Numerical solution of the bidomain equations

BY S. LINGE¹,², J. SUNDNES¹,³,*, M. HANSLIEN¹, G. T. LINES¹,³ AND A. TVEITO¹,³

¹Simula Research Laboratory, PO Box 134, 1325 Lysaker, Norway
²Telemark University College, PO Box 203, 3901, Porsgrunn, Norway
³Department of Informatics, University of Oslo, PO Box 1080, Blindern, Norway

Knowledge of cardiac electrophysiology is efficiently formulated in terms of mathematical models. However, most of these models are very complex and thus defeat direct mathematical reasoning founded on classical and analytical considerations. This is particularly so for the celebrated bidomain model that was developed almost 40 years ago for the concurrent analysis of extra- and intracellular electrical activity. Numerical simulations based on this model represent an indispensable tool for studying electrophysiology. However, complex mathematical models, steep gradients in the solutions and complicated geometries lead to extremely challenging computational problems. The greatest achievement in scientific computing over the past 50 years has been to enable the solving of linear systems of algebraic equations that arise from discretizations of partial differential equations in an optimal manner, i.e. such that the central processing unit (CPU) effort increases linearly with the number of computational nodes. Over the past decade, such optimal methods have been introduced in the simulation of electrophysiology. This development, together with the development of affordable parallel computers, has enabled the solution of the bidomain model combined with accurate cellular models, on geometries resembling a human heart. However, in spite of recent progress, the full potential of modern computational methods has yet to be exploited for the solution of the bidomain model. This paper reviews the development of numerical methods for solving the bidomain model. However, the field is huge and we thus restrict our focus to developments that have been made since the year 2000.

Keywords: cardiac electrophysiology modelling; numerical methods; parallel computing

1. Introduction

The bidomain equations are considered to represent cardiac electrophysiology accurately and have been applied successfully for a wide range of studies. A particularly large volume of research has been carried out on the study of

* Author for correspondence (sundnes@simula.no).

One contribution of 15 to a Theme Issue ‘The virtual physiological human: tools and applications’.
re-entrant arrhythmias and their treatment by externally applied electrical stimulus (e.g. Trayanova et al. 2006).

In spite of being applied successfully for a number of valuable studies, current research based on the bidomain model is slowed by a number of substantial challenges. Some of these challenges are directly related to the huge complexity and variability of living systems, and the consequent difficulty of developing mathematical models and acquiring the necessary experimental data. However, there are also severe challenges related to solving the equations efficiently. Although great advances have been made in both computer hardware and computational algorithms, computational time and, to some extent, memory usage remain a severe obstacle for biomedical research that is based on the bidomain model. The purpose of this review is to give an overview of recent advances and the current state of research in this area, and to indicate directions and requirements for future research. Owing to the large volume of research in this field, we will limit our discussion to developments after the year 2000.

(a) The mathematical model

The bidomain model was originally derived by Tung (1978), and can be formulated as a system of nonlinear ordinary and partial differential equations (ODEs and PDEs). For an introduction to this field, please see Keener & Sneyd (1998). The problem is commonly formulated in terms of the variables $u$, $v$ and $s$, which represent, respectively, the extracellular potential, the transmembrane potential and a vector of cellular state variables. The equations of the model are

\begin{align}
\frac{\partial s}{\partial t} & = F(t, s, v), \\
\nabla \cdot (M_i \nabla v) + \nabla \cdot (M_i \nabla u) & = \frac{\partial v}{\partial t} + I_{\text{ion}}(s, v), \\
\nabla \cdot ((M_i + M_e) \nabla u) + \nabla \cdot (M_i \nabla v) & = 0, \\
n \cdot (M_i \nabla (v + u)) & = 0
\end{align}

and

\begin{equation}
n \cdot (M_e \nabla u) = 0.
\end{equation}

Here, equation (1.1) is a system of ODEs that describes electrochemical reactions in the cells, while equations (1.2) and (1.3) describe propagation of the electrical signal through the cardiac tissue. The boundary conditions in equations (1.4) and (1.5) state that the normal components of intra- and extracellular currents are zero on the boundary, which is a standard assumption when considering an insulated heart or tissue sample. The model may easily be extended to include heart tissue surrounded by a conductive bath or a conductive body (e.g. Pullan et al. 2005; Sundnes et al. 2006), but, for the present study, it is sufficient to consider the insulated heart. The nonlinear term $I_{\text{ion}}(v, s)$ describes the ionic current across the cell membrane, while the symmetric tensors $M_i$ and $M_e$ represent tissue conductivities of the intra- and extracellular spaces, respectively. In the present formulation, the ionic current has been scaled with the cell membrane capacitance, and the conductivities have been scaled with both the membrane capacitance and the cell membrane area to cell volume ratio. See, for instance, Sundnes et al. (2006) for details.
In addition to the complexity of the involved equation systems, one may argue that the main source of the computational load in solving the bidomain model is the rapid dynamics of the cellular reactions described by equation (1.1) and the ionic current term in equation (1.2). The fast reactions lead to a narrow wavefront of activation that propagates through the cardiac tissue. Capturing these fast variations in space and time requires high resolution in the discretization, which results in large systems of equations that must be solved for each time step. From a numerical point of view, there are three main approaches to reducing the computational burden: (i) using adaptive and/or high-order discretization techniques in space and time, to reduce the number of unknowns in the resulting discrete systems, (ii) applying state-of-the-art solution methods (i.e. multi-level solvers) to solve the resulting systems in an optimal manner, and (iii) adapting the solvers for parallel hardware to take advantage of the increasingly available supercomputers and computing clusters. As will be demonstrated below, all of these approaches have been investigated to some extent for the case of the bidomain model.

