Chapter 1

Prologue

In this thesis we discuss and develop *adaptive* multigrid solvers for space–time discretisations of parabolic reaction diffusion equations with a potentially nonlinear forcing term. The latter present a broad class of partial differential equations that can be written in the form

$$\partial_t u - \operatorname{div}(D(x)\nabla u) = f(u) \tag{1.1}$$

for some $u = u(x,t)$ in a space domain over a time interval. We will postpone more rigorous definitions to the following chapters and for now simply assume the problem to be well posed which includes appropriately defined boundary conditions. This type of partial differential equation is used to describe a variety of physical phenomena. In its simplest form, we have a zero source term, i.e. $f = 0$. The resulting heat equation describes the variation of temperature in the space domain over time starting from a set of initial conditions. Other important applications with non-zero source terms, that can be described by (1.1), are chemical reactions over time, the development of populations in [1] or the propagation of wavefronts [2].

A particular instance of a traveling wave which we will particularly focus on and which originally motivated the topic of this thesis, is the propagation of electric signals in heart tissue. It can be modeled using the so-called monodomain equations which are also a reaction-diffusion system [3]. The contraction of our heart is governed by an electric impulse whose charge distribution is a wavefront front through our cell tissue. When trying to numerically approximate such a process one faces a number of challenges. One major challenge that arises is the multiscale range in space and time [4]. The overall domain in space and time is very large compared to the characteristic length-scale on which the rapid local changes of the current potential occur. Thus one requires a high accuracy in time and space. Accurate discretisations result in large systems of equations involving large numbers of degrees of freedom. Solving them in an robust and efficient way is an active area of interest and research [5].

In general when trying to numerically approximate the solution of a partial differential equation there is no unique way to do so and hence many design choices have to be made. A frequently used possibility is the method of lines approach, where one first discretises in space e.g. using finite elements and where the time variable remains continuous, which results in a system of ordinary differential equations that is then to be solved by an appropriate method [source]. A very common approach is to use a time stepping method [source]. That is one computes an approximation for all space elements or nodes at a certain time t_n and then uses those results or even preceding ones to compute the approximate solution at the next time t_{n+1}. This is the natural way to perform operations, because this is also how we move through time, sequentially, causality implies that the solution at a given time depends on the previous one but not the other way around. However in current technological development where there is no further significant increase in CPU clockspeed, the only way to really achieve a gain in computational power is through an increase in the number of processors. Therefore for this to actually translate to a computational speed up one requires algorithms to be more and more parallelisable, that is to allow for more operations to be performed at the same time. The time stepping approach outlined above contains inherently sequential processes, since only the space dimensions allow for parallelisation. As this saturates [source] there is no possibility for a further speed up. Thus it

makes sense to look for methods that utilise a parallelisation in space and time simultaneously. This in turn naturally leads to a space–time discretisation of the equation as a whole [6], which is also what we will be considering in this thesis, a large space–time system which we want to be able to mainly solve in parallel. A short discussion of the research done on this field so far, advantages and difficulties as well as some further references can be found in Section 3.1., while the particular discretisation we chose will be introduced in chapter 4.

It has been shown that large linear systems of equations are often most efficiently solved using iterative schemes [source]. Among them, multigrid methods represent an important and powerful class to approximate such solutions. In the case of sparse, symmetric, positive definite systems they even provide optimality in the sense that their complexity can be bounded by $O(N)$, where N is the number of degrees of freedom [7]. Unfortunately the behavior of multigrid algorithms in an indefinite or not symmetric setting is often not yet very well understood or not suitable [source], and convergence is generally not guaranteed [source]. Therefore we would like to aim for the construction of a system that can claim as many of these preferrable properties as possible. However most space-time solution methods do not give rise to symmetric positive definite systems [6] which is why we recast problem (1.1) as an optimisation probelm, an ansatz known as least squares finite element methods [8] and which will be first introduced in Section 3.4. It entails the construction of a minimisation problem whose solution coincides with the solution of the differential equation. Instead of solving the original problem we now apply a finite element approach in space-time to solve the auxiliary problem whose value for a given input u denotes an energy that we can minimise. In the linear case we are, due to the symmetry and positive definiteness of the system, guaranteed the existence of a global minimiser. In the nonlinear case we consider linearisations of the system which are generally not positive definite, but the symmtetry is maintained because of the commutativity of derivatives. The problem is non-convex but by successively reducing energy we can still find local minima. Hence for a nonlinear source or reaction term f, we additionally require an outer nonlinear iteration scheme which succuessively solves linearisations of the least squares functional. The non-linear solvers that were employed in the implementation section here are a damped Newton method [9] and a trust region method [10], and will be introduced in Section 3.2. We have convergence to a global minimiser in a neighbourhood of the solution, that is for a sufficiently good initial starting iterate the solution of the original problem is recovered.

