

**Proceedings of the Fifty-Eighth
European Study Group
Mathematics with Industry**

Utrecht, The Netherlands, January 29 – February 2, 2007

Editors:

Rob H. Bisseling
Karma Dajani
Tammo Jan Dijkema
Johan van de Leur
Paul A. Zegeling

Preface

This book presents the scientific proceedings of the 58th European Study Group Mathematics with Industry, held at Utrecht University in the Netherlands, from January 29 to February 2, 2007. Locally in the Netherlands, this study group is also known as the *Studieweek Wiskunde en Industrie 2007* (SWI2007).

During the week of the study group, 79 participants tried to solve six problems with a high mathematical content posed to them by industry. Most (but not all) of the participants were mathematicians, mainly from the Netherlands and Belgium, but also some from England, France, Ireland, and Poland. As expected, many applied mathematicians participated, some of them regular participants in study groups for years, but there were also a significant number of mathematicians with a more theoretical background. All of them were motivated by the pleasure of solving important real-life problems, hoping perhaps to make an immediate contribution to society, while having a good time with their mathematical colleagues.

In these scientific proceedings, the participants themselves present the problems as they see them, possible solutions, and the results obtained, in a format aimed at a scientific audience. The papers have mainly been written in the six weeks after the week of the study group. The scientific proceedings are the second part of the complete proceedings. The first part is a separate booklet written in Dutch by science writer Bennie Mols, and intended for a general audience.

The six problems originated in widely different areas. Two academic hospitals (AMC Amsterdam and UMC Utrecht) posed questions on state-of-the-art medical devices. The AMC asked for a mathematical model of the workings of a mechanical heart pump that can be used to help a patient recover after heart failure. The UMC asked for speeding up the adjustment of a new high-resolution 7-tesla MRI scanner to each individual patient. The current time needed for such an adjustment would be several hours of CPU time, making adjustment impractical; better calculations with improved numerics or analytics should reduce this to a couple of minutes.

KLM and Innogrow posed optimization problems. KLM hopes to minimize the total number of days that cabin crew are on stand-by duty as reserves needed to replace rostered crew in case of illness or other disruptions. Currently a replacement can cause further disruptions and the participants of the study group were asked how to limit this domino effect. Innogrow constructs closed greenhouses for agricultural crop such as tomatos. These greenhouses have closed windows and underground storage for surplus heat (in summer) and surplus cold (in winter). The question is how to minimize the energy expenditures and maximize the yield.

ASML manufactures machines for the production of computer chips. It posed a sampling problem occurring in one of the stages of the production process, namely the exposure of the photoresist on the silicon wafer to light of varying intensity. The aim here is to replace the commonly used light-mask by a grid of small mirrors performing the same task.

ING asked for a fast method for computing the price of financial options, taking fluctuations in both interest rate and stock volatility into account.

All problems were solved at least partially, and the details of the proposed solutions can be read in the papers of these scientific proceedings. For the UMC problem, a solution was found which indeed brings the required computation time down to practical levels. ING got three solution approaches for the price of one: as an exception, the different solutions to the ING problem were not written up in one paper, but presented as three separate papers with a common introduction by the editor of the problem, see the papers at the end of these proceedings.

We thank our main sponsors NWO and STW for their generous financial support in the framework *Wiskunde Toegepast* (Mathematics Applied) which makes the study group possible. We also thank them for pledging to support the study group in the coming period 2008–2012. We thank the CWI in Amsterdam for their offer to finance and print these proceedings, also in the coming years. Furthermore, we thank Utrecht University for all the support we got in many different ways. We want to express our gratitude to Hans Gooszen for excellent secretarial support, and to Celia Nijenhuis and Jasper van Winden for their successful promotional work. We are indebted to the European Consortium for Mathematics in Industry (ECMI), the *Stichting voor Industriële en Toegepaste Wiskunde* ITW, and the Geometry and Quantum Theory cluster QCT, who sponsored several social events during the week; these were much appreciated by the participants.

Finally, the biggest thanks go to the participants of the study week, who displayed willingness to collaborate, a drive to succeed, curiosity, good humour, and plenty of talent in tackling challenging problems. Thanks to you all!

Utrecht, December 2007

Rob Bisseling
Karma Dajani
Tammo Jan Dijkema
Johan van de Leur
Paul Zegeling
(organizing committee SWI2007)

Contents

Modeling a heart pump	7
<i>Vincent Creigen, Luca Ferracina, Andriy Hlod, Simon van Mourik, Krischan Sjauw, Vivi Rottschäfer, Michel Vellekoop, Paul Zegeling</i>	
Cabin crew rostering at KLM: optimization of reserves	27
<i>Marco Bijvank, Jarek Byrka, Peter van Heijster, Alexander Gnedin, Tomasz Olejniczak, Tomasz Świst, Joanna Zyprych, Rob Bisseling, Jeroen Mulder, Marc Paelinck, Heidi de Ridder</i>	
A sampling problem from lithography for chip layout	45
<i>Eric Cator, Tammo Jan Dijkema, Michiel Hochstenbach, Wouter Mulckhuysse, Mark Peletier, Georg Prokert, Wemke van der Weij, Daniël Worm</i>	
Optimizing a closed greenhouse	55
<i>Jaap Molenaar, Onno Bokhove, Lou Ramaekers, Johan van de Leur, Nebojša Gvozdenović, Taoufik Bakri, Claude Archer, Colin Reeves</i>	
Understanding the electromagnetic field in an MRI scanner	69
<i>Jan Bouwe van den Berg, Nico van den Berg, Bob van den Bergen, Alex Boer, Fokko van de Bult, Sander Dahmen, Katrijn Frederix, Yves van Gennip, Joost Hulshof, Hil Meijer, Peter in 't Panhuis, Chris Stolk, Rogier Swierstra, Marco Veneroni, Erwin Vondenhoff</i>	
The ING problem: a problem from the financial industry	91
<i>Cornelis W. Oosterlee</i>	
Three approaches to extend the Heston model	93
<i>Michael Muskulus</i>	
A semi closed-form analytic pricing formula for call options in a hybrid Heston–Hull–White model	101
<i>Karel in 't Hout, Joris Bierkens, Antoine P.C. van der Ploeg, Jos in 't Panhuis</i>	
Characteristic function of the hybrid Heston–Hull–White model	107
<i>Fang Fang, Bas Janssens</i>	

Modeling a heart pump

Vincent Creigen *Luca Ferracina*^{*} *Andriy Hlod*[†] *Simon van Mourik*[‡]
Krischan Sjauw[§] *Vivi Rottschäfer*[¶] *Michel Vellekoop*[‡] ^{||} *Paul Zegeling*^{**}

Abstract

In patients with acute heart failure, the heart can be assisted by the insertion of a mechanical device which takes over part of the heart's work load by pumping blood from the left ventricle, one of the heart chambers, into the aorta. In this project, we formulate a model that describes the effect of such a device on the cardiovascular dynamics. We show that data for the pressure–volume relationship within a heart chamber that have been obtained experimentally can be reproduced quite accurately by our model. Moreover, such experimental data can help in calibrating unknown parameters that specify the characteristics of the pump. A key parameter turned out to be the extra friction that is encountered by the blood flowing through the heart pump.

Key words: cardiovascular system, rotary heart pump.

1 Introduction

Nowadays it is possible to use mechanical devices to assist the pumping action of the heart in patients with cardiovascular problems. A particular class of these, left ventricular assist devices, move blood from the left ventricle (one of the heart chambers) into the aorta, to take over part of the heart's work load. Developments in the emerging field of cardiovascular medicine have led to the availability of a wide range of instruments, from temporary assist devices to devices for long-term support and from left or right ventricular assist devices to biventricular assist devices and even total artificial hearts. This report will focus on the effects of a left ventricular assist device called the Impella, manufactured by Abiomed Europe (GmbH, Aachen, Germany). Two versions are currently being used by the Academic Medical Centre (AMC) in Amsterdam: the Impella 2.5 and the Impella 5.0, which are able to produce a flow of 2.5 and 5.0 liters per minute, respectively. The problem considered here has been formulated by cardiologists and cardiothoracic surgeons from the AMC.

The insertion of a mechanical device in the cardiovascular system obviously influences the dynamics of the blood flow through the arterial system. The contraction and relaxation of the heart muscles in the heart chamber causes two valves, called the mitral and aortic valve, to open and close due to pressure differences. When modeling the cardiac cycle, one can distinguish four phases.

^{*}CWI, Amsterdam

[†]Technische Universiteit Eindhoven

[‡]Technische Universiteit Twente

[§]AMC, Amsterdam

[¶]Universiteit Leiden

^{||}corresponding author, M.H.Vellekoop@ewi.utwente.nl

^{**}Universiteit Utrecht

*Thanks to other participants who helped during the week.

1. Isovolumic contraction phase, when the mitral and aortic valves are closed. Pressure is being built up in this phase, until the left ventricular pressure rises sufficiently above the aortic pressure to open the aortic valve.
2. Ejection phase, when the mitral valve is closed and the aortic valve is open. Blood flows out of the chamber into the aorta.
3. Isovolumic relaxation phase, when the mitral and aortic valves are closed. Pressure in the chamber decreases until it is so low that the mitral valve opens.
4. Filling phase, when the mitral valve is open and the aortic valve is closed. Blood flows into the chamber, and the cycle then repeats itself.

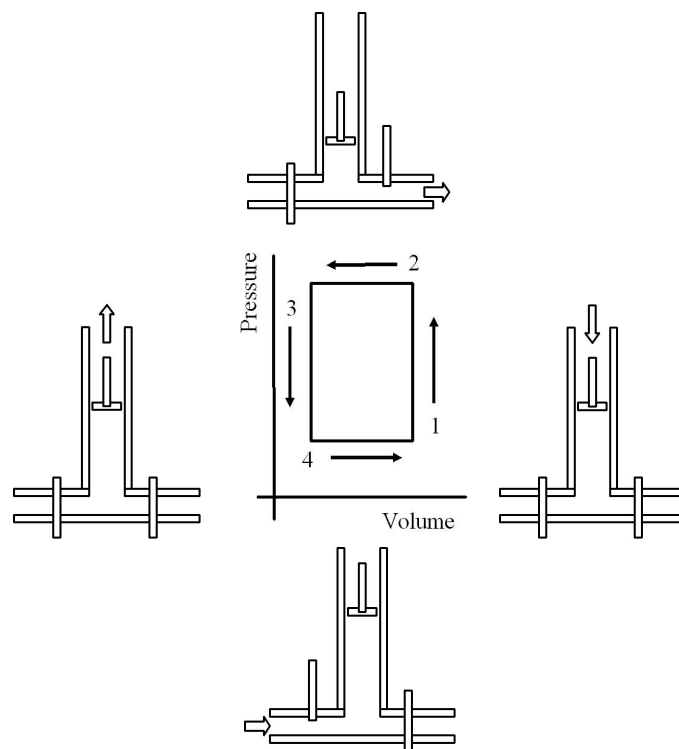


Table 1: The cardiac cycle.

The first two phases are known as diastole and the last two phases are known as systole.

In patients, the influence of a mechanical device which assists the heart during this process is difficult to quantify, since only a limited number of direct measurements can be performed on the cardiovascular system with and without the blood pump. Typically, only the pressure and volume of the blood in a heart chamber can be measured. The questions whether the pump really takes over a substantial part of the task of the heart and whether it reduces the amount of work the heart has to perform cannot be answered directly from such measurements. The amount of flow produced by the heart, the cardiac output, is 5 to 7 liters per minute in healthy people. It can be indirectly determined

by means of a method called thermodilution, which represents the gold standard in clinical practice. However, it is uncertain if this method is reliable in the presence of a continuous flow pump, since the normal physiology of the heart is altered, changing the premisses on which the method is based. Moreover, it is not possible to determine directly the individual contributions of the heart and the pump to their combined cardiac output, since detailed information about certain key parameters in the system is unavailable. Information about the cardiac output and the contribution of the pump are important to assess, and some form of mathematical modeling is therefore indispensable to obtain information about these quantities.

2 Problem Formulation

In this project, the study group was asked to develop a mathematical model for the influence of the Impella on the cardiovascular dynamics. More specifically, the group was asked to focus on the estimation of the blood flow through the pump during the cardiac cycle, since this is an important monitoring variable that can be used to establish how much the work load of the heart can be relieved. To validate the model and to make it possible to calibrate unknown parameters, a PV (Pressure–Volume) loop of a patient with and without the blood pump and an exact specification of the pump had been made available.

The distinguishing characteristic for this problem is the specific placement of the pump across the aortic valve. In order to obtain a realistic model, it is important that the pumping action of the heart itself, the dynamics of the pump, and the rest of the body, i.e. the arterial system, are modeled with sufficient accuracy. Moreover, the problem formulation suggests that the model should not be too complicated or detailed, since we would like to be able to explain the change in qualitative behaviour of the system in direct terms.

The structure of the paper is as follows. In the next section, the heart and the arterial system are modeled, using an analogy with electrical circuits. In section 4, some details concerning the operation of the pump are given. Section 5 combines all these elements in a full cardiovascular model and section 6 provides numerical results. We end by formulating our conclusions and recommendations for further research in the last section.

3 Modeling the Heart and Arterial Systems

The cardiovascular system can be described in terms of quite complex fluid dynamics. Different models have been developed to investigate it and various tools, ranging from rather simple to very sophisticated numerical techniques, have been employed; see for example [9, 11, 16] and references therein for an overview of computational methods in cardiovascular fluid dynamics.

A computationally cheap option to obtain information about the overall behaviour of the cardiovascular system is provided by so-called lumped parameter models [3, 13, 14, 15]. In these models, critical parameters are defined by taking averages over many different subsystems without distinguishing these subsystems themselves in too much detail. Such models proved to be very useful as a starting point for the investigation of arterial blood pressure and blood flow.

There is a close correspondence between the cardiovascular system and electrical circuits, which we intend to exploit here. A description of this correspondence will therefore be given in the following subsections.

Cardiovascular	Electrical
blood volume	electrical charge
flow rate (F)	current (I)
pressure (P)	potential (V)

Table 2: The analogy between the cardiovascular and electrical systems.

3.1 Mapping Cardiovascular Elements to Electrical Elements

In the analogy that we will use, electrical charge represents blood volume, while potential (difference) and currents correspond to pressure (difference) and flow rates. A particular vessel, or group of vessels, can be described by an appropriate combination of resistors, capacitors, and inductors. Blood vessels' resistance, depending on the blood viscosity and the vessel diameter, is modeled by resistors. The ability to accumulate and release blood due to elastic deformations, the so-called vessel compliance, is modeled by capacitors. The blood inertia is introduced using coils, and finally heart valves (forcing unidirectional flow) are modeled by diodes. In the following, we will explain in more detail how the different properties and parts of the cardiovascular system can be modeled by electrical components.

Vessel Resistance and Electrical Resistance

Blood flowing from wider arteries into smaller arterioles encounters a certain resistance. This resistance can be modeled as follows. Consider an ideal segment of a cylindrical vessel. The pressure difference between its two ends and the flow through the vessel depend on each other. Although this dependence will in general be nonlinear, for a laminar flow (which is the type of flow we are interested in) it can be accurately approximated by a linear relation. If we indicate by R_c the proportionality constant between the pressure difference P and the flow F then we can write

$$R_c = \frac{P}{F}. \quad (1)$$

Similarly, a resistor is an electronic component that resists an electric current by producing a potential difference between its end points. In accordance with Ohm's law, the electrical resistance R_e is equal to the potential difference V across the resistor divided by the current I through the resistor:

$$R_e = \frac{V}{I}. \quad (2)$$

Vessel Compliance and Capacitance

The walls of blood vessels are surrounded by muscles that can change the volume and pressure in the vessel. Consider the blood flow into such an elastic (compliant) vessel. We denote the flow into the vessel by F_i and the flow out of the vessel by F_o . Then the difference $F = F_i - F_o$ which corresponds to the rate of change of blood volume in the vessel is related to a change of pressure P inside the vessel. Assuming a linear relation, we have that

$$F = C_c \frac{dP}{dt}, \quad (3)$$

where C_c is a constant related to the compliance of the vessel.

The analogy with a capacitor is immediate. A capacitor is an electrical device that can store energy between a pair of closely-spaced conductors, so-called 'plates'. When a potential difference is applied to the capacitor, electrical charges of equal magnitude but opposite polarity build up on each plate. This process causes an electrical field to develop between the plates of the capacitor which gives rise to a growing potential difference across the plates. This potential difference V is directly proportional to the amount of separated charge Q (i.e. $Q = C_e V$). Since the current I through the capacitor is the rate at which the charge Q is forced onto the capacitor (i.e. $I = dQ/dt$), this can be expressed mathematically as:

$$I = C_e \frac{dV}{dt}, \quad (4)$$

where the constant C_e is the electrical capacitance of the capacitor.

Blood Inertia and Inductance

Since blood is inert, it follows that when a pressure difference is applied between the two ends of a long vessel that is filled with blood, the mass of the blood resists the tendency to move due to the pressure difference. Once more assuming a linear relation between the change of the blood flow (dF/dt) and the pressure difference P we can write

$$P = L_c \frac{dF}{dt}. \quad (5)$$

Note that this is the hydraulic equivalent of Newton's law, which relates forces to acceleration.

The inertia of blood can be modeled by a coil (also known as an 'inductor'), since the current in a coil cannot change instantaneously. This effect causes the relationship between the potential difference V across a coil with inductance L_e and the current I passing through it, which can be modeled by the differential equation:

$$V = L_e \frac{dI}{dt}. \quad (6)$$

Valves and Diodes

An ideal valve forces the blood to flow in only one direction. More specifically, it always stops the flow in one direction while it allows the blood to flow in the other direction, opposing only a small resistance (R_c) to the flow, as soon as the pressure difference is higher than a certain critical pressure P^* which is often taken to be zero. For this reason it is common use to model the action of a valve as follows:

$$F = \begin{cases} 0 & \text{if } P < P^* \\ P/R_c & \text{if } P \geq P^*. \end{cases} \quad (7)$$

The electrical analogue of a valve is a diode. In electronic circuits, a diode is a component that allows an electric current I to flow in one direction, but blocks it in the opposite direction. There are different models in the literature for diodes; we will use the idealized relationship corresponding to (7):

$$I = \begin{cases} 0 & \text{if } V < V^* \\ V/R_e & \text{if } V \geq V^*. \end{cases} \quad (8)$$

3.2 Other Relationships

One advantage of modeling the cardiovascular system by an electrical circuit is that Kirchhoff's laws for currents and potential differences can be applied:

- The sum of currents entering any junction is equal to the sum of currents leaving that junction (conservation of blood mass).
- The sum of all the voltages around a loop is equal to zero (pressure is a potential difference).

All these elements, or their nonlinear extensions, are used in different forms in the models for the heart and its environment, the arterial system. We summarize all the relationships in table 3.

In the following subsection, we discuss the relatively simple models for the cardiovascular system that are known as Windkessel models.

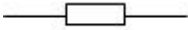


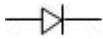
$P = FR_c$	vessel resistance	elec. resistance 	$V = IR_e$
$C_c \frac{dP}{dt} = F$	vessel compliance	elec. capacitance 	$C_e \frac{dV}{dt} = I$
$L_c \frac{dF}{dt} = P$	blood inertia	magnetic inductance 	$L_e \frac{dI}{dt} = V$
$F = \begin{cases} 0 & \text{if } P < 0 \\ P/R_c & \text{if } P \geq 0 \end{cases}$	valve	diode 	$I = \begin{cases} 0 & \text{if } V < 0 \\ V/R_e & \text{if } V \geq 0 \end{cases}$

Table 3: Analogy between electrical and cardiovascular behaviour.

3.3 Description of the Windkessel Model and its Use

The Windkessel model consists of ordinary differential equations that relate the dynamics of aortic pressure and blood flow to various parameters such as arterial compliance, resistance to blood flow and the inertia of blood. We discuss three forms of the model. In the different forms, the complexity of the model is increased by introducing extra components, each representing a characteristic of the cardiovascular system. Thus, closed-form solutions for the aortic pressure and the flow rate become increasingly difficult to obtain.

2-Module Windkessel Model

The first Windkessel model was put forward by Stephen Hales in 1733 [7]. He assumed that the arteries operate like a chamber in an old-fashioned hand-pumped fire engine (in German *Windkessel* pump) which smoothes the water pulses into a continuous flow. By conducting blood pressure experiments on various animals he was able to perform the first direct measurement of arterial blood pressure.

Hale's analogy between the cardiovascular system and the water pump was enhanced by the German physiologist Otto Frank [5]. His 2-module Windkessel model has since been applied in studies including chick embryos [17] and rats [8]. Figure 1 shows the 2-module Windkessel model

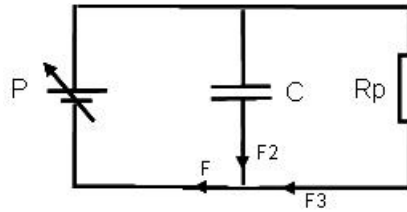


Figure 1: The 2-Module Windkessel model.

consisting of an electrical circuit with a capacitor C corresponding to the arterial compliance and a resistor R corresponding to the resistance to blood as it passes from the aorta to the narrower arterioles. This is referred to as the peripheral resistance. As explained in more detail in section 3.1, P and F represent the aortic pressure and the blood flow rate in the aorta, respectively, and both are functions of time, t . A differential equation in terms of P and F can be obtained using the equations given in section 3.2. These equations lead to

$$F = F_2 + F_3, \quad (9)$$

$$P = F_3 R, \quad (10)$$

$$C \frac{dP}{dt} = F_2. \quad (11)$$

Using Kirchoff's law for currents, we can eliminate the currents F_2 and F_3 from this equation, leading to:

$$F = \frac{P}{R} + C \frac{dP}{dt}.$$

This equation can be solved if we consider just the diastole period of the heartbeat in which the heart muscles relax, because during this period the left ventricle is expanding and $F = 0$. We then find

$$P = P(t_d) e^{-\frac{(t-t_d)}{RC}}. \quad (12)$$

Here it has been assumed that $P(t_d)$ is the blood pressure in the aorta at the starting time t_d of the diastole.

3-Module Windkessel Model

An extension of the 2-Module Windkessel model, the 3-Module Windkessel model, was formulated by the Swiss physiologist Ph. Broemser together with O. Franke and it was published in an article in

1930 [2]. This model, which is also known as the Broemser model, introduces an extra resistor R_a which represents the resistance encountered by blood as it enters the aortic valve. The corresponding electrical circuit is shown in figure 2.

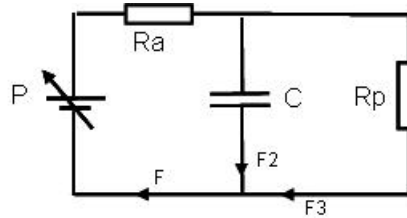


Figure 2: The 3-Module Windkessel model for the systemic circulation.

Adopting the same approach as before we obtain

$$\left(1 + \frac{R_a}{R_p}\right)F + R_a C_c \frac{dF}{dt} = \frac{P}{R_p} + C_c \frac{dP}{dt} \quad (13)$$

and during diastole, when F and its time derivative are zero, we may again simplify (13). This leads to the same expression for the aortic pressure during diastole as in the 2-Module Windkessel model.

4-Module Windkessel Model

The 4-Module version of the model was developed for the study of the systematic circulation in chick embryos [17] and pulmonary circulation (i.e. the circulation relating to the lungs) in cats [10] and dogs [6]. This model extends the 3-Module version by adding a coil L_c to represent the inertia of the blood, see figure 3. The approach used previously and some tedious but trivial algebra then

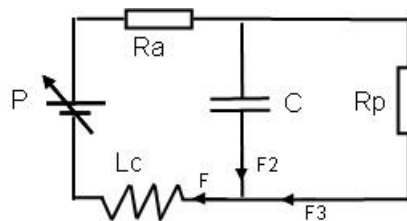


Figure 3: The 4-Module Windkessel model.

results in the following equation:

$$\left(1 + \frac{R_a}{R_p}\right)F + \left(R_a C_c + \frac{L_c}{R_p}\right) \frac{dF}{dt} + L_c C_c \frac{d^2 F}{dt^2} = \frac{P}{R_p} + C_c \frac{dP}{dt}.$$

In section 5, the elementary Windkessel models will be expanded to include the pumping action of both the heart and the mechanical pump. But first we will give a description of the pump itself and its most important operating characteristics.

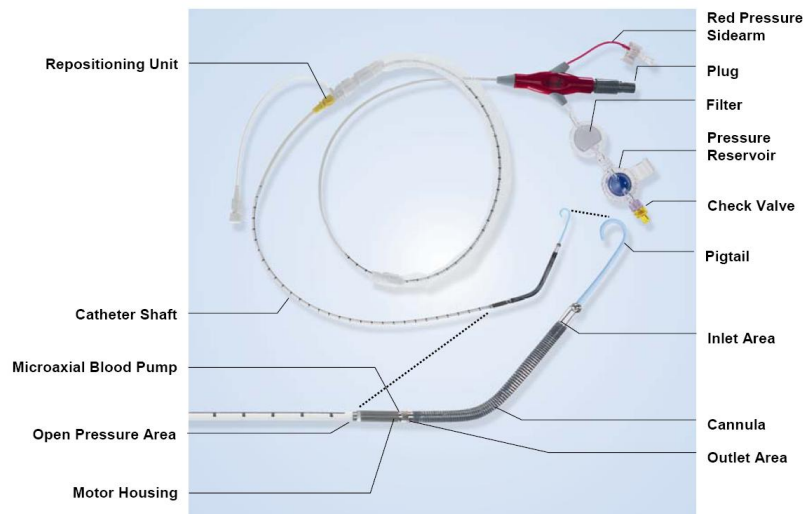


Figure 4: The Impella 2.5 LP device.

4 Description and Specification of the Pump

In this section, we give some specifications which will turn out to be relevant for the mathematical model of the blood pump. Part of the information about the heart pump was obtained from the instruction manual [1] and some was provided by Krischan Sjauw (AMC).

As described earlier, the Impella Recover LP 2.5 is a catheter mounted micro-axial rotary blood-pump, designed for short-term mechanical circulatory support. It is positioned across the aortic valve into the left ventricle, with its inlet in the left ventricle and its outlet in the aorta; see figure 5. The driving console of the pump allows management of the pump speed by 9 gradations and it displays the pressure difference between inflow and outflow, which gives an indication of the pump's position. Expelling blood from the left ventricle into the ascending aorta, the Impella is able to provide a flow of up to 2.5 litres per minute at its maximal rotation speed of 51000 rpm. To prevent aspirated blood from entering the motor, a purge fluid is delivered through the catheter to the motor housing by an infusion pump (see figure 6). Table 4 gives some further specifications.

The flow created by the pump depends mainly on the pressure difference between the outlet in the aorta and the inlet in the ventricle, and on the speed of the rotor. The flow decreases if the pressure difference increases or the rotor speed decreases. Figure 7 shows these dependencies, obtained experimentally, between the flow through the pump and the pressure difference for different pump speeds varying from the maximum possible 51000 rpm to 25000 rpm.

To describe the action of the rotary pump, we need to expand the Windkessel models discussed above to model the valves and the heart chamber, and the dynamics which lead to the cardiac cycle. An example of this is given in [4], where the left ventricle is described as a time varying capacitor with an elasticity function $E(t)$ for the heart, thereby providing a model for the heart capacitance which is time-varying. This time-varying function is calibrated to the end-systolic pressure and volume values and the end-diastolic pressure and volume values.

The pump itself is modeled as a bypass around the diode that represents the aortic valve with

Parameter	Value
Speed range	0 to 51000 rpm
Flow-Maximum	2.3±0.3 l/min
Catheter diameter	max. 4.2 mm (nom. 4.0 mm)
Length of invasive portion (w/o catheter)	130±3 mm
Voltage	max. 18 V
Power consumption	less than 0.99 A
Maximum duration of use	5 days

Table 4: Pump parameters.

resistance and inertia at both ends of the pump. We remark that this means that even when the aortic valve is closed, the pump can still pump blood from the left ventricle into the aorta and there may even occur some backflow: blood flowing back from the aorta into the left ventricle due to adverse pressure differences.

The rotator speed in the pump was assumed to be constant, since this was the explicit design objective of the pump considered here.

5 The Full Model

Our model for the environment of the heart and blood pump is based on a paper by Ursino [12], in which a nonlinear lumped parameter model of the cardiovascular system is proposed. The first order differential equations in this model describe pressures, volumes, and flows in the lumped subsystems.

These subsystems are the pulmonary arteries, pulmonary peripheral circulation, pulmonary veins, systemic arteries, peripheral systemic circulation, extrasplanchnic venous circulation, the left and right atrium of the heart, and the left and right ventricle of the heart. This means a distinction is made between veins, which carry blood to the heart, and arteries, which take blood from the heart to the organs, and between the pulmonary system, which corresponds to the lungs, the splanchnic system, which corresponds to abdominal internal organs, the peripheral system, corresponding to the outer part of the body, and the extrasplanchnic system, which corresponds to other organs.

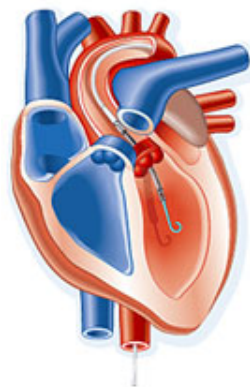


Figure 5: Placement of the pump between the aortic valves.

The Ursino model offers the possibility of parameter adjustments based on physical specifications for individual patients, whereas in a Windkessel model more parameters and state variables are modeled implicitly, and are therefore not adjustable in a straightforward way. Moreover, in the expanded model, blood flow in or out of the right and left ventricles is only possible if the corresponding valves are open. A valve opens one way, i.e. it is opened or closed, depending on the sign of the pressure difference over the valve.

In the model, a distinction is made between the capacitance, inertia, and other characteristics in the different components of the system. For example: the capacitance and resistance for blood flow obviously depend on the width of the arteries and veins, which may vary considerably within the human body. Different compartments of the system are denoted by different subscripts to make the model equations easier to read, see table 5. The electrical circuit corresponding to the model is given in figure 8. Notice that in this figure the time-varying pressures generated by the heart muscles in the left and right ventricle are indicated by capacitors with arrows through them, while we have also used the standard symbol for electrical earth to indicate a point of zero voltage, which corresponds to a reference pressure based on which all other pressure differences are stated.

The model for pressures, volumes and flows then becomes as follows. Conservation of mass and balance of forces in the different compartments lead to

$$\begin{aligned}\frac{dP_{pa}}{dt} &= \frac{1}{C_{pa}}(F_{o,r} - F_{pa}) \\ \frac{dF_{pa}}{dt} &= \frac{1}{L_{pa}}(P_{pa} - P_{pp} - R_{pa}F_{pa})\end{aligned}\quad (14)$$

$$\begin{aligned}\frac{dP_{pp}}{dt} &= \frac{1}{C_{pp}}(F_{pa} - \frac{P_{pp} - P_{pv}}{R_{pp}}) \\ \frac{dP_{pv}}{dt} &= \frac{1}{C_{pv}}(\frac{P_{pp} - P_{pv}}{R_{pp}} - \frac{P_{pv} - P_{la}}{R_{pv}})\end{aligned}\quad (15)$$

for the pulmonary arteries in the upper cycle in figure 8, while it leads to

$$\begin{aligned}\frac{dP_{sa}}{dt} &= \frac{1}{C_{sa}}(F_{o,l} - F_{sa}) \\ \frac{dF_{sa}}{dt} &= \frac{1}{L_{sa}}(P_{sa} - P_{sp} - R_{sa}F_{sa}) \\ \frac{dP_{sp}}{dt} &= \frac{1}{C_{sp} + C_{ep}}(F_{sa} - \frac{P_{sp} - P_{sv}}{R_{sp}} - \frac{P_{sp} - P_{ev}}{R_{ep}}) \\ \frac{dP_{ev}}{dt} &= \frac{1}{C_{ev}}(\frac{P_{sp} - P_{ev}}{R_{ep}} - \frac{P_{ev} - P_{ra}}{R_{ev}})\end{aligned}\quad (16)$$

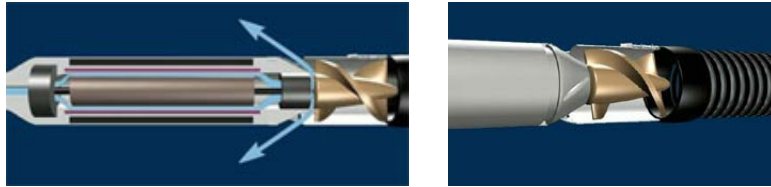


Figure 6: Left: Purge fluid preventing blood from entering motor housing. Right: Rotor positioned above the motor housing.

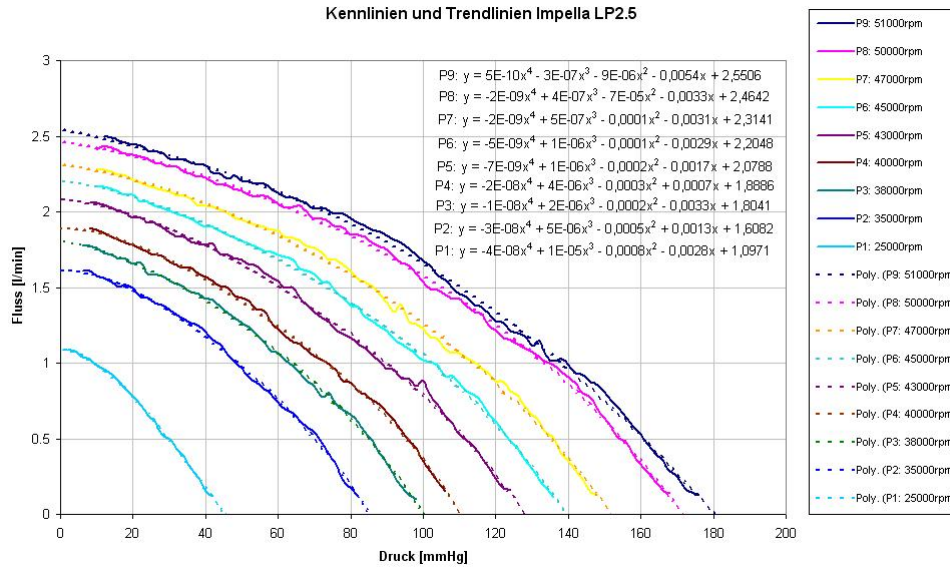


Figure 7: Dependencies between the flow through the Impella LP 2.5 pump and the pressure differences for different speeds of the motor.

for the lower cycle in that figure. Finally, for the left and right atrium

$$\frac{dP_{la}}{dt} = \frac{1}{C_{la}} \left(\frac{P_{pv} - P_{la}}{R_{pv}} - F_{i,l} \right) \quad (17)$$

$$\frac{dP_{ra}}{dt} = \frac{1}{C_{ra}} \left(\frac{P_{sv} - P_{ra}}{R_{sv}} + \frac{P_{ev} - P_{ra}}{R_{ep}} - F_{i,r} \right).$$

Here, $F_{i,l}$ and $F_{o,l}$ are the flow into and out of the left ventricle (in ml/s), and $F_{i,r}$ and $F_{o,r}$ are the flow into and out of the right ventricle. Assuming a known and constant total blood volume V_0 we can express the last remaining pressure, the splanchnic venous pressure P_{sv} , in terms of all the other

pa	pulmonary arteries	pp	pulmonary peripheral
pv	pulmonary veins	sa	systemic arteries
sp	systemic peripheral	ev	extrasplanchnic venous
sv	splanchnic venous	ep	extrasplanchnic peripheral
ra	right atrium	la	left atrium
rv	right ventricle	lv	left ventricle
i	in	o	out
l	left	r	right

Table 5: Subscripts of the variables.

