Anatomically accurate high resolution modeling of human whole heart electromechanics: A strongly scalable algebraic multigrid solver method for nonlinear deformation

Christoph M. Augustin a, Aurel Neic a, Manfred Liebmann b, Anton J. Prassl a, Steven A. Niederer c, Gundolf Haase b, Gernot Plank a,

a Institute of Biophysics, Medical University of Graz, Graz, Austria
b Institute for Mathematics and Scientific Computing, Karl-Franzens-University Graz, Graz, Austria
c Dept. Biomedical Engineering, Division of Imaging Sciences and Biomedical Engineering, King's College of London, London, United Kingdom

ARTICLE INFO

Article history:
Received 9 April 2015
Received in revised form 31 July 2015
Accepted 27 October 2015
Available online 2 November 2015

Keywords:
Cardiac electromechanics
Algebraic multigrid
Parallel computing
Whole heart model

ABSTRACT

Electromechanical (EM) models of the heart have been used successfully to study fundamental mechanisms underlying a heart beat in health and disease. However, in all modeling studies reported so far numerous simplifications were made in terms of representing biophysical details of cellular function and its heterogeneity, gross anatomy and tissue microstructure, as well as the bidirectional coupling between electrophysiology (EP) and tissue distension. One limiting factor is the employed spatial discretization methods which are not sufficiently flexible to accommodate complex geometries or resolve heterogeneities, but, even more importantly, the limited efficiency of the prevailing solver techniques which is not sufficiently scalable to deal with the incurring increase in degrees of freedom (DOF) when modeling cardiac electromechanics at high spatio-temporal resolution.

This study reports on the development of a novel methodology for solving the nonlinear equation of finite elasticity using human whole organ models of cardiac electromechanics, discretized at a high para-cellular resolution. Three patient-specific, anatomically accurate, whole heart EM models were reconstructed from magnetic resonance (MR) scans at resolutions of 220 μm, 440 μm and 880 μm, yielding meshes of approximately 184.6, 24.4 and 3.7 million tetrahedral elements and 95.9, 13.2 and 2.1 million displacement DOF, respectively. The same mesh was used for discretizing the governing equations of both electrophysiology (EP) and nonlinear elasticity. A novel algebraic multigrid (AMG) preconditioner for an iterative Krylov solver was developed to deal with the resulting computational load. The AMG preconditioner was designed under the primary objective of achieving favorable strong scaling characteristics for both setup and solution runtimes, as this is key for exploiting current high performance computing hardware.

Benchmark results using the 220 μm, 440 μm and 880 μm meshes demonstrate efficient scaling up to 1024, 4096 and 8192 compute cores which allowed the simulation of a single
1. Introduction

Electromechanical function of the heart emerges from a complex cascade of processes, which interact across a broad range of spatial and temporal scales. Despite the large body of experimental and theoretical research accumulated, our current understanding of how alterations in subcellular function affect whole organ pumping performance or, conversely, how changes at the systemic level influence subcellular function, remains incomplete. Computational modeling bears high promise as a methodology for integrating experimental data into a mechanistic framework which enables the quantitative observation of complex cause–effect relationships across a broad range of spatio-temporal scales.

While the potential gains from modeling approaches are significant, the challenges to be addressed are daunting. Historically, the development of EP and mechanical models of cardiac function proceeded rather independently than in tandem. The vast majority of EP modeling studies ignored any effects due to mechanical deformation, and, vice versa, most mechanical modeling studies did not represent explicitly any feedback of deformation on EP. While such a simplified EM coupling has proven suitable for addressing a variety of questions [85,53,60], the assumption of a unidirectional coupling between EP and mechanical deformation is inaccurate as there is strong evidence that EP is modulated by tissue distension via mechano–electric feedback (MEF) mechanisms [58,62,63]. However, from a model development point of view, it is attractive to assume that MEF can be neglected, as this allows to develop EP and mechanical models independently which reduces the complexity of implementation. This effective split into two separate sequentially executed solution steps is reflected in a notable divergence in the employed methodologies between EP and mechanics modeling communities.

State of the art organ scale EP finite element or finite volume models are discretized at a high spatio-temporal resolution using tetrahedral [112] or hybrid elements [100]. High resolutions are not only necessary to capture the fast transients and steep depolarization wavefronts, but also to resolve fine microscopic scale structural detail [93,122] as well as functional heterogeneities [78,18,57]. In contrast, mechanical models build on the assumption that cardiac deformation is governed by smoother spatial and slower temporal scales, suggesting that the use of much coarser spatio-temporal discretizations may yield sufficient accuracy. Thus the use of higher order cubic Hermite elements became popular [24] as they allow tessellation of stylized ventricular anatomy using a very small number of elements and they avoid volumetric locking, however, their major drawback is the limited capacity to accommodate complex geometries. This issue is being addressed now with the use of tetrahedral \( P_1 \rightarrow P_0, P_2 \rightarrow P_0 \) or mixed formulation \( P_2 \rightarrow P_1 \) elements, which have been introduced only very recently for modeling cardiac mechanics [43,42,32]. In general, finite element discretizations with tetrahedral and hexahedral elements of the same order showed similar results in terms of precision and efficiency [19].

Unsurprisingly, as a consequence the cost of model execution has been addressed quite differently as well. In EP modeling, due to a large number of DOF, a pressing need for strongly scalable iterative solvers has emerged [82,97], whereas for the lower dimensional mechanical models this has not been the case. While high resolution models and corresponding scalable solver algorithms have been developed for vascular applications [61,3], for cardiac mechanics direct solvers, executed on a very small number of compute cores [32], prevail.

The use of higher spatial resolutions and more flexible tetrahedral or hybrid finite element meshes is also driven by the needs of clinical modeling applications. Generation pipelines for individualized EM models rely increasingly on reconstruction from tomographic imaging [111] which provide an ever increasing level of anatomical detail [68]. Multimodal clinical imaging also facilitates more accurate tissue classification. Clinically important delineation of viable myocardium, fat deposits, cleft spaces and vascularization, as well as substrate abnormalities such as infarct scars [103] or the presence of fibrosis [76,77] is becoming feasible. These are known to be important factors influencing cardiac deformation, however, the spatial resolutions and methodology currently used impede their elucidation in modeling studies.

Spatial resolution also becomes an important factor when considering bidirectionally coupled, anatomically accurate EM models. A bidirectional link between EP and mechanics demands information being passed back and forth between EP and mechanics solver components. With tomographically reconstructed models the bidirectional projection of data may constitute a significant technical challenge when markedly different spatio-temporal resolutions are used. Either one resorts to using simplified, stylized anatomical representations for both physics to ensure an exact overlap of both domains, a refined electrical mesh is derived from spatial refinement of finite elements forming the mechanical mesh [15], or one opts to use the same high resolution mesh for both physics and accepts the increase in computational load. The advantage of the latter approach is that convergence studies with bidirectionally coupled EM models are feasible as discretization errors are governed by the spatio-temporal discretization of the EP problem. When attempting to build human whole organ EM models the use of high resolution meshes becomes particularly challenging due to the larger size of a human heart compared to other species used in many previous EM modeling studies [83], and the overall increase in DOF needed for discretizing both ventricles as well as the thin-walled atria [11].

This study reports on the development of a novel methodology for modeling of human whole organ cardiac electromechanics at a high para-cellular resolution. A clinical magnetic resonance (MR) scan was segmented to generate a
patient-specific, anatomically accurate whole heart EM model. Tetrahedral finite element meshes were generated from the segmented image stack at resolutions of 220 μm, 440 μm and 880 μm, resulting in meshes with approximately 184.6, 24.4 and 3.7 million tetrahedral elements and 95.9, 13.2 and 2.1 million displacement DOF, respectively. The same mesh was used for discretizing the governing equations of both electrophysiology (EP) and nonlinear elasticity. A novel algebraic multigrid (AMG) preconditioner for an iterative Krylov solver was developed to deal with the resulting computational load. The AMG preconditioner was designed under the primary objective of achieving favorable strong scaling characteristics for both setup and solution runtimes, as this is key for exploiting current high performance computing hardware.

Strong scaling benchmarks using the 220 μm, 440 μm and 880 μm meshes with 95.5, 13.2 and 2.1 million displacement DOF demonstrated efficient scaling up to 8192, 4096 and 1024 compute cores which allowed the simulation of a single heart beat in 235.3, 87.8 and 44.3 minutes, respectively. The efficiency of the method allows fast simulation cycles without compromising in terms of anatomical or biophysical detail, thus enabling modeling studies which elucidate the role of fine scale anatomical features, substrate heterogeneities or mechanistic inquiries into bidirectional feedback loops governing cardiac adaptation processes at a fundamental level. Moreover, meaningful verification of EM models becomes feasible as spatial resolutions can be refined sufficiently to gauge convergence errors.

2. Governing equations

2.1. Mechanical model

Cardiac tissue is mechanically characterized as a hyperelastic, nearly incompressible, orthotropic material with a non-linear stress–strain relationship. The deformation gradient \( \mathbf{F} \) describes the deformation \( \mathbf{U} \) of a body from the reference configuration \( \Omega_0(\mathbf{x}) \) to the current configuration \( \Omega(\mathbf{x}) \),

\[
F_{ij} = \frac{\partial x_i}{\partial X_j}, \quad i, j = 1, 2, 3.
\]

By convention, we denote \( J = \det \mathbf{F} > 0 \) and introduce the right Cauchy–Green tensor \( \mathbf{C} = \mathbf{F}^T \mathbf{F} \). The nearly incompressible behavior is modeled by a multiplicative decomposition of the deformation gradient, see, e.g., [31], of the form

\[
\mathbf{F} = J^{1/3} \mathbf{F}^*, \quad \mathbf{C} = J^{2/3} \mathbf{C}^*, \quad \text{with } \det \mathbf{F} = \det \mathbf{C} = 1,
\]

where \( J^{1/3} \) and \( J^{2/3} \) are associated with volumetric deformations, while \( \mathbf{F}^* \) and \( \mathbf{C}^* \) are associated with isochoric processes.

Mechanical deformation is governed by the stationary equilibrium equations given as

\[
- \text{Div} \{ \mathbf{FS(U,X)} \} = \mathbf{b}_0(\mathbf{x}) \quad \text{for } \mathbf{x} \in \Omega_0,
\]

where \( \mathbf{U}(\mathbf{x}, t) \) is the unknown deformation, \( \mathbf{S}(\mathbf{U}, \mathbf{X}) \) is the second Piola–Kirchhoff stress tensor, \( \mathbf{b}_0(\mathbf{x}) \) are the body forces and \( \text{Div} \) denotes the divergence operator in the reference configuration.

On the boundary \( \partial \Omega_0 = \Gamma_{0,D} \cup \Gamma_{0,N} \) Dirichlet displacement boundary conditions are imposed

\[
\mathbf{U}(\mathbf{x}) = \mathbf{U}_D(\mathbf{x}) \quad \text{on } \Gamma_{0,D},
\]

with \( \mathbf{U}_D(\mathbf{x}) \) being a prescribed displacement, and Neumann force boundary conditions

\[
\mathbf{FS(U,X)} \mathbf{N}_0(\mathbf{x}) = \mathbf{G}_0(\mathbf{x}) \quad \text{on } \Gamma_{0,N},
\]

where \( \mathbf{N}_0(\mathbf{x}) \) is the exterior normal vector and \( \mathbf{G}_0(\mathbf{x}) \) denotes a prescribed surface traction.

