Anisotropy and heterogeneity in finite deformation: resolving versus upscaling

Marcin Dabrowski

Thesis submitted for the degree of Philosophiae Doctor

Department of Physics, University of Oslo, Norway

April 2008

Preface

Some years back I was faced with the “what to do next?” dilemma. I then decided that a google search for “Yuri Podladchikov” would be the best solution to my problem. Soon after, I spent a month of an intense work at PGP following Yuri's finite element course. Half a year later, on the Fool's Day 2005 I started my PhD in Oslo. For the last three years, I have had a great pleasure to be a student of an extraordinary mathematician-physicist and a geologist at the same time. Yuri has opened my eyes to many scientific topics and I am convinced that it is just the beginning of what I can learn from him. I admire his incredible amount of enthusiasm for science and I have drawn upon this volcano of energy a lot. I would very much like to thank him for providing me with a constant influx of scientific ideas during my PhD. Many of these discussions started in Yuri's office and continued until late night at various spots in Oslo and surroundings. Admittedly, some of them would turn out deeply philosophical. Finally, I am indebted to Yuri for giving me a freedom to walk my first independent steps in the World of Science.

I had a pleasure to meet Dani Schmid for the first time during my short stay at PGP in 2004. When I was back in Oslo to start my PhD, it turned out that a Swiss and a Pole can get on quite well. From the very beginning, Dani generously shared his extensive knowledge of computing and all kinds of inclusion problems, and I am very grateful for his generosity. During my PhD, he has been very supportive of all my developments and I owe him many thanks in this respect. He is an excellent supervisor and I have learned from him not only on a professional account. I value his friendship and his mentoring very much and I am eagerly looking forward to our future collaboration.

Ray Fletcher is a very special person to me. I would like to thank him for his caring interest in my work. I am very much impressed and stimulated by Ray’s work. Having an opportunity to talk to him at PGP was great. Although very intense and mind breaking, these discussions helped me to look at problems from different angles, and to search for deeper motivations and insights

I know Marcin Krotkiewski longer than anybody else at PGP, yet we have not managed to bore each other (or should I speak for myself?). Since we are neighbors, we usually walk together back and forth between our homes and PGP, having plenty of time to discuss things. I would especially like to thank him for many answers we have found together, due to his razor-

sharp questioning mind. On a professional account, I am indebted to Marcin for his patience in explaining to me the high-performance computing issues.

I am thankful to Ebbe Hartz for taking me to the field campaign in Greenland. This was a fantastic scientific experience, great adventure, and I will always remember the striking beauty of this place. I would also like to thank Torgeir Anderson for taking me to several field trips in Norway and for his explanations of the mysteries hidden in the majestic Norwegian Caledonides.

PGP is a great place to work, with so many diverse minds around always ready for discussions. I have benefited a lot from its great scientific atmosphere and the opportunities I have been given here. In addition, it is a fantastic social environment and I would like to thank all the PGPeans for making my PhD time enjoyable. Whether during Healthy Sundays at Yuri's place, partying, skiing, sailing, playing cards or PS3, or simply going out together, it was always a great pleasure to have such a fantastic company around me.

I would especially like to thank my parents for their support, love and believing in me over the years.

I dedicate this work to my wife Ania. I am grateful to you for so many things. I thank you for your trust in me and your courage to start a new life following me to the north on a Fool's Day. Your unconditional love and support navigated me over the years. I owe you a lot (although, as you say, long sandy beach holidays would be enough) for your incredible patience during my PhD.

May 2008, Oslo

Marcin Dabrowski

Scientific Advisors:

Associate Professor Daniel W. Schmid

Professor Yuri Yu. Podladchikov

Contents

Paper summary

This thesis consists of three scientific papers in the main body and two papers in the appendix. The main body papers are arranged into a suite consisting of a tool description (Paper 1) and two studies concerning anisotropy evolution in systems subject to large deformation (Paper 2 & 3).

The first paper presents details of the developed finite element model implementation that allows for solving problems involving more than million degrees of freedom on a desktop computer on a minute scale. This tool is used in the subsequent studies to directly resolve analyzed models without introducing uncertainties related to a priori homogenization. In the second paper, we analyze strain pattern development around a rigid circular inhomogeneity embedded in a laminated host. Results obtained by direct resolution of isotropic layers are juxtaposed with a system evolution where the host is modeled as an equivalent anisotropic fluid. Numerical results are augmented by an analytical solution describing internal and external fields for an elliptical inclusion in an anisotropic medium subject to a uniform load. The original solution that is suitable for compressible elastic materials and arbitrary material parameters contrasts has been adjusted for the study purposes. In the third paper, we have developed a constitutive model for a two-phase composite that incorporates an effective anisotropy development due to an emerging shape preferred orientation. The final model combines a model of the shape evolution of an inclusion in an anisotropic host and an effective property scheme for a composite comprising aligned stretched inclusions. We have derived these models based on the analytical solution for an elliptical inhomogeneity embedded in a homogenous anisotropic host. Model predictions have been validated by finite element calculations.

The papers in the appendix are not directly related to the main topic of the anisotropy. Here, we study three-dimensional fold patterns (Paper 4) and analyze alternating direction implicit numerical schemes (Paper 5). Marcin Krotkiewski is the first author of the last paper.

A manuscript documenting my work with the analytical solution for a compressible elliptical inclusion in an anisotropic host is under preparation. During the duration of my Ph. D., I have worked with the internal and external Eshelby's solution for ellipsoidal inclusions. Ulrike Exner from Vienna University is working on a paper, where we have applied the limiting case of an elliptical inclusion embedded in a 3D matrix to model flanking structures. I have joined the Greenland field campaign organized in 2006 by Ebbe Hartz and a project concerning overpressure related to melting. I have contributed to a plasticity benchmark by Victorya Yarushina

with a finite element code based on the consistent tangent approach. Finally, a project with Espen Jettestuen devoted to Stokesian dynamics of rigid inclusions under a shear is under development. Below short descriptions of the scientific papers included in the thesis are provided.

Paper 1: MILAMIN: MATLAB-based FEM solver for large problems

This paper presents our implementation of a thermal and incompressible Stokes finite element solver in two dimensions. In our modeling, we employ unstructured computational meshes that are particularly useful in resolving complex geometries of heterogeneous geological materials. This ability of directly representing complicated and evolving model geometry, together with our tools that allow for an automatic mesh reconstruction during simulations, is crucial in the context of our later applications. Here, we are mainly focused on code efficiency. Our simulations involving numerous inclusions and complex anisotropy structures require exceptionally high resolutions. In addition, a large number of simulation steps is needed to accurately integrate system evolution to large strains that are often encounter in geological problems.

Therefore, we have set the goal of implementing an unstructured FEM code that is capable of solving thermal and mechanical problems involving one million degrees of freedom on a desktop computer in a minute. In the paper, we describe in details how we have accomplished this task using MATLAB environment. Firstly, we establish a standard version that already combines all the state of the art components required in a finite element implementation. This includes among others: efficient pre-processing, fast matrix assembly, exploiting matrix symmetry for storage, and employing the best available direct solver and reordering packages. Our particular formulation allows us to use the efficient Cholesky factorization also for the incompressible Stokes problem. In addition, we present an efficient implementation of boundary conditions. The clear structure of the code serves educational purposes well. Furthermore, in our optimized version we have improved the efficiency of the stiffness matrix calculations by minimizing the ratio of overhead to computation. Another priority was to avoid unnecessary data transfers and promote cache reuse, as memory speed is a major bottleneck on current computer architectures.

The MATLAB environment is especially suited for code development and testing. By paying attention to the strategies outlined in this article, MATLAB-based MILAMIN can also be used as a production tool for the analysis of two dimensional problems with millions of unknowns

within minutes. We provide the complete MILAMIN source code that can be downloaded as an electronic supplement to this paper.

Paper 2: Structure Development around a Rigid Circular Inclusion in an Anisotropic Host Subject to Simple Shear

In this paper, we discuss the structure formation around an inclusion embedded in a layered medium during a finite deformation. We study two different model categories: an explicit layering present in the host and an anisotropic matrix case.

Firstly, we analyze an instantaneous flow pattern around a rigid circular inclusion subject to a simple shear in a layering direction. Provided the host anisotropy factor corresponds to a viscosity ratio of weak and strong layers, the results obtained by the two approaches converge in a limit of thin layers. An analytical elastic solution describing internal and external fields for an elliptical inclusion embedded in an anisotropic host subject to a uniform far field loads has been introduced in details. We have adjusted the solution to a rigid circular inclusion case and for incompressible viscous materials. Numerical and analytical flow predictions coincide in a small inclusion limit, where effects related to proximity of computational model boundaries vanish. Nevertheless, these boundary effects lead to a decrease of the inclusion rotation rate, grow in importance with the anisotropy strength and in natural systems may significantly affect the inclusion motion in a presence of shear zone walls.

Our study of the instantaneous flow pattern provides a starting point for finite deformation simulations, where we have subjected the inclusion-host system to a simple shear of a magnitude five. Structure formation in the vicinity of the inclusion is automatically taken care of in the case of explicitly layered host and we have utilized a director field to model the anisotropy reorientation. Due to the structural development, the instantaneous flow pattern predicted for a homogeneous layering becomes invalid already after a small deformation. At the beginning, this is manifested by breaking of an orthotropic symmetry present in the initial flow field. A presence of the host layering has proven to have a pronounced impact on the inclusion rotation rate. Strong anisotropy leads to substantially reduced rotation rates once a shear strain magnitude of two is reached. This effect and related developments in the structural development in the matrix may yield finite strain structures that appear to be the result of substantially lower strains. The comparison of layered and anisotropic hosts reveals that the effective approach is good when inclusion motion is considered. However, the detailed structural evolution in a lay-

ered matrix, e.g. development of rootless folds, thickening of strong layers, deformation localization in weak layers can only be studied if the layers are explicitly resolved.

This study shows how inhomogeneity and anisotropy interact when subject to large deformation. The structural memory of the anisotropic host material proves essential with respect to the inclusion motion leading to the effective inclusion stabilization. The up-scaled behavior of the layered host materials can be approximated by the anisotropy in the thin layers limit. The presence of the coarse host layering results in a limited structure development. These results bear important implications for the strain analysis relying on the inclusion-host systems.

Paper 3: Structural anisotropy development of a two-phase composite subject to large deformation

This paper is devoted to the structural development and related mechanical effects in a two-phase composite characterized by the inclusion-host type of the geometry. Shape preferred orientation develops in an aggregate consisting of initially circular weak inclusions that has been subject to a shear strain. The overall mechanical response of such a medium becomes effectively anisotropic. An extreme case is provided by a perfectly layered medium, where shear and normal viscosity components correspond to the lower and upper theoretical bounds imposed on the effective viscosity. In this study, we examine shape evolution of an isolated inclusion embedded in an anisotropic host and derive a scheme predicting the effective normal and shear viscosity of a composite comprising aligned elliptical inclusions. Next, these models are combined into a set of ordinary differential equations that describe mechanical evolution of a composite rock for arbitrary deformation paths. The analytical predictions are compared to results of our finite element modeling, where we have directly resolved the shape and overall viscosity evolution for inclusion aggregates under a pure and simple shear.

We show that the shape evolution of an isolated inclusion is affected by a presence of the anisotropy in the host. In the pure shear case, an analytical formula describing the aspect ratio evolution for an initially circular inclusion has been derived. In comparison to the isotropic case, it predicts an enhanced inclusion stretching. In the simple shear case, the inclusion stretch is again stronger, but the ellipticity-inclination path is affected to a lesser degree. These results have been obtained assuming the host anisotropy of a constant strength (unrelated to the inclusion aspect ratio) and an orientation mimicking the inclusion axis.

The effective mechanical properties of inclusion assemblages have been studied numerically. Examined parameters include inclusion concentration, ellipticity and inclusion-host viscosity

ratio. We find that the effective viscosity varies non-linearly with inclusion concentration, but always falls within the theoretical bounds. For a given phase concentration, the results are significantly different depending on if this phase forms the matrix or the inclusions. The numerical results show that the effective properties of composites with phases interchanged are approximately reciprocal. Introducing aligned elongated inclusions in the host results in an overall anisotropy. Even for strongly elongated shapes, the computed anisotropy factors are significantly lower than in the laminate case, especially for large viscosity ratios. In the anisotropic case, the alternate effective viscosity components are approximately reciprocal after the phase swap. By varying the inclusion-host viscosity ratio, we find that the effective viscosities assume finite values in limiting rigid inclusion or incompressible void cases, even for concentrations of fifty percent. Details of the inclusion configuration exert a limited influence on the effective properties.

The numerical results have been utilized to validate analytical schemes predicting effective properties of such composites. The model developed by Fletcher (2004) in the spirit of the self-consistent averaging (SCA) technique provides a valuable contribution in this direction. It is free of phenomenological parameters and based on inclusion concentration, ellipticity and inclusion-host viscosity ratio predicts the overall normal and shear viscosity. We find that these predictions provide a good approximation to the numerical data up to twenty percent concentrations. This scheme is inherently symmetric with respect to a phase swap and cannot predict the observed dichotomy of the effective properties expressed at most for composites of the equal phase abundances. The other deficiency of this model is an unbound prediction for a limiting rigid inclusion case at the concentration of fifty percent and lack of it for higher concentrations in this case. We have derived another scheme based on the differential effective medium (DEM) approach that is suitable for the anisotropic composite case. The model is free from the shortcomings of the SCA scheme and provides a good prediction to the numerical data for a wide range of concentrations. We speculate that the SCA model is more suitable for multi-grain aggregates and becomes superior over the DEM scheme in this case.

The shape evolution model and DEM scheme have been combined into a set of equations that describe the shape and mechanical evolution of a two-phase composite rock under arbitrary load conditions. In particular, evolution of the stress component needed to drive the system can be extracted from the model. We have examined the weak inclusion case in a pure and simple shear. The model predicts a monotonous hardening in the pure shear case. In a simple shear, an initial hardening phase is followed by a pronounced weakening after a shear strain of magni-

tude one. We have performed corresponding numerical simulations for different inclusion concentrations. Despite complex shapes developing due to inclusion interactions, the analytical model captures the effective stress evolution well. Discrepancies are only observed in pure shear for densely packed inclusions and after significant strain. This has been attributed to a formation of localized deformation bands that may lead to an overall weakening. In a simple shear, S-C structures have formed and inclusion shapes are non-elliptical, but the analytical model provides a surprisingly good approximation to the overall stress evolution.

We conclude that our model is capable of predicting the overall shape and mechanical evolution of a two-phase composite, even for high concentration and large strains. We have observed localization phenomena on a grain scale in the numerical simulations. However, our overall constitutive model cannot exclusively explain localization on a larger scale, as long as a disconnected topology of an inclusion phase is maintained.

In this study, we have shown how the anisotropy originates and evolves in an initially isotropic but heterogeneous rock subject to finite deformation. The derived up-scaled analytical model has been validated by the numerical simulations. The predictions are accurate for high inclusion concentrations and large strains. Our finite strain analytical model improves the rheological laws employed in large scale tectonic simulations. The transient effects observed in the deformation experiments of heterogeneous materials can be understood better in our model framework. Structural development and related strain weakening is shown insufficient to explain the strain localization for linear materials in the ductile regime.

Paper 4 (Appendix): Evolution of large amplitude 3D fold patterns: a FEM study

We study fold patterns developing in three dimensions under different load conditions. The problem of the fold interference has been addressed many times in the literature, but mostly from a kinematic perspective. Analytical mechanical models proposed for three dimensional setups are limited to small amplitudes and cannot resolve the interaction of individual folds occurring at later development stages. Here, we present the results of finite element simulations, where we subject a perturbed planar inhomogeneity to perpendicular horizontal loads in different proportions. Initial amplification of the instability follows the analytical predictions derived by Fletcher (1991). We show that the fold pattern of a layer subject to equal loads is not a simple dome and basin structure (egg carton). Instead, a complicated hinge network develops, where individual folds are characterized by the aspect ratios of two. Finally, we specu-

late that the ellipticity of the folds could be utilized as a measure of the load proportions during the deformation. This paper has a strong technical component as the modeling required large numerical resolutions to capture sufficiently representative volumes and involved the unstructured computational meshes that allow integrating system evolution to large strains. Studying the incompressible Stokes problem for heterogeneous materials using body-fitting tetrahedral meshes in three dimensions necessitates development of tailored numerical schemes. We employ the minimal residual iterative method to solve a resulting linear system of equations and parallel computing is involved at this stage. Our code has been tested on a thousand CPUs and shows superb scaling.

Paper 5 (Appendix): Fractional Steps Methods for Transient Problems on Commodity Computer Architectures

In this paper, we investigate alternating direction implicit (ADI) methods and show an efficient implementation for shared memory machines. The ADI methods are characterized by a low operation count and our implementation is only limited by a memory bandwidth. We have been able to solve transient parabolic and hyperbolic problems involving a billion degrees of freedom. We pay a special attention to a material heterogeneity that is inherent to all geological models and is challenging from a numerical perspective. In addition, a possibility to employ the ADI schemes for solving elliptic problems is discussed. We recognize that boundary condition treatment becomes crucial in this context. Finally, two large-scale applications are presented that rely on real-world reservoir data.

Paper 1: MILAMIN: MATLAB-based FEM solver for large problems

Submitted to Geochemistry, Geophysics, Geosystems by

M. Dabrowski, M. Krotkiewski and D.W. Schmid

Abstract

The finite element method (FEM) combined with unstructured meshes forms an elegant and versatile approach capable of dealing with the complexities of problems in earth science. Practical applications often require high resolution models that necessitate advanced computational strategies. We therefore developed MILAMIN ("MILlion A MINute"), an efficient MATLAB implementation of FEM that is capable of setting up, solving, and post-processing two-dimensional problems with one million unknowns in one minute on a modern desktop computer. MILAMIN allows the user to achieve numerical resolutions that are necessary to resolve the heterogeneous nature of geological materials. In this paper we provide the technical knowledge required to develop such models without the need to buy a commercial FEM package, programming compiler-language code, or hiring a computer specialist.

It has been our special aim that all the components of MILAMIN perform efficiently, providing a well performing package. While some of the components rely on readily available routines, we develop others from scratch and make sure that all of them work together efficiently. One of the main technical focuses of this paper is the optimization of the global matrix computations. The performance bottlenecks of the standard FEM algorithm are analyzed. An alternative approach is developed that sustains high performance for any system size. Applied optimizations eliminate BLAS drawbacks when multiplying small matrices, reduce operation count and memory requirements when dealing with symmetric matrices, and increase data transfer efficiency by maximizing cache reuse. Applying loop interchange allows us to use BLAS on large matrices. In order to avoid unnecessary data transfers between RAM and CPU cache we introduce loop blocking.

The optimization techniques are useful in many areas as demonstrated with our MILAMIN applications for thermal and incompressible flow (Stokes) problems. We use these to provide performance comparisons to other open source as well as commercial packages and find that

MILAMIN is amongst the best performing solutions, both in terms of speed and memory usage.

The corresponding MATLAB source code for the entire MILAMIN, including input generation, FEM solver, and post-processing, is available from the authors (www.milamin.org) and can be downloaded as an electronic supplement.

Introduction

Geological systems are often formed by multi-physics processes interacting on many temporal and spatial scales. Moreover, they are heterogeneous and exhibit large material property contrasts. In order to understand and decipher these systems numerical models are frequently employed. Appropriate resolution of the behavior of these heterogeneous systems, without the (over-) simplifications of a priori applied homogenization techniques, requires numerical models capable of efficiently and accurately dealing with high resolution, geometry-adapted meshes. These criterions are usually used to justify the need for special purpose software (commercial FEM packages) or special code development in high performance compiler-languages such as C or FORTRAN. General purpose packages like MATLAB are usually considered not efficient enough for this task. This is reflected in the current literature. MATLAB is treated as an educational tool that allows for fast learning when trying to master numerical methods, e.g. the books by Kwon and Bang [2000], Elman et al. [2005], and Pozrikidis [2005]. MATLAB also facilitates very short implementations of numerical methods that give overview and insight, which is impossible to obtain when dealing with closed black-box routines, e.g. finite elements on 50 lines [*Alberty, et al.*, 1999], topology optimization on 99 lines [*Sigmund*, 2001], and mesh generation on one page [*Persson and Strang*, 2004]. However, while advantageous from an educational standpoint, these implementations are usually rather slow and run at a performance that is a fraction of the peak performance of modern computers. Therefore, the usual approach is to use MATLAB for prototyping, development, and testing only. This is followed by an additional step where the code is manually translated to a compiler language in order to achieve the memory and CPU efficiency required for high resolution models.

This paper presents the outcome of a project called “MILAMIN - MILlion A MINute” aimed at developing a MATLAB based FEM package capable of pre-processing, processing, and post-processing an unstructured mesh problem with one million degrees of freedom in two dimensions within one minute on a commodity personal computer. Choosing a native MATLAB implementation allows simultaneously for educational insight, easy access to computational

libraries and visualization tools, rapid prototyping and development, as well as actual two-dimensional production runs. Our standard implementation serves to provide educational insight into subjects such as implementation of the numerical method, efficient use of the computer architecture and computational libraries, code structuring, proper data layout, and solution techniques. We also provide an optimized FEM version that increases the performance of production runs even further, but at the cost of code clearness.

The MATLAB code implementing the different approaches discussed here is available from the authors (www.milamin.org) and can be downloaded as an electronic supplement to this paper.

Code Overview

A typical finite element code consists of three basic components: pre-processor, processor, post-processor. The main component is the processor, which is the actual numerical model that implements a discretized version of the governing conservation equations. The pre-processor provides all the input data for the processor and in the present case the main work is to generate an unstructured mesh for a given geometry. The task of the post-processor is to analyze and visualize the results obtained by the processor. These three components of MILAMIN are documented in the following sections.

Pre-Processor

Geometrically complex problems promote the use of interface adapted meshes, which accurately resolve the input geometry and are typically created by a mesh generator that automatically produces a quality mesh. The drawback of this approach is that one cannot exploit the advantages of solution strategies for structured meshes, such as operator splitting methods [e.g. ADI, *Fletcher*, 1997] or geometric multigrid [*Wesseling*, 1992] for efficient computation.

A number of mesh generators are freely available. However, none of these are written in native MATLAB and fulfill the requirement of automated quality mesh generation for multiple domains. DistMesh by Persson and Strang [2004] is an interesting option as it is simple, elegant, and written entirely in MATLAB. However, lack of speed and proper multi-domain support renders it unsuitable for a production code with the outlined goals. The mesh generator chosen is Triangle developed by Shewchuk [2007]. Triangle is extremely versatile and stable, and consists of one single file that can be compiled into an executable on all platforms with a standard C compiler. We choose the executable-based file I/O approach, which has the advantage that we can always reuse a saved mesh. The disadvantage is that the ASCII file I/O provided

by Triangle is rather slow, which, however, can be overcome by adding binary file I/O as described in the instructions provided in the MILAMIN code repository.

Processor

FEM Outline

In this paper we show two different physical applications of MILAMIN: steady state thermal problems and incompressible Stokes flow (referred to as mechanical problem). This section provides an outline of the governing equations and their corresponding FEM formulation. The numerical implementation and performance discussions follow in subsequent chapters.

Thermal problem

The strong form of the steady state thermal diffusion in the two-dimensional domain Ω is

$$\frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right)+\frac{\partial}{\partial y}\left(k\frac{\partial T}{\partial y}\right)=0 \text{ in } \Omega \tag{1}$$

where T is temperature, k is the conductivity, x and y are Cartesian coordinates. The boundary Γ of Ω is divided into two non-intersecting parts: $\Gamma=\Gamma_N \bigcup \Gamma_D$. Zero heat-flux is specified on Γ_N (Neumann boundary condition) and temperature $\overline{T}$ is prescribed on Γ_D (Dirichlet boundary condition).

The FEM is based on the weak (variational) formulation of partial differential equations, taking an integral form. For the purpose of this paper we only introduce the basic concepts of this method that are important from an implementation viewpoint. A detailed derivation of the finite element method and a description of the weak formulation of PDEs can be found in textbooks, e.g. Bathe [1996], Hughes [2000], and Zienkiewicz and Taylor [2000].

In FEM, the domain Ω is partitioned into non-overlapping element subdomains Ω_e, i.e. $\Omega=\bigcup_{e=1}^{nel}\Omega_e$, where *nel* denotes the number of elements. The basic two-dimensional element is a triangle. In the case of the thermal problem discrete temperature values are defined for the nodal points, which can be associated with element vertices, located on its edges, or even reside inside the elements. Introducing shape functions N_i that interpolate temperatures from the nodes T_i to the domains of neighboring elements, an approximation to the temperature field $\tilde{T}$ in Ω is defined as

$$\tilde{T}(x,y) = \sum_{i=1}^{nnod} N_i(x,y) T_i \tag{2}$$

where *nnod* is the number of nodes in the discretized domain.

Based on the weak formulation that takes the form of an integral over Ω, the problem can now be stated in terms of a system of linear equations. From a computational point of view it is beneficial to evaluate this integral as a sum of integrals over each element Ω_e. A single element contribution, the so-called "element stiffness matrix", to the global system matrix in the Galerkin approach for the thermal problem is given by

$$\mathbf{K}_{ij}^{e} = \iint_{\Omega_e} k^e \left(\frac{\partial N_i}{\partial x}\frac{\partial N_j}{\partial x} + \frac{\partial N_i}{\partial y}\frac{\partial N_j}{\partial y} \right) dxdy \tag{3}$$

where k^e is the element specific conductivity. Note that the shape function index in eqn. (3) corresponds to local numbering of element nodes and has to be converted to global node numbers before element matrix $\mathbf{K}^{\mathbf{e}}$ is assembled into the global matrix $\mathbf{K}$.

Mechanical problem

The strong form of the incompressible plane strain Stokes flow in Ω is

$$\begin{aligned}
\frac{\partial}{\partial x}\left(\mu\left(\frac{4}{3}\frac{\partial u_x}{\partial x} - \frac{2}{3}\frac{\partial u_y}{\partial y} \right) \right) + \frac{\partial}{\partial y}\left(\mu\left(\frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x} \right) \right) - \frac{\partial p}{\partial x} &= f_x \\
\frac{\partial}{\partial y}\left(\mu\left(\frac{4}{3}\frac{\partial u_y}{\partial y} - \frac{2}{3}\frac{\partial u_x}{\partial x} \right) \right) + \frac{\partial}{\partial x}\left(\mu\left(\frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x} \right) \right) - \frac{\partial p}{\partial y} &= f_y \qquad \text{in } \Omega \\
\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} + \frac{p}{\kappa} &= 0
\end{aligned} \tag{4}$$

where u_x and u_y are components of velocity, f_x and f_y are components of the body force vector field, p is pressure and μ denotes viscosity. In our numerical code the incompressibility constraint is achieved by penalizing the bulk deformation with a large bulk modulus κ. The boundary conditions are given as constraint velocity or vanishing traction components. In eqn. (4) we adhere to the divergence rather than Laplace form (in the latter different velocity components are only coupled through the incompressibility constraint) as we expect to deal with strongly varying viscosity. It is also worth noting that even for homogeneous models the computationally advantageous Laplace form may lead to serious defects if the boundary terms are not treated adequately [*Limache, et al.*, 2007]. Additionally our formulation, eqn. (4), and its

numerical implementation are also applicable to compressible and incompressible elastic problems due to the correspondence principle.