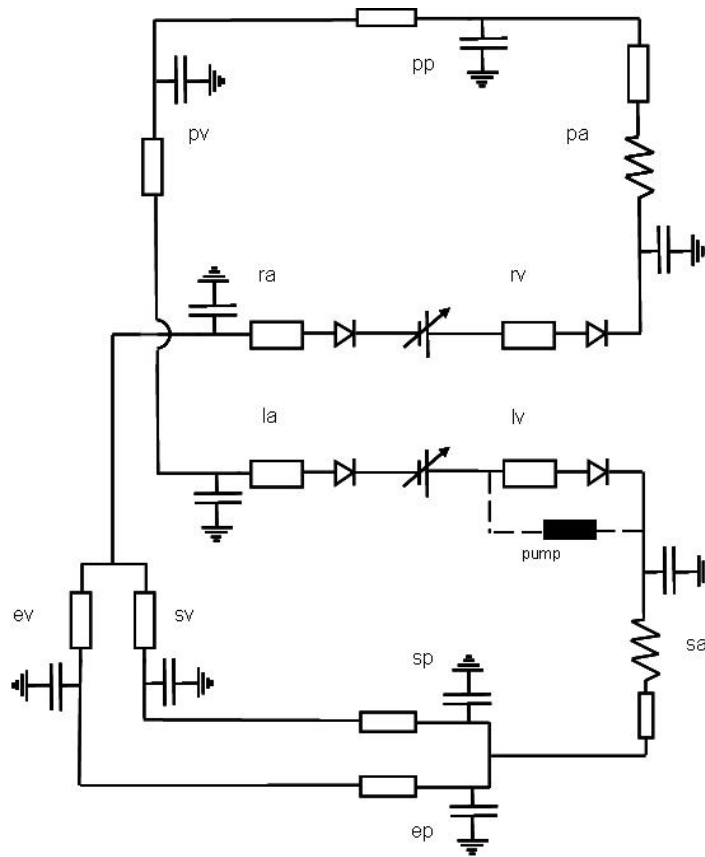


Figure 8: The full model.

pressures:

$$P_{sv} = \frac{1}{C_{sv}} [V_0 - C_{sa}P_{sa} - (C_{sp} + C_{ep})P_{sp} - C_{ev}P_{ev} - C_{ra}P_{ra} - V_{rv} - C_{pa}P_{pa} - C_{pp}P_{pp} - C_{pv}P_{cp} - C_{la}P_{la} - V_{lv} - V_u]. \quad (18)$$

The left and right ventricles are modeled using state variables that represent volumes instead of

pressures and the flows through the valves:

$$\begin{aligned}
\frac{dV_{rv}}{dt} &= F_{i,r} - F_{o,r} \\
F_{i,r} &= \begin{cases} 0 & \text{if } P_{ra} \leq P_{rv} \\ \frac{P_{ra} - P_{rv}}{R_{ra}} & \text{if } P_{ra} > P_{rv} \end{cases} \\
F_{o,r} &= \begin{cases} 0 & \text{if } P_{max,rv} \leq P_{pa} \\ \frac{P_{max,rv} - P_{pa}}{R_{rv}} & \text{if } P_{max,rv} > P_{pa} \end{cases} \\
\frac{dV_{lv}}{dt} &= F_{i,l} - F_{o,l} \\
F_{i,l} &= \begin{cases} 0 & \text{if } P_{la} \leq P_{lv} \\ \frac{P_{la} - P_{lv}}{R_{la}} & \text{if } P_{la} > P_{lv} \end{cases} \\
F_{o,l} &= \begin{cases} 0 & \text{if } P_{max,lv} \leq P_{sa} \\ \frac{P_{max,lv} - P_{sa}}{R_{lv}} & \text{if } P_{max,lv} > P_{sa}, \end{cases}
\end{aligned} \tag{19}$$

where the pressures and resistance in the ventricles are given by

$$\begin{aligned}
R_{lv} &= k_{r,lv} P_{max,lv} \\
P_{lv} &= P_{max,lv} - R_{lv} F_{o,l} \\
R_{rv} &= k_{r,rv} P_{max,rv} \\
P_{rv} &= P_{max,rv} - R_{rv} F_{o,r} \\
P_{max,lv}(t) &= \phi(t) E_{max} (V_{lv} - V_{u,lv}) + [1 - \phi(t)] P_{0,lv} (\exp(k_{E,lv} V_{lv}) - 1) \\
P_{max,rv}(t) &= \phi(t) E_{max} (V_{rv} - V_{u,rv}) + [1 - \phi(t)] P_{0,rv} (\exp(k_{E,rv} V_{rv}) - 1).
\end{aligned} \tag{20}$$

The parameter E_{max} is the ventricle elasticity at the instant of maximal contraction, and V_u is the unstressed ventricle volume. The constants k_r and k_E describe the ventricle resistance and the end-diastolic pressure–volume relationship for the heart. Parameter values that were not explicitly given by the AMC cardiologists were taken from [12].

The heart is activated by the ventricle activation function

$$\phi(t) = \begin{cases} \sin^2 \left[\frac{\pi T(t)}{T_{sys}(t)} u \right] & 0 \leq u \leq T_{sys}/T \\ 0 & T_{sys}/T \leq u \leq 1 \end{cases} \tag{21}$$

that steers $P_{max,lv}$, the isometric left ventricle pressure. The ventricle activation function is controlled by the baroreflex control system, which is a highly complex function of the sinus nerves.

For simplicity, we approximate the ventricle activation function by a simple sine function, which is shown in figure 9,

$$\phi(t) = \begin{cases} \sin(2\pi\omega) & 0 \leq \sin(2\pi\omega) \\ 0 & \sin(2\pi\omega) < 0, \end{cases} \tag{22}$$

with $\omega = 1.25 \text{ rad}$ the signal frequency which corresponds to the cardiac cycle.

The rotary pump is modeled as a tube which creates a pressure difference depending on rotational speed. It is assumed that the aortic valve closes perfectly around the tube, so when the valve is open, blood is allowed to flow through the aorta and the pump, and when the valve is closed, blood is only allowed to flow through the pump. The purge fluid is not modelled explicitly. The corresponding

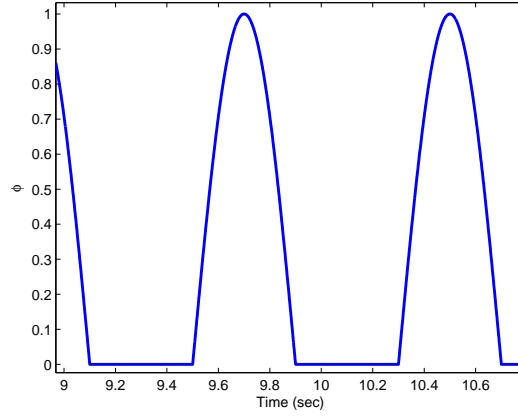


Figure 9: Approximated ventricle activation function ϕ .

equations for the outflow of the left ventricle are therefore modified to

$$F_{o,l} = \begin{cases} \frac{P_{max,lv} + P_{pump} - P_{sa}}{R_{pump}} & \text{if } P_{max,lv} \leq P_{sa} \\ \frac{P_{max,lv} + P_{pump} - P_{sa}}{R_{pump}} + \frac{P_{max,lv} - P_{sa}}{R_{lv}} & \text{if } P_{max,lv} > P_{sa}. \end{cases} \quad (23)$$

The first equation describes flow through the pump when the aortic valve is closed and the second equation describes the flow through the pump and the aorta when the aortic valve is open. The quantity and direction depend on the pressure difference between the left ventricle and the aorta, and on the pressure difference created by the pump. Here, R_{pump} is the flow resistance inside the pump.

As mentioned before, figure 7 shows the flow in l/min for the Impella LP 2.5 pump as a function of the pressure difference over the cannula. The different lines represent different rotational speeds. The exact experimental details were unknown to us, but it is our strong belief that the pressure difference at the horizontal axis is artificially induced. Figure 7 suggests that locally around an operating point a linear relation

$$F = \frac{P_{pump} - \Delta p}{R_{pump}} \quad (24)$$

for the flow through the cannula is reasonable. Here Δp is the artificial pressure difference at the horizontal axis of the experimental data, and R_{pump} is the flow resistance of the pump. We only consider the highest rotational speed of 51000 rpm here. Two data points from the experimental data available during the week (which differed from the ones presented in figure 7) were chosen to determine R_{pump} around the operating point, which resulted in $R_{pump} = 2.25 \text{ s} \cdot \text{mmHg/ml}$.

6 Numerical Results

The full model was simulated in Matlab with a Runge-Kutta difference scheme with varying timesteps. Figure 10 shows the PV loop (left) and the cardiac output (right) in millilitres per second for a normal patient with and without a pump. The pump pressure was taken to be constant at $P_{pump} = 25 \text{ mmHg}$ after calibration based on the experimental data from PV-loop measurements. This means that the pump creates a constant pressure difference of 25 mmHg , and therefore also implies that for pressure

differences between the left ventricle and the aorta that are higher than 25 mmHg , backflow through the pump arises. Other unknown constants were taken from [12]. We took initial conditions which

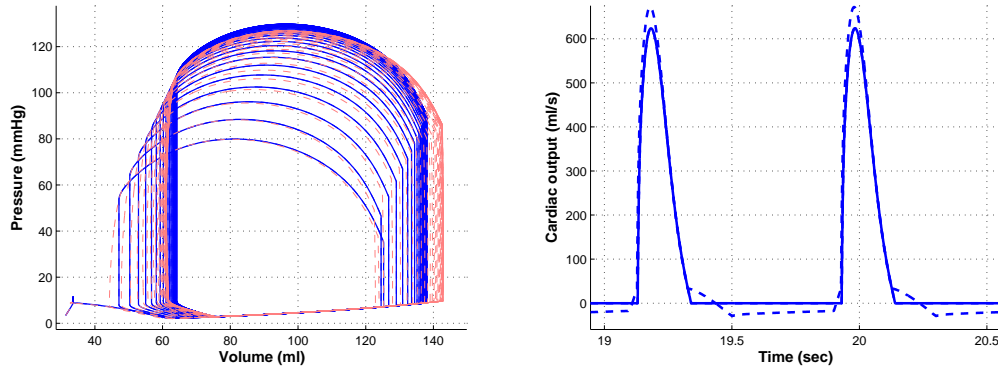


Figure 10: Left: Simulated PV loop. Right: Cardiac output for a normal patient with pump (dash-dotted line) and without pump (solid line). $P_{pump} = 25 \text{ mmHg}$.

differ slightly from the normal operating conditions of the heart to show that the dynamics converge to a stable limit cycle.

In the left plot of figure 10, we observe that starting from some initial value for P_{lv} and V_{lv} , the cardiovascular system stabilizes, and the PV loop converges to a steady cycle. The PV loop of the patient with pump shows a higher maximal volume, which is caused by backflow through the cannula during relaxation. The constant suction induced by the pump also allows less pressure buildup inside the left ventricle at the end of the systolic phase. The right plot of figure 10 shows that the peaks are higher for the patient with the pump, because the pump induces extra outflow during contraction. This increases the cardiac output. On the other hand, during relaxation there is a small backflow through the pump, decreasing the overall cardiac output, since the pressure P_{pump} is too small here to prevent backflow during relaxation.

We also simulated the system with a higher pump pressure of $P_{pump}=160 \text{ mmHg}$. Figure 11 shows experimental data for the PV loop of the heart of a patient with coronary artery disease, with and without the Impella LP 2.5 pump at the highest rotational speed. The left plot of figure 12 shows that for a normal heart the maximal volume of the left ventricle and the bottom left corner in the PV loop are shifted to the left after insertion of the pump, in correspondence to figure 11. The upper arcs in the PV loop of the simulated normal heart are absent in the measured PV loop of the weak heart. The right plot of figure 12 shows that there is no backflow anymore, and due to the constant outflow, the left ventricle volume is smaller at the end of the systolic phase, which results in a smaller peak in the cardiac output during contraction.

The area below the cardiac output graph equals the total cardiac output in ml , and a comparison of figures 10 and 12 shows that the total cardiac output has increased when a stronger pump is used. More specifically, the pump caused an increase from 73.9 ml in one cardiac cycle of 0.8 seconds without the pump to 79.3 ml when the pump was used, corresponding to an increase from 5.54 litres per minute to 5.95 litres per minute. This increase of approximately 7% is the result of almost 1.5 litres per minute extra cardiac output during the relaxation phase, and roughly 1 litre per minute less cardiac output during the contraction phase.

Further increase of P_{pump} to $P_{pump} = 1000 \text{ mmHg}$ gives negative volumes in the PV loop, which

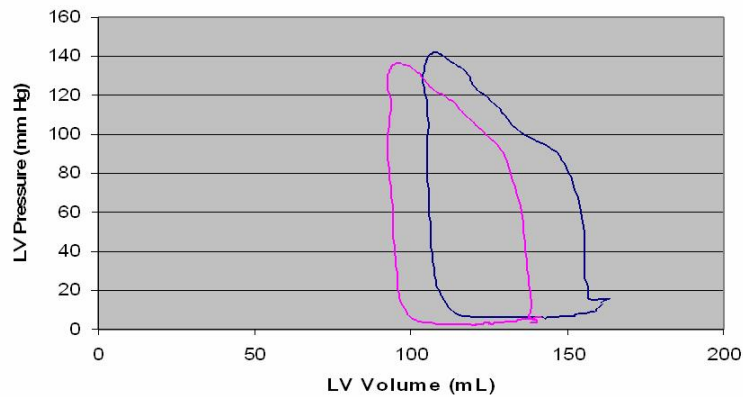


Figure 11: PV loop of the heart of a patient with coronary artery disease, with and without a pump (Measurements provided by AMC).

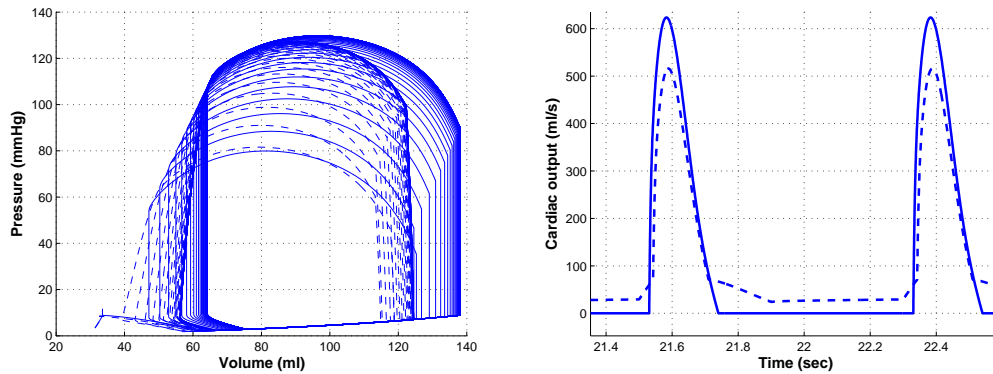


Figure 12: Left: PV loop, and Right: Cardiac output for a normal patient with pump (dash-dotted line) and without pump (solid line). $P_{pump} = 160 \text{ mmHg}$.

can be explained as follows. Due to the linearity of the differential equations, state variables are allowed to be negative. The resistances, compliances, and inertances are either measured for normal blood flow or calibrated to fit the model, and therefore the model is only realistic whenever the state variables do not deviate too much from the values of a normal blood flow.

7 Conclusions and Suggestions for Further Research

In this project, we have modeled the effect of a rotary blood pump on the behaviour of a cardiovascular system. It turns out that the resistance encountered by the blood flowing through the pump is a very important design parameter when one tries to calibrate such a model to existing experimental results for the pressure–volume relationships. By varying the pressures generated by the pump, we were able to see phenomena such as backflow through the pump in our simulations.

Under normal operating conditions we found an increase in the cardiac output by 7% as a result of the pump, which is the net result of a substantial increase in output during the relaxation phase

but also a substantial decrease in output during the contraction phase.

There is obvious room for more complex models in future research. We believe that such models should not necessarily involve more detailed modeling of the environment such as the artery systems, since its dynamics seem to be captured quite well, but should rather investigate the exact relationship between pressure and flow through the pump under more extreme circumstances. The design specification of the pump that we investigated during this project focusses on maintaining a constant rotational speed but one might easily envisage control systems for the pump in which other specifications are formulated, to enable an even more beneficial contribution from the mechanical device.

References

- [1] ABIOMED, Inc. *IMPELLA® RECOVER® LP 2.5 System -Û Instructions for Use*, April 2006.
- [2] P. Broemser and O. Franke. Über die Messung des Schlagvolumens des Herzens auf unblutigem Weg. *Zeitung fur Biologie*, 90:476–507, 1930.
- [3] L. De Pater and Van den Berg J.W. An electrical analogue of the entire human circulatory system. *Med. Electron. Biol. Eng.*, 42:161–166, 1964.
- [4] A. Ferreira, S. Chen, M.A. Simaan, J.R. Boston, and J.F. Antaki. A nonlinear state-space model of a combined cardiovascular system and a rotary pump. In *Proceedings of the 44th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC'05)*, pages 897–902, Seville, Spain, December 2005.
- [5] O. Frank. Die Grundform des arteriellen Pulses. *Zeitung fur Biologie*, 37:483–586, 1899.
- [6] B.J. Grant and L.J. Paradowski. Characterisation of pulmonary arterial input impedance with lumped parameter models. *Am. J. Physiol.*, 252:585–593, 1987.
- [7] S. Hales. *Statistical Essays II Haemostatics*. Innays and Manby, London, U.K., 1733.
- [8] P. Molino, C. Cerutti, C. Julien, G. Cuisinaud, M. Gustin, and C. Paultre. Beat-to-beat estimation of Windkessel model parameters in conscious rats. *Am. J. Physiol.*, 274:171–177, 1998.
- [9] J.T. Ottesen, M.S. Olufsen, and J.K. Larsen, editors. *Applied mathematical models in human physiology*. SIAM Monographs on Mathematical Modeling and Computation. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2004.
- [10] H. Piene and T. Sund. Does normal pulmonary impedance constitute the optimum load for the right ventricle? *Am. J. Physiol.*, 242:154–160, 1982.
- [11] C.A. Taylor and M.T. Draney. Experimental and computational methods in cardiovascular fluid mechanics. In *Annual review of fluid mechanics*. Vol. 36, pages 197–231. Palo Alto, CA, 2004.
- [12] M. Ursino. Interaction between carotid baroregulation and the pulsating heart: a mathematical model. *Am. J. Physiol.*, 275:H1733–H1747, 1998.
- [13] N. Westerhof, F. Bosman, J. Cornelis, and A. Noordergraaf. Analog studies of the human systemic arterial tree. *J. Biomech.*, 2:121–143, 1969.

-
- [14] N. Westerhof, G. Elzinga, and Sipkema P. An artificial arterial system for pumping hearts. *J. Appl. Physiol.*, 31:776–781, 1971.
 - [15] N. Westerhof and A. Noordergraaf. Arterial viscoelasticity: a generalized model. Effect on input impedance and wave travel in the systematic tree. *J. Biomech.*, 3(3):357–79, 1970.
 - [16] T. Yamaguchi, T. Ishikawa, K. Tsubota, Y. Imai, M. Nakamura, and T. Fukui. Computational blood flow analysis - new trends and methods. *Journal of Biomechanical Science and Engineering*, 1:29–50, 2006.
 - [17] M. Yoshigi and B.B. Keller. Characterisation of embryonic aortic impedance with lumped parameter models. *Am. J. Physiol.*, 273:19–27, 1997.

Cabin crew rostering at KLM: optimization of reserves

Marco Bijvank*[†] Jarek Byrka[‡] Peter van Heijster* Alexander Gnedin[§]
Tomasz Olejniczak[¶] Tomasz Świst[¶] Joanna Zyprych[¶] Rob Bisseling[§]
Jeroen Mulder^{||} Marc Paelinck^{||} Heidi de Ridder^{||}

Abstract

In this paper, we will discuss the issue of rostering jobs of cabin crew attendants at KLM. Generated schedules get easily disrupted by events such as illness of an employee. Obviously, reserve people have to be kept ‘on duty’ to resolve such disruptions. A lot of reserve crew requires more employees, but too few results in so-called secondary disruptions, which are particularly inconvenient for both the crew members and the planners. In this research we will discuss several modifications of the reserve scheduling policy that have a potential to reduce the number of secondary disruptions, and therefore to improve the performance of the scheduling process.

Key words: airline crew rostering, reserve duties, soft flights

1 Introduction

KLM (Koninklijke Luchtvaart Maatschappij N.V. also known as KLM Royal Dutch Airlines) has more than 100 aircraft and over 8,000 cabin flight attendants. Every week, a new roster is received by the cabin crew, which shows their assignments for the next several weeks. There are about 6,000 flights assigned to crew members each week. Besides the flights other assignments are rostered as well, such as trainings, days off and reserve duties.

1.1 Rostering of Flights

The assignment of cabin crew to flights is a difficult problem in which many different aspects have to be taken into consideration. First, the cabin crew is divided into four ranks: Senior Pursers, Pursers, Business Class Flight Attendants and Economy Class Flight Attendants. The last two ranks are sometimes denoted by the stripes (*band* in Dutch) on their sleeves: two stripes for the Business Class Flight Attendants (*2-bander* or 2B) and one stripe for the Economy Class Flight Attendants

*Vrije Universiteit, Amsterdam

[†]corresponding author, mbijvank@few.vu.nl

[‡]CWI, Amsterdam

[§]Universiteit Utrecht

[¶]Adam Mickiewicz University, Poland

^{||}KLM

⁰We would like to thank the other participants of the KLM group during SWI 2007. Especially, Niels Oosterling, Allard Veldman and Erik van Werkhoven.

(‘1-bander’ or 1B). There are certain regulations on how many crew members of a particular rank have to be on a certain flight. In general, there are only Pursers and 1-banders for flights within Europe (*short-haul* flights), while all four ranks have to be scheduled on intercontinental (*long-haul*) flights.

Second, the crew has to be qualified to fly on a particular aircraft type. There are in total six different aircraft types, and each crew member is qualified to fly a maximum of three different types. Crew members of the same rank and with the same qualifications are grouped into divisions. Currently, the KLM cabin crew members are divided into 17 divisions.

According to legal regulations and the collective labour agreement of the KLM cabin crew, each flight duty should be followed by a minimum number of hours of time off. The length of this time off depends on the characteristics of the duty such as duration, time of day, time difference between origin and destination. Also, after several flight duties the flight attendant is entitled to a number of days of leave, which also depends on the characteristics of the previous assignments.

The combining of flights in such a manner as to optimize the balance between duty time and leave is a specialized process due to the complexity of the regulations. This is why flights are combined into predefined patterns prior to the assignment process. A combination of flight duties followed by the appropriate number of days of leave is called a *pairing*. The length of a pairing can vary from three days up to seventeen days. The process of constructing these pairings is called the *pairing process*. This is performed by an application called *Carmen Crew Pairing*.

Pairings are assigned to specific crew members several weeks ahead of the actual day of execution. This is done with consideration of the rank and aircraft qualifications of the specific crew member. But also the flight preferences and requests for leave on specific days are taken into account. The requests are evaluated according to certain priority rules so as to avoid conflicts (such as a number of crew members requesting the same assignment). These requests result in assignments of pairings to crew members prior to the rostering of the remaining pairings, which is performed by the application *Carmen Crew Rostering*. See Kohl and Karish [3] for rostering algorithms exploited in such a system. Carmen produces a roster for a period of two weeks, the *published period*. The resulting roster is rather fixed and cannot change much. The roster after these two weeks is also published, however, it can still change quite a lot.

Currently, about 80% of all the flight assignments are assigned as requests. This method has a serious drawback, because it does not allow for Carmen to optimize the crew rostering with respect to efficient allocation of resources. Therefore, a new strategy is currently being introduced called *Preferential Bidding System*. It consists in giving preferences rather than requests for specific flights. A preference could refer to a single flight or something more general, like the preference to start and finish early or the preference for flights to the Far East. The Carmen Crew Rostering system will then try to comply with a maximum number of preferences while also optimizing the efficiency of the rosters.

1.2 Disruptions

In general, a schedule is subject to changes. The flight assignments can become disrupted. In order to handle these disruptions, reserve duties are assigned to crew members in-between their regular activities. The reserves can take over flights that have become vacant. This process of resolving disruptions and adapting the schedule is an *online procedure*. That is, as soon as a disruption is reported, a reserve is assigned to the disruption. When a reserve is assigned to a disrupted flight, the entire pairing or its remainder is assigned to the reserve crew member.

There are several types of disruptions. *Internal disruptions* deal with disruptions of individual

crew members, like illness. *External disruptions* affect more crew members, like short-term commercial changes, delays or problems on the day of operations. Next to these *primary disruptions*, there are also *secondary disruptions*. It often happens that the disrupted pairing consists of more days than the assigned reserve duty. In that case, the reassigned reserve crew member will no longer be able to perform the pairing which was originally assigned following his or her reserve block. As a consequence, this flight will become disrupted as well. This is what we call a secondary disruption. Such a disruption can cause another disruption, a *ternary disruption*. This *domino effect* will continue until a disrupted flight occurs outside the published period, or when it is assigned to a reserve who has enough days left to take over the remaining days of the disrupted pairing. This is illustrated with an example for three crew members in Figure 1. In the left figure, we see a possible roster produced by Carmen Crew Rostering. If on day 1 crew member 1 is disrupted, crew member 2 gets assigned to the disrupted flight. However, the pairing of crew member 2 starting on day 6 is also disrupted. Therefore, crew member 3 is assigned to this secondary disruption on day 6.

DAY1	DAY2	DAY3	DAY4	DAY5	DAY6	DAY7	DAY8	DAY9
FLT	FLT	FLT	FLT	FLT	OFF	OFF	FLT	FLT
RES	RES	RES	OFF	OFF	FLT	FLT	FLT	OFF
FLT	FLT	OFF	OFF	RES	RES	RES	RES	RES

DAY1	DAY2	DAY3	DAY4	DAY5	DAY6	DAY7	DAY8	DAY9
DIS	DIS	DIS	DIS	DIS	DIS	DIS	DIS	DIS
FLT	FLT	FLT	FLT	FLT	OFF	OFF	???	???
FLT	FLT	OFF	OFF	RES	FLT	FLT	FLT	OFF

Figure 1: The roster before and after the disruption of crew member 1.

Besides the concept of disruptions, crew members can also recover (after for instance illness). This means that they become available for duty again. After an internal disruption of a crew member his or her schedule is completely erased. As a consequence, a recovered crew member has no tasks left until the end of the published period and hence can be assigned new flights.

In the remainder of this paper, we refer to *flight blocks* when we talk about pairings. Note the difference between a *flight duty (reserve duty)* and a *flight block (reserve block)*. The first definition does not include days off (or days of leave), while the latter does.

Currently, a reserve block consists of five consecutive days of reserve -by duty followed by two days off. Each day, the reserve duty is restricted by a start time and end time which can vary from one reserve to the other. There are six different start times over a day and the end time is always eight and a half hours later. This means that a disrupted flight may only be assigned to those reserves of which the flight starts within the reserve duty. The number of reserve blocks (*level of reserves*) that needs to be assigned to crew members each day is determined at the beginning of each season. There is, however, no model available with which these decisions are made. At the moment, this is done by employees of the Planning Department of KLM who mostly use their past experience and good judgement.

The goal of this research is to come up with a good reserve strategy for the cabin crew of KLM. A reserve strategy is defined by the number of reserve blocks that have to start each time unit and the configuration of each reserve block, i.e., the length of the reserve duty, the number of days off, and where to locate the days off. The quality of a specific reserve strategy can be determined by the following performance measures:

- The number of secondary disruptions: this is a measure of the uncertainty for crew members about the assignment following the reserve block.

- The number of unused reserve days: these days will make it necessary to roster extra crew, making the reserve strategy more expensive.
- The number of ‘open’ days. These are days between the end of an assigned disruption and the remainder of the roster for the assigned reserve. In order to be able to utilize these days, it will mostly be necessary to reassign a part of the existing assignment of the crew member. This is not desirable if one wants to maintain the remainder of the assignment intact as much as possible.

There is an intricate trade-off between these three measures. For example, open days can be avoided at the expense of creating secondary disruptions, whereas unused reserve days can be avoided by assigning duties to reserves at the earliest possibility, thus creating more open days.

The rostering process of all KLM crew is quite complex. Figure 2 shows the relationships between the different components that are discussed in this section. The construction of the actual schedule is performed by Carmen. In this paper, our goal is to define the input for Carmen about the reserve duties that have to be assigned to crew members. Section 2 discusses the assumptions and simplifications we made. In Section 3, we propose several methods to determine how many reserve blocks should start on a particular day, and how long these reserve blocks should be. In Section 4, we discuss the *soft flight approach*, which can be seen as an extra constraint for Carmen to prevent the domino effect to occur. The main idea of this approach is to have a pairing after a reserve block assigned to a reserve crew member on duty for the same length of time. This pairing is called a soft flight. This guarantees that no ternary disruption occurs. The possibilities for other configurations for the reserve block are considered in Section 5. After producing a schedule, we should also have a method to evaluate the performance of the schedule. We therefore introduce an algorithm to analyze how different schedules perform in Section 6. In the final section, we obtain some numerical results on the different approaches and compare them.

2 Assumptions and Simplifications

Airline Crew Rostering problems belong to the most difficult problems to solve since they are NP-hard (see Ní Éigeartaigh and Sinclair [1]). There are, however, principles and heuristic rules that result in solutions that are good enough in practice. In order to find such principles we have to make some assumptions since not all details can be modelled properly.

The first simplification we make is that we do not distinguish between cabin crew, i.e., we use no ranking and no specific qualifications for aircraft type of the crew members. Consequently, all crew members are interchangeable which simplifies the assignment of reserves to disruptions. This simplification is only justified when the reserve blocks are rostered to members of each of the 17 divisions (as explained in Section 1) proportional to the number of crew members of each division required to perform the scheduled flights.

For reasons of simplicity we assume that all crew members work full-time. An employee working part-time is entitled to have more days of leave. Another simplification we make has to deal with the time units. We round everything to days. So, a reserve duty is for 24 hours on a day instead of eight and a half hours.

We assume that a disruption can only occur on the first day of a flight block. As mentioned in the previous section, we can distinguish short-haul and long-haul flights. Most of the time, a long-haul flight duty consists of only one flight, where a pairing on a short-haul flight is likely to consist of multiple flights. Consequently, a flight duty for a short-haul can be disrupted on each day while a

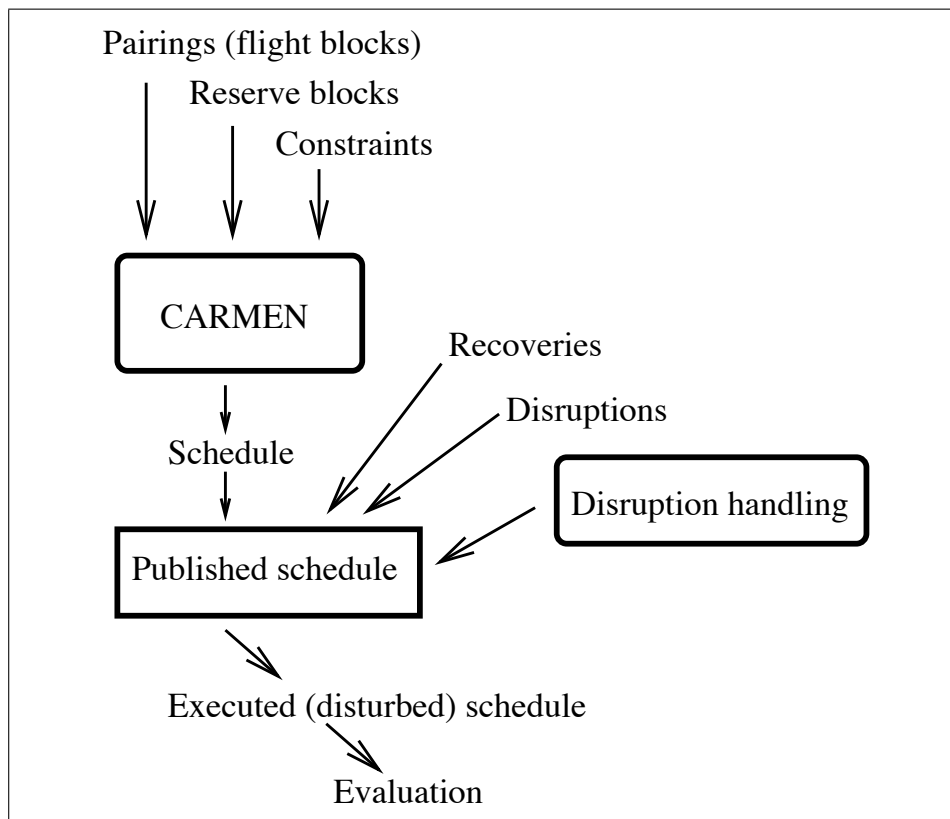


Figure 2: All aspects to be considered in the rostering problem.

flight duty for a long-haul can only be disrupted on the first day. Therefore, we make the assumption of considering long-haul flights only. Note that this assumption ignores the possibility of a crew member getting ill during his or her flight, or a malfunction of an aircraft at its foreign destination (flight blocks always start and end in Amsterdam). When we look at the length of short-haul flight blocks and the current reserve strategy, this assumption is not a problem. From historical data we know that about 29% of the flight blocks has a length of 6 days, 15% has length 8, 13% has length 7, 12% has length 11, 10% has length 10, 7% has length 9, and the remainder is small. The range of lengths is from 2 to 16 days.

Most short-haul flight blocks have duties of at most 5 days flying and 2 days off afterwards. This is exactly the same as the current configuration of the reserve blocks. Therefore, it is less likely that the current reserve strategy results in problems when these short-haul flights get disrupted. On the other hand, the current reserve strategy will most likely cause secondary disruptions when a long-haul flight is disrupted, since more than half of the long-haul flights exceed the length of the current reserve block. Therefore, the only way to deal with disruptions on long-haul flights without causing a secondary disruption is with recoveries. Since disruptions of short-haul flights are not a problem, it is reasonable to simply ignore short-haul flights completely in this research.

The final simplifications we make have to deal with the handling of internal disruptions, external disruptions, and recoveries. Internal disruptions result in less available crew members. Therefore,

they require actual reserves. On the other hand, external disruptions will most likely result in changes of the reserve crew, since the crew becoming available due to an external disruption can partly be used as a reserve again. A disrupted crew member becomes available again for services after a predictable (e.g., external disruption) or unpredictable (e.g., internal disruption) time period. The unpredictable recoveries are modelled as the start of a reserve duty with infinite length, i.e., a length equal to the published period of two weeks.

The main idea of this paper is to model the crew members that get disrupted as *workforce out-flow*, while crew members returning to service after a disruption (i.e., recovering) are modelled as *workforce in-flow*. In the long-run, by constant number of employees, the total workforce in-flow and out-flow is approximately equal – a *workforce conservation principle*. However, at each time instant there is a mismatch of the flows, which must be resolved by reserves.

It seems only wise that a disrupted crew member should return to his or her original roster as soon as possible, since this was found to be optimal. Hence, keeping as close to the original schedule as possible means staying close to optimality. This approach is advocated in Kohl *et al.* [4, Section 3.2], where closeness to the original schedule stands as a principal objective of disruption management. In order to use this principle, we have to make sure that secondary disruptions are prevented as much as possible. Otherwise, the original schedule is mixed up even more. Therefore, the reserve blocks must cover the long-haul flights. This can be achieved by rostering longer reserve blocks as compared to the current reserve strategy. In the next section, we develop different techniques that exploit this concept.

3 Level of Reserves

The previous section made clear that secondary disruptions have to be avoided as much as possible. With the current situation nearly every disrupted long-haul flight will cause a secondary disruption since there are no indefinite recoveries. Therefore, longer reserve blocks are proposed. In this section, we develop three techniques that determine the number of reserve blocks that has to start on a particular day with a certain length. The first technique is based on the concept of constructing reserve blocks that are copies of the long flight blocks, see Section 3.2. The second technique copies the flight blocks proportional to their occurrence, see Section 3.3. In the third technique, we use a more statistical model to construct the reserve blocks, see Section 3.4. But let us introduce some notation first.

3.1 Notation

A given flight schedule is defined by the number of crew members starting their flight block of type j at day k (denoted by S_{jk}). A type j can involve the characteristics length, rank, and aircraft type. We write $j \geq i$ if the characteristics of type j exceed (nonstrictly) the characteristics of type i (in the case of length or rank), or they are compatible (in the case of aircraft type). Plainly, $j \geq i$ means that a reserve of type j can be used to serve a disrupted flight of type i . In our simplified model, where we ignore rank and aircraft type, we can identify the type of block with its length, so that $j \geq i$ if and only if $j \geq i$.

We want to determine the number of crew members starting their reserve block of type j at day k (denoted by T_{jk}). Therefore, we have to model the disruptions and recoveries. Internal disruptions are assumed to be independent over time, as well as independent over all crew members. Consequently, each crew member can have an internal disruption with some probability p_{int} , where p_{int} is

based on the historical data. External disruptions are also assumed to be independent over time. For simplicity, external disruptions are analyzed for each flight attendant instead of per flight. Therefore, external disruptions are also independent over all crew members. Consequently, each crew member can get an external disruption with some probability p_{ext} , where p_{ext} is based on the historical data. We have to emphasize here that this is not always what happens in reality, but it is a useful simplification for an initial model.

The actual number of internal disruptions that requires a reserve of type j at day k is a random variable denoted by X_{jk} , where X_{jk} has a binomial (S_{jk}, p_{int}) distribution. In order to determine the number of recoveries used on a particular day, we assume independence over time. Consequently, we can define the number of recovered crew members on day k as a random variable Y_k with a probability distribution function $f_Y(y)$ with $y \in \{0, 1, 2, \dots\}$, and expectation $\mathbb{E}[Y]$, both independent of k .

3.2 Mirror Longest Flights

As mentioned in Section 2, we would like to exploit the idea of preferring long reserve blocks over short ones for long-haul flights. In this section we use a maximum length for the reserve block configuration. Several options are presented in Table 1.

option	1	2	3	4	5	6
length reserve duty	5	6	7	8	9	10
number of days off	2	2	3	3	3	4
length reserve block	7	8	10	11	12	14

Table 1: Different options for the configuration of the (longest) reserve block.

The actual number of reserve blocks that has to start with this configuration depends on the flight schedule. When there are more long flights, the reserve blocks should be long as well. But it is meaningless to schedule more reserve blocks of a particular length than there are flight blocks with at least this length.

The idea of this approach is to copy long flight blocks and replace the flight duty with a reserve duty. This process of copying is referred to as producing a *mirror*. Therefore, we first determine the number of flight blocks that has to start on day k (i.e., $\sum_j T_{jk}$) based on a minimal flight reserve cover ratio α ,

$$\frac{\sum_j T_{jk}}{\sum_j S_{jk}} \geq \alpha, \quad \forall k.$$

Currently KLM uses a fixed α equal to 4%. This means that we would like to copy at least 4% of the longest flights scheduled for day k as reserves, where the maximum length of the reserve blocks is restricted to those of the fixed configurations (as in Table 1). This approach assumes flight block types to be characterized by their length only (as mentioned in Section 2). Since it is preferred that disruptions of the longest flights are solved by recoveries, we do not copy the first $\mathbb{E}[Y]$ longest flights. This is beneficial because the longest flights have a higher probability of causing a secondary disruption, and recoveries cannot have secondary disruptions since they have no schedule yet. Note that this strategy only incorporates averages.