Below we can see a schematic overview of how these beforementioned core concepts are tied together in order to give rise to a comprehensive solution method.

Overview of the different Steps towards an Approximate Solution

1. Reformulate (1.1) as a mimimisation problem J whose solution coincides with the one of the original equation.
2. Discretise the problem using a space-time finite element approach
3. Derive a nonlinear iteration scheme or energy minimisation method (e.g. Newton or Trust Region method) where we solve a simplified problem using the current iterative solution in each step
4. Solve the arising linear system of equations using a multigrid method
5. Repeat Step 4 with the updated solution until a stopping criterion is met

These are the main ingredients that we will tie together in this thesis in order to develop an efficient, robust and accurate solver to tackle problems of type (1.1). It is a rather novel construction that has, to our knowledge, not been studied in this context and will therefore require further investigations before drawing any final conclusions on its utility. The mathematical methodologies will be introduced more thoroughly in Chapter 3, where we will also explain the particular choice for each of them in more detail, attempting to make use of their favourable properties while trying to avoid the pitfalls. In Chapter 4 we derive a proper problem formulation, which we will then discretise in order to derive linear systems of equations to be solved iteratively. Afterwards we introduce multrigrid methods in Chapter 5, especially discussing the particularities that arise due to the construction presented in Chapter 4. Chapter 6 then contains the numerical results we obtained for various test cases and discusses certain behaviors we observed during our work which will then be followed by a conclusion and an outlook in Chapter 7.

In order to really obtain a meaningful solution u we need a number of properties to be fulfilled. In each nonlinear iteration step the multigrid solver has to converge to the solution of the linearised least squares minimisation problem. In the outer iteration we need the nonlinear iteration scheme to converge to the minimiser of our non-linear functional whose solution as mentioned above is supposed to correspond to the solution of the original problem. However we are not ensured global convergence since the problem is in general non-convex.

Overall we are aiming for a better understanding of the versatility of space-time least squares finite element approaches in general and in combination with multigrid methods. *A focus will be given to the construction of a particular algebraic multigrid method that takes intrinsic properties related to the monodomain equation into account, developing an equally accurate but more efficient way through an adapted coarse grid construction.* To allow for a better understanding of the processes involved in this particular application the following chapter will give a brief insight into the functioning of the human heart, the transmission of electric potential through tissue, the different charge distribution within or between cells or cellular structures and how this can be turned into a mathematical model.

Chapter 2

Cardiac Electrophysiology

Our hearts are absolutely vital for our survival. While it normally functions with an incredible reliability and accuracy that does not even let us begin comprehend the complexity of the mechanisms involved, cardiovascular diseases are estimated to make up for more than 30% of all world wide's death [11]. Often this is related to abnormal heart contractions and thus understanding the processes involved in governing our heart beats is crucial to explaining heart failues. The heart acts as a double pump made out of muscle tissue that provides our bodies with freshly oxygenated blood [12]. A heart has about the size of a fist and sits between our two lungs. It consists of 4 chambers, the upper two atria, that is the left and right atrium and the lower two ventricles which have connections through four heart valves that can open and close respectively but only allow the blood to flow in one direction. The left and the right side of the heart is separated by a wall of tissue known as the atrioventicular septum. An intricate interplay of contraction and relaxation of the chambers governed through electrical stimuli lead to a stable blood flow that enables the replenishment of oxygen levels of the cells in our body. Below we can see a schematic image of the heart, where the arrows indicate the direction of blood flow.

