A VARIATIONAL MULTISCALE COMPUTATIONAL FRAMEWORK FOR REACTION-DOMINATED THERMO-CHEMO-MECHANICAL PROCESS MODELING IN MULTI-CONSTITUENT MATERIAL SYSTEMS

BY

MARCELINO ANGUIANO CHAVEZ

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Civil Engineering
with a concentration in Computational Science and Engineering
in the Graduate College of the
University of Illinois Urbana-Champaign, 2021

Urbana, Illinois

Doctoral Committee:

Professor Arif Masud, Chair and Director of Research
Professor C. Armando Duarte
Professor Oscar Lopez-Pamies
Professor Albert J. Valocchi
Assistant Professor Nikhil Chandra Admal

ABSTRACT

This dissertation develops a computational framework for modeling multi-constituent material systems characterized by the transport of reacting fluids through deformable solids, and their coupled, nonlinear, thermo-chemo-mechanical response in the reaction-dominated regime. This is accomplished through two major components of the work: (i) new robust variational multi-scale numerical methods that are consistently derived, and (ii) models for multi-physics processes in multi-constituent materials.

New robust numerical methods are developed via the variational multiscale (VMS) framework. Through the concept of fine scales in VMS, unresolved physics are recovered and embedded at the coarse scale level, improving stability and accuracy of the method. Focus is placed on fine scales that do not vanish at element boundaries (so-called "edge bubbles"). Using edge bubbles and an explicit time integration algorithm, a VMS Discontinuous Galerkin (VMDG) method is derived for multi-domain problems in elastodynamics where different subdomains can be solved synchronously and concurrently with minimal sharing of information. In addition, a new VMS method is introduced for the reaction-dominated regime of the diffusion–reaction equation. The proposed fine-scale basis consists of enrichment functions that may be nonzero at element edges. The method captures sharp boundary and internal layers, suppresses spurious oscillations, and better satisfies the maximum principle as compared to other existing methods. A priori mathematical analysis of the stability and convergence of the method is presented, and optimal rates of convergence are verified numerically.

The numerical methods developed in this work may be applied to many reaction-diffusion systems in mathematical models for coupled thermo-chemo-mechanical phenomena arising from different theoretical frameworks. Here, a model for thermo-chemo-mechanical response of open solid-fluid systems is presented in the context of mixture theory. Derivation starts from constituent-wise equations for balance of mass, momentum, and energy, accounting for energy in formation and breaking of chemical bonds. Interactions between different constituents are captured through interaction terms as per locally homogenized mixture theory. Satisfaction of the second law of thermodynamics is achieved by providing constitutive equations that guarantee non-negative entropy production. Resulting mathematical models yield transient diffusion-advection-reaction problems posed by systems of coupled, nonlinear, second-order partial differential equations (PDEs), whose solution require stable numerical methods.

Several numerical studies are presented to highlight stability, accuracy, and other features of the newly developed variational multiscale methods and thermo-chemo-mechanical models. Tests involve hypothetical as well as realistic materials with boundary layers, advancing reaction fronts, chemical swelling, and fingering phenomena.

ACKNOWLEDGEMENTS

I am unmeasurably indebted and thankful to my PhD advisor, Prof. Arif Masud, for sharing his passion and expertise in computational mechanics with me. His insight, guidance, patience, encouragement, and support made this work possible. I am also immensely grateful for his mentorship during my PhD, which helped me to grow academically and professionally, and I am thankful for the many stimulating conversations we shared.

I express my gratitude to the members of my doctoral committee: Prof. C. Armando Duarte, Prof. Oscar Lopez-Pamies, Prof. Albert J. Valocchi, and Prof. Nikhil Chandra Admal, for their valuable feedback and support. I am also grateful for the courses that they and other Faculty members taught, which made it possible for me to gain the required level of understanding in many topics relevant for my dissertation work.

I gratefully acknowledge the financial support provided by the Air Force Research Laboratory, National Science Foundation, Sandia National Laboratories, and the Department of Civil and Environmental Engineering at UIUC. I also thank the Illinois Campus Cluster Program for facilitating the computational resources needed to complete this dissertation.

I thank Joan Christian, Marissa Miller, and all the members of the administrative staff in the CEE Department at UIUC. They have all been incredibly helpful and kind to me in navigating the required steps at many stages during completion of my MS and PhD degrees.

I thank my colleagues in Prof. Masud’s research group with whom I had the pleasure to work: Dr. JaeHyuk Kwack, Dr. Harishanker Gajendran, Dr. Ahmad Al-Naseem, Dr. Pinlei Chen, Dr. Lixing Zhu, Soonpil Kang, Shoaib A. Goraya, Ignasius P. A. Wijaya, Ian Tuttle, and Sharbel Nashar. Their helpful discussions and feedback made my experience more valuable. I also extend my gratitude to other colleagues in the department for their helpful discussions and camaraderie: Dr. Oliver Giraldo Londoño, Dr. Fernando Gómez Sánchez, Dr. Aditya Kumar, Dr. Nathan Shauer, Bhavesh Shrimali, Kamalendu Ghosh, Alfredo Sánchez Rivadeneira, and others in the Computational Mechanics Lunch group, as well as other peers who were active members of the Structural Engineering Graduate Student Organization. I thank all of them for their diversity of ideas and viewpoints, as well as the sense of community we built.

To my wife, Eva. To my mother, Julieta. To my father, Marcelino (1956-2009). To my brothers, Guillermo, Carlos, and Enrique. And to all the people rooting for me on this journey.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1

CHAPTER 2: CHEMO-MECHANICAL COUPLING AND MATERIAL EVOLUTION IN FINITELY DEFORMING SOLIDS WITH ADVANCING FRONTS OF REACTIVE FLUIDS...22

CHAPTER 3: REDUCED MIXTURE MODEL AND ELASTIC RESPONSE OF CHEMICALLY SWOLLEN SOLIDS: APPLICATION TO SI OXIDATION AND LITHIATION...67

CHAPTER 4: MIXTURE MODEL FOR THERMO-CHEMO-MECHANICAL PROCESSES IN FLUID-INFUSED SOLIDS 95

CHAPTER 5: EXPLICIT VARIATIONAL MULTISCALE DISCONTINUOUS GALERKIN METHOD..151

CHAPTER 6: MODELING OF STEEP LAYERS IN SINGULARLY PERTURBED DIFFUSION-REACTION EQUATION VIA FLEXIBLE FINE-SCALE BASIS 187

CHAPTER 7: A VARIATIONAL MULTISCALE METHOD FOR TRANSIENT COUPLED THERMO-CHEMO-MECHANICAL PROBLEMS 249

CHAPTER 8: CONCLUDING REMARKS AND FUTURE DIRECTIONS 283

APPENDIX A: DETAILS OF ADDITIONAL DERIVATIONS IN CHAPTER 4................289

CHAPTER 1:
INTRODUCTION

1.1 Motivation

Coupled thermo-chemo-mechanical physical phenomena involving the transport of fluids through deforming solids arise in many problems of scientific and engineering interest. A first example is in the aerospace industry, where thermally resistant metal and ceramic materials are exposed to extreme oxidizing environments. Specifically, gas turbine and rocket engines employ thermal barrier coating systems of superalloys and ceramic matrix composites, the latter of which is also used in vehicle-structure thermal protection during reentry [1]. Some examples of thermal oxidation of aerospace ceramic materials can be found in [2–6]. Thermal oxidation is also a key process in the manufacturing of power and semiconductor devices [7–11], which is particularly important to develop next-generation devices that find applications in high-efficiency electronics, high-power electronics, and automotive parts [12]. Another relevant area is in the field of micro-self-actuators [13,14], where a functional mechanical response is triggered by other external stimuli (thermal, chemical, or electromagnetic). Other thermo-chemo-mechanical problems in active areas of research include biological growth [15–20], lithiation of electrodes in lithium ion batteries [21–25], processes in solid-oxide fuel cells [26,27], the response of elastomeric gels [28–30], Ziegler-Natta polymerization [31], and reactive porous media flows [32,33], among others.

In many of these systems, the reactive and thermal processes induce internal stresses and large local deformation (e.g., through thermal and chemical expansion) as well as material evolution due to chemical reactions. These factors may lead to material degradation or structural failure at the micrometer scale and above [34–36], which can extend all the way to geological length scales for problems in porous media flows. Therefore, it is important to develop accurate models that can be descriptive and predictive of the coupled response of the material systems under consideration. Given the length-scales of the problems of interest (from electronic and biological components in the scale of micrometers, to geophysical phenomena in the scale of kilometers), it is appropriate to model the domain of interest as a continuum. Such mathematical models yield problems that involve diffusion-advection-reaction processes, which are posed as initial-boundary value problems (IBVPs) for systems of coupled second-order partial differential equations (PDEs) [37–39]. The complexity of the governing system of equations increases with an increase in the mechanical deformation, the rate of chemical reactions, and the rate of mass transport. Evolving geometric and material nonlinearity, as well as coupled chemo-thermo-mechanical effects give rise to spatially localized phenomena, namely, boundary and moving internal layers, shear bands, and steep gradients that appear at the reaction fronts. Localized effects are also present in problems with discrete interfaces (e.g., material interfaces, fractures, and partitioned domains). Exact analytical solutions to IBVPs of this form are not known but for the simplest cases. Invariably, numerical approximations to the solutions are needed to generate descriptive and predictive results of engineering significance. For the class of problems described above involving the transport of fluid through a deforming solid domain, the finite element method (FEM) [40–42] is selected as the underlying method for finding numerical solutions for solid and fluid fields.

It is well known that the solutions obtained via the Galerkin FEM develop spurious oscillations in the presence of localized phenomena, engendered, for example, by boundary and/or internal layers in singularly perturbed second order PDEs [43,44]. These oscillations introduce global polluting error in the numerical solutions. Furthermore, in the case of coupled system of PDEs, solution with mixed FEM is possible only for specific combinations of approximating spaces, so that the discrete problem remains well posed by satisfying the *inf-sup* condition [40,41,45].

Therefore, the goal of this dissertation is to develop robust methods for numerical solution of coupled, nonlinear, transient second-order partial differential equations (PDEs) of diffusion-advection-reaction type. These methods are to be applied to problems of engineering interest involving material systems composed of deforming solids permeated by reactive fluids and subjected to different physical phenomena, such as mechanical, chemical, and thermal effects. In that regard, this dissertation aims to lay some of the foundational work to develop a framework for *simulation-based design* of engineered materials, manufacturing processes, devices, and structures in *extreme* environments. This is accomplished through two major components of the work. One component is the development of new variational multiscale numerical methods that are consistently derived and implemented, which enable computer simulation that is accurate, stable, robust, and cost-efficient. The other component is the adoption and extension of thermodynamically consistent models for multi-physics, multi-constituent, and multi-scale processes. The development of models along with methods and algorithms for their numerical implementation is crucial to gain important insight into the complex physical processes and to develop tools for engineering applications.

1.2 Background

1.2.1 Continuum Thermo-Chemo-Mechanical Models

Thermodynamically consistent formulations for multifield, nonlinear problems have been the subject of profound study by many authors spanning many years and we refer to seminal works in [37,46] for a historical perspective and a treatise on the subject. Significant contributions have also been made in the past decade to develop fully coupled, continuum models for thermo-chemo-mechanical phenomena that are thermodynamically consistent. Here, a select few are highlighted, and we refer to those and the references therein for a broader perspective on the subject. A coupled permeation and finite-deformation was treated in an isothermal context in [47], and a full thermo-chemo-mechanical coupled theory was presented in [48]. These works were applied to study problems in elastomeric gels [30,49] and oxidation of high-temperature alloys [5,50] through numerical simulation. Other recent work involves thermo-chemo-mechanical model for polymer curing [51], and thermodynamically consistent chemo-mechanical model for polymer oxidation [52]. A linear theory is developed in [53]. The work of Kannan and Rajagopal [54] provides a rigorous and detailed presentation on the thermodynamics of chemical reactions in multi-constituent systems.

Mathematical and computational analysis of problems where fluid and solid constituents coexist concurrently requires models that take into account disparate physical phenomena and methods that represent the interacting constituents either discretely or via homogenization techniques. Homogenization methods and mixture theories provide a framework for the modeling of coupled material systems that are comprised of more than one constituent material. These theories lead to computationally economic solution methods for this class of problems that would otherwise require discrete modeling approaches to represent each of the individual constituents along with the modeling of discrete material interfaces.

It is important to acknowledge some of the theoretical frameworks employed for modeling multi-phase systems and that use of the idea of diffuse interfaces. The seminal paper by Cahn and Hilliard [55] derives an equation for modeling phase transition and phase separation. The work by

Gurtin [56] discusses important equations describing the evolution of two-phase systems: the Ginzburg-Landau equation (also known as the Allen-Cahn equation), and the Cahn-Hilliard equation. In that work, this equations are a consequence and special cases of the introduction of a balance of configurational forces (or "microforces") as part of the basic governing laws (see [57] for a treatise on configurational forces). This framework is also known as the phase-field approach, which has been applied to couple Cahn-Hilliard equation with mechanical deformations in [58,59]. Furthermore, Heida et al. [60] have derived phase-field models for multi-phase fluid systems within the context of mixture theory and obtain Cahn-Hilliard type equations as special cases. These very brief overview suggests that one may start from different lines of thought and, under the appropriate assumptions, one still may arrive at equivalent models. Some of the work in this dissertation is within the context of mixture theory, which is one of several frameworks that lead to a coupled thermo-chemo-mechanical models that are thermodynamically consistent. However, it is emphasized that the numerical and computational methods derived and developed in this dissertation may also be applied to obtain numerical solutions to thermo-chemo-mechanical IBVPs that emanate from other theoretical frameworks to describe the physics of the problem.

As stated, this dissertation considers continuum level models for thermo-chemo-mechanical phenomena developed within the context of mixture theory, which provides a framework for modeling complex material systems composed of multiple constituents. The ideas associated to mixture theory emanate from as far back as the work of Fick [61] and Darcy [62]. The works of Bowen [63], and Atkin and Craine [64,65] are classical references on the subject. The book by Rajagopal and Tao [39] has also become a standard reference in the mixture theory literature. A more recent overview and account of historical developments is found in [66]. Ideas from mixture theory have been applied for problems in porous media flows [67], heterogeneous catalysis [68], biological growth [17–19], curing in polymer matrix composites [69], Ziegler-Natta polymerization [31], and diffusion of chemically active fluids within elastic solids [70], among others [65]. In addition, thermodynamically consistent, multiphysics mixture models accounting coupling between chemical, thermal, and/or mechanical effects based on mixture have been presented [71,72], where constitutive relations are derived by assuming maximization of the rate of entropy production. Stabilized numerical methods have also been developed to allow their numerical implementation in a finite element setting [70]. As noted in some of these references, the assumption of *co-occupancy* of the domain by the constituents is a strong one, which should be understood in a homogenized sense. However, this is tied to some outstanding issues within mixture theory related the ambiguity in the definition of some mixture quantities, as well as in the imposition of boundary conditions, as identified by [66,73]. Some work has been done to address these issues [74].

One reason that makes mixture theory appealing for the class of problems under consideration in this dissertation is that the conversion of the solid constituent from a pristine state to a reacted stated may be irreversible and is due to chemical reaction with an additional constituent – a fluid reactant. Mixture theory allows to model the transport of this fluid constituent within the deforming (and transforming) solid through its own set of balance laws for mass and momentum. A similar situation (although not explicitly studied in this work) is that in which the evolution of the solid may be mediated by a catalyst, which is not incorporated into the solid in its final state, but is needed to trigger or accelerate the phase transitions.

A key feature of mixture theory is that the classical balance laws for each constituent are augmented by including *interaction terms* that account for the interaction of each constituent with every other constituent. In a sense, mixture theory can be thought to facilitate physics-based, reduced-order-models for multi-constituent systems since it allows to capture the interactions between constituents at the continuum level, while the geometric intricacies of the interfaces between constituents at the smallest scales need not be explicitly modeled. The terms that capture the inter-constituent interactions are "reaction-like" (i.e., contain zeroth-order terms, which depend on the unknown fields but not their derivatives), so the governing equations are, in general, nonlinear advection-reaction or diffusion-advection-reaction equations.

1.2.2 Numerical Methods

Initial-boundary value problems (IBVPs) posed by PDEs are commonly solved using numerical methods such as the finite difference (FD) method an finite volume (FV) method, which is widely used in computational fluid dynamics (CFD) for solution of systems in the form of conservation laws. In the field of solid mechanics, where the mechanical deformation and stresses are of interest, the prevalent method is FEM. In this work, the FEM is chosen as the underlying numerical method because of its sound mathematical foundation, and because in the class of physical problems under consideration, the transport and reactive phenomena occur within a deforming solid domain, whose mechanical response is also of interest.

As it was touched upon earlier in this manuscript, the Galerkin FEM develops spurious oscillations engendered by boundary and/or internal layers in singularly perturbed second order problems (i.e. when the coefficient of the second-order term is relatively much smaller in comparison to the coefficients of lower order terms). This includes the cases of advection-dominated problems (dominant first-order term), and reaction-dominated ones (dominant zeroth-order term). Another source of error is poor approximation of weak discontinuities in the exact solution at material interfaces with complex geometry. These spurious oscillations are in violation of maximum principles for elliptic PDEs [75,76] and represent non-physical response when the solution fields represent strictly positive quantities (e.g., concentration and absolute temperature). Furthermore, in systems of coupled PDEs, oscillations in the numerical solution of one field may trigger instabilities in other fields and the entire system [77], causing computations to diverge and no numerical solution is found. An analysis that shows the origin of these spurious oscillations in the singularly perturbed diffusion-reaction equation can be found in [44,78]. A similar analysis for the origin of the spurious oscillations in the diffusion-advection equation for the advection-dominated regime is presented in section 7.6 of [42].

Several methods have been proposed to mitigate these issues with the Galerkin FEM. A historical perspective on stabilized methods is presented in [79]. One of the first such methods is the well-known Streamline Upwind Petrov-Galerkin (SUPG) [43], SUPG with discontinuity capturing (DC) [80], and the Galerkin Least-Squares (GLS) [81] methods. Although a considerable portion of this literature is mainly devoted to the advection-dominated case of the advection-diffusion equation, several methods have also been proposed to mitigate the issue of spurious oscillations in the reaction-dominated regime of the diffusion-reaction or advection-diffusion-reaction equations. These include an extension of SUPG with DC for the reaction-dominated regime as well [82], the Galerkin Gradient Least-Squares (GGLS) method in [44], and a modification of the GLS method for the case when reaction is dominant, in addition to a combination of the GLS and GGLS methods, as presented in [83]. The key idea underlying GGLS was to add a term to stabilize the discrete solution in the H^1 norm. This was followed by the Adjoint or Unusual Stabilized methods (USFEM) [84], wherein an $L_2(\Omega)$ term was subtracted to control the oscillations around steep fronts in reaction dominated cases. These methods are based on suitable selection of the stability parameter, which is shown to be critical for the stability of the method. Subsequent works on the topic include Sub-Grid Scale methods (SGS) [85], the

Variational Multiscale (VMS) method [86], Residual-Free Bubbles (RFB) method [87–89], and the use of bubble functions and Green's functions in fine-scale models [84,85,90]. Comparisons of the VMS, SGS, SUPG, and GLS for the diffusion-advection-reaction equation were done in [91,92], where, in particular, the reaction-dominated case was considered.

Some other classes of methods that are effective in suppressing spurious oscillations around boundary layers are found in the literature: The Galerkin Projected Residual (GPR) method was introduced in [93], which involves stabilization terms emanating from multiple projections of the residual weighted by coefficients that depend on matrices emanating from USFEM and GLS type methods. The Galerkin Enriched FEM (GEM), and Petrov-Galerkin Enriched Method (PGEM) (see [94] references therein) modify the standard FEM space of piecewise Lagrange polynomials with enrichments emanating from the solution of the strong form of the governing PDE in the elements and on element boundaries. In addition, corresponding adjoint stabilized methods (SGEM and SPGEM) were also presented. Furthermore, Finite Increment Calculus (FIC) has been used to derive methods such as the High Resolution Petrov-Galerkin (HRPG) (see [95] and references therein), where the built-in discontinuity capturing brings about stability around boundary and internal layers. Some other methods, such as the ones presented in [96,97], enforce satisfaction of maximum principles and non-negative constraints to suppress the non-physical undershoots in solution fields for which negative values have no physical meaning (e.g. concentration and absolute temperature).

1.2.2.1 Variational Multiscale Method

The development of robust numerical methods in this dissertation is based upon the well-established variational multiscale (VMS) framework. The VMS method was first proposed by Hughes and collaborators [85,86]. For an introductory overview of the subject, the interested reader is referred to [98] and references therein. The key idea in VMS is a multiscale decomposition of the solution field and its associated test function into a resolved (coarse) contribution and unresolved (fine) contribution. Several approaches to defining local approximations for the fine scales are presented in [86], including element Green's functions [99,100] and residual-free bubbles [87,101], whose equivalency was explored in [90], as well as hierarchical *p*-refinements. Another approach is the design of stabilization parameters appealing to asymptotic arguments, such as in [91,102]. The uniqueness of the multiscale decomposition was later formalized through the definition of projector operators by Hughes and Sangalli [103].

In this dissertation, we follow the approach used by Masud and collaborators, first presented in [104] and formalized in [105], which employs local polynomial bubbles in the vein of the hierarchical *p*-refinements discussed in [86]. That work has been the basis for stabilized methods for Navier-Stokes equations and turbulence models [106–109], stabilization in non-Newtonian flows [110,111], stabilized methods in incompressible elasticity [112], inelasticity [113], the development of interface coupling and Variational Multiscale Discontinuous Galerkin (VMDG) methods [114–118], and diffusion of chemically reacting fluids through elastic solids [70,77], to name a few representative examples.

The additive decomposition of the test functions into coarse and fine scale contributions, and the linear independence of the scales, leads to segregation of the problem a coarse-scale subproblem and a fine-scale subproblem. The fine scale subproblem is localized within each element and is used to solve for the local fine-scale field in terms of the coarse scale field through one of several possible approaches (element Green's function, Residual-Free Bubbles, polynomial bubbles, etc.). It is emphasized that the driving mechanism for the local or fine-scale problem is the residual of the Euler-Lagrange equation of the coarse scales. Then, the solution of the fine-scale field is embedded into the coarse-scale subproblem. Therefore, the numerical solution obtained in the VMS method accounts for the effect of the unresolved scales on the resolved ones.

VMS methods maintain accuracy and stability of the solution around the localized phenomena. One reason this is successful is because the consistently derived fine-scales are physics-driven, and they provide a vehicle for recovering unresolved "missing" physics and injecting them back into the coarse scale problem. In addition, it has been shown that the stabilized formulations of coupled PDEs admit arbitrary combinations of approximating spaces, circumventing the restrictions imposed by the *inf-sup* condition in mixed-field formulations [107,112,119]. From the computational point of view, these methods are scalable (i.e. easily parallelizable) because the fine scales are defined and solved locally within each element. Finally, the VMS method is equipped with a natural *a posteriori* error estimator through the definition of the fine-scales [86,120].

1.3 Dissertation Outline

The work in this dissertation develops both a mathematical model for thermo-chemo-mechanical coupled response of multi-constituent systems, as well as variationally consistent finite element methods for numerical solution of singularly perturbed second order partial differential equations in the reaction-dominated regime. The progress in these areas is presented in the remaining chapters and opens with Chapter 2, which introduces key ideas and initial work in the two main fronts: mathematical modelling of multi-constituent systems and variational multiscale finite element methods for numerical solution of the governing system of PDEs. Then, Chapter 3 and Chapter 4 present contributions regarding the mathematical modeling of the coupled multi-physics response of solid-fluid multi-constituent systems. Switching gears, the focus of the work in Chapter 5 and Chapter 6 is placed on the contributions regarding the development of robust numerical methods via the variational multiscale framework, utilizing the notion of mathematical fine-scales and exploring fine-scale models that do not vanish at element boundaries. Then, Chapter 7 ties the two components together by extending the ideas concerning numerical methods in Chapter 5 and Chapter 6, and applying them to the solution of initial-boundary value problems arising in the multiphysics response of multi-constituent systems which are described by extending and adapting the thermo-chemo-mechanical model from Chapter 4. Finally, overarching conclusions and directions for continued work are discussed in Chapter 8. An overview of each chapter is given as follows:

- Chapter 2 presents a stabilized method for coupled chemo-mechanical problems involving chemically reacting fluids flowing through deformable elastic solids. A mixture theory model is employed wherein kinematics is represented via an independent set of balance laws for each of the interacting constituents. The constitutive relations for the constituents in the mixture model are based on maximization of the rate of entropy production. When constitutive equations are substituted into the balance laws, they give rise to a system of coupled nonlinear PDEs. For large reaction rates, the system becomes singularly perturbed (reaction-dominated), which may exhibit boundary and/or internal layers. Presence of these features requires stable numerical methods, and we present a variational multiscale (VMS)-based stabilized finite element method for the initial-boundary value problem. Mathematical attributes of the method are investigated via a range of numerical test cases that involve diffusion of chemically reacting fluids through nonlinear elastic solids.
- Chapter 3 further elaborates on the mathematical modeling of chemo-mechanical coupling and material evolution across advancing fronts of reactive fluids from Chapter 2. A detailed analysis of the stress response of the chemically evolving solid is carried out, taking into

account the changes in the reference unstressed configurations of the unreacted and reacted solid. This yields a systematic way to determine the material parameters required for the material constitutive models. The formulation for the fluid constituent is simplified for the case of slow diffusion, which yields a reduced system of governing equations wherein coercivity of the continuum system is inherited by the discrete system. The resulting nonlinear model lends itself to consistent linearization and is implemented in a finite element method. The nonlinear system of coupled equations is solved monolithically using the Newton-Raphson algorithm and quadratic convergence is attained. The method is applied to investigate the bending-dominated response of a silicon wafer during thermal oxidation, and to the problem of cross-section evolution during lithiation of a silicon nanowire.

- Chapter 4 presents a new thermodynamically consistent model for thermo-chemo-mechanical processes in open systems consisting of nonlinear elastic solids that undergo large strains locally and are infused with reactive fluids. The interactions between different constituents of the material system are captured through locally homogenized mixture theory. The development of the formulation starts from constituent-wise balance equations for mass, momentum, and energy. The constitutive relations are derived to enforce non-negative entropy production to satisfy the second law of thermodynamics for the open system. The deformation due to thermal and chemical effects is carried out through a multiplicative split of the deformation gradient. Further assumptions on the fluid and solid material responses are introduced to specialize the model to a class of thermo-chemo-mechanical problems that is applicable to thermal oxidation of metallic and ceramic materials, lithiation of battery anodes, and other related problems. The model is amenable to a numerical implementation through the standard Galerkin finite element method in a computationally cost-efficient way. This is used to study the fully coupled thermo-mechanical, chemo-mechanical, and thermo-chemical effects in problems of interest. In addition, numerical results obtained are in reasonable agreement with experimental data for thermo-chemo-mechanical problems concerning the thermal oxidation of silicon carbide and FeCrAlY bond coat material in thermal barrier coatings.
- Chapter 5 presents a numerical method for developing inter-domain transmission conditions for transient problems in the finite element method where different parts of the

larger domain are assigned to different computational platforms and are solved synchronously and concurrently with minimal sharing of information. This is achieved via two key ingredients. The first is the variational multiscale method, which introduces the notion of mathematical fine scales that, in this chapter, are defined in terms of basis functions that are local within each element but do not vanish at the interface of discontinuity. The second ingredient is decoupling of the solution step for each subdomain via any off-the-shelf explicit time integration algorithm, enabling synchronous and concurrent solution across computing platforms. Such conditions are required for parallel solution of large-scale problems by domain decomposition or when different parts of the domain are modeled by different physical models implemented in separate codes. Such characteristics make the method ideally suited to leverage distributed and heterogeneous computational resources, even when they are part of "loosely" connected networks, such as through cloud computing technologies.

- Chapter 6 presents a new stabilized method for the diffusion–reaction equation which develops sharp boundary and/or internal layers for the reaction-dominated case (i.e. singularly perturbed case), as well as across discrete material interfaces. The method relies on an improved expression for the stabilization parameter that is derived via the fine-scale variational formulation facilitated by the variational multiscale (VMS) framework. The proposed fine-scale basis consists of enrichment functions which may be nonzero at element edges. The derived stabilization parameter enjoys spatial variation over element interiors and along inter-element boundaries that helps model the rapid variation of the residual of the Euler–Lagrange equations over the domain. This feature also facilitates consistent stabilization across boundary and internal layers as well as capturing anisotropic refinement effects. The method is able to better satisfy the maximum principle as compared to other existing methods. We present a priori mathematical analysis of the stability and convergence of the method for the diffusion–reaction equation. Optimal convergence on meshes comprised of linear triangles and bilinear quadrilateral elements are presented for smooth problems, as well as for problems with steep boundary layers. Stability and accuracy features of the method for problems with discontinuous forcing function, internal layers, boundary layers, and weak discontinuities (due to material interfaces in a heterogeneous domain) are shown and its performance on unstructured and distorted meshes comprised of quadrilateral and triangular elements is highlighted.

- Chapter 7 extends the method development carried out in Chapter 6. First, a VMS method is developed for the scalar transient diffusion-reaction equation with material evolution. Important consideration is placed in options regarding the time-evolution of the fine-scale field, as well other algorithmic choices. Then, these ideas are applied to develop a VMS method for an extension of the thermo-chemo-mechanical model presented in Chapter 4. The general VMS method considers multiscale split of both the fluid mass concentration and temperature fields. Numerical results are presented to highlight the features of the method: the so-called "bird's beak" problem is revisited, and fingering phenomena related to solid fuel combustion is studied.
- Chapter 8 presents a summary and concluding remarks of the work developed in this dissertation, as well as a discussion of future directions for continued extension of this research.

1.4 References

[1] N.P. Padture, Advanced structural ceramics in aerospace propulsion, Nat. Mater. 15 (2016) 804–809.

[2] M. Wilson, E. Opila, A Review of SiC Fiber Oxidation with a New Study of Hi-Nicalon SiC Fiber Oxidation, Adv. Eng. Mater. 18 (2016) 1698–1709.

[3] N.S. Jacobson, Corrosion of Silicon-Based Ceramics in Combustion Environments, J. Am. Ceram. Soc. 76 (1993) 3–28.

[4] A.W. Davis, A.G. Evans, Some Effects of Imperfection Geometry on the Cyclic Distortion of Thermally Grown Oxides, Oxid. Met. 65 (2006) 1–14.

[5] K. Loeffel, L. Anand, Z.M. Gasem, On modeling the oxidation of high-temperature alloys, Acta Mater. 61 (2013) 399–424.

[6] J. Smialek, N.S. Jacobson, Oxidation of High-Temperature Aerospace Materials, High Temp. Mater. Mech. (2014) 95–162.

[7] B.E. Deal, A.S. Grove, General Relationship for the Thermal Oxidation of Silicon, J. Appl. Phys. 36 (1965) 3770–3778.

[8] V.S. Rao, T.J.R. Hughes, On modelling thermal oxidation of Silicon I: Theory, Int. J.

Numer. Methods Eng. 47 (2000) 341–358.

[9] H. Kageshima, M. Uematsu, K. Shiraishi, Theory of thermal Si oxide growth rate taking into account interfacial Si emission effects, Microelectron. Eng. 59 (2001) 301–309.

[10] D. Goto, Y. Hijikata, S. Yagi, H. Yaguchi, Differences in SiC thermal oxidation process between crystalline surface orientations observed by in-situ spectroscopic ellipsometry, J. Appl. Phys. 117 (2015) 0–6.

[11] B. Stegemann, K.M. Gad, P. Balamou, D. Sixtensson, D. Vössing, M. Kasemann, H. Angermann, Ultra-thin silicon oxide layers on crystalline silicon wafers: Comparison of advanced oxidation techniques with respect to chemically abrupt SiO2/Si interfaces with low defect densities, Appl. Surf. Sci. 395 (2017) 78–85.

[12] X. She, A.Q. Huang, Ó. Lucía, B. Ozpineci, Review of Silicon Carbide Power Devices and Their Applications, IEEE Trans. Ind. Electron. 64 (2017) 8193–8205.

[13] S. Palagi, P. Fischer, Bioinspired microrobots, Nat. Rev. Mater. 3 (2018) 113–124.

[14] P. Yuan, J.M. McCracken, D.E. Gross, P. V Braun, J.S. Moore, R.G. Nuzzo, A programmable soft chemo-mechanical actuator exploiting a catalyzed photochemical water-oxidation reaction, Soft Matter. 13 (2017) 7312–7317.

[15] D. Ambrosi, S. Pezzuto, D. Riccobelli, T. Stylianopoulos, P. Ciarletta, Solid Tumors Are Poroelastic Solids with a Chemo-mechanical Feedback on Growth, J. Elast. 129 (2017) 107–124.

[16] D. Ambrosi, G.A. Ateshian, E.M. Arruda, S.C. Cowin, J. Dumais, A. Goriely, G.A. Holzapfel, J.D. Humphrey, R. Kemkemer, E. Kuhl, J.E. Olberding, L.A. Taber, K. Garikipati, Perspectives on biological growth and remodeling, J. Mech. Phys. Solids. 59 (2011) 863–883.

[17] D. Faghihi, X. Feng, E.A.B.F. Lima, J.T. Oden, T.E. Yankeelov, A coupled mass transport and deformation theory of multi-constituent tumor growth, J. Mech. Phys. Solids. 139 (2020) 103936.

[18] T.J. Truster, A. Masud, A unified mixture formulation for density and volumetric growth of multi-constituent solids in tissue engineering, Comput. Methods Appl. Mech. Eng. 314 (2017) 222–268.

[19] J.D. Humphrey, K.R. Rajagopal, A constrained mixture model for growth and remodeling of soft tissues, Math. Model. Methods Appl. Sci. 12 (2002) 407–430.

[20] J.D. Humphrey, K.R. Rajagopal, A constrained mixture model for arterial adaptations to a sustained step change in blood flow, Biomech. Model. Mechanobiol. 2 (2003) 109–126.

[21] J.W. Wang, Y. He, F. Fan, X.H. Liu, S. Xia, Y. Liu, C. Thomas, C.T. Harris, H. Li, J.Y. Huang, S.X. Mao, T. Zhu, Two-Phase Electrochemical Lithiation in Amorphous Silicon.pdf, Nano Lett. 13 (2013) 1–11.

[22] H. Yang, F. Fan, W. Liang, X. Guo, T. Zhu, S. Zhang, A chemo-mechanical model of lithiation in silicon, J. Mech. Phys. Solids. 70 (2014) 349–361.

[23] T. Jiang, S. Rudraraju, A. Roy, A. Van Der Ven, K. Garikipati, M.L. Falk, Multiphysics Simulations of Lithiation-Induced Stress in Li1+xTi2O4Electrode Particles, J. Phys. Chem. C. 120 (2016) 27871–27881.

[24] X. Zhang, Z. Zhong, A coupled theory for chemically active and deformable solids with mass diffusion and heat conduction, J. Mech. Phys. Solids. 107 (2017) 49–75.

[25] Y. Zhao, P. Stein, Y. Bai, M. Al-Siraj, Y. Yang, B.-X. Xu, A review on modeling of electro-chemo-mechanics in lithium-ion batteries, J. Power Sources. 413 (2019) 259–283.

[26] S.R. Bishop, Chemical expansion of solid oxide fuel cell materials: A brief overview, Acta Mech. Sin. 29 (2013) 312–317.

[27] B.A. Haberman, J.B. Young, Three-dimensional simulation of chemically reacting gas flows in the porous support structure of an integrated-planar solid oxide fuel cell, Int. J. Heat Mass Transf. 47 (2004) 3617–3629.

[28] N.A. Peppas, J.Z. Hilt, A. Khademhosseini, R. Langer, Hydrogels in biology and medicine: From molecular principles to bionanotechnology, Adv. Mater. 18 (2006) 1345–1360.

[29] T. Shoaib, A. Carmichael, R.E. Corman, Y. Shen, T.H. Nguyen, R.H. Ewoldt, R.M. Espinosa-Marzal, Self-adaptive hydrogels to mineralization, Soft Matter. 13 (2017) 5469–5480.

[30] S.A. Chester, L. Anand, A thermo-mechanically coupled theory for fluid permeation in elastomeric materials: Application to thermally responsive gels, J. Mech. Phys. Solids. 59 (2011) 1978–2006.

[31] A. Fasano, K. Kannan, A. Mancini, K.R. Rajagopal, Modelling Ziegler–Natta Polymerization in High Pressure Reactors, in: G. Capriz, P.M. Mariano (Eds.), Mater. Substruct. Complex Bodies, Elsevier Science Ltd, Oxford, 2007: pp. 206–237.

[32] C. Wei, P. Ortoleva, Reaction front fingering in carbonate-cemented sandstones, Earth-Science Rev. 29 (1990) 183–198.

[33] A. De Wit, Chemo-hydrodynamic patterns in porous media, Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 374 (2016) 20150419.

[34] G.P. Tandon, K. V. Pochiraju, G.A. Schoeppner, Modeling of oxidative development in PMR-15 resin, Polym. Degrad. Stab. 91 (2006) 1861–1869.

[35] J. Gonzalez, K. Sun, M. Huang, J. Lambros, S. Dillon, I. Chasiotis, Three dimensional studies of particle failure in silicon based composite electrodes for lithium ion batteries, J. Power Sources. 269 (2014) 334–343.

[36] S. Zhang, Chemomechanical modeling of lithiation-induced failure in high-volume-change electrode materials for lithium ion batteries, Npj Comput. Mater. 3 (2017) 1–10.

[37] C. Truesdell, W. Noll, The Non-Linear Field Theories of Mechanics, Third, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[38] M.E. Gurtin, E. Fried, L. Anand, The mechanics and thermodynamics of continua., Cambridge University Press, Cambridge, 2010.

[39] K.R. Rajagopal, L. Tao, Mechanics of Mixtures, World Scientific, 1995.

[40] S.C. Brenner, L.R. Scott, The Mathematical Theory of Finite Element Methods, 3rd ed., Springer New York, New York, NY, 2008.

[41] T.J.R. Hughes, The finite element method: linear static and dynamic finite element analysis, Dover Publications, Mineola, NY, 2000.

[42] T. Belytschko, W.K. Liu, B. Moran, K.I. Elkhodary, Nonlinear Finite Elements for Continua and Structures, Second, John Wiley & Sons, Chichester, West Sussex, UK, 2014.

[43] A.N. Brooks, T.J.R. Hughes, Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations, Comput. Methods Appl. Mech. Eng. 32 (1982) 199–259.

[44] L.P. Franca, E.G. Dutra Do Carmo, The Galerkin gradient least-squares method, Comput. Methods Appl. Mech. Eng. 74 (1989) 41–54.

[45] P.G. Ciarlet, The Finite Element Method for Elliptic Problems, Society for Industrial and Applied Mathematics, Paris, France, 2002.

[46] C. Truesdell, Rational Thermodynamics, Second, Springer New York, New York, NY,

1984.

[47] S.A. Chester, L. Anand, A coupled theory of fluid permeation and large deformations for elastomeric materials, J. Mech. Phys. Solids. 58 (2010) 1879–1906.

[48] K. Loeffel, L. Anand, A chemo-thermo-mechanically coupled theory for elastic–viscoplastic deformation, diffusion, and volumetric swelling due to a chemical reaction, Int. J. Plast. 27 (2011) 1409–1431.

[49] S.A. Chester, C. V. Di Leo, L. Anand, A finite element implementation of a coupled diffusion-deformation theory for elastomeric gels, Int. J. Solids Struct. 52 (2015) 1–18.

[50] K. Al-Athel, K. Loeffel, H. Liu, L. Anand, Modeling decohesion of a top-coat from a thermally-growing oxide in a thermal barrier coating, Surf. Coatings Technol. 222 (2013) 68–78.

[51] T. Sain, K. Loeffel, S. Chester, A thermo–chemo–mechanically coupled constitutive model for curing of glassy polymers, J. Mech. Phys. Solids. 116 (2018) 267–289.

[52] S. Konica, T. Sain, A thermodynamically consistent chemo-mechanically coupled large deformation model for polymer oxidation, J. Mech. Phys. Solids. 137 (2020) 103858.

[53] X. Zhang, Z. Zhong, A coupled theory for chemically active and deformable solids with mass diffusion and heat conduction, J. Mech. Phys. Solids. 107 (2017) 49–75.

[54] K. Kannan, K.R. Rajagopal, A thermodynamical framework for chemically reacting systems, Zeitschrift Für Angew. Math. Und Phys. 62 (2011) 331–363.

[55] J.W. Cahn, J.E. Hilliard, Free Energy of a Nonuniform System. I. Interfacial Free Energy, J. Chem. Phys. 28 (1958) 258–267.

[56] M.E. Gurtin, Generalized Ginzburg-Landau and Cahn-Hilliard equations based on a microforce balance, Phys. D Nonlinear Phenom. 92 (1996) 178–192.

[57] M.E. Gurtin, Configurational Forces as Basic Concepts of Continuum Physics, Springer New York, New York, NY, 2000.

[58] L. Anand, A Cahn–Hilliard-type theory for species diffusion coupled with large elastic–plastic deformations, J. Mech. Phys. Solids. 60 (2012) 1983–2002.

[59] C. V Di Leo, E. Rejovitzky, L. Anand, A Cahn–Hilliard-type phase-field theory for species diffusion coupled with large elastic deformations: Application to phase-separating Li-ion electrode materials, J. Mech. Phys. Solids. 70 (2014) 1–29.

[60] M. Heida, J. Málek, K.R. Rajagopal, On the development and generalizations of Cahn--

Hilliard equations within a thermodynamic framework, Zeitschrift Für Angew. Math. Und Phys. 63 (2012) 145–169.

[61] A. Fick, Ueber Diffusion, Ann. Phys. 170 (1855) 59–86.

[62] H.P.G. Darcy, Les Fontaines Publiques de la Ville de Dijon, PARIS, 1856.

[63] R.M. Bowen, Theory of Mixtures, in: A.C. Eringen (Ed.), Contin. Phys., Volume III, Elsevier, 1976: pp. 1–127.

[64] R.J. Atkin, R.E. Craine, Continuum theories of mixtures: Basic theory and historical development, Q. J. Mech. Appl. Math. 29 (1976) 209–244.

[65] R.J. Atkin, R.E. Craine, Continuum Theories of Mixtures: Applications, IMA J. Appl. Math. 17 (1976) 153–207.

[66] K. V Mohankumar, V. Průša, K. Kannan, A.S. Wineman, Remarks on continuum theory of mixtures: editorial to special issue on mixture theory, Int. J. Adv. Eng. Sci. Appl. Math. 9 (2017) 120–134.

[67] J. Siddique, A. Ahmed, A. Aziz, C. Khalique, A Review of Mixture Theory for Deformable Porous Media and Applications, Appl. Sci. 7 (2017) 917.

[68] O. Souček, V. Orava, J. Málek, D. Bothe, A continuum model of heterogeneous catalysis: Thermodynamic framework for multicomponent bulk and surface phenomena coupled by sorption, Int. J. Eng. Sci. 138 (2019) 82–117.

[69] H. Gajendran, R.B. Hall, A. Masud, K.R. Rajagopal, Chemo-mechanical coupling in curing and material-interphase evolution in multi-constituent materials, Acta Mech. 229 (2018) 3393–3414.

[70] R. Hall, H. Gajendran, A. Masud, Diffusion of chemically reacting fluids through nonlinear elastic solids: mixture model and stabilized methods, Math. Mech. Solids. 20 (2015) 204–227.

[71] R. Hall, K. Rajagopal, Diffusion of a fluid through an anisotropically chemically reacting thermoelastic body within the context of mixture theory, Math. Mech. Solids. 17 (2012) 131–164.

[72] R.B. Hall, A mixture-compatible theory of chemothermal deposition and expansion in n - constituent finitely deforming composite materials with initially circularly cylindrical microstructures, Math. Mech. Solids. 20 (2015) 228–248.

[73] V. Klika, A Guide through Available Mixture Theories for Applications, Crit. Rev. Solid

State Mater. Sci. 39 (2014) 154–174.

[74] H. Gajendran, R.B. Hall, A. Masud, Edge stabilization and consistent tying of constituents at Neumann boundaries in multi-constituent mixture models, Int. J. Numer. Methods Eng. 110 (2017) 1142–1172.

[75] R.C. McOwen, Partial Differential Equations: Methods And Applications, Print, Prentice-Hall, Upper Saddle, NJ, 2003.

[76] P.. Ciarlet, P.-A. Raviart, Maximum principle and uniform convergence for the finite element method, Comput. Methods Appl. Mech. Eng. 2 (1973) 17–31.

[77] M. Anguiano, H. Gajendran, R.B. Hall, K.R. Rajagopal, A. Masud, Chemo-mechanical coupling and material evolution in finitely deforming solids with advancing fronts of reactive fluids, Acta Mech. 231 (2020) 1933–1961.

[78] I. Harari, Stability of semidiscrete formulations for parabolic problems at small time steps, Comput. Methods Appl. Mech. Eng. 193 (2004) 1491–1516.

[79] L.P. Franca, G. Hauke, A. Masud, Revisiting stabilized finite element methods for the advective–diffusive equation, Comput. Methods Appl. Mech. Eng. 195 (2006) 1560–1572.

[80] T.J.R. Hughes, M. Mallet, M. Akira, A new finite element formulation for computational fluid dynamics: II. Beyond SUPG, Comput. Methods Appl. Mech. Eng. 54 (1986) 341–355.

[81] T.J.R. Hughes, L.P. Franca, G.M. Hulbert, A new finite element formulation for computational fluid dynamics: VIII. The galerkin/least-squares method for advective-diffusive equations, Comput. Methods Appl. Mech. Eng. 73 (1989) 173–189.

[82] T.E. Tezduyar, Y.J. Park, Discontinuity-capturing finite element formulations for nonlinear convection-diffusion-reaction equations, Comput. Methods Appl. Mech. Eng. 59 (1986) 307–325.

[83] I. Harari, T.J.R. Hughes, Stabilized finite element methods for steady advection-diffusion with production, Comput. Methods Appl. Mech. Eng. 115 (1994) 165–191.

[84] L.P. Franca, C. Farhat, Bubble functions prompt unusual stabilized finite element methods, Comput. Methods Appl. Mech. Eng. 123 (1995) 299–308.

[85] T.J.R. Hughes, Multiscale phenomena: Green's functions, the Dirichlet-to-Neumann formulation, subgrid scale models, bubbles and the origins of stabilized methods, Comput.

Methods Appl. Mech. Eng. 127 (1995) 387–401.

[86] T.J.R. Hughes, G.R. Feijóo, L. Mazzei, J.-B. Quincy, The variational multiscale method—a paradigm for computational mechanics, Comput. Methods Appl. Mech. Eng. 166 (1998) 3–24.

[87] F. Brezzi, A. Russo, Choosing bubbles for advection-diffusion problems, Math. Model. Methods Appl. Sci. 04 (1994) 571–587.

[88] L.P. Franca, A. Russo, Deriving upwinding, mass lumping and selective reduced integration by residual-free bubbles, Appl. Math. Lett. 9 (1996) 83–88.

[89] F. Brezzi, D. Marini, A. Russo, Applications of the pseudo residual-free bubbles to the stabilization of convection-diffusion problems, Comput. Methods Appl. Mech. Eng. 166 (1998) 51–63.

[90] F. Brezzi, L.P. Franca, T.J.R. Hughes, A. Russo, $b = \int g$, Comput. Methods Appl. Mech. Eng. 145 (1997) 329–339.

[91] R. Codina, Comparison of some finite element methods for solving the diffusion-convection-reaction equation, Comput. Methods Appl. Mech. Eng. 156 (1998) 185–210.

[92] G. Hauke, A. García-Olivares, Variational subgrid scale formulations for the advection–diffusion-reaction equation, Comput. Methods Appl. Mech. Eng. 190 (2001) 6847–6865.

[93] E.G. Dutra do Carmo, G.B. Alvarez, F.A. Rochinha, A.F.D. Loula, Galerkin projected residual method applied to diffusion–reaction problems, Comput. Methods Appl. Mech. Eng. 197 (2008) 4559–4570.

[94] H. Fernando, C. Harder, D. Paredes, F. Valentin, Numerical multiscale methods for a reaction-dominated model, Comput. Methods Appl. Mech. Eng. 201–204 (2012) 228–244.

[95] E. Oñate, P. Nadukandi, J. Miquel, Accurate FIC-FEM formulation for the multidimensional steady-state advection–diffusion–absorption equation, Comput. Methods Appl. Mech. Eng. 327 (2017) 352–368.

[96] K.B. Nakshatrala, M.K. Mudunuru, A.J. Valocchi, A numerical framework for diffusion-controlled bimolecular-reactive systems to enforce maximum principles and the non-negative constraint, J. Comput. Phys. 253 (2013) 278–307.

[97] J. Chang, K.B. Nakshatrala, Variational inequality approach to enforcing the non-negative constraint for advection–diffusion equations, Comput. Methods Appl. Mech. Eng. (2017).

[98] T.J.R. Hughes, G. Scovazzi, L.P. Franca, Multiscale and Stabilized Methods, in: Encycl.

Comput. Mech. Second Ed., John Wiley & Sons, Ltd, Chichester, UK, 2017: pp. 1–64.

[99] A.A. Oberai, P.M. Pinsky, A multiscale finite element method for the Helmholtz equation, Comput. Methods Appl. Mech. Eng. 154 (1998) 281–297.

[100] A.A. Oberai, P.M. Pinsky, A residual-based finite element method for the Helmholtz equation, Int. J. Numer. Methods Eng. 49 (2000) 399–419.

[101] A.L.G.A. Coutinho, L.P. Franca, F. Valentin, Numerical multiscale methods, Int. J. Numer. Methods Fluids. 70 (2012) 403–419.

[102] J. Principe, R. Codina, On the stabilization parameter in the subgrid scale approximation of scalar convection–diffusion–reaction equations on distorted meshes, Comput. Methods Appl. Mech. Eng. 199 (2010) 1386–1402.

[103] T.J.R. Hughes, G. Sangalli, Variational Multiscale Analysis: the Fine-scale Green's Function, Projection, Optimization, Localization, and Stabilized Methods, SIAM J. Numer. Anal. 45 (2007) 539–557.

[104] A. Masud, R.A. Khurram, A multiscale/stabilized finite element method for the advection–diffusion equation, Comput. Methods Appl. Mech. Eng. 193 (2004) 1997–2018.

[105] A. Masud, L.P. Franca, A hierarchical multiscale framework for problems with multiscale source terms, Comput. Methods Appl. Mech. Eng. 197 (2008) 2692–2700.

[106] A. Masud, R.A. Khurram, A multiscale finite element method for the incompressible Navier–Stokes equations, Comput. Methods Appl. Mech. Eng. 195 (2006) 1750–1777.

[107] A. Masud, R. Calderer, A variational multiscale method for incompressible turbulent flows: Bubble functions and fine scale fields, Comput. Methods Appl. Mech. Eng. 200 (2011) 2577–2593.

[108] R. Calderer, A. Masud, Residual-based variational multiscale turbulence models for unstructured tetrahedral meshes, Comput. Methods Appl. Mech. Eng. 254 (2013) 238–253.

[109] A. Masud, L. Zhu, Variationally derived closure models for large eddy simulation of incompressible turbulent flows, Int. J. Numer. Methods Fluids. 93 (2021) 2089–2120.

[110] J. Kwack, A. Masud, A stabilized mixed finite element method for shear-rate dependent non-Newtonian fluids: 3D benchmark problems and application to blood flow in bifurcating arteries, Comput. Mech. 53 (2014) 751–776.

[111] J. Kwack, A. Masud, K.R. Rajagopal, Stabilized mixed three-field formulation for a generalized incompressible Oldroyd-B model, Int. J. Numer. Methods Fluids. 83 (2017) 704–734.

[112] A. Masud, T.J. Truster, A framework for residual-based stabilization of incompressible finite elasticity: Stabilized formulations and F bar methods for linear triangles and tetrahedra, Comput. Methods Appl. Mech. Eng. 267 (2013) 359–399.

[113] K. Xia, A. Masud, A stabilized finite element formulation for finite deformation elastoplasticity in geomechanics, Comput. Geotech. (2009).

[114] A. Masud, T.J. Truster, L.A. Bergman, A unified formulation for interface coupling and frictional contact modeling with embedded error estimation, Int. J. Numer. Methods Eng. 92 (2012) 141–177.

[115] T.J. Truster, A. Masud, A Discontinuous/continuous Galerkin method for modeling of interphase damage in fibrous composite systems, Comput. Mech. 52 (2013) 499–514.

[116] T.J. Truster, A. Masud, Primal interface formulation for coupling multiple PDEs: A consistent derivation via the Variational Multiscale method, Comput. Methods Appl. Mech. Eng. 268 (2014) 194–224.

[117] T.J. Truster, P. Chen, A. Masud, Finite strain primal interface formulation with consistently evolving stabilization, Int. J. Numer. Methods Eng. 102 (2015) 278–315.

[118] A. Masud, P. Chen, Thermoelasticity at finite strains with weak and strong discontinuities, Comput. Methods Appl. Mech. Eng. 347 (2019) 1050–1084.

[119] R.A. Khurram, A. Masud, A Multiscale/stabilized Formulation of the Incompressible Navier–Stokes Equations for Moving Boundary Flows and Fluid–structure Interaction, Comput. Mech. 38 (2006) 403–416.

[120] A. Masud, T.J. Truster, L.A. Bergman, A variational multiscale a posteriori error estimation method for mixed form of nearly incompressible elasticity, Comput. Methods Appl. Mech. Eng. 200 (2011) 3453–3481.

CHAPTER 2:
CHEMO-MECHANICAL COUPLING AND MATERIAL EVOLUTION IN FINITELY DEFORMING SOLIDS WITH ADVANCING FRONTS OF REACTIVE FLUIDS[1]

2.1 Introduction

Transport of chemically reacting fluids through non-linear elastic solids is fundamental to many fields of science and engineering. A sample class of problems includes biological growth of tissue and tumors [1–3], development of bio-inspired devices [4], modeling of solid-oxide fuel cells [5,6], oxidation of ceramic materials [7–11], and battery lithiation [12], to name a few. These problems involve diffusion-reaction processes that result in chemical expansion and/or contraction of the constituents causing growth and/or recession of component materials. The complexity of the governing system of equations increases with an increase in the mechanical deformation, the rate of chemical reactions, and the rate of mass transport.

Mathematical and computational analysis of problems where fluid and solid constituents coexist concurrently requires models that take into account disparate physical phenomena and methods that represent the interacting constituents either discretely or via homogenization techniques. Homogenization methods and mixture theories provide a framework for the modeling of coupled material systems that are comprised of more than one constituent material. These theories lead to computationally economic solution methods for this class of problems that would otherwise require discrete modeling approaches to represent each of the individual constituents along with the modeling of discrete material interfaces. In our earlier work we employed mixture theory formalism (Hall et al [13]) and developed a one-dimensional stabilized finite element method that accounts for chemical reaction of two constituents: a fluid and a solid. A Variational Multiscale (VMS) stabilized method was developed where stabilization was performed on the fluid density field in the balance of mass equation. The fundamentals of the VMS framework were presented in [14–16]. This framework proved successful in numerical simulation of diffusion, slurry infiltration (SI), and the oxidation of PMR-15. However numerical instabilities were observed for the case of faster reaction rates. In addition, deformation that takes place due to chemo-mechanical expansion/contraction with accompanied mass growth and recession of the component materials was ignored.

---

[1] This Chapter has been adapted from "M. Anguiano, H. Gajendran, R.B. Hall, K.R. Rajagopal, A. Masud, Chemo-mechanical coupling and material evolution in finitely deforming solids with advancing fronts of reactive fluids, Acta Mech. 231 (2020) 1933–1961." The copyright owner has provided written permission to reprint the work.

In this chapter we adopt mixture theory which models complex multi-constituent materials by representing each constituent in a continuum sense in an overlapping domain, while retaining information about the interaction between these constituents through *interaction terms* that appear in the physical balance laws [17–21]. These terms that capture the inter-constituent interactions are "reaction-like" (i.e., contain zeroth-order terms, which depend on the unknown fields but not on their derivatives), so the governing equations are, in general, nonlinear advection-reaction or diffusion-advection-reaction equations. The present developments focus on the chemo-mechanically coupled response where chemical reactions and mass exchange between the constituents is associated with substantial geometric deformation. Since chemo-mechanical coupling term is non-zero, depending on the material classes, the intensity of these reactions can lead to large deformation (volume increase) in the transformed or reacted solid. In addition, the mechanical material properties of the solid constituent evolve with the progression of reaction because the reacted material is compositionally different from the unreacted material. From a mathematical viewpoint the higher reaction rates can give rise to dominant zeroth-order terms, in the sense that the coefficients of higher-order terms are small in comparison, which in turn can lead to singularly perturbed PDEs that are well known to exhibit boundary or moving internal layers. Consequently, depending on the width of the active reaction zone, which is a function of the chemical composition of the solid and fluid constituents, steep gradients in advecting fronts can appear. Accurate and stable modeling of these layers requires numerical methods that can capture steep gradients without overshoots or undershoots [22–26]. For a historic perspective of stabilized methods, the interested reader is referred to [27].

A significant contribution in the new numerical method presented here is the appearance of the reaction term in the stabilized form that is shown to play an important role in eliminating any nonphysical oscillations around the steep gradients in the zones with higher reaction rates. To make the discussion precise, we present Helmholtz free energies for each of the two constituents. These free energy expressions are otherwise arbitrary and only subject to the restrictions imposed by thermodynamic consistency of the mixture model. Numerical section shows that mixture theory leads to reduced-order models of the complex physical processes involved in multi-constituent material evolution. The resulting stabilized variational form of the mixture model results in a computationally economic method for capturing chemo-mechanically coupled behavior in finitely deforming solids that are permeated by reacting fluids. Test cases highlight that the elastic solid constituent undergoes material property evolution that is accompanied with swelling and growth. In particular we consider the oxidation of silicon (Si) as well as silicon carbide (SiC) in the test cases presented in the numerical section and compare our results with the data that is available in the literature.

An outline of this chapter is as follows: In section 2.2 we introduce the balance laws in the context of the mixture theory. Section 2.3 discusses the constitutive equations. The strong form of the problem is presented in section 2.4 and the weak form is derived in section 2.5. Section 2.6 presents the stabilized finite element method for chemo-mechanical problems. Numerical test cases are presented in Section 2.7 and Section 2.8 contains concluding remarks.

2.2 Balance Laws

We consider a mixture of two constituents, one fluid '*f*' and one solid '*s*' that co-occupy the domain $\Omega \in \mathbb{R}^n$ with boundary $\partial\Omega$, where n is the number of spatial dimensions. The balance of mass, the balance of linear momentum, and the balance of energy for constituent $\alpha \in \{f, s\}$ over the domain of interest Ω and for the time interval of interest $\mathrm{I} =]0,\mathrm{T}[$, are given in equations (2.1), (2.2), and (2.3), respectively.

$$\frac{D\rho^\alpha}{Dt} + \rho^\alpha \operatorname{div}(\mathbf{v}^\alpha) - m^\alpha = 0 \qquad \text{in } \Omega \times]0,\mathrm{T}[\qquad (2.1)$$

$$\rho^\alpha \frac{D\mathbf{v}^\alpha}{Dt} - \operatorname{div}(\mathbf{T}^\alpha) - \rho^\alpha \mathbf{b}^\alpha - \mathbf{I}^\alpha = \mathbf{0} \qquad \text{in } \Omega \times]0,\mathrm{T}[\qquad (2.2)$$

$$\rho^\alpha \frac{D\varepsilon^\alpha}{Dt} - \operatorname{tr}(\mathbf{T}^\alpha \mathbf{L}^\alpha) + \operatorname{div}(\mathbf{q}^\alpha) - \rho^\alpha r^\alpha - \varepsilon_S^\alpha = 0 \qquad \text{in } \Omega \times]0,\mathrm{T}[\qquad (2.3)$$

where $\frac{D}{Dt}$ denotes the material time derivative. For constituent α, ρ^α is the mass concentration, $\mathbf{v}^\alpha$ is the velocity, m^α is the rate of mass accretion, $\mathbf{T}^\alpha$ is the Cauchy stress tensor, $\mathbf{b}^\alpha$ is the body force per unit mass, and $\mathbf{I}^\alpha$ is the interactive force (i.e. the force exerted on constituent α due to interaction with the other constituents), ε^α is the internal energy, $\mathbf{L}^\alpha$ is the velocity gradient, $\mathbf{q}^\alpha$ is the heat flux vector, r^α is the radiant heat supply, and ε_S^α is the energy supply to constituent α due to interactions with the other constituents. Requiring the Cauchy stress tensor to be symmetric satisfies conservation of angular momentum. The constitutive equations employed here were developed to satisfy the second law of thermodynamics [28].

2.2.1 Arbitrary Lagrangian-Eulerian Formulation

We consider an unconstrained mixture model where the constituents exhibit motion relative to each other. We consider problems in which the solid constituent is comprised of a connected domain and the fluid diffuses through the pores at the meso-scale of the solid, or through otherwise empty space at the micro-scale (polymer networks, interstitials, micro-pores, etc.). Thus, an Arbitrary Lagrangian-Eulerian (ALE) formulation is well suited to model the fluid. The conservation laws are rewritten by expanding the material time derivative defined for ALE-based framework.

$$\frac{D(\cdot)}{Dt} = \frac{\partial(\cdot)}{\partial t} + (\mathbf{v} - \mathbf{v}^m)\cdot\nabla(\cdot) \tag{2.4}$$

where $(\cdot)$ is the indicated argument, $\mathbf{v}^m$ is the velocity of the referential or ALE domain and ∇ is the gradient with respect to ALE coordinates. In particular, the solid domain serves as the referential (or ALE) domain for the fluid constituent, thus resulting in the condition that $\mathbf{v}^m = \mathbf{v}^s$. From the perspective of the solid constituent, the material domain and the referential domain are one and the same, and we recover the standard Lagrangian formulation. However, from the perspective of the fluid constituent, the changing spatial domain is defined by the current configuration of the solid, and therefore ALE formulation is appropriate.

2.3 Mixture Constitutive Equations

We adopt the constitutive framework proposed in [28], which was derived by imposing the constraint of maximization of the rate of entropy production. The constitutive equations for the Cauchy stress in the solid and the fluid constituents as well as the interactive force for the volumetrically unconstrained mixture are given as

$$\mathbf{T}^s = \rho\mathbf{F}^s\left(\frac{\partial\psi}{\partial\mathbf{F}^s}\right)^T - \rho^s\left(g^s + \frac{\rho^f}{\rho}(\psi^f - \psi^s)\right)\mathbf{1} \tag{2.5}$$

$$\mathbf{T}^f = \rho^f\left(g^f + \frac{\rho^s}{\rho}(\psi^s - \psi^f)\right)\mathbf{1} + \mu^L\mathbf{A}^L\cdot\mathbf{D}^f \tag{2.6}$$

$$\begin{aligned}\mathbf{I}^f &= g^f \frac{\rho^s}{\rho}\nabla\rho^f - g^s \frac{\rho^f}{\rho}\nabla\rho^s - \nabla\left[\frac{\rho^s\rho^f}{\rho}(\psi^f - \psi^s)\right] \\ &\quad - (\nabla\theta)\frac{\rho^s\rho^f}{\rho}(\eta^f - \eta^s) - m^f(\mathbf{v}^f - \mathbf{v}^s) - \mu^L \mathbf{A}^v(\mathbf{v}^f - \mathbf{v}^s)\end{aligned} \tag{2.7}$$

where $\rho = \rho^f + \rho^s$ is the mass concentration of the mixture, $\mathbf{F}^s$ is the solid deformation gradient, $\mathbf{D}^f$ is two times the symmetric part of the fluid velocity gradient ($\mathbf{D}^f = \mathbf{L}^f + (\mathbf{L}^f)^{\mathrm{T}}$), θ is the absolute temperature of the mixture, $\mathbf{A}^L$ is the tensor of viscosity coefficients, and $\mathbf{A}^v$ is the tensor of drag coefficients. The entropy of constituent α is given by η^α, the Helmholtz free energy of constituent α is given by ψ^α, the Helmholtz free energy for the mixture is defined as the mass average of the energy of the constituents $\psi = (\rho^f\psi^f + \rho^s\psi^s)/\rho$, the term $g^\alpha \equiv \rho\,\partial\psi/\partial\rho^\alpha$ is the chemical potential for constituent α, and μ^L is the Lagrange multiplier enforcing maximization of the rate of entropy production given in [13,28]. Furthermore, the derivations in [28] arrive at the following restriction on the form in which the Helmholtz free energy depends on the state variables and the extent of reaction:

$$\rho\psi = h\left[\mathbf{F}^s, \rho^s, \rho^f, \theta\right] - \int \frac{\mu^L}{2}\left\{\Gamma_{kl}\Gamma_{mn}B^s_{op}K^0_{ijklmnop} + 2\dot{\Gamma}_{kl}A^\Gamma_{ijkl}\right\} d\Gamma_{ji} \tag{2.8}$$

where Γ is the extent-of-reaction tensor, $\dot{\Gamma}$ is its material derivative, $\mathbf{B}^s = \mathbf{F}^s\mathbf{F}^{s\mathrm{T}}$ is the left Cauchy-Green tensor, and $\mathbf{K}^0$ and $\mathbf{A}^\Gamma$ are higher dimensional tensors of material coefficients (not necessarily constants). The integral term in equation (2.8) corresponds to the part of the free energy that accounts for the chemo-mechanical coupling. The other contributions to the free energy of the mixture, such as the thermo-mechanical energy of the solid and the fluid, are contained in *h*.

Remark: The interactive force on the solid is obtained via application of Newton's third law, $\sum_\alpha\left(\mathbf{I}^\alpha + m^\alpha\mathbf{v}^\alpha\right) = 0$, which yields $\mathbf{I}^s = \mathbf{I}^f - m^f\mathbf{v}^f - m^s\mathbf{v}^s$.

Remark: Maximization of the rate of entropy production is not necessarily a physical principle, but it serves as a guiding methodology in determining the constitutive equations (see [28] and references therein).

Remark: The constitutive relations in this section are for the general Helmholtz free energy expressions. We will specify free energies for the particular material system in section 4, and the corresponding constitutive equations will reduce to a more concise form.

2.3.1 Chemical Reactions

We employ a simplified phenomenological reaction model, in which the rate of reaction is proportional to the concentration of reactants, namely the fluid and the amount of the unreacted solid substrate, given by

$$m^f = R_0 \left(\frac{\phi - \phi_0}{1 - \phi_0} \right) \rho^f \tag{2.9}$$

where R_0 is a material parameter. In this definition, we use the variable $\phi_0 \leq \phi \leq 1$, which represents the state of reaction and is a measure of the amount of unreacted solid material, which evolves according to equation (2.10). A similar treatment of the reaction term and the evolution of the internal variable can be found in [13].

$$\frac{D\phi}{Dt} = r_\alpha m^f \tag{2.10}$$

The assumption of conservation of mass during reaction dictates that the mass lost by the fluid is equal to the mass incorporated by the solid: $m^s + m^f = 0$. However, one can also consider chemical processes where byproducts of reactions may not be of engineering interest, and that their interaction with the rest of the system may be negligible. In that case we define

$$m^s = -r_m m^f \tag{2.11}$$

where r_m is the fraction of fluid mass (defined locally) that gets converted to solid. In addition, regarding the reaction terms, we assume the extent-of-reaction tensor to be isotropic, which gives

$$\dot{\Gamma} = \dot{\gamma}\mathbf{1} \quad \text{and} \quad \Gamma = \gamma\mathbf{1} \tag{2.12}$$

where γ is the variable for the extent of reaction. In order to satisfy the requirements on the reaction tensor stated in [28] we have

$$\dot{\gamma} = m^s \tag{2.13}$$

Combining equations(2.10), (2.11), and (2.13), the extent of reaction γ is related to the reaction state ϕ by the following equation

$$\gamma = \frac{r_m}{r_\alpha}(1-\phi) \tag{2.14}$$

Thus, only one variable (either γ or ϕ) needs to be tracked in the numerical implementation.

Remark: The chemo-mechanical effects include the expansion of the solid associated with the extent of the reaction $\boldsymbol{\Gamma}$ that is captured via the dependency of the Cauchy stress tensor in equation (2.5) on the extent of reaction, which results from the form of the Helmholtz free energy in equation (2.8), and the dependency of the mechanical properties on the reaction state.

2.4 Strong Form of the Problem

Consider the domain Ω with boundary $\partial\Omega$. In the present development thermal effects will not be considered, and we will focus on the coupled chemo-mechanical behavior of the system. Under these conditions the governing equations are the balance laws for mass and momentum of the individual constituents in the ALE framework, and they are as follows

$$\frac{\partial \rho^f}{\partial t} + \left(\mathbf{v}^f - \mathbf{v}^m\right)\cdot\nabla\rho^f + \rho^f \operatorname{div}\left(\mathbf{v}^f\right) - m^f = 0 \qquad \text{in } \Omega\times]0,\mathrm{T}[\tag{2.15}$$

$$\rho^f \frac{\partial \mathbf{v}^f}{\partial t} + \rho^f(\mathbf{v}^f - \mathbf{v}^m)\cdot\nabla\mathbf{v}^f - \operatorname{div}\left(\mathbf{T}^f\right) - \rho^f\mathbf{b}^f - \mathbf{I}^f = \mathbf{0} \qquad \text{in } \Omega\times]0,\mathrm{T}[\tag{2.16}$$

$$\frac{\partial \rho^s}{\partial t} + \rho^s \operatorname{div}\left(\mathbf{v}^s\right) - m^s = 0 \qquad \text{in } \Omega\times]0,\mathrm{T}[\tag{2.17}$$

$$-\operatorname{div}\left(\mathbf{T}^s\right) - \rho^s\mathbf{b}^s - \mathbf{I}^s = \mathbf{0} \qquad \text{in } \Omega\times]0,\mathrm{T}[\tag{2.18}$$

In order to make the discussion precise in regard to the development of the method, we select specific (but otherwise arbitrary) functional forms of Helmholtz free energies for the fluid and the solid.

$$\psi^f = \bar{R}\theta\left[\hat{c}_v - \ln\left(\frac{M_w\theta^{\hat{c}_v}}{\Phi A_v \rho^f}\right)\right] \tag{2.19}$$

$$\psi^s = (\rho^s J)^{-1}\hat{\psi} - \Lambda^s \tag{2.20}$$

where the fluid free energy corresponds to that of an ideal gas, and $\bar{R}$ is the ratio of the gas constant to the molecular weight of the fluid. The two terms that make up the free energy for the solid constituent are the strain energy function $\hat{\psi}$ (for an isotropic Neo-Hookean solid) and the chemo-mechanical coupling term Λ^s. These terms are, respectively:

$$\hat{\psi} = \frac{\mu^s}{2}\left(I_1 - 3\right) - \mu^s \ln\left(J\right) + \frac{k^s}{2}\left(J - 1\right)^2 \tag{2.21}$$

$$\Lambda^s = \int \frac{\mu^L}{2}\left(\Gamma_{kl}\Gamma_{mn}B^s_{op}K^0_{ijklmnop}\right)d\Gamma_{ji} \tag{2.22}$$

We note that (2.22) corresponds to the integral term in (2.8), where we have taken $A^{\Gamma}_{ijkl} = 0$ in order to simplify the formulation. Moreover, the tensor of chemo-mechanical coupling coefficients $K^0_{ijklmnop}$ is assumed to depend on the volumetric deformation and the density of the solid in the following form, which simplifies the derivation of the Cauchy stress of the solid.

$$K^0_{ijklmnop}\left(\rho^s, \mathbf{F}^s\right) = (\rho^s J)^{-1} C^0_{ijklmnop} \tag{2.23}$$

Furthermore, the isotropy of the extent of reaction, $\mathbf{\Gamma}$, allows the chemo-mechanical coupling tensor to be reduced to a second order tensor, which is assumed to be isotropic to keep the model simple.

$$C^0_{iikkmmop} = \bar{C}^0_{op} = \bar{c}^0\delta_{op} \tag{2.24}$$

Moreover, employing the assumptions discussed above and ignoring the dependence of the Lagrange multiplier on the extent of reaction, the chemo-mechanical term can be simplified as

$$\Lambda^s \approx (\rho^s J)^{-1}\frac{\mu^L}{6}\gamma^3\bar{c}^0\mathrm{tr}(\mathbf{B}^s) \tag{2.25}$$

We consider the fluid to be isotropic as well, and thus, the viscosity tensor in equation (2.6) takes the standard isotropic form

$$A^L_{ijkl} = \frac{\eta}{2}\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right) \tag{2.26}$$

where η is the dynamic viscosity of the fluid.

We relate the drag coefficient tensor to the solid diffusivity tensor $\mathbf{D}$ by reducing the problem to Fick’s diffusion. By comparing coefficients (as shown in the appendix of [13]) we get the following expression

$$\mathbf{A}^v = 2\bar{R}\theta\rho^f\mathbf{D}^{-1} \tag{2.27}$$

Embedding these simplifying assumptions in the constitutive model (2.5)-(2.7) for the unconstrained mixture, the expressions for the solid and fluid interactive forces and stresses are given as follows

$$\mathbf{T}^s = \mu^s J^{-1}\mathbf{F}\mathbf{F}^T + \left[-\mu^s J^{-1} + k^s (J-1)\right]\mathbf{1} - \frac{\mu^L}{3}\gamma^3 \bar{c}^0 J^{-1}\mathbf{F}\mathbf{F}^T \tag{2.28}$$

$$\mathbf{T}^f = -\bar{R}\theta\rho^f \mathbf{1} + \mu^L \eta \mathbf{D}^f \tag{2.29}$$

$$\mathbf{I}^f = -m^f(\mathbf{v}^f - \mathbf{v}^s) - 2\mu^L \bar{R}\theta\rho^f \mathbf{D}^{-1}(\mathbf{v}^f - \mathbf{v}^s) \tag{2.30}$$

$$\mathbf{I}^s = -\mathbf{I}^f - m^f \mathbf{v}^f - m^s \mathbf{v}^s \tag{2.31}$$

Remark: The material coefficients in the constitutive model for the solid constituent can in general be a function of the extent of reaction γ.

Remark: In the present chapter the Cauchy stress for the solid constituent is identical to that of a standard Neo-Hookean material, plus an additional term that depends on the extent of the reaction.

Remark: The partial derivatives of the Helmholtz free energy function for the solid with respect to the deformation and the extent of reaction tensors were omitted from the interactive force term so that no strain gradient terms appear in the final form. This simplification helps in employing the lowest order Lagrange elements in the numerical implementation of the mixture model.

The mathematical problem is posed in terms of the governing equations (2.15)-(2.18), the constitutive relations for stresses and interactive forces in (2.28)-(2.31), and the reaction and evolution equations in (2.9)-(2.11) and (2.14). In addition, appropriate initial and boundary conditions are applied. Substitution of the constitutive laws in the balance equations reveals that (a) the balance of mass for the fluid is a transient advection-reaction equation for the fluid mass concentration, (b) the balance of linear momentum for the fluid is a nonlinear advection-diffusion-reaction equation for the fluid velocity, and (c) the balance of linear momentum for the solid is a nonlinear diffusion-reaction equation for the solid displacement. These PDEs become singularly perturbed in the reaction dominated case [22–26], thereby requiring a numerical method that has enhanced stability and accuracy properties to control any non-physical instability in the numerical solution of the resulting system of equations

2.5 Weak Form of the Problem

We first note that the gradients of the solid density field ρ^s do not appear in this formulation. As a result, the balance of mass equation (2.17) for the solid constituent is not cast in a weak form. Rather it is treated algorithmically as an evolution equation at the integration points in the subsequent FEM development. Consequently, the unknown fields at the global level are the fluid mass concentration (ρ^f), fluid velocity ($\mathbf{v}^f$), and solid displacement ($\mathbf{u}^s$). We define the trial solution spaces for the unknown fields, namely the fluid mass concentration field ρ^f, the fluid velocity field $\mathbf{v}^f$, and the solid displacement field $\mathbf{u}^s$ as follows

$$\mathcal{S}^\rho = \left\{ \rho^f(\bullet,t) \middle| \rho^f(\bullet,t) \in H^1(\Omega), \rho^f(\bullet,t) = \rho_b^f \text{ on } \partial\Omega_g^\rho \times]0,T[\right\} \tag{2.32}$$

$$\mathcal{S}^f = \left\{ \mathbf{v}^f(\bullet,t) \middle| \mathbf{v}^f(\bullet,t) \in H^1(\Omega), v_i^f(\bullet,t) = \bar{v}_i^f \text{ on } \partial\Omega_i^f \times]0,T[\right\} \tag{2.33}$$

$$\mathcal{S}^s = \left\{ \mathbf{u}^s(\bullet,t) \middle| \mathbf{u}^s(\bullet,t) \in H^1(\Omega), u_i^s(\bullet,t) = \bar{u}_i^s \text{ on } \partial\Omega_i^s \times]0,T[\right\} \tag{2.34}$$

The corresponding test function spaces are:

$$\mathcal{V}^\rho = \left\{ g^f \middle| g^f \in H^1(\Omega), g^f = 0 \text{ on } \partial\Omega_g^\rho \right\} \tag{2.35}$$

$$\mathcal{V}^f = \left\{ \mathbf{w}^f \middle| \mathbf{w}^f \in H^1(\Omega), w_i^f = 0 \text{ on } \partial\Omega_i^f \right\} \tag{2.36}$$

$$\mathcal{V}^s = \left\{ \mathbf{w}^s \middle| \mathbf{w}^s \in H^1(\Omega), w_i^s = 0 \text{ on } \partial\Omega_i^s \right\} \tag{2.37}$$

The weak form of the problem is: Given the body force terms for the fluid and the solid $\mathbf{b}^\alpha$, find $\left\{\rho^f, \mathbf{v}^f, \mathbf{u}^s\right\} \in \mathcal{S}^\rho \times \mathcal{S}^f \times \mathcal{S}^s$ such that $\forall g^f \in \mathcal{V}^\rho$ and $\forall \mathbf{w}^\alpha \in \mathcal{V}^\alpha$:

$$\left(g^f, \frac{\partial \rho^f}{\partial t} + (\mathbf{v}^f - \mathbf{v}^m) \cdot \nabla \rho^f + \rho^f \operatorname{div}(\mathbf{v}^f) - m^f \right)_\Omega = 0 \tag{2.38}$$

$$\left(\nabla \mathbf{w}^f, \mathbf{T}^f \right)_\Omega + \left(\mathbf{w}^f, \rho^f \frac{\partial \mathbf{v}^f}{\partial t} + \rho^f (\mathbf{v}^f - \mathbf{v}^m) \cdot \nabla \mathbf{v}^f - \rho^f \mathbf{b}^f - \mathbf{I}^f \right)_\Omega - \left(\mathbf{w}^f, \mathbf{T}^f \mathbf{n} \right)_{\partial\Omega_h^f} = 0 \tag{2.39}$$

$$\left(\nabla \mathbf{w}^s, \mathbf{T}^s \right)_\Omega + \left(\mathbf{w}^s, -\rho^s \mathbf{b}^s - \mathbf{I}^s \right)_\Omega - \left(\mathbf{w}^s, \mathbf{T}^s \mathbf{n} \right)_{\partial\Omega_h^s} = 0 \tag{2.40}$$

We append initial conditions such that $\forall \mathbf{x} \in \Omega_0$:

$$\rho^\alpha(\mathbf{x}, 0) = \rho_0^\alpha(\mathbf{x}) \tag{2.41}$$

$$\mathbf{v}^\alpha(\mathbf{x}, 0) = \mathbf{v}_0^\alpha(\mathbf{x}) \tag{2.42}$$

$$\mathbf{u}^s(\mathbf{x},0)=\mathbf{u}_0^s(\mathbf{x}) \tag{2.43}$$

The boundary conditions are:

$$\rho^f=\rho_b^f \quad \text{on } \partial\Omega_g^\rho\times]0,\mathrm{T}[\tag{2.44}$$

$$v_i^f=\bar{v}_i^f \quad \text{on } \partial\Omega_i^f\times]0,\mathrm{T}[\tag{2.45}$$

$$\mathbf{u}^s=\bar{\mathbf{u}}^s \quad \text{on } \partial\Omega_g^s\times]0,\mathrm{T}[\tag{2.46}$$

$$\mathbf{T}^\alpha\mathbf{n}=\mathbf{t}^\alpha \quad \text{on } \partial\Omega_h^\alpha\times]0,\mathrm{T}[\tag{2.47}$$

where $\partial\Omega_i^\alpha\cap\partial\Omega_{h,i}^\alpha=\varnothing$, $\partial\Omega_i^\alpha\cup\partial\Omega_{h,i}^\alpha=\partial\Omega$, and $\partial\Omega_{h,i}^s\cap\partial\Omega_i^f$ is not necessarily empty.

2.6 Discrete Problem and Stabilized Variational Form

The discrete problem is formulated by substituting the trial solutions ($\rho^f,\mathbf{v}^f,\mathbf{u}^s$) and the weighting functions ($g^f,\mathbf{w}^\alpha$) by the discrete trial and weighting functions belonging to the finite-dimensional subsets of the solution and weighting function spaces: $\mathcal{S}^{\alpha h}\subset\mathcal{S}^\alpha$, and $\mathcal{V}^{\alpha h}\subset\mathcal{V}^\alpha$. The discrete mixed form that emanates from (2.38)-(2.40) needs to overcome some technical issues, namely: (i) arbitrary combinations of interpolating functions for the various fields may not satisfy the discrete *inf-sup* condition, (ii) for small reaction rates, the balance of mass of the fluid becomes an advection-dominated equation, (iii) for large reaction rates, the balance of mass of the fluid becomes a singularly perturbed equation (reaction dominated) and it may exhibit boundary and/or internal layers, depending on the source terms (or forcing) and boundary conditions, and (iv) for large reaction rates and/or low diffusivity (large drag), the balance of linear momentum for the fluid also becomes a singularly perturbed PDE, thus exhibiting boundary and/or moving internal layers. To address these technical points, we have employed stabilized finite element methods. We first present the simplest stabilized method that is developed by applying VMS stabilization to the balance of mass for the fluid. The general derivation of the stabilized method presented in (2.48) follows along [13]. However, in the present extension the reaction term is assumed to be a function of not only the coarse-scale fluid density field, but also of the fine-scale fluid density. This results in the appearance of the reaction terms in the fine-scale model, which on subsequent substitution in the coarse scale equation gets manifested in the expression of the weighting term χ^ρ as well as in the denominator of τ_e^ρ. The superposed $(\bullet)$ denotes coarse-scale quantities, which are associated with the standard FEM spaces.

2.6.1 VMS stabilized form for the balance of mass of the fluid constituent

A premise of the VMS framework is that the unknown field and its weighting function are assumed decomposed into coarse-scale and fine-scale components. In [13] we developed a VMS-based stabilized form for the balance of mass equation for the fluid. The reaction term was assumed to be a function of the coarse scale density field alone. The resulting method showed smoother solution in the reaction active zone without any numerical oscillations. When we extended this method to chemo-mechanical coupling with accompanied mechanical deformation, under higher reaction rates, overshoots and undershoots were observed around the steep reaction zones. Therefore, in the present developments the first extension made is that the reaction term is now a function of both coarse scale as well as fine scale density field.

$$\left(\hat{g}^f, \frac{\partial \hat{\rho}^f}{\partial t} + (\mathbf{v}^f - \mathbf{v}^m)\cdot\nabla\hat{\rho}^f + \hat{\rho}^f \mathrm{div}(\mathbf{v}^f) - \hat{m}^f\right)_\Omega + \sum_e \left(\chi^\rho, \tau_e^\rho \hat{R}^\rho\right)_{\Omega^e} = 0 \tag{2.48}$$

where

$$\hat{R}^\rho = \frac{\partial \hat{\rho}^f}{\partial t} + (\mathbf{v}^f - \mathbf{v}^m)\cdot\nabla\hat{\rho}^f + \hat{\rho}^f \mathrm{div}(\mathbf{v}^f) - \hat{m}^f \tag{2.49}$$

$$\chi^\rho = \hat{g}^f \frac{1}{\Delta t} + \hat{g}^f \mathrm{div}(\mathbf{v}^m) - \hat{g}^f \frac{\partial m^f}{\partial \rho^f} - (\mathbf{v}^f - \mathbf{v}^m)\cdot\nabla\hat{g}^f \tag{2.50}$$

$$\tau_e^\rho = -\left(b^e, b^e/\Delta t + (\mathbf{v}^f - \mathbf{v}^m)\cdot\nabla b^e + b^e \mathrm{div}(\mathbf{v}^f) - b^e \frac{\partial m^f}{\partial \rho^f}\right)_{\Omega^e}^{-1} b^e \left(b^e, 1\right)_{\Omega^e} \tag{2.51}$$

Equation (2.48) together with (2.39) and (2.40) constitute the resulting method for the discrete system of equations. The mathematical attributes of this stabilized method are numerically tested in section 2.7.3.

2.6.2 VMS stabilized form for the balance of linear momentum of the fluid constituent

Another numerical deficiency arises when the diffusivity of the unreacted solid is very low. In this case the balance of linear momentum for the fluid becomes a singularly perturbed equation where zeroth-order term becomes the dominant term. We follow along the lines of

[29,30] and carry out an overlapping additive decomposition of the velocity field (and its weighting function), that subsequently leads to stabilization of the velocity field, shown below.

$$\begin{aligned}&\left(\nabla\hat{\mathbf{w}}^f,\,\hat{\mathbf{T}}^f\right)_\Omega+\left(\hat{\mathbf{w}}^f,\,\rho^f\frac{\partial\hat{\mathbf{v}}^f}{\partial t}+\rho^f(\hat{\mathbf{v}}^f-\mathbf{v}^m)\cdot\nabla\hat{\mathbf{v}}^f-\rho^f\mathbf{b}^f-\hat{\mathbf{I}}^f\right)_\Omega\\&\qquad-\left(\hat{\mathbf{w}}^f,\,\mathbf{t}^f\right)_{\partial\Omega_h^f}+\sum_e\left(\chi^v,\,\tau_e^v\hat{\mathbf{R}}^v\right)_{\Omega^e}+\sum_e\left(\rho^f(\nabla\hat{\mathbf{v}}^f)\cdot\hat{\mathbf{w}}^f,\,\tau_e^v\hat{\mathbf{R}}^v\right)_{\Omega^e}=0\end{aligned}\tag{2.52}$$

where

$$\hat{\mathbf{R}}^v=\rho^f\frac{\partial\hat{\mathbf{v}}^f}{\partial t}+\rho^f\left(\hat{\mathbf{v}}^f-\mathbf{v}^m\right)\cdot\nabla\hat{\mathbf{v}}^f-\operatorname{div}(\hat{\mathbf{T}}^f)-\rho^f\mathbf{b}-\hat{\mathbf{I}}^f\tag{2.53}$$

$$\begin{aligned}\chi^v&=-\frac{1}{\Delta t}\hat{\mathbf{w}}^f+\operatorname{div}[2\mu\eta\cdot\operatorname{sym}(\nabla\hat{\mathbf{w}}^f)]\\&\quad+\operatorname{div}[\rho^f(\hat{\mathbf{v}}^f-\mathbf{v}^m)]\hat{\mathbf{w}}^f+\rho^f(\hat{\mathbf{v}}^f-\mathbf{v}^m)\cdot\nabla\hat{\mathbf{w}}^f+\frac{\partial\hat{\mathbf{I}}^f}{\partial\mathbf{v}^f}\hat{\mathbf{w}}^f\end{aligned}\tag{2.54}$$

Equation (2.52) together with (2.38) and (2.40) constitute the resulting method. The above equations can be further simplified by taking into account the low diffusivity (high drag coefficient) and high reaction rates (considered in this chapter), which result in a dominant interactive force term. Thus, only the quantities that are associated with the interactive force are retained in the weighting expression of the stabilization terms (as well as in the denominator of the stabilization tensor $\boldsymbol{\tau}_e^v$), which leads to a simplified form

$$\begin{aligned}&\left(\nabla\hat{\mathbf{w}}^f,\,\hat{\mathbf{T}}^f\right)_\Omega+\left(\hat{\mathbf{w}}^f,\,\rho^f\frac{\partial\hat{\mathbf{v}}^f}{\partial t}+\rho^f(\hat{\mathbf{v}}^f-\mathbf{v}^m)\cdot\nabla\hat{\mathbf{v}}^f-\rho^f\mathbf{b}^f-\hat{\mathbf{I}}^f\right)_\Omega\\&\qquad-\left(\hat{\mathbf{w}}^f,\,\mathbf{T}^f\mathbf{n}\right)_{\partial\Omega_h^f}+\sum_e\left(\frac{\partial\hat{\mathbf{I}}^f}{\partial\mathbf{v}^f}\hat{\mathbf{w}}^f,\,\boldsymbol{\tau}_e^v\hat{\mathbf{R}}^v\right)_{\Omega^e}=0\end{aligned}\tag{2.55}$$

where

$$\boldsymbol{\tau}_e^v=\left(b^e,\,-\frac{\partial\mathbf{I}^f}{\partial\mathbf{v}^f}b^e\right)_{\Omega^e}^{-1}b^e\left(b^e,\,1\right)_{\Omega^e}\tag{2.56}$$

The numerical features of this method are investigated via thermal oxidation of silicon carbide, presented in section 2.7.1, and the thermal oxidation of silicon in sections 2.7.2 and 2.7.6. To highlight the mathematical attributes of the stabilization of momentum balance equation of fluid, we have suppressed the stabilized terms for balance of mass of the fluid, presented in (2.48)-(2.51).

Remark: The more general form of the stabilized method is in equation (2.52), while the method implemented after applying simplifying assumptions is in equation (2.55).

2.7 Numerical Results

This section presents numerical test cases to highlight chemo-mechanical coupling in the mixture model and the stable response produced by the proposed numerical method. Chemical reactions induce two types of effects in the mechanical response: (i) the mechanical properties of the reacted solid evolve, and (ii) the chemo-mechanical coupling induces large deformations and volumetric growth that are accompanied with internal stresses. The method has been implemented with 1D linear and 3D trilinear Lagrange elements, employing sufficiently high quadrature rules for full numerical integration. Backward Euler method is employed for pseudo-time evolution of the internal variables model, as well as for the problems involving momentum balance of the fluid.

2.7.1 Thermal oxidation of silicon carbide (SiC)

The first test case is the thermal oxidation of silicon carbide (SiC), shown schematically in Figure 2.1. Oxidation of SiC is given by the following chemical reaction:

$$\mathrm{SiC}_{(s)} + 1.5\mathrm{O}_{2(g)} \rightarrow \mathrm{SiO}_{2(s)} + \mathrm{CO}_{(g)} \tag{2.57}$$

Solid SiC reacts with gaseous O_2 to produce solid silicon dioxide (SiO_2) and gaseous carbon monoxide (CO), which is assumed to be released to the surroundings. This chemical reaction is accompanied by volume expansion wherein the volume of the reacted oxide is 2.2 times that of the SiC substrate that it replaces [31].

We study the thermal oxidation of the *a*-face of 4H-SiC. The material parameters were estimated from the data in [32], [33], and [34], and are summarized in Table 2.1. The governing system of equations is given by the mass balance of the fluid and the balance of linear momentum of both the fluid and the solid. The VMS stabilized method in equation (2.55) is employed for the fluid balance of linear momentum. Because of the high reaction rate and low diffusivity of the solid material, the zeroth-order entities in the weighting slot of the stabilizing term are considered dominant.

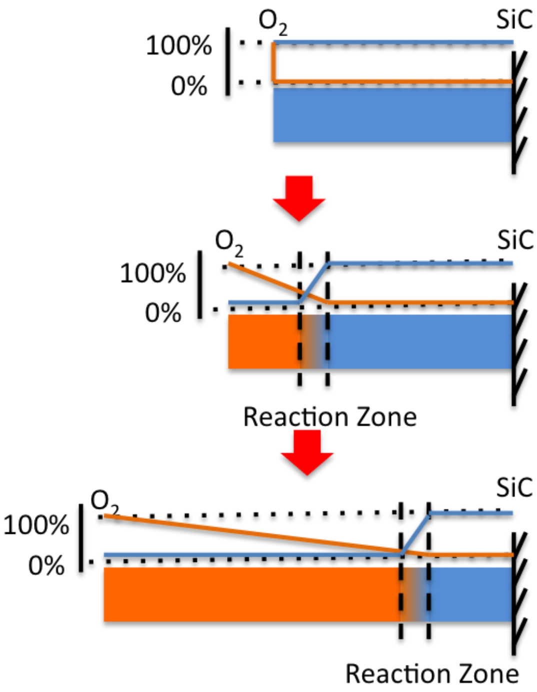


Figure 2.1. Schematic diagram of the thermal oxidation process in silicon carbide (SiC) showing SiC recession and SiO2 growth.

Table 2.1. Material Parameters for SiC Oxidation[2].

Parameter	Units	Value
Temperature (θ)	K	1373
Gas constant to O_2 molecular weight ratio ($\bar{R}$)	J / K-kg	259
O_2 reference concentration (ρ_T^f)	Kg/m^3	0.23
O_2 viscosity (η)	Kg/m-s	5.745e-5
Diffusivity in SiO_2 (D_{11})	m^2/s	1.6213e-12
SiC Shear modulus (μ^s)	Pa	180e9
SiC Lame second parameter (k^s)	Pa	120e9
SiO_2 Shear modulus (μ^s)	Pa	40.921e9
SiO_2 Lame second parameter (k^s)	Pa	12.158e9

---

[2] The value of coefficient for viscosity of oxygen here corresponds to the temperature of 1373 K and partial pressure 1 atm of pure oxygen. As oxygen permeates through the solid, the jump in concentration of oxygen between the solid and the ambient surrounding is given by Henry's Law. Accordingly, the coefficient used in the numerical calculation is 2.792e-7.

Table 2.1 (cont.)

SiC density (ρ_T^s)	Kg/m^3	3.215e3
Reaction rate (R_0)	s^{-1}	-5.6e5
Evolution parameter (α)	s-m^3/kg	2.603e-4
Reacted state (ϕ_0)	dimensionless	0.0
Mass conversion ratio (r_m)	dimensionless	0.41646
Chemo-mechanical coupling parameter ($\overline{c}^0$)	Pa-kg^3/m^9	51.8116

A domain of length L = 0.5 μm is discretized with 100 eight-node brick elements. We apply appropriate boundary conditions in the lateral directions to induce 1D deformation along the x-axis. The fluid mass concentration is prescribed at the left end of the domain (inlet, x = 0), and the value is linearly increased to $0.00486\rho_T^f$ (saturation concentration of oxygen in SiO_2) over a period of 10,000s to account for the time it takes for atmospheric O_2 to impregnate the solid SiO_2 phase at the interface. These parameters were estimated based on description provided in [33]. The fluid velocity is constrained to be zero at the right end of the domain (x = L), and Neumann BCs (depending on fluid concentration) are applied for the fluid at the inlet. The solid is fixed at the right end (x = L) but is free to move at the left end of the domain (x = 0).

The value of the reaction state variable ϕ =1 corresponds to the unreacted material, while a value of $\phi = \phi_0 = 0$ corresponds to the fully reacted state. Thus, the region where reaction is taking place is characterized as the zone with $\phi_0 < \phi < 1$. As time progresses and more oxygen diffuses through the solid domain to the reaction zone, the reaction front advances into the unreacted material, which corresponds to the steep gradient in the lines presented in Figure 2.2. In the numerical simulations the oxide layer thickness is measured as the distance in the current configuration between the inlet and the point within the reaction zone where $\phi = 0.5$. Figure 2.3 shows the evolution of the oxide layer thickness and compares it with the experimental data for the thermal oxidation of the *a*-face of 4H-SiC at 1100 C (1373 K) reported in [34]. The numerical results are in good agreement with the experimental data.

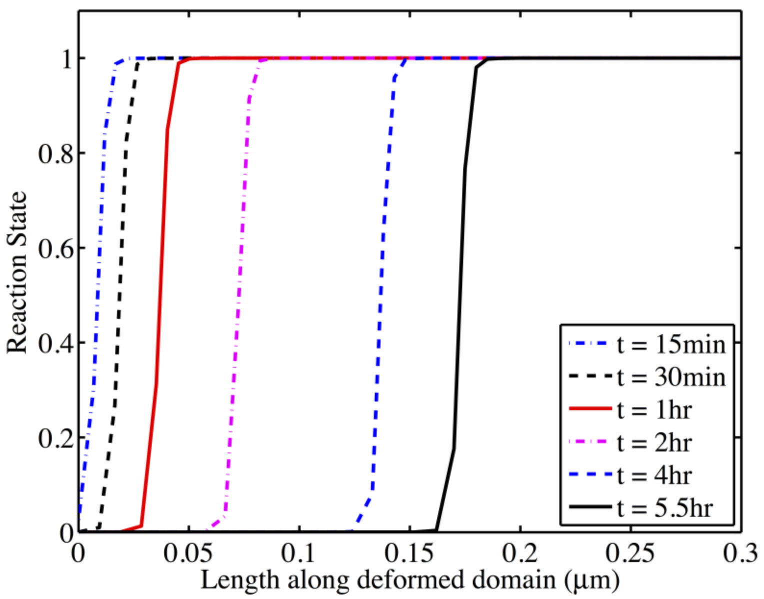


Figure 2.2. Evolution of the reaction state variable ϕ over the deformed domain at t=15min, 30 min, 1hr, 2hr, 4hr, and 5.5hr.

Figure 2.4 shows snapshots of the deformed domain corresponding to the time levels from Figure 2.2, where reaction zone is shown via colored contours of the reaction state variable. In Figure 2.5a, the nodally projected values of the transverse stress, induced due to lateral constraining of the nodes along the domain, are plotted for different time levels during problem evolution. For unconstrained isotropic expansion, the material would expand in every direction, but the domain is constrained in the transverse directions by the adjacent material, thus producing compressive stresses (negative values in Figure 2.5a) in the reacted zone. In addition, Figure 2.5b shows the values of the nodally projected axial stress.

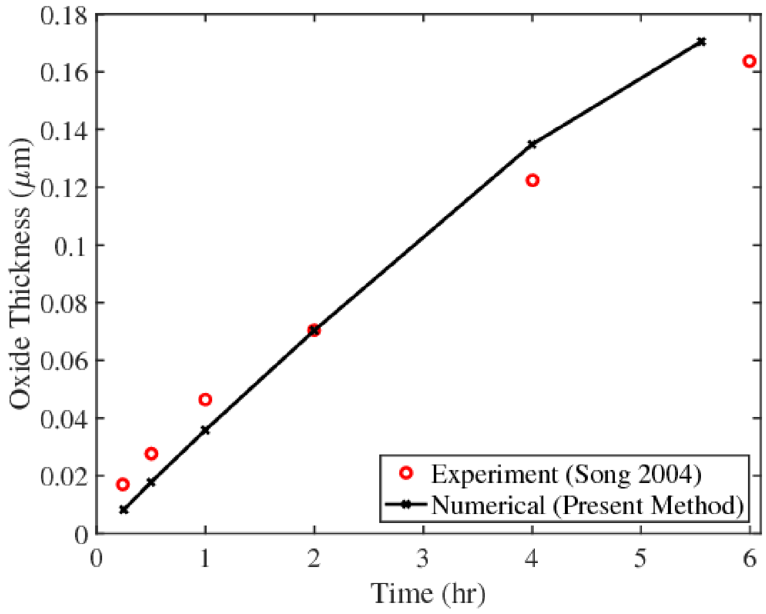


Figure 2.3. Evolution of oxide thickness and comparison with experimental data [34].

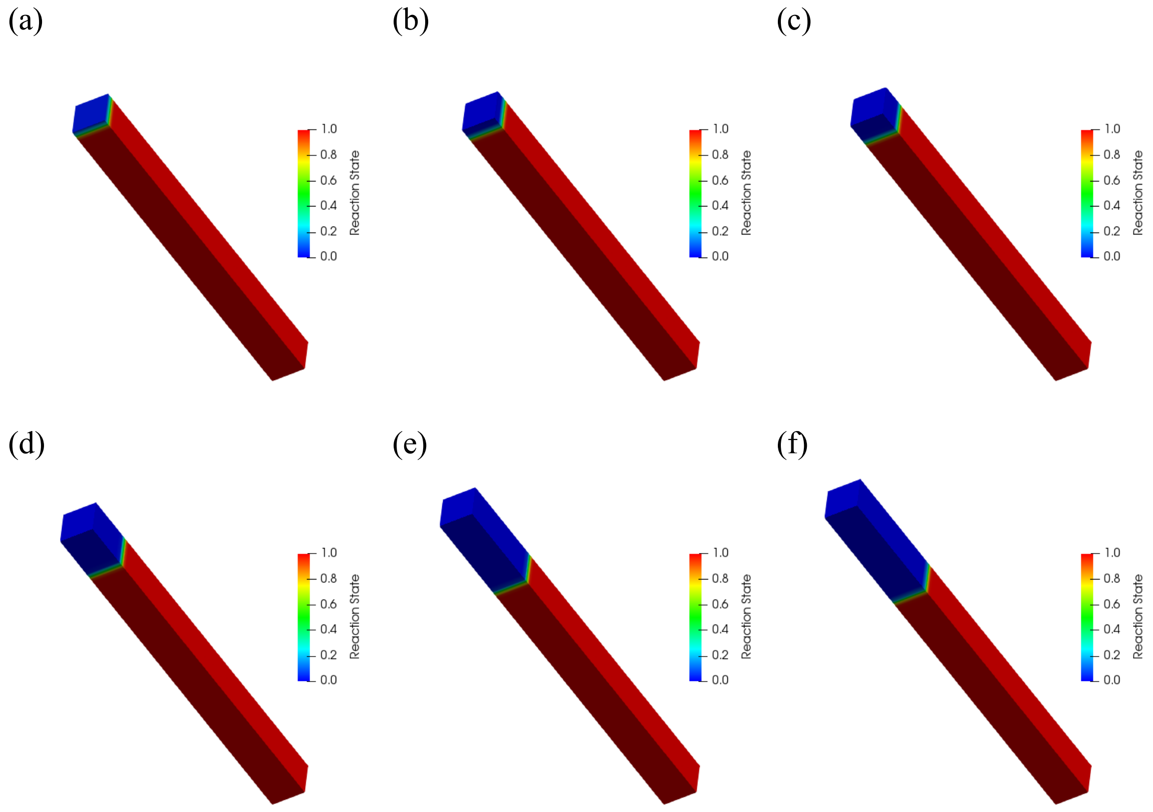

(a)
(b)
(c)
(d)
(e)
(f)
1.0
0.8
0.6
0.4
0.2
0.0
Reaction State

Figure 2.4. Snapshots of the oxidizing SiC bar at t = (a) 15min, (b) 30min, (c) 1hr, (d) 2hr, (e) 4hr, and (f) 5.5hr. The legend represents the reaction state where $\phi = 1.0$ corresponds to unreacted material (SiC), and $\phi = 0.0$, to fully reacted material (SiO2).

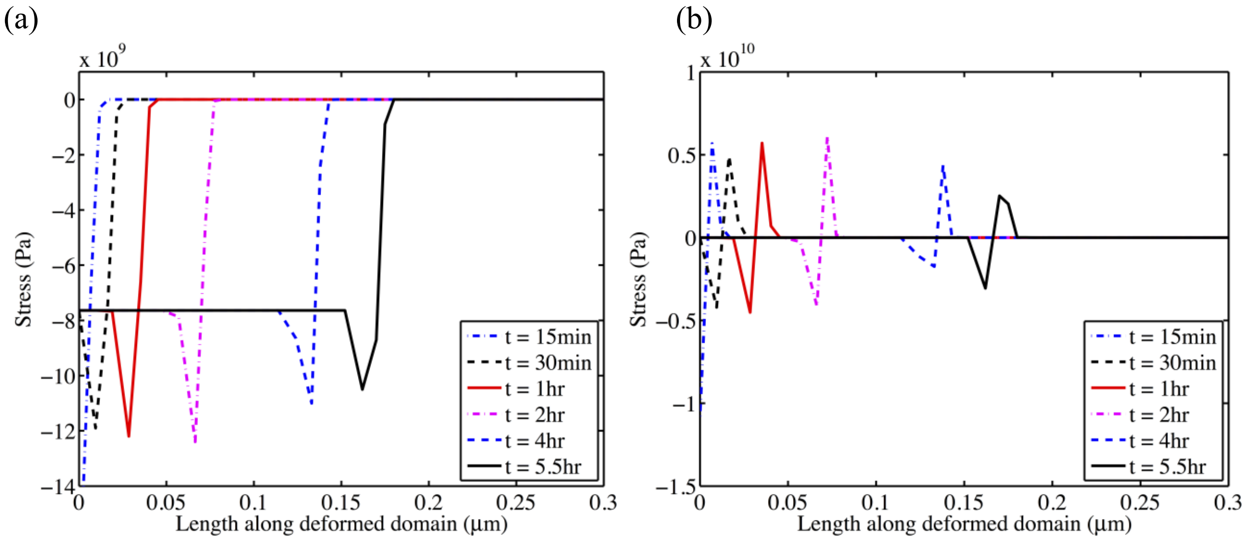


Figure 2.5. Snapshots of (a) the transverse stress and (b) the axial stress over the deformed domain at t = 15min, 30min, 1hr, 2hr, 4hr, and 5.5hr. The value stress is zero beyond 0.3 μm in both cases.

2.7.2 Oxidation-induced bending of silicon wafer

This test case considers thermal oxidation of a silicon (Si) wafer as described in [35]. The process of thermal oxidation of silicon (Si) follows the chemical reaction

$$\mathrm{Si}_{(s)} + \mathrm{O}_{2(g)} \rightarrow \mathrm{SiO}_{2(s)} \tag{2.58}$$

This reaction also induces an expansion in reacted solid, which is characterized by a volume increase of 120% [7]. The material parameters for Si and SiO_2 were determined based on the description in [35] and all the constitutive parameters are given in Table 2.2. The domain represents a slice of the wafer cross-section with dimension of 5μm × 1μm. The domain is discretized using a grid of eight-node brick elements with 25 elements along the x-direction and 20 along the y-direction (appropriate boundary conditions are applied to induce a 2D response). The horizontal displacement of the solid is set equal to zero along the left end of the domain (x = 0). The top surface of the domain is assumed to be exposed to an oxidizing environment with a O_2 partial pressure of 1atm at a temperature of 1373 K. The mass concentration of the fluid (O_2) at the top bounding surface is linearly increased to $0.00486\rho_T^f$ over a period of 10,000s as in the test case presented in section 2.7.1. The stabilized method presented in equation (2.55) is employed in this test case. The resulting stress values in this section are reported in units of dyn/cm^2 in order to facilitate comparison with the results in [35].

