Computer simulations of electrical behaviour in the whole ventricles have become commonplace during the last few years. The goals of this article are (i) to review the techniques that are currently employed to model cardiac electrical activity in the heart, discussing the strengths and weaknesses of the various approaches, and (ii) to implement a novel modelling approach, based on physiological reasoning, that lifts some of the restrictions imposed by current state-of-the-art ionic models. To illustrate the latter approach, the present study uses a recently developed ionic model of the ventricular myocyte that incorporates an excitation–contraction coupling and mitochondrial energetics model. A paradigm to bridge the vastly disparate spatial and temporal scales, from subcellular processes to the entire organ, and from sub-microseconds to minutes, is presented. Achieving sufficient computational efficiency is the key to success in the quest to develop multiscale realistic models that are expected to lead to better understanding of the mechanisms of arrhythmia induction following failure at the organelle level, and ultimately to the development of novel therapeutic applications.

Keywords: multiscale modelling; cardiac electrical activity; monodomain; bidomain; ordinary differential equation integration

1. Introduction

Computer simulations of electrical behaviour in the whole ventricles (Xie et al. 2004; Rodriguez et al. 2005; Potse et al. 2006; Ten Tusscher et al. 2007; Ashihara et al. 2008) or atria (Harrild & Henriquez 2000; Vigmond et al. 2001; Virag et al. 2002; Seemann et al. 2006) have become commonplace during the last few years.

*Author and address for correspondence: 216 CSEB, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA (ntrayanova@jhu.edu).

One contribution of 11 to a Theme Issue ‘The virtual physiological human: building a framework for computational biomedicine II’.
However, even when powerful state-of-the-art computational resources are used, attempts to integrate behaviour from the protein scale of ion channels to the organ scale of cardiac arrhythmias remain enormously challenging, and typically include significant trade-offs between representations of different types of complexities (subcellular processes versus structural complexities). To arrive at a computational efficiency in modelling the (patho)physiological processes in the heart, such that it permits the exploration of the parameter space of interest, the simplifications typically made are as follows.

— The geometry of the organ is represented in a stylized fashion (‘one heart geometry fits all’ approach; Xie et al. 2004; Rodriguez et al. 2005; Ashihara et al. 2008); only parts of the heart are modelled, such as slices across the ventricles (Meunier et al. 2002; Trayanova & Eason 2002; Hillebrenner et al. 2004) or wedges (Burton et al. 2006; Plank et al. 2008); and idealized geometries, such as myocardial slabs (Vigmond & Leon 1999; Cherry et al. 2003; Plank et al. 2005), sheets (Beaumont et al. 1998; Skouibine et al. 1999; Anderson & Trayanova 2001; Samie et al. 2001; Kneller et al. 2002; Weiss et al. 2005) or strands (Thomas et al. 2003; Qu et al. 2006), are used.

— In constraining the degrees of freedom, the choice of the computational mesh discretization often leads to (i) under-representation of (to the degree of fully ignoring) the finer details of the cardiac anatomy, such as endocardial trabeculations or papillary muscles, and (ii) the necessity to adjust ad hoc the tissue conductivity tensors in order to avoid the artificial scaling of the wavelength, thus compensating for the dependence of conduction velocities on grid granularity. That is, as the grid is coarsened with all other model parameters remaining unchanged, conduction velocity becomes reduced and thus the wavelength is diminished, with conduction block occurring above a certain spatial discretization limit.

— The myocardial mass is treated as a homogeneous continuum, without representing intramyocardial discontinuities such as vascularization, cleavage planes, or patches of fat or collagen.

— The specialized cardiac conduction system, i.e. the sinuatrial (SA) and atrioventricular (AV) nodes and the Purkinje network, is typically not represented in the whole-organ simulations, although a few exceptions exist (Berenfeld & Jalife 1998; Vigmond & Clements 2007; Ten Tusscher & Panfilov 2008).

— Myocardial membrane ion transport kinetics are modelled in a simplified fashion (Rogers & McCulloch 1994; Ten Tusscher & Panfilov 2006a). Reduced models preserve salient features such as excitability, refractoriness, electrical restitution, etc.; these are usually represented phenomenologically at the scale of the cell (so that they can be easily manipulated). Such approaches have led to important insights into the mechanisms by which action potential characteristics control the stability of electrical propagation (Weiss et al. 2006). Their limitation is, however, that phenomenologically represented parameters do not directly correspond to actual molecular structures or processes, and are thus incapable of accounting for many potentially arrhythmogenic mechanisms, such as propagation instabilities induced by instabilities in calcium cycling or by the altered metabolic state of the cell.
Clearly, the ultimate goal of modelling is to accurately represent the interplay between the subsystems responsible for the primary functions of the heart, including the electrophysiological, $\text{Ca}^{2+}$ handling, contractile and energetic components. This necessitates a significant level of complexity of the cell models, often achieved by implementing highly nonlinear control systems. Moreover, the ability to use simulations to gain a better understanding of cardiac pathophysiology requires a representation of the bidirectional feedback loops connecting the various subsystems. This presents additional computational concerns with respect to the coupling of processes that span very different temporal and spatial scales, described by equations with varying numerical stiffness.

A case in point is a recently described cell model, which combines mitochondrial bioenergetics with previously developed electrophysiological, $\text{Ca}^{2+}$ handling and contractile models, referred to as the excitation–contraction coupling and mitochondrial energetics (ECME) model (Cortassa et al. 2006). This cell model was developed as a framework for studying how the failure of the mitochondrial network of the cardiomyocyte can lead to action potential shortening or complete electrical inexcitability as a result of the depletion of the high-energy phosphate pool and the activation of the ATP-sensitive $K^+$ channels in the sarcolemma. Through a mechanism of reactive oxygen species (ROS)-induced ROS release (Cortassa et al. 2004), it was shown that the uncoupling of oxidative phosphorylation can induce oscillations of background $K^+$ currents (Aon et al. 2003) that were proposed to contribute to electrical heterogeneity and arrhythmias in whole hearts exposed to ischaemia-reperfusion (Akar et al. 2005). Compounds that could inhibit mitochondrial depolarization also prevented the post-ischaemic arrhythmias, demonstrating that a failure at the level of the mitochondrial scales to cause macroscopic desynchronization of cardiac electrical propagation. Thus, there is a strong motivation for developing multiscale models to study excitation–contraction–bioenergetic coupling, spanning from the organelle to the entire organ. Because the ECME model takes into account both fast Markovian ion channel gating kinetics (microsecond relaxation times) and slow enzymatic reactions associated with the tricarboxylic acid cycle (second-long relaxation times), an ordinary differential equation (ODE) solver, which explicitly accounts for the wide range of time-scales involved, would be ideally suited to speed up computational time. The technical challenges of scaling up a model with a temporal range spanning six orders of magnitude are formidable.

The goals of this article are (i) to review the techniques that are currently employed to model cardiac electrical activity in the heart, discussing the strengths and weaknesses of the various approaches, and (ii) to implement a novel modelling approach, based on physiological reasoning rather than on mathematical rigour, that lifts some of the restrictions imposed by ionic models of the most current generation, such as the ECME model. This article thus presents an efficient multiscale model of ventricular arrhythmogenesis that incorporates mitochondrial ionic channels and mitochondrial energetics. Achieving sufficient computational efficiency is the key to success in the quest to develop multiscale realistic models that are expected to lead to better understanding of the mechanisms of arrhythmia induction following failure at the organelle level and, ultimately, to the development of novel therapeutic applications.
2. Review of tissue- and organ-level modelling techniques

(a) The bidomain equations

Besides a few exceptions (Wang et al. 1996), the cardiac bidomain equations, or a simplified version referred to as the monodomain equations, are typically used to model cardiac bioelectric phenomena at the tissue and organ level. The bidomain equations (Plonsey 1988) represent the homogenization of cardiac tissue, replacing discrete components of the intracellular and extracellular tissue matrix, such as cells and gap junctions, with a continuum. Each point in space is considered to exist in two domains, the intra- and extracellular; the domains are superimposed in space and separated from each other by the cellular membrane.