(b) Focus of the present paper

Although, to date, the most successful approach for reducing the computational load of bidomain simulations has been to simplify or modify the model in some way, we limit our scope to computational methods for the full model. This choice is justified by the facts that: (i) the full model is superior for a number of important applications, in particular for studying the effects of external stimulus and defibrillation, and (ii) the computational techniques derived for the full model are also relevant for simplified models. The latter is particularly true for the monodomain model and various versions of the cable equation (Hodgkin & Rushton 1946; Roth & Krassowska 1998), and to a lesser extent for models such as the eikonal formulation (e.g. Keener 1991).

For research contributions prior to 2000, the reader may wish to consult the review by Henriquez (1993), the more recent review by Lines et al. (2003a) and the papers by Panfilov & Holden (1997) and Zipes & Jalife (2000). In addition, the book by Keener & Sneyd (1998) is highly recommended as a basis for understanding the material presented herein, and a general introduction to numerical methods for the bidomain model can be found in Sundnes et al. (2006). Other overviews of recent research can be found in Vigmond et al. (2008), the main emphasis of which is on solvers for linear systems, and in Clayton & Panfilov (2007), which gives an overview of computational methods based on finite-difference schemes. An overview of challenges connected to cardiac modelling can also be found in the work of Kerckhoffs et al. (2006), with emphasis on multi-scale (cell–tissue–organ–system) issues.

With few exceptions, all substantial improvements of computational methods for the bidomain model involve importing and implementing known methodology from other fields, rather than deriving new numerical techniques for the problem at hand. In many cases, this use of existing technology has required substantial adaptation and innovative approaches, but the applied methodology, as well as theoretical properties on convergence and other issues, is closely related to results from other fields. Given the long tradition and huge volume of computational research in other fields, we see it as natural and useful to
continue and strengthen this tradition. Future advances in computational
techniques for the bidomain model should therefore be based, to an increasing
extent, on reusing state-of-the-art methods from other scientific fields. This
requirement imposes certain restrictions on the choice of method and software for
the model. Pitt-Francis et al. (2008) postulated a number of such requirements
for software packages for cardiac modelling. Here, we define a similar set of
requirements for the computational methods, in order to ease the discussion of and
commenting upon the various methods. Owing to the intimate relation between
computational methods and software, the requirements overlap to some extent with
the ones defined in Pitt-Francis et al. (2008). We state that future computational
methods for the bidomain model should: (i) employ linear solvers that scale
optimally with the number of unknowns, (ii) be usable on parallel computers and
scale well with the number of processors employed, (iii) be flexible with respect to
the cell model used, (iv) be flexible with respect to geometry and boundary
conditions, (v) ease the reuse of existing computational methods from other fields,
(vi) allow a tidy and structured implementation in software, and (vii) allow efficient
and reliable software testing and debugging in modules. In the following discussion
of the various numerical approaches, we will refer to these requirements.

The rest of this paper is organized as follows. In §2, we discuss spatial and
temporal discretization techniques for the bidomain model. The main focus is on
techniques that are based on the finite-element method (FEM), since this
approach has clear advantages in handling anatomically realistic geometries.
In §3, we present the most significant contributions to solving the resulting linear
systems, with the main focus on order-optimal solvers based on multi-level
techniques. In §4, we present works relating to adaptive discretization
techniques, which are surprisingly few considering the potential computational
gains of this approach. Works on parallel solvers are presented in §5, and we give
a brief summary of the discussion in §6.

2. Discretization methods

Equations (1.1)–(1.5) constitute a complex system of nonlinear ODEs and PDEs,
for which it is difficult to derive efficient numerical methods. In this section, we
describe approaches that may be used to discretize the system. We focus on
recent developments. Most of the methods that we consider involve operator-
splitting or other semi-implicit approaches. This development is motivated by
the severe time-step restrictions that are imposed by stability requirements when
using explicit methods, and by the complexity of the equations that are involved,
which make it difficult to apply fully implicit methods.

In view of the requirements stated in §1b, the increasing use of operator-
splitting techniques is a natural and perhaps even necessary development, given
that the modular nature of these methods will ease the fulfilment of nearly all the
requirements. This will be discussed in more detail below.