The advantage of the *mirror longest flight approach* is that long reserve blocks are created. This approach, however, also has its disadvantages. Proportionally there are more long reserve blocks compared to the current situation. As a result, more reserve days are rostered as compared to the

current situation. To resolve this, either the cover ratio of 4% reserves starting on a particular day has to decrease, or the cover ratio should be based on the number of reserves and flights scheduled at a particular day instead of only on the ones starting at day k , i.e.

$$\frac{\sum_{l=1}^k \sum_{j=k-l+1}^{\infty} T_{jl}}{\sum_{l=1}^k \sum_{j=k-l+1}^{\infty} S_{jl}} \geq \alpha \quad \Rightarrow \quad \sum_j T_{jk} \geq \alpha \sum_{l=1}^k \sum_{j=k-l+1}^{\infty} S_{jl} - \sum_{l=1}^{k-1} \sum_{j=k-l+1}^{\infty} T_{jl}.$$

This will be referred to as the *alternative cover ratio*.

3.3 Mirror Flights Proportional

The mirror longest flight approach prefers long reserve blocks. A disruption of shorter flight blocks will result in partial usage of such long reserve blocks. But also less reserves will start at a particular day when a crew member is rostered as a reserve for around 4% of the time. Another technique to solve the problem is to copy (or mirror) the flights in the same way as discussed in Section 3.2, but the number of reserve blocks with a particular length should be proportional to the number of flight blocks starting that day with the same length. Both options for the cover ratio can be used as mentioned in Section 3.2. We can deal with the recoveries in the same way as well.

3.4 Statistical Model

When all flight block types are unique and cannot exceed another type ($j \not\geq i$ for $i \neq j$), only reserves of type j can be used for a disruption of a flight block with characteristic type j . In such a situation, the probability of requiring more reserves than available should be low, e.g., less than 5%:

$$P(X_{jk} > T_{jk}) < 0.05 \quad \Leftrightarrow \quad P(X_{jk} \leq T_{jk}) \geq 0.95 \quad \forall j, k. \quad (1)$$

In reality, however, characteristic types can exceed each other. In this paper, we assume flight and reserve block length to be the only characteristic, i.e., $i \geq j$ if and only if $i \geq j$. Consequently, disrupted flights of type j can only be resolved with reserves of type i if $i \geq j$ to exclude secondary disruptions. Furthermore, recoveries can be used as well. Therefore, Equation (1) can be reformulated as

$$P\left(X_{jk} \leq T_{jk} + Y_k + \sum_{i=j+1}^{\infty} (T_{ik} - X_{ik})\right) \geq 0.95, \quad \forall j, k. \quad (2)$$

Based on the central limit theorem (see Ross [5]), the binomial distribution for X_{jk} can be approximated by a normal distribution with mean $\mu_{jk} = S_{jk}p_{\text{int}}$ and variance $\sigma_{jk}^2 = S_{jk}p_{\text{int}}(1 - p_{\text{int}})$. For simplicity, the stochastic recovery variable Y_k is also assumed to be normally distributed, with parameters μ_{rec} and σ_{rec}^2 .

We can rewrite Equation (2) as

$$P\left(\sum_{i=j}^{\infty} X_{ik} - Y_k \leq \sum_{i=j}^{\infty} T_{ik}\right) \geq 0.95, \quad \forall j, k. \quad (3)$$

The left-hand side of this inequality represents the number of required reserves of at least length j , while the right-hand side represents the number of available reserves of at least this length. The probability of a mismatch should be less than 5%. Since the disruptions and recoveries are independent of crew members, the number of required reserves ($\sum_i X_{ik} - Y_k$) also has a normal distribution

with a mean equal to $\sum_{i=j}^{\infty} S_{ik} p_{\text{int}} - \mu_{\text{rec}}$ and a variance equal to $\sum_{i=j}^{\infty} S_{ik} p_{\text{int}} (1 - p_{\text{int}}) + \sigma_{\text{rec}}^2$. Therefore, Equation (3) equals

$$\frac{\sum_{i=j}^{\infty} T_{ik} - \left(\sum_{i=j}^{\infty} S_{ik} p_{\text{int}} - \mu_{\text{rec}} \right)}{\sqrt{\sum_{i=j}^{\infty} S_{ik} p_{\text{int}} (1 - p_{\text{int}}) + \sigma_{\text{rec}}^2}} \geq \Phi^{-1}(0.95), \quad (4)$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. This is equivalent to

$$T_{jk} \geq 1.645 \sqrt{\sum_{i=j}^{\infty} S_{ik} p_{\text{int}} (1 - p_{\text{int}}) + \sigma_{\text{rec}}^2} + \sum_{i=j}^{\infty} S_{ik} p_{\text{int}} - \mu_{\text{rec}} - \sum_{i=j+1}^{\infty} T_{ik}. \quad (5)$$

With this iterative procedure, we can determine T_{jk} by starting with the longest reserve block and then each time decreasing j and computing the next value of T_{jk} .

4 Soft flight approach

The previous section described several methods to determine the lengths and number of reserve blocks that start on a particular day. In this section, we will consider flight blocks following reserve blocks in a roster and give them a special status. We will call these special flight blocks *soft flight blocks* (SFBs). See also Figure 3. Note that this approach also deals with the input for Carmen, independent of the techniques discussed in the previous section. These SFBs have several special properties, and therefore, should be treated differently in the disruption management process. First of all, SFBs suffer the most from secondary disruptions. On the other hand, they account for just a few percent of all flight blocks (the number of SFBs is equal to the level of reserves, which is around 4% of the total flights at the moment). Hence they may potentially be treated with special care.

We propose to create dedicated reserve blocks to cover soft flight blocks. Although the idea here is similar as in the previous section, the method is quite different. First the level of reserves is determined, as in Section 3. Next they have to be assigned to crew members. In the SFB approach, we propose to start as many SFBs of a particular length as there are reserves starting that day with the same length. The latter is determined first. This principle is shown in Figure 3. Since reserves, and therefore, SFBs occur infrequently, most of these requests can easily be granted. If a dedicated reserve is not needed to prevent a secondary disruption, this will be known in advance and the reserve can then still serve as a regular reserve. This can be taken into account when determining the required level of reserves.

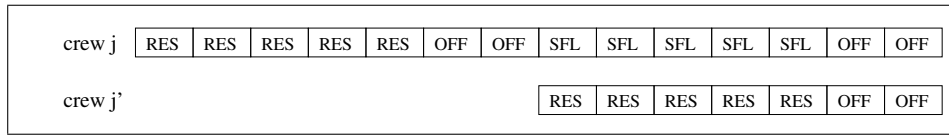


Figure 3: Soft flight block for crew member j , with a dedicated reserve j' .

Ensuring the availability of reserve blocks having the appropriate length to cover a disrupted SFB should stop the domino effect, i.e., a secondary disruption will almost never cause a ternary disruption. Currently, the end of the published period is often used to stop this domino effect. Stopping the domino effect could help to extend the planning horizon in the future.

One could go a step further and exploit the regularity of used reserve patterns together with the soft flight approach. We could pre-assign a potential secondary disruption on a SFB of crew member j to the crew member j' whose reserve block follows the reserve block of j . We could inform crew member j' in advance that he or she is the backup of j . It would probably be advantageous for j' to have this information. Additionally, such an automatization of handling secondary disruptions could simplify the online disruption handling process. Note that this is true since each day the level of reserves is about the same and therefore after a reserve block another reserve block starts for a different crew member.

5 Configuration of reserve blocks

As mentioned in Section 1, we can also change the configuration of a reserve block, i.e., the distribution of the off days in the reserve block. Currently, a reserve block consists of five days of reserve duty followed by two days of leave. In this section we discuss the possibility of changing this configuration of reserve blocks. To be more specific, we consider the possibility of moving the days off from the end of the block to the middle. It is desirable for the crew to have some days off at the end of a reserve period, but not necessarily all of the days off.

When we consider the configuration of a reserve block, we have to make a choice which days of the reserve block are most appropriate to be the starting days for serving disrupted flights. If the length of the disrupted flight is at most the length of the reserve block, we would prefer to give this flight to the person whose remaining length of reserve block just covers the flight. Consequently, no secondary disruption is caused and no open days occur in the roster. But if the disrupted flight does not fit to any available reserve block (which is almost always the case with disruptions of long-haul flights), then the choice of the reserve depends on the flight he or she has scheduled just after the reserve block. This is one more reason for introducing SFBs and controlling their lengths.

Imagine a situation where no reserve block is available to cover a particular disruption. Consequently, a secondary disruption occurs. In such a case, one would prefer to assign this disruption to a person at the end of his or her reserve block, such that serving the disrupted flight block would finish at the same time as the SFB of the reserve. The open days are minimized with such a principle. In the SFB approach, the secondary disruption would be served by its dedicated reserve crew member and both reserves could return to their original schedules afterwards.

It makes little sense to speculate what would be the best configuration of a reserve block without extensive testing. Just to illustrate the idea, one could consider the situation illustrated in Figure 4. Assume we use the SFB approach. The first line shows a current configuration of 5 days reserve and 2 days off. When we use the first day to cover disruptions without causing open days, we can use it for disrupted flights of length 7 and 14. When this does not occur, the reserve can be used on the second day for disruptions of 6 or 13 days, and so on. Disrupted flight blocks of length 8 or 9 always result in open days. Such flight blocks are quite common for long-haul flights, as we know from the historical data mentioned earlier. Therefore, we propose to shift the two days off to day 4 and 5 of the reserve block, as shown on the second line of Figure 4. Note that on the other hand the reordering of off days may increase the level of reserves needed, since off days will occur more often during a reserve block.

RES	RES	RES	RES	RES	OFF	OFF	SFL	SFL	SFL	SFL	SFL	OFF	OFF
7 14	6 13	5 12	4 11	3 10									
RES	RES	RES	OFF	OFF	RES	RES	SFL	SFL	SFL	SFL	SFL	OFF	OFF
7 14	6 13	5 12			9	8							
							RES	RES	RES	RES	RES	OFF	OFF

Figure 4: Two reserve blocks with different internal structure. The numbers under the reserve blocks indicate the length of a disruption which they can fit ‘perfectly’, that is, without causing open days in the roster (in combination with their soft flight block). If we want to optimize the perfect fittings, we need to determine which disruptions have the highest probability to occur. In particular, if there are more disruptions of length 8 and 9 than with length 3, 4, 10 and 11 we prefer the reserve block pattern : 3 days active / 2 days off / 2 days active.

6 Comparison of scheduling techniques

In Section 3, we have developed different approaches to determine the number of reserve blocks that have to start at a certain day with a particular length. In order to compare the different approaches, we developed an analytical comparison procedure, which is the subject of this section.

For the comparison procedure we use the same assumptions as made in Section 2, i.e., no specific qualifications for air crew members (no ranks, no particular qualifications for air craft type), only long-haul flights, recoveries are reserves of infinite length (i.e., equal to the publishing period). Another assumption we make, such that reserves can be assigned to disruptions more easily, is complete knowledge of the availability of crew members being on reserve and of all disruptions on a particular day before any reserve gets assigned to a disruption. In practice, this means that all disruptions are reported before any of them is resolved by assigning reserves. Also all people who recover and become available again are reported before this assignment takes place.

For the assignment of reserves to disruptions we want to have a fixed disruption handling scheme. This scheme should result in using the reserve capacity in such a way that secondary disruptions and open days are prevented as much as possible. Therefore, when a disruption of length j occurs, we check recoveries first. Otherwise, we want to use a reserve block of length j . If there is no such reserve block, we gradually increase the length of the desired reserve block, each time by one day. When all options of reserve blocks of at least j days are checked and no reserve is available, a secondary disruption occurs. Now, use a reserve block of length $j - 1$ and gradually decrease the length by one day until the disruption is resolved. This approach results in avoiding secondary disruptions as much as possible, and if they still occur making them as short as possible, thus minimizing the probability of ternary disruptions occurring.

Before we actually assign reserve blocks to disruptions, we make another simplification from reality. In reality, we distinguish between a crew member being allowed to be called as a reserve or having days off in a reserve block. In our comparison procedure, we do not want to keep track of this distinction. So, we only look at reserve blocks and flight blocks disregarding the days off.

Let us introduce some notations for the performance measures:

- U_k = the expected number of unused reserves on day k
- V_k = the expected number of secondary disruptions caused on day k
- W_k = the expected number of disruptions that cannot be resolved with recoveries and reserves on day k

The first two definitions have already been mentioned in Section 1. We added the third performance measure, to have an understanding whether there is a lack of reserve crew. In the comparison technique we predict the number of available reserve blocks of a particular length based on the assignment scheme as described above and on the probabilities for recoveries and disruptions.

For each day, we would like to keep track of all flights that can get disrupted. Therefore we define the sequence set D_k as all possible disruptions on day k , where $k \in \{1, \dots, K\}$ and K being the publishing period (i.e., the planning horizon). A possible disruption $d \in \{1, \dots, |D_k|\}$ on day k is identified by the length of the flight $l_d^{(k)}$ and by its probability of requiring a reserve $p_d^{(k)}$. There are as many possible disruptions of length l on day k as the number of scheduled flight blocks with this length (i.e., S_{jk}). Consequently, $|D_k|$ equals $\sum_j S_{jk}$. Each possible flight block gets initially disrupted with probability $p_{\text{int}} + (1 - p_{\text{int}})p_{\text{ext}} = 1 - (1 - p_{\text{int}})(1 - p_{\text{ext}})$. We order the flights in D_k such that $l_d^{(k)} \geq l_{d+1}^{(k)}$ (i.e., in decreasing order of their length).

Set the initial probabilities for having i reserve blocks available of length j starting at day k (denoted by p_{ijk}) as

$$p_{ijk} = \begin{cases} 1, & \text{if } i = T_{jk} \\ 0, & \text{otherwise} \end{cases} \quad \forall j, k$$

and for the availability of recoveries

$$p_{i,K+1,k} = f_Y(i) \quad \forall i, k.$$

Since disruptions occur, the probabilities of reserves being available (i.e., p_{ijk}) change over time. The analytical procedure to update these probabilities and to compute the performance measures is given in Appendix A.

7 Numerical Results

For the numerical results we used actual data of KLM. For simplicity, we assumed each day to be the same, i.e., $T_{jk} = T_{j,k+1}$ and $S_{jk} = S_{j,k+1}$. When we evaluated the data on long-haul flights, we observed the values as presented in Table 2 and Table 3, where $\mu_{\text{rec}} = 7.1$ and $\sigma_{\text{rec}}^2 = 8.353$.

j	S_{jk}	j	S_{jk}	j	S_{jk}	j	S_{jk}
1	0	5	8	9	27	13	13
2	8	6	108	10	38	14	3
3	0	7	49	11	46	15	5
4	0	8	55	12	12	16	2

Table 2: The number of flight blocks starting on a given day.

The probabilities for internal and external disruptions are $p_{\text{int}} = 0.06$ and $p_{\text{ext}} = 0.07$, respectively.

y	$f_Y(y)$	y	$f_Y(y)$	y	$f_Y(y)$
0	0	7	0.122905	14	0.011173
1	0.011173	8	0.139665	15	0.011173
2	0.033520	9	0.136872	16	0
3	0.064246	10	0.069832	17	0
4	0.092179	11	0.047486	18	0
5	0.103352	12	0.019553	19	0.002793
6	0.120112	13	0.013966		

Table 3: The probability distribution for the recoveries used to resolve disruptions.

Currently KLM uses a 4% flight reserve cover ratio where a reserve block consists of 5 days of duty and 2 days off. So, in total 15 reserve blocks start on a daily basis (4% of $\sum_j S_{jk} = 374$). Since the length of a reserve block is 7 days, there are 105 crew members scheduled as reserve each day. We would like to keep this number more or less fixed when comparing the different approaches discussed in Section 3 (rounding real values to integers causes some small fluctuations). The cover ratio and the alternative cover ratio are the same since each day is identical. For the different configuration approaches we get T_{jk} as in Table 4. Since all days are the same, the performance measures are also the same for each day, as presented in Table 4. The current policy with 15 reserves starting a reserve block of length 7 each day is reflected in our simplified model by option 1 in the table. The statistical method without the restriction of the current 105 crew members being planned as reserves is given in the last column, as ‘ideal statistical’. Notice that in our simplified model the number of needed reserves for the current policy is higher than the number available, by more than the accepted margin of 5% (about 5 disruptions), since $W_k = 18.64$, see column 2 in Table 4.

Table 5 presents a more detailed study of the requirements, only based on internal disruptions. As mentioned already, 374 flight blocks start each day, as well as 15 reserve blocks. Each day, there are on average around 7 recoveries. The expected number of internal disruptions of length j equals $p_{\text{int}} S_{jk}$ (column 3 in Table 5). On average, the 7 longest disruptions can be resolved with recovered crew members. The remaining disruptions have to be resolved with reserve crew (column 4 in Table 5). More than 17 reserves are required each day. So, this points to a lack of two reserves in our simplified model. This can also not be resolved with scheduling two disruptions in one reserve block. This is shown by multiplying the expected number of required reserves with their desired length, which represents the expected number of reserve days that have to be scheduled each day (column 5 in Table 5). This equals almost 120 reserves, compared to 105 in the current schedule.

8 Conclusions and Future Research

In this paper we developed several techniques to determine the number of reserve crew to be scheduled. The use of the statistical model is recommended since it incorporates the stochastic nature of disruptions and recoveries. This technique also uses a service level that can be interpreted quite easily, instead of the cover ratio. One topic for further research would be to investigate the effects of the SFB approach. This could be done by simulating the process of handling disruptions. Another option is to extend the comparison technique discussed in Section 6 to incorporate the configuration of the flight blocks and reserve blocks. It would also be interesting to look at more exact optimization methods like column generation and genetic algorithms (see for instance Guo [2] and Thiel [6]).

Our numerical results show that the statistical method halves the number of secondary disrup-

j	mirror longest flight						mirror proportional	statistical	ideal statistical
	opt 1	opt 2	opt 3	opt 4	opt 5	opt 6			
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	1
6	0	0	0	0	0	0	4	0	8
7	15	0	0	0	0	0	2	0	4
8	0	13	0	0	0	0	2	2	4
9	0	0	0	0	0	0	1	2	2
10	0	0	11	0	0	0	1	3	3
11	0	0	0	10	0	0	2	3	3
12	0	0	0	0	9	0	0	1	1
13	0	0	0	0	0	5	1	0	0
14	0	0	0	0	0	3	0	0	0
15	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0
$\sum_j T_{jk}$	15	13	11	10	9	8	13	11	26
$\sum_j j \cdot T_{jk}$	105	104	110	110	108	107	108	109	206
U_k	0.46	0.52	0.57	0.58	0.54	0.55	0.51	0.65	92.97
V_k	38.43	30.61	22.68	21.59	24.27	24.49	31.55	17.31	0.02
W_k	18.64	15.12	9.87	8.92	9.92	10.78	15.38	7.99	0

Table 4: The number of unused reserves U_k , secondary disruptions V_k , and unresolved disruptions W_k for the different approaches.

j	S_{jk}	disruptions	required reserves	scheduled reserve days
1	0	0	0	0
2	8	0.52	0.52	1.04
3	0	0	0	0
4	0	0	0	0
5	8	0.52	0.52	2.60
6	108	7.01	7.01	42.09
7	49	3.18	3.18	22.28
8	55	3.57	3.57	28.58
9	27	1.75	1.75	15.78
10	38	2.47	0.73	7.29
11	46	2.99	0	0
12	12	0.78	0	0
13	13	0.84	0	0
14	3	0.19	0	0
15	5	0.32	0	0
16	2	0.13	0	0
total	374	24.29	17.29	119.65

Table 5: The expected number of internal disruptions, required reserves, and scheduled reserve days.

tions compared to the current policy, comes close to the 5% accepted level of unresolved disruptions, while having the same low number of unused reserves. The current policy is already working quite reasonably in practice; the numerical results for our simplified model indicate that these can be improved by using the statistical model.

References

- [1] Ní Éigeartaigh, T. and Sinclair, D. (2000) The Airline Crew Rostering Problem, School of Computer Applications, Dublin City University, <http://www.computing.dcu.ie/research/papers/2000/0300.ps> .
- [2] Guo, Y. (2005) Decision Support Systems for Airline Crew Recovery, University of Paderborn, PhD thesis.
- [3] Kohl, N. and Karish, S.E. (2004) Airline crew rostering: problem types, modeling and optimisation, *Ann. Oper. Res.* **127**, 223-257.
- [4] Kohl, N., Larsen, A., Larsen, J., Ross, A. and Tiourine, S. (2004) Airline disruption management - perspectives, experiences and outlook, Carmen Research and Technology Report CRTR-0407, September 2004, http://www.carmensystems.co.uk/research_development/research_reports.htm .
- [5] Ross, S.M., 2003, *Introduction to Probability Models*, Eighth Edition, Academic Press, San Diego.
- [6] Thiel, M.P. (2005) *Team-oriented Airline Crew Scheduling and Rostering: Problem Description, Solution Approaches, and Decision Support*, Paderborn.

A Probabilities of having reserve capacity

```

1  for k = 1 to K
2    Vk = 0, Wk = 0
2    for d = 1 to ∑j Sjk
3      p-r = pd(k)
4      l = ld(k)
5      prob_ext = pext · (1 - p-r) / (1 - pext)
6      j = K + 1
7      prob_not_avail = p0jk
8      for i = 1 to ∞
9        pijk = { pi+1,j,k · p-r + pijk          if i = 0
                pi+1,j,k · p-r + pijk · (1 - p-r) otherwise
10     next i
11     p-r = prob_not_avail · p-r
12     for j = l to K
13       prob_not_avail? = p0jk
14       for i = 0 to ∞
15         pijk = { pi+1,j,k · p-r + pijk          if i = 0
                  pi+1,j,k · p-r + pijk · (1 - p-r) otherwise
15     next i
16     if (j > l)
17       daysleftj-l,k+l += p-r · (1 - prob_not_avail)
18     endif
19     p-r = prob_not_avail · p-r
20   next j
21   Vk = Vk + p-r
22   for j = l - 1 to 1
23     prob_not_avail = p0jk
24     for i = 0 to ∞
25       pijk = { pi+1,j,k · p-r + pijk          if i = 0
                  pi+1,j,k · p-r + pijk · (1 - p-r) otherwise
26     next i
27     pd(k+j) += p-r · (1 - prob_not_avail) / ∑i Si,k+j    ∀ d ∈ Dk+j
28     p-r = prob_not_avail · p-r
29   next j
30   Wk = Wk + p-r
31   for i = 0 to ∞
32     pilk = { pilk · (1 - prob_ext)          if i = 0
               pilk · (1 - prob_ext) + pi-1,l,k · prob_ext otherwise
33   next i
34   next d
35   Uk = ∑j ∑i i pijk
36   pi,j,k+1 = ∑l=0i pl,j+1,k · pi-l,j,k+1    ∀ i, j
37   pj = daysleftj,k+1 / ∑j Tjk
38   pi,j = (∑i Tmk) pji (1 - pj)∑m Tmk - i
39   pi,j,k+1 = ∑l=0i plj · pi-l,j,k+1    ∀ i, j
40 next k

```

For every day k , we treat every scheduled flight as a possible disruption $d \in \{1, \dots, \sum_j S_{jk}\}$. The impact of a disruption is represented by lines 3–33. Flight block d with length $l_d^{(k)}$ has an initial

probability $p_d^{(k)}$ of getting disrupted. Let us introduce the following probabilities:

- prob_not_avail = the probability that no reserve block of length j is available
- prob_ext = the probability that there is an external disruption
- p_r = the probability that a crew member on flight d is disrupted,
and still requires a reserve crew member

To resolve this disruption d we first check whether a recovered crew member is available (i.e., a reserve block of length $K+1$). Therefore, the probabilities p_{ijk} are adjusted according to line 9. When there is no recovery available, a reserve is still required, and p_r gets updated (line 11). Next, we check whether a reserve block of at least $l_d^{(k)}$ days is available (lines 12–20). No secondary disruption will occur in these situations. The procedure to update p_{ijk} is the same as on line 9. If a reserve block with length j ($j > l_d^{(k)}$) is available, the remaining reserve days of this reserve block ($j - l_d^{(k)}$ days) become available after resolving flight d (at day $k + l_d^{(k)}$). This is represented on line 17. If there is still no reserve crew found (with probability p_r), a secondary disruption occurs. Therefore V_k increases. Again the same update procedure is used to update p_{ijk} . A secondary disruption occurs at day $k + j$ since the reserve crew member cannot perform its scheduled flight block after his/her reserve block. Since it can affect any of the flights, all flights at day $k + j$ obtain a higher chance of getting disrupted (line 27). If there is still no reserve available, an emergency crew member has to be called (i.e., W_k increases).

When the disruption is external, the crew member becomes available to resolve other disruptions (line 32). When all disruptions on a day are taken care off, we can calculate the expected number of remaining reserves U_k . These reserves can be used the next day (line 36). The probability of having reserve blocks of j days remaining and becoming available at day $k+1$ equals p_j (line 37). The actual number of reserve blocks becoming available with j days remaining has a binomial distribution (line 38). These have to be added to the reserve crew (line 39).

A sampling problem from lithography for chip layout

Eric Cator^{*} *Tammo Jan Dijkema*[†] *Michiel Hochstenbach*[‡]
Wouter Mulckhuysen[§] *Mark Peletier*[‡] *Georg Prokert*[‡]
Wemke van der Weij[¶] *Daniël Worm*^{||}

Abstract

Given a list with simple polygons and a sampling grid geometry, we calculate the sample (greyscale) value of each grid pixel, such that the residual error in a certain norm is sufficiently small, taking into account that the amount of computational operations should be minimal.

Key words: sampling, Fast Fourier Transform, computational operations.

1 Introduction

1.1 Lithography

ASML is the worldwide leader in *lithographic techniques* for the semiconductor industry. Since the different steps in the lithography process are important for the discussion of this report, we describe them in some detail.

The main function of the lithographic system of ASML is to expose a silicon wafer with a pattern of given light intensity. This exposure step is embedded in the following sequence:

1. A designer creates the design of the desired pattern in a CAD system;
2. The CAD system perturbs the design to pre-correct for deficiencies in the optical system (see below);
3. For each layer of the chip a mask is created;
4. If needed, the material for the current layer of the chip is deposited on the wafer;
5. A photoresist coating is added on the wafer
6. The lithographic system exposes the photoresist on the wafer with the mask for the current layer;
7. The photoresist is developed;

^{*}Technische Universiteit Delft

[†]Universiteit Utrecht

[‡]Technische Universiteit Eindhoven

[§]ASML, Veldhoven, the Netherlands

[¶]CWI, Amsterdam

^{||}Universiteit Leiden

8. Those parts of the wafer where the photoresist is gone are further treated (i.e. etched).
9. The remainder of the photoresist is removed.

Steps 4 to 9 may be repeated as many as forty times, thus creating a landscape on the wafer with details at many different heights.

Only step 6 takes place within the lithography machines, and therefore only this step is under control here; the rest of this list is to be treated as a given, and this will be important below.

1.2 Maskless lithography

A new development is the creation of lithography systems that use an array of microscopic mirrors instead of the mask. Step 3 above is then skipped, and the output of step 2 is fed directly into step 6. The mirrors are then positioned in such a way that the resulting illumination pattern on the wafer corresponds to the pattern that a mask-based optical system *would* create.

Maskless lithography will not be competitive with mask-based lithography for production machines, since mask-based systems have higher throughput. Maskless systems will have the advantage, however, of eliminating the costly and time-consuming step of creating the mask. Therefore maskless systems may be beneficial in the pre-production (testing) phase. Since the client should be able to switch from testing on maskless systems to production on mask-based systems without changing the pattern, it is essential that the maskless systems act as a *drop-in replacement* for mask-based systems. In other words,

the maskless system has to imitate a mask-based machine *with all its imperfections*.

Some of the constraints posed to the Study Group stem from this requirement.

1.3 The pattern

The pattern that is created in steps 1 and 2 is defined as a collection of polygons. In mathematical terms we will describe this pattern by the indicator function χ_P of the union of these polygons, i.e.

$$\chi_P(x) := \begin{cases} 1 & x \text{ lies in some polygon} \\ 0 & \text{otherwise.} \end{cases}$$

For reasons of computational efficiency it is important that the pattern is given by polygons as these can be described easily by a sequence of the coordinates of their vertices, in the order that defines their boundary.

1.4 The question

As part of the ‘data path’ that transforms the machine input (the output of step 2 above) into the steering signal for the array of mirrors, a sampling step is necessary. In this step the collection of polygons is transformed into an array of intensity values (in $[0, 1]$) of the same size as the array of mirrors. In this sampling step, *aliasing* is an important problem (see Section 2 and Figures 1 and 2), and the original question as posed to the Study Group was to design an efficient algorithm to perform this sampling step with sufficient control on aliasing. In Figure 1 a part of a chip layout is shown. Clearly visible are the perturbations around corners that are introduced to correct for phenomena occurring in a mask-based machine. Figure 2 shows the desired rasterization of this chip layout. Gray values represent the intensity at which a square should be exposed.

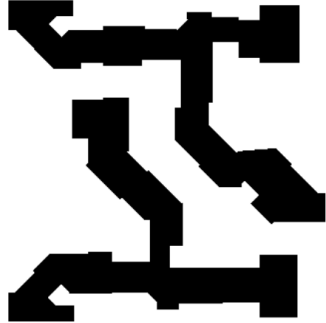


Figure 1: Part of a chip layout.

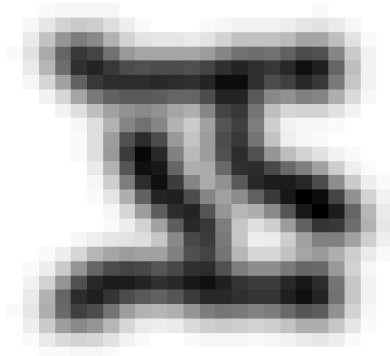


Figure 2: The desired rasterization of this chip layout.

2 Sampling, low-pass filters and aliasing

A well-known error source occurring in the processing of sampled signals is known as *aliasing*. In our context, it is best understood theoretically from the simple observation that sampling and (low-pass) filtering of a signal do not commute. To see this, interpret sampling as multiplication of a given signal in the space domain by a so-called shah distribution or Dirac comb

$$\text{III}_a := \sum_{k \in \mathbb{Z}} \delta(\cdot - ak)$$

where $\delta(\cdot - ak)$ denotes a shifted delta distribution and $1/a$ is the sampling rate. It follows from the Poisson summation formula that up to a scaling factor, the Fourier transform (denoted by a hat) of a shah distribution is a shah distribution with inverse sampling rate:

$$\widehat{\text{III}}_a = \frac{1}{a} \text{III}_{\frac{1}{a}}.$$

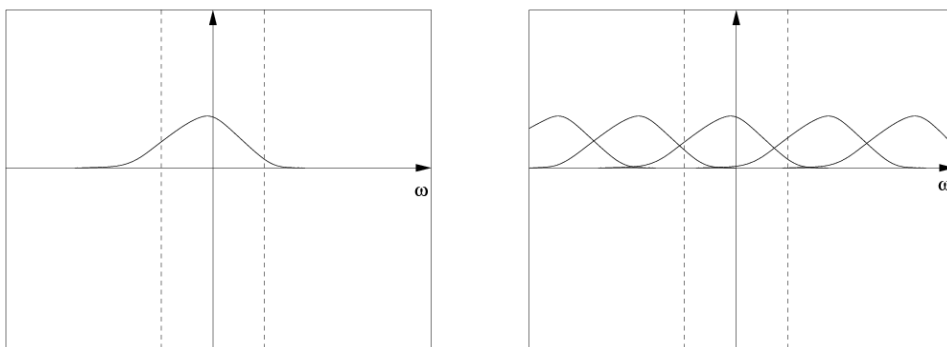


Figure 3: Spatial sampling in frequency domain representation (qualitatively). On the left the original signal is shown, on the right the resulting signal after sampling. This resulting signal consists of an (infinite) sum of scaled and translated copies of the original signal. The dashed lines show the low-pass filter bandwidth.

As pointwise multiplication in space domain corresponds to convolution in frequency domain, spatial sampling is represented in frequency domain by convolution with a shah distribution (Fig. 3). If the sampling rate is taken too small, spurious contributions of higher frequencies will appear in the low frequency band and persist even after applying a low-pass filter. For more details on sampling and the aliasing effect we refer to [1].

There are two possible remedies to avoid this:

1. use of a higher sampling rate
2. preprocessing of the input data to suppress high frequencies before sampling.

The first of these possibilities is impractical because of the prohibitively high computational effort. However, for a smaller model problem this approach can be used to obtain a “golden standard” for the purpose of comparison to results of more efficient algorithms.

The second possibility is known as applying an *anti-aliasing filter* and is used in practice by ASML. Schematically, the signal flow is described as in Fig. 2.



Figure 4: Signal flow scheme. The kernel of the anti-aliasing filter is denoted by K . In the frequency domain, the low-pass filter is applied by pointwise multiplication with the characteristic function χ_f of the set $\{|\omega| < f\}$. The application of the anti-aliasing kernel has to be corrected afterwards by pointwise multiplication by \hat{K}^{-1} .

As the main computational effort is in the anti-aliasing filter, it was the task of the study group to assess and, if possible, improve its efficiency. (It has to be remarked here that, of course, the anti-aliasing filter is a low-pass filter in itself, so one might expect either a high computational effort or aliasing effects for this filter as well. However, using the special structure of our data it is possible to achieve a higher efficiency than in a standard FFT-based algorithm.)

3 First approach: numerical convolution

In the case of a sample field of $100\mu\text{m} \times 50\mu\text{m}$ with $6 \cdot 10^6$ polygons of 20 vertices each and a grid pitch of $20\text{nm} \times 20\text{nm}$, the computational effort of the method currently used by ASML is estimated as to be about $1.9 \cdot 10^{13}$ operations, half of them being multiplications.

We propose the following different approach: introduce a second grid which refines the original (coarse) grid by a factor N in both directions. We will denote the gridpoints of the finer grid by P_{ij} with the understanding that P_{ij} is a point of the coarse grid if both i and j are multiples of N .

1. For simplicity we assume that the fine grid has total height and width $2LH$ where H is the distance between two coarse grid points. Let K be the kernel of the anti-aliasing filter as above and set

$$K_{ij} := K\left(\frac{iH}{N}, \frac{jH}{N}\right) \quad i, j \in \mathbb{Z}, \quad -NL \leq i, j \leq NL.$$

The calculation of these values has to be performed only once.

2. For all fine grid points P_{ij} , set

$$\chi_{ij} = \begin{cases} 1 & P_{ij} \text{ lies inside some polygon,} \\ 0 & \text{otherwise.} \end{cases}$$

To do this efficiently, one might use a triangulation of all polygons and determine for each of the resulting triangles the range of the indices i, j of points inside the triangle.

3. The convolution in a coarse grid point $P_{Nk,NI}$ can now be approximated by

$$(K * \chi_P)(P_{Nk,NI}) \approx \sum_{i,j=-NL}^{NL} K_{ij} \chi_{Nk+i,NI+j} \quad (1)$$

This approximation is extremely simple to implement and very flexible with respect to the choice of grid sizes and types of kernels used.

The necessary computational effort is dominated by the summations in (1) and is easily seen to be about $4N^2L^2$ additions per coarse grid point. The refinement factor N can be interpreted as an “oversampling rate”, and experiments might be needed to determine its optimal value. When the algorithm described here is applied with $N = 100$ to the sample field described above, $3.2 \cdot 10^{13}$ additions are needed. This is in the same order of size as the analytic approach, when additions and multiplications are considered to have the same computational cost.

In Figure 5, the difference is illustrated between an oversampling factor of $N = 50$ and $N = 100$. This figure indicates that the difference between an oversampling factor of 50 and one of 100 introduces only a small error. However, this error is mostly located at the boundary of polygons (where an error causes most damage to the chip)

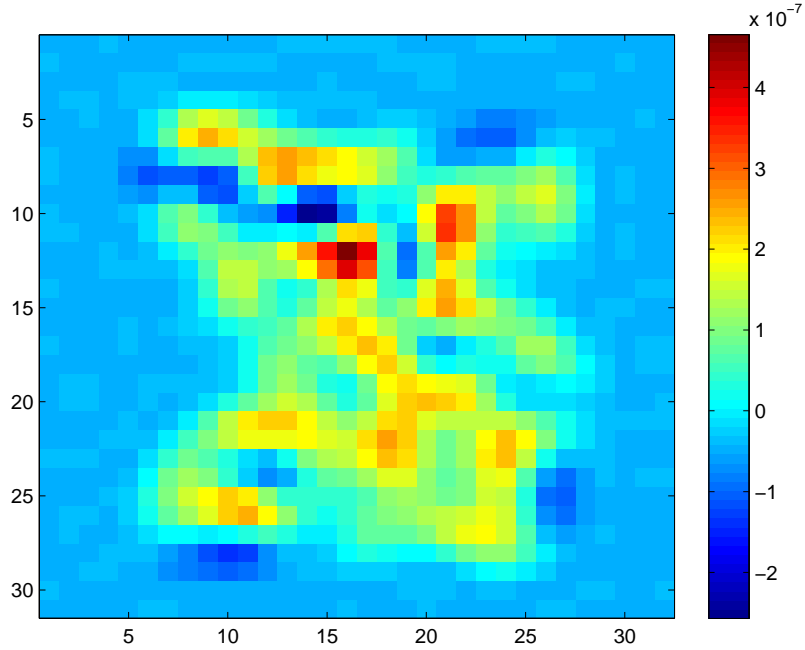


Figure 5: Oversampling effects.

