lean angle, respectively.

$$s_a = R_w \tan\phi; s_b = s_{b0} + R_w \tan\theta \tag{2.47}$$

Some more specific cases can be derived from the above situation.

1. If $s_{b0} = 0$, then the wheel center's coordinates are:

$$x_w = 0; y_w = 0; z_w = 0 \tag{2.48}$$

and the car body will experience a vertical displacement of:

$$H = R_w \sin\left[a\cos\left(\frac{\cos\theta\sin\phi}{\sqrt{\cos^2\phi + \cos^2\theta\sin^2\phi}}\sin\delta - \frac{\cos^2\phi\sin\theta\cos\theta}{\cos^2\phi + \cos^2\theta\sin^2\phi}\text{vers}\,\delta\right) - 1\right] \tag{2.49}$$

It is clear that such the configuration lowers the sprung mass when the wheel is steered away from the straight position.

2. If caster is zero, $\phi = 0$, then

$$x_w = s_{b0}\cos\theta\sin\delta \tag{2.50}$$

$$y_w = s_{b0} + R_w\tan\theta - R_w\sin\theta\cos\theta\text{vers}\,\delta \tag{2.51}$$

$$z_w = s_{b0}\sin\theta\cos\theta\text{vers}\,\delta \tag{2.52}$$

$$H = R_w\sin[a\cos(\sin\theta\cos\theta\text{vers}\,\delta)] - R_w - s_{b0}\sin\theta\cos\theta\text{vers}\,\delta \tag{2.53}$$

3. If lean angle is zero, $\theta = 0$, then

$$x_w = s_{b0}\cos\phi\sin\delta \tag{2.54}$$

$$y_w = s_{b0}(1 - \cos\delta) \tag{2.55}$$

$$z_w = -s_{b0}\sin\phi\sin\delta \tag{2.56}$$

$$H = R_w\sin[a\cos(\sin\phi sin\delta)] - R_w + s_{b0}\sin\phi\sin\delta \tag{2.57}$$

(2.53) and (2.57) indicate that car body can rise or drop as long as appropriate $s_{b0}$ is utilized. This feature is interesting as we can choose $s_{b0}$ independently with the orientation of the steering pivot.

## 2.5 Variable Caster Steering

In order to choose an angle between caster and lean to control camber, a comparison is made. A kinematic analysis is also presented to prove the feasibility of a variable caster mechanism.

### *2.5.1 A Comparison Between Caster and Lean Angle Based on Camber Generation*

Since the goal of the variable orientation is generating camber to adjust total lateral force, so whether a variable caster or KPI angle is the main source of the generated camber needs to be realized. To compare how much camber is generated by caster or by KPI angle when steering, the two following simplest configurations are considered:

In the first case: a variable KPI with zero caster. Substitute $\phi = 0$ in (2.27), we have:

$$\gamma = \frac{\pi}{2} - a\cos[\sin\theta\cos\theta(\cos\delta - 1)] \tag{2.58}$$

The second case: a variable caster with zero KPI. Substituting $\theta = 0$ in (2.27) yields:

$$\gamma = \frac{\pi}{2} - a\cos(\sin\phi\sin\delta) \tag{2.59}$$

The following are graphs illustrating how much camber of a steerable wheel generated by steering the wheel about the titled axis for different caster or *KPI* angles.

As can be seen from the graphs in Figs. 2.13, 2.14, 2.15, 2.16, 2.17, and 2.18, caster creates much more camber, for instant roughly ten times, than lean angle does for the same steering angle. Another striking feature is that a negative caster (the top titled rearward) always produces camber facilitating a turn regardless left or right wheel. Indeed, when the wheel is steered to the left ($\delta > 0$) about the axis having negative caster, the generated camber is negative (the wheel leans to the left). When the wheel is steered to the right ($\delta < 0$) with the same caster, the generated camber is positive (the wheel leans to the right). Furthermore, for any specific steering angle the generated camber can be positive or negative with a proper caster angle. This will enable us to choose an appropriate manner for varying caster if we not only increase but also decrease camber to satisfy a particular criterion of controlling problem. Thus, it is much better to use caster as a variable parameter when it comes to the generated-camber-view.

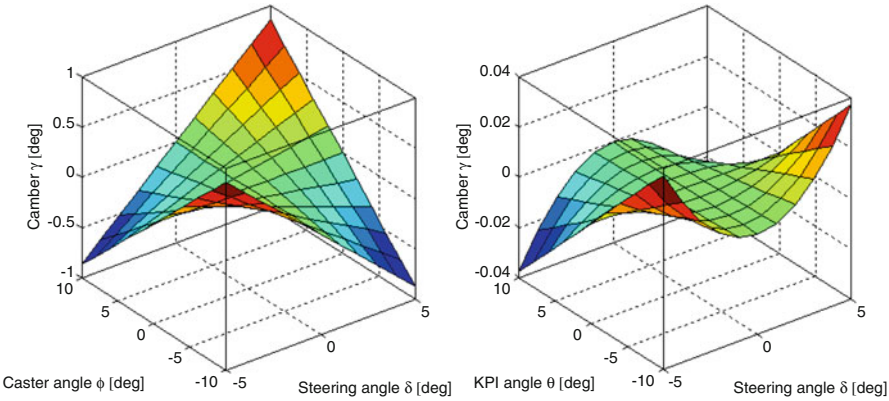



**Fig. 2.13** The generated camber for $\theta, \phi \in [-10°; 10°], \delta \in [-5°; 5°]$

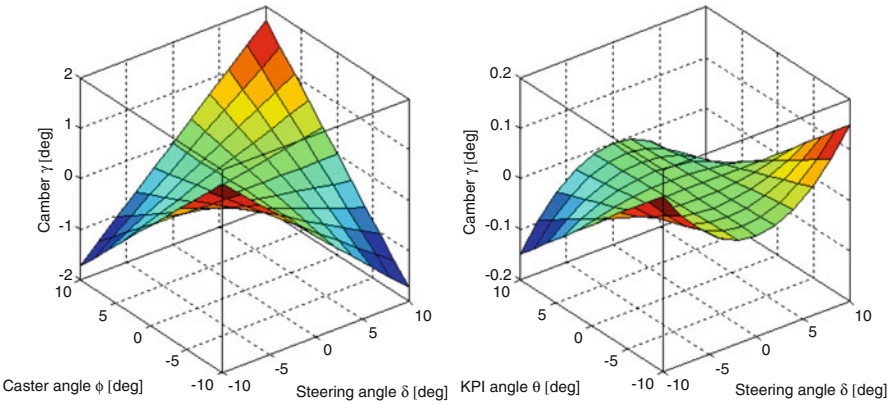



**Fig. 2.14** The generated camber for $\theta, \phi \in [-10°; 10°], \delta \in [-10°; 10°]$

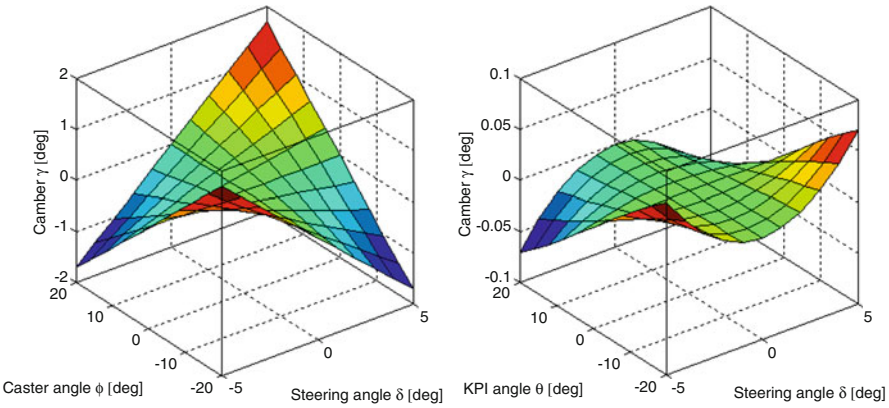



**Fig. 2.15** The generated camber for $\theta, \phi \in [-20°; 20°], \delta \in [-5°; 5°]$

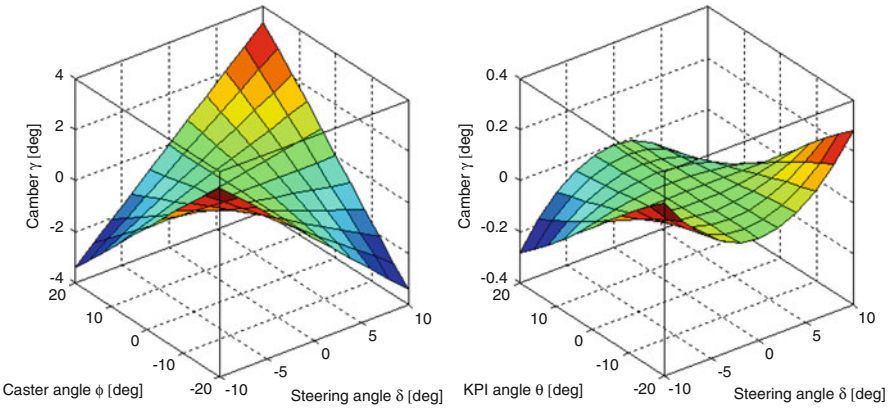



**Fig. 2.16** The generated camber for $\theta, \phi \in [-20°; 20°], \delta \in [-10°; 10°]$

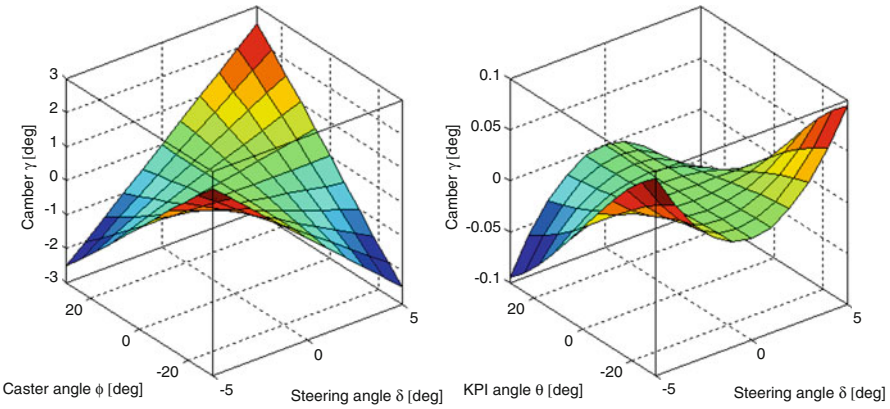



**Fig. 2.17** The generated camber for $\theta, \phi \in [-30°; 30°], \delta \in [-5°; 5°]$

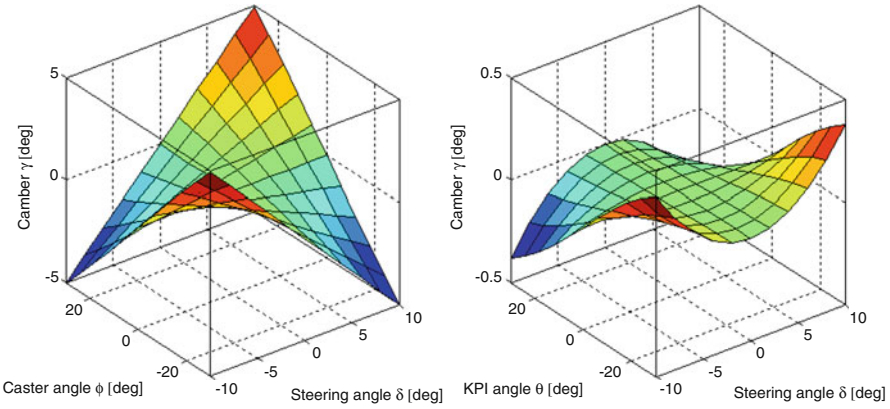



**Fig. 2.18** The generated camber for $\theta, \phi \in [-30°; 30°], \delta \in [-10°; 10°]$

### 2.5.2 Variable Caster and Car Body Displacement

The earlier results have proved that it is much better to employ variable caster than lean angle to gain camber of a steered wheel. However, steering about such a pivot leads to a raise or drop of car body with respect to the road (drop or raise the wheel compared to the car body, respectively), which is relating to self-centering ability and straight line stability of the car, especially at low speed. More specifically, it is desirable to have a steering axis that raise car body when the wheel is steered away, both left and right, from the straight position. Let us examine how much the car body is displaced vertically when using such a variable caster mechanism. When the wheel radius $R_w$ and lean angle are constant, the car body displacement is a function of variables, namely $\phi$, $\delta$, $s_a$, and $s_b$. The dependence on the different parameters provides the variety of options that can be used to adjust the displacement. Here, we only investigate the displacement in the most practical case when lean angle is kept zero. Caster $\phi$ is the only one made controllable, as shown in Fig. 2.19. As a result, (2.36) becomes:

$$H = R_w \sin[a \cos(\sin\phi \sin\delta)] - R_w + s_b \sin\phi \sin\delta \tag{2.60}$$

Below are how much the car body displaces when the wheel is steered with variable caster for different $s_b$ (Figs. 2.20, 2.21, 2.22, 2.23, and 2.24).

As illustrated by the graphs if the sign of caster is unchanged, saying negative, the displacement will never be able to reach a minimum at the straight position of the wheel (not steering) for a constant value $s_b$. The situation is even worse for a zero $s_b$ when the car body always falls while steering away from the middle

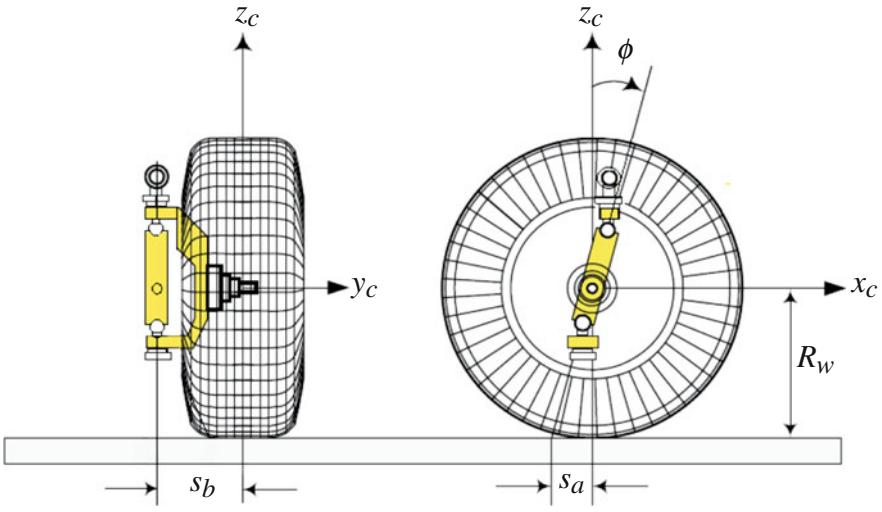



**Fig. 2.19** A practical variable configuration

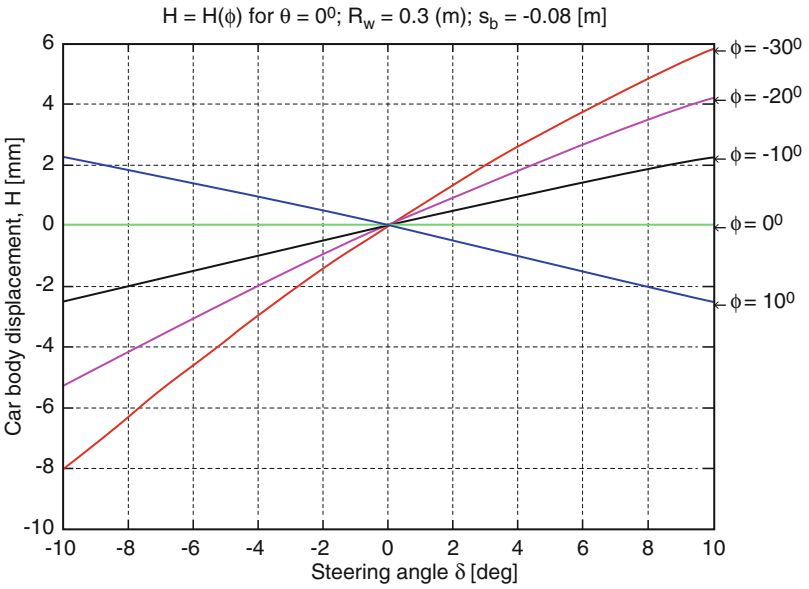



**Fig. 2.20** Vertical displacement of the car body for $s_b = -0.08$ m

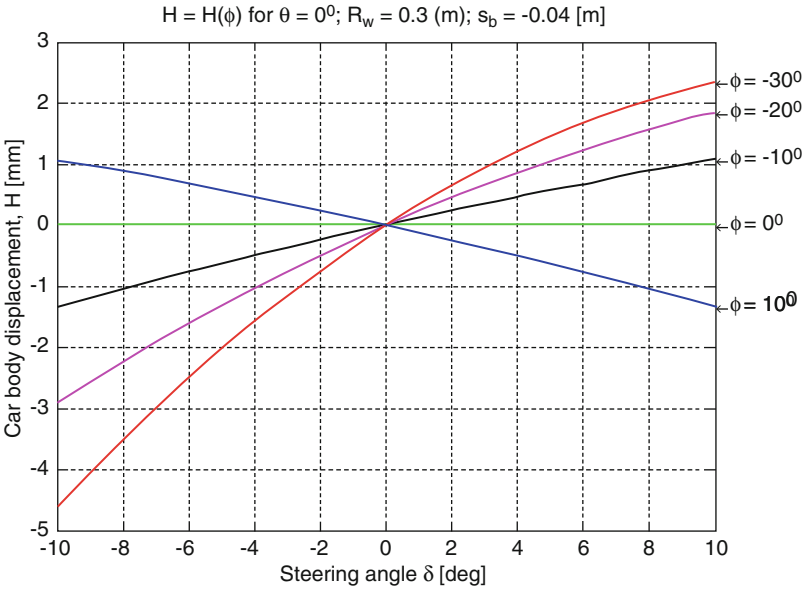



**Fig. 2.21** Vertical displacement of the car body for $s_b = -0.04$ m

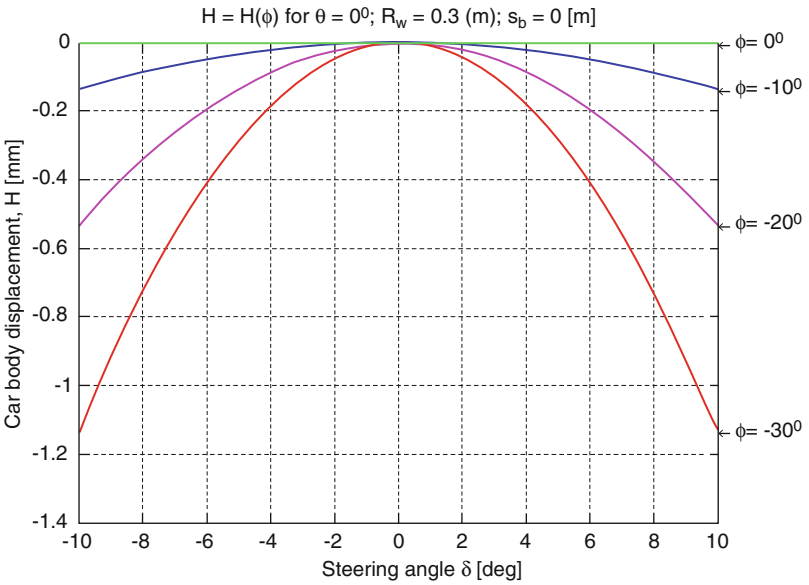



**Fig. 2.22** Vertical displacement of the car body for $s_b = 0\,\mathrm{m}$

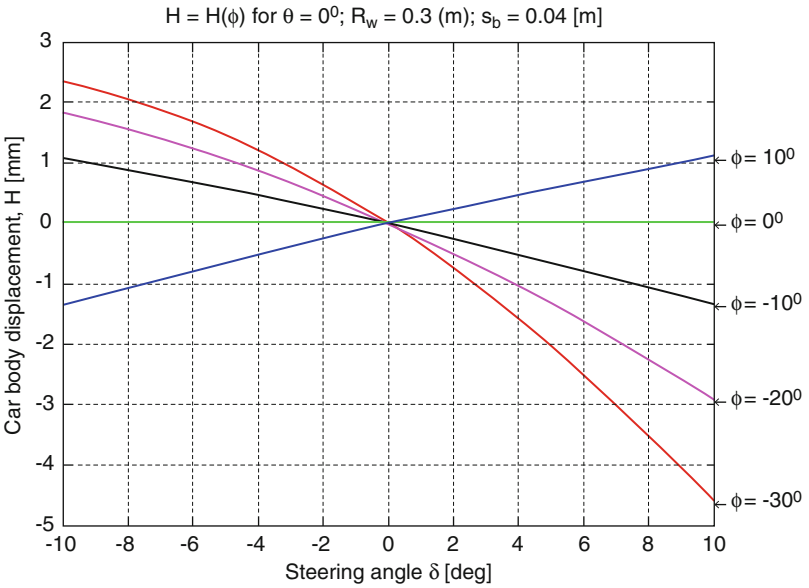



**Fig. 2.23** Vertical displacement of the car body for $s_b = 0.04\,\mathrm{m}$

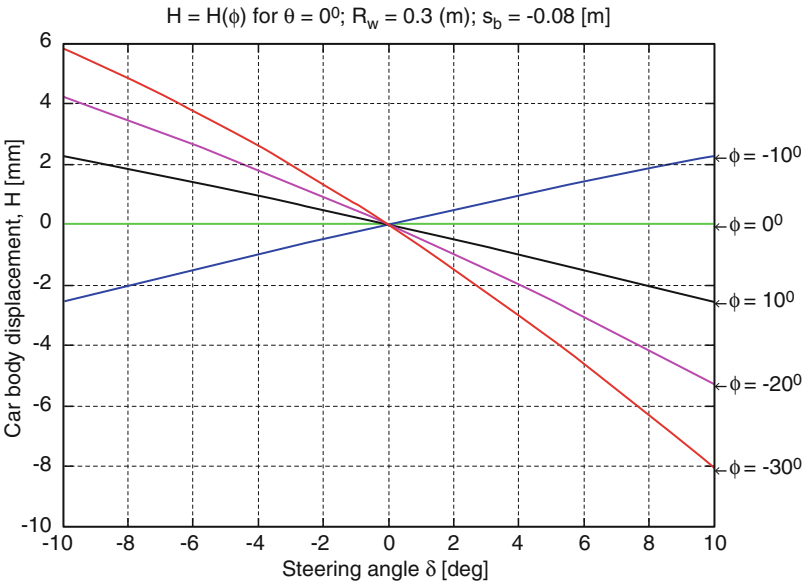



**Fig. 2.24** Vertical displacement of the car body for $s_b = 0.08$ m

position. This does not satisfy self-centering criterion. Fortunately, we can reverse the situation by providing a suitable variable $s_b$ to meet with the self-centering requirement. For example, if a negative caster is utilised to gain negative camber, the $s_b$ will need to be adjusted to negative for positive steering angle, and to positive for negative steering angle. Such a change will assure a desired displacement of car body when steering. This feature may be used and analyzed in future work when it comes to proposing a scheme of controlling caster.

## 2.6 Conclusion

A variation in camber is required to compensate the change in side-slip force during a turn of a car. For a steered wheel, camber can be produced by changing the caster or/and lean angle which are the representations of the steering axis orientation. In order to express the relationship between camber, caster, lean angle, and steering angle homogeneous transformation is employed with the assumptions that the tire is a flat disk; and the ground is rigid and flat. Having the relationship enabled us to compare between a variable caster and a variable lean angle based on different criteria. The comparison convinced that using caster as a variable/controllable parameter is much better and more practical than lean angle. From the examination, it is anticipated that an appropriate manner for varying caster together with a respective smart caster mechanism will be proposed to improve the safety, stability, and the manoeuvrability of the car.

## 2.7 Notations

| | |
|---|---|
| $C(x_c, y_c, z_c)$ | Wheel-body coordinate frame |
| $d_P$ | Position vector of the point $P$ |
| $H$ | The vertical displacement of point $C$ in car body |
| $\hat{I}, \hat{J}, \hat{K}$ | The unit vectors in the direction of $x_c$, $y_c$, $z_c$ of the $C$ frame |
| $\hat{i}, \hat{j}, \hat{k}$ | The unit vectors in the direction of $x_w$, $y_w$, $z_w$ of the $W$ frame |
| $KPI$ | Kingpin inclination |
| $L_b$ | Lateral offset at wheel center |
| $\pi_c$ | Caster plane |
| $\pi_I$ | KPI plane |
| $R_w$ | Tire radius |
| ${}^C r$ | Homogeneous representation of a position vector in the $C$ frame |
| ${}^W r$ | Homogeneous representation of a position vector in the $W$ frame |
| $R_{\hat{u},\delta}$ | Rodriguez rotation matrix |
| $s_a$ | Longitudinal location |
| $s_b$ | Lateral location |
| $s_{b\phi}$ | Lateral location as a function of $\phi$ |
| $T(x_t, y_t, z_t)$ | Tire coordinate frame |
| ${}^C T_w$ | Homogeneous matrix |
| $\hat{u}$ | The unit vector of the steering axis |
| ${}^C\hat{u}$ | The unit vector of the steering axis expressed in the $C$ frame |
| vers $\delta$ | $1 - \sin\delta$ |
| $W(x_w, y_w, z_w)$ | Wheel coordinate frame |
| $\gamma$ | Camber angle |
| $\delta$ | Steering angle |
| $\theta$ | Kingpin inclination angle |
| $\theta_0$ | Kingpin inclination angle associated with $\phi = 0$ |
| $\theta_\phi$ | Kingpin inclination angle associated with $\phi$ |
| $\rho$ | The angle between the normal vectors of tire plane and ground plane |
| $\phi$ | Caster angle |

## References

Jazar, R. N. (2010). *Theory of applied robotics: Kinematics, dynamics, and control* (2nd ed.). Boston, MA: Springer

Jazar, R. N. ( 2014). *Vehicle dynamics*. New York: Springer

Jazar, R. N., Subic, A., & Zhang, N. (2012). Kinematics of a smart variable caster mechanism for a vehicle steerable wheel. *Vehicle System Dynamics*, 50(12), 1861–1875

Milliken, W. F., Milliken, D. L., & Olley, M. (2002). *Chassis design : Principles and analysis*. Warrendale, PA: Society of Automotive Engineers

Reimpell, J., Stoll, H., & Betzler, J.W. (2000). *3-Wheel travel and elastokinematics*. Oxford: Butterworth-Heinemann

Wong, J. Y. (2001). *Theory of ground vehicles*. Hoboken, NJ: Wiley

# Chapter 3
# Design of Limaçon Gas Expanders

**Truong Phung, Ibrahim Sultan, and Alberto Boretti**

**Abstract** Limaçon is a simple and yet reliable technology, which can be employed to manufacture gas expanders to extract work or electrical power from low-grade heat or solar power resources; Limaçon technology can also be applied to small-scale power generation applications or can be used to improve energy efficiency of existing processes. Sultan (*Journal of Mechanical Design*, 787–793, 2006) and Sultan and Schaller (*Journal of Engineering for Gas Turbines and Power*, 2011) have come up with an optimum design of Limaçon-to-Limaçon expanders based on their thermodynamics performance; the design presented a model to calculate the cross-sectional area as well as the velocity of fluid flow through the inlet and discharge ports. However, the problem of optimum geometric characteristics of the inlet and discharge manifolds and the best parameters of various Limaçon embodiments are left to be solved. Additionally, the effect of the phase change on the expander performance is yet to be investigated. Aim of this document is to provide the framework of the Limaçon machines to achieve optimum expander geometries for power generation systems with various types of working fluids.

**Keywords** Limacon • Expanders • Renewable energy • Waste-heat recovery • Rotating machines

## 3.1 Introduction

The global drive of using energy more effectively and efficiently has fuelled the ever-increasing effort to obtain energy from low-grade heat resources by the research sector as well as the industry. Typically, in order to extract work from

T. Phung (✉) • I. Sultan
Engineering Department, Faculty of Science and Technology, Federation University Australia, Ballarat, VIC, Australia
e-mail: t.phung@federation.edu.au

A. Boretti
Department of Mechanical and Aerospace Engineering (MAE), Benjamin M. Statler College of Engineering and Mineral Resources, West Virginia University (WVU), Morgantown, WV, USA



R.N. Jazar, L. Dai (eds.), *Nonlinear Approaches in Engineering Applications*,
DOI 10.1007/978-3-319-27055-5_3

available heat resources, fluid processing machines, which may include rotary or reciprocating embodiments, are utilised. However, for the specific case of work extracting applications from low-grade heat sources, positive displacement expanders are preferred due to their ability to operate at reasonably low speeds and handle relatively small mass flow rate (Lemort et al. 2009).

The fluid processing machines can be divided into two main categories: positive displacement machine and non-positive displacement machines (turbo machines); they may be categorised as volume type and velocity type fluid processing machines as well (Qiu et al. 2011). Within each those two types, fluid processing machines are further divided into two sub-categories: one of which adds energy to the working fluids (i.e. pumps and compressors), the other one extracts energy from the working fluids (i.e. turbines and expanders). In each sub-category, the fluid processing machines are further split into reciprocating and rotary branches. A fluid processing machinery hierarchy is depicted in Fig. 3.1. Hence, it is obvious that the Limaçon type expanders investigated in this project belong to the rotary branch, nested under positive displacement machine type, which extract energy from the working fluids.

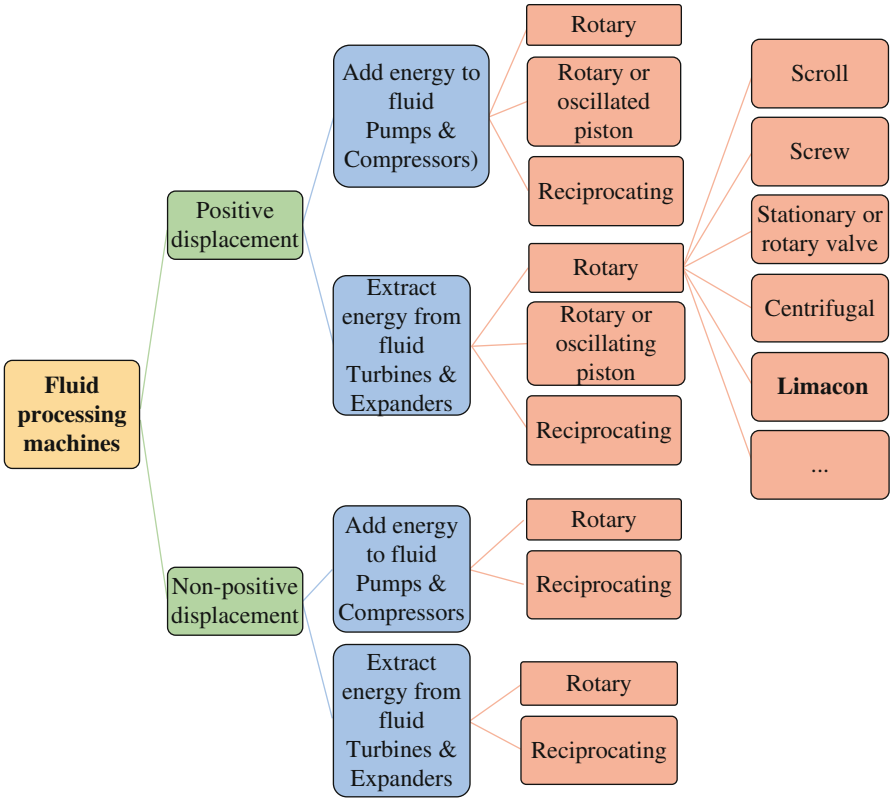



**Fig. 3.1** Fluid processing machines hierarchy

One huge advantage of the positive displacement rotary machines over the reciprocating, oscillated piston and radial piston machines is that they have far less stationary and moving parts. This advantage makes them lightweight, less complex and more compact in terms of physical size. Some of the examples of these rotary machines are the roots, Wankel and Limaçon. Of particular note is the well-known Wankel rotary engine, which consists of much fewer parts than the conventional piston engine. The flow across a Wankel machine is shown in Fig. 3.2 with the help of computational fluid dynamics (CFD) simulations. However, the unique housing-rotor arrangement of this machine requires accurate and sophisticated manufacturing methods. Up to now, Mazda is the only company known to put this design to mass-production (Mazda Motor Corporation 2014).

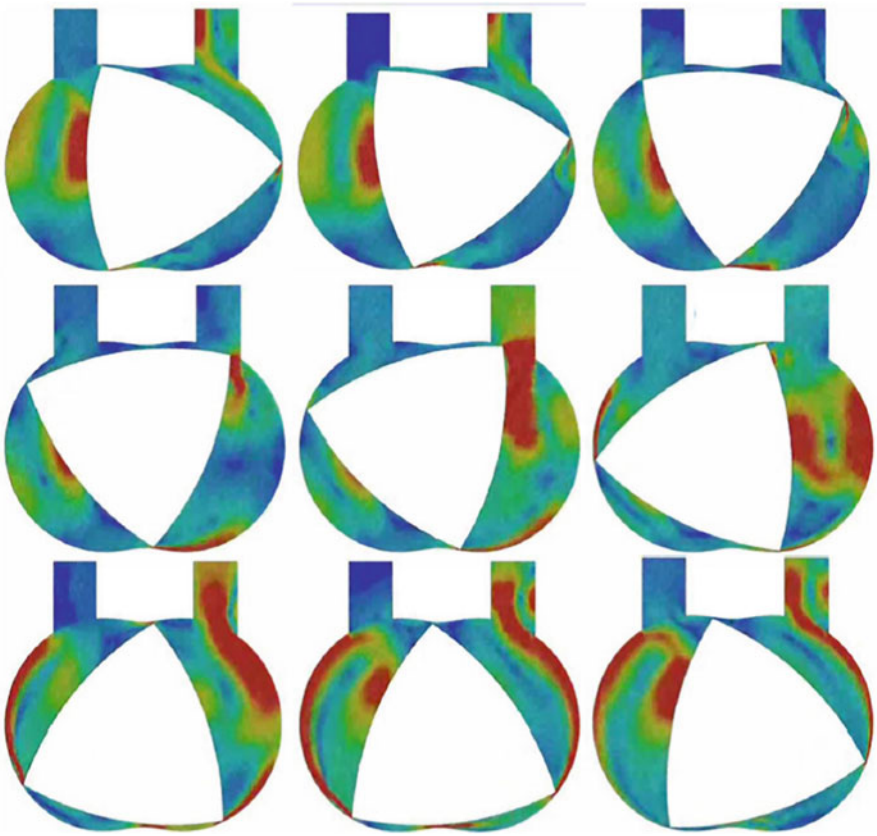


**Fig. 3.2** Sample CFD simulation of the flow in a Wankel engine. The line integral convolution (LIC) plots of velocity vectors representing flow around a Wankel engine rotor cover a complete rotor revolution. Top left is the inlet pipe, top right is the exhaust side. The rotor is rotating counter clockwise moving left to right, then top to bottom. In the Wankel, each one of the three lobes undergoes the four phases of a four stroke engine operation, with decreasing or increasing volumes following the rotation of the rotor. The simulations were performed by using the Star CCM+ fluid dynamic package (http://www.cd-adapco.com/products/star-ccm%C2%AE)

Limaçon expanders, despite the weight and size advantages they offer over the other rotary designs, to a certain extent share the complexity in housing and rotor manufacturing with the Wankel engines. However, with the advancement of additive manufacturing, the development as well as fabricating of Limaçon machines will be much simpler, accurate and cost efficient.

A Limaçon machine is shown in Fig. 3.3. The Limaçon machine's housing and rotor are based on the Limaçon curve, which has the form of $r = b + a.\cos(\theta)$ or

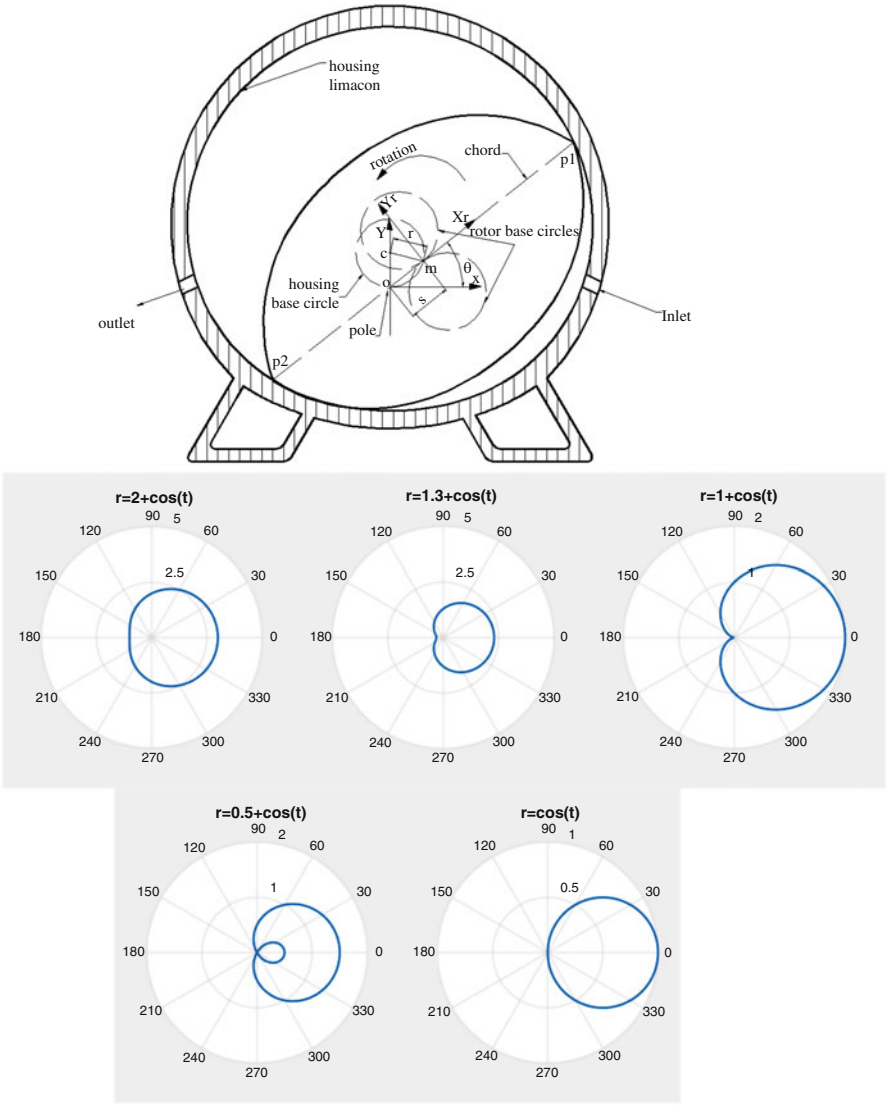



**Fig. 3.3** Limaçon machine and Limaçon machines' curve family

$r = b + a.\sin(\theta)$ in polar coordinates $(r, \theta)$. The values of the parameters $a$ and $b$ determine the shape of the curve. The $r = b + a.\cos(\theta)$ curve takes the $x$-axis as a mirror line, while the $r = b + a.\sin(\theta)$ takes $y$-axis as a mirror line. The profile of Limaçon housing and rotor can be formed by allowing a chord, p1p2, with length 2 L, to rotate and slide on a so-called base circle.

The fluid used to expand and transfer work to the Limaçon expander can be in either of the following forms: gas, liquid or two-phase (liquid and gas) (Sultan and Schaller 2011). The same fluid will also be used to carry heat from the boiler or evaporator to the expander to be processed.

## 3.2 Types of Gas Expanders and Their Role in Thermodynamic Cycles

The Limaçon technology in comparison with other energy conversion technologies such as turbine, rotary, reciprocating piston, rotary piston and rotary vane, is more compact in size and can be designed to run at relatively low speed and is capable of handling small mass flow rate. Additionally, the Limaçon machines are able to operate in multiphase flow conditions that can damage non-positive displacement machines. Despite all the advantages offered by this technology, the unavailability of the necessary manufacturing know-how to accurately machine the Limaçon housing and rotor profiles had in the past deprived Limaçon technology from gaining the publicity it deserved. However, the global drive of using existing energy resources effectively and efficiently combined with the rapid advancement of manufacturing technology will boost industry and research interest in the Limaçon technology.

As suggested by the literatures (Chen et al. 2010; Fukuta and Yanagisawa 2009; Hung et al. 1997; Lemort et al. 2009; Lemort et al. 2010; Persson 1990; Sutan 2012), the overall efficiency of energy recovery systems depends considerably on the performance of the expanders.

As has been noted above, the main focus of this project is the Limaçon technology, Limaçon machines and their applications especially in CHP and waste-heat recovery systems; therefore, the literature will also be mainly focused on this category. In addition to that, CFD, optimisation and multiphase flow modelling will also play a crucial role and need to be investigated. Types of gas expanders, other than Limaçon, which have been previously utilised in CHP or waste-heat recovery processes and their disadvantages, advantages and efficiencies, will be considered. This project aims to utilise the potential higher efficiency and lower manufacturing cost of the Limaçon expanders over other technologies in power recovery processes.

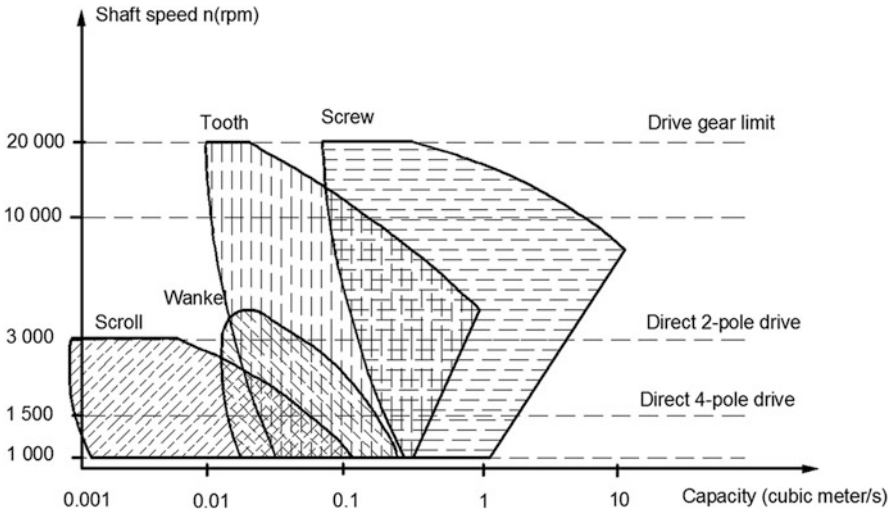



**Fig. 3.4** Example of a common performance map of different geometry concepts (after Persson 1990)

### *3.2.1 Types of Gas Expanders*

Over the years, various gas expander designs have been developed for power generation application in general and waste-heat recovery in particular. In the context of this project, the focus is drawn towards the expanders which are suitable for small-scale heat and power extracting applications such as: axial piston expanders, reciprocating piston expanders, rolling piston expanders, revolving vane expanders, screw expanders, scroll expanders and Limaçon expanders to name only a few.

Figure 3.4 from Persson (1990) shows how various expander designs perform base on their shaft speed and flow capacity. This figure can be considered as a rough guide for the selection of expander type for further design and evaluation.

#### 3.2.1.1 Reciprocating Piston Expanders

The most common reciprocating piston design is featured in the crank-slider mechanism employed for car engines and reciprocating compressors. Over the years, researchers from around the world have focused their attention to different geometric embodiments of the reciprocating piston machines such as: Gimstedt (1979), Glavatskaya et al. (2012) and Gridin and Mirkin (1971) who have designed and evaluated single acting piston expanders. On the other hand, Zhang et al. (2006) have developed the double acting piston expanders. These investigations along with many more published papers on the reciprocating technology can give us a sense of

the popularity, which this technology has achieved. It has been claimed that by the above authors that the expanders in this category can perform at a broad range of pressure ratios with an isentropic efficiency which varies from 50 % to 70 % (62 % in the case of double acting piston expanders) and a volumetric efficiency which ranges from 30 % to 60 %.

#### 3.2.1.2 Axial Piston Expanders

Grip (2009) has offered a geometrically different variation to the reciprocating piston machine. Grip's design is an axial piston expander, which is claimed to be able to run on biofuel or solar resources and to have the potential to apply to low-temperature Rankine cycle plants. In this design, the traditional reciprocating engine is rearranged so that the machines' pistons are spaced evenly around the crankshaft; similar approach can also be found in the engine proposed by Duke (1997). The axial reciprocating piston expanders' output power is limited to the power range below 1000 kW. Grip argues that although piston expanders suffer from higher mechanical losses, they have lower leakage and high part-load efficiency; piston expanders are preferred in some particular applications due to their capability of operating at low speeds. The unique design axial piston machine offers a possibility to significantly reduce the lateral force between the piston and the cylinder liners compared to the conventional reciprocating piston engines.

#### 3.2.1.3 Rolling Piston Expanders

Li et al. (2009) have designed and analysed a rolling piston-type expander, which can replace the throttling valve in the $CO_2$ cycle of refrigeration and air-conditioning systems. The expander helps recover power during the expansion process, hence, improve the system efficiency. The output energy generated by the expander can be fed back to the system as a second-stage compressor or used to drive other mechanisms, machines or applications.

Zheng et al. (2013) have carried out experiments in which a rolling piston expander has been utilised in a low-temperature organic Rankine cycle to generate electricity. Their proposed expander can operate between 350 and 800 rpm with the maximum power output of 0.35 kW and maximum isentropic efficiency of 43.3 % when the heat source temperature is below 90 °C.

#### 3.2.1.4 Rotary-Vane and Revolving-Vane Expanders

Amongst the potential expander concepts, the rotary-vane expanders have been developed and used for years in either waste-heat recovery or improving the process efficiency. Those expanders may include single-acting and double-acting expanders, which have inlet and outlet ports arranged symmetrically in the elliptical cylinder,

the vane slots and vanes are located on the rotor. Rotary-vane expander received a significant amount of attention from researchers and manufactures due to their simple structure and their simplicity in manufacturing compared to others. The calculated volumetric and isentropic efficiencies of such machines varied from 20 % to 70 % and from 20 % to 40 %, respectively. With different operational speeds, gap clearance values and eccentricity ratios, the experimental values of rotary-vane expanders have been reported to be as high as 70 % for volumetric efficiency and as high as 37 % for isentropic efficiency (Fukuta and Yanagisawa 2009; Yang et al. 2009a, b; Jia et al. 2009). By introducing high-pressure gas in to the vane slot, the isentropic efficiency may reach 45 % (Jia et al. 2010).

Subiantoro and Tiow (2009) and Subiantoro and Ooi (2010) have also designed and developed revolving vane (RV) expanders with an average power production of 650 W and mechanical efficiencies of 96.5 % and 95.7 % for RV-type I and RV-type II, respectively.

#### 3.2.1.5 Screw and Scroll Expanders

Screw expanders (helical rotor expanders) and scroll expanders have also been receiving considerable attention from researchers in the field of positive displacement machinery. Examples may be sought in the work of Wang, Wu et al. Tamura et al. Bowman and House who have designed and carried out experimentations on single and double screw expanders. Meanwhile, Wang et al. Lemort et al. and Xiaojun et al. have put their thoughts to scroll expanders. In the aforementioned papers, the authors have theoretically and experimentally explored the performance of screw and scroll expanders in CHP, micro-CHP, work-recovery and waste-heat recovery systems (Wang et al. 2011; Tamura et al. 1997; Bowman 1983; House 1976; Wang et al. 2009; Lemort et al. 2009; Xiaojun et al. 2004).

Peterson et al. (2008) in their paper have built and tested a small-scale regenerative cycle, which utilised moderate-grade heat source (150–400 °C). They experimented with a regenerative Rankine cycle using a scroll expander, which produced 256 W of gross power. The author argued that, despite of the lower than expected efficiency, the project was successful as an exploratory study of using scroll expander in small power generation plants. In addition to that, the proposed scroll expander also suffered from excessive leakage. In fact, leakage is known to be a major unresolved problem with all scroll expanders. This leakage problem significantly affects the volumetric and thermal efficiencies of these expanders. The quality of recovered work and other performance parameters of the scroll expanders had been previously addressed by Xiaojun et al. (2004).

#### 3.2.1.6 Turbo Expanders

The turbo machines can be used as gas expanders in power extracting applications. Ghosh et al. (2005) have developed a design procedure for the main components of

turbo expanders for cryogenic applications. These components are: turbine wheel, nozzle, diffuser, shaft, brake compressor and bearing. Additionally, Hou et al. (2004) experimented with different types of foil bearings for the turbo expanders; while Kucerija (1991) and McClure (1950) have had their turbo expander designs patented. In general, turbo expanders have received substantial attention from researchers in low-grade heat recovery.

### *3.2.2 The Rankine Cycle*

The Carnot cycle is a theoretical and reversible thermodynamic cycle proposed by Nicolas Leonard Sadi Carnot nearly two century ago. The Carnot cycle is the most efficient cycle for converting thermal energy to work. In reality, however, Carnot cycle is considered to be impractical due to its exceptionally slow isothermal heat transfer process.

However, the concept proposed by Carnot has been employed to develop a more realistic power generation system such as the Rankine cycle. The basic Rankine cycle is made up of four main components: a pump, a boiler or an evaporator, an expander and a condenser. Rankine cycle utilises a working fluid to transfer thermal energy from one component to the other, and uses an expander to extract work from that working fluids. With medium and large-scale plants, three groups of working fluids (wet fluids, dry fluids and isentropic fluids) may be used with the Rankine cycle. Micro Rankine cycle power plants are made available with organic fluids (refrigerants and/or hydrocarbons). Such plants are often referred to as "Organic Rankine Cycles power plant" (Tchanche et al. 2014).

There are two main types of Rankine cycles available: single fluid Rankine cycle and binary fluids Rankine cycle. A single fluid Rankine cycle utilises only one working fluid in the work extracting process. The working fluid can be used in different states such as: saturated, superheated or supercritical. Single fluid Rankine cycle is often less complex and has fewer components than the binary cycle. The binary Rankine cycles, on the other hand, with the purpose of increasing the output efficiency, have been developed with more than one working fluid loops. The Maloney and Robertson cycle, the Kalina cycle, the Uehara cycle and the Goswami cycle depicted in Fig. 3.5 are some members of the binary Rankine cycles category (Tchanche et al. 2014).

A basic single fluid Rankine cycle with four stages of change is depicted by a T-s diagram (temperature-entropy diagram) in Fig. 3.6:

- 1–2: Liquid pressurised in a feed pump (W-in)
- 2–3: Continuous heat addition to the working fluid in a boiler or heat exchanger (Q-in)
- 3–4: Isentropic (constant entropy) expansion of the working fluid in an expander(s) (W-out)
- 4–1: Continuous heat rejection in a condenser (Q-out)

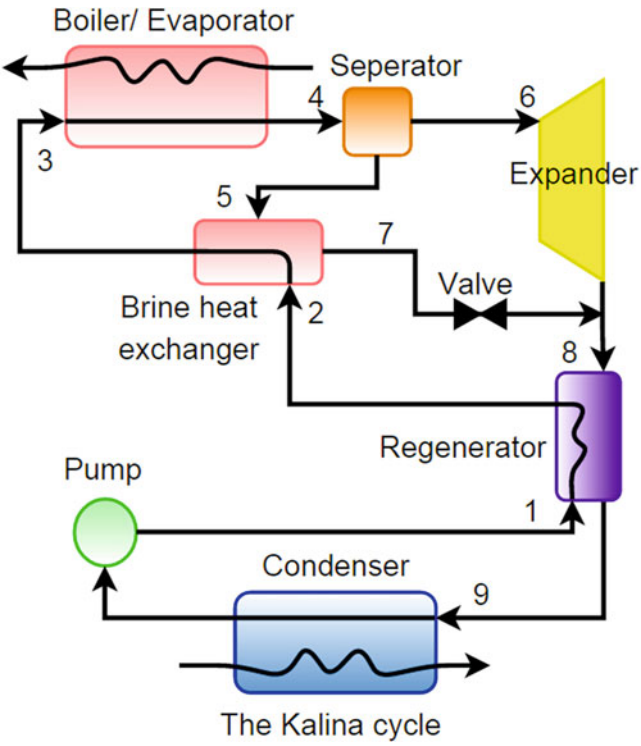



**Fig. 3.5** The Kalina and Goswami cycles (after Tchanche et al. 2014)

Rankine cycle is employed in most of the current thermal power plants, which mainly burn fossil fuels to produce necessary heat. Rankine cycle-based plants can also make use of heat from other thermal energy resources such as: biomass, bio fuels, geothermal, solar thermal and waste heat. In recent years, the ever-increasing demand of renewable energy and efficient energy usage has become the main drive of research effort in improving efficiency of existing power systems, and utilising low-grade or waste heat. Hung et al. Gao et al. Chen et al. and Quoilin et al. are some of the researchers who have put much effort into waste-heat recovery systems with the applications of different types of expanders, e.g. scroll, screw and rolling piston (Tchanche et al. 2014; Guo et al. 2014; Hung et al. 1997; Wang et al. 2010; Gao et al. 2012; Chen et al. 2006; Quoilin et al. 2011).

### *3.2.3 Micro-CHP and CHP-ORC Systems*

According to Qiu et al. expanders can be categorised as: dynamic (or velocity) type, and displacement (or volume type). Within these two types, various designs with

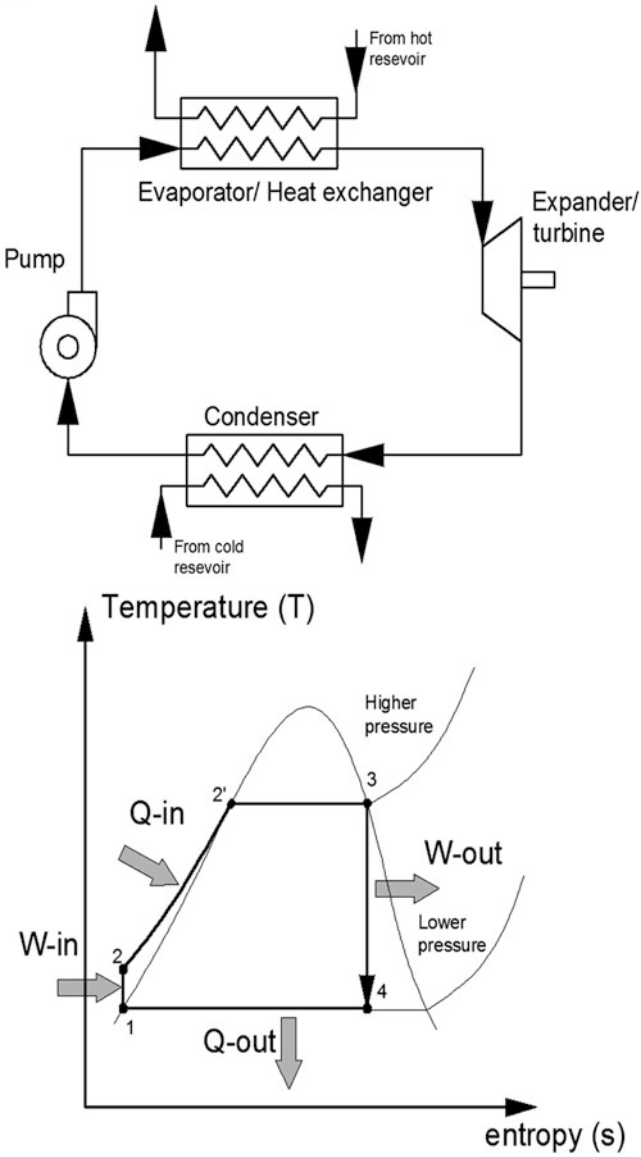



**Fig. 3.6** Basic Rankine cycle

different output efficiency have covered a wide range of output power, ranging from 1 to 50 kW, which falls within the realm of small-scale CHP systems. Additionally, Qiu et al. have also introduced and evaluated the performance of four different types of expanders such as: turbine expander, screw expander, scroll expander and air motor used as expander for household applications (Qiu et al. 2011).

In turbine expanders, the velocity of working fluid in vapour phase (or gas) at high temperature drives the blades, which in turn drive the output shaft. Turbine type expanders are generally used for output power greater than 50 kW due to the fact that the efficiency of such expander deteriorates gradually with the decrease of output power, and an unacceptably low efficiency reached at the output power of 10 kW. Besides, small size turbines are costly to make which prevents them from being utilised in small-scale power generation plants. (Qiu et al. 2011).

Screw expanders utilise the meshing of a male and a female rotors is another choice for small-scale CHP systems with organic Rankine cycle. As the rotors rotate, the volume trapped between the two rotors and the casing changes, working fluid is allowed to enter the space between the rotors and casing on one end and discharge at the other. During the process, the working fluid expands and transfers its energy to the rotors of the expander. However, screw expanders are not widely used as a consequence of difficulty in sealing the working fluid within the expanding chamber (Qiu et al. 2011).

CHP-ORC systems using scroll expanders have been successfully simulated and experimented by Lemort et al. (2009, 2010) and Wang et al. (2010). Wang et al. also mentioned that the efficiency of a scroll based system, which uses R134a as a working fluid may exceed 70 % over a wide range of rotor speeds and expansion ratios. Additionally, the CHP-ORC systems, which use scroll expander, are suitable for the output power range of 0.1–1 kW.

Vane type expanders can also be employed for CHP-ORC as has been featured in the work of Aoun (2010) who has also reported an isentropic efficiency of 80 % at 800 rpm using R11 as working fluid. On the other hand, the corresponding efficiency figures reported for the vane expanders vary from 50 % to 55 % within a speed range of 800–850 rpm. However, the efficiency figure reported by Aoun does not agree with the well-known fact that vane type expanders suffer from serious leakage problem. This is a problem which is highlighted by the comment of Qiu et al. (2011) that the rotational speed of a vane expander considerably influences the pressure and flow rate limit of the working fluid.

Interestingly, R11 and R113, which were employed for experimental evaluation of expanders as reported in many publications, have been phased out by the Montreal protocol (United Nation Environment Programme 1987) as gases that deplete the ozone layer.

### *3.2.4 The Trilateral Flash Cycle*

As described by Smith (1993), an ideal Trilateral Flash Cycle or Triangle Flash Cycle (TFC)—a modification of the ideal Carnot cycle—composed of three processes, two of which are the same as the Carnot cycle, namely the isentropic expansion to produce work, and the isothermal heat rejection to the cooling medium. It is obvious that the ideal TFC would have a lower thermal efficiency compared to the ideal Carnot cycle using the same hot and cold reservoirs. DiPippo (2007)

in a publication on thermal efficiency for geothermal binary plants has compared the efficiencies of a Carnot cycle with an ideal TFC and a practical TFC using the same hot source temperature of 150 °C. DiPippo (2007) also referred to some recent attempts to build machines that approach the ideal TFC, and argued that the ideal TFC is unachievable. In fact, a more realistic and irreversible version of TFC, depicted in Fig. 3.7, should often be considered.

Steffen et al. Fischer and Zamfirescu et al. revealed that TFC is amongst the three processes known for the conversion of low-temperature and waste heat resources to energy: the ORC, the Kalina cycle and TFC. With its ability to recover most of the heat from the heat source, the TFC can easily outperform the ORC and the Kalina cycles. Additionally, with purposefully selected inlet temperatures at various points of the plant, it has been proven that the exergy efficiency for power production of the TFC, which is the ratio of the net power output to the incoming exergy flow of the heat carrier, is from 14 % to 20 % higher than that of the ORC. Also the total efficiency of the TFC, which is the ratio of all the outgoing exergy flows to all incoming exergy flows, is from 1 % to 9 % higher than the ORCs in most cases. Moreover, Zamfirescu and Dincer have demonstrated that with certain setups and at given temperatures the calculated exergy efficiency of the TFC could be twice as much of the ORCs. However, TFC is a fairly new technology that will need further research and development (Steffen et al. 2013; Fischer 2011; Zamfirescu and Dincer 2008).

## 3.3 Limaçon of Pascal Curve and Limaçon Expanders

### *3.3.1 The Limaçon Curve from a Mechanical Point of View*

With the recent development of manufacturing and the advancement in the 3D (three dimensional) printing technology, the fabrication of Limaçon housing and rotor has never been more feasible. As previously mentioned, the polar coordinates form of the Limaçon curve can be described as follows:

$$r = b + a.\cos(\theta) \tag{3.1}$$

or

$$r = b + a.\sin(\theta) \tag{3.2}$$

They can also be transformed to the Cartesian coordinates by applying the two following equations:

$$\begin{cases} x = r\cos(\theta) \\ y = r\sin(\theta) \end{cases} \tag{3.3}$$

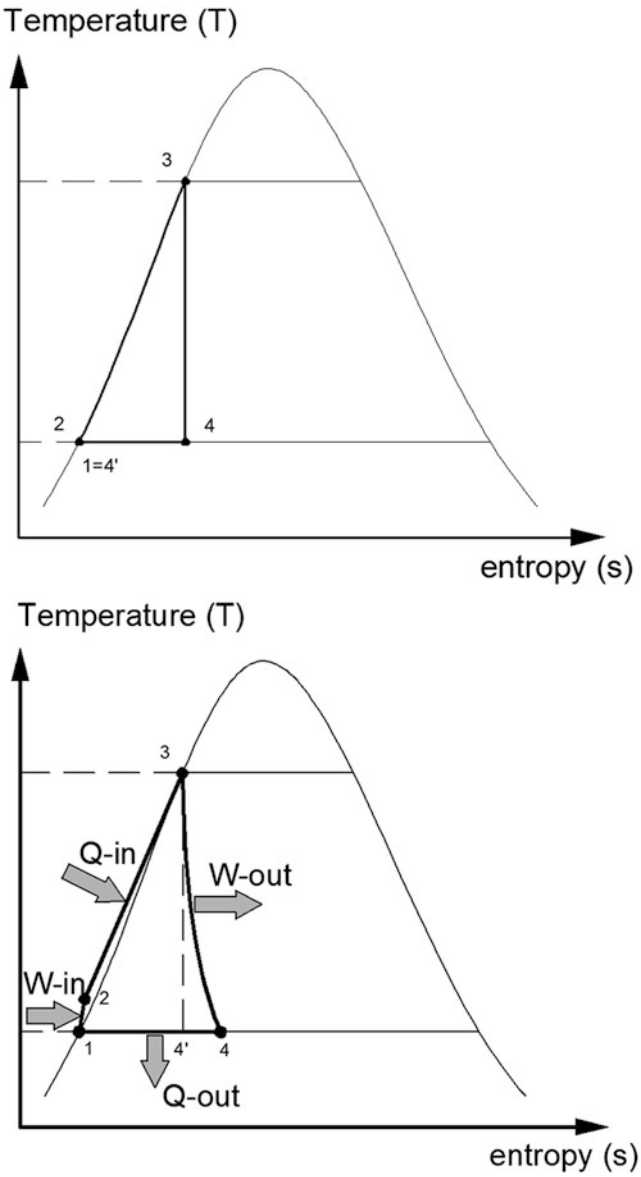



**Fig. 3.7** The ideal and practical trilateral flash cycle (TFC)

As has been mentioned in Sect. 3.1, when the machine is in motion, the rotor chord, p1p2, rotates and slides about the pole, o; the housing Limaçon is formed by the traces of the apex, p1 (or p2). Two coordinates XY and $X_rY_r$ are attached to the housing and rotor, respectively; The XY frame stays stationary with the housing

while the $X_rY_r$ frame rotates with the rotor chord, p1p2. The distance from the centre point, $m$, of the chord, p1p2, to the pole, $o$, is called the sliding distance, $s$, which is formulated as follows:

$$s = 2r \sin(\theta) \tag{3.4}$$

where $r$ is the radius of Limaçon base circle; $\theta$ represents the angle rotated by the chord, p1p2, measured in the right-hand sense from the $X$-axis to the $X_r$-axis. The radial distance, $R_h$, of the housing measured from the pole, $o$, to the apex along the chord, p1p2, may be expressed as:

$$R_h = 2r \sin(\theta) + L \tag{3.5}$$

where $L$ is half of the rotor chord length (as shown in Fig. 3.3 earlier in this document). Applying Eqs. (3.3)–(3.5), the machine housing can be expressed as:

$$\begin{cases} x_h = r \sin(2\theta) + L \cos(\theta) \\ y_h = r - r \cos(2\theta) + L \sin(\theta) \end{cases} \tag{3.6}$$

or

$$\begin{cases} x_h = L[b \sin(2\theta) + \cos(\theta)] \\ y_h = L[b[1 - \cos(2\theta)] + \sin(\theta)] \end{cases} \tag{3.7}$$

where $b = r/L$ is the Limaçon aspect ratio. The value of $b$ has to be less than 0.25 in order to ensure that the Limaçon profile is dimple and loop-free. A number of Limaçon machine embodiments have been proposed over the past 100 years. Despite the fact that these embodiments employ different mechanical drives (mechanisms), they all follow the same kinematical constrains imposed by the Eqs. (3.4)–(3.7). As such, they all feature the similar profiles for the rotor and for the housing, and exhibit the same volumetric and thermodynamic characteristics. However, these embodiments can be differentiated in terms of their distinctive mechanical characteristics such as: durability, dynamic force and vibration, manufacturability as well as ease of maintenance.

### *3.3.2 The Port Area Calculation*

To calculate the port area for the Limaçon machine, Sultan and Schaller (2011) have proposed a concept in which the two vectors $R_l$ and $R_t$ are used to define the port leading and trailing edges (Fig. 3.8).

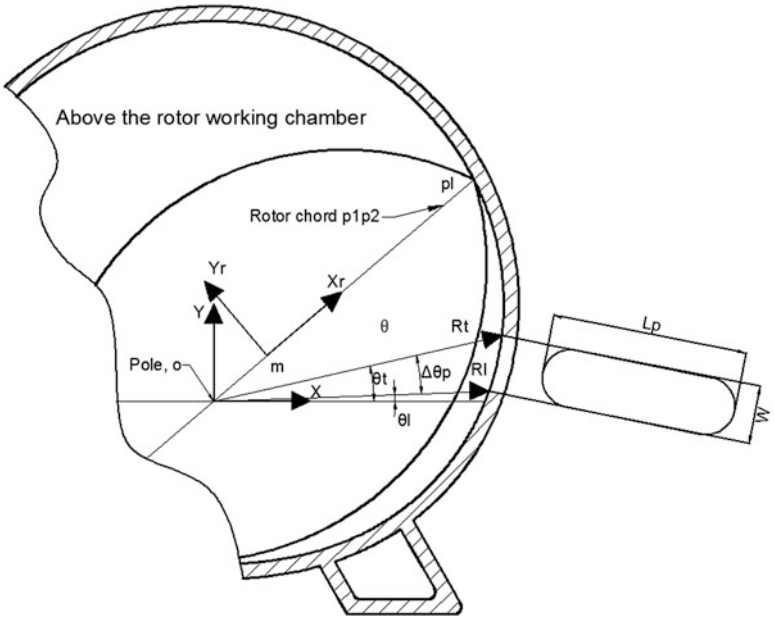



**Fig. 3.8** Ports geometric particulars

This concept can be applied to both inlet and discharge ports. The port edges vector can be defined as:

$$R_l = L\,(2b\sin\theta_l + 1)\,\widehat{R}_l \quad \text{and} \quad R_t = L\,(2b\sin\theta_t + 1)\,\widehat{R}_r \tag{3.8}$$

where $\theta_l$ and $\theta_t = \theta_l + \Delta\theta_p$ are the angular locations of the leading and trailing edges, $\widehat{R}_l = \begin{bmatrix} \cos\theta_l \\ \sin\theta_l \\ 0 \end{bmatrix}$ and $\widehat{R}_t = \begin{bmatrix} \cos\theta_t \\ \sin\theta_t \\ 0 \end{bmatrix}$ are the two unit vectors

The port length, $L_p$, is constrained by the rotor depth; the port width, $W$, can be calculated from the two edge vectors as follow:

$$W = |R_t - R_l| \tag{3.9}$$

If the port ends take the semi-circular shape, the port area, $A_\mathrm{f}$, can be calculated as:

$$A_f = L_p W - W^2\left(1 - \frac{\pi}{4}\right) \tag{3.10}$$

The relation between the rotor apices and the port can be represented by the two vectors $\mathbf{P}_\mathrm{l}$ and $\mathbf{P}_\mathrm{t}$ as follows:

$$\mathbf{P}_l = L\,(2b\sin\theta + 1)\,\widehat{\mathbf{X}}_r \quad \text{and} \quad \mathbf{P}_t = L\,(2b\ \sin\theta - 1)\,\widehat{\mathbf{X}}_r \tag{3.11}$$

Hence, the relative locations between the rotor leading apexes with the port edges can be defined by the two scalar quantities $s_l$ and $s_t$ as follows:

$$s_l = \left(\widehat{\mathbf{R}}_l \times \widehat{\mathbf{X}}_r\right) \bullet \left(\widehat{\mathbf{X}}_r \times \widehat{\mathbf{Y}}_r\right) \quad \text{and} \quad s_t = \left(\widehat{\mathbf{R}}_t \times \widehat{\mathbf{X}}_r\right) \bullet \left(\widehat{\mathbf{X}}_r \times \widehat{\mathbf{Y}}_r\right) \tag{3.12}$$

where $\widehat{X}_r$ and $\hat{Y}_r$ are given as:

$$\widehat{X}_r = \begin{bmatrix} \cos\theta \\ \sin\theta \\ 0 \end{bmatrix} \quad \text{and} \quad \widehat{Y}_r = \begin{bmatrix} -\sin\theta \\ \cos\theta \\ 0 \end{bmatrix} \tag{3.13}$$

Therefore, Sultan and Schaller suggest that the instantaneous value of the port area, $A_p$, can be determined by:

$$\text{for } s_l \geq 0 \begin{cases} \text{if } s_t \geq 0 & A_p = A_f \\ \text{if } s_t < 0 & A_p = (W_l/W)\, A_f \end{cases} \tag{3.14}$$

and

$$\text{for } s_l < 0 \begin{cases} \text{if } s_t \leq 0 & A_p = 0 \\ \text{if } s_t > 0 & A_p = (W_t/W)\, A_f \end{cases} \tag{3.15}$$

where $W_l = |\mathbf{P}_l - \mathbf{R}_l|$ and $W_t = |\mathbf{P}_t - \mathbf{R}_t|$.

### 3.3.3 *The Limaçon Mechanisms*

This section will feature a number of mechanisms, which have been proposed by designers to produce the Limaçon motion and create the profile suitable for fluid processing. In every design mentioned below, it is important to identify the elements of the mechanism, namely:

- the chord, which is the centre line of the rotor
- the chord centre point, $m$
- the Limaçon base circle
- the pole, $o$
- the instantaneous centre, $I$.

The pole, $o$, and the centre point, $m$, should always stay attached to the base circle so that the chord rotates at half of the angular velocity of the centre point, $m$, about the base circle centre. The instantaneous centre, $I$, should also fall on the base circle and is diametrically opposite to the centre point, $m$ (Sultan 2005).

#### 3.3.3.1 The Cam-Assisted Limaçon Mechanism

In the cam-assisted Limaçon mechanism introduced by Wheildon (1896), the lateral axis of the rotor is attached to the drive shaft at the pole. The two flat-face cam followers are perpendicularly attached to the chord and allowed to slide on a stationary circular cam. During motion, the centre point, stays on the base circle; the diameter of which is the distance between the centre of the stationary circular cam and the chord centre, when the chord is at its lowest horizontal position. As previously mentioned, the instantaneous centre of the mechanism will always fall on the base circle at a point which is diametrically opposite to the chord centre (Sultan 2005). This mechanism is presented in Fig. 3.9.

#### 3.3.3.2 The Double Slider Limaçon Mechanism

In the double slider mechanism proposed by Feyens (1927) and Frager and Menard (1954), similarly to the mechanism in which round pins are utilised instead of the sliders by Planche (1920), the double slider mechanism will certainly produce Limaçon motion as the pole, and the chord centre, would always remain attached

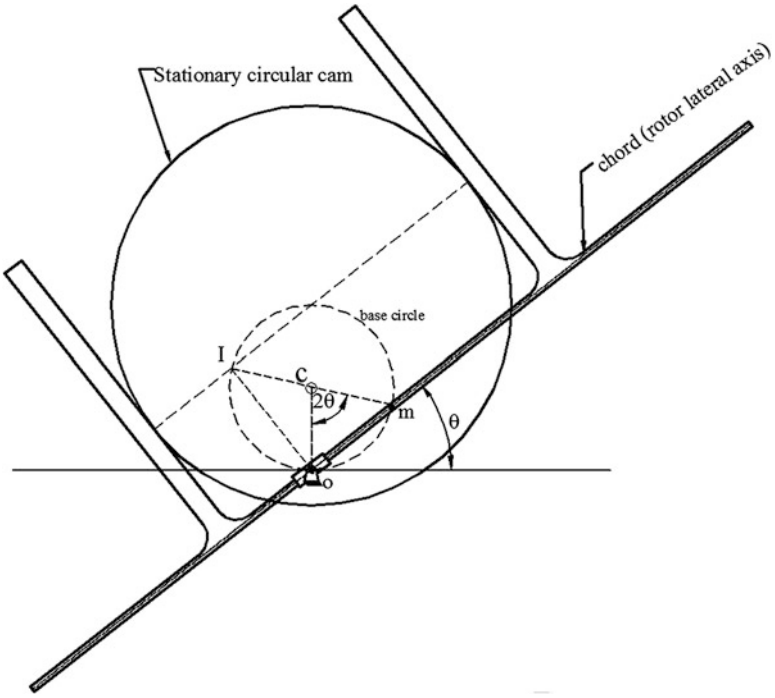



**Fig. 3.9** The cam-assisted Limaçon mechanism

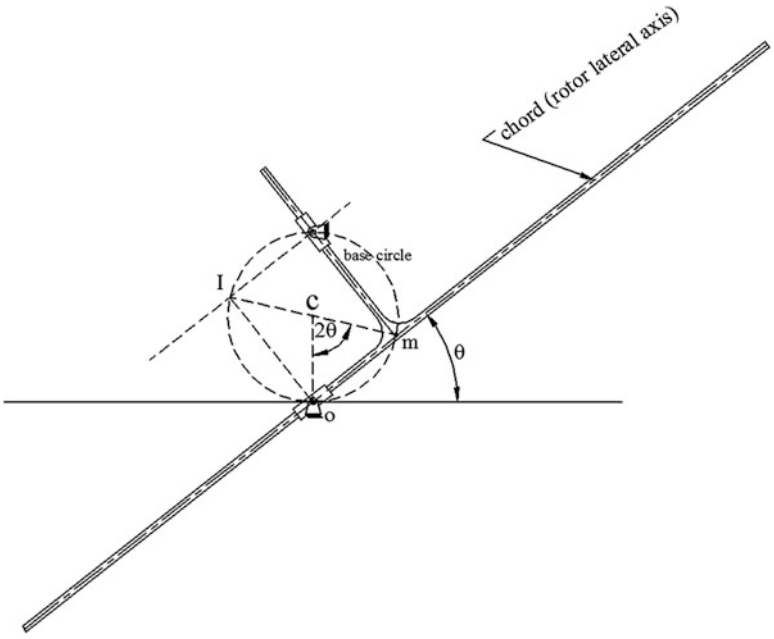



**Fig. 3.10** The double-slider Limaçon mechanism

to the base circle. The diameter of the base circle is equal to the distance between the two-slider pivots; and the drive or driven shaft is connected to the slider at the pole. This mechanism is also able to produce two compression—expansion cycles per every shaft revolution (Sultan 2005). This mechanism is presented in Fig. 3.10.

#### 3.3.3.3 The Epicycloid Gear System

Planche (1920) proposed a mechanism which employed an epicycloid gear system to produce the Limaçon motion for fluid processing applications. The epicycloid gear train utilised by Planche consists of a stationary gear and a ring gear of double the diameter of the stationary gear. A crank is used to connect the centre of the ring gear to the centre of the driveshaft to ensure that the angular velocity of the chord is half that of the driveshaft. Obviously, the base circle of this system is the pitch circle of the stationary gear. Although this mechanism produces the desired Limaçon motion, the constraints imposed by the combined effects of the aspect ratio $\left(b = \frac{r}{L} \leq 0.25\right)$ and the mechanical nature of gear design are likely to reflect themselves on the overall machine dimensions (Sultan 2005). This mechanism is presented in Fig. 3.11.

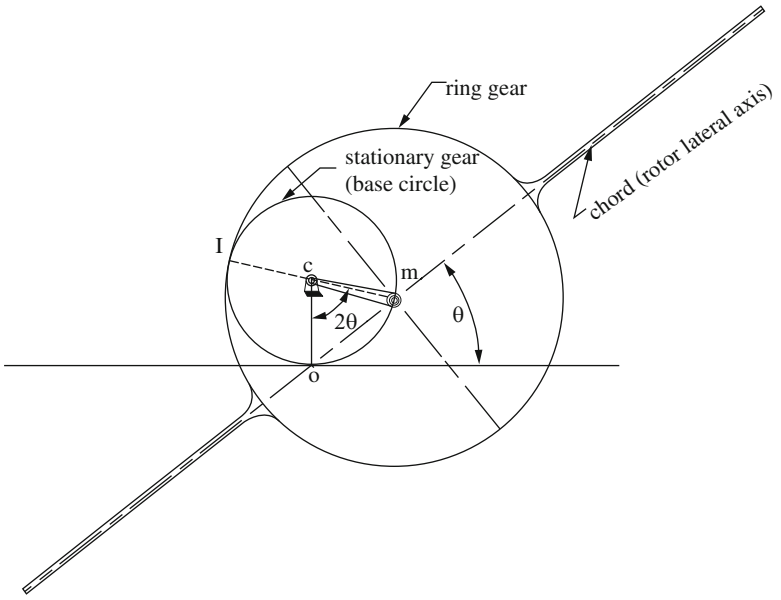



**Fig. 3.11** The epicycloid-gear mechanism

#### 3.3.3.4 Comparison Between Limaçon Mechanisms

Each of the above-mentioned mechanism has its advantages and disadvantages. As for the case of the cam-assisted and the double slider mechanisms (Wheildon 1896; Planche 1920; Feyens 1927; Frager and Menard 1954), both of them can produce two compression–expansion cycles per one revolution of the drive shaft. However, due to the nature of their sliding motion, these two mechanisms are likely to produce a considerable amount of heat and hence, require sophisticated bearing and lubrication design, which may impact their mechanical efficiency. Not to mention that the cam-assisted and the double slider mechanisms have to be manufactured to high levels of accuracy to ensure the roundness, parallelism and/or perpendicularity between different components. The epicycloid-gear mechanism (Planche 1920), on the other hand, employed the rolling motion to its design to accomplish much better mechanical efficiency and simplicity. Nevertheless, the epicycloid-gear mechanism can only produce one compression–expansion cycle per drive shaft revolution. Table 3.1 presents the comparison between those above Limaçon mechanisms.

### *3.3.4 Modelling of Limaçon Gas Expanders*

Despite the fact that Limaçon technology has been developed over a century ago (i.e. the rotary engine by Wheildon (1896)), it has not received the attention it deserved

**Table 3.1** Pros and cons of different Limaçon mechanisms

| | The cam-assisted mechanism | The double-slider mechanism | The epicycloid gear |
|---|---|---|---|
| Pros | – Can perform two compression–expansion strokes per shaft revolution (2:1)<br>– Higher volume flow rate | – Can perform two compression–expansion strokes per shaft revolution (2:1)<br>– Higher volume flow rate | – Utilised rolling motion, hence better mechanical efficiency<br>– Can utilise the technology developed for the Wankel engine |
| Cons | – Produce a significant amount of heat due to the use of sling motion on flat surfaces<br>– Has to be manufactured to relatively high accuracy levels | – Produce heat due to sliding motion on flat surfaces. Mechanical efficiency can be increased by introducing proper bearing and lubrication design.<br>– Has to be manufactured to high accuracy levels | – Can only produce one compression–expansion stroke per shaft revolution (1:1)<br>– Lower volume flow rate |

neither from the research community nor from industry. Recently, however, Sultan has drawn attention to this technology by developing the mathematical and design models to describe the working of such machines (2005, 2006a, b, 2008, 2012). Sultan also argued that the unavailability of the necessary technology to manufacture accurate Limaçon housing and rotor profile had held back the progress in this area. Sultan has published various papers on Limaçon-to-Limaçon and circoLimaçon machines, while Phung and Sultan have investigated the Limaçon-to-circular machines (Phung and Sultan n.d.).

### *3.3.5 Two-Phase Working-Fluid Flow Modelling*

In two-phase flow, the interaction between different phases of working fluids such as mixing, evaporating and condensation make it complicated to calculate the fluid properties. Those interactions are the reasons for a "semi-empirical" approach, which uses the simplified models in accordance with the correlations between phases, to be adopted. There are three basic models for two-phase fluid flow: the homogenous model, the separated flow model and the flow-pattern model. The homogeneous model is the simplest method to use although its accuracy is limited; in order to get a more accurate solutions, the separated flow model can be employed; however, the calculation process may be complicated due to the semi-empirical nature of the solution. The flow-pattern model is the most recent approach to solve two-phase flow problems and its accuracy is gradually being refined (Fairhurst 1983).

### 3.3.6 Homogenous Model

According to Fairhurst (1983), the homogenous model assumes that the two-phase fluid can be represented by a single-phase fluid with its properties derived from those of the two-phase following the below assumptions:

- Equal flow velocities for liquid and gas phases
- Thermodynamic equilibrium has been established between the phases.

Hence with the homogenous model, the properties of two-phase fluid can be represented as follows:

$$\frac{1}{\rho_h} = \frac{x}{\rho_g} + \frac{(1-x)}{\rho_l} \tag{3.16}$$

and

$$\frac{1}{\mu_h} = \frac{x}{\mu_g} + \frac{(1-x)}{\mu_l} \tag{3.17}$$

and

$$u_1 = u_g = u_l = \frac{\dot{m}}{A\rho_h} \tag{3.18}$$

where

- $\rho_h$, $\rho_g$, $\rho_l$: are density of the homogenous model, density of the gas phase and density of the liquid phase, respectively.
- $\mu_h$, $\mu_g$, $\mu_l$: viscosity of the homogenous model, viscosity of the gas phase and viscosity of the liquid phase, respectively.
- $u_h$, $u_g$, $u_l$: velocity of the homogenous model, velocity of the gas phase, velocity of the liquid phase, respectively.
- $\dot{m}$: mass flow rate
- $A$: flow cross-section area

(Fairhurst 1983; Shin et al. 2003; Wang et al. 2001)

The friction factor of the two-phase flow can be obtained by substituting the homogeneous fluid properties to a Moody diagram.

### 3.3.7 The Separated Flow Model

The separated flow model assumes that the two phases are separated and each one is governed by its own set of equations, which correspond to its phase. Fairhurst (1983) argues that: “knowledge of the frictional interactions between the phases and the

flow cross-section occupied by each phase". With this model, observed information is usually employed to determine the two-phase quantities at the regions of interest. Those quantities are otherwise not easily measured and usually unknown. The separated flow model also assumes that the thermodynamic equilibrium is achieved between phases and each phase has a constant velocity. Hence, the total flow is the flow of the liquid phase. The two-phase frictional gradient may be expressed in terms of the calculated single-phase gradient multiplied by a two-phase multiplier (Fairhurst 1983; Lockhart and Martinelli 1949; Iaccarino et al. 2003; Banerjee and Chan 1980).

### *3.3.8 The Flow-Pattern Model*

The flow-pattern model assumes that the two-phase flow exists in one of several configurations for which the basic equations of fluid mechanics can be solved. With this method, the knowledge of a particular flow regime must be obtained by visual observation or some other sophisticated technique (Fairhurst 1983; Kawahara et al. 2002; McQuillan and Whalley 1985).

## 3.4 Design of the Limaçon-to-Circular Fluid Processing Machine

Sultan (2006a, b, 2008) has developed two housing-rotor arrangements for the Limaçon machines such as: Limaçon-to-Limaçon and circoLimaçon. The housing and rotor of the Limaçon-to-Limaçon machines are manufactured of Limaçon curves while those of the circoLimaçon machines are developed from circular curves despite the fact that the motion of the rotor inside the housing still follows the Limaçon trajectory. The apices of the Limaçon-to-Limaçon rotor are always in contact with the housing, which reduces leakage and enables this machine to achieve a better thermodynamic performance than the circoLimaçon counterparts. However, it is quite complex to machine accurate Limaçon curves for both the housing and rotor. Hence, the development of the circoLimaçon embodiment can potentially reduce the manufacturing cost of the gas expander; yet significantly reduce the housing-rotor sealing effectiveness.

In order to set the balance between profile manufacturing cost and the machine sealing effectiveness, a new embodiment of Limaçon machines namely "the Limaçon-to-circular machines" have been proposed by Phung and Sultan (n.d.). This unique housing-rotor arrangement utilises both Limaçon and circular curves while still maintaining the Limaçon motion. Figure 3.12 depicts such a unique arrangement for the Limaçon machines. The Limaçon-to-circular embodiment has been designed with the prevention of housing-rotor interference in mind, and the

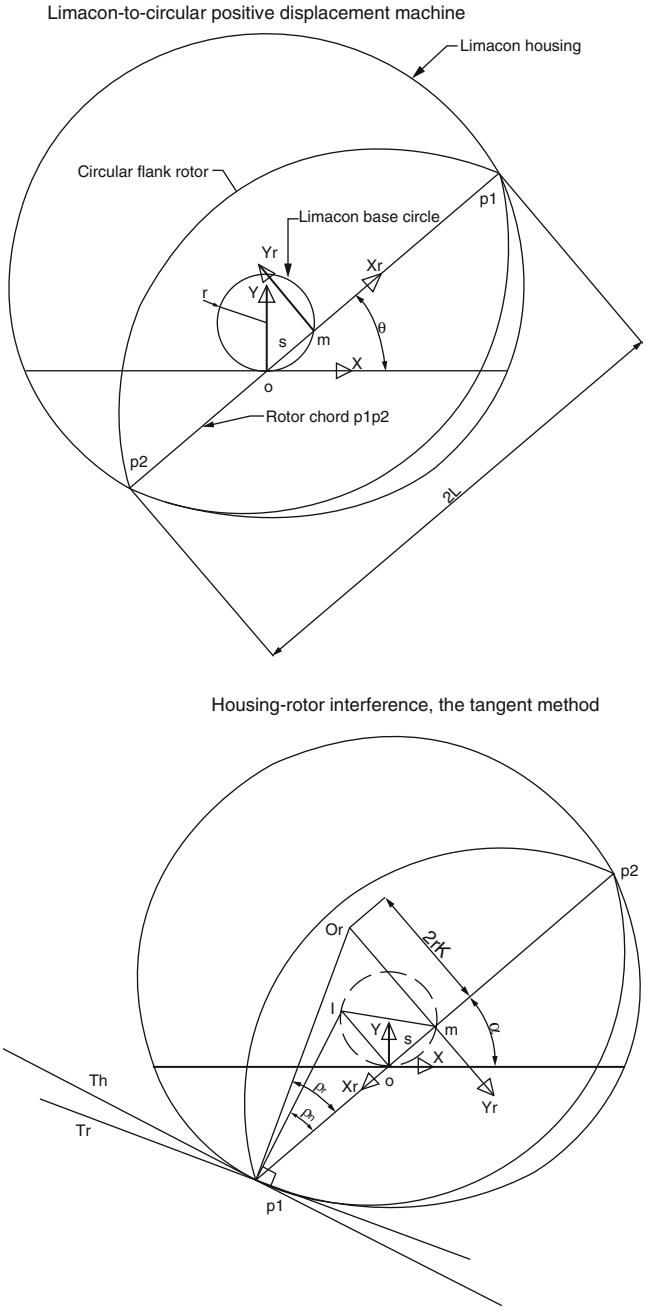



**Fig. 3.12** A Limaçon-to-circular machine and interference prevention—tangent method

design has also been subjected to an optimisation endeavour in order for different operational conditions to be accommodated (Phung and Sultan n.d.).

The sliding distance, $s$, radial distance, $R_h$ and the representation of the radial distance on $X$-axis and $Y$-axis, $x_h$ and $y_h$ are similar to that of the Limaçon-to-Limaçon machine as described early by Eqs. (3.4), (3.5) and (3.7). Since the rotor of the Limaçon-to-circular machine consists of two circular segments, the authors have introduced a small clearance, $C$, to prevent housing-rotor interference. Therefore, half the rotor length, $L_r$, can now be defined as:

$$L_r = L - C \tag{3.19}$$

And the radius of the rotor circular segments, $R$, may be calculated as $R = \sqrt{{L_r}^2 + (2rk)}$ where $k$ is a factor used to facilitate the design process within the machine housing-rotor interference prevention—tangent method. With an angle $\alpha = \theta - \pi$, the value of $k$ can be expressed as:

$$\frac{2rk}{L - C} > \frac{2r\cos(\alpha)}{L - 2r\sin(\alpha)} \tag{3.20}$$

or

$$k > \frac{1 - \frac{C}{L}}{\sqrt{1 - 4b^2}} \tag{3.21}$$

or

$$k = \frac{a(1 - C_L)}{\sqrt{1 - 4b^2}} \text{ with } a > 1 \quad \text{and} \quad C_L = \frac{C}{L} \tag{3.22}$$

The rotor segment radius, $R$, can then be manipulated and expressed as follows:

$$R = (1 - C_L)\sqrt{L^2 + \frac{4r^2a^2}{1 - 4b^2}} \tag{3.23}$$

One important value that determines the efficiency of positive displacement machine is the volume metric ratio, $R_{\text{vol}}$, which is the minimum chamber volume to the maximum chamber volume; Phung and Sultan have obtained such value as follows:

$$R_{\text{vol}} = \frac{\left[\left(b^2 + \frac{1}{2}\right)\pi - 4.b\right] - \left(1 - \frac{C}{L}\right)\left[\tan^{-1}\frac{\sqrt{1-4b^2}}{2ab}\left(1 + \frac{4a^2b^2}{1-4b^2}\right) - \frac{2ab}{\sqrt{1-4b^2}}\right]}{\left[\left(b^2 + \frac{1}{2}\right)\pi + 4.b\right] - \left(1 - \frac{C}{L}\right)\left[\tan^{-1}\frac{\sqrt{1-4b^2}}{2ab}\left(1 + \frac{4a^2b^2}{1-4b^2}\right) - \frac{2ab}{\sqrt{1-4b^2}}\right]} \tag{3.24}$$

After performing the optimisation procedure, Phung and Sultan have investigated different case studies for the Limaçon-to-circular machines based on operational conditions (Phung and Sultan n.d.).

## 3.5 Conclusions

The Limaçon machine is a simple mechanism to increase the work per crank train rotation of traditional reciprocating pistons expanders.

Limaçon-to-Limaçon compression–expansion machines have housings and rotors of profiles that are fully described by Limaçon curves.

We have reviewed the literature on the subject and described the mathematical framework for the design and optimisation of these devices.

## Nomenclature

| | |
|---|---|
| $a$ | Parameter |
| $A$ | Area |
| $b$ | Parameter |
| $C$ | Clearance |
| $k$ | Factor |
| $L$ | Length |
| $m$ | Mass |
| $Q$ | Heat |
| $r$ | Radius or coordinate |
| $P$ | Position vector |
| $R$ | Position vector |
| $s$ | Entropy or distance |
| $T$ | Temperature |
| $U$ | Velocity |
| $x, X$ | Coordinate or position vector |
| $y, Y$ | Coordinate |
| $W$ | Work or width |
| $\theta$ | Angle or coordinate |
| $\rho$ | Density |
| $\alpha$ | Angle |
| $\mu$ | Viscosity |

## References


Aoun, B. (2010). Micro combined heat and power operating on renewable energy for residential building. Paris.

Banerjee, S., & Chan, A. M. (1980). Separated flow models – I analysis of the averaged and local instantaneous formulations. *International Journal of Multiphase Flow*, 1–24.

Bowman, J. L. (1983). *Patent No. 4,412,796.* United States.

Canada, S., Cohen, G., Cable, R., Brosseau, D., & Price, H. (2005). *Parabolic trough organic—Rankine cycle solar power plant*. Colorado: National Renewable Energy Laboratory.

Chen, H., Goswami, D. Y., & Stefanakos, E. K. (2010). A review of thermodynamic cycles and working fluids for the conversion of. *Renewable and Sustainable Energy Reviews, 14*, 3059–3067.

Chen, Y., Lundqvist, P., Johanson, A., & Platell, P. (2006). A comparative study of the carbon dioxide transcritical power cycle compared with an organic Rankine cycle with R123 as working fluid in waste heat recovery. *Applied Thermal Engineering*, 2142–2147.

DiPippo, R. (2007). Ideal thermal efficiency for geothermal binary plants. *Geothermics*, 276–285.

Duke, N. S. (1997). *Patent No. 6,250,262.*

Fairhurst, C. (1983). Component pressure loss during two-phase flow. *Papers presented at International Conference on the Physical Modelling of Multi-Phase Flow*, 1–24.

Feyens, F. (1927). *Patent No. 1,802,887.* US.

Fischer, J. (2011). Comparison of trilateral cycles and organic Rankine cycles. *Energy*, 6208–6219.

Frager, M., & Menard, H. (1954). *Patent No. 3,029,741.* US.

Fukuta, M., & Yanagisawa, T. (2009). Performance of vane-type $CO_2$ expander and characteristic of transcritical expansion process. *HVAC&R Research*, 711–727.

Gao, H., Liu, C., He, C., Xu, X., Wu, S., & Li, Y. (2012). Performance analysis and working fluid selection of a supercritical organic Rankine cycle for low grade waste heat recovery. *Energies*, 3233–3247.

Ghosh, S. K., Seshaiah, N., Sahoo, R. K., & Sarangi, S. K. (2005). Designs of turboexpander for cryogenic applications. *Indian Journal of Cryogenics*.

Gimstedt, L. J. (1979). *Patent No. F02G 1/04.* International.

Glavatskaya, Y., Podevin, P., Lemort, V., Shonda, O., & Descombes, G. (2012). Reciprocating expander for an exhaust heat recovery Rankine cycle for passenger car application. *Energies*, 1751–1765.

Gridin, V. B., & Mirkin, A. Z. (1971). Study of operating processes of reciprocating expanders. *Plenum Publishing Corporation*, 19–21.

Grip, R. L. (2009). *A mechanical model of an axial piston machine*. Stockholm: Royal Institute of Technology.

Guo, C., Du, X., Yang, L., & Yang, Y. (2014). Organic Rankine cycle for power recovery of exhaust flue gas. *Applied Thermal Engineering*.

Hou, Y., Zhu, Z. H., & Chen, C. Z. (2004). Comparative test on two kinds of new compliant foil bearing for small cryogenic turbo-expander. *Cryogenics*, 69–72.

House, P. A. (1976). *Helical-rotor expander application for geothermal energy conversion*. Livermore, CA: Lawrence Livermore Laboratory.

Hung, T. C., Shai, T. Y., & Wang, S. K. (1997). A review of organic Rankine cycle (ORCs) for the recovery of low-grade waste heat. *Energy*, 661–667.

Iaccarino, G., Ooi, A., Durbin, P.A. & Behnia, M. (2003). Reynolds averaged simulation of unsteady separated flow. *International Journal of Heat and Fluid Flow*, 147–156.

Jia, X., Zhang, B., Pu, L., Guo, B., & Peng, X. (2010). Improved rotary vane expander for trans-critical $CO_2$ cycle by introducing high-pressure gas into the vane slots. *International Journal of Refrigeration*, 732–741.

Jia, X., Zhang, B., Yang, B., & Peng, X. (2009). Study of a rotary vane expander for the transcritical $CO_2$ cycle—Part II: Theoretical modelling. *HVAC&R Research*, 689–709.

Kawahara, A., Chung, P.-Y., & Kawaji, M. (2002). Investigation of two-phase flow pattern, void fraction and pressure drop in a microchannel. *International Journal of Multiphase Flow*, 1411–1435.

Kucerija. (1991). *Patent No. 5,003,782.* United States.

Lemort, V., Lebrun, J., & Quoilin, S. (2010). Experimental study and modeling of an Organic Rankine Cycle using scroll expander. *Applied Energy*, 1260–1268.

Lemort, V., Quoilin, S., Cuevas, C., & Lebrun, J. (2009). Testing and modeling a scroll expander integrated into an Organic Rankine Cycle. *Applied Thermal Engineering*, 3094–3102.

Li, M., Ma, Y., & Tian, H. (2009). A rolling piston-type two-phase expander in the transcritical $CO_2$ cycle. *HVAC&R Research*, 729–741.

Lockhart, R. W., & Martinelli, R. (1949). Proposed correlation of data for isothermal two-phase, two-component flow in pipes. *Chemical Engineering Process*, 39–48.

Mazda Motor Corporation. (2014, July 15). *What is the Rotary Engine?* Retrieved from www.mazda.com: http://www.mazda.com/stories/rotary/about/

McClure, A. W. (1950). *Patent No. 2,529,880.* United States.

McQuillan, K. W., & Whalley, P. B. (1985). Flow patterns in vertical two-phase flow. *International Journal of Multiphase Flow*, 161–175.

Persson, J.-G. (1990). Performance evaluation of fluid machinery during conceptual design. *Annals of the ClRP*, 137–140.

Peterson, R. B., Wang, H., & Herron, T. (2008). Performance of a small-scale regenerative Rankine power cycle employing a scroll expander. *Journal of Power and Energy*, 217–282.

Phung, T. H., & Sultan, I. A. (n.d.). On the design of the Limaçon-to-circular fluid processing machines.

Planche, B. R. (1920). *Patent No. 1,340,625.* US.

Qiu, G., Liu, H., & Riffat, S. (2011). Expanders for micro-CHP systems with organic Rankine cycle. *Applied Thermal Engineering*, 3301–3307.

Quoilin, S., Aumann, R., Grill, A., Schuster, A., Lemort, V., & Spliethoff, H. (2011). Dynamic modeling and optimal control strategy of waste heat recovery organic Rankine cycles. *Applied Energy*, 2183–2190.

Shin, B. R., Iwata, Y., & Ikohagi, T. (2003). Numerical simulation of unsteady cavitating flows using a homogenous equilibrium model. *Computational Mechanics*, 388–395.

Smith, I. (1993). Development of the trilateral flash cycle system part 1: Fundamental consideration. *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy*, 179–194.

Steffen, M., Löffler, M., & Schaber, K. (2013). Efficiency of a new triangle cycle with flash evaporation in a piston engine. *Energy*, 295–307.

Subiantoro, A., & Ooi, K. (2010). Design analysis of the novel revolving vane expander in a transcritical carbon dioxide refrigeration system. *Science Direct*, 675–685.

Subiantoro, A., & Tiow, O. K. (2009). Introduction of the revolving vane expander. *American Society of Heating and Air-Conditioning Engineering*, 801–816.

Sultan, I. A. (2005). The Limaçon of Pascal: mechanical generation and utilization for fluid processing. *The Journal of Mechanical Engineering Science*, 813–822.

Sultan, I. A. (2006a). Inverse geometric design for a class of rotary positive displacement machines. *Inverse Problems in Science and Engineering*, 127–139.

Sultan, I. A. (2006b). Profiling rotor for Limacon-to-Limacon compression-expansion machines. *Journal of Mechanical Design*, 787–793.

Sultan, I. A. (2008). A geometric design model for the circoLimacon positive displacement machines. *Journal of Mechanical Design*, 1–8.

Sultan, I. A., & Schaller, C. G. (2011). Optimum positioning of ports in the Limaçon gas expanders. *Journal of Engineering for Gas Turbines and Power*.

Sutan, I. A. (2012). Optimum design of Limaçon gas expanders based on thermodynamic performance. *Applied Thermal Engineering*, 188–197.

Tamura, I., Taniguchi, H., Sasaki, H., Yoshida, R., Sekiguchi, I., & Yokigawa, M. (1997). An analytical investigation of high-temperature heat pump system with screw expander for power recovery. *Energy Conversion Management*, 1007–1013.

Tchanche, B. F., Pétrissans, M., & Papadakis, G. (2014). Heat resources and organic Rankine cycle machines. *Renewable and Sustainable Energy Reviews*, 1185–1199.

United Nation Environment Programme. (1987, September). *United Nation Environment Programme – Ozone Secretariat.* Retrieved from The 1987 Montreal Protocol on Substances that Deplete the Ozone Layer: http://ozone.unep.org/new_site/en/Treaties/treaties_decisions-hb.php?sec_id=342&show_all

Wang, H., Peterson, R., Harada, K., Miller, E., Ingram-Goble, R., Fisher, L., et al. (2010). Performance of a combined organic Rankine cycle and vapor compression cycle for heat activated cooling. *Elsevier Energy*, 447–458.

Wang, H., Peterson, R. B., & Herron, T. (2009). Experimental performance of a compliant scroll expander for an organic Rankine cycle. *Journal of Power and Energy*, 863–871.

Wang, Z., Wang, C., & Chen, K. (2001). Two-phase flow and transport in the air cathode of proton exchange membrane fuel cells. *Journal of Power Sources*, 40–50.

Wang, W., Wu, Y.-t., Ma, C.-f., Liu, L.-d., & Yu, J. (2011). Preliminary experimental study of single screw expander prototype. *Applied Thermal Engineering*, 3684–3688.

Wheildon, W. M. (1896). *Patent No. 553,086.* United States.

Xiaojun, G., Liansheng, L., Yuanyang, Z., & Pengcheng, S. (2004). Research on a scroll expander used for recovering work in a fuel cell. *Journal of Thermodynamics*, 1–8.

Yang, B., Peng, X., He, Z., Guo, B., & Xing, Z. (2009a). Experimental investigation on the internal working process of a $CO_2$ rotary vane expander. *Applied Thermal Engineering*, 2289–2296.

Yang, B., Peng, X., Sun, S., Guo, B., & Xing, Z. (2009b). Study of a rotary vane expander for the transcritical $CO_2$ cycle—Part I: Experimental investigation. *HVAC&R Research*, 673–688.

Zamfirescu, C., & Dincer, I. (2008). Thermodynamic analysis of a novel ammonia–water trilateral Rankine cycle. *Thermochimica Acta*, 7–15.

Zhang, B., Peng, X., He, Z., Xing, Z., & Shu, P. (2006). Development of a double acting free piston expander for power recovery in transcritical $CO_2$ cycle. *Applied Thermal Engineering*, 1629–1636.

Zheng, N., Zhao, L., Wang, X., & Tan, Y. (2013). Experimental verification of a rolling-piston expander that applied for low-temperature organic Rankine cycle. *Applied Energy*, 1265–1274.

# Chapter 4
# Physical Networks' Approach in Train and Tram Systems' Investigation

**Milan Simic**

**Abstract** Large vehicles on rail networks are complex, nonlinear, engineering systems that perform translator motion. Apart from system components' nonlinearities, there are environmental influences as well, through railroad characteristics' changes on the long routes. On the one side, we have historical or legacy systems, still in place, and on the other, there are state-of-the-art train compositions that are capable of achieving extremely high ground speeds of transportation. In both cases reliability and safety are the primary objectives in system maintenance and system design. Global physical networks approach is presented here, and applied in this investigation. Physical quantities like speed, force, power, and energy are studied, monitored, and presented. The main contribution of this modeling application is in its capability to obtain various data from any part of the system which could be used for the improvement of overall system safety and reliability.

**Keywords** Physical network • Network elements • Train • Tram • Composition • Rail network • Traffic safety • System reliability

## 4.1 Introduction

We could easily say that the train is among the longest, mobile, mechanical engineering systems that humans can build, nowadays. Train compositions, that can stretch few kilometers, usually have distributed power units, i.e., locomotive engines positioned in front of, between wagons, and behind them. Such long and heavy transportation systems express nonlinearities of different types. That also depends heavily on the environment conditions. For example, friction between the rails and the wheels is changing drastically when the weather changes from the very dry and sunny conditions to heavy rain, or snow and ice on the rails. Those conditions are extremely important for the acceleration and braking, having in mind that we perform motion control of the overall mass of few $10^7$ kg. There are

M. Simic (✉)
School of Aerospace, Mechanical and Manufacturing Engineering (SAMME), RMIT University, Bundoora East Campus, Plenty Road, Bundoora, Melbourne, VIC 3083, Australia
e-mail: milan.simic@rmit.edu.au



R.N. Jazar, L. Dai (eds.), *Nonlinear Approaches in Engineering Applications*,
DOI 10.1007/978-3-319-27055-5_4

various mathematical models and simulations used in the systems' study and design (Ďurica 2015; A mathematical model for a train run). In various large and complex, nonlinear, engineering systems investigation, physical networks approach is one of the valuable tools adapted and already presented (Simic 2015). Application area is now extended and results are presented here. Velocities, forces, displacements, power, and energy have been monitored.

Initially, a physical networks' model of a single wagon is created taking in consideration the mass of the vehicle together with the stiffness and friction components of the interfacing elements between wagons and with the rails. Starting from that basic model, the whole train composition model, as a large network, is set up and simulations performed. Any physical quantity can easily be monitored and displayed. This gives valuable information on the subsystems and system design.

## 4.2 Physical Networks Approach in Modeling

Physical network is a geometrical structure of interconnected ideal network elements, with two connection points. Network elements represent mathematical relationships between two dependent system variables in the physical system. Engineering systems are built using various sets of elements, specific for the system type. Basic definitions, principles, and rules for solving equations represented by a network are independent of physical system which that network is representing.

Energy through the system is transferred and transformed using two types of system variables. ***Flow, f,*** type system variables are representing physical quantities that are traveling ***through*** the systems elements, and their connections. ***Potential, p,*** type variables are expressing the state established ***across*** the system elements and between any two network points, i.e., nodes. Potential of a point in the network is a relative quantity and depends on the reference point chosen. For example, speed is relative, as we know, and depends on the reference.

A simple physical network example is shown in Fig. 4.1. Looking from the energy point of view, there are passive and active network elements in each network. Generic name for any passive element is impedance, $\boldsymbol{Z}$, or admittance, $\boldsymbol{Y}$, where $\boldsymbol{Z = 1/Y}$. Active elements, or energy suppliers to the system, are ideal flow and potential sources, as well as initial conditions expressed as initial values of the flow through, and potential across network elements.

Ideal flow source, shown as ***F(t)*** in Fig. 4.1, has infinite internal resistance, called impedance, $\boldsymbol{Z_{in}}$, $\boldsymbol{Z_{in} -> \infty}$ while ideal potential source, ***P(t),*** has zero value of the internal impedance, $\boldsymbol{Z_{in} = 0}$. Flows through branches are shown, and labeled as $\boldsymbol{f_1}$ and $\boldsymbol{f_2}$, while the node potentials are marked as $\boldsymbol{p_1}$ and $\boldsymbol{p_2}$. In any physical network, product of instant values of two network variables, flow ***f(t)*** and potential ***p(t),*** is giving the value of the instant power *power(t),* that particular network element dissipates or stores, as given by Eq. (4.1).

$$power(t) = f(t) * p(t) \tag{4.1}$$

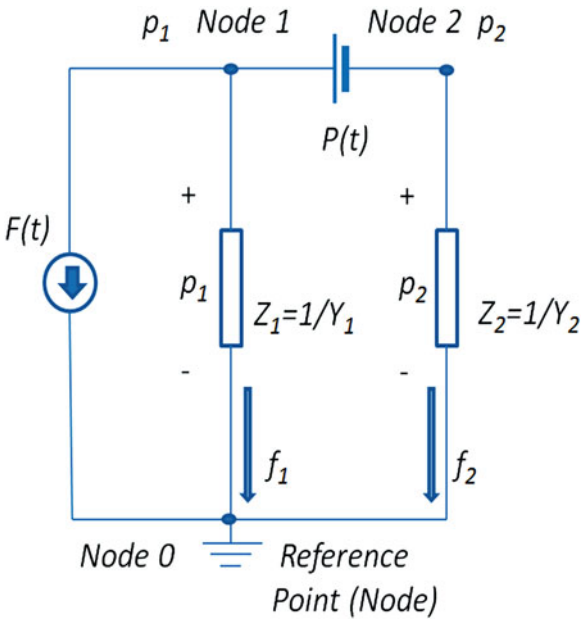



**Fig. 4.1** Generic physical network diagram: Ideal flow source is labeled as ***F(t)***, while ideal potential source is labeled as ***P(t)***

The unity of the nature is expressed in the extraordinary analogies of the differential equations used to represent various physical phenomena. The same type of equations, ordinary differential equations (ODE), is used for the study of mechanical systems with translation, mechanical systems with rotation, hydrodynamics, and for the electrical circuits. The theory of turbulence in liquids and the theory of friction in gases show great similarities with the electromagnetic theory. Network elements represent mathematical relationships between two dependent system variables in a physical system. There are three types of relationships: proportionality, differentiation, and integration. For example, in an electrical circuit those basic elements are resistor, capacitor, and the coil. Often, real systems are extremely complex, but they can be simplified, or they may have linear subsystems as their integral parts.

Examples of physical variables in an electric circuit are ***electrical current,*** as a ***through variable***, and ***electrical potential,*** as an ***across variable***. Current, $i$, through electrical conductors, or a network element, is directly related to the ***mechanical*** flow of electrons, i.e., charge, $dq$, over time $dt$, as per Eq. (4.2).

$$i = \frac{dq}{dt} \tag{4.2}$$

The charge of a single electron is $e = -1.602 \times 10^{-19}$ *coulomb*. Following that, the current of *1A* corresponds to the flow of $6.241509 \times 10^{18}$ electrons per second.

Electrical potential, $V_C$, for example, is related to the number of accumulated electrons, on capacitor $C$, plates, as given by Eq. (4.3),

$$V_C = \frac{1}{C}\int_0^T idt = \frac{1}{C}\int_0^T \frac{dq}{dt}dt = \frac{Q(T)}{C}; \quad \text{if } Q(0) = 0 \tag{4.3}$$

where $Q(T)$ refers to the accumulated charge over the period $T$ when initial charge on the capacitor plates, for $t = 0$, was *0 coulomb*. Potential difference between two nodes is called voltage. The universal reference point, in electric circuits, is ground potential of *0 V*.

We have ***force*** and ***velocity*** as network variables in a mechanical system with translation, or ***torque*** and ***angular speed*** in a mechanical system with rotation. In a hydraulic system we have ***flow*** and ***pressure***.

ODE with constant coefficients, $A_i$, $i = 0 - n$, are used for modeling various physical systems. An ODE, shown by Eq. (4.4), is a relation between two variables: independent variable $t$ and dependent $y = y(t)$, and the derivatives of $y$ as follows $\frac{dy}{dt}, \frac{d^2y}{dt^2}, \ldots, \frac{d^ny}{dt^n}$.

$$A_n\frac{d^ny}{dt^n} + A_{n-1}\frac{d^{n-1}y}{dt^{n-1}} + \cdots + A_1\frac{dy}{dt} + A_0y = f(t) \tag{4.4}$$

On the right-hand side, we can have any function of time, $f(t)$. The special case when $f(t) = 0$ is known as homogenous equation. Equation (4.4) is called ordinary because only one independent variable exists, which is usually time. It is linear because only the first exponent of dependent variable, or its derivatives, is present in the expression. Examples of network components in a generic physical network, then in an electrical circuit, mechanical system with translation and mechanical system with rotation, are given in Table 4.1. Equations for the stored energy and power losses are also presented.

In a generic physical system, integration, proportion, and differentiation of network variables are associated with elements labeled as *A, B*, and *C*. The general name for all of them is impedance, *Z*, or admittance *Y,* as already shown in Fig. 4.1. In an electrical circuit we have inductivity, *L*, conductivity *G*, i.e., resistivity, $R = 1/G$, and capacity *C*. Finally for the translation we have *k* for stiffness, *B* for friction, and *m* for mass of the object. Translator network variables are force, *F*, and velocity *v*. There are also other physical systems like thermal and fluids where analogies could easily be established, as given in (de Silva 2005 and Sanford 1965).