The following equations relate the intracellular potential $\phi_i$ to the extracellular potential $\phi_e$, through the transmembrane current density $I_m$:

$$\nabla \cdot \sigma_i \nabla \phi_i = \beta I_m, \quad (2.1)$$

$$\nabla \cdot \sigma_e \nabla \phi_e = -\beta I_m, \quad (2.2)$$

$$I_m = C_m \frac{\partial V_m}{\partial t} + I_{ion}(V_m, \nu) - I_{trans}, \quad (2.3)$$

$$V_m = \phi_i - \phi_e, \quad (2.4)$$

where $\sigma_i$ and $\sigma_e$ are the intracellular and extracellular conductivity tensors, respectively; $\beta$ is the surface-to-volume ratio of the cardiac cells; $I_{trans}$ is the current density of the transmembrane stimulus; $C_m$ is the membrane capacitance per unit area; $V_m$ is the transmembrane voltage, defined as $\phi_i - \phi_e$; and $I_{ion}$ is the current density of the ionic current, which depends on the transmembrane voltage and other state variables represented by $\nu$.

The bidomain conductivity tensors $\sigma_i$ and $\sigma_e$ are the result of a homogenization procedure (Henriquez 1993), through which the discrete nature of the tissue matrix is translated into a macroscopic representation of the conductive tissue properties. The conductivity tensors account, in an averaged macroscopic sense, for the observation that conduction velocity is faster along the direction of the fibre orientation, and slower in a direction transverse to it. In the intracellular space, this is a reflection of the discrete coupling network, where cell-to-cell connections (gap junctions) are found most frequently at the ends of a cell (along the long axis of the cell). Furthermore, in both intracellular and extracellular spaces, a preferred direction of conduction is a consequence of the cell geometry, which constrains current flow in both spaces. This property is referred to as anisotropy. The anisotropy ratios between the two domains have been shown to be unequal (Clerc 1976; Roberts et al. 1979; Roberts & Scher 1982; Roth 1997). This property has been demonstrated to be of major importance in the interaction between cardiac tissue and externally applied current (Sepulveda et al. 1989).

Different approaches to recast the bidomain equations have been suggested (Hooke et al. 1994). A particularly popular linear transformation is to add equations (2.1) and (2.2) and to use the definition of $V_m$ according to equation (2.4) to retain $\phi_e$ and $V_m$, the two experimentally measurable quantities of
interest, as the independent variables (Pollard et al. 1992):

\[
\nabla \cdot (\sigma_i + \sigma_e) \nabla V_e = -\nabla \cdot \sigma_i \nabla V_m - I_e, \tag{2.5}
\]

\[
\nabla \cdot \sigma_i \nabla V_m = -\nabla \cdot \sigma_i \nabla V_e + \beta I_m. \tag{2.6}
\]

In the most general case, when the heart is immersed in a conductive fluid (e.g. a tissue bath in an experimental context or a surrounding torso to model in vivo scenarios), a Poisson problem has to be additionally solved,

\[
\nabla \cdot \sigma_b \nabla \phi_e = I_e, \tag{2.7}
\]

where \(\sigma_b\) is the conductivity of the bath medium; and \(I_e\) is the volume density of the extracellular stimulus current, which is injected or withdrawn through electrodes located in the bath.

Boundary conditions at the tissue–bath interface enforce continuity of the normal component of the extracellular current and continuity of \(\phi_e\), whereas the intracellular domain is isolated there (the normal component of the intracellular current is zero). At the boundaries of the bath, the normal component of the extracellular current is zero. Without enforcing Dirichlet boundary conditions, the elliptic partial differential equation (PDE) components, i.e. equation (2.7), and the elliptic portion of the bidomain equations, i.e. equation (2.5), are singular, which may pose numerical difficulties. Typically, grounding electrodes are used in the extracellular domain, i.e. nodes in the mesh are chosen where \(\phi_e\) is fixed at zero value, to circumvent this problem. For most scenarios of practical interest this is not a limitation since the validation of simulation results with experimental data requires the use of grounding electrodes to match an experimental set-up. With certain iterative solver techniques, for instance Krylov subspace methods such as conjugate gradients (CG), this is not necessarily required (Potse et al. 2006).

(b) Discretization schemes and issues

Different spatial discretization techniques, such as the finite-difference method (FDM; Skouibine et al. 2000), the finite-volume method (FVM; Harrild & Henriquez 1997; Trew et al. 2005), the interconnected cable model (Leon & Roberge 1990), and the finite-element method (FEM; Rogers & McCulloch 1994; Vigmond et al. 2002; Rodriguez et al. 2005; Ashihara et al. 2008), have been applied to the cardiac bidomain problem. Spatial discretization is mainly governed by the spatial extent of the depolarization wavefront. Assuming that the upstroke of the action potential lasts for approximately 1 ms and that under physiological conditions the wavefront travels at a speed between 0.2 (transverse conduction) and 0.7 m s\(^{-1}\) (longitudinal conduction), the spatial extent of the wavefront is in the range of 0.2–0.7 mm. The grid resolution has to be smaller than the extent of the wavefront, otherwise spatial undersampling effects will influence the simulation results. In the following sections, ‘fine discretization’ will be used to refer to discretizations where the spatial resolution is well below the spatial extent of the wavefront, and ‘coarse’ if the resolution is well above it. A discretization step of 250 \(\mu\)m could be considered...
as the limit between the two; however, it is not strict. Under pathological conditions, for instance, when a very slow decremental conduction is modelled, even a resolution of 100 µm may be too coarse.

A more formal argument to estimate spatio-temporal discretization constraints is found by considering the Courant–Friedrichs–Lewy (CFL) condition (Courant et al. 1928). Briefly, assuming straight fibres and equal anisotropy ratio, the CFL condition, which governs the maximum possible time-step for stable integration using the forward Euler method, is approximated by

\[
\Delta t \leq \frac{\beta C_m \Delta x^2}{2(\sigma_i + \sigma_t)},
\]

(2.8)

where \( \sigma_i = \sigma_i \sigma_{es}/(\sigma_i + \sigma_{es}) \) with \( * = l, t \). Assuming a standard value of 1400 cm\(^{-1}\) for \( \beta \) and the conductivity values as published by Clerc (1976), the CFL condition predicts a maximum possible time-step of 45.5 µs for a 100 µm grid. For a 10 µm grid \( \Delta t \) decreases to 0.45 µs, which renders the forward Euler integration for the parabolic PDE computationally inefficient compared with an implicit scheme such as the Crank–Nicholson scheme discussed below.

Numerically, the bidomain equations can be solved as a coupled system (Vigmond et al. 2002), where transmembrane voltage and extracellular potential are solved for simultaneously. Alternatively, operator-splitting techniques are applied (Keener & Bogar 1998) to decouple the computing scheme into three components, such as an elliptic PDE, a parabolic PDE and a set of nonlinear ODEs (equations (2.9)–(2.12)). It has been shown that the decoupled scheme converges quickly against the coupled scheme by employing a block Gauss–Seidel iteration (Pennacchio & Simoncini 2002). However, in a number of studies, the components were essentially treated as independent (Weber dos Santos et al. 2004a, b; Sundnes et al. 2005; Austin et al. 2006; Plank et al. 2007); solutions were then found by leapfrogging between the decoupled components, where either \( V_m \) in equation (2.5) or \( \phi_e \) in equation (2.6) are considered as constant. In Vigmond et al. (2002), it was found that, with small error tolerance, the difference between coupled and decoupled approaches is negligible.

Discretizing the decoupled bidomain equations leads to a three-step scheme, which involves the solution of the parabolic PDE, the elliptic PDE and the nonlinear system of ODEs at each time-step. In the simplest case, both the parabolic PDE and the nonlinear ODE systems can be solved with an explicit forward Euler scheme (Vigmond et al. 2002),

\[
V^{k+1} = V^k + \Delta t A_i (V^{k+1} - V^k) + \Delta t A_e \phi_e^k,
\]

(2.9)

\[
V^{k+1} = V^k + \Delta t I_{ion}(V^{k+1}, \eta^k),
\]

(2.10)

\[
\eta^{k+1} = \eta^k + \Delta t g(V^{k+1}, \eta^k),
\]

(2.11)

\[
(A_i + A_e) \phi_e^{k+1} = A_i V^{k+1} + I_e,
\]

(2.12)

where \( A_i \) is the discretized \(-\nabla \cdot (\sigma_i \nabla)/(\beta C_m)\) operator, with \( \xi \) being either \( i \) (intracellular) or \( e \) (extracellular); \( \Delta t \) is the time-step; and \( V^k, \phi_e^k \) and \( \eta^k \) are the temporal discretizations of \( V_m, \phi_e \) and \( \eta \), respectively, at the time instant of \( k \Delta t \).
When the computational mesh is of fine discretization and thus the integration time-step of the parabolic PDE has to be reduced substantially to maintain stability, it is advantageous to employ the more expensive semi-implicit Crank–Nicholson scheme instead of the forward Euler scheme to solve the parabolic PDE, with

$$\left[1 + \frac{1}{2} \Delta t A_i\right] V^{k+1} = \left[1 - \frac{1}{2} \Delta t A_i\right] V^k - \Delta t A_e \phi^k + I_{\text{ion}}(V_m^k, \eta^k),$$

(2.13)

replacing equations (2.9) and (2.10).