(a) Spatial discretization

Spatial discretization of the bidomain model turns the system of PDEs and
ODEs given by equations (1.1)–(1.5) into a system of differential-algebraic
equations (DAEs) given by

Phil. Trans. R. Soc. A (2009)
\[ \frac{\partial s}{\partial t} = F(t, s, v), \]  
\[ \frac{\partial v}{\partial t} + I_{\text{ion}}(s, v) = Av + Bu \]  
and
\[ 0 = B^T v + Cu. \]

Here, \( v, u \) and \( s \) are vector quantities representing nodal values of the respective continuous fields in equations (1.1)–(1.3). Furthermore, \( A, B \) and \( C \) are matrices, which represent discrete versions of the diffusion operators in the original model. Note, in particular, the natural symmetry of the right-hand side of equations (2.2) and (2.3), which will arise regardless of the discretization method used as long as any Dirichlet boundary conditions are imposed correctly. Preserving a symmetric system will give substantial computational advantages when solving the linear systems.

A variety of methods can be applied to transform the continuous operators in equations (1.2) and (1.3) into the discrete matrix forms. For the present system, both the finite-difference method (FDM) and the FEM have been used extensively, and a number of studies have also been based on finite-volume methods (FVMs). Referring again to the requirements defined in §1, flexibility with respect to representing anatomically detailed geometries and potentially a mix of Dirichlet and Neumann conditions imposes certain restrictions on the choice of spatial discretization. Although FDMs can handle general geometries by applying domain-embedding techniques, the geometrical considerations clearly favour the more flexible FEMs and FVMs. The present discussion will therefore mainly focus on these methods. As noted above, an overview of FDMs for the bidomain model can be found in Clayton & Panfilov (2007).

The starting point for a spatial discretization is a geometrical model of the relevant anatomical structure. Several groups have created detailed geometrical models of the ventricles, which have subsequently been used in finite-element computations of cardiac electrophysiology and mechanics. The more detailed models include the contributions from the Auckland group (Nielsen et al. 1991; Legrice et al. 1997; Stevens et al. 2003), which presented models of geometry and fibre orientation for canine and porcine ventricles. The models were based on structured grids with cubic Hermite interpolation functions, which enable a smooth representation of ventricular geometry with relatively few elements. In addition to the geometrical models, the data from, for example, Nielsen et al. (1991) have been made available to other research groups and have been used for a number of finite-element models, for instance the unstructured tetrahedron-based model described in Sundnes et al. (2006).

Other models based on cubic Hermite interpolation include the rabbit ventricular model of Vetter & McCulloch (1998), which has been used extensively since it first appeared (e.g. the recent works of Rodriguez et al. (2005), Arevalo et al. (2007) and Vigmond & Clements (2007)).

As noted above, a clear advantage of the cubic Hermite and other high-order finite-element meshes is that the ventricular geometry can be represented with very few elements. This is convenient for fitting the meshes to measured geometrical data and for computations of ventricular mechanics. As described

*Phil. Trans. R. Soc. A* (2009)
by, for instance, Usyk et al. (2002), the nonlinear elasticity equation describing
cardiac mechanics imposes less strict requirements on discretization parameters
than the bidomain model, and computations may often be based on the same
coarse mesh that is used for fitting the geometry. However, this is not possible for
the bidomain computations, so the original mesh needs to be refined substantially.
This can be done by: (i) remeshing the geometrical model with a
finer, possibly unstructured mesh, as demonstrated in Arevalo et al. (2007),
(ii) uniform refinement of the structured finite-element mesh (Usyk et al. 2002),
or (iii) placing a deformable finite-difference grid over each element of the finite-
element mesh (Buist et al. 2003). Although the latter approach inherits some of
the computational advantages of the FDM, the added complexity of such a
hybrid approach seems to outweigh any potential benefits.

A different approach to geometrical modelling and finite-element discretization
was initiated by Hooks et al. (2002), who modelled a small transmural slab (0.8 ×
0.8 × 3.7 mm³) of the rat left ventricle. Although the geometry of the slab was
completely regular, the purpose of the study was to investigate the implications of
cardiac microstructure on electrical propagation and, in particular, on the tissue
response to an external stimulus. Tissue microstructure was represented in terms of
discontinuities (cleavage planes) by removing internal elements of an extremely
finely structured finite-element mesh, and by imposing homogeneous Neumann
conditions on the resulting internal boundaries. In total, the removed elements that
represented the cleavage planes comprised approximately 11 per cent of the
intracellular volume. The tissue discreteness represented by cleavage planes is
usually not included in cardiac bidomain models, but the results of Hooks et al.
suggest that this structural discontinuity might be important, at least in the study
of external tissue stimulus. This issue has yet to be investigated further by other
research groups, and one may argue that a definite conclusion regarding the
importance of representing cardiac microstructure has not been reached. However,
should it be concluded that this is essential for the study of certain important
phenomena, it will have clear implications for the computational load of the relevant
simulations, because the accurate representation of cardiac microstructure will lead
to very strict resolution requirements.

