Real-time haptic simulation of medical procedures involving deformations and device-tissue interactions
Christian Duriez

To cite this version:

HAL Id: tel-00785118
https://tel.archives-ouvertes.fr/tel-00785118
Submitted on 12 Mar 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Real-time haptic simulation of medical procedures involving deformations and device-tissue interactions

Prepared by Christian Duriez for the
Habilitation à diriger des recherches

Defended the 1st of February 2013

Committee:

Reviewers: Ming LIN Professor at the University of North Carolina, USA
           Cagatay BASDOGAN Professor at the Koc University, Istambul, Turkey
           Yohan PAYAN Research Director at CNRS, Grenoble

Examiners: Marco VICECONTI Professor at the University of Sheffield, UK
           Jean-Pierre RICHARD Professor at école Centrale de Lille
           Michel COSSON Professor at the University of Lille 2

Advisor: Stéphane COTIN Research Director at INRIA Lille and Strasbourg

President: Michel DE MATHELIN Professor at the University of Strasbourg
ACKNOWLEDGEMENTS

First, I dedicate this manuscript to Frédérique who brought me constant support during the writing, despite crazy schedules and to my three children Anatole, Siméon and Augustine.

Many thanks to all the people who led me to the research field and to the doctorate: Dominique Lamy, Christophe Chaillou, Patrick Dubois, Abderrahmane Kheddar, Claude Andriot and Frédéric Dubois.

Many thanks to Stéphane Cotin and Steeve Dawson for hosting me in their dream team at CIMIT Boston during my postdoc, and to all the people with whom I worked there.

Many thanks to Christophe Chaillou and to the whole ALCOVE team for the warmly welcome upon my arrival at INRIA in 2006. Special thanks to Laurent Grisoni for his help and for giving me the opportunity to be Ph.D. co-advisor with him.

Many thanks to Stéphane Cotin for sharing his passion of medical simulation, for all the crazy ideas and the thousands of projects! Thanks to him for getting me involved, upon his arrival at INRIA, in the creation of the team SHAMAN → S.H.A.M.A.N → SHACRA! Many thanks to the all the former and present members of the team. Many thanks to all the Jérémies of the team, the one who will be the boss soon and the one who is sharing my office with my ICAM-godfather that I thank too ! Many thanks to Anne Rejl for her help in the administrative jungle ! Thanks to each of you guys for your ideas, your support, your patience... and instead of reading this paragraph, have a look at the perspectives at the end of the document and go back to work because there is still a lot of work to do ;-) !

Many thanks to Max Dauchet and David Simplot-Ryl, former and present director of the INRIA center of Lille-North Europe for their constant support in the process of creating the SHACRA team and in the development of SOFA-framework. Many thanks to Sylvain Karpf (who wrote the first grant for SOFA ;-) !) and to all the people I work with on a daily basis whether at INRIA, at the center of INRIA Lille Nord-Europe, at LIFL, at LAGIS, and at the école centrale de Lille.

Many thanks to Julien Lenoir for his proofread on a large part this manuscript and to Igor Peterlik for the haptic chapter.

Many thanks to the reviewers and to the HDR committee members.
## Summary

### Introduction

1. Mechanical deformable models for real-time computation
   1.1 Related Works
   1.2 Real-time integration of a dynamic deformable models
   1.3 Wire and tube model based on beam theory
   1.4 Surface model based on shell theory
   1.5 Volume deformation of soft-tissues

2. Constraint-based modeling of biomechanical interactions
   2.1 Related work
   2.2 Computing non-smooth mechanics in real-time
   2.3 Compliance computation
   2.4 Constraint-based contact response
   2.5 Surgical interaction modeled by complementarity constraints

3. Haptic rendering and multithreading approaches
   3.1 Related Work
   3.2 Constraint-based haptic rendering for medical simulation
   3.3 Asynchronous simulation of deformable models with variable stiffness

4. Applications, research projects and related activities
   4.1 Ongoing research projects
   4.2 Technology development and transfer based on SOFA framework
   4.3 Other research activities

Conclusion and perspective

Publication List

Bibliography
INTRODUCTION

Medical interventions rely on technologies that have greatly advanced in recent years, thanks to a strong scientific effort in many disciplines. One major advance comes from the intensive use of medical images for both diagnosis, planning and guidance of interventions. Another advance that is emerging, assistance of robots during therapies such as the cyber knife\(^1\) that delivers radiotherapy while targeting the radiation on malignant tumors with a robotic arm, or the Da Vinci robot\(^2\) that allows for teleoperation of a laparoscopic minimally invasive surgery. Even if it is difficult to foresee what will be the future medical and surgical practices, we think that computer-based simulation will play a great role in the coming years.

The initial motivation for computer-based simulation of medical interventions was to address the problem of education and training of residents: It takes a long time to learn some procedures, that are more and more complex in a context where the risks need to be minimized. Simulations can offer a good learning in a safe environment, for acquiring basic skills and gradually improving the level of difficulty in order to, finally, be able to perform a full procedure. A database of rare pathologies can also be added to allow for full training of students during their residency.

Moreover, as the level of realism was increasing in these simulations, other motivations have also raised, which could increase the fields of use for simulation. If sufficiently predictive, patient data may be integrated the simulations for the planning of real complex interventions and/or for the final choices between possible cares. Simulation can also improve the guidance of real procedures that currently rely on imaging techniques that do not account for the deformations of the tissues. We also believe that simulation can help the control of surgical robots by predicting the deformations on the anatomy induced by a given trajectory of the robot.

Scientific objectives: For all these applications, we aim at simulating the procedures realized by the physician and their effect on the anatomical structures and tissues. A common element among the various targeted applications is the notion of interaction: the physicians (possibly through a robot) must be able to directly interact with the simulated environment. It implies that the simulations we develop are computed in real (or near real) time, and that the presence of a user in the loop is accounted for (through the use of dedicated hardware devices, haptic feedback and robust algorithms). Moreover, it requires to develop accurate models, that are able to reproduce the geometry of the anatomy, the biomechanical behavior of organs, the physiology of the patient, the interactions with the medical devices... Maintaining this real-time constraint while increasing the level of realism of the simulation so that it becomes predictive, is a huge scientific challenge.

\(^1\)http://www.cyberknife.com/
\(^2\)http://www.davincisurgery.com/davinci-surgery/
The work we have conducted in recent years, and which is described in this manuscript, is mainly focused on the improvement of numerical methods used in a real-time context.

For obtaining more realistic biomechanical models for soft-tissue simulation, we should use the type of methods that had proven their accuracy in computational mechanics. Thus, a good part of our work is to make possible the use of the finite element method in a real-time context. This is particularly challenging because finite element methods are known to be very computationally demanding. Moreover, we also know that much of the accuracy of these models depends on the numerical method associated with mechanical interactions between soft-tissues. Without a good treatment of these boundary conditions, precise modeling of the deformation is vain. Consequently, our research work is contributing to new precise, fast and generic methods based on mechanical constraints. Finally, haptic feedback is requiring strict real-time constraint at high refresh rates. We seek to improve the fidelity of haptic rendering by developing new methods dedicated to the control of haptic devices plugged to a computer-based medical simulation.

**Liver simulation example:** Throughout this manuscript, the simulation of the liver will be taken as the main example, to illustrate the concepts, algorithms and models. First, this choice is justified by the fact that we have done several works on this organ, for the research part of the SOFA project, for the European project Passport, leaded by IRCAD and an ongoing project in the context of the IHU of Strasbourg. Moreover, the liver is somehow, *the good example* as it covers several aspects (also illustrated on figure 1) on which we want to insist:

1. The liver is a complex organ composed by three main structures: the parenchyma (soft tissue, that have a hyper-visco-elastic behavior) is coupled with a dense vascular network (tree-like tubular structure with a stiffer behavior) and is surrounded by the Glisson capsule (a stiff collagenous membrane). It illustrates the need of relying on both volume, surface and line finite elements and the need of proposing coupling methods between them.

2. The boundary conditions are very important to model the deformations of the liver: on the upper part, the liver is compressed by the diaphragm on which it is attached through several ligaments including the falciform ligament. Moreover, it is linked to the stomach (also deformable) by the hepatogastric ligament. Hence, the liver is surrounded by the peritoneum (like most of the organs of the abdominal cavity) except on a large triangular surface called the bare area, through which the vena cava is in direct contact with the liver. Thus when studying the deformations of the liver, it can not be considered as an isolated organ. Our scientific contributions include advanced methods to model and compute the various mechanical interactions between soft tissues.

3. A wide variety of treatments exist for the various liver diseases, which generates
numerous possible models for mechanical interactions with medical instruments. For instance, there are huge challenges in modeling and simulating the contact and the grasping by the laparoscopy instruments or in modeling the suture during laparoscopy. The treatments inside the vascular network of the liver using catheters and guidewires also induce complex contact between these instruments and the vessel walls. Liver biopsy or other treatments by needle insertion are also very common, but are sometimes challenging for physicians. Modeling flexible needle insertion on liver tissues is also a great numerical challenge. We propose new approaches to provide physics-based simulation of these procedures.

4. Finally, in a context of learning the clinical practice, the simulation of the procedures on the liver often necessitates the use of haptic feedback. Indeed, during laparoscopy, as the visual feedback is only given through the use of an endoscope and a display on a remote control screen, the surgeons are used to feel with their hand, through the kinesthetic feedback. Moreover, during the insertion of a flexible needle, physicians are also used to feel the puncture and cutting forces. We contribute on algorithms that provide a high fidelity haptic rendering on our simulation tools.

Figure 1: Anatomy of the liver: (left) The liver is placed under the Diaphragm, close to the Stomach and the colon. (middle) Important network of blood vessels passes through its tissues. (right) The image illustrates the complexity of the boundary conditions: direct interactions with colon, gallbladder, stomach, pancreas will influence its behavior.

Outline of the manuscript: The work on modeling the deformations (related to the item 1. of this list) will be discussed in the chapter 1. The modeling of the interactions and the boundary conditions (item 2 and 3) is addressed in the works presented in the chapter 2. Haptic feedback algorithms dedicated to real-time medical simulations will be proposed in the chapter 3. Finally, if the liver is a good example, it is not the unique one. The last chapter (chapter 4) will illustrate other applications of interactive medical simulation. The manuscript ends with a conclusion and a discussion on the future works.
MECHANICAL DEFORMABLE MODELS FOR REAL-TIME COMPUTATION

Table of contents

1.1 Related Works ................................................................. 6
  1.1.1 The scientific problem .................................................. 7
  1.1.2 Curve and Surface models .......................................... 9
  1.1.3 Volume deformations .................................................. 12
1.2 Real-time integration of a dynamic deformable models .............. 16
  1.2.1 Time-stepping implicit integration: ................................. 17
  1.2.2 Quasi-dynamic consistent simulation ............................... 18
1.3 Wire and tube model based on beam theory ........................... 21
  1.3.1 Corotational beam model ............................................ 21
  1.3.2 Fast solver for real-time ............................................ 23
1.4 Surface model based on shell theory ................................ 25
  1.4.1 Shell model adapted to real-time simulation ..................... 25
  1.4.2 Shell model based on Bezier triangle geometry ................. 29
  1.4.3 Modeling of thin anatomical structures .......................... 30
1.5 Volume deformation of soft-tissues ................................ 33
  1.5.1 Multi-grid FEM solver for real-time simulation ................. 35
  1.5.2 Asynchronous Preconditioning techniques ....................... 38
  1.5.3 Vascularized organ model ......................................... 43
CHAPTER 1. MECHA. DEFORM. MODELS FOR REAL-TIME COMPUTATION

The accurate modeling of deformations holds a very important place in medical simulation. A large part of the realism of a simulation, in particular for surgical simulation, relies upon the ability to describe soft tissue response during the simulated intervention. Moreover, for simulating some procedures, like for interventional radiology, the surgical instruments (needle, catheter, coils, stents...) must also be modeled as deformable. Several approaches have been proposed over the past fifteen years to model solid’s deformations in real-time, usually based on the theory of elasticity and the finite element method (FEM).

These approaches, including the methods presented during my PhD Duriez (2004), were limited to linear models that can not capture deformations with large displacements. Thus, after my PhD, I decided to devote part of my research to robust models and algorithms for solving non-linear models of deformation, based on FEM, in real-time. This research, helped by the development of SOFA, has led to a framework of several FEM models adapted to their geometrical support (curve models, surface models, volume models). Our models are primarily built on corotational approaches, which are popular in computer graphics, in order to deal with the geometrical non-linearities of the deformations. For time integration, we mainly rely on implicit schemes, which remain stable with large time steps but ask for more intensive computation.

The first section of this chapter presents a short survey on the modeling of soft-tissue deformations for real-time applications. The choices on time integration will be discussed in the second section with a presentation of our contribution on the quasi-dynamic scheme. Then, we will present our contributions on several finite element models adapted for real-time simulation and user interactions. Finally, some contributions on numerical methods will be presented on the last section of the chapter. Illustrations of the achieved work will be shown on the example of the liver and liver procedures.

1.1 Related Works

This section presents some related work on modeling solid’s deformations in real-time. The goal, here, is not to browse all the (numerous) existing papers but to have an overview of the scientific issues, the existing models and algorithms. (We refer the reader to Nealen et al. (2006)*, for instance, for a broad survey on modeling solid deformations). To focus the discussion on a concrete case and justify our choices, we use the example of a liver simulation.
1.1. RELATED WORKS

1.1.1 The scientific problem

What do we need?

All solids are deformable! But the magnitude of the deformations depends on the external loads. That is why the choice of a mechanical model highly depends on its concrete case of use. For instance, for the real-time tracking of a tumor, in the context of radiotherapy, a rigid kinematic model for the liver is often used (see for instance Murphy (2004)). In the following, we will consider three main cases of use for the liver simulation, in which it is important to account for the deformations: (i) education: how to simulate a laparoscopic procedure, in real-time, in an interactive training context, (ii) planning of a real procedure: In the context of flexible needle insertion (biopsy, for instance), the simulation is used to help the choice of the best trajectory while targeting a given zone of the liver, (iii) per-operative guidance: the simulation of the deformations of the liver’s patient is computed and updated in real-time during a real procedure. The simulation can help the surgeon to visualize the motion of the liver more precisely or provide information for the control of surgical robots.

In these three contexts, the needs from the liver model are not completely the same: For the education purpose (i), the deformation must be realistic but not necessarily predictive, whereas the need for an absolute real-time execution is needed. In the context of planning (ii), the model must be accurate enough to be predictive. Its computation must be fast enough to be compatible with the medical pipeline but not necessarily real-time. In the context of pre-operative guidance (iii), the simulation must be both accurate and real-time but above all, the simulation needs to adapt to the real data obtained in real time on the patient (vision, medical imaging,....)

However, in these three cases, we need a physics-based model, that accounts for the non-homogeneities of the liver and interactions with surgical instruments. In the remainder of this section, we assume that we need to model the vessels of the liver, the liver capsule and also the deformations of surgical instrument like suture thread or flexible needle. In addition, we would like to make changes in the topology of the model, for example, in a resection case. Of course, we are looking for the best compromise between accuracy and computation time.

Macroscopic constitutive laws

Regarding the deformations, in our simulations, we aim at modeling the global behavior of tissues and medical devices during an intervention. A macroscopic point of view is thus used, so that it can be considered that the material is a continuum and that the properties of interest for the simulation (density, elasticity, etc...) vary continuously within the tissues (or the flexible devices). For most of the deformable surgical
devices (catheter, guide, flexible needles, implants...), obtaining the constitutive law is not really a research issue. These properties are known by most companies selling the devices or can be found using abacus, given the material of the device. Moreover, the deformations of these instruments are often quite small (so Hooke’s law is often sufficient).

A number of studies have attempted to measure these characteristics on tissues. For instance, indentation measures could be done like in Gerard et al. (2005) or by more complete measurements Rubod et al. (2008). But regarding the liver, it has been shown in Ottensmeyer et al. (2004) that the viscoelastic properties of this soft-tissue were altered by in vitro measurement (it was necessary to perfuse the liver to obtain accurate values) and problems of boundary conditions occur. These issues have led to the development of new instruments for in vivo soft tissue characterization (see, for instance, Ottensmeyer (2002), Schiavone et al. (2007) and Luboz et al. (2012)).

In addition, recent results in elastography are very promising: using ultrasound or magnetic resonance imaging, the dynamic response of the tissues to a known source of mechanical vibration is analyzed. Based on inverse problem resolution Eskandari et al. (2011) Salcudean et al. (2011), it can provides stiffness measurements but limited to small deformations. Thus, a 3D image is obtained with an evaluation of the elasticity (Young modulus) and the viscosity associated with each voxel. One current limitation is the difficulty to measure these mechanical properties on the whole organ: as the source of mechanical vibrations is often external, it does not reach deep regions of the anatomy.

Furthermore, even if macroscopic values are targeted, all body organs and tissues are constituted of cells. These cells are also composed of chemical components (elastin, collagen, water, ...), that play a role in this visco-elastic behavior of the tissues. Brieu et al. (2011) propose a new hyperelastic constitutive law based on multi-scale approaches to account for composition of the tissues.

Finally, to assess the models and the constitutive laws, and also to evaluate the potential simplifications that are needed to reach real-time constraints, some works have proposed to compare simulation results with well documented experiments, like the so-called ”truth-cube” Kerdok et al. (2003).

**Numerical integration methods for deformations**

The constitutive laws explained above describe the mechanical behavior on every material point of the deformable domain. They provide a mathematical formulation between the strain and stress tensors, defined as continuous fields inside the deformable media. But in simulation, an infinity of material points composing the deformable object can not be considered... There is a necessary step of discretization along with the definition of a given number of degrees of freedom (DoFs). Several methods exist to do this step, but the most popular and efficient one for the deformable models remains,
without any doubt, the finite element method (FEM) (see, for instance Przemieniecki (1985) Belytschko et al. (2000)) or derivative methods from FEM (XFEM Bordas et al. (2007), Galerkin methods or mesh-less approaches Nguyen et al. (2008)...) 

After the definition of an interpolation between these degrees of freedom, that is often build on a mesh of the domain, the FEM method is based on a spatial integration of the constitutive law and the strain derivatives, element by element. The problem is then assembled and solved using numerical methods which size is proportional to the number of DoFs. When used on well defined cases, the Finite Element Method converges to a unique solution. However, to reach this convergence, FEM often needs the use of many elements (and many DoFs). This method is consequently very demanding from a computational point of view and not easily compatible with real-time.

A first manner to reduce the complexity of the models is to benefit, when possible, from a particular geometry of the object (curvilinear objects like rods and tubes or surface objects like hollow organs or membranes).

1.1.2 Curve and Surface models

Wire-like structures

In the clinical context described above, we need to simulate medical instruments that behave like deformable rods and tubes (catheter, guide, flexible needle, suture..). The vessels contained in the liver can also be considered as deformable tubes, as well as the intestines that are close to the liver and that could be important to model, even with simplifications, in a liver simulation context (especially to provide accurate boundary conditions). The fact that the configuration of the object could be reduced to the deformation of a curve (eventually to the motion of a frame along a curve) allows to greatly reduce the complexity compared to a volume model.

There is numerous interesting related work on the modeling of rod structures! Again, the goal here is not to provide a complete survey but an overview of the existing models that can be adapted to our simulation problems. The reader can refer to Theetten et al. (2006) for a detailed discussion on existing models.

The main characteristics of wire-like structures include geometric non-linearities, high tensile strength and low resistance to bending. A first approach is to mimic this basic behavior using multibody dynamics model( Featherstone (1983)), like in Cotin et al. (2000) where a set of rigid elements are connected using spherical joints. More recent works have considered the use of angular springs. For instance Taskiran et Giidukbay (2005) have chosen to use a linear representation of angular springs to supply bending and torsion effects in hair modelling, which allowed them to simulate curly hair with good realism. Wang et al. (2007) successfully used a network of linear and angular springs to describe bending and twisting of catheters and guidewires in
From the geometrical point of view, splines are probably the most classical tool for animating these \textit{1D objects}. Assigning a physical behavior to splines has been introduced in Qin et Terzopoulos (1996). Nocent et Rémond (2001) defined the \textit{Dynamic Material Splines} (DMS), a full Lagrange-based simulation framework for splines. They considered spline control points as the degrees of freedom of the underlying continuous object. Lenoir et al. (2004a) introduced a curvature energy formulation for DMS that is not geometrically exact, but provides real-time manipulations, as well as adaptive simulations (see Lenoir et al. (2005)). Although real-time computation is possible, this model does not incorporate torsional energy terms. Using a background in mechanics consisting of elasticity and plasticity theories, Theetten et al. (2006) proposed a deformable model for one-dimensional objects: \textit{Geometrically Exact Dynamic Splines}. This approach addresses reversible and irreversible deformations, like stretching, twisting and bending, and can even detect fracture.

In robotics community, several recent works use Cosserat theory. Cosserat medium was first described in 1909 by Cosserat brothers Cosserat et Cosserat (1909). This medium is described by a set of oriented micro-solids. A first introduction of Cosserat’s rod theory in computer graphics to model cantilever objects is proposed in Pai (2002). A very accurate static solution of a cable simulation is proposed in Wakamatsu et Hirai (2004) by considering it as a succession of oriented frames and by minimizing its potential energy. Several works have proposed optimized solution of models based on Cosserat theory, like Grégoire et Schömer (2006), especially to include this model in the control loop of flexible robots, like in Torres et Alterovitz (2011).

In computational structure analysis, the use of beam theory is very popular Przemieniecki (1985). A virtual catheter model, based on a linear elasticity, was introduced by Nowinski et Chui (2001), using a set of finite beam elements. The beam theory is based on Kirchhoff rods hypothesis, that were used to develop the inextensible Super-Helices model Bertails et al. (2006). In general, the beam model provides accurate results but the quadratic complexity of the associated algorithm could be a key problem for real-time simulation, that we address in section 1.3.

Again, there is no direct conclusion for a model that would be perfectly suitable for what we want to simulate. Providing inextensible behavior in the model would be certainly a strength but it often necessitates the use of reduced coordinates (like the one used for the Cosserat model by Pai (2002) or the Super-Helices by Bertails et al. (2006)) that prevent from using band tri-diagonal (BTD) or band solvers that are faster. Otherwise, when using control points (spline models) or generalized coordinates frames (often used for beam models), inextension is provided by setting lagrange multipliers (so additional computations). But the advantage of not using reduced coordinates is that the behavior of any point on the curve only depends on a reduced number of DoFs, which is better for setting constraints with the environment. In the section 1.3, we present our approach, that mixes spline and beam approaches.
1.1. RELATED WORKS

Thin-structure deformation

The particularity of thin-object deformation is that their behavior can be abstracted through the motion of a surface. For the liver simulation, it can be used to model the mechanics of the capsule. The approach is suited for hollow organs like the stomach, the bladder... and also tubular organs like blood vessels, intestines... (it would replace wire-like approach, if more precision is needed: especially to get deformations that are normal to the curve.)

Numerous models are available in the literature to describe physics of thin objects, from fairly simple and naive approaches to more complex and thorough representations. Continuum mechanics provides several formulations able to accurately describe stresses occurring within thin objects. Most of them fall into one of the following two categories: plate theory or shell theory. Those theories have been a subject of interest in the mechanical community for decades. The difference between these two kinds of structures is very well explained by Liu et Wang (2003) and can be summarized by the fact that plate bending elements can only carry transversal loads while shells can undergo more complex deformations (see Figure 1.1).

![Figure 1.1: Illustration of a key difference between plate theory and shell theory: (Left) A typical deformation of the plate theory is illustrated: the thin object is deformed by transversal loads like its own weight, or a normal pressure on the surface. (Right) Conversely, a shell structure can carry loads in all directions, and therefore can undergo bending, twisting and in-plane deformation.](image)

Development of a satisfactory physical model that runs in real-time but produces visually convincing animation of thin objects has been a challenge in Computer Graphics, particularly in the area of cloth modelling. Rather than resorting to shell theory which involves the most complex formulations in continuum mechanics, previous works have often relied on discrete formulations. Early approaches to cloth modelling only considered in-plane deformation, and often relied on mass-spring models (see Provot (1995), Hammer et al. (2008) or Yu et Geng (2008) for instance).

Yet, such models are limited in their ability to describe certain behaviour, but do not integrate a constitutive law of a material. Another limitation of such models is the difficulty to derive spring stiffness (in particular for angular springs) from elastic properties (Young’s modulus and Poisson’s ratio). For these reasons, other approaches have been proposed. Among the different models introduced recently we can mention the work of Choi et al. (2007) and Bridson et al. (2003). Choi et al. proposed a
real-time simulation based on thin shells undergoing large deformations. The authors adopt the energy functions from the discrete shells proposed by Grinspun et al. (2003). For real-time integration of the governing equation, they adapted a modal analysis technique, called modal warping. The resulting simulations run in real-time even for large meshes, and the model can handle large bending and/or twisting deformations with acceptable realism. Bridson et al. followed the same energy approach to derive their bending model but improved the resolution of the equations by suggesting a novel mixed implicit/explicit integration scheme. They also presented a post-processing method for treating cloth-character collisions that preserves folds and wrinkles. Pabst et al. improved the bending modelling used by Bridson et al. to allow the integration of measured material data. A method to use thin shell dynamics with point sampled surfaces for efficient animation was proposed by Wicke et al. (2005) where the curvature of the shell is measured through the use of fibers.

1.1.3 Volume deformations

Non-FEM methods for Volume deformable objects

As FEM is particularly computationally expensive for volume deformations, one of the popular methods for real-time computation of the deformation of soft objects is based on the use of mass-spring models (see Meseure et Chaillou (2000) or Montgomery et al. (2002), for instance). In addition, such an approach is well suited to benefit from GPU-based acceleration, as demonstrated by (Sorensen et al., 2006). Another obvious advantage of using mass-spring network is their ability to handle topological changes at a reduced computational cost. Yet, although quite simple to implement and very fast to compute, mass-spring methods fail to properly characterize soft tissues deformation as they introduce artificial anisotropy through the choice of the mesh, and also make it difficult to relate spring stiffness to material properties such as Young modulus. This is the main drawback of mass-spring approach and in the context of liver simulation that we set above, it would prevent us from using this model for both planning and per-operative guidance. However, some recent work by Delingette (Delingette, 2008) has shown that such a relationship can be established in the case of hyperelastic constitutive laws. Using this approach, it is possible to compute spring’s stiffness on triangular and tetrahedral meshes relating to St Venant-Kirchhoff materials (one of the simplest hyperelastic law).

Another alternative is to rely on the shape matching approach presented in Müller et al. (2005): elastic forces are applied to the model to drive it towards its original shape. To conserve momentum, the rotation of the body is estimated using polar decompostion.
1.1. RELATED WORKS

Linear models and other precomputed methods

There are several instances in the literature of real-time simulation of deformable object which rely on linear elasticity. Initial work (see Bro-Nielsen et Cotin (1996), Cotin et al. (1999) or James et Pai (1999)) made the assumption of small displacements and often relied on precomputed responses based on Finite Element Models (FEM) and related techniques. The main benefit of approaches based on pre-computation is the important speed up that can be obtained, allowing not only real-time deformations but also haptic feedback. The first limitation is inherent to the small strain assumption: to keep a linear model, the rate of deformation must be very small (too small to be applicable on the liver case). Some methods have been proposed to extend the idea of pre-computation, but in the case of geometrically non-linear deformations by sampling a space of possible interactions, like in (Mahvash et Hayward, 2004) and could be applicable to our problem. The second issue is the impact of cutting on the pre-computed response. We listed in the specifications of the liver model, that we would need to simulate resection. While keeping a pre-inverted approach, some solutions have been proposed to solve the problem, using for instance Sherman-Morrison technique to update a pre-inverted stiffness matrix (Lee et al., 2005). These solutions are limited to linear deformable model: An update of a pre-computed non-linear model in case of cutting is, for now, too costly.

Model reduction

Some work aims at reducing explicitly the complexity of the deformable models. The idea is to reduce the number of unknown on the deformable system.

The first strategy is actually to reduce, as much as possible the degrees of freedom of the deformable model. It can be done by setting a skeleton inside the model to drive the deformations, like in Capell et al. (2002a). Skinning techniques could be used to model the deformations from a set of frames (see, for instance Kavan et Žará (2005) Rohmer et al. (2009)). In Gilles et al. (2011) and Faure et al. (2011), a skinning technique is incorporated in a meshless approach that uses a very small number of DoFs.

Another way of reducing the model is to use polar decomposition: it consists in using the most important eigenvalues of the mechanical system as degrees of freedom Basdogan (2001). These eigenvalues corresponds to the modes of vibration of the deformable object. Then the deformation can be computed in the sub-domain of the main deformation modes (like a low-pass filter). This type of decomposition in sub-domain was extended in Barbič et James (2005) for geometrical non-linearities and in An et al. (2008) for material non-linearities.

These strategies are particularly suitable to reduce the computation time. One difficulty is to guaranty that the model will remain valid when mechanical interactions
complex contact situations, cutting) will be applied. That is why some methods like in Faure et al. (2011) include a possible resampling of the DoFs, that is not - for now - automatic.

Explicit non-linear FEM

Another strategy to deal with real-time computation of the liver tissue deformations would be to base the computation on an explicit integration scheme, as proposed by Picinbono et al. (2000) or (Taylor et al., 2008) for instance. The main advantage is that the solving process only involves the mass matrix, which is diagonal if mass lumping is used. Very complex finite element model can be used, as no derivation of the internal forces is needed. Thus, the equations of motion can be decoupled and each degree of freedom can be solved independently. The resolution is then very quick and its parallelization is quite straightforward (Comas et al., 2008). Explicit integration methods are particularly well suited for some applications like the real-time non-rigid registration of the brain shift during surgery (Joldes et al., 2009a). In this case, only the steady state of the deformation is sought and the mass can be artificially increased in order to deal with stiff materials (Joldes et al., 2009b).

However, in the cases of use we have listed for the liver simulation, we aim at using the model in an interaction context. In this context, the transient states are also needed and the mass matrix can not be scaled any more. Moreover, liver tissues are often very soft but in most of pathologic cases, these tissues are much stiffer. In addition, for flexible needle insertion simulation, we also need to simulate the needle, which is particularly light and stiff. Given these type of ratio between mass and stiffness, the critical time step would be very small (typically $10^{-3}$ to $10^{-6}$ sec, sometimes less) and would limit the range of interactions that can be proposed to the user in realtime. For instance, only simple collision detection and response (i.e. limited to a rigid surface) can be handled in real-time. But more complex collision like between deformable models takes too much time to be handled in real-time with explicit schemes (it takes at least several msec for our method (Allard et al., 2010a) which is one of the fastest already published). Implicit integration allows the use of larger time steps (about $3.10^{-2}$ sec) without any stability issues even with complex user interactions. The counterpart is a heavier computation that needs to be processed at each time step.

We emphasize that to our opinion, the choice of the integration scheme highly depends on the application. In the SOFA framework, both explicit (Comas et al., 2008) and implicit (Faure et al., 2012) schemes are proposed. The issues regarding the integration schemes in the context of real-time application will be discussed in more details in section 1.2.
Fast implementation and efficient solvers for implicit FEM

Finding a solution for implicit deformable model based on FEM often amounts to solve a large system of nonlinear equations. Using Newton-Raphson approach, it could be processed by solving a series of sparse linear system. Then, efficient linear solvers are a key to obtain fast simulations.

In the literature, two families of algorithms are proposed: direct and iterative methods. First, direct solvers provide the solution in a fixed number of steps, initially by computing the actual inverse of the system matrix Bro-Nielsen et Cotin (1996), or more recently by creating a factorization that can then be used to compute the solution Barbić et James (2005). These methods are often too costly to be applied at each time-step except for small models (less than 300 nodes). They are also often used in combination with an approach to reduce the number of degrees of freedom of the model, either using condensation on surface nodes Bro-Nielsen et Cotin (1996), or reduced-coordinate models Barbić et James (2005).

The second class of methods is iterative algorithms Barrett et al. (1994), which starts from an initial estimate and iteratively refines it to approach the exact solution. The most popular algorithm is the Conjugate Gradient (CG) algorithm Shewchuk (1994), that is often used in works related to interactive simulation, like in Teschner et al. (1999), Otaduy et Gross (2007), Irving et al. (2007). Although in theory up to \( n \) iterations are necessary to achieve convergence (where \( n \) is the number of equations), in practice it is possible to stop the algorithm much sooner, and allows for a simple way of having a tradeoff between accuracy and computation time. However, while visually the result can appear sufficiently realistic, a premature end of the iterative process can introduce artificial damping. Also, the convergence rate can be slow for ill-conditioned problem (such as non-homogeneous objects). Despite the above limitations, the CG algorithm proved very efficient in most applications. It is also rather easy to implement, as it relies on only three basic operations: matrix–vector products, vector–vector inner products, and linear combination of vectors. A very interesting property of these operations is their inherent support for parallelism. Several works presented parallel implementations on CPU Parker et O’Brien (2009) Hermann et al. (2009)* and GPU Bolz et al. (2003) Buatois et al. (2007) and Allard et al. (2011a).

FEM is supposed to converge to a unique solution, so a fast computation on a coarse mesh could provide a first ”good” solution, that can be refined for more precise results. This strategy can be exploited by adaptive solvers, that add details using refinement functions on specific area of the model (see, for instance Grinspun et al. (2002) Capell et al. (2002b)). Thus the number of degrees of freedom is reduced for the same precision. However, these methods need an heuristic (for instance based on strain measurement), to find where to adapt the model. When the model undergoes large global deformation, the heuristic may activate all the refinement basis functions, and lead to heavy computations. In the opposite, multigrid solvers, presented for instance in Harrison et richard K. Gordon (1996) or in Georgii et Westermann (2006) take
full advantage of the hierarchy of grids to reduce solving time. An iterative process computes a correction at the coarser levels, taking into account the residual at the finest level. This correction is then transferred to the fine level. The majority of the iterations are performed at the coarse level (at low computation cost) and only a few iterations are performed at the finest level, which leads to a strong improvement of the performance.

1.2 Real-time integration of a dynamic deformable models

Deformable models are often modeled as second order mechanical systems with non-linear equations. Moreover, these models are often constrained by complex boundary conditions, like we mentioned previously for the liver. This section explains the numerical choices we make concerning the time integration of these complex systems in the context of a real-time simulation.

The first part is not a real contribution, but a presentation of the implicit scheme we have often adopted, that enables both good accuracy and good adaptation to the problems linked to the interactive context. In the second part, we briefly present a contribution, called ”Quasi-dynamic” that was presented in Theetten et al. (2007). This method allows to keep a consistent interactive simulation when the computation time is critical.