(i) Linear solver strategies

The PDEs are solved most efficiently with direct methods; however, such methods are limited to small grids only (Vigmond et al. 2002; Plank et al. 2007) because otherwise memory demands increase quickly, which, in turn, significantly increases the required number of operations per solver step. Although direct methods have been implemented to run in parallel environments (Amestoy et al. 2001; Li & Demmel 2003), typically they are harder to parallelize due to the fine-grained parallelism, which is communication intense. For large systems, iterative methods are mandatory because the memory requirements increase as $O(N^{5/3})$ and the complexity, in terms of required operations, increases as $O(N^{7/3})$, whereas the requirements for iterative solvers are $\approx O(N)$ and $\approx O(N^{5/3}\ln(\epsilon^{-1}))$, respectively, where $N$ is the system size and $\epsilon$ is the error tolerance (Meurant 1999). In general, linear systems with more than a few hundreds of thousands of unknowns can be considered as large. For instance, in Plank et al. (2007) a rabbit ventricular geometry immersed in a bath was discretized, resulting in 0.87 million unknowns. The linear system associated with this set-up could not be factorized on a desktop computer with 8 GB of RAM available.

When executing bidomain simulations on sequential computers, the main computational burden can be attributed to the solution of the elliptic problem and the set of ODEs (i.e. the ionic model). The complexity of an ionic model is determined by the number of state variables in the model and the stiffness of the ODE system. With simple ionic models, among which are phenomenological models such as FitzHugh–Nagumo (FitzHugh 1969) and its derivatives as well as electrophysiology-based (Hodgkin–Huxley-type) ionic models with a small number of state variables and without or with limited intracellular compartments and fluxes between them (Drouhard & Roberge 1987), the elliptic problem contributes more than 90 per cent to the overall workload. By contrast, for recent ionic models involving a large number (20–60) of state variables and very stiff ODEs as well as Markov state models (Iyer et al. 2004; Saucerman et al. 2004; Cortassa et al. 2006; Flaim et al. 2007; Mahajan et al. 2008), the ODE solution may even dominate the computations.

The parabolic problem is typically less of a concern. On coarser meshes, simple forward Euler steps can be employed to update $V_m$. In this case, the contributions of the diffusional component (i.e. the PDE) and the local membrane component to changes in $V_m$ can be updated separately, which renders the PDE linear. On finer grids, semi-implicit Crank–Nicholson schemes perform well. Even when relatively cheap iterative solvers are employed, the parabolic portion can be updated efficiently due to the diagonal dominance of the linear system.
For large systems, parallel computing approaches are necessary to reduce execution times. The parallel computing context alleviates the problem of solving the set of ODEs. State variables in an ionic model do not diffuse, which qualifies the ODEs as an embarrassingly parallel problem. No communication between processors is required to update the state variable and thus the parallel scaling of the ODE portion is linear. The parabolic problem is efficiently solved in parallel as well. Either only a forward Euler step is required (essentially a matrix–vector product, for which good scalability is expected), or the well-posed diagonally dominant linear system is solved efficiently with relatively cheap iterative methods, such as preconditioned CG. Typically, with an incomplete LU (ILU) preconditioner for the iterative CG solver, the parabolic problem can be solved in less than 10 iterations.

The elliptic PDE is the most challenging problem. Standard iterative solvers such as ILU–CG typically require several hundreds of iterations to converge (Plank et al. 2007), which makes this solution significantly more expensive than that of the parabolic system, although both systems share the same sparsity pattern. The parallel scaling of standard iterative solvers is fairly good (Plank et al. 2007); for instance, a parallel ILU–CG solver, where the system is decomposed by a block Jacobi preconditioner with ILU(0), i.e. an incomplete LU factorization with zero fill-in levels that preserves the sparsity pattern of the original matrix, used as a sub-block preconditioner, exhibits good parallel scaling. An example of good parallel scaling is presented in Plank et al. (2007), where parallel efficiency of approximately 92 per cent was achieved over the test range of processor numbers (up to 64 processors). With fewer number of processors, ILU(N) with N levels of fill-in tends to be more efficient; however, with an increasing number of processors, the efficiency of the preconditioning deteriorates since the preconditioner is applied only to the main diagonal block. This deterioration can be avoided by employing overlapping block preconditioners, such as alternating Schwarz methods; however, this increases the burden on the network interconnect owing to the increase in communication caused by the increased number of off-processor elements involved (Toselli & Widlund 2005). If the hardware lacks low-latency interconnects, this is undesirable since parallel efficiency levels off quickly with increasing number of processors.

It has been demonstrated in several recent studies (Weber dos Santos et al. 2004a,b; Austin et al. 2006; Plank et al. 2007) that multilevel preconditioners for CG methods both significantly improve the overall performance and exhibit reasonable parallel efficiency (better than 80%) for up to 128 processors. A generally applicable algebraic multigrid preconditioner (AMG) in conjunction with an iterative Krylov solver reduces the number of iterations per solver step by almost two orders of magnitude compared with ILU–CG (Plank et al. 2007). Although a single iteration with AMG is significantly more expensive than with ILU, the reduction in the number of iterations clearly favours a multilevel approach. For instance, in Plank et al. (2007) a speedup of 6 was reported. Using AMG–CG constitutes to date the most efficient method for solving the elliptic portion of the bidomain equations. Besides its computational efficiency, a major advantage of the AMG approach is that it is applied straightforwardly to unstructured grids, avoiding the complex task of explicitly generating nested representations of the geometry at different levels of resolution, as required by the geometric versions of multilevel preconditioners. Unstructured grids allow
smooth representations of the organ boundaries, which is of particular importance in simulations of cardiac defibrillation, where geometric discretizations with jagged boundaries inevitably evoke artificial surface polarizations. A more extensive review of numerical techniques to solve the cardiac bidomain equations can be found in Vigmond et al. (2008).

(c) Adaptivity and domain decomposition techniques

Substantial gains in computational efficiency can be achieved by explicitly taking into account the fact that high spatial and temporal resolution is needed only in the immediate vicinity of a propagating wavefront, where fast transitions in time are translated into steep gradients in space. Typically, steep wavefronts encompass a small portion of the overall domain, although this portion increases substantially when re-entrant activation patterns, particularly fibrillatory activity, are being modelled. Spatially adaptive methods that dynamically adjust grid resolution hold promise in increasing computational efficiency. Cherry et al. (2003) reported significant performance gains with an adaptive mesh refinement technique; however, the method, although very well suited for FDM and explicit stencils, is not easily modifiable to accommodate FEM or FVM discretizations and implicit stencils. A new class of methods loosely referred to as mortar FEM techniques (Pennacchio 2004) will possibly allow the use of FEM discretizations with spatial adaptivity; however, implementations using unstructured grids have not been reported yet.

Alternatively, simpler domain composition methods can be employed, where the region of interest is divided into several subdomains; each subdomain is then integrated at a different rate, the latter being a function of the ongoing activity in the domain itself and in the adjacent domains. However, such techniques introduce a significant bookkeeping overhead. Whether the reduction in floating point operations compensates for these extra costs, and to what degree, remains inconclusive. For instance, in Quan et al. (1998) a speedup of 17 was reported, whereas in Vigmond & Leon (1999) a speedup of only 3 could be achieved. Using different temporal resolutions on a per-domain base inevitably leads to load balancing issues for codes executed in parallel. That is, even though in a particular domain computations are executed quickly, the slowest domain dominates the overall execution time. A possible but non-trivial solution to address this problem is to migrate vertices between parallel partitions, although to our knowledge there are no simulator codes available today that are capable of achieving this in a robust manner.