Trew et al. (2005) introduced an FVM to give an improved representation of
the cleavage planes. The FVM is based on calculating and tracking the flux
between the mesh elements and allows the cleavage planes to be represented
without the need for removing elements. Instead, a cleavage plane can be
introduced with zero volume, simply by setting the flux over given element sides
to zero. It has been claimed that this approach offers several improvements over
a finite-element approach, although it is clear that the idea of cleavage planes as
zero-volume discontinuities could be achieved quite easily in a finite-element
mesh, simply by introducing internal no-flux boundaries without the removal of
elements. However, the FVM may have other computational advantages, as is
also pointed out by Trew et al., such as increased sparsity of the resulting linear
systems. Another FVM approach was presented by Penland et al. (2002),
who built on the FVM of Harrild et al. (2000) and applied structured
hexahedral meshes.

Harrild and Henriquez used their FVM (Harrild et al. 2000) to develop an
anatomically detailed simulation model of the atria, using the cell model of
Nygren et al. (1998) for the ion kinetics. Even though it was not derived from
imaging data, the model significantly advanced the realism of atrial modelling, including as it did a three-dimensional representation of both atria with major muscle bundles. Anisotropy was also incorporated, choosing conductivities to obtain realistic propagation velocities in the tissue components. A total of 248 264 hexahedral elements were used in the mesh, and, even with the monodomain model, this resulted in a substantial computational task. 

Zozor et al. (2003) presented a generalized finite-difference approach for the simulation of cardiac electrophysiology on an arbitrarily shaped three-dimensional monolayer. As emphasized by the authors, neglecting the layer thickness will ease the computational workload, and has particular relevance to the thin walls of the atria. Finite-difference equations, based on local flux conservation, were solved on surface triangular meshes. Several geometries were used, including an atrial geometry based on magnetic resonance imaging (MRI) images. A three-dimensional monolayer model was also used by Virag et al. (2002), who derived triangular Delauney meshes from MRI images of human atria. The same model has also recently been used by Jacquemet et al. (2005) for further investigations of atrial fibrillation.

A very detailed atrial model has more recently been presented by Seemann et al. (2006), who built the geometry from the visible female dataset (Spitzer & Whitlock 1998). These data are based on transversal cryosections from a female heart, with a section distance of 0.33 mm. An FEM with more than 18 million elements is used to solve the monodomain model and predict the spatio-temporal development of the transmembrane potential.

Although several methods allow the geometrical intricacies of cardiac muscles to be modelled, it is still extremely cumbersome to use real-world geometrical data. Open libraries of such data would indeed enable further realism in computational studies of cardiac electrophysiology; it would greatly simplify the entrance to this field for advanced research groups in scientific computing.

(b) Discretization in time

Mathematical models of the single cardiomyocyte are under intense development, as demonstrated by the development of a huge number of increasingly complex models. The CellML project (Lloyd et al. 2008) represents a substantial effort to present these models in a comprehensible manner. Although attempts have been made to introduce simplified, yet physiologically relevant, models, the main trend is definitely to address ever more complex models. This development obviously represents a moving target for the computational scientist. It lies beyond the scope of this paper to review that field and we will consequently restrict our attention to the generic system presented above.

Equations (2.1)–(2.3) are a DAE system of index 1 (see, for instance, Ascher & Petzold (1998) for a definition of DAE index). This has two important implications. The first one is that, being a DAE system, some degree of implicitness will always be required in order to fulfil the algebraic constraint in equation (2.3), which is commonly referred to as the elliptic part of the bidomain model. The second implication is that, since the index of the system is 1, the system is, in principle, fairly easy to solve, and the literature on solving ODEs and DAEs includes a large number of suitable methods. However, this is true only in principle, because the combination of highly nonlinear systems in
equation (1.1) and the huge dimensions of the matrices $A$, $B$ and $C$ in equations (2.2) and (2.3) makes the efficient implementation of even the simplest standard DAE solvers highly non-trivial. Fortunately, the low index of the system ensures that there are no hidden constraints or other problematic issues related to higher index DAEs (see Ascher & Petzold (1998) for details).

It is possible to employ fully implicit time-discretization methods to equations (2.1)–(2.3), but this approach makes it difficult to fulfil the requirements stated in §1b. In particular, it is difficult to imagine that requirements (iii) and (v)–(vii) can be met without splitting the equation system at some level. It would, of course, be possible to first apply a fully implicit time discretization and then split the resulting algebraic equations either on the level of nonlinear equations or the final linearized systems. This method may even allow the desired model flexibility and reuse of existing software, and may therefore be a desirable direction for future research because it would offer the advantage of higher order time-discretization methods. However, to date, most of the derived methods have been based on first splitting equations (2.1)–(2.3) into more manageable parts and then performing a more or less independent time discretization of the various subsystems. This kind of splitting, generally referred to as operator splitting, is very common in applications of simulations that are based on PDEs. Indeed, it is probably one of the most frequently reinvented algorithms in scientific computing. One reason for this is that the method is only rarely presented in textbooks (LeVeque (2007) being a prominent exception) and a reason for that is probably that the method is not straightforward to analyse. Still, it is an extremely versatile tool that no computational scientist should work without.