In analogy to the thermal problem we introduce the discrete spaces in order to approximate the velocity components and pressure

$$\begin{aligned}\tilde{u}_x(x,y) &= \sum_{i=1}^{nnod} N_i(x,y)u_x^i \\ \tilde{u}_y(x,y) &= \sum_{i=1}^{nnod} N_i(x,y)u_y^i \\ \tilde{p}(x,y) &= \sum_{i=1}^{np} \Pi_i(x,y)p^i\end{aligned} \tag{5}$$

where *np* denotes number of pressure degrees of freedom and Π_i are the pressure shape functions, which may not coincide with the velocity ones. In order to ensure the solvability of the resulting system of equations [inf-sup condition, see *Elman, et al.*, 2005], special care must be taken when constructing the approximation spaces. A wrong choice of the pressure and velocity discretization results in spurious pressure modes that may seriously pollute the numerical solution. Our particular element choice is the seven-node Crouzeix-Raviart triangle with quadratic velocity shape functions enhanced by a cubic bubble function and discontinuous linear interpolation for the pressure field [e.g., *Cuvelier, et al.*, 1986]. This element is stable and no additional stabilization techniques are required [*Elman, et al.*, 2005]. The fact that in our case the velocity and pressure approximations are autonomous leads to the so-called mixed formulation of the finite element method [*Brezzi and Fortin*, 1991].

Assuming that velocity degrees of freedom are followed by pressure ones in the local element numbering, the stiffness matrix for the Stokes problem is given by [e.g., *Bathe*, 1996]

$$\mathbf{K}^e = \begin{pmatrix} \mathbf{A} & \mathbf{Q}^T \\ \mathbf{Q} & -\kappa^{-1}\mathbf{M} \end{pmatrix} = \iint_{\Omega^e} \begin{pmatrix} \mu^e \mathbf{B}^T \mathbf{D}\mathbf{B} & -\mathrm{B}_{vol}^T \Pi^T \\ -\Pi \mathrm{B}_{vol} & -\kappa^{-1}\Pi\Pi^T \end{pmatrix} dxdy \tag{6}$$

where $\mathbf{B}$ is the so-called kinematic matrix transforming velocity into strain rate $\dot{\boldsymbol{\varepsilon}}$ (note the engineering convention for the shear strain rate)

$$\begin{pmatrix} \dot{\varepsilon}_{xx}(x,y) \\ \dot{\varepsilon}_{yy}(x,y) \\ \dot{\gamma}_{xy}(x,y) \end{pmatrix} = \mathbf{B}(x,y)u^e = \begin{pmatrix} \frac{\partial N_1}{\partial x}(x,y) & 0 & \dots \\ 0 & \frac{\partial N_1}{\partial y}(x,y) & \dots \\ \frac{\partial N_1}{\partial y}(x,y) & \frac{\partial N_1}{\partial x}(x,y) & \dots \end{pmatrix} \begin{pmatrix} u_x^1 \\ u_y^1 \\ \vdots \end{pmatrix} \tag{7}$$

The matrix $\mathbf{D}$ extracts the deviatoric part of the strain rate, converts from engineering convention to standard shear strain rate, and includes a conventional factor 2. The bulk strain rate is computed according to the equation $\dot{\varepsilon}_{vol} = \mathrm{B}_{vol} u^e$ and pressure is the projection of this field onto the pressure approximation space

$$p(x,y) = \kappa \Pi^T(x,y)\mathbf{M}^{-1}\mathbf{Q}u^e \tag{8}$$

With the chosen approximation spaces, the linear pressure shape functions Π are spanned by the corner nodal values that are defined independently for neighboring elements. Thus, it is possible to invert $\mathbf{M}$ on element level (the so-called static condensation) and consequently avoid the pressure unknowns in the global system. Since the pressure part of the right hand side vector is set to zero, this results in the following velocity Schur complement

$$(\mathbf{A} + \kappa \mathbf{Q}^{\mathbf{T}}\mathbf{M}^{-1}\mathbf{Q})u^e = f^e \tag{9}$$

Once the solution to the global counterpart of (9) is obtained, the pressure can be restored afterwards according to (8). The resulting global system of equations is not only symmetric, but also positive-definite as opposed to the original system (6). Unfortunately, the global matrix becomes ill-conditioned for penalty parameter values corresponding to a satisfactorily low level of the flow divergence. It is possible to circumvent this by introducing Powell and Hestenes iterations [*Cuvelier, et al.*, 1986] and keeping the penalty parameter κ moderate compared to the viscosity μ

$$\begin{aligned} &p^0 = 0 \\ &\textit{while } \max(\Delta p^i) > tol \\ &\quad u^i = (\mathbf{A} + \kappa \mathbf{Q}^{\mathbf{T}}\mathbf{M}^{-1}\mathbf{Q})^{-1}\left(f - \mathbf{Q}^{\mathbf{T}} p^i\right) \\ &\quad \Delta p^i = \kappa \mathbf{M}^{-1}\mathbf{Q}u^i \\ &\quad p^{i+1} = p^i + \Delta p^i \\ &\quad \textit{increment } i \\ &\textit{end} \end{aligned} \tag{10}$$

In the above iteration scheme the matrices **A**, **Q**, M represent global assembled versions rather than single element contributions.

Isoparametric elements

In order to exploit the full flexibility of FEM, we employ isoparametric elements. Each element in physical space is mapped onto the reference element with fixed shape, size, and orientation. This geometrical mapping between local (ξ,η) and global (x,y) coordinates of an element is realized using the same shape functions N_i that interpolate physical fields:

$$x(\xi,\eta) = \sum_{i=1}^{nnodel} N_i(\xi,\eta)x_i \tag{11}$$

$$\mathrm{y}(\xi,\eta) = \sum_{i=1}^{nnodel} N_i(\xi,\eta)y_i \tag{12}$$

where *nnodel* is the number of nodes in the element. The local linear approximation to this mapping is given by the Jacobian matrix J:

$$J = \begin{bmatrix} \dfrac{\partial x}{\partial \xi} & \dfrac{\partial x}{\partial \eta} \\ \dfrac{\partial y}{\partial \xi} & \dfrac{\partial y}{\partial \eta} \end{bmatrix} \tag{13}$$

The shape function derivatives with respect to global coordinates (x,y) are calculated using the inverse of the Jacobian and the shape function derivatives with respect to local coordinates (ξ,η):

$$\begin{Bmatrix} \dfrac{\partial N_i}{\partial x} & \dfrac{\partial N_i}{\partial y} \end{Bmatrix} = \begin{Bmatrix} \dfrac{\partial N_i}{\partial \xi} & \dfrac{\partial N_i}{\partial \eta} \end{Bmatrix} \begin{bmatrix} \dfrac{\partial x}{\partial \xi} & \dfrac{\partial x}{\partial \eta} \\ \dfrac{\partial y}{\partial \xi} & \dfrac{\partial y}{\partial \eta} \end{bmatrix}^{-1} \tag{14}$$

Thus, the element matrix from eqn. (3) is now given by:

$$K_{ij}^e = \iint_{\Omega_{ref}} k^e \left(\frac{\partial N_i}{\partial x}\frac{\partial N_j}{\partial x} + \frac{\partial N_i}{\partial y}\frac{\partial N_j}{\partial y} \right) |J| \, d\xi d\eta \tag{15}$$

where $|J|$ is the determinant of the Jacobian, taking care of the area change introduced by the mapping, and Ω_{ref} is the domain of the reference element. To avoid symbolic integration eqn. (15) can be integrated numerically:

$$K_{ij}^{e} = \sum_{k=1}^{nip} W_k k^e \left(\left(\frac{\partial N_i}{\partial x}\frac{\partial N_j}{\partial x} + \frac{\partial N_i}{\partial y}\frac{\partial N_j}{\partial y} \right) |J| \right) \Bigg|_{(\xi_k, \eta_k)} \tag{16}$$

Here the integral is transformed into a sum over nip integration points located at (ξ_k, η_k), where the individual summands are evaluated and weighted by point specific W_k. For numerical integration rules for triangular elements see e.g. [*Dunavant*, 1985]. The numerical integration of the element matrix arising in the mechanical case is analogous.

In the following we first show the straightforward implementation of the global matrix computation and investigate its efficiency. It proves to be unsuited for high performance computing in the MATLAB environment. We then introduce a different approach, which solves the identified problems. Finally, we discuss how to build sparse matrix data structures, apply boundary conditions, solve the system of linear equations and perform the Powell and Hestenes iterations.

Matrix Computation: Standard Algorithm

Code Fragment 1: Standard matrix computation.

```
% i) ELEMENT LOOP - MATRIX COMPUTATION
for iel = 1:nel

    % ii) FETCH DATA OF ELEMENT
    ECOORD_X = GCOORD(:,ELEM2NODE(:,iel));
    ED       =   D(Phases(iel));          % THERMAL
    EMu      =  Mu(Phases(iel));          % MECHANICAL
    ERho     = Rho(Phases(iel));

    % iii) INTEGRATION LOOP
    K_elem(:) = 0; % THERMAL
    A_elem(:) = 0; % MECHANICAL
    Q_elem(:) = 0;
    M_elem(:) = 0;
    Rhs_elem(:) = 0;

    P(2:3,:) = ECOORD_X(:,1:3);

    for ip=1:nip

        % iv) LOAD SHAPE FUNCTIONS DERIVATIVES FOR INTEGRATION POINT
        Ni       =    N
;
        dNdui       = dNdu{ip};

        % iv b) MECHANICAL: COMPUTE PRESSURE SHAPE FUNCTIONS FOR INTEGRATION POINT
        Pb(2:3)     = ECOORD_X*Ni;
        Pi          = P\Pb;

        % v) CALCULATE JACOBIAN, ITS DETERMINANT AND INVERSE
        J           = ECOORD_X*dNdui;
        detJ        = det(J);
        invJ        = inv(J);

        % vi) DERIVATIVES wrt GLOBAL COORDINATES
        dNdX        = dNdui*invJ;
```

```
        % vii   NUMERICAL INTEGRATION OF ELEMENT MATRICES
        weight = Ip_w(ip)*detJ;

        % vii a) THERMAL: STIFFNESS MATRIX
        K_elem       = K_elem + ED*(dNdX*dNdX');

        % vii b) MECHANICAL: STIFFNESS MATRIX AND RIGHT HAND SIDE
        B(1:2:end,1) = dNdX(:,1);
        B(2:2:end,2) = dNdX(:,2);
        B(1:2:end,3) = dNdX(:,2);
        B(2:2:end,3) = dNdX(:,1);
        Bvol         = dNdX';

        A_elem = A_elem + weight*EMu*(B'*D*B);
        Q_elem = Q_elem - weight*Bvol(:)*Pi';
        M_elem = M_elem + weight*Pi*Pi';

        Rhs_elem = Rhs_elem + weight*ERho*G*Ni';
    end

    % viii) STATIC CONDENSATION OF PRESSURE DOFS (MECHANICAL)
    invM_elem      = inv(M_elem);
    A_elem         = A_elem + PF*Q_elem*invM_elem*Q_elem';

    % ix) WRITE DATA INTO GLOBAL STORAGE
    K_all(:,iel)   = K_elem(indx_l); % THERMAL
    A_all(:,iel)   = A_elem(indx_l); % MECHANICAL
    Rhs_all(:,iel) = Rhs_elem(:);

    % ix b) MECHANICS: REUSE IN POWELL AND HESTENES ITERATIONS
    Q_all(:,iel)   = Q_elem(:);
    invM_all(:,iel)= invM_elem(:);
end
```

Algorithm description

The algorithm outlined in the above code fragment represents the straightforward implementation of the section "FEM Outline". We tried to use intuitive variable and index names; they are explained in "Appendix I: Variable Names". The details of the algorithm are described in the following (Roman numbers correspond to the comments in the code fragment).

i) The outermost loop of the standard algorithm is the element loop. Before the actual matrix computation, general element-type specific data such as integration points `IP_X` and weights `IP_w` are assigned. The derivatives of the shape functions `dNdu` with respect to the local (ξ,η) coordinates are evaluated in the integration points `IP_X`. All arrays used during the matrix computation procedure are allocated in advance, e.g. `K_all`.

ii) Inside the loop over all elements the code begins with reading element-specific information, such as indices of the nodes belonging to the current element, coordinates of the nodes, and element conductivity, viscosity and density.

iii) For each element the following loop over integration points performs numerical integration of the underlying equations, which results in the element stiffness matrix `K_elem[nnodel,nnodel]`. In the case of mechanical code additional matrices `A_elem[nedof, nedof]`, Q_`elem[nedof,np]` and `M_elem[np,np]` are required. All of the above arrays must be cleared before the integration point loop together with the right hand side vector `Rhs_elem`.

iv) Inside the integration point loop the pre-computed shape function derivatives `dNdui` are extracted for the current integration point. b) In the chosen element type the pressure is interpolated linearly in the global coordinates. Pressure shape functions `Pi` at an integration point are obtained as a solution of the system `P*Pi=Pb`, where the first equation enforces that the shape functions `Pi` sum to unity.

v) The Jacobian `J[ndim,ndim]` is calculated for each integration point by multiplying the element's nodal coordinates matrix `ECOORD_X[ndim,nnodel]` by `dNdui[nnodel,ndim]`. Furthermore its determinant, `detJ`, and inverse, `invJ[ndim,ndim]`, are obtained with the corresponding MATLAB functions.

vi) The derivatives versus global coordinates, `dNdx[nnodel, ndim]`, are obtained by `dNdx = dNdui*invJ` according to eqn. (14).

vii) The element thermal stiffness matrix contribution is obtained according to eqn. (16) and implemented as `K_elem = K_elem + weight*ED*(dNdX*dNdX')`. b) The kinematic matrix `B` needs to be formed, eqn. (7), and `A_elem`, `Q_elem` and `M_elem` are computed according to eqn. (6).

viii) The pressure degrees of freedom are eliminated at this stage. It is possible to invert `M_elem` locally because the pressure degrees of freedom are not coupled across elements, thus there is no need to assemble them into the global system of equations. It is worth noting that for large viscosity variations it may be beneficial to relate the penalty factor `PF` to element's viscosity to improve condition number of the global matrix.

ix) The lower (incl. diagonal) part of the element stiffness matrix is written into the global storage relying on the symmetry of the system. b) `Q_elem` and `invM_elem` matrices are stored for each element in order to avoid re-computing them during Powell and Hestenes iterations.

MATLAB provides a framework for scientific computing that is freed from the burden of conventional high-level programming languages, which require detailed variable declarations and do not provide native access to solvers, visualization, file I/O etc. However, the ease of code development in MATLAB comes with a loss of some performance, especially when certain recommended strategies are not followed: http://mathworks.com/support/solutions/data/1-15NM7.html. The more obvious performance considerations have already gone into the above standard implementation and we would like to point these out:

1) Memory Allocation / Variable Declaration: Although not formally required, it is advisable to explicitly declare variables including their size and type. If variables are not declared with their final size, but are instead successively extended (filled in) during loop evaluation, a large penalty has to be paid for the continuous and unnecessary memory management. Hence, all variables that could potentially grow in size during loop execution are pre-allocated; e.g., `K_all`. Variables such as `ELEM2NODE` that only have to store integer numbers should be declared accordingly, `int32` in the case of `ELEM2NODE` instead of MATLABs default variable type `double`. This reduces both the amount of memory required to store this large array and the time required to access it since less data must be transferred.

2) Data Layout: In order to facilitate memory access by the CPU it is important to have a proper data layout. For example, the indices of the nodes of each element must be stored in neighboring memory locations, and similarly the x-y-z coordinates of every node. The actual numbering of nodes and elements also has a visible effect on cache reuse inside the element loop, similarly to sparse matrix-vector multiplication problem [*Toledo*, 1997].

3) Avoid Multiple Data Transfer & Computation: Generally, statements should appear in the outermost possible loop in order to avoid multiple transfer and computation of identical data. This is why the integration point evaluated shape function derivatives with respect to local coordinates are pre-computed outside the element loop (as op-

posed to inside the integration loop) and the nodal coordinates are extracted before the integration loop.

Performance analysis

In order to analyze the performance of the standard matrix computation algorithm we run corresponding tests on an AMD Opteron system with 64bit Red Hat Enterprise Linux 4 and MATLAB 2007a using GoTo BLAS (http://www.tacc.utexas.edu/resources/software). This system has a peak performance of 4.4Gigaflops per core, i.e. it is theoretically capable of performing 4.4 billion double precision floating point operations per second (flops). The specific element types used are 6-node triangles (quadratic shape functions) with 6 integration points for the thermal problem and 7-node triangles with 12 integration points for the mechanical problem

In the case of the thermal problem test results are obtained for an unstructured mesh consisting of approximately 1 million nodes and 0.5 million elements. For this model the previously described matrix computation took 65 seconds, during which 324 floating point operations per integration point per element were calculated. This corresponds to 15 Megaflops (Mflops) or approximately 0.4% of the peak performance. Analysis of the code with MATLABs built-in profiler revealed that a significant amount of time was spent on the calculation of the determinant and inverse of the Jacobian. Therefore, in further tests these calls were replaced by explicit calculations of `detJ` and `invJ`. The final performance achieved by this algorithm was 30 Mflops, which is still less than one percent of the peak performance and equivalent to a peak CPU performance that was reached by commodity computers more than a decade ago.

Profiling the improved standard algorithm revealed that most of the computational time was spent on matrix multiplications. This means that the efficiency of the analyzed implementation depends mainly on the efficiency of dense matrix by matrix multiplications inside the integration point loop. In order to perform these calculations MATLAB uses hardware-tuned, high performance BLAS libraries (Basic Linear Algebra Subprograms, see http://www.netlib.org/blas/faq.html and Dongarra et al. [1990]), which reach up to 90% of the CPU peak performance; a value from which the analyzed code is far away.

The cause for this bad performance is that the matrix by matrix multiplications inside the integration point loop operate on very small matrices, for which BLAS libraries are known not to work well due to the introduced overhead (e.g. http://math-atlas.sourceforge.net/timing/36v34/OptPerf.html). Therefore the same observation can be

made when writing the standard algorithm in a compiler language such as C and relying on BLAS for the matrix multiplications, albeit the actual performance in this case is higher than in MATLAB. In C a possible solution is to explicitly write out the small matrix by matrix multiplications, which results in a more efficient code. In MATLAB, however, this is not a practical alternative as explicitly writing out matrix multiplications leads to unreadable code without substantial performance gains.

The above performance considerations apply equally to the mechanical code.

In conclusion, the standard algorithm is a viable option when writing compiler code. However, the achievable performance in MATLAB is unsatisfactory and we therefore developed a more efficient approach, which is presented in the following section.

Remark 1: Measuring code performance

Since no flops measure exists in MATLAB, the number of operations must be manually calculated based on code inspection and divided by the computational time. In order to provide more meaningful performance measures only the number of necessary floating point operations may be considered, e.g. the redundant computations of the upper triangular entries in the standard matrix contribute to the flop count, which artificially increases the measured performance. However, it is not necessarily the case that the algorithm with the lowest operation count is the fastest in terms of execution time. We therefore restrain from adjusting the actual flop counts in this paper.

Matrix Computation: Optimized Algorithm

In this section we explain how to efficiently compute the local stiffness matrices. This optimization strategy is common to both (thermal and mechanical) problems considered. For simplicity we present it on the example of the thermal problem. Overall performance benchmarks and application examples are provided for both types of problems in subsequent sections.

The small matrix by small matrix multiplications in the integration loop nested inside the loop over elements are the bottleneck of the standard algorithm. Written out in terms of loops, these matrix multiplications represent another three loops, totaling to five. Since the element loop exhibits no data dependency, it can be moved into the innermost three (out of five), effectively becoming part of small matrix by large matrix multiplication.

This loop reordering does not change the total amount of operations. However, the number of BLAS calls is greatly reduced (`ndim*nip` versus `nel*nip` in the standard approach), and the amount of computation done per function call is drastically increased. Consequently, the overhead problem vanishes leading to a substantial performance improvement. Unfortunately, the performance decreases once a certain number of elements is exceeded. The reason for this is

that the data required for the actual operation does not fit any longer into the CPUs cache, which inhibits cache reuse within the integration point loop. The remedy is to operate on blocks of elements of the size for which the observed performance is best. Once a block is processed, the results are written to the main memory and the data required by the next block is copied into the cache. Data required for every block should fit (reside) in the cache at all times. The ideal block size depends on the cache structure of a CPU and must be determined system and problem specifically. This computing strategy is called "blocking" and is implemented as a part of the optimized algorithm. Coincidentally, this entire approach to optimize the FEM matrix computation is similar to vector computer implementations [e.g., *Ferencz and Hughes*, 1998; *Hughes, et al.*, 1987; *Silvester*, 1988].

Code Fragment 2: Optimized Finite Element Global Matrix Computation

```
% i) BLOCK LOOP - MATRIX COMPUTATION
il = 1;
iu = nelblo;
for ib = 1:nblo

    % ii) FETCH DATA OF ELEMENTS IN BLOCK
    ECOORD_x = reshape( GCOORD(1,ELEM2NODE(:,il:iu)), nnodel, nelblo);
    ECOORD_y = reshape( GCOORD(2,ELEM2NODE(:,il:iu)), nnodel, nelblo);
    ED       = D(Phases(il:iu));

    % iii) INTEGRATION LOOP
    K_block(:)  = 0;
    for ip=1:nip

        % iv) LOAD SHAPE FUNCTIONS DERIVATIVES FOR INTEGRATION POINT
        dNdui       = dNdu{ip};

        % v) CALCULATE JACOBIAN, ITS DETERMINANT AND INVERSE
        Jx          = ECOORD_x'*dNdui;
        Jy          = ECOORD_y'*dNdui;
        detJ        = Jx(:,1).*Jy(:,2) - Jx(:,2).*Jy(:,1);

        invdetJ     = 1.0./ detJ;
        invJx(:,1)  = +Jy(:,2).*invdetJ;
        invJx(:,2)  = -Jy(:,1).*invdetJ;
        invJy(:,1)  = -Jx(:,2).*invdetJ;
        invJy(:,2)  = +Jx(:,1).*invdetJ;

        % vi) DERIVATIVES wrt GLOBAL COORDINATES
        dNdx        = invJx*dNdui';
        dNdy        = invJy*dNdui';

        % vii) NUMERICAL INTEGRATION OF ELEMENT MATRICES - ONLY LOWER TRIANGLE
        weight      = IP_w(ip)*detJ.*ED;
        indx = 0;
        for i = 1:nnodel
            for j = i:nnodel
                indx = indx  + 1;
                K_block(:,indx)  =   K_block(:,indx) + ...
                    (dNdx(:,i).*dNdx(:,j)+ dNdy(:,i).*dNdy(:,j)).*weight;
            end
        end
    end
```

```
    % ix) WRITE DATA INTO GLOBAL STORAGE
    K_all(:,il:iu) = K_block';

    % READJUST START, END AND SIZE OF BLOCK. REALLOCATE MEMORY
    il  = il + nelblo;
    if(ib==nblo-1)
        nelblo  = nel-iu;
        K_block = zeros(nelblo, nnodel*(nnodel+1)/2);
        invJx   = zeros(nelblo, ndim);
        invJy   = zeros(nelblo, ndim);
    end
    iu  = iu + nelblo;
end
```

Algorithm description

The above code fragment shows the actual implementation of the optimized matrix computation algorithm. The key operations are explained and compared to the standard algorithm in the following.

i) The outermost loop of the optimized matrix computation is the block loop. Before this loop is entered, required arrays (`IP_X, IP_w, dNdu`) are assigned and necessary variables are allocated.

ii) Inside the block loop the code begins with reading element specific information. Since we simultaneously operate on `nelblo` elements, all the corresponding global data blocks are copied into local arrays `ECOORD_x`, `ECOORD_y`, and `ED`, and are used repeatedly inside the integration loop.

iii) For the entire block of elements, the loop over integration points performs numerical integration of the element matrices `K_block[nelblo, nnodel*(nnodel+1)/2]`.

iv) As in the standard algorithm, every iteration of the integration point loop begins by reading pre-computed `dNdu` arrays.

v) The Jacobian of the standard algorithm, `J[ndim,ndim]`, is replaced by `ndim` matrices; `Jx[nelblo,ndim]` and `Jy[nelblo,ndim]`, containing the individual rows of the Jacobian evaluated at the actual integration point for all elements of the current block. `Jx` and `Jy` are calculated by multiplying the nodal coordinates by the shape function derivatives, e.g. `Jx[nelblo,ndim] = ECOORD_x[nnodel, nelblo]' *dNdui[nnodel, ndim]`. Thus, instead of `nelblo*nip` matrix multiplications of `dNdu[ndim,nnodel]` and `ECOORD_X[ndim,nnodel]`, `ndim*nip` multiplications

involving the larger matrices *ECOORD_x*, *ECOORD_y* are performed, i.e. the same work is done with less multiplications of larger matrices. Once the Jacobian is obtained, its determinant, *detJ*, and inverse, split into *invJx* and *invJy* , are explicitly computed using simple operations on vectors.

vi) The derivatives with respect to the global coordinates (x, y), *dNdx[nelblo,nnodel]* and *dNdy[nelblo,nnodel]*, are obtained by multiplying the *invJx* and *invJy* by the transpose of *dNdui*. Again, less multiplication calls involving larger matrices are performed.

vii) The local stiffness matrix contribution for all the elements in the block, *K_block[nelblo,nnodel*(nnodel+1)/2]*, is computed according to eqn. (16). Note that exploiting symmetry allows for calculation of only the lower triangle of stiffness matrices, which substantially reduces the operation count.

viii) After the numerical integration of *K_block* is completed, the results are written into the global storage *K_all,* again exploiting symmetry by storing only the lower triangular part.

ix) The number of elements remaining in the final block might be smaller than the *nelblo*. Consequently, *nelblo* and several arrays must be adjusted.

Performance analysis

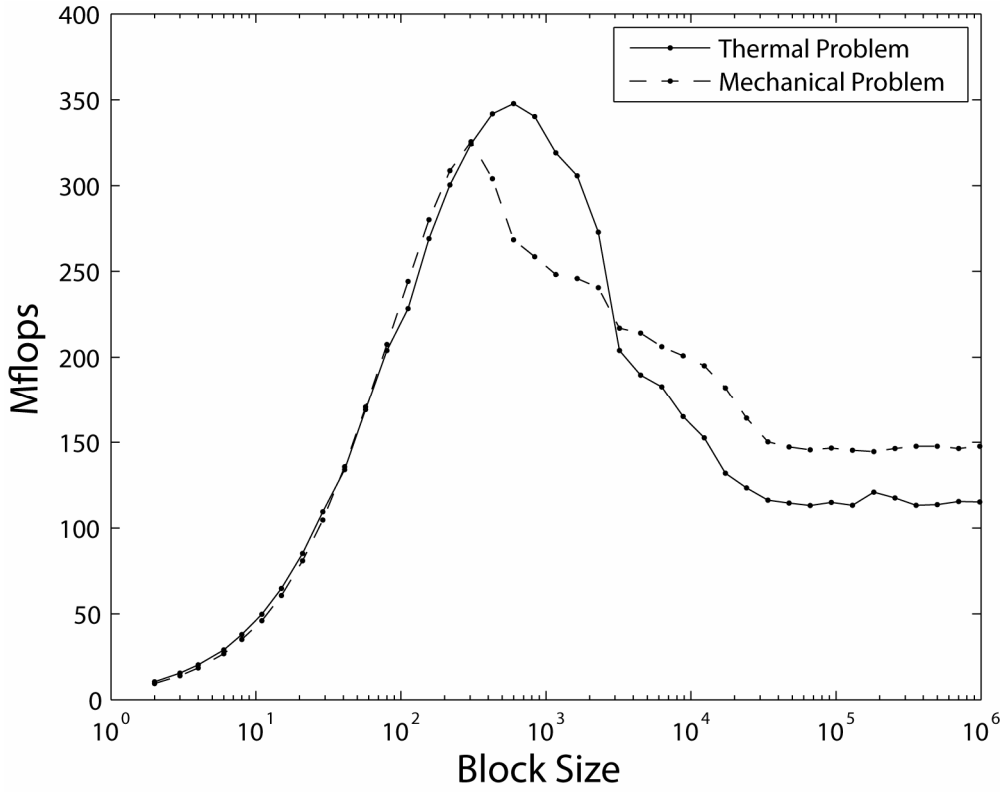


Figure 1 Performance of optimized matrix computation versus block size.