## 4.3 Mechanical System with Translation: Basic Model

Let us consider a mechanical system with translation. Two approaches in modeling of a basic network, which includes all three passive network elements and a power source, are presented here. Passive network elements are mass, ***m***, expressing

**Table 4.1** Various physical network components

| Description | Prototype | Electrical | Translation | Rotation |
|---|---|---|---|---|
| Through variable | ***Flow - f*** | ***Current - i*** | ***Force - F*** | ***Torque, or Momentum - M*** |
| Across variable | ***Potential - p*** | ***Voltage - u*** | ***Velocity - v*** | ***Angular Velocity - w*** |
| Element integration | Generic Element $A$ $f = A\int pdt$ | Inductivity $L$ $i = \frac{1}{L}\int udt$ | Stiffness $k$ $F = k\int vdt$ | Rotational Stiffness $k$ $M = k\int wdt$ |
| Accumulated energy | $\frac{f^2}{2A}$ | $\frac{Li^2}{2}$ | $\frac{F^2}{2k} = k\frac{x^2}{2}$ $x =$ distance | $\frac{M^2}{2k}$ |
| Element proportion | Generic element $B$ $f = Bp$ | Conductivity G/Resistivity $R$ $i = Gu = \frac{1}{R}u$ | Damping constant $B$ $F = Bv$ | Angular damping $D$ $M = Dw$ |
| Power dissipation | $fp = \frac{f^2}{B} = p^2B$ | $iu = i^2R = \frac{u^2}{R}$ | $Fv = \frac{F^2}{B} = v^2B$ | $Mw = \frac{M^2}{D} = w^2D$ |
| Element differentiation | Generic element $C$ $f = C\frac{dp}{dt}$ | Capacity $C$ $i = C\frac{du}{dt}$ | Mass $m$ $F = m\frac{dv}{dt}$ | Moment of Inertia $F = M\frac{dw}{dt}$ |
| Accumulated energy | $\frac{Cp^2}{2}$ | $\frac{Cu^2}{2}$ | $\frac{mv^2}{2}$ | $\frac{Mw^2}{2}$ |

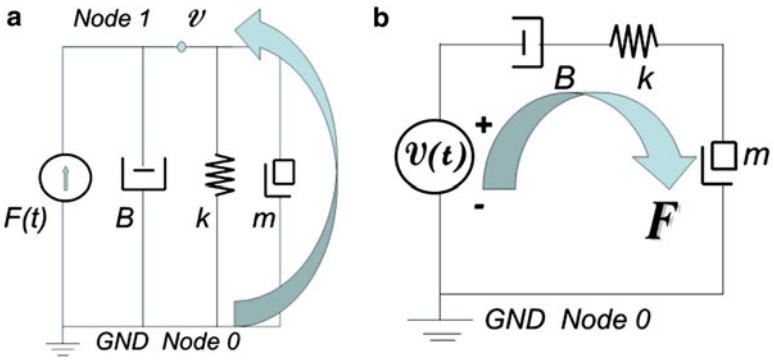



**Fig. 4.2** (**a**) Basic mechanical network model with ideal force power source ***F(t)***. (**b**) Basic mechanical network model with ideal velocity power source ***V(t)***

inertia, spring element with its stiffness, ***k***, and damping element, ***B.*** Active elements, presented here, are an ideal force source, ***F(t)***, as shown in the first layout, Fig. 4.2a, and an ideal velocity source, ***V(t),*** as shown in the second layout, Fig. 4.2b.

Basic translational motion mechanical network with ideal force source ***F(t)*** can be expressed with Eq. (4.5), i.e., Eq. (4.6) as follows:

$$Bv + m\frac{dv}{dt} + k\int vdt = F \tag{4.5}$$

or in an operator form

$$\left(B + m\frac{d}{dt} + k\int dt\right)v = F \tag{4.6}$$

On the other side, basic translational motion mechanical network with ideal velocity source ***V(t)*** can be expressed with Eq. (4.7), i.e., Eq. (4.8) as follows:

$$\frac{F}{B} + \frac{1}{k}\frac{dF}{dt} + \frac{1}{m}\int Fdt = v \tag{4.7}$$

or in an operator form

$$\left(\frac{1}{B} + \frac{1}{k}\frac{d}{dt} + \frac{1}{m}\int dt\right)F = v \tag{4.8}$$

In Fig. 4.2a the same velocity, $v$, is measured across all network elements as it is an ***across variable***. Since there are no initial condition shown, the sum of forces through passive elements equals the force supplied by ideal force source ***F(t).*** Analogue story is for the network shown in Fig. 4.2b, where the same force

is measured ***through*** all network elements. In this case the sum of all velocities measured ***across*** all passive elements equals to the velocity supplied by the ideal velocity source ***V(t).***

## 4.4 Modeling a Single Wagon

We will now consider a single wagon model. Vehicle has four pairs of wheels that support the body by springs and dry friction dampers. A wagon is shown in Figs. 4.3a and 4.4, while connection interface can be seen from Fig. 4.3b. Initial model design would include basic sub-networks as already shown in Fig. 4.2a, just without force power source component. We have eight sub-models, each with all three basic network elements. Since the velocity is same, in the node 1, we can use equivalent representation. Following that, in the next step of modeling we will represent the system with just two support points which very much correspond to the mechanical design. The system is nonlinear and multidimensional, but motion and vibrations along other axes might be subject of another investigation, since we now just consider the motion along the line connecting two points on the railroad. We will neglect gravitational stiffness since we will consider just one dimensional problem of translation along $x$ axis.


**Fig. 4.3** (**a**) A single wagon as an example for basic tram/train elements modeling; (**b**) Connection interface that will be modeled as a spring

**Fig. 4.4** A standard wagon with four sets of wheels

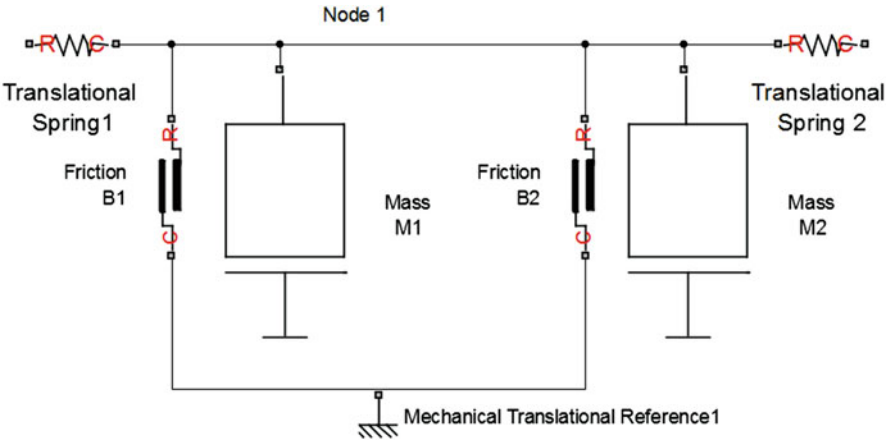



**Fig. 4.5** A Simulink model of a single wagon with 2 support points and connection interface

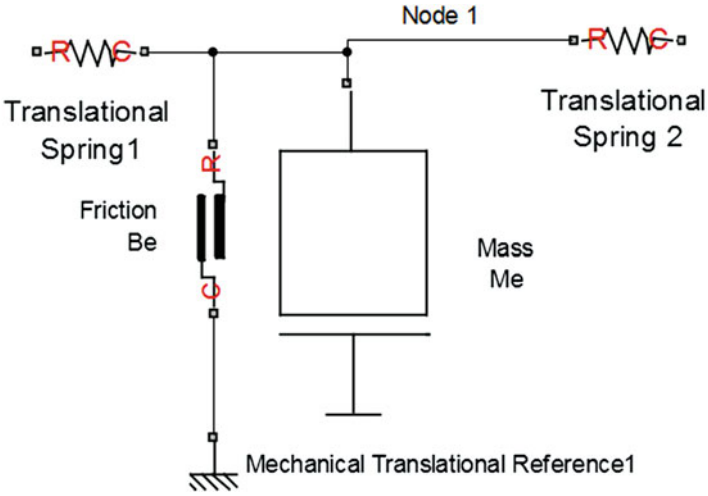



**Fig. 4.6** Final Simulink model of a single wagon with connection interface

Two support points model designed in Simulink is shown in Fig. 4.5. Connection interface is presented by translational springs. Further simplification of the model is shown in Fig. 4.6. Once again, since the velocity in the node 1 is applied ***across*** all network elements we can represent the network from Fig. 4.5 with the equivalent one given in Fig. 4.6. Total equivalent mass *Me* is the sum of M1 and M2, while the *Be* is the sum of B1 and B2. Comparing translational mechanical system to an electrical system, we can see that the ***mass*** of an object shows analogy with the ***capacity*** while the ***friction*** is similar to the ***conductivity***, i.e., reciprocal to the resistivity, $\boldsymbol{G = 1/R}$.

## 4.5 System of Two Wagons

Let us now consider the process of joining two wagons. One of them is in stationary state while the other one is approaching with the speed of $v_0 = 2$ m/s. Using the equivalent wagon model presented in Fig. 4.6 we have designed system's model of two wagons as shown in Fig. 4.7. The presented model is ready to run, but we should be able to monitor changes in the network variables. In our case, basic physical variables of interest are speed and force. Position, power, and energy could be monitored as well, based on the set of Eqs. (4.9), (4.10), and (4.11).

$$\text{Position} = \int_0^T v(t)dt \tag{4.9}$$

$$\text{Power} = p(t) = F(t) * v(t) \tag{4.10}$$

$$\text{Energy} = E = \int_0^T p(t)dt = \int_0^T F(t) * v(t) \tag{4.11}$$

In order to monitor physical quantities as mentioned above we need to introduce sensors and display devices. While measuring velocity as an ***across variable*** we have to place the sensor ***across*** network element. Our Ideal Translational Motion Sensor, as defined in Simulink environment, has to be placed across measurement node and the Mechanical Translational Reference, which is equivalent to mechanical

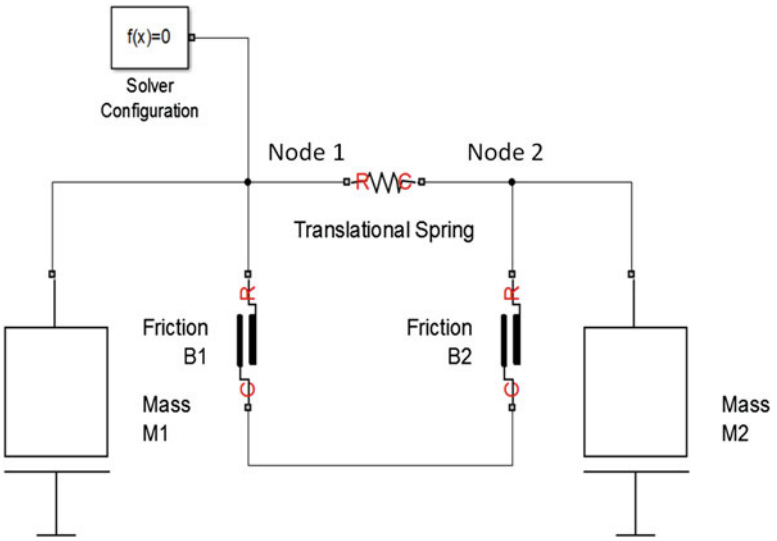



**Fig. 4.7** A Simulink model of two wagons

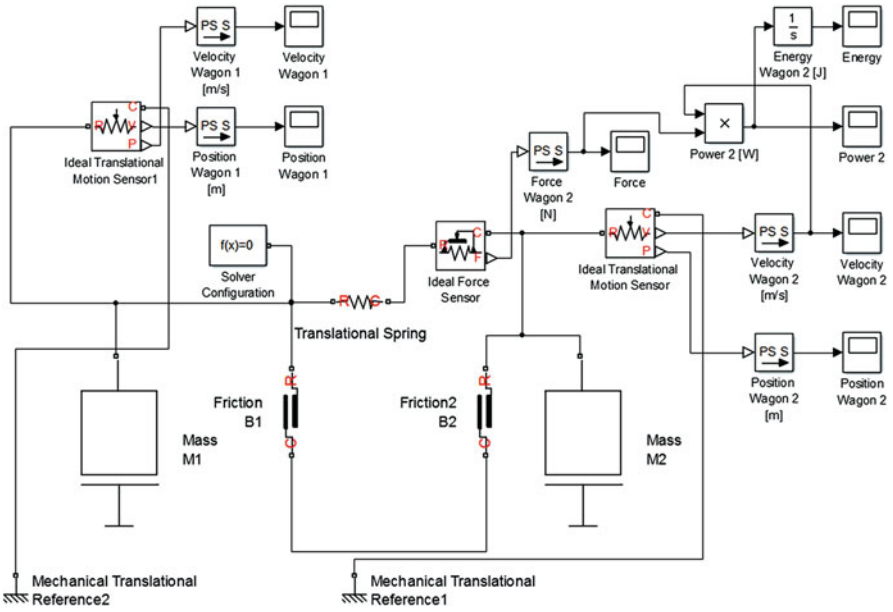



**Fig. 4.8** Final Simulink model of two wagons with sensors and monitors

motion ground, i.e., $V_{GND} = 0$ m/s. Opposite to that, measurement of a through variable, such as force in our case, requires placement of the sensor in the network. According to this a new simulation model is designed and presented in Fig. 4.8. Simulation results for the wagon one velocity, wagon two velocity, path traveled, i.e., position and force at the wagon two, are presented in Fig. 4.9.

We can see that the speed of the first wagon is going down from the initial value of $v_0 = 2$ m/s to 0.64 m/s, oscillating and then stabilizing at the value of 1 m/s as expected.

Similarly, after the impact, the speed of the second wagon is increasing from 0 m/s to 1.36 m/s, oscillating and then stabilizing at 1 m/s. The force is maximum just after the contact and then oscillating and going down to zero. The power diagram is shown in Fig. 4.10. Variables are expressed in SI systems units [W]. SI units are used in the whole paper. Finally energy carried by wagon 2 is shown in Fig. 4.11.

Since the friction elements are involved, we have inelastic collision of two masses, where one of them was stationary. Assumption was made that the masses are the same. Inelastic collisions do not conserve kinetic energy, but the conservation of moment is in place. After the collision two masses are joined together and travel in the same direction. In Eq. (4.12) speed is shown as a vector quantity. We can see

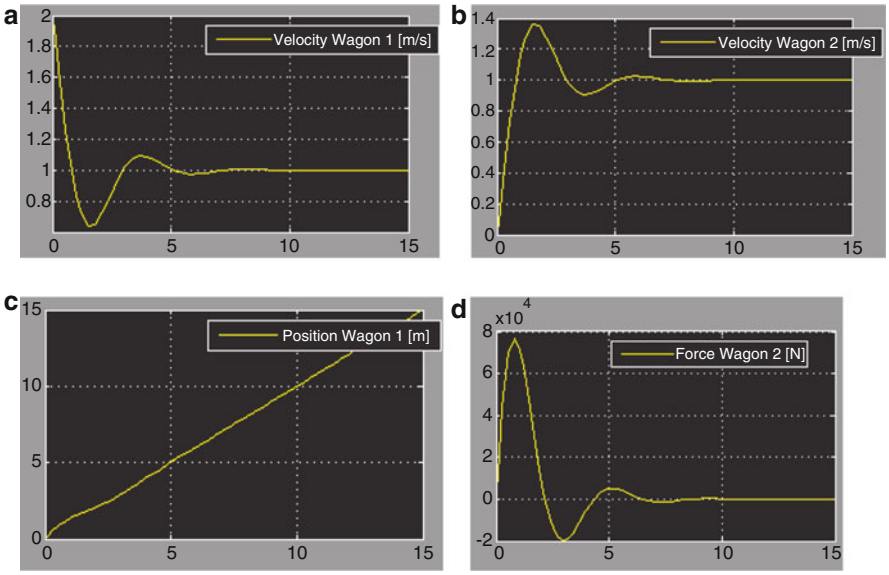
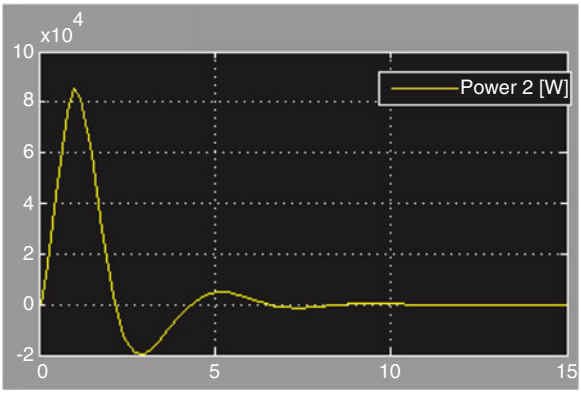



**Fig. 4.9** (**a**) Wagon 1 velocity. (**b**) Wagon 2 velocity. (**c**) Path traveled. (**d**) Force at wagon 2 node

**Fig. 4.10** Power measured at the wagon 2 node




that in the particular case of inelastic collision, as simulated by our model, expected final velocity should be half of the initial velocity. That can easily be verified by looking at Fig. 4.9a, b.

$$M1 = M2 = m,$$

$$m\mathbf{v}_0 = m\boldsymbol{v}_f + m\boldsymbol{v}_f = 2m\boldsymbol{v}_f$$

$$\boldsymbol{v}_f = \frac{1}{2}\mathbf{v}_0 \tag{4.12}$$

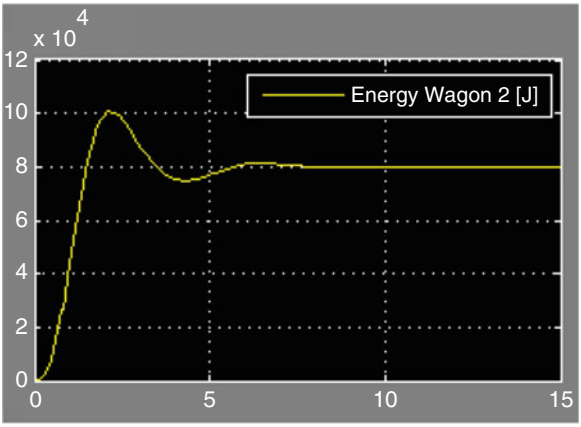



**Fig. 4.11** Kinetic energy carried by the wagon 2

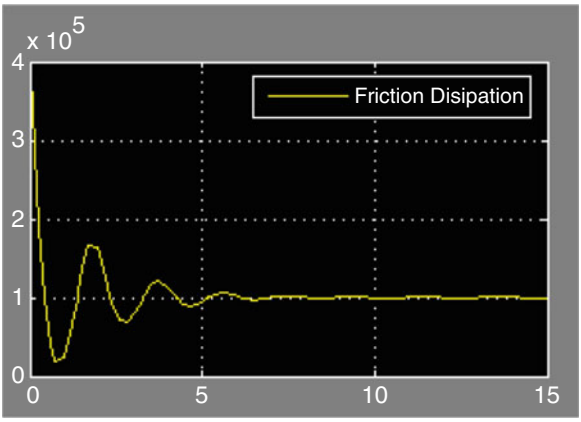



**Fig. 4.12** Energy dissipation in friction element

In Eq. (4.12) $v_0$ is an initial velocity of the wagon 1, as already given, while $v_f$ is the final value of the joint system velocity. We assumed that wagons have the same mass. Energy losses through one of the equivalent friction elements are calculated using expression $E = v^2 B$, as given in Table 4.1. Graph representing simulation results is given in Fig. 4.12.

The spring interfacing element is accumulating and releasing energy as per equation $E = k\frac{x^2}{2}$. Spring energy graph is given in Fig. 4.13.

Figure 4.14 shows distribution of kinetic and spring energy measured at the node 2, as labeled in Fig. 4.7.

Figure 4.15 presents final model of the systems with all sensors, calculations, and monitors shown.

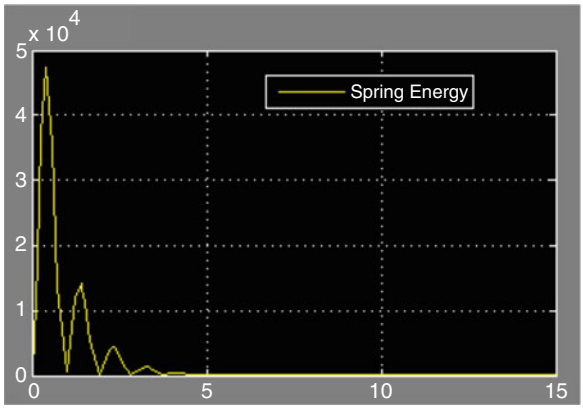



**Fig. 4.13** Spring energy

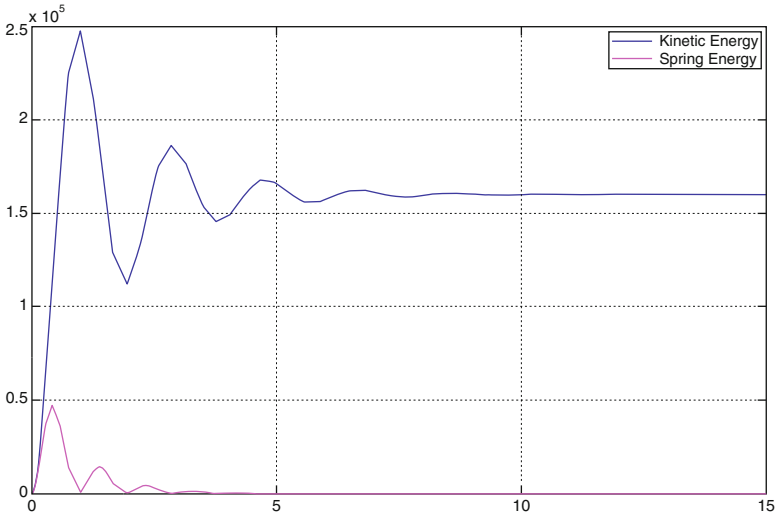



**Fig. 4.14** Kinetic and spring energy distribution measured at the node 2

## 4.6 Train Composition Model

We are now going to simulate whole train composition as shown in Fig. 4.16. The only difference now, comparing to previous modeling, is the way how we supply the energy to the system. Locomotive is simulated as a velocity source as shown in Fig. 4.17. The next Fig. 4.18 shows velocity pattern generated by the locomotive.

As with the previous model, we could add sensors, calculators, and monitors to trace changes in the physical quantities, in the various parts of this complex system.

System model is shown in Fig. 4.19. It can be loaded with more sensors and calculators for monitoring purposes.

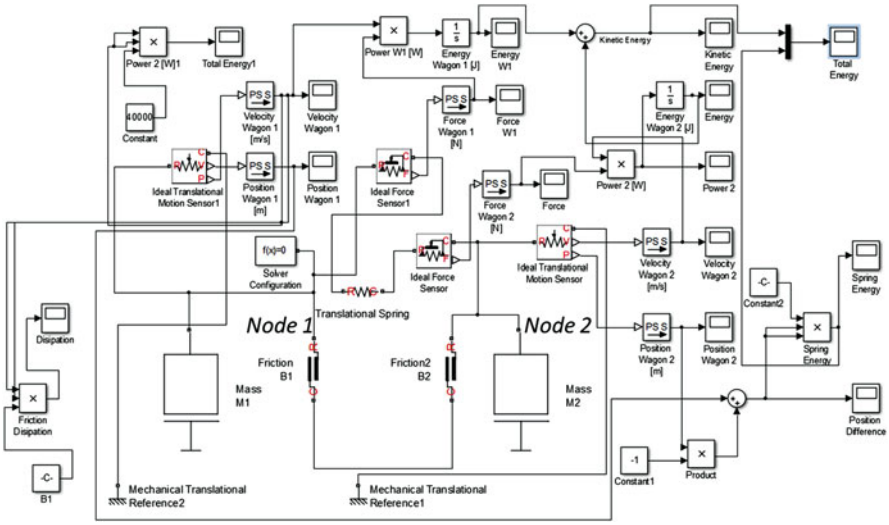



**Fig. 4.15** Two wagon model with sensors, calculators, and monitors attached

**Fig. 4.16** An ordinary train composition subject to simulation

As examples, forces and velocities in few network nodes are presented in Figs. 4.20 and 4.21.

Other physical quantities can easily be monitored, as already shown in the investigation of the less comprehensive system model, with two wagons and initial conditions present.

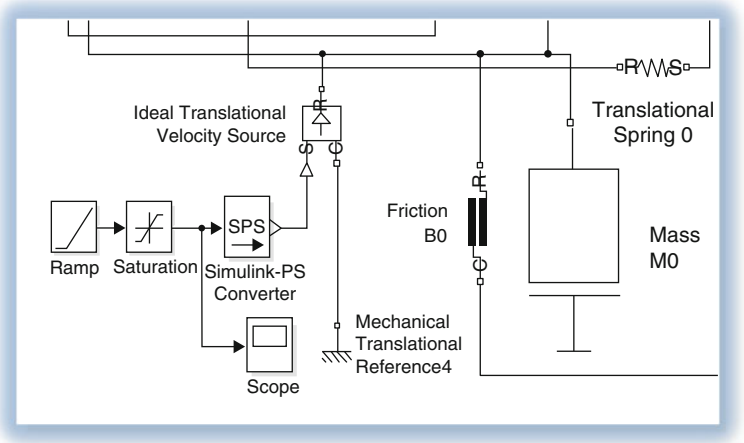



**Fig. 4.17** Velocity driving source

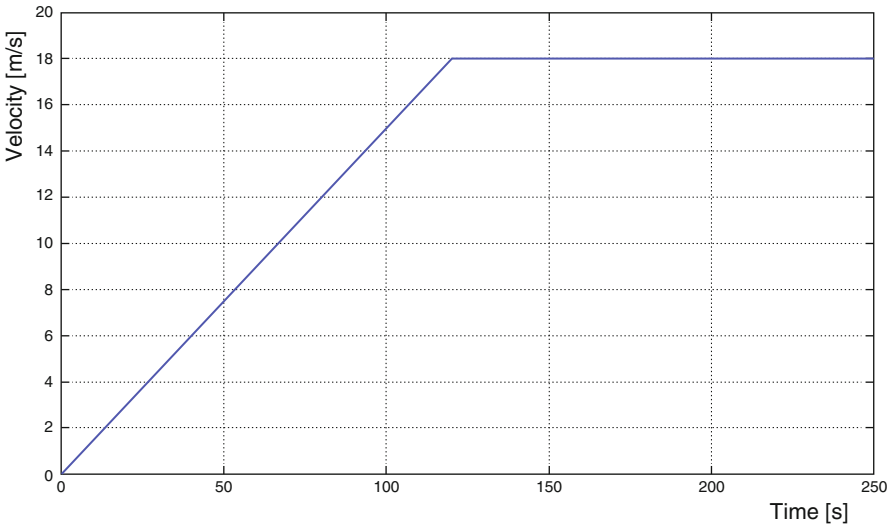



**Fig. 4.18** Velocity pattern

## 4.7 Conclusion

Based on ODE, which are a common way to express various physical systems, physical networks approach is a comprehensive and global tool to model and simulate all sorts of engineering systems. We can perform modeling of mechanical, electrical, or hydro systems easily. Using this approach we simply manage issues

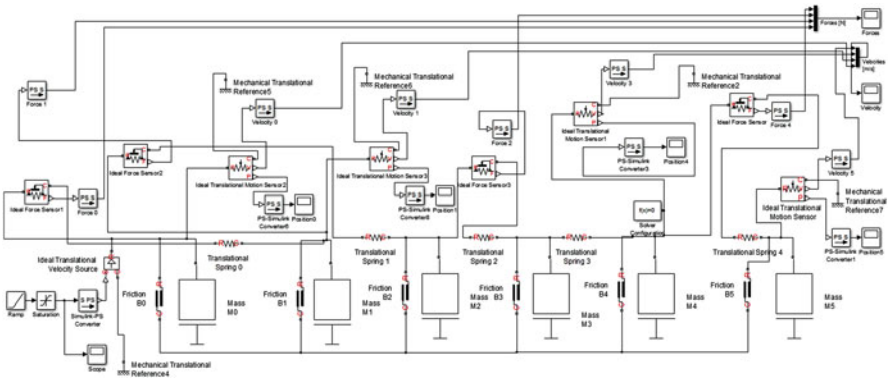


**Fig. 4.19** Basic Simulink model of a train composition as shown in Fig. 4.16

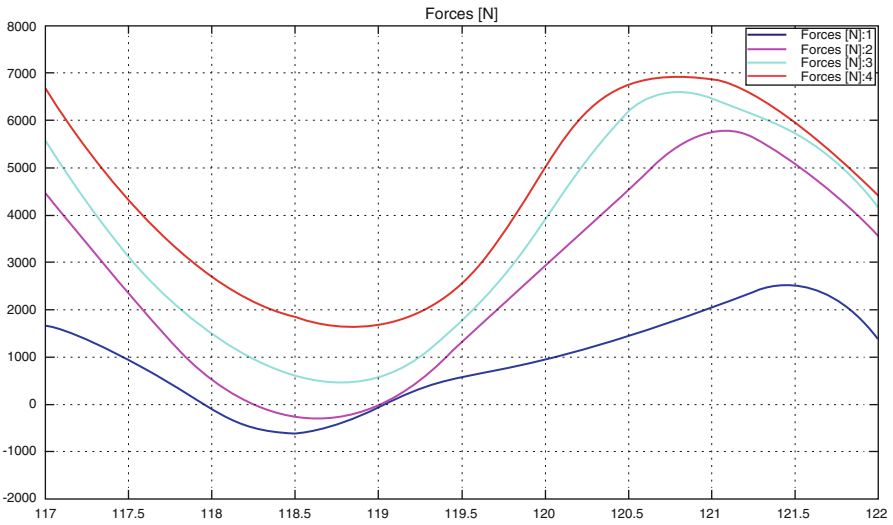



**Fig. 4.20** Forces in the different parts of the train composition system

with energy conversions, from one to another system. In each system two basic types of physical quantities exist, while energy conversions are conducted using various sensors and actuators.

As good examples of mechanical systems that conduct translation motion, train and tram systems were modeled using this approach. Train composition is an extremely nonlinear and multidimensional system. Modeling and simulation is presented in just one dimension, as per translation motion vector directions. It is shown how key quantities and performances of the system can be monitored. Motion in other directions, then vibrations and other phenomena, can also be investigated with more comprehensive modeling and simulations. That will be subject to the future physical network applications and presentations.

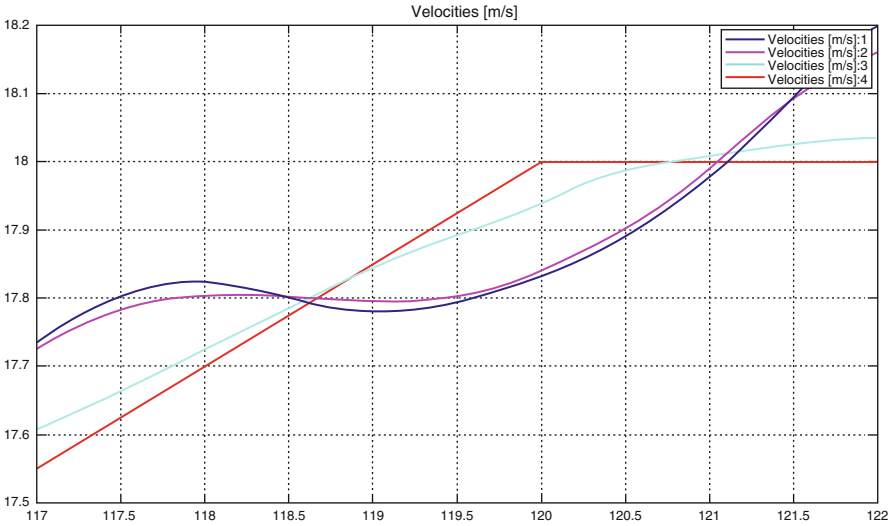



**Fig. 4.21** Velocities of the different parts of the train composition system

## References

*A mathematical model for a train run*. Available: http://diplom.utc.sk/wan/1881.pdf.

de Silva, C. W. (2005). *Mechatronics an integrated approach*. Boca Raton, FL: CRC Press LLC.

Ďurica, M. (2015). A train run simulation with the aid of the EMTP–ATP programme. In A. Abraham, P. Krömer, & V. Snasel (Eds.), *Afro-European conference for industrial advancement* (vol. 334, pp. 99–107). Switzerland: Springer.

Sanford, R. S. (1965). *Physical networks*. Englewood Cliffs, N.J.: Prentice-Hall.

Simic, M. (2015). Exhaust system acoustic modeling. In L. Dai & R. N. Jazar (Eds.), *Nonlinear approaches in engineering applications*. Cham; Heidelberg; New York; Dordrecht; London Springer, pp. 235–249.

# Chapter 5
# Characteristics of Preheated Bio-Oils Sprays

**Heena V. Panchasara**

**Abstract** Global atmosphere pollution has become a serious problem for today. The emissions from the combustion of fossil fuels contribute a notable part to this pollution. Environmental care together with the limited stock and growing prices of fossil fuels has given alternative fuels the potential to supplant a significant portion of fuel for combustion applications such as gas turbine engines and IC engines. Given a widespread of different biofuels available for combustion applications, the present study concentrates on atomization spray characteristics of vegetable oils. Vegetable oils have energy density, cetane number, heat of vaporization, and stoichiometric air/fuel ratio comparable to diesel fuel. Different techniques have been employed so far to improve on the physical properties of biooils and thus opening the doors to clean combustion. The high kinematic viscosity has an adverse effect on the combustion of vegetable oils, posing problems in the associated fuel supply line and injector system. Some well-known techniques to deal with high kinematic viscosity levels of neat vegetable oils include dilution, pyrolysis, micro-emulsion, and trans-esterification. These techniques, however, require additional energy input to improve the physical properties of the fuel. Preheating of the fuel is also one of the ways that reduces the kinematic viscosity to improve the atomization. Preheating is employed in the present study to investigate the spray characteristics in a non-evaporating spray as well as flame spray.

**Keywords** Bio-oil fuel • Vegetable fuel • Alternative fuel • Internal combustion engines • Fuel spray • Combustion system • Fuel injection system • Flare system

## Key Symbols and Key Words

| | |
|---|---|
| SMD | Sauter mean diameter |
| $\lambda$ | Evaporation constant for forced convection |
| $d$ | Droplet diameter |

H.V. Panchasara (✉)
School of Aerospace, Mechanical, and Manufacturing Engineering, RMIT University, Melbourne, VIC, Australia
e-mail: heena.panchasara@rmit.edu.au



R.N. Jazar, L. Dai (eds.), *Nonlinear Approaches in Engineering Applications*,
DOI 10.1007/978-3-319-27055-5_5

| | |
|---|---|
| PDPA | Phase Doppler particle analyzer |
| LPM | Liters per minute |
| $e$ | Emissivity |
| mm | Transverse or radial distance in millimeters |
| LHV | Latent heating value |
| $X$ | Radial distance |
| $Y$ | Transverse distance |
| RMS | Root mean square |
| $t$ | Time |
| $v$ | Velocity |
| $T$ | Temperature |
| $\mu$ | Microns |
| VO | Vegetable oil |
| CO | Carbon monoxide |
| NOx | Nitrogen diaoxide |
| $T_{\mathrm{f}}$ | Fuel temperature |
| LFE | Laminar flow element |

## 5.1 Introduction

The atomization and subsequent propagation of the fuel droplets, their vaporization and combustion are the most important processes concerning the formation of pollutants with the use of liquid fuels. For example, in diesel engines, gas turbine engines, and oil burners, the combustion rate of fuel is controlled by effective vaporization of the fuel. The liquid fuel atomization rate has a strong influence on vaporization rates because the total surface area of the fuel is increased greatly by the atomization process. The fundamental mechanisms of atomization have been under extensive experimental and theoretical study for more than a century (Liu and Reitz 1993). Still, one of the major thrusts in worldwide combustion research has been to gain insight into the physics of liquid fuel combustion in the primary zone of the combustor (Kneer et al. 1994).

Traditionally, the fundamental basis for the atomization process relies on either injecting the fuel under relatively high pressure into a relatively slow moving gas or subjecting the fuel to a high velocity air blast or a combination of two atomization mechanisms. The fuel injector assumes an important role in the combustor by providing some degree of fuel/air mixing close to the atomizer. Almost all combustor performance characteristics are strongly governed by the quality of the spray produced by the fuel atomizer. Calculation of the evaporation and reaction rates in the combustor involves the evaluation of parameters such as mean spray angle, range of drop sizes in the spray, and drop trajectory. It is a well-known fact that the achievement of efficient fuel atomization and rapid evaporation in addition to mixing with combustion air can have a significant effect on such parameters as emissions, exhaust gas temperature profiles, and pattern factor.

There are numerous studies on liquid breakup in the literature. As the relative velocity between the drop and gas increases, the drop breakup regimes such as bag breakup, shear breakup and catastrophic breakup are encountered. The jet breakup as summarized by (Liu and Reitz 1993) comprises of four liquid breakup regimes based on breakup drop size and unbroken jet length. Rayleigh breakup regime: first-wind induced breakup regime; second breakup regime, and the atomization breakup regimes which are encountered as the jet velocities increases. At lower jet velocities, the growth of the small disturbances on the liquid surface due to the interaction with the surrounding gas is believed to be the dominant reason for the liquid breakup. As the jet velocity is increased and the aerodynamic effect becomes more prominent, the breakup mechanisms become increasingly complex.

Air blast atomization is an attractive strategy for liquid fuel breakup in gas turbines. Principally, the air blast atomizer functions by employing the kinetic energy of a flowing airstream to break the fuel jet or sheet into ligaments and then drops. Air blast atomizers offer great advantages over pressure atomizers, since a finer spray can be produced with a lower fuel pumping pressure. Air blast atomization maximizes the interaction between the air and liquid flows by, taking advantage of high velocity airflow to produce fine drops in a well-distributed spray. The atomizing air also serves as emissions reduction strategy improving the fuel–air mixing in the combustor, to reduce soot particulates, CO, and NOx emissions.

The performance of a given spray combustion system depends not only on the fuel droplet size distribution but also on the spray spatial distribution and the interaction of the droplets with the gas turbulence that involves a physical mechanism that is yet to be well understood. For this reason spatially and temporally resolved information such as mean droplet size, droplet size distribution, mean drop velocity root mean square (RMS) velocities needs to be studied to understand most favorable spray conditions for optimal combustion performance.

Physical properties such as kinematic viscosity, surface tension, and volatility are the key parameters that affect the process of fuel atomization and evaporation. The liquid kinematic viscosity affects not only the drop size distribution of the spray but also the fuel injector pressure drop. An increase in kinematic viscosity lowers the Reynolds number, hindering the development of any natural instabilities in the fuel jet or sheet, which helps to further disintegrate the drops. These combined effects delay any further disintegration thus increasing the droplet sizes in the spray. Many alternative fuels are expected to have high kinematic viscosity which makes them difficult to atomize well and thus affecting the combustion efficiency.

A comprehensive study on turbulent diffusion flames using intrusive probing techniques was made by Onuma and Ogasawara (1975). They suggested that the spray flame structure is similar to that of a gaseous diffusion flame in turbulent flow. Chigier and McCreath (1974) used a non intrusive detection technique to measure the Sauter mean diameter (SMD), drop velocity, and number density of air assisted spray and spray flames. A series of experimental and numerical studies of air assisted sprays and spray flames have been made by (McDonell and Samuelsen 1991). Their observations concluded that the presence of fuel drops and reactions alters the structure of the gas-phase turbulence and that local clustering of drops

exists for both non-reacting and reacting cases. A large portion of the experimental research in liquid fuel combustion is focused on pressure atomization mainly in the diesel engine applications. Relatively few studies have been reported on air blast atomization and their potential optimum strategies in alternative fuel combustion. Moreover, very little attention has been given to the evaporation characteristics of the air blast atomized sprays of alternative fuels. Detailed studies on the characteristics of spray flames are necessary to mitigate environment problems and enhance the performance and efficiency of liquid bio fuel combustion systems.

The present work seeks to experimentally investigate the spray characteristics of the fuel droplets in a non-evaporating as well as flame spray conditions using the Phase Doppler Particle Analyzer (PDPA). An air blast atomizer is selected for the present investigation to generate the spray. The photographic image and the key structural features of a flame from a typical air blast atomizer are shown in Fig. 5.1. Larger drops are distributed mostly on the outer edge of the spray in region "C" and the smaller drops are located in the center region "A." These larger drops on the edge are affected by the presence of the flame. Due to slower evaporation rate of the large drops, a blue lean reaction sheath is formed inside the spray boundary as seen in the

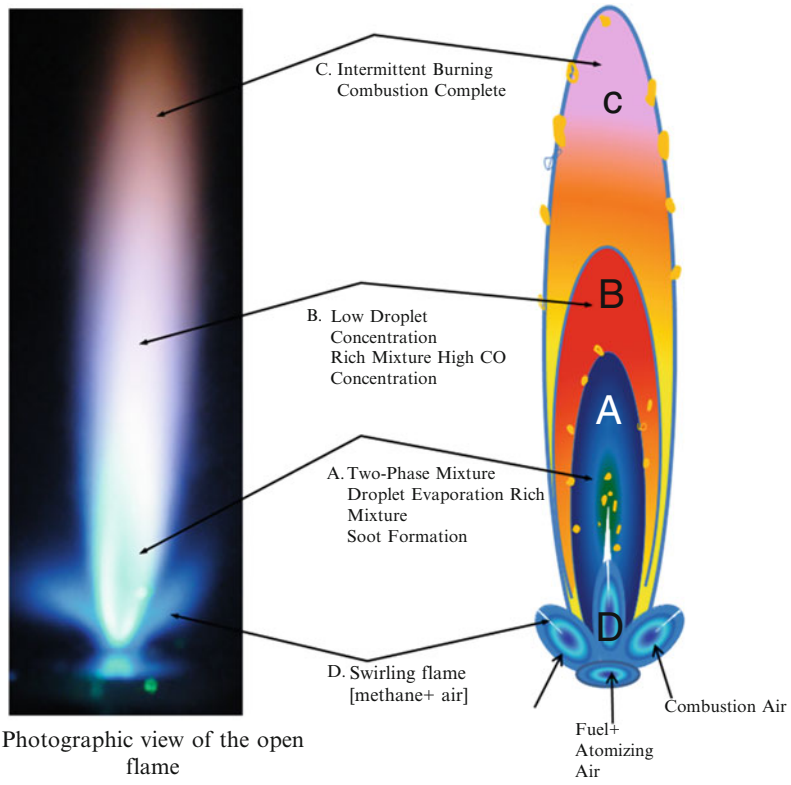



**Fig. 5.1** Flame structure of an air blast atomizer

photographic image. The main reaction zone is a mixture of fuel vapor and air that is sufficient for combustion. In the central core of the spray, as seen in region "A" the fuel vapor concentration is too rich to allow the chemical reactions to take place. Above the spray flame, the smaller droplets produce tiny blue streak flames as well as the larger drops that burn incompletely produces a yellow orange streak flame flying in all directions (regions "B" and "C"). Mostly on the outer periphery of the spray flame, single drop burning with self-sustained envelope flame is observed. The swirling air flow improves mixing and creates a homogenous combustible mixture and hence stabilizes the liquid fuel flame. This is shown in Fig. 5.1 by the region "D" which is a methane–air mixture flame. The swirling combustion air makes the liquid sheet unstable and hence helps any further disintegration in to smaller droplets that follow the flow. Figure 5.2 shows the droplet vaporization in spray flames. The spray

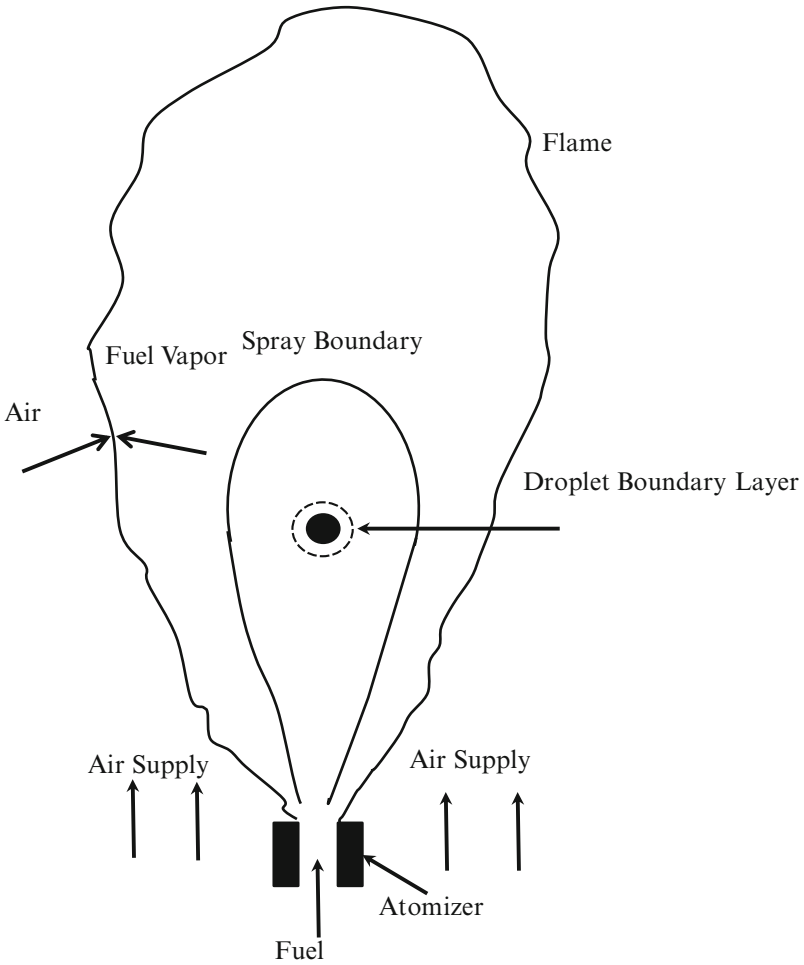



**Fig. 5.2** Vaporization of a typical droplet in an idealized spray flame

consists of initial cool and hot zones. Within the cool zone, heat transfer is restricted to radiation from the flame front to the droplet surface. In the hot zone, heat transfer takes place both by radiation from the flame front and by turbulent convection. The droplet diameter reduces by the $d^2$ law:

$$-d\left(d^2\right)/dt = \lambda$$

where $\lambda$ is the evaporation constant for forced convection.

The objective of the present work is to investigate the effects of combustion on spray and spray flames of bio-oil for which little experimental data are available. Experiments are conducted using unheated VO as well as heated VO at 100 °C. Measurements are obtained for both in open flame and an enclosed flame to simulate realistic gas turbine conditions. The mean axial and RMS velocities, SMD and drop size distribution data are acquired. The primary focus is placed on liquid fuel spray characteristics, and their effects on emissions. It can be envisioned that smaller droplets in the spray would lead to premixed combustion and hence lower emissions. The inferences from this study would aid in designing future liquid fuel combustors. Hereafter the details of the experimental set-up, results and discussion, and conclusions are presented in this chapter.

### *5.1.1 Experimental Set-Up and Procedure*

The experimental set-up is shown schematically in Fig. 5.3. It consists of the combustion, fuel injection, and flare systems as discussed below. PDPA mounted on a 3-way traversing system was used to acquire quantitative spray measurements in the cold flow as well as flame. In addition, emissions data were acquired by a continuous sampling probe and an infrared camera was used to record the combustor wall temperature.

### *5.1.2 Combustion System*

Combustion system was housed within a test chamber of dimensions 63.5 cm by 63.5 cm by 1.2192 m. The primary air enters the combustion system through a plenum filled with marbles to breakdown the large vortical structures. The air passes through a swirler into the mixing section, where the gaseous fuel is supplied during the startup. The reactants or combustion air enters the combustor through a swirler section shown schematically in Fig. 5.3. The swirler is used to enhance fuel–air mixing and it also helps to stabilize the flame. The swirler has six vanes positioned at 28° to the horizontal to produce swirl number of about 1.5. The bulk

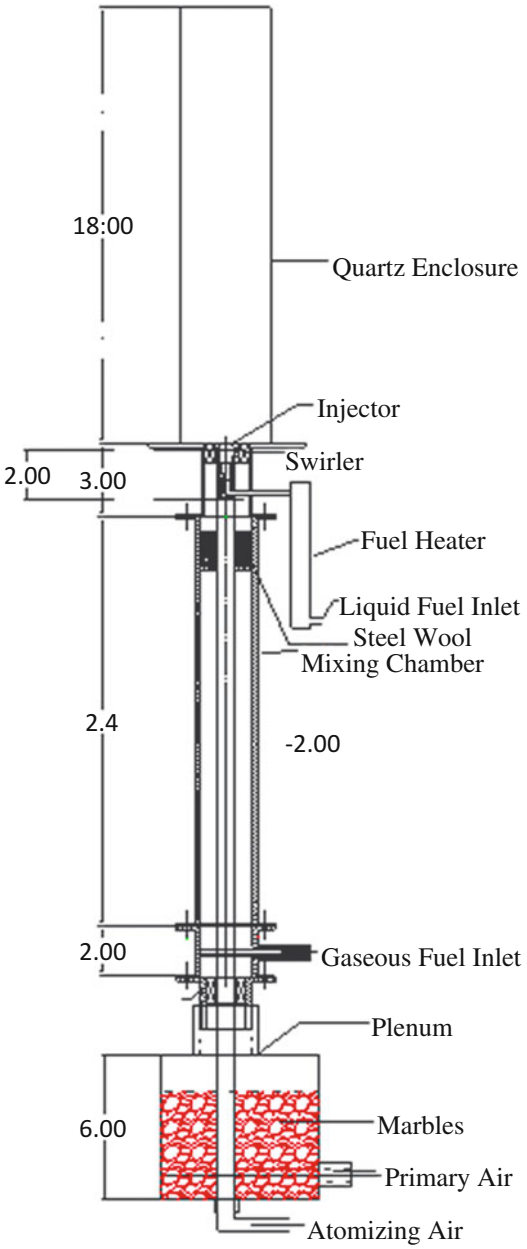



**Fig. 5.3** Schematic of the combustor experimental set-up

axial inlet velocity of the primary air is 1.9–2.1 m/s, which resulted in Reynolds number varying from 5960 to 6750. The combustor is enclosed within a 15 cm inside diameter, 46 cm long pentagonal enclosure, with two sides of quartz glass and remaining three sides of metal plates.

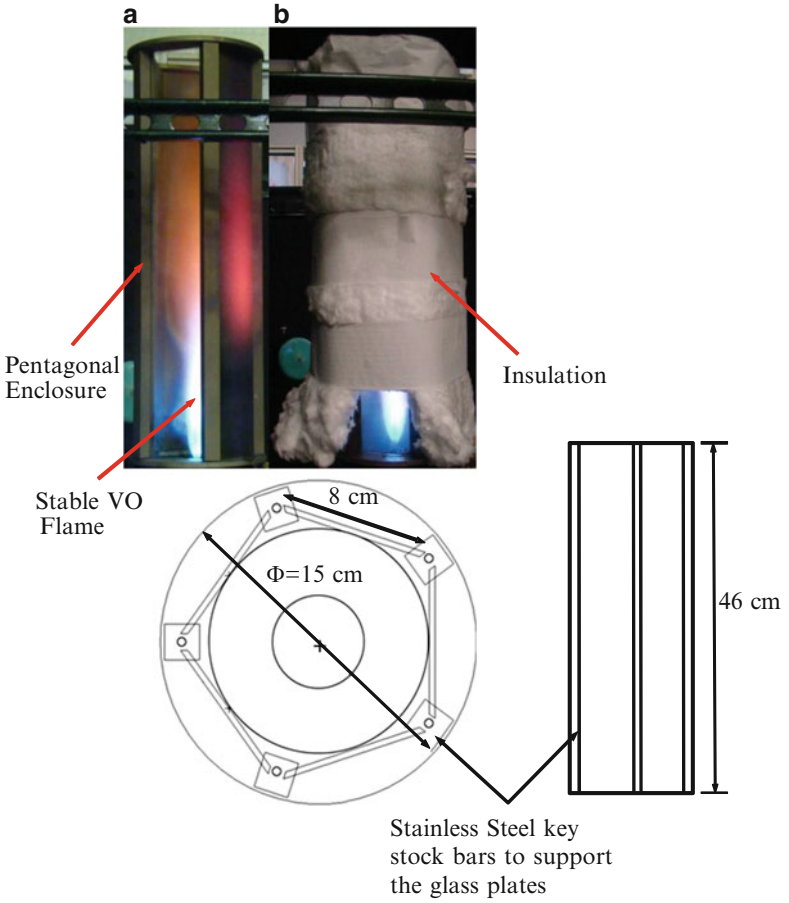



**Fig. 5.4** Photographic view of the enclosure (**a**) without and (**b**) with insulation and schematic of the top view and vertical view of the pentagonal enclosure

In order to have optical access and to be able to measure the spray data, a flat sided pentagonal enclosure was designed and fabricated in house. The photographic view of the enclosure with out and with insulation and the schematic of the top view of the pentagonal enclosure are shown in Fig. 5.4. Transparent quartz glass windows of 8.0 cm by 46.0 cm by 3 mm thick, grounded and polished to optical quality were used for the enclosure to allow optical access through the enclosure. The enclosure was insulated with 4 layers of insulation to minimize the heat loss from the flame. Alumina Mat type Buster Blanket insulation (1″ standard thickness) from Zircar with properties listed in Table 5.1 was used to insulate the enclosure. Buster blanket is a flexible, Hi-Alpha fiber insulation material. It is needled into a durable blanket with the addition of organic fiber reinforcement with temperature up to 1600 °C. These polycrystalline fiber blankets offer higher

**Table 5.1** Properties and characteristics of the insulating material (www.zircarzirconia.com)

| Nominal composition, wt% | Buster blanket |
|---|---|
| $Al_2O_3$ | 97 |
| $SiO_2$ | 3 |
| Organics | <3 |
| Trace elements | <0.5 |
| Melting point, °C (°F) | 2038 (3700) |
| Maximum use temperature, °C (°F) | 1600 (2912) |
| W/mK (BTU/h $ft^2$ °F/in.) at 1600 °C (2912 °F) | 0.476 (3.3) |
| Bulk density, g/cc | <4 |

temperature capability, less shrinkage, and greater chemical resistance than standard alumina–silica blankets. The insulation helped retaining heat inside the combustion area and hence improves flame stability. For optical access a small window of 10.16 cm by 10.16 cm was cut on the insulation on two glass windows at the bottom of the enclosure as observed in the photographic image of Fig. 5.4.

Air for combustion and atomization was supplied by an air compressor. The air passed through a pressure regulator, and a dehumidifier and water traps to remove the moisture. Then, the air was split into combustion air supply and atomizing air supply lines. The combustion air flow rate was measured by a laminar flow element (LFE) with reported calibration error of ±5 LPM. The pressure drop across the LFE was measured by a differential pressure transducer. An absolute pressure transducer was used to measure air pressure in the LFE. The atomizing airflow rate was measured by calibrated mass flow controller from Sierra (Model 810S-M., Mass-Track Flow Controller, 15–100 SLPM), with measurement uncertainty of ±0.5 LPM.

### *5.1.3 Fuel Injection System*

The fuel injector system runs through the plenum and the mixing chamber, as shown in Fig. 5.3. An O-ring within a sleeve is located at the bottom of the plenum to prevent air leakage. A commercial air blast atomizer (Delavan Siphon type SNA 0.20 nozzle) with its details shown schematically and photographically in Fig. 5.5 was used for the experiments. The injector creates a swirling flow of atomizing air to breakdown the fuel jet as the fluids exit the injector. The inside diameter of the fuel supply tube is about 0.4 mm.

The liquid fuel supplied by a peristaltic pump passes through a pulse dampener, a fuel filter, and an electric fuel heater (Infinity Fluids Corporations, CRES-ILB-12/24 inline water/liquid heater) as shown in Fig. 5.6. The heater uses a Proportional Derivative Controller (PID) control unit to control the fuel temperature within

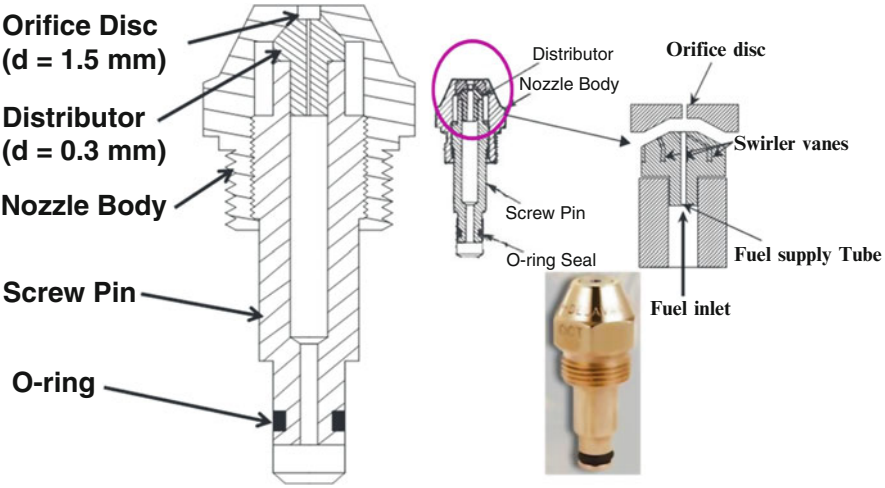



**Fig. 5.5** Air blast injector details

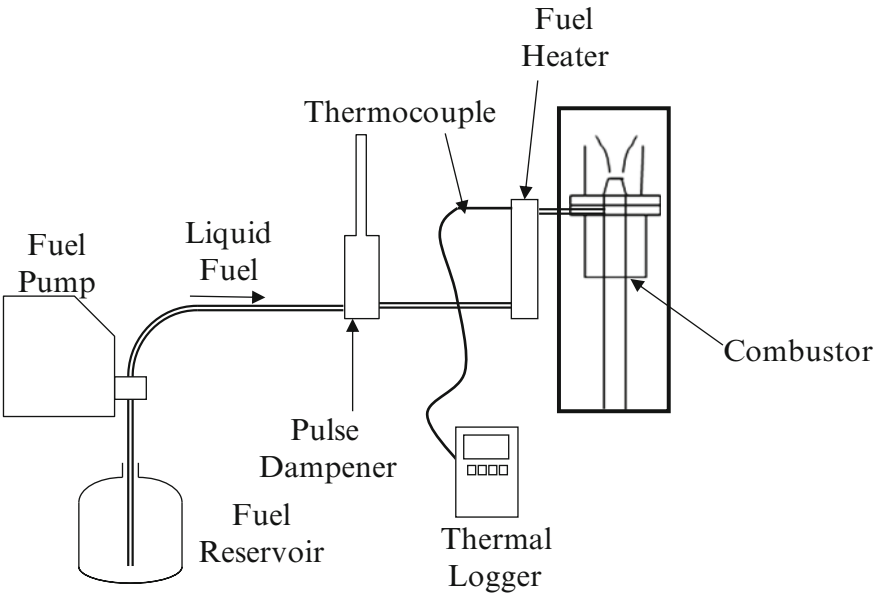



**Fig. 5.6** Schematic of liquid fuel supply system

accuracy of ±0.5 °C, measured at the heater outlet by a K-type thermocouple. Heated fuel enters fuel injector through 5 cm long tube of 5 mm ID to minimize the heat loss. The fuel temperature at the injector inlet was measured by a K-type thermocouple (Fig. 5.7).

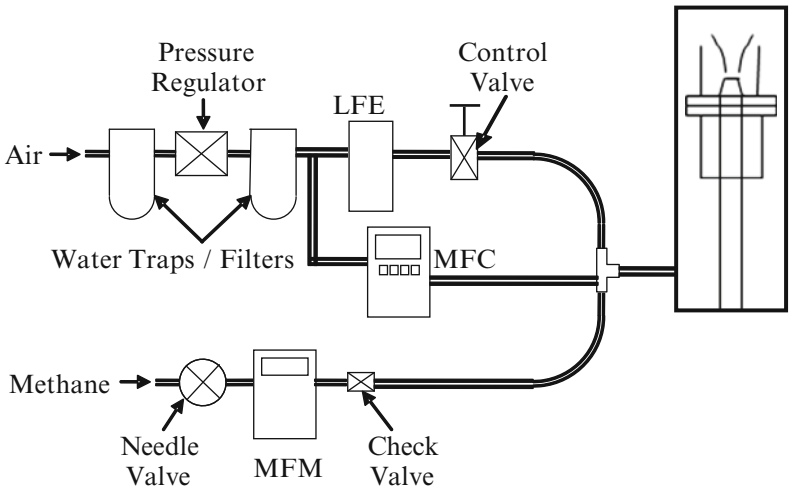



**Fig. 5.7** Schematic of gaseous fuel/air supply system

### *5.1.4 Flare System*

To burn the fuel droplets before they enter the exhaust system was a major challenge while conducting cold flow measurements. Thus a flare system that could successfully combust the liquid fuel upstream of the exhaust duct was designed in house. Several different approaches were attempted to create a flare system to burn most of the fuel droplets upstream of the exhaust duct while also avoiding bellowing of the spray in the operating area. Figure 5.8 shows the photographic view of the final flare system used with open cold spray. Spray measurements in the cold flow were acquired with and without swirling flow of combustion air. The flare system consists of 3 diffusion torches with co-flowing methane and air streams as shown in Fig. 5.8. The flare system was located at about 60.96 cm downstream of the open cold spray. The methane flow rate through the flare system was kept to about 40 SLPM while the air flow rate was about 300 SLPM. As seen from the photographic image, the open cold spray is encompassed in the long flame from the flare creating a combustible mixture with most of the fuel droplets burning upstream of the exhaust duct as shown in Fig. 5.8. Sufficient air was entrained to burn the fuel droplets using the co-flow air. The flares ensured that the fuel spray mixture passed through a sufficiently long flame so that enough residence time and high flame temperature would ensure proper droplet burning of the spray upstream of the exhaust duct.

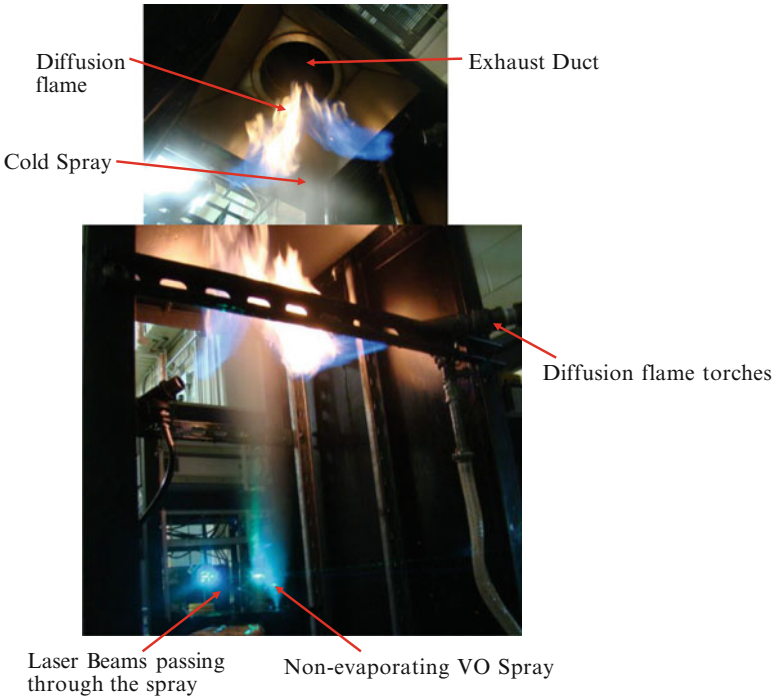



**Fig. 5.8** Photographic representation of the flare system

### *5.1.5 Emissions Measurement System*

The product gas was sampled continuously by a quartz probe (OD = 7.0 mm) attached to a three-way manual traversing system. The upstream tip of the probe was tapered to 1 mm ID to quench reactions inside the probe. The probe was traversed in the axial direction at the center of the combustor and in the radial direction at the combustor exit plane. The analyzer consists of a built-in sample pump which draws the flue gases through the stainless steel probe, pre-cooler, condensate trap and filter, flow meter, internal secondary filter, Teflon liquid blocker, and the sample collecting sensors as the basic built-in functions. Water traps mounted upstream of the emission analyzer prevent the sensors from the contamination. The dry sample was sent through the electrochemical analyzers to measure the concentrations of CO and NOx in ppm. As shown in Fig. 5.9, a NOVA gas analyzer was used to measure the concentrations of CO and NOx in the exhaust gas sample. The emission concentrations are measured using electrochemical sensors. Electrochemical "fuel cell" type sensors produce a small electrical output proportional to the gas being detected. The output signal is amplified and displayed on LCD digital panel meters. The emission analyzer measures the concentrations of CO in the range of 0–2000 ppm and NOx in the range of 0–1000 ppm. The analyzer also measured oxygen and

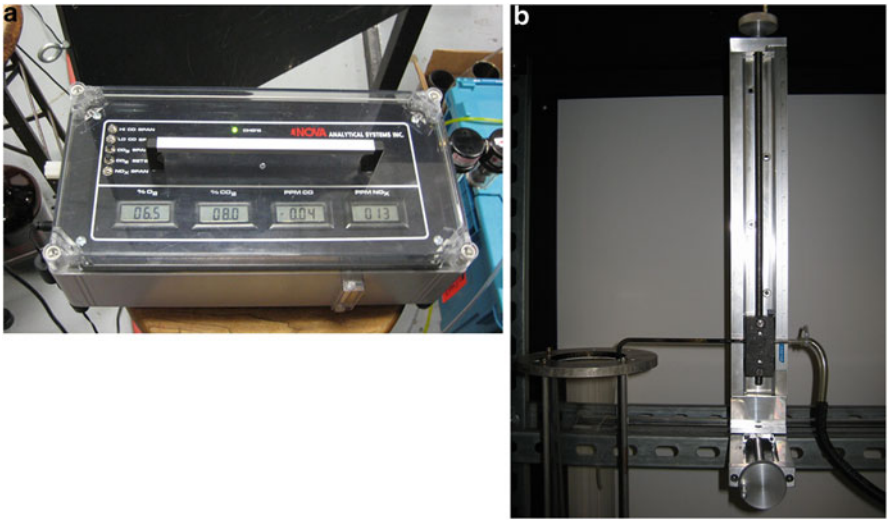



**Fig. 5.9** (**a**) Emissions analyzer; (**b**) Emissions measurement traversing system

carbon dioxide concentrations, which were used to cross-check the equivalence ratio computed from the measured fuel and air flow rates. The uncorrected emissions data on dry basis are reported with uncertainty of $\pm 2$ ppm.

### *5.1.6 Phase Doppler Particle Analyzer Set-Up*

A 2D PDPA was used for flow velocity and drop size measurements. PDPA is a point sampling device based on the light scattering interferometery principle. The laser beams from the transmitter probe intersect to form a sample measurement volume. Principally when a particle or drop passes through the beam intersection region, the scattered light forms a fringe pattern. Since the droplet is moving, the fringe pattern sweeps past the receiver aperture at the Doppler difference frequency, which is directly proportional to the drop velocity. The droplet size is measured by the phase shift of the light encoded in the spatial variation of fringes reaching two detectors after travelling paths of different lengths through the droplets. The phase shift is measured by the two detectors, each focused at a spatially distinct portion of the receiver lens. The spatial frequency of the fringe pattern is inversely proportional to the drop diameter. The phase shift between the Doppler burst signals from different detectors is proportional to the diameter of the spherical shaped droplet. There is no calibration required for the PDPA method since the drop size and the droplet velocity are dependant only on laser wavelength and optical configuration.

The schematic of the 2D PDPA system is shown in Fig. 5.10. The laser beam from a 2-W water-cooled argon–ion laser is separated into a pair of 514.5 nm

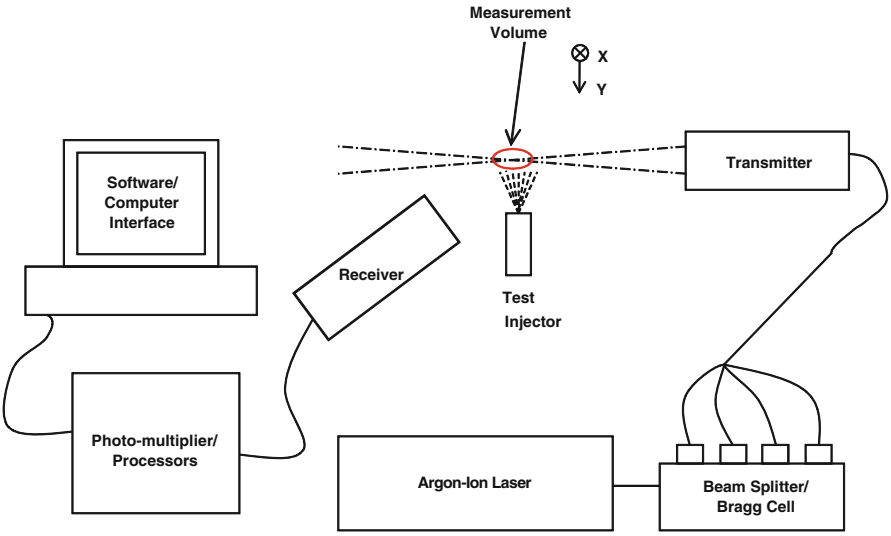



**Fig. 5.10** Schematic of the PDPA system

green beams and a pair of 488 nm blue beams using a beam separator assembly. One beam of each pair is shifted by a 40 MHz Bragg cell. Next, each beam is focused onto a fiber optic cable to deliver the beams to a 750 mm focal length PDPA transmitter. In order to be able to take flame spray measurements using PDPA, the lenses were replaced to the focal length of 750 mm (29 in.) for both transmitter and receiver probes so that the optical probes are kept away from the flame. Precisions Achromatic Doublet lenses from Newport Optics were used. The transmitter lens of 50.8 mm diameter, 750 mm focal length (Model No. PAC086) and the receiver lens of 76.2 mm diameter, 750 mm focal length (Model No. PAC046) are mounted on the respective probes. The PDPA receiver is set at an angle of 144° from the transmitter to collect the refracted light intensities from the spray. The detected signal is acquired by a data acquisition system, and analyzed using the TSI Flow Sizer software to obtain mean and RMS velocities, SMD, and drop size distribution data. The photographic view of the PDPA system integrated with the combustion system is shown in Fig. 5.11.

### *5.1.7 Traversing System*

Measurements were obtained by moving the PDPA system using a three-way traversing system, while the combustor was kept stationary. The PDPA system was traversed to acquire radial profiles at axial planes between $Y = 5$ and 75 mm, in 5 mm intervals for the cold spray as well as from $Y = 5$ and 40 mm, in 5 mm

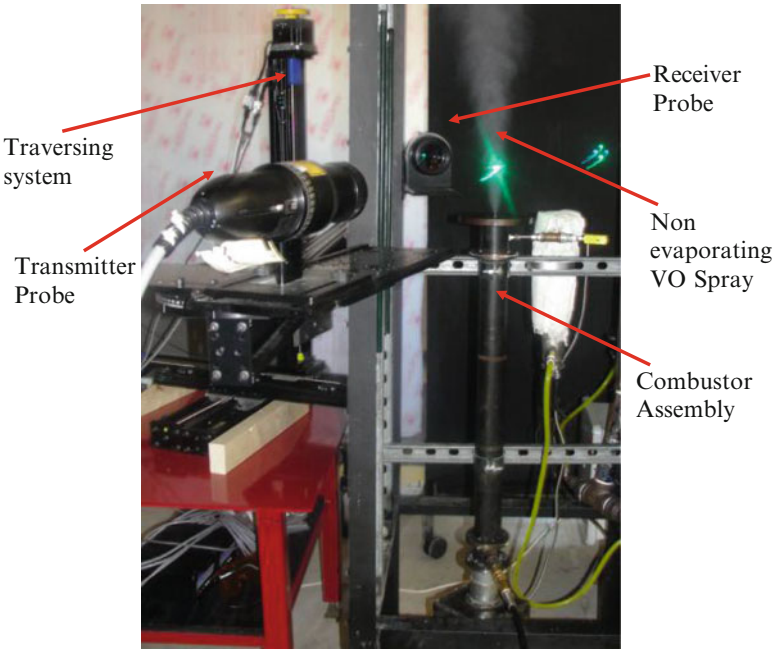



**Fig. 5.11** Photographic view of the PDPA system integrated with the combustor assembly

intervals for the flame spray measurements. The 3-way traversing system (Model No. MT10-0100-M02-31), from Velmex Inc., was used to mount the PDPA system. The schematic of the system is shown in Fig. 5.12. The plan view of the PDPA traversing mechanism in radial and axial direction is shown in Fig. 5.13. Each ($x$, $y$, and $z$) axis was traversed by the single shaft stepper motors Model No. PK296-03AA-A6-3/8. The $Y$-axis was traversed to measure the radial velocity and the $Z$-axis measured the axial velocity. The traversing system was bolt mounted on a carbon steel table of 8.0 cm by 8.0 cm and 82 cm high. The transmitter and the receiver probes were mounted on the rail attached to the $Z$ axis of the traversing system. To obtain a good signal quality, the angle between the transmitter and receiver was maintained at 144°. To obtain this angle, both the probes were set at 18° with respect to the optical rail attached to the $Z$ axis of the traversing system carrying the transmitter and receiver. To simplify the measurements in radial directions, we aligned the transmitter direction with respect to one of the traversing system axes, i.e., the $X$ axis. This demands the optical rail to be set at precisely the same angle with respect to $X$ axis. This arrangement of setting the optical rail at 18° facilitates to take radial location measurements along the $Y$ axis. The center of the spray pertained to the peak location of the axial velocity profile. Data rates of up to 30–40 kHz were obtained towards the center of the spray. The data rate and mean axial velocity both decreased to nearly zero as the detection volume reached the outer edge of the spray.

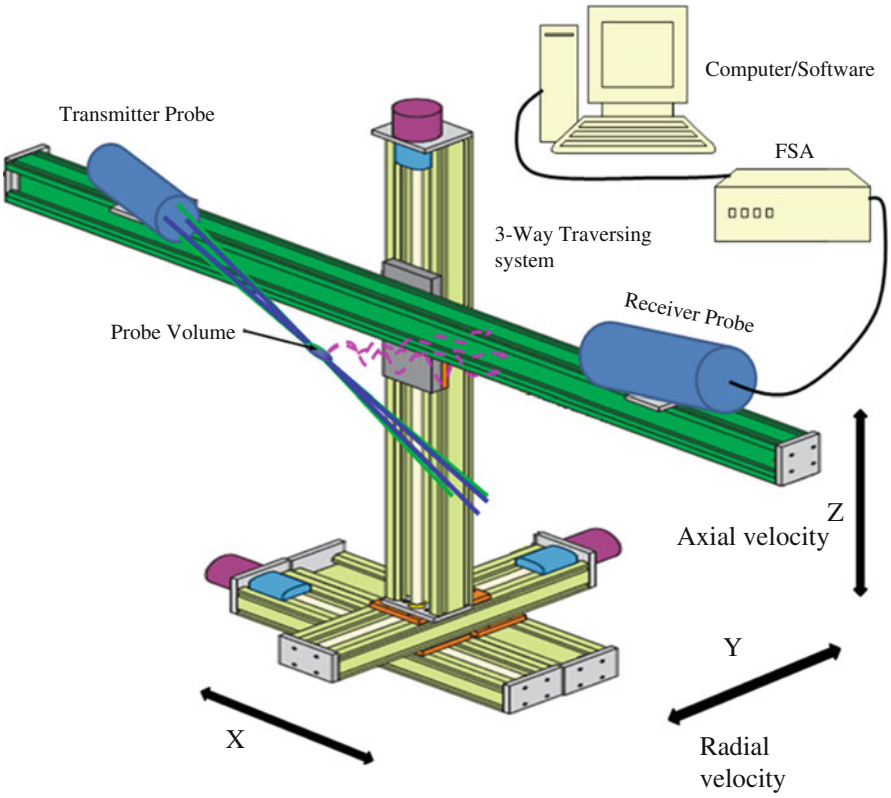



**Fig. 5.12** Experimental set-up of a PDPA system mounted on a 3-way traversing system

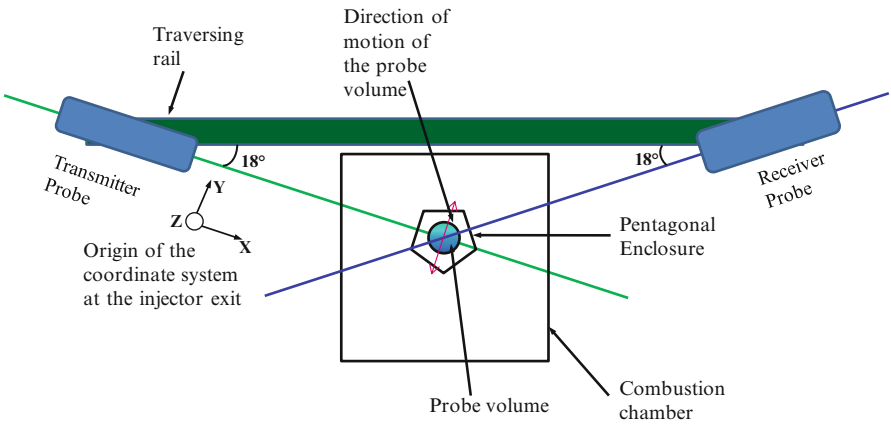



**Fig. 5.13** Plan view of the PDPA traversing mechanism in radial and axial coordinates

**Fig. 5.14** Photographic image of the MikroScan 7200v infrared camera

### *5.1.8 Infrared (IR) Imaging*

An infrared camera was used to measure the exterior surface temperature of the insulated enclosure as shown in the photographic image in Fig. 5.14. The camera used was from Mikron Infrared, Inc, Model No. 7200v. The IR imaging camera (MikroScan 7200V) is a non-contact, high sensitivity infrared radiometer. It measures the infrared radiation emitted by the target surface and converts this radiation into a two-dimensional image related to the temperature distribution at the target surface. The IR camera senses thermal energy that is emitted from the target object. Through the use of the software (MikroSpec), temperature variations over the area included in the field of view can be displayed. The total radiant emission of a blackbody is given by Stefan Boltzmann equation

$$W = aT^4.$$

The temperature of the blackbody can be obtained directly from the radiant energy of the blackbody by this equation. For normal objects, the right side of the above equation multiplied by the emissivity. The data processing software (MikroSpec) allows the user to view the temperature data of one or more points at selected locations anywhere within the field of view. The camera was mounted on a regular tripod. Two different temperature ranges were used to take IR images at different location on the enclosure. The ranges used were 0–500 °C and 0–2000 °C. The default emissivity value of 0.85 was used.

### 5.1.9 Test Conditions

Experiments were conducted for fixed VO flow rate of 12 LPM. The atomizing air flow rate through the injector was kept constant at about 25 LPM which corresponds to an ALR of 2.5. The combustion air flow rate was kept to about 125 LPM to maintain a total air flow rate of 150 LPM throughout the experiments. As a first step the cold open spray experiments were done with and without the swirling air to study the effect of swirling air on the spray. Then the open flame spray data were taken which required swirling air flow with small amounts of methane fuel to stabilize the flame. Experiments were conducted with unheated and heated (to 100 °C) VO. The PDPA system was traversed radially and axially to acquire flow and drop size measurements at different locations in the open cold spray as well as open flame. Next measurements were acquired in spray flame within an insulated enclosure. Again, the VO was supplied either at room temperature or at 100 °C.

The experiment was started by supplying gaseous methane and then, igniting the methane–air reactant mixture in the combustor. Next, the liquid fuel flow rate was gradually increased in steps of 2 mlpm to attain the desired value of 12 mlpm, while the methane flow rate was slowly decreased to zero. In this study, the volume flow rate of total air (combustion + atomizing) was constant at 150 SLPM. Experiments were conducted for fixed volume flow rate of fuel at 12 mlpm. The time required for the fuel to flow through the system and reach the atomizer was about 30 min. Another 0–40 min was required before the fuel temperature reached the set value.

Once a stable flame was achieved, the liquid fuel flow rate was gradually increased while the methane flow rate was gradually reduced. Initially the methane fuel was supplied at 18 LPM for about 1.5 h and then gradually reduced to about 3.88 LPM. This procedure took about 2.5 h before a stable VO flame at 12 mlpm was achieved without any condensation in the optical measurement area. The equivalence ratio ($\varphi$) was maintained constant at 0.89–0.9 throughout the experiment. The emissions measurements were taken for the spray flame along with the spray measurements using a 2-way velmex traversing system shown in Fig. 5.9b.

### 5.1.10 Results and Discussion

Experiments were conducted to measure the spray characteristics to demonstrate the (1) effects of swirling air flow on open cold spray, (2) effects of flame on open spray, and (3) effects of enclosure on spray flame and emissions. The effect of fuel inlet temperature is also studied for unheated and preheated VO at 100 °C. The following sections will discuss important observations from the current study. Results are shown in the form of contour plots and profiles of axial velocity, RMS velocity, SMD, and droplet diameter distribution. For each case, the contour plots will be discussed first followed by the profiles plots.

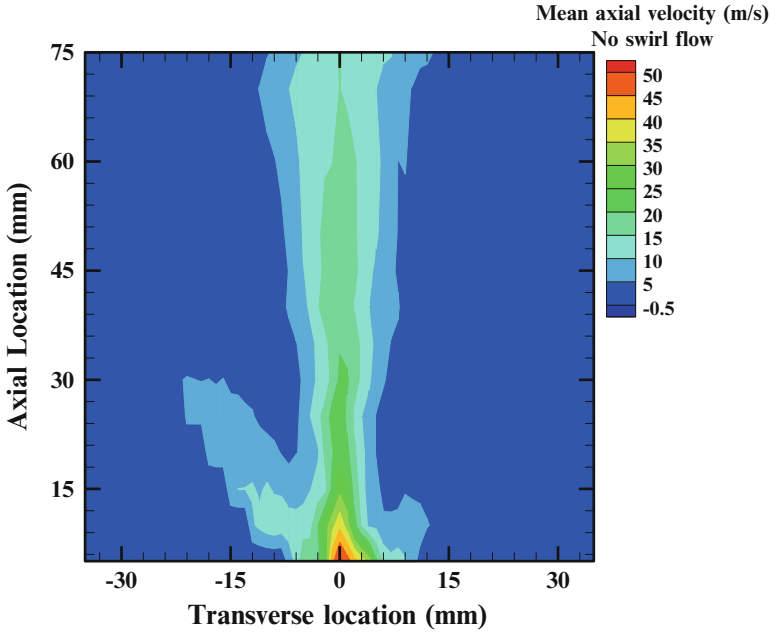



**Fig. 5.15** Axial velocity contour for cold spray without swirl

#### 5.1.10.1 Effect of Swirling Air Flow on Open Cold Spray

Axial Velocity Contours (Mean and RMS)

Figures 5.15 and 5.16 show the contour plots of axial velocity for cold spray with and without swirl, respectively. The plots were constructed using measurements taken for different axial and radial locations in the spray. The axial location was varied from 5 to 75 mm from the injector exit in steps of 5 mm interval. Transverse measurements were taken in steps of 1 mm across the spray. As mentioned earlier, the experimental system was kept stationary while the PDPA system was traversed radially and axially.

In general, for the contour plots shown in both the cases, the mean axial velocity peaks in the center of the spray, with the velocity magnitude diminishing towards the edge of the spray. The peak axial velocities are seen to be slightly higher in the near field axial locations from 5 to 20 mm in the cold spray with swirl as compared to the cold spray without swirl. Farther downstream in the mid field ($Y = 20$–$40$ mm) and far field ($Y = 45$–$75$ mm) locations in the spray, the peak axial velocities for cold spray without swirl are higher as compared to the cold spray with swirl. For both cases as the flow field diverges axially downstream, the velocity decreases with the loss of kinetic energy of the droplets. For the cold spray without swirl, the axial velocity in the midfield and far field locations is about 12 and 8 m/s higher

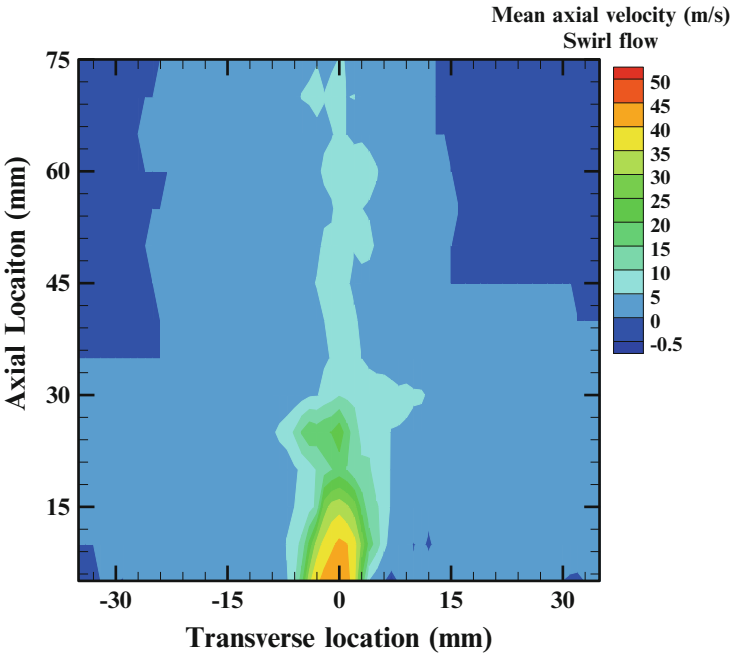



**Fig. 5.16** Axial velocity contour for cold spray with swirl

as compared to cold spray with swirl. Figure 5.16 reveals a well-pronounced flow recirculation zone near the outer edge of the spray between the radial locations of 10 and 40 mm. Note that the near field region of the spray, the values of axial velocity are negative. The presence of swirl widens the spread of the spray and increases the velocity near the injector exit while drastically reducing the velocity magnitude farther downstream compared to cold spray without swirl. Near the injector exit, the effect of swirl is to create high axial velocity owing to the increased airflow rate. However, farther downstream, the swirling air flow introduces additional turbulent mixing to widen the radial extent of the jet spray with sufficient reduction in axial velocity as compared to cold spray without swirl.

Figure 5.17 shows the mean axial velocity contour plot for cold spray with swirling air flow for preheated VO at 100 °C. As expected, the axial velocities show an increase in magnitudes compared to the unheated case. This increase in the axial velocity results from the heating of the atomizing air by heated VO. For preheated VO, the peak velocities at any given axial location in the spray are higher by about 10–15 m/s as compared to the unheated VO case in Fig. 5.8. With increase in fuel inlet temperature, the radial extent of the spray also decreases as compared to the unheated case. Higher axial momentum decreases the radial diffusion of the jet spray, resulting in a narrower spray in case of preheated fuel.

Figures 5.18, 5.19, and 5.20 show the contour plots for RMS axial velocity for no swirl flow, swirl flow, and cold spray with swirl flow for VO at 100 °C, respectively.

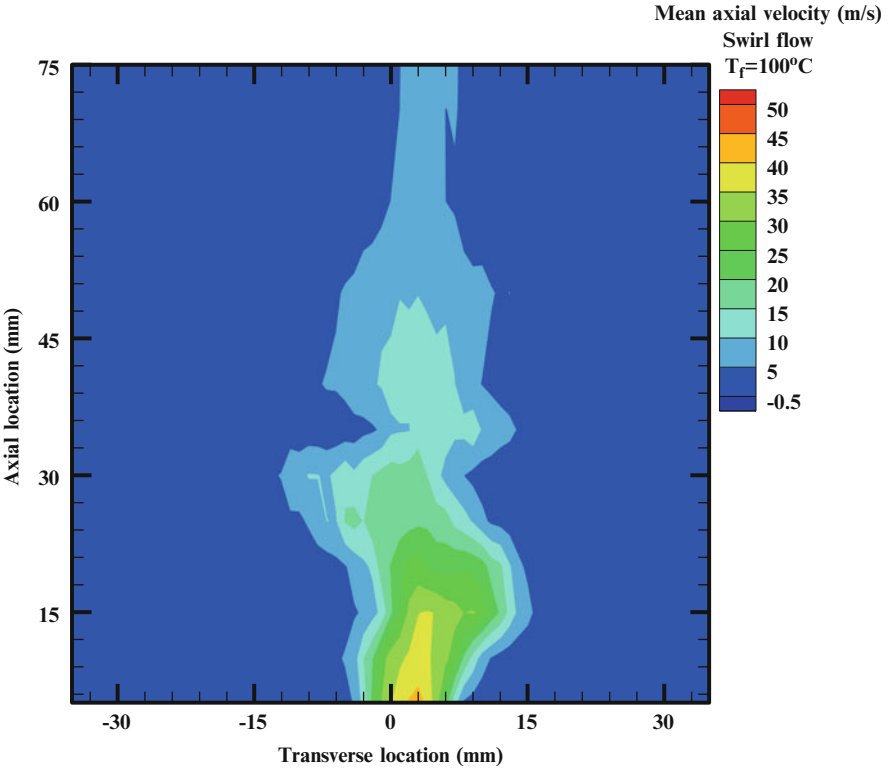



**Fig. 5.17** Axial velocity contour for cold spray with swirl at $T_f = 100\ ^\circ C$

As shown in Fig. 5.15, for cold spray without swirl flow, the peak RMS axial velocity value is observed to be of 18 m/s while for cold spray with swirl the peak RMS velocity in Fig. 5.19 is around 22 m/s. Both the plots in Figs. 5.17 and 5.19 show a central dip and peaks on either side of the jet due indicating shear interactions between the jet and surrounding gaseous phase in the near injector exit locations. The peak in the RMS axial velocity for cold spray with no swirl flow is observed at $X = 5$ mm for $Y = 5$–10 mm while for cold spray with swirl flow the RMS peaks at $X = 10$ mm for $Y = 5$–20 mm. Farther downstream for both the cases without and with swirl flow, the RMS velocities show more uniform distribution compared to the near injector plane locations. The RMS velocities trail off trend on the outside edge of the spray. The higher magnitude of RMS velocity in cold spray with swirl flow indicates improved inherent turbulent mixing due to the surrounding swirling air flow as compared to no swirl flow in Fig. 5.17. Figure 5.21 shows the RMS contour for cold spray with swirl for VO at 100 °C. Higher fuel inlet temperature shows a slight increase in RMS axial velocity compared to the unheated cold spray. The increase in the RMS axial velocity results from the heated fuel raising the temperature of the atomizing air to produce higher turbulence levels leading into smaller diameter droplets. The peak RMS value is about 25 m/s.

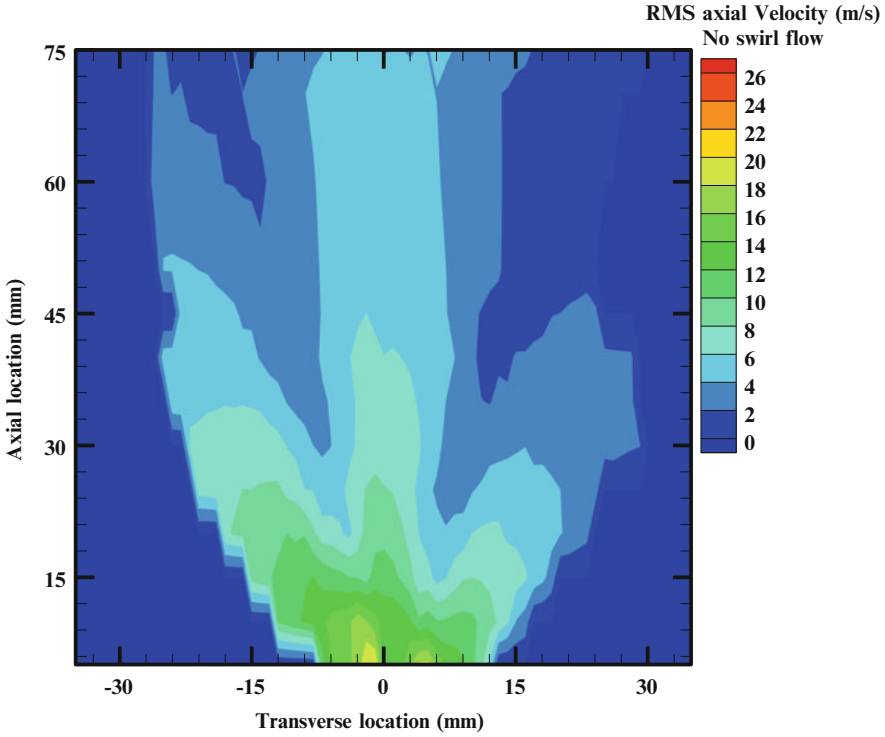



**Fig. 5.18** Axial RMS velocity contour for cold spray without swirl

SMD Contours

The effective method of expressing the quality of the atomization is the mean diameter. Probably, the most common form of mean droplet size is the SMD, which has the physical interpretation as the diameter of the drop having the same volume/surface area ratio as the entire spray. Figures 5.21 and 5.22 show the contour plots constructed from point-wise measurements acquired at different axial and radial locations.

Figures 5.21 and 5.22 show, respectively, the contour plots of SMD for open cold spray with and without co-flow swirling air. As seen from the plot the spray consists of the droplets with wide range of SMD values of 20–80 μm. The plot shows the distribution of the droplets in the various regions of the spray in the mid field ($Y = 20$–$40$ mm) and far field ($Y = 45$–$75$ mm) locations. The larger droplets of 55 μm and above diameter reside in the center core region near the injector exit ($Y = 5$–$15$ mm) as well as at the outer edge ($X = \pm 20$ mm) of the spray for $Y = 50$–$75$ mm. The presence of larger droplets in center of the spray can be attributed due to higher axial momentum and weaker mixing in the jet core region compared to the shear layer region of the spray, i.e., $X = \pm 5$–$15$ mm.

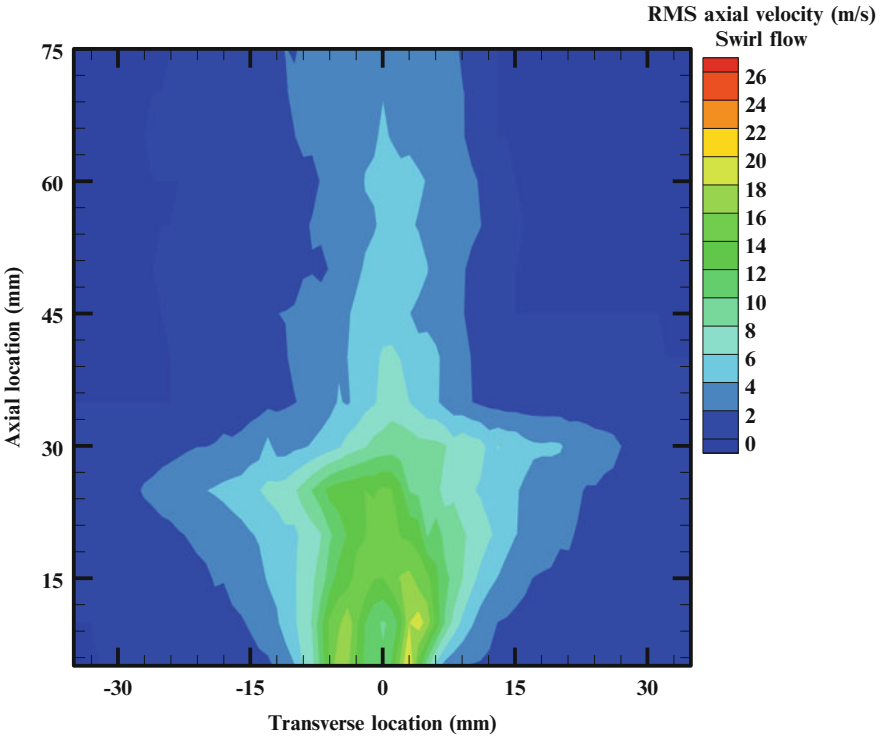



**Fig. 5.19** Axial RMS velocity contour for cold spray with swirl

Some larger droplets are observed towards the spray edge due to their higher momentum. At radial locations of $X = \pm 10$–25 mm and $Y = 30$–75 mm. As observed from the plot, the smaller and intermediate size droplets reside in the shear layer ($X = \pm 5$–15 mm) both near and far away from the injector exit. Droplets that escape from the center liquid core undergo further breakup which leads to formation of smaller droplets. However, farther downstream, more uniform and homogenous distribution of the droplet size is observed in the radial direction. For all the axial locations of the spray, the SMD values were observed to range from 20 to 45 μm.

Figure 5.22 shows the SMD contour for cold spray with swirling air flow and unheated VO. The SMD ranges from 20 to 60 μm. The radial spread in this case is seen to be about 30 mm as compared to 15 mm in case of no swirling air as shown in Fig. 5.19. As observed from the plot the larger droplets are located near the jet center, while the droplets diameter decreases away from the jet center. Droplet diameter in the near field location of the spray, i.e., $Y = 5$–15 mm, ranges from 50 to 60 μm. For downstream spray locations the droplet diameter gradually decreases and farther downstream of the spray at $Y = 60$–75 mm the SMD decreases to about 30–35 μm. The droplet diameter in the shear layer, i.e., $X = 10$–20 mm, is in the range of 20–35 μm. It can be seen from Figs. 5.21 and 5.22; that the SMD is smaller

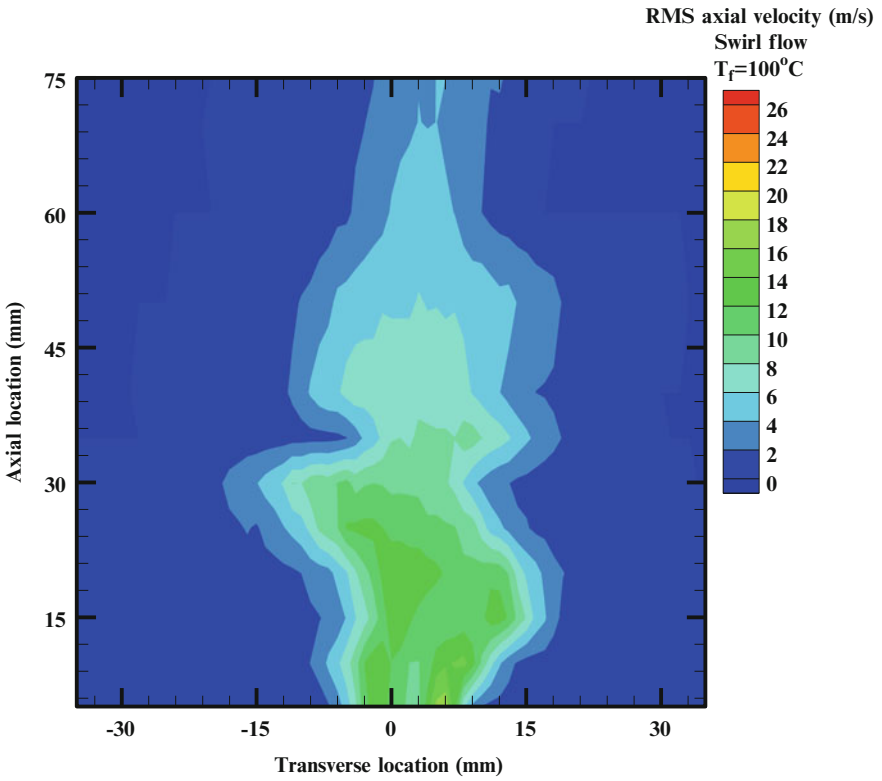



**Fig. 5.20** Axial RMS velocity contour for cold spray with swirl at $T_f = 100$ °C

for spray with swirling air compared to spray with no swirling air. The swirling air promotes secondary droplet breakup mechanism resulting into smaller SMD in the spray. Hence, the droplets escaping from the central jet core undergo further disintegration into smaller droplets ranging from 20 to 35 μm, due to the co-flowing swirling air. This process takes place for $X = 15$–45 mm for almost all the axial locations in the spray. The jet spreads wider in case of swirling air compared to without swirling air. Thus swirling air flow widens the spray and it also decreases the droplet diameter because of the secondary disintegration process.

Figure 5.23 shows the SMD contour plot for cold spray with swirling air for VO preheated to 100 °C. Results show similar general trends as discussed above for Figs. 5.21 and 5.22. Larger droplets in the central core region and smaller droplets in the shear layer region near and far field locations in the spray. The maximum SMD for this case however is around 40 μm which is significantly lower than the maximum SMD of 60 μm for the unheated VO cold spray. The difference in the results is attributed to the improved atomization due to increase in VO temperature. Increase in VO temperature improves the thermo physical properties of the VO, e.g., reducing the kinematic viscosity and surface tension

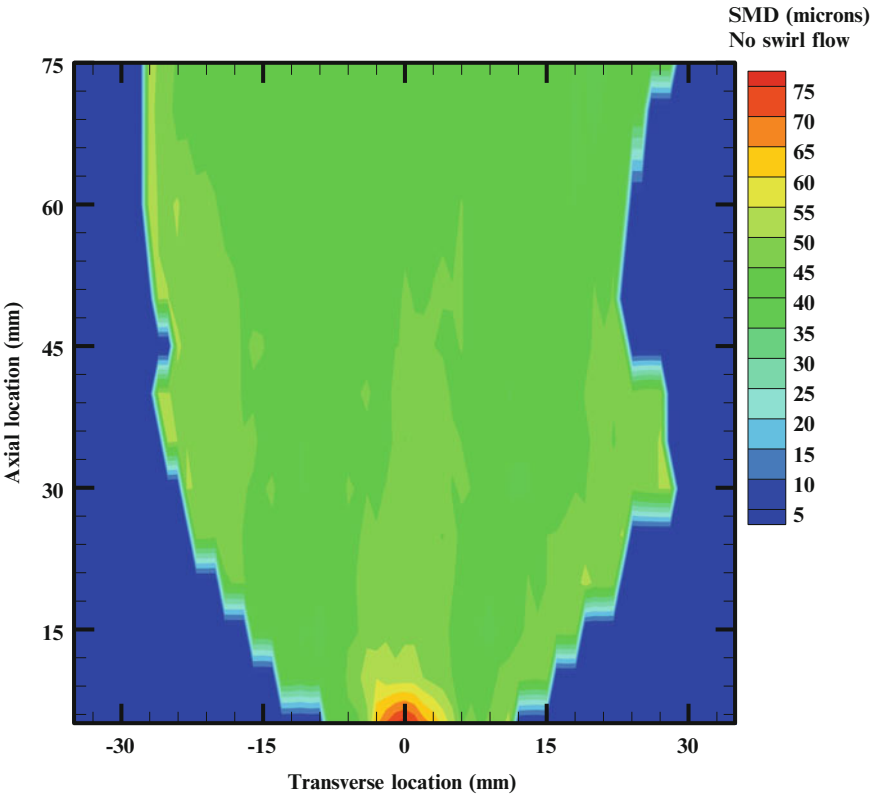



**Fig. 5.21** SMD contour for cold spray without swirling air

which makes it easier to break up the droplets producing finer spray. Increasing the VO temperature however requires thermal energy of about 0.53 % of the VO energy content (LHV = 37,000 kJ/kg) to achieve the benefits of improved spray characteristics.

### Transverse Profiles of Mean and RMS Axial Velocity

Figure 5.24a–z shows transverse profiles of mean axial velocity for cold spray with and without swirling air and unheated VO. The trend showing the effect of swirling air on cold spray is explained by Fig 5.24a at one axial location and for the rest of the axial locations the same trends follow. Figure 5.24a shows a dome shaped velocity profile for $Y = 5$ mm, for cold spray with and without swirl air flow. The mean axial velocity is highest at the center and it decreases towards the periphery of the spray. The peak axial velocity is similar for both the cases. But the general trend also observed in other profiles of Fig. 5.24 is that in the near field region, the cold

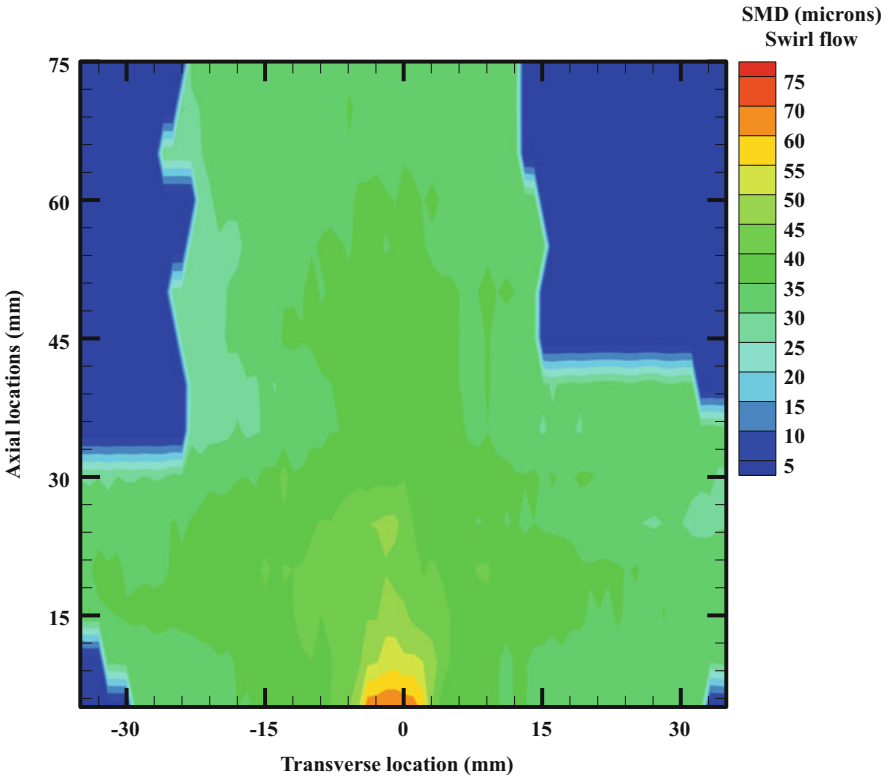



**Fig. 5.22** SMD contour for cold spray with swirling air

spray with swirling air flow shows marginally higher axial velocity, which decreases at farther downstream locations. The peak axial velocity near the injector exit region of the spray with swirling air is about 5–8 m/s higher compared to non-swirling air spray. From Fig. 5.24a–d, we can observe a well-pronounced recirculation zone in case of swirling air flow, characterized by negative velocities on either side of the spray center. The recirculation flow improves mixing of the fuel and air to further improve atomization. Swirling air flow is also the cause of the further disintegration the larger droplets. The swirling air flow tends to increase the radial spread of the spray compared to that for the non-swirling air flow. The recirculation zone can be identified from the dotted line in the plot and it can be seen at $X = 10$–$40$ mm for $Y = 5$–$20$ mm.

Further downstream, the peak axial velocity decreases for each case with non-swirl air flow resulting in higher velocities compared to the swirl air flow case. With increased axial distance, ($Y = 30$–$75$ mm), peak axial velocities show an increase for the non-swirl flow compared to the swirl flow. This effect is attributed to the intense mixing and radial spread of the spray with swirling airflow. While the non-swirling air spray maintains the high axial momentum because of poor mixing with the ambient air.

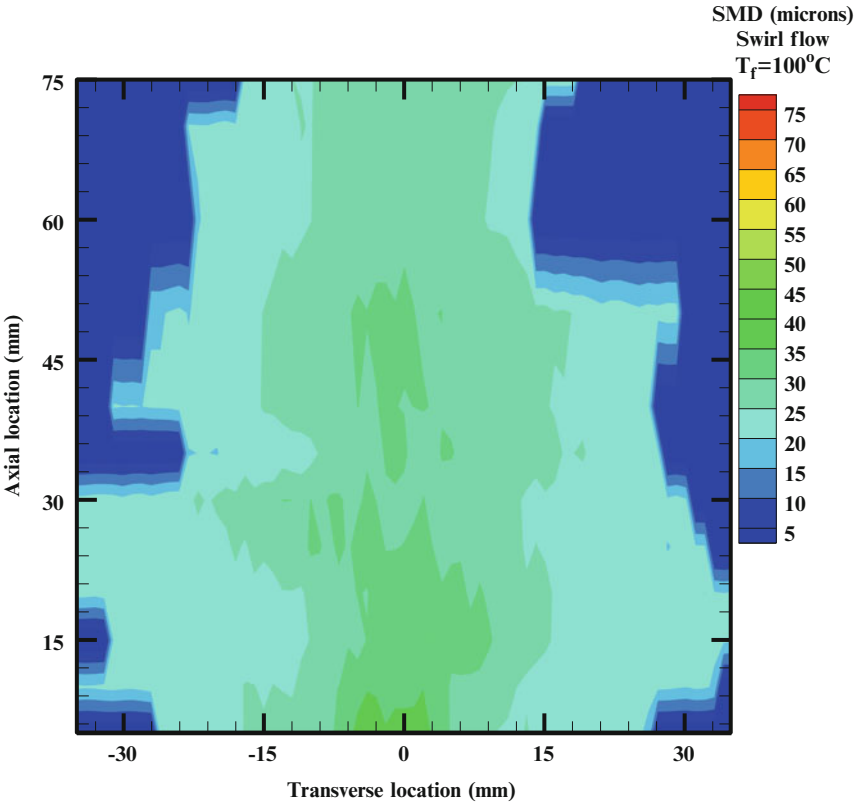



**Fig. 5.23** SMD contour for cold spray with swirling air at $T_f = 100$ °C

Figure 5.25 shows the effect of fuel temperature on cold spray. The Fig. 5.25a shows the axial velocity at $Y = 5$ mm, which shows a slight increase in peak axial velocity for VO at 100 °C mostly in the region near the injector exit in the spray. The peak velocities of about 45, 34, and 33 m/s are observed at respectively $Y = 5$, 10, and 15 mm in the near field region of the spray. A similar trend of mean axial velocity profile is observed for both the cases at all axial locations in the spray. Only a slight increase in mean axial velocity with heated VO is observed in the near and far field regions of the spray. The effect of fuel inlet temperature is more pronounced to reduce the droplet diameter which can be seen from the SMD profile plots.