Table 2.2. Material Parameters for Silicon (Si) Oxidation[3] in Section 2.7.2.

Parameter	Units	Value
Temperature (θ)	K	1373
Gas constant to O_2 molecular weight ratio ($\bar{R}$)	J / K-kg	259
O_2 reference concentration (ρ_T^f)	Kg/m^3	0.23
O_2 viscosity (η)	Kg/m-s	5.745e-5
Unreacted Diffusivity (D_{11}^s)	m^2/s	1.3786e-12
Unreacted Shear modulus (μ^s)	Pa	73.05e9
Unreacted Lame second parameter (k^s)	Pa	141.67e9
Reacted Diffusivity (D_{11}^s)	m^2/s	1.3786e-10
Reacted shear modulus (μ^s)	Pa	105.65e9
Reacted Lame second parameter (k^s)	Pa	12.158e9
Si density (ρ_T^s)	Kg/m^3	2.329e3
Reaction rate (R_0)	s^{-1}	-5.6e5
Evolution parameter (α)	s-m^3/kg	3.769e-4
Reacted state (ϕ_0)	dimensionless	0.0
Mass conversion ratio (r_m)	dimensionless	1
Chemo-mechanical coupling parameter ($\bar{c}^0$)	Pa-kg^3/m^9	20.798

In Figure 2.6 we present the spatial distribution and contours of the T_{11} component of the Cauchy stress for the solid (in units of dyn/cm^2) in the deformed configuration of the wafer at time t = 410s. The wafer's cross-section, which was originally straight, undergoes bending-dominated response, which is induced by the expanding SiO_2 formed along the top surface of the domain as a product of the chemical reaction. As the SiO_2 material expands, it pulls on the unreacted Si material that lies on the other side of the reaction zone, which induces a tensile stress in the T_{11} component. Figure 2.6 also shows black contour lines for stress values of 3e10, 1.8e10, 6e9, -6e9, -1.8e10, and -3e10 dyn/cm^2 that are picked to compare the state of stress and deformation with the data in the figure presented in plate 2 of [35]. This comparison is also shown in Figure 2.7 and Figure 2.8, where the data from plate 2 in [35] was digitized from an electronic version of the figure and appears as dashed or continuous black lines in the figures below while the data from the present chapter is indicated with discrete markers. We remark that the results in plate 2 of [35] correspond to time level of t = 100 min in that simulation, while a similar state of deformation was attained at time level t = 410s in the present chapter. The discrepancy may be attributed to the difference in the boundary conditions, namely the mass concentration and flux of O_2 at the top surface of the rectangular cross-section, which depends on the O_2 partial pressure and the temperature of the oxidizing environment.

---

[3] The value of the reacted shear modulus is determined such that the "effective" shear modulus of the reacted material $\mu_2^s - \mu^L \gamma_{max}^3 \bar{c}^0 / 3$ in equation (2.28) corresponds to the shear modulus of silicon dioxide employed in [35], which is 40.921 GPa.

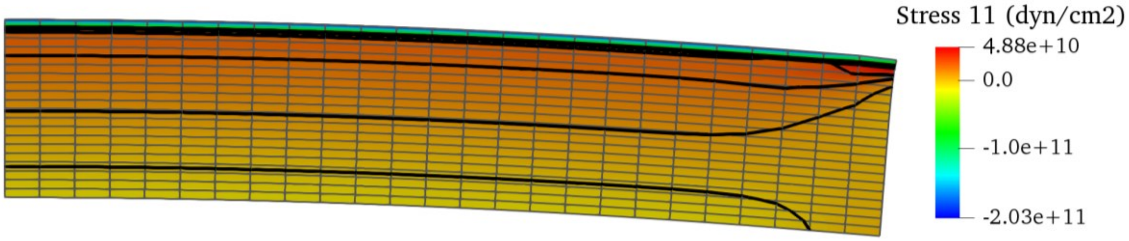


Figure 2.6. T_{11} component of Cauchy stress in the Si wafer in its deformed configuration at t = 410s.

Figure 2.7 and Figure 2.8 show a good agreement between the results reported in [35] and the present computations. However, one discrepancy is the location of the region delineated by the contour corresponding to a stress value of 3e10 dyn/cm^2, which is closer to the top surface in the present chapter than it is in the data presented in [35]. In addition, the magnitudes of the extreme values of the T_{11} component of stress are comparable, but they differ somewhat: 4.85e10 and -1.28e11 dyn/cm^2 in [35] vs. 4.88e10 and -2.03e11 dyn/cm^2 in the present chapter. A possible reason for the differences is that different models are used to represent the elastic response of the solid. Furthermore, the work in [35] accounts for viscoelasticity of the solid while in the present model we have neglected this effect.

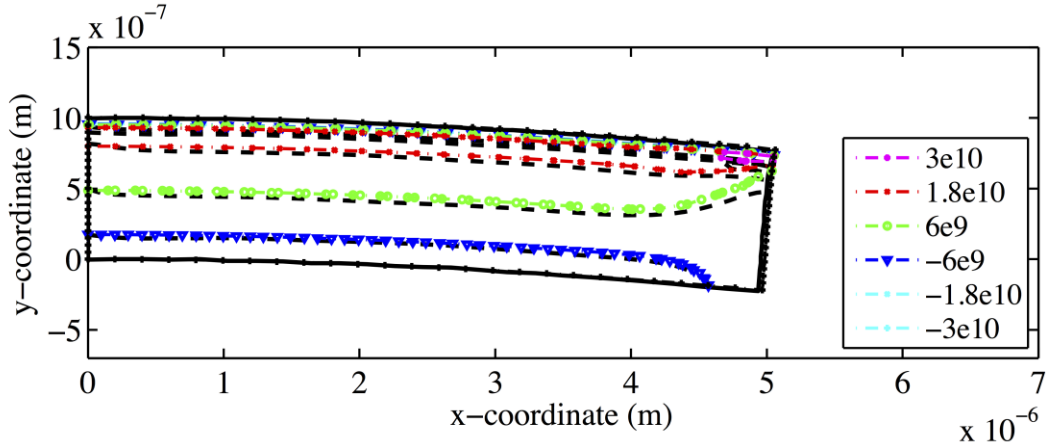


Figure 2.7. Comparison of the contours of the T_{11} component of stress with the data from [35].

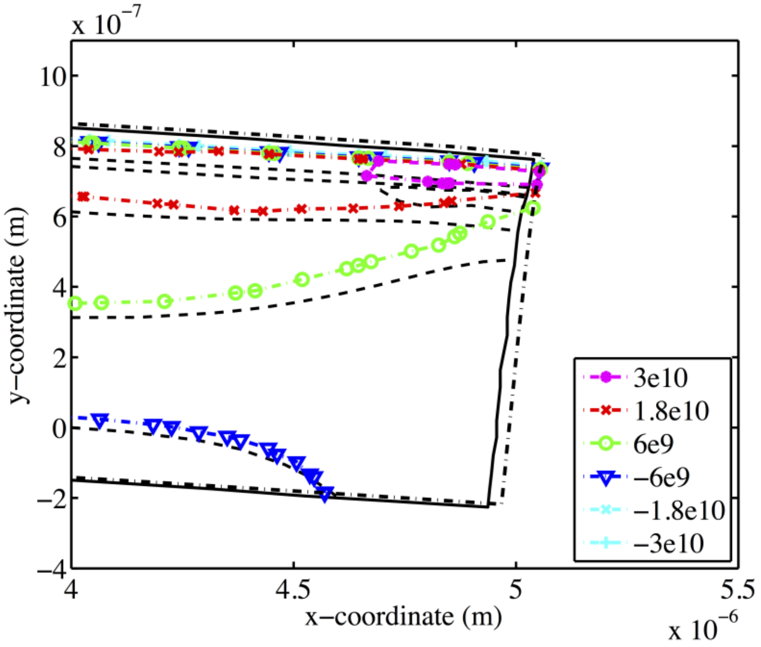


Figure 2.8. Close-up of the contours for the stress component T_{11} and comparison with the data from [35]

Figure 2.9 shows snapshots of the deformed configuration of the wafer cross-section and spatial distribution of the reaction state for time levels t = 500s, t = 1000s, t = 1500s, and t = 2000s. As the reaction proceeds, the advancing reaction moves farther down into the wafer. In the narrow reaction zone, the T_{11} component of stress goes from its maximum compressive value (on the oxide side) to its maximum tensile value (on the Si side).

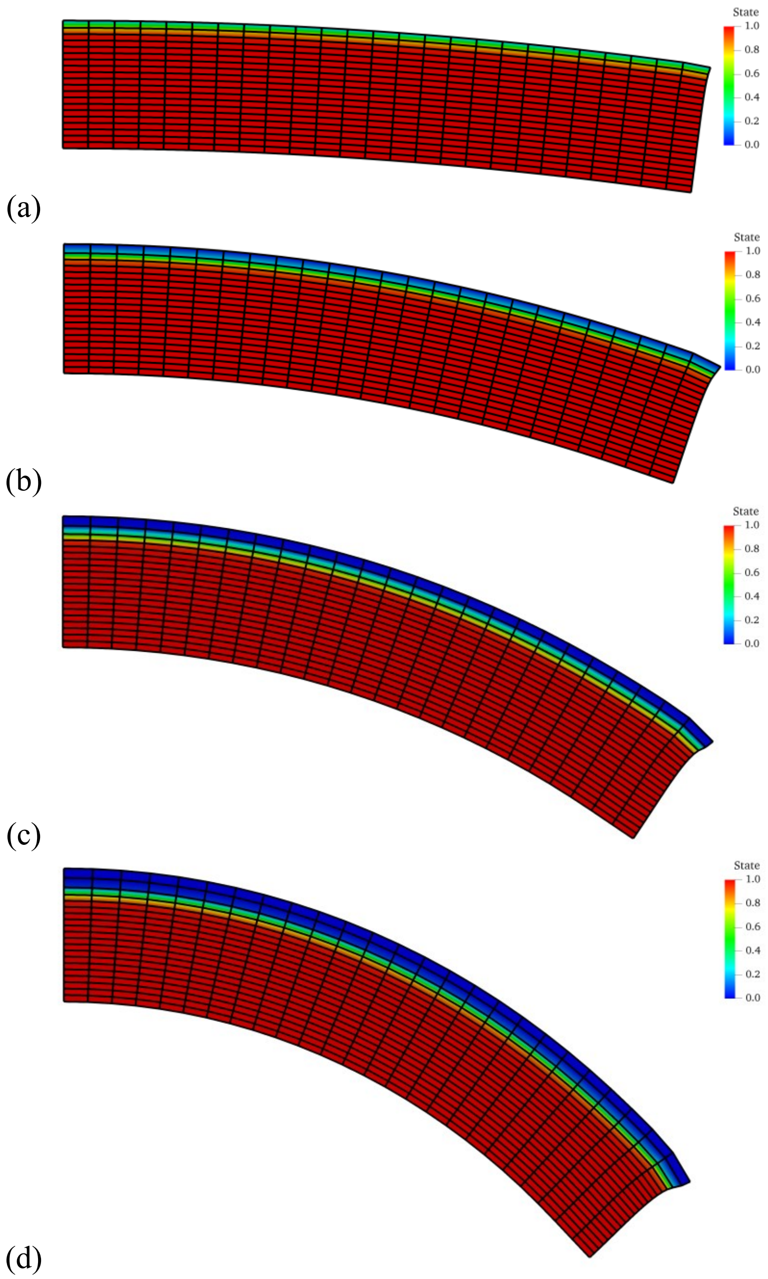


Figure 2.9. Snapshots of the deformation and reaction state contours at time levels (a) t = 500s, (b) t = 1000s, (c) t = 1500s, and (d) t = 2000s.

Figure 2.10 shows profiles of the T_{11} component of the Cauchy stress in the solid measured at x = 0 (the left end of the domain) along the y-axis in the current configuration for different time levels. The reacted material (near the top surface), undergoes compression (negative T_{11} stress values) because its expansion in the horizontal direction is constrained by the unreacted material across the advancing reaction front. On the other hand, the unreacted material adjacent to the reaction zone undergoes tension (positive T_{11} stress) as the expanding material tries to stretch it tangential to the advancing front. As oxidation continues, a larger region near the top of the wafer cross-section undergoes compression, while the tensile stresses across the reaction zone become larger in magnitude, which corresponds to further bending of the wafer.

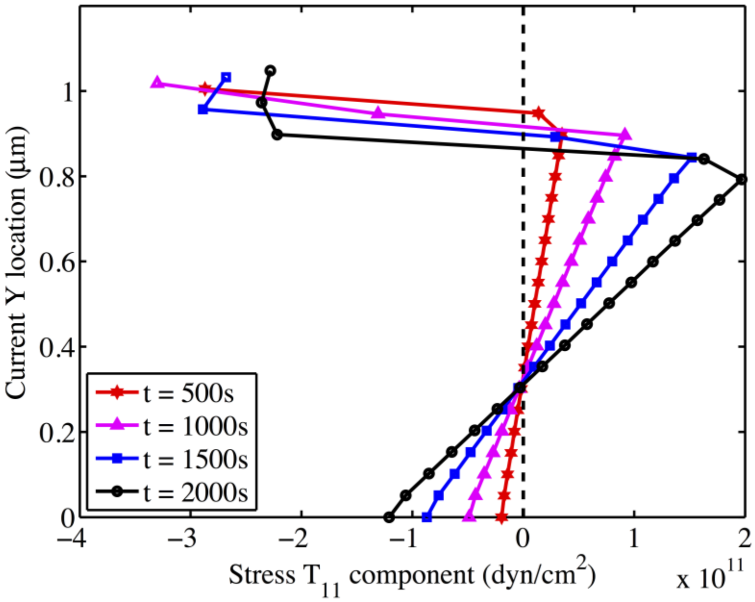


Figure 2.10. Spatial distribution along the y-axis of the nodally projected T_{11} component of stress at different time levels.

2.7.3 Isothermal chemo-mechanical response in 1D

This test case investigates the mathematical attributes of the stabilized method presented in equations (2.48)-(2.51) in section 6.1. We focus on the inclusion of the reaction term in the stabilized form presented in (2.48) to highlight the significance of this term for the higher intensity rates of chemical reaction. We keep the problem description simple and consider chemo-mechanical coupled response under isothermal conditions in a one-dimensional domain given by $\bar{\Omega} = [0, L]$ with L = 2μm. For this test problem, we adopt the one-dimensional constitutive model presented in reference [13] (i.e., equations (22), (23), (25), (30)-(33) therein). The temperature is taken to be spatially uniform and is held constant at 1373 K. Homogeneous Dirichlet boundary conditions at $x = L$ are imposed for the fluid velocity and solid displacement fields. The fluid density is prescribed at $x = 0$ and it is varied linearly from zero to its reference value over a period of 0.2s. A traction free boundary condition for the solid is imposed at $x = 0$. For this isothermal case, we solve for the fluid and solid balance of linear momentum, and the solid and fluid conservation of mass equations. The material parameters for the solid-fluid system are given in Table 2.3 and the spatial distribution of the computed solution fields at 0.2 s are shown in Figure 2.11. The computed fluid and solid densities are normalized by their reference values listed in Table 2.3. Figure 2.11a shows influx of oxygen at the left end and Figure 2.11b shows the evolution of spatial density where a steep gradient in the density of solid, which also marks the active zone of reaction, can clearly be seen. We observe a drop in solid density caused by the volumetric expansion of the solid, which is reflected in the solid displacement field (Figure 2.11d). The velocity of the fluid in the reaction zone is higher than in the rest of the domain due to steeper gradient in the concentration of the fluid and the chemical reaction that consumes the fluid in a narrow band and therefore locally increases its velocity. Figure 2.11d shows the expansion of the solid where the reacted material expands in the negative x-direction while the solid ahead of the reaction zone is still stationary.

Table 2.3. Material parameters for 1D isothermal problem in Section 2.7.3.

Quantity	Units	Value
Temperature (θ)	K	1373
Gas Constant to Molecular weight ratio ($\bar{R}$)	J / K-kg	259.8
Fluid reference concentration (ρ_T^f)	Kg/m^3	0.1077
Fluid viscosity (η)	Kg/m-s	5.745e-5
Solid diffusivity (D_{11})	m^2/s	8.1e-12
Young Modulus (E^s)	Pa	432.7e9
Solid reference concentration (ρ_T^s)	Kg/m^3	32.15
Reaction rate (R_0)	s^{-1}	-2e4
Evolution parameter (α)	s-m^3/kg	2.599e-4
Reacted state (ϕ_0)	dimensionless	0.91
Mass conversion ratio (r_m)	dimensionless	0.832
Chemo-mechanical coupling parameter ($\bar{c}^0$)	Pa-kg^3/m^9	1.15e4

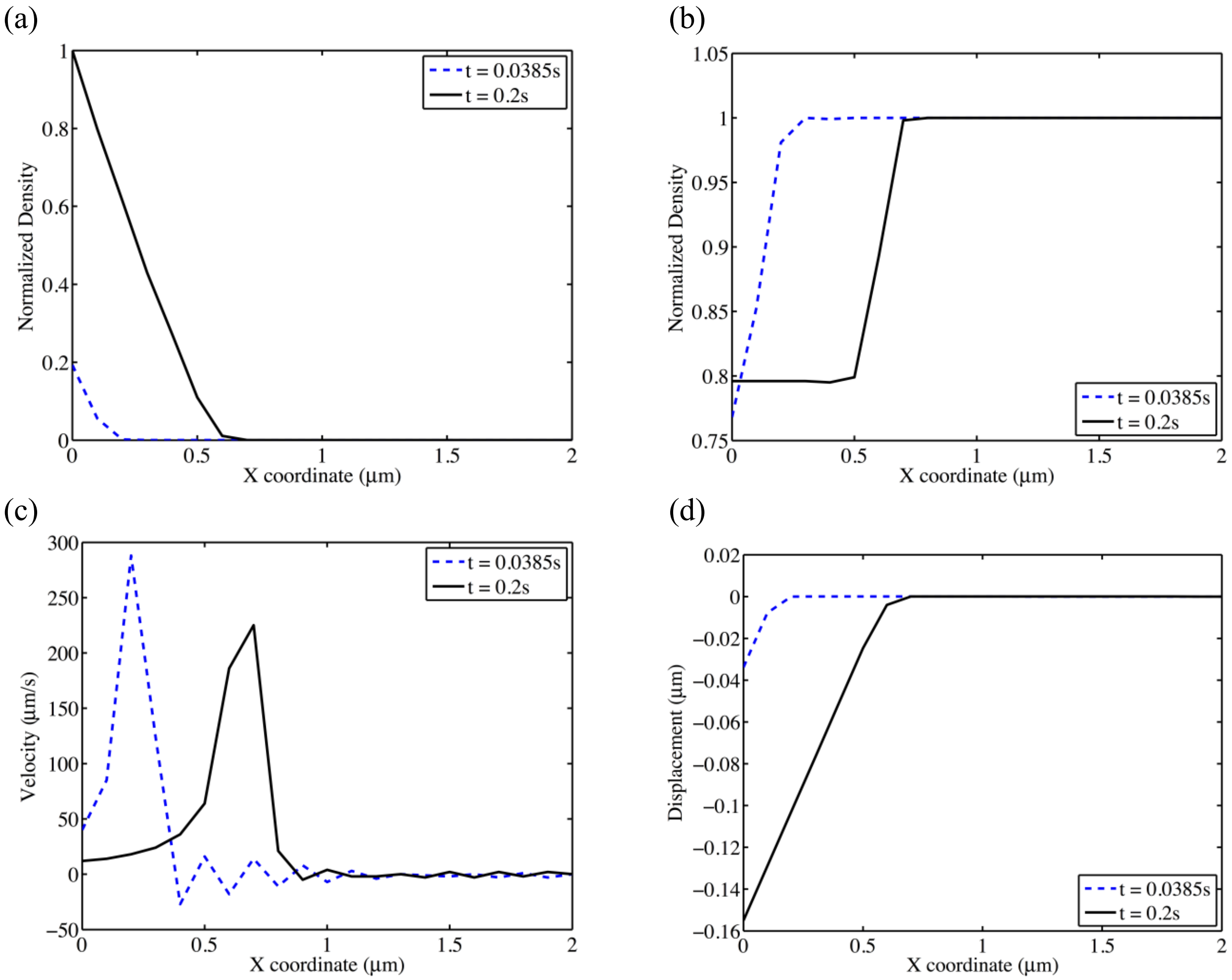


Figure 2.11. Solution fields of the 1D isothermal case along the undeformed domain for two different times: (a) fluid density; (b) solid density; (c) fluid velocity; and (d) solid displacement.

We remark that it is crucial to include the fine-scale reaction terms in the stabilized method (2.48) to be able to compute the solution in the entire time interval of interest. In fact, absence of these terms leads to numerical instabilities that grow out of control, resulting in non-convergence shortly after t = 0.0385s. Figure 2.12 shows a comparison of the velocity fields at t = 0.0385s obtained with the two stabilized methods: one that omits the fine-scale reaction terms in the stabilized form, as in [13], and the other from equation (2.48), which includes these terms. It is important to note that, although in both cases the stabilization is performed for the fluid density field, stability (or lack of it) also manifests itself in the other fields due to the coupled nature of the system of equations.

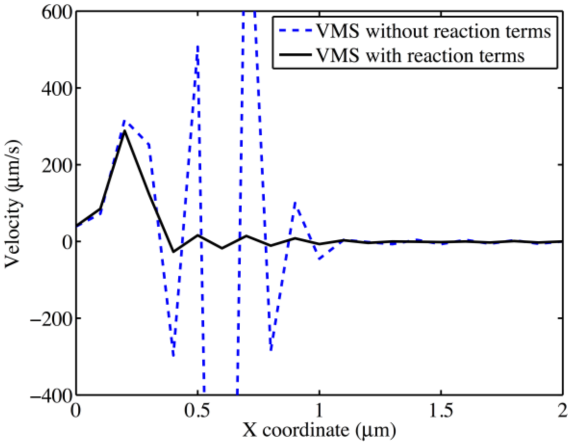


Figure 2.12. Comparison of fluid velocity field at t=0.0385s for the case in which the fine-scale reaction terms are omitted and the case in which they are included.

2.7.4 Chemically reacting flow through a rigid solid in 2D

This is a problem exhibiting a 2D response in a rectangular domain where the solid constituent is assumed to be rigid in order to highlight the response of the fluid constituent. In this test case, as well as in section 2.7.5, the dependence of the drag term on the fluid concentration is replaced by the fluid reference concentration, and the stabilized method employed is Streamline Upwind Petrov-Galerkin (SUPG) for the conservation of mass of the fluid constituent (interested reader is referred to [36,37] for discussion on the SUPG method). Figure 2.13 shows the schematic of the domain geometry and boundary conditions.

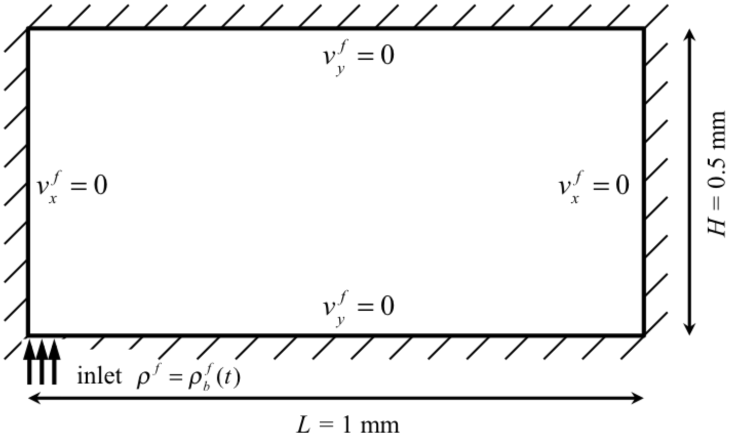


Figure 2.13. Schematic diagram of the problem: domain dimensions, solid boundary conditions, and the location and direction of the fluid (gas) inflow.

The discretization consists of one layer of 40×20 eight-node brick elements, constrained to exhibit a 2D response. No outflow and slip boundary conditions are applied to the fluid along the entire boundary, except at the inlet (part exposed to fluid), which lies at the bottom left corner of the domain with a width of *L*/10. The value of the fluid mass concentration at the inlet is increased linearly over a period of 10s until it reaches the reference value for the fluid concentration reported in Table 2.4, and it is held constant thereafter. The initial fluid mass concentration is zero elsewhere, and the initial solid mass concentration is uniform. The Backward Euler time integration scheme is used with a time step of 1s. The body force is ignored and the material parameters used are presented in Table 2.4. The solution fields at t = 300s for the fluid velocity in the x and y directions, fluid mass concentration, and solid mass concentration are shown in Figure 2.14. The spatial distribution of fluid concentration field is consistent with a diffusion problem and the fluid velocity field shows peak values near the inlet. In addition, there is a clear distinction between the solid densities for the fully reacted and unreacted materials. This shows that the solid density field can also serve as a surrogate indicator to distinguish between reacted and unreacted regions, while the true measure are the reaction state ϕ or the extent of reaction γ.

Table 2.4. Material parameters for the 2D problem in Section 2.7.4.

Quantity	Units	Value
Temperature (θ)	K	561
Gas Constant to Molecular weight ratio ($\bar{R}$)	J / K-kg	287
Fluid reference concentration (ρ_T^f)	Kg/m^3	0.23
Fluid viscosity (η)	Kg/m-s	1e-3
Solid diffusivity (D_{11})	m^2/s	1e-10
Shear Modulus (μ^s)	Pa	5e9
Lame second parameter (k^s)	Pa	3e9
Solid reference concentration (ρ_T^s)	Kg/m^3	10.0
Reaction rate (R_0)	s^{-1}	0.1
Evolution parameter (α)	s-m^3/kg	5.0
Reacted state (ϕ_0)	dimensionless	0.1
Mass conversion ratio (r_m)	dimensionless	1.0
Chemo-mechanical coupling parameter ($\bar{c}^0$)	Pa-kg^3/m^9	0

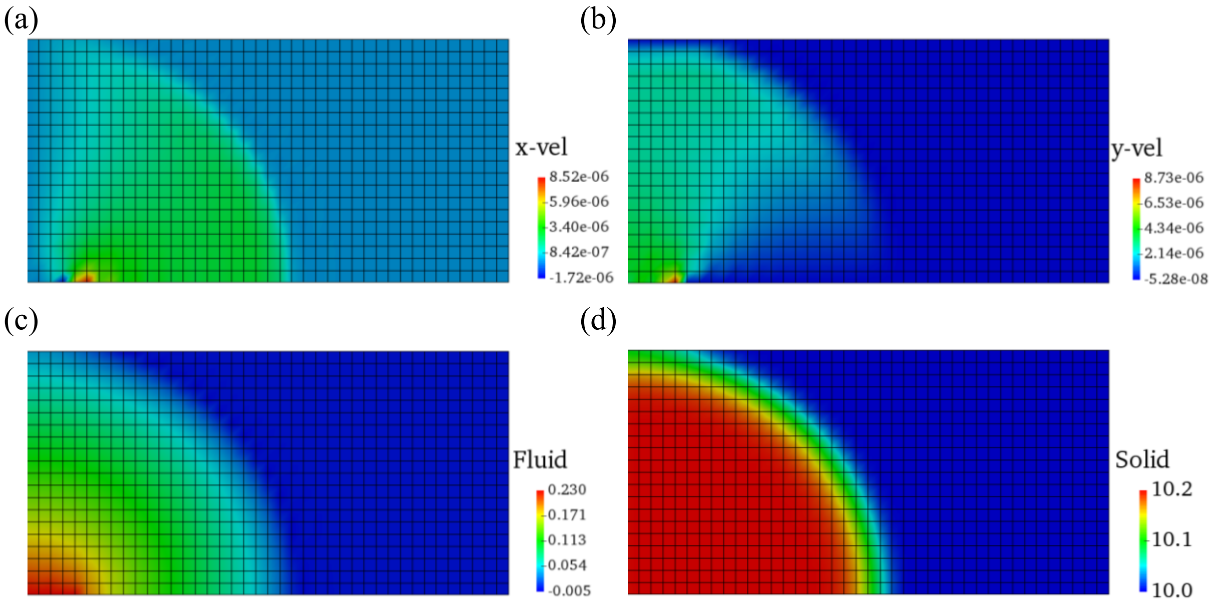


Figure 2.14. Solution fields at t = 300s: (a) x-component of fluid velocity, (b) y-component of fluid velocity, (c) fluid mass concentration, and (d) solid mass concentration.

2.7.5 Chemo-mechanical swelling in 2D

This test case investigates the impact of the chemo-mechanical coupling term in the constitutive equation for the Cauchy stress of the solid that leads to expansion of the solid as a function of the extent of reaction. The problem setup is similar to test case in section 2.7.4, however the displacement field of the solid is no longer held fixed and it evolves while accounting for the boundary conditions shown in Figure 2.15 in the reference configuration. The material parameters are given in Table 2.4, with the following specific changes: the chemo-mechanical coupling parameter is set equal to 1.0×10^{16} Pa-kg^3/m^9, the value of the solid diffusivity is assumed to depend linearly on the extent of the reaction, while the value for the fully reacted material is ten times higher than that of the unreacted material. In this test case, the stabilized method employed is SUPG for stability of the conservation of mass of the fluid equation.

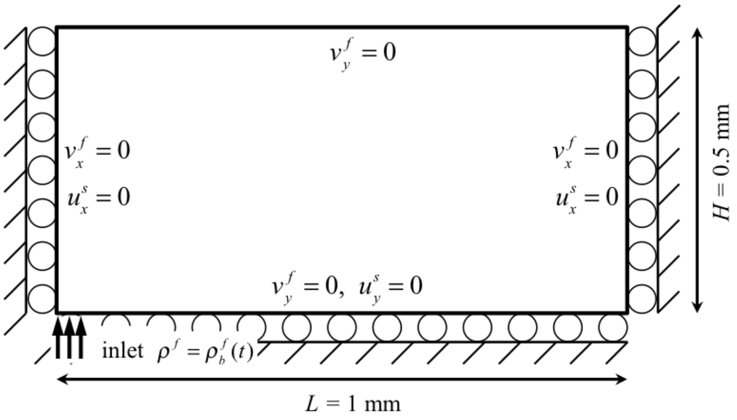


Figure 2.15. Schematic diagram of the chemo-mechanical swelling problem.

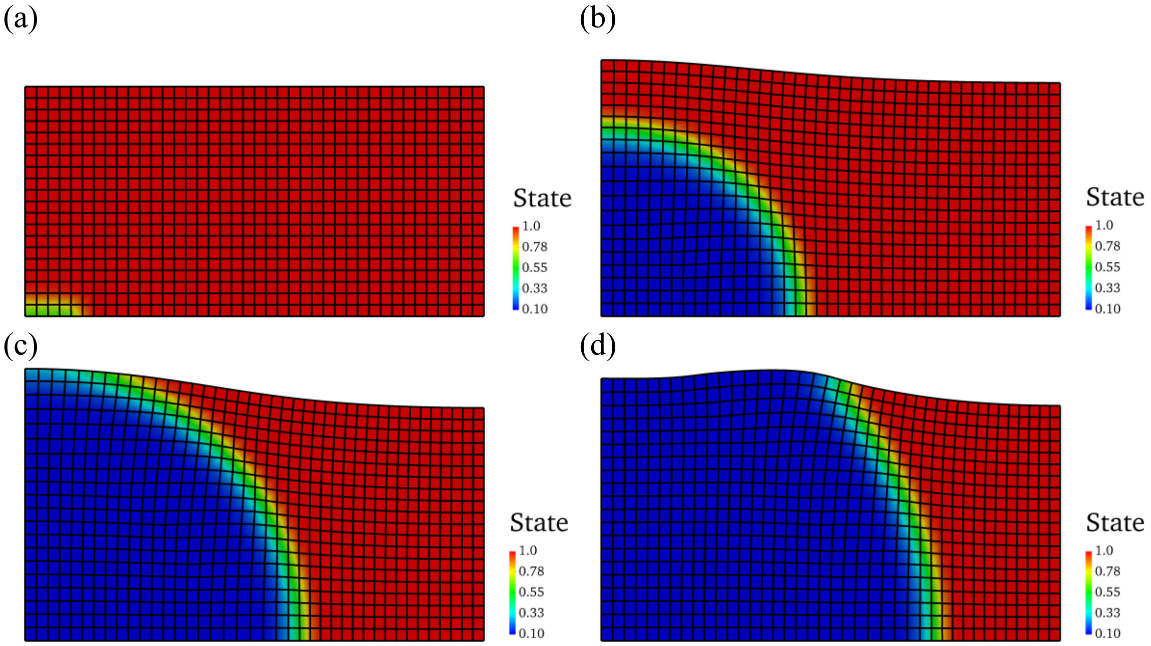


Figure 2.16. Reaction state in deformed configuration for the chemo-mechanical swelling problem at (a) 10s, (b) 300s, (c) 600s, and (d) 900s.

The snapshots in Figure 2.16 highlight the swelling of the material as the two constituents react. Particularly, since the top boundary deforms freely, the region directly above the inlet bulges out more as the reaction progresses, while the region away from the reaction zone is being slightly compressed by the expanding material. Spatial profiles of the top surface in the deformed configuration at different time levels are shown in Figure 2.17.

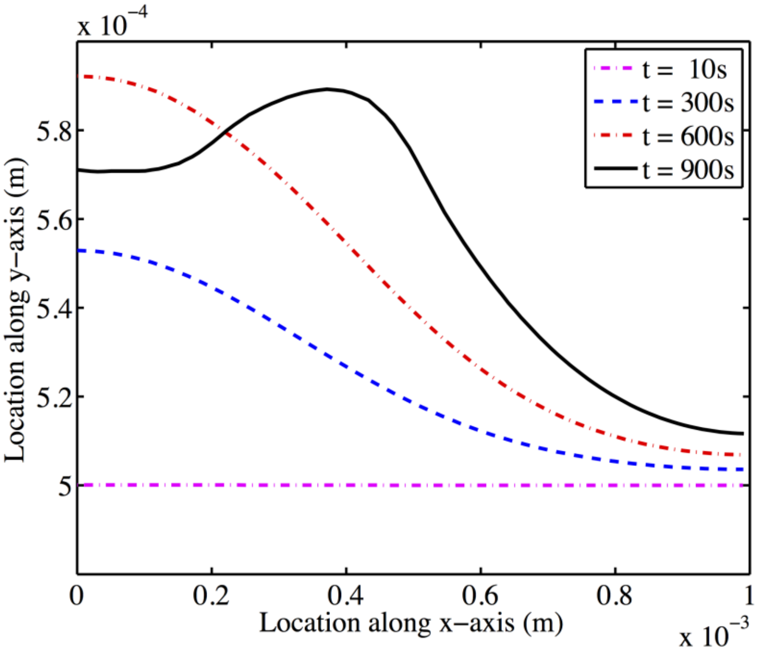


Figure 2.17. Spatial profiles of the top surface of the deformed configuration at different time levels for the chemo-mechanical swelling test.

The spatial distribution of stress components T_{11} and T_{22} are shown in Figure 2.18. These plots also show steep gradient in the stress profile across the reaction front, which can clearly be seen in the line plot of T_{11} along the bottom surface of the domain, shown in Figure 2.19. Here, this steep gradient is a product of the differential deformation of the solid: the expanding reacted solid stretches the unreacted material on the other side of the reaction front, and by the local self-equilibrium, the unreacted material constrains the free expansion of the reacted material. As a result, the unreacted material that lies on one side of the reaction front is under tension, while the reacted material across the front undergoes compression.

We also show numerical results for the case in which the Lagrange multiplier is approximated by a constant value of 0.5, which is the value of the Lagrange multiplier when and where chemical reaction is not taking place (i.e. $m^f = m^s = 0, \quad \dot{\mathbf{\Gamma}} = \mathbf{0}$), and it is held constant at all spatial points and for all times. Snapshots of the distribution of the reaction state variable at different points in time are shown in Figure 2.20.

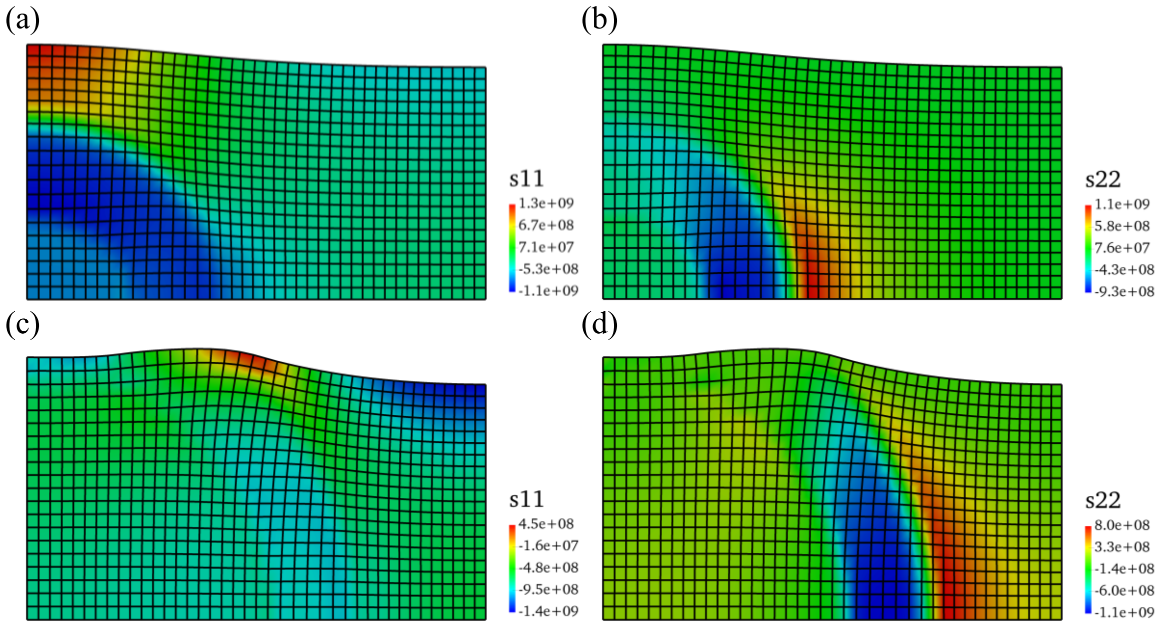


Figure 2.18. Stress component fields for the chemo-mechanical swelling problem: (a) T_{11} component at t = 300s, (b) T_{22} component at t = 300s, (c) T_{11} component at t = 900s, and (d) T_{22} component at t = 900s.

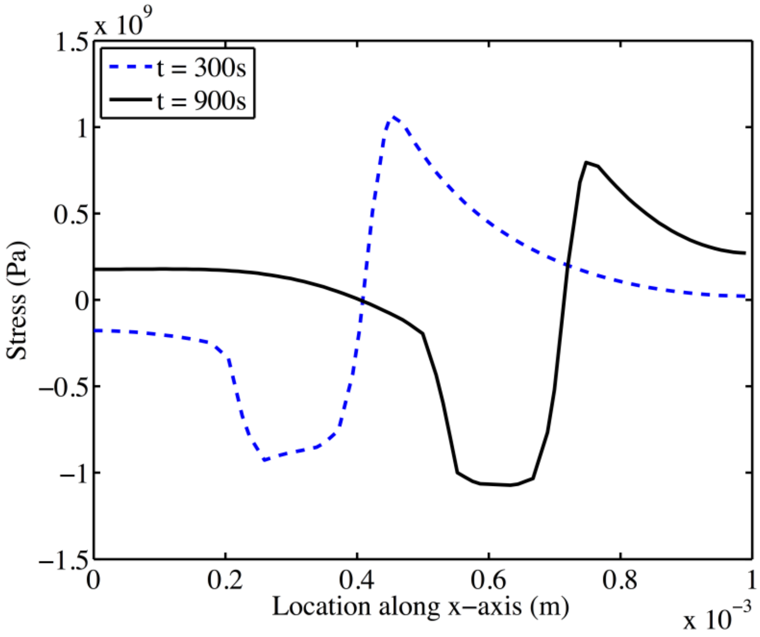


Figure 2.19. Variation in T_{22} component of the Cauchy stress tensor for the solid constituent along the x-axis (y = 0) at t = 300s and t = 900s for the chemo-mechanical swelling problem.

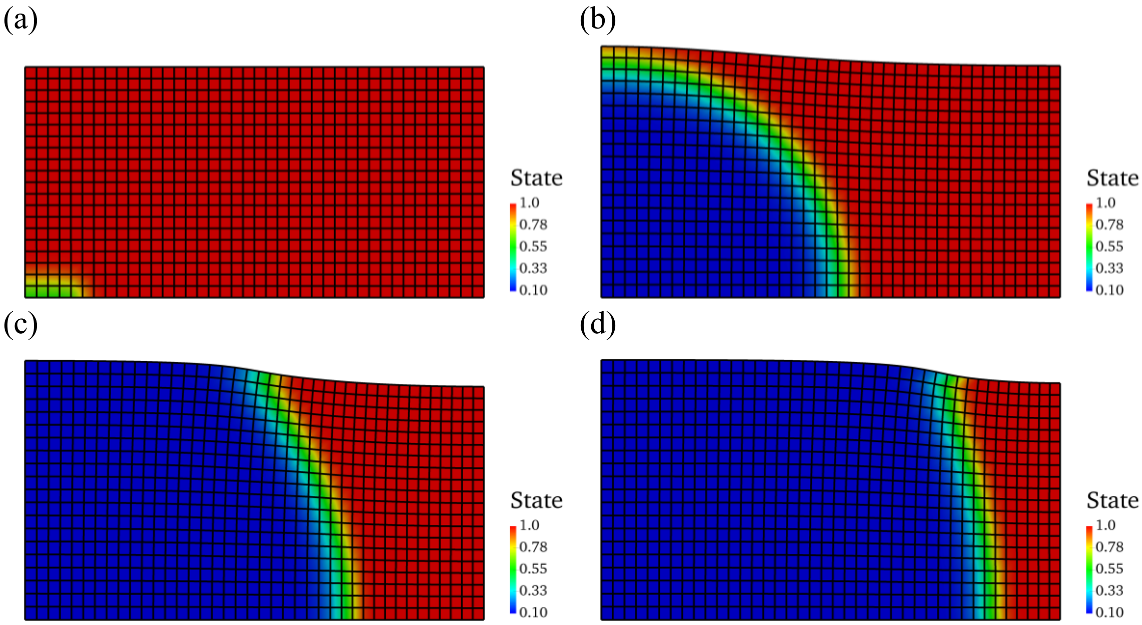


Figure 2.20. Reaction state in the deformed configuration for the chemo-mechanical swelling problem with constant Lagrange multiplier for time levels (a) 10s, (b) 300s, (c) 600s, and (d) 900s.

Figure 2.21 shows a comparison of the evolution of the vertical displacement at the top left corner of the domain for the cases in which the Lagrange multiplier is approximated by a constant value of 0.5, and when it is given by equation (2.59), derived from [28] and the simplifying assumptions in Section 2.3 and Section 2.4. The main difference is that the swelling deformation is monotonic for the constant Lagrange multiplier approximation.

$$\mu^L = \frac{1}{2} + \frac{\frac{1}{4}\dot{\gamma}\gamma^2\bar{c}^0\mathrm{tr}(\mathbf{B}^s) - m^f(\mathbf{v}^f - \mathbf{v}^s)\cdot(\mathbf{v}^f - \mathbf{v}^s)}{\eta\mathbf{D}^f\cdot\mathbf{D}^f + (\mathbf{v}^f - \mathbf{v}^s)\cdot\mathbf{A}^v(\mathbf{v}^f - \mathbf{v}^s) + \frac{1}{2}\dot{\gamma}\gamma^2\bar{c}^0\mathrm{tr}(\mathbf{B}^s)} \tag{2.59}$$

The spatial profiles of the top surface of the deformed configuration for this case are shown in Figure 2.22. Plots of stress distribution are shown in Figure 2.23, which also exhibit rapid variation in the value of stress across the reaction front. Figure 2.24 shows line plots for the T_{22} component of the Cauchy stress tensor along the x-axis of the domain at time levels t = 300s and t = 900s, where a variation of ~0.8GPa in the value of the stress across the reaction front can clearly be seen. This variation in T_{22} for the case of constant Lagrange multiplier is less than for the case when the Lagrange multiplier is allowed to vary. Comparing Figure 2.19 and Figure 2.24, the stress profiles are smoother in the latter.

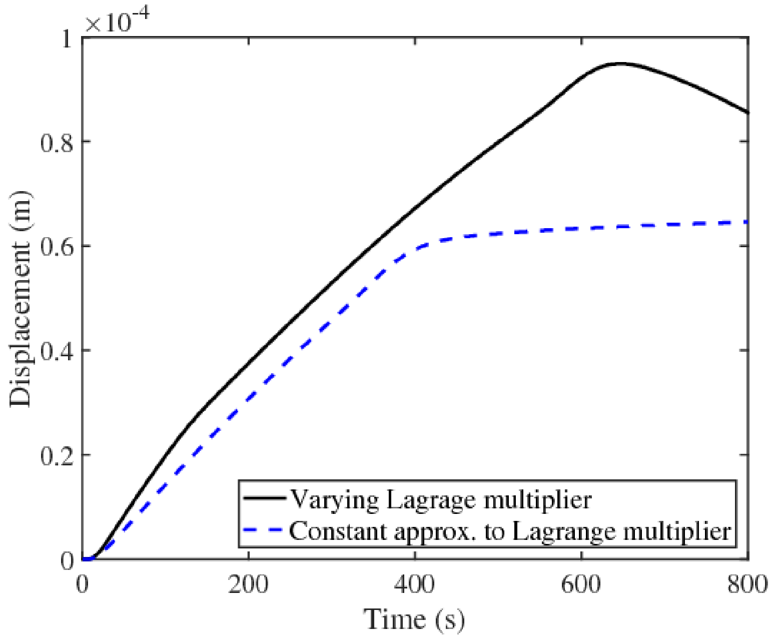


Figure 2.21. Vertical displacement as a function of time at the top left node for the chemo-mechanical swelling problem with constant Lagrange multiplier.

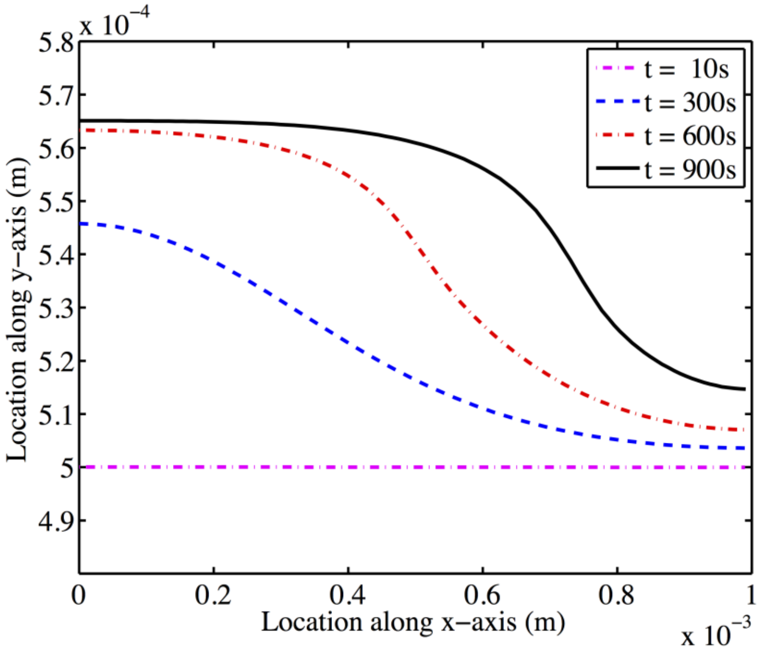


Figure 2.22. Profiles of the top surface of the domain at different time levels for chemo-mechanical swelling with constant Lagrange multiplier in the constitutive theory.

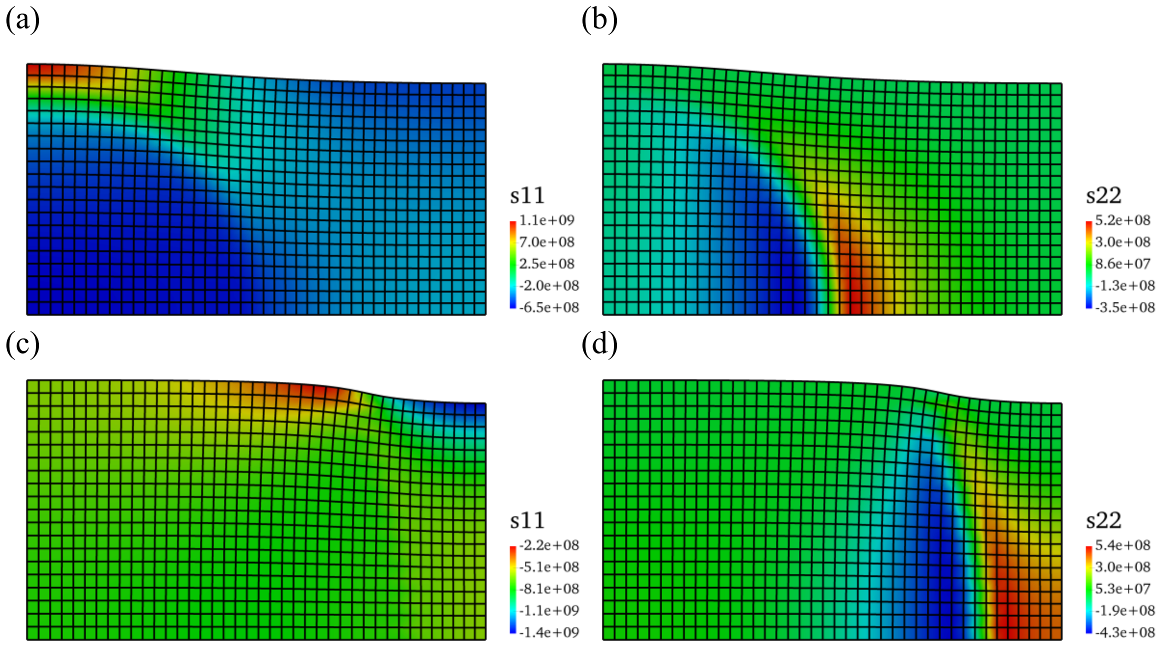


Figure 2.23. Spatial profiles of axial stress components for the chemo-mechanical swelling problem with constant Lagrange multiplier: (a) T_{11} component at t = 300s, (b) T_{22} component at t = 300s, (c) T_{11} component at t = 900s, and (d) T_{22} component at t = 900s.

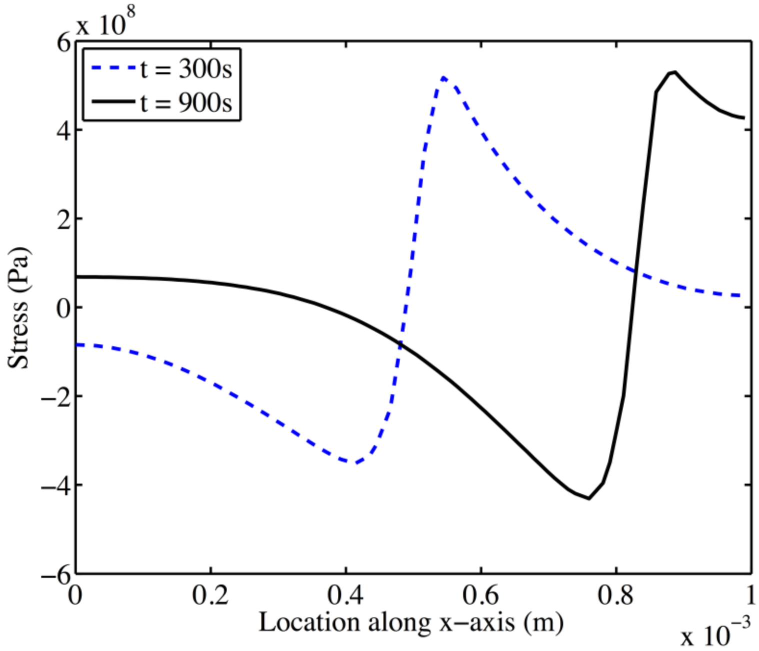


Figure 2.24. Variation in the T_{22} component of the Cauchy stress tensor for the solid constituent along the x-axis (at y = 0) at t = 300s and t = 900s for the chemo-mechanical swelling with constant Lagrange Multiplier.

To summarize, in this subsection we modeled two cases: (i) Employed the more complex form of the Lagrange multiplier in the constitutive theory, μ^L, as it appears in [28], and (ii) used a constant value of $\bar{\mu}^L = 0.5$ for the entropy Lagrange multiplier (note that the full expression is such that $\mu^L \geq 0.5$). Comparing Figure 2.16d and Figure 2.20d, it is noted that the front advanced further in the same simulated time (900s) for the case with constant Lagrange multiplier than the case where its full expression is used. The main reasons for this behavior is that the term associated with drag on the fluid, in the fluid interactive force given in equation (2.30), has μ^L as a factor. Therefore, the drag on the fluid is greater, on average, when the full expression of μ^L is used, and the fluid takes longer to reach the reaction front. From the numerical point of view, the computational cost of both cases is comparable up to the time level of 800s of the simulated time; however, the complexity of the Lagrange multiplier expression increases the difficulty in finding the consistent tangent for the numerical solution algorithm. From a physical standpoint, the authors are not aware of experimental evidence that rules out the swelling behavior obtained with the full expression for Lagrange multiplier or the swelling behavior corresponding to a constant value of the Lagrange multiplier. Thus, we have opted to use a constant value of the Lagrange multiplier for the simulations of thermal oxidation of silicon (Si) and silicon carbide (SiC) presented in this chapter, since the swelling behavior is monotonic, and it produces smoother stress profiles, as noted by comparing Figure 2.19 and Figure 2.24.

2.7.6 Bird's Beak Problem for Silicon (Si)

This test case, termed as the "bird's beak" problem, consists of thermal oxidation of a Si substrate that has been partially masked to control the location of oxide growth. The mask is made of silicon nitride (Si_3N_4) where we assumed that oxygen does not diffuse through or react with the mask. This technique is typically employed in the electronics industry to manufacture certain components such as transistors [7,38]. The material parameters for the solid materials were determined based on the description provided in [35]. The domain, shown in Figure 2.25, consists of a 2D silicon block 2 μm wide and 1 μm high. The nitride mask is a 1 μm wide by 0.2 μm high block that is placed on top of the right half of the silicon block. In [35,38], an initial oxide layer of 50 Å was considered to exist on top of the silicon block (and in between the silicon and the nitride). In this study, we replicate the geometry of the simulation in [35,38], but consider no initial oxide (the initial oxide region in [35,38] is replaced with Si) so that the full thermal oxidation process can be simulated and we are able to track the evolution of internal stress at the interface between oxide and silicon. The displacement of the solid is constrained in the normal direction at the left, right, and bottom surfaces of the domain. The boundary conditions for the fluid (O_2) mass concentration on exposed (unmasked) silicon surface are the same as for the Si wafer problem in section 2.7.2, and there is no outflow or inflow of oxygen anywhere else at the domain boundary.

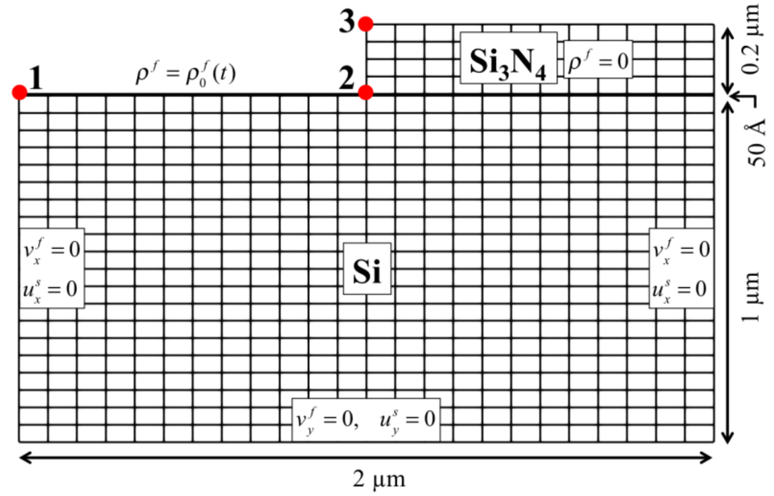


Figure 2.25. Discretization of the computational domain and location of points tracked for comparison.

Figure 2.26 shows snapshots of the location of the interface between reacted (SiO_2) material and unreacted (Si) material in the reference configuration for time levels t = 1130s, t = 1700s, t = 2200s, and t = 2600s. These time levels correspond to points in the evolution process where the interface location at the left end of the domain matches the location of the interface presented in Figure 6.15 in [38]. For this visualization, the location of the interface is assumed to lie within the reaction zone where 50% of expansion has taken place. This corresponds to a value of the reaction state variable of $\phi = 0.2$, where the blue color represents the reacted region with $\phi < 0.2$ and the red color is the region where $\phi > 0.2$.

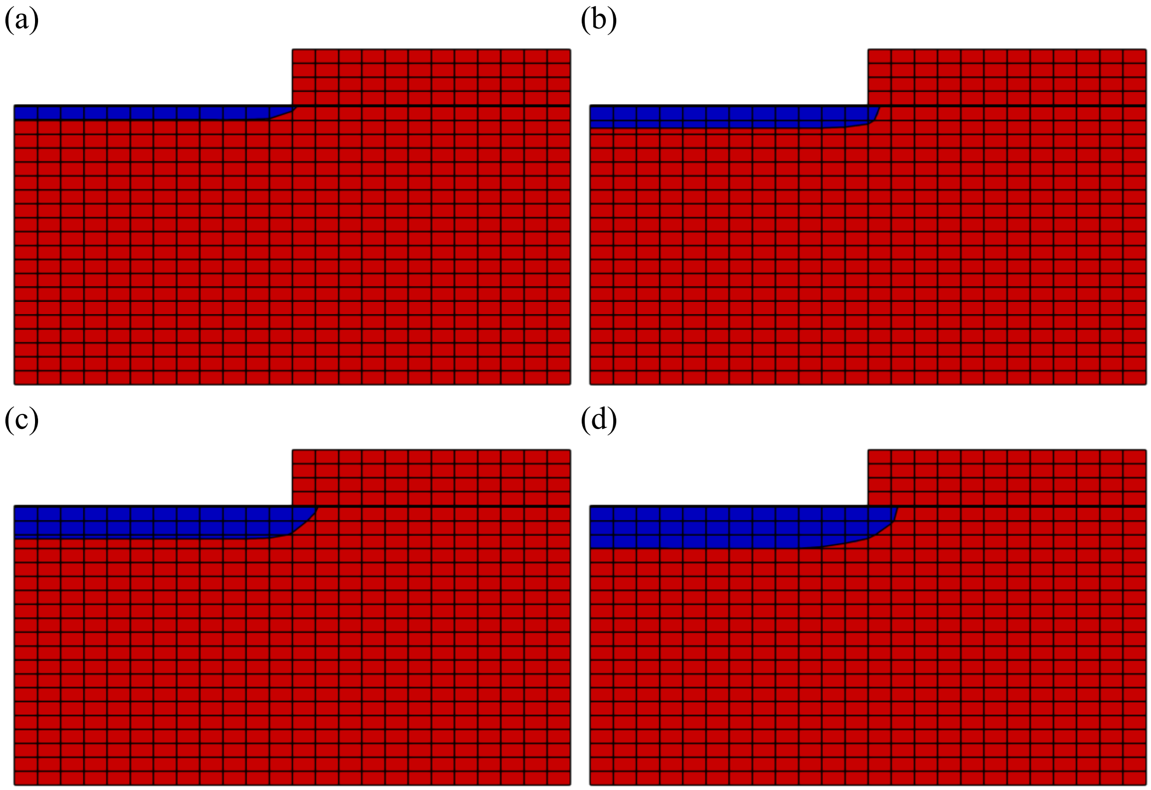


Figure 2.26. Snapshots of the interface location in the reference configuration for time levels: (a) t = 1130s, (b) t = 1700s, (c) t = 2200s and (d) t = 2600s

Figure 2.27 shows snapshots of the deformed configuration of the domain at time levels t = 1130s, 1700s, 2200s, and 2600s. These are comparable to the deformation states shown in [35,38]. As the reaction front advances and reaches the region underneath the nitride mask, the Si reacts into SiO_2 and expands, deforming locally but also exerting stress on the mask, causing it to deform as well. In addition, we observe that the reacted oxide material undergoes large compressive stresses compared to the unreacted material. This is because of the self-equilibrated state between reacted and unreacted materials where the reacted material swells as it transforms, while it is constrained by contiguous unreacted material in the x-direction.

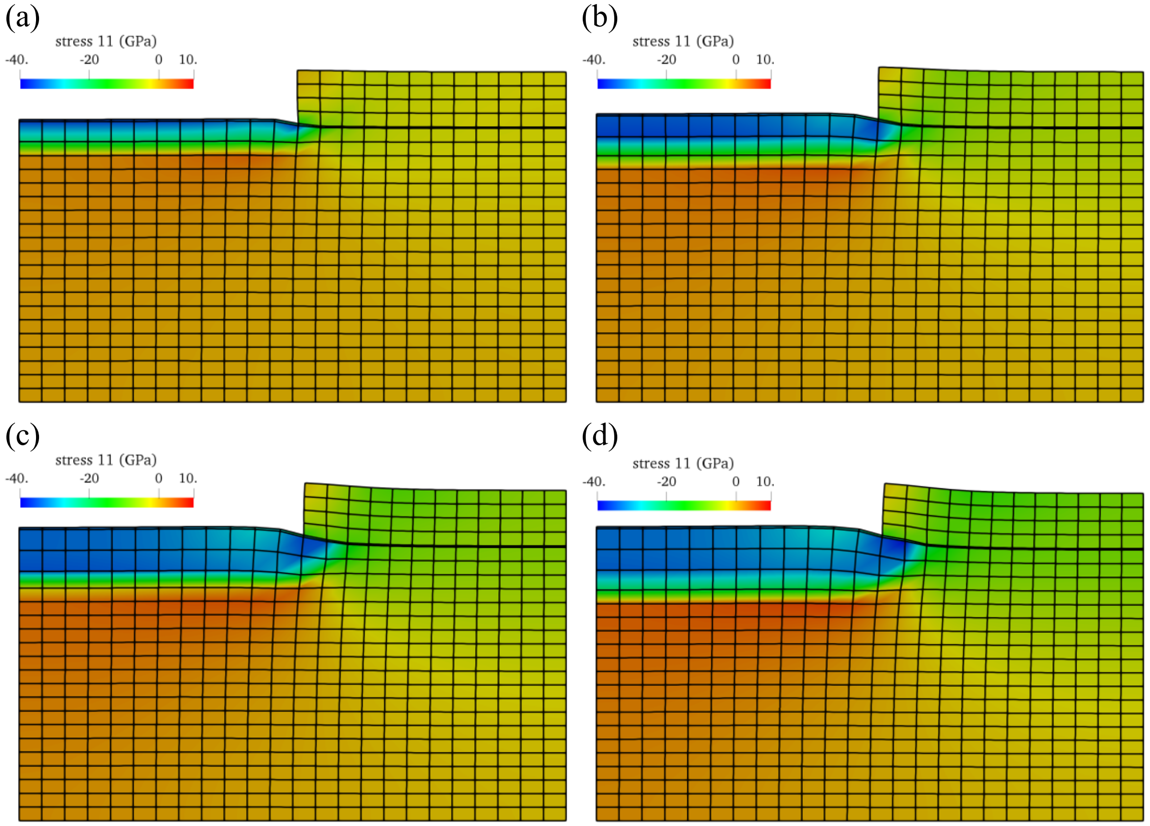


Figure 2.27. Snapshots of the spatial distribution of nodally-projected T_{11} component of stress in the deformed configuration for time levels (a) t = 1130s, (b) t = 1700s, (c) t = 2200s and (d) t = 2600s

Figure 2.28 shows the comparison of x- and y-components of the displacement field at three points labeled in Figure 2.25 with the digitized data extracted from the images in the figures presented in [35,38]. The numerical values and general trends are in reasonable agreement with comparable extent of the oxidation process for the selected time levels, as shown in the snapshots of Figure 2.27 and those in [35,38]. We remark that the time levels corresponding to a given oxidation state in the snapshots may not be in agreement with the work presented in [35,38] due to the difference in the assumptions on the oxidizing environment, as discussed earlier in the Si wafer test case presented in section 2.7.2.

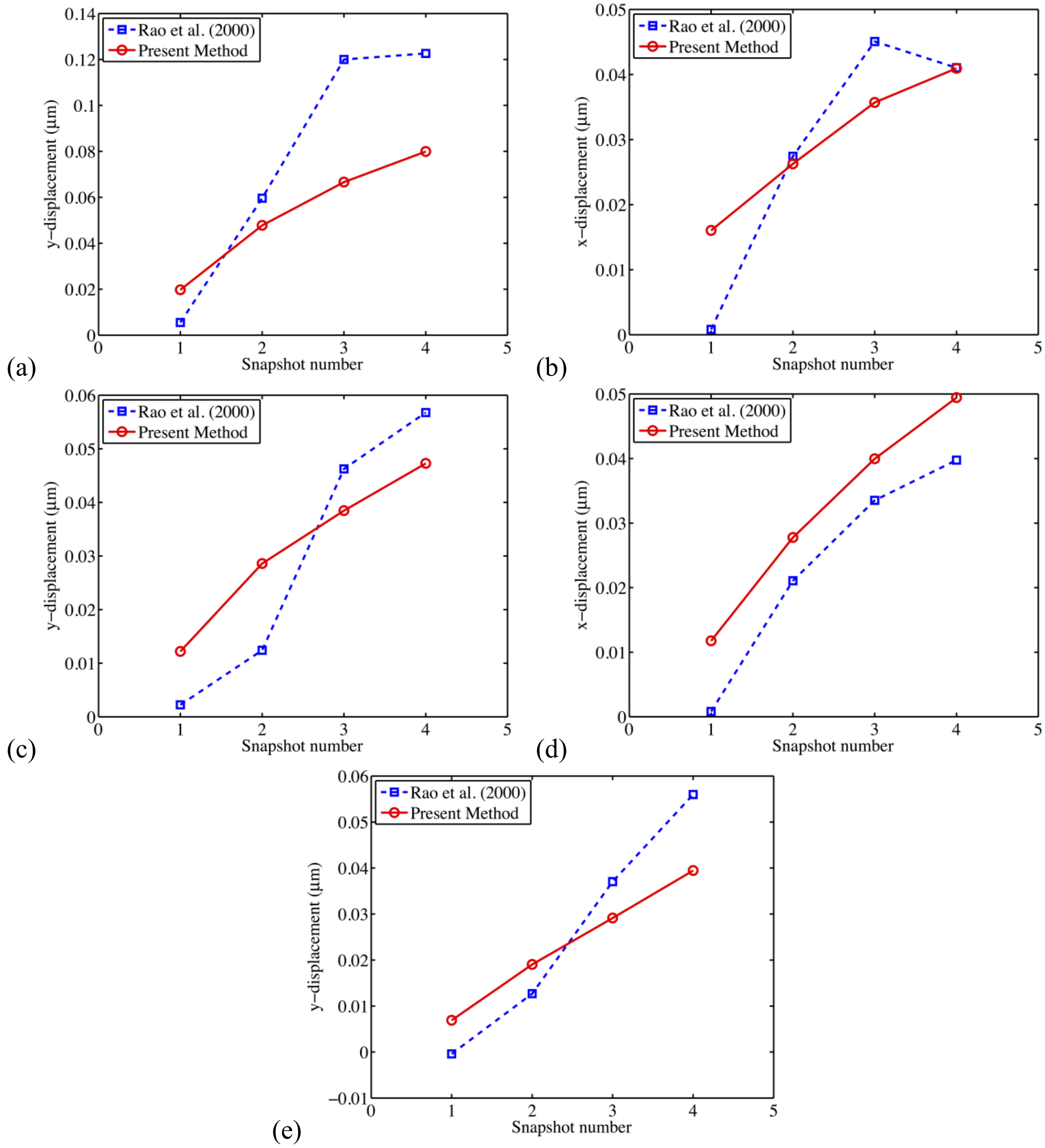


Figure 2.28. Comparison of the displacement field corresponding to each snapshot for: (a) y-displacement at point 1, (b) x-displacement at point 2, (c) y-displacement at point 2, (d) x-displacement at point 3, and (e) y-displacement at point 3.

Figure 2.29 shows profiles of the nodally-projected T_{11} component of the Cauchy stress tensor of the solid constituent along the y-axis (x = 0, along the left edge of the domain) in the current configuration. As in previous cases, it is observed that the reacted material (oxide) undergoes compression due to constrained expansion, and the unreacted material across the advancing reaction front undergoes tension as it gets stretched by the swelling of the oxide.

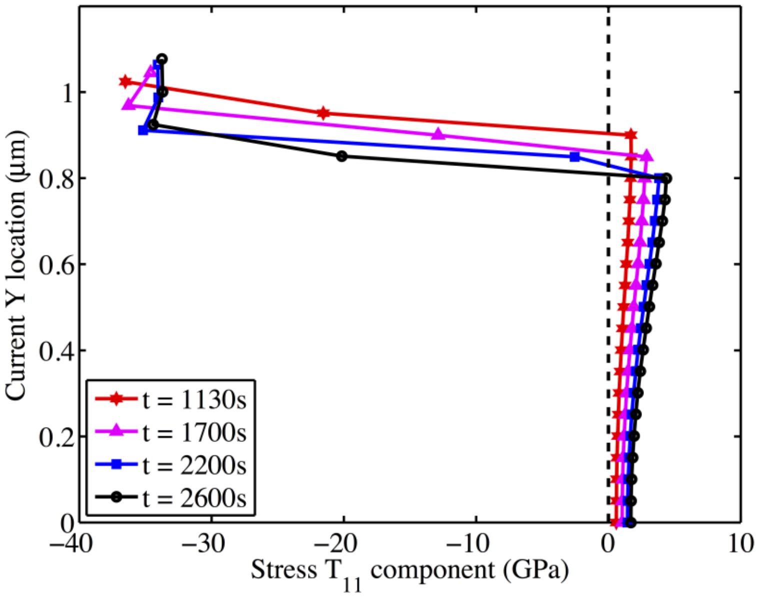


Figure 2.29. Profiles of the spatial distribution along the y-axis of the nodally-projected T_{11} component of stress at different time levels for the masked oxidation test case.

2.8 Conclusions

In this chapter we have presented an unconstrained mixture model for multi-constituent solid-fluid system that undergoes chemical reactions and associated mechanical deformations. The proposed model is well suited to material systems with multiple interacting constituents, wherein directly modeling all the inter-constituent interfaces at the micro scales becomes prohibitively expensive. It is shown that for such intricately complex systems, mixture theory provides a reduced order model by introducing interacting terms that capture inter-constituent interactions in the balance laws. The mixture theory model presented here is capable of capturing chemo-mechanical effects in solid-fluid mixtures, especially where chemically reacting fluid causes large expansions in the solid; a case that typically arises in problems like the oxidation of SiC materials in aerospace applications and Si in electronics manufacturing.

The mathematical description of the physical system via mixture representation involves interaction terms (such as inter-constituent mass transfer and momentum transfer) that contain zeroth order (reaction-like) terms. Depending on the material properties of the physical system, these zeroth-order terms can give rise to a system of singularly perturbed coupled PDEs, namely singular perturbations can manifest themselves as boundary or internal layers that could either be stationary or they could be traversing the domain. When the length-scales of these layers are smaller than the characteristic length-scale of the FEM discretization, instabilities are triggered that can result in the breakdown of the method. To mitigate the spurious oscillations, a VMS stabilized FEM was derived for fluid mass balance and momentum balance that has enhanced stability properties for multiple coupled nonlinear fields. We also employed SUPG based stabilization for the fluid density in some of the test cases.

The numerical results show that the mixture model is able to capture the chemical and mechanical effects, as well as their coupled response. Since the constituents are not modeled discretely, meshing is simple for complex microstructures when represented through the mathematical constructs in the mixture theory. From an algorithmic viewpoint all implementational issues for the proposed VMS method for the mixture model are addressed at the element level and consequently the method is easy to implement in a FE framework. Enhanced stability and accuracy permits using crude meshes that results in substantial computational economy as compared to the cost incurred if the constituents were discretely modeled. Numerical results are shown for a variety of chemo-mechanical problems in the finite deformation range. In all the cases a stable and accurate coupled response is achieved which is a numerical manifestation of the variational consistency and enhanced stability of the proposed method.

2.9 References

[1] T.J. Truster, A. Masud, A unified mixture formulation for density and volumetric growth of multi-constituent solids in tissue engineering, Comput. Methods Appl. Mech. Eng. 314 (2017) 222–268.

[2] J.D. Humphrey, K.R. Rajagopal, A constrained mixture model for growth and remodeling of soft tissues, Math. Model. Methods Appl. Sci. 12 (2002) 407–430.

[3] J.D. Humphrey, K.R. Rajagopal, A constrained mixture model for arterial adaptations to a sustained step change in blood flow, Biomech. Model. Mechanobiol. 2 (2003) 109–126.

[4] P. Yuan, J.M. McCracken, D.E. Gross, P. V Braun, J.S. Moore, R.G. Nuzzo, A programmable soft chemo-mechanical actuator exploiting a catalyzed photochemical water-oxidation reaction, Soft Matter. 13 (2017) 7312–7317.

[5] B.A. Haberman, J.B. Young, Three-dimensional simulation of chemically reacting gas flows in the porous support structure of an integrated-planar solid oxide fuel cell, Int. J. Heat Mass Transf. 47 (2004) 3617–3629.

[6] S.R. Bishop, Chemical expansion of solid oxide fuel cell materials: A brief overview, Acta Mech. Sin. 29 (2013) 312–317.

[7] V.S. Rao, T.J.R. Hughes, On modelling thermal oxidation of Silicon I: Theory, Int. J. Numer. Methods Eng. 47 (2000) 341–358.

[8] F. Lofaj, Y.S. Kaganovskii, Kinetics of WC-Co oxidation accompanied by swelling, J. Mater. Sci. 30 (1995) 1811–1817.

[9] T. Merzouki, E. Blond, N. Schmitt, M.-L. Bouchetou, T. Cutard, A. Gasser, Modelling of the swelling induced by oxidation in SiC-based refractory castables, Mech. Mater. 68 (2014) 253–266.

[10] T. Merzouki, E. Blond, N. Schmitt, Numerical study of the effects of refractory lining geometries on the swelling induced by oxidation, Finite Elem. Anal. Des. 108 (2016) 66–80.

[11] R.G. Munro, S.J. Dapkunas, Corrosion characteristics of silicon carbide and silicon nitride, J. Res. - Natl. Inst. Stand. Technol. 98 (1993) 607–631.

[12] H. Yang, F. Fan, W. Liang, X. Guo, T. Zhu, S. Zhang, A chemo-mechanical model of lithiation in silicon, J. Mech. Phys. Solids. 70 (2014) 349–361.

[13] R. Hall, H. Gajendran, A. Masud, Diffusion of chemically reacting fluids through nonlinear elastic solids: mixture model and stabilized methods, Math. Mech. Solids. 20 (2015) 204–227.

[14] T.J.R. Hughes, Multiscale phenomena: Green's functions, the Dirichlet-to-Neumann formulation, subgrid scale models, bubbles and the origins of stabilized methods, Comput. Methods Appl. Mech. Eng. 127 (1995) 387–401.

[15] T.J.R. Hughes, G.R. Feijóo, L. Mazzei, J.-B. Quincy, The variational multiscale method—a paradigm for computational mechanics, Comput. Methods Appl. Mech. Eng. 166 (1998) 3–24.

[16] F. Brezzi, L.P. Franca, T.J.R. Hughes, A. Russo, $b = \int g$, Comput. Methods Appl. Mech. Eng. 145 (1997) 329–339.

[17] C. Truesdell, Rational Thermodynamics, Second, Springer New York, New York, NY,

1984.

[18] R.M. Bowen, Theory of Mixtures, in: A.C. Eringen (Ed.), Contin. Phys., Volume III, Elsevier, 1976: pp. 1–127.

[19] R.J. Atkin, R.E. Craine, Continuum theories of mixtures: Basic theory and historical development, Q. J. Mech. Appl. Math. 29 (1976) 209–244.

[20] K.R. Rajagopal, L. Tao, Mechanics of Mixtures, World Scientific, 1995.

[21] I. Samohýl, Thermodynamics of irreversible processes in fluid mixtures : approached by rational thermodynamics, 1. Aufl., B.G. Teubner, Leipzig, 1987.

[22] L.P. Franca, E.G. Dutra Do Carmo, The Galerkin gradient least-squares method, Comput. Methods Appl. Mech. Eng. 74 (1989) 41–54.

[23] R. Codina, Comparison of some finite element methods for solving the diffusion-convection-reaction equation, Comput. Methods Appl. Mech. Eng. 156 (1998) 185–210.

[24] G. Hauke, A. García-Olivares, Variational subgrid scale formulations for the advection–diffusion-reaction equation, Comput. Methods Appl. Mech. Eng. 190 (2001) 6847–6865.

[25] H. Fernando, C. Harder, D. Paredes, F. Valentin, Numerical multiscale methods for a reaction-dominated model, Comput. Methods Appl. Mech. Eng. 201–204 (2012) 228–244.

[26] E.G. Dutra do Carmo, G.B. Alvarez, F.A. Rochinha, A.F.D. Loula, Galerkin projected residual method applied to diffusion–reaction problems, Comput. Methods Appl. Mech. Eng. 197 (2008) 4559–4570.

[27] A. Masud, Preface, Comput. Methods Appl. Mech. Eng. 193 (2004) iii–iv.

[28] R. Hall, K. Rajagopal, Diffusion of a fluid through an anisotropically chemically reacting thermoelastic body within the context of mixture theory, Math. Mech. Solids. 17 (2012) 131–164.

[29] A. Masud, R.A. Khurram, A multiscale finite element method for the incompressible Navier–Stokes equations, Comput. Methods Appl. Mech. Eng. 195 (2006) 1750–1777.

[30] A. Masud, R. Calderer, A variational multiscale method for incompressible turbulent flows: Bubble functions and fine scale fields, Comput. Methods Appl. Mech. Eng. 200 (2011) 2577–2593.

[31] R.S. Hay, Growth stress in SiO 2 during oxidation of SiC fibers, J. Appl. Phys. 111 (2012) 063527.

[32] L.L. Snead, T. Nozawa, Y. Katoh, T.S. Byun, S. Kondo, D.A. Petti, Handbook of SiC

properties for fuel performance modeling, J. Nucl. Mater. 371 (2007) 329–377.

[33] K. Kajihara, H. Kamioka, M. Hirano, T. Miura, L. Skuja, H. Hosono, Interstitial oxygen molecules in amorphous SiO 2. III. Measurements of dissolution kinetics, diffusion coefficient, and solubility by infrared photoluminescence, J. Appl. Phys. 98 (2005) 013529.

[34] Y. Song, S. Dhar, L.C. Feldman, G. Chung, J.R. Williams, Modified Deal Grove model for the thermal oxidation of silicon carbide, J. Appl. Phys. 95 (2004) 4953–4957.

[35] V.S. Rao, T.J.R. Hughes, K. Garikipati, On Modelling Thermal Oxidation of Silicon II: Numerical Aspects, Int. J. Numer. Methods Eng. 47 (2000) 359–377.

[36] A.N. Brooks, T.J.R. Hughes, Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations, Comput. Methods Appl. Mech. Eng. 32 (1982) 199–259.

[37] T.J.R. Hughes, M. Mallet, M. Akira, A new finite element formulation for computational fluid dynamics: II. Beyond SUPG, Comput. Methods Appl. Mech. Eng. 54 (1986) 341–355.

[38] V.S. Rao, On numerical modeling of thermal oxidation in silicon, Stanford University, 1997.

CHAPTER 3:
REDUCED MIXTURE MODEL AND ELASTIC RESPONSE OF CHEMICALLY SWOLLEN SOLIDS: APPLICATION TO SI OXIDATION AND LITHIATION[1]

3.1 Introduction

Many problems of scientific and engineering interest involve the diffusion of a fluid-like species into a reactive solid, producing mechanical deformation. Examples include oxidation [1,2], lithiation [3], hydration of gels [4], and tissue growth [5]. Another relevant area is in the field of micro-self-actuators [6], where a functional mechanical response is triggered by other external stimuli (thermal, chemical, or electromagnetic). The development of models along with algorithms for their numerical implementation is crucial to gain important insight into the complex physical processes and to develop tools for engineering applications. Some significant work done in these areas includes [7–11].

Mixture models for diffusion through polymeric solids undergoing large deformations [12,13] and for the class of materials that possess multiple natural configurations [14], have been proposed in the literature. Employing ideas from these works, advanced numerical methods for diffusion of chemically reacting fluids through nonlinear elastic solids were developed [11,15,16]. In [11,15], the authors adopted the constitutive theory developed in [13] wherein the solid constituent undergoes chemical evolution that also induces mechanical deformation. This can also be viewed in the context of materials with changing reference unstressed configuration relative to which the solid constituent undergoes elastic deformations. The extreme cases correspond to the solid in the pristine state, and the solid in its fully reacted, chemically swollen, unconstrained state. Following the seminal work of Truesdell and Noll [17], we investigate elastic deformation superimposed on the (finitely strained) chemically swollen solid and the associated stress response dictated by the constitutive equations reported in [11]. This analysis yields a relationship between the material parameters required by the model in that work and classical elastic constants (e.g. first Lamé parameter and shear modulus) of both the pristine and fully reacted solid materials. It is highlighted that determination of material parameters in terms of elastic constants typically reported in the literature (or easily determined through standard experimental procedures) is crucial to produce realistic, reliable, and accurate results. It is noted that in other models [18,19] the effect of superimposed elastic deformations is accounted for through a compositional deformation map. This results in a multiplicative split of the deformation gradient into an elastic contribution and an inelastic (chemical) contribution.

---

[1] This Chapter has been adapted from “M. Anguiano, A. Masud, Reduced mixture model and elastic response of chemically swollen solids: Application to Si oxidation and lithiation, Appl. Eng. Sci. 6 (2021) 100039.” The copyright owner has provided written permission to reprint the work.

In addition, many of the problems of interest are characterized by diffusive transport of the fluid species, as is the case when the fluid species can be modeled as “gas-like”. Examples include transport of interstitial oxygen molecules in oxidation processes [20], or the transport of lithium ions during the lithiation process [10]. In these situations, certain reasonable assumptions can be made to simplify the governing equations for the fluid flow without significant loss in the modeled physics. These assumptions reduce the complexity and the number of equations to be solved. Under appropriate assumptions, the balance of linear momentum for the fluid constituent becomes an algebraic equation in the fluid velocity field, which can be solved analytically in terms of the other fields. One important consequence is that the numerical implementation of the reduced model in a finite element method does not require a finite element discretization of the fluid velocity field. In 3D, this constitutes three fewer degrees of freedom per node which translates into computationally efficient algorithms while retaining engineering solutions that are comparable to what is obtained with the full fluid model.

The chapter is organized as follows: section 3.2 presents the governing equations. Section 3.3 presents a further analysis of the solid Cauchy stress terms and determination of the material parameters. Section 3.4 contains simplifying assumptions for the fluid model. Section 3.5 summarizes the reduced formulation for the governing system of equations. Numerical case studies for thermal oxidation and lithiation of silicon are included in section 3.6. Finally, discussion and concluding remarks are given in section 3.7.

3.2 Governing Equations

In this section, a summary of the formulation developed and utilized in [11] is presented. In the context of mixture theory, we consider two constituents: one solid (s) and one fluid (f). At time t, both constituents co-occupy the closed domain $\Omega_t \subset \mathbb{R}^3$, whose boundary is denoted by $\partial\Omega_t$. Since the fluid constituent moves relative to the deforming solid domain, an arbitrary Lagrangian-Eulerian (ALE) formulation is used to expand the time derivative of the governing equations for the fluid. The fluid balance of mass, fluid balance of linear momentum, solid balance of mass, and solid balance of linear momentum (neglecting solid inertial effects) are given, in that order, by

$$\frac{\partial \rho^f}{\partial t} + \left(\mathbf{v}^f - \mathbf{v}^s\right) \cdot \nabla \rho^f + \rho^f \operatorname{div}\left(\mathbf{v}^f\right) - m^f = 0 \qquad \text{in } \Omega_t \tag{3.1}$$

$$\rho^f \frac{\partial \mathbf{v}^f}{\partial t} + \rho^f (\mathbf{v}^f - \mathbf{v}^s) \cdot \nabla \mathbf{v}^f - \operatorname{div}\left(\mathbf{T}^f\right) - \rho^f \mathbf{b}^f - \mathbf{I}^f = \mathbf{0} \qquad \text{in } \Omega_t \tag{3.2}$$

$$\frac{\partial \rho^s}{\partial t} + \rho^s \operatorname{div}\left(\mathbf{v}^s\right) - m^s = 0 \qquad \text{in } \Omega_t \tag{3.3}$$

$$-\operatorname{div}\left(\mathbf{T}^s\right) - \rho^s \mathbf{b}^s - \mathbf{I}^s = \mathbf{0} \qquad \text{in } \Omega_t \tag{3.4}$$

where, for constituent $\alpha \in \{f, s\}$, ρ^α is its mass concentration (per unit volume in the deformed configuration), $\mathbf{v}^\alpha$ is its velocity, $\mathbf{T}^\alpha$ is its Cauchy stress tensor, and $\mathbf{b}^\alpha$ is its body force per unit mass. Likewise, m^α is the mass transfer rate (per unit volume in the deformed configuration) into constituent α from all other constituents, and $\mathbf{I}^\alpha$ is the interactive force or momentum transfer into constituent α from all other constituents. Note that ∇ and div denote the gradient and divergence operators with respect to the spatial coordinates $\mathbf{x}$.

In addition, progress of chemical reaction and material evolution state of the solid is tracked through a reaction state variable $0 \le \phi \le 1$, such that the unreacted, pristine, or virgin state of the solid corresponds to $\phi = 1$, while the fully transformed solid corresponds to $\phi = 0$. The reaction state variable may be defined via an evolution equation for its total/material time derivative $D^s\phi/Dt$. In order to determine the unknown fields ρ^f, $\mathbf{v}^f$, ρ^s, and $\mathbf{v}^s$, as well as the reaction state variable ϕ, suitable constitutive equations are required for m^f, $\mathbf{I}^f$, $\mathbf{T}^f$, m^s, $\mathbf{I}^s$, $\mathbf{T}^s$, and $D^s\phi/Dt$. The constitutive equation for the solid Cauchy stress $\mathbf{T}^s$ is discussed in section 3.3. The functional form of m^f and $D^s\phi/Dt$ may be problem dependent and it is left unspecified for now (specific forms are given in each of the numerical case studies in section 3.6). Considering "gas-like" fluid behavior, the following thermodynamically consistent constitutive equations were derived for $\mathbf{I}^f$ and $\mathbf{T}^f$

$$\mathbf{T}^f = -\frac{R\theta_0}{W_f} \rho^f \mathbf{1} + \bar{\eta} \mathbf{D}^f \tag{3.5}$$

$$\mathbf{I}^f = -m^f\left(\mathbf{v}^f - \mathbf{v}^s\right) - \mathbf{A}^v\left(\mathbf{v}^f - \mathbf{v}^s\right) \tag{3.6}$$

where R is the gas constant, W_f is the molar mass of the fluid, θ_0 is the isothermal temperature of the system, $\bar{\eta}$ is an effective viscosity, $\mathbf{D}^f$ is twice the symmetric part of the fluid velocity gradient, i.e., $\mathbf{D}^f = \nabla\mathbf{v}^f + \left(\nabla\mathbf{v}^f\right)^T$, $\mathbf{A}^v$ is an effective drag tensor, and $\mathbf{1}$ denotes the second order identity tensor. Mass and momentum are required to be conserved as they are transferred between constituents, which yields the following equations that are used to determine m^s and $\mathbf{I}^s$

$$\sum_{\alpha\in\{f,s\}} m^\alpha = 0 \qquad \Rightarrow \qquad m^s = -m^f \tag{3.7}$$

$$\sum_{\alpha\in\{f,s\}} \left(\mathbf{I}^\alpha + m^\alpha\mathbf{v}^\alpha\right) = 0 \qquad \Rightarrow \qquad \mathbf{I}^s = -\mathbf{I}^f - m^f\mathbf{v}^f - m^s\mathbf{v}^s \tag{3.8}$$

In the present chapter we limit our investigation to isothermal conditions, therefore the balance of energy is satisfied automatically. The balance of angular momentum is satisfied automatically as well since the constitutive equations for $\mathbf{T}^s$ and $\mathbf{T}^f$ yield symmetric tensors.

3.3 The Solid Cauchy Stress Response Function

The constitutive equation for the solid Cauchy stress used in [11] can be written as a function of the left Cauchy-Green deformation tensor $\mathbf{B}$ (and its invariants) as

$$\mathbf{T}^s = \bar{\mathbf{T}}(\mathbf{B}) = \left[\mu^s(\phi) - C^s(\phi)\right]J^{-1}\mathbf{B} + \left[\lambda^s(\phi)(J-1) - \mu^s(\phi)J^{-1}\right]\mathbf{1} \tag{3.9}$$

where $\mathbf{1}$ denotes the rank-two identity tensor and

$$\mathbf{B} = \mathbf{F}\mathbf{F}^T \tag{3.10}$$

$$J = \sqrt{\det\mathbf{B}} \tag{3.11}$$

where $\mathbf{F} = \partial\mathbf{x}/\partial\mathbf{X}^s$ is the solid deformation gradient. In this form, $\lambda^s(\phi)$, $\mu^s(\phi)$, and $C^s(\phi)$ are taken to be material parameters evolving with chemical reaction through their dependence on the reaction state ϕ. It is noted that $C^s(\phi)$ is associated with the chemo-mechanical coupling term in the Helmholtz free energy for the solid, and it is such that it vanishes for the unreacted material, i.e., for $\phi = 1$, $C^s(1) = 0$.

3.3.1 Stress-free reacted configurations.

As expected, the Cauchy stress in (3.9) vanishes for unstrained configurations of the unreacted (or pristine) solid material, i.e., $\mathbf{T}^s = \mathbf{0}$ for $\mathbf{B} = \mathbf{1}$ and $\phi = 1$ (recall $C^s(1) = 0$). Furthermore, it is also expected that an unconstrained, homogeneously reacted solid, would also be in a stress-free state. Let $\mathbf{F} = \mathbf{F}^\phi$ and $\mathbf{B} = \mathbf{B}^\phi$ denote the deformation gradient and left Cauchy-green deformation tensors associated with a deformation (from the reference configuration of the unreacted state) that corresponds to such a stress-free, reacted state. Therefore, one obtains the equation

$$\mathbf{0} = \left[\mu^s(\phi) - C^s(\phi)\right]\left(J^\phi\right)^{-1}\mathbf{B}^\phi + \left[\lambda^s(\phi)\left(J^\phi - 1\right) - \mu^s(\phi)\left(J^\phi\right)^{-1}\right]\mathbf{1} \tag{3.12}$$

Solving for $\mathbf{B}^\phi$ yields

$$\mathbf{B}^\phi = \frac{\left[\mu^s(\phi) - \lambda^s(\phi) J^\phi\left(J^\phi - 1\right)\right]}{\left[\mu^s(\phi) - C^s(\phi)\right]}\mathbf{1} \tag{3.13}$$

Consequently, $\mathbf{B}^\phi$ is a scalar multiple of the identity tensor, which corresponds to some uniform isotropic stretch, and therefore (3.13) is satisfied by any deformation gradient of the form

$$\mathbf{F}^\phi = \gamma\mathbf{Q} \tag{3.14}$$

where $\mathbf{Q}$ is a proper orthogonal tensor ($\mathbf{Q}\mathbf{Q}^T = \mathbf{1}$, $\det\mathbf{Q} = 1$), and $\gamma > 0$. It then follows from (3.10) that

$$\mathbf{B}^\phi = \gamma^2\mathbf{1} \tag{3.15}$$

and comparing with (3.13) yields

$$\gamma^2 = \frac{\left[\mu^s(\phi) - \lambda^s(\phi) J^\phi\left(J^\phi - 1\right)\right]}{\left[\mu^s(\phi) - C^s(\phi)\right]} \tag{3.16}$$

Moreover, recalling (3.11), it holds that

$$J^\phi = \sqrt{\det\mathbf{B}^\phi} = \gamma^3 \tag{3.17}$$

and substituting (3.17) into (3.16), it is obtained that (assuming all the material properties are known) γ must be a real positive root to a degree-6 polynomial, i.e.

$$\lambda^s(\phi)\left[\gamma^6 - \gamma^3\right] + \left[\mu^s(\phi) - C^s(\phi)\right]\gamma^2 - \mu^s(\phi) = 0 \tag{3.18}$$

Once a solution for γ is known, the stress-free reacted configuration is known up to a rigid translation and rotation. In particular, the deformation gradient of the stress-free reacted configuration is also known, up to a rigid rotation.

The following inequality constraint on the material parameters can be deduced from (3.16)

$$\mu^s(\phi) > \max\left(C^s(\phi), \lambda^s(\phi) J^\phi \left[J^\phi - 1\right]\right) \tag{3.19}$$

Note that this is valid for any reaction state ϕ for which a stress-free configuration with a homogenous deformation gradient exists.

3.3.2 Elastic response of the reacted solid

It is remarked that although a reacted configuration can be stress-free, it is, in general, strained. This is because the reference configuration is a stress-free state for the unreacted (or pristine) solid. Nevertheless, it is expected that the fully reacted material also behaves as an isotropic elastic solid. According to the theorem on superimposed elastic strain, stated in the classical work by Truesdell & Noll [17], “the response of a strained material subjected to an arbitrary further infinitesimal deformation is linearly elastic if and only if the given initial strain is maintained by a state of hydrostatic stress.” This holds for the stress-free reacted (and strained) configuration since a vanishing Cauchy stress is a state of hydrostatic stress. Therefore, we consider a deformation gradient corresponding to an infinitesimal strain superimposed on the reacted state

$$\mathbf{F}^\varepsilon \;=\; \frac{\partial \boldsymbol{\chi}^\varepsilon}{\partial \mathbf{X}} \;=\; \frac{\partial \boldsymbol{\chi}^\phi}{\partial \mathbf{X}} + \frac{\partial \mathbf{u}}{\partial \boldsymbol{\chi}^\phi}\frac{\partial \boldsymbol{\chi}^\phi}{\partial \mathbf{X}} \;=\; \mathbf{F}^\phi + \mathbf{H}\mathbf{F}^\phi \tag{3.20}$$

where $\boldsymbol{\chi}^\phi$ denotes the deformation mapping for the stress-free reacted configuration, $\mathbf{u}$ is an admissible infinitesimal displacement field applied to the stress-free reacted configuration, and $\mathbf{H}$ is the displacement gradient. The corresponding left Cauchy-Green deformation tensor is

$$\mathbf{B}^\varepsilon \;=\; \left(\mathbf{F}^\phi + \mathbf{H}\mathbf{F}^\phi\right)\left(\mathbf{F}^\phi + \mathbf{H}\mathbf{F}^\phi\right)^T \;=\; \mathbf{B}^\phi + \mathbf{B}^\phi\mathbf{H}^T + \mathbf{H}\mathbf{B}^\phi + \mathbf{H}\mathbf{B}^\phi\mathbf{H}^T \tag{3.21}$$

The Cauchy stress under this deformation is well approximated by taking a Taylor expansion of the stress around $\mathbf{B} = \mathbf{B}^\phi$, recalling (3.15), and dropping terms of order 2 and higher in $\mathbf{H}$

$$\mathbf{T}^s \;=\; \bar{\mathbf{T}}(\mathbf{B}^\varepsilon) \;\approx\; \bar{\mathbf{T}}(\mathbf{B}^\phi) + \boldsymbol{\mathcal{A}}^\phi : \gamma^2\left(\mathbf{H}^T + \mathbf{H}\right) \tag{3.22}$$

where the fourth order tensor $\boldsymbol{\mathcal{A}}^\phi$ is defined as

$$\boldsymbol{\mathcal{A}}^{\phi} = \left.\frac{\partial \bar{\mathbf{T}}}{\partial \mathbf{B}}\right|_{\mathbf{B}=\mathbf{B}^{\phi}} \tag{3.23}$$

And the action of a fourth other tensor on a second order tensor is defined using Einstein indicial notation as

$$\boldsymbol{\mathcal{A}}^{\phi} : \mathbf{G} \equiv \mathcal{A}^{\phi}_{ijpq} G_{pq} \tag{3.24}$$

Recalling the definition of $\mathbf{B}^{\phi}$ being the left Cauchy-Green deformation tensor for the stress-free reacted configuration, it follows that

$$\bar{\mathbf{T}}(\mathbf{B}^{\phi}) = \mathbf{0} \tag{3.25}$$

Then, the symmetric part of the superimposed displacement gradient is defined as follows

$$\boldsymbol{\varepsilon} = \frac{1}{2}\left(\mathbf{H} + \mathbf{H}^{T}\right) \tag{3.26}$$

Computing the tensor gradient in (3.23), and using equations (3.17), (3.25), and (3.26), the expression in (3.22) can be expanded as

$$\begin{aligned} \mathbf{T}^{s} \approx 2\gamma^{2}\boldsymbol{\mathcal{A}}^{\phi} : \boldsymbol{\varepsilon} = & \frac{2}{\gamma}\left[\mu^{s}(\phi) - C^{s}(\phi)\right]\boldsymbol{\varepsilon} \\ & + \left\{\frac{2}{\gamma^{3}}\mu^{s}(\phi) - \frac{2}{\gamma}\left[\mu^{s}(\phi) - C^{s}(\phi)\right] + \gamma^{3}\lambda^{s}(\phi)\right\}\operatorname{tr}(\boldsymbol{\varepsilon})\mathbf{1} \end{aligned} \tag{3.27}$$

From equation (3.27) one can deduce the expressions for the apparent elastic constants $\tilde{\mu}^{s}(\phi)$ and $\tilde{\lambda}^{s}(\phi)$ (shear modulus and Lamé's first parameter, respectively) of the reacted solid

$$\tilde{\mu}^{s}(\phi) = \frac{1}{\gamma}\left[\mu^{s}(\phi) - C^{s}(\phi)\right] \tag{3.28}$$

$$\tilde{\lambda}^{s}(\phi) = \frac{2}{\gamma^{3}}\mu^{s}(\phi) - \frac{2}{\gamma}\left[\mu^{s}(\phi) - C^{s}(\phi)\right] + \gamma^{3}\lambda^{s}(\phi) \tag{3.29}$$

Further constraints on the material parameters are obtained by invoking the so-called classic inequalities, which arise in connection with the arguments of plausibility in statics, work, and stability [17],

$$\tilde{\mu}^{s}(\phi) > 0 \tag{3.30}$$

$$3\tilde{\lambda}^{s}(\phi) + 2\tilde{\mu}^{s}(\phi) > 0 \tag{3.31}$$

3.3.3 Determination of material parameters

The functional form of $C^s(\phi)$, which emanates from the dependence of the chemo-mechanical coupling term on the reaction state ϕ, is given by

$$C^s(\phi) = \bar{C}^s(1-\phi)^3 \tag{3.32}$$

The other material parameters, which correspond to the Lamé parameters when $\phi = 1$ (unreacted or pristine state), were assumed in [11] to depend linearly on the reaction state ϕ.

$$\mu^s(\phi) = \bar{\mu}_1^s\phi + \bar{\mu}_2^s(1-\phi) \tag{3.33}$$

$$\lambda^s(\phi) = \bar{\lambda}_1^s\phi + \bar{\lambda}_2^s(1-\phi) \tag{3.34}$$

where $\bar{\lambda}_1^s$ and $\bar{\mu}_1^s$ are the Lamé's first parameter and shear modulus, respectively, of the unreacted (or pristine) solid. On the other hand, $\bar{\lambda}_2^s$ and $\bar{\mu}_2^s$ are material parameters for the fully reacted solid ($\phi = 0$). However, they do not necessarily correspond to the Lamé's first parameter and shear modulus of the reacted material. The form of the apparent Lamé parameter and shear modulus of the solid in its reacted state are given by equations (3.28) and (3.29).

When modeling reactive solids with chemically induced swelling, the mechanical material parameters (first Lamé parameter and shear modulus) for the unreacted (or pristine) solid and for the fully reacted solid are either known or possible to identify through experimental results. Furthermore, the final volumetric change due to complete chemical reaction is also known or is measurable through experiments. This yields the value of J^ϕ and, using (3.17), the value of γ for the fully reacted state $\phi = 0$. As already mentioned, in this model for the Cauchy stress the material parameters $\bar{\lambda}_1^s$ and $\bar{\mu}_1^s$ correspond, respectively, to the first Lamé parameter and the shear modulus of the unreacted (or pristine) solid. In order to determine the remaining material parameters ($\bar{\lambda}_2^s$, $\bar{\mu}_2^s$, and $\bar{C}^s$), one can consider the fully reacted state $\phi = 0$, for which the first Lamé parameter, shear modulus, and chemical stretch $\tilde{\lambda}^s$, $\tilde{\mu}^s$, and γ are known. Then, substituting (3.32) – (3.34) into (3.18), (3.28), and (3.29), and setting $\phi = 0$, yields a linear algebraic system of three equations and three unknowns ($\bar{\lambda}_2^s$, $\bar{\mu}_2^s$, and $\bar{C}^s$):

$$\bar{\mu}_2^s - \bar{C} = \gamma\tilde{\mu}^s \tag{3.35}$$

$$2\left(\frac{1}{\gamma^2}-1\right)\bar{\mu}_2^s+2\bar{C}+\gamma^4\bar{\lambda}_2^s=\gamma\tilde{\lambda}^s \tag{3.36}$$

$$\left(\gamma^2-1\right)\bar{\mu}_2^s-\gamma^2\bar{C}+\left(\gamma^6-\gamma^3\right)\bar{\lambda}_2^s=0 \tag{3.37}$$

We can write (3.35) to (3.37) in matrix form

$$\begin{bmatrix} 1 & -1 & 0 \\ 2\left(\gamma^{-2}-1\right) & 2 & \gamma^4 \\ \left(\gamma^2-1\right) & -\gamma^2 & \left(\gamma^6-\gamma^3\right) \end{bmatrix}\begin{bmatrix} \bar{\mu}_2^s \\ \bar{C} \\ \bar{\lambda}_2^s \end{bmatrix}=\begin{bmatrix} \gamma\tilde{\mu}^s \\ \gamma\tilde{\lambda}^s \\ 0 \end{bmatrix} \tag{3.38}$$

As a result, equation (3.38) allows to find the remaining material parameters in the constitutive equation for the Cauchy stress for the solid from values of parameters that exist in the literature or are straightforward to determine from experiments.

3.4 Reduced Model for the Fluid Constituent

Starting from the constitutive and governing equations in [11] and summarized in section 3.2, the following assumptions are made regarding the balance of linear momentum of the fluid: (i) the effect of body forces is negligible compared to other terms; (ii) the main contribution to the interactive force is from the drag term; (iii) the viscous contribution to the Cauchy stress for the fluid is small compared to the pressure term; and (iv) the inertial effects in the fluid are negligible. Assumptions (i) to (iv) are reflected in equations (3.39), (3.40), (3.41), and (3.42), respectively. It is noted that assumptions (iii) and (iv) are consistent with a slow speed of diffusion of the fluid into the solid constituent.

$$\mathbf{b}^f\approx\mathbf{0} \tag{3.39}$$

$$\mathbf{I}^f\approx-\mathbf{A}^v(\mathbf{v}^f-\mathbf{v}^s) \tag{3.40}$$

$$\mathbf{T}^f\approx-\frac{R}{W_f}\theta_0\rho^f\mathbf{1} \tag{3.41}$$

$$\rho^f\frac{D^f\mathbf{v}^f}{Dt}=\rho^f\frac{\partial\mathbf{v}^f}{\partial t}+\rho^f(\mathbf{v}^f-\mathbf{v}^s)\cdot\nabla\mathbf{v}^f\approx\mathbf{0} \tag{3.42}$$

Furthermore, the drag tensor $\mathbf{A}^v$ is assumed to have the following form

$$\mathbf{A}^v=\frac{R}{W_f}\theta_0\rho^f\mathbf{D}^{-1} \tag{3.43}$$

where $\mathbf{D}$ is a positive-definite diffusivity tensor, which in general may also depend on other state variables. The functional form of this dependency is left unspecified at this point, and the form of $\mathbf{D}$ is specialized for each of the case studies presented in section 3.6, accordingly. Equations (3.39) through (3.43) are substituted into the balance of linear momentum equation for the fluid constituent in equation (3.2). The resulting equation can be solved algebraically for the fluid velocity field to yield the following equation

$$\mathbf{v}^f = \mathbf{v}^s - \frac{1}{\rho^f}\mathbf{D}\nabla\rho^f \tag{3.44}$$

The balance of mass for the fluid in (3.1) is rewritten by gathering the fluid velocity terms as follows

$$\frac{\partial\rho^f}{\partial t} - \mathbf{v}^s \cdot \nabla\rho^f + \operatorname{div}(\rho^f \mathbf{v}^f) - m^f = 0 \tag{3.45}$$

The expression for fluid velocity given in equation (3.44) is embedded into the balance of mass equation for the fluid in (3.45). This yields a transient reaction-diffusion equation for the fluid mass concentration field, which is shown below. Remarkably, the advection terms cancel and do not appear in the final form.

$$\frac{\partial\rho^f}{\partial t} - \operatorname{div}\left(\mathbf{D}\nabla\rho^f\right) + \rho^f \operatorname{div}\left(\mathbf{v}^s\right) - m^f = 0 \tag{3.46}$$

The constitutive equation for the fluid mass transfer rate m^f is left unspecified at this point, and it is presented and specialized in each case study in the numerical section. It is noted that, in general, the mass transfer rate may be stress dependent ($\mathbf{T}^s$) and/or depend on the state variables ϕ, ρ^f, and $\mathbf{F}^s$.