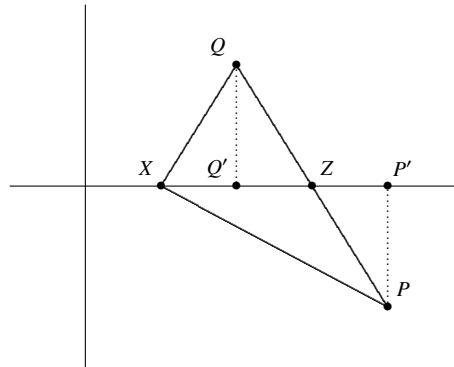


Figure 6: The triangle XPQ

4 Second approach: triangulation and look-up tables

As a second approach to the problem, we start with a triangulation of the data, so we are given the coordinates of the vertices of triangles T_1, \dots, T_N . Furthermore, we will use a two-dimensional normal kernel K with fixed bandwidth σ , so

$$K(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2} \frac{x^2+y^2}{\sigma^2}}.$$

The bandwidth σ is determined in such a way that the aliasing effect on the final coarse grid is negligible. Given a point $P \in \mathbb{R}^2$ of the coarse grid, we need to calculate the integral

$$I = \int_{T_i-P} K(x, y) dx dy$$

for each $1 \leq i \leq N$ such that T_i is close to P , in some precise sense.

The first step in approximating I is to make sure by applying a rotation that the triangle $T_i - P$ has at least one point on the (positive) x -axis. Because of the rotational symmetry of K , this will not change I . We denote this triangle by XPQ , where X is the point on the x -axis, and P and Q are the two other points.

4.1 Decomposing into right-angle triangles

We wish to decompose our triangle XPQ into four right-angle triangles. To this end, we define Z as the intersection of the line through PQ with the x -axis, P' as the projection of P onto the x -axis and Q' as the projection of Q onto the x -axis. If Z is not defined (or extremely far away), we will use a separate approach. Figure 6 depicts the situation.

Now consider XPQ as an oriented curve in \mathbb{R}^2 , where the order of the letters determines the orientation. We know that we can rewrite I as an integral over this curve. However, we can also add up curves, and it is easy to check that

$$XPQ = XPP' + ZP'P + XQ'Q + ZQQ'.$$

Clearly, these four curves all correspond to right-angle triangles. To calculate I , we can therefore determine the integral of K over the four right-angle triangles, and determine whether these integrals

should have a positive or a negative sign: if the orientation of XPP' is equal to the orientation of XPQ , the contribution of XPP' gets a positive sign, otherwise it gets a negative sign. The same holds for the other three right-angle triangles. Determining these orientations is a simple algebraic calculation.

4.2 Calculating the integral over right-angled triangles

We have reduced the problem of determining I to calculating the integral of K over a right-angle triangle with two vertices on the x -axis, among which is the vertex with the right angle. We can make sure by reflection in the y -axis that the x -coordinate x of the right angle is positive, so $x \geq 0$. Note that x either corresponds to P' or to Q' . Furthermore, by reflection in the x -axis, we can ensure that the point outside the x -axis has a positive y -coordinate, which we call $h \geq 0$. The x coordinate of the third point is given by $x + b$, where $b \in \mathbb{R}$. See Figure 7.

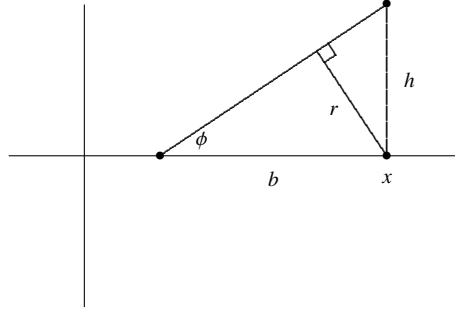


Figure 7: The right angle-triangle

We will use a different parametrization of the triangle (x, b, h) , as is shown in Figure 7. We define

$$r = \frac{|b|h}{\sqrt{b^2 + h^2}} \quad \text{and} \quad \phi = \arctan\left(\frac{h}{b}\right) \in (-\pi/2, \pi/2).$$

Note that in Figure 7, $b < 0$ and $\phi < 0$. The advantage of this parametrization lies in the fact we only need to consider a bounded subset of the parameter space, since if (x, r, ϕ) lies outside this set, $I(x, r, \phi)$ will either be very small, or almost equal to $I(x', r', \phi')$, where (x', r', ϕ') is an element of this bounded set. To see this, choose $R > 0$ such that

$$\int_{\{u^2+v^2>R^2\}} K(u, v) \, dudv < \varepsilon_0,$$

where ε_0 is a small, fixed constant, corresponding to the accuracy with which we need to evaluate I . First suppose $x > R$. Then if $b > R - x$, we have $I(x, r, \phi) < \varepsilon_0$. Otherwise, we get

$$I\left(R, \frac{|b| - (x - R)}{|b|} r, \phi\right) < I(x, r, \phi) < I\left(R, \frac{|b| - (x - R)}{|b|} r, \phi\right) + \varepsilon_0.$$

Now suppose $0 \leq x \leq R$ and fix $\phi \in (-\pi/2, \pi/2)$. We can define r_0 such that the hypotenuse of the triangle $I(x, r_0, \phi)$ is tangent to the circle with radius R . Clearly, if $r > r_0$, the set $I(x, r, \phi) \setminus I(x, r_0, \phi)$ falls outside of this circle. It is not hard to check that

$$r_0 = x - R \sin(\phi) \leq 2R.$$

So if $r > 2R$, we can choose $r = 2R$, and we will have

$$I(x, 2R, \phi) < I(x, r, \phi) < I(x, 2R, \phi) + \varepsilon_0.$$

4.3 Creating the “lookup table”

As we saw in the previous section, we only need to calculate $I(x, r, \phi)$ for $(x, r, \phi) \in [0, R] \times [0, 2R] \times (-\pi/2, \pi/2) := \Theta$. To do this, we define a grid on Θ , by dividing each of the three intervals in M parts, creating M^3 grid points. For all these points, we calculate I and also $\partial I/\partial x$, $\partial I/\partial r$ and $\partial I/\partial \phi$. This would allow us to calculate $I(x, r, \phi)$ by interpolation from the nearest grid point. Another possibility would be to just calculate I in all grid points, and use a weighted average over all nearby grid points as an approximation for $I(x, r, \phi)$.

To calculate $I(x, r, \phi)$, it is easier to work in the parametrization (x, b, h) , so

$$b = \frac{r}{\sin \phi} \quad \text{and} \quad h = \frac{r}{\cos \phi}.$$

Note that $\phi \approx 0$ is handled as an exception, since this means that the point Z is not well defined. We get, supposing that $b > 0$,

$$I = \frac{1}{2\pi\sigma^2} \int_x^{x+b} \int_0^{\frac{b+x-u}{b}h} e^{-\frac{1}{2} \frac{v^2+y^2}{\sigma^2}} dv du.$$

This means that

$$\begin{aligned} \frac{\partial I}{\partial x} &= -\frac{1}{2\pi\sigma^2} \int_0^h e^{-\frac{1}{2} \frac{x^2+y^2}{\sigma^2}} dy + \frac{1}{2\pi\sigma^2} \frac{h}{b} \int_x^{x+b} e^{-\frac{1}{2} \frac{b^2u^2+(b+x-u)^2h^2}{b^2\sigma^2}} du, \\ \frac{\partial I}{\partial b} &= \frac{1}{2\pi\sigma^2} \int_x^{x+b} \frac{(u-x)h}{b^2} e^{-\frac{1}{2} \frac{b^2u^2+(b+x-u)^2h^2}{b^2\sigma^2}} du, \\ \frac{\partial I}{\partial h} &= \frac{1}{2\pi\sigma^2} \int_x^{x+b} \frac{b+x-u}{b} e^{-\frac{1}{2} \frac{b^2u^2+(b+x-u)^2h^2}{b^2\sigma^2}} du. \end{aligned}$$

Using the formula

$$\frac{\partial I}{\partial r} = \frac{\partial I}{\partial b} \frac{\partial b}{\partial r} + \frac{\partial I}{\partial h} \frac{\partial h}{\partial r}$$

and its equivalent for $\partial I/\partial \phi$, we can create the entire lookup table.

If we have that Z is not well defined, meaning that PQ is parallel to the x -axis, we need to calculate the integral of K over the rectangle $P'PQQ'$ (and then subtracting the integral over $XP'P$ and XQQ'). Calling x the location of a point of the rectangle on the x -axis, $b \in \mathbb{R}$ the width and $h \geq 0$ the height, we can assume that $0 \leq x \leq R$, $b \in [-R, R]$ and $h \in [0, R]$, by neglecting anything outside the rectangle $[-R, R] \times [0, R]$. For the rectangle, we need to calculate (assuming again $b > 0$)

$$\begin{aligned} I &= \frac{1}{2\pi\sigma^2} \int_x^{x+b} \int_0^h e^{-\frac{1}{2} \frac{v^2+y^2}{\sigma^2}} dv du \\ &= \frac{1}{2\pi\sigma^2} \int_x^{x+b} e^{-\frac{1}{2} \frac{u^2}{\sigma^2}} du \int_0^h e^{-\frac{1}{2} \frac{v^2}{\sigma^2}} dv. \end{aligned}$$

This can be done by making a table of

$$\frac{1}{\sqrt{2\pi}\sigma} \int_x^{x+b} e^{-\frac{1}{2} \frac{u^2}{\sigma^2}} du$$

for a grid of M^2 points with $(x, b) \in [0, R] \times [-R, R]$. Of course also the derivative can be easily calculated and put in the table.

5 Conclusion

We have given two alternative approaches for the anti-aliasing filter that ASML currently uses, These approaches, however, do not clearly improve upon the current method. Our two approaches are both adjustments to this method. In our first approach, we replace the analytical approach used by ASML by numerical calculations on a finer grid, where the values of the kernel at those grid points can be calculated and stored beforehand. We get the same order of number of operations as the analytic approach used by ASML, if we choose $N = 100$, where N is the factor with which the original (coarse) grid has been refined in both directions.

Some numerical experiments have been done on this approach. For the masks in the test set it seems that N can be chosen smaller than 100 and still deliver the required accuracy. However, this is not necessarily true in general because patterns with worse aliasing behavior may exist. Further analysis / experiments will need to be done in order to determine the correct N . A disadvantage of this approach is that it requires a lot of memory to store the values of the kernel at the fine grid points, and this will cause the process to be slower.

In the second approach, we first need to triangulate the polygons, which can be done quickly, and many algorithms are available for this. Then we need to find for every triangle which grid points are close to it, and for each of those points find the right decomposition in rectangular triangles, which also should not take long, and find the required values of the integral in the look-up table. Here too, experiments need to be done to find out how large the look-up table should be to get the required accuracy, without needing too much memory.

References

- [1] BRACEWELL, R.: The Fourier Transform and its Applications, McGraw-Hill 1965

Optimizing a closed greenhouse

*Jaap Molenaar** *Onno Bokhove†* *Lou Ramaekers‡* *Johan van de Leur§*
Nebojša Gvozdenović¶ *Taoufik Bakri§* *Claude Archer||* *Colin Reeves***

Abstract

In this project we study the optimization of a closed greenhouse. Typical components of such a greenhouse are aquifers in which surplus heat (in summer) or cold (in winter) is stored. Water tanks are used to control short term variations in the heat and radiation supply due to the daily weather variability. A closed greenhouse may also yield an excess of electricity which is delivered to the public grid. The energy economics is determined by a considerable number of components. The full optimization of such a greenhouse involves the incorporation of market prices of crops and of electricity and gas, and the weather conditions. A formidable problem, as most of these inputs are stochastic variables. We therefore restrict ourselves to minimizing the energy costs given the heat and cold demand for a typical year. Discretizing the model equations on an hourly time basis, we show that this problem has a linear cost function which has to be optimized under linear conditions. Standard linear programming techniques are therefore applicable, guaranteeing our limited optimization problem of a closed greenhouse to be tractable.

Key words: optimization, energy conservation, linear programming, greenhouse, mathematical modelling

1 Introduction

The concept of using greenhouses has a long history. It allows farmers to grow products during seasons in which and at places where the climate conditions are not or far from optimal for the crop under consideration. In Europe this branch of agriculture is very dynamic. The competition is huge, especially as frontiers between the European countries become more and more open. Furthermore, the costs of energy are continuously increasing and governmental regulations to protect the environment of the greenhouses against negative influences become more and more severe. In 1997, within the framework of a project called ‘Greenhouse of the future’, a number of innovations were introduced in the classic greenhouse concept. These developments led to the idea of GeslotenKas or ‘closed greenhouse’. See Fig. 1.1 for a view from the outside and Fig. 1.2 for an inside view of such a closed greenhouse. Since 2003 this new type of greenhouse has been sold and further developed by Innogrow [1]. It is especially suitable for the growth of products that need climatic conditions that

*Wageningen Universiteit, corresponding author, jaap.molenaar@wur.nl

†Technische Universiteit Twente

‡Innogrow

§Universiteit Utrecht

¶CWI, Amsterdam

||Haute École Francisco Ferrer, Brussels, Belgium

**Coventry University, UK

vary only slightly, and for which CO₂ is an important growth factor. The new concept has economic advantages, especially for products that consume a lot of energy, such as some vegetables, roses, and orchids. The relevant growth factors in most greenhouses depend on temperature, humidity,



Figure 1.1: *Outside view of a Closed Greenhouse.*

and CO₂. By closing the windows of a greenhouse and providing it with an integrated climate and energy system, maximal control is obtained over these growth factors. For example, the CO₂ can be kept at a high level, which favours plant growth in general. Furthermore, the energy use can be kept under control, and thus optimized. An extra advantage of keeping the greenhouse closed is the lower susceptibility to diseases that might invade the greenhouse via the atmosphere. All of these aspects result in a lower energy use, together with an increase in production.

A greenhouse could be considered as a huge sun collector. Unfortunately, the radiation of the sun is not uniformly distributed over the year. In the summer there is an excess and in winter a lack of heat. Apart from the annual cycle, short-term temperature variations take place due to local weather conditions. Similar variations are found for humidity and CO₂. The idea of a closed greenhouse is to smooth these differences out by carefully controlling the local conditions.

1.1 Optimization of a closed greenhouse

The ultimate goal of optimizing the profit of a closed greenhouse includes two aspects:

- Maximizing crop growth and harvest at the right time.
- Minimizing operation costs, i.e., the energy demand to heat/cool the greenhouse.

These two aspects require the application of several submodels:

- A climate model, which determines the heating and cooling requirements of the greenhouse given the weather conditions.
- A crop model, which determines the development of the crop given the light, nutrition, CO₂ and humidity conditions.
- An energy cost model, also referred to as *utility model*, which calculates the costs if a certain energy strategy is followed.
- A market model, which predicts when the crop can be sold for a good price.

It has to be realized that several additional aspects are relevant, which make a complete optimization approach very hard, if not nearly impossible. For example, in practice the weather forecast is available for a few days. Within this time horizon an optimization tool that focuses on the energy expenditure could calculate the best strategy, and when time goes on and the weather forecast is updated, the strategy could be updated, too. However, such a short term strategy should be combined with long term issues such as crop and market developments. Furthermore, a highly restrictive condition, stemming from environmental safeguarding considerations, is that on average per year the total amount of heat/cold pumped into and released from the soil is vanishing: the system may not heat up or cool down the soil layer used for storage. So, this leads to a periodic boundary condition: after a year the situation in the storage devices must be the same as at the start of that year. Such a long term condition may have strong consequences for the short-term energy costs strategy.



Figure 1.2: *Inside view of a Closed Greenhouse.*

1.2 Goal of the project

In view of the considerations mentioned above the full treatment of optimizing the closed greenhouse is a huge task. To keep the project tractable we therefore defined a restricted goal: optimization of the energy costs in a typical year. For example, this implies that the influence of CO₂ is neglected. For temperature and radiation conditions, we take the average values over a long time period in the past. As for the crop development, we assume that a constant temperature and humidity are optimal. Typical heat/cold demands follow from these assumptions over the period of a year.

Similar assumptions hold for the energy prices. In summary, we assume to be given (on an hourly basis):

- the heat/cold demands of a ‘typical’ year, and
- the prices of electricity and gas.

Furthermore, the capacities of all storage, pump, and heat generation devices are known. Given these data, the *purpose of the project* is to: *find a heating/cooling strategy such that the costs are minimized, under the conditions that the reservoirs have the same heat/cold contents before and after the optimization period.*

In principle, the period of optimization must be a year, but for practical reasons we could start with a shorter period. It is expected that the models will yield useful information on:

- the capacities needed to cover the requirements in a typical year, and
- the best strategy for a typical year.

The insight, information, and mathematical tools gained from the model are useful in general, in particular in the following two ways:

- they allow to design new greenhouses based on the ‘closed greenhouse’ concept, and
- the mathematical techniques to be developed may directly be applied in a computer tool for the short-term optimization of the energy costs.

Finally, this report is organized as follows. In §2 we describe the components that constitute the energy sources of the closed greenhouse. In §3 the model presented is shown to be linear in the variables after discretization of the time window over which it is optimized. In §4 interim results are summarized, which are discussed in §5.

2 Energy (Utility) Model

In the present project, we focus on the energy needs of the closed greenhouse system. The purpose is to optimize the efficiency of its energy management, such that the costs are minimized. The system has a complicated network of devices to control the energy fluxes. In Fig. 2.1, a sketch is given of the energy devices in the closed greenhouse.

2.1 The simplified utility model

There are four basic demands: the heat and cold demands together with the demands for light and CO₂ for the crop. Here, as a first step we have decided to neglect the light and CO₂ demands.

Cold and hot water storages

Long term heat excesses in the greenhouse are stored in an underground aquifer, and short term heat excesses in a water reservoir alongside the greenhouse. Similar storage devices are used to store and release cold. The transfer of heat is performed by a heat exchanger.

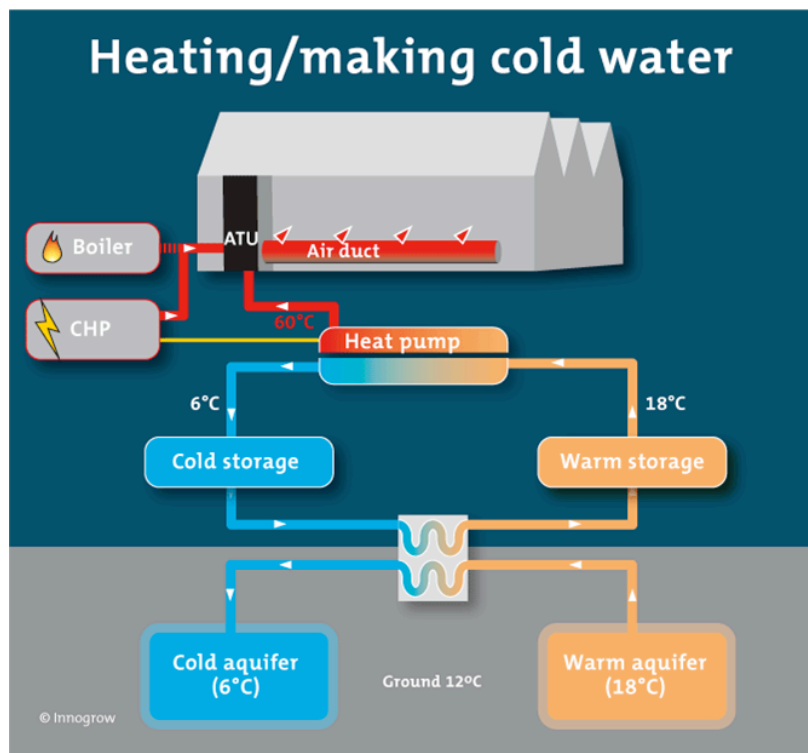


Figure 2.1: Schematic of the components involved in the energy needs of the Closed Greenhouse concept.

Heat from electricity

A consequence of our simplified model is that electricity is used only for driving the heat pump (ignoring the requirement for providing light at night). This pump has the property that it can only transfer heat and cold together, not independently. We have then to decide whether to buy electricity from the public grid or to produce it from the “combined heat power” (CHP) unit (also known as cogeneration). It is depicted in Fig. 2.2. The surplus of electricity can be delivered to the public electricity network, but the price of electricity delivered to the network is lower than the price of electricity bought from this network. Different day and night electricity prices must also be taken into account.

Heat from gas

The CHP device consumes gas (G^c). It produces not only electricity but also a heat quantity Q^c proportional to G^c . The system also contains a boiler that consumes gas and produces heat (Q^{bl}). The CO_2 generated by these two devices is not taken into account as the CO_2 demand for the crop is neglected.

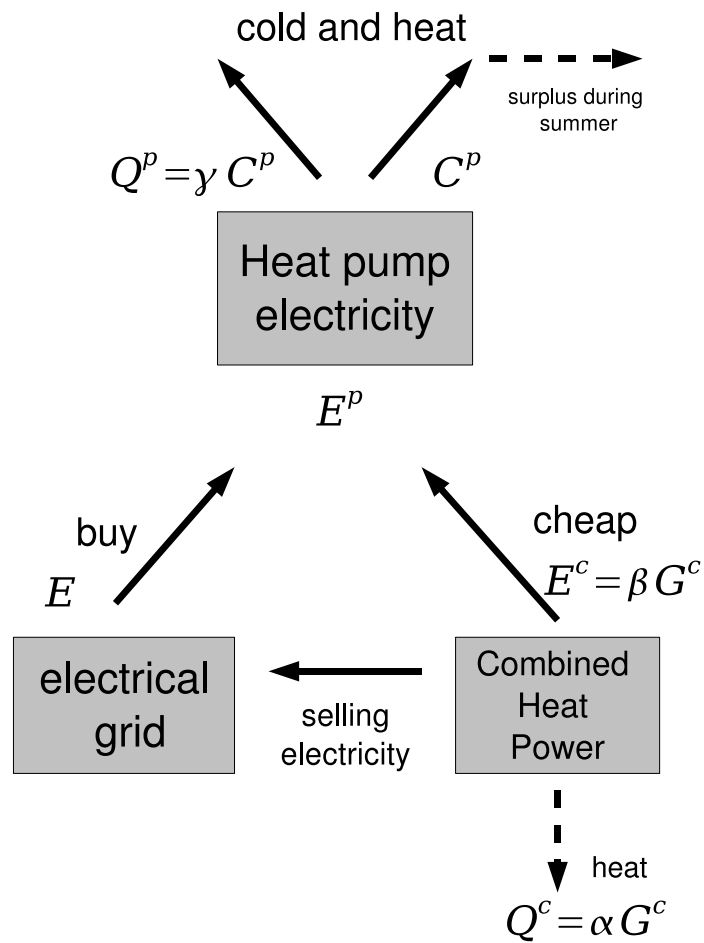


Figure 2.2: Part of the energy network of the Closed Greenhouse, showing the relations between the heat pump, the combined heat power unit and the external electrical grid.

2.1.1 Seasonal requirements

Accessing the aquifer is a very expensive initial investment, but later on it is a relatively cheap source of heat and cold, so it should be used as much as possible. What has to be done during the summer is quite clear. We only use the cold aquifer as a cheap source of cooling. Conversely, during winter, the aquifer is always configured as a heat supply. The difficulties arise with intermediate seasons where the aquifer may be switched from hot to cold according to needs. It must be noticed that switching the aquifer has a cost. During the 15 minutes after a switch, the water coming from the aquifer contains sand and cannot be used. Hence gas must be consumed to compensate the heat demand. We should also mention that during summer, we still have a demand for heat as well as cold. In addition, the air conditioning system needs to reheat greenhouse air after dehumidifying it.

2.1.2 Notation

It is clear that the system is quite intricate in view of the many couplings between the components. As for the notation, we denote the required (and prescribed) cooling demand in the greenhouse by $C = C(t)$ and the heating demand by $Q = Q(t)$.

In the system, the heating and cooling devices are uncoupled nearly everywhere, with the exception of the heat pump, which produces heat and cold at the same time. In the following, we denote an energy flux used for cooling by C . The source of such a flux is indicated with a superscript. E.g., C^p , C^a , and C^b are cooling fluxes stemming from the heat pump, the aquifer, and the cooling buffer, respectively. Heating fluxes are represented by Q . E.g., Q^p , Q^a , Q^b , Q^{bl} , Q^c indicate the heating contribution from the heat pump, the aquifer, the heating buffer, the boiler, and the CHP unit, respectively. The total heat flux from and to the aquifer is denoted by q^a , which can be positive (from) and negative (to). Electricity from the grid is denoted by E , it can be either positive (if supplied from grid to the greenhouse) or negative (the other way around). Electricity from the CHP is denoted by E^c . Gas used by the CHP is denoted by G^c and gas used by the boiler is related to the heat flux of the boiler Q^{bl} with efficiency and gas price conversion factors.

There are storage tanks or so-called *buffers* (relatively small compared to the aquifers), for short-term heating and cooling with fluxes C^b (cold temperature storage) and Q^b (warm temperature storage). The energy levels in the cold and hot buffer are H^{bc} and H^{bh} . The energy level in the aquifer is denoted by H^a .

2.2 Energy balances

In the following, we summarize the energy balances in the system. The total cooling demand $C = C(t)$ is supplied by cold fluxes from the heat pump, the aquifer and the cold buffer:

$$C = C^p + C^a + C^b. \quad (2.1)$$

The total heating demand $Q = Q(t)$ is supplied by heat fluxes from the heat pump, aquifer, hot buffer, boiler, and CHP:

$$Q = Q^p + Q^a + Q^b + Q^{bl} + Q^c. \quad (2.2)$$

The aquifer total flux q^a (a negative or positive value) is related to C^a and Q^a by

$$C^a = \max(-q^a, 0), \quad Q^a = \max(q^a, 0). \quad (2.3)$$

The energies in buffer and aquifer are related to the fluxes according to:

$$\frac{dH^a}{dt} = q^a, \quad (2.4)$$

$$\frac{dH^{bh}}{dt} = -Q^b - \mu_h H^{bh}, \quad (2.5)$$

$$\frac{dH^{bc}}{dt} = -C^b - \mu_c H^{bc}, \quad (2.6)$$

in which we introduced damping coefficients μ_h and μ_c associated with heat or cold losses. The heat pump has the special property that it produces heat and cold at the same time and at a fixed ratio:

$$Q^p = \gamma C^p \quad \text{with} \quad \gamma > 1. \quad (2.7)$$

The cooling power plus electrical demand of the heat pump equals its heating power:

$$C^p + E^p = Q^p. \quad (2.8)$$

The electricity used by the heat pump stems from the grid and the CHP (combined heat power):

$$E^p = E + E^c. \quad (2.9)$$

The gas used by the CHP yields both heat production and electrical power at fixed ratios:

$$\alpha G^c = Q^c, \quad \beta G^c = E^c, \quad (2.10)$$

with α and β the proportionality constants. In practice, the efficiency of the CHP is not 100 %, so that in general $\alpha + \beta < 1$. Because of efficiency considerations the boiler either operates at a maximum or is shut off; hence

$$Q^{bl} = 0 \quad \text{or} \quad Q^{bl} = \tilde{Q}^{bl} \quad (2.11)$$

with \tilde{Q}^{bl} a fixed value. If the CHP is working this happens at at least 50% capacity:

$$G^c = 0 \quad \text{or} \quad G^c \in [0.5, 1] \tilde{G}^c. \quad (2.12)$$

In summary, the balance equations of the system are

$$C = C^p + C^a + C^b, \quad (2.13a)$$

$$Q = Q^p + Q^a + Q^b + Q^{bl} + Q^c, \quad (2.13b)$$

$$C^a = \max(-q^a, 0), \quad (2.13c)$$

$$Q^a = \max(q^a, 0), \quad (2.13d)$$

$$\frac{dH^a}{dt} = q^a, \quad (2.13e)$$

$$\frac{dH^{bh}}{dt} = -Q^b - \mu_h H^{bh}, \quad (2.13f)$$

$$\frac{dH^{bc}}{dt} = -C^b - \mu_c H^{bc}, \quad (2.13g)$$

$$Q^p = \gamma C^p, \quad (2.13h)$$

$$C^p + E^p = Q^p, \quad (2.13i)$$

$$E^p = E + E^c, \quad (2.13j)$$

$$\alpha G^c = Q^c, \quad (2.13k)$$

$$\beta G^c = E^c, \quad (2.13l)$$

Note that there are 15 variables involved in (2.13): $C^p, C^a, C^b, Q^p, Q^a, Q^b, Q^{bl}, Q^c, q^a, H^a, H^{bh}, H^{bc}, E^p, E^c, E^p$. The range limits are denoted as:

$$0 \leq H^a \leq \tilde{E}^a, \quad 0 \leq H^{bh} \leq \tilde{H}^{bh}, \quad 0 \leq H^{bc} \leq \tilde{H}^{bc} \quad (2.14)$$

$$0 \leq C^{\dots} \leq \tilde{C}^{\dots}, \quad 0 \leq Q^{\dots} \leq \tilde{Q}^{\dots}, \quad (2.15)$$

where C^{\dots} and Q^{\dots} refer to any of the cold and heat fluxes.

2.3 Linear modelling

The aim is to formulate the problem as a linear program (LP), but some elements of the above formulation are impermissible. It should be clear that some of the equations do not allow a strictly linear model to be constructed. For example, (2.13c) and (2.13d) contain the (nonlinear) *max* function. Such a nonlinearity is circumvented by rewriting the constraints as

$$q^a = Q^a - C^a$$

and including a term

$$\lambda(Q^a + C^a) \tag{2.16}$$

in the cost function, where λ is a positive scalar. This ensures that at least one of the variables will be zero at an optimal solution of the LP; it can readily be seen that in the event that both are positive, the cost function can be decreased by reducing both values until the smaller value is zero, without affecting the constraint. This also relies on the fact that LP variables must be nonnegative—something guaranteed by any method, such as the simplex algorithm. The actual solution obtained will obviously depend on a particular choice of λ . We will defer discussion on this value, and its interpretation until §3.

The nonnegativity requirement also affects some of the variables: for example, the cold buffer flux as formulated is a directional variable with its value unconstrained in sign. This can be dealt with simply by decomposing it as

$$C^b = C^{b+} - C^{b-}$$

and then using a Boolean variable δ in a pair of constraints:

$$C^{b+} \leq M\delta \quad \text{and} \quad C^{b-} \leq M(1 - \delta),$$

where M stands for a suitably large scalar—larger than the maximum value that can feasibly be taken by the flux. The effect of this is that at most one can be positive, corresponding to the relevant direction of the flux. A similar approach is used for the hot buffer and the grid supply of electricity.

Finally, there are some either/or constraints that also need to be modelled using Booleans. For example, if the CHP is on, it must operate at more than 50% of its maximum capacity. This can be formulated as a pair of constraints:

$$G^c \leq \delta \tilde{G}^c \quad \text{and} \quad G^c \geq 0.5\delta \tilde{G}^c.$$

The situation is similar for the boiler, except that in this case, if it operates at all, it is at maximum capacity, so that only one constraint is needed:

$$Q^{bl} \leq \delta \tilde{Q}^{bl}.$$

2.4 Cost function

The cost function (in euros) is given by the following integral over the time window $[0, T]$ considered:

$$K = \int_0^T [f(t, \text{sign}(E)) E(t) + c_1 G^c(t) + c_2 Q^{bl}(t)] dt, \tag{2.17}$$

where the first term in the integrand $f(t, \text{sign}(E)) E(t)$ is the cost of the electricity and t is time. The coefficient $f(t, \text{sign}(E)) > 0$ attains four positive values and is thus boolean in nature. It takes

different values for day and night. Further, its sign indicates whether the electricity is taken from or delivered to the electricity grid. To wit:

$$f(t, E) = \begin{cases} \begin{cases} f_1 & \text{day hours} \\ f_2 & \text{night hours} \end{cases} & \text{for } E > 0 \\ \begin{cases} f_3 & \text{day hours} \\ f_4 & \text{night hours} \end{cases} & \text{for } E \leq 0 \end{cases} \quad (2.18)$$

with $f_1 > f_3$ and $f_2 > f_4$.

Further weak constraints are that the cold and warm buffers have no net influx over short periods, while the aquifer must remain in heat balance over a long period T_y , a year, say:

$$\int_0^T C^b(t') dt' \approx 0, \quad \int_0^T Q^b(t') dt' \approx 0, \quad \text{and} \quad \int_0^{T_y} q^a(t') dt' \approx 0. \quad (2.19)$$

Finally, we must remember to add the parameterized term from (2.16), which resurrects the question of the meaning of λ . Given its context, it should be clear that this relates to the cost of extracting heat from (or storing it in) the aquifer. A reasonable approximation could perhaps be determined after some lengthy calculations. Initially it was assumed by Innogrow that this cost is negligible, being dominated by the other sources, so it may be sufficient to try some rough estimates to determine the boundaries within which a particular solution remains optimal.

3 Discretization and linear programming

3.1 Linear programming

The most important insight for finding an optimization procedure for the cost function under the conditions given by the balance equations, is that both the cost function and the conditions are essentially linear in the variables. To see this for the cost function, we simply replace the integral by a Riemann sum taking as grid points hourly intervals. So, if we introduce as variables the hourly values of the 15 variables, the cost function is linear in this function space. Since the balance equations are instantaneous relations between variables, they are inherently linear in the hourly variables, too. Therefore we decided to apply linear programming techniques. The only complicating factor is that some variables are Boolean. Fortunately, standard linear programming techniques such as the 'simplex' method have been extended to incorporate Boolean variables. The idea of linear programming is to find the extremum of the cost function by ascending/descending along the nodes of a multi-dimensional simplex. Owing to the linearity this can be done in a systematic way using a simple 'greedy' algorithm to move from one node to the next. Although theoretically the computational complexity of this approach is not polynomial, practical experience over 5 decades has shown that the average-case performance is very good, and optimal solutions can be obtained relatively quickly. Adding a moderate number of Booleans degrades the performance a little; if very many Booleans are required then it may take a very long time to find an optimum. However, even in such cases solutions of excellent quality are usually found quickly—the problem is in proving that the quality is good!

3.2 Algebraic system in linear form

Note that by using hourly variables, the total number of variables is very high, especially if the time period is taken to be very long. For example, 15 variables taken on an hourly basis during a year leads to $15 \times 24 \times 365 = 131400$ variables in total. Furthermore, since for linear programming all variables must be nonnegative, we have to write some variables as the difference of two nonnegative variables. This increases the number of variables again. However, present day standard packages for linear programming may deal with huge numbers of variables.

In the end, the entire system of equations is rewritten as

$$C_i^b = C_i^{b^+} - C_i^{b^-}, \quad (3.1a)$$

$$Q_i^b = Q_i^{b^+} - Q_i^{b^-}, \quad (3.1b)$$

$$C_i = C_i^p + C_i^a + C_i^{b^+} - C_i^{b^-}, \quad (3.1c)$$

$$Q_i = Q_i^p + Q_i^a + Q_i^{b^+} - Q_i^{b^-} + Q_i^{bl} + Q_i^c, \quad (3.1d)$$

$$q_i^a = Q_i^a - C_i^a, \quad (3.1e)$$

$$H_i^{bc} = H_{i-1}^{bc} - C_{i-1}^{b^+} \Delta t + \Delta t C_{i-1}^{b^-} - \mu_c H_{i-1}^{bc} \Delta t, \quad (3.1f)$$

$$H_i^{bh} = H_{i-1}^{bh} - Q_{i-1}^{b^+} \Delta t + \Delta t Q_{i-1}^{b^-} - \mu_h H_{i-1}^{bh} \Delta t, \quad (3.1g)$$

$$Q_i^p = \gamma C_i^p, \quad (3.1h)$$

$$C_i^p + E_i^p = Q_i^p, \quad (3.1i)$$

$$E_i^p = E_i + E_i^c, \quad (3.1j)$$

$$E_i = E_i^+ - E_i^-, \quad (3.1k)$$

$$\beta G_i^c = E_i^c, \quad (3.1l)$$

$$\alpha G_i^c = Q_i^c. \quad (3.1m)$$

Here, Δt is the time step, usually one hour. Note that the (hourly) heat and cold demands Q_i and C_i are given. The domains of validity and Boolean variables are

$$-\tilde{q}^a \leq q_i^a \leq \tilde{q}^a, \quad (3.2a)$$

$$0 \leq E_i^+ \leq M \delta_{3i}, \quad (3.2b)$$

$$0 \leq E_i^- \leq M(1 - \delta_{3i}), \quad (3.2c)$$

$$0 \leq C_i^{b^+} \leq M \delta_{4i}, \quad (3.2d)$$

$$0 \leq C_i^{b^-} \leq M(1 - \delta_{4i}), \quad (3.2e)$$

$$0 \leq Q_i^{b^+} \leq M \delta_{5i}, \quad (3.2f)$$

$$0 \leq Q_i^{b^-} \leq M(1 - \delta_{5i}), \quad (3.2g)$$

$$Q_i^{bl} \leq \delta_{1i} \tilde{Q}^{bl}, \quad (3.2h)$$

$$G_i^c \leq \delta_{2i} \tilde{G}^c, \quad (3.2i)$$

$$G_i^c \geq 0.5 \delta_{2i} \tilde{G}^c, \quad (3.2j)$$

with $\delta_{1i} = 0, \delta_{2i} = 1, \delta_{3i} = \delta_{4i} = \delta_{5i} = 1$ if $E_i, C_i^{b^+}, Q_i^{b^+} > 0$ and with $\delta_{1i} = 1, \delta_{2i} = 0, \delta_{3i} = \delta_{4i} = \delta_{5i} = 0$ if $E_i, C_i^{b^+}, Q_i^{b^+} = 0$. M is a very large integer.

The cost function (2.17) is discretized in time, over intervals Δt over a period T such that $N \Delta t = T$. It then becomes

$$K = K(\mathbf{Q}) = \sum_{i=1}^N P_i^+ E_i^+ + P_i^- E_i^- + c_1 G_i^c + c_2 Q_i^{bl} + c_3 (Q_i^a + C_i^a) \quad (3.3)$$

with $P_i^+ = f_1$ or f_2 , and $P_i^- = -f_3$ or $-f_4$ for day and night prices of electricity intake or supply from the net. The cost function depends on all variables \mathbf{Q} in the desired period.