A variety of techniques have been applied to split the bidomain equations into more manageable parts. Early examples include the methods of Keener & Bogar (1998) and Qu & Garfinkel (1999). The latter work introduced second-order Strang (1968) splitting to the field of cardiac modelling, a method that was later studied in more detail by Trangenstein & Kim (2004). Both of the latter studies applied the Strang splitting to the monodomain model coupled to a Luo–Rudy phase 1 (Luo & Rudy 1991) cell model, but the approach may easily be extended to the full bidomain model, as described by Sundnes et al. (2005).

The effect of applying splitting methods, such as Strang splitting, to the bidomain model is that equation (2.2) is split into: (i) a system of linear ODEs that contain the diffusive terms and (ii) a spatially uncoupled nonlinear ODE system that describes the effect of the ionic current. This spatial uncoupling of the nonlinear term can also be achieved by a variety of semi-implicit methods, as applied to the monodomain model by ten Tusscher & Panfilov (2003) and to the full bidomain model by Austin et al. (2006a), Potse et al. (2006) and Plank et al. (2007).

Owing to the high level of complexity, it is difficult to perform a detailed mathematical analysis of the accuracy and stability of operator-splitting techniques, and few results of this kind have been reported in the literature. The application of operator splitting to the monodomain model was analysed by Schroll et al. (2007). Among other things, the authors concluded that the order of the reaction and diffusion steps was not irrelevant to the achieved error. Other analytical investigations include the work of Puwal & Roth (2007), who analysed the stability of a forward Euler time discretization combined with a standard
finite-difference space discretization of the bidomain equations. They concluded that the stability conditions were of a form similar to the stability restrictions of the monodomain model.

All of the operator-splitting approaches discussed above uncouple the cell-model ODE systems from the electrodiffusive parts of the bidomain model. This entails that each time step of the method involves solving the discretized diffusion model that results from equations (2.2) and (2.3), as well as advancing a large number of spatially uncoupled cell-model ODE systems. The only real constraint that applies to the choice of method for these two systems is that they must give at least the same order of accuracy as the overall splitting method. That is, if one is to obtain the overall second-order accuracy offered by Strang splitting, the two subsystems must be solved with methods that are at least second-order accurate. For first-order splitting methods, which constitute the majority of the methods presented above, no such restrictions apply.

We first consider the linear DAE systems obtained by applying splitting or semi-implicit time discretization to equations (2.2) and (2.3). This has commonly been solved by applying operator splitting once more, to obtain one scalar PDE for $v$ and one for $u$. Such an approach, based on a Gauss–Seidel-like algorithm, was presented in Lines et al. (2003b), and the FVM presented by Austin et al. (2006a) applied essentially the same time-discretization scheme. A similar method was used by Plank et al. (2005) and Potse et al. (2006), with the difference that they applied a forward Euler discretization of equation (2.2). The use of the explicit method leads to fairly strict restrictions on time steps, while the method applied in Lines et al. (2003b) and Austin et al. (2006a) does not seem to cause any such problems.

Given that no stability problems have been reported from applying a Gauss–Seidel approach to equations (2.2) and (2.3), this method is clearly an attractive choice. The method effectively splits the coupled PDE system into scalar PDEs, and this will ease the application of existing methods and thereby the fulfilment of the requirements stated in §1b. However, the separate solution of the equations may make it more difficult to obtain second-order accuracy, which is necessary to retain the overall accuracy of the Strang splitting scheme. This limitation motivates discretizing and solving the two equations fully coupled. This was first carried out by Keener & Bogar (1998), who applied a fully coupled solution of equations (2.2) and (2.3). A second-order version, which was based on a Crank–Nicolson discretization, was later presented in Sundnes et al. (2005).

In spite of the large body of research that can be found on ODE solvers for both stiff and non-stiff systems (e.g. Hairer & Wanner 1991; Hairer et al. 1991), significant efforts have been invested in developing special-purpose ODE solvers for the cell-model systems. The most important contribution in this respect is the so-called Rush–Larsen method (Rush & Larsen 1978), which has been widely used in the research community. This method exploits the fact that a large number of the equations of the cell-model systems describe so-called gate variables, and are written in the form

$$\frac{dm}{dt} = \alpha(1 - m) - \beta m.$$  

Here, the rate functions $\alpha$ and $\beta$ are nonlinear functions of the membrane potential, but we see that the equation is linear in $m$. By assuming that the
membrane potential is constant over a time step, this quasi-linearity can be exploited to derive an update formula based on the exact solution of scalar, linear ODEs. The method described in Rush & Larsen (1978) applied this update formula to all the gate variable equations and solved the remaining equations with the explicit Euler method. Although the overall algorithm is only first-order accurate, it showed a substantial improvement in stability and accuracy, resulting in significant central processing unit (CPU) time reduction, with almost no additional complexity with respect to its implementation.

Other authors have derived variations of the Rush–Larsen scheme. Whiteley (2006) exploited the quasi-linearity to obtain an explicit update formula based on the backward Euler method

\[ w^{n+1} = \frac{w^n + \Delta t a_w(v^{n+1})}{1 - \Delta t b_w(v^{n+1})}. \]  

This formula also leads to a first-order algorithm. As described in Sundnes et al. (2003), it is also possible to derive second-order schemes that are similar to the Rush–Larsen method.