3.5 Reduced System of Governing Equations

This section discusses additional simplifying assumptions on the governing equations from section 3.2 and summarizes the reduced system of governing equations. First, section 3.4 discussed some assumptions on the balance of linear momentum and constitutive equations for the fluid. Regarding the solid constituent, the interactive force and body force ($\mathbf{I}^s$ and $\rho^s\mathbf{b}^s$) are taken to be negligible as compared to the solid stress term. Therefore, the solid mass concentration field appears only in (3.3), and its gradient does not appear anywhere in the formulation. Provided the solid mass concentration field is not a quantity of interest, it is possible to reduce the set of unknown fields to the solid displacement field $\mathbf{u}^s$, fluid mass concentration field ρ^f, and reaction state variable ϕ. These fields are solved for using the system of equations given by

$$\frac{\partial \rho^f}{\partial t} - \operatorname{div}\left(\mathbf{D}\nabla\rho^f\right) + \rho^f \operatorname{div}\left(\mathbf{v}^s\right) - m^f = 0 \qquad \text{in } \Omega_t \tag{3.47}$$

$$\operatorname{div}\left(\mathbf{T}^s\right) = \mathbf{0} \qquad \text{in } \Omega_t \tag{3.48}$$

plus the constitutive equations for the solid Cauchy stress $\mathbf{T}^s$ defined in (3.9), the constitutive equations and functional forms of the fluid mass transfer rate m^f, reaction state variable evolution $D^s\phi/Dt$, and the diffusivity tensor $\mathbf{D}$, which are specialized for each case study in section 3.6. It is remarked that the fluid velocity field does not appear in the governing equations anymore. One important consequence for the numerical implementation of the model using the finite element method is that it is no longer required to discretize a PDE to solve for the fluid velocity field. In a 3D setting, this corresponds to three fewer degrees of freedom (DOFs) per node (those associated with the velocity field in 3D space) as compared to a model using the full fluid model instead of the reduced one from section 3.4. For a given mesh and monolithic solution scheme, fewer degrees of freedom per node result in a smaller matrix system, and therefore its solution incurs reduced computational cost.

Initial conditions and boundary conditions are provided as follows

$$\mathbf{u}^s = \mathbf{u}_0^s\left(\mathbf{X}\right) \qquad \text{in } \Omega_0 \tag{3.49}$$

$$\rho^f = \rho_0^f\left(\mathbf{X}\right) \qquad \text{in } \Omega_0 \tag{3.50}$$

$$\phi = \phi_0\left(\mathbf{X}\right) \qquad \text{in } \Omega_0 \tag{3.51}$$

$$u_i^s = u_{i,b}^s\left(t\right) \qquad \text{on } \partial\Omega_{t,D_i}^s \tag{3.52}$$

$$t_i^s = T_{ij}^s n_j = h_i\left(t\right) \qquad \text{on } \partial\Omega_{t,N_i}^s \tag{3.53}$$

$$\rho^f = \rho_b^f\left(t\right) \qquad \text{on } \partial\Omega_{t,D}^f \tag{3.54}$$

$$\left(\mathbf{D}\nabla\rho^f\right)\cdot\mathbf{n} = j\left(t\right) \qquad \text{on } \partial\Omega_{t,N}^f \tag{3.55}$$

where $\partial\Omega_{t,D_i}^s \cup \partial\Omega_{t,N_i}^s = \partial\Omega_{t,D}^f \cup \partial\Omega_{t,N}^f = \partial\Omega_t$ and $\partial\Omega_{t,D_i}^s \cap \partial\Omega_{t,N_i}^s = \partial\Omega_{t,D}^f \cap \partial\Omega_{t,N}^f = \varnothing$. In addition, $\mathbf{u}_0^s$, ρ_0^f, ϕ_0, $u_{i,b}^s$, h_i, ρ_b^f, and j are known functions.

Regarding the coupling between solid and fluid fields, note that fluid diffusion happens over the deforming domain, and solid velocity appears in fluid balance of mass equation. On the other hand, solid stress depends on the reaction state, which in turn is driven by the presence of fluid reactant.

3.6 Numerical Results

The model developed in this chapter and summarized in section 3.5 was implemented into a finite element (FE) code utilizing nodal Lagrangian polynomials as the basis functions (in particular, using 8-node hexahedral elements or "HEX8" elements). Quadrature rule of sufficiently high order is employed. The numerical solution of the governing system of equations is carried out in a fully coupled, monolithic way to ensure accuracy and stability. The nodal degrees of freedom are the solid displacements and the fluid mass concentration. It is highlighted that the governing equations in (3.47) and (3.48) are diffusion-type equations whose coercivity properties get inherited by the discrete problem. This allows for equal order interpolation of the underlying unknown fields (solid displacement and fluid mass concentration) during monolithic solution of the system.

A backward Euler method is used to discretize in time. The reaction state variable ϕ is treated as an internal variable, since its spatial gradient does not appear anywhere in the formulation, and it is solved for at the integration points through an ODE that is discretized with the backward Euler algorithm as well. At each time level, the nonlinear system is solved using a Newton-Raphson algorithm. The consistent tangent was implemented, which for a monolithic solution scheme, results in quadratic convergence rates when the residual vector varies smoothly with respect to the degrees of freedom. In 3D space, this reduced model has three degrees of freedom per node less compared to using the full fluid model, since there is no need to solve a PDE to find the fluid velocity field in the reduced formulation. For the same number of finite elements in a mesh, fewer degrees of freedom per node translates into a smaller matrix system, and therefore each solution requires less computational resources. To make this argument precise, consider a direct solver is used to solve the (unsymmetric) system, namely LU factorization, whose computational cost scales as $\mathcal{O}\left(n^3\right)$ [21], where n is the size of the matrix system. For 3D, using the reduced model with 4 degrees of freedom per node instead of 7 degrees of freedom in the full model results in up to a 67% reduction in computational cost.

The finite element method implementation of the model was employed to generate numerical results for two test cases. The first test case models the bending-dominated response of a silicon wafer during thermal oxidation. The second test case studies the response at the cross-section of a silicon nanowire during lithiation. The constitutive equations and functional forms of the fluid mass transfer rate m^f, reaction state variable evolution $D^s\phi/Dt$, and diffusivity tensor $\mathbf{D}$ are specialized for each case study.

3.6.1 Thermal oxidation of silicon wafer

The passive thermal oxidation of silicon at temperatures around 1000ºC proceeds according to the following chemical reaction

$$\mathrm{Si}_{(s)} + \mathrm{O}_{2(g)} \rightarrow \mathrm{SiO}_{2(s)} \tag{3.56}$$

This reaction induces a chemically driven volume expansion, where the produced (unconstrained) solid silicon dioxide (SiO_2) occupies 2.2 times the volume of the pure Si substrate that it replaces. A brief description of this process is as follows: interstitial molecular oxygen (O_2) diffuses from the ambient atmosphere through the oxide (SiO_2) to reach the unreacted Si substrate, where it is consumed by reaction according to (3.56) to produce more SiO_2, causing recession of the Si substrate and growth of the SiO_2 layer.

For modeling the thermal oxidation of Si in the context of the formulation presented herein, the interstitial O_2 is taken as the fluid constituent, while the solid constituent evolves from Si (for $\phi = 1$) to SiO_2 ($\phi = 0$). The functional form for the mass transfer rate for the fluid is defined as

$$m^f\left(\rho^f, \phi\right) = R_0 \phi \rho^f \tag{3.57}$$

where the effective reaction rate constant R_0 modulates the speed of chemical reaction in the bulk domain. In this test case it is taken as constant, but in general, it may accommodate temperature and stress dependence. The evolution equation for the reaction state is given by

$$\frac{D^s\phi}{Dt} = R_\phi \phi \rho^f \tag{3.58}$$

where the parameter R_ϕ can be determined based on R_0 and the stoichiometric relations for the governing chemical reaction. The value of R_0 and R_ϕ are taken to agree with the values used in

[11]. The diffusivity is taken to be isotropic and to have the following dependence on the reaction state

$$\mathbf{D} = d^s(\phi)\mathbf{1} = \left[\bar{d}_1^s\phi + \bar{d}_2^s(1-\phi)\right]\mathbf{1} \tag{3.59}$$

where the material parameters $\bar{d}_1^s$ and $\bar{d}_2^s$ are, respectively, the effective diffusivity of O_2 in Si and in SiO_2 at the reference temperature. Their values are taken to be the same as reported in [22].

Regarding the material parameters that appear in the Cauchy stress equation, $\bar{\lambda}_1^s$ and $\bar{\mu}_1^s$ are, respectively, the first Lamé parameter and shear modulus of Si, λ_{Si} and μ_{Si}. These are determined from the values reported[2] in [22]. The results from section 3.3 facilitate determination of the remaining material parameters ($\bar{\lambda}_2^s$, $\bar{\mu}_2^s$, and $\bar{C}^s$). They are computed from the value of the material properties of the fully reacted solid (SiO_2) using the matrix equation in (3.38), and taking $\tilde{\lambda}^s = \lambda_{\mathrm{SiO_2}}$, $\tilde{\mu}^s = \mu_{\mathrm{SiO_2}}$, and $\gamma = 1.3$ to account for the volumetric expansion, according to (3.23) (unconstrained SiO_2 occupies 2.2 times the volume of the original Si volume it replaces). Again, values for $\lambda_{\mathrm{SiO_2}}$ and $\mu_{\mathrm{SiO_2}}$ are determined from [22]. The values for all the model parameters employed for the thermal oxidation of Si are summarized in Table 3.1. It is noted that the values of the mechanical parameters are such that inequalities (3.19), (3.30), and (3.31) are satisfied.

Table 3.1. Model parameters for thermal oxidation of silicon (Si)

Parameter	Value	Units
R_0	-5.6×10^5	s^{-1}
R_ϕ	-211	$m^3\,kg^{-1}\,s^{-1}$
$\bar{d}_1^s$	1.38×10^{-12}	m^2/s
$\bar{d}_2^s$	1.38×10^{-10}	m^2/s
$\bar{C}^s$	56.8	GPa
$\bar{\lambda}_1^s$	93.0	GPa
$\bar{\mu}_1^s$	73.0	GPa

---

[2] The bulk and shear moduli for Si reported in [22] result in a negative Poison ratio, which is unrealistic for Si. Therefore, the values listed for bulk modulus are used in this work as the values for the first Lamé parameter of Si and SiO_2.

Table 3.1 (cont.)

$\bar{\lambda}_2^s$	3.45	GPa
$\bar{\mu}_2^s$	126	GPa

The model and material parameters presented herein are used to run the numerical test case of oxidation-induced bending of a Si wafer presented in [11,22]. Consider a domain representing the cross-section of a Si wafer whose top surface is exposed to an oxidizing environment of 1 atm O_2 partial pressure at 1100ºC (1373K). This corresponds to a saturation concentration of $\rho^f_{\text{sat}} = 1.22 \times 10^{-3}$ kg/m^3 of O_2 on the fully reacted boundary [20]. No fluid outflow ($j = 0$) is prescribed along all other surfaces. Taking advantage of the problem symmetries, only half the wafer cross-section is modeled, while applying appropriate symmetry boundary conditions at the left end of the domain (at $x = 0$). A schematic diagram of the domain and the boundary conditions is shown in Figure 3.1.

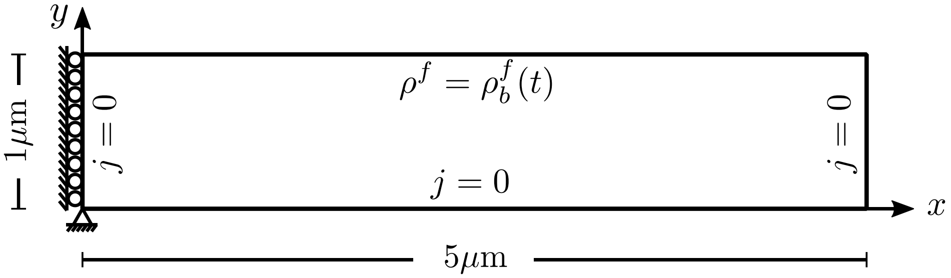


Figure 3.1. Schematic diagram of half of the Si wafer cross-section geometry along with the boundary conditions.

The computational domain is discretized using a mesh of 25×20×1 HEX8 elements. Appropriate boundary conditions are applied to induce a 2D plane strain response within the x-y plane. The Dirichlet boundary condition for the fluid mass concentration is applied gradually over a period of time $\Delta t_r = 1 \times 10^4$ s according to the equation $\rho^f_b(t) = (t/\Delta t_r)\rho^f_{sat}$. A backward Euler time marching algorithm is used with a fixed time step of $\Delta t = 10$ s.

The computed solution is compared with the published data [11,22] at the state of deformation where the position of the top-right corner of the domain in the deformed configuration has (x, y) coordinates $(5.063, 0.7741)\mu$m and the lower-right corner has coordinates $(4.967, -0.2178)\mu$m. This corresponds to $t = 6 \times 10^3$ s in [22], $t = 410$ s in [11], and $t = 510$ s in the present chapter. The difference in time in attaining the same state of deformation is attributed to discrepancies in the application of boundary conditions as well as the differences in the models employed. The contours of the T_{11}^{s} component of the stress are shown in Figure 3.2.

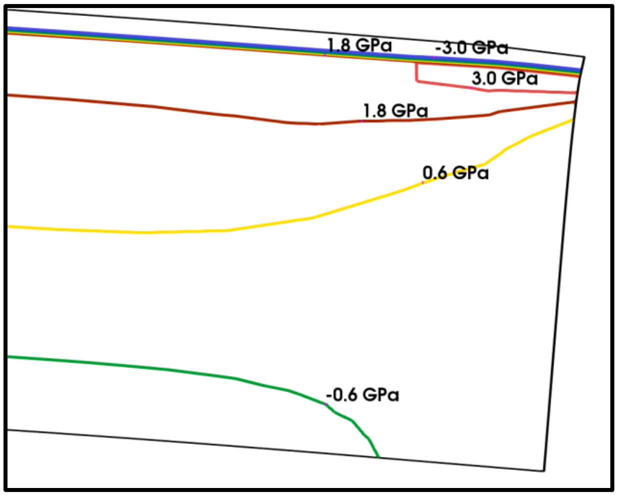


Figure 3.2. Close-up view of the contours of the T_{11}^{s} component of the solid Cauchy stress near the right end of the deformed domain at $t = 510$ s.

It is noted that the stress distribution, as evidenced by the stress contours in Figure 3.2, is in good agreement with those reported in [11,22]. In addition, the range of stress values for T_{11}^{s} (i.e., its maximum and minimum values) are also in reasonable agreement (see Table 3.2). The values obtained here are somewhat larger in magnitude, which is to be expected since the bulk modulus values reported in the reference were taken to be the values of first Lamé parameters in this chapter, which results in a stiffer material. As expected, for a comparable deformation state, a stiffer material develops larger stresses.

Table 3.2. Maximum and minimum values of the T_{11}^{s} component of Cauchy stress (in GPa)

	This chapter	Anguiano et al. [11]	Rao et al. [22]
maximum	5.24	4.88	4.85
minimum	-20.6	-20.3	-12.8

Simulation of the thermal oxidation of the Si wafer is carried out for a longer period of time, and Figure 3.3 shows the distribution of the reaction state field ϕ and of the von Mises stress in the deformed domain at four instances during the solution procedure corresponding to $t =$ 510s, 1000s, 2000s, and 3000s. As time progresses, the oxide layer grows and the pristine Si substrate recedes. As this happens, the expanding SiO_2 layer is constrained by the adjacent Si substrate at the interface between the two. As a result, the oxide layer undergoes compressive stress near the interface. On the other hand, the same expanding SiO_2 stretches the adjacent Si substrate, causing the top of the Si to undergo tensile stress. This, in turn, induces bending-dominated response as observed in the snapshots.

This observation is also reflected in Figure 3.4, which shows the profiles of the T_{11}^{s} component of the Cauchy stress along the reference y axis at the left end of the domain ($x = 0$). In this figure, it is easy to appreciate the rapid variation from negative (compressive) to positive (tensile) values of the T_{11}^{s} stress component across the transition zone of the material. Note that the neutral axis within the pristine substrate remains roughly at the same location. However, the extreme values of the T_{11}^{s} stress component within the Si substrate grow larger in magnitude. This induces a progressively larger internal bending moment, which results in the bending-dominated response as observed in the snapshots from Figure 3.3.

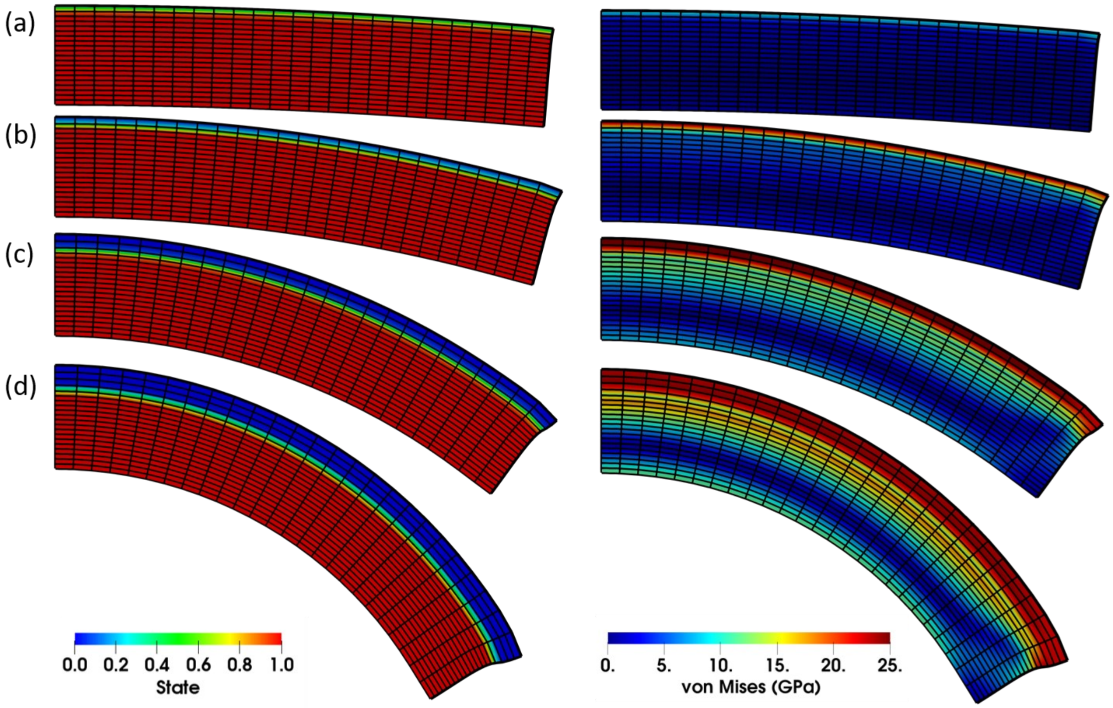


Figure 3.3. Snapshots of the distribution of the reaction state field (left) and of the von Mises stress (right) in the deformed configuration at four points in time: (a) 510s, (b) 1000s, (c) 2000s, and (d) 3000s.

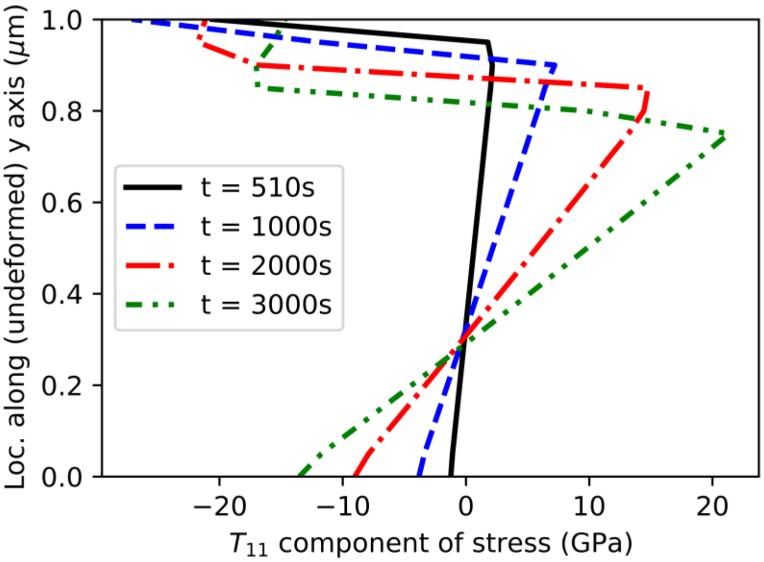


Figure 3.4. T_{11}^s component of the Cauchy stress in the solid along the y axis (left end of the domain, $x = 0$) in the undeformed configuration at four points in time: 510s, 1000s, 2000s, and 3000s.

The results of this section are in accordance with the results reported in [11] for the same problem description. However, the material parameters employed in the present chapter for the constitutive equation of the Cauchy stress for the solid, were determined in terms of the reported material parameters of the fully reacted solid (SiO_2) through equation (3.38). This equation provides the relation for the material parameters based on the rigorous analysis presented in section 3.3, instead of relying on a heuristic approach for material parameter determination. Therefore, discrepancies in the magnitude of the stress arose in connection to the use of different material parameters. Moreover, the reduced model presented in this chapter was employed to perform the numerical simulations. One consequence is the decrease in the incurred computational cost since the reduced model presented here leads to a numerical implementation with fewer degrees of freedom (DOFs) per node. Furthermore, Newton-Raphson algorithm achieves quadratic convergence in every time step during solution. Table 3.3 shows the residual norm reduction for Newton-Raphson algorithm for two timesteps during solution procedure, corresponding to the first and last snapshots from Figure 3.3.

Table 3.3. Residual norm reduction during solution procedure

Time step number	NR iteration counter	Residual Norm	Relative Norm
51	0	6.42E-4	1.00E0
	1	1.62E-6	2.53E-3
	2	2.49E-10	3.88E-7
	3	6.55E-16	1.02E-12
300	0	9.70E-4	1.00E0
	1	1.01E-6	1.04E-3
	2	9.48E-10	9.77E-7
	3	4.78E-15	4.93E-12

3.6.2 Lithiation of silicon nanowire

The electrochemical process in the lithium-ion (Li-ion) battery depends upon a Li intercalation reaction. Employing electrode materials with a high Li storage capacity is a key factor in increasing energy density of a battery. Unfortunately, a high capacity in Li storage is accompanied with large volume change and high stress in the materials. The purpose of this test case is to demonstrate that the model developed in this chapter, is also applicable to the class of problems involved with lithiation of silicon. A review of various approaches for the modeling of lithiation in the anodes of Li-ion batteries is presented in [10]. For this case study, the fluid constituent is taken to be the lithium (Li) ions, and the evolving solid constituent is taken to be pure Si in its unreacted state ($\phi = 1$), and $Li_{3.75}Si$ in its fully reacted/lithiated state ($\phi = 0$). The resulting chemically driven expansion is 300% of the original Si volume (i.e., $Li_{3.75}Si$ occupies 4 times the volume of the pure Si that it replaces). From this, the saturation concentration of Li in the fully lithiated $Li_{3.75}Si$ is estimated to be $\rho^f_{\text{sat}} = 542.6\,\text{kg/m}^3$. Following existing work [10,23–25], Li is assumed not to be consumed by a reaction, but its transport within Li_xSi is governed by a diffusivity with a nonlinear dependence on the amount of Li present. Therefore, a similar approach is adopted here, and in the context of the model from section 3.5, the following functional forms are employed: $m^f = 0$, $\phi = 1 - \rho^f / \rho^f_{\text{sat}}$, and the diffusivity is given by

$$\mathbf{D} = \left[\frac{1}{\phi^2} + (1-\phi) \right] \bar{\mathbf{D}} \tag{60}$$

where the nonlinear evolution of the diffusivity gives rise to a narrow phase transition zone, as observed in experiments. Note that as $\phi \to 0$, $\mathbf{D}$ grows unbounded. To avoid numerical instability arising because of unboundedness of $\mathbf{D}$, its value is capped at $\mathbf{D} = 2 \times 10^3 \bar{\mathbf{D}}$. Furthermore, the diffusivity tensor is taken to be anisotropic in order to account for the differences in crystallographic orientations $D_{\langle 110 \rangle} \approx 6D_{\langle 100 \rangle} \approx 60D_{\langle 111 \rangle}$. Specifically, we set the following: $\bar{D}_{11} = D_{\langle 110 \rangle}$, $\bar{D}_{22} = D_{\langle 111 \rangle}$ and $\bar{D}_{33} = \bar{D}_{11}$.

Similar to what was done for the Si oxidation, the mechanical parameters are defined as $\bar{\lambda}_1^s = \lambda_{\text{Si}}$ and $\bar{\mu}_1^s = \mu_{\text{Si}}$. The remaining parameters ($\bar{\lambda}_2^s$, $\bar{\mu}_2^s$, and $\bar{C}^s$) are determined using equation (3.38) (resulting from the rigorous analysis from section 3.3), and setting $\tilde{\lambda}^s = \lambda_{\text{Li}_{3.75}\text{Si}}$, $\tilde{\mu}^s = \mu_{\text{Li}_{3.75}\text{Si}}$, and $\gamma = 1.587$ (to account for the experimentally-observed chemical expansion). Young's moduli and Poisson's ratios as reported in [23,25] are $E_{\text{Si}} = 160\,\text{GPa}$, $E_{\text{Li}_{3.75}\text{Si}} = 40\,\text{GPa}$, $\nu_{\text{Si}} = 0.24$ and $\nu_{\text{Li}_{3.75}\text{Si}} = 0.22$. From these values, the corresponding first Lamé parameters and shear moduli are $\lambda_{\text{Li}_{3.75}\text{Si}} = 12.88\,\text{GPa}$ and $\mu_{\text{Li}_{3.75}\text{Si}} = 16.39\,\text{GPa}$. The values for all the model parameters employed for the lithiation of Si are summarized in Table 3.4. It is noted that the values of the mechanical parameters are such that inequalities (3.19), (3.30), and (3.31) are satisfied.

Table 3.4. Model parameters for the lithiation of silicon (Si)

Parameter	Value	Units
$\bar{\lambda}_1^s$	59.55	GPa
$\bar{\mu}_1^s$	64.52	GPa
$\bar{\lambda}_2^s$	1.289	GPa
$\bar{\mu}_2^s$	80.95	GPa
$\bar{C}^s$	54.94	GPa
ρ_{sat}^f	542.6	kg/m^3
$\bar{D}_{\langle 110 \rangle}$	1×10^{-12}	m^2/s

We consider the problem of lithiation through a cross-section of a silicon nanowire (SiNW) similar to the case presented in [23]. The domain under consideration is the cross section of a SiNW of radius $r_0 = 108\,\text{nm}$ before lithiation, whose outer surface is exposed to a fixed concentration of Li^+ which diffuses into the SiNW. Taking advantage of the problem symmetries, it suffices to model one quarter of the domain while imposing appropriate symmetry boundary conditions. A schematic diagram of the domain and boundary conditions are shown in Figure 3.5.

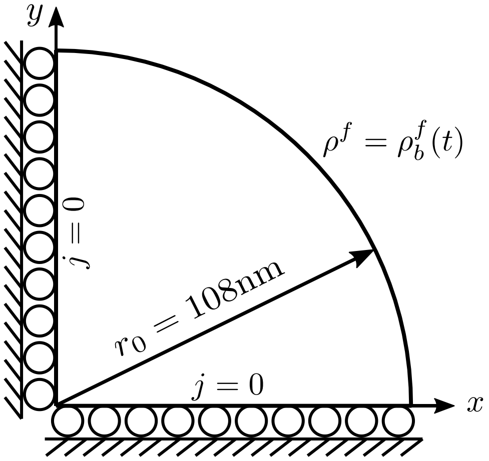


Figure 3.5. Schematic diagram of one quarter of the Si nanowire cross-section geometry along with boundary conditions.

The computational domain is discretized with a mesh of HEX8 elements with only one element through the thickness (longitudinal axis of nanowire). Total number of elements is 975. Appropriate boundary conditions are applied to induce a 2D plane strain response within the x-y plane. The Dirichlet boundary condition for the fluid mass concentration is applied gradually over a period of time $\Delta t_r = 1.1664 \times 10^{-5}\,\text{s}$ according to the equation $\rho_b^f(t) = (t/\Delta t_r)\rho_{sat}^f$. A backward Euler time marching algorithm is used with varying time step Δt presented in Table 3.5. The domain is taken to be completely lithiated by time $t_f = 4.201 \times 10^{-4}$ s. In reporting the results, time is normalized by the total time as $\bar{t} = t/t_f$.

Table 3.5. Time steps used during solution procedure of lithiation problem

Time interval (s)	Time step Δt	Number of time steps
0 to 1.050×10^{-5}	$\Delta t = \Delta t_r/10$	9
1.050×10^{-5} to 1.155×10^{-5}	$\Delta t = \Delta t_r/100$	9
1.155×10^{-5} to 1.190×10^{-5}	$\Delta t = \Delta t_r/1000$	30
1.190×10^{-5} to 1.285×10^{-4}	$\Delta t = \Delta t_r/100$	1000
1.285×10^{-4} to 4.201×10^{-4}	$\Delta t = \Delta t_r/10$	250

In order to allow for comparison with [23], Figure 3.6 presents snapshots of the solution in the deformed domain at four instants corresponding to the normalized times $\overline{t}$ = 0.01, 0.05, 0.1, and 0.3. The first field shown is the reaction state ϕ, where $\phi = 1$ corresponds to pristine Si and $\phi = 0$ corresponds to fully lithiated material ($Li_{3.75}Si$). The response is anisotropic, where the reaction front advances faster along the x-axis, corresponding to $\langle 110 \rangle$ direction of Si, while it is slower along the y-axis, corresponding to $\langle 111 \rangle$ direction of Si. This also results in the larger observed deformation along the x-axis. The second field is the maximum in-plane principal stress. During the lithiation process the reaction front advances and the magnitude of the stress increases due to the constrained expansion of the material. In addition, the value of the stress varies rapidly near the transition zone between lithiated and unlithiated material.

Figure 3.7 shows the time history evolution (in normalized time) of the reaction state variable ϕ as well as the hydrostatic and von Mises stress, for a point located along the x-axis at distance $r = 0.56 r_0 = 60.48$ nm from the origin in the undeformed configuration. Note that the reaction state varies rapidly at around $\overline{t} \approx 0.2$, which corresponds to the moment where the lithiation front advances through that point. In addition, this coincides with sudden changes in the evolution of the hydrostatic stress and the von Mises stress.

The results obtained for the reaction state field ϕ in Figure 3.6 and Figure 3.7are in reasonable agreement with the data reported in [23]. Nonetheless, the results for the stresses are somewhat different. This discrepancy is attributed to the fact that the model employed in the present chapter accounts for chemical and elastic deformations only. Meanwhile, the results reported in [23] account for plastic effects as well. In this process, plastic effects are relevant, since it is clear the stress field reaches large values that would result in plastic flow. Therefore, additional plastic deformations would relieve the excess stress (above the yielding point) and one would obtain a different deformation field and associated lower stress magnitudes than what is obtained in this chemical-purely-elastic analysis. The development of a model that also accounts for viscous and plastic effects will be carried out in a future work.

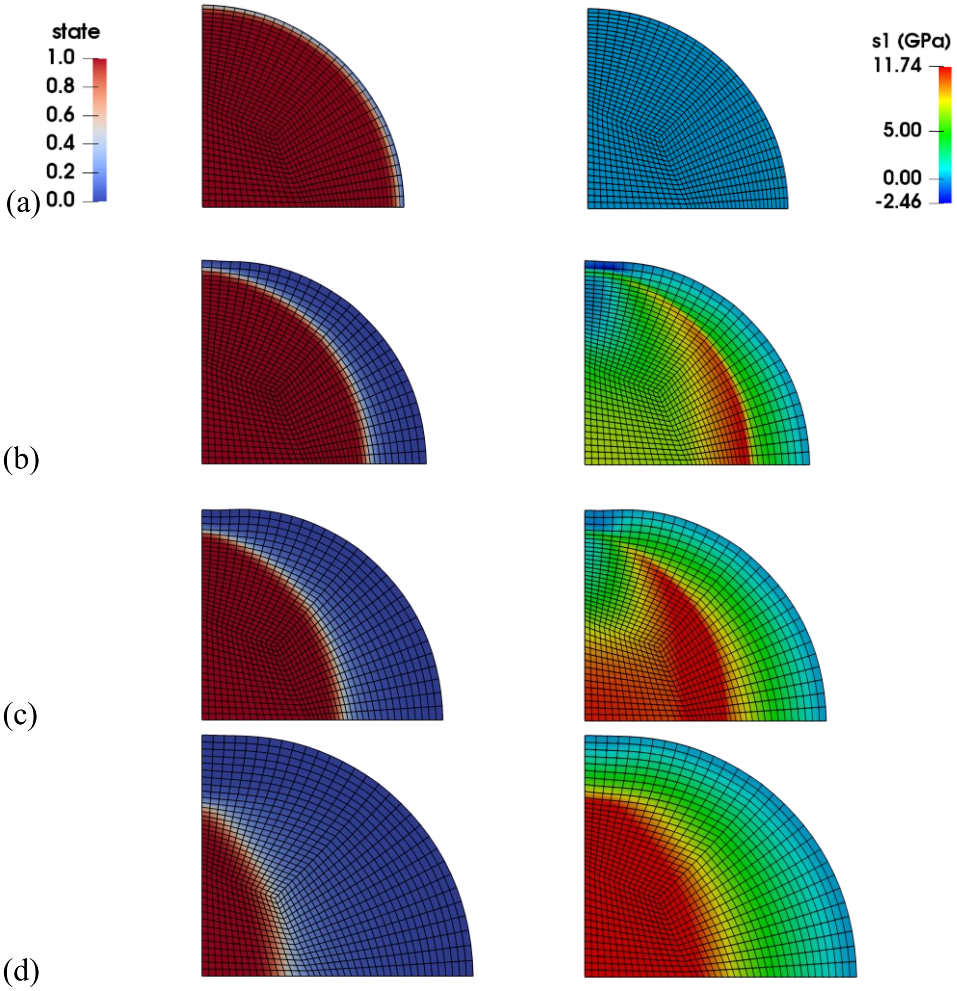


Figure 3.6. Snapshots of the distributions of the reaction state field ϕ and the in-plane principal stress field in the deformed configuration at four instants of time corresponding to (a) $\overline{t} = 0.01$, (b) $\overline{t} = 0.05$, (c) $\overline{t} = 0.1$, and (d) $\overline{t} = 0.3$.

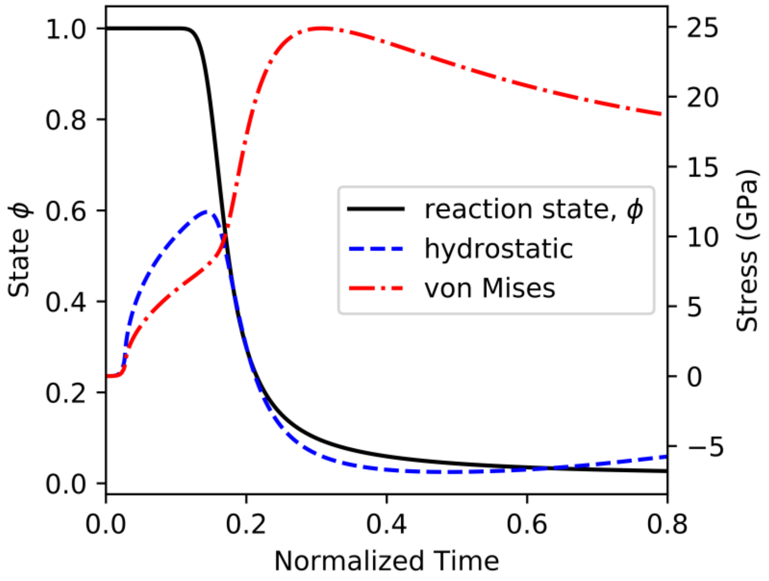


Figure 3.7. Evolution of the reaction state ϕ, the hydrostatic stress, and the von Mises stress at a point located a distance $r = 0.56r_0$ from the origin along the x-axis.

3.7 Conclusions

This chapter has elaborated on the formulation presented in [11] for chemo-mechanical coupling and material evolution of solids permeated by reactive fluids. For an unconstrained body, chemical deformations lead to an evolution of the unstressed configuration relative to which elastic deformations take place. For the model in [11], It is straight-forward to relate some of the material parameters appearing in the constitutive equation for the Cauchy stress of the solid constituent to the standard elastic constants of the pristine material. However, the relationship between the remaining material parameters and standard elastic constants of the fully reacted, chemically swollen solid was not evident. To address this, we first determined the form of stress-free configurations of the unconstrained chemically swollen solid according to the constitutive equation for the Cauchy stress of the solid. Then, using ideas from the seminal work of Truesdell and Noll [17], we investigated the stress response of the chemically swollen solid by considering an infinitesimal deformation superimposed on the finite deformations caused by the chemical reaction. This analysis yields inequality conditions on the material parameters as well as a system of equations that relates the unknown material parameters to the known elastic constants of the fully reacted solid and the observed chemical stretch. Determining the values for the material parameters of the model is crucial in order to perform realistic simulations. This result was employed in the numerical section to determine the material parameters in the model for the thermal oxidation of silicon and lithiation of silicon, based on the reported values in the literature for the first Lamé parameter and shear modulus of silicon oxide (SiO_2), and the Young's modulus and Poisson's ratio of fully lithiated silicon ($Li_{3.75}Si$).

In the present chapter we also introduced reasonable assumptions to yield a reduced model for the fluid constituent. The main simplifying assumptions correspond to the gas-like fluid behavior and its diffusion-dominated transport. These assumptions are reflected in the constitutive equations for the Cauchy stress and interactive force of the fluid, and in the balance of linear momentum for the fluid. The latter goes from a PDE to an algebraic equation that can be solved analytically for the fluid velocity field. The resulting expression for the fluid velocity is substituted in the fluid balance of mass, which yields a scalar diffusion-reaction equation for the fluid mass concentration. It is remarked that the advection-type term canceled automatically, through no additional assumptions. Regarding the solid constituent, this chapter assumed that the body forces and interactive forces due to fluid diffusion are negligible compared to the stress terms. This is a reasonable assumption for gas-like fluid behavior and slow speed of diffusion. Furthermore, it was concluded that the solid balance of mass equation may be disregarded when the solid density field is not a quantity of interest. This is made possible because the solid mass concentration field or its spatial gradient do not appear anywhere else in the model.

The resulting model consists of two governing equations (fluid mass balance and solid balance of linear momentum) plus the constitutive equations for solid stress, fluid mass transfer, and the reaction state variable. One advantage of the reduced model is that its numerical implementation has a reduced computational cost relative to the case when the full fluid model is employed. In 3D space, the numerical implementation in a finite element code has a total of 4 degrees of freedom per node (solid displacement in 3 spatial directions and fluid mass concentration) as well as a single internal variable (the reaction state ϕ). Contrasting it with the case of full fluid model (3 more DOFs associated with the components of fluid velocity, for a total of 7 DOFs), the reduced model results in up to 67% reduction in the computational cost. The system is solved monolithically at each time level through a Newton-Raphson (NR) algorithm. The full consistent tangent yields optimal quadratic rate of convergence for the NR algorithm when the residual vector is a smooth function of the degrees of freedom.

The finite element method for the model is applied to two problems of scientific and engineering interest that exhibit large chemo-mechanically-driven deformations. One is the thermal oxidation of a silicon wafer, and the other is the lithiation of a silicon nanowire. In both cases, the rigorous analysis from section 3.3 made it possible to determine the material properties required by the Cauchy stress constitutive equation based on the standard elastic moduli reported for the pristine and fully transformed solids, as well as the experimentally observed volumetric expansion. The results obtained for the deformation and stress fields are in reasonable agreement with those presented elsewhere in the literature. Discrepancies in the numerical values are in part due to the differences in the models employed for simulation.

3.8 References

[1] G.P. Tandon, K. V. Pochiraju, G.A. Schoeppner, Modeling of oxidative development in PMR-15 resin, Polym. Degrad. Stab. 91 (2006) 1861–1869.

[2] R.S. Hay, Growth stress in SiO 2 during oxidation of SiC fibers, J. Appl. Phys. 111 (2012) 063527.

[3] S. Zhang, Chemomechanical modeling of lithiation-induced failure in high-volume-change electrode materials for lithium ion batteries, Npj Comput. Mater. 3 (2017) 1–10.

[4] N.A. Peppas, J.Z. Hilt, A. Khademhosseini, R. Langer, Hydrogels in biology and medicine: From molecular principles to bionanotechnology, Adv. Mater. 18 (2006) 1345–1360.

[5] D. Ambrosi, G.A. Ateshian, E.M. Arruda, S.C. Cowin, J. Dumais, A. Goriely, G.A. Holzapfel, J.D. Humphrey, R. Kemkemer, E. Kuhl, J.E. Olberding, L.A. Taber, K. Garikipati, Perspectives on biological growth and remodeling, J. Mech. Phys. Solids. 59 (2011) 863–883.

[6] S. Palagi, P. Fischer, Bioinspired microrobots, Nat. Rev. Mater. 3 (2018) 113–124.

[7] S.A. Chester, L. Anand, A thermo-mechanically coupled theory for fluid permeation in elastomeric materials: Application to thermally responsive gels, J. Mech. Phys. Solids. 59 (2011) 1978–2006.

[8] K. Loeffel, L. Anand, Z.M. Gasem, On modeling the oxidation of high-temperature alloys, Acta Mater. 61 (2013) 399–424.

[9] T.J. Truster, A. Masud, A unified mixture formulation for density and volumetric growth

of multi-constituent solids in tissue engineering, Comput. Methods Appl. Mech. Eng. 314 (2017) 222–268.

[10] Y. Zhao, P. Stein, Y. Bai, M. Al-Siraj, Y. Yang, B.-X. Xu, A review on modeling of electro-chemo-mechanics in lithium-ion batteries, J. Power Sources. 413 (2019) 259–283.

[11] M. Anguiano, H. Gajendran, R.B. Hall, K.R. Rajagopal, A. Masud, Chemo-mechanical coupling and material evolution in finitely deforming solids with advancing fronts of reactive fluids, Acta Mech. 231 (2020) 1933–1961.

[12] K.R. Rajagopal, Diffusion through polymeric solids undergoing large deformations, Mater. Sci. Technol. 19 (2003) 1175–1180.

[13] R. Hall, K. Rajagopal, Diffusion of a fluid through an anisotropically chemically reacting thermoelastic body within the context of mixture theory, Math. Mech. Solids. 17 (2012) 131–164.

[14] K.R. Rajagopal, A.R. Srinivasa, On the thermomechanics of materials that have multiple natural configurations Part I: Viscoelasticity and classical plasticity, Zeitschrift Für Angew. Math. Und Phys. ZAMP. 55 (2004) 861–893.

[15] R. Hall, H. Gajendran, A. Masud, Diffusion of chemically reacting fluids through nonlinear elastic solids: mixture model and stabilized methods, Math. Mech. Solids. 20 (2015) 204–227.

[16] H. Gajendran, R.B. Hall, A. Masud, K.R. Rajagopal, Chemo-mechanical coupling in curing and material-interphase evolution in multi-constituent materials, Acta Mech. 229 (2018) 3393–3414.

[17] C. Truesdell, W. Noll, The Non-Linear Field Theories of Mechanics, Third, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[18] J.C. Simo, T.J.R. Hughes, Computational Inelasticity, Springer-Verlag, New York, 1998.

[19] K. Loeffel, L. Anand, A chemo-thermo-mechanically coupled theory for elastic–viscoplastic deformation, diffusion, and volumetric swelling due to a chemical reaction, Int. J. Plast. 27 (2011) 1409–1431.

[20] K. Kajihara, H. Kamioka, M. Hirano, T. Miura, L. Skuja, H. Hosono, Interstitial oxygen molecules in amorphous SiO 2. III. Measurements of dissolution kinetics, diffusion coefficient, and solubility by infrared photoluminescence, J. Appl. Phys. 98 (2005) 013529.

[21] M.T. Heath, Scientific Computing: An Introductory Survey, Second ed., Society for Industrial and Applied Mathematics, Philadelphia, PA, 2018.

[22] V.S. Rao, T.J.R. Hughes, K. Garikipati, On Modelling Thermal Oxidation of Silicon II: Numerical Aspects, Int. J. Numer. Methods Eng. 47 (2000) 359–377.

[23] H. Yang, S. Huang, X. Huang, F. Fan, W. Liang, X.H. Liu, L.-Q. Chen, J.Y. Huang, J. Li, T. Zhu, S. Zhang, Orientation-Dependent Interfacial Mobility Governs the Anisotropic Swelling in Lithiated Silicon Nanowires, Nano Lett. 12 (2012) 1953–1958.

[24] X.H. Liu, H. Zheng, L. Zhong, S. Huang, K. Karki, L.Q. Zhang, Y. Liu, A. Kushima, W.T. Liang, J.W. Wang, J.-H. Cho, E. Epstein, S.A. Dayeh, S.T. Picraux, T. Zhu, J. Li, J.P. Sullivan, J. Cumings, C. Wang, S.X. Mao, Z.Z. Ye, S. Zhang, J.Y. Huang, Anisotropic Swelling and Fracture of Silicon Nanowires during Lithiation, Nano Lett. 11 (2011) 3312–3318.

[25] H. Yang, F. Fan, W. Liang, X. Guo, T. Zhu, S. Zhang, A chemo-mechanical model of lithiation in silicon, J. Mech. Phys. Solids. 70 (2014) 349–361.

CHAPTER 4:
MIXTURE MODEL FOR THERMO-CHEMO-MECHANICAL PROCESSES IN FLUID-INFUSED SOLIDS

4.1 Introduction

Thermo-chemo-mechanical coupled physical phenomena arise in many problems of scientific and engineering interest. One important class of problems arise in the aerospace industry, where thermally resistant metal and ceramic materials are exposed to extreme oxidizing environments. Specifically, gas turbine and rocket engines employ thermal barrier coating systems of superalloys and ceramic matrix composites, the latter of which is also used in vehicle-structure thermal protection during reentry [1]. Some examples of thermal oxidation of aerospace ceramic materials can be found in [2–7]. Thermal oxidation is also a crucial process for the manufacturing of semiconductor materials [8–10]. Other thermo-chemo-mechanical problems of interest include the lithiation and delithiation of electrodes in lithium ion batteries [11,12], processes in solid-oxide fuel cells [13,14], the response of elastomeric gels [15,16] and bio-inspired devices [17], and Ziegler-Natta polymerization [18] among others. In many of these systems the reactive and thermal processes, as well as the internal stresses that develop as a consequence, may lead to large material deformation locally, material degradation, or structural failure at the micrometer scale and above [19–21]. Therefore, it is important to develop accurate continuum-level models that can be descriptive and predictive of the coupled response, but also are amenable to numerical implementation for computer simulation.

Thermodynamically consistent formulations for multifield, nonlinear problems have been the subject of profound study by many authors spanning many years and we refer to seminal works in [22,23] for a historical perspective and a treatise on the subject. Significant contributions have also been made in the past decade to develop fully coupled, continuum models for thermo-chemo-mechanical phenomena that are thermodynamically consistent. Here, a select few are highlighted, and we refer to those and the references therein for a broader perspective on the subject. The work of Kannan and Rajagopal [24] provides a rigorous and detailed presentation on the thermodynamics of chemical reactions in multi-constituent systems. In the present chapter we will consider open systems and relative motion between the constituents. Certain mathematical complexities arise when dealing with the thermodynamics of open systems, and the interested reader is referred to Bulíček et al. [25] for related discussion. A full thermo-chemo-mechanical coupled theory for permeation and finite deformation was presented in [26]. These works were applied to study problems in elastomeric gels [16,27] and oxidation of high-temperature alloys [5,28] through numerical simulation. Other recent work involves thermo-chemo-mechanical model for polymer curing [29], and a chemo-mechanical model for polymer oxidation [30]. A linear theory is developed in [31]. In some of these works, transport of the fluid-like reactive species is assumed to be diffusive at the onset. Here, full equations for the balance of mass and linear momentum of the reactive fluid-like agent are considered. Subsequently, those equations are reduced to the case of a single transient diffusion-reaction equation through appropriate assumptions.

In this chapter, a new, thermodynamically consistent, continuum level model for thermo-chemo-mechanical phenomena is developed within the context of mixture theory, which provides a framework for modeling material systems composed of multiple constituents [32–34]. A key feature of mixture theory is that the classical balance laws for each constituent are augmented by including source-type terms that account for the interaction of each constituent with every other constituent. In this sense, mixture theory can be thought to facilitate physics-based, reduced-order-models for multi-constituent systems since it allows to capture the interactions between constituents at the continuum level, while the geometric intricacies of the interfaces between constituents at the smallest scales need not be explicitly modeled. Ideas from mixture theory have been applied for problems in porous media flows [35], heterogeneous catalysis [36], biological growth [37–39], curing in polymer matrix composites [40], Ziegler-Natta polymerization [18], and diffusion of chemically active fluids within elastic solids [41], among others [42]. In addition, thermodynamically consistent, multiphysics mixture models accounting coupling between chemical, thermal, and/or mechanical effects based on mixture have been presented [43,44], where constitutive relations are derived by assuming maximization of the rate of entropy production. Stabilized numerical methods have also been developed to allow their numerical implementation in a finite element setting [41,45,46].

Although similar ideas regarding the mixture theory setting are employed in this chapter, there are also key differences in the development of the model presented herein. Earlier work accounted for energy exchange due to mass transfer between constituents. Here, we account for the energy associated with breaking and/or forming of bonds as well. Moreover, the solid deformation gradient can be expressed through a multiplicative decomposition into an inelastic and an elastic contribution, which has some advantages in developing functional forms for the constitutive relations. Furthermore, constitutive equations are derived to ensure non-negative entropy production for the open system in order to satisfy the second law of thermodynamics. This is done while accounting for entropy exchange between domain and surroundings through mass and energy exchange. Another key feature in this chapter is the derivation of a reduced formulation (summarized in section 4.7) that is amenable to computationally cost-efficient numerical implementation in a standard Galerkin finite element framework and accurately captures thermo-chemo-mechanical effects.

The outline of this chapter is as follows. Section 4.2 provides an overview of relevant kinematic relations in the context of mixture theory. Section 4.3 addresses the treatment of reaction kinetics. The balance equations are introduced in section 4.4. Thermodynamically consistent constitutive relations are derived in section 4.5. Further modeling assumptions as well as specific forms of the constitutive relations are discussed in section 4.6. Section 4.7 states the resulting system of governing equations as well as initial and boundary conditions. Section 4.8 presents numerical results from the finite element implementation of the model for test cases concerning thermal oxidation of silicon carbide (SiC) in ceramic matrix composites and semiconductor devices, as well as bond coat material FeCrAlY in thermal barrier coating of superalloys. Finally, section 4.9 provides further discussion and conclusions.

4.2 Kinematics

The material systems under consideration are composed of solid and fluid constituents, denoted by s and f respectively. In the context of mixture theory, all constituents are assumed to co-occupy spatial points $\mathbf{x}$ in the current configuration of the domain $\Omega \subset \mathbb{R}^{n_{sd}}$ with boundary $\partial\Omega$ at time t, where $n_{sd} \in \{1,2,3\}$ is the number of spatial dimensions. For any given constituent $\alpha \in \{s,f\}$ there is a corresponding reference configuration Ω_0^α at time $t = t_0$ with material points denoted by $\mathbf{X}^\alpha$, and a one-to-one deformation mapping $\boldsymbol{\chi}^\alpha : \Omega_0^\alpha \to \Omega$ such that

$$\mathbf{x} = \boldsymbol{\chi}^\alpha\left(\mathbf{X}^\alpha, t\right) \tag{4.1}$$

The displacement of constituent α is denoted by $\mathbf{u}^\alpha$ and is defined as

$$\mathbf{u}^\alpha := \boldsymbol{\chi}^\alpha\left(\mathbf{X}^\alpha, t\right) - \mathbf{X}^\alpha \tag{4.2}$$

The mappings (or at least the deformation mapping for the solid constituent) are assumed to be sufficiently smooth so that the deformation gradient of constituent α can be equivalently defined as follows in terms of the deformation mapping or the displacement field:

$$\mathbf{F}^{\alpha} := \frac{\partial \boldsymbol{\chi}^{\alpha}\left(\mathbf{X}^{\alpha}, t\right)}{\partial \mathbf{X}^{\alpha}} = \frac{\partial \mathbf{x}}{\partial \mathbf{X}^{\alpha}} = \mathbf{1} + \frac{\partial \mathbf{u}^{\alpha}}{\partial \mathbf{X}^{\alpha}} \tag{4.3}$$

in addition, the following notation is also introduced

$$J = \det \mathbf{F}^{s} \tag{4.4}$$

The relation between the current and reference configurations, and the deformation maps and of the constituents is illustrated in Figure 4.1.

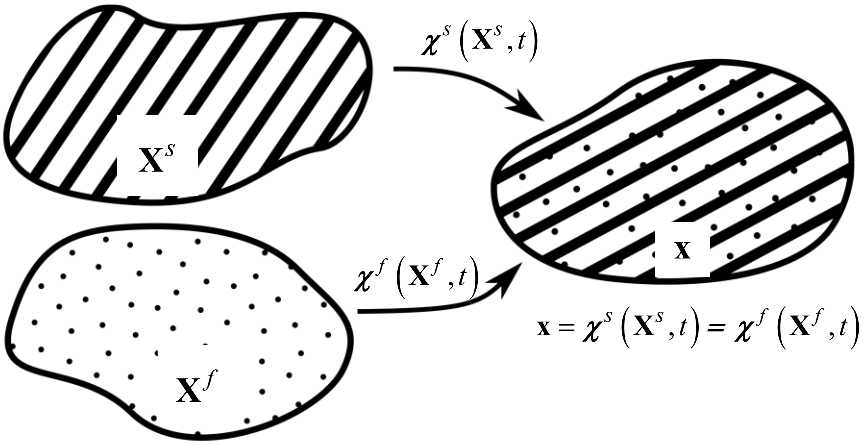

$\chi^{s}\left(\mathbf{X}^{s}, t\right)$
$\mathbf{X}^{s}$
$\mathbf{x}$
$\chi^{f}\left(\mathbf{X}^{f}, t\right)$
$\mathbf{x} = \chi^{s}\left(\mathbf{X}^{s}, t\right) = \chi^{f}\left(\mathbf{X}^{f}, t\right)$
$\mathbf{X}^{f}$

Figure 4.1. Schematic of the relationship between the reference configurations of a solid and a fluid constituent and the current configuration of the mixture.

The overlap of the constituents in the current configuration also leads to the following definition for the mass density (per unit volume in the current configuration) of the mixture

$$\rho := \sum_{\alpha} \rho^{\alpha} \tag{4.5}$$

where ρ^{α} is the mass concentration of constituent α per unit volume in the current configuration

The velocity of constituent α, $\mathbf{v}^{\alpha}$, can be equivalently defined in terms of the deformation mapping or the displacement field as:

$$\mathbf{v}^{\alpha} := \frac{\partial \boldsymbol{\chi}^{\alpha}\left(\mathbf{X}^{\alpha}, t\right)}{\partial t} = \frac{\partial \mathbf{u}^{\alpha}\left(\mathbf{X}^{\alpha}, t\right)}{\partial t} \tag{4.6}$$

In addition, the barycentric mean velocity of the mixture is defined below

$$\mathbf{v} := \frac{1}{\rho}\sum_{\alpha} \rho^{\alpha}\mathbf{v}^{\alpha} \tag{4.7}$$

The relative velocities of the constituents are defined as

$$\mathbf{w}^{\alpha} := \mathbf{v}^{\alpha} - \mathbf{v} \tag{4.8}$$

The spatial gradient of the constituent's velocity is given by:

$$\mathbf{L}^{\alpha} := \nabla \mathbf{v}^{\alpha} \tag{4.9}$$

where $\nabla := \frac{\partial}{\partial \mathbf{x}}$ is the spatial gradient operator (with respect to the current coordinate system).

The deformation of the solid constituent in the class of problems treated in this chapter is assumed to have elastic and inelastic (thermo-chemical) contributions, which are accounted for via a compositional deformation mapping defined as follows

$$\boldsymbol{\chi}^{s}\left(\mathbf{X}^{s},t\right) = \boldsymbol{\chi}^{e}\left(\boldsymbol{\chi}^{i}\left(\mathbf{X}^{s},t\right),t\right) = \boldsymbol{\chi}^{e} \circ \boldsymbol{\chi}^{i} \tag{4.10}$$

As a consequence of the compositional mapping, the deformation gradient can be represented through a multiplicative split, given by

$$\mathbf{F}^{s} = \mathbf{F}^{e}\mathbf{F}^{i} \tag{4.11}$$

The compositional mapping construct leads to the notion of intermediate configurations (intermediate between the reference or undeformed configuration and the current or deformed configuration), and the relationships between these configurations is shown in Figure 4.2, which is congruent with similar ideas in the case of elasto-plasticity, visco-elasticity, and thermo-mechanics. For a detailed discussion on intermediate or natural configurations, the interested reader is referred to the work of Rajagopal and Srinivasa [47].

The material time derivative for constituent α is the total time derivative of a quantity $a = \bar{a}(\mathbf{X}^{\alpha},t) = \hat{a}(\mathbf{x},t)$ from the point of view of a particle moving with constituent α and is the following

$$\frac{D^{\alpha}a}{Dt} = \frac{\partial \bar{a}(\mathbf{X}^{\alpha},t)}{\partial t} = \frac{\partial \hat{a}(\mathbf{x},t)}{\partial t} + \mathbf{v}^{\alpha} \cdot \nabla \hat{a}(\mathbf{x},t) \tag{4.12}$$

In particular, the solid rate of the deformation gradient is the following, using equations (4.3)-(4.9) and (4.12)

$$\frac{D^{s}\mathbf{F}^{s}}{Dt} = \mathbf{L}^{s}\mathbf{F}^{s} \tag{4.13}$$

The total time derivative of a quantity from the point of view of a particle moving with the mixture average velocity is

$$\frac{Da}{Dt} = \frac{\partial \hat{a}(\mathbf{x},t)}{\partial t} + \mathbf{v} \cdot \nabla \hat{a}(\mathbf{x},t) \tag{4.14}$$

The following useful identity arises by combining equations (4.7), (4.12), and (4.14)

$$\frac{Da}{Dt} = \frac{D^s a}{Dt} + \frac{\rho^f}{\rho}(\mathbf{v}^f - \mathbf{v}^s) \cdot \nabla a \tag{4.15}$$

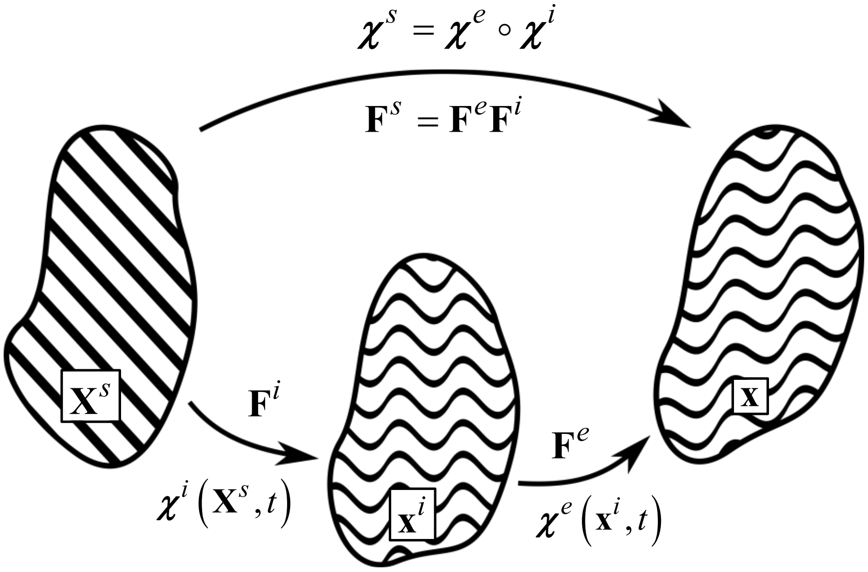

$\chi^s = \chi^e \circ \chi^i$
$\mathbf{F}^s = \mathbf{F}^e \mathbf{F}^i$
$\mathbf{X}^s$
$\mathbf{F}^i$
$\chi^i\left(\mathbf{X}^s, t\right)$
$\mathbf{x}^i$
$\mathbf{F}^e$
$\chi^e\left(\mathbf{x}^i, t\right)$
$\mathbf{x}$

Figure 4.2. Schematic the relationship between the undeformed, deformed, and intermediate configurations of the solid.

Finally, it is assumed that all constituents in the mixture are in thermal equilibrium with each other, i.e., they all have the same temperature field $\theta(\mathbf{x},t)$. Although it is possible to formulate a theory for mixtures where constituents each have a separate temperature field θ^α, assuming a single temperature field allows to simplify the development of the theory, and it is a reasonable assumption for problems where the constituents are well intermixed at length scales below that of a representative volume element.

4.2.1 Arbitrary Lagrangian-Eulerian formulation

For the class of problems of interest, the fluid constituent moves within the solid constituent. This relative motion is tracked using the Arbitrary Lagrangian-Eulerian framework, where the motion of the reference domain is given by the response of the solid constituent. Therefore, the solid reference coordinates are used as the independent coordinates, and the total time derivatives from the point of view of a particle traveling with the fluid are modified according to the ALE framework

$$\frac{D^f a}{Dt} = \frac{\partial a}{\partial t} + \left(\mathbf{v}^f - \mathbf{v}^s\right) \cdot \nabla a \tag{4.16}$$

On the other hand, this results in a Lagrangian formulation from the perspective of the solid constituent, and the total material derive from the point of view of a particle traveling with the solid reduces to a partial derivative

$$\frac{D^s a}{Dt} = \frac{\partial a}{\partial t} \tag{4.17}$$

4.3 Reaction Preliminaries

The class of problems considered in this chapter consist of a solid-fluid system where an unreacted solid species, denoted by S_1, comes into contact with a chemically reacting fluid denoted by F_1, to transform the solid into species S_2, possibly generating fluid byproducts of reaction denoted by F_2, according to the chemical equation[1].

$$aS_1 + bF_1 \to cS_2 + dF_2 \tag{4.18}$$

where *a, b, c*, and *d* are positive stoichiometric coefficients that balance the equation. Furthermore, we consider that the fluid species may enter or leave through the boundaries of the domain (but not the solid species).

First, we introduce some standard notation. The stoichiometric equation of a single chemical reaction involving N species $\mathcal{F}_i$, $i = 1, \ldots, N$ can be written in the form

$$\sum_{i=1}^{N} v_i \mathcal{F}_i = 0 \tag{4.19}$$

where v_i is the stoichiometric coefficient for the ith species, and it is negative for reactants and positive for products of reaction. The corresponding balance of mass during reaction is of the form

$$\sum_{i=1}^{N} v_i \mathcal{M}_i = 0 \tag{4.20}$$

---

[1] One may assume that the chemical reaction can proceed to completion, without the reverse reaction taking place. This can also be interpreted as a system of forward and backward reactions with very large equilibrium constant.

where $\mathcal{M}_i$ is the molar mass of the i th species. We note that equation (4.18) can be written in the form of equation (4.19) as

$$v_{S_1}S_1 + v_{F_1}F_1 + v_{S_2}S_2 + v_{F_2}F_2 = 0 \tag{4.21}$$

where $v_{S_1} = -a$, $v_{F_1} = -b$, $v_{S_2} = c$, and $v_{F_2} = d$.

In order to properly account for the effects of the chemical reaction, it is important to work with a measure of the amount of a species for which its rate of change is due to chemical reaction exclusively. We note that the species concentration (per unit volume in the current configuration) is not a suitable choice for material systems undergoing volume changes. This is because its rate of change has contributions from the rate of volume change even when reaction is not taking place. Another candidate is the specific concentration per unit mass of the mixture. As discussed in [24] for the case of closed systems with no relative motion of its constituents, the rate of change of this quantity is indeed only due to chemical reactions. In this chapter, we use the molar concentration per unit volume in a fixed reference configuration, denoted by $[\mathcal{F}_i]_R$ for species $\mathcal{F}_i$ whose rate of change is also such that it has contributions only from the chemical reaction. We choose the reference volume to be the volume in the reference configuration of the solid constituent Ω_0^s, therefore

$$[\mathcal{F}_i]_R = J[\mathcal{F}_i] \tag{4.22}$$

where $[\mathcal{F}_i]$ is the molar concentration of species $\mathcal{F}_i$ per unit volume in the current configuration.

Although we consider open systems in this chapter, we first recall that for closed systems the rate of reaction can be defined in terms of the rate of change in concentration (per unit reference volume) of any one of the chemical species in the reaction

$$\mathcal{R} := \frac{1}{v_i}\frac{d[\mathcal{F}_i]_R}{dt} \tag{4.23}$$

It is common to quantify the progress of the chemical reaction through the extent-of-reaction, ξ, also defined equivalently, for closed systems, in terms of any of the species involved in the chemical reaction.

$$\xi := \frac{[\mathcal{F}_i]_R - [\mathcal{F}_i]_{R,0}}{v_i} \tag{4.24}$$

where $\left[\mathcal{F}_i\right]_{R,0}$ is the given initial molar concentration (per unit reference volume) of species $\mathcal{F}_i$. It follows that:

$$\dot{\xi} = \mathcal{R} \tag{4.25}$$

In the case of open systems, one may no longer arbitrarily define the extent of reaction or the rate of reaction in terms of any of the species in the chemical reaction, since the changes in the amount of a species may be due to exchange of mass with the surroundings and not due to chemical reaction alone. Nevertheless, for the class of problems considered in this chapter, one can still define the extent of reaction and rate of reaction in terms of either the solid reactant or the solid product of reaction. This is because we consider problems where there is no solid mass exchange between the system and its surroundings and therefore, changes in their molar concentration (per unit reference volume) are still due to chemical reaction alone. Moreover, it is convenient to define a different (but equivalent) variable to track the reaction, the reaction state ϕ, to be bounded by $0 \le \phi \le 1$, which we define in terms of the concentration of the solid reactant S_1 as

$$\phi := \frac{\left[S_1\right]_R}{\left[S_1\right]_{R,0}} \tag{4.26}$$

Then, from the definition of the extent of reaction in equation (4.24) and equation (4.25), it follows that

$$\xi = \frac{\left[S_1\right]_{R,0}}{v_{S_1}}\left(\phi - 1\right) \tag{4.27}$$

$$\frac{D^s \phi}{Dt} = \frac{v_{S_1}}{\left[S_1\right]_{R,0}} \mathcal{R} \tag{4.28}$$

4.4 Balance Equations

The balance or conservation equations can be formulated for each constituent, from which the corresponding balance equations for the mixture as a whole can be obtained [33,43]. Equations (4.29) and (4.30) below are the constituent-wise balance of mass and balance of linear momentum in local form (per unit volume in the current configuration), while equation (4.31) is the local form of the balance of energy for the mixture as a whole, which in this case has been augmented with a term that accounts to the change in energy associated with the breaking and formation of bonds due to chemical reaction.

$$\frac{D^\alpha \rho^\alpha}{Dt} + \rho^\alpha \operatorname{div}(\mathbf{v}^\alpha) - m^\alpha = 0 \qquad \text{in } \Omega\times]0,\mathrm{T}[\tag{4.29}$$

$$\rho^\alpha \frac{D^\alpha \mathbf{v}^\alpha}{Dt} - \operatorname{div}(\mathbf{T}^\alpha) - \rho^\alpha \mathbf{b}^\alpha - \mathbf{I}^\alpha = \mathbf{0} \qquad \text{in } \Omega\times]0,\mathrm{T}[\tag{4.30}$$

$$\begin{aligned} &\rho \frac{D\varepsilon}{Dt} + \sum_\alpha \operatorname{div}\left(\rho^\alpha \varepsilon^\alpha \mathbf{w}^\alpha\right) + \operatorname{div}\mathbf{q} - \rho r - \sum_\alpha \operatorname{tr}\left(\mathbf{T}^\alpha \mathbf{L}^\alpha\right) \\ &\qquad + \frac{1}{2} m^f \left(\mathbf{v}^f - \mathbf{v}^s\right)\cdot\left(\mathbf{v}^f - \mathbf{v}^s\right) + \mathbf{I}^f \cdot \left(\mathbf{v}^f - \mathbf{v}^s\right) - c_\phi \frac{D^s \phi}{Dt} = 0 \qquad \text{in } \Omega\times]0,\mathrm{T}[\end{aligned} \tag{4.31}$$

where for constituent $\alpha \in \{s, f\}$, ρ^α is the mass concentration field (per unit volume in the current configuration), $\mathbf{v}^\alpha$ is the constituent velocity field, m^α is the net mass transfer rate to constituent α from all other constituents, $\mathbf{T}^\alpha$ is the Cauchy stress tensor, $\mathbf{b}^\alpha$ is the body force exerted per unit mass on constituent α, $\mathbf{I}^\alpha$ is the net interactive force (momentum transfer) on constituent α due to interactions with other constituents, ε^α is the specific internal energy (per unit mass), $\mathbf{w}^\alpha$ is the relative velocity defined in equation (4.8), and $\mathbf{L}^\alpha$ is the velocity gradient. In addition, ρ is the mixture mass concentration, $\mathbf{q}$ is the heat flux vector for the mixture, r is the radiant heat supply, and $\varepsilon := \rho^{-1} \sum_\alpha \rho^\alpha \varepsilon^\alpha$ is the specific internal energy of the mixture (per unit mass), ϕ is the ratio of current versus initial molar concentration of (unreacted) solid substrate in the solid reference configuration, and c_ϕ is the corresponding conjugate quantity, so that the last term in equation (4.31) represents a "chemical power" that accounts for the change of internal energy of the mixture due to the breaking and formation of chemical bonds. Further details of the derivations are included in Appendix A. For completeness, it is noted that the balance of angular momentum can be satisfied by providing a symmetric Cauchy stress tensor for each constituent, and satisfaction of the second law of thermodynamics has further implications on the form of the constitutive relations discussed in section 4.5.

4.4.1 Transfer of conserved quantities

The exchange of mass, momentum, and energy between constituents is assumed to be conservative. In other words, there is no spontaneous generation (or dissipation) of mass, momentum, nor energy as those quantities are transferred from one or multiple constituents to another. Therefore, the following is assumed to hold and is used in the derivation of the balance of energy and entropy inequality for the mixture as a whole

$$\sum_{\alpha} m^{\alpha} = 0 \tag{4.32}$$

$$\sum_{\alpha} \left(\mathbf{I}^{\alpha} + m^{\alpha} \mathbf{v}^{\alpha} \right) = 0 \tag{4.33}$$

$$\sum_{\alpha} \left(\varepsilon_S^{\alpha} + \mathbf{I}^{\alpha} \cdot \mathbf{v}^{\alpha} + m^{\alpha} \left(\varepsilon^{\alpha} + \frac{1}{2} \mathbf{v}^{\alpha} \cdot \mathbf{v}^{\alpha} \right) \right) = 0 \tag{4.34}$$

Equation (4.32) corresponds to mass conservation in the transfer of mass between constituents (i.e., there is no spontaneous mass generation or loss within the mixture); equation (4.33) corresponds to conservation of linear momentum in the interactions between constituents, and (4.34), to conservation of energy within the mixture as energy is transferred between constituents, where ε_S^{α} is the energy supplied to constituent α from every other constituent in the mixture.

4.4.2 Relation between m^{α} and reaction rate $\mathcal{R}$

The rate of mass transfer to constituent α, m^{α} is the amount of mass of constituent α that is generated or consumed by the chemical reaction per unit volume in the current configuration. Therefore, it is related to the rate of reaction (given in moles of reaction per unit volume in the reference configuration of the solid). If α is a constituent that corresponds to a single species in the chemical reaction, the relation is given by

$$m^{\alpha} = \frac{1}{\det \mathbf{F}^{s}} \nu_{\alpha} \mathcal{M}_{\alpha} \mathcal{R} \tag{4.35}$$

However, the solid constituent is associated with both the solid reactant and the solid product. In this case, it can be shown that the following relation holds:

$$m^{s} = \frac{1}{\det \mathbf{F}^{s}} \left(\nu_{S_1} \mathcal{M}_{S_1} + \nu_{S_2} \mathcal{M}_{S_2} \right) \mathcal{R} \tag{4.36}$$

By combining (4.35) and (4.36), it is noted that it is possible to express the mass transfer rate of the solid in terms of the mass transfer rate of the fluid reactant

$$m^{s} = r_s m^{f} \tag{4.37}$$

with

$$r_s = \frac{v_{S_1}\mathcal{M}_{S_1} + v_{S_2}\mathcal{M}_{S_2}}{v_f \mathcal{M}_f} \tag{4.38}$$

For a reaction with no fluid byproduct, $r_s = 1$ by equation (4.20).

Recalling that the rate of reaction $\mathcal{R}$ can be given in terms of the rate of the reaction state variable ϕ, we have that for any constituent α we have a relation of the form

$$m^\alpha = \beta_\phi^\alpha \frac{D^s\phi}{Dt} \tag{4.39}$$

Where $\beta_\phi^\alpha = \frac{1}{\det \mathbf{F}^s} \frac{v_\alpha \mathcal{M}_\alpha [S_1]_{R,0}}{v_{S_1}}$ if constituent α corresponds to a single chemical species, or $\beta_\phi^s = \frac{1}{\det \mathbf{F}^s} \frac{\left(v_{S_1}\mathcal{M}_{S_1} + v_{S_2}\mathcal{M}_{S_2}\right)[S_1]_{R,0}}{v_{S_1}}$ for the solid constituent, which is associated with both the solid reactant and solid product of reaction. In a closed system, with no relative diffusion of the constituents, it is clear that the mass concentration of constituent α is affected by the progress of the chemical reaction, as tracked by the evolution of the reaction state variable ϕ; therefore, an interpretation of the coefficient β_ϕ^α is as the rate of change in mass concentration of constituent α with respect to the change in reaction state ϕ

$$\beta_\phi^\alpha \equiv \frac{\partial \rho^\alpha}{\partial \phi} \tag{4.40}$$

4.5 Constitutive Equations

The second law of thermodynamics is given by the entropy inequality for the mixture as a whole in local form (see [33,43])

$$\rho \frac{D\eta}{Dt} + \sum_\alpha \operatorname{div}\left(\rho^\alpha \eta^\alpha \mathbf{w}^\alpha\right) + \frac{1}{\theta}\left(\operatorname{div}\mathbf{q} - \rho r\right) - \frac{\mathbf{q}\cdot\nabla\theta}{\theta^2} \geq 0 \qquad \text{in } \Omega\times]0,\mathrm{T}[\tag{4.41}$$

Where η^α is the specific entropy (per unit mass) of constituent α, and $\eta := \rho^{-1}\sum_\alpha \rho^\alpha \eta^\alpha$ is the specific entropy of the mixture as a whole. It is noted that the second term, $\sum_\alpha \operatorname{div}\left(\rho^\alpha \eta^\alpha \mathbf{w}^\alpha\right)$ is associated with entropy entering or leaving the open system through the boundary due to mass transport into or out of the domain. We assume the existence of a specific Helmholtz free energy potential[2] defined through the following Legendre transform of the internal energy

$$\psi = \varepsilon - \theta\eta \tag{4.42}$$

Which is related to the constituents' specific Helmholtz free energies ψ^α via

$$\psi = \frac{1}{\rho}\sum_\alpha \psi^\alpha \tag{4.43}$$

Using equation (4.42), and the relations of the mixture thermodynamic potentials to the constituents' thermodynamic potentials, equation (4.41) can be recast into the following form (details of the derivation are included in Appendix A)

$$\begin{aligned}&\rho\left(-\frac{D\psi}{Dt}-\eta\frac{D\theta}{Dt}\right)-\sum_\alpha \operatorname{div}\left(\rho^\alpha\psi^\alpha\mathbf{w}^\alpha\right)-\nabla\theta\cdot\left(\sum_\alpha\rho^\alpha\eta^\alpha\mathbf{w}^\alpha\right)\\&\quad+\sum_\alpha \operatorname{tr}\left(\mathbf{T}^\alpha\mathbf{L}^\alpha\right)-\frac{1}{2}m^f\left(\mathbf{v}^f-\mathbf{v}^s\right)\cdot\left(\mathbf{v}^f-\mathbf{v}^s\right)-\mathbf{I}^f\cdot\left(\mathbf{v}^f-\mathbf{v}^s\right)-\frac{\mathbf{q}\cdot\nabla\theta}{\theta} = \mathcal{D} \geq 0\end{aligned} \tag{4.44}$$

where $\mathcal{D}$ is the total dissipation production rate per unit volume in the current configuration.

In general, one may expect the constitutive equations to be functions of the independent variables $\rho^\alpha, \mathbf{F}^\alpha, \theta, \phi$ as well as their gradients and rates $\nabla\rho^\alpha, \nabla\mathbf{F}^\alpha, \nabla\theta, \nabla\phi, \dot{\rho}^\alpha, \mathbf{L}^\alpha, \dot{\theta}, \dot{\phi}$. Moreover, interactive mechanism between constituents may depend on their relative velocities $\mathbf{v}^\alpha - \mathbf{v}^\beta$. In addition, one could also consider dependence on a properly frame invariant form of the relative acceleration as well as higher spatial gradients of $\mathbf{F}^\alpha$ and $\mathbf{L}^\alpha$ [33]. In what follows, we will assume that the functional forms of the constitutive equations depend only on a subset of this list of possible arguments. The dependence of the specific Helmholtz free energies (per unit mass) on the state variables is assumed to be of the following form

$$\psi = \psi\left(\rho^f, \rho^s, \mathbf{F}^s, \phi, \theta\right) \tag{4.45}$$

$$\psi^f = \psi^f(\rho^f, \theta) \tag{4.46}$$

$$\psi^s = \psi^s\left(\rho^s, \mathbf{F}^s, \phi, \theta\right) = (\rho^s J)^{-1}\hat{\psi}^s(\mathbf{F}^s, \theta, \phi) \tag{4.47}$$

where $\hat{\psi}^s$ is the Helmholtz free energy of the solid per unit volume in the reference configuration, which is a convenient choice for developing constitutive equations for the solid constituent. Using chain rule to take mixture total time derivative and gradient of the Helmholtz free energy, equation (4.39), as well as other identities regarding mass concentration rate and gradients of the relative velocities, and grouping terms, equation (4.44) becomes (see details in Appendix A)

---

[2] Rajagopal and Srinivasa [63] point out that one may not always use a Legendre transform to obtain the Gibbs free energy from the Helmholtz free energy. It is further argued that for some systems only one of them may be defined.

$$
\begin{aligned}
&\operatorname{tr}\left\{\left[\mathbf{T}^s - J^{-1}\mathbf{F}^s\left(\frac{\partial\hat{\psi}^s}{\partial\mathbf{F}^s}\right)^T\right]\mathbf{L}^s\right\} + \operatorname{tr}\left\{\left[\mathbf{T}^f + (\rho^f)^2\frac{\partial\psi^f}{\partial\rho^f}\mathbf{1}\right]\mathbf{L}^f\right\} - \rho\left(\eta + \frac{\partial\psi}{\partial\theta}\right)\frac{D\theta}{Dt} \\
&\quad - \left[\frac{\mathbf{q}}{\theta} + \frac{\rho^f\rho^s}{\rho}\left(\eta^f - \eta^s + \frac{\partial\psi^f}{\partial\theta} - \frac{\partial\psi^s}{\partial\theta}\right)\left(\mathbf{v}^f - \mathbf{v}^s\right)\right]\cdot\nabla\theta \\
&\quad + \left[-\mathbf{I}^f - \frac{1}{2}m^f\left(\mathbf{v}^f - \mathbf{v}^s\right)\right]\cdot\left(\mathbf{v}^f - \mathbf{v}^s\right) \\
&\quad + \left[-\beta_\phi\left(\psi^f + \rho^f\frac{\partial\psi^f}{\partial\rho^f}\right) - J^{-1}\frac{\partial\hat{\psi}^s}{\partial\phi} + c_\phi\right]\frac{D^s\phi}{Dt} = \mathcal{D} \geq 0
\end{aligned}
\tag{4.48}
$$

We ensure non-negative entropy production for arbitrary processes by expressing the dependent quantities $\{\mathbf{T}^s,\ \mathbf{T}^f,\ \eta,\ \mathbf{q},\ \mathbf{I}^f,\ c_\phi\}$ as the sum of a contribution that automatically adds up to zero (denoted by superscript (q) below), and a contribution that guarantees non-negative dissipation (denoted by superscript (d) below). Furthermore, we assume that there is no dissipation associated with the temperature rate of change or with the solid velocity gradient (i.e., neglecting viscous effects). Then, the following general constitutive equations are obtained

$$\mathbf{T}^f = \mathbf{T}^{f,(q)} + \mathbf{T}^{f,(d)} \tag{4.49}$$

$$c_\phi = c_\phi^{(q)} + c_\phi^{(d)} \tag{4.50}$$

$$\mathbf{I}^f = \mathbf{I}^{f,(q)} + \mathbf{I}^{f,(d)} \tag{4.51}$$

$$\mathbf{T}^s = \mathbf{T}^{s,(q)} = J^{-1}\mathbf{F}^s\left(\frac{\partial\hat{\psi}^s}{\partial\mathbf{F}^s}\right)^T \tag{4.52}$$

$$\eta = \eta^{(q)} = -\frac{\partial\psi}{\partial\theta} \tag{4.53}$$

$$\eta^\alpha = -\frac{\partial\psi^\alpha}{\partial\theta} \tag{4.54}$$

$$\mathbf{q} = \mathbf{q}^{(d)} = -\boldsymbol{\kappa}\nabla\theta \tag{4.55}$$

$$\mathbf{T}^{f,(q)} = -(\rho^f)^2\frac{\partial\psi^f}{\partial\rho^f}\mathbf{1} \tag{4.56}$$

$$\mathbf{T}^{f,(d)} = \mu^f \ \mathrm{sym}\mathbf{L}^f \tag{4.57}$$

$$c_\phi^{(q)} = \beta_\phi \left(\psi^f + \rho^f \frac{\partial \psi^f}{\partial \rho^f} \right) + J^{-1} \frac{\partial \hat{\psi}^s}{\partial \phi} \tag{4.58}$$

$$c_\phi^{(d)} = \overline{c}_\phi \, \mathrm{sign}\left(\frac{D^s \phi}{Dt} \right) \tag{4.59}$$

$$\mathbf{I}^{f,(q)} = -\frac{1}{2} m^f \left(\mathbf{v}^f - \mathbf{v}^s \right) \tag{4.60}$$

$$\mathbf{I}^{f,(d)} = -\mathbf{A}^v (\mathbf{v}^f - \mathbf{v}^s) \tag{4.61}$$

where the tensor of heat conductivities $\boldsymbol{\kappa}$ and the drag tensor $\mathbf{A}^v$ are positive definite, and together with the fluid viscosity $\mu^f > 0$ and $\overline{c}_\phi > 0$, are material parameters that in general could depend on the state variables as well.

Remark: Equation (4.53) and the relations between the thermodynamics potentials of the mixture and the constituents lead to identity (4.54), causing all but the $\mathbf{q}$ term to vanish in the coefficient of $\nabla \theta$ from (4.48). As a result, $\mathbf{q}$ has no quasi-conservative contribution and is a purely dissipative quantity. Relaxing the assumption that η have a purely quasi-conservative contribution leads to an additional temperature rate contribution to $\mathbf{q}$ and gives rise to hyperbolic-type heat conduction.

Remark: The current formulation is capable of accommodating the effects of viscoelasticity and plasticity in the solid constituent.

As an alternative to (4.59), the physical meaning of $c_\phi^{(d)}$ can be interpreted based on the following constraint for satisfaction of (4.48)

$$c_\phi^{(d)} \frac{D^s \phi}{Dt} \geq 0 \tag{4.62}$$

Recall that the sign of $D^s \phi / Dt$ corresponds to the instantaneous direction in which the chemical reaction from (4.18) progresses (negative for the forward reaction, and positive for the reverse reaction, as per the definition of ϕ). As a consequence, the role of $c_\phi^{(d)}$ in (4.62) is consistent with the concept of change in Gibbs energy of reaction, or chemical affinity, (per unit volume in the current configuration, per unit change in ϕ), whose sign determines the direction in which a given chemical reaction proceeds, i.e., whether a given chemical reaction is thermodynamically favored. Therefore, we establish the relation

$$c_{\phi}^{(d)} \equiv -\frac{\partial(\rho g)}{\partial \phi} \tag{4.63}$$

where g is the specific Gibbs energy potential (per unit mass of the mixture). A negative sign arises because the Gibbs free energy of the products is less than that of the reactants when the forward reaction is thermodynamically favored. In that case, the change in Gibbs free energy of reaction is negative, and the change in ϕ is also negative for the forward reaction, then $\partial(\rho g)/\partial\phi > 0$. But then, $c_{\phi}^{(d)}$ must be negative in order to satisfy (4.62).

For further discussion on De Donder affinity see [24]. Assuming a Gibbs energy potential can be defined for this system, one could adopt a Gibbs energy formulation where the Gibbs-Duhem relation would follow from the assumption of the Gibbs potential being a homogeneous function of degree one on the amount of chemical species and Euler's theorem for homogeneous functions (see [24]). Also see [36] for a similar relation in a Helmholtz formulation, which is a direct consequence of the implications of the Gibbs-Duhem relation (assuming both Gibbs and Helmholtz potentials can be defined through a Legendre transform).

4.6 Modeling Assumptions

In this section, further modeling assumptions are introduced with the intent to bring the general thermo-chemo-mechanical theory presented in the preceding sections into a specialized form for application to a certain class of problems of scientific and engineering interest. Specialization of the equations to a specific material system allows numerical implementation of the model in order to test its descriptive and predictive capabilities through computer simulation. In developing the specific forms of the constitutive equations, one must consider the principles of observer invariance[3] and material symmetries [23]. In what follows, we directly present relations in terms of frame-invariant measures of independent state variable (e.g., solid stretch). Furthermore, the model is specialized to the case of isotropic material behavior for ease of presentation. Nevertheless, one may develop constitutive equations for anisotropic materials along similar lines, while ensuring they satisfy material symmetry invariance. Other assumptions made in this section relate to negligible inertial effects, gas-like behavior of the fluid constituent, and relatively low (or dilute) mass concentration of the fluid within the solid, among others, which are reasonable approximations in the class of problems analyzed in section 4.8. These further idealizations have mathematical implications and are reflected on the final form of the governing equations through the derivations that follow.

---

[3] It is common to require the constitutive relations to be invariant under a change of frame (so called material frame indifference or objectivity), but as touched upon in [24], there is some controversy around this issue, and Galilean frame invariance may suffice in some cases.

4.6.1 Solid deformation

The solid inertial terms, body forces, and interactive forces are taken as negligible and dropped from equation (4.30). In addition, the inelastic deformation consists only of thermochemical contributions, which are assumed to be isotropic, and results in the following form of the deformation gradient

$$\mathbf{F}^i = \mathbf{F}^{tc} = \lambda^{tc}(\phi,\theta)\mathbf{1} \tag{4.64}$$

Thus, the total form of the deformation gradient is

$$\mathbf{F}^s = \lambda^{tc}\mathbf{F}^e \tag{4.65}$$

And the elastic deformation gradient is

$$\mathbf{F}^e = \left(\lambda^{tc}\right)^{-1}\mathbf{F}^s \tag{4.66}$$

where specific forms for the thermo-chemical stretch is defined as

$$\lambda^{tc}(\phi,\theta) = 1 + \alpha_\theta^s(\phi)\left[\theta - \theta_0\right] + \alpha_\phi^s\left[1 - \phi\right] \tag{4.67}$$

and the dependence of the coefficient of thermal expansion on the reaction state is given by

$$\alpha_\theta^s(\phi) = \bar{\alpha}_\theta^{s1}\phi + \bar{\alpha}_\theta^{s2}\left(1 - \phi\right) \tag{4.68}$$

where $\bar{\alpha}_\theta^{s1}$ and $\bar{\alpha}_\theta^{s2}$ are the coefficients of thermal expansion for the unreacted and reacted solid, respectively, at the reference temperature θ_0.

4.6.2 Temperature form of the balance of energy

Consider the mixture balance of energy in equation (4.31) and combining it with the dissipation equation in (4.44), the balance of energy is rewritten as

$$
\begin{aligned}
&\rho\frac{D\varepsilon}{Dt}+\operatorname{div}\left(\sum\rho^{\alpha}\varepsilon^{\alpha}\mathbf{w}^{\alpha}\right)+\operatorname{div}\mathbf{q}-\rho r-\mathcal{D}-\frac{\nabla\theta\cdot\mathbf{q}}{\theta}\\
&+\rho\left(-\frac{D\psi}{Dt}-\eta\frac{D\theta}{Dt}\right)-\operatorname{div}\left(\sum\rho^{\alpha}\psi^{\alpha}\mathbf{w}^{\alpha}\right)-\nabla\theta\cdot\sum\rho^{\alpha}\eta^{\alpha}\mathbf{w}^{\alpha}=0
\end{aligned}
\tag{4.69}
$$

Then, using the identities between the thermodynamic potentials to write the mixture internal energy in terms of the Helmholtz free energy and entropy, and applying chain rule and the constitutive equations, the balance of energy becomes the following (see Appendix A for details)

$$
\begin{aligned}
&-\rho^{s}\theta\frac{\partial^{2}\psi^{s}}{\partial\theta^{2}}\frac{D^{s}\theta}{Dt}-\rho^{f}\theta\frac{\partial^{2}\psi^{f}}{\partial\theta^{2}}\frac{D^{f}\theta}{Dt}+\operatorname{div}\mathbf{q}-\rho r+\mathbf{I}^{f,(d)}\cdot\left(\mathbf{v}^{f}-\mathbf{v}^{s}\right)\\
&-\operatorname{tr}\left\{\theta\frac{\partial\mathbf{T}^{s,(q)}}{\partial\theta}\mathbf{L}^{s}\right\}-\operatorname{tr}\left\{\left(\theta\frac{\partial\mathbf{T}^{f,(q)}}{\partial\theta}+\mathbf{T}^{f,(d)}\right)\mathbf{L}^{f}\right\}-\left(\theta\frac{\partial c_{\phi}^{(q)}}{\partial\theta}+c_{\phi}^{(d)}\right)\frac{D^{s}\phi}{Dt}\quad=\quad 0
\end{aligned}
\tag{4.70}
$$

For the class of problems of interest, it is assumed that the fluid heating terms are negligible compared with the solid heat capacity, heat conduction, and chemical heating terms. In addition, the elastic heating is also neglected, and no external radiant heat supply or sink is considered. Therefore, the energy balance equation in temperature form reduces to

$$
-\rho^{s}\theta\frac{\partial^{2}\psi^{s}}{\partial\theta^{2}}\frac{D^{s}\theta}{Dt}-\left(\theta\frac{\partial c_{\phi}^{(q)}}{\partial\theta}+c_{\phi}^{(d)}\right)\frac{D^{s}\phi}{Dt}+\operatorname{div}\mathbf{q}\quad=\quad 0
\tag{4.71}
$$

4.6.3 Enthalpy of reaction

Recalling (4.54) and (4.58), the following identity is obtained

$$
\frac{\partial c_{\phi}^{(q)}}{\partial\theta}\quad=\quad-\beta_{\phi}\frac{\partial\left(\rho^{f}\eta^{f}\right)}{\partial\rho^{f}}-\frac{\partial\left(\rho^{s}\eta^{s}\right)}{\partial\phi}
\tag{4.72}
$$

Recall from (4.40) that $\beta_{\phi}\equiv\partial\rho^{f}/\partial\phi$ may be interpreted as the ratio of the change in ρ^{f} to the change in ϕ due to chemical reaction, therefore

$$
\frac{\partial c_{\phi}^{(q)}}{\partial\theta}\quad\equiv\quad-\frac{\partial\left(\rho\eta\right)}{\partial\phi}
\tag{4.73}
$$

and consequently, the rate of heat production (per unit volume in the current configuration) due to chemical reaction (i.e., the coefficient of the rate of ϕ in the temperature equation) can be written as follows

$$
\theta\frac{\partial c_{\phi}^{(q)}}{\partial\theta}+c_{\phi}^{(d)}=-\theta\frac{\partial\left(\rho\eta\right)}{\partial\phi}-\frac{\partial\left(\rho g\right)}{\partial\phi}=-\rho\frac{\partial}{\partial\phi}\left(\theta\eta+g\right)=-\frac{\partial\left(\rho h\right)}{\partial\phi}
\tag{4.74}
$$

where h is the specific enthalpy of reaction of the mixture. It is reasonable to expect that the rate of heat release (or absorption) per unit reference volume be constant in ϕ, since one mole of reaction should produce the same amount of heat regardless of how much of the material has previously reacted or is yet to react. Then, the enthalpy term can be taken to be

$$-\frac{\partial(\rho h)}{\partial\phi} = J^{-1}\bar{h}_{\phi} \tag{4.75}$$

where $\bar{h}_{\phi}$ is the change in enthalpy of reaction per unit reference volume, and is, in general, a function of the absolute temperature θ. We take the common relation for the enthalpy of reaction at a given temperature in terms of typically known quantities, such as the molar enthalpies of formation and molar heat capacities of the chemical species as follows:

$$\bar{h}_{\phi}(\theta) = \Delta h_{rxn}\big|_{\theta_{ref}} + \Delta c_p\left(\theta - \theta_{ref}\right) \tag{4.76}$$

and

$$\Delta h_{rxn}\big|_{\theta_{ref}} = \frac{[S_1]_{R,0}}{v_{S_1}}\sum_{i=1}^{N} v_i\, H_{f,i}\big|_{\theta_{ref}} \tag{4.77}$$

$$\Delta c_p = \frac{[S_1]_{R,0}}{v_{S_1}}\sum_{i=1}^{N} v_i\, C_{p,i}\big|_{\theta_{ref}} \tag{4.78}$$

where $H_{f,i}\big|_{\theta_{ref}}$ is the standard enthalpy of formation of species $\mathcal{F}_i$ at some temperature θ_{ref}, v_i is its corresponding stoichiometric coefficient, $[S_1]_{R,0}$ is the initial molar concentration (per unit volume in the reference configuration) of the solid reactant (species S_1), v_{S_1} is its stoichiometric coefficient, and $C_{p,i}$ is the molar heat capacity at constant pressure of species $\mathcal{F}_i$.

4.6.4 Rate of reaction

The chemical affinity, or Gibbs free energy of reaction, dictates if a given forward reaction is thermodynamically favored (or spontaneous), or if it is the reverse reaction that is favored. In other words, it defines the sign of $D^s\phi/Dt$. However, the rate of reaction itself remains undefined. A large class of chemical reactions are described by the following standard elementary reaction kinetics of the rate equation in terms of the concentration of the reactants (a similar form is also used in [36])

$$\mathcal{R} = k(\theta)[S_1]^x[F_1]^y \tag{4.79}$$

where $[\cdot]$ denotes molar concentration per unit volume in the current configuration[4], x and y are exponents that depend on the specific reaction, θ is the temperature, and $k(\theta)$ is a temperature dependent reaction rate coefficient. In general, reaction could also be stress dependent, as discussed in [48]. However, this effect is not considered in this chapter but will be addressed in the future. Temperature dependence is assumed to follow Arrhenius law

$$k(\theta) = k_0 \exp\left[\frac{-E_a}{R}\left(\frac{\theta_0 - \theta}{\theta_0 \theta}\right)\right] \tag{4.80}$$

where k_0 is the rate coefficient at the reference temperature θ_0, E_a is the activation energy for the reaction, and R is the gas constant. It is convenient to write the reaction rate $\mathcal{R}$ in terms of the reaction state variable ϕ and the fluid mass concentration ρ^f. Substituting (4.22), (4.26), (4.80), into (4.79), using the relation $[F_1] = \rho^f / \mathcal{M}_f$, and assuming first-order dependence of the reaction rate on reactant concentrations (i.e., $x = y = 1$ in equation (4.79)), the following expression for the reaction rate is obtained

$$\mathcal{R} = r_0 \exp\left[\frac{-E_a}{R}\left(\frac{\theta_0 - \theta}{\theta_0 \theta}\right)\right] J^{-1}\phi\rho^f \tag{4.81}$$

where

$$r_0 = \frac{[S_1]_{R,0}}{\mathcal{M}_f} k_0 \tag{4.82}$$

Then, substituting equation (4.81) into the equation for the total time derivative of the reaction state $D^s\phi/Dt$ from equation (4.28), one obtains the following

$$\frac{D^s\phi}{Dt} = r_\phi \exp\left[\frac{-E_a}{R}\left(\frac{\theta_0 - \theta}{\theta_0 \theta}\right)\right] J^{-1}\phi\rho^f \tag{4.83}$$

where

---

[4] Although the definition of the reaction rate is in terms of concentrations per unit volume in some reference configuration, its magnitude depends on the concentration per unit volume in its current configuration since this determines the actual likelihood of molecular interactions between reactants.

$$r_\phi = \frac{v_{s_1}}{\mathcal{M}_f} k_0 \tag{4.84}$$

In addition, the rate of mass transfer can be written as

$$m^\alpha = r_m^\alpha \exp\left[\frac{-E_a}{R}\left(\frac{\theta_0 - \theta}{\theta_0 \theta}\right)\right] J^{-2} \phi \rho^f \tag{4.85}$$

where

$$r_m^\alpha = J \beta_\phi^\alpha \frac{v_{s_1}}{\mathcal{M}_f} k_0 \tag{4.86}$$

For the fluid reactant, one has the following

$$m^f = r_m^f \exp\left[\frac{-E_a}{R}\left(\frac{\theta_0 - \theta}{\theta_0 \theta}\right)\right] J^{-2} \phi \rho^f \tag{4.87}$$

$$r_m^f = v_f \left[S_1\right]_{R,0} k_0 \tag{4.88}$$

4.6.5 Specific form of Helmholtz free energy

The fluid will be modeled as an ideal gas, and its corresponding specific Helmholtz free energy is given by

$$\psi^f(\rho^f, \theta) = A + \frac{R\theta}{\mathcal{M}_f}\left[\hat{c}_v - \ln\left(\frac{\theta^{\hat{c}_v}}{\omega \rho^f}\right)\right] \tag{4.89}$$

where R is the gas constant, $\mathcal{M}_f$ is the fluid molar mass (or molecular weight), $\hat{c}_V$ is the non-dimensional heat capacity at constant volume, which is $3/2$ for monoatomic gases and $5/2$ for diatomic gases, and A, ω are parameters that depend on the specific gas and can be determined using the standard entropy of formation and change in enthalpy of formation for the fluid substance.

Assuming that the solid is unstressed and undeformed in its reference configuration (i.e., when $\mathbf{F} = \mathbf{1}$, $\theta = \theta_0$, and $\phi = 1$), the dependence of the solid constituent's specific Helmholtz free energy on the state variables is chosen as follows

$$\psi^s(\rho^s, \mathbf{F}^s, \phi, \theta) = (\rho^s J)^{-1} \hat{\psi}^s\left(\mathbf{F}^s, \theta, \phi\right) \tag{4.90}$$

$$\hat{\psi}^s\left(\mathbf{F}^s, \theta, \phi\right) = \hat{\psi}^s_{thermal}(\theta, \phi) + \hat{\psi}^s_{mech}(\mathbf{F}^e, \phi) \tag{4.91}$$

$$\hat{\psi}^{s}_{thermal}(\theta;\phi) = B + h_0^s(\phi) - \theta\eta_0^s(\phi) + c_0^s(\phi)\left[\theta - \theta_0 - \theta\ln\left(\frac{\theta}{\theta_0}\right)\right] \tag{4.92}$$

$$\hat{\psi}^{s}_{mech}(\mathbf{F}^e;\phi) = \bar{W}(\bar{\mathbf{C}}^e;\phi) + U(J^e;\phi) \tag{4.93}$$

$$\bar{W}(\bar{\mathbf{C}}^e;\phi) = \frac{1}{2}\mu^s(\phi)\left[\mathrm{tr}\bar{\mathbf{C}}^e - 3\right] \tag{4.94}$$

$$U(J^e;\phi) = \frac{1}{2}k^s(\phi)\left[\ln J^e\right]^2 \tag{4.95}$$

$$\bar{\mathbf{C}}^e = (J^e)^{-2/3}(\mathbf{F}^e)^T\mathbf{F}^e \tag{4.96}$$

$$J^e = \det\mathbf{F}^e \tag{4.97}$$

where $\mathbf{F}^e$ is as in equation (4.110). This is consistent with a thermo-elastic solid. In addition, response to chemical changes (swelling and material evolution) are also captured. Material parameters could be functions of temperature θ as well, but only the dependence on reactions state ϕ is considered in order to keep presentation simple. Furthermore, this modeling assumption is accurate when material parameters are less sensitive to temperature changes than to changes due to chemical reaction, or when the process involves a higher degree of chemical change and temperature changes are negligible in comparison. The functional dependence of the material parameters in (4.92), (4.94) and (4.95) on the reaction state ϕ are set as follows

$$\mu^s(\phi) = \left[1 + \alpha_\phi^s(1-\phi)\right]^3\left[\bar{\mu}_1^s\phi + \bar{\mu}_2^s(1-\phi)\right] \tag{4.98}$$

$$k^s(\phi) = \left[1 + \alpha_\phi^s(1-\phi)\right]^3\left[\bar{k}_1^s\phi + \bar{k}_2^s(1-\phi)\right] \tag{4.99}$$

$$c_0^s(\phi) = \left[1 + \alpha_\phi^s(1-\phi)\right]^3\left[\rho_1^s\bar{c}_1^s\phi + \rho_2^s\bar{c}_2^s(1-\phi)\right] \tag{4.100}$$

$$\eta_0^s(\phi) = \left[1 + \alpha_\phi^s(1-\phi)\right]^3\left[\frac{\rho_1^s\bar{\eta}_1^s}{\mathcal{M}_{s_1}}\phi + \frac{\rho_2^s\bar{\eta}_2^s}{\mathcal{M}_{s_2}}(1-\phi)\right] \tag{4.101}$$

$$h_0^s(\phi) = \left[1 + \alpha_\phi^s(1-\phi)\right]^3\left[\frac{\rho_1^s\bar{h}_1^s}{\mathcal{M}_{s_1}}\phi + \frac{\rho_2^s\bar{h}_2^s}{\mathcal{M}_{s_2}}(1-\phi)\right] \tag{4.102}$$

where $\bar{\mu}_1^s$, $\bar{k}_1^s$, ρ_1^s, $\bar{c}_1^s$, $\mathcal{M}_{s_1}$, $\bar{\eta}_1^s$, and $\bar{h}_1^s$ are respectively the shear modulus, bulk modulus, reference density, specific heat capacity, molar weight, entropy of formation, and change in enthalpy of formation for the unreacted solid, while the quantities with the subscript "2" are the ones that correspond to the reacted solid, all of which are standard material properties that are commonly reported (or otherwise are straight-forward to obtain) for many materials.

4.6.6 Specific Forms of Constitutive Equations

Using the specific form of the Helmholtz free energies for the constituents, given in terms of equations (4.90)-(4.97), and substituting them in the general form of the constitutive equations, the corresponding specific forms of those relations are obtained. First, by substituting in the definition of the quasi-conservative contribution (pressure term) to the fluid Cauchy stress from (4.56)

$$\mathbf{T}^{f,(q)} = -\frac{R}{\mathcal{M}_f}\theta\rho^f \mathbf{1} \tag{4.103}$$

Substituting for the solid Cauchy stress from (4.52)

$$\mathbf{T}^{s} = J^{-1}\left[\mu^s(\phi)\mathrm{dev}\left(\bar{\mathbf{B}}^e\right) + k^s(\phi)\ln J^e \mathbf{1}\right] \tag{4.104}$$

where

$$\mathrm{dev}\left(\bar{\mathbf{B}}^e\right) = \bar{\mathbf{B}}^e - \frac{1}{3}\mathrm{tr}\left(\bar{\mathbf{B}}^e\right)\mathbf{1} \quad ; \qquad \bar{\mathbf{B}}^e = (J^e)^{-2/3}\mathbf{B}^e \quad ; \qquad \mathbf{B}^e = \mathbf{F}^e\mathbf{F}^{eT} \tag{4.105}$$

Now, for the solid heat capacity, it is first assumed that the following holds, which is a reasonable assumption for most condensed matter (solids and liquids)

$$\left|-c_0^s(\phi)\frac{1}{\theta}\right| >> \left|3\left[3+\ln J^e\right]\left(\lambda^{tc}\right)^{-2}\left[\alpha_\theta^s(\phi)\right]^2 k^s(\phi)\right| \tag{4.106}$$

And the expression for the solid heat capacity is given by

$$-\rho^s\theta\frac{\partial^2\psi^s}{\partial\theta^2} = J^{-1}c_0^s(\phi) \tag{4.107}$$

4.6.7 Reduced Formulation for the Fluid Constituent

The work in [49] presented a reduced formulation of a fluid constituent, under certain assumptions, in the case of isothermal conditions. Here, a similar derivation is carried out for the non-isothermal case. The following assumptions are made regarding the balance of linear momentum of the fluid: (i) the effect of body forces is negligible compared to other terms; (ii) the main contribution to the interactive force is due to the drag term; (iii) the viscous contribution to the fluid Cauchy stress is small compared to the pressure term; and (iv) the inertial effects of the fluid are negligible. Assumptions (i) to (iv) are captured by equations (4.108), (4.109), (4.110), and (4.111), respectively. It is noted that assumptions (iii) and (iv) are consistent with a slow speed of diffusion of the fluid into the solid constituent.

$$\mathbf{b}^f \approx \mathbf{0} \tag{4.108}$$

$$\mathbf{I}^f \approx -\mathbf{A}^v(\mathbf{v}^f - \mathbf{v}^s) \tag{4.109}$$

$$\mathbf{T}^f \approx \mathbf{T}^{f,(q)} = -\frac{R}{\mathcal{M}_f}\theta\rho^f\mathbf{1} \tag{4.110}$$

$$\rho^f\frac{D^f\mathbf{v}^f}{Dt} = \rho^f\frac{\partial\mathbf{v}^f}{\partial t} + \rho^f(\mathbf{v}^f - \mathbf{v}^s)\cdot\nabla\mathbf{v}^f \approx \mathbf{0} \tag{4.111}$$

Furthermore, the drag tensor $\mathbf{A}^v$ is assumed to have the following form

$$\mathbf{A}^v = \frac{R}{\mathcal{M}_f}\theta\rho^f\mathbf{D}^{-1} \tag{4.112}$$

where $\mathbf{D}$ is a positive-definite diffusivity tensor, which in general may also depend on other state variables. The functional form of this dependencies is left unspecified at this point, and the form of $\mathbf{D}$ is specialized to different specific forms for each case study in section 4.8, accordingly. Equations (4.108) through (4.112) are substituted into the balance of linear momentum equation for the fluid constituent in equation (4.30). Assuming that $\nabla\rho^f/\rho^f \gg \nabla\theta/\theta$, the resulting equation can be solved algebraically for the fluid velocity field to yield the following.

$$\mathbf{v}^f = \mathbf{v}^s - \frac{1}{\rho^f}\mathbf{D}\nabla\rho^f \tag{4.113}$$

On the other hand, the balance of mass for the fluid constituent from equation (4.29) can be written in the following alternate form by using the definition of the ALE material time derivative in equation (4.16), and where the fluid velocity terms are gathered into a single term.

$$\frac{\partial\rho^f}{\partial t} - \mathbf{v}^s\cdot\nabla\rho^f + \operatorname{div}(\rho^f\mathbf{v}^f) - m^f = 0 \tag{4.114}$$

The fluid velocity expression from equation (4.113) is embedded into the alternate form of the fluid balance of mass in equation (4.114). This yields a transient reaction-diffusion equation for the fluid mass concentration field, which is shown below. Remarkably, the advection-like solid velocity terms cancel each other out and do not appear in the final form.

$$\frac{\partial\rho^f}{\partial t} - \operatorname{div}\left(\mathbf{D}\nabla\rho^f\right) + \rho^f\operatorname{div}\left(\mathbf{v}^s\right) - m^f = 0 \tag{4.115}$$

4.7 Reduced System of Governing Equations

Applying the modeling assumptions from section 4.6 to the balance laws, and taking the interactive force on the solid as well as the solid inertial effects to be negligible compared to the stress divergence term, the resulting system of equations is

$$\frac{D^s\rho^s}{Dt}+\rho^s\operatorname{div}(\mathbf{v}^s)-m^s=0 \qquad \text{in } \Omega\times]0,T[\qquad (4.116)$$

$$-\operatorname{div}(\mathbf{T}^s)=\mathbf{0} \qquad \text{in } \Omega\times]0,T[\qquad (4.117)$$

$$\frac{\partial\rho^f}{\partial t}-\operatorname{div}\left(\mathbf{D}\nabla\rho^f\right)+\rho^f\operatorname{div}\left(\mathbf{v}^s\right)-m^f=0 \qquad \text{in } \Omega\times]0,T[\qquad (4.118)$$

$$J^{-1}c_0^s(\phi)\frac{D^s\theta}{Dt}-J^{-1}\overline{h}_\phi\frac{D^s\phi}{Dt}+\operatorname{div}\mathbf{q} \;=\; 0 \qquad \text{in } \Omega\times]0,T[\qquad (4.119)$$

in addition to the evolution equation for the reaction state variable, given by

$$\frac{D^s\phi}{Dt}=r_\phi\exp\left[\frac{-E_a}{R}\left(\frac{\theta_0-\theta}{\theta_0\theta}\right)\right]J^{-1}\phi\rho^f \qquad \text{in } \Omega\times]0,T[\qquad (4.120)$$

the mass transfer terms are given by

$$m^f=r_m^f\exp\left[\frac{-E_a}{R}\left(\frac{\theta_0-\theta}{\theta_0\theta}\right)\right]J^{-2}\phi\rho^f \qquad (4.121)$$

$$m^s=r_s m^f \qquad (4.122)$$

and the constitutive equations for the heat flux and solids stress are

$$\mathbf{q}=-\boldsymbol{\kappa}\nabla\theta \qquad (4.123)$$

$$\mathbf{T}^s=J^{-1}\left[\mu^s(\phi)\operatorname{dev}\left(\overline{\mathbf{B}}^e\right)+k^s(\phi)\ln J^e\mathbf{1}\right] \qquad (4.124)$$

Remark: the gradient of the solid mass concentration field does not appear anywhere in the formulation. Therefore, the solid mass concentration field can be solved for via an ordinary differential equation in time at the integration points of a finite element discretization.