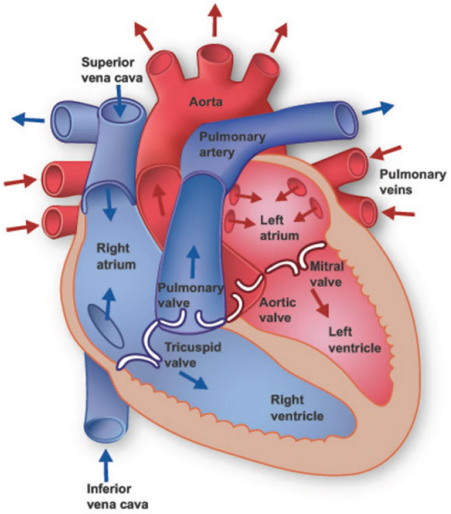

Figure 2.1: Scheme of the Heart [3]

The blood flow through the different chambers of the heart occurs in repeating cycles. After circulating through the body low oxygenated blood flows back into the heart through our veins and enters into the right atrium, which contracts once it is full. This contraction causes a pressure built up and pushes the tricuspid valve open. The blood rushes into the right ventricle, whose walls, once filled, also begin to contract, the pressure within rises again, which shuts tricuspid valve and opens the pulmonary valve to the pulmonary artery from where the blood reaches the lungs and replenishes it oxygen stocks. Afterwards it returns to the left side of the heart from the pulmonary veins to the left atrium, which again, once it is completely filled, contracts and hereby opens the miral valve and forces the blood into the left ventricle. The left

ventricle then pumps the oxygenated blood through the aortic valve into the aorta from where it flows into different parts in the body to supply cells with oxygen and nutrients before returning to the right atrium and repeating its cycle. The mitral and tricuspid valves open, and the aortic and pulmonic vavles close while the ventricles fill with blood. In contrast the mitral and tricuspid valves shut, and the aortic and pulmonic valves open during ventricular contraction. This particular sequence makes sure that all ventricles are filled up to capacity before pumping and that blood flows only in one direction. For references and further information see [13], while the following sections are based on [3].

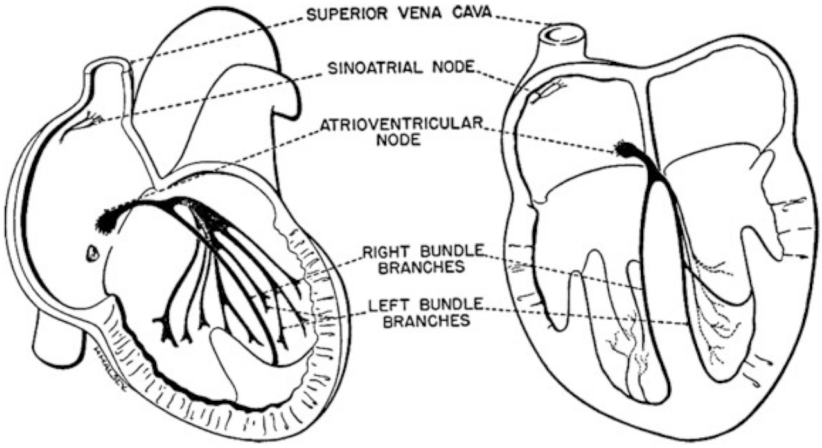

Figure 2.2: Overview of the Hearts Conduction System [3]

The heart contractions are initiated by an electric activiation, that is a depolarizing transistory membrane current which raises the transmembrane potential from its resting value of about -90 to -80 mV to small positive values. This potential describes the difference in the electric potential between the interior and exterior of the cell. The increase is followed by a repolarization current which sends the transmembrane potential back to its resting value. The initial electrical stimulus is generated by the sinoatrial node which is located on the right atrium close to the superior vena cava and possesses the ability to excite its cells autonomously. The frequency of its stimuli is dependent on the parasympathetic nervous system and hormonal factors but under normal health and stress conditions ranges from about 60-100 times per minute. The signal is then transmitted through the surrounding cells and cardiac conduction pathways to the various chambers of the heart. It first propagates to the right atrium and through Bachmann's bundle to the left atrium where it stimulates the cardiac muscle cells of the atria to contract. The activation front then travels to the atrioventricular node situated at the base of the atria. The cells there have a relatively slow conduction velocity and therefore cause a delay in the transmission which is timed this way to achieve optimal pump activity. From the atrioventricular node the stimulus reaches specialised fibres in the bundle of His and the Purkinje network that branch in the left and right bundle onto the inner surface of the ventricals. Again causing a contraction of the cardiac muscle tissue.