The main reason for the popularity of the Rush–Larsen method and similar approaches is probably that it combines the implementational complexity of explicit solvers with significantly improved accuracy and stability. Owing to the fact that the original method applied a forward Euler discretization for all fully nonlinear ODEs, it suffers from stability problems when applied to the stiffest cell-model systems. More recent modifications avoid many of these problems, but, for some cell models, they have still been shown to perform less well than standard ODE solvers, such as implicit Runge–Kutta methods (see Sundnes et al. 2001, 2003).

Many of the general-purpose ODE solvers have been implemented in well-tested, high-quality software that is publicly available. Although reusing such software components has a number of advantages, it is a non-trivial task because of certain characteristic features of the problem at hand. Recall that each time step typically involves initializing millions of ODE systems with the solution of the PDEs, then integrating the systems over a very short time interval, and finally communicating the updated membrane potentials back to the PDE solver. This means that initializing the solvers and copying solutions may take up a substantial portion of the computation time, unless these tasks are implemented carefully. General-purpose ODE solvers are normally not optimized for these tasks, since they are typically developed for solving a single ODE system over an extended time interval, and, in this case, the initialization cost is negligible. Reusing general-purpose ODE software in a splitting algorithm for the bidomain model will therefore normally be inefficient, unless one is able to modify and optimize the solver initialization.

3. Solving the linear equations

All the discretization techniques introduced above require the solution of linear systems of equations. Solution methods, based on operator splitting, may typically give rise to two different categories of linear systems: (i) small systems that result from solving the nonlinear ODEs of equation (2.1) with implicit
solvers and a Newton iteration scheme (e.g. Sundnes et al. 2006) and (ii) large systems that arise from equations (2.2) and (2.3). The systems that arise from equation (2.1) are normally solved with direct methods and will not be considered here. Instead, we devote our attention to the large linear systems from equations (2.2) and (2.3), which arise from the discretization of the diffusion operators in equations (1.2) and (1.3).

Owing to the strict resolution requirements described above, and the increased focus on large-scale three-dimensional simulations, the linear systems from equations (2.2) and (2.3) are potentially huge, and it is crucial to employ methods that scale well as the size of the problem increases. Ideally, the methods should have order-optimal behaviour, which means that the computational load should increase linearly with the number of unknowns. From a practical point of view, this means that a simulation of, for instance, 300 cm$^3$ of active cardiac tissue should take approximately 300 times more CPU time than a simulation of 1 cm$^3$ of tissue. Traditional linear solvers, such as direct methods and simple iterative methods (e.g. Greenbaum 1997), will not yield such a linear increase, but have been used successfully for bidomain simulations on small- to medium-sized domains (e.g. Muzikant & Henriquez 1998; Meunier et al. 2002; Vigmond et al. 2002; Ashihara et al. 2003).

Multi-level solvers are the only group of linear solvers that can be shown to give order-optimal behaviour for linear systems that arise from the discretization of PDEs. These solvers were introduced to the field of cardiac modelling by Keener & Bogar (1998), who applied a geometric multi-grid (GMG) as a stand-alone solver for the block linear systems that arise from a coupled discretization of equations (2.2) and (2.3). The computational domain considered was a simplified two-dimensional domain. The use of multi-level methods to realistically shaped three-dimensional domains was introduced by Sundnes et al. (2002), where the linear system was solved with the conjugate gradient (CG) method and a GMG block preconditioner. Experiments on two- and three-dimensional domains indicated order-optimal or close to order-optimal convergence properties. This was later confirmed theoretically in Mardal et al. (2007). The performance of the GMG-preconditioned CG method was later investigated by dos Santos et al. (2004), who also developed a parallel version. This method will be commented upon below.

Although the GMG method has proved to be an effective tool for large-scale bidomain simulations, it suffers from some limitations when applied to geometrically complex computational domains. GMG is based on a hierarchy of successively coarser grids, on which the residual of the solution is restricted and smoothed for each iteration (see Briggs et al. (2000) for details). The limitation of this approach is that even the coarsest grid must give a reasonably accurate representation of the geometry, and this is difficult to achieve for simulations on a realistic heart geometry.

A more flexible tool for irregularly shaped domains is the algebraic multi-grid (AMG) method (Brandt et al. 1984; Brandt 1986), which mimics the steps of the GMG, but without explicitly generating the grid hierarchies. The central multi-grid operations of coarse-grid restriction, smoothing and interpolation back on the fine grid are, instead, constructed directly from the information contained in the linear system matrix. This gives increased geometrical flexibility because only a fine grid needs to be represented. It also has known advantages for
handling problems with discontinuous or strongly varying coefficients, as these features will be reflected in the system matrix and therefore in the constructed restriction and interpolation operators. The AMG was first studied in the context of bidomain simulations by Austin et al. (2005), who compared the performance with a related method named black-box multi-grid (BoxMG; see Dendy 1982). Similar to the AMG, the BoxMG method employs matrix-dependent interpolation, but a key difference is that the BoxMG assumes an underlying grid hierarchy and therefore can only be applied to structured grids. A performance comparison of the two methods used as stand-alone solvers revealed a slight performance advantage of the BoxMG. In a related study, Austin et al. (2006a) investigated the performance of the GMG and BoxMG for bidomain problems with discontinuous conductivities and concluded that the BoxMG shows more robust performance than the GMG in this setting. However, a similar robustness with only a slight increase in computational load was obtained by using the GMG as a preconditioner for the CG method. The BoxMG method has also been used by Austin et al. (2006b) for homogenizing (upscaling) high-resolution imaging data that are to be used in bidomain simulations.

