Optimization of duty cycles in magnetic resonance imaging systems

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Optimization of duty cycles in Magnetic Resonance Imaging systems

E.N. Ivanov

This work has been carried out under the auspices of the Engineering Mechanics research school.

Optimization of duty cycles in Magnetic Resonance Imaging systems

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Technische Universiteit Eindhoven, op gezag van de
rector magnificus, prof.dr.ir. C.J. van Duijn, voor een
commissie aangewezen door het College voor
Promoties in het openbaar te verdedigen
op dinsdag 27 maart 2012 om 16.00 uur

door

Evgeniy Nikolaevich Ivanov

geboren te Sint-Petersburg, Rusland

Dit proefschrift is goedgekeurd door de promotor:

prof.dr.ir. J.E. Rooda

Copromotor:
dr. A.Y. Pogromsky

Preface

This thesis is the final result of my PhD research at the Systems Engineering Group of the Mechanical Engineering Department at the Eindhoven University of Technology.

My PhD research has been performed as part of the Darwin research project. Darwin is a cooperation between Philips Healthcare, Philips Research, the Embedded Systems Institute, the Eindhoven University of Technology and several other partners. Finding a proper balance between the academic research and the real-life industrial problems of Philips Healthcare was one of the many interesting challenges that this project has offered to me.

At this point, I would like to thank special people who contributed to the research as presented in this thesis and supported me during these years.

First of all, I would like to thank professor Koos Rooda for giving me the opportunity and facilities to perform this PhD project within the Systems Engineering Group and for his attentive supervision during the project. I would also like to thank my co-promotor Alexandr Pogromsky for his dedicated coaching, co-authoring all articles in this thesis and sharing his expertise with me. Furthermore, I would like to thank professor Ivo Adan, professor Jan Tijmen Udding and professor Ger Koole, for their consent to be a part of the reading committee as well as their valuable comments that enabled me to improve the quality of my dissertation.

Next, I would like to thank people from Philips Healthcare for posing challenging industrial problems, sharing their expertise, coaching and supporting this research. Especially, my gratitude goes to Frank Benschop for guiding me through initial phase of the project and introduction to Philips Healthcare and MRI systems; to Geran Peeren for sharing his highest expertise; to Johan van den Brink for his coaching, contribution to several articles, help with the experimental setup and his contagious enthusiasm throughout the project; to Andre van Est for his

help in the final stage of this project and special thanks for organizing experiments in a hospital.

The Darwin project brought together an interesting mix of people from different backgrounds to form a team. I enjoyed being part of that team and would like to thank all Darwin project members, not only for their cooperation and fruitful discussions, but also for all the fun we had during social events. Special thanks to our Darwin coaches: Pierre van den Laar, David Watts, Teade Punter and Pierre America for the coordination of the project, their support and valuable advices.

I thank my colleagues at the Systems Engineering Group of the Eindhoven University of Technology for the pleasant working atmosphere. Especially, I thank Konstantin Starkov for his friendship and endless support that kept me motivated and productive. Special gratitude to Mieke Lousberg for her personal interest and for helping me with all the administrative affairs that I encountered during the project. Furthermore, I would like to thank all the students of this group who helped me in the research as part of their bachelor or master assignment. Results of those studies found their place in Chapters 4 and 6 of this dissertation.

On a more personal note, I would like to thank my parents. Dear Ma and Pa, I thank you, I have achieved a lot. Thank you for your love, patience and support! Дорогие мои мама и папа, я добился очень многого благодаря Вам. Огромное Вам спасибо за Вашу любовь, терпение и поддержку!

A special word of gratitude goes to my beloved wife Marina for her endless love and support.

Evgeniy Ivanov

February, 2012

Summary

Optimization of duty cycles in Magnetic Resonance Imaging systems

About 30 years ago the first commercial Magnetic Resonance Imaging (MRI) scanner was installed at the Hammersmith Hospital in London. This revolutionary technique made it possible to image tissues surrounded by bone. This was a big advantage in comparison to X-ray based imaging methods. However, resolution of the first magnetic resonance images was low and the scanning time was long, due to problems of weak signal and high sensitivity to the patient motion. Since then a lot of research has been done to improve the overall performance of the machines. In mid 90s, fast imaging techniques were developed that had a tremendous impact on the popularity of MRI among other medical imaging methods.

Nowadays there are a lot of clinical imaging applications where MRI overtakes the X-ray successors. Moreover, MRI is believed to be harmless to the patient, because no ionizing radiation is utilized. However, the main disadvantages of MRI are strong magnetic field, extreme expense, and relatively long examination time when compared to X-ray. The first factor imposes high safety standards that must be respected in an MRI scanner room, whereas the last two factors prevent hospitals from fast investments return. Moreover, due to high demand on MRI examinations, the patient waiting lists in hospitals are often several weeks long. This backlog decreases patient satisfaction.

In this dissertation, a new approach to reduce the examination time of MRI systems is described. The time reduction is accomplished by dividing parts of the MRI examination into segments that are then intermixed. The intermixing algorithms are based on scheduling technique from the field of Operations Research.

There are a number of physical parameters that restrict performance of MRI systems, such as temperature of MRI hardware during the examination. Also, due to electromagnetic effects inside the bore of MRI scanner, the temperature

of patient's body can get close to an uncomfortable level. In current practice, all these duty cycle limitations are modeled and verified before the MRI examination starts. Then, if necessary, the MRI examination time is prolonged, in order not to exceed the temperature limits. Typical MRI examination consists of several discrete parts, i.e., scans. Different types of scans impose different duty cycle limitations. The approach proposes that the examination can be divided into small segments that are rescheduled in such a way that the adverse effects of duty cycle limited scans are reduced by non-limited scans.

In this thesis, several scheduling algorithms are described that were designed to deal with different kinds of duty cycle limitations and to improve performance of MRI systems. The algorithms were verified on a large number of MRI examinations. According to collected statistics, time of MRI examinations can be reduced by up to 22%. As a result, the capacity of one MRI system can be increased by up to 4 patients per day. Moreover, special MRI experiments were carried out to validate the algorithms.

Finally, the thesis presents an approach to patient flow modeling in MRI departments in hospitals. The patient flow is modeled by means of queuing theory in order to uncover bottlenecks. Then, discrete-event computer simulations are performed to overcome limitations of the classical queuing theory assumptions. The current hospitals practice demonstrates that the MRI scanners are not always the bottleneck in the overall examinations workflow. The resulting models can be utilized to predict patient flow for various layouts of MRI departments and appointment scheduling strategies. Based on these detailed models, recommendations on improving MRI departments' workflow can be derived.

The results of this study can be used to optimize performance of MRI departments in hospitals or free-standing imaging centers. First, the MRI scanning time can be reduced. Second, the patient flow can be optimized that yields the overall MRI examination time reduction. This will result in better patient comfort and faster return on investments in MRI equipment.

Contents

Chapter 1

Introduction

About 30 years ago the first commercial Magnetic Resonance Imaging (MRI) scanner was installed at the Hammersmith Hospital in London, [Greenblatt et al. 1997]. This revolutionary imaging technique provided ability to image tissues surrounded by bone (e.g. gray and white matter of brain) which was a big advantage in comparison to X-ray based imaging methods. However, resolution of the first MRI images and the scan time left much to be desired. This was due to problems of weak signal and high sensitivity to the patient motion. Since then a lot of research has been done to improve the overall performance of the machines. In mid 90s, the fast imaging techniques were developed, such as gradient echo and phased array coil technology, that had a tremendous impact on the popularity of MRI among other medical imaging methods, [Mcrobbie et al. 2007].

Nowadays MRI has become an essential part of clinical diagnostic imaging. MRI scanners generate highly detailed images of soft-tissue structures, organs, blood vessels, and the brain matter. They are capable of revealing even the most subtle differences in patient's body tissues.

In Figure 1.1(a) one can see a modern MRI scanner by Philips Healthcare. The main and the largest part of the scanner is a magnet to generate strong static magnetic field. In this figure the outstanding part of the scanner is a table, where patient is positioned, which is called patient support table. There are more components inside the scanner: gradient coils, magnetic shielding, etc. Gradient coils together with gradient amplifiers are parts of magnetic gradient system. This system generates magnetic field gradients that are used for encoding spatial

information on the nuclei within a tissue sample by local variations of magnetic field.

A typical MRI examination procedure is as follows. At the beginning of the examination, an examinee is positioned on the patient support table. Afterwards technicians place Radio Frequency (RF) coils close the examined region of the patient's body, see Figure 1.1(b). Then the tabletop with the patient is moved inside the bore of the MRI scanner. During MRI examination, a number of RF pulses are emitted by these RF coils, and protons inside the examined region of the body are excited and brought into phase. Then the protons gradually release the absorbed energy, which is measured by RF receiver system and mathematically reconstructed on a computer to form highly detailed images.
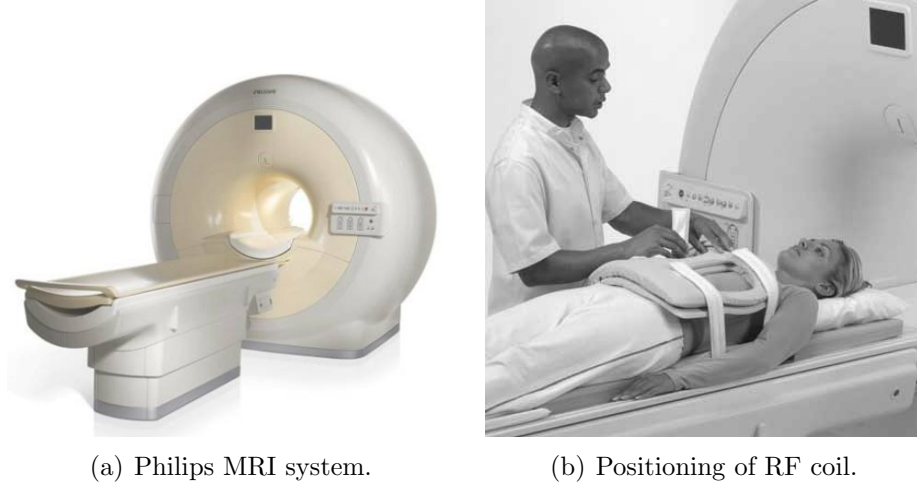
(a) Philips MRI system.   (b) Positioning of RF coil.

Figure 1.1: MRI examination.

After MRI examination is finished, the patient support tabletop is moved out of the bore, the RF coils are removed, and the patient is released. In contrast to X-ray systems, there is no need to develop the images, and they are ready for the radiologist's expertise straightaway.

More detailed insights into MRI hardware and the examination process follows in Chapter 2 of this dissertation. A broader view on MRI examination is presented in Chapter 6, where not only the scanning process itself, but complete patient path in MRI department of hospital is addressed.

1.1   MRI system performance limitations

Due to high demand on MRI examinations, the patient waiting lists in hospitals are often several weeks long, see e.g. [Canada 2008]. This backlog decreases

patient satisfaction.

There are several factors that limit performance of modern MRI systems. The most tangible are: RF heat dissipation level, temperature of gradient amplifiers, and temperature of gradient coils. The first one is patient health-related, whereas other two are MRI hardware-related. These three major performance hitters are briefly introduced in this section.

During MRI examination, RF pulses can increase radiation level that in turn can result in heating of patient's body. The amount of the RF power dissipated per kilogram of patient's body is called Specific energy Absorption Rate (SAR) level. SAR level depends on RF field characteristics such as power of the RF pulses and their frequency. There are several SAR safety limits prescribed by International Electrotechnical Commission (IEC) that should not be exceeded during MRI examinations, [CEI/IEC 60601-2-33 2008]. These safety limits vary per region of the patient's body. For instance, for the head the SAR limits are quite strict, whereas for extremities the SAR limits are looser.

Temperatures of gradient amplifiers and coils depend on amplitude and frequency of magnetic gradient waveforms. In order not to damage MRI hardware, safety limits are adopted on these temperatures as well. For instance, in Philips MRI software all the gradient waveforms are verified prior to examination not to exceed the temperature limits, otherwise some extra time (so-called *'dead time'*) is introduced into the gradient waveforms. During MRI examination, temperatures of amplifiers and coils are often measured by hardware sensors to stay in the safety bounds, otherwise the examination is immediately halted.

Therefore, MRI examination time is often prolonged not to overheat the coils and the amplifiers and not to exceed safety levels of SAR. In order to overcome these performance limitations and reduce the examination time, a new method is introduced and investigated in this dissertation. This method is based, inter alia, on insights into the internal structure of MRI examination, which is introduced in the following section.

1.2 Optimization of performance of MRI examination

In this section the essence of our method is introduced. The method is based on the structure of MRI examinations, which is briefly described here. Detailed insights into the MRI examination structure follows in Chapter 2.

All MRI examinations have a discrete structure, which is illustrated in Figure 1.2. This figure should be viewed from top to bottom as follows. A typical MRI examination consists of several scans. Each scan in turn consists of multiple components, which are commonly denoted by 'TR'. Every TR defines a profile of RF

pulses and magnetic gradient waveforms to be executed during the examination.

With Philips Healthcare MRI systems all the details about MRI examination (like the sequence of scans, TRs, etc.) are stored in so-called *ExamCard* files. Having an ExamCard file in the system, MRI operator can execute a predefined sequence of scans necessary for particular diagnostics.
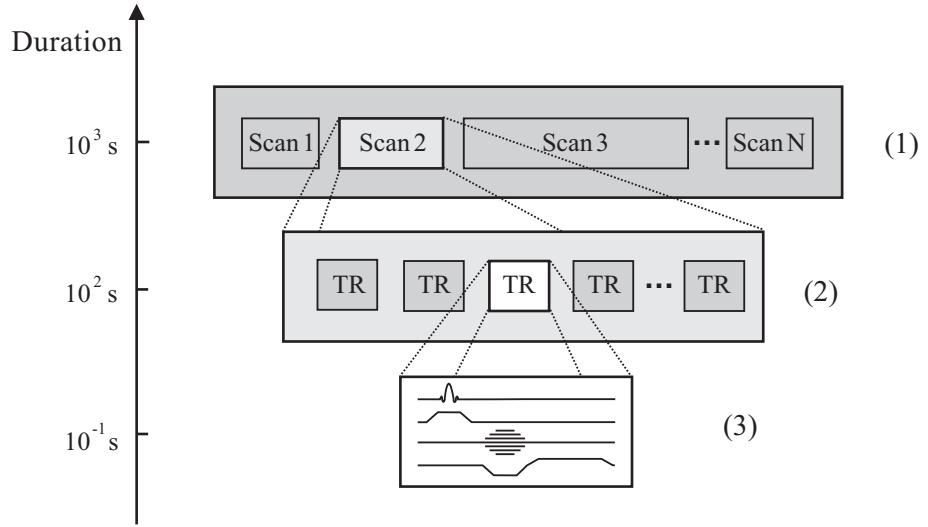
Figure 1.2: MRI examination structure: (1) MRI examination consists of several scans, (2) each scan contains multiple components, TRs, (3) every TR defines a profile of RF pulses and magnetic gradient waveforms.

Typical MRI examinations consist of scans with various performance limitations as well as scans that are not limited at all, which are ordered in a certain sequence. We propose that during MRI examination, MRI scans should be divided into segments and then intermixed in such a way that adverse effects of performance-limited scans are reduced by non-limited scans. Suchwise, the duration of the limited scans is reduced and the overall performance of MRI systems is optimized.

This scans-intermixing method is developed in the following chapters and several scheduling algorithms are described that were designed to deal with different types of performance limitations.

1.3 Improvement of evolvability of MRI systems

Our work was carried out as a part of the DARWIN project at Philips Healthcare, which was focused on improvement of evolvability of Philips MRI systems. In this section the contribution of this thesis to improvement of evolvability of MRI systems is described.

1.3.1 Generations of MRI systems

Durations of MRI scans strongly depend on hardware of the MRI systems, such as amplifiers and coils. This dependency is hardcoded in Philips Healthcare MRI software, e.g. in the part that is responsible for resolving examination performance limitations. For every new MRI hardware the software is modified, e.g. parameters in thermal models are adjusted not to overheat particular amplifier or gradient coil during the MRI scans execution.

In every new generation of MRI systems the hardware is updated, and as a consequence Philips Healthcare software is updated as well. Moreover, a variation of hardware is present even within the same family of the MRI systems. For instance, slightly different modifications of a particular amplifier are used with Achieva 1.5T systems (this is also relevant for Achieva 3T systems). Accordingly, there are a lot of versions of software drivers of the amplifiers together with corresponded thermal models for every particular MRI system type. The fact that every time the MRI software should be adjusted to the hardware reduces *evolvability* of Philips MRI systems, i.e., complicates the task of upgrading generations of the MRI systems. This increases workload on Philips application specialists and can reduce time-to-market of new products.

1.3.2 Simplifying software modification

Modification of MRI software can be simplified by utilizing scan segments intermixing algorithms, which are developed in this thesis, for preprocessing MRI ExamCards. In case of minor differences in hardware of the amplifiers, there is no need to modify the temperature-related parameters in the MRI software, but to modify them in the intermixing algorithms. The ExamCards can be intermixed only one time per amplifier type. This is less time consuming than to patch MRI software to change sequence protocol parameters of every particular MRI system that utilizes that amplifier type.

Modification of gradient coil thermal parameters of the MRI software can be avoided in a similar way, i.e., by preprocessing the ExamCards with the scan intermixing algorithms, see Figure 1.3. Especially when the difference in the temperatures of the coils is minor, that can be a case when different liquid cooling systems are used with the same MRI scanner type. Thereby modification of the values in the software that are related to the coils and amplifiers temperatures can be replaced by preprocessing the ExamCards.

To summarize, the task of MRI software adaptation to new gradient chain hardware can be simplified or even avoided by means of utilizing our intermixing algorithms to preprocess MRI ExamCards. The algorithms developed in this thesis can compensate for minor differences in MRI hardware by intermixing the scans in the MRI examination based on thermal parameters of amplifiers and
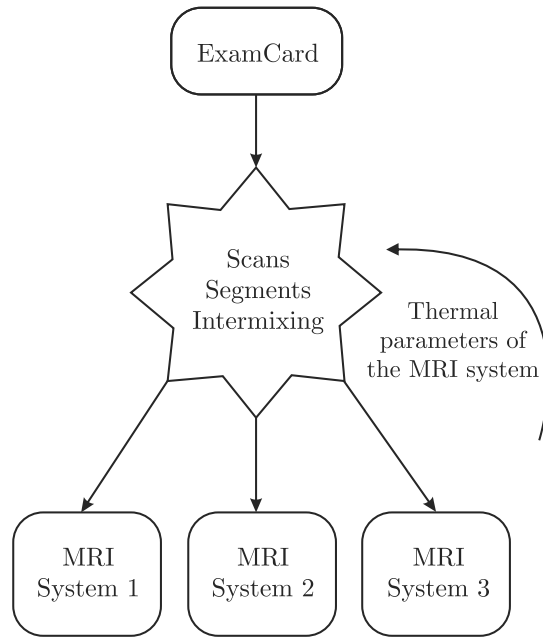
Figure 1.3: ExamCards are preprocessed by scan segments intermixing algorithm that is adaptable to various MRI hardware.

coils. This has a positive impact on the evolvability of the Philips MRI systems. For further reading on the evolvability topic see [Ivanov et al. 2010c; van de Laar and Punter 2010].

1.4   Contribution of this thesis

The main contributions of the research presented in this thesis are:

- A new approach to reduction of examination time of the MRI systems is introduced and investigated.

- Several scheduling algorithms are designed to solve the examination time reduction problem with respect to various MRI performance limitations.

- Queuing theory and discrete-event computer simulations are used to model patient flow in MRI departments in hospitals.

- All these approaches to MRI examination time reduction are validated by 'proof of concept' experiments.

The research in this thesis was carried out as a part of DARWIN project at Philips Healthcare under the responsibilities of the Embedded Systems Institute[1] (ESI).

---

[1]ESI web site: http://www.esi.nl/

This project is partially supported by the Dutch Ministry of Economic Affairs under the BSIK program.

As in all project organized by ESI, one of the goals of the DARWIN project was to provide a 'proof of concept' showing that the proposed approaches and techniques are applicable as well as profitable in industrial practice. In this project, so-called 'industry as laboratory', [ESI 2006], research concept is utilized. This means that the research is performed in close relation with actual system development activities of an industrial partner, which is Philips Healthcare for the DARWIN project.

Such close cooperation with industry has pros and cons, when compared to typical academic research projects. For instance, one of the sources of information for the project are (internal) technical reports that can contain ambiguous and inconsistent information, which should be accepted and dealt with in some way. On the other hand, the answers to all the research questions should be supported by 'proof of concept' experiments that shows their industrial applicability and profitability.

Due to practical context and constraint of this PhD project, this research has a strong engineering component. The engineering nature of this project is also reflected in the thesis.

1.5 Outline of this thesis

This PhD thesis is organized as follows. After an introduction into basic concepts of Magnetic Resonance and MRI system itself (Chapter 2), a formal statement of MRI examination performance optimization problem is given in Chapter 3. In following Chapters 4 and 5, two main instances of the problem are investigated, solutions are elaborated, verified by simulations and validated by 'proof of concept' experiments. Finally, Chapter 6 contains a piece of research that represents a 'helicopter-view' on the problem of MRI examination time reduction. It is focused not only on the duty cycle of the MRI system, but on patient flow in MRI departments of hospitals in general.

A quick overview of content of the chapters is given below:

In Chapter 2
The concept of Magnetic Resonance is introduced and relevant parts of hardware of the MRI scanner are described. Structure of MRI examination and MRI performance limitations are described in details, which are referenced to throughout the dissertation.

In Chapter 3
MRI examination performance optimization problem is posed as a number of scheduling problems. Then a comprehensive literature survey on the related scheduling problems is provided.

In Chapter 4
The first instance of the MRI performance optimization problem is investigated that is related to the SAR level limitations. Several scheduling algorithms are derived to solve the problem. These algorithms were validated on Philips Healthcare ExamCards. The results of the validation are provided at the end of the chapter.

In Chapter 5
The second instance of the MRI performance optimization problem is investigated that addresses thermal limitations of gradient subsystem of the MRI scanner. New scheduling algorithms are designed to solve the problem. Due to the dynamic nature of the thermal limitations, the algorithms are completely different from the algorithms in the previous chapter. The thermal scheduling algorithms are supported by 'proof of concept' experiments carried out on Philips MRI system. Outcome of the experiments is provided at the end the chapter.

In Chapter 6
Patient flow in MRI departments of hospitals is investigated to find bottlenecks and means to reduce average time that patients spend in queues waiting for the MRI examination. The patient flow is modeled twofold: as a multi-phase queueing network and as a discrete-event system. Literature survey on similar queueing problems in health care is provided. Various capacity limitations and patient assignment strategies in the MRI departments are addressed. Finally, several recommendations to improve the patient flow are given.

In Chapter 7
Research of this thesis is summarized, outlined and concluding remarks are drawn. Benefits of the introduced methods and designed algorithms to health care industry are discussed. In the tail-end of the thesis, recommendations for future research are given.

All the chapters were (partly) published in a number of refereed conference papers, a journal paper, and as a book chapter (by Springer). Lists of these publications can be found in the abstracts to the corresponding chapters.

Chapters 2 and 3 are recommended to be read in the first place, since they provide comprehensive background of the problem. There are a lot of references to these

insights throughout the PhD thesis. An exception is Chapter 6 that addresses a distinct view on the MRI examination time reduction problem. That chapter is quite independent from the previous chapters (in particular, from Chapters 2 and 3), therefore it can be read separately. Concept-wise Chapter 6 represents logical generalization of the work in the preceding part of the PhD thesis, therefore it is located at the end of the thesis.

Chapter 2

MRI system overview

The purpose of this chapter is to describe the origins and the framework of the problems we are dealing with in the study. The outline of the chapter is as follows. First, the basic principles of Magnetic Resonance are introduced. Second, the relevant parts of the MRI hardware are described in detail. Third, the MRI scanning process itself together with the ExamCards approach are described. Finally, the duty cycle limitations of the MRI scanning process are described. Several parts of this chapter were published in [Ivanov et al. 2010c].

2.1 Basic principles of MRI

In current section the basic principles of MRI are briefly described. A more detailed explanation can be found e.g. in [Weishaupt et al. 2006].

About 60% of the human body consists of water [Guyton 1991]. Most of the tissues are made up of water for more than 70% (e.g. muscle tissue and blood). The properties and amount of water in tissue vary greatly with disease and injury. Magnetic resonance (MR) is an imaging method that is very sensitive to these variations of water in tissues, thus it can be used for medical diagnostics.

The molecule of water contains two hydrogen atoms. A hydrogen atom (^{1}H) is electrically neutral atom, which consists of a single proton nucleus and has a single electron forming an orbit around the proton. An intrinsic property of all protons is spin, which means that protons rotate about their axes like spinning

tops. Basically, the proton is a rotating mass with an electrical charge, thus it has a magnetic moment. In a similar way many other atoms have magnetic moments. Due to natural thermal motion of atoms, their tiny magnetic fields

(a) $B_0 = 0$, $M = 0$.   (b) $B_0 \gg 0$, $M > 0$.
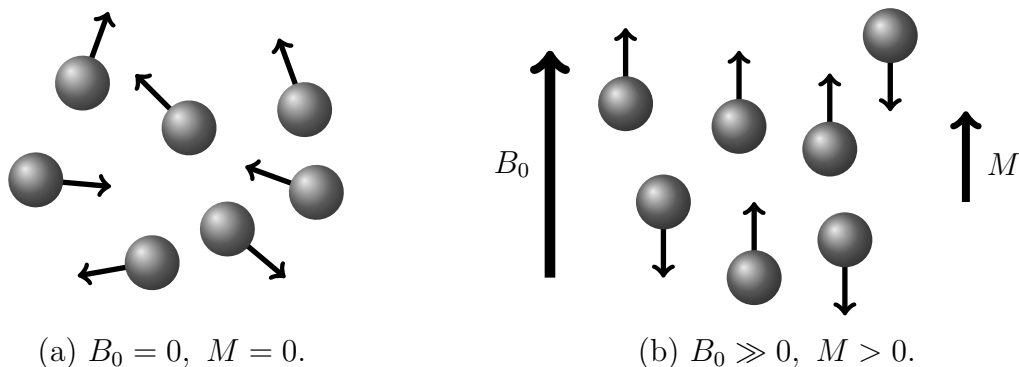
Figure 2.1: Nuclear spins orientation: (a) original, and (b) with applied external magnetic field, B_0. The net magnetization vector, M, increases with the applied field strength.

are pointing in random directions, thus can not be measured. When atoms are exposed to an external magnetic field, B_0, the magnetic moments align with the direction of the field see Figure 2.1. In a similar way the compass needle align with the magnetic field of the Earth. In the field of modern MRI scanner less than 0.001% of the atoms in the patient's body are aligned. Most of the aligned atoms are parallel to the external field, though a part of them is antiparallel to the field. The vector sum of their magnetic moments is called *net magnetization* vector and is denoted by M, see Figure 2.1(b). Practically, it is impossible to measure the net magnetization with current technology, because it is very weak in comparison to the strength of the external field.
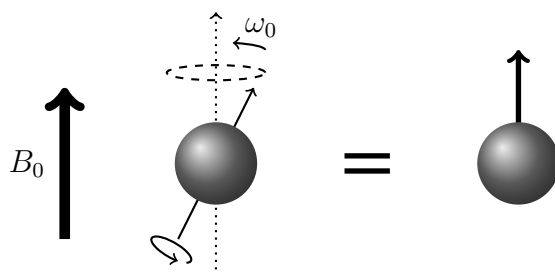
Figure 2.2: Precession of magnetic moment of proton: movement of the axis of the magnetic moment around B_0 axis due to a torque.

A phenomena called *resonance* is utilized to obtain the information about the aligned atoms. The magnetic moments do not only align with the external field but undergo precession, in a similar way as spinning tops in the gravitation field

do, see Figure 2.2. The characteristic speed of nuclei precession is proportional to the strength, B_0, of the external field and is called *Larmor frequency*:

$$\omega_0 = \gamma \cdot B_0, \tag{2.1}$$

where ω_0 is the Larmor frequency in MHz, the magnetic field strength (i.e., magnetic flux density) B_0 is in tesla (T), and γ is a constant specific to particular nucleus (for protons $\gamma_p = 42.58$ MHz/T, [Weishaupt et al. 2006]).

By applying an electromagnetic wave of the same frequency as the Larmor frequency (for protons in 1.5 T field the frequency is $\omega_0 = \gamma_p \cdot 1.5 = 63.9$ MHz) the energy can be introduced into the system. In MRI scanners the required electromagnetic wave is applied by a transmit coil, i.e., antenna. The energy absorption process is called excitation (the protons are exited to a higher energy state) and results in change of the directions of the magnetic moments of the atoms. If the electromagnetic pulse is strong enough and long enough the net magnetization vector, M, can rotate on 90° and become perpendicular to the B_0 field. After the pulse is over, the net magnetization vector, M, starts rotating or precessing about the B_0 field vector, and finally aligns back in parallel with the field. When the magnetic moments of the atoms perform this motion, it has an effect of an electrical generator. Therefore an alternating voltage is induced in a receiver coil of the MRI scanner. This voltage is called the *MR signal*. After processing of the MR signal, the MR images are generated.

The angle to which the net magnetization vector M is rotated or tipped relative to the main magnetic field direction B_0 via the application of a RF excitation pulse is called the *flip angle* and denoted by α. Flip angles between 0° and 90° are typically used in gradient echo sequences, whereas 90° and a series of 180° RF pulses are used in spin echo sequences (see Section 2.3 for details on the pulse sequences).

2.2 MRI system hardware

In this section the MRI system hardware is described. The section starts from a general overview of the hardware. Then, the parts that are relevant for this study are discussed in more detail in the subsections.

In hospitals the MRI system is typically distributed between three rooms of the MRI department. These rooms are: magnet room, technical room, and the control room, see Figure 2.3. The MRI scanner itself is located in the magnet room. An air-conditioned technical room houses supporting electronics, e.g. racks with amplifiers and a liquid cooling system cabinet (LCC). The operator's console is located in the control room. In the latter room the MRI technicians can select parameters of the scans and view or post-process MR images.

Patients are examined in the magnet room. A strong magnetic field imposes high safety standards that must be respected in that room. Especially from ferro-magnetic projectiles, that can damage the scanner or hurt the patient. Thus all the ferromagnetic items, and credit cards should be left out of the magnet room.
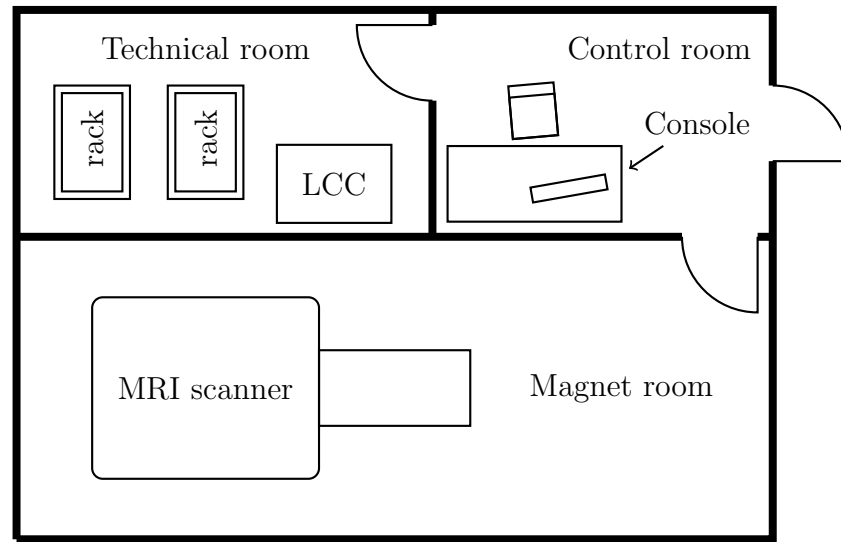
Figure 2.3: MRI rooms plan.

The MRI system hardware consists of:

- Magnet that produces a strong static magnetic field.

- Magnetic field gradient system that consists of gradient coils and the corresponding gradient amplifier.

- Radiofrequency transmit and receive coils together with amplifiers and supporting electronics to excite and detect the MR signal.

- Various computers for acquisition, scanner control, as well as image processing, display, and archiving.

- Patient support table, comfort and positioning system.

- Radiofrequency and magnetic shieldings.

- Physiological monitoring equipment to measure patient's ECG and respiratory cycle.

The components of the MRI system are presented in Figure 2.4. In the following subsections the parts of the MRI system that are relevant for our research are described in more detail.
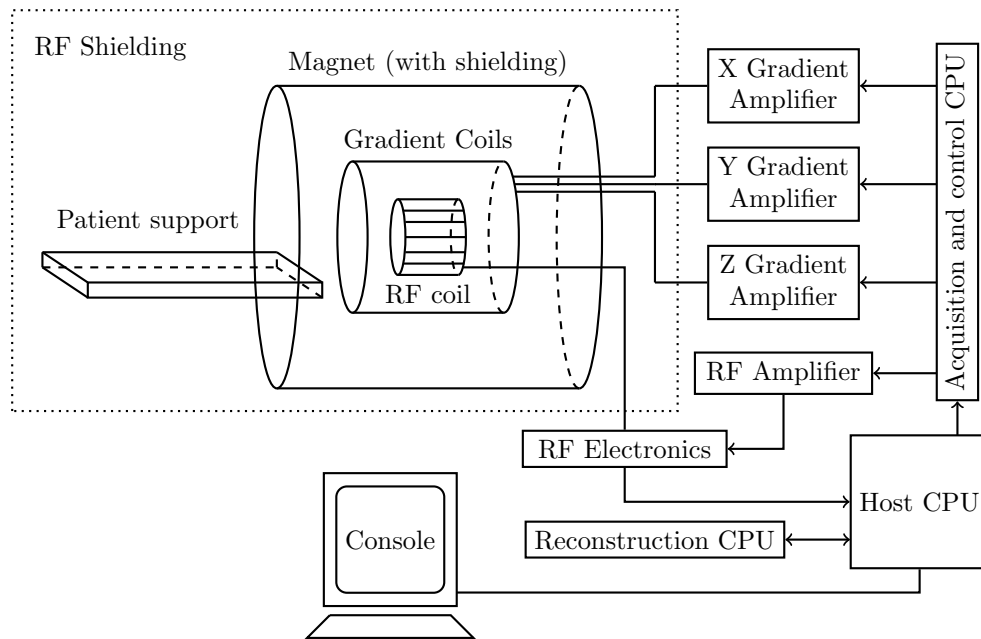
Figure 2.4: Basic MRI hardware.

2.2.1 The magnet

The magnet is the main, and the most expensive component of the MRI system. Magnets for clinical MRI are available in field strengths ranging from approximately 0.01 to 3 tesla (T). Additionally, for research laboratories magnets with even higher fields are available.

For comparison: Earth's magnetic field is 60 000 times weaker than a field of 3 T magnet, which is the most popular one in the modern clinical MRI systems. A higher field strength results in higher net magnetization and potentially higher signals and signal-to-noise ratios (SNR). However, in practice the gains in SNR from the field strength are often offset by losses due to RF attenuation in the patient's body.

The magnets that are used in modern mid and high-field systems (0.5 tesla and higher) are superconducting magnets, which require extreme cooling with liquid helium as a cryogenic fluid. A modern superconducting magnet typically weighs 3000 – 4000 kg together with the cryogens [Mcrobbie et al. 2007].