In order to illustrate the performance of the optimized matrix computation systematic tests were run with the same 1 million node problem that was used for the performance analysis of the standard algorithm. Since larger matrices resulting from larger block sizes should yield better BLAS efficiency, the performance in Mflops is plotted versus the number of elements in a block, see Figure 1. This plot confirms the arguments for the introduction of the blocking algorithm. Starting from approximately the performance of the standard algorithm, a steady increase can be observed up to ~350 Mflops, which on the test system is reached for a block with ~1000 elements for thermal problem. Further increase of the block size leads to a performance decrease towards a stable level of ~120 Mflops due to lack of cache reuse in the integration point loop. Compared to the standard version, the optimized matrix computation achieves a 20-fold speedup in terms of flops performance. Since the optimized algorithm performs less operations (computation of only lower triangular part of symmetric element matrix), its execution time is actually more than 30 times faster.

The achieved 350Mflops efficiency corresponds to only ~8 % of the peak CPU performance. Profiling the code revealed that for the test problem approximately half of the time was spent on reading and writing of variables from and to RAM (e.g. nodal coordinates and element ma-

trices). This value is constrained by the memory bandwidth of the hardware, which on current computer architectures is often a bigger bottleneck than the CPU performance. Compared to C implementations, the optimized matrix computation performance is better than the straightforward standard algorithm using BLAS, but more than a factor 3 slower than what can be achieved by explicitly writing out the matrix multiplications.

In the case of the mechanical code, the peak flops performance is similar. Note that in this case the optimal blocksize is smaller due to the larger workspace of the method, see Figure 1.

Matrix Assembly: Triplet to Sparse Format Conversion

The element stiffness matrices stored in `K_all` must now be assembled into the global stiffness matrix `K`. The row and column indices (`K_i` and `K_j`) that specify where the individual entries of `K_all` have to be stored in the global system are commonly known as the triplet sparse matrix format [e.g., *Davis*, 2006]. Since we only use lower triangular entries, special care must be taken so that the indices referring to the upper triangle are not created, see Code Fragment 3. Note that `K_i` and `K_j` hold duplicate entries, and the purpose of the MATLAB `sparse` function is to sum and eliminate them.

While creation of the triplet format is fast, the call to `sparse` gives some concerns. MATLABs `sparse` implementation requires that `K_i` and `K_j` are of type double, which is memory- and performance-wise inefficient. In addition, `sparse` itself is rather slow, especially if compared to the time spent on the entire matrix computation. The equivalent function sparse2, provided by T. Davis within the CHOLMOD package (http://www.cise.ufl.edu/research/sparse/SuiteSparse), is substantially faster and does not require a conversion of the coefficients to double precision. The code fragment below presents in detail how to create a global system matrix.

Code Fragment 3: Global sparse matrix assembly

```
% CREATE TRIPLET FORMAT INDICES
indx_j = repmat(1:nnodel,nnodel,1); indx_i = indx_j';
indx_i = tril(indx_i); indx_i = indx_i(:);indx_i = indx_i(indx_i>0);
indx_j = tril(indx_j); indx_j = indx_j(:);indx_j = indx_j(indx_j>0);

K_i = ELEM2NODE(indx_i,:);
K_j = ELEM2NODE(indx_j,:);

K_i      = K_i(:);
K_j      = K_j(:);

% SWAP INDICES REFERRING TO UPPER TRIANGLE
indx     = K_i < K_j;
tmp      = K_j(indx);
K_j(indx) = K_i(indx);
K_i(indx) = tmp;
```

```
K_all  = K_all(:);

% CONVERT TRIPLET DATA TO SPARSE MATRIX
K      = sparse2(K_i, K_j, K_all);
clear K_i K_j K_all;
```

The triplet format is converted into the sparse matrix *K* with one single call to *sparse2*. Assembling smaller sparse matrices for blocks of elements and calling sparse consecutively would reduce the workspace for the auxiliary arrays; however, it would also slow down the code. Therefore, as long as the K_i, K_j and K_all arrays are not the memory bottleneck, it is beneficial to perform the global conversion. Once *K* is created, the triplet data is cleared in order to free as much memory as possible for the solution stage. In case of the mechanical code $\mathbf{Q}$ and $\mathbf{M}^{-1}$ matrices are stored in sparse format for later reuse in the Powell and Hestenes iterations.

Remark 2: Symbolic approach to sparse matrix assembly

It is worth mentioning that in general the auxiliary arrays can be altogether avoided with a symbolic approach to sparse matrices. While the idea of sparse storage is the elimination of zero entries, in a symbolic approach all possible non-zero entries are stored and initialized to zero. During the computation of element stiffness matrices, global locations of their entries can be found at a small computational cost, and corresponding values are incrementally updated. Also, this symbolic storage pattern can be reused between subsequent time steps, as long as the mesh topology is not changed. Unfortunately, this improvement can not be implemented in MATLAB as zero entries are automatically deleted.

Boundary conditions

The implemented models have two types of boundary conditions: natural and Dirichlet. While the former automatically results from the FEM discretization, the latter must be specified separately, which usually leads to a modification of the global stiffness matrix. These modifications may, depending on the implementation, cause loss of symmetry, changes in the sparsity pattern and row addressing of *K*, all of which can lead to a badly performing code.

An elegant and sufficiently fast approach is to separate the degrees of freedom of the model into *Free* (indices of unconstraint degrees of freedom) *Bc_ind*, where Dirichlet boundary conditions with corresponding values *Bc_val* are applied. Since the actual solution values in the *Bc_ind* are known, the corresponding equations can be eliminated from the system of equations by modifying the right hand side of the remaining degrees of freedom accordingly. This is implemented as:

Code Fragment 4: Boundary conditions for the thermal problem

```
Free         = 1:nnod;
Free(Bc_ind)= [];
TMP          = K(:,Bc_ind) + cs_transpose(K(Bc_ind,:));
Rhs          = Rhs -  TMP*Bc_val';
K            = K(Free,Free);
T            = zeros(nnod,1);
T(Bc_ind)    = Bc_val;
```

Since only the lower part of the global matrix is stored, we need to restore the remaining parts of the columns by transposing the adequate rows.

System Solution

We have ensured that the global system of linear equations under consideration is symmetric, positive-definite, and sparse. It has the form:

$$\mathbf{K}T = \mathrm{Rhs} \tag{17}$$

where **K** is the stiffness matrix, T the unknown temperature vector, and Rhs is the right hand side. One of the fastest and memory efficient direct solvers for this type of systems is CHOLMOD, a sparse supernodal Cholesky factorization package developed by T. Davis [*Chen, et al.*, submitted; *Davis and Hager*, 2005; submitted]; see the report by Gould et al. [2007]. Newer versions of MATLAB (2006a and later) use this solver, which is substantially faster than the previous implementation. When symmetric storage is not exploited, CHOLMOD can be invoked through the backslash operator: `T=K\Rhs` (make sure that the matrix `K` is numerically symmetric, otherwise MATLABs will invoke a different, slower solver).

However, it is best to use CHOLMOD and the related parts by installing the entire package from the developers SuiteSparse website (http://www.cise.ufl.edu/research/sparse/SuiteSparse). This provides access to `cholmod2`, which is capable of dealing with only upper triangular input data and pre-computed permutation (reordering) vectors. SuiteSparse also contains `lchol`, a Cholesky factorization operating only on lower triangular matrices, which is faster and more memory efficient than MATLABs `chol` equivalent. Reusing the Cholesky factor `L` during the Powell and Hestenes iterations in the mechanical problem greatly reduces the computational cost of achieving a divergence free flow solution.

The mentioned re-use of reordering data is possible as long as the mesh topology remains identical, which even in our large strain flow calculations is the case for many time steps. The reordering step decreases factorization fill-in and consequently improves memory and CPU effi-

ciency [*Davis*, 2006], but is a rather costly operation compared to the rest of the Cholesky algorithm. Different re-ordering schemes can be used, and we compare two of them in Figure 2: AMD (Approximate Minimum Degree) and METIS (http://glaros.dtc.umn.edu/gkhome/views/metis). While AMD is faster during the reordering steps, it results in slower Cholesky factorization and forward and back substitution. If the reordering can be reused for a large number of steps, it is recommended to rely on METIS, which is accessible in MATLAB through the SuiteSparse package.

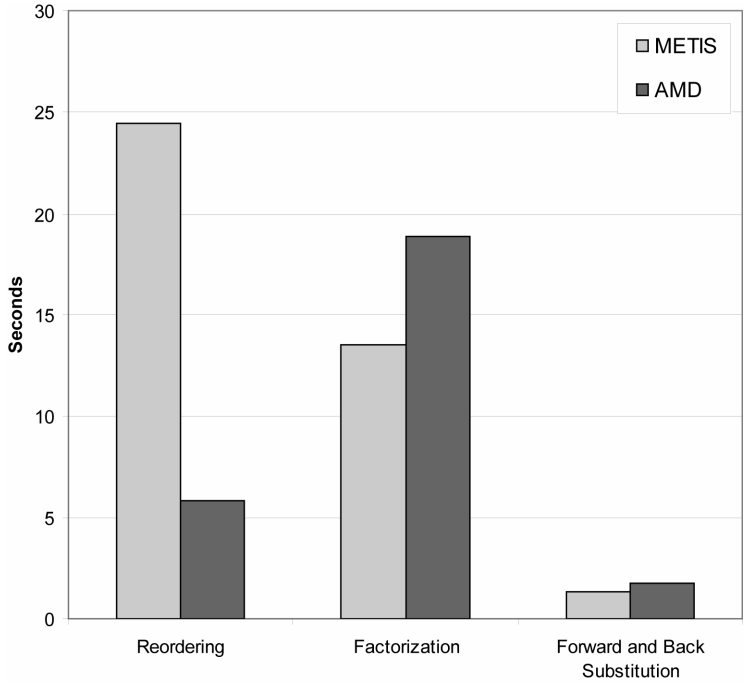


Figure 2

Performance analysis of the different steps of the Cholesky algorithm with different reorderings for our one million degrees of freedom thermal test problem.

Powell and Hestenes Iterations

In the case of the thermal code, the solution vector is obtained by calling forward and back substitution routines with the Cholesky factor and the adequately permuted right hand side vector. During the second substitution phase the upper Cholesky factor is required. However, instead of explicitly forming it through the transposition of the stored lower factor, it is advantageous to call the `cs_ltsolve` that can operate on the lower factor and performs the needed task of the back substitution.

In the MILAMIN flow solver the incompressibility constraint is achieved through an iterative penalty method, i.e. the bulk part of the deformation is suppressed with a large bulk modulus (penalty parameter) κ. In a single step penalty method there is a trade off between the incom-

pressibility of the flow solution and the condition number of the global equation system. This can be avoided by using a relatively small κ, which ensures a good condition number and then iteratively improving incompressibility of the flow. Note that for the chosen Crouzeix-Raviart element, pressure is discontinuous between elements and the corresponding degrees of freedom can be eliminated element-wise (no global system solution required). Pressure increments can be computed with the velocity solution vector and stored $\mathbf{Q}$ and $\mathbf{M}^{-1}$ matrices. These pressure increments are sent to the right hand side of the system and accumulated in the total pressure. The code fragment for these so-called Powell and Hestenes iterations is given below.

Code Fragment 5: Powell and Hestenes iterations

```
while (div_max>div_max_uz  && uz_iter<uz_iter_max)
    uz_iter = uz_iter + 1;
    %FORWARD AND BACK SUBSTITUTION
    Vel(Free(perm)) = cs_ltsolve(L,cs_lsolve(L,Rhs(Free(perm))));

    %COMPUTE QUASI-DIVERGENCE
    Div = invM*(Q*Vel);

    %UPDATE RHS
    Rhs = Rhs - PF*(Q'*Div);

    %UPDATE TOTAL PRESSURE
    Pressure    = Pressure + PF*Div;

    %CHECK INCOMPRESSIBILITY
    div_max     = max(abs(Div(:)));
end
```

Post-Processor

The results of a numerical model are only useful if fast and precise analysis and visualization is possible. One of the main aspects to achieve this is to avoid loops. For triangular meshes *trisurf* is the natural choice for two and three dimensional data visualization as it employs the usual FEM structures: connectivity (*ELEM2NODE*), coordinates (*GCOORD*), and data (*T*). This allows for visualization of FEM models with more than one million elements in less than one second.

A problem that often arises is the visualization of discontinuous data, such as pressure in mixed formulations of deformation problems. The remedy is to simply abandon the nodal connectivity and to create a new one, where physical nodes are listed separately for every element that accesses them. The same can also be done for other meshes than triangular ones by creating the corresponding connectivity (*ELEM2NODE*) and calling:

Code Fragment 6: Post-processor

```
patch('faces', ELEM2NODE,'vertices',GCOORD','facevertexcdata',T);
shading interp;
```

MILAMIN Performance Analysis

Overall Performance

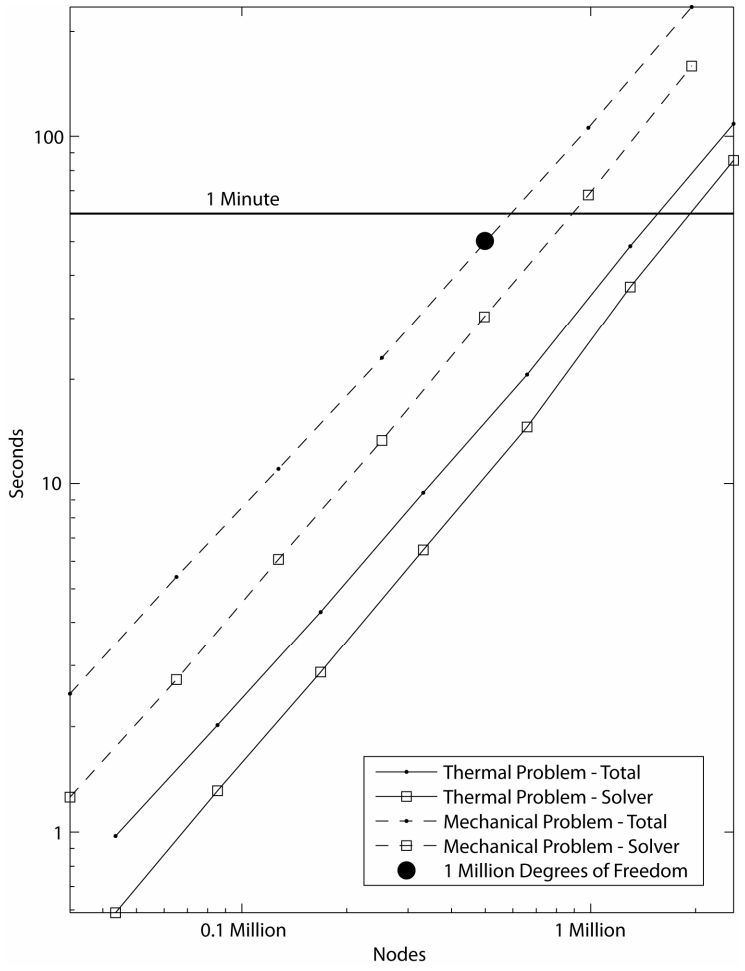


Figure 3

Overall performance results for MILAMIN given for total time spent on problem, and the solver contribution.

The overall performance of MILAMIN versus the number of nodes is analyzed in Figure 3. The goal of MILAMIN to perform a complete FEM analysis for one million unknowns in one minute is reached for the thermal as well as the mechanical problem. All components of MILAMIN scale linearly with the number of nodes; the only exception is the direct solver, which shows super-linear scaling. The performance details are discussed in the following sections.

Component Performance

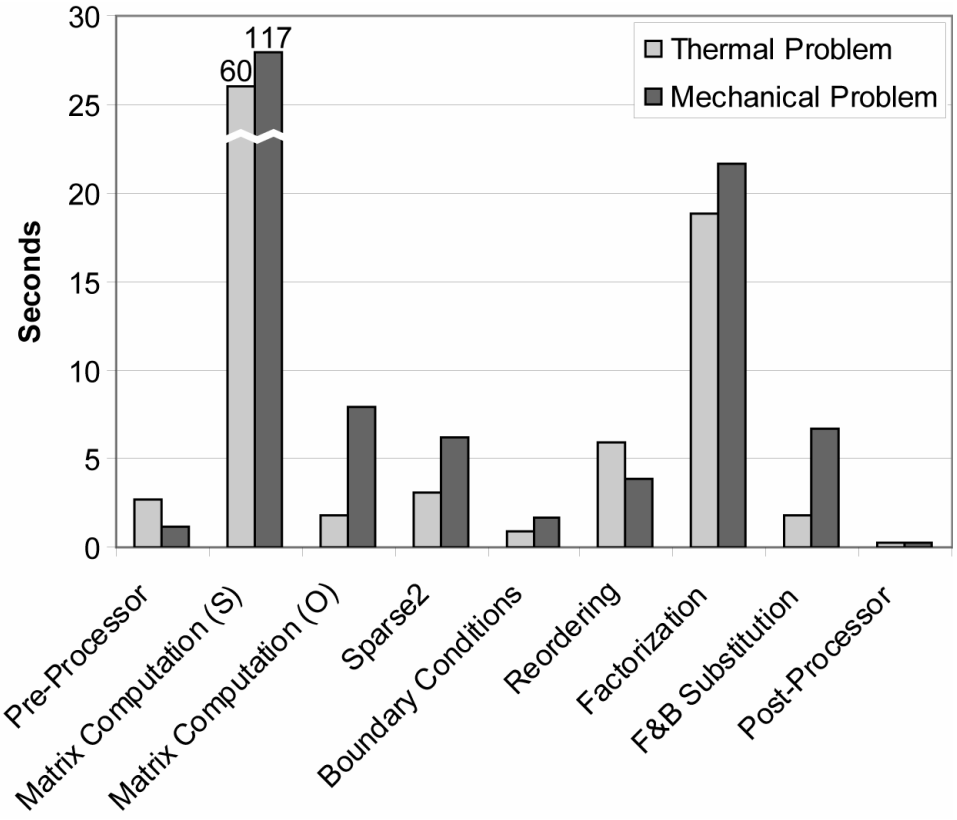


Figure 4

Overall performance of MILAMIN split up into the individual components for thermal and mechanical test problems with one million degrees of freedom. The timing for the matrix computation is given for the standard (S) and the optimized (O) algorithm. Note that the forward and backward (F&B) substitution timing also contains three Powell and Hestenes iterations in the case of the mechanical problem.

Figure 4 shows the total amount of time for the one million degrees of freedom (DOFs) test problems split into the individual components of MILAMIN. The contributions of the boundary conditions and post-processor are minor. The time taken by the pre-processor is also not relevant, especially if the same (Langrangian) mesh is used for many time steps. A major achievement of MILAMIN is the performance of the optimized matrix computation that is more than 15-30fold better than the standard algorithm. The matrix assembly done by `sparse2` is one of the major contributors to the total time, but cannot be optimized without a major change in the way MATLAB operates on sparse matrices, see Remark 2. Finally the three components of the Cholesky solver take substantial time.

The time taken by the first part of the Cholesky solver, the reordering, can often be neglected for practical applications. During non-linear material and time step iterations the mesh topol-

ogy remains the same as long as no remeshing is performed, and the permutation vector can be reused if the SuiteSparse package is employed.

The second component of the Cholesky solver is the factorization. This step takes most of the total MILAMIN execution time. However, the efficiency achieved by CHOLMOD is close to the optimal CPU performance. For further optimization one could consider other types of solvers such as iterative ones. Yet, preconditioned iterative methods or algebraic multigrid are less robust (especially for large material contrasts as targeted here) and perform better only for large systems, see section "Comparison to Other Software" below. These methods are the only option in the case of most three-dimensional problems, because the scaling of factorization time and memory requirements for direct solvers is much worse than in 2D. However, for two dimensional problems direct solvers are the best choice for resolutions on the order of one million degrees of freedom, especially for positive definite systems that can be solved with Cholesky factorizations. Moreover, it is problems of this size where our optimizations greatly reduce the total solution time. Such numerical resolutions are often sufficient in 2D to solve challenging problems and the achieved performance allows for studies with large number of time steps.

The third part of the Cholesky solver is the forward and backward substitution and does not contribute substantially in the case of thermal problems. For mechanical problems several Powell and Hestenes iterations are required to enforce incompressibility, each issuing a forward and back substitution call plus other computations. The time spent on the Powell and Hestenes iteration is not negligible, but the strategy chosen to deal with incompressibility is clearly advantageous to other strategies that would not allow the use of Cholesky solvers, see for example the results for FEMLAB using UMFPACK in section "Comparison to Other Software" below.

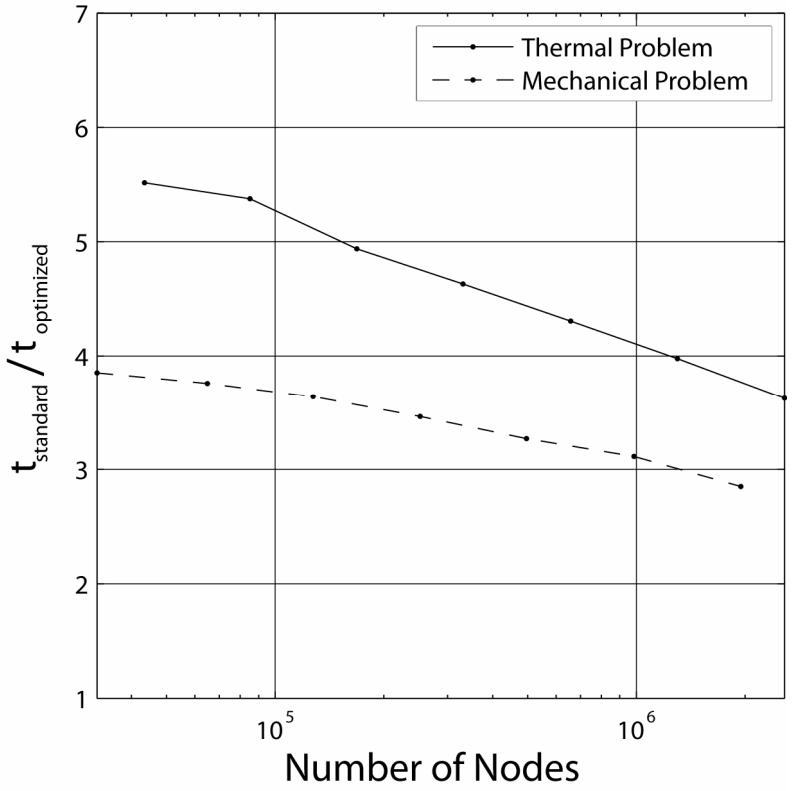


Figure 5

Achieved MILAMIN speedup for all operations that need to be performed for every time step; see text for details.

A final analysis of the overall speedup achieved by MILAMIN is shown in Figure 5where we depict the ratio of the total time $t_{standard}/t_{optimized}$ for the thermal and mechanical code. In this speedup analysis we define the total time as the sum of the time needed to compute and assemble the global matrix, apply boundary conditions, factorize and solve the system of equations, and perform the Powell and Hestenes iterations (incompressible Stokes flow). Thus mesh generation, post-processing, and reordering, which do not need to be performed for every time step, do not enter this analysis. For our target system sizes the achieved speedups reach approximately 3 and 4 for the mechanical and thermal codes, respectively. Hence, the performance gains due to the developed MILAMIN package are substantial. The scaling with respect to system size shows that the speedup decreases with increasing number of nodes. This is due to the super-liner scaling of the direct solver, which starts to dominate the total execution time for very large systems.

Memory Requirements

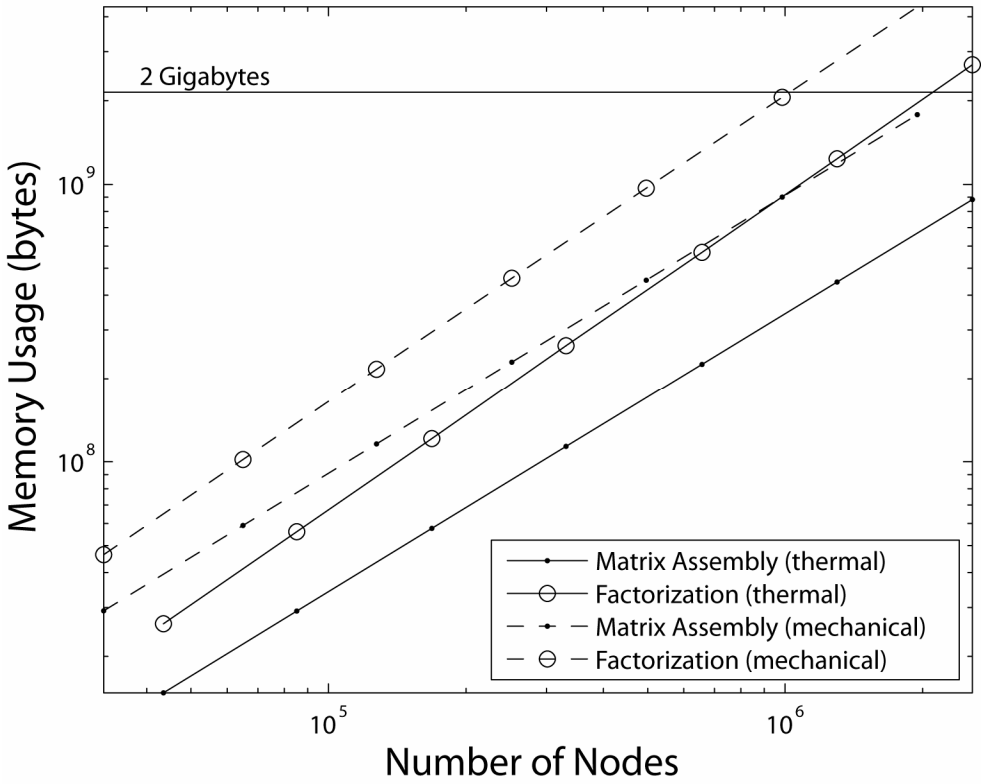


Figure 6

Memory requirements of the thermal and mechanical versions of MILAMIN.

Besides CPU performance the available memory (RAM) is the other parameter that determines the problem size that can be solved on a specific machine. The memory requirements of MILAMIN are presented in Figure 6. Within the studied range of systems sizes, all data allocated during the matrix computation and assembly requires substantially less memory than the solution stage. Thus, the auxiliary arrays such as `K_i`, `K_j` and `K_val` are not a memory bottleneck and it is indeed beneficial to perform conversion to sparse format globally. Note that the amount of memory required during the factorization stage depends strongly on reordering used. This analysis is only approximate as the workspace of the external routines (`lchol`, `sparse2`, etc) is not taken into account. On 2Gb RAM computers we are able to solve systems consisting of 1.65 and 0.65 million nodes for the thermal and mechanical problems, respectively.