If we again refer to the requirements stated in §1b, the AMG method stands out as particularly attractive for efficient bidomain simulations. While all the considered multi-level methods have the potential of order-optimal convergence and hence fulfil requirement (i), the AMG method is clearly more flexible than both the GMG and BoxMG with respect to geometries, and the black-box nature of the AMG facilitates the reuse of software that was developed in other fields (requirements (iv) and (v)). The latter point was recently exploited by Plank et al. (2007), who developed a parallel bidomain solver based on the BoomerAMG (Henson & Yang 2002) preconditioner. This will be commented on in greater detail below.

An alternative to the multi-level methods was suggested by Deo & Vigmond (2005), who suggested a CG preconditioner based on the Arnoldi method. Numerical experiments showed that the method was superior to a standard incomplete LU preconditioner and showed a linear increase in CPU time as the size of the problem increased. The Arnoldi preconditioner was later extended to a cascaded preconditioner that involved a few iterations of successive over-relaxation, which further improved the performance (see Deo et al. 2005, 2007). The linear relation between CPU time and problem size for the Arnoldi preconditioner indicates order-optimal behaviour similar to that of the multi-level solvers, and Deo et al. (2007) suggested that the method is considerably faster than existing multi-level solvers. However, the superior performance has yet to be confirmed for large-scale computations and, in contrast to the case of the multi-grid solvers, the order-optimal convergence has not been confirmed theoretically.

4. Adaptive solvers

In §§2 and 3, we discussed two alternative ways to speed up bidomain computations: reducing the number of discrete equations and solving the equations more efficiently. As commented in §2, high-order discretization in space and time has been explored to some extent, and there are indications that these approaches may be effective in reducing the computational burden.

Phil. Trans. R. Soc. A (2009)
An alternative approach for reducing the number of discretization nodes is to apply adaptive discretization techniques, where either the resolution or the interpolation order, or both, is adapted to the local characteristics of the solution. The interest in adaptivity is motivated by the characteristics of the solution, which has large spatial and temporal variations around the activation wavefront and very smooth characteristics elsewhere. The strict resolution requirements listed above are therefore primarily dictated by the behaviour of the solution in a relatively small area, which calls for using different resolutions in different regions and at different time intervals. Time-adaptive schemes have been used for the bidomain model and related models for decades (e.g. Rush & Larsen 1978), whereas spatially adaptive discretization schemes have appeared more recently. Both the time-adaptive and space-adaptive schemes have mostly been based on adaptive resolution (h-adaptivity) rather than adaptive interpolation order (p-adaptivity). Furthermore, in contrast to time-discretization techniques and solvers for linear systems, most of the adaptive methods introduced to date are directly applicable to, and equally relevant for, both the monodomain and bidomain models. For this reason, the following discussion will include a number of papers entirely focused on the monodomain model.

A time-adaptive scheme with error control was suggested by Rousseau & Kapral (2000), who used the classical technique of step doubling (e.g. Ascher & Petzold 1998) to estimate the local error in a forward Euler scheme and chose the local time step to satisfy a given error tolerance. The method was derived and analysed for the monodomain model coupled to a FitzHugh–Nagumo cell model. Cherry et al. (2003) presented a space–time-adaptive algorithm for three-dimensional monodomain simulations of cardiac electrophysiology in anisotropic and inhomogeneous tissue. The work was an extension of a previous algorithm (Cherry et al. 2000) that enabled similar simulations to be performed for isotropic tissue in two dimensions. This method was based on an adaptive mesh-refinement algorithm that had previously been introduced for the integration of hyperbolic PDEs (Berger & Oliger 1984) and which used estimates of derivatives in space and time to select appropriate refinements of the mesh. An explicit method was used for the time discretization of the diffusion equation, in combination with a finite-difference spatial discretization on successively refined patches of structured Cartesian meshes. Substantial performance increases were obtained compared with a fixed mesh, even when applied to the complex activation patterns of re-entrant arrhythmias.

A straightforward approach to adaptivity was tested by Lines et al. (2003b), where the full bidomain model was addressed with an FEM on an unstructured mesh. The mesh was refined in the area around the activation wavefront, which was located by estimating the gradient of the membrane potential v throughout the finite-element mesh. The adaptive scheme was combined with the multi-grid method, so that the grid hierarchy was constructed by successive refinements in the area around the wavefront, instead of the uniform refinements used in, for example, Sundnes et al. (2002). The same criterion for refinement was employed by Whiteley (2007), who named this pragmatic yet effective approach physiology driven adaptivity. Whiteley’s method was based on a fixed coarse grid and a moving, variable fine grid that followed the activation wavefront. Both methods employed adaptivity in space and time, and gave a substantial increase in...
computational performance. A similar approach, which only included adaptivity in time, was employed by Colli Franzone et al. (2005), who chose the time step to control the variation in transmembrane potential over each time step.