For all of the aforementioned test conditions, RMS axial velocities are shown in Fig. 5.24a–z. The ratio of RMS over the mean axial velocity represents the turbulence intensity. Figure 5.24b shows the RMS axial velocity at $Y = 5$ mm for cold spray with and without swirl. In general the profiles for both the cases show a dip in the center of the spray, which is also observed at $Y = 10$, 15, and 20 mm. This result is attributed to the intense turbulent fluctuations in the shear region of the spray where the respective gradients reach the highest values to dominate the

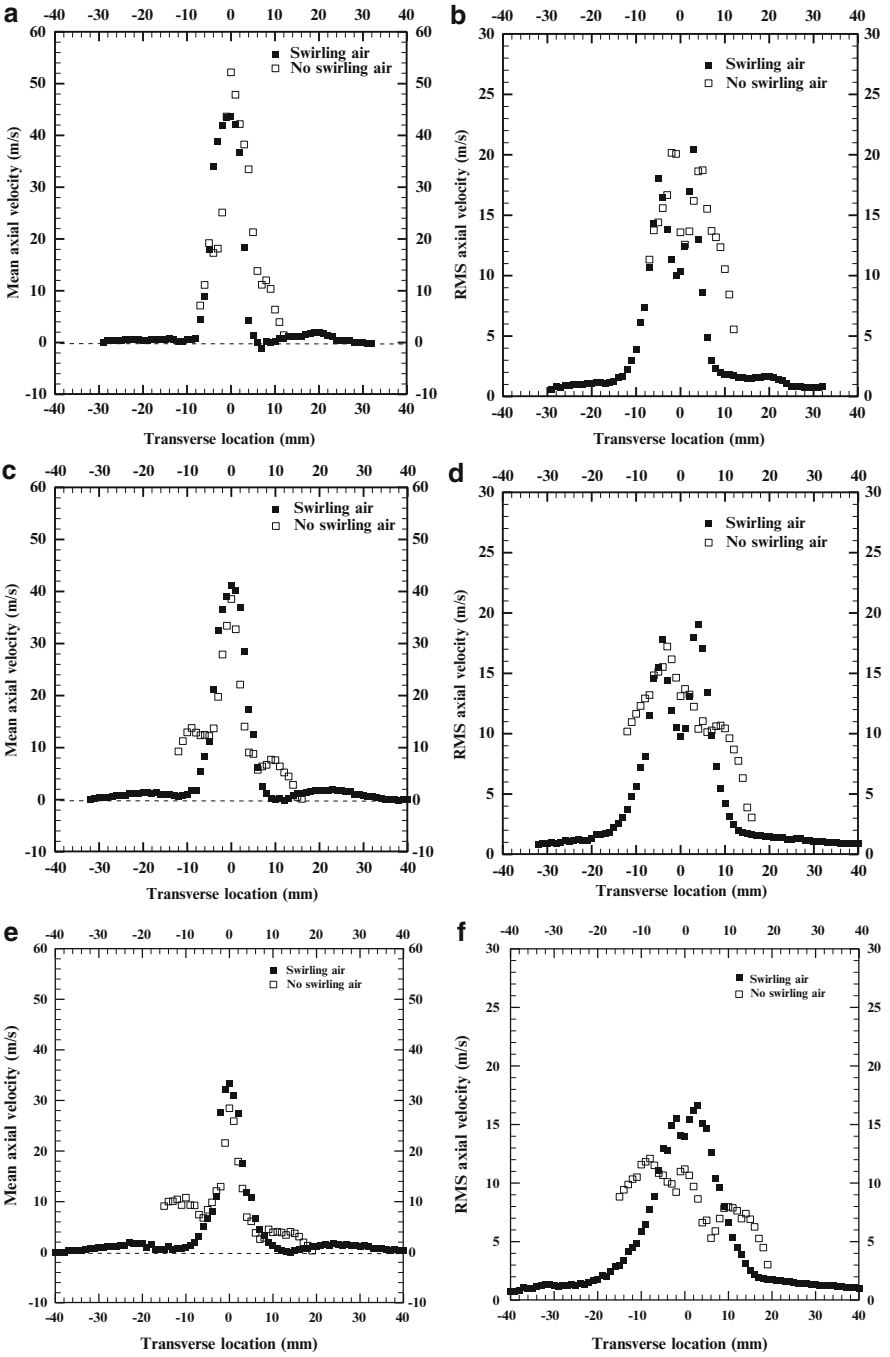



**Fig. 5.24** (continued)

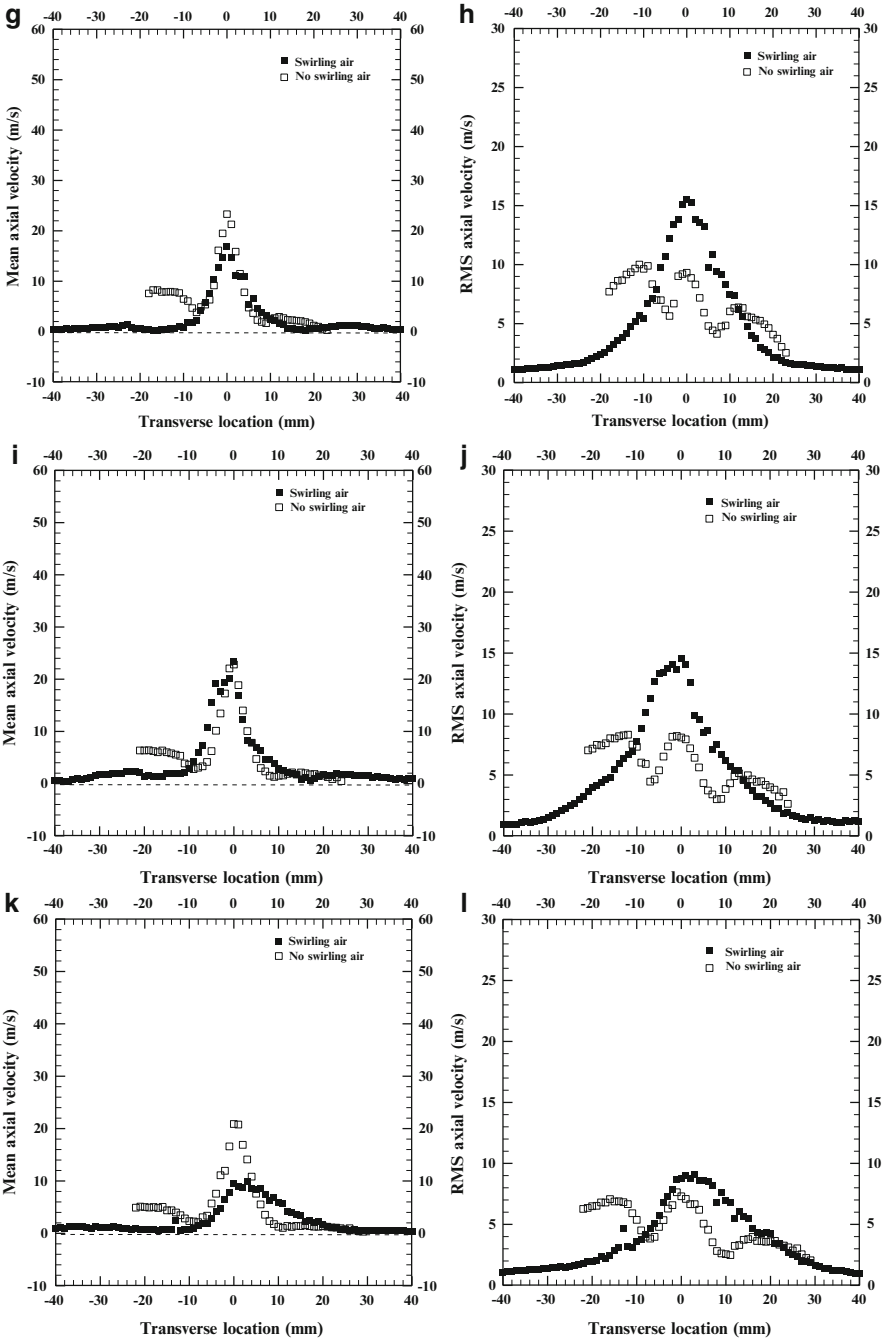



**Fig. 5.24** (continued)

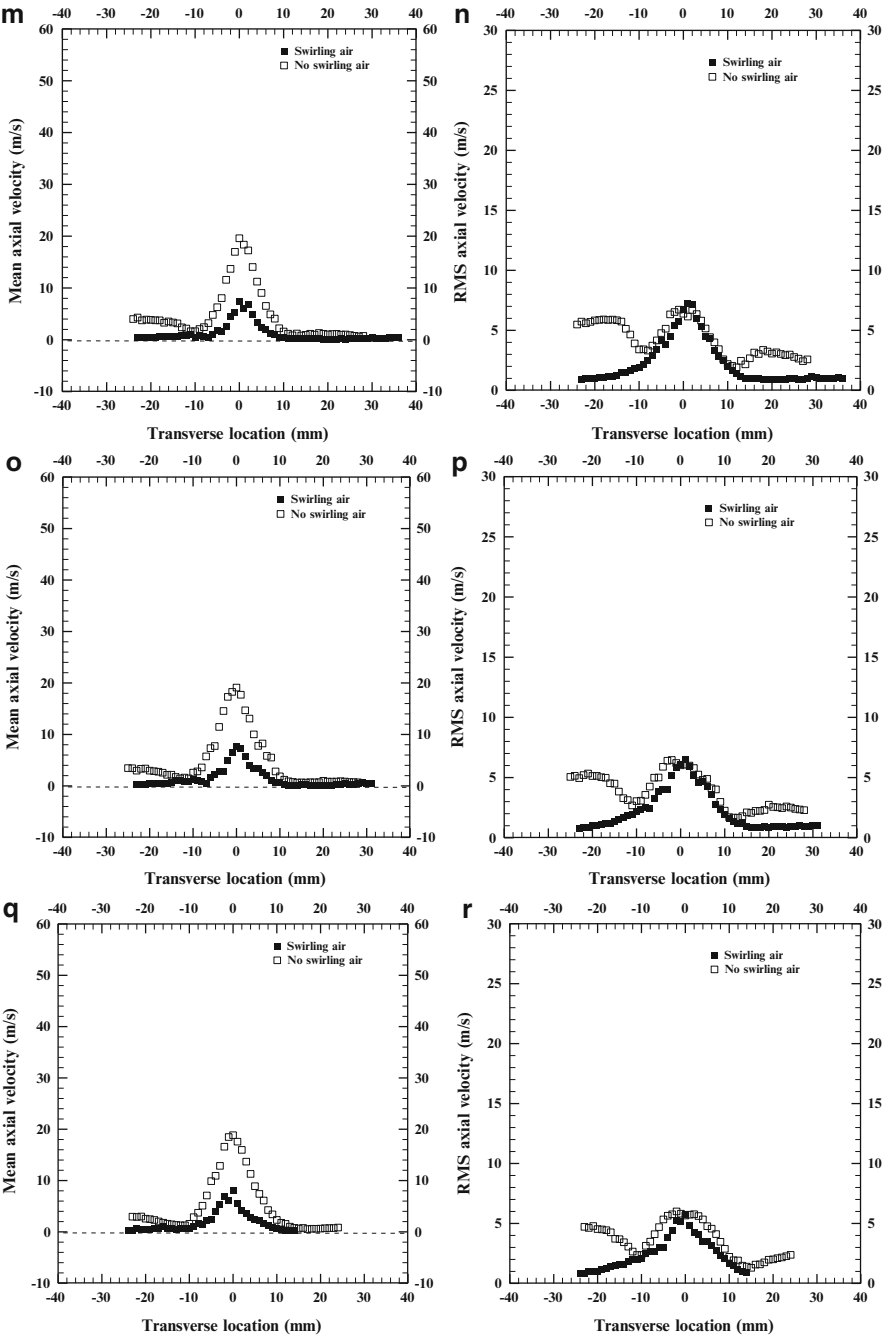



**Fig. 5.24** (continued)

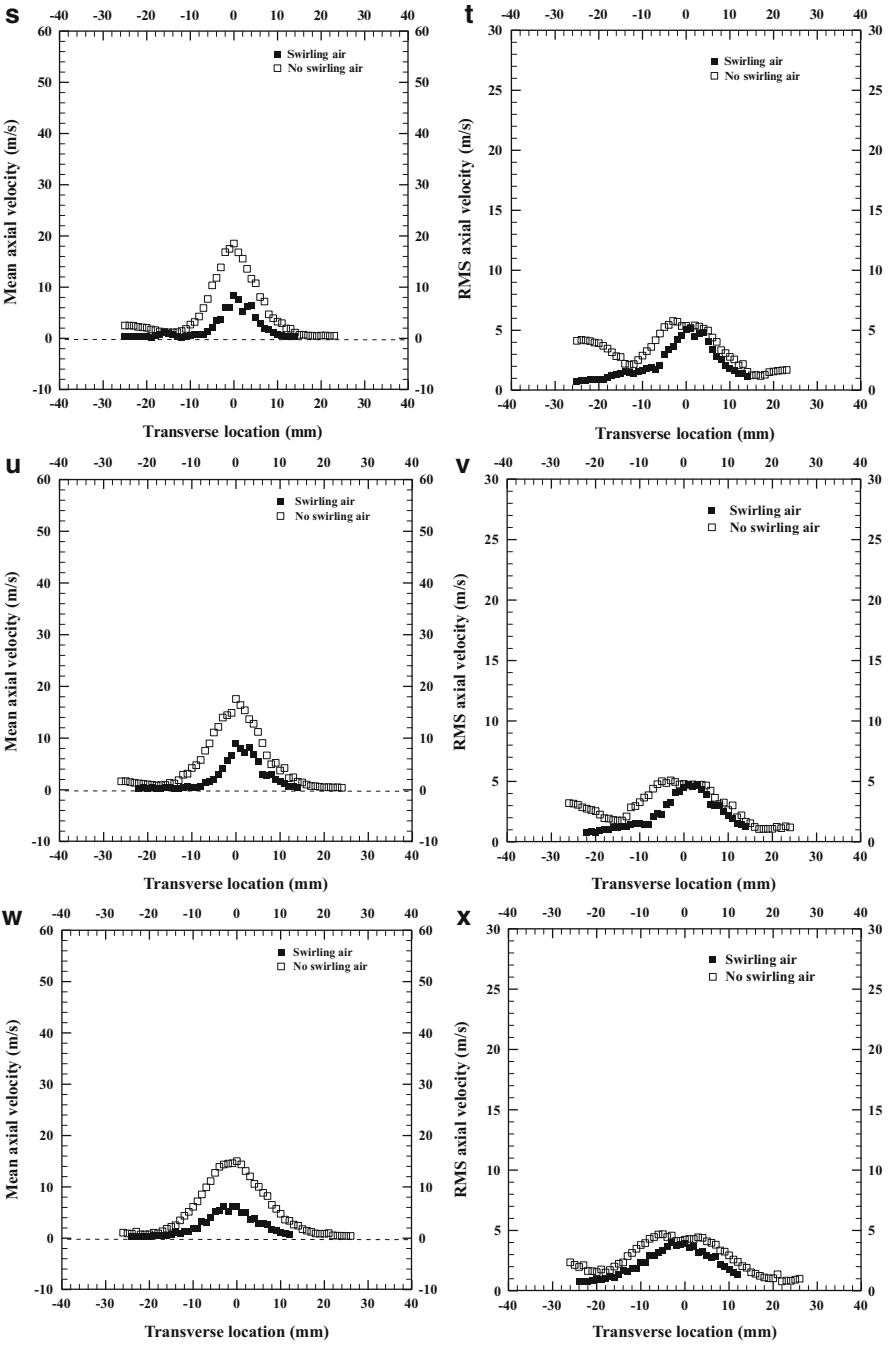



**Fig. 5.24** (continued)

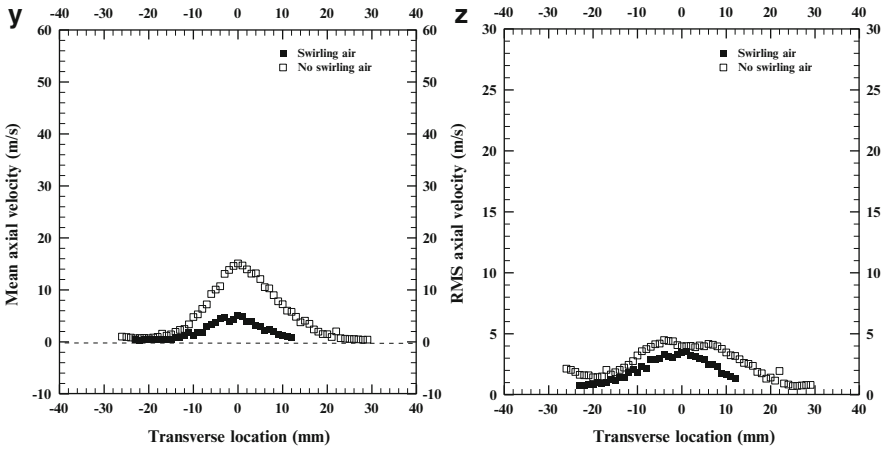



**Fig. 5.24** (**a** and **b**) Transverse profiles of mean axial velocity and RMS axial velocity for cold spray with and without swirl at $Y = 5$ mm. (**c** and **d**) Transverse profiles of mean axial velocity and RMS axial velocity for cold spray with and without swirl at $Y = 10$ mm. (**e** and **f**) Transverse profiles of mean axial velocity and RMS axial velocity for cold spray with and without swirl at $Y = 15$ mm. (**g** and **h**) Transverse profiles of mean axial velocity and RMS axial velocity for cold spray with and without swirl at $Y = 20$ mm. (**i** and **j**) Transverse profiles of mean axial velocity and RMS axial velocity for cold spray with and without swirl at $Y = 25$ mm. (**k** and **l**) Transverse profiles of mean axial velocity and RMS axial velocity for cold spray with and without swirl at $Y = 30$ mm. (**m** and **n**) Transverse profiles of mean axial velocity and RMS axial velocity for cold spray with and without swirl at $Y = 35$ mm. (**o** and **p**) Transverse profiles of mean axial velocity and RMS axial velocity for cold spray with and without swirl at $Y = 40$ mm. (**q** and **r**) Transverse profiles of mean axial velocity and RMS axial velocity for cold spray with and without swirl at $Y = 45$ mm. (**s** and **t**) Transverse profiles of mean axial velocity and RMS axial velocity for cold spray with and without swirl at $Y = 50$ mm. (**u** and **v**) Transverse profiles of mean axial velocity and RMS axial velocity for cold spray with and without swirl at $Y = 60$ mm. (**w** and **x**) Transverse profiles of mean axial velocity and RMS axial velocity for cold spray with and without swirl at $Y = 70$ mm. (**y** and **z**) Transverse profiles of mean axial velocity and RMS axial velocity for cold spray with and without swirl at $Y = 75$ mm

turbulence generation rate. Further downstream the turbulence level begins to decay owing to the diffusion and reduction in kinetic energy of the droplets. In particular, turbulent kinetic energy attenuation is significant in the shear layer because of the additional dissipation introduced by the presence of the droplets. As seen in the Fig. 5.24a, the magnitude of the RMS axial velocity values for the swirl air flow as well as non-swirl flow is almost overlapping. This result is consistent with the higher axial mean velocities as discussed previously in the near field regions of the spray. A local maximum is seen at radial location $X = 5$ mm, indicating that the faster and smaller drops have higher RMS axial velocity values. For $Y = 30$–$75$ mm, the dip in the RMS axial velocity is no longer observed and the profiles are more uniform indicating shear layer merging with the centerline at these downstream locations in

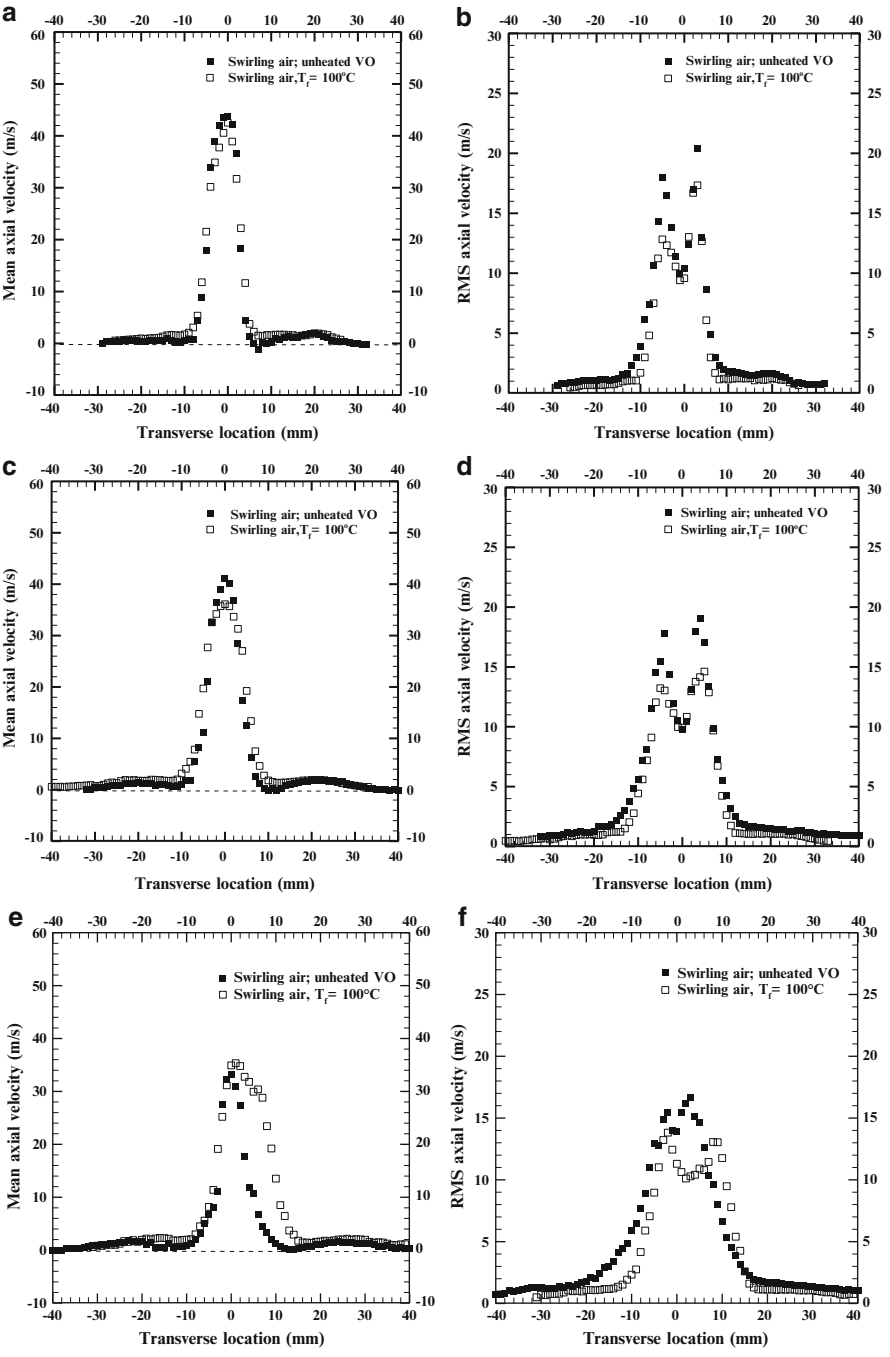



**Fig. 5.25** (continued)

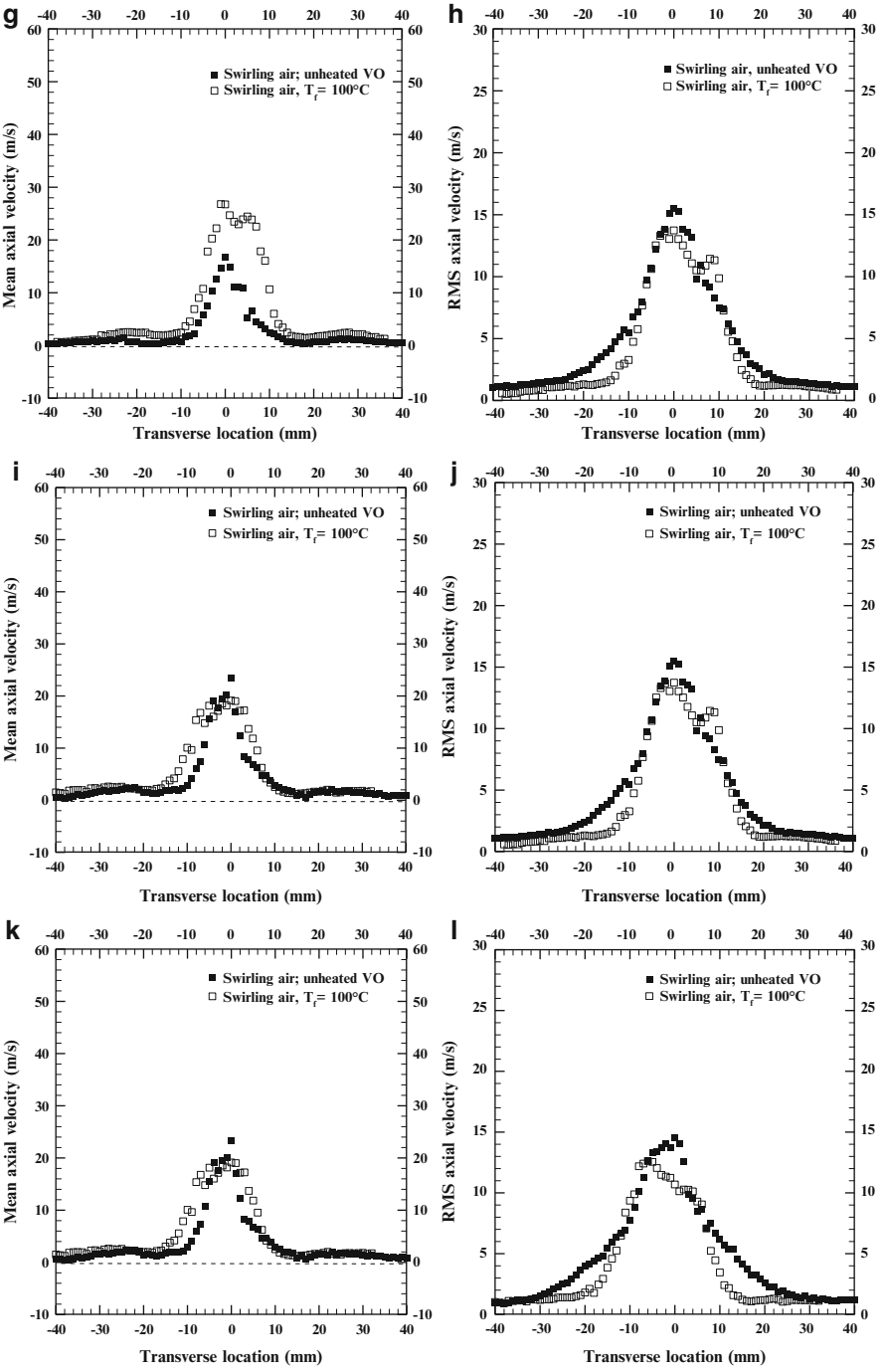



**Fig. 5.25** (continued)

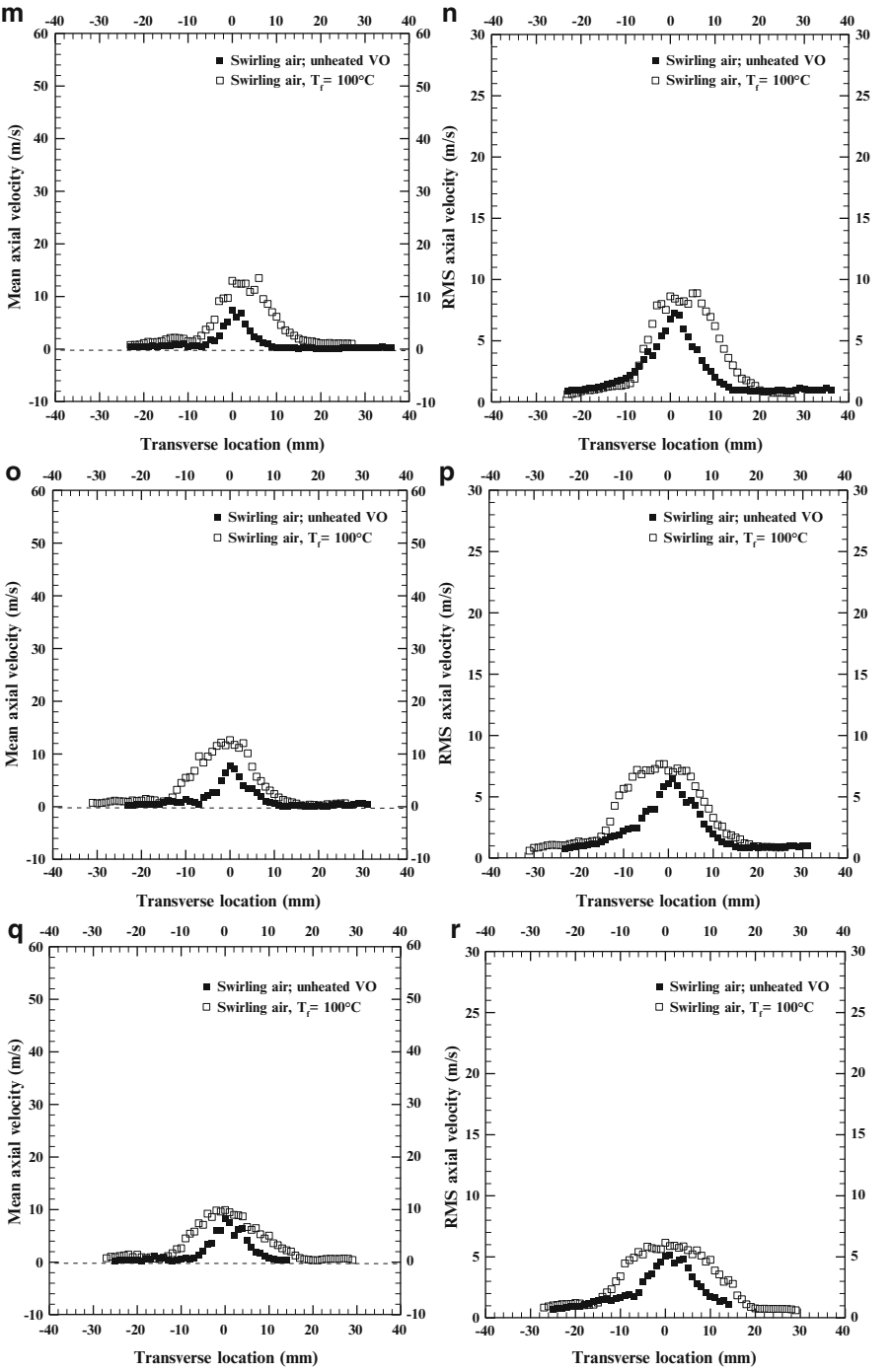



**Fig. 5.25** (continued)

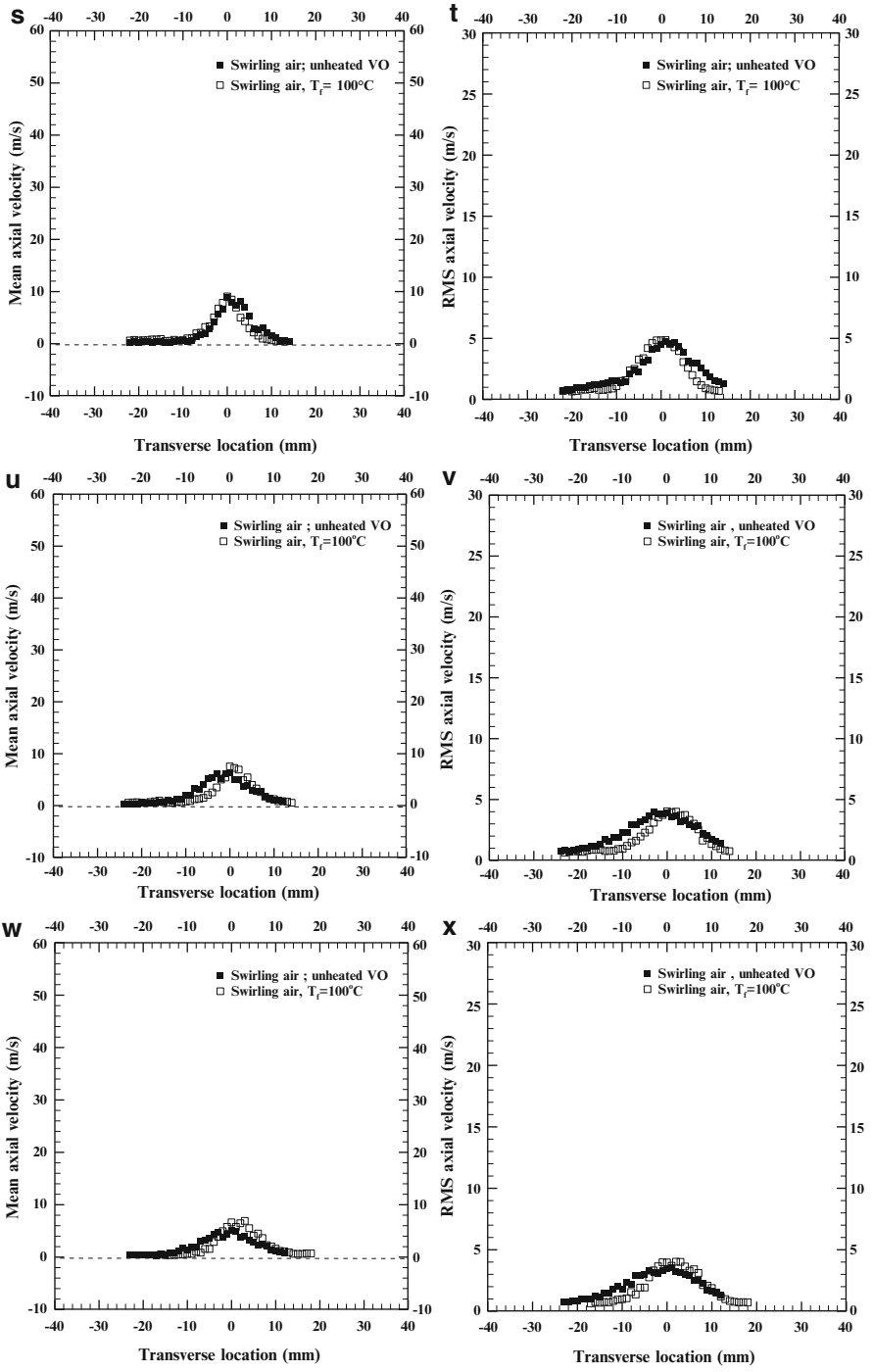



**Fig. 5.25** (continued)

the spray. The RMS axial velocity values for $Y = 30, 35, 40$ mm are about 10, 8, and 7 m/s for both the cases. Further downstream at $Y = 60$, 70 and 75 mm, the RMS axial velocity of about 3–5 m/s is similar for swirl air flow and non-swirl air flow, respectively.

### Transverse Profiles of SMD

Figure 5.26 shows the radial profiles of SMD at different axial locations in the spray to demonstrate the effect of swirl air flow on the cold spray. Figure 5.26a shows the SMD profile for cold spray with and without swirl air flow at of $Y = 5$ mm. In general, the SMD values are seen to be higher at the centerline with a diminishing trend of smaller droplets towards the edge of the spray. For the near field location in the spray, i.e., $Y = 5$ mm, SMD values at the center is around 75 μm. Moving away from the center, the droplet diameter decreases. A slight difference is observed in the SMD values for cold spray with and without swirling air, the latter showing smaller droplets. The profile for cold spray with swirl is tapered radially at $X = 20$ mm, since the swirling co-flow air tends to spread the spray and further disintegrate the droplets to form smaller drops. The radial extent of the spray is wider for swirl air flow case compared to the without swirl case, i.e., 55 mm for swirl case and 30 mm for non-swirl spray. SMD values at $Y = 5$ mm range from 25 to 72 μm for cold spray with swirl air flow, while that of cold flow without swirl air flow ranges from 40 to 80 μm. Similar trend is observed for other near field locations in the spray, i.e., at $Y = 10$, 15, and 20 mm. The maximum SMD is nearly the same at $Y = 10$, 15, and 20 mm with and without swirl air flow. However, the magnitude of the SMD decreases in the axial direction.

**Fig. 5.25** (**a** and **b**) Transverse profiles of mean axial velocity and RMS axial velocity for swirling air cold spray; unheated and VO at $T_f = 100$ °C, $Y = 5$ mm. (**c** and **d**) Transverse profiles of mean axial velocity and RMS axial velocity for swirling air cold spray; unheated and VO at $T_f = 100$ °C, $Y = 10$ mm. (**e** and **f**) Transverse profiles of mean axial velocity and RMS axial velocity for swirling air cold spray; unheated and VO at $T_f = 100$ °C, $Y = 15$ mm. (**g** and **h**) Transverse profiles of mean axial velocity and RMS axial velocity for swirling air cold spray; unheated and VO at $T_f = 100$ °C, $Y = 20$ mm. (**i** and **j**) Transverse profiles of mean axial velocity and RMS axial velocity for swirling air cold spray; unheated and VO at $T_f = 100$ °C, $Y = 25$ mm. (**k** and **l**) Transverse profiles of mean axial velocity and RMS axial velocity for swirling air cold spray; unheated and VO at $T_f = 100$ °C, $Y = 30$ mm. (**m** and **n**) Transverse profiles of mean axial velocity and RMS axial velocity for swirling air cold spray; unheated and VO at $T_f = 100$ °C, $Y = 35$ mm. (**o** and **p**) Transverse profiles of mean axial velocity and RMS axial velocity for swirling air cold spray; unheated and VO at $T_f = 100$ °C, $Y = 40$ mm. (**q** and **r**) Transverse profiles of mean axial velocity and RMS axial velocity for swirling air cold spray; unheated and VO at $T_f = 100$ °C, $Y = 50$ mm. (**s** and **t**) Transverse profiles of mean axial velocity and RMS axial velocity for swirling air cold spray; unheated and VO at $T_f = 100$ °C, $Y = 60$ mm. (**u** and **v**) Transverse profiles of mean axial velocity and RMS axial velocity for swirling air cold spray; unheated and VO at $T_f = 100$ °C, $Y = 70$ mm. (**w** and **x**) Transverse profiles of mean axial velocity and RMS axial velocity for swirling air cold spray; unheated and VO at $T_f = 100$ °C, $Y = 75$ mm

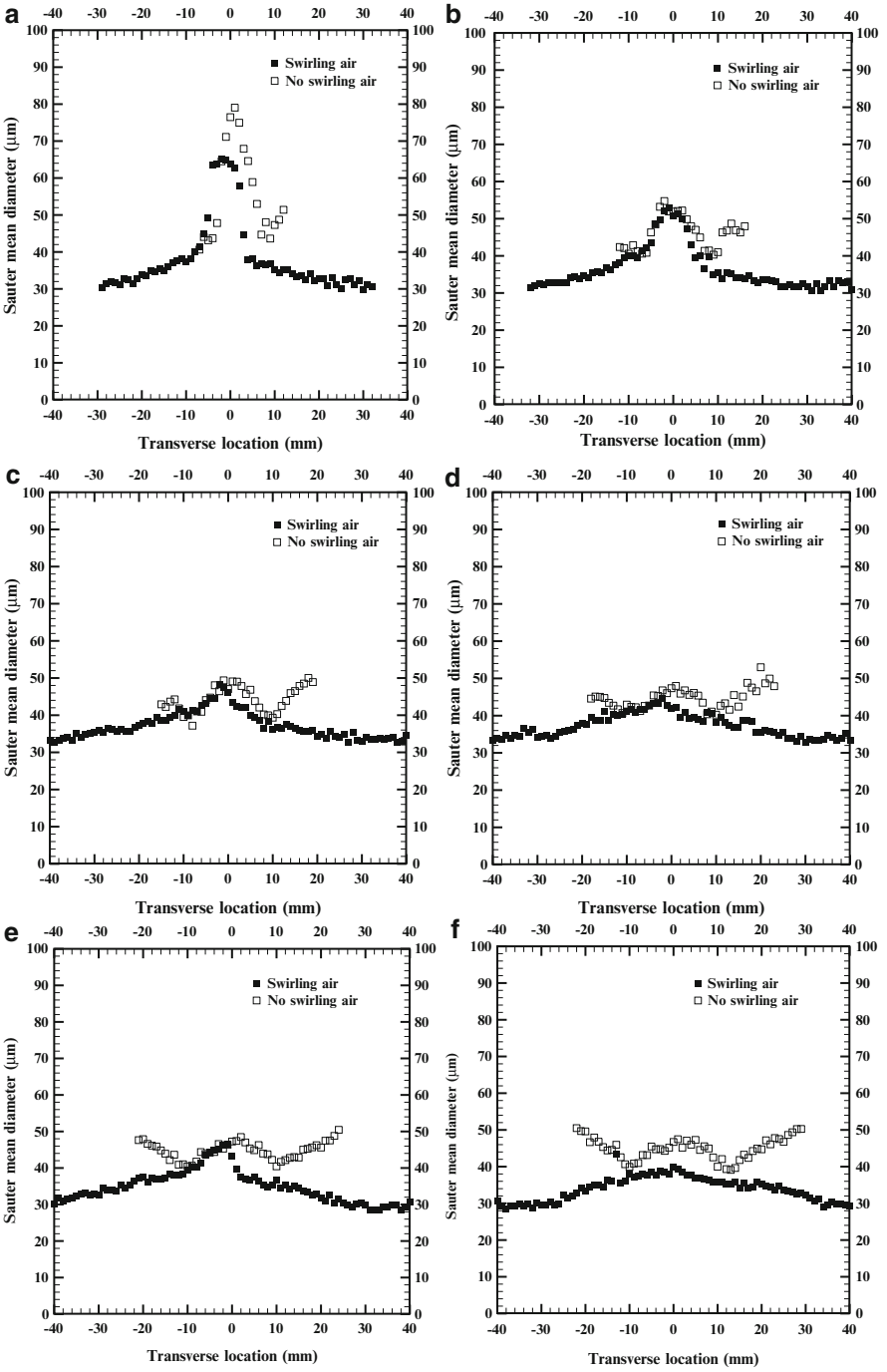



**Fig. 5.26** (continued)

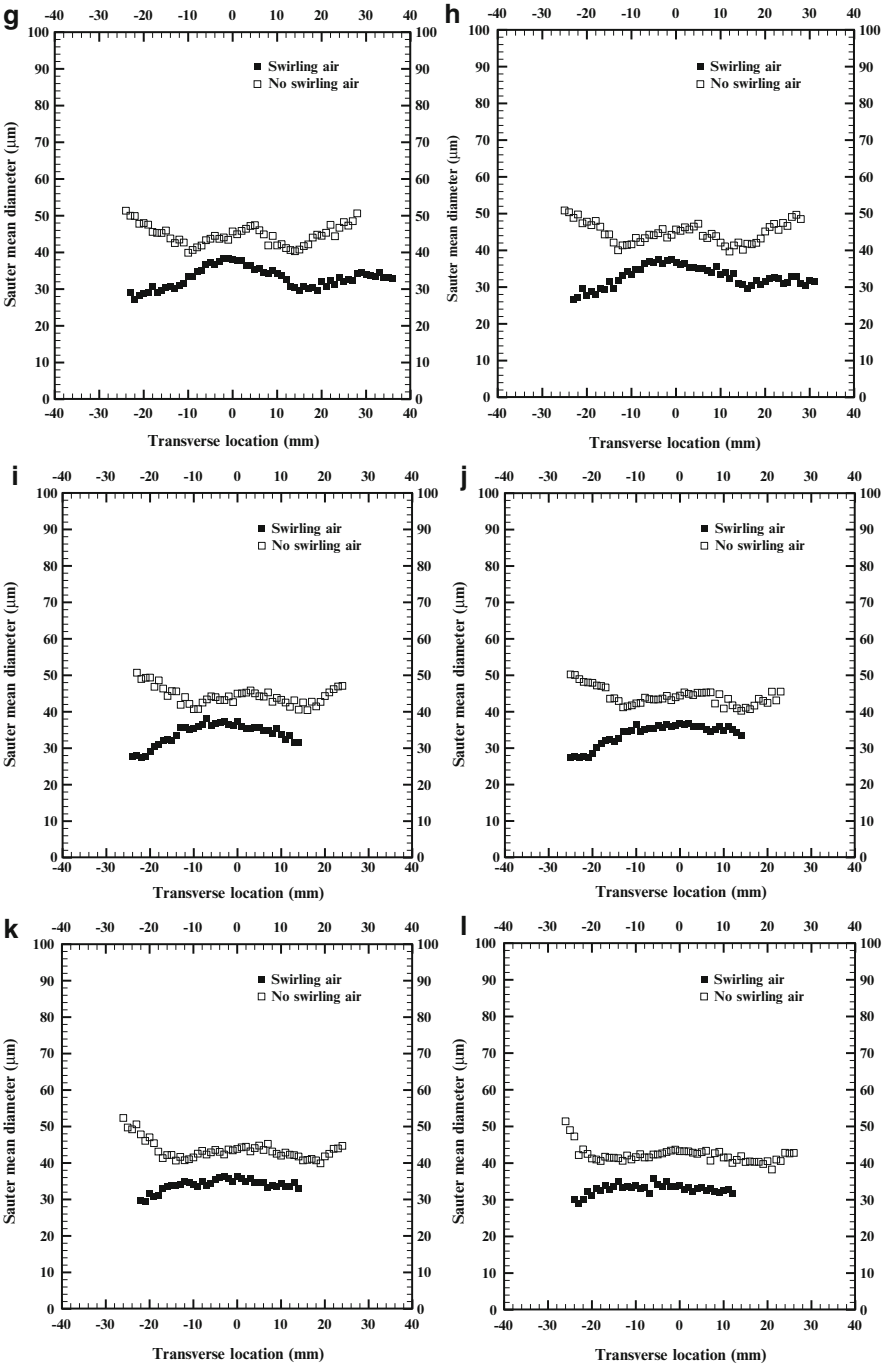



**Fig. 5.26** (continued)

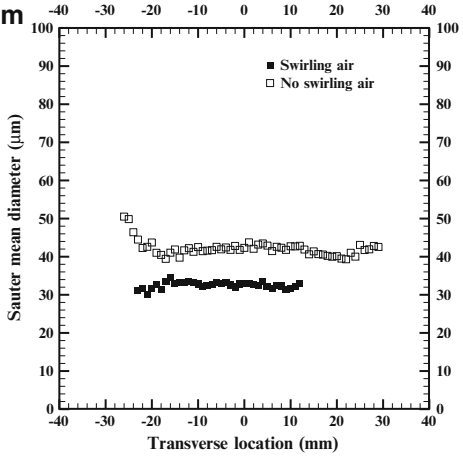



**Fig. 5.26** (**a** and **b**) Transverse profiles of SMD for cold spray with and without swirl at $Y = 5$ and 10 mm. (**c** and **d**) Transverse profiles of SMD for cold spray with and without swirl at $Y = 15$ and 20 mm. (**e** and **f**) Transverse profiles of SMD for cold spray with and without swirl at $Y = 25$ and 30 mm. (**g** and **h**) Transverse profiles of SMD for cold spray with and without swirl at $Y = 35$ and 40 mm. (**i** and **j**) Transverse profiles of SMD for cold spray with and without swirl at $Y = 45$ and 50 mm. (**k** and **l**) Transverse profiles of SMD for cold spray with and without swirl at $Y = 60$ and 70 mm. (**m**) Transverse profiles of SMD for cold spray with and without swirl at $Y = 75$ mm

The noticeable difference in the droplet diameter is observed from $Y = 30$–75 mm in both the cases. The co-flow swirling spray decreases the SMD by about 25 μm as compared to the non-swirling spray. For co-flow swirling spray there is a decrease of about 25 μm in SMD for $Y = 30$–75 mm. In the same region the SMD for the non-swirling spray decreases from 57 to 45 μm. Thus the swirl air flow in the cold spray decreases the droplet diameter which would improve atomization and combustion.

Figure 5.27 shows the effect of fuel inlet temperature on radial profiles for SMD in the spray. With the preheated VO at 100 °C, the maximum SMD decreases to 40 μm from 70 μm at $Y = 5$ mm. The range of SMD at this axial location is about 17 μm as compared to the wider range of 38 μm in case of unheated VO spray. The radial extent with the smaller droplets is greater in case of preheated VO spray compared to the unheated VO spray. Increase in fuel inlet temperature produces a spray with droplet diameter ranging from 18 to 40 μm; whereas the droplet diameter range in the unheated fuel spray is from 26 to 70 μm. Evidently the effect of fuel inlet temperature is seen at almost all axial locations in the spray ranging from near field to far field locations .The higher fuel inlet temperature decreases the fuel kinematic viscosity and surface tension, which aids in improving the atomization characteristics to produce finer droplets. Increased fuel inlet temperature decreases the SMD and its range as well as the radial spread of the spray as seen from the mean and RMS axial velocity profiles.

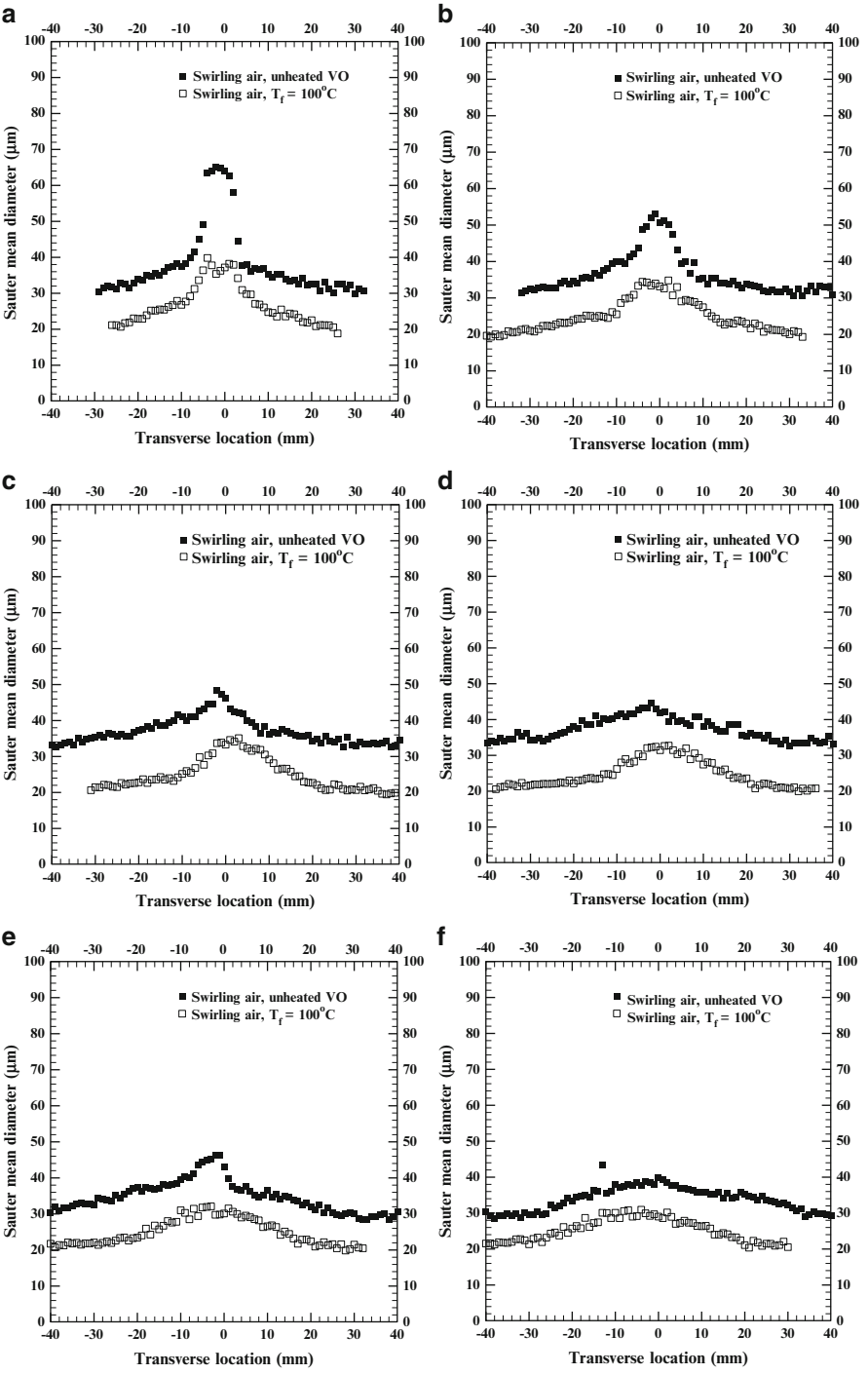



**Fig. 5.27** (continued)

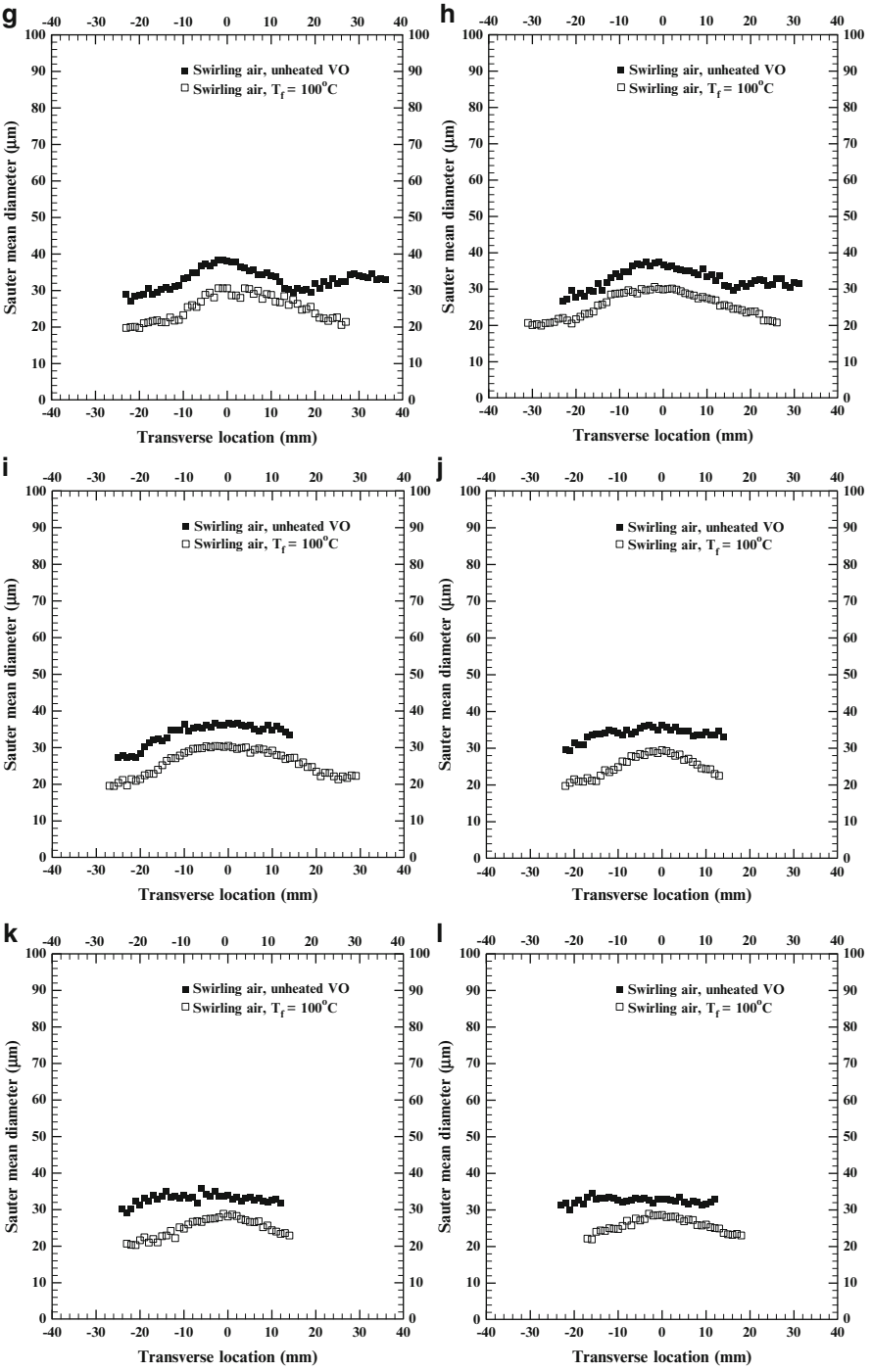



**Fig. 5.27** (continued)

Droplet Diameter Distribution Profiles

To better understand the spray structure, Fig. 5.28 compares the droplet size distribution profiles at the center of the injector exit. Figure 5.28a shows the droplet size distribution profiles at the $Y = 5$ mm and $X = 0$ mm. The SMD is 64 and 76 μm, respectively, for cold spray with swirling air and without swirling air. As seen from the plot, cold flow with swirling air shows a higher percentage of smaller diameter droplets in the range of < 50 μm as compared to the non-swirling spray. The largest droplet diameter is about 200 μm for both swirling and non-swirling spray, although very few larger diameter droplets are observed in case of the swirling spray. For swirling spray, the SMD distribution shows most of the droplets with diameter of about 100 μm or less while for non-swirling spray the distribution spread out with a higher percentage of larger diameter droplets. The greater number of smaller diameter resulted in the smaller SMD in case of swirling spray as compared to non-swirling spray.

Figure 5.28b shows the distribution profiles at $Y = 40$ mm, $X = 0$ mm which is considered here as the mid field location in the spray. Close to 90 % of the droplets in case of swirling spray are smaller than 100 μm. The largest diameter in non-swirling spray is about 150 μm. A significant difference in the distribution count is seen for droplets of diameter <35 μm, with swirling flow showing higher percentage compared to the non-swirling flow. The presence of these larger droplets in case of non-swirling spray is the cause of higher SMD of about 46 μm compared to that of 36 μm in case of swirling spray. Comparatively, the near field region of the spray shows greater number of larger droplets than at the axial locations away from the injector exit.

Figure 5.28c shows the distribution profile for the swirling and non-swirling spray at $Y = 60$ mm, $X = 0$ mm, i.e. at a far field location in the spray. The profiles for both the cases are very similar, still showing greater number of larger droplets in case of non-swirling spray. The highest droplet diameter in case of non-swirling spray is 150 μm as compared to that of 80 μm for swirling flow. The presence of these larger droplets dominates the SMD in case of non-swirling spray showing higher overall SMDs compared to the cold spray with swirling air flow. The SMDs

**Fig. 5.27** (**a** and **b**) Transverse profiles of SMD for swirling air cold spray; unheated and VO at $T_f = 100$ °C, $Y = 5$ mm, and $Y = 10$ mm. (**c** and **d**) Transverse profiles of SMD for swirling air cold spray; unheated and VO at $T_f = 100$ °C, $Y = 15$ mm, and $Y = 20$ mm. (**e** and **f**) Transverse profiles of SMD for swirling air cold spray; unheated and VO at $T_f = 100$ °C, $Y = 25$ mm, and $Y = 30$ mm. (**g** and **h**) Transverse profiles of SMD for swirling air cold spray; unheated and VO at $T_f = 100$ °C, $Y = 35$ mm, and $Y = 40$ mm. (**i** and **j**) Transverse profiles of SMD for swirling air cold spray; unheated and VO at $T_f = 100$ °C, $Y = 50$ mm, and $Y = 60$ mm. (**k** and **l**) Transverse profiles of SMD for swirling air cold spray; unheated and VO at $T_f = 100$ °C, $Y = 70$ mm, and $Y = 75$ mm

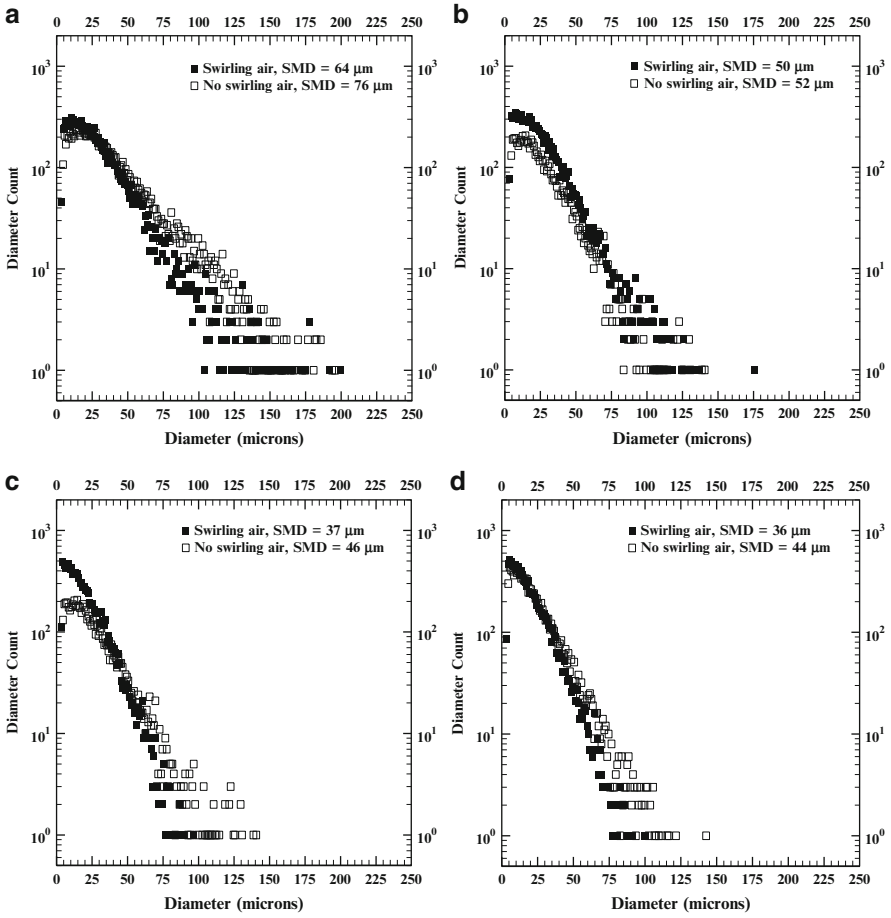



**Fig. 5.28** (**a** and **b**) Droplet distribution profile for cold spray with and without swirl at $Y = 5$ mm, $X = 0$ mm and $Y = 10$ mm, $X = 0$ mm. (**c** and **d**) Droplet distribution profile for cold spray with and without swirl at $Y = 40$ mm, $X = 0$ mm and $Y = 60$ mm, $X = 0$ mm

are 36 and 44 μm for swirling spray and non-swirling spray, respectively. As explained earlier, the droplet distribution is narrower with greater number of smaller drops at far field regions of the spray than that observed at the near field locations, which is consistent with the higher SMD near the injector exit at the center of the spray and smaller SMD farther downstream. Close to the injector exit the spray is dense and due to the presence of various droplet sizes the distribution becomes rather wider displaying discrete maximum than that father downstream center of the spray.

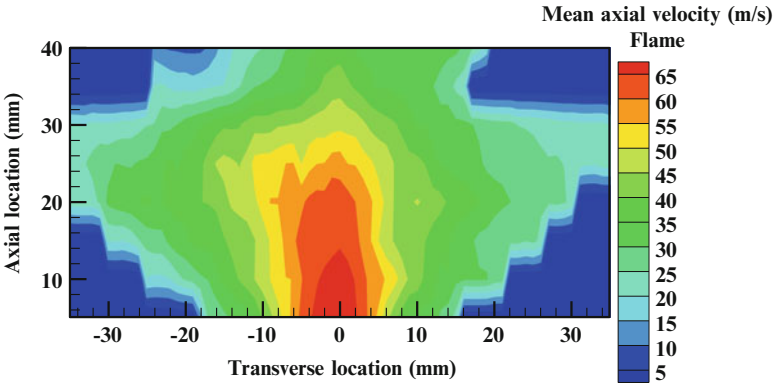



**Fig. 5.29** Axial velocity contour for flame spray

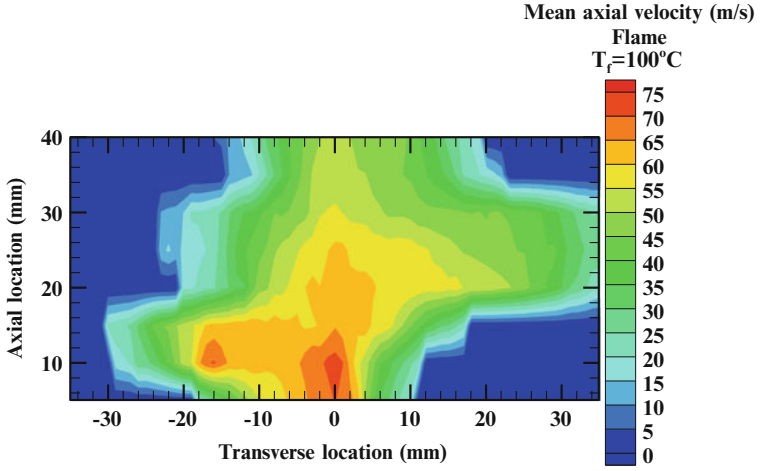



**Fig. 5.30** Axial velocity contour for flame spray for VO at 100 °C

#### 5.1.10.2 Effect of Flame on Open Spray

Velocity Contours (Mean and RMS)

Figures 5.29 and 5.30 show, respectively, the mean axial velocity contour for spray flame of unheated and preheated VO. The measurements were acquired for $Y = 5$–40 mm in the flame spray. The measurements in the flame spray downstream of this axial location were exceedingly difficult because of the high intensity flame radiation and extremely low data rate. For the PDPA system, the signal-to-noise ratio was too small to obtain any accurate information at low data rates. In general the mean axial velocity exhibits a peak at the center and decays gradually in radial direction, similar to the cold spray results discussed above. However, for the spray

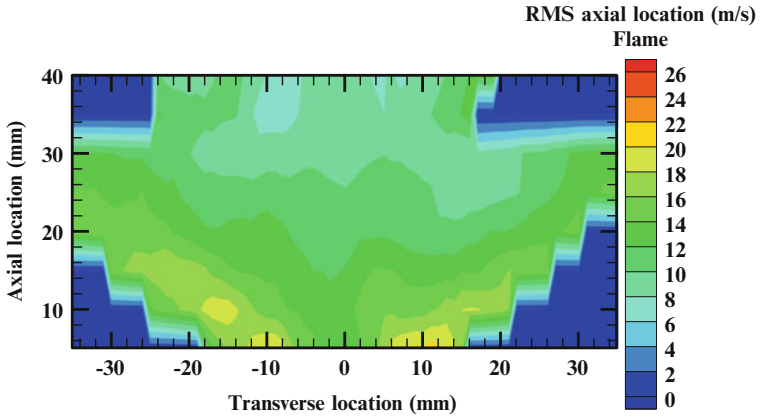



**Fig. 5.31** Axial RMS velocity contour for flame spray

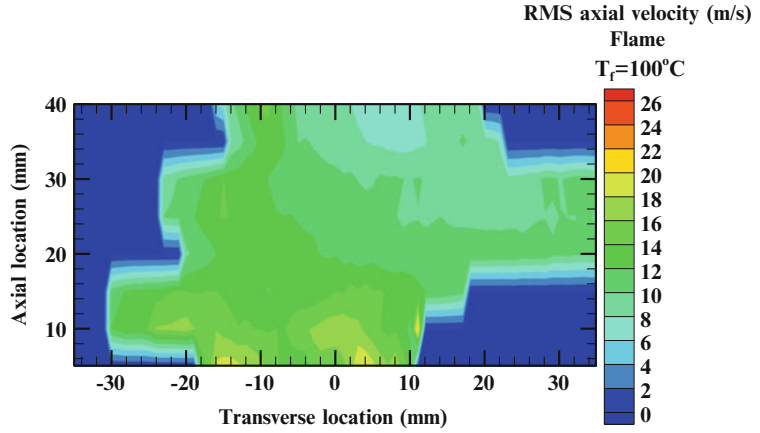



**Fig. 5.32** Axial RMS velocity contour for flame spray for VO at 100 °C

flame the peak velocities are higher because of the increased temperature resulting from heat release. Unlike cold spray with swirl flow, the negative velocities for the droplets are not measured in flame spray, since they evaporate completely (Figs. 5.31 and 5.32).

The mean axial velocity increases drastically in the flame spray compared to the cold spray. The difference of about 35–40 m/s is observed in the peak mean axial velocity for both the cases. The rapidly moving smaller drops in the mid spray region are attributed to the smaller drag drop forces and volumetric expansion of the hot gases. Slow moving larger drops in the outer edge of the spray result from high momentum of such drops. Similar trend is observed for the flame spray for VO at 100 °C as well. The peak mean axial velocities show a marginal difference

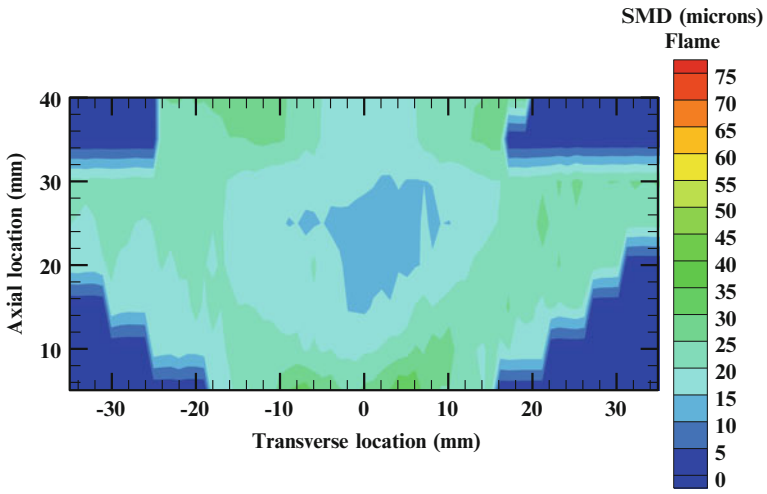



**Fig. 5.33** SMD contour for flame spray

for unheated and heated VO. Results show that fuel preheating increases mean axial velocity leading to improved atomization. Similar improvement is also observed in the combusting sprays compared to the non-burning cold sprays.

SMD Contours

The SMD contour plot for open flame with unheated VO is presented in Fig. 5.33. The larger droplets are seen close to the injector exit between $Y = 5$ and 10 mm, and also on the outer edge at farther downstream locations. The magnitude of the SMD ranges from about 6–37 μm. The maximum SMD of 37 μm in case of flame spray is significantly smaller compared to that of 60 μm for cold spray as seen in Fig. 5.22. The decrease in the SMD is seen in the near field, i.e., $Y = 5$–10 mm. The decrease in the droplet size can be attributed to the droplets passing through the flame zone. The high temperatures in the flame tend to reduce the droplet diameter throughout the spray including the core region near the jet center where larger droplet sizes are observed in the cold spray. The SMD range in case of flame is narrower than the cold spray indicating high evaporation rate for the former case. Farther downstream of the injector exit, SMD decreases rapidly indicating faster evaporation of droplets in the flame. Large droplets do not evaporate completely and thus they are still able to migrate towards the outer edge of the spray.

Figure 5.34 shows the SMD contour of an open flame for preheated VO at 100 °C. In this case the maximum SMD decreases to about 30 μm compared to 37 μm for unheated VO. The flame tends to reduce the radial jet spread and produces a narrower range of SMD indicating droplet evaporation associated with high flame and fuel temperatures.

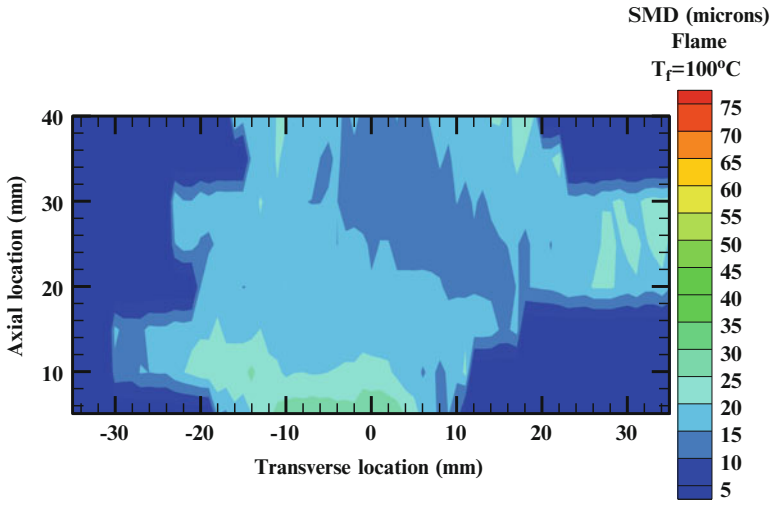



**Fig. 5.34** SMD contour for flame spray for VO at 100 °C

Transverse Profiles of Mean and RMS Axial Velocity

Figure 5.35 shows the mean axial velocity profiles in the cold spray and spray flame. The measurements were taken for axial locations of 5–40 mm in the flame, because of the low data rate at $Y > 40$ mm, indicating nearly complete fuel vaporization at these locations. In general, the mean axial velocity peaks at the center and decreases towards the edge of the spray, similar to the trends in the cold spray discussed above. Figure 5.35a compares the mean axial velocity for cold spray and flame spray at $Y = 5$ mm. The flame spray shows peak velocity of 70 m/s as compared to that of 45 m/s in the cold spray. Thermal expansion in the flame results in droplets with higher mean axial velocity. At $Y = 30$ mm, the peak axial velocity decreases to about 52 m/s as compared to peak axial velocity of 14 m/s for cold spray. The decrease in the axial velocity is attributed to the spray extending over a wider region at downstream locations. Figure 5.35h shows that at $Y = 40$ mm, the peak axial velocity in the flame spray decreased to about 38 m/s compared to that of 10 m/s in the cold spray.

Figure 5.36 shows the effect of flame on spray for VO at 100 °C. The observed trend for the axial velocities is similar to as discussed above. The mean axial velocity at $Y = 5$ mm, peaks to 75 m/s in the flame spray as compared to peak value of 48 m/s in the cold spray. At $Y = 10$ mm and $Y = 15$ mm, the profiles show a dip in the jet center while peak velocity on either side of centerline at $X = 10$ mm. Farther downstream in the spray at $Y = 30$ mm and $Y = 40$ mm, the flame spray velocity peaks at 58 and 50 m/s compared to 20 and 15 m/s in case of cold spray.

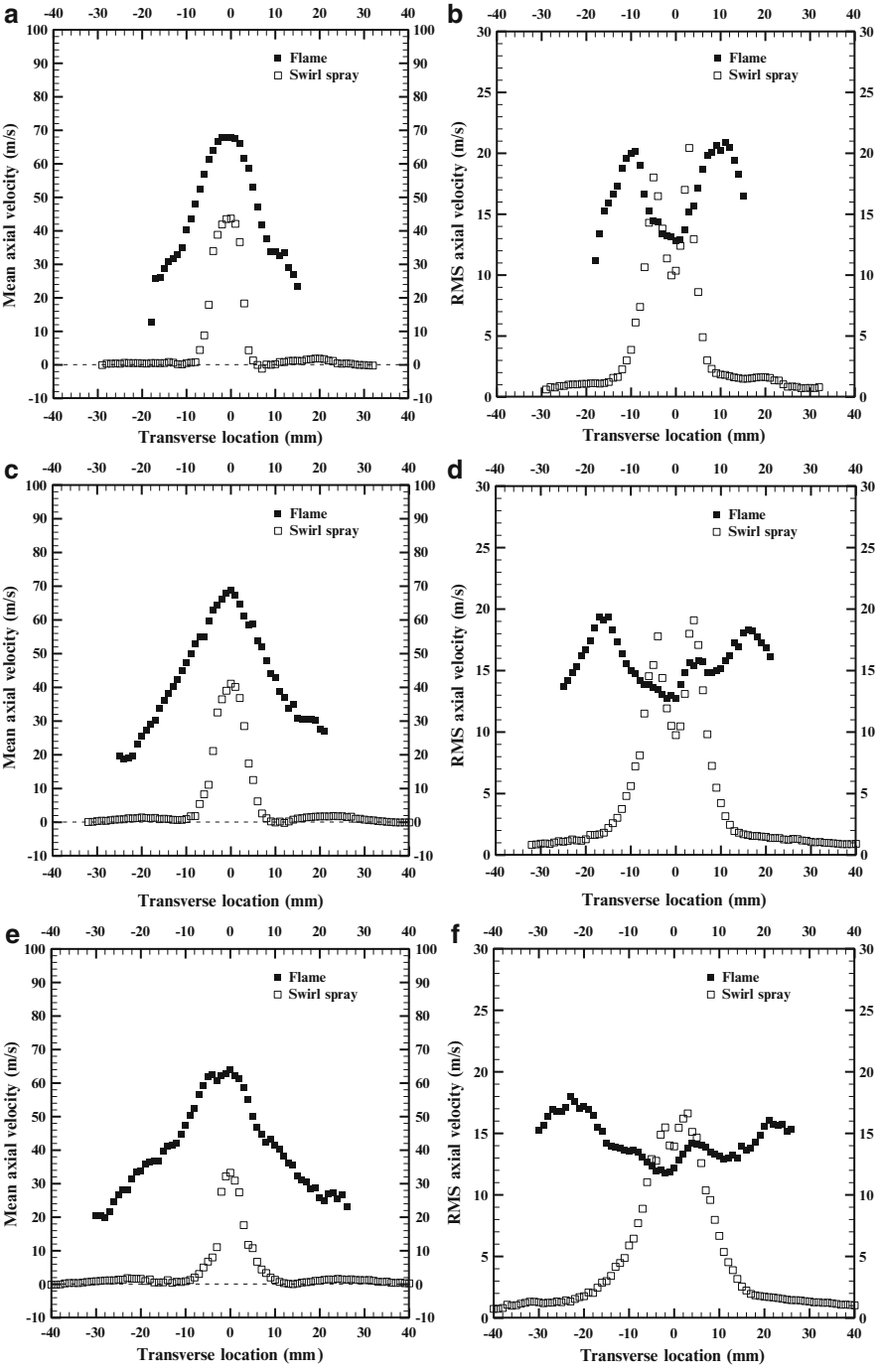



**Fig. 5.35** (continued)

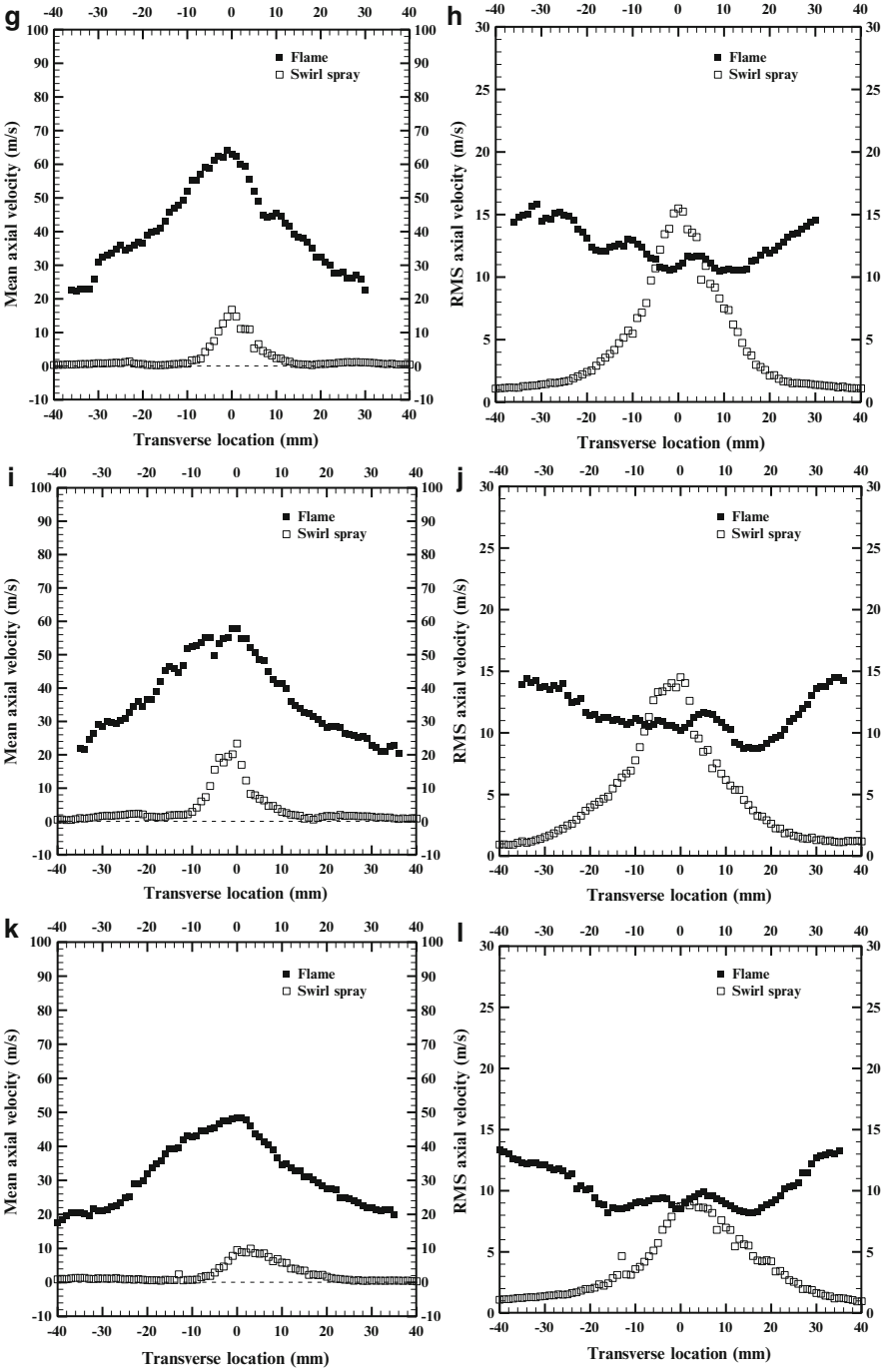



**Fig. 5.35** (continued)

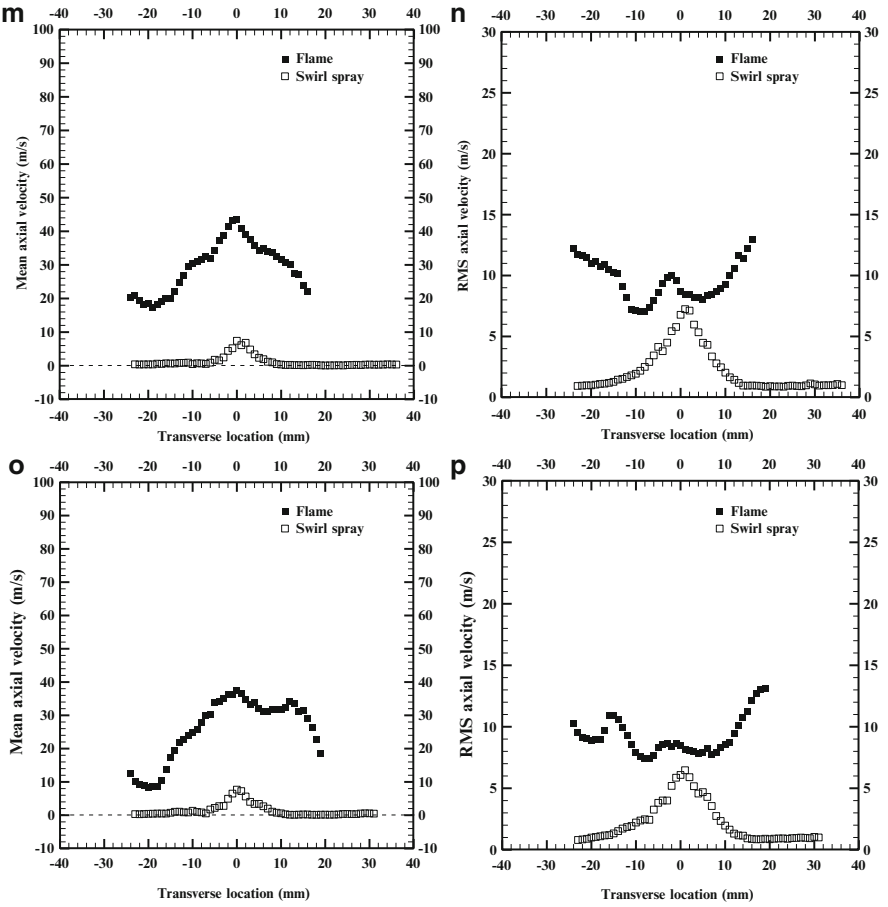



**Fig. 5.35** (**a** and **b**) Transverse profiles of axial mean velocity and axial RMS velocity for flame spray cold spray at $Y = 5$ mm. (**c** and **d**) Transverse profiles of axial mean velocity and axial RMS velocity for cold spray with and without swirl at $Y = 10$ mm. (**e** and **f**) Transverse profiles of axial mean velocity and axial RMS velocity for cold spray with and without swirl at $Y = 15$ mm. (**g** and **h**) Transverse profiles of axial mean velocity and axial RMS velocity for cold spray with and without swirl at $Y = 20$ mm. (**i** and **j**) Transverse profiles of axial mean velocity and axial RMS velocity for cold spray with and without swirl at $Y = 25$ mm. (**k** and **l**) Transverse profiles of axial mean velocity and axial RMS velocity for cold spray with and without swirl at $Y = 30$ mm. (**m** and **n**) Transverse profiles of axial mean velocity and axial RMS velocity for cold spray with and without swirl at $Y = 35$ mm. (**o** and **p**) Transverse profiles of axial mean velocity and axial RMS velocity for cold spray with and without swirl at $Y = 40$ mm

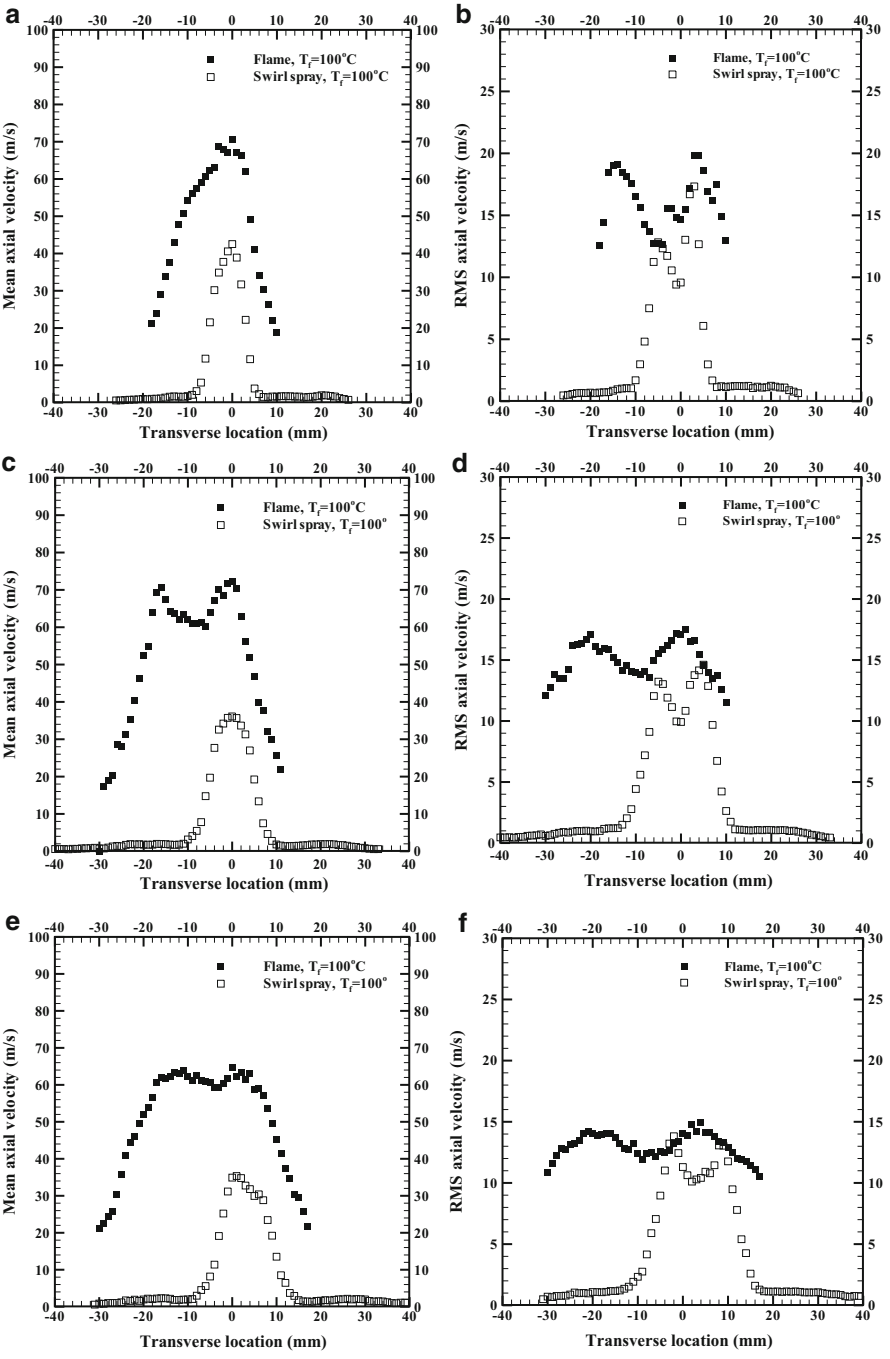



**Fig. 5.36** (continued)

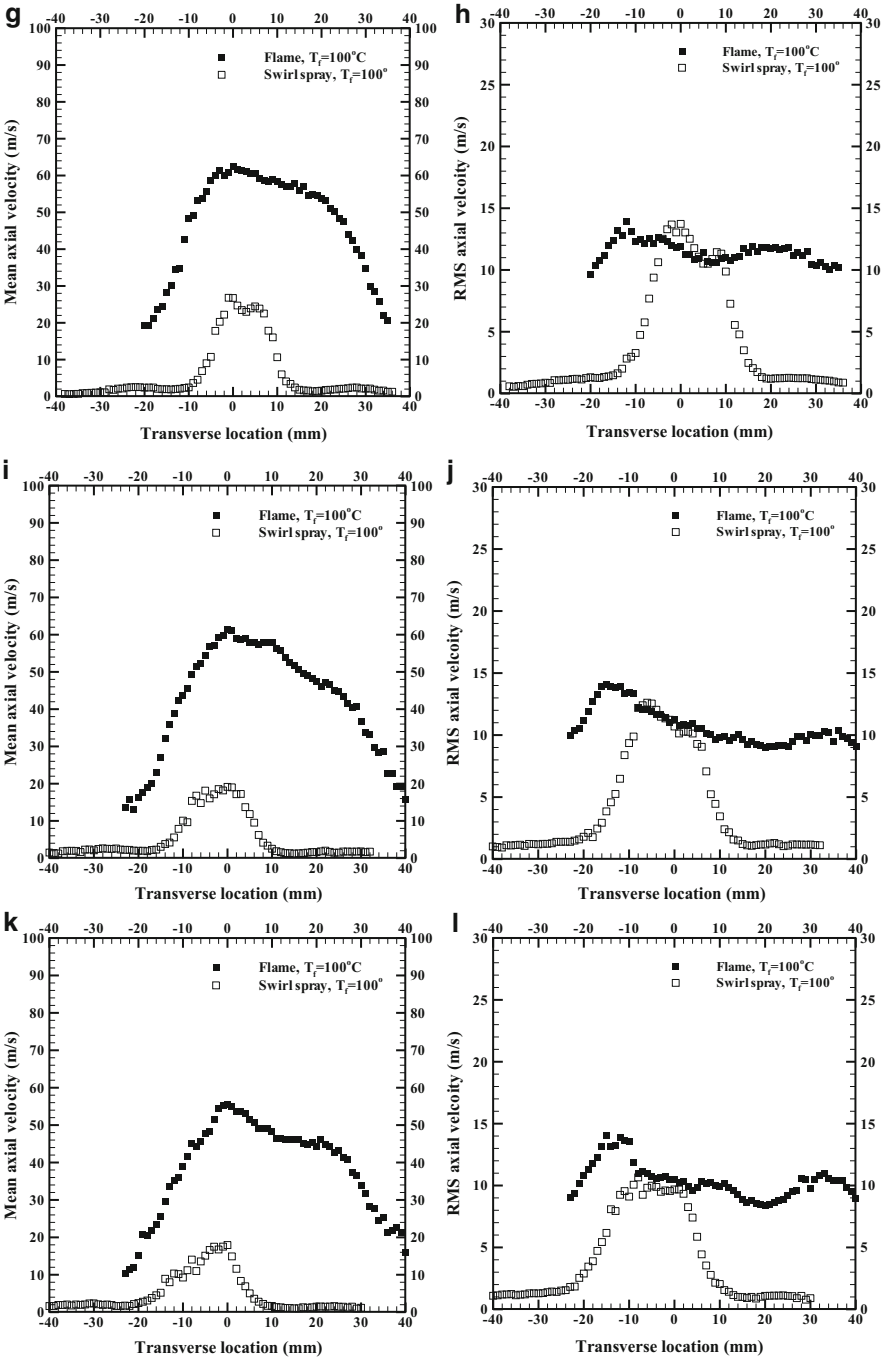



**Fig. 5.36** (continued)

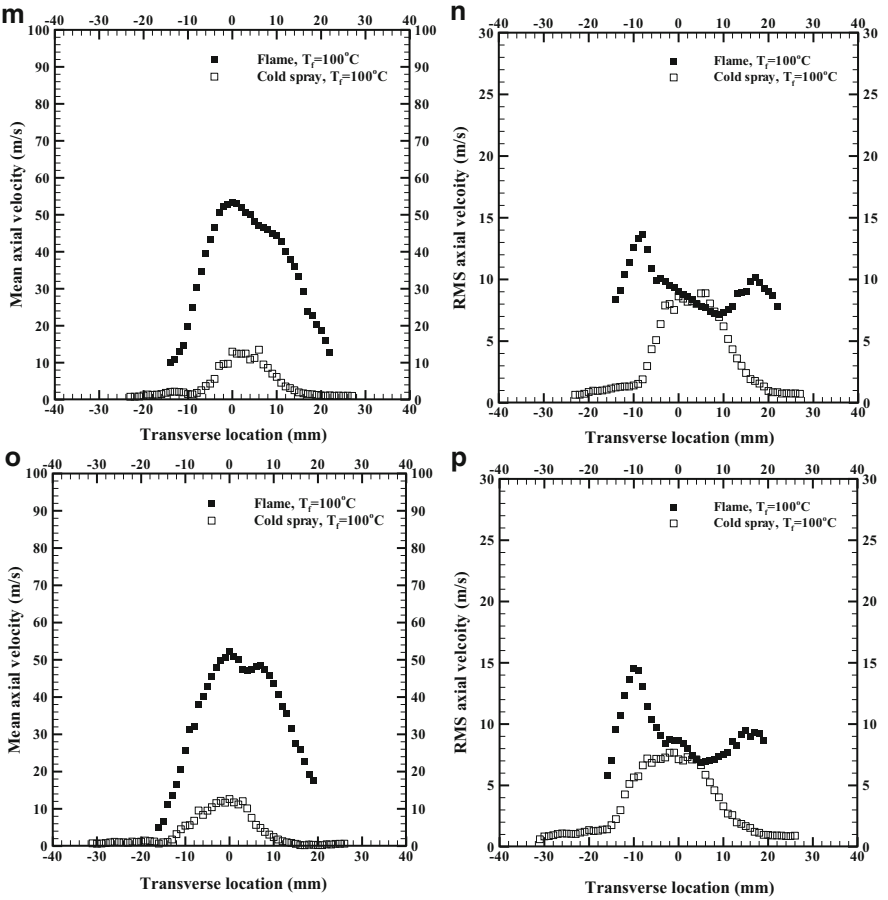



**Fig. 5.36** (**a** and **b**) Transverse profiles of mean axial velocity and RMS axial velocity for flame spray cold spray at $Y = 5$ mm for $T_f = 100$ °C. (**c** and **d**) Transverse profiles of mean axial velocity and RMS axial velocity for flame spray cold spray at $Y = 10$ mm for $T_f = 100$ °C. (**e** and **f**) Transverse profiles of mean axial velocity and RMS axial velocity for flame spray cold spray at $Y = 15$ mm for $T_f = 100$ °C. (**g** and **h**) Transverse profiles of mean axial velocity and RMS axial velocity for flame spray cold spray at $Y = 20$ mm for $T_f = 100$ °C. (**i** and **j**) Transverse profiles of mean axial velocity and RMS axial velocity for flame spray cold spray at $Y = 25$ mm for $T_f = 100$ °C. (**k** and **l**) Transverse profiles of mean axial velocity and RMS axial velocity for flame spray cold spray at $Y = 30$ mm for $T_f = 100$ °C. (**m** and **n**) Transverse profiles of mean axial velocity and RMS axial velocity for flame spray cold spray at $Y = 35$ mm for $T_f = 100$ °C. (**o** and **p**) Transverse profiles of mean axial velocity and RMS axial velocity for flame spray cold spray at $Y = 40$ mm for $T_f = 100$ °C

Figure 5.35 shows that the RMS axial velocity is observed in the flame spray as well, as discussed above for cold spray. The flame spray profiles show a dip in the center and peaks on either side of the centerline which represent the flame locations at $X = 15$–$28$ mm, unlike the cold spray having a decaying trend with peak at the

center. This is attributed due to the presence of higher turbulent fluctuations in case of the flame spray as compared to that of the cold spray. The magnitude of RMS axial velocity ranges from 10 to 22 m/s which is consistent to the mean axial velocity discussed in the previous sections. The maximum RMS axial velocity values of about 22, 18, and 16 m/s are observed at $Y = 5$, 10, and 15 mm, respectively. For the flame spray, the RMS axial velocity is higher than those of the cold spray on the outer edge of the spray while they are almost similar in magnitude in the center of the spray.

Figure 5.36 shows the similar trend for flame spray for preheated VO at 100 °C. At the near field locations, the profiles peak nearly at same values for both flame spray and cold spray at $Y = 5$, 10, and 15 mm. The flame spray profiles show a slight asymmetry at the center which can be attributed due to the asymmetry in the flame. The magnitude in the RMS axial velocity values varies from 5 to 20 m/s for the axial locations of $Y = 5$–40 mm in the flame spray. As discussed above the RMS values are smaller in case of cold spray compared to that of flame.

### Transverse Profiles of SMD

Transverse profiles of SMD for flame spray and cold spray are presented in Fig. 5.37. Figure 5.37a shows the effect of flame on spray at axial location of $Y = 5$ mm. For the flame spray the profiles of SMD show smaller droplets in the center and slightly larger droplets on each side, and gradual decrease in droplet diameter in the transverse direction. This trend is same for near field axial locations of $Y = 5$, 10, 15, and 20 mm. The maximum SMD for the near field location in the flame spray is observed to be about 35 μm at $Y = 5$ mm, 25 μm at 10 mm, and 15 μm at 15 mm. Farther downstream, a noticeable change in the SMD trend is observed; smaller droplets are at the center while the larger droplets on the outer edge. This result is attributed to the fact that the droplets at these locations are passing through the flame zone. The size difference farther away from the centerline is observed to be larger because at the outer edge of the spray the smaller droplets evaporate faster compared to the large droplets and the larger droplets with high inertia penetrate out of the flame. The escaped larger droplets are believed to burn in diffusion mode. Comparatively the SMD range for the burning spray is about 35 μm smaller than that of the cold spray for near field locations. The presence of flame increases the evaporation rate of the droplets hence producing smaller droplets. The radial spread for the flame spray is narrower compared to cold spray indicating that the droplets in case of cold spray possess higher initial momentum hence penetrating farther and wider.

Figure 5.38 shows the radial profiles for SMD for fuel inlet temperature of 100 °C for flame and cold spray. Similar trend is observed for all the cases in Fig. 5.37 as discussed above. The effect of higher fuel inlet temperature is to decrease the fuel kinematic viscosity and surface tension, hence improving atomization characteristics by decreasing the mean droplet size. For cold spray the maximum SMD is 45 μm compared to 30 μm for the flame spray at $Y = 5$ mm. The SMD

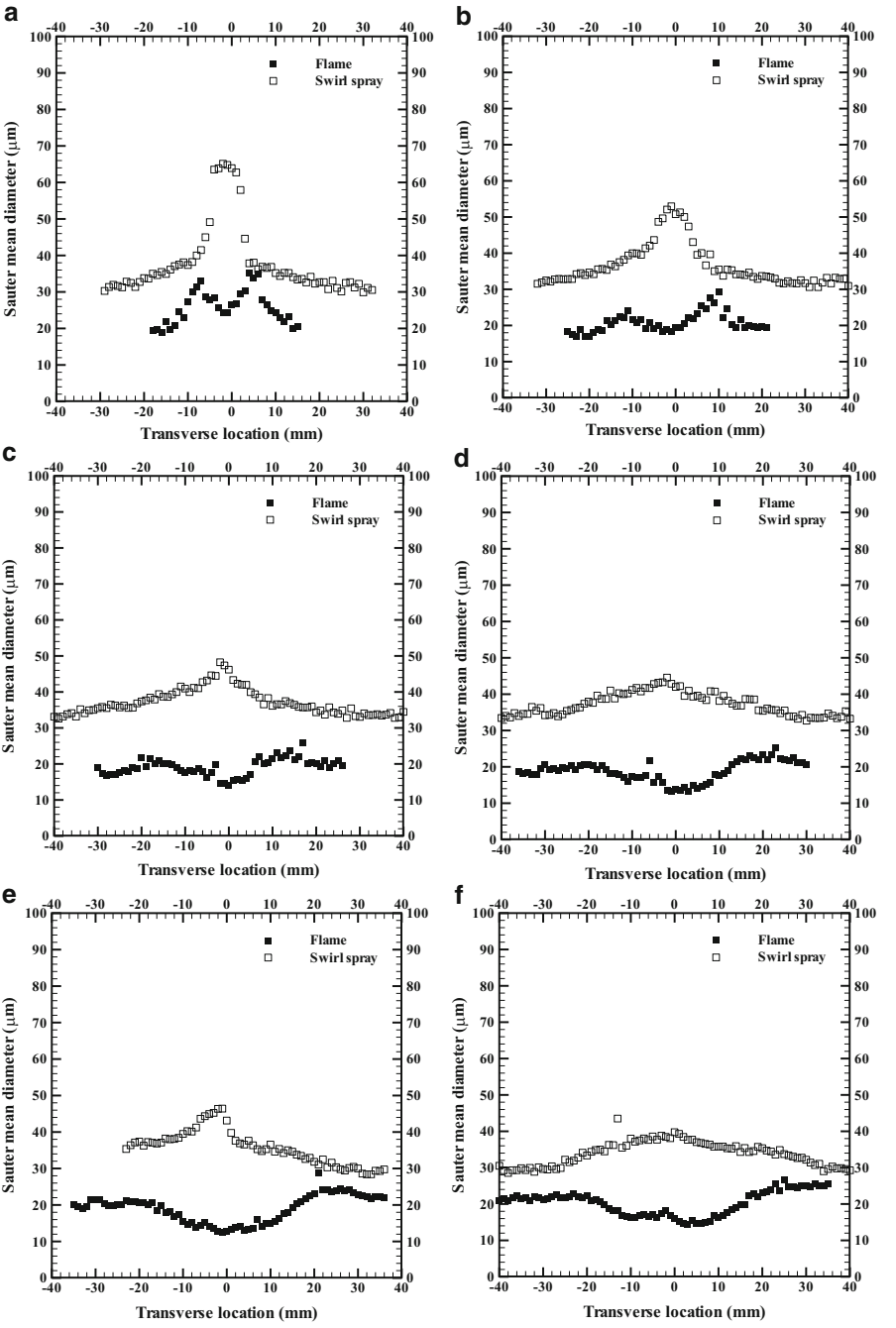



**Fig. 5.37** (continued)

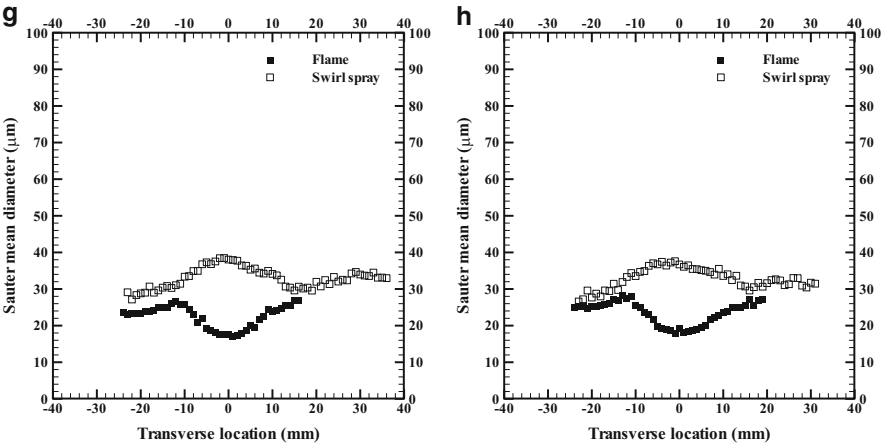



**Fig. 5.37** (**a** and **b**) Transverse profiles of SMD for flame spray and cold spray at $Y = 5$ and 10 mm. (**c** and **d**) Transverse profiles of SMD for flame spray and cold spray at $Y = 15$ and 20 mm. (**e** and **f**) Transverse profiles of SMD for flame spray and cold spray at $Y = 25$ and 30 mm. (**g** and **h**) Transverse profiles of SMD for flame spray and cold spray at $Y = 35$ and 40 mm.

values decrease farther downstream in the spray compared to the near field locations. And this trend decreases at farther downstream locations in the spray. Evidently the higher fuel inlet temperature helped burning of the droplets much faster compared to unheated case discussed in Fig. 5.37. The minimum SMD measured for the flame spray is reduced to about 10 μm compared to 30 μm in the cold spray. The minimum for unheated flame spray case was about 20 μm. Also, the axial locations in the flame spray were measured to only about 40 mm, since most of the droplets were evaporated by the time they reached farther downstream. Hence the data rate was significantly low to measure any signal from the droplets above this axial locations in the spray.

### Droplet Diameter Distribution Profiles

Figure 5.39 shows the droplet diameter distribution profiles for flame spray and cold spray at near and far field locations of $Y = 5$, 10, 35, and 40 mm and $X = 0$ mm. Figure 5.39a shows the distribution profile for near the injector exit in the center of the spray at $Y = 5$ mm. Results show a significant difference in the droplet diameter distribution for flame and cold spray. The flame spray is observed to have most of the droplets of $\leq$50 μm while large droplets with >100 μm are found in the cold spray. The largest droplet in the flame spray is about 100 μm where as for the cold spray it is about 200 μm. For the flame spray the droplets tend to start vaporizing due to the high flame temperature zones, impact of phase interactions, and combustion producing smaller droplets and narrower distribution compared to the wider spread

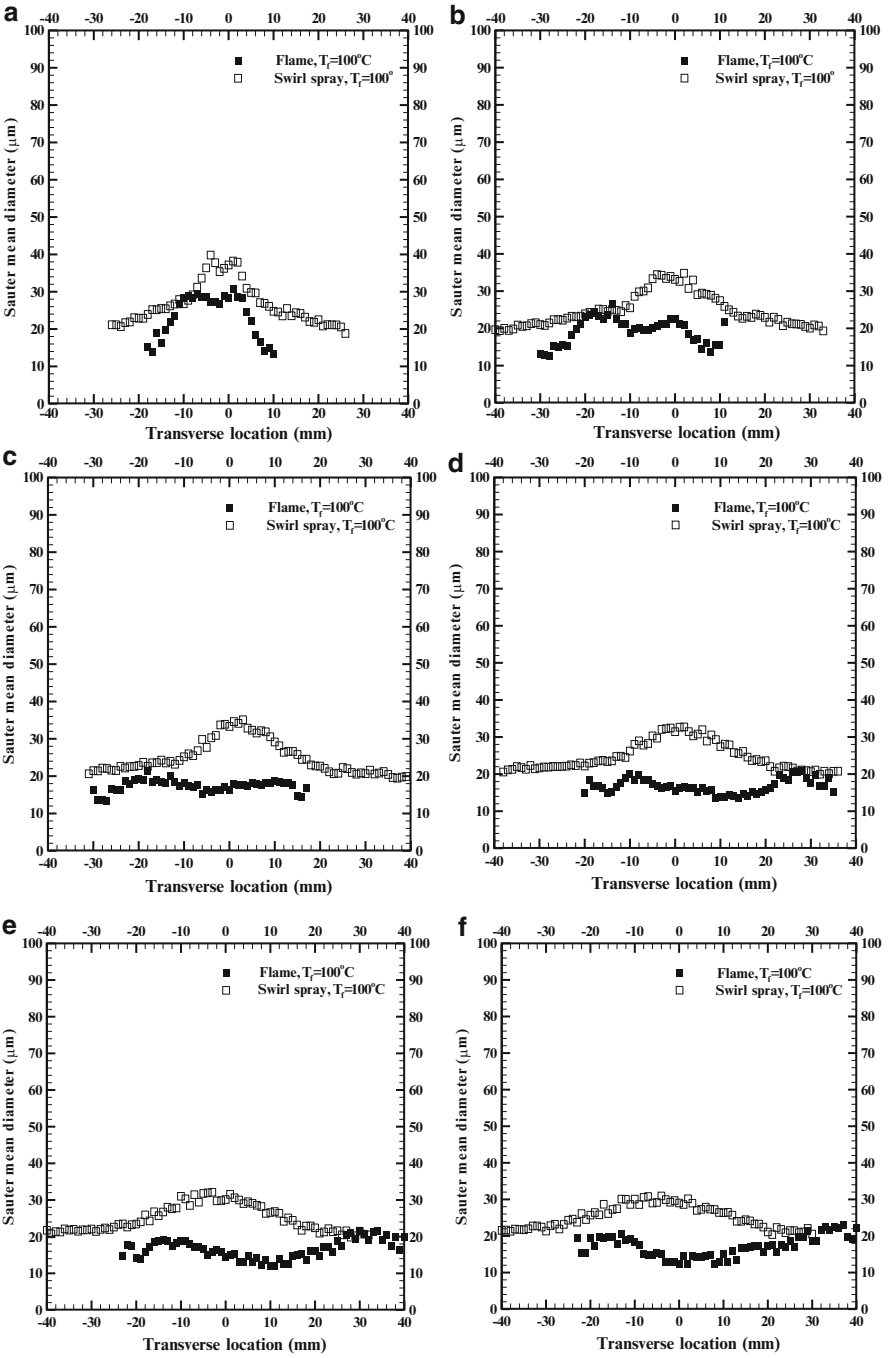



**Fig. 5.38** (continued)

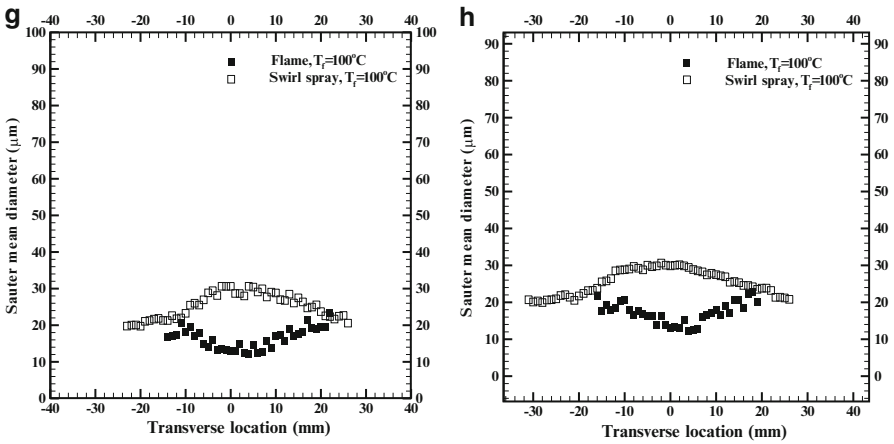



**Fig. 5.38** (**a** and **b**) Transverse profiles of SMD for flame spray and cold spray at $Y = 5$ and 10 mm for $T_f = 100$ °C. (**c** and **d**) Transverse profiles of SMD for flame spray and cold spray at $Y = 15$ and 20 mm for $T_f = 100$ °C. (**e** and **f**) Transverse profiles of SMD for flame spray and cold spray at $Y = 25$ and 30 mm for $T_f = 100$ °C. (**g** and **h**) Transverse profiles of SMD for flame spray and cold spray at $Y = 35$ and 40 mm for $T_f = 100$ °C

distribution as observed in the cold spray. The SMD for flame spray and cold spray is respectively 26 and 64 μm, indicating rapid evaporation of the droplets in the flame compared to the cold spray.

Figure 5.39b, c shows, respectively, the droplet diameter distribution profiles for $Y = 10$ and 35 mm and $X = 0$ mm, respectively. At $Y = 10$ mm, $X = 0$ mm, the flame spray has most of the droplets of diameter <30 μm while the cold spray has majority of droplets 130 μm diameter. The largest diameter for flame spray is 75 μm while that of cold spray is 175 μm. The SMD is 19 and 50 μm, respectively, for flame spray and cold spray. The higher count of larger droplets dominates the higher SMD for cold spray. The larger numbers of smaller droplets in case of flame spray result in a decrease in SMD. At $Y = 35$ mm as shown in Fig. 5.39c, the largest droplet diameter in cold spray is 125 μm where as that in the flame spray is 60 μm. The respective SMD for these cases is 17 and 38 μm for flame spray and cold spray. Again there is a higher count of smaller droplets in case of flame spray resulting in the overall decrease in the SMD. As observed the distribution gets narrower with increase in axial distance, due to the droplets getting evaporated rapidly and combusted in the flame locations and leading to smaller number of larger droplets. The distribution profile is narrower in for far field locations in the spray for both the cases compared to the locations near the nozzle exit. Moreover the number density of the droplet further downstream decreases in the flame spray due to the droplet evaporation and burning.