4.7.1 Thermo-chemo-mechanical coupled effects

Through the stated set of governing and constitutive equations, there are several coupled phenomena that we are able to capture. Specifically, there are three categories of coupling effects: (i) thermo-mechanical effects, (ii) chemo-mechanical effects, and (iii) thermo-chemical effects.

The thermo-mechanical effects include the thermal expansion of the solid, captured via the Cauchy stress tensor dependency on the temperature implied by the Helmholtz free energy function; the effect of the rate of deformation on the temperature, through the coupling term in the full balance of energy equation (this was neglected in the reduced formulation); and the dependency of the mechanical properties of the solid on the temperature.

Chemo-mechanical effects include the expansion of the solid associated with the extent of the reaction, captured via the dependency of the Cauchy stress tensor on the extent of reaction implied by the Helmholtz free energy function; and the dependency of the mechanical properties on the reaction state. Thermo-chemical effects include the heat generated by the exothermic reaction along the reaction zone, which is a moving heat source captured by the heat source term dependency on the mass reaction rate; the dependency of the rate of reaction on temperature since the reaction rate constant is a function of temperature; and the dependency of thermal properties of the solid on the reaction state.

4.7.2 Initial and boundary conditions

Imposition of boundary conditions for the constituents of a mixture is not a trivial task and it is subject of debate. For the Neumann boundary conditions of the mechanical field (i.e., tractions), it is usually the case that only the total traction on the mixture is known. However, this is generally not enough information to determine what part of the traction is applied to each constituent. Different physics-based assumptions can be made depending on the class of problems to address, and we refer to the work of Rajagopal and Tao [33], and Mohankumar et al. [34], for further discussion. A variety of boundary conditions are discussed in the work of Prasad and Rajagopal [50]. The work of Souček et al [51] deals with generating boundary conditions using a thermodynamical approach. The work in [38,52] presents numerical methods for consistent tying of constituents' tractions at Neumann boundaries.

In the modeling assumptions of the preceding section, the fluid constituent was taken to behave as an ideal gas, which is compressible and much more compliant than the solid constituent. Furthermore, we consider initial conditions of the problem that are far from a fully saturated state. As a consequence, the partial traction on the homogenized fluid is expected to be negligible compared to the that of the homogenized solid. Therefore, the partial traction of the homogenized solid is assumed to be approximately equal to the total traction on the mixture. On the other hand, the pressure in an ideal gas is proportional to its mass concentration, and the difference in concentration (or partial pressure) between the fluid in the surrounding environment and the fluid within the solid phase is enough to induce a flux across the boundary. As a result, the fluid diffusion relative to the solid is driven by the fluid mass concentration gradient.

We provide appropriate initial and boundary conditions so that the mathematical problem given by the system of PDEs and ODEs defined by (4.116)-(4.124) is well-posed. Considering the system to be in its initial state at time $t = t_0$, initial conditions for the fields across the domain Ω_0 to be

$$\rho^s(\mathbf{X}, t_0) = \rho_0^s(\mathbf{X}) \qquad \forall \mathbf{X} \in \Omega_0 \tag{4.125}$$

$$\phi(\mathbf{X}, t_0) = \phi_0(\mathbf{X}) \qquad \forall \mathbf{X} \in \Omega_0 \tag{4.126}$$

$$\mathbf{u}^s(\mathbf{X}, t_0) = \mathbf{u}_0^s(\mathbf{X}) \qquad \forall \mathbf{X} \in \Omega_0 \tag{4.127}$$

$$\rho^f(\mathbf{X}, t_0) = \rho_0^f(\mathbf{X}) \qquad \forall \mathbf{X} \in \Omega_0 \tag{4.128}$$

$$\theta(\mathbf{X}, t_0) = \theta_0(\mathbf{X}) \qquad \forall \mathbf{X} \in \Omega_0 \tag{4.129}$$

where all the initial fields on right hand side (with zero subscript) of (4.125)-(4.129) are given, known functions. In addition, Dirichlet and Neumann boundary conditions are prescribed for the solid displacement field as

$$\mathbf{u}^s(\mathbf{x}, t) = \mathbf{u}_D^s(\mathbf{x}, t) \qquad \forall \mathbf{x} \in \partial\Omega_{\mathbf{u}}^D \tag{4.130}$$

$$\mathbf{t}^s(\mathbf{x}, t) = \mathbf{T}^s \cdot \mathbf{n} = \mathbf{h}(\mathbf{x}, t) \qquad \forall \mathbf{x} \in \partial\Omega_{\mathbf{u}}^N \tag{4.131}$$

where $\partial\Omega_{\mathbf{u}}^D \cup \partial\Omega_{\mathbf{u}}^N = \partial\Omega$ and $\partial\Omega_{\mathbf{u}}^D \cap \partial\Omega_{\mathbf{u}}^N = \varnothing$, and $\mathbf{u}_D^s$, $\mathbf{h}$ are known, given functions. For the fluid mass concentration field, Dirichlet, Neumann, and Robin boundary conditions may be prescribed

$$\rho^f(\mathbf{x}, t) = \rho_D^f(\mathbf{x}, t) \qquad \forall \mathbf{x} \in \partial\Omega_\rho^D \tag{4.132}$$

$$\mathbf{n} \cdot \mathbf{D}\nabla\rho^f = j_N(\mathbf{x}, t) \qquad \forall \mathbf{x} \in \partial\Omega_\rho^N \tag{4.133}$$

$$\mathbf{n} \cdot \mathbf{D}\nabla\rho^f = c_\rho\left(\rho_{sat}^f - \rho^f\right) \qquad \forall \mathbf{x} \in \partial\Omega_\rho^R \tag{4.134}$$

where $\partial\Omega_\rho^D \cup \partial\Omega_\rho^N \cup \partial\Omega_\rho^R = \partial\Omega$ and their pairwise intersections are empty. Moreover, ρ_D^f is given concentration a the Dirichlet boundary, j_N is a given mass flux across the boundary (positive for inflow, negative for outflow), ρ_{sat}^f is a saturation mass concentration, and $c_\rho > 0$ is the dissolution rate, which may be dependent on temperature, reaction state, and state of stress or deformation. Likewise, boundary conditions for the temperature field are given by

$$\theta(\mathbf{x},t) = \theta_D(\mathbf{x},t) \qquad \forall \mathbf{x} \in \partial\Omega_\theta^D \tag{4.135}$$

$$\mathbf{n} \cdot \mathbf{q} = q_N(\mathbf{x},t) \qquad \forall \mathbf{x} \in \partial\Omega_\theta^N \tag{4.136}$$

$$\mathbf{n} \cdot \mathbf{q} = h_\theta \left(\theta_R - \theta\right) \qquad \forall \mathbf{x} \in \partial\Omega_\theta^R \tag{4.137}$$

where $\partial\Omega_\theta^D \cup \partial\Omega_\theta^N \cup \partial\Omega_\theta^R = \partial\Omega$ and their pairwise intersections are empty, θ_D is a prescribed temperature field at the Dirichlet boundary, q_N is a known heat flux across the Neumann boundary, θ_R is a reference ambient temperature, and h_θ is the heat transfer coefficient.

General Robin boundary conditions are written as some linear combination of the unknown field and its gradient. Here, the form of such linear combination for the boundary condition corresponds to the so-called “convection” boundary condition in heat conduction problems. In the case of the fluid, this boundary condition is also used to model the rate of dissolution of a given solute at the surface of the dissolving medium.

4.8 Numerical Implementation and Results

In this section, numerical results for different test cases are presented. The goal of this section is to demonstrate the applicability of the formulation developed in the preceding sections to a class of coupled thermo-chemo-mechanical problems involving fluid-infused solids and to showcase different aspects of the formulation developed in this chapter, such as finite strains, and coupled physics response, as well as its numerical implementation. In what follows, we drop the superscript s of the solid displacement for ease of notation. The unknown fields are the solid mass concentration field ρ^s, reaction state ϕ, solid displacement $\mathbf{u}$, fluid mass concentration ρ^f, and temperature θ. We write the weak form of the problem using an updated Lagrangian formulation wherein the unknown fields are taken to be functions of the material coordinates of the solid constituent $\mathbf{X}^s$ and time t, quantities are integrated over the deformed (or current) configuration Ω, and gradients are taken with respect to the spatial coordinate $\mathbf{x}$. First, we define the spaces of admissible trial and test functions for the fields whose spatial gradients appear in the governing system of equations ($\mathbf{u}$, ρ^f, and θ)

$$S_{\mathbf{u}} := \left\{ \mathbf{u} \,|\, \mathbf{u} \in H^1,\ \mathbf{u} = \mathbf{u}_D \text{ on } \partial\Omega^D_{\mathbf{u}} \right\} \tag{4.138}$$

$$S_{\rho} := \left\{ \rho^f \,|\, \rho^f \in H^1,\ \rho^f = \rho^f_D \text{ on } \partial\Omega^D_{\rho} \right\} \tag{4.139}$$

$$S_{\theta} := \left\{ \theta \,|\, \theta \in H^1,\ \theta = \theta_D \text{ on } \partial\Omega^D_{\theta} \right\} \tag{4.140}$$

$$V_{\mathbf{u}} := \left\{ \hat{\mathbf{w}} \,|\, \hat{\mathbf{w}} \in H^1,\ \hat{\mathbf{w}} = \mathbf{0} \text{ on } \partial\Omega^D_{\mathbf{u}} \right\} \tag{4.141}$$

$$V_{\rho} := \left\{ \hat{g} \,|\, \hat{g} \in H^1,\ \hat{g} = 0 \text{ on } \partial\Omega^D_{\rho} \right\} \tag{4.142}$$

$$V_{\theta} := \left\{ \hat{\varphi} \,|\, \hat{\varphi} \in H^1,\ \hat{\varphi} = 0 \text{ on } \partial\Omega^D_{\theta} \right\} \tag{4.143}$$

The weak form of the problem can be stated as follows: given all the material parameters, and appropriate boundary and initial conditions, find $\left\{\mathbf{u}, \rho^f, \theta\right\} \in S := S_{\mathbf{u}} \times S_{\rho} \times S_{\theta}$ such that for all $\left\{\hat{\mathbf{w}}, \hat{g}, \hat{\varphi}\right\} \in V := V_{\mathbf{u}} \times V_{\rho} \times V_{\theta}$, the following system of equations holds

$$R_{\mathbf{u}}\left(\hat{\mathbf{w}}\ ;\ \mathbf{u},\ \rho^f,\ \theta;\ \phi\right) \ = \ \left(\frac{\partial \hat{w}_i}{\partial x_j},\ T^s_{ij}\right)_{\Omega} - \left(\hat{w}_i,\ h_i\right)_{\partial\Omega^N_{u_i}} \ = \ 0 \tag{4.144}$$

$$\begin{aligned} R_{\rho}\left(\hat{g}\ ;\ \mathbf{u},\ \rho^f,\ \theta;\ \phi\right) \ = \ & \left(\hat{g},\ \frac{\partial \rho^f}{\partial t} + \rho^f \operatorname{div}\left(\mathbf{v}^s\right) - m^f\right)_{\Omega} + \left(\frac{\partial \hat{g}}{\partial x_i},\ D_{ij}\frac{\partial \rho^f}{\partial x_j}\right)_{\Omega} \\ & - \left(\hat{g},\ j_N\right)_{\partial\Omega^N_{\rho}} - \left(\hat{g},\ c_{\rho}\left(\rho^f_{sat} - \rho^f\right)\right)_{\partial\Omega^R_{\rho}} \ = \ 0 \end{aligned} \tag{4.145}$$

$$\begin{aligned} R_{\theta}\left(\hat{\varphi}\ ;\ \mathbf{u},\ \rho^f,\ \theta;\ \phi\right) \ = \ & \left(\hat{\varphi},\ J^{-1} c^s_0(\phi)\frac{D^s\theta}{Dt} - J^{-1}\bar{h}_{\phi}\frac{D^s\phi}{Dt}\right)_{\Omega} - \left(\frac{\partial \hat{\varphi}}{\partial x_i},\ q_i\right)_{\Omega} \\ & + \left(\hat{\varphi},\ q_N\right)_{\partial\Omega^N_{\theta}} + \left(\hat{\varphi},\ h_{\theta}\left(\theta_R - \theta\right)\right)_{\partial\Omega^R_{\theta}} \ = \ 0 \end{aligned} \tag{4.146}$$

where $\left(u, v\right)_{\omega} := \int_{\omega} uv\, d\omega$ and Einstein summation convention is used over the indices $i, j \in \left\{1, 2, 3\right\}$. The system of equations is complemented by the evolution equation for the solid mass concentration in (4.116), the evolution equation for the reaction state variable ϕ from (4.120), and the constitutive equations for the fluid mass transfer m^f from equation (4.121), the heat flux $\mathbf{q}$ from equation (4.123), and the Cauchy stress of the solid from equation (4.124). We recall that the Cauchy stress of the solid is a function of $\mathbf{u}$, θ, and ϕ through the gradient of the total and inelastic deformation and the evolution of material properties.

Spatial discretization is carried out by seeking trial solutions in finite-dimensional subsets $\left\{\mathbf{u}^h, \rho^{f,h}, \theta^h\right\} \in S^h \subset S$ such that for all $\left\{\hat{\mathbf{w}}^h, \hat{g}^h, \hat{\varphi}^h\right\} \in V^h \subset V$,

$$R_{\mathbf{u}}\left(\hat{\mathbf{w}}^h \ ; \ \mathbf{u}^h, \ \rho^{f,h}, \ \theta^h; \phi\right) \ = \ 0 \tag{4.147}$$

$$R_{\rho}\left(\hat{g}^h \ ; \ \mathbf{u}^h, \ \rho^{f,h}, \ \theta^h; \phi\right) \ = \ 0 \tag{4.148}$$

$$R_{\theta}\left(\hat{\varphi}^h \ ; \ \mathbf{u}^h, \ \rho^{f,h}, \ \theta^h; \phi\right) \ = \ 0 \tag{4.149}$$

For the standard finite element method, S^h and V^h are piecewise polynomial spaces defined for a non-overlapping partition of the domain into finite elements K. We use nodal Lagrangian finite elements, where a generic test or trial function σ is expanded in terms of finite element shape functions as

$$\sigma\left(\mathbf{X}^s, t\right) = \sum_{A=1}^{nel} \bar{\sigma}_A\left(t\right) N_A\left(\mathbf{X}^s\right) \tag{4.150}$$

where nel is the number of element nodes, N_A is the shape function associated to node A, and $\bar{\sigma}_A$ is a nodal degree of freedom associate to node A. Integrals are computed numerically using a Gauss quadrature rule of sufficiently high order.

The time interval of interest is discretized into N subintervals $\left[t_n, \ t_{n+1}\right]$ with $n = 0, ..., N-1$ and $\Delta t_{n+1} = t_{n+1} - t_n$. The following notation is used for the time discrete approximation of a generic time-continuous field σ.

$$\sigma_{n+1} \approx \sigma\left(t_{n+1}\right) \tag{4.151}$$

The backward Euler time marching algorithm is employed, and rate quantities are given by

$$\mathbf{v}^s = \frac{\mathbf{u}_{n+1} - \mathbf{u}_n}{\Delta t_{n+1}} \tag{4.152}$$

$$\frac{\partial \rho^f}{\partial t} = \frac{\rho^f_{n+1} - \rho^f_n}{\Delta t_{n+1}} \tag{4.153}$$

$$\frac{D^s \rho^s}{Dt} = \frac{\rho^s_{n+1} - \rho^s_n}{\Delta t_{n+1}} \tag{4.154}$$

$$\frac{D^s \theta}{Dt} = \frac{\theta_{n+1} - \theta_n}{\Delta t_{n+1}} \tag{4.155}$$

$$\frac{D^s\phi}{Dt} = \frac{\phi_{n+1} - \phi_n}{\Delta t_{n+1}} \tag{4.156}$$

The discrete version of the evolution equation for ρ^s from (4.116) yields the following update equation for ρ^s_{n+1} at the integration points

$$\rho^s_{n+1} = \frac{\rho^s_n + \Delta t_{n+1} m^s_{n+1}}{1 + \operatorname{div}\left(u_{n+1} - u_n\right)} \tag{4.157}$$

Similarly, the discrete version of the ODE for ϕ from equation (4.120) yields the following update equation for ϕ_{n+1} at the integration points.

$$\phi_{n+1} = \frac{\phi_n}{1 - \Delta t_{n+1} r_\phi \exp\left[\frac{-E_a}{R}\left(\frac{\theta_0 - \theta_{n+1}}{\theta_0 \theta_{n+1}}\right)\right] J^{-1}_{n+1} \rho^f_{n+1}} \tag{4.158}$$

Thus, the formulation from section 4.7 was implemented in a finite element code using three-dimensional 8-noded, trilinear hexahedral elements (HEX8) for all DOFs. Therefore, the total number of DOFs in this formulation is 5 DOFs per node , comprising three components for solid displacement vector $\mathbf{u}$, the mass concentration of the fluid ρ^f, and the temperature θ. The solid mass concentration ρ^s and the reaction state variable ϕ are obtained by numerically solving their corresponding ODEs at the integration points. We remark that one needs to solve for only a single solid displacement field, and it is not required to solve an additional PDE for the fluid velocity field. In contrast, there would be 11 DOFs per node if one had to solve additional PDEs for another solid constituent’s displacement field and for the fluid velocity field. In other words, the formulation summarized in section 4.7 enables size reduction of the system that is to be solved numerically, which results in a computationally cost-effective formulation. Time integration was carried out through the backward Euler algorithm, both for the time discretization of the DOFs as well as for the evolution of internal variables at the integration points. At each timestep, the nonlinear system of equations is solved using the Newton-Raphson algorithm, wherein the consistent tangent was derived and implemented, and quadratic rate of convergence is achieved. The unknown fields and governing equations were normalized during the solution procedure to maintain good conditioning of the system and to ensure same level of accuracy of the solution in all fields.

The numerical test cases show that standard Galerkin FEM with equal order interpolation for all the fields works for the model from section 4.7 and the material parameters under consideration. For discussion regarding finite elements for mixed-field problems we refer the interested reader to Chapter 4 of [53]. For stabilized methods dealing with incompressibility in nonlinear regime, see [54]. For problems with highly dominant advective or reactive effects, more sophisticated stabilized methods are required, such as in the solution of coupled chemo-mechanical problems presented in [41,46]. Problems where multiple solid constituents are modeled independently, and the solid interactive forces are also significant, require further algorithmic considerations regarding splitting of traction boundary conditions and determination of the interactive force, which are discussed in [38,40,52]. A stabilized method for the mixed form of convective-diffusive heat transport is presented in [55].

4.8.1 Coupled thermo-chemo-mechanical 1D response

This test case serves the purposes of highlighting the coupled effects (i.e. thermo-chemical, thermo-mechanical, and chemo-mechanical effects) in a simple setting: one-dimensional response of a representative material. The diffusivity and heat conductivity tensors are assumed isotropic, which are taken to depend only on reaction state according to the following equation

$$\mathbf{D} = D\mathbf{1}, \quad D(\phi) = \bar{D}_1\phi + \bar{D}_2(1-\phi) \tag{4.159}$$

$$\boldsymbol{\kappa} = \kappa\mathbf{1}, \quad \kappa(\phi) = \bar{\kappa}_1^s\phi + \bar{\kappa}_2^s(1-\phi) \tag{4.160}$$

where $\bar{D}_1$ and $\bar{D}_2$ are the diffusivities for the pristine and the fully reacted material, respectively, and $\bar{\kappa}_1^s$, $\bar{\kappa}_2^s$ are their thermal conductivities. Here, the temperature and stress dependence are ignored, but these effects can be straight-forward to capture by redefining the functional form of these tensors. The values of the relevant material properties for this case study are given in Table 4.1.

Table 4.1. Material parameters for a representative material

Quantity	Units	Value
Unreacted solid volumetric heat capacity ($\rho_1^s\bar{c}_1^s$)	J m^{-3} K^{-1}	2.3e5
Unreacted solid heat conductivity ($\bar{\kappa}_1^s$)	W m^{-1} K^{-1}	2e-7
Unreacted solid coefficient of thermal expansion ($\bar{\alpha}_\theta^{s1}$)	K^{-1}	1e-3
Unreacted solid diffusivity ($\bar{D}_1$)	m^2 s^{-1}	1e-12
Unreacted solid shear modulus ($\bar{\mu}_1^s$)	Pa	70e9
Unreacted solid bulk modulus ($\bar{k}_1^s$)	Pa	100e9
Reacted solid specific heat capacity ($\rho_2^s\bar{c}_2^s$)	J m^{-3} K^{-1}	2.3e5
Reacted solid heat conductivity ($\bar{\kappa}_2^s$)	W m^{-1} K^{-1}	2e-7
Reacted solid coefficient of thermal expansion ($\bar{\alpha}_\theta^{s2}$)	K^{-1}	1e-3
Reacted solid diffusivity ($\bar{D}_2$)	m^2 s^{-1}	20e-12
Reacted solid shear modulus ($\bar{\mu}_2^s$)	Pa	35e9
Reacted solid bulk modulus ($\bar{k}_2^s$)	Pa	50e9
Reaction constant (r_m)	s^{-1}	-1e3
Reaction constant (r_ϕ)	m^3 kg^{-1} s^{-1}	-5e6
Coefficient of chemical expansion (α_ϕ)	dimensionless	3.5e-1
Reaction activation energy (E_a)	J mol^{-1}	8.314
Gas constant (R)	J K^{-1} mol^{-1}	8.314
Standard change in enthalpy of reaction (Δh_{rxn})	J m^{-3}	-1.5e8
Reaction change in heat capacity (Δc_p)	J m^{-3} K^{-1}	1e1
Reference temperature (θ_0)	K	1e3

Consider a thin sheet of this material of thickness $L = 0.5\mu\text{m}$ with one side exposed to an oxidizing environment, and the other side being insulated and fixed to a surface. The representative computational domain is taken as a slender square column of material through the thickness of the sheet. The side faces are constrained due to the surrounding material, and this is captured through idealized boundary conditions that induce a one-dimensional response within the computational domain. The geometry, and boundary and initial conditions for the normalized fields of the computational domain are shown in Figure 4.3, noting that the temperature field is normalized by a reference value of 1000 K, the solid displacement field is normalized by L, and the fluid concentration field is normalized by a reference value of 1e-4 kg/m^3.

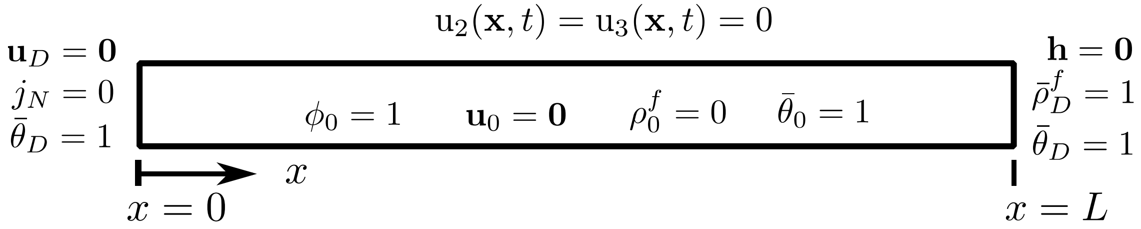


Figure 4.3. Schematic diagram of the side view of the computational domain and boundary conditions.

The problem is solved in the time interval $I=[0,\ 3.1]\,\mathrm{s}$ using a nonuniform timestep size as follows: a timestep size Δt_1 is used for the time subinterval $I_1=[0,\ 0.1)\mathrm{s}$, for the subinterval $I_2=(0.1,\ 0.6)\mathrm{s}$ the timestep used is $\Delta t_2=10\Delta t_1$, and a third timestep size $\Delta t_3=10\Delta t_2$ is employed for the reminder of the time interval of interest. The spatial domain is discretized with a single row of equal size 8-node hexahedral elements along the axial direction. In order to perform checks on the convergence of the numerical solution, the problem is solved with different levels of refinement in the spatial and temporal discretizations, as given by the total number of finite elements (nel) and timesteps (nt) employed. Figure 4.4 shows numerical results for a range of the time history of the normalized displacement at the right end of the domain, in which it reaches its peak value.

The numerical results from Figure 4.4 suggest numerical convergence towards a unique solution, with refinement in space and time. Furthermore, the error in the maximum displacement of the coarser discretizations is within acceptable tolerance (below 2%) for the purposes of this numerical test case of a hypothetical material. The numerical results that follow are for a discretization with 20 elements and 175 total timesteps.

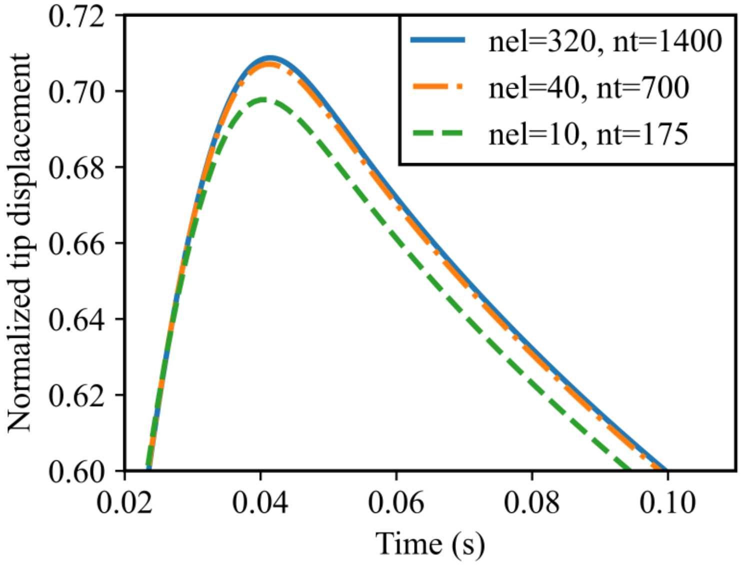


Figure 4.4. Extract from the time history of the normalized displacement at the right end of the domain obtained for numerical simulations with different number of finite elements (nel) and timesteps (nt).

Figure 4.5 shows snapshots of the distribution of the 22-component of the Cauchy stress field in the deformed configuration of the domain at different times corresponding to (a) the initial state, (b) active reaction zone in the middle of the domain, (c) solid reactant is depleted, and (d) the steady state. The expansion along the axial direction is due to the coupled thermo-mechanical and chemo-mechanical effects. Furthermore, recall the thermal and chemical contributions to the deformation are isotropic in the model presented herein, while the domain is constrained in the perpendicular directions by the surrounding material (idealized through fixed displacement boundary conditions), which gives rise to compressive stress. These stress values corresponding to each snapshot are also plotted in Figure 4.6.

Regarding the evolution of the 22-component of Cauchy stress field in Figure 4.6, it is noted that it is zero at time $t = 0$. At time $t = 0.015\text{s}$ there is a higher compressive stress in the right half of the domain, corresponding to the constrained chemical and thermal expansion taking place, while the stress value reduces towards zero in the left end, where no chemical expansion and very small thermal expansion has occurred. At time $t = 0.6\text{s}$, the entire solid has fully reacted and the entire domain has chemically swollen, while the heat produced by the reaction has not yet completely dissipated, leading to higher temperatures in the middle of the domain, which in turn result in larger stresses due to larger thermal expansion in that region. Finally, at $t = 1\text{s}$, a steady state is reached, in which heat produced by reaction has dissipated and there is a uniform compressive stress throughout the domain associated with the constrained chemical expansion everywhere in the domain.

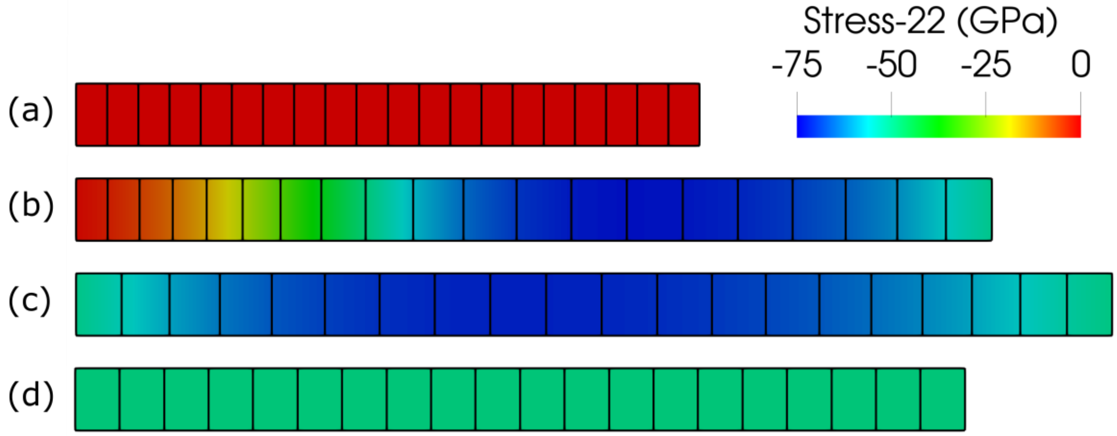


Figure 4.5. Snapshots of the distribution of the 22-component of Cauchy stress in the deformed configuration at times (a) 0s (initial), (b) 0.015s (mid-reaction), (c) 0.06s (depleted solid reactant), and (d) 1s (steady state).

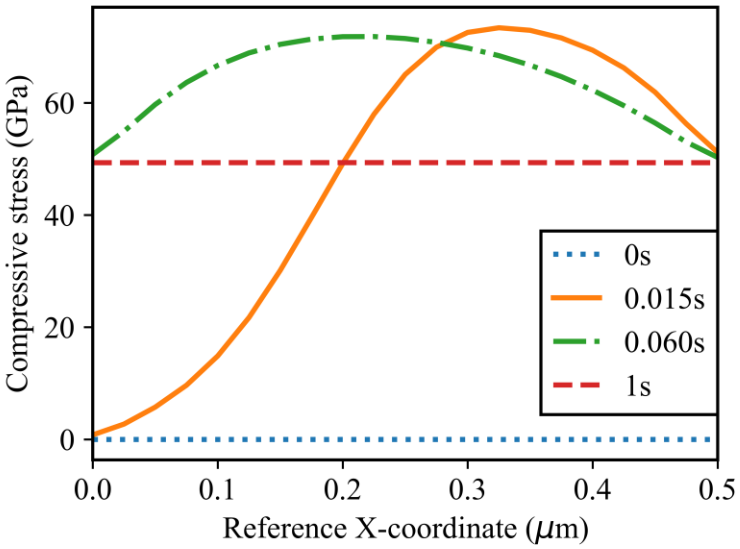


Figure 4.6. Distribution of the 22-component of Cauchy stress (positive for compression) vs the reference X coordinate at times (a) 0s (initial), (b) 0.015s (mid-reaction), (c) 0.06s (depleted solid reactant), and (d) 1s (steady state).

In order to further highlight the nature of the thermo-chemo-mechanical coupled response, consider the state of the solution corresponding to Figure 4.5b. The normalized temperature field $\bar{\theta}$, normalized fluid mass concentration field $\bar{\rho}^f$, and reaction state field ϕ are plotted along the deformed configuration at time $t = 0.015\text{s}$ in Figure 4.7. At this point during the solution procedure, the right end of the domain has fully reacted: the solid reactant is completely depleted

($\phi = 0$) and it has been replaced by the reacted solid (product of reaction). On the other hand, the left end of the domain remains in a virgin state ($\phi = 1$). At the same time, the fluid has diffused into the domain, and the region where the fluid concentration is nonzero ($\bar{\rho}^f > 0$) and there is solid reactant available ($\phi > 0$) is the zone where chemical reaction is actively taking place, fluid is being consumed by reaction, and the solid material is evolving from its unreacted (or virgin) state to its reacted state. In addition, the chemical reaction here is exothermic, and heat is generated by active chemical reaction faster than it is dissipated due to conduction to the boundary, which leads to an increase of temperature within the domain.

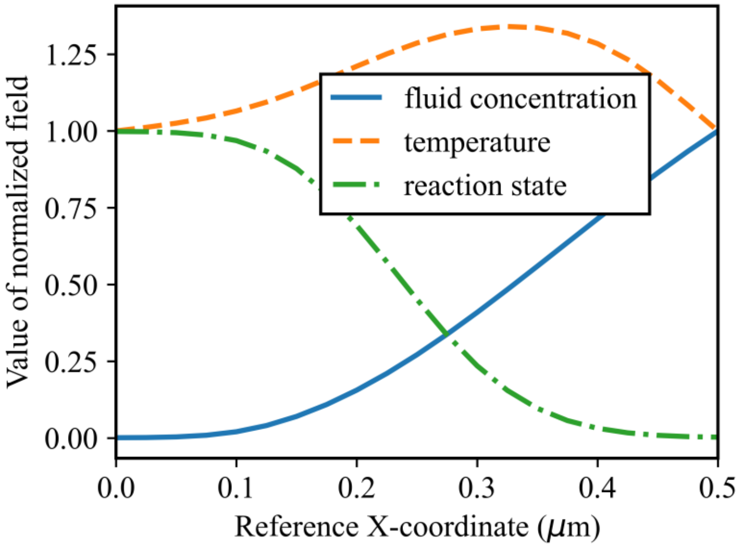


Figure 4.7. Plot of normalized fluid concentration, normalized temperature, and reaction state fields along the deformed configuration at time $t = 0.015\text{s}$.

The time histories up to 0.5s of the normalized displacement at the right end of the domain, and of the normalized fluid concentration at the left end of the domain are shown in Figure 4.8. Note that the quantities asymptote to constant values as time increases, signaling that a steady state is reached. Furthermore, there is a peak in the tip displacement which is the point where the combination of chemical and thermal expansion are the greatest, which then asymptotes to a nonzero value since the chemical deformation is permanent, while the thermal deformation vanishes as the heat dissipates and temperature goes back to its initial value. It is worth noting that the maximum value of the normalized fluid concentration at the left end is higher ($\max_t \left(\bar{\rho}^f \big|_{X=0} \right) = 1.01$) than its boundary condition at the right end ($\bar{\rho}^f \big|_{X=L} = 1$). This is a consequence of the fact that the fluid is diffusing through the deformed solid, and when the domain contracts as it cools back down, it does so faster than the fluid diffuses. Therefore, it is possible to have regions in the domain where the concentration of fluid is temporarily higher than the concentration at the inlet.

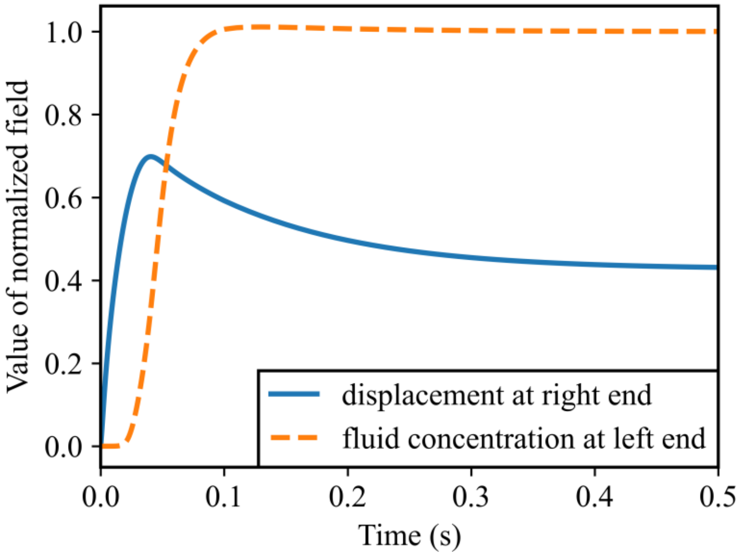


Figure 4.8. Time history of normalized displacement at right end of the domain, and of normalized fluid concentration at the left end of the domain.

The most noticeable coupled effects that are showcased by this test case are the expansion due to chemical reaction (chemo-mechanical), the thermal expansion (thermo-mechanical), and the production of heat by chemical reaction (thermo-chemical). Furthermore, there are other coupled effects that are taking (or could take) place in this problem, but due to the magnitudes of the material parameters, are not as prominently evident: evolution of mechanical material properties (chemo-mechanical), temperature dependence of reaction rate and evolution of thermal material properties (thermo-chemical). In addition, recall that it was assumed in section 4.6 that the mechanical heating due to rate of deformation is negligible, and therefore the reduced system of governing equations does not account for it.

4.8.2 Thermal oxidation of silicon carbide

In this numerical test case, we apply the model developed in this chapter to the thermal oxidation of silicon carbide (SiC). Ceramic matrix composites (CMCs) with SiC fibers and/or matrix are key components in many components of rocket engines (including main combustion chamber and exhaust nozzle) as well as vehicle-structure thermal protection during reentry [1]; therefore the study of its thermal oxidation is subject of interest [56]. Moreover, properties of SiC make it suitable for next-generation power and semiconductor devices, where thermal oxidation is a crucial step in the manufacturing process of metal-oxide semiconductor field effect transistors (MOSFETS), and finds applications in high-efficiency electronics, high-power electronics, and automotive parts [57]. SiC is also used in nuclear power and automotive applications. The oxidation of SiC follows the chemical reaction

$$\mathrm{SiC}_{(s)} + 1.5\mathrm{O}_{2(g)} \rightarrow \mathrm{SiO}_{2(s)} + \mathrm{CO}_{(g)} \tag{4.161}$$

where one mol of silicon carbide (SiC) reacts with 1.5 moles of gaseous molecular oxygen (O_2) to form one mol of solid silicon dioxide (a.k.a. silica, SiO_2) and one mol of gaseous carbon monoxide (CO), which quickly diffuses out of the system. In this process, the newly formed SiO_2 occupies 2.2 times the volume of the SiC substrate it replaces. A finite strain model, such as the one formulated in this chapter, is required to accurately describe material response undergoing local deformation of this magnitude. A schematic of the oxidation process with oxide growth and substrate recession is shown in Figure 4.9.

Remark: In this test case, we use the isotropic formulation from section 4.7 to model 4H-SiC, which is an anisotropic material. However, the boundary and initial conditions are such that there is only a 1D response in the solution fields. Nevertheless, we recognize that using an isotropic model for an anisotropic material is a gross approximation, and a more general constitutive model would be more appropriate in more general problems.

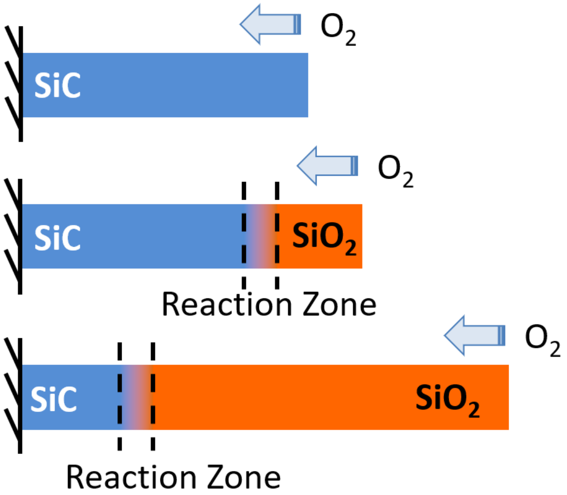


Figure 4.9. Schematic diagram of the thermal oxidation process in silicon carbide (SiC) showing SiC recession and SiO_2 growth.

We consider a thin plate of 4H-SiC with crystal orientation so that the surface corresponding to its C-face is exposed to a high temperature oxidizing environment with an oxygen partial pressure of 1 atm. The temperature is held at a constant uniform value of $\theta = 1100$ ºC (1373 K) throughout the oxidation process. Because the simulation is run under isothermal conditions, there is no need to solve the balance of energy equation to find the temperature field. Therefore, the problem is reduced to a chemo-mechanical problem and we need only the material parameters required in the solution of the system posed by the balance of mass of the fluid and the balance of linear momentum for the solid. The diffusivity is taken to be of the form given in equation (4.159)[5]. Robin boundary conditions are considered for the fluid mass concentration at the exposed boundary. Therefore, it is required to specify the saturation concentration of the fluid, ρ^f_{sat}, and the dissolution rate, c_ρ. The dissolution rate of O_2 is taken to depend on the reaction state according to the following equation

$$c_\rho(\phi) = \bar{c}_{\rho,1}\phi + \bar{c}_{\rho,2}(1-\phi) \tag{4.162}$$

where $\bar{c}_{\rho,1}$ and $\bar{c}_{\rho,2}$ are the O_2 dissolution rates in the pristine and in the fully reacted material. The values of the required material parameters are given in Table 4.2.

Table 4.2. Material parameters for oxidation of C-face of 4H-SiC under isothermal conditions[6].

Quantity	Units	Value
Unreacted solid volumetric heat capacity ($\rho^s_1\bar{c}^s_1$)	J m^{-3} K^{-1}	not required
Unreacted solid heat conductivity ($\bar{\kappa}^s_1$)	W m^{-1} K^{-1}	not required
Unreacted solid coefficient of thermal expansion ($\bar{\alpha}^{s1}_\theta$)	K^{-1}	not required
Unreacted solid diffusivity ($\bar{D}_1$)	m^2 s^{-1}	1e-12
Unreacted solid dissolution rate ($\bar{c}_{\rho,1}$)	m s^{-1}	7.5e-3

---

[5] Oxygen diffusion occurs predominantly in the region of the domain that has already been oxidized to SiO_2. Diffusivity of O_2 in SiO_2 is known to be temperature and stress dependent. However, isothermal conditions are applied in this test case. Moreover, numerical results show that there is a uniform and constant state of stress in most of the fully oxidized region. Therefore, the value of the diffusivity in the oxide $\bar{D}_2$ can be taken as a constant, which is to be understood as its effective value under the given constant and uniform temperature and state of stress.

[6] These values were determined with the help of several references [60,61,64,65]. In particular, the estimates for the diffusivities, dissolution rates, and the reaction constant r_m were set so that a reasonable fit is obtained with a separate set of experimental data in [61] corresponding to an oxidizing environment with a 0.02 atm O_2 partial pressure.

Table 4.2 (cont.)

Unreacted solid shear modulus ($\bar{\mu}_1^s$)	Pa	180e9
Unreacted solid bulk modulus ($\bar{k}_1^s$)	Pa	250e9
Reacted solid specific heat capacity ($\rho_2^s \bar{c}_2^s$)	J m^{-3} K^{-1}	not required
Reacted solid heat conductivity ($\bar{\kappa}_2^s$)	W m^{-1} K^{-1}	not required
Reacted solid coefficient of thermal expansion ($\bar{\alpha}_\theta^{s2}$)	K^{-1}	not required
Reacted solid diffusivity ($\bar{D}_2$)	m^2 s^{-1}	2e-12
Reacted solid dissolution rate ($\bar{c}_{\rho,2}$)	m s^{-1}	2.4e-5
Reacted solid shear modulus ($\bar{\mu}_2^s$)	Pa	30e9
Reacted solid bulk modulus ($\bar{k}_2^s$)	Pa	35e9
Reaction constant (r_m)	s^{-1}	-0.1333e7
Reaction constant (r_ϕ)	m^3 kg^{-1} s^{-1}	-0.3467e3
Coefficient of chemical expansion (α_ϕ)	dimensionless	0.3
Reaction activation energy (E_a)	J mol^{-1}	not required
Gas constant (R)	J K^{-1} mol^{-1}	8.314
Standard change in enthalpy of reaction (Δh_{rxn})	J m^{-3}	not required
Reaction change in heat capacity (Δc_p)	J m^{-3} K^{-1}	not required
Reference temperature (θ_0)	K	1373
Saturation concentration of the fluid (ρ_{sat}^f)	kg m^{-3}	2.44e-5

The computational domain may be defined as a column of material, representing the material through the thickness of a plate. For formation of thin oxide scales, the in-plane deformation is constrained by the surrounding material, and the response can be approximated as one-dimensional. In fact, accurate models have been proposed for oxide thickness growth that assume a one-dimensional steady-state response [58,59], although they may fail to capture the development of internal stresses that induce plate bending observed on similar experiments in silicon after the oxide layer grows thicker [9,60]. Here, the computational domain of length

$L = 120$ nm is discretized with a row of 400 HEX8 finite elements. The geometry of the computational domain, as well as boundary and initial conditions, are shown in Figure 4.10. The initial conditions are $\mathbf{u}_0(\mathbf{X}) = \mathbf{0}$, $\rho_0^f(\mathbf{X}) = 0$, and $\phi_0(\mathbf{X}) = 1$. For the displacement field, homogeneous Dirichlet boundary conditions (i.e., fixed) are applied at the left end ($x = 0$), and homogeneous Neumann boundary conditions (i.e., traction free) at the right end ($x = L$). For the fluid mass concentration field, homogeneous Neumann boundary conditions (i.e., no outflow) are applied at the left end ($x = 0$), and Robin boundary conditions are applied at the right end ($x = L$). No-outflow boundary conditions are applied for the fluid in the remaining parts of the boundary, and the displacement in the directions perpendicular to the *x*-axis is constraint to be zero everywhere (analogous to plane strain conditions). Collectively, these boundary conditions induce a one-dimensional response. Regarding time discretization, we used 20 timesteps of 0.01 s, 20 timesteps of 1 s, 20 timesteps of 10 s, and a time step of 100 s thereon until a final time of 5 h is reached. The solution fields are the normalized displacement and normalized fluid concentration. The displacement field is normalized by dividing by the domain size L, and the fluid concentration field is normalized by dividing with the saturation concentration ρ_{sat}^f. Therefore, the Robin boundary condition for the normalized fluid concentration field utilizes the normalized saturation concentration value as $\bar{\rho}_{sat}^f = 1$.

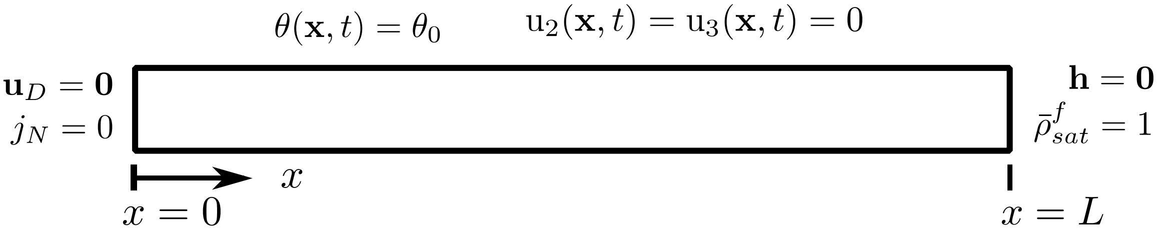
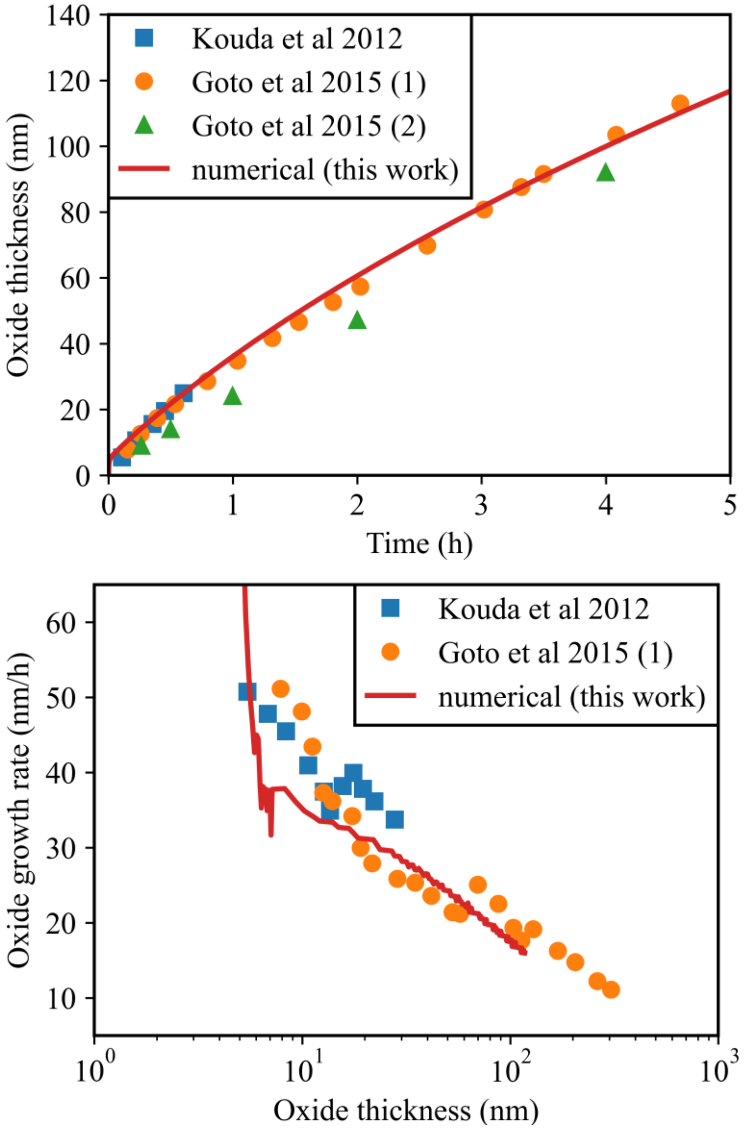

$\theta(\mathbf{x}, t) = \theta_0$
$\mathrm{u}_2(\mathbf{x}, t) = \mathrm{u}_3(\mathbf{x}, t) = 0$
$\mathbf{u}_D = \mathbf{0}$
$j_N = 0$
$\mathbf{h} = \mathbf{0}$
$\bar{\rho}_{sat}^f = 1$
x
$x = 0$
$x = L$

Figure 4.10. Schematic diagram of the computational domain and boundary conditions.

Figure 4.11 shows a comparison between the numerical results and several experimental results [61,62] for the evolution of the oxide thickness with time as well as the oxide growth rate as a function of oxide thickness. The thickness of the oxide at the end of a time step is measured as the length of the region in the deformed configuration where $\phi < 0.75$[7]. The growth rate was determined through numerical differentiation (using a central difference formula) of the oxide-thickness-vs-time data. The numerical results are in good agreement with experimental data.

---

[7] This value was chosen as part of the calibration process to estimate some of the material parameters in Table 4.2



(a)

(b)

Figure 4.11. Comparison of numerical results with experimental data [61,62] for the dry thermal oxidation of C-face of 4H-SiC at 1373 K (1100 ºC) and 1 atm O_2 partial pressure, (a) oxide thickness vs time, and (b) oxide growth rate vs thickness.

In addition, we show snapshots of the solution fields for different times during the solution procedure (10 min, 1hr, and 5hr). Figure 4.12 shows the growth of the oxide layer and recession of the substrate material. The distribution of the T_{11} (axial) and T_{22} (transverse) component of the Cauchy stress at the times corresponding to those snapshots are shown in Figure 4.13. It is noted that there is a sharp variation in the stress field near the reaction front. In addition, the compressive transverse stress in the fully oxidized region has a constant uniform value, which is associated with the chemical expansion being constrained in the transverse directions to the axis of the computational domain.

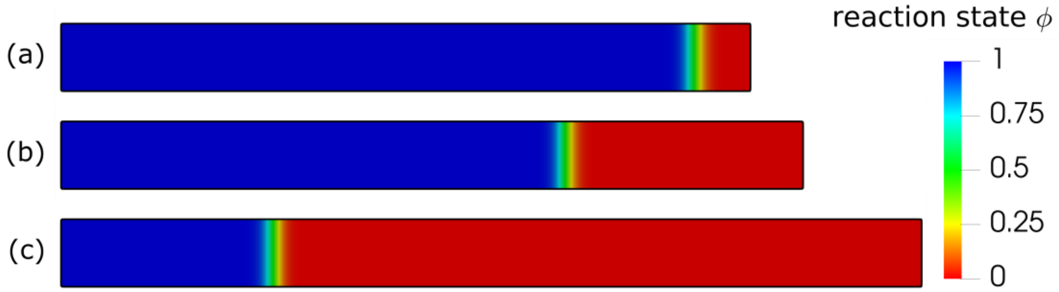


Figure 4.12. Snapshots of distribution of reaction state in the deformed configuration at (a) t = 10min, (b) t = 1hr, and (c) t = 5hr.

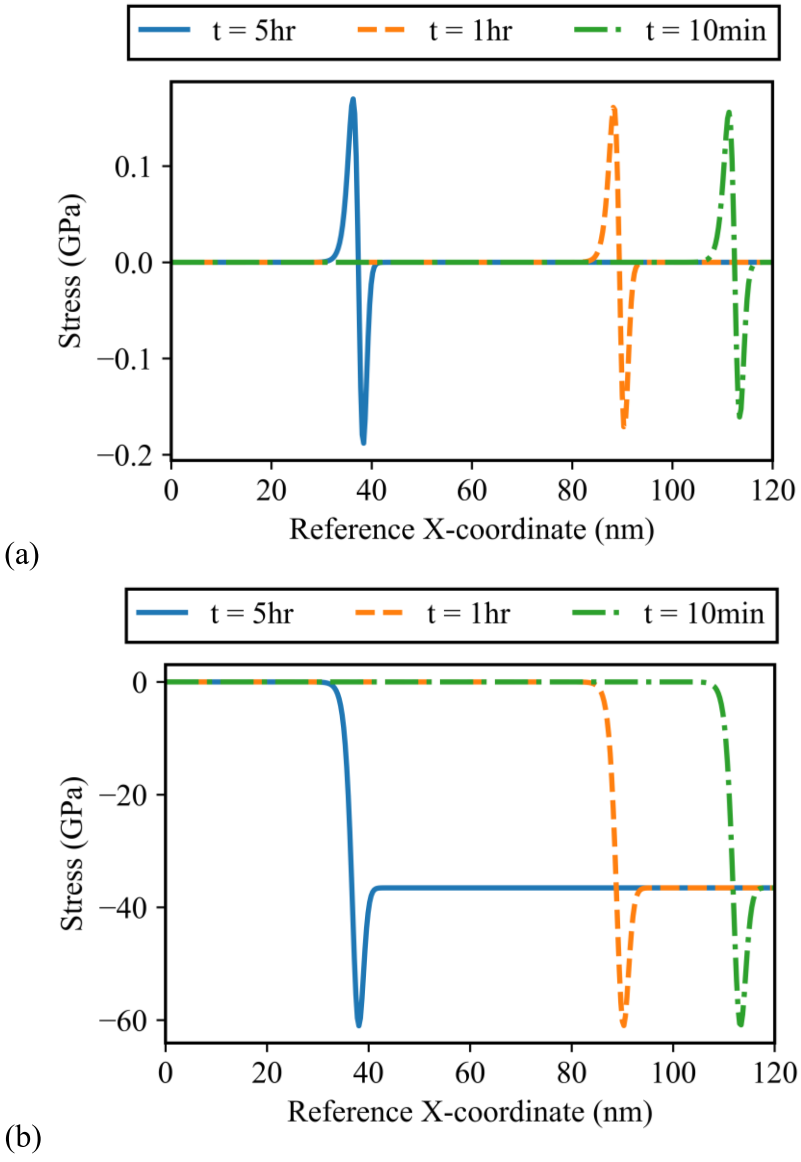


Figure 4.13. Snapshots of the distribution of (a) the T_{11} and (b) the T_{22} components of the Cauchy stress at different times during solution procedure.

4.8.3 Thermal cycling of FeCrAlY plate with groove

In this test case we consider the thermal cycling and oxidation of a plate made of a FeCrAlY alloy. This alloy is used as a metallic bond coat as part of thermal barrier coating (TBC) systems for superalloys, and it forms a thermally grown oxide (TGO) in oxidizing environments at high temperatures, which may lead to debonding [28]. Therefore, this process is critical for the aerospace industry, which relies in this class of materials for components in the gas turbine or rocket engines [1]. In this test case, the domain geometry consists of a plate of material that has a groove on its surface, similar to the problem discussed in [5]. Here, the diffusivity is assumed isotropic and given by the following equation, which accounts for reaction state and temperature dependence

$$\mathbf{D}\left(\phi,\theta\right) = D(\phi,\theta)\mathbf{1} \tag{4.163}$$

$$D(\phi,\theta) = \bar{D}_1 \exp\left[\frac{Q_{s,1}}{R\theta_0}\left(1-\frac{\theta_0}{\theta}\right)\right]\phi + \bar{D}_2 \exp\left[\frac{Q_{s,2}}{R\theta_0}\left(1-\frac{\theta_0}{\theta}\right)\right]\left(1-\phi\right) \tag{4.164}$$

where $Q_{s,1}$ and $Q_{s,2}$ are the activation energies of diffusion for the unreacted and reacted material, and $\bar{D}_1$, $\bar{D}_2$ are the effective diffusivities of the unreacted and reacted material at the reference temperature θ_0. During thermal cycling, the temperature history is prescribed at all points in the domain; therefore, it is not required to solve for the balance of energy equation to find the temperature field. Here, Robin boundary conditions will not be considered, so the values of the dissolution rates are not needed. The required material parameters are given in Table 4.3.

Table 4.3. Material parameters for thermal cycling of FeCrAlY alloy[8].

Quantity	Units	Value
Unreacted solid volumetric heat capacity ($\rho_1^s \bar{c}_1^s$)	J m^{-3} K^{-1}	not required
Unreacted solid heat conductivity ($\bar{\kappa}_1^s$)	W m^{-1} K^{-1}	not required
Unreacted solid coefficient of thermal expansion ($\bar{\alpha}_\theta^{s1}$)	K^{-1}	14.3e-6
Unreacted solid diffusivity ($\bar{D}_1$)	m^2 s^{-1}	1.6e-12
Unreacted solid activation energy of diffusion ($Q_{s,1}$)	J mol^{-1}	100e3

---

[8] Material parameters for the model presented in this chapter were determined from those reported for the model in [5].

Table 4.3 (cont.)

Unreacted solid dissolution rate ($\bar{c}_{\rho,1}$)	m s^{-1}	not required
Unreacted solid shear modulus ($\bar{\mu}_1^s$)	Pa	77.775e9
Unreacted solid bulk modulus ($\bar{k}_1^s$)	Pa	165.55e9
Reacted solid specific heat capacity ($\rho_2^s \bar{c}_2^s$)	J m^{-3} K^{-1}	not required
Reacted solid heat conductivity ($\bar{\kappa}_2^s$)	W m^{-1} K^{-1}	not required
Reacted solid coefficient of thermal expansion ($\bar{\alpha}_\theta^{s2}$)	K^{-1}	not required
Reacted solid diffusivity ($\bar{D}_2$)	m^2 s^{-1}	1.6e-12
Unreacted solid activation energy of diffusion ($Q_{s,2}$)	J mol^{-1}	100e3
Reacted solid dissolution rate ($\bar{c}_{\rho,2}$)	m s^{-1}	not required
Reacted solid shear modulus ($\bar{\mu}_2^s$)	Pa	166.1e9
Reacted solid bulk modulus ($\bar{k}_2^s$)	Pa	253.1e9
Reaction constant (r_m)	s^{-1}	-12.5
Reaction constant (r_ϕ)	m^3 kg^{-1} s^{-1}	-3.256e-3
Coefficient of chemical expansion (α_ϕ)	dimensionless	0.20546
Reaction activation energy (E_a)	J mol^{-1}	100e3
Gas constant (R)	J K^{-1} mol^{-1}	8.314
Standard change in enthalpy of reaction (Δh_{rxn})	J m^{-3}	not required
Reaction change in heat capacity (Δc_p)	J m^{-3} K^{-1}	not required
Reference temperature (θ_0)	K	1423

The domain considered is a plate where several grooves have been carved on its top surface at regular spacing (see [4,5]). Taking advantage of the symmetries of the problem, the computational domain represents a planar cross-section of the plate including half a groove, and its discretization is shown in Figure 4.14 using a single layer of HEX8 elements. The time step is kept a $\Delta t = 30s$ throughout the solution procedure.

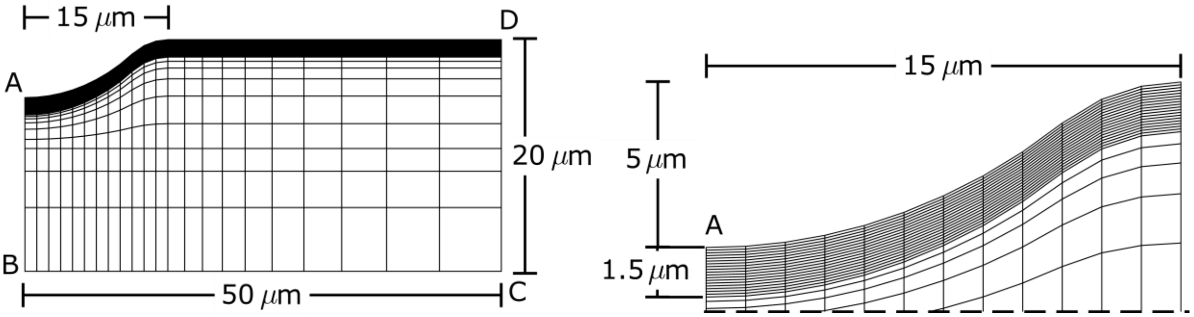


Figure 4.14. Schematic of the computational domain (left) and close-up of the groove (right).

The domain is subjected to 24 thermal cycles, where the temperature history for a single cycle is given in Figure 4.15. In addition, there is an initial 2 min artificial cooldown to match the process simulated in [5]. The initial conditions are $\mathbf{u}_0(\mathbf{X}) = \mathbf{0}$, $\theta_0(\mathbf{X}) = 1423$ K, $\rho_0^f(\mathbf{X}) = 0$, and $\phi_0(\mathbf{X}) = 1$. For the fluid concentration field, edge AB is the Dirichlet boundary where the prescribed history of fluid concentration $\rho_D^f(t)$ is described for a single cycle[9] in Figure 4.15. The fluid concentration at the boundary is kept as zero during the initial cool down step. In the remainder of the boundary, we apply no-outflow conditions. For the displacement field, we apply zero horizontal displacement along AB and zero vertical displacement along BC. The horizontal displacement along edge DC is prescribed so that it corresponds to the thermal expansion of the substrate, i.e., $u_2|_{DC} = \bar{\alpha}_\theta^{s1}(\theta - \theta_0) \times 50\ \mu\text{m}$. Finally, out-of-plane deformation is prescribed in the entire domain so that it also corresponds to the thermal expansion of the substrate, i.e., $F_{33} = 1 + \bar{\alpha}_\theta^{s1}(\theta - \theta_0)$. Traction free boundary conditions are applied for the remaining parts of the boundary for which the displacement is not prescribed.

Figure 4.16 shows the shape of the deformed groove after thermal cycling and its comparison with other numerical simulations and experiments. There is reasonably good agreement between the results obtained in this chapter and the results in the literature. Note that the magnitude of the deformation smaller in the numerical results obtained in this chapter. The main reason for this discrepancy is that plastic and viscoplastic effects are not accounted for in the present model. This is also reflected in Figure 4.17, where the magnitude of the stresses is very large since the response of the material is elastic. Here, the maximum compressive stress in the oxide is ≈ -160 GPa, while the value reported in [5] is ≈ -5.3 GPa. Once plastic effects are included, the excess stress would be relieved by larger plastic deformations, resulting in both larger magnitude of the deformation and lower magnitude of stresses. Nevertheless, key features of the solution are still captured by the model in this chapter, namely the material evolution and formation of an oxide layer as well as the inelastic deformations associated with the chemical reactions. The value obtained for the oxide layer thickness (measured to the location where $\phi = 0.2$) is 1.5 μm.

---

[9] The solubility of interstitial oxygen in the solid is a function of temperature. Therefore, the fluid concentration at the Dirichlet boundary depends nonlinearly on the temperature history. The fluid concentration history is approximated as piecewise linear to simplify the imposition of the fluid boundary condition.

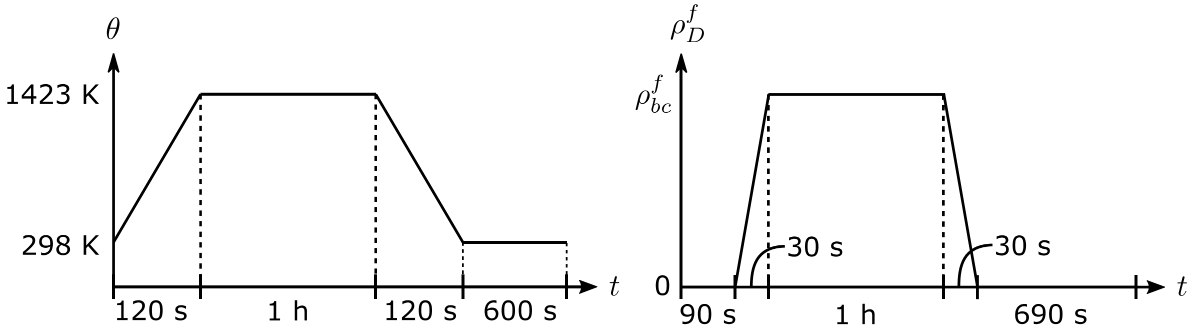


Figure 4.15. Prescribed history of uniform temperature (left) and concentration at fluid Dirichlet boundary (right), where $\rho_{bc}^f = 4.5414 \times 10^{-2}$ kg/m^3.

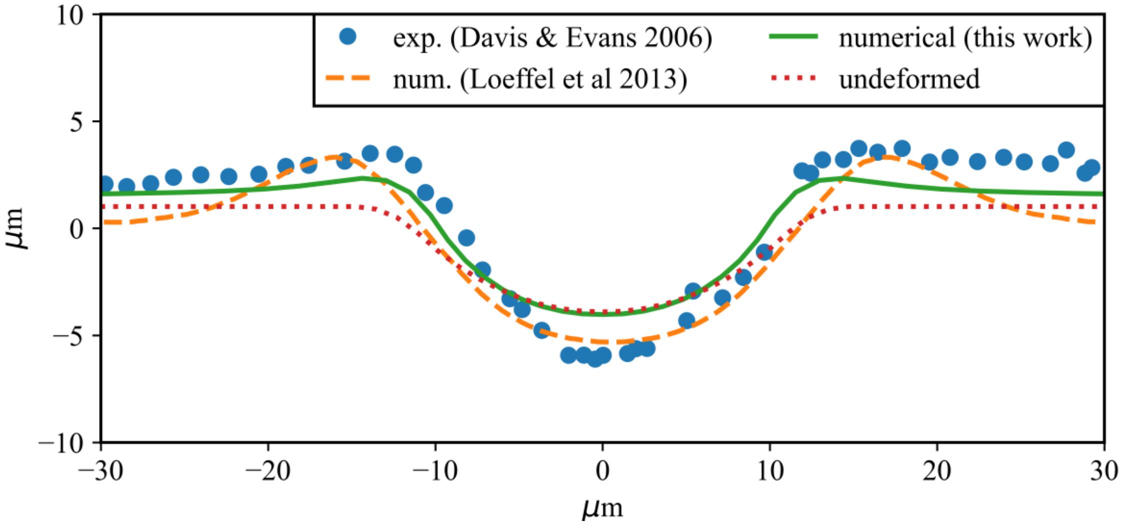


Figure 4.16. Comparison of numerical results obtained for the profile of the deformed groove geometry with the literature [4,5].

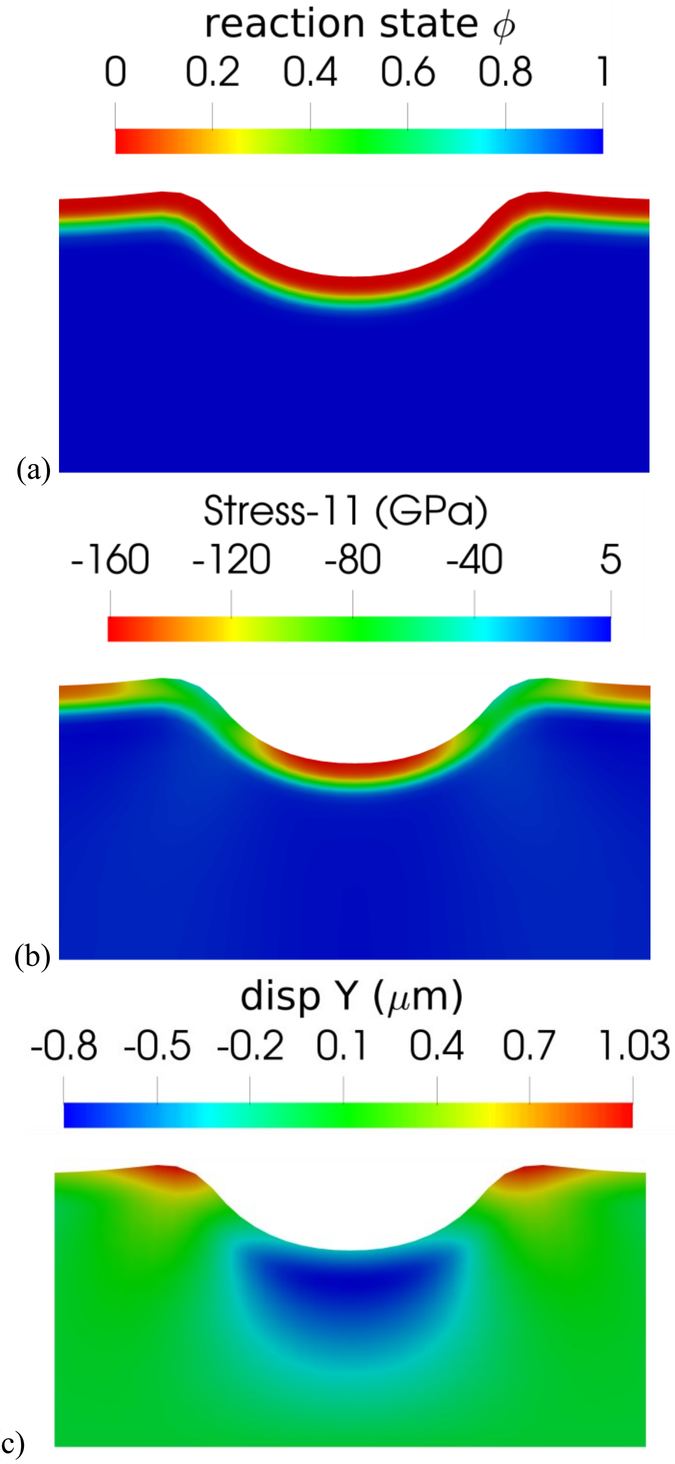


Figure 4.17. Deformed configuration of the domain showing distribution of (a) the reaction state, (b) T_{11} component of stress, and (c) vertical displacement.

4.9 Conclusions

A new, thermodynamically consistent, mixture model for fully coupled thermo-chemo-mechanical processes in solid-fluid material systems was derived. There is a large class of problems of engineering significance in the aerospace and semiconductor industries, among many others, that exhibit a thermo-chemo-mechanical response under different stimuli. The material systems in these problems are also characterized by being comprised of multiple constituents. Therefore, mixture theory provided a framework to develop a model where interactions between constituents are captured at the continuum level. The subscale details of the interfaces where different constituents come into contact need not be explicitly modeled. As such, the resulting formulation may be thought of a physics-based, reduced-order model for multi-constituent interaction. The derivation of thermodynamically consistent constitutive relations is done in a way that ensures non-negative entropy production in the open system and satisfies the second law of thermodynamics.

Further assumptions were made to specialize the formulation to the case where the fluid reactant is gas-like in nature, and this results in a system of PDEs with fewer unknown fields that need to be solved. In addition, specific functional forms for the Helmholtz free energies of the fluid and solid constituents are provided, from which constitutive equations are derived. A key element in developing a tractable constitutive theory is the multiplicative split of the deformation gradient, stemming from the assumption of a compositional mapping, which accounts for and segregates the effects of the elastic and inelastic mechanical response.

Furthermore, the form of the resulting model has consequences in terms of its numerical implementation. The model is amenable to numerical solution via the standard Galerkin finite element method with equal order interpolation for all fields. The state variables for solid density and reaction state may be treated as internal variables and solved for via discretized ODEs at the integration points of a finite element code. This is a direct result from the fact that gradients of those fields do not appear in the final form of the specialized formulation. This has the benefit of reducing the number of degrees of freedom that must be solved for in the FEM, which reduces the computational cost of solving fully coupled thermo-chemo-mechanical boundary value problems.

The numerical test cases presented showcase how the model presented herein is able to capture fully coupled thermo-chemo-mechanical processes in solid-fluid material systems. Effects such as swelling due to chemical reaction, material evolution, material degradation, heat generation/consumption for exothermic/endothermic reactions, and influence of mechanical deformation on fluid transport, among others. In particular, the formulation in this chapter was applied to model the thermal oxidation and deformation of silicon carbide (SiC) and of metallic bond-coat alloys such as FeCrAlY during thermal cycling. These materials are used in gas turbine engines, rocket engines, and vehicle-structure thermal protection during reentry, as well as in the manufacturing of semiconductor devices in the case of SiC. In both cases, reasonably good agreement with experimental data was attained, which serves as validation of the descriptive and predictive capabilities of the formulation presented in this chapter.

4.10 References

[1] N.P. Padture, Advanced structural ceramics in aerospace propulsion, Nat. Mater. 15 (2016) 804–809.

[2] F. Lofaj, Y.S. Kaganovskii, Kinetics of WC-Co oxidation accompanied by swelling, J. Mater. Sci. 30 (1995) 1811–1817.

[3] N.S. Jacobson, Corrosion of Silicon-Based Ceramics in Combustion Environments, J. Am. Ceram. Soc. 76 (1993) 3–28.

[4] A.W. Davis, A.G. Evans, Some Effects of Imperfection Geometry on the Cyclic Distortion of Thermally Grown Oxides, Oxid. Met. 65 (2006) 1–14.

[5] K. Loeffel, L. Anand, Z.M. Gasem, On modeling the oxidation of high-temperature alloys, Acta Mater. 61 (2013) 399–424.

[6] J. Smialek, N.S. Jacobson, Oxidation of High-Temperature Aerospace Materials, High Temp. Mater. Mech. (2014) 95–162.

[7] T. Merzouki, E. Blond, N. Schmitt, M.-L. Bouchetou, T. Cutard, A. Gasser, Modelling of the swelling induced by oxidation in SiC-based refractory castables, Mech. Mater. 68 (2014) 253–266.

[8] M. Aronovici, G. Bianchi, L. Ferrari, M. Barbato, S. Gianella, G. Scocchi, A. Ortona, Heat and Mass Transfer in Ceramic Lattices During High-Temperature Oxidation, J. Am. Ceram. Soc. 98 (2015) 2625–2633.

[9] Dah-Bin Kao, J.P. McVittie, W.D. Nix, K.C. Saraswat, Two-dimensional silicon oxidation experiments and theory, in: 1985 Int. Electron Devices Meet., IRE, 1985: pp. 388–391.

[10] H. Kageshima, M. Uematsu, K. Shiraishi, Theory of thermal Si oxide growth rate taking into account interfacial Si emission effects, Microelectron. Eng. 59 (2001) 301–309.

[11] T. Jiang, S. Rudraraju, A. Roy, A. Van Der Ven, K. Garikipati, M.L. Falk, Multiphysics Simulations of Lithiation-Induced Stress in Li1+xTi2O4Electrode Particles, J. Phys. Chem. C. 120 (2016) 27871–27881.

[12] H. Yang, F. Fan, W. Liang, X. Guo, T. Zhu, S. Zhang, A chemo-mechanical model of lithiation in silicon, J. Mech. Phys. Solids. 70 (2014) 349–361.

[13] S.R. Bishop, Chemical expansion of solid oxide fuel cell materials: A brief overview, Acta Mech. Sin. 29 (2013) 312–317.

[14] B.A. Haberman, J.B. Young, Three-dimensional simulation of chemically reacting gas flows in the porous support structure of an integrated-planar solid oxide fuel cell, Int. J. Heat Mass Transf. 47 (2004) 3617–3629.

[15] T. Shoaib, A. Carmichael, R.E. Corman, Y. Shen, T.H. Nguyen, R.H. Ewoldt, R.M. Espinosa-Marzal, Self-adaptive hydrogels to mineralization, Soft Matter. 13 (2017) 5469–5480.

[16] S.A. Chester, L. Anand, A thermo-mechanically coupled theory for fluid permeation in elastomeric materials: Application to thermally responsive gels, J. Mech. Phys. Solids. 59 (2011) 1978–2006.

[17] P. Yuan, J.M. McCracken, D.E. Gross, P. V Braun, J.S. Moore, R.G. Nuzzo, A programmable soft chemo-mechanical actuator exploiting a catalyzed photochemical water-oxidation reaction, Soft Matter. 13 (2017) 7312–7317.

[18] A. Fasano, K. Kannan, A. Mancini, K.R. Rajagopal, Modelling Ziegler–Natta Polymerization in High Pressure Reactors, in: G. Capriz, P.M. Mariano (Eds.), Mater. Substruct. Complex Bodies, Elsevier Science Ltd, Oxford, 2007: pp. 206–237.

[19] G.P. Tandon, K. V. Pochiraju, G.A. Schoeppner, Modeling of oxidative development in PMR-15 resin, Polym. Degrad. Stab. 91 (2006) 1861–1869.

[20] J. Gonzalez, K. Sun, M. Huang, J. Lambros, S. Dillon, I. Chasiotis, Three dimensional studies of particle failure in silicon based composite electrodes for lithium ion batteries, J. Power Sources. 269 (2014) 334–343.

[21] S. Zhang, Chemomechanical modeling of lithiation-induced failure in high-volume-change electrode materials for lithium ion batteries, Npj Comput. Mater. 3 (2017) 1–10.

[22] C. Truesdell, Rational Thermodynamics, Second, Springer New York, New York, NY, 1984.

[23] C. Truesdell, W. Noll, The Non-Linear Field Theories of Mechanics, Third, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[24] K. Kannan, K.R. Rajagopal, A thermodynamical framework for chemically reacting systems, Zeitschrift Für Angew. Math. Und Phys. 62 (2011) 331–363.

[25] M. Bulíček, J. Málek, V. Průša, Thermodynamics and Stability of Non-Equilibrium Steady States in Open Systems, Entropy. 21 (2019).

[26] K. Loeffel, L. Anand, A chemo-thermo-mechanically coupled theory for elastic–viscoplastic deformation, diffusion, and volumetric swelling due to a chemical reaction, Int. J. Plast. 27 (2011) 1409–1431.

[27] S.A. Chester, C. V. Di Leo, L. Anand, A finite element implementation of a coupled diffusion-deformation theory for elastomeric gels, Int. J. Solids Struct. 52 (2015) 1–18.

[28] K. Al-Athel, K. Loeffel, H. Liu, L. Anand, Modeling decohesion of a top-coat from a thermally-growing oxide in a thermal barrier coating, Surf. Coatings Technol. 222 (2013) 68–78.

[29] T. Sain, K. Loeffel, S. Chester, A thermo–chemo–mechanically coupled constitutive model for curing of glassy polymers, J. Mech. Phys. Solids. 116 (2018) 267–289.

[30] S. Konica, T. Sain, A thermodynamically consistent chemo-mechanically coupled large deformation model for polymer oxidation, J. Mech. Phys. Solids. 137 (2020) 103858.

[31] X. Zhang, Z. Zhong, A coupled theory for chemically active and deformable solids with mass diffusion and heat conduction, J. Mech. Phys. Solids. 107 (2017) 49–75.

[32] R.J. Atkin, R.E. Craine, Continuum theories of mixtures: Basic theory and historical development, Q. J. Mech. Appl. Math. 29 (1976) 209–244.

[33] K.R. Rajagopal, L. Tao, Mechanics of Mixtures, World Scientific, 1995.

[34] K. V Mohankumar, V. Průša, K. Kannan, A.S. Wineman, Remarks on continuum theory of mixtures: editorial to special issue on mixture theory, Int. J. Adv. Eng. Sci. Appl. Math. 9 (2017) 120–134.

[35] J. Siddique, A. Ahmed, A. Aziz, C. Khalique, A Review of Mixture Theory for Deformable Porous Media and Applications, Appl. Sci. 7 (2017) 917.

[36] O. Souček, V. Orava, J. Málek, D. Bothe, A continuum model of heterogeneous catalysis:

Thermodynamic framework for multicomponent bulk and surface phenomena coupled by sorption, Int. J. Eng. Sci. 138 (2019) 82–117.

[37] J.D. Humphrey, K.R. Rajagopal, A constrained mixture model for growth and remodeling of soft tissues, Math. Model. Methods Appl. Sci. 12 (2002) 407–430.

[38] T.J. Truster, A. Masud, A unified mixture formulation for density and volumetric growth of multi-constituent solids in tissue engineering, Comput. Methods Appl. Mech. Eng. 314 (2017) 222–268.

[39] D. Faghihi, X. Feng, E.A.B.F. Lima, J.T. Oden, T.E. Yankeelov, A coupled mass transport and deformation theory of multi-constituent tumor growth, J. Mech. Phys. Solids. 139 (2020) 103936.

[40] H. Gajendran, R.B. Hall, A. Masud, K.R. Rajagopal, Chemo-mechanical coupling in curing and material-interphase evolution in multi-constituent materials, Acta Mech. 229 (2018) 3393–3414.

[41] R. Hall, H. Gajendran, A. Masud, Diffusion of chemically reacting fluids through nonlinear elastic solids: mixture model and stabilized methods, Math. Mech. Solids. 20 (2015) 204–227.

[42] R.J. Atkin, R.E. Craine, Continuum Theories of Mixtures: Applications, IMA J. Appl. Math. 17 (1976) 153–207.

[43] R. Hall, K. Rajagopal, Diffusion of a fluid through an anisotropically chemically reacting thermoelastic body within the context of mixture theory, Math. Mech. Solids. 17 (2012) 131–164.

[44] R.B. Hall, A mixture-compatible theory of chemothermal deposition and expansion in n - constituent finitely deforming composite materials with initially circularly cylindrical microstructures, Math. Mech. Solids. 20 (2015) 228–248.

[45] M. Anguiano, H. Gajendran, R.B. Hall, A. Masud, Coupled Anisothermal Chemomechanical Degradation Solutions in One Dimension, in: A. Arzoumanidis, M. Silberstein, A. Amirkhizi (Eds.), Challenges Mech. Time Depend. Mater. Vol. 2, Springer International Publishing, Cham, 2018: pp. 5–9.

[46] M. Anguiano, H. Gajendran, R.B. Hall, K.R. Rajagopal, A. Masud, Chemo-mechanical coupling and material evolution in finitely deforming solids with advancing fronts of reactive fluids, Acta Mech. 231 (2020) 1933–1961.

[47] K.R. Rajagopal, A.R. Srinivasa, On the thermomechanics of materials that have multiple natural configurations Part I: Viscoelasticity and classical plasticity, Zeitschrift Für Angew. Math. Und Phys. ZAMP. 55 (2004) 861–893.

[48] A.B. Freidin, E.N. Vilchevskaya, I.K. Korolev, Stress-assist chemical reactions front propagation in deformable solids, Int. J. Eng. Sci. 83 (2014) 57–75.

[49] M. Anguiano, A. Masud, Reduced mixture model and elastic response of chemically swollen solids: Application to Si oxidation and lithiation, Appl. Eng. Sci. 6 (2021) 100039.

[50] S.C. Prasad, K.R. Rajagopal, On the Diffusion of Fluids Through Solids Undergoing Large Deformations, Math. Mech. Solids. 11 (2006) 291–305.

[51] O. Souček, M. Heida, J. Málek, On a thermodynamic framework for developing boundary conditions for Korteweg-type fluids, Int. J. Eng. Sci. 154 (2020) 103316.

[52] H. Gajendran, R.B. Hall, A. Masud, Edge stabilization and consistent tying of constituents at Neumann boundaries in multi-constituent mixture models, Int. J. Numer. Methods Eng. 110 (2017) 1142–1172.

[53] T.J.R. Hughes, The finite element method: linear static and dynamic finite element analysis, Dover Publications, Mineola, NY, 2000.

[54] A. Masud, T.J. Truster, A framework for residual-based stabilization of incompressible finite elasticity: Stabilized formulations and F bar methods for linear triangles and tetrahedra, Comput. Methods Appl. Mech. Eng. 267 (2013) 359–399.

[55] M. Ayub, A. Masud, A New Stabilized Formulation For Convective-Diffusive Heat Transfer, Numer. Heat Transf. Part B Fundam. 44 (2003) 1–23.

[56] M. Wilson, E. Opila, A Review of SiC Fiber Oxidation with a New Study of Hi-Nicalon SiC Fiber Oxidation, Adv. Eng. Mater. 18 (2016) 1698–1709.

[57] X. She, A.Q. Huang, Ó. Lucía, B. Ozpineci, Review of Silicon Carbide Power Devices and Their Applications, IEEE Trans. Ind. Electron. 64 (2017) 8193–8205.

[58] Y. Song, S. Dhar, L.C. Feldman, G. Chung, J.R. Williams, Modified Deal Grove model for the thermal oxidation of silicon carbide, J. Appl. Phys. 95 (2004) 4953–4957.

[59] D. Goto, Y. Hijikata, Unified theory of silicon carbide oxidation based on the Si and C emission model, J. Phys. D. Appl. Phys. 49 (2016) 225103.

[60] R.S. Hay, Growth stress in SiO 2 during oxidation of SiC fibers, J. Appl. Phys. 111 (2012)

063527.

[61] K. Kouda, Y. Hijikata, S. Yagi, H. Yaguchi, S. Yoshida, Oxygen partial pressure dependence of the SiC oxidation process studied by in-situ spectroscopic ellipsometry, J. Appl. Phys. 112 (2012).

[62] D. Goto, Y. Hijikata, S. Yagi, H. Yaguchi, Differences in SiC thermal oxidation process between crystalline surface orientations observed by in-situ spectroscopic ellipsometry, J. Appl. Phys. 117 (2015) 0–6.

[63] K.R. Rajagopal, A.R. Srinivasa, A Gibbs-potential-based formulation for obtaining the response functions for a class of viscoelastic materials, Proc. R. Soc. A Math. Phys. Eng. Sci. 467 (2011) 39–58.

[64] K. Kajihara, H. Kamioka, M. Hirano, T. Miura, L. Skuja, H. Hosono, Interstitial oxygen molecules in amorphous SiO 2. III. Measurements of dissolution kinetics, diffusion coefficient, and solubility by infrared photoluminescence, J. Appl. Phys. 98 (2005) 013529.

[65] L.L. Snead, T. Nozawa, Y. Katoh, T.S. Byun, S. Kondo, D.A. Petti, Handbook of SiC properties for fuel performance modeling, J. Nucl. Mater. 371 (2007) 329–377.

CHAPTER 5:
EXPLICIT VARIATIONAL MULTISCALE DISCONTINUOUS GALERKIN METHOD[1]

5.1 Introduction

We present a numerical method for developing inter-domain transmission conditions for transient problems in the finite element method where different parts of the larger domain are assigned to different computational platforms and are solved synchronously and concurrently with minimal sharing of information. Such conditions are required for parallel solution of large-scale problems by domain decomposition or when different parts of the domain are modeled by different physical models implemented in separate codes. Such characteristics make the method ideally suited to leverage distributed and heterogeneous computational resources, even when they are part of "loosely" connected networks, such as through cloud computing technologies.

Issue of coupling of meshes representing multiple domains also arises in multiphysics interacting problems that have two or more geometrically distinct regions that interact at their boundaries through the transfer of motion, forces, and heat [1,2]. An important consideration in the coupling of different models on these adjoining meshes is the exchange of information across the interfaces between the domains. Various approaches have appeared in the literature that employ interface operators that minimizes some (norm) measure of the difference between the values and/or derivatives of the fields across the interfaces. For the case of elastodynamics, Nitsche methods [3–7] and mortar methods [8–11] have been proposed and successfully applied in the literature. Other non-conforming methods have been studied and used in practice for elastic wave propagation phenomena, such as the Discontinuous Galerkin method and its extension to spectral element methods [12,13]. A drawback in Nitsche-type methods is the need for user-defined penalty parameters that must be carefully tuned to ensure stability of the formulation. In the case of mortar methods, the concepts of so-called slave-master interfaces may introduce a bias in the solution, while Lagrange multiplier fields yield a mixed-field formulation that may raise issues regarding the so-called *inf-sup* condition and introduce additional unknowns to be solved for, increasing the cost of computation.

---

[1] This Chapter has been adapted from "M. Anguiano, P. Kuberry, P. Bochev, A. Masud, Synchronous and concurrent multi-domain computing method for cloud computing platforms, SIAM J. Sci. Comput. (2021) *In press*." The copyright owner has provided written permission to reprint the work.

From a mathematical viewpoint, coupling of surfaces of two elastic bodies involves unknown kinematic variables (displacement, velocity, and acceleration) as well as the interfacial stress terms that ensure the continuum requirement of the continuity of flux. In other words, these approaches rely on a Lagrange multiplier field to enforce the constraint of continuity (zero-jump) of a primal field over the interface between multiple subdomains [14–17]. This leads to mixed-field problems that are restricted by the classical inf-sup condition on the admissible spaces of interpolation functions for approximating the displacement and stress-type quantities. Within the context of standard numerical approaches, staggered methods [18–20] have been proposed where different fields are represented via different discretizations. However, issues related to the order of accuracy and unconditional stability of the resulting method become stumbling blocks for developing stable and convergent methods.

One approach to address this issue is to eliminate or condense out the Lagrange multiplier field by some other means in order to obtain a primal field formulation. For instance, in [21] the Lagrange multiplier is eliminated algebraically before time discretization by considering the time differentiation of the displacement continuity constraint, effectively enforcing no jump in acceleration. This is achieved at the cost of requiring matrix inversions that are needed to relate old state information to traction forces between subdomains. Another successful approach has been to derive an expression for the Lagrange multiplier field that couples the subdomains by appealing to the notion of multiple physical or mathematical scales, which we will adopt in this chapter.

We employ the Variational Multiscale Discontinuous Galerkin (VMDG) method [22–24] for interface coupling of multiple subdomains. This method has several attractive features for this class of problems: (i) it is free from tunable, user-defined parameters because interface transmission conditions are consistently derived; (ii) it is unbiased in its treatment of the subdomains at either side of an interface; and (iii) it yields a primal field formulation that does not require additional degrees of freedom for Lagrange multipliers, also circumventing issues regarding the so-called *inf-sup* condition in mixed-field problems. The VMDG method developed by Masud and coworkers was first introduced in [22] in the context of quasi-static linear elasticity. This VMDG approach was formalized and rigorously applied to general coupling of other PDEs across interfaces in [23]. This chapter has been extended to accommodate interfacial damage and debonding in composites [1]; heterogenous modeling of porous media flows [25]; finite strain regime [24]; debonding, damage, and friction at finite strains [26]; and thermoelasticity [27,28].

Work in the context of infinitesimal strain elastodynamics has been presented in [2,29], where the formulation is first discretized in time using the Newmark method, and then in space using the VMDG methodology. In [2], the extension of the matrix form of the problem to generalized-α time integrator is shown, but the numerical results are obtained with the (implicit) average acceleration or trapezoidal rule (Newmark method with $\beta = 1/4$, $\gamma = 1/2$), whereas in [29] the numerical results use either an implicit (average acceleration or trapezoidal rule) or explicit (central differences $\beta = 0$, $\gamma = 1/2$) methods within the Newmark family.

As a novel contribution, we derive a VMDG formulation for linear elastodynamics for a *general* explicit time integration scheme. The derivation of the method is carried out for the semi-discrete formulation (discrete in space and continuous in time), so that the final formulation admits *any* time integration scheme of choice. Then, using an *explicit* method enables the decoupling of the solution procedure for each subdomain, while only sharing information from previous (known) solution steps, and where the VMDG method facilitates the consistent transmission conditions at the interfaces without requiring costly matrix inversions. We demonstrate numerically that optimal convergence rates are attained for the spatial and temporal errors. Moreover, additional numerical results are presented to show that the interface does not introduce spurious modes or otherwise adversely impact the stability and energy conservation attributes of the underlying spatial and temporal discretizations.

An outline of this chapter is as follows: in section 5.2 we state the strong form of the problem, in section 5.3 we cast it in the corresponding weak form, and in section 5.4 we show the variationally consistent derivation of the semidiscrete formulation, where the transmission conditions that couple dynamics across the interfaces arise naturally in this derivation. Considerations regarding time discretization are discussed in section 5.5. In section 5.6 we present numerical results to showcase accuracy, convergence, and stability for the explicit method, where computations for each subdomain are performed separately, only communicating interface information from previous time levels. Finally, conclusions are drawn in section 5.7.

5.2 Strong Form

In this chapter, we will consider a domain $\Omega \subset \mathbb{R}^d, d = 2,3$ composed of two non-overlapping, simply-connected subdomains $\Omega^{(1)}$ and $\Omega^{(2)}$ with boundaries $\partial\Omega^{(1)}$ and $\partial\Omega^{(2)}$

respectively, and sharing a common interface $\Gamma_I = \partial\Omega^{(1)} \cap \partial\Omega^{(2)}$, as shown in Figure 5.1. We assume that the rest of each domain's boundary is partitioned in Dirichlet $\Gamma_g^{(\alpha)}$ and Neumann $\Gamma_h^{(\alpha)}$ parts such that $\Gamma_g^{(\alpha)} \cup \Gamma_h^{(\alpha)} = \partial\Omega^{(\alpha)} / \Gamma_I$ and $\Gamma_g^{(\alpha)} \cap \Gamma_h^{(\alpha)} = \varnothing$ for $\alpha = 1,2$.

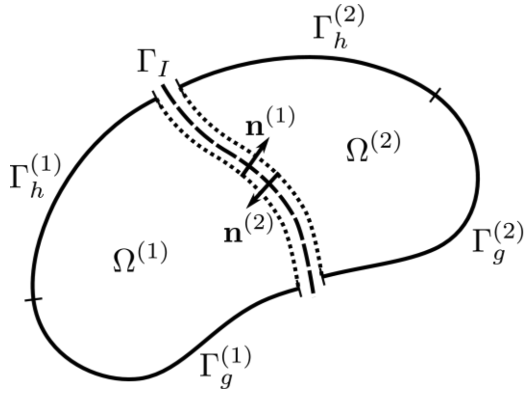


Figure 5.1. Schematic of domain, subdomains, and interface.

We want to solve the initial-boundary value problem of linear elastodynamics in $\Omega = \Omega^{(1)} \cup \Omega^{(2)}$ over a time interval $I =]0,T[$. The strong form of the problem is given by the following set of PDEs, boundary and initial conditions, and transmission conditions along Γ_I, for $\alpha = 1,2$:

$$-\rho^{(\alpha)}\ddot{\mathbf{u}}^{(\alpha)} + \nabla\cdot\boldsymbol{\sigma}^{(\alpha)} + \mathbf{b}^{(\alpha)} = \mathbf{0} \qquad \text{in } \Omega^{(\alpha)}\times]0,T[\tag{5.1}$$

$$\mathbf{u}^{(\alpha)} = \mathbf{g}^{(\alpha)}(\mathbf{x},t) \qquad \text{on } \Gamma_g^{(\alpha)}\times]0,T[\tag{5.2}$$

$$\boldsymbol{\sigma}^{(\alpha)}\mathbf{n}^{(\alpha)} = \mathbf{h}^{(\alpha)}(\mathbf{x},t) \qquad \text{on } \Gamma_h^{(\alpha)}\times]0,T[\tag{5.3}$$

$$\mathbf{u}^{(\alpha)}(\mathbf{x},0) = \mathbf{u}_0^{(\alpha)}(\mathbf{x}) \qquad \text{in } \Omega^{(\alpha)} \tag{5.4}$$

$$\dot{\mathbf{u}}^{(\alpha)}(\mathbf{x},0) = \dot{\mathbf{u}}_0^{(\alpha)}(\mathbf{x}) \qquad \text{in } \Omega^{(\boldsymbol{\alpha})} \tag{5.5}$$

$$\mathbf{u}^{(1)} - \mathbf{u}^{(2)} = [\![\mathbf{u}]\!] = \boldsymbol{\zeta}(\mathbf{x},t) \qquad \text{on } \Gamma_I \tag{5.6}$$

$$\boldsymbol{\sigma}^{(1)}\mathbf{n}^{(1)} = \boldsymbol{\lambda}(\mathbf{x},t) = -\boldsymbol{\sigma}^{(2)}\mathbf{n}^{(2)} \qquad \text{on } \Gamma_I \tag{5.7}$$

where, for subdomain α, $\rho^{(\alpha)}$ is the density, $\mathbf{u}^{(\alpha)}$ is the displacement field, $\dot{\mathbf{u}}^{(\alpha)}$ is its time derivative (i.e., velocity), $\ddot{\mathbf{u}}^{(\alpha)}$, the second time derivative (i.e., acceleration), $\mathbf{b}^{(\alpha)}$ is the body force field, $\boldsymbol{\sigma}^{(\alpha)}$ is the Cauchy stress tensor, which in the context of linear elasticity is given by $\boldsymbol{\sigma}^{(\alpha)} = \boldsymbol{\sigma}(\mathbf{u}^{(\alpha)}) = \mathbf{C}^{(\alpha)}\boldsymbol{\varepsilon}(\mathbf{u}^{(\alpha)})$, wherein $\mathbf{C}^{(\alpha)}$ is the major- and minor-symmetric, fourth-order tensor of material moduli on $\Omega^{(\alpha)}$, and $\boldsymbol{\varepsilon}(\mathbf{u}) = \frac{1}{2}(\nabla\mathbf{u} + (\nabla\mathbf{u})^T)$ is the symmetric part of the displacement gradient. We will also use the shorthand $\mathbf{t}^{(\alpha)} = \mathbf{t}(\mathbf{u}^{(\alpha)}) = \boldsymbol{\sigma}^{(\alpha)}\mathbf{n}^{(\alpha)}$ to denote the traction, where $\mathbf{n}^{(\alpha)}$ is the outward unit normal. Furthermore, $\mathbf{g}^{(\alpha)}$ is the prescribed displacement on the Dirichlet boundary, $\mathbf{h}^{(\alpha)}$, the prescribed traction on the Neumann boundary, $\mathbf{u}_0^{(\alpha)}$ and $\dot{\mathbf{u}}_0^{(\alpha)}$ are given initial conditions on the displacement and velocity fields, $\boldsymbol{\zeta}$ is a prescribed gap[2] at the interface Γ_I, and $\boldsymbol{\lambda}$ is the value of the traction across the interface that enforces the gap $[\![\mathbf{u}]\!] = \boldsymbol{\zeta}$, where the jump operator is defined, without loss of generality, as $[\![\mathbf{u}]\!] = \mathbf{u}^{(1)} - \mathbf{u}^{(2)}$.

5.3 Weak Form

The weak form of the problem is constructed by taking the inner product of vector-valued test functions $\mathbf{w}^{(\alpha)}$ with the balance of linear momentum equation and integrating over the domain, as well as taking the inner product of another vector-valued test function $\boldsymbol{\mu}$ with the continuity equation and integrating along the common interface. First, appropriate spaces of functions are defined

$$\mathcal{S}_t^{(\alpha)} = \{\, \mathbf{u}(\mathbf{x},t) \mid \mathbf{u}(\cdot,t) \in H^1(\Omega^{(\alpha)}) \text{ and } \mathbf{u} = \mathbf{g}^{(\alpha)} \text{ on } \Gamma_g^{(\alpha)} \} \tag{5.8}$$

$$\mathcal{V}^{(\alpha)} = \{\, \mathbf{w}(\mathbf{x}) \mid \mathbf{w} \in H^1(\Omega^{(\alpha)}) \text{ and } \mathbf{w} = \mathbf{0} \text{ on } \Gamma_g^{(\alpha)} \} \tag{5.9}$$

$$\mathcal{Q} = \{\, \boldsymbol{\mu}(\mathbf{x}) \mid \boldsymbol{\mu} \in H^{-1/2}(\Gamma_I) \} \tag{5.10}$$

After applying integration by parts, the weak form of the problem can be stated as follows: given $\rho^{(\alpha)}$, $\mathbf{b}^{(\alpha)}$, $\mathbf{g}^{(\alpha)}$, $\mathbf{h}^{(\alpha)}$, $\mathbf{u}_0^{(\alpha)}$, $\dot{\mathbf{u}}_0^{(\alpha)}$, and $\boldsymbol{\zeta}^{(\alpha)}$; find $\mathbf{u}^{(\alpha)} \in \mathcal{S}_t^{(\alpha)}$ and $\boldsymbol{\lambda} \in \mathcal{Q}$ such that $\forall \mathbf{w}^{(\alpha)} \in \mathcal{V}^{(\alpha)}$ and $\forall \boldsymbol{\mu} \in \mathcal{Q}$, for $\alpha = 1,2$, such that

$$\begin{aligned} &\sum_{\alpha=1}^{2}\Big[\big(\mathbf{w}^{(\alpha)} \,,\, \rho^{(\alpha)}\ddot{\mathbf{u}}^{(\alpha)}\big)_{\Omega^{(\alpha)}} + \big(\boldsymbol{\varepsilon}(\mathbf{w}^{(\alpha)}) \,,\, \boldsymbol{\sigma}(\mathbf{u}^{(\alpha)})\big)_{\Omega^{(\alpha)}}\Big] - \big([\![\mathbf{w}]\!] \,,\, \boldsymbol{\lambda}\big)_{\Gamma_I} \\ &= \sum_{\alpha=1}^{2}\Big[\big(\mathbf{w}^{(\alpha)} \,,\, \mathbf{b}^{(\alpha)}\big)_{\Omega^{(\alpha)}} + \big(\mathbf{w}^{(\alpha)} \,,\, \mathbf{h}^{(\alpha)}\big)_{\Gamma_h^{(\alpha)}}\Big] \end{aligned} \tag{5.11}$$

---

[2] The gap parameter in (5.6) may be used to introduce some interface physics, such as damage and fracture (we refer to [1,2,29] for further details). Although this gap parameter is set equal to zero and dropped in section 5.6, we have chosen to retain this term in the derivations through section 5.4 for completeness of the presentation and for consistency with the earlier work.

$$(\boldsymbol{\mu}\ ,\ \boldsymbol{\zeta}-[\![\mathbf{u}]\!])_{\Gamma_I}=0 \tag{5.12}$$

where $(\mathbf{v}\ ,\ \mathbf{f})_{\Omega}=\int_{\Omega}\mathbf{v}\cdot\mathbf{f}\,\mathrm{d}\Omega$ is the $L^2(\Omega)$ inner product of $\mathbf{f}$ and $\mathbf{v}$. Rather than numerically solving for the interfacial Lagrange multiplier field, at this point our objective is to derive $\boldsymbol{\lambda}$ in terms of the interfacial physics of the problem. Consequently, there are three objectives

1. Eliminate the Lagrange multiplier field to obtain a primal field formulation.
2. Account for kinematic and constitutive quantities from the two subdomains in the coupling terms.
3. Segregate the coupling terms such that they can be assigned independently to each subdomain and executed concurrently on separate computational platforms while still accounting for the physics from the counterpart domain running simultaneously.