The values c_1, c_2 represent the actual costs of gas for the CHP and boiler respectively, while c_3 is what we earlier called λ —the variable cost of using the aquifer.

In summary, the procedure is to minimize cost function (3.3) of the system (3.1) with booleans (3.2) with linear programming (LP).

3.2.1 Possible refinements

The above formulation assumes that we can treat periods (of whatever length) independently, which is probably not entirely true. A few extra constraints came to light in subsequent discussions: for instance, the aquifer cannot work at full capacity after being switched from the cooling to the heating mode or vice versa. A reasonable estimate is that the maximum capacity in such cases would be 75% of its normal value. This necessitates inter-period constraints. For example, in period i we have a new Boolean variable δ_{6i} which is 1 (resp. 0) if the aquifer is in heating (resp. cooling) mode in period i , and constraints

$$Q_i^a \leq \bar{q}^a \delta_{6i} \quad C_i^a \leq \bar{q}^a (1 - \delta_{6i}).$$

This ensures that at most one of Q^a, C^a is positive, and both are bounded from above. Then we have to consider whether the aquifer was in the heating mode in period $i - 1$ or not. If it was, we can use the full capacity, otherwise only 75%. The constraint

$$Q_i^a - 0.25 \delta_{6(i-1)} \bar{q}^a \leq 0.75 \bar{q}^a$$

will model this situation. There is an analogous constraint for the cooling mode.

4 Optimization experience

For the parameters we used the following values [2]

$$\begin{aligned} (\bar{C}, \bar{Q}) &= (70, 43) \text{ W/m}^2, & (C_{max}^{obs}, Q_{max}^{obs}) &= (629, 124) \text{ W/m}^2, \\ \alpha &= 0.501, & \beta &= 0.42, \\ COP &= 4.2, & \delta &= 0.97, \\ c_1 &= 0.0262 \text{ Euro/kWh}, \\ c_2 &= 0.029 \text{ Euro/kWh}, \\ (f_1, f_2, f_3, f_4) &= (0.109, 0.059, 0.085, 0.042) \text{ Euro/kWh}. \end{aligned} \quad (4.1)$$

The optimization program was run with the above values and c_3 running over the values 0.001, 0.01 and 0.1. An initial attempt was made to verify the formulation using the simple LP package LINDO—a “student” version of which is freely downloadable [3]. This version proved sufficient to deal with a day with a known demand profile divided into 3 periods of 8, 10 and 6 hours. With

$c_3 = 0.001$, the LP was poorly scaled and gave rise to numerical problems, but eventually a solution was obtained. Although the actual numerical values changed a little over the range of values used for c_3 , it was encouraging that the overall structure of the different solutions (i.e., which sources were used or ignored) tended to remain fairly stable, and agreed in broad terms with what would have been expected. Nonetheless, it suggested that there is a need for a more accurate estimate of the value of c_3 in order to obtain a better approximation of the optimal solution.

Later it was possible to use another software package—CPLEX, which implements linear and integer programming techniques in a significantly more sophisticated way [4]. For example, it can take care of nonnegative variables implicitly, without the need for the ‘tricks’ used above, thus reducing the number of variables. It is also considerably faster than LINDO. Hence, it was possible to discretize to hourly periods, and to optimize the model over a time horizon of a year. The results were satisfactory, but gave rise to several further problems that suggested the modelling needs some fine-tuning. Although the costs were of the right order of magnitude, they suggested values that were above those currently incurred without optimization! Also in some cases the LP became infeasible after several periods had elapsed. These suggest that the existence of unsuspected inter-period constraints needs to be examined further, and the accuracy of all parameter values needs to be considered again.

5 Conclusions

We conclude that the greenhouse energy optimization problem can be formulated as a linear programming problem including a number of Boolean variables. To find the solution standard techniques can be applied. The computation times strongly depend on the time period over which is optimized. In principle it must be possible to take a full year as optimization period. This allows the application of the boundary conditions that are in force, namely that the net influx in the aquifers averaged over a year must vanish. It would also make it possible to find the optimal strategy for a year with “typical” weather conditions, i.e., averaged over a long period. These insights in turn allow the designers to use “typical” dimensions for the capacities of all components, when they are designing a new greenhouse. Furthermore, since the optimization is very fast for short periods, this project makes clear that optimization of the energy costs in a closed greenhouse can easily be performed using standard software. However, further work needs to be done in terms of understanding and modelling the inter-period connections, and the effect of other assumptions made in the initial formulation.

References

- [1] See the website <http://www.innogrow.nl/>
- [2] Lou Ramaekers, Innogrow, personal communication (2007).
- [3] See the website <http://www.lindo.com/>
- [4] See the website <http://www.ilog.com/products/cplex/>

Understanding the electromagnetic field in an MRI scanner

Jan Bouwe van den Berg^{*} *Nico van den Berg*[†] *Bob van den Bergen*[‡]
Alex Boer[‡] *Fokko van de Bult*[§] *Sander Dahmen*[‡] *Katrijn Frederix*[¶]
Yves van Gennip^{||} *Joost Hulshof*^{*} *Hil Meijer*^{**} *Peter in 't Panhuis*^{||}
Chris Stolk^{**} *Rogier Swierstra*[‡] *Marco Veneroni*^{||} *Erwin Vondenhoff*^{||}

Abstract

In this note, we study the magnetic field pattern in an MRI scanner, in order to ultimately improve the resolution of the image. We model the situation in 2-D, with a simplified model for the patient, consisting of two regions bounded by ellipses with constant dielectric properties. The solution to the Maxwell equations is described in terms of two different bases: Bessel and Mathieu functions. By expansions in Bessel (cylindrical) modes, that are matched at the boundaries, the magnetic field can be computed in a few seconds on a PC or Mac. By optimizing the distribution of antenna currents the homogeneity of the magnetic field can be improved.

Key words: MRI scanner, Maxwell equations, Bessel functions, Mathieu functions

1 Introduction

In an MRI scanner a patient is placed in a strong constant magnetic field. We give a brief heuristic (mixed classical and quantum-mechanical) description of the main physical process. The magnetic field aligns all dipoles (for practical purposes, the spin-1/2 hydrogen nuclei) in the patient's body parallel or anti-parallel to this field. However, this alignment is not perfect and therefore the dipoles

^{*}Vrije Universiteit, Amsterdam

[†]UMC, Utrecht

[‡]Universiteit Utrecht

[§]Universiteit van Amsterdam

[¶]Katholieke Universiteit Leuven, Belgium

^{||}Technische Universiteit Eindhoven

^{**}Technische Universiteit Twente

precess around the radial axis of the field with a typical frequency, called the Larmor frequency, proportional to the size of the external field.

Subsequently, electromagnetic waves are created in the cavity of the scanner (by sending currents through antennas placed inside the MRI scanner), which excite the hydrogen dipoles. The frequency of these waves is chosen to be the Larmor frequency to maximize the number of excitations. The electromagnetic waves temporarily cause some of the dipoles to leave their parallel/anti-parallel state, and when the wave field subsides, these dissidents “fall back” to a parallel/anti-parallel state. This causes the emission of a photon, a scanner detects these photons, and from this information a computer constructs an image of the object or person in the MRI scanner. What is effectively measured here, is the density of dipoles with fixed Larmor frequency — in practice, hydrogen nuclei.

As the strength of the magnetic field increases, so does the Larmor frequency, and hence the frequency of the electromagnetic waves. Advances in superconducting magnet design and in MRI technology have increased the field strength to such an extent that nowadays the wavelength of the electromagnetic waves is of the order of the size of the patient. This causes the field to be significantly altered from the field in an empty MRI scanner if a patient is inserted. Moreover, the homogeneity of the magnetic field is reduced. This has the disadvantage that the received image is distorted: it will show too many hydrogen atoms where the induced field is big and too few where the field is small. Moreover, the induced electric field may lead to significant currents which heat up the patient.

Therefore, the question is how to create an induced electromagnetic field whose magnetic field is homogeneous and whose electric field is small. As constraints we assume the geometry of the MRI scanner (i.e., size of the scanner, size of the patient, location of the antennas etc.) to be given, and we are only allowed to change the phases and amplitudes of the currents through the antennas. In practice, changing the size of the scanner would be prohibitively expensive, and one cannot really change the size of the patients either. However, the antennas could be moved around a bit, but for mathematical simplicity we do not consider this option here.

To answer this question we first have to be able to find a “simple” expression for the induced field generated by the antennas when a given current runs through them. After some introductory remarks about electromagnetic waves in Sections 2 and 3 we present two methods to obtain an approximation of the induced field. In Section 4 we discuss an expression using Bessel functions. This has the advantage that Bessel functions are well-known and many good numerical packages exist for them. In Section 5 we consider an option using Mathieu functions, which are less easy to work with numerically, but fit the geometry of the situation better. Finally, in Section 6, we consider how to optimize the currents, given the fields generated by the individual antennas, such as to maximize the homogeneity of the induced magnetic field.

2 Geometric considerations

The outer cylinder of the MRI scanner itself has, to good approximation, a rotation and translation symmetry along the central axis. We choose coordinates so that the z -direction corresponds to the translation symmetry, and $x = y = 0$ on the central axis. The edge of the scanner will therefore be a circle at radius r (typically $r \approx 35$ cm).

We consider a cross-section of the patient’s abdomen. In this region the patient is modeled by two (confocal) ellipses. The outer one denotes a surrounding fat layer, while the inner one denotes the inside of the patient with organs and muscles and bones. In each layer we consider the electromagnetic properties to be constant; in particular, we use an average of the electromagnetic

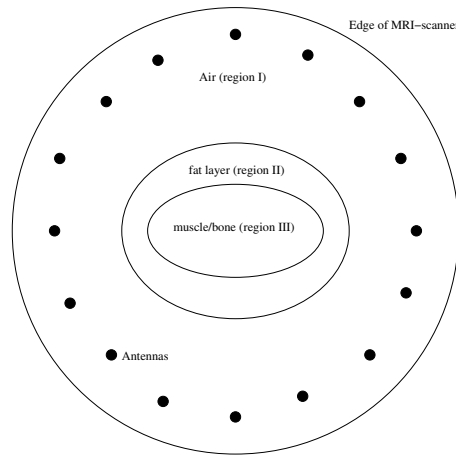


Figure 2.1: Cross-section of the MRI-scanner.

properties of the different tissues in the inside layer. The typical size of the inner ellipse is an outer radius of 15 cm and an eccentricity of 0.85, while the outer ellipse can vary a lot, but is generally not more than 10 cm thick.

Finally, the antennas are modeled as point/line sources located in a circle at a distance of a few centimeters from the edge of the scanner. The currents running through the antennas will have constant frequency, and we assume that the electromagnetic properties of the different layers are time-independent for this fixed frequency.

3 Maxwell equations

We will show that in the case at hand, the Maxwell equations, which describe the electromagnetic field in general, reduce to a single Helmholtz equation in each of the three regions. The theory used is well-known, so we shall go through the derivation rapidly. For references on our notation, consult [1].

We begin by writing down the Maxwell equations for the four fields: the electric field E , the magnetic field B , the so-called displacement field D and the auxiliary field H ¹. The free charge and the current are denoted by ρ and J respectively.

$$\nabla \cdot B = 0, \quad \nabla \times E + \frac{\partial B}{\partial t} = 0, \quad (3.1a)$$

$$\nabla \cdot D = \rho, \quad \nabla \times H - \frac{\partial D}{\partial t} = J. \quad (3.1b)$$

The equations on the first line are known as the *homogeneous Maxwell equations*, and are universally valid. Those on the second line – the *inhomogeneous Maxwell equations* – depend on free charges and currents, and on the different materials in our system.

By assumption, our materials are linear and isotropic, which means that, in each region, D and H are scalar multiples of E and B respectively. We write $D = \epsilon E$ and $H = \mu B$, where the dielectric

¹Although different authors may use different names for these fields, there is consensus about the symbols.

constant ϵ and the magnetic permeability μ are constants of the materials in the different regions. In the situation of the MRI scanner, μ is constant throughout, and equal to μ_0 , the permeability of the vacuum.

Next, we discuss the boundary conditions. Since the conductivity in the inner regions (patient) is much smaller than that of the metal cylinder, the losses in the latter will be small, hence we will assume it to be superconducting. Further, we will neglect the penetration depth of the current in the metal. Let n be a vector normal to the boundary between two media 1 and 2, and let $D_i, E_i, i = 1, 2$, be the displacement and electric fields, respectively, then

$$n \cdot (D_1 - D_2) = \Sigma, \quad n \times E_1 = n \times E_2.$$

Here Σ is the free surface charge between the media. In particular, the *tangential* component of E vanishes on the outer boundary (i.e., the MRI scanner itself). The boundary conditions for B and H are similar, and, since μ is constant, they imply that B is continuous on the entire domain.

The two inhomogeneous Maxwell equations (3.1b) imply the conservation law

$$\frac{\partial \rho}{\partial t} + \nabla \cdot J = 0.$$

In the current term J , we must distinguish between the externally imposed current J_{ext} , and the current J_{ind} induced in the medium by the electric field. We assume that the induced currents are governed by Ohm's law, i.e. $J_{\text{ind}} = \sigma E$, where σ is the conductivity of the medium. We note that different sources approach this in different ways: some include the induced currents in the J , others redefine ϵ , which is then commonly referred to as the complex permittivity. The externally imposed current is assumed to consist of N line sources, located at positions (x_l, y_l) to be specified later:

$$J_{\text{ext}} = J_{\text{ext}}(x, y, t) = \sum_{l=1}^N \begin{pmatrix} 0 \\ 0 \\ I_l \exp(i\omega t) \delta(x - x_l) \delta(y - y_l) \end{pmatrix}, \quad (3.2)$$

where the constants I_l determine the amplitude and the phase of the currents (obviously, in reality all the physical quantities are real, but this complex formalism simplifies the formulas).

The homogeneous Maxwell equations (3.1a) imply the existence of potential functions: a *vector potential* A and a *scalar potential* Φ . They are related to the electromagnetic fields via

$$B = \nabla \times A, \quad E = -\nabla \Phi - \frac{\partial A}{\partial t}. \quad (3.3)$$

We note that A and Φ are not uniquely defined: if (A, Φ) are potentials, then one can check that $(A + \nabla f, \Phi - \frac{\partial f}{\partial t})$ are potentials of the same fields, for any function f . This non-uniqueness, called *gauge-freedom*, can be used to impose some conditions on A and Φ . There are several choices for this, but we will use the so called Lorenz²-gauge:

$$\nabla \cdot A + \frac{1}{c^2} \frac{\partial \Phi}{\partial t} = 0, \quad \text{where} \quad \frac{1}{c^2} = \epsilon \mu.$$

This gauge is chosen so that the inhomogeneous equations would separate if the domain were homogeneous, which our domain is not; nevertheless, we stick with the Lorenz gauge.

²Although often erroneously called Lorentz gauge, supposedly after Hendrik Lorentz, it was in fact Ludwig Lorenz who first published the idea.

Based on the symmetries of the problem, the expression (3.2) for J_{ext} , and physical intuition, we now take the *Ansatz*

$$\Phi = 0, \quad \text{and} \quad A = \begin{pmatrix} 0 \\ 0 \\ A_z(x, y, t) \end{pmatrix}.$$

For notational convenience, we write x for the pair (x, y) , and we replace A_z by A , hence $A_z(x, y, t)$ becomes $A(x, t)$, with $x \in \mathbb{R}^2$. The governing equation for the vector potential is

$$\epsilon\mu \frac{\partial^2 A}{\partial t^2} = \Delta A + \mu J, \quad (3.4)$$

which is to hold globally, i.e., A is continuously differentiable throughout the domain (in particular across the boundaries between the regions), except at the antennas, where singularities occur. As boundary condition we take $A = 0$ on the outer boundary.

In view of the time dependence of the currents through the antennas, we expect waves of fixed frequency ω . This means that all functions under consideration (and in particular A) are a product of a function that depends only on space, and $e^{i\omega t}$. The spatial part of a function will be denoted by the same letter as the function itself, e.g., the electric field is from now on of the form $Ee^{i\omega t}$. We thus replace $\frac{\partial A}{\partial t}$ by $i\omega A$ and $\frac{\partial^2 A}{\partial t^2}$ by $-\omega^2 A$.

Recalling that

$$J_{\text{ind}} = \sigma E = -\frac{\partial A}{\partial t} = -i\omega A$$

and by combining (3.4) and (3.2) we obtain an elliptic (Helmholtz) equation, for $\vec{x} \in \Omega \subset \mathbb{R}^2$,

$$(\Delta + \zeta^2)A = -\sum_{l=1}^N c_l \delta(\vec{x} - \vec{x}_l), \quad A|_{\partial\Omega} = 0, \quad (3.5)$$

for the nonzero component of the vector potential. Here

$$c_l = \mu I_l,$$

and

$$\zeta^2 = \epsilon\mu\omega^2 - i\sigma\omega$$

is (a complex) constant on each of the regions. We will also use the notation $\zeta_k^2 = \epsilon_k\mu_0\omega^2 - i\sigma_k\omega$ for the three regions $k = 1, 2, 3$. Note that to good approximation $\epsilon_1 = \epsilon_0$, the dielectric constant of vacuum/air.

Equation (3.5) indeed has a (unique) solution, from which we can recover the fields B and E using (3.3). It is easily checked that these fields then satisfy the Maxwell equations as well as the boundary conditions, hence they represent a solution to our original problem. In fact, general PDE theory for Maxwell equations implies that it is the unique solution. We are thus left with the task of finding the solution of the Helmholtz equation (3.5).

4 Cylindrical modes

Within each region the coefficients ϵ , σ , and μ are constant. For the constant coefficient PDE, an infinite number of solutions can easily be found by separation of variables. Since the Helmholtz equation (3.5) is linear, with an inhomogeneous right-hand side, we solve the equation for one antenna at a time, and we may restrict our attention to $c_l = 1$. Based on this, the strategy in this section will be as follows:

1. In region I (see Figure 2.1 for the geometry of the various regions), write $A = \tilde{A} + F$, with F a fundamental solution. The function \tilde{A} then satisfies a homogeneous PDE, with inhomogeneous boundary conditions.
2. For each of the regions, find a set of basis functions that satisfy the homogeneous PDE.
3. Write A and \tilde{A} as a finite linear combination of the basis functions in each region. Each linear combination automatically satisfies the PDE. Choose the coefficients such that the boundary conditions are satisfied, or at least as well as possible, since only a finite number of modes can be used in practice.

We will use polar coordinates (even if the domain and the different regions (as well as the patient) are elliptic). Furthermore, we drop the tilde (on A in region I) from the notation.

The fundamental solution $F = F(\vec{x}, \vec{x}_l)$ satisfies

$$(\Delta + \zeta_1^2)F = -\delta(\vec{x} - \vec{x}_l),$$

where $\zeta_1 = \omega \sqrt{\epsilon_0 \mu_0}$ as explained in the previous section. The solution is readily given by

$$F = -\frac{1}{4} Y_0(\zeta_1 |\vec{x} - \vec{x}_l|),$$

where Y_0 is the 0-th order Bessel function of the second kind (notice that F is not uniquely determined since any smooth solution of the homogeneous PDE can be added to it). With the source at the antenna position, we have

$$F = -\frac{1}{4} Y_0(\zeta_1 \rho),$$

where $\rho(r, \theta) = r^2 + R_{\text{ant}}^2 - 2R_{\text{ant}}r \cos(\theta - \theta_{\text{ant}})$ is the distance to the antenna, which is positioned at $(R_{\text{ant}}, \theta_{\text{ant}})$, in polar coordinates.

To find the basis functions we use separation of variables in polar coordinates, which can be found in many textbooks on PDEs. Substituting the *Ansatz* $A(r, \theta) = R(r)\Theta(\theta)$, we find that Θ must satisfy the eigenvalue problem

$$\Theta'' + \lambda\Theta = 0, \quad \Theta \text{ is } \pi\text{-periodic},$$

with solutions $\Theta = e^{in\theta}$, $\lambda = n^2$, $n \in \mathbb{Z}$, while R must satisfy the equation

$$\frac{d^2 R}{dr^2} + \frac{1}{r} \frac{dR}{dr} - \frac{n^2}{r^2} R + \zeta^2 R = 0.$$

This equation has two linearly independent solutions for each n :

$$J_n(\zeta r) \quad \text{and} \quad Y_n(\zeta r),$$

where J_n and Y_n are the Bessel functions of order n of the first and second kind, respectively. The basis functions are therefore $\phi_{k,n}(r)e^{in\theta}$ and $\psi_{k,n}(r)e^{in\theta}$, with

$$\begin{aligned} \phi_{k,n}(r) &= \kappa_{k,n} J_n(\zeta r), \\ \psi_{k,n}(r) &= \tilde{\kappa}_{k,n} Y_n(\zeta r), \end{aligned}$$

where $k = 1, 2, 3$ represents the region, and $\kappa_{k,n}$ and $\tilde{\kappa}_{k,n}$ are normalization constants.

In regions I, II and III we now use the following representation of the solution:

$$\begin{aligned}
A_1(r, \theta) &= F(\rho) + \sum_{j=-n_1}^{n_1} a_{1,j} \phi_{1,j}(r) e^{ij\theta} + \sum_{j=-\tilde{n}_1}^{\tilde{n}_1} b_{1,j} \psi_{1,j}(r) e^{ij\theta} \\
&\stackrel{\text{def}}{=} \Phi_1(r, \theta) \cdot a_1 + \Psi_1(r, \theta) \cdot b_1 \\
A_2(r, \theta) &= \sum_{j=-n_2}^{n_2} a_{2,j} \phi_{2,j}(r) e^{ij\theta} + \sum_{j=-\tilde{n}_2}^{\tilde{n}_2} b_{2,j} \psi_{2,j}(r) e^{ij\theta} \\
&\stackrel{\text{def}}{=} \Phi_2(r, \theta) \cdot a_2 + \Psi_2(r, \theta) \cdot b_2 \\
A_3(r, \theta) &= \sum_{j=-n_3}^{n_3} a_{3,j} \phi_{3,j}(r) e^{ij\theta} \\
&\stackrel{\text{def}}{=} \Phi_3(r, \theta) \cdot a_3,
\end{aligned}$$

where we define a_k and b_k , $k = 1, 2, 3$, to be column vectors containing the $a_{k,j}$ and $b_{k,j}$, and Φ_1 to be a row vector containing the functions $\phi_{1,j}(r) e^{ij\theta}$, and similarly for the other Φ 's and Ψ 's. The ψ 's are omitted in region III, as they would cause an undesirable pole in the solution. The choice of the number of basis functions $n_k, k = 1, 2, 3$ and $\tilde{n}_k, k = 1, 2$ is discussed below.

Using the boundary conditions we try to match the solutions at the boundaries between the different regions and thus obtain the coefficients $a_{k,j}$ and $b_{k,j}$. These conditions are evaluated using a uniform distribution in the angle along the cavity (C) and ellipses (E_1 and E_2). This leads to the following over-determined system

$$\begin{pmatrix}
\Phi_1|_C & \Psi_1|_C & 0 & 0 & 0 \\
\Phi_1|_{E_1} & \Psi_1|_{E_1} & -\Phi_2|_{E_1} & -\Psi_2|_{E_1} & 0 \\
\frac{\partial \Phi_1}{\partial n}|_{E_1} & \frac{\partial \Psi_1}{\partial n}|_{E_1} & -\frac{\partial \Phi_2}{\partial n}|_{E_1} & -\frac{\partial \Psi_2}{\partial n}|_{E_1} & 0 \\
0 & 0 & \Phi_2|_{E_2} & \Psi_2|_{E_2} & -\Phi_3|_{E_2} \\
0 & 0 & \frac{\partial \Phi_2}{\partial n}|_{E_2} & \frac{\partial \Psi_2}{\partial n}|_{E_2} & -\frac{\partial \Phi_3}{\partial n}|_{E_2}
\end{pmatrix}
\begin{pmatrix}
a_1 \\
b_1 \\
a_2 \\
b_2 \\
a_3
\end{pmatrix}
=
\begin{pmatrix}
-F|_C \\
-F|_{E_1} \\
-\frac{\partial F}{\partial n}|_{E_1} \\
0 \\
0
\end{pmatrix}. \quad (4.1)$$

Here $\frac{\partial}{\partial n}$ denotes the derivative in the direction normal to the boundary, and $\Phi_1|_C$ is the matrix with as its rows Φ_1 evaluated at (many) different points in C . To be precise, the elements of the matrix $\Phi_1|_C$ are given by $(\Phi_1|_C)_{j,k} = \phi_{1,-n_1-1+k}(r_j) e^{i(-n_1-1+k)\theta_j}$, in which (r_j, θ_j) are the points along C , and $k = 1, \dots, 2n_1 + 1$. The other blocks in the matrix in (6) are defined similarly. The coefficients a_1, a_2, a_3, b_1, b_2 are obtained by solving the system (4.1) using a least squares approach.

The above algorithm was implemented in Matlab. In Figure 4.1 a plot of the field is given, for the following parameters. We set the outer cylinder to have a radius of 0.34 m, and the semi-axes of the ellipses to be 0.2 m, 0.125 m, and 0.175 m, 0.1 m for the outer and inner ellipse, respectively. The antenna was located at $\theta_{\text{ant}} = 0$, $R_{\text{ant}} = 0.315$ m, with unit current. Furthermore, we set the frequency at the Larmor frequency $\omega = 300$ MHz, and set material constants of $\epsilon_r = 5, \sigma = 0.076$, for the fatty layer, and $\epsilon_r = 5, \sigma = 0.4$ for the interior part of the body; see [5]. The number of the different modes used was (20, 20, 13, 13, 13) for $(n_1, \tilde{n}_1, n_2, \tilde{n}_2, n_3)$. With lower orders very similar pictures were obtained. The computation of the coefficients was done in a few seconds on a PC. Evaluating all the basis functions to compute an image took more time, in the order of a minute depending on the resolution and the number of basis functions involved. The remaining errors in the boundary values were small, up to a few percent of the values of the fundamental solution on the circle and outer ellipse. A second example is shown in Figure 4.2, where we have two antennas, located at $\theta = 3\pi/8$ and $\theta = 11\pi/8$. The field is seen to have a hard time penetrating the body.

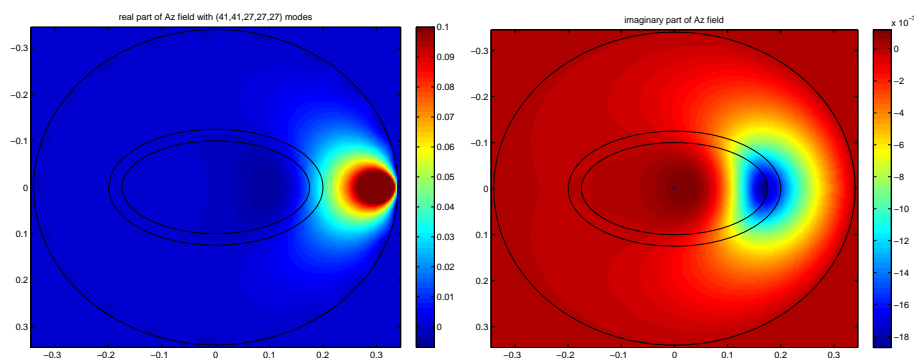


Figure 4.1: Real and imaginary part of the simulated A field

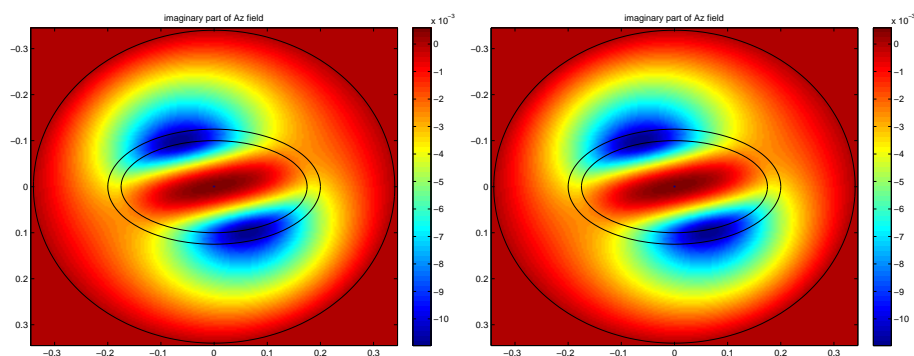


Figure 4.2: Real and imaginary part of the simulated A field, with two antennas.

Two remarks can be made about the numerics. First, the normalization is important, since Bessel functions are badly scaled on the domain of interest. Also, the rows in the system (4.1) that correspond to derivatives are normalized differently from the other rows. Second, the method breaks down because of rank deficiency of the matrix if the number of basis functions is too large. By choosing suitable normalization, quite a large number can be handled, whereas with a poor choice for normalization the rank deficiency occurs already for much smaller numbers of basis functions.

To conclude, the expansion in these cylindrical modes leads to a very fast algorithm. A comparison with results from finite-difference calculations could provide a final check of the results.

5 Mathieu functions

5.1 Elliptic coordinates

Let $\Omega \subset \mathbb{R}^n$ be a bounded open set, we now focus on the Helmholtz equation:

$$\Delta A + \zeta^2 A = 0, \quad \text{in } \Omega, \quad (5.1)$$

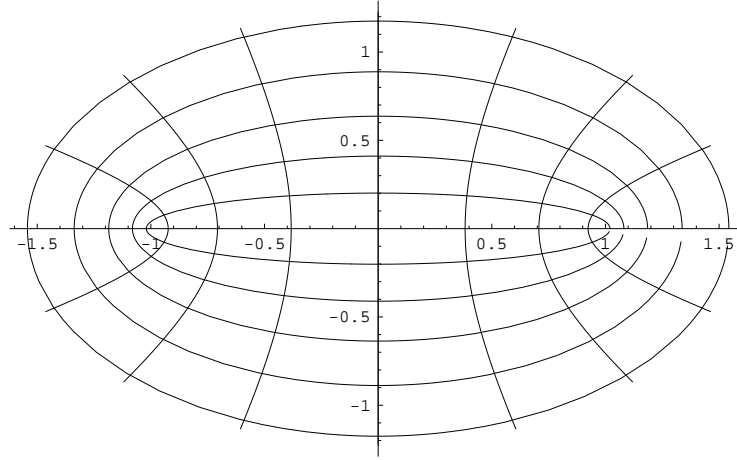


Figure 5.1: Elliptic coordinate grid: the ellipses are curves of constant ξ , the hyperboles curves of constant η .

for $\zeta \in \mathbb{C}$ specified in the previous sections. Since the specific domains we are interested in are elliptical, we introduce the elliptic coordinates, thus simplifying the form of the boundary conditions. We set

$$\begin{aligned} x &= a \cosh \xi \cos \eta, \\ y &= a \sinh \xi \sin \eta, \end{aligned}$$

where $2a$ is the distance between the foci, $\xi \geq 0$ and $\eta \in [-\pi, \pi)$. An impression of an elliptic coordinate grid is given in Figure 5.1.

Note that the ellipticity of the ellipse for given ξ equals $1/\cosh(\xi)$ and hence quickly becomes zero as $\xi \rightarrow \infty$. Equation (5.1) in elliptic coordinates becomes

$$\frac{2}{a^2(\cosh 2\xi - \cos 2\eta)} \left(\frac{\partial^2 A(\xi, \eta)}{\partial \eta^2} + \frac{\partial^2 A(\xi, \eta)}{\partial \xi^2} \right) + \zeta^2 A(\xi, \eta) = 0, \quad \text{in } \Omega. \quad (5.2)$$

Now we use the standard separation of variables technique, i.e., we look for a solution of the particular form $A(\xi, \eta) = X(\xi)Y(\eta)$. Then the partial differential equation (5.2) results in two linear ordinary differential equations:

$$\begin{cases} \frac{X''(\xi)}{X(\xi)} + \frac{1}{2}\zeta^2 a^2 \cosh 2\xi = \lambda, \\ \frac{Y''(\eta)}{Y(\eta)} - \frac{1}{2}\zeta^2 a^2 \cos 2\eta = -\lambda, \end{cases} \quad (5.3)$$

where λ is the separation constant. These equations are known as the *radial* and the *angular* Mathieu equations, respectively (see e.g. [2], [3], [4] and the references therein). Note that the radial equation can be obtained from the angular equation by the substitution $\eta = -i\xi$, hence the solutions of the radial equations are just those of the angular part with imaginary argument.

As any linear second order differential equation these equations have two independent solutions, and we can choose as basis of the solution space an even and an odd solution. Denote these by $C(\lambda, q, \eta)$ (even) and $S(\lambda, q, \eta)$ (odd) respectively, with (traditionally) $q = \frac{1}{4}\zeta^2 a^2$. The angular equation only has 2π -periodic solutions for specific values of λ , that are denoted by the two (q -dependent)

sequences $\lambda_k(q)$ ($k \geq 0$) and $\mu_k(q)$ ($k \geq 1$) respectively³, called the characteristic values. The sequences $\lambda_k(q)$ and $\mu_k(q)$ are defined such that

$$C_k(q, \eta) = C(\lambda_k(q), q, \eta), \quad S_k(q, \eta) = S(\mu_k(q), q, \eta),$$

are the 2π -periodic solutions, continuous in q , with $\lim_{q \rightarrow 0} C_k(q, \eta) = \cos(k\eta)$ and $\lim_{q \rightarrow 0} S_k(q, \eta) = \sin(k\eta)$. For $q = 0$ we have $\lambda_k = \mu_k = k^2$, but for $q \neq 0$ all λ_k and μ_k are generically different. For positive real q we have $\lambda_0 < \mu_1 < \lambda_1 < \mu_2 < \lambda_2 < \mu_3 < \dots$, see [3].

The angular part of our solution $Y(\eta)$ should be 2π -periodic, hence we only retain these C_k 's and S_k 's as solutions for that part. Furthermore, using the eigenvalues λ_k and μ_k one can write down the radial solutions $X(\xi)$. Thus, we find that the system of equations (5.3) has the general solution:

$$A(\xi, \eta) = \sum_k \left(\alpha_k C(\lambda_k(q), q, -i\xi) + \beta_k S(\lambda_k(q), q, -i\xi) \right) C_k(q, \eta) \\ + \left(\gamma_k C(\mu_k(q), q, -i\xi) + \delta_k S(\mu_k(q), q, -i\xi) \right) S_k(q, \eta),$$

where the coefficients $\alpha_k, \beta_k, \gamma_k, \delta_k$ have to be determined in each of the specific domains by matching boundary conditions and imposing the correct symmetry and regularity for A .

In the domain containing the origin we must also ensure that the solution remains continuously differentiable on the line connecting the two focal points. For the even angular solutions, this implies that the derivative at $\xi = 0$ of the radial part must vanish (for it changes sign when passing the line between the focal points). For odd angular solutions, the value of the radial part itself must vanish on this line. More concretely, we lose the $S(\lambda_k, q, -i\xi)C_k(q, \eta)$ and $C(\mu_k, q, -i\xi)S_k(q, \eta)$ solutions on this domain, or alternatively $\beta_k = \gamma_k = 0$ on this domain.

5.2 Matching

It still remains to determine the constants $\alpha_k, \beta_k, \gamma_k$ and δ_k on all domains. Denote the solutions on the different domains by $A_1 + F, A_2, A_3$ for the domains I, II, and III, respectively, and the corresponding constants by $\alpha_{k,j}, \beta_{k,j}$ etc. for $j = 1, 2, 3$.⁴ Here F is the solution in free space for the Helmholtz equation including the sources. We set ξ_1 to be the outer boundary of the scanner, and ξ_2 and ξ_3 to be the radii of the outer (resp. inner) ellipse of the patient. The boundary conditions then become

$$A_1 + F = 0 \quad \text{on } \xi = \xi_1, \quad (5.4a)$$

$$A_2 = A_1 + F \quad \text{on } \xi = \xi_2, \quad (5.4b)$$

$$A_3 = A_2 \quad \text{on } \xi = \xi_3, \quad (5.4c)$$

$$\partial_\xi A_2 = \partial_\xi A_1 + \partial_\xi F \quad \text{on } \xi = \xi_2, \quad (5.4d)$$

$$\partial_\xi A_3 = \partial_\xi A_2 \quad \text{on } \xi = \xi_3, \quad (5.4e)$$

Heuristically, this will give us '5 sets' of conditions that can be used to determine the '10 sets' of constants $\alpha_{j,k}$ ($j = 1, 2, 3$), $\beta_{j,k}$ ($j = 1, 2$) etc..

Concerning condition (5.4a), as written down here we force our solution to vanish at an elliptical boundary, while the actual MRI scanner is circular. However, since the ellipticity of ξ -levels

³In the literature, these sequences are usually called $a_k(q)$ and $b_k(q)$, but in this paper we already use those symbols for different purposes.

⁴The indexing is a bit different from the one in the previous section.

decreases with ξ , the shape of the ellipse for ξ_1 is already close to a circle, and this will provide a rather good approximation. With a little extra work the matching can be performed on a circular interface as well.

The idea of the matching is to take a basis of the angular part of the solution on each interface, write the solutions on both sides in terms of this basis and then match the corresponding coefficients. When we started researching this method we hoped that the angular Mathieu functions in the different regions would give the same basis, so that we would not have to perform any base-change before the actual matching. This would reduce the boundary conditions to equations involving only 4 or 3 parameters each. However, as the functions C_k and S_k do depend on q , we were out of luck. This forces us to perform a basis transformation on the angular part of at least one side for each interface.