Spatial adaptivity using a non-conforming mortar element method was proposed by Pennacchio (2001, 2004). A particular feature of this method is that the fine grid of the depolarization region can be chosen independently of coarse mesh, and the equality of the two solutions is satisfied weakly.

Reuse of existing methodology and software tools, as suggested in the requirements stated in §1b, becomes increasingly relevant also in the context of adaptivity. Colli Franzone et al. (2006a) presented a three-dimensional solver with adaptivity in space and time, based on the KARDOS software (Erdmann et al. 2002). This solver has recently been extended by Deuflhard et al. (in press) to an unstructured mesh of a human heart geometry, and its performance has been investigated for the challenging application of ventricular fibrillation.

5. Parallel solvers

Unless radical improvements are obtained either in discretization methods or the performance of single CPUs, an efficient solution of the bidomain equations can only be obtained by exploiting parallel hardware. This has been a key topic for several years, but recent hardware development has increased its relevance. While, in the fairly recent past, the performance of computer processors increased at a rapid rate due to increases in the clock rate, it is now primarily increased by increasing the number of processor cores (e.g. Sutter 2005). Thus, the most recent increases in performance can only be fully used by parallel algorithms.

As described in §2, the application of operator-splitting techniques leads to a spatial uncoupling of the cell-model ODE system equation (2.1), which entails that these can be integrated independently for each time step. This has important consequences for parallel computing because this part of the problem becomes trivial to parallelize on shared-memory architectures and is also very simple on distributed memory systems, such as the widely available clusters. A more challenging task is to efficiently distribute and solve the linear systems in equations (2.2) and (2.3).

Examples of early contributions to parallel solution approaches are Fisher & Thakor (1991), Pollard & Barr (1991), Winslow et al. (1993), Ng et al. (1995), Saleheen et al. (1997), Quan et al. (1998) and Cai & Lines (2002). Common to most of these contributions is that they considered explicit solution schemes for the monodomain model, which can be parallelized very efficiently, and that they used shared-memory architectures. An early contribution to distributed memory approaches was made by Porras et al. (2000), who compared four different parallel schemes for the two-dimensional monodomain equations using a Beeler–Reuter cell model (Beeler & Reuter 1977).

The work of Vigmond et al. (2002), which was addressed above, also included parallel computations on shared-memory architectures with two to four processors and compared the performance of a number of direct and iterative methods.
Given the complexity of implementing efficient parallel solvers, reusing existing software becomes even more attractive than for serial codes. An early contribution in this context was made by Colli Franzone & Pavarino (2004), who developed a parallel bidomain solver based on the PETSc library (http://www.mcs.anl.gov/petsc; Balay et al. 2007). This solver was later used to study repolarization and the distribution of the duration of action potential, elaborating on the effect of transmural heterogeneities (see Colli Franzone et al. 2005, 2006b). The PETSc library was also used by dos Santos et al. (2004), to implement a parallel GMG-preconditioned CG solver for the linear systems that arise from equation (2.3). Murillo & Cai (2004) used the PETSc library to devise a parallel bidomain solver that was based on a fully implicit time discretization. Although the simple FitzHugh–Nagumo model was used, this approach included the additional challenge of solving large nonlinear systems of equations, as no spatial uncoupling of the cell-model ODEs was obtained. More recently, Plank et al. (2007) have developed a similar solver using PETSc in combination with the BoomerAMG (Henson & Yang 2002) preconditioner.

6. Concluding remarks

Electrophysiology research based on the bidomain model is facing a number of obstacles, of which the huge computational burden stands out as one of the most important. Although the model has been used in a large number of interesting studies, the full potential of the simulations can only be reached if the computation time can be reduced dramatically. In §1, we stated a couple of hours as a necessary goal for the computation time of full bidomain simulations, whereas other authors have introduced real-time simulations as a natural goal (Kerckhoffs et al. 2006).

It is our opinion that the necessary reductions in computation time can only be obtained by successfully combining a number of the techniques discussed so far; high-order discretization methods must be combined with adaptivity in space and time, and the resulting linear systems must be solved with order-optimal solvers that scale well on large parallel computers. To date, a number of important contributions have been made to all of these fields, but they have yet to be combined successfully. However, this previously formidable task is growing increasingly realistic by the existence of well-designed and well-tested software and methods for generating solutions, and the increasing use of these components for bidomain simulations.

References


Rodriguez, B., Li, L., Eason, J. C., Efimov, I. R. & Trayanova, N. A. 2005 Differences between left and right ventricular chamber geometry affect cardiac vulnerability to electric shocks. *Circ. Res.* 97, 168–175. (doi:10.1161/01.RES.0000174429.00987.17)