The deformation of cardiac tissue is governed by imposed external loads such as intracavitary and pericardial pressures, and active stresses intrinsically generated during contraction. The total stress \( \mathbf{S} \) is additively decomposed according to

\[
\mathbf{S} = \mathbf{S}_p + \mathbf{S}_a,
\]

where \( \mathbf{S}_p \) and \( \mathbf{S}_a \) refer to the passive and active stresses, respectively. Passive stresses are modeled based on the constitutive equation

\[
\mathbf{S}_p = 2 \frac{\partial \Psi(\mathbf{C})}{\partial \mathbf{C}},
\]

where \( \Psi \) is an orthotropic, invariant-based strain-energy function [49,30], which results in a frame-independent stress tensor. The strain-energy function \( \Psi \) is additively composed of three functions

\[
\Psi(\mathbf{C}) = \Psi_{vol}(J) + \Psi_{iso}(\mathbf{C}) + \Psi_{aniso}(\mathbf{C}, f_0, s_0),
\]

where \( \Psi_{vol}(J) \) is the volumetric contribution to the hyperelastic response while \( \Psi_{iso} \) and \( \Psi_{aniso} \) are two volume-preserving parts which relate to the isotropic and an anisotropic behavior, respectively. The prevailing myocyte orientation, referred to as the fiber axis, is denoted as \( f_0 \) and the sheet axis, which is perpendicular to \( f_0 \) and parallel to the collagen layers which
envelop bundles of myocytes, is \( \mathbf{s}_0 \). Together with the sheet-normal axis \( \mathbf{n}_0 \), defined to be orthogonal to the sheet and the fiber orientations this forms a right-handed orthonormal set of basis vectors [49]. The volumetric contribution is given as

\[
\Psi_{\text{vol}}(f) = \frac{\kappa}{2} \ln(f^2),
\]

with \( \kappa > 0 \) being a penalty parameter for enforcing the nearly incompressible behavior of the tissue. The volume-preserving isotropic part is based upon an exponential model [25]

\[
\Psi_{\text{iso}} = \frac{a}{2b} \exp \left[ b(T_1 - 3) \right], \quad T_1 = \text{tr}(\mathbf{T}).
\]

where \( a > 0 \) is a stress-like and \( b > 0 \) a dimensionless material parameter, and the anisotropic function is described by \( \Psi_{\text{aniso}} = \Psi_f + \Psi_s + \Psi_{fs} \) and

\[
\begin{align*}
\Psi_f(\mathbf{T}, \mathbf{f}_0) &= \frac{af}{2b} \left\{ \exp[b(\mathbf{I}_f - 1)^2] - 1 \right\}, \\
\Psi_s(\mathbf{T}, \mathbf{s}_0) &= \frac{as}{2bs} \left\{ \exp[b_s(\mathbf{I}_s - 1)^2] - 1 \right\}, \\
\Psi_{fs}(\mathbf{T}, \mathbf{f}_0, \mathbf{s}_0) &= \frac{af_s}{2bs} \left\{ \exp[b_s(\mathbf{I}_s)^2] - 1 \right\},
\end{align*}
\]

with the invariants \( T_f := \mathbf{f}_0 \cdot \mathbf{T} \mathbf{f}_0, T_s := \mathbf{s}_0 \cdot \mathbf{T} \mathbf{s}_0 \) and \( T_{fs} := \mathbf{f}_0 \cdot \mathbf{T} \mathbf{s}_0 \). All parameters, the stress-like \( a_f, a_s, a_{fs} \) as well as the dimensionless \( b_f, b_s, b_{fs} \) are considered to be positive. Moreover, the anisotropic responses (11) and (12) only contribute for \( T_f > 1 \) or \( T_s > 1 \), respectively. Note that in the anisotropic model (11)–(13) sheet-normal properties and corresponding shear terms are omitted as they are either expressible in terms of other invariants or were shown to be of minor significance in shear tests, see [49, 55].

Stresses due to active contraction are assumed to act along the fiber orientation \( \mathbf{f}_0 \) in the reference configuration. Thus, the active Cauchy stress can be defined as

\[
\mathbf{S}_a = \mathbf{S}_a(\mathbf{f}_0 \cdot \mathbf{Cf}_0)^{-1} \mathbf{f}_0 \otimes \mathbf{f}_0,
\]

where \( \mathbf{S}_a \) is the scalar active stress induced in fiber direction \( \mathbf{f}_0 \), see Section 2.3.

2.2. Electrophysiology model

The spread of electrical activation and repolarization is described by the bidomain equations. Cast in the elliptic–parabolic form, these are given in material coordinates by

\[
\begin{align*}
- \text{Div} \left[ J F^{-1} (\mathbf{\sigma}_i + \mathbf{\sigma}_e) F^{-\top} \nabla \phi_e \right] - \text{Div} \left[ J F^{-1} \mathbf{\sigma}_i F^{-\top} \nabla V_m \right] &= 0, \\
- \text{Div} \left[ J F^{-1} \mathbf{\sigma}_i F^{-\top} \nabla V_m \right] - \text{Div} \left[ J F^{-1} \mathbf{\sigma}_e F^{-\top} \nabla \phi_e \right] &= -\beta I_m + \beta I_i, \\
\frac{\partial \eta}{\partial t} &= \mathbf{g}(V_m, \eta, \lambda),
\end{align*}
\]

where \( \mathbf{\sigma}_i \) and \( \mathbf{\sigma}_e \) are the intracellular and extracellular conductivity tensors, respectively;

\[
V_m = \phi_i - \phi_e
\]

is the transmembrane voltage, with \( \phi_i \) and \( \phi_e \) being the intracellular and extracellular potentials, respectively; \( \beta \) is the bidomain surface to volume ratio;

\[
I_m = C_m \frac{\partial V_m}{\partial t} + I_{\text{ion}}(V_m, \eta, \lambda)
\]

is the transmembrane current density; \( C_m \) is the membrane capacitance per unit area; \( I_{\text{ion}} \) is the membrane ionic current density, which depends on \( V_m \), a set of state variables \( \eta \), and the stretch ratio \( \lambda = \mathbf{f}_0 \cdot \mathbf{Cf}_0 \), defined as the ratio between current muscle length and reference length in the unloaded state measured along the fiber direction; \( I_i \) is an intracellular current density, specified on a per membrane area basis, which serves as a stimulus to initiate propagation; and \( V \) is the gradient operator in the reference configuration.

A computationally less costly monodomain model can be derived from the bidomain model under the assumption that intracellular and interstitial conductivity tensors can be related by \( \mathbf{\sigma}_m = \alpha \mathbf{\sigma}_e \). In this scenario Eq. (15) can be neglected and the intracellular conductivity tensor in Eq. (16) is replaced by the harmonic mean tensor

\[
\mathbf{\sigma}_m = \mathbf{\sigma}_i \mathbf{\sigma}_e (\mathbf{\sigma}_i + \mathbf{\sigma}_e)^{-1}.
\]
That is, the monodomain equation is given by

$$\text{Div} \left[ jF^{-1}\sigma_m F^{-\top} \nabla V_m \right] = -\beta I_m + \beta I_{tr},$$

where $I_{tr}$ is a transmembrane stimulus current.

At tissue boundaries, no-flux boundary conditions are imposed on $\phi_i$ and $\phi_e$, i.e.,

$$\int_{\partial\Omega} \sigma_i \nabla \phi_i \, d\Gamma = \int_{\partial\Omega} \sigma_e \nabla \phi_e \, d\Gamma = 0$$

holds. Note that for the numerical examples in Section 6 any effects of deformation upon the distribution of potentials is neglected. Under the conditions considered in here, feedback mechanisms of deformation upon current flow are of minor importance as the period in time where the magnitude of gradients $\|\sigma_i \nabla \phi_i\|$ and $\|\sigma_e \nabla \phi_e\|$ is large, is temporally disjoint from deformation [89]. Thus, in Eqs. (15)–(17), $\mathbf{F}$ is replaced by the identity matrix $\mathbf{I}$ and $\mathbf{f} = \det(\mathbf{F})$ is replaced by 1.

2.3. Active stress model

Most physiological models of active stress generation are based on sets of ordinary differential equations of the form

$$\frac{\partial \mathbf{z}}{\partial t} = \mathbf{h}(\mathbf{z}, \lambda, \dot{\lambda}) = \mathbf{h}(\mathbf{z}, \eta, \mathbf{U}, \dot{\mathbf{U}})$$

(22)

$$S_{\lambda} = f(\mathbf{z}, \lambda, \dot{\lambda}) = f(\mathbf{z}, \eta, \mathbf{U}, \dot{\mathbf{U}}),$$

(23)

where the cellular mechanisms of active stress generation are represented by the state vector $\mathbf{z}$, which depends on the electrophysiological state of the cell $\eta$, stretch ratio $\lambda$, and stretch rate $\dot{\lambda}$.

The dependency on $\lambda$ and $\dot{\lambda}$ accounts for length and velocity dependence of active stress generation while the dependency on the electrophysiological state $\eta$, in particular on the cytosolic calcium concentration $[\text{Ca}^{2+}]$, models excitation-contraction coupling.

Conversely, deformation exerts influence over cellular EP via mechano-electric feedback (MEF) mechanisms which are mediated through the length dependence of $\eta$ through Eq. (17), mainly via $\lambda$ dependent modulation of troponin C binding affinity for calcium, the stretch dependent modulation of ion channel conductivity through Eq. (19), and changes in intracellular and interstitial current flow through Eqs. (15)–(16). If any of these MEF mechanisms is under study, coupling between EP and deformation is bidirectional. In such scenarios, often referred to as strong coupling, the two physics cannot be solved sequentially since deformation influences EP. Hence, they must be solved in a coupled fashion which can be achieved using a monolithic approach or, as used in this study, by updating the variables involved in the coupling when solving for deformation.

3. Finite element method for cardiac electromechanics

3.1. Temporal and spatial discretization

Temporal dynamics of cardiac EP and mechanical deformation are governed by very different time scales, suggesting that the use of different time steps for the two physics is advantageous to obtain a computationally efficient numerical scheme. Even when considering the EP problem alone, the time scales involved differ significantly between reaction and diffusion terms. Fast transients such as the upstroke of the action potential are governed by time constants in the $\mu$s range whereas slower processes occur at the order of tens or hundreds of $\mu$s up to ms. Mechanical deformation can be assumed to occur at similar time scales as slower EP processes such as the cytosolic calcium transients. As mass momentum is not considered here, there is no explicit time dependence in the stationary equilibrium equation (3). Efficient numerical schemes aim to exploit these differences in temporal resolution by using appropriate update intervals [92].

In this study, we conform to the following conventions. Time is discretized according to $t^k = k \cdot \Delta t$ where $\Delta t$ is considered to be a global synchronization clock. Time steps used in the individual subproblems are multiples $n$ of $\Delta t$ with $n \in \mathbb{N}$. In our numerical scheme, three different time steps are used: $\Delta t_{ode}$, to discretize the systems of ODEs in Eq. (17) and Eq. (22) representing EP reaction terms and active stress generation; $\Delta t_{par}$, to discretize the parabolic diffusion terms given by the system of PDEs in Eq. (16); and $\Delta t_{mech}$, to update the current configuration given by the PDE in Eq. (3). Unless otherwise noted, throughout this paper we use $\Delta t_{ode} = \Delta t_{par} = \Delta t$ and $\Delta t_{mech} = n \cdot \Delta t$. Time is discretized by $t^k = k \cdot \Delta t$ for solving the bidomain equations and $t^\ell = \ell \cdot \Delta t_{mech}$ for referring to the time-varying states of deformation. The instants $t^k = k$ and $t^\ell$ are equal for $k = n \cdot \ell$.