Comparison to Other Software

In this section we compare MILAMIN to different available commercial and free software solving similar test problems. Table 1 presents run times for a thermal problem with ca. 1 million degrees of freedom. The model setup consists of a box with a circular hole (zero flux) and

a circular inclusion of ten times higher conductivity than the matrix. The outer boundaries are set to Dirichlet conditions representing a linearly varying temperature field.

Table 1

Performance results for different software packages for the thermal problem. T1 and T2 stand for linear and quadratic triangles, Q1 and Q2 for linear and quadratic quadrilateral elements, respectively.

Software	Matrix Computation and Assembly	Solve	Solver Type
ABAQUS, T2	80	260	Proprietary
FEMLAB, T2	18	40	UMFPACK
		45	TAUCS
		52	PARDISO
		58	SPOOLES
		240	ICCG
		500	AMG-CG
		1000	SSOR-CG
		2500	PCG
FEAPpv, fortran, T2	7	712	PCG
OOFEM, C++, T1	36	400	ICCG
TOCHNOG, C\C++, T2	15	1711	BiCG
AFEM@matlab, T1	25	19	MATLAB \
IFISS, Q2	999	57	MATLAB \
IFISS, Q1	464	30	MATLAB \
MILAMIN std, T2	65	24	CHOLMOD2 (AMD)
MILAMIN opt, T2	5	24	CHOLMOD2 (AMD)

The software that entered the test are commercial finite element packages, ABAQUS [*SIMULIA*] and FEMLAB [*COMSOL*], and open source packages FEAPpv [*Zienkiewicz and Taylor*], OOFEM [*Patzak*], and TOCHNOG [*Roddeman*] for compiler languages, and AFEM@matlab [*Chen and Zhang*], and IFISS [*Silvester, et al.*] for MATLAB. For the solution stage we used a wide range of direct solvers including UMFPACK [*Davis*], TAUCS [*Toledo, et al.*], PARDISO [*Schenk and Gärtner*], SPOOLES [*Ashcraft, et al.*], CHOLMOD [*Davis*], and the MATLAB backslash operator (\). We also compared different implementations of iterative solvers such as Conjugate Gradients preconditioned with Jacobi (PCG), Symmetric Successive Over-Relaxation (SSOR-CG), Incomplete Cholesky (ICCG), and Algebraic Multigrid (AMG-CG), and a Biconjugate Gradients solver preconditioned with Jacobi (BiCG).

A number of other MATLAB based packages are available, which, however, could not enter our table because they are simply incapable of solving the test problem in a reasonable amount of time and the amount of RAM available. From the MATLAB packages that entered the performance comparison AFEM seemingly excels with high performance. However, AFEM is

specifically developed to operate with linear triangles solving the Poisson problem. This allows AFEM to employ only one integration point and the amount of work performed is substantially less than for isoparamteric quadratic elements, although the actual number of elements is higher for the test problem with a fixed number of nodes. IFISS is another MATLAB based package capable of solving Poisson and incompressible Navier-Stokes problems based on linear and quadratic quadrilateral meshes. Despite its aim of being a vectorized code, the performance IFISS is not optimal. This is partly due to a badly performing boundary condition implementation. The matrix computation and assembly performance of the compile language and commercial codes is quite reasonable, with FEAP being the clear leader. However, none of the tested packages is as fast for the matrix computation and assembly as the optimized version of MILAMIN and even the standard version of MILAMIN is performing quite reasonably in comparison.

The analysis of the solver times confirms our previous statement that for the studied 2D problems direct solvers (CHOLMOD, UMFPACK, TAUCS, PARDISO, SPOOLES) are the best choice with CHOLMOD being the best in the group. Iterative solvers, even if equipped with good preconditioners, like incomplete Cholesky or AMG, are not competitive with respect to the direct solvers for the targeted problem size.

Table 2

Performance results for different software packages for the mechanical problem.

Software	Matrix Computation and Assembly	Solve	Solver Type
IFISS Q2-P1 (5e5 DOFs)	340	298	MATLAB \
FEMLAB 3.3 T2+P-1 (2e5 DOFs)	7	66	UMFPACK
		186	ILU-GMRES
MILAMIN (opt) T2+P-1 (1e6 DOFs)	15	34	CHOLMOD (AMD)

A performance comparison of MILAMIN for a mechanical test problem is given in Table 2. The domain is again a box containing a circular hole (free surface) and a circular inclusion with a ten times higher viscosity than the matrix. The outer boundaries are set to Dirichlet conditions representing pure shear deformation. The number of available packages to solve incompressible Stokes problems with heterogeneous material is greatly reduced compared to the thermal problem. In fact the IFISS package is not capable of dealing with heterogeneous materials and we used here an iso-viscous model. In the case of FEMLAB we had to employ the special MEMS module, which provides an incompressible Stokes application mode. However,

even with this specialized module we were unable to fit the test problem into the 2Gb RAM and therefore the results are provided for a five times smaller problem size. MILAMIN outperforms IFISS as well as FEMLAB both in terms of matrix computation and assembly, and the solution time. The latter demonstrates that iterative penalty approach chosen in MILAMIN and the resulting possibility to use a Cholesky solver (symmetric and positive definite system) is superior to other approaches.

Applications

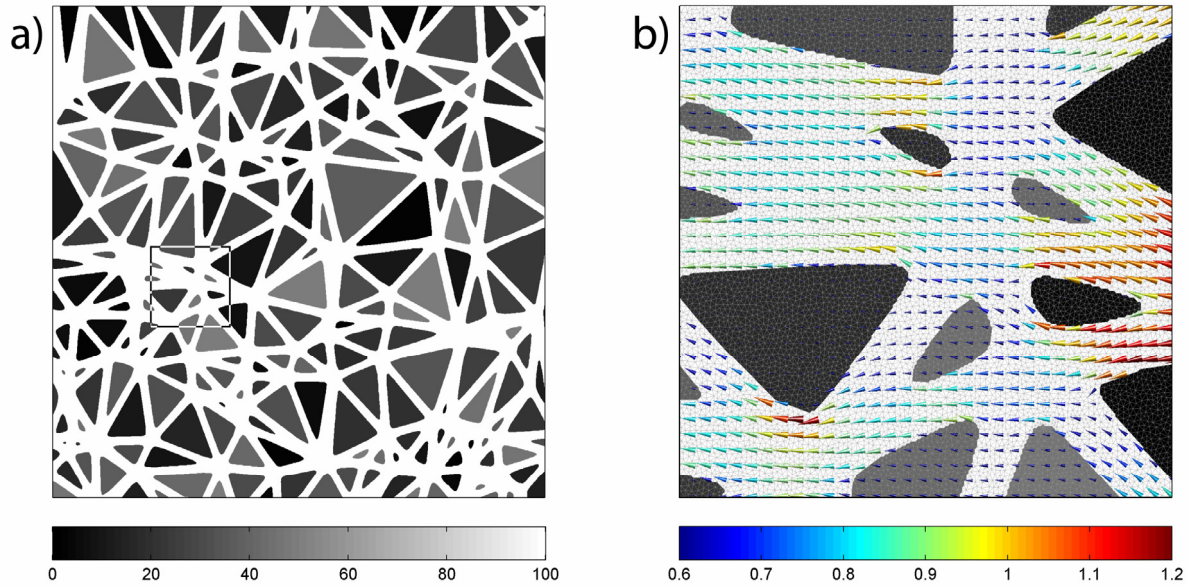


Figure 7

Illustration of a one million node application problem modeled with MILAMIN. Steady state diffusion is solved in a heterogeneous rock with channels of high conductivity. Heat flow is imposed by a horizontal thermal gradient; i.e. T(left boundary)=0, T(right boundary)=1, top and bottom boundary conditions are zero flux. a) Conductivity distribution. b) Flux visualized by cones and colored by magnitude. Normalization versus flux in homogeneous medium with conductivity of the channels. Background color represents the conductivity. Triangular grid is the finite element mesh used for computation. Note that this picture only corresponds to a small subdomain of a) (see square outline).

The power of MILAMIN to perform high resolution calculation for heterogeneous problems is illustrated with a thermal and a mechanical application example. Figure 7 shows the heat flux through a heterogeneous rock requiring approximately one million nodes to resolve it. Figure 8 shows a mechanical application of MILAMIN. Gravity driven incompressible Stokes flow is used to study the interaction of circular inclusions with different densities leading to a stratification of the material, see movie.

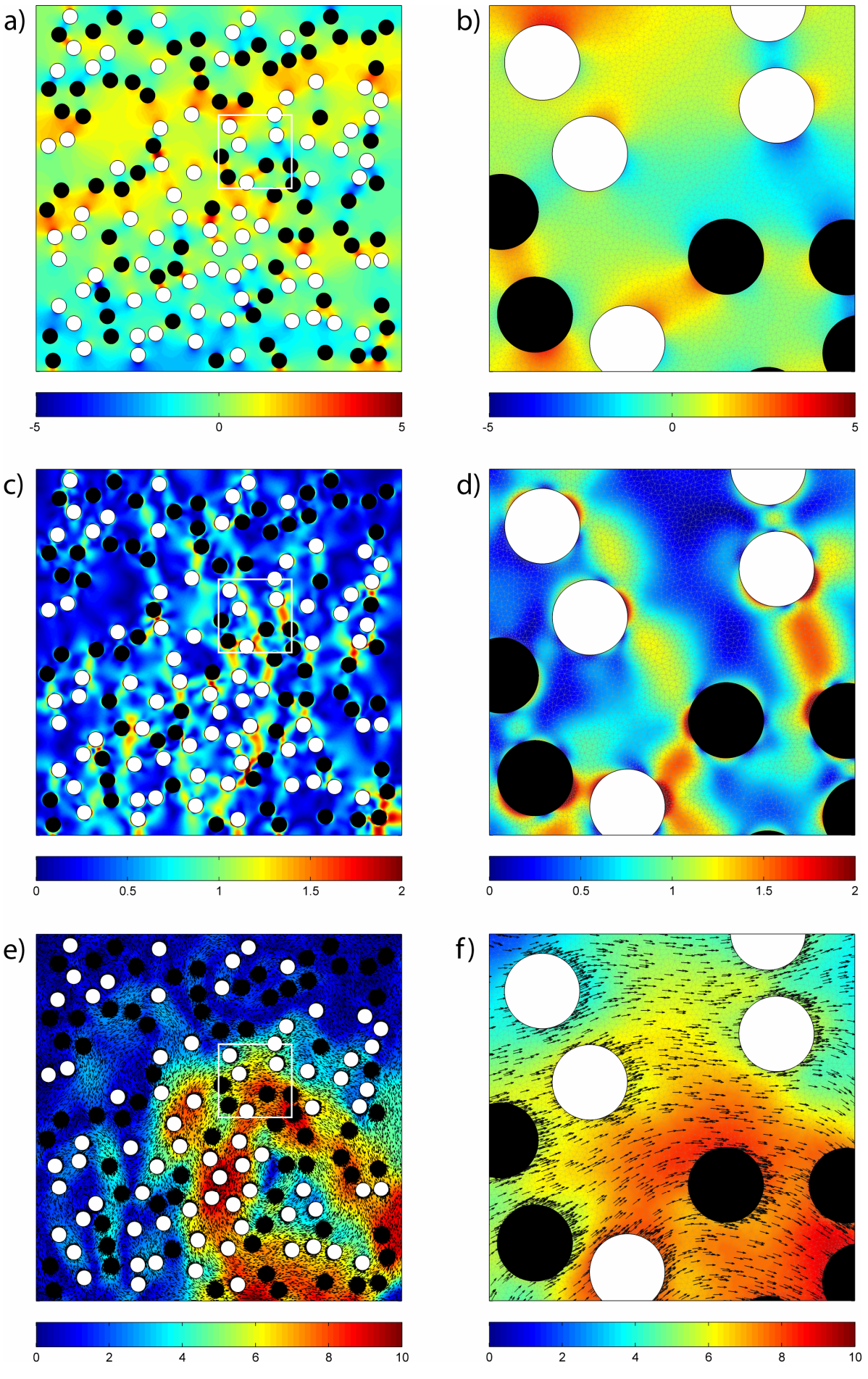

a)
-5
0
5
b)
-5
0
5
c)
0
0.5
1
1.5
2
d)
0
0.5
1
1.5
2
e)
0
2
4
6
8
10
f)
0
2
4
6
8
10

Figure 8

Mechanical application example. Circular inclusions in box subjected to vertical gravity field. Black (heavy) and white (light) inclusions have the same density contrast with respect to the matrix. They are hundred times more viscous than the matrix. First row shows (unsmoothed) pressure perturbations, second row maximum shear strain rate, and the third row the magnitude of the velocity field with superposed velocity arrows (random positions). All values are normalized by the corresponding maximum value generated by a single inclusion of the same size centered in the same box. The first column shows the entire domain, the second column a zoom-in with superposed finite element mesh according to the white square.

Movie 1

Time evolution of a model similar to what is shown in Figure 8. Viscosity ratio 1000:1.

MILAMIN allows not only to study the overall response of the system, but also resolves the details of the flow pattern around the heterogeneities. Note that no pressure oscillations problems can be observed that are often caused by the incompressibility constraint [e.g. *Pelletier, et al.*, 1989].

The MILAMIN strategies and package are applicable to a much broader class of problems than illustrated here. For example, transient thermal problems require only minor modifications to the thermal solver. As already mentioned the mechanical solver is devised in a way that compressible and incompressible elastic problems can be easily treated, simply by variable substitution. Coupled thermo-mechanical problems, arising for example in mantle convection, only require that the developed thermal and mechanical models are combined in the same time loop. This results in an unstructured, Lagrangian mantle convection solver capable of efficiently dealing with hundreds of thousands of nodes [cf., *Davies, et al.*, 2007].

Conclusions

We have demonstrated that it is possible to write an efficient native MATLAB implementation of the finite element method and achieved the goal to setup, process, and post-process thermal and mechanical problems with one million degrees of freedom in one minute on a desktop computer.

In our standard implementation we have combined all the state of the art components required in a finite element implementation. This includes among others: efficient pre-processing, fast matrix assembly, exploiting matrix symmetry for storage, and employing the best available direct solver and reordering packages. MATLAB-specific optimizations include proper memory management (pre-allocation of arrays) and data structures, explicit type declaration for in-

teger arrays, and efficient implementation of boundary conditions. In the case of the mechanical application the chosen penalty method together with the particular element type allows us to use the efficient Cholesky factorization to solve the incompressible flow problem. The clear structure of the code serves the educational purposes well. The results of our software comparison show that our standard version performs surprisingly efficient even compared to packages implemented in compiler languages.

Furthermore, in our optimized version we have improved the efficiency of the stiffness matrix calculations, which resulted in an overall execution speedup of approximately 4 times with respect to the standard version. This has been done by minimizing the ratio of overhead (BLAS and MATLAB) to computation. Another priority was to avoid unnecessary data transfers and promote cache reuse, as memory speed is a major bottleneck on current computer architectures. Particular optimizations to the matrix computation algorithm include 1) increased performance of the BLAS operations by interchanging loops and operating on large matrices, 2) reducing the total operation count by exploiting the symmetry of the system, and 3) facilitating cache reuse through the introduction of blocking.

Our implementation of the matrix computation achieves a sustained performance of 350 Mflops for any system size. Any further performance improvements to this part of the code are irrelevant, since even for smallest systems the matrix computation takes now only a fraction of the total solution time, with the solver being the bottleneck.

By paying attention to the strategies outlined in this article, MATLAB-based MILAMIN can not only be used as a development and prototype tool, but also as a production tool for the analysis of two dimensional problems with millions of unknowns within minutes. The complete MILAMIN source code is available from the authors and can be downloaded as an electronic supplement to this paper.

Acknowledgements

This work was supported by the Norwegian Research Council through a Centre of Excellence grant to PGP. We would like to thank Tim Davis, the author of the SuiteSparse package, for making this large suite of tools available and giving us helpful comments. We would also like to thank J. R. Shewchuk for making the mesh generator Triangle freely available. We are grateful to Antje Keller for her help regarding code benchmarking. The manuscript benefited from the reviews by Boris Kaus and Eh. Tan and the editorial work of Peter van Keken. Fi-

nally, we would like to thank Yuri Podladchikov for his never ending enthusiasm and stimulation.

References

Alberty, J., et al. (1999), Remarks around 50 lines of Matlab: short finite element implementation, *Numerical Algorithms*, *20*, 117-137.
Ashcraft, C., et al. SPOOLES, www.netlib.org/linalg/spooles/spooles.2.2.html.
Bathe, K.-J. (1996), *Finite element procedures*, XIV, 1037 pp., Prentice-Hall International, London etc.
Brezzi, F., and M. Fortin (1991), *Mixed and hybrid finite elements methods*, ix, 350 pp., Springer-Verlag, New York.
Chen, L., and C. Zhang AFEM@matlab, www.mathworks.com/matlabcentral/fileexchange.
Chen, Y., et al. (submitted), Algorithm 8xx: CHOLMOD, supernodal sparse Cholesky factorization and update/downdate.TR-2006-005, *ACM Trans. Math. Software*.
COMSOL FEMLAB, 3.3, www.femlab.com.
Cuvelier, C., et al. (1986), *Finite element methods and Navier-Stokes equations*, XVI, 483 pp., Reidel, Dordrecht a. o.
Davies, D. R., et al. (2007), Investigations into the applicability of adaptive finite element methods to two-dimensional infinite Prandtl number thermal and thermochemical convection, *Geochemistry Geophysics Geosystems*, *8*, -.
Davis, T. A. CHOLMOD, www.cise.ufl.edu/research/sparse/cholmod.
Davis, T. A. UMFPACK, www.cise.ufl.edu/research/sparse/umfpack.
Davis, T. A. (2006), *Direct Methods for Sparse Linear Systems*, SIAM
Davis, T. A., and W. W. Hager (2005), Row modifications of a sparse Cholesky factorization, *Siam Journal on Matrix Analysis and Applications*, *26*, 621-639.
Davis, T. A., and W. W. Hager (submitted), Dynamic supernodes in sparse Cholesky update/downdate and triangular solves. TR-2006-004, *ACM Trans. Math. Software*.
Dongarra, J. J., et al. (1990), A Set of Level 3 Basic Linear Algebra Subprograms, *Acm Transactions on Mathematical Software*, *16*, 1-17.
Dunavant, D. A. (1985), High Degree Efficient Symmetrical Gaussian Quadrature-Rules for the Triangle, *International Journal for Numerical Methods in Engineering*, *21*, 1129-1148.
Elman, H. C., et al. (2005), *Finite elements and fast iterative solvers with applications in incompressible fluid dynamics*, 400 pp., Oxford University Press, New York.
Ferencz, R. M., and T. J. R. Hughes (1998), Implementation of element operations, in *Handbook of numerical analysis*, edited by P. G. Ciarlet and J. L. Lions, pp. 39-52, North-Holland ; Distributors for the United States and Canada, Elsevier Science Pub. Co., Amsterdam ; New York
New York.
Fletcher, C. A. J. (1997), *Computational Techniques for fluid dynamics*, 3 ed., Springer, Berlin Heidelberg.
Gould, N. I. M., et al. (2007), A numerical evaluation of sparse direct solvers for the solution of large sparse symmetric linear systems of equations, edited, p. 10, ACM Press.
Hughes, T. J. R. (2000), *The finite element method : linear static and dynamic finite element analysis*, xxii, 682 p. pp., Dover Publications, Mineola, NY.
Hughes, T. J. R., et al. (1987), Large-Scale Vectorized Implicit Calculations in Solid Mechanics on a Cray X-Mp/48 Utilizing Ebe Preconditioned Conjugate Gradients, *Computer Methods in Applied Mechanics and Engineering*, *61*, 215-248.
Kwon, Y. W., and H. Bang (2000), *The finite element method using MATLAB*, 2nd ed., 607 pp., CRC Press, Boca Raton.

Limache, A., et al. (2007), The violation of objectivity in Laplace formulations of the Navier-Stokes equations, *International Journal for Numerical Methods in Fluids*, *54*, 639-664.
Patzak, B. OOFEM 1.7, www.oofem.org.
Pelletier, D., et al. (1989), Are FEM solutions of incompressible flows really incompressible - (or How simple flows can cause headaches), *International Journal for Numerical Methods in Fluids*, *9*, 99-112.
Persson, P. O., and G. Strang (2004), A simple mesh generator in MATLAB, *Siam Review*, *46*, 329-345.
Pozrikidis, C. (2005), *Introduction to finite and spectral element methods using MATLAB*, 653 pp., Chapman & Hall/CRC, Boca Raton.
Roddeman, D. Tochnog, feb11_2001, http://sourceforge.net/projects/tochnog.
Schenk, O., and K. Gärtner PARDISO, www.pardiso-project.org.
Shewchuk , J. (2007), Triangle, 1.6, www.cs.cmu.edu/~quake/triangle.html.
Sigmund, O. (2001), A 99 line topology optimization code written in Matlab, *Structural and Multidisciplinary Optimization*, *21*, 120-127.
Silvester, D. J. (1988), Optimizing Finite-Element Matrix Calculations Using the General Technique of Element Vectorization, *Parallel Computing*, *6*, 157-164.
Silvester, D. J., et al. IFISS, 2.2, www.maths.manchester.ac.uk/~djs/ifiss.
SIMULIA Abaqus, 6.6-1, www.simulia.com/products/abaqus_fea.html.
Toledo, S. (1997), Improving the memory-system performance of sparse-matrix vector multiplication, *Ibm Journal of Research and Development*, *41*, 711-725.
Toledo, S., et al. TAUCS, www.tau.ac.il/~stoledo/taucs.
Wesseling, P. (1992), *An introduction to multigrid methods*, 284 pp., Wiley, Chichester etc.
Zienkiewicz , O. C., and R. L. Taylor FEAPpv, 2.0, www.ce.berkeley.edu/~rlt/feappv.
Zienkiewicz, O. C., and R. L. Taylor (2000), *The finite element method*, 5th ed., Butterworth-Heinemann, Oxford ; Boston.

Appendix I: Variable Names

Table 3: MILAMIN variables. Note: "aeib" stands for "all elements in block".

Variable group	Variable	Size	Description
Variable size	ndim	1	Number of dimensions
	nel	1	Number of elements
	nnod	1	Number of nodes
	nnodel	1	Number of nodes per element
	nedof	1	Number of thermal or velocity degrees of freedom per element
	np	1	Number of pressure degrees of freedom per element
	nip	1	Number of integration points per element
	nelblo	1	Number of elements per block
	nblo	1	Number of blocks
	npha	1	Number of material phases
	nbc	1	Number of constraint degrees of freedom
	nfree	1	Number of unconstraint degrees of freedom
Mesh	ELEM2NODE	[nnodel, nel]	Connectivity
	Phases	[1, nel]	Phase of elements
	GCOORD	[ndim, nnod]	Global coordinates of nodes
Integration points, shape functions and their derivatives	IP_X	[ndim, nip]	Local coordinates of integration points
	IP_w	[1, nip]	Weights of integration points
	N	{nip*[nnodel, 1]}	Cell array of nip entries of shape functions Ni evaluated at integration points
	dNdu	{nip*[nnodel, ndim]}	Cell array of nip entries of shape functions derivatives wrt local coordinates dNdui evaluated at integration points
Geometry	ECOORD_X	[ndim, nnodel]	Global coordinates of nodes in element
	J	[ndim, ndim]	Jacobian in integration point
	invJ	[ndim, ndim]	Inverse of Jacobian
	detJ	1 or [nelblo,1]	Determinant of Jacobian (or aeib)
	dNdX	[nnodel, ndim]	Shape function derivatives wrt global coordinates in integration point
	ECOORD_x, ECOORD_y	[nnodel, nelblo]	Global x and y-coordinates for nodes (aeib)
	Jx, Jy	[nelblo, ndim]	1st (x) and 2nd (y) row of Jacobian in integration point (aeib)
	invJx, invJy	[nelblo, ndim]	1st (x) and 2nd (y) column of inverse of Jacobian (aeib)
	dNdx, dNdy	[nelblo, nnodel]	Shape function derivatives wrt global x and y-coordinate (aeib)
Auxiliary arrays	indx_l	[nedof*(nedof+1)/2,1]	Indices extracting lower part of element matrix
Boundary conditions	Free	[1, nfree]	Unconstraint degrees of freedom
	Bc_ind	[1, nbc]	Constraint degrees of freedom
	Bc_val	[1, nbc]	Constraint boundary values
Solution	perm	[1,nfree]	Permutation vector reducing factorization fill-in
	L	[nfree, nfree]	Sparse lower Cholesky factor of global stiffness matrix
	Rhs	[nfree, 1]	Global right hand side vector
THERMAL			
Materials	D	[npha,1]	Conductivities for different phases
	ED	1 or [nelblo,1]	Conductivity of element (or aeib)
Matrix calculations	K_elem	[nnodel, nnodel]	Element stiffness matrix
	K_block	[nelblo, nnodel*(nnodel+1)/2]	Flattened element stiffness matrices (aeib)
Triplet storage	K_i	[nnodel*(nnodel+1)/2, nel]	Row indices of triplet sparse format for K_all
	K_j		Column indices of triplet sparse format for K_all
	K_all		Flattened element stiffness matrices for all elements
Solution stage	K	[nfree, nfree]	Sparse global stiffness matrix (only lower part)
	T	[nnod, 1]	Unknown temperature vector

MECHANICAL			
Materials	Mu, Rho	[npha,1]	Viscosity and density for different phases
	EMu, ERho	1 or [nelblo,1]	Viscosity and density of element (or aeib)
Matrix calculations	Pi	[np,1]	Pressure shape functions in integration point
	P	[np, np]	Auxiliary matrix containing global coordinates of the corner nodes
	Pb	[np,1]	Auxiliary vector containing global coordinates of integration point
	B	[nedof, ndim*(ndim+1)/2]	Kinematic matrix
	A_elem	[nedof, nedof]	Element stiffness matrix (velocity part)
	Q_elem	[np, nedof]	Element divergence matrix
	M_elem	[np, np]	Element pressure mass matrix
	invM_elem	[np, np]	Inverse of element pressure mass matrix
	Rhs_elem	[ndim, nedof]	Element right hand side vector
	PF	1	Penalty factor
	GIP_x, GIP_y	[1,nelblo]	Global x and y-coordinates of integration point (aeib)
	Pi_block	[nelblo, np]	Pressure shape functions in integration point (aeib)
	A_block	[nelblo, nedof*(nedof+1)/2]	Flattened element stiffness matrices (aeib)
	Q_block	[nelblo, nedof*np]	Flattened element divergence matrices (aeib)
	M_block	[nelblo, np*(np+1)/2]	Flattened element pressure mass matrices (aeib)
	invM_block	[nelblo, np*np]	Flattened inverses of element pressure mass matrices (aeib)
	Rhs_block	[nelblo, nedof]	Element right hand side vectors (aeib)
Triplet storage	Rhs_all	[nedof, nel]	Element right hand side vectors for all elements
	A_i	[nedof*(nedof+1)/2, nel]	Row indices of triplet sparse format for A_all
	A_j		Column indices of triplet sparse format for A_all
	A_all		Flattened element stiffness matrices for all elements
	Q_i	[nedof*np, nel]	Row indices of triplet sparse format for Q_all
	Q_j		Column indices of triplet sparse format for Q_all
	Q_all		Flattened element divergence matrices for all elements
	invM_i	[np*np, nel]	Row indices of triplet sparse format for invM_all
	invM_j		Column indices of triplet sparse format for invM_all
	invM_all		Flattened inverses of element pressure mass matrices for all elements
Solution stage	A	[nfree, nfree]	Sparse global stiffness matrix (only lower part)
	Q	[np*nel, ndim*nnod]	Sparse divergence matrix
	invM	[np*nel, np*nel]	Sparse pressure mass matrix
	Div	[nel*np, 1]	Quasi-divergence vector
	Vel	[ndim*nnod, 1]	Unknown velocity vector
	Pressure	[nel*np, 1]	Unknown pressure vector

Paper 2: Structure Development around a Rigid Circular Inclusion in an Anisotropic Host Subject to Simple Shear

Submitted to Journal of Structural Geology by M. Dabrowski and D.W. Schmid

Abstract

We study the initial stages and the finite strain evolution of the behavior of a rigid circular inclusion embedded in an anisotropic matrix subjected to anisotropy parallel simple shear. The model is two dimensional, plane strain, and linear viscous materials are assumed. The initial instantaneous flow stages are studied with an analytical solution that is based on Willis (1964). The finite strain evolution of the inclusion and the structural development are studied with a finite element method model. This model also allows for studying the effect of explicitly layered host material and the correspondence to effective anisotropic approximations and sheds light on the issue of upscaling versus resolving layered medium. Investigated aspects include the inclusion motion, flow field rearrangement and structural development in the host. Our results show that matrix anisotropy has a first order effect on the motion of a rigid heterogeneity subject to shear and the development of structures around it. We are able to demonstrate that this is the case for even weakly anisotropic hosts if the total inclusion rotation is considered. Strong anisotropy leads to substantially reduced rotation rates once a shear strain magnitude of 2 is reached. This effect and related developments in the structural development in the matrix may yield finite strain structures that appear to be the result of substantially lower strains. Furthermore, perturbation flow ranges are much larger in anisotropic material and therefore boundary effects and flow confinement and interaction distances much larger, which also have to be considered when interpreting natural structures. The comparison of layered and anisotropic hosts reveals that the effective approach is good when inclusion motion is considered. However, the detailed structural evolution in a layered matrix, e.g. development of rootless folds, thickening of strong layers, deformation localization in weak layers, and break in orthotropic symmetry, can only be studied if the layers are explicitly resolved.