2.1 Electrical Activity on the Cellular Level

The heart's walls can be subdivided into 3 different layers; the inner endocardium which surrounds the the heart chambers; the outer endocardium which protects and delimits the heart from other parts of the body and the predominant middle layer consisting of cardiac muscle tissue called myocardium. This is where the conduction of the electric potential and the heart contractions mainly take place. Myocardium is made up of sheets of cells, where each one is roughly of a cylindrical shape with a size ranging from 100-150 μm by 30-40 μm. (diff source 50-150, 10-20) They are organised in a way similar to a brick wall and joined together at the ends by intercalated disks turning them into long fibres. The disks allow for easy ion movement between the cells and thus allowing for a rapid transmission of electrical impulses. Each cardiomyocyte that is each cardiac muscle cell contains bundles of myofibrils which are protein fibres which can slide past each other, making it possible for the tissue to contract. The cell's membrane is called sarcolemma and contains certain transmembrane channels whose opening and closing is governed through electric stimuli (mainly transversal cell direction?). The intercalated disks allow the transit of ions through channels called gap junctions which are predominantly in longitudinal fiber direction. Due to the varying density of the gap junctions in the different directions there is an anisotropic propagation of the electric potential throughout the tissue which complicates its simulation.

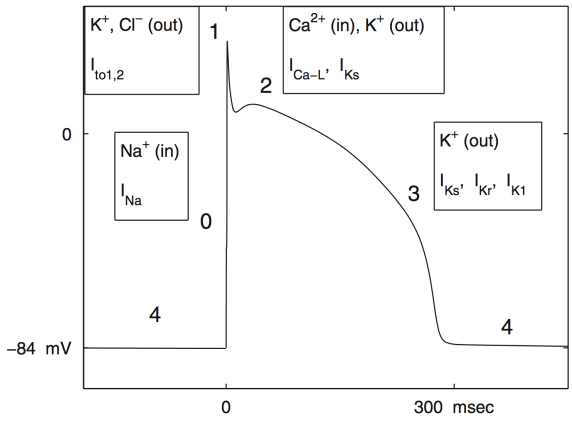

Figure 2.3: Different Phases of Cardiac Action Potential [3]

In this figure we can see a standard ventricular action potential in its main phases, that is the electric charge distribution a cell goes through over time. Following [3] we will have a brief look at the different stages that occur and what they entail.

Phase 0: Depolarization of the cell by opening of Na^+ channels of the sarcolemma which leads to rapid inflow of Na^+ ions into the cell. Hence, the transmembrance potential passes from negative to positive values.

Phase 1: Outward flow of K^+ and Cl^- ions after the inactivaiton of Na^+ channels, which causes a rapid decrease of the potential

Phase 2: Governed by an inward as well as an outward current of Ca^{2+} and K^+ ions respectively such that there almost is a balance in the potential

Phase 3: Repolarization of the cell by closing of the Ca^{2+} channels while outward current of K^+ ions is maintained therefore returning the potential to negative values.

Phase 4: The potential remains at a constant negative value. Some channels are kept open to allow for keeping the right inter-and extracellular charge balance. The cardiomyocyte stays in

this resting state until the next stimulation.

The stimuli are under normal conditions about 0.6-1 seconds apart, which means that the cell is about half the time in its resting phase. They travel as a wave through the cardiac tissue, where one cell excites the next. After this brief description of the processes involved in the functioning of the human heart, let us in the subsequent section turn towards the question of how to adequately represent them in a model and what the particular difficulties are that arise.

2.2 The Monodomain Equation

The propagation of these stimuli is usually modeled with either a version of the bidomain equation or monodomain equation. The former was first developed in the late 1970's and is the more comprehensive one of the two still posing many computational challenges in its implementation and execution. Therefore one often relies on a simpler monodomain model which in the vast majority of applications leads to a very similar solution and can therefore be considered as an adequate approximation [14]. The discrete cellular structure is replaced by an averaged continuous model, giving rise to a parabolic reaction-diffusion equation derived from the cable equation maintaining a conservation of charge. A formulation of the model taken from [15] then reads

$$(\mathcal{X}C_m)\partial_t u - \nabla \cdot (D(x)\nabla u) = (\mathcal{X}C_m) f_{\text{ion}}(u,z) + I_{\text{ext}}(x,t) \qquad (x,t) \in \mathcal{S} \times [0,T] \tag{2.1}$$