We considered two methods. One method is to use as a basis the Mathieu functions $C_k(q, \cdot)$ and $S_k(q, \cdot)$ for the value of q on one of the two sides of the interface. Since these Mathieu functions are orthonormal with respect to the standard L^2 inner product given by

$$\langle f, g \rangle = \frac{1}{\pi} \int_0^{2\pi} f(t)g(t)dt,$$

we can find the expansion of the functions $C_k(q', \cdot)$ and $S_k(q', \cdot)$ in this basis by simply calculating the inner products $\langle C_k(q', \cdot), C_l(q, \cdot) \rangle$ and $\langle S_k(q', \cdot), S_l(q, \cdot) \rangle$ (since arguments of oddness of the integrand show that $\langle C_k(q', \cdot), S_l(q, \cdot) \rangle = 0$). Unfortunately, not much is known about the coefficients obtained in this way.

Another method is to express both sets of solutions in the basis given by cosines and sines. The advantage is that the expansion of the angular Mathieu functions in cosines and sines has been studied previously. The disadvantage is that we now have to perform two basis transformations, which implies we make approximation/numerical errors twice. We will expand on this method further (the analysis of the first method is quite similar).

We need to find an expansion

$$C_k(q, x) = \sum_{j=0}^{\infty} c_j(k, q) \cos(jx),$$

where sines do not occur since $C_k(q, x)$ is even. Indeed, since C_k is either π -periodic or anti-periodic depending on the parity of k , only the $\cos(jx)$ occur where $k-j$ is even. Good algorithms to calculate these coefficients exist. Similarly we want the coefficients in

$$S_k(q, x) = \sum_{j=1}^{\infty} s_j(k, q) \sin(jx),$$

where again $s_j(k, q) = 0$ if $k-j$ is odd.

We remark that most of these coefficients are quite small, namely both $c_j(k, q) = \mathcal{O}\left(\frac{q^{k-j/2}}{k^{k-j/2}}\right)$ and $s_j(k, q) = \mathcal{O}\left(\frac{q^{k-j/2}}{k^{k-j/2}}\right)$ for fixed k , with small constants. For the specific values of q encountered in this problem we can therefore approximate $C_k(q, x)$ very well by $\cos(kx)$ for large enough k (for example $c_{k-2}(k, q) = q/4(k-1) + \mathcal{O}\left(\frac{q^3}{k^3}\right)$ if $k > 2$).

Matching the solution on each interface can now be done by equating the coefficients of each $\cos(jx)$ and $\sin(jx)$. For example, for the continuity on the interface between the second and third

layer we obtain

$$\begin{aligned} & \sum_k [\alpha_{k,2} C(\lambda_k(q_2), q_2, -i\xi_3) + \beta_{k,2} S(\lambda_k(q_2), q_2, -i\xi_3)] c_j(k, q_2) \\ &= \sum_k \alpha_{k,3} C(\lambda_k(q_3), q_3, -i\xi_3) c_j(k, q_3), \end{aligned}$$

as a relation between the coefficients of $\cos(jx)$ on both sides of the interface (recall that $\beta_{k,3} = 0$). Similarly the continuous radial differentiability at the interface gives us the relation

$$\begin{aligned} & \sum_k [\alpha_{k,2} C'(\lambda_k(q_2), q_2, -i\xi_3) + \beta_{k,2} S'(\lambda_k(q_2), q_2, -i\xi_3)] c_j(k, q_2) \\ &= \sum_k \alpha_{k,3} C'(\lambda_k(q_3), q_3, -i\xi_3) c_j(k, q_3), \end{aligned}$$

where the C' and S' are the derivatives of the radial solution.

In order to take F into account we will have to express that solution also in terms of cosines and sines on the interfaces. Since the solution F is generally not immediately given in terms of elliptical coordinates, there is no simple formula expanding this function in terms of cosines and sines. However we can always numerically calculate the Fourier coefficients to express

$$F(\xi_2, \eta) = \sum_j \psi_{j,2c} \cos(j\eta) + \psi_{j,2s} \sin(j\eta),$$

and a similar equation on the outer boundary.

If the assumption that the circular outer boundary can be approximated by an ellipse of sufficiently low eccentricity fails we can also use a similar calculation to express the basis functions of the solution on region I in terms of Fourier coefficients on an actual circle. However, calculating these coefficients involves (numerically) calculating many integrals for each basis function, and since each integral involves the slightly intractable Mathieu functions this could become computationally expensive.

Note that the resulting system of equations splits in four systems of equations. Indeed, we have one set of equations relating the $\alpha_{k,i}$ and $\beta_{k,i}$ for even k , one set for $\alpha_{k,i}$ and $\beta_{k,i}$ with odd k , one set for $\gamma_{k,i}$ and $\delta_{k,i}$ with even k and one set for $\gamma_{k,i}$ and $\delta_{k,i}$ with odd k . These sets of equations correspond to solutions which are even/odd in η (i.e. symmetrical or anti-symmetrical with respect to reflection in the horizontal axis) and π periodic/anti-periodic (i.e. symmetrical or anti-symmetrical with respect to the vertical axis). Since all sets are very similar we will focus on the first one.

5.3 Approximation

In order to obtain a finite set of equations we only consider a small number K of modes. This means we set $\alpha_{k,i} = 0$ and $\beta_{k,i} = 0$ for $k \geq K$. Note that here we only consider the equations between the $\alpha_{k,i}$ and $\beta_{k,i}$ with k even (as announced in the previous section), so in particular we forget about $\gamma_{k,i}$ and $\delta_{k,i}$. To obtain a system with the right amount of equations we moreover only consider the equations related to the coefficients of $\cos(jx)$ with $j \leq K$. Indeed, in the equations of the coefficient of $\cos(jx)$ for $j > K$ the terms with $\alpha_{j,i}$ are very dominant since $c_j(j, q_i) \approx 1$, while $c_j(k, q_i) \ll 1$ for $k \neq j$ and $j > K$, so including that equation without including the $\alpha_{j,i}$ term would probably lead to very bad results.

We now have to solve the system of equations

$$\begin{pmatrix} A_1(\xi_3) & 0 & 0 \\ A_1(\xi_2) & -A_2(\xi_2) & 0 \\ 0 & A_2(\xi_3) & -A_3(\xi_3) \\ A_1'(\xi_2) & -A_2'(\xi_3) & 0 \\ 0 & A_2'(\xi_3) & -A_3'(\xi_3) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \beta_1 \\ \alpha_2 \\ \beta_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} -F(\xi_3) \\ -F(\xi_2) \\ 0 \\ -F'(\xi_2) \\ 0 \end{pmatrix}.$$

Here $A_m(\xi_n)$ denotes the matrix

$$A_m(\xi_n) = \{C(\lambda_k(q_m), q_m, -i\xi_n)c_j(k, q_m), S(\lambda_k(q_m), q_m, -i\xi_n)c_j(k, q_m)\}_{0 \leq k, j \leq K, k, j \text{ even}}$$

and similarly for $A_m(\xi_n)'$ (containing the derivatives of C and S). Moreover

$$\alpha_n = (\alpha_{n,k})_{0 \leq k \leq K, k \text{ even}},$$

and finally

$$-F(\xi_n) = (-\psi_{j,nc})_{0 \leq j \leq K, j \text{ even}}.$$

The problem thus involves solving a system of $5[(k+1)/2]$ equations in as many variables. Considering the matrix is nearly sparse (i.e. involves many small terms as $c_l(k, q)$ is small for $|l-k| > 0$) this should be feasible, but we have not implemented it.

A numerical problem might occur since the columns in the matrix corresponding to $\alpha_{1,k}$ and $\beta_{1,k}$ are nearly identical, so the matrix becomes almost singular and likewise for $\alpha_{2,k}$ and $\beta_{2,k}$. The problem is that $C(\lambda_k(q), q, -i\xi)$ and $S(\lambda_k(q), q, -i\xi)$ (the basis of the solutions to the radial Mathieu equation) are very similar for large ξ . Indeed they are the generalizations of the hyperbolic cosines and sines, which both behave as $\exp(x)/2$ for large x . While $C(\lambda_k(q), q, -i\xi)$ and $S(\lambda_k(q), q, -i\xi)$ do not behave like $\exp(\xi)/2$ for large ξ , they still are very similar. In order to find a less near-singular matrix it would therefore be good to find a different basis of the radial solution to the Mathieu equations. Indeed the arguments given above apply to any basis of solutions (in the radial part), so the method would not need to change.

One convenient basis would be the one which intuitively is associated to $\exp(x)$ and $\exp(-x)$, namely $C(\lambda_k(q), q, -i\xi) + S(\lambda_k(q), q, -i\xi)$ and $C(\lambda_k(q), q, -i\xi) - S(\lambda_k(q), q, -i\xi)$. Unfortunately we have been unable to find any good algorithms to actually calculate these functions (other than calculating C and S and taking their difference, which does not behave well numerically as the difference of these two functions is much smaller than their values itself).

6 Optimization

We continue our semi-explicit approach to the problem and consider here the task of making the field in the patient as uniform as possible. We combine the separation-of-variables method in polar coordinates (i.e., using Bessel functions) from Section 4, with the handling of the boundary conditions between regions from Section 5. In particular, we use a discrete Fourier transform technique to match the solutions in the different regions at their common boundaries. Furthermore, we refrain from “normalizing” the Bessel functions (as was done in Section 4). Instead, we precondition the matrix that governs the matching of the Fourier modes, which reduces rank deficiencies (mentioned in Sections 4 and 5).

More precisely, for each of the N antennas, uniformly distributed on a circle of radius R_{ant} , we solve the Helmholtz equation with source term (3.5) using $2M + 1$ angular modes in the Fourier-Bessel expansion in each region (and correspondingly also $2M + 1$ modes in the Fourier expansion for the matching conditions at the boundaries), i.e.,

$$A = \sum_{m=-M}^M (a_{k,m} J_m(\zeta_k r) + b_{k,m} Y_m(\zeta_k r)) e^{im\theta},$$

for the three regions $k = 1, 2, 3$. We will only consider optimization of the field in (a subregion of) the core of the patient (region III). There, only Bessel functions of the first kind contribute ($b_{3,m} = 0$), and for the j^{th} antenna we denote the coefficients $a_{3,m}$ from now on by a_m^j . In region III, the expression for the potential then becomes

$$A = \sum_{l=1}^N c_l \sum_{m=-M}^M a_m^l J_m(\zeta r) e^{im\theta}, \quad (6.1)$$

where $\zeta = \zeta_3$, and the complex amplitudes $c_l = \mu_0 I_l$ appear as complex control parameters.

The magnetic field corresponding to A is given by

$$\mathbf{B} = \begin{pmatrix} B^x \\ B^y \\ 0 \end{pmatrix} = \nabla \times \begin{pmatrix} 0 \\ 0 \\ A \end{pmatrix} = \begin{pmatrix} \frac{\partial A}{\partial y} \\ -\frac{\partial A}{\partial x} \\ 0 \end{pmatrix}.$$

The part of the field that we are interested in is $B^+ = B^x + iB^y$, since this is the (polarized) combination that turns the spins. This field, induced by the antennas, is often called B_1^+ to distinguish it from the much bigger constant field B_0 in the axial direction (generated by the superconducting magnet). We are thus interested in

$$B^+ = B^x + iB^y = \frac{\partial A}{\partial y} - i \frac{\partial A}{\partial x} = e^{i\theta} \left(\frac{1}{r} \frac{\partial A}{\partial \theta} - i \frac{\partial A}{\partial r} \right).$$

Using the identities

$$\begin{aligned} J_{m-1}(r) + J_{m+1}(r) &= \frac{2m}{r} J_m(r), & \text{and} \\ J_{m-1}(r) - J_{m+1}(r) &= 2 \frac{dJ_m}{dr}(r), \end{aligned}$$

for Bessel functions, the expression (6.1) for the potential implies that the approximation of B^+ in the inner region III is

$$B^+ = \sum_{l=1}^N c_l \sum_{m=-M}^M i \zeta a_m^l J_{m+1}(\zeta r) e^{i(m+1)\theta}. \quad (6.2)$$

We have attempted to find those values of c_l for which B^+ , rather than $|B^+|$, is as uniformly distributed as possible. We make this choice because this problem is *much* easier to solve and still leads to reasonably uniform $|B^+|$. The reason it is easier is that minimizing the variation in B^+ , as formulated below, is in essence a least square problem, i.e., a *linear* algebra problem, while minimizing the variation in $|B^+|$ is a fully *nonlinear* optimization problem. Even if one eventually would like to optimize $|B^+|$, it would not be a bad idea to start that optimization procedure from

the easily computable configuration that optimizes B^+ . Let us also remark that, to keep the method tractable, we did not consider the issue of minimizing the electric field.

Recalling that $\overline{f_{\Omega} B^+} = \int_{\Omega} B^+ / \int_{\Omega} 1$ is the average of B^+ on Ω , a simple and apparently satisfying approach is to use

$$\frac{\overline{f_{\Omega} |B^+ - \overline{f_{\Omega} B^+}|^2}}{|\overline{f_{\Omega} B^+}|^2}$$

as a measure for the variation in B^+ . We can rewrite this as

$$\frac{\overline{f_{\Omega} |B^+ - \overline{f_{\Omega} B^+}|^2}}{|\overline{f_{\Omega} B^+}|^2} = \frac{\overline{f_{\Omega} |B^+|^2}}{|\overline{f_{\Omega} B^+}|^2} - 1.$$

Since this expression is clearly invariant under scalings of B^+ , and since $\int_{\Omega} 1$ is just the (fixed) measure of Ω , one may reformulate the problem as finding the minimizer of

$$\min \left\{ \int_{\Omega} |B^+|^2 : \overline{f_{\Omega} B^+} = 1 \right\},$$

i.e., minimization is over all c_l such that $\overline{f_{\Omega} B^+} = 1$.

We first consider the case where the optimization domain Ω is a disk D of radius ρ around the origin, where ρ is sufficiently small, so that the domain lies entirely in region III. This choice of domain reduces the integral formulas considerably (we will review the general case below). In view of (6.2) the constraint becomes

$$\overline{f_{\Omega} B^+} = \frac{2}{\rho^2} \sum_{l=1}^N c_l a_{-1}^l i \zeta \int_0^{\rho} J_0(\zeta r) r dr = 1.$$

The expression to be minimized reduces to (complex conjugation denoted by a star)

$$\begin{aligned} \int_D |B^+|^2 &= \sum_{l,k,m,n} c_l c_k^* a_m^l a_n^{k*} \zeta \zeta^* \int_D J_{m+1}(\zeta r) J_{n+1}^*(\zeta r) e^{i(m+1)\theta} e^{-i(n+1)\theta} \\ &= \sum_{l,k,m} c_l c_k^* a_m^l a_m^{k*} \zeta \zeta^* 2\pi \int_0^{\rho} |J_{m+1}(\zeta r)|^2 r dr \\ &= \sum_{m=-M}^M \left| \sum_{l=1}^N c_l a_m^l \zeta \sqrt{2\pi \int_0^{\rho} |J_{m+1}(\zeta r)|^2 r dr} \right|^2. \end{aligned}$$

To ease notation we introduce

$$q_{ml} = a_m^l \zeta \sqrt{2\pi \int_0^{\rho} |J_{m+1}(\zeta r)|^2 r dr},$$

and the matrix $Q = (q_{ml})$, as well as the vector $c = (c_l)$. Then

$$\int_D |B^+|^2 = \sum_m \left| \sum_l c_l q_{ml} \right|^2 = |Qc|^2.$$

To write the constraint in linear algebra terms as well, define the vector $p = (p_l)$ with

$$p_l = \frac{2}{\rho^2} a_{-1}^l i \zeta \int_0^\rho J_0(\zeta r) r dr.$$

Then we may reformulate the problem as finding the least-square solution of $Qc = 0$ under the constraint $p^T c = 1$.

With a final reformulating step the constraint can be absorbed into the matrix, namely define

$$\tilde{Q} = \begin{pmatrix} p^T \\ Q \end{pmatrix},$$

and let e_1 be the standard unit vector. Find the least-square solution of $\tilde{Q}x = e_1$, say $x = \tilde{c} = (\tilde{c}_l)$, then, using the linearity of the problem, it follows that the optimal $c = (c_l)$ of the constraint problem above is given by a rescaled version of \tilde{c} , namely

$$c_l = \frac{1}{p^T \tilde{c}} \tilde{c}_l.$$

For general domains Ω this can be generalized as follows. Let us describe the method for a two-dimensional integral using polar coordinates, but it is straightforward to extend. We first want to discretize the integral. Let Ω lie inside some (large) disk D_R , and let Δr and $\Delta \theta$ be discretization step sizes, so that $n_1 = \frac{R}{\Delta r}$ and $n_2 = \frac{2\pi}{\Delta \theta}$ are integers. The grid points are now given by $r^{j_1} = (j_1 - \frac{1}{2})\Delta r$ and $\theta^{j_2} = j_2\Delta \theta$ for $1 \leq j_1 \leq n_1, 1 \leq j_2 \leq n_2$. Let I be an enumeration of all the grid points that lie inside Ω , and $1 \leq i \leq N_\Omega$ (with $N_\Omega \leq n_1 n_2$) indexes I , i.e., $(r_i, \theta_i) \in \Omega$.⁵ Then

$$\int_\Omega f(r, \theta) \approx \Delta r \Delta \theta \sum_{i=1}^{N_\Omega} f(r_i, \theta_i) r_i,$$

where the final “weight” r_i is due to the polar-coordinate Jacobian. With this discretization in place, we may write, with $\delta = \Delta r \Delta \theta$

$$\int_\Omega |B^+|^2 \approx \sum_{l,k,m,n,i} c_l c_k^* a_m^l a_n^{k*} |\zeta|^2 \delta J_{m+1}(\zeta r_i) J_{n+1}^*(\zeta r_i) r_i e^{i(m+1)\theta_i} e^{-i(n+1)\theta_i}.$$

To simplify notation we introduce

$$g_{im} = J_{m+1}(\zeta r_i) e^{i(m+1)\theta_i} \sqrt{r_i \delta},$$

and the matrix $G = (g_{im})$, as well as $h_{ml} = a_m^l \zeta$ and $H = (h_{ml})$. Then the above expression reduces to

$$\int_\Omega |B^+|^2 \approx \sum_{l,k,m,n,i} c_l c_k^* h_{ml} h_{nk}^* g_{im} g_{in}^* = \sum_i \left| \sum_{m,l} g_{im} h_{ml} c_l \right|^2 = |GHc|^2.$$

The remainder of the argument is now analogous (with $Q = GH$) to the case $\Omega = D$.

⁵A formal description is as follows: let $I = \{(j_1, j_2) | (r_{j_1}, \theta_{j_2}) \in \Omega\}$, and N_Ω is the number of elements in the set I . Then there are “enumeration” functions $\tilde{j}_1(i)$ and $\tilde{j}_2(i)$ such that $I = \{(\tilde{j}_1(i), \tilde{j}_2(i)) | 1 \leq i \leq N_\Omega\}$. Now set $r_i = r^{\tilde{j}_1(i)}$ and $\theta_i = \theta^{\tilde{j}_2(i)}$.

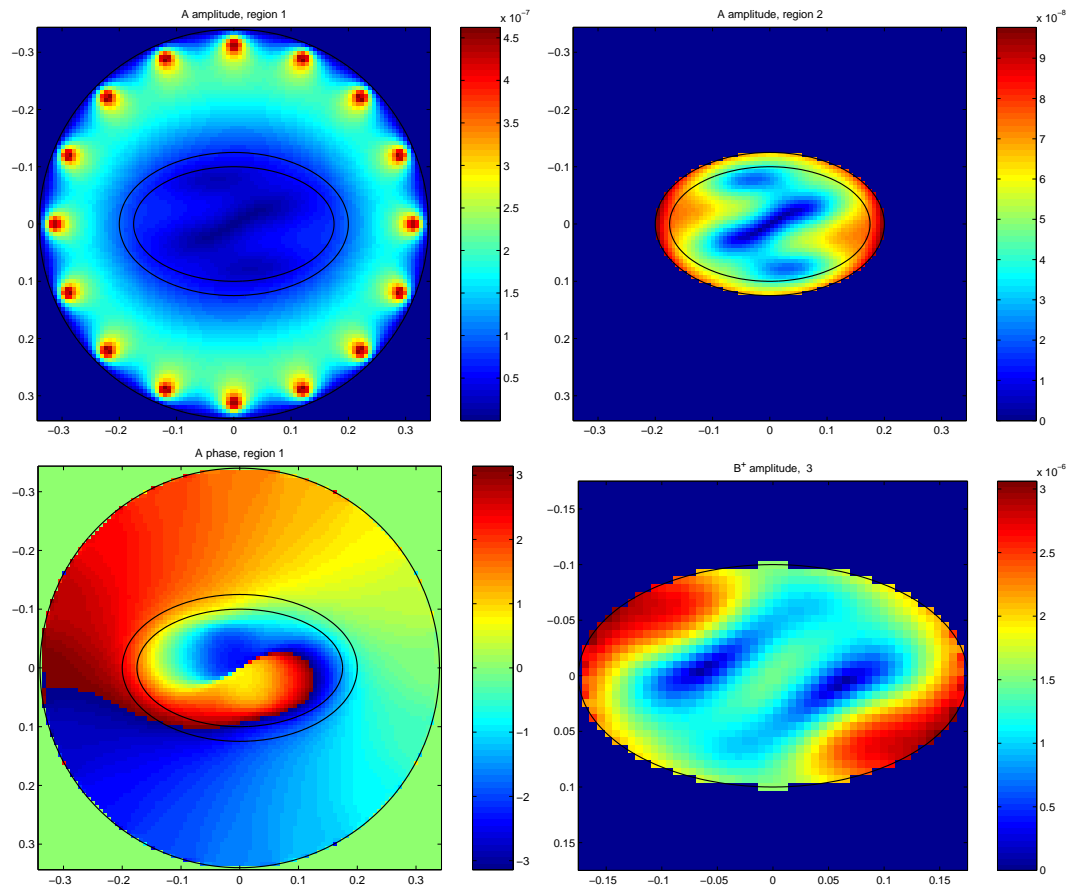


Figure 6.1: The amplitude and phase of the fields A and B^+ in the different regions. The antenna currents are *not* optimized.

6.1 Results

The computational parameters in the simulation were chosen as follows: 65 Fourier-Bessel modes are used in each region (for both types of Bessel functions), and 2^8 discretization points on each boundary. The optimization computation in this case only takes a few seconds. For the physical parameters we used realistic values provided by the problem presenters: $\epsilon_2 = 10\epsilon_0$ and $\sigma_2 = 0.076$ for the fatty layer, and $\epsilon_3 = 34\epsilon_0$ and $\sigma_3 = 0.4$ for the interior part of the body. The parameters that determine the geometry of the MRI scanner are the same as in Section 4. Note that since the problem is linear the absolute size of the fields is fairly irrelevant (although it is of course important in practice), since it can be tuned by an arbitrary multiplicative constant.

For comparison, we first look at the non-optimized fields in Figure 6.1. There the amplitudes of the currents in the antennas are all equal and the phase is rotated uniformly ($I_l = I_1 e^{2\pi i(l-1)/16}$). We indeed see the phase of A nicely rotating, while the field does not penetrate the body very well. Looking at the crucial B^+ field, we see that its amplitude is nonuniform and very small in certain central parts of the body.

In Figure 6.2 we have optimized (in the sense explained above) the B^+ field in a disk of radius

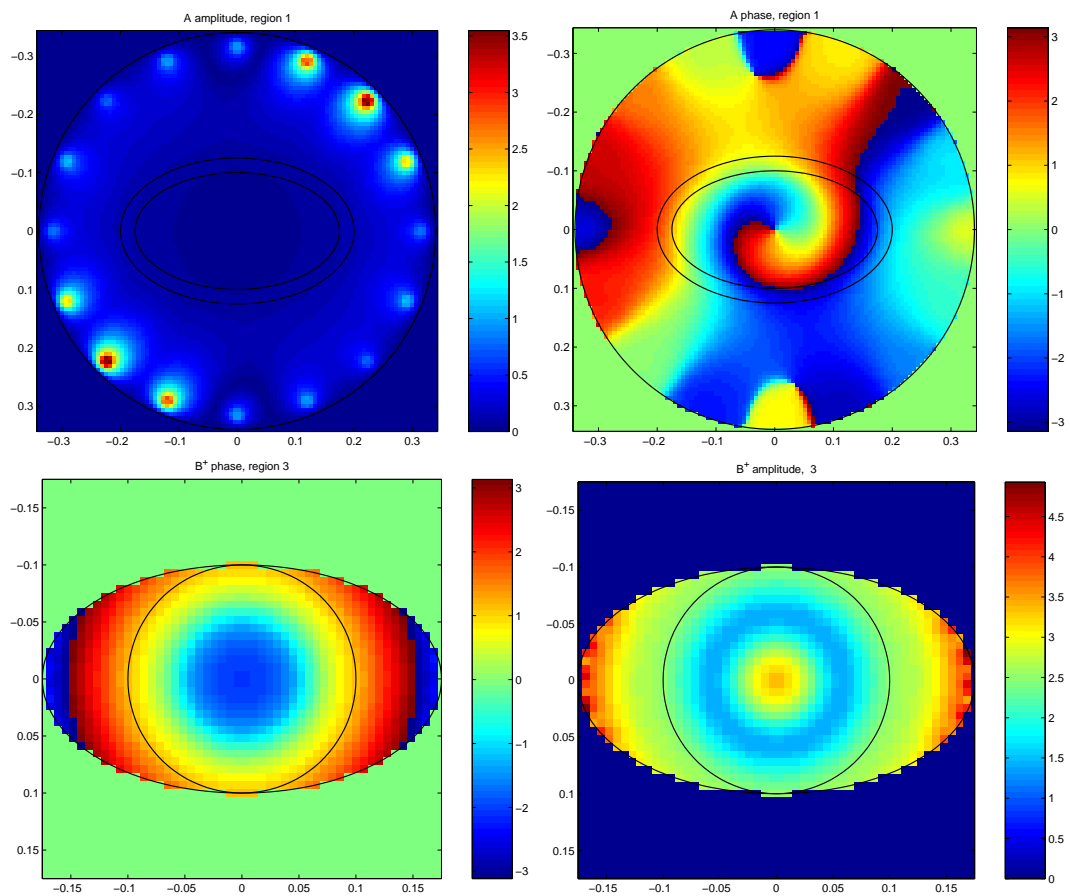


Figure 6.2: The amplitude and phase of the fields A and B^+ in the different regions. The antenna currents are optimized as to make B^+ optimally uniform in the indicated disk inside the inner ellipse.

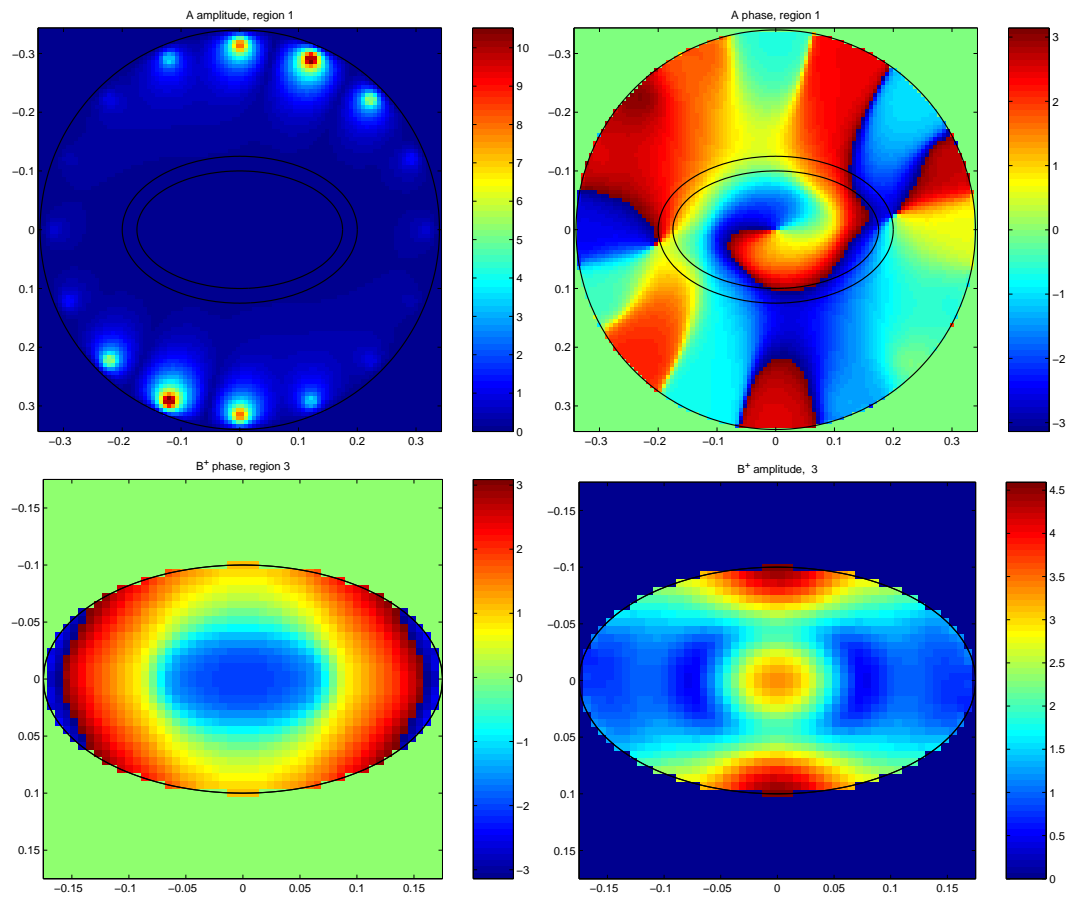


Figure 6.3: The amplitude and phase of the fields A and B^+ in the different regions. The antenna currents are optimized as to make B^+ optimally uniform in the entire inner ellipse.

0.1 m around the origin (the optimization region is indicated in the figure). We see that although our method makes B^+ optimally uniform, in fact the amplitude $|B^+|$ is much more uniform than the phase. This is a bit unexpected, but it is very welcome in view of the original aim of making this amplitude uniform, and we are pleasantly surprised by how uniform the amplitude of the field is: the fluctuations are within a factor 2. When we look at the antennas, we see quite a spread in current amplitudes, and the complicated phase pattern illustrates the subtleness of the optimal configuration.

Next we optimize the B^+ field on the entire inner ellipse (the interesting not-fat part of the body). In Figure 6.3 we see that the amplitude of the resulting field is less uniform, which is to be expected since we are trying to make it uniform on a bigger domain, but the results are still much better than the non-optimized case in Figure 6.1.

Finally, we optimize on a domain whose size is in between the disk and the entire ellipse. The results are depicted in Figure 6.4, where also the domain of optimization is indicated. The results show a fairly uniform $|B^+|$: within a factor 3 over the entire ellipse representing the inner region of the body. This demonstrates that the relatively simple optimization procedure performs satisfactorily even on non-circular domains. We note however that perhaps the optimization in Figure 6.2 is

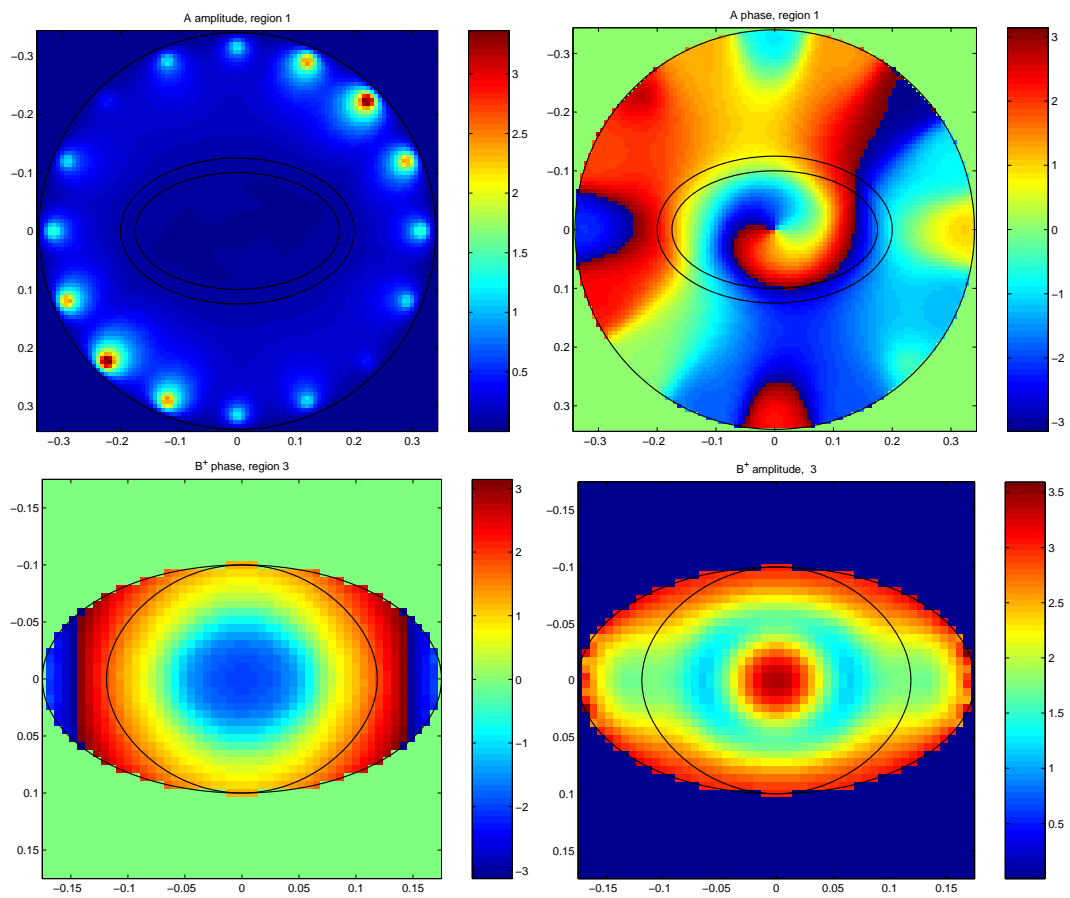


Figure 6.4: The amplitude and phase of the fields A and B^+ in the different regions. The antenna currents are optimized as to make B^+ optimally uniform in the indicated domain.

preferable. Furthermore, the pattern in the (amplitude and phase of) antenna currents is quite similar to Figure 6.2.

7 Concluding remarks

After completing the report we asked Ir. Bob van den Bergen and Dr. Ir. Nico van den Berg of the Department of Radiotherapy of the University Medical Center in Utrecht to write the concluding remarks on the results obtained for the problem they submitted to the study group:

The UMC Utrecht problem consisted of finding a semi-analytical method to calculate and optimize rapidly the radiofrequency (RF) field of an MRI scanner. According to the UMC Utrecht this goal has been fully achieved. The developed model allows an evaluation of the full electromagnetic field in less than a minute. This enables an on-the-fly optimization procedure for patients. At the moment we are designing the RF hardware to implement this procedure for our 7 Tesla MRI scanner.

As an extra bonus we obtained a rapid optimization method which is based on a simple least squares method in stead of the conventionally applied non-linear optimization procedures which suffer from lengthy calculation times and local minima. Currently, we are using this to study the ultimate RF homogeneity as a function of various physical parameters. Furthermore, the short computation time opens up the new possibility to find the optimal coil geometry in an automated fashion. Concluding, we can state that the SWI workshop has been a great success for the UMC Utrecht and has resulted in much new research.

The UMC Utrecht would like to express their gratitude for being able to take part in this workshop. It is quite unique that such mathematical talent and knowledge is brought together to solve such a complex modelling problem in the medical industry. The rigorous mathematical methodology of the participants applied to this physical problem has been an eye-opener. On a more personal level, we would like to thank to all the participants for their work and great character. Enlighten all these mathematical heathens out there!

Acknowledgments

There are plenty of people to thank for their help. First of all we thank the organizers for providing a comfortable atmosphere and an inspiring program. We also thank Nico van den Berg and Bob van den Bergen (from the UMC Utrecht) for proposing this stimulating problem, and for their ceaseless patience and good humor. We also thank experts Hans Duistermaat and Ernst van Faassen, for their comments on PDEs and Maxwell theory, respectively. Finally, the participants thank each other for the enthusiastic cooperation and good spirit which characterized the week.

References

- [1] J. D. Jackson. Classical electrodynamics, 3rd ed. John Wiley & Sons, Inc, 1999.
- [2] E. Mathieu. Mémoire sur le mouvement vibratoire d'une membrane de forme elliptique. *Jour. de Math. Pures et Appliquées*, 13:137–203, 1868.
- [3] J. C. Gutiérrez Vega. Theory and numerical analysis of the Mathieu functions. *Lecture notes*, Monterrey, 2003.

- [4] J. C. Gutiérrez Vega, R. M. Rodríguez-Dagnino, M. A. Meneses-Nava, S. Chávez-Cerda. Mathieu functions, a visual approach. *Am. J. Phys.*, 71(3): 233–242, 2003.
- [5] <http://www.fcc.gov/cgi-bin/dielec.sh>

The ING problem: a problem from the financial industry

*Cornelis W. Oosterlee**

In the 2007 Mathematics with Industry workshop, ING posed a challenging problem from financial mathematics. For a system of stochastic differential equations, representing an advanced model for asset prices, the question was whether a closed, or semi-closed, form of a pricing formula for call options could be derived. The asset price model of interest was the so-called hybrid Heston–Hull–White model.

The industrial interest comes from the fact that valuing and risk-managing derivatives demands fast and accurate prices. As the financial models used in practice are becoming increasingly complex, efficient solution methods have to be developed to cope with such models. Needless to say that working with a closed form solution is highly efficient.

The basis of modern option pricing theory is found in the famous Black–Scholes model, which itself is based on a one-factor stochastic model for asset prices,

$$dS_t = rS_t dt + \sqrt{v}S_t dW_t.$$

Here S_t denotes asset price, and W_t denotes Brownian motion. Interest rate r and ‘volatility’ \sqrt{v} are assumed to be constant in this model which is a major model simplification. Based on this model derivatives, like options, can be priced highly efficiently.