Introduction

Most geological materials are heterogeneous, with the heterogeneities ranging from sub-grain to kilometer scales. Preferentially oriented crystal lattices of rock forming minerals may give

rise to an overall anisotropy (e.g. Kocks et al. 2000), but this effect may be equally well related to shape alignment of its heterogeneous constituents or ultimately layering. The relation between heterogeneity and anisotropy has been recently quantified by Treagus (2003) and Fletcher (2004). The study of the deformation in and around heterogeneities and in anisotropic materials represents an important cornerstone of structural geology. Examples include rotation and shape evolution of inclusions (e.g. Bilby & Kolbuszewski 1977, Eshelby 1957, e.g. 1976, Jeffery 1922, Jezek et al. 1996, Marques et al. 2005, Mulchrone 2007, Schmid & Podladchikov 2004), pattern formation around inclusions (e.g. Bons et al. 1997, Kocher & Mancktelow 2005, Mandal et al. 2005, Masuda & Mizuno 1996, Passchier & Simpson 1986, Rosas et al. 2002, Schmid & Podladchikov 2005), interactions between heterogeneities (e.g. Arbaret et al. 2001), single and multilayer folding (e.g. Biot 1961), chevron folding (e.g. Bayly 1970), formation of kink bands or internal instabilities in general (e.g. Cobbold et al. 1971).

Studies where both a distinct heterogeneity and material anisotropy are present and their mutual interaction is investigated are relatively few (e.g., Kocher & Mancktelow 2006, Kocher et al. 2006, Muhlhaus et al. 2002a, Muhlhaus et al. 2002b). In the work of Kocher and Mancktelow (2006) a flanking structure is approximated as a weak elliptical inclusion that is embedded in an anisotropic matrix and consequently subject to shear. A finding is that the motion of such an inclusion is identical to the case of a strongly elliptical weak inclusion embedded in an isotropic matrix, which is equivalent to that of a passive marker. This is rather surprising as one could expect that the intricate development of the anisotropy structure around such an inclusion could substantially influence its behavior, similar to how matrix anisotropy influences folding (Kocher et al. 2006). The question arises to which extent the findings of Kocher and Mancktelow (2006) hold for other cases of inclusion-matrix systems.

We study the behavior of a rigid circular inclusion in simple shear as a function of matrix anisotropy. The initial stages where the trace of the matrix anisotropy is undisturbed are analyzed with an analytical solution. This is not possible for the finite strain evolution due to the heterogeneous evolution of the anisotropy structure of the matrix and we address these stages with a finite element method (FEM) model. In particular we investigate i) the flow evolution in the matrix as a function of anisotropy, ii) the motion of the inclusion, iii) the development of structures in the anisotropic matrix, and iv) the correspondence between explicitly layered matrix material and homogenously anisotropic material.

Mathematical Model

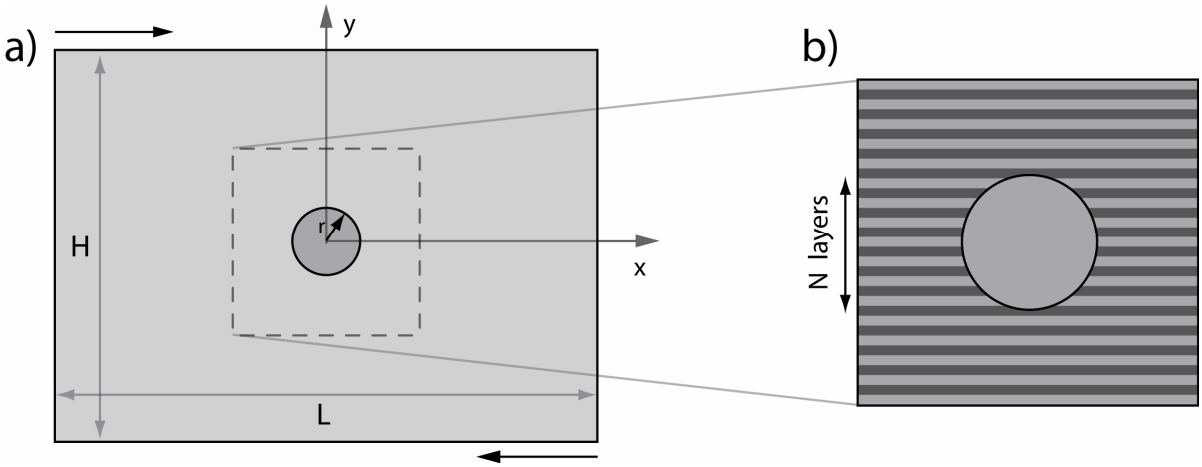


Figure 1

Schematic models of the two different categories of inclusion-host systems: a) A rigid circular inclusion of a radius r embedded in a homogeneous anisotropic host with the anisotropy trace aligned with the x-direction. In the analytical setup the system is loaded at infinity. In the case of the numerical setup a rectangular computational domain of a height H and a width L is used and loads are prescribed on its boundaries. b) In the second model category, competent and incompetent layers are explicitly present in the host. The number of individual layers across the inclusion is varied throughout the model instances.

We study the flow perturbation around a circular inclusion embedded in a viscous host subjected to a simple shear in two different kinds of setups, namely the homogeneous and explicitly layered host cases (see Figure 1). Our analysis is restricted to two-dimensions and the plane strain condition is present throughout it. The inclusion is modeled as a rigid object and the complete coherence at the inclusion and host interface is assumed. We allow for the mechanical anisotropy of a host material given by a shear μ_s and normal μ_n viscosity for the shearing parallel and longitudal to the anisotropy trace, respectively. Using the Cartesian coordinate system aligned with the anisotropy direction, the constitutive relationship for an incompressible orthotropic or transversally isotropic medium reads

$$\begin{aligned} \sigma_{xx} &= -p + 2\mu_n D_{xx} \\ \sigma_{yy} &= -p + 2\mu_n D_{yy} \\ \sigma_{xy} &= 2\mu_s D_{xy} \end{aligned} \tag{1}$$

where p and σ denote pressure and the stress tensor, respectively. The strain rate tensor D is defined by the usual kinemetic relations

$$D_{xx} = \frac{\partial v_x}{\partial x}, \quad D_{yy} = \frac{\partial v_y}{\partial y}, \quad D_{xy} = \frac{1}{2}\left(\frac{\partial v_x}{\partial y} + \frac{\partial v_y}{\partial x} \right) \tag{2}$$

where v is the velocity vector field. Throughout the constitutive equations (1), we have assumed the incompressibility condition complemented by the plane strain assumption that yield the following constraints on the strain rate components

$$D_{xx} + D_{yy} = 0, \quad D_{xz} = D_{yz} = D_{zz} = 0 \tag{3}$$

Finally, the momentum conservation for a homogeneous anisotropic medium with a vanishing body force under a creeping flow regime (Stokes flow) and subject to the introduced assumptions is given in differential form by

$$\begin{aligned} -\frac{\partial p}{\partial x} + \frac{\partial}{\partial x}\left(2\mu_n \frac{\partial u_x}{\partial x} \right) + \frac{\partial}{\partial y}\left(\mu_s \left(\frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x} \right) \right) = 0 \\ -\frac{\partial p}{\partial y} + \frac{\partial}{\partial y}\left(2\mu_n \frac{\partial u_y}{\partial y} \right) + \frac{\partial}{\partial x}\left(\mu_s \left(\frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x} \right) \right) = 0 \end{aligned} \tag{4}$$

It is worth noting that an arbitrarily oriented anisotropy trace with respect to the chosen global reference frame necessitates the tensorial transformation of the constitutive relationship (1). The latter one is provided in the second part of our paper where we allow for a non-uniform evolution of the anisotropy direction in the host during progressive deformation. In this case the momentum conservation given by (4) is not valid any more. However, it is automatically taken care of in our numerical solver that relies on a general integral formulation of the mechanical equilibrium and the appropriate constitutive relationship.

It has been previously noted that the overall anisotropy of a property such as viscosity may stem from an intrinsic mechanical anisotropy of rock constituents or is a collective behavior of adequately structured but otherwise isotropic phases. In this study, we presume that the mechanical anisotropy in the host is derived from the presence of alternating incompetent and competent layers of viscosity μ_1 and μ_2, respectively. In the first category of the studied models, we postulate that the layer thickness is negligible in comparison to the inclusion size and the host layering serves as a microscopic motivation. However, such a layering of a host material is indeed present in the second category of our models that is referred as the explicitly layered host case (see Figure 1).

We frequently use a conventional anisotropy factor δ defined as a ratio of the normal and shear viscosity that provides a measure of the anisotropy strength

$$\delta = \frac{\mu_n}{\mu_s} \tag{5}$$

It is convenient to relate this anisotropy factor to a viscosity ratio and phase fraction f of the weak and strong component in an explicitly layered system in order to make a quantitative link between the two different model categories. A simple calculation shows that the overall mechanical response of such a layered medium is described by a shear μ_s and normal μ_n viscosity that replace the isotropic viscosities μ_1 and μ_2 (Biot 1965).

$$\mu_n = f_1\mu_1 + f_2\mu_2, \quad \mu_s = \left(\frac{f_1}{\mu_1} + \frac{f_2}{\mu_2} \right)^{-1} \tag{6}$$

The shear and normal viscosity are given by the weighted harmonic and arithmetic mean, respectively. Thus, the normal viscosity is never smaller than the shear viscosity irrespective of the phase abundances. In addition, these viscosities coincide with the theoretical lower and upper bounds on an effective viscosity in a composite linear medium (Hill 1952). In this study we consider equal area fractions of the competent and incompetent phase, i.e. $f_1 = f_2 = 0.5$. In this particular case, the anisotropy factor of a composite consisting of two predefined phases assumes the maximum value that reads

$$\delta = \frac{1}{4}\left(\sqrt{m} + \frac{1}{\sqrt{m}} \right)^2 \tag{7}$$

where we have introduced a viscosity ratio factor $m = \mu_2 / \mu_1$. In order to provide the remaining part of the analyzed link, we invert the relation (7) to obtain

$$m = \frac{\sqrt{\delta} + \sqrt{\delta - 1}}{\sqrt{\delta} - \sqrt{\delta - 1}} \tag{8}$$

where we assume $\delta \geq 1$.

Instantaneous flow pattern

The basis for the understanding of the finite strain evolution around a rigid inclusion in an anisotropic matrix is the initial instantaneous flow stage. We study this with analytical and finite element methods. This allows us to establish the basic characteristics of the studied system, benchmark the numerical model, quantify boundary effects, and put bounds on the applicability of anisotropic material approximation.

Analytical solution

The analytical solution for an elliptical inclusion in an anisotropic matrix, which covers arbitrary viscosity ratios, compressibility, and general uniform far field loads was derived by Willis (1964). Here, we look at the end member case where a rigid circular inclusion with radius r is embedded in a homogeneously anisotropic matrix and subjected to a far field simple shear with rate $2D_{xy}^{\infty}$ that is parallel to the direction of the anisotropy.

The complete perturbation velocity field outside the inclusion can be derived (see "Appendix I: Analytical solution" for details):

$$V = iD_{xy}^{\infty}\frac{r}{4\gamma^2}\left(\gamma\left(\sqrt{X^{+}-4\gamma}-\sqrt{X^{-}+4\gamma}+X^{-}-X^{+}\right)+\sqrt{\bar{X}^{+}-4\gamma}+\sqrt{\bar{X}^{-}+4\gamma}+\bar{X}^{-}+\bar{X}^{+}\right) \quad (9)$$

Note that we use complex notation for the coordinates, $X = x + iy$, and perturbation velocity, $V = v_x + iv_y$. Furthermore, the following coordinate transformation is introduced

$$X^{\pm} = X \pm \gamma\bar{X} \quad (10)$$

and γ is related to the anisotropy factor by

$$\gamma^2 = \frac{\sqrt{\delta}-1}{\sqrt{\delta}+1} \quad (11)$$

The strain rate field can be obtained upon differentiation of the eqn. (9) according to the formulas (2).

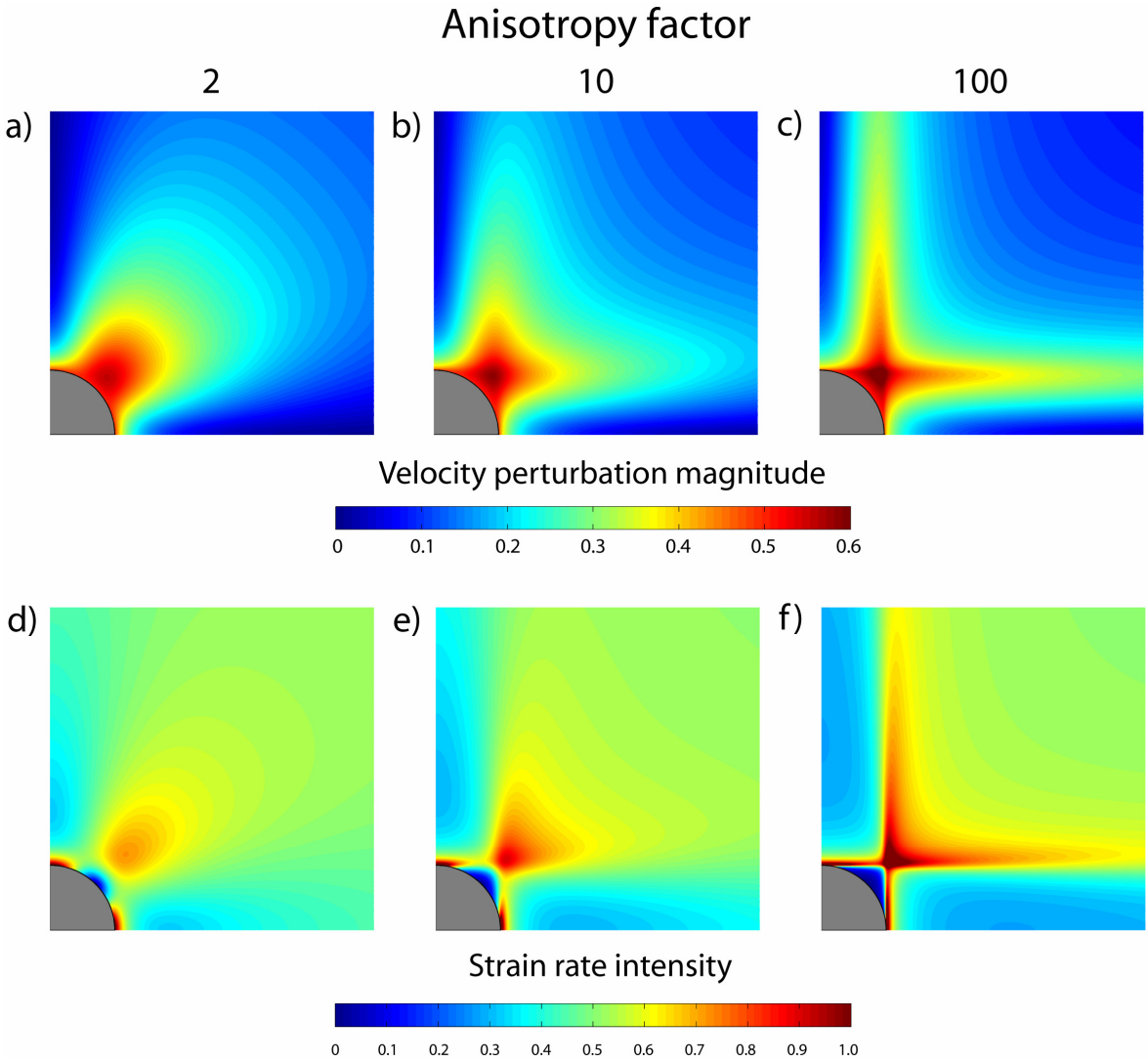


Figure 2

Velocity perturbation magnitude (a, b and c) and strain rate intensity (d, e and f) fields around a rigid circular inclusion embedded in an infinite homogenous anisotropic host and subjected to a simple shear with the unit rate at infinity. Columns show results for anisotropy factors 2, 10 and 100, respectively.

The magnitude of the flow perturbation field and strain rate intensity for different anisotropy factors is depicted in Figure 2. The results show that the anisotropy of the host material has strong impact on the characteristics of the perturbation flow around the inclusion. The first column shows the weak anisotropy case that exhibits relatively smooth variations similar to the isotropic case (cf. Schmid & Podladchikov 2003). With increasing anisotropy these fields start to focus in the principal directions of the anisotropy. The strain rate maxima adjacent to the inclusion are enhanced and gradually spread in the direction of the anisotropy. For high anisotropy factors these maxima merge and form high strain rate segments that follow the outline of the bounding square to the inclusion. The interior of this square is shielded from the deforma-

tion through the mentioned zones of a localized deformation. Thus, effectively the rigid inclusion becomes a square. The secondary strain rate maxima, initially present at some distance from the inclusion, are also enhanced with an increasing anisotropy factor and migrate inwards towards the corners of the bounding square. Moreover, the initial high strain rate oval regions split and penetrate outwards into the host as conjugate bands that sharpen and approach the anisotropy direction with increasing its strength. In contrary, the initial secondary low strain rate regions extend according to their original elongation that coincides with the anisotropy direction. The resulting bands of a suppressed shearing inherit the width of the inclusion.

It is evident based on a symmetry that the rotation rate of a circular inclusion in an orthotropic medium subjected to a pure shear vanishes. Thus, in the case of simple shear the rotation rate of a circular inclusion equals the magnitude of the background flow vorticity, which is half of the simple shear rate. Furthermore, a general result stating that the rotation rate of an elliptical inclusion in an incompressible anisotropic medium mimics the isotropic case has been derived (R.C. Fletcher, publication submitted). Later, we will show that the notion of the host homogeneity is crucial for this result to hold.

In order to gain further insights into the characteristics of the perturbation flow we have derived the infinite anisotropy factor limit of eqn. (9). The flow field can now be expressed as

$$\begin{aligned} v_x &= \begin{cases} D_{xy}^{\infty}\,\mathrm{sgn}(y)\sqrt{y^2-1} & \|y\| \geq 1 \\ 0 & \|y\| < 1 \end{cases} \\ v_y &= \begin{cases} D_{xy}^{\infty}\,\mathrm{sgn}(x)\sqrt{x^2-1} & \|x\| \geq 1 \\ 0 & \|x\| < 1 \end{cases} \end{aligned} \tag{12}$$

In this degenerate case only the off-diagonal components of the strain rate tensor survive. The strain rate intensity and vorticity of the field (12) is shown in Figure 3 together with the quiver plots of the total flow and the corresponding one with the background rotation removed.

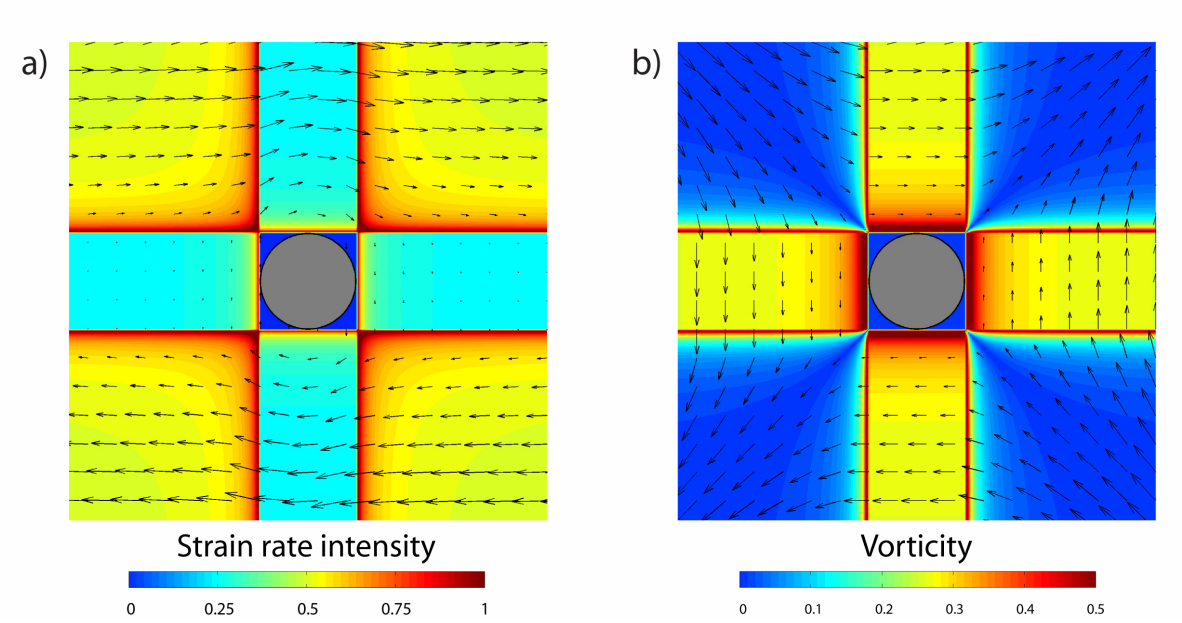


Figure 3

Strain rate intensity (a) and vorticity magnitude field (b) in the limit of the infinite anisotropy strength. The flow field is visualized with the overlying quiver plots.in (b) the uniform background vorticity component of the simple shear flow is subtracted from both the field values and the vector field

By inspection of the flow field, we identify three distinct regions:

1. The velocity field vanishes in the square domain that outlines the inclusion. Thus, this confirms our observations that in the limit of an infinite anisotropy factor the problem of a rigid circular inclusion subjected to a simple shear load aligned with the anisotropy direction is equivalent to the problem of a rigid square inclusion.

2. Sharp bands forming a cross propagate from the center along the anisotropy directions and the width of these bands is dictated by the inclusion size. Neglecting the background rotational component, flow vectors within these bands are perpendicular to their boundaries. At some distance from the inclusion, the increase of the velocity magnitude is closely approximated by a linear function and results in a simple shear flow of a rate equal to half of the far field value. Consequently, the strain rate intensity and vorticity magnitude assume a quarter of a far field simple shear rate. Towards the rim of the inclusion, both the strain rate intensity and vorticity diverge within the bands.

3. The flow field in the remaining part of the host outside the cross is quickly adjusted to the far field uniform simple shear flow. Again, the strain rate and vorticity diverge towards the bands.

Numerical solution

In this section we present results obtained with a modified version of MILAMIN, our implicit FEM code that is capable of solving incompressible Stokes flow. In both studied cases, anisotropic and layered matrix, large numerical resolutions are required in order to resolve either flow complexities or layer geometry. The optimizations that have been implemented into MILAMIN allow for solving mechanical problems around one million of nodes on a desktop computer in the order of a minute (Dabrowski et al. in press). Details of the implementation and other numerical considerations can be found in "Appendix II: Finite Element Model".

Homogeneous anisotropic host

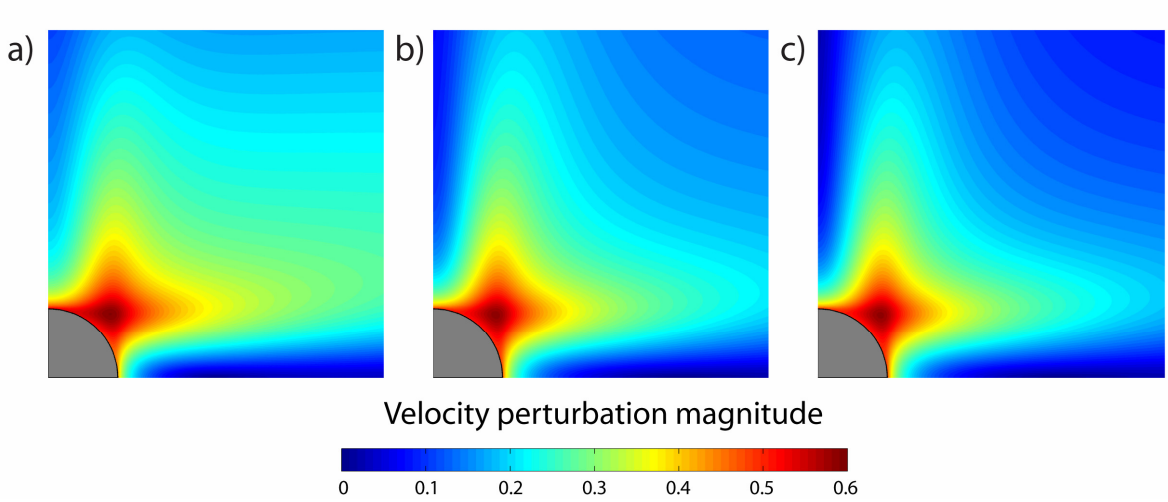


Figure 4

Magnitude of perturbation velocity obtained with FEM for anisotropy factor 10. a) Inclusion size/box height ratio 0.1, box aspect ratio 1, b) Inclusion size/box height ratio 0.1, box aspect ratio 2, c) Inclusion size/box height ratio 0.05, box aspect ratio 2. Note: Actual numerical models are always larger than what is displayed.

The first set of numerical models investigates the correspondence between the analytically obtained results and the numerical ones, Figure 4. The setup consists of a box with varying aspect ratio (width/height) and a circular inclusion located in the center (see Figure 1). The inclusion is given a high viscosity so that it effectively behaves like a rigid object (we have used four orders higher viscosity than the normal viscosity of the host material). Top and bottom boundary conditions are constant (simple shear) velocities and periodicity is enforced on the lateral walls.