Another important parameter of the magnet is magnetic field homogeneity that describes the quality or uniformity of its field. This characteristic is crucial to generating images with high SNR and low distortions. After manufacture, due to design constraints the magnetic field is nonuniform, and the magnets are *shimmed* by placing pieces of iron into the bore to optimize homogeneity. In modern systems also *dynamic shimming* is used: inhomogeneities caused by patient presence are corrected with gradient coils during the scan execution.
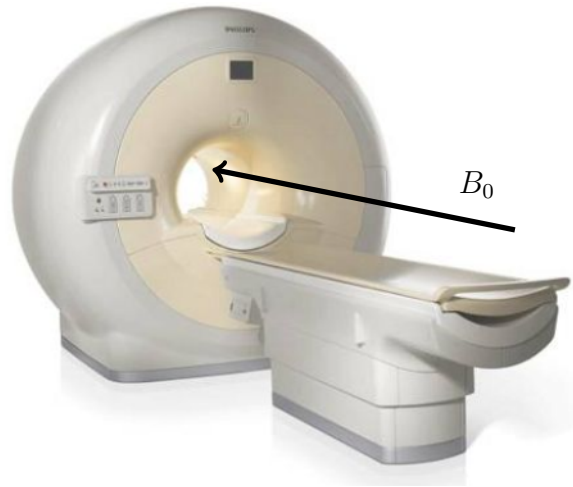
Figure 2.5: Philips Achieva 3.0T closed bore system. (Image is from http://www.healthcare.philips.com.)

The main field vector, B_0, of the magnet points either horizontally or vertically depending on whether the MRI system is with *closed* or *open bore*. The closed bore systems with cylindrical in shape superconducting magnets are the most common ones in clinical diagnostics, see Figure 2.5. However, some patients can experience claustrophobia from being inside the bore of these magnets. That is why the open bore systems, see Figure 2.6, are also popular.

Figure 2.6: Philips Panorama HFO open bore system. (Image is from http://www.healthcare.philips.com.)

2.2.2 Magnetic gradient system

The magnetic field gradient system is also known as *Gradient Chain*. It consists of gradient coils, gradient amplifiers and the cooling system for the amplifiers and the coils. The components of the magnetic gradient system are described below.

Gradient coils

The localization of MR signal in the body is accomplished by generating spatial variations in magnetic field strength across the patient. The gradient fields are produced by a set of three separate gradient coil layers in the bore of the magnet. These coils alter the magnetic field strength linearly along the orthogonal axises, namely x-, y-, and z-axes. Gradient coils are built into the bore of the magnet. In a typical cylindrical closed magnet the direction along the bore is called z-axes, the left-right direction is called x-axis, and the top-bottom direction is called y-axis.

(a) X coil   (b) Y coil   (c) Z coil
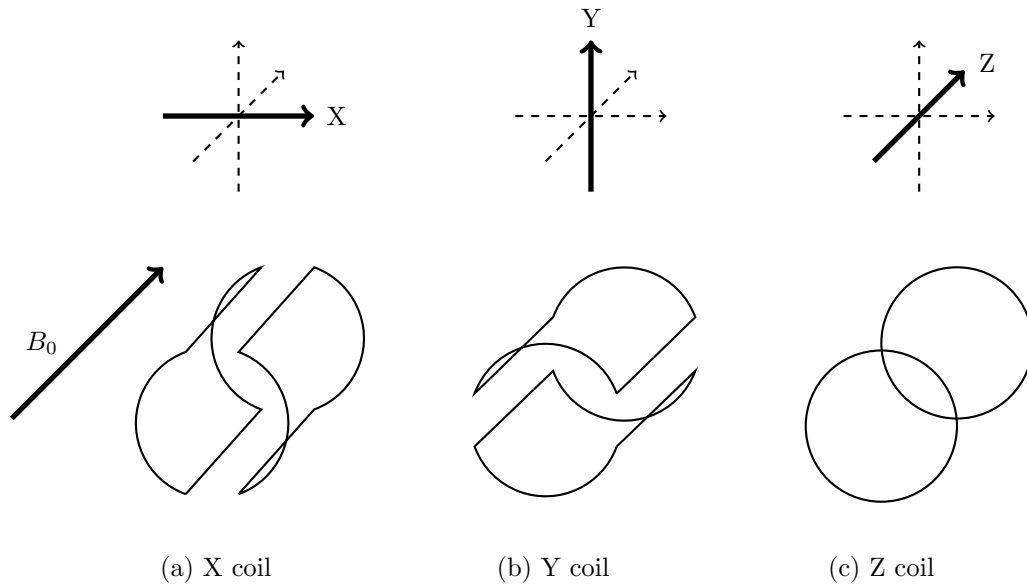
Figure 2.7: Gradient coils components design for cylindrical closed bore MRI systems.

Linear variation in field along the z-axis is produced by a so-called Maxwell coil, which is a pair of coils separated by $\sqrt{3}$ times their radius, see Figure 2.7(c). The current in these two coils flows in opposite directions that produces very linear gradient. The linear gradient in x- and y-axes is usually produced by sets of so-called Golay coils, see Figures 2.7(a) and 2.7(b). One can find a detailed description of the gradient coils design, e.g. in [Jin 1999].

During the pulse sequences the gradients are rapidly switched. Therefore, eddy currents are induced in nearby conducting components, such as magnet cryostat and other coils. This also results in a 'clanging' noise on which patients often complain. Special pre-emphasizing of the gradient waveforms is performed to reduce the eddy current effects.

Gradient amplifiers

Gradient amplifiers are parts of magnetic gradient system. These are typically high-power audio frequency amplifiers. The same kind of amplifiers are used in concert sound systems, see Figure 2.8. Gradient amplifiers generate electrical currents of several hundred Amperes together with voltages that need to be applied to the coils to produce the magnetic gradient fields.

       (a) Amplifier   (b) Rack with amplifiers
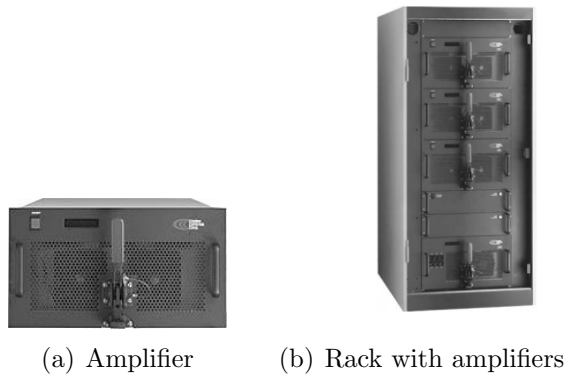
Figure 2.8: Gradient amplifiers: (a) Copley high-power audio frequency amplifier, (b) rack with liquid-cooled amplifiers. (Images are from http://www.copleycontrols.com.)

Different types of amplifiers can be used together with MRI systems. The most common amplifier type is switch-mode amplifier, which consists of several H-bridges[1] combined together in various structures, see e.g. [Li et al. 2008]. In a simple case the H-bridges in the amplifier are placed in parallel, and every bridge delivers an equal current.

In Philips Healthcare MRI systems, one or two racks with the gradient amplifiers are located in the technical room, see Figure 2.3. Each rack contains three amplifiers corresponding to the x-,y-,z- gradient coils, see Figure 2.4. In case of two racks the amplifiers are usually connected in parallel, in order to double the performance.

---

[1]An H-bridge is an electronic circuit which enables a voltage to be applied across a load in either direction. The circuit diagram resembles the letter 'H.'

Cooling system

Due to the high currents, extra power is dissipated in the gradient coils and amplifiers. Consequently they require cooling, either by air or by water. In Philips MRI systems the water cooling is used both for gradient coils and amplifiers. In the Philips Achieva MRI system, see Figure 2.5, only z-coil is cooled directly, whereas x- and y-coils are cooled indirectly. A Liquid Cooling Cabinet (LCC) is located in the Technical room, see Figure 2.3. It supplies cooling fluid (water or helium) with appropriate temperature, flow and pressure to the gradient z-coil, gradient amplifiers and helium compressor.

2.2.3   Radiofrequency coils

The radiofrequency (RF) coils are used for excitation of the nuclei within the patient's tissue (transmit coils) and for MR signal detection (receive coils). Often the same coil is used both for transmitting and for receiving. A body RF coil is usually integrated into the bore of the MRI scanner, which is enough for preliminary scans and some general body examinations. For head and extremities examinations, as well as body examinations that require higher contrast, additional RF coils are used, see Figure 2.9.

(a) Quadrature head RF coil.   (b) Surface abdominal RF coil.

Figure 2.9: Radio frequency coils: (a) volume coil, and (b) surface coil. (Images are from http://www.healthcare.philips.com.)

These coils are positioned close the examined region of the body, which yields a better SNR than the standard body coil (up to 3-5 times better, [Philips 2005]). There are many different designs of RF coils. However, most of them can be divided into two categories: volume coils and surface coils. Surface coils are positioned on top of the region being imaged, see Figure 2.9(b). They are good

in imaging the areas which lie close to the surface of the body. If the region of interest is far from the surface, then a volume coil is required. The volume coils fit the entire region, such as head, see Figure 2.9(a).

The MR signals are very weak, thus the RF receive coils are high sensitive. If the same coils are used for transmitting and receiving, then a transmit/receive switch is used, which protects the receiver circuitry from the high voltages applied during the transmitting.

The generated RF pulses have frequencies determined by the Larmor equation (equation (2.1) in Section 2.1): from 63 MHz for 1.5 T systems, up to 128 MHz for 3 T systems. These frequencies are in the range of FM transmitters. Therefore special RF shielding is built into the walls of the magnet room (i.e., Faraday cage) to prevent adverse electrical interference from external sources. Even the door in the magnet room should be closed during the examination to complete the Faraday cage.

2.2.4   Computer system

There are three computers in an MRI system to control and coordinate all the processes. The acquisition and control computer is used for controlling scanner hardware (e.g. gradient, RF systems and the physiological monitoring equipment). The reconstruction computer is used for creation of 2D and 3D MR images out of the raw MR data. The host computer is used for general coordination of all the processes, see Figure 2.4. The operator's console and image archives are connected to the host computer.

2.3   The scanning process

In this section the MRI scanning process is described. The concepts of slice selection, spatial encoding, and pulse sequences are introduced.

2.3.1   Slice selection and Spatial Encoding

In typical 2D imaging, the MR scanner produces cross-sectional images of the patient body. The excitation RF pulse is delivered only to the *slice* to be imaged, but not to the whole body. Therefore the patient body (or the region of the body) is imaged slice-by-slice. The thickness of the slice is selected to be smaller than the smallest structure to be imaged. Usually a couple of millimeters is enough. However, in modern 3 T MRI the slice thickness can be less than a tenth of a millimeter. (In 3D imaging the entire volume is excited instead of the slices, however the 3D MR images are out of scope of this section.)

It was mentioned in Section 2.2.1 that in closed bore MRI scanners the magnetic field vector, B_0, is directed along the body axis of the patient being imaged, see Figure 2.5. In Section 2.2.2 the x-, y-, and z-gradients were described.
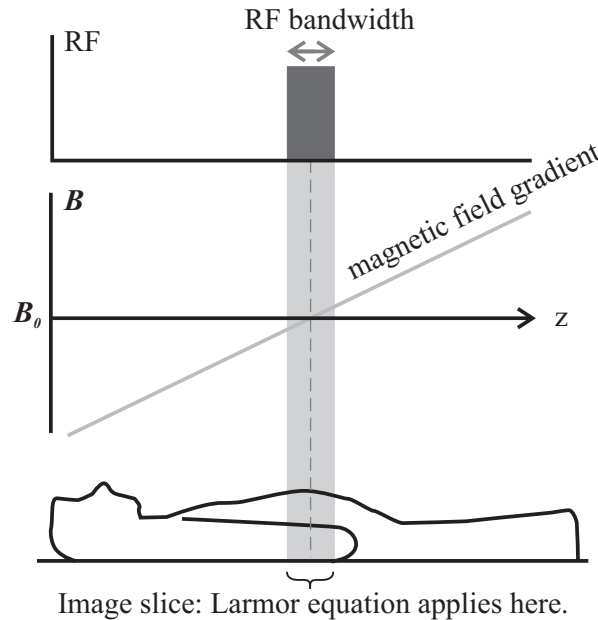
Figure 2.10: Selective excitation of an image slice.

The *slice selection* process is based on the magnetic resonance phenomenon and particularly on the Larmor equation, see Section 2.1. According to equation (2.1), the Larmor frequency, ω_0, linearly depends on the magnetic field strength, B_0. The z-gradient coil creates linear gradient in the magnetic field along the patient body (z-axis). Therefore the Larmor frequencies of the nuclei also become different along the patient body. When a narrow bandwidth RF pulse is applied, only the protons in the particular slice of the body are excited, which Larmor frequencies are equal to the frequency of the RF pulse, see Figure 2.10. The z-coordinate of the excited protons is then known. In order to get a complete image of the region of the body the slice-selection process is repeated several times, and each time the RF pulse frequency is slightly shifted. A time period TR (Repetition Time) occurs between the application of one RF pulse and the next.

The x- and y- gradients are used for the *spatial encoding* of the MR signal emanating from the excited slice. This process is also based on the Larmor equation. The y-gradient is used for *phase encoding*. For every slice the y-gradient pulse is applied a number of times, each time with different amplitude. When the pulse is applied, the phases of the spins are changed: the individual protons have their phases altered according to their relative position along the y-axis. The higher the amplitude of the y-gradient the more evident is the difference in phases of the

spins. Then the x-gradient is applied during MR signal acquisition to encode signals into different frequencies, depending on their position toward the x-gradient, i.e., *frequency encoding*. For more details on slice selection and spatial encoding processes see, e.g. [Mcrobbie et al. 2007; Weishaupt et al. 2006].

Spatial domain k space
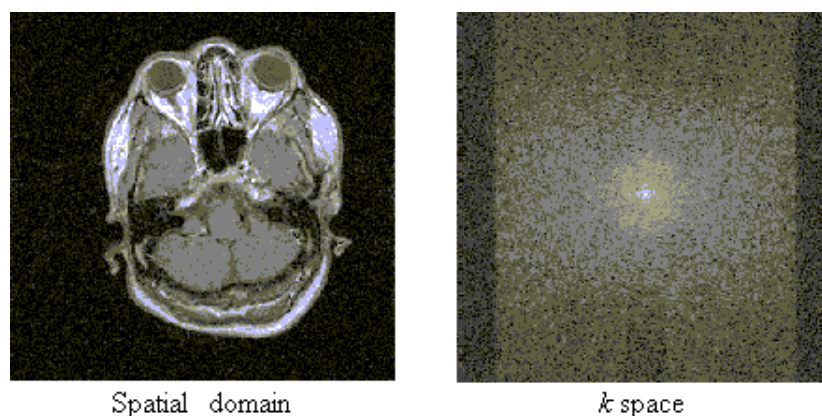
Figure 2.11: Spatial domain and the corresponding k-space. (Images are from http://m-pss.org.)

Raw MR data are stored in a table of phase amplitudes against acquired MR signals. This table represents a so-called *k-space*, which was first introduced by [Twieg 1983]. The k-space has the same number of columns and rows as the final image does, see Figure 2.11. During the scanning process, the k-space is filled with the raw data one line per TR. After the table is filled, 2D Fourier transformation is applied. Therefore, the spatial coordinates of the excited protons are extracted out of the raw data and the MR image is generated.

2.3.2 Pulse Sequences

In order to obtain MR images, special *pulse sequences* are executed on the MRI systems. There are several purposes of these pulse sequences: to generate contrast between different tissues, to provide spatial localization, to excite atomic nuclei, and to be sensitive to some dynamic parameters, e.g. blood perfusion or contrast liquid flow. Moreover, the artifacts should be avoided, and the atomic region should be covered within a particular time frame, such as breath-hold time.

Every pulse sequence is a set of defined RF and gradient pulses to excite a slice and localize the MR signal. MR pulse sequence timing diagrams are used to demonstrate time intervals between RF pulses, their flip angles (see Section 2.1) and amplitudes and shapes of the gradient waveforms. In Figure 2.12 one can see the diagram for a Gradient Echo sequence. Each line of the diagram belongs to a different hardware component. The first line is needed for the RF transmit coils, the next three lines are for gradient coils (slice selection z-gradient,
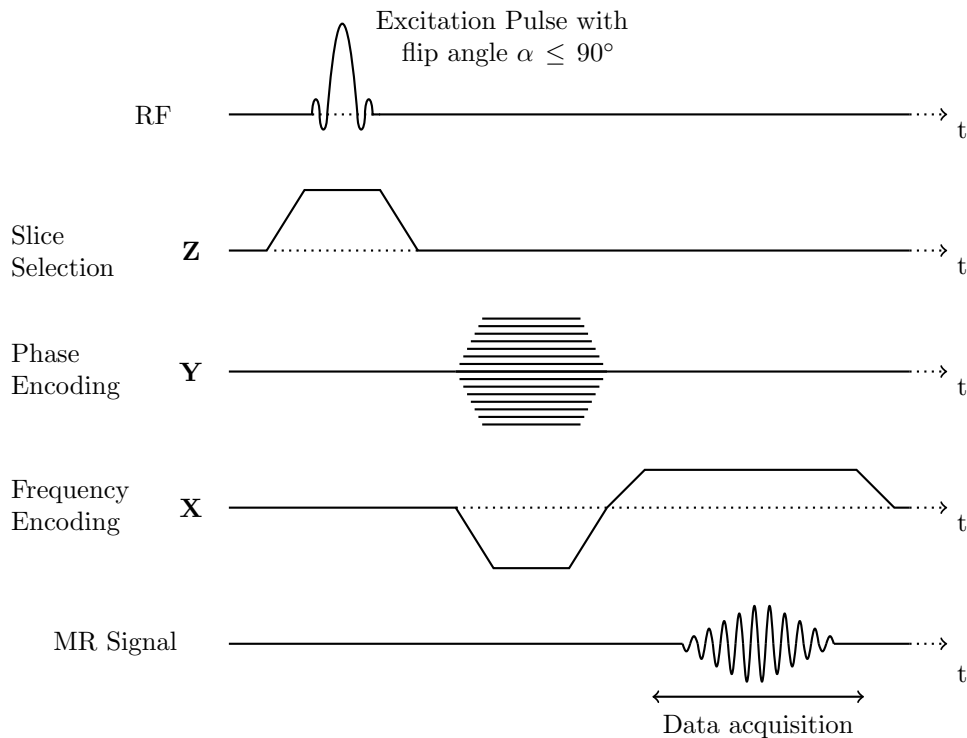
Figure 2.12: Pulse sequence diagram for a Gradient Echo sequence (one TR).

phase encoding y-gradient, frequency encoding x-gradient), the last (optional) line demonstrates of MR signal (after sampling) that is detected by the RF receive coils. On each line, the vertical axis corresponds to the amplitude and the horizontal axis to the time. Gradient Echo, see Figure 2.12, along with Spin Echo and Inversion Recovery are mainstays of MRI — the majority of advanced pulse sequences are based on these three simple sequences.

Every MR pulse sequence consists of several steps:

- **Excitation of the target region**. Achieved by combining the slice selection gradient (usually z-gradient) and excitation RF pulse with a particular flip angle α.

- **Phase encoding**. The phase encoding gradient (usually y-gradient) is switched repeatedly, each time with a different strength, to create the number of phase shifts across the image.

- **MR signal (echo) formation**. Achieved either by applying a RF pulse (Spin Echo sequences), or by reversing the direction of the magnetic field gradient (Gradient Echo sequences).

- **Data acquisition**. First, frequency encoding gradient (usually x-gradient) is switched, then the MR signal is recorded.

In Gradient Echo sequences the target region is excited by a RF pulse with flip angle $\alpha < 90°$, and the Echo is formed by a gradient pulse, which reverse the direction of the magnetic field gradient. In Spin Echo sequences the slice is excited by a RF pulse with flip angle $\alpha = 90°$, whereas the echo formation is achieved by applying one or more 180° RF pulses. Inversion Recovery pulse sequence is similar to Spin Echo one, but an additional 180° RF pulse (inversion pulse) precedes the excitation 90° RF pulse, which has some benefits, as to enhance the contrast of particular tissues while suppressing the signals of the other ones.

A lot of advanced pulse sequences, with more sophisticated RF pulse profiles and gradient waveforms, are used in MRI clinical practice, and many of them are based on the aforesaid three sequences. Some additional techniques are used either to increase the contrast or to reduce MRI scan time. For instance, a *multislice imaging* technique is used to reduce the scan time: basically, several slices are excited during one repetition time. Therefore, more than a dozen of slices instead of only one can be acquired in the time of one TR. Consequently, the number of RF pulses per TR in such a sequence is much higher, which yields higher RF deposition in a patient. Some other advanced pulse sequences include gradient waveforms with high amplitudes and/or short periods, which can result in temperature increase of the gradient hardware components. Before an MRI examination starts, all the pulse sequences are checked not to overheat the hardware and not to harm the patient. Otherwise the pulse sequences are modified to comply with some safety guidelines, which often results in reduction of MRI duty cycles. These duty cycle limitations are discussed in detail in Section 2.5. Structure of typical clinical MRI examinations is discussed in Section 2.4.

2.4 MRI examination structure

2.4.1 Components of MRI examination

Nowadays a typical clinical MRI examination takes about 10 to 30 min and consists of several different scans, i.e., scanning protocols. Each scanning protocol describes a pulse sequence, see Section 2.3.2. This pulse sequence is repeated a lot of times[2]. For the Spin Echo based pulse sequences, the *repetition time* (TR) of this pulse sequence is typically 200 to 2000 milliseconds (ms). For the Gradient Echo based sequences the TR can be as short as 20 ms.

According to [Bernstein et al. 2004], the MRI scan time T_{scan} can be calculated as follows:

$$T_{scan} = \text{TR} \times N_{phase} \times \text{NEX} \times N_{acq}, \qquad (2.2)$$

---

[2]Usually, the number of repetitions is a factor of 256, because fast Fourier transform algorithm (for 2D Fourier transform) is efficient for data sets of power-of-two size.

where N_{phase} is the number of phase-encoding gradient steps, NEX is the number of *signal averages*, and N_{acq} is the number of *passes*, i.e., acquisitions to obtain all the required slices of the patient body.

The signal averaging is performed to suppress the effects of random variations or random artifacts: each measurement is repeated NEX times and the results are averaged. The number of passes N_{acq} is the rounded ratio of how many slices are to be excited and how many of them can be excited per one TR. This value also depends on the duty cycle limitations, which are described in Section 2.5.
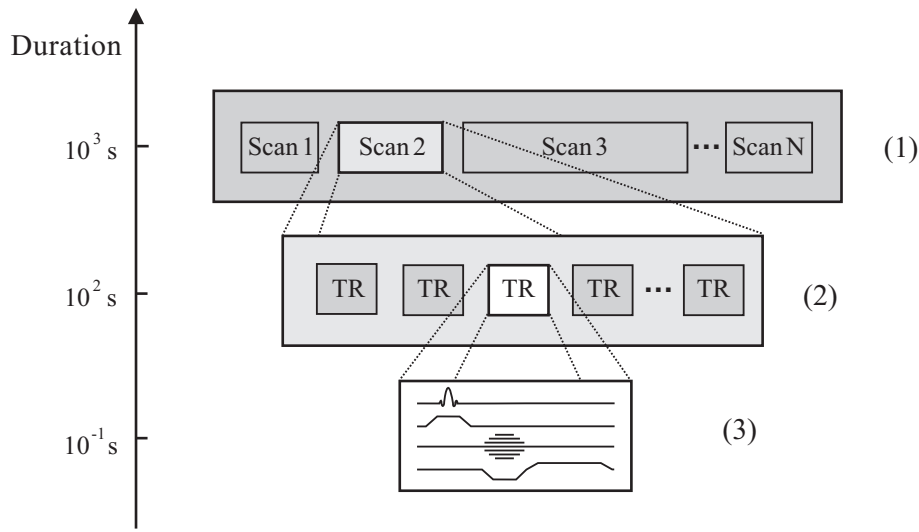
Figure 2.13: MRI examination structure: (1) MRI examination consists of several scans, (2) each scan contains a lot of repetitions of a pulse sequence of length TR, (3) the pulse sequence is a set of defined RF and magnetic gradient pulses.

A hierarchical structure of a typical MRI examination is presented in Figure 2.13. The vertical axis (at the left hand side) displays the order of magnitudes of the durations of the MRI examination components. The duration of the MRI examination itself is in order of dozens of minutes (10^3 s), the duration of the scans is in order of minutes (10^2 s), and the average duration of pulse sequences is in order of hundreds of milliseconds (10^{-1} s).

To be more precise, the average and the maximum duration of one MRI scan is 2 min and 23 min respectively. The average duration of the MRI examination is about 20 minutes, whereas the duration of 90% of MRI examinations ranges from 10 to 30 minutes. This is the so-called *gross* duration of the MRI examination, which includes in-between scan setup times (the setup times will be discussed in Chapter 3). For more detailed statistics on the durations of MRI examinations and the scans, see Table 2.1. Since the distributions are neither normal nor symmetric (they are positively skewed), the percentile ranges are used instead of

Statistic	Examination duration	Scan duration	Number of scans
Mean	21 min	2 min	8.5
Max	> 1 hour	23 min	> 20
90%	10 − 30 min	< 4 min	5 − 14
95%	10 − 40 min	< 5 min	5 − 17
99%	< 1 hour	< 7 min	< 20

Table 2.1: Descriptive statistics of duration of the MRI examinations, duration of the scans, and number of scans per examination, respectively. Since the distributions are neither normal nor symmetric (they are positively skewed), the percentile ranges are used instead of the variances.

the variances. These statistics are based on 3 months data set from a hospital in The Netherlands. There are 2 MRI scanners in that hospital, and the data set contains more than 2000 MRI examinations with 8.5 scans per examination on average.

2.4.2 MRI examinations with ExamCards

There are many kinds of MRI examinations that can produce 2D and 3D images of different parts of patient's body. For a given MRI examination the scanning region of the patient's body generally remains the same during the whole examination. For example, in a brain examination the scanning region is the brain, and the different scans during that examination obtain images of different parts of the brain. Therefore, the RF coils are assembled only once, before the examination starts, and in-between-scans time is relatively small.

Before MRI examination starts, a set of scans necessary for the diagnosis must be defined in the MRI software. In old generations of Philips MRI scanners technologists in hospitals had to select each scan manually from a large list of options. This is so-called 'scan-by-scan' MRI examination planning, which is not time-efficient.

Not all the scans during the MRI examination are necessary for the radiologist to make the diagnosis itself. There is always some amount of preliminary technical scans, such as survey and reference scans at the beginning of each examination. However, those scans do not take much time, e.g. a survey scan is about 15 s long.

Nowadays, routine MRI examinations in Philips Healthcare are facilitated by the *ExamCards* concept. An ExamCard is a binary encoded executable file which contains instructions for the Philips MRI scanner and all the information on the various scans necessary for the MRI examination. This information (e.g. scans,

duration, TR-length, SAR[3] levels) can be extracted from those files and converted to readable XML-files (dump files) with special Philips Healthcare software.
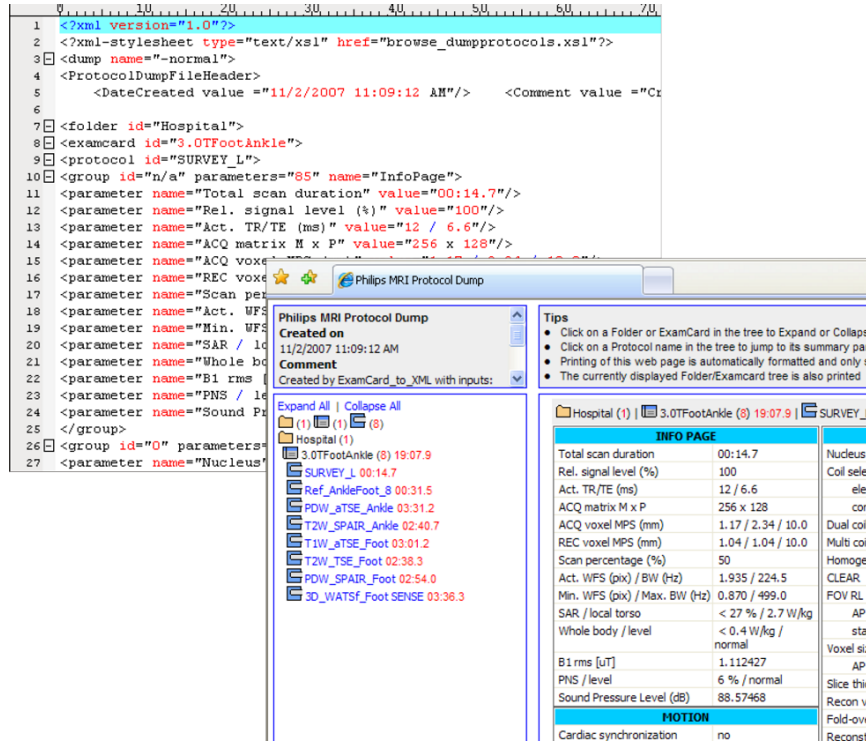
Figure 2.14: XML-dump ExamCard file and Philips Healthcare software.

In Figure 2.14 we can see an example of a 3.0 T Foot Ankle ExamCard. The upper window is the dump XML-file, and the lower window is the Philips software used to display the file in a structured way. On the left hand side of the lower window we can see the names and the durations of the scans in the ExamCard, one can also see them in Table 2.2. On the right hand side of the lower window we can see the detailed parameters of the scans (e.g., the SAR levels and TR lengths). The Ankle examination is 19 min long and consists of 8 scans:

- Two preliminary scans (`SURVEY L` and `Ref AnkleFoot 8`),

- Six regular scans (from `PDW aTSE Ankle` to `3D WATSf Foot SENSE`).

The duration of the preliminary scans is less then one minute, whereas the average duration of a regular scan in the Ankle examination is about 3 min, see Table 2.2. This complies with the typical durations of the MRI scans, which were presented in the previous section, see Table 2.1.

---

[3]The term SAR will be explained in Section 2.5.1.

Protocol name	Protocol duration [min]	TR length [ms]	relative SAR [%]
SURVAY L	00:14.7	12	27
Ref AnkleFoot 8	00:31.5	4	4
PDW aTSE Ankle	03:31.2	3200	66
T2W SPAIR Ankle	02:40.7	2678	82
T1W aTSE Foot	03:01.2	737	100
T2W TSE Foot	02:38.3	2932	100
PWD SPAIR Foot	02:54.0	3000	62
3D WATSf Foot SENSE	03:36.3	20	13

Table 2.2: 3.0 T Foot Ankle ExamCard. The *SAR* term will be explained in Section 2.5.1.

The ExamCards that were utilized in our research were downloaded from NetForum, [NetForum 2004–2010] a web facility where hospitals that use Philips MRI systems can share their routine ExamCards. Those ExamCards were created by hospitals for their own purposes, and there is a high probability that they represent a set of the ExamCards that are actually used in hospitals on a daily basis.

Currently, most of the MRI hospitals use the ExamCards with the Philips systems, whereas the rest still use the old-fashioned scan-by-scan MRI examination planning. Philips Healthcare recommends the hospitals to switch to the new ExamCards-based workflow because it results in significant examination time decrease. This is because in the old-fashioned scan-by-scan way radiologists lose a lot of time on selection of the scans and tweaking their parameters before the MRI examination start. Also in the new ExamCards the modern techniques for speeding up the examination are utilized, e.g. pMRI [Blaimer et al. 2004].

Practically, the results of our research are only applicable to the MRI hospitals that use ExamCards or other predefined sets of scans for the MRI examinations. This is because the proposed algorithms need to know all the scans that will be in the examination before the actual examination starts.

2.5   Duty cycle limitations

In clinical practice the MRI examinations consist of different scans (i.e., scanning protocols). The detailed structure of the scans is discussed in Section 2.4.

Some types of scans can overheat gradient amplifiers or coils, whereas others can exceed SAR level limitation (SAR stands for Specific energy Absorption Rate, it will be discussed in detail in next section). These types of scans are called *gradient-limited* and *SAR-limited* respectively. In current practice, an amount of

dead time is included into every TR of the limited scan to keep the temperature of the amplifiers, coils and the SAR level below the limits. During those dead times the system does no useful job, but waits idle for the SAR level or the temperature to go down.

The MRI scan time is determined by equation (2.2) in Section 2.4.1. The insertion of the dead time is accomplished either by reducing the number of slices per TR (thus increasing the N_{acq}) or by extending the duration of the TR. Anyway, this results in extension of the scan time and, consequently, the overall MRI examination time.

2.5.1 Specific energy absorption rate

The *Specific energy Absorption Rate* (SAR) level is the main factor that influences the temperature of patient's body during the MRI examination. This is a quantity to measure 'warming' effect of RF electromagnetic fields in the frequency range 100 kHz – 10 GHz. The 'warming' effects of low frequency (< 100 kHz) electric and magnetic fields are negligible, [ICNIRP 1998].

From a certain point of view, the electromagnetic effects in the bore of MRI-scanner are the same as in a microwave oven. The excitation RF pulses are generated by RF transmit coils during the scan execution (see Sections 2.2.3 and 2.3.2). These pulses increase the radiation level that can result in heating patient's body.

The SAR indicates how much RF power is being dissipated in patient's body and it is measured in watts per kilogram of tissue (W/kg). SAR can be calculated from the RF field within the tissue as follows:

$$SAR = \frac{\sigma E^2}{\rho}, \qquad (2.3)$$

where σ is the sample electrical conductivity, E is the root mean square of RF field B_1, and ρ is the sample density [Jin 1999]. Therefore, SAR is proportional to the squared RF transmit field B_1 averaged over time. In turn, it can be calculated from amplitudes and durations of the RF pulses in the scan, and how many there are per TR.

Body temperature rises of more then $1 - 2$ °C can have adverse health affects such as heat stroke and heat exhaustion, see [ACGIH 1996]. On the other hand, there have been several experimental studies with human volunteers exposed to RF field in MRI systems, see [Magin et al. 1992; Shellock and Grues 1987], that demonstrated that exposure for up to 30 min, under conditions in which SAR level was less then 4 W/kg, caused an increase in the body temperature of less than 1 °C.

To avoid harmful effects on the patient health, the International Electrotechnical

Averaging time	6 minutes				
	Whole body SAR	Partial body SAR	Head SAR	Local SAR	
Body Region $\rightarrow$	whole body	exposed part	head	head/trunk	extremities
Operating Mode $\downarrow$	(W/kg)	(W/kg)	(W/kg)	(W/kg)	(W/kg)
Normal	2	$2-10$	3.2	10	20
1st Level Controlled	4	$4-10$	3.2	10	20
2nd Level Controlled	>4	$>(4-10)$	>3.2	>10	>20
Short-term SAR	The SAR limit over any 10 s period shall not exceed three times the stated values				

Table 2.3: IEC standard on SAR limits.

Commission (IEC), based on the aforesaid and some other studies, has adopted limits for a safe exposure to the RF energy emitted during the MRI examination. The IEC posed both long-term and short-term limits of the SAR level during the examination, [CEI/IEC 60601-2-33 2008]. The long-term limits prescribe a maximum SAR level averaged over the duration of the MRI scan (to be more precise, the averaging time is 6 min). This restricts the body tissue temperature rises to 0.7 °C or 1 °C for *normal* and *1st level controlled* operation modes, respectively. For instance, for the whole body scans *1st level controlled* mode the maximum SAR level is 4 W/kg, see Table 2.3. The short-term limits in turn prescribe that the SAR over any 10 s period shall not exceed 3 times its long-term limits (thus 12 W/kg for the whole body scans).

Before an MRI examination starts, all the scans within the examination are tested to comply with the SAR limits. If the SAR level of a scan exceeds the limits, the pulse sequences of the limited scan are modified to fulfill the IEC regulations.

There are a few ways to decrease the SAR level of a SAR-limited scan. If originally short RF pulses with high SAR level were used, then the geometry of the RF pulses can be modified. The SAR level is proportional to the square of the RF field B_1, which is in turn proportional to the square of the Larmor frequency or B_0 and the flip angle α:

$$\text{SAR} \ \propto \ B_1^2 \ \propto \ \alpha^2 B_0^2. \tag{2.4}$$

Since the SAR level is proportional to α^2, the longer RF pulses with lower flip angles can be used during the scan execution to reduce the SAR level, see Figure 2.15. However these manipulations can potentially affect the contrast of the MR image, see [Gai and Zur 2007].