A global time step of $\Delta t = 5 \mu$s is used for time discretization of the bidomain equations while a larger time step of $\Delta t_{mech} = 1$ ms is used to solve the mechanical model, owing to its slower dynamics. Thus, the system of ODEs (17) and (22) as well as elliptic PDE (38) and parabolic PDE (42) are integrated in $n$ time steps of size $\Delta t$ from $t_{m}^{\ell-1}$ to $t_{m}^{\ell}$ using $\lambda$ and $\dot{\lambda}$ derived from $\mathbf{U}$ and $\dot{\mathbf{U}}$ at time $t_{m}^{\ell-1}$.
We employ piecewise continuous linear ansatz functions for both the electric potentials in Eqs. (38)–(42) as well as the displacements in Eq. (32). To account for the near incompressibility condition (9) the idea of element-by-element static condensation is applied [50,104]. As we use tetrahedral elements for domain tessellation, this leads to $P_1 - P_0$ elements for the nonlinear deformation problem. Although this type of element is prone to volumetric and shear locking, as discussed in the limitations section 7.5, it is commonly used for cardiac electromechanics applications [5,101]. The same spatial discretization is used for both physics (see Fig. 4). Hence, the projection of coupling variables between electrical and mechanical grid is avoided.

3.2. Variational formulation and discretization of nonlinear elasticity problems

In the numerical examples considered in this study no Neumann boundary conditions were applied (see Section 5.3), and body forces $b_0$ were neglected. We introduce $w = \begin{bmatrix} q; \xi \end{bmatrix}^T$ which is the combined electrophysiological (17) and myofilament (22) state vector and $q = \begin{bmatrix} g; h \end{bmatrix}^T$ which is the respective right-hand side term. Then, the boundary value problem (3)–(5) combined with the cellular EP model (17) and the myofilament model (22) is formally equivalent to the equations

$$\frac{dw}{dt} = q(w, V_m, U),$$

$$\langle A_0(U), V \rangle_{\Omega_0} = 0,$$  \hspace{1cm} (24)

which is valid for all smooth enough vector fields $V$ vanishing on the Dirichlet boundary $\Gamma_{0,D}$ [20, Theorem 2.6-1]. The nonlinear operator $A_0$ is induced by the stress tensor representation (6) and $(\cdot, \cdot)_{\Omega_0}$ is the related duality pairing. The left hand side of the variational equation (25) has the physical interpretation of the rate of internal mechanical work and is given by

$$\langle A_0(U), V \rangle_{\Omega_0} := \int_{\Omega_0} S(w(U), U) : \Sigma(U, V) \ dX,$$  \hspace{1cm} (26)

with the second Piola–Kirchhoff stress tensor

$$S(w(U), U) = S_p(U) + S_a(w(U), U)$$  \hspace{1cm} (27)

and $\Sigma(U, V)$ is the directional derivative of the Green–Lagrange strain tensor, see [48]. Choosing a time discretization as described in Section 3.1 we compute the stress tensor (27) at time step $\ell$ using

$$S(w^\ell(U^\ell), U^\ell) = S_p(U^\ell) + S_a(w^\ell(U^\ell), U^\ell)$$  \hspace{1cm} (28)

where $U^\ell$ is the displacement at time step $\ell$ and $w^\ell(U^\ell)$ is an approximation of $w^\ell$ which is computed similarly to [42] using a forward Euler predictor step as

$$w^\ell(U^\ell) = w^\ell + q(w^\ell + t^\ell, U^\ell) \Delta t$$

where $t^\ell$ is the previous electrical time step, i.e.

$$t^\ell = \ell \Delta t_{\text{mech}} - \Delta t = (\ell \cdot n - 1) \Delta t.$$  \hspace{1cm} (29)

That means that a prediction is repeatedly computed from the previous electrical time step, which is then corrected in each iteration of the Newton loop to account for the feedback of deformation upon the EP state of the tissue.

To apply the finite element method (FEM) we consider an admissible decomposition of the computational domain $\Omega \subset \mathbb{R}^3$ into $M$ tetrahedral elements and introduce a conformal finite element space

$$X_h \subset H^1(\Omega_0), \ N = \dim X_h$$

of piecewise polynomial continuous basis functions $\varphi_i$. Using Newton’s method and a Galerkin finite element discretization results in solving the nonlinear discretized system to find $\delta U_h \in [X_h]^3$, $\delta U_h = 0$ on $\Gamma_{0,D}$ such that

$$\langle \delta U_h, A_0(U^\ell_{v,h}) \rangle_{\Omega_0} = -\langle A_0(U^\ell_{v,h}), \nu_h \rangle_{\Omega_0},$$  \hspace{1cm} (29)

$$U^\ell_{v+1,h} = U^\ell_{v,h} + \delta U_h$$  \hspace{1cm} (30)

holds for all $\nu_h \in [X_h]^3$, $\nu_h = 0$ on $\Gamma_{0,D}$ at time step $\ell$ and $v$ is the index of the Newton iteration. The left hand side of (29), omitting $(\cdot)_h$ to increase readability, is computed by

$$\langle \delta U, A_0(U^\ell_{v,h}) \rangle_{\Omega_0} = \int_{\Omega_0} \nabla \delta U S(U^\ell) : \nabla V \ dX + \int_{\Omega_0} \nabla \delta U : \nabla V \ dX.$$  \hspace{1cm} (31)
For more details see [2,3,48]. The standard finite element method now yields a linear system of equations which is equivalent to the discretized variational formulation (29)–(30). Finally, we have to solve

$$K(U_{v}^\ell) \delta U = -K(U_{v}^\ell), \quad U_{v+1}^\ell = U_{v}^\ell + \delta U,$$

(32)

with the solution vector $U_{v}^\ell \in \mathbb{R}^{3N}$ at the $v$-th Newton step and the increment $\delta U \in \mathbb{R}^{3N}$. The tangent stiffness matrix $K' \in \mathbb{R}^{3N \times 3N}$ is calculated according to

$$K'(U_{v}^\ell)[j, l] := \langle \varphi_{i}, A_{0}'(U_{v}^\ell)\varphi_{j} \rangle_{\Omega_{0}},$$

(33)

and the terms of the right hand side are constructed by

$$K(U_{v}^\ell)[j] := \langle A_{0}(U_{v}^\ell), \varphi_{j} \rangle_{\Omega_{0}}.$$

(34)

Because of symmetry properties of the stress tensor $S$ and the elasticity tensor $C$, the operator $A_{0}'(U_{v}^\ell)$ is self-adjoint and thus the tangent stiffness matrix $K'$ is symmetric. On the other hand, we cannot guarantee that $A_{0}'(U_{v}^\ell)$ is always an elliptic operator, see, e.g., [75]. Nevertheless, we use the conjugate gradient (CG) method as the preferred Krylov solver, and in the case that ellipticity fails and the CG method is not converging we switch to the more robust GMRES method.

### 3.3. Variational formulation and discretization of the bidomain equation

The corresponding weak form to the elliptic–parabolic system (15)–(17) is to find sufficiently smooth $(V_{m}, \phi_{e}, \eta)$ such that

$$\int_{\Omega_{0}} (\sigma_{1} + \sigma_{e}) \nabla \phi_{e} \nabla q_{1} \, dX + \int_{\Omega_{0}} \sigma_{1} \nabla V_{m} \nabla q_{1} \, dX = 0$$

(35)

for all $q_{1} \in H^{1}(\Omega_{0})$ with $\int_{\Omega_{0}} q_{1}(X) \, dX = 0$,

$$\int_{\Omega_{0}} \beta_{C_{m}} \frac{\partial V_{m}}{\partial t} q_{2} \, dX + \int_{\Omega_{0}} \sigma_{1}(\nabla V_{m} + \nabla \phi_{e}) \nabla q_{2} \, dX + \int_{\Omega_{0}} \beta I_{\text{ion}}(V_{m}, \eta) q_{2} \, dX = 0$$

(36)

for all $q_{2} \in H^{1}(\Omega_{0})$ and

$$\int_{\Omega_{0}} \frac{\partial \eta}{\partial t} q_{3} \, dX + \int_{\Omega_{0}} g(V_{m}, \eta, \lambda) q_{3} \, dX = 0,$$

(37)

for all $q_{3} \in L^{2}(\Omega_{0})$.

For more details including regularity assumptions and the proof of existence and uniqueness of a solution of the weak system (35)–(37), we refer to [65].

To reduce the complexity of the problem, the elliptic (35) and parabolic portions (36) of the bidomain equations are decoupled and solved sequentially [118]. First, the elliptic system (35) is solved, then an operator splitting approach is employed to subsequently solve the parabolic system [106, 96, 108]. This allows us to treat reaction and diffusion terms separately. The reaction term represented by the ODE system (37) is solved using an accelerated Rush–Larsen technique, see [92], and the diffusion problem is solved using a Crank–Nicolson scheme. A Galerkin finite element approach and a time discretization as described in Section 3.1 with $t^{k} = k\Delta$ yields the following linear system of equations to obtain $(\phi_{e}^{k+1}, V_{m}^{k+1}, \eta^{k+1}) \in \mathbb{R}^{N} \times \mathbb{R}^{N} \times \mathbb{R}^{N}$ such that

$$K_{i+e}^{k+1} = -K_{i} V_{m}^{k},$$

(38)

$$\eta_{f}^{k+1} = \eta_{f}^{k} e^{-\frac{\Delta t}{\tau_{f}}} + \eta_{f}^{\infty} \left(1 - e^{-\frac{\Delta t}{\tau_{f}}} \right)$$

(39)

$$\eta_{s}^{k+1} = \eta_{s}^{k} + \xi_{s}^{k+1} \Delta t$$

(40)

$$V_{m}^{k+\frac{1}{2}} = V_{m}^{k} - \frac{\Delta t}{C_{m}} I_{\text{ion}}(V_{m}^{k}, \eta^{k+1}, \lambda^{(-1)}).$$

(41)

$$(\bar{M}_{i} + \frac{1}{2} K_{i}) V_{m}^{k+\frac{1}{2}} = -K_{i} \left( \frac{1}{2} V_{m}^{k+\frac{1}{2}} + \phi_{e}^{k} \right) + \bar{M}_{i} V_{m}^{k+\frac{1}{2}},$$

(42)

where
\[
\mathbf{K}_{i+\epsilon}[j, i] := \int_{\Omega_0} (\sigma_i + \sigma_e) \nabla \varphi_i \nabla \varphi_j \, d\mathbf{X},
\]
(43)
\[
\mathbf{K}_i[j, i] := \int_{\Omega_0} \sigma_i \nabla \varphi_i \nabla \varphi_j \, d\mathbf{X},
\]
(44)
\[
\tilde{\mathbf{M}}_i[j, i] := \frac{\beta C_m}{\Delta t} \int_{\Omega_0} \varphi_i \varphi_j \, d\mathbf{X},
\]
(45)

\[\eta = [\eta, \eta] \text{ with } \eta \text{ being fast acting gating variables solved for by evaluating an analytical solution, } \tau \text{ and } \eta_{\infty}, \text{ are functions of the rate coefficients which govern channel gating kinetics, and } \eta \text{ are slower acting states, integrated with a forward Euler method [92]. From this construction and } \sigma = \sigma^T, \ \sigma_e = \sigma_e^T \text{ we conclude that the matrices (43)–(45) are symmetric and positive definite. The CG method is therefore applicable to solve the linear systems of equations (38) and (42).}

4. Linear solvers

All linear solvers used in this study were implemented in the publicly available numerical package Parallel Toolbox (pt) [72,44,73]. The cardiac bidomain equations and the nonlinear deformation equation were solved in parallel using the Cardiac Arrhythmia Research Package (CARP) [117] as an outer framework which was interfaced with the pt library. Technical aspects have been described elsewhere [80]. CARP makes use of the MPI based library PETSc [7] as the basic infrastructure for handling parallel matrices and vectors. Parallel partitioning of unstructured grids relies upon the graph based decomposition ParMetis [54,55].