Figure 5.40 shows the droplet diameter distribution profiles for the flame spray for preheated VO at 100 °C. The largest droplet diameter for cold spray is observed to be 140 μm where as for flame spray it is 90 μm. The increase in fuel inlet

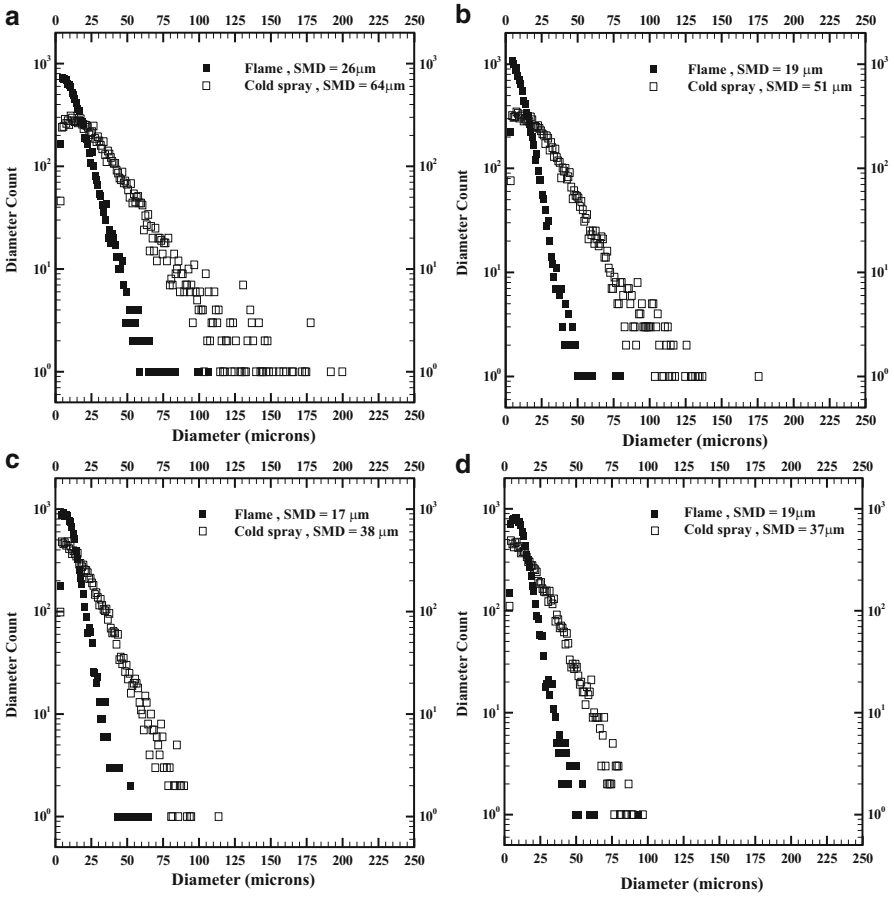



**Fig. 5.39** (**a** and **b**) Droplet distribution profile for flame spray and cold spray at $Y = 5$ mm, $X = 0$ mm and $Y = 10$ mm, $X = 0$ mm. (**c** and **d**) Droplet distribution profile for flame spray and cold spray at $Y = 35$ mm, $X = 0$ mm and $Y = 40$ mm, $X = 0$ mm

temperature produces a narrow distribution profile with lower range of droplet diameter and hence with overall smaller SMD. The SMDs for both the cases is 28 and 37 μm for flame spray and cold spray, respectively. At $Y = 10$ mm, $X = 0$ mm the largest SMD for flame spray is of about 75 μm as well as for cold spray is of about 115 μm, respectively. This decreases the global SMD to 22 and 33 μm for flame spray and cold spray. The increase in fuel inlet temperature aids to further reduce the SMD as compared to the unheated case. As we move axially farther downstream, the droplets distribution becomes narrower and the number density also reduces especially significantly for the flame spray attributed to the fact that the droplets get vaporized in the flame zone. Also, the preheating of the oil produces the finer droplets to begin with since it improves the thermo physical properties of the fuel itself, and hence the finer droplets get vaporized and consumed faster.

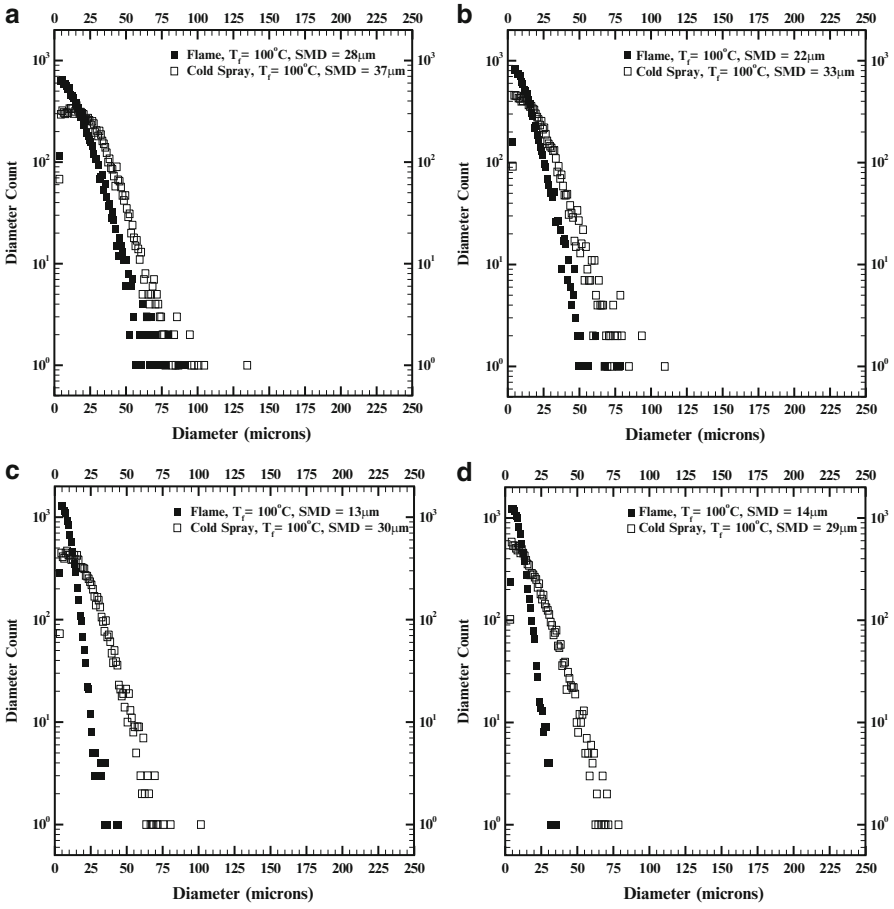



**Fig. 5.40** (**a** and **b**) Droplet distribution profile for flame spray and cold spray at $Y = 5$ mm, $X = 0$ mm and $Y = 10$ mm, $X = 0$ mm for $T_f = 100$ °C. (**c** and **d**) Droplet distribution profile for flame spray and cold spray at $Y = 35$ mm, $X = 0$ mm and $Y = 40$ mm, $X = 0$ mm for $T_f = 100$ °C

Farther downstream at $Y = 35$ mm, the largest droplet diameter is of about 100 μm for cold spray and 40 μm for flame spray producing the SMD of 31 and 13 μm, respectively. Similar to the previous discussions the far field region in the spray has narrower distribution with smaller droplet number density compared to that of the center of the spray. This is consistent to that of the SMD profiles with dense center liquid jet near the injector exit showing higher SMD values and decreasing trend as we move axially downstream with a wider and more dispersed spray. Also the fact that the droplets in case of flame spray get consumed faster and get evaporated rapidly, mostly the smaller ones compared to the larger drops and hence globally being responsible for smaller SMD values.

#### 5.1.10.3 Effect of Enclosure on Spray Flame

Experiments were done to study the effect enclosure on flame spray characteristics and combustion emissions. The enclosure was insulated to provide thermal feedback to the VO flame and maintain a stable VO flame during all experimental conditions. It required about 2 h to preheat the enclosure with a stable VO flame before the methane flow rate was reduced to about 3.8 SLPM which corresponds to $\varphi = 0.89$. This mode of operation resulted in no condensation of the droplets on the glass window of the enclosure. Practically in the continuous combustion operations of gas turbine applications the enclosed flames are used, and hence the present work also incorporates enclosure effect on the spray characteristics and emissions. At steady state insulation around the enclosure, retained sufficient heat within the system to raise the VO temperature at the injector exit to 90–100 °C

Transverse Profiles of Mean and RMS Axial Velocity

Figure 5.41 shows the mean axial velocity profiles for enclosed flames of VO preheated to 100 °C and 150 °C. As seen in Fig. 5.41a, at $Y = 5$ mm, mean axial velocity shows a peak at the center and gradual decrease towards the edge of the spray. The peak mean axial velocity for both cases is nearly the same. For both cases, the axial velocity peaks at 70 m/s, and both the profiles almost overlap each other, with slight difference towards the edge of the spray. At $Y = 10$ and 15 mm the profiles are again observed to be overlapping, although at higher fuel inlet temperature the spread of the spray is observed to be narrower as compared to VO at 100 °C. The transverse spray for VO at 100 °C is extended widely from −40 to 30 mm while for VO at 150 °C, the spread narrows down to a −15 to 20 mm for $Y = 5$, 10, and 15 mm axially. Farther downstream, the peak axial velocity decreased down to 60 m/s for VO at 150 °C and 50 m/s for VO at 100 °C. Unlike velocities, a significant difference is observed in the droplet sizes as shown in the SMD profiles for the enclosed flame. In enclosure, the co-flow swirling air is the combustion air for the flame without any ambient entrainment. In open flame there is no heat feedback but there is a lot of ambient entrainment, while in enclosure heat feedback is there due to insulation.

Figure 5.41b shows the RMS axial velocity profiles for enclosed flame. For VO at 100 °C at $Y = 5$ mm, the peak RMS axial velocity is 22 m/s while for VO at 150 °C the peak RMS axial velocity is 20 m/s. For VO at 100 °C, the profiles show a local minima at the jet center line while double peaks at flame locations, i.e., $X = -30$–15 mm and $Y = 10$–35 mm. The magnitude of the RMS axial velocity ranges from about 2–22 m/s for $Y = 5$ mm to $Y = 35$ mm. Slight asymmetry in the RMS axial velocity profiles is observed at far field locations in the spray i.e. $Y = 25$–35 mm. Comparatively for VO at 150 °C, the RMS axial velocity profiles look more symmetrical with local minima dip in the centerline as well as double peaks on the flame locations. This result is attributed to the fact that the jet center is subjected to the small turbulent fluctuations with lower RMS values compared to

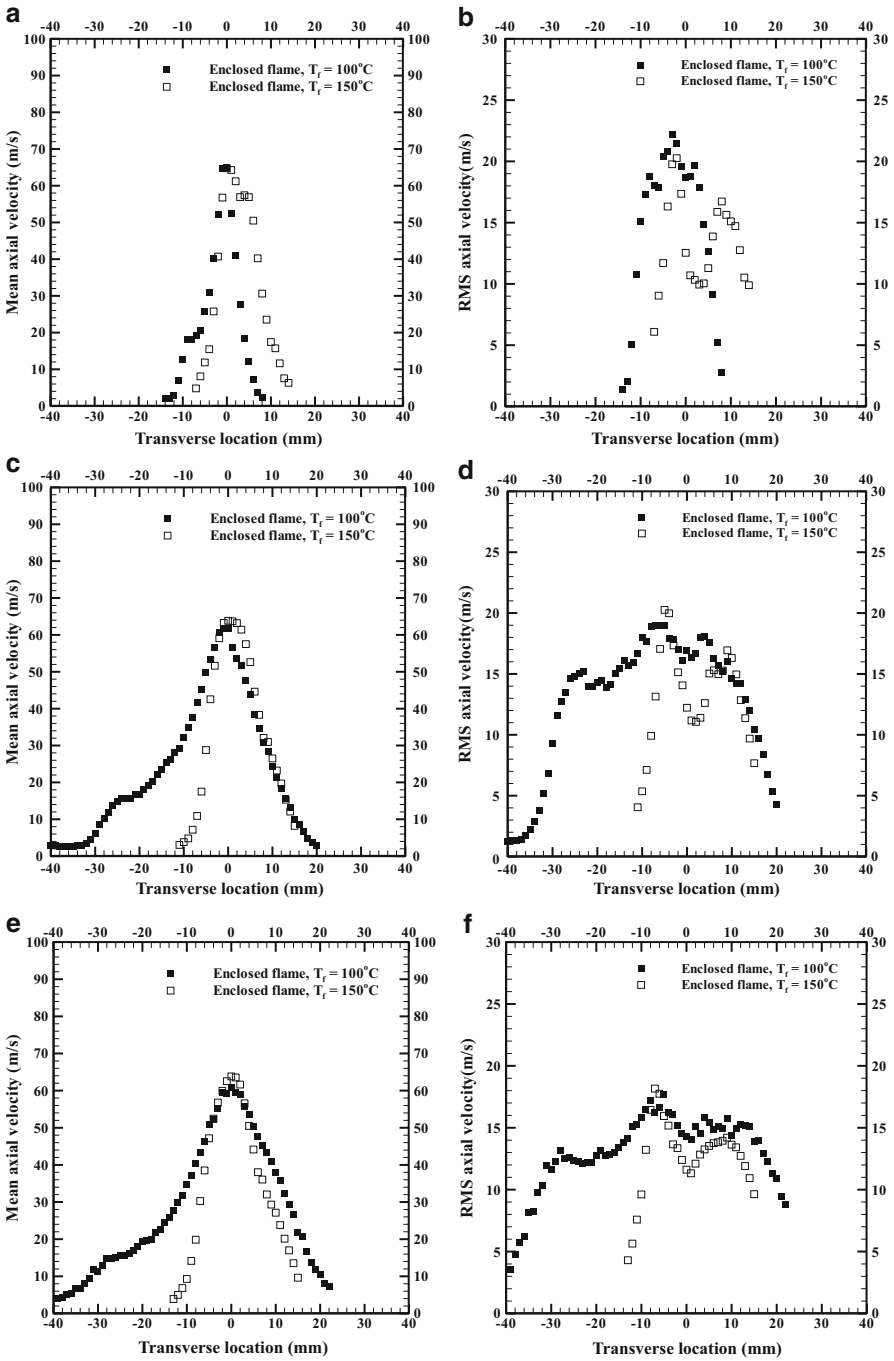



**Fig. 5.41** (continued)

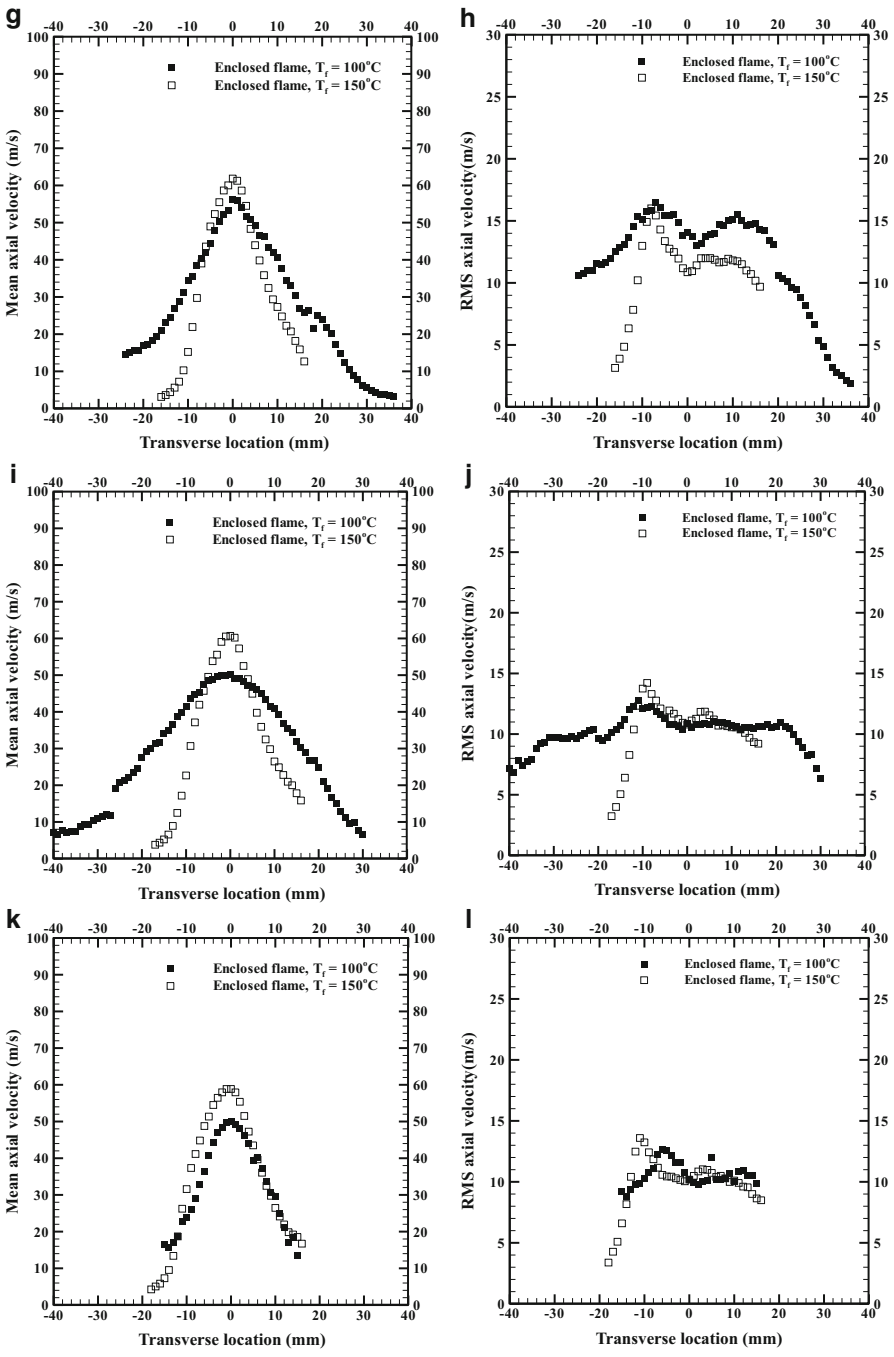



**Fig. 5.41** (continued)

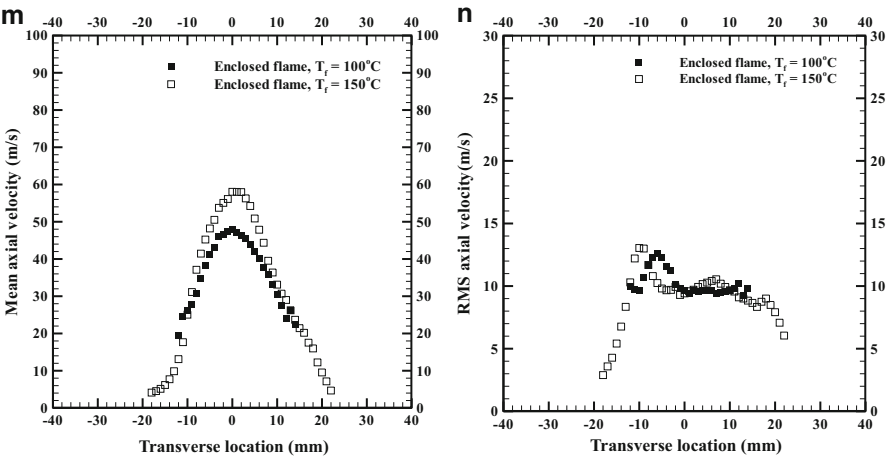



**Fig. 5.41** (**a** and **b**) Transverse profiles of mean axial velocity and RMS axial velocity for enclosed flame for $T_f = 100$ °C and $T_f = 150$ °C at $Y = 5$ mm. (**c** and **d**) Transverse profiles of mean axial velocity and RMS axial velocity for enclosed flame for $T_f = 100$ °C and $T_f = 150$ °C at $Y = 10$ mm. (**e** and **f**) Transverse profiles of mean axial velocity and RMS axial velocity for enclosed flame for $T_f = 100$ °C and $T_f = 150$ °C at $Y = 15$ mm. (**g** and **h**) Transverse profiles of mean axial velocity and RMS axial velocity for enclosed flame for $T_f = 100$ °C and $T_f = 150$ °C at $Y = 20$ mm. (**i** and **j**) Transverse profiles of mean axial velocity and RMS axial velocity for enclosed flame for $T_f = 100$ °C and $T_f = 150$ °C at $Y = 25$ mm. (**k** and **l**) Transverse profiles of mean axial velocity and RMS axial velocity for enclosed flame for $T_f = 100$ °C and $T_f = 150$ °C at $Y = 30$ mm. (**m** and **n**) Transverse profiles of mean axial velocity and RMS axial velocity for enclosed flame for $T_f = 100$ °C and $T_f = 150$ °C at $Y = 35$ mm

that at the edge of the spray. Near the injector exit, i.e., at $Y = 5$–15 mm, the peak RMS axial velocity for both the cases overlaps with no significant difference in the velocity values. At locations farther downstream, the peak RMS axial velocities for 150 °C are higher at $X = 20$ mm, indicating higher velocity fluctuations compared to VO at 100 °C.

### Transverse Profiles of SMD

Figure 5.42 shows the transverse profiles of SMD for enclosed VO flame at 100°c and 150 °C. Figure 5.42a shows the SMD profile at $Y = 5$ mm. The profiles show peak SMD value at the center of the spray. The maximum SMD for VO at 100 °C is 34 μm, while for VO at 150 °C is 32 μm. At $Y = 15$ and 20 mm, the range of SMD values decreases. The profiles no longer show a well-defined central peak as compared to near the injector exit. SMD values range from 10 to 25 μm for VO at 100 °C while for VO at 150 °C the SMD ranges from 18 to 28 μm. Farther downstream, at $Y = 25$–35 mm, the larger droplets are observed to have moved towards the periphery of the spray causing higher SMDs at the outer edge of the

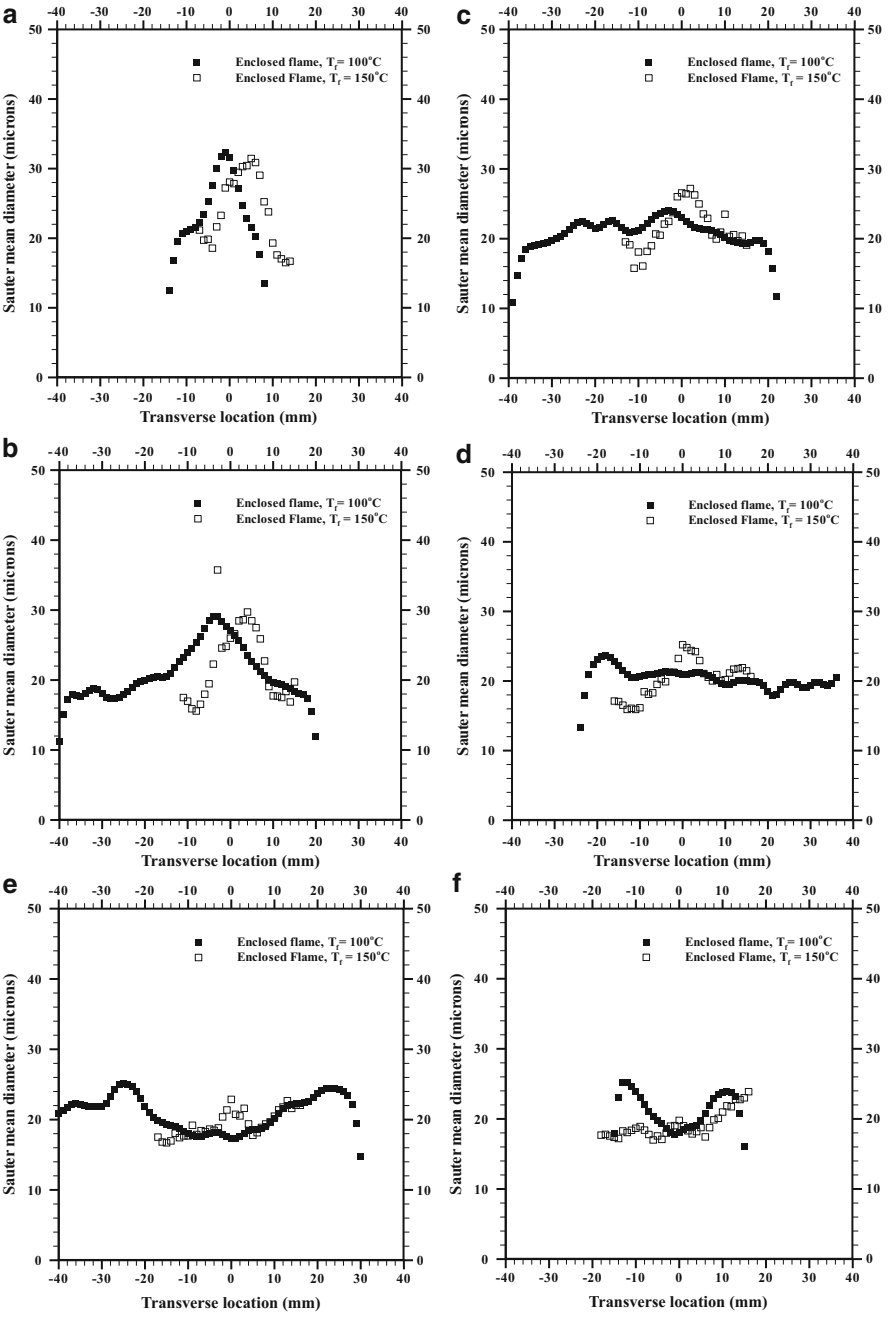



**Fig. 5.42** (continued)

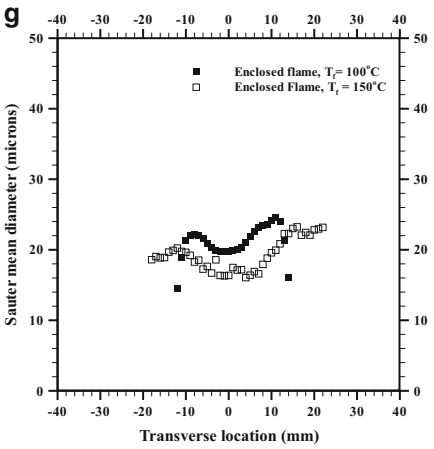



**Fig. 5.42** (**a** and **b**) Transverse profiles of SMD for enclosed flame for $T_f = 100$ °C and $T_f = 150$ °C at $Y = 5$ mm and $Y = 10$ mm. (**c** and **d**) Transverse profiles of SMD for enclosed flame for $T_f = 100$ °C and $T_f = 150$ °C at $Y = 15$ mm and $Y = 20$ mm. (**e** and **f**) Transverse profiles of SMD for enclosed flame for $T_f = 100$ °C and $T_f = 150$ °C at $Y = 25$ mm and $Y = 30$ mm. (**g**) Transverse profiles of SMD for enclosed flame for $T_f = 100$ °C and $T_f = 150$ °C at $Y = 35$ mm

spray as seen in Fig. 5.44e, f. The profiles show a central depression as compared to central peak observed in the near injector locations ($Y = 5$ and 10 mm). For VO at 100 °C, the SMD ranges from 16 to 25 μm, which is larger compared to 18–22 μm for VO at 150 °C. For VO at 150 °C, the transverse distribution of SMD is narrower compared to VO at 100 °C. This result is attributed to the higher fuel inlet temperature and improved vaporization of the droplets thus reducing the SMD at downstream locations in the flame. At $Y = 35$ mm, the SMD profile for VO at 100 °C by 5 μm compared to that for VO at 150 °C. For VO at 100 °C, the SMD values show a decrease of about 12 μm with axial distance while that for VO at 150 °C, the SMD decreases by about 15 μm.

### Droplet Diameter Distribution Profiles

Figure 5.43 shows the droplet size distribution profiles for enclosed VO flame at 100 and 150 °C, respectively. At $Y = 5$ mm, $X = 0$ mm, the largest diameter for VO 100 °C is 150 μm while that for VO at 150 °C is 60 μm. For VO at 150 °C, a narrow distribution profile is observed with greater number of smaller droplets. 90 % of the droplet sizes are in the range of <50 μm. The presence of few large drops highly influences the SMD in case of lower fuel inlet temperature, i.e., VO at 100 °C. The SMD for VO at 150 °C is 28 μm while for VO at 100 °C is 32 μm. At $Y = 10$ mm, $X = 0$ mm, the distribution profile is almost similar for the drops <30 μm. A minor difference in droplet diameter distribution is observed in the droplets >50 μm. The lower fuel inlet temperature, i.e., VO at 100 °C, has more large droplets compared to

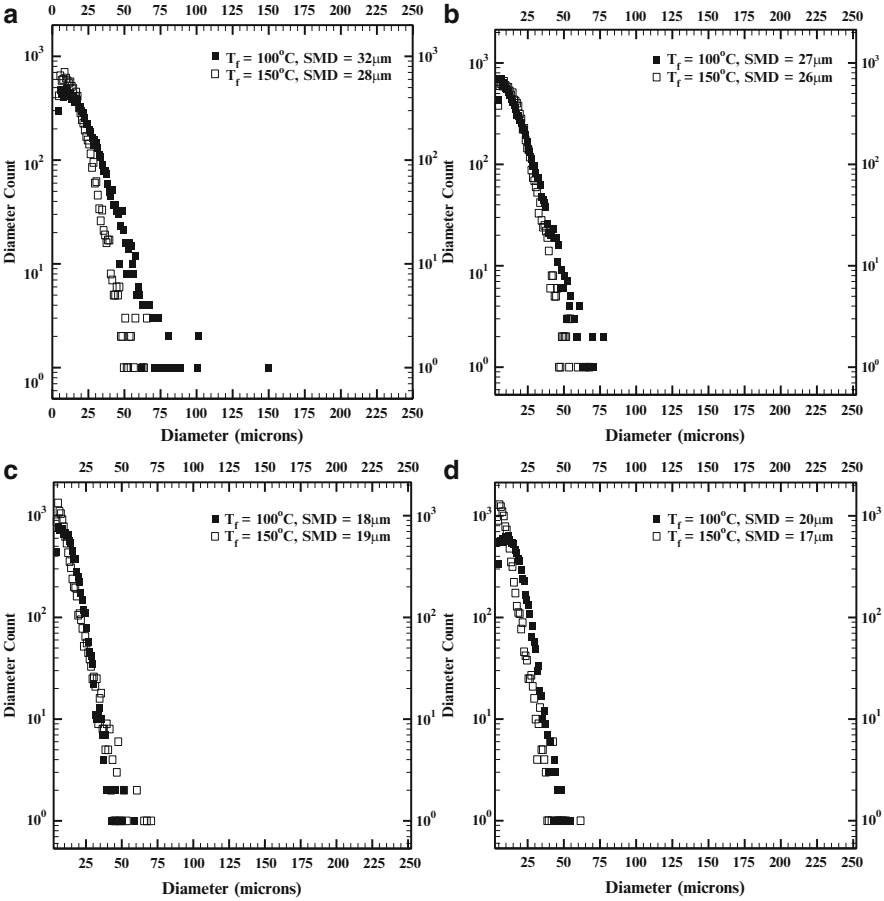



**Fig. 5.43** (**a** and **b**) Droplet distribution profile for enclose flame for $T_f = 100$ °C and $T_f = 150$ °C at $Y = 5$ mm, $X = 0$ mm and $Y = 10$ mm, $X = 0$ mm. (**c** and **d**) Droplet distribution profile for enclosed flame for $T_f = 100$ °C and $T_f = 150$ °C at $Y = 30$ mm, $X = 0$ mm and $Y = 35$ mm, $X = 0$ mm

that for VO at 150 °C. The SMD for VO at 100 °C is 27 μm and for VO at 150 °C is 25 μm. Largest diameter for VO at 150 °C is of about 60 μm and for VO at 100 °C is of about 80 μm, respectively.

At $Y = 35$ mm, $X = 0$ mm, VO at 150 °C shows higher count of smaller droplets for drops <11 μm compared to that of VO T 100 °C. The largest droplet for VO 150 °C is 70 μm and for VO 100 °C 58 μm, respectively. For drops >40 μm, the distribution in case of VO 150 ° C is higher than that of VO at 100 ° C. The respective SMDs for both the cases are 20 and 18 μm which are influenced by the presence of few large droplets in the case of VO at 150 °C. At $Y = 35$ mm, $X = 0$ mm, the largest diameter for VO at 150 °C is 60 μm where as for VO 100 °C is 55 μm. The respective SMDs for both the cases are 16 and 20 μm, which is attributed to the fact of larger percentage of smaller droplets for higher

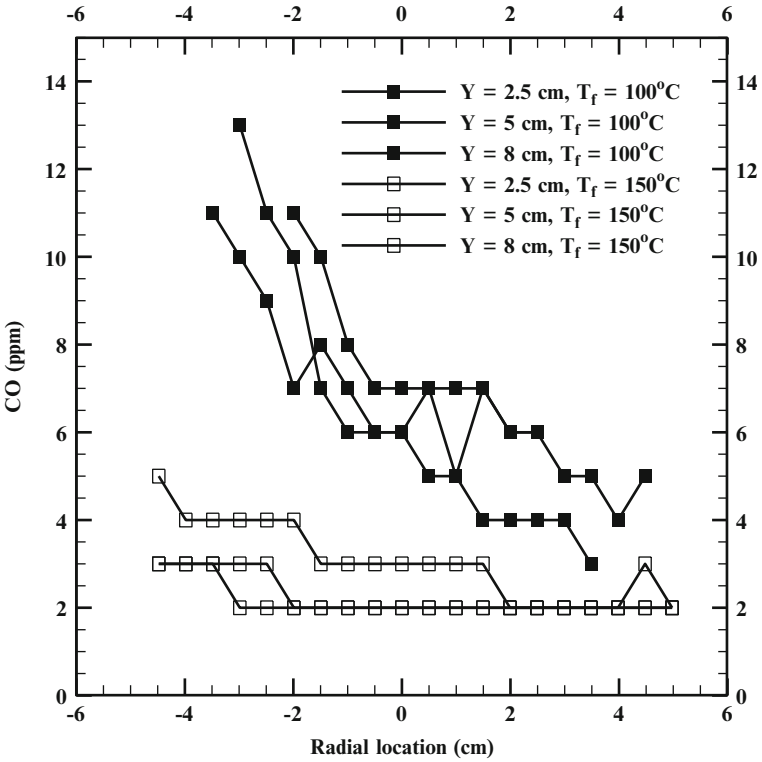



**Fig. 5.44** Radial profiles of CO for enclosed VO flame at 100 and 150 °C

fuel inlet temperature, i.e., VO at 150 °C compared to that of lower fuel inlet temperatures. As the axial distance increases the droplet distribution is observed to be narrower and also more number of smaller droplets is observed for both the cases. Comparatively the higher fuel inlet temperature shows improved atomization with higher percentage of smaller drops compared to that of lower fuel inlet temperature.

Effect of Enclosed Spray on Combustion Emissions

The emissions of CO and NOx were measured for three different axial locations in the flame. The radial profiles are plotted for axial locations at $Y = 2.5$, 5, and 8 cm. The combustor enclosure is of a regular pentagonal shape of about 15 cm measured diagonally. Figure 5.44 shows the CO profile for VO at 100 and 150 °C at $Y = 2.5$, 5, and 8 cm at the combustor exit plane. The CO emissions for VO at 150 °C range from 2 to 5 ppm while those of VO at 100 °C range from 4 to 12 ppm. The CO emissions at the combustor exit plane indicate that sufficient flow mixing has taken place, forming a homogenous product gas mixture. The CO emissions show a

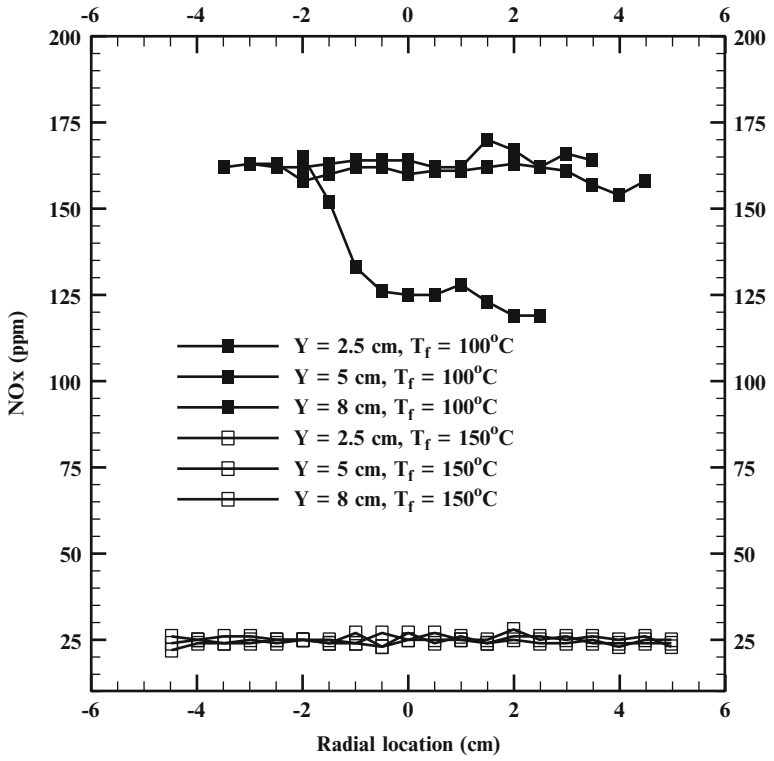



**Fig. 5.45** Radial profiles of NOx for enclosed VO flame at 100 and 150 °C

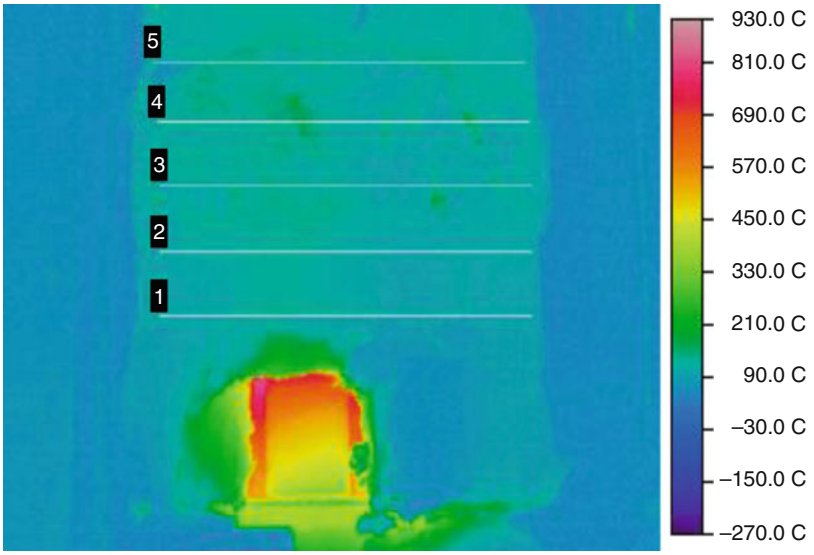
minor decreasing trend in axial direction but mostly they are in similar range. The decrease in emissions for VO at 150 °C is about 8 ppm compared to VO at 100 °C, which is attributed to the smaller droplet diameter for flame with higher fuel inlet temperature.

Figure 5.45 shows the NOx emissions profile at the aforementioned axial locations in the flame. A significant reduction in NOx emissions is observed for VO at 150 °C compared to VO at 100 °C. The NOx emissions for VO at 150 °C are constant at about 25 ppm, where as for VO at 100 °C, the NOx emissions are significantly higher, in the range of 120–160 ppm. The NOx emissions for different fuel temperatures are shown in chapter 3, which show a similar trend of emissions for a given ALR, i.e., for VO at 99 °C, the NOx emissions range from 130 to 155 ppm. An increase in fuel inlet temperature decreases the kinematic viscosity of the fuel hence improving the atomization process, having significant effect on emissions. Spray droplet characteristics cannot be directly compared to the emissions, since the emissions data were taken at the axial locations farther downstream near the combustor exit plane. Most of the droplets were consumed and vaporized in the flame and the PDPA system could not detect the droplets, because of the low data rate after $Y = 40$ mm. Near the injector exit, the CO emissions could

not be measured because it was not possible to quench the reactions consuming CO. The reactions near the injector exit are supposed to be incomplete, due to continuous combustion and the emissions probing within the flame should give detectable CO emissions. However in our case the probe was exposed to a high temperature environment, which allowed the reactions to proceed further within the sampling probe. Since we did not use a water-cooled sampling probe, the measurement of CO emissions was not possible within the flame.

Enclosure Exterior Surface Temperature Distributions

Figure 5.46 shows the infrared camera image of the enclosure. The exterior surface temperature data is inferred from the thermal radiation captured by the infrared image as shown in Fig. 5.45. Image reveals the transverse profiles of temperature across the enclosure surface at 5 different axial locations. The surface temperature profile shows lower temperature for points 1 and 2 and higher for the points 3 and




| | Axial location (mm) | Min C | Max C | Avg C | Range C |
|---|---|---|---|---|---|
| Line - 1 | 166.00 | 82.3 | 146.5 | 113.3 | 64.2 |
| Line - 2 | 222.00 | 95.3 | 149.8 | 119.0 | 54.5 |
| Line - 3 | 280.00 | 89.2 | 156.5 | 125.7 | 67.3 |
| Line - 4 | 336.00 | 90.1 | 180.2 | 131.1 | 90.1 |
| Line - 5 | 388.00 | 80.4 | 153.9 | 127.0 | 73.5 |

**Fig. 5.46** IR image of the enclosure surface temperature

4 and again lower for point 5. This decrease can be observed as point 5 is exposed to exit of the combustor close to the exhaust ambient air while for points 1 and 2 the heat loss is seen to occur due to the uninsulated glass area. Point 4 has the higher average temperature of about 132 °C which indicated that most of the heat is retained in the flame zone at that axial location. The average temperature as seen from the figure varies from a minimum of 113–127 °C. The minimum temperature observed on each axial location would in fact be higher than the indicated because of variation in view factor, since the enclosure is not a flat surface facing normal to the viewing direction. The principle on which IR camera works is based on the radiosity readings, and hence the view factor, emissivity of the object in the field of view would affect the temperature readings. The IR image provided with an idea on how the exterior surface temperature varied. The red contour region in the figure shows the highest temperature which is for the steel bars used to support the glass in the enclosure. Again a significant heat loss is there due to the surface exposed to the ambient air and also the small window kept uninsulated for optical access. However, the insulation helped retaining heat with minimum heat loss through the surface and thus allowing minimum supply of methane to achieve a stable VO flame throughout the experimental duration. The temperature profile at five different axial locations is shown in Fig. 5.47.

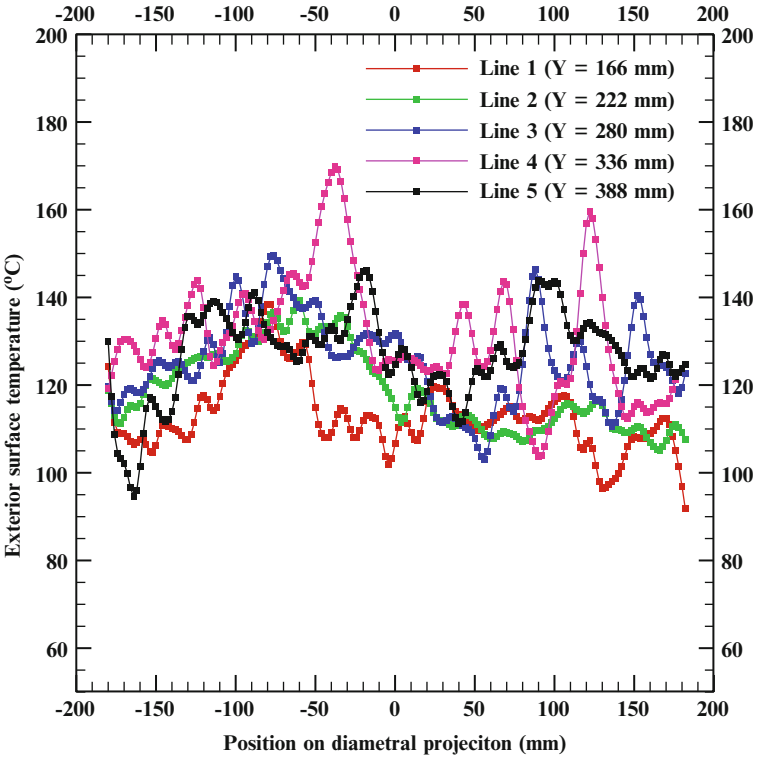



**Fig. 5.47** Profile plot for exterior surface temperature at different axial location on the enclosure

## 5.2 Conclusions

The present study investigates the spray characteristics of heated and unheated vegetable oil. The measurements of mean axial velocity, RMS axial velocity, SMD, and droplet diameter distributions are presented to explain the effect of swirling air on a non-evaporating spray, effect of flame on spray, and effect of enclosure on spray.

- The mean axial velocity for the swirling air flow is observed to be higher than the non-swirling air flow. The axial velocity further decrease in the farther downstream locations in the spray. The swirling spray shows higher axial velocity in the near field region of the spray; whereas for far field regions the non-swirling spray shows the axial velocity profiles with higher magnitudes. Similar trend is observed for RMS axial velocity profiles; higher RMS axial velocity near the injector exit compared to farther downstream locations in the spray. And higher RMS axial velocity in the center of the spray while lower RMS axial velocity at the edge of the spray.
- Swirling air flow in non-evaporating spray lowers the SMD. The reduction is significantly in the far field locations of the spray, i.e., $Y = 25$ mm and higher. In the near field location of the spray, i.e., $Y = 5$–$20$ mm, the swirling air flow tends to increase the radial extent of the spray hence further disintegrating the droplets. The effect of swirling air flow is more prominent near the injector exit while the spray flattens out farther downstream, overall improving the atomization compared to the non-swirling air flow.
- Higher flame temperatures tend to accelerate the droplets leading into increase in mean axial velocity at all the measured axial locations compared to the cold spray. The mean axial velocity peaked at the center while they showed a decreasing trend at the outer edges of the spray. RMS axial velocity increases for flame spray compared to cold spray. The droplet diameter is observed to be smaller in case of flame spray compared to the cold spray. SMD significantly reduces with axial distance. A decrease of about 40–50 μm is observed in the SMD for the flame spray compared to cold spray with swirl air flow. This result is attributed to the improved vaporization because of high flame temperatures versus cold spray.
- The effect of fuel inlet temperature improves the fuel kinematic viscosity and surface tension hence improving the atomization leading to reduction in SMD. This result is observed in both cold spray and also for the flame spray with heated and unheated VO. Increase in mean and RMS axial velocity is observed for higher VO temperature compared to unheated VO. The greater number of larger droplets in case of unheated VO dominates the higher droplet diameter compared to VO at 100 °C. Higher fuel inlet temperatures aids to improve droplet vaporization and hence combustion reactions thus improving the overall spray characteristics .This further improves chemical reactions thus reducing the emissions due to the thermal feedback from the flame delineating the effect of heat release rate on the fuel vaporization.

- The effect of enclosed flame was seen on two different fuel inlet temperatures with higher fuel inlet temperature producing finer droplets lowering the SMD, increasing the mean and RMS axial velocity. The insulated enclosure helped to provide heat feedback which enabled the PDPA measurements avoiding any fuel condensation in the measurement area. The confinement of the flame due to enclosure reduced the radial spread compared to open flame conditions. There was no significant difference in the mean axial velocities and SMD values.
- Based on this study we conclude that the smaller drop size distribution results in lower emissions of CO and NOx. Larger droplets tend to burn in diffusion mode compared to finer drops that lead to premix mode of combustion. Larger droplets that are burning in diffusion mode results into higher local temperatures and higher flame temperatures result in increase in NOx and the more likely chances of fuel pyrolysis, fuel coking problems, and fuel decomposition resulting in higher CO emissions.
- It can be seen from the present investigation, that the flame environment enhances the reduction of droplet size and smaller droplet size distribution, significantly altering the mixing and entrainment due to high flame temperature, as well as faster fuel vaporization as a secondary effect of high temperatures consuming larger droplets.
- A significant change is not observed in the SMD and axial velocity for open and enclosed flame because of the addition of large amount of methane to sustain a stable VO flame in open conditions which results, in significant heat feedback in the near field region of the flame spray, thus neutralizing the influence of the insulated enclosure, effect that is observed in other cases.

## References

Chigier, N. A., & McCreath, C. G. (1974). Combustion of droplets in sprays. In *Fourth International Colloquium on Gas Dynamics of Explosions and Reactive Systems* (1973). San Diego: University of California; *Acta Astronautica*, 1, 687–710.

Kneer, R., Willmann, M., Zeitler, R., Wittig, S., & Collin, K. H. (1994). Phase Doppler particle sizing inside a gas turbine combustor. *International Journal of Energetic Materials and Chemical Propulsion, 3*(1), 301–312.

Liu, A. B., & Reitz, R. D. (1993). Mechanisms of air-assisted liquid atomization. *Atomization and Sprays, 3*, 55–75.

McDonell, V. G., & Samuelsen, S. (1991). Gas and drop behavior in reacting and non-reacting airblast atomizer sprays. *Journal of Propulsion and Power, 7*, 684–691.

Onuma, Y., & Ogasawara, M. (1975). Studies on the structure of a spray combustion flame. In Symposium (International) on Combustion (Vol. 15, No. 1, pp. 453–465).

# Part II
# Vibrations Applications

# Chapter 6
# An Active Vibration Control Strategy for Simply Supported Beams

**L. Dai, L. Sun, and X. Wang**

**Abstract** An active control strategy based on the fuzzy sliding mode control (FSMC) is developed in this chapter for controlling large-amplitude chaotic vibrations of an Euler–Bernoulli beam with simply supported boundary conditions. The geometric nonlinearity of the simply supported beam is considered, and a sinusoidal external excitation is applied on the beam. The Hamilton's principle is applied to establish the nonlinear equations of motion of the beam, and a third-order Galerkin method is then implemented to discretize the nonlinear governing equation from the equations of motion. Corresponding to the nonlinear three-dimensional system via the Galerkin method, an active control strategy is proposed for suppressing the vibrations of the system. A comparison between the vibrations of a single-dimensional system and a three-dimensional system is performed. As found in the chapter, the higher order vibrations have significant influence on the entire vibration of the beam and must be considered in controlling the vibrations of the beam. The proposed active control strategy shows effectiveness and applicability in controlling the beam's chaotic vibrations of large amplitude.

**Keywords** Euler–Bernoulli beam • Active control strategy • Nonlinear multi-dimensional dynamic system • Chaotic vibration

## Key Symbols

| | |
|---|---|
| $l$ | The length of the beam |
| $h$ | The thickness of the beam |
| $b$ | The width of the beam |
| $x$ | The horizontal axis |
| $u$ | The displacement along the $x$-axis |
| $z$ | The vertical axis |
| $w$ | The displacement along the $z$-axis |

L. Dai (✉) • L. Sun • X. Wang
Industrial Systems Engineering, University of Regina, Regina, SK S4S 0A2, Canada
e-mail: liming.dai@uregina.ca



R.N. Jazar, L. Dai (eds.), *Nonlinear Approaches in Engineering Applications*,
DOI 10.1007/978-3-319-27055-5_6

| | |
|---|---|
| $\overline{r}$ | A position vector of any point $(x, z)$ of the beam without any deformation |
| $\overrightarrow{i}$ | The unit vector of the fixed Cartesian coordinate along the $x$-axis |
| $\overline{k}$ | The unit vector of the fixed Cartesian coordinate along the $z$-axis |
| $\Delta$ | The deformation of the beam |
| $\overline{R}$ | The displacement field of the beam |
| $u_0$ | The displacement component along the $x$-axis |
| $w_0$ | The displacement component along the $z$-axis |
| $\rho$ | The density of the beam |
| $V$ | The volume of the beam |
| $T$ | The kinetic energy of the beam |
| $\varepsilon_{11}$ | The von Karman-type equation of strains of large deflection associated with $\overline{R}$ |
| $U$ | The total strain energy of the beam |
| $Q_{11}$ | The elastic coefficient in the same direction with $\varepsilon_{11}$ |
| $q$ | The distributed periodic external excitation |
| $W$ | The virtual work done by the external excitation force |
| $L$ | The Lagrangian function |
| $\Phi$ | The general form of a nonlinear governing equation |
| $U$ | The control input |
| $\Delta F(w,)$ | The unknown external disturbance |
| $\mathbf{W}$ | The column vector in the Galerkin discretization |
| $w_p$ | The response of a selected point on the beam along the $x$-axis |
| $\kappa$ | The control parameter governing the sliding surface |
| $\gamma$ | The periodicity ratio |
| $NOP$ | The total number of periodic points that are overlapping |
| $n$ | The number of all the points forming the Poincaré map |
| $\xi(K)$ | The number of the points overlapping $K$th point in the Poincaré map |
| $\prod$ | The symbol for multiplication |
| $t_0$ | A given time |
| $T$ | A period of a periodic loading |

## 6.1 Introduction

In the literatures available, the linear and nonlinear dynamics of beams has been extensively investigated in the last decades. The frequency of transverse bending of a flat clamped–clamped beam subjected to an axial force was presented (Bouwstra and Geijselaers 1997), and it has been revealed that there is a significant compliance of the support compared to typical compliance of micro-bridges. An investigation into the response of a resonant micro-beam to an electric actuation was presented (Younis and Nayfeh 2002), and the dynamic responses were conducted and it has been shown that an inaccurate representation of the system nonlinearities may lead to an erroneous prediction of the frequency response. In the next year, an analytical

approach and a reduced order modal was employed in the investigation of the linear and nonlinear behaviors of electrically actuated micro-beam (Younis et al. 2003), and it has been discovered that the numerical models for solving the boundary-value problem may become unstable especially when the pull-in voltage is approached. A nonlinear modal analysis approach based on the invariant manifold method was proposed (Xie et al. 2003), and it has been found that the method proposed can save computation effort without any accuracy loss.

With the investigations conducted into the dynamic response of beams, a focus has been put on the control design to suppress the nonlinear responses of beams. In 1992, a control theory (Utkin 1992), namely the theory of sliding modes, was proposed and it has been pointed out that the control theory is of high efficiency in the control of multi-dimensional systems operating under uncertain conditions. After that, a fuzzy sliding mode control (FSMC), which originated from the traditional sliding mode control (SMC), was proposed and a nonlinear system, which was modeled as a beam, has been presented to demonstrate the effectiveness of the control design in suppressing the nonlinear response of the system (Yau and Kuo 2006; Kuo et al. 2007). The chaotic response of a micro-mechanical resonator with electrostatic forces applied on both sides has been investigated using Melnikov function, and a robust control with both fuzzy and sliding mode control involved has been applied to stabilize the vibration of the target system (Haghighi and Markazi 2009). It should be noticed: in the previous works available, the system investigated is usually the governing equation derived with the 1st-order Galerkin's method. Therefore, a control strategy was proposed for nonlinear vibration control of a laminated beam in multi-dimensional form (Dai and Sun 2012), via the design of multiple controllers corresponding to the multiple vibration modes. However, as can be seen from the paper, the control strategy proposed by the authors is applied with two controllers. Nevertheless, the application of vibration control in real world can only be realized with a single controller instead of controlling each vibration mode. A control strategy utilizing a single controller is therefore needed.

In this chapter, the equations of vibration of a simply supported beam are to be established based on the Hamilton's principle. A multi-dimensional system is to be derived for modeling the vibration of the beam, employing non-dimensionalization and discretization. With the concepts of the FSMC design, an active control strategy will be developed corresponding to the governing equations of the beam. The chaotic motion of the Euler–Bernoulli beam is to be discovered and determined via the PR method by Dai and Singh (1997, 1998), and the effects of each of the vibration modes on the responses of the beam will be emphasized and compared with that of a single-dimensional system. The active control strategy will be applied in controlling the large-amplitude chaotic vibrations of the beam at a selected point via a single controller, rather than multiple controllers in the previous work. The effectiveness of the control strategy will be emphasized in the chapter.

## 6.2 Equations of Motion

The simply supported beam investigated in this chapter is sketched in Fig. 6.1. The governing equations of motion of the beam are to be derived based on the Hamilton's principle. As can be seen from Fig. 6.1, the length of the beam is given as $l$, the width of the beam is $b$, and the thickness of the beam is $h$. The $x$-axis is along the axial direction of the beam. The displacement of any point of beam along the $x$- and $z$-axes are designated with $u$ and $w$.

Starting from the origin of the beam, a position vector, $\vec{r}$, of any point $(x, z)$ of the beam without any deformation is given as,

$$\vec{r} = x\vec{i} + z\vec{k},$$

where $\vec{i}$ and $\vec{k}$ are the unit vectors of the fixed Cartesian coordinate shown in the Fig. 6.1.

The deformation is given in the following,

$$\vec{\Delta} = \left(u_0 - z\frac{\partial w_0}{\partial x}\right)\vec{i} + w_0\vec{k}.$$

Thus, the displacement field of the beam can be derived as,

$$\vec{R} = \vec{r} + \vec{\Delta} = u\vec{i} + w\vec{k} = \left(x + u_0 - z\frac{\partial w_0}{\partial x}\right)\vec{i} + (z + w_0)\vec{k},$$

where $u_0$ and $w_0$ are the displacement components along the $x$- and $z$-directions, respectively, of a point on the beam.

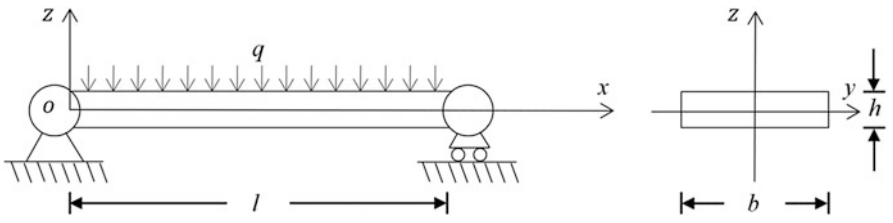



**Fig. 6.1** The sketch of the Euler–Bernoulli beam

Taking the total differentiation of $\vec{R}$ with respect to the time $t$, the following can be obtained,

$$\frac{d\vec{R}}{dt} = \left(\frac{dx}{dt} + \frac{du_0(x,t)}{dt}\right)\vec{i} + \frac{dw_0(x,t)}{dt}\vec{k}.$$

Hence, the kinetic energy of the beam is expressed as,

$$T = \int_V \frac{1}{2}\rho \frac{d\vec{R}}{dt}\frac{d\vec{R}}{dt}dV = b\int_{-\frac{h}{2}}^{+\frac{h}{2}}\int_0^l \frac{1}{2}\rho \frac{d\vec{R}}{dt}\frac{d\vec{R}}{dt}dxdz, \tag{6.1}$$

where $\rho$ denotes the density of the beam.

The von Karman-type equations of strains of large deflection associated with the displacement field, normal to the cross section of the beam along the $x$-direction, can be given by,

$$\varepsilon_{11} = \frac{\partial u_0}{\partial x} + \frac{1}{2}\left(\frac{\partial w_0}{\partial x}\right)^2 - z\frac{\partial^2 w_0}{\partial x^2}.$$

Therefore, the total strain energy of the beam can be given by,

$$U = \int_V \frac{1}{2}\left(Q_{11}\varepsilon_{11}\varepsilon_{11}\right)dV = b\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_0^l \frac{1}{2}\left(Q_{11}\varepsilon_{11}\varepsilon_{11}\right)dxdz, \tag{6.2}$$

where $Q_{11}$ represents the elastic coefficient in the same direction with $\varepsilon_{11}$.

The virtual work done by the external excitation force is given below,

$$W = b\int_0^l qw_0 dx. \tag{6.3}$$

In the following analysis, the Hamilton's principle will be employed to obtain the nonlinear equations of motion for the beam. The mathematical statement of the Hamilton's principle is given by,

$$\int_{t_1}^{t_2}\delta L\, dt + \int_{t_1}^{t_2}\delta W\, dt = 0,$$

where the total Lagrangian function $L$ is given by,

$$L = T - U.$$

Then, the first term in the mathematical statement of the Hamilton's principle can be developed in the following,

$$\begin{aligned}
\int_{t_1}^{t_2}\delta L dt &= \int_{t_1}^{t_2}\delta\left(T-U\right)dt \\
&= \int_{t_1}^{t_2}\int_{V}\rho\frac{d\vec{R}}{dt}\frac{d\delta\vec{R}}{dt}dVdt - \int_{t_1}^{t_2}\int_{V}Q_{11}\varepsilon_{11}\delta\varepsilon_{11}dVdt \\
&= \int_{V}\int_{t_1}^{t_2}\rho\left(\frac{d\vec{R}}{dt}\right)\left(\frac{d\delta\vec{R}}{dt}\right)dtdV - \int_{t_1}^{t_2}\int_{V}Q_{11}\varepsilon_{11}\delta\varepsilon_{11}dVdt \\
&= 0 - \int_{V}\int_{t_1}^{t_2}\rho\frac{d^2\vec{R}}{dt^2}\delta\vec{R}dtdV - \int_{t_1}^{t_2}\int_{V}Q_{11}\varepsilon_{11}\delta\varepsilon_{11}dVdt \\
&= -b\int_{t_1}^{t_2}\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_{0}^{l}\rho\delta\left(u\vec{i}+w\vec{k}\right)\left(\frac{d^2u}{dt^2}\vec{i}+\frac{d^2w}{dt^2}\vec{k}\right)dxdzdt \\
&\quad - b\int_{t_1}^{t_2}\left(\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_{0}^{l}Q_{11}\delta\varepsilon_{11}\varepsilon_{11}dxdz\right)dt \\
&= -b\int_{t_1}^{t_2}\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_{0}^{l}\rho\delta\left(x+u_0-z\frac{\partial w_0}{\partial x}\right)\frac{d^2u}{dt^2}dxdzdt \\
&\quad - b\int_{t_1}^{t_2}\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_{0}^{l}\rho\delta\left(z+w_0\right)\frac{d^2w_0}{dt^2}dxdzdt \\
&\quad - b\int_{t_1}^{t_2}\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_{0}^{l}Q_{11}\delta\left(\frac{\partial u_0}{\partial x}+\frac{1}{2}\left(\frac{\partial w_0}{\partial x}\right)^2-z\frac{\partial^2 w_0}{\partial x^2}\right)\varepsilon_{11}dxdzdt\,, \\
&= L_1+L_2+L_3
\end{aligned}$$

where $L_1$, $L_2$, and $L_3$ are given in the following,

$$L_1 = -b\int_{t_1}^{t_2}\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_{0}^{l}\rho\delta\left(x+u_0-z\frac{\partial w_0}{\partial x}\right)\frac{d^2u}{dt^2}dxdzdt, \tag{6.4a}$$

$$L_2 = -b\int_{t_1}^{t_2}\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_{0}^{l}\rho\delta\left(z+w_0\right)\frac{d^2w_0}{dt^2}dxdzdt, \tag{6.4b}$$

$$L_3 = -b\int_{t_1}^{t_2}\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_{0}^{l}Q_{11}\delta\left(\frac{\partial u_0}{\partial x}+\frac{1}{2}\left(\frac{\partial w_0}{\partial x}\right)^2-z\frac{\partial^2 w_0}{\partial x^2}\right)\varepsilon_{11}dxdzdt\,, \tag{6.4c}$$

Equations (6.4a), (6.4b), and (6.4c) can be further developed as follows,

$$
\begin{aligned}
L_1 &= -b\int_{t_1}^{t_2}\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_0^l \rho\delta\left(x+u_0-z\frac{\partial w_0}{\partial x}\right)\frac{d^2u}{dt^2}dxdzdt \\
&= -b\int_{t_1}^{t_2}\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_0^l \rho\frac{d^2u}{dt^2}\delta u_0 dxdzdt + b\int_{t_1}^{t_2}\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_0^l \rho z\frac{d^2u}{dt^2}\frac{\partial\delta w_0}{\partial x}dxdzdt \\
&= -b\int_{t_1}^{t_2}\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_0^l \rho\frac{d^2u}{dt^2}\delta u_0 dxdzdt + 0 - b\int_{t_1}^{t_2}\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_0^l \rho z\frac{d^2}{dt^2}\left(\frac{\partial u}{\partial x}\right)\delta w_0 dxdzdt \\
&= -b\int_{t_1}^{t_2}\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_0^l \rho\left(\frac{d^2x}{dt^2}+\frac{d^2u_0}{dt^2}-z\frac{d^2}{dt^2}\left(\frac{\partial w_0}{\partial x}\right)\right)\delta u_0 dxdzdt \\
&\quad + 0 - b\int_{t_1}^{t_2}\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_0^l \rho z\frac{d^2}{dt^2}\left[\left(1+\frac{\partial u_0}{\partial x}-z\frac{\partial^2 w_0}{\partial x^2}\right)\right]\delta w_0 dxdzdt \\
&= -b\int_{t_1}^{t_2}\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_0^l \rho\left(\frac{d^2x}{dt^2}+\frac{d^2u_0}{dt^2}+0\right)\delta u_0 dxdzdt \\
&\quad - b\int_{t_1}^{t_2}\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_0^l \left[0+0-\rho z^2\frac{\partial^2}{\partial x^2}\left(\frac{d^2w_0}{dt^2}\right)\right]\delta w_0 dxdzdt \\
&= -b\int_{t_1}^{t_2}\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_0^l \rho\left(\frac{d^2x}{dt^2}+\frac{d^2u_0}{dt^2}\right)\delta u_0 dxdzdt \\
&\quad + b\int_{t_1}^{t_2}\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_0^l \left[\rho z^2\frac{\partial^2}{\partial x^2}\left(\frac{d^2w_0}{dt^2}\right)\right]\delta w_0 dxdzdt, \qquad (6.5a)
\end{aligned}
$$

$$
\begin{aligned}
L_2 &= -b\int_{t_1}^{t_2}\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_0^l \rho\delta\left(z+w_0\right)\frac{d^2w_0}{dt^2}dxdzdt \\
&= -b\int_{t_1}^{t_2}\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_0^l \rho\frac{d^2w_0}{dt^2}\delta w_0 dxdzdt, \qquad (6.5b)
\end{aligned}
$$

$$
\begin{aligned}
L_3 &= -b\int_{t_1}^{t_2}\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_0^l Q_{11}\delta\left(\frac{\partial u_0}{\partial x}+\frac{1}{2}\left(\frac{\partial w_0}{\partial x}\right)^2-z\frac{\partial^2 w_0}{\partial x^2}\right)\varepsilon_{11}dxdzdt \\
&\quad -b\int_{t_1}^{t_2}\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_0^l Q_{11}\left(\frac{\partial \delta u_0}{\partial x}+\frac{\partial w_0}{\partial x}\frac{\partial \delta w_0}{\partial x}-z\frac{\partial^2 \delta w_0}{\partial x^2}\right)\varepsilon_{11}dxdzdt \\
&= -b\int_{t_1}^{t_2}\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_0^l \left(Q_{11}\varepsilon_{11}\frac{\partial \delta u_0}{\partial x}+Q_{11}\varepsilon_{11}\frac{\partial w_0}{\partial x}\frac{\partial \delta w_0}{\partial x}-zQ_{11}\varepsilon_{11}\frac{\partial^2 \delta w_0}{\partial x^2}\right)dxdzdt \\
&= 0+b\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_0^l Q_{11}\frac{\partial \varepsilon_{11}}{\partial x}\delta u_0 dxdz+0+b\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_0^l Q_{11}\frac{\partial}{\partial x}\left(\varepsilon_{11}\frac{\partial w_0}{\partial x}\right)\delta w_0 dxdz \\
&\quad +0+0+b\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_0^l zQ_{11}\frac{\partial^2 \varepsilon_{11}}{\partial x^2}\cdot\delta w_0 dxdz \\
&= +b\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_0^l Q_{11}\left(\frac{\partial^2 u_0}{\partial x^2}+\frac{\partial w_0}{\partial x}\frac{\partial^2 w_0}{\partial x^2}-z\frac{\partial^3 w_0\,(x,t)}{\partial x^3}\right)\delta u_0 dxdz \\
&\quad +b\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_0^l Q_{11}\left(\frac{\partial^2 u_0}{\partial x^2}+\frac{\partial w_0}{\partial x}\frac{\partial^2 w_0}{\partial x^2}-z\frac{\partial^3 w_0}{\partial x^3}\right)\frac{\partial w_0}{\partial x}\delta w_0 dxdz \\
&\quad +b\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_0^l Q_{11}\left(\frac{\partial u_0}{\partial x}+\frac{1}{2}\left(\frac{\partial w_0}{\partial x}\right)^2-z\frac{\partial^2 w_0}{\partial x^2}\right)\frac{\partial^2 w_0}{\partial x^2}\delta w_0 dxdz \\
&+b\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_0^l\left(zQ_{11}\frac{\partial^3 u_0}{\partial x^3}+zQ_{11}\frac{\partial^2 w_0}{\partial x^2}\frac{\partial^2 w_0}{\partial x^2}+zQ_{11}\frac{\partial w_0}{\partial x}\frac{\partial^3 w_0}{\partial x^3}-z^2Q_{11}\frac{\partial^4 w_0}{\partial x^4}\right)\delta w_0 dxdz \\
&= +b\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_0^l Q_{11}\left(\frac{\partial^2 u_0}{\partial x^2}+\frac{\partial w_0}{\partial x}\frac{\partial^2 w_0}{\partial x^2}\right)\delta u_0 dxdz \\
&\quad +b\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_0^l Q_{11}\left(\frac{\partial^2 u_0}{\partial x^2}+\frac{\partial w_0}{\partial x}\frac{\partial^2 w_0}{\partial x^2}\right)\frac{\partial w_0}{\partial x}\delta w_0 dxdz \\
&\quad +b\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_0^l Q_{11}\left(\frac{\partial u_0}{\partial x}+\frac{1}{2}\left(\frac{\partial w_0}{\partial x}\right)^2\right)\frac{\partial^2 w_0}{\partial x^2}\delta w_0 dxdz \\
&\quad -b\int_{-\frac{h}{2}}^{\frac{h}{2}}\int_0^l z^2Q_{11}\frac{\partial^4 w_0}{\partial x^4}\delta w_0 dxdz.
\end{aligned}
\tag{6.5c}
$$

From Eqs. (6.3), (6.5a), (6.5b), and (6.5c), the nonlinear governing equation of the investigated simply supported beam can be derived in the following form,

$$
-\rho h\frac{d^2x}{dt^2}-\rho h\frac{d^2u_0}{dt^2}+Q_{11}h\frac{\partial^2 u_0}{\partial x^2}+Q_{11}h\frac{\partial w_0}{\partial x}\frac{\partial^2 w_0}{\partial x^2}=0, \tag{6.6a}
$$

$$-\rho h\frac{d^2w_0}{dt^2}+\rho\frac{h^3}{12}\frac{\partial^2}{\partial x^2}\left(\frac{d^2w_0}{dt^2}\right)+Q_{11}h\frac{\partial^2u_0}{\partial x^2}\frac{\partial w_0}{\partial x}$$

$$+Q_{11}h\frac{\partial u_0}{\partial x}\frac{\partial^2w_0}{\partial x^2}+Q_{11}h\frac{3}{2}\left(\frac{\partial w_0}{\partial x}\right)^2\frac{\partial^2w_0}{\partial x^2}-\frac{h^3}{12}Q_{11}\frac{\partial^4w_0}{\partial x^4}+q=0. \tag{6.6b}$$

Based on the reference (Younis and Nayfeh 2002) and Eq. (6.6a), it can be obtained as

$$\frac{\partial u_0}{\partial x}=-\frac{1}{2}\left(\frac{\partial w_0}{\partial x}\right)^2+\frac{1}{2l}\int_0^l\left(\frac{\partial w_0}{\partial x}\right)^2dx+\frac{\rho}{Q_{11}}\frac{d^2x}{dt^2}\left(x-\frac{l}{2}\right). \tag{6.7}$$

With the substitution of Eq. (6.7) into Eq. (6.6b), the nonlinear differential governing equation of the beam in $z$-direction is derived as

$$-\rho\frac{d^2w_0}{dt^2}+\int_0^l\frac{1}{2l}Q_{11}\left(\frac{\partial w_0}{\partial x}\right)^2dx\frac{\partial^2w_0}{\partial x^2}=0. \tag{6.8}$$

To validate the governing equation Eq. (6.8) and facilitate the numerical simulations in the consequent sections, the following non-dimensional variables are introduced,

$$\overline{t}=t\sqrt{\frac{\frac{1}{12}bh^3Q_{11}}{I_0bl^4}},\quad \overline{x}=\frac{x}{l},\quad \overline{w}_0=\frac{w_0}{h}. \tag{6.9}$$

Introducing the non-dimensional variables shown in Eq. (6.9) into Eq. (6.8), the non-dimensional governing equation of the investigated nonlinear simply supported beam can be expressed as,

$$-\frac{d^2\overline{w}_0}{d\overline{t}^2}+A\frac{\partial^2}{\partial\overline{x}^2}\left(\frac{d^2\overline{w}_0}{d\overline{t}^2}\right)+B\int_0^1\left(\frac{\partial\overline{w}_0}{\partial\overline{x}}\right)^2d\overline{x}\frac{\partial^2\overline{w}_0}{\partial\overline{x}^2}-C\frac{\partial^4\overline{w}_0}{\partial x^4}+\overline{q}=0. \tag{6.10}$$

where

$$A=\frac{h^2}{12l^2},\quad B=\frac{Q_{11}h^2}{2l^4\rho\tau^2},\quad C=\frac{h^2Q_{11}}{12l^4\rho\tau^2},\quad \overline{q}=\frac{q}{\rho h^2\tau^2}.$$

## 6.3 Series Solutions

Based on the Galerkin method of discretization, the transverse displacement $\overline{w}_0$ is expanded in a series form, in terms of a set of comparison functions as,

$$\overline{w}_0=\sum_{n=1}^{\infty}\phi_n(\overline{x})\,\overline{w}_n(\overline{t}). \tag{6.11}$$

Corresponding to the boundary conditions of the Euler–Bernoulli beam, $\phi_n(\overline{x})$ can be given as follows,

$$\phi_n(\overline{x}) = \sin(n\pi\overline{x}). \tag{6.12}$$

Substitute the series solution of Eq. (6.12) into Eq. (6.11), and to assist the following presentation, replace $\phi_n$, $w_n$, $_n$, $\ddot{w}_n$, and $q$ for $\phi_n(\overline{x})$, $\overline{w}_n(\overline{t})$, $\frac{d\overline{w}_n}{d\overline{t}}$, $\frac{d^2\overline{w}_n}{d\overline{t}^2}$, and $\overline{q}$, respectively, and

$$w_1 = w_{1,1}, w_2 = w_{2,1}, w_3 = w_{3,1}$$

$$\dot{w}_{1,1} = w_{1,2} \dot{w}_{2,1} = w_{2,2}, \dot{w}_{3,1} = w_{3,2}.$$

Therefore, with the application of the Galerkin method at $n = 1$ and $n = 3$, the discretized governing equations of the simply supported beam can be obtained in the following,

$$\begin{cases} \dot{w}_{1,1} = w_{1,2} \\ \dot{w}_{1,2} = \eta_0(\chi_0 + \zeta_0) \end{cases}, \tag{6.13a}$$

$$\begin{cases} \dot{w}_{1,1} = w_{1,2} \\ \dot{w}_{1,2} = \eta_1(\chi_1 + \zeta_1) \\ \dot{w}_{2,1} = w_{2,2} \\ \dot{w}_{2,2} = \eta_2(\chi_2 + \zeta_2) \\ \dot{w}_{3,1} = w_{3,2} \\ \dot{w}_{3,2} = \eta_3(\chi_3 + \zeta_3) \end{cases}, \tag{6.13b}$$

where,

$$\eta_0 = -\frac{1}{\left(-\frac{1}{2} - \frac{1}{2}\pi^2 A\right)},$$

$$\chi_0 = -\frac{1}{2}\pi^4 C w_{1,1} - \frac{1}{4}\pi^4 B w_{1,1}^3,$$

$$\zeta_0 = \frac{2q}{\pi},$$

$$\eta_1 = -\frac{1}{\left(-\frac{1}{2} - \frac{1}{2}\pi^2 A\right)}, \eta_2 = -\frac{1}{\left(-\frac{1}{2} - 2\pi^2 A\right)}, \eta_3 = -\frac{1}{\left(-\frac{1}{2} - \frac{9}{2}\pi^2 A\right)}$$

$$\chi_1 = -\pi^4 B w_{1,1} w_{2,1}^2 - \frac{1}{2}\pi^4 C w_{1,1} - \frac{9}{4}\pi^4 B w_{1,1} w_{3,1}^2 - \frac{1}{4}\pi^4 B w_{1,1}^3,$$

$$\chi_2 = -4\pi^4 B w_{2,1}^3 - 8\pi^4 C w_{2,1} - 9\pi^4 B w_{2,1} w_{3,1}^2 - \pi^4 B w_{2,1} w_{1,1}^2,$$

$$\chi_3 = -9\pi^4 B w_{3,1} w_{2,1}^2 - \frac{81}{2}\pi^4 C w_{3,1} - \frac{81}{4}\pi^4 B w_{3,1}^3 - \frac{9}{4}\pi^4 B w_{3,1} w_{1,1}^2,$$

$$\zeta_1 = \frac{2q}{\pi}, \zeta_2 = 0, \zeta_3 = \frac{2q}{3\pi}.$$

## 6.4 Control Strategy Design

With the developed governing equations, boundary conditions, and the solutions of the governing equations, an active vibration control strategy can be developed. Based on the previous works (Yau and Kuo 2006; Kuo et al. 2007), the proposed active control strategy is developed for the vibration control of a multi-dimensional nonlinear dynamic system shown in Eq. (6.13).

For a nonlinear governing equation in the following general form,

$$\ddot{w} = \Phi\left(w, \dot{w}, t\right), \tag{6.14}$$

If $U$ is given as the control input and $\Delta F(w,)$ as the unknown external disturbance applying on beam, the governing equation Eq. (6.14) for the nonlinear Euler–Bernoulli beam with the control input and the external disturbance can be given by,

$$\ddot{w} = \Phi\left(w, \dot{w}, t\right) + U + \Delta F\left(w, \dot{w}\right). \tag{6.15}$$

With the application of the control, it is expected that the vibration of the beam can be controlled.

If the $n$th Galerkin method is applied in the discretization of the governing equation given in Eq. (6.15), a series of 2nd-order ordinary differential equations considering the control input $U$ and the unknown external disturbance $\Delta F(w,)$ will be derived as follows,

$$\begin{cases} \dot{w}_{1,1} = w_{1,2} \\ \dot{w}_{1,2} = \phi_1\left(\mathbf{W}, t\right) + u_1 + \Delta f_1\left(\mathbf{W}, t\right) \\ \dot{w}_{2,1} = w_{2,2} \\ \dot{w}_{2,2} = \phi_2\left(\mathbf{W}, t\right) + u_2 + \Delta f_2\left(\mathbf{W}, t\right) \\ \vdots \\ \dot{w}_{i,1} = w_{i,2} \\ \dot{w}_{i,2} = \phi_i\left(\mathbf{W}, t\right) + u_i + \Delta f_i\left(\mathbf{W}, t\right) \\ \vdots \\ \dot{w}_{n,1} = w_{n,2} \\ \dot{w}_{n,2} = \phi_n\left(\mathbf{W}, t\right) + u_n + \Delta f_i\left(\mathbf{W}, t\right), \end{cases} \tag{6.16}$$

where $\varphi_i(\mathbf{W}, t)$, $u_i$, and $\Delta f_i(\mathbf{W}, t)$ represent the expressions of $\Phi(w, , t)$, $U$, and $\Delta F(w, )$ after the application of the Galerkin discretization. With the Galerkin discretization, the column vector $\mathbf{W}$ in Eq. (6.16) is given below,

$$\mathbf{W} = [w_{1,1}\ w_{1,2}\ w_{2,1}\ w_{2,2}\ \cdots\ w_{i,1}\ w_{i,2}\ \cdots\ w_{n,1}\ w_{n,2}]^T.$$

Considering the response of a point on the beam along the $x$-direction, based on Eq. (6.16) and the expression in Eq. (6.11), the non-dimensional response of the selected point $w_p$ can be given as,

$$w_p = \sum_{n=1}^{\infty} \phi_n \left(x_p\right) w_n(t), \tag{6.17}$$

where $x_p$ denotes the location of the selected point.

For a desired reference vibration expressed as,

$$w_r = \Psi(t), \tag{6.18}$$

the control input $U$ can be given as,

$$U = U_{eq} - U_r, \tag{6.19}$$

where $U_{eq}$ and $U_r$ are expressible as,

$$U_{eq} = -\left(\left(\dot{w}_p - \dot{\Psi}\right) + \kappa \left(w_p - \Psi\right)\right), U_r = k_{fs} U_{fs}. \tag{6.20}$$

In Eq. (6.20), $\kappa$ designates the control parameter governing the sliding surface, $k_{fs}$ is given as $|\Delta F\left(w, \dot{w}\right)| < k_{fs} \in R^+$, and the value of $U_{fs}$ depends on the fuzzy rule shown in Table 6.1, in which PB, PM, PS, ZE, NS, NM, and NB represent 1, 2/3, 1/3, 0, −1/3, −2/3, and −1, respectively.

With the control strategy demonstrated in Eqs. (6.15)–(6.20), the vibration control of the beam expressed with the general governing equation Eq. (6.14) can be realized.

**Table 6.1** The fuzzy rule of $U_{fs}$

| $U_{fs}$ | | $U_{eq}$ | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | PB | PM | PS | ZE | NS | NM | NB |
| $\frac{dU_{eq}}{dt}$ | PB | NB | NB | NB | NB | NM | NS | ZE |
| | PM | NB | NB | NB | NM | NS | ZE | PS |
| | PS | NB | NB | NM | NS | ZE | PS | PM |
| | ZE | NB | NM | NS | ZE | PS | PM | PB |
| | NS | NM | NS | ZE | PS | PM | PB | PB |
| | NM | NS | ZE | PS | PM | PB | PB | PB |
| | NB | ZE | PS | PM | PB | PB | PB | PB |

Take the beam governed by Eq. (6.10) as an example. Use the control strategy developed and apply the control input as shown in Eq. (6.15), the governing equation with the control input for the beam can be given by the following expression,

$$-\frac{d^2\overline{w}_0}{d\overline{t}^2}+A\frac{\partial^2}{\partial\overline{x}^2}\left(\frac{d^2\overline{w}_0}{d\overline{t}^2}\right)+B\int_0^1\left(\frac{\partial\overline{w}_0}{\partial\overline{x}}\right)^2 d\overline{x}\frac{\partial^2\overline{w}_0}{\partial\overline{x}^2}$$

$$-C\frac{\partial^4\overline{w}_0}{\partial x^4}+\overline{q}-U-\Delta F\left(w,\dot{w}\right)=0. \tag{6.21}$$

With the application of the 3rd-order Galerkin method, Eq. (6.21) may have the following form,

$$\begin{cases}\dot{w}_{1,1}=w_{1,2}\\ \dot{w}_{1,2}=\eta_1\left(\chi_1+\zeta_1\right)+u_1+\Delta f_1\left(\mathbf{W},t\right)\\ \dot{w}_{2,1}=w_{2,2}\\ \dot{w}_{2,2}=\eta_2\left(\chi_2+\zeta_2\right)+u_2+\Delta f_2\left(\mathbf{W},t\right)\\ \dot{w}_{3,1}=w_{3,2}\\ \dot{w}_{3,2}=\eta_3\left(\chi_3+\zeta_3\right)+u_3+\Delta f_3\left(\mathbf{W},t\right),\end{cases} \tag{6.22}$$

where $u_1$, $u_2$, and $u_3$ are derived as follows through the 3rd-order Galerkin method,

$$u_1=0.83U, u_2=0\cdot U, u_3=0.36U.$$

In the next section, it will be demonstrated in the numerical simulation that the actual response of the beam at a selected point can be well synchronized with a desired reference signal in the case that $u_2$ equals to zero.

## 6.5 Nonlinear Response Characterization

The responses of a selected point on the simply supported beam considered are investigated in this section. With the numerical simulations performed, chaotic vibrations of the selected point are discovered, and the comparison is provided between the vibrations of the established single-dimensional dynamic system and the multi-dimensional dynamic system. To facilitate efficient numerical calculation in this chapter, the fourth-order P-T method (Dai 2008) is implemented.

It should be noticed, to evaluate the nonlinear characteristics of the system considered, a characteristic diagnosing method named Periodicity Ratio (P-R) method is employed. The P-R method developed by Dai and Singh (1997, 1998) has shown great advantages in diagnosing the nonlinear behavior such as periodic, quasi-periodic, and chaotic responses for a nonlinear system. The P-R criterion is applied based on Poincaré maps of the nonlinear system considered. The P-R value

is a criterion for analyzing the nonlinear dynamic behavior with the considerations of the overlapping points in comparing with the total number of points in the Poincare map. The Periodicity Ratio (P-R) $\gamma$ is defined as:

$$\gamma = \lim_{n\to\infty} \frac{NOP}{n}.$$

*NOP* in the equation denotes the total number of periodic points which are overlapping points and $n$ represents the number of all the points forming the Poincaré map. *NOP* can be obtained by the formula shown below:

$$NOP = \xi(1) + \sum_{k=2}^{n} \xi(K) P \left( \prod_{L=1}^{K-1} \left\{ X_{\ KL} + \dot{X}_{KL} \right\} \right)$$

In the equation above, $\xi(K)$ represents the number of points overlapping the $K$th point in the Poincaré map, and $\prod$ is the symbol for multiplication. $X_{KL}$,$_{KL}$, and $P$ are functions expressed as,

$$X_{KL} = \left| X\left(t_0 + KT\right) - X\left(t_0 + LT\right) \right|,$$

$$\dot{X}_{KL} = \left| \dot{X}\left(t_0 + KT\right) - \dot{X}\left(t_0 + LT\right) \right|,$$

$$P(\vartheta) = \begin{cases} 0 & if\ \vartheta = 0 \\ 1 & if\ \vartheta \neq 0 \end{cases}.$$

where, $t_0$ is a given time, and $T$ is a period of a periodic loading.

With the criterion $\gamma$ such defined, if the vibration is perfectly periodic, $\gamma$ equals one; if value $\gamma$ approaches zero, the vibration is then quasi-periodic or chaotic. When $\gamma$ falls between 0 and 1, theoretically, the vibration is neither periodic nor chaotic. With this single value criterion, the dynamic behavior of a nonlinear dynamic system can be conveniently characterized.

The parameters of the beam are given as those from the work,

$$Q_{11} = 127 \times 10^9\ \text{Pa}, 1 = 2\ \text{m}, \rho = 7800\ \text{kg/m}^3,$$

and the non-dimensional external excitation is

$$q = 5.58 \sin(5.90t).$$

The non-dimensionalized initial conditions, corresponding to the displacements described by Eq. (6.13) after the implementation of the 3rd-order Galerkin method, are taken as,

$$w_{1,1}(0) = 0.01, w_{1,2}(0) = 0.05, w_{2,1}(0) = 0.005, w_{2,2}(0) = 0.025$$

$$w_{3,1}(0) = 0.003, w_{3,2}(0) = 0.02.$$

$w_p$, the transverse displacement at a selected point $p$ located at 1.8 m from the origin of the Cartesian coordinates shown in Fig. 6.1, is expressed corresponding to Eq. (6.13) as below,

$$w_* = \sum_{n=1}^{1} \phi_n(x_p) w_{n,1} = 0.309 w_{1,1}, \tag{6.23a}$$

$$w_p = \sum_{n=1}^{3} \phi_n(x_p) w_{n,1} = 0.309 w_{1,1} - 0.588 w_{2,1} + 0.809 w_{3,1}. \tag{6.23b}$$

A chaotic response of the beam occurs as shown in Fig. 6.2 for $w_*$ and Fig. 6.3 for $w_p$, while the developed control strategy is not applied. The wave diagram, 2-D phase diagram, and Poincaré map are shown, respectively, in Figs. 6.2a and 6.3a, Figs. 6.2b and 6.3b, and Figs. 6.2c and 6.3c. With the utilization of the P-R method, the typical chaotic cases shown in Figs. 6.2 and 6.3 can be determined with the P-R value approaching to 0 in both cases.

In Figs. 6.2 and 6.3, one may notice in Fig. 6.2a the maximum amplitude of the chaotic response of the beam is about 1.35, and $w_*$ is derived from Eq. (6.13a), which is obtained via the first-order Galerkin method. However, in Fig. 6.3a it can be discovered that the maximum amplitude is close to 1.85, which is much larger than the one in Fig. 6.2a by almost 37 %.

Besides, from Fig. 6.4a–c, it can be learned: although the first vibration mode in Fig. 6.4a is larger than those of the other two vibration modes as shown in Fig. 6.4b, c, the amplitudes of the other two vibration modes are obviously not negligible. Actually it can be found that the other two vibration modes also significantly contribute to the actual response of the selected point.

Thus, the development of a multi-dimensional dynamic system is necessary for the accurate prediction of the dynamics of the simply supported beam subjected to a sinusoidal external excitation.

## 6.6 Chaotic Vibration Control

### *6.6.1* $\kappa = 100, k_{fs} = 5000$

Considering that the displacement shown in Fig. 6.3a is non-dimensional and the amplitude is actually 1.85 times the thickness of the beam, the maximum amplitude showing in Fig. 6.3a is very large. Therefore, the large-amplitude chaotic vibration

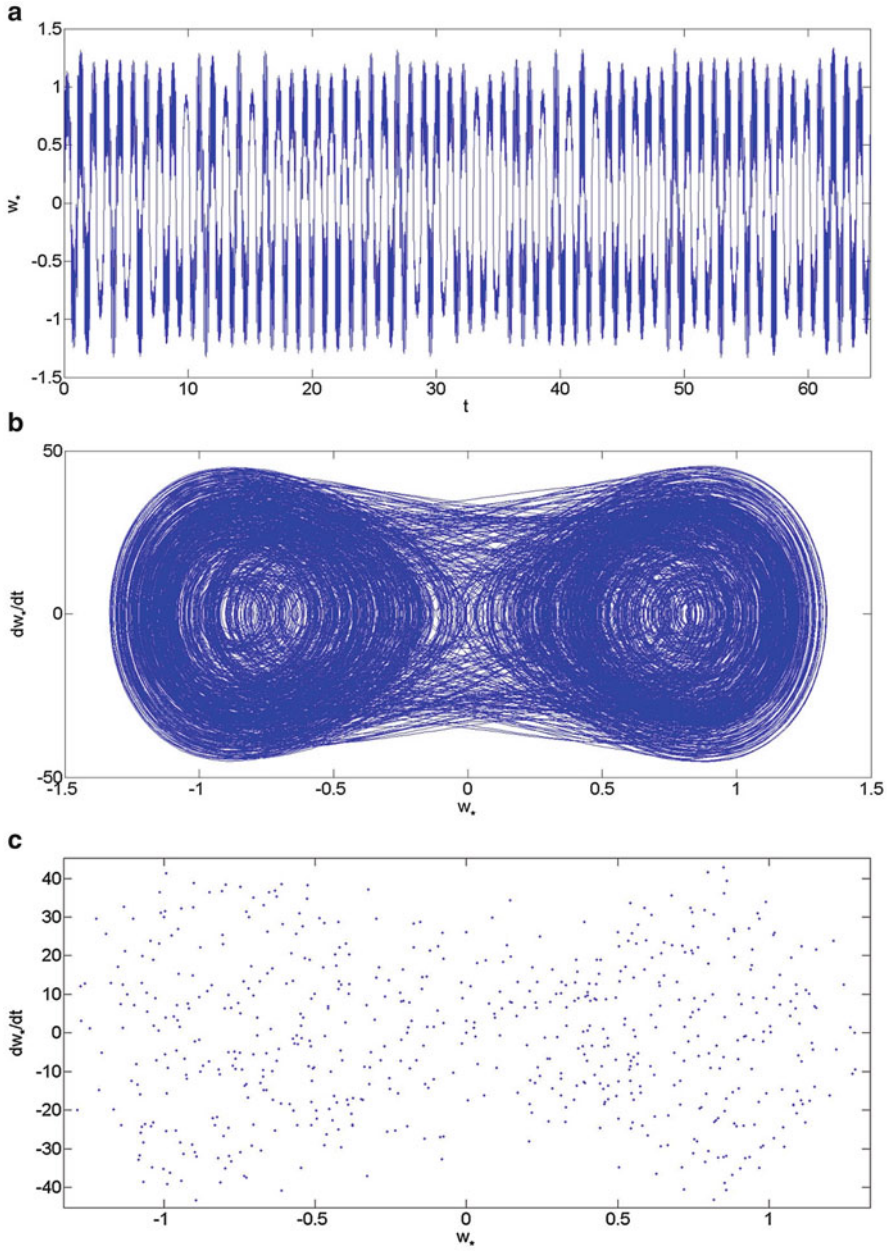



**Fig. 6.2** (**a**) The wave diagram of $w_*$ derived from Eq. (6.23a). (**b**) The 2-D phase diagram of $w_*$ derived from Eq. (6.23a). (**c**) The Poincaré map of $w_*$ derived from Eq. (6.23a)

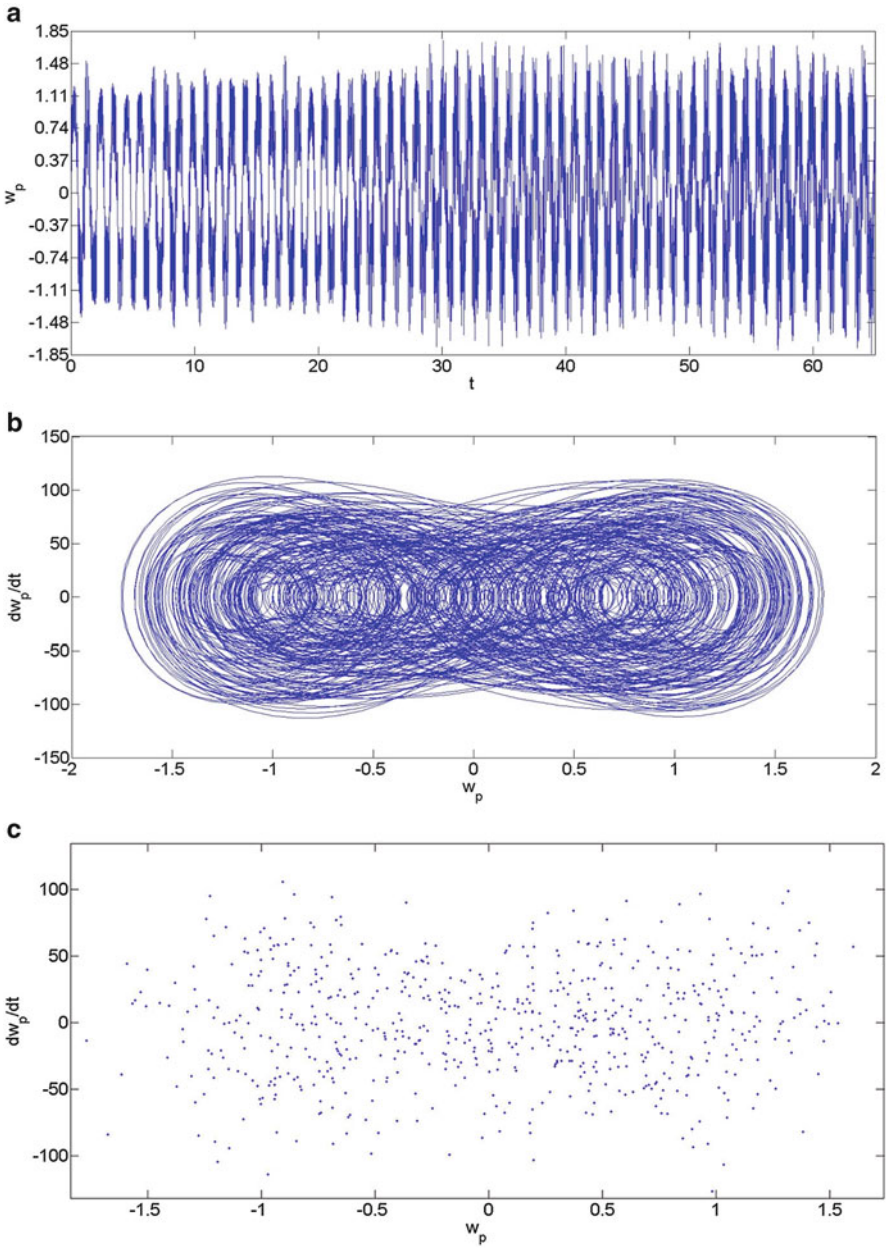



**Fig. 6.3** (**a**) The wave diagram of $w_p$ derived from Eq. (6.23b). (**b**) The 2-D phase diagram of $w_p$ derived from Eq. (6.23b). (**c**) The Poincaré map of $w_p$ derived from Eq. (6.23b)

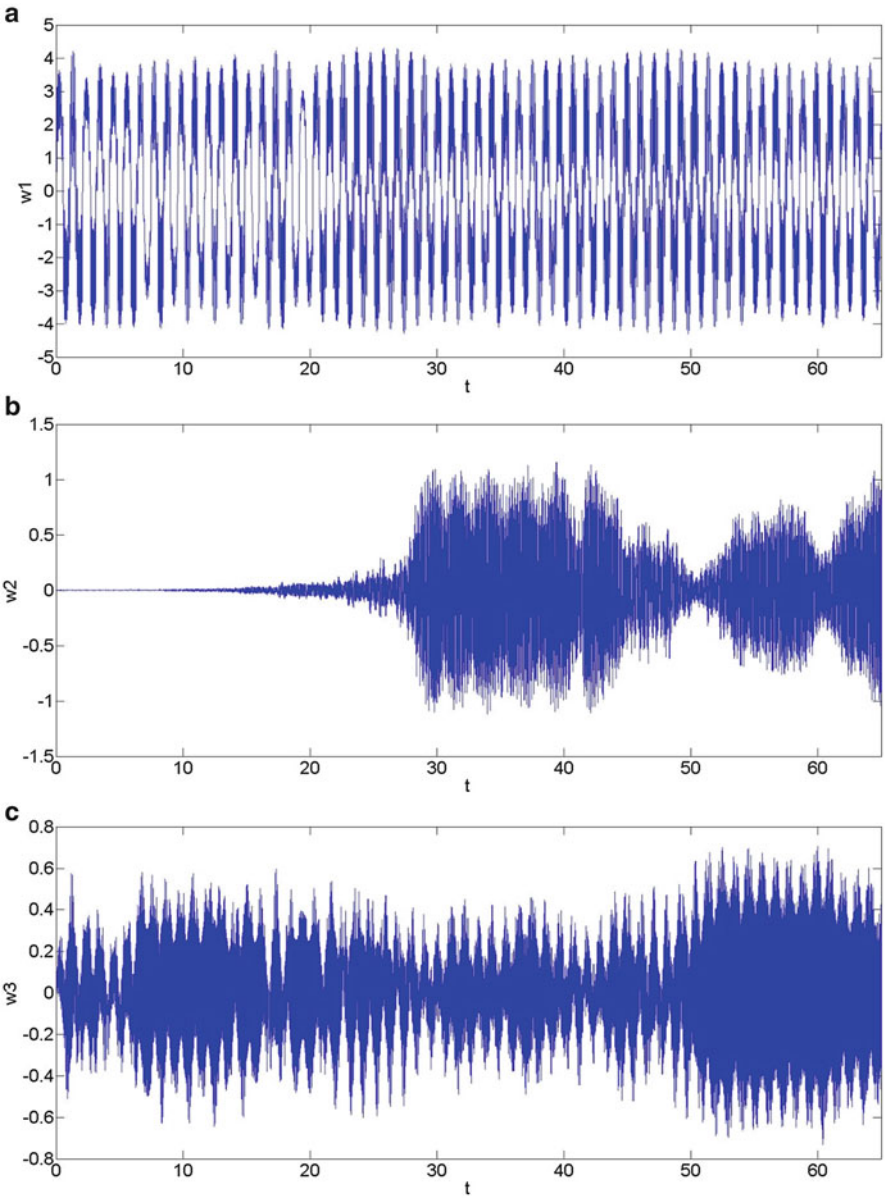



**Fig. 6.4** (**a**) The wave diagram of $w_1$ without the application of the control strategy. (**b**) The wave diagram of $w_2$ without the application of the control strategy. (**c**) The wave diagram of $w_3$ without the application of the control strategy

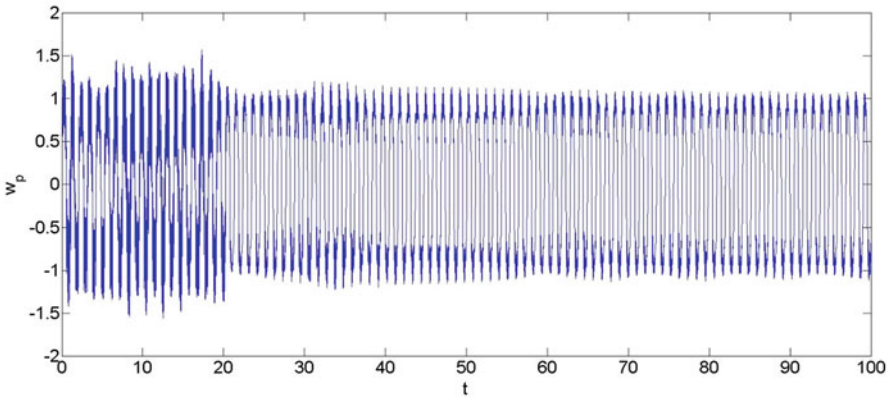



**Fig. 6.5** The wave diagram of $w_p$ with the application of the control strategy for the control parameters given in Eq. (6.24)

of the beam requires to be suppressed. The proposed active control strategy is found not only effective in reducing the amplitude of the response, but also synchronizing the vibration with the given frequency of the desired reference signal.

As shown in Fig. 6.5, the proposed control strategy is applied at $t = 20$, and the control parameters $\kappa$ and $k_{fs}$ are given in Eq. (6.24)

$$\kappa = 100, k_{fs} = 5000. \tag{6.24}$$

The desired reference signal and the unknown external disturbance take the following values,

$$w_r = 1.0 \sin(5.90t), \Delta F(w, \dot{w}) = 0.01 \sin\left(w_p\right). \tag{6.25}$$

As can be seen from Fig. 6.5, the maximum amplitude of the response of the beam is reduced significantly from about 1.85 to 1.0. The synchronization of the response of the beam also makes the response of the selected point become a periodic one.

From Fig. 6.6a–c, it should be noticed that the responses of the first three vibration modes of the beam are indeed affected with the application of proposed control strategy.

Figure 6.7 is presented to fully demonstrate the effectiveness of the proposed control strategy. In Fig. 6.7, the difference is small between the actual responses of the beam and that of the reference signal, and the synchronization between the actual responses of the beam at the selected point and the reference signal shows the significant effectiveness of the proposed control strategy.

In Fig. 6.8, the control input required for the vibration control of the selected point on the beam is given. From Fig. 6.8, the maximum control required for continuously maintaining the synchronization is around 5000, and it presents a periodic profile.

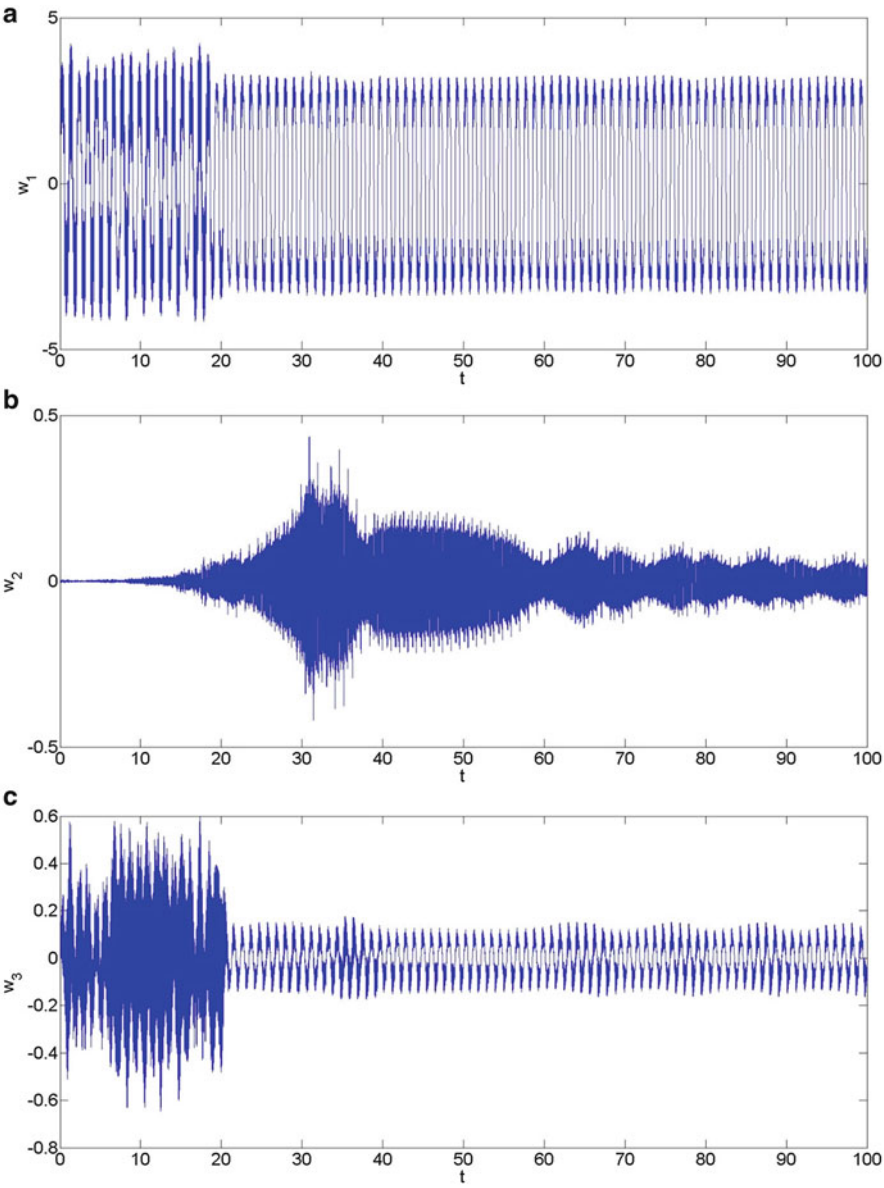



**Fig. 6.6** (**a**) The wave diagram of $w_1$ with the application of the control strategy for the control parameters given in Eq. (6.24). (**b**) The wave diagram of $w_2$ with the application of the control strategy for the control parameters given in Eq. (6.24). (**c**) The wave diagram of $w_3$ with the application of the control strategy for the control parameters given in Eq. (6.24)

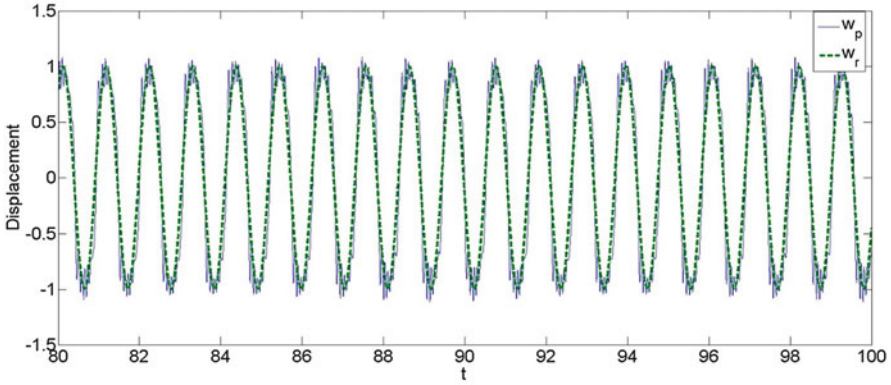



**Fig. 6.7** The wave diagram of $w_p$ (the *blue continuous line*) and the reference signal $w_r$ (the *green dash line*) for the control parameters given in Eq. (6.24)

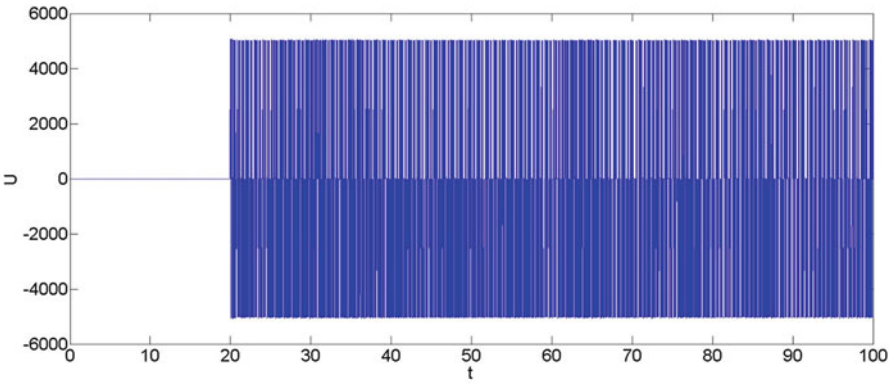



**Fig. 6.8** The control input $U$ for the control parameters given in Eq. (6.24)

### 6.6.2 $\kappa = 50, k_{fs} = 50$

In Fig. 6.8, the control input required for the vibration control of the selected point on the beam is around 5000, which means a very high and continuous control cost is required. Therefore, the efficiency of the control strategy should be considered for the practical application of the control strategy in engineering fields.

To test the efficiency of the control strategy, another set of control parameters with smaller values are given in the following,

$$\kappa = 50, k_{fs} = 50, \tag{6.26}$$

and the desired reference signal and the unknown external disturbance still take the same values given in Eq. (6.25).

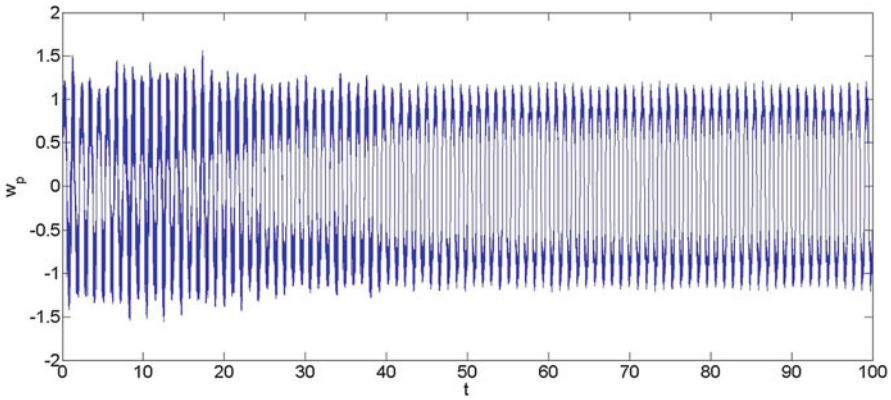



**Fig. 6.9** The wave diagram of $w_p$ with the application of the control strategy for the control parameters given in Eq. (6.26)

As can be seen from Fig. 6.9, the maximum amplitude of the response of the beam is again reduced significantly from about 1.85 to around 1.1. The synchronization of the response of the beam also makes the response of the selected point become a periodic one. However, the amplitude synchronization has not been achieved well since the difference becomes larger between the actual responses of the selected point on the beam and the desired reference signal.

From Fig. 6.10a–c, it should be noticed that the responses of the first three vibration modes of the beam are indeed affected with the application of proposed control strategy, in the case of the lower control parameters given in Eq. (6.26).