If the modification of the shapes of the RF pulses does not result in a sufficient SAR decrease, then another method is applied: some amount of dead time is added after each TR of the SAR-limited scans in order not to exceed the SAR constraints. The disadvantage of the latter method is that it results in prolongation of the MRI examination time.
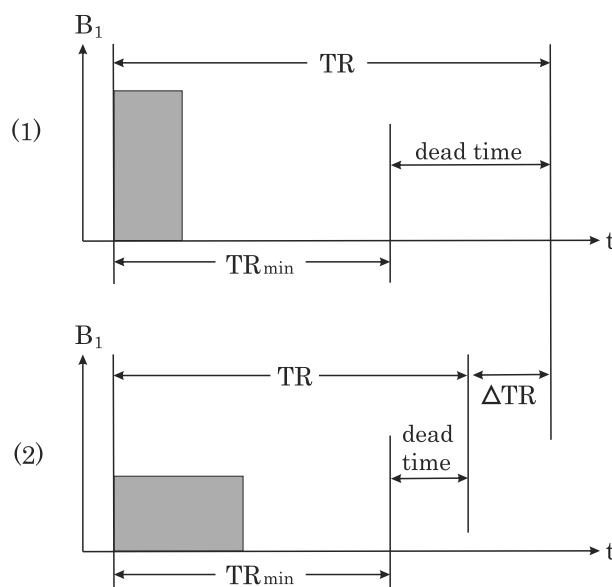
Figure 2.15: Changing the shape of the RF pulse can decrease the SAR level and hence reduce the dead time: (1) RF pulse with a standard flip angle, (2) RF pulse with a lower flip angle.

Philips Healthcare uses specific formulas to forecast the SAR level during the MRI examination. The formulas that are used in the Philips MRI software are more conservative than the IEC limitations. For example, the 'worst case' of the MRI patient is assumed (from the anatomic point of view). Also some delays during scan execution are excluded from the SAR level calculations (such as waiting time due to synchronization with heart beat or breathing).

Another aspect of the Philips Healthcare MRI software, is that the SAR levels of the scans are represented in percentages with respect to the corresponding IEC limitations. This is the so-called *relative SAR level*; an example of the relative SAR levels of scans from an Ankle ExamCard one can see in Table 2.2 on page 28. This notation will be frequently used in Chapter 4.

There is a tendency to increase the strength of the magnets in new generations of MR-scanners. For instance, the modern generation of MRI scanners use 3 T and even 7 T magnets instead of 1.5 T ones (see Section 2.2.1). The SAR level increases with the square of the magnetic field B_0, see equation (2.4). In consequence, the number of SAR-limited clinical MRI scans is expected to increase.

2.5.2 Gradient amplifiers temperature

The gradient amplifiers generate high currents, whereas the gradient coils transform these currents into magnetic gradients. A detailed description of the mag-

netic gradient subsystem is in Section 2.2.2.

During execution of the MRI examination the gradient amplifiers should not be overheated, i.e., the temperature of transistors in the gradient amplifiers must not exceed some limit

$$T_{\text{amp}} \leq T_{\text{amp}}^{\max}, \tag{2.5}$$

which is specified by the amplifier manufacturer. The temperature of the transistors depends on the input current.

Special liquid cooling systems are used to keep the temperature of the amplifiers below the limitation, see Section 2.2.2. In case of overheating, the scan execution is aborted in order not to damage the amplifiers hardware. The scan abortion is extremely undesirable, thus special mechanisms are utilized by Philips Healthcare to avoid it.

Different types of amplifiers are used together with MRI systems. The most common amplifier type is a switch-mode amplifier, which consists of several H-bridges[4] combined together in various structures, see e.g. [Li et al. 2008]. In a simple case, the H-bridges in the amplifier are placed in parallel and every bridge delivers an equal current.
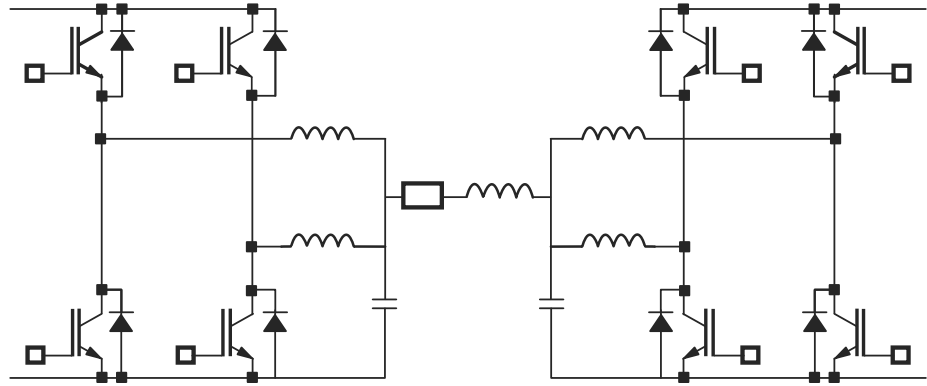
Figure 2.16: H-bridges in parallel structure.

A duality between heat transfer and electrical phenomena is well-known, see e.g. [Cengel 2002]. For each transistor in the H-bridge, the junction temperature can be approximated by means of a special modeling technique called Resistor-Capacitor (RC) thermal networks, see e.g. [Hauck and Bohm 2000]. The RC thermal network of transistor consists of a sequence of $i = 1 \ldots k$ elementary RC thermal circuits, where k is the number of sections in the transistor. The elementary RC thermal circuit with heat capacitance of C_i and a heat resistance of θ_i is presented in Figure 2.17. The heat power fed to the system is denoted

---

[4]For the definition of an H-bridge see the footnote on page 18.
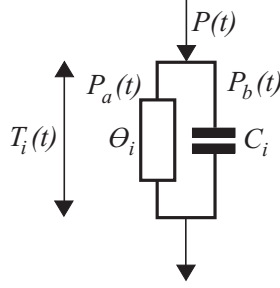
Figure 2.17: Elementary RC thermal circuit.

by $P(t)$, and the temperature difference over the ith RC circuit (section of the transistor) by $T_i(t)$. The heat flow through the heat resistance is denoted by P_a, whereas the heat flow through the heat capacitance by P_b. The following first-order differential equation can be easily obtained, which is an adopted Fourier's Law:

$$P(t) = P_a(t) + P_b(t) = C_i \frac{d\, T_i(t)}{dt} + \frac{T_i(t)}{\theta_i}. \tag{2.6}$$

The general solution of this differential equation follows (assigning $\tau_i = \theta_i C_i$):

$$T_i(t) = e^{-t/\tau_i}\, T_i(0) + \frac{1}{\tau_i} \int\limits_0^t e^{\frac{-t+t'}{\tau_i}}\, \theta_i P(t')\, \mathrm{d}\, t', \tag{2.7}$$

where θ_i is the heat resistance ($^\circ C/W$), C_i is the heat capacitance ($J/^\circ C$), and $\tau_i = \theta_i C_i$ is the time constant ($J/W = s$).

In a simple case of constant power dissipation $P(t) \equiv P$, the heat resistance θ_i can express the temperature rise T_{ss}^i to which the section will heat up to in steady state:

$$T_{ss}^i = \theta_i\, P. \tag{2.8}$$

The heat capacitance C_i is a measure of the heat required to change the temperature of the section. The larger the C_i, the slower the temperature of the section rise for the same power dissipation $P(t)$. For different sections in a transistor, the values of C_i can differ in order of one-two magnitudes.

The temperature of the transistor junction in the H-bridge is a sum of temperature differences over k RC thermal circuits (sections) and the temperature T_{cp} of the cold plate, which is cooled by water:

$$T_{\mathrm{amp}}(t) = \sum_{i=1}^{k} T_i(t) + T_{cp}, \tag{2.9}$$

where k is the number of sections in the transistor.

The heat power $P(t)$ depends quadratically on the input current $I(t)$ that is scaled gradient waveforms of the scans $G(t)$, see Figure 2.18. Each type of scan, has a unique gradient waveform $G(t)$, which consists of the gradient pulses. The
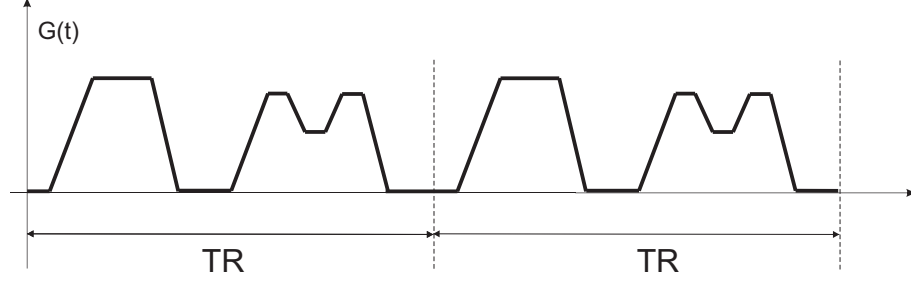
Figure 2.18: Gradient waveforms.

gradient waveforms are periodic functions, where the period is one *repetition time* TR. The function $G(t)$ represents the strength and direction of the magnetic field gradients, see Section 2.4. In general this function is trapezoid-like, however in practice the shape can be different.

2.5.3   Gradient coils temperature

The last MRI scan execution limitation addressed in this chapter is the temperature of the gradient coils. The coils transform the currents generated by the external gradient amplifiers into magnetic gradients, which was addressed in Section 2.2.2. During the MRI scan execution the gradient coils can also be overheated, because of the high primary currents in each coil and the eddy currents that are induced by one neighboring coil to another.

The dynamics of the gradient coils temperature is similar to the amplifiers one, but is more complicated. The main difference is that eddy currents are also taken into account. Another difference is that the maximum temperature is calculated for so-called *hot spots*. These hot spots are limited number of regions known in advance with high local current density that causes maximum local Joule heating in those regions.

In order, not to damage the gradient coils, their maximum temperature should be kept below some limitation, $T_{\text{coils}}^{\max}$. It is enough to keep only the maximum of the hot spots temperature below that limitation:

$$\max\left(T_{\text{coils}}\right) = \max_i\left(T_{\text{hotspot},i}\right) \leq T_{\text{coils}}^{\max}. \tag{2.10}$$

In practice there are less then a dozen of hot spots with different temperatures, e.g. due to symmetry.

The input currents I_x, I_y and I_z pass through all three coils x, y and z accordingly to generate the magnetic field gradients in three dimensions. The temperatures

of the coils depend on the input (primary) currents and eddy currents that are induced by one neighboring coil to another.

The temperature dynamics of the hot spot can be also modeled with the RC thermal circuits, see Figure 2.17. In contrast to the amplifier temperature model (equations (2.7) – (2.9)) the power $P(t)$ quadratically depends not only on the primary currents, but on the eddy currents as well. For each hot spot in a coil the temperature rise formula is the sum of the temperature rises due to the primary and all the eddy currents that are induced on the coil [Peeren 2009]:

$$\Delta T(t) = \sum_{\substack{a=x,y,z \\ i=0,\ldots,K_a}} \frac{R_{a,i}\,\Theta_{a,i}}{\tau_{\mathrm{therm},a,i}} \int_0^t \exp\left(\frac{-t+t'}{\tau_{\mathrm{therm},a,i}}\right) s_{a,i}^2(t')\ \mathrm{d}t', \qquad (2.11)$$

where K_a is the number of the eddy currents $s_{a,i}(t)$ that are induced in the a-coil with $a = x, y, z$; $i = 1, \ldots, K_a$. The primary currents are denoted by $s_{a,0}(t) = I_a$ in order to keep the expression uniform. The variable $R_{a,i}$ is the electrical resistance, whereas $\Theta_{a,i}$ is the heat resistance between the eddy current circuit and the hot spot. The thermal time constants $\tau_{\mathrm{therm}} = \Theta_{a,i}C_{a,i}$ are in order of minutes.

The heat resistances $\Theta_{a,i}$ and heat capacitors $C_{a,i}$ are calculated in advance for every primary and eddy current in the hot spots based on experiments. For various gradient coils designs the number and the location of the hot spots, as well as their temperature parameters are different.

Equations (2.5) – (2.11) were used as an input for a part of the scan segments intermixing algorithms that were designed during our research. These algorithms are presented in Chapter 5.

Chapter 3

Scheduling of MRI tasks

In this chapter an approach to reduce the examination time of MRI systems is introduced and supported by a literature review.

Due to duty cycle limitations, which were described in Section 2.5, the MRI examination time usually prolongates. This extra time can take up to the half of the scan time. Different techniques are used to reduce the duration of MRI examinations without compromising the image quality.

A lot of papers have been published in recent years considering the optimization of duty cycles of MRI scanners. There are several software-based approaches to the RF duty cycles optimization, those are based on RF pulse redesign, see e.g. [Gai and Zur 2007; Hargreaves et al. 2004]. Other papers consider improving design of MRI hardware as the solution to the problem, e.g., improving design of the MRI coils, see [Mueller et al. 2009; Poole et al. 2008; Trakic et al. 2009; van den Brink et al. 2003].

In certain cases, the prolongation of the examination time can be avoided without modification of the MRI hardware and without redesign of original RF pulses and/or gradient waveforms. We introduce a new comprehensive approach to MRI examination time reduction: the time reduction is accomplished by dividing parts of the MRI examination into segments that are then intermixed.

In this study the time reduction is posed as a scheduling problem. Since this approach does not influence the MRI hardware and pulses/waveforms design, it can be combined with conventional approaches to the duty cycle optimization

(1) | TR | dead time | TR | dead time | TR | dead time |

(2) | TR | TR | TR | dead time | TR | TR | TR | dead time |
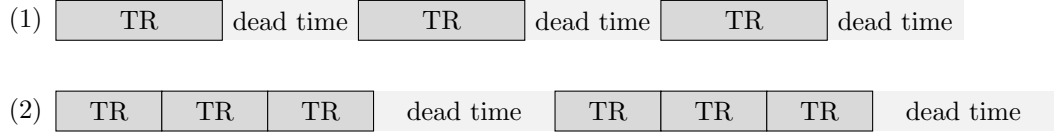
Figure 3.1: In current practice, an equal amount of dead time is inserted either (1) at the end of every TR, or (2) at the end of every fixed sequence of TRs of the limited scan.

problem.

The outline of the chapter is as follows. First, the scan segments intermixing approach is introduced. Second, the problem is stated formally in the scheduling framework. Finally, scheduling literature is reviewed on two main instances of the problem.

3.1   Scan segments intermixing

To understand the approach described in this section, some insights into the MRI examination structure and the duty cycle limitations are desirable. This information is available in Sections 2.4 and 2.5 of the previous chapter.

An MRI scan is called *resource-limited* if it utilizes a resource throughout the scanning process. The resources here stand for duty cycles of RF, gradient amplifiers or coils. Currently, when the resource is expended, a *dead time* is included into the resource-limited scan when the system just waits idle for the resource to become available again. Typically, an equal amount of dead time is inserted either at the end of every TR, or at the end of every fixed sequence of TRs of the resource-limited scan, see Figure 3.1. In some scans the dead time is up to the half of the scan time that provides a big opportunity for optimization. Different types of scans impose requirements on different resources. The scans that are SAR-limited do not impose severe requirements on the gradient system. On the other hand, the diffusion-weighted protocols (DWI) are often gradient-limited, whereas the RF deposition during these scans is typically low.

We propose, that during the examination, MRI scans can be divided into segments and then intermixed in such a way that the dead time of one scan would be exploited for the useful job of another one. Switching between the segments of different scans is not immediate. It takes some time, which is called *setup time*. During the setup time several pulse sequences of the next scan are executed in order to re-establish the steady-state, see [Weishaupt et al. 2006]. For different combinations of scans the setup times are different; however, it takes on average from 1 to 1.5 s to switch from one scan segment to another. The setup time is usually less than the total dead time in the limited scans, thus time reduction
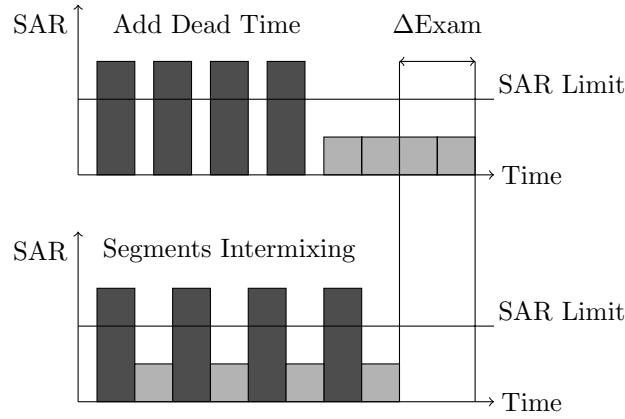
Figure 3.2: Intermixing SAR-limited scan (dark) with not SAR-limited one (light): ΔExam is the examination time reduction after the scans segments intermixing (for the sake of brevity the setup times are omitted).

can be gained.

The segments size depends on the limitation type. The time constants for the SAR level limitation is in order of magnitudes smaller than for the gradient amplifiers and coils temperature limitation (seconds against minutes), thus the scheduling algorithms for these types of limitations can be designed separately. An example of an intermixing strategy for the SAR level limitation is presented in Figure 3.2.

Obviously, not all the types of the scans can be intermixed, e.g. 'breath-holding' scans are excluded, as well as scans with a contrast agent injection (see Section 6.2). Also all the survey scans do not participate in the intermixing because of their relatively short duration (see Section 2.4.2), as well as some other scans that require special treatment. In practice, about half of scans from an examination can be selected for the intermixing process, and only these types of scans are considered onward.

The intermixing of the scans should not significantly increase the image obtaining cycle, which consists of two parts: image acquisition and reconstruction. The reconstruction is performed partly in parallel with the acquisition cycle: part of the data is being analyzed before the scan is finished. Moreover, the reconstruction is much faster than the acquisition, since it is only limited by the speed of the reconstruction computer, see Section 2.2.4.

3.2 Scheduling framework

In this section the problem is stated formally in the scheduling framework. Therefore scheduling notations are utilized; the explanation of the notation can be

found, e.g., in [Pinedo 2008].

The scan segments from here on are called *jobs* and scans are called *job families*. *Job* is a term from the scheduling theory that denotes some activity (task) that should be processed on the machine (MRI scanner) without interruption[1]. Moreover, there are *setup times* between jobs and they are different for the jobs from different job families. A scheduling algorithm specifies the sequence in which all the jobs must be processed to fulfill a specific criterion, e.g. a minimum total processing time (i.e., makespan).

Consider N jobs j, $j = 1, \ldots, N$ that belong to F distinct job families (scans). Let n_i denote the number of jobs in family i, the following equality is satisfied:

$$N = n_1 + n_2 + \cdots + n_F. \tag{3.1}$$

The completion time of job j in the schedule is denoted by C_j. Let (g, j) refer to job j from family g, $j = 1, \ldots, n_g$; it has a processing time $p_{(gj)}$.

Assumption 1 *To all intents and purposes jobs within single family are equal and can be arbitrarily interchanged. For each job family $g \in \{1, \ldots, F\}$ the processing times of the jobs within the family are identical:*

$$p_{(gj)} = p_g, \qquad g = 1, \ldots, F, \ j = 1, \ldots, n_g. \tag{3.2}$$

According to Assumption 1, the order of the jobs in the families is of no importance. Therefore the schedule $L = (j_1, \ldots, j_N)$ that specifies the order in which the jobs are processed can be unambiguously described by a switching sequence σ over the job families:

$$\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_N), \quad \sigma_k \in \{1, \ldots, F\}. \tag{3.3}$$

3.2.1 Specific absorption rate case

In the case of SAR limitations, the time reduction problem can be formulated as a problem of scheduling a single machine with sequence-dependent family setup times. *Job* here stands for several TRs in a row of a scan with excluded dead time, see Figure 3.1. All the jobs are available for processing at time zero (i.e., no release dates restrictions are involved). The objective is to minimize the makespan, $C_{\max}$, for the set of released orders.

A setup time is required on each occasion when the machine switches from processing jobs in one family to jobs in another family. These setup times are *sequence-dependent*, i.e., they depend on the families of the preceding and the current jobs. Additionally, *sequence-independent* setup times are required when

---

[1]preemptive scheduling is not relevant for this study.

the machine processes jobs from the same family; these setup times stand for the dead time, and they are separable in the sense that can be skipped if the next job is from a different job family.

Following a common three-field notation for theoretic scheduling problems, see e.g. [Allahverdi et al. 2008; Graham et al. 1979; Pinedo 2008], the problem can be denoted as follows:

$$1 \mid \text{ST}_{\text{si}}, \text{ST}_{\text{sd}, b} \mid C_{\max} . \tag{3.4}$$

In the first part of the notation (3.4) the machine environment is described: 1 stands for single machine; the second part expresses the jobs characteristics: $\text{ST}_{\text{sd}, b}$ stands for sequence-dependent family (or batch) setup times, whereas ST_{si} stands for sequence-independent job setup times (i.e., dead time); the last part of the notation describes the objective function: $C_{\max} = \max_{j=1,\dots,N} \{C_j\}$, that is, makespan.

In the simplest case, for single machine scheduling problems *without* sequence-dependent setup times and job families $(1 \mid \mid C_{\max})$ the makespan is independent of the sequence and equal to the sum of the processing times: $C_{\max} = \sum p_i$. When sequence-dependent setups are involved, the makespan *does* depend on the schedule. A literature review on existing approaches to the (3.4) problem follows in Section 3.3.

3.2.2 Hardware temperature constraints case

In the case of hardware temperature limitations, a *job* stands for a relatively long scan segment with multiple TRs. The length of these segments is in the order of a minute. Therefore the in-between scan setup times can be omitted, since their length is typically in the order of a second. All the jobs are available for processing at time zero, as well as in the SAR case.

In contradistinction to the SAR case, the scan segments switching policy depends on a complex dynamic function, that is temperature of several parts of MRI hardware — gradient amplifiers and gradient coils. During scans execution, the temperatures of amplifiers should be kept below $T_{\text{amp}}^{\max}$, whereas the temperatures of coils should not exceed $T_{\text{coils}}^{\max}$, see Sections 2.5.2 and 2.5.3 of the previous chapter. All these temperature constraints must be satisfied simultaneously:

$$\mathbf{T}(t) \leq \mathbf{T}^{\max}, \tag{3.5}$$

where $\mathbf{T}^{\max}$ is a vector of the temperature constraints for the amplifiers and the coils hot spots, whereas $\mathbf{T}(t)$ is a vector function of the corresponding temperatures during the MRI examination. Here the inequality relation is *component-wise*.

Additionally a *dummy* job family is introduced to deal with the dead time during the execution of the MRI examination. Therefore equation (3.1) should be

rewritten as follows:

$$N = n_0 + n_1 + \cdots + n_F, \tag{3.6}$$

where n_0 is the number of jobs in the *dummy* family, and F is the number of regular job families (i.e., scans). The *dummy* family is special in a sense that its size, n_0, is not predefined and can be arbitrary large.

The objective is to find σ a switching sequence over the job families

$$\sigma = (\sigma_0, \sigma_1, \ldots, \sigma_N), \quad \sigma_k \in \{0, \ldots, F\}, \tag{3.7}$$

for which all the jobs from the job families are processed, while keeping the number of the *dummy* jobs, n_0, minimal. By decreasing the number of the *dummy* jobs the maximum completion time is decreased. Thereby our problem is equivalent to the problem of the makespan minimization:

$$C_{\max} \to \min.$$

Notice that switching sequence σ can contain multiple jobs in a row from the same job family, e.g. a long sequence of the *dummy* jobs. Practically, duration of the *dummy* job, p_0, is used as a discretization step to construct 'dead time' of the required length. The value of p_0 is a design parameter that can be varied, so a trade-off can be achieved between potential time reduction and the efficiency of the scheduling algorithms.

3.3   Literature review

This section provides an overview of scheduling methods that are relevant for the problem of reduction of the MRI examination time. Two branches of scheduling theory are investigated. In the SAR limitations case, scheduling problems that include setup times are of interest. In the hardware temperature limitations case, literature on scheduling problems with temperature constrains is reviewed.

3.3.1   Scheduling problems with setup times

The interest in scheduling problems with setup times began in mid-1960s, and they were soon recognized among the most difficult classes of scheduling problems. Even for small systems, the complexity of the scheduling problems with sequence-dependent setup times is unreachable for existing theories, see [Pinedo 2008]. For instance, if there is a single family of jobs, all the jobs within the family require sequence-dependent setup times, and the objective is the makespan minimization:

$$1 \mid \mathrm{ST_{sd}} \mid C_{\max},$$

the problem can be represented as a *Traveling Salesman Problem* (TSP), see e.g. [Bianco et al. 1993; Fischetti et al. 1993]. However, the TSP is known to be strongly NP-hard and the algorithms that give an exact solution are numerically expensive, see [Pinedo 2008].

Comprehensive reviews of scheduling problems that involve setup times can be found in [Allahverdi et al. 1999], [Allahverdi et al. 2008], [Potts and Kovalyov 2000] and [Yang 1999]. These surveys demonstrate that a lot of research has been done in recent years regarding this type of scheduling problems. Different machine environments, job characteristics, and objective functions were investigated by researchers from all over the world. Throughout the past two decades, hundreds of studies have been published regarding scheduling with setup considerations. Only a small subset of these problems is efficiently solvable (i.e., not NP-hard); for the rest heuristic algorithms are being designed.

Makespan, $C_{\max}$, was not very popular in those studies among the objective functions. According to the aforesaid literature surveys, problems that are close to the one we are dealing with (see Section 3.2.1) were only addressed in [Monma and Potts 1989] and [van der Veen et al. 1998].

In the first study by [Monma and Potts 1989], authors draw conclusions on the complexity of several single machine scheduling problems, including the $1|\,\mathrm{ST}_{\mathrm{sd},b}\,|C_{\max}$ problem. Several assumptions are made in this study: setup times within job families are assumed to be zero, whereas sequence-dependent setup times between job families are assumed to satisfy the *triangle inequality*, that is $s_{ac} \leq s_{ab} + s_{bc}$, $a, b, c \in \{1, \ldots, F\}$. The problem, therefore, reduces to one of scheduling the families so that the total setup time is minimized. This problem is then equal to the *metric* TSP, with number of nodes equal to the number of job families. However, the aforesaid assumptions are not always satisfied for the scan segments intermixing problem we are dealing with.

In the second study by [van der Veen et al. 1998], a scheduling problem is investigated, which can be denoted by the same nomenclature as the problem we are dealing with (nomenclature (3.4) in Section 3.2.1). The authors consider that a certain additional *resource* is required to process a job (they refer to this resource as *template*). The jobs are partitioned in several families according to the resources they require; the number of the resources does not exceed the number of jobs. After a job has been processed, the *dead time* is required to restore the resource (e.g. to cool down or to decrease the SAR level). During this dead time the job normally remains in the machine. However, this dead time is separable in the sense that it can be skipped if the next job is from a different job family (i.e., requires another resource). In this case the first resource will be restored 'off-line', while processing the jobs from another family, and only a sequence-dependent family setup time is required to switch.

The authors demonstrate that this problem is equal to a special case of the TSP,

which is solvable in polynomial time. They also provide a multi-step algorithm that solves the problem in $O(N \log N)$ time. To some extent their algorithm can be used as an approach to the problem of scheduling the segments of scans while keeping the SAR level below the limitations. However, the authors make an assumption that the dead time is negligible compared to the processing times, which implies that the machine can switch back to the job family whenever that is required. This is not satisfied for the SAR case, because it takes multiple TRs of a non-limited scan to reduce the SAR level below the limitations (i.e., several jobs of the next family to restore the resource of the previous family).

In a study by [Ozden et al. 1985] a problem was investigated that can be denoted by the following nomenclature:

$$1 \mid \mathrm{ST_{sd}}, \mathrm{ST_{sd,\,b}} \mid \mathrm{TST},$$

where $\mathrm{ST_{sd,\,b}}$ stands for sequence-dependent family setup times, whereas $\mathrm{ST_{sd}}$ denotes that job setup times within each family are also sequence-dependent. The objective function is Total Setup Time (TST), which is obviously equal to the $C_{\max}$ function in the single machine case (since $C_{\max} = \sum p_i + \mathrm{TST}$ and the processing times, p_i, $i = 1, \ldots, N$, are fixed). The authors have developed a dynamic programming-based formulation of this problem, and demonstrated its efficiency by means of simulation. The algorithm is based on an assumption that setup times within job families are much smaller compared to the family setup times. However, as it was mentioned above, such an assumption is not satisfied for the problem we are dealing with.

In a study by [Psaraftis 1980] a problem of sequencing families of identical jobs was considered, which is applicable to the scan segments intermixing case, since Assumption 1 guarantees that the processing times of the jobs within the families are identical. The authors have developed a dynamic programming approach to the sequencing problem. The cost function (objective) they have considered is very general, thus can be reduced to the $C_{\max}$ one. In Section 4.3.1 we adopt their method to the scan segments intermixing problem. However the efficiency of the dynamic programming algorithm is not high.

To the best of our knowledge, no other approaches to the SAR case of the scan segment intermixing problem (see Section 3.2.1) exist in the present-day scheduling literature. Therefore in our research this scheduling problem was investigated and efficient algorithms were designed in order to solve it. The results of this research are presented in Chapter 4.

3.3.2 Scheduling problems with temperature constrains

In this section the literature on the scheduling problem with the thermal constraints is reviewed. For ease of exposition, the notations from Sections 2.5.2 – 2.5.3 and Section 3.2 are used.

Scheduling problems similar to the one we are dealing with arise in applications to embedded microprocessors. These problems are of tasks sequencing in modern processors with on-chip temperature constrains, namely *Temperature Aware Scheduling* (TAS) problems. Temperature dynamics in microprocessors is similar to the one in amplifiers and coils, and in most of the papers it is described by similar RC thermal models, see [Skadron et al. 2002].

Since power density continues to increase exponentially with semiconductor technology scaling, the rise in on-chip temperatures can shorten the lifetime of electrical circuits and degrade the performance of modern processors. Recently there has been significant interest in thermal management in microprocessors and other embedded multi-tasking systems. Various techniques have been proposed as solutions of the *Dynamic Thermal Management* (DTM) problem (reviews of these techniques see, e.g. in [Brooks and Martonosi 2001; Jayaseelan and Mitra 2009; Skadron et al. 2004]). One of the cooling solutions embedded into modern processors is *dynamic voltage/frequency scaling* (DVFS) that allows processor to select from multiple voltage/frequency states. However most of the DTM techniques are reactive in nature, that is, response mechanisms are invoked only when the temperature crosses a threshold. This response leads to a reduction in temperature, which is typically accompanied by a degradation in performance of the processor, see [Srinivasan and Adve 2003].

There are several papers considering the *predictive* DTM. The very first one was by [Srinivasan and Adve 2003], where thermal dynamic models were utilized to anticipate the future thermal behavior and take actions to avoid overheating. However the scope of that paper was limited to the single-tasking workloads, thus the effect of multiple tasks was not considered in that paper.

In a paper by [Rao et al. 2006] the shape of optimal processor speed profile was derived to maintain the temperature below a maximum limit. The authors assumed that the speed of the processor can be changed smoothly over time and used the calculus of variations technique to solve the DTM problem analytically. However this assumption is not applicable to our case.

In a paper by [Zhang and Chatha 2007] a problem that is similar to our case is addressed. Namely, minimization of execution time of a periodic task set on a processor subject to thermal constraints. Unlike our problem, the execution order of the tasks is predefined. The authors considered a processor with one *sleep* state that is characterized by a low power consumption and steady state temperature T_{ss}. The processor can go into that sleep state on completion of a job (i.e., task), and before the start of the next job. This is very similar to the *dead time* insertion during the scanning process. The authors considered a set of N jobs with the initial temperature $T(0)$ and the peak temperature constraint $T_{\max}$. The maximal number of $N+1$ sleep states can be interleaved with this sequence: one sleep state before and one after every regular job. The duration of each

sleep state is selected from a range of discrete values with a fixed step size (for instance $\{0, 10, 20, \ldots, 100\}$). Moreover the voltage/frequency (hence the power) of each job can be selected from a predefined set of values (using DVFS). Since the set of jobs is executed periodically, an additional constraint was imposed that the final temperature after one complete execution of the jobs set must not exceed the initial temperature $T(0)$. The sequence of jobs is considered to be predefined, hence the only way to control the temperature profile is to scale the voltages/frequencies of the jobs and to insert the sleep states between the jobs. The authors proved that the execution time minimization problem subject to the aforesaid constraints is NP-hard by a reduction from the multiple choice knapsack problem. The authors also developed polynomial time approximation algorithms.

In paper by [Jayaseelan and Mitra 2008] a temperature-aware task sequencing of a periodic *heterogeneous* task executing on a processor under timing constraints is addressed (where heterogeneous tasks are tasks with different thermal profiles). The authors claim that it is the first approach that exploits task reordering for thermal management. They observed that, in general, tasks vary significantly in power consumption. This results in variation of the thermal profiles of the tasks. The temperature profile and the peak temperature of a set of heterogeneous tasks are highly dependent on the execution sequence of the tasks on a processor. They observed $9.02\,^\circ C$ difference in the peak temperatures between 'worst' and 'best' execution sequence of the same task set with 8 tasks.

For a given set of tasks (not necessary periodic) and initial temperature $T(0)$ the problem to find a sequence of the tasks with minimal peak temperature $T_{\max}$ was shown to be NP-hard by a reduction from the bottleneck traveling salesman problem (bottleneck TSP). The authors also proposed an heuristic algorithm to solve that problem. The heuristic is based on alternating *hot* and *cold* tasks. Whether the task is *hot* or *cold* is defined by a special metric that takes into account the steady state temperature of the task, its duration, and some other factors. The efficiency of their task sequencing algorithm is $O\left(N\left(\log N\right)^2\right)$, where N is the number of tasks. Finally, the authors extended their algorithm with iterative voltage/frequency scaling (using DVFS) and *sleep* states insertion (similar to the work by [Zhang and Chatha 2007] reviewed upward).

The authors of [Jayaseelan and Mitra 2008] considered a problem of minimizing peak temperature under time constraint, which is different from the problem of minimizing execution time under peak temperature constraint (in some sense, these problems are dual). The task set was considered to be periodic, i.e. repeating itself infinitely; the temperatures of the tasks were considered to be in the steady state. However these assumptions are not satisfied for the scan segments sequencing, since the number of scan segments is limited and is not necessary equal for different scans: in equation (3.6), n_i and n_j can be not equal for some $i \neq j$. Also scan sequencing algorithms should be able to take advantage of a *cold*

pre-steady state phase, which is relatively long. Thereby the heuristic algorithms published in [Jayaseelan and Mitra 2008] are not directly applicable to the scan segments intermixing problem.

Recently, a paper by [Zhang and Chatha 2010] was published, where a problem of throughput maximization (i.e., minimization of execution time) for a periodic task set subject to peak temperature constraint $T_{\max}$ is addressed. Sequencing of tasks, insertion of *sleep* states, and DVFS were combined in that paper in a general approach to the DTM problem.
The authors claim that the problem in general is NP-hard, though its several special instances (e.g. homogeneous tasks case) can be solved efficiently. This observation implies that the general case of the scan segments sequencing problem subject to thermal constraints is also NP-hard. The authors propose an heuristic algorithm to solve the general instance of the thermal aware task sequencing problem. To some extent this heuristic can be used as an approach to the scan segments sequencing problem we are dealing with.

To the best of our knowledge, the heuristic by [Zhang and Chatha 2010] is the closest approach to the temperature restricted case of the scans segments sequencing problem (see Section 3.2.2) existing in the present-day scheduling literature. The results of our research and the algorithms that we have developed to solve the problem are presented in Chapter 5.

<div align="right">

Chapter 4

</div>

Scheduling with setup times and SAR constraints

In this chapter the first case of the MRI examination time reduction problem is described, namely the SAR level limitations case. For the insights into the problem see Sections 2.2.3 and 2.5.1 in Chapter 2. An introduction into our approach to the time reduction problem with the corresponding literature review is presented in Chapter 3.

Most parts of this chapter were published in [Ivanov et al. 2009a],[Ivanov et al. 2009c] and [Ivanov et al. 2010b].

The outline of this chapter is as follows. In the first section the SAR-related equations are presented. In the next section the formal problem statement to the SAR limitations case is posed. Then in subsections of Section 4.3 the algorithms that we have designed to solve the problem are described. In the last section the results of validation of the algorithms on Philips MRI examination protocols (ExamCards) are provided.