Details on the methods employed for solving the bidomain equations in parallel have been reported previously [91,118, 80]. Briefly, the elliptic portion of the bidomain equations is solved using the algebraic multigrid preconditioner ptAMG [72] with an iterative CG solver method (ptAMG-PCG). The parabolic portion of the bidomain equation is solved using an \(\omega\)-Jacobi preconditioner, also in combination with the iterative CG solver (\(\omega\)-PCG), using an \(\omega\) of 2/3 and two iterations per CG iteration [80]. The ptAMG-PCG method used for solving the scalar elliptic PDE turned out to be of very limited efficiency for solving the nonlinear deformation problem, thus motivating an adaptation of the AMG preconditioner for nonlinear biomechanical applications. The Krylov solver of choice remained the CG method. In the case that ellipticity fails, we switch to the GMRES method (see section 3.2). An extensive description of the underlying concepts is presented in the following section.

Other AMG implementations such as hypre/BoomerAMG [45] or ML [33] are suitable for the electrical and mechanical problems described in this paper. However, using these packages, we were not able to match the overall performance of our own AMG implementation. This is not surprising as those packages are designed to accommodate a wide range of applications, whereas our own much simpler implementation is tuned for a small set of applications in cardiac modeling.

4.1. Algebraic multigrid design

For a detailed introduction to the AMG method we refer to [114,107]. The initial design of the ptAMG preconditioner was geared towards solving linear systems arising from the discretization of scalar PDEs such as the elliptic PDE embedded in the bidomain equations. To be suitable for systems of equations where each node in a finite element mesh is associated with more than one DOF, the AMG setup had to be altered.

We assume an equation system \(\mathbf{K} u = b\) with \(\mathbf{K} \in \mathbb{R}^{3N \times 3N}\) and \(u, b \in \mathbb{R}^{3N}\) as in (32). The total number of unknowns is \(3N\) where \(N\) represents the number of finite element grid points and \(3\) is the number of physical unknowns per grid point. Classical multigrid coarsening methods would compute the coarse grid selection based on the discrete operator \(\mathbf{K}\), taking into account the matrix graph alone [9], or additionally also the magnitude of matrix entries [102]. However, in our application, this would lead to different coarse grids for the physical unknowns of one grid point which can translate into bad convergence behavior.

As in [39], this problem is circumvented by computing an auxiliary matrix \(\tilde{\mathbf{K}} \in \mathbb{R}^{N \times N}\), defined as

\[
\tilde{\mathbf{K}}[i, j] := \| K_{ij} \|_F
\]
(46)
where \(K_{ij} \in \mathbb{R}^{3 \times 3}\) denotes the subblock of \(\mathbf{K}\) which couples the DOF associated with a finite element grid point \(i\) with the DOF of the grid point \(j\), and \(\| \cdot \|_F\) is the Frobenius matrix norm. A coarsening scheme inspired by the Ruge–Stüben scheme [102] is applied to the auxiliary matrix using

\[
\tilde{\mathbf{K}}[i, j] \geq \epsilon \tilde{\mathbf{K}}[i, i]
\]
(47)
as strong connection criterion for the edge \((i, j)\). The set of FE grid point indices \(T := \{0, \ldots, N - 1\}\) is split into the disjoint subsets of coarse grid \(C \subset T\) and fine grid indices \(F \subset T\). The prolongation operator \(P\) given by
interpolates from $C$ onto $T$. The coarse grid $C \subset T$ can be mapped directly via the identity $I_C$ while the fine grid $F \subset T$ needs to be interpolated from the coarse grid via $P_F$. The interpolation weight $n_i$ is computed as the cardinality of the set of all coarse grid points interpolating any $i \in F$:

$$n_i := |\{j \in C : \tilde{K}[i, j] > \epsilon \tilde{K}[i, i]\}|$$

(49)

Then, the coarse to fine interpolation $P_F$ is defined as

$$(P_F)_{ij} := \begin{cases} 
(1/n_i) I_{3 \times 3} & \text{if } \tilde{K}[i, j] > \epsilon \tilde{K}[i, i] \\
0 & \text{otherwise}
\end{cases}$$

(50)

with $i \in F$ and $j \in C$. Finally, the coarse grid operator $K_c$ is defined as the triple matrix product

$$K_c := P^T C P.$$  

(51)

By default, the multigrid coarsening in ptAMG is continued until no strong connections between the remaining unknowns are found. The inverse of the resulting coarse grid system can therefore be approximated by the inverse $3 \times 3$ block diagonal. Alternatively, the coarsening process can be configured to finish earlier at a finer level. The resulting coarse grid system needs to be inverted by a direct solver. While using a direct coarse-grid solver can greatly reduce the number of solver iterations, the overall setup and solving time was consistently worse in our numerical benchmarks. One pre- and post-smoothing iteration with a Jacobi-smoother is used in a multigrid V-cycle as a preconditioner step. The inverse $3 \times 3$ block diagonal of the system matrix is used as the approximate inverse in the Jacobi iteration.

4.2. Strongly scalable parallelization

Devising iterative solvers and suitable preconditioners with favorable strong scaling properties, i.e., where execution time reduces when the number of compute cores used increases, is of pivotal importance when attempting to model cardiac electromechanics at high spatial resolutions. Departing from an initial implementation of the ptAMG method described above, strong scaling experiments were performed, in which an entire single heart beat was simulated with a varying number of processors. Parallel performance and scalability was gauged based on the following performance metrics:

1. **Real time lag factor $\eta$**: Execution time divided by the time span simulated.
2. **Speedup $S$**: The reduction in execution time achieved when using $N_p$ parallel processes relative to a reference run using $N_{pr}$ processes

$$S = \frac{T_{pr}}{T_p}.$$  

(52)

3. **Parallel efficiency $E$**: The algorithm’s ability to reduce execution time as $N_p$ increases, defined as the ratio of the fractional reduction in execution time to the fractional increase in $N_p$.

$$E = \frac{T_p}{T_{pr}} \frac{N_{pr}}{N_p}.$$  

(53)

where $T_p$ refers to the execution time when using $N_p$ processes, and $T_{pr}$ refers to the same quantity measured in a reference simulation using $N_{pr}$ processes. An $E$ value of 1 is referred to as linear scaling, because computation time is reduced by the same factor as $N_p$ is increased. $E$ values below 1 are known as sub-linear scaling whereas those above 1 are known as super-linear scaling.

Whenever parallel efficiency degraded, parallel profiling data was generated and analyzed using the Scalasca toolset \cite{34} (Score-P, CUBEx, Scalasca) to identify the scaling bottlenecks. Based on these analyses, the following parallelization techniques were implemented:

1. Data exchange on the interfaces between individual subdomains was reworked to better load balance the cost of accumulation of interface data, as shown previously for solving a scalar elliptic PDE \cite{74}.
2. Parallel data layout of the multigrid hierarchy was modified to increase the effective network bandwidth of the data exchange across subdomain interfaces. Parallel profiling has shown that a combination of small average message sizes and dense communication patterns significantly decrease the effective network bandwidth on the intermediate multigrid levels. A strategy to counter this phenomenon is to redistribute the parallel data of the intermediate multigrid levels onto fewer active processes \cite{6}, thus reducing communication density and increasing average message size at the same time. The reasoning of this approach is, that the increase in effective network bandwidth outweighs the loss of computational units.
5.1.1. The approach implemented in ptAMG is to systematically reduce the number of active processes on each multigrid level. A constant group size \( g > 1 \) is introduced. On any multigrid level, the range of active processes is split into groups of size \( g \). One process of each group, declared to be the master, will be active on the next coarser multigrid level, while the other processes are declared idle. The idle processes redistribute their current coarse grid system to their group’s master, thus reducing the number of active processes on any multigrid level to \( 1/g \) of the previous level.

The main advantage of this approach is that there is no need for tuning the data redistribution process to individual use cases (e.g. problem size, number of computational units). The optimal choice for \( g \) depends only on the rate by which the unknowns are reduced during coarsening. By default \( g \) is set to 2.

3. The overall AMG algorithm was split into steps which depend on the matrix sparsity pattern (and thus on the FE graph) and those which depend on the actual matrix entries. When solving Eq. (29), the system matrix is updated in each iteration of the Newton solver, but the sparsity structure remains unchanged as our method is not \( h \)-adaptive. Therefore only those setup steps which depend on the matrix values have to be executed.

5. Four chamber electromechanical model

5.1. Generation of whole heart anatomy model

A processing pipeline consisting of multiple stages was implemented, incorporating segmentation, tissue classification, geometry smoothing and resampling, FE mesh generation and tag transfer, and the assignment of orthotropic eigenaxes. The development of the model generation pipeline aimed at achieving maximum flexibility, allowing the generation of complete four chamber models of the heart including the connected major vessels. Such models are preferable from a modeling point of view as they allow the application of more realistic mechanical boundary conditions [32].

5.1.1. Image acquisition

The anatomical model was derived from an end diastolic 3D balanced steady state free precession (SSFP) cardiac magnetic resonance imaging acquisition in a sagittal orientation with whole-heart coverage and an isotropic resolution of 1.3 mm. The image was processed using automatic model based segmentation of the left and right ventricle chambers, left and right atrial chambers and the left ventricle myocardium wall [90].

5.1.2. Segmentation and tissue classification

As the thin walled atrial walls and the right ventricular wall could not be delineated directly from the image stacks, these walls were introduced by dilating the atrial blood pool by 2 voxels and a 4 voxel right ventricular wall was created by dilating the corresponding chamber [21]. Classification tags were added to further discriminate between different anatomical regions and tissue types relevant to the subsequent model parametrization. As there is no region-specific signature in the acquired imaging data which would allow for an automatic discrimination, a semi-automatic approach was implemented using various tools such as CAIP (Fraunhofer, MEVIS) and itsknap (University of Pennsylvania, USA).

5.1.3. Anatomy smoothing

Depending on the resolution of a given medical image stack, the geometric appearance of the heart’s surfaces in a generated mesh can be jagged, which is ideal neither for visual representation nor for computational accuracy. These staircasing artefacts were reduced by smoothing and upsampling the segmented image stack using a variational technique [47].

Briefly, a marching cubes algorithm, as implemented in the VTK library [119], was used to generate a triangulated surface representation of the heart’s anatomy. In an iterative variational optimization procedure, the generated surfaces were smoothed. The method avoided volumetric shrinking and limited the maximum nodal displacements to \( \pm0.5 \) voxels. The smoothed surface mesh was resampled at an isotropic resolution of 110 \( \mu \)m using the 2D Cairo library [16] and stacked together to generate the smoothed 3D volume. Note that only the anatomical mask was smoothed, while previously assigned classification tags were preserved.