5.4 Variational Multiscale Formulation

In this section, we derive the semi-discrete form (discrete in space, continuous in time) of the problem, and employ the VMS methodology to recover a primal field formulation with consistently derived transmission conditions for interface coupling of subdomains. Similar derivations can be found in [23,29], however, those derivations are either quasi-static, or assume an *a priori* time discretization, to arrive at a fully discrete form. Here, we retain a time-continuous form in order to keep formulation general so that any explicit time integration algorithm can be applied in combination with the VMDG method. We now outline the key steps of the derivation: First, the primary field (displacements) and its associated test function will be decomposed into coarse- and fine-scale contributions, which gives rise to a coarse- and a fine-scale weak subproblems of the balance of linear momentum equation. By expanding the fine-scale field in terms of bubble functions, and with the appropriate assumptions, the fine-scale problem is solved for the fine-scale field in terms of the coarse scales and $\boldsymbol{\lambda}$. Then, this solution is substituted in the weak form of the continuity equation. Arguments about the stability of the problem and the function space of $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$, the solution for $\boldsymbol{\lambda}$ is found in terms of the coarse-scale displacement field. Finally, both the fine displacement field and $\boldsymbol{\lambda}$ are substituted in the coarse-scale weak form of the balance of linear momentum equation, in which the only unknown that remains is the coarse-scale displacement field. This modified coarse-scale problem also includes interface terms containing parameters that capture the effect of the fine scales and the traction at the interface, and provide the coupling between the subdomains.

5.4.1 Multiscale Decomposition

Following the VMS framework [30,31], the displacement field and its corresponding test function are additively decomposed into linearly independent coarse-scale and fine-scale contributions as follows:

$$\mathbf{u}^{(\alpha)} = \hat{\mathbf{u}}^{(\alpha)} + \tilde{\mathbf{u}}^{(\alpha)}, \qquad \mathbf{w}^{(\alpha)} = \hat{\mathbf{w}}^{(\alpha)} + \tilde{\mathbf{w}}^{(\alpha)} \tag{5.13}$$

where the coarse scales, denoted by a " $\hat{}$ " ("hat"), are associated with the standard finite element spaces, i.e., $\hat{\mathbf{u}}^{(\alpha)} \in \mathcal{S}_t^{h(\alpha)}$ and $\hat{\mathbf{w}}^{(\alpha)} \in \mathcal{V}^{h(\alpha)}$, where $\mathcal{S}_t^{h(\alpha)}$ and $\mathcal{V}^{h(\alpha)}$ are finite dimensional subsets of $\mathcal{S}_t^{(\alpha)}$ and $\mathcal{V}^{(\alpha)}$, and they are comprised of continuous, piecewise polynomial functions with compact support for the commonly used case of Lagrangian finite elements. On the other hand, the fine scales are denoted by a " $\tilde{}$ " ("tilde"), and they are defined in the sections ahead. Differentiating the displacement field in time, the following is obtained

$$\dot{\mathbf{u}}^{(\alpha)} = \dot{\hat{\mathbf{u}}}^{(\alpha)} + \dot{\tilde{\mathbf{u}}}^{(\alpha)}, \qquad \ddot{\mathbf{u}}^{(\alpha)} = \ddot{\hat{\mathbf{u}}}^{(\alpha)} + \ddot{\tilde{\mathbf{u}}}^{(\alpha)} \tag{5.14}$$

In order to simplify the formulation, we assume that the effect of the time variation of the fine-scale field is negligible, so that

$$\dot{\mathbf{u}}^{(\alpha)} \approx \dot{\hat{\mathbf{u}}}^{(\alpha)}, \qquad \ddot{\mathbf{u}}^{(\alpha)} \approx \ddot{\hat{\mathbf{u}}}^{(\alpha)} \tag{5.15}$$

Substituting (5.13) and (5.15) into the weak form of the balance of linear momentum (5.11), and using the linearity of the operators on the test function, we segregate the problem into a coarse-scale and a fine-scale problem.

5.4.2 Fine Scale Problem

After applying integration by parts one more time, the fine scale problem in each subdomain, for $\alpha = 1,2$, can be written as:

$$\begin{aligned}\left(\boldsymbol{\varepsilon}(\tilde{\mathbf{w}}^{(\alpha)}) \,,\, \boldsymbol{\sigma}(\tilde{\mathbf{u}}^{(\alpha)})\right)_{\Omega^{(\alpha)}} &= \left(\tilde{\mathbf{w}}^{(\alpha)} \,,\, \hat{\mathbf{R}}^{(\alpha)}\right)_{\Omega^{(\alpha)}} + \left(\tilde{\mathbf{w}}^{(\alpha)} \,,\, \mathbf{h}^{(\alpha)} - \mathbf{t}(\hat{\mathbf{u}}^{(\alpha)})\right)_{\Gamma_h^{(\alpha)}} \\ &\quad + \left(\tilde{\mathbf{w}}^{(\alpha)} \,,\, (-1)^{\alpha-1}\boldsymbol{\lambda} - \mathbf{t}(\hat{\mathbf{u}}^{(\alpha)})\right)_{\Gamma_I}\end{aligned} \tag{5.16}$$

where $\hat{\mathbf{R}}^{(\alpha)}$ is the residual of the coarse-scale Euler-Lagrange equations:

$$\hat{\mathbf{R}}^{(\alpha)} = \mathbf{b}^{(\alpha)} - \rho^{(\alpha)}\ddot{\hat{\mathbf{u}}}^{(\alpha)} + \nabla \cdot \boldsymbol{\sigma}(\hat{\mathbf{u}}^{(\alpha)}) \tag{5.17}$$

In general, the exact fine-scale field belongs to an infinite-dimensional function space. However, in order to develop the numerical method and find a solution of (5.16) for the fine-scale field, we restrict the fine-scales to a finite dimensional space in the following sections.

5.4.3 Bubble Function

Before defining a basis for the fine scale space, we introduce some definitions that will be helpful in describing the regions of support for such basis. We consider the discretized interface Γ_I, which is partitioned into segments γ_s defined by the discretizations at either side of the interface, such that $\bigcup \gamma_s = \tilde{\Gamma}_s$. The γ_s are line segments in 2D, and general polygons in 3D[3]. The segments γ_s in turn are used to define sectors $\omega_s^{(\alpha)}$ that lie in the interior of elements at either side of the interface. This is illustrated for 2D triangular elements in Figure 5.2. The segments γ_s and sectors $\omega_s^{(\alpha)}$ are defined analogously for 2D quadrilaterals and 3D elements. Furthermore, the sectors $\omega_s^{(\alpha)}$ are re-scaled in the direction orthogonal to γ_s to maintain acceptable aspect ratios (refer to [23] for further details).

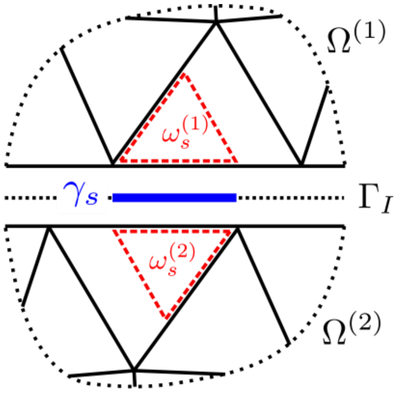


Figure 5.2. Interface segment γ_s, and sectors $\omega_s^{(\alpha)}$.

We approximate the spatial variation of the fine scales by bubble functions $b_s^{(\alpha)}(\mathbf{x})$ defined on sectors $\omega_s^{(\alpha)}$ as follows:

---

[3] Identification of interface segments required for numerical integration is a nontrivial problem for general interface geometries, especially in 3D. Earlier works on VMDG [22], as well as the present work, adopt ideas from [37]. Although not employed in this work, we note that there exist open source tools that can be used to generate the required common mesh refinement at the interface, such as R3D [38], TempestRemap [39], and MOAB [40].

$$\tilde{\mathbf{u}}^{(\alpha)}\Big|_{\omega_s^{(\alpha)}} = b_s^{(\alpha)}\bar{\mathbf{u}}_s^{(\alpha)}, \qquad \tilde{\mathbf{w}}^{(\alpha)}\Big|_{\omega_s^{(\alpha)}} = b_s^{(\alpha)}\bar{\mathbf{w}}_s^{(\alpha)} \tag{5.18}$$

Here, $\bar{\mathbf{w}}_s^{(\alpha)}$ and $\bar{\mathbf{w}}_s^{(\alpha)}$ are constant coefficient vectors. The fine scales $\tilde{\mathbf{u}}$ are thus spanned by the basis $\{b_s\mathbf{e}_i\}$, where $\{\mathbf{e}_i\}_{i=1}^{d}$ is the canonical basis on $\mathbb{R}^d$. We define $b_s^{(\alpha)}$ to be zero at the boundaries of $\omega_s^{(\alpha)}$ that are not part of the interface Γ_I (it is nonzero at γ_s). For simplicity, we refer to these as edge bubble functions, although strictly speaking they are not the 1D bubble function associated with γ_s. A sample edge bubble function is shown in Figure 5.3.

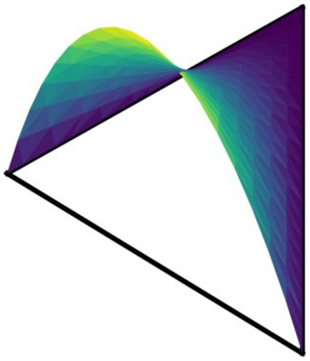


Figure 5.3. Sample edge bubble function for 3-node triangular element (T3).

5.4.4 Solution of the Fine Scale Problem

We assume, along the lines of [23], that the coarse scales residual $\hat{\mathbf{R}}^{(\alpha)}$ is nearly orthogonal to the fine scale bubble functions. This is a reasonable assumption since the residual is expected to be small for a stable solution, and the contribution of the residual in the domain interior term is small compared to the interface contribution.

$$\left(\tilde{\mathbf{w}}^{(\alpha)}\ ,\ \hat{\mathbf{R}}^{(\alpha)}\right)_{\Omega^{(\alpha)}} \approx 0 \tag{5.19}$$

In addition, note that by the definition of the bubble function support, $b_s^{(\alpha)}\Big|_{\Gamma_h} = 0$, then:

$$\left(\tilde{\mathbf{w}}^{(\alpha)}\ ,\ \mathbf{h}^{(\alpha)} - \mathbf{t}(\hat{\mathbf{u}}^{(\alpha)})\right)_{\Gamma_h} = 0 \tag{5.20}$$

Hence, using (5.19) and (5.20), and using the fact that the fine-scale basis consists of bubbles with local, non-overlapping support, the fine scale problem is localized and reduced to the following

$$\left(\boldsymbol{\varepsilon}(b_s^{(\alpha)}\bar{\mathbf{w}}_s^{(\alpha)})\ ,\ \boldsymbol{\sigma}(b_s^{(\alpha)}\bar{\mathbf{u}}_s^{(\alpha)})\right)_{\omega_s^{(\alpha)}} \quad = \quad \left(b_s^{(\alpha)}\bar{\mathbf{w}}_s^{(\alpha)}\ ,\ (-1)^{\alpha-1}\boldsymbol{\lambda} - \mathbf{t}(\hat{\mathbf{u}}^{(\alpha)})\right)_{\gamma_s} \quad \forall\bar{\mathbf{w}}_s^{(\alpha)} \tag{5.21}$$

Solving for the fine-scale coefficients gives

$$\bar{\mathbf{u}}_s^{(\alpha)} = \left[\left(\boldsymbol{\varepsilon}(b_s^{(\alpha)}\mathbf{e}_i) \,,\, \boldsymbol{\sigma}(b_s^{(\alpha)}\mathbf{e}_j) \right)_{\omega_s^{(\alpha)}} \mathbf{e}_i \otimes \mathbf{e}_j \right]^{-1} \left(b_s^{(\alpha)} \,,\, (-1)^{\alpha-1}\boldsymbol{\lambda} - \mathbf{t}(\hat{\mathbf{u}}^{(\alpha)}) \right)_{\gamma_s} \tag{5.22}$$

Recalling (5.18), the fine scales are recovered by multiplying the fine scale coefficient, given by (5.22), with the corresponding bubble function $b_s^{(\alpha)}$. In addition, the following approximations are made:

1. We use the mean value theorem (MVT) for integrals to "pull out" the traction terms from the boundary integral term in the fine scale coefficient, and approximate those traction terms with their fully varying form instead of evaluating them at the point guaranteed by the MVT.

$$\begin{aligned} \left(b_s^{(\alpha)} \,,\, (-1)^{a-1}\boldsymbol{\lambda} - \mathbf{t}(\hat{\mathbf{u}}^{(\alpha)}) \right)_{\gamma_s} &= \int_{\gamma_s} b_s^{(\alpha)} \mathrm{d}\Gamma \left[(-1)^{a-1}\boldsymbol{\lambda} - \mathbf{t}(\hat{\mathbf{u}}^{(\alpha)}) \right]\Big|_{\mathbf{x}=\mathbf{c}\in\gamma_s} \\ &\approx \int_{\gamma_s} b_s^{(\alpha)} \mathrm{d}\Gamma \left[(-1)^{a-1}\boldsymbol{\lambda} - \mathbf{t}(\hat{\mathbf{u}}^{(\alpha)}) \right]\Big|_{\gamma_s} \end{aligned} \tag{5.23}$$

2. We approximate the value of the bubble at the segment γ_s by its average value.

$$b_s^{(\alpha)}\Big|_{\gamma_s} \approx \left[\mathrm{meas}(\gamma_s) \right]^{-1} \int_{\gamma_s} b_s^{(\alpha)} \mathrm{d}\Gamma \tag{5.24}$$

With these approximations, we find the following expression for the fine scales over each interface segment γ_s:

$$\tilde{\mathbf{u}}^{(\alpha)}\Big|_{\gamma_s} = \boldsymbol{\tau}_s^{(\alpha)} \left[(-1)^{\alpha-1}\boldsymbol{\lambda} - \mathbf{t}(\hat{\mathbf{u}}^{(\alpha)}) \right]\Big|_{\gamma_s} \tag{5.25}$$

where,

$$\boldsymbol{\tau}_s^{(\alpha)} = \left[\left(\boldsymbol{\varepsilon}(b_s^{(\alpha)}\mathbf{e}_i) \,,\, \boldsymbol{\sigma}(b_s^{(\alpha)}\mathbf{e}_j) \right)_{\Omega^{(\alpha)}} \mathbf{e}_i \otimes \mathbf{e}_j \right]^{-1} \left[\mathrm{meas}(\gamma_s) \right]^{-1} \left(\int_{\gamma_s} b_s^{(\alpha)} \mathrm{d}\Gamma \right)^2 \tag{5.26}$$

Remark: For purposes of the coarse-scale problem discussed in subsection 5.4.6, it is assumed that the domain interior contributions of the fine scale are negligible compared to the interface contributions. Thus, the expression for the fine scales over $\omega_s^{(\alpha)}$ is inconsequential.

Remark: The approximations serve the purpose of leading to an expression where the fine scales are expressed in terms of the constant stabilizing tensor $\boldsymbol{\tau}_s^{(a)}$, and a varying part containing the unknown interfacial traction $\boldsymbol{\lambda}$, over each γ_s.

At this point, the fine scales are expressed purely in terms of the coarse scale displacements, the interfacial traction $\boldsymbol{\lambda}$, and the stabilizing parameter $\boldsymbol{\tau}_s^{(\alpha)}$, which contains geometric and material information.

5.4.5 Solution for the Lagrange multiplier

We substitute the multiscale decomposition from equation (5.13) into the weak form of the continuity equation (5.12) to obtain

$$\left(\boldsymbol{\mu}\ ,\ \boldsymbol{\zeta} - [\![\hat{\mathbf{u}}]\!] - [\![\tilde{\mathbf{u}}]\!]\right)_{\Gamma_I} = 0 \qquad \forall \boldsymbol{\mu} \in \mathcal{Q} \tag{5.27}$$

In mixed field problems, the VMS framework improves the stability of the problem and allows to circumvent issues with the Babuska-Brezzi (inf-sup) condition. Therefore, with relaxed conditions on the discrete Lagrange multiplier space, we choose it to correspond to piecewise constant functions on $\tilde{\Gamma}_I$, constant over each segment γ_s. Thus, equation (5.27) can be localized over each segment γ_s to yield

$$\left[\boldsymbol{\zeta} - [\![\hat{\mathbf{u}}]\!] - [\![\tilde{\mathbf{u}}]\!]\right]\Big|_{\gamma_s} = \mathbf{0} \tag{5.28}$$

Expanding the fine-scale jump term and substituting the fine-scale expression from equation (5.25) results in

$$\left[\boldsymbol{\zeta} - [\![\hat{\mathbf{u}}]\!] - \boldsymbol{\tau}_s^{(1)}\left(\boldsymbol{\lambda} - \mathbf{t}(\hat{\mathbf{u}}^{(1)})\right) + \boldsymbol{\tau}_s^{(2)}\left(-\boldsymbol{\lambda} - \mathbf{t}(\hat{\mathbf{u}}^{(2)})\right)\right]\Big|_{\gamma_s} = \mathbf{0} \tag{5.29}$$

This is solved for $\boldsymbol{\lambda}$ to obtain the following expression for the Lagrange multiplier at each segment γ_s

$$\boldsymbol{\lambda}\big|_{\gamma_s} = \boldsymbol{\tau}_s\left(\boldsymbol{\zeta} - [\![\hat{\mathbf{u}}]\!]\right) + \left\{\mathbf{t}(\hat{\mathbf{u}})\right\} \tag{5.30}$$

where,

$$\boldsymbol{\tau}_s = \left(\boldsymbol{\tau}_s^{(1)} + \boldsymbol{\tau}_s^{(2)}\right)^{-1} \tag{5.31}$$

$$\left\{\mathbf{t}(\hat{\mathbf{u}})\right\} = \boldsymbol{\delta}^{(1)}\boldsymbol{\sigma}^{(1)}\mathbf{n}^{(1)} - \boldsymbol{\delta}^{(2)}\boldsymbol{\sigma}^{(2)}\mathbf{n}^{(2)} \tag{5.32}$$

$$\boldsymbol{\delta}^{(\alpha)} = \boldsymbol{\tau}_s\boldsymbol{\tau}_s^{(\alpha)} \tag{5.33}$$

Remark: It follows from the definition of $\boldsymbol{\delta}^{(\alpha)}$ that $\boldsymbol{\delta}^{(1)} + \boldsymbol{\delta}^{(2)} = \mathbf{I}$

5.4.6 Modified Coarse Scale Problem

The coarse scale problem in equation (5.34) follows along the same lines as the fine scale one by using the linearity of the weak forms with respect to test functions and the independence of the scales.

$$\begin{aligned}&\sum_{\alpha=1}^{2}\left[\left(\hat{\mathbf{w}}^{(\alpha)}\ ,\ \rho^{(\alpha)}\ddot{\mathbf{u}}^{(\alpha)}\right)_{\Omega^{(\alpha)}}+\left(\boldsymbol{\varepsilon}(\hat{\mathbf{w}}^{(\alpha)})\ ,\ \boldsymbol{\sigma}(\mathbf{u}^{(\alpha)})\right)_{\Omega^{(\alpha)}}-\left([\![\hat{\mathbf{w}}]\!]\ ,\ \boldsymbol{\lambda}\right)_{\Gamma_I}\right]\\ =\ &\sum_{\alpha=1}^{2}\left[\left(\hat{\mathbf{w}}^{(\alpha)}\ ,\ \mathbf{b}^{(\alpha)}\right)_{\Omega^{(\alpha)}}+\left(\hat{\mathbf{w}}^{(\alpha)}\ ,\ \mathbf{h}^{(\alpha)}\right)_{\Gamma_h^{(\alpha)}}\right]\end{aligned}\tag{5.34}$$

Next, equation (5.13) is substituted into (5.34) and integration by parts is applied on the fine-scale term. Recall that, by definition, the fine scales vanish on $\Gamma_h^{(\alpha)}$ and on element boundaries not coincident with the interface Γ_I. Finally, denoting the union of element interiors in subdomain α as $\tilde{\Omega}^{(\alpha)}$, one obtains the following

$$\begin{aligned}&\sum_{\alpha=1}^{2}\left[\left(\hat{\mathbf{w}}^{(\alpha)}\ ,\ \rho^{(\alpha)}\ddot{\hat{\mathbf{u}}}^{(\alpha)}\right)_{\Omega^{(\alpha)}}+\left(\boldsymbol{\varepsilon}(\hat{\mathbf{w}}^{(\alpha)})\ ,\ \boldsymbol{\sigma}(\hat{\mathbf{u}}^{(\alpha)})\right)_{\Omega^{(\alpha)}}\right]-\left([\![\hat{\mathbf{w}}]\!]\ ,\ \boldsymbol{\lambda}\right)_{\Gamma_I}\\ +\ &\sum_{\alpha=1}^{2}\left[-\left(\nabla\cdot\boldsymbol{\sigma}(\hat{\mathbf{w}}^{(\alpha)})\ ,\ \tilde{\mathbf{u}}^{(\alpha)}\right)_{\tilde{\Omega}^{(\alpha)}}+\left(\mathbf{t}(\hat{\mathbf{w}}^{(\alpha)})\ ,\ \tilde{\mathbf{u}}^{(\alpha)}\right)_{\Gamma_I}\right]\\ =\ &\sum_{\alpha=1}^{2}\left[\left(\hat{\mathbf{w}}^{(\alpha)}\ ,\ \mathbf{b}^{(\alpha)}\right)_{\Omega^{(\alpha)}}+\left(\hat{\mathbf{w}}^{(\alpha)}\ ,\ \mathbf{h}^{(\alpha)}\right)_{\Gamma_h^{(\alpha)}}\right]\end{aligned}\tag{5.35}$$

We assume that the element interior contributions of the fine scales are negligible in comparison with their interface contribution, and therefore that term is dropped. Note that the support of the fine scales is defined only on sectors adjacent to the interface Γ_I. Furthermore, the element interior term is exactly zero for linear elements. The fine-scale expression from (5.25) is then substituted in (5.35) to yield

$$\begin{aligned}&\sum_{\alpha=1}^{2}\left[\left(\hat{\mathbf{w}}^{(\alpha)}\ ,\ \rho^{(\alpha)}\ddot{\hat{\mathbf{u}}}^{(\alpha)}\right)_{\Omega^{(\alpha)}}+\left(\boldsymbol{\varepsilon}(\hat{\mathbf{w}}^{(\alpha)})\ ,\ \boldsymbol{\sigma}(\hat{\mathbf{u}}^{(\alpha)})\right)_{\Omega^{(\alpha)}}\right]-\left([\![\hat{\mathbf{w}}]\!]\ ,\ \boldsymbol{\lambda}\right)_{\Gamma_I}\\ +\ &\sum_{\alpha=1}^{2}\left[\left(\mathbf{t}(\hat{\mathbf{w}}^{(\alpha)})\ ,\ \boldsymbol{\tau}_s^{(\alpha)}\left[(-1)^{\alpha-1}\boldsymbol{\lambda}-\mathbf{t}(\hat{\mathbf{u}}^{(\alpha)})\right]\right)_{\Gamma_I}\right]\\ =\ &\sum_{\alpha=1}^{2}\left[\left(\hat{\mathbf{w}}^{(\alpha)}\ ,\ \mathbf{b}^{(\alpha)}\right)_{\Omega^{(\alpha)}}+\left(\hat{\mathbf{w}}^{(\alpha)}\ ,\ \mathbf{h}^{(\alpha)}\right)_{\Gamma_h^{(\alpha)}}\right]\end{aligned}\tag{5.36}$$

By substituting the expression for the Lagrange multiplier from (5.30), the only unknown fields in the resulting modified coarse-scale problem are the coarse-scale displacement and acceleration.

Expanding and rearranging the interface terms, the modified coarse-scale problem can be written as

$$
\begin{aligned}
&\sum_{\alpha=1}^{2}\left[\left(\hat{\mathbf{w}}^{(\alpha)}\ ,\ \rho^{(\alpha)}\ddot{\hat{\mathbf{u}}}^{(\alpha)}\right)_{\Omega^{(\alpha)}}+\left(\boldsymbol{\varepsilon}(\hat{\mathbf{w}}^{(\alpha)})\ ,\ \boldsymbol{\sigma}(\hat{\mathbf{u}}^{(\alpha)})\right)_{\Omega^{(\alpha)}}\right]\\
-\ &\left([\![\hat{\mathbf{w}}]\!]\ ,\ \boldsymbol{\tau}_s(\boldsymbol{\zeta}-[\![\hat{\mathbf{u}}]\!])\right)_{\Gamma_I}\ -\ \left([\![\hat{\mathbf{w}}]\!]\ ,\ \{\mathbf{t}(\hat{\mathbf{u}})\}\right)_{\Gamma_I}\\
+\ &\left(\{\mathbf{t}(\hat{\mathbf{w}})\}\ ,\ (\boldsymbol{\zeta}-[\![\hat{\mathbf{u}}]\!])\right)_{\Gamma_I}\ -\ \left([\![\mathbf{t}(\hat{\mathbf{w}})]\!]\ ,\ \boldsymbol{\delta}_s[\![\mathbf{t}(\hat{\mathbf{u}})]\!]\right)_{\Gamma_I}\\
=\ &\sum_{\alpha=1}^{2}\left[\left(\hat{\mathbf{w}}^{(\alpha)}\ ,\ \mathbf{b}^{(\alpha)}\right)_{\Omega^{(\alpha)}}+\left(\hat{\mathbf{w}}^{(\alpha)}\ ,\ \mathbf{h}^{(\alpha)}\right)_{\Gamma_h^{(\alpha)}}\right]
\end{aligned}
\tag{5.37}
$$

where we have used the symmetry of the stabilizing tensors, the fact that $\boldsymbol{\delta}_s^{(1)}+\boldsymbol{\delta}_s^{(2)}=\mathbf{I}$ and the identity

$$\boldsymbol{\delta}_s=\boldsymbol{\tau}_s^{(1)}\boldsymbol{\delta}_s^{(2)}=\boldsymbol{\tau}_s^{(2)}\boldsymbol{\delta}_s^{(1)}=\left[\left(\boldsymbol{\tau}_s^{(1)}\right)^{-1}+\left(\boldsymbol{\tau}_s^{(2)}\right)^{-1}\right]^{-1}\tag{5.38}$$

which is straight-forward to verify through algebraic manipulations. In addition, the traction jump was defined for ease of notation as

$$[\![\mathbf{t}]\!]=\mathbf{t}^{(1)}+\mathbf{t}^{(2)}=\boldsymbol{\sigma}^{(1)}\mathbf{n}^{(1)}+\boldsymbol{\sigma}^{(2)}\mathbf{n}^{(2)}\tag{5.39}$$

Remark: The traction jump term is neglected in the implementations discussed in [23,29]. In this chapter, we have chosen to retain it; however, the numerical cases presented were solved with and without the traction terms, and the error and the convergence rates are comparable for both implementations.

Remark: In this chapter, we will only consider cases with $\boldsymbol{\zeta}=\mathbf{0}$. Allowing a nonzero prescribed gap is a feature that can be employed to develop more elaborate coupling methods, and subsequently, embed sophisticated constitutive models at the interface[4].

To recapitulate, the semi-discrete form in equation (5.37) (discrete in space, continuous in time) was obtained via the VMDG method and the interfacial terms on the left hand side of (5.37) are the VMDG interface terms. Equation (5.37) represents a system of ordinary differential equations and, taking $\boldsymbol{\zeta}=\mathbf{0}$, can be written in matrix form by means of the standard FEM assembly process as

---

[4] See comments in footnote 2.

$$\begin{bmatrix} \mathbf{M}^{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}^{(2)} \end{bmatrix} \begin{bmatrix} \ddot{\mathbf{d}}^{(1)}(t) \\ \ddot{\mathbf{d}}^{(2)}(t) \end{bmatrix} + \begin{bmatrix} \mathbf{K}^{(1)} + \mathbf{K}_{\Gamma_I}^{(11)} & \mathbf{K}_{\Gamma_I}^{(12)} \\ \mathbf{K}_{\Gamma_I}^{(21)} & \mathbf{K}^{(2)} + \mathbf{K}_{\Gamma_I}^{(22)} \end{bmatrix} \begin{bmatrix} \mathbf{d}^{(1)}(t) \\ \mathbf{d}^{(2)}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{F}^{(1)}(t) \\ \mathbf{F}^{(2)}(t) \end{bmatrix} \tag{5.40}$$

where, for subdomain α, $\mathbf{d}^{(\alpha)}(t)$ is the subdomain-level vector of (time-dependent) nodal displacements, and $\mathbf{M}^{(\alpha)}$ and $\mathbf{K}^{(\alpha)}$ are the standard mass and stiffness matrices, while the $\mathbf{K}_{\Gamma_I}$ terms are matrix operators that correspond to the interface terms in (5.37).

Remark: Equation (5.40) was written in this form to simplify notation and presentation of ideas. However, the interface terms in (5.37) are associated only with degrees of freedom (DOFs) near the interface Γ_I, thus the $\mathbf{K}_{\Gamma_I}$ matrices in (5.40) have very few non-zero entries, and a practical implementation of the method can exploit this fact.

Remark: The off-diagonal block matrices $\mathbf{K}_{\Gamma_I}^{(12)}$ and $\mathbf{K}_{\Gamma_I}^{(21)}$ in equation (5.40) provide the coupling between subdomain (1) and subdomain (2). If an implicit time integration algorithm is used to find the solution of (5.40), the off-diagonal $\mathbf{K}_{\Gamma_I}$ terms appear on the left hand side matrix. Then, one would require solving the system of linear equations for both subdomains in a monolithic way. On the other hand, if an explicit time integration algorithm is used, then $\mathbf{d}^{(1)}(t)$ and $\mathbf{d}^{(2)}(t)$ are replaced by values from the known data pool (so-called "predictors"). Then, the stiffness term is moved to the right-hand side of the equation, and only the mass matrix term remains on the left-hand side. Note that in the mass matrix for the entire system, the off-diagonal block matrices are zero. Therefore, solution for the acceleration vector of the entire system can be decoupled into the solution for the acceleration vector of each subdomain. Because of this decoupling, the solution step can be performed concurrently on separate computational platforms for each subdomain. This is further discussed in the following section.

5.5 Time Discretization

In seeking numerical solutions to the system in equation (5.40), the time interval of interest $]0,T[$ is partitioned into N subintervals (t_n, t_{n+1}), with $t_n < t_{n+1}$, $n \in \mathbb{N}$, $n < N$. In this chapter, we consider all time subintervals to be of equal size $t_{n+1} - t_n = \Delta t = T / N$. The ideas here can be extended for variable time step during the solution procedure. In addition, another extension is the case of different time step size for each subdomain. The approximations for the acceleration, velocities and displacements at a given time level n are denoted as

$$\mathbf{a}_n^{(\alpha)} \approx \ddot{\mathbf{d}}^{(\alpha)}(t_n) \qquad \mathbf{v}_n^{(\alpha)} \approx \dot{\mathbf{d}}^{(\alpha)}(t_n) \qquad \mathbf{d}_n^{(\alpha)} \approx \mathbf{d}^{(\alpha)}(t_n) \tag{5.41}$$

Within the time marching scheme, equation (5.40) can be used to solve algebraically for the nodal accelerations at a given step (or fractional step) n in terms of the displacements. When the displacements are replaced by quantities that belong to the *known* data pool at that point of the solution procedure, the method becomes explicit. In this case, the solution of the subdomain acceleration vectors is decoupled and can be carried out separately, which makes the method amenable for parallel computation on multiple platforms concurrently. The *decoupled* system of equations is given by

$$\mathbf{a}_n^{(1)} = \left(\mathbf{M}^{(1)}\right)^{-1}\left[\mathbf{F}^{(1)}(t_n) - \left(\mathbf{K}^{(1)} + \mathbf{K}_{\Gamma_I}^{(11)}\right)\breve{\mathbf{d}}_n^{(1)} - \mathbf{K}_{\Gamma_I}^{(12)}\breve{\mathbf{d}}_n^{(2)}\right] \tag{5.42}$$

$$\mathbf{a}_n^{(2)} = \left(\mathbf{M}^{(2)}\right)^{-1}\left[\mathbf{F}^{(2)}(t_n) - \left(\mathbf{K}^{(2)} + \mathbf{K}_{\Gamma_I}^{(22)}\right)\breve{\mathbf{d}}_n^{(2)} - \mathbf{K}_{\Gamma_I}^{(21)}\breve{\mathbf{d}}_n^{(1)}\right] \tag{5.43}$$

where the superposed "$\breve{\ }$" ("breve") is used to indicate that, at the moment of solving for $\mathbf{a}^{(\alpha)}$ at step (or fractional step) n, $\breve{\mathbf{d}}_n^{(\alpha)}$ in (5.42) and (5.43) is given in terms of known quantities.

Remark: If a non-diagonal mass matrix is used, it needs to be factored only once at the beginning of the solution procedure. If the mass matrix is diagonalized, the method is fully explicit.

Remark: Prior to solving for $\mathbf{a}_n$ in subdomain (1), it is required to pass the known $\breve{\mathbf{d}}_n^{(2)}$ from subdomain (2) to (1), and vice-versa for $\mathbf{a}_n^{(2)}$.

Therefore, the form given by (5.42) and (5.43) accommodates standard explicit stencils, such as central differences, and methods that use estimates at fractional steps, such as explicit Runge-Kutta methods. Furthermore, other explicit methods can be constructed through the notion of a predictor-corrector formulation [32]. Starting from an implicit method, the displacement can be replaced by an estimate (*predictor*) in terms of known quantities in order to solve for the accelerations, and subsequently employ the original formula to recover the displacements (*corrector*). This gives rise to some split step methods. For instance, starting with an (implicit) backward difference stencil (5.44), an explicit split step method is obtained through a predictor-corrector formulation

$$\mathbf{a}_{n+1} = \frac{\mathbf{d}_{n+1} - 2\mathbf{d}_n + \mathbf{d}_{n-1}}{\Delta t^2}, \qquad \mathbf{v}_{n+1} = \frac{\mathbf{d}_{n+1} - \mathbf{d}_n}{\Delta t} \tag{5.44}$$

Rewriting to isolate $\mathbf{d}_{n+1}$ defines the "corrector" for the displacement, in terms of the acceleration at time level $n+1$

$$\mathbf{d}_{n+1} = \Delta t^2 \mathbf{a}_{n+1} + 2\mathbf{d}_n - \mathbf{d}_{n-1} \tag{5.45}$$

The predictor is set in terms of the known data pool, for instance $\breve{\mathbf{d}}_{n+1} = \mathbf{d}_n$ (other estimates can also be used). This results in a split method because in using (5.42) and (5.43) to solve for $\mathbf{a}_{n+1}$, the displacement corresponds to time level n, while the external forcing terms are for time level $n+1$. Finally, $\mathbf{d}_{n+1}$ is obtained by using the corrector formula from equation (5.45).

5.6 Numerical Results

In this section, we consider numerical test cases in 2D to showcase the consistency, accuracy, and stability of the VMS interface coupling formulation for explicit elastodynamics. Throughout this section, the material parameters used are the same for both subdomains, with $\rho = 1$, Young's modulus $E = 2$, and Poison's ratio $\nu = 0.3$, or equivalently, Lamé parameters $\mu = 1.154$ and $\lambda = 0.7692$. The discretizations $\mathcal{T}_h$ of the polygonal domains consist of meshes comprised of three-node triangular (T3) elements (unless specified otherwise), K, where $\mathcal{T}_h$ has an associated mesh size parameter h defined as

$$h = \max_{K \in \mathcal{T}_h} (\mathrm{diam}(K)) \tag{5.46}$$

In addition, the time interval is divided into N equal time steps, N being the smallest positive integer such that

$$\Delta t = T / N \le \frac{1}{2} \cdot \frac{h_{min}}{2\mu + \lambda} \tag{5.47}$$

where $h_{min} = \min_{K \in \mathcal{T}_h} (\mathrm{diam}(K))$. The bubble function for a triangular sector $\omega_s^{(\alpha)}$ is defined in natural coordinates as $b_s^{(\alpha)} = 4\eta(1 - \xi - \eta)$ (see Figure 5.3) taking the parent axis ξ to be parallel to γ_s. The method is implemented in FreeFEM [33], a high-level FEM software, and the systems associated with each subdomain are solved separately utilizing diagonal mass matrices (lumped mass) for each subdomain.

5.6.1 Convergence

A numerical convergence study is performed for two domain geometries: the first one is formed by the union of two circles of unit diameter, one centered at the origin and the other shifted right by $\sqrt{2}/2$, shown in Figure 5.4a. The second case is a unit square with its bottom-left corner located at $(0,1/2)$ and subdivided by a chevron interface, as shown in Figure 5.4b. In both cases, the discretizations for each subdomain are non-matching (but conforming) at the interface Γ_I. Progressively refined meshes are generated by splitting every element into 4 identical elements, so that h and h_{min} (and Δt as per equation (5.47)) are halved at each refinement level. The meshes shown in Figure 5.4 correspond to the first (coarsest) level of refinement.

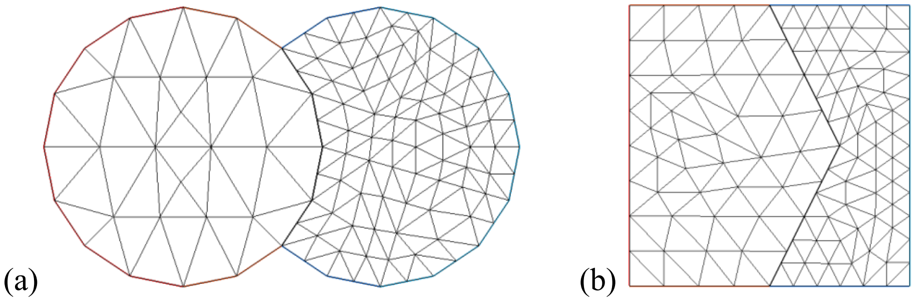


Figure 5.4. Domains partitioned by an interface and different discretizations in each subdomain: (a) "Circles" and (b) "Chevron".

We apply initial and boundary conditions, as well as body force field, consistent with the following manufactured solution

$$\mathbf{u}^{ex}(\mathbf{x},t) = \sin(x)\sin(y)\left[\cos(\pi t) \quad t\right]^T \tag{5.48}$$

In the case of the "circles" domain, the Dirichlet boundary is defined to be the quarter of the circumference farthest from the common interface for both subdomains (i.e., leftmost and rightmost arcs), and at the vertical sides for the case of the square domain (i.e., at $x = 0.5$ and $x = 1.5$). In addition, Neumann boundary conditions are applied to the top and bottom quarter circumference of the subdomains of the two circles, and at the horizontal sides for the square (i.e., $y = 0$ and $y = 1$). Solution is carried out for the time interval of interest $]0,T[$ with $T = 0.5$. Two explicit time integration schemes are used: first is central differences and the second one is a split step method based on a predictor-corrector modification of backward differences, discussed in section 5.5. Initial conditions, and information about time steps previous to $t = 0$ (needed to "kickstart" the time integration algorithm), are computed based on the exact solution. Error is measured by its L^2-norm and H^1-seminorm in space, and its L^2-norm in time. More precisely, we define

$$|||\,\mathbf{e}\,|||_{H1L2}=\left(\sum_{n=1}^{N}\Delta t\,|\,\mathbf{u}(\mathbf{x},t_n)-\mathbf{u}^{ex}(\mathbf{x},t_n)\,|_{H^1(\Omega)}^2\right)^{\frac{1}{2}} \tag{5.49}$$

$$|||\,\mathbf{e}\,|||_{L2L2}=\left(\sum_{n=1}^{N}\Delta t\,\|\,\mathbf{u}(\mathbf{x},t_n)-\mathbf{u}^{ex}(\mathbf{x},t_n)\,\|_{L^2(\Omega)}^2\right)^{\frac{1}{2}} \tag{5.50}$$

The convergence of the error measures in equations (5.49) and (5.50) is plotted in Figure 5.5 and Figure 5.6.

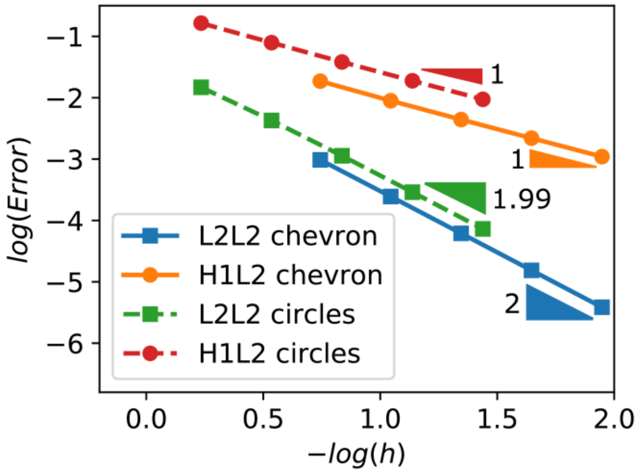


Figure 5.5. Error convergence plots with central difference time integration algorithm.

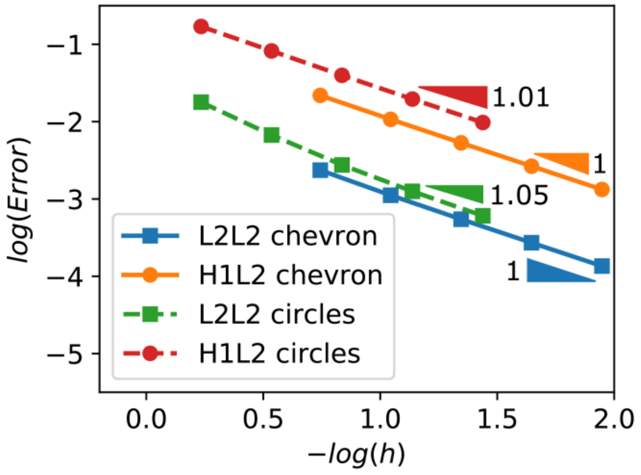


Figure 5.6. Error convergence plots with modified backward difference time integration algorithm.

The rates of convergence correspond to the optimal asymptotic rates expected. Note that the asymptotic rate is governed by the order of accuracy of the least accurate discretization, either the spatial or temporal one. In particular, the method based on backward differences is first order accurate, therefore, the convergence rates in Figure 5.6 are at most 1. On the other hand, the central differences method is second order accurate and the rates in Figure 5.5 are at most 2, but the rates associated with the spatial $H^1(\Omega)$-seminorm remain order 1 (consistent with linear shape functions).

Figure 5.7 and Figure 5.8 show contours of the magnitude of the error $\mathbf{e} = \mathbf{u} - \mathbf{u}^{ex}$ at time $t = T$. They correspond to the solution with central difference time integration and the third level of refinement for each computational domain. Note that the magnitude of the error near the interface is on the same order as the magnitude of the error elsewhere in the domain. In other words, the interface formulation does not introduce errors beyond the error inherent in the discretization.

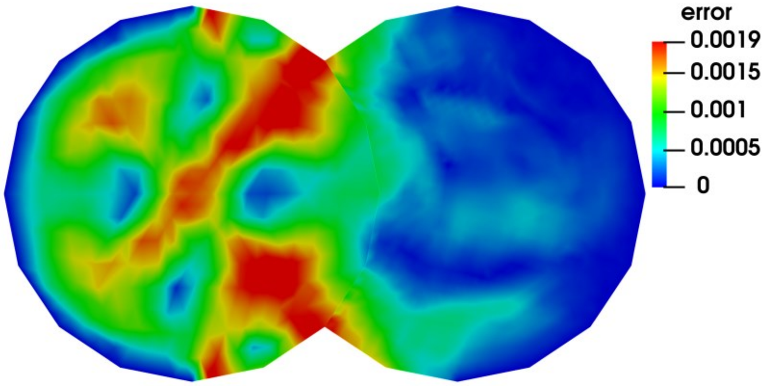


Figure 5.7. Distribution of error magnitude in "circles" domain.

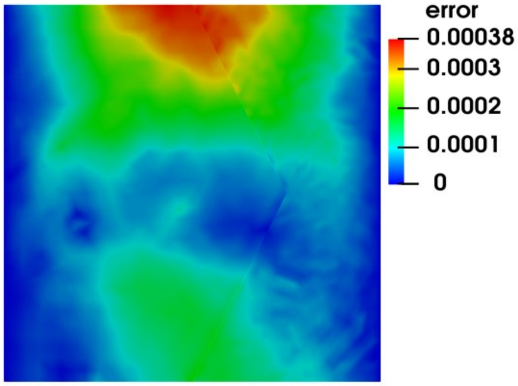


Figure 5.8. Distribution of error magnitude in "chevron" domain.

As a further test of the robustness of the explicit VMDG method, we now consider a domain with a chevron interface (identical to Figure 5.4b), but with a 1:10 element size ratio between the left and right subdomains as shown in Figure 5.9a. For this computational domain, we also apply initial and boundary conditions, as well as body force field, consistent with the same manufactured solution as in equation (5.48). We study convergence of the numerical solution with mesh refinement, while maintaining the 1:10 element size ratio at the interface. Central difference method is used for time integration and the time step size is selected according to (5.47). Optimal rates of convergence are still obtained: quadratic in the L2L2 norm and linear in the H1L2 norm (see Figure 5.9b).

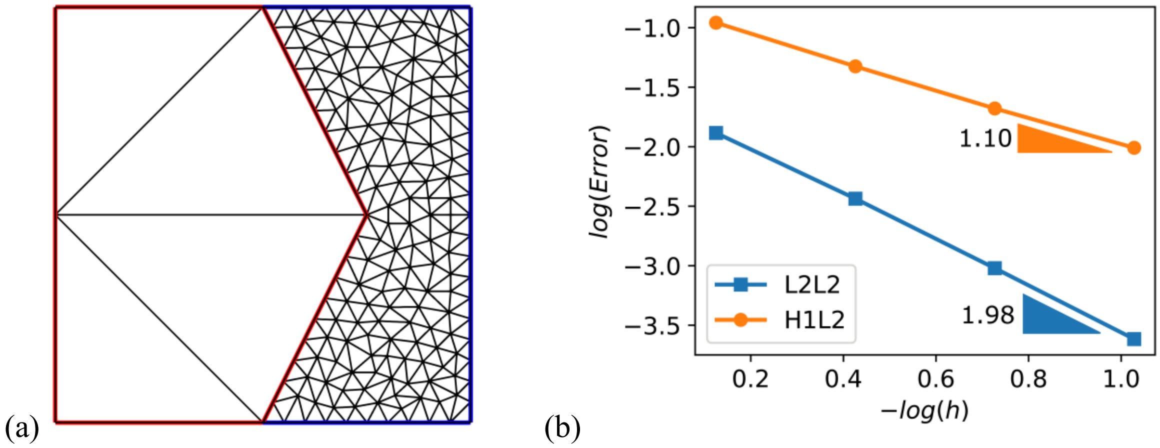


Figure 5.9. (a) Domain with chevron interface and 1:10 mesh size ratio and (b) convergence study.

Finally, we look into a case with higher order elements[5]. Specifically, we consider again the domain with a chevron interface as in Figure 5.4b and central difference in time, but with quadratic 6-node triangular (T6) elements in space. Because central difference is a second order accurate method, we choose a very small time step so that the time discretization error is small relative to the spatial discretization error and that the error norms in (5.49) and (5.50) probe into and reflect the rate of convergence associated with the order of accuracy of the spatial discretization. Therefore, we select a large number $M = 1 \times 10^8$ to set the total time to $T = 0.01/M$ and the time step is determined by picking the largest integer N such that.

$$\Delta t = T / N \le \frac{1}{2M} \cdot \frac{h_{min}^2}{2\mu + \lambda} \tag{5.51}$$

The results of the convergence study are shown in Figure 5.10, where it is observed that optimal rates of convergence for quadratic elements are attained: 3 in L2L2 norm and 2 in H1L2 norm.

---

[5] Results for higher-order elements have been presented for the quasi-static case in 2D and 3D in [22,23], where optimal convergence rates were recovered.

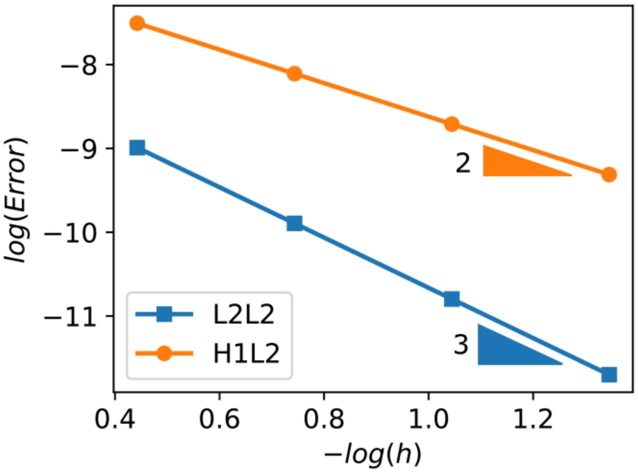


Figure 5.10. Error convergence plots for T6 elements.

5.6.2 Energy

For this section, we consider an unforced, free-vibration problem on a unit square with its bottom-left corner at the origin. Homogeneous Dirichlet boundary conditions are prescribed along the line $x = 0$, and traction-free (homogeneous Neumann) boundary conditions are applied to the remainder of the domain boundary. The final time is $T = 5$ and the initial conditions are given as

$$\mathbf{u}_0 = 0.1xy\begin{bmatrix}1 & 1\end{bmatrix}^T, \qquad \dot{\mathbf{u}}_0 = \mathbf{0} \tag{5.52}$$

We employ a central difference algorithm in time with $N = 325$ and $\Delta t \approx 0.0154$. The solution is computed for a spatial discretization with an interface as in Figure 5.4b, as well as for a reference "monolithic" mesh (i.e., without interface) shown in Figure 5.11. The reference mesh has an h_{min} similar to that of the chevron mesh, and the number of time steps used is $N = 359$. The strain, kinetic, and total energy of the system are respectively given by

$$U(\mathbf{d}_n^{(\alpha)}) = \frac{1}{2}\mathbf{d}_n^{(\alpha)T}\mathbf{K}\mathbf{d}_n^{(\alpha)} \tag{5.53}$$

$$T(\mathbf{v}_n^{(\alpha)}) = \frac{1}{2}\mathbf{v}_n^{(\alpha)T}\mathbf{M}\mathbf{v}_n^{(\alpha)} \tag{5.54}$$

$$E(\mathbf{d}_n^{(\alpha)}, \mathbf{v}_n^{(\alpha)}) = U(\mathbf{d}_n^{(\alpha)}) + T(\mathbf{v}_n^{(\alpha)}) \tag{5.55}$$

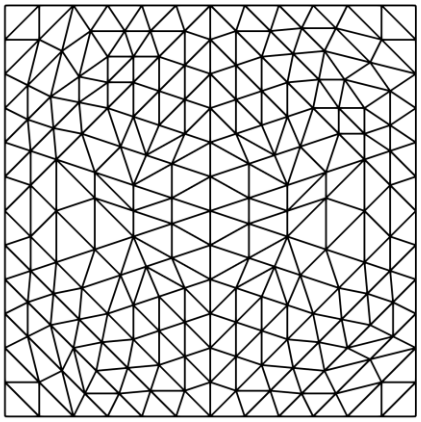


Figure 5.11. Reference monolithic mesh.

In addition, when the domain is partitioned by an interface, one should account for the work done by the tractions at the interface that tie the subdomains together. The work done by those tractions over a single time step is estimated as the work done by the force at the mid-step over the displacement increment for that time step, i.e.,

$$\begin{aligned} \Delta W_{n+1}^{\Gamma_I} &= \left(\mathbf{d}_n^{(1)} - \mathbf{d}_{n+1}^{(1)}\right)^T \left(\mathbf{K}_{\Gamma_I}^{(11)}\mathbf{d}_{n+1/2}^{(1)} + \mathbf{K}_{\Gamma_I}^{(12)}\mathbf{d}_{n+1/2}^{(2)}\right) \\ &+ \left(\mathbf{d}_n^{(2)} - \mathbf{d}_{n+1}^{(2)}\right)^T \left(\mathbf{K}_{\Gamma_I}^{(22)}\mathbf{d}_{n+1/2}^{(2)} + \mathbf{K}_{\Gamma_I}^{(21)}\mathbf{d}_{n+1/2}^{(1)}\right) \end{aligned} \tag{5.56}$$

Where $\mathbf{d}_{n+1/2}^{(\alpha)} = \frac{1}{2}\left(\mathbf{d}_n^{(\alpha)} + \mathbf{d}_{n+1}^{(\alpha)}\right)$. Therefore, the total work done by the interfacial forces up to t_{n+1} is given by

$$W_{n+1}^{\Gamma_I} = \sum_{i=0}^{n} \Delta W_{i+1}^{\Gamma_I} \tag{5.57}$$

The time history of the system strain energy, kinetic energy, work at the interface, and total energy plus work ($E + W$) are shown in Figure 5.12. Consistent with initial conditions, at $t = 0$, all the energy is stored strain energy, while the kinetic energy is zero. During dynamic evolution of the problem, energy is converted from strain energy to kinetic energy and vice versa. Throughout, the work done at the interface remains close to zero in comparison with the strain and kinetic energy contributions, and the total energy plus work remains fairly constant. Figure 5.13 contains snapshots of the deformed configuration of the solution at time levels $t = 0$ (initial conditions), 1.66, 3.34, and 5. Some minor gaps and overlaps develop at the interface since continuity is weakly enforced through the coupling formulation.

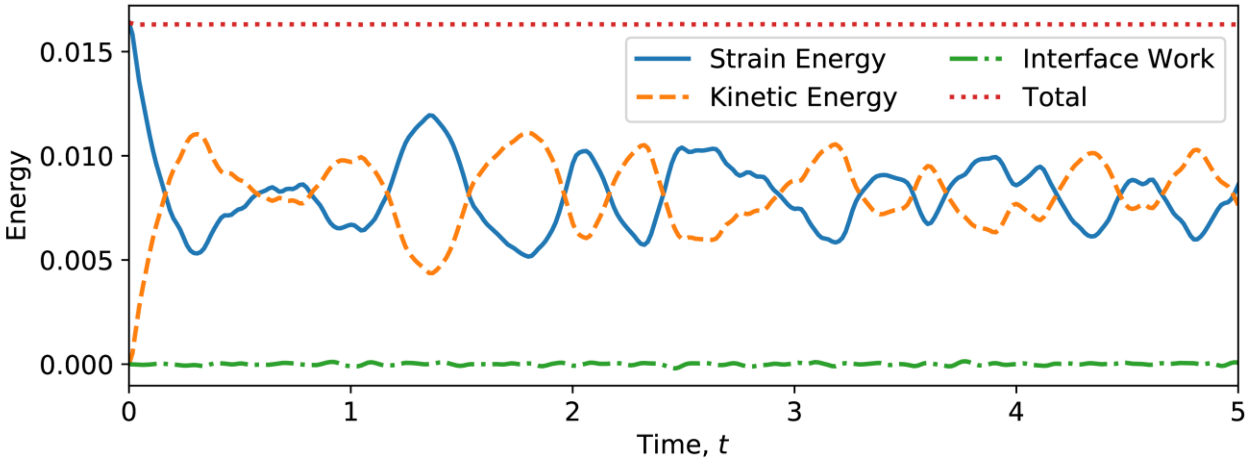


Figure 5.12. Time history of energy and work.

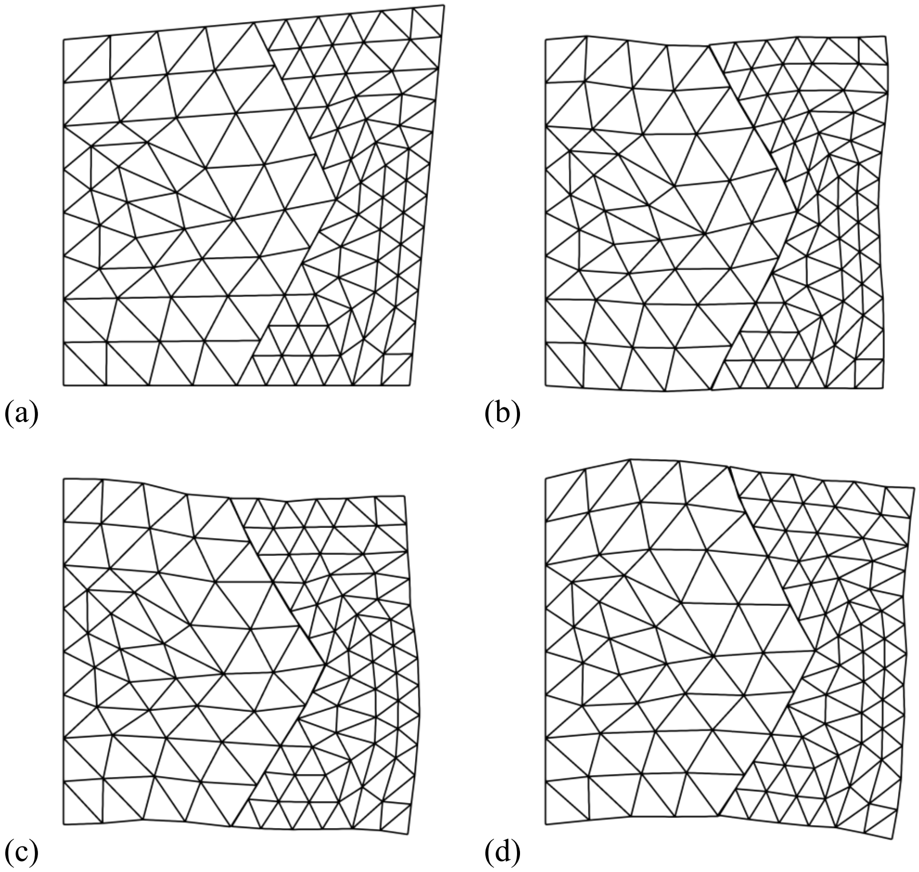


Figure 5.13. Snapshots of the deformed configuration at (a) $t = 0$, (b) $t = 1.66$, (c) $t = 3.34$, and (d) $t = 5$.

The effect of the VMDG interface coupling method on the conservation properties of the solution are studied through the time history of several quantities related to the energy. First, we remark that, *regardless* of whether there is an interface in the domain, the algorithmically conserved quantity for time integration schemes in the Newmark-β family of methods with γ = 1/2 and no external forcing function is given by (see [32])

$$Q(\mathbf{d}_n, \mathbf{v}_n, \mathbf{a}_n) = E(\mathbf{d}_n, \mathbf{v}_n) + (\Delta t)^2 \left(\beta - \frac{1}{4} \right) T(\mathbf{a}_n) \tag{5.58}$$

For the well-known (implicit) trapezoidal or average acceleration method ($\gamma = 1/2$, β = 1/4), the algorithmically conserved quantity coincides exactly with the total energy $E(\mathbf{d}_n, \mathbf{v}_n)$. Meanwhile, for the (explicit) central difference method ($\gamma = 1/2$, $\beta = 0$) that has been used to obtain numerical results in this chapter, the algorithmically conserved quantity is given by $E(\mathbf{d}_n, \mathbf{v}_n) - \frac{1}{4}(\Delta t)^2 T(\mathbf{a}_n)$. Furthermore, we define a few additional quantities for ease of notation. In the case of a monolithic domain (no internal interface), we set

$$E_n = E(\mathbf{d}_n, \mathbf{v}_n), \qquad Q_n = Q(\mathbf{d}_n, \mathbf{v}_n, \mathbf{a}_n) \tag{5.59}$$

For a domain traversed by an interface, we set:

$$E_n^{(\alpha)} = E(\mathbf{d}_n^{(\alpha)}, \mathbf{v}_n^{(\alpha)}), \qquad \hat{E}_n = E_n^{(1)} + E_n^{(2)}, \qquad E_n^W = \hat{E}_n + W_n^{\Gamma_I} \tag{5.60}$$

$$Q_n^{(\alpha)} = Q(\mathbf{d}_n^{(\alpha)}, \mathbf{v}_n^{(\alpha)}, \mathbf{a}_n^{(\alpha)}), \qquad \hat{Q}_n = Q_n^{(1)} + Q_n^{(2)}, \qquad Q_n^W = \hat{Q}_n + W_n^{\Gamma_I} \tag{5.61}$$

where $E_n^{(\alpha)}$ is the total energy (strain energy plus kinetic energy) of subdomain α, $\hat{E}_n$ is the combined energy of both subdomains, E_n^W is the total energy of the subdomains plus the total work done by the forces at the interface, and Q_n^W is the algorithmically conserved quantity accounting for the work done by the forces at the interface. Then, we normalize these quantities ($\hat{E}_n$, E_n^W, Q_n^W) as well as those defined for the reference (monolithic) mesh in equation (5.59) by their corresponding initial value at time level $n = 0$. The time history of the normalized quantities is shown in Figure 5.14. The following remarks are in order:

1. The normalized energy for the reference mesh (*without* interface) is not exactly conserved – its value starts at 1 and then varies around a constant value that is slightly less than 1 (i.e., there is no further decay either). This is to be expected because the time integration scheme used was central difference method, which corresponds to Newmark-β method with γ = 1/2 and $\beta = 0$, for which total energy E_n is not algorithmically conserved. Recall that the

only member of the Newmark-β family that exactly conserves total energy algorithmically is the trapezoidal rule or average acceleration method ($\gamma = 1/2$, $\beta = 1/4$).

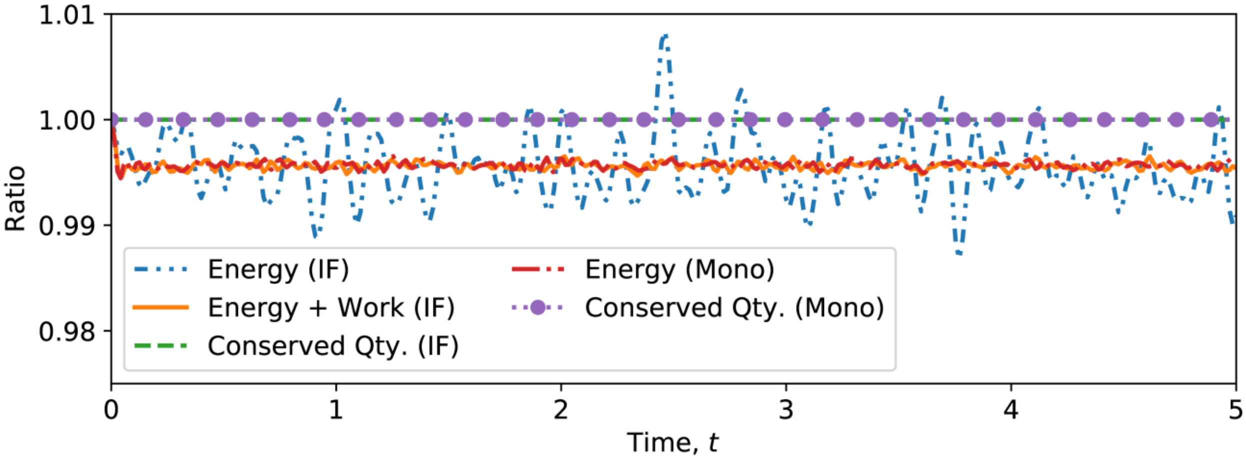


Figure 5.14. Time history of normalized quantities for interface (IF) and monolithic (Mono) meshes.

2. The plot of the normalized total energy plus interfacial work E_n^W for the mesh with interface follows closely that of the normalized total energy E_n of the reference (monolithic) mesh. In other words, the VMDG interface coupling method *does not* affect (and rather inherits) the energy conservation properties of the underlying time integration scheme.
3. In the case of a domain with an interface and multiple subdomains, computing the total energy of the system by just combining the total energy of each subdomain (i.e., $\hat{E}_n$) may be an overly simplistic approach. In fact, that approach fails to account for the work done by the forces at the interface, which couple or "tie" the subdomains together. In Figure 5.14, this gets reflected in that, although both the normalized values of $\hat{E}_n$ and E_n^W (which accounts for the interfacial work) oscillate around the same constant value, $\hat{E}_n$ has larger overshoots and undershoots.
4. The normalized algorithmically conserved quantities for the reference (monolithic) mesh (Q_n) and for the mesh with an interface (Q_n^W) are verified to be exactly conserved. Their plots overlap in a horizontal straight line at a constant value of 1. This verifies once more that the consistently-derived VMDG interface coupling method inherits the energy conservation properties of the underlying time integration scheme.

5.6.3 Eigenvalues

In this section, we study the influence of the VMDG interface formulation on the eigenvalues (or natural frequencies) of the problem. Recall that eigenvalues λ_l represent the square of the natural frequencies ω_l, i.e. $\lambda_l = (\omega_l)^2$. First, we focus on the largest eigenvalue (corresponding to the largest natural frequency), which constrains the critical time step size for explicit time integration. Secondly, we study the lowest eigenvalues (corresponding to the lowest natural frequencies, or largest natural period) to quantify the numerical dispersion (commonly measured by the relative period error) introduced by the VMDG method for interface coupling.

5.6.3.1 Largest Eigenvalue

We consider additional discretizations of the domain described in section 5.6.2 to study the influence of the interface formulation on the natural frequencies ω^h of the problem. We solved numerically for the largest eigenvalues of the system from equation (5.40) for the unconstrained domain discretized with the meshes in Figure 5.4b and Figure 5.11, as well as for two more meshes with a chevron interface in Figure 5.15. These latter two are such that the number of degrees of freedom (NDOF) is similar to that of the monolithic mesh, but one has matching nodes at the interface, while there is a slight mismatch in the other. The results are summarized in Table 5.1.

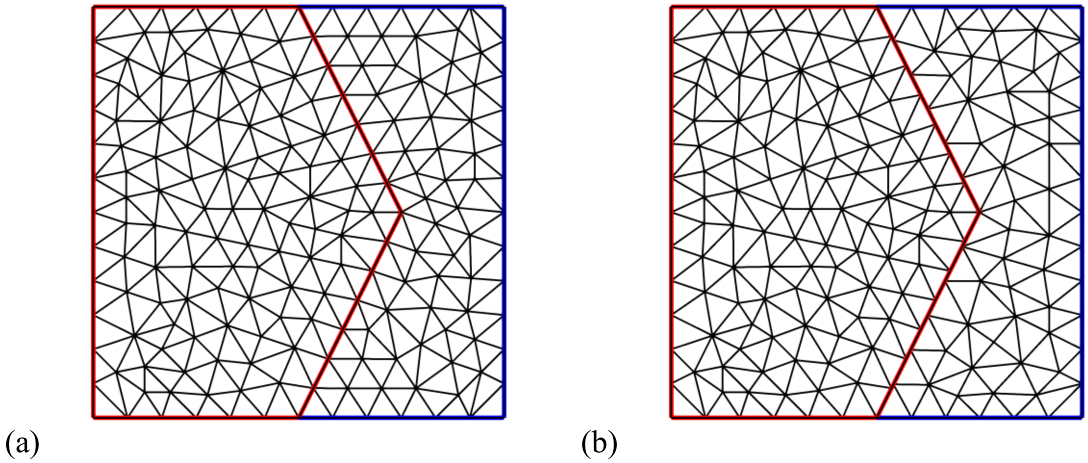

(a) (b)

Figure 5.15. Additional chevron meshes: (a) Matching, and (b) Nonmatching.

Table 5.1. Maximum Natural Frequencies (Eigenvalues).

Figure	Mesh	NDOF	$\left(\omega_{max}^{h}\right)^{2}$
Figure 5.11	monolithic	402	3712
Figure 5.15a	matching	416	3306
Figure 5.4	non-matching	256	3468
Figure 5.15b	non-matching	398	6167

Note that $\left(\omega_{max}^{h}\right)^{2}$ is similar for the monolithic mesh and the matching chevron mesh, having a similar number of degrees of freedom as well. Therefore, no spurious high frequencies are introduced by the interface formulation in the case of matching discretization of the subdomains. On the other hand, $\left(\omega_{max}^{h}\right)^{2}$ for the non-matching chevron mesh from Figure 5.4b has a similar value as well, despite having fewer degrees of freedom and similar h_{min} value, and the non-matching mesh from Figure 5.15b has a value of $\left(\omega_{max}^{h}\right)^{2}$ that is almost twice that of the other cases, despite having a similar number of degrees of freedom and h_{min} value. In other words, for subdomains with non-matching discretizations at the interface, the maximum eigenvalues are larger than expected. We note that node mismatch at the interface Γ_{I} can lead to relatively small segments γ_{s}. Therefore, these preliminary results suggest that having small segments γ_{s} (relative to h_{min}) may give rise to large natural frequencies (relative to a monolithic mesh with comparable h_{min}), which constrains the critical time step size Δt_{crit} for explicit time integration.

Consequently, we further studied the change in the maximum eigenvalue of the system with respect to the mismatch at the interface. To that effect, we consider the following 2 cases for discretizing a unit square domain: (1) a case with a single mismatched node at the interface (Figure 5.16), and (2) a case with three mismatched nodes at the interface (Figure 5.17).

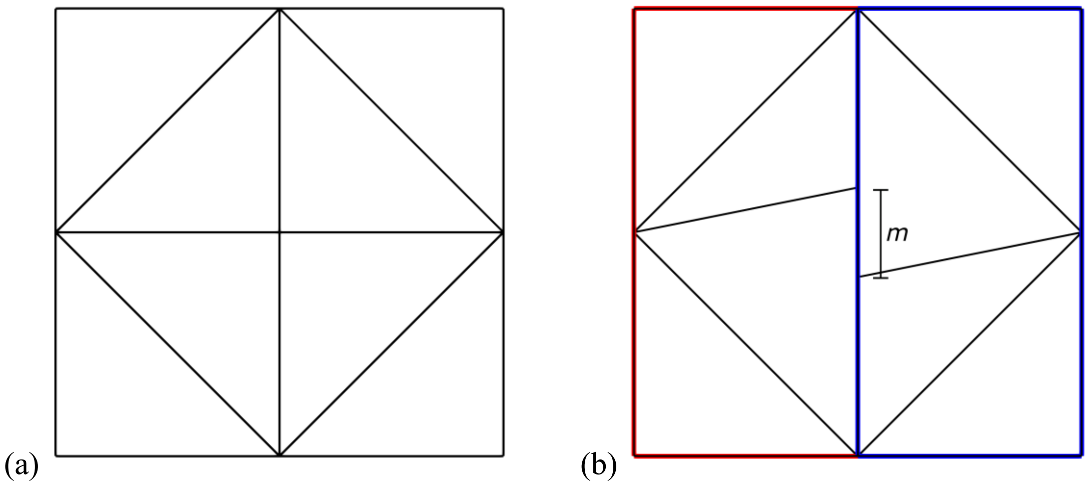


Figure 5.16. Meshes used for eigenvalue study with a single interface mismatch (``case 1"): (a) Monolithic/Matching, and (b) Interface mismatch m.

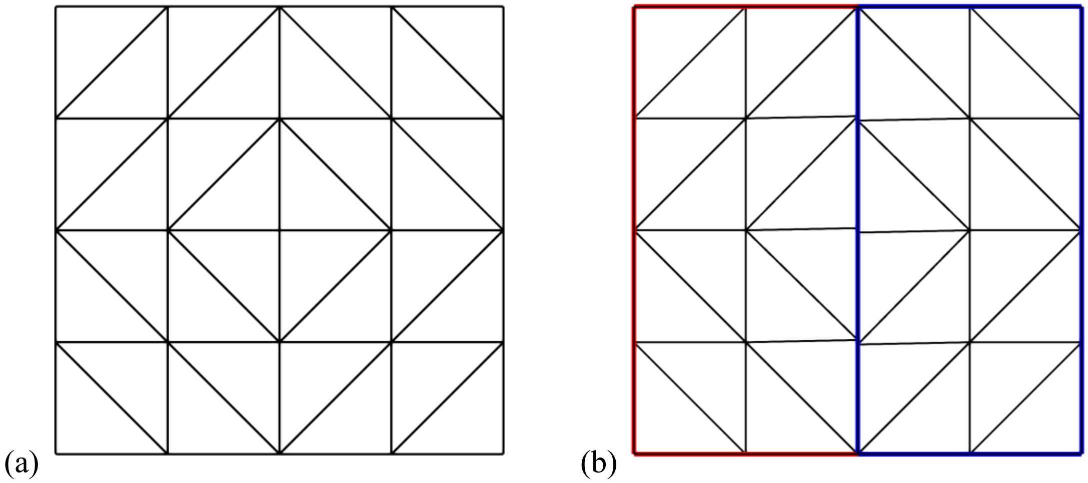


Figure 5.17. Meshes used for eigenvalue study with 3 mismatched nodes at the interface (``case 2"): (a) Monolithic/Matching and (b) Interface mismatch $m = 0.01$.

For both cases, the maximum eigenvalue $\left(\omega_{max}^{h}\right)^{2}$ of the system is computed for a monolithic mesh (no interface), a mesh with matching interface (mismatch $m = 0$), and meshes with varying mismatch size m at the interface. The results are summarized in Table 5.2 and Table 5.3.

Table 5.2. Maximum Eigenvalues for Case 1.

NDOF	Mismatch	$\left(\omega_{max}^{h}\right)^{2}$
18	monolithic	84.88
24	0	104.36
24	1.00×10^{-1}	189.88
24	1.29×10^{-2}	208.93
24	1.67×10^{-3}	221.40
24	2.15×10^{-4}	221.97
24	1.00×10^{-5}	222.05

Table 5.3. Maximum Eigenvalues for Case 2.

NDOF	Mismatch	$\left(\omega_{max}^{h}\right)^{2}$
50	monolithic	354.41
60	0	417.44
60	1.00×10^{-1}	740.86
60	1.29×10^{-2}	857.76
60	1.67×10^{-3}	884.03
60	2.15×10^{-4}	887.66
60	1.00×10^{-5}	888.18

In addition, we computed and then plotted the ratio of the maximum eigenvalue of a mismatched interface to that of the corresponding matching mesh in Figure 5.18. The following remarks are in order:

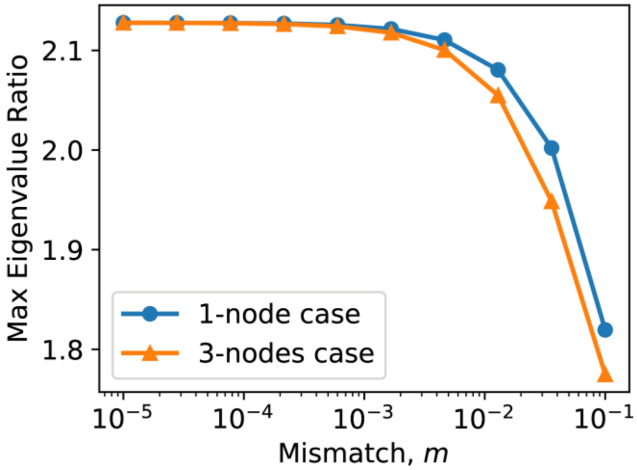


Figure 5.18. Normalized maximum eigenvalue vs mismatch size.

1. The maximum eigenvalue of the meshes with matching interface ($m=0$) is larger than their corresponding monolithic mesh's maximum eigenvalue. This is expected since a mesh with interface has more DOFs than its corresponding monolithic mesh.
2. In fact, comparing the results only for the meshes with *matching* interface ($m=0$) with their corresponding monolithic mesh, the increase in DOFs is a factor of $24/18\approx1.33$ and

$60/50 = 1.2$ for case 1 and case 2, respectively. On the other hand, the factor increases in maximum eigenvalues $\left(\omega_{max}^{h}\right)^{2}$ are $104.36 / 84.88 \approx 1.23$ and $417.44 / 354.41 \approx 1.18$, respectively, or in terms of maximum natural frequencies ω_{max}^{h}, 1.11 and 1.09. Thus, the scaling of maximum eigenvalues/natural frequencies associated with an increase in NDOF due to the presence of an interface is comparable to (and possibly less than) the scaling of maximum eigenvalues associated with an increase in NDOFs due to uniform refinement in a monolithic mesh.

3. When nodes from either side of the interface are mismatched, the maximum eigenvalue is larger than that of a mesh with matching nodes at the interface. However, as the mismatch distance m goes to zero, the value of the maximum eigenvalue asymptotes to a finite upper bound (Figure 5.18). For the test cases discussed here, the value of this upper bound is roughly 2.13 times the maximum eigenvalue of the mesh with matching nodes at the interface.
4. Recall that the critical time step for the Newmark method with $\beta = 0$, $\gamma = 1/2$ (central difference method) is given by (see [32])

   $$\Delta t \leq \frac{2}{\omega_{max}^{h}} \tag{5.62}$$

   Therefore, the critical time step is decreased only by a factor of at most $\sqrt{2.3} \approx 1.46$, or, equivalently, the total number of time steps must be increased only by a factor of no more than 1.46 when there are small mismatches at the interface. As a result, the effect of node mismatches at the interface does not represent a major concern regarding the time step size for the implementation of the explicit VMDG method for synchronous and concurrent parallel transient multi-domain computation.

5.6.3.2 Smallest Eigenvalues

For a given vibration mode number l, its period T_l is related to the natural frequency ω_l by the relation $T_l = 2\pi / \omega_l$. Spatial and temporal discretizations introduce errors in the solution, and the relative period error is used as a measure of dispersion in the numerical methods [32,34,35]. Given the straight-forward relations between periods T_l, natural frequencies ω_l, and eigenvalues λ_l, it is sufficient to report on the results for the eigenvalues λ_l, since the results for the other quantities follow immediately. Furthermore, we can split the error into three contributions: (i) the error due to spatial discretization using FEM in the semi-discrete problem relative to the exact solution, (ii) the error introduced by the presence of an interface and the use of the VMDG method for interface coupling relative to a monolithic semi-discrete FEM solution, and (iii) the error introduced by the numerical time integration algorithm in the fully discrete solution, relative to the exact solution of the semi-discrete (time continuous, spatially discrete) problem.

Regarding the error in (iii), comparisons of the dispersion properties (as measured by the relative period error) for several implicit and explicit time integration methods have been documented in the literature [32,34,35]. Furthermore, for (i), the dispersion error introduced by the spatial discretization via the Galerkin FEM can be quantified in terms of the rate of convergence of the error in the discrete eigenvalues relative to the exact eigenvalues of the continuous problem. The standard error estimate for eigenvalues in the Galerkin FEM is [32,36]

$$\lambda_l^h - \lambda_l \le c h^{2k} \lambda_l^{k+1} \tag{5.63}$$

where c is a constant independent of h and λ_l, λ_l is the lth exact eigenvalue, λ_l^h is its counterpart in the spatially discrete, time-continuous solution, and k is the largest complete polynomial degree spanned by the FE basis functions. Note that for the case of linear elements ($k = 1$), the rate of convergence with respect to h is quadratic. We also remark that the error rate of convergence in the frequency $\omega_l^h = \sqrt{\lambda_l^h}$, is the same as that in the eigenvalue λ_l^h (see [32]).

Finally, in order to quantify the contribution to the error (ii), we compare the relative error in the 4 smallest (nonzero) eigenvalues of a domain discretized using a monolithic mesh to those of a mesh with a matching interface using VMDG. For this study, λ_l^h denotes the lth nonzero eigenvalue of the monolithic mesh, while $\overline{\lambda}_l^h$ denotes the corresponding eigenvalue in the mesh with a matching interface. The meshes considered in this study are progressive refinements of those in Figure 5.16 (with mismatch $m = 0$ for the matching interface case), obtained by partitioning each element into four elements with each increase in refinement level (as done in going from Figure 5.16a to Figure 5.17a, for instance). The eigenvalue error convergence results are shown in Figure 5.19.

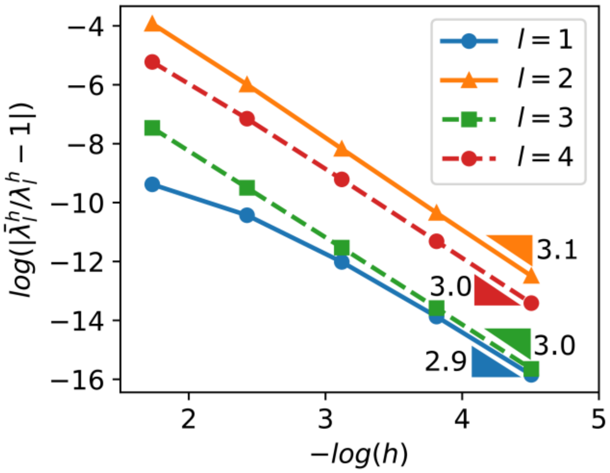


Figure 5.19. Convergence of error in 4 lowest (nonzero) eigenvalues of mesh with matching interface relative to eigenvalues in a corresponding monolithic mesh.

This numerical convergence study shows that the error between the eigenvalues in the VMDG problem and the eigenvalues of the monolithic problem has a *cubic* rate of convergence for linear elements, which is faster than the quadratic rate of convergence from the standard error estimate in FEM. In other words, the dispersion introduced by the VMDG method is expected to become insignificant compared to that introduced by the underlying FEM discretization of the subdomains, as well as the one introduced by the choice of time integration algorithm. Again, as in the case of energy conservation (or dissipation) properties, the fully discrete method inherits the dispersion properties of the selected time integration algorithm.

5.7 Conclusions

We have presented an explicit VMDG method for synchronous and concurrent multi-domain parallel computations. The coupling of different computational domains across an embedded interface is achieved by consistently deriving transmission conditions for transient problems. The interface coupling terms arise naturally from the derivation that follows after employing the variational multiscale method to decompose the solution field into coarse and fine scales. Moreover, the result is a formulation in primal form, where the only unknown that needs to be solved for is the coarse scale field. The effect of the fine scales and Lagrange multiplier is captured through the additional coupling terms that are derived via the VMDG method. The solution procedure at a given time step is decoupled for each subdomain by means of explicit time-integration algorithm. Optimal convergence rates are verified numerically by considering a measure of error that accounts for the spatial and temporal discretization error, which confirms the consistency of the method. Furthermore, the magnitude of the error at the interface is in the same order of magnitude as (or less than) the error in the interior of the domain. Consequently, the interface formulation does not introduce additional errors and the accuracy of the method is dictated by the underlying discretization. Furthermore, the energy conservation behavior of the coupled subdomains is in agreement with the results for a monolithic mesh without an interface, once the work done by the interfacial forces that tie the subdomains is accounted for. By using an explicit method for time integration, the solution procedure for each subdomain is decoupled from one another, only subject to sharing of information from previously computed steps. Nevertheless, the consistently derived VMDG interface terms capture the interaction between the subdomains. These features of the method enable synchronous and concurrent solution of multiple subdomains that may reside in separate computational platforms, with minimal information sharing.

5.8 References

[1] T.J. Truster, A. Masud, A Discontinuous/continuous Galerkin method for modeling of interphase damage in fibrous composite systems, Comput. Mech. 52 (2013) 499–514.

[2] T.J. Truster, A. Masud, Discontinuous Galerkin Method for Frictional Interface Dynamics, J. Eng. Mech. 142 (2016) 1–12.

[3] J. Nitsche, Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind, in: Abhandlungen Aus Dem Math. Semin. Der Univ. Hambg., 1971: pp. 9–15.

[4] A. Hansbo, P. Hansbo, An unfitted finite element method, based on Nitsche's method, for elliptic interface problems, Comput. Methods Appl. Mech. Eng. 191 (2002) 5537–5552.

[5] C. Annavarapu, M. Hautefeuille, J.E. Dolbow, A robust Nitsche's formulation for interface problems, Comput. Methods Appl. Mech. Eng. 225–228 (2012) 44–54.

[6] A. Seitz, W.A. Wall, A. Popp, Nitsche's method for finite deformation thermomechanical contact problems, Comput. Mech. 63 (2019) 1091–1110.

[7] U. Albocher, I. Harari, Spectral aspects of Nitsche's method on nonconforming meshes, Mech. Res. Commun. (2020) 103611.

[8] T.W. McDevitt, T.A. Laursen, A mortar-finite element formulation for frictional contact problems, Int. J. Numer. Methods Eng. 48 (2000) 1525–1547.

[9] T.A. Laursen, M.A. Puso, J. Sanders, Mortar contact formulations for deformable–deformable contact: Past contributions and new extensions for enriched and embedded interface formulations, Comput. Methods Appl. Mech. Eng. 205–208 (2012) 3–15.

[10] A. Popp, W.A. Wall, Dual mortar methods for computational contact mechanics – overview and recent developments, GAMM-Mitteilungen. 37 (2014) 66–84.

[11] M. Hiermeier, W.A. Wall, A. Popp, A truly variationally consistent and symmetric mortar-based contact formulation for finite deformation solid mechanics, Comput. Methods Appl. Mech. Eng. 342 (2018) 532–560.

[12] P.F. Antonietti, I. Mazzieri, A. Quarteroni, F. Rapetti, Non-conforming high order approximations of the elastodynamics equation, Comput. Methods Appl. Mech. Eng. 209–212 (2012) 212–238.

[13] J.D. De Basabe, M.K. Sen, M.F. Wheeler, Elastic wave propagation in fractured media using the discontinuous Galerkin method, Geophysics. 81 (2016) T163–T174.

[14] K.C. Park, C.A. Felippa, R. Ohayon, Partitioned formulation of internal fluid–structure interaction problems by localized Lagrange multipliers, Comput. Methods Appl. Mech. Eng. 190 (2001) 2989–3007.

[15] K.C. Park, C.A. Felippa, A variational principle for the formulation of partitioned structural systems, Int. J. Numer. Methods Eng. 47 (2000) 395–418.

[16] M.R. Ross, C.A. Felippa, K.C. Park, M.A. Sprague, Treatment of acoustic fluid–structure interaction by localized Lagrange multipliers: Formulation, Comput. Methods Appl. Mech. Eng. 197 (2008) 3057–3079.

[17] M.R. Ross, M.A. Sprague, C.A. Felippa, K.C. Park, Treatment of acoustic fluid–structure interaction by localized Lagrange multipliers and comparison to alternative interface-coupling methods, Comput. Methods Appl. Mech. Eng. 198 (2009) 986–1005.

[18] C.A. Felippa, K.C. Park, C. Farhat, Partitioned analysis of coupled mechanical systems, Comput. Methods Appl. Mech. Eng. 190 (2001) 3247–3270.

[19] P.B. Nakshatrala, K.B. Nakshatrala, D.A. Tortorelli, A time-staggered partitioned coupling algorithm for transient heat conduction, Int. J. Numer. Methods Eng. 78 (2009) 1387–1406.

[20] T. He, K. Zhang, An Overview of the Combined Interface Boundary Condition Method for Fluid–Structure Interaction, Arch. Comput. Methods Eng. 24 (2017) 891–934.

[21] K. Peterson, P. Bochev, P. Kuberry, Explicit synchronous partitioned algorithms for interface problems based on Lagrange multipliers, Comput. Math. with Appl. 78 (2019) 459–482.

[22] A. Masud, T.J. Truster, L.A. Bergman, A unified formulation for interface coupling and frictional contact modeling with embedded error estimation, Int. J. Numer. Methods Eng. 92 (2012) 141–177.

[23] T.J. Truster, A. Masud, Primal interface formulation for coupling multiple PDEs: A consistent derivation via the Variational Multiscale method, Comput. Methods Appl. Mech. Eng. 268 (2014) 194–224.

[24] T.J. Truster, P. Chen, A. Masud, Finite strain primal interface formulation with consistently evolving stabilization, Int. J. Numer. Methods Eng. 102 (2015) 278–315.

[25] G. Hlepas, T. Truster, A. Masud, A heterogeneous modeling method for porous media flows, Int. J. Numer. Methods Fluids. 75 (2014) 487–518.

[26] P. Chen, T.J. Truster, A. Masud, Interfacial stabilization at finite strains for weak and strong discontinuities in multi-constituent materials, Comput. Methods Appl. Mech. Eng. 328 (2018) 717–751.

[27] A. Masud, P. Chen, Thermoelasticity at finite strains with weak and strong discontinuities, Comput. Methods Appl. Mech. Eng. 347 (2019) 1050–1084.

[28] P. Chen, I.P.A. Wijaya, I. Tuttle, A. Masud, Interfacial coupling across incompatible meshes in a monolithic finite-strain thermomechanical formulation, Comput. Math. with Appl. 79 (2020) 3068–3091.

[29] S.C. Aduloju, T.J. Truster, A variational multiscale discontinuous Galerkin formulation for both implicit and explicit dynamic modeling of interfacial fracture, Comput. Methods Appl. Mech. Eng. 343 (2019) 602–630.

[30] T.J.R. Hughes, G.R. Feijóo, L. Mazzei, J.-B. Quincy, The variational multiscale method—a paradigm for computational mechanics, Comput. Methods Appl. Mech. Eng. 166 (1998) 3–24.

[31] A. Masud, L.P. Franca, A hierarchical multiscale framework for problems with multiscale source terms, Comput. Methods Appl. Mech. Eng. 197 (2008) 2692–2700.

[32] T.J.R. Hughes, The finite element method: linear static and dynamic finite element analysis, Dover Publications, Mineola, NY, 2000.

[33] F. Hecht, New development in freefem++, J. Numer. Math. 20 (2012) 251–266.

[34] J. Chung, G.M. Hulbert, A Time Integration Algorithm for Structural Dynamics With Improved Numerical Dissipation: The Generalized-α Method, J. Appl. Mech. 60 (1993) 371.

[35] G.L. Goudreau, R.L. Taylor, Evaluation of numerical integration methods in elastodynamics, Comput. Methods Appl. Mech. Eng. 2 (1973) 69–97.

[36] W.G. Strang, G.J. Fix, An analysis of the finite element method, Prentice-Hall, Englewood Cliffs, N.J., 1973.

[37] M.A. Puso, T.A. Laursen, A mortar segment-to-segment contact method for large deformation solid mechanics, Comput. Methods Appl. Mech. Eng. 193 (2004) 601–629.

[38] D. Powell, T. Abel, An exact general remeshing scheme applied to physically conservative voxelization, J. Comput. Phys. 297 (2015) 340–356.

[39] P.A. Ullrich, M.A. Taylor, Arbitrary-Order Conservative and Consistent Remapping and a Theory of Linear Maps: Part I, Mon. Weather Rev. 143 (2015) 2419–2440.

[40] T.J. Tautges, C. Ernst, C. Stimpson, R.J. Meyers, K. Merkley, MOAB : a mesh-oriented database., Albuquerque, NM, and Livermore, CA (United States), 2004.

CHAPTER 6:
MODELING OF STEEP LAYERS IN SINGULARLY PERTURBED DIFFUSION-REACTION EQUATION VIA FLEXIBLE FINE-SCALE BASIS[1]

6.1 Introduction

The Galerkin Finite Element method develops spurious oscillations engendered by boundary and/or internal layers in singularly perturbed diffusion-reaction equations (i.e. when the coefficient of the zeroth-order term is much larger than the coefficient of the second order term). An analysis that shows the origin of these spurious oscillations in the singularly perturbed diffusion-reaction equation can be found in [1,2]. There exists extensive literature in the subject of stabilized methods to control spurious oscillations, and some of the first include the Streamline-Upwind Petrov-Galerkin (SUPG) [3], SUPG with discontinuity capturing (DC) [4], and the Galerkin Least-Squares (GLS) [5] methods. Although a considerable portion of this literature is mainly devoted to the advection-dominated case of the advection-diffusion equation, several methods have also been proposed to mitigate the issue of spurious oscillations in the reaction-dominated regime of the diffusion-reaction or advection-diffusion-reaction equations. These include an extension of SUPG with DC for the reaction-dominated regime as well [6], the Galerkin Gradient Least-Squares (GGLS) method in [1], and a modification of the GLS method for the case when reaction is dominant, in addition to a combination of the GLS and GGLS methods, as presented in [7]. The key idea underlying GGLS was to add a term to stabilize the discrete solution in the H^1 norm. This was followed by the Adjoint or Unusual Stabilized methods (USFEM) [8], wherein an $L_2(\Omega)$ term was subtracted to control the oscillations around steep fronts in reaction dominated cases. These methods are based on suitable selection of the stability parameter, which is shown to be critical for the stability of the method. Subsequent works on the topic include Sub-Grid Scale methods (SGS) [9], the Variational Multiscale (VMS) method [10], Residual-Free Bubbles (RFB) method [11–13], and the use of bubble functions and Green's functions in fine-scale models [8,9,14]. Comparisons of the VMS, SGS, SUPG, and

---

[1] This Chapter has been adapted from "A. Masud, M. Anguiano, I. Harari, Modeling of steep layers in singularly perturbed diffusion–reaction equation via flexible fine-scale basis, Comput. Methods Appl. Mech. Eng. 372 (2020) 113343." And "M. Anguiano, A. Masud, Variational Multiscale Method with Flexible Fine-Scale Basis for Diffusion-Reaction Equation: Built-In a Posteriori Error Estimate and Heterogenous Coefficients, in: 14th WCCM-ECCOMAS Congress, CIMNE, 2021." The copyright owners have provided written permission to reprint the work.

GLS for the diffusion-advection-reaction equation were done in [15,16], where, in particular, the reaction-dominated case was considered.

Some other classes of methods that are effective in suppressing spurious oscillations around boundary layers in singular diffusion-reaction problems are found in the literature: The Galerkin Projected Residual (GPR) method was introduced in [17], which involves stabilization terms emanating from multiple projections of the residual weighted by coefficients that depend on matrices emanating from USFEM and GLS type methods. The Galerkin Enriched FEM (GEM), and Petrov-Galerkin Enriched Method (PGEM) (see [18] references therein) modify the standard FEM space of piecewise Lagrange polynomials with enrichments emanating from the solution of the strong form of the governing PDE in the elements and on element boundaries. In addition, corresponding adjoint stabilized methods (SGEM and SPGEM) were also presented. Furthermore, Finite Increment Calculus (FIC) has been used to derive methods such as the High Resolution Petrov-Galerkin (HRPG) (see [19] and references therein), where the built-in discontinuity capturing brings about stability around boundary and internal layers for the singular diffusion-reaction case. The work in [20] introduces a variationally derived DC method that was shown to work for travelling steep gradients in the advection-dominated regime, and relies on enriching the definition of a fine-scale bubble with a discontinuity within the element domain.

In this chapter, we employ the VMS framework [10] to consistently and variationally derive a stabilized method for the diffusion-reaction equation. Stabilization is facilitated by a fine-scale model that is derived via local solution of the fine-scale variational formulation. The driving mechanism for the local or fine-scale problem is the residual of the Euler-Lagrange equation of the coarse scales. One approach that has been successfully applied to many problems in fluids and solids involves applying the bubble-based approach directly at the fine-scale level using simple polynomial bubbles [21–25]. Furthermore, application of this idea to problems involving chemically-reacting solids permeated by fluids is presented in [26,27], where the governing equations contain reaction-type terms. Numerical results in the present chapter show that the VMS method with standard polynomial bubbles reduces the oscillations in the diffusion-reaction equation but does not completely eliminate them. Thus, still within the VMS framework, in section 6.5 we introduce a fine-scale model with increased flexibility, that adapts locally and automatically to the restrictions and directionality imposed by the mesh and the internal and/or boundary layers, in addition to mitigating the spurious oscillations engendered by boundary and internal layers. Furthermore, the new method is more effective in satisfying the maximum principle, i.e., the spurious oscillations around boundary and internal layers are suppressed in most cases, and their magnitude is substantially reduced in comparison to other methods in the remaining cases. The increased flexibility of the proposed fine-scale model is due to the use of a basis with multiple local functions, which are not necessarily zero at element boundaries (i.e., they are not all “bubble-like”). The importance of capturing directionality in the diffusion-reaction problem has also been recognized in [28], where a stabilization parameter was designed to consider the directionality of the boundary or internal layers. However, the current approach differs in that this feature arises naturally by virtue of the fine-scale model. In addition, sub-scale models that do not vanish at element boundaries were also used in other works, although not necessarily in the context of the diffusion-reaction equation. In [29], local enrichments are used, which emanate from solution of the local strong form of the problem, and interelement continuity is enforced via Lagrange multipliers; in [30], a nested-grid approach is employed and continuity is weakly enforced with a penalty term; in [31], polynomial functions and discontinuous Galerkin (DG) ideas are used at the fine-scale level; and in [32,33] the fine-scale contribution along element boundaries is used to consistently develop interfacial coupling terms. Although the fine-scale model used in this chapter also does not vanish at element boundaries, the proposed approach is distinct in that the element boundary terms are assumed to be negligible, and interelement continuity of the fine-scales is not explicitly enforced. Other methods, such as the Weak Galerkin (WG) method [34] also rely on discontinuous spaces where inter-element continuity is not explicitly enforced through DG-type terms; yet, the jump terms remain bounded and the formulation is well-posed due to a rigorous definition of the discrete weak gradient operator.

The remainder of this chapter is organized as follows: in section 6.2 we state the problem posed by the diffusion-reaction equation and in section 6.3 we present the VMS stabilized method. In section 6.4 we show the derivation of the stabilization terms following the VMS framework. In section 6.5, the proposed fine scale model is introduced, and we study the effects of multiple edge functions and nonpolynomial bubbles on the stabilization parameter. Section 6.5.1 presents the structure of stabilization parameter τ, and section 6.5.2 elaborates on its behavior. Section 6.6 contains the error analysis for the method. Section 6.7 discusses the built-in a posteriori error estimation inherent to VMS methods. Section 6.8 discusses an optional modification to the method in problems with varying coefficients. Section 6.9 presents numerical test cases, and concluding remarks are summarized in section 6.10.

6.2 The Diffusion-Reaction Equation

Consider the closed domain $\Omega \subset \mathbb{R}^{n_{sd}}$ where $n_{sd} \geq 1$ is the number of spatial dimensions. We denote the boundary of the domain by $\partial\Omega$. Let $u : \Omega \to \mathbb{R}$ denote an unknown scalar field in the domain. Consider the diffusion-reaction equation with Dirichlet boundary condition given by

$$-\varepsilon\Delta u + \sigma u = f \qquad \text{in } \Omega \tag{6.1}$$

$$u = g \qquad \text{on } \partial\Omega \tag{6.2}$$

where $\varepsilon \in \mathbb{R}^+$ is the diffusion coefficient, $\sigma \in \mathbb{R}^+$ is the reaction coefficient, $f : \Omega \to \mathbb{R}$ is the forcing function, and g is the prescribed Dirichlet boundary condition. In this chapter ε and σ are assumed constants. We define the linear operator $\mathcal{L} := -\varepsilon\Delta + \sigma$, so equation (6.1) can be written as

$$\mathcal{L}u = f \qquad \text{in } \Omega \tag{6.3}$$

We further define the appropriate spaces of functions

$$\mathcal{S} := \{u \mid u \in H^1(\Omega)\ ,\ u = g \text{ on } \partial\Omega\} \tag{6.4}$$

$$\mathcal{V} := \{w \mid w \in H^1(\Omega)\ ,\ w = 0 \text{ on } \partial\Omega\} \tag{6.5}$$

and formulate the weak form of the problem: Given ε, σ, f, and g, find $u \in \mathcal{S}$ such that

$$(\nabla w,\ \varepsilon\nabla u)_\Omega + (w,\ \sigma u)_\Omega = (w, f)_\Omega \qquad \forall w \in \mathcal{V} \tag{6.6}$$

where $(v_1, v_2)_\Omega = \int_\Omega v_1 \cdot v_2 d\Omega$ denotes the L^2-inner product over the domain Ω. Approximations to the weak form are formulated by the Galerkin method, where functions are taken from finite-dimensional subspaces of (6.4) and (6.5). The formal statement is: Given ε, σ, f, and g, find $u^h \in \mathcal{S}^h \subset \mathcal{S}$ such that

$$(\nabla w^h,\ \varepsilon\nabla u^h)_\Omega + (w^h,\ \sigma u^h)_\Omega = (w^h, f)_\Omega \qquad \forall w^h \in \mathcal{V}^h \subset \mathcal{V} \tag{6.7}$$

where $\mathcal{S}^h$ and $\mathcal{V}^h$ are finite-dimensional subsets of the test and trial spaces respectively. In the context of the finite element method (FEM), these spaces consist of $C^0(\Omega)$ piecewise-differentiable functions $C^1(K)$, commonly polynomials, with compact support defined over non-overlapping subdomains or elements, K. Rewriting equation (6.7), we obtain the following Galerkin method

$$B_{Gal}(w^h, u^h) = L_{Gal}(w^h) \qquad \forall w^h \in \mathcal{V}^h \subset \mathcal{V} \tag{6.8}$$

$$B_{Gal}(w^h, u^h) = (\nabla w^h,\ \varepsilon \nabla u^h)_\Omega + (w^h,\ \sigma u^h)_\Omega \tag{6.9}$$

$$L_{Gal}(w^h) = (w^h, f)_\Omega \tag{6.10}$$

We introduce a non-dimensional parameter termed as the Damkohler number, α, which is a measure of the relative influence of reactive effects versus diffusive effects in the problem. In this chapter, we define the Damkohler number as

$$\alpha = \sqrt{\frac{\sigma h^2}{\varepsilon}} \tag{6.11}$$

where h is a certain length-scale associated with the discretized domain. In the context of the finite element method (FEM), it is useful to define an element Damkohler number, in which the length-scale h is taken to be the characteristic length-scale of the element, commonly taken as the element diameter or the element side length.

6.3 VMS Stabilized Method

It is well known that the Galerkin FEM fails to satisfy the maximum principle in the presence of boundary (or internal) layers of length scales that are shorter than the element characteristic length scale, h. These boundary layers appear for singularly perturbed problems, where the diffusion coefficient is much smaller than the reaction coefficient or, more precisely, when the Damkohler number is larger than some critical value α_{cr}, in particular for the 1D linear case $\alpha > \alpha_{cr} = \sqrt{6}$, as shown in [1,2].

To alleviate the issue of oscillations around the regions of steep gradients that pollute the solution elsewhere, we employed the Variational Multiscale (VMS) framework to derive the stabilized method presented below:

$$B_{VMS}(w^h, u^h) = L_{VMS}(w^h) \qquad \forall w^h \in \mathcal{V}^h \subset \mathcal{V} \tag{6.12}$$

$$B_{VMS}(w^h, u^h) = B_{Gal}(w^h, u^h) - \sum_K \left((-\varepsilon \Delta w^h + \sigma w^h),\ \tau^e(-\varepsilon \Delta u^h + \sigma u^h) \right)_K \tag{6.13}$$

$$L_{VMS}(w^h) = L_{Gal}(w^h) + \sum_K \left((\varepsilon \Delta w^h - \sigma w^h),\ \tau^e f \right)_K \tag{6.14}$$

where τ^e is the stabilization parameter that arises naturally from consistent derivation facilitated by the fine-scale variational form in the VMS framework. The definition of τ^e is derived in section 6.4 and is given in equation (6.42) in terms of a set of basis functions, which are discussed in section 6.5, while a simplified and approximate expression for τ^e, suitable in the regime of very large Damkohler number ($\alpha \to \infty$), is given in equation (6.61) for 4-node quadrilateral elements.

We can rewrite (6.12) by moving all the terms to the left-hand-side and substituting equations (6.13), (6.14), (6.9) and (6.10), to obtain

$$\underbrace{\left(\nabla w^h,\ \varepsilon \nabla u^h\right) + \left(w^h,\ \sigma u^h\right) - \left(w^h, f\right)}_{\text{Standard Galerkin FEM}} + \underbrace{\sum_K \left(\underbrace{\varepsilon \Delta w^h - \sigma w^h}_{-\mathcal{L} = -\mathcal{L}^*},\ \tau^e \overline{R} \right)_K}_{\text{Stabilizing Term}} = 0 \tag{6.15}$$

where $\overline{R} = \mathcal{L} u^h - f = -\varepsilon \Delta u^h + \sigma u^h - f$ denotes the residual of the Euler-Lagrange equation for the trial solution u^h over element interiors. Thus, this method is residual-based and satisfies consistency *ab initio*.