4.1 SAR Calculations

In Section 2.5.1 the SAR level limitations were described. The general equation to calculate the SAR level is (2.3) on page 29. The RF power dissipation (SAR) in a patient depends on the following factors:

1. spatial distribution of the RF transmit field B_1 (depends on the RF coil);

2. patient's position in that field (depends on the RF coil);

3. patient's shape, size, and conductivity distribution;

4. strength of the RF field B_1 as a function of time.

Typically, the SAR is calculated as square of RF field B_1 averaged over the duration of the scan with a conversion factor. Other dependencies (which are handled by assuming either a 'standard' or a 'worsts-case' patient) are aggregated in the conversion factor.

From Section 2.4 in Chapter 2 it is known that every scan consists of multiple pulse sequences; each pulse sequence of a scan has the same duration, i.e. repetition time (TR). To estimate the SAR level during the scan it is enough only to calculate the square of RF field B_1 averaged over the duration of a single TR of the scan:

$$\text{SAR} = \frac{C}{\text{TR}} \int_{\text{TR}} B_1^2(t) \, \mathrm{d}t \,, \tag{4.1}$$

where C stands for the conversion factor that aggregates all the other dependences (the value of C depends on the RF coil).

In Section 2.3.2 it was mentioned that RF pulses take place at the beginning of the TR. In Section 3.1 the process of dead time insertion was described: the dead time is typically added at the end of every TR, or a sequence of TRs.

Consider $\text{TR}_1 = t_1 - t_0$ as a minimal required repetition time and $\text{TR}_2 = t_2 - t_0$, $t_0 < t_1 < t_2$, as the repetition time after addition of the dead time, see Figure 4.1. A straightforward computation of the relative SAR level of the extended TR, based on equation (4.1), follows.
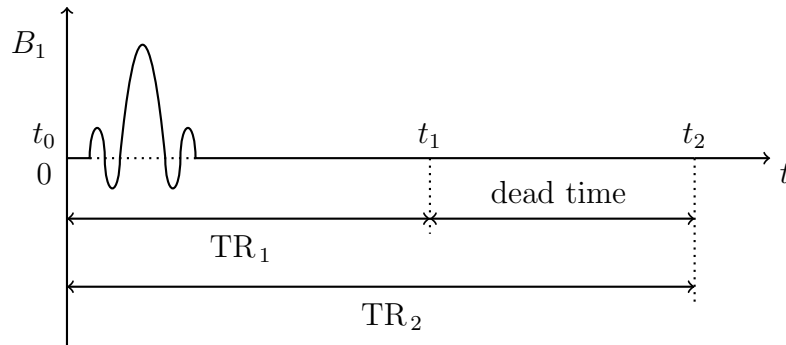
Figure 4.1: TR before and after addition of the dead time.

$$\text{SAR}_1 \;=\; \frac{C}{t_1 - t_0} \int\limits_{t_0}^{t_1} B_1^2(t)\,\mathrm{d}t\,,$$

$$\text{SAR}_2 \;=\; \frac{C}{t_2 - t_0} \int\limits_{t_0}^{t_2} B_1^2(t)\,\mathrm{d}t \;=\; \frac{C}{t_2 - t_0} \left(\int\limits_{t_0}^{t_1} B_1^2(t)\,\mathrm{d}t + \int\limits_{t_1}^{t_2} B_1^2(t)\,\mathrm{d}t \right) =$$

$$= \Big[\, B_1(t) = 0,\; t_1 < t \le t_2 \,\Big] \;=\; \frac{C}{t_2 - t_0} \int\limits_{t_0}^{t_1} B_1^2(t)\,\mathrm{d}t\,.$$

Here SAR_1 and SAR_2 are the SAR levels of TR_1 and TR_2 respectively. Since $B_1^2(t) > 0$ always holds, the ratio of these SAR levels is

$$\frac{\text{SAR}_2}{\text{SAR}_1} = \frac{t_1 - t_0}{t_2 - t_0} = \frac{\text{TR}_1}{\text{TR}_2}\,. \tag{4.2}$$

Equation (4.2) demonstrates that the SAR level is inverse proportional to the length of the TR. Therefore, doubling the TR reduces the specific absorption rate by half. Similar calculations can be performed to derive the average SAR level after alternating segments of two different scans.
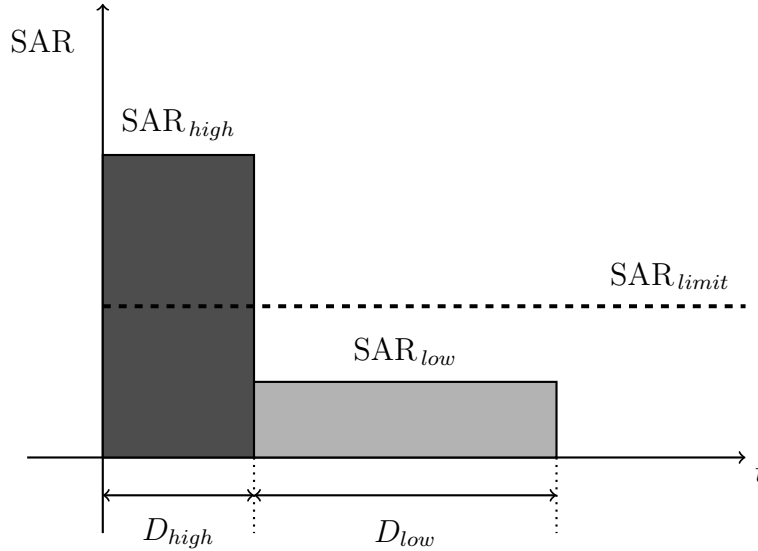
Figure 4.2: Alternating two scan segments of different scans with different levels of SAR to keep the average SAR level below the limit.

Let us consider two segments of different scans with different levels of SAR. A *high-SAR* segment and a *low-SAR* segment with the SAR_{high} and SAR_{low} levels

of SAR respectively. Given the duration of the high-SAR scan segment D_{high} we will calculate the duration D_{low} of the low-SAR scan segment to keep the average SAR level over the time period $D_{high} + D_{low}$ below some limit SAR_{limit}, $SAR_{low} < SAR_{limit} < SAR_{high}$, see Figure 4.2. Since the SAR level is constant throughout each of the segments, see equation (4.1), the equality is simple:

$$D_{high}\, SAR_{high} + D_{low}\, SAR_{low} = (D_{high} + D_{low})\, SAR_{limit}\,,$$
$$D_{low}\, (SAR_{limit} - SAR_{low}) = D_{high}\, (SAR_{high} - SAR_{limit})\,,$$
$$D_{low} = D_{high}\, \frac{SAR_{high} - SAR_{limit}}{SAR_{limit} - SAR_{low}}\,. \tag{4.3}$$

In a particular case, when the SAR level of the high-SAR scan segment exceeds two times the limit $SAR_{high} = 2\, SAR_{limit}$, whereas the SAR level of the low-SAR scan segment is only a half of the limit $SAR_{low} = \frac{1}{2}\, SAR_{limit}$, from equation (4.3) it follows that the duration of the low-SAR scan segment should be two times longer the duration of the high-SAR one:

$$D_{low} = D_{high}\, \frac{2\, SAR_{limit} - SAR_{limit}}{SAR_{limit} - \frac{1}{2}\, SAR_{limit}} = D_{high}\, \frac{SAR_{limit}}{\frac{1}{2}\, SAR_{limit}} = 2\, D_{high}\,.$$

In case if the setup time takes place, then the D_{low} represents the duration of the low-SAR scan together with the setup time.

According to the IEC guidelines, [CEI/IEC 60601-2-33 2008], the SAR level of a high-SAR scan SAR_{high} shall not exceed three times the corresponding long-term SAR limit, whereas the IEC short-term SAR limit prescribes that the duration of the high-SAR scan segment D_{high} should not exceed 10 s:

$$SAR_{high} \leq 3 \cdot SAR_{limit}\,, \tag{4.4}$$
$$D_{high} \leq 10\ \text{s}\,, \tag{4.5}$$

for the details on the long-term and the short-term SAR limits see Section 2.5.1. Since each scan segment consists of several TRs, the following inequality holds:

$$\# TR_{high} \equiv \frac{D_{high}}{TR_{high}} \leq \left\lfloor \frac{10\ \text{s}}{TR_{high}} \right\rfloor\,, \tag{4.6}$$

where $\# TR_{high}$ is the number of TRs of a high-SAR scan segment, and $\lfloor \cdot \rfloor$ stands for the integer part of the ratio. Inequality (4.6) limits the maximum number of TRs in a high-SAR segment that can be executed without inserting any dead time in-between.

4.2   Problem statement

In this section a formal problem statement for the SAR limitations case of the time reduction problem is presented. The scheduling framework was introduced

in Section 3.2 and a nomenclature for the problem was presented in Section 3.2.1 of the previous chapter. The notations introduced in those sections will be used onward.

There are N jobs j, $j = 1, \ldots, N$ that belong to F different job families; n_i denotes the number of jobs in family i and equality (3.1) from Section 3.2 is satisfied. The setup times within job families depend on the *resources*. These resources are considered to deal with constrains (e.g. SAR level) during the MRI scans execution. The SAR raises during execution of each scan segment, and it has to be decreased either by adding a *dead time* after the segment or by switching to a scan with a lower SAR level, see Section 3.1.

Each job family g can depend on a resource R_i, $i \in \{1, \ldots, m\}$, $m \leq F$. The processing of jobs (g, j), $j \in \{1, \ldots, n_g\}$ from the family g can only be started if the resource R_i is available. Different job families can depend on identical resources; this dependence is denoted by mapping $R(\cdot)$ as follows:

- $R(g) = R_i$ if job family g depends on resource R_i

- $R(g) = 0$ if job family g does not depend on any resources.

If the job family g depends on resource $R(g) = R_i$, then after each job within the family a sequence-independent setup time $\tau_g > 0$ is necessary to restore the resource (if $R(g) = 0$ then $\tau_g = 0$); this setup time stands for the *dead time* in a resource-limited scan (for the details see Chapter 3). The setup time τ_g is separable, in the sense that during this setup time the machine can switch to another job family h if the family does not depend on the same resource, and the resource $R(h)$ is available.

The switching to another family is not immediate: a sequence-dependent family setup time $s_{gh} > 0$ is required. When such an inter-family switching occurs, the machine should process jobs from family h for at least τ_{gh} time, i.e., *cooling time*, to restore the resource $R(g)$ of the previous job family g. Only after the τ_{gh} time the machine is able to switch back to the jobs from family g or to any other family (if $R(g) = 0$ then $\tau_{gh} = 0$, $\forall h \in \{1, \ldots, F\}$). This requirement is motivated by the fact, that it takes multiple TRs of a non-limited scan to reduce the SAR level below the limitations (i.e., several jobs of the next job family to restore the resource of the previous job family).

It is disallowed to switch between any families until τ_{gh} time has passed. This requirement can be restrictive for a case with two and more resources, since an opportunity to restore the resource of the previous family with a combination of jobs from different families is not utilized. For a case with single resource this requirement does not impose any side limitations on the solution.
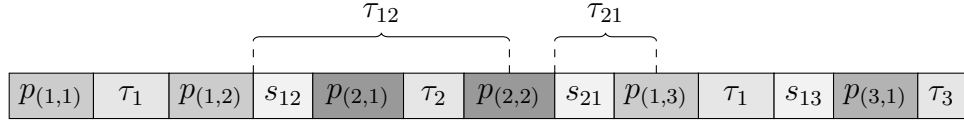
Figure 4.3: Example with three job families, the first and the third job families depend on the same resource $R(1) = R(3) = R_1$, whereas the second job family depends on different resource $R(2) = R_2$.

The objective of the scheduling problem is to minimize of the total production time (makespan):

$$C_{\max} = \max_{j=1,\ldots,N} C_j \;\; \rightarrow \;\; \min. \tag{4.7}$$

A three-field scheduling notation for this problem (see e.g. [Graham et al. 1979]) is provided in Section 3.2.1; the literature on similar problems is reviewed in Section 3.3.1 of the previous chapter.

To illustrate the aforesaid notation two examples are analyzed below:

Example with two different resources:

An example with two resources and three job families is presented in Figure 4.3. Here job families 1 and 3 depend on the same resource $R(1) = R(3) = R_1$, whereas job family 2 depends on a different resource $R(2) = R_2$.

Thereby during processing the job family 1 instead of performing setup τ_1 the machine switches to the job family 2 and process it for a period of τ_{12} time units (two jobs here with the processing times $p_{(2,1)}$ and $p_{(2,2)}$ respectively). Afterward it switches back to the job family 1 and processes it for at least τ_{21} time units (one job here). Finally, the machine switches from the job family 1 to the job family 3, which depends on the same resource R_1, thus the setup time τ_1 cannot be omitted and both setups (τ_1 and s_{13}) take place.

Example with a single resource – SAR:

In this example the notation of the current section will be linked to the SAR calculations in Section 4.1, therefore the terms like SAR level, TR, high-SAR and low-SAR scan are widely used onward. The explanation of these terms can be found in that section.

Consider set H of job families that depend on the resource 'SAR', and set L of job families that do not depend on any resources: $R(h) = \mathrm{SAR}, \forall h \in H$, and $R(l) = 0, \forall l \in L$. Processing, setup, cooling and dead times of the jobs from the

families are defined as follows:

$$
\begin{aligned}
p_{(l,\cdot)} &= \mathrm{TR}_l, \\
p_{(h,\cdot)} &= D_h, \\
\tau_{hl} &= D_l, \\
\tau_{lh} &= 0, & h \in H \text{ and } l \in L \\
s_{hl} &= f(h,l), \\
s_{lh} &= f(l,h), \\
\tau_l &= 0.
\end{aligned}
\tag{4.8}
$$

Here TR_l is the TR length of the low-SAR scan, i.e., scan that does not depend on the resource SAR; D_h is the duration of the high-SAR scan segment, which is selected to satisfy the inequalities (4.5) – (4.6). Whereas D_l is the duration of the low-SAR scan segment; given the D_h of the precessing scan, the D_l can be calculated by equation (4.3). The family setup times s_{hl}, s_{lh} are determined by some nonnegative mapping $f : H \cup L \times H \cup L \to [0,\infty)$ that depends on physical parameters of the preceding and the following scan, and f is predefined for each pair of the job families (i.e., each pair of the scans). The dead time τ_l and the cooling time τ_{lh} of the non-limited scan are zeros by the definition (see page 54).

Finally, the dead time τ_h, $h \in H$ of the SAR limited scan segments can be calculated by substituting $\mathrm{SAR}_{high} = \mathrm{SAR}_h$, $\mathrm{SAR}_{low} = 0$, $D_{high} = D_h$ into equation (4.4):

$$
\tau_h = D_h \left(\frac{\mathrm{SAR}_h - \mathrm{SAR}_{limit}}{\mathrm{SAR}_{limit}} \right),
\tag{4.9}
$$

where $h \in H$, whereas SAR_{limit} is the long-term IEC limit for the SAR level of the scan, see Section 2.5.

To comply with the short-term SAR limits, the SAR levels of the scan segments should satisfy equation (4.4). If the SAR level of a scan segment $h^* \in H$ exceeds the limit, i.e., $\mathrm{SAR}_{h^*} > 3\,\mathrm{SAR}_{limit}$, then the TR of the scan TR_{h^*} should be modified by inserting the dead time. The length of the extended repetition time $\mathrm{TR}_{h^*}^{ext}$ is straightforwardly determined by equation (4.2) with a substitution

$$
\begin{aligned}
\mathrm{SAR}_1 &= \mathrm{SAR}_{h^*}, \quad \mathrm{SAR}_2 = 3\,\mathrm{SAR}_{limit}, \\
\mathrm{TR}_1 &= \mathrm{TR}_{h^*}, \quad \mathrm{TR}_2 = \mathrm{TR}_{h^*}^{ext},
\end{aligned}
$$

as follows:

$$
\mathrm{TR}_{h^*}^{ext} = \mathrm{TR}_{h^*} \, \frac{\mathrm{SAR}_{h^*}}{3\,\mathrm{SAR}_{limit}}.
\tag{4.10}
$$

The reduction of the SAR level of the limited scans below the short-term limitations, by modifying the TRs of the limited scans according to equation (4.10), should be performed before scheduling of the segments. The extended repetition

times $\text{TR}_{h^*}^{ext}$, $h^* \in H$ and the reduced SAR levels 3 SAR_{limit} of these scans are then used to calculate the parameters of the corresponding jobs (4.8) – (4.9).

Thereby if the SAR levels and TR lengths of the scans in an MRI examination are given and the scan setup times are known, then all the parameters of the jobs in the families in the sets H, L can be calculated by equations (4.8) – (4.10). There is a choice when selecting the duration of the high-SAR segments D_h, $h \in H$. On the other hand, the longer the D_h, the shorter the examination length is, thus an 'optimal' duration is

$$D_h = \text{TR}_h \left\lfloor \frac{10 \text{ s}}{\text{TR}_h} \right\rfloor,$$

which is the longest duration that satisfies both inequalities (4.5) – (4.6). Here $\lfloor \cdot \rfloor$ stands for the integer part of the ratio.

Algorithms that we have designed to solve the scheduling problem with the objective function (4.7) are presented in the following section.

4.3 Scheduling algorithms

In this section algorithms are described that have been designed to solve the scheduling problem. These algorithms, for a given set of jobs $j = 1, \ldots, N$, which belong to $F \geq 2$ different job families, and the corresponding set of resources $R_1, \ldots, R_m$, $1 \leq m \leq F$, generate a sequence σ^* with minimal makespan:

$$\sigma^* = \arg \min_{\sigma \in \{\text{All sequences of the jobs}\}} C_{\max}. \tag{4.11}$$

All the parameters relevant for the SAR limitation case $p_a, p_b, \tau_a, \tau_b, n_a, n_b, s_{ab}, s_{ba},$ τ_{ab}, τ_{ba} (that were defined in Sections 3.2 and 4.2) are given for all pairs $a \neq b$ of job families $a, b \in \{1, \ldots, F\}$.

4.3.1 Dynamic programming approach

In this section a dynamic programming approach to the problem is described, which is similar to the one from [Psaraftis 1980], see Section 3.3.1.

The following assumption is quite natural for the MRI systems, where any switching between different scans takes some extra time.

Assumption 2 *The sequence-dependent family setup times are strictly positive:*

$$s_{ab} > 0, \qquad a \neq b, \ a, b \in \{1, \ldots, F\}. \tag{4.12}$$

This assumption is posed to eliminate some unnecessary switchings between the job families.

Let us define **the optimal value function** $V(L, k_1, \ldots, k_F)$ as the minimum total setup time to process k_i jobs of family i, $i = 1, \ldots F$ that are waiting to be processed, given that a job belonging to family L is just started.

The following **recursive relation** is satisfied:

$$V(L, k_1, \ldots, k_F) = \begin{cases} 0, & \text{if } k_1 = \cdots = k_F = 0 \\ \min_{j \in \mathbb{J}} \Big[f(L, j, k_1, \ldots, k_F) + \\ \qquad + V(j, k_1, \ldots, k_j - 1, \ldots, k_F) \Big], & \text{otherwise} \end{cases} \tag{4.13}$$

where $\mathbb{J} = \{j : k_j > 0\}$ is the set of under-processed job families. A prescribed function $f(a, b, k_1, \ldots, k_F)$ defines the incremental setup time of processing a job from family b immediately after a job from family a (where $a, b \in \{1, \ldots, F\}$). For our problem function f is defined as

$$f(a, b, k_1, \ldots, k_F) = \begin{cases} \tau_a, & a = b \\ s_{ab}, & a \neq b, \ R(a) \neq R(b), \ \tau_{ab}^* \geq k_b \\ \tau_a + s_{ab}. & \text{otherwise} \end{cases} \tag{4.14}$$

Here $\tau_{ab}^* = \left\lceil \dfrac{\tau_{ab} - s_{ab}}{p_b + \tau_b} \right\rceil$ is the minimum number of jobs from family b needed to restore the resource of family a, where $\lceil \cdot \rceil$ stands for the ceiling of the ratio.

The algorithm to solve the dynamic programming problem is as follows:

- Start from $k_1 = \cdots = k_F = 0$, where $V = 0$, $\forall L \in \{1, \ldots, F\}$.

- Process the k-vector in lexicographical order.

- At each F-tuple, apply equation (4.13) for all possible $L \in \{1, \ldots, F\}$.

- Each time equation (4.13) is applied, record

$$\mathrm{F}_{next}(L, k_1, \ldots, k_F) := \arg\min_{j \in \mathbb{J}} \big[\cdot \big],$$

  the best next family to process.

- Arbitrarily set $\mathrm{F}_{next}(L, 0, \ldots, 0) := 0$, $\forall L \in \{1, \ldots, F\}$.

The initial conditions of this problem are specified by two parameters: L_0 and $(n_1, \ldots, n_F)$. The latter parameter specifies initial values of k-vector. This determines that all the jobs in the families are available to be processed. Here n_i is the number of jobs in family $i \in \{1, \ldots, F\}$, see Section 3.2.

The value of $L_0 \in \{1, \ldots, F\}$ specifies the initial job family, which job is being processed just prior to processing the first job of the set. If no initial job is specified ($L_0 = 0$), then the first job, L_1, is the one that minimizes $V(L_1, n_1, \ldots, n_F)$.

Thereby, for a given set of initial conditions L_0 and $(n_1, \ldots, n_F)$, the algorithm should be executed single time to fill the array F_{next} with the values. Then, the algorithm moves forward N steps, from $(L_0, n_1, \ldots, n_F)$ to the terminal state $(L_N, 0, \ldots, 0)$, using the data from F_{next}.

The complexity of the algorithm is in the order of

$$F^2 \prod_{i=1}^{F}(1 + n_i). \tag{4.15}$$

In a simple case, when all the families have the equal amount of jobs $n = N/F$, the complexity of the algorithm can be rewritten as $F^2 (1 + n)^F$. This function is polynomial with respect to n, the average number of jobs in a family, but exponential with respect to F, the number of families.

4.3.2   Scheduling job families in pairs

In the previous section a dynamic programming approach to the problem was described. However, the performance of that algorithm is typically low, since the efficiency of the recursion procedure is exponential with respect to the number of families.

In practice, an additional assumption can be imposed, that can reduce the complexity of the problem, which was originally NP-hard (see Section 3.3.1).

Assumption 3 *No more than two job families can be intermixed simultaneously. Technically, for each $k \leq N$ the number of different job families in the head subsequence $\{f_1, \ldots, f_k\} \subset \sigma$ that have been started, but have not been processed yet, does not exceed two:*

$$\# \left\{ g \,\middle|\, 0 < \sum_{i=1}^{k} \chi_g(f_i) < n_g, g = 1 \ldots F \right\} \leq 2, \tag{4.16}$$

where $\chi_g \to \{0, 1\}$ is the characteristic function of subset of the jobs that belong to the family g.

Furthermore, this assumption makes the examination time less sensitive with respect to the scan execution failures (e.g., due to the patient motion), as no more then two scans have to be re-executed in this case.

According to Assumption 3, the algorithms that solve the optimization problem should work with pairs of the job families. On a general level, the sequence can be specified in which the pairs of the job families will be processed. Then for each particular pair an optimal intermixing of jobs can be derived.

After intermixing a pair of job families $a, b \in \{1 \dots F\}$ typically several jobs from one of the families will be left over. They can be intermixed with the next job family c. Let n_g^* denote a number of jobs from family g that still needs to be processed. Initially, when the machine starts working $n_g^* = n_g$, $g = 1 \dots F$. Afterward, the n_g^* decreases while processing the jobs from the families.

We have designed a scheduling algorithm that consists of two parts. First, a sequence π in which the pairs of the job families will be processed is selected. Secondly, the jobs in the pairs are intermixed to minimize the $C_{\max}$ of each pair in the sequence π.

In the first part of the algorithm, the selection of the sequence π can be done either in a brute-force way or by a heuristic greedy search. Thereby, there are actually two scheduling algorithms, depending on the way the sequence π is selected:

- the *brute-force* algorithm,

- the *heuristic greedy* algorithm.

The description of these algorithms follows.

In the *brute-force* algorithm, almost all the permutations of $\pi = (1, \dots, F)$ are tried out (only permutations that lead to the same solutions are excluded). The resulting number of job sequences $\#\{\pi\}$ is in the order of magnitude of $F!$. Consequently, the complexity is $O(F!)$, which is worse than exponential for large F.

The *heuristic greedy* algorithm in each step searches for a pair of job families that after intermixing will gain maximal time reduction $T(g, h)$, $g, h \in \{1, \dots, F\}$.

1. All pairs of the families are looked over to find the most 'gainful' one, that is, $T(\cdot, \cdot) \to \max$. The number of the pairs is $\binom{F}{2}$, i.e., the number of 2-element subsets of an F-element set. Here $\binom{\cdot}{\cdot}$ stands for binomial coefficients.

2. Next step, all the remaining job families are tested to find the one that will gain maximal time reduction $T(g, \cdot)$ after intermixing with n_g^* residual jobs of the previous family.

3. Algorithm finishes after $F - 2$ repetitions of Step 2, when all F families are processed.

The total number of job family sequences is:

$$\#\{\pi\} = \binom{F}{2} + \binom{F-2}{1} + \cdots + \binom{1}{1} = (F-1)^2, \qquad (4.17)$$

and the complexity is $O(F^2)$, which is polynomial with respect to F.

For a small number of jobs families $F \leq 5$ the brute-force and both the greedy algorithms are comparable, and perform slightly better than the exponential-time one from the previous section. (As will be shown below, the scheduling algorithms are at most polynomial with respect to $n = N/F$, the average number of jobs in the family). For a larger number of job families F the greedy algorithm significantly outperforms others. Practically, there are about 5 scans in a clinical MRI examination that can be intermixed, see Section 2.4.1 and Section 3.1 for the details. Therefore all three algorithms are applicable.

The second part of the algorithm is as follows. For a given pair of job families (g, h) and the last job l in the previous pair (for the first pair $l = 0$), the optimal sequence $\phi(g, h)$ is designed in such a form:

$$(l, \phi(g,h)) = (l, \underbrace{g, h, h, h}_{k^* \leq k_g + k_h}, \overbrace{g, g}^{k_g}, \underbrace{h, h, h}_{k_h}, \overbrace{g, g}^{k_g}, \underbrace{h, h, h}_{k_h}, \ldots$$
$$\ldots, \overbrace{g, g}^{k_g}, \underbrace{h, h, h}_{k_h}, \overbrace{g, g}^{k_g}, \underbrace{h, h, h}_{k_h}), \qquad (4.18)$$

where

$$k_g = \arg \min_{k \in \mathbb{N}} \left(s_{hg} + k \left(p_g + \tau_g \right) - \tau_{hg} > 0 \right),$$
$$k_h = \arg \min_{k \in \mathbb{N}} \left(s_{gh} + k \left(p_h + \tau_h \right) - \tau_{gh} > 0 \right), \qquad (4.19)$$

are the minimal numbers of jobs to restore the resources $R(h)$ and $R(g)$ respectively. The following inequalities should hold:

$$k_g < n_g^*, \qquad k_h < n_h^*, \qquad (4.20)$$
$$\tau_g > 0 \quad \text{or} \quad \tau_h > 0, \qquad (4.21)$$
$$\tau_g + \tau_h - (s_{gh} + s_{hg}) > 0, \qquad (4.22)$$

otherwise the pair of the job families cannot be intermixed and must be processed sequentially.

The inequality (4.22) guarantees that there will be a time reduction for every switching from one family to another and back (because the overall family setup time $s_{gh} + s_{hg}$ is less than the dead time $\tau_g + \tau_h$ eliminated by the switching).

The obvious strategy for designing the $\phi(g,h)$ is to maximize the number of such switchings. In this case, switching to the next job family should occur as soon as the resource of the previous job family is restored.

If inequalities (4.20) – (4.22) are satisfied, then intermixing of the job families takes place, and one of the families (either g or h) will be completely processed. Suppose, that this family is g, i.e., $\left\lfloor \frac{n_g^*}{k_g} \right\rfloor \leq \left\lfloor \frac{n_h^*}{k_h} \right\rfloor$, where $\lfloor \cdot \rfloor$ stands for the integer part, and the equality

$$\sum_{i=1}^{n} \chi_g(\phi_i) = n_g^*$$

holds for $\phi(g,h) = (\phi_1, \ldots, \phi_n)$. Sequence $\phi(g,h)$ consists of

$$\left\lfloor \frac{n_g^*}{k_g} \right\rfloor \leq M_{gh} \leq \left\lfloor \frac{n_g^*}{k_g} \right\rfloor + 1$$

switchings from family g to h and of

$$\left\lfloor \frac{n_g^*}{k_g} \right\rfloor - 1 \leq M_{hg} \leq \left\lfloor \frac{n_g^*}{k_g} \right\rfloor$$

switchings from family h to g. The time reduction $T(g,h)$ for these switchings is as follows:

$$\begin{aligned}
T(g,h) &= M_{gh}(\tau_g - s_{gh}) + M_{hg}(\tau_h - s_{hg}) = \\
&= M_{hg}\underbrace{(\tau_g + \tau_h - (s_{gh} + s_{hg}))}_{>0} + \underbrace{(M_{gh} - M_{hg})}_{0 \leq \cdot \leq 2}(\tau_g - s_{gh}).
\end{aligned} \quad (4.23)$$

The first $k^* \leq k_g + k_h$ jobs in the switching sequence $\phi(g,h)$, see equation (4.18), are selected by brute-force from the rest of $a_g = n_g^* - k_g \left\lfloor \frac{n_g^*}{k_g} \right\rfloor$ jobs in family g and $a_h, (0 \leq a_h \leq k_h)$ available jobs in family h, with the objective to maximize $T(g,h)$.

The designed sequence $\phi(g,h)$ gains maximal time reduction for intermixing of the pair of the job families (g,h):

$$T(g,h) \rightarrow \max.$$

The new number of jobs that needs to be processed in the subsequent intermixing is specified as follows:

$$n_h^* := n_h - \left(k_h \left\lfloor \frac{n_g^*}{k_g} \right\rfloor + a_h\right),$$
$$n_g^* := 0, \qquad l := h.$$

The complexity of the algorithm for intermixing a pair of job families is as follows:

$$O\left(\frac{2^{k^*}}{\sqrt{k^*}} + \frac{n}{k^*}\right), \tag{4.24}$$

where $n = N/F$ is the average number of jobs in a family and $k^* = k_g + k_h$ is the minimal number of jobs in both families to restore the resources of each other; k^* is an order of magnitudes smaller than n. Assuming $k^* = O(\ln(n))$, the efficiency is close to linear with respect to n.

Finally, both algorithms are designed: the general one that specifies the sequence π in which the pairs of the job families are to be processed, and the local one that intermixes each pair in the sequence in an optimal way.

The resulting sequence σ^* in which the jobs in the schedule are processed is as follows:

$$\sigma^* = \phi(\pi_1, \pi_2), \phi(\pi_{r_1}, \pi_3), \dots, \phi(\pi_{r_{F-1}}, \pi_F), \tag{4.25}$$

where π_{r_i}, $i = 1, \dots, F-1$ is the job family that remains underprocessed after intermixing of the ith pair of job families.

The sequence σ^* satisfies the Assumption 3 and guarantees the minimal $C_{\max}$.

For the heuristic greedy algorithm, the overall complexity is polynomial with respect to the number of job families F and close to linear with respect to n, the average number of jobs in a family.

The algorithm was implemented as a computer program that was tested on a PC with standard configuration as of 2009 (2 GHz dual core processor, 1.5 GB RAM). For all the test ExamCards (see Appendix A) the execution time was almost instant: few seconds. Therefore we can conclude that the algorithm can be used for an 'on-line' scan segments scheduling even in MRI examinations with a large number of scans.

4.4   Validation on ExamCards

In this section the results of application of the algorithm to the Philips MRI examination protocols (ExamCards) are presented.

An ExamCard is a file that contains all the information on the several scans that are necessary for the MRI examination. These are durations of the scans, TR-lengths, SAR levels, gradient waveforms, etc. For more information about the ExamCards see Section 2.4.2.

The scan segments scheduling algorithm from Section 4.3.2 was applied to 1.5 tesla (T) and 3 T ExamCards of Philips Achieva 1.5T and 3.0T systems accordingly. One can find an image of Philips Achieva 3.0T system in Figure 2.5 on page 16.

The relative MRI examination time reduction was calculated for a set of various ExamCards. These ExamCards were downloaded from NetForum[1], which is a web facility where hospitals that utilize Philips MRI systems can share their routine ExamCards.

The ExamCards available in NetForum were created by hospitals for their own purposes, and there is a high probability that those represent a set of the examinations that are performed in hospitals on a daily basis.

In total 57 ExamCards were downloaded, where 31 were for 1.5T MR system and 26 for the 3.0T system. However, the time reduction algorithm was applicable to less than a half of those ExamCards, see Table 4.1. The other part of the Exam-Cards either did not have any SAR limited scans or contained scans that cannot be intermixed with others (e.g. breath-hold protocols, protocols with contrast agent injection). A detailed list of the downloaded ExamCards is provided in Appendix A.

	Achieva 1.5T	Achieva 3.0T
Total amount	31	26
Containing SAR limited scans	8	19
Containing changeable SAR limited scans	7	17

Table 4.1: ExamCards per MRI system type.

The downloaded ExamCards are from various clinical areas (with various anatomies). Table 4.2 provides an overview of the number of ExamCards per MRI system for different clinical areas. Some ExamCards are used in more than one clinical area (see Appendix A), and in the table those are counted several times.

Clinical area	Achieva 1.5T	Achieva 3.0T
Body	7 (3)	5 (5)
Cardiac	7 (0)	1 (1)
Musculoskeletal	2 (0)	7 (5)
Neuro	10 (2)	8 (4)
Oncology	5 (2)	1 (1)
Pediatric	1 (1)	1 (1)
Vascular	5 (2)	5 (2)

Table 4.2: Number of used ExamCards per MRI system for different clinical areas. In brackets are the numbers of ExamCards that contain changeable SAR limited scans.

According to the IEC short-term SAR limitations for MRI examinations, see [CEI/IEC 60601-2-33 2008], the SAR over any 10 s period shall not exceed 3

---
[1]MRI NetForum community: http://netforum.medical.philips.com

times its long-term limits (see section 2.5.1). In order to comply with upcoming IEC edition, the limitations in our calculations were assumed to be even stronger. We considered that the SAR over any 5 s can not be exceeded more than 2 times (i.e., 200% relative SAR level). The results of calculation of relative time reduction are represented in Figures 4.4 – 4.9.
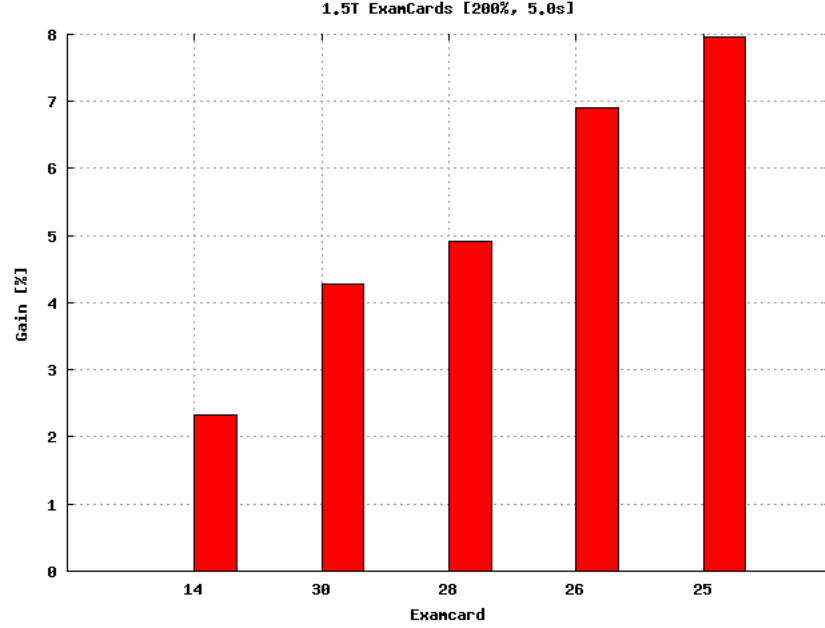

Figure 4.4: Examinations with maximum relative time reduction (as a percentage of the original examination length) for 1.5 T MRI systems.