5.1.4. Finite element mesh generation

Segmentation and tagged image stacks were fed into the image based unstructured mesh generation software Tarantula (CAE Software Solutions, Eggenburg, Austria) which builds fully unstructured, boundary fitted, locally refined, hex-dominant hybrid tessellations consisting of hexahedra, prisms, pyramids and tetrahedra [95].

Hybrid meshes were converted into purely tetrahedral element meshes. Classification tags were carried over from the tagged image stacks to the finite element mesh to be used later for the assignment of electrical and mechanical tissue properties during finite element matrix assembly. For the sake of numerical testing, three finite element meshes of the same whole heart geometry were generated at different spatial resolutions in multiples of the image stack resolution of 110 \( \mu \)m. A fine model (FM) was discretized at an average \( h \) of \( 220 \) \( \mu \)m, a medium model (MM) at \( h \approx 440 \) \( \mu \)m, and a coarser model (CM) at \( h \approx 880 \) \( \mu \)m.

While the CM with a resolution \( h \approx 880 \) \( \mu \)m is, to our knowledge, the highest resolution mesh used for modeling cardiac mechanics, for solving the bidomain equations with acceptable accuracy its resolution is far too coarse. With physiological EP models as used here, predictions of conduction velocity would carry unacceptably large errors. At slower conduction
velocities, as they govern conduction in a transmural direction, even an artificial conduction block may occur, thus rendering such a model unsuitable for bidirectionally coupled EM modeling. Nonetheless, this coarse model is deliberately included in this study to report performance and scalability limits since such smaller models are suitable for being used in weakly coupled electromechanical applications where EP and mechanics can be treated in isolation. Further, results obtained with the CM model are informative for simulating smaller animal models with appropriately higher spatial resolution.

Cardiac tissues are orthotropic in both electrical [17] as well as mechanical properties [49]. Since measuring the eigenaxes of the orthotropic tensor in vivo is not yet clinical routine and generates sparse and often noisy data sets, a previously developed Laplace–Dirichlet rule based method (LDRB) is employed [8] to assign the principal eigenaxes along fiber direction $f_0$, sheet direction $s_0$, and sheet-normal direction $n_0$ to each finite element in the ventricles. In the atria no fibers were assigned. The atria were modeled as an unexcitable, non-contracting, isotropic hyperelastic material, that had the principal role of providing more physiological boundary conditions in this model.

5.2. Modeling electrophysiology and active stress generation

Cellular dynamics were described by the recent Grandi–Pasqualini–Bers (GPB) model of the human ventricular myocyte [38], which was coupled to the Land−Niederer (LN) active stress model [69]. To allow for strong coupling, i.e., to account for length effects on the cytosolic calcium transient, the equation governing the binding of calcium to the low affinity regulatory sites on troponin in the myofilament model (see Eq. (114) in the Appendix of [38])

$$\frac{d[TnC_L]}{dt} = k_{onTnC} [Ca^{2+}] \left( \hat{B}_{TnC_L} - [TnC_L] \right) - k_{offTnC} [TnC_L]$$  \hspace{0.5cm} (54)

is replaced by the analogous Eq. (1) of the LN active stress model [69], given as

$$\frac{d\text{TRPN}}{dt} = k_{TRPN} \left( \frac{[Ca^{2+}]}{[Ca^{2+]_{150}(\lambda)}} \right)^{n_{TRPN}} (1 - \text{TRPN}) - k_{TRPN} \text{TRPN}.$$  \hspace{0.5cm} (55)

In the combined GPB+LN model (54) is replaced by (55), scaled by the total buffer concentration $\hat{B}_{TnC_L}$ assumed to be 70 μmol·L$^{-1}$ in the GPB model. This scaling is necessary to correctly relate fractional occupancy, used by the LN myofilament model, with the concentration of troponin bound calcium, used by the GPB model. That is, the state TnC state vector $\eta$ is replaced by an appropriately scaled state TRPN, taken from the LN state vector, $\zeta$. Therefore, the number of states in the combined state vector $w$ is 58, which is one less than the sum of the number of states of the GPB state vector $\eta$ (54 states) and the LN state vector $\zeta$ (5 states). Parameters of the GPB model were left unaltered whereas parameters of the LN model were adapted to give human myocardium tension transients when coupled to the GPB calcium transient [110] (see Table 1). This study sought to find an optimal set of parameters for combining cytosolic calcium transients, as predicted by the GPB model, with active stress transients, as predicted by the LN model.

In single cell experiments, both the unaltered GPB model and the combined GPB+LN model were paced at a cycle length of 500 ms for a duration of 60 seconds. In the combined GPB+LN model the stretch ratio was kept constant at $\lambda = 1.0$ over the entire protocol. In both models, all state variable transients were plotted over the entire pacing experiment to ensure that the system had settled close to a stable limit cycle. The effect of the altered troponin buffering equation in the GPB+LN model relative to the GPB model is illustrated in Fig. 1. The effect of strong coupling in the GPB+LN model under fixed stretch conditions in the range $\lambda \in [1.0, 1.2]$, with the same pacing protocol, is illustrated in Fig. 2. The state vector at the end of the pacing protocol is stored to be used later for initializing the organ model. The same GPB+LN model is used throughout the ventricles, known electrophysiological heterogeneities remained unaccounted for.

---

Fig. 1. Comparison of $V_m$, TnC, and $[Ca^{2+}]$ single cell traces between the GPB model (blue) and the GPB+LN model at steady state with $\lambda = 1$ (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
5.3. Mechanical boundary conditions

Homogeneous Dirichlet boundary conditions were applied at the termini of the meshed pulmonary veins and superior and inferior vena cava (Fig. 3). In addition, an elastic apical boundary condition was applied by attaching a small fragment of soft material to the apex of the heart, and applying homogeneous Dirichlet boundary conditions to the opposite side of this material. The heart was modeled without attachment to the vascular system, with left and right ventricles remaining unpressurized over the entire cardiac cycle. This setup therefore modeled a Langendorff perfused preparation of a human heart [79].

5.4. Material parameters

The bidomain equations were parametrized to produce conduction velocities of 0.6 m s$^{-1}$, 0.4 m s$^{-1}$ and 0.2 m s$^{-1}$ along the orthotropic eigenaxes $f_0$, $s_0$ and $n_0$ of the tissue. While simulating action potential propagation in the FM ($h = 220 \mu m$) was possible with convergence errors that were well within the bounds of measurement uncertainties (2.8%, 4.5% and 19.4% for $v_f$, $v_s$ and $v_n$ with the numerical settings used in this study), significant errors incurred at coarser mesh resolutions, particularly at slower conduction velocities. Without appropriate adjustments of conductivities, artificial conduction block would occur in the CM for propagation along the $n_0$ axis. Therefore, to minimize differences in mechanical response induced by deviations in activation pattern intracellular conductivities were varied as a function of $h$ to arrive at comparable conduction velocities for FM, MM and CM [23]. Using these tuned conductivity values (Table 1) led to the same conduction velocities for each mesh resolution under planar wavefront propagation conditions. It is worth noting that the necessary variations were well within the uncertainty of conductivity measurements reported in the literature [22,99,98] for both the FM and MM model, but not the CM model as an $\bar{h} = 880 \mu m$ is insufficient when using a biophysically detailed model of cellular dynamics.

5.5. Benchmark setup

The combined electromechanical four chamber model was used to benchmark performance and scalability of the AMG solver. At time $t = 0$ the endocardium of both ventricles was stimulated with a transmembrane current of 200 $\mu A cm^{-2}$ for
Fig. 3. Mechanical boundary conditions: homogeneous Dirichlet boundary conditions were enforced at the termini of the meshed pulmonary veins and venae cavae, and at the bottom of the soft material attached to the apex.

Table 1
Summary of electrical and mechanical material parameters. Bidomain conductivities were varied as a function of mean resolution \( \bar{h} \) to keep conduction velocities at prescribed values.

<table>
<thead>
<tr>
<th>Used for</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passive stress</td>
<td>( \kappa = 1000 \text{ kPa} )</td>
<td>( \alpha_1 = 8.535 \text{ kPa} )</td>
</tr>
<tr>
<td></td>
<td>( \sigma = 1.333 \text{ kPa} )</td>
<td>( \alpha_2 = 2.564 \text{ kPa} )</td>
</tr>
<tr>
<td></td>
<td>( b = 9.242 [\text{--}] )</td>
<td>( \alpha_3 = 0.417 \text{ kPa} )</td>
</tr>
<tr>
<td>Active stress</td>
<td>( T_{\text{ref}} = 117.1 \text{ mNm}^2 \text{mm}^{-2} )</td>
<td>( [\text{Ca}^{2+}]_{\text{SO}} = 0.52 \text{ μmolL}^{-1} )</td>
</tr>
<tr>
<td></td>
<td>( \sigma_{\text{TRPN}} = 1.54 [\text{--}] )</td>
<td>( k_{\text{TRPN}} = 0.14 \text{ ms}^{-1} )</td>
</tr>
<tr>
<td>Bidomain, 220 μm</td>
<td>( \beta = 1400 \text{ cm}^{-1} )</td>
<td>( \sigma_{\text{il}} = 0.23415 \text{ S m}^{-1} )</td>
</tr>
<tr>
<td></td>
<td>( C_{\text{m}} = 1.0 \text{ μF cm}^{-2} )</td>
<td>( \sigma_{\text{il}} = 0.08603 \text{ S m}^{-1} )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \sigma_{\text{il}} = 0.015933 \text{ S m}^{-1} )</td>
</tr>
<tr>
<td>Bidomain, 440 μm</td>
<td>( \beta = 1400 \text{ cm}^{-1} )</td>
<td>( \sigma_{\text{il}} = 0.29431 \text{ S m}^{-1} )</td>
</tr>
<tr>
<td></td>
<td>( C_{\text{m}} = 1.0 \text{ μF cm}^{-2} )</td>
<td>( \sigma_{\text{il}} = 0.14047 \text{ S m}^{-1} )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \sigma_{\text{il}} = 0.03521 \text{ S m}^{-1} )</td>
</tr>
<tr>
<td>Bidomain, 880 μm</td>
<td>( \beta = 1400 \text{ cm}^{-1} )</td>
<td>( \sigma_{\text{il}} = 1.08396 \text{ S m}^{-1} )</td>
</tr>
<tr>
<td></td>
<td>( C_{\text{m}} = 1.0 \text{ μF cm}^{-2} )</td>
<td>( \sigma_{\text{il}} = 0.76053 \text{ S m}^{-1} )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \sigma_{\text{il}} = 0.09040 \text{ S m}^{-1} )</td>
</tr>
</tbody>
</table>

2 ms, initiating the transmural activation of the ventricles. Electromechanical activity was simulated for 500 ms to cover all phases of an entire heart beat. The same protocol was applied to all three models (CM, MM and FM), with the only alteration being the use of different conductivities (Table 1) to arrive at sufficiently similar electrical activation patterns. Simulations were performed with varying numbers of processes \( N_{\text{p}} \) for each model, with the range of \( N_{\text{p}} \) scaled by overall problem size (32–2048, 128–4096, and 512–8192 processes for CM, MM and FM, respectively). For all runs, time spent on solving the individual components of the elliptic PDE (bidomain equations), parabolic PDE and ODE systems was recorded. For the mechanics solver, time spent on matrix assembly and solver time was measured. Furthermore, the number of iterations was stored for both linear and nonlinear solves. The nonlinear mechanical problem was solved using Newton’s method with an absolute \( \ell^2 \) norm error reduction of the residual (34) of \( \varepsilon = 10^{-6} \). For all linear subproblems, we used the preconditioned CG method with an absolute error reduction of \( \varepsilon = 10^{-8} \).