Compared to classical simulations of deformable models, our goal is to allow a virtual interaction (as transparent as possible) with the models. Indeed, in the liver simulation context we defined above, the physician should be able to virtually touch or manipulate liver through instruments and change the course of the simulation. There is a user is in the loop. Consequently, we do not have a perfect control of the boundary conditions, the intensity of the solicitations, the delays between the real gesture and its virtual execution. In this context, implicit integration scheme provides unconditional stability, with the main drawback of obtaining a large system of equations that need to be solved quickly.

In the context of interactive physics-based simulations, it is often admitted that only a few milliseconds are available to compute the next state. Or, at least, it is assumed that a high update frequency is required to produce the feeling of interactivity. To introduce this section, let’s examine more closely this problem of update frequency when a user interacts with the simulation.

Let’s consider this simple example: a user can touch a dynamic deformable model through the use of a rigid probe. Whatever numerical scheme is used to integrate the second order differential system of equations, a time step $h$ is chosen. At each step, the position of the rigid probe is measured before integration and a resolution with the deformable mechanical system is performed. If the computation time $t_c$ of one
1.2. REAL-TIME INTEGRATION OF A DYNAMIC DEFORMABLE MODELS

step exceeds \( h \) (for instance \( t_c = 2h \)), the (real) time elapsed between two successive measured positions is equal\(^1\) to \( t_c \), whereas the simulated time between these two positions is only \( h \). It means that the simulated velocity of the probe is twice faster in the simulation than it is in reality! This creates an artificial excessive kinetic energy that may completely distort the results!

So, we can state that the real-time constraint for an interactive simulation is primarily a problem of accuracy!

We often present our choices as a tradeoff between accuracy and real-time performance but, as explained above, we can finally justify them by a single problem of accuracy... Even if we need to simplify the model to make it real-time, the error introduced by these simplifications should be put in perspective with the errors that would be caused by the failure of real-time. In our choice of time integration, we prefer schemes that support constant and large time steps while maintaining the computations as simple as possible. These reasons led us to use low order schemes, like implicit Euler, coupled with a principle of a unique linearization per time step.

1.2.1 Time-stepping implicit integration:

Let’s consider a generic dynamic deformable model. Equations used to model the dynamic behavior of bodies have led to a synthetic formulation, given by the Newton’s second law:

\[
\dot{\mathbf{v}} = \mathbf{M}(\mathbf{q}) \mathbf{v} = \mathbf{P}(t) - \mathbf{F}(\mathbf{q}, \mathbf{v}) + \mathbf{H}^T \lambda
\]

where \( \mathbf{q} \in \mathbb{R}^n \) is the vector of generalized degrees of freedom (for instance, displacement of a mesh), \( \mathbf{M}(\mathbf{q}) : \mathbb{R}^n \rightarrow \mathbb{M}^{n \times n} \) is the inertia matrix, \( \mathbf{v} \in \mathbb{R}^n \) is the vector of velocity. \( \mathbf{F} \) represents internal forces applied to the simulated object depending on the current state and \( \mathbf{P} \) gathers external forces. \( \mathbf{H}^T \lambda \in \mathbb{R}^n \) is the vector of constraint forces contribution.

\( \mathbf{M}(\mathbf{q}) \) and \( \mathbf{F}(\mathbf{q}, \mathbf{v}) \) are derived from the physics-based deformable model. \( \mathbf{H} \) is the matrix containing the constraint directions and \( \lambda \) the vector of lagrange multipliers containing the constraint force intensities.

Implicit schemes provide several advantages, in particular improved stability with relatively large time steps. This is particularly relevant for interactive simulations involving contacts with virtual devices controlled by an operator.

Using time-stepping methods, the time step is fixed and there is no limitation on the number of discontinuity that could happen during a time step (Anitescu et al. (1999)), but low-order integration schemes should be used. This could lead to excessive dissipation if the time step is too large. However it provides stable simulations.

\(^1\) quasi equal would be more appropriate if we integrate the variable delays caused by the measure of the position itself.
Let’s consider the time interval \([t_i, t_f]\) which length is \(h = t_f - t_i\). We have:

\[
M(v_f - v_i) = \int_{t_i}^{t_f} (P(t) - F(q, v)) \, dt + h \mathbf{H}^T \lambda
\]  \hspace{1cm} (1.2)

\[
q_f = q_i + \int_{t_i}^{t_f} v \, dt
\]  \hspace{1cm} (1.3)

To evaluate integrals \(\int_{t_i}^{t_f} (P(t) - F(q, v, t)) \, dt\) and \(\int_{t_i}^{t_f} v \, dt\) we chose the following implicit Euler integration scheme:

\[
M(v_f - v_i) = h (P(t_f) - F(q_f, v_f)) + h \mathbf{H}^T \lambda_f
\]  \hspace{1cm} (1.4)

\[
q_f = q_i + hv_f
\]  \hspace{1cm} (1.5)

\(F\) is a non-linear function, we apply a Taylor series expansion to \(F\) and make the first order approximation:

\[
F(q_i + dq, v_i + dv) = f_i + \frac{\delta F}{\delta q} dq + \frac{\delta F}{\delta v} dv
\]  \hspace{1cm} (1.6)

Using \(dq = q_f - q_i = hv_f\) and \(dv = v_f - v_i\), we obtain:

\[
\begin{pmatrix}
M + h \frac{\delta F}{\delta v} + h^2 \frac{\delta^2 F}{\delta q^2} \\
\frac{\delta F}{\delta q}
\end{pmatrix} \begin{pmatrix} dv \\ x \end{pmatrix} = -h^2 \frac{\delta F}{\delta q} v_i - h (f_i + p_f) + h \mathbf{H}^T \lambda_f
\]  \hspace{1cm} (1.7)

where \(p_f\) is the value of function \(P\) at time \(t_f\). The only unknown values are the Lagrange multipliers \(\lambda\) but their computation is detailed in section 2.2. In the remainder of this section, we will refer to this system using matrix \(A\) and vector \(b\).

### 1.2.2 Quasi-dynamic consistent simulation

This subsection is an abstract of the contribution presented in Theetten et al. (2007) which is dedicated to precise and interactive simulation of mechanical splines. For this work, we introduce a new method based on a simple idea: for a coherent simulation of a spline, when it is not possible to do the computation in real-time, we switch from a dynamic model to a static model. In the following, we generalize the concept to any deformable model.

First, we will separate two different notions that are often mixed:

- An interactive simulation is a simulation that has a refresh rate that is sufficiently fast to allow a continuous and transparent interaction with the user (typically 30fps for a visual interaction).
- A real-time simulation is a simulation for which there is a notion of time in the models and the simulated time is the same than the execution time.
If the refresh rate of the computation is fast enough, we can have an interactive real-time simulation but, one can also imagine a real-time simulation with very long time step that would not be interactive.

Moreover, if there is a notion of time in the models computed in an interactive simulation (like dynamic models), real-time becomes a strong constraint. Indeed, as explained above\(^2\), if the simulation is not computed in real-time, some artifact can appear because the gesture of the user may appear to be artificially fast (or slow, depending on the case) compared to the simulated objects.

The aim of quasi-dynamic simulation is to propose a precise interactive mechanical simulation that follows these principles using a dynamic switch between dynamic and static models. Indeed, static models are 0-order differential equation systems, so they are not influenced by the time.

On the other hand, dynamic models provide dynamic transitions between rest states and they are even necessary to describe some physical behavior like oscillations. However, due to the use of an implicit scheme, these oscillations are damped when we are forced to use large time step because of the computation time. At a certain point, there is no interest of using a dynamic model anymore.

In our method, we propose to switch the simulation from dynamic to quasi-static or from quasi-static to dynamic, depending on the possibility (or the non-possibility) of fulfilling the real-time constraint (that has only a meaning in the case of a dynamic model).

**Dynamic-static switch:** From the model point of view, switching from dynamic equations to static equations consists, first, in suppressing the term \(\mathbf{M}(\mathbf{q})\dot{\mathbf{v}}\) in equation 1.1. When real-time response is not reached in the simulation, this inertial term will add artificial kinetic energy, as explained in the introduction of section 1.2. Second, we should also remove the dependency to velocity (so to time) in the internal force \(\mathbf{F}(\mathbf{q}, \mathbf{v}) \rightarrow \mathbf{F}(\mathbf{q})\), in order to get a pure static model. We notice that removing this dependency is less important since this term does not add artificial energy in the system, but actually, reduce it. Moreover, we could estimate the velocity with a measure of the computation time. Anyway, on the basis of a pure static system, the generic deformable model becomes:

\[
\mathbf{P} - \mathbf{F}(\mathbf{q}) + \mathbf{H}^T \lambda = 0 \tag{1.8}
\]

We use the Taylor series expansion (1.6) to obtain a unique linearization per simulation step. Note that we update the measure of the user’s tool position at each simulation step. This new position can change the external forces \(\mathbf{P}\) (for instance when using

\(^2\)see the introduction of section 1.2
penalty method) or the value of the constraints \( \lambda \) (when using constraint-based approaches).

\[
P - f_{i-1} + H^T \lambda = \frac{\delta F}{\delta q} dq
\]  

(1.9)

When the switch is performed from static to dynamic, the initial conditions (position and velocity) must be given. It means that, even during the static simulation, a velocity is evaluated. This evaluation is performed using the computation time \( t_c \) between two successive equilibrium state: \( v = dq/t_c \). However, the stiffness matrix \( \frac{\delta F}{\delta q} \) is singular. The Lagrangian constraints \( H \) suppress the singularity if they remove the rigid motion of the object. On the opposite, the mass matrix \( M(q) \) is always defined.

Consequently, we define two heuristic rules for a decision of a switch between static and dynamic: The first heuristic is based on the system singularity: if the rigid-body motion of the deformable object is not removed by constraints, we switch to dynamic model. The second heuristic is the temporal consistency: if the computation of the dynamic model can not fulfill real-time constraint, or if we detect that the dynamic model is over-damped by the use of too large time step (compared to the eigenfrequency of vibration of the deformable object), we switch to quasi-static model. This second heuristic is associated with an hysteresis to avoid a permanent switch between static and dynamic models.

**Discussion:** The quasi-dynamic method is highlighted in this manuscript because it is representative of specific problems we encounter when dealing with physics-based interactive simulation. For the liver simulation, it would be particularly suited for the per-operative context, in which both real-time and accuracy are needed. New but similar issues will be faced when providing haptic feedback on such simulations as detailed in chapter 3.
1.3 Wire and tube model based on beam theory

Medical procedures on the liver are a good illustration of the need of a *wire-like* model adapted to real-time simulation. Indeed, we developed a first model based on beam theory in the context of interventional radiology, for catheter and guidewire. This type of intervention is often realized for liver cancer treatments. The approach was associated with a fast solver described in Cotin et al. (2005). An extended version of the work was presented in Duriez et al. (2006a). In more recent work, the model was also used to model the surgical thread in Guebert et al. (2009) for suture (important for the simulation of laparoscopic surgery) and for flexible needles insertion in Duriez et al. (2009) (which can be used for liver biopsy simulation). Recently, we have proposed to couple the beam elements with volume model to simulate the effects of the vascularization on the liver deformation mechanics in Peterlik et al. (2012), as detailed in section 1.5.

These contributions are also closely linked to compliance computation in the constraint solving process, which is detailed in section 2.3.1.

This section focuses on a model for instruments or anatomical structures for which the length is greater than the other transverse dimensions. Examples of objects with this geometry are abundant in medical simulation. The particular nature of such objects generally leads to large geometric deformations, which is notorious for being computationally demanding. Our approach, initially based on a linear beam analysis, extends this representation by a series of optimizations particularly suited for real-time animation. By using a corotational approach, our model can handle the important geometric non-linearity due to large changes in the shape of the object. A substructure analysis allows further optimization of the solver for collision or self-collision response. Finally, we present the method used to couple the mechanical behavior of the liver vessels, modeled by beam elements with the FEM model of the parenchyma.

The remainder of this section is organized as follows. First, we present the FEM beam model based on a corotational formulation. Then, we focus on the contributions dedicated to the optimization of the solving process. Finally, we show, that the discretization of the beam elements can be adapted dynamically, in real-time, and the beam elements can be combined to model coaxial deformable objects.

### 1.3.1 Corotational beam model

To model the deformation of any solid body whose geometry and mechanical characteristics are similar to a wire, rod or beam, we use a representation based on three-dimensional beam theory Przemieniecki (1985), where the elementary stiffness matrix $K_e$ is a $12 \times 12$ symmetric matrix that relates angular and spacial positions of each end of a beam element to the forces and torques applied to them:
The stiffness matrix \( \mathbf{K}_e \) is initially calculated in local coordinates, it is necessary to introduce transformation matrices changing the reference frame from a local to a global coordinate system. In order to determine the stiffness property of the complete structure, a common reference frame must be established for all unassembled structural elements so that all the displacements and their corresponding forces will be referred to a common (global) coordinate system. We need a matrix relationship between the variation of the position \( \delta \mathbf{u} \) in the local coordinate system and the variation of the node position \( \delta \mathbf{q} \) in the global coordinates (see Fig 1.2). This relationship is expressed by the matrix equation:

\[
\delta \mathbf{u} = \mathbf{\Lambda}(\mathbf{q}) \, \delta \mathbf{q}
\]

where \( \mathbf{\Lambda}(\mathbf{q}) \) is a matrix obtained from the direction cosines of angles between the local and global coordinate systems.

The linearization of the element force-displacement equation is obtained in global coordinates:

\[
\delta \mathbf{F}(\mathbf{q}) = \left[ \mathbf{\Lambda}^T(\mathbf{q}) \mathbf{K}_e \mathbf{\Lambda}(\mathbf{q}) \right] \delta \mathbf{q}
\]
1.3. WIRE AND TUBE MODEL BASED ON BEAM THEORY

Figure 1.2: Initial (blue) and deformed (red) configurations. (a) The initial configuration of a catheter could be curved. Thus, for each beam, we store an initial displacement $u_0$ based on the position of the tip node in the local frame of the base node. (b) Actual deformations can be measured locally in the local reference frame using $u$. The variation of this local position is linked to the variation of the node’s position $\delta q$ in the global frame.

The model can be combined with the computation of a mass matrix to obtain mechanical dynamic model. We then use the equation 1.7 to compute the system. One important feature of this model is that the interpolation is performed segment by segment between only two frames. Additionally, these frames are the independent Degrees of Freedom (DoFs) of the system.

1.3.2 Fast solver for real-time

In the beam model, as each element’s DoFs are in global coordinates, and as the topology is a simple sequence of segments, the resulting global stiffness matrix is block-tri-diagonal. The structure of the matrix $K$ (or the matrix $A$ from equation 1.7 if mechanical dynamics is used) is composed by $6 \times 6$ block elements placed on the diagonal $B_i$ and $6 \times 6$ Block elements $A_{ij}$ on the first upper and first lower diagonal:

\[
\begin{bmatrix}
  B_1 & A_{12} & 0 & \ldots & \ldots & 0 \\
  A_{12}^T & B_2 & A_{23} & \ldots & \ldots & \vdots \\
  0 & A_{23}^T & B_3 & A_{34} & \ldots & \vdots \\
  \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\
  \vdots & & \ddots & \ddots & \ddots & \vdots \\
  0 & \ldots & \ldots & 0 & A_{(n-2)(n-1)}^T & B_{n-1} & A_{(n-1)(n)} \\
  \end{bmatrix}

\begin{bmatrix}
  x_1 \\
  x_2 \\
  x_3 \\
  \vdots \\
  x_{n-1} \\
  x_n \\
  \end{bmatrix}

= \begin{bmatrix}
  b_1 \\
  b_2 \\
  b_3 \\
  \vdots \\
  b_{n-1} \\
  b_n \\
  \end{bmatrix}

(1.13)

This type of system can be solved very efficiently, and allows for the real-time computation of wire models composed of hundreds of elements.

In the following, the method is described quickly. It perfectly illustrates the fact that obtaining a real-time model of deformation often necessitates both performance in
CHAPTER 1. MECHA. DEFORM. MODELS FOR REAL-TIME COMPUTATION

the mechanical model and in the resolution process. Sometimes it requires to adapt directly the numerical solvers to obtain the desired performance.

The method is based on a technique that is similar to domain decomposition. Without loss of generality, let us cut and zoom on the matrix at a given block $i$ as we did on block $i = 2$ on the equation 1.13 (vertical and horizontal lines). We suppose that $x_{i-1}$ is known:

\[
\begin{bmatrix}
B_i & A_{i,i+1}\n
A^T_{i,i+1} & B_{i+1}
\end{bmatrix}
\begin{bmatrix}
x_i \\
x_{i+1}
\end{bmatrix}
= \begin{bmatrix}
\frac{b_i - A^T_{i-1,i}x_{i-1}}{b_{i+1} - A_{i+1,i+2}x_{i+2}} \\
\frac{b_i}{b_{i+1}}
\end{bmatrix}
\tag{1.14}
\]

We can decompose the first line of this system with two separate resolutions:

\[
B_i x_i^{(1)} = b_i, \quad B_i x_i^{(2)} = -A_{i,i+1} x_{i+1}, \quad x_i = x_i^{(1)} + x_i^{(2)}
\]

$x_i^{(1)}$ can be computed if we know $b_i$ (so, if we suppose that we know the value of $x_{i-1}$) whereas $x_i^{(2)}$ necessitates the solution of block $i + 1$ to be computed.

The algorithm relies on a first process called boundary fixation, from $i = 1$ to $i = N - 1$ that computes $x_i^{(1)}$ while accumulating values in $b_i$. Then, the solution $x_N$ can be found for the last node. Finally, the algorithm iterate from $i = N - 1$ to $i = 1$ to compute $x_i^{(2)}$ and the final value of $x_i$. More details are available in Duriez et al. (2006a). The resulting algorithm for solving the structure is $O(n)$ if $n$ the number of beam elements.

The algorithm can also be adapted to obtain the diagonal blocks of the inverse matrix $C = K^{-1}$ (or $C = A^{-1}$ in dynamics). Let’s inverse the system presented in equation 1.14:

\[
\begin{bmatrix}
x_i \\
x_{i+1}
\end{bmatrix}
= \begin{bmatrix}
B_i^{-1} + H_{i,i+1} C_{i+1,i+1} H_{i+1,i} & -B_i^{-1} C_{i+1,i+1} H_{i+1,i} \\
-C_{i+1,i} A^T_{i+1,i} B_i^{-1} & H_{i,i+1}
\end{bmatrix}
\begin{bmatrix}
\frac{b_i}{b_{i+1}} \\
\frac{b_i}{b_{i+1}}
\end{bmatrix}
\tag{1.15}
\]

The computation of $C_{i+1}$ (2nd line and 2nd column of the matrix above) can be done while processing the different blocks in ascending order, whereas the computation of $B_i^{-1} + H_{i,i+1} C_{i+1,i+1} H_{i+1,i}$ can be obtained by processing the blocks in descending order. It leads to a similar algorithm that provides the tri-diagonal blocks of the inverse matrix with a $O(n)$ complexity.

Conclusion: The model and this algorithm was mainly used for the computation of interventional radiology instruments (see figure 1.3), and can also be used for modeling suture thread. Recently, we have proposed a new implementation where the sampling of the model can be adapted in real-time. In future work, we need to improve the adaptation of the model to soft anatomical structures (such as blood vessels like in section 1.5.3).
1.4 Surface model based on shell theory

The work done on interactive simulation of shell models has started during the PhD thesis of Olivier Comas. The contributions include a corotational formulation of a shell element Comas et al. (2010a) and its adaptation to simulate, in real-time, the deformation of hollow anatomical structure Comas et al. (2010c). This topic is still active: Tomas Golembiovsky, who started his PhD in September 2011, has recently proposed a new formulation based on Bezier Triangles, in Golembiovsky et Duriez (2012), that is shortly presented here.

1.4.1 Shell model adapted to real-time simulation

We propose to define a triangular shell element by combining a two-dimensional inplane membrane energy, with an off-plane energy for describing bending and twist (see figure 1.4). To allow for real-time simulation, a computationally efficient formulation is needed. We therefore propose to extend the co-rotational idea introduced by Felippa (2000) and applied to real-time simulation of volumetric objects over the last few years Muller et Gross (2004). Corotational models offer a good trade-off between computational efficiency and accuracy by allowing small deformations but large displacements. We propose to improve and extend in the same way the plate model first introduced by Przemieniecki Przemieniecki (1985). Once combined with an in-plane membrane formulation we obtain an accurate, yet computationally efficient, shell finite element method featuring both membrane and bending energies.
CHAPTER 1. MECHA. DEFORM. MODELS FOR REAL-TIME COMPUTATION

Figure 1.4: A triangular shell element can be defined as a combination of a triangular in-plane membrane element (left) and a triangular thin plate in bending (right). The different degrees of freedom $v$ and $u$ of both models are illustrated above.

**Triangular elastic membrane**

The computation of the triangular elastic membrane stiffness matrix is often detailed in books related to FEM deformable model, like Przemieniecki (1985). The interpolation inside the element is linear, and the stiffness matrix $K_e$ can be computed as follows:

$$K_e = \int_V J \chi J^T dV \quad (1.16)$$

where $J$ is a matrix that provides the strain-displacement relation and $\chi$ embodies the material’s behaviour. We assume that the local deformations of each triangle remain limited during the simulation and a linear constitutive law is sufficiently accurate. Thus in the simple case of Hooke’s law, assuming the plane stress hypothesis, we have:

$$\chi = \frac{E}{(1-\nu^2)} \begin{bmatrix} 1 & \nu & 0 \\ \nu & 1 & 0 \\ 0 & 0 & \frac{1}{2}(1-\nu) \end{bmatrix} \quad (1.17)$$

The stiffness matrix in the global frame is eventually obtained using the rotation matrix of the element: $K = RK_eR^T$ where $R$ describes the rotation of the (triangular) element with respect to its initial configuration.

**Triangular plate bending:** Let’s assume that we have a function of the deflection $u_z(x,y)$ for each point of coordinate $(x,y)$ on the triangle. In the flat-plate theory the strains are computing by using:

$$e_{xx} = -z \frac{\partial^2 u_z}{\partial x^2} \quad e_{yy} = -z \frac{\partial^2 u_z}{\partial y^2} \quad e_{xy} = -2z \frac{\partial^2 u_z}{\partial x \partial y} \quad (1.18)$$
1.4. SURFACE MODEL BASED ON SHELL THEORY

As second derivatives are involved in the computation, a polynomial interpolation is used rather than a linear interpolation:

\[ u_z = c_1 + c_2 x + c_3 y + c_4 x^2 + c_5 xy + c_6 y^2 + c_7 x^3 + c_8 xy^2 + c_9 y^3 \] (1.19)

This deflection is associated to the transverse displacements (along \( z \) axis) and the rotations of the nodes, as shown on figure 1.4. For instance, here, for point 1:

\[ u_1 = (u_z)_{x_1,y_1} \quad u_2 = \left( \frac{\partial u_z}{\partial y} \right)_{x_1,y_1} \quad u_3 = -\left( \frac{\partial u_z}{\partial x} \right)_{x_1,y_1} \] (1.20)

Thanks to this interpolation we can compute a stiffness matrix for the bending part of the form:

\[ K_e = (C^{-1})^T \int_e \mathbf{D}(x,y)^T \chi \mathbf{D}(x,y) dV \quad C^{-1} \] (1.21)

Where \( \mathbf{C} \) is a constant matrix, that depends on the initial position of the triangle vertices, and \( \mathbf{D} \) a non-constant matrix that depends on \((x,y)\) values (see the details in Comas et al. (2010a)). Consequently, the integration is carried out numerically using Gauss points. The quadrature chosen for the element was based on 3 points located at the middle of each edge of the triangle.

Finally the shell element is a combination of the elastic membrane model and plate bending model.

Implementation and Validation: In practical terms, the different computations associated with each triangular shell element can be described as follows:

1. Compute the rotation matrix \( \mathbf{R} \) from global to triangle (local) frame
2. Compute the local displacement vector \( \mathbf{u} = \{ v_1, v_2, 0, u_2, u_3, v_3, v_4, 0, u_5, u_6, v_5, v_6, 0, u_8, u_9 \} \) for each of the 3 nodes of the triangle (the nodes have 6 degrees-of-freedom). As we are in a co-rotational framework the normal displacements \( u_1, u_4, u_7 \) of all 3 nodes are null in the local frame of the triangle.
3. Compute matrix \( \mathbf{D}_i \) at each Gauss point \( i \)
4. The strain-displacement matrix at each Gauss point \( i \) is computed with \( \mathbf{J}_i = \mathbf{D}_i \mathbf{C}^{-1} \)
5. Compute the local stiffness matrix \( \mathbf{K}_e \) of the element as \( \mathbf{K}_e = \sum_{i=1}^{3} \mathbf{J}_i \chi \mathbf{J}_i^T \)
6. Transform the local element stiffness matrix into the global frame and add it to a global stiffness matrix
The co-rotational shell formulation is successfully implemented into the open-source framework SOFA. The framework allows the comparison with some theoretical results, like the one reported by Zhongnian (1986) to assess the quality in modeling bending. The test carried out uses a square shape mesh clamped on all four edges and loaded uniformly using increasing load values $q$ (ranging from 1 to $5N/m^2$). The maximum deflection $z_{\text{max}}$ is compared with theoretical results:

<table>
<thead>
<tr>
<th>$q$</th>
<th>$z_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1218</td>
</tr>
<tr>
<td>2</td>
<td>0.2475</td>
</tr>
<tr>
<td>3</td>
<td>0.3747</td>
</tr>
<tr>
<td>4</td>
<td>0.5050</td>
</tr>
<tr>
<td>5</td>
<td>0.6374</td>
</tr>
</tbody>
</table>

Table 1.1: Comparison of our shell model with theoretical results on the bending of a square plate. An error of less than 1% was found between our simulation and theoretical results.

**Computation times:** We perform several tests at different resolutions to measure computation times (Fig. 1.5). A set of shells are resisting a uniform pressure load and solved using a Conjugate Gradient (CG) iterative solver. Implicit integration allows for large time steps (40ms) and the computation is real-time for 800 shell elements and a reasonable error criterion (5%). When the computation time must be bounded (critical real-time applications), one can fix the number of CG iterations to, for instance, 100 and remains real-time for 1000 shell elements. However, in that case the accuracy of the results is not checked.

**Discussion:** We obtain very good results with this model, that allows for a good tradeoff between performance and accuracy. The fact that the interpolation for bending is cubic allows the use of large triangle. In the same time, the real-time constraint
limits the number of triangle elements that can be computed simultaneously. In or-
der to use the model for anatomical structures, a mesh reconstruction is needed to
minimize the number of triangles (see following subsection).

The model has also two drawbacks: (1) it is not perfectly symmetric, as equation 1.19 is
not symmetric for $x$ and $y$ and (2) how nodes are numbered can have a small influence
on the results. Moreover, the model has 6 degrees of freedom per node but only 5 are
concerned by the stiffness matrix (the rotation along $z$ axis is "free"). It could result
some problem of bad-conditioning and some issue with stability (the frames tend to
rotate along $z$ axis when the deformation becomes large). One solution to this problem
could be to remove this degree of freedom (have only 5 DOFs per node) but it would
result a more complex kinematic law and would not solve the problem of symmetry.

1.4.2 Shell model based on Bezier triangle geometry

Recently, in Golembiovsky et Duriez (2012), we have proposed a new formulation for
the shell elements based on Bezier triangle geometry. This new formulation addresses
noted drawback of the corotational shell element presented above.

Extending our previous work, we have designed a shell element that uses an interpola-
tion based on Bézier polynomials. Our element solves the problems presented above:
(i) it uses all 6 degrees of freedom, (ii) the deformations are completely symmetric and
(iii) the boundary between elements is continuous ($C^1$ on nodes, $C^0$ on edges).

![Figure 1.6: Two stages of interpolation: First, from triangular with 6DOF nodes (a) a mesh
of control points (b) is computed. Then based on definition of the Bézier triangle a surface
(c) is interpolated.](image)

But the novelty is also coming from the use of a two stage interpolation (see figure
1.6):

1. A kinematic link between 6 DoF nodes of element and nodes of Bézier control
mesh is defined

2. Interpolation function of cubic Bézier triangle is applied to interpolate on the
surface of the element.
Cubic Bézier triangle has 10 control points with 3 DoFs per point, for a total of 30 DoFs per triangle. Our technique is based on a reduction of the number of DoFs (using 3 nodes with 6 DoFs) while keeping the high degree of the interpolation function. To do that, we propose some kinematic links between the 3 nodes with 6 DoFs and the Bézier control points that is illustrated on Figure 1.7 (Left).

Figure 1.7: From 6DoFs nodes to Bézier interpolation: (Left) Vertex and edge control points are rigidly attached to the nearest corner node, The central point position blends the contribution of the 3 nodes. (Right) Comparison of tube deformation between coarse and fine models: (up) the model is composed of only 208 shell elements whereas (down) it has 3038 triangles. In both case, the visualization is realized using a model composed of 3038 triangles.

These kinematic links are also derived for interpolating the velocity of the surface points and also to distribute the forces applied on the element surface back to 6DoFs nodes of the element. This allows us to use smooth mapped surface for visualization and collision detection (see the figure 1.7 (right)) even with very coarse mechanical models (so very fast to compute).

For now, the model still relies on linear elasticity and on a separation between in-plane and bending energies. Yet, the constitutive laws of soft anatomical structures are often more complex. In Ubach et al. (2010), it is shown that a hyperelastic formulation can be successfully derived from shell elements based on Bézier cubic triangle interpolation. We plan to provide this extension in a close future and also to adapt the computation process of the shell elements on GPU.

1.4.3 Modeling of thin anatomical structures

In this section, we present some contributions that help the use of the shell elements for thin anatomical structures like membranes and hollow organs. In order to apply the shell theory in the context of medical simulation, our method proposes to base the geometrical reconstruction of the organ on the shape functions of the shell elements. Moreover, we also use these continuous shape functions to handle the interactions
1.4. SURFACE MODEL BASED ON SHELL THEORY

with other types of deformable tissues. The technique is valid for both models of shell (corotational and Bézier triangle) using the adequate shape functions.

**Physics-based reconstruction using shell elements:** Because the surface of an anatomical structure has a physical meaning, we propose to patch the surface with triangular elements whose interpolation makes use of the same shape function designed for our shell FEM formulation. Let’s assume that we have a high resolution triangular mesh obtained from a binary segmented image of the organ we want to simulate (via a Marching Cube algorithm for instance). We have proposed a method to create a mesh of good quality, featuring the optimal number of shell elements while staying as close as possible to our targeted geometry.

![Figure 1.8](image.png)

Figure 1.8: (a) the targeted high resolution Glisson’s capsule mesh (8,000 triangles). (b) the one-sided Hausdorff distance error map after applying only one iteration of our algorithm to the coarse mesh (1,200 shells).

While many flat triangles are required to describe highly curved surfaces, fewer triangular shell elements are needed to describe the given geometry with the same precision since they can be curved. Consequently, the first step of the method is an important decimation of the high resolution mesh, using quadric edge collapse technique implemented in Meshlab CNR (2005).

Then the method improves the quality of the mesh and reduces its distance to the high resolution triangular mesh by using heuristics that were first presented in Saupin et al. (2007) and are detailed in Comas et al. (2010c). This approach has been applied to the modeling of the Glisson’s capsule, as presented in figure 1.8. Using relatively coarse meshes of curved shells, we can approximate with good accuracy, the complex geometry of this capsule (the error is expressed as a percentage of the diagonal of the object’s bounding box).

**Mechanical interactions and coupling with the curved surface of shells:** The practical interest of modeling complex behaviors such as bending and twisting would remain fairly low for medical simulation if contacts and constraints were not handled properly. In the case of liver simulation, for instance, it is important to handle contacts with the surface of the liver (for contacts with instruments). We could also use it for
the hollow organs in its surrounding environment (stomach, transverse colon...) but it would necessitate to model the mechanical interactions.

![Diagram](image)

Figure 1.9: (a, b) The triangle formed by the three vertices of the shell has been recursively subdivided 3 times to allow more accurate rendering and collision detection. (c) The shape function is used to distribute an external force $F$ onto the triangle nodes.

In order to detect the collisions, or, in general, the interactions with the bent surface, we can chose a more detailed mesh. We can use a subdivision approach or an unstructured strategy. The vertices of the new mesh are computed using the shape function (see equation 1.19 or Bézier polynomials) and the position and rotation of the corresponding element nodes. An example using subdivision is illustrated in Fig. 1.9 (a) and (b). The process then use classical collision detection algorithm, working on flat triangles, using the detailed mesh.

We put aside the problem of modeling these interactions (related to collision response issues that are addressed in chapter 2). After solving the constraints, we still need to distribute the forces applied on the curved triangles. Indeed, as our surface model is composed of curved shell elements, the forces need to be distributed between linear forces and torques onto its three vertices (see figure 1.9 (c)). In Comas et al. (2010c), we propose a method based on virtual work principle that is also used for the Bézier element in Golembioveyk et Duriez (2012).

An example of medical procedure to illustrate this point is angioplasty. Angioplasty is the technique of mechanically widening a narrowed or obstructed blood vessel. This technique is used typically for atherosclerosis pathologies, but can also be used for specific pathologies on the liver (complications of liver transplantation).

An empty and collapsed balloon on a guide wire is passed into the narrowed location and then inflated to a fixed size. The balloon crushes the fat deposits, opening up the blood vessel to improve the blood flow. As a proof of concept we tried to simulate an angioplasty (Fig. 1.10). The blood vessel is modelled using the shell FEM formulation described in this paper and the deposits are simulated with a tetrahedral elements and are fixed to the interior wall of the blood vessel. When the balloon inflates, it crushes the deposits which apply a pressure onto the curved surface of shells modelling the interior wall. The forces are then distributed onto the mechanical nodes of the blood
1.5 Volume deformation of soft-tissues

Figure 1.10: Simulation of an angioplasty procedure. (a, c): A collapsed stent is inserted into the blood vessel. (b, d): The stent is crushing the fatty deposits which creates a pressure onto the interior wall and widens the blood vessel.

vessel, which widens the blood vessel as expected.

Conclusion: We propose a framework for real-time modeling of thin anatomical structures based on shell finite elements. In future work, we will investigate the use of adaptive strategy to improve the precision when dealing with localized high curvatures. Moreover, Mouhamadou Diallo, who is doing his PhD in collaboration with the Laboratoire de Mecanique de Lille, has recently used the shells to model the deformations of the pelvic system (see section 4.1.5) and will work on incorporating more advanced constitutive laws in the model.