Figure 4 shows that the velocity perturbation patterns display a close resemblance to the distribution predicted analytically (cf. Figure 2b). However, due to the boundaries located at a finite distance from the inclusion discrepancies are to be expected. The boundary effect is still quite

strong for an inclusion size that corresponds to 10% of the domain height. We have found that increasing the aspect ratio of the box from one to two can substantially reduce the boundary effect, Figure 4b. However, further reduction of the boundary effect requires changing the inclusion size rather than the aspect ratio of the box, Figure 4c.

It has been noticed that the boundaries also influence the rotation rate of the inclusion. This effect is again to be expected, since the lateral and horizontal wall boundary conditions are not equivalent in these models. The rotation rate of the inclusion provides an important measure of the boundary effect; the analytically predicted value being 0.5. In the case of the anisotropy factor 10 and the three setups depicted in Figure 4, the rotation rate is 0.445, 0.485 and 0.492, respectively. With increasing strength of anisotropy the boundary effect become stronger; e.g. the rotation rate equals 0.435 for the setup in Figure 4c with anisotropy factor 100.

We conclude that using the inclusion size of 5% of the domain height and the box aspect ratio of two provides a sufficient setup to avoid a strong influence of the boundary effects for moderate anisotropy factors while still being able to resolve it numerically.

Layered host

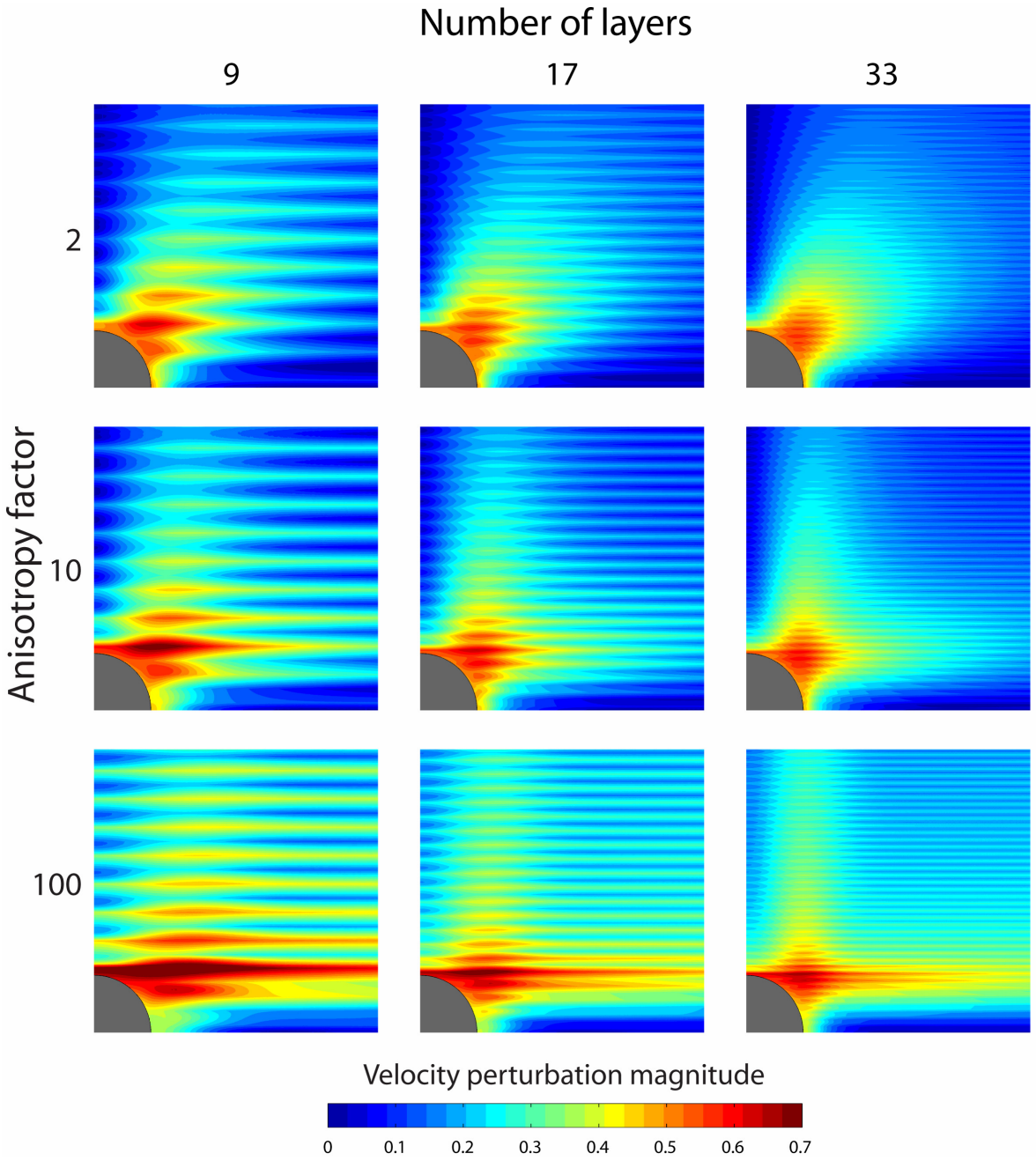


Figure 5

Perturbation flow magnitude around a rigid circular inclusion subjected to simple shear deformation at the boundaries (not shown on these portions). Inclusion size/box height ratio is 0.05 and box aspect ratio is 2 for all models. Isotropic competent and incompetent layers are present in the host material. The top-most point of the inclusion is within a competent layer. Number of layers across the inclusion is fixed in the columns to 9, 17 and 33. Viscosity ratio of the strong and weak layers is constant in the rows and is set to ca. 6, 38, 380; values that correspond to anisotropy factor 2, 10 and 100, respectively.

To study the correspondence between the anisotropic and layered host cases, we compare the influence of the finite layer thickness on the perturbation flow pattern by systematically varying layer thickness/inclusion size ratio. Hence, we introduce equally spaced weak and competent layers in the host material. Following the findings of the previous section the box has aspect ratio two (width/height) and the effectively rigid inclusion is 5% of the box height subjected to the same set of boundary conditions. In order to facilitate comparison with the results obtained with the anisotropic code (anisotropy factors 2, 10 and 100), the viscosity ratio of the layers has been set to ~6, 38 and 380 according to the eqn. (8). We have analyzed the cases of 9, 17 and 33 layers across the inclusion, which results in 180, 340 and 660 layers in the whole model, respectively. The lower- and uppermost points of the inclusion are located in the middle of the strong layers. In this symmetric configuration translation of the inclusion is inhibited.

A comparison of Figure 2 and Figure 5 shows that the perturbation flow converges towards the anisotropic host solution in the limit of the thin layering. Already for the analyzed 33 layers case, the similarity of the flow perturbation pattern and the anisotropic solution is rather evident. In addition, weaker viscosity contrast between the layers promotes the applicability of the anisotropic medium approach. Similar conditions describing the validity of representation by a continuous media were already formulated by Biot (1965, p. 191).

Our FEM results indicate that up-scaling of the mechanical behavior of a layered medium using effective anisotropic material properties is a valid approach. Nevertheless, discrepancies exist between the two model categories. We find that the major difference between the two is the lack of orthotropic symmetry in the layered host case. By introducing explicit layering one direction of the anisotropy is favored and it is clearly reflected in the flow perturbation pattern. Here, the horizontal high strain band is developed stronger than the vertical one with shearing focused into the weak layers. The latter observation suggests the possibility of inclusion-host decoupling, which is likely to be strong for coarser layering.

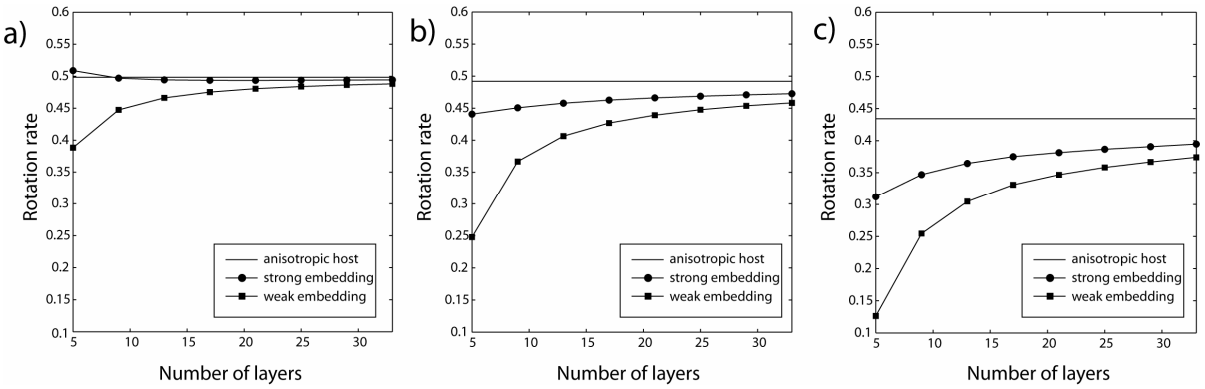


Figure 6

Rotation rate of a rigid circular inclusion embedded in a layered host ffor different anisotropy factors:a) 2,b) 10, c) 100. Results are shown for strong and weak embedding as a function of number of layers across the inclusion. Reference rotation rates obtained for a layer-free anisotropic host, which deviate from 0.5 due the influence of boundaries.

Figure 6 shows the rotation rate of the inclusion as a function of layer thicknesses in the matrix. Apart from the weakly anisotropic model, Figure 6a, the rotation rate decreases with coarsening the layers. A decrease in the rotation rate is consistent with the previously mentioned shear localization into the weak layers in the vicinity of the inclusion. For completeness, we have analyzed the case of the weak embedding when the marginal layers are incompetent. This results in consistently lower rotation rates than in the strong embedding case and again the rotation rate decreases as the layering coarsens. For anisotropy factor 100, a value as low as ca. 0.12 has been obtained.

Finite strain evolution

We now analyze the finite strain evolution of the studied models. Due to the structural development in the matrix the anisotropy will emerge heterogeneously and it is therefore no longer feasible to obtain analytical solutions and we restrict the analysis to FEM models. We first extend the constitutive relations to the case of an arbitrarily orientated anisotropy trace with respect to the flow directions and present the update rule for the anisotropy orientation that is concordant with the finite strain evolution of a layered medium. Then, we present the results of numerical experiments where a simple shear deformation of magnitude five is reached.

Director theory

The constitutive relationship (1) requires that the anisotropy directions are aligned with the axes of the global coordinate system. During progressive deformation the flow perturbation in the vicinity of the inclusion will lead to heterogeneous advection and reorientation of the initially homogeneously oriented anisotropy traces. First, we need to provide the constitutive law

for an anisotropic fluid in a form that covers the case of an arbitrarily directed anisotropy. Under our two-dimensional assumptions the normal to the layering or the axis of a transversally isotropic host is expected to remain within the analyzed plane. Introducing the inclination φ of the anisotropy trace with respect to the global Cartesian coordinate system xy and applying the tensorial transformation rules, we obtain the following constitutive relationship for an incompressible transversally isotropic medium:

$$\begin{pmatrix} \sigma_{xx} \\ \sigma_{xy} \end{pmatrix} = \begin{pmatrix} -p \\ 0 \end{pmatrix} + \begin{pmatrix} \mu_n \cos^2 2\varphi + \mu_s \sin^2 2\varphi & (\mu_n - \mu_s)\sin 2\varphi \cos 2\varphi \\ (\mu_n - \mu_s)\sin 2\varphi \cos 2\varphi & \mu_n \sin^2 2\varphi + \mu_s \cos^2 2\varphi \end{pmatrix} \begin{pmatrix} D_{xx} \\ D_{xy} \end{pmatrix} \tag{13}$$

In addition to the constitutive relationship and the usual conservation laws, we require the evolution equation of the anisotropy orientation described in general by the so-called director field $\vec{n}(\vec{x},t)$ that represents the normal to the anisotropy plane. In the context of layered materials, the evolution law of the director field that coincides with the material derivative of the normal to a surface element is postulated. The transformation of the surface element $\vec{n}da$ under the deformation gradient F is given by Nanson's formula

$$n_k' da' = JF_{ik}^{-1} n_i da \tag{14}$$

where J is Jacobian that is equal to 1 in the incompressible case. By applying the material time derivative and decomposing the velocity gradient into strain rate D and spin W, we arrive at

$$\frac{Dn_k}{Dt} = [W_{kp} - (D_{mk} n_m n_p - D_{pl} n_l n_k)] n_p \tag{15}$$

Eqn. (15) coincides with the evolution law postulated in the frame of the Ericksen transversely istotropic fluid describing uniaxial nematic liquid with the tumbling parameter set to one.

Numerical results

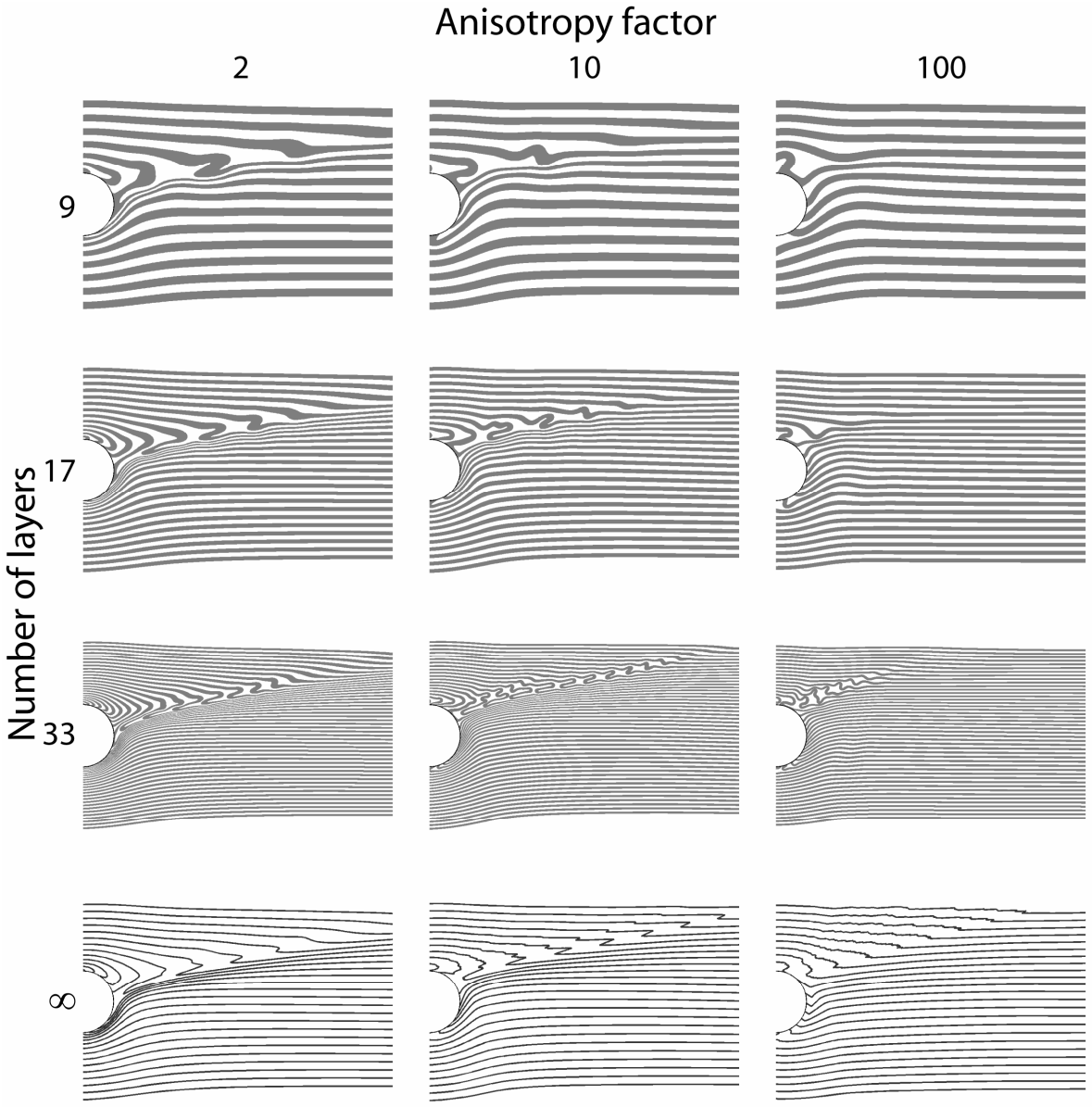


Figure 7

Matrix structures after simple shear deformation of magnitude 5. Small portion of the actual models is displayed. The anisotropy factor is constant in columns and the number of layers across the inclusion is fixed in rows.

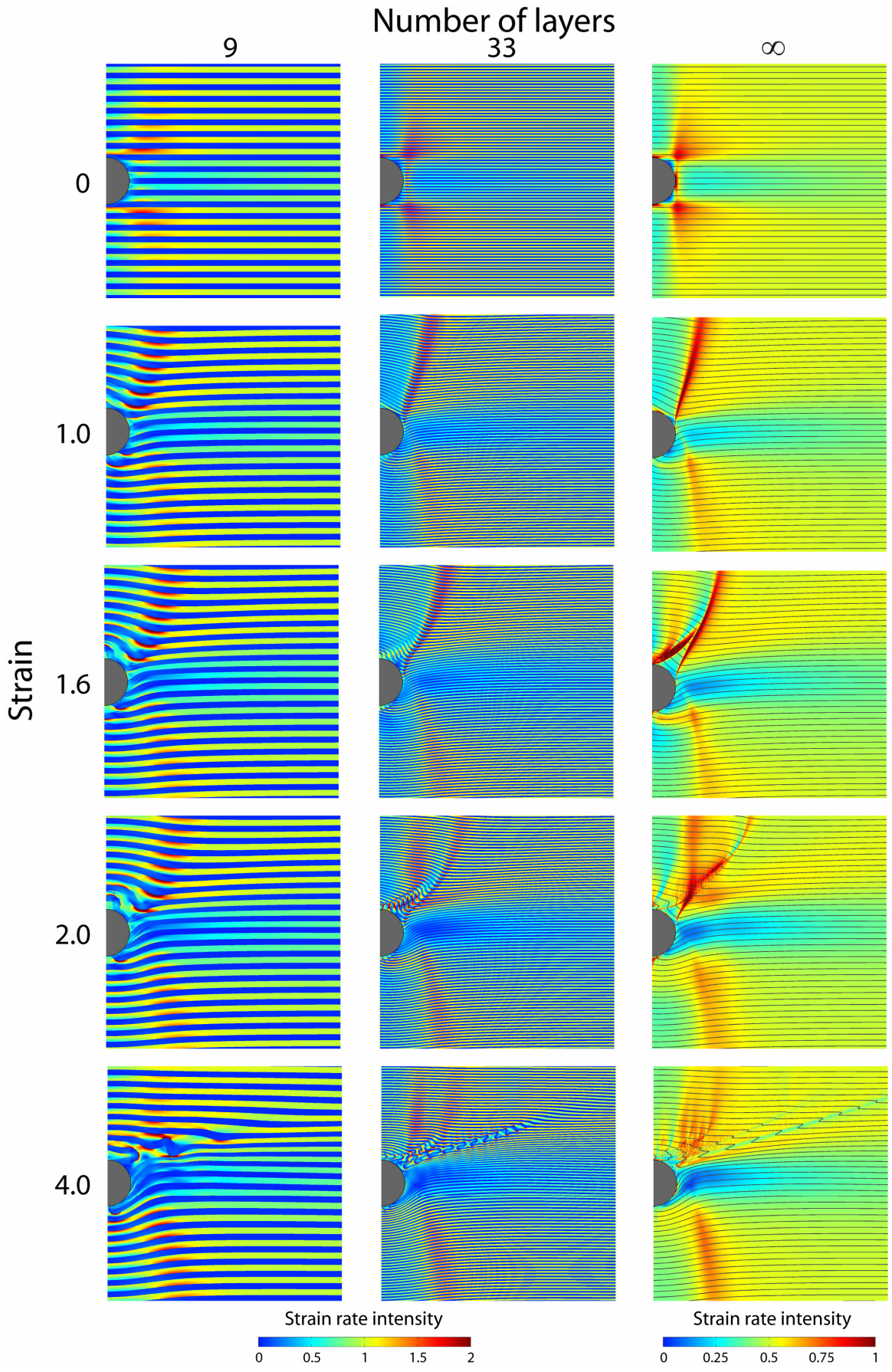
Finite strain models have been performed for the all instantaneous models that are presented in Figure 2 and Figure 5, i.e. for matrix material that is truly anisotropic or has 9, 17 and 33 competent across the inclusion and anisotropy factors 2, 10 and 100. The time integration of the anisotropy direction field is based on the director theory that has been described in the previous section. The structures developed after a simple shear of magnitude 5 are presented in Figure 7.

The results for an anisotropic host are shown in the bottom row. The pattern developed in the weakly anisotropic case (anisotropy factor 2) is expectedly similar to the isotropic case. A distinct band of tightly folded markers exist, which is predominately inclined at 20 degrees to the shear plane. In the vicinity of the inclusion markers and the deformation band are strongly wrapped by the inclusion rotation and mimic δ clast geometry (Passchier & Simpson 1986). At a distance of ca. seven times the inclusion radius only slight marker deflections are still present in the host. In the anisotropy factor 10 case, the total rotation of the inclusion as indicated by the wrapping of the cross-cutting markers is reduced. The structure becomes more angular and a second, weaker band of folds is present. The maximum fold amplitude is reduced but the perturbation penetrates deeper into the host. The orientation of the main deformation band is slightly steeper. For the anisotropy factor 100 even stronger inhibition of the inclusion rotation is observed. Fold amplitudes are markedly reduced and the previous deformation bands are hardily recognizable. In the sector between ca. 40 and 80 degrees to the shear plane marker deflections become ubiquitous and the localization is effectively pervasive.

Explicit layering of the host material essentially does not affect the total rotation of the inclusion if compared to the anisotropic model. For weak host anisotropy also the structural evolution is almost identical, because the layers behave as passive markers. However, in hosts with stronger anisotropy the presence of layering causes a reduction of the perturbation range and smoother fold hinges for thicker layers. In the extreme case of 9 layers and anisotropy factor 100 the structure formation is restricted to layers that cross cut the inclusion. Contrary, for the finest lamination and the anisotropy factor 10 the obtained pattern is very similar to the anisotropic host case, even including the formation of the secondary band of localized deformation.

Figure 8

Strain rate intensity field after a simple shear of magnitude 0, 1, 1.6, 2 and 4. The results are shown for the anisotropic and layered host case with 9 and 33 layers across the inclusion. The anisotropy factor equals 10 and the viscosity ratio is set accordingly. In the case of the anisotropic host an arbitrarily spaced set of thin lines is used as passive markers. Note the difference in the colorbars between the two model categories.


Number of layers
9
33
∞
Strain
0
1.0
1.6
2.0
4.0
Strain rate intensity
0
0.5
1
1.5
2
Strain rate intensity
0
0.25
0.5
0.75
1

In Figure 8 the detailed structural evolution for the models with the anisotropy factor 10 is presented together with strain rate intensity fields. We analyze the anisotropic host case first.

- Shear Strain 1: The orthogonal symmetry present in the onset of deformation is visibly broken. At this stage the deformation is strongly localized into a distinct band inclined at a high angle to the shearing direction. This band coincides with the axis of the most prominent structure visible in the host and formed by the gentle flexure of the markers. The initially present horizontal pair of the strain rate maxima located in the vicinity of the inclusion has disappeared and the horizontal low strain rate band has strengthened.
- Shear Strain 1.6: A secondary band of localized deformation has developed and links the inclusion and the primary high strain rate band. The secondary band is contained within a well developed kink band that together with deflections related to the primary band encloses a structure similar to a box fold.
- Shear Strain 2: The high strain rate bands have merged into a more diffuse zone that is limited in outreach and inclined at predominantly at 45 degrees to the shearing direction with a tendency to steepen towards the tips. Two subsidiary high strain rate bands have propagated vertically and horizontally from the main band in the neighborhood of the inclusion, but these zones are diffuse and the flow is localized to a lesser degree. The initially horizontal band of low strain rate is now strongly enhanced. The structure is dominated by the kink band that has propagated and broadened partly in expense of the box fold interior.
- Shear Strain 4: The development of numerous diffuse bands of elevated shearing predominates. The primary kink band has been sheared and reoriented towards the shear plane and is associated with a low strain rate. However, the kink band has not actively propagated much further into the matrix. A secondary kink band of smaller range and amplitude has developed behind the primary kink band.

The overall patterns of structure evolution and strain rate intensity are retained in the case of a layered host. The discrepancies grow with respect to the anisotropic case with increasing strain and thickness of the individual layers. The magnitude of the strain rate intensity in the incompetent layers is approximately twice higher than in the anisotropic case and virtually vanishes in the competent layers. The case of 33 layers across the inclusion mimics even the development of a secondary kink band through isoclinal bending of the competent layers. These fold stacks are in the subsequent shearing stages pulled apart (in the direction of the axial trace) and

less inclined segments escape the tight folding stage. As for the anisotropic matrix no active propagation of the deformation bands is visible and even unfolding can be observed. However, the presence of the strong layers causes the localized deformation to become less pronounced leading to smaller deflections within the folded bands. This effectively results in a shorter perturbation range in the structure with increasing the layer thickness, which can be observed in the first column of Figure 8.

For completeness, we performed experiments where the analytical flow field for a homogenously anisotropic host is persistently reused to integrate the structural evolution in the host. The results show that for large shear strains virtually no difference between the isotropic and anisotropic model develops, which substantiates the approach to analyze the finite strain evolution with FEM models.

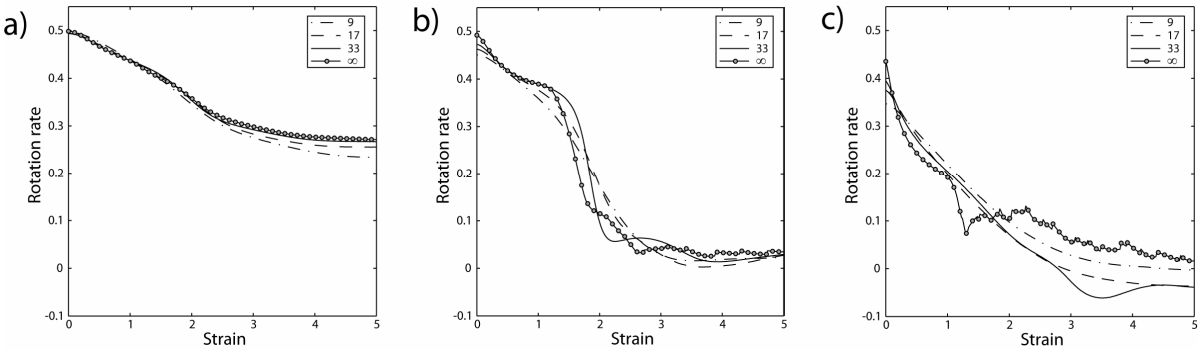


Figure 9

Rotation rate of the inclusion for the anisotropic and layered host cases. The results are presented for (effective) anisotropy factors of 2 (a), 10 (b) and 100 (c).