6. Results

6.1. Model generation and baseline simulations

Finite element meshes were generated from the segmented image stack at three different average mesh resolutions \( \bar{h} \) of 220 μm, 440 μm and 880 μm. The resulting FM, MM and CM geometry models are visualized in Fig. 4 and metrics for these meshes are summarized in Table 2.
Prior to numerical benchmarking, baseline simulations were carried out for each model using the same settings except for the bidomain conductivities, which were modified as a function of $\tilde{h}$ (Table 1). Maps of local activation time ($T_{act}$) were computed from $V_m$ to validate that the electrical activation sequence of the ventricles was consistent between models. As shown in Fig. 5, the total activation time of the ventricles was around 45 ms. These activation times which are shorter than anticipated for a physiological activation sequence in a human, are due to the simple endocardial stimulation profile which covered the entire endocardia of both ventricles. This is different from a physiological case where the Purkinje system stimulates the endocardia only over apical and mid segments, but not over the basal segments which are not covered by the Purkinje system. Visual comparison of activation isochrones for later activation times around 35–45 ms suggests that the choice of $\tilde{h}$-dependent conductivities led to activation sequences which were indeed very similar between the models (Fig. 5). As expected for the case of closely matching activation sequences, close agreement was also found between models for the electrophysiological quantities $V_m$, $\phi_e$ and $[Ca_i]$. Similarly, the mechanical quantities $\|u\|$, $S_s$, $S_p$ and $\lambda$, computed from (25) at a temporal resolution of 1 ms were also compared visually between models. Differences in global mechanical performance of the left ventricle were gauged by tracking the time course of the left ventricular cavity volume $V_{LV}(t)$ over an entire heartbeat. A representative result is shown for the FM model in Fig. 6.

6.2. Solver configuration

Solver parameters were chosen based on the experience gained during the baseline simulations. The parameter $\epsilon$, which is used during the AMG setup to decide which nodes are strongly connected, was set to 0 for both the electrics and me-
mechanics solvers. With $\epsilon = 0$, the AMG setup is based on the matrix graph alone. While this choice had a negative impact on the linear solver convergence, it significantly improved parallel load balancing and, consequently, overall performance. Compared to $\epsilon = 0$, choosing $\epsilon > 0$ yields a larger parallel variability in size of the linear systems arising at the various levels of the multigrid hierarchy, particularly for uneven parallel distributions of the Dirichlet boundaries. These variations in linear system size led to poorly balanced computational load and thus increased MPI waiting time. The number of pre- and post-smoothing sweeps was set to 1. A larger number of smoothing sweeps would improve convergence but overall solving time would be longer as the reduction in iterations is insufficient to compensate for the increased computational cost of the additional smoothing sweeps. Similarly, the use of direct solvers for inverting the AMG coarse-grid was omitted for both the mechanical problem and the elliptic problem. While direct coarse-grid solvers can improve convergence significantly (depending on the number of multigrid levels used), the baseline benchmarks also showed a reduced overall parallel efficiency of both setup and solving phases. Instead, the coarse-grid solving strategy was to continue coarsening until a $1 \times 1$ or $3 \times 3$ system was reached for the elliptic and mechanics problems respectively. The most important factor for improving strong scalability was the grouping of parallel data at the intermediate multigrid levels onto fewer active computing units, as described in Section 4.2. The reduction in communication density and the increase in average message size turned out to be beneficial for $N_p > 320$. This effect was even greater for $N_p > 640$, above which scalability stalled with a group size of $g = 1$, meaning data redistribution was turned off (Fig. 7). In terms of overall solving time, a group size of $g = 2$ proved to be optimal; a more aggressive reduction of the computing units $g = 3, 4$ was not beneficial.

The performance of the linear solver in terms of convergence and cost of individual solver steps was monitored by recording iteration count, $N_{it}$, and time per iteration, $T_{it}$. Over the range of processes $N_p$ tested, only a very minor dependency of $N_{it}$ was observed. Thus only one representative graph for $N_p = 512$ is shown to illustrate the course of $N_{it}$ over an entire cardiac cycle (Fig. 8). $N_{it}$ increased with the overall size of the system to be solved, which shows that ptAMG does not fully succeed in preserving convergence independently of the spatial resolution. Nevertheless, the increase in iteration counts is moderate compared to the increase in spatial resolution and thus linear system size from CM to FM (also compare Table 2). The number of nonlinear Newton iterations $N_{Ne}$ was monitored as well. Depending on the phase of deformation $N_{Ne}$ varied in the range between 3 to 7 Newton iterations, largely independent of the problem size (Fig. 8).

6.3. Strong scaling benchmarks

The final baseline setup, which simulated EP and mechanics of an entire heartbeat over a basic cycle length of 500 ms, was used for performing a strong scaling benchmark. The goal was to determine the strong scaling performance of the major contributing factors (finite element matrix assembly time and solver time) as well as the overall scaling. The initial number of used processes $N_p$ was chosen sufficiently large to be able to fit the electromechanical model into memory. The required $N_p$ was 32, 128 and 512 for CM, MM and FM respectively. Due to its highly parallel nature, the matrix assembly showed linear scaling over almost all $N_p$ probed. With the chosen ptAMG-PCG solver configuration for the deformation problem,
strong scaling with high efficiency was achieved over a wide range of processor counts (Fig. 9, left panel). For the CM model a parallel efficiency above 69% was achieved as far as 512 processes. From 1024 processes parallel efficiency started to degrade, which can be explained by low computational load on each of the very small local partitions ($\approx 2500$ DOF per process). The MM model showed an efficiency above 64% up to 2048 processes. Parallel efficiency saturated at $N_p = 4096$ ($\approx 4000$ DOF per partition). The FM model showed strong scaling for all $N_p$ probed. At $N_p = 8192$ (with $\approx 13000$ DOF per partition) parallel efficiency was still at 58%. Further increasing $N_p > 8192$ could not be tested, due to an unresolved fault of the MPI library on the available HPC system. Measured timings and parallel efficiency values are listed in Table 3. Overall, for electromechanical simulations using the monodomain equations to represent EP, computing costs are clearly dominated by solving the deformation equation which is between 3× and 7× as expensive as the matrix assembly costs. In comparison, the costs of solving the parabolic PDE and the system of ODEs which make up the monodomain model are negligible.

When considering a bidomain model to represent EP in an electromechanical simulation, the distribution of costs is markedly different. As shown in the right panel of Fig. 9, computational costs of solving the elliptic portion of the bidomain equation alone is comparable to solving an electromechanical problem with a monodomain EP model. In addition, solving of parabolic and elliptic PDEs can be seen to be far less scalable as solving the deformation problem on the same grid. This is due to the smaller number of degrees of freedom per node which leads to a linear system to be solved that is smaller by a factor of 3× and a computational load that is smaller by a factor of 9×. Therefore, the critical surface to volume ratio of a local partition which translates to a communication to compute load ratio, is hit earlier, at $\approx N_p/3$. This confirmed by Fig. 9, where scaling efficiency can be seen to drop markedly at $N_p = 512, 2048$ and 4096 for CM, MM and FM respectively.
Table 3
Strong scaling benchmark results for a monodomain electromechanical simulation of a heartbeat with CM, MM and FM model: table columns are number of cores $N_p$, and timings for matrix assembly (MA), solving (Solve), monodomain (Mono), total time and parallel efficiency $E$.

<table>
<thead>
<tr>
<th>$N_p$</th>
<th>MA</th>
<th>Solve</th>
<th>Mono</th>
<th>Total</th>
<th>$E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CM</td>
<td>32</td>
<td>5746 s</td>
<td>25 088 s</td>
<td>12 852 s</td>
<td>43 687 s</td>
</tr>
<tr>
<td></td>
<td>64</td>
<td>2928 s</td>
<td>13 054 s</td>
<td>6 381 s</td>
<td>22 763 s</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>1488 s</td>
<td>6 946 s</td>
<td>3 290 s</td>
<td>11 724 s</td>
</tr>
<tr>
<td></td>
<td>256</td>
<td>771 s</td>
<td>4 006 s</td>
<td>2 009 s</td>
<td>6 787 s</td>
</tr>
<tr>
<td></td>
<td>512</td>
<td>404 s</td>
<td>2 372 s</td>
<td>1 175 s</td>
<td>3 951 s</td>
</tr>
<tr>
<td></td>
<td>1024</td>
<td>213 s</td>
<td>1 633 s</td>
<td>812 s</td>
<td>2 658 s</td>
</tr>
<tr>
<td>MM</td>
<td>128</td>
<td>9 503 s</td>
<td>50 122 s</td>
<td>20 042 s</td>
<td>79 667 s</td>
</tr>
<tr>
<td></td>
<td>256</td>
<td>4 835 s</td>
<td>25 752 s</td>
<td>10 683 s</td>
<td>41 270 s</td>
</tr>
<tr>
<td></td>
<td>512</td>
<td>2 484 s</td>
<td>13 545 s</td>
<td>5 541 s</td>
<td>21 572 s</td>
</tr>
<tr>
<td></td>
<td>1024</td>
<td>1 293 s</td>
<td>8 163 s</td>
<td>3 030 s</td>
<td>12 486 s</td>
</tr>
<tr>
<td></td>
<td>2048</td>
<td>716 s</td>
<td>5 212 s</td>
<td>1 889 s</td>
<td>7 817 s</td>
</tr>
<tr>
<td></td>
<td>4096</td>
<td>461 s</td>
<td>3 563 s</td>
<td>1 246 s</td>
<td>5 270 s</td>
</tr>
<tr>
<td></td>
<td>8192</td>
<td>1 743 s</td>
<td>9 363 s</td>
<td>3 014 s</td>
<td>14 120 s</td>
</tr>
</tbody>
</table>

Fig. 9. Strong scaling profile of an electromechanical model using a monodomain (left) or bidomain (right) EP model. Red, blue and green traces refer to CM, MM and FM simulation runs respectively, over a time frame of 500 ms. On the left hand side the assembly time for mechanics (33)–(34) and the solve time (i.e. preconditioner setup and solving) for all linearized systems of equations (32) is given. In addition, we show the total solving time of a monodomain electromechanical simulation. Depicted in the right plot is the computational time for all elliptic systems (38), the computational time for all ODE systems (39)–(40) and the computational time for all parabolic systems (42). In these three cases the computational time includes assembling, preconditioner setup and solving times. Solid lines correspond to the total solving time of a bidomain electromechanical simulation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

6.4. Impact of length dependence upon calcium

The effect of length dependent calcium buffering of troponin C on cytosolic calcium was investigated by comparing two sets of simulations using the same settings except for the reference value $[Ca^{2+}]_{T50}(\lambda)$ in Eq. (55). In the strong coupling scenario $[Ca^{2+}]_{T50}(\lambda)$ was used as described in [69], i.e.,

$$[Ca^{2+}]_{T50}(\lambda) = [Ca^{2+}]_{T50}^{ref}(1 + \beta_1(\lambda - 1))$$

whereas in the weak coupling case $[Ca^{2+}]_{T50}(\lambda)$ was kept constant at the value corresponding to $\lambda = 1.0$, i.e., $[Ca]_{T50}(\lambda) = [Ca^{2+}]_{T50}^{ref}$. Fig. 10 illustrates the effect of length dependent calcium binding affinity on calcium transients, stretch ratio $\lambda$ and active stress $S_a$.