The motivation behind using more general processes is the simple fact that the Black–Scholes model is not able to reproduce important empirical features of asset returns and at the same time provide a reasonable fit to the so-called implied volatility surfaces observed in option markets. Over the past few years it has been shown that several models that incorporate stochastic volatility are, at least to some extent, able to reproduce the volatility skew or smile. The particular model we will consider here is a more advanced form of the well-known Heston stochastic volatility model. The model is a generalization as also the interest rate is modeled by a stochastic differential equation. The hybrid Heston–Hull–White asset price model reads:

$$\begin{cases} dS_t &= r_t S_t dt + \sqrt{v_t} S_t dW_{1,t}, \\ dv_t &= \kappa(\eta - v_t) dt + \lambda \sqrt{v_t} dW_{2,t}, \\ dr_t &= (\theta(t) - ar_t) dt + \sigma dW_{3,t} \end{cases}$$

for $0 \leq t \leq T$ with T the maturity of the option. Here S_t , v_t , r_t denote the random variables asset price, its variance and interest rate, respectively, at time $t \geq 0$. The model constitutes an extension of the well-known Black–Scholes model as the volatility and the interest rate both evolve randomly over time. The quantities κ , η , λ , a , σ are positive real constants, that can be calibrated to market data. Furthermore, $\theta(t)$ is a deterministic, continuous, positive function of time which can be chosen

*CWI, Amsterdam, and Technische Universiteit Delft, editor ING problem, c.w.oosterlee@cwi.nl

as to match the so-called term structure of interest rates. Finally, $W_{1,t}$, $W_{2,t}$, $W_{3,t}$ denote Brownian motions with a positive covariance matrix

$$\text{var}_{\mathbb{P}}(\tilde{\mathbf{W}}_t) := \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix} t.$$

By means of the risk-neutral valuation formula, the price of any European option can be written as an expectation of the discounted payoff of this option. Starting from this representation one can apply several techniques to calculate the price itself. Broadly speaking one can distinguish two types of methods: Solution of the corresponding partial differential equation (PDE) or stochastic differential equation (SDE) by integration. Both solution approaches may rely on techniques from numerical mathematics, including Monte Carlo simulation, in particular when pricing early exercise options or complex option contracts.

Quite a few mathematicians took up this ING challenge, and during the week three subgroups were formed, each approaching the problem from a different side. A particular challenge here was that some Dutch professors in financial mathematics in earlier attempts were not able to come up with a closed form option pricing solution for this particular model. It is therefore no surprise that the problem in its full generality could not be solved within the workshop week. However, three high-quality approaches with interesting insights are presented hereafter that allow for different dependency structures. We believe that based on the results in the contributions the pricing of options under the dynamics of the complete hybrid Heston–Hull–White model, such as by classical Monte Carlo simulation, can be significantly accelerated¹

On behalf of the group participants we would like to thank in particular Dr. Antoine van der Ploeg from ING for his detailed technical note on the problem and for his assistance during the workshop.

¹We would also like to point to work by N. Kunitomo and Y-J Kim, which can be found at <http://www.e.u-tokyo.ac.jp/~kunitomo/Effects.pdf>, which contains interesting aspects for our problem, but which we did not study during the workshop.

Three approaches to extend the Heston model

*Michael Muskulus**

1 Introduction

The stock price in the Heston model [8] is given by the following stochastic differential equation

$$dS_t = rS_t dt + \sqrt{v_t} S_t dW_{1,t}, \quad S_0 > 0,$$

where $r > 0$ denotes the risk-free interest rate, which is assumed to be constant in time. Since S_t follows a geometric Brownian motion, it is advantageous to consider $X_t = \ln S_t$ instead. By the Itô–Doëblin formula one then has

$$dX_t = d \ln S_t = \left(r - \frac{1}{2}v_t\right)dt + \sqrt{v_t}dW_{1,t}.$$

The volatility of the instantaneous stock returns dS_t/S_t follows the process

$$dv_t = \kappa(\eta - v_t)dt + \lambda \sqrt{v_t}dW_{2,t}, \quad v_0 > 0,$$

in which $\kappa > 0$ determines the speed of adjustment of the volatility towards its theoretical mean $\eta > 0$, and $\lambda > 0$ is the second-order volatility, i.e., the variance of the volatility. Note that this has exactly the form as the Cox–Ingersoll–Ross (CIR) [6] interest rate process.

The money-market account evolves according to the ordinary differential equation $dM_t = rM_t dt$ with solution $M_t = M_0 e^{rt}$. The importance of the Heston model comes from the fact that it allows for a semi-analytical solution in terms of characteristic functions (see Section 3).

2 Extension of the Heston model

Although the Heston model incorporates stochastic volatility, the fixed interest rate is an unrealistic assumption. Let us therefore consider (following [14]) a generalized Hull–White process [9] for the interest rate,

$$dr_t = (\theta_t - ar_t)dt + \sigma dW_{3,t},$$

where $\theta_t > 0$, $t \in \mathbb{R}$, is a nonconstant drift term. Usually, stock rate, volatility, and interest rate are correlated; a phenomenon known as the leverage effect [2, 3]. Assume that

$$dW_{i,t}dW_{j,t} = \rho_{ij}dt,$$

*Universiteit Leiden, muskulus@math.leidenuniv.nl

*We would like to thank the other participants of our group: Joris Bierkens, Fang Fang, Karel in 't Hout, David Kan, Coen Leentvaar, Kees Oosterlee.

where

$$C = (\rho_{ij})_{1 \leq i, j \leq 3} = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{21} & 1 & \rho_{23} \\ \rho_{31} & \rho_{32} & 1 \end{pmatrix}$$

is a constant¹ covariance matrix, and therefore positive semi-definite. In fact, for the application in finance, we can assume that C is nonsingular².

From the spectral theorem of linear algebra we see that C , being positive definite and symmetric, has a unique matrix square root $A = (a_{ij})_{1 \leq i, j \leq 3}$, such that

$$C = U\Sigma U^t = (U\Sigma^{1/2})(U\Sigma^{1/2})^t = AA^t, \quad (2.1)$$

where $U\Sigma U^t$ is the singular-value decomposition of C . Explicitly, we have

$$\sum_{k=1}^3 a_{ik}a_{jk} = \rho_{ij}, \quad \text{for all } i, j = 1, 2, 3.$$

There now exist adapted, independent Brownian motions $B_{i,t}$, $i = 1, 2, 3$, such that

$$dW_{i,t} = \sum_{j=1}^3 a_{ij} dB_{j,t},$$

and the general model we consider here is the following:

$$dS_t = r_t S_t dt + \sqrt{v_t} S_t a_{1i} dB_{i,t} \quad \text{or} \quad dX_t = (r_t - \frac{1}{2}v_t)dt + \sqrt{v_t} a_{1i} dB_{i,t} \quad (2.2)$$

$$dv_t = \kappa(\eta - v_t)dt + \lambda \sqrt{v_t} a_{2j} dB_{j,t} \quad (2.3)$$

$$dr_t = (\theta_t - ar_t)dt + \sigma a_{3k} dB_{k,t}, \quad (2.4)$$

where the Einstein convention for summation of repeated indices is used. The money market account develops according to

$$M_t = M_0 \exp\left(\int_0^t r_s ds\right).$$

In this generality, the model is probably not solvable (semi-) analytically. Therefore three different constraints, arising from different strategies are discussed that lead to partial solutions.

3 Independent interest process

The first simplification is to assume that the interest rate process r_t evolves independently from the stock price and volatility processes S_t and v_t , keeping the correlation between the latter two,

$$\begin{aligned} dW_{1,t}dW_{2,t} &= \rho dt \\ dW_{1,t}dW_{3,t} &= dW_{2,t}dW_{3,t} = 0. \end{aligned}$$

¹The decomposition of correlated Brownian motions into independent ones we are about to describe is also possible if $C = C(t)$ is an adapted process in time.

²This is possible since we will never have a perfectly linear relation between the driving Brownian motions of stock price, volatility, and interest rate — this would be rather contradictory to the assumption of stochasticity, and in such a case we could do with a simpler model than the one considered.

The first relation can be rewritten³ as

$$dW_{1,t} = \rho dW_{2,t} + \sqrt{1 - \rho^2} dW'_{2,t},$$

where $W'_{2,t}$ is another Brownian motion, independent of $W_{2,t}$.

Define the integrated interest $R_t = \int_0^t r_t dt$. We want to find the European call option price at maturity time T , given an initial stock price S_0 , volatility v_0 and interest rate r_0 (and initial time $t = 0$),

$$\begin{aligned} C_T(S_0, v_0, r_0) &= \mathbb{E}[e^{-R_T} (S_T - K)^+ | S_0, v_0, r_0] \\ &= \mathbb{E}[e^{-R_T} S_T \cdot 1_{(\ln S_T > \ln K)}] - K \mathbb{E}[e^{-R_T} \cdot 1_{(\ln S_T > \ln K)}] \\ &= \mathbb{E}[e^{-R_T} S_T] \frac{\mathbb{E}[e^{-R_T} S_T \cdot 1_{(\ln S_T > \ln K)}]}{\mathbb{E}[e^{-R_T} S_T]} - K \mathbb{E}[e^{-R_T}] \frac{\mathbb{E}[e^{-R_T} \cdot 1_{(\ln S_T > \ln K)}]}{\mathbb{E}[e^{-R_T}]}, \end{aligned}$$

where $x^+ = \max(0, x)$ denotes the positive part of x , and 1_A is the indicator function of the event A . Note that under the risk-neutral measure the process $(e^{-R_t} S_t)_{t \geq 0}$ is a martingale, such that $\mathbb{E}[e^{-R_t} S_t] = S_0$.

Define an (analytic) function

$$\Psi(z) = \mathbb{E}[e^{-R_T + z \ln S_T}], \quad z \in \mathbb{C},$$

such that

$$\Psi(0) = \mathbb{E}[e^{-R_T}] = P(r_0, T)$$

is the discount price function, i.e., the price of a zero-coupon bond at time T .

Consider now the two (scaled) characteristic functions

$$\begin{aligned} \Phi_1(z) &= \frac{\Psi(1 + iz)}{\Psi(1)} = \frac{\mathbb{E}[e^{-R_T} S_T e^{iz \ln S_T}]}{\mathbb{E}[e^{-R_T} S_T]} \\ \Phi_2(z) &= \frac{\Psi(iz)}{\Psi(0)} = \frac{\mathbb{E}[e^{-R_T} e^{iz \ln S_T}]}{\mathbb{E}[e^{-R_T}]} \end{aligned}$$

for two distribution functions F_1, F_2 .

The particular form of these functions is a consequence of the generalized Bayes theorem [12, pg. 231] for conditional expectations, when we require

$$C_T(S_0, v_0, r_0) = S_0 \int_{\ln K}^{\infty} dF_1(x) - KP(r_0, T) \int_{\ln K}^{\infty} dF_2(x). \quad (3.1)$$

Fourier inversion⁴ then allows to numerically evaluate the probability distributions [4], such that

³This is nothing else than the two-dimensional analogue of the matrix square root decomposition, Eq. (2.1).

⁴The inversion formula goes back to Gurland [7], who showed that

$$F(x) + F(x - 0) = 1 - \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{e^{-iux} \Phi(u)}{iu} du,$$

where the integral has to be interpreted as a Cauchy principal value. For (left-) continuous $F(x)$ this reduces to

$$P(X \leq x) = F(x) = \frac{1}{2} + \frac{1}{2\pi} \int_0^{\infty} \frac{\Phi(-u)e^{iux} - \Phi(u)e^{-iux}}{iu} du,$$

such that

$$P(X \geq \ln K) = 1 - F(\ln K) = \frac{1}{2} - \frac{1}{2\pi} \int_0^{\infty} \frac{\Phi(-u)e^{iu \ln K} - \Phi(u)e^{-iu \ln K}}{iu} du.$$

the option pricing function at time t is

$$C_{T-t}(S_t, v_t, r_t) = S_t \left(\frac{1}{2} - \frac{1}{2\pi} \int_0^\infty \frac{\Phi_1(-u)e^{iu \ln K} - \Phi_1(u)e^{-iu \ln K}}{iu} du \right) - KP(r_t, T-t) \left(\frac{1}{2} - \frac{1}{2\pi} \int_0^\infty \frac{\Phi_2(-u)e^{iu \ln K} - \Phi_2(u)e^{-iu \ln K}}{iu} du \right).$$

The remaining work is to find an expression for $\Psi(z)$. This method is due to Scott [11], and we just follow his calculations (see Appendix for details), to arrive at

$$\Psi(z) = e^{-z(v_0 + \kappa\eta T)} \cdot \mathbb{E}[e^{(z-1)R_T}] \cdot \mathbb{E}[e^{wV_T + z\frac{\rho}{\lambda}v_T}], \quad (3.2)$$

where we used the integrated volatility $V_t = \int_0^t v_t dt$, and

$$w = (z-1)z\frac{1}{2}(1-\rho^2) + z\left(\frac{\rho}{\lambda}\kappa - \frac{1}{2}\rho^2\right).$$

From the theory of Bessel bridges [10, 5] we have the following closed form for the second expectation:

$$\mathbb{E}[e^{-s_1 V_T - s_2 v_T} | v_0] = e^{a_T - b_T v_0}, \quad \text{Re } s_i \geq 0, \quad i = 1, 2,$$

where

$$a_T = 2\kappa\eta \cdot \ln \frac{2\gamma e^{\frac{1}{2}(\kappa-\gamma)T}}{2\gamma e^{-\gamma T} + (\kappa + \gamma + s_2)(1 - e^{-\gamma T})}$$

$$b_T = \frac{(1 - e^{-\gamma T})(2s_1 - \kappa s_2) + \gamma s_2(1 + e^{-\gamma T})}{2\gamma e^{-\gamma T} + (\kappa + \gamma + s_2)(1 - e^{-\gamma T})},$$

and $\gamma = \sqrt{\kappa^2 + 2s_1}$. The parameters κ and η are taken from the volatility process:

$$dv_t = \kappa(\eta - v_t)dt + \lambda\sqrt{v_t}dW_{2,t}, \quad v_0 > 0. \quad (3.3)$$

This almost solves the problem, since we still need to find an expression for the first expectation in Eq. (3.2).

If we now replace⁵ the generalized Hull–White interest rate process with a CIR type interest process,

$$dr_t = (\theta - ar_t)dt + \sigma\sqrt{r_t}dW_{3,t},$$

then this is also of the above form (3.3) (replacing κ by a , and η by θ/a), giving us a semi-analytical solution.

4 Constrained correlations

We now present an alternative method. Consider the model (2.2-2.4) again.

The change of variable $S_t = \exp(X_t)$ leads to $G_t(X_t, \cdot, \cdot, \cdot) = C_t(S_t, \cdot, \cdot, \cdot)$, such that $(e^{-R_t}G_t)$ is a martingale (under the appropriate, equivalent risk-neutral measure). Following the strategy of the

⁵In fact, it should be possible to arrive at a similar expression for the (standard) Hull–White interest rate process, too, by following the lines of the proof of above formula in [10, 5]. This is one possible direction for future research.

multi-dimensional Feynman-Kac theorem for independent Brownian motions [13], we expand the differential $d(e^{-R_t}G_t)$ in dt and $dB_{i,t}$ terms ($i = 1, 2, 3$), and set the dt term equal to zero, leading⁶ to the following PDE:

$$\begin{aligned} r_t G_t = & \frac{\partial G_t}{\partial t} + (r_t - \frac{1}{2}v_t) \frac{\partial G_t}{\partial X_t} + \kappa(\eta - v_t) \frac{\partial G_t}{\partial v_t} + (\theta_t - ar_t) \frac{\partial G_t}{\partial r_t} \\ & + \frac{1}{2}v_t \frac{\partial^2 G_t}{\partial X_t^2} + \frac{1}{2}\lambda^2 v_t \frac{\partial^2 G_t}{\partial v_t^2} + \frac{1}{2}\sigma^2 \frac{\partial^2 G_t}{\partial r_t^2} \\ & + \lambda\rho_{12}v_t \frac{\partial^2 G_t}{\partial X_t \partial v_t} + \sigma\rho_{13} \sqrt{v_t} \frac{\partial^2 G_t}{\partial X_t \partial r_t} + \lambda\sigma\rho_{23} \sqrt{v_t} \frac{\partial^2 G_t}{\partial v_t \partial r_t}. \end{aligned}$$

The ansatz⁷

$$G_t = e^{A(T-t)+v_t B(T-t)+r_t C(T-t)+\sqrt{v_t} D(T-t)+iuX_t}$$

now gives⁸ the following system of equations:

$$\begin{aligned} \frac{dA}{dt} &= \theta_t C(t) + \frac{1}{2}\sigma^2 C^2(t) + \kappa\eta B(t) + \frac{1}{2}\lambda\sigma\rho_{23} D(t)C(t) + \frac{1}{8}\lambda^2 D^2(t) \\ \frac{dB}{dt} &= -\frac{iu}{2} - \frac{u^2}{2} - \kappa B(t) + \lambda\rho_{12}iuB(t) + \frac{1}{2}\lambda^2 B^2(t) \\ \frac{dC}{dt} &= iu - aC(t) \\ \frac{dD}{dt} &= iu\sigma\rho_{13}C(t) - \frac{1}{2}\kappa D(t) + iu\frac{1}{2}\lambda\rho_{12}D(t) + \lambda\sigma\rho_{23}B(t)C(t) + \frac{1}{2}\lambda^2 B(t)D(t) \\ 0 &= 8D(t)(4\kappa\eta - \lambda^2) \end{aligned}$$

which is a system of ODEs, either (i) if we set

$$\lambda = 2\sqrt{\kappa\eta} \quad (\text{Forced volatility variance}),$$

or (ii) if we set $D(t) = 0$. The latter is possible, if we let $B(t) = -iu\frac{\rho_{13}}{\rho_{23}}\frac{1}{\lambda}$, which gives us two constraints on the parameters (from $\frac{dB}{dt} = 0$):

$$\rho_{23} = \frac{2\kappa}{\lambda}\rho_{13}, \quad \rho_{12} = \frac{4\kappa^2 + \lambda^2}{4\kappa\lambda} \quad (\text{Forced volatility correlation}).$$

In this case, the equation in $A(t)$ can be integrated easily, since $C(t)$ is readily available,

$$C(t) = \frac{iu}{a} (e^{at} - 1), \quad \text{when } C(0) = 0.$$

Furthermore, if θ_t is assumed constant, the solution is given analytically by the characteristic function of G_t , as in the solution of the Heston model.

⁶Note that $a_{ik}dB_{k,t} \cdot a_{jl}dB_{l,t} = a_{ik}a_{jl}\delta_{kl}dt = a_{ik}a_{jk}dt = \rho_{ij}dt$, where δ_{kl} is the Kronecker delta.

⁷Which fulfills the necessary boundary condition $G_T = e^{iuX_T}$, given the initial conditions $A(0) = B(0) = C(0) = D(0) = 0$.

⁸Use that

$$\begin{aligned} \frac{\partial G_t}{\partial v_t} &= G_t \left[B(t) + \frac{1}{2\sqrt{v_t}} D(t) \right] \\ \frac{\partial^2 G_t}{\partial v_t^2} &= G_t \left[B^2(t) + \frac{B(t)D(t)}{\sqrt{v_t}} + \frac{1}{4v_t} D^2(t) - \frac{1}{4(v_t)^{3/2}} D(t) \right]. \end{aligned}$$

5 Volatility-interest coupling

The third method discussed considers an interest rate process that is coupled⁹ to the volatility, via

$$dr_t = (\theta_t - ar_t)dt + \sigma \sqrt{v_t} a_{3k} dB_{k,t}.$$

The Feynman–Kac partial differential equation for the martingale ($e^{-R_t} G_t$) then reads

$$\begin{aligned} r_t G_t = & \frac{\partial G_t}{\partial t} + (r_t - \frac{1}{2}v_t) \frac{\partial G_t}{\partial X_t} + \kappa(\eta - v_t) \frac{\partial G_t}{\partial v_t} + (\theta_t - ar_t) \frac{\partial G_t}{\partial r_t} \\ & + \frac{1}{2}v_t \frac{\partial^2 G_t}{\partial X_t^2} + \frac{1}{2}\lambda^2 v_t \frac{\partial^2 G_t}{\partial v_t^2} + \frac{1}{2}\sigma^2 v_t \frac{\partial^2 G_t}{\partial r_t^2} \\ & + \lambda v_t \frac{\partial^2 G_t}{\partial X_t \partial v_t} \rho_{12} + \sigma v_t \frac{\partial^2 G_t}{\partial X_t \partial r_t} \rho_{13} + \lambda \sigma v_t \frac{\partial^2 G_t}{\partial v_t \partial r_t} \rho_{23}. \end{aligned}$$

Following Heston, we make a similar ansatz for the characteristic function:

$$G_t = e^{A(T-t) + B(T-t)v_t + C(T-t)r_t + iuX_t}.$$

Grouping together terms with v_t , respectively r_t , we get the following system of ordinary differential equations,

$$\begin{aligned} \frac{dA}{dt} &= \kappa\eta B(t) + \theta_t C(t) \\ \frac{dB}{dt} &= b_0 + b_1 B(t) + \frac{1}{2}\lambda^2 B(t)^2 + \frac{1}{2}\sigma^2 C(t)^2 \\ &\quad + \lambda\sigma\rho_{23} B(t)C(t) + iu\lambda\rho_{12} C(t) \\ \frac{dC}{dt} &= (iu - 1) + aC(t) \end{aligned}$$

where $b_0 = -\frac{1}{2}iu(1 - iu)$, and $b_1 = iu\sigma\rho_{13} - \kappa$.

The initial conditions are $A(0) = B(0) = C(0) = 0$, and the last equation has solution:

$$C(t) = \frac{1 - iu}{a} (e^{-at} - 1).$$

The second equation is a Riccati equation of form

$$\frac{dB(t)}{dt} = \frac{1}{2}\lambda^2 B(t)^2 + g(t)B(t) + h(t)$$

with coefficient functions

$$\begin{aligned} g(t) &= g_0 + g_1 e^{-at} \\ h(t) &= h_0 + h_1 e^{-at} + h_2 e^{-2at} \end{aligned}$$

⁹The form of this coupling is only motivated by the mathematical structure. In fact, whether this coupling is of any value in the modelling of real-world finance, is quite unclear, though one might expect it not to be.

where, setting $q = (1 - iu)$,

$$\begin{aligned} g_0 &= iu\sigma\rho_{13} - \kappa - \lambda\sigma\frac{q}{a}\rho_{23}, & h_0 &= -\frac{1}{2}iuq + \frac{q^2\sigma^2}{2a^2} - iu\lambda\frac{q}{a}\rho_{12} \\ g_1 &= \lambda\sigma\frac{q}{a}\rho_{23}, & h_1 &= iu\lambda\frac{q}{a}\rho_{12} - \frac{q^2\sigma^2}{a^2} \\ & & h_2 &= \frac{q^2\sigma^2}{2a^2}. \end{aligned}$$

Although the quadratic term $B(t)^2$ makes it impossible to split this equation into real and imaginary parts, there exists¹⁰ an analytical solution of this equation in terms of Whittaker functions [1], such that it can be evaluated efficiently with tabulated values. Yet the equation for $A(t)$ makes it necessary to solve the whole system numerically. Still, this is more efficient than integration of the partial differential equation or direct (Monte-Carlo) simulation, and makes this approach also interesting.

6 Discussion

In this short note we have discussed three different ways of obtaining efficient solutions to extensions of the Heston model. Unfortunately, the page limitation in this contribution does not allow for numerical experiments with these methods.

A The method of Scott

Write

$$\begin{aligned} \ln S_t &= \int_0^t r_s ds + \int_0^t \sqrt{v_s} \left(\rho dW_{2,s} + \sqrt{1 - \rho^2} dW'_{2,s} \right) - \frac{1}{2} \int_0^t v_s ds \\ &= R_t + \left(\rho \int_0^t \sqrt{v_s} dW_{2,s} - \frac{1}{2} \rho^2 \int_0^t v_s ds \right) \\ &\quad + \left(\sqrt{1 - \rho^2} \int_0^t \sqrt{v_s} dW'_{2,s} - \frac{1}{2} (1 - \rho^2) \int_0^t v_s ds \right) \\ &= R_t + \eta_t + \xi_t. \end{aligned}$$

Since v_s develops independently from $dW'_{2,s}$, we can calculate

$$\begin{aligned} \mathbb{E}[\xi_t | W_{2,t}] &= -\frac{1}{2}(1 - \rho^2)V_t, \\ \text{Var}[\xi_t | W_{2,t}] &= (1 - \rho^2)V_t, \end{aligned}$$

where $V_t = \int_0^t v_s ds$.

Furthermore, we now can use $\sqrt{v_t} dW_{2,t} = \frac{1}{\lambda}(dv_t - \kappa(\eta - v_t)dt)$ to write

$$\eta_t = \frac{\rho}{\lambda}(v_t - v_0 - \kappa\eta t + \kappa V_t) - \frac{1}{2}\rho^2 V_t.$$

¹⁰The commercial software package MAPLE can be used to derive the analytical solution of this ODE.

Considering $\Psi(z) = \mathbb{E} \left[e^{-R_T + z \ln S_T} \right]$, we see that $\Psi(z) = \mathbb{E} \left[e^{(z-1)R_T} \right] \cdot \mathbb{E} \left[e^{z\xi_T + \eta_T} \right]$. Now ξ_T , being an Itô integral, is normally distributed. Therefore $e^{z\xi_T}$ has a log-normal distribution, such that

$$\mathbb{E}[e^{z\xi_T} \mid W_{2,t}] = e^{(z-1)z\frac{1}{2}(1-\rho^2)V_t} \quad (\text{conditional on } W_{2,t})$$

and we arrive at the formula given in the text, Eq. (3.2).

References

- [1] G. E. Andrews, R. A. Askey, and R. Roy. *Special Functions*, volume 71 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 2001.
- [2] F. Black. Studies of stock price volatility changes. In *Proceedings of the 1976 Meetings of the Business and Economic Statistics Section*, pages 177–181. American Statistical Association, 1976.
- [3] J.-P. Bouchaud, A. Matacz, and M. Potters. Leverage effect in financial markets: The retarded volatility model. *Phys. Rev. Lett.*, 87(22):228701–228704, 2001.
- [4] Peter Carr and Dilip B. Madan. Option valuation using the Fast Fourier Transform. *Journal of Computational Finance*, 2(4):61–73, 1998.
- [5] Marc Chesney, Robert J. Elliott, and Rajna Gibson. Analytical solutions for the pricing of American bond and yield options. *Mathematical Finance*, 3(3):277–294, 1993.
- [6] J. C. Cox, J. E. Ingersoll, and S. A. Ross. An intertemporal general equilibrium model of asset prices. *Econometrica*, 53(2):363–384, 1985.
- [7] J. Gurland. Inversion formulae for the distribution of ratios. *The Annals of Mathematical Statistics*, 19(2):228–237, 1948.
- [8] S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343, 1993.
- [9] J. C. Hull and A. White. Pricing interest-rate derivative securities. *Rev. Finan. Stud.*, 3(4):573–592, 1990.
- [10] J. Pitman and M. Yor. A decomposition of Bessel bridges. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 59:425–457, 1982.
- [11] L. O. Scott. Pricing stock options in a jump-diffusion model with stochastic volatility and interest rates: Applications of Fourier inversion methods. *Mathematical Finance*, 7(4):413–424, 1997.
- [12] A. N. Shiryaev. *Probability*, volume 95 of *GTM*. Springer, second edition, 1996.
- [13] S. E. Shreve. *Stochastic Calculus for Finance II: Continuous-Time Models*. Springer, 2004.
- [14] A.P. C. van der Ploeg and J. Tistaert. Improving an equity option pricing model for ING: Deriving a call option pricing formula in the hybrid Heston–Hull–White model. Technical report, ING Corporate Market Risk Management, 2007.

A semi closed-form analytic pricing formula for call options in a hybrid Heston–Hull–White model

Karel in 't Hout^{*} *Joris Bierkens*[†] *Antoine P.C. van der Ploeg*[‡]
Jos in 't Panhuis[§]

1 Introduction

We consider the valuation of European call options under the general Heston–Hull–White asset pricing model. The model constitutes an extension of the well-known Black–Scholes model [3] where the volatility and the interest rate both evolve randomly over time. The process for the variance v_t has been proposed by Heston [5]. The process for the interest rate r_t was formulated by Hull and White [6] and forms a generalization of the Vasicek model [8]. In this contribution we assume that the process $W_{3,t}$ is independent from $W_{1,t}$ and $W_{2,t}$. The two Brownian motions $W_{1,t}$, $W_{2,t}$ are allowed to be correlated; their correlation is denoted by $\rho \in [-1, 1]$.

The purpose of this note is to derive an analytic pricing formula in semi closed-form for European call options under the Heston–Hull–White asset pricing model. The availability of such a pricing formula is particularly useful in a calibration procedure. In practice, option pricing models are calibrated to a large number of market-observed call option prices. It is important that such a parameter estimation procedure is fast. Therefore a (near) closed-form call option pricing formula is very desirable.

Our analysis in this note follows the lines of Heston [5]. The formula that we obtain forms a direct extension of Heston's pricing formula for call options, which can quickly be evaluated.

2 A semi closed-form analytic formula for call option prices

Let $C(t, s, v, r)$ denote the price of a European call option at time $t \in [0, T]$ given that at this time the asset price equals s , its variance equals v and the interest rate equals r .

From standard no-arbitrage arguments it follows that C satisfies the parabolic partial differential equation (PDE)

$$\begin{aligned} 0 = & \frac{\partial C}{\partial t} + \frac{1}{2}s^2v\frac{\partial^2 C}{\partial s^2} + \frac{1}{2}\lambda^2v\frac{\partial^2 C}{\partial v^2} + \frac{1}{2}\sigma^2\frac{\partial^2 C}{\partial r^2} + \rho\lambda sv\frac{\partial^2 C}{\partial s\partial v} \\ & + rs\frac{\partial C}{\partial s} + \kappa(\eta - v)\frac{\partial C}{\partial v} + (\theta(t) - ar)\frac{\partial C}{\partial r} - rC, \end{aligned} \tag{2.1}$$

^{*}Universiteit Antwerpen, Belgium, karel.inthout@ua.ac.be

[†]Universiteit Leiden, joris@math.leidenuniv.nl

[‡]ING Corporate Market Risk Management/Quants, Amsterdam, antoine.van.der.Ploeg@ingbank.com

[§]Technische Universiteit Eindhoven, j.c.h.w.panhuis@tue.nl

^{*}Other participants: Claude Archer, Chun Dong Chau, Zhengwei Han, Remco van der Hofstad, David Kun

for $0 \leq t < T$, $s > 0$, $v > 0$, $-\infty < r < \infty$. This PDE can be viewed as a time-dependent advection–diffusion–reaction equation on an unbounded, three-dimensional spatial domain. The payoff of a call option yields the terminal condition

$$C(T, s, v, r) = \max(0, s - K), \quad (2.2)$$

where $K > 0$ is the strike price of the call option. Further, a boundary condition at $s = 0$ holds,

$$C(t, 0, v, r) = 0 \quad (0 \leq t < T). \quad (2.3)$$

We note that at $v = 0$ no condition is specified.

It is convenient to first apply a change of variables. Define

$$\hat{C}(t, x, v, r) = C(t, e^x, v, r). \quad (2.4)$$

Then \hat{C} satisfies the PDE

$$\begin{aligned} 0 = & \frac{\partial \hat{C}}{\partial t} + \frac{1}{2}v \frac{\partial^2 \hat{C}}{\partial x^2} + \frac{1}{2}\lambda^2 v \frac{\partial^2 \hat{C}}{\partial v^2} + \frac{1}{2}\sigma^2 \frac{\partial^2 \hat{C}}{\partial r^2} + \rho\lambda v \frac{\partial^2 \hat{C}}{\partial x \partial v} \\ & + (r - \frac{1}{2}v) \frac{\partial \hat{C}}{\partial x} + \kappa(\eta - v) \frac{\partial \hat{C}}{\partial v} + (\theta(t) - ar) \frac{\partial \hat{C}}{\partial r} - r\hat{C} \end{aligned} \quad (2.5)$$

for $0 \leq t < T$ on the spatial domain $(x, v, r) \in \mathbb{R} \times (0, \infty) \times \mathbb{R}$ with terminal condition

$$\hat{C}(T, x, v, r) = \max(0, e^x - K). \quad (2.6)$$

As in [5], we guess a solution of the form similar to the Black–Scholes formula:

$$\hat{C}(t, x, v, r) = e^x P_1(t, x, v, r) - KB(t, r) P_2(t, x, v, r). \quad (2.7)$$

Here $B(t, r)$ denotes the time- t value of a zero-coupon bond that pays off 1 at maturity, given that at time t the short rate equals r . It satisfies the PDE

$$0 = \frac{\partial B}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 B}{\partial r^2} + (\theta(t) - ar) \frac{\partial B}{\partial r} - rB \quad (2.8)$$

for $0 \leq t < T$, $r \in \mathbb{R}$ and a semi closed-form solution is given by

$$B(t, r) = e^{b(t, r)}, \quad (2.9a)$$

$$\begin{aligned} b(t, r) = & -\frac{r}{a} (1 - e^{-a(T-t)}) - \frac{1}{a} \int_t^T \theta(s) (1 - e^{-a(T-s)}) ds \\ & + \frac{\sigma^2}{2a^2} \left(T - t + \frac{2}{a} e^{-a(T-t)} - \frac{1}{2a} e^{-2a(T-t)} - \frac{3}{2a} \right). \end{aligned} \quad (2.9b)$$

By linearity, the guess (2.7) satisfies the PDE (2.5) if its two constituent terms satisfy (2.5). As such, P_1 satisfies the PDE

$$\begin{aligned} 0 = & \frac{\partial P_1}{\partial t} + \frac{1}{2}v \frac{\partial^2 P_1}{\partial x^2} + \frac{1}{2}\lambda^2 v \frac{\partial^2 P_1}{\partial v^2} + \frac{1}{2}\sigma^2 \frac{\partial^2 P_1}{\partial r^2} + \rho\lambda v \frac{\partial^2 P_1}{\partial x \partial v} + \\ & (r + \frac{1}{2}v) \frac{\partial P_1}{\partial x} + [\kappa(\eta - v) + \rho\lambda v] \frac{\partial P_1}{\partial v} + (\theta(t) - ar) \frac{\partial P_1}{\partial r}, \end{aligned} \quad (2.10)$$

and by invoking (2.8), P_2 satisfies

$$0 = \frac{\partial P_2}{\partial t} + \frac{1}{2}v \frac{\partial^2 P_2}{\partial x^2} + \frac{1}{2}\lambda^2 v \frac{\partial^2 P_2}{\partial v^2} + \frac{1}{2}\sigma^2 \frac{\partial^2 P_2}{\partial r^2} + \rho\lambda v \frac{\partial^2 P_2}{\partial x \partial v} + (r - \frac{1}{2}v) \frac{\partial P_2}{\partial x} + \kappa(\eta - v) \frac{\partial P_2}{\partial v} + \left[\theta(t) - ar + \sigma^2 \frac{\partial b}{\partial r} \right] \frac{\partial P_2}{\partial r}. \quad (2.11)$$

Further, (2.6) yields for the PDEs (2.10), (2.11) the terminal conditions

$$P_j(T, x, v, r) = 1 \quad (x > \ln K) \quad , \quad P_j(T, x, v, r) = 0 \quad (x < \ln K) \quad (2.12)$$

for $j = 1, 2$, respectively.

From the undiscounted, multidimensional version of the Feynman–Kac Theorem (cf. [7]) it follows that the solutions P_1, P_2 to (2.10), (2.11) with (2.12) can be written as expectations of the indicator function corresponding to (2.12), and thus can be regarded as probabilities¹. We next derive semi closed-form formulas for P_1 and P_2 by solving for their characteristic functions. From these characteristic functions the probabilities P_1, P_2 can be retrieved with the inversion theorem (cf. [4, 5]):

$$P_j(t, x, v, r) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \operatorname{Re} \left[\frac{e^{-iu \ln K} f_j(t, x, v, r; u)}{iu} \right] du \quad \text{for } j = 1, 2 \quad (2.13)$$

where $i^2 = -1$.

The Feynman–Kac theorem directly yields that the functions f_1, f_2 satisfy the same PDEs (2.10), (2.11), respectively, but with the terminal condition

$$f_j(T, x, v, r; u) = e^{iux}. \quad (2.14)$$

For f_1 we guess a solution of the form (cf. [5])

$$f_1(t, x, v, r; u) = \exp[F_1(t; u) + G_1(t; u)v + H_1(t; u)r + iux]. \quad (2.15)$$

Substituting this into the PDE (2.10), it follows by perusal of the coefficients of v, r and 1 that (2.15) is a solution if the functions F_1, G_1, H_1 satisfy the system of ordinary differential equations (ODEs)

$$F_1'(t) + \kappa\eta G_1(t) + \theta(t)H_1(t) + \frac{1}{2}\sigma^2 H_1(t)^2 = 0, \quad (2.16a)$$

$$G_1'(t) + \frac{1}{2}ui - \frac{1}{2}u^2 + (\rho\lambda ui + \rho\lambda - \kappa)G_1(t) + \frac{1}{2}\lambda^2 G_1(t)^2 = 0, \quad (2.16b)$$

$$H_1'(t) + ui - aH_1(t) = 0, \quad (2.16c)$$

with the terminal condition $F_1(T) = G_1(T) = H_1(T) = 0$.