3. Review of ionic model computation techniques

(a) Integration in standard form

The dynamics of the processes that underlie the action potential in a cardiac cell are typically described by a set of ODEs. These are initial-value problems of the form

\[
\frac{d\mathbf{y}}{dt} = \mathbf{f}(\mathbf{y}, t), \quad \mathbf{y}(t = 0) = \mathbf{y}_0, \tag{3.1}
\]
where \( y \) is a vector and \( f \) is a vector-valued function. The solution to the above equation is typically found as

\[
y(t + dt) = y(t) + \int_{t}^{t+dt} f(y(\tau), \tau) d\tau.
\] (3.2)

Different approaches can be used to approximate the integral over the nonlinear vector function \( f \). Here, all the components of the vector \( y \) are updated simultaneously with all the components of \( f(y) \) integrated simultaneously. General-purpose packages incorporating a large set of standard integrators are currently available and allow a convenient integration in a vector form (Cohen & Hindmarsh 1996). Standard integration techniques include explicit (or forward) and implicit (or backward) methods. Explicit methods are popular since they are easy to implement; however, the order of this class of methods is one, which often results in insufficient accuracy. Approaches to overcome this weakness include the use of several previously computed solutions (multistep methods) or additional intermediate solutions in the interval \([t, t + dt] \) (Runge–Kutta methods) to update the current solution. The more sophisticated implicit backward methods such as backward differentiation formula (BDF) or implicit Runge–Kutta methods (e.g. Rosenbrock methods; Rosenbrock 1963) have superior stability properties and allow larger time-steps; however, they are computationally expensive and, in general, robust implementation of these methods is difficult. The use of BDF methods requires the solution of a linear system of equations iteratively, by either fixed-point methods or versions of the Newton–Raphson methods. In this case, Jacobian matrices have to be either analytically determined and repetitively evaluated, or numerically approximated (Hairer & Wanner 2004). Many currently used integrators incorporate advanced features such as variable time-stepping and error control, where a time-step is chosen such that the local error per step is below a prescribed tolerance level.

For the particular purpose of integrating ODEs representing the myocardial action potential, standard methods have not been established and many different techniques are currently employed. Overall, although it is generally believed that implicit methods, owing to the stiffness of membrane kinetic variables, are advantageous and should be preferred over forward methods, the vast majority of implementations rely on forward methods.

**Component-wise integration**

Although the use of standard integrators is convenient for single-cell action potential simulations, it becomes cumbersome when ionic model integrators are coupled with PDE solvers for calculation of the potentials at the tissue and organ level. Therefore, as described in Maclachlan et al. (2007), in the field of cardiac modelling, it is common to split the vector formulation in equation (3.1) into a set of equations, where each ODE is integrated separately. In this component form, a system of coupled ODEs with \( N \) state variables decouples into \( N \) ODEs of the form

\[
\frac{dy_i}{dt} = f_i(y, t),
\] (3.3)
where \( 0 \leq i < N \) and \( f_i \) refers to the component \( i \) of the vector function \( f(y) \) (Maclachlan et al. 2007). Any standard method discussed above is well suited for component-wise integration as well. The application of implicit BDF methods is easier with component-wise integration since only scalar Newton iterations are required.

Possibly the most popular component-wise integration approach was proposed by Rush & Larsen (1978), and it has been widely applied (Vigmond & Leon 1999; Vigmond et al. 2002; Ten Tusscher & Panfilov 2006b) ever since. When the set of ODEs is decoupled and integrated sequentially in a component form, each state variable \( y_i \) is updated, with all other state variables \( y_j (j \neq i) \) held constant. Many of the ODEs comprising an ionic model, typically all gating equations in Hodgkin–Huxley-type models, but also ODEs in Markov state formulations, can be written in the form

\[
\frac{dy_i}{dt} = A(y) y_i + B(y),
\]

where all variables belonging to the subvector \( \bar{y} \) are considered constant during an integration step. Hence, the nonlinear ODE reduces to a linear ODE with constant coefficients, for which an analytic solution can be found. With \( A \) and \( B \) being constant, the analytical solution is

\[
y_i(t + \Delta t) = \left( y_i(t) + \frac{A}{B} \right) e^{A \Delta t} - \frac{A}{B}.
\]

This analytical solution is then used to compute the solution at the next time-step, \( t + \Delta t \). Numerical analysis has revealed that the Rush–Larsen method offers significant stability and accuracy benefits over the forward Euler method (Maclachlan et al. 2007). Any portions \( f_i \) of the vector function \( f(y, t) \) that cannot be written in a linear form need to be integrated in the fully nonlinear form. In the original Rush–Larsen formulation, the forward Euler method was applied to solve the fully nonlinear form. Therefore, the Rush–Larsen method has been classified as a non-standard FDM with forward Euler (NSFD w/FE; Maclachlan et al. 2007). Several suggestions have been made to further improve the method and overcome its weaknesses (Maclachlan et al. 2007). First, stability issues due to the use of the forward Euler method for the nonlinear components could be addressed by the use of implicit methods. Second, the accuracy of the method, originally first order, could be improved to second by a more expensive scheme involving the computation of midpoints. Although it has been shown that such elaborate approaches yield additional benefits for some ionic models, for instance the Courtemanche model of the human atrial cell (Courtemanche et al. 1998), where stiffness properties are less critical, no performance benefits or even worse performance has been observed with models that have pronounced stiffness properties, such as the canine ventricular ionic model by Winslow et al. (1999).

(c) Acceleration techniques

Regardless of the particular method, the expense in ionic model integration is in the evaluation of \( f_i(y, t) \), requiring the calculation of hundreds of computationally expensive expressions (mostly exponential and logarithmic functions). Many of these expressions depend on a single variable or they can be

\[Phil. Trans. R. Soc. A (2008)\]
recast into products of terms that depend on a single variable only. The physiological range of these variables is typically well known a priori, thus it is convenient to precompute and tabulate the expensive terms and then simply look up their values when required. For instance, assuming that in equation (3.5) the terms $A/B$ and $\exp(A\Delta t)$ depend only on one state variable $y_j$ and that no adaptive time-stepping scheme is employed, i.e. $\Delta t$ is constant throughout the simulation, then the functions $l_0(y_j) = A(y_j)/B(y_j)$ and $l_1(y_j) = \exp(A(y_j)\Delta t)$ can be precomputed and used for table lookup (TLU). As shown in Vigmond et al. (2003), recasting equation (3.5) leads to a more efficient TLU scheme,

$$y_i(t + \Delta t) = y_i(t) e^{A\Delta t} + \frac{A}{B}(e^{A\Delta t} - 1), \quad (3.6)$$

where one $y_j$-dependent lookup operation is required to determine the values of $l_0(y_j)$ and $l_1(y_j)$. The solution of the ODE can then be updated with one multiplication and one sum operation only (see algorithm 3.1).

**Algorithm 3.1.** NSFD w/FE integration of the linear part of the vector function $f$ using TLU.

```plaintext
for $k=0, \ldots, T_{\text{end}}/\Delta t$ do
  for $n=0, \ldots, N_{\text{cells}}$ do
    $l_0 = \text{TLU}(y_j)$;
    $y_i[n] = y_i[n] + 1[0] + 1[1]$;
  end for
end for
```

The TLU, in its most efficient form, requires one multiplication and one typecast operation only (nearest-neighbour TLU) or two multiplications and two sum operations (TLU with interpolation). Furthermore, if data structures are chosen properly for lookup tables and state variables, excellent cache performance can be achieved with this method.

(d) Tissue-level aspects of ionic model computation

On a current standard desktop computer, single-cell ionic model simulations, no matter how complex the particular ionic model is, can be executed faster than real time and memory requirements are negligible. With regard to organ-level simulations, however, real time is orders of magnitude too slow for a satisfactory performance. For instance, assuming that a human heart is sufficiently finely discretized by 10 million vertices, with real-time performance the simulation of one second of activity would last for approximately 116 days. Using supercomputers with tens or hundreds of CPUs scales the simulation time down to a few hours. However, considering that this is only the ODE part of a tissue-level simulation and that only a time interval of one second is simulated, it becomes apparent that, even on supercomputers, real-time performance is much too slow to allow meaningful computational studies, where a large parameter space is explored and where simulations of several seconds to minutes of electrophysiological activity are desired.
Clearly, performance plays a key role in tissue/organ simulations whereas it is of minor importance for single-cell simulations. Therefore, ionic model implementations used in tissue-/organ-level simulations are typically not straightforward extensions of single-cell codes. Ionic state variables in tissue models become vectors, memory demands increase, and the flexibility needed to allow implementation of spatial heterogeneities in model parameters or the use of different ionic models with different parameter settings in one organ model requires a careful software engineering design.

At the tissue/organ level, there are specific issues that prevent concepts successfully applied at the single-cell level from being carried over to tissue simulations without major adjustments. These issues can be summarized as follows.