7. Discussion

This study describes novel methodology for modeling human whole heart electromechanics at a high spatial resolution. Such high resolutions are beneficial for several reasons: (i) cardiac anatomy can be faithfully represented with high geomet-
ric fidelity; (ii) achieving fully, or at least nearly, converged simulation results becomes feasible in strong coupling scenarios where mechanical deformation influences the electrophysiological state of cells via MEF mechanisms; and (iii) both fine scale structural heterogeneities such as vessels, fat deposits, scarred tissue and fibrosis as well as functional heterogeneities due to spatial variation in micro-perfusion, metabolic state or protein expression can be explicitly included in a model. The computational burden imposed by such models is significantly larger when compared to lower dimensional models as they are standard in cardiac mechanics applications [59].

This issue of computational cost has been addressed by focusing on a massively parallel iterative solving method for large deformation problems, allowing us to exploit the full potential of modern HPC hardware. In particular, we developed a novel AMG preconditioner for a CG solver for application to nonlinear mechanics. The primary design objective was to achieve favorable strong scaling characteristics for both setup and solution times up to a sufficiently large core count, such that execution times were reduced and simulations were computationally tractable. Depending on the chosen resolution, benchmark results demonstrated efficient scaling up to 1024, 4096 and 8192 compute cores which allowed the simulation of a single heart beat in 44.3, 87.8 and 235.3 minutes with the CM, MM and FM models respectively. The achieved efficiency makes sufficiently short simulation cycles feasible without compromise of anatomical or biophysical detail.

7.1. Computational costs

The computational burden imposed by high resolution electromechanical models demands the most efficient numerical approaches in order to keep simulations tractable. In the cardiac EP modeling community, two approaches are currently investigated, the use of spatio-temporal adaptivity [10,26], polynomial adaptivity [1], or, as in this study, the use of strongly scalable solvers which reduce execution times by engaging a larger number of computational units, be it traditional CPUs [82,97], or acceleration devices such as GPUs [80], or a combination of both parallelization and adaptivity [27]. Solving deformation problems in cardiac or vascular modeling applications relies more on reducing cost by resorting to lower dimensional EP models with slow upstroke velocities combined with simplified active stress models [85,5,42] and weak unidirectional coupling to allow for coarse spatial discretization of mechanics [85,5,42]. In terms of solver methods used, most studies relied upon direct solvers which tend to become unfeasible for high resolution 3D problems due to linear system size and sparsity pattern [46]. Exceptions exist where strongly scalable solvers were employed, so far only in vascular modeling studies [61,13] or for a small number of cores [42]. Another approach to reduce costs would be the use of grids of different resolution, a fine mesh for discretizing the bidomain equations to capture the steep wave fronts, and a coarser mesh for discretizing the deformation equation, as mechanical solutions are supposed to be much smoother in space.

Departing from our previously developed bidomain solver infrastructure for which strong scaling properties have been reported in detail [82,91,80], we aimed to achieve similar scalability for solving nonlinear deformation problems. As shown in Fig. 9, left panel, scaling efficiencies of 51%, 47% and 58% (Table 3) could be achieved when using 1024, 4096 and 8192 cores for executing the CM, MM and FM model, respectively. Remember that the coarse model (CM) was discretized at an average $\hat{h}$ of $\approx$ 880 µm, the medium model (MM) at $\hat{h}$ $\approx$ 440 µm, and the fine model (FM) at $\hat{h}$ $\approx$ 220 µm, see also Table 2. Scalability of the two major cost factors, matrix assembly and solving of the linear systems, are shown in Fig. 9. As expected, matrix assembly scales almost linearly as this is an embarrassingly parallel problem. Solver scalability levels off when hitting a critical surface-to-volume ratio between communication, $\propto$ to the surface of the local partition, and compute cost, $\propto$ to the volume of the local partition. When compared to the scalability of solving the monodomain equation, the much better scalability of the deformation solver becomes apparent. Noticeable degradation in scalability starts at lower $N_p$ of 256, 1024 and 4096 cores for CM, MM and FM model, respectively. This is due to the less favorable compute load to communication ratio of a scalar PDE with one degree of freedom per node as compared to the three degrees of freedom.
per node in the deformation equation. Moreover, a significant portion of the mechanics solver work is dedicated to matrix assembly which is computationally quite expensive, but highly scalable as there is no communication cost involved in this step.

While for electromechanical problems the combination of monodomain and deformation suffices, there are applications of practical relevance where a bidomain formulation may be more appropriate, such as investigating the influence of deformation on the T-wave of the ECG [56]. For this sake scaling data for bidomain simulations were added as well as a reference. As shown in Fig. 9, right panel, scalability of the elliptic and parabolic solve is comparable but solving the elliptic PDE is roughly one order of magnitude more expensive than solving the parabolic PDE. Solving the biophysically detailed GPB+LN model is more costly than the parabolic solve, but since state variables \( \eta \) do not diffuse, the ODE solver step is embarrassingly parallel, which is reflected in a linear scaling graph.

A note of caution is warranted when comparing execution times for EP and mechanics, as a much finer time step of \( \Delta t = 5 \mu s \) was used for EP than for mechanics with \( \Delta t = 1 \text{ ms} \), that is, the EP problem was solved 200× more frequently than the deformation problem. Therefore, a single solve step of the deformation problem is 470–560× and 75–160× more expensive than a single solver step of the monodomain and bidomain equation, respectively.

Preliminary data suggest that further significant improvements in scalability are achievable by exploiting shared memory parallelism in a hybrid MPI-OpenMP approach. The main scaling benefits are a result of the shorter latencies of shared memory data access and the more efficient use of interconnect hardware in terms of latency and bandwidth as less messages are passed, each message being of larger size. Another potentially effective way of enhancing scalability is to employ latency hiding Krylov methods [35,36].

While both approaches are promising, a detailed exploration of their suitability for cardiac electromechanics applications is beyond the scope of this study.

7.2. Representing cardiac anatomy

Cardiac anatomy plays a role of pivotal importance in electrophysiological and mechanical function. Computer models therefore aim to represent anatomy as faithfully as possible, however due to limitations in image acquisition, image segmentation and registration techniques and the difficulty in tessellating anatomically complex 3D structures into discrete finite element models of sufficient mesh quality, most reported studies resorted to using simplified, idealized FE models. In earlier modeling studies, the left ventricle was modeled in isolation as an idealized geometry of cylindrical or ellipsoidal shape [14,81]. Later studies used more elaborate finite element models which incorporated anatomically realistic biventricular geometries based on histological sections [40,115,113,105]. Due to the cost of generating detailed anatomical models, only one representative heart per species was available in these studies. While such a “one heart fits all” approach proved to be of great utility for studying generic mechanisms, this approach is limited in its capability to make predictions for individual patients, due to the wide anatomical inter-subject variability of the heart. Therefore, driven by major advancements in magnetic resonance (MR) and diffusion tensor (DT) MR imaging, a new trend emerged over the past few years towards models derived from tomographic medical images. Traditionally, the most widely used approach is based on using cubic Hermite finite elements to approximate the biventricular geometry [86,116,41]. Techniques were developed to register cubic Hermite template models with segmented medical image stacks [66]. More recently, the cubic Hermite method has also been applied to generating models of the thin walled atria, but significant processing efforts and manual steering is necessary [37]. The main advantage of this cubic Hermite element based method is its capability to approximate cardiac geometry with a limited number of finite elements, keeping compute and memory costs tractable even when using modest computing hardware. However, with regard to creating biophysically detailed and anatomically accurate models coarse mesh approaches pose three challenges. Firstly, they are limited in their ability to capture fine anatomical structures, including papillary muscle, thin walled atria and trabeculation. Secondly, they are limited in their ability to capture regional heterogeneities, including ion channel gradients, peri infarct zones, local fiber dispersion and vasculature. Finally, most approaches rely on the adaptation of template meshes which may not conform to the breadth of patient anatomies. Moreover, finding an optimal tradeoff between geometric fitting accuracy, mesh quality and stability for coarse cubic Hermite based models under large deformations can be challenging [67]. These issues can all be addressed through the use of fine resolution unstructured tetrahedral or hybrid meshes [100,42], generated directly from image segmentations. The ability to generate models automatically from segmentations improves the robustness of the model generation process. This will be beneficial for future clinical applications and will pave the way towards making models available to non-expert users.

Advanced image acquisition and processing [93,13] as well as unstructured mesh generation pipelines [95] have been developed for such purposes. These methods permit the automatic construction of anatomical FE models which represent individual geometries at an unprecedented level of geometric fidelity. So far, such detailed models have been used in electrophysiological modeling studies [12], but not yet in any mechanical modeling studies. Gurev and co-workers reported a method which combined an anatomically detailed biventricular tetrahedral FE geometry model for simulating electrophysiology, with a cubic Hermite FE model for mechanics [41]. The method allows wrapping a cubic Hermite template model around the detailed tetrahedral FE model. While efforts were made to minimize differences between the meshes, due to the significantly lower level of detailed of the mechanical model, achieving a perfect overlap was impossible. This increases the complexity, of the model generation and simulation process, particularly in strongly coupled electro-mechanical sim-
ulations where projection operators would have to be constructed to transfer data between two FE spaces of imperfect overlap.

An alternative method based on CAD reconstruction of hearts, which uses a four chamber representation plus the attached large vessels similar to that used in this study, has been reported recently [5]. As details on the model generation process have not been published yet, it remains to be seen whether a CAD based approach is viable for generating tomodigraphically reconstructed FE models. This study has similarities with ours, in particular that \( P_1 - P_0 \) tetrahedral elements were used, facilitating an accurate approximation of the gross anatomy and the use of the same FE mesh for both electrophysiology and mechanics. The major difference of this study from ours is that the low resolution used precludes the use of physiologically realistic models of cellular dynamics such as the GPB+LN model used in our study.

The four chamber heart model used in our study represents all details which can be delineated in a standard clinical image stack. The mesh generation methodology draws from previously developed tools [95], which permits, in theory, to represent any discernible structure of the heart down to the microscopic level [93]. All FE meshes in this study were generated fully automatically, without any manual intervention. Any anatomical deviations of the model are within the uncertainty of the segmentation. Generating the CM, MM and FM meshes on a desktop computer took 51, 81 and 205 minutes respectively.