1.5 Volume deformation of soft-tissues

Volume model deformations can be accurately modeled using FEM approach. But the complexity, in term of the number of nodes and elements is difficultly compatible with real-time. It leads to the use of optimized solver combined with computations on GPU, as we did in Courtecuisse et al. (2011b).

In this chapter, a first attempt on using multi-resolution approach for real-time is presented. It was developed during the Ph.D. of Guillaume Saupin and published in Saupin et al. (2007). Moreover, we discuss a new numerical method, based on asynchronous preconditionning, presented in Courtecuisse et al. (2010), that fixes the bad-conditioning associated to inhomogeneities of the soft-tissues. It was developed during the Ph.D. of Hadrien Courtecuisse. Finally, we also proposed a new coupling strategy for vascularized organs, like the liver, as presented in Peterlik et al. (2012), which was realized during the post-doctorate fellowship of Igor Peterlik.

One of the main limitations of the work presented during my phD was that linearity of the volume deformation models used for the real-time simulation. Using linear models,
there was a strong limit on the range of deformation modeled with accuracy.

Let’s take again the liver simulation example. In the context of using medical simulation for planning or per-operative guidance, more realistic models of the patients’ liver anatomy and behaviour are required. The geometry and tissue properties of the liver model need to be patient-specific and obtained without complex additional procedure; Soft tissue behaviour, which is highly non-homogeneous due to the presence of a dense network of blood vessels, needs to demonstrate a predictive capability, yet should be compatible with real-time computation; Stable interactions with the surrounding anatomy and with medical devices should be handled; In the particular case of an hepatectomy, the different types of dissection performed on soft tissues should be simulated.

Addressing all these requirements is very challenging. The parenchyma of the liver, as a volume model, needs generally to be sample with more DoFs than surface or curve model. Moreover, to capture precise material behavior on a large range of volume deformation, we need to rely on non-linear strain tensors coupled with realistic constitutive laws. Additionally, some features could have a large impact on the computation time: inhomogeneities often lead to badly conditioned systems, stable interactions often leads to the use of implicit solver that are more computationally demanding, dissections ask for topological changes that are often costly and could also lead to bad conditioning.

Given these challenges, there is a relatively large gap between models typically used in the field of biomechanics and approaches traditionally chosen to obtain real-time computation of volume deformable bodies. Although more complex to implement, FEM methods are the usual choice in biomechanics for numerically solving the partial differential equations of constitutive laws. Therefore, we seek to have a framework able to use these non-linear FEM in a real-time context, even using a relatively simple constitutive law. In Courtecuisse et al. (2011b), we showed that the choice of a linear elastic material, combined with a GPU implementation, can enable very fast computations: we used the co-rotational method, introduced by Felippa (2000) that provides a relatively simple way to handle large displacements and geometrical non-linearities. However, recently, it has been shown in Marchesseau et al. (2010) that it is possible to use a fast optimization for modeling visco-hyperelastic properties of the liver in real-time using the so called MJED technique.

Whatever the constitutive model chosen is (elastic or visco-hyperelastic), these works proved that it is possible to use a FEM framework in real-time for the liver simulation. However there are at least three main issues that we adress in the contributions presented in the following: First, we try to increase the number of DoFs that are simulated in real-time, in order to reach a certain numerical convergence, using a multi-grid approach. Second, we address the problem of bad-conditioning by a new preconditioning technique that relies on a method called Compliance Warping, that is compatible with both elastic and hyperelastic models and with cutting by removing
elements. Third, using this preconditioner, we propose a new way of modeling the inhomogeneities caused by the blood vessels network of the liver. The approach relies on a coupling between the volume model of the parenchyma and the beam model (presented in section 1.3) for the blood vessels.

1.5.1 Multi-grid FEM solver for real-time simulation

Multigrid approach is based on the presupposition that the finite element method applied to the deformable model with its constitutive law converges to a unique solution. Using a hierarchy of mesh, a first solution can be obtained quickly on the coarser mesh and can be propagated at the finer mesh levels. Then an iterative solver on the finer mesh is initialized with a guess that is supposedly very close to the solution. This leads to fast convergence rate at the finer mesh while doing most of the iterations on the coarser mesh (that are much faster to do).

Multigrid solvers are a very promising technique in our context. However, these multigrid methods still suffer, when used to simulate large deformations, from two pitfalls, depending on the kind of grids hierarchy used. If embedded grids are used, approximating complex geometries becomes difficult, when unstructured grids hierarchy is used, solving speed-up is reduced by the necessity to update coarser levels stiffness matrices.

In this work, published in Saupin et al. (2007), we introduce a new hierarchical mesh generator which can build a hierarchy of topologically embedded grids that approximates a complex geometry. Then, we also improve the computation time on embedded grid by taking advantage of the stiffness matrix sparsity pattern to efficiently update coarse matrices.

Hierarchical Mesh Generator

Embedded hierarchy of 3D meshes is built on a coarse mesh \( \Omega_0 \) obtained using a decimation tool. Then, for each level of the hierarchy, each tetrahedron is subdivided in 8 new tetrahedra, and new nodes are introduced in the middle of each edge (see Figure 1.11).

We use external decimation tools to build the coarse approximation \( \partial \Omega_0 \) of the boundary target \( \partial \Omega \). This initial coarse boundary mesh \( \partial \Omega_0 \) is generated using an external software (for instance CGAL Library \(^3\)).

Due to the construction scheme above, the geometry approximated by each level of the mesh hierarchy is exactly the same that the one of the coarsest mesh \( \Omega_0 \) as each new nodes \( n_{\text{new}} \) of

\(^3\)CGAL (Computational Geometry Algorithms Library): http://www.cgal.org/
the boundary of $\Omega_{i+1}$ already belongs to the boundary of $\Omega_i$.

Thus, we introduce a heuristic for a geometrical approximation step in the refinement scheme: each new node $n_{\text{new}}$ that belongs to the boundary of the mesh is moved in order to fit a target boundary geometry $\partial \Omega$.

The proposed heuristic is based on very simple geometrical rules, that are described in Saupin et al. (2007). It has been tested with success on various meshes with more complex surfaces than the surface of the liver (see Figure 1.12 to 1.17). This technique allows for obtaining a hierarchy of embedded grids while obtaining a fine mesh that is very close to the original mesh.

**Multigrid Solver with fast update of matrices** Multigrid solvers are a well known class of solvers that use a grids hierarchy to efficiently and quickly solve differential equations. When used in a finite elements context, grids are referred as meshes. The method uses restriction and prolongation operators $R$ and $P$, which allow respectively to switch from fine to coarse resolution and inversely, are used to build the problem at a coarser resolution, and to transfer a result from one level to the other.

For instance if we try to solve $A_i x_i = b_i$ on the fine level, we can compute an initial guess $x_i^0$ by solving $A_{i-1} x_{i-1} = b_{i-1}$ on the coarse level and then transfer the solution to the fine level: $x_i^0 = Px_{i-1}$, where $A_{i-1} = RA_i P$. There are several existing schemes adapted to multi-grid approach (like the V-Cycle or the W-Cycle etc..) but they all rely on the same principle:

1. Compute residual: $r_i = A_i x_i^0 - b_i$
2. transfer residual to coarse mesh: $r_{i-1} = R r_i$. 
3. Find \( \mathbf{v}_{i-1} \) so that \( \mathbf{A}_{i-1} \mathbf{v}_{i-1} = \mathbf{r}_{i-1} \)

4. Transfer correction to fine mesh : \( \mathbf{v}_i = \mathbf{P} \mathbf{v}_{i-1} \).

5. Apply correction : \( \mathbf{x}_i = \mathbf{x}_0^i + \mathbf{v}_i \). (\( \mathbf{x}_i \) is non-smooth as the residual \( \mathbf{r}_i \) wasn’t)

6. Smooth \( \mathbf{A}_i \mathbf{x}_i = \mathbf{b}_i \) using \( \alpha \) Gauss-Seidel relaxations.

The prolongation operator \( \mathbf{P} \) is usually an interpolating operator. The advantage of embedded mesh, is that this interpolation is very simple: each new node of the fine mesh \( \Omega_i \) is the average of its two neighbors nodes in \( \Omega_{i-1} \). Consequently, \( \mathbf{P} \) is a very sparse matrix with a specific pattern. The restriction operator is usually \( \mathbf{R} = \mathbf{P}^T \).

The stumbling point in the algorithm is the computation of the coarse matrix \( \mathbf{A}_{i-1} = \mathbf{R} \mathbf{A}_i \mathbf{P} \). The models used in simulations usually have thousands of tetrahedra which implies huge \( \mathbf{A}_i \) matrices. Therefore the product \( \mathbf{R} \mathbf{A}_i \mathbf{P} \) is time consuming, even though \( \mathbf{R}, \mathbf{P} \) and \( \mathbf{A}_i \) are highly sparse, because when they are stored in the sparse format, matrix product is very inefficient. This problem is recurrent in every multigrid solver, especially on unstructured grids, like in Georgii et Westermann (2006), Shi et al. (2006).

However, the \( \mathbf{P}, \mathbf{R} \) and \( \mathbf{A}_i \) have the remarkable property that their sparsity pattern is fixed for the whole simulation. And this important property allows some precomputing operations which improves the computation time.

We propose to compute this product in two steps :

\[
\mathbf{C} = \mathbf{A}_i \star \mathbf{P} \quad (1.22)
\]

\[
\mathbf{A}_{i-1} = \mathbf{R} \star \mathbf{C} \quad (1.23)
\]

Where the \( \star \) operator, our optimized matrix product, works as follow :

1. Once the sparcity patterns of \( \mathbf{P} \) and \( \mathbf{A}_i \) are known, we compute the intersection for the equation (1.22) by finding for each element \( c_{k,n} \) of the matrix \( \mathbf{C} \) the list of pairs \( p = (\text{index}(p_{k,l}), \text{index}(a_{m,n}^i)) \) such that \( p_{k,l} \neq 0 \) and \( a_{m,n}^i \neq 0 \). These lists of pairs are stored in an array to compute the \( c_{k,n} \) elements. In addition to each list, we store the index of the corresponding \( c_{k,n} \) in the \( \mathbf{C} \) matrix.

2. Each time we need to compute \( \mathbf{C} \), we move through the array of list of pairs, do the necessary products, and store the result at the given index.

The speed up observed on a bi Dual Core 2 Duo 5130 using our method is presented in table 1.18, and compared with the naive methods: using Intel MKL sparse matrix product, and using unstructured grids (barycentric approach) from Georgii et Westermann (2006). We observed a 2.0 to 3.0 speed up factor.

Despite very encouraging results, this work was not pursued because Guillaume then worked on a new way of solving contact for deformable with geometrical non-linearities. Moreover this work was implemented outside SOFA (that was not adapted for that at
that time, which is not the case today). However, let’s emphasize that the algorithm of mesh generation was partially reused by Olivier Comas for the shell elements (see section 1.4.3). I have chosen to highlight this work in this document because I think that multi-grid approach is still a good direction of research. It could be combined, for instance, with the preconditioning technique presented below or with algorithms for contact response.

### 1.5.2 Asynchronous Preconditioning techniques

The use of parallelism is particularly well-suited to our context: theoretically, more computation are doable for the same time interval. Moreover, we achieve performance gain by taking full advantage of the heterogeneous architecture of current hardware. In other words, by adapting our algorithms to multi-threaded approach, we could possibly gain accuracy while maintaining computational time adapted to real-time. This section is a perfect illustration of this strategy: a preconditioner is computed asynchronously in a separate thread. It leads to fast update of the deformable model even in the case of bad conditioning, such as in the presence of strong inhomogeneities. Moreover, in chapter 2, we will show how it was implemented on GPU to compute a good approximation of the inverse matrix used by the contact response algorithms.

**Brief background on Preconditioners** Conjugate Gradient is a very good solver in our context: as it is based on matrix vector product, it does not require the construction of the matrix of the system, it can be parallelized on GPU Allard et al. (2011b) and it provides an easy tradeoff between accuracy and computation time. However, it is well known in the literature that in some bad cases like non-homogeneous simulations or high material stiffnesses, the convergence of the CG becomes very slow.

A preconditioner is used to reduce the condition number of the system, by defining an approximation of the system matrix $A$ which is easier to invert than $A$. Solving equation (1.7) with a preconditioner $P$ can be written as:

$$P^{-1}Ax = P^{-1}b$$

(1.24)
1.5. VOLUME DEFORMATION OF SOFT-TISSUES

The condition number of this modified problem is $\kappa(P^{-1}A)$. If we choose $P$ sufficiently close to $A$, then $(P^{-1}A)$ will be close to the identity matrix, its condition number close to 1, ensuring a fast convergence.

When $P$ is not close to $A$, like inverse diagonal, block diagonal (used in Choi et Ko (2002) and Boxerman et Ascher (2004)) or even symmetric successive over-relaxation (SSOR) (see Hauth et al. (2003)), the preconditioning technique provides limited overhead of computation at each iteration of the CG but usually a limited impact on the reduction of the number of iterations.

When $P$ is close to $A$, like the Incomplete Cholesky (IC) (see Hauth et al. (2003)), the preconditioner has two overheads: A major one corresponds to the factorization or the inversion of $P$ which is done one time before the iteration. The second overhead is the application of the preconditioner at each iteration. However, the number of iterations is greatly reduced.

Multi-threaded asynchronous preconditioning The idea of this contribution is to exploit the temporal coherence of deformable objects simulation: $A$ does not change radically between successive time steps. $P^{-1}$ could be a good preconditioner during several steps so the factorization of $P$ could be put in a separate thread. Then the benefit is that the computation factorization of $P$ does not affect the computation time of the CG any more.

The approach is presented in figure 1.19(left): the simulation thread executes the step using a non-built version of the CG (above) and a second separate thread builds matrix $P$ and computes a new factorization. When this factorization (invert $P_t$ on the Figure) is done, a notification is sent to the simulation so that the CG can use this new factorization.

The method is said to be asynchronous because there is no synchronization between the two threads: the launch of a new factorization can simply be triggered by the end of the preceding computation. This provides to the CG the latest available preconditioner, while the thread associated to the preconditioner almost never stops computing.

Overall, the proposed multi-threaded approach allows to desynchronize the preconditioner from the simulation to produce good inverse approximations of the system matrix continuously. Even if there is a good temporal coherence, the preconditioner could be quite old when it is used by the CG. Indeed, it was computed using a matrix built several steps previously. Defining $\Delta t$ as the average update period in simulation time, each preconditioner will be used from time $t + \Delta t$ to time $t + 2\Delta t$.

Local rotations warping In the worst case, the preconditioner used in the simulation is $2\Delta t$ old. In some case, it is too long to prevent a divergence of the quality of the preconditioning.
We analyzed that what breaks the temporal coherence of $A$, in the first place, is the geometrical non-linearities in the strain formulation due to the local rotations. Thus, we estimate the rotation for each node of the model, relative to their position at the time the preconditioner was computed. The rotations of all the nodes are stored in a block diagonal rotation matrix $R$. This rotation matrix is then used to warp the factorized matrix of the preconditioner. Equation (1.24) is replaced by:

$$RP^{-1}R^T Ax = RP^{-1}R^T b$$

This technique was inspired by the compliance warping method developed during the PhD of Guillaume Saupin Saupin et al. (2008b) for contact problems with a corotational formulation. When the deformation model uses a corotational formulation, we exploit the computed rotations on each element, using for each node the mean rotation from neighboring elements. But we also applied the method on hyperelastic deformations using the MJED implementation from Marchesseau et al. (2010) available in SOFA and we rely on shape matching Müller et al. (2005), to compute an estimate of the rotations of each node. The rotation value $R$ is estimated thanks to the computation of a rotation value for each node, computed during the current time step $R_{t+\Delta t}$, and thanks to the value of the same rotation $R_t$ at time $t$ when the preconditioner computation was launched (see figure 1.19 (right)).

$$R = R_t^{-1} R_{t+\Delta t}$$

This warping of the preconditioner matrix greatly improves its quality (a shown in Figure 1.20) and the overhead is relatively small since $R$ is a 3 $\times$ 3 block diagonal matrix.
1.5. VOLUME DEFORMATION OF SOFT-TISSUES

Figure 1.20: Performances of different preconditioning schemes during the simulation of a co-rotational FEM beam bending under gravity, measured as (a) the number of iterations required by the CG to converge, and (b) the computation time per simulation time-step. All preconditioners use an Incomplete Cholesky factorization, but with different update strategies. CG: no preconditioner, PCG-0: factorization done at the beginning, never actualized, PCG-1: factorization done at each time step, PCG-30: factorization done each 30 steps, PCG-30 (async): asynchronous update of the preconditioner using a separate thread, PCG-30 (async+ warp) asynchronous update with local rotations warping

Results  We first evaluate the warping technique that is introduced in this method for both Corotational and MJED models. In Figure 1.21, we report the performance obtained with or without warping, on different deformation. Using our asynchronous preconditioner method with warping, significant improvements of the quality of the preconditioner are obtained when geometrical non-linearities occur (Scenario 2 and Scenario 3). Indeed, the warping method divides the required number of iterations up to a factor of 5 on co-rotational FEM, and 1.5 for MJED.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Model</th>
<th>No preconditioner</th>
<th>Asynchronous preconditioner</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>No warping</td>
<td>Warping</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Iter.</td>
<td>Time</td>
</tr>
<tr>
<td>Corot.</td>
<td>257.44</td>
<td>110.97</td>
<td>6.91</td>
</tr>
<tr>
<td>MJED</td>
<td>117.36</td>
<td>13.74</td>
<td>4.51</td>
</tr>
<tr>
<td>Corot.</td>
<td>252.63</td>
<td>107.40</td>
<td>13.58</td>
</tr>
<tr>
<td>MJED</td>
<td>179.28</td>
<td>20.03</td>
<td>9.52</td>
</tr>
<tr>
<td>Corot.</td>
<td>264.20</td>
<td>119.39</td>
<td>20.70</td>
</tr>
<tr>
<td>MJED</td>
<td>119.61</td>
<td>16.32</td>
<td>10.33</td>
</tr>
</tbody>
</table>

Figure 1.21: Influence of rotation warping for different solicitations applied on different deformation models

We propose to evaluate the convergence using different preconditioners on non homogeneous problems. We present performances evaluation in Fig. 1.22. The method is tested with various preconditioners (Jacobi, SSOR, Cholesky, Incomplete Cholesky).
First, these preconditioners are updated at each step of the simulation (Standard). Then asynchronous factorization is combined warping technique (asynch+warp).

<table>
<thead>
<tr>
<th>Preconditioner</th>
<th>CG iterations</th>
<th>Computation time (ms)</th>
<th>Invert</th>
<th>Solve</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-\text{CG})</td>
<td>654.03</td>
<td>–</td>
<td>126.44</td>
<td>126.44</td>
<td></td>
</tr>
<tr>
<td>Jacobi</td>
<td>356.96</td>
<td>.15</td>
<td>86.42</td>
<td>86.58</td>
<td></td>
</tr>
<tr>
<td>SSOR</td>
<td>116.72</td>
<td>30.60</td>
<td>56.61</td>
<td>86.21</td>
<td></td>
</tr>
<tr>
<td>Cholesky</td>
<td>1.00</td>
<td>39.27</td>
<td>3.13</td>
<td>42.40</td>
<td></td>
</tr>
<tr>
<td>IC</td>
<td>6.07</td>
<td>107.24</td>
<td>7.29</td>
<td>114.54</td>
<td></td>
</tr>
<tr>
<td>\text{async} + \text{warp})</td>
<td>357.99</td>
<td>.16</td>
<td>93.18</td>
<td>93.34</td>
<td></td>
</tr>
<tr>
<td>SSOR</td>
<td>116.76</td>
<td>.26</td>
<td>62.70</td>
<td>62.97</td>
<td></td>
</tr>
<tr>
<td>Cholesky</td>
<td>6.02</td>
<td>.06</td>
<td>11.92</td>
<td>11.98</td>
<td></td>
</tr>
<tr>
<td>IC</td>
<td>8.71</td>
<td>.03</td>
<td>10.76</td>
<td>10.79</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1.22: Performances evaluation of a non homogeneous beam composed of 3000 elements, falling under gravity.

The results on Figure 1.22 show that the asynchronous and warped version of the Cholesky and Incomplete Cholesky provide strong increase of the computation time in such situation (about $10 \times$ faster than the CG without preconditioner and about $4 \times$ faster than the Standard Cholesky). Indeed, their lack of efficiency found in previous works comes from the factorization cost. With our method this factorization cost is hidden in another thread. In addition, the warping technique permits to limit the divergence of the approximation even on a non homogeneous object. This allows the simulation of both homogeneous and heterogeneous objects with about the same level of accuracy, as even highly non-homogenous simulations converge within less than 10 iterations. We also tested it in an interactive application visible in Figure 1.23. As it was published in a computer graphics workshop, we use a deformable armadillo (not a human soft-tissue) composed of 3650 tetrahedral elements with a non homogenous

Figure 1.23: Interactive virtual manipulation of a highly non-homogeneous deformable object
stiffness. The user can interactively move and rotate to experience the responsiveness and smoothness of the simulation. In this example, while a non-preconditioned version would not converge in a reasonable time for interactive rates, our method was able to maintain a framerate between 50 and 80 FPS.

This preconditionning technique allows for noticeable improvements of the computation time, especially for non-homogenous cases. It is perfectly adapted to real-time and interactive simulation and applicable on both elastic and hyperelastic models. A concrete medical application is presented in the following: an approach that takes into account the rigidity of the vessels in the liver deformable model.

1.5.3 Vascularized organ model

As stated in the introduction, we need to increase the accuracy on the biomechanical model of the liver to enable planning and per-operative approaches. An important amount of work exists regarding the modeling of the liver and its biomechanical properties. For instance in Kerdok et al. (2006) a visco-elastic model of the liver is proposed as well as experimentally measured parameters. Gao et al. (2009) also report on a study to determine soft tissue properties of the liver. Yet, none of the existing approaches take into account the biomechanical influence of the vascular structures of the liver.

In this work, we evaluate this influence through in vitro experiments and we build a real-time composite model of vascularized organs such as the liver. The model takes into account separate constitutive laws and separate numerical representations for the parenchyma and vessels: the parenchyma uses a volume deformation model whereas the vessels are modeled using beam elements. The method provides a coupling between them. A numerical validation is provided by comparing the results with a standard, already converged, FEM simulation.

**Vessel Model**  As a first step towards a mechanical model of vessels, a continuous representation of vascular structures is constructed. Using segmented data of a vessel tree, points along centerlines of each vessel branch are selected. This task can be performed either manually or automatically, e.g. with VMTK. Each branch can be represented as a series of cubic Bezier curves fitted to the centerline points: the series begins at a starting point of the branch and ends either at another branching point or at the endpoint of the branch. The interpolated points are used as nodes of serially linked beam elements, similarly as proposed in section 1.3. The model takes into account the particular nature of vessels through specific cross section profiles and moments of inertia.
Mechanical coupling between vessel and parenchyma  The parenchyma is discretized with tetrahedral mesh and deformed using standard FEM. Since no relative motion between the vessels and parenchyma is observed in reality, the interaction between the two structures can be modeled with an holonomic constraint: Thus, we can reduce the DoFs of the system by constraining the vessel node positions to follow the displacements of the parenchyma. Reciprocally, the force contribution due to the deformation of the vessel is propagated to the parenchyma. The mapping of forces is based on the principle of virtual work.

Each node of the volume model has 3DoFs whereas each node of the beam elements has 6DoFs (3 translations and 3 rotations). Consequently the mapping is based on a polar decomposition of the tetrahedron motion, as illustrated in figure 1.24:

![Figure 1: Mapping between 6DoF beam node and tetrahedron in initial and rotated positions.](image)

**In vitro Measurements of Vascularized Sample Response**  While some studies on liver tissue have been conducted in the past, few actually focused on the evaluation of mechanical properties of hepatic veins. In Umale et al. (2011) tensile tests on porcine hepatic veins are presented. Nevertheless there is no experiments evaluating the influence of vascularization on mechanical behavior and elastic response of the soft tissue. To address this gap, we conducted a series of tensile tests on fresh porcine liver on homogeneous and heterogeneous samples with single straight vessel.

The experimental setup is shown in Fig. 1.25(Left). Obtaining reliable measurements was particularly challenging as during loading, either tissue damage or glue failure occurred in many samples. Only three representatives results were selected for the quality of the experiment (one homogeneous sample and two heterogeneous). The apparent Young modulus measured on the three samples (3.5 kPa without vessel, 8.54kPa with a vessel of diameter 5 mm and 14.41kPa with a vessel of diameter of 8 mm), indicates that the presence of the vein inside the sample affects significantly
1.5. VOLUME DEFORMATION OF SOFT-TISSUES

The increase in apparent stiffness seems proportional to the size of the vessel, even if the number of good experiments were too small to have statistics.

**Numerical Simulation of Vascularized Tissue** We performed several numerical tests to validate the coupling method and compare with experiments. In the simulation, the parenchyma is modeled using a corotational non-linear elastic model (most studies agree on a viscoelastic behavior but we are focusing on the static equilibrium under some specific loading conditions). For modeling the vessel, we compared two different numerical strategies:

- the first is *conventional* and consist in using graded tetrahedral meshes having 37,000 and 52,000 elements to discretize the thin vessel wall correctly (samples with smaller and larger vessel, respectively). Parenchyma and vessels are discretized in a unique mesh but with different stiffness. The Young modulus is based on estimated values from experiments 3.5 kPa for the parenchyma vs. 1.1MPa and 1.4 MPa (respectively) for the small and the large vessel walls (see Figure 1.25 middle)

- the second uses the method of coupling between tetrahedral elements for the parenchyma and beam elements for the vessel walls. We tested with a mesh of 1160 elements for the parenchyma and 2 beams for the vessel. We did a second test with only 170 elements for the parenchyma and still 2 beams for the vessel (see Figure 1.25 right).

The objective is to validate the numerical accuracy of the composite model. The number of elements used in the first strategy corresponds to a solution for which FEM has converged. It can be used as a reference to verify that the second strategy is accurate. For the mesh with 1160 elements, the relative errors between the force responses computed by standard and composite FEM was 6.5% and 2.45% for smaller and larger vein, respectively. For the very simplified mesh with 170 elements, the relative error in force responses of 9% and 4.6%. The accuracy of the composite model is emphasized by the fact that the error was significantly lower for samples with large
Figure 1.26: Vascularized model of liver: (a) initial position with mapped vascular trees, (b)(c) deformation of liver (stiffness 3500 kPa and 15000 kPa, respectively) under gravity (green: non-vascularized model, red: vascularized model)

vascularization that has more important influence on the overall elastic response.

In order to demonstrate the efficiency of the algorithm, we performed a simulation of entire liver: here, the parenchyma was discretized into 2620 tetrahedra and two vascular trees were modeled inside the liver: hepatic vein (composed of 257 beams) and hepatic portal vein (modeled with 57 beams). The tetrahedral mesh was fixed on surface close to the entrance of the hepatic vein. Gravitational loading was applied on the liver, resulting in large displacements. Two different simulations were performed: first, the liver was modeled without any embedded vascularization, second, both hepatic vein and portal hepatic vein were included in the composite model. For both cases, refresh rate of 60 FPS was achieved on PC with CPU Intel CPU i7 running at 2.00 GHz. To help the convergence, we make use of the preconditioner presented above. The resulting deformations are depicted in Fig. 1.26, showing an important difference for vascularized and non-vascularized organ.

The experiments presented in the evaluation section confirm that vascular structures play an important role in tissue behavior. However, we demonstrate that we have a numerical solution, compatible with real-time, to simulate the influence of vascular tree in the mechanical behavior of the liver.

Conclusion: We have highlighted three numerical strategies for the simulation of volume soft-tissues in real-time: Two are dedicated to the solver with a multigrid approach and the asynchronous preconditioner technique. The last one is a numerical coupling of the vascular stiffness with the soft-tissue models. The problem of modeling soft-tissue deformations is still very active: we have recently started a Ph.D. (Julien Bosman) on modeling the connective tissues. Moreover, in the context of the collaboration with the Laboratoire de Mécanique de Lille (Ph.D. of Zhifan Jiang and Mouhamadou Diallo) we will develop models based on more complex constitutive laws and inverse simulation tools to retrieve the parameters from patient data.
Chapter 2

CONSTRAINT-BASED MODELING OF BIOMECHANICAL INTERACTIONS

Table of contents

2.1 Related work .................................................. 49
2.2 Computing non-smooth mechanics in real-time .................... 54
  2.2.1 Contact and Friction laws .............................. 54
  2.2.2 Time stepping for non-smooth dynamics of deformable objects .. 56
  2.2.3 Constraint-based response .............................. 57
2.3 Compliance computation ........................................ 59
  2.3.1 Compliance of curved and flexible instruments modeled with beam 60
  2.3.2 Compliance approximation for surface and volume models .......... 61
  2.3.3 Compliance computation based on the Asynchronous Preconditioner approach ........................................ 64
2.4 Constraint-based contact response ................................ 69
  2.4.1 Contact mapping ........................................ 70
  2.4.2 Volume contact Constraints at Arbitrary Resolution ............. 73
2.5 Surgical interaction modeled by complementarity constraints ........ 77
  2.5.1 Constraint-based interaction model ........................ 78
  2.5.2 Unified processing of constraints .......................... 80
  2.5.3 Simulation results of needle and suture interactions with soft-tissues 81
To accurately model soft tissue deformations, the simulation must account for the intrinsic behavior of the targeted organ, but also for its biomechanical interactions with surrounding tissues or with medical devices. For instance, while the biomechanical behavior of the liver has been well studied, few works exist regarding the mechanical interactions between the anatomical structures and more generally to the boundary conditions. For tissue-tool interactions, most approaches rely on a simple contact models, and rarely account for friction. While this simplification can produce plausible results in the case of an interaction between the end effector of a laparoscopic instrument and the surface of an organ, it is generally an incorrect approximation. As we move towards simulations for planning or per-operative guidance, accurately modeling the interactions with the medical devices will take an increasingly important place. In laparoscopic surgery, the main challenge lies in the modeling of interactions between anatomical structures as well as between the instruments and the surface of an organ. During the different steps of a procedure, organs slide against each other or are linked by connective tissues, while respiratory, cardiac and patient motion also generate permanent interactions.

**Needs and contributions:** The context of real-time execution often leads to strong simplifications in the modeling: usual real-time simulations are composed of a single organ with simple boundary conditions that does not take into account all these existing interactions. Yet, if the boundary conditions of the deformable model are ill-defined, it could lead to important errors on the mechanical behavior, that would ruin the benefit of using accurate deformable models such as FEM. This aspect is often ignored or underestimated in the previous work related to real-time simulation. But, there are strong numerical issues related to this topic: modeling and solving these interactions is related to non-smooth mechanics domain (due to sudden changes on the velocity fields). This domain requires dedicated numerical schemes that are presented in the context of real-time simulation in this chapter.

The processing of these multiple interactions must include the mechanical behavior of the tissues. For instance, when the liver is pushed by the diaphragm, the contact response on the other side (with the stomach or on the bare area) is more impacted if the liver is stiff (liver cirrhosis for instance). This aspect is particularly challenging and we have several contributions for obtaining, through the compliance operator, a measure of the mechanical coupling between the interactions. Moreover, we have introduced a generic use of Lagrange multipliers in the context of the SOFA framework and have adapted the processing to obtain the fewest possible number of contact constraints when dealing with collision response. We have also extended the use of complementarity constraints, popularized for contact modeling, and introduced this modeling for other type of interaction like the insertion of a needle or a suture in soft tissues.
2.1 Related work

Modeling the various mechanical interactions that take place during a procedure is related to many different domains including collision detection and response, non-smooth mechanics and constraint-based simulation. Again, it is not the goal here to present an exhaustive survey of the related work. In contrast, if we would limit this overview to the works dedicated to this topic in the real-time medical simulation domain, the number of existing work would be very low. Moreover, the numerical solution provided would be often simplistic. Consequently, this short survey firstly cites the main groups of existing methods independently of their application domain. Secondly, it focuses on simulating some procedures (interventional radiology, needle insertion, suture) for which the choice of the interaction model is overriding.

While keeping the context of simulating procedures on the liver, we can illustrate the importance of providing methods and models to simulate the mechanical interactions on the liver model: interaction with the surrounding anatomical structures, interactions due to a suture or the insertion of a long needle or interactions with other surgical instruments (collision, contact eventually cutting) such as catheters and guides used in interventional radiology or laparoscopic grasper.

Simulating these interactions necessitate (i) to detect them, through the use of a collision detection algorithm, (ii) to model them through the use of dedicated laws, (iii) to solve them with adapted numerical methods. We will get into the related works on these 3 challenging topics (independently from the applications) before concentrating on the existing work on the simulations of (iv) interventional radiology and (v) suture and needle simulation.

Collision detection: The amount of work done on collision detection alone is quite large. Collision detection typically involves hierarchical bounding volume data struc-
tures, such as bounding boxes Gottschalk et al. (1996) or spheres Hubbard (1995). Alternatively to collision detection, some works have proposed algorithms that compute the minimal distance(s) between objects like, Lin et Canny (1991), Fisher et Lin (2001), Johnson et Willemsen (2003). In addition to collision for rigid object, significant work has been done to address deformable models: van den Bergen (1997), James et Pai (2004) and self-contact Volino et Magnenat-Thalmann (1995), Provot (1997). See the survey by Teschner et al. (2005) for an excellent overview of different methods. More recently, techniques which exploiting the GPU have been proposed as this can result in important speed improvements through parallelization Vassilev et al. (2001); Heidelberger et al. (2003, 2004); Baciu et Wong (2004); Wong et Baciu (2005); Otaduy et al. (2004); Govindaraju et al. (2007).

Collision detection returns either penetration depth or a minimal distance between meshes, or pairs of primitives (between which distance constraints can be formulated to resolve the contacts) or interpenetration volumes (Heidelberger et al.). Building a contact formulation based on interpenetration volume is less obvious. This is the subject of a contribution presented after. Continuous Collision methods are also able to provide the time and the geometrical configuration when objects are colliding (Redon et al. (2003), Govindaraju et al. (2007) Tang et al. (2008b).)

Contact response models: The Signorini’s law is known in continuous media mechanics (see for instance N. Kikuchi (1988)), to deal with contacts between deformable bodies. Moreau popularized this model and extended it to dynamic rigid objects collisions Moreau (1966). Moreover, he introduced the use of quadratic programming for solving it. He also introduced the use of variational inequalities to solve the problem with tangential friction laws, like Coulomb’s law, to reproduce static and dynamic friction Moreau (1974).