The relative time reduction was calculated as a percentage of the original examination time. In Figures 4.4 – 4.5 the ExamCards with maximal gain are presented. In those bar plots the ExamCards are labeled with numbers, the corresponding names of the ExamCards and further details (e.g., clinical areas) can be found in Appendix A. From those figures one can conclude that the maximum gain for 3 T examinations is about 2 times larger then for the 1.5 T ones: 22% against 8%. For the average gain a similar conclusion can be made. This is partly because 3 T examinations more often contain SAR-limited scans, see Table 4.1. We assume that for the 7 T MRI systems the time reduction increases proportionally.

Some distinctions can be observed between the examinations from different clinical areas. The most 'gainful' ExamCards (where the maximum relative time reduction was obtained) were from the *body* and the *neuro* clinical areas. The *neuro* is recognized among the most popular clinical areas of the MRI systems, which is quite promising. However we do not have enough data to calculate the weighted average based on the statistics of the clinical usage of those ExamCards.
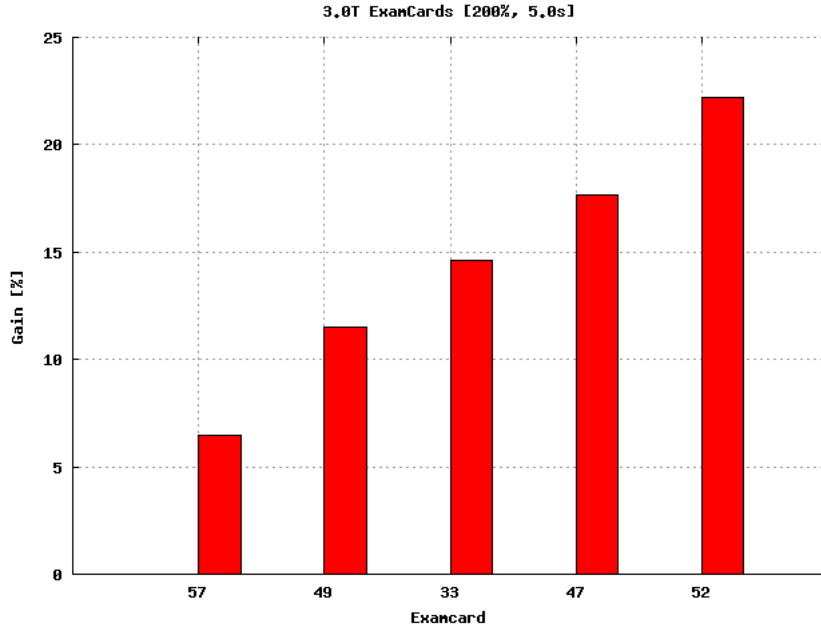
Figure 4.5: Examinations with maximum relative time reduction (as a percentage of the original examination length) for 3 T MRI systems.

Furthermore a sensitivity analysis was performed for various SAR limitations. The results of such analysis for an ankle and a brain examinations are presented in Figures 4.6 and 4.7. Those plots display the relative time reduction of the examinations as the percentage of the original durations of the examinations (that were about 15 min long). On the upper sensitivity analysis plots 4.6(a) and 4.7(a) the duration of high-SAR segment varies on the x-axis, whereas the maximum SAR level is fixed to 200%. On the lower sensitivity analysis plots 4.6(b) and 4.7(b) the dual situation is displayed: the size of the high-SAR segments is fixed to 5 s whereas the maximum allowed SAR level varies from 160% up to 300%.

One can see that the curves saturate, but the rates of convergence are different for different ExamCards. Although the selection of the high-SAR segment size between 5 and 10 s seems to be a reasonable compromise, whereas the maximum SAR level is better to keep as high as it is allowed by the IEC limitations.

The sensitivity analysis was performed for all the ExamCards that contained SAR limited scans. The results of these calculations for different MRI systems are presented in Figures 4.8 – 4.9. The variability in the shapes of the curves can be explained by the discrete structure of scans: different lengths of the TRs influence the optimal high-SAR segment size. Also different durations and SAR-levels of the scans within the ExamCard influence the total gain.

In certain cases the gain can be even negative, this means that the total setup time is greater than the dead time. For instance, this happens when the high-
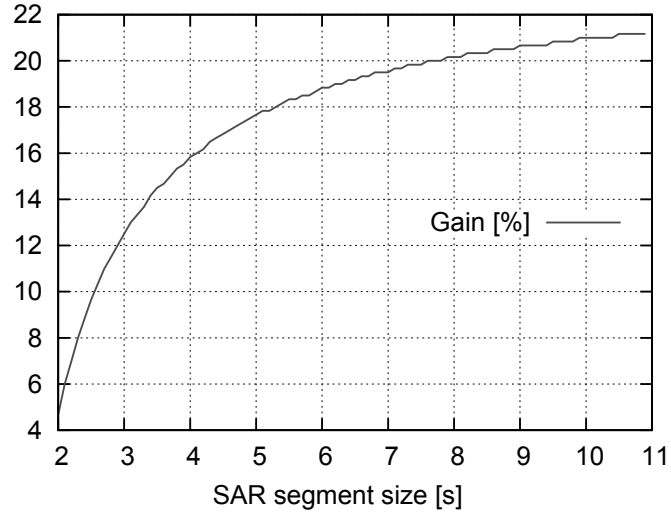
(a) Max SAR is fixed to 200%
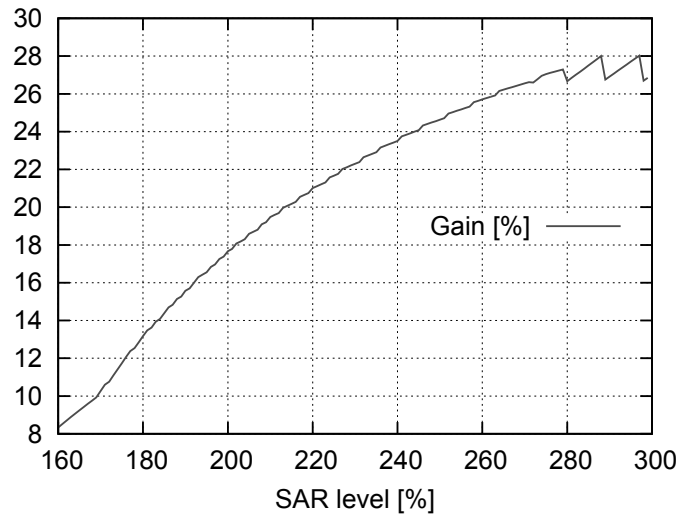
(b) high-SAR segment length is less than 5 s

Figure 4.6: Sensitivity plots of relative time reduction (Gain) for 3 T Ankle ExamCard: (a) max SAR level is fixed to 200%; (b) max high-SAR segment size is fixed to 5 s.
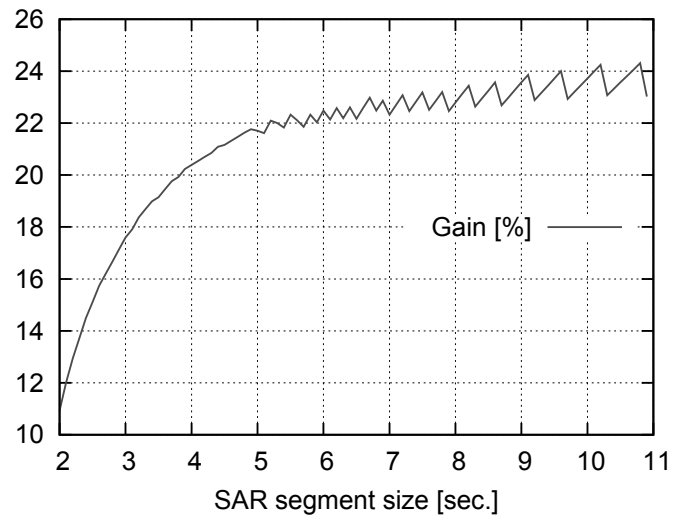
(a) Max SAR is fixed to 200%
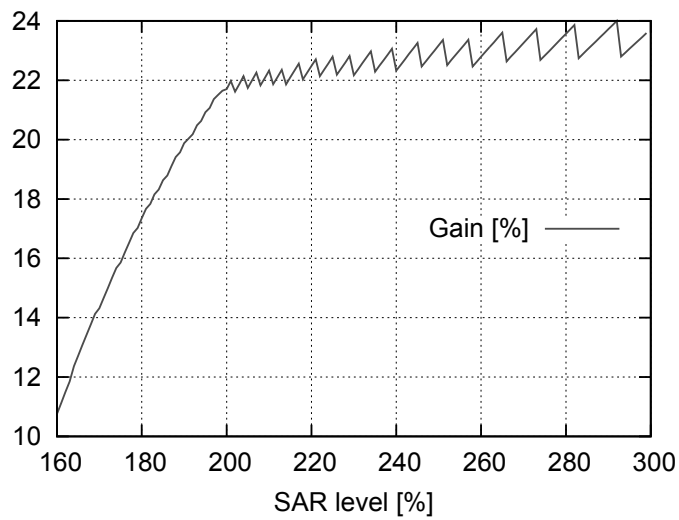
Figure 4.7: Sensitivity plots of relative time reduction (Gain) for 3 T Brain ExamCard: (a) max SAR level is fixed to 200%; (b) max high-SAR segment size is fixed to 5 s.

SAR segment length is chosen to be very short, thus too many switchings between the scans take place and the total setup time exceeds the eliminated dead time.

From the results presented above one can see that even for the limitations stronger than posed by current IEC standards edition the relative time reduction obtained by the algorithm is up to 22% on routine MRI examinations on 1.5 T and 3 T MRI systems. Moreover, the time reduction is expected to increase for the examinations on 7 T MRI systems.
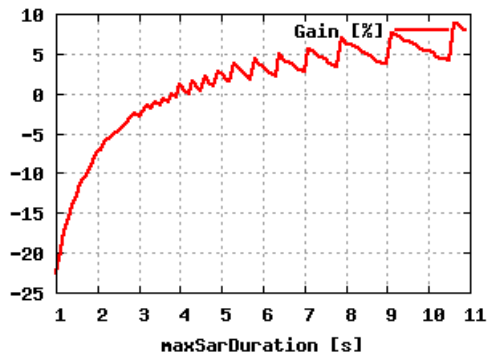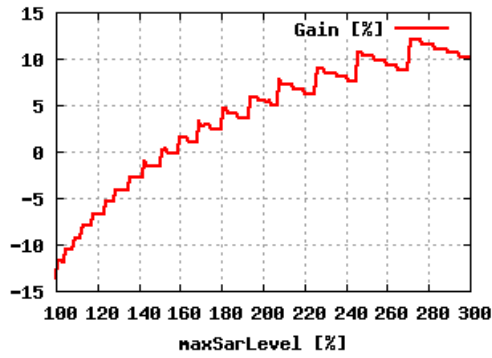
(a) ExamCard 28
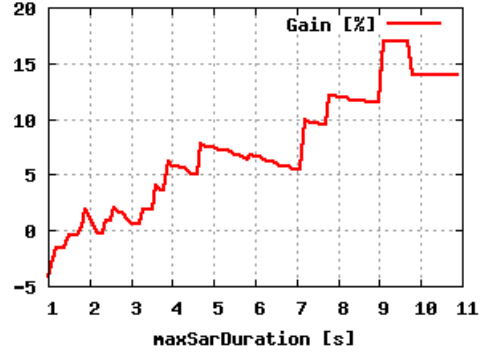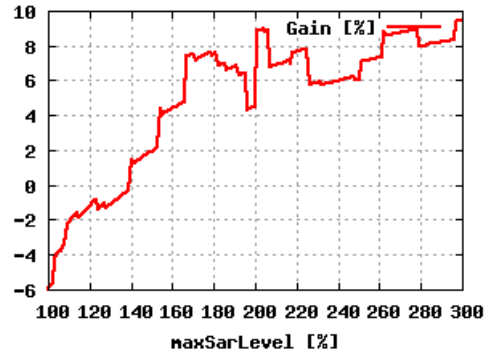
(b) ExamCard 26
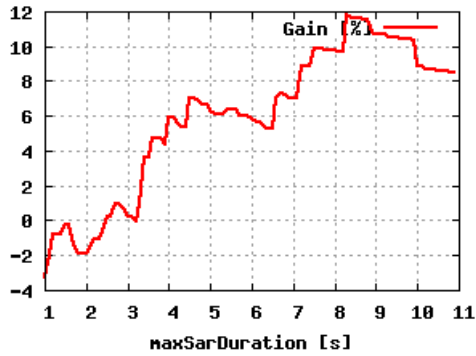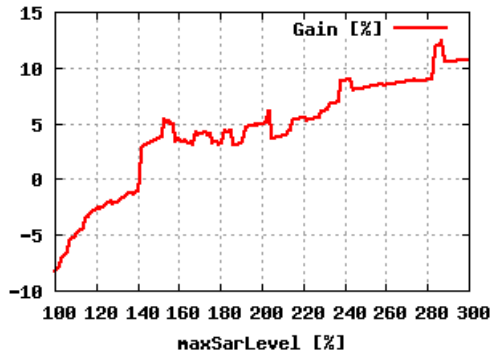
(c) ExamCard 25
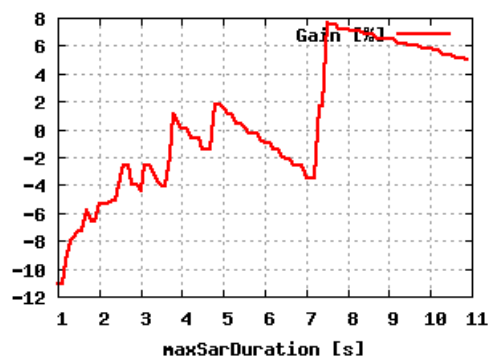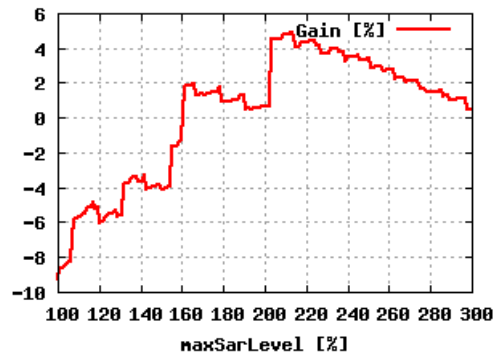
(d) ExamCard 30

Figure 4.8: Sensitivity plots of relative time reduction (i.e., Gain) for various 1.5 T ExamCards.
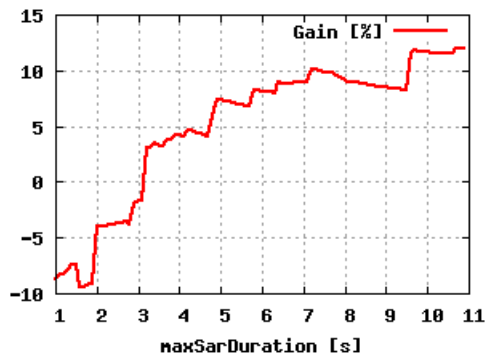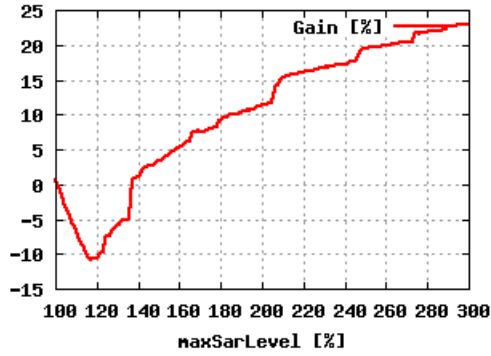
(a) ExamCard 47
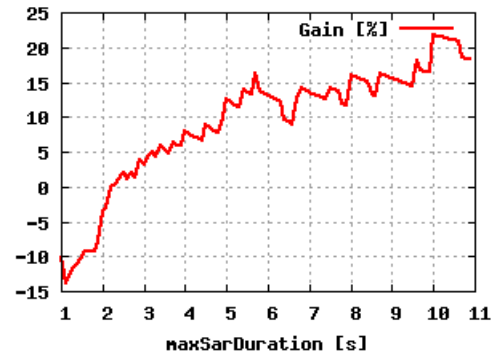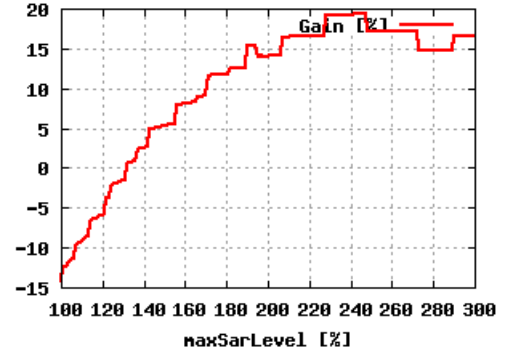
(b) ExamCard 33
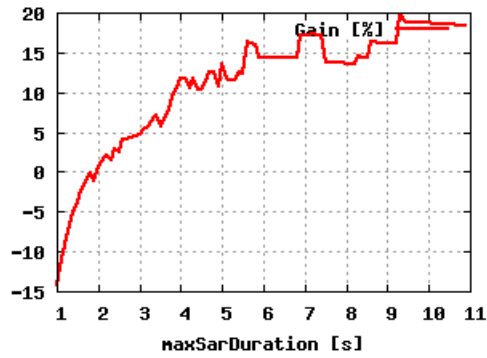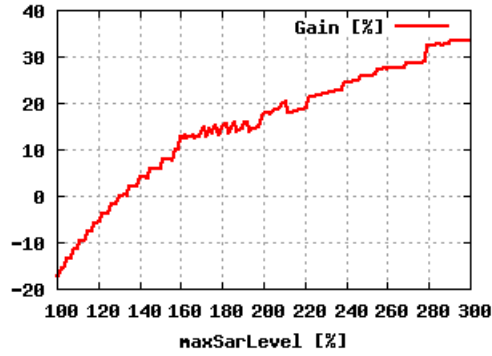
(c) ExamCard 52
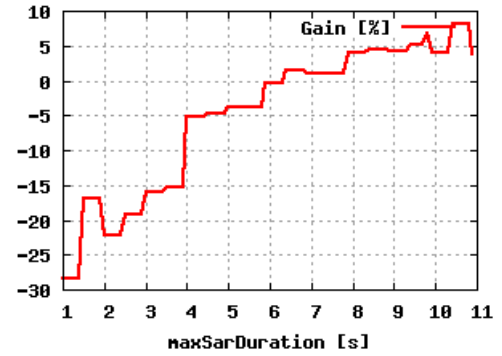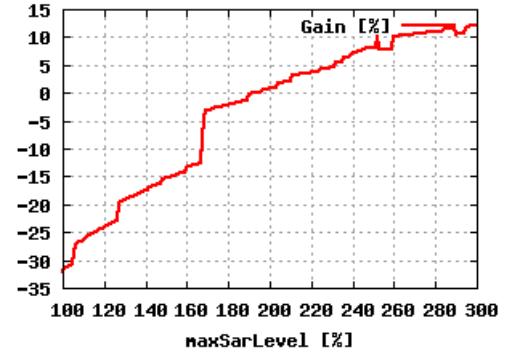
(d) ExamCard 57

Figure 4.9: Sensitivity plots of relative time reduction (i.e., Gain) for various 3 T ExamCards.

Chapter 5

Scheduling with hardware temperature constraints

In this chapter the Gradient Chain system hardware temperature case of the MRI examination time reduction problem is described. This hardware includes gradient amplifiers and gradient coils. These components can overheat during the MRI examination. For the insights into the problem see Sections 2.2.2, 2.5.2, 2.5.3, and Chapter 3. For the literature review on the corresponding scheduling problem see Section 3.3.2. Most parts of this chapter were published in [Ivanov et al. 2009b],[Ivanov et al. 2009c] and [Ivanov et al. 2010b].

The outline of the chapter is as follows. First, the formal problem statement is presented. Next, our algorithms that solve the problem are described. Finally, the results of 'proof of concept' MRI experiments are reported that demonstrate the time reduction obtained by application of the algorithms.

5.1 Problem statement

The problem of scan segments scheduling with hardware temperature constraints was introduced in Sections 3.2 and 3.2.2 of Chapter 3. In the current section the problem is posed in more detail with respect to the temperature dynamics of the MRI system hardware.

As it was stated in Chapter 3, the temperature of the amplifiers and coils depend

on the sequence in which the job families are processed. These temperatures should be kept below the maximal level $T_{\text{amp}}^{\max}$ for amplifiers and $T_{\text{coils}}^{\max}$ for coils. Moreover, all the temperature constraints must be satisfied simultaneously, see equation (3.5) in Section 3.2.2. The setup times between the job families can be omitted, since the processing times of the jobs, i.e., scan segments, is in order of magnitudes longer than the setup times (minutes against few seconds).

There are either one or two amplifiers in the MRI system, see Section 2.2.2. In the case of two amplifiers, they are connected in parallel, thereby the power load is supposed to be uniformly distributed between the amplifiers. For the sake of brevity, we assume without loss of generality that there is only one amplifier. Temperatures of the amplifier $T_{\text{amp}}(k)$ and *hot spots* of the coils $T_{\text{hotspot},h}(k)$ after processing the kth job in sequence σ, see equation (3.7) in Section 3.2, can be calculated by the following recurrent equation:

$$\begin{cases} \mathbf{T}(0) = \mathbf{T}^0 \triangleq \left[T_{\text{amp}}^0, T_{\text{hotspot},1}^0, \ldots, T_{\text{hotspot},H}^0 \right]^{\text{T}} \\ \mathbf{T}(k) = A_{\sigma_k} \cdot \mathbf{T}(k-1) + B_{\sigma_k} \ . \end{cases} \tag{5.1}$$

Here $\mathbf{T}(\mathrm{k})$ denotes vector of temperatures of the amplifier and all the hot spots of the coils after processing the kth job. $\mathbf{T}^0$ is the vector of initial temperatures, with H being the number of the hot spots in the gradient coils ($[\ \cdot\ ]^{\text{T}}$ denotes a column vector). The maximal temperature during processing of the kth job is expressed by $\mathbf{T}(k-1) + M_{\sigma_k}$ and should not exceed the limits:

$$\mathbf{T}(k-1) + M_{\sigma_k} \leq \mathbf{T}^{\max}, \tag{5.2}$$

where $\mathbf{T}^{\max}$ is the vector of the temperature constraints, the same as in equation (3.5) in Section 3.2.2. The parameters B_{σ_k} and M_{σ_k} are vectors, whereas A_{σ_k} is a diagonal matrix. Here the relations $<$ and $\leq$ are *component-wise.*

The parameters A_{σ_k}, B_{σ_k}, and M_{σ_k} are predefined for each job family σ_k, where $\sigma_k \in \{1, \ldots, F\}$. Moreover, the following inequalities hold:

$$\begin{aligned} \mathbb{O} < A_{\sigma_k} < \mathbb{I}, \\ [0, \ldots, 0]^{\text{T}} < B_{\sigma_k} \leq M_{\sigma_k}, \end{aligned} \tag{5.3}$$

The initial temperature conditions $\mathbf{T}^0$ and constraints $\mathbf{T}^{\max}$ are also given:

$$\mathbf{T}^0 \ \leq \mathbf{T}^{\max}. \tag{5.4}$$

As for the *dummy* job family ($\sigma_k = 0$), see Section 3.2.2, the parameters and initial conditions are as follows:

$$\begin{aligned} A_0 &\triangleq \varepsilon < \mathbb{I}, \\ B_0 &\triangleq [0, \ldots, 0]^{\text{T}}, \\ M_0 &\triangleq [0, \ldots, 0]^{\text{T}}. \end{aligned} \tag{5.5}$$

Here $\mathbb{O}$, $\mathbb{I}$, and ε denote diagonal $H + 1$-by-$H + 1$ matrices with zeroes, ones, and relatively small elements on the main diagonal (and zeroes elsewhere), respectively. All the vectors in equations (5.1) – (5.5) have dimension of $H + 1$.

For each particular type of scan $f \in \{1, \dots F\}$ the matrix A_f, the vectors B_f and M_f can be easily calculated knowing the gradient waveform $G_f(t)$ (hence the input power $P_f(t)$ and the primary currents $I_x(t), I_y(t), I_z(t)$) and the length t_f of the scan segment. The first elements of the diagonal matrix A_f and of the vectors B_f, M_f are responsible for the temperature dynamics of the amplifier. They can be calculated by the following equations, [Ivanov et al. 2009c]:

$$A_f[1,1] = \exp\left(\frac{-t_f}{\tau}\right),$$

$$B_f[1] = \frac{1}{\tau} \int_0^{t_f} e^{\frac{-(t_f - t')}{\tau}} \theta P_f(t') \mathrm{d}\, t' + S_f \left(1 - A_f[1,1]\right),$$

$$M_f[1] = \max_{0 \le t \le t_f} \frac{1}{\tau} \int_0^{t} e^{\frac{-(t - t')}{\tau}} \theta P_f(t') \, \mathrm{d}\, t' + S_f,$$

(5.6)

where $\tau \triangleq \max_i \tau_i$ is the longest of the time constants from equation (2.7) in Section 2.5.2. Parameter $\theta = \theta_i$ is the corresponding thermal resistance from the same equation. Temperatures of other sections of the transistor with lower time constants $\tau_i < \tau$ are considered to be in the steady state. (Practically, one of the time constants in the duty cycles of the MRI amplifier is in order of magnitudes greater than others.) These temperatures are aggregated in the parameter S_f:

$$S_f = \sum_{i \in \{1, \dots, k\} \setminus \{m\}} T_{ss}^i, \qquad m = \arg \max_i \tau_i.$$

(5.7)

Here the steady state temperatures T_{ss}^i are known for each section i in the transistor; in a case of constant power dissipations, they are expressed by equation (2.8) in Section 2.5.2.

The rest of the vectors and the matrix elements $B_f[h], M_f[h], A_f[h,h]$, $h = 2, \dots, H + 1$ can be calculated based on equation (2.11) for the temperature rise in the gradient coils in a similar way, see Section 2.5.3:

$$A_f[h,h] = \sum_{\substack{i = 0, \dots, K_a \\ a = x,y,z}} \exp\left(\frac{-t_f}{\tau_{\mathrm{therm},a,i}}\right),$$

$$B_f[h] = \Delta T(t_f),$$

$$M_f[h] = \max_{0 \le t \le t_f} \Delta T(t).$$

(5.8)

where $h = 2, \ldots, H + 1$ ranges through all the hot spots; $A_f[a, b] = 0$ if $a \neq b$; the function $\Delta T(t)$ and the parameters $\tau_{\mathrm{therm},a,i}$ and K_a are the same as in equation (2.11).

The objective is to find a switching sequence σ, see equation (3.7) in Section 3.2.2, for which all the jobs from the job families are processed, while keeping the number of *dummy* jobs minimal. Therefore

$$\sum_{n=1}^{N} \chi_g(\sigma_n) = n_g, \qquad g \in \{1, \ldots, F\}$$
$$\sum_{n=1}^{N} \chi_0(\sigma_n) = n_0 \longrightarrow \min.$$
(5.9)

has to be satisfied. Here $\chi_g \to \{0, 1\}$ is the characteristic function of the jobs that belong to the family g:

$$\chi_g(f) = \begin{cases} 1 & \text{if } f = g, \\ 0 & \text{if } f \neq g. \end{cases}$$
(5.10)

It is important to notice, that by decreasing the number of *dummy* jobs the maximum completion time is decreased. Thereby our problem is equivalent to the problem of the makespan minimization:

$$C_{\max} \to \min.$$

5.2   Scheduling algorithms

For the literature review on the scheduling problems with thermal constraints the reader is referred to Section 3.3.2 of Chapter 3. Based on that review, a conclusion can be made that the problem of jobs scheduling with respect to thermal constraints in general is NP-hard, and so no polynomial time algorithm is available for it. However in the case of scans segments scheduling the NP-hardness of the problem is not crucial, since the number of the jobs (i.e., segments) is typically low. The length of each segment is in order of a minute, whereas a typical duration of MRI examination ranges from 10 to 30 min, see Section 2.4. This implies a relatively small number of jobs. Moreover, not all the scans are intermixable during the examination that reduces the number of jobs even further, see Section 3.1. Therefore the problem can be solved by a backtracking algorithm, see e.g. [Cormen et al. 1989], [Knuth 1997], in a reasonable time.

5.2.1 Backtracking algorithm

The backtracking algorithm systematically searches for a solution to the problem among all the available cases. Solution σ is built incrementally, and partial candidates σ' are rejected if they cannot be completed to a valid solution. If a partial candidate σ' is rejected, the algorithm backtracks by removing the trailing value from the σ', and then proceeds by trying to extend σ' with alternative values. The solution space, Θ, for the problem is as follows:

$$\Theta = \underbrace{F \times F \times \cdots \times F}_{N \text{ times}} = F^N, \tag{5.11}$$

where N is defined by equation (3.6); each solution $\sigma \in \Theta$. The traversal of the solution space can be represented by a depth-first traversal of a tree. Not all the branches of the searching tree are traversed and a large number of candidates is eliminated with a single test. The portion of the solution space that needs to be traversed highly depends on the *validity criteria*, i.e, the test used in checking for acceptable solution candidates.

A typical backtracking algorithm consists of four procedural parameters: *First, Next, Reject*, and *Accept*. These procedures should do the following:

- *First(s)* generates a first extension of a candidate s,

- *Next(s)* generates a next alternative extension of the candidate,

- *Reject(s)* returns *TRUE* if the partial candidate s is rejected,

- *Accept(s)* returns *TRUE* if valid solution s is found

The algorithm is initialized with an empty vector $S = [\,]$ and proceeds backtracking until a solution is found (accepted) or all the partial candidates are rejected. The second parameter of the algorithm is the vector T of the amplifier and the coils hot spots temperatures.

In the algorithms global constants-vectors $\mathbf{T}^{\max}$ and n, as well as variables k and *Solution* are used. The constant $\mathbf{T}^{\max}$, the same as in equation (5.2), represents the vector of limitations for the temperature of the amplifiers and coils. The size of the vectors T and $\mathbf{T}^{\max}$ is $H + 1$, where H is the number of hot spots (see Section 2.5.3).

The constant $n = [n_1, n_2, \ldots, n_F, n_{F+1}]$ is a vector of size $F + 1$ where $n_i, i \in \{1, \ldots, F\}$ denotes the number of job from family i to be processed and $n_{F+1} = n_0$ represents a number of *dummy* jobs to cool down the amplifiers.

The variable k is the vector of size $F + 1$ where $k_i, i \in \{1, \ldots, F + 1\}$ denotes the number of jobs in each family that has been already processed on the current

---

Procedure 5.1 BACKTRACK1(S,T)

Require: Vectors S, T

 1: **if** REJECT(S,T) **or** *Solution* $\neq [\,]$ **then**
 2:     **return**
 3: **else if** $k = n$ **then**
 4:     *Solution* $\leftarrow S$
 5:     **return**
 6: **end if**
 7: $V \leftarrow$ FIRST(S)
 8: **while** $V \neq [\,]$ **do**
 9:     BACKTRACK1$(V,\ T \cdot A_{V[end]} + B_{V[end]})$
10:     $k[V[end]] \leftarrow k[V[end]] - 1$
11:     $V \leftarrow$ NEXT(V)
12: **end while**

---

step of the algorithm. In the variable *Solution* the first found valid solution is stored.

The procedures *First — Accept* are listed below.

The procedure *First* searches for a first job family that has jobs to be processed:

Procedure 5.2 FIRST(S)

Require: Vector S
 1: $j \leftarrow \{$ first $i : k[i] < n[i]\}$
 2: **if** $j = \emptyset$ **then**
 3:     **return** $[\,]$
 4: **else**
 5:     $k[j] \leftarrow k[j] + 1$
 6:     **return** $S \cup j$
 7: **end if**

The procedure *Next* searches for the next job that has not exceeded the family size, where $S[end]$ denotes the last element of the partial candidate vector S:

The procedure *Accept* checks if the solution is found. In our case it only checks if $k = n$, so it is written inline: Procedure 5.1, lines 3 – 5.

The procedure *Reject* abandons partial candidates if they do not satisfy the restrictions.

This algorithm traverses all possible solutions. However, a large number of job families F and a large number of jobs N can result in large computational efforts. For that case a slight modification of the backtracking procedure is designed

---

Procedure 5.3 NEXT(S)

Require: Vector S
  1: $j \leftarrow \{$ first $i > S[end] : k[i] < n[i]\}$
  2: **if** $j = \emptyset$ **or** $S = [\,]$ **then**
  3:    **return** $[\,]$
  4: **else**
  5:    $k[j] \leftarrow k[j] + 1$
  6:    $S[end] \leftarrow j$
  7:    **return** S
  8: **end if**

Procedure 5.4 REJECT(S,T)

Require: Vectors S, T
  1: **if** $T + M_{S[end]} > \mathbf{T}^{\max}$ **then**
  2:    **return** TRUE
  3: **else**
  4:    **return** FALSE
  5: **end if**

(Procedures 5.5 and 5.6) that backtracks no more than one step back.

Finally, the main body of the algorithm with initialization of all the parameters is presented in Algorithm 5.1. The input for the algorithm is as follows:

- vector $n' = [n'_1, \ldots, n'_F]$ of size F, where $n'_i, i \in \{1, \ldots, F\}$ denotes the number of jobs from family i to be processed,

- thermal parameters of the gradient system: initial temperatures $\mathbf{T}^0$ and the maximum allowable temperature $\mathbf{T}^{\max}$,

- matrices A, B, M with the thermal parameters for the dynamical system equations (5.1) – (5.2).

The complexity of the algorithm in the heuristic case is $O(H \cdot N \cdot n_0)$, where N is the number of jobs, see equation (3.1), H is the number of hot spots, and n_0 is the minimum size of the *dummy* jobs family. If the linear search for n_0 in the Algorithm 5.1 (lines 2 and 7) is replaced by a binary search algorithm (see e.g. [Knuth 1997]), then the overall complexity will reduce to $O(H \cdot N \cdot \log n_0)$. Notice that n_0 is not the size of the input, but the size of the output. The value of n_0 in not known in advance, but it can be regulated by a design parameter p_0, i.e. the processing time of the *dummy* job (see Section 3.2.2).

To summarize, in this section we described two algorithms that are designed to solve the scheduling problem. The first algorithm is based on a complete

---

Procedure 5.5 BACKTRACK2(S,T)

Require: Vectors S, T

 1: **if** REJECT2(S,T) **or** $Solution \neq [\,]$ **then**
 2:    **return**
 3: **else if** $k = n$ **then**
 4:    $Solution \leftarrow S$
 5:    **return**
 6: **end if**
 7: $V \leftarrow$ FIRST(S)
 8: **while** $V \neq [\,]$ **do**
 9:    BACKTRACK2$(V,\ T \cdot A_{V[end]} + B_{V[end]})$
10:    $V \leftarrow$ NEXT(V)
11: **end while**

Procedure 5.6 REJECT2(S,T)

Require: Vectors S, T

 1: **if** $T + M_{S[end]} > \mathbf{T}^{\max}$ **then**
 2:    $k[S[end]] \leftarrow k[S[end]] - 1$
 3:    **return** TRUE
 4: **else**
 5:    **return** FALSE
 6: **end if**

Algorithm 5.1 BT ALGORITHM$(\mathbf{T}^{\max}, \mathbf{T}^0, A, B, M, n')$

Require: Vectors $\mathbf{T}^{\max}, \mathbf{T}^0, n'$; Matrix A, B, M; Scalars F

 1: $Solution \leftarrow [\,]$
 2: $i \leftarrow 0$
 3: **while** $Solution = [\,]$ **do**
 4:    $n \leftarrow n' \cup i$
 5:    $k \leftarrow [\ \underbrace{0, 0, \ldots, 0}_{F \text{ times}}\ ]$
 6:    $\left.\begin{array}{l}\text{BACKTRACK1}([\,], \mathbf{T}^0) \\ \text{BACKTRACK2}([\,], \mathbf{T}^0)\end{array}\right\}$ either exact or heuristic
 7:    $i \leftarrow i + 1$
 8: **end while**
 9: **return** $Solution$

---

backtracking procedure (Procedure 5.1), whereas the second one is a heuristic that backtracks no more than one step back (Procedure 5.5). Both the algorithms were implemented as MATLAB scripts, see [MATLAB 2007], and tested on a number of ExamCards. These tests demonstrated that in most cases the heuristic backtracking procedure provides time reduction close to the exact algorithm, but the solution in not always optimal.