(i) Organ-level simulations are almost exclusively carried out in parallel computing environments in order to achieve reasonable performance, whereas parallelization is pointless for single-cell codes. Since state variables do not diffuse, the solution of the system of ODEs qualifies as an embarrassingly parallel problem. That is, parallelization is easily achieved by simply partitioning the problem into evenly sized chunks, with no communication required. In tissue-level codes, the ODE solver is coupled to the solution of the PDEs, adding complexity. In the most complex case, where a bidomain formulation including a bath is discretized on an unstructured grid and where globally defined quantities other than $V_m$ (for instance, strain or interstitial concentrations) are linked to sets of different ODE systems representing different types of tissues (SA and AV node, Purkinje network, ventricular tissue, with spatial heterogeneities in transmural and apicobasal directions), the partitioning of the ODE system and its interaction with the global quantities becomes a truly challenging problem. Addressing parallelization and domain decomposition issues will become even more important in the near future because any standard desktop computer will be essentially a parallel computer with several computational cores. Approaches that do not address these issues will not benefit from future generations of computers and thus will be of very limited use.

(ii) In single-cell simulations, computational efficiency can be improved by allowing adaptivity in time-stepping, with $dt$ being a function of $y_i$ or $dy_i$. In tissue-level simulations this becomes problematic because ODE time-stepping has to be synchronized with the update intervals of the PDEs. Similar to the approach of solving bidomain equations in the decoupled form, where the elliptic portion is solved and hence the extracellular potential $\phi_e$ is updated at a slower rate than the parabolic portion updating $V_m$, the ODEs can be updated at a different rate than the PDEs. This is likely to work for small time-steps, with the ODE time-step chosen to be smaller than that of the PDE. Such an approach is advantageous if the stiffness associated with the fast-changing internal state variables enforces a time-step well below the time-step limit imposed by the mesh ratio of the parabolic portion of the bidomain equations. Furthermore, for time-adaptive ODE methods, the solver time depends largely on the current state of the ionic model. Therefore, solution time is larger in
portions of the tissue grid where activation is ongoing, and shorter in portions where the tissue is quiescent or at plateau/repolarization. In tissue simulations, particularly when re-entrant phenomena are simulated, all phases of the action potential coexist at any time. Therefore, adaptive ODE time-stepping introduces a load imbalance in parallel codes, where partitions with the fastest processes and thus the smallest time-steps become the limiting factor preventing linear parallel scaling of the ODE system.

(iii) For current computer architectures, cache performance is the key factor. Efforts to reduce the overall number of flops are not nearly as efficient as the careful layout of data structures, which increases the cache performance of the code. For single-cell simulators, code efficiency is of minor importance. No matter how complex an ionic model is, it will always fit into cache.

(iv) Single-cell codes aim to increase the maximum time-step while preserving stability and accuracy. Requirements are, however, very different in tissue-/organ-level simulations. The maximum time-step that allows stable integration could be limited by the mesh ratio and not the ODEs. For instance, using the NSFD w/FE technique, the Courtemanche model of the human atrial action potential (Courtemanche et al. 1998) can be integrated with a time-step of 400 \( \text{ms} \). However, when incorporated in a tissue model of a fine discretization, the time-step has to be reduced well below this number; the particular value depends on the method used to solve the parabolic PDE (Weber dos Santos et al. 2004a, b). That is, the potential of expensive ODE integration schemes, which allow large time-steps, cannot be fully exploited since PDE time-step constraints limit \( \Delta t \) to values for which cheaper ODE integrators perform well without any stability problems. On the other hand, it has been demonstrated that methods that increase the stability of an integration scheme allow for larger time-steps; however, they lead to increasingly inaccurate solutions with increasing \( \Delta t \). For instance, in Whiteley (2006, 2007) tissue simulations were conducted with time-steps from 100 to 500 \( \mu\text{s} \). Although the numerical scheme was stable, oscillations were observed immediately after the action potential upstroke (Whiteley (2006); figure 1c, d). Such oscillations may lead to spurious transients in internal state variables, which, in turn, will be reflected in \( V_m \). For instance, owing to the tight coupling between \( V_m \) and intracellular calcium handling, oscillations in \( V_m \) at the end of an action potential upstroke would give rise to artificial calcium transients; these are purely numerical artefacts and are not related to any intrinsic properties of the model. We demonstrate this in figure 1. Hence, such large time-steps have to be used with great care. In Whiteley (2006), the largest time-step used in a fully implicit ODE solver that resulted in reasonably accurate solutions (a small artefactual ripple in \( V_m \) during early repolarization was visible, though) turned out to be 100 \( \mu\text{s} \). The same simulation can be carried out with the NSFD w/FE method with the same time-step without any stability problems. In addition, a single NSFD w/FE step implemented efficiently could take only a fraction of the execution time required for an implicit BDF solver step, thus...
allowing one to execute tens to hundreds of NSFD w/FE solver steps within the time-frame required for a single BDF step. Thus, it can be concluded that, in most cases of practical interest, BDF methods are not likely to help in reducing the computational burden.

4. A new modelling paradigm for complex ionic models based on temporal multiscale decoupling

(a) The need for a new paradigm

Although computationally efficient implementations of versions of the NSFD w/FE method continue to be an excellent choice for many recent ionic models (Mahajan et al. 2008), for other recent models such as the ECME ionic model (Cortassa et al. 2006), their application becomes impracticable. For instance, with the fastest extended NSFD method implementation to date (Maclachlan et al. 2007), it took 25.77 s to execute a simulation of over 500 ms of activity in a single cell using the Winslow et al. (1999) canine ventricular model (no specific detail regarding the implementation was provided). The same study showed that fully implicit methods indeed allow much larger time-steps, but the overall performance was not better and the accuracy was strikingly inferior to that of NSFD w/FE.

The reason why some recent models have imposed a much higher burden on the ODE integrators than others stems from the current trend to develop more realistic (and hence, predictive) mechanistic formulations. This has led to the incorporation of processes whose dynamics occur on a faster time-scale and require a large number of state variables to faithfully reproduce experimentally observed phenomena. This is the case for formulations where Hodgkin–Huxley-type ionic channel representations have been replaced by Markov process descriptions (Clancy & Rudy 1999). Taking this approach has led to the
development of a series of integrative cardiac myocyte models that can explain the mechanisms underlying observations, such as action potential prolongation in heart failure (Winslow et al. 1999), the coexistence of high gain and graded sarcoplasmic reticulum Ca release (Greenstein & Winslow 2002), the role of beta-adrenergic stimulation on excitation–contraction coupling (Greenstein et al. 2004), and how properties of the cardiac dyad influence excitation–contraction coupling gain (Greenstein et al. 2006).

The wider range of time-scales captured in such ionic models increases the overall model stiffness primarily owing to the very fast time constants that govern intermediate states of these Markov state models. In a number of these ionic models (Jafri et al. 1998; Winslow et al. 1999; Cortassa et al. 2006; Greenstein et al. 2006), the fastest processes are linked to formulations describing the intrinsically fast dynamics of the L-type calcium channel and the ryanodine receptor, RyR, as well as the calcium dynamics in the dyadic space, i.e. the restricted space between the membrane of the T-tubules, where the L-type calcium channels reside predominantly, and the co-localized RyR in the membrane of the junctional sarcoplasmic reticulum. The positive feedback loop between the L-type calcium currents and the calcium release controlled by RyR is referred to as calcium-induced calcium release (CICR). Descriptions of CICR processes stiffen the ODE system considerably, as illustrated in figure 2. The time-scales governing CICR are often in the sub-microsecond range. Using the NFSD w/FE method enforces a time-step as small as 0.2 μs to successfully integrate these equations.

To avoid issues with ODE integration that are related to the stiffness of the system of equations resulting from detailed descriptions of CICR, a simplified phenomenological description could be used. Such descriptions employ a simple set of equations that reproduce the profile of a typical calcium-release signal with the time course of the release ‘hard coded’ into the equations. However, because CICR is the process that is at the very core of myocyte contractile function, and because understanding how alterations in CICR impact integrative function in both normal and diseased cells is a fundamental goal of building mathematical models, the use of phenomenological descriptions of CICR or any subcellular process hinders progress towards elucidating the integrative mechanisms at play in the cardiac myocyte. In order to be considered predictive, ionic models must be developed that accurately capture the underlying biophysical mechanisms of experimentally observed phenomena.

In the following sections, we present a novel ODE integration scheme that builds on the NSFD w/FE method but overcomes its limitations to improve computational efficiency and thus enable the use of ionic models such as the ECME model in tissue- and organ-level simulations. The needed gains in efficiency are obtained by taking advantage of two key observations.