Figure 6.11 is presented to fully demonstrate the effectiveness of the proposed control strategy. In Fig. 6.11, the difference is larger between the actual responses of the beam and that of the reference signal than that in the case given in Sect. 6.6.1. However, it should be noticed that the synchronization still works well considering the small difference between the actual responses and the desired reference signal.

In Fig. 6.12, the control input required for the vibration control of the selected point on the beam is given. From Fig. 6.12, the maximum control required for continuously maintaining the synchronization is around 110, and it presents a periodic profile. Comparing to the control input shown in Sect. 6.6.1, the maximum control input required has been reduced by about 98 %. That is, the efficiency of the proposed control strategy can be greatly enhanced at the cost of a highly exact synchronization.

### *6.6.3* $\kappa = 10, k_{fs} = 50$

In Sect. 6.6.2, it can be learned that properly lowering the control parameters may greatly enhance the efficiency of the proposed control strategy. However, the

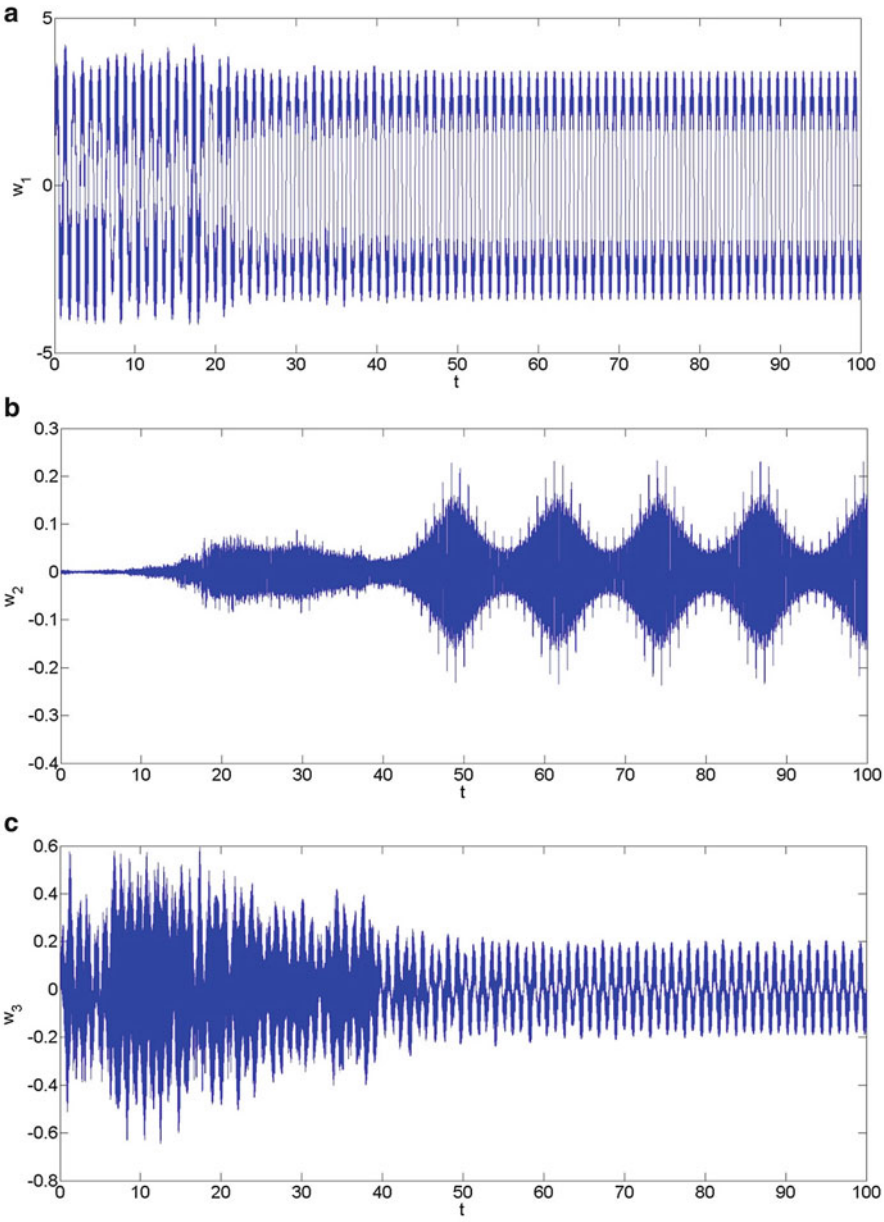



**Fig. 6.10** (**a**) The wave diagram of $w_1$ with the application of the control strategy for the control parameters given in Eq. (6.26). (**b**) The wave diagram of $w_2$ with the application of the control strategy for the control parameters given in Eq. (6.26). (**c**) The wave diagram of $w_3$ with the application of the control strategy for the control parameters given in Eq. (6.26)

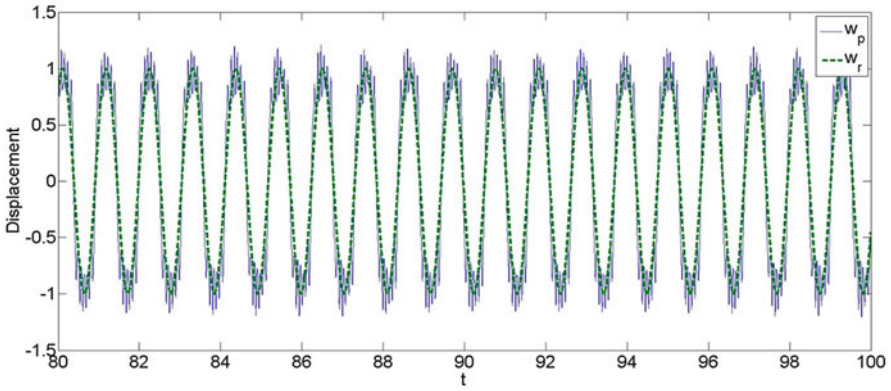



**Fig. 6.11** The wave diagram of $w_p$ (the *blue continuous line*) and the reference signal $w_r$ (the *green dash line*) for the control parameters given in Eq. (6.26)

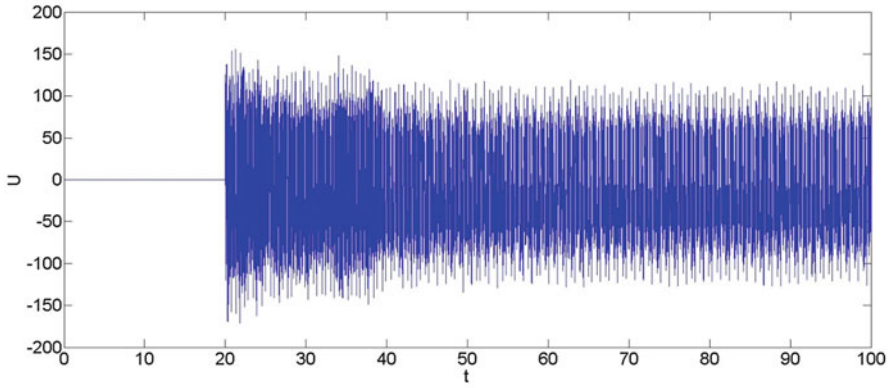



**Fig. 6.12** The control input $U$ for the control parameters given in Eq. (6.26)

control process may become very sensible to reducing the control parameters. In some cases, simply reducing the control parameters may not enhance the control efficiency, and may even lead to a response that may never get stabilized.

For instance, take the set of control parameters with smaller values as shown below,

$$\kappa = 10, k_{fs} = 50, \tag{6.27}$$

and the desired reference signal and the unknown external disturbance still take the same values given in Eq. (6.25).

In this case, it should be noticed that the value given in Eq. (6.27) features a smaller value of $\kappa$.

As can be seen from Fig. 6.13, the amplitude of the actual responses at the selected point on the beam can never get stabilized in the simulation. The

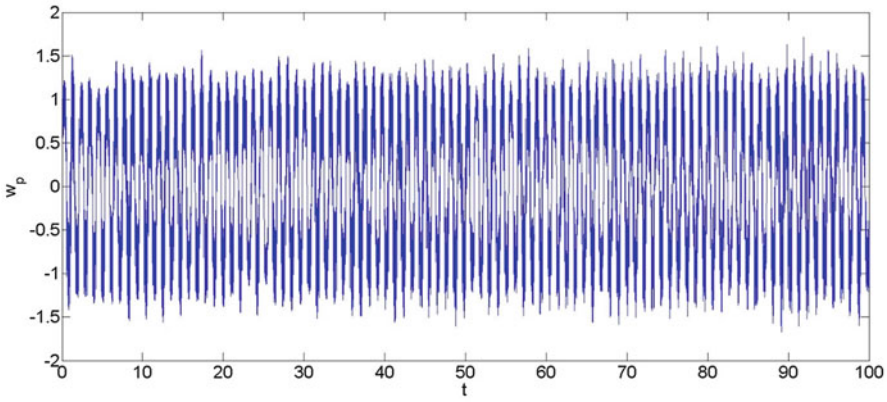



**Fig. 6.13** The wave diagram of $w_p$ with the application of the control strategy for the control parameters given in Eq. (6.27)

synchronization of the response of the beam cannot be achieved since the amplitude synchronization cannot be met. Besides, the responses of the selected point present a non-periodic vibration from the beginning to the end of the simulation.

From Fig. 6.14a–c, it should be noticed that the responses of the first three vibration modes of the beam are indeed affected with the application of proposed control strategy, in the case of the lower control parameters given in Eq. (6.27). However, the application of the control strategy with the lower control parameters does not stabilize any of the first three selected vibration modes.

Figure 6.15 is presented to fully demonstrate the ineffectiveness of the proposed control strategy with a lower control parameter. In Fig. 6.15, the difference is very significant between the actual responses of the beam and that of the reference signal in comparing with that in the case given in Sects. 6.6.1 and 6.6.2. Such difference demonstrates that simply reducing the control parameters will not present an expected synchronization. In this specific, simply lowering the control parameters will never lead to a stabilized response.

In Fig. 6.16, the control input required for the vibration control of the selected point on the beam is given. From Fig. 6.16, the maximum control required for continuously maintaining the synchronization is around 150, and it presents a non-periodic profile. Comparing to the control inputs shown in Sect. 6.6.2, the maximum control input required has not been reduced. That is, only focusing on the efficiency of the proposed control strategy will not only destabilize the response of the selected, but also increase the control inputs on the contrary.

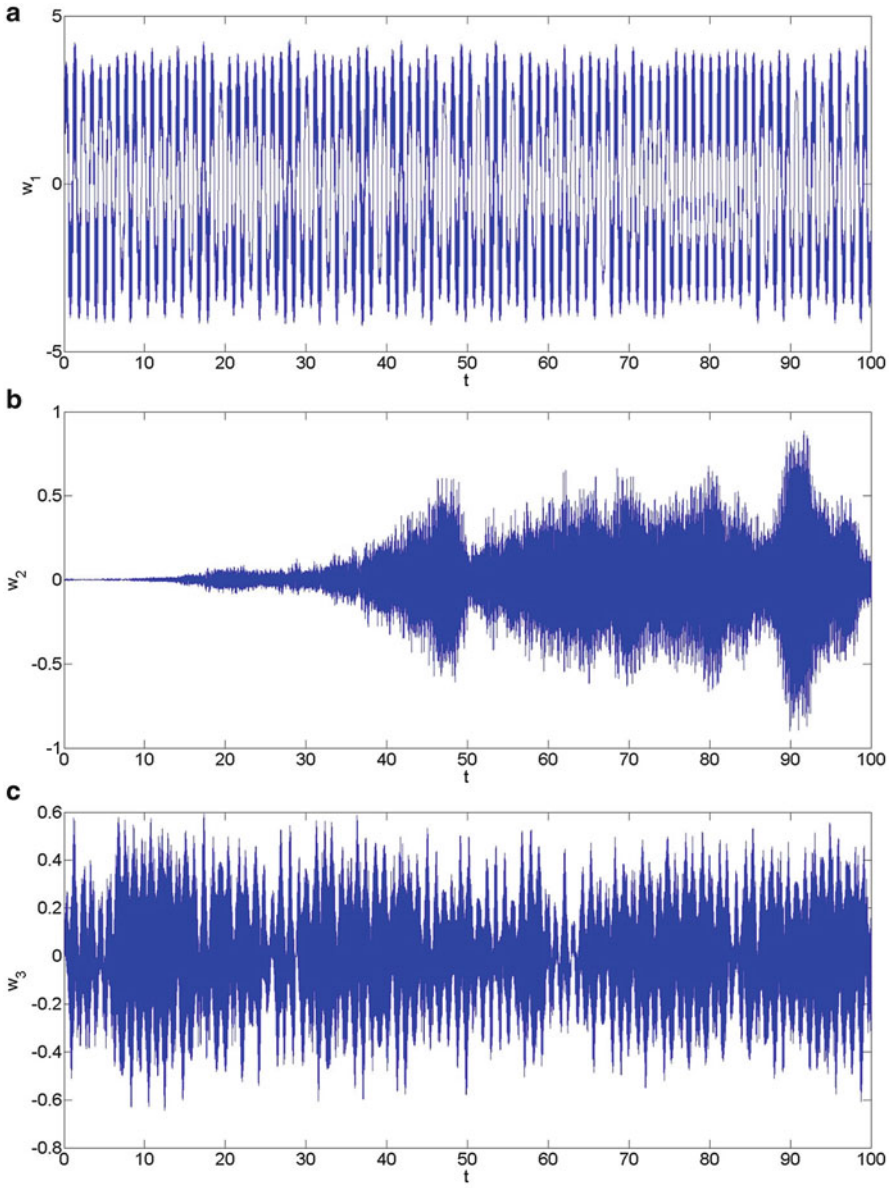



**Fig. 6.14** (**a**) The wave diagram of $w_1$ with the application of the control strategy for the control parameters given in Eq. (6.27). (**b**) The wave diagram of $w_2$ with the application of the control strategy for the control parameters given in Eq. (6.27). (**c**) The wave diagram of $w_3$ with the application of the control strategy for the control parameters given in Eq. (6.27)

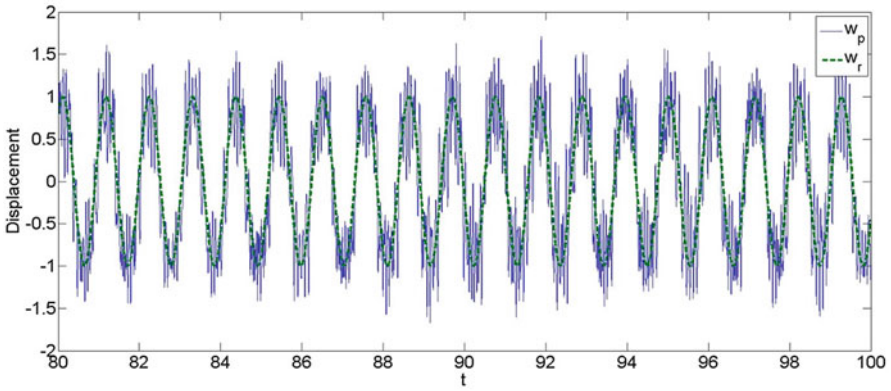



**Fig. 6.15** The wave diagram of $w_p$ (the *blue continuous line*) and the reference signal $w_r$ (the *green dash line*) for the control parameters given in Eq. (6.27)

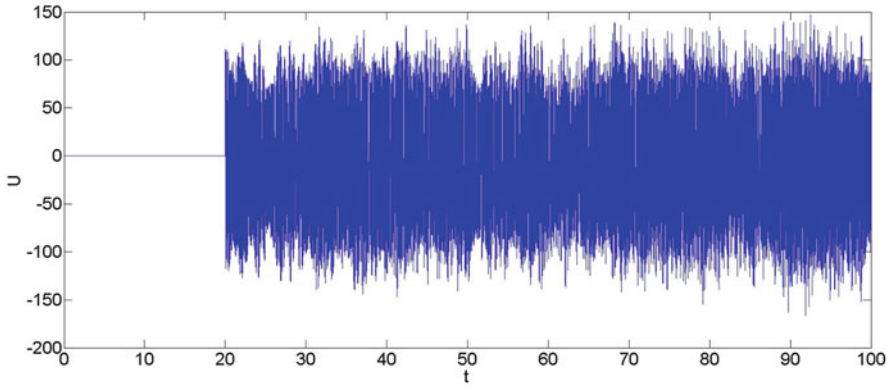



**Fig. 6.16** The control input $U$ for the control parameters given in Eq. (6.27)

## 6.7 Conclusion

An active control strategy is developed in this chapter for controlling the nonlinear vibrations of a simply supported beam. A nonlinear multi-dimensional dynamic system via the higher order Galerkin method is developed for modeling the nonlinear vibrations of the beam. Development of such control strategy is significant, and in fact there is no control strategy found in the literature available for controlling the chaotic vibrations of a multi-dimensional dynamic system. With the findings of the chapter, the following can be concluded:

1. To better and accurately analyze the vibrations of a simply supported beam subject to sinusoidal external excitation, the contributions of the higher order vibrations are significant especially when chaotic vibrations need to be taken

into consideration. As shown in the present chapter, the effects of the higher order vibrations must be considered in controlling the vibrations of a nonlinear beam system.
2. The control strategy developed in the chapter is effective in controlling the large-amplitude chaotic vibrations of the Euler–Bernoulli beam subjected to external excitations.
3. The frequency of the reference signal should be the same as the one of the external sinusoidal excitation, in order to facilitate the application of control strategy. This feature also helps the determination of the reference signal, since the value of the frequency of the reference can be specified with the external sinusoidal excitation.
4. A continuous control input is required for continuously applying the control strategy due to the existence of the external excitation. In addition, the variation of the control input would present a periodic profile in Fig. 6.8, once the synchronization of the actual response of the selected point is completed.

## References

Bouwstra, S., & Geijselaers, B. (1997). On the resonance frequencies of microbridges. *Proceedings of the 6th International Conference on Slide-State Sensors and Actuators (TRANSDUCERS'91)*, *2*, 1141–1144.

Dai, L. (2008). *Nonlinear dynamics of piecewise constant systems and implementation of piecewise constant arguments*. New Jersey: World Scientific.

Dai, L., & Singh, M. C. (1997). Diagnosis of periodic and chaotic responses in vibratory systems. *Journal of the Acoustical Society of America, 102*(6), 3361–3371.

Dai, L., & Singh, M.C. (1998). Periodic, quasiperiodic and chaotic behavior of a driven froude pendulum *International Journal of Non-linear Mechanics,33*(6), 947–965.

Dai, L., & Sun, L. (2012). On the fuzzy sliding mode control of nonlinear motions in a laminated beam. *Journal of Applied Nonlinear Dynamics, 1*(3), 287–307.

Haghighi, H. H., & Markazi, A. H. D. (2009). Chaos prediction and control in MEMS resonators. *Community of Nonlinear Science, 15*(10), 3091–3099.

Kuo, C. L., Shieh, C. S., Lin, C. H., & Shih, S. P. (2007). Design of an Adaptive fuzzy sliding-mode controller for chaos synchronization. *International Journal of Nonlinear Science, 8*(4), 631–636.

Utkin, V. I. (1992). *Sliding modes in control and optimization*. Berlin: Springer.

Xie, W. C., Lee, H. P., & Lim, S. P. (2003). Nonlinear dynamics analysis of MEMS switches by nonlinear modal. *Nonlinear Dynamics, 31*(3), 243–256.

Yau, H. T., & Kuo, C. L. (2006). Fuzzy sliding mode control for a class of chaos synchronization with uncertainties. *International Journal of Nonlinear Sciences and Numerical Simulation, 7*(3), 333–338.

Younis, M. I., Abdel-Rahman, E. M., & Nayfeh, A. H. (2003). A reduced-order model for electrically actuated microbeam-based MEMS. *Journal of Microelectromechanical Systems, 12*(5), 672–680.

Younis, M. I., & Nayfeh, A. H. (2002). A study of the nonlinear response of a resonance of an electric actuation. *Nonlinear Dynamics, 31*(1), 91–117.

# Chapter 7
# Nonlinear Forced Vibration of Nanobeams

**Mohammad M. Aghdam and Hamed Niknam**

**Abstract** In this chapter, the effects of nonlocal stress field on the beam excited with distributed and concentrated harmonic loads are investigated. The Euler–Bernoulli assumptions together with Von Karman geometric type nonlinearity are used to formulate the strain field. The stress field is written based on Eringen's nonlocal theory. Applying the Hamilton principle, the governing partial differential equation is obtained. Implementing Galerkin technique together with Ritz averaging method, closed form solutions are presented for both time and frequency responses of the beam.

**Keywords** Eringen's nonlocal theory • Nonlinear forced vibration • Nonlinear frequency response • Galerkin technique • Ritz averaging method

## 7.1 Introduction

Application of macro-scale beams subjected to dynamic loading is extensive in various engineering structures. These dynamics loads may be idealized as harmonic excitations in different structures. Depending on the amplitude of excitation and material properties, nonlinearity may occur in the problem. Up to now, many studies have attempted to investigate nonlinear behavior of beams in forced vibration. Benamar et al. (1991) studied effects of nonlinearity on the mode shapes and natural frequency of thin elastic structures. Forced response and stability of a beam with fully clamped edges was studied by Bennett and Eisley (1970) by implementing a multiple-degree-of-freedom approach. Busby and Weingarten (1972) applied finite element formulation to obtain an approximate solution for nonlinear equation of

M.M. Aghdam (✉)
Department of Mechanical Engineering, Thermoelasticity Center of Excellence, Amirkabir University of Technology, Hafez Ave, Tehran, Iran
e-mail: aghdam@aut.ac.ir

H. Niknam
Department of Mechanical Engineering, Amirkabir University of Technology, Hafez Ave, Tehran, Iran
e-mail: hamed.niknam@aut.ac.ir



R.N. Jazar, L. Dai (eds.), *Nonlinear Approaches in Engineering Applications*,
DOI 10.1007/978-3-319-27055-5_7

forced vibration of beams. Including longitudinal displacement and inertia, Mei and Decha-Umphai (1985) applied finite element method to study nonlinear forced vibration of beams. Large deflection of cantilever beams subjected to an end force using homotopy perturbation method was studied by Wang et al. (2012). Shooshtari and Rafiee (2011) used multiple scale method to study forced vibration of clamped–clamped functionally graded beam. Other types of boundary conditions are, also, discussed by Hosseini et al. (2012). Şimşek and Kocatürk (2009) analyzed the behavior of FG Euler–Bernoulli beam under concentrated moving harmonic load. Other studies on large amplitude forced vibration behavior of beams may be also found in (Azrar et al. 1999, 2002; Kanani et al. 2014; Lewandowski 1997a, b; Ribeiro and Petyt 1999; Shi et al. 1997; Şimşek 2010a, b).

Nanostructures have attracted much attention in the past two decades through wide range of application and superior mechanical properties. Their small dimension, however, resulted different elastic and dynamics behavior which are mainly caused by existence of quantum effects and high surface to volume ratio (Valcárcel et al. 2008). These size effects cannot be considered through classical continuum theory; therefore, different continuum-based theories have been proposed to include the effect of small dimensions in the analysis (Aifantis 1999; Eringen 2010; Gurtin et al. 1998; Yang et al. 2002). Among them, Eringen's nonlocal elasticity theory (Eringen 2010) has received much attention in research communities and was applied by various researchers to study different behavior of structural elements. In most of the cases, comparing their results with atomic-based approach reveals a reasonably good agreement (Ansari et al. 2011; Duan et al. 2007; Kiani 2010).

Many researchers have used nonlocal elasticity theory to investigate vibrational behavior of different types of beams. Reddy and Pang (2008) reformulated Euler–Bernoulli and Timoshenko beam theories using the nonlocal differential constitutive relations of Eringen. Aydogdu (2009) and Reddy (2007) investigated the effect of nonlocal parameter on bending, buckling, and vibration of nanobeams using different strain fields. Thai (2012) and Thai and Vo (2012) applied some modification in the beam theories and obtained results for vibration behavior of nanobeams. Some other investigation in the field of vibration of nanobeams can be found elsewhere (Adhikari et al. 2013; Eltaher et al. 2012, 2013; Niknam and Aghdam 2015; Şimşek and Yurtcu 2013).

Investigations on the forced vibration behavior of nanobeams are restricted to few articles. Şimşek and Kocatürk (2009) studied the behavior of FG nanobeam subjected to moving harmonic load. Şimşek (2010c), also, used nonlocal elasticity theory to investigate the behavior of SWCNTs in forced vibration in which it was shown that increasing nonlocal parameter leads to increase in dynamic deflection of the beam. Frequency response of linear forced vibration of CNTs is obtained by Karaoglu and Aydogdu (2010). Simsek (2011) investigated the effects of nonlocality on the forced vibration of elastically connected double-carbon nanotube under moving nanoparticle in which it was shown that the stiffness of elastic layer has a significant role in the behavior of DCNTs. More recently, Lei et al. (2013) studied the dynamic characteristic of damped viscoelastic nanobeams. Uymaz (2013) studied the forced vibration of nonlocal FG beams using Navier solution

in which no nonlinearity is observed. However, to the best of authors' knowledge, none of the previous works have included effect of nonlinearity and large amplitude deflection on nonlocal beams.

In the present study, the Eringen's nonlocal elasticity together with Euler–Bernoulli beam theory is implemented to obtain the governing equation of the beam. The nonlinearity is modeled according to Von Karman geometric type nonlinearity. The excitation is considered to be either distributed or concentrated harmonic loading. Galerkin technique together with Ritz averaging method is applied to the governing equation in order to obtain frequency and time response of the beam. Finally, the effects of the presence of nonlocality on the time and frequency response of nanobeam are illustrated through figures and tables.

## 7.2 Governing Equations

In the classical elastic continuum theory, the stress field at a point depends only on the strain field at the same point. However, according to Eringen's nonlocal elasticity theory, stress field at a point is dependent on strains at all other points of the body. Therefore, the following equation is proposed by Eringen (2010) to describe the stress field at nano-scale.

$$\mathcal{L}\sigma^{nl} = \left(1 - \mu\nabla^2\right)\sigma^{nl} = T,\ \mu = e_0^2 a^2 \tag{7.1}$$

In which $\mathcal{L} = \left(1 - \mu\nabla^2\right)$ and $\nabla^2$ are nonlocal differential operator and Laplacian operator, respectively. $e_0$ is a constant related to the material of the beam and $a$ is internal characteristic length, e.g., lattice parameter, wavelength.

In Fig. 7.1a, b, a nanobeam with length $L$, thickness $h$, and width $b$ is depicted. According to the Euler–Bernoulli beam theory, the axial and transverse displacements of any point of the beam are, respectively, considered as

$$u\left(x, z, t\right) = U\left(x, t\right) + z\frac{\partial W\left(x, t\right)}{\partial x} \tag{7.2a}$$

$$w\left(x, z, t\right) = W\left(x, t\right) \tag{7.2b}$$

in which $U$ and $W$ are the axial and transverse displacements of any point on the neutral axis of the beam. Considering Von Karman geometric type nonlinearity, the only nonzero term of strain tensor may be written as:

$$\varepsilon_{xx} = \frac{\partial u}{\partial x} + \frac{1}{2}\left(\frac{\partial W}{\partial x}\right)^2 = \frac{\partial U}{\partial x} + z\frac{\partial^2 W}{\partial x^2} + \frac{1}{2}\Pi\left(\frac{\partial W}{\partial x}\right)^2 \tag{7.3}$$

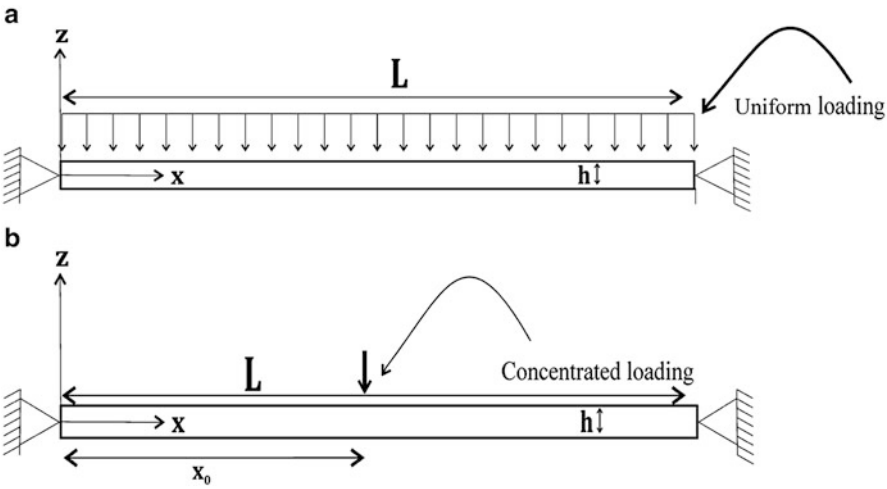



**Fig. 7.1** Schematic view of nanobeam subjected to: (**a**) uniform distributed loading (**b**) concentrated loading

The parameter Π in (7.3) is introduced to include nonlinearity effects in the analysis and is equal to 0 or 1 for linear and nonlinear analyses, respectively. Applying Hamilton principle, the governing equations are obtained as:

$$\frac{\partial N^{nl}}{\partial x} = \rho A \frac{\partial^2 U}{\partial t^2} \tag{7.4a}$$

$$\frac{\partial^2 M^{nl}}{\partial x^2} - \frac{\partial}{\partial x}\left(N^{nl}\frac{\partial W}{\partial x}\right) + \rho A \frac{\partial^2 W}{\partial t^2} - \rho I \frac{\partial^4 W}{\partial t^2 \partial x^2} = f(x,t) \tag{7.4b}$$

where $f(x,t)$ is the applied external transverse distributed force in unit length. $N^{nl}$ and $M^{nl}$ are the force and moment resultants, $A$ and $I$ refer to area and second moment of the cross section of the beam and are defined as:

$$\left(N^{nl},\ M^{nl}\right) = b\int_{-\frac{h}{2}}^{\frac{h}{2}} (1,\ z)\,\sigma_{xx}^{nl} dz \tag{7.5}$$

$$(A,\ I) = \left(bh,\ \frac{bh^3}{12}\right) \tag{7.6}$$

where $\sigma_{xx}^{nl}$ is the normal stress in the $x$-direction. Neglecting the effect of in-plane and rotational inertia, one may rewrite Eq. (7.4) as

$$\frac{\partial N^{nl}}{\partial x} = 0 \tag{7.7a}$$

$$\frac{\partial^2 M^{nl}}{\partial x^2} - N^{nl}\frac{\partial^2 W}{\partial x^2} + \rho A\frac{\partial^2 W}{\partial t^2} = f(x,t) \tag{7.7b}$$

Using generalized Hooke's law, the local value of stress, $\sigma_{xx}$, is obtained as

$$\sigma_{xx} = E\varepsilon_{xx} \tag{7.8}$$

where $E$ is the Young modulus. Substituting Eq. (7.8) into (7.5), the local force and moment resultants, $N$ and $M$, may be obtained as

$$N = EA\left(\frac{\partial U}{\partial x} + \frac{1}{2}\Pi\left(\frac{\partial W}{\partial x}\right)^2\right) \tag{7.9a}$$

$$M = EI\frac{\partial^2 W}{\partial x^2} \tag{7.9b}$$

Considering the fact that $N^{nl}$ is a constant value throughout the beam Eq. (7.7a) and integrating Eq. (7.1) over the beam cross section, it can be concluded that

$$N^{nl} = N = EA\left(\frac{\partial U}{\partial x} + \frac{1}{2}\Pi\left(\frac{\partial W}{\partial x}\right)^2\right) \tag{7.10a}$$

$$M^{nl} - \mu\frac{\partial^2 M^{nl}}{\partial x^2} = M \tag{7.10b}$$

Substituting Eqs. (7.10b) and (7.9b) in (7.4b) leads to

$$M^{nl} = EI\frac{\partial^2 W}{\partial x^2} + \mu\left(N\frac{\partial^2 W}{\partial x^2} - \rho A\frac{\partial^2 W}{\partial t^2} + f(x,t)\right) \tag{7.11}$$

For a beam with constrained edges at $x = 0$ and $x = l$ ($U(0,t) = U(l,t) = 0$), $N$ is also found to be

$$N = \Pi\frac{EA}{2l}\int_0^l \left(\frac{\partial W}{\partial x}\right)^2 dx \tag{7.12}$$

Now, substituting $M^{nl}$ from Eq. (7.11) in Eq. (7.4b) it is obtained that

$$(EI + \mu N)\frac{\partial^4 W}{\partial x^4} - N\frac{\partial^2 W}{\partial x^2} - \mu\rho A\frac{\partial^4 W}{\partial x^2 \partial t^2} + \rho A\frac{\partial^2 W}{\partial t^2} = f(x,t) - \mu\frac{\partial^2 f(x,t)}{\partial x^2} \tag{7.13}$$

Finally, the following dimensionless parameters are introduced to perform various parametric analyses

$$\widehat{x} = \frac{x}{l};\ \widehat{W} = \frac{W}{r};\ \widehat{t} = t\sqrt{Er^2/\rho l^4};\ \widehat{f} = \frac{fl^4}{EIr};\ \widehat{\mu} = \frac{\mu}{l^2}$$

in which $r = \sqrt{\frac{I}{A}}$ is the radius of gyration of the cross section. Substituting the dimensionless parameters in (7.13), the dimensionless form of the equation of motion is derived as

$$\frac{\partial^2 \widehat{W}}{\partial \widehat{t}^2} - \mu \frac{\partial^4 \widehat{W}}{\partial \widehat{x} \partial \widehat{t}^2} + \left(1 + \pi \frac{\widehat{\mu}}{2} \int_0^1 \left(\frac{\partial \widehat{W}}{\partial \widehat{x}}\right)^2 d\widehat{x}\right) \frac{\partial^4 \widehat{W}}{\partial \widehat{x}^4} - \left(\prod \frac{1}{2} \int_0^1 \left(\frac{\partial \widehat{W}}{\partial \widehat{x}}\right) d\widehat{x}\right) \frac{\partial^2 \widehat{W}}{\partial \widehat{x}^2} = \widehat{f}\left(\widehat{x}, \widehat{t}\right) - \mu \frac{\partial^2 \widehat{f}\left(\widehat{x}, \widehat{t}\right)}{\partial \widehat{x}^2} \tag{7.14}$$

It is worth mentioning that the dimensionless form of the governing equation reveals that the effect of nonlocality depends only on the length of the beam. Therefore, using nonlocal elasticity and Euler–Bernoulli beam theory results in the fact that both height and width of the beam have no influence on the nonlocality of the problem.

## 7.3 Solution Procedure

In this section, as a first step, the Galerkin technique is applied to the governing Eq. (7.14) so that a time-dependent ordinary differential equation is obtained. Then, the Ritz averaging technique is implemented to find the characteristic equation.

### *7.3.1 Galerkin Technique*

According to Galerkin technique, it is assumed that the transverse displacement field, $\hat{W}$, has a form of $\widehat{W}\left(\widehat{x}, \widehat{t}\right) = \varnothing\left(\widehat{x}\right) q\left(\widehat{t}\right)$; in which $\varnothing\left(\widehat{x}\right)$ is the basic function that satisfies the essential boundary conditions of the beam and $q\left(\widehat{t}\right)$ is the time-dependent generalized coordinate. In the present study, $\varnothing\left(\widehat{x}\right)$ is considered to be the linear beam modes of vibration; which are introduced in Table 7.1 for different types of boundary conditions. Symbol *S* and *C* stand for a simply supported and clamped edge, respectively; for instance, *CS* refers to a beam with clamped edge at $\widehat{x} = 0$ and simply supported at $\widehat{x} = 1$.

Applying Galerkin technique (Azrar et al. 1999) to Eq. (7.16), the final time-dependent governing equation may be obtained as

$$\ddot{q} + \omega_0^2 q + \gamma q^3 = f_0\left(\widehat{t}\right) \tag{7.15}$$

**Table 7.1** The mode shape functions of a uniform beam for the different boundary conditions (Kanani et al. 2014)

| B.C. | Mode shape function |
|---|---|
| SS | $\sin(\pi x)$ |
| CS | $(\cosh(qx)-\cos(qx))-\frac{\cosh(q)-\cos(q)}{\sinh(q)-\sin(q)}(\sinh(qx)-\sin(qx))\,;q=3.9266$ |
| CC | $(\cosh(qx)-\cos(qx))-\frac{\cosh(q)-\cos(q)}{\sinh(q)-\sin(q)}(\sinh(qx)-\sin(qx))\,;q=4.7300$ |

which is the well-known Duffing equation as in nonlinear forced oscillators. In Eq. (7.15) upper dots denote differentiation with respect to dimensionless time and constant parameters $\omega_0^2$, $\gamma$, and $f_0$ are defined as

$$\omega_0^2=\frac{\int_0^1\phi^{(4)}\phi d\widehat{x}}{\int_0^1[\phi-\widehat{\mu}\phi'']\,\phi d\widehat{x}} \tag{7.16a}$$

$$\gamma=\frac{\int_0^1\left[\left(\frac{1}{2}\int_0^1\phi'^2d\widehat{x}\right)\left(\widehat{\mu}\phi^{(4)}-\phi''\right)\right]\phi d\widehat{x}}{\int_0^1[\phi-\widehat{\mu}\phi'']\,\phi d\widehat{x}} \tag{7.16b}$$

$$f_0=\frac{\int_0^1\left(\widehat{f}\left(\widehat{x},\widehat{t}\right)-\widehat{\mu}\frac{\partial^2\widehat{f}\left(\widehat{x},\widehat{t}\right)}{\partial\widehat{x}^2}\right)\phi d\widehat{x}}{\int_0^1[\phi-\widehat{\mu}\phi'']\,\phi d\widehat{x}} \tag{7.16c}$$

$\omega_0$ is known to be the natural frequency of equivalent linear system. Two types of excitation force are considered in this study; uniformly distributed harmonic force, $\widehat{f}^d\left(\widehat{x},\widehat{t}\right)$, and concentrated harmonic force, $f^c\left(\widehat{x},\widehat{t}\right)$, which are formulated as:

$$\widehat{f}^d\left(\widehat{x},\widehat{t}\right)=F_0^d\cos\left(\Omega\widehat{t}\right) \tag{7.17a}$$

$$\widehat{f}^c\left(\widehat{x},\widehat{t}\right)=F_0^c\cos\left(\Omega\widehat{t}\right)\delta\left(\widehat{x}-\widehat{x}_0\right) \tag{7.17b}$$

in which $\delta$ is the Dirac function (Kanwal 1998), $F_0^d$ and $F_0^c$ are the amplitudes of uniformly distributed and concentrated harmonic forces, respectively, and $\Omega$ is the dimensionless frequency of excitation. Therefore, using (7.16c), $f_0$ for two types of excitation force will be obtained as

$$f_0^d=F_0^{-d}\cos\left(\Omega\widehat{t}\right) \tag{7.18a}$$

$$f_0^c=F_0^{-c}\cos\left(\Omega\widehat{t}\right) \tag{7.18b}$$

where $F_0^{-d}$ and $F_0^{-c}$ are defined to be

$$F_0^{-d} = F_0^d \int_0^1 \phi d\widehat{x} \tag{7.19a}$$

$$F_0^{-c} = F_0^c \left(\phi\left(\widehat{x}_0\right) - \widehat{\mu}\phi''\left(\widehat{x}_0\right)\right) \tag{7.19b}$$

Equation (7.19b) is obtained by the usage of properties of Dirac delta function (Kanwal 1998).

### 7.3.2 Ritz Averaging Method

In order to solve the time-dependent nonlinear governing equation by Ritz averaging method (Genta 2009), $q\left(\widehat{t}\right)$ is considered to be

$$q\left(\widehat{t}\right) = Y\cos\left(\Omega\widehat{t}\right) \tag{7.20}$$

where $Y$ is the dimensionless amplitude of vibration. Substituting (7.20) and (7.18a, 7.18b) in (7.15) leads to

$$-Y\Omega^2\cos\left(\Omega\widehat{t}\right) + Y\omega_0^2\cos\left(\Omega\widehat{t}\right) + \gamma Y^3\cos^3\left(\Omega\widehat{t}\right) - \overline{F}_0\cos\left(\Omega\widehat{t}\right) = 0 \tag{7.21}$$

in which $\overline{F}_0$ can be either $\overline{F}_0^d$ or $\overline{F}_0^c$.

Now, according to Ritz averaging technique, it is written that

$$\int_0^{\frac{2\pi}{\Omega}} \left[-Y\Omega^2\cos(\Omega\widehat{t}) + Y\omega_0^2\cos(\Omega\widehat{t}) + \gamma Y^3\cos^3(\Omega\widehat{t}) - \overline{F}_0\cos(\Omega\widehat{t})\right]\cos(\Omega\widehat{t})d\widehat{t} = 0 \tag{7.22}$$

which yields

$$\frac{3\gamma}{4}Y^3 + \left(\omega_0^2 - \Omega^2\right)Y - \overline{F}_0 = 0 \tag{7.23}$$

which is interpreted as the frequency response of the nonlinear beam.

In the case of linear analysis, $\Pi = 0$, leads to $\gamma = 0$. Therefore, the frequency response can be obtained as

$$Y = \frac{\overline{F}_0}{\omega_0^2 - \Omega^2} \tag{7.24}$$

## 7.4 Results and Discussion

In this section, results are presented to investigate the effect of nonlocality on the forced vibration behavior of nanobeams. First of all, in order to ensure that the coefficients of nonlinear governing equation have been obtained accurately, the determined natural frequency of nanobeam based on the presented solution procedure is compared with available data in literature (Şimşek 2014). It should be noted that the following formula is presented in (Şimşek 2014) to determine nonlinear natural frequency

$$\omega_{nl}^2 = \omega_0^2 + \frac{3}{4}\gamma A^2 \tag{7.25}$$

where $A$ is the mid-point amplitude of vibration. It is observed that the nonlinear natural frequency is only dependent on these coefficients. Predictions for natural frequency ratio together with results of (Şimşek 2014) which are based on the Eq. (7.25) are tabulated in Table 7.2. Excellent agreement can be observed between both approaches. This verifies accuracy of the presented procedure to obtain coefficients of nonlinear time-dependent governing equation.

In Fig. 7.2, time response of the beam obtained by MATLAB ODE45 solution is compared with the predictions of the presented solution procedure in Sect. 7.3. Again, excellent agreement can be seen between the two solution techniques.

**Table 7.2** Frequency ratio ($\omega_{nl}/\omega_0$) for nonlocal beam

| B.C. | $A$ | $\widehat{\mu} = 0$ | | $\widehat{\mu} = 0.25$ | | $\widehat{\mu} = 1$ | |
|---|---|---|---|---|---|---|---|
| | | Present work | Şimşek (2014) | Present work | Şimşek (2014) | Present work | Şimşek (2014) |
| *SS* | 1 | 1.08972 | 1.08972 | 1.28457 | 1.28457 | 1.74300 | 1.74300 |
| | 2 | 1.32287 | 1.32287 | 1.89751 | 1.89751 | 3.02526 | 3.02526 |
| | 3 | 1.63935 | 1.63935 | 2.61748 | 2.61748 | 4.39800 | 4.39800 |
| | 4 | 2.00000 | 2.00000 | 3.37671 | 3.37671 | 5.79731 | 5.79731 |
| *CC* | 1 | 1.05517 | 1.05517 | 1.50557 | 1.50557 | 2.39308 | 2.39308 |
| | 2 | 1.20563 | 1.20563 | 2.46313 | 2.46313 | 4.46177 | 4.46177 |
| | 3 | 1.42143 | 1.42143 | 3.52148 | 3.52148 | 6.59862 | 6.59862 |
| | 4 | 1.67755 | 1.67755 | 4.61173 | 4.61173 | 8.75385 | 8.75385 |
| *CS* | 1 | 1.09950 | 1.09950 | 1.51273 | 1.51273 | 2.35080 | 2.35080 |
| | 2 | 1.35510 | 1.35510 | 2.48062 | 2.48062 | 4.37093 | 4.37093 |
| | 3 | 1.69755 | 1.69755 | 3.54899 | 3.54899 | 6.46037 | 6.46037 |
| | 4 | 2.08452 | 2.08452 | 4.64908 | 4.64908 | 8.56856 | 8.56856 |

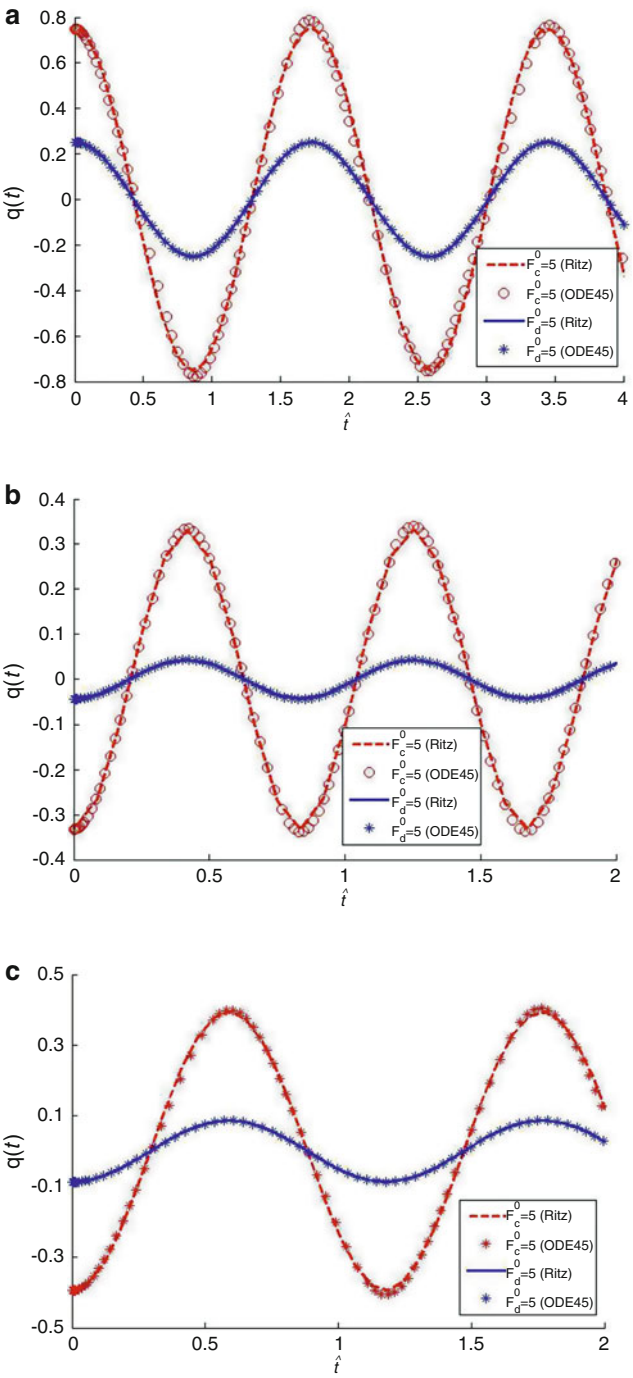



**Fig. 7.2** Time response of nanobeam subjected to distributed and concentrated harmonic loading with $\hat{\mu} = 0.5$, $\hat{x}_0 = 0.5$, and $\Omega = 0.9\omega_0$. (**a**) *SS* (**b**) *CC* (**c**) *CS*

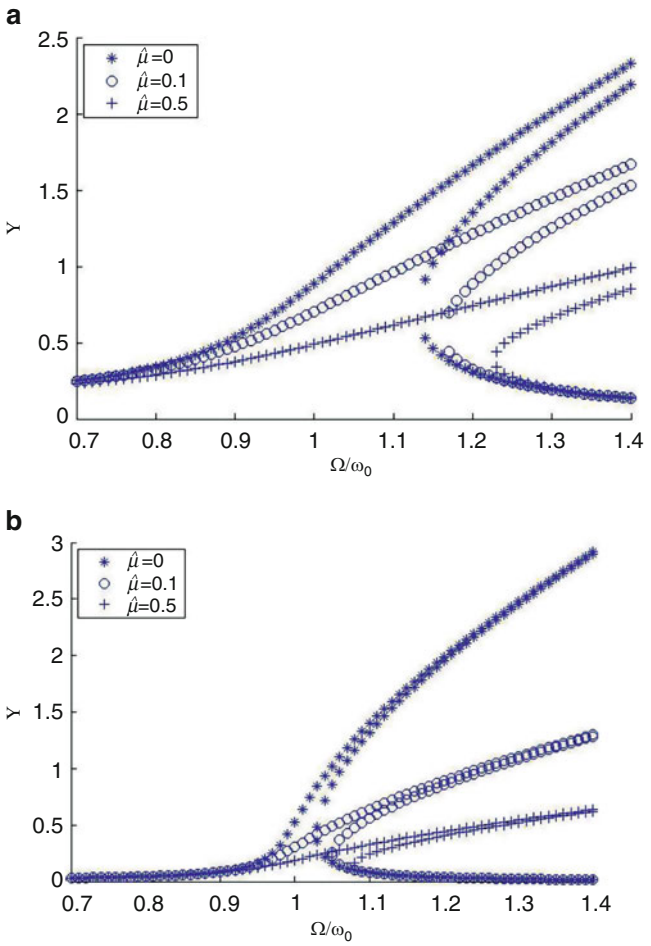



**Fig. 7.3** Nonlinear frequency response of nanobeam subjected to distributed harmonic loading with $F_d^0 = 10$. (**a**) *SS* (**b**) *CC*

The effects of nonlocal parameter, $\widehat{\mu}$, on the nonlinear frequency response of nanobeam subjected to distributed and concentrated harmonic loading are, respectively, presented in Figs. 7.3a, b and 7.4a, b. These figures reveal that increasing nonlocal parameter leads to more nonlinearity in the vibration behavior. Therefore, for the same value of excitation frequency ratio, $\frac{\Omega}{\omega_0}$, the amplitude of vibration for nonlocal beam is higher than that of local beam.

Figure 7.5a–c illustrates the same effect on the linear frequency response of the nanobeams for various boundary conditions. It is interesting to note that different trends can be observed in linear and nonlinear behavior of the beams. For instance, in the case of distributed loading, the nonlocal parameter, $\widehat{\mu}$, has no effects on the

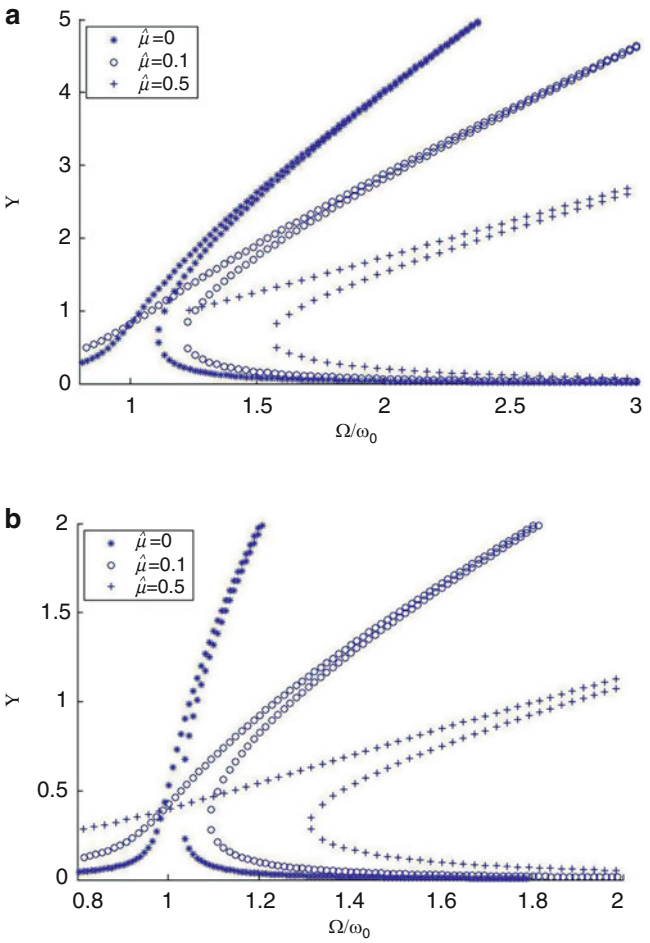



**Fig. 7.4** Nonlinear frequency response of nanobeam subjected to concentrated harmonic loading with $F_c^0 = 5$ and $\hat{x}_0 = 0.5$. (**a**) *SS* (**b**) *CC*

vibration amplitude, $Y$, versus excitation frequency ratio, $\frac{\Omega}{\omega_0}$, while it has a strong influence in the case of nonlinear response as shown in Fig. 7.3. However, in the case of concentrated load at the mid-point of the beam, the nonlocal parameter affects both linear and nonlinear frequency responses in two different ways. In this case, while increasing nonlocal parameter leads to lower amplitudes for a constant excitation frequency ratio in the nonlinear analysis, it has an opposite effect for the linear analysis. It should also be noted that similar conclusions can be obtained from Eqs. (7.16a, 7.16b, 7.16c), (7.19a, 7.19b), and (7.24) for the case of linear analysis.

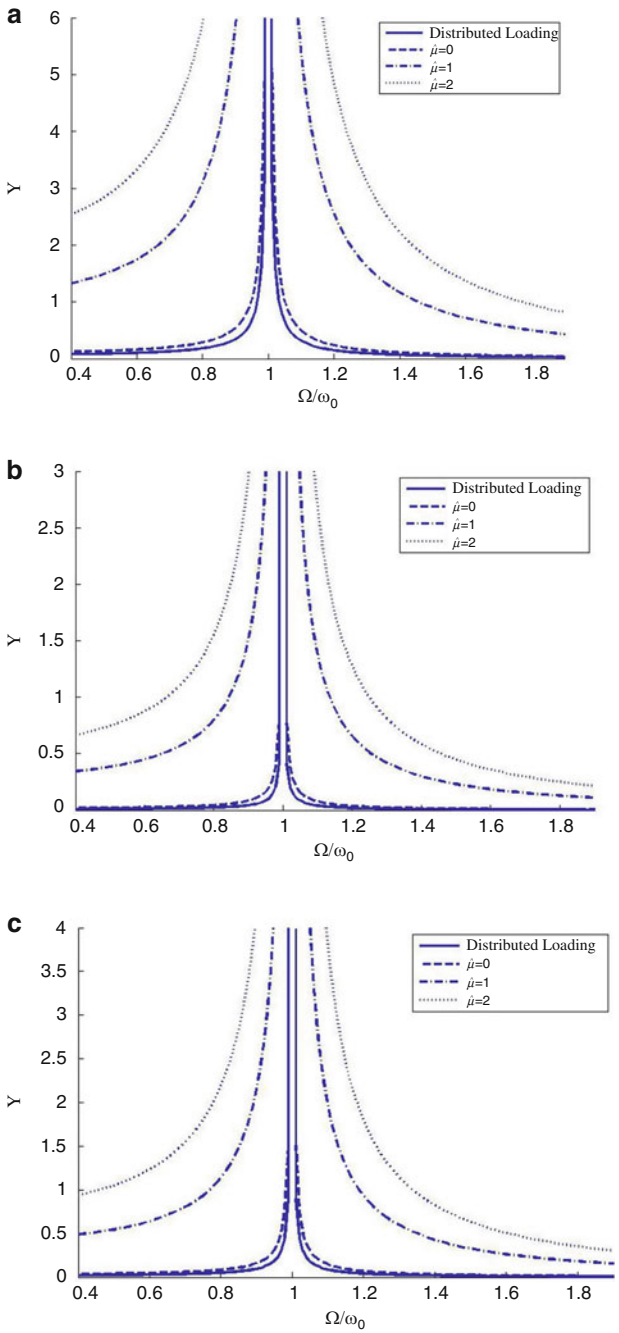



**Fig. 7.5** Nonlinear frequency response of nanobeam subjected to both distributed and concentrated harmonic loading with $F^0 = 5$ and $\hat{x}_0 = 0.5$. (**a**) *SS* (**b**) *CC* (**c**) *CS*

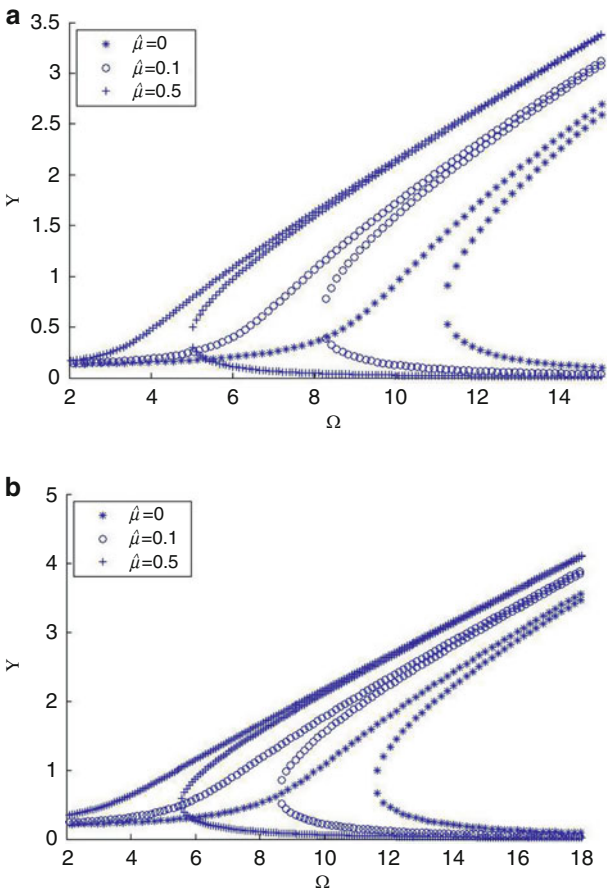



**Fig. 7.6** Nonlinear vibration amplitude of *SS* nanobeam versus non-dimensional excitation frequency, Ω, subjected to (**a**) distributed loading with $F_d^0 = 10$ (**b**) concentrated loading with $F_c^0 = 10$ and $\hat{x}_0 = 0.5$

Figures 7.3, 7.4, and 7.5 in which frequency responses are plotted vs. frequency ratio may result in an erroneous conclusion that the presence of nonlocal parameter has a uniform influence on the amplitude of the vibration. This is due to the fact that nonlocal parameter affects the linear natural frequency which is used to define the excitation frequency ratio, $\frac{\Omega}{\omega_0}$. In order to clarify the case, similar analysis as presented in Figs. 7.3, 7.4, and 7.5 is also regenerated and shown in Figs. 7.6 and 7.7 using dimensionless excitation frequency, Ω, instead of frequency ratio, $\frac{\Omega}{\omega_0}$. It can be concluded from the figures that the effect of nonlocal parameter on the amplitude of the vibration is different for low and high excitation frequencies.

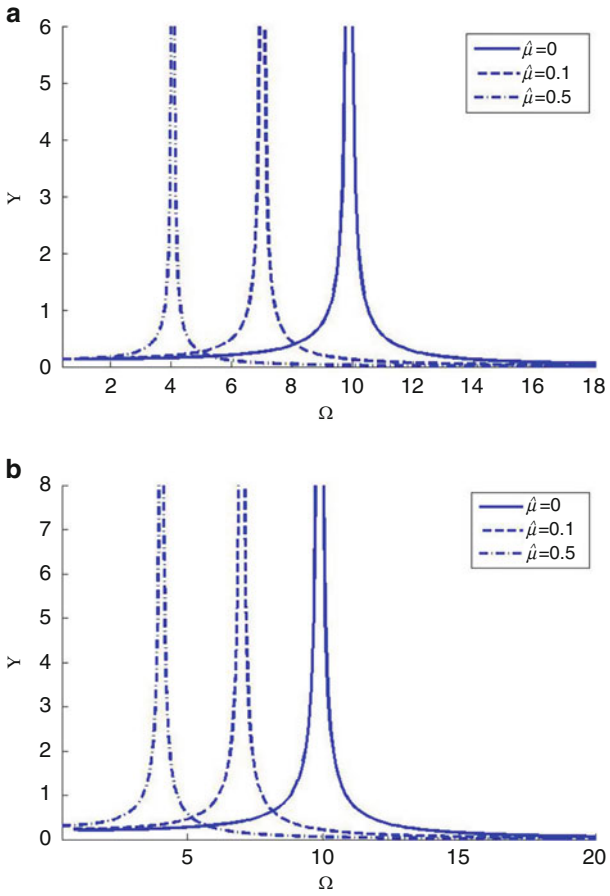



**Fig. 7.7** Linear vibration amplitude of *SS* nanobeam versus non-dimensional excitation frequency, Ω, subjected to (**a**) distributed loading with $F_d^0 = 10$ (**b**) concentrated loading with $F_c^0 = 10$ and $\hat{x}_0 = 0.5$

Figure 7.8a–c presents the time response of mid-point of nanobeams based on the nonlinear analysis performed in Sect. 7.3.2 for different sets of boundary conditions. The increasing effect of nonlocal parameter on the amplitude of vibration for certain dimensionless excitation frequency, Ω, can be seen in this figure.

The effect of location of the concentrated force on the nonlinear frequency response is depicted in Fig. 7.9. It can be seen in the figure that the two branches of the nonlinear frequency response are closer to each other as the concentrated force moves towards the edges.

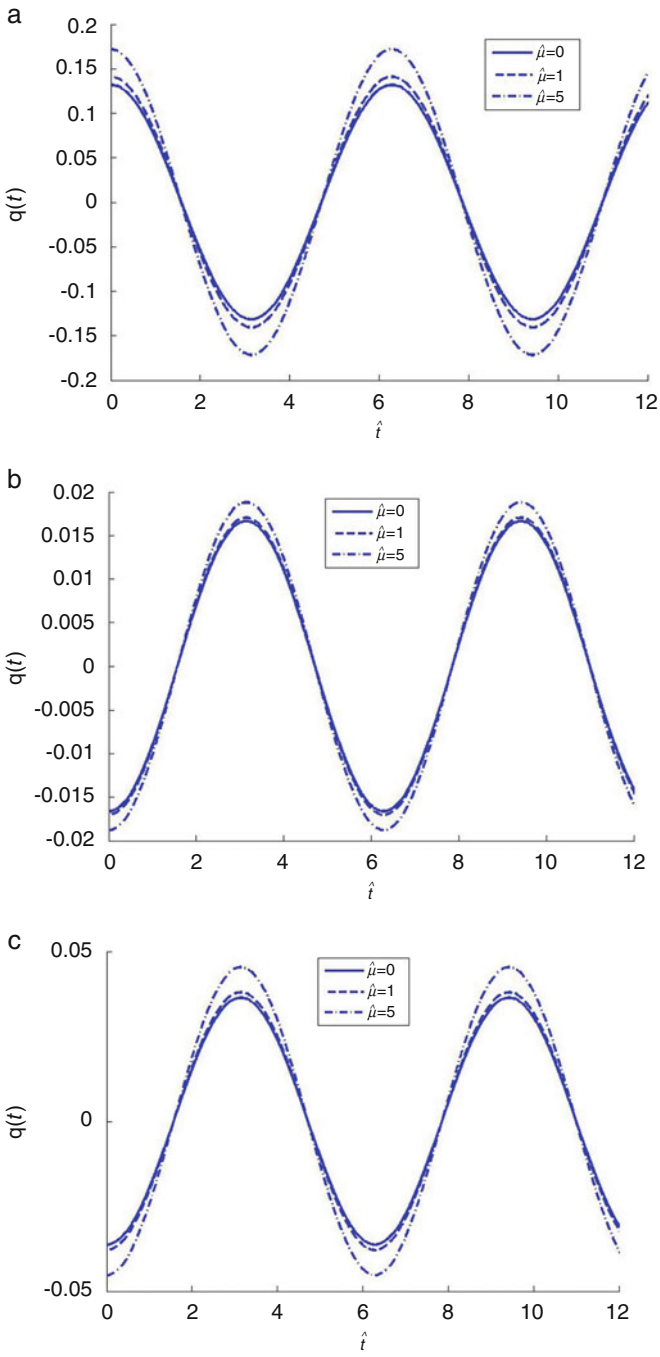



**Fig. 7.8** Time response of nanobeam subjected to distributed harmonic loading with $F_d^0 = 10$ and $\Omega = 1$. (**a**) *SS* (**b**) *CC* (**c**) *CS*

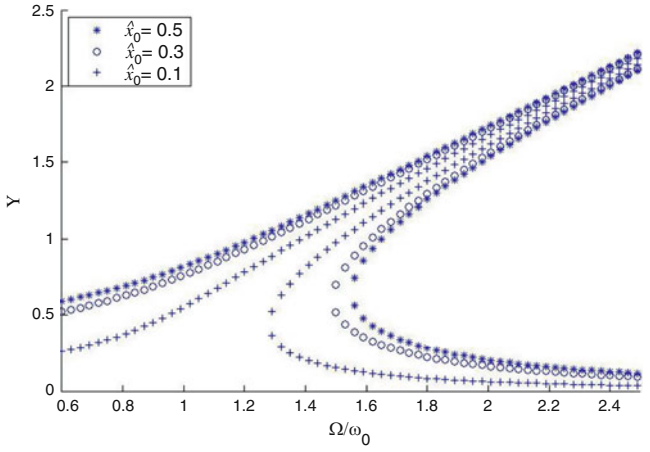



**Fig. 7.9** Nonlinear frequency response of *SS* nanobeam subjected to concentrated harmonic loading with $F_c^0 = 5$ and $\hat{\mu} = 0.5$

Finally, in Fig. 7.10a, b, the variation of amplitude of vibration versus nonlocal parameter for different types of excitation frequency is depicted. It is seen that the dimensionless vibration amplitude varies in opposite way for low and high nonlocal parameters. This is due to the fact that linear natural frequency is dependent on nonlocal parameter, $\widehat{\mu}$, therefore, there is critical value for $\widehat{\mu}$ where the excitation frequency is the same as linear natural frequency.

## 7.5 Conclusion

In this article, the effects of nonlocal stress field on the beam excited with distributed and concentrated harmonic loads are investigated. The Euler–Bernoulli assumptions together with Von Karman geometric type nonlinearity are used to formulate the strain field. The stress field is written based on Eringen's nonlocal theory. Applying the Hamilton principle, the governing partial differential equation is obtained. Implementing Galerkin technique together with Ritz averaging method, closed form solutions are presented for both time and frequency responses of the beam. In results, it is demonstrated that a beam with higher nonlocal parameter exhibits more nonlinear behavior. It is also shown that, the effect of nonlocality on the vibration amplitude is different for various values of excitation frequency.

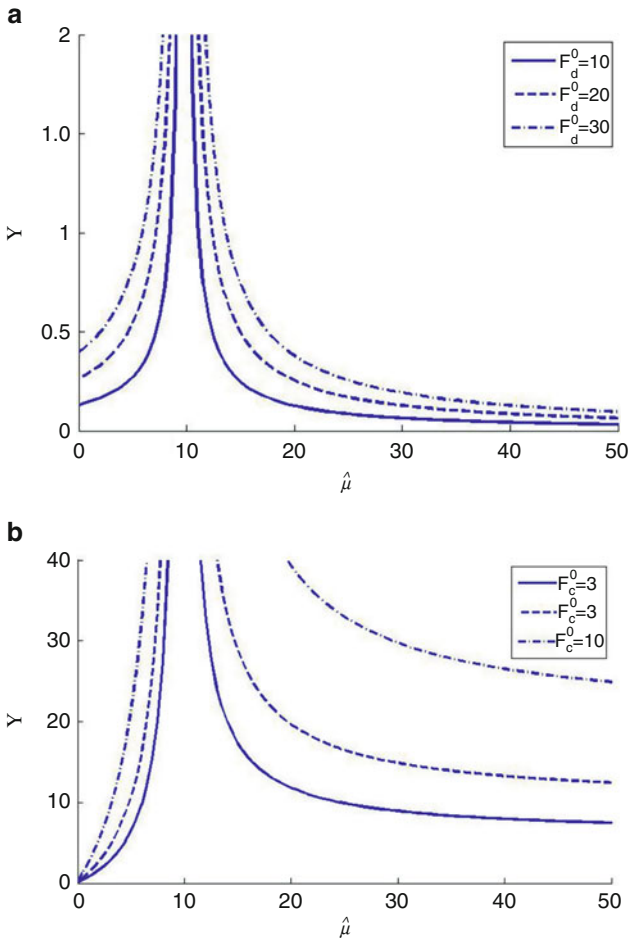



**Fig. 7.10** Variation of linear non-dimensional amplitude, $Y$, versus nonlocal parameter, $\hat{\mu}$, for *SS* beam subjected to (**a**) distributed loading with excitation frequency $\Omega = 1$ (**b**) concentrated loading with $\hat{x}_0 = 0.5$ with excitation frequency $\Omega = 1$

## References

Adhikari, S., Murmu, T., & McCarthy, M. A. (2013). Dynamic finite element analysis of axially vibrating nonlocal rods. *Finite Elements in Analysis and Design, 63*, 42–50.

Aifantis, E. C. (1999). Strain gradient interpretation of size effects. *International Journal of Fracture, 95*, 299–314.

Ansari, R., Rouhi, H., & Sahmani, S. (2011). Calibration of the analytical nonlocal shell model for vibrations of double-walled carbon nanotubes with arbitrary boundary conditions using molecular dynamics. *International Journal of Mechanical Sciences, 53*, 786–792.

Aydogdu, M. (2009). A general nonlocal beam theory: Its application to nanobeam bending, buckling and vibration. *Physica E: Low-dimensional Systems and Nanostructures, 41*, 1651–1655.

Azrar, L., Benamar, R., & White, R. G. (1999). Semi-analytical approach to the non-linear dynamic response problem of s–s and c–c beams at large vibration amplitudes part I: General theory and application to the single mode approach to free and forced vibration analysis. *Journal of Sound and Vibration, 224*, 183–207.

Azrar, L., Benamar, R., & White, R. G. (2002). A semi-analytical approach to the non-linear dynamic response problem of beams at large vibration amplitudes, part II: Multimode approach to the steady state forced periodic response. *Journal of Sound and Vibration, 255*, 1–41.

Benamar, R., Bennouna, M. M. K., & White, R. G. (1991). The effects of large vibration amplitudes on the mode shapes and natural frequencies of thin elastic structures part I: Simply supported and clamped-clamped beams. *Journal of Sound and Vibration, 149*, 179–195.

Bennett, J. A., & Eisley, J. G. (1970). A multiple degree-of-freedom approach to nonlinear beam vibrations. *AIAA Journal, 8*, 734–739.

Busby, H. R., Jr., & Weingarten, V. I. (1972). Non-linear response of a beam to periodic loading. *International Journal of Non-Linear Mechanics, 7*, 289–303.

Duan, W. H., Wang, C. M., & Zhang, Y. Y. (2007). Calibration of nonlocal scaling effect parameter for free vibration of carbon nanotubes by molecular dynamics. *Journal of Applied Physics, 101*, 024305–024307.

Eltaher, M. A., Alshorbagy, A. E., & Mahmoud, F. F. (2013). Vibration analysis of Euler–Bernoulli nanobeams by using finite element method. *Applied Mathematical Modelling, 37*, 4787–4797.

Eltaher, M. A., Emam, S. A., & Mahmoud, F. F. (2012). Free vibration analysis of functionally graded size-dependent nanobeams. *Applied Mathematics and Computation, 218*, 7406–7420.

Eringen, A. C. (2010). *Nonlocal continuum field theories*. Berlin: Springer.

Genta, G. (2009). Forced response of conservative nonlinear systems. In G. Genta (Ed.), *Vibration dynamics and control* (pp. 481–499). New York: Springer.

Gurtin, M. E., Weissmüller, J., & Larché, F. (1998). A general theory of curved deformable interfaces in solids at equilibrium. *Philosophical Magazine A, 78*, 1093–1109.

Hosseini, S. M., Mareishi, S., Kalhori, H., & Rafiee, M. (2012). Large amplitude free and forced oscillations of functionally graded beams. *Mechanics of Advanced Materials and Structures, 21*, 255–262.

Kanani, A. S., Niknam, H., Ohadi, A. R., & Aghdam, M. M. (2014). Effect of nonlinear elastic foundation on large amplitude free and forced vibration of functionally graded beam. *Composite Structures, 115*, 60–68.

Kanwal, R. P. (1998). *Generalized functions theory and technique: Theory and technique*. Berlin: Springer.

Karaoglu, P., & Aydogdu, M. (2010). On the forced vibration of carbon nanotubes via a non-local Euler—Bernoulli beam model. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 224*, 497–503.

Kiani, K. (2010). Longitudinal and transverse vibration of a single-walled carbon nanotube subjected to a moving nanoparticle accounting for both nonlocal and inertial effects. *Physica E: Low-dimensional Systems and Nanostructures, 42*, 2391–2401.

Lei, Y., Murmu, T., Adhikari, S., & Friswell, M. I. (2013). Dynamic characteristics of damped viscoelastic nonlocal Euler–Bernoulli beams. *European Journal of Mechanics - A/Solids, 42*, 125–136.

Lewandowski, R. (1997a). Computational formulation for periodic vibration of geometrically nonlinear structures—part 1: Theoretical background. *International Journal of Solids and Structures, 34*, 1925–1947.

Lewandowski, R. (1997b). Computational formulation for periodic vibration of geometrically nonlinear structures—part 2: Numerical strategy and examples. *International Journal of Solids and Structures, 34*, 1949–1964.

Mei, C., & Decha-Umphai, K. (1985). A finite element method for non-linear forced vibrations of beams. *Journal of Sound and Vibration, 102*, 369–380.

Niknam, H., & Aghdam, M. M. (2015). A semi analytical approach for large amplitude free vibration and buckling of nonlocal FG beams resting on elastic foundation. *Composite Structures, 119*, 452–462.

Reddy, J. N. (2007). Nonlocal theories for bending, buckling and vibration of beams. *International Journal of Engineering Science, 45*, 288–307.

Reddy, J. N., & Pang, S. D. (2008). Nonlocal continuum theories of beams for the analysis of carbon nanotubes. *Journal of Applied Physics, 103*, 023611-1:16.

Ribeiro, P., & Petyt, M. (1999). Non-linear vibration of beams with internal resonance by the hierarchical finite-element method. *Journal of Sound and Vibration, 224*, 591–624.

Shi, Y., Lee, R. Y. Y., & Mei, C. (1997). Finite element method for nonlinear free vibrations of composite plates. *AIAA Journal, 35*, 159–166.

Shooshtari, A., & Rafiee, M. (2011). Nonlinear forced vibration analysis of clamped functionally graded beams. *Acta Mechanica, 221*, 23–38.

Şimşek, M. (2010a). Non-linear vibration analysis of a functionally graded Timoshenko beam under action of a moving harmonic load. *Composite Structures, 92*, 2532–2546.

Şimşek, M. (2010b). Vibration analysis of a functionally graded beam under a moving mass by using different beam theories. *Composite Structures, 92*, 904–917.

Şimşek, M. (2010c). Vibration analysis of a single-walled carbon nanotube under action of a moving harmonic load based on nonlocal elasticity theory. *Physica E: Low-dimensional Systems and Nanostructures, 43*, 182–191.

Şimşek, M. (2011). Nonlocal effects in the forced vibration of an elastically connected double-carbon nanotube system under a moving nanoparticle. *Computational Materials Science, 50*, 2112–2123.

Şimşek, M. (2014). Large amplitude free vibration of nanobeams with various boundary conditions based on the nonlocal elasticity theory. *Composites Part B: Engineering, 56*, 621–628.

Şimşek, M., & Kocatürk, T. (2009). Free and forced vibration of a functionally graded beam subjected to a concentrated moving harmonic load. *Composite Structures, 90*, 465–473.

Şimşek, M., & Yurtcu, H. H. (2013). Analytical solutions for bending and buckling of functionally graded nanobeams based on the nonlocal Timoshenko beam theory. *Composite Structures, 97*, 378–386.

Thai, H.-T. (2012). A nonlocal beam theory for bending, buckling, and vibration of nanobeams. *International Journal of Engineering Science, 52*, 56–64.

Thai, H.-T., & Vo, T. P. (2012). A nonlocal sinusoidal shear deformation beam theory with application to bending, buckling, and vibration of nanobeams. *International Journal of Engineering Science, 54*, 58–66.

Uymaz, B. (2013). Forced vibration analysis of functionally graded beams using nonlocal elasticity. *Composite Structures, 105*, 227–239.

Valcárcel, M., Simonet, B. M., & Cárdenas, S. (2008). Analytical nanoscience and nanotechnology today and tomorrow. *Analytical and Bioanalytical Chemistry, 391*, 1881–1887.

Wang, Y.-G., Lin, W.-H., & Liu, N. (2012). A homotopy perturbation-based method for large deflection of a cantilever beam under a terminal follower force. *International Journal for Computational Methods in Engineering Science and Mechanics, 13*, 197–201.

Yang, F., Chong, A. C. M., Lam, D. C. C., & Tong, P. (2002). Couple stress based strain gradient theory for elasticity. *International Journal of Solids and Structures, 39*, 2731–2743.

# Chapter 8
# Analytical Solutions for Generalized Duffing Equation

**Mohammad M. Aghdam and Ali Fallah**

**Abstract** Structures like beams, plates, shells, and sectors have a key role in different fields of science and engineering. Reliable design of structures requires detailed understanding of their behavior in different service conditions including statics, dynamics, and vibration. For instance, dynamics and vibration characteristics of structures such as natural frequencies, mode shapes, and damping ratios are among interesting properties. Furthermore, in most severe conditions, linear theories are not enough to accurately capture real structural behavior. Thus, investigation of nonlinear and/or large amplitude vibration of structures which is normally governed by system of nonlinear differential equations is necessary.

In this chapter, large amplitude vibration of engineering structures such as beams, plate, and sectors is investigated. In all cases, the vibrational behavior of the structure is governed by the well-known Duffing equation. In addition to classical cubic nonlinearity in Duffing equation, inclusion of quadratic nonlinearity in the governing equation is also necessary for some special cases. Therefore, the generalized form of Duffing equation with both odd and even nonlinearities should be considered. In this chapter, an approximate analytical solution for generalized Duffing equation with both odd and even higher order nonlinearities is presented. Validity of the presented solution is examined through some numerical examples.

**Keywords** Duffing equation • Nonlinear differential equation • He's variational method • Perturbation theory

M.M. Aghdam (✉)
Department of Mechanical Engineering, Thermoelasticity Center of Excellence, Amirkabir University of Technology, Hafez Ave, Tehran, Iran
e-mail: aghdam@aut.ac.ir

A. Fallah
Department of Mechanical Engineering, Sharif University of Technology, Tehran, Iran
e-mail: a_fallah@mech.sharif.edu



R.N. Jazar, L. Dai (eds.), *Nonlinear Approaches in Engineering Applications*,
DOI 10.1007/978-3-319-27055-5_8

## 8.1 Introduction

The Duffing equation is a common and prevailing mathematical model for modeling nonlinear phenomenon such as engineering structures oscillation. The common form of Duffing equation is recognized with a cubic nonlinear term which is added to governing equation of motion of a simple mass–spring oscillator, as (Nayfeh and Mook 2007):

$$m\ddot{u} + Ku + K'u^3 = F(t) \tag{8.1}$$

where $m$, $K$, $K'$, and $F(t)$ are mass of oscillator, linear spring stiffness, nonlinear spring stiffness, and external time-dependent force, respectively (Fig. 8.1).

Furthermore, in some cases the governing equation of motion of oscillators is more general and complicated than common Duffing equation Eq. (8.1). For example, in large amplitude vibration of laminated beam (Baghani et al. 2011; Fallah and Aghdam 2012a) or functionally graded beams (Fallah and Aghdam 2011; Fallah and Aghdam 2012b), quadratic nonlinearity also appears in the governing equation due to material inhomogeneity of these structures. Moreover, it can be shown that the nonlinear governing equation of motion of any conservative oscillator can be transformed to generalized form of Duffing equation which has both higher order odd and even nonlinear terms. Thus, solution of generalized Duffing equation is desired by research community.

Like most of nonlinear differential equations, obtaining an analytical exact solution for Duffing equation is either impossible or very complicated whenever it is possible. Thus, just remain only two options on the table for solution of this equation, numerical methods and approximate closed form solutions.

Numerical solutions are very common in nonlinear analysis. There are various types of methods which are employed in nonlinear analysis such as finite element

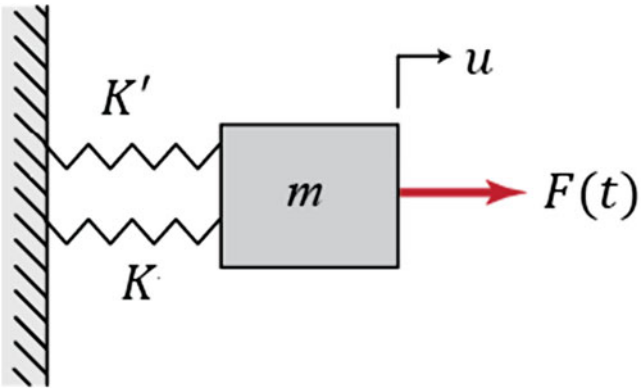



**Fig. 8.1** Nonlinear mass–spring system

method and finite difference method in conjunction with Newton–Raphson method. For the case of nonlinear initial value ordinary differential equation (ODE) like Duffing equation, some other methods are proposed. For example, methods based on Taylor series, one-step methods such as Euler method, multi-stage single-step methods such as Runge–Kutta methods, and linear multi-step methods (Aghdam et al. 2015).

Although the numerical methods solve the complex problems some time easier than other methods, they could not help when a parametric study is desired. For example, consider nonlinear oscillation of a structure like beam. If the time response of beam displacement at any point is interested, numerical solution can be used easily. But, if the aim of analysis is to find nonlinear oscillation frequency and its relation with beam geometrical parameter, the numerical methods are useless. Thus, closed form solutions even if they are approximate solutions are more preferred due to various advantages such as ease of parametric studies and considering of physics of the problem.

Various approximate analytical solution methods are proposed for the solution of nonlinear initial value ODEs up to now. For example, one can mention multiple time scale method and Lindstedt–Poincaré method. In these methods perturbation technique is employed to expand the field variable or natural frequency in terms of some time-dependent functions and a very small arbitrary parameter which is called bookkeeping parameter. For example, the displacement in these two methods is expanded as:

$$u(t) = \epsilon u_1(t) + \epsilon^2 u_2(t) + \cdots + \epsilon^n u_n(t) \tag{8.2}$$

where $u_n(t)$ and $\epsilon$ are time-dependent functions and bookkeeping parameter, respectively. Substitution of the considered series in the main equation and collecting the term with equal power of $\epsilon$ lead to a system of linear ODE which can be solved one after another for $u_n(t)$. More detail about the solution process of these methods can be found in Nayfeh and Mook (2007).

In these methods, it is assumed that the vibration amplitude is small but finite, thus, if vibration amplitude becomes large the results of these methods are no longer valid. Moreover, existence of bookkeeping parameter which is a physical interpretation less variable in the final form of the response is not desired. In order to overcome this shortage some new techniques are proposed, for instance, homotopy perturbation method (He 2005, 2006; Cai et al. 2006; Rafei and Ganji 2006; Siddiqui et al. 2006), the modified Lindstedt–Poincaré method (He 2002; Liu 2005) which are more accurate and can be used for moderately large amplitude vibration. Although the solution of the obtained linear ODEs in the mentioned methods is much simpler than solution of the main nonlinear ODE, this process becomes tedious and complicated in some cases.

Another types of approximate analytical solution methods are introduced in recent decades which are technically based on the variational methods. For example, one can mention the variational iteration method (He and Wu 2006, 2007a; Momani and Abuasad 2006), the harmonic balance method (Beléndez et al. 2007; Chen and

Liu 2007), the energy balance method (He 2003; D'Acunto 2006), the parameter-expansion method (He 2001; Shou and He 2007; Tao 2009), and He's variational method (He 2004, 2007b; Liu 2009). The solution process of these new methods is much simpler with respect to old methods. Some of these methods such as He's variational method and energy balanced method can be only used for conservative systems while other such as variational iteration method can be used for general type of nonlinear initial value ODE. Since these methods use variational method instead of perturbation method, the bookkeeping parameter is not included in the solution process and the final form of the response. In this study, application of He's variational method to obtain solution for the Duffing equation is examined.

In this chapter, based on the He's variational method simple and accurate analytical expression for time response and nonlinear frequency of an oscillator with generalized Duffing equation of motion is presented. Validity of the presented solution is examined through some numerical examples.

## 8.2 He's Variational Method

Consider a general conservative nonlinear oscillator in the form of:

$$\ddot{u} + f(u) = 0,\ u(0) = A,\ \dot{u}(0) = 0 \tag{8.3}$$

Variational form of Eq. (8.3) can be obtained using the semi-inverse method (He 2007b):

$$J(u) = \int_0^{T/4} \left\{ -\frac{1}{2}\dot{u}^2 + F(u) \right\} dt \tag{8.4}$$

where $T$ is period of the nonlinear oscillator and $\partial F/\partial u = f(u)$. Considering the initial conditions, the approximate solution of Eq. (8.3) can be considered as:

$$u(t) = A\ \cos(\omega t) \tag{8.5}$$

where $a$ and $\omega$ are the amplitude and frequency of the oscillator, respectively. Substituting Eq. (8.5) into Eq. (8.4) results in:

$$J(A, \omega) = \int_0^{\frac{T}{4}} \left\{ -\frac{1}{2}A^2\omega^2 \sin^2(\omega t) + F(A\cos(\omega t)) \right\} dt \tag{8.6}$$

Under the transformation $\tau = \omega t$, Eq. (8.6) can be written as:

$$J(A,\omega) = \frac{1}{\omega}\int_0^{\frac{\pi}{2}} \left\{-\frac{1}{2}A^2\omega^2 \sin^2(\tau) + F(A\cos(\tau))\right\} d\tau \tag{8.7}$$

Based on the Ritz method, in order to determine $\omega$ the conditions $\frac{\partial J}{\partial A} = 0$; $\frac{\partial J}{\partial \omega} = 0$ must be satisfied. However, it is shown (He 2007b) that identifying relationship between the frequency and amplitude of a nonlinear oscillator requires:

$$\frac{\partial J}{\partial A} = 0 \tag{8.8}$$

## 8.3 Application of the He's Variational Method

The governing equation of motion of conservative system in general form is:

$$\ddot{x} + f(x) = 0 \tag{8.9}$$

where $f(x)$ is a general nonlinear function of $x(t)$. It is worth mentioning that motion of oscillator is oscillatory just in its center neighborhood, so it is necessary to find the center points of the governing ODE. To this end, the singular points which are the roots of $f(x)$ must be found at first. In order to find out that a singular point is either center point or saddle point, the sign of first derivative of $f$ with respect to its argument at that point must be evaluated. At the center points the first derivative of $f$ is positive while at saddle point it is negative (Nayfeh and Mook 2007).

The origin must be shifted to a center point like, $x = x_0$, as:

$$u = x - x_0 \tag{8.10}$$

Thus, the equation of motion Eq. (8.9) in new coordinate system is:

$$\ddot{u} + f(u + x_0) = 0 \tag{8.11}$$

In order to simplify the equation and solution process, the $f$ can be expanded around the center point as:

$$f(u + x_0) = \sum_{i=1}^{n} \propto_i u^i \tag{8.12}$$

where

$$\propto_i = \frac{1}{n!} f^{(n)}(x_0) \tag{8.13}$$

and $f^{(n)}$ denotes the $n$th derivative of $f$ with respect to its argument. Thus, the general form of nonlinear governing equation of motion of a conservative oscillator can be rewritten as:

$$\ddot{u} + \propto_1 u + \propto_2 u^2 + \propto_3 u^3 + \propto_4 u^4 + \propto_5 u^5 + \cdots + \propto_n u^n = \ddot{u} + \sum_{i=1}^{n} \propto_i u^i = 0 \tag{8.14}$$

which can be also called generalized Duffing equation. In continue, analytical solution for the generalized Duffing equation is presented based on the He's variational method.

With accordance to Eq. (8.4) variational form of Eq. (8.14) can be written as:

$$J(u) = \int_0^{T/4} \left\{ -\frac{1}{2}\dot{u}^2 + \sum_{i=1}^{n} \frac{1}{i+1} \propto_i u^{i+1} \right\} d\tau \tag{8.15}$$

Substituting approximate solution Eq. (8.5) into Eq. (8.15) variational form of Eq. (8.14) can be rewritten as:

$$J(A, \omega) = \frac{1}{\omega} \int_0^{\pi/2} \left\{ -\frac{1}{2} A^2 \omega^2 \sin^2(\tau) + \sum_{i=1}^{n} \frac{1}{i+1} \propto_i A^{i+1} \cos^{i+1}(\tau) \right\} d\tau \tag{8.16}$$

Stationary condition with respect to $A$ leads to:

$$\frac{\partial J}{\partial A} = \frac{1}{\omega} \int_0^{\pi/2} \left\{ -A\omega^2 \sin^2(\tau) + \sum_{i=1}^{n} \propto_i A^i \cos^{i+1}(\tau) \right\} d\tau = 0 \tag{8.17}$$

The nonlinear frequency can be determined from Eq. (8.17) as:

$$(\omega_{\mathrm{NL}})^2 = \frac{\sum_{i=1}^{n} \propto_i A^{i-1} \int_0^{\pi/2} \cos^{i+1}(\tau)}{\int_0^{\frac{\pi}{2}} \sin^2(\tau)\, d\tau} = \frac{4}{\pi} \sum_{i=1}^{n} \propto_i \beta_{i+1} A^{i-1} \tag{8.18}$$

in which $\beta_i$'s are defined as:

$$\beta_m = \int_0^{\pi/2} \cos^m(\tau) = \frac{\sqrt{\pi}}{2} \frac{\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{m+2}{2}\right)} \tag{8.19}$$

where $\Gamma$ is the Gamma function. With some mathematical simplification Eq. (8.19) can be written as:

$$\beta_m = \begin{cases} \frac{2\times4\times6\times\cdots\times2k}{3\times5\times7\times\cdots(2k+1)}, m = 2k+1 \\ \frac{1\times3\times5\times\cdots(2k-1)}{2\times4\times6\times\cdots\times2k}\frac{\pi}{2}, m = 2k \end{cases}; m \geq 2 \tag{8.20}$$

Therefore, nonlinear frequency and approximate time response can be readily determined as:

$$\omega = \sqrt{\frac{4}{\pi}\sum_{i=1}^{n}\propto_i \beta_{i+1}A^{i-1}} = \sqrt{\propto_1 + \frac{4}{\pi}\sum_{i=2}^{n}\propto_i \beta_{i+1}A^{i-1}} \tag{8.21}$$

$$u = A\ \cos\left(\sqrt{\propto_1 + \frac{4}{\pi}\sum_{i=2}^{n}\propto_i \beta_{i+1}A^{i-1}}\ t\right) \tag{8.22}$$

## 8.4 Results and Discussion

In previous sections analytical expressions for both nonlinear natural frequency and time response of generalized Duffing equation are presented which in spite of existence relations are very simple. The aim of this section is to demonstrate the accuracy and validity of these simple but powerful expressions for nonlinear analysis. Various examples are considered in this section to investigate the order of accuracy of the presented expressions for both large and small value of vibration amplitude. Results of presented expressions are compared and validated with either exact solutions whenever exists or previous works. Some practical problems are also included in the examples to prove that the presented expressions can be used in engineering problems.

### *8.4.1 Example 1*

At first, consider a simple pendulum as shown in Fig. 8.2 with the governing equation of motion as:

$$\ddot{\theta} + \frac{g}{L}\sin(\theta) = 0,\ \theta(0) = \theta_0,\ \dot{\theta}(0) = 0 \tag{8.23}$$

where $\theta$, $g$, $L$, and $\theta_0$ are pendulum angle with respect to vertical axis, acceleration of gravity, pendulum length, and initial angle of pendulum with respect to vertical axis, respectively.

In linear vibration, it is assumed that the pendulum disperses a very small angle from the vertical axis, thus the nonlinear term in the Eq. (8.23) can be approximated

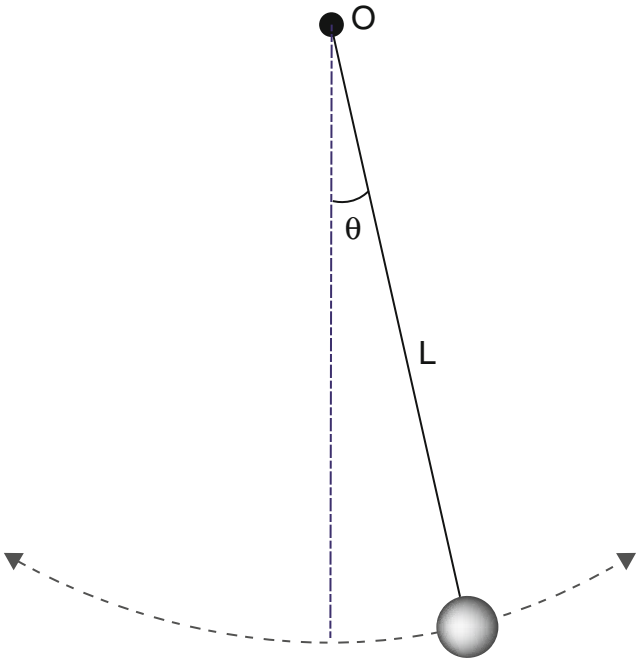



**Fig. 8.2** Simple pendulum configuration

as $\sin(\theta) \cong \theta$ and the main nonlinear ODE reduced to simple linear ODE, i.e. $\ddot{\theta} + \theta = 0$. However, for the case of large or moderately large values of $\theta$, this assumption is no longer valid, thus, the main nonlinear ODE must be solved.

The singular points of the nonlinear ODE Eq. (8.23) are:

$$\sin(\theta) = 0 \rightarrow \theta = 0, \pi, 2\pi, 3\pi \ldots \tag{8.24}$$

Evaluation of sign of first derivative of $f$ at singular point leads to:

$$\left.\frac{df}{d\theta}\right|_{\theta=0,2\pi,4\pi\ldots} = \cos(\theta)|_{\theta=0,2\pi,4\pi\ldots} = 1 > 0 \tag{8.25}$$

$$\left.\frac{df}{d\theta}\right|_{\theta=\pi,3\pi,5\pi\ldots} = \cos(\theta)|_{\theta=\pi,3\pi,5\pi\ldots} = -1 < 0 \tag{8.26}$$

Thus, points $\theta = 0, 2\pi, 4\pi$ are center points while points $\theta = \pi, 3\pi, 5\pi$ are saddle points. It is worth mentioning that from physical point of view, $\theta = 0, 2\pi, 4\pi$ are equal to one physical position, thus only $\theta = 0$ is considered in analysis procedure.

According to Eq. (8.13) with considering only the first seven terms of the expansion series, the new form of the Eq. (8.23) at $\theta = 0$ neighborhood is:

$$\ddot{\theta} + \propto_1 \theta + \propto_3 \theta^3 + \propto_5 \theta^5 + \propto_7 \theta^7 = 0 \tag{8.27}$$

where

$$\propto_i = \frac{g}{i!L} \frac{d^i (\sin(\theta))}{d\theta^i}\bigg|_{\theta=0} \tag{8.28}$$

Based on the Eq. (8.21) and Eq. (8.22), nonlinear natural frequency and time-dependent response of the pendulum are as:

$$\omega_{\mathrm{NL}} = \sqrt{\propto_1 + \frac{3}{4}\propto_3 {\theta_0}^2 + \frac{5}{8}\propto_5 {\theta_0}^4 + \frac{35}{64}\propto_7 {\theta_0}^6} \tag{8.29}$$

$$\theta(t) = \theta_0 \cos(\omega_{\mathrm{NL}} t) = \theta_0 \cos\left(\sqrt{\propto_1 + \frac{3}{4}\propto_3 {\theta_0}^2 + \frac{5}{8}\propto_5 {\theta_0}^4 + \frac{35}{64}\propto_7 {\theta_0}^6}\, t\right) \tag{8.30}$$

In order to investigate the accuracy and validity of Eq. (8.29) the relative error of nonlinear frequency for both small and large amplitude should be studied. Fortunately, for this case exact nonlinear frequency can be calculated as (Lima and Arun 2006):

$$\omega_{\mathrm{ex}} = \frac{2\pi}{T_{\mathrm{ex}}} \tag{8.31}$$

$$T_{\mathrm{ex}} = 4\sqrt{\frac{L}{2g}} \int_0^{\theta_0} \frac{1}{\sqrt{\cos(\theta) - \cos(\theta_0)}} d\theta \tag{8.32}$$

Since the integral in Eq. (8.32) cannot be evaluated analytically, some authors rewritten it in form of elliptic integrals and evaluate the $T_{\mathrm{ex}}$ using the existing numerical tables for value of elliptic integrals. On the other hand, the integral can be approximated with an acceptable accuracy using the following expression (Nelson and Olsson 1986):

$$T_{\mathrm{ex}} = 2\pi\sqrt{\frac{L}{g}}\left(1 + \frac{1}{16}\theta_0^2 + \frac{11}{3072}\theta_0^4 + + \frac{173}{737280}\theta_0^6 + \cdots\right) \tag{8.33}$$

Nonlinear natural frequency of pendulum for different values of $\theta_0$ together with its corresponding error with respect to exact value, i.e. $\frac{(\omega_{\mathrm{NL}} - \omega_{\mathrm{ex}}) \times 100}{\omega_{\mathrm{ex}}}$ is presented in Table 8.1. To examine the effect of number of terms of expansion series which is used on the results accuracy, nonlinear frequency obtained by considering 3, 5, and 7 terms of expansion series is reported in this table. Results show that for small value of amplitude the error is negligible. But as the amplitude increased, the value of error also increased as it was expected. Furthermore, from the results in can be concluded that to achieve accurate results even in large amplitude vibrations (Error

**Table 8.1** Nonlinear natural frequency of simple pendulum

| | $\theta_0 = 0.5$ | | $\theta_0 = 1.0$ | | $\theta_0 = 2.0$ | |
|---|---|---|---|---|---|---|
| $N$ | $\omega_{NL}$ | Error % | $\omega_{NL}$ | Error % | $\omega_{NL}$ | Error % |
| 3 | 3.0812 | 0.01 | 2.9283 | 0.26 | 2.2136 | 6.50 |
| 5 | 3.0817 | 0.00 | 2.9370 | 0.04 | 2.3910 | 0.99 |
| 7 | 3.0817 | 0.00 | 2.9368 | 0.03 | 2.3767 | 0.39 |

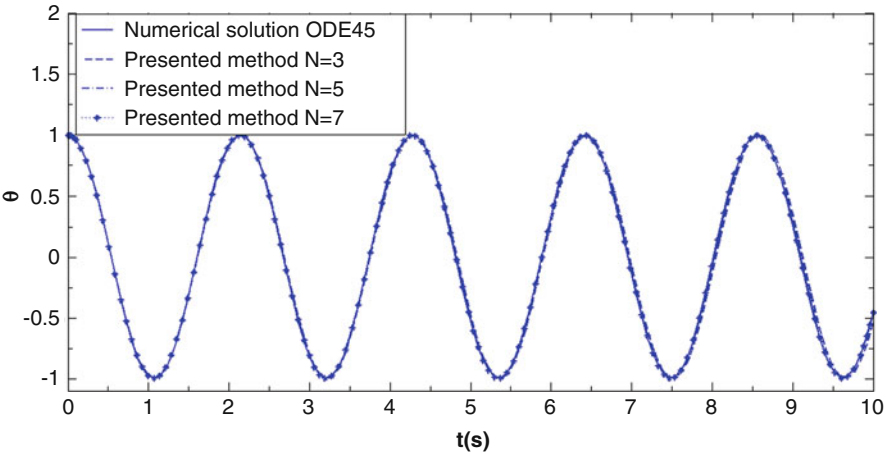



**Fig. 8.3** Time response of simple pendulum, $L = 1, g = 9.8, \theta_0 = 1.0$

<1 %), considering only the first five term of expansion series is adequate. Finally, it can be seen that, in spite of that the presented analytical expression for nonlinear natural frequency is very simple, it is valid and accurate.

Accuracy of time response obtained by Eq. (8.30) is also important. To check the time response accuracy, variation of $\theta$ with respect to time considering the first three, five, and seven terms of expansion series of $f$ in Eq. (8.23) is presented in Figs. 8.3 and 8.4 for $\theta_0 = 1.0$ and $\theta_0 = 2.0$, respectively. Numerical solution of Eq. (8.23) obtained by Runge–Kutta method (MATLAB ODE45) is also included for comparison. For $\theta_0 = 1.0$, results of presented method for $N = 3$, 5, and 7 are in good agreement with numerical solution. However, for larger vibration amplitude, i.e. $\theta_0 = 2.0$, the presented method with $N = 3$ leads to inaccurate time response while in case of $N = 5$, 7, time response is in good agreement with numerical solution. Thus, it can be concluded that considering the first five term of series expansion is enough to achieve acceptable accuracy for both nonlinear natural frequency and time response even in large vibration amplitude.

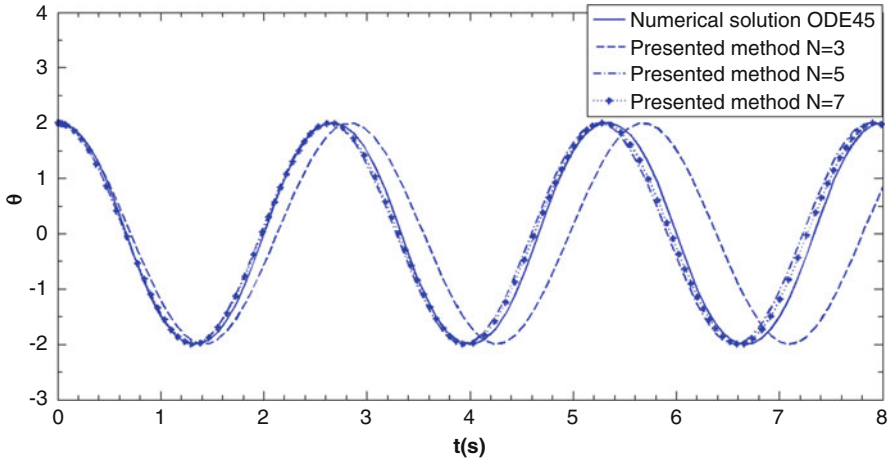



**Fig. 8.4** Time response of simple pendulum, $L = 1, g = 9.8,\ \theta_0 = 2.0$

### *8.4.2 Example 2*

Consider the following nonlinear equation:

$$\ddot{u} + \propto_1 u + \propto_3 u^3 + \cdots + \propto_n u^n = 0,\ u(0) = A,\ \dot{u}(0) = 0 \tag{8.34}$$

where $n > 1$ is an odd integer. For this case, nonlinear natural frequency is obtained using Eq. (8.21) as:

$$\omega_{\mathrm{NL}} = \sqrt{\propto_1 + \sum_{i=3,5,7,..}^{n} 2\frac{1 \times 3 \times 5 \times \cdots (2k-1)}{2 \times 4 \times 6 \times \cdots \times 2k} \propto_n A^{n-1}; i = 2k-1} \tag{8.35}$$

which is exactly the same as what Younesian et al. (2010) obtained by energy balanced method. It is noticeable to mention that for a nonlinear ODE which has only odd nonlinear terms like Eq. (8.34), the exact nonlinear frequency can be computed as (Younesian et al. 2010):

$$\omega_{\mathrm{ex}} = \frac{2\pi}{T_{\mathrm{ex}}} \tag{8.36}$$

$$T_{\mathrm{ex}} = 4\int_0^{\frac{\pi}{2}} 1/\sqrt{\sum_{i=1,3,5,..}^{n} \frac{1}{k} \propto_i A^{i-1} \left( \sum_{j=0,2,4}^{i-1} \sin^{\mathrm{j}}(\tau) \right)} d\tau;\ i = 2k-1 \tag{8.37}$$

For example, in the case of $n = 5$, exact frequency is calculated from:

$$\omega_{ex} = \frac{2\pi}{4\int_0^{\frac{\pi}{2}} \frac{1}{\sqrt{\propto_1 + \frac{1}{2}\left(1+\sin^2(\tau)\right)\propto_3 A^2 + \frac{1}{3}\left(1+\sin^2(\tau)+\sin^4(\tau)\right)\propto_5 A^4}} d\tau} \tag{8.38}$$

By the way, it must be noticed that the integral in Eqs. (8.37) and (8.38) cannot be evaluated analytically and must be solved with numerical methods.

The relative error of nonlinear natural frequency obtained by Eq. (8.38) for the large and small values of the amplitude can be, respectively, determined as:

$$\lim_{A\to\infty} \frac{\omega_{\mathrm{NL}}}{\omega_{\mathrm{ex}}} = 1.0585 \tag{8.39}$$

$$\lim_{A\to 0} \frac{\omega_{\mathrm{NL}}}{\omega_{\mathrm{ex}}} = 1 \tag{8.40}$$

Therefore, for any value of $A > 0$ the maximum error of nonlinear natural frequency obtained by Eq. (8.35) for $n = 5$, i.e. $\frac{|\omega_{\mathrm{NL}} - \omega_{\mathrm{ex}}| \times 100}{\omega_{\mathrm{ex}}}$ is less than 5.8 % which shows the excellent accuracy of the presented analytical expressions.

For $n = 3$, Eq. (8.34) turns to be the well-known classical Duffing equation. The nonlinear frequency in terms of amplitude of oscillation can be obtained as:

$$\omega = \sqrt{\propto_1 + \frac{3}{4}\propto_3 A^2} \tag{8.41}$$

which is exactly equal to relations obtained by the variational iteration method (He 2007a) and the parameter-expansion method (Shou and He 2007).

### *8.4.3 Example 3*

As another example, consider the following nonlinear ODE with both cubic and quadratic nonlinearity as:

$$\ddot{u} + \propto_1 u + \propto_2 u^2 + \propto_3 u^3 = 0 \tag{8.42}$$

This equation can be used for mathematical modeling of nonlinear free vibration of asymmetrically laminated beams (Baghani et al. 2011; Fallah and Aghdam 2012a) and functionally graded beams (Fallah and Aghdam 2011; Fallah and Aghdam 2012b). In spite of isotropic beams, material inhomogeneity leads to an additional quadratic term with respect to classical Duffing equation.

It is obvious that $u = 0$ is center point of this equation, thus the origin is located on the center point already. Using Eq. (8.21), nonlinear frequency can be obtained as:

**Table 8.2** Frequency ratio ($\omega_{NL}/\omega_L$) of simply supported functionally graded beam

| $A$ | $\omega_{NL}/\omega_L$ | |
|---|---|---|
| | Present | Gunda et al. (2011) |
| 0 | 1.000 | 1.000 |
| 0.5 | 1.008 | 1.008 |
| 1.0 | 1.048 | 1.048 |
| 1.5 | 1.117 | 1.118 |
| 2 | 1.210 | 1.211 |

$$\omega_{NL} = \sqrt{\propto_1 + \frac{8}{3\pi}\propto_2 A + \frac{3}{4}\propto_3 A^2} = \sqrt{\propto_1 + 0.8488\propto_2 A + 0.75\propto_3 A^2} \tag{8.43}$$

$$\omega_L = \sqrt{\propto_1} \tag{8.44}$$

On the other hand, Gunda et al. (2011) employed Rayleigh–Ritz method for this problem and obtained nonlinear frequency as:

$$\omega = \sqrt{\propto_1 + \frac{5}{6}\propto_2 A + \frac{3}{4}\propto_3 A^2} = \sqrt{\propto_1 + 0.8333\propto_2 A + 0.75\propto_3 A^2} \tag{8.45}$$

which is seemed to be in good agreement with the frequency obtained by presented expression. To investigate accuracy of results of the presented method, frequency ratio $\omega_{NL}/\omega_L$ for a simply supported functionally graded beam based on both Eqs. (8.43) and (8.45) is presented in Table 8.2. Quantitative analysis of results proved that the nonlinear frequency obtained by presented method and those obtained by Gunda et al. (2011) are in good agreement.

### *8.4.4 Example 4*

Consider the following nonlinear equation:

$$\ddot{u} - u + u^n = 0,\ u(0) = A,\ \dot{u}(0) = 0 \tag{8.46}$$

where $n > 1$ is a positive integer. It is obvious that $u = 1, 0$ are the singular points of the equation. It follows that:

$$\left.\frac{df}{du}\right|_{u=0} = nu^{n-1} - 1\big|_{u=0} = -1 < 0 \tag{8.47}$$

$$\left.\frac{df}{du}\right|_{u=1} = nu^{n-1} - 1\big|_{u=1} = n - 1 > 0 \tag{8.48}$$

Thus, $u = 0$ is saddle point while $u = 1$ is center. Thus the origin must be shifted to the center point using the following transformation, as:

$$v = u - 1 \tag{8.49}$$

substituting Eq. (8.49) into Eq. (8.46) leads to:

$$\ddot{v} + (n-1)\, v + \sum_{i=2}^{n} \binom{n}{i} v^n = 0,\ v(0) = \overline{A} = A - 1,\ \dot{v}(0) = 0 \tag{8.50}$$

which is a nonlinear equation with both odd and even nonlinearity. The nonlinear frequency can be obtained using Eq. (8.21). In case of $n = 3$, Eq. (8.50) is turn into:

$$\ddot{v} + 2v + 3v^2 + v^3 = 0(0) = \overline{A} = A - 1,\ \dot{v}(0) = 0 \tag{8.51}$$

The nonlinear frequency can be obtained as:

$$\omega_{\mathrm{NL}} = \sqrt{2 + \frac{8}{\pi}\overline{A} + \frac{3}{4}\overline{A}^2} \tag{8.52}$$

### *8.4.5 Example 5*

As a final example, the equation of the motion of the human eardrum is considered (Mickens 1981):

$$\ddot{u} - \alpha_1 u + \alpha_2 u^2 = 0,\ u(0) = A,\ \dot{u}(0) = 0 \tag{8.53}$$

It can be shown that the center point of the oscillation is $u = \alpha_1/\alpha_2$. Therefore, in order to shift the origin to the location of center, v is defined as:

$$v = u - \frac{\alpha_1}{\alpha_2} \tag{8.54}$$

Substituting Eq. (8.54) into Eq. (8.53) leads to:

$$\ddot{v} + \alpha_1 v + \alpha_2 v^2 = 0,\ v(0) = \overline{A} = A - \frac{\alpha_1}{\alpha_2},\ \dot{v}(0) = 0 \tag{8.55}$$

Thus, nonlinear frequency can be obtained based on the presented solution method by using Eq. (8.21) as:

$$\omega = \sqrt{\propto_1 + \frac{8}{3\pi} \propto_2 \overline{A}} \tag{8.56}$$

## 8.5 Conclusion

In this study, nonlinear analysis of the generalized form of Duffing equation which is one of the most famous and applicable nonlinear ODEs in engineering is considered. Moreover, it is shown that the governing equation of conservative systems can be also transformed to the form of generalized Duffing equation. Then, simple approximate analytical expressions for both nonlinear natural frequency and time response of generalized Duffing equation are presented. Using various examples the accuracy of the presented expressions is demonstrated. Results of the present study are validated and compared with available exact solutions and previous studies results.

## References

Aghdam, M. M., Ali, F., & Pouriya, H. (2015). Nonlinear initial value ordinary differential equations, Chapter 5. In *Nonlinear approaches in engineering applications*. Berlin: Springer.

Baghani, M., Jafari Talookolaei, R. A., & Salarieh, H. (2011). Large amplitudes free vibrations and post-buckling analysis of unsymmetrically laminated composite beams on nonlinear elastic foundation. *Applied Mathematical Modelling, 35*, 130–138.

Beléndez, A., Hernández, A., Beléndez, T., Álvarez, M. L., Gallego, S., & Ortuño, M. (2007). Application of the harmonic balance method to a nonlinear oscillator typified by a mass attached to a stretched wire. *Journal of Sound Vibration, 302*, 1018–1029.

Cai, X. C., Wu, W. Y., & Li, M. S. (2006). Approximate period solution for a kind of nonlinear oscillator by He's perturbation method. *International Journal of Nonlinear Sciences and Numerical Simulation, 7*, 109–112.

Chen, Y. M., & Liu, J. K. (2007). A new method based on the harmonic balance method for nonlinear oscillators. *Physics Letters A, 368*, 371–378.