We have already noticed that the total rotation of the inclusion is significantly affected by the anisotropy. It is not surprising then that the development of the matrix structure as a function of accumulated shear has a strong influence on the inclusion rotation rate, Figure 9. a) to c) correspond to increasing (effective) anisotropy factors. Anisotropic matrix results are compared with explicitly layered runs of varying number of layers. The rotation rate is characterized by a decreasing trend for the studied strain range in all analyzed cases. The deviation from the homogenous isotropic case increases with increasing anisotropy and already for a factor ten the inclusion effectively stagnates after a shear strain of magnitude 2. It is worth noticing that an even antithetic motion is developed in the explicitly layered case corresponding to the anisotropy factor 100. An impact of the finite thickness layering on the inclusion rotation rate is noticeable but not substantial. The characteristics observed in the "Instantaneous flow pattern"

section such as a slower rotation for the coarsely layered host in the strongly anisotropic case are not preserved during finite strain simulations.

The discrepancy between the curves obtained for an anisotropic host and explicitly layered ones in the strongest anisotropy case may be partly explained by numerical problems. The anisotropy structure in the host becomes increasingly complex with strain and is difficult to resolve properly during remeshing stages (occasional sharp changes visible on the rotation rate curve). We suspect that main cause is the strong (internal) instability that lacks a length scale and is therefore mesh-size dependent.

The rotation rate curves have been integrated in order to inspect the overall rotation of an inclusion at the end of the simulations. In an isotropic host the expected rotation of the inclusion would be 145° after a shear strain of magnitude 5; the results for the anisotropic host are 100°, 50° and 20° for the studied anisotropy factors of 2, 10 and 100, respectively.

Discussion

Structure development

We have presented the analytical solution that describes the flow field around a rigid circular inclusion embedded in an anisotropic host. The analytical results show that with increasing anisotropy factor the flow localizes into narrowing bands that propagate towards the anisotropy directions. Performing finite strain simulations up to a simple shear of magnitude 5 by persistently reusing this solution, we have found that the structural development in the host is virtually insensitive to the anisotropy strength. Yet, the results of our FEM simulations that allow for the heterogeneous anisotropy reorientation during the deformation show significant modifications of the host structure due to the anisotropy. The major differences introduced by the anisotropy include the limited rotation of the inclusion and effectively suppressed wrapping of the markers adjacent to it, and sharpening of the marker deflections in the major deformation band.

During the structural evolution we have noticed the appearance of the subsidiary bands of localized shearing and the formation and propagation of related kink bands. The kink bands undergo an initial active growth stage that is followed by band stiffening and transposition towards the shearing direction. The fact that the propagation is inhibited at the later stages of the band evolution raises the general question of how far structural perturbations can propagate orthogonal to the shear plane in anisotropic rocks. One could envision a self-driven chain of perturbation–advection events that can propagate endlessly. However, this mechanism cannot be observed in our simulations.

Another effect is that stronger anisotropy leads to a reduction of fold amplitudes in behalf of a more pervasive development of internal instabilities. The latter requires high enough anisotropy factor and favorable orientation of the anisotropy trace. Again, this is only the case in the vicinity of the inclusion but not elsewhere.

Explicit layering in the host remarkably reduces the perturbation range, which is caused by the bending stiffness of the stronger layers. However, at no stage the bending of layers seems to be one of active folding (cf. Biot 1961) and we may speculate that either the anisotropy is not strong enough (anisotropy factor 10) or the number of explicit layers is not high enough (anisotropy factor 100, 33 layers). In the coarsest models the perturbation range is reduced to a degree that only a few open folds are formed in the inclusion vicinity. Folds like the one present in Figure 7 for the anisotropy factor 10 and 9 layers are likely to be transported intact away from the inclusion along the weak layers; avoiding isoclinal development. For even larger strains these folds may appear rootless.

The derived analytical solution is strictly only valid for unperturbed host anisotropy. Our experiments show that in the case of simple shear reasonable approximations to the true structural evolution are limited to shear strains of magnitude <0.5.

Rotation rate

In the first part of our study we have presented a closed-form analytical solution describing the flow field around a rigid circular inclusion embedded in a homogeneously anisotropic host subject to a simple shear flow in the far field. The instantaneous motion of such an object undisputedly mimics spinning in the far field. A more general result has previously been derived (R.C. Fletcher, submitted) showing that the rotation rate of a rigid elliptical inclusion is not affected by the host anisotropy and follows the formula derived by Jeffery (1922). The corresponding FEM models show some deviations from the predicted analytical rotation rates. Here, the rigid circular inclusion rotates slower than expected and this effect is strengthened by an increase in anisotropy magnitude or inclusion radius as compared to the model size. These variations are caused by the relative proximity of the boundaries and by the different treatment of lateral (periodic) and top and bottom (impermeable) walls in our numerical setup.

For an explicitly layered host the rotation rate of an inclusion is additionally affected by layer thickness with respect to inclusion size. It also depends on the embedding type, i.e. if the bottom and top-most points of the inclusion are embedded in weak or strong layers. The extreme cases are given by a strongly embedded inclusion in a weakly anisotropic host when the rota-

tion rate slightly exceeds the far field value and a weakly embedded inclusion in a strongly anisotropic host with the rotation rate assuming ca. thirty percent of the expected value.

Marques et al. (2005) have studied the influence confinement on the rotation of rigid elliptical clasts in shear zones. They demonstrated that clast stabilization and shape preferred orientation result from confinement. Our results concerning the distances over which boundaries influence the behavior of inclusions in an anisotropic matrix and the large strain evolution suggest that the results obtained by Marques et al. (2005) are even more relevant in anisotropic materials, which large strain shear zone rocks often are. Hence, estimation of flow conditions (e.g., Marques et al. 2007) should consider possible confinement effects even if impermeable walls are not present in the immediate vicinity of clasts.

Boundary effects are responsible for the slight deviations from the equivalent isotropic matrix results at the onset of our simulations. However, the rotation rate evolution for the anisotropic matrix case shows marked differences to the isotropic one, Figure 9, which are attributed to structure development in the matrix and not to boundary effects. Already an anisotropy factor of 2 produces significantly lower rotation rates. With even stronger anisotropy the rotation essentially ceases once a shear strain of 2 is reached. While the actual number of layers across the inclusion matter for the initial stages of inclusion rotation the finite strain evolution is largely independent of this parameter. As long as the layering is approximately shear plane parallel the inclusion may be decoupled from the matrix above and below by the localized shearing in the adjacent weak layers. Accumulated deflection due to inclusion rotation deactivates this effect.

The question arises how inclusions with different aspect ratios or finite viscosity ratio behave in an anisotropic matrix. The motion of rigid elliptical inclusion is likely to exhibit a dependence on its initial orientation. Preliminary results show that an elliptical inclusion initially aligned with an anisotropy trace exhibits antithetic rotation under simple shear. Moreover, irrespective of initial orientation moderately elliptical inclusions stabilize at a shallow antithetic angle to the shear plane. For the initial motion of a deformable elliptical inclusion Eshelby's conjecture of an uniform straining within inclusion still holds in an orthotropic surrounding medium (Willis 1964). However, it seems unlikely that Eshelby's conjecture remains applicable during the progressive deformation where the anisotropy in the host is deflected. However, it has been reported in the literature that in the extreme case of a flanking structure, modeled as an elongated weak elliptical inclusion, embedded in an anisotropic medium the motion of the inclusion remains passive as in an isotropic host case (Kocher & Mancktelow 2006).

Instability and length scale

In this study we have used an anisotropy model that is not equipped with a finite length scale and corresponds to the case of an infinitely thin layering present in a material. To investigate the evolution of the system with a finite layer thickness, we have explicitly resolved individual layers. This approach is advantageous as it is free of any built-in uncertainties inherent to all up-scaling techniques and yields relevant results to cases where relatively coarse layering/banding is present. The disadvantage is, however, the need for large numerical resolution. A characteristic length scale may be restored by introducing a bending stiffness into the calculations. Such a scheme, relying on a couple stress (micro-polar) formulation and incorporating a bending stiffness, has been proposed (Muhlhaus et al. 2002a). Besides the obvious rationale that a definite length scale of the layering exists for real systems, employing such a code circumvents the mesh sensitivity problem inherent to an anisotropic code lacking a length scale.

Yet, equipping an anisotropic material with a bending stiffness will only partly solve the problem of approximating a layered material. Figure 7 shows that the layers in the explicitly layered matrix change their thickness during the deformation. Furthermore, we have seen that it matters how exactly the inclusion is embedded in the layered material, i.e. top and bottom in strong or weak layers. These kinds of effects can only be studied by direct resolution of the layers; anisotropic upscaling will not be able to give corresponding insight. Nevertheless, some parameters such as the rotation rate as a function of strain are largely independent of the exact structure development in the matrix and may readily be studied with an upscaling approach. One may envision hybrid models where the focus of interest is resolved explicitly and the more distal parts are treated with an effective material model.

Conclusions

We have studied a rigid circular inclusion embedded in an anisotropic host and subject to far field simple shear. At the onset of the simulations the anisotropy in the host is assumed homogeneous and a relevant analytical solution has been presented. The system evolution has been computed with a finite element code and models with an anisotropic or explicit layered host have been employed. Studied aspects include the inclusion motion, flow field rearrangement and structural development in the host.

Our results show that matrix anisotropy has a first order effect on the motion of a rigid heterogeneity subject to shear and the development of structures around it. Already an anisotropy factor of 2 is sufficient to substantially decrease the total rotation of a rigid circular inclusion.

Further increase in the strength of anisotropy leads to an effective cease of rotation once a shear strain of magnitude 2 is reached. Cross-cutting markers are consequently markedly less deflected in the direct vicinity of the inclusion, giving the impression of smaller accumulated strain. Expectedly, box folds and kink bands develop around the inclusion. The range of the perturbation flow increases with increasing anisotropy but the amplitude is reduced. For strong anisotropy factors the distinct zones of localized deformation disappear in behalf of pervasive internal instability with vanishing amplitudes.

The increased perturbation flow range leads to a stronger sensitivity to the applied boundary conditions. Already in the initial stage of a homogenously anisotropic matrix with factor 100 the rotation rate of the inclusion is reduced by more than 10 percent even when the inclusion diameter is only 1/20 of the shear zone width. Explicit layering intensifies this effect.

The major effect of explicit layering on the structural development is the reduction of the perturbation range. In the coarse limit the formation of the band forming isoclinal recumbent folds is prevented and open folds result instead. These may in subsequent stages be advected away from the inclusion and appear as rootless structures. These and related effects such as competent layer thickening and deformation localization in weak layers cannot be predicted in the frame of the up-scaling anisotropic approach. However, in the limit of thin layering the effective anisotropy provides a robust representation of a layered medium and the discrepancies between the two approaches vanish with decreasing viscosity ratio of layer materials.

Field studies that look at the behavior of heterogeneities in anisotropic materials should consider the following. In anisotropic materials the effects of flow confinement and interaction are substantially stronger. Furthermore, inclusion rotation, foliation wrapping, and deflection amplitudes may yield severe strain underestimations when natural structures are interpreted without the effects of anisotropy taken into account.

Acknowledgments

This work was supported by a Center of Excellence grant from the Norwegian Research Council to PGP.

Appendix I: Analytical solution

We provide an outline of a solution strategy for a deformable elliptical inclusion embedded in a compressible anisotropic elastic host subject to a uniform straining in the far field. For simplicity, we consider only an isotropic inhomogeneity. For a detailed derivation readers are re-

ferred to the original work of Willis (1964). The solution builds on the Airy stress function and complex coordinates formulation of elasticity (e.g. Green & Zerna 1968), Cauchy's theory of residues and the eigenstrains theory and the Eshelby's method of an equivalent inclusion (e.g. Eshelby 1957, Mura 1987). Readers are encouraged to check the mentioned references for the in-depth presentations of the topics.

The Airy stress function methodology is a classical approach applicable to planar elastostatic problems. Upon introduction of the Airy potential Ψ for the Cartesian coordinates

$$\sigma_{xx} = \frac{\partial^2 \Psi}{\partial y^2}, \quad \sigma_{xy} = -\frac{\partial^2 \Psi}{\partial x \partial y}, \quad \sigma_{yy} = \frac{\partial^2 \Psi}{\partial x^2} \tag{16}$$

the mechanical equilibrium equations are identically satisfied for a vanishing body force field. The constitutive relationships for a transversally isotropic elastic medium is given by

$$\begin{aligned} \sigma_{xx} + \sigma_{yy} &= 2(\lambda + \mu_n)\left(D_{xx} + D_{yy}\right) \\ \sigma_{xx} - \sigma_{yy} &= 2\mu_n\left(D_{xx} - D_{yy}\right) \\ \sigma_{xy} &= 2\mu_s D_{xy} \end{aligned} \tag{17}$$

where λ is the first Lame parameter and μ_n and μ_s are normal and shear elastic moduli. By inverting eqn. (17) and substituting the result into the two-dimensional compatibility equation

$$\frac{\partial^2 D_{xx}}{\partial y^2} + \frac{\partial^2 D_{yy}}{\partial x^2} = 2\frac{\partial^2 D_{xy}}{\partial x \partial y} \tag{18}$$

the Airy stress function is shown to satisfy the following partial differential equation

$$\frac{\partial^4 \Psi}{\partial x^4} + 2\left(1 + 4\frac{\delta - 1}{\kappa + 1}\right)\frac{\partial^4 \Psi}{\partial x^2 \partial y^2} + \frac{\partial^4 \Psi}{\partial y^4} = 0 \tag{19}$$

where $\kappa = \left(\lambda + 3\mu_n\right)/\left(\lambda + \mu_n\right)$. In the Airy stress function approach the vector field problem is reduced to the scalar field one, but the order of the partial differential equation is raised. Eqn. (19) reduces to the biharmonic equation for the isotropic medium case.

Introducing the complex coordinates $X = x + iy$ eqn. (19) can be reformulated as

$$\frac{\partial^4 \Psi}{\partial X^4} - 2A\frac{\partial^4 \Psi}{\partial X^2 \partial \bar{X}^2} + \frac{\partial^4 \Psi}{\partial \bar{X}^4} = 0 \tag{20}$$

where $A = (\mu_n + \kappa\mu_s)/(\mu_n - \mu_s)$.

Firstly, we seek for the most general real solution of eqn. (20). This solution admits the following form that involves arbitrary functions $\Omega(X+\gamma\bar{X})$ and $\omega(X-\gamma\bar{X})$

$$\Psi=\Omega(X+\gamma\bar{X})+\bar{\Omega}(\bar{X}+\bar{\gamma}X)+\omega(X-\gamma\bar{X})+\bar{\omega}(\bar{X}-\bar{\gamma}X) \tag{21}$$

where γ is a root of the equation:

$$\gamma^4-2A\gamma^2+1=0 \tag{22}$$

The roots are distinct if normal and shear moduli are different, If γ is a root then its reciprocal and their negatives are roots as well.

For our particular case, the deviatoric stress components are given by

$$\sigma_{yy}-\sigma_{xx}+2i\sigma_{xy}=4\left(\Omega''(X+\gamma\bar{X})+\gamma^2\bar{\Omega}''(\bar{X}+\bar{\gamma}X)+\omega''(X-\gamma\bar{X})+\gamma^2\bar{\omega}''(\bar{X}-\bar{\gamma}X)\right) \tag{23}$$

and the displacements are integrated accordingly

$$\begin{aligned}2\left(v_x+iv_y\right)=&\left\{\Omega'(X+\gamma\bar{X})-\omega'(X-\gamma\bar{X})\right\}\left\{\left(\frac{1}{\mu_s}-\frac{1}{\mu_n}\right)-\gamma^2\left(\frac{1}{\mu_s}+\frac{1}{\mu_n}\right)\right\}\Big/\gamma+\\&\left\{\Omega'(X+\gamma\bar{X})+\omega'(X-\gamma\bar{X})\right\}\left\{\gamma^2\left(\frac{1}{\mu_s}-\frac{1}{\mu_n}\right)-\left(\frac{1}{\mu_s}+\frac{1}{\mu_n}\right)\right\}\end{aligned} \tag{24}$$

At this stage, we introduce an auxiliary problem of finding the stress distribution in a planar anisotropic medium with an elliptical domain subject to a uniform eigenstrain D^0. The eigenstrain is a generic name given to non-elastic contributions to a total strain

$$D^{el}=D-D^0 \tag{25}$$

Thermal strains provide an example of eigenstrains. The action of a uniform eigenstrain applied to any domain is equivalent to loading the system with the layer of body-force over the interface between the domain and the host. The second derivatives of Ω and ω that describe the stress response to a body-force $F=F_x+iF_y$ located at a point Z of an anisotropic plane are

$$\begin{aligned}\Omega''(X+\gamma\bar{X})&=\frac{\alpha F+\beta\bar{F}}{2\pi}\frac{1}{(X+\gamma\bar{X})-(Z+\gamma\bar{Z})}\\\omega''(X-\gamma\bar{X})&=\frac{-\alpha F+\beta\bar{F}}{2\pi}\frac{1}{(X-\gamma\bar{X})-(Z-\gamma\bar{Z})}\end{aligned} \tag{26}$$

where

$$\alpha = \frac{-\gamma\left\{\left(1-\gamma^2\right)\mu_n + \left(1+\gamma^2\right)\mu_s\right\}}{4(\mu_n-\mu_s)\left(1-\gamma^4\right)}, \quad \beta = \frac{\left\{\left(1-\gamma^2\right)\mu_n - \left(1+\gamma^2\right)\mu_s\right\}}{4(\mu_n-\mu_s)\left(1-\gamma^4\right)} \tag{27}$$

Now, it suffices to integrate the loads corresponding to the eigenstrain around a given contour

$$\Omega'' = AI_1 + BI_2 \tag{28}$$

where

$$I_1 = \oint \frac{dZ}{\left(X+\gamma\bar{X}\right)-\left(Z+\gamma\bar{Z}\right)}, \qquad I_2 = \oint \frac{d\bar{Z}}{\left(X+\gamma\bar{X}\right)-\left(Z+\gamma\bar{Z}\right)} \tag{29}$$

and

$$\begin{aligned} A &= \frac{1}{2\pi i}\left(\alpha\frac{\sigma_{xx}^0+\sigma_{yy}^0}{2} + \beta\frac{\sigma_{xx}^0-\sigma_{yy}^0}{2}\right) - \frac{1}{2\pi}\beta\sigma_{xy}^0 \\ B &= \frac{1}{2\pi i}\left(-\alpha\frac{\sigma_{xx}^0-\sigma_{yy}^0}{2} - \beta\frac{\sigma_{xx}^0+\sigma_{yy}^0}{2}\right) - \frac{1}{2\pi}\alpha\sigma_{xy}^0 \end{aligned} \tag{30}$$

The ω'' is obtained by replacing γ with $-\gamma$ in eqn. (28). In the general case of an elliptical inclusion, the integration involves the coordinate transformation. Upon the transformation the integrals (29) become

$$\begin{aligned} I_1 &= \frac{-1}{(a+b)+\gamma e^{-2i\theta}(a-b)}\oint \frac{(a+b)Z-(a-b)/Z}{Z^2-\left(X+\Gamma\bar{X}\right)Z+\Gamma}dZ \\ I_2 &= \frac{-e^{-2i\theta}}{(a+b)+\gamma e^{-2i\theta}(a-b)}\oint \frac{(a-b)Z-(a+b)/Z}{Z^2-\left(X+\Gamma\bar{X}\right)Z+\Gamma}dZ \end{aligned} \tag{31}$$

where a, b denote ellipse axes, θ inclination angle and

$$\Gamma = \frac{(a-b)+\gamma e^{-2i\theta}(a+b)}{(a+b)+\gamma e^{-2i\theta}(a-b)} \tag{32}$$

The integration (31) is taken around the unit circle. By employing the Cauchy's theory of residues, it can be shown that within the inclusion

$$I_1 = -2\pi i\frac{a+b}{(a+b)+\gamma e^{-2i\theta}(a-b)}, \quad I_2 = -2\pi i\frac{(a-b)e^{-2i\theta}}{(a+b)+\gamma e^{-2i\theta}(a-b)} \tag{33}$$

and for the points outside the inclusion

$$I_1 = 2\pi i \frac{1}{(a+b)+\gamma e^{-2i\theta}(a-b)} \frac{\left(X+\Gamma\bar{X}\right)-\left\{\left(X+\Gamma\bar{X}\right)^2-4\Gamma\right\}^{\frac{1}{2}}}{2\left\{\left(X+\Gamma\bar{X}\right)^2-4\Gamma\right\}^{\frac{1}{2}}}\left(a+b-\left.(a-b)\right/\Gamma\right)$$

$$I_2 = 2\pi i \frac{e^{-2i\theta}}{(a+b)+\gamma e^{-2i\theta}(a-b)} \frac{\left(X+\Gamma\bar{X}\right)-\left\{\left(X+\Gamma\bar{X}\right)^2-4\Gamma\right\}^{\frac{1}{2}}}{2\left\{\left(X+\Gamma\bar{X}\right)^2-4\Gamma\right\}^{\frac{1}{2}}}\left(a-b-\left.(a+b)\right/\Gamma\right) \tag{34}$$

Eqn. (33) predicts the uniform stress field within the inclusion. Thus, the Eshelby's conjecture holds for an elliptical inclusion also in the anisotropic case. Plugging in (33) into (28) , repeating it for ω" and using (23), the relation between the constrained strain and the prescribed eigenstrain is obtained.

The method of an equivalent inclusion links the auxiliary and the original inhomogeneity problems. An elliptical inclusion of elastic parameters λ^{incl}, μ^{incl} different from those of the matrix is subject to the eigenstrain D^* and we seek the constrained strain D^C. Finding the eigenstrain D^0 that produces the similar constrained strain in an 'equivalent' inclusion (having host elastic properties) and requiring the same stresses in both cases reduces the problem of an inhomogeneity to the auxiliary problem. The condition is given by

$$\begin{aligned} 2\mu_n\left(D_{ij}^C - D_{ij}^0\right)+\lambda\left(D^C - D^0\right) &= 2\mu^{incl}\left(D_{ij}^C - D_{ij}^*\right)+\lambda\left(D^C - D^*\right) \quad i=j \\ 2\mu_s\left(D_{ij}^C - D_{ij}^0\right) &= 2\mu^{incl}\left(D_{ij}^C - D_{ij}^*\right) \quad i\neq j \end{aligned} \tag{35}$$

where symbols without indices correspond to the trace of a tensor. Eliminating D^C, we obtain the relation between D^0 and D^*. Thus, the inhomogeneity problem is indeed reduced to the auxiliary one of the eigenstrain D^0 applied to an elliptical region of an anisotropic space.

In the instance of uniform loads σ^∞ prescribed at infinity, perturbation fields around the inclusion may be considered as originating from a particular eigenstrain applied to it. This eigenstrain corresponds to the misfit between the far field strains D^∞ and strains that inclusion would acquire if subjected directly to the far field stress σ^∞. The final relations become now

$$\begin{aligned} 2\mu_n\left(D_{ij}^C - D_{ij}^0\right)+\lambda\left(D^C - D^0\right) &= 2\mu^{incl}\left(D_{ij}^C + D_{ij}^\infty\right)+\lambda\left(D^C + D^\infty\right)-\sigma_{ij}^\infty \quad i=j \\ 2\mu_s\left(D_{ij}^C - D_{ij}^0\right) &= 2\mu^{incl}\left(D_{ij}^C + D_{ij}^\infty\right)-\sigma_{ij}^\infty \quad i\neq j \end{aligned} \tag{36}$$

In the rigid inclusion case $\lambda^{incl} = \mu^{incl} \to \infty$ we arrive at the expected result of perturbation strain compensating the far field one i.e.: $D_{ij}^{C} = -D_{ij}^{\infty}$. Under the assumptions of a circular inclusion and an incompressible viscous flow case (velocity field is identified with displacements and $\kappa \to 1$), we arrive at the following eigenstrain component D_{xy}^{0} (in our simple shear setup diagonal entries of the far field strain rate are set to zero)

$$D_{xy}^{0} = -\frac{1+\sqrt{\delta}}{\sqrt{\delta}} D_{xy}^{\infty} \tag{37}$$

The external velocity and stress (strain rate) fields can be now obtained by plugging in the calculated eigenstrain into eqn. (34), (28), repeating it for ω" and using formulas (23) and …

Appendix II: Finite Element Model

A large number of required timesteps and high discretization level exceeding million degrees of freedom present in our models necessitate the use of an efficient implementation of FEM. In this study we have utilized our unstructured mesh FEM code MILAMIN implemented entirely in MATLAB. We have employed the mixed formulation of FEM that allows us to tackle incompressible flow problems. In this formulation pressure field is approximated independently from the velocity field and corresponding degrees of freedom are explicitly present in the model. A particular element type that we have utilized is the seven-node Crouzeix-Raviart triangle with the pressure field interpolated by a discontinuous linear function. The incompressible constraint is enforced through the penalty approach and performing Hestenes-Powell iterations allows us to use a moderate penalty factor. By eliminating the pressure degrees of freedom on the element level, we gain the possibility to operate on a positive-definite global matrix. The direct solvers are tailored for such systems and Cholesky factorizations are well known for their high performance. Details concerning implementation and applied optimizations have been describe in our other paper (Dabrowski et al. in press).

A reliable mesh generator is required to create high quality unstructured meshes fitting internal and external interfaces present in computational models. For this purpose, we have chosen TRIANGLE software developed by Shewchuk (2007). We find this software fast, robust and flexible. The file input/output facilitates the coupling with our MATLAB code. Furthermore, a possibility of an element area control existing in TRIANGLE allows us to utilize better the power of an unstructured mesh approach employed in this study. Specifically, finite element

meshes have been strongly refined around the inclusion to facilitate resolution of flow patterns around the inclusion in detail.

Technical issues that have arisen during finite strain runs need some attention. By employing an unstructured mesh FEM, we have a possibility to update the mesh geometry according to the computed velocity field and continue with calculations. This Lagrangian approach has been used in both categories of the studied models. Advantageously, the advection of the material properties defined in integrations points such as anisotropy trace inclination is automatically taken care of. But a computational mesh must be occasionally regenerated to improve its quality that deteriorates after a number of updates. For the layered host case, we only need to track internal and external interfaces. However, inclinations of the anisotropy trace need to be interpolated to new integration points during this stage for the other category of our models. We have implemented an interpolation scheme in a spirit of krigging methods. The utilized scheme relies on identifying a prescribed amount of neighboring integration points (belonging to an old mesh), forming a radial basis functions and computing the corresponding weights in a way that enforces a collocation. Next, the radial basis functions are evaluated for the point of interest (i.e. new integration point) and together with the computed weights and inclinations defined in old integration points are used to compute a new value of the anisotropy orientation. During computations some precautions need to be taken to avoid the ambiguity around [0 180] transition.

References

Arbaret, L., Mancktelow, N. S. & Burg, J. P. 2001. Effect of shape and orientation on rigid particle rotation and matrix deformation in simple shear flow. *Journal of Structural Geology* **23**(1), 113-125.

Bayly, M. B. 1970. Viscosity and Anisotropy Estimates from Measurements on Chevron Folds. *Tectonophysics* **9**(5), 459-&.

Bilby, B. A. & Kolbuszewski, M. L. 1977. Finite Deformation of an Inhomogeneity in 2-Dimensional Slow Viscous Incompressible-Flow. *Proceedings of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences* **355**(1682), 335-353.

Biot, M. A. 1961. Theory of Folding of Stratified Viscoelastic Media and Its Implications in Tectonics and Orogenesis. *Geological Society of America Bulletin* **72**(11), 1595-1620.

Biot, M. A. 1965. *Mechanics of incremental deformations : theory of elasticity and viscoelasticity of initially stressed solids and fluids, including thermodynamic foundations and applications to finite strain*. Wiley, New York.