In a typical MRI application the number of job families F is less then 10 and the number of jobs N is less than 100. For these input parameters both the algorithms find solutions in a reasonable time. However the heuristic performs in order of magnitudes faster than the complete backtracking: seconds against several minutes. Based on the performance, we conclude that the heuristic version can be implemented in the MRI software for 'online' scheduling of the scans segments, whereas the exact algorithm is mostly suitable for 'offline' pre-processing of the ExamCards.

5.2.2   Sequencing a periodic set of scan segments

In this section a heuristic approach to the scan segments sequencing problem is designed that is based on the 'thermal aware' processor task sequencing algorithm by [Zhang and Chatha 2010]. In that paper a heuristic algorithm is proposed for sequencing of a periodic set of processor tasks subject to peak temperature constraint $T_{\max}$, see Section 3.3.2. After some adaptation this heuristic algorithm can be applied to sequence the segments of the scans.

The original algorithm by [Zhang and Chatha 2010] searches for a periodic solution to the sequencing problem. That is, a sequence of jobs from all the job families is designed that after multiple repetitions does not exceed the temperature limitations, moreover the amount of the *dummy* jobs in that sequence should be as low as possible, see equation (5.9) in Section 5.1.

In the case of scan segments sequencing, the periodic solution can only be found if all the numbers of the jobs in the families n_j, $j \in 1, \ldots, F$ are multiples of each other. However this is not always satisfied, since n_j are arbitrary numbers. Therefore in the general case the solution-searching procedure should be modified. At first, the algorithm will be designed for a simple case with an equal number of jobs in the families. Then it will be generalized to the case of arbitrary-sized families.

Job families of equal size

Consider a special case, where all the job families contain the same number n of jobs, i.e. the following equation is satisfied:

$$n_i = n_j = n \qquad \forall i, j \in 1, \ldots, F,$$

where $n = N/F$. The last equation does not state the *dummy* job family size, which can be different, i.e., inequality $n_0 \neq n$ can be satisfied (or not).

The solution sequence σ, see Section 3.2.2, is designed to be periodic. It consists out of n repetition of the same sequence π of the job families:

$$\sigma = \underbrace{\pi, \; \pi, \; \ldots, \; \pi}_{n \text{ times}}. \tag{5.12}$$

Here sequence π contains only one job from each job family (except for the *dummy* job family):

$$\chi_g(\pi) = 1, \qquad g \in \{1, \ldots, F\},$$
$$\chi_0(\pi) \leq F + 1,$$

where χ_g is the characteristic function of the jobs that belong to the family g, see equation (5.10). The *dummy* job can be inserted in between any pair of regular job families, which implies that the number of the *dummy* jobs in the sequence can be as much as $F + 1$.

Therefore it is enough to find sequence π of the job families

$$\pi = \; \pi_1, \; \pi_2, \; \ldots, \; \pi_{F'}, \qquad F \leq F' \leq 2F + 1 \tag{5.13}$$

that after n repetitions does not exceed the temperature constraints.

In order to do this, first the thermal profile of a job sequence should be defined. It can be done the same way as by [Jayaseelan and Mitra 2008]. Figure 5.1 demonstrates a thermal profile of a repeating sequence of two jobs from different job families, here only the amplifier temperature is displayed.

After multiple iterations the temperature reaches a steady state where the thermal profile of the job sequence exhibits a recurring pattern. There are several constraints that are satisfied for the recurring thermal profile in the steady state: the initial and the final temperatures of a job π_j in the sequence remain the same for all the executions of the sequence in steady state. The temperatures at the beginning and the end of the sequence are equal in the steady state.

This algorithm first classifies jobs into *cool* and *hot* based on their steady state temperatures. The steady state temperature of a job from family f can be straightforwardly calculated by substituting $\mathbf{T}(k) = \mathbf{T}(k-1) \triangleq \mathbf{T}_f^{ss}$ and $f = \sigma_k$ in equation (5.1), since the temperature at the beginning of the job is equal to the temperature at the end (by the definition of the steady state temperature):

$$\mathbf{T}(k) = A_f \cdot \mathbf{T}(k) + B_f.$$

Hence the steady state temperatures of the amplifier and hot spots of the coils are expressed by the vector

$$\mathbf{T}_f^{ss} \triangleq \left[T_{ss}^1, T_{ss}^2, \ldots, T_{ss}^{H+1} \right]^{\mathrm{T}}, \tag{5.14}$$
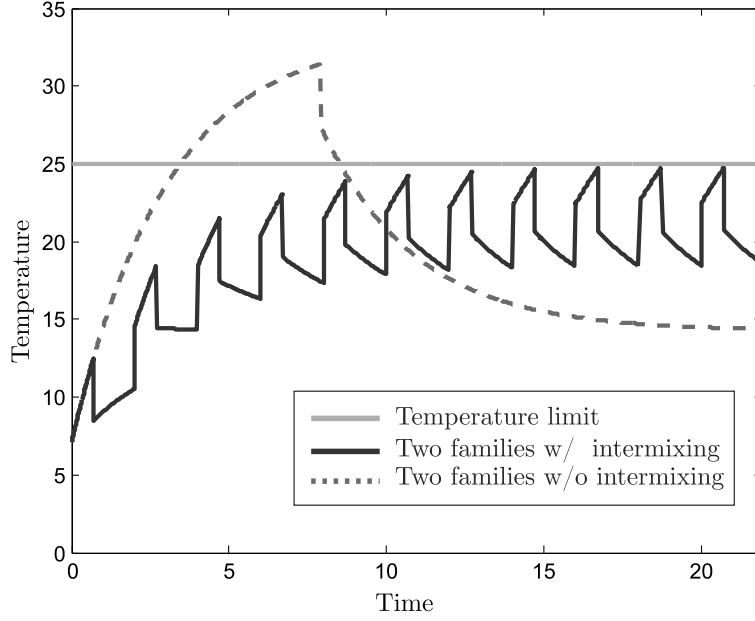
Figure 5.1: Thermal profiles of two job families with and without intermixing. Both profiles are without the dead time.

the components of which are defined by the following equations:

$$T_{ss}^h = \frac{B_f[h]}{1 - A_f[h, h]}, \qquad h = 1, \ldots, H + 1. \tag{5.15}$$

Here the parameters $A_f[h, h]$ and $B_f[h]$ are from equations (5.6) – (5.8).
It should be mentioned that in a case of constant power dissipation $P_f(t) \equiv P$, equation (5.15) for the amplifier steady state temperature ($h = 1$) reduces to a simple equation (2.8) from Section 2.5.2 with addend S_f from equation (5.7):

$$
\begin{aligned}
T_{ss}^1 = \frac{B_f[1]}{1 - A_f[1, 1]} &= \frac{\frac{1}{\tau} \int_0^{t_f} \exp \frac{-(t_f - t)}{\tau} \, \theta P \, \mathrm{d} t}{1 - \exp\left(\frac{-t_f}{\tau}\right)} + S_f = \\
&= \frac{\theta P \left(\exp\left(\frac{t_f}{\tau}\right) - 1\right) \exp\left(\frac{-t_f}{\tau}\right)}{1 - \exp\left(\frac{-t_f}{\tau}\right)} + S_f = \frac{\theta P \left(1 - \exp\left(\frac{-t_f}{\tau}\right)\right)}{1 - \exp\left(\frac{-t_f}{\tau}\right)} + S_f = \\
&= \theta P + S_f.
\end{aligned}
$$

Classification of the job families into *cool* and *hot* is performed by a component-wise comparison of the steady state temperatures vector $\mathbf{T}_f^{ss}$, which is defined by

equations (5.14) – (5.15), with the vector of the maximum temperatures $\mathbf{T}^{\mathrm{max}}$:

$$\text{job family } f \text{ is } \begin{cases} cool & \text{if } \mathbf{T}_f^{ss} \leq \mathbf{T}^{\mathrm{max}} \\ hot & \text{if } \mathbf{T}_f^{ss} \nleq \mathbf{T}^{\mathrm{max}}. \end{cases} \tag{5.16}$$

Here the binary relations are defined as follows. Consider a set $\mathbb{V}^{H+1}$ of vectors of dimension $H + 1$ the components of which are from the set of real numbers $\mathbb{R}$; $\mathbf{T}_f^{ss}$ and $\mathbf{T}^{\mathrm{max}} \in \mathbb{V}^{H+1}$. The component-wise vector inequality

$$\mathbf{v_a} \leq \mathbf{v_b}, \qquad \mathbf{v_a}, \mathbf{v_b} \in \mathbb{V}^{H+1} \tag{5.17}$$

is satisfied if and only if all the components of vector $\mathbf{v_a}$ are no more than components of vector $\mathbf{v_b}$. If any component of vector $\mathbf{v_b}$ is greater than the corresponding component of vector $\mathbf{v_a}$, then this is denoted by $\mathbf{v_a} \nleq \mathbf{v_b}$.

After the jobs families are classified into the *cool* and the *hot* ones, the algorithm generates sequences L_{cool} and L_{hot} that contains all the *cool* and *hot* job families respectively.

The jobs in sequence L_{cool} are sorted in the decreasing order of their power consumption. This order minimizes the temperature at the completion of the job set, see Lemma 2 from [Zhang and Chatha 2010]. Sequence L_{cool} will be used further to lower the temperatures of the *hot* jobs.

The sequence of the *hot* jobs L_{hot} is formed as follows. The *hot* jobs are sorted in the increasing order of their *maximal 'feasible' initial temperatures*. Here 'feasible' are initial temperatures of amplifier and coils that still enables processing of the job, that is, no thermal constraints are violated during processing of the job.

For each particular job family f the vector of these initial temperatures $\mathbf{T}_f^{\mathrm{init}}$ can be directly calculated from equation (5.2):

$$\mathbf{T}_f^{\mathrm{init}} = \mathbf{T}^{\mathrm{max}} - M_f, \tag{5.18}$$

where $\mathbf{T}_f^{\mathrm{init}} \triangleq \left[T_{\mathrm{amp}}^{\mathrm{init}}, T_{\mathrm{hotspot},1}^{\mathrm{init}}, \ldots, T_{\mathrm{hotspot},H}^{\mathrm{init}} \right]^{\mathrm{T}} \in \mathbb{V}^{H+1}$. Since $\mathbf{T}^{\mathrm{max}}$ is constant, $\mathbf{T}_f^{\mathrm{init}}$ are unambiguously defined by M_f for each job family f.

Let us consider a set of vectors $\mathbb{V}^{H+1}$ with components from $\mathbb{R}$. It can be qualified as a vector space with scalars from $\mathbb{R}$, if addition of the vectors is defined component-wise and scalar multiplication is defined by multiplication each component of the vector on the scalar:

$$\mathbf{v_a} + \mathbf{v_b} \triangleq [v_a^1 + v_b^1, \ v_a^2 + v_b^2, \ \ldots, \ v_a^{H+1} + v_b^{H+1}]^{\mathrm{T}},$$
$$x \cdot \mathbf{v_a} \triangleq [x \cdot v_a^1, \ x \cdot v_a^2, \ \ldots, \ x \cdot v_a^{H+1}]^{\mathrm{T}},$$

where $x \in \mathbb{R}$ and $\mathbf{v_a} = [v_a^1, \ldots, v_a^{H+1}]^{\mathrm{T}}$, $\mathbf{v_b} = [v_a^1, \ldots, v_a^{H+1}]^{\mathrm{T}} \in \mathbb{V}^{H+1}$.

The norm on the vector space $\mathbb{V}^{H+1}$ over $\mathbb{R}$ can be defined as follows:

$$\|\mathbf{v}\| \triangleq \max_h \left\{ \left| \frac{v^h}{T_h^{\max}} \right| \right\}, \qquad h \in \{1, \ldots, H+1\}, \ \mathbf{v} \in \mathbb{V}^{H+1} \tag{5.19}$$

where $T_1^{\max} = T_{\mathrm{amp}}^{\max}$, whereas $T_h^{\max} = T_{\mathrm{coils}}^{\max}$, $h = 2, \ldots, H+1$, see equations (2.5) and (2.10) in Section 2.5. It is easy to proof that this norm is equivalent to the classical *maximum norm* $\|\mathbf{v}\|_\infty = \max \left\{ |v^1|, \ldots, |v^{H+1}| \right\}$. Given two job families f and g, the norms of the vectors $M_f, M_g \in \mathbb{V}^{H+1}$ can be compared: either $\|M_f\| \leq \|M_g\|$, or $\|M_g\| \leq \|M_f\|$, or both the inequalities are satisfied.

From the finite set of job families $\mathcal{F} = \{1, \ldots, F\}$ a subset $\mathcal{F}_{\max} \subseteq \mathcal{F}$ can be extracted with vectors M_f, $f \in \mathcal{F}_{\max}$ having the maximal norm.

$$\mathcal{F}_{\max} = \left\{ f \ : \ \|M_f\| = \max_g \max_h \left\{ \left| \frac{M_g^h}{T_h^{\max}} \right| \right\}, \ \begin{array}{l} g \in \mathcal{F} \\ h \in \{1, \ldots, H+1\} \end{array} \right\} \tag{5.20}$$

Job families from $\mathcal{F}_{\max}$ have the lowest *maximal 'feasible' initial temperatures*, since the maximality of M_f norm implies that $\mathbf{T}_f^{\mathrm{init}}$, $f \in \mathcal{F}_{\max}$ have components with the lowest relative temperatures with respect to the corresponding temperature constraints ($T_{\mathrm{amp}}^{\max}$ or $T_{\mathrm{coils}}^{\max}$). This can be shown by component-wise dividing of both sides of equation (5.18) by $\mathbf{T}^{\max}$.

Therefore one of the jobs from $\mathcal{F}_{\max}$ goes first to sequence L_{hot}. Since all the jobs in $\mathcal{F}_{\max}$ can be considered equivalent, they can be enumerated, e.g., in lexicographical order, and added to the head of L_{hot} in that order.

Afterwards, a set $\mathcal{F}' \triangleq \mathcal{F} \setminus \mathcal{F}_{\max}$ is considered, which contains the rest of the *hot* job families. Job families with the lowest *maximal 'feasible' initial temperatures* are selected by means of equation (5.20) with $g \in \mathcal{F}'$ and added to L_{hot} in some order, e.g., lexicographical. These jobs families are then subtracted from $\mathcal{F}'$ and a new set is considered $\mathcal{F}'' \triangleq \mathcal{F}' \setminus \mathcal{F}_{\max}$. The procedure is repeated at most F times before all the *hot* jobs are in L_{hot} sorted in the increasing order of their *maximal 'feasible' initial temperatures*.

As soon as the sequences L_{cool} and L_{hot} are specified, the main part of the algorithm starts. This part is the same as in [Zhang and Chatha 2010]. The initial temperature is set to be maximal $\mathbf{T}^0 := \mathbf{T}^{\max}$. For periodic solutions that approach steady state temperatures, this assignment can be done without loss of generality, see Theorem 1 by [Zhang and Chatha 2010].

First, the algorithm lowers the temperatures only with the *cool* job families. The *hot* job families are picked one by one from the head of the L_{hot} sequence and inserted into the L_{cool} sequence if only if the temperature constraints are not violated. When the cool tasks can not lower the temperature, the *dead time* is added right before the remaining *hot* job families. This procedure will be described in detail below.

Sequence $L_{cool} = l_1, l_2, \ldots, l_{|L_{cool}|}$ is executed and the final temperatures at each cool job family is recorded in a sequence

$$\mathbf{T}^{\mathrm{fin}} = \mathbf{T}^{\mathrm{fin}}_{l_1}, \mathbf{T}^{\mathrm{fin}}_{l_2}, \ldots, \mathbf{T}^{\mathrm{fin}}_{l_{|L_{cool}|}} \tag{5.21}$$

Here $l_i \in \{1, \ldots, F\}$ is the index of job family that is in the ith position in L_{cool}, whereas $\mathbf{T}^{\mathrm{fin}}_{l_i} \in \mathbb{V}^{H+1}$. Temperature profile of the amplifier after execution of this sequence L_{cool} is presented in Figure 5.2.
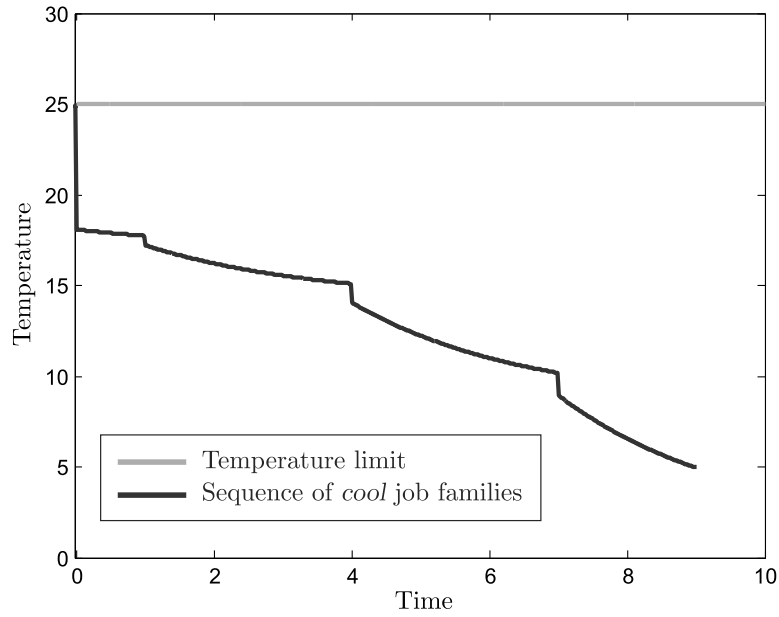
Figure 5.2: Temperature drop after execution of sequence L_{cool} of *cool* job families.

Next step, the hottest job is taken from the head of L_{hot}. Let this job belong to job family f. Then the initial temperatures vector of this job family is $\mathbf{T}^{\mathrm{init}}_f$.

If $\left\| \mathbf{T}^{\max} - \mathbf{T}^{\mathrm{fin}}_{l_{|L_{cool}|}} \right\| < \left\| \mathbf{T}^{\max} - \mathbf{T}^{\mathrm{init}}_f \right\| \triangleq \|M_f\|$, then all the jobs from L_{cool} are not enough to execute the *hot* job after L_{cool} without violating the temperature constraints. In this case an amount of the *dead time* is added right after L_{cool} and before the *hot* job to enable its execution.

Let us consider a favorable case, when

$$\left\| \mathbf{T}^{\max} - \mathbf{T}^{\mathrm{fin}}_{l_{|L_{cool}|}} \right\| \geq \left\| \mathbf{T}^{\max} - \mathbf{T}^{\mathrm{init}}_f \right\| \triangleq \|M_f\|.$$

This implies that there are enough jobs in sequence L_{cool} to enable execution of the *hot* job from family f.
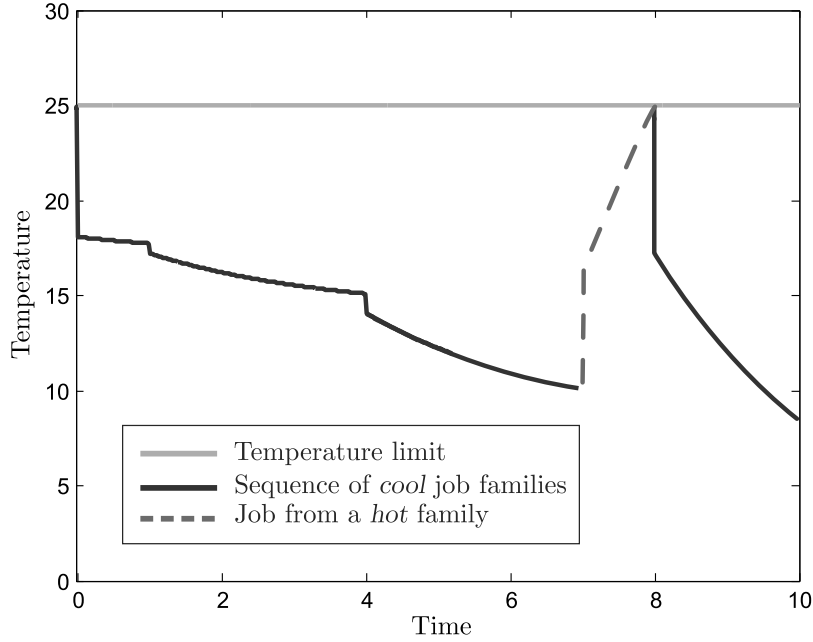
Figure 5.3: Temperature profile after insertion of the first *hot* job from the head of L_{hot} into the *cool* jobs sequence L_{cool}.

The algorithm searches for the ith position in $\mathbf{T}^{\text{fin}}$ for which the following inequality is satisfied:

$$\left\| \mathbf{T}^{\max} - \mathbf{T}^{\text{fin}}_{l_i} \right\| \geq \left\| \mathbf{T}^{\max} - \mathbf{T}^{\text{init}}_{f} \right\| > \left\| \mathbf{T}^{\max} - \mathbf{T}^{\text{fin}}_{l_{i-1}} \right\| .$$

Then the *hot* job is inserted between l_i and l_{i+1} in L_{cool}, see Figure 5.3. Next, the jobs execution sequence π, see equation (5.13), is updated. The first i jobs are deleted from L_{cool} and added to the tail of π. The first *hot* job f is deleted from L_{hot} and added to the tail of π. Moreover, $\mathbf{T}^0$ is updated by the final temperatures of the current schedule π, i.e. the temperatures after execution of the *hot* job in Figure 5.3. This procedure is repeated until all the *cool* jobs that can lower the temperature of the *hot* ones are utilized, i.e. L_{cool} is empty.

After this the only option left to lower the temperatures of the *hot* jobs is to add the *dead time*. An amount of the *dead time* is inserted right before each remaining *hot* task.

The computational complexity of this algorithm is

$$O\left(H \cdot F^2 \right), \tag{5.22}$$

the same complexity as the algorithm by [Zhang and Chatha 2010].

Job families of arbitrary sizes

Let us consider the general case of job families with different number of jobs

$$\exists\, i \neq j : \; n_j \neq n_i \qquad i, j \in 1, \ldots, F.$$

Then there exists a job family j_m with the lowest number of jobs

$$n_{j_m} \leq n_j, \qquad \forall j \in 1, \ldots, F.$$

In this case, after n_{j_m} repetitions of sequence π (see equation (5.13)) generated by the previous algorithm, all the jobs from the family j_m will be expended. After that, a new periodic solution will be built up out of the residual job families. This procedure will be repeated several times before all the job families expire.

In the worst case scenario, with all the job families having different amounts of jobs, this procedure is repeated F times. Therefore the worst-case complexity of the algorithm is in order of

$$O\left(H \cdot F^3 \right), \tag{5.23}$$

which is polynomial with respect to the number of job families F.

To compare, the worst-case complexity of the backtracking algorithm (BT) from Section 5.2.1 can be as high as $O\left(2^N \cdot H \cdot n_0 \right)$, where $N = n\,F$ is the total number of jobs in the families (BT stands for Algorithm 5.1 on page 78). However, in the BT a lot of subtrees are eliminated from the depth-first searching tree with a single test, if the partial solution-candidate violates the thermal constraints. This is similar to the *backtracking with cutoff* from [Stone and Sipala 1986]. The average number of nodes visited in their algorithm is only $O(d)$, where d is the depth of the searching tree (in our case $d = N$). This implies that the average complexity of the BT algorithm can be as low as $O(N \cdot H \cdot n_0)$.

To summarize, both the algorithms, the backtracking (BT) and the periodic-solutions based (PS), are suitable for intermixing the scan segments with respect to temperature constraints. The PS algorithm performs better in the worst-case scenario, whereas the BT algorithm is able to find more accurate solutions and the average performance is acceptable for a moderate number of jobs N (especially the performance of the heuristic version of the BT algorithm). Both the algorithms can be implemented in the MRI software for scheduling of the scan segments, however for the large-scale problems with a large number of scan segments the computational time of the PS algorithm is lower.

5.3 MRI experiments

Calculations of MRI examination time reduction were performed for a big set of routine brain examinations. Those calculations revealed that the most limited

from the temperature point of view are so-called *diffusion* brain examinations. Those diffusion examinations were also the most 'gainful': the time reduction achieved was up to 25%, whereas the average time reduction for brain examinations was about 10%.

Therefore for the 'proof of concept' experiment an examination with a diffusion scan was selected. This experiment was carried out on Achieva 3.0T Philips MRI system. An ExamCard with three following scans was selected (for information on ExamCards see Section 2.4.2):

- **DTI** (gradient duty cycle limited)

- **T2W** (non-limited)

- **Flair** (SAR limited, without any limitations on the gradient system)

The duration of the original examination was 11 min 57 s, including 7 min long **DTI** scan. The duration of one TR of the **DTI** scan was about 200 milliseconds (ms) with 67 ms of dead time due to the gradient temperature limitations (about one third of the TR). Gradient waveforms of the **DTI** scan are presented in Figure 5.4 (the time scale is one TR).
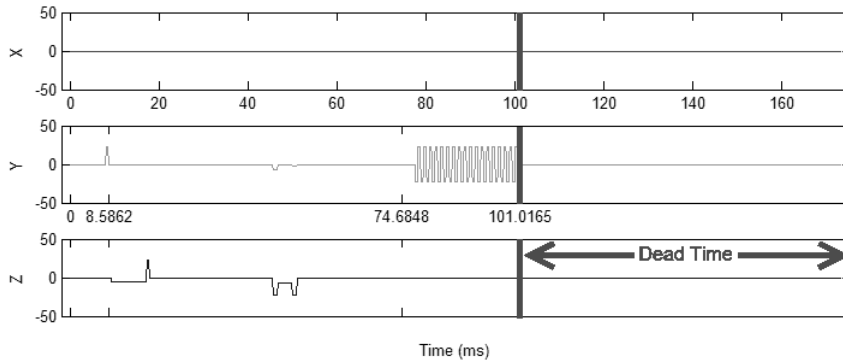
Figure 5.4: Gradient waveforms of the gradient-limited DTI protocol: the dead time is about 1/3 of the TR.

During the experiment, a custom ExamCard was created with the gradient-limited **DTI** scan being divided into three parts according to the output of the backtracking algorithm (Algorithm 5.1 from Section 5.2.1), and those three parts were intermixed with the rest of the scans, see Figure 5.5.

After being intermixed by the algorithm, the duration of the MRI examination was reduced to 9 min and 36 s. This was due to elimination of the dead time in the gradient-limited **DTI** scan. The relative time reduction achieved on the diffusion examinations was:

$$\frac{11\min 57\,\mathrm{s} - 9\min 36\,\mathrm{s}}{11\min 57\,\mathrm{s}} \cdot 100\% = 20\%. \tag{5.24}$$
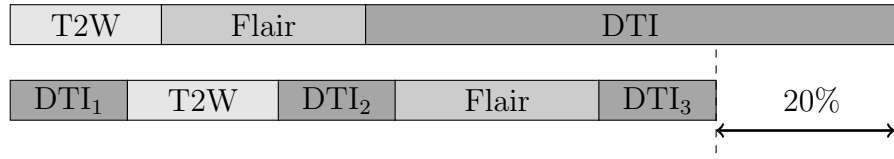
Figure 5.5: MRI experiments with the diffusion examination: the relative time reduction after intermixing is 20%.

A similar time reduction can be achieved on the majority of the brain diffusion examinations. The size of the slitted segments is relatively big (several minutes), consequently the effort to implement this intermixing algorithms in MRI software should be lower compared to the SAR case.

Chapter 6

Patient flow in MRI departments

In this chapter an approach to patient flow management in Magnetic Resonance Imaging (MRI) departments of hospitals is described. The approach consists of several steps. First, the patient flow is modeled by means of queuing theory in order to uncover bottlenecks. Then discrete-event computer simulations are performed to overcome limitations of the classical queuing theory assumptions. The resulting models are utilized to predict patient throughput and waiting times for various workflows and appointment scheduling strategies. Based on these detailed models, recommendations on improving MRI departments workflow are derived. The cost efficiencies are recognized through increased MRI scanners utilization, even with only modest workflow improvements. The approach can easily be extended to other Diagnostic Imaging departments (radiography, computed tomography, etc.). This way of working will help to reduce costs in public clinical imaging facilities. Most parts of this chapter were published in [Ivanov et al. 2010a].

6.1 Introduction

Over the last decade, demand on MRI examinations has sufficiently increased. Consequently, the number of MRI scanners per hospital increased. Moreover, free-standing imaging facilities were built, the only purpose of which is to provide the diagnostic imaging service. The cost of modern MRI scanners in Europe is several million euro, thus the question of investments return is important for

hospitals and facilities that utilize this diagnostic imaging equipment. However, due to inefficient capacity utilization, patient waiting lists are often several weeks long. This backlog decreases patient satisfaction and hospitals revenue.

Some research has been done recent years to reduce the scan time of MRI systems, e.g. [Blaimer et al. 2004; Ivanov et al. 2010b]. Hardware and software of MRI systems is constantly improving, and newer generations of MRI systems are able to perform scans faster and with higher resolution. However, scheduling the MRI patient's appointments is often performed manually by a receptionist at the front desk. This often results in inefficient MRI capacity planing: utilization of MRI scanners in many hospitals is less than 50%. In current practice of the hospitals, MRI examination appointment slot is about 30–45 min, whereas the scanning phase is only 15–20 min in average. The rest of the time is spent on pre- and post-scanning activities (e.g. patient preparation and aftercare), and on waiting for the next patient. As a result, the patient backlogs in hospitals are often more than two weeks long.

The research for this chapter was motivated by Philips Healthcare, The Netherlands. Recent years Philips has started to provide *utilization services* that help hospitals to extract the most out of the purchased equipment, e.g. to reduce the MRI exam duration.

In this chapter we describe an approach to patient flow management in MRI departments that is based on queuing theory supported by discrete-event computer simulation. One of the earliest applications of queuing theory to health care appointment management was in [Bailey 1954], where the queuing process in hospital outpatient departments was investigated. The theoretical results were promising, and many papers have followed, like [Gupta et al. 1971] and [Brahimi and Worthington 1991], a detailed survey of the literature can be found in [Brandeau et al. 2004; Fomundam and Herrmann 2007]. In those studies, the hospital appointment scheduling was addressed together with the capacity planing.

The queuing theory models are usually simple and provide general results on the patient flow. However, discrete-event simulations can help to describe more details about the system and to overcome the limitations of the classical queuing theory assumptions. A number of studies of patient flow and capacity planing in hospitals combine results of the queuing theory together with simulations, e.g. [Albin et al. 1990; Tucker et al. 1999]. An extensive survey of discrete-event simulations in health care can be found in [Jun et al. 1999].

In this chapter we model a patient flow in MRI department as a multiphase queuing system. There were several studies regarding operational research and queuing theory for the Radiology department, e.g. [Reinus et al. 2000; Rosenquist 1987; Tattoni et al. 2009]. The amount of papers considering specifically MRI department is very limited, see [Green et al. 2006]. None of the studies considered the phases of the MRI examination as a queuing system.

The outline of the chapter is as follows. In Section 6.2 the patient flow in MRI departments is described. In Section 6.3 the queuing models of the system and the discrete-event simulations are presented. In Section 6.4 the obtained results are discussed.

6.2 Patient flow in MRI departments

In order to perform an MRI examination a patient should proceed through different steps (i.e., phases).

First of all, the patient should schedule an appointment. On the examination day, once the patient has reached the MRI department, he/she will be accepted by a receptionist and asked to wait in the waiting area, see Figure 6.9. A technician will then show up and accompany the patient to one of the available changing rooms. There the patient should take off the outerwear and leave all magnetizable objects (e.g. keys, jewelry, credit cards, belt, etc.). If the patient has mobility problems he/she may require help of a technician inside the changing room. After leaving the changing room, the patient will proceed to the staging area to wait before a MRI room is available.
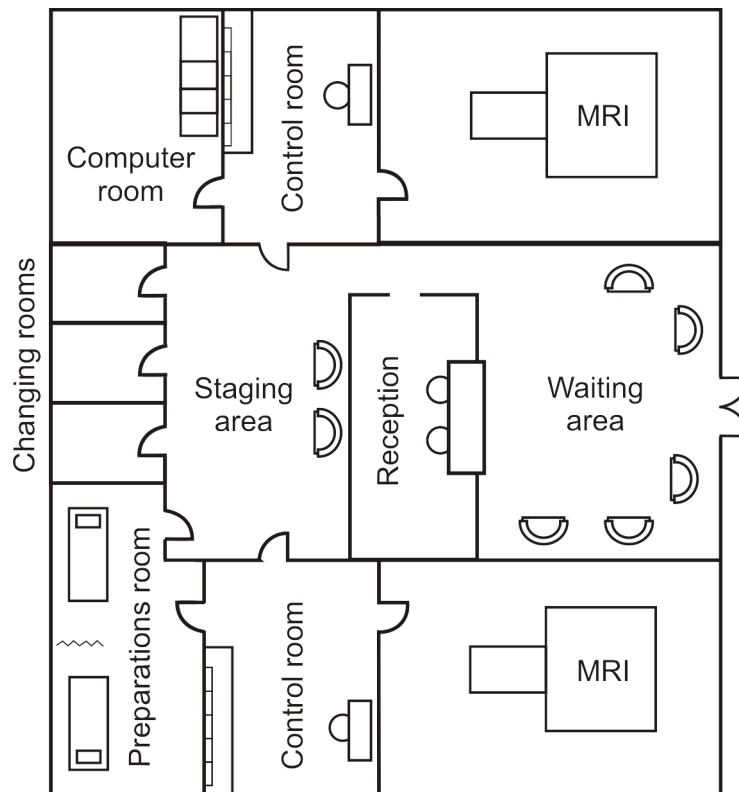
Figure 6.1: A typical MRI department plan.

Certain MRI examinations require a contrast agent injection to improve the image quality. In that case, special *contrast preparations* are required. The patient is accompanied from the staging area to a room where intravenous (IV) line preparations take place. (Usually these preparations take place in a dedicated room, however in some hospitals everything is done inside the magnet room; in our research we consider the dedicated room case.) During the contrast preparations a small IV cannula needle (catheter) is inserted into his/her vein in the hand or arm by the radiologist, technologist or a nurse, see Figure 6.2(a). The contrast agent will be injected later during the scanning process using tubing through the needle into the body, see Figure 6.2(b).

(a) IV cannula needle.   (b) Contrast agent injection.
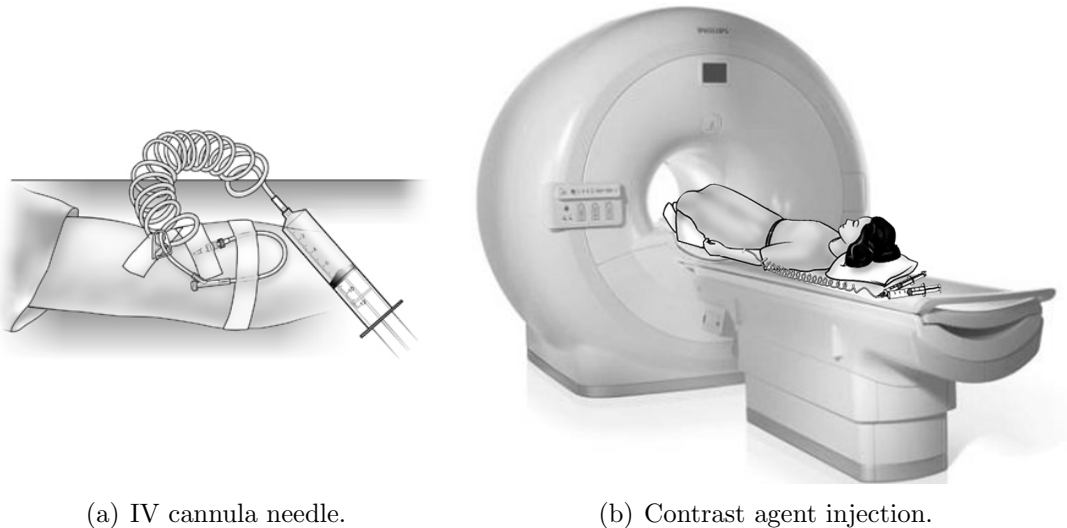
Figure 6.2: MRI examination with contrast agent: (a) IV line preparations can be done in advance, outside the magnet room; (b) contrast agent is injected during the MRI examination, usually during one of the last scans.