(i) There is a wide range of time constants involved in the system of ODEs, ranging from the sub-microsecond range to hundreds of milliseconds (figure 2). Decoupling the ODE system into a suitable number of subsystems that are governed by similar time constants, and using different time-steps and/or different integrators for each subsystem could lead to substantial performance gains. We refer to this approach as temporal multiscale decoupling (TMSD).
A salient feature of ionic models such as the ECME model is the extreme stiffness of the system of ODEs, which is, however, only present during very brief phases of the action potential. By making simplifications to reduce this stiffness, the integrator can be relieved from having to track this extremely fast dynamics. We refer to this technique as dynamic reformulation of the ODEs.

(b) Formalization of the TMSD approach

In this section, we present the general TMSD approach, which we later apply to the ECME model to provide a specific illustration of its implementation. The essence of the approach is that fast state variables are considered constant as long as the difference in time-scales with the rest of the system is sufficiently large. In applying the temporal multiscale decoupling approach, state variables that act at a similar time-scale are grouped together, and blockwise updates of the different groups are performed using different update frequencies without any impact on the overall solution. However, special care is taken of equations where $f_i$ depends on fluxes. In this case, fluxes are integrated and adjusted to a particular update frequency when used outside of the group in which they were computed. Often such groups can be mapped to subsystems of the cell, for instance the mitochondria in the ECME model (Cortassa et al. 2006). Temporal changes of state variables in this subsystem occur at time-scales that are at least three orders of magnitude slower than those of the RyR subsystem and thus infrequent updates of the state variables in this subsystem can be executed without any loss of information. Since updates of slow variables require the same expensive evaluations of a group of functions $f_n$ of the vector function $f$, not updating all state variables at the same rate (figure 2) leads to significant savings in terms of execution time.

The wide range of time-scales involved in these ionic models can be exploited in a systematic manner to arrive at more efficient integration schemes, where different components of the state vector are decoupled according to a temporal multiscale modelling approach. In the simplest case, the state vector $y$ is split into $M$ portions and each partial state vector $y_m$ is integrated either with the same time-step $\Delta t$, when fixed time-step integration schemes are employed, or if an adaptive integration method with a varying internal time-step is used, the integrator output is synchronized at a desired $\Delta t$. With respect to the ODE integration frameworks presented in the review above, each subvector can then be integrated either in the standard form or component-wise in the non-standard form. Such decoupling has several benefits.

— Each group $y_m = f_m(y)$ is integrated with the method that is best suited. Stiff partial state vector systems with very fast components can be integrated with an expensive ODE solver, whereas partial state vectors with slow components can be integrated with much cheaper methods.
— When implicit backward integrators in the standard form are employed, the nonlinear systems that have to be solved are much smaller and thus easier to solve. For instance, solving the ODEs of the ECME model in the standard form involves the solution of a system with a sparse Jacobian of dimension
By decoupling and employing the method described here to the RyR subsystem, only a 4×4 system has to be solved. The evaluation of the Jacobian is much simpler in this case because evaluation of the portions of the vector function \( \mathbf{f} \) that are not pertinent to the subsystem \( \mathbf{f}_m \) can be skipped.

A formal description of the TMSD method for three groups is provided in algorithm 4.1. Each group is integrated using a different time-step with every outer time-step being a multiple of the time-step of the innermost loop.
Algorithm 4.1. TMSD integration of a system of ODEs using time-steps $\Delta t_0$, $\Delta t_1$ and $\Delta t_2$. State variables $\bar{y}$ are considered constant. Primed and double-primed fluxes $J'$ and $J''$ are used to update state variables with a different time-step in an outer loop.

\begin{verbatim}
for $k=0$, ..., $T_{End}/\Delta t_2$ do
  $J'_0 = 0$;
  $J''_0 = 0$;
  $t_2 = k\Delta t_2$;
  for $j=0$ to $\Delta t_2/\Delta t_1$ do
    $t_1 = t_2 + j\Delta t_1$;
    $J'_0 = 0.0$;
    for $i=0$, ..., $\Delta t_1/\Delta t_0$ do
      $t_0 = t_1 + i\Delta t_0$;
      $\dot{y}_0 = f_0(y_0(t_0), y_1(t_1), y_2(t_2))$;
      $J'_0 = (1/\Delta t_0) \int_{t_1}^{t_1+\Delta t_1} J_0(y_0(\tau_0), \bar{y}_1(t_1), \bar{y}_2(t_2)) d\tau_0$;
    end for
    $\dot{y}_1 = f_1(\bar{y}_0(t_0), y_1(t_1), y_2(t_2), J'_0)$;
    $J'_1 = (1/\Delta t_2) \int_{t_2}^{t_2+\Delta t_2} J_1(\bar{y}_0(t_0), y_1(\tau_1), \bar{y}_2(t_2)) d\tau_1$;
    $J''_0 = (1/\Delta t_2) \int_{t_2}^{t_2+\Delta t_2} J'_0(\tau_1) d\tau_1$;
  end for
  $\dot{y}_2 = f_2(\bar{y}_0(t_0), y_1(t_1), y_2(\tau_2), J''_0, J'_1) d\tau_2$;
end for
\end{verbatim}

(c) Dynamic reformulation of the ODEs

The main reason for the increased computational cost of the current state-of-the-art ionic models that use Markov chains is their stiffness and, to a much lesser extent, the large number of state variables. With Markov state models, the stiffness of the governing equations may vary substantially depending on the state vector $y$, since the Jacobian is not a constant but a function of $y$. That is, during a particular interval, a problem may be stiff, but it may be non-stiff during other intervals. For instance, with the ECME model, stiffness seems to be a problem only for a few microseconds during the onset and upstroke of the calcium transient in the dyad; other than that, the ECME model can be treated as other models are, with the minor difference that the number of state variables is two to three times higher than that of most standard ionic models. During the phases of pronounced stiffness, the integration time-step has to be reduced to values as small as 0.2 $\mu$s, even when expensive implicit ODE integrators are employed (Maclachlan et al. 2007).

In other words, this problem can be circumvented neither with more elaborate ODE integrators nor with the TMSD approach, since the inner loop of the TMSD integrator is tied to the same small time-step. Depending on the number of Markov state variables relative to the overall size of the set of ODEs, the inner
loop may impose a computational burden that renders organ-level applications computationally intractable or at least extremely expensive. In such cases, we suggest that the equations themselves be modified to reduce the stiffness of the system without changing the solution. This is again achieved by taking advantage of the separation of time-scales using the rationale as follows.

(i) Processes that are sufficiently faster than the minimum time-scale of interest can be assumed to occur instantaneously. In this case, an ODE $\frac{dy}{dt} = f(y, \bar{y}, t)$ can be replaced by an algebraic equation by assuming that $\frac{dy}{dt} = 0$, which yields $y = g(\bar{y}, t)$.

(ii) In a Markov state model, rapid equilibrium assumptions can be made if the transition between two connected states is substantially faster than other transition rates in the Markov network. For instance, in the Markov state model of RyR in the ECME model (Jafri et al. 1998; Cortassa et al. 2006), for a transition between the closed state $P_C$ and the open state $P_O$, (figure 3), one can assume that $k^+ \lbrack \text{Ca}^{2+}\rbrack_{ss} P_C = k^- P_O$, where $k^+$ and $k^-$ are rate constants and $[\text{Ca}^{2+}]_{ss}$ is the calcium concentration in the dyad, and use the combined state $P_C \cup P_O$ instead of the individual states $P_C$ and $P_O$. This approach has been used in a stochastic implementation of a Markov state RyR model (Greenstein & Winslow 2002), and in the development of a simplified (yet mechanistically detailed) model of CICR in the cardiac myocyte (Greenstein et al. 2006); however, unlike here, an integrator in standard form was used.