$$\partial_t z = g(u,z) \qquad (x,t) \in \mathcal{S} \times (0,T] \tag{2.2}$$

$$n^T D(x) \nabla u = 0 \qquad (x,t) \in \partial\mathcal{S} \times [0,T] \tag{2.3}$$

$$u(x,0) = u_0(x),\ z(x,0) = z_0(x) \qquad x \in \mathcal{S} \tag{2.4}$$

where the above terms describe the following

$u(x,t)$: transmembrane potential
$z(x,t)$: auxiliary variables
$D := D_{\text{in}}^{-1}(D_{\text{in}} + D_{\text{ex}})D_{\text{ex}}^{-1}$, where D_{in} and D_{ex} are intracellular and extracellular conductivities
$f_{\text{ion}}(u,z)$: transmembrane ionic current
$I_{\text{ext}}(x,t)$: source currents
C_m: membrane capitance
$\mathcal{X}$: surface-to-volume ratio

The diffusion term $\nabla \cdot (D(x)\nabla u)$ represents the spread of current through gap junctions and cardiac tissue while the reaction term $(\mathcal{X}C_m) f_{\text{ion}}(u,z)$ describes the flux of ions across the myocyte membrane. The auxiliary variables z capture the dynamics of the cell and the cell membrane. D_{in} and D_{ex} are tensors representing the conductivity given by the local microstructure of the tissue. In the monodomain equation one assumes equal anisotropy ratios in the intra-and extracellular domain. This assumption is not made in the bi-domain equations and therefore gives rise to two different coupled reaction diffusion equations, instead of one.

One version of the above monodomain equation is the FitzHugh-Nagumo model where we neglect the system of ordinary differential equations of auxiliary variables z by setting them to zero. This leads to the neglect of modeling the recovery of the action potential, however the model still captures the propagation of the steep wavefront, which is the crucial part of the process. One also assumes that are no source currents $I_{\text{ext}}(x,t)$. The ionic currents are modeled as cubic functions of the transmembrane potential, that is

$$f_{\text{ionic}}(u) = (u - u_{\text{rest}})(u - u_{\text{thres}})(u - u_{\text{depol}}), \tag{2.5}$$

where $u_{\text{rest}} \leq u_{\text{thres}} \leq u_{\text{depol}}$. The scalar u_{rest} describes the resting potential, which for simplicity we will assume to be zero, u_{depol} describes the depolarisation potential which we will assume to be equal to one. And u_{thres} is the threshold value in between. Hence we can see that f_{ionic} has three roots, in zero, one and u_{thres}, which are then fixed points of the FitzHugh–Nagumo model and it can be shown, that the two in zero and one are stable fixed points, whereas u_{thres} is unstable [16]. Hence depending on the initial conditions u_0 the solution will quickly tend to either zero or one, unless it is exactly α, creating the steep activation front, which will through the diffusion term lead to a propagation of the wavefront over time.

There are a number of difficulties that arise when trying to numerically approximate a solution to this problem. One is the beforementioned challenging task of dealing with large differences in spatial and temporal scaling due to the complex multiscale structure. The combination of fast reaction dynamics and relatively low conductivity lead to steep gradients in the solution. As mentioned before the microscopic scale the electric conduction of excitation fronts happens through ion channels of cellular membranes which are on a scale of the order of 0.1 mm. On the other hand the overall size cardiac tissues invovled entails a size of several centimeters which leads to a spatial spread factor of up to 10^3. Similarly for the time parametrisation we have an even larger spread factor. A normal heartbeat takes about 1 second, that is one full cycle, whereas the step excitation front that is described in phase 0 in the previous section ranges on a much shorter time scale therefore requiring time steps within a range of about 0.1 to 500 milliseconds for accurate representation [4].

But the sheer size of the problem is not the only difficulty, as mentioned in the prologue there are many ways to discretise the domain, various methodologies of how to represent the differential operators and how to approximate the arising linear and nonlinear systems of equations. As we have tried to reason for the particular choices we have made, the following chapter is meant to serve as an introduction to these mathematical methodologies we apply, outline their underlying principles, demonstrate their functioning using standard examples and point out their relevance and applicability for other problems.