Solving Signorini’s and Coulomb’s laws on deformable models often requires a huge amount of computation. In a context of interactive simulation, an alternative is to use simplified approaches, based on penalty method. This method uses the geometrical criterion of interpenetration, coupled with a contact spring to compute the contact forces necessary to avoid objects interpenetration. This method has been widely used in computer graphics and in medical simulation context Fisher et Lin (2001); Deguet et al. (1998); Meseure et al. (2003) However, this method, although quite efficient, suffers from many limitations. The choice of the spring stiffness is very tedious, and is even more difficult when multiple bodies interact. The springs have no physical meaning and create inconsistent forces and instabilities. Using implicit integration, it leads to non-linear stiff problem. Some of these problems have been overcome in a very recent work Tang et al. (2012) (which is based on SOFA !)

While it is possible to solve contact response using penalty method, it has long been recognized that there are advantages to formulating contact with constraints. The contact constraints are unilateral and result in a linear complementarity problem (LCP).
The addition of Coulomb friction involves non-linear constraints relating the tangential force to the normal force and results in a non-linear complementarity problem (NLCP). Most work in computer graphics uses a pyramid discretization of the friction cone to formulate the problem as an LCP Baraff (1991); Milenkovic et Schmidl (2001), but it is also possible to use the exact cone (a nonlinear complementarity problem) and to compute the solution using iterative methods Alart et Curnier (1991) Jourdan et al. (1998). Anisotropic friction can also be simulated Pabst et al. (2009). While formulating the frictional contact equations at the acceleration level requires the use of event-driven method (the integration is stopped at the time of impact), some related papers have introduced the use of time-stepping approaches, with a velocity level formulation Stewart et Trinkle (1996); Anitescu et al. (1999); Liu et Wang (2003). A velocity formulation is also preferred since it avoids the Painlevé paradox on rigid bodies. But position based simulation could be also relevant in the context of deformable bodies or quasi-rigid objects Duriez et al. (2006b).

The friction contact problem can result in an NP-hard combinatorial problem to identify the active constraints Baraff (1991). Some work solves the problem using a sequence of Quadratic Programming (QP) problems Kaufman et al. (2008), or by reconstituting the temporal sequence of contact Harmon et al. (2009). Baraff et Witkin (1998) and Otaduy et al. (2009) use implicit contact constraints to simulate contact response in cloth simulation. LCPs have also been used for fluid-solid coupling Batty et al. (2007). Other relevant work on friction includes fast approximate models in video game systems Parker et O’Brien (2009).

This type of approaches are also proposed in a biomechanics context: Delp (2001) models the contact between implants using complementarity constraints, in the context of dynamic simulation of total knee replacement motion. Miller et al. (2005) propose a framework to simulate grasping with the goal of creating a biomechanically realistic human hand model. More recently, Stavness et al. (2011) propose this approach to simulate the complex coupled jaw-tongue-hyoid dynamics.

**Solving process** The methods used to solve contact equations can be classified into several categories:

- **(N)LCP methods vs. QP methods:** (N)LCP methods solve the contact problem in the constraint space: Using a technique that can be assimilated to domain decomposition method, each contacting objects are firstly solved separately and then, the (N)LCP process solves the contact equations between adjacent objects Stewart et Trinkle (1996); Duriez et al. (2006b); Otaduy et al. (2009). QP methods build and solve a single system of equalities and inequalities (equalities for the mechanical models of the objects and inequalities are used for contact and friction) Redon et al. (2002); Kaufman et al. (2008). However, in the context of interactive simulation of contacting Finite Element models with thousands of degrees of freedom, it is more efficient not to build a single system with all the objects in contact.
• **Direct solver vs. Iterative solver:** NLCP can be solved using Gauss Seidel and/or Newton iterative algorithms [Alart et Curnier (1991); Jourdan et al. (1998)] whereas Lemke’s direct method can be used to solve LCP and necessitates a simplification of the friction cone [Anitescu et al. (1999); Lloyd (2005)]. For QP, both direct method (for instance active-set strategy) and iterative methods (special implementation of conjugate gradient) can be used. [Kaufman et al. (2008)] use a QP direct solver, but alternatively solve nonpenetration and friction constraints to avoid scaling and non-convexity problems associated with traditional direct methods, while improving on the slow convergence that can be typical with iterative methods.

• **Time-stepping vs. Event-driven:** When solids come into contact, there is a jump in the velocity of the contacting points. Therefore, the acceleration of these points is not defined anymore and the collision event enters in the non-smooth mechanics. Event-driven techniques stop the time integration at the time of impact [Redon et al. (2003)] whereas time-stepping techniques rely on a constant time integration and a formulation of the mechanics using velocity variation and impulse [Stewart et Trinkle (1996); Anitescu et al. (1999)]. Event-driven methods are usually more precise but require continuous collision detection and are adapted to simulation with a few number of collision events. The accuracy of the time-stepping scheme depends on the size of the time step but are adapted to discrete collision detection. To reinforce the precision, especially to avoid strictly the interpenetrations, position-level constraints can also be used directly [Duriez et al. (2004b)] or in a post-stabilization framework [Cline et Pai (2003)].

Note that in [Duriez et al. (2004b)], position-level constraints necessitate the use of algorithms able to evaluate the minimal distance(s) between contacting objects. In general, even if the problem of collision detection and response are usually addressed separately in related work, there is often a strong link between them.

After this short survey of existing collision detection and response algorithms, we now focus on specific related work: the simulation of interventional radiology and the interaction models used for suture and flexible needle insertion.

**Interventional radiology simulation** A first approach in modeling the catheter was in introduced by [Lawton et al. (2000)] using thin rod elements and *wall potential* model. This type of approach, derived from penalty models are also used in [Nowinski et Chui (2001) and in Alderliesten et al. (2004)]. The penalty approach is justified by the local elastic behavior of the vessel walls and the fact that the vessel can be considered as more rigid than the catheter. It is also chosen in [Wang et al. (2007)] for its low computational expense in a real-time simulation context.

We first proposed to rely on constraint-based approach for contact response in [Duriez et al. (2006a)] using an optimized iterative LCP approach described in the following. In [Dequidt et al. (2007)], the collision detection is optimized using an implicit surface representation of the vessels. The work is then extended to include self-collisions
for the simulation of intracranial aneurysm treatments by coil embolization Dequidt et al. (2009). A Gaussian deformation, close to the Boussinesq approximation model, allowed to take into account the very small deformation of the aneurysm wall. It will be detailed in the section 2.4 of this chapter. A similar approach was proposed by Spillmann et Harders (2009) that uses inelastic rods to model the behavior of the interventional radiology instruments.

**Needle insertion and suture simulation** Even if flexible needle insertion and suture simulations have common modeling problems (i.e. interaction between 1d model and soft tissues), they are often treated separately in papers. Suture simulation have additional modeling problems: the large deformations of the thread, the potential self-collisions and the knot tying.

Pioneering works concerning needle insertion were presented by DiMaio et Salcudean (2003, 2005); Alterovitz et al. (2005). A recent survey by Abolhassani et al. (2007b) summarizes existing methods. In general, the interaction between needle and tissue combines different physical phenomena. Three different forces are often associated to the insertion of a needle in soft tissues: puncture force, cutting force and friction force. Several studies have identified these forces through a series of experiments Simone et Okamura (2002); Okamura et al. (2004), and were used to parameterize existing models Crouch et al. (2005); Dehghan et al. (2008). Remeshing process of tissue models remains a difficulty to obtain interactive models. Furthermore, Marchal et al. (2007) proposes to rely on a discrete meshless model that is compatible with interactive frame rates. Some studies have been presented for needle steering Abolhassani et al. (2007a); Tang et al. (2008a).

Two approaches, simultaneously published at Siggraph and Miccai in 2009 have proposed to use constraint-based modeling to capture the mechanical interaction between flexible needle and soft-tissues. The first, proposed by Chentanez et al. (2009) uses stick slip friction models. The solving process is based on an active sets algorithm and necessitates the resolution of multiple conjugate gradients before finding the correct configuration. Our approach, proposed in Duriez et al. (2009), is based on several complementarity constraints to capture puncture, cutting and friction forces. No remeshing is needed and the solving process, based on a generic NLCP iterative solver, is usable at interactive rates. This approach will be detailed in section 2.5.

A short state of the art on modeling the deformations of a suture thread is available in section 1.1.3. Here, we focus on the mechanical interactions.

As for previous work directly related to suturing simulation, LeDuc et al. (2003) proposed an interaction model where the suture can pierce the mesh only at vertices, which can then slide with friction from one suture node to the next. Similar approaches were used in later studies by Zhang et al. (2007). Nageotte et al. (2005a) build a path planner based solely on a geometrical modeling of a suturing task. A few other works consider the subdivision of the mesh at the collision point Marshall et al.
Knots tying is an important aspect of the medical suturing procedures and requires tedious learning. Previous works extensively studied the interaction required for knot-tying Phillips et al. (2002); Lenoir et al. (2002); Spillmann et Teschner (2008). Suture simulations integrating such capability of tying knots with surgical thread are described in Brown et al. (2004); Marshall et al. (2005); Shi et Payandeh (2008) for instance. These approaches either do not use physically-based deformable models, or lower their accuracy in order to achieve interactive rates. Recently, Bertails-Descoubes et al. (2011) have proposed to rely on non-smooth mechanics to model thin elastic rods in contact, in the presence of friction. A very stable demonstration of knot tying is demonstrated, but not in real-time.

2.2 Computing non-smooth mechanics in real-time

Collision response on mechanical objects leads to discontinuities in the velocities of the colliding points. For such discontinuous event, the acceleration is not defined: the problem belongs the field of non-smooth mechanics. In the following section, we present the time-stepping approach that we are using when we need to integrate such events in the context of real-time simulation, especially on deformable bodies. With time-stepping, we get constant time steps, which is particularly important in the context of real-time applications. We combine it with a low order numerical integration based on backward Euler with a single linearization per step (already presented in section 1.2). Finally, we use a constraint-based approach, based on solving complementarity problems, to process the collision response or any other interaction that enters in the field of non-smooth mechanics.

2.2.1 Contact and Friction laws

We are using the Signorini’s law to solve a set of contacts between two bodies named $D_1$ and $D_2$. Surfaces $S_1$ on $D_1$ and $S_2$ on $D_2$ are defined as part of boundaries where particles are potentially in contact. For each particle $P_1$ of $D_1$ a neighbor particle $P_2$ of $D_2$ is associated to test the contact between $D_1$ and $D_2$ (see figure 2.1 left ). Let $f_n^{(1)}(P_1)$ be the contact force exerted on $D_1$ in $P_1$ by body $D_2$ in the direction of contact $n$. The action/reaction principle gives: $f_n^{(1)}(P_1) = -f_n^{(2)}(P_2)$. We use $f_n^{(1)}(P_1) = \lambda_n$ in the following. The normal $n$, chosen arbitrarily, is directed toward the inside of $D_1$. The gap between the two objects at $P_1$ is: $\delta_n(P_1) = P_2P_1 \cdot n$. In

\footnote{It could have been the direction $-n$ of $P_1P_2$. By the arbitrary choice of the direction $n$, the problem is solved considering the unknown forces applied at points $P_1$ on $D_1$. It is exactly the same by using the opposite direction and taking the unknown forces applied at point $P_2$ on $D_2$.}
2.2. COMPUTING NON-SMOOTH MECHANICS IN REAL-TIME

Figure 2.1: Contact and Friction laws: (Left) contact between deformable objects, (right) Coulomb’s friction law

the following we use \( \delta_n = \delta_n(P_1) \). The Signorini contact model indicates that there is a complementarity relation\(^2\) between this gap \( \delta_n \) and the contact force \( \lambda_n \), that is:

\[
0 \leq \delta_n \perp \lambda_n \geq 0 \quad (2.1)
\]

This model has several physical justifications including non interpenetration and no sticking force. Moreover the contact force is null if the points are not strictly in contact. Dynamic problems often use a velocity formulation of this law. However, this formulation is valid only during the time of contact:

\[
0 \leq \dot{\delta}_n(t) \perp \lambda_n \geq 0 \quad \text{if} \quad \delta_n(t) = 0 \quad (2.2)
\]

\( \dot{\delta}_n(t) \) describes the relative velocity between \( D_1 \) and \( D_2 \) along \( n \) at the contact point. Using Signorini’s law, the contact space is along the normal and creates a frictionless response. Coulomb’s friction law describes the macroscopic behavior in the tangent contact space. In this law, the reaction force is included in a cone which height and direction is given by the normal force (see fig 2.1 - right). If the reaction force is strictly included inside the cone, objects stick together, otherwise, the reaction force is on the cone’s border and objects are slipping along the tangential direction. In this last case, the friction force must be directed along the direction of motion.

\[
\begin{align*}
\delta_{\tau} = 0 \Rightarrow \|\lambda_{\tau}\| &< \mu \|f_{\tau}\| \quad \text{(stick)} \\
\delta_{\tau} \neq 0 \Rightarrow \lambda_{\tau} = -\mu \|\lambda_n\| \frac{\delta_{\tau}}{\|\delta_{\tau}\|} = -\mu \|\lambda_n\| \vec{T} \quad \text{(slip)}
\end{align*}
\]

During 3D slipping motion (also called dynamic friction), the tangential direction is unknown. We only know that the tangential force and the tangential velocity are opposite along a direction that is to be found. It creates a non-linearity in addition to the complementarity state stick/slip. Signorini’s law and Coulomb’s law are also valid in a multi-contact case. However, to solve these laws on every contact, we have to consider the coupling that exists between them. This coupling comes from the intrinsic mechanical behavior of deformable objects.

\(^2\) Noted \( \perp \), this relation state that one of the two values \( \delta_n \) or \( f_n \) must be null.
2.2.2 Time stepping for non-smooth dynamics of deformable objects

Let’s suppose that the deformations are non-linear and rely on the Finite Element Method (as we described in chapter 2) or on any other physics-based model (including rigid bodies or articulated rigid bodies). The dynamic behavior of mechanical solids leads to a system of differential equations. In section 1.2.1, we propose a time-stepping scheme, based on backward Euler integration and a single linearization of the internal forces per time step. When using this scheme we obtain the following system:

\[
\begin{pmatrix}
M + h \frac{\delta F}{\delta v} + h^2 \frac{\delta^2 F}{\delta q^2} \\
\end{pmatrix} \begin{pmatrix}
v \\
\end{pmatrix} = -h^2 \frac{\delta F}{\delta q} v_i - h (f_i + p_f) + h \mathbb{H}^T \lambda
\]  

with the mass \(M\) and the internal forces \(F(q, v)\) derived from the physics-based model. \(q \in \mathbb{R}^n\) is the vector of generalized degrees of freedom and \(v \in \mathbb{R}^n\) is the vector of velocities. \(f_i\) is the value of \(F\) with the positions and velocities at the beginning of the time step \(h\) whereas \(p_f\) represent the external forces at the end of the step. Finally, \(\mathbb{H}^T\) is a matrix that contains the direction of the \(m\) constraints and \(\lambda \in \mathbb{R}^m\) is the vector of contact forces, that we aim at computing.

From collision or proximity detection, we have a set of potential contact spots \(\alpha = 1...n_c\) and we can find their associate frame \(\mathfrak{F}_\alpha = [n_\alpha, t_\alpha, s_\alpha]\). In that space, we will measure the relative displacement \(\delta_\alpha\) and velocity \(\dot{\delta}_\alpha\) between colliding objects in order to use contact and friction laws. For every contact between two object, we can build a mapping function \(A\) that links the positions in the contact space to the motion space:

\[
\delta_\alpha = A_\alpha(q_1, t) - A_\alpha(q_2, t) \tag{2.5}
\]

with \(A_\alpha(q, t)\) the mapping function which depends on the contact \(\alpha\) and the positions \(q_1\) and \(q_2\) of the two colliding objects. To obtain a kinematic relation between the two spaces (contact, motion), we use a linearization of equation (2.5). Note that we have developed an original and very generic technique, in SOFA, to map the collision points that is described in section 2.4.2.

If \(\mathbb{H}_\alpha(q) = \frac{\partial A_\alpha}{\partial q}\), we obtain, at time \(t\) for each contact:

\[
\dot{\delta}_\alpha(t) = \mathbb{H}_\alpha(q_1) v_1(t) - \mathbb{H}_\alpha(q_2) v_2(t) \tag{2.6}
\]

where \(\mathbb{H}\) corresponds to the transposed of the matrix used in the equation 2.4. For simplicity, we often take the direction of contact based on the positions at the beginning of the time step. We suppose that this matrix does not change during the contact response process: \(\mathbb{H}^T \lambda_f = \mathbb{H}(q_1)^T \lambda_f\). In the following, this matrix \(\mathbb{H}(q_1)\) is noted \(\mathbf{H}\) to emphasize that it is constant during the time step.
2.2.3 Constraint-based response

In the following, we present how the laws of contact (2.2) and friction (2.3) are solved while taking into account the dynamic equation (1.7) between 2 contacting objects. To solve these laws, we use a Lagrange Multipliers approach and a single linearization by time step. For both interacting objects we applied the equation 1.7:

\[
\begin{align*}
A_1 dv_1 &= b_1 + hH_1^T \lambda \\
A_2 dv_2 &= b_2 + hH_2^T \lambda
\end{align*}
\] (2.7)

In order to solve \( \lambda \) the process is performed by the following steps.

**Step 1**: interacting objects are solved independently while setting \( \lambda = 0 \). We obtain what we call the free motion \( dv_1^{\text{free}} \) and \( dv_2^{\text{free}} \) for each object. After integration, we obtain \( q_1^{\text{free}} \) and \( q_2^{\text{free}} \).

**Step 2**: the constraint laws are linearized:

\[
\delta = A_\alpha (q_1^{\text{free}}) - A_\alpha (q_2^{\text{free}}) + hH_1 dv_1^{\text{cor}} + hH_2 dv_2^{\text{cor}}
\] (2.8)

With \( dv_1^{\text{cor}} \) and \( dv_2^{\text{cor}} \) being the unknown corrective motions when solving equation 2.7 with \( b_1 = b_2 = 0 \). When gathering equations 2.7 and 2.8, we have:

\[
\delta = \delta^{\text{free}} + h^2 \left[ H_1 A_1^{-1} H_1^T + H_2 A_2^{-1} H_2^T \right] \lambda
\] (2.9)

Together with Signorini’s law (equation 2.1), we obtain a LCP. If it is combined with Coulomb’s law (equation 2.3), a NLCP (Non-linear complementarity problem) is obtained.

If the collision detection is based on minimal distances algorithms, position-based formulation can be used, because the potential contacts can be set before being active in the simulation. However, if a discrete collision algorithm is used, based on interpenetration detection, it has been observed that an approach based on velocities is more stable (sometimes combined with a post-stabilization on positions). In that case, we use:

\[
\dot{\delta} = H_1 v_1^{\text{free}} - H_2 v_2^{\text{free}} + h \left[ H_1 A_1^{-1} H_1^T + H_2 A_2^{-1} H_2^T \right] \lambda
\] (2.10)

We obtain the value of \( \lambda \) using a Gauss-Seidel algorithm dedicated to the NLCP created by contact and friction equations. Considering a contact \( \alpha \), among \( m \) instantaneous contacts, one can write the behavior of the model in contact space:

\[
\frac{\delta}{\text{unknown}} - W_{\alpha \alpha} \lambda_{\alpha} = \sum_{\beta=1}^{\alpha-1} W_{\alpha \beta} \lambda_{\beta} + \sum_{\beta=\alpha+1}^{m} W_{\alpha \beta} \lambda_{\beta} + \delta_{\alpha}^{\text{free}}
\] (2.11)
where $W_{\alpha,\beta}$ is a compliance matrix that models the coupling between contact points $\alpha$ and $\beta$. For each contact $\alpha$, this method solves the contact equations by considering the others contact points ($\alpha \neq \beta$) as "frozen". The new value of $\lambda_\alpha$ is given by solving the Signorini’s law and the Coulomb’s law on this contact (see Duriez et al. (2006b) for details). The algorithm will be generalized to other type of constraints in section 2.5.

**Step 3**: When the value of $\lambda$ is available, the corrective motion is computed:

$$
\begin{align*}
    x_{1,t+h} &= x_{1}^{\text{free}} + h\mathbf{d}v_1^{\text{cor}} & \text{with} & \quad \mathbf{d}v_1^{\text{cor}} = A_1^{-1}H_1^T\lambda \\
    x_{2,t+h} &= x_{2}^{\text{free}} + h\mathbf{d}v_2^{\text{cor}} & \text{with} & \quad \mathbf{d}v_2^{\text{cor}} = A_2^{-1}H_2^T\lambda
\end{align*}
$$

After applying this correction, the state of the two objects fulfills the constraints expressed on each contact point. For instance, we have proven (for instance in Duriez et al. (2008)) that this approach can be used to simulate physical grasping using friction-contact equations.

**Discussion**: This scheme is particularly well-suited for contact between deformable models: It can handle multiple contacts, each model is being computed separately during the free motion, and more importantly, each model can use a dedicated solver that is adapted to the structure of its matrix $A$. This last feature is particularly relevant when using wire-like instruments modeled with beams (like described previously in section 1.3), for which $A$ is Block-Tri-Diagonal. This aspect is detailed in section 1.3.2.

The approach has also several limitations:

- The contact directions are constant during the time-step, which could create undesirable behaviors. Yet, without proposing a general framework, we have solved this problem in the particular case of rigid contact on a model represented by an implicit surface issued from a CT-scan in Syllebranque et Duriez (2010) (see section ??).

- **Free motion** can be very far from final motion if contact constraints are strong. We have proposed a variation of this scheme in Peterlik et al. (2011b) by replacing the free motion by a predictive motion, that is based on an estimate of $\lambda$ based on the previous time steps.

- Equations 2.9 and 2.12 involve the inverse of matrix $A$, which changes at every time step. In particular, obtaining $W = HA^{-1}H^T$ in real-time is doubly challenging as soon as the objects involve more than a few hundred nodes and/or if the collision response involves more than hundred simultaneous contacts. This last point is a real scientific lock and in a real-time context, it is very often the bottleneck of the approach. I have devoted a large part of my research work on this problem and some results are presented in the following.
2.3 Compliance computation

When using non-linear models of deformation, it is a big challenge to get, in real-time, the compliance matrix $W$ (that involves the inverse matrix of the system). Yet, this matrix is necessary to build the (N)LCP before solving the contact constraints.

First, we introduce a solution dedicated to catheter navigation in blood vessels, that relies on the specific structure (Bloc Tri Diagonal) of the beam model matrix. This work was mainly developed during my post-doc and was published in Cotin et al. (2005) and in Duriez et al. (2006a).

Then, we present some approaches that are based on an approximation of the compliance for modeling the vessel walls Dequidt et al. (2009) and for modeling volume deformable model, using the compliance warping method Saupin et al. (2008b). This approach, that keeps a precomputation step but introduces a correction based on the current configuration, was developed by Guillaume Saupin during his Ph.D. and extended to cutting in Courtecuisse et al. (2011b).

Finally, a more recent work, based on an asynchronous update of the compliance matrix, was developed by Hadrien Courtecuisse during his Ph.D. (Courtecuisse et al. (2011a)). This approach benefits from a parallel computation on GPU and provides both fast and accurate results. It brings a pertinent answer to this deadlock.

Obtaining the inverse matrix projected in the contact space $W = h^2HA^{-1}H^T$ is particularly challenging in a real-time context, when $A$ is changing at every step. Yet, this matrix brings an important information during the solving process of the contact constraints: For a force $\lambda_\beta$ applied on a given contact $\beta$, the induced displacement on any other contact $\alpha$ is obtained through a very simple operation: $\delta_\alpha = W_{\alpha \beta} \lambda_\beta$. This type of operation is needed in the Gauss-Seidel algorithm (see equation 2.23) that we used for solving contact constraints, but it would also be necessary for using a direct LCP solver like the Lemke’s algorithm.

As it provides a direct link between contact forces and induced displacement, $W$ is called a compliance matrix in this manuscript. In the literature this matrix is also called Delassus operator or flexibility matrix.

In the following, we present an optimization for computing $W$ when using a beam model on wire-like instruments. The computation reduces to a linear complexity for both compliance computation and iterative solving process of the NLCP. Second, we define a new contact model based on an approximation of the compliance matrix, that still relies on accurate contact and friction laws. Some examples of approximation of $W$ are presented, for both surface models (i.e. vessel walls) and volume models (i.e. soft tissues). Finally, we present a very promising approach based on the Asynchronous Preconditioner presented above which allows for both accurate results even on inhomogeneous tissues and real-time computation thanks to GPU implementation. This approach could be used for any type of deformable model.
2.3.1 Compliance of curved and flexible instruments modeled with beam

As presented in section 1.3, FEM beam model can be used to simulate the deformations of catheters or guidewires during interventional radiology procedure. A key point for simulating the navigation of these devices inside the blood vessels is the inclusion of the contacts with the vessel walls. An accurate collision response, based on Signorini’s and Coulomb’s laws, is particularly relevant for this type of application.

In the context of modeling curved and flexible instruments, we have presented, in section 1.3.2, a fast solver based on the Block-Tri-Diagonal (BTD) structure of matrix $A$. Indeed, the solver allows to compute the block-diagonal elements of matrix $A^{-1}$ with a linear complexity $O(n)$. However, the contacts can appear all the way along the catheter and are all coupled by the mechanical model of the device. Thus the matrix $W$ will be fully filled and the computation would possibly involve the whole matrix $A^{-1}$ (it is the case when the device is in contact over its entire length, which is far from being an unrealistic scenario). It leads to a quadratic complexity for obtaining $W$ before solving the NLCP. Moreover, the algorithm for solving the NLCP will also have (in the best case) a quadratic complexity.

To reduce the computation of building and solving the NLCP at the same time, we adapt the iterative algorithm based on a Gauss-Seidel scheme. In equation (2.23), the terms $W_{\alpha\beta}$ of matrix $W$ (with $\alpha \neq \beta$) appear in a frozen part of the equation. Only $W_{\alpha\alpha}$ is absolutely necessary when visiting each contact at each iteration. If $\alpha$ is a friction contact at node $n$ along the the normal and tangential directions $^nH_\alpha = \{n_\alpha, t_\alpha, s_\alpha\}$, this local compliance is:

$$W_{\alpha\alpha} = ^nH_\alpha (A^{-1})_{nn} ^nH_\alpha^T$$

(2.13)

The computation of $W_{\alpha\alpha}$ can be very efficient, since we can compute all the block diagonal elements $(A^{-1})_{nn}$ with a linear complexity (see equation 1.15). The frozen part of the equation (2.23) can be rewritten using three summations: the first one groups the other contacts that appear on the same node, the second groups the contacts that appear on the upstream nodes ($i \in [0...n]$) of the model and the third groups the contacts on the downstream nodes ($i \in [n...N-1]$) with $N$ the total number of nodes).

$$\delta_\alpha - W_{\alpha\alpha}\lambda_\alpha = ^nH_\alpha^T \left[ (A^{-1})_{nn} r_n + \sum_{i=1}^{n-1} (A^{-1})_{ni} r_i + \sum_{i=n+1}^{N-1} (A^{-1})_{ni} r_i \right] + \delta_{\text{free}}$$

(2.14)

computed by substructure decomposition

Where $r_i$ is the sum of the contact reactions on the node $i$, i.e. $r_i = \sum_\beta ^\beta H_\beta^T \lambda_\beta$. If the contacts are sorted in ascending order along the structure, then the computation of equation 2.14 can follow a similar algorithm than the one proposed in section 1.3.2: the contact forces are accumulated along the structure and the complexity of one iteration
2.3. COMPLIANCE COMPUTATION

Figure 2.2: Catheter navigation inside the vascular network. Reducing the complexity of the collision response algorithm allows for real-time simulation of complex configurations.

of the Gauss-Seidel like algorithm stays linear $O(M + N)$ for an identical result (if $M$ is the number of contact and $N$ the number of nodes.).

Very recently, the approach has been adapted for self-collision response (which is not effortless, because obtaining $W_{\alpha\alpha}$ involves then several non-consecutive blocks of matrix $A^{-1}$). But the algorithm has not yet been published.

The $W$ is the addition of the compliances of the two contacting objects. If we suppose that the vessels are perfectly rigid, they do not add any compliance in $W$. However, if we model the deformation of vessel walls, we need to compute their contribution to $W$.

2.3.2 Compliance approximation for surface and volume models

In a context where boundary conditions play a great role in the final results of the simulation, we have proposed a new approach for the contact response. The main idea is to provide an approximation for the computation of the compliance while keeping the maximum precision on the contact and friction laws. Somehow, it is the dual of the strategy used in penalty methods which simplify the contact and friction laws with no simplification of the deformation model.

Modeling contacts with vessel walls: In the context of interventional radiology, coil embolization is a standard procedure to treat aneurysms that appear in the vascular network. We have studied the coil embolization in a neurovascular context, but coil embolization can also be realized on many other parts of the vasculature, including the arteries of the liver Stambo et al. (2004). Simulating coil embolization requires to accurately model contacts that occur between the coil and the aneurysm wall. The contact model must provide the following features: first, to account for the stick and
slip transitions that take place during the coil deployment; second, to use the mechanical model of the instrument (i.e. the coil) and in some case the global displacements of the vessels; third, to include a local compliant behavior for the vessel wall that is really thin in an aneurysm.

The model presented above already accounts for the two first features. The stick slip transitions are provided by the friction contact model, and the contact response is based on the compliance of the beam model that can be used to capture the behavior of the coil Dequidt et al. (2008). The global motions of the vessels can be integrated during the free motion, to adapt the interpenetration measure $\delta^{\text{free}}$.

But as the compliance computation only comes from the coil, the contact model will not generate any additional deformations on the vessel walls. Yet, the aneurysm wall is particularly thin, and consequently the coil will deform it slightly and locally at the contact spots. To model this local compliance, we use an approach that is close to Boussinesq model.

Let’s consider a contact $\alpha$ on the surface of the aneurysm, and $H_\alpha$ the matrix of the contact frame $[n \ t \ s]$. The mechanical coupling of this contact with a contact $\beta$ (with frame $H_\beta$) can be evaluated with the following $3 \times 3$ matrix:

$$W_{(\alpha,\beta)} = \frac{g(d_{ij})}{e} H^T_\alpha H_\beta$$

where $e$ is an elasticity parameter that is homogeneous to young modulus and $g(d_{ij})$ is a Gaussian function of the distance, defined on the surface, between contact point $\alpha$ and $\beta$. The Gaussian function allows a fall-off of the coupling with increasing distance between the contact points. This model is close to the Boussinesq approximation which provides a distribution of the normal contact stress from the elasticity of the surface, around a point of contact (Pauly et al., 2004).

The global compliance matrix $W$ used to solve the contact is the sum of the coil compliance (provided by equation 2.14) and the aneurysm wall compliance. The results provided by this contact response are illustrated in figure 2.3. The compliance of the aneurysm wall improves the stability of the stick-slip transitions and the realism of the coil embolization. The overall processing of a simulation step, when including dozens
2.3. COMPLIANCE COMPUTATION

of friction contact points remains compatible with an interactive execution (around 30 frames per second for a coil with 100 beams). This is a first simple example showing that the compliance used for the contact response can be modeled with a dedicated model. This approach was published with more details in Dequidt et al. (2009).

Compliance Warping: The computation of the matrix in equation 2.9 requires the computation of matrix $W_i = h^2HA^{-1}HT$ for compliant object $i$. With deformable objects undergoing large displacements, $A^{-1}$ must be updated when $M$, $K$ and $B$ change according to the non-linear deformations. This computation is highly time-consuming, especially for volume model that have a large number of DOFs.

In Saupin et al. (2008b), we proposed to use a good approximation of the matrix $A^{-1}$ instead of its real value. This approximation uses a precomputed value in the rest shape configuration $A_0^{-1}$. To account for the geometrical non-linearities induced by the deformation, we evaluate a rotation at each node, from its rest shape to its current shape. We approximate the value of the compliance matrix $C \approx h^2A^{-1}$ using the following equation:

$$C = h^2RA_0^{-1}R^T$$  \hspace{1cm} (2.16)

Where $R$ is a $3 \times 3$ block diagonal matrix gathering the rotations associated with object nodes. If using a corotational model, the rotation evaluation is done using an average of the frames computed for the elements in the direct neighborhood of the point. Other approaches, like shape matching presented in Müller et al. (2005) can also be used to obtain a geometrical evaluation of the rotation at each nodes. This simplification significantly speeds up the computation of the matrix of Equation 2.9. The speed up is enforced by using the sparsity of matrices $R$ and $H$. This warping using $R$ is illustrated in figure 2.4.

A key point of the approach is to use the same approximation when applying corrective motion. Thus equation 2.12 becomes (for object 1, for instance):

$$x_{1,t+h} = x_{1}^{\text{free}} + C_1 H_1^T \lambda$$  \hspace{1cm} (2.17)

This guaranties the non-penetration and the stick-slip transitions of the contacting objects at the end of the time step. The results are illustrated in figure 2.5 with an example of a physics based grasping of the liver during a laparoscopic procedure.

GPU Implementation The constraint process still requires the computation of $W$ at each time step of the simulation. The size of this square matrix depends on the
number of constraints simultaneously involved. Its computation could quickly become a bottleneck of real-time execution when a lot of contacts are involved. Fortunately, the majority of required operations can be massively parallelized.

Thus, in Courtecuisse et al. (2011b), we have proposed a new GPU implementation of the compliance warping. The method was also extended to the case of topological changes (for instance, when cutting). The approach is based on a decomposition of all topological changes in 3 basic operations: element removal, element subdivision, element addition (see figure 2.6). These topological changes are mapped back to the initial configuration. The Sherman-Morrisson-Wodoebury formulas are used to update the values of $A_0^{-1}$ in order to maintain a coherent compliance model. The implementation of the formulas can be parallelized and were implemented on GPU. If an average of 2 nodes are concerned by topological changes at each step, 22 msec are needed at each step for the update on a mesh with 1,607 elements and 64 msec for a mesh with 3,874 elements. Moreover, in section 3.2, we will demonstrate the compatibility of the algorithm with haptic feedback.

2.3.3 Compliance computation based on the Asynchronous Preconditioner approach

The benefit of the compliance warping approach is to show that it is possible to get a precise collision response when solving Signorini and Coulomb laws using an approximation of the compliance matrix.
2.3. COMPLIANCE COMPUTATION

Even if the method is very efficient from a computational point of view, it has several issues: (i) its lack of accuracy when the geometrical non-linearities become large, (ii) its limitation to material with a linear constitutive law, (iii) the number of topological changes per time step must be very small to maintain real-time computation.