7.3. Bidirectional coupling

From a physiological point of view, strong or bidirectional coupling is important as this mediates acute and regulatory MEF, i.e., changes in electrophysiological state of a cell as a function of tissue deformation, measured as stretch ratio \( \lambda \). From a modeling point of view, strong coupling adds technical complexity as electrophysiology and mechanics cannot be treated independently, necessitating the solution of a coupled mono- or bidomain and deformation equation system. Three major mechanisms mediate MEF, by which deformation alters the electrophysiological state of the tissue: (i) deformation alters geometry and thus electrical gradients \( \nabla \phi_i \) and \( \nabla \phi_e \) respectively; (ii) currents through stretch activated channels \( I_{\text{SAC}}(V_m, \lambda) \) directly alter transmembrane voltage \( V_m \) which indirectly influences upon cellular signaling; (iii) stretch alters binding affinity of calcium to troponin C which influences cytosolic calcium transients and diastolic calcium levels. As calcium is involved in many important signaling pathways in a cell this effect may play important roles in both acute and long term regulation of contractility. While our framework would easily accommodate all three MEF pathways, for the sake of testing the scalability of the developed solver methods we restricted ourselves to the feedback through calcium binding affinity of troponin C. In the single heart beat scenario, effects on electrical gradients are very small as electrical gradients are large only prior to contraction, but are very small during contraction when cells are at the plateau of the action potential or slowly repolarize \([120]\). While the existence of stretch activated channels in cardiac muscle has been postulated \([52]\), in models they are largely treated phenomenologically as a “black box”, as the molecular mechanisms which relate stretch to channel conductance remain poorly understood. Various mathematical structures for \( I_{\text{SAC}} \) currents have been published, but parametrized for other species than humans \([109]\) or without parametrization at all \([64]\).

As illustrated in Fig. 10, distribution of stretch ratio \( \lambda \) and active stress \( S_a \) during a heartbeat is quite heterogeneous. This is particularly apparent around the apex where the ventricles were elastically anchored, as well as along sites of transition between actively contracting myocardium and passive structures such as the superior vena cava or the interface between ventricles and atria which were modeled as mechanically passive material in this study. Differences between strong coupling case TnC(\( \lambda \)) and weak coupling case TnC(TnC(\( \lambda \)) in terms of \( \lambda \) and \( S_a \) were rather minor (Fig. 10). Overall, stresses were slightly higher in the weak coupling case. This was due to the length dependent calcium binding affinity being higher compared to the TnC(\( \lambda \)) case where under the given loading conditions and in absence of any caviary pressures the stretch \( \lambda \) was smaller than 1 during contraction, entailing a comparably lower calcium-bound [TnC] and thus less force generation. As length dependence of active stress in the LN model is modeled as \( S_a \propto h(\lambda) \times \text{XB}(\text{TnC}(\lambda)) \) where \( h(\lambda) \) and XB refer to filament overlap and fraction of force generating crossbridges, in the weak coupling case length dependent force generation is mediated only by filament overlap \( h(\lambda) \).

The observed heterogeneity in \( \lambda(\chi) \) translated into a marked heterogeneity in cytosolic calcium [Ca\(^{2+}\)](\( \chi \)) only in the strong coupling case TnC(\( \lambda \)), whereas in the weak coupling case a TnC(TnC(\( \lambda \)) heterogeneity at the same instant in time was marginal (compare left and right in [Ca\(^{2+}\)] panel of Fig. 10). It is important to note that this effect on calcium heterogeneity can only be captured in a strongly coupled model. In most weak coupling scenarios, EP models such as the GPB model are coupled with stress models such as the LN model in a way where calcium buffered [TnC] is tracked by both EP and stress model using different equations. That is, the GPB model tracks TnC concentration based on Eq. (54) and the LN model based on Eq. (55). This is not only inconsistent from a physical point of view, such approaches may also miss a potentially important feedback mechanism. While this study is far from being an exhaustive evaluation the effects of strong coupling, as this was beyond our scope, these results suggest that such effects may be non-negligible, depending on the particular question under study.

Furthermore, it is assumed that deformation is governed by slower time scales and that solutions are much smoother in space compared to electrophysiology which is characterized by fast transients and steep wave fronts. This notion, if correct, suggests that coarser spatio-temporal discretization might be suitable, thereby reducing computational cost when solving the deformation equations. The exact choice of time and space step for mechanics and the resulting errors are likely to depend on the scenario to be modeled and the level of numerical accuracy one aims to achieve. While exhaustive
convergence studies are missing for the complex models as used in this study, numerical work with simplified 1D strand models suggest that spatio-temporal discretization in a strongly coupled scenario also requires fairly fine time and space steps [120]. It therefore remains to be seen whether or not the use of a coarser mesh for mechanics pays off in terms of reduced execution time and overall modeling cost. Additional costs would be incurred due to the need of generating two meshes of different spatial resolution which, ideally, perfectly overlap in space. In practice, this is difficult to achieve when modeling complex structures such as an anatomically realistic four chamber representation of the human heart. Due to the complex geometry, the surfaces of both meshes will be different, requiring extrapolation to transfer coupling variables between the two grids. In a parallel context, a coarser mesh for mechanics will be partitioned differently from the mesh used for EP, causing an increase in the communication cost of the projection operation. Using the same mesh for both EP and mechanics, as implemented in this study or as used by others [5], may be less efficient in weak coupling scenarios where a single static offline projection of data such as activation time or calcium transients can be used as an input for mechanics, but in strong coupling scenarios this may not be the case. Thorough numerical convergence testing studies are necessary in future work to address the issue of spatio-temporal discretization for strongly coupled electromechanical simulations. Such convergence studies will lead to a robust quantification of errors, allowing us to find a balanced tradeoff between computational cost and accuracy.

7.4. Convergence testing

In an ideal numerical scheme for solving PDEs, the solution depends only on the chosen parameters, but not any other factors such as the spatial discretization step \( h \). While it is feasible to construct such numerical schemes, in practice this is rarely the case. From an applied point of view, the relevant aspect is not to drive discretization errors to zero. Rather, it is more important to approximate the discretization error and contrast it with the typically large uncertainties in parameters and model assumptions. By considering all these factors, we are able to find a balanced tradeoff between computational cost, model parametrization and accuracy.

In the EP modeling communities this issue has been addressed, starting with very early pioneering studies [94], by spatio-temporal convergence testing to choose \( h \) and \( dt \) appropriately. An extension of this approach brought the introduction of \( N \)-version benchmarking to verify whether \( N \) different and independent implementations yield comparable results [84]. For solving the monodomain equations it has been demonstrated that all \( N = 19 \) independent codes used by the EP modeling community yielded the same result, if \( h \) and \( dt \) were chosen to be sufficiently small. For electromechanical models, convergence testing or \( N \)-version benchmarking has not yet been established. In a recent study an \( N \)-version benchmark was carried out for solving a cardiac deformation problem, but mechanical deformation was driven by a time triggered active stress models, i.e. a rather simplified weakly coupled model was considered. Moreover, with \( N = 2 \) the number of independent implementations was quite small. In a strong coupling context, as the modeling work in this study, testing for convergence of electromechanical simulations is more challenging as any errors in the driving EP solution will feed into the solution of the deformation problem. With biophysically detailed models of cellular dynamics such as the GPB model, very fine spatial resolutions are key to keep discretization errors within acceptable bounds. To minimize the impact of errors of the electrical activation pattern on mechanical deformation, conductivities were fitted to travel at prescribed velocities for each of the CM, MM and FM models. As shown in Fig. 5, this approach indeed results in very similar activation sequences, though deviations are visible in the CM model, indicating that this approach may not be suitable for even coarser discretizations as done by others [42,5]. While we did not perform a thorough convergence study, as the complex geometry used did not lend itself well to such an investigation, the effect of mesh resolution on global mechanical performance was evaluated by comparing the computed left ventricular cavity volume over time between models. As shown in Fig. 11, volumes at end-diastole are not discernibly different, but noticeable discrepancies arise during systole, which do not change monotonically with \( h \).

In weak coupling cases using activation time triggered mechanics [42] or when simplified phenomenological models of FitzHugh–Nagumo type are used [5], electromechanical convergence is likely to be less of a concern. However, such models are not suitable for studying biophysical mechanisms of excitation–contraction coupling and MEF at the organ scale. The mesh resolutions used in such studies are unsuitable for use with physiologically detailed models of cellular dynamics such as the GPB. At such coarse resolutions artificial conduction block can occur, impeding action potential propagation. This is not the case with simple phenomenological models as the upstroke of the action potential is orders of magnitude slower. As shown in [5], the upstroke of the used EP model lasts around 50 ms whereas the upstroke duration of a biophysically realistic model lasts around 1 ms. Such slow upstroke velocities translate into less steep wave fronts, which can be resolved on coarser meshes, reducing computational costs significantly.

In the context of convergence testing of strongly coupled electromechanical models the presented numerical approach is of pivotal importance. The demonstrated excellent scaling properties enable the use of ultra-high spatial resolutions which are necessary for generating high fidelity reference simulation results. The complex models used in this study, while well suited for scaling benchmarks, are less suitable for rigorous convergence testing. Simpler geometries, as used in standard EP benchmarks [84], lend themselves more easily to such tests since spatio-temporal resolution can be varied over domains of exact overlap in space.
7.5. Limitations

The focus of this study was to design and implement a strongly scalable iterative solver which is sufficiently efficient to enable execution of high resolution strongly coupled whole heart EM models. Other issues which have not been addressed and may limit the applicability of the proposed method in different modeling scenarios include:

1. A penalty formulation and element-by-element static condensation is used to ensure a nearly incompressible material behavior. Depending on the penalty parameter $\kappa$ this may lead to very ill-conditioned problems. However, with the choice of $\kappa = 1000$ kPa (Table 1) and the design of the AMG method as described in Section 4, we were able to solve the linearized systems within a reasonable number of iterations (Fig. 8). Nevertheless, as the iteration numbers of the linear solver increase with $\kappa$, a mixed approach [4,71] may be preferable.

2. Simple linear tetrahedral elements were employed for both the electrophysiological and the mechanical models. While this is a valid and accepted approach for the bidomain equations, the resulting $P_1 - P_0$ elements for solid mechanics are known to be prone to volumetric and shear locking. A mixed formulation with quadratic shape functions for displacement and a linear hydrostatic pressure field, i.e. $P_2 - P_1$ tetrahedral elements, would help to avoid these problems [28,32,87]. Such an approach is well suited for large deformation problems and shows the same performance as a mixed formulation using $Q_2 - P_1$ hexahedral elements [19]. However, as we wanted to stick to our principle of utilizing the same grid for both physics, the usage of higher order elements for mechanics would lead to a highly unbalanced relation between the size of the mechanical and electrical systems. Hence, a reasonable employment of $P_2 - P_1$ elements requires different spatial resolutions for mechanics and electrics and an appropriate projection of coupling data between the electrical and the mechanical grid. The construction of such projections and therefore the mixed approach are beyond the scope of this specific paper and left to future work.

3. A Langendorff perfused setup was modeled where the ventricles were not pressurized. This is different from a working heart model where isovolumetric phases are enforced when valves are closed and pressure–volume relations are governed by the cardiovascular system or the atria when valves are open.

4. In vivo, the heart is enveloped by the pericardium, which is elastically attached to surrounding tissue via the mediastinum. This important aspect of cardiac function has not been accounted for [32].

Addressing some of these limitations yield block-structured systems that require special treatment. Such systems can be solved iteratively using specially designed AMG methods [51,70,121] or properly chosen block preconditioners, generally based on a Schur complement approach [28,29,42]. While the solver approach presented here is not directly applicable to such problems, it still constitutes an important building block for solving block systems. However, conceiving AMG solvers for block system or building Schur complement preconditioners is beyond the scope of this study.

Acknowledgements

This project has received funding from the Austrian Science Fund (FWF) within the SFB Mathematical Optimization and Applications in Biomedical Sciences and from the European Union’s Seventh Framework Programme for research, technological development and demonstration under grant agreement No. 611232. We acknowledge PRACE for awarding us access to resource SuperMUC based in Germany at LRZ (grant CAMEL), and, partially, ARCHER based in the UK at EPCC (project e384).