D'Acunto, M. (2006). Determination of limit cycles for a modified van der Pol oscillator. *Mechanics Research Communications, 33*, 93–98.

Fallah, A., & Aghdam, M. M. (2011). Nonlinear free vibration and post-buckling analysis of functionally graded beams on nonlinear elastic foundation. *European J Mechanics-A/Solids, 30*, 571–583.

Fallah, A., & Aghdam, M. M. (2012a). Large amplitude free vibration and post buckling analysis of asymmetrically laminated beams on Pasternak foundation. *Nonlinear Engineering, 1*, 35–48.

Fallah, A., & Aghdam, M. M. (2012b). Thermo-mechanical buckling and nonlinear free vibration analysis of functionally graded beams on nonlinear elastic foundation. *Composites Part B: Engineering, 43*, 1523–1530.

Gunda, J. B., Gupta, R. K., Janardhan, G. R., & Rao, G. V. (2011). Large amplitude vibration analysis of composite beams: Simple closed-form solutions. *Composites Part B: Engineering, 93*, 870–879.

He, J. H. (2001). Bookkeeping parameter in perturbation methods. *Journal of Nonlinear Sciences and Numerical Simulation, 2*, 257–264.

He, J. H. (2002). Modified Lindstedt-Poincare methods for some strongly non-linear oscillations-part I: Expansion of a constant. *International Journal of Nonlinear Sciences and Numerical Simulation, 37*, 309–314.

He, J. H. (2003). Determination of limit cycles for strongly nonlinear oscillators. *Physical Reviews Letters, 90*, 174301–174304.

He, J. H. (2004). Variational principles for some nonlinear partial differential equations with variable coefficients. *Chaos Solitons Fractals, 19*, 847–851.

He, J. H. (2005). Application of homotopy perturbation method to nonlinear wave equations. *Chaos Solitons Fractals, 26*, 695–700.

He, J. H. (2006). Homotopy perturbation method for solving boundary value problems. *Physics Letters A, 350*, 87–88.

He, J. H. (2007a). Variational iteration method-Some recent results and new interpretations. *Journal of Computational and Applied Mathematics, 207*, 3–17.

He, J. H. (2007b). Variational approach for nonlinear oscillators. *Chaos Solitons Fractals, 34*, 1430–1439.

He, J. H., & Wu, X. H. (2006). Construction of solitary solution and compacton-like solution by variational iteration method. *Chaos Solitons Fractals, 29*, 108–113.

Lima, F. M. S., & Arun, P. (2006). An accurate formula for the period of a simple pendulum oscillating beyond the small angle regime. *American Journal of Physics, 74*, 892–895.

Liu, J. F. (2009). He's variational approach for nonlinear oscillators with high nonlinearity. *Computers & Mathematics with Applications, 58*, 2423–2426.

Liu, H. M. (2005). Approximate period of nonlinear oscillators with discontinuities by modified Lindstedt-Poincare method. *Chaos Solitons Fractals, 23*, 577–579.

Mickens, R. E. (1981). *An introduction to non-linear oscillations*. Cambridge: Cambridge University Press.

Momani, S., & Abuasad, S. (2006). Application of He's variational iteration method to Helmholtz equation. *Chaos Solitons Fractals, 27*, 1119–1123.

Nayfeh, A. H., & Mook, D. T. (2007). *Nonlinear oscillations*. New York: Wiley.

Nelson, R., & Olsson, M. G. (1986). The pendulum-rich physics from a simple system. *American Journal of Physics, 54*, 112–121.

Rafei, M., & Ganji, D. D. (2006). Explicit solutions of Helmholtz equation and fifth-order KdV equation using homotopy perturbation method. *International Journal of Nonlinear Sciences and Numerical Simulation, 7*, 321–328.

Shou, D. H., & He, J. H. (2007). Application of parameter-expanding method to strongly nonlinear oscillators. *Journal of Nonlinear Sciences and Numerical Simulation, 8*, 121–124.

Siddiqui, A. M., Mahmood, R., & Ghori, Q. K. (2006). Thin film flow of a third grade fluid on a moving belt by He's homotopy perturbation method. *International Journal of Nonlinear Sciences and Numerical Simulation, 7*, 7–14.

Tao, Z. L. (2009). Frequency–amplitude relationship of nonlinear oscillators by He's parameter-expanding method. *Chaos Solitons Fractals, 41*, 642–645.

Younesian, D., Askari, H., Saadatnia, Z., & KalamiYazdi, M. (2010). Frequency analysis of strongly nonlinear generalized Duffing oscillators using He's frequency_amplitude formulation and He's energy balance method. *Computers & Mathematics with Applications, 59*, 3222–3228.

# Part III
# Engineering Applications

# Chapter 9
# Nonlinear Deformations of Soft Tissues for Surgery Simulation

**Yongmin Zhong, Bijan Shirinzadeh, Julian Smith, Chengfan Gu, and Aleksandar Subic**

**Abstract** Soft tissue deformation is of great importance to virtual reality based surgery simulation. This paper presents a new methodology for modelling of nonlinear soft tissue deformation from the physicochemical viewpoint of soft tissues. This methodology converts soft tissue deformation into nonlinear chemical–mechanical interaction. Based on this, chemical diffusion of mechanical load and non-rigid mechanics of motion are combined to govern the dynamics of soft tissue deformation. The mechanical load applied to a soft tissue to cause a deformation is incorporated in chemical diffusion and distributed among mass points of the soft tissue. A chemical diffusion model is developed to describe the distribution of the mechanical load in the tissue. Methods are established for construction of the diffusion model on a 3D tissue surface and derivation of internal forces from the distribution of the mechanical load. Real-time interactive deformation of virtual human organs with force feedback has been achieved by the proposed methodology for surgery simulation. The proposed methodology not only accommodates isotropic, anisotropic and inhomogeneous materials by simply modifying diffusion coefficients, but it also accepts local and large-range deformation.

**Keywords** Soft tissue deformation • Virtual reality • Surgery simulation • Chemical diffusion and force feedback

Y. Zhong (✉) • C. Gu
School of Aerospace, Mechanical and Manufacturing Engineering, RMIT University, Melbourne, VIC, Australia
e-mail: yongmin.zhong@rmit.edu.au

B. Shirinzadeh
Department of Mechanical and Aerospace Engineering, Monash University, Melbourne, VIC, Australia

J. Smith
Department of Surgery, Monash Medical Centre, Monash University, Melbourne, VIC, Australia

A. Subic
Swinburne Research & Development, Swinburne University of Technology, Hawthorn, VIC, Australia
e-mail: aleksandar.subic@rmit.edu.au



R.N. Jazar, L. Dai (eds.), *Nonlinear Approaches in Engineering Applications*,
DOI 10.1007/978-3-319-27055-5_9

## 9.1 Introduction

Simulation of soft tissues is a challenging research topic in surgery simulation. Surgery simulation requires soft tissues react to the applied forces in a realistic fashion and in real time. However, due to the complex nonlinear behaviours of soft tissues, it is difficult to handle both of these conflicting requirements (Costa 2012; Zhu and Gu 2012; Zhong et al. 2012). Currently, most of the existing deformation methods are fully built on an elastic model. The common methods such as mass-spring (San-Vicente et al. 2012; Patete et al. 2013; Omar et al. in press), FEM (Finite Element Method) (Taylor et al. 2011; Zhang et al. 2014; Mafi and Sirouspour 2014) and BEM (Boundary Element Method) (Zhu and Gu 2012; Duarte et al. 2011) are mainly built on linear elasticity. Linear elastic models are simple and allow reduced runtime computations. However, they only allow displacements to be less than 10 % of the soft object size (Fung 1993), and thus cannot handle the nonlinear behaviours of soft tissues. Although nonlinear elastic models can handle nonlinear material properties, it is very difficult to satisfy the real-time requirement of soft tissue deformation due to the complexity and extremely expensive computational nature of nonlinear elastic models. So far, research on large-range nonlinear deformations of soft tissues is insufficient owing to the complex soft tissue behaviour and insufficient tissue response information (Taylor et al. 2011).

This paper presents a new methodology for modelling of nonlinear soft tissue deformation by using chemical–mechanical interaction. This methodology combines nonlinear chemical diffusion of mechanical load with non-rigid mechanics of motion to govern the simulation dynamics of soft tissue deformation. The mechanical load applied to a soft tissue to cause a deformation is incorporated in chemical diffusion as a chemical load, and is further distributed among mass points of the soft tissue. A chemical diffusion model is developed to describe the distribution of the mechanical load in the soft tissue. Numerical methods are presented for construction of the chemical diffusion model on a 3D tissue surface. A gradient method is established for deriving internal forces from the distribution of the mechanical load. Interactive deformation of anatomical models of human body with force feedback has been achieved for surgery simulation. Examples are presented to demonstrate the efficacy of the proposed methodology.

## 9.2 Related Work

A considerable amount of research efforts have been directed towards the area of soft tissue simulation. The most popular method is the mass-spring model (San-Vicente et al. 2012; Patete et al. 2013; Omar et al.), in which a deformable object is discretized into a system of mass points connected by springs. The advantage of the mass-spring model is that the computation is less time consuming

and the algorithm is easier to implement. However, the mass-spring model does not allow accurate modelling of material properties, and increasing the number of springs leads to a stiffer system.

FEM (Taylor et al. 2011; Zhang et al. 2014; Mafi and Sirouspour 2014) and BEM (Zhu and Gu 2012; Duarte et al. 2011) are the typical methods focused on accurate modelling of soft tissue deformation. In FEM or BEM, rigorous mathematical analysis based on continuum mechanics is applied to accurately model the mechanical behaviours, namely the stress–strain relationship of a soft tissue by decomposing it into a number of volume elements or boundary elements. However, these methods are computationally expensive and only the simplest variants such as linear shape functions have been used in soft tissue simulation. Although the computational performance can be improved by the explicit FEM (Zhang et al. 2014; Mafi and Sirouspour 2014) and pre-computation technique (Peterlik et al. 2010), the use of linear elasticity cannot accommodate nonlinear large-range deformation.

Studies were reported to handle large-range nonlinear deformation. These methods are not built on real nonlinear elasticity, and they are mainly focused on incorporation of nonlinear elements into linear elastic models to achieve large-range deformation. Picinbono et al. reported an explicit nonlinear FEM to handle large-range deformation by using nonlinear strains (Picinbono et al. 2003). However, the use of quadric strains generally requires very expensive computations for real-time simulation, and the runtime assembly of all the force terms for every element limits the interactivity to only a few hundred elements. Although the expensive computations caused by the use of the nonlinear strains can be handled by utilizing a subspace integration method (Barbic and James 2005), the use of nonlinear strains can only model geometric nonlinearity rather than nonlinear material properties. Schwartz et al. reported a FEM model by introducing two nonlinear Lame material constants for soft tissue simulation (Schwartz et al. 2005). However, the improvement does not comply with the constitutive laws of materials, since the Lame material constants are the inherent properties of materials and they are not changeable constants. The stiffness warping method was also reported to handle nonlinear large-range deformation (Etzmuss et al. 2003; Choi and Ko 2005; Cakir and Yazici 2009). However, the warping method can only handle geometric nonlinearity due to the use of the geometric improvement based on linear elasticity. Liu et al. reported a method by incorporating nonlinear internal forces into the linear mass-spring model to deal with nonlinear properties of soft tissues (Patete et al. 2013). Although it can handle large-range deformation, this method is still limited to the linear strain–stress relationship. Zhu and Gu also reported a hybrid deformable model, in which the traditional BEM was combined with the mass-spring model to deal with large-range deformation (Zhu and Gu 2012). However, due to the use of BEM, this method still replies on the pre-computation technique to achieve the real-time performance.

Recently, the authors studied soft tissue deformation by using a physical process, such as autowaves (Zhong et al. 2006a), cellular neural network (Zhong et al. 2006b), heat conduction (Zhong et al. 2010) and reaction–diffusion process

(Wu and Herzog 2002). These methods mainly rely on the analogies between elastic deformation and the nonlinear physical processes. Although experiments are conducted to verify the analogies, additional computational theoretical work is required to further support the analogies.

## 9.3 Chemical Diffusion of Mechanical Load

Behaviours of soft tissues are not only mechanically governed, but they also depend on physicochemical events (Lai et al. 1991). In fact, soft tissues form a system consisting of different particles such as mobile ions, and the soft tissue behaviours are associated with the ion concentration in the interstitium (Myers et al. 1984). When a rise in ion concentration is imposed on a tissue previously at equilibrium, ions diffuse into the tissue and move through the tissue matrix. The gradients of the concentrations are the driving forces for ion movement, and the movement of ions controls most of the compressive properties of soft tissues (Myers et al. 1984; Chou and Pagano 1967). Therefore, from the physicochemical viewpoint, soft tissue behaviours are an effect induced by ion diffusion activities.

According to the continuum theory, conservation of ions requires

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \tag{9.1}$$

where $\rho$ is the ion density at time $t$, $\nabla\cdot$ is the divergence operator and $\mathbf{v}$ is the average velocity of ion diffusion.

Since

$$\rho = MC \tag{9.2}$$

where $C$ is the ion concentration and $M$ is the ion mass.

Equation (9.1) becomes

$$\frac{\partial C}{\partial t} + \nabla \cdot (C\mathbf{v}) = 0 \tag{9.3}$$

The ion diffusion may be described by Fick's law

$$-\omega \nabla C = C\mathbf{v} \tag{9.4}$$

where $\omega$ is the diffusion coefficient and $\nabla$ is the gradient operator.

Combining Eq. (9.4) with Eq. (9.1) results in the governing equation of the chemical diffusion process

$$\frac{\partial C}{\partial t} = \nabla \cdot (\omega \nabla C) \tag{9.5}$$

If ions are generated by a chemical load, Eq. (9.5) becomes

$$\frac{\partial C}{\partial t} + H = \nabla \cdot (\omega \nabla C) \tag{9.6}$$

where $H$ represents the ion density under a chemical load.

When a soft tissue is deformed under a mechanical force, work is done by the mechanical force. The deformation is the consequence of applying the load generated by the mechanical force to the soft tissue. The load is absorbed in the soft tissue in the form of strain energy to deform the soft tissue away from its natural state (Chou and Pagano 1967; Sadd 2005). According to the first law of thermodynamics, the mechanical load applied to the soft tissue can be treated as an equivalent chemical load to inject ions into the tissue. Therefore, the ion density $H$ can be defined as the strain energy density

$$H = \int \boldsymbol{\sigma} \, d\varepsilon \tag{9.7}$$

where $\boldsymbol{\sigma}$ is the stress and $\boldsymbol{\varepsilon}$ is the strain at the contact point.

Since the mechanical load is only applied at the contact point, the resultant ion density is set only at the contact point where the mechanical load is applied, while the values of the ion density at other points are set to zero. In addition, it is not difficult to see from Eqs. (9.6) and (9.7) that soft tissue deformation is described as chemical diffusion of the mechanical load, which is applied to a soft tissue to cause a deformation.

To solve Eq. (9.6), it is necessary to determine the boundary conditions. The boundary conditions determine the character of the ion interchange over the boundary. Here, we choose the Neumann boundary condition as shown in Eq. (9.8), which implies that there is no energy loss at the boundary.

$$\frac{\partial C}{\partial \mathbf{N}_\Gamma} = 0 \tag{9.8}$$

where $\mathbf{N}_\Gamma$ is an outward normal vector at the boundary $\Gamma$.

## 9.4 Model Establishment

The chemical diffusion model can be constructed on a 3D surface or a 3D volume. For the sake of simplicity and without loss of generality, we consider the construction of the chemical diffusion model on a 3D surface.

The construction of the chemical diffusion model on a regular grid such as a rectangular grid is straightforward. The discrete chemical diffusion model on a rectangular grid can be easily established by using a finite difference scheme. For an internal point $\mathbf{P}_{i,j}$ in a rectangular grid, the discrete chemical diffusion model at point $\mathbf{P}_{i,j}$ may be written as

$$\begin{aligned} C_{i,j}(t+\Delta t)=C_{i,j}(t)&+\frac{2\omega C_{i+1,j}(t)\Delta t}{\left\|\overrightarrow{\mathbf{P}_{i,j}\mathbf{P}_{i+1,j}}\right\|\left(\left\|\overrightarrow{\mathbf{P}_{i-1,j}\mathbf{P}_{i,j}}\right\|+\left\|\overrightarrow{\mathbf{P}_{i,j}\mathbf{P}_{i+1,j}}\right\|\right)}+\frac{2\omega C_{i-1,j}(t)\Delta t}{\left\|\overrightarrow{\mathbf{P}_{i-1,j}\mathbf{P}_{i,j}}\right\|\left(\left\|\overrightarrow{\mathbf{P}_{i-1,j}\mathbf{P}_{i,j}}\right\|+\left\|\overrightarrow{\mathbf{P}_{i,j}\mathbf{P}_{i+1,j}}\right\|\right)}\\ &+\frac{2\omega C_{i,j+1}(t)\Delta t}{\left\|\overrightarrow{\mathbf{P}_{i,j}\mathbf{P}_{i,j+1}}\right\|\left(\left\|\overrightarrow{\mathbf{P}_{i,j-1}\mathbf{P}_{i,j}}\right\|+\left\|\overrightarrow{\mathbf{P}_{i,j}\mathbf{P}_{i,j+1}}\right\|\right)}+\frac{2\omega C_{i,j-1}(t)\Delta t}{\left\|\overrightarrow{\mathbf{P}_{i,j-1}\mathbf{P}_{i,j}}\right\|\left(\left\|\overrightarrow{\mathbf{P}_{i,j-1}\mathbf{P}_{i,j}}\right\|+\left\|\overrightarrow{\mathbf{P}_{i,j}\mathbf{P}_{i,j+1}}\right\|\right)}\\ &-\frac{2\omega C_{i,j}(t)\Delta t}{\left\|\overrightarrow{\mathbf{P}_{i-1,j}\mathbf{P}_{i,j}}\right\|\left\|\overrightarrow{\mathbf{P}_{i,j}\mathbf{P}_{i+1,j}}\right\|}-\frac{2\omega C_{i,j}(t)\Delta t}{\left\|\overrightarrow{\mathbf{P}_{i,j-1}\mathbf{P}_{i,j}}\right\|\left\|\overrightarrow{\mathbf{P}_{i,j}\mathbf{P}_{i,j+1}}\right\|}-H_{i,j}\Delta t \end{aligned} \tag{9.9}$$

where $\left\|\overrightarrow{\mathbf{P}_{i-1,j}\mathbf{P}_{i,j}}\right\|$ and other similar terms represent the magnitudes of vector $\overrightarrow{\mathbf{P}_{i-1,j}\mathbf{P}_{i,j}}$ and other similar vectors, and $\Delta t$ is a constant time step.

To construct the chemical diffusion model on an irregular grid such as a triangular grid, we subdivide the triangular grid into a finite number of non-overlapping control volumes, over which energy conservation is enforced in a discrete sense. Therefore, the construction of the chemical diffusion model on a triangular grid can be achieved by discretizing the chemical diffusion model on each control volume. Figure 9.1 shows a control volume (surrounded by the dot lines) constructed around point $\mathbf{P}_i$. The control volume consists of the centroids of the triangles adjacent to point $\mathbf{P}_i$ and the midpoints of the edges adjacent to point $\mathbf{P}_i$.

Considering Eq. (9.6) over the control volume shown in Fig. 9.1, and applying Gauss formula yield

$$\int_{S_i}\left(\frac{\partial C}{\partial t}+H\right)\mathrm{d}S=\int_{B_i}(\omega\nabla C)\cdot\mathbf{n}\mathrm{d}B \tag{9.10}$$

where $S_i$ is the measure of the control volume at point $\mathbf{P}_i$, $B_i$ is the closed boundary of $S_i$, "•" represents the dot product of two vectors and $\mathbf{n}$ is the outward unit normal vector at the boundary $B_i$.

The left side of Eq. (9.10) can be approximated as

$$\int_{S_i}\left(\frac{\partial C}{\partial t}+H\right)\mathrm{d}S=\left(\frac{\partial C_i}{\partial t}+H_i\right)S_i \tag{9.11}$$

Substituting Eq. (9.11) into Eq. (9.10) yields

$$\left(\frac{\partial C_i}{\partial t}+H_i\right)S_i=\int_{B_i}(\omega\nabla C_i)\cdot\mathbf{n}\mathrm{d}B \tag{9.12}$$

where $C_i$ is the ion concentration at point $\mathbf{P}_i$.

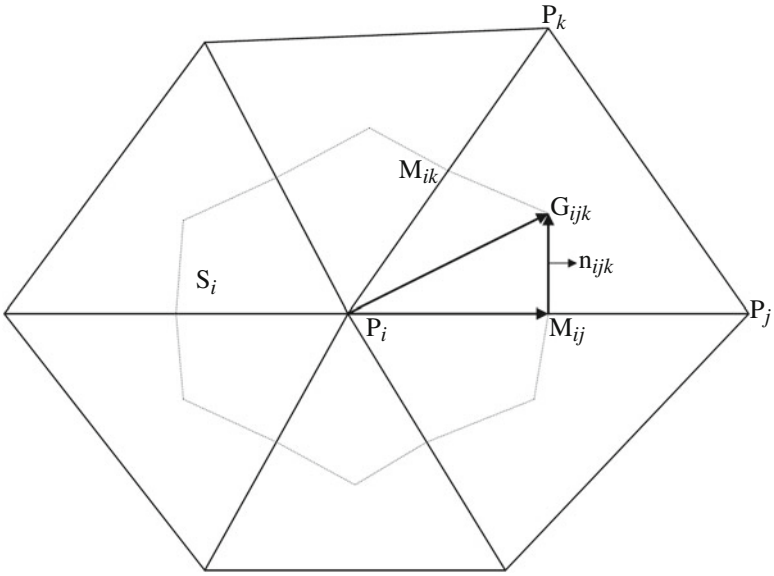



**Fig. 9.1** A control volume over an irregular grid: the control volume consists of the centroids of the triangles adjacent to point $\mathbf{P}_i$ and the midpoints of the edges adjacent to point $\mathbf{P}_i$

By discretizing $\frac{\partial C_i}{\partial t}$, Eq. (9.12) may be rewritten as

$$C_i\left(t+\Delta t\right)=C_i(t)+\frac{\Delta t}{S_i}\int_{B_i}\left(\omega\nabla C_i(t)\right)\cdot\mathbf{n}\mathrm{d}B-H_i\Delta t \tag{9.13}$$

By discretizing $B_i$, Eq. (9.13) becomes

$$C_i\left(t+\Delta t\right)=C_i(t)+\frac{\Delta t}{S_i}\sum_{i\in N\left(\mathbf{P}_i\right)}\sum_{k\in N\left(\mathbf{P}_j\right)\cap N\left(\mathbf{P}_i\right)}\int_{M_{ij}G_{ijk}}\left(\omega\nabla C_i(t)\right)\cdot\mathbf{n}_{ijk}\mathrm{d}B-H_i\Delta t \tag{9.14}$$

where $N(\mathbf{P}_i)$ is the set of the neighbouring points of point $\mathbf{P}_i$ and $N(\mathbf{P}_j)$ is the set of the neighbouring points of point $\mathbf{P}_j$.

The gradient $\nabla C_i(t)$ over the small control volume defined by the triangle $\Delta\mathbf{P}_i\mathbf{P}_j\mathbf{P}_k$ may be written as

$$\begin{aligned}\nabla C_i(t)=&\frac{C_j(t)-C_i(t)}{\left\|\overrightarrow{\mathbf{P}_i\mathbf{P}_j}\times\overrightarrow{\mathbf{P}_i\mathbf{P}_k}\right\|^2}\left(\left\|\overrightarrow{\mathbf{P}_i\mathbf{P}_k}\right\|^2\overrightarrow{\mathbf{P}_i\mathbf{P}_j}-\left(\overrightarrow{\mathbf{P}_i\mathbf{P}_j}\cdot\overrightarrow{\mathbf{P}_i\mathbf{P}_k}\right)\overrightarrow{\mathbf{P}_i\mathbf{P}_k}\right)\\&+\frac{C_k(t)-C_i(t)}{\left\|\overrightarrow{\mathbf{P}_i\mathbf{P}_j}\times\overrightarrow{\mathbf{P}_i\mathbf{P}_k}\right\|^2}\left(\left\|\overrightarrow{\mathbf{P}_i\mathbf{P}_j}\right\|^2\overrightarrow{\mathbf{P}_i\mathbf{P}_k}-\left(\overrightarrow{\mathbf{P}_i\mathbf{P}_j}\cdot\overrightarrow{\mathbf{P}_i\mathbf{P}_k}\right)\overrightarrow{\mathbf{P}_i\mathbf{P}_j}\right)\end{aligned} \tag{9.15}$$

where "×" denotes the cross product of two vectors.

Substituting (9.15) into (9.14), there is

$$C_i\left(t+\Delta t\right)=C_i(t)+\frac{\omega\Delta t}{6S_i}\sum_{j\in N_i}\sum_{k\in N_i\cap N_j}\lambda_{ijk}\left(C_j(t)-C_i(t)\right)+\mu_{ijk}\left(C_k(t)-C_i(t)\right)-H_i\Delta t \tag{9.16}$$

where

$$\begin{aligned}\lambda_{ijk}&=\frac{\left\|\overrightarrow{\mathbf{P}_i\mathbf{P}_k}+\overrightarrow{\mathbf{P}_j\mathbf{P}_k}\right\|}{\left\|\overrightarrow{\mathbf{P}_i\mathbf{P}_j}\times\overrightarrow{\mathbf{P}_i\mathbf{P}_k}\right\|^2}\left(\left\|\overrightarrow{\mathbf{P}_i\mathbf{P}_k}\right\|^2\overrightarrow{\mathbf{P}_i\mathbf{P}_j}-\left(\overrightarrow{\mathbf{P}_i\mathbf{P}_j}\cdot\overrightarrow{\mathbf{P}_i\mathbf{P}_k}\right)\overrightarrow{\mathbf{P}_i\mathbf{P}_k}\right)\cdot\mathbf{n}_{ijk}\\ \mu_{ijk}&=\frac{\left\|\overrightarrow{\mathbf{P}_i\mathbf{P}_k}+\overrightarrow{\mathbf{P}_j\mathbf{P}_k}\right\|}{\left\|\overrightarrow{\mathbf{P}_i\mathbf{P}_j}\times\overrightarrow{\mathbf{P}_i\mathbf{P}_k}\right\|^2}\left(\left\|\overrightarrow{\mathbf{P}_i\mathbf{P}_j}\right\|^2\overrightarrow{\mathbf{P}_i\mathbf{P}_k}-\left(\overrightarrow{\mathbf{P}_i\mathbf{P}_j}\cdot\overrightarrow{\mathbf{P}_i\mathbf{P}_k}\right)\overrightarrow{\mathbf{P}_i\mathbf{P}_j}\right)\cdot\mathbf{n}_{ijk}\end{aligned} \tag{9.17}$$

Finally, the unit normal vector at the boundary $\overrightarrow{\mathbf{M}_{ij}\mathbf{G}_{ijk}}$ of the small control volume may be written as

$$\mathbf{n}_{ijk}=\frac{\left(\overrightarrow{\mathbf{P}_i\mathbf{P}_k}+\overrightarrow{\mathbf{P}_j\mathbf{P}_k}\right)\times\left[\overrightarrow{\mathbf{P}_i\mathbf{P}_j}\times\left(\overrightarrow{\mathbf{P}_i\mathbf{P}_j}+\overrightarrow{\mathbf{P}_i\mathbf{P}_k}\right)\right]}{\left\|\left(\overrightarrow{\mathbf{P}_i\mathbf{P}_k}+\overrightarrow{\mathbf{P}_j\mathbf{P}_k}\right)\times\left[\overrightarrow{\mathbf{P}_i\mathbf{P}_j}\times\left(\overrightarrow{\mathbf{P}_i\mathbf{P}_j}+\overrightarrow{\mathbf{P}_i\mathbf{P}_k}\right)\right]\right\|} \tag{9.18}$$

## 9.5 Internal Forces and Deformation Dynamics

Since the gradient of the concentration is the driving force for ion movement, the internal force can be described as

$$\mathbf{F}=-\omega\nabla_{\mathbf{P}}C \tag{9.19}$$

where "$\nabla_P$" represents the gradient with respect to the change in position.

For a regular grid, the internal force at a point can be directly obtained by discretizing the gradient operator at the point using a finite difference scheme. Therefore, the internal force at point $P_i$ on a regular grid is:

$$\mathbf{F}_i=\omega\sum_{j\in N(\mathbf{P}_i)}\frac{\left|C_{\mathbf{P}_j}-C_{\mathbf{P}_i}\right|}{\left\|\overrightarrow{\mathbf{P}_i\mathbf{P}_j}\right\|}\mathbf{U}_{ij} \tag{9.20}$$

where $\mathbf{U}_{ij}=\frac{\overrightarrow{\mathbf{P}_i\mathbf{P}_j}}{\left\|\overrightarrow{\mathbf{P}_i\mathbf{P}_j}\right\|}$ and $\left|C_{\mathbf{P}_j}-C_{\mathbf{P}_i}\right|$ is the magnitude of the concentration change between point $\mathbf{P}_i$ and point $\mathbf{P}_j$.

For an irregular grid such as a triangular grid, since the discretization of the gradient operator is represented as Eq. (9.15), the internal force at point $\mathbf{P}_i$ is

$$\mathbf{F}_i = \omega \sum_{j \in N(\mathbf{P}_i)} \sum_{k \in N(\mathbf{P}_i) \cap N(\mathbf{P}_j)} \frac{C_j(t) - C_i(t)}{\left\| \overrightarrow{\mathbf{P}_i\mathbf{P}_j} \times \overrightarrow{\mathbf{P}_i\mathbf{P}_k} \right\|^2} \left( \left( \overrightarrow{\mathbf{P}_i\mathbf{P}_j} \cdot \overrightarrow{\mathbf{P}_i\mathbf{P}_k} \right) \overrightarrow{\mathbf{P}_i\mathbf{P}_k} - \left\| \overrightarrow{\mathbf{P}_i\mathbf{P}_k} \right\|^2 \overrightarrow{\mathbf{P}_i\mathbf{P}_j} \right)$$
$$+ \omega \sum_{j \in N(\mathbf{P}_i)} \sum_{k \in N(\mathbf{P}_i) \cap N(\mathbf{P}_j)} \frac{C_k(t) - C_i(t)}{\left\| \overrightarrow{\mathbf{P}_i\mathbf{P}_j} \times \overrightarrow{\mathbf{P}_i\mathbf{P}_k} \right\|^2} \left( \left( \overrightarrow{\mathbf{P}_i\mathbf{P}_j} \cdot \overrightarrow{\mathbf{P}_i\mathbf{P}_k} \right) \overrightarrow{\mathbf{P}_i\mathbf{P}_j} - \left\| \overrightarrow{\mathbf{P}_i\mathbf{P}_j} \right\|^2 \overrightarrow{\mathbf{P}_i\mathbf{P}_k} \right) \tag{9.21}$$

During the simulation, after computing the internal forces for each node, the Lagrange's equation of motion is used to update the node positions, and thus the following relationship may be written

$$m_i \frac{d^2\mathbf{P}_i}{dt^2} + \gamma_i \frac{d\mathbf{P}_i}{dt} + \mathbf{F}_i = \mathbf{G}_i \tag{9.22}$$

where $\mathbf{P}_i$ is the position vector of node $i$ at time $t$, $m_i$ and $\gamma_i$ are the mass and damping constants of node $i$, respectively, $\mathbf{F}_i$ is the net internal force applied to node $i$ at time $t$ and $\mathbf{G}_i$ is the external force applied to node $i$ at time $t$.

Equation (9.22) is solved by using an explicit integration scheme which does not require matrix inversion for updating each vertex position, and also has a simple implementation.

## 9.6 Implementation Results and Discussions

Experiments have been conducted to investigate the effect of isotropic, anisotropic and inhomogeneous deformation as well as nonlinear load-deformation of soft tissues. Interactive deformation of anatomical models of human body using a haptic device for surgery simulation is studied, and the comparison with the existing deformation methods is also discussed in this section.

### *9.6.1 Isotropic, Anisotropic, Inhomogeneous, Local and Large-Range Deformations*

Experiments have been conducted to investigate isotropic, anisotropic and inhomogeneous deformation. Figure 9.2 illustrates the deformation of the isotropic material with 900 grid points ($\omega = 0.08$, mass = 10.0 and damping = 10.0). As shown in Fig. 9.2a, b, the material has a cubic shape with round corners in its rest state. Figure 9.2c, d are two different views of the material deformed under a compressive force of 3 kg.

Anisotropic deformation can be achieved by setting different diffusion coefficients in different directions. Figure 9.3 illustrates the deformation of an anisotropic

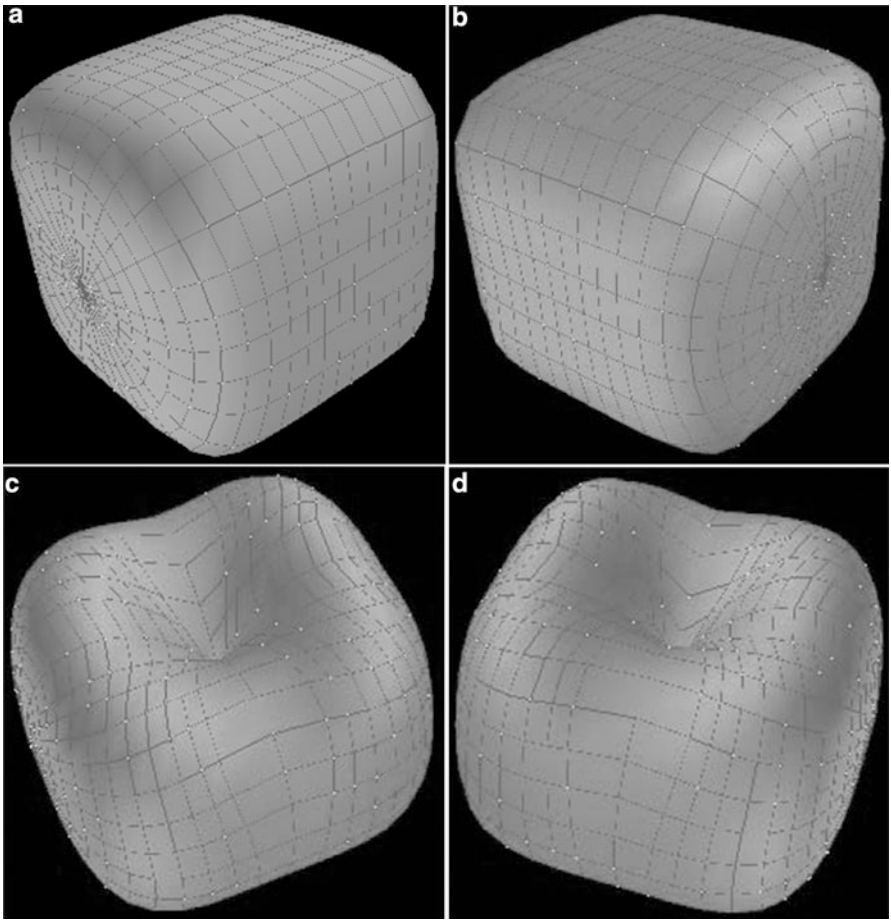



**Fig. 9.2** Deformation of an isotropic material with a cubic and round-corner shape in its rest state: the material has a common diffusion coefficient ($\omega = 0.08$) at each point

material under the same external force as Fig. 9.2. The material has the same parameters as the material in Fig. 9.2 except that diffusion coefficients are different in the different parametric directions ($\omega$ in the latitudinal direction $= 0.08$ and $\omega$ in the longitudinal direction $= 0.24$). Compared to the isotropic deformation shown in Fig. 9.2, Fig. 9.3 shows that the material is deformed more in the longitudinal direction. This demonstrates the proposed methodology can simulate anisotropic deformation by simply setting different diffusion coefficients in different directions.

Inhomogeneous deformation can also be simulated by setting different diffusion coefficients at different points. Figure 9.4 illustrates the deformation of an inhomogeneous material under the same external force as Fig. 9.2. The material has the same parameters as the material in Fig. 9.2 except that the white portions ($\omega = 0.08$)

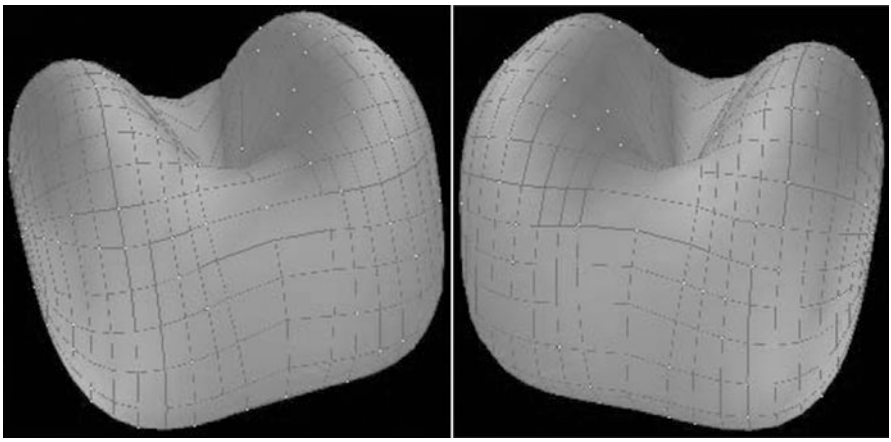


**Fig. 9.3** Deformation of an anisotropic material by setting different diffusion coefficients in different directions ($\omega$ in the latitudinal direction $= 0.08$ and $\omega$ in the longitudinal direction $= 0.24$)

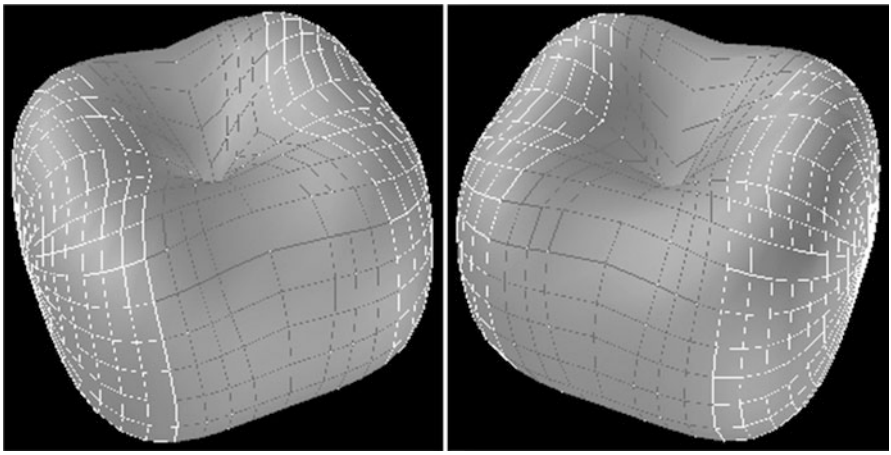


**Fig. 9.4** Deformation of an inhomogeneous material by setting different diffusion coefficients at different points ($\omega$ in the *white portions* $= 0.08$ and $\omega$ in the *grey portion* $= 0.24$)

have a different diffusion coefficient from the grey portion ($\omega = 0.24$). As shown in Fig. 9.4, some of the white portions are deformed correspondingly during the deformation, and they also highlight the difference from the isotropic deformation shown in Fig. 9.2. Compared to the anisotropic deformation shown in Fig. 9.3, Fig. 9.4 illustrates that the material is deformed differently at the different points with different diffusion coefficients.

Local deformation can also be achieved with the proposed methodology by setting a very low diffusion coefficient to the points where small deformation is

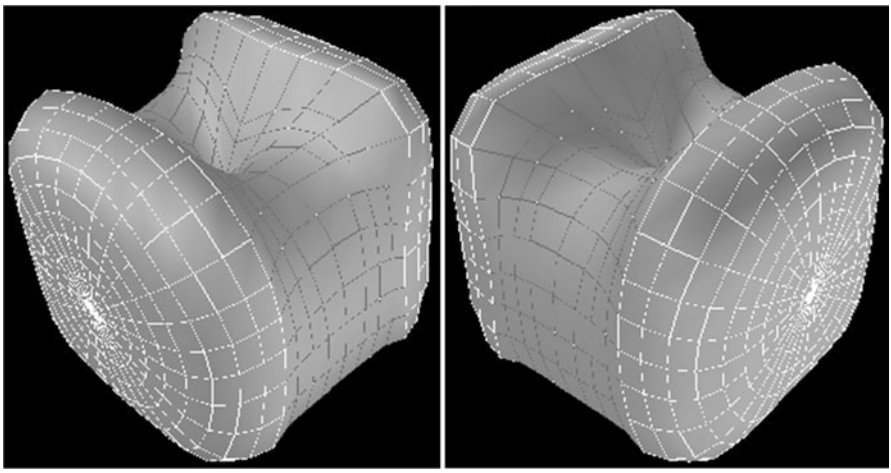


**Fig. 9.5** Local deformation by using a very low diffusion coefficient: the *white portions* with a very low diffusion coefficient are not deformed

expected. Figure 9.5 illustrates an example of local deformation under the same external force as Fig. 9.2, in which the white portions with a very low diffusion coefficient are not deformed.

The proposed methodology has been further verified with real soft tissues. Figure 9.6 illustrates the simulated deformation behaviours of the proposed methodology against the experimental deformation behaviours of a lamb kidney. It can be seen that the simulation curve is in agreement with the experimental curve, and both deformations vary nonlinearly with the applied forces. This demonstrates that the proposed methodology can exhibit the mechanical behaviours of soft tissues. The nonlinear load–deformation relationship also reveals that the proposed methodology can accommodate large-range deformation.

### *9.6.2 Human Organ Deformation with Haptic Feedback*

Laparoscopic surgery is much more difficult for surgeons than classical surgery. Surgeons lose the third dimension and use two long special surgical tools instead of their hands. Therefore, they must use all the clues provided to them to understand and reconstruct the scene in their minds. Force feedback is one of such clues, and its introduction in the surgery simulation greatly improves the overall realism of the simulation.

Real-time soft tissue simulation with force feedback for use in surgery simulation has been achieved by integration of a haptic device (PHANToM from Sensable Technologies) into the proposed methodology. The graphical rendering is achieved by the OpenGL graphics library, and the haptic rendering is achieved by the

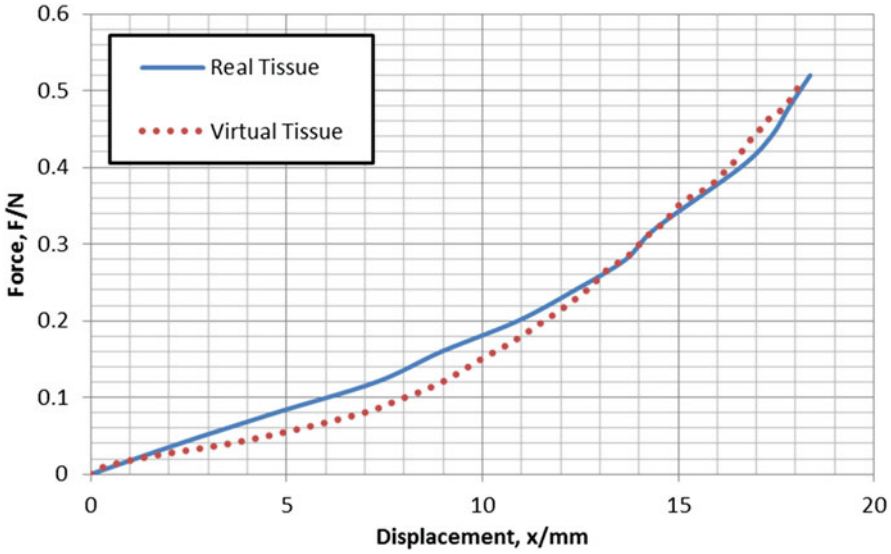



**Fig. 9.6** Comparison between virtual and real soft tissue behaviours

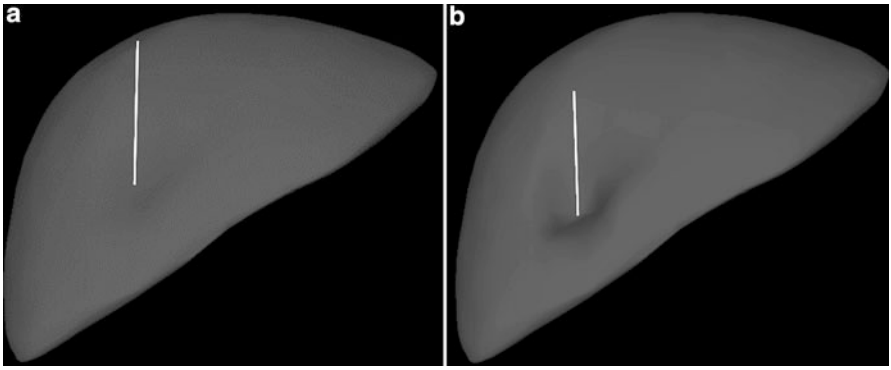



**Fig. 9.7** Deformations of a virtual human liver model with haptic feedback by using a virtual probe

OpenHaptic toolkit from Sensable Technologies. Real-time interactive deformation of virtual human organs with force feedback has been performed by the proposed methodology. Figure 9.7 shows the deformation process of a virtual human liver model with force feedback by using a virtual probe and the final deformation result.

**Table 9.1** Computational performance

| Numbers of points | 261 | 405 | 650 | 900 | 1100 |
|---|---|---|---|---|---|
| Computational time (ms) | 4.3 | 8.5 | 12.4 | 24.3 | 39.8 |
| Frame rate (fps) | 232.5 | 117.6 | 80.6 | 41.2 | 25.1 |

### *9.6.3 Computational Performance and Discussions*

The proposed methodology has been implemented on an Intel Pentium (R) 2.8 GHz and 1.0G memory PC. With one external stimulus, the graphical update rate utilizing different numbers of mesh points is provided in Table 9.1. From Table 9.1, it can be seen that the graphical update rate is decreased with the increment of the mesh points. The visually satisfactory refresh rate of 30 Hz to maintain a realistic visual feedback (Sadd 2005) is achieved by meshes with less than about 1000 grid points. Since the correct visualization of an object such as the human liver requires at least 600 grid points (Zhang et al. 2014), this is sufficient to provide realistic visual feedback with the proposed methodology.

The haptic device requires forces to be updated at the rate of at least 1000 Hz for realistic force feedback (OpenHaptics ToolKit-Programmer's and Sensable 2004). It is observed that the force update rate is above 1000 Hz when the number of mesh points is relatively small. When the computational speed cannot cope with the haptic refresh rate any more, force extrapolation (Picinbono et al. 2002) is employed to improve the realism of force feedback by generating the missing forces from the previous time step.

The main difference of the proposed methodology from most of the existing deformation methods is that soft tissue simulation is carried out from the physic-ochemical viewpoint of soft tissues. It converts soft tissue deformation into a chemical–mechanical interaction problem, and thus the complex and expensive elastic computations are avoided. Compared with the common deformation methods such as mass-spring, FEM and BEM, the proposed methodology can deal with large-range deformation, while the common deformation methods can only handle small deformation. Compared with the few methods based on nonlinear strains for large-range deformations, such as explicit nonlinear FEM, large-range deformation is achieved by the nonlinear load–deformation relationship rather than the use of nonlinear strain in the explicit nonlinear FEM. The proposed methodology can handle both local and large-range deformation. In contrast, local deformation is difficult to achieve due to the use of nonlinear strain in the explicit nonlinear FEM. Furthermore, the proposed methodology can easily accommodate anisotropic and inhomogeneous deformation by simple modification of diffusion coefficients.

## 9.7 Conclusions

This paper presents a new methodology to conduct simulation of nonlinear soft tissue deformation from the physicochemical viewpoint of soft tissues. This methodology converts soft tissue deformation into chemical–mechanical interaction, and thus the complex and expensive elastic computations are avoided. It features chemical diffusion of mechanical load and non-rigid mechanics of motion to govern the simulation dynamics of soft tissue deformation. The proposed methodology can not only accommodate isotropic, anisotropic and inhomogeneous materials through simple modification of diffusion coefficients, but it can also accommodate local and large-range deformation.

Future research work will focus on the global validation of simulated deformation against real in vivo data of soft tissues. In fact, the validation should be based on in vivo measurement data on the mechanical properties of living tissues, due to the different behaviours between living tissues and non-living tissues. The measurement of in vivo mechanical data of soft tissues would be of great value for a full validation but is much more challenging to achieve experimentally. A minimally invasive in vivo measurement method for acquisition of mechanical data of soft tissues will be established for a full validation of deformation models.

## References

Barbic, J., & James, D. L. (2005). Deformable models: Real-time subspace integration for St. Venant-Kirchhoff deformable models. *ACM Transactions on Graphics, 24*(3), 982–990.

Cakir, O., & Yazici, R. (2009). Real-time cutting simulation based on stiffness-warped FEM. In: *24th International Symposium on Computer and Information Sciences* (pp. 721–724), North Cyprus.

Choi, M. G., & Ko, H. S. (2005). Modal warping: Real-time simulation of large rotational deformation and manipulation. *IEEE Transactions on Visualization and Computer Graphics, 11*(1), 91–101.

Chou, P. C., & Pagano, N. J. (1967) *Elasticity*. Princeton, NJ: D. Van Nostrand Company.

Costa, I. F. (2012). A novel deformation method for fast simulation of biological tissue formed by fibers and fluid. *Medical Image Analysis, 16*, 1038–1046.

Duarte, V., Gonzalez, Y., & Cerrolaza, M. (2011). Boundary element simulation of bone tissue. *International Journal of Biomedical Engineering and Technology, 5*(2–3), 211–228.

Etzmuss, O., Gross, J., & Strasser, W. (2003). Deriving a particle system from continuum mechanics for the animation of deformable objects. *IEEE Transactions on Visualization and Computer Graphics, 9*(4), 538–550.

Fung, Y. C. (1993) *Biomechanics: Mechanical properties of living tissues* (2nd ed.). New York: Springer.

Lai, W. M., Hou, J. S., & Mow, V. C. (1991). A triphasic theory for the swelling and deformation behaviors of articular cartilage. *Journal of Biomedical Engineering, 113*(3), 245–258.

Mafi, R., & Sirouspour, S. (2014). GPU-based acceleration of computations in nonlinear finite element deformation analysis. *International Journal for Numerical Methods in Biomedical Engineering, 30*(3), 365–381.

Myers, E. R., Lai, W. M., & Mow, V. C. (1984). A continuum theory and an experiment for the ion-induced swelling behavior of articular cartilage. *ASME Journal of Biomechanical Engineering, 106*(2), 151–158.

Omar, M. N., Zhong, Y., Jazar, R. N., Subic, A., Smith, J., & Shirinzadeh, B. (2015). Soft tissue modelling with conical springs. *Bio-Medical Materials and Engineering, 26,* S207–S214.

OpenHaptics ToolKit-Programmer's Guide, Sensable Technologies (2004).

Patete, P., Iacono, M. I., Spadea, M. F., Trecate, G., Vergnaghi, D., Mainardi, L. T., et al. (2013). A multi-tissue mass-spring model for computer assisted breast surgery. *Medical Engineering and Physics, 35*(1), 47–53.

Peterlik, I., Sedef, M., Basdogan, C., et al. (2010). Real-time visio-haptic interaction with static soft tissue models having geometric and material nonlinearity. *Computers and Graphics – UK*, *34*(1), 43–54.

Picinbono, G., Lombardo, J.-C., Delingette, H., & Ayache, N. (2002). Improving realism of a surgery simulator: Linear anisotropic elasticity, complex interactions and force extrapolation. *Journal of Visualization and Computer Animation, 13*(3), 147–167.

Picinbono, G., Delingette, H., & Ayache, N. (2003). Non-linear anisotropic elasticity for real-time surgery simulation. *Graphical Models, 65*(5), 305–321.

Sadd, M. H. (2005). *Elasticity: Theory, applications, and numerics*. Amsterdam: Elsevier Butterworth Heinemann.

San-Vicente, G., Aguinaga, I., & Celigueta, J. T. (2012). Cubical mass-spring model design based on a tensile deformation test and nonlinear material model. *IEEE Transactions on Visualization and Computer Graphics, 18*(2), 228–241.

Schwartz, J.-M., Denninger, M., Rancourt, D., Moisan, C., & Laurendeau, D. (2005). Modelling liver tissue properties using a non-linear visco-elastic model for surgery simulation. *Medical Image Analysis, 9*(2), 103–112.

Taylor, Z. A., Crozier, S., & Ourselin, S. (2011). A reduced order explicit dynamic finite element algorithm for surgical simulation. *IEEE Transactions on Medical Imaging, 30*(9), 1713–1721.

Wu, J. Z., & Herzog, W. (2002). Simulating the swelling and deformation behaviour in soft tissues using a convective thermal analogy. *Biomedical Engineering Online, 1*(8), 8.

Zhang, J., Wang, J., Wang, X., & Feng, D. (2014). The adaptive FEM elastic model for medical image registration. *Physics in Medicine and Biology, 59*(1), 97–118.

Zhong, Y., Shirinzadeh, B., Alici, G., & Smith, J. (2006a). An autowave based methodology for deformable object simulation. *Computer-Aided Design, 38*(7), 740–754.

Zhong, Y., Shirinzadeh, B., Alici, G., & Smith, J. (2006b). A neural network methodology for deformable object simulation. *IEEE Transactions on Information Technology in Biomedicine, 10*(4), 749–762.

Zhong, Y., Shirinzadeh, B., Smith, J., & Gu, C. (2010). Thermal-mechanical based soft tissue deformation for surgery simulation. *International Journal of Advanced Robotics, 24*, 1719–1739.

Zhong, Y., Shirinzadeh, B., Smith, J., & Gu, C. (2012). Soft tissue deformation with reaction-diffusion process for surgery simulation. *Journal of Visual Languages and Computing, 23*(1), 1–12.

Zhu, B., & Gu, L. (2012). A hybrid deformable model for real-time surgical simulation. *Computerized Medical Imaging and Graphics, 36*(5), 356–365.

# Chapter 10
# Strength and Contraction Speed of Muscle Groups: An Application of the Lambert Function

**Franz Konstantin Fuss and Ming Adin Tan**

**Abstract** The aim of this book chapter is to find and evaluate an analytical solution for a non-linear ODE that shows the sum of the function, its derivative and the product of function and derivative equals unity. Consequently we will use the solution of the ODE to develop a model of limb movement based on muscle parameters, and to define muscle performance parameters from limb segment movements. Furthermore, the solution can be applied to experimentally derive muscle group parameters for muscle diagnostics, as well as for scaling of the speed of biological systems based on muscle analysis.

**Keywords** Lambert function • Tree function • Muscle model • Muscle dynamics • Muscle performance • Muscle power • Curve fitting • Hill's curve • Force–velocity relationship • Froude number • Scaling of dynamic systems • Driven pendulum

## Key Symbols

| | |
|---|---|
| $A$ | Physiological cross-sectional area of a muscle |
| $a$ | Linear acceleration |
| $a$, $b$ | Coefficients of a Matlab function |
| $a$, $b$, $c$, $d$ | Coefficients of an ordinary differential equation |
| $c_F$ | Muscle force constant |
| COM | Centre of mass |
| $c_{R1}$, $c_{R2}$ , $c_{R3}$, $c_{R4}$ | Constants of the rotational muscle model |
| $c_T$ | Muscle torque constant |
| $c_{T1}$, $c_{T2}$ , $c_{T3}$, $c_{T4}$ | Constants of the translational muscle model |
| $c_v$ | Muscle (linear) velocity constant |

F.K. Fuss (✉) • M.A. Tan
School of Aerospace, Mechanical, and Manufacturing Engineering, RMIT University, Melbourne, VIC, Australia
e-mail: franz.fuss@rmit.edu.au



R.N. Jazar, L. Dai (eds.), *Nonlinear Approaches in Engineering Applications*,
DOI 10.1007/978-3-319-27055-5_10

| | |
|---|---|
| $c_\omega$ | Muscle angular velocity constant |
| d | Differential operator |
| e | Base of natural logarithm (2.71828183 . . . ) |
| $F$ | Force; muscle force |
| $F_{\mathrm{CF}}$ | Centrifugal force |
| $F_G$ | Gravitational force |
| $F_I$ | Inertial force |
| $Fr$ | Froude number |
| $ft$ | Relative amount of fast twitch fibres |
| $g$ | Gravitational acceleration |
| $h$ | Height |
| $I$ | Moment of inertia |
| IC | Instant centre |
| $k$ | Radius of gyration |
| $L$ | Length of pendulum, leg length |
| $l_f$ | Optimal muscle fascicle length |
| $l_m$ | Muscle moment arm |
| log | Natural logarithm |
| $\log_{10}$ | Decadic logarithm |
| $m$ | Mass; unspecific constant in Lambert's equations |
| $m, p, q$ | Constants in Lambert's equations |
| $n$ | Summation counter |
| ODE | Ordinary differential equation |
| $P$ | Power |
| $p$ | Power of scaling factor; unspecific constant in Lambert's equations |
| $r$ | Radius; distance between COM and IC |
| $s$ | Scaling factor |
| $st$ | Relative amount of slow twitch fibres |
| $t$ | Time |
| $T$ | Torque |
| T | Tree function |
| $v$ | Linear velocity |
| $V$ | Volume |
| W | Lambert function |
| $\mathrm{W}_0$ | Principal branch of Lambert function |
| $\mathrm{W}_{-1}$ | Negative branch of Lambert function |
| $x$ | Variable of a function |
| $\dot{x}$ | Time derivative of $x$ |
| $y$ | Variable of a function |
| $\alpha$ | Angular acceleration; unspecific constant in Euler's equations |
| $\alpha, \beta, \upsilon, \omega$ | Constants in Euler's equations |
| $\varphi$ | Angular frequency of a pendulum |
| $\pi$ | 3.14159265359 |
| $\theta$ | Angle |

| | |
|---|---|
| $\rho$ | Density |
| $\sigma_{\max}$ | Maximal physiological muscle stress (0.3 MPa) |
| $\tau$ | Period |
| $\omega$ | Angular velocity; unspecific constant in Euler's equations |
| $\infty$ | Infinity |

## 10.1 Introduction

The Lambert W function is defined as follows:
if

$$y = x\mathrm{e}^{x} \tag{10.1}$$

then

$$x = \mathrm{W}(y) = \mathrm{W}\left(x\mathrm{e}^{x}\right) \tag{10.2}$$

and is named after Johann Heinrich Lambert, a Swiss mathematician, who lived in the eighteenth century (1728–1777). Lambert's work on the series solution of trinomial equations of the structure

$$x^{m} + px = q \tag{10.3}$$

(first equation on p. 156 of Lambert 1758), re-written as

$$x = q + x^{m} \tag{10.4}$$

(second equation on p. 226 of Lambert 1770) attracted the interest of Euler who transformed Eq. (10.4) to

$$x^{\alpha} - x^{\beta} = (\alpha - \beta)\,\upsilon x^{\alpha+\beta} \tag{10.5}$$

(second equation on p. 29 of Euler 1783), by substituting $x$ for $x^{-\beta}$, $q$ for $(\alpha - \beta)\upsilon$, and $m$ for $\alpha/\beta$. Note that Corless et al. (1996) stated that $m = \alpha\beta$, thereby overlooking the division sign between $\alpha$ and $\beta$. Applying these substitutions to Eq. (10.4)

$$x^{-\beta} = (\alpha - \beta)\,\upsilon + \left(x^{-\beta}\right)^{m} \tag{10.6}$$

$$x^{-\beta} = (\alpha - \beta)\,\upsilon + x^{-\beta m} \tag{10.7}$$

$$x^{-\beta} = (\alpha - \beta)\,\upsilon + x^{-\beta\frac{\alpha}{\beta}} \tag{10.8}$$

$$x^{-\beta} = (\alpha - \beta)\,\upsilon + x^{-\alpha} \tag{10.9}$$

$$x^{-\beta} - x^{-\alpha} = (\alpha - \beta)\,\upsilon \tag{10.10}$$

and multiplying by $x^{\alpha}\, x^{\beta}$

$$x^{-\beta}x^{\alpha}x^{\beta} - x^{-\alpha}x^{\alpha}x^{\beta} = (\alpha - \beta)\,\upsilon x^{\alpha}x^{\beta} \tag{10.11}$$

$$x^{\alpha} - x^{\beta} = (\alpha - \beta)\,\upsilon x^{\alpha}x^{\beta} \tag{10.12}$$

results in Eq. (10.5). Equation (10.5) is also referred to as '*Lambert's Transcendental Equation*' in Wolfram MathWorld (Weisstein 2015), although it is due to Euler. Euler (1783) refers in his paper, written in Latin, to the 'series Lambertina' (Lambert's series), by probably coining the eponym of the W function. In fact, Eq. (10.5) results in the Lambert function when following Euler's derivation. Applying Euler's case II ($\beta = \alpha$, p. 33, Euler 1783) to Eq. (10.5), by setting $\alpha = \beta + \omega$ and letting $\omega \to 0$, such that $\beta = \alpha$,

$$x^{\beta+\omega} - x^{\beta} = (\omega)\,\upsilon x^{2\beta+\omega} \tag{10.13}$$

$$x^{\beta}x^{\omega} - x^{\beta} = \omega\upsilon x^{2\beta}x^{\omega} \tag{10.14}$$

and multiplying by $x^{-\beta}$,

$$x^{\omega} - 1 = \omega\upsilon x^{\beta}x^{\omega} \tag{10.15}$$

yields

$$\frac{x^{\omega} - 1}{\omega} = \upsilon x^{\beta+\omega} \tag{10.16}$$

If $\omega \to 0$,

$$\lim_{\omega\to 0} \frac{x^{\omega} - 1}{\omega} = \log x \tag{10.17}$$

and thus

$$\log x = \upsilon x^{\beta} = \upsilon x^{\alpha} \tag{10.18}$$

(p. 33, Euler 1783). When setting $\alpha = 1$ (p. 39, Euler 1783),

$$\log x = \upsilon x \tag{10.19}$$

Transforming Eq. (10.19)

$$\log x = -x\,(-\upsilon) \tag{10.20}$$

$$-\log x = \mathrm{e}^{\log x}\,(-\upsilon) \tag{10.21}$$

$$-\log x\,\mathrm{e}^{-\log x} = (-\upsilon) \tag{10.22}$$

and applying Eqs. (10.1) and (10.2) yields

$$-\log x = \mathrm{W}\,(-\upsilon) \tag{10.23}$$

and thus

$$\log x = -\mathrm{W}\,(-\upsilon) \tag{10.24}$$

The right-hand term of Eq. (10.24) is identical to the tree function T (Eq. (3.1) of Janson et al. 1993)

$$-\mathrm{W}\,(-\upsilon) = \mathrm{T}\,(\upsilon) = \sum_{n\geq 1} \frac{n^{n-1}}{n!}\upsilon^n \tag{10.25}$$

(Corless et al. 1996; Janson et al. 1993). Euler solved Eq. (10.19) with exactly the same series expansion (first equation on p. 39 of Euler 1783). Although Lambert solved his trinomials with series expansions as well (Lambert 1758; Lambert 1770), only Euler's transformed Eq. (10.5) leads directly to the Lambert W function.

The Lambert W function comprises of two branches; the principal branch $\mathrm{W}_0$ and the negative branch $\mathrm{W}_{-1}$, (Fig. 10.1). It can be further divided into three segments as detailed in Fig. 10.1.

Segment 1: principle branch $\mathrm{W}_0$, positive segment: 1st quadrant: $0 \leq x \leq \infty$; $0 \leq \mathrm{W}(x) \leq \infty$.
Segment 2: principle branch $\mathrm{W}_0$, negative segment: 3rd quadrant: $-1\mathrm{e}^{-1} \leq x \leq 0$; $-1 \leq \mathrm{W}(x) \leq 0$.
Segment 3: negative branch $\mathrm{W}_{-1}$: 3rd quadrant: $-1\mathrm{e}^{-1} \leq x \leq 0$; $-1 \geq \mathrm{W}(x) \geq -\infty$.

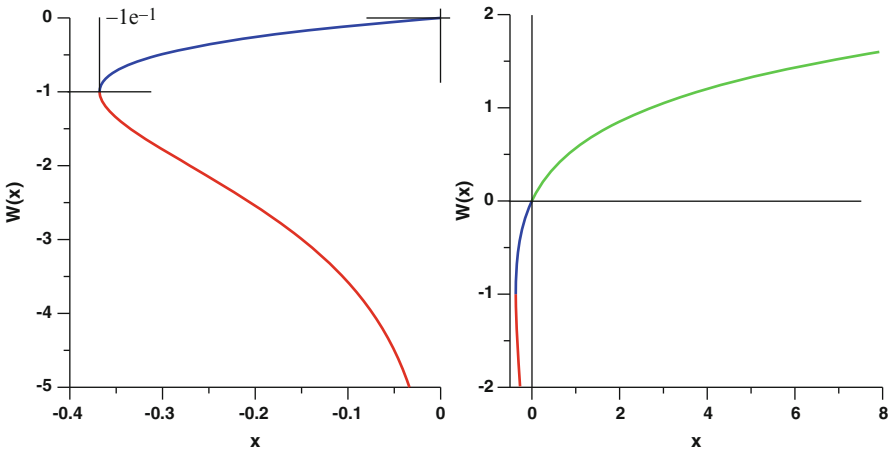



**Fig. 10.1** Segments of the Lambert function: *red*: negative branch; *blue*: negative segment of principle branch; *green*: positive segment of principle branch (color figure online)

The Lambert function is a solution for differential equations, such as delay differential equations (Corless et al. 1996) and ordinary differential equations (ODE) of the following structure:

$$a\frac{\mathrm{d}x}{\mathrm{d}t} + x\frac{\mathrm{d}x}{\mathrm{d}t} + bx = c \tag{10.26}$$

used for describing

(1) enzyme-catalysed reactions with single-substrate mechanisms following the Michaelis-Menten rate law for enzyme kinetics (Segel and Slemrod 1989; Williams 2010), and
(2) the maximal possible muscle speed when moving a limb segment (Fuss and Tan 2006).

Equation (10.26) has the structure of the ODE shown in Eqs. (10.27) and (10.29).

The aim of this book chapter is to find and evaluate an analytical solution for the ODE of Eqs. (10.27) and (10.29). Consequently we will use the solution of the ODE to develop a model of limb movement based on muscle parameters, and to define muscle performance parameters from limb segment movements. Furthermore, the solution can be further applied to experimentally derive muscle group parameters for muscle diagnostics, as well as for scaling of the speed of biological systems based on muscle analysis.

## 10.2 Non-Linear ODE

The ODE to be solved can be described in its simplest form as

$$x_t + \dot{x}_t + x_t\dot{x}_t = 1 \tag{10.27}$$

where $x_t$ is a function of time and $\dot{x}_t$ is its first derivative. The structure of the ODE shows that the sum of the function, its derivative and the product of function and derivative equals unity. The non-linearity of the ODE is given by the product $x_t\dot{x}_t$.

$$x + \frac{\mathrm{d}x}{\mathrm{d}t} + x\frac{\mathrm{d}x}{\mathrm{d}t} = 1 \tag{10.28}$$

The general form of this ODE is

$$ax_t + b\dot{x}_t + cx_t\dot{x}_t = d \tag{10.29}$$

with coefficients $a$, $b$, $c$, and $d$.

$$ax + b\frac{\mathrm{d}x}{\mathrm{d}t} + cx\frac{\mathrm{d}x}{\mathrm{d}t} = d \tag{10.30}$$

The ODE is solved as follows:

$$\frac{\mathrm{d}x}{\mathrm{d}t}(b + cx) = d - ax \tag{10.31}$$

$$\frac{b + cx}{d - ax}\mathrm{d}x = \mathrm{d}t \tag{10.32}$$

$$\int_{x_1}^{x_2} \frac{b + cx}{d - ax}\mathrm{d}x = t_2 - t_1 \tag{10.33}$$

Solving the integral by considering the initial conditions: $x_1 = 0$, and $t_1 = 0$:

$$\left| -\frac{(ab + cd)\log(ax - d) + acx}{a^2} \right|_0^x = t \tag{10.34}$$

$$|(ab + cd)\log(ax - d) + acx|_0^x = -a^2 t \tag{10.35}$$

$$\log\left|(ax - d)^{(ab+cd)}\mathrm{e}^{acx}\right|_0^x = -a^2 t \tag{10.36}$$

$$\log\frac{(ax - d)^{(ab+cd)}\mathrm{e}^{acx}}{(-d)^{(ab+cd)}} = -a^2 t \tag{10.37}$$

$$(ax-d)^{(ab+cd)}\mathrm{e}^{acx} = \mathrm{e}^{-a^2t}(-d)^{(ab+cd)} \tag{10.38}$$

$$(ax-d)^{(ab+cd)}\mathrm{e}^{acx} = -\mathrm{e}^{-a^2t}d^{ab+cd} \tag{10.39}$$

At that point, it becomes evident that $x$ on the left side of Eq. (10.39) appears as a multiplier of the exponential function as well as its exponent. The solution is therefore a Lambert function, to be applied after transforming the equation. Taking the '$ab+cd$'-root of Eq. (10.39)

$$(ax-d)\,\mathrm{e}^{\frac{acx}{ab+cd}} = -\mathrm{e}^{-\frac{a^2t}{ab+cd}}d \tag{10.40}$$

$$(ax-d) = -\mathrm{e}^{-\frac{a^2t}{ab+cd}}\mathrm{e}^{-\frac{acx}{ab+cd}}d \tag{10.41}$$

$$(ax-d) = -\mathrm{e}^{\frac{-a^2t-acx}{ab+cd}}d \tag{10.42}$$

Introducing '$cd$–$cd$' into the exponent of the exponential function allows separating the '$x$'- and '$t$'-arguments of the exponent such that the '$x$'-argument of the exponent becomes identical to the left-hand side of the equation

$$\frac{acx-cd}{c} = -\mathrm{e}^{\frac{cd-cd-a^2t-acx}{ab+cd}}d \tag{10.43}$$

$$\frac{acx-cd}{c} = -\mathrm{e}^{\frac{cd-acx}{ab+cd}}\mathrm{e}^{\frac{-cd-a^2t}{ab+cd}}d \tag{10.44}$$

$$\frac{acx-cd}{c}\mathrm{e}^{-\frac{cd-acx}{ab+cd}} = -\mathrm{e}^{\frac{-cd-a^2t}{ab+cd}}d \tag{10.45}$$

$$\frac{acx-cd}{c\,(ab+cd)}\mathrm{e}^{\frac{acx-cd}{ab+cd}} = -\frac{\mathrm{e}^{\frac{-cd-a^2t}{ab+cd}}}{ab+cd}d \tag{10.46}$$

$$\frac{acx-cd}{ab+cd}\mathrm{e}^{\frac{acx-cd}{ab+cd}} = -\frac{\mathrm{e}^{\frac{-cd-a^2t}{ab+cd}}}{ab+cd}cd \tag{10.47}$$

On the left-hand side of Eq. (10.47), multiplier and exponent of the exponential function are identical such that Eq. (10.47) can be solved with a Lambert function:

$$\frac{acx-cd}{ab+cd} = \mathrm{W}\left(-\frac{\mathrm{e}^{\frac{-cd-a^2t}{ab+cd}}}{ab+cd}cd\right) \tag{10.48}$$

By transforming and simplifying

$$\frac{acx-cd}{ab+cd} = \mathrm{W}\left(-\frac{\mathrm{e}^{\frac{-cd-a^2t+(ab+cd)\log(cd)}{ab+cd}}}{ab+cd}\right) \tag{10.49}$$

$$\frac{acx-cd}{ab+cd} = \mathrm{W}\left(-\mathrm{e}^{\frac{-cd-a^2t+(ab+cd)\log(cd)-(ab+cd)\log(ab+cd)}{ab+cd}}\right) \tag{10.50}$$

$$\frac{acx-cd}{ab+cd} = \mathrm{W}\left(-\mathrm{e}^{\frac{-cd-a^2t+(ab+cd)[\log(cd)-\log(ab+cd)]}{ab+cd}}\right) \tag{10.51}$$

$$\frac{acx-cd}{ab+cd} = \mathrm{W}\left(-\mathrm{e}^{\frac{-cd-a^2t+(ab+cd)\log\frac{cd}{ab+cd}}{ab+cd}}\right) \tag{10.52}$$

$$acx = (ab+cd)\,\mathrm{W}\left(-\mathrm{e}^{\frac{-cd-a^2t+(ab+cd)\log\frac{cd}{ab+cd}}{ab+cd}}\right) + cd \tag{10.53}$$

$$x = \frac{ab+cd}{ac}\mathrm{W}\left(-\mathrm{e}^{\frac{-cd-a^2t+(ab+cd)\log\frac{cd}{ab+cd}}{ab+cd}}\right) + \frac{d}{a} \tag{10.54}$$

$$x = \frac{ab+cd}{ac}\mathrm{W}\left(-\mathrm{e}^{-\frac{a^2t}{ab+cd}-\frac{cd}{ab+cd}+\frac{(ab+cd)\log\frac{cd}{ab+cd}}{ab+cd}}\right) + \frac{d}{a} \tag{10.55}$$

$$x = \frac{ab+cd}{ac}\mathrm{W}\left(-\mathrm{e}^{-\frac{a^2t}{ab+cd}-\frac{cd}{ab+cd}+\log\frac{cd}{ab+cd}}\right) + \frac{d}{a} \tag{10.56}$$

$$x = \frac{ab+cd}{ac}\mathrm{W}\left(-\mathrm{e}^{-\frac{a^2t}{ab+cd}-\frac{cd}{ab+cd}}\,\mathrm{e}^{\log\frac{cd}{ab+cd}}\right) + \frac{d}{a} \tag{10.57}$$

$$x = \frac{ab+cd}{ac}\mathrm{W}\left(-\mathrm{e}^{-\frac{a^2t}{ab+cd}-\frac{cd}{ab+cd}}\frac{cd}{ab+cd}\right) + \frac{d}{a} \tag{10.58}$$

we obtain

$$x = \frac{ab+cd}{ac}\mathrm{W}\left(-\frac{cd}{ab+cd}\mathrm{e}^{-\frac{cd}{ab+cd}}\mathrm{e}^{-\frac{a^2t}{ab+cd}}\right) + \frac{d}{a} \tag{10.59}$$

Equation (10.59) is the final solution of ODE (10.30). When evaluating Eq. (10.59), it becomes evident that $\frac{ab+cd}{ac}$ is a multiplier of the Lambert function, and $\frac{d}{a}$ is an asymptotic value, identical to $x_{\max}$ as shown below. The argument of the Lambert function comprises of a constant, $-\frac{cd}{ab+cd}\mathrm{e}^{-\frac{cd}{ab+cd}}$, and an inverse exponential

function of the independent variable $t$. If $t = 0$, the latter exponential function reduces to 1:

$$x_{t=0} = \frac{ab+cd}{ac}\mathrm{W}\left(-\frac{cd}{ab+cd}\mathrm{e}^{-\frac{cd}{ab+cd}}\right) + \frac{d}{a} = \frac{ab+cd}{ac}\left(-\frac{cd}{ab+cd}\right) + \frac{d}{a} = -\frac{d}{a} + \frac{d}{a} = 0 \tag{10.60}$$

If $t$ approaches infinity, the inverse exponential function of the independent variable $t$ reaches 0, and $x$ approaches asymptotically its maximal value, $x_{\max}$, at $\frac{d}{a}$.

$$\begin{aligned} &\lim_{t\to\infty} x = \tfrac{ab+cd}{ac}\mathrm{W}\left(-\tfrac{cd}{ab+cd}\mathrm{e}^{-\frac{cd}{ab+cd}}\mathrm{e}^{-\infty}\right) + \tfrac{d}{a} \\ &= \tfrac{ab+cd}{ac}\mathrm{W}\left(-\tfrac{cd}{ab+cd}\mathrm{e}^{-\frac{cd}{ab+cd}}\tfrac{1}{\infty}\right) + \tfrac{d}{a} = \tfrac{ab+cd}{ac}\mathrm{W}(0) + \tfrac{d}{a} = \tfrac{d}{a} = x_{\max} \end{aligned} \tag{10.61}$$

Plotting the value of the Lambert function of Eq. (10.59) against the value of its argument, it becomes evident that the values are located on the negative segment of the principal branch, ranging from $-\frac{cd}{ab+cd}\mathrm{e}^{-\frac{cd}{ab+cd}}$ to 0 on the horizontal axis (Fig. 10.2), and from $-\frac{cd}{ab+cd}$ to 0 on the vertical axis (Fig. 10.2). The negative limit of $-1\mathrm{e}^{-1}$ on the horizontal axis can be reached only if

$$-\frac{cd}{ab+cd} = -1 \tag{10.62}$$

Equation (10.62) is true only if $a$ or $b$ is 0. From a practical point of view, however, Eq. (10.59) is not defined if $a = 0$. This implies that if the second argument of ODE (10.29), i.e. $b\dot{x}_t$, approaches 0, then the value on the horizontal axis approaches the negative limit of $-1\mathrm{e}^{-1}$. The other limit of the negative segment of the principal branch is 0, which is applicable only if

$$-\frac{cd}{ab+cd} = 0 \tag{10.63}$$

Equation (10.63) is true only if $c$ or $d$ is 0. From a practical point of view, however, Eq. (10.59) is not defined if $c = 0$. This implies that if the third argument of ODE (10.29), i.e. $cx_t\dot{x}_t$, or the product of function $x$ and its derivative, is integral to the Lambert function solution of ODE (10.29). Conversely, if $d$ approaches 0, $x$ becomes constant in Eq. (10.59) at a limit of 0, and is no longer dependent on $t$. The time $t$, however, reaches singularity if $d = 0$, according to Eq. (10.37).

Therefore, $-1 \leq -\frac{cd}{ab+cd} \leq 0$.

As

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{W}(t) = \frac{\mathrm{W}(t)}{t\mathrm{W}(t)+t} \tag{10.64}$$

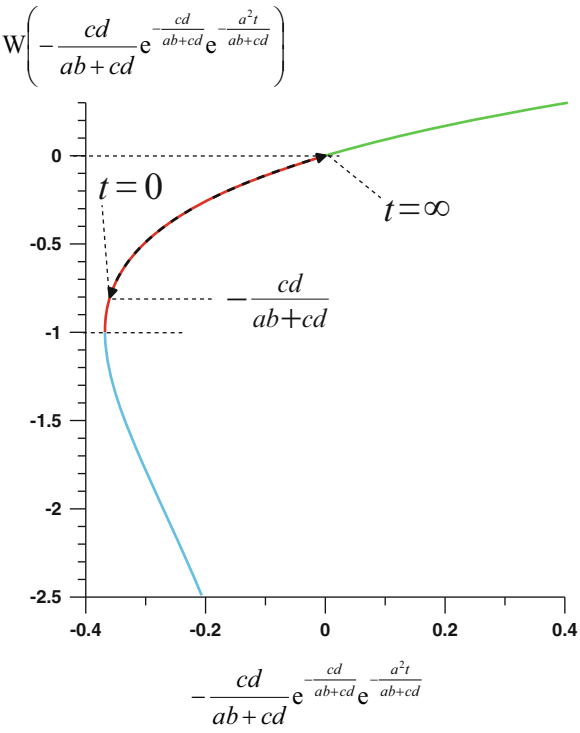



**Fig. 10.2** Value of the Lambert function of Eq. (10.59) against the value of its argument, for $-cd/(ab+cd) = -0.8$; *blue*: negative branch; *red*: negative segment of principle branch; *green*: positive segment of principle branch (color figure online)

the first derivative of Eq. (10.59) is

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\frac{ab+cd}{ac}\frac{a^2 W\left(-\frac{cd}{ab+cd}e^{-\frac{cd}{ab+cd}}e^{-\frac{a^2t}{ab+cd}}\right)}{(ab+cd)\,W\left(-\frac{cd}{ab+cd}e^{-\frac{cd}{ab+cd}}e^{-\frac{a^2t}{ab+cd}}\right)+(ab+cd)} \tag{10.65}$$

The initial, and also maximum, gradient at $t=0$ is therefore: $\dot{x}_{\max} = \frac{d}{b}$. If $b$ approaches 0, i.e. at the condition of Eq. (10.62), then the initial gradient approaches infinity.

The asymptotic behaviour of $x$ against $t$, according to Eq. (10.59), is shown in Fig. 10.3.

## 10.3 Model of Limb Movements Based on Muscle Force–Velocity Relationship

When moving a limb segment, the torque $T$ generated by a muscle or muscle group accelerates the segment.

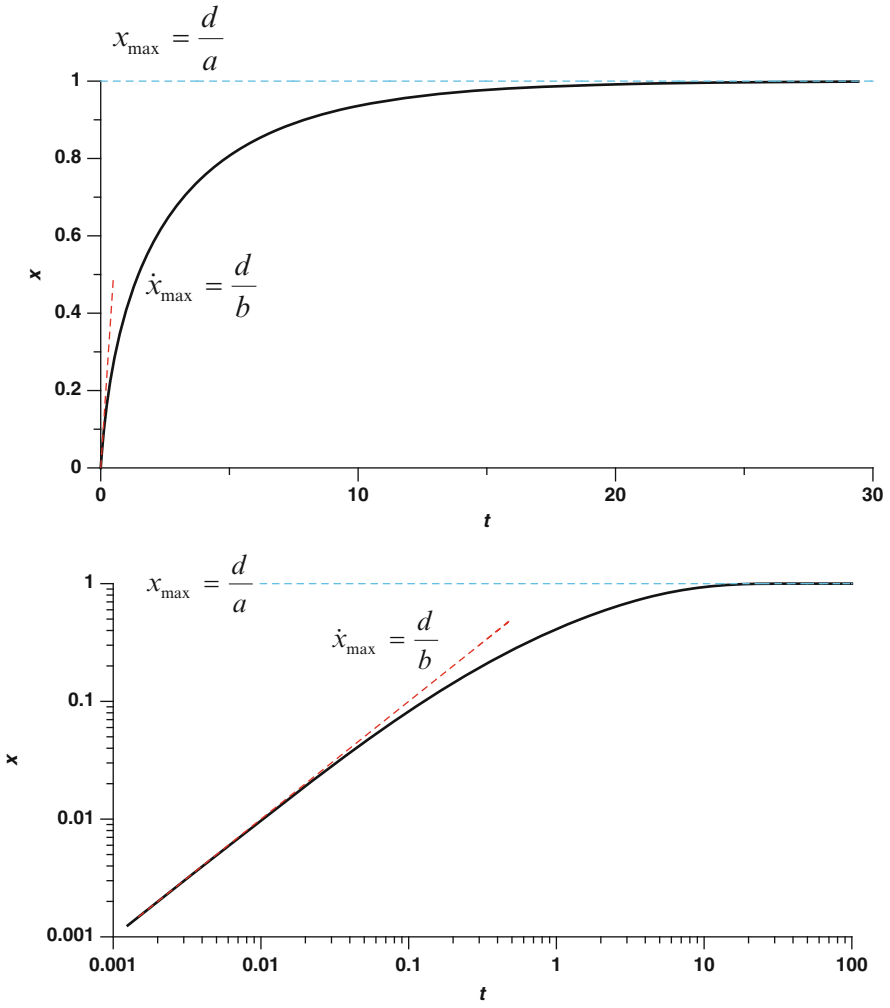



**Fig. 10.3** $x$ against $t$, according to Eq. (10.59), for coefficients of $a = b = d = 1$, $c = 4$

$$T = \alpha I_{IC} \tag{10.66}$$

where $\alpha$ is the angular acceleration of the limb segment, and $I_{IC}$ is its moment of inertia about the instant centre IC (axis of that joint about which the segment rotates). According to the parallel axis theorem,

$$I_{IC} = I_{COM} + r^2 m = k^2 m + r^2 m = m\left(k^2 + r^2\right) \tag{10.67}$$

where $I_{COM}$ is its moment of inertia about the centre of mass (COM); $r$ is the distance between the COM and the IC; $m$ is the mass of the limb segment; and $k$ is the radius of gyration about the COM.

The muscle torque $T$ is the cross-product of muscle force $F$ and moment arm $l$:

$$T = F \times l_m \tag{10.68}$$

The linear acceleration $a$ of the shortening muscle is calculated from its moment arm $l$ and the angular acceleration $\alpha$ of a rotating limb segment:

$$a = \alpha l_m \tag{10.69}$$

Equating Eqs. (10.66) and (10.68), by considering Eqs. (10.67) and (10.69), yields

$$\frac{a}{l_m} m \left(k^2 + r^2\right) = F l_m \tag{10.70}$$

Solving for $a$:

$$a = \frac{F {l_m}^2}{m\left(k^2 + r^2\right)} \tag{10.71}$$

The muscle force $F$ in Eqs. (10.68) and (10.71), however, is a function of the muscle velocity $v$, according to Hill's equation (Hill 1970), and $v$ increases with $a$.

The muscle force $F$ in Hill's equation (10.72) is calculated from the maximal force $F_{max}$ a muscle can produce under static conditions (isometric contraction at $v = 0$), from the muscle velocity $v$, and from two constants of the force–velocity equation, the force constant $c_F$, and the velocity constant $c_v$. Hill denotes these constants 'a' and 'b'. However, in this book chapter, we use $c_F$ and $c_v$ instead, to avoid confusion with the muscle acceleration $a$, and in order to distinguish linear/translational constants ($c_F$ and $c_v$) from angular/rotational ones ($c_T$ and $c_\omega$).

$$F = \frac{F_{max} c_v - c_F v}{c_v + v} \tag{10.72}$$

Solving Eq. (10.72) for $v$ yields

$$v = \frac{c_v \left(F_{max} - F\right)}{F + c_F} \tag{10.73}$$

The maximal velocity $v_{max}$ is the one a muscle can reach under zero load. Thus, the basic relationship is that $v_{max}$ occurs at $F = 0$, and $F_{max}$ at $v = 0$. The ratios of $c_F/F_{max}$ as well as $c_v/v_{max}$ are identical

$$\frac{c_F}{F_{max}} \equiv \frac{c_v}{v_{max}} \approx 0.25 \tag{10.74}$$

This becomes evident when $F$ approaches zero, $v$ becomes $v_{\max}$ in Eq. (10.73). Typical values for $c_F/F_{\max}$ were found to be about 0.25 for a variety of species (Hill 1938; Close 1964; Close and Hoh 1967). These ratios will be set to 0.25 for the subsequent calculations.

The maximal force $F_{\max}$ a muscle can produce depends on the physiological cross-sectional area $A$, and on the maximal physiological stress $\sigma_{\max}$. The latter is a physiological constant (Wells 1965; Jayes and Alexander 1982; Nygaard et al. 1983) of about 0.3 MPa.

$$\sigma_{\max} = \frac{F_{\max}}{A} = 0.3 \text{ MPa} \tag{10.75}$$

Substitution of Eqs. (10.75) and (10.72) into (10.71) delivers

$$a = \frac{(\sigma_{\max} A c_v - c_a v)\, {l_m}^2}{(c_v + v)\, m\, (r^2 + k^2)} \tag{10.76}$$

To simplify the above equation, we let

$$c_{T4} = \frac{{l_m}^2}{m\,(r^2 + k^2)} \tag{10.77}$$

Substituting $c_{T4}$ into Eq. (10.76) and rearranging, we obtain

$$a c_v + a v + v c_F c_{T4} - \sigma_{\max} A c_v c_{T4} = 0 \tag{10.78}$$

Note that the subscript $_T$ of $c_{T4}$ and the subsequent $c_{T1}$, $c_{T2}$ and $c_{T3}$ denotes linear/translational constants, related to $v$ and $F$ (in contrast to $\omega$ and $T$ further down, where constants are denoted with the subscript $_R$).

After defining further constants

$$c_{T1} = c_v \tag{10.79}$$

$$c_{T2} = c_{T4} c_F \tag{10.80}$$

$$c_{T3} = \sigma_{\max} A c_v c_{T4} \tag{10.81}$$

where

$$c_{T3} = 4 c_{T1} c_{T2} \tag{10.82}$$

we obtain the following non-linear ODE:

$$v c_{T2} + a c_{T1} + a v = 4 c_{T1} c_{T2} \tag{10.83}$$

$$v c_{T2} + \frac{\mathrm{d}v}{\mathrm{d}t} c_{T1} + \frac{\mathrm{d}v}{\mathrm{d}t} v = 4 c_{T1} c_{T2} \tag{10.84}$$

Equations (10.83) and (10.84) have the same structure as ODEs (10.29) and (10.30), and the coefficients of $a$, $b$, $c$, and $d$ of ODEs (10.29) and (10.30) correspond to $c_{T2}$, $c_{T1}$, 1 and $4c_{T1}c_{T2}$ in Eqs. (10.83) and (10.84). Consequently, Eq. (10.59) can be directly transformed for solving Eqs. (10.83) and (10.84) for $v$:

$$v = 5 c_{T1} \mathrm{W}\left( -\frac{4}{5} \mathrm{e}^{-\frac{4}{5}} \mathrm{e}^{-\frac{c_{T2} t}{5 c_{T1}}} \right) + 4 c_{T1} \tag{10.85}$$

which corresponds to the solution found by Fuss and Tan (2006). Equation (10.85) defines $v$ as being located on Hill's curve, i.e. as the maximal possible $v$ at any given $F$ according to Hill's equation (10.72) and (10.73).

The constants derived from Eq. (10.85) are (with the constants of the general ODE in square brackets):

- $$\left[ \frac{cd}{ab + cd} \right] = 4/5 = 0.8 \tag{10.86}$$

- $$v_{\max} = \left[ \frac{d}{a} \right] = 4 c_{T1} \tag{10.87}$$

- $$a_{\max} = \left[ \frac{d}{b} \right] = 4 c_{T2} \tag{10.88}$$

- $$F_{\max} = 4 c_{T2} \frac{m\left(r^2 + k^2\right)}{l^2} \tag{10.89}$$

- $$P_{\max} = 0.09549 F_{\max} v_{\max} = 1.52786 c_{T1} c_{T2} \frac{m\left(r^2 + k^2\right)}{{l_m}^2} \tag{10.90}$$

where $P_{\max}$ is the maximum power which can be produced in theory; the constant of 0.09549 used in Eq. (10.90) is derived as follows:

Power $P$ is the product of force $F$ and velocity $v$.

$$P = Fv \tag{10.91}$$

Substituting Eq. (10.72) into (10.91) yields

$$P = v\frac{F_{\max}c_v - c_F v}{c_v + v} = \frac{F_{\max}c_v v - c_F v^2}{c_v + v} \tag{10.92}$$

The maximal power $P_{\max}$ occurs at that velocity $v_{P\max}$, where the gradient of the power–velocity curve is zero:

$$\frac{\mathrm{d}P}{\mathrm{d}v} = \frac{(F_{\max}c_v - 2c_F v)(c_v + v) - \left(F_{\max}c_v v - c_F v^2\right)}{(c_v + v)^2} = 0 \tag{10.93}$$

$$F_{\max}{c_v}^2 + F_{\max}c_v v - 2c_F v c_v - 2c_F v^2 - F_{\max}c_v v + c_F v^2 = 0 \tag{10.94}$$

$$v^2 + 2vc_v - \frac{F_{\max}}{c_F}{c_v}^2 = 0 \tag{10.95}$$

Solving for $v$, considering that $v$ must be positive, yields

$$v_{P\max} = c_v\left(\sqrt{1 + \frac{F_{\max}}{c_F}} - 1\right) \tag{10.96}$$

As $c_v = 0.25\, v_{\max}$, and $F_{\max}/c_F = 4$

$$v_{P\max} = 0.25 v_{\max}\left(\sqrt{5} - 1\right) \tag{10.97}$$

$$v_{P\max} = 0.25 v_{\max}\left(\sqrt{5} - 1\right) = v_{\max}\frac{\sqrt{5} - 1}{4} = 0.309017 v_{\max} \tag{10.98}$$

Substituting Eq. (10.98) into (10.72) yields

$$F_{P\max} = 0.309017 F_{\max} \tag{10.99}$$

Consequently:

$$P_{\max} = F_{P\max}v_{P\max} = 0.309017^2 F_{\max}v_{\max} = 0.0954915 F_{\max}v_{\max} \tag{10.100}$$

General problems associated with Eq. (10.85) are:

(1) The gravitational force and moment are not taken into account. Doing that would require the constantly changing inclination angle of the limb segment to be factored into the ODE, which changes the first order ODE to a second order one. The problem can be solved by moving the limb segment in a plane perpendicular to the gravitational acceleration vector, and having the joint axis aligned in parallel to the gravitational acceleration vector.

(2) There is no single muscle velocity $v$, as in most limb segment movements multiple muscles (or parts of a muscle with different geometries) are involved. The solution to the problem is to consider the angular velocity of the limb segment instead of the linear velocities of individual muscles involved. The angular velocity then represents the overall averaged speed of the system instead of individual speeds of its components.
(3) The moment arm of muscles changes throughout the movement; furthermore, the number of moment arms corresponds to the number of muscles involved, and the actual length of each moment arm is difficult to measure. The problem can be solved by considering the angular velocity and the torque produced by a limb segment, instead of linear velocity and force, as then the knowledge of moment arms is no longer required, e.g. for calculating $F_{\max}$ according to Eq. (10.89).
(4) The force–length relationship of muscles is not taken into account; the force does not only change with velocity but also with the length of the muscle. This relationship does not significantly influence the model, as the force drops faster due to the increasing velocity, than the shortening of the muscle would imply. Moreover, most muscles work on the ascending limb of the force—displacement curve (Murray et al. 2000), i.e. the plateau segment where the maximal force is produced corresponds to the position where the movement starts.

In order to overcome problems (2) and (3), the model is changed subsequently from $v$ and $F$ to $\omega$ and $T$.

$$\alpha = \frac{T}{m\,(r^2 + k^2)} \tag{10.101}$$

$$\frac{c_F}{F_{\max}} = \frac{c_F l_m}{F_{\max} l_m} = \frac{c_T}{T_{\max}} = 0.25 \tag{10.102}$$

where $c_T$ is the torque constant.

$$\frac{c_v}{v_{\max}} = \frac{c_v / l_m}{v_{\max} / l_m} = \frac{c_\omega}{\omega_{\max}} = 0.25 \tag{10.103}$$

where $c_\omega$ is the angular velocity constant.

$$T = \frac{T_{\max} c_\omega - c_T \omega}{c_\omega + \omega} \tag{10.104}$$

$$\alpha = \frac{T_{\max} c_\omega - c_T \omega}{(c_\omega + \omega)\, m\,(r^2 + k^2)} \tag{10.105}$$

$$c_{R4} = \frac{1}{m\,(r^2 + k^2)} \tag{10.106}$$

$$c_{R1} = c_\omega \tag{10.107}$$

$$c_{R2} = c_T c_{R4} \tag{10.108}$$

$$c_{R3} = T_{\max} c_\omega c_{R4} = T_{\max} c_{R1} c_{R4} = 4 c_T c_{R1} c_{R4} = 4 c_{R1} c_{R2} \tag{10.109}$$

The non-linear ODE results in

$$c_{R2}\omega + c_{R1}\alpha + \alpha\omega = 4 c_{R1} c_{R2} \tag{10.110}$$

$$c_{R2}\omega + c_{R1}\frac{\mathrm{d}\omega}{\mathrm{d}t} + \frac{\mathrm{d}\omega}{\mathrm{d}t}\omega = 4 c_{R1} c_{R2} \tag{10.111}$$

the solution of which is

$$\omega = 5 c_{R1} \mathrm{W}\left(-\frac{4}{5}\mathrm{e}^{-\frac{4}{5}}\mathrm{e}^{-\frac{c_{R2}t}{5c_{R1}}}\right) + 4 c_{R1} \tag{10.112}$$

The constants derived from Eq. (10.112) are (with the constants of the general ODE in square brackets), if other than Eqs. (10.86)–(10.90):

- $$\omega_{\max} = \left[\frac{d}{a}\right] = 4 c_{R1} \tag{10.113}$$

- $$\alpha_{\max} = \left[\frac{d}{b}\right] = 4 c_{R2} \tag{10.114}$$

- $$T_{\max} = 4 c_{R2} m \left(r^2 + k^2\right) \tag{10.115}$$

- $$P_{\max} = 0.09549 T_{\max} \omega_{\max} = 1.52786 c_{R1} c_{R2} m \left(r^2 + k^2\right) \tag{10.116}$$

The moment of inertia $I$ according to Eq. (10.67) can be calculated from the data of Drillis and Contini (1966) or with the help of video techniques (e.g. Hatze and Baca 1992). If this is too inaccurate or too difficult, then the maximal torque and power can be expressed as per unit moment of inertia

$$T_{\max}/I = \alpha_{\max} = 4 c_{R2} \tag{10.117}$$

$$P_{\max}/I = 0.09549\,\omega_{\max}T_{\max}/I = 0.09549\,\omega_{\max}\alpha_{\max} = 1.52786 c_{R1}c_{R2} \quad (10.118)$$

$T_{\max}$ and $\omega_{\max}$ are not necessarily ultimate muscle constants, as muscle force, torque and speed decrease with fatigue, muscle and nerve damage, and pain. The parameters $T_{\max}$ and $\omega_{\max}$ rather describe the current state of muscle condition for diagnostic purposes, which become important in the following section.

## 10.4 Muscle Group Diagnostics

The problem associated with muscle group diagnostics is that, although $F_{\max}$ and $T_{\max}$ can be determined statically, i.e. through isometric contraction at $v = 0$, $v_{\max}$ of the muscle and $\omega_{\max}$ of the limb segment are unknown. Therefore, also the maximal power produced by the muscle is unknown as well. The power curve, however, can be established at different, maximal possible, torques and velocities, e.g. with a dumbbell. However, fatigue is likely to set in during such an exercise. Each repetition of this exercise at various speeds and torques results in a single datum on the power curve, and a reasonable number of data are required for fitting them with Eq. (10.92). The solution to that problem is to record the maximum speed (angular velocity) of a limb segment over several repetitions (e.g. 10 times), and take only the maximum $\omega$-values at any timestamp for fitting them with Eq. (10.112).

First of all, it will be shown subsequently, that the maximal muscle velocity cannot be reached by muscles within the given motion range of a limb segment, exemplified by upper arm and thigh muscle groups (Table 10.1 and Fig. 10.4).

The analytical model of Eq. (10.85) was tested in three conditions:

(1) Extension of the shank + foot segment driven by the quadriceps.
(2) Flexion of the forearm + hand segment driven by biceps, brachialis and brachioradialis.
(3) Extension of the forearm + hand segment driven by the triceps.

The input parameters of the model are listed in Table 10.1.

The velocity–angle graphs (Fig. 10.4) show that the limb segments (1) shank driven by quadriceps, and forearm driven by (2) flexors and (3) extensors, move at 0.5 $v_{\max}$ after an angular displacement of 25° (0.07 revolutions), 293° (0.81 revolutions), and 599° (1.66 revolutions), respectively (Fig. 10.4b). The segments move at 0.9 $v_{\max}$ after 285° (0.79 revolutions), 3400° (9.4 revolutions), and 6935° (19.3 revolutions); $v_{\max}$ is reached only after 3.9, 46.5 and 94.9 revolutions, respectively, for the data listed in Table 10.1. Biological joints, however, have a maximal range of motion of about half a revolution. In the specific case of the knee and elbow joint, the active range of motion is 140°, after which the velocity would be 0.793 $v_{\max}$, 0.390 $v_{\max}$, and 0.305 $v_{\max}$ for (1) shank driven by quadriceps, and forearm driven by (2) flexors and (3) extensors, respectively. These 'peak' velocities

**Table 10.1** Input parameters of the analytical model of Eq. (10.85)

| Muscle (group) | Quadriceps | Biceps brachii + brachialis + brachioradialis | Triceps brachii |
|---|---|---|---|
| $A$ ($m^2$) | 0.00935 (Winter 1990) | 0.00131 (An et al. 1984) | 0.00188 (An et al. 1984) |
| $F_{max}$ (N) | 2805 | 393 | 564 |
| $c_F$ | 701.25 | 98.25 | 141 |
| $l_m$ (m) | 0.04 (Fuss 1993) | 0.03223 (Murray et al. 2000) | 0.02 (An et al. 1984; Murray et al. 2000) |
| $m$ (kg) | 3.9 | 1.68 | 1.68 |
| $r$ (m) | 0.2335 | 0.1867 | 0.1867 |
| $k$ (m) | 0.1450 | 0.1111 | 0.1111 |
| $I$ (kg $m^2$) | 0.2946 | 0.0793 | 0.0793 |
| $l_f$ (m) | 0.069 (Winter 1990) | 0.123 (Murray et al. 2000) | 0.103 (Murray et al. 2000) |
| $v_{max}$ (m/s) | 0.759 | 1.353 | 1.133 |
| $c_v = c_{T1}$ | 0.18975 | 0.33825 | 0.28325 |
| $c_{T2}$ | 3.80811 | 1.28707 | 0.71126 |

$l_m$ = muscle moment arm (average), $l_f$ = optimal muscle fascicle length; $l_m$ of upper arm flexors and $l_f$ of quadriceps and upper arm flexors were weighted according to the physiological cross sections $A$ of the individual muscles or muscle heads; $m$, $r$, and $k$ were calculated for an individual of 70 kg body mass and 1.75 m body height based on the data of Drillis and Contini (1966); $v_{max} = l_f$ (16 *ft* + 6 *st*) (Herzog 1999), where *ft* and *st* are the relative amount of fast and slow twitch fibres, which produce a velocity of 16 $l_f$/s and 6 $l_f$/s, respectively (Spector et al. 1980); in the quadriceps and biceps, *ft*/*st* is 1 (Bottinelli et al. 1997; Nygaard and Sanchez 1982); in the triceps, *ft*/*st* is 0.67–4 (Sjogaard 1982), and was set to 1; due to the considerable range of *ft*/*st*, in the triceps, as well as 0.25–2.33 in the vastus lateralis (Sjogaard 1982), *ft*/*st* was also set to 1 in brachialis and brachioradialis

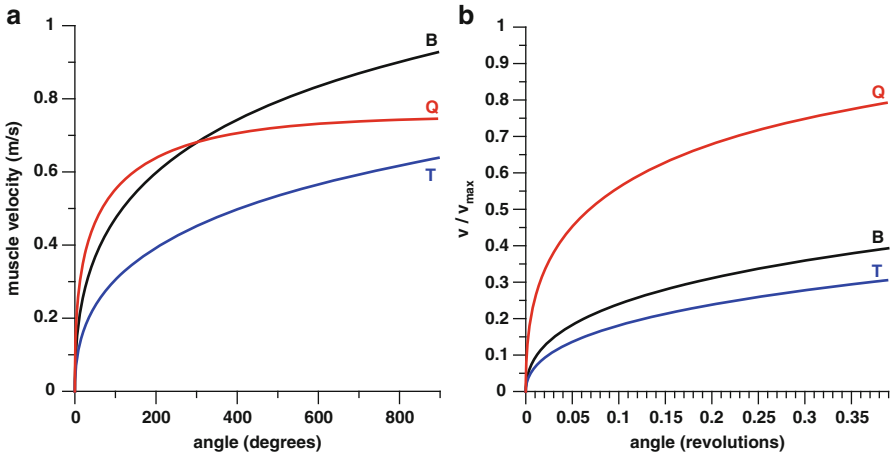



**Fig. 10.4** Muscle velocity against angle of limb segment (**a**), and normalised velocity against angle (**b**); B = forearm flexors (biceps, brachialis, brachioradialis), T = forearm extensor (triceps), Q = shank extensor (quadriceps)

at the end of the 140° motion range are theoretical ones, as the peak should occur well before the end of the motion range, in order to allow sufficient time and angle for decelerating the velocity down to zero.

These results clearly show that the maximal muscle velocity is purely a theoretical value, which can be reached under extreme experimental conditions, but never in reality. The experimental conditions are, as applied in Hill's experiments (Hill 1938), to cut the muscle at one end from the bone, such that the load applied to the muscle is basically zero and effectively only corresponds to its mass times the muscle acceleration. However, muscles are attached to bones at both ends, as they serve to move limb segments. Limb segments are characterised by a specific mass concentration and distribution, which accounts for the mass and moment of inertia to be moved by the muscle. The nature of the velocity–angle (Fig. 10.4) relationship, i.e. its asymptotic nature anticipates that $v_{max}$ is a theoretical value. This becomes even more evident based on the fact, that joints cannot perform a full revolution and thus cannot rotate indefinitely.

For determining muscle diagnostic parameters, the following experiment was conducted. A participant of the RMIT Smart Cricket Ball study (Fuss et al. 2011, 2012, 2014; Fuss and Smith 2013, 2014) held the ball firmly in his hand, by moving the forearm horizontally at maximal speed, but keeping the upper arm immobile. This standard exercise was used to check the functionality of the smart ball before collecting bowling data. The Smart Cricket Ball study was approved by the Ethics Committee of RMIT University (BSEHAPP 13-12). The ball records the angular velocity with three orthogonal high-speed MEMS gyros (Fuss et al. 2011, 2012, 2014; Fuss and Smith 2013, 2014). The raw data representing the angular velocity were converted to rad/s and the resultant angular velocity was calculated. Nine data sets were recorded at 500 Hz (Fig. 10.5a). At each time stamp, the maximum

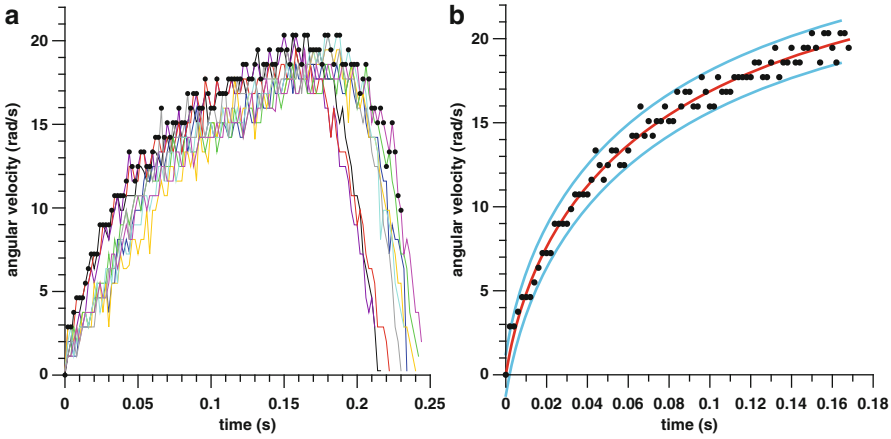



**Fig. 10.5** Angular velocity against time of nine elbow flexion experiments at maximally possible speed (**a**) with maxima indicated as *black dots*; (**b**): fit curve (*red*) of the maxima using Eq. (10.119) and observation prediction bounds (*blue*) at confidence level 95 % (color figure online)

datum was taken out of the nine data sets (Fig. 10.5a), and the maxima were fitted (Fig. 10.5b) with Matlab (v 2010; by Mathworks, Natick, MA, USA) using Eq (10.112).

The time and angular velocity data shown in Fig. 10.5a were imported in the Matlab curve fit tool (cftool) and a custom fit equation of the structure

$$f(x)=5^{*}a^{*}\left(\text{lambertw}\left(-\exp\left(\left(-b^{*}x+a^{*}\left(5^{*}\log(0.8)-4\right)\right)/\left(5^{*}a\right)\right)\right)\right)+4^{*}a \tag{10.119}$$

where $a = c_{R1}$, and $b = c_{R2}$, was used with fit boundaries of 0.1–$\infty$ and 1–$\infty$ for $a$ and $b$, respectively.

The result obtained was:

Coefficients (with 95 % confidence bounds):

$$a = 6.103 \ \ (5.851, 6.354)$$

$$b = 189.7 \ \ (176.3, 203)$$

Goodness of fit:

SSE (sum of the squared errors of the prediction function): 31.83
$R^2$ (coefficient of determination): 0.9847
RMSE (root mean squared error): 0.6193

In addition to that, the residual standard deviation was 0.6155 rad/s. When using Eqs. (10.113), (10.114), (10.117), and (10.118), the following parameters can be calculated: $\omega_{max} = 24.4$ rad/s; $T_{max}/I = \alpha_{max} = 758.8$ Nm kg$^{-1}$m$^{-2}$ or rad/s$^2$; and $P_{max}/I = 1768$ W kg$^{-1}$m$^{-2}$.

## 10.5 Scaling of Muscle Function

Scaling the speed of biological systems was undertaken for the first time by Alexander (1976, 1991), specifically for predicting the speed of dinosaurs from trackways. Alexander used the Froude number approach, predicting that an increase in size of a biological system by a factor of $x$, improves the locomotion speed by a factor of $\sqrt{x}$.

The naval architect William Froude introduced a non-dimensional parameter that represented the dynamic similarity of boats of different hull lengths (Froude 1874). This parameter, i.e. the Froude number, is defined as $v^2/gL$ where $v$ is velocity, $g$ is gravitational acceleration, and $L$ is the hull length (Vaughan and O'Malley 2005).

For biological systems, limbs are considered to be pendulums, with limb or limb segment moving as swinging pendulum (or inverted pendulum). The Froude number of the pendulum's speed is defined as the ratio of centrifugal force $F_{CF}$ to gravitational force $F_G$ (Kram et al. 1997).

$$Fr = \frac{F_{CF}}{F_G} = \frac{m{\omega_{max}}^2 L}{mg} = \frac{v^2}{Lg} \tag{10.120}$$

where $\omega_{max}$ is the maximum angular velocity of a pendulum, $L$ is the length of the pendulum, $m$ is the point mass at the end of $L$, $g$ is the gravitational acceleration, and $v$ is the maximum linear velocity of the point mass.

$\omega_{max}$ of a pendulum is defined as

$$\omega_{max} = \theta_0 \varphi \tag{10.121}$$

where $\theta_0$ is half the range of motion of the pendulum, and $\varphi$

$$\varphi = \frac{2\pi}{\tau} \tag{10.122}$$

is the angular frequency of a pendulum with period $\tau$, where

$$\tau \approx 2\pi \sqrt{\frac{L}{g}} \tag{10.123}$$

for small oscillation angles.

$\varphi$ is therefore

$$\varphi = \sqrt{\frac{g}{L}} \tag{10.124}$$

and $\omega_{max}$ is

$$\omega_{max} = \theta_0 \sqrt{\frac{g}{L}} \tag{10.125}$$

Substituting Eq. (10.125) into (10.120) yields

$$Fr = \frac{F_{CF}}{F_G} = \frac{m{\omega_{max}}^2 L}{mg} = \frac{{\theta_0}^2 g}{g} = {\theta_0}^2 \tag{10.126}$$

Equation (10.126) explains that the Froude number is independent of the pendulum's length and mass, and depends only on its range of motion ($2\theta_0$). The Froude number is therefore a constant, independent of scaling in terms of size and mass.

According to Eq. (10.120), biological systems of different sizes move at the same Froude number, and their velocity

$$v = \sqrt{FrLg} \tag{10.127}$$

scales with the square root of $L$. A biological system of twice or thrice the size does not move twice or thrice as fast but only $\sqrt{2}$ or $\sqrt{3}$ times as fast. This applies, of course, only to the condition that limbs (e.g. legs) act as a pendulum. According to Eq. (10.125), biological systems can speed up only when increasing range of motion ($2\theta_0$) of their limb pendulums. This correlates with the fact that the faster the movement speed, the longer is the stride length.

The only problem in this approach is that limbs do not work like a ***simple harmonic oscillator with free oscillations***, but rather as a ***driven oscillator with forced oscillations***. The difference is that in a free oscillator all forces acting on the system originate from the COM (gravitational, inertial and centrifugal forces) and the pivot (reaction force), whereas a driven oscillator is additionally subjected to external forces acting on parts of the system other than COM or IC. These are the ***muscle forces*** acting on limb bones. It is therefore questionable whether the Froude number concept is applicable to muscle-driven biological systems.