The main issues (on this method but also on the other existing alternative presented in Otaduy et al. (2009) based on a Gauss-Seidel like relaxation) come from the growing gap between the approximation $C$ and the actual compliance matrix $h^2A^{-1}$. If the approximation is not correct, the mechanical coupling between contact points that allows to compute an accurate collision response is no more correct. This issue is illustrated in figure 2.7: while the constraint-based formulation will ensure a contact-free configuration at the end of each time-step in all cases (as Signorini’s law is always fulfilled), the local deformation near contacts obtained without the correct compliance can be invalid.

Figure 2.7: Force repartition for homogeneous and inhomogeneous objects on which collisions are solved with and without mechanical coupling. Stiffer part are shown in red.

Yet, in section 1.5, we presented a preconditioning method that relies on an asynchronous factorization of the matrix $A$. From the conditioning number point of view, we have demonstrated that the solution provided by this asynchronous method is very close to the inverse of $A$, as only few iterations are necessary for the preconditioned conjugate gradient to converge.

Thus, if $P$ (the preconditionner) remains a good approximation of the system, it can be re-used to build the compliance matrix. That way, we guarantee a contact-free configuration at the end of the time step with almost the exact mechanical coupling of elements taken into account. To build the LCP matrix, we combine eq. (2.9) and (1.25) and we rely on a LDL preconditionner. Thus, we compute for each object:

$$h^2HA^{-1}H^T \approx h^2HRP^{-1}R^TJ^T = h^2J(LDL^T)^{-1}J^T \quad \text{with} \quad J = HR \quad (2.18)$$

This computation can be implemented in three steps: The step 1 consists in applying the local rotations since the last update of $H$. This operation is inexpensive because it is done by computing the product of a block-diagonal matrix $R$ with a sparse matrix into matrix $J$ with the same sparsity structure. During step 2, we compute the product of $J^T$ with the inverse of the factorization. This can be achieve by computing rows independently, each requiring two triangular solves with one row of $J$. In step 3, the
resulting matrix is multiplied by matrix $J$ to obtain the final contribution to the LCP matrix.

The number of successive solves in step 2. is proportional to the number of constraints (3 per friction contact). They can therefore become expensive in complex scenarios. The good news is that the solve for each constraint is independent. Therefore, we propose to implement them on GPU in parallel, reducing the level of parallelism that must be achieved within each solve. This maps nicely to the two-level SIMD architecture of today’s GPU where synchronizations within a group of cores is fast, whereas global synchronizations over multiple groups is much more prohibitive. Thus, we perform the resolution on the factorized matrix $LDL^T$ for each column of $J^T$ on separate groups of core. This is a first level of parallelism that we can extract from the algorithm (see figure 2.8). A second level of parallelism is obtained at the level of the each resolution with the factorized matrix.

Figure 2.8: First level of parallelism achieved for solving a Sparse Triangular System with multiple right hand side vector on GPU. A second level of parallelism (illustrated on the 4th Multiprocessor) is extracted by using a multi-threaded strategy at the level of each system.

**Validation and Results:** Even if the asynchronous preconditioner is close to matrix $A^{-1}$, it is still an approximation. We measured the error introduced by the approach compared to an exact factorization at each time step. We produced a simulation involving a non homogeneous deformable disc (the center is stiffer while the periphery is soft): the disc is contacting a sphere that push it through a small hole. There is a strong coupling between the contacts on the sphere and the contacts on the border of the hole.

Figure 2.9: Difference in position over time introduced by several approximation as compared to an exact factorization.
2.3. COMPLIANCE COMPUTATION

The difference of vertices position was measured when using a diagonal compliance matrix, which cancels the coupling between the contacts. We also tested the example with no update of the compliance being equivalent to the compliance warping, and multi-thread corresponding to our contribution where the preconditioner is updated as soon as possible. To demonstrate the convergence of the approach, we also tested the results with an update of the factorization every 20 steps and with an update every 10 steps. The results are shown in Fig 2.12 and demonstrate that inaccurate approximations do not lead to the correct behavior, whereas our approach is very precise if there is frequent update of the preconditioner (less than every 10 steps), which is coherent with the performance measured on models that we want to simulate.

![Figure 2.10: Simulated lens (blue) with reference (green) after 0.8 second of simulation.](image)

**GPU implementation:** We measured the computation time\(^3\) (see Fig. 2.11) required to build the compliance matrix with our GPU-based algorithm compared to a sequential CPU implementation. It can be observed that the computation time for solving less than 100 instantaneous contacts remains almost constant with the GPU algorithm, which allows to quickly take advantage of the Graphic processor.

Indeed, until 130 contacts the computation units of the GPU are not fully exploited, and each resolution is computed in parallel. Beyond this number, some of the computation units will compute several successive solves. However, the GPU is able to overlay waiting time due to synchronizations and read/write in memory, with computations for another solve. Thus, solving 260 contacts is only 1.5 times slower than 130.

\(^{3}\)All the following results were obtained on a quad-core Intel® Core™ i7 3.07 GHz CPU with a Nvidia® GeForce® GTX 580 GPU.

![Figure 2.11: Comparison between GPU and CPU performance to compute the compliance matrix according to the number of contacts.](image)
Overall, this solution is very performant from both precision and computation time point of views. It allows to compute contact constraints on heterogeneous deformable models in real time. Many applications can be derived from this new technique. For instance, in section 4.2.2, we present a simulation of the extraction of the eye lens during a particular procedure of the cataract surgery. But, in the following we continue to illustrate the method using an example of liver surgery.

**Simulation of liver resection**  In some cases, hepatectomy may be considered for the treatment of tumors localized inside the liver. Simulators of this procedure has already been developed in the past (Bourquain et al. (2002) Lamadé et al. (2002)). The originality of our approach is that our simulation is based on patient specific data. Indeed, the meshes of the organs are obtained from a semi-automatic segmentation of a CT (see Soler et al. (2001) for details). We simulate 5 deformable bodies in interactions (liver, stomach, colon, intestines and diaphragm). Each organ is composed of several hundred of nodes and thousand of elements with complex shapes composed of several thousand of triangles. The interactions are mostly modeled using frictional contacts (except the interaction between the diaphragm and the liver). This is currently a limitation because the real interactions between these soft-tissues are even more complex and involve connective tissues. Frictional contacts with surgical instruments can also be simulated as well as cutting (by removing tetrahedral elements, re-meshing is currently not supported).

The collision detection is performed by the method introduced in Allard et al. (2010b) and presented in the following section. This simulation runs at a frequency of 25 FPS, including during cutting. The time distribution within a time step is as follow: 30% for free motion (for the computation of the deformation models) and only 20% for building the compliance matrix. It demonstrates a certain equilibrium in the computation time between these two bottlenecks of our simulation. The remaining time is used for collision detection, Gauss-Seidel algorithm, corrective motion and propagation of the deformations on the visual meshes.

The important fact is that the liver model is no more isolated from its anatomical deformable neighborhood. Without a credible approach to the boundary conditions, it is useless to try improving the accuracy of the FEM model of the liver. This method is an important step towards an answer of this issue. But more work is still needed, especially in the constraint models of interaction between the anatomical structures (today, only contacts are simulated) and in the modeling of connective tissues (ignored for the moment). However, thanks to the asynchronous preconditioner approach, we can obtain the compliance of each soft-tissue at the level of its interface, which is necessary to solve frictional contact but can also be used for more complex constraint modeling (see 2.5).
2.4. CONSTRAINT-BASED CONTACT RESPONSE

In this section, two contributions are presented. They both concern the modeling of the contact response based on constraints. The first one is dedicated to a very generic way of setting the contact constraints whatever mechanical model or collision approach is used. It is closely related to one of the most difficult point on which I contributed in SOFA: make compatible the use of contact constraints with SOFA mappings that are a central component of the framework. The concept is presented in Duriez et al. (2008) and generalized more recently in a book chapter related to SOFA: Faure et al. (2012).

The second contribution is a realistic response for collision algorithms based on interpenetration volume that can be implemented very efficiently on GPU. The approach is based on a re-formulation of Signorini’s law using this interpenetration volumes and on an arbitrary discretization of these volume contact constraints. It has been published in Allard et al. (2010a).

Conclusion Computing the compliance in constraint space $W$ is an issue in a real-time context. We present a fast and exact solution for 1D structures. For 2D and 3D structures, we propose different ways of approximating the compliance, based on simplified model or on precomputation. Finally, the most promising approach is based on asynchronous preconditionner computed on GPU which is more accurate and still fast. In the following, we present how we set the contact constraints between all these possible mechanical objects (rigid, deformable, 1D, 2D, or 3D...) and also how we can reduce the number of contact constraints between volume deformable models.

2.4 Constraint-based contact response

Figure 2.12: Simulation of an hepatectomy: the approach allows to model the liver in contact with its anatomical environment and with the surgical tools. The user can virtually cut the liver and grasp the resected part using the laparoscopic devices.
2.4.1 Contact mapping

The work presented in this section is a generalization of the concept of Contact Skin-ning presented in Duriez et al. (2008). This has been the support for the implementation of the constraints inside the SOFA framework.

Indeed, in SOFA we introduced the concept of Mapping that allows to define several representations for the same object. Given the position of the degrees of freedom $\mathbf{q}$, one can define a new geometry (for collision detection or for visualization, for instance), that is kinematically mapped to $\mathbf{q}$: the position of any point of this geometry is given by $\mathbf{p} = \mathcal{J}(\mathbf{q})$, where $\mathcal{J}$ is the mapping function (which is sometimes non-linear).

![Figure 2.13: A simulated Liver. Left: The liver displayed in its environment. Right: Three representations are used for the liver: one master model for the internal deformable mechanics, one for the collisions, and one for the visualization. Mappings (black arrows) are used to propagate positions ($\mathbf{q}$) and velocities ($\mathbf{v}$) from master to slaves, while forces ($\mathbf{f}$) are propagated in the opposite direction.](image)

Consider the deformable model of a liver shown in the left of Figure 2.13. It is surrounded by different anatomical structures (including the diaphragm, the ribs, the stomach, the intestines, etc.) and is also in contact with a grasper (modeled as an articulated rigid chain). In SOFA, the liver is decomposed in three different models: The first is used to represent its internal mechanical behavior, which may be computed using Finite Element Method (FEM) or other models. The geometry of this model is optimized for the internal force computations, typically using a reduced number of well-shaped tetrahedra for speed and stability. The second model is used for collision detection: the best trade-off between precision and speed may require another geometrical model. Finally, the third model is used for the visualization and probably needs a smoother and more detailed geometry.
2.4. CONSTRAINT-BASED CONTACT RESPONSE

Figure 2.14: Mappings between the DOFs and the contact points. Left (top to bottom): the internal model of the liver is based on Finite Element model. A triangular mesh is mapped for collision detection with the surface. The two contact points found by the collision detection (with the grasper) are mapped on the collision model. Right (bottom to top): the contact points are also mapped on the collision model of the grasper. This collision model is a simplification of the grasper shape and is mapped on the rigid body frames. The motion of these frames is mapped on the state of the joints which are the independent DOFs of the grasper.

We can recursively apply several mappings, when necessary. For instance, the grasper in contact with the liver can be computed with an articulated body model. The degrees of freedom would be the generalized coordinates. A first mapping will allow to obtain the position and orientation of rigid bodies of the grasper while a second one can be used to map collision models on these rigid bodies (triangle vertices in figure 2.14).

Concerning the collision, the mappings allow to detect the contacts with a unified representation of collision models even if the underlying mechanical model is completely different. The contact points can also be mapped on the collision model in order to place the contact points between the vertices of the collision mesh, for instance, as illustrated in figure 2.14.
If a function $\mathbf{J}(\mathbf{q})$ maps the positions from a *parent* model to a *child* model, the velocities of the child $\mathbf{v}_p$ can be mapped in a similar way using $\mathbf{v}_p = \mathbf{Jv}$ with $\mathbf{J} = \frac{\partial \mathbf{X}}{\partial \mathbf{q}}$.

If we apply the principle of virtual work, we can map the force from the child model $\mathbf{f}_c$ to the parent model, using $\mathbf{f} = \mathbf{J}^T \mathbf{f}_c$.

Figure 2.15: Contact process using constraints: A unilateral constraint is placed at the level of the contact points. The constraint direction is mapped to the degrees of freedom of the objects to obtain matrix $\mathbf{H}^T$. The *ConstraintCorrections* components compute the compliance to obtain equation 2.9. The Constraint solver found a new value of $\lambda$ which is sent to the *ConstraintCorrections* to compute an adequate corrective motion. The Master Solver is placed at the root of the simulation graph to impose the steps of the simulation process.

One of the contribution of this work is to use this principle to set contact constraints between all these types of deformable models (1D, 2D, 3D) and also with any other mechanical object or structure (rigid, articulated bodies...). Indeed, in the contact process described above, we have also defined an equation of mapping (see equation 2.5) between the degrees of freedom and the contact space. For a contact point $\alpha$
2.4. CONSTRAINT-BASED CONTACT RESPONSE

along direction \( n \), between two collision models that are mapped using the SOFA principle. Thus, as illustrated in figure 2.15, we need to apply the successive mapping from the degrees of freedom \((i = 0)\) to the contact points \((i = n)\) in order to evaluate violation \( \delta^{\text{free}} \). Indeed, the value of \( A_{\alpha}(q^{\text{free}}) \) used in the computation of \( \delta^{\text{free}} \) depends on the mapping functions: \( A_{\alpha}(q^{\text{free}}) = n_{\alpha}^T \cdot n \mathcal{J}(...i \mathcal{J}(...0 \mathcal{J}(q^{\text{free}}))) \). Whereas, we go through the mappings in the reverse order to build the matrix of contact directions \((H^T)_{\alpha} = (0 \mathcal{J}^T) \ldots (i \mathcal{J}^T) \ldots (n \mathcal{J}^T) n_{\alpha} \). As the matrices \( i \mathcal{J} \) are often highly sparse, we use a dedicated structure and functions to optimize the transmission through the mappings.

Additional components were implemented in SOFA to allow for a very generic approach when dealing with contact constraints. First, the AnimationLoop allows to impose the different steps of the constraint-based process (Free-Motion, Collision detection, Evaluation of the violation, Setting the constraints, Solving, Corrective-Motion). The way constraints are solved is separated in a Constraint Solver to allow different strategies for solving constraints. Finally a component called ConstraintCorrection allows to implement separately the various models of compliance that can be used to solve the constraints and that are presented in the previous section.

Finally, we obtain a very generic framework, that is able to compute collision response between any type of mechanical models (rigid, deformable, articulated...) as soon as they can apply a motion to collision models through the use of one or several mappings.

2.4.2 Volume contact Constraints at Arbitrary Resolution

We introduce a new method for simulating frictional contact between volume objects using interpenetration volume constraints. When applied to complex geometries, our formulation results in dramatically simpler systems of equations than those of traditional mesh contact models. Contact between highly detailed meshes can be simplified to a set of unilateral constraint equations, and accurately processed at arbitrary geometry-independent resolution with simultaneous sticking and sliding across contact patches.

The basic idea is to detect and minimize the intersection volume between polyhedral meshes. The intersection volume is computed on the GPU based on bounding pixels in the rasterization of the object surfaces into Layered Depth Images (LDI) Heidelberger et al. (2003, 2004). One LDI in an arbitrary direction is sufficient to detect collisions between volumetric objects. But, computing the gradient of the intersection volume requires three LDIs in mutually orthogonal directions Faure et al. (2008). At left in Figure 2.16 is a depiction of LDI volume models using red, green, and blue pixels to denote \( x, y, \) and \( z \) viewing directions, respectively. The rasterization can be done in any direction, but for simplicity we assume an orthogonal projection along one of the primary axes. The corresponding intersection volumes are shown in the second image of the same figure.
CHAPTER 2. CONSTRAINT-BASED MODEL OF BIOMECHA. INTERACTIONS

Thus, the volume computed using a $z$ projection is

$$ V = a \sum_{(i,j) \in S^+_z} z^+_{ij} - a \sum_{(i,j) \in S^-_z} z^-_{ij}, $$

(2.19)

where $a$ is the area of a pixel, $z^+_{ij}$ and $z^-_{ij}$ are the upper and lower pixel depths, and the sets $S^+_z$ and $S^-_z$ respectively contain the pixel locations $(i,j)$ of the upper and lower contact surfaces.

For each axis, the sums over pixels for the volume and gradient computation are done simultaneously. This is done three times, once per LDI. The volume derivative is accumulated with respect to the coordinate of each vertex is accumulated into the gradient vector:

$$ \frac{\partial V}{\partial \vec{p}} = \left( \frac{\partial V}{\partial p_1} \ldots \frac{\partial V}{\partial p_n} \right) \quad \text{and} \quad \frac{\partial V}{\partial \vec{q}} = \left( \frac{\partial V}{\partial q_1} \frac{\partial V}{\partial q_2} \frac{\partial V}{\partial q_3} \right), $$

(2.20)

where $x, y, z$ are the three successive LDI viewing directions. When the independent DOFs $\vec{q}$ are not directly the vertex positions $\vec{p}$, as in the case of rigid bodies or detailed surfaces embedded in coarse deformable models, we use the mapping strategy of SOFA:

$$ \frac{\partial V}{\partial \vec{q}} = \frac{\partial V}{\partial \vec{p}} \frac{\partial \vec{p}}{\partial \vec{q}}. $$

The collision detection provides a measure of the interpenetration volume and allows to compute a gradient of this measure. Yet, the Signorini’s law is based on a signed distance between colliding pairs of points. Consequently, we reformulate the Signorini condition in order to obtain an equivalent complementarity problem between contact...
reaction pressures and interpenetration volume. We also need to adapt the formulation to be compatible with Coulomb’s friction model. This volume-based formulation can directly use the LDI collision output to define a very efficient constraint-based collision response.

**Equivalence** Recall that each contact $\alpha$ must verify the condition $0 \leq \delta_\alpha(q) \perp \lambda_\alpha \geq 0$. This Signorini’s condition can be rewritten at a velocity level for any active contact $0 \leq \dot{\delta}_\alpha(q, \dot{q}) \perp \lambda_\alpha \geq 0$ if $\delta_\alpha = 0$. In practice, as the LDI method is a discrete collision detection, only non-zero intersection volume can be detected (if the intersection is zero, there is simply no contact). Thus, as many other approaches that relies on discrete collision detection, we solve the system at the velocity level, when a contact is detected ($\delta_\alpha < 0$). In practice, a post-stabilization step can reduce the amount of intersection.

We observe that $\dot{\delta}_\alpha(q, \dot{q}) = \mathbf{H}_\alpha(q) \dot{q}$ where $\mathbf{H}_\alpha$ is the separation gradient of contact $\alpha$. Thus Signorini’s law, at the velocity level, is respected when the 3 following conditions are simultaneously fulfilled:

$$
\mathbf{H}_\alpha \dot{q} \cdot \lambda_\alpha = 0 \quad \lambda_\alpha \geq 0 \quad \mathbf{H}_\alpha \dot{q} \geq 0
$$

We can associate a small area $\mathcal{A}_\alpha > 0$ with each contact $\alpha$, and observe that the Signorini’s law is also respected with the 3 following conditions:

$$
\mathcal{A}_\alpha \mathbf{H}_\alpha \dot{q} \cdot \lambda_\alpha \frac{1}{\mathcal{A}_\alpha} = 0 \quad \lambda_\alpha \frac{1}{\mathcal{A}_\alpha} \geq 0 \quad \mathcal{A}_\alpha \mathbf{H}_\alpha \dot{q} \geq 0
$$

where $\rho_\alpha \equiv \lambda_\alpha \frac{1}{\mathcal{A}_\alpha}$ is a pressure and $\frac{\partial \mathcal{V}_\alpha}{\partial q} \equiv \mathcal{A}_\alpha \mathbf{H}_\alpha$ is the volume gradient (thus, $\dot{\mathbf{V}}_\alpha = \frac{\partial \mathcal{V}_\alpha}{\partial q} \dot{q}$). Because the inequalities of the separation distance constraint still apply, we have equivalent linear complementarity problem for the pressure and volume-based constraint formulation,

$$
\begin{cases}
0 \leq \rho_\alpha \perp \dot{\mathbf{V}}_\alpha \geq 0 \\
\dot{\mathbf{V}}_\alpha = \left[ \frac{\partial \mathcal{V}_\alpha}{\partial q} \right] \mathbf{h} \mathbf{A}^{-1} \left[ \frac{\partial \mathcal{V}_\alpha}{\partial q} \right]^T \rho_\alpha + \dot{\mathbf{V}}^\text{free}_\alpha
\end{cases}
$$

where $\dot{\mathbf{V}}^\text{free}_\alpha = \frac{\partial \mathcal{V}_\alpha}{\partial q} \dot{q}^\text{free}$ can be measured using the free motion.

**Friction** The repulsion pressure direction is given by the gradient vector $\frac{\partial \mathcal{V}}{\partial p}$. It provides the approximate direction of the contact normal vector $\mathbf{n}_\alpha$, after the division of the gradient vector by an estimation of an area of contact $\mathcal{A}_\alpha$. For friction, we need a similar measure of the tangential directions. We then compute two orthogonal unit vectors $\mathbf{t}_1\alpha$ and $\mathbf{t}_2\alpha$ spanning the tangent plane using Gram-Schmidt orthogonalization. To spread these tangential directions on the vertices involved in the contact $\alpha$ and build the tangential Jacobians, we need to associate a weight for each vertex. This weight is chosen according to the associated contact surface area, which we approximate by the
volume gradient with respect to the vertex position projected onto the volume-contact normal.

We validated the sampling approach on the example of a snake pulled laterally illustrated in Figure 2.18. We first tested the same example with a unique contact constraint (one cell in the regular grid) and with more and more constraints.

**Arbitrary Sampling** We divide the intersection volume using a regular grid aligned with the LDI directions, as shown in the second image of Figure 2.16. The volume of contact $V_\alpha$ and the gradient are estimated separately for each cell of the grid. We obtain three contact constraints (1 normal and 2 tangential directions) per cell where two objects are colliding. Consequently, it is easy to tune the spatial resolution of the contact: if precision is needed, we use a fine grid whereas a coarse grid could provide good visual results with very few constraints but less accuracy.

With a single contact constraint, the non-penetration constraint was still fulfilled but the snake spins around the contact center located approximately in the middle of its body. When the number of contact constraints increases, the snake model gently evolves to a nice stable bended shape. This configuration corresponds to the converged position (less than 0.35% error) obtained using classical contact constraints at each vertex of the snake mesh.

**Application to hepatic simulation based on medical data** The grasping simulation of the liver, presented in Figure 2.16 includes contact response with surrounding deformable structures (stomach, intestines). We emphasize that realistic grasping is obtained by using frictional contact constraints and the boundary conditions of the liver model now include the interactions with other deformable structures.

It is based on real data, after a reconstruction process realized by IRCAD. The organs...
2.5. **SURGICAL INTERACTION MODELED BY COMPLEMENTARITY CONSTRAINTS**

slightly intersect each other at initialization time, but our volume-based method is robust to initial intersections and has no problem separating the organs on the first time step, thanks to a post-stabilization process that corrects the volume of interpenetration. The scene includes 550,000 triangles, 9,500 degrees of freedom. Each step is simulated in 35ms including rendering with a mean of only 20 frictional contact constraints (60 constraints). To compute the compliance matrix, we use the method presented on section 2.3.3.

**Conclusion**  Simulating contact and friction between deformable bodies in real-time is a great challenge that involves the modeling of the deformations, the collision detection and the constraint process that is used to define and solve the contact equations. With the contact mapping framework inside SOFA, we have developed a generic solution that allows to define the contact equations independently from the collision model and the type of mechanical models involved in the collision. For performance purpose, we have developed a contact response method that is able to use the informations of a fast GPU-based collision algorithm and more importantly, that allows to dramatically reduce the number of contact constraints with a good accuracy.

### 2.5 Surgical interaction modeled by complementarity constraints

Several medical procedures rely on the insertion of one or more flexible needle(s) such as biopsy, brackytherapy, cryotherapy or of another slender medical devices such as an electrode in deep-brain stimulation. One other very basic surgical task is the suture, that is particularly challenging from the simulation point of view. The key contribution presented on this section is the use of the complementarity constraints (used for contact and friction) mixed with bilateral constraints to model these interactions.

This work has been done during the Ph.D. thesis of Christophe Guebert and the post-doc of Maud Marchal. It was published in the following conference papers: Guébert et al. (2008) (generalization of the constraint solver for mixed non-linear complementarity problem), Duriez et al. (2009) (flexible needle insertion modeling), and Guebert et al. (2009) (Suture simulation).

Developing this type of models is still an active research. It could be very interesting to improve the realism of training simulators but it has many potential applications for planning and interactive guidance (see for instance sections 4.1.3 and 4.1.6). It is also closely related to constraint-based haptic feedback (see section ??).

As presented above, the Signorini and the Coulomb laws describing the contact and friction between mechanical models can be solved using a set of complementarity constraints. These constraints allow to describe several *states* in the constraint law:
the contact can be *active* when \( \lambda > 0 \) or can be *non-active* when \( \delta > 0 \). Similarly in the friction law, the state can be *stick* if the force lies inside the friction cone or *slip* if the force lies on the surface of the cone.

We propose to use these complementarity constraints to develop a new model for the insertion of a needle inside the soft tissues or for suture simulation. The formulation is generic and independent of both needle and tissue models. Our approach relies on using complementarity theory to describe the non-smooth mechanical phenomena that occurs during the insertion.

More precisely, we develop a set of specifically defined constraints to simulate different physical phenomena that have also several *states* such as puncturing the surface of soft-tissues, cutting the tissues along the needle path, static and dynamic friction created by the tissues on the needle or on the suture. Each constraint is dynamically created on the soft tissue model without requiring any remeshing thanks to the use of the mapping technique already used for contact.

In the following, we first present the formulation of our modeling approach, then we quickly present the solving process. The section ends with the results obtained on needle insertion and on suture simulation.

### 2.5.1 Constraint-based interaction model

Our main contribution in this study is a new set of interaction models to describe the mechanical phenomena occurring during needle insertion or suturing tasks.

The following notations are used: \( \delta \) is a measure of a distance between a current state and a target state that will be redefined for each constraint. \( \lambda \) represents the force used to solve the constraint (applied by the tissue on the needle or the thread). In general, \( \delta \) and \( \lambda \) are both unknown.

**Puncturing constraint:** Puncturing illustrates the interest of using complementarity theory to model the constraint. Three successive steps of the interaction can be defined with a set of inequalities, as illustrated in Fig. 2.19. Here, \( \delta = PQ \cdot \bar{n} \) where \( Q \), the tip of the needle and \( P \) is the projection of \( Q \) on the tissue surface. \( \bar{n} \) is the surface normal vector at point \( P \).

During **step 1**, \( Q \) is only approaching the tissue surface. The gap \( \delta \) is positive (\( \delta \geq 0 \)) and the interaction force must be null (\( \lambda = 0 \)). During **step 2**, \( Q \) is touching without piercing the tissue surface. The gap between \( P \) and \( Q \) is null (\( \delta = 0 \)) and the interaction force is necessarily positive in the direction of the surface normal (\( \lambda \geq 0 \)). The value of this force is strictly inferior to a puncturing force threshold \( \lambda \leq f_p \). During **step 3**, the needle tip enters in the tissue media, the gap along the constraint direction is negative (\( \delta \leq 0 \)) and the constraint force is equal to the threshold (\( \lambda = f_p \)).
The puncture constraint can be applied several times during the simulation if the needle passes through different tissue layers: different values for the threshold $f_p$ can be defined in order to simulate different tissue behaviors. A constraint on each lateral direction is associated to the puncture constraint. When contact is activated (step 2), lateral motion is constrained by friction constraint. If the tip of the needle hits a bone for example, $f_p$ is very high and never reached by $\lambda$. The needle slips along the surface of the bone. On the contrary, when puncture states are activated (step 3), lateral motion (along the surface tangential directions) is made impossible by using a bilateral constraint.

Other complementarity constraints: Similarly to the puncture model, we can use complementarity constraints to reproduce other mechanical interactions that takes place during needle or suture insertion.

- The cutting force: it corresponds to the force needed to penetrate a tissue structure. If this force is not reached, the needle stops its progression in the soft tissue. We model it using a constraint which is similar to the puncture, by replacing $f_p$ by a cutting force that can be measured experimentally (it depends on the tissue density, the size of the needle and its sharpness). This force disappears if the needle is re-inserted at the same location.

- Dry and dynamic friction: It corresponds to the resistance to the motion when the needle is retracted or re-inserted (the cutting force then vanishes). As for friction contact model, we rely on the Coulomb’s law but the model needs to be integrated along the surface of the needle that is inserted inside the tissue. Indeed, experimentally, it is observed that the friction force increases with the length of the needle that is inserted.

- Friction contact: the different layers of tissues that are penetrated by the needle
are sometimes in contact. In the case of the suture, many contacts appear on
the suture thread and need an adequate response. The thread is colliding the
surface of the organs, the other surgical tools (such as the laparoscopic grasper
if it applies), and is sometimes self-colliding (namely when knot tying).

For all these constraints it is possible to describe a graph between $\delta$ and $\lambda$, as it is done
for the puncture in the figure 2.19. This graph will be used in the solving process.

**Dynamic path inside the tissue:** A direction is associated to the needle tip in order
to constrain its lateral motion. To obtain needle steering due to bevel tip needle,
a specific orientation of the cutting direction in the tip frame is defined. When
the needle is inserted more further, more constraints are set to drive the sliding of the
needle inside the soft-tissue. For each constraint location, one point of the tissue
is constrained to slide along the curve of the needle. For that, we can use usual
bilateral and holonomic constraints that are regularly sampled along the needle curve
(in practice, the sampling is slightly less dense than the discretization of the needle
deformation model). The same model could be applied to the suture. Indeed, using
the beam model, we have a unique deformable model for the needle and the suture
(just by adapting the mechanical parameters of the beam elements accordingly).

### 2.5.2 Unified processing of constraints

The processing of this mixed non linear complementarity problem uses a generalization
of the Gauss-Seidel algorithm that was used for contact. As it is the case for friction
contact problem, the main issue is first to get the compliance matrix in the constraint
space. Here it is the sum of the needle (or the suture) compliance $W_n$ and the soft
tissue compliance $W_t$:

$$
\delta = \begin{bmatrix}
W_n & W_t
\end{bmatrix}
\begin{bmatrix}
H_n \left( \frac{M_n}{h^2} + \frac{\partial F_n}{\partial v_n} + \frac{\partial F_n}{\partial q_n} \right)^{-1}
H_n^T + H_t \left( \frac{M_t}{h^2} + \frac{\partial F_t}{\partial v_t} + \frac{\partial F_t}{\partial q_t} \right)^{-1}
H_t^T
\end{bmatrix}
\lambda + \delta_{free}
$$

where $h$ is the time step and $\left( \frac{M}{h^2} + \frac{\partial F}{\partial v} + \frac{\partial F}{\partial q} \right)$ the dynamic tangent matrix. Based on
this formulation, $W_n$ is obtained by the method detailed in section 2.3.1 and $W_t$ can be
computed using compliance warping (see section 2.3.2) or asynchronous preconditioner
approach (see section 2.3.3).

To find the constraint force value, we continue to rely on a Gauss-Seidel like algorithm.
Considering a constraint $\alpha$, among $m$ instantaneous constraints, one can reuse equation
2.23 (that was introduced for friction contact response). Here, instead of a contact,
2.5. SURGICAL INTERACTION MODELED BY COMPLEMENTARITY CONSTRAINTS

we have a generic constraint \( \alpha \):

\[
\delta_{\alpha} - [W_{\alpha,\alpha}]\lambda_{\alpha} = \sum_{\beta=1}^{\alpha-1} [W_{\alpha,\beta}]\lambda_{\beta} + \sum_{\beta=\alpha+1}^{m} [W_{\alpha,\beta}]\lambda_{\beta} + \delta^{\text{free}}_{\alpha} = \delta^{-}_{\alpha}
\] (2.23)

where \([W_{\alpha,\beta}]\) is the value of the matrix \( W = W_n + W_t \) at line \( \alpha \) and column \( \beta \). It models the coupling between constraint points \( \alpha \) and \( \beta \). At each iteration of the Gauss-Seidel solver, each constraint \( \alpha \) is visited and a new estimate of \( \lambda_{\alpha} \) is performed while "freezing" the contributions of \( \lambda_{\beta} (\alpha \neq \beta) \). The new estimate of \( \lambda_{\alpha} \) is found at the intersection of the characteristic graph of each constraint with the line of equation (2.23). Note that for all mechanical models, the local compliance of a point is always positive: \([W_{\alpha,\alpha}] > 0\), so we obtain a unique solution for \( \lambda_{\alpha} \) (see Fig 2.20). The convergence of the algorithm is given by a threshold error on the values of vector \( \delta \).

Figure 2.20: Graph intersection example: it allows to find \( \lambda_{\alpha} \) given \( W_{\alpha,\alpha} \) and \( \delta^{-}_{\alpha} \)

2.5.3 Simulation results of needle and suture interactions with soft-tissues

Figure 2.21: (Left) simulated force at the base of a needle during an insertion sequence. (Right) measurements in Dehghan et al. (2008), courtesy of E. Dehghan, ECE, Vancouver, Canada.

Needle Insertion: We conducted a first validation of our model by reproducing several experiments presented in previous papers. In Dehghan et Salcudean (2007), the
authors insert and retract multiple times a needle into a phantom, measuring the force applied at the base of the needle. In figure 2.21 we compare their measures with the simulation of that experiment using our model. We can identify (1) contact and puncture, (2) friction increasing proportionally to the penetration distance and (3) relaxation. After being partially retracted (4) the needle is inserted again along the same path (5); therefore no cutting force is applied. During the last and complete retraction (6), the friction force decreases as constraints are removed.

Beveled-needles can be supported in our model. We reproduce in simulation the experiments conducted by Webster et al. (2006). We first look for the parameters to recreate the single bend needle path in which a needle with a bevel tip of $45^\circ$ is inserted without spin into a rigid phantom. Keeping the same parameters, we compare our simulation of the double bend needle path with their experimental measures. The needle is first inserted one third of the distance into the phantom, then spun $180^\circ$, and finally inserted the remaining two thirds. Figure 2.22 shows that our model fits experimental measures.

![Figure 2.22: (Left) Insertion simulation of a flexible and beveled needle into a rigid tissue. (Right) Comparison to the experimental model of Webster et al. (2006).](image)

For obstacle avoidance, surgeons may use thin and flexible needles, even with symmetric tips. However, flexible needle insertion and navigation deep into the tissue is difficult, and calls for robotic insertion and rigorous path planning algorithms Glozman et Shoham (2004). One possible path for a needle insertion to a target and avoiding an obstacle is shown in figure 2.23, from a simulation using an haptic device for the manipulation of the needle.