Since the linear operator $\mathcal{L}$ is self-adjoint, the form of the resulting stabilization term is consistent with both GLS and Adjoint methods (e.g. USFEM, VMS). The difference in sign can be absorbed by the definition of the stabilization parameter τ^e. The resulting method can be thought of as GLS with $\tau^e < 0$, or as an Adjoint method with $\tau^e > 0$. For (bi)linear elements, the higher order gradients vanish, which results in equation (6.16) that is implemented for the numerical test cases in the numerical section of this chapter (this is also the resulting form of the methods compared in Figure 6.3).

$$\underbrace{\left(\nabla w^h,\ \varepsilon \nabla u^h\right) + \left(w^h,\ \sigma u^h\right) - \left(w^h, f\right)}_{\text{Standard Galerkin FEM}} + \underbrace{\sum_K \left(-\sigma w^h,\ \tau^e (\sigma u^h - f)\right)_K}_{\text{Stabilizing Term}} = 0 \tag{6.16}$$

6.4 Consistent Derivation of τ^e via the VMS Framework

This section addresses the derivation and definition of the stabilization parameter τ^e for the stabilized method in (6.12)-(6.14). First, consider the weak form from (6.6) and following the Variational Multiscale (VMS) framework [9,10,35,36], it is possible to define a unique additive decomposition of the test and trial functions into linearly independent coarse-scale and fine-scale contributions. This multi-scale decomposition is made precise by introducing the notion of a projection operator $\mathcal{P}_{\mathcal{V}} : \mathcal{V} \rightarrow \bar{\mathcal{V}}$ such that $\forall w \in \mathcal{V}$, $\bar{w} = \mathcal{P}_{\mathcal{V}} w \in \bar{\mathcal{V}}$, and $w' = w - \mathcal{P}_{\mathcal{V}} w \in \mathcal{V}'$ where $\mathcal{V}' = \text{Null}(\mathcal{P}_{\mathcal{V}})$ resulting in the decomposition of the test function space $\mathcal{V} = \bar{\mathcal{V}} \oplus \mathcal{V}'$. Similarly, the space of trial solutions is also decomposed as $\mathcal{S} = \bar{\mathcal{S}} \oplus \mathcal{S}'$ via a projection operator $\mathcal{P}_{\mathcal{S}} : \mathcal{S} \rightarrow \bar{\mathcal{S}}$ such that $\forall u \in \mathcal{S}$, $\bar{u} = \mathcal{P}_{\mathcal{S}} u \in \bar{\mathcal{S}}$ and $u' = u - \mathcal{P}_{\mathcal{S}} u \in \mathcal{S}'$. The resulting additive decomposition can also be written as

$$u = \bar{u} + u' \tag{6.17}$$

$$w = \bar{w} + w' \tag{6.18}$$

where $\overline{(\cdot)}$ denotes the coarse scales, which are associated with the finite-dimensional standard finite element spaces (i.e. $\bar{u} = u^h \in \mathcal{S}^h = \bar{\mathcal{S}}$ and $\bar{w} = w^h \in \mathcal{V}^h = \bar{\mathcal{V}}$). On the other hand, $(\cdot)'$ denotes the fine-scale trial and test functions, $u' \in \mathcal{S}'$ and $w' \in \mathcal{V}'$ respectively, where $\mathcal{S}'$ and $\mathcal{V}'$ are infinite-dimensional spaces. Using the linearity of the test function slot in the weak form (6.6), and the linear independence of $\bar{w}$ and w', a coarse-scale problem and a fine-scale problem are obtained

Coarse-Scale Problem

$$\left(\nabla \bar{w},\ \varepsilon \nabla u\right)_{\Omega} + \left(\bar{w},\ \sigma u\right)_{\Omega} = \left(\bar{w}, f\right)_{\Omega} \qquad \forall \bar{w} \in \bar{\mathcal{V}} \tag{6.19}$$

Fine-Scale Problem

$$\left(\nabla w',\ \varepsilon \nabla u\right)_{\Omega} + \left(w',\ \sigma u\right)_{\Omega} = \left(w', f\right)_{\Omega} \qquad \forall w' \in \mathcal{V}' \tag{6.20}$$

The coarse scale problem can be re-written as follows by using (6.17) and writing domain integrals as sum of integrals over domain interiors

$$\begin{aligned} &\left(\nabla \bar{w},\ \varepsilon \nabla \bar{u}\right)_{\Omega} + \left(\bar{w},\ \sigma \bar{u}\right)_{\Omega} \\ &+ \sum_{K} \left[\left(\nabla \bar{w}\big|_K,\ \varepsilon \nabla u'\big|_K\right)_K + \left(\bar{w}\big|_K,\ \sigma u'\big|_K\right)_K \right] = \left(\bar{w}, f\right)_{\Omega} \end{aligned} \tag{6.21}$$

where the notation $(\cdot)\big|_K$ is used to denote the restriction of a field to its definition over element K. Furthermore, using integration by parts

$$
\begin{aligned}
&\left(\nabla\bar{w},\ \varepsilon\nabla\bar{u}\right)_\Omega + \left(\bar{w},\ \sigma\bar{u}\right)_\Omega \\
&+\sum_K \left[-\left(\varepsilon\Delta\,\bar{w}\big|_K,\ u'\big|_K\right)_K + \left(\sigma\,\bar{w}\big|_K,\ u'\big|_K\right)_K + \left(\varepsilon\mathbf{n}\cdot\nabla\,\bar{w}\big|_K,\ u'\big|_K\right)_{\partial K} \right] = \left(\bar{w}, f\right)_\Omega
\end{aligned}
\tag{6.22}
$$

Rearranging in order to gather fine-scale terms, the following form of the coarse-scale problem is obtained

$$
\begin{aligned}
&\left(\nabla\bar{w},\ \varepsilon\nabla\bar{u}\right)_\Omega + \left(\bar{w},\ \sigma\bar{u}\right)_\Omega \\
&+\sum_K \left(-\varepsilon\Delta\,\bar{w}\big|_K + \sigma\,\bar{w}\big|_K,\ u'\big|_K\right)_K + \sum_K \left(\varepsilon\mathbf{n}\cdot\nabla\,\bar{w}\big|_K,\ u'\big|_K\right)_{\partial K} = \left(\bar{w}, f\right)_\Omega
\end{aligned}
\tag{6.23}
$$

On the other hand, the domain interior integrals in the fine-scale problem from equation (6.20) can be re-written as the sum of integrals over element interiors to yield

$$
\begin{aligned}
&\sum_K \left[\left(\nabla\, w'\big|_K,\ \varepsilon\nabla\,\bar{u}\big|_K\right)_K + \left(w'\big|_K,\ \sigma\,\bar{u}\big|_K\right)_K \right] \\
&+\sum_K \left[\left(\nabla\, w'\big|_K,\ \varepsilon\nabla\, u'\big|_K\right)_K + \left(w'\big|_K,\ \sigma\, u'\big|_K\right)_K \right] = \sum_K \left[\left(w'\big|_K, f\right)_K \right]
\end{aligned}
\tag{6.24}
$$

In order to develop a numerical method, one is required to find an approximation and appropriate finite-dimensional spaces for the fine-scale test and trial functions, $w'\big|_K \approx \tilde{w}'\big|_K \in \mathcal{V}_K$ and $u'\big|_K \approx \tilde{u}'\big|_K \in \mathcal{S}_K$ respectively. In this chapter, the first approximation made is to require that the fine-scale problem in (6.24) be satisfied elementwise

$$
\begin{aligned}
&\left(\nabla\,\tilde{w}'\big|_K,\ \varepsilon\nabla\,\bar{u}\big|_K\right)_K + \left(\tilde{w}'\big|_K,\ \sigma\,\bar{u}\big|_K\right)_K \\
&+\left(\nabla\,\tilde{w}'\big|_K,\ \varepsilon\nabla\,\tilde{u}'\big|_K\right)_K + \left(\tilde{w}'\big|_K,\ \sigma\,\tilde{u}'\big|_K\right)_K = \left(\tilde{w}'\big|_K, f\right)_K \qquad \forall\,\tilde{w}'\big|_K \in \mathcal{V}_K
\end{aligned}
\tag{6.25}
$$

Using integration by parts on the first term gives

$$
\begin{aligned}
&-\left(\tilde{w}'\big|_K,\ \varepsilon\Delta\,\bar{u}\big|_K\right)_K + \left(\tilde{w}'\big|_K,\ \sigma\,\bar{u}\big|_K\right)_K + \left(\tilde{w}'\big|_K,\ \varepsilon\mathbf{n}\cdot\nabla\,\bar{u}\big|_K\right)_{\partial K} \\
&+\left(\nabla\,\tilde{w}'\big|_K,\ \varepsilon\nabla\,\tilde{u}'\big|_K\right)_K + \left(\tilde{w}'\big|_K,\ \sigma\,\tilde{u}'\big|_K\right)_K = \left(\tilde{w}'\big|_K, f\right)_K
\end{aligned}
\tag{6.26}
$$

Rearranging to isolate the terms with fine-scale trial function on the left-hand-side results in equation (6.27)

$$
\left(\nabla\,\tilde{w}'\big|_K,\ \varepsilon\nabla\,\tilde{u}'\big|_K\right)_K + \left(\tilde{w}'\big|_K,\ \sigma\,\tilde{u}'\big|_K\right)_K = -\left(\tilde{w}'\big|_K,\ \bar{R}\big|_K\right)_K - \left(\tilde{w}'\big|_K,\ \varepsilon\mathbf{n}\cdot\nabla\,\bar{u}\big|_K\right)_{\partial K}
\tag{6.27}
$$

where $\bar{R}\big|_K = \mathcal{L}\,\bar{u}\big|_K - f = -\varepsilon\Delta\,\bar{u}\big|_K + \sigma\,\bar{u}\big|_K - f$ denotes the residual of the coarse-scale Euler-Lagrange equation restricted to the element interior.

The finite-dimensional test and trial spaces for the fine-scales $\mathcal{S}_K$ and $\mathcal{V}_K$ are defined in terms of a set of n_b basis functions $B_K = \{b_1^e,\ b_2^e,\ \dots,b_{n_b}^e\}$ as

$$\tilde{\mathcal{V}}_K = \tilde{\mathcal{S}}_{\_K} = \operatorname{span} B_K \tag{6.28}$$

Then, the fine-scale test and trial functions can be expanded as a linear combination of the basis functions within each element

$$\tilde{u}'\big|_K = \sum_{p=1}^{n_b} \hat{u}_p^e b_p^e \tag{6.29}$$

$$\tilde{w}'\big|_K = \sum_{q=1}^{n_b} \hat{w}_q^e b_q^e \tag{6.30}$$

The fine-scale problem is rewritten by substituting (6.29) and (6.30) into (6.27), and employing Einstein indicial notation

$$\begin{aligned}\Big[\left(\nabla b_q^e,\ \varepsilon\nabla b_p^e\right)_K + \left(b_q^e,\ \sigma b_p^e\right)_K\Big]\hat{u}_p^e \\ = -\left(b_q^e,\ \bar{R}\big|_K\right)_K - \left(b_q^e,\ \varepsilon\mathbf{n}\cdot\nabla\bar{u}\big|_K\right)_{\partial K} \qquad \forall b_q^e \in B_K\end{aligned} \tag{6.31}$$

where indices $p,q \in \{1,\dots,n_b\}$. This defines an $n_b \times n_b$ linear system for the element fine-scale coefficients

$$\mathbf{A}\hat{\mathbf{u}} = \mathbf{f} \tag{6.32}$$

where

$$\left[\hat{\mathbf{u}}\right]_p = \hat{u}_p^e \tag{6.33}$$

$$\left[\mathbf{f}\right]_q = -\left(b_q^e,\ \bar{R}\big|_K\right)_K - \left(b_q^e,\ \varepsilon\mathbf{n}\cdot\nabla\bar{u}\big|_K\right)_{\partial K} \tag{6.34}$$

$$\left[\mathbf{A}\right]_{qp} = \left(\nabla b_q^e,\ \varepsilon\nabla b_p^e\right)_K + \left(b_q^e,\ \sigma b_p^e\right)_K \tag{6.35}$$

Solving for the coefficients $\hat{u}_p^e$

$$\hat{u}_p^e = -\left[\mathbf{A}^{-1}\right]_{pq}\Big[\left(b_q^e,\ \bar{R}\big|_K\right)_K + \left(b_q^e,\ \varepsilon\mathbf{n}\cdot\nabla\bar{u}\big|_K\right)_{\partial K}\Big] \tag{6.36}$$

Then, the fine-scale field is recovered by multiplying the coefficients with the basis

$$\tilde{u}'\big|_K = -\left[\mathbf{A}^{-1}\right]_{pq}\Big[\left(b_q^e,\ \bar{R}\big|_K\right)_K + \left(b_q^e,\ \varepsilon\mathbf{n}\cdot\nabla\bar{u}\big|_K\right)_{\partial K}\Big] b_p^e \tag{6.37}$$

A second simplifying approximation is to drop the boundary integral term to obtain

$$\tilde{u}'\big|_K \approx -\left[\mathbf{A}^{-1}\right]_{pq}\left(b_q^e,\ \bar{R}\big|_K\right)_K b_p^e \tag{6.38}$$

Even though this term is nonzero for the definition of the fine scales that is presented in the next section, this approximation is reasonable for singularly perturbed diffusion-reaction problems, in which the diffusion coefficient is much smaller than the reaction coefficient.

Remark: The boundary term in (6.37) drops automatically when the fine-scale basis vanishes at the element boundaries, as is the case with the quadratic bubble commonly used in VMS. The fine-scale basis used for the numerical tests in this chapter does not vanish identically to zero at element boundaries, so neglecting the boundary term is an approximation. A more detailed examination of the omission of the boundary terms will be carried out in future work.

We proceed by using the mean value theorem for integrals on the coarse-scale residual integral

$$\left(b_q^e,\ \bar{R}\big|_K\right)_K = \left(b_q^e,\ 1\right)_K \bar{R}(\mathrm{x}_c) \tag{6.39}$$

for some $\mathrm{x}_c \in K$. Since, in general, this x_c is not known *a priori*, we retain the spatially-varying coarse-scale residual $(b_\beta^e,\ \bar{R})_K \approx (b_\beta^e,\ 1)_K \bar{R}$, which is an approximation that is commonly made in the literature (see for instance [23,26,37]). As a result, the expression for the fine-scale field reduces to

$$\begin{aligned}\tilde{u}'\big|_K &\approx -\left[\mathbf{A}^{-1}\right]_{pq}\left(b_q^e,\ 1\right)_K \bar{R}\big|_K b_p^e \\ &\approx -\tau^e \bar{R}\big|_K\end{aligned} \tag{6.40}$$

where

$$\tau^e = \left[\mathbf{A}^{-1}\right]_{pq}\left(b_q^e,\ 1\right)_K b_p^e \tag{6.41}$$

and in matrix notation

$$\tau^e = \left(\mathbf{b}^e\right)^T \mathbf{A}^{-1}\hat{\mathbf{b}}^e \tag{6.42}$$

where

$$\mathbf{b}^e = [b_1^e, b_2^e, \ldots, b_{n_b}^e]^T \tag{6.43}$$

$$\hat{\mathbf{b}}^e = \int_K \mathbf{b}^e dK \tag{6.44}$$

$$\mathbf{A} = \mathbf{K} + \mathbf{M} \tag{6.45}$$

$$[\mathbf{M}]_{qp} = (b_q^e, \sigma b_p^e)_K = \int_K \sigma b_q^e b_p^e dK \qquad q, p \in \{1, ..., n_b\} \tag{6.46}$$

$$[\mathbf{K}]_{qp} = (\nabla b_q^e, \varepsilon \nabla b_p^e)_K = \int_K \varepsilon \nabla b_q^e \cdot \nabla b_p^e dK \qquad q, p \in \{1, ..., n_b\} \tag{6.47}$$

Remark: The stabilization parameter $\tau^e(\mathrm{x})$ is a scalar and it is spatially-varying, i.e., it depends on **x** through the vector of basis functions $\mathbf{b}^e(\mathrm{x})$, while the integral terms are constant for each element, but may be different from element to element.

We now consider the coarse-scale problem in (6.23) where by similar arguments as before, we drop the element boundary terms and embed the fine-scale solution from equation (6.40) to obtain the following stabilized form

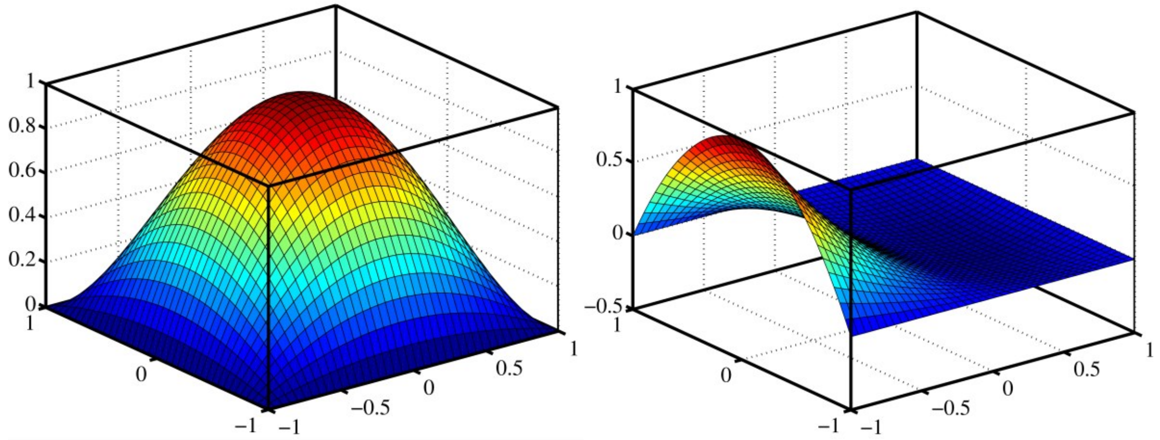
$$\left(\nabla \bar{w},\ \varepsilon \nabla \bar{u}\right)_\Omega + \left(\bar{w},\ \sigma \bar{u}\right)_\Omega - \left(\bar{w}, f\right)_\Omega + \sum_K \left[\left(\varepsilon \Delta \left. \bar{w} \right|_K - \sigma \left. \bar{w} \right|_K ,\ \tau^e \left. \bar{R} \right|_K \right)_K \right] = 0 \tag{6.48}$$

This is identical to equation (6.15), since the coarse scales are given in terms of the finite element spaces (recall $\bar{u} = u^h$, $\bar{w} = w^h$). The final step in the development of the method is to provide a precise representation of the fine-scale model by specifying the set of fine-scale basis functions, which is done in the following section.

Remark: The approximations done in the derivation presented in this section serve the purpose of simplifying the implementation and producing a linear method for a linear problem, while retaining consistency as well as robust stability and accuracy attributes demonstrated in the numerical section.

6.5 A Fine-Scale Model with Multiple Basis Functions

In [21–27,37,38] a single bubble that is zero at the element boundaries was used to represent the fine scales. The example of a biquadratic bubble typically used with 4-node quadrilaterals is shown in Figure 6.1a. In this chapter we consider a fine-scale basis comprised of multiple functions that are not necessarily zero at the element boundary and are not necessarily enforced to be continuous across element interfaces. This amounts to a fine-scale representation with increased flexibility to be able to capture solutions that exhibit steep gradients.



(a) (b)

Figure 6.1. (a) Standard bi-quadratic bubble and (b) quadratic edge function for Q4 element.

In order to define a possible fine-scale basis, first consider the underlying spatial discretization of the domain $\Omega \subset \mathbb{R}^{n_{sd}}$ using Lagrange (nodal) finite elements capable of reproducing complete polynomials of order up to p, whose set of shape functions is given by $B_p = \{P_1^e,\ P_2^e,\ \dots,\ P_{n_p}^e\}$ and each shape function is associated with a node from the element nodal set $X_p = \{\boldsymbol{\xi}_1,\ \boldsymbol{\xi}_2,\ \dots,\ \boldsymbol{\xi}_{n_p}\}$. As a general scheme for defining the set of fine-scale basis functions B_K for such element of arbitrary polynomial order and spatial dimension, we adopt the idea of hierarchical *p*-refinement as discussed in [10]. The set $B_K = \left\{b_1^e,\ b_2^e,\ \dots,\ b_{n_b}^e\right\}$ is chosen to be a subset of the shape functions from a Lagrange element of order $q > p$ of the same family of elements whose set of shape functions is given by $B_q = \{N_1^e,\ N_2^e,\ \dots,\ N_{n_q}^e\}$ associated with nodes $X_q = \{\boldsymbol{\eta}_1,\ \boldsymbol{\eta}_2,\ \dots,\ \boldsymbol{\eta}_{n_q}\}$. More specifically, $B_K := \left\{N_i^e \mid \boldsymbol{\eta}_i \notin X_p\right\}$, i.e., the set of higher order Lagrange basis functions corresponding to the nodes that are not in the nodal set of the underlying FE discretization. Another possible choice is to fix $q = p + 1$ and $B_K := \left\{N_i^e \mid \boldsymbol{\eta}_i \notin X_1\right\}$, i.e., removing only the shape functions associated with the element corner vertices from B_q, and keeping the rest in B_K.

For instance, in the case of 1D linear elements, the basis of the fine scales reduces to a single quadratic function, which is the standard quadratic bubble. For 2D bilinear elements (Q4), the basis for the fine scales consists of the shape functions from the Q9 element corresponding to nodes 5, 6, 7, 8, and 9, as sequenced in the element natural coordinates, shown in Table 6.1.

There, the first 4 functions correspond to the 4 edge functions (so-called "edge bubbles" used in [32,33]) and have the generic shape shown in Figure 6.1b, and the fifth function corresponds to the quadratic bubble shown in Figure 6.1a. For 2D linear triangles (T3), the fine-scales basis is the set of shape functions of the T6 triangle corresponding to nodes 4, 5, and 6 (see Table 6.2). In 3D, for 4-node tetrahedra and 8-node brick elements, the fine-scale basis will include quadratic "face-bubble" functions (in the terminology of [33]), edge functions, as well as actual "bubble-like" interior functions. In general, this yields a variational multiscale method with multiple functions in the fine-scale basis and is denoted as VMSm hereon. Furthermore, numerical results are presented for the case of Q4 and T3 elements in 2D to highlight the features of the method, while maintaining the presentation of ideas simple.

Remark: The idea of having a single non-zero-edge function (Figure 6.1b) as a fine-scale representation has been used in [32,33,39] to develop a stabilized discontinuous Galerkin (DG) method for interfacial coupling. Here, we use similar functions as part of the fine scale basis, but do not enforce the inter-element continuity of the fine scales.

Table 6.1. Basis for the fine scales in the Q4 element (in element natural coordinates).

Fine-scale basis function	Functional form	Equivalent shape function (from Q9)
b_1^e	$-\frac{1}{2}(1-\xi)(1+\xi)(1-\eta)\eta$	N_5^e
b_2^e	$\frac{1}{2}\xi(1+\xi)(1-\eta)(1+\eta)$	N_6^e
b_3^e	$\frac{1}{2}(1-\xi)(1+\xi)(1+\eta)\eta$	N_7^e
b_4^e	$-\frac{1}{2}\xi(1-\xi)(1-\eta)(1+\eta)$	N_8^e
b_5^e	$(1-\xi)(1+\xi)(1-\eta)(1+\eta)$	N_9^e

Table 6.2. Basis for the fine scales in the T3 element (in element natural coordinates).

Fine-scale basis function	Functional form	Equivalent shape function (from T6)
b_1^e	$4\xi(1-\xi-\eta)$	N_4^e
b_2^e	$4\xi\eta$	N_5^e
b_3^e	$4\eta(1-\xi-\eta)$	N_6^e

One feature of this choice of fine-scale basis is that the magnitude of the fine-scale field along each element edge is associated with separate basis functions. This also allows the response of the fine-scale field along the element ξ -axis to be somewhat independent from its response along the η -axis. In contrast, the symmetries of the standard quadratic bubble cause the profile of the fine-scale field along the ξ - and η -axes within each element to be identical, which can be overly constraining in some situations. The added feature of the fine-scale basis proposed here endows the method with increased flexibility for capturing sharp layers and directionality in the discrete problem, which is further explored in the following sections and specifically highlighted in section 6.9.2.

As noted in section 6.4, computation of the stabilization parameter associated with a fine-scale basis having n_b functions incurs the cost of solving a small ($n_b \times n_b$) linear system within each element. As the problem scales, the number of small matrices increases with the number of elements, but the size of each small matrix stays the same. In other words, the number of local problems to be solved is increased, but the size of each local problem is same; therefore, this additional cost is $O(n)$, where n is the number of elements (or, roughly equivalently, the number of degrees of freedom). On the other hand, the cost of solving the matrix at the global level does not change because the number of elements is not affected by the choice of fine-scale basis. Assembly and solution of the global system of equations therefore remains the dominant factor in the cost of computation. In the best-case scenario, a direct sparse Cholesky decomposition in 2D takes $O(n^{1.5})$ work and in 3D takes $O(n^2)$ work, while a preconditioned conjugate gradient (CG) method is $O(n^{1.25})$ in 2D, and $O(n^{1.17})$ in 3D [40]. These are optimal estimates, and in practice, iterative methods may see their convergence slowed down. Even for these cases, as the problem size scales the leading term remains the work associated with solution of the global level linear system, since the proposed method introduces only $O(n)$ work. Moreover, this $O(n)$ work does not need to be done sequentially and it can easily be parallelized. From the perspective of modern computer architecture with resident memory on the processing nodes, even this step of locally solving a small element level problem does not incur the gather-scatter communication cost. Therefore, the additional computational cost for the proposed method is not significant, and the total computational cost is on par (to leading order term) with other similar methods.

6.5.1 Structure of the stabilization parameter

As an example of the resulting shape of the stabilization parameter τ^e for a Q4 element, consider Figure 6.2, which was obtained from a linear combination of functions in Table 6.1. Note that by virtue of the choice of basis, the fine-scale model may be nonzero over element boundaries, but it vanishes to zero at the element nodes. The resulting shape resembles the "canopy" functions mentioned in [41].

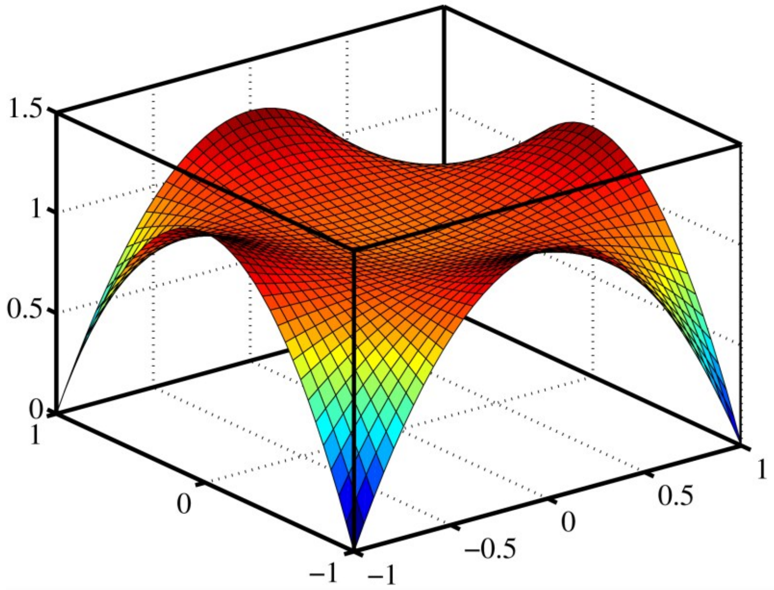


Figure 6.2. Example of the fine-scale representation after a linear combination of the standard bubble and 4 edge functions.

Remark: The elements coincident with the Dirichlet boundary would have non-zero fine scales between the end-nodes along that edge. In our implementation we have not explicitly set fine scales at Dirichlet boundaries to be exactly zero.

In order to simplify the subsequent analysis, we will consider the average value of the stabilization parameter in (6.42), over the element subdomain K. The average value of the stabilization parameter over an element, $\bar{\tau}^e$, is given by the following equation

$$\bar{\tau}^e = \left(\bar{\mathbf{b}}^e\right)^T \mathbf{A}^{-1}\,\hat{\mathbf{b}}^e \tag{6.49}$$

where

$$\bar{\mathbf{b}}^{\mathbf{e}} = [\mathrm{meas}(K)]^{-1}\int_K \mathbf{b}^{\mathbf{e}} dK \tag{6.50}$$

is the vector of fine-scale basis functions' average values, and $\mathbf{b}^e$, $\hat{\mathbf{b}}^e$, and $\mathbf{A}$ are as in (6.43), (6.44), and (6.45) respectively.

Let us consider the structure of the stabilization parameter in one-dimension. Although the fine-scale basis proposed in this chapter for the case of 1D linear elements coincides with the standard quadratic bubble, it is worthwhile to work out the structure of the stabilization parameter, τ^e, of the 2D case via a 1D analog in order to reduce the complexity. For the 1D analog, consider a 1D element of length h, aligned along the x-axis. The 1D analog (or projection) for the 2D fine-scale basis is taken as the standard shape functions of a quadratic element, namely

$$\left.\begin{aligned} b_1^e &= \frac{2}{h^2}(x-h/2)(x-h) \\ b_2^e &= \frac{2}{h^2}(x-h/2)x \\ b_3^e &= -\frac{4}{h^2}(x-h)x \end{aligned}\right\} \tag{6.51}$$

From which the following quantities are obtained by recalling (6.43)-(6.47) and (6.50)

$$\mathbf{b}^{\mathbf{e}} = [b_1^e \quad b_2^e \quad b_3^e]^T \tag{6.52}$$

$$[\bar{\mathbf{b}}^e]_q = \left[\mathrm{meas}(K)\right]^{-1}\int_K b_q^e dK \quad\Rightarrow\quad \bar{\mathbf{b}}^e = \frac{1}{6}\left[1 \quad 1 \quad 4\right]^T \tag{6.53}$$

$$[\hat{\mathbf{b}}^e]_q = \int_K b_q^e dK \quad\Rightarrow\quad \hat{\mathbf{b}}^e = \frac{h}{6}\left[1 \quad 1 \quad 4\right]^T \tag{6.54}$$

$$\left[\mathbf{K}\right]_{qp} = (\nabla b_q^e,\ \varepsilon\nabla b_p^e)_K \quad\Rightarrow\quad \mathbf{K} = \frac{\varepsilon}{3h}\begin{bmatrix} 7 & 1 & -8 \\ 1 & 7 & -8 \\ -8 & -8 & 16 \end{bmatrix} \tag{6.55}$$

$$[\mathbf{M}]_{qp} = (b_q^e,\ \sigma b_p^e)_K \qquad \Rightarrow \quad \mathbf{M} = \frac{\sigma h}{30}\begin{bmatrix} 4 & -1 & 2 \\ -1 & 4 & 2 \\ 2 & 2 & 16 \end{bmatrix} \tag{6.56}$$

$$\mathbf{A} = \mathbf{K} + \mathbf{M} \qquad \Rightarrow \quad \mathbf{A} = \frac{\varepsilon}{3h}\begin{bmatrix} 7 & 1 & -8 \\ 1 & 7 & -8 \\ -8 & -8 & 16 \end{bmatrix} + \frac{\sigma h}{30}\begin{bmatrix} 4 & -1 & 2 \\ -1 & 4 & 2 \\ 2 & 2 & 16 \end{bmatrix} \tag{6.57}$$

Then, substituting (6.53), (6.54) and (6.57) into (6.49), the average value of the stabilization parameter in the given element is

$$\begin{aligned} \bar{\tau}^e &= \left(\bar{\mathbf{b}}^e\right)^T \mathbf{A}^{-1}\hat{\mathbf{b}}^e \\ &= \frac{1}{\sigma} \end{aligned} \tag{6.58}$$

Remark: The 1D analog results in a very simple form of the average value of the stabilization parameter $\bar{\tau}^e$, where the stabilization parameter is independent of the diffusion coefficient since the matrix **K** is singular, which is due to the fact that the 1D projection employed for the 2D fine-scale basis is a complete polynomial space. However, this is not the case in the actual two-dimensional case, since the fine-scale basis is not a complete polynomial space over the element domain, and thus, the matrix **K** is not singular in 2D.

In Table 6.3, we show the resulting expression for the average value of the stabilization parameter for different choices of fine-scale basis in 1D. In a similar way, it is possible to work out the expressions for $\bar{\tau}^e$ in the 2D case for a square Q4 element, by substituting the choice of fine-scale basis into equation (6.49). Some of these expressions are included in Table 6.4.

Table 6.3. Average value of the stabilization parameter for different choices of fine-scale basis in 1D

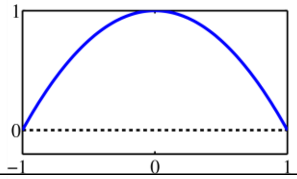
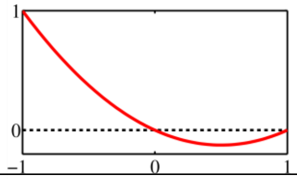
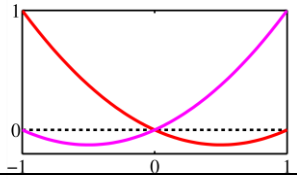
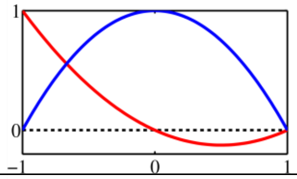
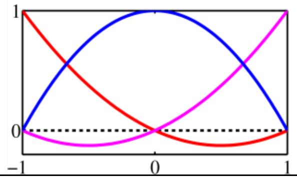
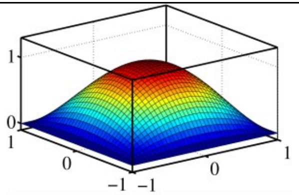
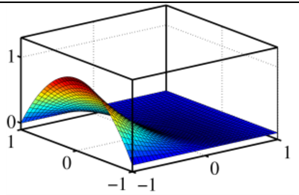
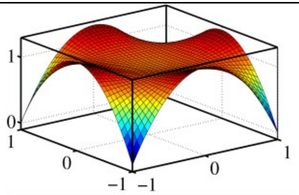
Plots of the Basis Functions	Average value of the element stabilization parameter
	$\bar{\tau}^e\sigma = \dfrac{5\alpha^2}{6\alpha^2+60}$
	$\bar{\tau}^e\sigma = \dfrac{5\alpha^2}{24\alpha^2+420}$
	$\bar{\tau}^e\sigma = \dfrac{5\alpha^2}{9\alpha^2+80}$
	$\bar{\tau}^e\sigma = \dfrac{8\alpha^4+240\alpha^2}{9\alpha^4+312\alpha^2+720}$
	$\bar{\tau}^e\sigma = 1$

Table 6.4. Average value of the stabilization parameter for different choices of fine-scale basis in 2D (square Q4 element)[2].

Plots of the Basis Functions	Average value of the element stabilization parameter
	$$\bar{\tau}^e \sigma = \frac{25\alpha^2}{36\alpha^2 + 720}$$
	$$\bar{\tau}^e \sigma = \frac{25\alpha^2}{144\alpha^2 + 3960}$$
	$$\bar{\tau}^e \sigma = \frac{35\alpha^4 + 2700\alpha^2}{36\alpha^4 + 2880\alpha^2 + 7200}$$

For non-square quadrilateral elements, as well as for anisotropic diffusion, there is an increase in the complexity of the problem. However, the expression for the stabilization parameter in (6.42), combined with the choice of fine-scale basis discussed in section 6.5, can naturally capture the additional intricacy. For example, Table 6.5 shows expressions for the average value of the stabilization parameter $\bar{\tau}^e$ over a rectangular element of size h_x by h_y. Damkohler numbers are defined for each direction as follows

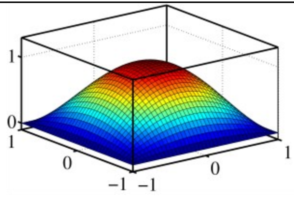
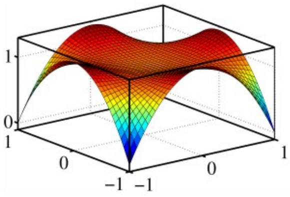
$$\alpha_x = \sqrt{\frac{\sigma h_x^2}{\varepsilon}} \tag{6.59}$$

$$\alpha_y = \sqrt{\frac{\sigma h_y^2}{\varepsilon}} \tag{6.60}$$

---

[2] The last row of Table 6.4 corresponds to the basis from Table 6.1. The plot shown corresponds to an example of a linear combination of that basis.

Table 6.5. Average value of the stabilization parameter for different choices of fine-scale basis in 2D (rectangular Q4 element)[3].

Plots of the Basis Functions	Average value of the element stabilization parameter
	$\bar{\tau}^e \sigma = \dfrac{25\alpha_x^2\alpha_y^2}{36\alpha_x^2\alpha_y^2 + 360\alpha_x^2 + 360\alpha_y^2}$
	$\bar{\tau}^e \sigma = \dfrac{5\begin{pmatrix} 3600\alpha_x^4 + 25200\alpha_x^2\alpha_y^2 + 3600\alpha_y^4 \\ +480\alpha_x^4\alpha_y^2 + 480\alpha_x^2\alpha_y^4 + 7\alpha_x^4\alpha_y^4 \end{pmatrix}}{36\begin{pmatrix} 6000\alpha_x^2 + 6000\alpha_y^2 + 600\alpha_x^4 + 3800\alpha_x^2\alpha_y^2 \\ +600\alpha_y^4 + 70\alpha_x^4\alpha_y^2 + 70\alpha_x^2\alpha_y^4 + \alpha_x^4\alpha_y^4 \end{pmatrix}}$

Furthermore, when the diffusivity is negligible compared to the reactive effects, i.e., for very large Damkohler number, the contribution from the diffusion term in equation (6.47) can be dropped and the (spatially-varying) stabilization parameter over each element $\tau^e(\mathrm{x})$ simplifies to the following expression in element natural coordinates for Q4 elements.

$$\tau^e(\mathrm{x}(\xi)) \approx \frac{-25\xi^2\eta^2 + 5\xi^2 + 5\eta^2 + 15}{16\sigma} \tag{6.61}$$

Remark: This term is amenable for direct implementation into the element subroutine of any existing FEM code, resulting in a simple approximated form of the method presented herein and that is reasonable for problems with very large Damkohler number ($\alpha \to \infty$).

We emphasize that the stabilization parameter from equation (6.42) is the one to be implemented in a general computer code and it is used to generate the numerical results in section 6.9. The analysis presented in this section utilized the average value from equation (6.49) only with the purpose of elucidating the scaling and structure of the stabilization parameter, as well as how it is affected by the choice of fine-scale basis.

---

[3] The last row of Table 6.5 corresponds to the basis from Table 6.1. The plot shown corresponds to an example of a linear combination of that basis.

6.5.2 Behavior of stabilization parameter with Damkohler number

To study the behavior of the stabilization term obtained through the derivation in the preceding section, consider a square element and a reaction coefficient σ equal to 1. For this analysis, we take the average value of the stabilization parameter over the element, $\bar{\tau}^e$, given by equation (6.49). This would make our analysis compatible with that of other methods proposed in the literature wherein stabilization parameter is given as a constant with no spatial variation within an element.

In Figure 6.3 we compare this variationally derived expression of the stabilization parameter with other definitions for τ^e reported in the literature [15,18,28,42], which are presented in Table 6.6, for varying value of the Damkohler number, α. All the methods in the references for which τ^e is compared with, reduce to the same stabilized form as in equation (6.16) for the scalar diffusion-reaction equation with constant coefficients and linear or bilinear elements.

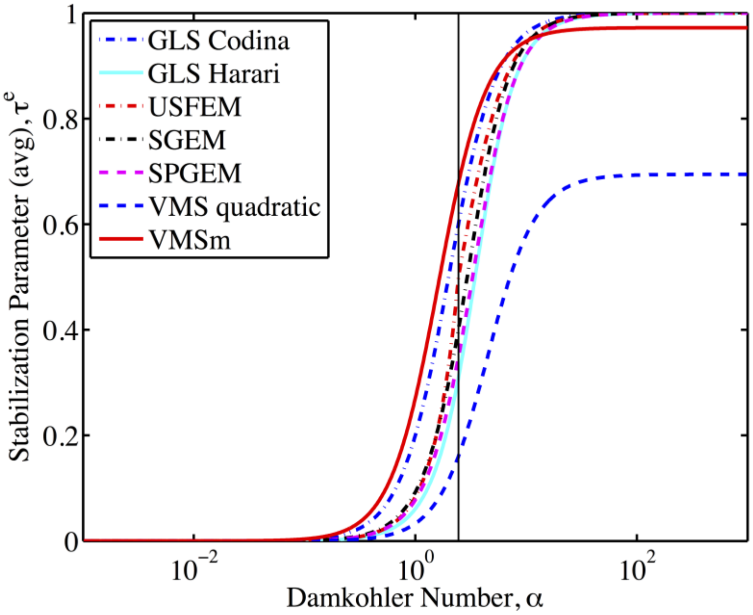


Figure 6.3. Average value of the stabilization parameter τ^e vs. Damkohler number (α) over a square Q4 element with $\sigma = 1$ for different methods. The vertical line marks the value of the critical Damkohler number $\alpha_{cr} = \sqrt{6}$.

From Figure 6.3, the values of the stabilization parameter are higher when the problem is singularly perturbed (small diffusivity, high Damkohler numbers, or reaction-dominated), and lower in the diffusion dominated regime (large diffusivity or low Damkohler number). Note as well that for VMS with the standard biquadratic bubble, $\bar{\tau}^e$ asymptotes to a maximum value of $25/36 \approx 0.7$, while the parameters from other methods are closer to 1.0 in the reaction-dominated regime. This is indicative of the shortcoming of the standard biquadratic bubble in suppressing oscillations in the singular diffusion-reaction problem for large α , since the scaling of the stabilization term is less than what is suggested in the other methods.

Table 6.6. Stabilization parameters for different stabilized methods[4].

Method	Stabilization Parameter (rewritten in terms of α)
GLS Codina	$\tau^e \sigma = \dfrac{\alpha^2}{4+\alpha^2}$
GLS Harari	$\tau^e \sigma = 1 + \dfrac{8}{\alpha^2} \cdot \dfrac{1-\cosh\left(\alpha\sqrt{3}/2\right)}{2+\cosh\left(\alpha\sqrt{3}/2\right)}$
USFEM	$\tau^e \sigma = \dfrac{\alpha^2}{\alpha^2 \max\{\ 1\ ,\ 6/\alpha^2\ \}+6}$
SGEM	$\tau^e \sigma = 1 - \dfrac{3(\sinh(\alpha)^2 - \alpha^2)}{(\alpha \sinh(\alpha))^2}$
SPGEM	$\tau^e \sigma = 1 - \dfrac{4(\cosh(\alpha)-1)}{\alpha^2(1+\cosh(\alpha))}$
VMS quadratic	$\bar{\tau}^e \sigma = \dfrac{25\alpha^2}{36\alpha^2+720}$
VMSm	$\bar{\tau}^e \sigma = \dfrac{35\alpha^4+2700\alpha^2}{36\alpha^4+2880\alpha^2+7200}$

One possible argument is that the issues with the standard quadratic bubble can be addressed by choosing a different functional form such that its average value asymptotes to a higher value as $\alpha \to \infty$, without having to deal with the complexity introduced by a basis with multiple functions as the one in Table 6.1. However, the numerical results in section 6.9 demonstrate that the suggested asymptotic behavior of the stabilization parameter alone does not guarantee that oscillations will be suppressed. To rigorously test this, we defined another VMS method, termed here as VMSroot, where the fine-scale basis consists of a single bubble function such that the resulting average value of the stabilization parameter asymptotes to the same value as the one in VMSm for the case of square Q4 elements (see Figure 6.4). The fine-scale bubble function for VMSroot in the case of Q4 elements is shown in Figure 6.5 and is given by equation (6.62), while its counterpart for T3 elements is given by equation (6.63)

---

[4] The expressions for τ^e from different papers have been rewritten in terms of Damkohler number α as defined in equation (6.11) in order to facilitate the comparison.

$$b^e_{\mathrm{Q4}} = [(1-\xi)(1+\xi)(1-\eta)(1+\eta)]^{0.1} \tag{6.62}$$

$$b^e_{\mathrm{T3}} = [27\xi\eta(1-\xi-\eta)]^{0.1} \tag{6.63}$$

These functions are non-polynomial and were obtained by raising a polynomial bubble function to the 0.1 power. A similar and more sophisticated approach to define a non-polynomial bubble function was presented in [43].

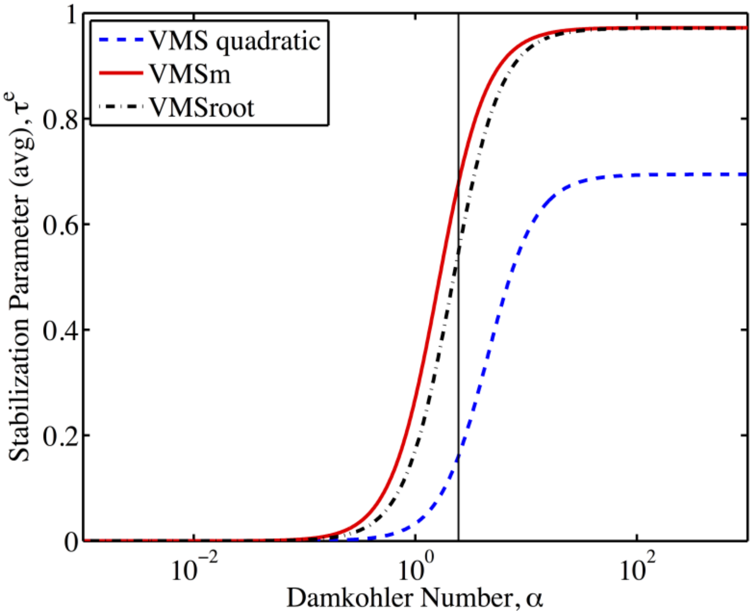


Figure 6.4. Average value of the stabilization parameter $\bar{\tau}^e$ over a square Q4 element with varying Damkohler number, $\sigma = 1$, for different VMS methods. The vertical line marks the value of the critical Damkohler number $\alpha_{cr} = \sqrt{6}$.

Regarding the asymptotic behavior of the stabilization parameter for VMSm shown in Figure 6.3 and Figure 6.4, we note that it improves significantly on the standard quadratic bubble, but it does not tend to a value of 1 in the reaction-dominated zone, as other stabilized methods do. This discrepancy is likely connected to the use of an incomplete polynomial space of basis functions for the fine-scale field (the basis in Table 6.1 cannot represent all polynomials of up to order 2). Nevertheless, numerical tests in section 6.9 show that the method is effective in suppressing spurious oscillations. We also reiterate that the presented approach has the added built-in feature of automatically capturing directionality in the problem (direction of boundary layer and uneven element aspect ratio) via the terms in the definition of τ^e that account for material parameters and the gradients of basis functions. This is discussed in more detail in section 6.9, in view of the numerical results from section 6.9.2. Another difference between our expression and that of the various definitions of τ^e is their transition region from the reaction-dominated to the diffusion-dominated regimes. In this case, we note that the proposed approach yields the leftmost curve, indicating the possibility of being overly diffusive for problems in this transition zone. We believe that this may be remediated by incorporating the element boundary terms in (6.37) that were neglected during the derivation. This will be studied further in a future work.

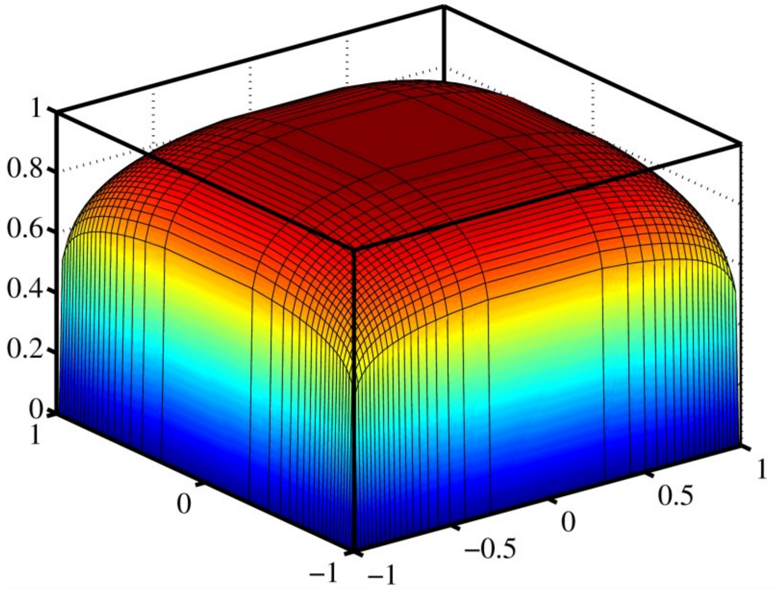


Figure 6.5. Non-polynomial bubble for Q4 element.

6.6 Error Analysis

In this section we show *a priori* mathematical analysis of the stability (Lemma 6.2) and convergence (Lemma 6.4) of the method. The analysis is done in general for a stabilized method of the form given in (6.12)-(6.14). However, in order to simplify the analysis, the stabilization parameter τ^e is assumed constant within each element, and for the case of the spatially-varying stabilization parameter τ^e from equation (6.41) or (6.42), it is replaced by its average value over the element $\overline{\tau}^e$ given in equation (6.49). We assume a convex polygonal domain Ω discretized into element subdomains K of side length h. We employ some preliminary results, beginning with the following inverse estimate.

$$\left\|\Delta w^h\right\|_{0,K}^2 \quad \leq \quad \frac{C_{inv}}{h^2}\left\|\nabla w^h\right\|_{0,K}^2 \qquad \forall w^h \in \mathcal{V}^h \tag{6.64}$$

where $\left\|\bullet\right\|_{0,K}$ denotes the L_2-norm over the element subdomain K. As given in [44], some specific values of the inverse estimate constant are $C_{inv} \geq 0$ for right isosceles T3 and square Q4 elements, $C_{inv} \geq 48$ for T6, and $C_{inv} \geq 24$ for Q9. An additional preliminary result is Proposition 6.1 below.

Proposition 6.1. For some positive constant $\beta < 1$ (strictly less than 1) independent of σ, ε, and h, and for α as defined in (6.11), the inequality

$$\overline{\tau}^e \leq \frac{\beta}{\sigma}\left(\frac{\alpha^2}{\alpha^2 + C_{inv}}\right) \tag{6.65}$$

is satisfied by the average value of the element stabilization parameter induced by the proposed fine-scale basis for the case of 2D square Q4 elements and right isosceles T3 elements.

Proof. Consider a square Q4 element and a right isosceles T3 element, both of side length h. Employing the fine-scale basis functions from Table 1 for Q4 and from Table 2 for T3 in equation (6.49), and carrying out the calculations, the following expressions for the average value of the stabilization parameter are obtained

$$\overline{\tau}^e_{\mathrm{Q4}} \quad = \quad \frac{2700\alpha^2 + 35\alpha^4}{36\left(200 + 80\alpha^2 + \alpha^4\right)\sigma} \tag{6.66}$$

$$\overline{\tau}^e_{\mathrm{T3}} \quad = \quad \frac{1500\alpha^2 + 15\alpha^4}{16\left(450 + 105\alpha^2 + \alpha^4\right)\sigma} \tag{6.67}$$

Recalling that $\alpha \geq 0$ since $\sigma, \varepsilon, h \geq 0$, it is straight-forward to verify that $\overline{\tau}^e$ is bounded by a more compact expression as follows

$$\overline{\tau}_{\mathrm{Q4}}^{e} \;=\; \frac{1}{\sigma}\left[\frac{2700\alpha^2+35\alpha^4}{36\left(200+80\alpha^2+\alpha^4\right)}\right] \;\le\; \frac{1}{\sigma}\left[\frac{35\alpha^2}{36\alpha^2+93}\right] \tag{6.68}$$

$$\overline{\tau}_{\mathrm{T3}}^{e} \;=\; \frac{1}{\sigma}\left[\frac{1500\alpha^2+15\alpha^4}{16\left(450+105\alpha^2+\alpha^4\right)}\right] \;\le\; \frac{1}{\sigma}\left[\frac{15\alpha^2}{16\alpha^2+72}\right] \tag{6.69}$$

Therefore, it is possible to write the upper bound in a generic form for both cases as

$$\overline{\tau}^{e} \;\le\; \frac{1}{\sigma}\left[\frac{(n-1)\alpha^2}{n\alpha^2+p}\right] \tag{6.70}$$

with $p > n > 1$, where $n = 36$ and $p = 93$ in the case of square Q4 elements, and $n = 16$ and $p = 72$ in the case of right isosceles T3 elements. Furthermore, the expression in (6.70) can be bounded by

$$\overline{\tau}^{e} \;\le\; \frac{1}{\sigma}\left[\frac{(n-1)\alpha^2}{n\alpha^2+p}\right] \;\le\; \frac{\chi}{\sigma}\left[\frac{\alpha^2}{\alpha^2+C_{inv}}\right] \tag{6.71}$$

with $\chi = \max\left((n-1)C_{inv}/p,\ (n-1)/n\right)$. From [44], in the case of right isosceles T3 and square Q4 elements, it is required only that $C_{inv} \ge 0$ and hence it is always possible to choose $C_{inv} \le p/n$ so that $\chi = (n-1)/n < 1$. Then, by taking $\beta = \chi$, which is independent of σ, ε, and h, inequality (6.65) is satisfied. □

Remark. Inequality (6.65) is a sufficient condition on the value of the stabilization parameter τ^e to ensure stability and convergence of any method of the form given in (6.12)-(6.14), as is proven below in Lemmas 6.2 and 6.4. Extension of those Lemmas for the proposed method in 3D and/or higher order elements is straightforward by verifying that the fine-scale basis for those cases induces a stabilization parameter that satisfies inequality (6.65) as is shown in Proposition 6.1 for the case of right isosceles T3 and square Q4 elements in 2D.

Lemma 6.1 (Orthogonality of error). Let $e = u - u^h$ be the total error between the exact solution u and the numerical approximation u^h. Then, the bilinear form in (6.13) satisfies the following

$$B_{VMS}(w^h, e) = 0 \qquad \forall w^h \in \mathcal{V}^h \tag{6.72}$$

Proof. By the standard arguments of consistency of the method and the linearity of the bilinear form on the trial solution slot, result (6.72) is obtained directly, i.e., e is orthogonal to the test function space $\mathcal{V}^h$ with respect to the bilinear form $B_{VMS}(\cdot,\cdot)$. □

Lemma 6.2 (Stability). Given positive constant values for the coefficients σ and ε, and assuming τ^e is constant within each element and satisfies inequality (6.65), then

$$B_{VMS}(u^h,u^h) \;\geq\; \sum_K C_s\left[\sigma\left\|u^h\right\|_{0,K}^2+\varepsilon\left\|\nabla u^h\right\|_{0,K}^2\right] \qquad \forall u^h \in \mathcal{V}^h \tag{6.73}$$

where C_s is a positive constant independent of σ, ε, and h.

Proof. Recall the definition of the bilinear form in equation (6.13).

$$\begin{aligned} B_{VMS}(u^h,u^h) &= \sigma(u^h,u^h)+\varepsilon(\nabla u^h,\nabla u^h)-\sum_K \tau^e(\sigma u^h-\varepsilon\Delta u^h \;,\; \sigma u^h-\varepsilon\Delta u^h)_K \\ &= \sum_K\left[\sigma\left\|u^h\right\|_{0,K}^2+\varepsilon\left\|\nabla u^h\right\|_{0,K}^2-\tau^e\sigma^2\left\|u^h\right\|_{0,K}^2-\tau^e\varepsilon^2\left\|\Delta u^h\right\|_{0,K}^2\right] \\ &\quad +\sum_K 2\tau^e\sigma\varepsilon(u^h,\Delta u^h)_K \end{aligned} \tag{6.74}$$

Invoking Cauchy-Schwarz and Arithmetic-Mean (AM) Geometric-Mean (GM) inequalities for the last term on the RHS.

$$\begin{aligned} B_{VMS}(u^h,u^h) &\geq \sum_K\left[\sigma\left\|u^h\right\|_{0,K}^2+\varepsilon\left\|\nabla u^h\right\|_{0,K}^2-\tau^e\sigma^2\left\|u^h\right\|_{0,K}^2-\tau^e\varepsilon^2\left\|\Delta u^h\right\|_{0,K}^2\right] \\ &\quad +\sum_K \tau^e\sigma\varepsilon\left(-\frac{1}{\gamma}\left\|u^h\right\|_{0,K}^2-\gamma\left\|\Delta u^h\right\|_{0,K}^2\right) \\ &= \sum_K\left[\left(1-\tau^e\sigma-\frac{\tau^e\varepsilon}{\gamma}\right)\sigma\left\|u^h\right\|_{0,K}^2+\varepsilon\left\|\nabla u^h\right\|_{0,K}^2\right] \\ &\quad +\sum_K\left(-\tau^e\varepsilon^2-\gamma\tau^e\sigma\varepsilon\right)\left\|\Delta u^h\right\|_{0,K}^2 \end{aligned} \tag{6.75}$$

where $\gamma > 0$ is an arbitrary constant. Using the inverse estimate on the Laplacian term.

$$\begin{aligned} B_{VMS}(u^h,u^h) &\geq \sum_K \left[\left(1-\tau^e\sigma-\frac{\tau^e\varepsilon}{\gamma} \right) \sigma \left\| u^h \right\|_{0,K}^2 + \varepsilon \left\| \nabla u^h \right\|_{0,K}^2 \right] \\ &\quad + \sum_K \left(-\tau^e\varepsilon^2 - \gamma\tau^e\sigma\varepsilon \right) \frac{C_{inv}}{h^2} \left\| \nabla u^h \right\|_{0,K}^2 \\ &= \sum_K \left(1-\tau^e\sigma-\frac{\tau^e\varepsilon}{\gamma} \right) \sigma \left\| u^h \right\|_{0,K}^2 \\ &\quad + \sum_K \left(1-\tau^e\varepsilon\frac{C_{inv}}{h^2} - \gamma\tau^e\sigma\frac{C_{inv}}{h^2} \right) \varepsilon \left\| \nabla u^h \right\|_{0,K}^2 \end{aligned} \tag{6.76}$$

Choosing $\gamma = h^2/C_{inv}$ (with $C_{inv} > 0$), and using the definition of the Damkohler number α from equation (6.11)

$$B_{VMS}(u^h,u^h) \geq \sum_K \left(1-\tau^e\sigma\left(\frac{\alpha^2 + C_{inv}}{\alpha^2} \right) \right) \left[\sigma \left\| u^h \right\|_{0,K}^2 + \varepsilon \left\| \nabla u^h \right\|_{0,K}^2 \right] \tag{6.77}$$

Using inequality (6.65)

$$B_{VMS}(u^h,u^h) \geq \sum_K (1-\beta) \left[\sigma \left\| u^h \right\|_{0,K}^2 + \varepsilon \left\| \nabla u^h \right\|_{0,K}^2 \right] \tag{6.78}$$

where letting $C_s = 1-\beta$ completes the proof. □

To continue with the error analysis, we define $\tilde{u}^h \in \mathcal{V}^h$ to be an interpolant of the exact solution u. The total error can be written as $e = \eta + e_h$, where $\eta = u - \tilde{u}^h$ is the interpolation error and $e_h = \tilde{u}^h - u^h \in \mathcal{V}^h$ is the FE error.

Lemma 6.3 (FE error estimate). Under the conditions on σ, ε, and τ from Lemma 6.2, the stability norm of the FE error can be bounded in terms of the interpolation error as

$$\sum_K \left[\sigma \left\| e_h \right\|_{0,K}^2 + \varepsilon \left\| \nabla e_h \right\|_{0,K}^2 \right] \leq \sum_K C_{FE} \left[\sigma \left\| \eta \right\|_{0,K}^2 + \varepsilon \left\| \nabla \eta \right\|_{0,K}^2 + h^2\varepsilon \left\| \Delta\eta \right\|_{0,K}^2 \right] \tag{6.79}$$

where C_{FE} is a positive constant independent of σ, ε, and h.

Proof. Using the stability result on e_h and the orthogonality of the error with respect to the bilinear form from Lemma 6.1

$$\begin{aligned}\sum_K C_s\left[\sigma\|e_h\|_{0,K}^2+\varepsilon\|\nabla e_h\|_{0,K}^2\right]&\le B_{VMS}(e_h,e_h)\\&= B_{VMS}(e-\eta,e_h)\\&= B_{VMS}(e,e_h)-B_{VMS}(\eta,e_h)=-B_{VMS}(\eta,e_h)\\&\le \left|B_{VMS}(\eta,e_h)\right|\end{aligned}\tag{6.80}$$

Using the definition of the bilinear form and triangle inequality

$$\begin{aligned}&\left|B_{VMS}(\eta,e_h)\right|\\&\quad\le \sum_K\sigma\left|(\eta,e_h)_K\right|+\sum_K\varepsilon\left|(\nabla\eta,\nabla e_h)_K\right|+\sum_K\tau^e\left|(\sigma\eta-\varepsilon\Delta\eta\ ,\ \sigma e_h-\varepsilon\Delta e_h)_K\right|\end{aligned}\tag{6.81}$$

Using Cauchy-Schwartz and the AM-GM inequalities for each term on the RHS

$$\begin{aligned}&\left|B_{VMS}(\eta,e_h)\right|\\&\quad\le \sum_K\frac{\sigma}{2}\left[\gamma_1\|\eta\|_{0,K}^2+\frac{1}{\gamma_1}\|e_h\|_{0,K}^2\right]+\sum_K\frac{\varepsilon}{2}\left[\gamma_2\|\nabla\eta\|_{0,K}^2+\frac{1}{\gamma_2}\|\nabla e_h\|_{0,K}^2\right]\\&\qquad+\sum_K\frac{\tau^e}{2}\left[\gamma_3\|\sigma\eta-\varepsilon\Delta\eta\|_{0,K}^2+\frac{1}{\gamma_3}\|\sigma e_h-\varepsilon\Delta e_h\|_{0,K}^2\right]\end{aligned}\tag{6.82}$$

where $\gamma_1,\gamma_2,\gamma_3>0$ are arbitrary positive constants. Using triangle inequality and AM-GM inequality on the last two terms on the RHS and gathering common terms.

$$\begin{aligned}&\left|B_{VMS}(\eta,e_h)\right|\\&\quad\le \sum_K\frac{\sigma}{2}\left[\gamma_1\|\eta\|_{0,K}^2+\frac{1}{\gamma_1}\|e_h\|_{0,K}^2\right]+\sum_K\frac{\varepsilon}{2}\left[\gamma_2\|\nabla\eta\|_{0,K}^2+\frac{1}{\gamma_2}\|\nabla e_h\|_{0,K}^2\right]\\&\qquad+\sum_K\frac{\tau^e}{2}\gamma_3\left(\frac{1+\gamma_4}{\gamma_4}\sigma^2\|\eta\|_{0,K}^2+(1+\gamma_4)\varepsilon^2\|\Delta\eta\|_{0,K}^2\right)\\&\qquad+\sum_K\frac{\tau^e}{2}\frac{1}{\gamma_3}\left(\frac{1+\gamma_5}{\gamma_5}\sigma^2\|e_h\|_{0,K}^2+(1+\gamma_5)\varepsilon^2\|\Delta e_h\|_{0,K}^2\right)\\&\quad=\sum_K\left[\left(\frac{\gamma_1\gamma_4+\gamma_3(1+\gamma_4)\tau\sigma}{2\gamma_4}\right)\sigma\|\eta\|_{0,K}^2+\frac{\gamma_2\varepsilon}{2}\|\nabla\eta\|_{0,K}^2+\frac{\tau^e\gamma_3(1+\gamma_4)\varepsilon}{2}\varepsilon\|\Delta\eta\|_{0,K}^2\right]\\&\qquad+\sum_K\left[\left(\frac{\gamma_3\gamma_5+\gamma_1(1+\gamma_5)\tau\sigma}{2\gamma_1\gamma_3\gamma_5}\right)\sigma\|e_h\|_{0,K}^2+\frac{\varepsilon}{2\gamma_2}\|\nabla e_h\|_{0,K}^2+\frac{(1+\gamma_5)\tau^e\varepsilon}{2\gamma_3}\varepsilon\|\Delta e_h\|_{0,K}^2\right]\end{aligned}\tag{6.83}$$

where $\gamma_4,\gamma_5>0$ are arbitrary positive constants. Using the inverse estimate (6.64) on the last term of the RHS, and gathering common terms

$$\begin{aligned}
&\left|B_{VMS}(\eta, e_h)\right| \\
&\quad\le \sum_K \left\{ \left(\frac{\gamma_1\gamma_4 + \gamma_3(1+\gamma_4)\tau^e\sigma}{2\gamma_4} \right) \sigma \|\eta\|_{0,K}^2 + \frac{\gamma_2\varepsilon}{2}\|\nabla\eta\|_{0,K}^2 + \frac{\tau^e\gamma_3(1+\gamma_4)\varepsilon}{2}\varepsilon\|\Delta\eta\|_{0,K}^2 \right\} \\
&\qquad + \sum_K \left\{ \left(\frac{\gamma_3\gamma_5 + \gamma_1(1+\gamma_5)\tau^e\sigma}{2\gamma_1\gamma_3\gamma_5} \right) \sigma \|e_h\|_{0,K}^2 + \left(\frac{1}{2\gamma_2} + \frac{(1+\gamma_5)\tau^e\varepsilon}{2\gamma_3}\frac{C_{inv}}{h^2} \right) \varepsilon \|\nabla e_h\|_{0,K}^2 \right\}
\end{aligned} \tag{6.84}$$

Using inequality (6.65) and choosing

$$\gamma_1 = \gamma_2 = \beta\gamma_3 \qquad \gamma_3 = \frac{1+\beta}{\beta C_s} \qquad \gamma_4 = \gamma_5 = \frac{\alpha^2}{C_{inv}} \tag{6.85}$$

where β is as in inequality (6.65), C_s is the constant coefficient from the stability norm (Lemma 6.2), and C_{inv} is the constant for the inverse estimate (6.64), the expression in equation (6.84) can be bounded such that the following result is obtained

$$\begin{aligned}
&\left|B_{VMS}(\eta, e_h)\right| \\
&\quad\le \sum_K \left\{ \left(\frac{1+\beta}{C_s} \right) \sigma \|\eta\|_{0,K}^2 + \left(\frac{1+\beta}{2C_s} \right) \varepsilon \|\nabla\eta\|_{0,K}^2 + \left(\frac{1+\beta}{2C_s} \right) h^2 \varepsilon \|\Delta\eta\|_{0,K}^2 \right\} \\
&\qquad + \sum_K \left\{ \left(\frac{C_S(1-\beta)}{2} \right) \sigma \|e_h\|_{0,K}^2 + \left(\frac{C_s(1-\beta)}{2} \right) \varepsilon \|\nabla e_h\|_{0,K}^2 \right\}
\end{aligned} \tag{6.86}$$

Recall $0 < \beta < 1$ and (6.80)

$$\begin{aligned}
&\sum_K C_s \left[\sigma \|e_h\|_{0,K}^2 + \varepsilon \|\nabla e_h\|_{0,K}^2 \right] \\
&\qquad\qquad \le \sum_K \left(\frac{2}{C_s} \right) \left\{ \sigma \|\eta\|_{0,K}^2 + \varepsilon \|\nabla\eta\|_{0,K}^2 + h^2 \varepsilon \|\Delta\eta\|_{0,K}^2 \right\} \\
&\qquad\qquad\quad + \sum_K \left(\frac{C_s}{2} \right) \left\{ \sigma \|e_h\|_{0,K}^2 + \varepsilon \|\nabla e_h\|_{0,K}^2 \right\}
\end{aligned} \tag{6.87}$$

Moving the e_h terms from the RHS to the LHS, and dividing by the constants

$$\begin{aligned}
&\sum_K \left[\sigma \|e_h\|_{0,K}^2 + \varepsilon \|\nabla e_h\|_{0,K}^2 \right] \\
&\qquad\qquad \le \sum_K \frac{4}{C_s^2} \left\{ \sigma \|\eta\|_{0,K}^2 + \varepsilon \|\nabla\eta\|_{0,K}^2 + h^2 \varepsilon \|\Delta\eta\|_{0,K}^2 \right\}
\end{aligned} \tag{6.88}$$

Letting $C_{FE} = 4/C_s^2$ and noting it is independent of σ, ε, and h completes the proof. □

In order to develop the estimate on the total error, we make use of the following estimate on the interpolation error for sufficiently smooth exact solution u.

$$\|\eta\|_{s,K} \leq Ch^{k+1-s}\|u\|_{k+1,K} \tag{6.89}$$

where $\|\cdot\|_{s,K}$ is the H^s-norm over the domain K, and k is the order of the highest complete polynomial in the approximating space over a single element subdomain K. Thus

$$\begin{aligned} \|\eta\|_{0,K}^2 &\leq C_0 h^{2k+2}\|u\|_{k+1,K}^2 \\ \|\nabla\eta\|_{0,K}^2 &\leq \|\eta\|_{1,K}^2 \leq C_1 h^{2k}\|u\|_{k+1,K}^2 \\ \|\Delta\eta\|_{0,K}^2 &\leq \|\eta\|_{2,K}^2 \leq C_2 h^{2k-2}\|u\|_{k+1,K}^2 \end{aligned} \tag{6.90}$$

Lemma 6.4 (Convergence estimate). Under the conditions on σ, ε, and τ from Lemma 6.2, the total error e is bounded by the norm of the exact solution u as follows

$$\sigma\|e\|_{0,\Omega}^2 + \varepsilon\|\nabla e\|_{0,\Omega}^2 \leq C_e h^{2k}(\sigma h^2 + \varepsilon)\|u\|_{k+1,\Omega}^2 \tag{6.91}$$

where C_e is a positive constant independent of σ, ε, and h.

Proof. By triangle inequality and AM-GM inequality we get the following:

$$\begin{aligned} \sigma\|e\|_{0,\Omega}^2 + \varepsilon\|\nabla e\|_{0,\Omega}^2 &\leq \sigma\left[\|\eta\|_{0,\Omega}^2 + \|e_h\|_{0,\Omega}^2\right] + \varepsilon\left[\|\nabla\eta\|_{0,\Omega}^2 + \|\nabla e_h\|_{0,\Omega}^2\right] \\ &= \sum_K\left[\sigma\|\eta\|_{0,K}^2 + \varepsilon\|\nabla\eta\|_{0,K}^2\right] + \sum_K\left[\sigma\|e_h\|_{0,K}^2 + \varepsilon\|\nabla e_h\|_{0,K}^2\right] \end{aligned} \tag{6.92}$$

Using Lemma 6.2

$$\begin{aligned} \sigma\|e\|_{0,\Omega}^2 + \varepsilon\|\nabla e\|_{0,\Omega}^2 &\leq \sum_K\left[\sigma\|\eta\|_{0,K}^2 + \varepsilon\|\nabla\eta\|_{0,K}^2\right] \\ &\quad + \sum_K C_{FE}\left[\sigma\|\eta\|_{0,K}^2 + \varepsilon\|\nabla\eta\|_{0,K}^2 + h^2\varepsilon\|\Delta\eta\|_{0,K}^2\right] \\ &\leq \sum_K \hat{C}\left[\sigma\|\eta\|_{0,K}^2 + \varepsilon\|\nabla\eta\|_{0,K}^2 + h^2\varepsilon\|\Delta\eta\|_{0,K}^2\right] \end{aligned} \tag{6.93}$$

where $\hat{C} = C_{FE} + 1$. Using interpolation error estimates from (6.90)

$$
\begin{aligned}
&\sigma \|e\|_{0,\Omega}^2 + \varepsilon \|\nabla e\|_{0,\Omega}^2 \\
&\quad \le \sum_K \hat{C}\left[\sigma C_0 h^{2k+2} \|u\|_{k+1,K}^2 + \varepsilon C_1 h^{2k} \|u\|_{k+1,K}^2 + h^2 \varepsilon C_2 h^{2k-2} \|u\|_{k+1,K}^2 \right] \\
&\quad \le \sum_K \hat{C} \max(C_0\, ,\, C_1 + C_2) h^{2k} \|u\|_{k+1,K}^2 \left(\sigma h^2 + \varepsilon\right) \\
&\quad \le \sum_K C_e h^{2k} (\sigma h^2 + \varepsilon) \|u\|_{k+1,K}^2 \\
&\quad = C_e h^{2k} (\sigma h^2 + \varepsilon) \|u\|_{k+1,\Omega}^2
\end{aligned}
\tag{6.94}
$$

where $C_e = \hat{C}\max(C_0, C_1 + C_2)$ is independent of σ, ε, and h. □

Remark: This result means that the convergence rate of the unknown field in the $H^1(\Omega)$-norm in the diffusion-dominated case is 'k', and therefore it would be 'k+1' in the $L_2(\Omega)$-norm.

Remark: For the reaction-dominated case the rate of convergence in the $L_2(\Omega)$-norm is 'k+1'. The numerical tests presented in Figure 6.24 show that we attain this rate asymptotically as the mesh is refined.

6.7 Built-in *A Posteriori* Error Estimate

A feature of VMS methods is that they come equipped with an error estimation framework, as discussed in [45]. To see this, consider the multiscale decomposition of the exact solution in equation (6.17), where the fine scale field u' can be interpreted as the error e between the exact solution u and the coarse scale field $\bar{u}$.

$$
e = u - \bar{u} = u' \tag{6.95}
$$

In the derivation of VMS methods, the weak form is decomposed into a coarse scale subproblem and a fine scale subproblem. Solution of the coarse scale subproblem to find the coarse scale field depends on the fine scales, therefore $\bar{u} = \hat{u}\left(u'\right)$. Invariably, in order to develop a numerical method, it is required to make approximations regarding the fine scales. In the present method, the fine scales are approximated locally within each element K as

$$
u'\big|_K \approx \tilde{u}'\big|_K = \tau^e\, \bar{R}\big|_K \tag{6.96}
$$

where $\bar{R} = \mathcal{L}\bar{u} - f = -\varepsilon\Delta\bar{u} + \sigma\bar{u} - f$ is the residual of the coarse scale Euler-Lagrange equations, and τ^e is given by (6.42).

The numerical solution u^h to (6.12) represents an approximation to the exact coarse scales, i.e. $\bar{u} \approx u^h$, because of two factors: (i) in deriving the modified coarse scale problem, some boundary terms were neglected, and (ii) the fine scale model is an approximation of the exact fine scales. Therefore, the total error in the numerical solution $e^h = u - u^h$ can be split into a local error e_L associated with the error between the exact and approximated fine scales, and a global pollution error e_G. Moreover, the local error can be further split into an explicit term (given by the modeled fine scales) and an implicit correction. Therefore, the total error estimate is given as

$$e^h = \widehat{e}_L + \tilde{e}_L + \tilde{e}_G \tag{6.97}$$

where the explicit error corresponds to the computed fine scales.

$$\widehat{e}_L\big|_K = \tilde{u}'\big|_K = \tau^e \, \bar{R}\big|_K \tag{6.98}$$

For details on the computation of the implicit parts of the a posteriori error estimation ($\tilde{e}_L$ and $\tilde{e}_G$), the interested reader is referred to [45]. Numerical results for the spatial distribution of the local explicit error are shown in section 6.9.1.

6.8 Selective VMSm

Depending on the definition of the fine-scale basis B_K, one arrives at different values of τ^e through equations (6.42)-(6.47). It is also noted that computation of τ^e requires the solution of a $n_b \times n_b$ linear system. In this section, and for 4-node quadrilateral elements (Q4), we denote as τ_m^e the stabilization parameter arising from the fine-scale basis given in Table 6.1, i.e., $B_{K,m} = \{b_1^e, b_2^e, b_3^e, b_4^e, b_5^e\}$. On the other hand, we use τ_s^e to denote the stabilization parameter associated with a fine-scale basis consisting of only the standard quadratic bubble, i.e., $B_{K,s} = \{b_5^e\}$. We further define a Selective VMSm method for problems with heterogeneous coefficients. In this method, the stabilization parameter τ^e within a given element K is determined according to the value of the Damkohler number in that element α_K as follows

$$\tau^e = \begin{cases} \tau_m^e, & \alpha_K \geq \alpha_{cr} \\ \tau_s^e. & \alpha_K < \alpha_{cr} \end{cases} \tag{6.99}$$

The purpose of the definition in (6.99) is to reserve the use of the more flexible (but more computationally expensive) fine scale basis in regions of the domain where steep gradients are expected (i.e., reaction dominated regions with large Damkohler number), while using a single standard bubble in the diffusion dominated parts of the domain, for computational cost savings.

6.9 Numerical Results

In this section, we apply the VMS method with multiple fine-scale basis functions (VMSm) given by equation (6.12), with the fine-scale basis in Table 6.1 and Table 6.2 for 4-node quadrilaterals and 3-node triangles, respectively, in order to solve the boundary value problem posed by (6.1). The domain is defined as the unit square $\Omega = [0,1]\times[0,1]$. We explore different domain discretizations, forcing function f, and diffusion coefficient, ε. The reaction coefficient is set as $\sigma = 1$. In the VMS methods (VMSm, VMSroot, and VMS quadratic) used to generate the results in this section, the full, spatially-varying stabilization parameter $\tau^e(\mathrm{x})$ from equation (6.42) was implemented.

6.9.1 Structured quadrilaterals and triangles

We discretize the domain by structured meshes of 4-node (Q4) and 3-node (T3) elements. The meshes are shown in Figure 6.6 (*h* = 1/20). The forcing function is set to $f = 1$ and the diffusivity is $\varepsilon = 1e-6$, with corresponding $\alpha = 50$. We compare results obtained from the Galerkin FEM method given by equation (6.8), the VMS method with standard bubble, and VMS with multiple fine-scale basis functions (VMSm), and other methods (see Figure 6.7, Figure 6.8, Figure 6.9, and Figure 6.10).

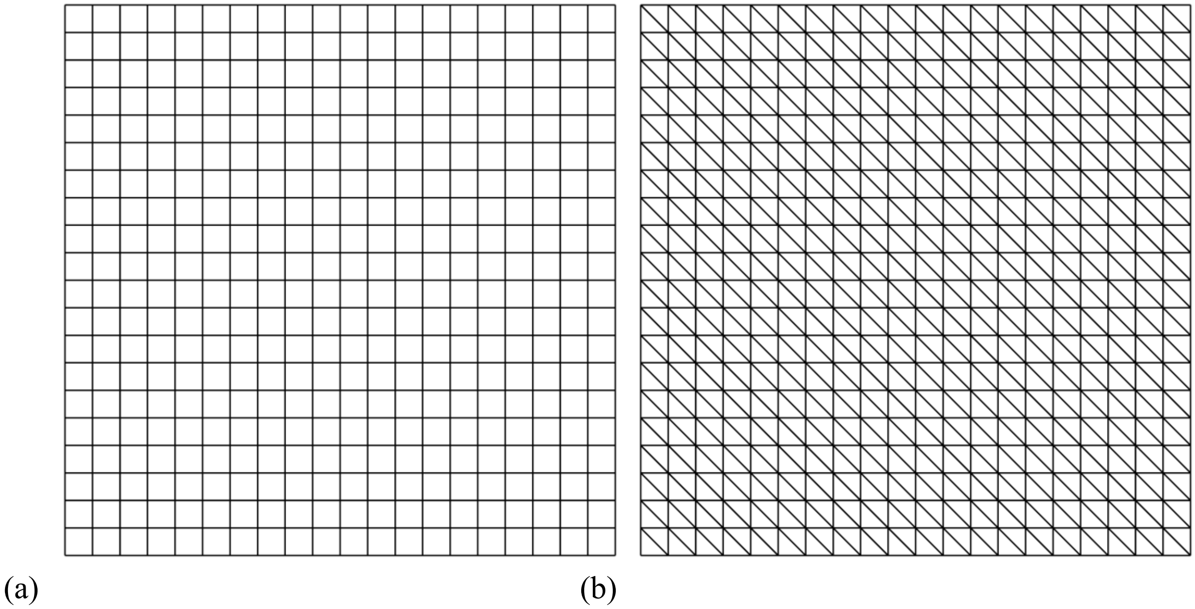


Figure 6.6. (a) Structured mesh of 20×20 4-node quadrilaterals (Q4 elements). (b) Structured mesh of 800 (‘20×20’) 3-node triangles (T3 elements).

The Galerkin method and the VMS method with the standard bubble show oscillations engendered by a boundary layer around the domain. The Galerkin solution peaks at 1.605 (Q4 and T3), while the VMS solution peaks at 1.261. On the other hand, the VMSm solution eliminates the oscillatory behavior and the peak value does not go over 1.0 for 4-node elements, while it peaks at 1.024 for 3-node elements. Figure 6.9 and Figure 6.10 show cross-sections of the solution for quadrilaterals and triangles, respectively. In addition, since the problem parameters are constant over the domain, and for the Q4 element case all elements are identical, the stabilization parameter τ is exactly the same in every element, and it is of the shape shown in Figure 6.11.

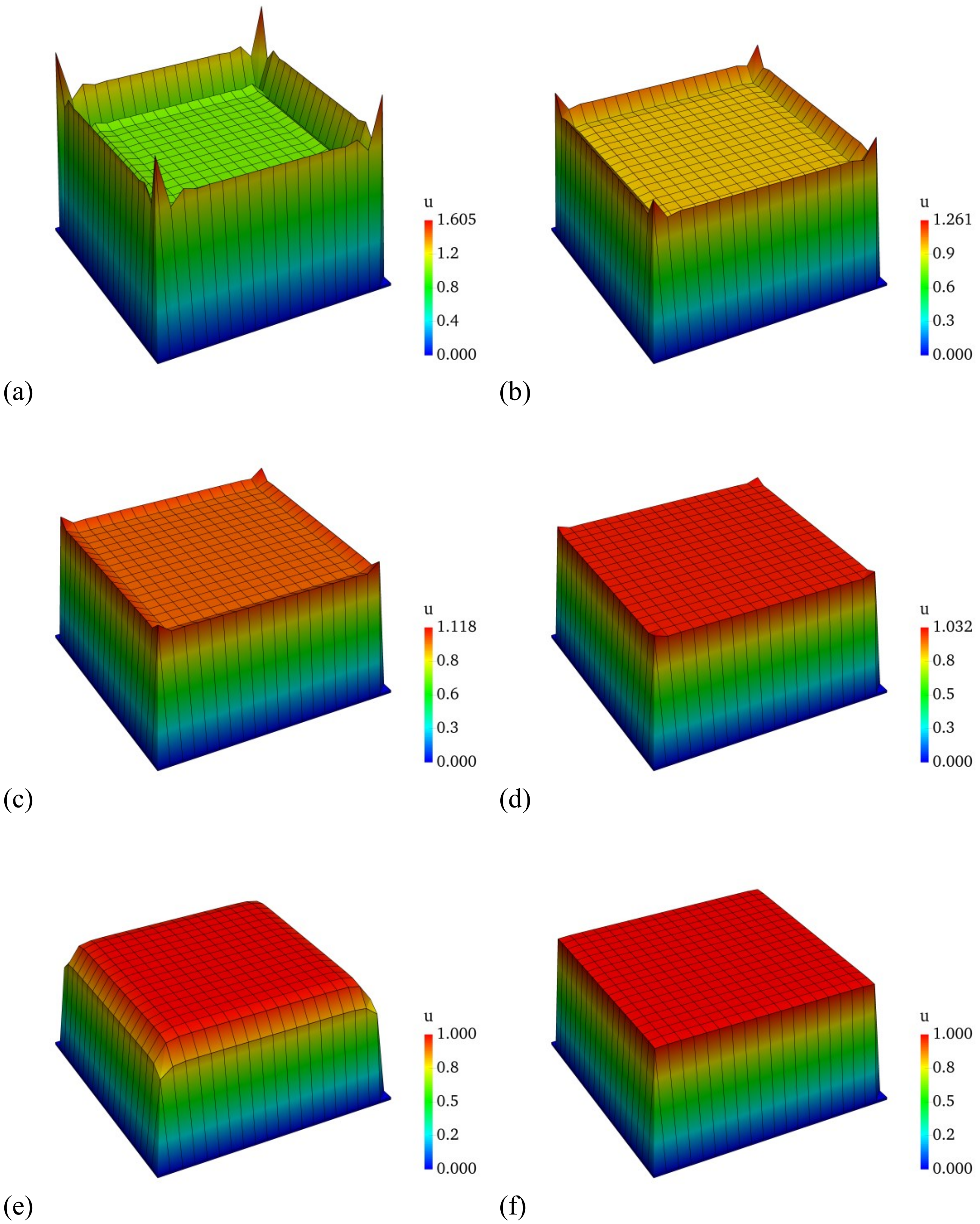

u
1.605
1.2
0.8
0.4
0.000
(a)
u
1.261
0.9
0.6
0.3
0.000
(b)
u
1.118
0.8
0.6
0.3
0.000
(c)
u
1.032
0.8
0.5
0.3
0.000
(d)
u
1.000
0.8
0.5
0.2
0.000
(e)
u
1.000
0.8
0.5
0.2
0.000
(f)

Figure 6.7. Solutions for Q4 mesh with different methods: (a) Galerkin, (b) VMS quadratic bubble, (c) GLS Harari's tau, (d) GGLS, (e) VMSroot, and (f) VMSm.

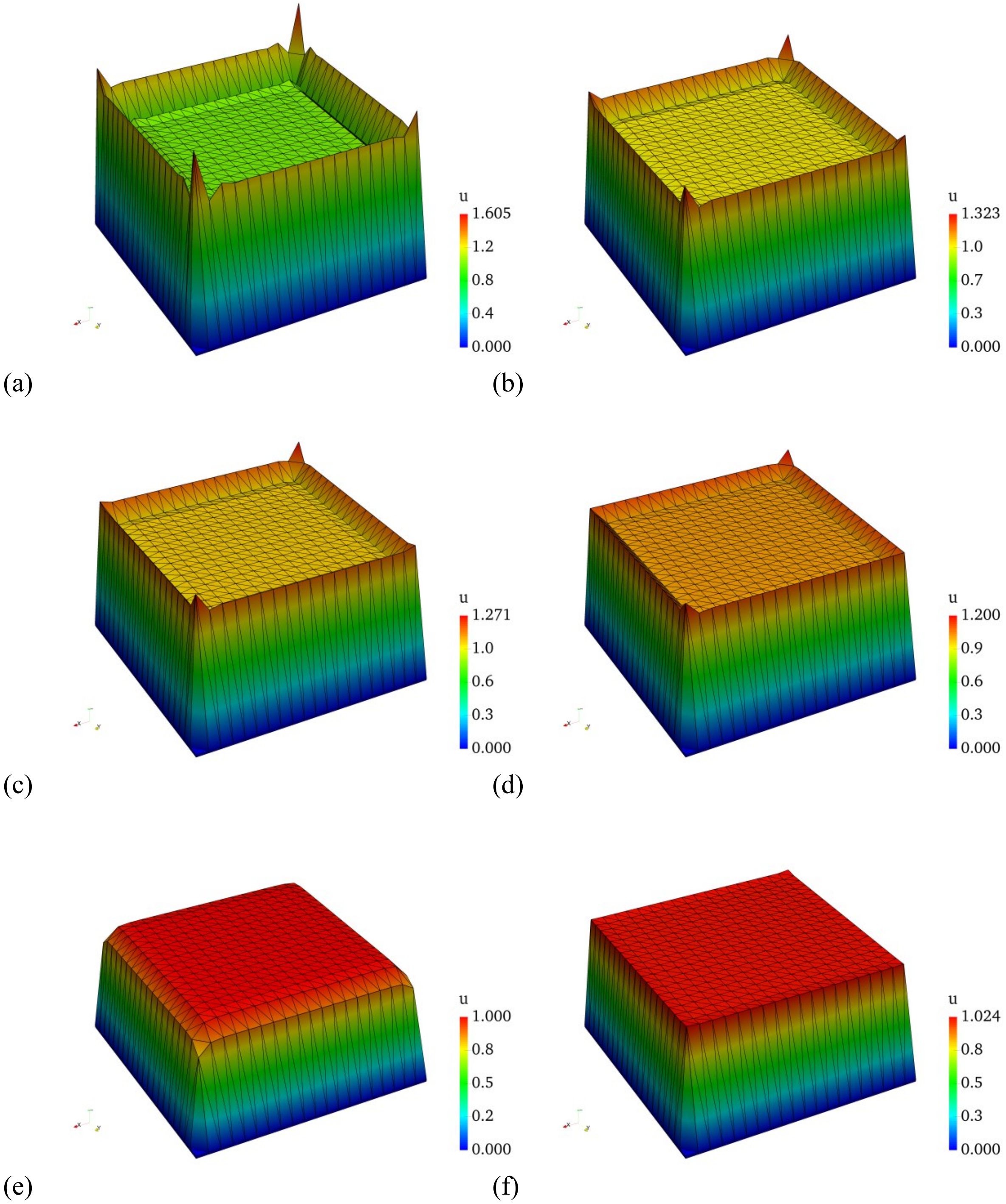
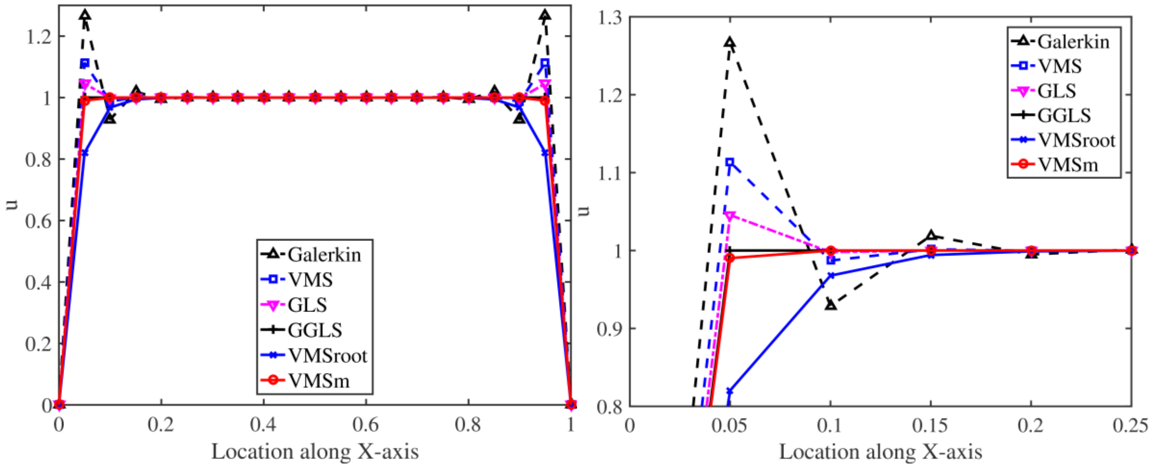
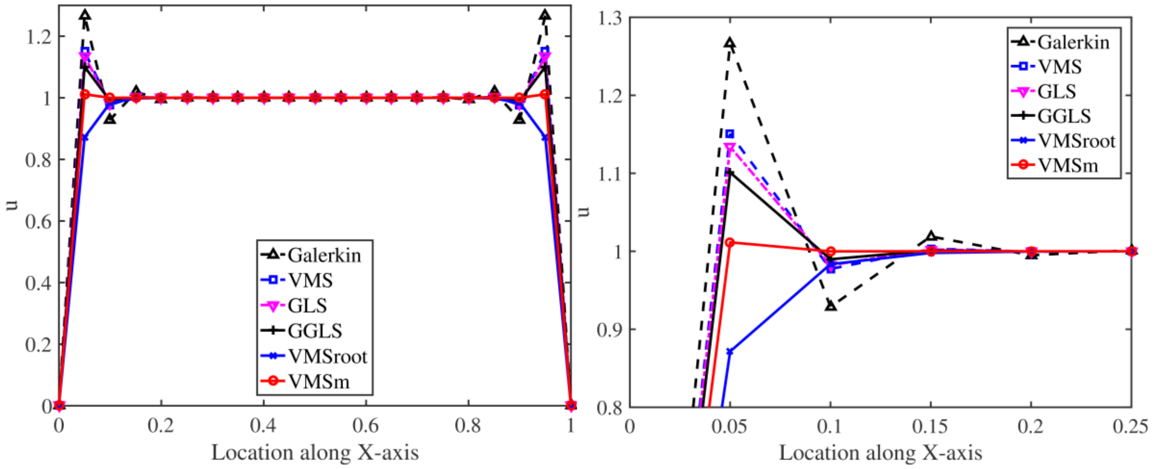

u
1.605
1.2
0.8
0.4
0.000
(a)
u
1.323
1.0
0.7
0.3
0.000
(b)
u
1.271
1.0
0.6
0.3
0.000
(c)
u
1.200
0.9
0.6
0.3
0.000
(d)
u
1.000
0.8
0.5
0.2
0.000
(e)
u
1.024
0.8
0.5
0.3
0.000
(f)

Figure 6.8. Solutions for T3 mesh with different methods: (a) Galerkin, (b) VMS quadratic bubble, (c) GLS Harari's tau, (d) GGLS, (e) VMSroot, and (f) VMSm.



(a) (b)

Figure 6.9. (a) Comparison of the solution along the line y = 0.5 for different methods (Q4 elements), and (b) close-up of oscillations.



(a) (b)

Figure 6.10. (a) Comparison of the solution along the line y = 0.5 for different methods (T3 Elements), and (b) close-up of oscillations.

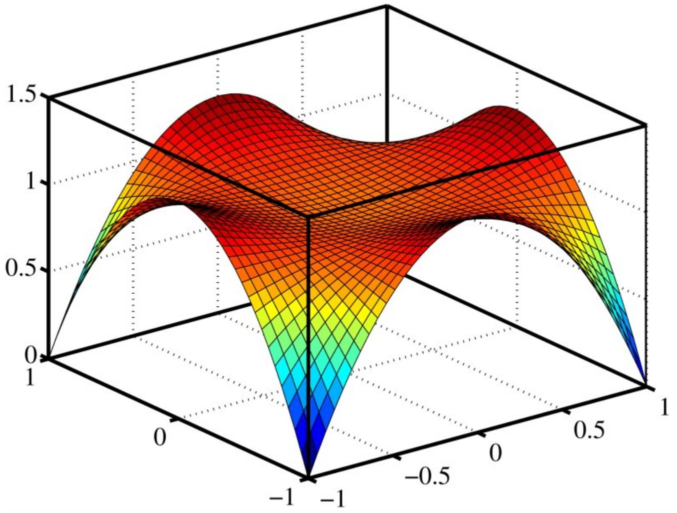


Figure 6.11. Shape of the stabilization parameter τ^e for high Damkohler number across both perpendicular length scales in the element (element natural coordinates).

As discussed in section 6.7, the fine-scale field itself represents an error estimate. In particular, the numerical fine scales (expanded in terms of the fine scale basis) represent the so-called local explicit error $\widehat{e}_L$ in the notation of [45]. Therefore, Figure 6.12 shows the distribution of the compute fine scales (or the local explicit error $\widehat{e}_L$) in the domain. It is noted that this error estimate coincides with the notion that the largest errors occur in the boundary layer region. Furthermore, one may follow the framework provided in [45] to compute additional more accurate implicit estimates of local and global error, as well as compute the norms or seminorms (in $H^1(\Omega)$ or $L^2(\Omega)$) of these errors to verify their convergence. Nevertheless, we remark that the explicit local error already provides a reasonable estimate for the distribution of error in the domain.

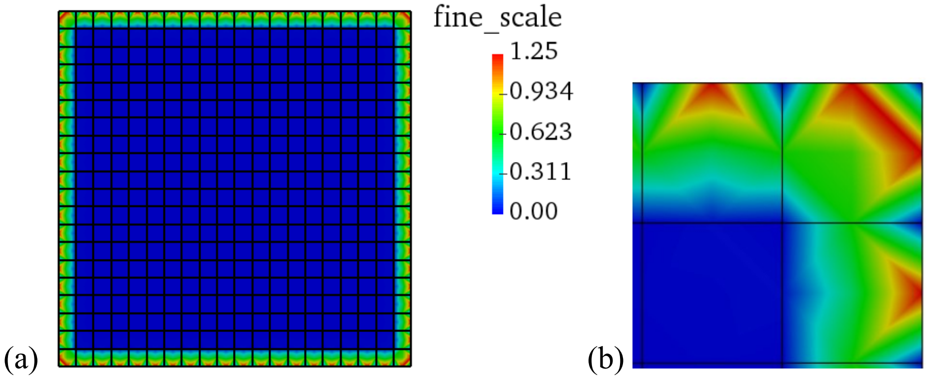


Figure 6.12. Distribution of the local explicit error (or numerical fine scale field) in (a) the entire domain and (b) close-up of the top-right corner of the domain.

6.9.2 Anisotropic *h*-refinement

Considering now the parameters from section 6.9.1, we change the diffusivity to $\varepsilon = 2e-4$, and use a mesh with anisotropic refinement of 10×40 elements (Figure 6.13). This generates a regular mesh of rectangular elements with aspect ratio of 1:4. The Damkohler number is now different in the two perpendicular directions, with α_y = 1.77 < $\sqrt{6}$ in the direction discretized with 40 elements (*h* = 1/40), and α_x = 7.07 > $\sqrt{6}$ along the direction discretized with 10 elements (*h* = 1/10). We compare results from the Galerkin method and several stabilized methods as in section 6.9.1. Figure 6.14 shows a 3D surface plot of the solution, Figure 6.15 and Figure 6.16 show solutions cross-sections along the lines *y* = 0.5 and *x* = 0.5, respectively.

Due to the anisotropic refinement, even the Galerkin method is able to capture the boundary layers along the y-axis due to the higher element refinement along that direction. However, oscillations along the other direction (x-axis) persist (Figure 6.15). On the other hand, the VMSm solution captures the boundary layer jump over a single element across the coarser element length (x-axis) without any overshoots or undershoots, but the boundary layer in the other direction is slightly diffused (Figure 6.16). One explanation for the over-diffusivity in this specific situation is the missing boundary terms discussed previously. In the region and direction where the mesh is finer, the Damkohler number is closer to its critical value $\alpha_{cr} = \sqrt{6}$, and the diffusive and reactive effects are comparable in magnitude. Thus, the assumption of negligible boundary terms at the element boundaries may have been violated. Nonetheless, reasonably accurate and stable solutions are still obtained with the present method.

This test case also highlights the advantages of using the full, spatially-varying stabilization parameter τ^e induced by the proposed fine-scale basis. In the numerical results for square Q4 elements from section 6.9.1, some methods whose stabilization parameter is a constant within the element, such as GGLS or the one denoted as GLS Harari, and VMS methods with a single bubble, such as VMSroot, are able to suppress oscillations to a considerable degree (Figure 6.7c, 7d, and 7e). However, those same methods fail to suppress the oscillations for the test case in this section (Figure 6.14c, 13d, 13e, and Figure 6.15), where $\alpha_x \neq \alpha_y$, i.e., the Damkohler numbers along the x-direction is different from the one along the y-direction. On the other hand, the VMSm method, with multiple fine-scale basis functions and spatially-varying stabilization parameter, is able to completely suppress oscillations for both test cases: square Q4 elements in section 6.9.1 (Figure 6.7f), and for the test case with anisotropic features in this section (Figure 6.14f). Moreover, this is despite the methods' similarities in the asymptotic behavior of the stabilization parameter's average value $\overline{\tau}^{e}$ from Figure 6.3 and Figure 6.4, where the value for "GLS Harari" asymptotes to 1.0 in the reaction-dominated regime, while it asymptotes to 0.97 approximately for both VMSroot and VMSm. Therefore, robust performance gained for the anisotropic case is associated with the proposed fine-scale basis with multiple functions and the corresponding spatially varying stabilization parameter, because it has the flexibility to deal with the anisotropy of this test case by capturing the directionality of the problem. This is also manifested in the shape of the stabilization parameter τ^{e}. For the (isotropic) problem with square Q4 elements of section 6.9.1, the shape of the VMSm stabilization parameter in Figure 6.11 is symmetric under 90º rotations. However, that is not the case for the anisotropic problem in this section. Instead, the shape of the stabilization parameter automatically adapts and changes to the one shown in Figure 6.17, which has a higher variation along the axis aligned with the x-direction (along which the mesh is coarser and requires more stabilization), than along the axis aligned with the y-direction (which is the direction of higher mesh refinement).

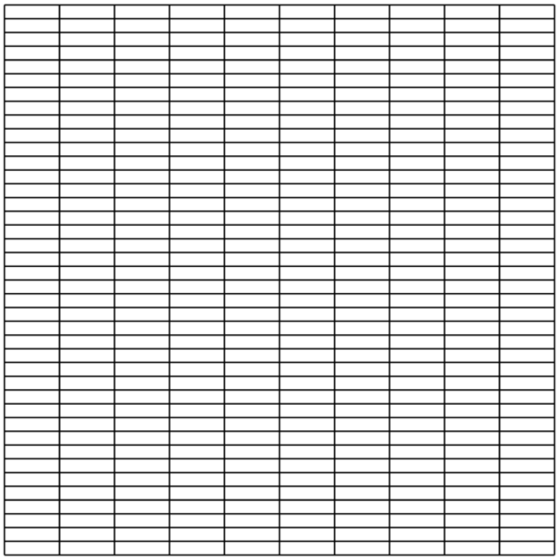


Figure 6.13. Structured mesh of 10×40 4-node rectangles (Q4 elements).

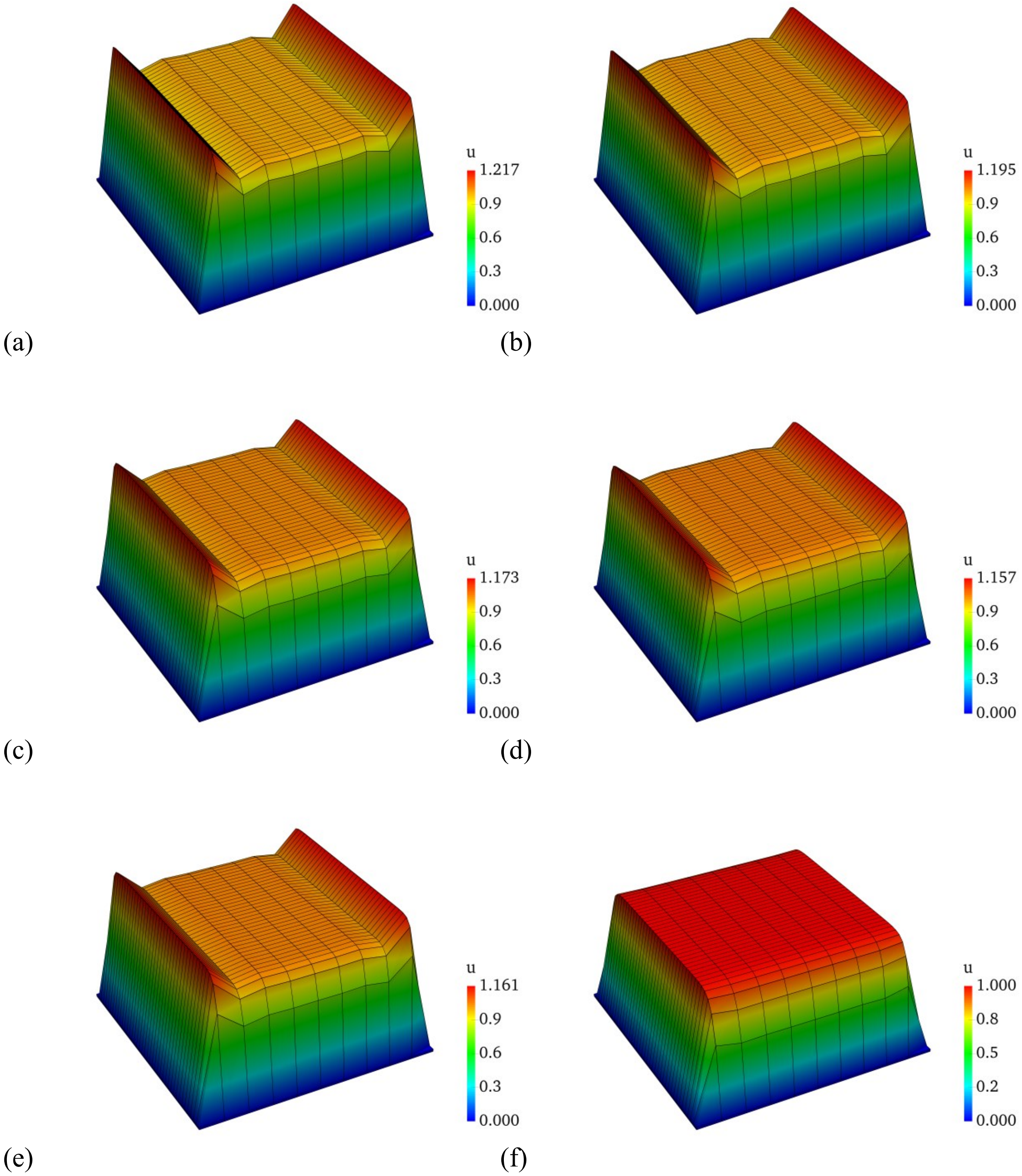
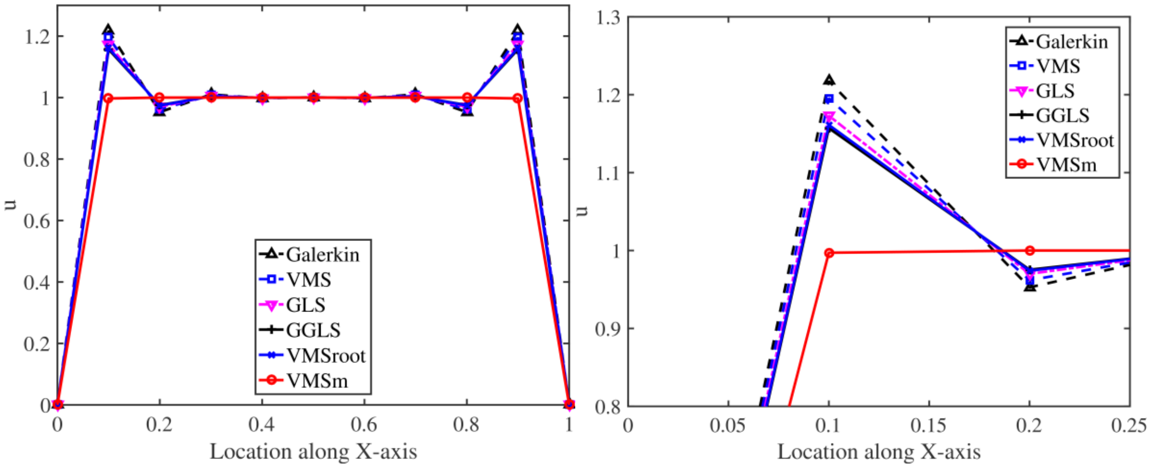
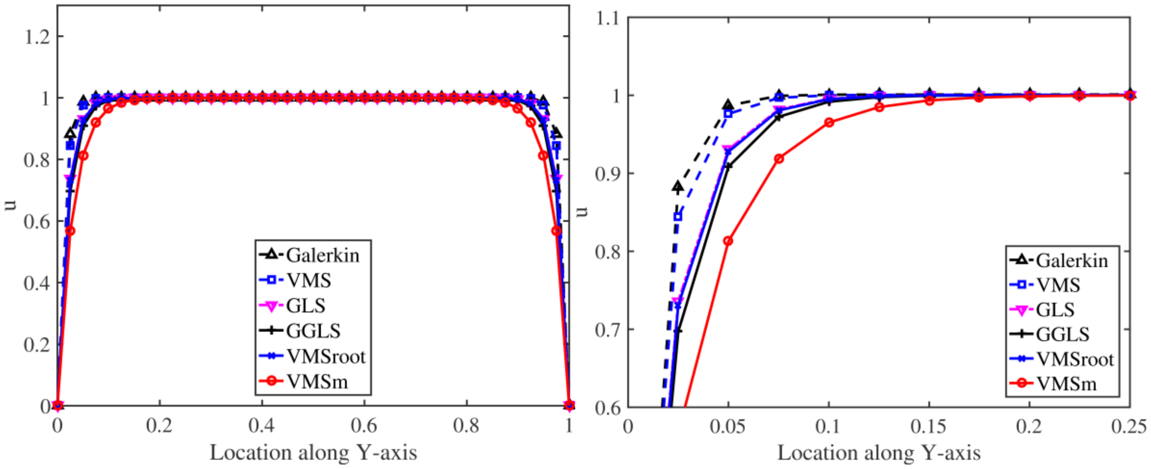


Figure 6.14. Solution for anisotropic Q4 mesh with different methods: (a) Galerkin, (b) VMS quadratic bubble, (c) GLS Harari's tau, (d) GGLS, (e) VMSroot, and (f) VMSm.



(a) (b)

Figure 6.15. (a) Comparison of the solution along the line y = 0.5 for different methods, and (b) close-up of oscillations.



(a) (b)

Figure 6.16. (a) Comparison of the solution along the line x = 0.5 for different methods (Q4 elements), and (b) close-up of oscillations.