Table 10.2 compares physical quantities of muscle-driven systems and pendulums in terms of scaling factors. If the ***scaling factor*** is $x$, then the system is $x$ times as large as a system with scaling factor of 1. The ***power of the scaling factor*** depends on the quantity. If the physical quantity is the mass, then the power is 3, and hence the mass increases with $x^3$. These two systems in Table 10.2 have different ***powers*** for forces, linear and angular velocity and acceleration, torques, power, and for the Froude number. In pendulums, the Froude number is constant, independent of scaling factors, whereas in muscle-driven systems, the (theoretical) Froude number, calculated from Eq. (10.120; applicable to pendulums) is variable

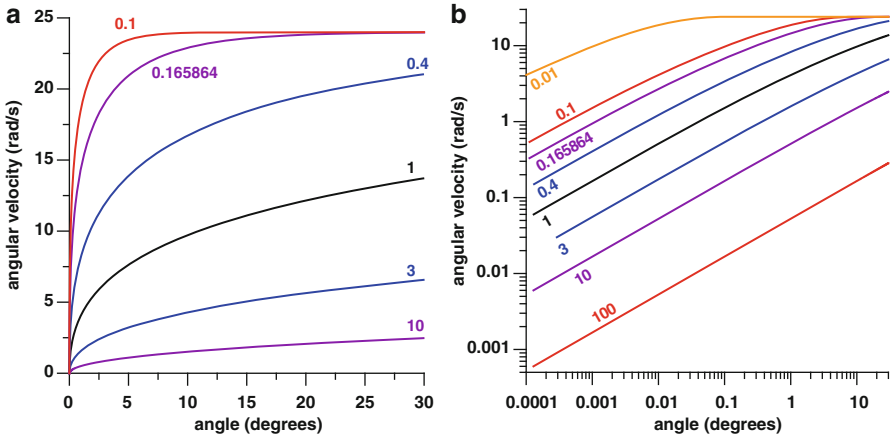



**Fig. 10.6** Angular velocity against movement angle of a limb segment moving over an arc of 30°; the curves are labelled with the magnitude of the scaling factor

and scales to the power of 1. Therefore, the Froude number, for estimating the speed of biological systems of different sizes, is not applicable.

Subsequently, the scaling is evaluated with a realistic example: a limb segment moves through an arc over 60°, 30° each required for accelerating and decelerating the segment. The muscle group parameters were selected at $c_{R1} = c_\omega = 6$ and $c_{R2} = c_T/I = 200$, i.e. realistic data taken from the experiment above. The maximal possible speed of the limb segment is therefore 24 rad/s. The average linear velocity $v_a$ over the shortest distance between the endpoints of the segment at the beginning and end of the movement is proportional to

$$v_a \propto \frac{s}{t_{30}} \tag{10.128}$$

where $t_{30}$ is the time required for moving the segment through an arc of 30°, calculated numerically from Eq. (10.112), and $s$ is the scaling factor.

Figure 10.6 shows how fast the angular velocity builds up over 30°. At scaling factors $s > 0.165864$ (found empirically through numerical calculation), the segment does not reach $\omega_{\max}$ of 24 rad/s at the end of the given 30° arc (cf. Fig. 10.4). At very small scaling factors, $\omega_{\max}$ is reached instantaneously.

The proportional average linear velocity $v_a$ is shown in Fig. 10.7. At large scaling factors, the average velocity flattens out and seems to reach a limit. This stands in contrast to the Froude number approach (Fig. 10.8), where the segment is considered a pendulum with free oscillations. The velocity predicted from the Froude number $Fr$ is overestimated at large and small scaling factors, compared to the velocity of a muscle-driven limb segment (Fig. 10.8). The scaling exercise shown in Table 10.2 predicts the scaling factor of the linear speed of a pendulum directly, without utilising the Froude number: 0.5; i.e. when increasing the size of a biological system

**Table 10.2** Scaling of physical quantities ($\rho$ = density)

| Muscle-driven system | | | Pendulum with point mass | | |
|---|---|---|---|---|---|
| Physical quantity | Equation | Power $p$ of scaling factor | Physical quantity | Equation | Power $p$ of scaling factor |
| Length, height, width, radius, radius of gyration, moment arm, distance, stride length, muscle fibre length | $L$, $h$, $w$, $r$, $k$, $l$, $x$, $S$, $lf$ | 1 | Length, height, width, radius, radius of gyration, moment arm, distance, stride length | $L$, $h$, $w$, $r$, $k$, $l$, $x$, $S$ | 1 |
| Area, physiological cross section of muscle | $A = x^2$ | $2 = (2 \bullet 1)$ | Area | $A = x^2$ | $2 = (2 \bullet 1)$ |
| Volume | $V = x^3$ | $3 = (3 \bullet 1)$ | Volume | $V = x^3$ | $3 = (3 \bullet 1)$ |
| Mass | $m = \rho V$ | $3 = (0 + 3 \bullet 1)$ | Mass | $m = \rho V$ | $3 = (0 + 3 \bullet 1)$ |
| Angle | $\theta$ | 0 | Angle | $\theta$ | 0 |
| Muscle force | $F = \sigma_{max} A$ | $2 = (0 + 2)$ | Gravitational force, inertial force, centrifugal force | $F_G = mg$ | $3 = (3 + 0)$ |
| | | | | $F_I = ma$ | $3 = (3 + 0)$ (for $a$ see below) |
| | | | | $F_{CF} = mr\omega^2$ | $3 = (3 + 1 + 2 \bullet [-0.5])$ (for $\omega$ see below) |
| Theoretical maximal contraction speed of muscle | $v_{max} = c\, l_f$<br>$c$ = constant; 6 in slow twitch fibres, 16 in fast twitch fibres; cf. Table 10.1 | $1 = (1)$ | Maximal linear velocity of pendulum | $v_{max} = \omega L = \theta_0 \varphi L = \theta_0 L\sqrt{g/L} = \theta_0\sqrt{gL}$ | $0.5 = (0 + 0/2 + 1/2)$ |

| | | | | | |
|---|---|---|---|---|---|
| Theoretical maximal angular velocity of limb segment | $\omega_{max} = v/l_m$ (constant value) also applies to $c_\omega = c_{R1} = \omega_{max}/4$ | $0 = (1-1)$ | Maximal angular velocity of pendulum | $\omega_{max} = \theta_0 \varphi = \theta_0 \sqrt{g/L}$ | $-0.5 = (0 + 0/2 - 1/2)$ |
| Moment of inertia | $I = m\ (k^2 + r^2)$ | $5 = (3 + 2 \bullet 1)$ | Moment of inertia | $I = m\ L^2$ | $5 = (3 + 2 \bullet 1)$ |
| Muscle torque | $T_m = F_m l_m$ also applies to $c_T = T_{max}/4$ | $3 = (2 + 1)$ | Torque of gravitational force and inertia | $T_{GF} = F_G L$ | $4 = (3 + 1)$ |
| | | | | $T_{IF} = F_I L = maL$ | $4 = (3 + 0 + 1)$ |
| | | | | $T_I = \alpha\ I = mL^2\alpha = maL$ | $4 = (3 + 0 + 1)$ (for $a$ see below) |
| Angular acceleration of limb segment | $\alpha = T/I$<br>$\alpha_{max} = 4c_{R2} =$<br>$= 4\ c_T c_{R4} = T_{max}/I$ | $-2 = (3-5)$<br>$-2 = (3-5)$ | Maximal angular acceleration of pendulum | $\alpha_{max} = \theta_0 \varphi^2 = \theta_0\ 2\pi\ g/L$ | $-1 = (0 + 0 + 0 - 1)$ |
| Maximal linear acceleration of muscle | $a_{max} = \alpha\ l_m$ | $-1 = (-2 + 1)$ | Maximal linear acceleration of pendulum | $a_{max} = \alpha_{max} L =$<br>$\theta_0\ 2\pi g$<br>(constant value, no variable) | $0 = (-1 + 1)$<br>$0 = (0 + 0 + 0)$ |
| Power of muscle | $P = F_m v =$<br>$= T_m \omega$ | $3 = (2 + 1)$<br>$3 = (3 + 0)$ | Power of pendulum | $P = F\ v =$<br>$= T\ \omega$ | $3.5 = (3 + 0.5)$<br>$3.5 = (4 - 0.5)$ |
| '*Theoretical maximal Froude number*' | $Fr = {v_{max}}^2/(gL)$<br>(not constant, variable) | $1 = (2 \bullet 1 - 0 - 1)$ | Froude number | $Fr = v^2/(gL)$<br>$Fr = {\theta_0}^2$<br>(constant value) | $0 = (2 \bullet 0.5 - 0 - 1)$<br>$0 = (2 \bullet 0)$ |

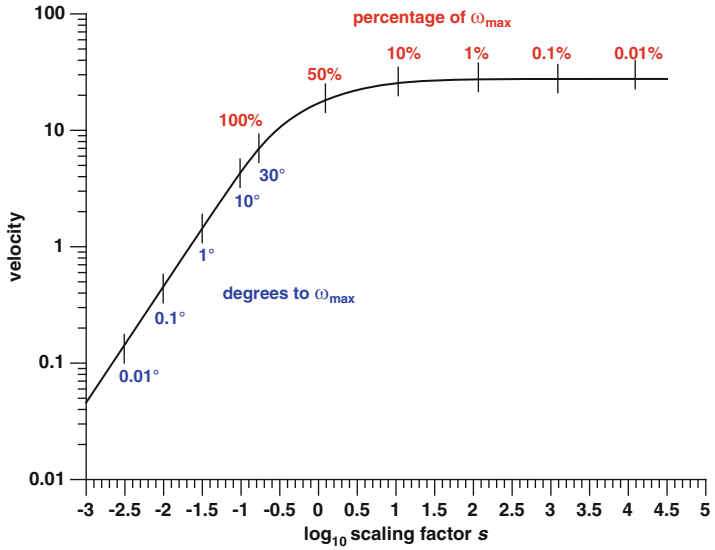



**Fig. 10.7** Velocity from Eq. (10.128) against the logarithm of the scaling factor; *blue labels*: angle in degrees required to reach the maximal angular velocity; *red labels*: cases where the maximal angular velocity cannot be reached indicated by the percentage of the maximal angular velocity which can be reached within the 30° arc of movement (color figure online)

by a factor $x$, then the speed increases by $\sqrt{x}$. Scaling of biological muscle-driven systems suggests that the linear speed scales with the power of 1. However, this applies to the maximal muscle speed only, and it takes a certain amount of time to reach this speed, if it can be reached at all (cf. Fig. 10.6). Therefore, the average speed does not exactly follow the power of 1. In fact, the speed scales with powers between 0 and 1, distributed around 0.5 (Fig. 10.9). For very small scaling factors, the power converges to 1 (power of maximal muscle velocity, which increases linearly with the scaling factor), and for very large ones, the power asymptotes to 0 (i.e. the velocity reaches a limit and becomes constant).

$$v_n = {s_n}^p \tag{10.129}$$

where $s_n$ = normalised scaling factor (normalised to that $s$ used as a reference for calculating $Fr$) and $v_n$ = normalised speed ($v_n = 1$ at $s_n = 1$). The power $p$ of scaling factor is calculated from

$$p = \frac{\log_{10} v_n}{\log_{10} s_n} \tag{10.130}$$

The relationship of the velocity calculated for a muscle-driven system and the one calculated from the Froude number $Fr$ depends on the scaling factor which serves as a reference for the velocity which the reference Froude number is calculated from.

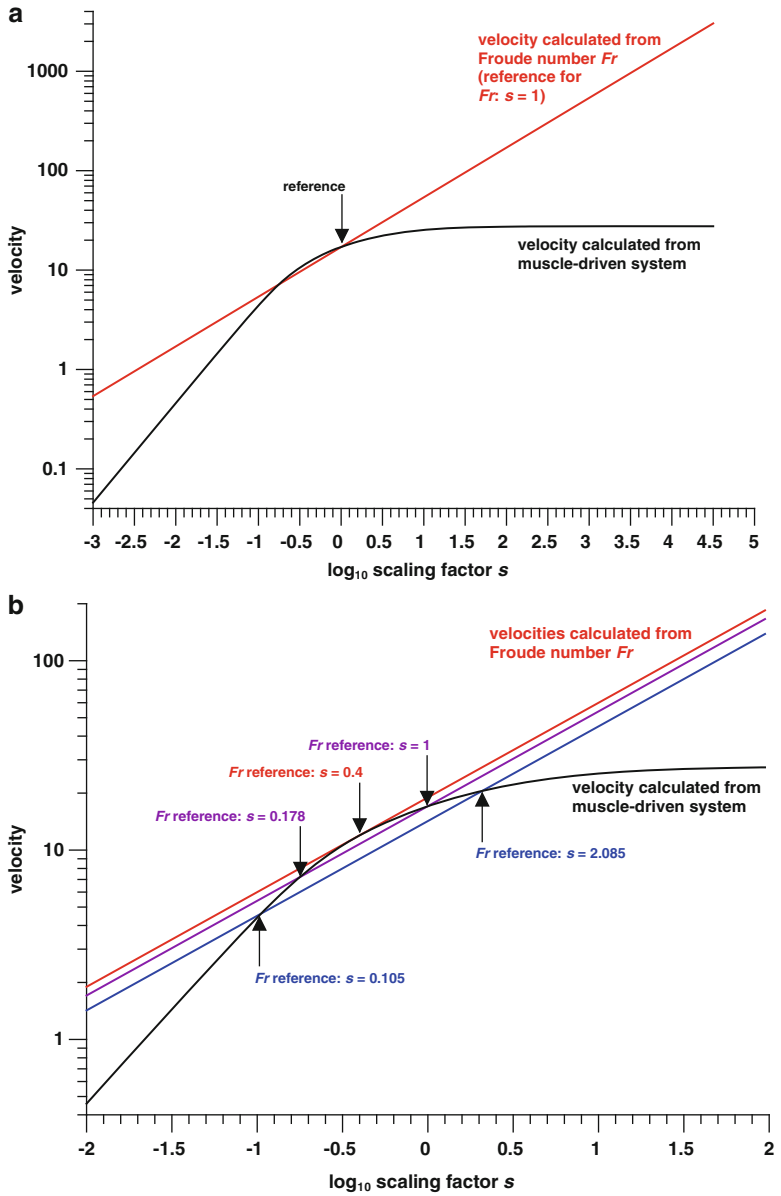



**Fig. 10.8** Velocity from Eq. (10.128) against the logarithm of the scaling factor; comparison of muscle model and Froude number approach; (**a**) Froude number calculated from the reference scaling factor of 1; (**b**) Froude number calculated from different references

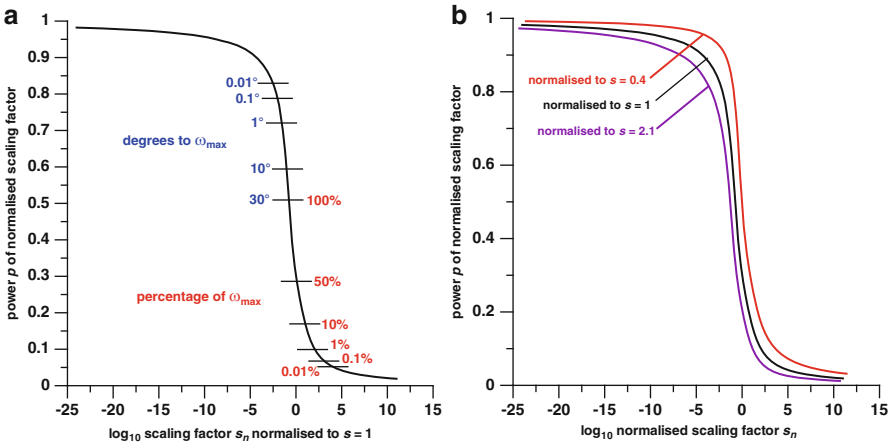



**Fig. 10.9** Power of the normalised scaling factor against the logarithm of the normalised scaling factor; (**a**) *blue labels*: angle in degrees required to reach the maximal angular velocity; *red labels*: cases where the maximal angular velocity cannot be reached indicated by the percentage of the maximal angular velocity which can be reached within the 30° arc of movement; (**b**) power of scaling factor calculated from three different reference scaling factors (color figure online)

The highest velocity estimated from the *Fr* occurs when referenced to a scaling factor of 0.4, where the *Fr* velocity curve is tangent to the shoulder of the muscle-driven system velocity. For other scaling references, the *Fr*-estimated velocity is smaller and the *Fr*-velocity curve intersects the muscle-driven system velocity curve twice, i.e. there are two reference scaling factors per reference *Fr*.

As the power $p$ of the scaling factor $s$ is determined from the normalised scaling factor and normalised speed according to Eqs. (10.129) and (10.130), the power shifts depending on the reference scaling factor (Fig. 10.9). If the reference scaling factor is 1 (Fig. 10.8a), then for scaling factors from 0.13 to 1.525, the velocity predicted from the Froude number lies within the $\pm 10$ % range of the velocity predicted for biological muscle-driven systems.

The fact that the gravitational force was not included in the original ODE (10.83) does not significantly influence the speed of the biological system. Although the gravitational force does not influence the Froude number of the pendulum, it does have an effect on the swing of limbs, due to generation of a torque. However, during locomotion, the swinging movement of the limbs is symmetrical about the vertical plane. There are two halves to this symmetry where one half requires the muscle to work in the direction of the gravitational force and the other half against. The overall effect of the gravitational force is thus cancelled out. In addition, these two halves of the symmetry during the swing coincide with the acceleration and deceleration of the limb by the muscle.

During scaling, muscle forces are changing at a factor of 2 whereas the pendulum forces are changing at a factor of 3. In addition, the mass and moment of inertia are scaled at factors of 3 and 5, respectively, whilst the muscle force, which is dependent

on the muscle's cross-sectional area will lag significantly due to the scaling factor of 2. Consequently, when scaling biological systems, the velocity will reach an asymptotic value, and thereby scales with powers of 0, in spite of increasing the size scaling factor. Hence the Froude number approach overestimates the velocity.

## 10.6 Summary

An ODE with the structure of $ax + b\frac{\mathrm{d}x}{\mathrm{d}t} + cx\frac{\mathrm{d}x}{\mathrm{d}t} = d$ solves with a Lambert function.

This ODE corresponds to the one of a limb segment driven by muscle force and results in an analytical solution.

Based on this analytical solution, the muscle performance parameters can be determined experimentally by fitting the experimental data with the equation of the analytical solution.

This method is useful for muscle function diagnostics, and it even works at conditions of reduced muscle speed or force, such as under fatigue.

The maximal theoretical muscle speed cannot be reached at large scaling factors.

The velocity of biological systems estimated from Froude number scaling with a power 0.5, hinges on the scaling of pendulum with free oscillations; however, limbs and limb segments are muscle-driven links with forced oscillations, scaling with powers between 0 and 1, but not exactly 0.5.

The velocity estimated via the Froude number overestimates the actual muscle-driven limb velocity.

The usage of the Froude number for estimating the velocity of biological systems is mechanically incorrect and an over-simplified approach whose accuracy depends on the reference scaling factor.

## References

Alexander, R. M. (1976). Estimates of speeds of dinosaurs. *Nature, 261*, 129–130.

Alexander, R. M. (1991). How dinosaurs ran. *Scientific American, 264*(4), 130–136.

An, K. N., Kwak, B. M., Chao, E. Y., & Morrey, B. F. (1984). Determination of muscle and joint forces: A new technique to solve the indeterminate problem. *Journal of Biomechanical Engineering, 106*(4), 364–367.

Bottinelli, R., Narici, M., Pellegrino, M. A., Kayser, B., Canepari, M., Faglia, G., et al. (1997). Contractile properties and fiber type distribution of quadriceps muscles in adults with childhood-onset growth hormone deficiency. *The Journal of Clinical Endocrinology and Metabolism, 82*, 4133–4138.

Close, R. (1964). Dynamic properties of fast and slow skeletal muscles of the rat during development. *The Journal of Physiology, 173*, 74–95.

Close, R., & Hoh, J. F. Y. (1967). Force: Velocity properties of kitten muscles. *Journal of Physiology, 192*(3), 815–822.

Corless, R. M., Gonnet, G. H., Hare, D. E. G., Jeffrey, D. J., & Knuth, D. E. (1996). On the Lambert W function. *Advances in Computational Mathematics, 5*, 329–359.

Drillis, R., & Contini R. (1966). *Body segment parameters*. BP174-945, Tech. Rep. No. 1163.03. New York: School of Engineering and Science, New York University.

Euler, L. (1783). De serie Lambertina plurimisque eius insignibus proprietatibus. *Acta Academiae Scientarum Imperialis Petropolitanae of 1779, 3*(2), 29–51 (printed 1783).

Froude, W. (1874). On useful displacement as limited by weight of structure and of propulsive power. *Transactions of the Institute of Naval Architects, 15*, 148–155.

Fuss, F. K. (1993). Helical axis surface of the knee joint. In: *Proceedings of the XIVth Congress of the International Society of Biomechanics* (pp. 438–439), Paris.

Fuss, F. K., Ferdinands, R. E. D., Doljin, B., & Beach, A. (2014). Development of a smart cricket ball and advanced performance analysis of spin bowling. In: *Proceedings of the International Conference in Sports Science & Technology (ICSST)* (pp. 588–595), 11–12 December 2014, Singapore. doi:10.13140/2.1.1222.1766.

Fuss, F. K., Lythgo, N., Smith, R. M., Benson, A. C., & Gordon, B. (2011). Identification of key performance parameters during off-spin bowling with a smart cricket ball. *Sports Technology, 4*(3–4), 159–163. doi:10.1080/19346182.2012.726225.

Fuss, F. K., & Smith, R. M. (2013). Should the finger pressure be well distributed across the seam in seam bowling? A problem of precession and torque. *Procedia Engineering, 60*, 453–458.

Fuss, F. K., & Smith, R. M. (2014). Accuracy performance parameters of seam bowling, measured with a smart cricket ball. *Procedia Engineering, 72*, 435–440.

Fuss, F. K., Smith, R. M., & Subic, A. (2012). Determination of spin rate and axes with an instrumented cricket ball. *Procedia Engineering, 34*, 128–133.

Fuss, F. K., & Tan, M. A. (2006). Maximal muscle velocity: Reality or theory? In: *Proceedings of the 15th International Conference on Mechanics in Medicine and Biology ICMMB*, Singapore, 6–8 December 2006.

Hatze, H., & Baca, A. (1992). Contact-free determination of human body segment parameters by means of videometric image processing of an anthropomorphic body model. In A. G. Tescher (Ed.), *Applications of digital image processing XV* (Proceedings of SPIE, Vol. 1771, pp. 536–546).

Herzog, W. (1999). Muscle. In: B. M. Nigg & W. Herzog (Eds.), *Biomechanics of the musculoskeletal system*. Chichester: Wiley.

Hill, A. V. (1938). The heat of shortening and the dynamic constants of muscle. *Proceedings of the Royal Society of London. Series B, Biological Sciences, 126*(843), 136–195.

Hill, A. V. (1970). *First and last experiments in muscle mechanics*. London: Cambridge University Press.

Janson, S., Knuth, D. E., Łuczak, T., & Pittel, B. (1993). The birth of the giant component. *Random Structures and Algorithms, 4*, 233–358.

Jayes, A. S., & Alexander, R. McN. (1982). Estimates of mechanical stresses in leg muscles of galloping Greyhounds (Canis familiaris). *Journal of Zoology, 198*, 315–328.

Kram, R., Domingo, A., & Ferris, D. P. (1997). Effect of reduced gravity on the preferred walk-run transition speed. *The Journal of Experimental Biology, 200*(4), 821–826.

Lambert, J. H. (1758). Observationes variae in mathesin puram. *Acta Helvetica, physico-mathematico-anatomico-botanico-medica, 3*, 128–168.

Lambert, J. H. (1770). Observations Analytiques. *Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres, 1*, 225–244.

Murray, W. M., Buchanan, T. S., & Delp, S. L. (2000). The isometric functional capacity of muscles that cross the elbow. *Journal of Biochemistry, 33*, 943–952.

Nygaard, E., Houston, M., Suzuki, Y., Jorgensen, K., & Saltin, B. (1983). Morphology of the brachial biceps muscle and elbow flexion in man. *Acta Physiologica Scandinavica, 117*, 287–292.

Nygaard, E., & Sanchez, J. (1982). Intramuscular variation of fiber types in the brachial biceps and the lateral vastus muscles of elderly men: How representative is a small biopsy sample? *Anatomical Record, 203*, 451–459.

Segel, L. A., & Slemrod, M. (1989). The quasi-steady-state assumption: A case study in perturbation. *SIAM Review, 31*, 446–477.

Sjogaard, G. (1982). Capillary supply and cross-sectional area of slow and fast twitch muscle fibres in man'. *Histochemistry, 76*, 547–555.

Spector, S. A., Gardiner, P. F., Zernicke, R. F., Roy, R. R., & Edgerton, V. R. (1980). Muscle architecture and force-velocity characteristics of cat soleus and medial gastrocnemius: Implications for motor control. *Journal of Neurophysiology, 44*, 951–960.

Vaughan, C. L., & O'Malley, M. J. (2005). Froude and the contribution of naval architecture to our understanding of bipedal locomotion. *Gait and Posture, 21*, 350–362.

Weisstein, E. W. (2015). Lambert's transcendental equation. From MathWorld—a Wolfram web resource. Available: http://mathworld.wolfram.com/LambertsTranscendentalEquation.html

Wells, J. B. (1965). Comparison of mechanical properties between slow and fast mammalian muscle. *The Journal of Physiology, 178*, 252–269.

Williams, B. W. (2010). The utility of the Lambert function W[a exp(a-bt)] in chemical kinetics. *Journal of Chemical Education, 87*, 647–651.

Winter, D. A. (1990). *Biomechanics and motor control of human movement*. Chichester: Wiley.

# Chapter 11
# Optoelectronics Modelling of Thin Film Solar Cells

**Nithya Saiprasad, Stefania Castelletto, and Alberto Boretti**

**Abstract** This book chapter summarizes novel designs of thin film solar cells and describes the methods used to compute their optical and electrical performance. The latest trends proposed as well as the state-of-the-art simulation software tools are discussed and explained.

**Keywords** Solar cells • Plasmonics • Optical simulations • Electric simulation • Finite difference time domain • Maxwell's equation • Poisson equation

## List of Symbols

| | |
|---|---|
| $\lambda$ | Wavelength |
| $\nu$ | Frequency |
| $\omega$ | Angular frequency |
| $h$ | Planck's constant $= 6.62606957 \times 10^{-34}$ m$^2$kg/s, $\hbar = h/\pi$ |
| $E_g$ | Energy gap |
| F, $F_\lambda$ | Total and spectral Solar Radiance |
| $E$ | Electric field |
| $\Xi$ | Solar cell efficiency |
| B, $B_\lambda$ | Radiance, spectral radiance |
| $q$ | Charge |
| $I$ | Current |
| $V$ | Voltage |
| $J$ | Current density |
| $c$ | Speed of light in air or vacuum $3.00 \times 10^8$ m/s |

N. Saiprasad • S. Castelletto (✉)
School of Aerospace, Mechanical and Manufacturing Engineering, RMIT University, PO Box 71, Bundoora, VIC 3083, Australia
e-mail: stefania.castelletto@rmit.edu.au

A. Boretti
Department of Mechanical and Aerospace Engineering (MAE), Benjamin M. Statler College of Engineering and Mineral Resources, West Virginia University (WVU), P.O. Box 6106, 325 Engineering Sciences Building, Morgantown, WV 26506, USA



R.N. Jazar, L. Dai (eds.), *Nonlinear Approaches in Engineering Applications*,
DOI 10.1007/978-3-319-27055-5_11

| | |
|---|---|
| $I_{sc}$ | Short circuit current |
| $V_{oc}$ | Open circuit voltage |
| $\varepsilon$ | Dielectric constant |
| OE | Optical efficiency |
| EE | Electrical efficiency |
| IQE | Internal quantum efficiency |
| EQE | External quantum efficiency |
| $\phi_p$ | Photon flux |
| $\gamma$ | Sun light concentration |
| $n$ | Refractive index |
| $\eta$ | Quantum efficiency of spontaneous emission in radiative recombination centres |
| $L_{abs}$ | Absorbed power |

## 11.1 Introduction

The high demand for efficient conversion of solar energy has brought about many methods (Akimov and Koh 2010) to improve the design, which has led to a varieties of solar cells structures with different fabrication types and optimization methods. The major factors that weigh negatively against deployment of latest solar cell technology and design are its costs and a yet lacking mass production and off-the-shelf fabrication methods, in addition to the still overall very low single solar cell efficiency.

In the recent years, various approaches have been implemented and tested to improve their efficiency mainly by improving the photon harvesting, the photon trapping and by reducing the reflectance on the photovoltaic cells. A relatively intense field of research in the above directions, has been associated to the use of plasmonic nanoparticles to improve light trapping and surface texturing (Chao et al. 2010; Losurdo et al. 2009; Sharma et al. 2014; Zhang et al. 2014). However, plasmonic metallic particles introduce parasitic absorption and recombination due to free carriers' concentration in addition to produce the excitation of favourable localized surface plasmonic resonances in metal nanoparticle arrays. The plasmonic enhancement is typically also narrow bandwidth and in specific visible region of the spectra. In addition, plasmonic nanoparticles are not active material for the photovoltaic processes, thus there is a need of optimization of the density of nanoparticles to achieve the maximum efficiency (Pillai and Green 2010), which is not a methodology achievable in a commercial scale production.

In this chapter, we will first cover the basics of Sun radiation and photovoltaic effect, and then discuss recent thin film solar cells designs and modelling approaches, that have been intensively investigated for improving the efficiency of thin film solar cells, based on surface texturing or plasmonic nanoparticles.

## 11.2 Description of the Source of Sun Energy: Sun Spectral Irradiance

The sun behaves much like a high temperature blackbody radiation source with a total power density of $5.961 \times 10^7$ W/m$^2$ at its surface. The total flux of solar radiation reaching the top of the earth's atmosphere, per unit surface normal to the solar beam at the mean distance between the sun and the earth, is reduced to 1353 W/m$^2$ ($\pm 1.6$ %). The solar spectrum can be approximated by a blackbody radiator at a temperature $T_{\text{sun}} = 5762 \pm 50$ K, using Planck's radiation law for the spectral radiance $B_\lambda$ (W nm$^{-1}$ m$^{-2}$ sr$^{-1}$), given by:

$$B_\nu(T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{\left(\frac{h\nu}{K_B T}\right)} - 1}$$

Or

$$B_\lambda(\lambda) = \frac{2hc^2}{\lambda^5} \frac{1}{e^{\left(\frac{h\nu}{K_B T}\right)} - 1},$$

where $T$ is the absolute temperature, $K_B$ is the Boltzmann constant, $h$ is the Planck's constant, $c$ is the velocity of light in vacuum or free space, $\lambda$ is the wavelength of light and $\nu$ is the frequency of light.

The solar radiation spectra at the top of the atmosphere and at the sea level are different due to the light absorption in the atmosphere.

The sun illumination can be measured by the spectral irradiance, defined as the amount of power incident per unit area per unit wavelength:

$$F_\lambda(\lambda) = 2B_\lambda(\lambda) \quad \left[\text{W/m}^2\text{nm}\right],$$

while considering the entire spectrum, the total power per unit area, i.e. the total solar irradiance is given by:

$$F = \int_0^\infty F_\lambda(\lambda)$$

For a known surface area, A, of a solar cell, the total absorbed power over all wavelengths is $L_{\text{abs}} = F \cdot A$.

Measurements of the sun spectral irradiance are performed in different conditions and are referred to air mass (AM) conditions.

In Fig. 11.1 we report some reference Sun spectra used in the evaluation of solar cells performance.

The AM is defined as AM$= \frac{1}{\cos\theta}$, where $\theta$ is the sun zenith angle; when the sun is directly overhead and the spectral irradiance is measured on a horizontal surface after the atmosphere at the sea level, the Air Mass $= 1$ or AM1 (AM0 is referred to

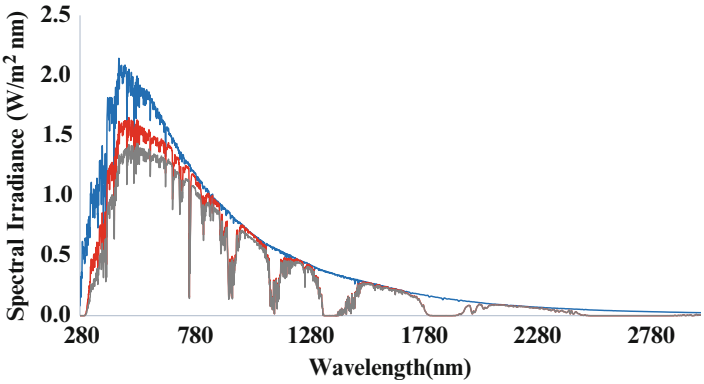



**Fig. 11.1** Solar Standard Radiation Spectra (ASTM G173) represent: extraterrestrial spectral irradiance (AM0, *blue*), Direct Normal Spectral Irradiance (AM1.5D *grey*) and the Global Total Spectral Irradiance (AM1.5G *red*) measured on tilted surface at the 37° and for specific atmospheric conditions

the spectral irradiance measured at the entrance of the Earth's atmosphere), while AM = 1.5 corresponds to the sun spectral irradiance on a tilted surface at 37° and with zenith at $\theta = 48.2°$.

However, there are two distinctive AM 1.5 spectra measured, 1.5D and 1.5G, corresponding to direct radiation and the sum of both direct and diffuse radiation, as non-concentrating solar panels are sensitive also to non-direct Sun radiation. These standard solar spectra are fundamental common reference for evaluating photovoltaic (PV) devices with respect to their performance if typically measured using artificial sources of light.

## 11.3 The Photovoltaic Effect and Its Efficiency in Solar Cells

The well-known principle of a photosensitive device like a solar cell is the excitation of charges by the absorbed sunlight and transfer of this energy to create electrons and holes in the semiconductor conduction and valence band. Figure 11.2a presents the energy diagram of a single junction solar cell.

By absorbing an incident sunlight photon (red arrow) with energy, $\hbar\omega$, an energized electron is excited from the valence band (V-band) to the higher energy levels of the conduction band (C-band), if the incident photon energy exceeds the band gap of semiconductor, $E_g$, thus $\hbar\omega > E_g$. The energy gap or band gap, $E_g$, is an energy range in a solid where no electron states or electron energy level can exist in the hypothesis of ideally pure material. The excited electrons from the V-band leave behind mobile vacancies, called holes, which are positively charged. The excited electron from the upper energy level of the C-band excited state first thermalizes (yellow) by decaying to the lowest energy level in the conduction band,

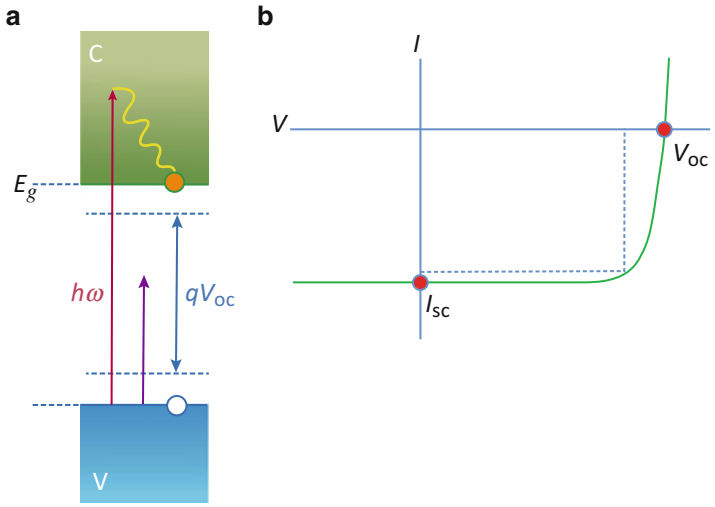



**Fig. 11.2** Illustration of a single junction Solar cell and its electrical characteristics. (**a**) Energy diagram. (**b**) Exemplary electrical current–voltage output of a solar cell. Reprinted by permission from Macmillan Publishers Ltd: Nature Materials (Harry A. Atwater & Albert Polman, Photonic design principles for ultrahigh-efficiency photovoltaics, Nature Materials 11, 174–177 (2012) doi:10.1038/nmat3263), copyright (2012)

before being available as free carriers. Therefore, when light strikes a solar cell, the photon energy may be dissipated as heat (no electron is excited), or it may energize an electron making it is available to flow as current. Light with energy below the band gap (purple arrow, $\hbar\omega < E_g$) is not absorbed. The efficiency of a solar cell is therefore limited by the formation of a thermalized electron-hole pair per absorbed photon above $E_g$.

In Fig. 11.2b the typical current–voltage ($I$–$V$) characteristics curve of the solar cell is shown. The short circuit current ($I_{sc}$) is the current available (generated) in the solar cell when the voltage is zero, it is determined by the generation and collection of charge carriers when photons are incident and it is directly related to the cell conversion efficiency from incident photons to electrical current. The open circuit voltage, $V_{oc}$, is the maximum available voltage when the current is zero in a solar cell and its maximization is key for achieving higher conversion efficiencies, as described later. The dotted lines indicate the maximum power point of a solar cell. Once an electron–hole pair is generated, the electron and hole should be separated and driven to collection at separate electrodes because of the potential at a p-n junction within the semiconductor. In a standard p-n junction, this occurs by diffusion: the charge carriers diffuse to a p-n junction within the semiconductor and are separated. In a p-i-n device the entire device drives the carriers to separation by drift rather than diffusion. Once the carriers are separated they are collected by the electrodes, generating current.

Active photovoltaic absorbers are based on semiconductors with various band gap energies, which selectively absorb only certain spectral range of photons. In addition distinction is made from direct and indirect band gap semiconductors. Indirect band gap semiconductors such as Si, which is the most used for visible photon absorption, require thick films (several microns) before the photons are absorbed and can generate carriers. Therefore Si-solar cells need to be 'optically thick' to allow high light absorption and high photocarrier current production.

In Fig. 11.3a, the standard AM1.5 solar spectrum is compared to its absorbed spectrum by a single pass 2 μm-thick crystalline Si film. Due to the Si band gap the spectral region between 600 and 1100 nm is not greatly absorbed. Industrial developed conventional crystalline Si-solar cells are in fact produced with larger thickness (180–300 μm). On the other hand, the carrier diffusion length ($L_d$) must be much larger than the material thickness to allow efficient photocarriers diffusion from the location where they have been generated and towards the p-n junction (Fig. 11.3b). This requirement is difficult to meet in thick solar cells, as charge carriers generated at distance larger than $L_d$ from the p-n junction are prone to recombination with other charges in the bulk material. Therefore a compromise must be reached in the solar cell design and materials adoption to match opposed constraints given by optical absorption thickness and carrier collection length.

The above considerations brought to the recent advent of the so-called second generation solar cells, based on thin films, as well as to "third generation" based on ultrathin films, this last also invoked to reduce the cost associated to material synthesis at the expense of reduced efficiency in the film by optical losses (lack of absorption, thermalization, reflection, transmission), surface and bulk charges recombination.

To aim to an efficient and cost-effective solar cell, absorption should be in the entire range of solar spectrum with optical minimal losses and the collection of the charge carriers should be improved.

## 11.4 Solar Cell Design Trends

Based on the previous section considerations, research has moved to identify pathways to reduce optical losses (the major issue in second and third generation cells) by increasing the light trapping mechanism within the photovoltaic cells. The very first approach was based on increasing the path length of the incident photon and by depositing efficient antireflection coating on the front surface and introduce back reflectors.

Increasing the path length has been achieved using surface texturing, surface coating and lately surface plasmonic.

The textures mainly include upright or inverted pyramids and moth eye. Random textures have shown a significant improvement in the efficiencies of such solar cells. However, structural types like pyramids are more suited for thicker mono-crystalline films compared to thin films, as thin films are a few nanometres or few microns in thickness.

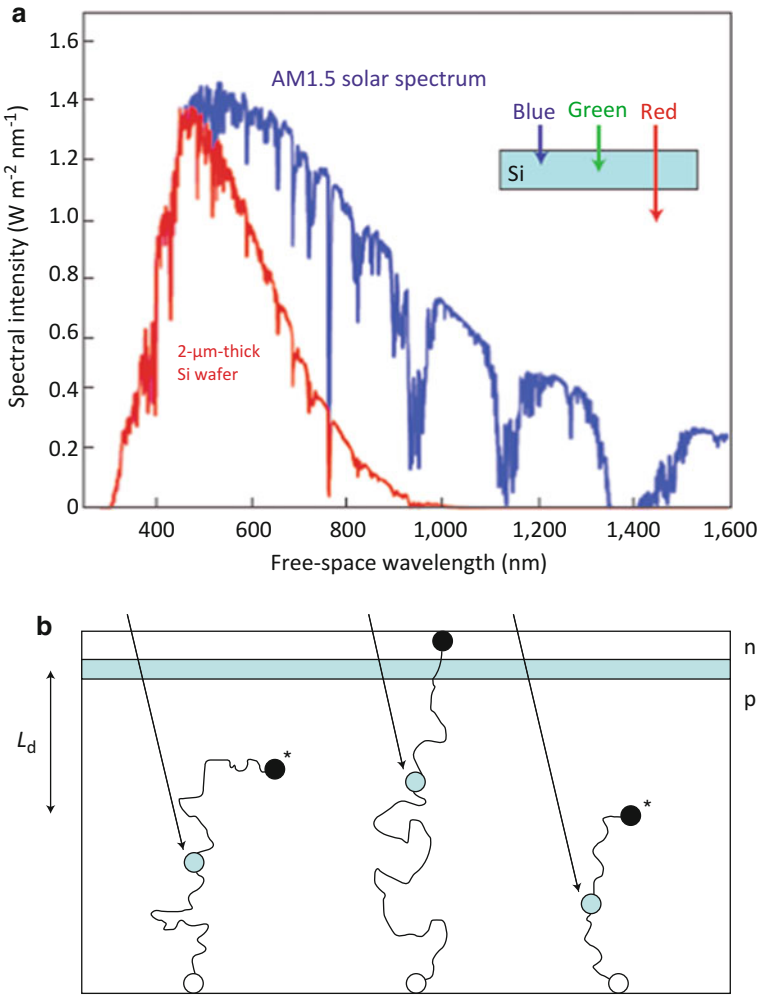



**Fig. 11.3** Illustration of the effects of band gap absorption and carrier diffusion in a Si-solar cell. (**a**), AM1.5 solar spectrum, and its single pass absorption spectrum with no reflection in a 2 μm-thick crystalline Si film (**b**), Illustration of typical carrier diffusion from their localized generation in the bulk away from the p-n junction more than $L_d$ (*blue dots*), showing the cases where recombination with free carries occurs in the bulk (*asterisk*). Reprinted by permission from Macmillan Publishers Ltd: Nature Materials (Harry A. Atwater & Albert Polman, Plasmonics for improved photovoltaic devices, Nature Materials 9, 205–213 (2010) doi:10.1038/nmat2629), copyright (2010)

To absorb nearly all of the incident sunlight, as described in the previous section, a solar cell is typically engineered to be optically thick and textured to improve light trapping.

When a photon is incident on the material, depending on the angle, there is a possibility of reflection from the surface, total internal reflection with multiple reflections, scattering and transmission through the active layer and further reflection from the back surface. Bare silicon has a high surface reflection of over 30 %, which can be reduced by texturing and by applying antireflection coatings to the surface. The same is valid for the back surface to trap for longer time the light within the cell.

Surface texturing includes creating morphological roughness on the surface of the thin film (Alongkarn et al. 2011). The corrugation on the surface (either on the front side or rear side or both) can be created during the manufacturing of the thin films (Sai et al. 2010). This corrugation increases the path length of incident photon and hence multiple reflections take place. Some of the studies show that this texturing improves the efficiency of the thin film especially when they are used on a tandem cell with such textures on front and back (Escarré et al. 2012). However the major drawback is that the thickness of the thin film solar cell will be very much reduced in doing so, leading to lesser active material absorption.

Two types of textures are usually adopted, on the front surface or both front and back surfaces to increase the incident photon path and scattering.

Surface coating has shown a significant enhancement in the light trapping efficiency. From white paint being cheapest to aluminium, silver, polyvinyl butyral (PVB) and many other metals are used to study the enhancement efficiency in light trapping. However white paint is normally used for a large scale production, but silver as a back reflector is widely studied. The enhancement efficiency when studied with the different type of antireflection coating PVB has to be 8.3 % for a-Si:H solar cell whilst 10.1 % enhancement is seen with an a-Si:H/mc-Si:H tandem solar cell (Moulin et al. 2013; Zhu et al. 2013).

The absorption of light cannot be enhanced in a randomly textured material by a factor larger than $4n^2/\sin^2\theta$ ($n$ is the refractive index of the material and $\theta$ is the angle of the emission cone in the medium surrounding the cell). This is generally referred as limit of light trapping or Yablonovitch's limit (Yu et al. 2010). The Yablonovitch's limit can be reached and eventually surpassed in thin film solar cells by creating randomly submicron (few hundreds nanometers) textured surfaces. In fact, this limit is not valid at the nanoscale and it can be substantially exceeded by confining the optical modes in the active layer at the sub-wavelength scale. This recent observation opened new avenues for improved light harvesting moving from mesa scale surface texturing to sub-wavelength nano-structuring and towards other methods to improve the light harvesting based on plasmonic effects.

Surface plasmonic effects are a more recent nano-optics discovery and therefore many studies have emerged in this effect potential application in engineering solar cells. Metal nanoparticles or thin metallic nanostructured are used to produce plasmonic effects when deposited on the surface of thin film solar cells or fabricated in place of the back reflectors, respectively.

There are different methods to use surface plasmonic effects to enhance light trapping ability in thin film solar cells.

One method is by incorporating nanostructures as sub-wavelength scattering elements to trap the sun's rays in the absorbing layers. The sub-wavelength nanoparticles scatter the incident photon at different angles and thus increase the path of the light in the absorption layer and enhance the short circuit current density. In the intermediate surface between a metal and a dielectric or semiconductor, the electrons excited by incident photons are called localized surface plasmons (LSPs).

The excitation of the LSPs, by enhancing the strong local field around the nanoparticles, aid in increasing the absorption around the semiconductor due to high localization of the incident light in the near field. This includes the mostly studied enhancement of light absorption facilitated by gap modes between the nanoparticles.

Due to the shape and size of these nanoparticles, optical excitation of LSP occurs at particular frequencies. However, it is known that metals also act as recombination centres, thus using metal nanoparticles here lead to further recombination centres or defects at the interface of these nanoparticles and semiconductor. The variation of the size and shapes and the distance of these nanoparticles from the surface also play important roles on the light trapping efficiencies.

Another method is to excite surface plasmon polaritons (SPPs) at the interface of metal and semiconductors, where light is converted into SPPs, which are electromagnetic waves that travel along the semiconductor-absorber layer and the metal back contact interface. The electromagnetic SPP fields, which are localized near this interface, are smaller than the wavelength of light, and this happens near the plasmon resonance frequency. At the metal and semiconductor interface, the SPPs can efficiently trap and guide the light to the semiconductor layer.

Figure 11.4 illustrates the above described different geometries used in plasmonic light trapping for thin film solar cells. In Fig. 11.4a, the metallic nanoparticles placed on the solar cell surface are used as scatterers of the incident photons, thus increasing the photons trapping within the cell and the probability of the light to be absorbed by the active semiconductor layer. This preferential scattering of sunlight is trapped onto the semiconductor thin film by making use of high angle and multiple scattering, thereby increasing the effective optical length of the solar

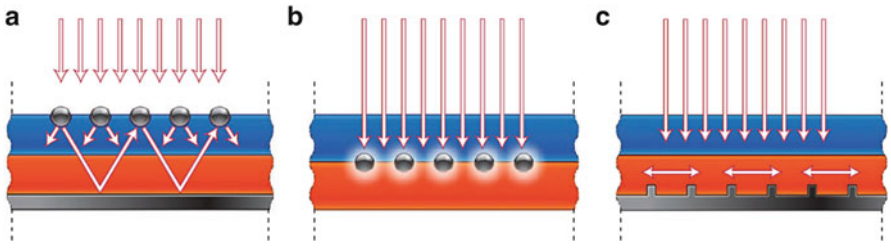



**Fig. 11.4** Plasmonic light-trapping geometries for thin film solar cells based on different effects. (**a**) Scattering effects. (**b**) Excitation of LSPs in metal nanoparticles. (**c**) Excitation of SPPs at the metal/semiconductor interface. Reprinted by permission from Macmillan Publishers Ltd: Nature Materials (Harry A. Atwater & Albert Polman, Plasmonics for improved photovoltaic devices, Nature Materials 9, 205–213 (2010) doi:10.1038/nmat2629), copyright (2010)

cell. In Fig. 11.4b, a different mechanism is illustrated where the incident photons striking on the metal nanoparticles generate resonant LSPs, due to free electron in the metal. Resonant LSP excitations in the nanoparticles embedded in the semiconductor produce a strong local field enhancement around the metal nanoparticles that increases the absorption in the surrounding semiconductor material. These surface plasmons therefore enhance in the near field, the absorption of the incident photons, only closely to the nanoparticles inserted in the active semiconductor. In Fig. 11.4c, SPPs are excited along the interface of metal–semiconductor by the incident trapped photons. In this geometry the incident solar photons are absorbed along the lateral direction of the solar cell extending the extension of the optical absorption length. The metal back surface of the film which is corrugated combines the light with the SPPs to propagate in the same plane as that of the semiconductor.

Noble metals showing plasmonic effects possess negative value of the real part of the dielectric constant. Amongst the noble metals, silver and gold have shown LSPs and SPPs in visible and near infrared regions. Due to their high cost these metals are not ideal for large scale applications and thus aluminium has been used for trapping the incident photon efficiently also in the UV regions. The major challenge at present is that, when metals are used to produce surface plasmons, the photon energy incident on the particles is also converted into heat, thus reducing the cell efficiency further. In addition, the metals provide free carriers to the device thus increase recombination losses at the front and back surfaces. This major drawback can be partially reduced by TCO (transparent conducting oxides) as intermediate layer between the metal nanoparticles and active layer. From the recent literature numerous studies have focussed on the design of the optimum and most efficient metals, nanoparticles shape, providing surface plasmonic effect and intermediate dielectric layer. There have been several studies where metal oxides like ZnO, IndiumTinOxide (ITO) were used along with other metals, to reduce the losses of light due to metal absorption, which occurs in the near infrared region and limits the beneficial plasmonic effect of the material for increasing the cell efficiency.

More recently it has been discovered that some heavily doped metal oxides amongst them ZnO doped with aluminium and other materials like nitrides (see Fig. 11.5) can exhibit plasmonic effects, without introducing losses as in metal. The use of these reduced loss plasmonic materials has recently been proposed to replace metals in devices relying on plasmonic effects, solar cells being one of them. A recent study shows that nanometric film and nanoparticles of heavily doped aluminium ZnO possesses plasmonic resonance in the NIR (Naik et al. 2012). Figure 11.5 presents metamaterial (specifically engineered materials) that could be used for solar cells applications or other applications, where metals are needed. It represents a metamaterial map, where materials are classified on the basis of the optical properties of conducting materials, such as carrier concentration (maximum doping concentration for semiconductors), carrier mobility and interband losses. The carrier concentration is responsible for a negative real permittivity, while high carrier mobilities and interband losses are associated to various material losses. The ideal material for plasmonics would have almost zero interband losses, though such a lossless metal remains abstract. Semiconductors can have plasmonic properties

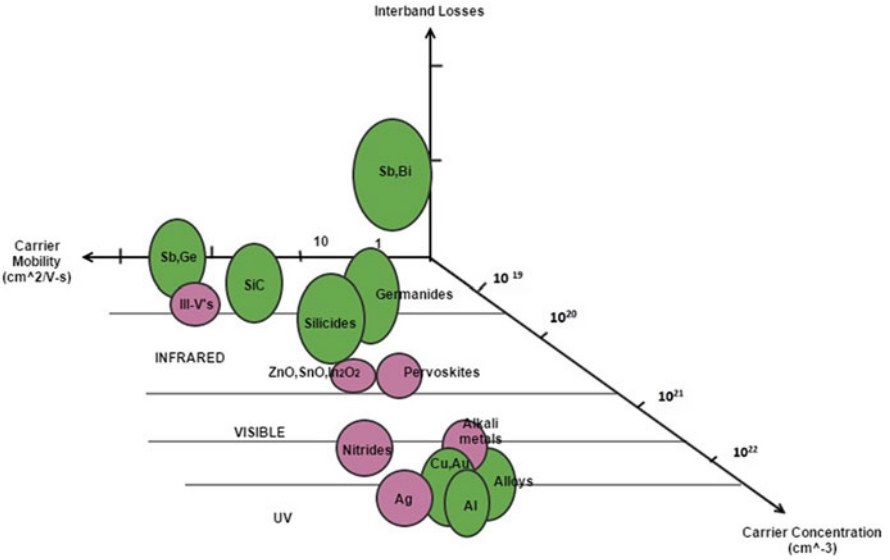



**Fig. 11.5** Metamaterial map (after Boltasseva and Atwater 2011)

and are suitable for applications at near-infrared (NIR) and mid-infrared (MIR) wavelengths. Plasmonics in the visible range with transparent conductive oxide require higher doping. To reduce the carrier concentration in metals it possible to mix them with non-metals obtaining silicides, germanides, borides, nitrides, oxides and metallic alloys. The optical properties of titanium nitride approach those of gold for plasmonic applications above 550 nm, thus ideal for plasmonics solar cell or other light managing architectures, such as described later in Fig. 11.7.

In Fig. 11.5, spheres represent materials with low interband losses, and ellipses represent materials with larger interband losses in the identified part of the electromagnetic spectrum.

Plasmonic effects are combined with solar cells mostly based on amorphous Si thin film. Amorphous hydrogenated silicon solar cells (a Si:H) are in fact preferred than crystalline silicon (c Si)thin film solar cells as they are easy and cheaper to fabricate and that they still absorb light in the visible region of solar spectrum (with band gap 1.7 eV compared to c-Si 1.1 eV). Light induced degradation is pronounced in a Si:H solar cells and losses due to the sp2-dangling bonds in the lattice structure make their efficiency smaller when compared to crystalline silicon.

Figure 11.6 outlines some plasmonic solar cell design principles sprung up in the recent years. Figure 11.6a shows a plasmonic tandem solar cell, which is sandwiching the solar cells of various bandgap active materials to harness a wider range of solar spectrum. Figure 11.6b depicts the ultrathin quantum dots layers along the intermediate layers in the metals like Ag/insulator/metal SPP waveguide to couple the SPP modes and thereby enhancing the absorption of incident photon in quantum dots solar cells. In Fig. 11.6c the light is focussed between the nanoscale

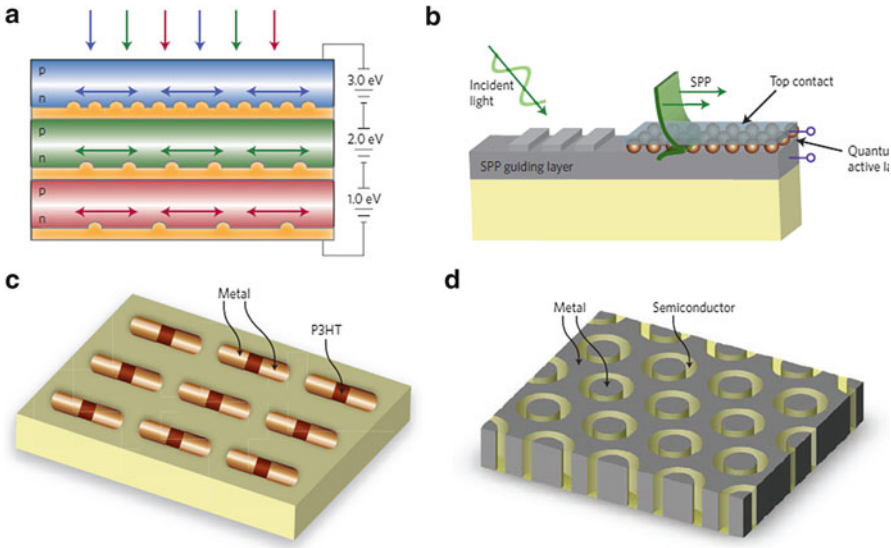



**Fig. 11.6** Plasmonic solar cell designs. (**a**) A tandem multi-junction solar cell geometry with each a SPPs at the rear surface. (**b**) Plasmonic ultrathin quantum-dot layers solar cell where enhanced photo absorption in the mediated by coupling light to SPPs modes in Ag propagating in the plane of the interface between metal and the active-dots layer. (**c**) Array of optical antenna formed by an heterostructure combining metal and poly (3-hexylthiophene) (P3HT). (**d**) Array of coaxial holes in a metal film that support localized Fabry–Perot plasmon modes. Reprinted by permission from Macmillan Publishers Ltd: Nature Materials (Harry A. Atwater & Albert Polman, Plasmonics for improved photovoltaic devices, Nature Materials 9, 205–213 (2010) doi:10.1038/nmat2629), copyright (2010)

gap of an optical antenna array made from an axial heterostructure of metal and poly (3-hexylthiophene) (P3HT), thus producing photocurrent in P3HT semiconductor. Figure 11.6d shows an array of coaxial holes in a metal film. These holes are filled with inexpensive semiconductors possessing short lifetime of minority carriers and aid localized Fabry–Perot plasmon modes. The metals along the inner and outer layers of these coaxial holes collect the charge carriers. This setup can enhance the field up to a factor of 50 and thereby increasing non-linear photovoltaic conversion effects.

Apart from record single junction solar cell efficiency (around 28 %) and multi-junction solar cells under concentrated illumination (around 44 %), to achieve ultrahigh efficiency solar cells the main limitation in current design that needs to be addressed is related to thermodynamics losses (the creation of a thermalized electron-hole after a photon absorption) (Atwater and Polman 2012). Thermodynamics effects in fact can describe the limit of the maximum $V_{oc}$, that in the case of the assumption of radiative recombination limit and full collection of all carriers, is given by:

$$q\,V_{\mathrm{oc}} = E_{\mathrm{g}}\left(1 - \frac{T_c}{T_{\mathrm{sun}}}\right) - kT\left[\ln\frac{\Omega_{\mathrm{emit}}}{\Omega_{\mathrm{sun}}} + \ln\frac{4n^2}{C} - \ln\gamma\right],$$

where $T_{\mathrm{c}}$ is the cell temperature, $T_{\mathrm{sun}}$ is the Sun temperature, $\Omega_{\mathrm{sun}}$ is the Sun solid angle ($\Omega_{\mathrm{sun}} = 6 \times 10^{-5}$sr), $\Omega_{\mathrm{emit}} = 4\pi$ is the solid angle of spontaneous re-irradiated photons from the semiconductor, $n$ is the refractive index and $\gamma$ is the sun light concentration. The terms in square brackets are increase of entropy (thus reducing $V_{\mathrm{oc}}$), specifically the term $\ln\frac{\Omega_{\mathrm{emit}}}{\Omega_{\mathrm{sun}}}$ is an entropy increase due to the absorption of a photon from the Sun and re-irradiation from the active material in the cell, $\ln\frac{4n^2}{\gamma}$ is an entropy increase due to optical losses or poor absorption, and $\ln\eta$ is associated to radiative recombination of electrons or holes. For a planar solar cell $\gamma = 1$, corresponding to a $V_{\mathrm{oc}}$ reduction, light trapping will increase $\gamma$ and improve the current at the cell. When the Yablonovitch's limit is reached $\gamma = 4n^2$, thus vanishing the entropy term $\ln\frac{4n^2}{\gamma}$, while $\gamma$ can be increased by nanostructuring. $\ln\eta$ is an increase of entropy due to the quantum efficiency of radiative recombination processes within the semiconductors, in fact free electrons and holes can recombine and emit photons due to some material defects close by after they are generated. The non-radiative recombination is responsible for a quantum efficiency of the process $< 1$ , typically in Si it is 10 %. The non-radiative decay increases heat dissipation. Using light management strategies it is possible to reduce these entropy terms and beat the Shockley–Queisser limit of 33 %, given for $\eta = 1$ and ideal full light absorption and trapping, reaching for instance 40 % for single junction solar cells.

Figure 11.7 presents recently proposed light management designs to achieve ultrahigh efficiency. Figure 11.7a shows the use of parabolic reflectors which deflects the spontaneous emission along the solid angle to reach the Sun's disc. Figure 11.7b shows that by using planar metamaterial it is possible to direct light. Figure 11.7c shows that by using nanoparticles as Mie scatterers on the surface this improves the light trapping. In Fig. 11.7d the spontaneous emission rate can be increased (thus increase $\eta$ and consequently $V_{\mathrm{oc}}$) by enhancing the optical density of states in the metal–dielectric–metal waveguide or semiconductor–dielectric–semiconductor slot waveguide.

Figure 11.8 presents a multi-junction solar cells. In Fig. 11.8a semiconductors possessing various bandgaps are brought together to harness energy from different regions of solar spectrum, thus reducing the loss due to thermalization. In the given diagram the blue dotted lines represent the Fermi level modification that provides a larger $V_{\mathrm{oc}}$ for the different bandgaps. The yellow dots are electrons. In Fig. 11.8b shows a horizontal design of multi-junctions based on an array of in-parallel single junctions fabricated by using techniques such as epitaxial lift off and printing of semiconductor layers. This could be improved by adding a layer of micro or nanophotonic spectrum splitting. To minimize losses due to entropy, each of the semiconductor layers can be integrated with one of the above light management structures and can be optimized individually.

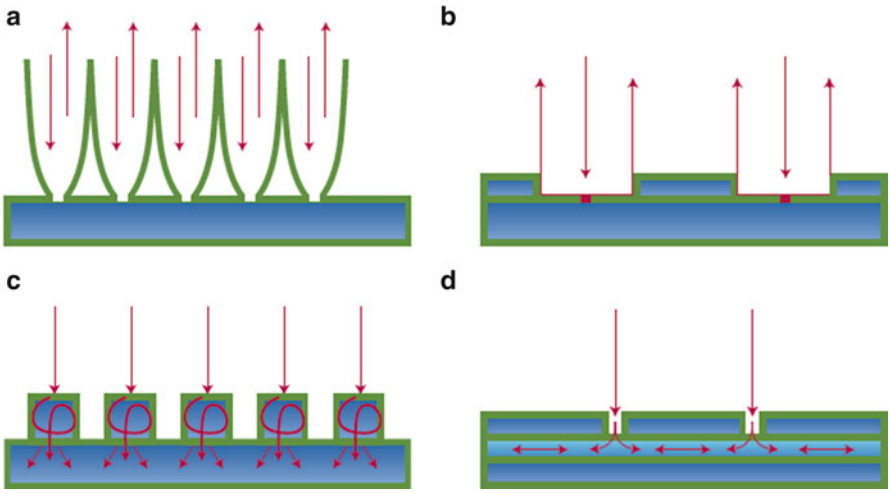



**Fig. 11.7** Light-harvesting based on surface light reflectors and collectors. (**a**) 3D parabolic light reflectors to redirect spontaneous emission to the Sun and thus reducing spontaneous emission heat dissipation within the material. (**b**) Planar meta-material light-director structures. (**c**) Mie-scattering surface nanostructures to increase light trapping. (**d**) Metal–dielectric–metal waveguide or semiconductor–dielectric–semiconductor slot waveguide to enhance the optical density of states of the spontaneous emission from radiative recombination of charges and therefore reducing the non-radiative channels, responsible for heat dissipation. Reprinted by permission from Macmillan Publishers Ltd: Nature Materials (Harry A. Atwater & Albert Polman, Photonic design principles for ultrahigh-efficiency photovoltaics, Nature Materials 11, 174–177 (2012) doi:10.1038/nmat3263), copyright (2012)

## 11.5 Numerical Methods to Analyse Photovoltaic Devices

Many computational tools have been developed to study photovoltaic devices, and many of them are available as commercial simulation packages. One of the most widely used is commercial software by Lumerical (www.lumerical.com) and includes two packaged coupled together, a Maxwell's solver using the finite difference time domain approach (FDTD) describing the optical field, and another product computing the electric field from the optical field (DEVICE). A Lumerical FTDT optical simulation basically computes the fraction of photons absorbed in the semiconductor material from the incident photons at the source, called optical efficiency (OE). This result gives an upper limit to the electrons reaching the electrode. A Lumerical DEVICE electric simulation compute the fraction of electrons collected at the electrodes from the electrons undergoing recombination effects or subject to charge traps in the material. The electrical simulation provides the electrical efficiency (EE) by solving the Poisson and Continuity Equations. The quantum efficiency (QE) is the product of these two efficiencies, $QE = OE \times EE$.

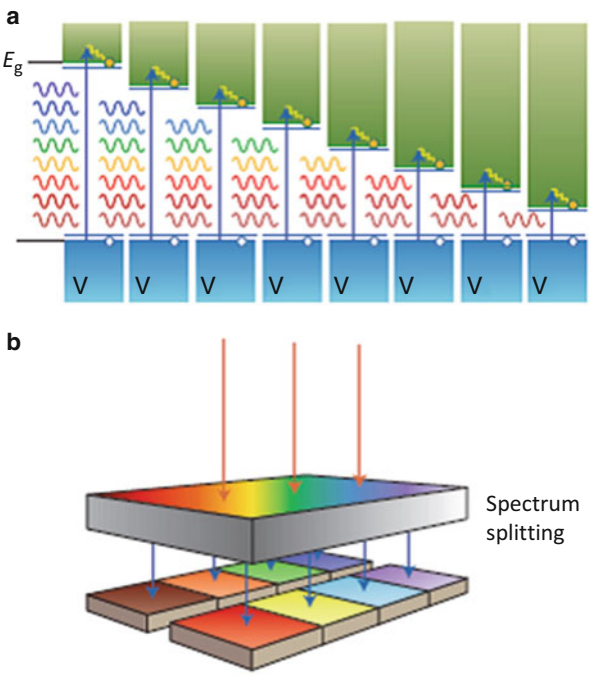



**Fig. 11.8** Multi-junction solar cell. (**a**) Energy diagram using different bandgap semiconductors to convert different portions of the solar spectrum and increase available $V_{oc}$. (**b**) Horizontal architecture for array of single junction cells with Spectrum splitting layers. Reprinted by p/emission from Macmillan Publishers Ltd: Nature Materials (Harry A. Atwater & Albert Polman, Photonic design principles for ultrahigh-efficiency photovoltaics, Nature Materials 11, 174–177 (2012) doi:10.1038/nmat3263), copyright (2012)

The following section describe in details the mathematics of these two simulations approaches, optical field and electric field simulations, common also to other products, underlining the approximations used.

By using simulations, different designs may be analysed quickly and inexpensively, as the construction of prototypes involving sophisticated materials and nanoscale structures are everything but simple and experimental campaign are lengthy and expensive. Obviously, as continuous validation is a necessary prerequisite for any simulation to be trusted, a certain amount of experiments are always necessary, but their number can be reduced with the help of the simulations. The application of nanoparticles responsible for plasmonic effects, surface texturing and antireflection coating for different materials and cells design to improve the performance of solar cell can be studied by using these opto-electric simulations.

By optical simulation based on FDTD, it is possible to compute the photons spatial absorption in the substrate region of the geometry. The electric model calculates the electron collected rate, returning the efficiency of the modelled solar cell.

There are two types of efficiency, often defined and used to provide performance of solar cells in the literature, they are referred to internal quantum efficiency (IQE) and external quantum efficiency (EQE). The EQE coincides with the QE defined above.

IQE is the ratio of amount of charges collected by the solar cell to the number photons per second absorbed.

$$\text{IQE} = (\text{Electrons collected/s}) \,/\, (\text{Photons absorbed/s})$$

When an electromagnetic light, with electric field $E(r, \omega)$, is incident on a material, the amount of power absorbed per unit volume and at angular frequency $\omega$ not only depends on the nature of incident electromagnetic spectrum but also the nature of the material (dielectric permittivity, $\varepsilon(r, \omega)$) given by

$$L_{\text{abs}}(r, \omega) = -\frac{1}{2}\omega \left|E(r, \omega)\right|^2 \text{Im}\left\{\varepsilon(r, \omega)\right\}.$$

As one photon energy is $\hbar\omega$ ($h\nu$), the photon flux absorbed per unit volume and angular frequency is

$$\phi_{p,\text{abs}}(r, \omega) = \frac{L_{\text{abs}}(r, \omega)}{\hbar\omega} = -\frac{\pi}{h}\left|E(r, \omega)\right|^2 \text{Im}\left\{\varepsilon(r, \omega)\right\}.$$

To obtain the photon flux absorbed for the complete spectrum, the above equation has to be integrated for the entire frequencies of the Sun spectrum within AM1.5.

$$\phi_{p,\text{abs}}(r) = \int \phi_{p,abs}(r, \omega)\, d\omega$$

The total photon flux absorbed is considered equivalent to the electron generation rate, assuming that each absorbed photon forms an electron–hole. This is a strong approximation. Other absorption mechanisms that do not contribute to the generation of electron–hole, such as free-carrier absorption, are neglected by this approximation but could be included.

This leads to the internal IQE if the photon absorbed flux is integrated over the cell total volume,

$$\text{IQE} = \frac{I}{q\phi_{p,\text{abs}}},$$

where $\phi_{p,\text{abs}} = \int \phi_{p,\text{abs}}(r)dr$, $I$ is the current collected at the electrodes of the device and $J = I/A$, is the current density.

The EQE is defined as the ratio of charges collected rate to incident photons rate or incident photon flux, $\phi_{p,\text{incident}}$ :

$$\text{EQE} = (I/q) \,/\, \left(\phi_{p,\text{incident}}\right)$$

EQE can also be defines in terms of current density, $J$, of a monochromatic source as

$$\text{EQE} = \frac{hc}{q\lambda_0}\frac{J(\lambda_0)A}{P_{\text{incident}}},$$

where the incident photon flux is defined as the total power incident per unit area ($A$) and per photon energy $\frac{hc}{q}$ of the photosensitive device for a monochromatic source of wavelength $\lambda_0$. In the case of a non-monochromatic source with power spectrum, $P_{\text{incident}}(\omega)$, the incident photon flux is:

$$\phi_{p,\text{incident}} = \frac{1}{A}\int_0^\infty \frac{P_{\text{source}}(\omega)}{\hbar\omega}\mathrm{d}\omega,$$

thus

$$\text{EQE} = \frac{1}{q}\frac{\text{JA}}{\int_0^\infty \frac{P_{\text{source}}(\omega)}{\hbar\omega}\mathrm{d}\omega}.$$

EQE takes into account the optical losses due to reflection and transmission and the electrical losses, hence the EQE is lower than IQE. The QE or EQE is also frequency-dependent, requiring that the electrical response should be simulated using the electron generation rate or photon flux absorbed, $\phi_{p,\text{abs}}(r,\omega)$, for each specified ω. However, an single electrical simulation with a broadband optical input cannot be decomposed into set of electrical responses for each optical wavelengths λ because the electrical response is determined by a system of non-linear equations. As opposite a single broadband optical simulation can be used to calculate the generation rate for a collection of spectra. When running the simulation therefore using a broadband source, first the optical efficiency is computed $\text{OE} = \phi_{p,\text{abs}}/\phi_{p,\text{incident}}$, by FDTD optical simulation, thus determining the generated electrons given by $\phi_{p,\text{abs}}(r)$, then it used to determine the EE and the QE by later performing the electrical simulations. In the electrical simulation procedure the EE can also be determined over a range of operating conditions, including changes in bias voltage and ambient temperature. Often in the literature designers assume an IQE = 100 %, so that the ideal current at the electrode is $I = q\phi_{p,\text{abs}}$. In this case, the EQE can be calculated solely from an optical simulation as $E = \frac{\phi_{p,\text{abs}}}{\int_0^\infty \frac{P_{\text{source}}(\omega)}{\hbar\omega}\mathrm{d}\omega}$, which is a large approximation and results should be carefully considered.

Finally the photovoltaic efficiency, $\Xi = \frac{P_{\max}}{P_{\text{in}}}$, defined as the ratio of the maximum power, that can be delivered to a load compared to the input power can be deduced from the electrical simulations.

The output power of the device is calculated under load: for a constant current, the voltages across the resistance are directly related as $V = IR$. The generated power is then determined as the product of the voltage across the load and the current absorbed by the load.

The photovoltaic efficiency can be calculated from the electrical simulation, the effect of changing the load resistance can be simulated, and the power can calculated as a function of applied bias thus the maximum power can then be determined. By knowing the source power over entire spectrum, the input power can be determined.

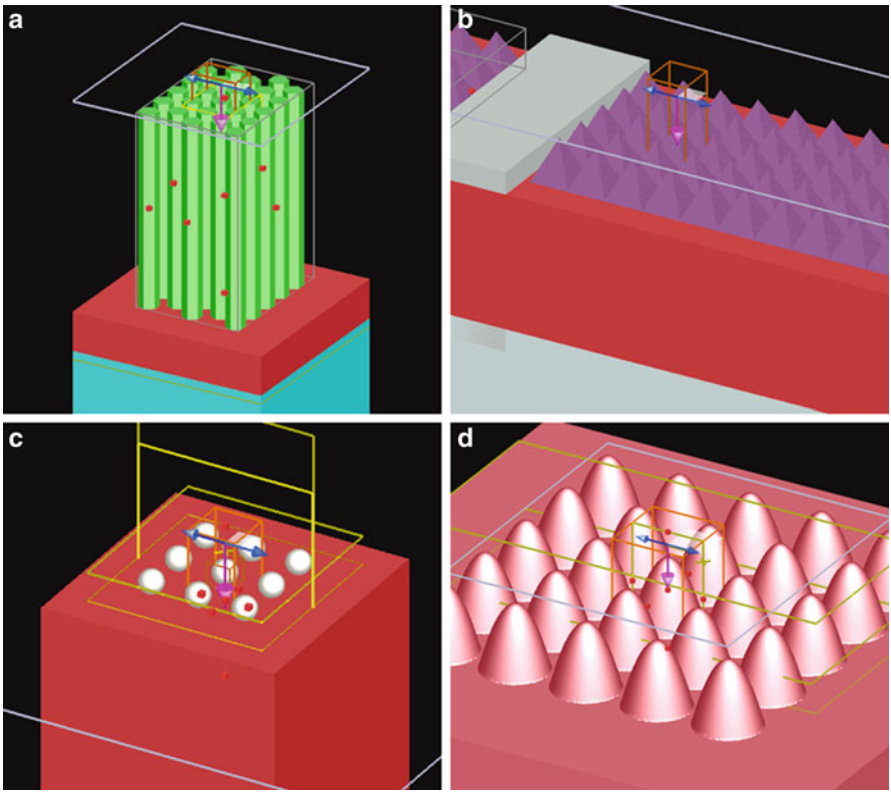



**Fig. 11.9** Sample FDTD (www.lumerical.com) models. (**a**) InAs pillars on silicon substrate solar cell. (**b**) $TiO_2$ pyramids on silicon substrate solar cells. (**c**) Plasmonic solar cell with suspended nano particles of plasmonic materials (Gold, Silver). (**d**) Moth eye silicon solar cell

It can be concluded that both optical and electrical simulations are necessary to determine the generation rate from the power absorbed and introducing the recombination effects from the electrical response, the maximum power delivered by the cell can be determined thus efficiently computing its efficiency.

Microscale and nanoscale structures of the solar cell and the broadband response of the materials can be efficiently addressed using FDTD Solutions.

Figure 11.9 presents fours very popular models developed by using Lumerical FDTD (www.lumerical.com). Figure 11.9a shows an InAs pillars solar cell, b: a $TiO_2$ pyramids on silicon substrate solar cells, c: a plasmonic solar cell with suspended nano particles of plasmonic materials, d: shows a moth eye silicon solar cell. By using a Maxwell's equation solver, optical simulations may be easily set up to study the influence of particles of any shape and any material placed on top of the silicon substrate or also embedded in the silicon substrate in any array.

## 11.6 Conclusions

This book chapter has discussed the physics of solar cells as well as the design issues of thin film solar cells. It has also described the methods used to compute the optical and electrical operation of the solar cell. The latest trends proposed as well as the state-of-the-art simulation software tools are discussed and explained. This book chapter is an up-to-date introduction to what can be done with simulations to optimize these devices, with particular emphasis to thin film structures and their combination with plasmonics effects.

## References

Akimov, Y. A., & Koh, W. S. (2010). Design of plasmonic nanoparticles for efficient subwavelength light trapping in thin-film solar cells. *Plasmonics, 6*(1), 155–161. doi:10.1007/s11468-010-9181-4.

Alongkarn, C., Catherine, W. W. L., Nazir, K. P., & Stefan, Z. (2011). Wave-optical studies of light trapping in submicrometre-textured ultra-thin crystalline silicon solar cells. *Journal of Physics D: Applied Physics, 44*, 1–4.

Atwater, H. A., & Polman, A. (2010). Plasmonics for improved photovoltaic devices. *Nature Materials, 9*, 205–213. doi:10.1038/nmat2629.

Atwater, H. A., & Polman, A. (2012). Photonic design principles for ultrahigh-efficiency photovoltaics. *Nature Materials, 11*, 174–177. doi:10.1038/nmat3263.

Boltasseva, A., & Atwater, H. A. (2011). Materials science. Low-loss plasmonic metamaterials. *Science, 331*(6015), 290–291.

Chao, C. C., Wang, C. M., & Chang, J. Y. (2010). Spatial distribution of absorption in plasmonic thin film solar cells. *Optics Express, 18*(11), 11763–11771.

Escarré, J., Söderström, K., Despeisse, M., & Nicolay, S. E. (2012). Geometric light trapping for high efficiency thin film silicon solar cells. *Solar Energy Materials and Solar Cells, 98*, 185–190.

Losurdo, M., Giangregorio, M. M., Bianco, G. V., Sacchetti, A., Capezzuto, P., & Bruno, G. (2009). Enhanced absorption in Au nanoparticles/a-Si:H/c-Si heterojunction solar cells exploiting Au surface plasmon resonance. *Solar Energy Materials and Solar Cells, 93*(10), 1749–1754.

Moulin, E., Paetzold, U. W., Bittkau, K., Ermes, M., Ding, L., Fanni, L., et al. (2013). Thin-film silicon solar cells applying optically decoupled back reflectors. *Materials Science and Engineering: B, 178*(9), 645–650.

Naik, G. V., Liu, J., Kildishev, A. V., Shalaev, V. M., & Boltasseva, A. (2012). Demonstration of Al:ZnO as a plasmonic component for near-infrared metamaterials. *Proceedings of the National Academy of Sciences of the United States of America, 109*(23), 8834–8838. doi:10.1073/pnas.1121517109.

Pillai, S., & Green, M. A. (2010). Plasmonics for photovoltaic applications. *Solar Energy Materials and Solar Cells, 94*(9), 1481–1486.

Sai, H., Jia, H., & Kondo, M. (2010). Impact of front and rear texture of thin-film microcrystalline silicon solar cells on their light trapping properties. *Journal of Applied Physics, 108*(4), 044505.

Sharma, M., Pudasaini, P. R., Ruiz-Zepeda, F., Vinogradova, E., & Ayon, A. A. (2014). Plasmonic effects of Au/Ag bimetallic multispiked nanoparticles for photovoltaic applications. *ACS Applied Materials & Interfaces, 6*(17), 15472–15479.

Zhang, Y., Stokes, N., Jia, B., Fan, S., & Gu, M. (2014). Towards ultra-thin plasmonic silicon wafer solar cells with minimized efficiency loss. *Scientific Reports, 4*, 4939.

Zhu, H., Huang, Y., Mai, Y., Wang, Y., Yin, J., Feng, Y., et al. (2013). Study of silicon thin film solar cells with different back-reflectors. *Physica Status Solidi (a), 210*(6), 1143–1148.

Yu, Z., Raman, A., & Fan, S. (2010). Fundamental limit of nanophotonic light trapping in solar cells. *Proceeding of the National Academic of Science, 107*(41), 17491–17496.

# Chapter 12
# Nonlinear Filtering Based on Model Prediction

**Shesheng Gao, Yan Zhao, Yongmin Zhong, Aleksandar Subic, and Rezar Jazar**

**Abstract** Nonlinear filtering is of great importance in many applied areas. As a typical nonlinear filtering algorithm, the unscented Kalman filter (UKF) has the merits such as simplicity in realization, high filtering precision, and good convergence. However, its filtering performance is very sensitive to system model error. To overcome this limitation, this paper presents a new UKF for state estimation in nonlinear systems. This algorithm integrates model prediction into the process of the traditional UKF to improve the filtering robustness. This algorithm incorporates system driving noise in system state by increasing the state space dimension to expand the input of system state information to the system. The system model error is constructed by model prediction to rectify the system estimation from the traditional UKF. Simulation and experimental analyses have been conducted, showing that the proposed filtering algorithm is superior to the existing nonlinear filtering algorithms such as the EKF and traditional UKF in terms of accuracy.

**Keywords** Nonlinear filtering • Unscented Kalman filter • Model-error prediction • Dimension increase

## 12.1 Introduction

The problem of nonlinear filtering has its origins in the areas of tracking and signal processing. Nevertheless, the underlying setting is extremely general and is ubiquitous in many applied areas such as integrated navigation system, geodetic

S. Gao • Y. Zhao
School of Automatics, Northwestern Polytechnical University, Xi'an 710072, China

Y. Zhong (✉) • R. Jazar
School of Aerospace, Mechanical and Manufacturing Engineering, RMIT University, Melbourne, VIC, Australia
e-mail: yongmin.zhong@rmit.edu.au

A. Subic
Swinburne Research & Development, Swinburne University of Technology, Hawthorn, VIC, Australia
e-mail: aleksandar.subic@rmit.edu.au



R.N. Jazar, L. Dai (eds.), *Nonlinear Approaches in Engineering Applications*,
DOI 10.1007/978-3-319-27055-5_12

positioning, and automatic control, where random processes are used to model complex dynamical phenomenon. In essence, nonlinear filtering is to estimate the state of a nonlinear and non-Gaussian stochastic system from noisy measurement data. However, this process inevitably involves a gross error when there is a deviation between the theoretical model and the practical model, leading to the biased or even divergent filtering solution (Einicke 2012). The unscented Kalman filter (UKF), which is a typical nonlinear filter, is such a particular case. It suffers from the problem that the filtering performance is very sensitive to system model error.

This paper presents a new UKF algorithm for nonlinear filtering. The algorithm incorporates model prediction in the process of UKF. It augments system driving noise to system state to increase the input information of system state. The proposed algorithm overcomes the limitation of the traditional UKF that the filtering performance is sensitive to system model error. It can effectively resist the disturbance of model error on system state estimation and enhance the robustness of the filtering process. Experiments and comparison analysis with the existing methods have been conducted to comprehensively evaluate the performance of the proposed filtering algorithm.

## 12.2 Related Work

The traditional Kalman filter uses statistical characteristics of the system model to determine estimates recursively. However, the optimality of this filter heavily depends on linearity. A significant amount of research efforts have been dedicated to nonlinear filtering. The EKF is an approach to recursive nonlinear estimation. It is obtained by first-order linearization of nonlinear models such that the traditional KF can be applied. However, the EKF has the well-known drawbacks such as the large linearization error, derivation of Jacobian matrices, and poor robustness against the system model uncertainty (Boutayeb and Aubry 1999; Julier and Uhlmann 1997; Julier et al. 1995; Doucet et al. 1998).

The particle filter is an optimal recursive Bayesian filtering method based on Monte Carlo simulation, aiming to produce a sample of independent random variables distributed according to the conditional probability distribution (Doucet et al. 2001; Van der Merwe et al. 2000; Oppenheim et al. 2008). This method provides a complete representation for the posterior distribution of the state, enabling the easy calculation of any statistical estimate such as the mean, modes, kurtosis, and variance. Consequently, it can deal with nonlinear system models and non-Gaussian noise. However, the phenomenon of particle degeneracy may occur in the approximation process, and the accuracy largely depends on the choice of the importance sampling density and resampling scheme. The approximation process may diverge if a dynamic system has very small noise or the measurement noise has very small variance (Oppenheim et al. 2008; Wang 2006). The computational complexity also depends on the number of samples in the process of state estimation.

The UKF is a derivative-free filter. It combines the concept of unscented transform with the linear update structure of the Kalman filtering. This method is based on statistical approximations of system equations without requiring the calculation of the Jacobian matrix (Julier and Uhlmann 2004). It can capture the posterior mean and covariance accurately to the second order of Taylor series expansion for any nonlinearity, leading to a superior performance with an equivalent computational complexity (Pan et al. 2005). Further, the use of unscented transformation can avoid random sampling error caused by any sampling method such as the Monte Carlo simulation, thus dramatically reducing the number of points without the trade-off of the filtering accuracy (Julier 1998). However, similar to the EKF, the UKF is sensitive to system model uncertainty. In the presence of model uncertainties such that the input data cannot reflect the real model, the filtering solution of the UKF will be deteriorated or even divergent (Yang and Li 2011; Jwo and Lai 2009; Gao et al. 2015).

In the recent years, the robust adaptive filter has been used in integrated navigation system to control the influences due to both system state models and measurement singularities. Yang et al. reported a robust adaptive filter by combining the robust maximum likelihood estimation with the adaptive filtering process to adaptively adjust the weight matrix of predicted parameters according to the difference between system measurement and model information (Yang and Gao 2006; Yang et al. 2001). This filter can be adaptively converted into the classical Kalman filter, adaptive Kalman filter, and Sage filter by modifying the weight matrix and adaptive factor. However, the filtering method is difficult to estimate state parameters at the epochs with insufficient measurement information. They also developed a robust adaptive filter with multiple adaptive factors (Yang and Cui 2008). Although the robustness is improved by using multiple adaptive factors, the use of multiple adaptive factors causes an extra computational load, as it requires the number of measurements at all calculation epochs be larger than the number of state components. Ding et al. reported a process noise scaling method to monitors the filtering process using covariance matching for improving the robustness of adaptive filtering (Ding et al. 2007). However, this method cannot optimally distribute noises to each individual source. Gao et al. also reported a robust adaptive filter for SINS/SAR (Strap-down Inertial Navigation System/Synthetic Aperture Radar) integrated navigation system (Gao et al. 2011). However, this method may not guarantee that the robust adaptive factor constructed from predicted residuals is optimal for achieving the best filtering result. In general, the iterative process involved in the robust adaptive filtering requires reliable state estimate to calculate the covariance matrix of measurement noise. If the state estimate is disturbed by kinematic model error and measurements, it is difficult to obtain the reliable equivalent covariance matrix for describing the characteristics of measurement noise (Sayed and Rupp 2010).

Model predictive is a method to determine the minimum model error during the estimation process, where the model error is not limited to Gaussian noise characteristics (Goldenstein 2004). Crassidis et al. reported a nonlinear model prediction algorithm (Crassidis and Markley 1997). This algorithm estimates the model error

by comparing predictions with measurements. Subsequently, it remedies the filter conditions according to the model error to obtain the estimation of system state. Due to the real-time estimation and correction of system error model, the MPF algorithm is capable of handling large and dynamic errors of a nonlinear system model, leading to a continuous model to avoid discrete jumps in system state estimate. However, this algorithm requires the calculation of partial derivatives, leading to an increased computational load. Fang et al. reported a method to estimate the model error for an INS/GPS integrated navigation system by combining the EKF with model prediction (Fang and Gong 2010). Huang et al. studied the closed-loop robust stability of a nonlinear model predictive filter coupled with an extended Kalman filter, showing that the estimation error dynamics of the EKF are input-to-state stable in the presence of nonvanishing perturbations (Huang et al. 2013). However, due to the use of the EKF, these two methods suffer from the problems of the EKF in handling nonlinear systems, such as the requirement of calculating the Jacobian matrix and the low accuracy due to the use of linear approximation.

## 12.3 Model Prediction Based UKF

### *12.3.1 Prediction of System Model Error*

Consider the state and measurement equations of a nonlinear system

$$\dot{\widehat{\mathbf{x}}}(t) = f\left(\widehat{\mathbf{x}}(t)\right) + G\left(\widehat{\mathbf{x}}(t)\right)\underline{\mathbf{d}}(t) \tag{12.1}$$

$$\widehat{\mathbf{y}}(t) = \underline{h}\left(\widehat{\mathbf{x}}(t)\right) \tag{12.2}$$

where both $f(\cdot)$ and $h(\cdot)$ are continuous and differentiable nonlinear functions, and $\underline{\mathbf{d}}(t) \in R^r$ represents the model-error vector. $\mathbf{x}(t) \in R^n$ is the state vector, $\widehat{\mathbf{x}}(t)$ the state estimate vector of $\mathbf{x}(t)$, $G\left(\widehat{\mathbf{x}}(t)\right) \in R^{n\times r}$ the model-error distribution matrix, $\mathbf{y}(t) \in R^m$ the measurement vector, and $\hat{\mathbf{y}}(t)$ the estimated output vector.

Equation (12.2) can be discretized as

$$\tilde{\mathbf{z}}(t_k) = h\left(\widehat{\mathbf{x}}(t_k)\right) + \mathbf{v}_k \tag{12.3}$$

where $\tilde{\mathbf{z}}(t_k)$ is the discrete expression of $\hat{\mathbf{y}}(t)$, and $\mathbf{v}_k \in R^{m\times 1}$ is the measurement noise vector of the system.

The cost function is constructed as by summing the weighted sum square of the measurement-minus-estimate residuals and the weighted sum square of the model correction term

$$J\left(\underline{\mathbf{d}}(t)\right)=\frac{1}{2}\left\{\tilde{\mathbf{z}}\left(t+\Delta t\right)-\widehat{\mathbf{y}}\left(t+\Delta t\right)\right\}^{\mathrm{T}}\mathbf{R}^{-1}\left\{\tilde{\mathbf{z}}\left(t+\Delta t\right)-\widehat{\mathbf{y}}\left(t+\Delta t\right)\right\}+\frac{1}{2}\underline{\mathbf{d}}^{\mathrm{T}}(t)\mathbf{W}\underline{\mathbf{d}}(t) \tag{12.4}$$

where $\mathbf{W}$ is the system weight matrix and $\mathbf{R}$ is measurement covariance matrix of $\mathbf{v}_k$.

By minimization of (12.4), we can obtain the system model error

$$\underline{\mathbf{d}}(t)=-\left\{\left[\mathbf{\Lambda}\left(\Delta t\right)\mathbf{S}\left(\widehat{\mathbf{x}}\right)\right]^{\mathrm{T}}\mathbf{R}^{-1}\left[\mathbf{\Lambda}\left(\Delta t\right)\mathbf{S}\left(\widehat{\mathbf{x}}\right)\right]+\mathbf{W}\right\}^{-1}\left[\mathbf{\Lambda}\left(\Delta t\right)\mathbf{S}\left(\widehat{\mathbf{x}}\right)\right]^{\mathrm{T}}\mathbf{R}^{-1}\left[\mathbf{l}\left(\widehat{\mathbf{x}},\Delta t\right)-\tilde{\mathbf{z}}\left(t+\Delta t\right)+\widehat{\mathbf{y}}(t)\right] \tag{12.5}$$

where the $i$th element of $\mathbf{l}\left(\widehat{\mathbf{x}}(t),\Delta t\right)$ is defined as

$$l_i\left(\widehat{\mathbf{x}}(t),\Delta t\right)=\sum_{k=1}^{p_i}\frac{\Delta t^k}{k!}L_f^k\left(h_i\right) \tag{12.6}$$

where $L_f^k(h_i)$ is the $k$th-order Lie derivative.

$\mathbf{\Lambda}\left(\Delta t\right)\in\mathbf{R}^{m\times m}$ is a diagonal matrix with elements defined by

$$\lambda_{ii}=\frac{\Delta t^{p_i}}{p_i!},\quad i=1,2,\cdots,m \tag{12.7}$$

where $p_i\ \ (i=1,2,\cdots,m)$ is the lowest order of the derivative of $h\left(\widehat{\mathbf{x}}(t)\right)$.

$\mathbf{S}\left(\widehat{\mathbf{x}}\right)\in\mathbf{R}^{m\times r}$ is a matrix with each row defined by

$$\mathbf{s}_i=\left[L_{g_1}\left(L_f^{p_i-1}\left(h_i\right)\right),\cdots,L_{g_r}\left(L_f^{p_i-1}\left(h_i\right)\right)\right],\quad i=1,2,\cdots,m \tag{12.8}$$

Letting $\mathbf{A}=\mathbf{\Lambda}\left(\Delta t\right)\mathbf{S}\left(\widehat{\mathbf{x}}\right)$, (12.5) can be simplified as

$$\underline{\mathbf{d}}(t)=-\left\{\mathbf{A}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{A}+\mathbf{W}\right\}^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{R}^{-1}\left[\mathbf{l}\left(\widehat{\mathbf{x}},\Delta t\right)-\tilde{\mathbf{z}}\left(t+\Delta t\right)+\widehat{\mathbf{y}}(t)\right] \tag{12.9}$$

### *12.3.2 Expanded System State*

Consider the discrete state equation and measurement equation of a navigation system

$$\mathbf{X}_k=f\left(\mathbf{X}_{k-1}\right)+\mathbf{\Gamma}_{k-1}\mathbf{w}_{k-1} \tag{12.10}$$

$$\mathbf{Z}_k=h\left(\mathbf{X}_k\right)+\mathbf{v}_k \tag{12.11}$$

where $\mathbf{X}_k$ is the estimated state vector, $f(\cdot)$ and $h(\cdot)$ are nonlinear functions of the system state, $\mathbf{\Gamma}_{k-1}$ is the noise driving matrix, $\mathbf{w}_{k-1}$ is the state noise vector of the system state model, $\mathbf{Z}_k$ is the measurement vector, and $\mathbf{v}_k$ is the measurement noise vector.

Let us first consider the expansion of the space dimension of the system state vector. In order to take into account the effect of system driving noise on the system state model, the noise is added in the system state. Thus, the expanded system state vector and covariance matrix can be expressed as

$$\mathbf{X}_k^a = \left[\mathbf{X}_k^{\mathrm{T}} \;\; \underline{\mathbf{d}}_k^{\mathrm{T}} \;\; \mathbf{v}_k^{\mathrm{T}}\right]^{\mathrm{T}} \tag{12.12}$$

$$\mathbf{P}_k^a = \begin{bmatrix} \mathbf{P}_k^x & & \\ & \mathbf{Q}_k^d & \\ & & \mathbf{R}_k^v \end{bmatrix} \tag{12.13}$$

where the dimension of $\mathbf{X}_k^a$ is $L = n + q + p$, $q$ is the dimension of $\mathbf{Q}$, and $p$ is the dimension of $\mathbf{R}$.

As mentioned previously, the UKF is very sensitive to system model error, thus unable to deal with the problem of error estimation. In order to overcome this limitation, this paper adopts model prediction to determine a reliable model error $\underline{\mathbf{d}}(t)$ to improve the robustness of the UKF. Subsequently, by replacing $\mathbf{w}_{k-1}$ in (12.10) with $\underline{\mathbf{d}}_{k-1}$ and substituting the estimated value of the expanded state into the state and measurement equations, we can get

$$\mathbf{X}_k^a = f\left(\mathbf{X}_{k-1}^a\right) + \mathbf{\Gamma}_{k-1}\underline{\mathbf{d}}_{k-1} \tag{12.14}$$

$$\mathbf{Z}_k^a = h\left(\mathbf{X}_k^a\right) + \mathbf{v}_k \tag{12.15}$$

where $\underline{\mathbf{d}}_k$ and $\mathbf{v}_k$ can be assumed as a white noise process. This is because if the random error is not a white noise process, it can be transformed into a white noise process by decreasing the sampling rate.

Therefore, the following conditions hold for $\underline{\mathbf{d}}_k$ and $\mathbf{v}_k$

$$\begin{cases} \mathrm{E}\left[\underline{d}_k\right] = 0, & \mathrm{COV}\left[\underline{d}_k, \underline{d}_j^{\mathrm{T}}\right] = \mathbf{Q}_k\delta_{kj} \\ \mathrm{E}\left[v_k\right] = 0, & \mathrm{COV}\left[v_k, v_j^{\mathrm{T}}\right] = \mathbf{R}_k\delta_{kj} \\ & \mathrm{COV}\left[\underline{d}_k, v_j^{\mathrm{T}}\right] = 0 \end{cases} \tag{12.16}$$

### *12.3.3 Filtering Algorithm*

The proposed model prediction based UKF includes the following steps:

(1) Calculate the sigma point $\boldsymbol{\chi}_k$ of the state vector

$$\boldsymbol{\chi}_k = \begin{bmatrix} \overline{\mathbf{x}}_{k-1}^a \\ \overline{\mathbf{x}}_{k-1}^a + \sqrt{(L+I)\cdot \overline{\mathbf{P}}_{x,k-1}^a} \\ \overline{\mathbf{x}}_{k-1}^a - \sqrt{(L+I)\cdot \overline{\mathbf{P}}_{x,k-1}^a} \end{bmatrix} \tag{12.17}$$

(2) Calculate the system model error $\underline{\mathbf{d}}_{k-1}$ from (12.9)

$$\underline{\mathbf{d}}(t) = -\left\{\mathbf{A}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{A} + \mathbf{W}\right\}^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{R}^{-1}\left[\mathbf{l}\left(\widehat{\mathbf{x}}, \Delta t\right) - \tilde{\mathbf{z}}\left(t + \Delta t\right) + \widehat{\mathbf{y}}(t)\right] \tag{12.18}$$

(3) The time update equations are

$$\boldsymbol{\chi}_{k,k-1}^a = f\left(\boldsymbol{\chi}_{k-1}^x\right) + \boldsymbol{\chi}_{k-1}^d \tag{12.19}$$

$$\widehat{\mathbf{x}}_{k,k-1}^a = \sum_{i=0}^{2L} W_i^{(m)} \cdot \boldsymbol{\chi}_{i,k,k-1}^a \tag{12.20}$$

$$\mathbf{P}_{k,k-1}^a = \sum_{i=0}^{2L} W_i^{(c)} \cdot \left(\boldsymbol{\chi}_{i,k,k-1}^a - \widehat{\mathbf{x}}_{k,k-1}\right) \cdot \left(\boldsymbol{\chi}_{i,k,k-1}^a - \widehat{\mathbf{x}}_{k,k-1}\right)^{\mathrm{T}} \tag{12.21}$$

(4) The measurement update equations are

$$\mathbf{Z}_k^a = h\left(\boldsymbol{\chi}_{k,k-1}^{(z)}\right) + \boldsymbol{\chi}_{k,k-1}^{(v)} \tag{12.22}$$

$$\widehat{\mathbf{Z}}_k = \sum_{i=0}^{2L} W_i^{(m)} \cdot \mathbf{Z}_{i,k}^a \tag{12.23}$$

$$\mathbf{P}_{x_k z_k} = \sum_{i=0}^{2L} W_i^{(c)} \cdot \left(\boldsymbol{\chi}_{i,k,k-1}^a - \widehat{\mathbf{x}}_{k,k-1}\right) \cdot \left(\mathbf{Z}_{i,k}^a - \widehat{\mathbf{Z}}_k\right)^{\mathrm{T}} \tag{12.24}$$

$$\mathbf{P}_{z_k} = \sum_{i=0}^{2L} W_i^{(c)} \cdot \left(\mathbf{Z}_{i,k}^a - \widehat{\mathbf{Z}}_k\right) \cdot \left(\mathbf{Z}_{i,k}^a - \widehat{\mathbf{Z}}_k\right)^{\mathrm{T}} \tag{12.25}$$

$$\overline{\mathbf{P}}_{z_k} = \gamma_{ij}\mathbf{P}_{z_k} \tag{12.26}$$

$$\mathbf{K}_k = \mathbf{P}_{x_k z_k}\overline{\mathbf{P}}_{z_k}^{-1} \tag{12.27}$$

$$\widehat{\mathbf{x}}_k^a = \widehat{\mathbf{x}}_{k,k-1}^a + \mathbf{K}_k\left(\mathbf{Z}_k^a - \widehat{\mathbf{Z}}_k\right) \tag{12.28}$$

$$\mathbf{P}_k^a = \mathbf{P}_{k,k-1}^a - \mathbf{K}_k\overline{\mathbf{P}}_{z_k}\mathbf{K}_k^{\mathrm{T}} \tag{12.29}$$

In (12.24), $W_i^{(m)}$ and $W_i^{(c)}$ are the weight matrices, which are represented as

$$\begin{cases} W_0^{(m)} = \frac{I}{(L+I)} \\ W_0^{(c)} = \frac{I}{(L+I)} + \left(1 - \alpha^2 - \beta\right) \\ W_i^{(m)} = W_i^{(c)} = \frac{0.5}{(L+I)} \qquad i = 1, 2, \cdots, 2L \end{cases} \tag{12.30}$$

where $L + I = 3$. When $I < 0$, the semi-regularity of (12.30) cannot be guaranteed. In this case, a scale correction (Gao et al. 2011) is applied to the obtained sigma point set to make (12.30) semi-regular. In the most cases, $10^{-4} \leq \alpha \leq 1$, and $\beta = 2$ is optimal for Gaussian distribution (Julier et al. 1995).

In (12.26), $\gamma_{ii}$ represents the main diagonal element of prior weight matrix $\gamma$. If the measurement information contains abnormality, the corresponding weighted matrix is adjusted by the following IGG III weight function to improve the filtering performance

$$\gamma_{ii} = \begin{cases} 1 & \left|\tilde{V}_i\right| \leq k_0 \\ \frac{k_0}{\left|\tilde{V}_i\right|}\left(\frac{k_1 - \tilde{V}_i}{k_1 - k_0}\right)^2 & k_0 < \left|\tilde{V}_i\right| \leq k_1 \\ 0 & \left|\tilde{V}_i\right| > k_1 \end{cases} \tag{12.31}$$

where $\tilde{V}_i$ is standardized residuals, $k_0$ and $k_1$ are constants, $k_0 = 1.0 \sim 1.5$ and $k_1 = 3.5 \sim 8.0$.

$$\gamma_{ij} = \sqrt{\gamma_{ii}}\sqrt{\gamma_{jj}} \tag{12.32}$$

## 12.4 Performance Evaluation and Discussions

Experiments have been conducted to comprehensively evaluate and analyze the performance of the proposed filtering algorithm (named the MP-UKF). The comparison analysis with the existing filtering algorithms such as the EKF and UKF is also discussed in this section.

### *12.4.1 Simulations and Analysis*

The univariate non-stationary growth model (UNGM) was adopted to evaluate the performance of the proposed algorithm. The state and measurement equations of the UNGM are described as

$$x(t) = 0.5x(t-1) + \frac{2.5x(t-1)}{1+[x(t-1)]^2} + 8\cos[1.2(t-1)] + w(t) \tag{12.33}$$

$$Z(t) = x(t)/20 + v(t) \tag{12.34}$$

where both $w(t)$ and $v(t)$ are the zero-mean Gaussian noises, and their intensity variances $Q$ and $R$ are

$$Q = \mathrm{cov}\left[w(t), w(\tau)^{\mathrm{T}}\right] = 10 \tag{12.35}$$

$$R = \mathrm{cov}\left[v(t), v(\tau)^{\mathrm{T}}\right] = 1 \tag{12.36}$$

The initial state and its variance are assumed as

$$x(0) = 0.1 \text{ and } P(0) = 2 \tag{12.37}$$

By expanding the state space dimension, we get

$$x^a(0) = \begin{bmatrix} 0.1 & d & 0 \end{bmatrix}^{\mathrm{T}} \text{and } \mathbf{P}^a(0) = \begin{bmatrix} 2 & & \\ & 10 & \\ & & 1 \end{bmatrix}^{\mathrm{T}} \tag{12.38}$$

where $d$ is calculated by (12.9).

It can be seen from Fig. 12.1 that the estimated values of the EKF are within $(-20, 10)$, leading to large deviations from the real values. The estimated values of the UKF are within $(-15, 5)$. Although the UKF improves the estimation

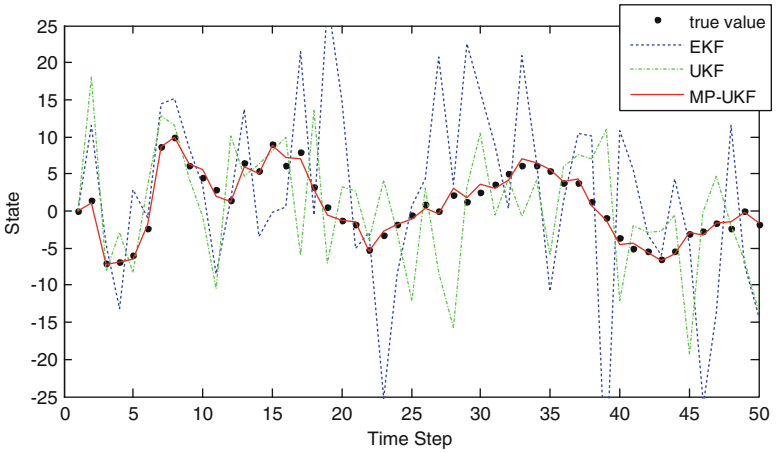



**Fig. 12.1** Estimation of the three filtering methods

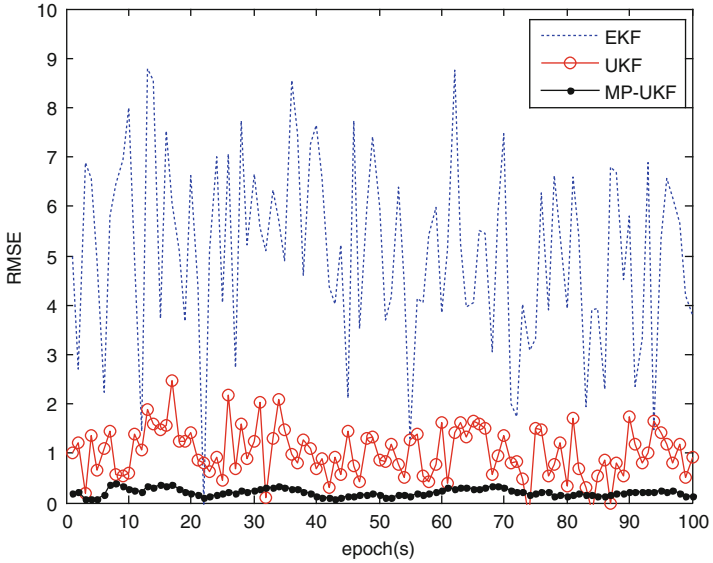



**Fig. 12.2** Root mean square errors of the three algorithms for the estimation

performance of the EKF, there still are pronounced deviations between the estimated and real values. In contrast, the proposed MP-UKF accurately captures the true values, leading to the estimate error within $(-5, 5)$. Figure 12.2 shows the root mean square errors (RMSE) of the three filtering algorithms. The RMSE is within (1, 8) for the EKF, (0.5, 2.5) for the UKF, and (0, 0.5) for the RMP-AUKF. Therefore, it is evident that the MP-UKF outperforms the other two.

### *12.4.2 Experiments and Analysis*

Practical experiments were conducted to evaluate the performance of the proposed filtering algorithm for a SINS/CNS/SAR (Strap-down Inertial Navigation System/Celestial Navigation System/Synthetic Aperture Radar) integrated navigation system. The navigation coordinate is E-N-U (East-North-Up) geography coordinate system. The SINS/CNS/SAR integrated navigation system is

$$\mathbf{X}_k = f\left(\mathbf{X}_{k-1}\right) + \mathbf{\Gamma}_{k-1}\underline{\mathbf{d}}_{k-1} \tag{12.39}$$

$$\mathbf{Z}_k = \mathbf{H}_k\mathbf{X}_k + \mathbf{v}_k \tag{12.40}$$

where $f(\cdot)$ is the nonlinear function of the state vector $\mathbf{X}$ and is described as

$$f(\mathbf{X}(t)) = \begin{bmatrix} \boldsymbol{C}_{nb}^{-1}\left[\left(I-\boldsymbol{C}_{n}^{n'}\right)\widehat{\omega}_{in}^{n}+\boldsymbol{C}_{n}^{n'}\delta\omega_{in}^{n}-\boldsymbol{C}_{b}^{n'}\delta\omega_{ib}^{b}\right] \\ \left[I-\left(\boldsymbol{C}_{n}^{n'}\right)^{\mathrm{T}}\right]\boldsymbol{C}_{b}^{n'}\widehat{\boldsymbol{f}}_{sf}^{b}+\left(\boldsymbol{C}_{n}^{n'}\right)^{\mathrm{T}}\boldsymbol{C}_{b}^{n'}\delta\boldsymbol{f}_{sf}^{b} \\ -\left(2\delta\omega_{ie}^{n}+\delta\omega_{en}^{n}\right)\times\widehat{\boldsymbol{V}}-\left(2\widehat{\omega}_{ie}^{n}+\widehat{\omega}_{en}^{n}\right) \\ \times\delta\boldsymbol{V}+\left(2\omega_{ie}^{n}+\omega_{en}^{n}\right)\times\delta\boldsymbol{V}+\delta\boldsymbol{g} \\ \frac{\widehat{V}_N}{\widehat{R}_M+h}-\frac{\left(\widehat{V}_N-\delta V_N\right)}{\left(\widehat{R}_M-\delta R_M\right)+\left(\widehat{h}-\delta h\right)} \\ \frac{\widehat{V}_E\sec L}{\widehat{R}_N+h}-\frac{\left(\widehat{V}_E-\delta V_E\right)\sec\left(\widehat{L}-\delta L\right)}{\left(\widehat{R}_N-\delta R_N\right)+\left(\widehat{h}-\delta h\right)} \\ \delta V_U \end{bmatrix} \tag{12.41}$$

The state variables are

$$\mathbf{X} = \left[\phi_E \;\; \phi_N \;\; \phi_U \;\; \delta v_E \;\; \delta v_N \;\; \delta v_U \;\; \delta L \;\; \delta\lambda \;\; \delta h \;\; \varepsilon_{bx} \;\; \varepsilon_{by} \;\; \varepsilon_{bz} \;\; \nabla_x \;\; \nabla_y \;\; \nabla_z\right]^{\mathrm{T}} \tag{12.42}$$

where $(\phi_E, \phi_N, \phi_U)$ is the attitude error, $(\delta v_E, \delta v_N, \delta v_U)$ is the velocity error, $(\delta L, \delta\lambda, \delta h)$ the position error, $(\varepsilon_{bx}, \varepsilon_{by}, \varepsilon_{bz})$ the gyro constant drift, and $(\nabla_X, \nabla_Y, \nabla_Z)$ the accelerometer zero-bias.