![Figure 2.23: Simulating an needle insertion to a target, with obstacle avoidance.](image)

In-vivo tissues are inherently non-homogeneous and the motion of thin medical tools can be deflected by harder regions inside overall soft tissues. We simulate this behaviour using two tissues of different stiffness (see Fig. 2.24). While being inserted in the soft tissue, the needle collides the surface of the stiffer tissue. As the force at
the tip is lower than the puncture threshold, the needle slide along the surface of this object. The needle and the soft tissues deform, until the puncture force is reached. These deformations vary depending on the stiffness of the needle.

Figure 2.24: Influence of the stiffness of the needle on its deformation during an insertion.

In the brachytherapy procedure, surgeons sometimes insert multiple needles in order to rigidify a region of the soft tissue. Then, the insertion of additional needles is more precise, with less deformations of both the needle and the tissue. Using the previous setup, we first insert five needles into the soft region before inserting a needle all the way to the target. Figure 2.25 shows that the error of insertion can be dramatically reduced.

Figure 2.25: The insertion of multiple needles rigidify the tissue around them, lowering the amount of deformation and making latter insertions more precise.

**Suture simulation** is still a work in progress. Even if the modeling is similar, other issues arise when dealing with suture:

- More constraints are instantaneously simulated: for instance there are a lot of situations with self-collision, including simulating knot. Moreover, there is contact response between the wall of soft-tissues, contact and grasping with the laparoscopic tools...

- The constraints are tighter: the suture thread is clamping the soft tissues. Stability and robustness in any case are difficult to obtain and the resolution process could face additional difficulties to converge.

- The suture thread has much larger deformations that can still be handled by the beam model presented in section 1.3 but necessitates the introduction of an adaptive strategy in the discretization.

- The dynamic of the suture thread is very fast whereas the dynamic of the soft tissue is low. In a real-time context, the two models should not be simulated
at the same frequency. This issue is particularly challenging when dealing with haptic feedback. A partial answer to this problem is proposed in the section 3.3.

Figure 2.26: Illustration of some constraints in the suture model: (a) When we stop to pull on the thread, the suture point is opening again. (b) Thanks to dry friction model, the suture point is maintained closed. (c) Examples of contacts occurring during suture: contact between the soft tissue walls, between the soft tissue and the thread and self-collision of the suture thread

Anyway, we have obtained interesting preliminary results for the proof of concept. For instance, figures 2.26(a) and (b) illustrate the interest of friction (and particularly dry friction model) for the realism of the simulation. On the same figure (c), we show an example that combines self collision of the suture thread and strong interactions with the soft tissue.

As illustrated in figure 2.27, a basic suture task can be executed by driving directly the trajectory of the needle, like a manual suture, but the needle can also be grasped and driven by a laparoscopic tool.

Figure 2.27: Suture simulation: (left) several steps of the simulation of a suturing task. (right) snapshot of the suture model in a laparoscopic simulator (Courtesy of Digital Train-ers)

We measure various simulations including the basic suturing task simulation when three stitches have been made and the suture is being pulled to join together the
2.5. SURGICAL INTERACTION MODELED BY COMPLEMENTARITY CONSTRAINTS

<table>
<thead>
<tr>
<th></th>
<th>Contact</th>
<th>Friction</th>
<th>Suture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free Motion Suture</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Free Motion Tissue</td>
<td>15</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Build $W$ Suture</td>
<td>46</td>
<td>43</td>
<td>90</td>
</tr>
<tr>
<td>Build $W$ Tissue</td>
<td>1</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Gauss-Seidel</td>
<td>4</td>
<td>8</td>
<td>20</td>
</tr>
<tr>
<td>Correction</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 2.1: Timings (ms) for several examples, in each step of our method. The simulations measured are: Contact (figure 2.26), Friction (figure 2.26.b) and Suture (figure 2.27).

two parts of the tissue. All timings are on a bi-Xeon 2.66GHz with 2GB RAM and show that the simulation can still be computed at interactive rates. It must be noted that when these measures were realized the computation of $W$ for the suture was not optimized using the method presented in section 2.3.1.

Conclusion  This work on needle insertion and suture simulation illustrates the benefit that could bring the use of complementarity constraints to accurately model the mechanical interactions. These constraints can be added dynamically and solved using a generic implementation of a Gauss-Seidel algorithm based on graph intersection.

In future work, we plan to apply the complementarity constraints modeling on other type of interaction such as cutting inside a soft-tissue, or to describe complex interactions between soft-anatomical structures, like adhesive contacts.
### Haptic Rendering and Multithreading Approaches

#### Table of contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Related Work</td>
<td>90</td>
</tr>
<tr>
<td>3.2</td>
<td>Constraint-based haptic rendering for medical simulation</td>
<td>93</td>
</tr>
<tr>
<td>3.2.1</td>
<td>6DoF rendering of contact and drilling on rigid CT-scan model</td>
<td>94</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Multirate solving of contact constraints</td>
<td>98</td>
</tr>
<tr>
<td>3.2.3</td>
<td>Multirate Compliant Mechanisms</td>
<td>100</td>
</tr>
<tr>
<td>3.3</td>
<td>Asynchronous simulation of deformable models with variable stiffness</td>
<td>107</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Motivations: Why adapting the refresh rates to each object ?</td>
<td>107</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Asynchronous quasi-static deformable models in contact</td>
<td>108</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Further extensions of the approach</td>
<td>113</td>
</tr>
</tbody>
</table>
With interactive simulation, the course of the computations can be modified in real-time while integrating events from the real world. This is their main interest and it is essential for surgical training: the gestures of the user must be captured in order to be reproduced by virtual medical instruments on the simulated environment with a good fidelity. When one of these instruments comes into contact with some models of a soft-tissue, instantaneous deformations must be computed. This requires real-time computation, as already discussed in this manuscript. It is sometimes essential that this visual feedback of the contact be enhanced by haptic rendering\(^1\) so that the surgeon can really feel the contact.

Haptic rendering is often provided by reversible articulated robotic devices. The user holds the end effector of the device which is often designed to resemble the tools used during the procedure. Different control strategy are possible to obtain a stable haptic interaction Adams et Hannaford (1999) including impedance, admittance or hybrid controls. During the master thesis (before Ph.D.), we proposed in Duriez et al. (2001) an other strategy where the robotic device provides an indirect feedback on a real instruments (in that case, a needle for amniocentesis).

There are several issues for providing haptic rendering from a real-time physics-based simulation of a complex environment:

- A first challenge is stability: Indeed, haptic feedback could artificially add some energy inside the simulation and creates instabilities, if the control is not passive. Additionally, the simulation must be very robust to all possible gestures of a user and remain stable in any case.

- As the delays and the stiffness have a direct impact on stability, the haptic control needs to be computed at least 1kHz for contacts with a moderate stiffness (for very stiff contacts or to capture vibrations, even higher refresh rates are needed Florens et Urma (2006)). This requirement is far from the real-time refresh rate of 30Hz needed for comfortable visual feedback (without haptics).

- The feedback must be as transparent as possible: the haptic rendering must reproduce the forces of the real world and the control methods employed for coupling the haptic device to the simulation should prevent haptic artifact such as excessive damping.

- In the particular case of a medical procedure simulation, the haptic feedback should account for a large variety of mechanical interactions. These interactions are potentially very complex and of different nature, depending on the instrument and on the nature of the intervention. In the case of needle insertion, for instance, the interactions are in addition closely linked to deformations of the soft tissues.

\(^1\)Sometimes the other type of feedback, like tactile, auditive or temperature feedback can also be integrated (see for instance Guiatni et al. (2012)).
Contributions: To address these challenging issues, we propose haptic rendering methods that seek, above all, to reproduce accurately the mechanical interactions between surgical instruments and organs. In all cases, the force cue felt by the physician comes mainly from these interactions. Subsequently, their accurate physical modeling is of prime importance in order to provide the user with high fidelity haptic feedback.

As presented in the previous chapter, these interactions are coupled with the mechanical model of the objects (deformable or rigid) through the compliance matrix in the constraint space. This compliance will have a key role for the computation of the force feedback with our methods. Moreover, we will use the same modeling (based on (non)-linear complementarity problem or (N)LCP) and solving process (based on Gauss-Seidel) to deal with the mechanical interactions.

For deformable bodies, when it is not possible to update this compliance at haptic rates, we will rely on an asynchronous and multi-threaded approach. On the opposite, for stiff contacts like in dental simulation, the states and the directions of the constraints could change quickly, so it is important to have fast update of the (N)LCP. The problem is even more complex when having stiff and deformable interactions in the same haptic simulation. In that case, a deformable body may create very quick changes in the force feedback, like a cable or a wire that is suddenly stretched. We have thus proposed a new asynchronous approach dedicated to interacting deformable bodies (for the moment limited to quasi-static), so that stiff deformable bodies can be simulated at higher rates than the soft tissues.

Outline of the chapter: The first section is dedicated to a literature review of the methods related to haptic rendering of medical simulations. Then, in section 2, we present the haptic rendering methods based on the process of mechanical constraints. We show that the approach is compatible with constraints between rigid objects (with a 6DoFs dental simulation example) and also with deformable objects, by desynchronizing the computation of the compliance matrix. Finally, in section 3, we present a new approach of interactive simulation where a stiff deformable object can be computed at high refresh rates, while interacting with a soft object computed with low refresh rates.
3.1 Related Work

Haptic rendering in the context of medical simulation is again at the intersection of several fields including robotics, control, human-computer interaction, physics based modeling, collision detection and response... For an exhaustive study on haptic rendering, which is a domain of research by itself, we refer to a dedicated book by Lin et Otaduy (2008) in which I wrote a chapter dedicated to deformable models Duriez (2008).

This section will focus on computer haptic methods related to our contributions that are mainly focused on obtaining haptic feedback from a simulation-based environment or mechanical solids (deformable and/or rigid). It is largely inspired by the section dedicated to related work of our recent publication Peterlik et al. (2011b) that unifies our previous works on haptic rendering of interacting solids objects.

Haptic rendering enables a physical interaction with simulated objects of a virtual environment. Somehow, it provides a ground truth on the accuracy of the physics-based models used in an interactive simulation. However, the link between the quality of haptic feedback and the accuracy of the interactive simulation is not direct. Indeed, as the real-time constraints of haptic feedback are even more severe, it is difficult to maintain the same level of quality in the modeling.

Yet, simulation dynamics and haptic rendering share the same fundamentals: in both cases, mechanical forces must be computed. The manipulated object interacts with others and generates reaction forces that feed the physics-based engine, and more importantly the controller of the haptic display. In the last case, the computation of the reaction forces is constrained by (i) the need to meet high frequency refresh rates, (ii) the stability of the device control law, and (iii) the fidelity of the haptic rendering as perceived by the user. Several approaches have been proposed to deal with some or all these constraints.

In Colgate et al. (1995), the concept of virtual coupling is proposed. Originally designed to improve the haptic rendering of virtual walls, it can be regarded as an artificial mechanical coupling which connects the haptic device to the virtual object. Parameters can be tuned to guarantee the stability of the haptic interaction but it often introduces ghost damping forces. Since the position of the manipulated virtual object is updated from the haptic display handle position, penetration between virtual objects occurs. A god-object approach is introduced in Zilles et Salisbury (1995) to increase the perception of stiffness through visually enforcing the non-penetration. In this case the haptic interface is represented by a 3DoF virtual model (the god-object) which conforms to the virtual environment. The method is extended by so called virtual proxy in Ruspini et al. (1997). These methods can be gathered as virtual coupling network as noted in Adams et al. (1998) and Adams et al. (2002) to further improve the stability and performance of the rendering. Extensions of the god-object to 6DoF (constraining position and orientation) is made in Ortega et al. (2007). The latter is
based on a minimization of the kinetic distance between the actual position of the haptic device handle and the virtual proxy. However, god-object approaches are limited to contact constraints and do not account for other types of mechanical constraints.

The virtual mechanisms concept was proposed for an easy design of haptic joints in the context of rigid bodies manipulation and prototyping in Nahvi et al. (1998). It was initially designed for computing haptics in computer-assisted force reflecting bilateral teleoperation to improve both robot maneuverability and stability Joly et Andriot (1995). In Kosuge et al. (1995), a task-oriented virtual mechanism is designed so that its dynamic behavior matches that of the actual remote robot. Using virtual mechanisms, it is possible to impose various motion constraints to the teleoperator including non-linear ones, such as coupling between translations and rotations, for instance. The use of haptic virtual mechanisms for modeling of articulated bodies is demonstrated in Constantinescu et al. (2006) and Beenakkers et al. (2007) where the simulation of serial and closed chains is studied. To achieve more complex set of constraints, the concept of virtual mechanisms has been extended to unilateral contacts by Merlhiot (2009). The particularity of this method, which makes it close to our needs, is that it can process various types of mechanical interactions. However, virtual mechanisms are always implemented for rigid bodies, therefore the entire solving process is performed at high rates which is generally not possible with biomechanical models of soft-tissues.

The methods proposed to model the haptic interactions with deformable bodies are often tailored to given specific cases and can hardly be generalized to a large range of constraints. In the pioneering works Bro-Nielsen et Cotin (1996); Cotin et al. (1999), a displacement-driven interaction is used instead of contact modeling. The positions are applied as bilateral constraints (equality conditions) and solved by a method of Lagrange multipliers. The approach is based on a superposition principle, and was further extended Picinbono et Lombardo (1999), by using a force extrapolation method. Other methods, such as a finite element model handling geometric nonlinearities is proposed in Zhuang et Canny. (1999). The model employs mass-lumping and explicit time integration for real-time simulation of dynamic behavior. But due to explicit integration, the rendering is only possible on very soft is objects.

Similarly, the small area paradigm relies on linear modeling Popescu et Compton (2003), for which the equality boundary conditions affects only a small number of surface nodes. Through a simple update of inverse stiffness matrix, the response forces can be computed at each step of the haptic loop. In Mahvash et Hayward (2004), Sedef et al. (2006) and Peterlik et al. (2010), methods based on precomputations are proposed; stable haptic rendering is achieved, since the response forces are efficiently calculated from precomputed data, e.g. by interpolation, performed directly in the haptic loop. Although viscoelastic and non-linear models are employed, only point-based interaction is considered due to limitations given by the precomputations.

Basdogan et Srinivasan (2002) propose a scheme for simulating force reflecting deformable objects based on a simultaneous computation of force and displacements on
the physics-based model. The computation of the FEM model is shorten using spectral Lanczos decomposition method. In James et Pai (2005), a unified approach to the interaction with elastostatic contact simulation is presented. In this case, the contact resolution is based on capacitance matrix which relates the imposed displacements and response forces\(^2\). The method is used for single-point as well as grasping interaction, where haptic rates are achieved using precomputed Green functions. In Barbič et James (2008), the contact problem is solved using a penalty-based method allowing for multiple contacts and self-collisions. Both tool and obstacle are deformable and simulated by finite elements methods optimized thanks to model reduction. However, as already mentioned in the previous chapter, the penalty-based methods cannot guarantee the non-interpenetration, they are very sensitive to the choice of the penalty parameters and do not integrate properly static/dynamic friction. In the context of haptic rendering, it could lead to additional problems with stability.

Therefore, based on an example of virtual snap-in simulations Duriez et al. (2003), (where stiff contacts and deformations are closely linked) we introduced Signorini’s model for contact handling in the field of haptic rendering. In Duriez et al. (2004a), we proposed a first multi-threaded approach based on solving a Linear Complementarity Problem\(^3\) at low rates while updating the constraints without changing their status, at high rates. In Duriez et al. (2006b), friction contact response synchronized with deformations is proposed. The computation performance is obtained through the use of a precomputed and condensed compliance matrix. Models are limited to small displacements deformations but the global corotational approach allows for large global transformations. However, this approach is not realistic with soft-tissue models.

To overcome the limited refresh rate when dealing with deformable bodies, one strategy is to implement an intermediate representation of the constraints provided by the virtual environment, see e.g. Adachi et al. (1995), Mark et al. (1996) or Garre et Otaduy (2009). This representation allows for separating the haptic rendering from the physics engine, but at the cost of using a simplified model of the simulation Forest et al. (2004). In Jacobs et al. (2010) the simplified model relies on the linearization of the non-linear deformable that runs at low rates.

When trying to combine constraint-based approaches on deformable with intermediate representation to handle non-linearities, the main issue is to solve the constraints on a simplified model. This problem will bring us, again, to the compliance matrix which would be somehow the ideal simplified model as it provides the condensation of the mechanical behavior in the constraint space. For instance, in Garre et Otaduy (2009), a mixed-LCP formulation is used to solve both unilateral and bilateral constraints. The formulation relies on a simplified inverse matrix (which is similar to the compliance matrix) as only diagonal blocks are considered. Thus, indirect contacts cannot be rendered and the computation of the direct contact force is not always

\(^2\)This matrix is similar to the compliance matrix \(W\) described in the previous chapter, except that \(W\) is projected in the constraint space.

\(^3\)LCP are introduced in the previous chapter.
3.2. CONSTRAINT-BASED HAPTIC RENDERING FOR MEDICAL SIMULATION

accurate, especially for light and stiff objects.

On the contrary, in the method presented in Saupin et al. (2008a) and its extensions presented in the following section, we rely on an asynchronous computation, at low rates, of the full compliance matrix in the contact space. An approximation is used if the exact compliance is not computable in real-time at low rate, but as presented in section 2.3, we can get quite precise approximation thanks to GPU. This intermediate representation is completed by a set of constraints between mechanical models that are re-solved at high rates using the last position of the device, which makes the method very close to virtual mechanism technique.

3.2 Constraint-based haptic rendering for medical simulation

This section present haptic rendering algorithms which are compatible with the constraint-based models of mechanical interactions presented in Chapter 2. The section begins with a presentation of a work conducted with a company called Didhaptic in order to improve the quality of the haptic rendering of a commercial dental simulator Syllebranque et Duriez (2010). Here, the purpose is to demonstrate that the constraint-based formulation of the interactions can be adapted to rigid objects and extended to drilling. Moreover, to improve the haptic rendering of stiff contacts, we introduce an implicit formulation of the contact directions. Then, we come back to the liver example to illustrate the contributions on haptic rendering of mechanical interactions on soft-bodies. This work has been a long process, including the contributions of Guillaume Saupin during his Ph.D. and of Igor Peterlik, during his postdoc. In Saupin et al. (2008a) we have laid the foundation of the framework but the method was limited to contact constraints and the computation of compliance was relying only on compliance warping (see section 2.3). In Peterlik et al. (2011b), we have extended the approach to other type of interactions with a more generic formulation that is very close to the virtual mechanisms presented above. We called this method the multirate compliant mechanism. Very recently, in Courtecsis et al. (2013), we have used this approach in combination with the GPU method presented at section 2.3.3.

To obtain haptic feedback of good quality on medical simulations, we have focused our efforts on the rendering of mechanical interactions. Indeed, these interactions are the main source of forces felt by the physician during a procedure. In the simulation, the interactions are modeled by a set of equations that are linked to the compliance matrix of the interacting objects and to the constraint’s laws.

When rigid simulation is considered, the compliance matrix is linked to the mass and inertia of the objects and to the 6DoFs stiffness of the coupling spring. In this case, there is no real issue of updating the compliance at high rates.
CHAPTER 3. HAPTIC RENDERING AND MULTITHREADING APPROACHES

But when the simulation involves complex deformable model, it is not possible to build the compliance at high haptic rates. Our approach consists in sharing the constraint equations and the compliance between a haptic control loop (at high rates \(\approx 1 \text{kHz}\)) and the simulation (at low rates \(\approx 30 \text{Hz}\)). The constraints are built and solved in the simulation and then, re-solved in the high-frequency haptic loop in order to impose these constraints directly in the control loop. With this approach, it is possible to describe the specific behavior of various medical devices while relying on a unified method for solving the mechanical interactions between deformable/rigid objects and haptic rendering.

![Image](image.jpg)

Figure 3.1: Compliant mechanisms technique. The simulation shares the mechanical compliance of the objects and the constraints between them. The constraint response is being computed at low rate within the simulation and at high rates within a separate haptic thread. A 6-DoF damped spring is still used to couple the position of the device to its position in the simulation.

But first, let’s begin by introducing the use of advanced constraint modeling for haptic rendering in a case of mechanical interactions between rigid models. From a certain point of view, this context is simpler because multi-rate computation is not needed. However, obtaining haptic rendering of rigid contact with high fidelity is also challenging because the states and the directions of the constraints evolve quickly.

3.2.1 6DoF rendering of contact and drilling on rigid CT-scan model

For this section, we leave the context of hepatic interventions to explore the haptic feedback in the context of dentistry. More precisely, we aim at providing the haptic rendering of contact and drilling on the jawbone for implant surgery.

The virtual object (here the virtual drill) is attached by a 6DoF damped spring to the idealized reference frame (the position of the device). The constraints are given by the collision detection output between the drill geometry and the virtual environment. It is based on a distance map extracted from a CT-scan. This way the user sees only the realistic non-penetrating configurations between the virtual object and its environment. In our algorithm the collision detection is interactively called in the collision response algorithm so that there is no need for a continuous collision detection and we can rely on a fast distance map.
3.2. CONSTRAINT-BASED HAPTIC RENDERING FOR MEDICAL SIMULATION

Figure 3.2: Illustration of 3D distance map extraction from CT-scan datas (here for one slice): (a) Original CT-scan slice (b) Binarization (c) Border extraction (d) 3D distance map to surface (e) Signed 3D distance map to surface (green = outside, positive, and red = inside, negative)

Collision detection between jawbone and drill models: The map is precomputed off-line and updated on-line at high rates during drilling. It is extracted from a CT-scan of a real patient and discretized homogeneously using voxel samples of a size about 0.2mm\(^4\). First, we binarize each CT-slice to obtain an inside/outside map. Then, we assemble those binarized slices to get a volume block and extract borders. Finally, for each of these border voxels, we compute a distance field in neighborhood (within a given distance \(d\)). Inside the volume, the minimal distance is inverted. This process is illustrated in Figure 3.2. The result is a discrete potential field holding minimal distance to the object surface that can be interpolated inside each voxel using a trilinear interpolation. Using this approach, we obtain an implicit surface of the jawbone that can be evaluated very quickly (if \(\chi\) is the potential function evaluated at point \(P\), \(\chi(P) < 0\) if and only if the point \(P\) is inside the jawbone).

For the collision detection with the virtual drill, we use the method proposed by McNeely et al. (1999). The drill is modeled using pointshell: a set of surface point samples associated with inward-pointing surface normals. The collision computation consists in a simple evaluation of the potential function and the contact normal is taken from the pointshell normal. As the collision technique is not continuous, we could miss some collision events. To avoid this problem we integrate the collision detection in the multi-contact process. In the haptic loop, we rely on a Gauss-Seidel iterative solver. At each iteration, each point \(\alpha\) of the drill is visited and the following steps are performed:

1. The position of \(\alpha\) is computed \(P_\alpha = P_\alpha^{\text{free}} + H_\alpha (K + hB)^{-1} r\) where \(r = \sum_{\beta=1}^{\alpha-1} H_\beta \lambda_\beta + \sum_{\beta=\alpha+1}^{N} H_\beta \lambda_\beta\) is a 6DoF wrench vector \(r\) that is the sum of the contribution of all pointshells except \(\alpha\). It corresponds to the application of the Gauss-Seidel strategy (see equation 2.23 in section 2.2.3) to the rigid drill model. Here \(r\) can be accumulated instead of being recomputed for each pointshell \(\alpha\), leading to a linear complexity.

\(^4\)CT-Scan resolutions vary from 0.1mm to 1mm. We resample the CT-Scan using trilinear interpolation.
2. A collision detection is performed by evaluating $\delta = \chi(P) = 0$ and if $\chi(P) \leq 0$ evaluate $F = \{ n, t, s \}$, the normal and the tangential directions of the contact. If $\chi(P) > 0$, go to the following pointshell.

3. The local compliance $W_{aa}$ is updated using the frame $F$.

4. If drilling, an erosion model is computed (described in the following). The parameters of the model include the density of the bone, $W_{aa}$ and $\delta$. The output is a variation of the interpenetration $D\delta$. A CSG (Constructive Solid Geometry) algorithm updates the distance map.

5. The contact response $\lambda$ is computed by solving Signorini’s law and Coulomb’s friction on pointshell $a$, given the current interpenetration $\delta - D\delta$ and the local compliance $W_{aa}$. Note that the friction is cancelled when drilling.

Figure 3.3: Computation process for one time step: (a) In-between collision position $x_c$, (b) For each colliding pointshell, an erosion rate is computed, (c) the associated interpenetration distance is eroded and the computation follows with the Gauss-Seidel algorithm except that the new drill position $x_{go}$ can penetrate the inside jawbone with a depth equal to the shrinkage of $\delta$, (d) Afterwards, the distance map is updated using an algorithm close to the CSG substraction, (e) to obtain fast and precise geometrical results on the implicit surface, (f) Finally, the algorithm restart for a new time step.

**Erosion computation**. Using a CT-Scan of a real patient jawbone and the method presented in Sjögren et al. (2002), we compute a volumetric density map. We use this map to estimate locally the density, i.e. the quantity of material that has to be
removed to erode one voxel. The local erosion follows an analytical model that is close\(^5\) to the one proposed by Agus et al. (2002). After the integration of the erosion analytical model, we store a new density for each concerned voxel. If the density on a voxel vanishes, then we shrink the interpenetration distances by the size of this voxel (it corresponds to the value of \(D_\alpha\)). As the interpenetration distances were eroded, the solution does not provide a collision-free configuration: the drill tool penetrates the jawbone surface according to the erosion computation.

When the jawbone is eroded, we need to update the distance field map that we use for the collision computation and interpenetration evaluation. We precompute another distance field map for the drilling tool\(^6\) (see figure 3.3), but we inverse the values of this map (positive values become negative and vice-versa). Then, we perform a CSG algorithm with a very simple operation on all voxels of the map that are close to the drilling tool.

**Results:** The presented algorithms are implemented on a surgery training simulator solution, named VirTeaSy Implant Pro, which is commercialized by the French company Didhaptic\(^7\) (cf Fig. 3.4). The haptic system is composed of an implantology hand-piece fixed on a 6DoF Desktop, Haption force feedback device.

Figure 3.4: The surgery training simulator solution VirTeaSy Implant Pro - Didhaptic: (a) global view (b) haptic device (c) Virtual environment (d) Virtual working zone closeup

In the paper Syllebranque et Duriez (2010), we show that the haptic rendering is able to reproduce the cortical breakthrough. The jawbone, like most of human bones, is composed of two different layers: the core, called spongious bone, is protected by a harder one, the cortical bone. Thus, when the cortical is about to be pierced (cortical breakthrough), the dentist must be careful not to apply an excessive force that would quickly drive in the drill tool too deeply. As the algorithm relies on physics and is

---

5It only differs because the shape of our drill tool models are different.

6We use a method similar to the one we developed for the jawbone (cf Fig. 3.2), except that binarization step is made from the tool profile.

7www.didhaptic.com
based on real data, we can accurately reproduce this particular difficulty of this surgery as well as the perturbations provided by the jawbone inhomogeneities.

Another feature of the algorithm is the precision needed for a good training: the final hole is obtained by using several drills and by increasing progressively the diameter, in order to avoid cracking the bone (see figure 3.5). Consequently, the geometrical precision of the contact response is a key feature. For instance, a drill must not be able to enter a hole made by a smaller one. The global convergence criterion of the contact response must then be small enough to ensure such behaviour. In our tests, we use a precision of $10^{-2}$ mm with 0.2 mm square voxels.

![Figure 3.5: The diameter of the hole is progressively increased. On the right, the erosion model allows to estimate the temperature of the bone during drilling. The physician need to avoid burning the surface of the bone which complicates the placement of the implant.](image)

For the computation time, in scenarios with many constraints, the contact algorithm takes between 2 ms and 3 ms to converge. At this maximum computation time, 63 pointshells were detected simultaneously in collision. The rate obtained by the drilling part is related to the number of voxels involved in the distance map update. In the worst case, about 10000 voxels are in intersection with the tool and are updated in 3 ms and the global refresh rate is near 166 Hz. As it appears during drilling, the contacts are less stiff which is compatible with this decrease of performance. In future work, to maintain a fast refresh rates, the contact algorithm and the update of the distance map will be computed in parallel using a multi-rate approach. Precisely, let’s introduce a multirate approach for haptic simulation of deformable objects.

### 3.2.2 Multirate solving of contact constraints

In the approach published in Saupin et al. (2008b), we aim at providing haptic rendering of contacts with soft-tissues undergoing larger deformations, i.e. we included the geometrical non-linearities of the deformation models. It extends the approach developed in Duriez et al. (2006b) that was limited to linear deformable models. For solving contact response in the simulation, we rely on compliance warping, as already explained in section 2.3.2. Using the method, an approximation of the compliance is
available in real-time but with large time steps (±30 ms) that are far from the haptic rates. Moreover, the method should include the collision detection and the solving process of the deformation models which are already difficult to perform at 30 Hz.

Thus, we employ a multi-rate implementation for the coupling scheme between the device and the simulation (see Figure 3.6): After collision detection and computation of the deformable models, the LCP is built and solved in the simulation, exactly as presented in section 1.2. But it is also shared (step 1) with a separate haptic loop, computed at high rates by another thread. In the haptic loop, the position of the device is refreshed and allows to update the free motion of the virtual solid that is driven by the motion of the device $\Delta q_h$. This free motion provides new violation values in the constraint space: $\delta_{\text{free}}^{\text{new}} = H\Delta q_h$ (step 2). Given this new value of $\delta_{\text{free}}^{\text{new}}$, we compute a new value for the contact response (step 3) and send it to the force feedback $r = H^T \lambda$. Step 2 and 3 are repeated in a loop at 1kHz.

Inside the haptic loop, most of the computation time is spent by solving the LCP. To improve the solver, we rely on the temporal consistency: we use the last computed solution of the LCP as an initial guess (which is one of the simulation when step 1 is applied). As it is close to the solution, it converges quickly. Moreover, as we use an iterative solver, we can introduce a time criterion in the Gauss Seidel algorithm.
in order to stop the iterations before the end of the haptic step. It provides the best possible solution in the given limited time (less than 1ms)

**Discussion:** The method uses a 6 DoFs damped spring between the position of the device and the constrained simulated position (in both loops). The compliance of this spring is computed and integrated in the resolution process of the constraints. For a good rendering, the stiffness parameter needs to be tuned so that the compliance of the spring is negligible compared to the compliance of the contacting deformable models.

An other important point is the type of contact detection: The contact constraints are set in the simulation before being actually processed and rendered in the haptic rendering. It is important to have a collision detection that somehow anticipates the contact appearing in the haptic loop between two steps of the simulation loop. To have this property, the method relies on the computation of local minimal distances (proximity queries) between meshes: if two points on colliding meshes have a distance that is inferior to a given *alarm distance*, we set a potential contact between them. Consequently, if the virtual tool is approaching an obstacle in the simulation, the contact is set and can be activated in the haptic loop (for instance if the user continues its motion towards the obstacle) before collision in the simulation.

### 3.2.3 Multirate Compliant Mechanisms

**Motivation and Contribution** In the two methods described above, the constraints that are rendered mainly involve contact constraints. These approaches are not sufficiently generic to account for force feedback of complex various mechanical interactions on both rigid and deformable solids.

In section 2.5, we show that complex procedure, like (flexible) needle insertion through several layers of soft tissues, necessitates the use of a set of constraint-based models to capture the physics of the interaction. In addition, if we consider the procedures on the liver like the needle biopsy or a suture under laparoscopy (which were the main
examples in section 2.5), in both cases, the haptic feedback is very important for the physician.

No existing method proved to cover haptic rendering for all possible task scenarios in interactive surgical simulations. Therefore, the method presented in this paper attempts to fill this gap by proposing a generic framework for solving complex interactions between various medical devices and anatomical structures, and subsequently computing the associated haptic rendering. It accounts for large deformations of the tool as well as the organs while at the same time, it handles arbitrary types of constraints, such as unilateral contacts, non-linear friction as well as bilateral ones.

From the computer haptic viewpoint, the approach may be classified as relying on an intermediate representation. The dynamics of the virtual objects is computed at low-rate in the simulation and the interaction forces are modeled and solved using constraints, exactly like in section 2.5. But these forces are re-computed at high-rate, in the haptic loop based on an intermediate representation shared between the two loops. This intermediate representation includes (i) a physics-based mechanical coupling between the different interactions (based on the compliance of the interacting solids’ model) and (ii) a set of kinematic links that capture the behavior of the interactions and which are modeled by using constraint laws). We call this intermediate representation the multirate compliant mechanisms (see figure 3.8):

- The term *mechanisms* emphasizes that we can support an extensive number of interaction types as soon as it relies on an adequate force and/or motion transmission model.

- The term *compliant* is used because the mechanical coupling between the interacting objects is based on physics, more precisely on compliance matrices, that are computed with FEM models for deformable or with 6DoFs models for rigid bodies.

- The mechanisms are simulated at low rates. However, they are also shared with the haptic loop, where they are recomputed at high rates for an intuitive and passive control. Since the same mechanisms are computed at two different refresh rates, we add the term *multirate*.

Overall, Fig. 3.8 illustrates the motivation of the method: soft tissue deformation and needle bending are computed in real-time at a low frequency (~ 30Hz) in one thread, while another thread computes at high frequency (> 500Hz) the mechanisms defined for this problem. As a result, the force felt by the user describes correctly the interaction of the needle with the two soft tissues. We think that this contribution unifies the God-Object approach and Virtual Mechanisms (with a virtual environment) while extending these methods to deformable objects.
CHAPTER 3. HAPTIC RENDERING AND MULTITHREADING APPROACHES

Genericity of the concept To show the genericity of the approach, let us now consider the case of a laparoscopic procedure (see Fig. 3.9). The instruments are inserted through a small incision in the abdominal wall. They are generally considered as mechanically linked with the operative field through a fixed insertion point, and the design of dedicated haptic rendering devices often reflects this hypothesis. In reality, the degrees of freedom of the laparoscopic device are constrained by a cylindrical joint (Fig. 3.9 (a)) corresponding to the trocar that is fixed on the abdominal wall and through which the instruments slide with friction. Therefore, a more accurate haptic rendering could be computed by modeling the deformation of the abdominal wall (for instance with a FEM model updated at a low frequency, see Fig 3.9(b)), and by modeling precisely the interaction, in the one hand between the instrument and the trocar, and in the other hand, the instrument with the body’s tissues. If we suppose that we use a 6DoFs haptic device\(^8\), we could constrain the motion of the haptic device using a compliant mechanism (Fig. 3.9 (c)). In order to match the mechanical constraints described in the simulation, this multirate compliant mechanism will act directly on the control of the interface, at high frequency rates.