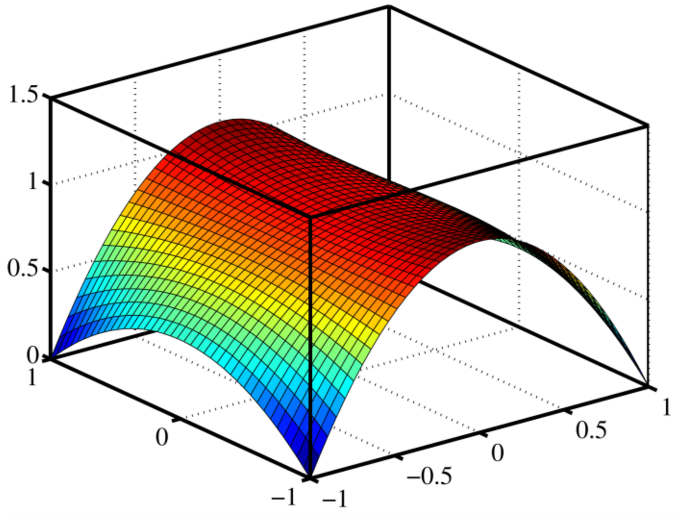
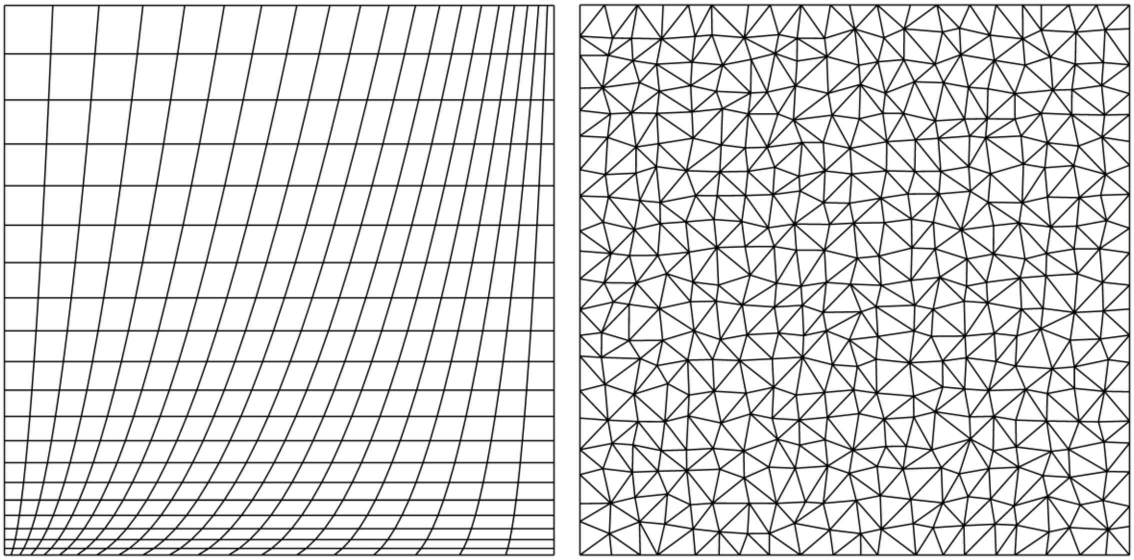


Figure 6.17. Shape of the stabilization parameter τ^e over a typical element (element natural coordinates) for problem in section 6.9.2.

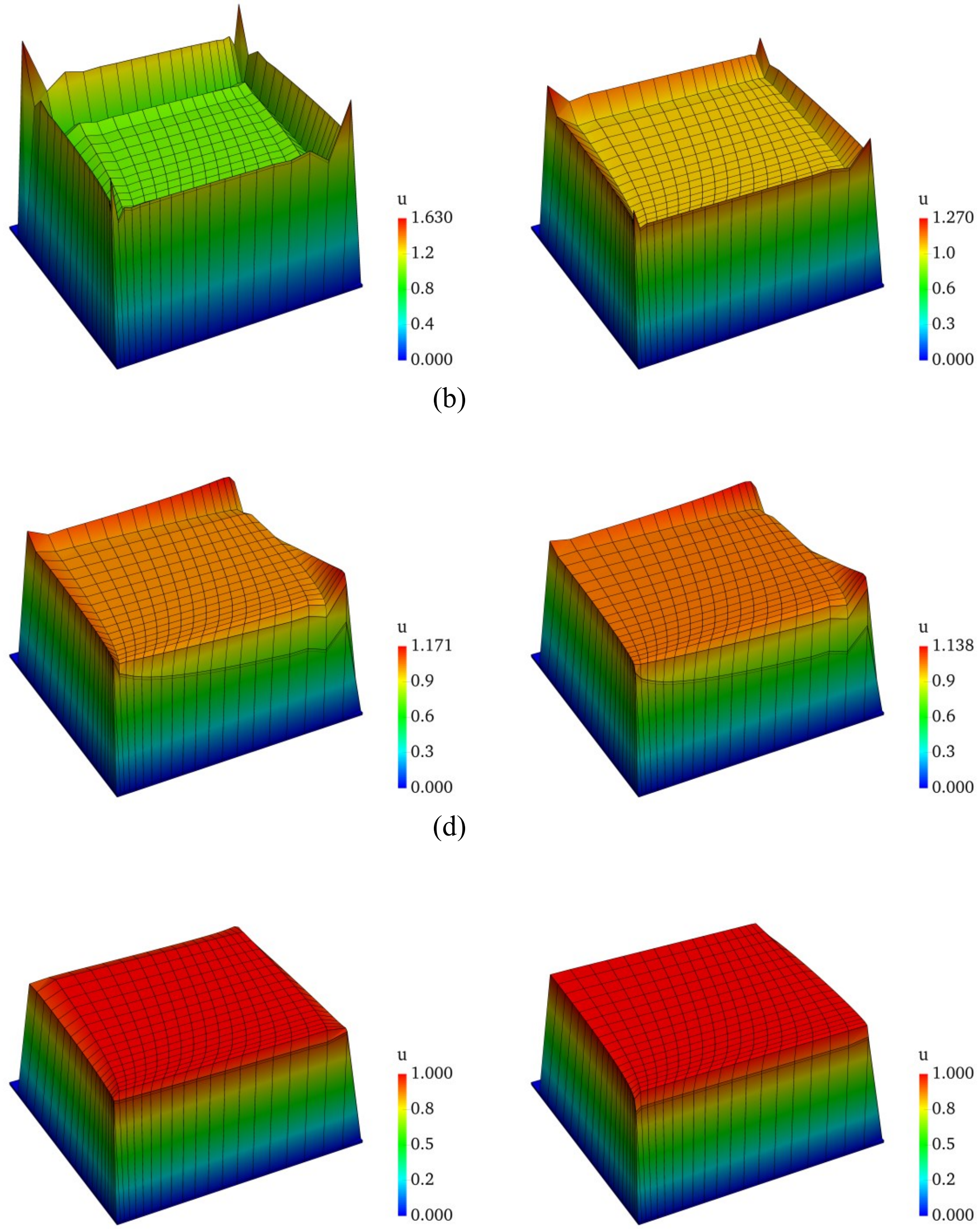
6.9.3 Distorted quadrilaterals and unstructured triangles

We consider the problem and parameters in section 6.9.1 again, only changing the mesh to the distorted ones shown in Figure 6.18. Note that near the corners of the mesh comprised of quadrilaterals we observe 4 situations: fine discretization, coarse discretization, and slender elements oriented in perpendicular directions. The second case is an unstructured triangular mesh. These two cases are representative of the type of discretizations that can be encountered in practical applications, and thus it is worthwhile to explore the behavior of the method in these cases. Note that since the element size and shape vary, also does the characteristic length and the Damkohler number associated with that element. Numerical results are shown in Figure 6.19 and Figure 6.20.



(a) (b)

Figure 6.18. (a) Mesh of distorted 20×20 4-node elements (Q4). (b) Unstructured mesh of 400 3-node triangles (T3).


u
1.630
1.2
0.8
0.4
0.000
u
1.270
1.0
0.6
0.3
0.000
u
1.171
0.9
0.6
0.3
0.000
u
1.138
0.9
0.6
0.3
0.000
u
1.000
0.8
0.5
0.2
0.000
u
1.000
0.8
0.5
0.2
0.000

(a) (b)

(c) (d)

(e) (f)

Figure 6.19. Solutions for the distorted Q4 mesh with different methods: (a) Galerkin, (b) VMS quadratic bubble, (c) GLS Harari's tau, (d) GGLS, (e) VMSroot, and (f) VMSm.

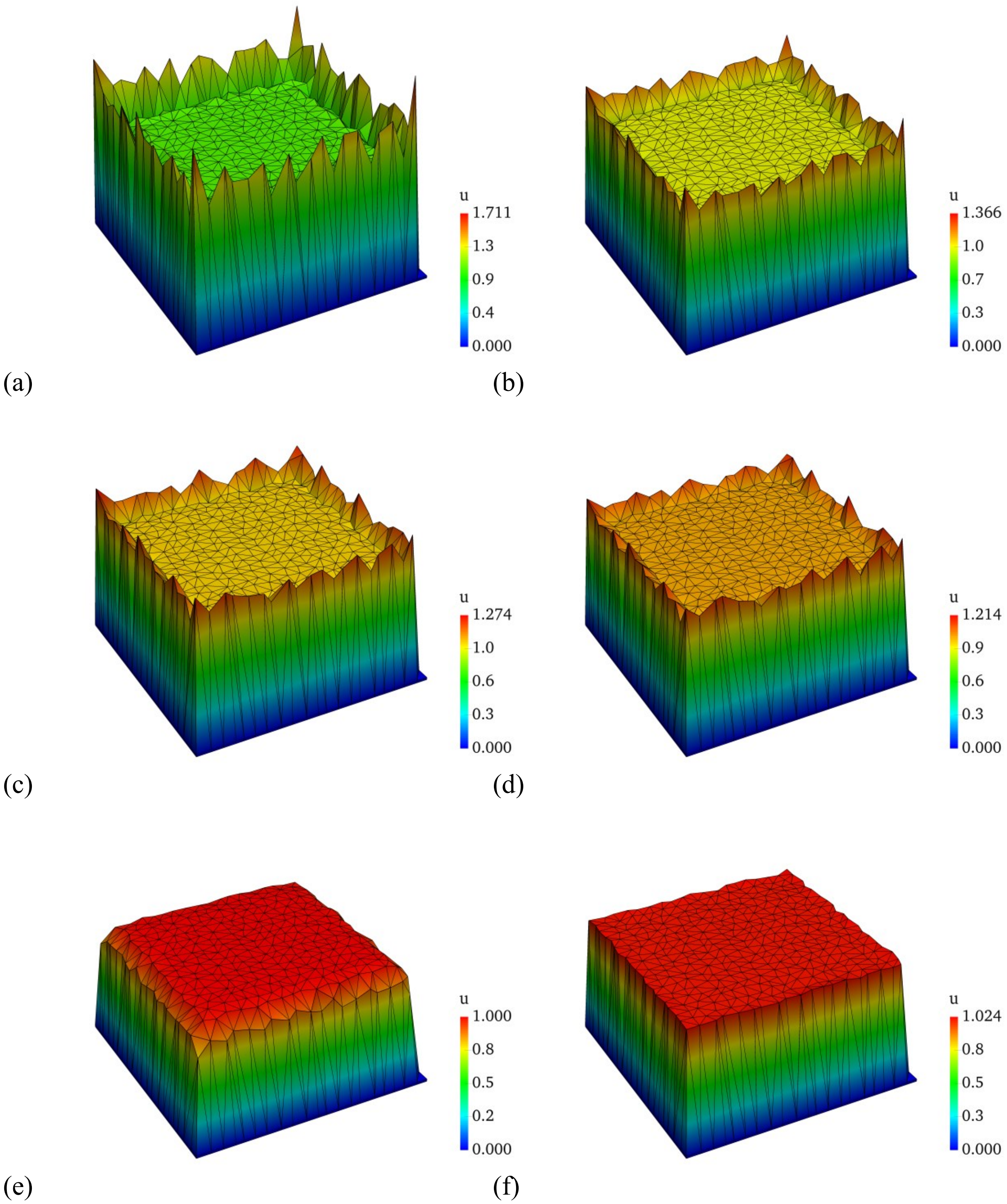

u
1.711
1.3
0.9
0.4
0.000
(a)
u
1.366
1.0
0.7
0.3
0.000
(b)
u
1.274
1.0
0.6
0.3
0.000
(c)
u
1.214
0.9
0.6
0.3
0.000
(d)
u
1.000
0.8
0.5
0.2
0.000
(e)
u
1.024
0.8
0.5
0.3
0.000
(f)

Figure 6.20. Solutions for the T3 unstructured mesh: (a) Galerkin, (b) VMS quadratic bubble, (c) GLS Harari's tau, (d) GGLS, (e) VMSroot, and (f) VMSm.

6.9.4 Discontinuous forcing function and internal layers

For this test case, we use the structured mesh of 4-node elements shown in Figure 6.6, and $\varepsilon = 1\text{e-}6$ (α = 50). The forcing function is now given by the discontinuous function in (6.100) and it generates an internal layer.

$$f = \begin{cases} 1, & \text{if } y < y_c \\ 0, & \text{otherwise} \end{cases} \tag{6.100}$$

We found the numerical solution for y_c = 0.50, 0.54, and 0.55. The first and last values coincide with element interfaces, while 0.54 falls within an element. A 3D surface plot of the numerical solution with the discontinuity at y = 0.50 is presented in Figure 6.21. It shows that the boundary layers are still captured within a single element, while the internal layer spans over two elements. Figure 6.22 shows a cross-section of the solution for the three cases. For the cases where the discontinuity coincides with the element interface, we see that the internal layer is captured within one element on either side of the discontinuity. It is also notable that the solutions in which the discontinuity is at 0.50 and 0.54 coincide, and that the internal layer does not shift, until the discontinuity falls on the next inter-element boundary.

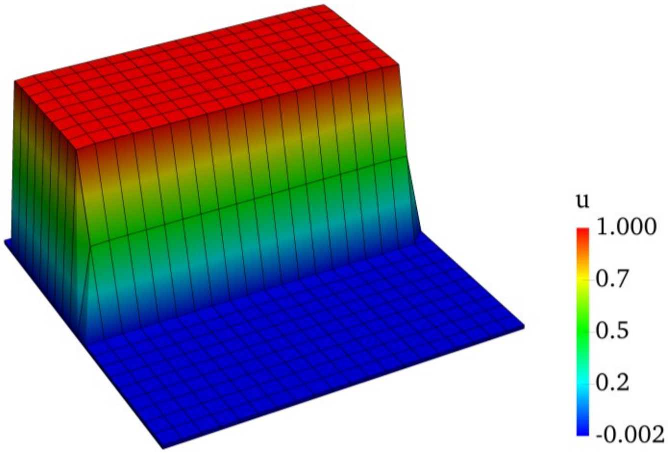


Figure 6.21. Solution for a discontinuous forcing term where the location of the discontinuity is at $y_c = 0.50$.

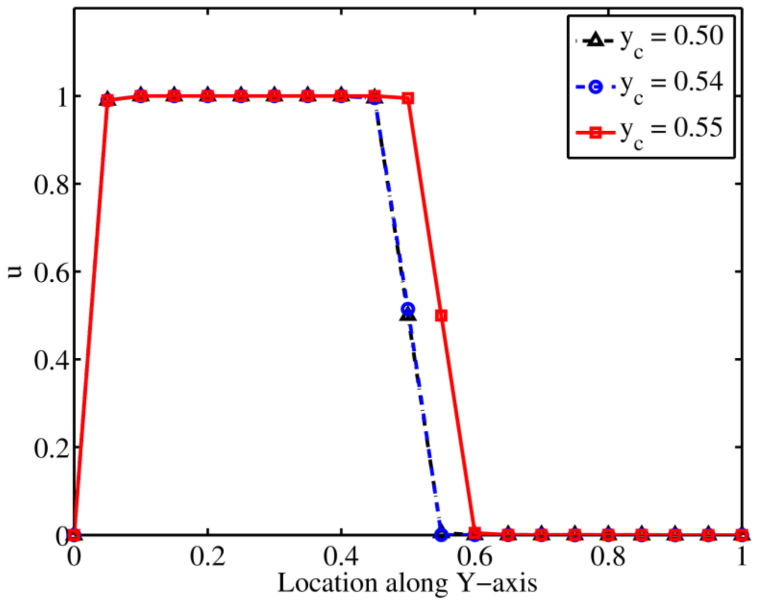


Figure 6.22. Numerical solutions at $x = 0.5$ for forcing terms with discontinuities at 0.50, 0.54, and 0.55. Note the solutions for 0.50 and 0.54 overlap.

We also consider the case where the discontinuity in the body force is not aligned with the mesh. In this case, the body force is defined as follows

$$f = \begin{cases} 1, & \text{if } y < -0.5x + 0.75 \\ 0, & \text{otherwise} \end{cases} \tag{6.101}$$

Comparison of the results with different fine-scale models is shown in Figure 6.23. The method with a single quadratic bubble for the fine scales is not able to suppress the oscillations either near the boundary layers or across the internal layer. The VMSroot method suppresses the oscillations with a slight over-diffusion at the boundary and internal layers. This issue is successfully overcome by the VMSm method, which can sharply capture the steep gradients as shown in Figure 6.23.Once more, this is evidence that the spatial variation of the stabilization parameter τ^e plays an important role, since the average value of τ^e is approximately the same for both VMSroot and VMSm at $\alpha = 50$ (Figure 6.4). Moreover, the flexibility gained by employing multiple fine-scale basis functions with non-zero trace allows the method to reproduce steep gradients sharply.

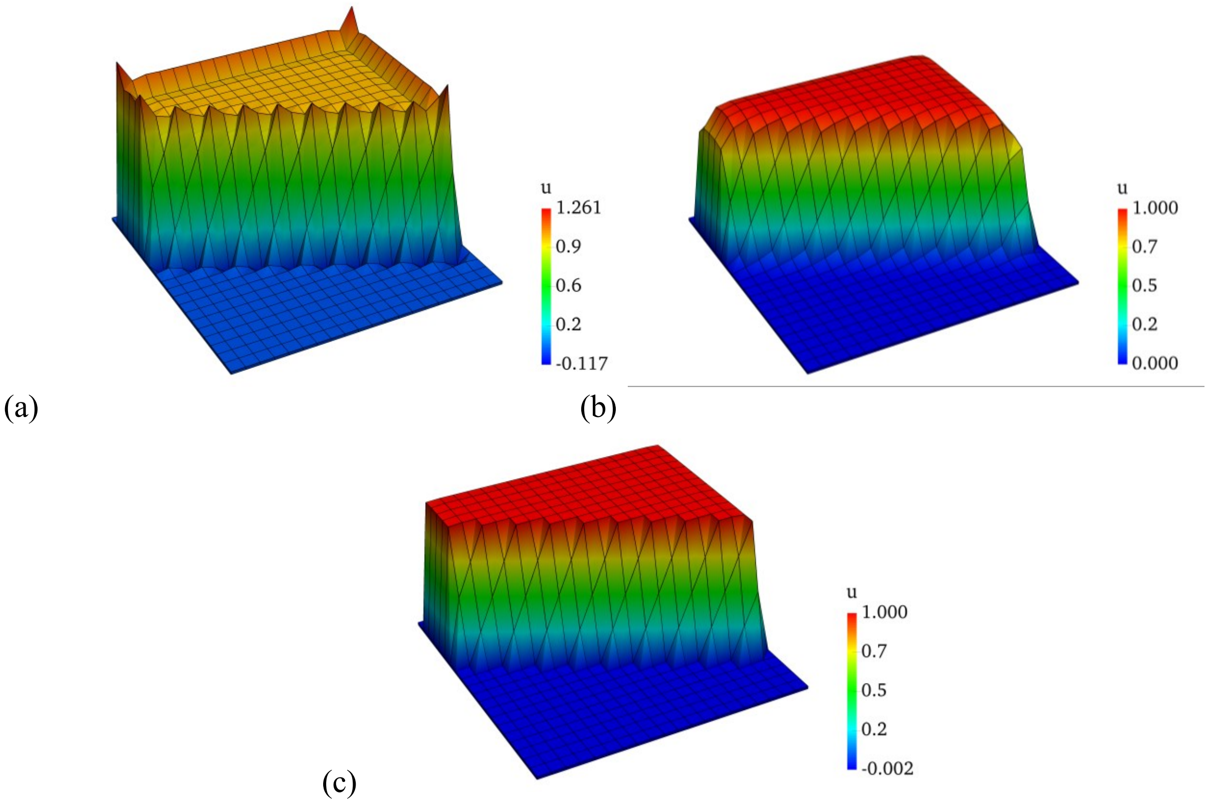


Figure 6.23. Solution for non-aligned discontinuous body force (a) VMS quadratic bubble, (b) VMSroot, and (c) VMSm.

6.9.5 Convergence rate study

To perform a convergence rate study we apply the forcing function given by (6.102), which corresponds to the exact solution in (6.103) over the bi-unit square with homogeneous Dirichlet boundary conditions.

$$f = \sigma\left[1-\cosh^{-2}\left(\sqrt{\frac{\sigma}{4\varepsilon}}\right)\cosh\left(\sqrt{\frac{\sigma}{\varepsilon}}\left[x-\frac{1}{2}\right]\right)\cosh\left(\sqrt{\frac{\sigma}{\varepsilon}}\left[y-\frac{1}{2}\right]\right)\right] \tag{6.102}$$

$$u = \left[1-\frac{\cosh\left(\sqrt{\frac{\sigma}{\varepsilon}}\left[x-\frac{1}{2}\right]\right)}{\cosh\left(\sqrt{\frac{\sigma}{4\varepsilon}}\right)}\right]\left[1-\frac{\cosh\left(\sqrt{\frac{\sigma}{\varepsilon}}\left[y-\frac{1}{2}\right]\right)}{\cosh\left(\sqrt{\frac{\sigma}{4\varepsilon}}\right)}\right] \tag{6.103}$$

The exact solution exhibits boundary layers when the diffusivity is small compared to the reaction coefficient. We perform the convergence rate studies for *h*-refinement in structured meshes with $\sigma = 1$ for 2 cases: (i) a smooth problem ($\varepsilon = 1$), and (ii) a problem with boundary layers ($\varepsilon = 1\text{e}-4$). For the 4-node elements, we use uniform meshes of 10x10, 20x20 (Figure 6.6), 40x40, 80x80, and 160x160 elements. Meanwhile, for the 3-node triangular elements, we consider uniform tessellations like the one shown in Figure 6.6, of 200, 800, 3200, 12800, and 51200 elements. For this study, where the meshes used are uniform and structured, the element length scale to be employed in the convergence plots is taken as the length of element side aligned with the domain boundary. For both Q4 and T3 elements, this corresponds to h = 1/10, 1/20, 1/40, 1/80, and 1/160.

For each mesh, we compute the error of the numerical solution u^h with respect to the exact solution u in the L_2 norm as well as in the H^1 norm and normalize it by the L_2 norm and the H^1 norm of the exact solution, respectively. The results are shown in Figure 6.24 for the L_2 norm, and in Figure 6.25 for the H^1 norm. The latter shows results only for the $\varepsilon = 1\text{e}-4$ case. The y-axis displays the natural logarithm of the relative error

$$\log\left(\frac{\left\|u-u^h\right\|}{\left\|u\right\|}\right) \tag{6.104}$$

where $\|\cdot\|$ denotes L_2- or H^1-norm, as appropriate, which are defined by $\left\|v\right\|_{L_2(\Omega)} = \sqrt{\int_\Omega v^2 d\Omega}$ and $\left\|v\right\|_{H^1(\Omega)} = \sqrt{\int_\Omega (v^2 + v_{,x}^2 + v_{,y}^2) d\Omega}$, respectively.

Given the dependence of the Damkohler number on the element length-scale, mesh refinement causes the Damkohler number to decrease. Thus, the problem goes from reaction dominated with boundary layers of length-scales smaller than the elements ($\alpha > \sqrt{6}$), to the diffusion dominated regime, where the boundary layers are captured across several elements ($\alpha < \sqrt{6}$).

In Figure 6.24, we see that both kinds of elements (Q4 and T3) converge optimally (rate of convergence of 2) for the smooth problem ($\varepsilon = 1$). For the problem with $\varepsilon = 1\text{e}-4$, convergence is initially slow in the region where the element size leads to large Damkohler numbers and boundary layers arise, whereas the convergence accelerates with h-refinement as the Damkohler number transitions to diffusion dominated regime and convergence rate becomes near optimal. The vertical dashed line marks the value of the element length scale h that corresponds to the critical value of the Damkohler number $\alpha_{cr} = \sqrt{6}$ for the problem with

$\varepsilon = 1\mathrm{e}-4$. On the other hand, for $\varepsilon = 1$, the problem already lies in the diffusion dominated regime (i.e. the value of h that corresponds to the critical value of the Damkohler number in that case, lies far to the left, outside the scope of the plot).

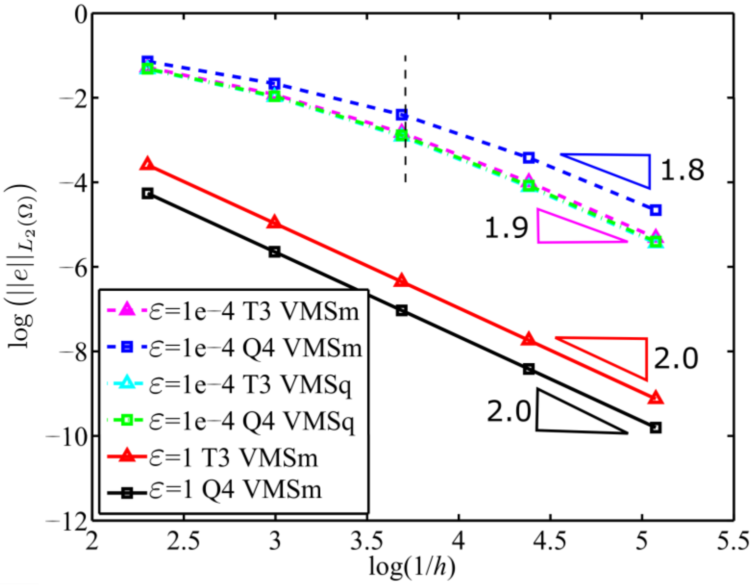


Figure 6.24. Log-log plot showing the convergence rate of the L_2 norm of the error (normalized by the L_2 norm of the solution) with mesh refinement for Q4 and T3 elements, for a smooth solution ($\varepsilon = 1$), and a solution with boundary layers ($\varepsilon = 1\text{e-}4$) for multiple fine-scale basis functions (VMSm) and the standard quadratic bubble (VMSq).

Remark: The absolute error of the VMSm with T3 is less than that with Q4 elements, as opposed to in most other methods. One possible reason is that for a patch defined by 4 nodes (one Q4 or two T3 elements), the fine scales are expanded in terms of 5 functions for one Q4, and 6 functions for two T3 elements, and therefore the T3 discretization gives rise to a more flexible fine-scale basis.

In Figure 6.25 we show the convergence of the error in the H^1 norm for the VMS methods with multiple fine-scale basis functions as well as with the standard quadratic bubble and it is compared with the error in the H^1 norm for the Galerkin method. Note that all the lines are practically overlapping. Here, convergence is initially suboptimal, and it asymptotically achieves near-optimal convergence rates, which are reported in Table 6.7.

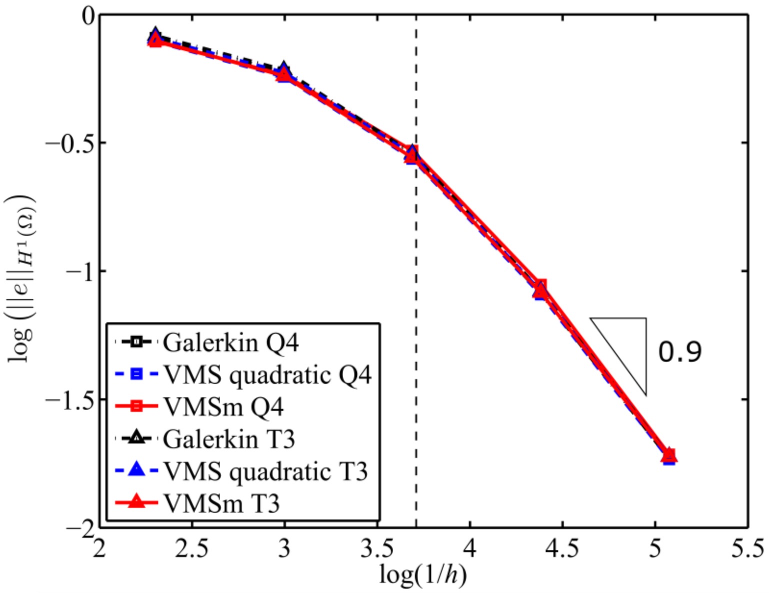


Figure 6.25. Log-log plot showing the convergence of the error in the H1 norm (normalized by the H1 norm of the solution) with mesh refinement for Q4 and T3 elements, $\varepsilon = 1\text{e-}4$. Vertical dashed line marks the critical value of the Damkohler number $\alpha_{cr} = \sqrt{6}$

Table 6.7. Convergence Rates for Different Methods

Method	Convergence rate in H^1 norm
Galerkin Q4	0.931
VMS quadratic Q4	0.926
VMSm Q4	0.954
Galerkin T3	0.930
VMS quadratic T3	0.928
VMSm T3	0.923

6.9.6 Maximum principle

Application of the maximum principle to the diffusion-reaction equation in (6.1) dictates that the solution u must satisfy the inequality $u \leq f/\sigma = 1$ at any point in Ω . Using the same meshes and parameters from section 6.9.1 we vary the diffusion coefficient between 1e-3 to 1e-7, which gives different Damkohler numbers (between 1.6 and 160) and record the maximum value of the computed solution for different methods. These results are plotted in Figure 6.26.

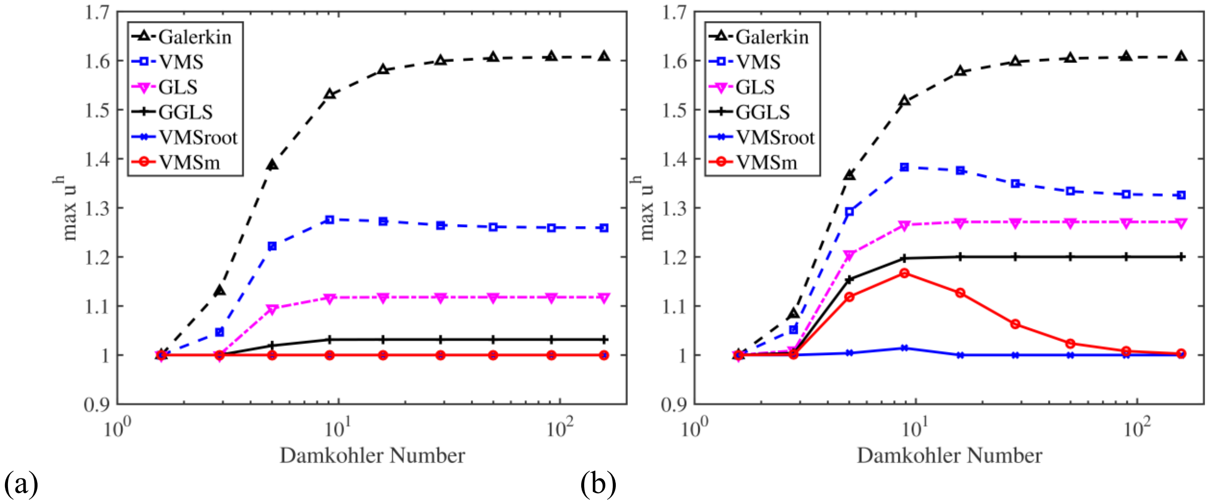


Figure 6.26. Deviation from the maximum principle for different methods on meshes of (a) Q4 and (b) T3 elements.

The different methods give numerical solutions that are in violation of the maximum principle for $\alpha > \sqrt{6}$. Only the VMSm and VMSroot methods are able to satisfy the maximum principle for the range of Damkohler number explored for discretizations utilizing Q4 elements. In the case of meshes comprised of T3 elements, we observe that the VMSm and VMSroot are in compliance with the maximum principle initially, and also asymptotically for larger values of the Damkohler number, but there is an intermediate range of Damkohler number values for which the maximum principle is not exactly satisfied, although the deviation is larger for VMSm. This is reminiscent of the so-called resonance error observed in [18] for the PGEM, but where the GEM method is shown to eliminate oscillations for the case of T3 elements in a crisscross mesh (for this study, the mesh from Figure 6.6b was used instead). The deviation from the maximum principle in Q4 elements with VMSroot is comparable to that of VMSm, while for T3 elements, VMSroot has only a slight deviation as compared to VMSm for the intermediate range where resonance occurs. However, the VMSroot method may produce over-diffused solutions, while VMSm captures steep layers more sharply, and VMSroot also fails to capture anisotropic features such as in section 6.9.2, as opposed to the VMSm formulation.

6.9.7 Heterogeneous Coefficients: Checkerboard Problem

In this problem, we consider the scalar diffusion reaction equation with heterogeneous coefficients, i.e. $\sigma = \sigma(\mathbf{x})$ and $\varepsilon = \varepsilon(\mathbf{x})$. Moreover, there are strong discontinuities in the spatial variation of these material parameters as in equations (6.105) and (6.106), in a checkerboard pattern (Figure 6.27). The problem domain is the unit square $\Omega = [0,1] \times [0,1]$, and we impose homogeneous Dirichlet boundary conditions (Figure 6.27) and a constant forcing function $f = 1$ over the entire domain.

$$\sigma(\mathbf{x}) = \begin{cases} \sigma_1 & \text{if } \mathbf{x} \text{ in black material} \\ \sigma_2 & \text{if } \mathbf{x} \text{ in white material} \end{cases} \tag{6.105}$$

$$\varepsilon(\mathbf{x}) = \begin{cases} \varepsilon_1 & \text{if } \mathbf{x} \text{ in black material} \\ \varepsilon_2 & \text{if } \mathbf{x} \text{ in white material} \end{cases} \tag{6.106}$$

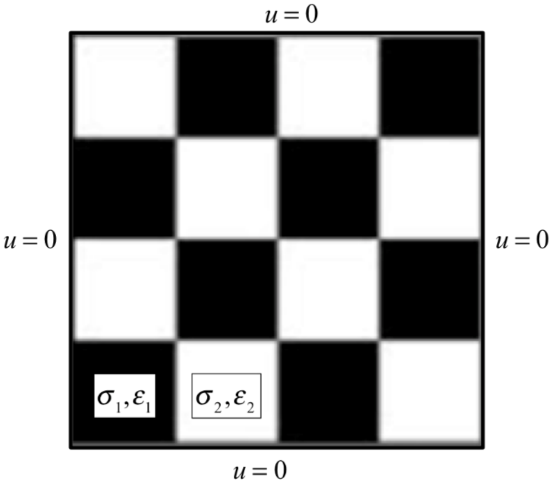
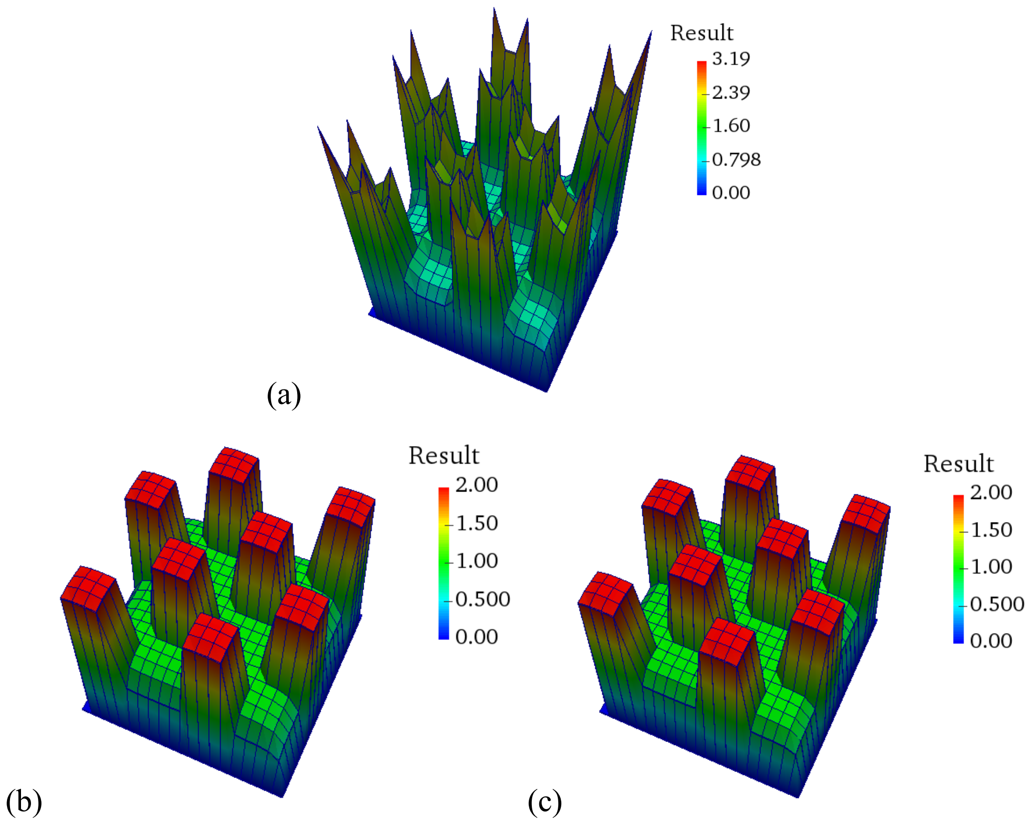


Figure 6.27. Schematic drawing of the domain for the checkerboard problem.

The domain is discretized using a 20×20 grid of 4-noded quadrilateral elements. Then, the element length scale is $h_K = 1/20 = 0.05$, and the Damkohler number for each material region in the checkerboard can be defined as in equation (6.107).

$$\alpha_1^2 = \frac{\sigma_1 h_K^2}{\varepsilon_1}, \quad \text{and} \quad \alpha_2^2 = \frac{\sigma_2 h_K^2}{\varepsilon_2} \tag{6.107}$$

The values of the coefficients are $\sigma_1 = 0.5$, $\varepsilon_1 = 1\text{e}-6$, $\sigma_2 = 1$, and $\varepsilon_2 = 1\text{e}-3$; consequently, the values of the Damkohler numbers are $\alpha_1^2 = 1250 > \alpha_{cr}^2$ and $\alpha_2^2 = 2.5 < \alpha_{cr}^2$. Thus, the black regions are *reaction-dominated,* and the white regions are *diffusion-dominated.* The numerical results are shown in Figure 6.28 for different methods.



(a)

(b)

(c)

Figure 6.28. Solution field obtained with different methods: (a) Galerkin, (b) VMSm, and (c) Selective VMSm.

The Galerkin solution shows spurious oscillations near the edges of the black squares of the checkerboard patten. These oscillations are suppressed by the VMSm and Selective VMSm methods. Furthermore, we remark that the maximum principle requires the exact solution to satisfy $u \leq \max f/\sigma = 2$, which the Galerkin FEM solution does not, while the VMS solutions with flexible fine scale basis (VMSm and Selective VMSm) do satisfy it. Therefore, the VMSm method is robust for application to heterogeneous material properties, even with strong discontinuities in the spatial distribution of material coefficients.

For this case, computation of τ_m^e at the element level requires solution of a 5×5 system of linear equations, whereas computation of τ_s^e requires solution of a single linear equation. In practice, the additional cost incurred by solving the 5×5 systems in each element is negligible compared to other parts of the solution procedure (e.g., assembly, global problem solution). However, in higher dimensions and for higher order elements, the fine-scale basis may require more functions, and solution of the fine-scale problem may no longer be insignificant. In those situations, one may adopt the Selective VMSm approach to reduce computational cost.

6.10 Conclusions

Within the VMS framework, we have presented a new, more flexible approach for defining fine-scale models. The main characteristics of the proposed model are: (i) the fine scales are expanded with a basis that is comprised of multiple functions, (ii) those basis functions are not necessarily zero at element interfaces, and (iii) the fine scales are not explicitly enforced to be continuous across element interfaces. This results in a VMS method with multiple functions for the fine-scale basis that we have denoted VMSm for ease of notation. The improved fine-scale model of the VMSm inherits the error estimation features of VMS methods and the explicit local error estimate can be used to provide information about the distribution of the error within the domain. As one may expect, the error is localized near the boundary layers as shown in the numerical results. Since local solution of the fine-scale variational problem yields the fine-scale model that is proportional to the residual of the Euler-Lagrange equations of the coarse scale, pre-multiplied by the stabilization parameter, and that the coarse-scale fields are expanded in terms of interpolation functions that are continuous, the local discontinuity of the fine-scale basis functions only makes the stabilization parameter discontinuous across the element boundaries. It is important to note that this is the case even when element-wise mean value of the stabilization parameter is employed in non-uniform meshes.

The method presented herein was implemented for singular diffusion-reaction problems that exhibit boundary and internal layers with steep gradients. The method is able to suppress the spurious oscillations observed in the Galerkin FEM for diffusion-reaction equation in the reaction-dominated regime. The oscillations persist in VMS methods when the standard quadratic bubble is employed for expanding the fine-scale field, as well as in other methods such as GLS and GGLS when applied to this class of problems. The VMSm method is able to capture boundary layers within a single element, as shown in the numerical results from sections 6.9.1 to 6.9.4, and internal layers within two elements, as shown in the numerical test case in section 6.9.4. Although the results show a diffusive response in certain cases, this may be a consequence of neglecting the contribution of the boundary terms during the derivation of the stabilization parameter τ^e. Nevertheless, we show optimal convergence of the method for smooth problems, as well as for problems with boundary layers. We also show the stability and accuracy of the method for problems with discontinuous forcing function and internal layers (section 6.9.4), and its performance with distorted quadrilateral elements and unstructured meshes of triangular elements (section 6.9.3). In addition to this attributes, section 6.9.6 shows that the VMSm has improved performance regarding the maximum principle, as compared to other stabilized methods. Moreover, the method has the flexibility to automatically provide stabilization as needed along different directions independently, as shown in the numerical section 6.9.2 for a mesh with anisotropic refinement. The VMSm method is also robust for application to heterogeneous diffusion-reaction problems with strong discontinuities in the material coefficients. For this situation, a simple and optional modification to the method, termed as the Selective VMSm, allows for some computational cost savings while retaining all the benefits of the VMSm method.

6.11 References

[1] L.P. Franca, E.G. Dutra Do Carmo, The Galerkin gradient least-squares method, Comput. Methods Appl. Mech. Eng. 74 (1989) 41–54.

[2] I. Harari, Stability of semidiscrete formulations for parabolic problems at small time steps, Comput. Methods Appl. Mech. Eng. 193 (2004) 1491–1516.

[3] A.N. Brooks, T.J.R. Hughes, Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations, Comput. Methods Appl. Mech. Eng. 32 (1982) 199–259.

[4] T.J.R. Hughes, M. Mallet, M. Akira, A new finite element formulation for computational fluid dynamics: II. Beyond SUPG, Comput. Methods Appl. Mech. Eng. 54 (1986) 341–355.

[5] T.J.R. Hughes, L.P. Franca, G.M. Hulbert, A new finite element formulation for computational fluid dynamics: VIII. The galerkin/least-squares method for advective-diffusive equations, Comput. Methods Appl. Mech. Eng. 73 (1989) 173–189.

[6] T.E. Tezduyar, Y.J. Park, Discontinuity-capturing finite element formulations for nonlinear convection-diffusion-reaction equations, Comput. Methods Appl. Mech. Eng. 59 (1986) 307–325.

[7] I. Harari, T.J.R. Hughes, Stabilized finite element methods for steady advection-diffusion with production, Comput. Methods Appl. Mech. Eng. 115 (1994) 165–191.

[8] L.P. Franca, C. Farhat, Bubble functions prompt unusual stabilized finite element methods, Comput. Methods Appl. Mech. Eng. 123 (1995) 299–308.

[9] T.J.R. Hughes, Multiscale phenomena: Green's functions, the Dirichlet-to-Neumann formulation, subgrid scale models, bubbles and the origins of stabilized methods, Comput. Methods Appl. Mech. Eng. 127 (1995) 387–401.

[10] T.J.R. Hughes, G.R. Feijóo, L. Mazzei, J.-B. Quincy, The variational multiscale method—a paradigm for computational mechanics, Comput. Methods Appl. Mech. Eng. 166 (1998) 3–24.

[11] F. Brezzi, A. Russo, Choosing bubbles for advection-diffusion problems, Math. Model. Methods Appl. Sci. 04 (1994) 571–587.

[12] L.P. Franca, A. Russo, Deriving upwinding, mass lumping and selective reduced integration by residual-free bubbles, Appl. Math. Lett. 9 (1996) 83–88.

[13] F. Brezzi, D. Marini, A. Russo, Applications of the pseudo residual-free bubbles to the stabilization of convection-diffusion problems, Comput. Methods Appl. Mech. Eng. 166 (1998) 51–63.

[14] F. Brezzi, L.P. Franca, T.J.R. Hughes, A. Russo, $b = \int g$, Comput. Methods Appl. Mech. Eng. 145 (1997) 329–339.

[15] R. Codina, Comparison of some finite element methods for solving the diffusion-convection-reaction equation, Comput. Methods Appl. Mech. Eng. 156 (1998) 185–210.

[16] G. Hauke, A. García-Olivares, Variational subgrid scale formulations for the advection–diffusion-reaction equation, Comput. Methods Appl. Mech. Eng. 190 (2001) 6847–6865.

[17] E.G. Dutra do Carmo, G.B. Alvarez, F.A. Rochinha, A.F.D. Loula, Galerkin projected residual method applied to diffusion–reaction problems, Comput. Methods Appl. Mech. Eng. 197 (2008) 4559–4570.

[18] H. Fernando, C. Harder, D. Paredes, F. Valentin, Numerical multiscale methods for a reaction-dominated model, Comput. Methods Appl. Mech. Eng. 201–204 (2012) 228–244.

[19] E. Oñate, P. Nadukandi, J. Miquel, Accurate FIC-FEM formulation for the multidimensional steady-state advection–diffusion–absorption equation, Comput. Methods Appl. Mech. Eng. 327 (2017) 352–368.

[20] A. Masud, A.A. Al-Naseem, Variationally derived discontinuity capturing methods: Fine scale models with embedded weak and strong discontinuities, Comput. Methods Appl. Mech. Eng. 340 (2018) 1102–1134.

[21] A. Masud, G. Scovazzi, A heterogeneous multiscale modeling framework for hierarchical systems of partial differential equations, Int. J. Numer. Methods Fluids. 65 (2011) 28–42.

[22] A. Masud, R.A. Khurram, A multiscale finite element method for the incompressible Navier–Stokes equations, Comput. Methods Appl. Mech. Eng. 195 (2006) 1750–1777.

[23] A. Masud, R. Calderer, A variational multiscale method for incompressible turbulent flows: Bubble functions and fine scale fields, Comput. Methods Appl. Mech. Eng. 200 (2011) 2577–2593.

[24] A. Masud, K. Xia, A variational multiscale method for inelasticity: Application to superelasticity in shape memory alloys, Comput. Methods Appl. Mech. Eng. 195 (2006) 4512–4531.

[25] K.B. Nakshatrala, A. Masud, K.D. Hjelmstad, On finite element formulations for nearly incompressible linear elasticity, Comput. Mech. 41 (2007) 547–561.

[26] R. Hall, H. Gajendran, A. Masud, Diffusion of chemically reacting fluids through nonlinear elastic solids: mixture model and stabilized methods, Math. Mech. Solids. 20 (2015) 204–227.

[27] M. Anguiano, H. Gajendran, R.B. Hall, K.R. Rajagopal, A. Masud, Chemo-mechanical coupling and material evolution in finitely deforming solids with advancing fronts of reactive fluids, Acta Mech. 231 (2020) 1933–1961.

[28] J. Principe, R. Codina, On the stabilization parameter in the subgrid scale approximation of scalar convection–diffusion–reaction equations on distorted meshes, Comput. Methods Appl. Mech. Eng. 199 (2010) 1386–1402.

[29] C. Farhat, I. Harari, L.P. Franca, The discontinuous enrichment method, Comput. Methods Appl. Mech. Eng. 190 (2001) 6455–6479.

[30] N.C.B. Arruda, R.C. Almeida, E.G. Dutra do Carmo, Discontinuous subgrid formulations for transport problems, Comput. Methods Appl. Mech. Eng. 199 (2010) 3227–3236.

[31] C. Coley, J.A. Evans, Variational multiscale modeling with discontinuous subscales: analysis and application to scalar transport, Meccanica. 53 (2018) 1241–1269.

[32] T.J. Truster, A. Masud, Primal interface formulation for coupling multiple PDEs: A consistent derivation via the Variational Multiscale method, Comput. Methods Appl. Mech. Eng. 268 (2014) 194–224.

[33] T.J. Truster, P. Chen, A. Masud, Finite strain primal interface formulation with consistently evolving stabilization, Int. J. Numer. Methods Eng. 102 (2015) 278–315.

[34] J. Wang, X. Ye, A weak Galerkin finite element method for second-order elliptic problems, J. Comput. Appl. Math. 241 (2013) 103–115.

[35] T.J.R. Hughes, G. Sangalli, Variational Multiscale Analysis: the Fine-scale Green's Function, Projection, Optimization, Localization, and Stabilized Methods, SIAM J. Numer. Anal. 45 (2007) 539–557.

[36] T.J.R. Hughes, G. Scovazzi, L.P. Franca, Multiscale and Stabilized Methods, in: Encycl. Comput. Mech. Second Ed., John Wiley & Sons, Ltd, Chichester, UK, 2017: pp. 1–64.

[37] A. Masud, R.A. Khurram, A multiscale/stabilized finite element method for the advection–diffusion equation, Comput. Methods Appl. Mech. Eng. 193 (2004) 1997–2018.

[38] A. Masud, L.P. Franca, A hierarchical multiscale framework for problems with multiscale source terms, Comput. Methods Appl. Mech. Eng. 197 (2008) 2692–2700.

[39] S.C. Aduloju, T.J. Truster, A variational multiscale discontinuous Galerkin formulation for both implicit and explicit dynamic modeling of interfacial fracture, Comput. Methods Appl. Mech. Eng. 343 (2019) 602–630.

[40] M.T. Heath, Scientific Computing: An Introductory Survey, Second ed., Society for Industrial and Applied Mathematics, Philadelphia, PA, 2018.

[41] C. Oskay, Variational multiscale enrichment method with mixed boundary conditions for modeling diffusion and deformation problems, Comput. Methods Appl. Mech. Eng. 264

(2013) 178–190.

[42] L.P. Franca, F. Valentin, On an improved unusual stabilized finite element method for the advective–reactive–diffusive equation, Comput. Methods Appl. Mech. Eng. 190 (2000) 1785–1800.

[43] A. Masud, J. Kwack, A stabilized mixed finite element method for the first-order form of advection–diffusion equation, Int. J. Numer. Methods Fluids. 57 (2008) 1321–1348.

[44] I. Harari, T.J.R. Hughes, What are C and h?: Inequalities for the analysis and design of finite element methods, Comput. Methods Appl. Mech. Eng. 97 (1992) 157–192.

[45] A. Masud, T.J. Truster, L.A. Bergman, A variational multiscale a posteriori error estimation method for mixed form of nearly incompressible elasticity, Comput. Methods Appl. Mech. Eng. 200 (2011) 3453–3481.

CHAPTER 7:
A VARIATIONAL MULTISCALE METHOD FOR TRANSIENT COUPLED THERMO-CHEMO-MECHANICAL PROBLEMS.

7.1 Introduction

A thermodynamically consistent model was presented for thermo-chemo-mechanical processes in deformable solids permeated by reactive fluids. In the reaction-dominated regime, the governing system of equations produces sharp traveling reaction fronts across the deforming domain. Fully resolving this traveling internal layers would require mesh refinement throughout the entire domain if a standard Galerkin Finite Element Method is used, which may be prohibitively expensive in terms of cost of computation. On the other hand, A variational multiscale method (VMS) for reaction dominated problems was presented in Chapter 6. This Chapter extends the method development carried out in Chapter 6 and applies it to thermo-chemo-mechanical problems. First, a VMS method is developed for the scalar transient diffusion-reaction equation with material evolution. Important consideration is placed in options regarding the time-evolution of the fine-scale field, as well other algorithmic choices. Then, these ideas are applied to develop a VMS method for an extension of the thermo-chemo-mechanical model presented in Chapter 4. The general VMS method considers multiscale split of both the fluid mass concentration and temperature fields. Numerical results are presented to highlight the features of the method: the so-called "bird's beak" problem is revisited, and fingering phenomena related to solid fuel combustion are studied.

7.2 Transient Reaction-Diffusion with Material Evolution.

Consider a domain $\Omega \in \mathbb{R}^{n_{sd}}$ with boundary $\partial\Omega = \partial\Omega_D \cup \partial\Omega_N$, $\partial\Omega_D \cap \partial\Omega_N = \varnothing$, a time interval of interest $I = [0,T]$, and a scalar field $u : \Omega \times I \rightarrow \mathbb{R}^+$. We consider a scalar transient (first order in time) reaction-diffusion equation where the problem parameters evolve with reaction. The weak form of the problem is: Given appropriate spaces $\mathcal{S}_t$ and $\mathcal{V}$, find $u \in \mathcal{S}_t$ such that $\forall w \in \mathcal{V}$ the following holds

$$\left(\frac{\partial w}{\partial x_i}, \quad \varepsilon_{ij}(\phi) \frac{\partial u}{\partial x_j} \right)_\Omega + \left(w, \quad \frac{\partial u}{\partial t} + \sigma(\phi, u) \right)_\Omega = (w, \quad f)_\Omega + (w, \quad h)_{\partial\Omega_N} \tag{7.1}$$

where the trial and test spaces, accounting for homogeneous Dirichlet boundary conditions, are

$$\mathcal{S}_t = \left\{ u(x,t) \mid u(x,t) \in H^1(\Omega) \text{ and } u(x,t) = 0 \ \forall x \in \partial\Omega_D \right\} \tag{7.2}$$

$$\mathcal{V} = \left\{ w(x) \mid w(x) \in H^1(\Omega) \text{ and } w(x) = 0 \ \forall x \in \partial\Omega_D \right\} \tag{7.3}$$

Subject to initial and boundary conditions

$$u(x, t=0) = u_0(x) \qquad x \in \Omega \tag{7.4}$$

$$u(x,t) = u_D(x,t) \qquad x \in \partial\Omega_D \tag{7.5}$$

$$n_i \varepsilon_{ij}(\phi) \frac{\partial u}{\partial x_j} = h(x,t) \quad x \in \partial\Omega_N \tag{7.6}$$

Where n_i is the outward unit normal to $\partial\Omega_N$. The material evolution is tracked by the reaction state variable $\phi : \Omega \times I \to \mathbb{R}^+$, which is governed by the following ODE

$$\dot{\phi} = r(\phi, u) \tag{7.7}$$

$$r(\phi,u) \le 0, \qquad r(0,u) = 0, \qquad r(\phi,0) = 0 \tag{7.8}$$

$$\phi(x, t=0) = \phi_0 \ge 0 \tag{7.9}$$

We introduce the following notation

$$\mathcal{R}(u,v,w) = \left(\frac{\partial w}{\partial x_i}, \quad \varepsilon_{ij}(\phi) \frac{\partial u}{\partial x_j} \right)_\Omega + \left(w, \quad v + \sigma(\phi,u)\right)_\Omega - \left(w, \quad f\right)_\Omega - \left(w, \quad h\right)_{\partial\Omega_N} \tag{7.10}$$

where $v = \partial u / \partial t$.

7.2.1 Time discretization

The time domain is partitioned into N subintervals, where the nth subinterval is $[t_{n-1}, t_n]$ delimited by points t_n. We introduce the following notation:

$$u_n(x) \approx u(x, t_n) \tag{7.11}$$

$$\phi_n(x) \approx \phi(x, t_n) \tag{7.12}$$

$$v_n(x) \approx \left. \frac{\partial u(x,t)}{\partial t} \right|_{t=t_n} \tag{7.13}$$

We consider a time discretization scheme of the PDE in equation (7.1) where we advance the solution variables from t_n to t_{n+1}, assuming that the quantities at t_n and all preceding time points are known. We assume that the time discretization algorithm has the following form[1]

$$\mathcal{R}\left(u_{n+\alpha}, v_{n+\alpha}, w\right) = 0 \tag{7.14}$$

where

$$v_{n+1} = c_0 v_{n+\alpha} + \breve{v}_{n+1} \tag{7.15}$$

$$u_{n+1} = c_1 \Delta t v_{n+\alpha} + \breve{u}_{n+1} \tag{7.16}$$

$$u_{n+\alpha} = c_2 u_{n+1} + \breve{u}_{n+\alpha} \tag{7.17}$$

The quantities with a superimposed "breve" are linear combinations of the known data pool up to and including time level t_n. This yields explicit methods when $c_1 = 0$. Here, we will consider implicit methods. For instance, we can recover the backward Euler method by setting

$$c_0 = 1, \qquad \breve{v}_{n+1} = 0, \qquad c_1 = 1, \qquad \breve{u}_{n+1} = u_n, \qquad c_2 = 1, \qquad \breve{u}_{n+\alpha} = 0 \tag{7.18}$$

which yields

$$v_{n+\alpha} = v_{n+1} = \left(u_{n+1} - u_n\right)/\Delta t, \quad u_{n+\alpha} = u_{n+1} \tag{7.19}$$

We apply the same time discretization to the evolution equation:

$$\dot{\phi}_{n+\alpha} = r\left(\phi_{n+\alpha}, u_{n+\alpha}\right) \tag{7.20}$$

where

$$\dot{\phi}_{n+1} = c_0 \dot{\phi}_{n+\alpha} + \breve{\phi}_{n+1} \tag{7.21}$$

$$\phi_{n+1} = c_1 \Delta t \dot{\phi}_{n+\alpha} + \breve{\phi}_{n+1} \tag{7.22}$$

$$\phi_{n+\alpha} = c_2 \phi_{n+1} + \breve{\phi}_{n+\alpha} \tag{7.23}$$

We note the following algorithmic relationship between $\phi_{n+\alpha}$ and $u_{n+\alpha}$. First, substituting (7.20) into (7.22), and then substituting into (7.23) we obtain:

$$\phi_{n+\alpha} = c_2 \left(c_1 \Delta t \cdot r\left(\phi_{n+\alpha}, u_{n+\alpha}\right) + \breve{\phi}_{n+1}\right) + \breve{\phi}_{n+\alpha} \tag{7.24}$$

Then, taking the partial derivative with respect to $u_{n+\alpha}$ on both sides

---

[1] This form is general enough that it can accommodate BDF, Newmark-type, Generalized-α, and explicit RK methods through appropriate definitions of the constants and known "predictors".

$$\frac{\partial \phi_{n+\alpha}}{\partial u_{n+\alpha}} = c_2 c_1 \Delta t \left(\frac{\partial r}{\partial u_{n+\alpha}} + \frac{\partial r}{\partial \phi_{n+\alpha}} \frac{\partial \phi_{n+\alpha}}{\partial u_{n+\alpha}} \right) \tag{7.25}$$

Gathering and solving for $\partial \phi_{n+\alpha} / \partial u_{n+\alpha}$

$$\frac{\partial \phi_{n+\alpha}}{\partial u_{n+\alpha}} = c_2 c_1 \Delta t \left(1 - c_2 c_1 \Delta t \frac{\partial r}{\partial \phi_{n+\alpha}} \right)^{-1} \frac{\partial r}{\partial u_{n+\alpha}} \tag{7.26}$$

If the method is explicit $c_1 = 0$, then $\partial \phi_{n+\alpha} / \partial u_{n+\alpha} = 0$. We note that $\phi_{n+\alpha}$ is a *nonlinear* function of $u_{n+\alpha}$ through the implicit relation in (7.24) if the function r actually depends on ϕ or if it depends on u in a nonlinear way. By using an implicit time integration method, the discrete problem becomes not only implicit, but also nonlinear.

7.3 Variational Multiscale Method (VMS)

Here, we consider a generic solution step for the governing weak form, that is, we are concerned with solution of:

$$\mathcal{R}\left(u_{n+\alpha}, v_{n+\alpha}, w\right) = 0 \tag{7.27}$$

For generality and ease of notation, we drop the subscripts $n+\alpha$ in the derivation that follows. In this section we apply the VMS method [1,2] to the transient diffusion-reaction equation with material evolution. In particular, we follow the developments in [3,4]. The derivation of the VMS method begins with the assumption that the test and trial functions admit an additive decomposition into a coarse scale contribution and fine scale contribution.

$$u = \bar{u} + u' \tag{7.28}$$

$$w = \bar{w} + w' \tag{7.29}$$

The uniqueness of such decomposition may be formalized by introducing appropriate projection operators as discussed in [5]. In addition, one may consider a multiscale split of the time dependent term as follows

$$\dot{u} = v = \bar{v} + v' \tag{7.30}$$

The coarse scales are associated to the standard finite element spaces of functions. Using linearity of test function slot and linear independence of coarse and fine scale spaces, the weak form gives rise to a coarse-scale subproblem and a fine-scale subproblem

Coarse scale subproblem

$$\left(\frac{\partial \bar{w}}{\partial x_i},\quad \varepsilon_{ij}\left(\phi\right)\frac{\partial u}{\partial x_j}\right)_{\Omega} + \left(\bar{w},\quad v + \sigma\left(\phi, u\right)\right)_{\Omega} - \left(\bar{w},\quad f\right)_{\Omega} - \left(\bar{w},\quad h\right)_{\partial\Omega_N} = 0 \tag{7.31}$$

Fine scale subproblem

$$\left(\frac{\partial w'}{\partial x_i},\quad \varepsilon_{ij}\left(\phi\right)\frac{\partial u}{\partial x_j}\right)_{\Omega} + \left(w',\quad v + \sigma\left(\phi, u\right)\right)_{\Omega} - \left(w',\quad f\right)_{\Omega} - \left(w',\quad h\right)_{\partial\Omega_N} = 0 \tag{7.32}$$

7.3.1 Fine scale subproblem

To find a numerical solution for the fine-scale field in a computationally efficient way, we require that the fine-scale subproblem be satisfied at the element level.

$$\left(\frac{\partial w'}{\partial x_i},\quad \varepsilon_{ij}\left(\phi\right)\frac{\partial u}{\partial x_j}\right)_{K} + \left(w',\quad v + \sigma\left(\phi, u\right)\right)_{K} + \left(w',\quad f\right)_{K} - \left(w',\quad h\right)_{\partial K \cap \partial\Omega_N} = 0 \tag{7.33}$$

For an explicit time integration algorithm, it is possible to solve for the fine scale velocity directly (in terms of the coarse scales and the known data pool including fine scales), and then update the fine-scale field itself according to the time integration algorithm. On the other hand, if the time integration algorithm is implicit, then we are required to linearize the element level weak form around the coarse scale trial field to find an approximate solution.

As part of the linearization, we note that ϕ depends on the unknown field u at the solution time level $t_{n+\alpha}$ when the time integration algorithm is implicit. In developing the algorithm, one may choose to assume that the internal variable ϕ depends only on the coarse scale contribution $\bar{u}$ and is independent of the fine scales u'. For completeness, the derivation that follows assumes that ϕ does depend on the fine scale contribution u'. However, for ease of implementation, we adopt the algorithmic choice of assuming that ϕ depends only on the coarse scale contribution $\bar{u}$ for the purposes of the numerical test cases presented in this Chapter. The corresponding method is recovered by setting to zero the $\partial\phi/\partial u$ terms that appear in the linearization process below. An analogous assumption is made in the VMS method presented in [6], where the fluid mass reaction term was assumed to depend only on the coarse scale contribution to the fluid mass concentration field.

Solution of the linearized weak form is also a valid approximation process for the fine scales in the explicit time integration case. The left-hand-side of equation (7.33) is linearized with respect to the scalar field u in a neighborhood of $\bar{u}$ in the direction of the fine scale field u' to yield

$$\begin{aligned}
&\left(\frac{\partial w'}{\partial x_i},\ \ \bar{\varepsilon}_{ij}\frac{\partial \bar{u}}{\partial x_j}\right)_K + \left(w',\ \ \bar{v} + \sigma\left(\bar{\phi},\bar{u}\right)\right)_K - \left(w',\ \ f\right)_K - \left(w',\ \ h\right)_{\partial K\cap\partial\Omega_N} \\
&\qquad + \left(\frac{\partial w'}{\partial x_i},\ \ \left[\left.\left(\frac{\partial \varepsilon_{ij}}{\partial \phi}\frac{\partial \phi}{\partial u}\right)\right|_{\bar{u}} u'\right]\frac{\partial \bar{u}}{\partial x_j}\right)_K + \left(\frac{\partial w'}{\partial x_i},\ \ \bar{\varepsilon}_{ij}\frac{\partial u'}{\partial x_j}\right)_K \\
&\qquad\qquad + \left(w',\ \ v' + \left[\left.\left(\frac{\partial \sigma}{\partial \phi}\frac{\partial \phi}{\partial u}\right)\right|_{\bar{u}} + \frac{\partial \sigma}{\partial u}\right]u'\right)_K = 0
\end{aligned} \tag{7.34}$$

To simplify the expression, we introduce the following notation

$$\hat{a}_i\left(\bar{u}\right) = \left.\frac{\partial \varepsilon_{ij}}{\partial \phi}\frac{\partial \phi}{\partial u}\right|_{\bar{u}}\frac{\partial \bar{u}}{\partial x_j} \tag{7.35}$$

$$\hat{\varepsilon}_{ij}\left(\bar{u}\right) = \bar{\varepsilon}_{ij} \tag{7.36}$$

$$\hat{\sigma}\left(\bar{u}\right) = \left.\left[\frac{\partial \sigma}{\partial \phi}\frac{\partial \phi}{\partial u} + \frac{\partial \sigma}{\partial u}\right]\right|_{\bar{u}} \tag{7.37}$$

$$\hat{A}^K\left(w', u'; \bar{u}\right) = \left(\frac{\partial w'}{\partial x_i},\ \ \hat{\varepsilon}_{ij}\left(\bar{u}\right)\frac{\partial u'}{\partial x_j}\right)_K + \left(\frac{\partial w'}{\partial x_i},\ \ \hat{a}_i\left(\bar{u}\right)u'\right)_K + \left(w',\ \ \hat{\sigma}\left(\bar{u}\right)u'\right)_K \tag{7.38}$$

$$\hat{M}^K\left(w', v'\right) = \left(w',\ \ v'\right)_K \tag{7.39}$$

$$\hat{F}^K\left(w'; \bar{u}, \bar{v}\right) = \left(\frac{\partial w'}{\partial x_i},\ \ \bar{\varepsilon}_{ij}\frac{\partial \bar{u}}{\partial x_j}\right)_K + \left(w',\ \ \bar{v} + \sigma\left(\bar{\phi},\bar{u}\right)\right)_K - \left(w',\ \ f\right)_K - \left(w',\ \ h\right)_{\partial K\cap\partial\Omega_N} \tag{7.40}$$

The last equation can be rewritten by applying integration by parts on the first term

$$\left(\frac{\partial w'}{\partial x_i},\ \ \bar{\varepsilon}_{ij}\frac{\partial \bar{u}}{\partial x_j}\right)_K = -\left(w',\ \ \frac{\partial}{\partial x_i}\left[\bar{\varepsilon}_{ij}\frac{\partial \bar{u}}{\partial x_j}\right]\right)_K + \left(w',\ \ n_i\bar{\varepsilon}_{ij}\frac{\partial \bar{u}}{\partial x_j}\right)_{\partial K} \tag{7.41}$$

$$\hat{F}^K\left(w'; \bar{u}, \bar{v}\right) = \left(w',\ \ \bar{R}\right)_K + \left(w',\ \ n_i\bar{\varepsilon}_{ij}\frac{\partial \bar{u}}{\partial x_j}\right)_{\partial K} - \left(w',\ \ h\right)_{\partial K\cap\partial\Omega_N} \tag{7.42}$$

where the residual of the coarse scale Euler-Lagrange equation is

$$\bar{R} = \bar{v} - \frac{\partial}{\partial x_i}\left[\bar{\varepsilon}_{ij}\frac{\partial \bar{u}}{\partial x_j}\right] + \sigma\left(\bar{\phi},\bar{u}\right) - f \tag{7.43}$$

Then, the linearized fine-scale subproblem can be written as

$$\hat{M}^K\left(w',\dot{u}';\overline{u}\right)+\hat{A}^K\left(w',u';\overline{u}\right)=-\hat{F}^K\left(w';\overline{u}\right) \tag{7.44}$$

Assuming a basis for the fine scale test and trial functions $\mathbf{b}=\left[b_1^K,\quad ...,\quad b_{n_b}^K\right]$ provides the spatial discretization for both u' and $\dot{u}'$ fields

$$\mathbf{M}\hat{\mathbf{v}}+\mathbf{A}\hat{\mathbf{u}}=\mathbf{f} \tag{7.45}$$

Where

$$\left[\hat{\mathbf{u}}\right]_p=\hat{u}_p^K \tag{7.46}$$

$$\left[\hat{\mathbf{v}}\right]_p=\hat{v}_p^K \tag{7.47}$$

$$\left[\mathbf{f}\right]_p=-\hat{F}^K\left(b_p^K;\overline{u},\overline{v}\right) \tag{7.48}$$

$$\left[\mathbf{A}\right]_{pq}=\hat{A}^K\left(b_p^K,b_q^K;\overline{u}\right) \tag{7.49}$$

$$\left[\mathbf{M}\right]_{pq}=\hat{M}^K\left(b_p^K,b_q^K\right) \tag{7.50}$$

where indices p , q $=1,...,n_b$ In combination with a time discretization algorithm, the governing equation for the fine scales (7.45) is solved to find the vectors of fine scale coefficients $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$. Then the fine scale fields within the element is recovered by taking

$$\left.\tilde{u}'\right|_K=\hat{u}_p^K\left(\overline{u},\overline{v}\right)b_p^K \tag{7.51}$$

$$\left.\tilde{v}'\right|_K=\hat{v}_p^K\left(\overline{u},\overline{v}\right)b_p^K \tag{7.52}$$

7.3.2 Coarse scale subproblem

Once the solution for the fine-scale field is obtained, it is embedded into the coarse scale subproblem. One alternative is to directly substitute the solution; however, in practice, it is often convenient to linearize the coarse scale problem in a manner similar to the way in which the linearized fine scale subproblem is obtained. First, recall the coarse scale subproblem

$$\left(\frac{\partial\overline{w}}{\partial x_i},\quad \varepsilon_{ij}\left(\phi\right)\frac{\partial u}{\partial x_j}\right)_\Omega+\left(\overline{w},\quad \dot{u}+\sigma\left(\phi,u\right)\right)_\Omega=-\left(\overline{w},\quad f\right)_\Omega+\left(\overline{w},\quad h\right)_{\partial\Omega_N} \tag{7.53}$$

Then, we expand the scalar field u in terms of the coarse-scale and fine-scale contributions

$$\begin{aligned}&\left(\frac{\partial\overline{w}}{\partial x_i},\quad \varepsilon_{ij}\left(\phi\left(\overline{u}+\tilde{u}'\left(\overline{u},\overline{v}\right)\right)\right)\frac{\partial\left[\overline{u}+\tilde{u}'\left(\overline{u},\overline{v}\right)\right]}{\partial x_j}\right)_\Omega\\&+\left(\overline{w},\quad \overline{v}+\tilde{v}'\left(\overline{u},\overline{v}\right)+\sigma\left(\phi\left(\overline{u}+\tilde{u}'\left(\overline{u},\overline{v}\right)\right),\overline{u}+\tilde{u}'\left(\overline{u},\overline{v}\right)\right)\right)_\Omega=-\left(\overline{w},\quad f\right)_\Omega+\left(\overline{w},\quad h\right)_{\partial\Omega_N}\end{aligned} \tag{7.54}$$

The linearized coarse scale subproblem is given by

$$\begin{aligned}
&\left(\frac{\partial \bar{w}}{\partial x_i},\quad \bar{\varepsilon}_{ij}\frac{\partial \bar{u}}{\partial x_j}\right)_\Omega + \left(\bar{w},\quad \bar{v}+\sigma\left(\bar{\phi},\bar{u}\right)\right)_\Omega - \left(\bar{w},\quad f\right)_\Omega - \left(\bar{w},\quad h\right)_{\partial\Omega_N} \\
&+\left(\frac{\partial \bar{w}}{\partial x_i},\quad \left[\left.\left(\frac{\partial \varepsilon_{ij}}{\partial \phi}\frac{\partial \phi}{\partial u}\right)\right|_{\bar{u}}\frac{\partial \bar{u}}{\partial x_j}\right]\tilde{u}'\left(\bar{u},\bar{v}\right)\right)_\Omega + \left(\frac{\partial \bar{w}}{\partial x_i},\quad \bar{\varepsilon}_{ij}\frac{\partial \tilde{u}'\left(\bar{u},\bar{v}\right)}{\partial x_j}\right)_\Omega \\
&+\left(\bar{w},\quad \tilde{v}'\left(\bar{u},\bar{v}\right)+\left.\left(\frac{\partial \sigma}{\partial \phi}\frac{\partial \phi}{\partial u}+\frac{\partial \sigma}{\partial u}\right)\right|_{\bar{u}}\tilde{u}'\left(\bar{u},\bar{v}\right)\right)_\Omega \qquad\qquad = 0
\end{aligned} \tag{7.55}$$

This expression can be rewritten as

$$\mathcal{R}\left(\bar{u},\bar{v},\bar{w}\right)+\mathcal{W}\left(\bar{u},\bar{v},\bar{w}\right)=0 \tag{7.56}$$

where

$$\mathcal{R}\left(\bar{u},\bar{v},\bar{w}\right)=\left(\frac{\partial \bar{w}}{\partial x_i},\quad \bar{\varepsilon}_{ij}\frac{\partial \bar{u}}{\partial x_j}\right)_\Omega + \left(\bar{w},\quad \bar{v}+\sigma\left(\bar{\phi},\bar{u}\right)\right)_\Omega - \left(\bar{w},\quad f\right)_\Omega - \left(\bar{w},\quad h\right)_{\partial\Omega_N} \tag{7.57}$$

$$\begin{aligned}
\mathcal{W}\left(\bar{u},\bar{v},\bar{w}\right)=&\left(\frac{\partial \bar{w}}{\partial x_i},\quad \hat{\varepsilon}_{ij}\left(\bar{u}\right)\frac{\partial \tilde{u}'\left(\bar{u},\bar{v}\right)}{\partial x_j}\right)_\Omega + \left(\frac{\partial \bar{w}}{\partial x_i},\quad \hat{a}_i\left(\bar{u}\right)\tilde{u}'\left(\bar{u},\bar{v}\right)\right)_\Omega \\
&+\left(\bar{w},\quad \tilde{v}'\left(\bar{u},\bar{v}\right)+\hat{\sigma}\left(\bar{u}\right)\cdot\tilde{u}'\left(\bar{u},\bar{v}\right)\right)_\Omega
\end{aligned} \tag{7.58}$$

where the operator $\mathcal{R}$ is the one defined in equation (7.10), but in equation (7.58) it is applied to the coarse scales. Continuing with the derivations, one may rewrite (7.58) by applying integration by parts to the first term on the right-hand side as follows

$$\begin{aligned}
\left(\frac{\partial \bar{w}}{\partial x_i},\quad \hat{\varepsilon}_{ij}\frac{\partial \tilde{u}'}{\partial x_j}\right)_\Omega &= \sum_K\left(\frac{\partial \bar{w}}{\partial x_i},\quad \hat{\varepsilon}_{ij}\frac{\partial \tilde{u}'}{\partial x_j}\right)_K \\
&= -\sum_K\left(\frac{\partial}{\partial x_j}\left[\hat{\varepsilon}_{ij}\frac{\partial \bar{w}}{\partial x_i}\right],\quad \tilde{u}'\right)_K + \sum_K\left(\frac{\partial \bar{w}}{\partial x_i},\quad \hat{\varepsilon}_{ij}n_j\tilde{u}'\right)_{\partial K}
\end{aligned} \tag{7.59}$$

which is used to obtain the following

$$\begin{aligned}
\mathcal{W}\left(\bar{u},\bar{v},\bar{w}\right)=&\sum_K\left(\bar{w},\quad \tilde{v}'\right)_K \\
&+\sum_K\left(-\frac{\partial}{\partial x_j}\left[\hat{\varepsilon}_{ij}\frac{\partial \bar{w}}{\partial x_i}\right]+\hat{a}_i\frac{\partial \bar{w}}{\partial x_i}+\hat{\sigma}\bar{w},\quad \tilde{u}'\right)_K + \sum_K\left(\frac{\partial \bar{w}}{\partial x_i},\quad \hat{\varepsilon}_{ij}n_j\tilde{u}'\right)_{\partial K}
\end{aligned} \tag{7.60}$$

Equation (7.56) along with the identities in equations (7.57) and (7.60) constitute the modified coarse scale subproblem, where $\mathcal{R}\left(\bar{u},\bar{v},\bar{w}\right)$ corresponds to the standard Galerkin method and $\mathcal{W}\left(\bar{u},\bar{v},\bar{w}\right)$ contains the VMS terms.

7.3.3 Simplifying assumptions

Some approximations can be made in order to facilitate the solution of the modified coarse scale subproblem and its algorithmic implementation in a computer code. In summary, these are as follows:

1. Neglect boundary terms in $\mathcal{W}\left(\bar{u},\bar{v},\bar{w}\right)$ from (7.60) and in $\hat{F}^K\left(w';\bar{u},\bar{v}\right)$ from (7.42). In fact, this approximation is exact when fine scales vanish at element boundaries.
2. Pull the coarse scale residual term in $\hat{F}^K\left(w';\bar{u},\bar{v}\right)$ outside of the integral, by appealing to arguments related to the mean value theorem.
3. Assume $\tilde{v}'_n = \tilde{u}'_n = 0$ for every time step (i.e., fine scales are zero at the beginning of the timestep). Alternative, assume $\tilde{v}' = 0$ (i.e., fine scales are quasi-static or reach their steady state for each timestep).

Note that these approximations affect only the VMS terms as follows

$$\mathcal{W}\left(\bar{u},\bar{v},\bar{w}\right) \approx \tilde{\mathcal{W}}\left(\bar{u},\bar{v},\bar{w}\right) = \sum_K \left(\bar{w},\quad \tilde{v}'\right)_K + \sum_K \left(-\frac{\partial}{\partial x_j}\left[\hat{\varepsilon}_{ij}\frac{\partial \bar{w}}{\partial x_i}\right] + \hat{a}_i\frac{\partial \bar{w}}{\partial x_i} + \hat{\sigma}\bar{w},\quad \tilde{u}' \right)_K \tag{7.61}$$

In addition, for the solution step for the fine scale field, the right-hand side of the linearized fine scale subproblem in (7.42) is modified as

$$\hat{F}^K\left(w';\bar{u},\bar{v}\right) \approx \tilde{F}^K\left(w';\bar{u},\bar{v}\right) = \left(w',\quad 1\right)_K \bar{R}\left(x\right) \tag{7.62}$$

Correspondingly, the fine scale forcing vector in (7.48) is modified as

$$\left[\mathbf{f}\right]_p \approx -\tilde{F}^K\left(b_p^K;\bar{u},\bar{v}\right) = -\left(b_p^K,\quad 1\right)_K \bar{R}\left(x\right) \tag{7.63}$$

Which can be rewritten as

$$\mathbf{f} \approx -\hat{\mathbf{b}}\bar{R}\left(x\right) \tag{7.64}$$

where

$$\left[\hat{\mathbf{b}}\right]_p = \int_K b_p^K dK \tag{7.65}$$

7.3.4 Backward Euler time integration

We specialize the fine scale subproblem for the case in which time discretization is carried out using the Backward Euler algorithm. Recall this may be done by setting

$$\begin{aligned} &c_0 = 1, \quad c_1 = 1, \quad c_2 = 1, \\ &\breve{v}_{n+1} = 0, \quad \breve{u}_{n+1} = u_n, \quad \breve{u}_{n+\alpha} = 0 \end{aligned} \tag{7.66}$$

which yields

$$v_{n+\alpha} = v_{n+1} = \left(u_{n+1} - u_n\right)/\Delta t, \quad u_{n+\alpha} = u_{n+1} \tag{7.67}$$

Therefore, the fine scale subproblem becomes

$$\frac{1}{\Delta t}\mathbf{M}_{n+1}\left(\hat{\mathbf{u}}_{n+1} - \hat{\mathbf{u}}_n\right) + \mathbf{A}_{n+1}\hat{\mathbf{u}}_{n+1} = \mathbf{f}_{n+1} \tag{7.68}$$

which may be rewritten as

$$\mathbf{A}^*_{n+1}\hat{\mathbf{u}}_{n+1} = \mathbf{f}^*_{n+1} \tag{7.69}$$

where

$$\mathbf{A}^*_{n+1} = \mathbf{A}_{n+1} + \frac{1}{\Delta t}\mathbf{M}_{n+1}, \qquad \mathbf{f}^*_{n+1} = \mathbf{f}_{n+1} + \frac{1}{\Delta t}\mathbf{M}_{n+1}\hat{\mathbf{u}}_n \tag{7.70}$$

$$\left[\mathbf{A}^{K*}_{n+1}\right]_{pq} = \left(\frac{\partial b^K_p}{\partial x_i}, \ \hat{\varepsilon}_{ij}\left(\bar{u}\right)\frac{\partial b^K_q}{\partial x_j}\right)_K + \left(b^K_p, \ \left[\hat{\sigma}\left(\bar{u}\right) + 1/\Delta t\right]b^K_q\right)_K \tag{7.71}$$

$$\left[\mathbf{f}^*_{n+1}\right]_p = -\left(b^K_p, \ 1\right)_K\left[\bar{R}_{n+1} - \tilde{u}'_n/\Delta t\right] \tag{7.72}$$

Therefore, the fine scales for element K can be written in terms of a so-called stabilization parameter τ^K_{n+1}.

$$\tilde{u}'_{n+1}\big|_K = -\tau^K_{n+1}\left[\bar{R}_{n+1} - \tilde{u}'_n\big|_K/\Delta t\right] \tag{7.73}$$

$$\tau^K_{n+1} = \mathbf{b}^K \cdot \left(\mathbf{A}^{K*}_{n+1}\right)^{-1}\hat{\mathbf{b}}^K \tag{7.74}$$

Substituting in the linearized coarse scale subproblem

$$\mathcal{R}_{n+1}\left(\bar{u}, \bar{v}, \bar{w}\right) + \sum_K\left(\bar{w}, \ \left[\tilde{u}'_{n+1} - \tilde{u}'_n\right]/\Delta t\right)_K + \sum_K\left(-\frac{\partial}{\partial x_j}\left[\hat{\varepsilon}_{ij}\frac{\partial \bar{w}}{\partial x_i}\right] + \hat{\sigma}\bar{w}, \ \tilde{u}'_{n+1}\right)_K = 0 \tag{7.75}$$

$$\mathcal{R}_{n+1}\left(\bar{u},\bar{v},\bar{w}\right) + \sum_K \left(-\frac{\partial}{\partial x_j}\left[\hat{\varepsilon}_{ij}\frac{\partial \bar{w}}{\partial x_i}\right] + \left[\hat{\sigma} + 1/\Delta t\right]\bar{w}, \quad -\tau^K_{n+1}\left[\bar{R}_{n+1} - \tilde{u}'_n/\Delta t\right]\right)_K - \sum_K \left(\bar{w}, \quad \tilde{u}'_n/\Delta t\right)_K = 0 \tag{7.76}$$

where $\mathcal{R}_{n+1}$ denotes the time-discrete counterpart of operator $\mathcal{R}$. Some remarks are in order:

- This final form accounts for simplifications 1 through 3 in section 7.2.2.
- For the assumption of quasi-static fine-scales, all the terms with factor $1/\Delta t$ are dropped from (7.76) as well as from (7.71).
- For fully evolving fine-scales, it is necessary to store and carry the value of $\tilde{u}'$ at the integration points from one time step to the next.
- The nonlinear equation in (7.76) is solved with an iterative method such as Newton-Raphson (NR) or a Quasi-Newton method. In order to find the next solution iterate $\bar{u}^{(i+1)}_{n+1}$ which would require a linearization around the most recent iterate $\bar{u}^{(i)}_{n+1}$. However, finding the consistent tangent term for τ^K_{n+1} is not a trivial task in the genera non-linear regime with evolving coefficients.
- For low order elements, the second order spatial derivatives appearing in (7.76) (including the one within $\bar{R}_{n+1}$) are dropped.

7.3.5 Additional considerations for coercivity analysis

The idea of pulling the coarse-scale residual out of the integral as in point 2 from section 7.2.3 is commonly applied in practice. Not doing so yields a method that is still residual based (and therefore consistent), but not in the typical form of classical stabilized methods. In this section we make some remarks that are useful for the coercivity analysis of this case. To simplify the presentation of the ideas, consider the linear, scalar, diffusion-reaction equation with no transient effects (i.e., steady state). The resulting classical stabilization term obtained from VMS would be of the following form

$$-\sum_K \left(-\varepsilon\Delta u^h + \sigma u^h, \quad \underbrace{\mathbf{b}\cdot\mathbf{A}^{-1}\hat{\mathbf{b}}}_{\tau}\left(-\varepsilon\Delta u^h + \sigma u^h\right)\right)_K \tag{7.77}$$

where

$$[\mathbf{A}]_{pq} = \left(\frac{\partial b_p^K}{\partial x_i}, \quad \varepsilon_{ij} \frac{\partial b_q^K}{\partial x_j} \right)_K + \left(b_p^K, \quad \sigma b_q^K \right)_K \tag{7.78}$$

However, if the coarse residual remains "inside" the integral, the VMS term is given by the following expression written in indicial notation

$$-\sum_K \left(-\varepsilon \Delta u^h + \sigma u^h, \quad b_i^K A_{ij}^{-1} \left(b_j^K, \quad -\varepsilon \Delta u^h + \sigma u^h \right)_K \right)_K \tag{7.79}$$

Recall that the matrix $\mathbf{A}^K$ is defined in term of some integrals over K. As a result, $\mathbf{A}^K$ is constant over K. The same applies for the other integral term in the second slot of the bilinear form. Therefore, they can be pulled out from the bilinear form to yield

$$-\sum_K \left(-\varepsilon \Delta u^h + \sigma u^h, \quad b_i^K \right)_K A_{ij}^{-1} \left(b_j^K, \quad -\varepsilon \Delta u^h + \sigma u^h \right)_K \tag{7.80}$$

In order for the coercivity results derived in [4] to hold, the following inequality must be satisfied

$$\left(-\varepsilon \Delta u^h + \sigma u^h, \quad b_i^K \right)_K A_{ij}^{-1} \left(b_j^K, \quad -\varepsilon \Delta u^h + \sigma u^h \right)_K < \frac{\beta}{\sigma} \left(\frac{\alpha^2}{\alpha^2 + C_{inv}} \right) \left\| -\varepsilon \Delta u^h + \sigma u^h \right\|_{L^2(K)}^2 \tag{7.81}$$

In order to show this, we begin by noting that $u^h\big|_K \in P_m(K)$, i.e., the trial function u^h is expanded in terms of the FE basis, which within a single element K, corresponds to polynomials of degree of at most m. Moreover, since $u^h\big|_K \in P_m(K)$, then $\Delta u^h\big|_K \in P_m(K)$ as well. Then, for constant ε and σ over the element K, one has that $-\varepsilon \Delta u^h + \sigma u^h \in P_m(K)$. Consequently, we can rewrite this expression in terms as the FE shape functions:

$$-\varepsilon \Delta u^h + \sigma u^h = a_i N_i \tag{7.82}$$

Now, we note that the norm term in the right-hand side of (7.81) is, by definition, given by

$$\left\| -\varepsilon \Delta u^h + \sigma u^h \right\|_{L^2(K)}^2 := \left(-\varepsilon \Delta u^h + \sigma u^h, \quad -\varepsilon \Delta u^h + \sigma u^h \right)_K \tag{7.83}$$

Using equations (7.82) and (7.83), we can rewrite inequality (7.81) as:

$$\left(a_p N_p, \quad b_i^K \right)_K A_{ij}^{-1} \left(b_j^K, \quad a_q N_q \right)_K < \frac{\beta}{\sigma} \left(\frac{\alpha^2}{\alpha^2 + C_{inv}} \right) \left(a_p N_p, \quad a_q N_q \right)_K \tag{7.84}$$

Which we can be further rewritten as

$$a_p \underbrace{\left[\frac{\beta}{\sigma} \left(\frac{\alpha^2}{\alpha^2 + C_{inv}} \right) \left(N_p, \quad N_q \right)_K - \left(N_p, \quad b_i^K \right)_K A_{ij}^{-1} \left(b_j^K, \quad N_q \right)_K \right]}_{M_{pq}} a_q \quad > \quad 0 \quad \forall a_i \in \mathbb{R}^{nel} \tag{7.85}$$

Therefore, the problem of proving inequality (7.81) is reduced to showing that the matrix with entries M_{pq} is positive definite. This is convenient since M_{pq} is computable, (i.e., it is defined in terms of known quantities). Then, in principle, we can compute its eigenvalues and verify that they are all positive, which would ensure positive definiteness of $\mathbf{M}$ and prove inequality (7.81).

For instance, consider the case in which K is a square 4-node quadrilateral (Q4) element with side length h. Define the entries of matrix $\mathbf{M}$ as follows

$$M_{pq} = C_1 \left(N_p, \ N_q \right)_K - \left(N_p, \ b_i^K \right)_K A_{ij}^{-1} \left(b_j^K, \ N_q \right)_K \tag{7.86}$$

where C_1 is some positive constant, N_p are the standard Q4 element shape functions, b_i^K correspond to the fine scale basis defined in [4] for Q4 elements, and $\mathbf{A}$ is as defined in (7.78). Note that $\mathbf{M}$ is a 4×4 matrix since the Q4 element has 4 shape functions N_p. The software package Mathematica is used to compute the 4 eigenvalues of (7.86), which are found to be

$$\lambda_1 = \lambda_2 = \frac{h^2}{72\sigma} \left[\frac{\alpha^2 \left(6\sigma C_1 - 5 \right) + 132\sigma C_1}{\alpha^2 + 22} \right] \tag{7.87}$$

$$\lambda_3 = \frac{h^2}{144\sigma} \left[\frac{\alpha^4 \left(36\sigma C_1 - 35 \right) + 180\alpha^2 \left(16\sigma C_1 - 15 \right) + 7200\sigma C_1}{\alpha^4 + 80\alpha^2 + 200} \right] \tag{7.88}$$

$$\lambda_4 = \frac{C_1 h^2}{36} \tag{7.89}$$

where α is the Damkohler number given by

$$\alpha^2 = \frac{\sigma h^2}{\varepsilon} \geq 0 \tag{7.90}$$

For $\alpha = 0$, all eigenvalues λ_i of $\mathbf{M}$ are positive. For $\alpha > 0$, all eigenvalues are guaranteed to be positive if[2]

$$\sigma C_1 \geq \frac{35}{36} \tag{7.91}$$

Therefore, for the coercivity results from [4] to hold, the following 2 inequalities must hold

$$\beta \left(\frac{\alpha^2}{\alpha^2 + C_{inv}} \right) \geq \frac{35}{36} \tag{7.92}$$

$$\beta < 1 \tag{7.93}$$

---

[2] This is a sufficient condition to ensure positive definiteness of **M**.

Recall that $C_{inv} \geq 0$ for Q4 elements [7]; therefore, we can choose $C_{inv} = 0$ and $\beta = 35/36$ to simultaneously satisfy inequalities (7.92) and (7.93), and the coercivity results from [4] are thus extended to the VMS method wherein the coarse scale residual is kept within the integral in the right-hand side of the fine-scale subproblem.

Remark: Extending a similar analysis for nonlinear problems requires further considerations, since the coarse scale residual may no longer be in the space of functions spanned by the element shape functions and (7.82) would no longer hold.

7.4 Application to Thermo-Chemo-Mechanical Systems

We consider the system of equations from the thermo-chemo-mechanical model in Chapter 4. Furthermore, the governing equations are extended to include a term to account for heat dissipation in the solid, and advection terms for the mass concentration and temperature fields. Consider a deforming solid domain $\Omega \in \mathbb{R}^{n_{sd}}$ at time $t \in]0,T[$, with boundary $\partial\Omega$. At time $t = 0$ the domain is taken to occupy its reference configuration $\Omega = \Omega_0$. The unknown fields are the solid displacement $\mathbf{u}$, the fluid mass concentration ρ, the temperature of the mixture θ, and the reaction state variable ϕ. It is noted that no gradients of ϕ appear in the formulation from Chapter 4, and the evolution of ϕ is given by and ODE in terms of the other unknown fields. For details regarding this evolution equation, as well as its discrete counterpart, and other constitutive equations, the interested reader is referred to Chapter 4. This section will skip over those details and focus on the derivation of a VMS method for the thermo-chemo-mechanical system. Now, we define the appropriate spaces of trial and test functions for the weak form of the problem:

$$\mathcal{S}_t^{\mathbf{u}} = \left\{ \mathbf{u}(\mathbf{x},t) \mid \mathbf{u} \in H^1(\Omega),\ u_i = u_i^D \text{ on } \partial\Omega_{u_i}^D \right\} \tag{7.94}$$

$$\mathcal{S}_t^{\rho} = \left\{ \rho(\mathbf{x},t) \mid \rho \in H^1(\Omega),\ \rho = \rho^D \text{ on } \partial\Omega_{\rho}^D \right\} \tag{7.95}$$

$$\mathcal{S}_t^{\theta} = \left\{ \theta(\mathbf{x},t) \mid \theta \in H^1(\Omega),\ \theta = \theta^D \text{ on } \partial\Omega_{\theta}^D \right\} \tag{7.96}$$

$$\mathcal{V}^{\mathbf{u}} = \left\{ \hat{\mathbf{w}}(\mathbf{x}) \mid \hat{\mathbf{w}} \in H^1(\Omega),\ \hat{w}_i = 0 \text{ on } \partial\Omega_{u_i}^D \right\} \tag{7.97}$$

$$\mathcal{V}^{\rho} = \left\{ \hat{g}(\mathbf{x}) \mid \hat{g} \in H^1(\Omega),\ \hat{g} = 0 \text{ on } \partial\Omega_{\rho}^D \right\} \tag{7.98}$$

$$\mathcal{V}^{\theta} = \left\{ \hat{\varphi}(\mathbf{x}) \mid \hat{\varphi} \in H^1(\Omega),\ \hat{\varphi} = 0 \text{ on } \partial\Omega_{\theta}^D \right\} \tag{7.99}$$

The system of equations is written in weak form as follows: Find $\{\mathbf{u}, \rho^f, \theta\} \in \mathcal{S}_t^{\mathbf{u}} \times \mathcal{S}_t^{\rho} \times \mathcal{S}_t^{\theta}$, such that $\forall \{\hat{\mathbf{w}}, \hat{g}, \hat{\varphi}\} \in \mathcal{V}^{\mathbf{u}} \times \mathcal{V}^{\rho} \times \mathcal{V}^{\theta}$ the following system of equations holds

$$R_{\mathbf{u}}\left(\hat{\mathbf{w}}\ ;\ \mathbf{u},\ \rho^f,\ \theta;\ \phi\right) = \left(\frac{\partial \hat{w}_i}{\partial x_j},\ T_{ij}^s\right)_\Omega - \left(\hat{w}_i,\ h_i\right)_{\partial\Omega_{u_i}^N} = 0 \tag{7.100}$$

$$\begin{aligned} R_\rho\left(\hat{g}\ ;\ \mathbf{u},\ \rho^f,\ \theta;\ \phi\right) = &\left(\hat{g},\ \frac{\partial \rho^f}{\partial t} + v_i^\rho \frac{\partial \rho^f}{\partial x_i} + \rho^f \operatorname{div}\left(\mathbf{v}^s\right) - m^f\right)_\Omega \\ &+ \left(\frac{\partial \hat{g}}{\partial x_i},\ D_{ij}\frac{\partial \rho^f}{\partial x_j}\right)_\Omega \\ &- \left(\hat{g},\ j_N\right)_{\partial\Omega_\rho^N} - \left(\hat{g},\ c_\rho\left(\rho_{sat}^f - \rho^f\right)\right)_{\partial\Omega_\rho^R} = 0 \end{aligned} \tag{7.101}$$

$$\begin{aligned} R_\theta\left(\hat{\varphi}\ ;\ \mathbf{u},\ \rho^f,\ \theta;\ \phi\right) = &\left(\hat{\varphi},\ J^{-1} c_0^s(\phi)\frac{D^s\theta}{Dt} + v_i^\theta \frac{\partial \theta}{\partial x_i} - h_{rxn} + s_\theta\left(\theta - \theta_R\right)\right)_\Omega \\ &- \left(\frac{\partial \hat{\varphi}}{\partial x_i},\ q_i\right)_\Omega + \left(\hat{\varphi},\ q_N\right)_{\partial\Omega_\theta^N} + \left(\hat{\varphi},\ h_\theta\left(\theta_R - \theta\right)\right)_{\partial\Omega_\theta^R} = 0 \end{aligned} \tag{7.102}$$

where

$$h_{rxn} = \overline{h} m^f \tag{7.103}$$

We remark the following:

- For a chemical reaction in which the fluid constituent is consumed, and where the divergence of solid velocity is negligible in comparison, the coefficient of linearized zeroth order term is positive. Thus, the governing equation is a transient diffusion-advection-reaction equation.
- For a chemical reaction that is temperature dependent and produces heat, the coefficient of the zeroth order term in the linearized temperature equation may be negative. In that case, and depending on its magnitude, the equation may no longer have the form of a classical transient diffusion-advection-reaction equation. Instead, it has the character of a transient Helmholtz equation with propagating waves.
- As discussed in [2], for propagating waves, and as the magnitude of the coefficient of the zeroth order terms becomes larger, the dominant source of error becomes the element residual.

To develop a VMS method for the system in (7.100)-(7.102), we consider an additive decomposition of the trial functions for the fluid mass concentration field and the temperature field, as well as their corresponding test functions, into linearly independent coarse scale ($\bar{u}$) and fine scale ($\tilde{u}$) contributions.

$$\rho^f = \bar{\rho}^f + \tilde{\rho}^f \tag{7.104}$$

$$\theta = \bar{\theta} + \tilde{\theta} \tag{7.105}$$

$$\hat{g} = \bar{g} + \tilde{g} \tag{7.106}$$

$$\hat{\varphi} = \bar{\varphi} + \tilde{\varphi} \tag{7.107}$$

where the coarse scale contributions are associated to the standard FEM basis functions, i.e., the coarse scale trial and test functions belong to finite-dimensional spaces $\{\bar{\rho}^f, \bar{\theta}\} \in \bar{\mathcal{S}}_t^\rho \times \bar{\mathcal{S}}_t^\theta \subset \mathcal{S}_t^\rho \times \mathcal{S}_t^\theta$ and $\{\bar{g}, \bar{\varphi}\} \in \overline{\mathcal{V}}^\rho \times \overline{\mathcal{V}}^\theta \subset \mathcal{V}^\rho \times \mathcal{V}^\theta$, consisting of basis function with compact support defined over a discretization of the domain Ω into non-overlapping elements. The fine scale trial and test functions are defined in terms of fine-scale bases local to the element as discussed in the following section. Using the linearity of the test function slot and linear independence of the scales, the weak form may be segregated into coarse scale subproblem and a fine scale subproblem, which are discussed in the following subsections.

7.4.1 Fine scale subproblem

The fine scale subproblems for the fluid balance of mass and the temperature equation are localized to each element. In addition, the fine scales are assumed to vanish at the Neumann and Robin boundaries. The resulting coupled fine scale subproblem is given as follows

$$\begin{aligned} &R_\rho^K\left(\tilde{g}\ ;\ \mathbf{u},\ \rho^f,\ \theta;\ \phi\right) \\ &= \left(\tilde{g},\ \ \frac{\partial \rho^f}{\partial t} + v_i^\rho \frac{\partial \rho^f}{\partial x_i} + \rho^f \operatorname{div}\left(\mathbf{v}^s\right) - m^f\right)_K + \left(\frac{\partial \tilde{g}}{\partial x_i},\ \ D_{ij}\frac{\partial \rho^f}{\partial x_j}\right)_K \end{aligned} \tag{7.108}$$

$$\begin{aligned} &R_\theta^K\left(\tilde{\varphi}\ ;\ \mathbf{u},\ \rho^f,\ \theta;\ \phi\right) \\ &= \left(\tilde{\varphi},\ \ J^{-1} c_0^s(\phi) \frac{D^s \theta}{Dt} + v_i^\theta \frac{\partial \theta}{\partial x_i} - h_{rxn} + s_\theta\left(\theta - \theta_R\right)\right)_K - \left(\frac{\partial \tilde{\varphi}}{\partial x_i},\ \ q_i\right)_K \end{aligned} \tag{7.109}$$

The linearized equations are given by

$$R_\rho^K\left(\tilde{g}\ ;\ \mathbf{u},\ \bar{\rho}^f,\ \bar{\theta};\ \phi\right) + \mathcal{D}_\rho R_\rho^K + \mathcal{D}_\theta R_\rho^K = 0 \tag{7.110}$$

$$R_\theta^K\left(\tilde{\varphi}\ ;\ \mathbf{u},\ \rho^f,\ \theta;\ \phi\right)+\mathcal{D}_\rho R_\theta^K+\mathcal{D}_\theta R_\theta^K=0 \tag{7.111}$$

where the linearized terms are obtained via the variational derivative defined as

$$\mathcal{D}_\rho R_\rho^K\left(\tilde{\varphi}\ ;\ \mathbf{u},\ \rho^f,\ \theta;\ \phi;\delta\rho^f\right)=\left[\frac{d}{d\varepsilon}R_\rho^K\left(\tilde{\varphi}\ ;\ \mathbf{u},\ \rho^f+\varepsilon\delta\rho^f,\ \theta;\ \phi\right)\right]_{\varepsilon=0} \tag{7.112}$$

$$\mathcal{D}_\rho R_\theta^K\left(\tilde{\varphi}\ ;\ \mathbf{u},\ \rho^f,\ \theta;\ \phi;\delta\rho^f\right)=\left[\frac{d}{d\varepsilon}R_\theta^K\left(\tilde{\varphi}\ ;\ \mathbf{u},\ \rho^f+\varepsilon\delta\rho^f,\ \theta;\ \phi\right)\right]_{\varepsilon=0} \tag{7.113}$$

$$\mathcal{D}_\theta R_\theta^K\left(\tilde{\varphi}\ ;\ \mathbf{u},\ \rho^f,\ \theta;\ \phi;\delta\theta\right)=\left[\frac{d}{d\varepsilon}R_\theta^K\left(\tilde{\varphi}\ ;\ \mathbf{u},\ \rho^f,\ \theta+\varepsilon\delta\theta;\ \phi\right)\right]_{\varepsilon=0} \tag{7.114}$$

$$\mathcal{D}_\theta R_\rho^K\left(\tilde{\varphi}\ ;\ \mathbf{u},\ \rho^f,\ \theta;\ \phi;\delta\theta\right)=\left[\frac{d}{d\varepsilon}R_\rho^K\left(\tilde{\varphi}\ ;\ \mathbf{u},\ \rho^f,\ \theta+\varepsilon\delta\theta;\ \phi\right)\right]_{\varepsilon=0} \tag{7.115}$$

which yield

$$\begin{aligned}\mathcal{D}_\rho R_\rho^K&=\left(\tilde{g},\ \frac{\partial\tilde{\rho}^f}{\partial t}\right)_K+\left(\frac{\partial\tilde{g}}{\partial x_i},\ D_{ij}\frac{\partial\tilde{\rho}^f}{\partial x_j}\right)_K\\&\quad+\left(\tilde{g},\ v_i^\rho\frac{\partial\tilde{\rho}^f}{\partial x_i}\right)_K+\left(\tilde{g},\ \left[\operatorname{div}\left(\mathbf{v}^s\right)-\frac{\partial\hat{m}^f}{\partial\rho}\right]\tilde{\rho}^f\right)_K\end{aligned} \tag{7.116}$$

$$\mathcal{D}_\theta R_\rho^K=\left(\tilde{g},\ -\frac{\partial\hat{m}^f}{\partial\theta}\tilde{\theta}\right)_K+\left(\frac{\partial\tilde{g}}{\partial x_i},\ \frac{\partial D_{ij}}{\partial\theta}\frac{\partial\rho^f}{\partial x_j}\tilde{\theta}\right)_K \tag{7.117}$$

$$\begin{aligned}\mathcal{D}_\theta R_\theta^K=&\left(\tilde{\varphi},\ J^{-1}c_0^s(\phi)\frac{D^s\tilde{\theta}}{Dt}\right)_K+\left(\frac{\partial\tilde{\varphi}}{\partial x_i},\ \kappa\frac{\partial\tilde{\theta}}{\partial x_i}\right)_K\\&+\left(\tilde{\varphi},\ v_i^\theta\frac{\partial\tilde{\theta}}{\partial x_i}\right)_K+\left(\tilde{\varphi},\ \left[s_\theta-\frac{\partial\hat{h}}{\partial\theta}\right]\tilde{\theta}\right)_K\end{aligned} \tag{7.118}$$

$$\mathcal{D}_\rho R_\theta^K=-\left(\tilde{\varphi},\ \frac{\partial\hat{h}_{rxn}}{\partial\rho}\tilde{\rho}^f\right) \tag{7.119}$$

The coarse scale residual in the fine scale subproblem is written as

$$\begin{aligned}&R_\rho^K\left(\tilde{g}\ ;\ \mathbf{u},\ \bar{\rho}^f,\ \bar{\theta};\ \phi\right)\\&=\left(\tilde{g},\ \frac{\partial\bar{\rho}^f}{\partial t}+v_i^\rho\frac{\partial\bar{\rho}^f}{\partial x_i}+\bar{\rho}^f\operatorname{div}\left(\mathbf{v}^s\right)-\bar{m}^f\right)_K+\left(\frac{\partial\tilde{g}}{\partial x_i},\ \bar{D}_{ij}\frac{\partial\bar{\rho}^f}{\partial x_j}\right)_K\\&=\left(\tilde{g},\ \bar{R}_\rho\right)_K+\left(\tilde{g},\ n_i^K\bar{D}_{ij}\frac{\partial\bar{\rho}^f}{\partial x_j}\right)_{\partial K}\end{aligned} \tag{7.120}$$

$$
\begin{aligned}
&R_\theta^K\left(\tilde{\varphi}\ ;\ \mathbf{u},\ \bar{\rho}^f,\ \bar{\theta};\ \phi\right)\\
&=\left(\tilde{\varphi},\quad J^{-1}c_0^s(\phi)\frac{D^s\bar{\theta}}{Dt}+v_i^\theta\,\frac{\partial\bar{\theta}}{\partial x_i}-\bar{h}_{rxn}+s_\theta\left(\bar{\theta}-\theta_R\right)\right)_K-\left(\frac{\partial\tilde{\varphi}}{\partial x_i},\quad \bar{q}_i\right)_K\\
&=\left(\tilde{\varphi},\quad \bar{R}_\theta\right)_K-\left(\tilde{\varphi},\quad n_i^K\bar{q}_i\right)_{\partial K}
\end{aligned}
\tag{7.121}
$$

where

$$
\bar{R}_\rho=\frac{\partial\bar{\rho}^f}{\partial t}-\frac{\partial}{\partial x_i}\left(\bar{D}_{ij}\,\frac{\partial\bar{\rho}^f}{\partial x_j}\right)+v_i^\rho\,\frac{\partial\bar{\rho}^f}{\partial x_i}+\bar{\rho}^f\operatorname{div}\left(\mathbf{v}^s\right)-\bar{m}^f
\tag{7.122}
$$

$$
\bar{R}_\theta=J^{-1}c_0^s(\phi)\frac{D^s\bar{\theta}}{Dt}+\frac{\partial\bar{q}_i}{\partial x_i}+v_i^\theta\,\frac{\partial\bar{\theta}}{\partial x_i}-\bar{h}_{rxn}+s_\theta\left(\bar{\theta}-\theta_R\right)
\tag{7.123}
$$

Utilizing the same finite-dimensional basis for the fine-scale test and trial functions for both the fine-scale fluid mass concentration and the fine-scale temperature fields, $b_p^{K,\rho}$ and $b_r^{K,\theta}$ the element level fine scale problems have the form

$$
\begin{bmatrix}\mathbf{M}_\rho & \mathbf{0}\\ \mathbf{0} & \mathbf{M}_\theta\end{bmatrix}\begin{bmatrix}\dot{\boldsymbol{\rho}}^K\\ \dot{\boldsymbol{\theta}}^K\end{bmatrix}+\begin{bmatrix}\mathbf{A}_{\rho\rho} & \mathbf{A}_{\rho\theta}\\ \mathbf{A}_{\theta\rho} & \mathbf{A}_{\theta\theta}\end{bmatrix}\begin{bmatrix}\boldsymbol{\rho}^K\\ \boldsymbol{\theta}^K\end{bmatrix}=\begin{bmatrix}\mathbf{F}_\rho^K\\ \mathbf{F}_\theta^K\end{bmatrix}
\tag{7.124}
$$

Where

$$
\left[\mathbf{F}_\rho^K\right]_p=-R_\rho^K\left(b_p^{K,\rho};\ \mathbf{u},\ \bar{\rho}^f,\ \bar{\theta};\ \phi\right)
\tag{7.125}
$$

$$
\left[\mathbf{F}_\theta^K\right]_r=-R_\theta^K\left(b_r^{K,\theta};\ \mathbf{u},\ \bar{\rho}^f,\ \bar{\theta};\ \phi\right)
\tag{7.126}
$$

$$
\left[\mathbf{M}_\rho\right]_{pq}=\left(b_p^{K,\rho},\quad b_q^{K,\rho}\right)_K
\tag{7.127}
$$

$$
\left[\mathbf{M}_\theta\right]_{rs}=\left(b_r^{K,\theta},\quad J^{-1}c_0^s(\phi)b_s^{K,\theta}\right)_K
\tag{7.128}
$$

$$
\begin{aligned}
&\left[\mathbf{A}_{\rho\rho}\right]_{pq}\\
&=\left(\frac{\partial b_p^{K,\rho}}{\partial x_i},\quad D_{ij}\,\frac{\partial b_q^{K,\rho}}{\partial x_j}\right)_K+\left(b_p^{K,\rho},\quad v_i^\rho\,\frac{\partial b_q^{K,\rho}}{\partial x_i}\right)_K+\left(b_p^{K,\rho},\quad \left[\operatorname{div}\left(\mathbf{v}^s\right)-\frac{\partial\hat{m}^f}{\partial\rho}\right]b_q^{K,\rho}\right)_K
\end{aligned}
\tag{7.129}
$$

$$
\begin{aligned}
&\left[\mathbf{A}_{\theta\theta}\right]_{rs}\\
&=\left(\frac{\partial b_r^{K,\theta}}{\partial x_i},\quad \kappa_{ij}\,\frac{\partial b_s^{K,\theta}}{\partial x_j}\right)_K+\left(b_r^{K,\theta},\quad v_i^\theta\,\frac{\partial b_s^{K,\theta}}{\partial x_i}\right)_K+\left(b_r^{K,\theta},\quad \left[s_\theta-\frac{\partial\hat{h}}{\partial\theta}\right]b_s^{K,\theta}\right)_K
\end{aligned}
\tag{7.130}
$$

$$\begin{aligned}\left[\mathbf{A}_{\rho\theta}\right]_{ps} &= \mathcal{D}_{\theta} R_{\rho}^{K}\left(b_{p}^{K,\rho};\ \mathbf{u},\ \bar{\rho}^{f},\ \bar{\theta};\ \phi; b_{s}^{K,\theta}\right) \\ &= \left(b_{p}^{K,\rho},\ \ -\frac{\partial \hat{m}^{f}}{\partial \theta} b_{s}^{K,\theta}\right)_{K} + \left(\frac{\partial b_{p}^{K,\rho}}{\partial x_{i}},\ \ \frac{\partial \bar{D}_{ij}}{\partial \theta} \frac{\partial \bar{\rho}^{f}}{\partial x_{j}} b_{s}^{K,\theta}\right)_{K}\end{aligned} \tag{7.131}$$

$$\begin{aligned}\left[\mathbf{A}_{\theta\rho}\right]_{rp} &= \mathcal{D}_{\rho} R_{\theta}^{K}\left(b_{r}^{K,\theta};\ \mathbf{u},\ \bar{\rho}^{f},\ \bar{\theta};\ \phi; b_{q}^{K,\rho}\right) \\ &= -\left(b_{r}^{K,\theta},\ \ \frac{\partial \hat{h}_{rxn}}{\partial \rho} b_{q}^{K,\rho}\right)_{K}\end{aligned} \tag{7.132}$$

Neglecting the off-diagonal terms $\mathbf{A}_{\rho\theta}$ and $\mathbf{A}_{\theta\rho}$, enables decoupled solution of the fine-scale coefficients for the fine-scale fluid concentration and fine-scale temperature fields. In that case, the fine-scale subproblem for each fine-scale field takes the general form discussed in section 7.3. Furthermore, one may modify the fine scale subproblem in (7.124) according to the simplifying assumptions discussed in section 7.3.3.