If the patient is allergic to the contrast agent, that step will be skipped and the patient will be directly called into the available MRI room.

Inside the MRI room some extra preparations are required. First, the patient lays down on the patient support table of the MRI scanner. Next, MRI coils are placed close to the scanning region to improve the image quality, see Figure 2.9 on page 19. Then, the patient support table is moved inside the bore of the scanner and the MRI process starts. After the scanning is over, the coils are taken off and the patient is then accompanied to his/her changing room. Finally, after getting dressed, the patient will leave the imaging facility.

Duration of each phase of MRI examination varies from patient to patient. Therefore, each of the phases can become a bottleneck and result in extra waiting time.

6.3 Queuing theory and discrete-event simulation

In this study two approaches are used in order to optimize the patient flow: queuing theory and discrete-event simulation. First, a multiphase queuing model of the MRI department is designed and analyzed to provide simple recommendations on capacity planing and appointments scheduling in the MRI departments. Then, the basic queuing models are enriched by details and assumptions that deviate from the classical queuing theory results. These more sophisticated models are analyzed by means of discrete-event simulation.

6.3.1 Queuing models

Queuing theory is a mathematical approach to study and analyze waiting lines. In health care the goal of queuing analysis is the cost minimization. According to [Ozcan 2005] the costs can be subdivided into two groups: *capacity costs* (e.g. cost of facilities, equipment and salaries of the employees) and *waiting costs*. Therefore queuing models in health care are used to achieve trade-off between health care service capacity and the costs of keeping patients waiting.

(a) Separate queue lines   (b) Single queue line

Figure 6.3: Patient assignment policies in MRI department.

Modern hospitals and free-standing imaging facilities often have several MRI scanners, though the substitutability of those scanners is not used by the receptionists during assignment of the patients. A queue can be formed as separate lines to each MRI scanner, see Figure 6.3(a), or it can be formed as a single line

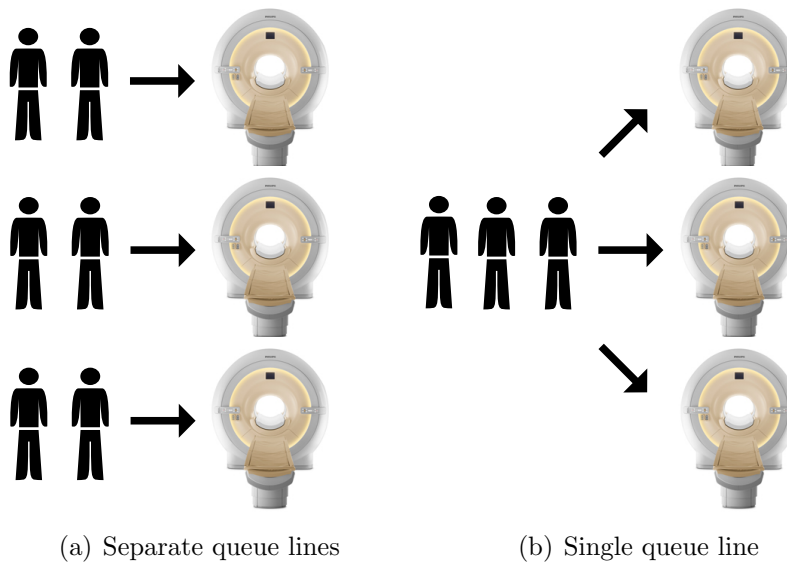to all the MRI scanners, see Figure 6.3(b). According to results of the classical queuing theory, see e.g. [Medhi 2002], in the separate lines patients often lose more time because of service variability.

In current practice of hospitals the patients are assigned to the MRI scanners in advance, therefore separate lines are formed. We propose to optimize the patient flow by using a common queue of patients to the MRI scanners.

Let us consider an example to numerically demonstrate the difference in the waiting times for the assignment policies.

Suppose that hospital has three MRI scanners. Patients arrive randomly to the MRI department of the hospital according to the Poisson process. Patient inter-arrival time is exponentially distributed with mean t_a of 7 min. The coefficient of variation (ratio of the standard deviation to the mean) for the interarrival time, c_a, equals to one. The average scan time, t_0, is characterized by a general distribution, e.g. by a symmetric triangular distribution with minimum $a = 12$ min, and maximum $b = 18$ min. The squared coefficient of variation, c_0^2, for a symmetric triangular distribution can be calculated as follows:

$$c_0^2 = \frac{1}{6} \left(\frac{b-a}{b+a} \right)^2 = \frac{1}{150}. \tag{6.1}$$

If a separate schedule is made for each scanner — 3 servers, each modeled as $M/G/1$, according to [Kendall 1953] nomenclature — then, assuming that a patient is randomly assigned (with equal probability) to one of the three scanners, the patients interarrival time to the scanner will be exponentially distributed with mean $7 \cdot 3 = 21$ min. The following equation [Medhi 2002] expresses the mean queuing time, φ_1, for a single scanner:

$$\varphi_1 \approx \frac{c_a^2 + c_0^2}{2} \cdot \frac{u}{1-u} \cdot t_0, \tag{6.2}$$

where, $u = t_0/t_a$, is the utilization of the scanner. According to equation (6.2), each patient will wait on average $\varphi_1 = 18.9$ min in a queue and the total flow time (queuing time plus examination time) is $\varphi_{\text{total},1} = \varphi_1 + t_0 = 18.9 + 15 = 33.9$ min on average.

Now suppose that an alternative assignment policy is applied: the patients wait for the first scanner available of the three, which is modeled as $M/G/3$. Equation (6.3) from [Hopp and Spearman 2001] express the mean queuing time, φ_2, for $m = 3$ identical parallel scanners:

$$\varphi_2 \approx \frac{c_a^2 + c_0^2}{2} \cdot \frac{u^{\sqrt{2(m+1)}-1}}{m(1-u)} \cdot t_0, \tag{6.3}$$

where, $u = t_0/(mt_a)$, is utilization. In this case the average queuing time is $\varphi_2 = 4.8$ min and the average total flow time is $\varphi_{\text{total},2} = \varphi_2 + t_0 = 19.8$ min.

Therefore, with the aforesaid assumptions, the average waiting time in the queue with the proposed assignment policy is 4 times less than the original one: 4.8 against 18.9. Note, that this time reduction can be gained without additional investments.

In Section 6.2 it was mentioned that MRI examination process consists of several phases. The duration of each phase varies from patient to patient. Therefore, each of these phases can become a bottleneck in the patient flow. Bottlenecks cause patients to wait and MRI examinations time to increase. More about the bottleneck analysis in health care can be found in [Langabeer 2007].

There are two well-known bottlenecks in the patient flow in MRI departments. Namely, the number of changing rooms and the number of MRI technicians. This results in unnecessary waiting time: even if the examination room is ready, a patient has to wait for a technician to come or for a changing room to free up.

In this chapter the assumption is made that the number of MRI technicians is sufficient, therefore only capacity limitations are considered.

A queuing model of the patient flow in MRI department is presented in Figure 6.4. The patient flow is represented as a queuing network that consists of several servers and buffers. Two waiting rooms, three MRI scanners, and an optional intermediate phase of contrast preparation are modeled as servers; whereas the waiting and staging areas are modeled as buffers. This queuing system is reentrant, because each patient should visit the changing room twice: before and after the MRI scanner (see Section 6.2).
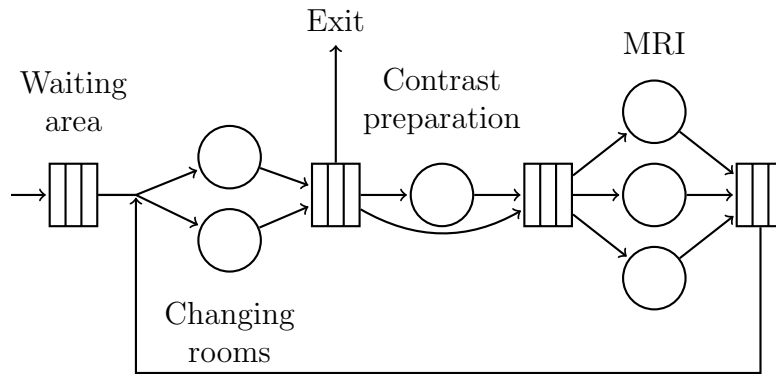
Figure 6.4: Queuing model of patient flow in MRI departments.

The queuing model of the patient flow (see Figure 6.4) can be analyzed to find the causes of variability in the system. There are two main causes of variability in queuing systems: variability in arrival pattern, and variability in service rates. The variability tends to accumulate throughout the network, even if the patients originally arrive according to predefined appointments, see [Kuehn 1979]. For instance, the variations in clothes changing and contrast preparation phases increase the variability in the arrival time of the patients to the staging area.

Lateness, earliness and no-shows of the patients will increase the variability even further.

Considering the variability factors mentioned above, the MRI departments of the hospitals can benefit from the proposed patient assignment policy.

6.3.2   Discrete-event simulation

There are several limitations of the classical queuing theory that can be overcome with simulation.

A major assumption of the theory is that patient interarrival time is exponentially distributed. This is satisfied for imaging facilities that work on walk-in basis, e.g. many facilities in China. For MRI departments in European and American hospitals where patients arrive according to predefined appointments, this assumption is not satisfied.

Classical queuing theory fails to model that a changing room only becomes available when a patient has re-entered and left it for the second time; during the scan time the patient's belongings are still in the changing room, making it unavailable for other patients. We call this a *blocking behavior*. Closed or semi-open Jackson's networks approach, see [Chen and Yao 2001], can be used to model a system with the blocking behavior to some extent. However, all the limitations mentioned above are fully applicable to this approach that significantly decreases the value of such a modeling.

It was thus decided to proceed with simulation. The discrete-event simulations were performed by means of χ process algebra, see [van Beek et al. 2006]. The purpose of the simulations was to derive answers to simple capacity planning questions, like:

How many changing rooms are required in order not to limit the patient throughput?

Simulation makes it possible to compare complex layouts of MRI department. For instance, MRI departments of some Japanese hospitals have only one changing room, see Figure 6.5. This room has multiple lockers inside, and several patient can leave their clothes there; though only one patient is allowed to be inside the changing room at the same time. The results of comparison of different changing room layouts are presented in Section 6.4.

Few words on our assumption for the patient arrival process. As stated previously, in most European and American hospitals the patients arrive according to predefined appointments. This arrival process was considered to be deterministic, thus patient lateness, earliness and no shows were not taken into account.

It was additionally assumed that only 40% of the patients require contrast material injection, whereas the other 60% just skip the contrast preparation phase
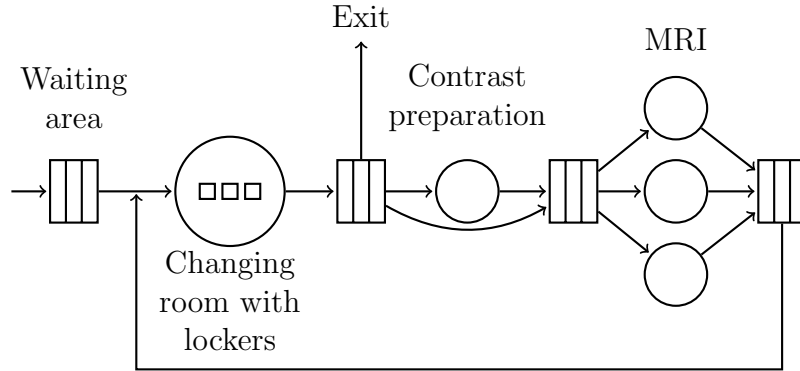
Figure 6.5: Model of MRI patient flow: one changing room with several lockers inside.

and immediately proceed to the staging area buffer.

6.4 Results and discussion

In this section the results of discrete-event simulations are presented. For details on implementation of the simulation models in χ process algebra, the reader is referred to [Pagoria 2011]. The question addressed is:

How different layouts of changing rooms affect the patient flow in MRI departments?

The data for the MRI processing, contrast preparation and changing clothes service times were collected from a hospital in The Netherlands and since the samples were small, symmetric triangle distributions were adopted. The rounded values of the distributions parameters together with the coefficients of variation in square, c_0^2, see equation (6.1), are presented in Table 6.1.

Service times	min	max	c_0^2
Changing room	1	3	0.042
Contrast preparation	2	4	0.019
MRI room	12	18	0.006

Table 6.1: Descriptive statistics of service times (in min).

In total, four different layouts were investigated: two for a department with 3 MRI scanners, and two for a department with 2 MRI scanners. For each number of MRI scanners the layout with multiple separate changing rooms (see Figure 6.4) was compared to the layout with a multi-locker changing room (see Figure 6.5.)

The maximal patient throughput, δ, was calculated for different number of changing rooms and lockers. Then, for a fixed interarrival rate the total flow times were estimated. Results of these simulations are presented in Figures 6.6 – 6.7 and Tables 6.2 – 6.3.

(a) Dedicated changing rooms.   (b) Single changing room with lockers.

Figure 6.6: Maximal throughput, δ, against number of (a) changing rooms or (b) lockers for systems with 3 MRI scanners.

From Figure 6.6(a) one can see, that for a system with three MRI scanners it is enough to have four separate changing rooms, in order not to limit the throughput. The added value of the 5$^{\text{th}}$ changing room is small, and of the 6$^{\text{th}}$ one is negligible.

From comparison of Figures 6.6(a) and 6.6(b) one can conclude that there is no significant difference in terms of patient throughput among the system with several changing rooms and the system with several lockers in a single changing room. Therefore, for a department with three MRI scanners it is enough to have only one changing room with four lockers inside. However, the total average flow time for the system with single room is up to 9% higher, see Table 6.2.

Total flow time	Amount	
	4	5
Lockers	23.01	22.49
Changing rooms	21.61	20.65
Difference in %	6%	9%

Table 6.2: Total average flow times for system with 3 MRI scanners and fixed interarrival time $t_a = 5.5$ min.

Similar conclusions can be drawn from Figures 6.7(a) and 6.7(b) for the systems with two MRI scanners. The required amount of separate changing rooms or

lockers is three. The added value of the 4th changing room or locker is negligible.

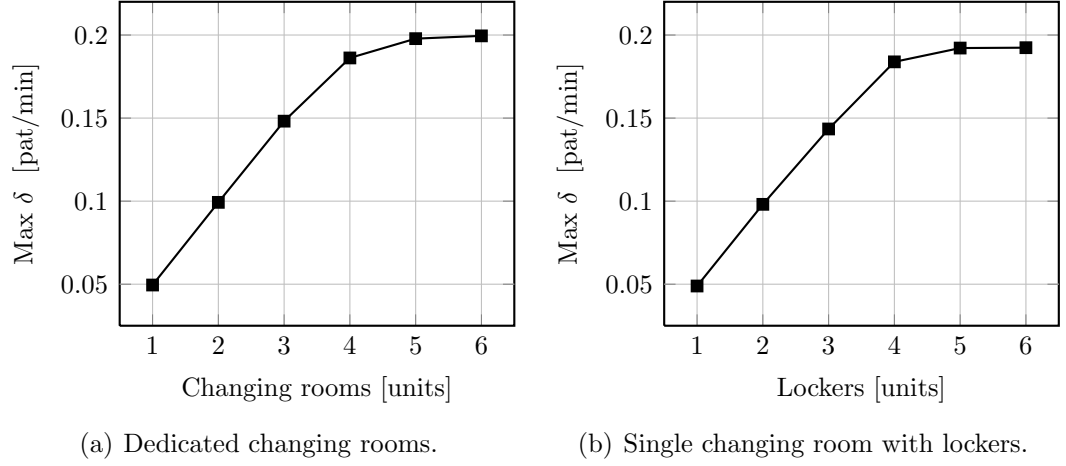(a) Dedicated changing rooms.   (b) Single changing room with lockers.
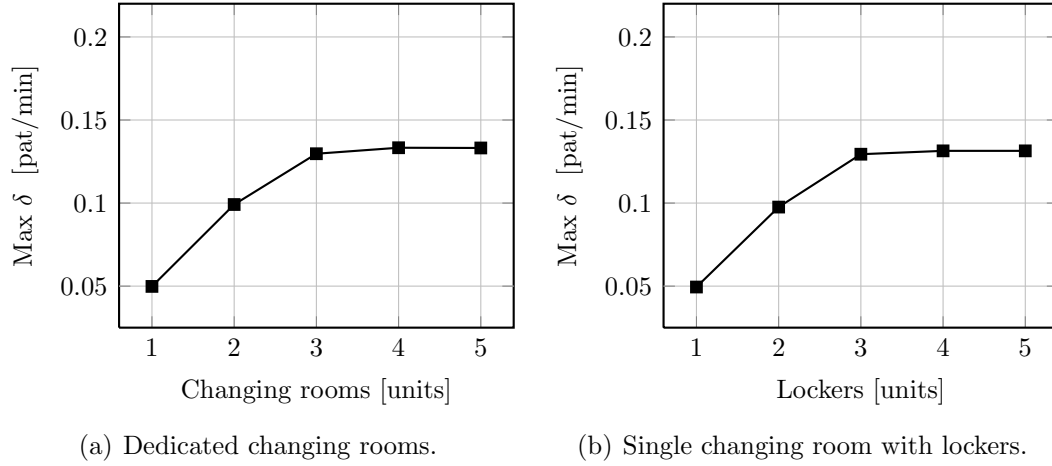
Figure 6.7: Maximal throughput, δ, against number of (a) changing rooms or (b) lockers for systems with 2 MRI scanners.

In Table 6.3 the average flow times for the systems with two MRI scanners are presented. For the compared layouts the difference in the flow times is a couple of percents. These results can be promising for hospitals that have space restric-

Total flow time	Amount
	3
Lockers	23.30
Changing rooms	22.84
Difference in %	2%

Table 6.3: Total average flow times for system with 2 MRI scanners and fixed interarrival time $t_a = 7.8$ min.

tions. The investment in extra lockers is smaller than construction of additional clothes changing rooms, whereas the difference in patient flow is negligible.

A simple 'rule of thumb' can be proposed for hospitals with a relatively small number of MRI systems: in order not to limit the throughput is enough to have one more clothes changing room (or locker) than the number of the MRI systems in the hospital.

An example of application of the approach was presented above. One of the capacity limitations of the MRI department, i.e., cloth changing rooms capacity, was considered in detail. The approach can be straightforwardly applied to other capacity limitations. For instance, some MRI scanners have detachable patient support tables. If a hospital purchases extra tables, the part of patient in-room

preparations (e.g. positioning of MRI coils) can be made in advance. This process can be easily described by queuing theory models and validated by discrete-event simulations. In a similar way the MRI staffing limitations can be modeled: either one or two MRI technicians are required in every phase of the workflow. However, the resulting models are more sophisticated, and only discrete-event simulation can give realistic estimations of the patient throughput and the waiting times in the system with the MRI staff.

6.5   Experiments

In Section 6.3.1 an assumption was made that MRI examination duration satisfies triangular distribution with certain parameters. This assumption was based on a simple estimation and can be far from the real situation in the hospitals. In order to obtain more accurate data on distribution of the MRI examination duration, a number of measurements of the patient flow was performed in a radiology department of one of the hospitals in The Netherlands.

A group of students from TU/e spent two full working days in the radiology department of St. Antonius Hospital in Nieuwegein. This radiology department has two MRI systems and four changing rooms. During these two days the students were working in cooperation with MRI technicians; they were observing the flow of patients in the radiology department and measuring the timings of the patients. Two days of data were collected for both MRI systems, see Appendix B.
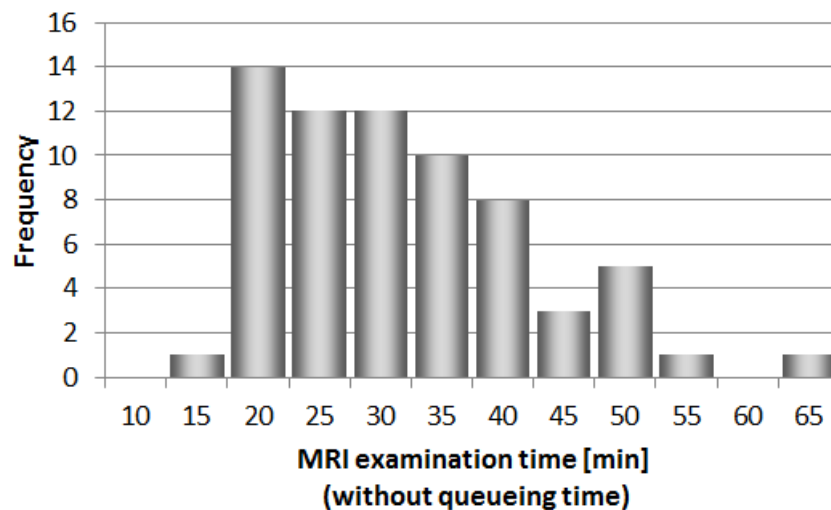

Figure 6.8: Distribution of MRI examination duration based on two days of data from St. Antonius hospital, The Netherlands.

Based on the collected data of 67 examinations, a distribution of MRI examination duration was calculated. In Figure 6.8 the histogram of this distribution is

displayed.

Similar statistics were collected for patient arrival times to the MRI department and the times that patients spent in the changing rooms prior to the MRI examinations, see [Pagoria 2011].

During this experiment, it was observed that patients in the hospital are assigned to each MRI system separately. This corresponds to assignment policy in Figure 6.3(a). And this assignment policy results in extra queueing time of the patients in the MRI department. Applying the queueing theory methods, it is possible to calculate the reduction of average patient queuing time if that hospital switches to the single-line assignment policy, see Figure 6.3(b). These calculations are done below.

Consider a queueing system with two MRI scanners and without limitation on the changing rooms. The MRI scanners are the bottlenecks in that system. Based on the collected statistics, the coefficient of variation of the patients interarrival time $c_a = 0.37$, whereas the coefficient of variation of the MRI examination time $c_0 = 0.33$. The average duration of MRI examination in that hospital $t_0 = 30$ min. Reduction in the average patient queuing time can be expressed by difference between equations (6.2) and (6.3) as follows:

$$\varphi_{\text{red}} = \varphi_1 - \varphi_2 = \frac{c_a^2 + c_0^2}{2} \cdot \frac{u}{1-u} \cdot t_0 \ - \ \frac{c_a^2 + c_0^2}{2} \cdot \frac{u^{\sqrt{2(m+1)}-1}}{m(1-u)} \cdot t_0 =$$

$$= 3.7 \left(\frac{u}{1-u} - \frac{u^{\sqrt{6}-1}}{2(1-u)} \right).$$

A plot of φ_{red} against utilization values higher than 0.5 is presented in Figure 6.9, where queuing time reduction ranges from 2 min up to 20 min for low and high utilization values of the MRI systems respectively.

6.6 Appointment slots planning

Another approach to reduce variability in the system and consequently patient queuing times is to plan patient appointment slots based on expected MRI examination time. Nowadays MRI hospitals use some rules to plan appointment slot durations when schedule MRI examinations, however these rules are quite general. This results in appointment slots far from actual durations of the MRI examinations. On the contrary, scheduling of non-realistic time slots can either cause extra queuing time or underutilization of MRI systems.

We proposed to plan the appointment slots based on distribution of actual MRI examination durations. To investigate this approach, 90 days of data of MRI scanners usage were collected from St. Antonius hospital and Philips Healthcare
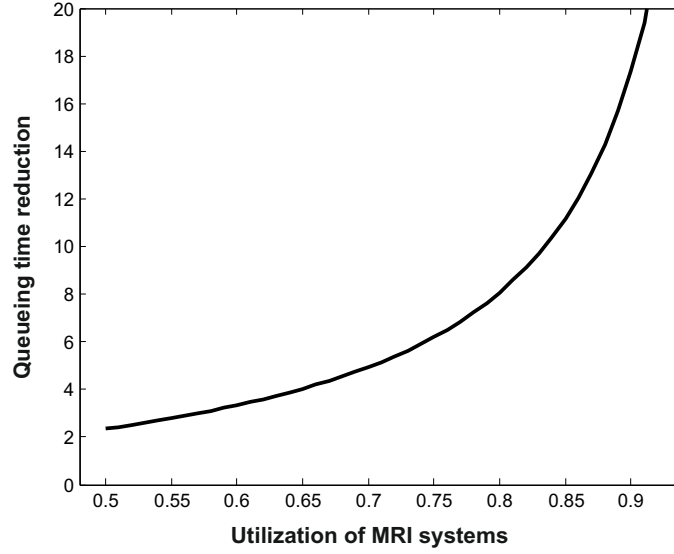
Figure 6.9: Estimated benefit of common queue in the radiology department with two MRI systems: reduction in average queuing time φ_{red} against utilization u of the systems.

data base. Afterwards, the examinations were stratified by the following characteristics: anatomical regions and ExamCards. This stratification resulted in multiple groups of MRI examinations, where each group had its own distribution of examination durations.

MRI scanners are used to obtain images of different anatomical regions of patient body, e.g. the brain, the breast, the shoulder. Although the individual scans can vary from one ExamCard to another, examinations of the same anatomical region often have similar sequence of scans and share a part of the scans.

Two strategies for patient appointment slots planning were compared by discrete-event simulations:

A. An appointment slot is planned based on distribution of examination durations of the **anatomical region** to be scanned.

B. An appointment slot is planned based on distribution of examination durations of the **ExamCard** to be used.

Both scheduling strategies are based on the same principle, which is described below.

At first, all the 90 days of the hospital data were processed and the examinations were divided into several groups.
For strategy A, the groups were based on anatomical regions. All the regions that made up less than 3% of the total population or had less than 20 measurements were grouped together in a group called *remainder*.
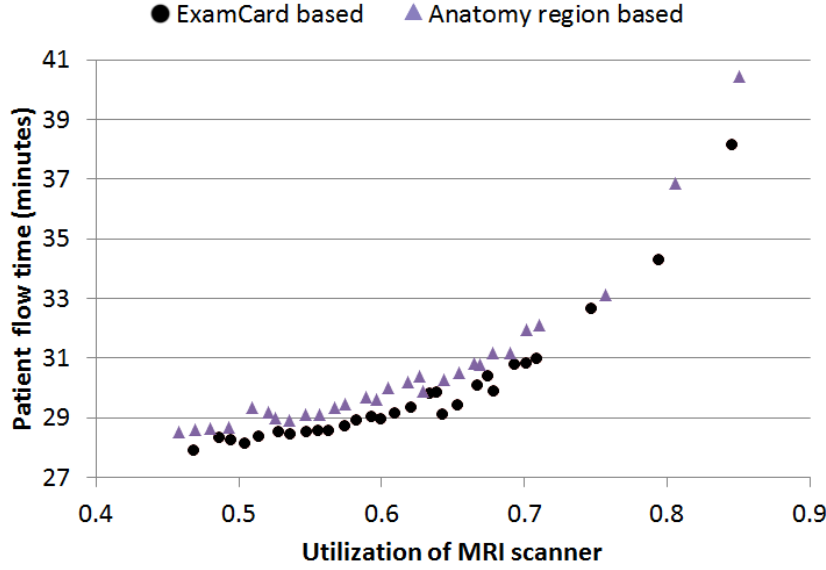
Figure 6.10: Average patient flow time versus utilization when comparing two appointment slots planning strategies: based on anatomical regions and based on ExamCards.

For strategy B, every single ExamCard that had enough data points was grouped separately. ExamCards that did not have enough measurements were grouped together based on anatomical regions if possible, otherwise grouped together in a *remainder* ExamCards group.

Afterwards, distributions of MRI examination durations[1] were calculated for every group. The observed distributions were fitted to log-normal distributions. These log-normal distributions (one distribution per group, and one for the *remainder* group) were then used to determine duration of patient appointment slots.

For each anatomical region (strategy A) or ExamCard (strategy B) the appointment slots are planned in such a way that cumulative distribution function $F(\cdot)$ of the corresponding log-normal distribution is equal to a fixed number p:

$$F(\text{slot duration}) = p. \tag{6.4}$$

The idea is to select a duration of the slot, for which the probability that the examination is finished before the end of the slot equals to p. (With an assumption that random variable of examination duration is described by that log-normal distribution). This number p is a design parameter for the schedule, which can be used to determine tradeoff between system utilization and average flow times of the patients. For instance, if $p < 0.5$ then there is a high probability that most

---

[1]Here, examination duration includes changeover time of the patients and room turnaround time, e.g. changing the coils between exams.

of the examinations will take more time than the reserved slots, which can result in long patient flow times; if p is too large (e.g. $p = 0.9$) than MRI scanners can be underutilized, which can result in increased backlog of patients in the hospital waiting list.

A discrete-event simulation model of the radiology department of St. Antonius hospital was used to compare both the scheduling strategies. This model included two MRI systems and four changing rooms: two rooms per each MRI system. A series of simulations was performed for parameter p ranging from 0.5 to 0.99. For all the simulations the same schedule of 10,000 patients was inputed. In the model the patients were assumed to arrive exactly according to the appointment time, therefore no variability was introduced by patient lateness or earliness. For details on realization of the simulation model in χ process algebra the reader is referred to [Berkien 2010].

Results of the simulations for one of the MRI systems are presented in Figure 6.10. From this figure one can see that ExamCards-based appointment planning strategy (strategy B) is more favorable for the average flow time than the anatomical region-based strategy (strategy A). These results were expected, since the slots based on ExamCards represent more precisely the actual durations of the examinations. It is important to notice that the difference in the flow time between the two strategies is less than 2 minutes even for high utilization numbers. This difference can be negligible when compared with the appointment slot duration.

We conclude that planning of patient appointment slots based on distributions of the examination durations of the anatomical regions (e.g. brain, breast, shoulder) is a fast and easy option that results in average patient flow time close to the average flow time when the appointment slots are based on distributions of examination durations of individual ExamCards.

6.7   Conclusion

In this chapter an approach for patient flow management in Magnetic Resonance Imaging (MRI) departments of hospitals was described. This approach combines queuing theory models with discrete-event simulations. The patient flow in MRI departments was modeled as a multi-phase queuing network. Two patient arrival processes were considered: Poisson arrivals and deterministic arrivals.
An alternative patient assignment policy was proposed that reduces queuing times. The simulation part of the approach was demonstrated on capacity limitation problems: several layouts of MRI department were considered in detail. The results are promising for hospitals that have space restrictions. The cost efficiencies are recognized through increased MRI scanners utilization, even with only modest workflow improvements. The approach can easily be extended to other diagnostic imaging departments (radiography, computed tomography, etc.).

Chapter 7

Conclusions and recommendations

This chapter summarizes and outlines the contributions made in this dissertation. Herein conclusions are drawn about the methods presented in the dissertation and suggestions for future research are given. Also benefits of the introduced methods and designed algorithms to the MRI industry are discussed.

7.1 Conclusions

In this dissertation, two approaches to improvement of productivity of MRI systems were investigated. In the first approach, only MRI duty cycles were considered, whereas the second one covered the complete period of time a patient spends in MRI department of a hospital.

7.1.1 MRI duty cycles optimization

First and main part of this thesis (from Chapter 3 to Chapter 5) was dedicated to elaboration of the approach to MRI duty cycles optimization. It was focused on MRI hardware and processes occurring in MRI systems during the scanning time. In this novel approach, the time reduction was accomplished by dividing parts of the MRI examination into relatively small segments that are then inter-

mixed in such a way that duty cycles of the system are optimized against various limitations.

In total, two duty cycle limitations were investigated, which are the main performance hitters for the MRI systems:

- SAR level,

- Gradient system hardware temperatures.

In Chapter 3, the problem of examination time reduction was posed as two separate scheduling problems: one scheduling problem for each MRI duty cycle limitation type. In Chapters 4 and 5 a number of scheduling algorithms was designed to deal with the duty cycle limitations and to improve performance of MRI systems.

The algorithms were verified on a large number of MRI examinations. According to collected statistics, duration of MRI examinations can be reduced by up to 22% for the SAR limitations and up to 25% for the gradient system limitations. As the result, the capacity of single MRI system can be increased by up to 4 patients per day. Algorithms for gradient system temperature limitation case were additionally validated by 'proof of concept' experiments.

Efficiency of each algorithm was calculated. There is at least one algorithm with relatively hight efficiency for each type of duty cycle limitation. This enables 'online' intermixing of the scan segments during MRI examinations.

Moreover, the approach developed in this dissertation can be combined with existing hardware-wise solutions to the MRI duty cycle optimization problem (e.g. modern ones: [van den Brink et al. 2003], [Poole et al. 2008], [Trakic et al. 2009], [Mueller et al. 2009]) because it does not influence the design of MRI hardware.

Detailed results for two types of duty cycle limitations are given below, followed by remarks on evolvability of Philips MRI systems.

SAR duty cycles

Three scheduling algorithms were designed to solve the problem with respect to SAR duty cycle limitations, and three different algorithmic methods were respectively applied. These methods are:

1. Dynamic programming,

2. Brute-force,

3. Greedy algorithm (heuristic).

Efficiency of these three algorithms was compared, and the best performing one is the greedy algorithm. Its complexity is polynomial with respect to the number of scans in examination and close to linear with respect to number of segments in the scan. Therefore this algorithm can be used for 'online' intermixing of scan segments: computing with a modern CPU will be almost instant. However the greedy algorithm is based on a heuristic, thus solutions can be imprecise. For high precision, other two exact algorithms can be used.

The greedy algorithm was applied to 57 typical Philips MRI examination protocols (so-called 'ExamCards'), however only 24 of them contained intermixable SAR limited scans. MRI examination time reduction was calculated for these 27 protocols. Achieved relative time reduction was up to 8% for examination protocols of 1.5T MRI system and up to 22% for protocols of 3.0T MRI system. For novel 7.0T MRI systems the time reduction is expected to increase.

Gradient system temperature

Two scheduling algorithms were designed to solve the duty cycle optimization problem with respect to temperature constrains on gradient chain system. These algorithms were base on the following methods respectively:

1. Backtracking (BT),

2. Periodic-solutions based (PS).

Both the algorithms, the BT and the PS, were shown to be suitable for intermixing the scan segments with respect to temperature constraints. The PS algorithm performs better in the worst-case scenario, whereas the BT algorithm is able to find more accurate solutions and the average performance is acceptable for a moderate number of scan segments.
Both the algorithms can be implemented in the MRI software for 'online' scheduling of the scan segments, however for the large-scale problems with a large number of scans and scan segments the computational time of the PS algorithm was shown to be lower.

Calculations of MRI examination time reduction were performed for a large set of routine brain examinations. Those calculations revealed that the most limited from the temperature point of view are *diffusion* brain examinations. Those diffusion examinations were also the most 'gainful': the time reduction achieved was up to 25%, whereas the average time reduction for brain examinations was about 10%. For the 'proof of concept' experiment an examination with a diffusion scan was selected. This experiment was carried out on Achieva 3.0T Philips MRI system. Registered time reduction was 20% of the original duration of the MRI examination.

The duration of the scan segments for gradient system temperature limitations is relatively long (several minutes), consequently the effort to implement these intermixing algorithms in MRI software should be lower compared to the SAR limitations case.

Evolvability of MRI systems

Algorithms designed for the gradient temperature limitations case can also help to improve evolvability of Philips MRI systems. These algorithms can compensate for minor differences in MRI hardware by intermixing the scans in the MRI examination based on thermal parameters of gradient amplifiers and coils.

Modification of the values in the MRI software that are related to the thermal parameters of coils and amplifiers can be replaced by preprocessing of ExamCards with these algorithms. Thereby the task of MRI software parameters modification can be simplified or even avoided. This has positive impact on evolvability of Philips MRI systems, and reduces workload on Phlilips application specialists.