Figure 3. $[\text{Ca}^{2+}]_{ss}$-dependent toggling of RyR subsystem models. Depending on a threshold calcium level $[\text{Ca}^{2+}]_{ss,th}$ in the dyad, different formulations to describe the RyR subsystem were used. For low $[\text{Ca}^{2+}]_{ss}$ concentrations with $[\text{Ca}^{2+}]_{ss} < [\text{Ca}^{2+}]_{ss,th}$, when rate coefficients are slow, the full-blown formulation A is employed (a), while for $[\text{Ca}^{2+}]_{ss} \geq [\text{Ca}^{2+}]_{ss,th}$, the simpler formulation B is employed (b). Furthermore, while using formulation B, the ODE describing $[\text{Ca}^{2+}]_{ss}$ is replaced by an algebraic equation. (c) Different combinations of the presented formulations. Reference trace A (grey line) is generated using formulation A and the ODE to update $[\text{Ca}^{2+}]_{ss}$ over the entire cycle. For trace B (black dots), formulation A is initially employed, but replaced with B upon $[\text{Ca}^{2+}]_{ss} \geq [\text{Ca}^{2+}]_{ss,th}$. No noticeable differences are observed between these two traces. Trace B* (black line) is obtained employing the same methodology as in trace B, but an algebraic formulation is used during all phases of the cycle, including the initial phase of low $[\text{Ca}^{2+}]_{ss}$ level and relatively slow rate coefficients. In this case, noticeable deviations in the timing of $[\text{Ca}^{2+}]_{ss}$ are observed; however, the $[\text{Ca}^{2+}]_{ss}$ transient waveform and amplitude are similar with reference to trace A.
(d) Implementation of a TMMD ODE integrator with dynamic reformulation of ODEs for the ECME model

As part of this study, the suggested TMMD ODE numerical integration technique with dynamic reformulation of ODEs was implemented for the ECME model and validated against the well-established integrator used in the original paper (Cortassa et al. 2006); the latter relies on an implicit BDF method provided through the CVODE library (Cohen & Hindmarsh 1996). Our validated TMMD code used three different groups, which we integrated at time-steps of 2.5, 20 and 100 µs. The smallest time-step was applied to integrate the equations related to $[Ca^{2+}]_{ss}$ and the Markov state models representing RyR and the L-type calcium current. Standard Hodgkin–Huxley-type gating variables and other ionic fluxes were integrated at a time-step of 20 µs using the NSFD w/FE integrator. The very slow processes linked to mitochondrial metabolism and force generation were integrated with the largest time-step of 100 µs.

Besides using the TMMD approach, dynamic reformulation was applied to the equations for RyR and $[Ca^{2+}]_{ss}$, but not to the Markov state model representing the L-type calcium current. Details of the procedure are illustrated in figure 3.

Formulations for the RyR subsystem were switched based on the transition rates of the $[Ca^{2+}]_{ss}$ level. For $[Ca^{2+}]_{ss}$ below a threshold level, $[Ca^{2+}]_{ss,th}$, the full-blown Markov state model of the RyR (formulation A, figure 3a) was employed and the ODE that governs the calcium flux balance in this compartment was integrated to update $[Ca^{2+}]_{ss}$. Since there are no $[Ca^{2+}]_{ss}$ transients, large time-steps could be implemented to speed up execution time. Fast transients in the RyR states are triggered by the quick rise in $[Ca^{2+}]_{ss}$. As soon as $[Ca^{2+}]_{ss}$ rose above the threshold level $[Ca^{2+}]_{ss,th}$, rapid equilibrium assumptions were made, and a simplified model was employed to update the RyR states (formulation B, figure 3b). The combined state $P_{Oi} \cup P_{Ci}$ substituted the individual states $P_{Oi}$ and $P_{Ci}$, where the probability of $P_{Oi}$ relative to the new combined state was determined by $P(O_i|O_1 \cup C_1) = \alpha$ with $\alpha$ being determined from $k_a [Ca^{2+}]^{n}_{ss} P_{Ci} = k_f P_{Oi}$ and $P_{Oi} \cup P_{Ci} = P_{Oi} + P_{Ci}$, i.e. $P_{Oi} = \alpha P_{Oi} \cup P_{Ci}$. The threshold value for switching between the formulations was chosen in such a way that the rate coefficients governing the transitions between $P_{Ci}$ and $P_{Oi}$ were at least 10 times faster than all other processes. Specifically, we used a threshold $k_a [Ca^{2+}]^{n}_{ss} = 100$ m s$^{-1}$, which translated into a threshold concentration of $[Ca^{2+}]_{ss,th} = 0.095$ mM.

When switching to formulation B, further numerical stability benefits were gained by replacing, with an algebraic equation, the ODE that governs the $[Ca^{2+}]_{ss}$ transient as a function of the fluxes into and out of the dyad. This was accomplished by assuming that $[Ca^{2+}]_{ss}$ equilibrates rapidly, i.e. the $Ca^{2+}$ influx into the dyadic space and the $Ca^{2+}$ efflux out of it are equal. This is equivalent to assuming that $d[Ca^{2+}]_{ss}/dt = 0$ holds true, resulting in the reduction of the ODE governing the transients into an algebraic equation. This assumption is well justified because $Ca^{2+}$ diffusion out of the dyad occurs far more rapidly than the L-type Ca channel and RyR gating processes (Hinch et al. 2004; Greenstein et al. 2006). As evidenced by figure 3c, this modification does not lead to any differences in results (compare traces A and B in figure 3c), but allowed a much larger time-step, 2.5 µs. Otherwise, for stability reasons, the integrator was forced into a small time-step of 0.2 µs. However, for $[Ca^{2+}]_{ss}$ close to the base level, employing the algebraic equation to update $[Ca^{2+}]_{ss}$ results in small
deviations. In this case, the onset of the $[\text{Ca}^{2+}]_{\text{ss}}$ transient occurs approximately 5 μs earlier than that in the original model (compare traces A and B with B* in figure 3c).

Finally, longer single-cell simulations, over 1 min, were executed to ensure the long-term stability of the scheme. It was established, for all 50 state variables of the model, that the results were indiscernible from those obtained with the reference implementation.

5. Implementation of the new paradigm in organ-level simulations with the ECME model

The gain in execution speed obtained by using TMSD with the dynamic reformulation of several ODEs versus the classical NSFD w/FE scheme turned out to be at least two orders of magnitude. Although the exact speedup depended on several factors, such as platform, compiler, optimization and implementation details, differences in execution time between a straightforward NSFD w/FE implementation of the ECME model, using an integration time-step of 0.2 μs, and the TMSD integrator with dynamic reformulation reached a factor of up to 285. This is mainly attributed to gains in the maximum stable time-step size that can be used to integrate the RyR state equations, while less frequent updates of state variables related to the metabolic state and force generation helped to reduce execution times further by reducing the number of expensive evaluations of the function $f$.

To assess to what degree the gain in efficiency in the ECME model was sufficient to allow simulations at the tissue and organ levels, we carried out benchmark simulations in tissue consisting of one million cells. The benchmark was executed without coupling the cells, i.e. the parabolic PDE was not solved and no spatial domain was discretized. This problem size was chosen (i) by considering that one million degrees of freedom suffice to discretize any small mammalian heart up to the size of a rabbit at a sufficiently fine spatial resolution (approx. 200 μm) and (ii) to ensure that realistic organ-level simulation is tested where cache performance will play an important role.

Execution times to simulate one second of activity were measured on a standard desktop computer with four CPUs for the ECME model and several other widely used models, from a simple ventricular model with seven state variables, the MBRDR model (Drouhard & Roberge 1987), through the Courtemanche model of the human atrial action potential (Courtemanche et al. 1998) and the Luo–Rudy guinea-pig ventricular model in its latest revision (Faber & Rudy 2000; Luo & Rudy 2006), to the most recent Markov state model of the rabbit ventricular cell by Mahajan et al. (2008). For all models other than the ECME, we used the NSFD w/FE method with the acceleration techniques described in the review portion of this article. Although some of these models could be integrated with significantly larger time-steps in a single-cell mode, all models were integrated with a time-step of 20 μs, assuming that this value is a constraint imposed by the CFL conditions of the parabolic PDE.

Benchmark results are summarized in table 1. Even with all the optimizations implemented, the ECME model is, by a wide margin, computationally significantly more expensive than all the other models in this benchmark.
Nonetheless, the performance is sufficient to investigate phenomena occurring on the time-scale of several seconds to minutes when simulations are executed on a supercomputer. Since the ODE computations scale linearly on a parallel computer, the time spent on solving the ODEs can be scaled down to execute the computations within a reasonable time. For instance, using 512 processors, the execution time will be reduced by a factor of 128 relative to this benchmark. That is, in 2.34 min, 1 s of activity can be simulated; 1 min of activity would be executed in 140 min.

Finally, we coupled the ECME model with the implemented TMSD ODE integration scheme with dynamic reformulation of ODEs to the ECME model in an organ-level simulation. An apical pacing sequence was simulated in a monodomain guinea-pig-size ventricular model (approx. 75e^3 degrees of freedom) using a proof-of-concept implementation of a TMSD ODE integrator with dynamic reformulation. On a current standard desktop computer, the execution time was approximately 1e^3 times slower than real time (i.e. simulation of 1 ms of activity lasted for approx. 1 s). Shown are the epicardial maps of the distributions of (a) transmembrane voltage, $V_{m}$, (b) calcium concentration in the dyadic space, $[Ca^{2+}]_{ss}$, (c) mitochondrial membrane potential, $\Delta\psi$, and (d) the concentration of permissive tropomyosin with zero crossbridges, $P_0$. All variables are shown at 40 ms after the pacing pulse delivery, except for $P_0$, which is shown at a later instant, 70 ms.