7.4.2 Modified coarse scale subproblem

In order to embed the fine scale field for the density and the temperature, the coarse scale problems for the temperature equation and the fluid balance of mass equation are linearized as follows

$$0 = R_{\rho}\left(\hat{\varphi}\ ;\ \mathbf{u},\ \rho^{f},\ \theta;\ \phi\right) \approx R_{\rho}\left(\hat{\varphi}\ ;\ \mathbf{u},\ \bar{\rho}^{f},\ \bar{\theta};\ \phi\right) + \mathcal{D}_{\rho} R_{\rho} + \mathcal{D}_{\theta} R_{\rho} \tag{7.133}$$

$$0 = R_{\theta}\left(\hat{\varphi}\ ;\ \mathbf{u},\ \rho^{f},\ \theta;\ \phi\right) \approx R_{\theta}\left(\hat{\varphi}\ ;\ \mathbf{u},\ \bar{\rho}^{f},\ \bar{\theta};\ \phi\right) + \mathcal{D}_{\theta} R_{\theta} + \mathcal{D}_{\rho} R_{\theta} \tag{7.134}$$

The functional derivatives are computed, and the arbitrary variations of each field are substituted with the corresponding fine-scale field, which is taken to vanish at the domain boundary

$$\begin{aligned}\mathcal{D}_{\rho} R_{\rho} &= \sum_{K}\left(\hat{g},\ \ \frac{\partial \tilde{\rho}^{f}}{\partial t}\right)_{K} + \sum_{K}\left(\mathcal{L}_{\rho}^{*}\left(\hat{g}\right),\ \ \tilde{\rho}^{f}\right)_{K} \\ &+ \sum_{K}\left(n_{i}^{K} v_{i}^{\rho} \hat{g},\ \ \tilde{\rho}^{f}\right)_{\partial K} + \sum_{K}\left(n_{i}^{K} D_{ji} \frac{\partial \hat{g}}{\partial x_{j}},\ \ \tilde{\rho}^{f}\right)_{\partial K}\end{aligned} \tag{7.135}$$

$$\mathcal{D}_{\theta} R_{\rho} = \left(\hat{g},\ \ -\frac{\partial \hat{m}^{f}}{\partial \theta} \tilde{\theta}\right)_{\Omega} + \left(\frac{\partial \hat{g}}{\partial x_{i}},\ \ \frac{\partial D_{ij}}{\partial \theta} \frac{\partial \bar{\rho}^{f}}{\partial x_{j}} \tilde{\theta}\right)_{\Omega} \tag{7.136}$$

$$\mathcal{D}_\theta R_\theta = \sum_K \left(\hat{\varphi},\ \ J^{-1} c_0^s(\phi) \frac{D^s \tilde{\theta}}{Dt} \right)_K + \sum_K \left(\mathcal{L}_\theta^*(\hat{\varphi}),\ \ \tilde{\theta} \right)_K$$
$$+ \sum_K \left(n_i^K v_i^\theta \hat{\varphi},\ \ \tilde{\theta} \right)_{\partial K} + \sum_K \left(n_i^K \kappa_{ji} \frac{\partial \hat{\varphi}}{\partial x_j},\ \ \tilde{\theta} \right)_{\partial K} \tag{7.137}$$

$$\mathcal{D}_\rho R_\theta = -\left(\hat{\varphi},\ \ \frac{\partial \hat{h}_{rxn}}{\partial \rho} \tilde{\rho}^f \right) \tag{7.138}$$

where

$$\mathcal{L}_\rho^*(\hat{g}) = -\frac{\partial}{\partial x_j}\left(D_{ij} \frac{\partial \hat{g}}{\partial x_i} \right) - \frac{\partial}{\partial x_i}\left(v_i^\rho \hat{g} \right) + \left(\operatorname{div}\left(\mathbf{v}^s \right) - \frac{\partial \hat{m}^f}{\partial \rho} \right) \hat{g} \tag{7.139}$$

$$\mathcal{L}_\theta^*(\hat{\varphi}) = -\frac{\partial}{\partial x_j}\left(\kappa_{ij} \frac{\partial \hat{\varphi}}{\partial x_i} \right) - \frac{\partial}{\partial x_i}\left(v_i^\theta \hat{\varphi} \right) + \left(s_\theta - \frac{\partial \hat{h}_{rxn}}{\partial \theta} \right) \hat{\varphi} \tag{7.140}$$

In addition, for the case when diffusivity dependence on temperature is negligible

$$\mathcal{D}_\theta R_\rho = \left(\hat{g},\ \ -\frac{\partial \hat{m}^f}{\partial \theta} \tilde{\theta} \right)_\Omega \tag{7.141}$$

In order to pose the discrete problem, we first use a superscript h to denote the discrete spaces, which are finite dimensional subsets of the solution and test function spaces, and are taken to be the standard finite element spaces $\mathcal{S}_t^{\mathbf{u},h} \subset \mathcal{S}_t^{\mathbf{u}}$, $\mathcal{S}_t^{\rho,h} \subset \mathcal{S}_t^{\rho}$, $\mathcal{S}_t^{\theta,h} \subset \mathcal{S}_t^{\theta}$, $\mathcal{V}^{\mathbf{u},h} \subset \mathcal{V}^{\mathbf{u}}$, $\mathcal{V}^{\rho,h} \subset \mathcal{V}^{\rho}$, and $\mathcal{V}^{\theta,h} \subset \mathcal{V}^{\theta}$. In particular, it is noted that the coarse scale spaces already correspond to the finite element spaces: $\bar{\mathcal{S}}_t^{\rho} = \mathcal{S}_t^{\rho,h}$, $\bar{\mathcal{S}}_t^{\theta} = \mathcal{S}_t^{\theta,h}$, $\bar{\mathcal{V}}^{\rho} = \mathcal{V}^{\rho,h}$, and $\bar{\mathcal{V}}^{\theta} = \mathcal{V}^{\theta,h}$. The statement of the discrete problem is: Given appropriate initial and boundary conditions, find $\{\mathbf{u}^h, \rho^{f,h}, \theta^h\} \in \mathcal{S}_t^{\mathbf{u},h} \times \mathcal{S}_t^{\rho,h} \times \mathcal{S}_t^{\theta,h}$, such that $\forall \{\hat{\mathbf{w}}^h, \hat{g}^h, \hat{\varphi}^h\} \in \mathcal{V}^{\mathbf{u},h} \times \mathcal{V}^{\rho,h} \times \mathcal{V}^{\theta,h}$ the system of equations posed by replacing the unknown fields with their discrete counterparts into (7.100), (7.133), (7.134) holds.

Remark: The final system of equations is such that an updated Lagrangian formulation is used for the solid constituent: a Lagrangian mesh where nodes are identified with material points that move with the deformation, and although the gradients and integrals are written over the deformed configuration, the reference coordinates are the independent coordinate system. On the other hand, the fluid constituent is treated through an Arbitrary Lagrangian-Eulerian (ALE) formulation, wherein the fluid transport occurs over the deforming solid domain. For a further discussion of

Lagrangian meshes (both “total” and “updated” formulations) as well as of the ALE formulation, the interested reader is referred to [8].

7.5 Numerical Results

In this section, numerical test cases are presented to highlight the stability and accuracy of the numerical method. The problems under consideration involve advancing reaction fronts that traverse the solid domain. First, we consider the so-called “bird’s beak” problem that arises during thermal oxidation step in the manufacturing of semiconductor devices. Secondly, we study the development of fingering phenomena related to problems in solid fuel combustion.

Solution of the nonlinear system employs a quasi-Newton method in which approximations to the unsymmetric tangent are updated through Broyden’s method as implemented in the finite element software package FEAP (Finite Element Analysis Program). To improve the conditioning of the discrete system of equations, the degrees of freedom are the normalized values of the fields in the thermo-chemo-mechanical method by some reference value of each corresponding field. In addition, the residual of each equation is also normalized by a reference value. In this section, a superimposed bar denotes a normalized value of the field unless otherwise specified.

The method was implemented for 3D trilinear hexahedral elements (HEX8). Nevertheless, numerical results explore 2D response, for which the domain is defined using a layer of a single element through the thickness in the third spatial dimension. The out of plane displacements are constrained (analogous to plane strain conditions), while periodic boundary conditions are applied for the other fields to “tie” the top and bottom faces of the 3D domain.

7.5.1 Bird’s beak problem

In this test case we revisit the so-called “bird’s beak” problem, which was also treated in [9,10]. Here, the formulation derived in section 7.4 is employed. The multiscale decomposition is applied only to the trial function for the fluid mass concentration field and its associated test function. The test case presented here differs from the one in [9] in several regards:

- The current formulation includes non-isothermal effects through the solution of the balance of energy in temperature form.

- The mathematical model is the reduced model from Chapter 4, wherein the kinematics of the deformation are captured through a multiplicative split of the deformation gradient into an inelastic and elastic contributions.
- In addition, the reduced formulation from Chapter 4 assumes a fluid species that can be modeled as an ideal gas with a slow speed of diffusion. As a result, the fluid velocity field needs not be solved through a PDE, which reduces the number of degrees of freedom (and associated computational cost) in the numerical method.
- Robin boundary conditions, instead of time dependent Dirichlet BCs, are applied on the exposed surface, which constitutes a more realistic approximation of the dissolution process of ambient oxygen into the solid substrate.
- The material parameters employed in the simulation correspond to those determined in Chapter 4 for which there is a good agreement with experimental data for the 1D response therein.
- The VMS method uses a fully evolving fine-scale field and the coarse scale residual remains within the integral on the right-hand side of the fine scale subproblem.

The computational domain is shown in Figure 7.1 and there are two main parts to it. The silicon substrate, and a silicon nitride mask that partially covers the surface of the substrate. The substrate is discretized with a uniform grid of 20×24 elements. For the mask, there is a thin 10 nm layer of a single element thick and 12 elements along the horizontal direction, while the rest of the mask block is discretized with a uniform grid of 3×12 elements. For this numerical study, no pad oxide layer was considered. The boundary conditions are as follows. The solid displacement is set to zero in the normal direction along the left, right, and bottom. For the fluid mass concentration field, homogeneous Dirichlet boundary condition are applied along the boundaries of the nitrate mask, Robin boundary conditions are applied along the top boundary of the Si substrate, and no fluid mass outflow is imposed along the left, right, and bottom boundaries of the domain. The diffusivity of oxygen in the silicon nitride mask is much less than that in the silicon substrate. As a simplification of the model consistent with that observation, the fluid mass concentration degrees of freedom are constrained to be zero within the mask subdomain. No heat outflux is imposed along the left, right, and bottom boundaries. Dirichlet boundary conditions of $\bar{\theta} = 1$ for the temperature field are applied in the remaining parts of the boundary. Time integration is carried out via the backward Euler algorithm, using 50 timesteps with $\Delta t = 100$ s, followed by 500 timesteps with $\Delta t = 1000$ s.

Figure 7.2 shows snapshots of the distribution of the reaction state field and the location of the advancing reaction front in the deformed domain at several times during solution procedure. It is noted that the solid silicon substrate undergoes large swelling deformation due to chemical reaction. Moreover, the nitride mask is deformed by the swelling substrate underneath it. The reaction front (i.e., the variation in the reaction state field) is captured across the length of roughly a single element.

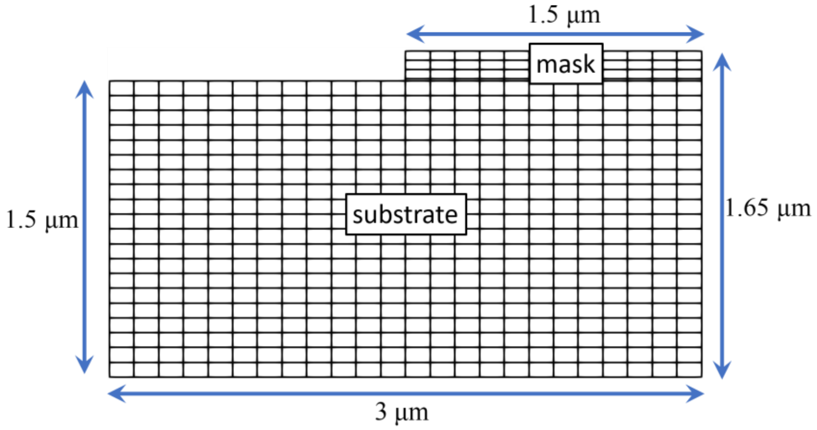


Figure 7.1. Schematic diagram of the computational domain for the bird's beak problem.

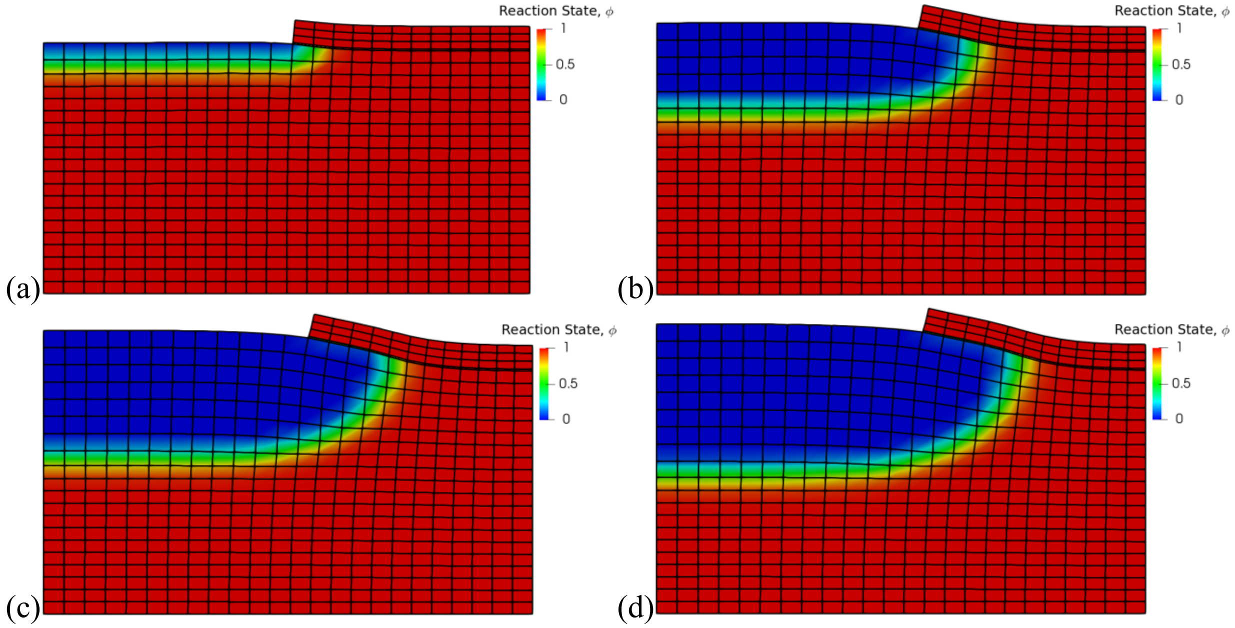


Figure 7.2. Snapshots showing the distribution of the reaction state field and the location of the advancing reaction front in the deformed domain at timesteps number (a) 50 $(t = 5\times10^3\ \text{s})$, (b) 200 $(t = 1.55\times10^5\ \text{s})$, (c) 400 $(t = 3.55\times10^5\ \text{s})$, and (d) 550 $(t = 5.05\times10^5\ \text{s})$.

The numerical results in Figure 7.3c and Figure 7.3d show that the magnitude of the coarse scale residual and fine-scale fields (respectively) is the largest around the advancing front, highlighting the role of the VMS formulation in capturing the solution in this region. In addition, distribution of the Von Mises stress is shown in Figure 7.3a, and it is the larger in the already reacted solid oxide region, due to the chemical swelling being constraint my the boundary conditions and adjacent unreacted material. Furthermore, there are also large stress values developed in the mask due to the deformation it undergoes as a result of the swelling substrate underneath it. This large stress may lead to failure of the mask and lead to complications in the its removal stage which results in imperfections and lower quality of the manufactured semiconductor device.

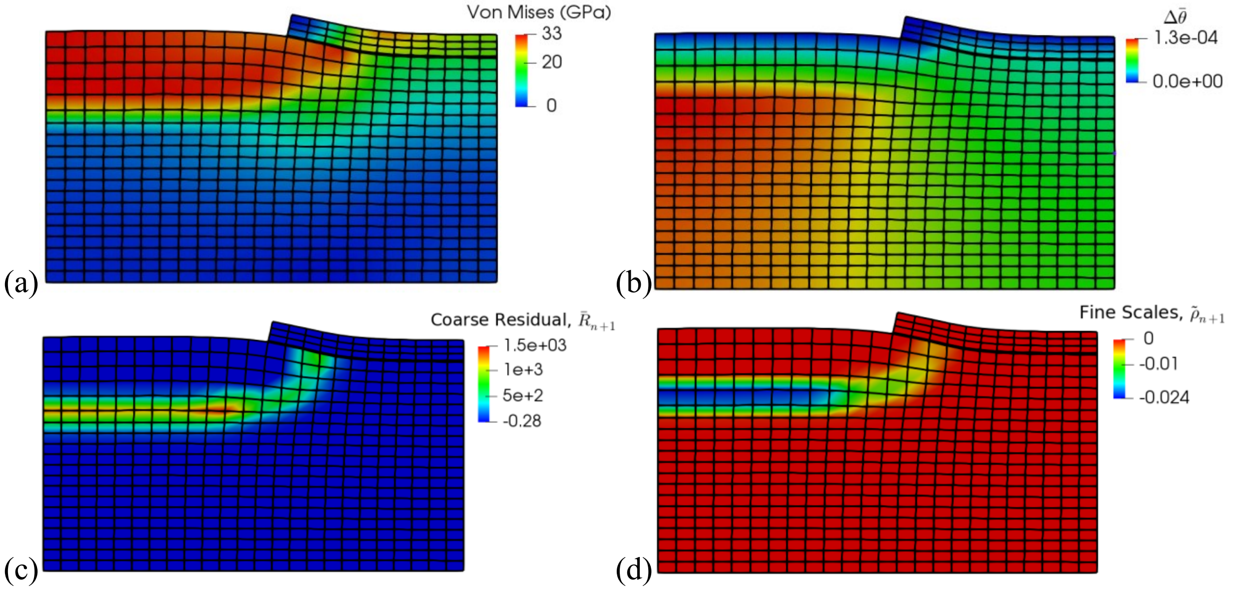


Figure 7.3. Distribution of several fields over the domain at timestep 200 or $t = 1.55 \times 10^5$ s: (a) Von Misses stress, (b) change in normalized temperature, (c) coarse scale residual, and (d) magnitude of the fine scales.

Using the values for change in enthalpy of reaction for the oxidation of silicon into silicon carbide, there is no change in the temperature field to within 5 significant figures. Therefore, the numerical results shown correspond to a case in which the volumetric heat capacity and heat conductivity where reduced by 4 orders of magnitude while the change in enthalpy of reaction was increased by 4 orders of magnitude as well. Even with these modified parameters, the normalized change in the temperature field shown in Figure 7.3b is minimal. Therefore, prior modeling of the problem as an isothermal process is a justifiable assumption.

Figure 7.4 shows the evolution of the normalized fluid mass concentration at the exposed surface of the silicon substrate. Robin boundary conditions were applied for the fluid mass concentration at this boundary, which are more realistic in capturing the dependence of the dissolution rate at the surface on the current concentration of the solute in the bulk of the solid. It is noted that the fluid reaches more than 90% of its saturation concentration by the 25 min mark, and close to 99% of the saturation concentration at the 50 min mark.

Figure 7.5 shows the total vertical displacement at the left end of the exposed surface of the silicon substrate, as well as the vertical displacement at the bottom-left corner of the nitride mask, which is an indicator of the magnitude of the deformation that the mask undergoes. It is noted that there is a period of rapid growth at the beginning of the thermal oxidation process, followed by a tapering of the rate of growth to reach a quasi-linear rate.

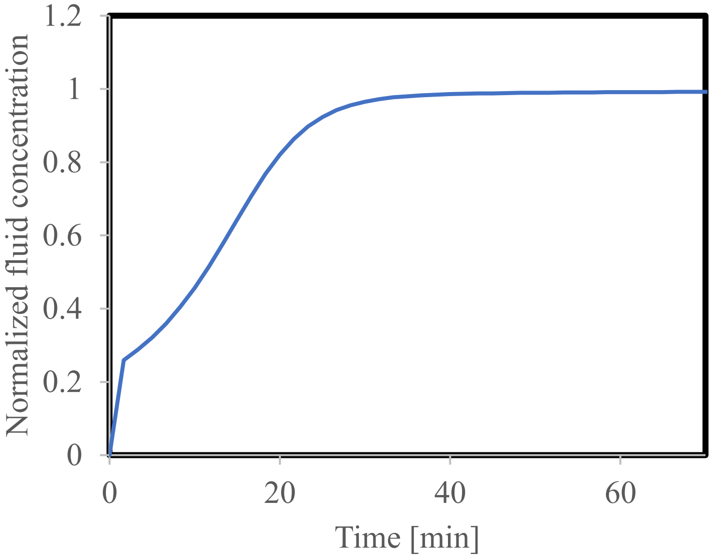


Figure 7.4. Evolution of the normalized fluid concentration at the exposed surface of the substrate.

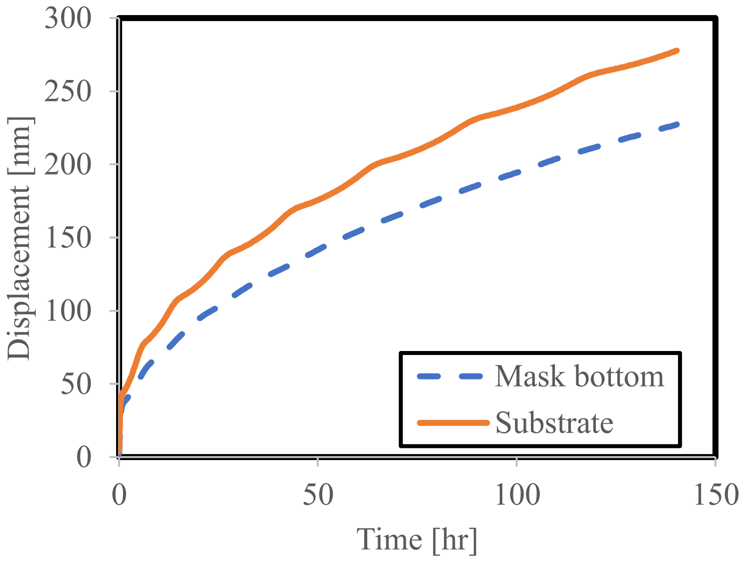


Figure 7.5. Evolution of the vertical displacement at the top surface of the substrate (as measured at its top-left corner), and at the bottom-left corner of the nitride mask.

The effect of the mesh on the results is studied by comparing the results with those obtained with a finer discretization of the domain. The finer mesh utilizes a uniform gride of 40×48 elements in the Si substrate block. We compare the evolution of the reaction zone thickness (or interphase[3] layer thickness) for the two meshes in Figure 7.6. The reaction zone thickness was measured as the distance, along the left end of the domain in the reference configuration, between the locations where the reaction state takes the values $\phi = 0.1$ and $\phi = 0.9$. It is noted that thickness remains fairly constant for both meshes, i.e., the interphase does not become more diffuse as the problem evolves. Some variation in the thickness is observed in the figure, which is attributed to how the values of the reaction state ϕ are interpolated as the reaction zone moves from one element to the next. Moreover, it is noted that the characteristic lengths of the elements along the vertical direction are 75 nm and 37.5 nm for the coarse and fine meshes respectively. Thus, the interphase thickness occupies roughly two elements in either mesh. This is consistent with what was observed for the VMS method in [4], where sharp intermediate layers were captured across two elements. We remark that the size of the reaction zone is determined by the diffusion and reaction constants inherent to the physical problem, and when its size is smaller than that of the element, the VMS method developed for diffusion-reaction problems is able to capture the variation across this zone across two-elements. For even finer discretizations, the entire interphase layer thickness may ultimately be resolved across several elements.

---

[3] The term "interphase" is used deliberately to highlight that there is no discrete interface, but a transition zone between pure Si and pure SiO_2. The existence of such intermediate layer of non-zero thickness is also reported in the experimental results for the oxidation of silicon carbide (SiC) in [17]

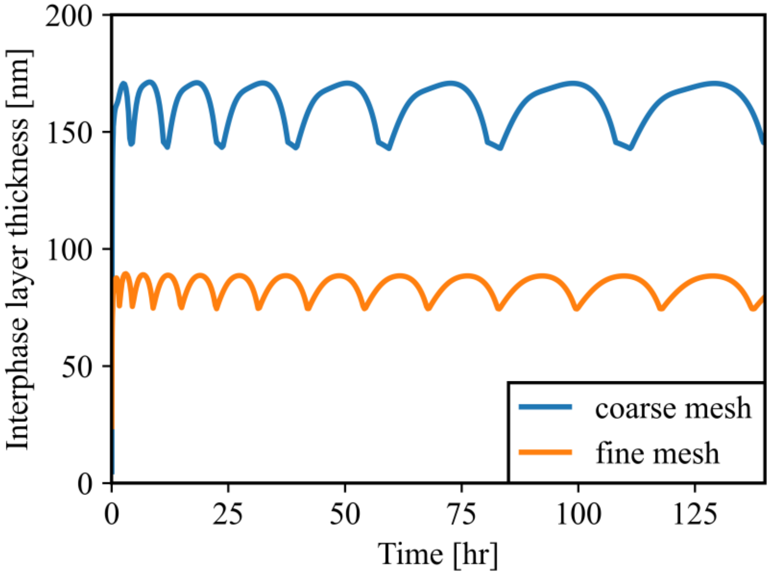


Figure 7.6. Evolution of the reaction zone (or interphase layer) thickness for two levels of mesh refinement.

Figure 7.7 shows a comparison of the evolution of reaction front location for two levels of refinement, measured as the distance in the reference configuration along the left boundary of the domain from the exposed surface to the location where $\phi = 0.75$. For both meshes, the advancing reaction front follows a similar trend. However, in the same amount of time, the reaction front moves further into the domain for the coarse mesh than the fine mesh. Other than the standard errors due to a coarser discretization, a possible source of error is the algorithmic assumption that the evolution of the reaction state variable does not depend on the fine scale field. Accounting for the fine scales in the evolution of the reaction state variable may improve the accuracy of the method in coarser meshes, at the cost of increasing the complexity of accurately computing the consistent tangent. This will be investigated in future work.

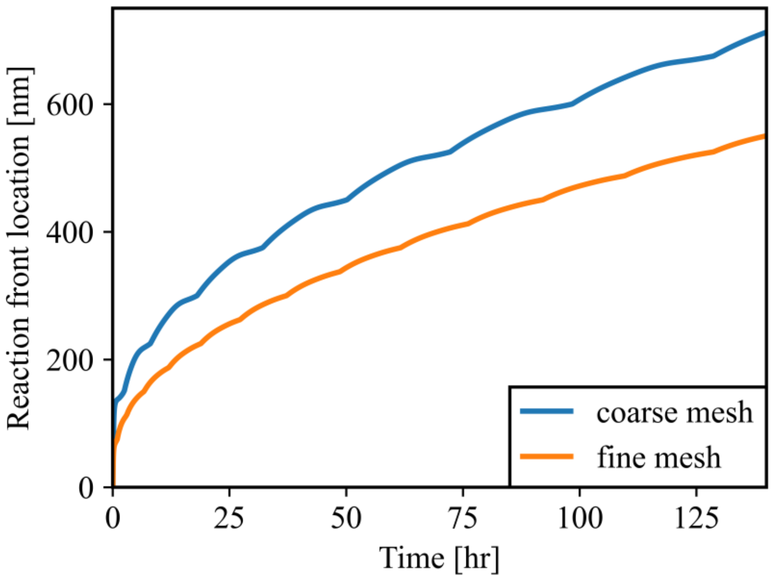


Figure 7.7. Evolution of the location of the reaction front for two levels of mesh refinement.

7.5.2 Fingering in (rigid) solid fuel combustion

Literature reporting fingering phenomena in solid fuel combustion includes the experimental results of Zik and Moses regarding smoldering combustion of a thin solid [11], and subsequent studies that develop mathematical models and present numerical results via explicit finite difference methods. Study of analytical traveling wave solutions is discussed [12], a study on flame spread speed by Funashima et al [13], and the work in [14], where it is shown that the mathematical model does not require advection terms for fingering to occur. Rather, the system is susceptible to develop a fingering instability for small values of the Lewis number, which is the ratio of the effective thermal diffusivity to the effective mass diffusivity of the fluid. At higher advection (higher Peclet number), the supply of oxidant is large enough that the reaction front advances uniformly, even for small values of the Lewis number.

For different material parameters, we consider the thermo-chemical problem of solid fuel combustion wherein the solid domain is assumed to be rigid. The parameters are similar to those of the problem studied by Funashima [13], except for the activation energy of the chemical reaction, which is taken to be half as much. The heat conductivity and heat dissipation coefficients are varied slightly to induce different conditions for fingering in the problem. Lower heat conductivity and higher heat dissipation coefficient make the problem more critical [13,14]. The Lewis number in the problems discussed below is such that $Le < 0.2$ in all cases.

Two computational domains are considered, the first is a smaller domain of size $10\text{ mm}\times 5\text{ mm}$ with a finer discretization into a grid of 100×50 elements, yielding a characteristic element size of $h = 0.1\text{ mm}$. The second computational domain is larger $(20\text{ mm}\times 5\text{ mm})$ discretized into a grid of 60×30 elements, so that the characteristic element size is $h = 0.33\text{ mm}$. Initial conditions for the temperature field are set to $\bar{\theta}_0 = 1.45$ (i.e. 45% hotter than the reference temperature) on a small portion of the left end of the domain, with a perturbation of total variation between 1.4 and 1.5 in the form of a cosine wave of 4 or 3 periods along the y axis. Backward Euler time integration algorithm is employed. The problem is solved using 200 timesteps with $\Delta t = 1\times 10^{-4}\text{ s}$, and $\Delta t = 1\times 10^{-3}\text{ s}$ thereafter.

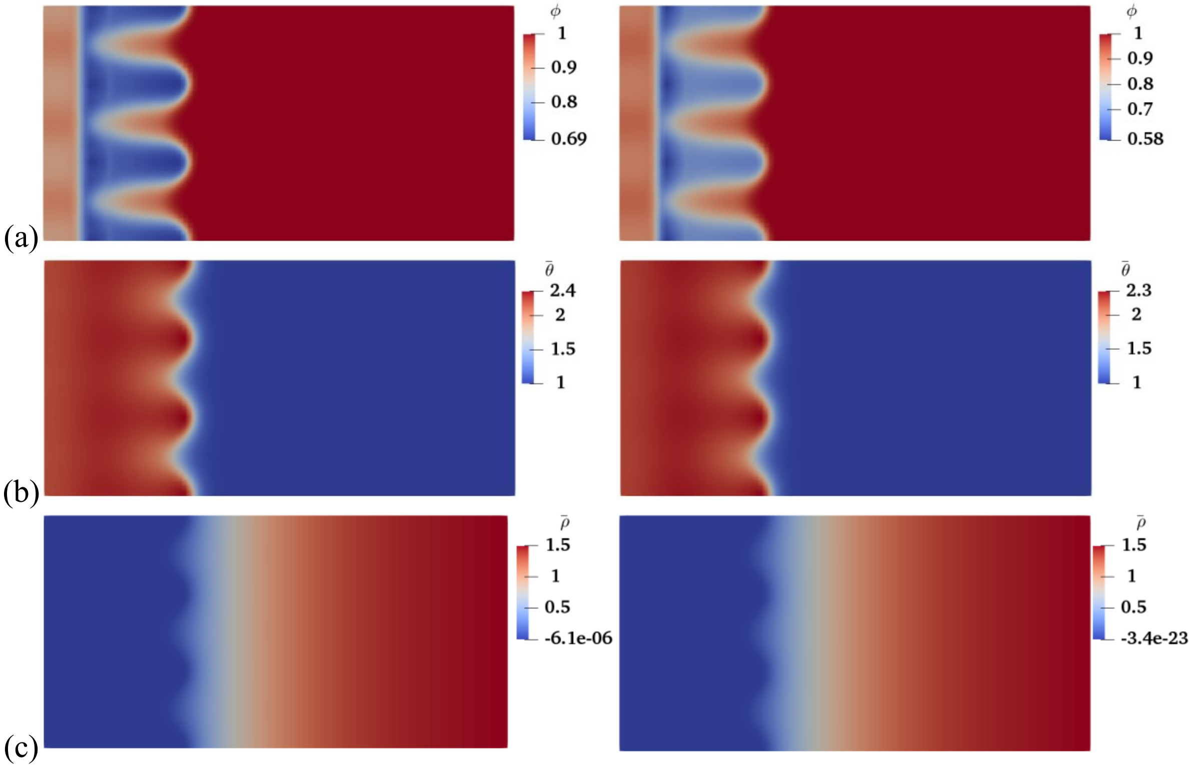


Figure 7.8. Comparison of (a) burnt patterns, (b) temperature field, and (c) normalized concentration field after 300 timesteps for a Galerkin FEM (left) and VMS method (right) in a fine mesh.

Figure 7.8 shows comparison of the burnt patters, normalized temperature field, and normalized fluid concentration field after 300 timesteps for the Galerkin FEM method and the VMS method in the smaller computational domain with a finer discretization. The results are in good qualitative agreement.

Figure 7.9a shows a comparison of the burnt patters as given by the reaction state variable. Regions where the reaction state variable is less than 1 are regions where reaction took place. In both cases there is finger splitting and flame extinction. However, for material parameter set 1, a stable configuration for finger growth is reached, whereas for material parameter set 2 computations eventually blow up. The only differences between these material parameter sets are that for set 2, there is a mild increase in the assumed enthalpy of reaction of about 10%, and a tenfold increase in the rate of heat dissipation in the bulk.

Figure 7.9b and c show the corresponding distribution of the normalized temperature field and normalized fluid concentration field. It is noted that non-physical undershoots appear near the fronts where there are temperature values below 1 (reference temperature in the domain) as well as negative fluid concentrations. This type of undershoot is expected for sharp fronts in the Galerkin FEM.

The solution is also computed utilizing the VMS method discussed in section 7.4, where the solution of the fine scale subproblem is decoupled and fine scales are assumed zero at the beginning of the timestep. This VMS method suppresses the non-physical undershoots and produces a much more stable solution as shown in Figure 7.10. However, the burnt pattern obtained is very different from that obtained with the standard Galerkin method. In fact, there are faint broad fingers in the burnt pattern, but not defined finger splitting or flame extinction. In addition, the front has not advance as far into the domain. Here, the VMS method is yielding a stable yet more diffuse solution. One important issue may be the fact that for this set of parameters, the temperature equation has the character of a Helmholtz equation instead of that of a diffusion-reaction equation. The two differ by the sign of the coefficient in the zeroth order term in the linearized problem. For the case of diffusion-reaction, the coefficient of the zeroth order term is positive, while for the Helmholtz equation, a negative real coefficient corresponds to a wave/oscillatory solution. It is also well known that the Galerkin FEM with low order elements suffers from dispersion errors in the Helmholtz equation for high wave number [2,15,16].

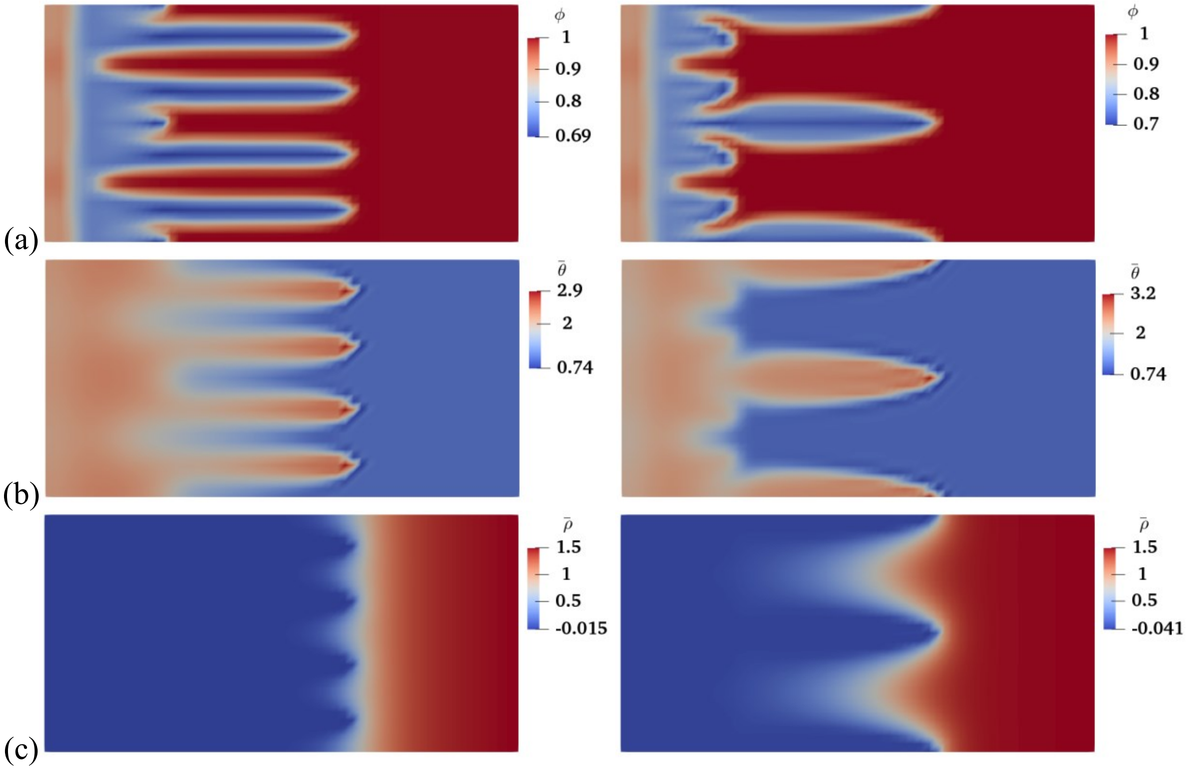


Figure 7.9. Comparison of (a) burnt patterns, (b) temperature field, and (c) normalized concentration field for material parameter set 1 after 700 timesteps (left) and set 2 after 444 timesteps (right)

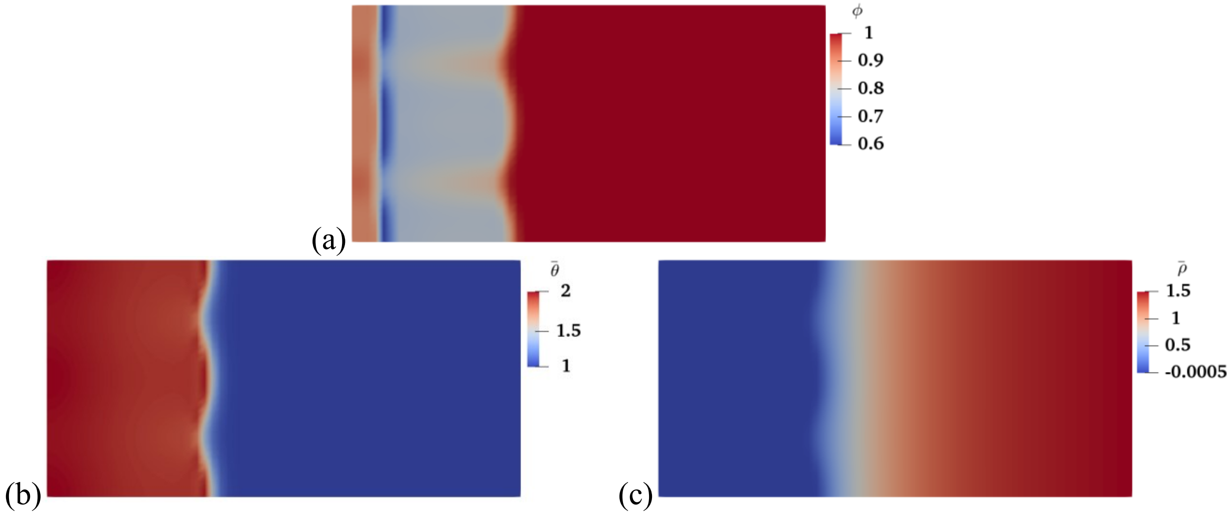


Figure 7.10. (a) Burnt pattern, (b) normalized temperature field, and (c) normalized concentration field for VMS solution with material parameter set 2

7.6 Conclusions

This Chapter developed a VMS method for application to thermo-chemo-mechanical problems. First, the scalar transient diffusion-reaction equation with material evolution is considered. Rigorous derivation within the VMS framework follows and several algorithmic choices are discussed. One important result is that it is shown that the coercivity results from [4] for linear Q4 element with flexible fine scale basis also hold when one makes the algorithmic choice of keeping the coarse scale residual within the integral on the right-hand side of the fine scale subproblem.

Furthermore, a VMS method is derived for an extension of the thermo-chemo-mechanical model presented in Chapter 4. The general VMS method is applied to the *system* of PDEs and considers multiscale split of both the fluid mass concentration and temperature fields. As an algorithmic simplification, one may choose to neglect the coupling of the fine-scale fields in the solution of the discrete fine scale subproblem. Numerical results are presented to highlight the features of the method. In the revisited bird's beak problem is revisited, using the reduced model yields computational cost savings, reaction front is captured across the length of roughly a single element, and the assumption of isothermal conditions is validated by verifying minimal variation of the temperature field for the non-isothermal problem. Moreover, numerical results show that the magnitude of the coarse scale residual and fine-scale fields is the largest around the advancing front, highlighting the role of the VMS formulation in capturing the solution in this region.

The numerical methods where applied to the simulation of fingering phenomena related to solid fuel combustion. The model and numerical methods are able to replicate stable finger growth as well as finger splitting and flame extinction. However, for critical material parameters, solutions grow unstable and computations diverge in the case of the Galerkin FEM. In contrast, the VMS method presented here does not develop into an unstable state, but the solution it finds for coarse meshes is overly diffusive and the same finger-splitting and flame extinction events are not observed. Possible remedies include implementing a selective VMS method as discussed in Chapter 6, where the VMS terms are active only where the reaction takes on critical values. In addition, another approach is to solve the full coupled fine-scale problem to obtain better approximations for the fine scale fields.

Nevertheless, the results for fingering phenomena are encouraging about the relevance of these developments as they relate to problems in combustion of rocket propellant. Enabling chemical expansion of the solid component represents a first step in modeling that kind of phenomena.

7.7 References

[1] T.J.R. Hughes, Multiscale phenomena: Green's functions, the Dirichlet-to-Neumann formulation, subgrid scale models, bubbles and the origins of stabilized methods, Comput. Methods Appl. Mech. Eng. 127 (1995) 387–401.

[2] T.J.R. Hughes, G.R. Feijóo, L. Mazzei, J.-B. Quincy, The variational multiscale method—a paradigm for computational mechanics, Comput. Methods Appl. Mech. Eng. 166 (1998) 3–24.

[3] A. Masud, L.P. Franca, A hierarchical multiscale framework for problems with multiscale source terms, Comput. Methods Appl. Mech. Eng. 197 (2008) 2692–2700.

[4] A. Masud, M. Anguiano, I. Harari, Modeling of steep layers in singularly perturbed diffusion–reaction equation via flexible fine-scale basis, Comput. Methods Appl. Mech. Eng. 372 (2020) 113343.

[5] J.A. Evans, T.J.R. Hughes, G. Sangalli, Enforcement of constraints and maximum principles in the variational multiscale method, Comput. Methods Appl. Mech. Eng. 199 (2009) 61–76.

[6] R. Hall, H. Gajendran, A. Masud, Diffusion of chemically reacting fluids through nonlinear elastic solids: mixture model and stabilized methods, Math. Mech. Solids. 20 (2015) 204–227.

[7] I. Harari, T.J.R. Hughes, What are C and h?: Inequalities for the analysis and design of finite element methods, Comput. Methods Appl. Mech. Eng. 97 (1992) 157–192.

[8] T. Belytschko, W.K. Liu, B. Moran, K.I. Elkhodary, Nonlinear Finite Elements for Continua and Structures, Second, John Wiley & Sons, Chichester, West Sussex, UK, 2014.

[9] M. Anguiano, H. Gajendran, R.B. Hall, K.R. Rajagopal, A. Masud, Chemo-mechanical coupling and material evolution in finitely deforming solids with advancing fronts of reactive fluids, Acta Mech. 231 (2020) 1933–1961.

[10] V.S. Rao, T.J.R. Hughes, K. Garikipati, On Modelling Thermal Oxidation of Silicon II: Numerical Aspects, Int. J. Numer. Methods Eng. 47 (2000) 359–377.

[11] O. Zik, E. Moses, Fingering instability in solid fuel combustion: The characteristic scales of the developed state, Symp. Combust. 27 (1998) 2815–2820.

[12] E.R. Ijioma, H. Izuhara, M. Mimura, Traveling Waves in a Reaction-Diffusion System Describing Smoldering Combustion, SIAM J. Appl. Math. 77 (2017) 614–637.

[13] K. Funashima, A. Masuyama, K. Kuwana, G. Kushida, Opposed-flow flame spread in a narrow channel: Prediction of flame spread velocity, Proc. Combust. Inst. 37 (2019) 3757–3765.

[14] Y. Uchida, K. Kuwana, G. Kushida, Experimental validation of Lewis number and convection effects on the smoldering combustion of a thin solid in a narrow space, Combust. Flame. 162 (2015) 1957–1963.

[15] L.P. Franca, C. Farhat, A.P. Macedo, M. Lesoinne, Residual-free bubbles for the Helmholtz equation, Int. J. Numer. Methods Eng. 40 (1997) 4003–4009.

[16] A.A. Oberai, P.M. Pinsky, A residual-based finite element method for the Helmholtz equation, Int. J. Numer. Methods Eng. 49 (2000) 399–419.

[17] K. Kouda, Y. Hijikata, S. Yagi, H. Yaguchi, S. Yoshida, Oxygen partial pressure dependence of the SiC oxidation process studied by in-situ spectroscopic ellipsometry, J. Appl. Phys. 112 (2012).

CHAPTER 8:
CONCLUDING REMARKS AND FUTURE DIRECTIONS

8.1 Concluding Remarks

This dissertation developed a framework for robust computer simulation of reaction-dominated thermo-chemo-mechanical processes involving deformable solids and reactive fluids. There are two main areas of work that are brought together and in which progress is made

1. The development of new variational multiscale numerical methods that are consistently derived. The methods recover unresolved physics to enable accurate, stable, robust, and cost-efficient numerical solutions.
2. The adoption, derivation, and extension of thermodynamically consistent models at the continuum level for multi-constituent material systems.

The work done here in these areas provides advances for the *simulation-based design* of engineered materials, manufacturing processes, devices, and structures in *extreme* environments. In this regard, key contributions of this work include

1. An explicit VMS method for multi-domain parallel computation in elastodynamics, wherein interdomain transmission conditions are facilitated by the VMDG method and the notion of edge bubbles as a fine scale model that does not vanish on element boundaries. The decoupling of the solution step for each subdomain is accomplished via explicit time integration scheme. It is numerically shown that the VMDG interface formulation in the method preserves energy conservation properties of the underlying time integration scheme, that the dispersion error is negligible compared to the underlying FEM discretization, and that increase in maximum frequency due to node mismatch at the interface is bounded.
2. A VMS method the diffusion-reaction equation in the reaction-dominated regime. The salient feature of the method is a fine scale model with increased flexibility consisting of multiple basis functions that do not vanish at element boundaries. A rigorous mathematical analysis with a priori error estimate is provided. Optimal error convergence rates are numerically verified. The method is numerically shown to be stable and robust in sharply capturing boundary and internal layers.

3. A thermodynamically consistent model for multi-constituent open systems characterized by evolving deformable solids permeated by chemically reacting fluids. The model accounts for energy associated with the formation and breaking of bonds. It also captures chemical deformation through a multiplicative split of the deformation gradient, which is convenient for the development of constitutive equations. Second law of thermodynamics is satisfied by providing constitutive equations that guarantee non-negative entropy production according to the Clausius-Duhem inequality for an open multi-constituent system. Furthermore, assuming the fluid constituent can be modeled as an ideal gas with a slow speed of diffusion, a reduced model is obtained that is well-suited for numerical implementation through the standard Galerkin FEM.
4. A VMS method for transient thermo-chemo-mechanical problems with material evolution in which both the fluid mass concentration field and the temperature field are decomposed into coarse scale and fine scale contributions. Derivation takes into account coupling in the solution of the fine scale subproblem and in embedding the fine scale solution into the coarse scale subproblem. Some simplifying approximations and algorithmic choices are discussed such as dependency of internal variables on fine-scale fields and time-dependent fine scales.
5. Numerical studies that provide insight into several problems in active areas of research. This include lithiation of silicon nanowires, which is important to the development of high-capacity batteries; thermal oxidation of silicon and silicon carbide, both relevant to the manufacturing of current and future semiconductor devices; and thermal oxidation of silicon carbide and thermal alloys, as well as studies on fingering phenomena related to solid fuel combustion, pertaining to applications in the aerospace industry.

In addition to the major contributions, there are other concluding remarks worth stating. The first is concerning the treatment of the material evolution and the computational cost of the alternative approaches. Throughout this work, the material evolution of the solid constituent is described by the reaction state variable. The gradient of this variable does not appear in the formulation and its evolution is given by an ODE in terms of the other state variables. Therefore, in the numerical implementation, the reaction state variable is treated as an internal time history variable whose evolution is given by solving its discrete ODE at the integration points. It is noted that the approach used here corresponds to a diffuse-interface approach because the chemical reaction is treated as a bulk process. Other theories for multi-phase systems and phase transition in fluids and solids also consider a diffuse interface, such as in the case of phase-field theories, closely related to the concept of balance of configurational forces, and which can be used to derive Cahn-Hilliard-type equations and their generalizations to multiphysics processes (see discussion in Chapter 1). In these other theories, gradients of the so-called "order parameter" (used to describe the presence of the phases in the system) do appear in the formulation. This requires the discretization of the order parameter in space, which would increase the number of degrees of freedom per node in the FEM and the computational cost of the model. Therefore, the treatment utilized in this dissertation involves computational cost savings in comparison.

Other important conclusions that are drawn from this work are summarized in the points below:

- There is a large class of coupled thermo-chemo-mechanical problems that involve the transport of chemically reactive species through evolving deformable solids. They arise in many areas of active research in engineering science such as aerospace applications, energy production and storage, manufacturing of electronics, biological growth and bioinspired devices, micro-self-actuators, and porous media flows. Some characteristics of these problems are multiple constituents, advancing reaction fronts, localized phenomena, and chemical deformations.
- Mixture theory framework allows to model multiple constituents as a continuum in the same domain, while retaining terms that capture the interaction between constituents. Thus, the geometric details of the microstructure and interfaces between constituents need not be explicitly modeled. In this sense, the resulting formulation may be thought of as a physics-based, reduced-order model for multi-constituent interaction.
- The governing equations containing the interaction terms give rise to a system of coupled, nonlinear, transient diffusion-advection reaction PDEs. For a range of processes and material parameters, as in the reaction- or advection-dominated cases, these PDEs become singularly perturbed and give rise to localized features such as internal and boundary layers.
- The Variational Multiscale framework allows for the consistent derivation of stable and accurate methods for mathematically non-smooth problems. Furthermore, the local nature

of the fine scale subproblem is amenable to scalable parallel implementation of these methods in a computer code.

- In particular, fine-scale models that do not vanish at element boundaries provide increased flexibility in the ability of the method to recover missing physics, reproduce the rapid variation of the coarse scale residual of the Euler-Lagrange equation, and capture directionality in the discrete problem. Furthermore, non-vanishing fine-scale models can also to develop transmission conditions to couple PDEs across interfaces, as done in the Variational Multiscale Discontinuous Galerkin (VMDG) methods.
- The ensuing VMS methods suppress spurious oscillations around localized features, reduce deviation of the numerical solution from maximum principle, and maintain the sharpness of boundary and intermediate layers.

8.2 Future Directions

The work in this dissertation employed the concept of fine scales facilitated by the VMS framework to develop stable methods for reaction-dominated regimes and for interface coupling in multi-domain problems. There remains a lot of potential in exploiting the concept of fine scales facilitated by the VMS framework to develop methods with other attributes (e.g., enforcement of constraints, uncertainty quantification, bridging of scales and models). Likewise, the methods developed in this research work have a wide range of applications that are yet to be fully explored. For instance, problems in *bio-engineering*, such as biological tissue growth and remodeling, as well as curing during extrusion-based *additive manufacturing* (AM), have significant parallels with the processes that were studied in this work: reaction-driven swelling, material evolution, multiple constituents, multi-physics response, and localized effects. These parallels also exist in outstanding problems in the *energy* sector, such as modeling of lithium ion batteries, reactive porous media flows, and performance of solid-oxide fuel cells (SOFCs). Moreover, *physics-informed machine learning* has emerged as a powerful tool in scientific computing for many applications, which can be leveraged to complement well-established approaches in this work. Therefore, it is proposed to move this research work forward by grouping several lines of work into two fronts:

1. **Method development:**
    1. Building on the innovations presented for addressing instability in the singularly perturbed diffusion-reaction equation, it is worth investigating the extension of those ideas to the advection-dominated case. The aim is to develop a consistent approach, free from user-defined parameters, for robust stabilization of the general diffusion-advection-reaction equation.
    2. The coupled and nonlinear nature of the PDEs being solved represent a challenge to the accuracy and stability of the numerical methods used to solve them. In the pursuit of robust numerical methods that can overcome these challenges, future work will explore the behavior of fine scales in time. This includes the evolution of the spatial fine-scale field in time as well as employing similar VMS multi-scaling ideas for the time domain.
    3. The physical problems of interest can be subject to constraints such as non-negativity of the field, conservation of certain quantities, incompressibility, and satisfaction of maximum principles. The VMS framework has facilitated a way to recover a primal field formulation for constrained problems. This line of thought may be explored as an avenue for enforcement of constraints in the class of problems discussed here.
    4. Explore the use of the fine-scale concept as a bridge between physics-informed machine-learning and uncertainty quantification tools on one side, and deterministic continuum physics models on the other.
2. **Extensions to physical models.**
    1. Account for plastic and viscous effects, which are crucial to accurately capture the response of biomaterials, high-temperature solids, and extrusion-based AM technologies. In addition, consider fully anisotropic material response.
    2. Revisit the fluid formulation to account for liquid flows, which may be modeled as incompressible and thus introduce an additional pressure field.
    3. Augment models with the driving magnetic and electric fields as well as the corresponding coupled response, such as in piezoelectric materials.

iv. Incorporate physics-informed machine-learning tools to the constitutive modeling framework, both in parameter determination and as substitute for unknown functional forms.

APPENDIX A:
DETAILS OF ADDITIONAL DERIVATIONS IN CHAPTER 4

A.1 Mixture Balance of Energy

The local form of the constituent-wise balance of energy is given as follows as per [1,2]

$$\frac{\partial}{\partial t}\left(\rho^{\alpha}\varepsilon^{\alpha}\right)+\operatorname{div}\left(\rho^{\alpha}\varepsilon^{\alpha}\mathbf{v}^{\alpha}\right)-\operatorname{tr}\left(\mathbf{T}^{\alpha}\mathbf{L}^{\alpha}\right)+\operatorname{div}\mathbf{q}^{\alpha}-\rho^{\alpha}r^{\alpha}-m^{\alpha}\varepsilon^{\alpha}-\varepsilon_{S}^{\alpha}=0 \tag{A.1}$$

where $\mathbf{q}^{\alpha}$ is the heat flux in constituent α. However, in the case of the solid constituent that goes through material evolution due to chemical reactions, equation (A.1) is augmented with a term that accounts for the energy associated with breaking and formation of chemical bonds

$$\frac{\partial}{\partial t}\left(\rho^{s}\varepsilon^{s}\right)+\operatorname{div}\left(\rho^{s}\varepsilon^{s}\mathbf{v}^{s}\right)-\operatorname{tr}\left(\mathbf{T}^{s}\mathbf{L}^{s}\right)+\operatorname{div}\mathbf{q}^{s}-\rho^{s}r^{s}-m^{s}\varepsilon^{s}-\varepsilon_{S}^{s}-c_{\phi}\frac{D^{s}\phi}{Dt}=0 \tag{A.2}$$

Then, the balance of energy for the mixture is obtained by adding together the constituent-wise local form of the balance of energy for all constituents, which yields

$$\sum_{\alpha}\begin{bmatrix}\frac{\partial}{\partial t}\left(\rho^{\alpha}\varepsilon^{\alpha}\right)+\operatorname{div}\left(\rho^{\alpha}\varepsilon^{\alpha}\mathbf{v}^{\alpha}\right)\\ -\operatorname{tr}\left(\mathbf{T}^{\alpha}\mathbf{L}^{\alpha}\right)+\operatorname{div}\mathbf{q}^{\alpha}-\rho^{\alpha}r^{\alpha}-m^{\alpha}\varepsilon^{\alpha}-\varepsilon_{S}^{\alpha}\end{bmatrix}-c_{\phi}\frac{D^{s}\phi}{Dt}=0 \tag{A.3}$$

Furthermore, consider the following definitions of the mixture level quantities

$$\varepsilon:=\frac{1}{\rho}\sum_{\alpha}\rho^{\alpha}\varepsilon^{\alpha} \tag{A.4}$$

$$\mathbf{q}:=\sum_{\alpha}\mathbf{q}^{\alpha} \tag{A.5}$$

$$r:=\frac{1}{\rho}\sum_{\alpha}\rho^{\alpha}r^{\alpha} \tag{A.6}$$

In addition, assuming energy is conserved as it is transferred between constituents

$$\sum_{\alpha}\left(\varepsilon_{S}^{\alpha}+\mathbf{I}^{\alpha}\cdot\mathbf{v}^{\alpha}+m^{\alpha}\left(\varepsilon^{\alpha}+\frac{1}{2}\mathbf{v}^{\alpha}\cdot\mathbf{v}^{\alpha}\right)\right)=0 \tag{A.7}$$

Using (A.4)-(A.7), equation (A.3) is rewritten as

$$\begin{aligned}&\frac{\partial}{\partial t}\left(\rho\varepsilon\right)+\operatorname{div}\mathbf{q}-\rho r-c_{\phi}\frac{D^{s}\phi}{Dt}\\&+\sum_{\alpha}\left[\operatorname{div}\left(\rho^{\alpha}\varepsilon^{\alpha}\mathbf{v}^{\alpha}\right)-\operatorname{tr}\left(\mathbf{T}^{\alpha}\mathbf{L}^{\alpha}\right)+\mathbf{I}^{\alpha}\cdot\mathbf{v}^{\alpha}+\frac{1}{2}m^{\alpha}\mathbf{v}^{\alpha}\cdot\mathbf{v}^{\alpha}\right]=0\end{aligned} \tag{A.8}$$

Recall the definition of the mixture total time derivative given in (4.14), then

$$\begin{aligned}\rho\frac{D\varepsilon}{Dt} &= \rho\frac{\partial\varepsilon}{\partial t}+\rho\mathbf{v}\cdot\nabla\varepsilon \\ &= \rho\frac{\partial\varepsilon}{\partial t}+\operatorname{div}\left(\rho\varepsilon\mathbf{v}\right)-\varepsilon\mathbf{v}\cdot\nabla\rho-\rho\varepsilon\operatorname{div}\mathbf{v}\end{aligned} \tag{A.9}$$

Therefore,

$$\begin{aligned}\frac{\partial}{\partial t}\left(\rho\varepsilon\right) &= \varepsilon\frac{\partial\rho}{\partial t}+\rho\frac{\partial\varepsilon}{\partial t} \\ &= \varepsilon\frac{\partial\rho}{\partial t}+\rho\frac{D\varepsilon}{Dt}-\operatorname{div}\left(\rho\varepsilon\mathbf{v}\right)+\varepsilon\mathbf{v}\cdot\nabla\rho+\rho\varepsilon\operatorname{div}\mathbf{v} \\ &= \rho\frac{D\varepsilon}{Dt}-\operatorname{div}\left(\rho\varepsilon\mathbf{v}\right)+\varepsilon\left[\frac{\partial\rho}{\partial t}+\mathbf{v}\cdot\nabla\rho+\rho\operatorname{div}\mathbf{v}\right]\end{aligned} \tag{A.10}$$

Recall that due to conservation of mass for the mixture, we have that

$$\frac{\partial\rho}{\partial t}+\mathbf{v}\cdot\nabla\rho+\rho\operatorname{div}\mathbf{v}=0 \tag{A.11}$$

Therefore,

$$\frac{\partial}{\partial t}\left(\rho\varepsilon\right)=\rho\frac{D\varepsilon}{Dt}-\operatorname{div}\left(\rho\varepsilon\mathbf{v}\right) \tag{A.12}$$

Substituting (A.12) into (A.8) and rearranging yield

$$\begin{aligned}&\rho\frac{D\varepsilon}{Dt}+\operatorname{div}\mathbf{q}-\rho r-c_{\phi}\frac{D^{s}\phi}{Dt}+\sum_{\alpha}\operatorname{div}\left(\rho^{\alpha}\varepsilon^{\alpha}\mathbf{v}^{\alpha}\right)-\operatorname{div}\left(\rho\varepsilon\mathbf{v}\right) \\ &+\sum_{\alpha}\left[-\operatorname{tr}\left(\mathbf{T}^{\alpha}\mathbf{L}^{\alpha}\right)+\mathbf{I}^{\alpha}\cdot\mathbf{v}^{\alpha}+\frac{1}{2}m^{\alpha}\mathbf{v}^{\alpha}\cdot\mathbf{v}^{\alpha}\right]=0\end{aligned} \tag{A.13}$$

Recalling (A.4), and the definition of the relative velocity in (4.8), we have that

$$\sum_{\alpha}\operatorname{div}\left(\rho^{\alpha}\varepsilon^{\alpha}\mathbf{v}^{\alpha}\right)-\operatorname{div}\left(\rho\varepsilon\mathbf{v}\right)=\sum_{\alpha}\operatorname{div}\left(\rho^{\alpha}\varepsilon^{\alpha}\mathbf{w}^{\alpha}\right) \tag{A.14}$$

and substituting (A.14) into (A.13)

$$\begin{aligned}&\rho\frac{D\varepsilon}{Dt}+\sum_{\alpha}\operatorname{div}\left(\rho^{\alpha}\varepsilon^{\alpha}\mathbf{w}^{\alpha}\right)+\operatorname{div}\mathbf{q}-\rho r-c_{\phi}\frac{D^{s}\phi}{Dt} \\ &+\sum_{\alpha}\left[-\operatorname{tr}\left(\mathbf{T}^{\alpha}\mathbf{L}^{\alpha}\right)+\mathbf{I}^{\alpha}\cdot\mathbf{v}^{\alpha}+\frac{1}{2}m^{\alpha}\mathbf{v}^{\alpha}\cdot\mathbf{v}^{\alpha}\right]=0\end{aligned} \tag{A.15}$$

Using again the assumption that mass and momentum are conserved when transferred among constituents, i.e.

$$\sum_{\alpha} m^{\alpha} = 0 \tag{A.16}$$

$$\sum_{\alpha} \left(\mathbf{I}^{\alpha} + m^{\alpha} \mathbf{v}^{\alpha} \right) = 0 \tag{A.17}$$

then, equation (A.13) is re-written as

$$\begin{aligned} &\rho \frac{D\varepsilon}{Dt} + \sum_{\alpha} \operatorname{div}\left(\rho^{\alpha} \varepsilon^{\alpha} \mathbf{w}^{\alpha} \right) + \operatorname{div}\mathbf{q} - \rho r - \sum_{\alpha} \operatorname{tr}\left(\mathbf{T}^{\alpha} \mathbf{L}^{\alpha} \right) \\ &\qquad + \frac{1}{2} m^{f} \left(\mathbf{v}^{f} - \mathbf{v}^{s} \right) \cdot \left(\mathbf{v}^{f} - \mathbf{v}^{s} \right) + \mathbf{I}^{f} \cdot \left(\mathbf{v}^{f} - \mathbf{v}^{s} \right) - c_{\phi} \frac{D^{s}\phi}{Dt} = 0 \end{aligned} \tag{A.18}$$

A.2 Clausius-Duhem Entropy Inequality

The local form of the entropy inequality can be written as follows [1,2]

$$\frac{\partial}{\partial t}\left(\sum_{\alpha} \rho^{\alpha} \eta^{\alpha} \right) + \sum_{\alpha} \operatorname{div}\left(\rho^{\alpha} \eta^{\alpha} \mathbf{v}^{\alpha} \right) + \operatorname{div}\left(\sum_{\alpha} \frac{\mathbf{q}^{\alpha}}{\theta} \right) - \sum_{\alpha} \frac{\rho^{\alpha} r^{\alpha}}{\theta} \quad \geq \quad 0 \tag{A.19}$$

Note that the second term, $\sum_{\alpha} \operatorname{div}\left(\rho^{\alpha} \eta^{\alpha} \mathbf{v}^{\alpha} \right)$, is associated with entropy entering or leaving the open system through the boundary due to mass transport into or out of the domain. Then, the mixture entropy is defined as

$$\eta := \frac{1}{\rho} \sum_{\alpha} \rho^{\alpha} \eta^{\alpha} \tag{A.20}$$

Using (A.5), (A.6), and (A.20), equation (A.19) is re-written as

$$\frac{\partial}{\partial t}(\rho\eta) + \sum_{\alpha} \operatorname{div}\left(\rho^{\alpha} \eta^{\alpha} \mathbf{v}^{\alpha} \right) + \operatorname{div}\left(\frac{\mathbf{q}}{\theta} \right) - \frac{\rho r}{\theta} \quad \geq \quad 0 \tag{A.21}$$

Analogously to (A.9)-(A.12), the following identity is obtained

$$\frac{\partial}{\partial t}(\rho\eta) = \rho \frac{D\eta}{Dt} - \operatorname{div}(\rho\eta\mathbf{v}) \tag{A.22}$$

Furthermore, we have that

$$\operatorname{div}\left(\frac{\mathbf{q}}{\theta} \right) = \frac{1}{\theta} \operatorname{div}\mathbf{q} - \frac{\mathbf{q} \cdot \nabla\theta}{\theta^{2}} \tag{A.23}$$

Substituting (A.22) and (A.23) in (A.21) yields

$$\rho \frac{D\eta}{Dt} - \operatorname{div}(\rho\eta\mathbf{v}) + \sum_{\alpha} \operatorname{div}\left(\rho^{\alpha} \eta^{\alpha} \mathbf{v}^{\alpha} \right) + \frac{1}{\theta} \operatorname{div}\mathbf{q} - \frac{\mathbf{q} \cdot \nabla\theta}{\theta^{2}} - \frac{\rho r}{\theta} \quad \geq \quad 0 \tag{A.24}$$

Furthermore, using an identity analogous to (A.14), the inequality is re-written as

$$\rho \frac{D\eta}{Dt} + \sum_{\alpha} \text{div}\left(\rho^{\alpha}\eta^{\alpha}\mathbf{w}^{\alpha}\right) + \frac{1}{\theta}\text{div}\mathbf{q} - \frac{\mathbf{q}\cdot\nabla\theta}{\theta^{2}} - \frac{\rho r}{\theta} \quad \geq \quad 0 \tag{A.25}$$

Multiplying inequality (A.25) with the absolute temperature θ, and substituting the heat source term using the energy balance from equation (A.15), the following is obtained

$$\begin{gathered} \rho\theta\frac{D\eta}{Dt} - \rho\frac{D\varepsilon}{Dt} + \theta\sum_{\alpha}\text{div}\left(\rho^{\alpha}\eta^{\alpha}\mathbf{w}^{\alpha}\right) - \sum_{\alpha}\text{div}\left(\rho^{\alpha}\varepsilon^{\alpha}\mathbf{w}^{\alpha}\right) \\ -\frac{\mathbf{q}\cdot\nabla\theta}{\theta} + c_{\phi}\frac{D^{s}\phi}{Dt} - \sum_{\alpha}\left[-\text{tr}\left(\mathbf{T}^{\alpha}\mathbf{L}^{\alpha}\right) + \mathbf{I}^{\alpha}\cdot\mathbf{v}^{\alpha} + \frac{1}{2}m^{\alpha}\mathbf{v}^{\alpha}\cdot\mathbf{v}^{\alpha}\right] \quad \geq \quad 0 \end{gathered} \tag{A.26}$$

To make progress, the specific Helmholtz free energy of the mixture is defined as

$$\psi := \varepsilon - \theta\eta \tag{A.27}$$

Likewise, the specific Helmholtz free energy of each constituent is defined as

$$\psi^{\alpha} = \varepsilon^{\alpha} - \theta\eta^{\alpha} \tag{A.28}$$

From (A.4), (A.20), (A.27), and (A.29), it follows that

$$\psi = \frac{1}{\rho}\sum_{\alpha}\rho^{\alpha}\psi^{\alpha} \tag{A.29}$$

Furthermore, also from (A.27)

$$\frac{D\varepsilon}{Dt} = \frac{D\psi}{Dt} + \eta\frac{D\theta}{Dt} + \theta\frac{D\eta}{Dt} \tag{A.30}$$

and from (A.28)

$$\text{div}\left(\rho^{\alpha}\varepsilon^{\alpha}\mathbf{w}^{\alpha}\right) = \text{div}\left(\rho^{\alpha}\psi^{\alpha}\mathbf{w}^{\alpha}\right) + \theta\text{div}\left(\rho^{\alpha}\eta^{\alpha}\mathbf{w}^{\alpha}\right) + \rho^{\alpha}\eta^{\alpha}\mathbf{w}^{\alpha}\cdot\nabla\theta \tag{A.31}$$

Substituting (A.30) and (A.31) into (A.26)

$$\begin{gathered} -\rho\left(\frac{D\psi}{Dt} + \eta\frac{D\theta}{Dt}\right) - \sum_{\alpha}\text{div}\left(\rho^{\alpha}\psi^{\alpha}\mathbf{w}^{\alpha}\right) - \left(\frac{\mathbf{q}}{\theta} + \sum_{\alpha}\rho^{\alpha}\eta^{\alpha}\mathbf{w}^{\alpha}\right)\cdot\nabla\theta \\ +c_{\phi}\frac{D^{s}\phi}{Dt} - \sum_{\alpha}\left[-\text{tr}\left(\mathbf{T}^{\alpha}\mathbf{L}^{\alpha}\right) + \mathbf{I}^{\alpha}\cdot\mathbf{v}^{\alpha} + \frac{1}{2}m^{\alpha}\mathbf{v}^{\alpha}\cdot\mathbf{v}^{\alpha}\right] \quad \geq \quad 0 \end{gathered} \tag{A.32}$$

The divergence term can be expanded by using product rule and recalling the definition of the mixture velocity in (4.7), the relative velocity in (4.8), and the velocity gradient in (4.9), which results in

$$\sum_{\alpha}\text{div}\left(\rho^{\alpha}\psi^{\alpha}\mathbf{w}^{\alpha}\right) = \sum_{\alpha}\left[\left(\psi^{\alpha} - \psi\right)\left(\rho^{\alpha}\text{tr}\mathbf{L}^{\alpha} + \mathbf{w}^{\alpha}\cdot\nabla\rho^{\alpha}\right) + \rho^{\alpha}\mathbf{w}^{\alpha}\cdot\nabla\psi^{\alpha}\right] \tag{A.33}$$

Now, the Helmholtz free energy terms will be expanded using chain rule and their dependence on the state variables as given in equations (4.45)-(4.47). First, the gradient terms in (A.33) can be expanded as

$$\nabla \psi^{f} = \nabla \rho^{f} \frac{\partial \psi^{f}}{\partial \rho^{f}} + \nabla \theta \frac{\partial \psi^{f}}{\partial \theta} \tag{A.34}$$

$$\nabla \psi^{s} = \nabla \rho^{s} \frac{\partial \psi^{s}}{\partial \rho^{s}} + \nabla \theta \frac{\partial \psi^{s}}{\partial \theta} + \nabla \phi \frac{\partial \psi^{s}}{\partial \phi} + \nabla \mathbf{F}^{s} : \frac{\partial \psi^{s}}{\partial \mathbf{F}^{s}} \tag{A.35}$$

Then, the time derivative term from (A.32) is expanded as

$$\frac{D\psi}{Dt} = \sum_{\alpha} \frac{\partial \psi}{\partial \rho^{\alpha}} \frac{D\rho^{\alpha}}{Dt} + \frac{\partial \psi}{\partial \theta} \frac{D\theta}{Dt} + \frac{\partial \psi}{\partial \phi} \frac{D\phi}{Dt} + \frac{\partial \psi}{\partial \mathbf{F}^{s}} : \frac{D\mathbf{F}^{s}}{Dt} \tag{A.36}$$

Combining the definition of the relative velocity from (4.8), the velocity gradient from (4.9), the mixture and constituent total time derivatives in (4.12) and (4.14), and the constituent balance of mass in (4.29), the following identity is obtained

$$\frac{D\rho^{\alpha}}{Dt} = -\mathbf{w}^{\alpha} \cdot \nabla \rho^{\alpha} - \rho^{\alpha} \operatorname{tr}\left(\mathbf{L}^{\alpha}\right) + m^{\alpha} \tag{A.37}$$

Furthermore, also using the definitions of the solid rate of the deformation gradient (4.13), the mixture and constituent total time derivatives in (4.12) and (4.14), it holds that

$$\frac{D\phi}{Dt} = \frac{D^{s}\phi}{Dt} - \mathbf{w}^{s} \cdot \nabla \phi \tag{A.38}$$

$$\frac{D\mathbf{F}}{Dt} = \frac{D^{s}\mathbf{F}}{Dt} - \mathbf{w}^{s} \cdot \nabla \mathbf{F}^{s} = \mathbf{L}^{s}\mathbf{F}^{s} - \mathbf{w}^{s} \cdot \nabla \mathbf{F}^{s} \tag{A.39}$$

Now, substituting (A.37)-(A.39) into (A.36); substituting (A.34) and (A.35) into (A.33); and substituting the resulting expressions into (A.32), the entropy inequality becomes the following

$$\begin{aligned}
&-\rho\sum_{\alpha}\frac{\partial\psi}{\partial\rho^{\alpha}}\left(-\mathbf{w}^{\alpha}\cdot\nabla\rho^{\alpha}-\rho^{\alpha}\mathrm{tr}\left(\mathbf{L}^{\alpha}\right)+m^{\alpha}\right)\\
&-\rho\left(\frac{\partial\psi}{\partial\theta}\frac{D\theta}{Dt}+\frac{\partial\psi}{\partial\phi}\left[\frac{D^{s}\phi}{Dt}-\mathbf{w}^{s}\cdot\nabla\phi\right]+\frac{\partial\psi}{\partial\mathbf{F}^{s}}:\left[\mathbf{L}^{s}\mathbf{F}^{s}-\mathbf{w}^{s}\cdot\nabla\mathbf{F}^{s}\right]+\eta\frac{D\theta}{Dt}\right)\\
&-\sum_{\alpha}\left[\left(\psi^{\alpha}-\psi\right)\left(\rho^{\alpha}\mathrm{tr}\mathbf{L}^{\alpha}+\mathbf{w}^{\alpha}\cdot\nabla\rho^{\alpha}\right)+\rho^{\alpha}\mathbf{w}^{\alpha}\cdot\nabla\rho^{\alpha}\frac{\partial\psi^{\alpha}}{\partial\rho^{\alpha}}+\rho^{\alpha}\mathbf{w}^{\alpha}\cdot\nabla\theta\frac{\partial\psi^{\alpha}}{\partial\theta}\right]\\
&-\rho^{s}\mathbf{w}^{s}\cdot\left(\nabla\phi\frac{\partial\psi^{s}}{\partial\phi}+\nabla\mathbf{F}^{s}:\frac{\partial\psi^{s}}{\partial\mathbf{F}^{s}}\right)-\left(\frac{\mathbf{q}}{\theta}+\sum_{\alpha}\rho^{\alpha}\eta^{\alpha}\mathbf{w}^{\alpha}\right)\cdot\nabla\theta\\
&+c_{\phi}\frac{D^{s}\phi}{Dt}-\sum_{\alpha}\left[-\mathrm{tr}\left(\mathbf{T}^{\alpha}\mathbf{L}^{\alpha}\right)+\mathbf{I}^{\alpha}\cdot\mathbf{v}^{\alpha}+\frac{1}{2}m^{\alpha}\mathbf{v}^{\alpha}\cdot\mathbf{v}^{\alpha}\right]\;\geq\;0
\end{aligned}\tag{A.40}$$

The terms can be rearranged by gathering terms whose coefficients are gradients or time derivatives of the state variables as follows

$$\begin{aligned}
&\mathrm{tr}\left\{\left[\mathbf{T}^{s}-\rho\mathbf{F}^{s}\left(\frac{\partial\psi}{\partial\mathbf{F}^{s}}\right)^{T}+\rho^{s}\left(\rho\frac{\partial\psi}{\partial\rho^{s}}+\psi-\psi^{s}\right)\mathbf{1}\right]\mathbf{L}^{s}\right\}\\
&+\mathrm{tr}\left\{\left[\mathbf{T}^{f}+\rho^{f}\left(\rho\frac{\partial\psi}{\partial\rho^{f}}+\psi-\psi^{f}\right)\mathbf{1}\right]\mathbf{L}^{f}\right\}\\
&-\rho\left(\eta+\frac{\partial\psi}{\partial\theta}\right)\frac{D\theta}{Dt}-\left(\frac{\mathbf{q}}{\theta}+\sum_{\alpha}\left(\eta^{\alpha}+\frac{\partial\psi^{\alpha}}{\partial\theta}\right)\rho^{\alpha}\mathbf{w}^{\alpha}\right)\cdot\nabla\theta\\
&-\sum_{\alpha}\left[\mathbf{I}^{\alpha}\cdot\mathbf{v}^{\alpha}+\frac{1}{2}m^{\alpha}\mathbf{v}^{\alpha}\cdot\mathbf{v}^{\alpha}\right]+\left(c_{\phi}-\rho\frac{\partial\psi}{\partial\phi}\right)\frac{D^{s}\phi}{Dt}-\rho\sum_{\alpha}\frac{\partial\psi}{\partial\rho^{\alpha}}m^{\alpha}\\
&+\sum_{\alpha}\left(\rho\frac{\partial\psi}{\partial\rho^{\alpha}}+\psi-\psi^{\alpha}-\rho^{\alpha}\frac{\partial\psi^{\alpha}}{\partial\rho^{\alpha}}\right)\mathbf{w}^{\alpha}\cdot\nabla\rho^{\alpha}\\
&+\left(\rho\frac{\partial\psi}{\partial\phi}-\rho^{s}\frac{\partial\psi^{s}}{\partial\phi}\right)\mathbf{w}^{s}\cdot\nabla\phi+\mathbf{w}^{s}\cdot\nabla\mathbf{F}^{s}:\left(\rho\frac{\partial\psi}{\partial\mathbf{F}^{s}}-\rho^{s}\frac{\partial\psi^{s}}{\partial\mathbf{F}^{s}}\right)\;\geq\;0
\end{aligned}\tag{A.41}$$

In addition, recalling (4.46), (4.47), and (A.29), we have that

$$\frac{\partial\psi}{\partial\rho^{\alpha}}=\frac{1}{\rho}\left(-\psi+\psi^{\alpha}+\rho^{\alpha}\frac{\partial\psi^{\alpha}}{\partial\rho^{\alpha}}\right)\tag{A.42}$$

$$\frac{\partial\psi}{\partial\rho^{s}}=-\frac{1}{\rho}\psi\tag{A.43}$$

$$\frac{\partial\psi}{\partial\phi}\;=\;\frac{\rho^{s}}{\rho}\frac{\partial\psi^{s}}{\partial\phi}\;=\;\frac{1}{\rho J}\frac{\partial\hat{\psi}^{s}}{\partial\phi}\tag{A.44}$$

$$\frac{\partial \psi}{\partial \mathbf{F}^s} = \frac{\rho^s}{\rho}\frac{\partial \psi^s}{\partial \mathbf{F}^s} = \frac{\rho^s}{\rho}\left[-\psi^s \left(\mathbf{F}^s\right)^{-T} + \left(\rho^s J\right)^{-1} \frac{\partial \hat{\psi}^s}{\partial \mathbf{F}^s} \right] \tag{A.45}$$

Using identities (A.42)-(A.45) in (A.41), the last three terms vanish and the following results

$$\begin{aligned}
&\operatorname{tr}\left\{ \left[\mathbf{T}^s - J^{-1}\mathbf{F}^s \left(\frac{\partial \hat{\psi}^s}{\partial \mathbf{F}^s} \right)^T \right] \mathbf{L}^s \right\} + \operatorname{tr}\left\{ \left[\mathbf{T}^f + \left(\rho^f\right)^2 \frac{\partial \psi^f}{\partial \rho^f}\mathbf{1} \right] \mathbf{L}^f \right\} \\
&- \rho\left(\eta + \frac{\partial \psi}{\partial \theta} \right)\frac{D\theta}{Dt} - \left(\frac{\mathbf{q}}{\theta} + \sum_\alpha \left(\eta^\alpha + \frac{\partial \psi^\alpha}{\partial \theta} \right)\rho^\alpha \mathbf{w}^\alpha \right)\cdot \nabla\theta \\
&- \sum_\alpha \left[\mathbf{I}^\alpha \cdot \mathbf{v}^\alpha + \frac{1}{2} m^\alpha \mathbf{v}^\alpha \cdot \mathbf{v}^\alpha \right] + \left(c_\phi - J^{-1}\frac{\partial \hat{\psi}^s}{\partial \phi} \right)\frac{D^s\phi}{Dt} - \rho \sum_\alpha \frac{\partial \psi}{\partial \rho^\alpha} m^\alpha \geq 0
\end{aligned} \tag{A.46}$$

Finally, using equations (A.16), (A.17), (A.42), (A.43), and recalling the connection between mass transfer rate and evolution of reaction state, given by equation (4.39), the sums over constituents are also expanded to yield

$$\begin{aligned}
&\operatorname{tr}\left\{ \left[\mathbf{T}^s - J^{-1}\mathbf{F}^s \left(\frac{\partial \hat{\psi}^s}{\partial \mathbf{F}^s} \right)^T \right] \mathbf{L}^s \right\} + \operatorname{tr}\left\{ \left[\mathbf{T}^f + (\rho^f)^2 \frac{\partial \psi^f}{\partial \rho^f}\mathbf{1} \right] \mathbf{L}^f \right\} - \rho\left(\eta + \frac{\partial \psi}{\partial \theta} \right)\frac{D\theta}{Dt} \\
&\quad - \left[\frac{\mathbf{q}}{\theta} + \frac{\rho^f \rho^s}{\rho}\left(\eta^f - \eta^s + \frac{\partial \psi^f}{\partial \theta} - \frac{\partial \psi^s}{\partial \theta} \right)\left(\mathbf{v}^f - \mathbf{v}^s\right) \right]\cdot \nabla\theta \\
&\quad - \left[\mathbf{I}^f + \frac{1}{2} m^f \left(\mathbf{v}^f - \mathbf{v}^s\right) \right]\cdot\left(\mathbf{v}^f - \mathbf{v}^s\right) \\
&\quad + \left[c_\phi - \beta_\phi \left(\psi^f + \rho^f \frac{\partial \psi^f}{\partial \rho^f} \right) - J^{-1}\frac{\partial \hat{\psi}^s}{\partial \phi} \right]\frac{D^s\phi}{Dt} = \mathcal{D} \geq 0
\end{aligned} \tag{A.47}$$

A.3 Temperature Form of the Balance of Energy

In order to derive the temperature form of the balance of energy, some useful identities are required. First, for the total dissipation, $\mathcal{D}$, recall the Clausius-Duhem form of the entropy inequality in (4.48). The dependent quantities are substituted in terms of their quasi-conservative (q) and dissipative (d) contributions given in equations (4.49)-(4.61), the total dissipation is given by

$$\mathcal{D} = -\mathbf{I}^{f,(d)}\cdot\left(\mathbf{v}^f - \mathbf{v}^s\right) + \operatorname{tr}\left(\mathbf{T}^{f,(d)}\mathbf{L}^f\right) + c_\phi^{(d)}\frac{D^s\phi}{Dt} - \frac{\mathbf{q}\cdot\nabla\theta}{\theta} \tag{A.48}$$

Furthermore, using (A.27), and the constitutive equation for the mixture entropy from (4.53), the following is obtained

$$\rho\frac{D\varepsilon}{Dt} \quad = \quad \rho\left(\frac{D\psi}{Dt}+\eta\frac{D\theta}{Dt}+\theta\frac{D\eta}{Dt}\right) \quad = \quad \rho\left(\frac{D\psi}{Dt}+\eta\frac{D\theta}{Dt}\right)+\rho\theta\frac{D}{Dt}\left(-\frac{\partial\psi}{\partial\theta}\right) \qquad \text{(A.49)}$$

Now, recall the mixture energy balance from equation (4.31), and the entropy equation from (4.44), reproduced here for clarity and convenience of the derivation

$$\begin{aligned}&\rho\frac{D\varepsilon}{Dt}+\sum_{\alpha}\operatorname{div}\left(\rho^{\alpha}\varepsilon^{\alpha}\mathbf{w}^{\alpha}\right)+\operatorname{div}\mathbf{q}-\rho r-\sum_{\alpha}\operatorname{tr}\left(\mathbf{T}^{\alpha}\mathbf{L}^{\alpha}\right)\\ &\qquad+\frac{1}{2}m^{f}\left(\mathbf{v}^{f}-\mathbf{v}^{s}\right)\cdot\left(\mathbf{v}^{f}-\mathbf{v}^{s}\right)+\mathbf{I}^{f}\cdot\left(\mathbf{v}^{f}-\mathbf{v}^{s}\right)-c_{\phi}\frac{D^{s}\phi}{Dt}=0\end{aligned} \qquad \text{(A.50)}$$

$$\begin{aligned}&\rho\left(-\frac{D\psi}{Dt}-\eta\frac{D\theta}{Dt}\right)-\sum_{\alpha}\operatorname{div}\left(\rho^{\alpha}\psi^{\alpha}\mathbf{w}^{\alpha}\right)-\nabla\theta\cdot\left(\sum_{\alpha}\rho^{\alpha}\eta^{\alpha}\mathbf{w}^{\alpha}\right)\\ &\quad+\sum_{\alpha}\operatorname{tr}\left(\mathbf{T}^{\alpha}\mathbf{L}^{\alpha}\right)-\frac{1}{2}m^{f}\left(\mathbf{v}^{f}-\mathbf{v}^{s}\right)\cdot\left(\mathbf{v}^{f}-\mathbf{v}^{s}\right)-\mathbf{I}^{f}\cdot\left(\mathbf{v}^{f}-\mathbf{v}^{s}\right)-\frac{\mathbf{q}\cdot\nabla\theta}{\theta} \quad = \quad \mathcal{D}\end{aligned} \qquad \text{(A.51)}$$

Solving for the stress power term, $\sum_{\alpha}\operatorname{tr}\left(\mathbf{T}^{\alpha}\mathbf{L}^{\alpha}\right)$, in the entropy equation and substituting it in the mixture energy balance equation, the following is obtained

$$\begin{aligned}&\rho\frac{D\varepsilon}{Dt}+\sum_{\alpha}\operatorname{div}\left(\rho^{\alpha}\varepsilon^{\alpha}\mathbf{w}^{\alpha}\right)+\operatorname{div}\mathbf{q}-\rho r-\mathcal{D}-\frac{\mathbf{q}\cdot\nabla\theta}{\theta}\\ &+\rho\left(-\frac{D\psi}{Dt}-\eta\frac{D\theta}{Dt}\right)-\sum_{\alpha}\operatorname{div}\left(\rho^{\alpha}\psi^{\alpha}\mathbf{w}^{\alpha}\right)-\nabla\theta\cdot\sum_{\alpha}\rho^{\alpha}\eta^{\alpha}\mathbf{w}^{\alpha}=0\end{aligned} \qquad \text{(A.52)}$$

Substituting equations (A.31), (A.48), and (A.49) in equation (A.52) results in

$$\begin{aligned}&\rho\theta\frac{D}{Dt}\left(-\frac{\partial\psi}{\partial\theta}\right)+\theta\sum_{\alpha}\operatorname{div}\left(\rho^{\alpha}\eta^{\alpha}\mathbf{w}^{\alpha}\right)+\operatorname{div}\mathbf{q}-\rho r\\ &+\mathbf{I}^{f,(d)}\cdot\left(\mathbf{v}^{f}-\mathbf{v}^{s}\right)-\operatorname{tr}\left(\mathbf{T}^{f,(d)}\mathbf{L}^{f}\right)-c_{\phi}^{(d)}\frac{D^{s}\phi}{Dt} \quad = \quad 0\end{aligned} \qquad \text{(A.53)}$$

The divergence term is expanded analogously to equation (A.33) and constitutive equation for the constituents' entropy from equation (4.54) is substituted to obtain

$$\begin{aligned}&-\rho\theta\frac{D}{Dt}\left(\frac{\partial\psi}{\partial\theta}\right)-\theta\sum_{\alpha}\left[\left(\frac{\partial\psi^{\alpha}}{\partial\theta}-\frac{\partial\psi}{\partial\theta}\right)\left(\rho^{\alpha}\operatorname{tr}\mathbf{L}^{\alpha}+\mathbf{w}^{\alpha}\cdot\nabla\rho^{\alpha}\right)+\rho^{\alpha}\mathbf{w}^{\alpha}\cdot\nabla\left(\frac{\partial\psi^{\alpha}}{\partial\theta}\right)\right]\\ &+\operatorname{div}\mathbf{q}-\rho r+\mathbf{I}^{f,(d)}\cdot\left(\mathbf{v}^{f}-\mathbf{v}^{s}\right)-\operatorname{tr}\left(\mathbf{T}^{f,(d)}\mathbf{L}^{f}\right)-c_{\phi}^{(d)}\frac{D^{s}\phi}{Dt} \quad = \quad 0\end{aligned} \qquad \text{(A.54)}$$

Furthermore, the time derivative and gradient of the partial derivative of Helmholtz free energy with respect to temperature are expanded using chain rule in view of (4.45)-(4.47), in a way analogous to (A.34)-(A.36), to yield

$$
\begin{aligned}
&-\rho\theta\left(\sum_{\alpha}\frac{\partial^2\psi}{\partial\rho^\alpha\partial\theta}\frac{D\rho^\alpha}{Dt}+\frac{\partial^2\psi}{\partial\theta^2}\frac{D\theta}{Dt}+\frac{\partial^2\psi}{\partial\phi\partial\theta}\frac{D\phi}{Dt}+\frac{\partial^2\psi}{\partial\mathbf{F}^s\partial\theta}:\frac{D\mathbf{F}^s}{Dt}\right)\\
&-\theta\sum_{\alpha}\left[\left(\frac{\partial\psi^\alpha}{\partial\theta}-\frac{\partial\psi}{\partial\theta}\right)\left(\rho^\alpha\mathrm{tr}\mathbf{L}^\alpha+\mathbf{w}^\alpha\cdot\nabla\rho^\alpha\right)+\rho^\alpha\mathbf{w}^\alpha\cdot\left(\nabla\rho^\alpha\frac{\partial^2\psi^\alpha}{\partial\rho^\alpha\partial\theta}+\nabla\theta\frac{\partial^2\psi^\alpha}{\partial\theta^2}\right)\right]\\
&-\theta\rho^s\mathbf{w}^s\cdot\left(\nabla\phi\frac{\partial^2\psi^s}{\partial\phi\partial\theta}+\nabla\mathbf{F}^s:\frac{\partial^2\psi^s}{\partial\mathbf{F}^s\partial\theta}\right)\\
&+\mathrm{div}\mathbf{q}-\rho r+\mathbf{I}^{f,(d)}\cdot\left(\mathbf{v}^f-\mathbf{v}^s\right)-\mathrm{tr}\left(\mathbf{T}^{f,(d)}\mathbf{L}^f\right)-c_\phi^{(d)}\frac{D^s\phi}{Dt} \quad = \quad 0
\end{aligned}
\tag{A.55}
$$

Substituting (A.37)-(A.39), as well as the analogous identity for the mixture time derivative of the temperature term and gathering common terms yields

$$
\begin{aligned}
&-\rho\theta\frac{\partial^2\psi}{\partial\theta^2}\frac{D^s\theta}{Dt}+\theta\mathrm{tr}\left\{\frac{\partial}{\partial\theta}\left[-\rho\mathbf{F}^s\left(\frac{\partial\psi}{\partial\mathbf{F}^s}\right)^T+\rho^s\left(\rho\frac{\partial\psi}{\partial\rho^s}-\psi^s+\psi\right)\mathbf{1}\right]\mathbf{L}^s\right\}\\
&+\theta\mathrm{tr}\left\{\frac{\partial}{\partial\theta}\left[\rho^f\left(\rho\frac{\partial\psi}{\partial\rho^f}-\psi^f+\psi\right)\mathbf{1}\right]\mathbf{L}^f\right\}\\
&+\theta\frac{\partial^2}{\partial\theta^2}\left[\rho\psi\mathbf{w}^s-\rho^s\psi^s\mathbf{w}^s-\rho^f\psi^f\mathbf{w}^f\right]\cdot\nabla\theta\\
&-\rho\theta\frac{\partial}{\partial\theta}\left(\frac{\partial\psi}{\partial\phi}\right)\frac{D^s\phi}{Dt}-\rho\theta\sum_{\alpha}\frac{\partial^2\psi}{\partial\rho^\alpha\partial\theta}m^\alpha\\
&+\mathrm{div}\mathbf{q}-\rho r+\mathbf{I}^{f,(d)}\cdot\left(\mathbf{v}^f-\mathbf{v}^s\right)-\mathrm{tr}\left(\mathbf{T}^{f,(d)}\mathbf{L}^f\right)-c_\phi^{(d)}\frac{D^s\phi}{Dt}\\
&+\theta\sum_{\alpha}\frac{\partial}{\partial\theta}\left[\rho\frac{\partial\psi}{\partial\rho^\alpha}-\psi^\alpha+\psi-\rho^\alpha\frac{\partial\psi^\alpha}{\partial\rho^\alpha}\right]\mathbf{w}^\alpha\cdot\nabla\rho^\alpha\\
&+\theta\frac{\partial}{\partial\theta}\left(\rho\frac{\partial\psi}{\partial\phi}-\rho^s\frac{\partial\psi^s}{\partial\phi}\right)\mathbf{w}^s\cdot\nabla\phi+\theta\mathbf{w}^s\cdot\nabla\mathbf{F}^s:\frac{\partial}{\partial\theta}\left(\rho\frac{\partial\psi}{\partial\mathbf{F}^s}-\rho^s\frac{\partial\psi^s}{\partial\mathbf{F}^s}\right) \quad = \quad 0
\end{aligned}
\tag{A.56}
$$

Using (A.42)-(A.45), as well as the relations of the constituents' mass transfer rate m^α in (4.32) and (4.39), the last three terms vanish (analogously to the derivation in Appendix B), and the remaining terms are rewritten to give

$$
\begin{aligned}
&-\rho\theta\frac{\partial^2\psi}{\partial\theta^2}\frac{D^s\theta}{Dt}+\theta\mathrm{tr}\left\{\frac{\partial}{\partial\theta}\left[-J^{-1}\mathbf{F}^s\left(\frac{\partial\hat{\psi}^s}{\partial\mathbf{F}^s}\right)^T\right]\mathbf{L}^s\right\}\\
&+\theta\mathrm{tr}\left\{\frac{\partial}{\partial\theta}\left[(\rho^f)^2\frac{\partial\psi^f}{\partial\rho^f}\mathbf{1}\right]\mathbf{L}^f\right\}+\theta\frac{\partial^2}{\partial\theta^2}\left[\rho\psi\mathbf{w}^s-\rho^s\psi^s\mathbf{w}^s-\rho^f\psi^f\mathbf{w}^f\right]\cdot\nabla\theta\\
&+\theta\frac{\partial}{\partial\theta}\left(-\beta_\phi\left(\psi^f+\rho^f\frac{\partial\psi^f}{\partial\rho^f}\right)-J^{-1}\frac{\partial\hat{\psi}^s}{\partial\phi}\right)\frac{D^s\phi}{Dt}\\
&+\mathrm{div}\mathbf{q}-\rho r+\mathbf{I}^{f,(d)}\cdot\left(\mathbf{v}^f-\mathbf{v}^s\right)-\mathrm{tr}\left(\mathbf{T}^{f,(d)}\mathbf{L}^f\right)-c_\phi^{(d)}\frac{D^s\phi}{Dt}\quad=\quad 0
\end{aligned}
\tag{A.57}
$$

By the definitions of the constituent and mixture time derivatives in (4.12) and (4.14), as well as the definition of the relative velocity in (4.8), we have that

$$
\frac{D^s\theta}{Dt}=\frac{D^f\theta}{Dt}+\left(\mathbf{w}^s-\mathbf{w}^f\right)\cdot\nabla\theta
\tag{A.58}
$$

Using (A.29) and (A.58), equation (A.57) can be rewritten as

$$
\begin{aligned}
&-\rho^s\theta\frac{\partial^2\psi^s}{\partial\theta^2}\frac{D^s\theta}{Dt}-\rho^f\theta\frac{\partial^2\psi^f}{\partial\theta^2}\frac{D^f\theta}{Dt}\\
&+\theta\mathrm{tr}\left\{\frac{\partial}{\partial\theta}\left[-J^{-1}\mathbf{F}^s\left(\frac{\partial\hat{\psi}^s}{\partial\mathbf{F}^s}\right)^T\right]\mathbf{L}^s\right\}+\theta\mathrm{tr}\left\{\frac{\partial}{\partial\theta}\left[(\rho^f)^2\frac{\partial\psi^f}{\partial\rho^f}\mathbf{1}\right]\mathbf{L}^f\right\}\\
&+\theta\frac{\partial}{\partial\theta}\left(-\beta_\phi\left(\psi^f+\rho^f\frac{\partial\psi^f}{\partial\rho^f}\right)-J^{-1}\frac{\partial\hat{\psi}^s}{\partial\phi}\right)\frac{D^s\phi}{Dt}\\
&+\mathrm{div}\mathbf{q}-\rho r+\mathbf{I}^{f,(d)}\cdot\left(\mathbf{v}^f-\mathbf{v}^s\right)-\mathrm{tr}\left(\mathbf{T}^{f,(d)}\mathbf{L}^f\right)-c_\phi^{(d)}\frac{D^s\phi}{Dt}\quad=\quad 0
\end{aligned}
\tag{A.59}
$$

Using the constitutive equations for the quasi-conservative contributions $\mathbf{T}^{s,(q)}$, $\mathbf{T}^{f,(q)}$, and $c_\phi^{(q)}$ from equations (4.52), (4.56), and (4.58) yields

$$
\begin{aligned}
&-\rho^s\theta\frac{\partial^2\psi^s}{\partial\theta^2}\frac{D^s\theta}{Dt}-\rho^f\theta\frac{\partial^2\psi^f}{\partial\theta^2}\frac{D^f\theta}{Dt}+\mathrm{div}\mathbf{q}-\rho r+\mathbf{I}^{f,(d)}\cdot\left(\mathbf{v}^f-\mathbf{v}^s\right)\\
&-\mathrm{tr}\left\{\theta\frac{\partial\mathbf{T}^{s,(q)}}{\partial\theta}\mathbf{L}^s\right\}-\mathrm{tr}\left\{\left(\theta\frac{\partial\mathbf{T}^{f,(q)}}{\partial\theta}+\mathbf{T}^{f,(d)}\right)\mathbf{L}^f\right\}-\left(\theta\frac{\partial c_\phi^{(q)}}{\partial\theta}+c_\phi^{(d)}\right)\frac{D^s\phi}{Dt}\quad=\quad 0
\end{aligned}
\tag{A.60}
$$

A.4 References

[1] K.R. Rajagopal, L. Tao, Mechanics of Mixtures, World Scientific, 1995.

[2] R. Hall, K. Rajagopal, Diffusion of a fluid through an anisotropically chemically reacting thermoelastic body within the context of mixture theory, Math. Mech. Solids. 17 (2012) 131–164.