The SINS and SAR are combined as sub-filter 1, and its measurement is

$$\mathbf{Z}_1 = \begin{bmatrix} \phi_E - \phi_{\mathrm{SE}} \\ (L - L_S) R_M \\ (\lambda - \lambda_S) R_N \cos L \end{bmatrix} \tag{12.43}$$

The SINS and CNS are combined as sub-filter 2, and its measurement is

$$\mathbf{Z}_2 = \begin{bmatrix} \phi_E - \phi_{\mathrm{CE}} \\ \phi_N - \phi_{\mathrm{CN}} \\ \phi_U - \phi_{\mathrm{CU}} \end{bmatrix} \tag{12.44}$$

The measurement of the SINS/CNS/SAR integrated navigation system is described as

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{bmatrix} \tag{12.45}$$

The measurement matrix of the SINS/CNS/SAR integrated navigation system is

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \end{bmatrix}_{6\times 10} \tag{12.46}$$

where

$$\mathbf{H}_1 = \begin{bmatrix} 0_{1\times 2}\ 1\ 0_{1\times 3} & 0 & 0 & 0_{1\times 7} \\ 0_{1\times 2}\ 0\ 0_{1\times 3} & R_M & 0 & 0_{1\times 7} \\ 0_{1\times 2}\ 0\ 0_{1\times 3} & 0 & R_N \cos L & 0_{1\times 7} \end{bmatrix} \tag{12.47}$$

$$\mathbf{H}_2 = \begin{bmatrix} -\cos\psi_I & \sin\psi_I & 0\ 0_{1\times 12} \\ \frac{\sin\psi_I}{\cos\theta_I} & -\frac{(\cos\psi_I - 2\sin\theta_I \sin\gamma_I \cos\gamma_I \sin\psi_I)}{\cos\theta_I} & 0\ 0_{1\times 12} \\ -\sin\psi_I \tan\theta_I & \cos\psi_I \tan\theta_I & -1\ 0_{1\times 12} \end{bmatrix} \tag{12.48}$$

Suppose the Earth's rotation angular velocity is $\omega_{\text{ie}} = 15°/\text{h}$, the radius of the Earth is $R_e = 6378$ km, and the acceleration of gravity is selected as $\text{g} = 9.780\ \text{m/s}^2$. The initial position of the aircraft was at East longitude 109°, North latitude 34°, and altitude 1000 m. The initial attitude angle error was $\phi_E = \phi_N = \phi_U = 0.5'$, the velocity error $\delta V_E = \delta V_N = \delta V_U = 0.1$ m/s, and the position error $\delta\lambda = 1''$, $\delta L = 1''$, $\delta h = 10$ m. The gyro's random drift and walk were 0.01 °/h and $0.001°/\sqrt{\text{h}}$. The accelerometer's zero offset and random walk were $10^{-4}$g and $10^{-5}\text{g}/\sqrt{\text{s}}$. The horizontal positioning accuracy of SAR was 25 m, the heading angle of mean square error by SAR was 0.1°, and the measurement accuracy of CCD (Charge-Coupled Device) star sensor was 20″. The sampling cycles of SINS, SAR, and CNS were 0.01, 1, and 0.5 s. The filtering period was 1 s. The measurement noise was $\boldsymbol{v} = [0.5\ \ 0.2\ \ 0.2\ \ 0.5\ \ 0.5\ \ 0.5]^{\text{T}}$.

As shown in Figs. 12.3, 12.4, and 12.5, the EKF algorithm has the largest attitude angle error with the heading error (–1.5′, 1.25′), roll angle error (−1′, 1.5′), and pitch angle error (−1.25′, 1.5′). The portions of the curve, which are beyond (−1′, 1′),

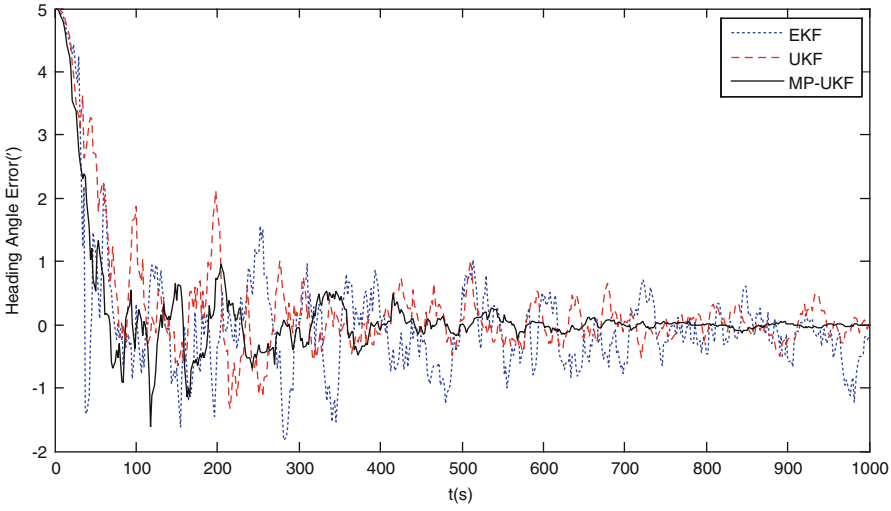



**Fig. 12.3** Heading angle error of the SINS/CNS/SAR integrated navigation system

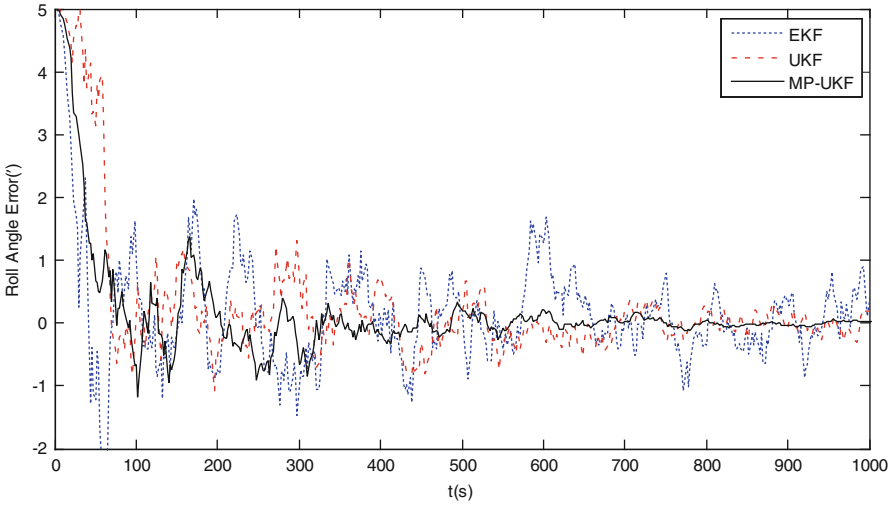



**Fig. 12.4** Roll angle error of the SINS/CNS/SAR integrated navigation system

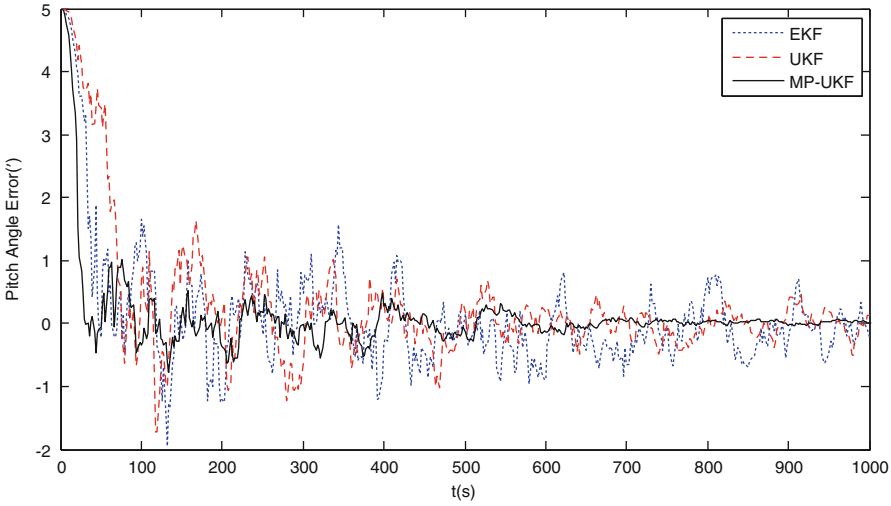



**Fig. 12.5** Pitch angle error of the SINS/CNS/SAR integrated navigation system

involve a significant change in amplitude. The attitude angle error is improved by the UKF algorithm with all the heading, roll, and pitch angle errors within $(-1', 1')$. In contrast, the attitude angle error by the proposed filtering algorithm is reduced to $(-0.3', 0.4')$.

Figures 12.6 and 12.7 show the similar trend as Figs. 12.3, 12.4, and 12.5. With the EKF, the velocity error is largest, which is within (−1.5 m/s, 1.5 m/s). With the UKF, both the east and north velocity errors are within (−1 m/s, 0.9 m/s). However,

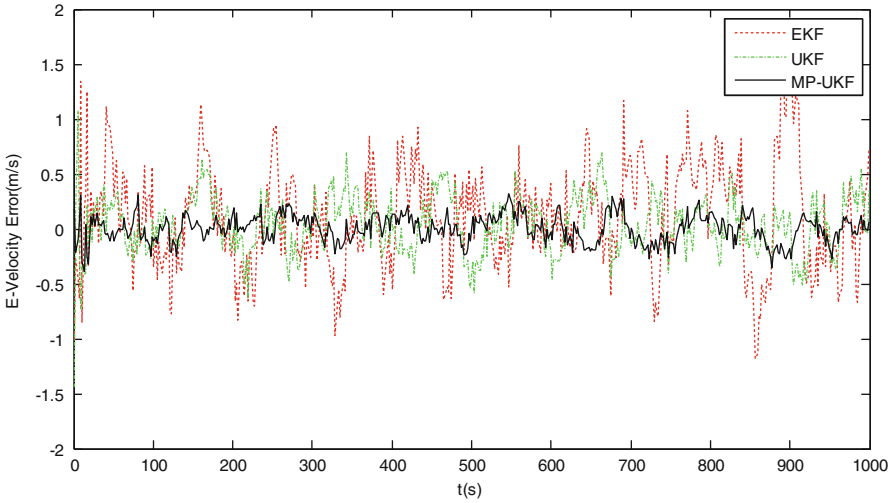



**Fig. 12.6** E-velocity error of the SINS/CNS/SAR integrated navigation system

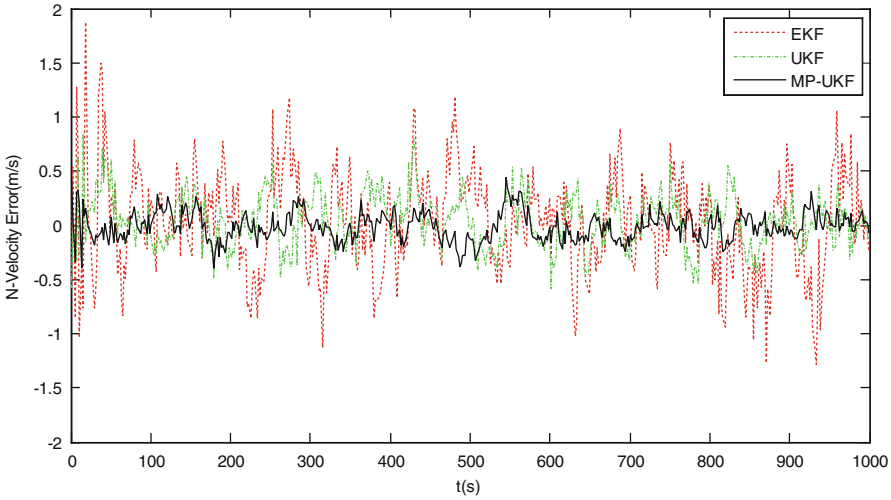



**Fig. 12.7** N-velocity error of the SINS/CNS/SAR integrated navigation system

the proposed filter algorithm reduces the velocity error to (−0.2 m/s, 0.3 m/s). As shown in Figs. 12.8 and 12.9, the velocity errors in latitude and longitude for the EKF are gradually reduced by the UKF and proposed MP-UKF. The estimate errors of the three filters are within $(-0.9'', 0.8'')$, $(-0.5'', 0.5'')$, and $(-0.2'', 0.2'')$, respectively. Table 12.1 shows the detailed comparison for the position errors of the three filters.

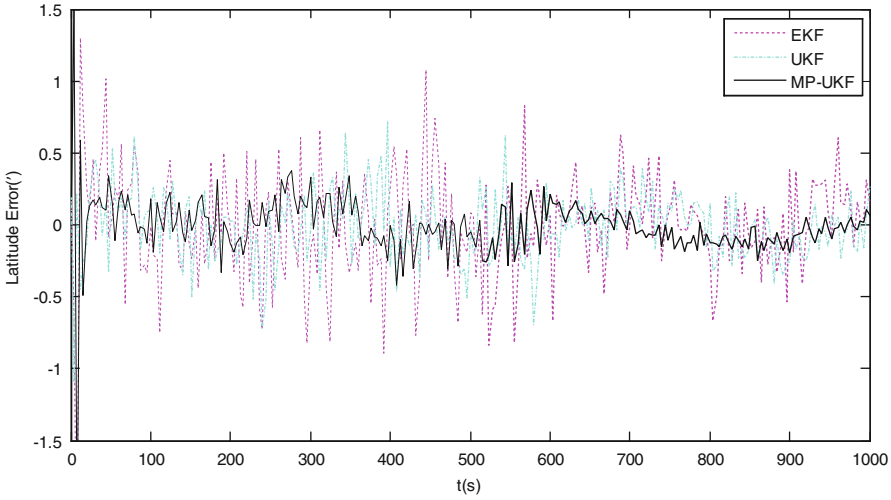



**Fig. 12.8** Latitude error of the SINS/CNS/SAR integrated navigation system

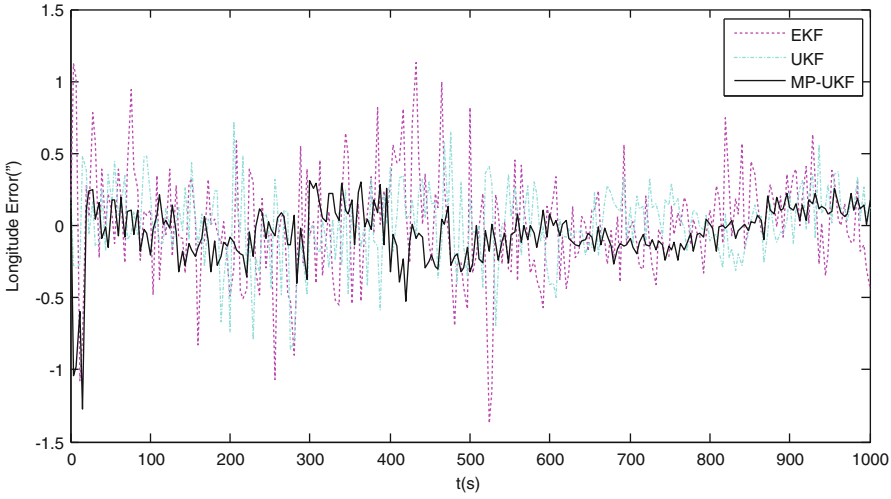



**Fig. 12.9** Longitude error of the SINS/CNS/SAR integrated navigation system

**Table 12.1** Position error statistics of the SINS/CNS/SAR integrated navigation system

| Filtering methods | Mean error | Standard deviation | Position error |
|---|---|---|---|
| EKF | 0.538 | 1.143 | $-0.9'' \sim 0.8''$ |
| UKF | 0.326 | 0.601 | $-0.5'' \sim 0.5''$ |
| MP-UKF | 0.107 | 0.207 | $-0.2'' \sim 0.2''$ |

From the above results and analysis, it is evident that the proposed algorithm can significantly reduce navigation error. The achieved navigation accuracy is much higher than those of the EKF and UKF.

## 12.5 Conclusions

This paper presents a nonlinear filtering algorithm for state estimation in nonlinear systems. By taking into account the effect of system driving noise on system state, this algorithm adds system driving noise in system state to expand the input of system state information to the system. Subsequently, it incorporates model prediction in the UKF filtering process to overcome the UKF limitation that the performance is sensitive to system model error. Simulation and experimental results demonstrate that the proposed filtering algorithm can significantly reduce navigation error. The achieved accuracy is much higher than those of the EKF and UKF.

Future work will focus on improving the computational performance of the proposed algorithm. Despite the improved filtering accuracy, the increase of the state space dimension in the proposed algorithm causes an extra computational load to the filtering process. It is expected to establish a strategy to simplify the computational process of the proposed algorithm by studying the relationship between the accuracy and computational time in terms of the degree of nonlinearity.

## References

Boutayeb, M., & Aubry, D. (1999). A strong tracking extended Kalman observer for nonlinear discrete-time systems. *IEEE Transactions on Automatic Control, 44*(8), 1550–1556.

Crassidis, J. L., & Markley, F. L. (1997). Predictive filtering for nonlinear systems. *Journal of Guidance Control and Dynamics*, *20*(3), 566–572.

Ding, W., Wang, J., & Rizos, C. (2007). Improving adaptive Kalman estimation in GPS/INS integration. *Journal of Navigation, 60*(3), 517–529.

Doucet, A. *On sequential simulation-based methods for Bayesian filtering*, Technical Report 310. (1998). Department of Engineering, Cambridge University.

Doucet, A., de Freitas, J. F. G., & Gordon, N. J. (2001). An introduction to sequential Monte Carlo methods, sequential Monte Carlo methods in practice. In A. Doucet, J. F. G. de Freitas & N. J. Gordon (Eds.). New York: Springer.

Einicke, G. A. (2012). *Smoothing, filtering and prediction - estimating the past, present and future*. Rijeka, Croatia: InTech.

Fang, J., & Gong, X. (2010). Predictive iterated Kalman filter for INS/GPS integration and its application to SAR motion compensation. *IEEE Transactions on Instrumentation and Measurement, 59*(4), 909–915.

Gao, S., Zhong, Y., & Li, W. (2011). Robust adaptive filtering method for SINS/SAR integrated navigation system. *Aerospace Science and Technology, 15*(6), 425–430.

Gao, S., Hu, G., & Zhong, Y. (2015). Windowing and random weighting-based adaptive unscented Kalman filter. *International Journal of Adaptive Control and Signal Processing, 29*(2), 201–223.

Goldenstein, S. (2004). A gentle introduction to predictive filters. *Revista de Informática Teórica e Aplicada (RITA)*, *XI* (1), 61–89.

Huang, R., Patwardhan, S. C., & Biegler, L. T. (2013). Robust stability of nonlinear model predictive control with extended Kalman filter and target setting. *International Journal of Robust and Nonlinear Control, 23*(11), 1240–1264.

Julier, S. (1998). A skewed approach to filtering. *Proceedings of SPIE, 3373*, 271–282.

Julier, S. J, & Uhlmann, J. K. (1997). A new extension of the Kalman filter to nonlinear systems. *Proceedings of SPIE*, *3068*, 182–193.

Julier, S. J., & Uhlmann, J. K. (2004). Unscented filtering and nonlinear estimation. *Proceedings of The IEEE, 92*(3), 401–422.

Julier, S. J., Uhlmann, J. K., & Durrant-Whyte, H. F. (1995). A new approach for filtering nonlinear systems. In: *The Proceedings of the American Control Conference* (pp. 1628–1632). Seattle, Washington.

Jwo, D. J., & Lai, S. Y. (2009). Navigation integration using the fuzzy strong tracking unscented Kalman filter. *Journal of Navigation, 62*(2), 303–322.

Oppenheim, G., Philippe, A., & de Rigal, J. (2008). The particle filters and their applications. *Chemometrics and Intelligent Laboratory Systems, 91*(1), 87–93.

Pan, Q., Yang, F., Ye, L., Liang, Y., & Cheng, Y. (2005). Survey of a kind of nonlinear filters - UKF. *Control and Decision, 20*(5), 481–489.

Sayed, A. H., & Rupp, M. (2010). Robust issues in adaptive filtering. In V. K. Madisetti (Ed.), *Digital signal processing fundamentals* (Vol. 20, pp. 1–20). Boca Raton, FL: Taylor & Francis.

Van der Merwe, R., Doucet, A., De Freitas, N., & Wan, E. (2000). The unscented particle filter. In: *Advances in neural information processing systems* (Vol. 4, pp. 351–357).

Wang, L. (2006). Fixed parameter estimation method using Gaussian particle filter. Lecture Notes in Computer Science, *4115/2006*, 121–129.

Yang, Y., & Cui, X. (2008). Adaptively robust filter with multi adaptive factors. *Survey Review, 40*(309), 260–270.

Yang, Y., & Gao, W. (2006). An optimal adaptive Kalman filter. *Journal of Geodesy, 80*(4), 177–183.

Yang, W. B., & Li, S. Y. (2011). Autonomous navigation filtering algorithm for spacecraft based on strong tracking UKF. *Systems Engineering and Electronics, 33*(11), 2485–2491.

Yang, Y., He, H.-B., & Xu, G. (2001). Adaptively robust filtering for kinematic geodetic positioning. *Journal of Geodesy, 75*(2–3), 109–116.

# Chapter 13
# Solutions for Path Planning Using Spline Parameterization

**M. Elbanhawi, M Simic, and R. Jazar**

**Abstract** This chapter presents a B-spline path synthesis approach for non-holonomic car-like vehicles. Recent findings in robotic research suggest that $C^2$ continuity is a realistic representation of robotic paths. A novel solution that guarantees $C^2$ parametric continuity for a cubic path, whilst satisfying kinodynamic constraints imposed on the vehicle, is proposed here. This chapter investigates path synthesis for front wheel steered vehicles. It is required to generate paths with continuous, bounded curvatures. Our approach is based on clamped B-spline curves. We leverage the curves' properties to resolve the considered problem. In order to ensure continuity, a single curve segment is utilized for the entire path. Path curvature is formulated with respect to the B-spline curve parameters. This approach mimics human operating and minimizes disturbances on passengers. We present here a method of curvature estimation that is more accurate and time efficient than previously published solutions. Two analytical bounding solutions are proposed. Both methods are shown to minimize path length and deviation, whilst maintaining curvature and parametric continuity.

## 13.1 Introduction

Human drivers display remarkable abilities when controlling vehicles in a highly reactive manner and with impeccable precision. Even more impressive is the implicit consideration of the vehicle and road parameters. Researchers have shown that humans use specific visual cues to identify road curvature and endeavor to match it. It is attempted to analyze to draw inspiration from the steering commands, used by operators of varying driving experience, to generate likewise natural paths. This approach can be combined with a path planning algorithm to generate paths with natural smooth trajectories for autonomous vehicles. This will circumvent the need to rely on computationally intensive planning algorithms that are based on forward model integration. Humans have been controlling the steering wheel for

M. Elbanhawi (✉) • M. Simic • R. Jazar
School of Aerospace, Mechanical and Manufacturing Engineering,
RMIT University, Melbourne, VIC, Australia
e-mail: mohamed.elbenhawi@rmit.edu.au; milan.simic@rmit.edu.au; reza.jazar@rmit.edu.au



R.N. Jazar, L. Dai (eds.), *Nonlinear Approaches in Engineering Applications*,
DOI 10.1007/978-3-319-27055-5_13

the past century. Currently self-driving is emerging as technology that promises to improve our lives greatly. Cars are underactuated systems designed to facilitate their control for operators. This simplification of actuation has adverse effects when attempting to automate such systems, leading to the appearance of nonholonomic constraints (Jazar 2008). Researchers have shown that drivers rely on certain visual landmarks to assess the path curvature prior to attempting to steer towards it (Land and Lee 1994; Land and Horwood 1995). The majority of studies, conducted on human steering, focus on modeling and predicating it, using control systems theories (Donges 1978; MacAdam 1981; Prokop 2001). The aim of this chapter is to employ an efficient spline parameterization method to synthesize paths that mimic human steering controls, in order to generate motions that feel natural and familiar to human passengers. It is not uncommon for robotics researchers to draw inspiration by observing natural behavior (Donghyun et al. 2014; Lentink 2014). We hope that the results, presented in this chapter, can be employed in robotics to efficiently plan spline-based paths, similar to Yang et al. (2014), in addition to mimicking human steering, improving passenger comfort and explicitly considering the limitations of the car.

Advances in sensing technology, computer vision, communications, and computational power have contributed towards the development of autonomous agents in a wide range of fields. Self-driving ground vehicles are used in military, urban transportation, industrial and agricultural applications. Unmanned Aerial Vehicles (UAVs) and Micro Aerial Vehicles (MAVs) are considered as a cost-effective, safe, and efficient choice for several military and civil applications. Robotic platforms are currently equipped with multiple sensors, which enable them to sense their surroundings and localize themselves in reference to their environment, goals, and obstacles. Path planning is a widely studied, fundamental task for mobile robots. Robot navigation mandates a strategy that steers it from its current location, through the environment, whilst avoiding obstacles towards its goal.

Classical planning algorithms such as A* algorithm (Hart et al. 1968), Voronoi diagrams (Canny 1985), visibility graphs (Asano et al. 1985), and cell decompositions (Brooks and Lozano-Perez 1985) produce piecewise linear paths. These paths consist of subsequent waypoints joined by straight lines. Potential field methods guide the robot towards its goal by applying attractive forces, toward the goal, and repulsive, away from obstacles (Khatib 1986). Potential field methods tend to produce oscillating paths in narrow passages (Koren and Borenstein 1991). Sampling-based motion planning algorithms, such as Rapidly exploring Random Trees (RRT) (LaValle 1998) and Probabilistic Roadmap Method (PRM) (Kavraki et al. 1996), rely on stochastic sampling to efficiently explore the search space. Resulting paths from randomization are suboptimal and require post-processing to improve their quality (Elbanhawi and Simic 2014c). Motion planning using state lattices is disadvantaged by discretization (Pivtoraiko et al. 2009; Pivtoraiko and Kelly 2011). Coarse discretization leads to loss of completeness whilst high fidelity subdivision increases the computational time of the planner, especially in highly dimensional scenarios. Homotopy class optimization of trajectories were proposed (Zucker et al. 2013). These methods do not discuss curvature continuity and the

performance is dependent on the optimization algorithm. Optimization methods are not immune from local minima and are not guaranteed to converge.

Agile robots, such as omnidirectional, differential-drive robots, and quadrotors, are capable of traversing piecewise linear paths. Such paths require stationary turns, at every waypoint, to change heading towards the subsequent waypoint. This approach is inefficient with regard to time, energy, and jerk considerations. The motion of some robots, such as car-like vehicles and fixed-wing UAVs, is highly constrained. Nonholonomic robots must be considered in the planning procedure, as they cannot follow piecewise linear paths. Minimum turning radius constraints impose further limitations on the path, which is often represented by maximum curvature restrictions.

Traditionally, Dubins paths or Reeds and Shepp's (Xuan-Nam et al. 1994; Reeds and Shepp 1990) are used in path smoothing for vehicles with minimum turning radius constraint in a two-dimensional space. Configurations are joined by sets of primitives consisting of circular arcs and straight lines. The amalgamation of circular arcs and lines results in discontinuities in curvature. Clothoids may appear to be suitable for path smoothing, as they are characterized by continuous curvature (Fraichard and Scheuer 2004). However, clothoids generation is challenging, as they have no closed-form expression. High order splines (11th order) and polynomials (26th order) have been proposed for clothoids approximation (Wang et al. 2001; Meek and Walton 2004; Walton and Meek 2005; McCrae and Singh 2009; Montes et al. 2008). Recent work has enabled the real-time approximation of clothoids under bounded length and orientation limitations (Brezak and Petrovic 2013). Consequently, they are still not suited for real-time replanning and highly dimensional scenarios.

Curvature discontinuities result in overactuation, slipping, localization errors (Magid et al. 2006), passenger discomfort (Gulati and Kuipers 2008), mechanical wear and failure (Berglund et al. 2010; Maekawa et al. 2010), and control instability (Lau et al. 2009; Roth and Batavia 2002). Subsequently, achieving continuous curvature is advantageous in applications that involve carrying sensitive cargos such as human passengers (Gulati and Kuipers 2008) or heavy loads in mining applications (Berglund et al. 2010; Maekawa et al. 2010) and those which require precise localization such as agricultural applications to minimize the impact of the vehicle on crops (McPhee and Aird 2013; Sabelhaus et al. 2013; Alshaer et al. 2013) or energy loss minimization for MAVs with battery size restrictions (Myung et al. 2007).

In our earlier work, we proposed an evaluating and bounding B-spline paths approximate solution. We show that humans control vehicles with continuous commands and generate paths that obey the vehicles' kinematic constraints. We propose the premise of using a single B-spline curve to generate paths that resemble human driving and obey the vehicle's constraints. This is achieved by defining the curvature of a B-spline segment in terms of the parameters of its corresponding control polygon, which in this case is assumed to be a linear path generated by a path planning algorithm.

In this work we improve B-spline based motion planning by proposing efficient methods for segment curvature evaluation and analytical bounding. The characteristics of B-splines are exploited to present two solutions for continuous curvature bounding, which can be combined together, or used separately. The novelty of our proposal is that it is not limited to a plane or a dimension; it is not subject to orientation, length, or control polygon restrictions. It guarantees continuity throughout the path whilst preserving real-time performance. We also show that it is possible to plan the trajectory of a robot with nonholonomic constraints and maintain parametric continuity.

This chapter is organized as follows. Sect. 13.2 lists the current related research in path smoothing. The problem is formally described in Sect. 13.3. B-spline curve synthesis is introduced in Sect. 13.4. We address curvature continuity, segment curvature formulation, and curvature bounding in 2D in Sect. 13.5. Our findings are validated and compared with previous work using simulation experiments, as given in Sect. 13.6. Chapter is concluded in Sect. 13.7.

## 13.2 Related Work

There are two separate problems addressed in this chapter. Firstly, planning a geometric curve with curvature bounds given in Sect. 13.4. Secondly, maintaining parametric continuity of the generated trajectory. The authors could not ascertain any literature that combined these two problems. There are only approaches that address each issue separately. The benefits of synthesizing kinodynamically feasible and continuous paths are well studied in robotics (McPhee and Aird 2013; Magid et al. 2006; Gulati et al. 2009; Gulati and Kuipers 2008; Maekawa et al. 2010; Lau et al. 2009; Myung et al. 2007; Sabelhaus et al. 2013; Alshaer et al. 2013). However, current solutions given in the literature fail to guarantee $C^2$ continuity with curvature bounds for nonholonomic mobile robots.

Similarly, in trajectory generation literature there are multiple solutions to optimally generate bounded trajectories for given geometric curves with regard to time (Balkcom and Mason 2002; Wu et al. 2000) and jerk considerations (Guarino Lo Bianco 2013). Others approaches considered curvature and acceleration bounds as parameters in optimization problems (Johnson and Hauser 2012; Kunz and Stilman 2013; Sachin et al. 2014). However these methods, adversely, provided no discussion on the parametric continuity problem and often led to $C^1$ continuity only. The approach developed by Velenis and Tsiotras (2008) for vehicles is limited to velocity continuity and acceleration bounds, and ignored acceleration continuity which will undoubtedly lead to a jerky and uncomfortable ride (Gulati and Kuipers 2008; Guarino Lo Bianco 2013). Control laws proposed for unicycle robots ignored acceleration and curvature bounds (Lapierre et al. 2007; Sgorbissa and Zaccaria 2010; Morro et al. 2011).

Dubins paths and circular arcs were commonly used for robot planning despite their curvature discontinuity. Dubins paths were generated under the assumption that the vehicle maintains constant linear velocity. Multiple circular segments have been proposed for UAV path smoothing (Anderson et al. 2005). This approach was limited to planar scenarios and produces discontinuous paths. Bézier curves were commonly used for path smoothing. The order of Bézier is dependent on the number of control points, which resulted in limiting them to maintain a low order curve (Jolly et al. 2009; Lau et al. 2009; Kwangjin and Sukkarieh 2010). A comparative navigational-based analysis showed that Bézier curves poorly interpolate a linear path as opposed to B-splines (Elbanhawi et al. 2015). Limiting the number of control points of Bézier curves results in the need to join multiple curve segments. Discontinuities arise at the joint of two Bézier segments.

A condition for Bézier curve geometric, $G^2$, continuity was presented in Walton et al. (2003). However, this condition had no closed-form solution. Kwangjin et al. (Kwangjin and Sukkarieh 2010; Kwangjin et al. 2013) provided a solution for a particular case of the $G^2$ planar condition. This approach was used for upper bounded curvature smoothing algorithms. It is still limited to a plane and, in fact, incapable of considering different curvature bounds in horizontal and vertical planes. Fixed-wing UAVs have different turning angle (horizontal plane) and climbing angle (vertical plane) limitations. Barsky and Derose (1990) proposed geometric continuity, $G^k$, as condition for ensuring that curve endpoints had the same directions, not the values. The work in Kwangjin and Sukkarieh (2010) and Kwangjin et al. (2013) fails to guarantee velocity and acceleration continuity, which are more realistic for robotics than geometric continuity. Pan et al. (2012) have shown that only $C^2$ parametric continuities of acceleration and velocity are suitable for real robots and provided a shortcutting algorithm that guarantees continuity in most scenarios, but fails to address the maximum curvature constraint. A more recent personation is on trajectory planning for trailer cars with continuous velocities (Ghilardelli et al. 2014).

The advantages of B-splines for real-time planning have been shown (Dyllong and Visioli 2003; Elbanhawi and Simic 2012). A genetic algorithm was employed to select the location of a fixed number of control points, for a single B-spline curve (Nikolos et al. 2003). This guaranteed the continuity of the curve. However, having a constant number of control points reduced the robustness of the generated path. B-splines were used for generating smooth paths for passenger transporting robots (Gulati and Kuipers 2008). That approach is limited to a 2D setting and robots with no curvature bounds. Similarly, a 3D B-spline smoothing algorithm was presented that did not consider curvature continuity or upper bounds (Koyuncu and Inalhan 2008). Several optimization algorithms are limited to 2D offline B-spline smoothing and curvature bounding (Berglund et al. 2010; Maekawa et al. 2010). A B-spline shortcutting algorithm was proposed, which used multiple segments; however, it did not guarantee continuity in all segments and did not consider the maximum curvature (Pan et al. 2012). In our earlier work we provided an approximate B-spline based approach to the presented problem and did not consider acceleration and velocity bounds (Elbanhawi et al. 2015). Research findings were implemented, in

real time, on experimental vehicle (Elbanhawi and Simic 2014b). On the other hand, here we present solution of the problem analytically and consider kinodynamic constraints.

## 13.3 Problem Statement

We consider front wheel steered vehicles, referred to as car-like robots. It is common to model the vehicle using the bicycle model, which assumes identical steering angles on both sides (Jazar 2008), as shown in Fig. 13.1. This model has been shown to be adequate for modeling the global kinematic motion of front wheel steered vehicles (Campion et al. 1996). The advantage of our approach is in solving two closely related problems simultaneously, which are generally decoupled in robotic literature. We aim to synthesize a curve that satisfies conditions given by Eqs. (13.1–13.4) and maintain parametric continuity. The vehicle's Cartesian coordinates ($x$, $y$) and heading angle $\theta$ are measured from the center of the rear axle relative to a global frame. The length between the front and back wheel is referred to as wheel base, $W$. The two actuation commands are linear velocity, $v$, and steering angle $\Phi$. It is clear that the vehicle is underactuated as it has two controls and three degrees of freedom, i.e. it is not fully controllable (Ogata 2010). The velocity components in the x and y directions, $v_x$ and $v_y$, are constrained as given in Eq. (13.1). This nonholonomic condition is often referred to as the rolling without slipping constraint.

$$\upsilon_x \sin(\theta) - \upsilon_y \cos(\theta) = 0 \tag{13.1}$$

Consider a planning algorithm that produces a path consisting of, $n-1$, straight lines joining successive, $n$, waypoints, $P = [P_1, P_2 \ldots P_n]$, where $P_i = (P_{xi}, P_{yi}, P_{zi})$ for $i = [1, 2, .., n-1, n]$. It is required to generate a curve, $c(u)$, that closely

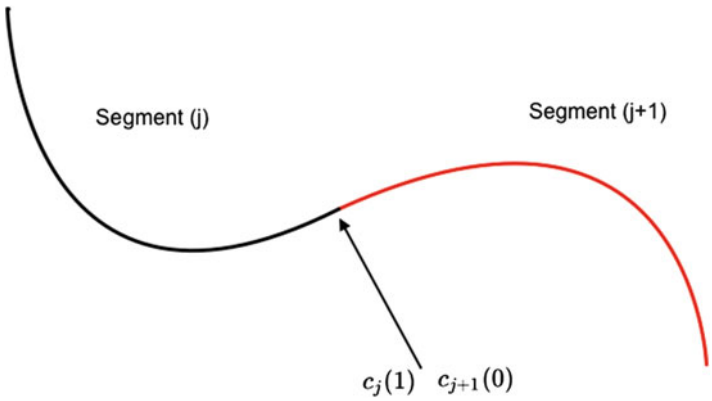



**Fig. 13.1** Joining two path segments

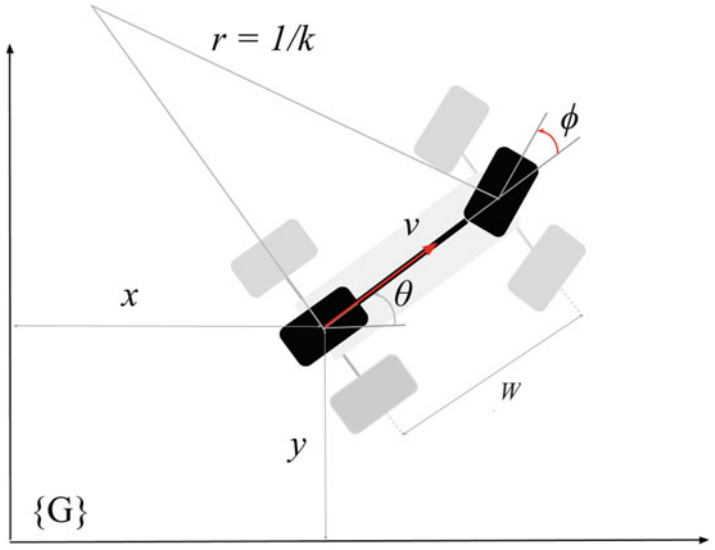



**Fig. 13.2** Bicycle model for front wheel steered vehicles

follows straight lines path, where $u$ is the normalized path length parameter. It is an independent variable in the range of [0,1] for any curve, $c(u)$. Parameter $u$ takes the value $u = 0$ at the beginning of the segment and reached the value $u = 1$ at the end. The generated curve must satisfy the following imposed constraints.

Path continuity at the endpoints of two curve segments must be addressed; such a situation is illustrated in Fig. 13.1. For two consecutive curve segments $c_j(u)$ and $c_{j+1}(u)$, $C^k$ parametric continuity could then be defined as shown in Eq. (13.2), according to Farin (2002), where $k$ is a positive integer denoting the order of the parametric continuity.

$$\frac{\partial^i c_j + (0)}{\partial u^i} = \frac{\partial^i c_j(1)}{\partial u^i}, \forall i = 1, 2, 3..k \tag{13.2}$$

The curvature of the path must not exceed the maximum curvature, $K_{\max}$, at any point. The curvature, $k(u)$, along a path is defined as follows in Eq. (13.3), where $c(u) = [x(u), y(u)]$ and the first and second order derivatives with respect to $u$ are $c'(u) = [x'(u), y'(u)]$ and $c''(u) = [x''(u), y''(u)]$.

$$k(u) = \frac{x'(u)y''(u) - x''(u)y'(u)}{\left(x'(u)^2 + y'(u)^2\right)^{3/2}} \tag{13.3}$$

The minimum radius of curvature, $r_{\min}$, in a plane restricts the curvature of the path to $K_{\max}$ in that plane. For car-like robots, the curvature constraint is a result

of the maximum steering, $\Phi_{\max}$ angle due to the mechanical construction of the vehicle, as shown in Eq. (13.4). In three-dimensional scenarios for aerial vehicles, the maximum yaw and pitch angles in the horizontal and vertical planes must be considered separately.

$$k_{\max} = \frac{1}{r_{\min}} = \frac{\tan(\phi_{\max})}{W} \tag{13.4}$$

## 13.4 Spline Primitives

B-splines are vector-valued parametric curves, initially proposed by Schoenberg (1946). B-splines and Non-uniform Rational B-splines (NURBS) are commonly used for Computer Aided Design (CAD) applications as a result of their efficient synthesis and robustness (Farin 1992). In addition to CAD, they have been utilized in reverse engineering (Ma and Kruth 1998; Piegl and Tiller 2001), finite element analysis (Hughes et al. 2008), machining (Cheng et al. 2002; Sungchul and Taehoon 2003), medical imaging (Zhang et al. 2007), computer vision (Biswas and Lovell 2008), bio-inspired data fitting (Jones and Adamatzky 2014), and signal processing (Unser et al. 1993). As discussed in earlier sections, their use in robotics is fairly recent.

A $p$-th degree B-spline curve, $c(u)$, is defined by $n$ control points and a knot vector $\hat{u}$, evaluated by Eq. (13.5). The length of the one-dimensional knot vector, $m$, is equal to $n + p + 1$. Normalized path length parameter, $u$, is simply referred to as the path parameter (Farin 2002).

$$c(u) = \sum_{i=0}^{n} N_{i,p}(u) P_i \tag{13.5}$$

$P_i$ is the $i$-th control point, which is in turn influenced by a corresponding basis function. The number of basis functions therefore mirrors the number of control points, $n$. $N_{n,i}(u)$ is the $i$-th B-spline basis function, which is defined using the Cox-de Boor recursive algorithm (De Boor 1972). First order basis functions are evaluated using Eq. (13.6) based on the predefined knot vector. Higher order functions are computed by the recursive substitution in Eq. (13.7).

$$N_{i,0}(u) = \begin{cases} 1 & u \in [\widehat{u}_i, \widehat{u}_{i+1}) \\ 0 & else \end{cases} \tag{13.6}$$

$$N_{i,p}(u) = \frac{u - \widehat{u}_i}{\widehat{u}_{i+p} - \widehat{u}_i} N_{i,p-1}(u) + \frac{\widehat{u}_{i+p+1} - u}{\widehat{u}_{i+p+1} - \widehat{u}_{i+1}} N_{i+1,p-1}(u) \tag{13.7}$$

We have previously shown B-splines properties that render them as superior to other parametric curves, for the task of robot navigation (Elbanhawi et al. 2015). The curve's degree, $p$, is independent of the number of control points, $n$. This allows the possibility of using a single curve for the entire path smoothing without imposing limitations on the number of control points. It is in contrast to Bézier curve methods (Jolly et al. 2009; Kwangjin and Sukkarieh 2010; Lau et al. 2009; Kwangjin et al. 2013) where the number of control points is predefined. Modification of control points affects the curve shape locally and does not change the rest of the path. This enables the local control of the path for smoothing or obstacle avoidance purposes. A clamped B-spline curve follows its control polygon more closely in comparison with a Bézier curve of the same order. Clamping is achieved by having $(p + 1)$ multiplicity of the initial and final knots, $\hat{u}$ (Farin 2002). Knot multiplicity ensures that the curve passes through the initial and final control points.

Despite the beneficial properties that characterize B-splines, maintaining path continuity and controlling its curvature are nontrivial issues. They continue to challenge the use of B-splines in robotic path planning applications (Lau et al. 2009; Elbanhawi and Simic 2014a; Pan et al. 2012). It must be noted that both Beziers and B-spline are essentially combinations of polynomials. In principle, there should exist control laws, or conditions, that are capable of generating parametrically continuous trajectories using Bezier curves as well. The authors could not identify such methods in literature. Consequently, we have utilized the existing benefits of B-splines for motion planning.

## 13.5 Curvature Bounding

### *13.5.1 Parametric Continuity*

The challenge of path continuity stemmed from the linking of two separate path segments. Primitives such as circular arcs, polynomials, and clothoids were not flexible enough to represent a path using a single segment. The number of control points, which were usually predefined prior to smoothing, governs the order of a Bézier curve. Consequently, multiple Bézier curves must be linked for smoothing a single piecewise linear path.

The order of a B-spline curve is independent of the number of control points in the path, as already mentioned. In theory, it is possible to smooth a path using a single curve of a predefined order. The single B-spline curve approach was adopted for UAV planning; however, the number of control points was fixed (Nikolos et al. 2003). The region in which planning is conducted and path shape robustness are significantly limited by fixing the number of control points. The work by Jolly et al. (2009) is based on rapid replanning with a short planning horizon and relies on four control points. We did not pose any restrictions on the number of control points, apart from that the number of control points must exceed the degree of the curve, $p$.

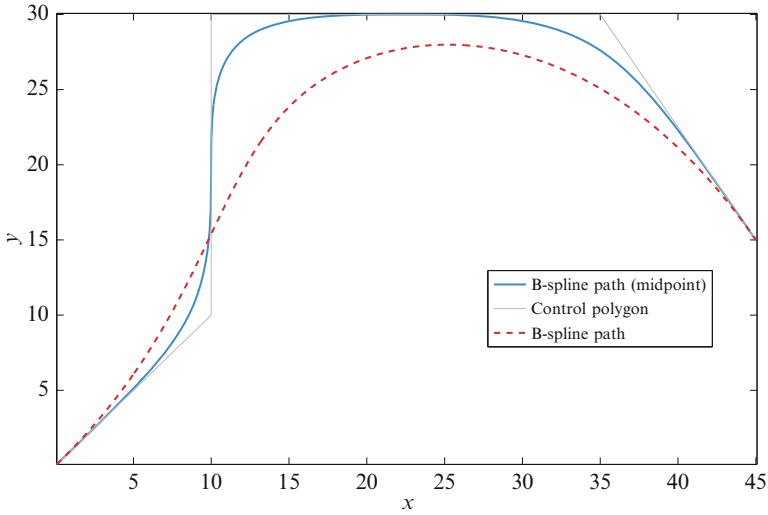



**Fig. 13.3** *Midpoint* insertion improves the path proximity of B-splines without compromising parametric continuity. It forces the curve (*blue*) tangency to the edge of the control polygon (*black*) unlike the unmodified B-spline curve (*red*)

The local control property of B-spline enables the modification of a curve segment without changing the entire path. The necessity for re-routing commonly results from obstacle detection or smoothing purposes.

Despite the superiority of B-splines, over a Bézier curve of the same order, in closely following the shape of a path, they still deviate from the original path (control polygon). Ideally, the curve would follow the original linear path and smoothly cut corners when turning is needed. It is desired to maintain proximity to the originally planned straight-line path as it is more likely to be collision free. This was achieved by forcing the tangency of the curve to the sides of the control polygon. B-spline tangency to collinear control points is leveraged to ensure the close following of the original path. Systematic midpoint insertion, between every two successive points, effectively transformed control polygon edges into lines connecting three control points, thus forcing the curve's tangency to the edges. The effect of midpoint insertion is illustrated in Fig. 13.3. It is worth highlighting that in both cases, a single curve segment was used for smoothing. That guarantees continuity along the path. This avoids the need to address parametric continuity at union points, as illustrated in Fig. 13.1. The curvature and higher order derivatives do not exhibit any abrupt changes after adding midpoints. This can be validated by comparing the resulting trajectories given in Fig. 13.4.

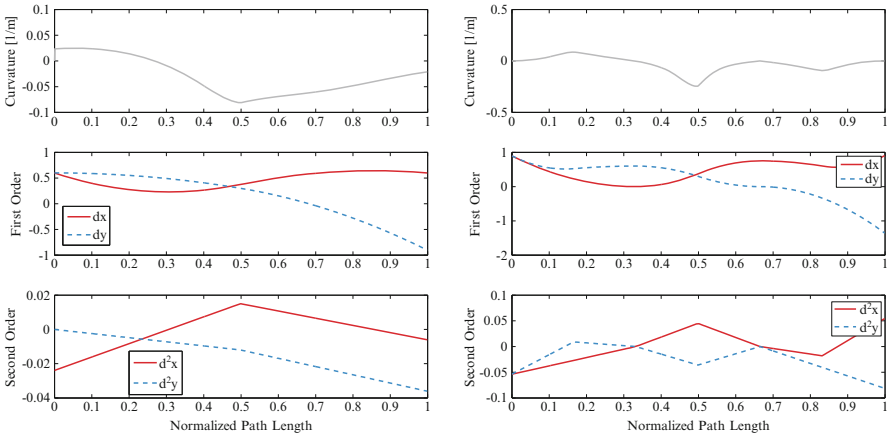



**Fig. 13.4** Parametric continuity was maintained before (*left*) and after (*right*) midpoint insertion as a result of using a single B-spline segment

## *13.5.2 Curvature Evaluation*

### 13.5.2.1 Path Segmentation

Our aim was to control the curvature, $k$, of a B-spline curve. Specifically, it was required to maintain the curvature below the maximum curvature bound, $K_{\max}$. Systematic midpoint insertion allowed for the definition of a repeated segment throughout the path (see for Fig. 13.5a illustration). The segment consists of two intersecting control edges and a total of five control points (including two midpoints). It was required to define B-spline paths curvature in terms of their corresponding segment parameters. This enabled the isolation of each segment and local modifications of its parameters by leveraging the local support property of B-splines. Smoothing modifications will be proposed to ensure maximum curvature bounds are obeyed.

The parameters of the reoccurring control segment are the side length, $L$, the angle between segment sides, $\alpha$, and the length ratio of both sides, $r$, as illustrated in Fig. 13.5b. In our earlier work (Elbanhawi et al. 2015), segments of equal sides were assumed, $r = 1$, which overestimated the curvature of the path and resulted in attaining approximate solutions. The use of the length ratio parameter, $r$, is presented to enable a more precise evaluation of the curvature. Position vectors describing the five control points of the segment can be defined with respect to the parameters of the same segment and are given in Eq. (13.8).

$$P = \begin{pmatrix} P_x \\ P_y \end{pmatrix} = \begin{pmatrix} L, \frac{L}{2}, 0, r\frac{L}{2}\cos(\alpha), rL\cos(\alpha) \\ 0, 0, 0, r\frac{L}{2}\sin(\alpha), rL\sin(\alpha) \end{pmatrix} \tag{13.8}$$

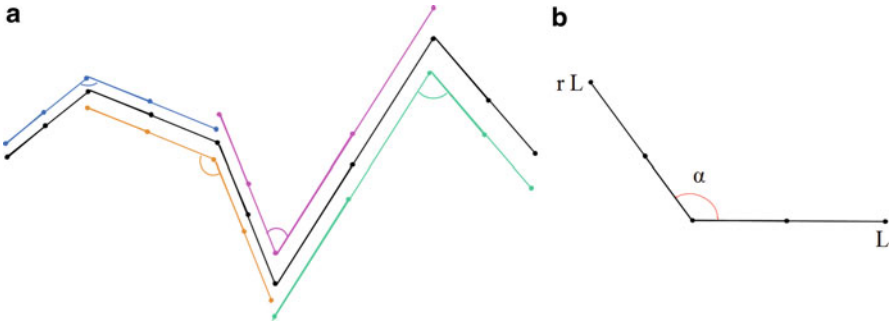



**Fig. 13.5** (**a**) The notion of a reoccurring control segment through the path. A segment consists of two intersecting *straight lines* and five *control points*. (**b**) The parameters of a single segment

The cubic B-spline curve, $p = 3$, has five control points, $n = 5$, and, $m = 9$, knots with four initial and final multiplicity for clamping, $\widehat{u} = [0, 0, 0, 0, 0.5, 1, 1, 1, 1]$. Initial order basis functions were evaluated using Eq. (13.6). Following that, basis functions, *N(u)*, were computed, using the Cox-deBoor Algorithm by recursive evaluation of Eq. (13.7). The following third order basis functions can be defined as given in the set of Eq. (9).

$$N_{0,3} = \frac{(1-2u)^3}{2} \tag{13.9a}$$

$$N_{1,3} = 3u^3 - 3u^2 + 1 \tag{13.9b}$$

$$N_{2,3} = 6u^2 - 6u - 1 \tag{13.9c}$$

$$N_{3,3} = -6u^3 + 12u^2 - 6u + 1 \tag{13.9d}$$

$$N_{4,3} = \frac{(2u-1)^3}{2} \tag{13.9e}$$

In order to define the curvature of a segment in terms of its parameters, $k = f(r, L, \alpha)$, the position vectors of the segment, Eq. (13.8), and basis functions, Eq. (9), were substituted in the curve Eq. (13.5). The curve was defined as a function of its corresponding segment parameters, $c(u) = \begin{bmatrix} x(u) \\ y(u) \end{bmatrix} = \begin{bmatrix} x(r, L, \propto) \\ y(r, L, \propto) \end{bmatrix}$, $x(u)$, $x(u)$ and $y(u)$ are given in Eq. (10).

$$x(u) = \frac{(1-2u)^3}{2} * L + \left(6u^3 - 6u^2 + 1\right) * \frac{L}{2} + \left(-6u^3 + 12u^2 - 6u + 1\right) * \frac{rL\cos(\alpha)}{2} + \frac{(2u-1)^3}{2} * rL\cos(\alpha) \tag{13.10a}$$

$$y(u) = \left(-6u^3 + 12u^2 - 6u + 1\right) * \frac{rL\sin(\alpha)}{2} + \frac{(2u-1)^3}{2} * rL\sin(\alpha) \tag{13.10b}$$

For a given segment, its parameters, *r, L, α* are constant and known prior to a curvature query. The first and second order derivatives with respect to the path parameter, *u*, are derived below from equations set (13.11a–13.11d).

$$x'(u) = 3L\left(u^2(r\cos(\alpha) - 1) + 2u + 1\right) \tag{13.11a}$$

$$x''(u) = 6L(u(r\cos(\alpha) - 1) + 1) \tag{13.11b}$$

$$y'(u) = 3Lr\sin(\alpha)\,u^2 \tag{13.11c}$$

$$y''(u) = 6Lr\sin(\alpha)\,u \tag{13.11d}$$

The curvature expression $k = f(r, L, \alpha)$ in Eq. (13.12) was obtained by substituting the curve and its first and second order derivatives from Eq. (11) into Eq. (13.3). It can be noted that when substituting by $r = 1$, in Eq. (13.12) we get the same expression derived in Elbanhawi et al. (2015). Prior to introducing the parameter, *r*, curvature evaluations were approximate and the accuracy of the maneuvers could not be ascertained.

$$k(u) = \frac{2ru(u-1)\sin(\alpha)}{3L(u^4(r^2 - 2r\cos(\alpha) + 1) + 4u^3(r\cos(\alpha) - 1) - 2u^2(r\cos(\alpha) - 3 - 4u + 1))^{3/2}} \tag{13.12}$$

#### 13.5.2.2 Segment Curvature Evaluation

Midpoint insertion ensured the curve's tangency to the polygon edges, which resembled paths generated by human operators in the experiments conducted by Elbanhawi et al. (2015). Subsequently, the curvature of the path started, at $u = 0$, and finished, at $u = 1$, with $k = 0$. Curvature peaked to $k_{peak}$ at some point, $u_{peak}$, in between $u = [0,1]$. In order to limit path curvature to the maximum value of $K_{max}$, the peak curvature, $k_{peak}$, of the segments must be evaluated first. The point, $u_{peak}$, along the parametric path length, *u*, where the curvature peaks, was found by solving Eq. (13.13). Then $k_{peak}$ was computed by substituting $u_{peak}$ in Eq. (13.12).

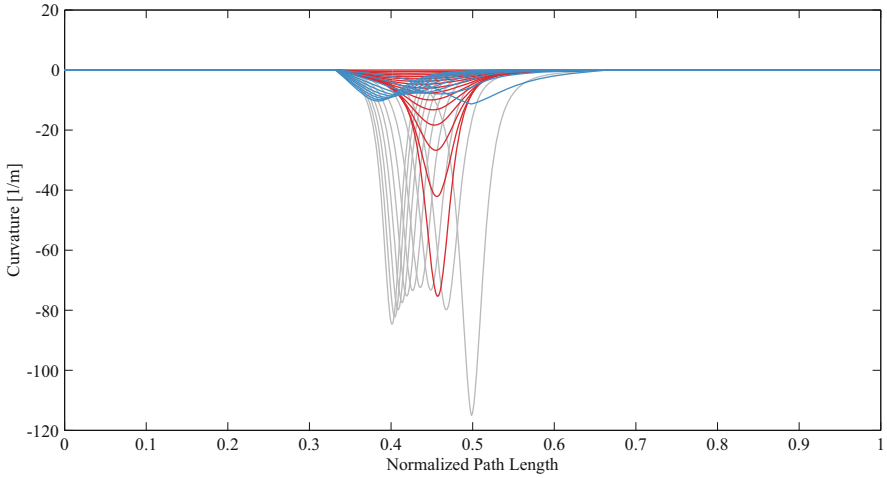



**Fig. 13.6** Changing segment parameters shifts the position of the curvature peaks. In all cases, curvature profile is continuous with a singular peak

$$\frac{dk(u)}{du} = 0 \tag{13.13}$$

For every path segment, a singular curvature peak exists, as shown in Fig. 13.6. The red profiles shown the influence of changing the segment angle whilst maintaining fixed length and ratio. The location of the peak curvature was entirely dependent on the length and ratio. For a large angle (blue) and fixed length, the ratio changed both the position and the value of the peak curvature. Similarly, for a much smaller segment angle (grey) the length ratio was still influential on both the peak value and position.

Solving Eq. (13.13) for $u_{\text{peak}}$ can prove to be a computationally intensive task, particularly when $k_{\text{peak}}$ had to be evaluated multiple times during each query of the path planning procedure. One useful observation is that the location of $u_{\text{peak}}$ is dependent on the segment angle, $\alpha$, and length ratio, $r$, as highlighted in Fig. 13.7. We note that whilst $u_{\text{peak}}$ is dependent on $r$ and $\alpha$ only, the peak curvature value, $k_{\text{peak}}$, is still dependent on $r$, $\alpha$, and $L$. It was possible to store $u_{\text{peak}}$ values in a lookup table of equal intervals from $r = 1$ to $10$ and $\alpha = 0$ to $\pi$. The required values can be interpolated. To maintain a sparse lookup table we use the property in Eq. (13.14), which can be observed from Fig. 13.7. In our case, retaining a lookup table (*less than* 10 kB *in size*) produced curvature values of $10^{-3}$ accuracy.

$$\text{If } 0 < r < 1, \text{ then } u_{\text{peak}\left(r,\,\alpha\right)} = 1 - u_{\text{peak }(1/r,\,\alpha)} \tag{13.14}$$

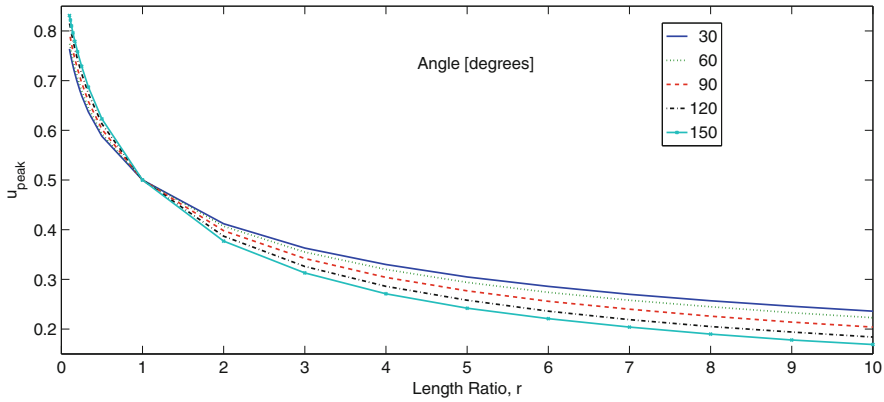



**Fig. 13.7** Parametric length location, $u_{\text{peak}}$, of the peak curvature, $k_{\text{peak}}$, is dependent on the segment angle, $\alpha$, and the length ratio, $r$. It can be noted that when length ratio is $0 < r < 1$, $u_{\text{peak}} > 0.5$ and when $r > 1$, $u_{\text{peak}} < 0.5$. This results from the observation that $u_{\text{peak}}$ is shifted towards the shorter segment edge

### *13.5.3 Curvature Bounding*

In this section, two analytical solutions for curvature bounding are presented. They ensured peak segment curvature does not exceed the maximum curvature, $k_{\text{peak}} \leq K_{\max}$. This confirms that the path is feasible, having shown in the previous section that each path segment has a single peak. The first solution was relaxed ensuring a smooth curvature. The second solution was strict to minimize deviation from the original control polygon. It was possible to combine both conditions in different segments, on account of B-spline local support property, with minimal effect on other segments. Both conditions were designed to make certain that the path was contained within the convex hull of the original control polygon to reduce the probability of the obstacle collision. The presented solutions ensure feasibility through continuous transformation of the original path such that both paths, solution and original, are homotopic. Nonetheless, guaranteeing that the path is collision free was not addressed in this work. We assume that this work will eventually be combined within a planning framework and will not be restricted to path smoothing.

#### 13.5.3.1 Single Peak Solution

Consider the single control segment, shown in Fig. 13.8, whose corresponding B-spline curvature violates the maximum curvature condition. The segment consists of two lines $l_{n,\,n+2}$, joining point *(n)* and point $(n+2)$, and $l_{n+2,\,n+4}$, joining point $(n+2)$ and point $(n+4)$, shown as solid black lines. Point *(o)* is the intersection point between $l_{n,\,n+4}$ (thin grey line) and line $l_{o,\,n+2}$ (dotted blue line) which is passing through point $(n+2)$ and is orthogonal to $l_{n,\,n+4}$.

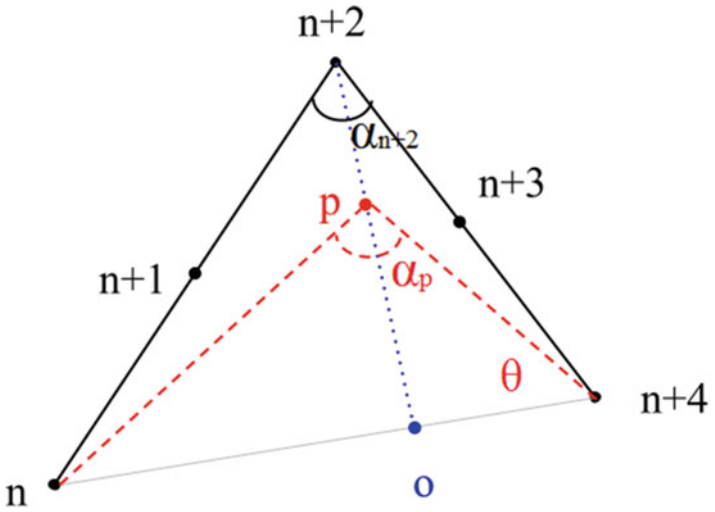



**Fig. 13.8** First smoothing solution; it is required to find the point *(P)* along the line (*dotted blue line*), joining point $(n+2)$ and point *(o)*, that ensures the curvature, $k_{\text{peak}}$, does not exceed $K_{\max}$

The current curvature, $k_{n+2}$, and segment angle, $\alpha_{n+2}$, are known, and $k_{n+2} > K_{\max}$. Assume that point $(n+2)$ is shifted towards point *(o)*, along the line, $l_{o,\,n+2}$, whilst points *(n)* and $(n+4)$ are unchanged and the midpoints $(n+1)$ and $(n+3)$ are re-computed accordingly. Finally at $\alpha_{n+2} = \pi$, $k_o = 0$. It is required to find the nearest point $(p)$, at which $k_p = K_{\max}$, as point $(n+2)$ is being shifted towards *(o)* along $l_{o,\,n+2}$. The minimum angle $\alpha_{\text{p}}$ lies between $\alpha_o = \pi$ and $\alpha_{n+2}$ as given by Eq. (13.15). We define $\overline{l}_{i,j}$as the Euclidean distance between two points $(i)$ and $(j)$ whose Cartesian coordinates are known.

Assuming, Line $l_{o,\,n+2}$ is parameterized between $P_{n+2}$ and $P_o$ using $\widehat{l} = [0, 1]$. It is required the value of $\widehat{l}$ where the Point $(p)$ satisfies the curvature requirement. Firstly, $P_p$ is given as follows:

$$P_p = P_{n+2}\left(1 - \widehat{l}\right) + P_0\widehat{l} \tag{13.15}$$

In every iteration, the curvature is evaluated until the $k_p = K_{\max}$ condition is satisfied. To optimize the search, we can estimate the initial point where the curvature may be equal to $K_{\max}$. This is achieved by knowing that at $\widehat{l} = 0$, $k = k_p$ and at $\widehat{l} = 1$, $k = 0$

$$\widehat{l}_{\text{init}} = 1 - \left|\frac{K_{\max}}{K_p}\right| \tag{13.16}$$

An example of curvature bounding is shown in Fig. 13.9 using this solution. The resulting curvature has a single segment as shown in Fig. 13.10 and was bound to 0.14 $\text{m}^{-1}$. Curvature continuity was maintained in both cases.

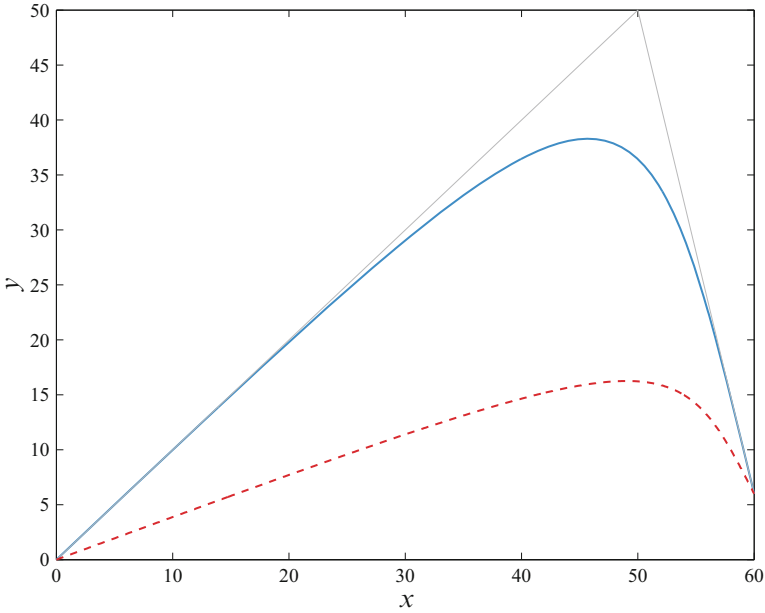



**Fig. 13.9** Bounding using single peak solution. The original path is *blue* and new path is *red*

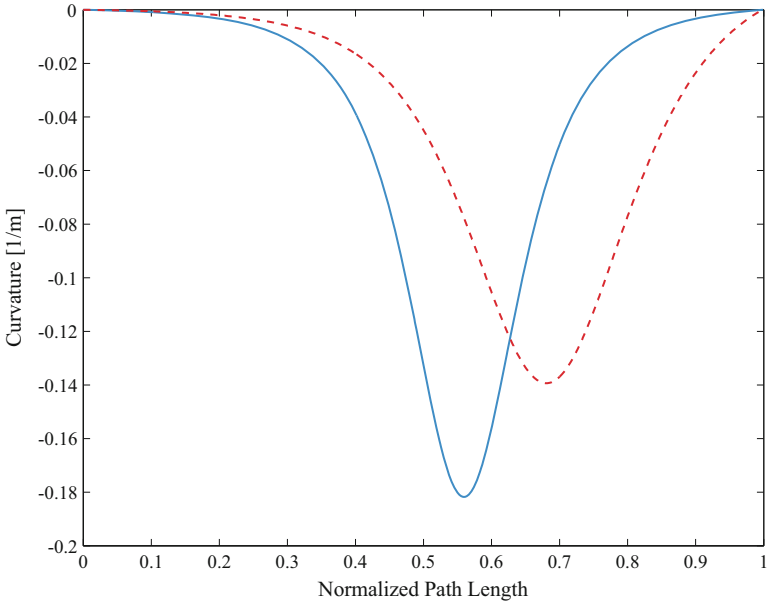



**Fig. 13.10** Resulting curvature profiles before (*blue*) and after (*red*) bounding

**Table 13.1** Comparing related methods

| Method | Curvature bounds | Curve | Path | Continuity |
|---|---|---|---|---|
| Nikolos et al. (2003) | Yes | B-spline | 3D | $C^2$ |
| Anderson et al. (2005) | Yes | Arcs/line | 2D | No |
| Gulati and Kuipers (2008) | No | B-spline | 2D | No |
| Koyuncu and Inalhan (2008) | No | B-spline | 3D | No |
| Jolly et al. (2009) | No | Bézier | 2D | $C^2$ |
| Lau et al. (2009) | No | Bézier | 2D | No |
| Berglund et al. (2010) | Yes | B-spline | 2D | No |
| Maekawa et al. (2010) | Yes | B-spline | 2D | No |
| Pan et al. (2012) | No | B-spline | 3D/2D | $C^2/C^1$ |
| Kwangjin (2013) | Yes | Bézier | 2D | $G^2$ |
| Huh and Chang (2014) | No | Polynomial | 2D | $G^2$ |
| Proposed | Yes | B-spline | 2D/3D | $C^2$ |

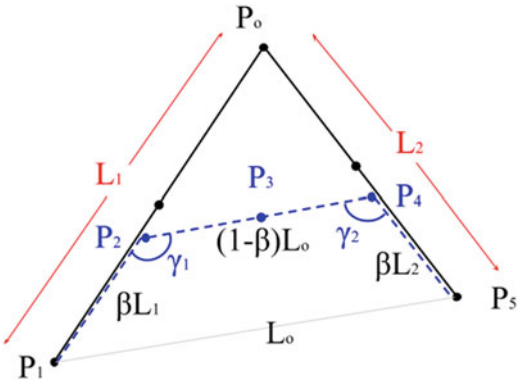



**Fig. 13.11** Second smoothing solution; it is required to find the value of $\beta$ that ensures curvature bounding in both segments and minimizes the total path length

### 13.5.3.2 Double Peak Solution

In this section we proposed a different approach for the same problem considered in the previous section. The curvature of a control segment, $P_1$, $P_0$, $P_5$ and their midpoints in Fig. 13.1 exceeds $K_{max.}$ Segment, $P_1$, $P_0$, $P_5$, is decomposed into two segments, $P_1$, $P_2$, $P_4$ (segment 1) and $P_2$, $P_4$, $P_5$ (segment 2). Line segment $\overline{P_2P_4}$ is constructed to be parallel to edge $\overline{P_1P_5}$. As a result, triangles $\Delta P_1P_0P_5$ and $\Delta P_2P_0P_4$ are similar and the ratio between their side lengths is $(1-\beta)$, where $0<\beta<1$. Segment 1 and 2 parameters can be described in terms of $\beta$, where segment angles are constant, as given in Table 13.1 (Fig. 13.11).

By substituting the segment parameters, given in Table 13.2, in Eq. (13.12), it is possible to find a range for $\beta$, subset of set [0,1), in which both segments curvatures are less than $K_{max}$. Firstly, we compute a separate range for each segment 1 and 2 $[\beta_{min1}, \beta_{max1}]$ and $[\beta_{min2}, \beta_{max2}]$. These computations are efficient by virtue of using the lookup table in the previous section. The allowable range for $\beta$ is $[\max(\beta_{min1}, \beta_{min2}), \min(\beta_{max1}, \beta_{max2})]$.

**Table 13.2** Segment Parameter

| Parameter | Symbol | Segment 1 | Segment 2 |
|---|---|---|---|
| Length ratio | $r$ | $\frac{(1-\beta)L0}{\beta L1}$ | $\frac{(1-\beta)L0}{\beta L_2}$ |
| Edge Length | $L$ | $\beta L_1$ | $\beta L_2$ |
| Segment Angle | $\alpha$ | $\gamma_1$ | $\gamma_2$ |

We nominate the $\beta$ value that minimizes the total length. Now, new segment control points, $P_2$, $P_3$, $P_4$, can be computed, where for any control point we have $P_i = (x_i, y_i)$.

$$\beta = \arg\min\left(\beta\left(L_1 + L_2 - L_0\right), \beta \in [\beta_{\min}, \beta_{\max}]\right) \tag{13.17}$$

$$P_2 = \beta P_1 + (1-\beta) P_0 \tag{13.18}$$

$$P_4 = \beta P_5 + (1-\beta) P_0 \tag{13.19}$$

A midpoint is inserted between the two added points based on the ratio between the lengths of both, such that if both lines are equal, $r = 1$, the midpoint is equidistant between them.

$$P_3 = \frac{r}{r+1}\left(P_4 - P_2\right) + P_2 \tag{13.20}$$

An example of curvature bounding is shown in Fig. 13.12 using this solution. The resulting curvature has two segments as shown in Fig. 13.13 and was bound to 0.14 $\mathrm{m}^{-1}$. Curvature continuity was maintained in both cases.

## 13.6 Results

### *13.6.1 Curvature Evaluation*

To efficiently evaluate a segment's curvature we proposed storing the peak curvature position $u_{\text{peak}}$ in a sparse lookup table and evaluating the curvature using the segment parameters. We conducted 1000 queries, for a range of segment parameters where $r$ and $L = [1\text{ m}, 10\text{ m}]$ in steps of 1 m and $\alpha$ was $= [30°, 180°]$ in steps of 15°. The time performance of this evaluation method was compared with solving Eq. (13.13). From the results, given in Table 13.3, it is clear that this method is more efficient. Comparing with previously published research results (Elbanhawi et al. 2015), which assumed equal segment length, we show that this approach has better accuracy. The results are illustrated in Fig. 13.14 and given in Table 13.4.

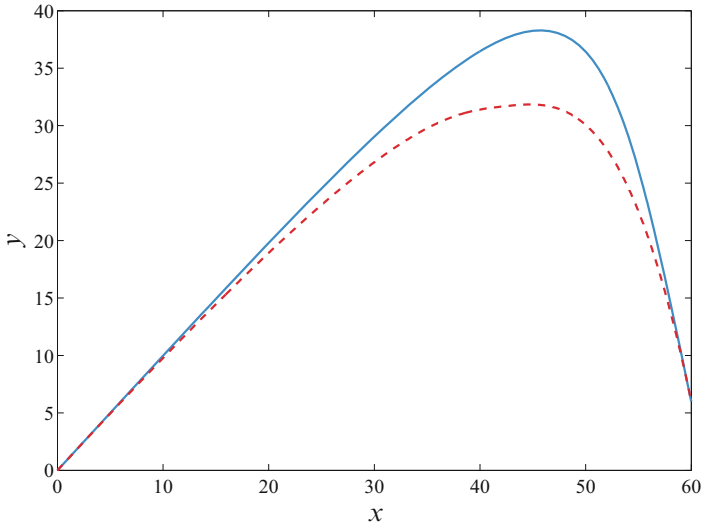



**Fig. 13.12** Bounding using double peak solution. The original path is *blue* and the feasible path is *red*

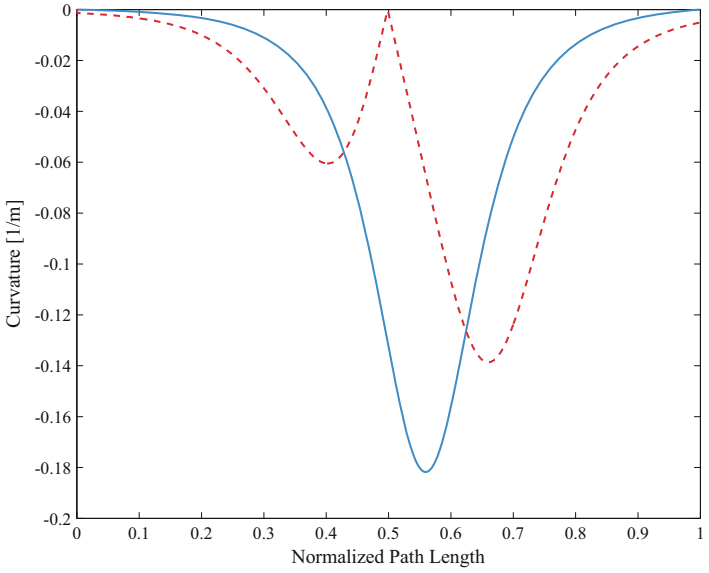



**Fig. 13.13** Resulting curvature profiles before (*blue*) and after (*red*) bounding

**Table 13.3** Curvature evaluation time performance for 1000 queries

| Query time | Analytical Solution | Lookup table |
|---|---|---|
| Mean (ms) | 122.16 | 0.91 |
| Standard deviation (ms) | 6.95 | 0.34 |

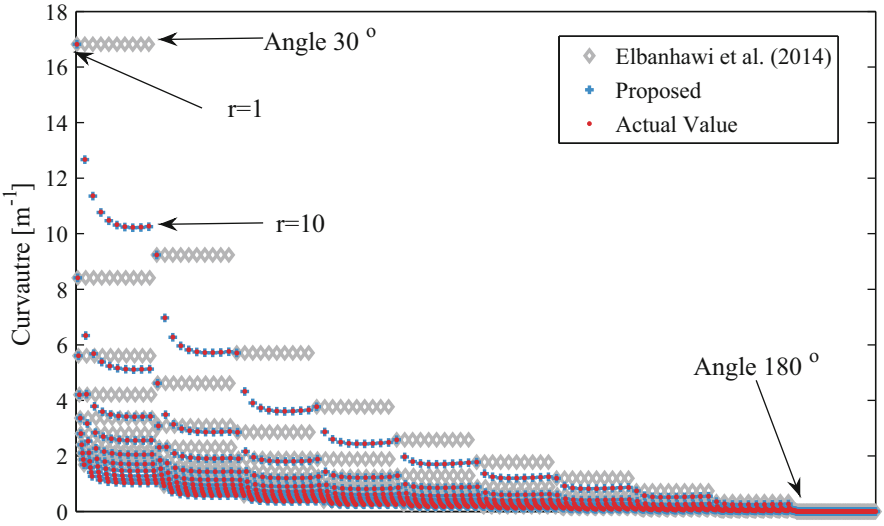



**Fig. 13.14** Curvature evaluation errors of proposed lookup table compared to Elbanhawi et al. (2015)

**Table 13.4** Curvature evaluation errors

| Error ($mm^{-1}$) | Proposed | Elbanhawi et al. (2015) |
|---|---|---|
| Mean | 0.87 | 384.25 |
| Standard deviation | 1.738 | 788.45 |
| Maximum | 19.34 | 6592.10 |

### *13.6.2 Curvature Bounding*

In this section we compared the presented bounding solutions to our earlier work in Elbanhawi et al. (2015). Two different examples were used as shown in Figs. 13.15 and 13.18. The linear reference paths are assumed to result from a planning algorithm. It can be noted that the proposed solutions maintain the curve within the convex hull of the original reference path. In both cases, the curvature is successfully bounded, to 0.2 and 0.15 $m^{-1}$ successively, and its continuity is maintained as shown in Figs. 13.16–13.18. The proposed solutions reduce the deviation from the original path and the total path length, outperforming our earlier work as detailed in Tables 13.5 and 13.6. Solution (1) results in a low frequency single peak curvature profile as opposed to solution (2), which may have a better impact on passenger comfort in autonomous cars, as suggested in Gulati and Kuipers (2008) and Turner and Griffin (1999). On the other hand, solution (2) minimizes deviation from the reference paths and as a consequence minimizing the risk of collision.

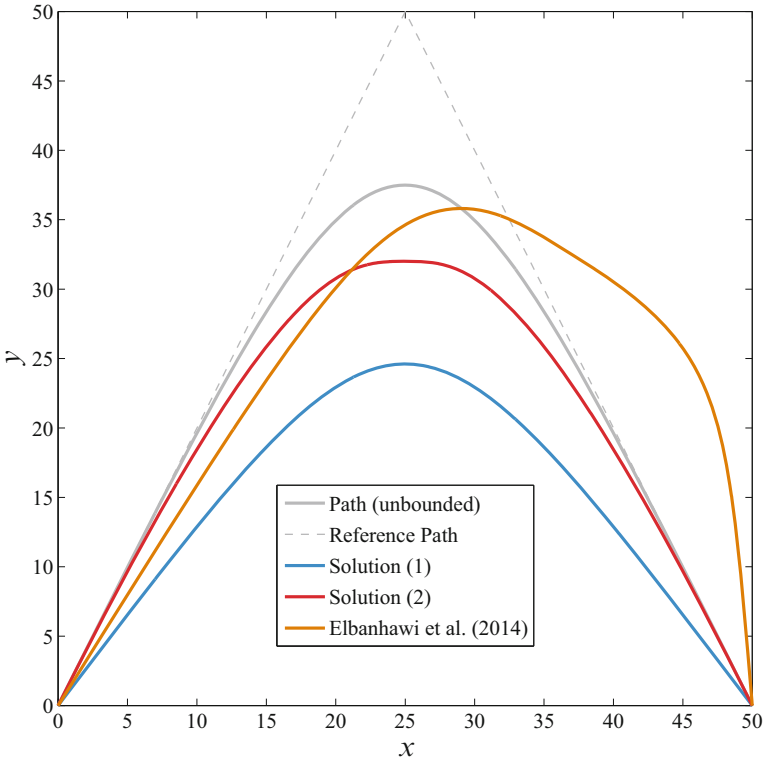



**Fig. 13.15** Example 1: Bounding paths using different methods

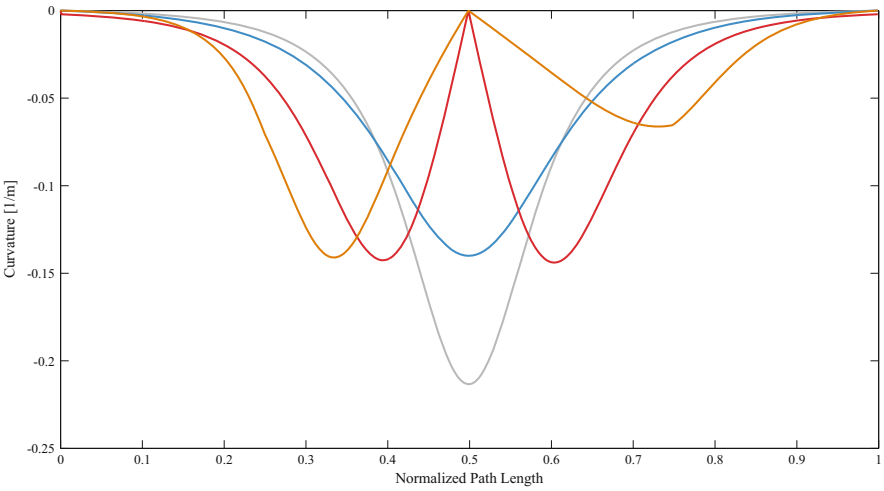



**Fig. 13.16** Example 1: Resulting curvature profiles

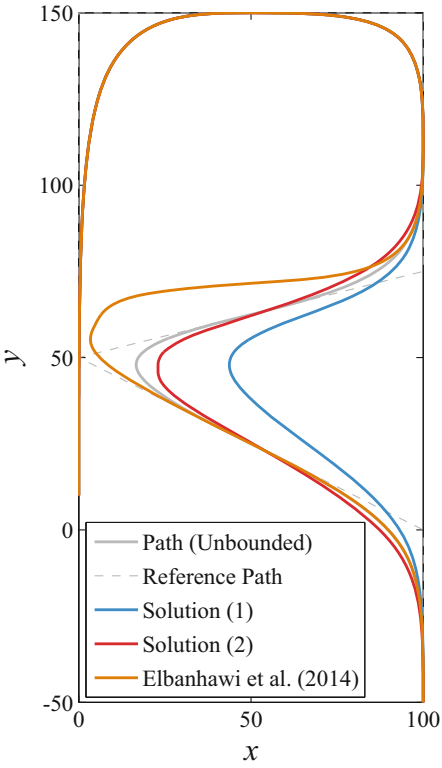



**Fig. 13.17** Example 2: Bounding paths using different methods

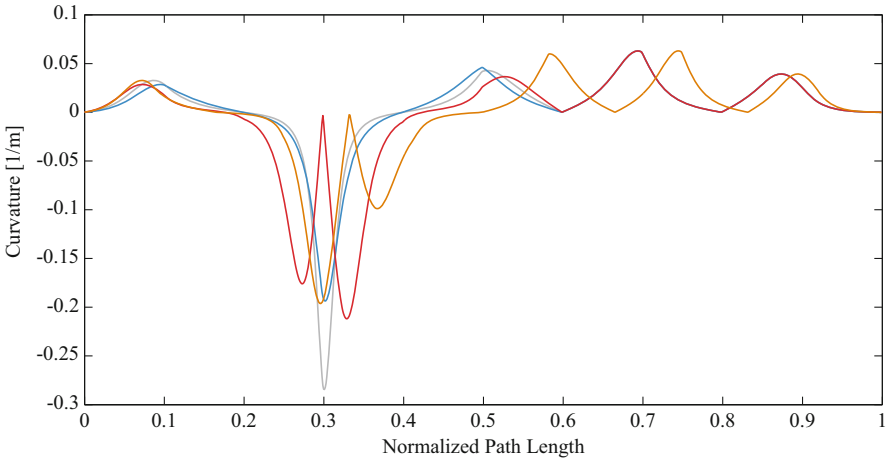



**Fig. 13.18** Example 2: Resulting curvature profiles

**Table 13.5** Example 1: Resulting path lengths and deviation

| Path | Length (m) | Deviation mean (m) | Deviation maximum (m) |
|---|---|---|---|
| Linear path | 94.33 | – | – |
| Reference B-spline | 92.28 | – | – |
| Elbanhawi et al. (2015) | 112.59 | 12.48 | 17.19 |
| Solution (1) | 71.78 | 8.58 | 12.89 |
| Solution (2) | 84.42 | 2.94 | 5.50 |

**Table 13.6** Example 2: Resulting path lengths and deviation

| Path | Length (m) | Deviation mean (m) | Deviation maximum (m) |
|---|---|---|---|
| Linear Path | 579.88 | – | – |
| Reference B-spline | 507.35 | – | – |
| Elbanhawi et al. (2015) | 535.28 | 20.15 | 45.04 |
| Solution (1) | 464.47 | 6.48 | 27.03 |
| Solution (2) | 506.27 | 6.04 | 23.99 |

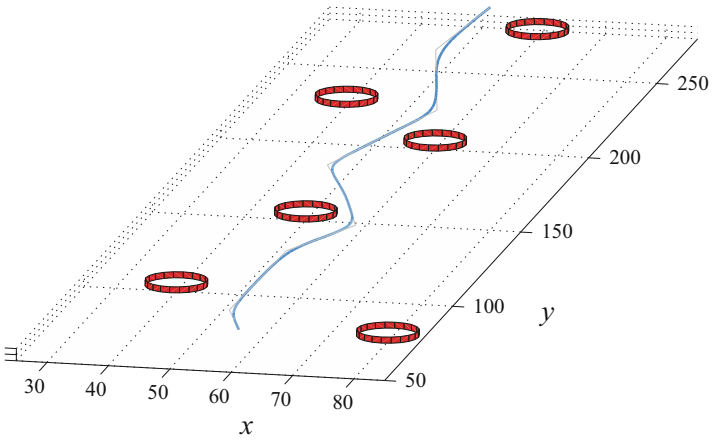



**Fig. 13.19** Example 3: Kinodynamic motion amongst obstacles

Example (3) highlights the ability of the proposed method to generate a feasible path amongst obstacles. The benefit of maintaining the curve within the convex hull of the path is apparent in this example. The linear path was generated from an RRT algorithm (Elbanhawi and Simic 2014b). The resulting B-spline path amongst obstacles is illustrated in Fig. 13.19. Post-processing RRT algorithms have been shown to improve path quality and produce fairly consistent results. Nonetheless these methods do not guarantee the path is collision free. The resulting trajectory is given in Fig. 13.20. It is clear the multi-segment path maintains curvature and parametric continuity.

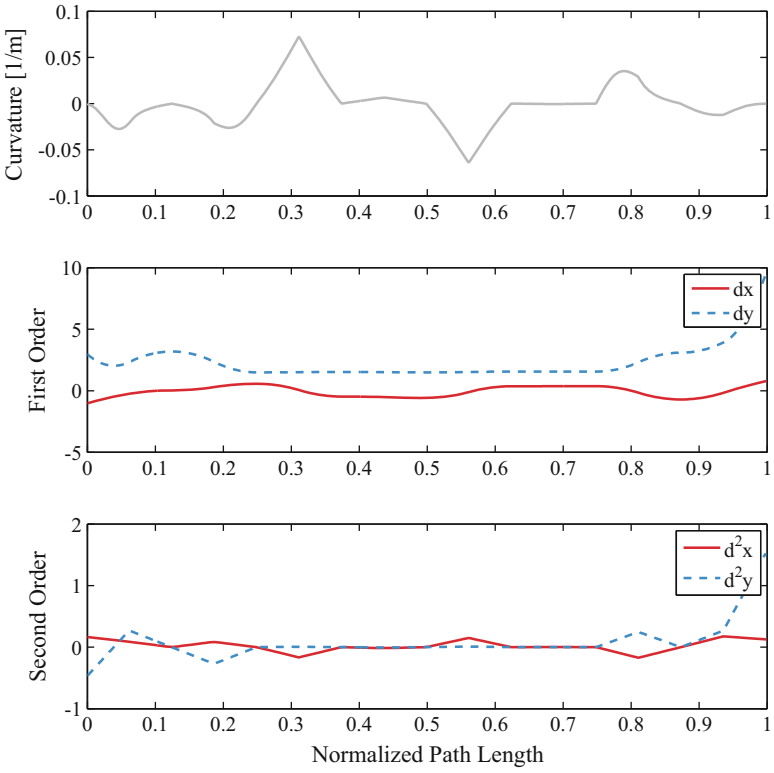



**Fig. 13.20** Example 3: Resulting path maintains parametric continuity

## 13.7 Conclusion

An approach to continuous curvature robot path smoothing, which satisfies the maximum curvature bounds and parametric continuity, is presented here. B-spline curves have been proposed for this task. In this chapter we offer the following contributions:

- Maintaining path parametric $C^2$ continuity, by using a single B-spline curve segment with midpoint insertion, to generate more realistic robot paths (Pan et al. 2012). No limitations were posed on the number of control points, for the B-spline curve, enabling a more robust representation of the path, unlike the work in Nikolos et al. (2003) and Jolly et al. (2009).
- Two analytical solutions are offered, formulating the path curvature in terms of a predefined path segment's parameters. They modify the path to limit its curvature to the maximum kinodynamic curvature and satisfy the vehicle's constraints. Our previous publications presented an introduction to the more advanced solutions (Elbanhawi et al. 2015; Elbanhawi and Simic 2014b).

Based on the presented numerical and experimental results, we show that this approach:

- Improved accuracy of segment curvature evaluation.
- Accelerated segment curvature evaluation.
- Decreased path length compared to reference spline and linear path
- Decreased deviation from reference path
- Bounded curvature to desired value whilst maintaining parametric continuity.

The proposed method results in paths that lie within the convex hull of the linear path, with no undesirable oscillations in the path. This produced realistic commands with continuous velocity and acceleration.

This approach relies on smoothing a path defined by successive waypoints, which are generated by a planning algorithm. As presented here, smoothing is considered as a post-planning procedure. Consequently, obtaining an obstacle free smooth path cannot be guaranteed. In many cases, when collision is detected, replanning is required (Koyuncu and Inalhan 2008). Several researchers examine collision detection for parametric curves (Kwangjin and Sukkarieh 2010; Pan et al. 2012). Circumventing the need for replanning can be achieved by incorporating the smoothing process within the planning framework. The benefits of guiding the search by the reachability of the robot have been revealed (Shkolnik et al. 2009; Jaillet et al. 2011). In this context, the reachable set can be computed using the efficient curvature evaluation method presented here. Several comments have to be made regarding the presented contributions.

The benefit of solving $C^2$ parametric continuity problem with maximum curvature constraint and employing the results to mimic human steering is the possibility to combine this parameterization, within any planning framework, such as an RRT or A* algorithm for autonomous vehicles. We predict improvements in human comfort as a result of mimicking human steering, in addition to other claims made by researchers, with regard to continuous curvature paths.

The proposed midpoint insertion algorithm is used to simplify the smoothing algorithm; a more generalized approach would include the location of the inserted point, as a function of the segment angle, but the benefits of doing that are not clear. That could be the subject of other investigation.

In the practical implementation through experiments, we demonstrated that the closely following control polygon method is advantageous over the other algorithms. This benefit comes from the assumption that the path planning algorithm generates a collision free piecewise linear path, which is then used by our smoothing algorithm, as shown in Elbanhawi and Simic (2014b).The results can be developed within the context of a recently developed sampling-based algorithm (Elbanhawi and Simic 2014c), which employs efficient collision checking procedures.

We expect that the outcomes of presented research can be integrated within an efficient planning framework, in which the spline-parameterized motions feel natural to passengers and improve their comfort. Passenger comfort and natural paths are obviously subjective terms that require a large sample of human volunteers for validation. The promising simulations' results, presented here, will be followed

by field tests using prototype ground vehicles and UAVs. We plan to validate the concept of graceful motions and curvature continuity, with regard to passenger comfort, by conducting full-scale field experiments.

**Acknowledgements** M. Elbanhawi acknowledges the financial support of the Australian Postgraduate Award (APA) and Research Training Scheme (RTS).

## References

Alshaer, B. J., Darabseh, T. T., & Alhanouti, M. A. (2013). Path planning, modeling and simulation of an autonomous articulated heavy construction machine performing a loading cycle. *Applied Mathematical Modelling, 37*(7), 5315–5325. doi:http://dx.doi.org/10.1016/j.apm.2012.10.042

Anderson, E. P., Beard, R. W., & McLain, T. W. (2005). Real-time dynamic trajectory smoothing for unmanned air vehicles. *IEEE Transactions on Control Systems Technology, 13*(3), 471–477. doi:10.1109/TCST.2004.839555.

Asano, T., Asano, T., Guibas, L., Hershberger, J., & Imai, H. (1985, October 21–23). Visibility-polygon search and euclidean shortest paths. In *26th annual symposium on Foundations of Computer Science* (pp. 155–164). doi:10.1109/SFCS.1985.65

Balkcom, D. J., & Mason, M. T. (2002). Time optimal trajectories for bounded velocity differential drive vehicles. *The International Journal of Robotics Research, 21*(3), 199–217. doi:10.1177/027836402320556403.

Barsky, B. A., & Derose, T. D. (1990). Geometric continuity of parametric curves: constructions of geometrically continuous splines. *IEEE Computer Graphics and Applications, 10*(1), 60–68. doi:10.1109/38.45811.

Berglund, T., Brodnik, A., Jonsson, H., Staffanson, M., & Soderkvist, I. (2010). Planning smooth and obstacle-avoiding B-spline paths for autonomous mining vehicles. *IEEE Transactions on Automation Science and Engineering, 7*(1), 167–172. doi:10.1109/TASE.2009.2015886.

Biswas, S., & Lovell, B. C. (2008). *Bézier and splines in image processing and machine vision*. Berlin: Springer.

Brezak, M., & Petrovic, I. (2013). Real-time approximation of clothoids with bounded error for path planning applications. *IEEE Transactions on Robotics, 99*, 1–9. doi:10.1109/TRO.2013.2283928.

Brooks, R. A., & Lozano-Perez, T. (1985). A subdivision algorithm in configuration space for find path with rotation. *IEEE Transactions on Systems, Man and Cybernetics, SMC-15*(2), 224–233. doi:10.1109/TSMC.1985.6313352.

Campion, G., Bastin, G., & Dandrea-Novel, B. (1996). Structural properties and classification of kinematic and dynamic models of wheeled mobile robots. *IEEE Transactions on Robotics and Automation, 12*(1), 47–62. doi:10.1109/70.481750.

Canny, J. A. (1985). Voronoi method for the piano-movers problem. In *Proceedings of IEEE international conference on robotics and automation* (pp. 530–535), March 1985. doi:10.1109/ROBOT.1985.1087297

Cheng, M. Y., Tsai, M. C., & Kuo, J. C. (2002). Real-time NURBS command generators for CNC servo controllers. *International Journal of Machine Tools and Manufacture, 42*(7), 801–813. doi:http://dx.doi.org/10.1016/S0890-6955(02)00015-9

De Boor, C. (1972). On calculating with B-splines. *Journal of Approximation Theory, 6*(1), 50–62.

Donges, E. (1978). A two-level model of driver steering behavior. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 20*(6), 691–707. doi:10.1177/001872087802000607.

Donghyun, K., Cheongjae, J., & Frank, C. P. (2014). Kinematic feedback control laws for generating natural arm movements. *Bioinspiration & Biomimetics, 9*(1), 016002.

Dyllong, E., & Visioli, A. (2003). Planning and real-time modifications of a trajectory using spline techniques. *Robotica, 21*(05), 475–482. doi:10.1017/S0263574703005009.

Elbanhawi, M., & Simic, M. (2012). Robotics application in remote data acquisition and control for solar ponds. *Applied Mechanics and Materials, 253–255*, 705–715. doi:10.4028/http://www.scientific.net/AMM.253-255.705.

Elbanhawi, M., & Simic, M. (2014a). Examining the use of B-splines in parking assist systems. *Applied Mechanics and Materials, 490–491*(1), 1025–1029. doi:10.4028/http://www.scientific.net/AMM.490-491.1025.

Elbanhawi, M., & Simic, M. (2014b). Sampling-based robot motion planning: A review. *Access, IEEE, 2*, 56–77. doi:10.1109/ACCESS.2014.2302442.

Elbanhawi, M., & Simic, M. (2014b). Randomised kinodynamic motion planning for an autonomous vehicle in semi-structured agricultural areas. *Biosystems Engineering, 126*(0), 30–44. doi:http://dx.doi.org/10.1016/j.biosystemseng.2014.07.010

Elbanhawi, M., Simic, M., & Jazar, R. (2015). Continuous path smoothing for car-like robots using B-spline curves. *Journal of Intelligent and Robotic Systems, 80*(1), 23–56. doi:10.1007/s10846-014-0172-0

Farin, G. (1992). From conics to NURBS: A tutorial and survey. *IEEE Computer Graphics and Applications, 12*(5), 78–86. doi:10.1109/38.156017.

Farin, G. (2002). *Curves and surfaces for CAGD. Computing*. San Francisco: Morgan Kaufmann.

Fraichard, T., & Scheuer, A. (2004). From Reeds and Shepp's to continuous-curvature paths. *IEEE Transactions on Robotics, 20*(6), 1025–1035. doi:10.1109/TRO.2004.833789.

Ghilardelli, F., Lini, G., & Piazzi, A. (2014). Path Generation Using η4-splines for a truck and trailer vehicle. *IEEE Transactions on Automation Science and Engineering, 11*(1), 187–203. doi:10.1109/TASE.2013.2266962.

Guarino Lo Bianco, C. (2013). Minimum-jerk velocity planning for mobile robot applications. *IEEE Transactions on Robotics, 29*(5), 1317–1326. doi:10.1109/TRO.2013.2262744.

Gulati, S., & Kuipers, B. (2008, May 19–23). High performance control for graceful motion of an intelligent wheelchair. In *IEEE international conference on robotics and automation, ICRA 2008* (pp. 3932–3938). doi:10.1109/ROBOT.2008.4543815

Gulati, S., Jhurani, C., Kuipers, B., & Longoria, R. (2009, October 10–15). A framework for planning comfortable and customizable motion of an assistive mobile robot. In *IEEE/RSJ international conference on intelligent robots and systems, IROS 2009* (pp. 4253–4260). doi:10.1109/IROS.2009.5354172

Hart, P. E., Nilsson, N. J., & Raphael, B. (1968). A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics, 4*(2), 100–107. doi:10.1109/TSSC.1968.300136.

Hughes, T. J. R., Reali, A., & Sangalli, G. (2008). Duality and unified analysis of discrete approximations in structural dynamics and wave propagation: Comparison of p-method finite elements with k-method NURBS. *Computer Methods in Applied Mechanics and Engineering, 197*(49–50), 4104–4124. doi:http://dx.doi.org/10.1016/j.cma.2008.04.006

Huh, U. Y., & Chang, S. R. (2014). A $G^2$ continuous path-smoothing algorithm using modified quadratic polynomial interpolation. *International Journal of Advanced Robotic Systems, 11*(25). doi:10.5772/57340

Jaillet, L., Hoffman, J., van den Berg, J., Abbeel, P., Porta, J. M., & Goldberg, K. (2011, September 25–30). EG-RRT: Environment-guided random trees for kinodynamic motion planning with uncertainty and obstacles. In *IEEE/RSJ international conference on intelligent robots and systems (IROS)* (pp. 2646–2652). doi:10.1109/IROS.2011.6094802

Jazar, R. N. (2008). *Vehicle dynamics: Theory and application*. Berlin: Springer.

Johnson, J., & Hauser, K. (2012, May 14–18). Optimal acceleration-bounded trajectory planning in dynamic environments along a specified path. In *IEEE international conference on robotics and automation (ICRA)* (pp. 2035–2041). doi:10.1109/ICRA.2012.6225233

Jolly, K. G., Sreerama Kumar, R, & Vijayakumar, R. (2009). A Bezier curve based path planning in a multi-agent robot soccer system without violating the acceleration limits. *Robotics and Autonomous Systems, 57*(1), 23–33. doi:http://dx.doi.org/10.1016/j.robot.2008.03.009

Jones, J., & Adamatzky, A. (2014). Material approximation of data smoothing and spline curves inspired by slime mould. *Bioinspiration & Biomimetics, 9*(3), 036016. doi:10.1088/1748-3182/9/3/036016.

Kavraki, L. E., Svestka, P., Latombe, J. C., & Overmars, M. H. (1996). Probabilistic roadmaps for path planning in high-dimensional configuration spaces. *IEEE Transactions on Robotics and Automation, 12*(4), 566–580. doi:10.1109/70.508439.

Khatib, O. (1986). Real-time obstacle avoidance for manipulators and mobile robots. *International Journal of Robotics Research, 5*(1), 90–98. doi:10.1177/027836498600500106.

Koren, Y., & Borenstein, J. (1991, April 9–11). Potential field methods and their inherent limitations for mobile robot navigation. In *IEEE international conference on proceedings of robotics and automation* (pp. 1398–1404, Vol. 1392). doi:10.1109/ROBOT.1991.131810

Koyuncu, E., & Inalhan, G. (2008, September 22–26). A probabilistic B-spline motion planning algorithm for unmanned helicopters flying in dense 3D environments. In *IEEE/RSJ international conference on intelligent robots and systems, IROS 2008* (pp. 815–821). doi:10.1109/IROS.2008.4651122

Kunz, T., & Stilman, M. (2013). Time-optimal trajectory generation for path following with bounded acceleration and velocity. *Robotics: Science and Systems*, 209.

Kwangjin, Y. (2013, May 8–31). An efficient Spline-based RRT path planner for non-holonomic robots in cluttered environments. In *International conference on unmanned aircraft systems (ICUAS)* (pp. 288–297). doi:10.1109/ICUAS.2013.6564701

Kwangjin, Y., & Sukkarieh, S. (2010). An analytical continuous-curvature path-smoothing algorithm. *IEEE Transactions on Robotics, 26*(3), 561–568. doi:10.1109/TRO.2010.2042990.

Kwangjin, Y., Jung, D., & Sukkarieh, S. (2013). Continuous curvature path-smoothing algorithm using cubic Bezier spiral curves for non-holonomic robots. *Advanced Robotics, 27*(4), 247–258. doi:10.1080/01691864.2013.755246.

Land, M., & Horwood, J. (1995). Which parts of the road guide steering? *Nature, 377*(6547), 339–340.

Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature, 369*(6483), 742–744.

Lapierre, L., Zapata, R., & Lepinay, P. (2007). Combined path-following and obstacle avoidance control of a wheeled robot. *The International Journal of Robotics Research, 26*(4), 361–375. doi:10.1177/0278364907076790.

Lau, B., Sprunk, C., & Burgard, W. (2009, October 10–15). Kinodynamic motion planning for mobile robots using splines. In *IEEE/RSJ international conference on intelligent robots and systems, IROS 2009* (pp. 2427–2433). doi:10.1109/IROS.2009.5354805

LaValle, S. M. (1998). *Rapidly-exploring random trees: A new tool for path planning* (Computer Science Dept.), Iowa State University, 98–11

Lentink, D. (2014). Bioinspired flight control. *Bioinspiration & Biomimetics, 9*(2), 020301. doi:10.1088/1748-3182/9/2/020301.

Ma, W., & Kruth, J. P. (1998). NURBS curve and surface fitting for reverse engineering. *International Journal of Advanced Manufacturing Technology, 14*(12), 918–927. doi:10.1007/BF01179082.

MacAdam, C. C. (1981). Application of an optimal preview control for simulation of closed-loop automobile driving. *IEEE Transactions on Systems, Man and Cybernetics, 11*(6), 393–399. doi:10.1109/TSMC.1981.4308705.

Maekawa, T., Noda, T., Tamura, S., Ozaki, T., & Machida, K-I. (2010). Curvature continuous path generation for autonomous vehicle using B-spline curves. *Computer-Aided Design, 42*(4), 350–359. doi:http://dx.doi.org/10.1016/j.cad.2009.12.007

Magid, E., Keren, D., Rivlin, E., & Yavneh, I. (2006, October 9–15). Spline-based robot navigation. In *IEEE/RSJ international conference on intelligent robots and systems* (pp. 2296–2301). doi:10.1109/IROS.2006.282635

McCrae, J., & Singh, K. (2009). Sketching piecewise clothoid curves. *Computers & Graphics, 33*(4), 452–461. doi:http://dx.doi.org/10.1016/j.cag.2009.05.006

McPhee, J. E., & Aird, P. L. (2013). Controlled traffic for vegetable production: Part 1. Machinery challenges and options in a diversified vegetable industry. *Biosystems Engineering*, *116*(2), 144–154. doi:http://dx.doi.org/10.1016/j.biosystemseng.2013.06.001

Meek, D. S., & Walton, D. J. (2004). An arc spline approximation to a clothoid. *Journal of Computational and Applied Mathematics*, *170*(1), 59–77. doi:http://dx.doi.org/10.1016/j.cam.2003.12.038

Montes, N, Herraez, A., Armesto, L., & Tornero, J. (2008, May 19–23). Real-time clothoid approximation by Rational Bezier curves. In *IEEE international conference on robotics and automation, ICRA 2008* (pp. 2246–2251). doi:10.1109/ROBOT.2008.4543548

Morro, A., Sgorbissa, A., & Zaccaria, R. (2011). Path following for unicycle robots with an arbitrary path curvature. *IEEE Transactions on Robotics, 27*(5), 1016–1023. doi:10.1109/TRO.2011.2148250.

Myung, H., Kuffner, J., & Kanade, T. (2007, April 10–14). Efficient two-phase 3D motion planning for small fixed-wing UAVs. In *IEEE international conference on robotics and automation* (pp. 1035–1041). doi:10.1109/ROBOT.2007.363121

Nikolos, I. K., Valavanis, K. P., Tsourveloudis, N. C., & Kostaras, A. N. (2003). Evolutionary algorithm based offline/online path planner for UAV navigation. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 33*(6), 898–912. doi:10.1109/TSMCB.2002.804370.

Ogata, K. (2010). *Modern control engineering*. Upper Saddle River: Prentice Hall.

Pan, J., Zhang, L., & Manocha, D. (2012). Collision-free and smooth trajectory computation in cluttered environments. *The International Journal of Robotics Research, 31*(10), 1155–1175. doi:10.1177/0278364912453186.

Piegl, L. A., & Tiller, W. (2001). Parametrization for surface fitting in reverse engineering. *Computer-Aided Design, 33*(8), 593–603. doi:http://dx.doi.org/10.1016/S0010-4485(00)00103-2

Pivtoraiko, M., & Kelly, A. (2011, September 25–30). Kinodynamic motion planning with state lattice motion primitives. In *IEEE/RSJ international conference on intelligent robots and systems (IROS)* (pp. 2172–2179). doi:10.1109/IROS.2011.6094900

Pivtoraiko, M., Knepper, R. A., & Kelly, A. (2009). Differentially constrained mobile robot motion planning in state lattices. *Journal of Field Robotics, 26*(3), 308–333.

Prokop, G. (2001). Modeling human vehicle driving by model predictive online optimization. *Vehicle System Dynamics, 35*(1), 19–53. doi:10.1076/vesd.35.1.19.5614.

Reeds, J. A., & Shepp, L. A. (1990). Optimal paths for a car that goes both forward and backward. *Pacific Journal of Mathematics, 145*(2), 367–393.

Roth, S., & Batavia, P. (2002, July 26–27). *Evaluating path tracker performance for outdoor mobile robots*. Paper presented at the Automation Technology for Off-Road Equipment, Chicago, Illinois, USA.

Sabelhaus, D., Röben, F., Meyer zu Helligen, L. P., & Schulze Lammers, P. (2013). Using continuous-curvature paths to generate feasible headland turn manoeuvres. *Biosystems Engineering, 116*(4), 399–409. doi:http://dx.doi.org/10.1016/j.biosystemseng.2013.08.012

Sachin, P., Jia, P., Abbeel, P., & Goldberg, K. (2014, August 3–5). *Planning curvature and torsion constrained ribbons in 3D with application to intracavitary brachytherapy*. Paper presented at the International Workshop on the Algorithmic Foundations of Robotics (WAFR), Boğaziçi University, İstanbul, Turkey.

Schoenberg, I. J. (1946). Contributions to the problem of approximation of equidistant data by analytic functions - A. On the problem of smoothing or graduations a 1st class of approximation formulae. *Quarterly of Applied Mathematics, 4*(1), 45–99.

Sgorbissa, A., & Zaccaria, R. (2010, October 18–22). 3D path following with no bounds on the path curvature through surface intersection. In *IEEE/RSJ international conference on intelligent robots and systems (IROS)* (pp. 4029–4035). doi:10.1109/IROS.2010.5653235

Shkolnik, A., Walter, M., & Tedrake, R. (2009, May 12–17). Reachability-guided sampling for planning under differential constraints. In *IEEE international conference on robotics and automation, ICRA '09.* (pp. 2859–2865). doi:10.1109/ROBOT.2009.5152874

Sungchul, J., & Taehoon, K. (2003, June 4–6). Tool-path generation for NURBS surface machining. In *Proceedings of the American control conference* (pp. 2614–2619, Vol. 2613). doi:10.1109/ACC.2003.1243471

Turner, M., & Griffin, M. J. (1999). Motion sickness in public road transport: The effect of driver, route and vehicle. *Ergonomics, 42*(12), 1646–1664. doi:10.1080/001401399184730.

Unser, M., Aldroubi, A., & Eden, M. (1993). B-spline signal processing. I. Theory. *IEEE Transactions on Signal Processing, 41*(2), 821–833. doi:10.1109/78.193220.

Velenis, E., & Tsiotras, P. (2008). Minimum-time travel for a vehicle with acceleration limits: Theoretical analysis and receding-horizon implementation. *Journal of Optimization Theory and Applications, 138*(2), 275–296. doi:10.1007/s10957-008-9381-7.

Walton, D. J., & Meek, D. S. (2005). A controlled clothoid spline. *Computers & Graphics, 29*(3), 353–363. doi:http://dx.doi.org/10.1016/j.cag.2005.03.008

Walton, D. J., Meek, D. S., & Ali, J. M. (2003). Planar G2 transition curves composed of cubic Bézier spiral segments. *Journal of Computational and Applied Mathematics, 157*(2), 453–476. doi:10.1016/s0377-0427(03)00435-7.

Wang, L. Z., Miura, K. T., Nakamae, E., Yamamoto, T., & Wang, T. J. (2001). An approximation approach of the clothoid curve defined in the interval [0, π/2] and its offset by free-form curves. *Computer-Aided Design, 33*(14), 1049–1058. doi:http://dx.doi.org/10.1016/S0010-4485(00)00142-1

Wu, W., Chen, H., & Woo, P.-Y. (2000). Time optimal path planning for a wheeled mobile robot. *Journal of Robotic Systems, 17*(11), 585–591. doi:10.1002/1097-4563(200011)17:11<585::AID-ROB1>3.0.CO;2-7.

Xuan-Nam, B., Boissonnat, J-D, Soueres, P., & Laumond, J. P. (1994, May 8–13). Shortest path synthesis for Dubins non-holonomic robot. In *Proceedings of IEEE international conference on robotics and automation* (pp. 2–7, Vol. 1). doi:10.1109/ROBOT.1994.351019

Yang, K., Moon, S., Yoo, S., Kang, J., Doh, N., Kim, H., et al. (2014). Spline-based RRT path planner for non-holonomic robots. *Journal of Intelligent and Robotic Systems, 73*(1–4), 763–782. doi:10.1007/s10846-013-9963-y.

Zhang, Y., Bazilevs, Y., Goswami, S., Bajaj, C. L., & Hughes, T. J. R. (2007). Patient-specific vascular NURBS modeling for isogeometric analysis of blood flow. *Computer Methods in Applied Mechanics and Engineering, 196*(29–30), 2943–2959. doi:http://dx.doi.org/10.1016/j.cma.2007.02.009

Zucker, M., Ratliff, N., Dragan, A. D., Pivtoraiko, M., Klingensmith, M., Dellin, C. M., et al. (2013). CHOMP: Covariant Hamiltonian optimization for motion planning. *The International Journal of Robotics Research, 32*(9–10), 1164–1193. doi:10.1177/0278364913488805.

# Index





R.N. Jazar, L. Dai (eds.), *Nonlinear Approaches in Engineering Applications*,
DOI 10.1007/978-3-319-27055-5