7.1.2   Patient flow in MRI departments

The last part of this thesis (Chapter 6) presented an approach to patient flow modeling in MRI departments in hospitals. Several queueing and discrete-event models of the patient flow were designed, which can be used to predict patient flow for various layouts of MRI departments, patient appointment scheduling and assignment strategies.

For hospitals that have several MRI scanners, two patient assignment strategies were compared: separate queues to each scanner and common queue to all the MRI scanners. In current practice, most of the hospitals assign patients to separate queues. We recommended to use a common queue of the patients and every time assign the patient to the first available scanner. This strategy reduces average waiting time, thus beneficial for patient satisfaction.

Also capacity limitation problems were investigated: several layouts of MRI department were considered in detail and discrete-event simulations were performed. The results of the simulations are promising for hospitals with space restrictions, like majority of hospitals in Japan.

Methods and models used in this study can be combined with optimization of duty cycles of the MRI systems, in order to improve productivity of MRI departments in hospitals. First, the MRI scanning time is reduced, then the patient flow is optimized that yields the overall MRI examination time reduction.

7.2 Benefits to health care industry

Research in this dissertation was focused on productivity of MRI scanners. Importance of productivity is widely accepted in industries like production and manufacturing. When a customer is purchasing a machinery, he is usually concerned about its efficiency and, in particular, its throughput. However for medical equipment the importance of productivity is not tangible at first glance. This is because many customers for this equipment are non-profit hospitals, like hospitals in The Netherlands and majority of hospitals in America, see [Walker 2005]. Nevertheless non-profit hospitals are also concerned about return of investments.

Over the last decades, a new business model of 'free-standing imaging centers' has been developing. These are for-profit organizations offering diagnostic imaging service like MRI and X-ray. The free-standing imaging centers together with for-profit hospitals are fully concerned about high throughput and low cycle times of the diagnostic imaging systems they utilize.

The price of a modern MRI scanner is about \$2 million, therefore customers want to maximize utilization of such an expensive piece of medical equipment. For instance, in 2006 in Canada average operation time of MRI systems was 71 hours per week. This utilization is much higher than for other types of medical diagnostic equipment including X-ray computed tomography (CT). However due to the long duration of MRI examinations, the average number of MRI examinations per machine per year is 41% lower than for CT scanners (all statistics data is from [Health Information 2007]). Consequently productivity of MRI scanners is a concern for the customers of this medical equipment.

The benefit of higher productivity of MRI tools is twofold:

- Higher system throughput increases profitability and reduces time return of investments in MRI equipment.

- Faster systems increases overall quality of service: both examination and queueing time of patients is reduced.

These bullets are important both for for-profit and non-profit medical organizations.

For-profit free-standing imaging centers and hospitals are definitely interested in higher throughput of their MRI systems. In Section 7.1.1 it was stated that our methods can increase throughput of MRI systems by up to 4 patients per day. An average cost of MRI examination in the USA is around \$2500[1]. Simple calculations demonstrate that every single additional patient per day throughout the year results in $\$2500 \times 365 \simeq \$1$ million extra revenue annually for these centers in the USA. On the other hand, the utilization of the systems in these

---

[1]data from http://www.comparemricost.com/

centers is often kept below critical level to enable fast processing of new patients. This utilization level is maintained for higher quality of service. With higher machine speed, the queueing time also goes down and quality of service improves.

For **non-profit** hospitals the situation is different, but the need in productivity is also high. For instance, in case of emergency the duration of MRI examination is crucial and can be even vital. Also for claustrophobic patients the time they spend in the bore of the MRI scanner should be maximally reduced to avoid panic attacks. But even for regular patients long waiting and examination times are exhausting, and result in undesirable stress both for the patient and for MRI technicians.

We conclude that the implementation of the methods developed in this dissertation will reduce return of investments time, increase quality of service together with patient satisfaction and reduce stress on employees of MRI departments.

7.3   Recommendations for further research

Different MRI performance limitations

In this dissertation, three major MRI examination performance hitters were investigated: SAR level, temperature of gradient amplifiers and temperature of gradient coils. However there are more factors that limit performance of MRI systems, like acoustic noise level during MRI examination and peripheral nerve stimulation of patients in MRI systems. Latter two are the patient comfort and safety issues.

The limitation on acoustic level is as follows. During the MRI examination the pulse sequences are executed and the gradients are rapidly switched. Therefore, eddy currents are induced in nearby conducting components, such as magnet cryostat and other coils. Large forces act on the mechanical parts of MRI system so that these start oscillating during gradient switching. This results in a 'clanging' noise on which patients often complain to technicians. The phenomenon is pretty similar to the one that drives loud speakers. If no actions are taken, the nose level can exceed comfortable threshold.

The fast switching of magnetic gradients generates electric field that can lead to peripheral nerve stimulation (PNS) of a patient in the MRI systems. The threshold depends on the duration of the stimulation. The IEC prescribes limitations on the gradient switching to avoid harmful effects on the patient health in the same way like it does for the SAR level, see [CEI/IEC 60601-2-33 2008]. Several PNS prediction models are available in literature, see e.g. [Ham et al. 1997].

Currently gradient waveforms are preprocessed by algorithms to reduce the noise to a comfortable level, and to avoid peripheral nerve stimulation of patients. Re-

sulting waveforms are often longer than the original ones, thus MRI examination time is extended.

It would be interesting to investigate if some intelligent intermixing of the scan segments can reduce the gradient noise and/or prevent PNS without prolongation of the MRI examination duration.

Effective Processing Time

In this dissertation, patient flow in MRI department was investigated using conservative models from Queueing theory and as a discrete-even system. However the Effective Processing Time (EPT) method can be more convenient to model the flow of the patients, since it requires much less measurements to understand the system behavior, see [Jacobs et al. 2003]. The EPT has also an advantage of including all sources of variability and capacity losses due to various outages like machine breakdown and setup times in single distribution of machine effective processing time.

It would be interesting to combine the EPT method together with automatic data collection systems (databases etc.) in hospitals to crunch all the data and generate effective processing time distributions for various layouts of radiology departments. Based on the collected data, sophisticated recommendations on improvement of the patient flow can be provided.

Appendix A

List of ExamCards

This appendix contains lists of ExamCards that were used for the SAR reduction calculations that are described in Chapter 4. The ExamCards for Achieva 1.5T MRI system are listed in Table A.1, whereas the ExamCards for the Achieva 3T MRI scanner are listed in Table A.2.

#	ExamCard	# SAR limited sc.	# non-SAR limited sc.	clinical area (anatomy)
1	Fast Cervical Spine	0	5	neuro
2	C Vascular rings	0	8	cardiac
3	C Coarctation	0	11	cardiac
4	C ASD and PAPVR	0	11	cardiac
5	DWIBS R1 2	0	1	oncology, body
6	Total Neuro	0	8	neuro
7	Spine Diffusion	0	5	neuro
8	Fast Lumbar Spine	1	4	neuro
9	Carotiden Cologne	1 (survey)	4	vascular
10	T2 Star Myocardium	0	2	cardiac
11	Cardiac InterActive	0	3	cardiac
12	Basic Cardiac Function	0	7	cardiac
13	SLEH ARVD ExamCard-V4	0	7	cardiac
14	Q-FLOW OF SMA SMV	3	4	vascular
15	Abdomen FFE BH HR	0	3	oncology, body
16	MENINGITIS BRAIN	2	4	pediatric, neuro
17	SENSE Cervical Spine	0	6	neuro
18	Shoulder ArthroMR	0	10	musculoskeletal
19	BRACHIAL PLEXUS	0	6	neuro
20	PIRIFORMIS	0	7	neuro
21	Liver-Dynamic	0	13	body
22	PROSTATE	0	11	oncology, body
23	VASOVIST PERIPHERAL	0	8	vascular
24	Brain singleshot	0	5	neuro
25	ABDO PELV16	4	10	oncology, body
26	ABDOMEN16	3	9	oncology, body
27	Fast spectroscopic imaging 15T	0	3	neuro
28	NASA	1	3	vascular
29	1.5TFoot Ankle	0	9	musculoskeletal
30	PELVIC FLOOR LAXITY	2	5	body
31	TRANCE R11 1 3	0	5	vascular

Table A.1: Summary of used ExamCards for Achieva 1.5 T MRI scanner.

#	ExamCard	# SAR limited sc.	# non-SAR limited sc.	clinical area (anatomy)
32	Fast Knee Manchester	2	3	musculoskeletal
33	Liver Dynamic	3	4	body
34	brain tumor Manchester	1	7	neuro
35	Cervical Spine HNP	2	3	neuro
36	Braintumor protocol Bonn	1	9	neuro
37	Whole Brain CE-MRA	0	4	vascular
38	3T ShoulderExamCard	3 (1x survey)	3	musculoskeletal
39	3T orbits Manchester	0	6	neuro
40	Angio Aorta	1	4	vascular
41	YONSEI DTI	0	6	neuro
42	BBA MOBI FLEX dualinj	2	5	vascular
43	30T NV8 Carotids SF2	0	5	vascular
44	Brain single shot	0	6	neuro
45	T30 Pelvis Imaging	2	2	oncology, body
46	Knee 3D iso	1 (survey)	2	musculoskeletal
47	HEMANGIOMA	3	6	body
48	Knee	4	3	musculoskeletal
49	MRCP	3	3	body
50	BILAT BREAST	1	7	body
51	3.0T Fast Knee	2 (1x survey)	5	musculoskeletal
52	FAST STROKE	2	4	neuro
53	Fast spectroscopic im. 30T	0	4	neuro
54	3D MOTSA COW	0	3	vascular
55	ROUTINE KNEE	1 (survey)	7	musculoskeletal
56	OHSU PEDS CARDIAC	2	8	cardiac, pediatric
57	3.0TFootAnkle	2	6	musculoskeletal

Table A.2: Summary of used ExamCards for Achieva 3 T MRI scanner.

<div align="right">

Appendix B

</div>

<div align="right">

Hospital data

</div>

This appendix contains data collected in radiology department of St. Antonius Hospital in Nieuwegein, The Netherlands. This radiology department has two MRI systems and four changing rooms.

Two days of data were collected for both MRI systems: the start and finish time of MRI examination for each patient. These data are presented in Figures B.1 and B.2.

MRI 1		
patient	exam start	exam end
1	8:20:50	8:39:18
2	8:40:10	9:19:20
3	9:19:56	9:49:15
4	9:51:35	10:10:55
5	10:12:25	10:49:35
6	10:50:00	11:12:30
7	11:14:30	12:03:45
8	12:08:30	12:28:37
9	12:29:00	13:05:00
10	13:15:30	13:49:55
11	13:53:15	14:29:15
12	14:31:50	15:11:55
13	15:13:50	15:33:00
14	15:33:10	15:59:30
15	15:59:50	16:30:40
16	16:30:45	16:50:25
17	16:54:50	17:17:00

MRI 2		
patient	exam start	exam end
1	8:14:00	9:07:44
2	9:09:00	9:27:20
3	9:27:50	9:50:20
4	9:51:00	10:19:15
5	10:19:40	10:39:00
6	10:40:30	11:21:40
7	11:24:20	12:02:00
8	12:03:00	12:49:30
9	12:50:50	13:21:50
10	13:22:10	13:59:00
11	14:00:00	14:33:02
12	14:33:30	14:57:40
13	14:58:30	15:13:30
14	15:15:20	15:51:30
15	15:52:00	16:39:30
16	16:43:00	17:03:00
17	17:07:30	17:30:50
18	17:32:30	17:58:00

Figure B.1: Start and finish times of MRI examinations, 1st day in the St. Antonius hospital.

MRI 1		
patient	exam start	exam end
1	8:17:38	8:40:25
2	8:42:50	9:17:10
3	9:18:08	9:35:45
4	9:36:05	9:52:17
5	9:53:35	10:23:20
6	10:24:33	10:46:00
7	10:51:15	11:39:20
8	11:42:13	12:16:33
9	12:17:25	12:52:07
10	12:53:00	13:21:20
11	13:23:08	13:41:01
12	13:43:41	14:11:59
13	14:15:02	14:40:37
14	14:42:46	15:08:58
15	15:09:03	15:31:08
16	15:31:15	15:49:48
17	15:50:23	16:05:40
18	16:08:16	16:34:30

MRI 2		
patient	exam start	exam end
1	8:13:00	8:56:22
2	8:56:57	9:43:40
3	9:46:35	10:48:40
4	10:49:25	11:21:57
5	11:23:50	11:43:20
6	11:46:00	12:21:00
7	12:23:18	12:56:55
8	13:17:50	13:35:10
9	13:40:35	14:08:50
10	14:09:30	14:33:13
11	14:33:54	14:47:45
12	14:48:00	15:15:40
13	15:18:06	15:53:27
14	16:02:20	16:23:30

Figure B.2: Start and finish times of MRI examinations, 2nd day in the St. Antonius hospital.

Bibliography

ACGIH. Threshold limit values for chemical substances and physical agents and biological exposure indices. In *American Conference of Governmental Industrial Hygienist*, 1996.

S. L. Albin, J. Barrett, D. Ito, and J. E. Mueller. A queueing network analysis of a health center. *Queueing Systems*, 7(1):51–61, 1990.

A. Allahverdi, J. N. D. Gupta, and T. Aldowaisan. A review of scheduling research involving setup considerations. *Omega*, 27:219–239, 1999.

A. Allahverdi, C. T. Ng, T. C. E. Cheng, and M. Y. Kovalyov. A survey of scheduling problems with setup times or costs. *European Jornal of Operational Research*, 187:985–1032, 2008.

N. T. J. Bailey. A study of queues and appointment systems in hospital out-patient departments, with special reference to waiting times. *Journal of the Royal Statistical Society*, 14:185–199, 1954.

B. Berkien. Patient flow modeling of radiology departments of hospitals. Bachelor's thesis, Eindhoven University of Technology, Department of Mechanical Engineering, Systems Engineering Group, 2010.

M. A. Bernstein, F. K. King, and J. X. Zhou, editors. *Handbook of MRI Pulse Sequences*. Elsevier, 2004.

L. Bianco, A. Mingozzi, and S. Ricciardelli. The travelling salesman problem with cumulative costs. *Networks*, 23:81–91, 1993.

M. Blaimer, F. Breuer, M. Mueller, R. M. Heidemann, M. A. Griswold, and P. M. Jakob. SMASH, SENSE, PILS, GRAPPA: how to choose the optimal method. *Topics in magnetic resonance imaging : TMRI*, 15(4):223–236, August 2004. ISSN 0899-3459. URL `http://view.ncbi.nlm.nih.gov/pubmed/15548953`.

M. Brahimi and D. J. Worthington. Queueing models for out-patient appointment systems — a case study. *Journal of the Operational Research Society*, 42 (9):733–746, 1991.

M. L. Brandeau, F. Sainfort, and W. P. Pierskella. *Operations Research and Health Care: A Handbook of Methods and Applications*. Kluwer's International Series, 2004.

D. Brooks and M. Martonosi. Dynamic thermal management for high-performance microprocessors. In *High-Performance Computer Architecture, 2001. HPCA. The Seventh International Symposium on*, pages 171–182, 2001. doi: 10.1109/HPCA.2001.903261.

H. Canada, editor. *Healthy Canadians a Federal Report on Comparable Health Indicators*. Health Canada, 2008.

CEI/IEC 60601-2-33. *Medical Electrical Equipment – Part 2-33: Particular requirements for basic safety and essential performance of magnetic resonance equipment for medical diagnosis*. International Electrotechnical Commission, Geneva, Switzerland, 2.2 edition, Apr. 2008.

Y. A. Cengel. *Heat Transfer: A Practical Approach*. McGraw-Hill Companies, 2nd edition, 2002. ISBN 0072458933.

H. Chen and D. D. Yao. *Fundamentals of Queueing Networks*. Springer, first edition, 2001.

T. H. Cormen, C. E. Leiserson, and R. L. Rivest. *Introduction to Algorithms*. The MIT Press and McGraw-Hill Book Company, 1989. ISBN 0-262-03141-8, 0-07-013143-0.

ESI. ESI research agenda on embedded systems engineering. Technical report, Embedded Systmems Institute, 2006.

M. Fischetti, G. Laporte, and S. Martello. The delivery man problem and cumulative matroids, operations research. *Operations Research*, 41:1055–1064, 1993.

S. Fomundam and J. W. Herrmann. A survey of queuing theory applications in healthcare. Technical report, The Institute for Systems Research, 2007.

N. D. Gai and Y. Zur. Design and optimization for variable rate selective excitation using an analytic RF scaling function. *Journal of Magnetic Resonance*, 189(1):78–89, 2007. ISSN 1090-7807. doi: 10.1016/j.jmr.2007.08.017.

R. Graham, E. Lawler, J. Lenstra, and A. Kan. Optimization and approximation in deterministic sequencing and scheduling: a survey. *Annals of Discrete Mathematics*, 5:287–326, 1979. doi: 10.1016/S0167-5060(08)70356-X.

L. V. Green, S. Savin, and B. Wang. Managing patient service in a diagnostic medical facility. *Operations Research*, 54(1):11–25, 2006. doi: 10.1287/opre. 1060.0242.

S. H. Greenblatt, Dagi, and M. H. Epstein. *A History of Neurosurgery: In its Scientific and Professional Contexts*. Thieme/AANS, 1 edition, 1997. ISBN 1879284170.

I. Gupta, J. Zoreda, and N. Kramer. Hospital manpower planning by use of queueing theory. *Health Services Research*, 6, 1971.

A. C. Guyton. *Textbook of Medical Physiology*. Elsevier Health Sciences, 8 edition, 1991.

C. L. G. Ham, J. M. L. Engels, G. T. van de Wiel, and A. Machielsen. Peripheral nerve stimulation during MRI: Effects of high gradient amplitudes and switching rates. *J. Magn. Reson. Imaging*, 7(5):933–937, 1997. ISSN 1522-2586. doi: 10.1002/jmri.1880070524.

B. A. Hargreaves, C. H. Cunningham, D. G. Nishimura, and S. M. Conolly. Variable-rate selective excitation for rapid mri sequences. *Magn. Reson. Med.*, 52(3):590–597, 2004. doi: 10.1002/mrm.20168.

T. Hauck and T. Bohm. Thermal RC-network approach to analyze multi-chip power packages. In *Sixteenth Annual IEEE Semiconductor Thermal Measurement and Management Symposium*, pages 227 –234, 2000. doi: 10.1109/STHERM.2000.837088.

Health Information. *Medical Imaging in Canada*. Canadian Institute for Health Information, 2007.

W. J. Hopp and M. L. Spearman. *Factory Physics*. McGraw-Hill, Boston, 2 edition, 2001.

ICNIRP. Guidelines for limiting exposure to time-varying electric, magnetic, and electromagnetic fields (up to 300 GHz). *Health Physics*, 74(4):494–522, 1998. International Commission on Non-Ionizing Radiation Protection.

E. N. Ivanov, A. Y. Pogromsky, and J. E. Rooda. Scheduling with sequence dependent setup times in application to magnetic resonance imaging scans processing. In *Proc. IEEE Control Applications, (CCA). Intelligent Control, (ISIC)*, pages 867–872, July 8–10, 2009a. doi: 10.1109/CCA.2009.5280713.

E. N. Ivanov, A. Y. Pogromsky, and J. E. Rooda. Scheduling with dynamic constraints in application to magnetic resonance imaging scans processing. In *Proc. IEEE Control Applications, (CCA). Intelligent Control, (ISIC)*, pages 873–877, July 8–10, 2009b. doi: 10.1109/CCA.2009.5280980.

E. N. Ivanov, J. E. Rooda, A. Y. Pogromsky, and J. S. Brink. Optimization of duty cycle of magnetic resonance imaging scanners. In *IEEE conference IECON*, 2009c.

E. N. Ivanov, G. Pagoria, A. Y. Pogromsky, and J. E. Rooda. Patient flow management of MRI departments in hospitals. In *Proceedings of Quaderni della XIV Summer School Francesco Turco*, Porto Giardino, Monopoli, Italy, 2010a.

E. N. Ivanov, A. Y. Pogromsky, J. S. Brink, and J. E. Rooda. Optimization of duty cycles for MRI scanners. *Concepts in Magnetic Resonance Part B: Magnetic Resonance Engineering*, 37B(3):180–192, 2010b. ISSN 1552-504X. doi: 10.1002/cmr.b.20167.

E. N. Ivanov, A. Y. Pogromsky, J. S. Brink, and J. E. Rooda. *Scheduling in MRI scans processing*, chapter 14, pages 227–243. Embedded Systems. Springer, October 2010c.

J. H. Jacobs, L. F. P. Etman, E. J. J. van Campen, and J. E. Rooda. Characterization of operational time variability using effective process times. *IEEE Transactions on Semiconductor Manufacturing*, 16(3):511–520, Aug. 2003. ISSN 0894-6507. doi: 10.1109/TSM.2003.815215.

R. Jayaseelan and T. Mitra. Temperature aware task sequencing and voltage scaling. In *Computer-Aided Design, 2008. ICCAD 2008. IEEE/ACM International Conference on*, pages 618–623, 2008. doi: 10.1109/ICCAD.2008.4681641.

R. Jayaseelan and T. Mitra. Temperature aware scheduling for embedded processors. In *VLSI Design, 2009 22nd International Conference on*, pages 541–546, 2009. doi: 10.1109/VLSI.Design.2009.42.

J. Jin. *Electromagnetic analysis and design in magnetic resonance imaging*. London : CRC Press, 1999.

J. B. Jun, S. H. Jacobson, and J. R. Swisher. Application of discrete-event simulation in health care clinics: A survey. *The Journal of the Operational Research Society*, 50(2):109–123, 1999.

D. G. Kendall. Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded markov chain. *Annals of Mathematical Statistics*, 24(3):338–354, 1953.

D. E. Knuth. *The Art of Computer Programming*, volume 3. Addison-Wesley, third edition, 1997.

P. Kuehn. Approximate analysis of general queuing networks by decomposition. *Communications, IEEE Transactions on*, 27(1):113–126, 1979.

J. R. Langabeer. *Health Care Operations Management: A Quantitative Approach to Business and Logistics*. Jones & Bartlett Publishers, first edition, 2007.

S. Li, B. Cao, D. Bi, and X. Jiang. Stacked high/low voltage level H-bridge circuit for gradient amplifier of mri system. In *Proceedings of ICEMS2008 The 11th International Conference on Electrical Machines and Systems October 17-20, 2008 Wuhan, China*, pages 2154 –2158, oct. 2008.

R. L. Magin, R. P. Liburdy, and B. Persson. Biological effects and safety aspects of nuclear magnetic resonance imaging and spectroscopy. In *Ann.*, number 5 in 649, page 402. New York Academy of Sciences, N.Y., 1992.

MATLAB. *version 7.5.0.342 (R2007b)*. The MathWorks Inc., Natick, Massachusetts, 2007.

D. W. Mcrobbie, E. A. Moore, M. J. Graves, and M. R. Prince. *MRI from Picture to Proton*. Cambridge University Press, 2 edition, March 2007.

J. Medhi. *Stochastic Models in Queueing Theory*. Academic Press, 2 edition, 2002.

C. L. Monma and C. N. Potts. On the complexity of scheduling with batch setup times. *Operations Research*, 37(5):798–804, 1989. ISSN 0030364X.

M. F. Mueller, M. Blaimer, F. Breuer, T. Lanz, A. Webb, M. Griswold, and P. Jakob. Double spiral array coil design for enhanced 3D parallel MRI at 1.5 Tesla. *Concepts in Magnetic Resonance Part B: Magnetic Resonance Engineering*, 35B(2):67–79, April 2009. ISSN 15525031. doi: 10.1002/cmr.b.20137.

NetForum. Philips Healthcare, 2004–2010. URL `http://netforum.medical.philips.com/`. Philips NetForum Community.

Y. A. Ozcan. *Quantitative Methods in Health Care Management: Techniques and Applications*. Jossey-Bass, first edition, 2005.

M. Ozden, P. J. Egbelu, and A. V. Iyer. Job scheduling in a group technology environment for a single facility. *Computers & Industrial Engineering*, 9(1): 67–72, 1985. doi: 10.1016/0360-8352(85)90037-3.

G. Pagoria. Analysis and performance improvement of mri departments. Master's thesis, Eindhoven University of Technology, Department of Mechanical Engineering, Systems Engineering Group, 2011.

G. Peeren. Personal communication, 2009.

Philips. Basic principles of MR imaging. NL, 2005. Philips Medical Systems.

M. L. Pinedo. *Scheduling: Theory, Algorithms, and Systems.* Springer, Berlin, third edition, 2008.

M. Poole, H. S. Lopez, and S. Crozier. Adaptively regularized gradient coils for reduced local heating. *Concepts in Magnetic Resonance Part B: Magnetic Resonance Engineering*, 33B(4):220–227, 2008. ISSN 1552-504X. doi: 10.1002/cmr.b.20125. School of Information Technology and Electrical Engineering, University of Queensland, St. Lucia, Brisbane QLD 4072, Australia.

C. N. Potts and M. Y. Kovalyov. Scheduling with batching: A review. *European Journal of Operational Research*, 120(2):228–249, 2000. doi: 10.1016/S0377-2217(99)00153-8.

H. N. Psaraftis. A dynamic programming approach for sequencing groups of identical jobs. *Operations Research*, 28(6):1347–1359, 1980.

R. Rao, S. Vrudhula, and C. Chakrabarti. An optimal analytical solution for processor speed control with thermal constraints. In *in Proc. Intl' Symp. Low Power Electronics and Design (ISLPED)*, pages 292–297, 2006.

W. R. Reinus, A. Enyan, P. Flanagan, B. Pim, D. S. Sallee, and J. Segrist. A proposed scheduling model to improve use of computed tomography facilities. *Journal of Medical Systems*, 24(2):61–76, April 2000.

C. J. Rosenquist. Queueing analysis: A useful planning and management technique for radiology. *Journal of Medical Systems*, 11(6):413–419, 1987.

F. G. Shellock and J. V. Grues. Temperature, heart rate, and blood pressure changes associated with clinical imaging at 1.5 t. *Radiology*, 163:259–267, 1987.

K. Skadron, T. Abdelzaher, and M. R. Stan. Control-theoretic techniques and thermal-rc modeling for accurate and localized dynamic thermal management. In *HPCA '02: Proceedings of the 8th International Symposium on High-Performance Computer Architecture*, page 17, Washington, DC, USA, 2002. IEEE Computer Society.

K. Skadron, M. R. Stan, K. Sankaranarayanan, W. Huang, S. Velusamy, and D. Tarjan. Temperature-aware microarchitecture: Modeling and implementation. *ACM Trans. Archit. Code Optim.*, 1(1):94–125, 2004. doi: http://doi.acm.org/10.1145/980152.980157.

J. Srinivasan and S. V. Adve. Predictive dynamic thermal management for multimedia applications. In *ICS '03: Proceedings of the 17th annual international conference on Supercomputing*, pages 109–120, New York, NY, USA, 2003. ACM. ISBN 1-58113-733-8. doi: 10.1145/782814.782831.

H. S. Stone and P. Sipala. The average complexity of depth-first search with backtracking and cutoff. *IBM Journal of Research and Development*, 30(3): 242 –258, may. 1986. ISSN 0018-8646. doi: 10.1147/rd.303.0242.

S. Tattoni, D. Giannico, and M. M. Schiraldi. Operations management techniques for resource optimization in health care structures. In *Sustainable Development: The Role of Industrial Engineering*, pages V.13–V.18, Porto Giardino, Monopoli, Italy, 2009. Quaderni della XIV Summer School Francesco Turco.

A. Trakic, B. K. Li, E. Weber, H. Wang, S. Wilson, and S. Crozier. A rapidly rotating RF coil for MRI. *Concepts in Magnetic Resonance Part B: Magnetic Resonance Engineering*, 35B(2):59–66, April 2009. ISSN 15525031. doi: 10. 1002/cmr.b.20136.

J. B. Tucker, J. E. Barone, J. Cecere, R. G. Blabey, and C. K. Rha. Using queueing theory to determine operating room staffing needs. *The Journal of trauma*, 46(1):71–79, 1999.

D. B. Twieg. The k - trajectory formulation of the NMR imaging process with applications in analysis and synthesis of imaging methods. *Medical Physics*, 10(5):610–622, 1983.

D. A. van Beek, K. L. Man, M. A. Reniers, J. E. Rooda, and R. R. H. Schiffelers. Syntax and consistent equation semantics of hybrid chi. *Journal of Logic and Algebraic Programming*, 68(1-2):129–210, 2006.

P. van de Laar and T. Punter. *Views on Evolvability of Embedded Systems*. Embedded Systems. Springer, October 2010.

J. S. van den Brink, Y. Watanabe, C. K. Kuhl, T. Chung, R. Muthupillai, M. V. Cauteren, K. Yamada, S. Dymarkowski, J. Bogaert, J. H. Maki, C. Matos, J. W. Casselman, and R. M. Hoogeveen. Implications of SENSE MR in routine clinical practice. *European Journal of Radiology*, 46(1):3–27, 2003. ISSN 0720-048X. doi: 10.1016/S0720-048X(02)00333-9.

J. A. A. van der Veen, G. J. Woeginger, and S. Zhang. Sequencing jobs that require common resources on a single machine: A solvable case of the tsp. *Mathematical Programming*, 82(1):235–254, June 1998. doi: 10.1007/BF01585874.

D. M. Walker. Gao testimony before committee on ways and means, house of representatives. Technical report, Comptroller General of the United States, 2005.

D. Weishaupt, V. D. Koechli, and B. Marincek. *How does MRI work? An Introduction to the Physics and Function of Magnetic Resonance Imaging.* Springer, 2 edition, 2006.

W.-H. Yang. Survey of scheduling research involving setup times. *International Journal of Systems Science*, 30(2):143–155, 1999. doi: 10.1080/002077299292498.

S. Zhang and K. Chatha. Approximation algorithm for the temperature-aware scheduling problem. In *Computer-Aided Design, 2007. ICCAD 2007. IEEE/ACM International Conference on*, pages 281–288, 2007. doi: 10.1109/ICCAD.2007.4397278.

S. Zhang and K. S. Chatha. Thermal aware task sequencing on embedded processors. In *DAC '10: Proceedings of the 47th Design Automation Conference*, pages 585–590, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0002-5. doi: http://doi.acm.org/10.1145/1837274.1837418.

Samenvatting

Duty cycle optimalisatie van Magnetic Resonance Imaging systemen

Dertig jaar geleden werd de eerste commerciële Magnetic Resonance Imaging (MRI) scanner geïnstalleerd in het Hammersmith Ziekenhuis in Londen. Deze revolutionaire techniek maakte het mogelijk om weefsel omgeven door bot af te beelden. Dit was een groot voordeel vergeleken met op röntgenstraling gebaseerde beeldvormingsmethoden. Echter was de resolutie van de eerste beelden op basis van magnetische resonantie erg laag en was de benodigde scantijd lang ten gevolge van een zwak signaal en de gevoeligheid voor bewegingen van de patiënt. Er is ondertussen veel onderzoek gedaan om de algehele performance van de machines te verbeteren. Halverwege de jaren '90 zijn snelle beeldvormingstechnieken ontwikkeld welke een enorme invloed hebben gehad op de populariteit van MRI in vergelijking met andere medische beeldvormingsmethoden.

Tegenwoordig zijn er een groot aantal klinische beeldvormingsapplicaties waarbij de opvolgers van de röntgenstraling worden ingehaald door MRI. Bovendien wordt MRI gezien als onschadelijk voor de patiënt, aangezien geen ioniserende straling wordt gebruikt. Echter zijn de belangrijkste nadelen van MRI de sterke magnetische velden, de extreem hoge kosten en de relatief lange onderzoekstijd in vergelijking met beeldvorming met behulp van röntgenstraling. De eerste factor legt strenge veiligheidseisen op die moeten worden nageleefd in de MRI scanner ruimte. De laatste twee factoren leiden tot lange terugverdientijden voor ziekenhuizen. Bovendien zijn de wachtlijsten in ziekenhuizen vaak lang (vaak tot zelfs enkele weken) door de grote vraag naar MRI onderzoeken. Dit heeft een negatieve invloed op de tevredenheid van de patiënt.

In dit proefschrift wordt een nieuwe aanpak beschreven om de benodigde onderzoekstijd voor MRI systemen te verkorten. De tijdsreductie wordt gerealiseerd door het opsplitsen van het MRI onderzoek in segmenten welke vervolgens opnieuw worden ingepland. De planningsalgoritmes zijn gebaseerd op schedulingtechnieken uit het veld van de Operations Research.

127

Er zijn een aantal fysieke parameters welke de performance van MRI systemen beperken, zoals bijvoorbeeld de temperatuur van de MRI hardware tijdens het MRI onderzoek. Tevens kan, ten gevolge van elektromagnetische effecten in de opening van de MRI scanner, de lichaamstemperatuur van de patiënt tot een onaangenaam niveau stijgen. Bij de huidige manier van werken worden deze duty cycle beperkingen gemodelleerd en geverifieerd voorafgaand aan de start van het MRI onderzoek. Vervolgens wordt, indien nodig, de onderzoekstijd verlengd om te voorkomen dat de temperatuurlimieten worden overschreden. Een typisch MRI onderzoek bestaat uit een aantal discrete stappen, ook wel scans genoemd. Elk type scan leidt tot andere duty cycle beperkingen. De aanpak houdt in dat het MRI onderzoek wordt verdeeld in segmenten welke op een zodanige manier worden ingepland dat de ongunstige effecten van de duty cycle gelimiteerde scans worden gereduceerd door niet-gelimiteerde scans.

In dit proefschrift worden diverse scheduling-algoritmes beschreven welke ontwikkeld zijn om met verschillende soorten duty cycle beperkingen om te kunnen gaan en de performance van MRI systemen te verbeteren. De algoritmes zijn geverifieerd aan de hand van een groot aantal MRI onderzoeken. Uit het onderzoek is gebleken dat de benodigde tijd voor MRI onderzoeken gereduceerd kan worden met maximaal 22%. Dit resulteert in een vergroting van de capaciteit van een MRI systeem met maximaal 4 patiënten per dag. Bovendien zijn speciale MRI experimenten uitgevoerd om de algoritmes te valideren.

Tot slot wordt in dit proefschrift een aanpak beschreven om de patiëntenstroom op MRI afdelingen in ziekenhuizen te modelleren. De patiëntenstroom wordt gemodelleerd door middel van wachtrijtheorie om de bottlenecks te identificeren. Vervolgens worden discrete-event computersimulaties uitgevoerd om de beperkingen van de aannames in de klassieke wachtrijtheorie te omzeilen. De huidige manier van werken in ziekenhuizen laat zien dat de MRI scanners niet altijd de bottleneck zijn in de totale werkstroom van onderzoeken. De ontwikkelde modellen kunnen worden gebruikt om de patiëntenstroom te voorspellen voor diverse indelingen van MRI afdelingen en voor verschillende strategieën waarop afspraken worden ingepland. Op basis van deze gedetailleerde modellen kunnen aanbevelingen worden afgeleid om de werkstroom binnen MRI afdelingen te verbeteren.

The resultaten van deze studie kunnen worden gebruikt om de prestaties van MRI afdelingen in ziekenhuizen of onafhankelijke beeldvormingscentra te optimaliseren. Allereerst kan de MRI scantijd worden gereduceerd. Daarnaast kan de patiëntenstroom worden geoptimaliseerd met als resultaat een kortere MRI onderzoekstijd. Dit leidt weer tot grotere tevredenheid van de patiënten en een kortere terugverdientijd van MRI apparatuur.

Curriculum Vitae

Evgeniy Nikolaevich Ivanov was born on the 25th of March, 1983 in Saint-Petersburg, Russia (former Leningrad, USSR). After finishing Lyceum of Physics and Mathematics no. 366 in 2000, he studied Mathematics and Mechanics in Saint Petersburg State University, Russia. In 2005 he graduated from department of Theory of Probability and Mathematical Statistics of the State University. In 2006 he started a PhD project at Systems Engineering group of Mechanical Engineering faculty in Eindhoven University of Technology, The Netherlands. The latter research was carried out in strong cooperation with Philips Healthcare B.V., The Netherlands. Since November 2010 he is working as an engineer at ASML Netheralnds B.V.