Figure 4. Application of the TMSD ODE integrator with dynamic reformulation of ODEs to the ECME model in an organ-level simulation. An apical pacing sequence was simulated in a monodomain guinea-pig-size ventricular model (approx. 75e^3 degrees of freedom) using a proof-of-concept implementation of a TMSD ODE integrator with dynamic reformulation. On a current standard desktop computer, the execution time was approximately 1e^3 times slower than real time (i.e. simulation of 1 ms of activity lasted for approx. 1 s). Shown are the epicardial maps of the distributions of (a) transmembrane voltage, $V_{m}$, (b) calcium concentration in the dyadic space, $[Ca^{2+}]_{ss}$, (c) mitochondrial membrane potential, $\Delta\psi$, and (d) the concentration of permissive tropomyosin with zero crossbridges, $P_0$. All variables are shown at 40 ms after the pacing pulse delivery, except for $P_0$, which is shown at a later instant, 70 ms.

Nonetheless, the performance is sufficient to investigate phenomena occurring on the time-scale of several seconds to minutes when simulations are executed on a supercomputer. Since the ODE computations scale linearly on a parallel computer, the time spent on solving the ODEs can be scaled down to execute the computations within a reasonable time. For instance, using 512 processors, the execution time will be reduced by a factor of 128 relative to this benchmark. That is, in 2.34 min, 1 s of activity can be simulated; 1 min of activity would be executed in 140 min.

Finally, we coupled the ECME model with the implemented TMSD ODE integration scheme with dynamic reformulation to tissue-/organ-level bidomain codes to conduct true organ-level simulations. A guinea-pig-size ventricular geometry model, which represents the rabbit ventricular geometry used in our previous studies (Trayanova et al. 2002; Rodriguez et al. 2005; Ashihara et al. 2008), scaled half in size, was chosen to facilitate the execution of the test on a standard desktop computer with reasonable performance. The discretization of the myocardial volume at an average element size of approximately 200 μm led to approximately 75e^3 degrees of freedom. The guinea-pig ventricles were paced at the apex to initiate a simple activation sequence (figure 4). Using a current desktop computer (Dual Opteron X2), the simulation of 1 ms of activity

Table 1. Summary of execution times $T_{exec}$ in minutes, to simulate 1 s of activity in a tissue model with one million cells running on a standard desktop computer for various ionic models. (A global time-step of 20 μs was used for all models.)

<table>
<thead>
<tr>
<th>Model</th>
<th>$T_{exec}$/s (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MBRDR</td>
<td>11</td>
</tr>
<tr>
<td>Courtemanche</td>
<td>24</td>
</tr>
<tr>
<td>LRdII</td>
<td>54</td>
</tr>
<tr>
<td>Mahajan</td>
<td>125</td>
</tr>
<tr>
<td>ECME</td>
<td>300</td>
</tr>
</tbody>
</table>

Figure 4. Application of the TMSD ODE integrator with dynamic reformulation of ODEs to the ECME model in an organ-level simulation. An apical pacing sequence was simulated in a monodomain guinea-pig-size ventricular model (approx. 75e^3 degrees of freedom) using a proof-of-concept implementation of a TMSD ODE integrator with dynamic reformulation. On a current standard desktop computer, the execution time was approximately 1e^3 times slower than real time (i.e. simulation of 1 ms of activity lasted for approx. 1 s). Shown are the epicardial maps of the distributions of (a) transmembrane voltage, $V_{m}$, (b) calcium concentration in the dyadic space, $[Ca^{2+}]_{ss}$, (c) mitochondrial membrane potential, $\Delta\psi$, and (d) the concentration of permissive tropomyosin with zero crossbridges, $P_0$. All variables are shown at 40 ms after the pacing pulse delivery, except for $P_0$, which is shown at a later instant, 70 ms.

Nonetheless, the performance is sufficient to investigate phenomena occurring on the time-scale of several seconds to minutes when simulations are executed on a supercomputer. Since the ODE computations scale linearly on a parallel computer, the time spent on solving the ODEs can be scaled down to execute the computations within a reasonable time. For instance, using 512 processors, the execution time will be reduced by a factor of 128 relative to this benchmark. That is, in 2.34 min, 1 s of activity can be simulated; 1 min of activity would be executed in 140 min.

Finally, we coupled the ECME model with the implemented TMSD ODE integration scheme with dynamic reformulation to tissue-/organ-level bidomain codes to conduct true organ-level simulations. A guinea-pig-size ventricular geometry model, which represents the rabbit ventricular geometry used in our previous studies (Trayanova et al. 2002; Rodriguez et al. 2005; Ashihara et al. 2008), scaled half in size, was chosen to facilitate the execution of the test on a standard desktop computer with reasonable performance. The discretization of the myocardial volume at an average element size of approximately 200 μm led to approximately 75e^3 degrees of freedom. The guinea-pig ventricles were paced at the apex to initiate a simple activation sequence (figure 4). Using a current desktop computer (Dual Opteron X2), the simulation of 1 ms of activity
was executed in approximately 1.2 s for a monodomain formulation and in approximately 5 s for a bidomain formulation. That is, on a standard desktop computer in 1 hour execution time, approximately 3 s of activity could be simulated with a monodomain formulation and 0.72 s with a bidomain formulation. These numbers clearly demonstrate the potential of the new methods presented in this study. Again, when parallel computers are employed to execute simulations, much longer observation periods, of the order of tens of seconds to minutes, can be studied with ionic models as complex as the ECME model within a reasonable time.

6. Concluding remarks on the new paradigm

This article has presented a review of various numerical techniques to model electrical activity in the heart, spanning the spatial scales from subcellular mechanisms to the entire organ. Furthermore, a new version of the NSFD w/FE method was developed, which takes advantage of the very different time-scales of the processes represented in recent ionic models of very high complexity, such as the ECME model, to obtain substantial increases in computational efficiency. In Maclachlan et al. (2007), it was reported that the simulation of 500 ms activity in a single cell modelled by the Winslow et al. (1999) canine ventricular model lasted for 25 s. The ECME model used in this study employed the same set of equations to represent the L-type calcium current and RyR subsystem, which are by far the computationally most expensive components of the model. Although direct comparison is difficult, the ECME model can be judged as computationally more expensive than the canine model by Winslow et al. The performance we obtained in this study for the ECME model translates into 36 ms to simulate 500 ms activity in a single cell and, thus, is significantly more efficient than the standard NSFD w/FE method.

This is the first attempt at an implementation based on the TMSD ODE integration method with dynamic reformulation of ODEs. Although the performance is satisfactory and opens new avenues to employ highly detailed cell models in organ-level simulations, this is a first step only and further significant performance improvements can be expected. The choice of decomposing the set of ODEs into groups and the maximum possible time-step for each group cannot be determined automatically and needs manual tuning. In general, this is a significant disadvantage of the TMSD technique since intimate knowledge of the model is required to tune up the method for best performance. Decoupling an ionic model into different subsystems requires a non-trivial effort and careful examination of the maximum rates of change of different state variables. The effort to develop ionic models based on the suggested methods is further complicated by the fact that at least two implementations of an ionic model are required, a very accurate reference implementation such as an implicit BDF integrator in the standard form and the TMSD implementation itself. At the current stage of development, the method is sufficiently efficient to allow parameter sweeps in organ modelling studies of small mammalian hearts, even when high-performance computing (HPC) environments are not available. To simulate larger mammalian hearts or for questions that require the use of the bidomain formulation, the use of HPC hardware is mandatory. Today, with the
availability of HPC environments, there are virtually no limitations in using highly detailed Markov chain ionic models, and the testing of various simulation scenarios as well as explorations of parameter spaces indeed appear feasible.

This work was supported by Marie Curie Fellowship MC-OIF 040190 funded by the European Commission and SFB grant F3210-N18 from the Austrian Science Fund (FWF) to G.P.; NIH grants R37-HL54598, R33-HL87345 and P01-HL081427 to B.O’R.; NIH grants R33 HL87345, N01-HV28180, PO1 HL081427 and P01-HL077180 to R.L.W.; and NIH grants R01-HL063195, R01-HL082729 and R01-HL067322 and NSF grant CBET-0601935 to N.A.T.

References


*Phil. Trans. R. Soc. A* (2008)


*Phil. Trans. R. Soc. A* (2008)