Complex interactions, like grasping and suturing, can be modeled by more complex constraint laws, and/or by combining several compliant mechanisms together. As an example, we have shown that the grasping of a deformable object can be modeled using complementarity constraints Saupin et al. (2008b). In other cases, it is necessary to

\(^8\)Several haptic devices that are dedicated to laparoscopic procedure, like the Xitact IHP device (now owned by Mentice Corporation) have only 4DoFs: the entry point in the abdomen is supposed to be fixed. But other devices, like the Freedom 7S, are 6DoFs haptic device with additional graspers. We suppose that we use such a device. See the bibliography of Guiatni et al. (2012) for details.
3.2. CONSTRAINT-BASED HAPTIC RENDERING FOR MEDICAL SIMULATION

Figure 3.9: Multirate compliant mechanism applied to laparoscopy: (a) The motion of the laparoscopic tool is constrained by its interaction with the abdominal wall. (b) A complete simulation of the operative field can be setup and a FEM model of the abdominal wall is computed at low rates. (c) From the simulation, we extract the mechanical compliance of both abdominal soft tissues and articulated system and the interaction joint (the mechanical constraint) between them. This forms the mechanism that is recomputed at high rates by the haptic controller.

... combine several advanced compliant mechanisms together in order to feedback the influence of indirect constraints (i.e., constraints that are not directly applied to the coupled virtual object as illustrated in Fig. 3.10).

Figure 3.10: Indirect constraint influence: in this example the haptic feedback mainly comes from the contact between the two tori. Thus, compliant mechanisms are dynamically set by the simulation for both friction contacts between the grasper and the deformable torus but also for the contacts between the two tori.

Setting multirate mechanisms As mentioned before, interactive medical simulations using accurate deformable and complex interaction models can hardly run at haptic refresh rates in the same loop. In general, the refresh rate of these simulations range typically between 30 Hz and 60 Hz (vision rendering rates). While using a fast intermediate representation for haptic, it is important to keep an accurate modeling of the forces between interacting bodies. Thus, we propose to share the constraint-based models computed in the simulation with the haptic control loop in a unified formalism called multirate compliant mechanisms.
The definition of the intermediate representation based on compliant mechanisms relies on:

1. the constraints law $\text{law}(\delta_\alpha, \lambda_\alpha)$ for each constraint $\alpha$ and their value during the free motion $\delta^{\text{free}}_\alpha$, and

2. the compliance of the simulated models given by equation (2.22).

The same mechanisms are computed in the simulation and in the haptic loop: We use the unified processing described in the section 2.5.2. There is no model simplification or reduction to provide the user with high quality haptic rendering. The only difference is that the Gauss-Seidel in the haptic loop can only iterates in a very short period of time (less than 1ms). We could improve its convergence by providing a good initial guess issued from the simulation or from the previous computation of the haptic loop.

The figure 3.12 illustrates the main steps of both simulation and haptic loops as well as the data that are shared between the two loops. As well as explained in the scheme 3.7 we use proximity detection in order to anticipate contact contact constraints.

Note that when convergence is reached in the haptic loop (which is almost always the case in our experiments) the constraint law defined for each mechanism is fulfilled for the haptic rendering. It means that the haptic feedback is directly linked to the modeling of the interactions. More details about the algorithms and the implementations are provided in Peterlik et al. (2011b). We emphasize that the method is highly linked to the compliance computation on the simulation. In the original implementation, we were using the Compliance Warping method. Recently, in Courtecuisse et al. (2013) we have used the GPU implementation of the asynchronous preconditionner for contact. Without any change in the method, this new version provides even more realism in the haptic feedback.

**Results and discussion**  Several tests were performed to evaluate the method. First, we reproduced some experiments that were described in the literature. For instance,
3.2. CONSTRAINT-BASED HAPTIC RENDERING FOR MEDICAL SIMULATION

![Diagram of communication between haptic and simulation loops]

Figure 3.12: Communication between haptic and simulation loops

In Dehghan et al. (2008), they measure the forces when inserting a needle in a phantom cube made of silicone. We compared the force profile that is obtained in the haptic loop with the data measured. The results show that the profile in the simulation is very similar to the experimental results.

![Snapshots of needle insertion simulation]

Figure 3.13: Four snapshots of the simulation of needle insertion. Haptic rendering allows to feel the puncturing, the cutting, the friction and the sliding forces due to the interaction between the flexible needle and both abdominal wall and liver models.

Second, we realized the simulation of a liver biopsy in order to illustrate that the method allows for stable and realistic haptic rendering on quite complex simulation. the needle is first inserted into a deformable object representing the abdominal wall. After puncturing the wall, the needle is further inserted into 3D deformable model of liver (see figure 3.13). This simplified simulation allows to illustrate the different interactions (penetration constraint, cutting force at the tip of the needle, friction forces...) that are modeled and applied to the haptic rendering. Then we used the same method in a more realistic simulation of a liver biopsy.
The data set comes from a real-patient and the geometries were reconstructed by the IRCAD\textsuperscript{9} team. In addition to the interaction constraints between the needle and the liver, the contact constraints between the liver and its environment can also be simulated and improve the haptic rendering quality.

According to our experimental studies, the threshold frequency of the simulation loop is about 30 Hz, which could be sustained even for complex simulation with one hundred volume constraints. Figure 3.14: Photo of the liver biopsy simulator

If simulation is running at a frequency superior to this threshold, both stability and accuracy of the interaction is guaranteed. However, for lower frequencies, the interaction becomes instable and inaccurate. It can be explained by two factors: First, the timed Gauss-Seidel method does not give reliable results as the initial guess provided by the simulation is too far from the optimal solution associated with the actual position of the haptic device. Second, the constraints are created with a delay that leads to visually inconsistent simulation.

**Conclusion and future work** This section presents several strategies that we have employed to provide haptic feedback on solid simulation (from rigid to non-linear deformable models). The main advantage of our approach is to have an accurate intermediate representation of the deformation in the haptic loop through the compliance matrix. Moreover, the haptic rendering is directly linked to the processing of the constraint equations, thus to the model chosen for the mechanical interactions. If the mechanical interactions are modeled carefully with complementarity constraints, the haptic feedback is quite convincing.

Currently, the method was only tested with corotational FEM model, in future work, we plan to test and integrate the method with fast hyperelastic FEM models. Moreover, some limitation remains, like the difference of frame rates between the simulation and the haptic loop: there is the assumption that the direction of the constraints remains constant between two simulation steps. If the simulation is too slow, this assumption is not valid and some artificial bumps can be felt in the haptic rendering. Consequently, it remains a central point to optimize the simulation (proximity detections, FEM models, constraint process...) as much as possible.

A second limitation appears when simulating light deformable objects with high stiffness properties. Indeed, it leads to high-frequency dynamics: from one haptic step to an other, the compliance could change suddenly. This problem is partially addressed in the following section.

\textsuperscript{9}IRCAD: www.ircad.fr
3.3 Asynchronous simulation of deformable models with variable stiffness

In the following section, we present a contribution that is related to haptic feedback but that could have a more general impact on real-time simulation domain. Indeed, each deformable model has its own mechanical properties and its own computation cost. Thus, each deformable model has a small range of possible time-steps that are ideal in a real-time context. Yet, as soon as interacting, objects must be simulated at the same rates to have stable and reliable results.

In the paper Peterlik et al. (2011a) that is summarized in this section, we propose a new strategy. A stiff object with few degrees of freedom can be simulated at higher rates than a complex FEM model of soft-tissue, while maintaining a good quality in the simulation of the interactions. However, the method is limited to quasi-static behavior, so that the time step chosen for each object does not appear in the measure of the compliance.

3.3.1 Motivations: Why adapting the refresh rates to each object?

Solids are generally separated in two categories: rigid objects and deformable ones. Indeed, from the modeling point of view there a big difference. Rigid objects have only 6 DoF whereas the number of DoFs of a deformable model is defined by its discretization. However, from the physics, a rigid object is just a solid object that is sufficiently stiff, in a given context, for considering that its deformation can be neglected.

Stiffer objects have a dynamic response at high frequency, it means that they should be simulated at high rates. Additionally, their models are generally (not always) simpler because they undergo smaller deformations. On the contrary, complex soft-tissues such as anatomical tissues have a low dynamics but need complex and detailed FEM deformable models which require long time steps. However, if stiff and soft solids are interacting, models must be simulated at the same rates to have stable and reliable results. In a non-real time context, small time step would be chosen to have a precise integration of the stiff model and we would have the computation complexity of the soft object for each time steps. In a real-time context, the situation is more tricky: each model could be simulated in real-time separately. Nevertheless, the ranges of possible time-steps, for which the models are real-time and accurately modeled, are not intersecting.

Let’s take the example, illustrated by Fig 3.15, of a virtual thread attached at one end to a fixed support and controlled at the other end by the haptic device. A characteristic of a thread or very thin cable is that it has a very low bending energy but high stretching energy (high stiffness). As a consequence, a nearly null haptic force is generated during bending. But when the user pulls, on the thread a stiff force appears
very suddenly when the thread is straightened out. The haptic rendering of this example can be compared to the well known problem of virtual walls: to obtain a stiff rendering, one needs to increase the frequency of its simulation in the haptic control loop. In other words, in our example, the simulation rate of the entire thread needs to be very high (e.g. 1kHz, about the same frequency as for virtual walls) to be able to reproduce this rapid change in stiffness during the manipulation of the virtual thread. Assuming this is possible for the thread model, the rest of the simulation (the other deformable models, the collision detection and response) will most likely need to be updated at slower rates (i.e. ≈ 50Hz) due to their complexity. This example is not unique in real-time simulation of deformable objects and is not limited to haptic problem. There is a practical need of adapting the refresh rate of each object depending on the characteristic of its deformations.

3.3.2 Asynchronous quasi-static deformable models in contact

In Peterlik et al. (2011a), we present a first attempt to solve this problem that is limited, for now, to quasi-static models in a context of haptic rendering. We introduce a technique that maintains a coherency between interacting deformable models that are simulated in real-time but at two different frequencies. Similarly to the technique employed for haptic rendering, the method is based on sharing constraints equations between the high rate and the low rate loops. The shared compliance matrix allows to have a prediction (computable at high frequency) of the behavior in the constraint space, for an object simulated at low frequency. Using it, the motion of an object simulated at high frequency is always constrained by low frequency objects, leading to coherent behaviors in both loops.

To illustrate the method, we use two different interacting deformable models: a curved deformable object (guided by the haptic device motion) is simulated at high rates (such as 1kHz) and a volumetric obstacle is computed in a low-rate loop (e.g. 50 Hz). The model used for both objects allows for large deformations.

It should be emphasized that the method is not limited to this scenario: the only
limitation is the compatibility of the computation model with high refresh rates. But, for instance, volumetric linear deformable models, computed at high rates can be employed instead of the beam model.

Throughout this section, following notation is used: $\mathbf{W}_v$ and $\mathbf{W}_c$ corresponds to the compliance matrix of the volumetric and the curve model, respectively. The left superscript $H$ indicate that the compliance $\mathbf{H} \mathbf{W}_c$ is computed at high rates while the left subscript is used if the compliance is buffered between the thread. $\mathbf{L}^{-1} \mathbf{H} \mathbf{W}_v$ is computed at low rates but is accessed from the high-rate thread via buffer.

**Deformable models and contact process** The non-linear deformation of the volumetric object is computed using a corotational finite element approach, whereas the model used to simulate the deformation of a curve-like object at high rates is modeled using serially-linked beam elements (see chapter 2). Both models relies on quasi-static model, so that the compliance matrices do not rely on any time step $h$ parameter. Indeed, when computing $\mathbf{W}$ in equation 2.9, the matrix $\mathbf{A}$ contains $h$. It would prevent from adding $\mathbf{W}_v$ computed with $h_{\text{low}}$ to $\mathbf{W}_c$ computed with $h_{\text{high}}$. Here, the compliance of an object is obtained by computing $\mathbf{W} = \mathbf{H} \mathbf{K}^{-1} \mathbf{H}^T$.

Collision detection is often one of the bottlenecks of the simulation performance. It seems more adapted to place this process in the low rate loop. But to avoid synchronization problems during the interactions between low-rate and high-rate object, a proxy of the highrate body is introduced. It can be regarded as an interface of the curve object that provides information needed for contact collision and response computed at low rates.

The contact process begins with a proximity queries algorithm that is computed at low rates. For the proxy, we use the position that was stored at the end of the previous low step. All detected pairs of close points are considered as being potentially in contact. As previously, the directions of the contacts are stored in the matrices $\mathbf{H}^T$ for each object involved in the interaction. For the contact response, we apply the Signorini’s law and Coulomb’s friction. However, like the multirate compliant mechanism approach, the method here is generic and other type of constraints used to model the interactions can be used.

**Multirate computational model:** In order to solve these constraints in the low rate loop, we need the compliance but we also need $\delta_{\text{free}}$ (see equation 2.8) that is based on the free position of both objects. For the volumetric object, nothing changes in the computation of $\mathbf{x}_{\text{free}}$ compared to previous approaches. For the curve deformable model, we assign the free motion that is computed at high rate $H \mathbf{x}_{\text{free}}$. As the loops are not synchronized, we use the last buffered free position, that does not necessary perfectly correspond to same simulation time. We use this free position to evaluate the constraint violations in the low rate loop.
The compliance equation of the constraint problem is written the following way (still in the low rate loop):

\[
\delta = \left[ L W_v + H W_c \right] \lambda + \delta_{\text{free}} \quad (3.1)
\]

The compliance in the constraint space of the curve model \( H W_c \) is computed at high rates but buffered to be accessible from the low-rate thread. The constraints are first solved by the Gauss-Seidel-like algorithm at low rates (see section 2.5.2), resulting in solution vector \( L \lambda \). Then, the positions of both volumetric model \( x_v \) and curve object proxy \( L x_c \) are updated using \( x = x_{\text{free}} + A^{-1} H^T \{ L \lambda \} \). Note that the position computed for the proxy is now synchronized with the position of the volumetric object. It is then stored to be used in the proximity queries in the next time step.

Figure 3.16: Schematic visualization of the computational model.

Moreover, like the methods presented in the previous section, the constraint problem is also shared with the high-rate loop and recomputed according to the curve object deformations. In that case, the compliance equation in the constraint space is written:

\[
\delta = \left[ L H W_v + H W_c \right] \lambda + \delta_{\text{free}} \quad (3.2)
\]

where \( L H W_v \) is the buffered compliance value for the volumetric object. On the contrary, for the curve object, the compliance is updated according to the last position in the high rate loop. \( H W_c = \{ L H H_c \} \{ H H^{-1} \} \{ L H H^T \} \). Note that the contact directions \( H \) are computed at low rates and buffered to be used at high rates. The scheme depicted on figure 3.16 summarize the exchanges between the two loops.

**Haptic rendering** Obviously, the two loops (at high and low rates) are launched in separate threads. The haptic feedback is synchronized with the high rates loop which
allows for using a virtual coupling on the curve model without any unwanted force rendering due to delay. However, since the quasi-static deformations of the high-rate object is performed directly in the haptic thread, deformation model must be fast enough to keep the haptic rate. Another bottleneck is the constraint solver that may not reach convergence in the allocated time at high rates. To solve this problem, we use a timed version of the Gauss-Seidel algorithm (that does not wait for convergence if the time limit is reached, like introduced in section 3.2.2).

**Results** In the following, experiments are presented in order to demonstrate the multi-rate approach.

First, the validation of *action-reaction* principle is performed using two beam models having the identical physical properties, each being calculated at different frequency. The low-rate curve-like object is fixed in the space, whereas the high-rate one is attached to the low-rate body on one end and to the haptic device on the other. Although each model is calculated at different frequency, the deformations of both objects is identical for arbitrary position of the haptic device as shown in Fig. 3.17.

![Figure 3.17: Validation of action-reaction principle. The deformations of low-frequency object (green) and high-frequency object (red) are the same.](image)

Second, we implement a snap-in scene, depicted in Fig. 3.18. In this example, the high-rate object is represented by a stiff but deformable clamp attached to the haptic device via coupling spring. The clip is modeled with 20 corotational beam elements and nodes with positional and rotational degrees of freedom were used in the model. As the obstacle, a deformable cylinder composed of 103 tetrahedral elements is simulated. During the snap operation, up to 20 friction contact constraints are created.

This example allows to test the haptic rendering with various rigidities of both cylinder and clamp. It demonstrates that we are able to model complex contact problem between objects which are simulated at various refresh rates. Moreover, even with stiff mechanical properties for the clamp, the haptic rendering remains stable; it should be noted, however, that the stiffness of the damped spring used for virtual coupling needs to be adapted accordingly.
Further, our approach is validated using a scene with thread and deformable cylinder. The thread is modeled with 10 beam elements (computed at 1000 Hz) and cylinder described in the previous example was used (simulated at 50 Hz). One end of the thread was attached to the cylinder, while the other end was manipulated by the haptic interface. Beside the bilateral constraint between the thread and cylinder, some contacts between the two objects were detected and resolved resulting in more complex interaction. Despite the complexity of the simulation, both stable and realistic haptic force feedback was delivered to the user.

The same simulation is performed for the scenario where the thread is also computed at low rates (50 Hz). In this case it was observed that the simulation becomes rapidly unstable. This instability appeared mainly in the moment when the user rapidly pulled the thread, changing its configuration from loosely bended to stretched.

We think that the beam model displays an important change in stiffness during very short time when the deformation is changed from bending to stretching. To handle this transition correctly, the simulation must be run at high rates, as was demonstrated by our experiments. Nevertheless, with increasing stiffness of the thread, even the frequency of 1 kHz might not be sufficient and higher rates would be needed to guarantee the stability during the pulling procedure.
3.3.3 Further extensions of the approach

The method allows for accurate modeling of constraints between non-linear deformable models computed at two different frequencies. The method is particularly interesting in the context of haptic rendering of stiff deformable objects, but may be generalized to dynamic simulation of deformable bodies with heterogeneous stiffness properties. Another potential application is the control of a robot in a soft environment.

Extension to dynamics Currently, the most important limitation of the method is the hypothesis of a quasi-static modeling of the deformations. The use of quasi-static scenario highly simplifies the problem because, as explained previously, dynamic models generate differential equations that need to be integrated in time with a numerical method. In the context of real-time, the time step \( h \) used in the numerical time-integration method needs to be chosen accordingly to the computation time. It means that the two interacting objects that are not computed at the same refresh rates would not use the same time step \( h \). This creates additional difficulties that are not solved at the moment.

Indeed, in equations 3.1 or 3.2, when using the compliances in (respectively) the low and the high rate loops, we add values computed at low and high rates. When compliance matrices \( W \) are computed from a dynamic model, like in Equation 2.9, they rely on the inverse of a matrix \( A \) which depends on \( h \), the time step used in the time integration scheme. In that case the addition of a compliance \( L_W(h_{low}) \) computed at low rate and a compliance \( H_W(h_{high}) \) computed at high rates is incorrect.

Moreover, the state of a static mechanical system is defined by the sole position. In the computational model depicted in Figure 3.16, only the free position is shared between the two loops. If dynamic models are employed, then both velocities and positions needs to be shared. However, it could lead to incoherent results because the integration schemes at low-rate loops have the natural tendency to create damping. It means that the velocities computed at the low-rate and high-rate loops are not always coherent. Indeed, in the low-rate loop, the simulation accounts only for the low frequency variations of the velocity field.

During the master thesis of François Dervaux (2012), we made a precise list of all these problems and found first answers (like the use of numerical low-pass filters), but we have not yet led to a unified solution that works in all cases.

Extension to robot control In the context of robotic surgery, we could build some bridge between interactive medical simulation and robot control. For instance, using interactive FEM approach, one can simulate the deformations induced by the robot on the surgical environment. When complex interactions take place in a deformable environment, like during a suturing task in laparoscopy Nageotte et al. (2005b), it
would be very useful to be able to simulate in real-time the behavior in order to update the path-planning inside the control loop. Moreover, surgical robotics relies more and more on flexible structures that are also deforming when contacting the soft-tissues.

In the same time, robots have generally a simple model (deformable or rigid) that allows for the control of its actuators at high rates. So, the simulation of the interactions between a robot and a soft environment seems to be a good application for the method described in this section. It also seems to be particularly promising in a context of teleoperation: As well as virtual mechanisms in Joly et Andriot (1995) were proposed in a context of teleoperation in a rigid environment, we could impose some constraints on the motions of a robot according to the deformation of the soft environment. Unlike the context of haptic rendering described in section 3.2, this would necessitate modeling the robot inside the simulation. In that case, it could be advantageous to compute the model of the robot at high rates, while simulating its environment at low rates.
Applications, research projects and related activities

Table of contents

4.1 Ongoing research projects ........................................... 116
  4.1.1 New tools for interventional neuroradiology ..................... 117
  4.1.2 Training and rehearsal of procedures for cardiac arrhythmia ..... 119
  4.1.3 Brain shift during Deep Brain Stimulation (DBS) procedures . . 121
  4.1.4 Middle ear surgery simulation and robotic assistance .......... 123
  4.1.5 Simulation of a pathologic pelvic system ....................... 125
  4.1.6 Percutaneous procedures for liver tumor ablation .............. 127
4.2 Technology development and transfer based on SOFA framework ... 128
  4.2.1 Development of the framework SOFA .......................... 129
  4.2.2 Cataract simulation ........................................... 130
  4.2.3 InSimo: a spin-off from Inria dedicated to medical simulation . 133
4.3 Other research activities ........................................... 134
  4.3.1 Scientific Animation .......................................... 134
  4.3.2 Contracts / Funding ......................................... 136
  4.3.3 Teaching and scientific diffusion ............................. 137
In the previous chapters, we focus on the technical aspects and challenges in medical simulation. We often refer to liver surgery to illustrate the simulation techniques and their potential applications. However, no concrete example of a simulator is exposed yet. This is the purpose of this chapter, which presents some research projects and application developments. In our team, we share the desire to put our research work on the hands of physicians. Therefore, we need to develop concrete applications, that can be tested on prototype of simulators.

The realization of concrete applications, even prototypes, is difficult to do alone in a lab. The research in our field is very multidisciplinary and one of the key aspects is to build a network of research groups which have competencies in related fields. The most important is probably to have in each project (at least) one physician who is motivated by introducing simulation in his field, and has the taste of research (and a lot of patience...). But other expertise could be very important depending of the type of chosen application, like medical image analysis, geometry algorithms, biomechanics (experimental measures, new constitutive laws...), realistic real-time rendering, haptic device,...

From the beginning of SOFA, there was this idea of facilitating cooperations inside the community to help the development of new projects. This objective is maybe one of those that is best fulfilled today.

In the first section of this chapter, current research projects are introduced as well as the corresponding collaborations. These ongoing projects are dedicated to specific applications for which we expose briefly the interest of the simulation. The variety of these projects, for very different surgeries, shows the potential impact of simulation in the coming years.

In a second section dedicated to technology development and transfer, we will first discuss the contribution towards the creation and the improvement of the framework SOFA. Then, an example of an application on cataract surgery, which is more mature for being transferred, is quickly exposed. Finally, I present the objective of inSimo, a start-up company that is currently created by our research team and in which I am involved as scientific advisor.

The third section details the several activities related to the job of researcher: European and National founding, scientific animation, reviews, activities for diffusion of science...

4.1 Ongoing research projects

In the following, we will present more concrete examples of the use of simulation in a clinical context. These projects are the motivation of the research work that is described above. The initial motivation of real-time medical simulation was to have
4.1. ONGOING RESEARCH PROJECTS

pedagogic simulators. The more recent projects are more dedicated to the use of simulation for planning or assistance.

4.1.1 New tools for interventional neuroradiology

The development of a high-fidelity simulation system for interventional neuroradiology (minimally invasive treatment of brain vascular disease), is a good example to illustrate the three main impacts that simulation could have on complex medical procedure.

(i) Training simulation: we developed EVE, an endovascular simulation dedicated to the training of neuroradiologists, especially to provide realistic catheter and guide navigation through physics based collision response with vessel walls.

(ii) Simulation for planning: we started the project SIMPLE with the Magrit team in order to provide a simulation tool that would help the preparation of coil embolization. The collaboration with this team on this subject is still active.

(iii) Per-operative guidance: Recently, Jérémie Dequidt, started a project called IDEAS in which the team is involved. The goal is to use simulation to reduce the Xrays sent to the patient and to provide 3D visualization of the instruments. This work was published in several journals and MICCAI conference papers.

Many pathologies of the brain vessels (Ischemic strokes, aneurysms, arteriovenous malformations... ) are now often treated using interventional neuroradiologic therapies which rely on the insertion and navigation of catheter devices inside the vessels. Instead of open surgery, it allows to reach the lesion of the arteries with therapeutic devices through a catheter. The treatment is delivered directly within the closed brain, using only image-based guidance. Consequently, the dedicated skill of instrument navigation and the thorough understanding of vascular anatomy are critical to avoid devastating complications which could result from poor visualization or poor technique. These procedures require an intricate combination of visual and tactile feedback, a very good knowledge of the anatomy, a good comprehension of each individual pathology, and extensive training periods. Moreover, the training is not possible on cadaver (no blood flow) nor on animal (no animal has a brain which is close to the human’s one). Providing a simulation based training system was our initial motivation for the EVE project, developed at CIMIT, Boston. An other motivation was the decision by the FDA to require all physicians who wish to treat carotid disease using catheter-based techniques, to train to proficiency before performing high-risk procedures in the cerebral circulation. For the needs of the project, we developed the beam approach detailed in section 1.3 that provides a physics-based model for the instruments (catheter and guide) as well as the collision response based on compliance in section 2.3.1 (see Cotin et al. (2005), Lenoir et al. (2006) and Duriez et al. (2006a)). These algorithms were implemented more recently in SOFA and new functionalities were added for the simulation of coil embolization.
Indeed, in the particular case of intracranial aneurysms, the treatment is delivered by filling the localized widening of the artery with a set of coils to prevent a rupture due to the weakened arterial wall. Considering the location of the lesion, close to the brain, and its very small size, the procedure requires a combination of planning and excellent technical skills. In the project named SIMPLE: SIMulation for PlAnning an Embolization, and, more recently, through the project SOFA-intermed, we have a long-term collaboration with the team MAGRIT (M.O Berger, E. Kerrien, and R.Anxionnat), in order to provide precise and fast simulation of coil embolization. For the project we have adapted and validated the coil model Dequidt et al. (2008) and the collision response (including self-collision) Dequidt et al. (2009). An approach to add blood flow simulation was proposed during the PhD of W.Yiyi in the team Wei et al. (2012). But one main issue is to obtain a good geometrical model of the brain vascular network from patient data. Apart from the problems of segmentation, the final reconstruction must be smooth (to avoid collision response artifacts), optimized for collision detection, and topologically coherent (no possible entry between two neighboring vessels)... A solution to this problem was recently proposed in Yureidini et al. (2012).

Obtaining precise and even predictive simulation would be interesting in order to help the choice of the first coil that is going to be delivered inside the aneurysm. Indeed, depending on the size and the shape of the aneurysm on the pre-operative 3D image, the physician makes the non-easy choice of the first coil. As illustrated in figure 4.1, there are different types of detachable coils that are pre-shaped in order to fill at maximum the aneurysm cavity. A bad choice of this first coil can seriously complicate the following of the intervention. To assist this choice, a simulation tool could be used during the planning of the procedure. We are moving gradually towards this goal and the current simulation is visually realistic and can provide predictive results in very simple case (like when the coil is obviously too small or too large). We still need more investigation for an assessment of the results.

In parallel, the new direction is taken by the project IDeaS financed by ANR (french research agency) and piloted by J. Dequidt. The goal is to reduce the number of
fluoroscopic images (Xrays) during the procedure and to allow for a 3D visualization of catheter, guidewire and coils. The aim of this project is to have a smart dialogue between per-operative images and simulation. We could use the simulation to provide the missing information, not contained on the images, and in return, we can use the image information to apply some corrections on the simulation.

### 4.1.2 Training and rehearsal of procedures for cardiac arrhythmia

Catheter-based procedures are also widely used in cardiology. In this work, we develop an interactive framework for rehearsal and training in the context of Cardiac Resynchronization Therapy (CRT). To this end, an interactive and real-time electrophysiology model of the heart is developed to fit patient-specific data. The solution that is currently developed relies on two main contributions. (i) We have used the latest GPU computing techniques to obtain an efficient implementation of cardiac electrophysiology. This computation is coupled to mechanical simulation of the heart and produces realistic motions (Talbot et al. (2012b)). (ii) We built a simulator prototype, based on catheter navigation algorithms, to reproduce a radio-frequency ablation which is a procedure currently performed for ventricular tachycardia (Talbot et al. (2012a)). This project is piloted by Hervé Delingette (INRIA Asclepios team, Sophia-Antipolis) with whom we collaborate on the PhD thesis work of Hugo Talbot.

Cardiac arrhythmia consists in an abnormal electrical activity of the heart muscle (myocardium). Such pathologies can occur upon changes in the heart structure following a coronary artery disease (heart attack) or as chronic consequences of hypertension, diabetes or cardiomyopathy. Several therapies may be pursued depending on the type of arrhythmia. For ventricular tachycardia, radiofrequency ablation of the ventricles may be performed, whereas cases of severe heart failure with bundle branch blocks may be treated through Cardiac Resynchronization Therapy (CRT).

These treatments are realized under complex minimally invasive procedure, realized under image control with dedicated catheters. They need to be performed by experienced and highly skilled cardiologists. Therefore, building training systems for performing this endovascular procedure would benefit both junior electrophysiologists, to train for those procedures, but also experienced electrophysiologists, to rehearse some complex procedures.

There exists a limited number of commercial endovascular simulators Simbionix (2012); Mentice (2012) but they are somewhat limited because they rely on pre-stored electrophysiology data but not on electromechanical modeling. In this work, we propose an innovative framework towards the interactive simulation of a cardiac electromechanical model. With the constant improvement of computational methods Smith et al. (2011), patient-specific cardiac modeling is now being considered as a promising support on the therapy and pathophysiological understanding. Physiological models can not only
reproduce the cardiac motion and electrophysiology signals but also predict in silico the impact of therapies Sermesant et al. (2012) thus serving as an additional guidance for cardiologists.

Our framework is mainly composed of two physiological models. First, we use the model of Mitchell et Schaeffer (2003) to describe the cardiac electrophysiology, because it has only two variables and only depends on few parameters while capturing the main properties of restitution of the action potential duration and conduction velocities. Second, we chose the Bestel-Clément-Sorine model to simulate the electromechanical coupling due to its compact nature and its consistency with the law of thermodynamics (balance of energy). The advantage in having a compact mechanical model, i.e. a model with few variables and parameters, lies in the ability to calibrate its parameters from a limited set of patient specific data as shown in Marchesseau et al. (2012) for the cardiac mechanics.

Our main contribution lies in adapting and optimizing these models and the solver algorithms in order to reach a fully interactive electromechanical simulation of the heart. In our team, we concentrate on obtaining real-time simulation of the electrophysiology model whereas the Asclepios team provides the interactive mechanical model, the coupling method and the patient specific data. The development of the simulator involves both teams.

The current results are illustrated in figure 4.2. From the 3D image of the patient’s heart, we build a volume mesh of the cardiac tissues and we extract the cardiac fibers. Thanks to an optimized GPU implementation of the Mitchell Schaeffer model, we decrease the computation time to reach real-time computation. Using a catheter model, we can virtually navigate inside the heart cavity and activate, in real-time, an electrical stimulation at the level of the tip of the catheter. We can also simulate an ablation and its effect on the electrophysiology. Even if it is not real-time yet, the coupling algorithm provides the deformations of the heart within a short period of time.

Figure 4.2: (Left) View of MRI images coupled with a 3D representation of the heart. (Middle) Simulation of the electrical stimulation launched at catheter tip. (Right) Cardiac geometry contracting with the current action potential and the rest geometry (black lines).
4.1.3 Brain shift during Deep Brain Stimulation (DBS) procedures

During DBS procedure, targeting a specific part of the brain is rendered difficult due to a series of brain shifts that take place during and after the intervention. In the project, we aim at introducing a biomechanical simulation of the intra and postoperative stages of the procedure in order to determine lead deformation and electrode migration due to brain shift. An other motivation, is to propose an approach that overcomes the limitations of current planning software that ignore brain shift.

This work is supported by the French National Agency for Research (ANR-Acoustic) and leaded by Pierre Jannin (INSERM). Alexandre Bilger is PhD candidate on this project. We have already obtained significant results for the modeling of the electrode migration after the procedure Bilger et al. (2011) and on a proposal of simulation based planning Bilger et al. (2012).

Deep Brain Stimulation (DBS) is a modern surgical treatment of brain disorders such as Parkinson’s disease or dystonia. This procedure consists in the placement of a micro-electrode in the subthalamic area, deep into the brain. The placement of the electrode is crucial to maximize outcomes and to prevent adverse effects. However, two major problems arise during this process: first, the planning stage does not account for the brain shift that takes place during surgery. Depending on the amplitude of the brain shift, the effective location for the electrode can be quite remote from the planned location (5 mm or more) Munckhof et al. (2010) requiring extensive mapping of the area to determine the final location. As a consequence, this stage of the procedure usually lasts several hours, while the patient is awake, and increases the likelihood of complications due to the number of insertions to reach the targeted area Deep-Brain Stimulation for Parkinson’s Disease Study Group (2001). The second problem, also linked to the amplitude of the brain shift, takes place several days or weeks after the surgery. As reported in Munckhof et al. (2010) a post-operative electrode displacement and deformation may appear as the brain returns to its initial position when the subdural air introduced during surgery is resorbed (see Fig. 4.3).

In this context, we aim at proposing a global approach that can be used to adjust the planned trajectory, to determine a potential post-operative electrode migration, and to propose alternative strategies to minimize its amplitude.

As a first step in this direction, our paper Bilger et al. (2011) introduces an original and unified approach to model the brain behavior during a DBS procedure. The focus of the paper is not set on a specific biomechanical model of the brain, but rather on a complete framework that is able to simulate intra-cranial fluid loss, subdural air invasion, brain shift and electrode migration and curvature. The work takes great care in defining the boundary conditions of the deformable model of the brain: the
CHAPTER 4. APPLICATIONS AND RELATED ACTIVITIES

Figure 4.3: Post-operative (left) and follow-up (right) CT scans. The post-operative scan illustrates the brain shift at the end of the procedure. The follow-up scan emphasizes the deformation of the electrode due to the inverse brain shift, leading to an upward migration of the electrode away from its initial location after the craniotomy (black cross). For large brain shifts, the electrode can move of up to 5 mm. Courtesy of Munckhof et al. (2010).