For f_2 we guess a solution of the form (cf. [2, 5])

$$f_2(t, x, v, r; u) = \exp[F_2(t; u) + G_2(t; u)v + H_2(t; u)r + iux - b(t, r)]. \quad (2.17)$$

Substituting this into the PDE (2.11) and using (2.8),(2.9), it follows analogously as above that (2.17) is a solution if the functions F_2, G_2, H_2 satisfy the system of ODEs

$$F_2'(t) + \kappa\eta G_2(t) + \theta(t)H_2(t) + \frac{1}{2}\sigma^2 H_2(t)^2 = 0, \quad (2.18a)$$

$$G_2'(t) - \frac{1}{2}ui - \frac{1}{2}u^2 + (\rho\lambda ui - \kappa)G_2(t) + \frac{1}{2}\lambda^2 G_2(t)^2 = 0, \quad (2.18b)$$

$$H_2'(t) + ui - aH_2(t) - 1 = 0, \quad (2.18c)$$

¹We omit the details, which are completely analogous to those explained in [5].

with the terminal condition $F_2(T) = G_2(T) = H_2(T) = 0$.

The equations (2.16c), (2.18c) are easy to solve. Let $\delta_1 = 0$, $\delta_2 = 1$. Then

$$H_j(t; u) = \frac{ui - \delta_j}{a} \left(1 - e^{-a(T-t)}\right) \quad \text{for } j = 1, 2. \quad (2.19)$$

The equations (2.16b), (2.18b) are identical² to the first line of equation (A7) in [5] and closed-form solutions were obtained in loc. cit. For completeness, we include these formulas here. Let

$$\alpha = \kappa\eta, \quad \beta_1 = \kappa - \rho\lambda, \quad \beta_2 = \kappa, \quad \gamma_1 = \frac{1}{2}, \quad \gamma_2 = -\frac{1}{2}$$

and for $j = 1, 2$

$$d_j = \sqrt{(\beta_j - \rho\lambda ui)^2 - \lambda^2(2\gamma_j ui - u^2)}, \quad g_j = \frac{\beta_j - \rho\lambda ui + d_j}{\beta_j - \rho\lambda ui - d_j}.$$

Then the solutions to (2.16b), (2.18b) are given by

$$G_j(t; u) = \frac{\beta_j - \rho\lambda ui + d_j}{\lambda^2} \left[\frac{1 - e^{d_j(T-t)}}{1 - g_j e^{d_j(T-t)}} \right] \quad \text{for } j = 1, 2. \quad (2.20)$$

The equations (2.16a), (2.18a) can finally be solved by integration. Using the result from [5] for the integral of G_j , it follows that

$$\begin{aligned} F_j(t; u) = & \frac{\alpha}{\lambda^2} \left\{ (\beta_j - \rho\lambda ui + d_j)(T - t) - 2 \ln \left[\frac{1 - g_j e^{d_j(T-t)}}{1 - g_j} \right] \right\} \\ & + \frac{ui - \delta_j}{a} \int_t^T \theta(s) (1 - e^{-a(T-s)}) ds \\ & + \frac{\sigma^2}{2} \left(\frac{ui - \delta_j}{a} \right)^2 \left(T - t + \frac{2}{a} e^{-a(T-t)} - \frac{1}{2a} e^{-2a(T-t)} - \frac{3}{2a} \right) \end{aligned} \quad (2.21)$$

for $j = 1, 2$. Of course, for many functions θ the integral in (2.21) may be explicitly computed.

The formulas (2.4), (2.7), (2.9), (2.13), (2.15), (2.17), (2.19), (2.20), (2.21) together constitute the semi closed-form pricing formula for European call options under the hybrid asset pricing model. This pricing formula is easily seen to be a proper extension of Heston's formula, upon considering $\theta(t) \equiv ar_0$ and $\sigma = 0$.

If the integrals in (2.9b), (2.21) involving $\theta(s)$ can be explicitly computed, the pricing formula consists of two single integrals over u , see (2.13). Otherwise, one has an additional single integral over s ,

$$\int_t^T \theta(s) (1 - e^{-a(T-s)}) ds.$$

Note the useful property that the latter integral does not depend on u . In all cases, the pricing formula can be quickly approximated to any accuracy with a suitable numerical integration method. For a discussion of some computational issues relevant to the pricing formula, we refer to the paper [1] on the Heston formula.

²With the proper change of notation and removing a typo in [5].

Finally, we remark that two issues are not addressed in this note, namely whether the solution obtained above is unique and whether it satisfies the condition (2.3). These two issues are left for future research. We note that it is plausible that the probability $P_2(t, x, v, r)$ in (2.7) vanishes as $x \rightarrow -\infty$, and therefore that (2.3) holds. But, this requires a careful analysis of course.

References

- [1] H. Albrecher, Ph. Mayer, W. Schoutens, and J. Tistaert, *The little Heston trap*, Wilmott Magazine, January Issue (2007) 83–92.
- [2] G. Bakshi, C. Cao, and Z. Chen, *Empirical performance of alternative option pricing models*, J. Finan. **LII** (1997) 2003–2049.
- [3] F. Black and M. Scholes, *The pricing of options and corporate liabilities*, J. Polit. Econ. **81** (1973) 637–654.
- [4] J. Gil-Pelaez, *Note on the inversion theorem*, Biometrika **38** (1951) 481–482.
- [5] S. L. Heston, *A closed-form solution for options with stochastic volatility with applications to bond and currency options*, Rev. Finan. Stud. **6** (1993) 327–343.
- [6] J. C. Hull and A. White, *Pricing interest rate derivative securities*, Rev. Finan. Stud. **3** (1990) 573–592.
- [7] S. E. Shreve, *Stochastic Calculus for Finance II*, Springer (2004).
- [8] O. A. Vasicek, *An equilibrium characterization of the term structure*, J. Finan. Econ. **5** (1977) 177–188.

Characteristic function of the hybrid Heston–Hull–White model

*Fang Fang**

Bas Janssens†

In our contribution the goal is to find the analytic solution of the characteristic function (ch.f.) of x_T , given the initial data under the hybrid Heston–Hull–White model. That is, we want to find a closed form expression for

$$\Phi(\omega; x_0, v_0, r_0) := \mathbb{E}(\exp(i\omega x_T) | x_0, v_0, r_0).$$

A first observation on the model is the following: If x_t satisfies

$$dx_t = (r_t - \frac{1}{2}v_t)dt + \sqrt{v_t}d\tilde{W}_{1,t},$$

then $S_t = \exp(x_t)$ satisfies the Heston–Hull–White model, as can be seen by applying Itô's lemma. This paper has a twofold aim:

- Solve the problem under the assumption $\rho_{13} = \rho_{23} = 0$.
- Solve the problem under the assumption $\rho_{23} = 0$, and under the additional assumption that $\kappa\eta = \lambda^2/4$, in which case $\sqrt{v_t}$ is governed by an Ornstein–Uhlenbeck process.

It is organized as follows: in section 1, we decompose the three correlated Wiener processes into three independent ones, and establish some notation. In section 2, we eliminate two noises by exploiting the Gaussianity of the r_t -distribution, as well as the fact that x_t does not occur on the r.h.s. of the equations. In section 3, we obtain the ch.f. of x_T in the aforementioned two cases.

1 Reformulating the Model

With the assumption that $\rho_{23} = 0$, we can write $\tilde{W}_{i,t}$, $i = 1, 2, 3$, as a sum of *independent* processes $W_{i,t}$:

$$\tilde{W}_{3,t} = W_{3,t}$$

$$\tilde{W}_{2,t} = W_{2,t}$$

$$\tilde{W}_{1,t} = \alpha_1 W_{1,t} + \alpha_2 W_{2,t} + \alpha_3 W_{3,t},$$

where $\alpha_2 = \rho_{12}$, $\alpha_3 = \rho_{13}$, and $\alpha_1^2 + \alpha_2^2 + \alpha_3^2 = 1$. Thus the model is reformulated as

$$dx_t = (r_t - \frac{1}{2}v_t)dt + \alpha_1 \sqrt{v_t}dW_{1,t} + \alpha_2 \sqrt{v_t}dW_{2,t} + \alpha_3 \sqrt{v_t}dW_{3,t} \tag{1.1}$$

$$dv_t = \kappa(\eta - v_t)dt + \lambda \sqrt{v_t}dW_{2,t} \tag{1.2}$$

$$dr_t = (\theta(t) - ar_t)dt + \sigma dW_{3,t}. \tag{1.3}$$

*Technische Universiteit Delft, f. fang@ewi.tudelft.nl

†Universiteit Utrecht, janssens@math.uu.nl

Equation (1.2) gives $\sqrt{v_t}dW_{2,t} = (dv_t - \kappa(\eta - v_t)dt)/\lambda$. Insert it into (1.1) to obtain

$$dx_t = r_t dt + \left(\frac{\alpha_2 \kappa}{\lambda} - \frac{1}{2}\right)v_t dt - \frac{\alpha_2 \kappa \eta}{\lambda} dt + \frac{\alpha_2}{\lambda} dv_t + \alpha_1 \sqrt{v_t} dW_{1,t} + \alpha_3 \sqrt{v_t} dW_{3,t} \quad (1.4)$$

(cf. [1].) We introduce the notation

$$R_t := \int_0^t r_s ds \quad \text{and} \quad V_t := \int_0^t v_s ds.$$

Equation (1.4) is then integrated to

$$x_T - x_0 = R_T + \left(\frac{\alpha_2 \kappa}{\lambda} - \frac{1}{2}\right)V_T - \frac{\alpha_2 \kappa \eta}{\lambda} T + \frac{\alpha_2}{\lambda}(v_T - v_0) + \alpha_1 \int_0^T \sqrt{v_t} dW_{1,t} + \alpha_3 \int_0^T \sqrt{v_t} dW_{3,t}.$$

From now on, unless otherwise specified, all expectations are understood to be conditioned on x_0, r_0 and v_0 , i.e.

$$\mathbb{E}(X) := \mathbb{E}(X|x_0, v_0, r_0).$$

Using the tower property of conditional expectations, we have

$$\begin{aligned} \Phi(\omega; x_0, v_0, r_0) &= \mathbb{E} \left\{ \mathbb{E} \left[e^{i\omega(x_T - x_0)} | \mathcal{R}_T, \{v_s; s \in [0, T]\} \right] \right\} \\ &= \mathbb{E} \left\{ \exp(i\omega[R_T + \left(\frac{\alpha_2 \kappa}{\lambda} - \frac{1}{2}\right)V_T - \frac{\alpha_2 \kappa \eta}{\lambda} T + \frac{\alpha_2}{\lambda}(v_T - v_0)]) \right. \\ &\quad \times \mathbb{E} \left[\exp(i\omega[\alpha_1 \int_0^T \sqrt{v_t} dW_{1,t} + \alpha_3 \int_0^T \sqrt{v_t} dW_{3,t}]) | \mathcal{R}_T, \{v_s; s \in [0, T]\} \right] \left. \right\}. \end{aligned} \quad (1.5)$$

Note that v_t and R_T are only driven by their own noises, but $\Phi(\omega)$ is still driven by all three Wiener processes.

2 Elimination of Two Noises

As the title suggests, two driving noises will be eliminated in this section.

2.1 Distribution of R_T

The dynamics of the interest rate r_t can be rewritten as follows:

$$\begin{aligned} dr_t &= (\theta(t) - ar_t)dt + \sigma dW_{3,t} \\ d(e^{at}r_t) &= e^{at}\theta(t)dt + e^{at}\sigma dW_{3,t} \\ r_\tau &= e^{-a\tau}r_0 + \int_0^\tau \theta(s)e^{a(s-\tau)}ds + \sigma \int_0^\tau e^{a(s-\tau)}dW_{3,s}. \end{aligned}$$

Thus for $R_T := \int_0^T r_\tau d\tau$, we have nested integrals. Fubini's theorem yields

$$\int_0^T \left(\int_0^\tau \theta(s)e^{a(s-\tau)}ds \right) d\tau = \frac{1}{a} \int_0^T \theta(s)(1 - e^{a(s-T)})ds.$$

For the stochastic part, we have

$$\int_0^T \left(\int_0^\tau e^{a(s-\tau)} dW_{3,s} \right) d\tau = \frac{1}{a} \int_0^T (1 - e^{a(s-T)}) dW_{3,s}.$$

Patching these together, we obtain

$$R_T = F_{r_0, a, \theta}(T) + \frac{\sigma}{a} \int_0^T (1 - e^{a(s-T)}) dW_{3,s}, \quad (2.1)$$

with

$$F_{r_0, a, \theta}(T) := \frac{r_0}{a} (1 - e^{-aT}) + \frac{1}{a} \int_0^T \theta(s) (1 - e^{a(s-T)}) ds.$$

Since the Itô integral in (2.1) is a weighted Wiener process, R_T has a Gaussian distribution with mean $F(T)$ and variance

$$\text{Var}(T) := \frac{\sigma^2}{a^2} \int_0^T (1 - e^{a(s-T)})^2 ds = \frac{\sigma^2}{a^2} \left(T - \frac{2}{a} (1 - e^{-aT}) + \frac{1}{2a} (1 - e^{-2aT}) \right).$$

2.2 The Correlation

Recall the expression for Φ in (1.5). Let us first focus on the inner expectation, which is conditioned on R_T and the complete path $\{v_s\}_{0 \leq s \leq T}$.

We fix a function $w : [0, T] \rightarrow \mathbb{R}^+$ such that $v_s = w(s)$ and introduce the notation $W_i(f) := \int_0^T f(s) dW_{i,s}$.

Since there is no restriction on x_s , the process $W_{1,t}$ is still a Wiener process. With fixed $v_s = w(s)$, the random variable $W_{1,t}(\sqrt{w})$ is just a weighted Brownian motion. For $W_{3,t}$ however, there is a restriction, since we have fixed R_T . When we define $g(s) := (1 - e^{a(s-T)})$, we have fixed

$$W_{3,t}(g) = \frac{a}{\sigma} (R_T - F_{r_0, a, \theta}(T)).$$

Apart from the fixed x_0 , r_0 and v_s , this is the *only* relevant restraint. And since $W_{3,t}(f)$ is independent from $W_{3,t}(g)$ if $f \perp g$ we simply decompose

$$W_{3,t}(\sqrt{w}) = W_{3,t}(\sqrt{w_{\parallel}}) + W_{3,t}(\sqrt{w_{\perp}})$$

with

$$\sqrt{w_{\parallel}} = \frac{\langle \sqrt{w}, g \rangle}{\langle g, g \rangle} g \quad \text{and} \quad \sqrt{w_{\perp}} = \sqrt{w} - \frac{\langle \sqrt{w}, g \rangle}{\langle g, g \rangle} g.$$

Thus, for fixed v_s and R_T , we know that $W_{3,t}(\sqrt{w})$ is Gaussianly distributed with mean

$$\frac{a}{\sigma} (R_T - F_{r_0, a, \theta}(T)) \times \frac{\langle \sqrt{w}, g \rangle}{\langle g, g \rangle}$$

and variance

$$\langle \sqrt{w_{\perp}}, \sqrt{w_{\perp}} \rangle = \langle \sqrt{w}, \sqrt{w} \rangle - \frac{\langle \sqrt{w}, g \rangle^2}{\langle g, g \rangle},$$

where $\langle g, g \rangle = T - \frac{2}{a} (1 - e^{-aT}) + \frac{1}{2a} (1 - e^{-2aT})$. If we define

$$\mu_T := \int_0^T (1 - e^{a(s-T)}) \sqrt{v_s} ds,$$

we have

$$W_{3,t}(\sqrt{w}) \sim N\left(\frac{a\mu_T}{\sigma\langle g, g \rangle}(R_T - F(T)), V_T - \frac{\mu_T^2}{\langle g, g \rangle}\right).$$

Recall that $W_{1,t}$, $W_{2,t}$ and $W_{3,t}$ are independent Wiener processes. The process R_t is only driven by $W_{3,t}$, and v_t only by $W_{2,t}$. For fixed R_T and $v_s = w(s)$, we therefore know that $W_{1,t}(\sqrt{w})$ and $W_{3,t}(\sqrt{w})$ are independent Gaussians. The conditional characteristic function (CCF) of their sum,

$$\phi(\omega; R_T, v_s) := \mathbb{E}\left[\exp\left(i\omega(\alpha_1 W_{1,t}(\sqrt{w}) + \alpha_3 W_{3,t}(\sqrt{w}))\right) | R_T, \{v_s; s \in [0, T]\}\right],$$

is therefore the product of the individual CCF's for $\alpha_1 W_{1,t}(\sqrt{w})$ and $\alpha_3 W_{3,t}(\sqrt{w})$. These we know; the characteristic function of a Gaussian with mean μ and variance u is

$$f_{\mu,u}(\omega) = \exp\left(i\mu\omega - \frac{u}{2}\omega^2\right).$$

Adding the means and variances of the two independent Gaussians, we write

$$\phi(\omega; R_T, v_s) = f_{\alpha_3 \frac{a\mu_T}{\sigma\langle g, g \rangle}(R_T - F(T)), (\alpha_1^2 + \alpha_3^2)V_T - \alpha_3^2 \frac{\mu_T^2}{\langle g, g \rangle}}(\omega)$$

which only depends on R_T , μ_T and V_T .

Returning to (1.5), we have

$$\begin{aligned} \Phi(\omega; x_0, v_0, r_0) &= \mathbb{E}\left[\exp\left(i\omega\left(R_T + \left(\frac{\alpha_2\kappa}{\lambda} - \frac{1}{2}\right)V_T - \frac{\alpha_2\kappa\eta}{\lambda}T + \frac{\alpha_2}{\lambda}(v_T - v_0)\right)\right)\right. \\ &\quad \times \exp\left(i\omega\left(\alpha_3 \frac{a\mu_T}{\sigma\langle g, g \rangle}[R_T - F(T)]\right)\right) \\ &\quad \left. \times \exp\left(-\frac{1}{2}\omega^2\left((\alpha_1^2 + \alpha_3^2)V_T - \alpha_3^2 \frac{\mu_T^2}{\langle g, g \rangle}\right)\right)\right]. \end{aligned} \quad (2.2)$$

We now use the tower rule to get an inner expectation conditioned on μ_T , V_T and v_T . That is:

$$\Phi(\omega; x_0, v_0, r_0) = \mathbb{E}(\dots) = \mathbb{E}(\mathbb{E}(\dots | \mu_T, V_T, v_T)).$$

Recall that R_T is independent of V_T , v_T , or μ_T . So what remains as the inner expectation is $\mathbb{E}(e^{i\omega c(R_T - F(T))} | \mu_T, V_T, v_T)$, with $c = (1 + \alpha_3 \frac{a}{\sigma\langle g, g \rangle})\mu_T$. Since $R_T - F(T)$ is $N(0, \frac{\sigma^2}{a^2}\langle g, g \rangle)$ -distributed, we have

$$\mathbb{E}(\exp(i\omega c(R_T - F(T))) | \mu_T, V_T, v_T) = f_{0, \frac{\sigma^2}{a^2}\langle g, g \rangle}(c\omega) = \exp\left(-\frac{\sigma^2}{2a^2}\langle g, g \rangle c^2 \omega^2\right).$$

In writing this out, the term $-\frac{1}{2}\frac{\alpha_3^2}{\langle g, g \rangle}\mu_T^2\omega^2$ mysteriously vanishes;

$$\begin{aligned} \Phi(\omega; x_0, v_0, r_0) &= \exp\left(i\omega F(T) - i\omega \frac{\alpha_2\kappa\eta}{\lambda}T - \frac{1}{2}\omega^2 \frac{\sigma^2}{a^2}\langle g, g \rangle\right) \\ &\quad \times \mathbb{E}\left[e^{i\omega \frac{\alpha_2}{\lambda}(v_T - v_0)} \times e^{i\omega\left[\left(\frac{\alpha_2\kappa}{\lambda} - \frac{1}{2}\right) + \frac{1}{2}i\omega(\alpha_1^2 + \alpha_3^2)\right]V_T} \times e^{i\omega\left[i\omega\alpha_3 \frac{\sigma}{a}\right]\mu_T}\right]. \end{aligned} \quad (2.3)$$

We simplify the expression by introducing the notations

$$C_0 := e^{i\omega F(T) - i\omega \frac{\alpha_2\kappa\eta}{\lambda}T - \frac{1}{2}\omega^2 \frac{\sigma^2}{a^2}\langle g, g \rangle}, \quad C_1 := \frac{\alpha_2}{\lambda},$$

$$C_2 := \frac{\alpha_2 \kappa}{\lambda} - \frac{1}{2} + \frac{1}{2} i \omega (\alpha_1^2 + \alpha_3^2), \quad C_3 := i \omega \alpha_3 \sigma / a,$$

and

$$Z_T - Z_0 := C_1(v_T - v_0) + C_2 V_T + C_3 \mu_T. \quad (2.4)$$

Thus we have

$$\Phi(\omega; x_0, v_0, r_0) = C_0 \mathbb{E}[\exp(i\omega(Z_T - Z_0))]. \quad (2.5)$$

Therefore, finding the ch.f. of x_T at ω is equivalent to finding the ch.f. of $Z_T - Z_0$ at ω (where Z_t still depends on ω through the constants $C_i(\omega)$).

Integrating dv_t over $[0, T]$ gives

$$v_T - v_0 = \int_0^T \kappa(\eta - v_s) ds + \int_0^T \lambda \sqrt{v_s} dW_{2,s}.$$

Substituting this into (2.4) yields

$$Z_T - Z_0 = \int_0^T [C_1 \kappa(\eta - v_s) + C_2 v_s + C_3 g(s) \sqrt{v_s}] ds + C_1 \int_0^T \lambda \sqrt{v_s} dW_{2,s}. \quad (2.6)$$

Equivalently, the dynamics of Z_t read

$$dZ_t = [C_1 \kappa \eta + (C_2 - C_1 \kappa) v_t + C_3 g(t) \sqrt{v_t}] dt + C_1 \lambda \sqrt{v_t} dW_{2,t}. \quad (2.7)$$

We have a small subtlety here. In principle, $g(s, T)$ depends both on s and on T . In the dynamics of Z_t , this would give rise to terms involving $\partial g / \partial T$. We circumvent this problem by defining, for each fixed T , a process $t \mapsto \hat{Z}(T)_t$, which is defined according to equation (2.7), in which $g(s, T)$ has fixed T . Then $Z_T = \hat{Z}(T)_T$. From now on, we work with equation (2.7), omitting the hats.

All in all, we are left with Z_t , which is driven by a single noise, $W_{2,t}$.

3 Analytic Solution

We denote the ch.f. of Z_T conditioned on \mathcal{F}_t by

$$\Psi_t(\omega; \mathcal{F}_t) := \mathbb{E}[\exp(i\omega Z_T) | \mathcal{F}_t].$$

By definition, Ψ_t is a martingale: $\mathbb{E}[d\Psi_t | \mathcal{F}_t] = 0$. It is clear that Ψ_t depends only on $Z_t, v_t, \sqrt{v_t}, t$. Therefore, Itô's lemma yields, setting $\tau = T - t$:

$$\begin{aligned} d\Psi(\omega; Z_t, v_t, \sqrt{v_t}, \tau) &= -\frac{\partial \Psi}{\partial \tau} dt + \frac{\partial \Psi}{\partial Z_t} dZ_t + \frac{\partial \Psi}{\partial v_t} dv_t + \frac{\partial \Psi}{\partial \sqrt{v_t}} d\sqrt{v_t} + \frac{1}{2} \frac{\partial^2 \Psi}{\partial Z_t^2} (dZ_t)^2 \\ &\quad + \frac{1}{2} \frac{\partial^2 \Psi}{\partial v_t^2} (dv_t)^2 + \frac{1}{2} \frac{\partial^2 \Psi}{\partial \sqrt{v_t}^2} (d\sqrt{v_t})^2 + \frac{\partial^2 \Psi}{\partial Z_t \partial v_t} dZ_t dv_t \\ &\quad + \frac{\partial^2 \Psi}{\partial Z_t \partial \sqrt{v_t}} dZ_t d\sqrt{v_t} + \frac{\partial^2 \Psi}{\partial v_t \partial \sqrt{v_t}} dv_t d\sqrt{v_t}. \end{aligned} \quad (3.1)$$

Given the dynamics of $\sqrt{v_t}$, this gives rise to a PDE for Ψ . We proceed with the two cases in which we can solve this.

3.1 Case 1: $\rho_{23} = 0$ and $\rho_{13} = 0$

This is essentially the Heston model. Indeed, it is immediately clear from equation (1.1) that $x_t = x_{H,t} + R_t - r_0t$, where $x_{H,t}$ denotes the logarithmic price in the Heston model. Therefore,

$$\Phi(\omega; x_0, v_0, r_0) = \chi(\omega)\Phi_H(\omega; x_0, v_0, r_0),$$

with $\chi(\omega) = e^{i\omega F(T) - i\omega r_0 T - \frac{1}{2}\omega^2 \frac{v_0}{\sigma^2} \langle g, g \rangle}$ the characteristic function of $R_T - r_0T$.

3.2 Case 2: $\rho_{23} = 0$ and $\kappa\eta = \lambda^2/4$

We will now solve a different set of equations:

$$dx_t = (r_t - \frac{1}{2}v_t)dt + \Theta_t d\tilde{W}_{1,t} \quad (3.2)$$

$$d\Theta_t = -\beta\Theta_t dt + \delta dW_{2,t} \quad (3.3)$$

$$dr_t = (\theta(t) - ar_t)dt + \sigma W_{3,t} \quad (3.4)$$

The relevance is as follows: if we set $B_{2,t} := \int_0^t sn(\Theta_s) dW_{2,s}$, with $sn(x)$ the sign of x , then $t \mapsto B_{2,t}$ is again a Wiener process by Levy's Martingale characterization of Brownian motions. We then see that $v_t = \Theta_t^2$ satisfies (the 2nd equation of) the hybrid Heston–Hull–White model for the particular case $\kappa\eta = \lambda^2/4$.

Indeed, $d\Theta_t^2 = (\delta^2 - 2\beta\Theta_t^2)dt + 2\delta\Theta_t dW_{2,t}$ and $dW_{2,t} = sn(\Theta_t)dB_{2,t}$, so that

$$dv_t = (\delta^2 - 2\beta v_t)dt + 2\delta\sqrt{v_t}dB_{2,t}.$$

(Recall that $\sqrt{v_t} = |\Theta_t| = sn(\Theta_t)\Theta_t$.) This requires $\lambda = 2\delta$, $\kappa = 2\beta$ and $\kappa\eta = \delta^2$, and thus $\kappa\eta = \lambda^2/4$. This subclass of the model, in which Θ_t is an Ornstein–Uhlenbeck process, was in fact Heston's way of justifying the more general model [2]. Notice that we have changed $\sqrt{v_t}$ into Θ in equation (3.2). This seems to be essential unless $\rho_{13} = 0$.

We change to a new measure, under which Θ_t is simply Brownian motion [3]. Define the Radon–Nikodym derivative as

$$M_\tau = \exp(-Y_\tau) := \exp\left(-\int_0^\tau \frac{-\beta\Theta_t}{\delta} dW_{2,t} - \frac{1}{2}\int_0^\tau \frac{\beta^2\Theta_t^2}{\delta^2} dt\right).$$

Substitute $\int_0^\tau \Theta_t dW_{2,t}$ by $\frac{1}{2\delta}[(v_\tau - v_0) + \int_0^\tau 2\beta\Theta_t^2 dt - \delta^2\tau]$, we have

$$Y_\tau = -\frac{\beta}{2\delta^2}(v_\tau - v_0) - \frac{\beta^2}{2\delta^2}V_\tau + \frac{\beta}{2}\tau.$$

By Girsanov's theorem, if $\mathbb{E}\left[\exp\left(\int_0^T \frac{\beta^2}{\delta^2}v_t dt\right)\right] < \infty$, then

$$\hat{B}_\tau := \int_0^\tau \frac{-\beta\Theta_t}{\delta} dt + W_{2,\tau}; \quad \tau \leq T$$

is a Brownian motion w.r.t. $d\mathbb{Q} = M_T d\mathbb{P}$, where \mathbb{P} denotes the old measure. Moreover, in terms of \hat{B}_t , the process Θ_t has the representation of

$$d\Theta_t = \delta d\hat{B}_t.$$

Thus

$$dv_t = d\Theta_t^2 = 2\Theta_t d\Theta_t + d\Theta_t^2 = 2\delta\Theta_t d\hat{B}_t + \delta^2 dt.$$

The derivation of equation (2.7) goes through word for word in this new system, except that all \sqrt{v} 's are replaced by Θ 's. (In particular, μ should be defined as $\mu_t = \int_0^t g(s)\Theta_s ds$.) We obtain

$$dZ_t = \left[C_1\kappa\eta + (C_2 - C_1\kappa)\Theta_t^2 + C_3g(t)\Theta_t \right] dt + C_1\lambda\Theta_t dW_{2,t}. \quad (3.5)$$

For the expectation w.r.t. \mathbb{P} , one has

$$\begin{aligned} \mathbb{E}^{\mathbb{P}} [\exp(i\omega(Z_T - Z_0))] &= \int_{\Omega} \exp(i\omega(Z_T - Z_0)) M_T^{-1} d\mathbb{Q} \\ &= \int_{\Omega} \exp(i\omega(Z_T - Z_0)) \exp(Y_T) d\mathbb{Q} := \mathbb{E}^{\mathbb{Q}} [\exp(i\omega(\tilde{Z}_T - \tilde{Z}_0))], \end{aligned}$$

with (we use v_t for Θ_t^2 and V_t for its integral)

$$\tilde{Z}_T - \tilde{Z}_0 = (C_1 + i\frac{\beta}{2\omega\delta^2})(v_T - v_0) + (C_2 + i\frac{\beta^2}{2\omega\delta^2})V_T + C_3\mu_T - i\frac{\beta}{2\omega}T,$$

and

$$d\tilde{Z}_t = \left[C_1\delta^2 + (C_2 + i\frac{\beta^2}{2\omega\delta^2})v_t + C_3g(t)\Theta_t \right] dt + (2\delta C_1 + i\frac{\beta}{\omega\delta})\Theta_t d\hat{B}_t.$$

Let us then set $\tau := T - t$. Our ansatz for Ψ will be

$$\Psi(\tilde{Z}_t, v_t, \Theta_t, \tau) = \exp \left[C(\tau) + D(\tau)\Theta_t + E(\tau)v_t + i\omega\tilde{Z}_t \right], \quad (3.6)$$

with initial conditions

$$C(0) = 0, D(0) = 0 \text{ and } E(0) = 0.$$

For this we have

$$\begin{aligned} \frac{\partial \Psi}{\partial \tau} / \Psi &= \frac{\partial C}{\partial \tau} + \frac{\partial D}{\partial \tau} \Theta_t + \frac{\partial E}{\partial \tau} v_t, & \frac{\partial \Psi}{\partial \tilde{Z}_t} / \Psi &= i\omega, \\ \frac{\partial \Psi}{\partial v_t} / \Psi &= E, & \frac{\partial \Psi}{\partial \Theta_t} / \Psi &= D, \\ \frac{\partial^2 \Psi}{\partial \tilde{Z}_t^2} / \Psi &= -\omega^2, & \frac{\partial^2 \Psi}{\partial v_t^2} / \Psi &= E^2, \\ \frac{\partial^2 \Psi}{\partial \Theta_t^2} / \Psi &= D^2, & \frac{\partial^2 \Psi}{\partial v_t \partial \tilde{Z}_t} / \Psi &= i\omega E, \\ \frac{\partial^2 \Psi}{\partial \Theta_t \partial \tilde{Z}_t} / \Psi &= i\omega D, & \frac{\partial^2 \Psi}{\partial \Theta_t \partial v_t} / \Psi &= DE. \end{aligned}$$

Substituting these into (3.1), factoring out Ψ and remembering $\mathbb{E}(d\Psi) = 0$, we come to the following equation:

$$\begin{aligned} 0 &= \left(-\frac{\partial C}{\partial \tau} + i\omega C_1\delta^2 + \delta^2 E + \frac{1}{2}\delta^2 D^2 \right) \\ &+ v_t \left(-\frac{\partial E}{\partial \tau} + i\omega C_4 - \frac{1}{2}\omega^2 C_5^2 + 2\delta^2 E^2 + 2i\omega\delta C_5 E \right) \\ &+ \Theta_t \left(-\frac{\partial D}{\partial \tau} + i\omega C_3 g(t) + i\omega\delta C_5 D + 2\delta^2 ED \right), \end{aligned} \quad (3.7)$$

with $C_4 := C_2 + i\frac{\beta^2}{2\omega\delta^2}$ and $C_5 := 2\delta C_1 + i\frac{\beta}{\omega\delta}$. Since (3.7) has to hold for every v_t and Θ_t , we obtain three ODEs:

$$-\frac{\partial C}{\partial \tau} + i\omega C_1 \delta^2 + \delta^2 E + \frac{1}{2} \delta^2 D^2 = 0 \quad (3.8)$$

$$-\frac{\partial E}{\partial \tau} + i\omega C_4 - \frac{1}{2} \omega^2 C_5^2 + 2\delta^2 E^2 + 2i\omega\delta C_5 E = 0 \quad (3.9)$$

$$-\frac{\partial D}{\partial \tau} + i\omega C_3 g(t) + i\omega\delta C_5 D + 2\delta^2 ED = 0. \quad (3.10)$$

With $\gamma := \delta \sqrt{-2i\omega C_4}$, we find

$$E(\tau) = e_+ \frac{1 - \exp[2\delta^2(e_+ - e_-)\tau]}{1 - \frac{e_\pm}{e_-} \exp[2\delta^2(e_+ - e_-)\tau]}, \quad (3.11)$$

$$D(\tau) = \frac{i\omega C_3 e^{\gamma\tau}}{\frac{e_\pm}{e_-} e^{2\gamma\tau} - 1} \times \left(\left(\frac{1}{\gamma} (e^{-\gamma\tau} - 1) - \frac{1}{a + \gamma} (e^{-(a+\gamma)\tau} - 1) \right) + \frac{e_\pm}{e_-} \left(\frac{1}{\gamma} (e^{\gamma\tau} - 1) - \frac{1}{\gamma - a} (e^{(\gamma-a)\tau} - 1) \right) \right), \quad (3.12)$$

$$C(\tau) = (e_+ + i\omega C_1) \delta^2 \tau - \frac{1}{2} \log\left(\frac{e_\pm}{e_-} e^{2\delta^2(e_+ - e_-)\tau} - 1\right) + \frac{1}{2} \int_0^\tau D^2(s) ds. \quad (3.13)$$

We briefly sketch how to arrive at this. Reformulate (3.9) as

$$\frac{d}{d\tau} [\log(E - e_+) - \log(E - e_-)] = 2\delta^2(e_+ - e_-),$$

with $e_\pm = \frac{-i\omega C_5 \pm \sqrt{-2i\omega C_4}}{2\delta}$. This yields (3.11). For (3.12), we first solve the homogeneous equation

$$\frac{dD_0}{d\tau} = i\omega\delta C_5 D_0 + 2\delta^2 E D_0.$$

Explicitly,

$$D_0(\tau) = \exp(i\omega\delta C_5 \tau + 2\delta^2 \int_0^\tau E(s) ds),$$

and¹

$$\int_0^\tau E(s) ds = e_+ \tau - \frac{1}{2\delta^2} \log\left(\frac{e_+}{e_-} e^{2\delta^2(e_+ - e_-)\tau} - 1\right).$$

Thus

$$D_0(\tau) = \frac{\exp((2\delta^2 e_+ + i\omega\delta C_5)\tau)}{\frac{e_\pm}{e_-} \exp(2\delta^2(e_+ - e_-)\tau) - 1}.$$

From ‘variation of constants’, we see $D(\tau) = i\omega C_3 D_0(\tau) \int_0^\tau g(T-s) D_0^{-1}(s) ds$, with $g(T-s) = (1 - e^{-as})$. The result of this laborious but simple integration is shown above. (One uses $2\delta^2(e_+ - e_-) = 2\gamma$ and $2\delta^2 e_\pm + i\omega\delta C_5 = \pm\gamma$.) The result for $C(\tau)$ is obtained by integration, where we have left the term $\int_0^\tau D^2(s) ds$ intact.

Substituting the above into (3.6) and (2.5), we obtain the analytic solution for ch.f. of x_T given x_0, v_0 and r_0 .

¹In principle, the logarithm should be interpreted carefully, in the sense that the function should not jump at branch cuts. However, since it occurs inside an exponential eventually, this remark belongs in a footnote.

4 Conclusion

Towards a solution to the problem of finding the characteristic function of the Heston–Hull–White model, we have made the following observations:

- With ρ_{13} and ρ_{23} equal to zero, the problem is essentially equivalent to Heston’s model, and can be solved.
- With $\rho_{23} = 0$ and $\kappa\eta = \lambda^2/4$, but with arbitrary ρ_{13} , the problem can also be solved. This is an extension of the model by Stein and Stein [4]. It is tractable because the process underlying the volatility is an Ornstein–Uhlenbeck process, and not a Bessel process as in the Heston model. This is exactly the special case used by Heston to motivate the general model.

With only $\rho_{23} = 0$, the problem seems to be less simple. Still, we have been able to eliminate two out of three driving noises, which may result in faster numerical simulation.

Acknowledgement

We would like to thank Coen Leentvaar for valuable discussion.

References

- [1] M. Broadie, Ö. Kaya, ‘*Exact Simulation of Stochastic Volatility and other Affine Jump Diffusion Processes*’, *Operations Research* 54, 217–231 (2006).
- [2] S.L. Heston, ‘*A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bonds and Currency Options*’, *Rev. Financial Stud.* 6, 327–343, (1993).
- [3] J. Pitman, M. Yor, ‘*A Decomposition of Bessel Bridges*’, *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 59, 425–457, (1982).
- [4] E.M. Stein, J.C. Stein, ‘*Stock Price Distributions with Stochastic Volatility; An Analytic Approach.*’ *Rev. Financial Stud.* 4, 727–752 (1991).