Mechanical interaction with cannula and electrode (that uses the method described in section 2.5), the contacts between the brain and the inner part of the skull and falx cerebri, the effect of the cerebro-spinal fluid loss and air invasion. Results of the simulation, illustrated in figure 4.4, exhibit qualitatively consistent results compared to the clinical studies reported in the literature. It shows that we are able to reproduce the phenomenon of electrode migration using simulation.

Figure 4.4: Screenshot showing the deflection of the right electrode after the cannula removal (left) and after CSF recovery (right).

We realize a second step towards the use of simulation for planning of DBS in a work detailed in Bilger et al. (2012). We propose an approach to overcome the limitations of current planning software that ignores brain shift. In particular, we consider the motion of vascular structures in order to prevent risks of dissecting a vessel during the procedure. Facing the difficulty to produce an exact brain shift prediction, we propose to build a brain shift aware risk map, which embeds the vascular motion. We use simulation to extrapolate on any patient the clinical studies providing statistics on the displacement of anatomical landmarks during the procedure. The risk map can be directly integrated into present planning software or into automatic path planning algorithms (developed by Caroline Essert from iCube-IGG team in Strasbourg) to have a prediction of the optimal electrode trajectories.
4.1. ONGOING RESEARCH PROJECTS

4.1.4 Middle ear surgery simulation and robotic assistance

Ossicular surgery requires a high dexterity manipulation of the anatomical structures due to the size of the components and risks of trauma. In this project, we aim at proposing a simulation software of middle ear surgery dedicated to residents. Moreover, in parallel, the partners of this project are developing a new type of ossicular surgery based on a tele-operated robot and would benefit from the use of a simulation for rehearsal or control of the robot.

In the context of this project, we are co-advising Guillaume Katzmitcheff, which is a PhD student paid by Collin (a company developing new solutions for middle ear surgery). The project is piloted by the INSERM team of the Beaujon hospital. The current results of this work focus on an accurate finite-element model of a human ossicular chain which is validated through comparison with experiments. This research work illustrate that simulation could have multiple interests (education, planning, robot control), but it necessitates to assess the quality of the models used for the anatomical structures.

The ossicles are the three smallest bones in human body. They are contained within the middle ear space and serve to transmit sounds from the air to the fluid-filled labyrinth (cochlea) that is the final sensor. Middle ear surgery is a common means of hearing rehabilitation in conductive losses and provides a significant benefit. Ossicular repair requires dexterity and delicacy in the manipulation of the involved anatomical structures, since each ossicle measures a few millimeters and weighs a few milligrams.

Surgical training is confronted to the necessity of allowing a junior surgeon to perform the gestures in order to enhance eye-hand coordination and proprioceptive experience in a realistic environment and at the same time to maintain a high level of security and quality for the patient. This training is partly insured by temporal bone dissection laboratories, but a temporal bone specimen can reproduce neither disease nor realistic environment. In this context, a training simulator could have very positive impact.

Moreover, this surgery which is particularly delicate, was chosen as an objective for the development of a robotic system Kazmitcheff et al. (2011). A simulator would be
also an interesting training tool for the use of the robot. This simulation, if sufficiently accurate and based on patient data, could be included in a planning tool of the robot in order to test new surgical strategies that are made possible with the robot. An other possible use is the control of the teleoperation by using additional constraints issued from the simulation (to avoid collision with anatomical structures).

Figure 4.6: (Left) Tetrahedral model of the middle ear components used in the simulation. (Right) Comparison of the peak-to-peak displacement frequency response of the finite element method (thick dashed line) and the experimental data

To reach these objectives, a finite-element model of a human ossicular chain is developed for a surgical middle ear (Figure 4.6) The model is composed of the 3 ossicles (malleus, incus and stapes), the tensor tympani tendon, the stapedial tendon and the tympanic membrane. As these structures have very different stiffness properties, it leads to a badly-conditioned system. We use the asynchronous preconditioner described in the section 1.5.2 to improve the convergence and obtain real-time computations. The mechanical parameters are tuned accordingly to published data and the model is real-time, if large time steps are used ($\approx 30$ ms). The model is compared to some measures available in the literature. We assess the accuracy of the model using two very different measures: one is the transfer function analysis (small deformation, high frequency) and the other evaluates the static force pressure (large displacement, low frequency). It shows good fitting with the published results of these measures. A complete study on these results will soon be submitted to publication.
4.1.5 Simulation of a pathologic pelvic system

The use of synthetic implants for deficient soft tissues reinforcement in the case of prolapse treatment is sometimes unsatisfactory because of the mechanical behavior of soft tissues that are involved. The challenge of the project is to be able to perform non-destructive in vivo characterization and a simulation of soft tissues personalized to each patient, to propose a better planning of surgical implant placement.

This project is piloted by Professor Michel Cosson from the THAIS team (Hospital of Lille & INSERM) and Professor Mathias Brieu (Mechanics lab in Lille - LML). We joint the project recently (one year ago) to facilitate the use of SOFA for the simulation of the tissue behavior, with new constitutive laws proposed by the LML team and to provide our help on multi-organs simulation. A first numerical challenge is the modeling and the simulation of multi-organ deformable structures extracted from patient image data (MRI). The second challenge is the use of inverse method to identify the mechanical properties of the tissues. In this context, we are co-advising two PhD candidates, Zhifan Jiang and Mouhamadou Diallo.

This project aims at developing a numerical tool that is able to analyze and reproduce, by simulation, the pelvic mobilities observed on a dynamic MRI. Thanks to comparison between simulation and real patient data, we would like to characterize the mechanical properties of soft tissues, pathologic or not. In other words, the approach must enable the identification of the mechanical properties of the patient’s tissues using inverse analysis.

The context of the project affords to consider on one hand the difficulties linked to the large transformations occurring in soft tissues, and, on the other hand, the problems of contact between organs. In vitro biomechanical characterization results were already obtained through destructive methods by the team of Matthias Brieu. Based on these measures, they have built a functional numerical model with very few parameters to allow for accurately studying this problem.

We joint the project to facilitate the use of SOFA in this context: SOFA has the main advantage to be open-source so it allows to develop any type of model, constitutive laws etc., in a free manner (which is not the case with commercial softwares that are often used for FEM problems). Moreover, the architecture and the algorithms of SOFA are particularly adapted for fast simulation of multi-organ simulation. In this project, real-time is not the ultimate goal but very fast simulation is particularly interesting: inverse methods often requires to launch many simulations.

Before reaching this result, it will be necessary to address scientific issues:

- Personified anatomical model: Nowadays, 3D MRI of patients are routinely performed before a procedure that involves the placement of a synthetic implant. However, these MRI are not very precise and the complete reconstruction of the anatomy from these data is not trivial. We need to fuse the geometrical informa-
tion of the reconstruction with a-priori known knowledge related to biomechanics: the connections between organs, the boundary conditions, etc...

- Pelvic system mechanics: some unknowns are presently remaining particularly concerning the description of ligamentous system involved for which there is no consensus in the literature. These structures are often hidden in the 3D MRI but play a great role in the final behavior of the tissues. Globally, the modeling of the boundary interface between organs is not trivial.

- Personified mechanical parameters: For each 3D data set, we have a dynamic MRI of the patient who is asked to cough. It allows to observe the mobilities of the pelvic system (currently, only on a 2D plane). We plan to develop inverse method to identify some mechanical properties of the tissues based on the correlation between simulation and dynamic MRI. The results of this type of method on a such complex system are difficult to anticipate. Additional measures could also be used (like elastography) to fuse and assess the results.

Figure 4.7: Automatic segmentation of the bladder (red contour), the vagina (pink contour) and the rectum (blue contour) on a dynamic MRI image.

Currently, two PhD candidates are working on this project. Zhifan Jiang started a PhD one year ago on Image analysis and correlation. One objective is to build the mesh required for finite element simulation with a minimum of manual interventions. Moreover, the other goal is to track the displacements of organs under stress, from the dynamic MRI images (see 4.7) and to be able to compare these displacements to the simulation, which is the needed information of inverse method.

Mouhamadou Diallo also started a PhD one year ago on the mechanical modeling of the pelvic system. The first part of the work is dedicated to the use of deformable models currently available in SOFA, especially the shell model presented in section 1.4. Then he will evaluate the precision (compared to theoretical results or to equivalent models in commercial softwares) and if necessary, to propose some numerical improvements. Finally, he will work on the implementation of advanced constitutive laws in the FEM framework of SOFA and will initiate the work on inverse method.
4.1.6 Percutaneous procedures for liver tumor ablation

Tumor ablation based on needles insertion is a difficult procedure, realized under 2D or 3D image guidance. One particular difficulty is the fact that the liver is constantly moving due to breathing and that the soft-tissues are deformed by the insertion of the needle. Consequently, the targeted tumors change location compared to the initial planning. In this project, we propose to use the simulation for anticipating these motions during planning and/or for adapting the trajectory during the procedure. Simulation could also help the use of a robot to execute this procedure.

Scientifically, we will build on the method presented in section 2.5 for simulating, in real-time, flexible needle insertion in soft-tissues. This project, that has started since september 2012, is realized and supported by the IHU of Strasbourg (university hospital institute, which includes our research team) which is focused on research for improving abdominal procedures. For this project, we work with the iCube-IGG team (an other team of IHU) and the Hospital of Strasbourg.

Needles are used in many forms of medical diagnosis and treatment, from tissue biopsies to placement of radioactive seeds for cancer treatment. With current advances in the field of medical imaging, it is now possible to perform such procedures under 2D or 3D image guidance, allowing the assessment of the needle placement. In addition, the development of new thin, flexible needles that can be steered through deformable tissues around tissue obstacles, permit to reach specified anatomical targets that would otherwise not be accessible under percutaneous procedures. One key difficulty remains: soft tissue motion, either due to breathing or deformation induced by the needle, changes the location of the initial target. Either when using image guidance, or robotic control of the needle insertion, this remains a major obstacle. In this project we propose to address this issue by developing an advanced path planning method which accounts for both tissue and needle deformations, avoids anatomical obstacles, and maximizes chances to reach the target. Such a method would have direct applications in pre-operative planning, per-operative guidance, and control for robotics. Our approach will combine advanced modeling of the needle, liver tissue deformation, tissue-tool interactions, and planning algorithms. By offering the ability to accurately plan percutaneous procedures while accounting for tissue and needle deformation, our approach has the potential to improve targeting accuracy for a wide range of procedures, including biopsy, and tumor ablation. These advances could also improve the outcome of existing procedures and enable needle-based procedures for conditions that currently require open surgery or systemic treatment.

In section 2.5, we have proposed a new, generic, method for simulating the insertion of thin and flexible medical tools into soft tissues. We obtained real-time frame rate while modeling the geometrical non-linearities of the tissue and needle deformations. Our
method can handle complex scenarios where needle steering, non-homogeneous tissues and interactions between different needles can be combined. It has been validated by comparing the simulation results to the experiments on fantsoms that were published in the literature. However, the approach must be improved to include the specificities of the applications:

- **Simulation of needle insertion in a deformable liver model.** We are currently developing a specific version of previous simulations (which were done using simple geometric shapes). This simulation uses an average liver model (in terms of its shape and material properties) and a needle models based on actual needles used in percutaneous procedures (see figure 4.8)

- **Validation of interaction model using silicone phantoms.** We plan to work on a series of experiments for assessing the accuracy of the interaction model between the needle and deformable materials. For this, we will use actual needles used in percutaneous procedures and silicone models with similar characteristics than a liver. From the comparison between simulations and experiments, we will adjust both models and parameters.

- **Trajectory correction** We will develop a new method, based on the mechanical modeling (hopefully validated by the measures), to adjust the orientation of the needle at the point of insertion in order to reduce the distance between the planned path and the actual path. This method will use an inverse approach that we will need to optimize to obtain real-time results. This last point is a key scientific challenge.

### 4.2 Technology development and transfer based on SOFA framework

In this section, we discuss the technology developments and transfer activities. All our developments are done on the open source SOFA platform that we develop and
which is used by many teams around the world. This development is a collective effort shared by three teams at INRIA, which is not common for researchers. Then, we show a more advanced example of a simulation in SOFA framework with cataract surgery. This realization, conducted as a team, is about to be transferred in order to be used in a very ambitious simulator of a new generation. Finally, we present InSimo, a startup that is dedicated to medical simulation, which uses the technologies developed in SOFA (both private and public code) and for which I am scientific advisor.

4.2.1 Development of the framework SOFA

SOFA\(^1\) is an Open Source framework primarily targeted at real-time simulation, with an emphasis on medical simulation. The development of SOFA started 7 years ago, from a common vision of several teams at INRIA and of the CIMIT simulation group (see Allard et al. (2007b)). In recent years, SOFA framework has begun to be renown worldwide as an open-source libraries of reference in the field of interactive (bio)-mechanical simulation. SOFA facilitates collaborations between specialists from various domains, by decomposing complex simulators into components designed independently, and providing a simple way of building simulation prototypes. Each component encapsulates one of the aspects of a simulation, such as the degrees of freedom, the forces and constraints, the differential equations, the main loop algorithms, the linear solvers, the collision detection algorithms or the interaction devices... The simulated objects can be represented using several models, each of them optimized for a different task such as the computation of internal forces, collision detection, haptics or visual display as described in Faure et al. (2012).

For the development of this platform, we had a constant and important support from INRIA through the technological development projects (ADT) and also the building of a national research initiative and network (SOFA-Intermed) on medical simulation. Working in parallel on the technological and scientific aspects has increases the variety of the functionalities and improved the performance and extensibility for attracting a growing number of users. SOFA was downloaded more than 135000 times. We have about 75 contributors (with 35 active in 2011). There is today more than 500 thousands lines of code (framework and plugins). Many research teams around the world are using SOFA (for the ones we know): TIMC (France), IRCAD (France), CSIRO (Australie), ICL (UK), Kitware (USA), Bandung Institute of Technology (Indonesia), KIST & KAIST (Korea) TU-Berlin (Germany), Technicka univerzita vo Zvolene (Slovakia), Universidad de las Ciencias Informaticas (Spain), University of Twente (Netherland), University of North Carolina (USA), Thomas Jefferson University (USA), Bradley University (USA), Siemens (USA and Germany)... Although sometimes difficult to identify, other research groups, working on different scope are publishing very good papers based on SOFA (for instance, recently, Tang et al. (2012) at SIGGRAPH, and Mansi et al. (2012) in Medical Image Analysis).

\(^1\)web-site of SOFA: www.sofa-framework.org
Participating to the SOFA-project from the quasi-beginning was a chance. As a researcher, this is not common to be involved in such an important development project. Activities around SOFA include designing and coding, teaching the concepts to every new student or engineer, following the new developments, answering the questions on the mailing list, reporting and solving bugs, doing demos, finding additional funding,... This requires a large commitment with a long term view. For instance, I have spent almost a year for nicely integrating in SOFA framework an other collision response than penalty method... But this effort is not lost!! The platform now allows to build new simulations very quickly and all simulation can benefit from constraint-based collision response (like the caduceus simulation shown in figure 4.9). Moreover, now, it reduces the needed time to integrate new results: for instance, for the asynchronous compliance matrix (method presented in section 2.3.3), no additional modification of the collision response process was necessary. For advanced project, like cataract surgery project (presented below), it allows to incorporate such a new result very quickly in a demo.

In the coming years, we will focus on improving the quality of SOFA and build a real consortium of users and contributors.

![Figure 4.9: (Left) default simulation in the Release Candidate Version 1 (RC1) of SOFA. (Right) Use of SOFA on an advanced prototype of cataract surgery simulation.](image)

### 4.2.2 Cataract simulation

Historically, there is a collaboration between the simulation team in Lille and the ophthalmology department of the hospital in Lille, directed by J.F Rouland (see Santerre et al. (2007)). In the recent years, we have developed a simulator for cataract surgery, illustrated in figure 4.9 (right), which is based on SOFA. We want to introduce a new generation of simulator where it is possible to make some mistakes while continuing the simulated procedure. Indeed, current simulator\(^2\) are similar to computer-game

scenarios: when an error is detected, the simulation is stopped (like a *game over*). From a pedagogic point of view, it is important not to panic after an error and to continue performing the procedure while compensating for the error. To achieve this goal, we want to rely on a simulation fully based on physics-based models.

Typically, related works offer a combination of simulation software that aims at reproducing some parts of the cataract surgery procedure and hardware devices that allow the surgeons for realistic interaction with the simulation. For instance, Weber et al. (2006) propose a mechanical model of the capsule in order to simulate the capsulorhexis\(^3\), whereas Choi et al. (2009) and Santerre et al. (2007) focus on the phacoemulsification\(^4\). Some other works have combined multiple cataract surgery steps in their simulators like Khalifa et al. (2006) or Agus et al. (2006). However, these steps are not fused in a common simulation: it allows to simulate the incision inside the eye, but when simulating the capsulorhexis or the phacoemulsification, the shape of the incision is standardized.

Using SOFA, we have also reproduced the main steps of the cataract surgery: Incision, capsulorhexis, phacoemulsification and lens implant deployment. For improved realism, our simulation of cataract surgery relies on geometrically non-linear finite element models (FEM) using tetrahedral elements for the lens, triangular membrane elements for the capsule and shell elements to simulate the lens implant deployment (see figure 4.10). The interactions can be simulated using precise collision response and advanced

\(^3\)Capsulorhexis is a technique used to remove the lens capsule.

\(^4\)During phacoemulsification, eye's internal lens is emulsified with an ultrasonic handpiece and aspirated from the eye.
constraints models described in chapter 2. This work has resulted in two publications at MICCAI (Comas et al. (2010b) and Courtecuisse et al. (2011a)).

Moreover, we have developed a working environment and devices that are as close as possible to real operating rooms. The environment consists in a microscope where stereo glasses have been fixed in order to render the virtual scene in 3D, a full-body mannequin where the eyes can support cataract surgery. Monitors are also available in the room to provide additional informations (like vital constants) or feedbacks. The tools are tracked in real-time and their 3D position is used to feed our real-time simulation. Figure 4.11 illustrates the various components that are part of the training system and how they are connected.

![Figure 4.11: Functional view and picture of the training system. The system relies on devices used in real procedures (microscope, pedals...), replicas of surgical tools that are tracked in real-time and on a simulation engine that renders a realistic and interactive simulation of cataract surgery.](image)

**MSCIS Simulation:** More recently, we have been involved in the HelpMeSee project ([www.helpmesee.org/programs.aspx](http://www.helpmesee.org/programs.aspx)) which objective is to restore the eyesight to 20 million children and adults by developing cataract surgery in the developing countries. Even if the foundation HelpMeSee raises enough money to build a massive treatment program for 20 million free surgeries, they will quickly run out of eye surgeons to perform this surgery.

To solve this problem, they launch an innovative training program that can create an eye surgeon very quickly (the objective is 6 months !). The training program is based on a surgical simulator that will enable our surgeons to perform thousands of surgeries virtually, before they ever operate on a single patient. The chosen surgery for cataract is slightly different in order to reduce the costs: phacoemulsification is replaced by MSCIS technique (Manual small incision cataract surgery): The eye-lens is removed in one block, and replaced by a rigid lens implant as illustrated in figure 4.12. This simulation was a very good example of application for the compliance
4.2. TECHNOLOGY DEVELOPMENT AND SOFA FRAMEWORK

Figure 4.12: Simulation of the extraction of the eye lens during MSICS (from Courtecuisse et al. (2011a))

computed based on the asynchronous preconditioner, which is presented in section 2.3.3 (heterogeneities of the deformable lens, deformable/deformable contacts).

After two research contracts between HelpMeSee and INRIA, and including new partners involved in the project (Moog, Sensegraphics), we have greatly improved the realism of the simulation and optimized the computation (as illustrated in figure 4.13). We have also included haptic rendering, which is very important for MSCIS, based on the algorithms described in section 3.2. The main challenge will be now to integrate all stages of the simulation while keeping very good performances.

Figure 4.13: (Left) Cutting of the conjonctiva, (Middle) Cutting a tunnel to introduce instruments and remove the lens (Right) Capsulorhexis

4.2.3 InSimo: a spin-off from Inria dedicated to medical simulation

InSimo is an innovative company in the field of software for medical simulation that will be created in the coming weeks. Its mission is to provide software tools constituting the core of the simulation engine for a new generation of medical simulators. Using the software package of InSimo, the simulators will be developed faster, will be more accurate, and will expand their range of applications to planning or per-operative guidance.

It is a spin-off from INRIA which is based on the technologies developed in our team.
and more generally within SOFA. InSimo will accelerate the transfer of our technologies to the medical simulation market which is asking for new solutions. The startup company has an original strategy: instead of selling directly simulators, its primary objective is to be the provider of a strategic component of all medical simulators, which is the physics-based engine.

The company is created by Jérémie Allard, a research scientist of our team, who has greatly contributed to the development of SOFA, and two former engineers of our team, Pierre-Jean Bensoussan and Juan Pablo de la Plata Alcalde. They will lead the new company. Stephane Cotin and I are also part of the founders. As scientific advisor, our role in this new company will be to orientate the technological and scientific choices from new research trends. We can also have a role of ”teachers”, sharing our expertise on simulation with the employees of inSimo.

InSimo can play a major role to increase the size of this market by providing solutions that are closer to the needs of the physicians. This company is the outcome of an ambitious project that we made in constructing the project Sofa. Now the challenge is to move from a technology that is an academic success to commercial products! InSimo needs to develop an economically viable company, and to prove the impact of its products for the quality of care for future patients.

More personally, I was proud of the experience with the company Didhaptic that sells the simulator Virteasy, based on an algorithm for dental simulation that I co-invented (presented on section 3.2.1). I am now very pleased that a company is created to bring our research work on the market of medical simulation. I think that helping the creation and the development of innovative companies from emerging technologies is also part of our job. For me, this is not the less interesting part!

### 4.3 Other research activities

#### 4.3.1 Scientific Animation

**Vice-leader of Shacra team:** I have participated to the creation project of the team Shacra piloted by Stéphane Cotin. Since september 2011, Stéphane is mostly based in Strasbourg, so I am in charge of a part of the administrative work and the scientific animation of the team in Lille.

**Responsible for the Colloquium Polaris** Researchers in computer science and automation of Lille form a dynamic and growing scientific community. There is three entity for research: the INRIA center of Lille-North Europe, the computer science lab (LIFL), and automation and signal processing lab (LAGIS) of the University of Lille. In order to federate the community, it was decided in 2009, to launch a colloquium, named POLARIS (http://www.colloquiumpolaris.fr/)
Max Dauchet, the former director of the INRIA center of Lille, asked me to take the responsibility of this colloquium. It consists in organizing a series of invited talks, in Lille, with renowned research scientists in computer science and automatic science. The choice of the invited speakers is decided in coordination with the scientific committee of the colloquium, composed of the heads of the three entities (Sophie Tison for LIFL, Philippe Vanheeghe for LAGIS and David Simplot-Ryl for INRIA). I also have the help of the communication department of INRIA and LIFL for the concrete organization (poster, emails, website...)

After two years and a half of existence, the colloquium is a success: the number of people who regularly assist to the invited talks is globally increasing (between 60 and 100 attendees each time). Moreover, people have asked for more regularity: this year we have a program with a talk per month except during summer.

**Ph.D. advisor and member of Ph.D committee**

I have co-advised the following PhD thesis:

- Guillaume Saupin, *Vers la simulation interactive réaliste de corps déformables*, PhD thesis, Université de Lille, 2008 (supervision ratio: 60% with Laurent Grisoni)

- Christophe Guébert, *Simulation temps réel de suture de modèles déformables*, PhD thesis, Université de Lille, 2010 (supervision ratio: 70% with Laurent Grisoni)

- Olivier Comas, *Real-time Soft Tissue Modelling on GPU for Medical Simulation*, PhD thesis, Université de Lille, 2010 (supervision ratio: 30% with Stéphane Cotin and CSRIO Australia)


I am currently advising the

I have participated to Ph.D. defenses of Nadjet Talbi (University of Evry, december 2008, examiner) and Mohamed Guiatni (University of Evry, May 2009, examiner).

**Journal / Conference and reviewing activities:** In the last three years, I have been reviewer for the following conference and journals:

- Proceedings of IEEE,
- Medical Image Analysis (MedIA), Elsevier.
- IEEE Transaction on Medical Imaging (TMI),
- IEEE Transactions on Biomedical Engineering (TBME-EMBS)
IEEE Transaction on Visualization and Computer Graphics (TVCG),
IEEE Transaction on Haptics (ToH),
Computer Methods in Applied Mechanics and Engineering (CMAME),
Computer Methods and Programs in Biomedicine (CMPB),
Progress in Biophysics and Molecular Biology journal,
Computer Animation and Virtual Worlds journal,
Computer and Graphics,
Eurographics Conference 2010 and 2012,
Eurohaptics 2012,
IEEE World Haptics Conference 2011 and 2013
International Conference on Control, Automation, Robotic and Vision (ICARCV 2012),
Joint Virtual Reality Conference (JVRC 2011)
IEEE Virtual Reality Conference (VR 2012),
IEEE Symposium on 3D User Interfaces (3DUI 2011)
Medical Image Computing and Computer Assisted Intervention (MICCAI 2011 and 2012),
International Symposium on Biomedical Simulation (SBMS 2010)

I have also been member of the following committees:

International Symposium on Biomedical Simulation (ISBMS 2010)
IEEE World Haptics Conference Conference 2013

I will organize the next edition of the symposium on Virtual Reality Interaction and Physical Simulation (VRIPHYS 2013) in Lille.

Finally, I was reviewer of 3 scientific projects of ANR (French research agency).

4.3.2 Contracts / Funding

European Initiatives: PASSPORT (PATient Specific Simulation and PreOperative Realistic Training for liver surgery), was a 3-year project (2008 - 2011) related to the objectives of the Virtual Physiological Human ICT-2007.5.3 objective. Indeed, PASSPORT’s aim was to develop patient-specific models of the liver which integrates anatomical, functional, mechanical, appearance, and biological modelling. Many results presented in this manuscript are the results of PASSPORT project, like the vascularized organ (section 1.5.3) and the volume-based collision response (section 2.4.2).
4.3. OTHER RESEARCH ACTIVITIES

**National Initiatives:** I have participated to several projects financed by ANR: VOR-TISS, REACTIVE, ACOUSTIC and IDeaS. The two last projects are still active. The main objective of ACOUSTIC is to develop an innovative strategy based on models for helping decision-making process during surgical planning in Deep Brain Stimulation (see section 4.1.3). IDeaS is a project targeted at per-operative guidance for interventional radiology procedures (see section 4.1.1).

**INRIA Initiatives:** I have also participated start projects financed by INRIA: The project named SIMPLE on simulation for planning of embolization with the team MAGRIT (see section 4.1.1), the ADT SOFA: Large Scale Initiatives to have financial credits for engineers on software development with the teams ASCLEPIOS and EVASION, the AEN SOFA-Intermed which was the research part of the SOFA-project that involved several teams of INRIA.

**Bilateral contract with private sector:** With the company Digital Trainers, I have signed a research contract based on our research on suture simulation (see section 2.5) The contract is including a 2 years license for the code and the research contract is dedicated to include this solution into their simulator and to improve the stability and robustness of the algorithms. More recently, I have participated to the project dedicated to MSCIS cataract surgery with the foundation HelpMeSee (see section 4.2.2).

4.3.3 Teaching and scientific diffusion

I was invited in 2011 to give a keynote lecture on medical simulation in the summer school on surgical robotics in Montpellier.

I am teaching the Finite Element Method at ICAM (Institut Catholique d’Arts et Métiers) which is an engineer school in Lille (about 64h per year). Moreover, I am also teaching a course on Interactive Simulation in the master of the University of Lille 1 called IVI (Image, vision and Interaction) (8 12h per year). Finally, each year, I have a 2h conference-course on Medical Simulation at the Ecole Centrale de Lille.

I have often participated to special events dedicated to diffusion of science to a large public: like the *Nuit des chercheurs* or *La fête de la science*... I love to share my passion for science and research, especially with children and young people!
CONCLUSION AND PERSPECTIVE

As discussed in chapter 1, an important part of our research is dedicated to the development of accurate models that remain compatible with real-time computation. Our main strategy is to propose some algorithms that allow for the fast execution of the Finite Element Method using implicit integration scheme with large time steps. Depending on the geometrical feature (curve, surface or volume) of the deformable objects, different models and optimizations are proposed. Thanks to the implementation of dedicated algorithms, the use of FEM models composed of several thousands of elements can be simulated in real-time even when badly conditioned.

The FEM models permit to increase the realism of training systems, and they also build a bridge toward the development of patient-specific preoperative planning as well as per-operative guidance. There is still work to do on the numerical aspects, for instance, the domain discretization of an interactive simulation is rarely addressed as a research topic; we often only try to simulate the greatest number of elements in real-time. Moreover, we should allow the use of more complex constitutive laws that are closer to the properties of the soft-tissues. Another limitation of the present solutions is the difficulty to use adequate parameters for these models, while taking into account the specificity of each patient.

In chapter 2, we focus on the problem of boundary conditions of the soft-tissues and on the modeling of the mechanical interactions with the medical devices. We justify the use of a time stepping approach, which is derived from existing works developed in the context of non-smooth mechanics, and on constraint-based solving process adapted to complementarity problems. In this context, we highlight the role of the compliance matrix and proposed several strategies to compute it in real-time with non-linear models. We also propose new ways of defining contact constraints, to be more independent from the model used for collision detection and to reduce the number of necessary constraints to model the contact. Finally, we introduce new models for the mechanical interaction with medical devices, based on complementarity constraints.

The proposed solutions greatly increase the possibilities of user inputs and the realism of interactive simulations. Nevertheless, we still do not have a pertinent modeling for a set of important interactions, like cutting, tearing or burning the tissues or for some boundary conditions, like the role of connective tissues between organs. Moreover, mechanical interactions between solids depend on both mechanical models of these solids and also on models and parameters used for these interactions (friction, puncturing force of the needle etc. ..). One important limitation of our work is that we still lack a methodology to validate these approaches, and to measure the influence of all parameters on the final results.

In chapter 3, we present new methods and algorithms for haptic rendering in the context of interactive medical simulation. Our method aims at accurately reproducing the haptic feedback of the mechanical interactions between surgical instrument and organs. The approach builds on the constraint-based process described in chapter 2, and proposes multi-rate
strategies to fulfill the strong real-time constraints or haptic feedback.

Our results show that the intermediate representation, called multi-rate compliant mechanisms is relevant. Moreover, we have made a significant step towards the objective of allowing a full multi-rate simulation with an independent time step and refresh rate for each deformable solid. Again, we need a methodology for validation: Even if the haptic rendering corresponds to the solution of the constraint problem computed in the simulation, we didn’t yet validate the whole method by confronting the forces transmitted by the device by a series of measure on real data.

In chapter 4, we demonstrate the wide spectrum of possible applications for medical simulation, through the description of a number of ongoing projects. All these projects are using the SOFA-framework and they benefit from our efforts for providing new numerical methods in SOFA that produce realistic simulations in real-time. We also highlight that the step for moving from training simulators towards the planning or the per-operative guidance using simulation, is in obtaining more predictive simulations, with a level of accuracy that is mastered.

The perspectives of this work are numerous. Among all, three challenges and a new application are highlighted in the following.

The first challenge is to better adapt the numerical methods to the needs of the simulation. Even if we made good progress in providing FEM models and constraint-based methods in a real-time context, the distribution of the computational effort is still far from optimal. For instance, the discretization of the meshes used in FEM is often static and is arbitrarily determined by the real-time constraints. Optimization strategies for positioning the degrees of freedom, for adjusting the sampling, or for dynamically adapting the models to the solicitations should allow to improve the precision without increasing the computation time. This process should also address the problem of precise cutting. In addition, a deeper work on discretization choices for modeling interactions should also be undertaken.

A second challenge is the improvement of the accuracy combined with a validation of the simulation results, in order to, ultimately, have some proofs that we can rely on a simulation as a predictive tool. First, we need to validate the implementation and the numerical methods: we should compare more carefully, on a representative database of tests, the results provided by SOFA with other non real-time Finite Element softwares and with well-known theoretical results. We started this type of study in Marchal et al. (2008) but this work has not been extended. Moreover, many small details in the anatomical modeling are often neglected: small ligaments, connective tissues, fat, abdominal wall, skin, nerves etc... We should not ignore these structures without a preliminary study of their mechanical influence. More generally, we should use highly detailed anatomical atlas to obtain reference simulations, that we could compare with the measures obtained in reality. Then, we could validate numerically the simplifications or the approximations that are made to achieve real-time computation and we would have a better understanding of the sources of error.

The third challenge is the parametrization of the models adapted to each individual patient. Simulations often contain many parameters. We should master their number and their final
influence on the results using sensitivity studies. We need to focus on the models that have less parameters and/or with parameters that can be obtained in the literature or by a measure that we could aquire in routine. For instance, we should explore the use of last elastographic devices that allows for nominal measures of tissue stiffness. However, all parameters will not always be accessible directly by a measure. An alternative is to rely on the deformations observed on the medical images to adapt the parameters of the models. For per-operative guidance, corrections of several parameters of the models will need to be done in real-time to be able to correct a discrepancy between the simulation results and the behaviors observed in reality. The next challenge is to investigate inverse methods and observers of non-linear systems to make them compatible with real-time!

For future work, we want to build bridges between our simulation tools and robotics. An example of application is the use of simulation for the control of a deformable robot inside the anatomy. Deformable robots are well-suited to surgical procedures: they can be passed through a natural orifice, they can be molded with a unique material compatible with medical imaging, they can be less dangerous when colliding with fragile anatomical structures... But, deformation induces a huge number of degrees of freedom. Consequently, there is a challenge in the control of these deformable robots. Moreover, in the particular case of a surgical procedure, the interactions with the anatomy could create some additional deformations that can not be ignored by the control. Recently, we have obtained preliminary results on the control of deformable robots (Duriez (2013)) by using a FEM model for the robot computed in an interactive simulation. This is a first result that it is far from the final application but is an encouraging step towards this goal. To succeed, in addition to the FEM model of the robot, we will again need to have a biomechanical model of patient that is both accurate and fast to compute.
Publication List

Book chapters


Journal papers


Conference papers


Patents


BIBLIOGRAPHY


BIBLIOGRAPHY