Bons, P. D., Barr, T. D. & tenBrink, C. E. 1997. The development of delta-clasts in non-linear viscous materials: A numerical approach. *Tectonophysics* **270**(1-2), 29-41.

Cobbold, P. R., Cosgrove, J. W. & Summers, J. M. 1971. Development of Internal Structures in Deformed Anisotropic Rocks. *Tectonophysics* **12**(1), 23-&.

Dabrowski, M., Krotkiewski, M. & Schmid, D. W. in press. MILAMIN: MATLAB-based FEM solver for large problems. *Geochemistry, Geophysics, and Geosystems* **in press**.

Eshelby, J. D. 1957. The determination of the elastic field of an ellipsoidal inclusion, and related problems. *Proceedings of the Royal Society of London Series A- Mathematical and Physical Sciences* **241**(1226), 376-396.

Fletcher, R. C. 2004. Anisotropic viscosity of a dispersion of aligned elliptical cylindrical clasts in viscous matrix. *Journal of Structural Geology* **26**(11), 1977-1987.

Ghosh, S. K. & Ramberg, H. 1976. Reorientation of Inclusions by Combination of Pure Shear and Simple Shear. *Tectonophysics* **34**(1-2), 1-70.

Green, A. E. & Zerna, W. 1968. *Theoretical elasticity*. Clarendon Press, Oxford.

Hill, R. 1952. The Elastic Behaviour of a Crystalline Aggregate. *Proceedings of the Physical Society of London Section A* **65**(389), 349-355.

Jeffery, G. B. 1922. The motion of ellipsoidal particles immersed in a viscous fluid. *Proceedings of the Royal Society of London* **A102**, 161-179.

Jezek, J., Schulmann, K. & Segeth, K. 1996. Fabric evolution of rigid inclusions during mixed coaxial and simple shear flows. *Tectonophysics* **257**(2-4), 203-221.

Kocher, T. & Mancktelow, N. S. 2005. Dynamic reverse modelling of flanking structures: a source of quantitative kinematic information. *Journal of Structural Geology* **27**(8), 1346-1354.

Kocher, T. & Mancktelow, N. S. 2006. Flanking structure development in anisotropic viscous rock. *Journal of Structural Geology* **28**(7), 1139-1145.

Kocher, T., Schmalholz, S. M. & Mancktelow, N. S. 2006. Impact of mechanical anisotropy and power-law rheology on single layer folding. *Tectonophysics* **421**(1-2), 71-87.

Kocks, U. F., Tomé, C. N. & Wenk, H.-R. 2000. *Texture and anisotropy : preferred orientations in polycrystals and their effect on materials properties*. Cambridge University Press, Cambridge.

Mandal, N., Samanta, S. K. & Chakraborty, C. 2005. Numerical models of flow patterns around a rigid inclusion in a viscous matrix undergoing simple shear: implications of model parameters and boundary conditions. *Journal of Structural Geology* **27**(9), 1599-1609.

Marques, F. O., Schmid, D. W. & Andersen, T. B. 2007. Applications of inclusion behaviour models to a major shear zone system: The Nordfjord-Sogn Detachment Zone in western Norway. *Journal of Structural Geology* **29**(10), 1622-1631.

Marques, F. O., Taborda, R. M. & Antunes, J. V. 2005. 2D rotation of rigid inclusions in confined bulk simple shear flow: a numerical study. *Journal of Structural Geology* **27**(12), 2171-2180.

Masuda, T. & Mizuno, N. 1996. Computer modelling of mantled porphyroclasts in Newtonian and non-Newtonian simple shear viscous flows. *Journal of Structural Geology* **18**(12), 1487-1491.

Muhlhaus, H. B., Dufour, F., Moresi, L. & Hobbs, B. 2002a. A director theory for visco-elastic folding instabilities in multilayered rock. *International Journal Of Solids And Structures* **39**(13-14), 3675-3691.

Muhlhaus, H. B., Moresi, L., Hobbs, B. & Dufour, F. 2002b. Large amplitude folding in finely layered viscoelastic rock structures. *Pure And Applied Geophysics* **159**(10), 2311-2333.

Mulchrone, K. F. 2007. An analytical solution in 2D for the motion of rigid elliptical particles with a slipping interface under a general deformation. *Journal of Structural Geology* **29**(6), 950-960.

Mura, T. 1987. *Micromechanics of defects in solids*. Kluwer Academic Publ., Dordrecht.

Passchier, C. W. & Simpson, C. 1986. Porphyroclast Systems as Kinematic Indicators. *Journal of Structural Geology* **8**(8), 831-843.

Rosas, F., Marques, F. O., Luz, A. & Coelho, S. 2002. Sheath folds formed by drag induced by rotation of rigid inclusions in viscous simple shear flow: nature and experiment. *Journal of Structural Geology* **24**(1), 45-55.

Schmid, D. W. & Podladchikov, Y. Y. 2003. Analytical solutions for deformable elliptical inclusions in general shear. *Geophysical Journal International* **155**(1), 269-288.

Schmid, D. W. & Podladchikov, Y. Y. 2004. Are isolated stable rigid clasts in shear zones equivalent to voids? *Tectonophysics* **384**(1-4), 233-242.

Schmid, D. W. & Podladchikov, Y. Y. 2005. Mantled porphyroclast gauges. *Journal of Structural Geology* **27**(3), 571-585.

Shewchuk , J. 2007. Triangle.

Treagus, S. H. 2003. Viscous anisotropy of two-phase composites, and applications to rocks and structures. *Tectonophysics* **372**(3-4), 121-133.

Willis, J. R. 1964. Anisotropic Elastic Inclusion Problems. *Quarterly Journal of Mechanics and Applied Mathematics* **17**(2), 157-&.

Paper 3: Mechanical Anisotropy Development of a Two-Phase Composite Subject to Large Deformation

To be submitted by M. Dabrowski, D.W. Schmid and Yu.Yu. Podladchikov

Abstract

Evolution of overall mechanical properties has been demonstrated in large strain deformation experiments. Strain softening is frequently employed in geodynamic simulations. In this paper, we quantify the structural and mechanical evolution of a two-phase composite rock subject to pure and simple shear. An inclusion-host type of geometry is assumed, we focus on the weak inclusion scenario and both materials obey a linear viscous behavior. Finite deformation leads to a shape preferred orientation development that results in an overall mechanical anisotropy. We derive the shape evolution model based on an analytical solution for an isolated elliptical inclusion embedded in an anisotropic host and subject to a uniform far field load. The presence of a strong anisotropy in the host leads to an enhanced inclusion stretching. A differential effective medium type of scheme predicting an overall anisotropic viscosity of a composite consisting of aligned elliptical inclusions is proposed and validated by finite element modeling. A comparison with an existing self-consistent averaging scheme is given and the new scheme is shown to provide an improved estimate of the effective normal and shear viscosity for high inclusion concentrations. The two models are combined into a final set of equations describing evolution of a two-phase rock under a shear. Hardening is predicted in pure shear. In simple shear, the hardening phase is followed by a pronounced softening after a shear strain of one, irrespective of inclusion concentration. Numerical simulations resolving evolution of inclusion-host systems under pure and simple shear demonstrate the high accuracy of our model prediction. The shape evolution model provides a sufficient approximation to the shape preferred orientation developing in an aggregate of interacting inclusions. Both in pure and simple shear, deformation localizes into conjugate trails of inclusions leading to formation of complex sigmoidal inclusion shapes. The strain weakening due to the structural development in a two-phase material obeying a linear viscous rheology is insufficient to explain alone strain localization in rocks that are heterogeneously populated with inclusions.

Introduction

Material heterogeneity and deformation are manifested and interlinked on virtually all scales in geology. Structures present on a grain, rock or continent scale evolve during tectonic episodes and in turn deformation paths are influenced by the structural development. For a given scale in the hierarchy, models are sought that provide an overall description of such mechanical and structural coupling and alleviate the need for direct resolution of complex underlying processes. Constitutive laws provide a particularly useful example of the up-scaling technique, where the atomistic approach is replaced by the continuum mechanics framework. The rock sample scale are particular in this hierarchy of scales. Processes operating at this level become complex due to the presence of grain boundaries and multi-phase components. The steady-state regime is usually only achieved after a long transient stage and is not expected to provide a sufficient description framework alone. The overall characteristics of the transient regime are of first-order importance with regard to large scale geodynamic processes. However, material models employed in the global tectonic simulations are typically based on phenomenological strength-strain laws. Quantifying the mechanical evolution during the transient stage and relating it to intrinsic system properties provides a clear improvement of the large-scale models.

In a plethora of geological structures originating during deformation episodes, a shape (SPO) or lattice (LPO) preferred orientation of rock constituents are of primary importance. The bulk of crustal materials shows penetrative compositional layering and crystal lattices of rock-forming minerals like quartz or feldspars are often clustered in sheared rocks. The anisotropy of the seismic wave velocity observed in subcrustal rocks can be attributed to the alignment of olivine grains emerging in a mantle flow (Dawson & Wenk 2000, McNamara et al. 2002). Strong compositional layering has been described in peridodite bodies (Spengler et al. 2006) and numerical models predict the formation of layered structures for the lower parts of the mantle (Madi et al. 2005).

In this study we focus on mechanical effects of the phase geometry rearrangement during the progressive straining of a two-phase composite. It has been recognized that the overall mechanical response of a heterogeneous rock may become anisotropic due to the SPO development. In the limiting case, a laminate of equal phase abundances exhibits a maximal degree of the anisotropy, where shear and normal viscosities assume values corresponding to the lower (Reuss) and upper (Voigt) theoretical bounds (Biot 1965). This material model is not suited to study the transient stage of the anisotropy evolution during the SPO build up and other analytical estimates of the overall anisotropy incorporating a finite magnitude of SPO have been de-

veloped (Fletcher 2004, Treagus 2003). A stiffening in a pure shear has been suggested for a two-phase particulate and an oscillatory evolution of the overall anisotropy has been anticipated for deformable competent inclusions in a simple shear flow (Treagus & Treagus 2002). Quantification of the mechanical response evolution of a two-phase composite subject to different deformation paths is a primary goal of this study.

Structure development in a composite rock provides an interesting problem on its own. The impact of the developing mechanical anisotropy and grain interaction on the phase geometry evolution remains poorly constrained. Strain localization, instability development and inclusion-host topology reversals are particularly important regarding the effective mechanical properties of a composite medium. Additionally, structures developed in composite rocks may be employed for strain magnitude estimates.

An important aspect of SPO development is related to strain localization on a larger scale in a heterogeneous rock mass. Formation of ductile shear zones is a long-standing topic in the geosciences (e.g. Bruhn & Burlini 2005, Regenauer-Lieb & Yuen 2003). Advocated mechanisms promoting strain localization in a ductile regime apart from LPO or SPO development include: shear heating, grain-size reduction, metamorphic reactions or partial melting. Whereas different observation and models substantiate particular mechanisms, strain localization in rock materials is likely to emerge from a synergy of several processes. Tackling this complex phenomenon requires the quantification and assessment of individual mechanisms in the first place. Here, we study the localization potential of the developing SPO and related anisotropy.

Isolated inclusion in an anisotropic host subject to pure and simple shear

In this section, we investigate the instantaneous response of an elliptical inclusion immersed in an anisotropic host and subject to a pure and simple shear using an analytical solution. The rates of ellipticity and inclination change are integrated to obtain finite strain evolution paths for an isolated inclusion under different load conditions. Inclusion and host are modeled as incompressible linear viscous materials.

The ultimate goal of this paper is to derive a model of a composite rock consisting of aligned inclusions, where the anisotropy is entirely determined by the inclusion orientation and aspect ratio. However, at this point we assume that the anisotropy is independent of the inclusion ellipticity to facilitate the study of its impact on the inclusion motion. Alternatively, an intrinsic anisotropy may be present in the host, but then the anisotropy reorientation rate is not necessarily coincident with that of an inclusion.

Using a Cartesian coordinate system aligned with the anisotropy direction, the constitutive relationship for an incompressible anisotropic medium reads

$$\tau_{xx} = 2\mu_n \dot{\varepsilon}_{xx}, \quad \tau_{yy} = 2\mu_n \dot{\varepsilon}_{yy}, \quad \tau_{xy} = 2\mu_s \dot{\varepsilon}_{xy} \tag{1}$$

where τ denotes the deviatoric stress and μ_n and μ_s are normal and shear components of viscosity. The strain rate tensor $\dot{\varepsilon}$ is defined by the usual kinemetic relations.

$$\dot{\varepsilon}_{xx} = \frac{\partial v_x}{\partial x}, \quad \dot{\varepsilon}_{yy} = \frac{\partial v_y}{\partial y}, \quad \dot{\varepsilon}_{xy} = \frac{1}{2}\left(\frac{\partial v_x}{\partial y} + \frac{\partial v_y}{\partial x} \right) \tag{2}$$

where v is the velocity vector field. The incompressibility condition complemented by the plane strain assumption yield the following constraints on the strain rate components

$$\dot{\varepsilon}_{xx} + \dot{\varepsilon}_{yy} = 0, \quad \dot{\varepsilon}_{xz} = \dot{\varepsilon}_{yz} = \dot{\varepsilon}_{zz} = 0 \tag{3}$$

The elastic problem of an elliptical inclusion embedded in an anisotropic host of and subject to a uniform far field load admits an analytical solution (Willis 1964). This solution predicts a constant strain rate within the inclusion. In our study, it suffices to discuss the incompressible limit, and an isotropic elliptical inclusion embedded coaxially with the host anisotropy directions. The elastic and viscous solutions coincide in this case according to the correspondence principle. The inclusion strain rate components are (e.g., Fletcher submitted to JSG)

$$\dot{\varepsilon}_{xx}^{incl} = \frac{\sigma^2 + 2\sqrt{\delta}\sigma + 1}{\sigma^2 + 2R_n\sqrt{\delta}\sigma + 1} \dot{\varepsilon}_{xx}^{\infty} \tag{4}$$

$$\dot{\varepsilon}_{xy}^{incl} = \frac{\sigma^2 + 2\sqrt{\delta}\sigma + 1}{R_n\delta\sigma^2 + 2\sqrt{\delta}\sigma + R_n\delta} \dot{\varepsilon}_{xy}^{\infty} \tag{5}$$

and the vorticity magnitude is given by

$$\omega^{incl} = \omega^{\infty} + \frac{\left(R_n\delta - 1\right)\left(\sigma^2 - 1\right)}{R_n\delta\sigma^2 + 2\sqrt{\delta}\sigma + R_n\delta} \dot{\varepsilon}_{xy} \tag{6}$$

where $\delta = \mu_n^{host} / \mu_s^{host}$ denotes the host anisotropy factor defined as a ratio of a normal and shear viscosity, $R_n = \mu^{incl} / \mu_n^{host}$ is the inclusion-host viscosity ratio, σ is the aspect ratio of the inclusion, and the upperscripts $incl$ and ∞ indicate inclusion and far field strain rate components, respectively. The inclusion and far field strain rate components are given in the inclusion axis reference system.

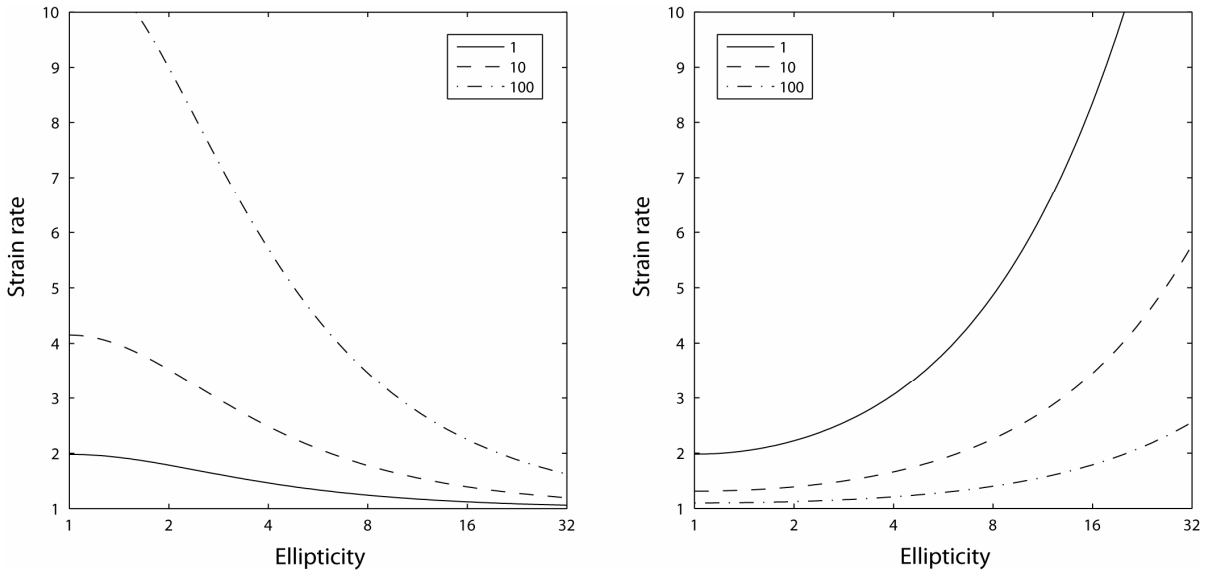


Figure 1

Inclusion strain rate components $\dot{\varepsilon}_{xx}^{incl}$ a) and $\dot{\varepsilon}_{xy}^{incl}$ (b) as a function of inclusion ellipticity and host anisotropy (individual curves). Inclusion viscosity is set to 0.01 of μ_n^{host} in (a) and μ_s^{host} in (b).

Inclusion strain rate components are shown in Figure 1 as a function of inclusion ellipticity and host anisotropy factor for an inclusion that is 100 times weaker than the host normal (a) and shear (b) viscosity. The normal strain rate component decreases, while the shear component increases with increasing the inclusion ellipticity. The presence of the host anisotropy counteracts the effect of ellipticity. Further increasing the inclusion-host viscosity contrast has no significant impact on the curves. The normal strain rate component in the anisotropic case yields values not admissible in the isotropic case.

We now focus on the finite deformation of an initially circular inhomogeneity embedded in an anisotropic host. An elliptical object subject to a uniform strain rate remains elliptical and this allows us to reuse the analytical solution during the integration. For a general homogeneous far field shear, the motion of an elliptical inclusion is governed by the following equations (e.g., Bilby & Kolbuszewski 1977)

$$\dot{\sigma} = 2\sigma\dot{\varepsilon}_{xx}^{incl} \tag{7}$$

$$\dot{\varphi} = \omega^{incl} + \frac{\sigma^2 + 1}{\sigma^2 - 1}\dot{\varepsilon}_{xy}^{incl} \tag{8}$$

where φ is the inclination of the inclusion. The far field strain rates need to be transformed to the current inclusion reference system, before the inclusion strain rate components and vorticity can be evaluated according to (4), (5) and (6).

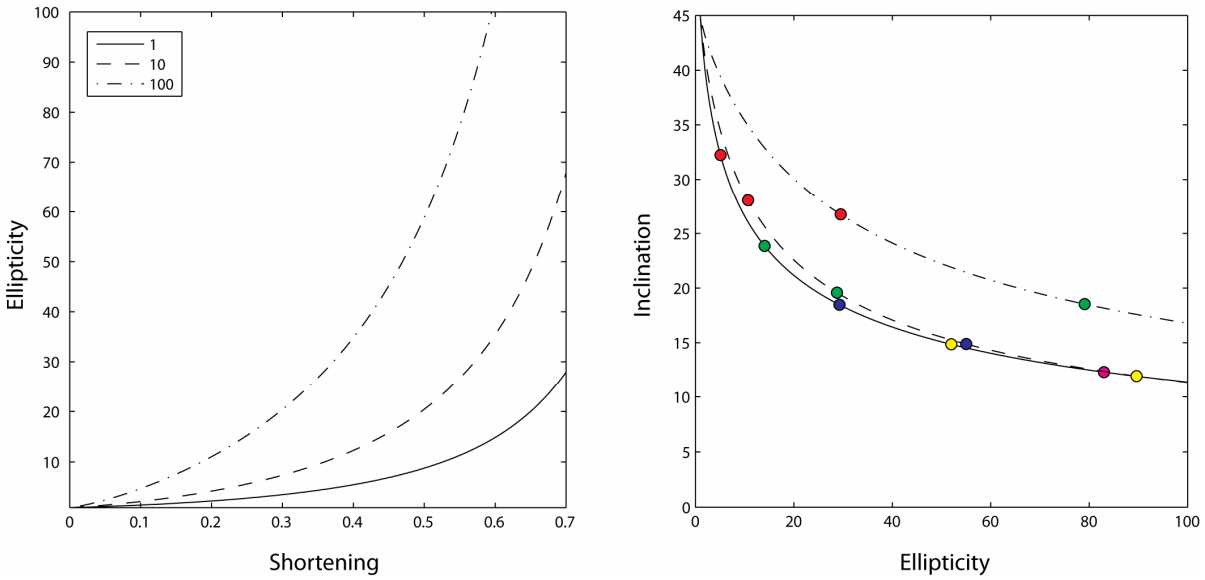


Figure 2

Shape evolution of an initially circular inclusion subject to pure (a) and simple (b) shear. The inclusion viscosity is 0.01 of the normal viscosity in the host. The results are obtained for the anisotropy factors 1, 10 and 100. In the simple shear case, the inclination-ellipticity paths are shown together with markers indicating the unit increment of a simple shear

The aspect ratio evolution of an initially circular inclusion subject to a pure shear of a rate $\dot{q}$ has been found upon integration to follow

$$\ln\sigma + \left(R_n - 1\right)\ln\left(1 + \frac{\left(\sigma - 1\right)}{\sigma\left(\sqrt{\delta} - \sqrt{\delta - 1}\right) + 1}\left(1 - \left(\sqrt{\delta} + \sqrt{\delta - 1}\right)\right)^{\frac{\sqrt{\delta}}{\sqrt{\delta - 1}}}\right) = 2q \tag{9}$$

The inclusion angle φ is constant in time and is determined by the extension direction. Note that the result presented by Bilby et al. (1975) can be reproduced by taking the isotropic limit of (9)

$$\ln\sigma + (R - 1)\frac{\sigma - 1}{\sigma + 1} = 2q \tag{10}$$

where $R = \mu^{incl} / \mu^{host}$ is the ratio of the isotropic viscosities. The pure shear shape evolution in the weak inclusion case is presented in Figure 2a for anisotropy factors 1, 10 and 100. The results show that due to the anisotropy of the host the inclusion stretching is enhanced as expected based on the instantaneous rates. After a shorting of 50% an initially circular inclusion assumes aspect ratio of 10 in the isotropic case. In the anisotropic cases of factors 10 and 100, the ellipticity yields 20 and 60, respectively.

In the simple shear case, a numerical integration of (7) and (8) has been performed. The far field strain rate components for a simple shear of a rate $\dot{s}$ resolved in the inclusion reference frame are given by

$$\dot{\varepsilon}_{xx}^{\infty} = \frac{\dot{s}}{2}\sin(2\varphi), \quad \dot{\varepsilon}_{xy}^{\infty} = \frac{\dot{s}}{2}\cos(2\varphi), \quad \omega^{\infty} = -\frac{\dot{s}}{2} \tag{11}$$

and consequently, eqn. (7) and (8) are coupled. The results are shown in Figure 2b. Again, the presence of the host anisotropy enhances inclusion stretching. In the initial deformation stages, the total rotation of the inclusion is also amplified if we compare the isotropic and anisotropy factor ten curves. Unlike in the aspect ratio case, further strengthening of the anisotropy has no significant impact on the inclusion rotation. The increased rotation rate of the inclusion in the anisotropic cases reduces time spent by the inclusion in the vicinity of the instantaneous stretching direction of the flow. Thus, the anisotropy impact on the inclusion stretching is not as pronounced as in the pure shear case. The inclination-ellipticity path for the anisotropy factor hundred is significantly different than the remaining ones and in comparison a higher ellipticity is reached for a given inclination.

Effective anisotropy of a composite consisting of aligned elliptical inclusions

Our goal in this paper is to develop a quantitative model of the structural and mechanical evolution for a two-phase composite initially consisting of circular weak inclusions. Above we have found that the shape evolution of an isolated inclusion is significantly affected by the host anisotropy both in pure and simple shear. However, these findings rely on a constant anisotropy in the host that is unrelated to SPO. In this section, we study the relation between the ellipticity σ and concentration f of the isotropic inclusions of viscosity μ^{incl} embedded in the isotropic host of viscosity μ^{host} and the overall normal μ_n^{eff} and shear μ_s^{eff} viscosity of the composite. We introduce an existing scheme predicting the overall behavior of such a medium, derive a new method and compare the two approaches to our finite element results

The effective viscosities are defined in terms of spatial averages of stress and strain rate components

$$\begin{aligned} \mu_n^{eff} &= \langle \tau_{xx} - \tau_{yy} \rangle / 2\langle \dot{\varepsilon}_{xx} - \dot{\varepsilon}_{yy} \rangle \\ \mu_s^{eff} &= \langle \tau_{xy} \rangle / 2\langle \dot{\varepsilon}_{xy} \rangle \end{aligned} \tag{12}$$

where the coordinate system is aligned with the inclusion axes and the spatial averaging denoted by $\langle \ldots \rangle$ is over a representative domain containing sufficient number of inclusions. The effective anisotropy factor δ^{eff} is defined as a ratio of the overall viscosities $\mu_n^{eff} / \mu_s^{eff}$. Prescribing uniform overall strain rate $\langle \dot{\varepsilon}_{ij} \rangle$ in the model, the kinematic boundary conditions become

$$v_i = \langle \dot{\varepsilon}_{ij} \rangle x_j \tag{13}$$

The relation between inclusion $\boldsymbol{\mu}^{incl}$ and host $\boldsymbol{\mu}^{host}$ viscosity, inclusion concentration f and the effective viscosity $\boldsymbol{\mu}^{eff}$ can be obtained now

$$\left(\mu_{ijkl}^{host} - \mu_{ijkl}^{eff} \right) \langle \dot{\varepsilon}_{kl} \rangle = f \left(\mu_{ijkl}^{host} - \mu_{ijkl}^{incl} \right) \langle \dot{\varepsilon}_{kl}^{incl} \rangle \tag{14}$$

This is an exact result showing that only the average inclusion strain rate $\langle \dot{\varepsilon}_{ij}^{incl} \rangle$ is required to evaluate the effective viscosity (e.g., Nemat-Nasser & Hori 1993).

Analytical models: SCA and DEM

A number of techniques predicting overall properties of heterogeneous materials has been derived (Christensen 2005, Nemat-Nasser & Hori 1993). In the paper, we focus on the self-consistent averaging (SCA) and the differential effective medium (DEM) approaches.

In the dilute limit, individual inclusions may be considered as effectively isolated objects. In this case the inclusion strain rate is given by taking the isotropic limit of the analytical solution (4) and (5). The normal and shear viscosity of a dilute composite consisting of aligned inclusions of ellipticity σ can be estimated by plugging

$$\frac{\dot{\varepsilon}_{xx}^{incl}}{\dot{\varepsilon}_{xx}^{\infty}} = \frac{1+\eta}{R+\eta}, \quad \frac{\dot{\varepsilon}_{xy}^{incl}}{\dot{\varepsilon}_{xy}^{\infty}} = \frac{1+\eta}{1+\eta R} \tag{15}$$

into eq. (14), where $\eta = 1/2\left(\sigma + \sigma^{-1}\right)$ is the shape factor.

The SCA has been recently applied to a composite consisting of aligned ellipses by Treagus (2003) and Fletcher (2004). In this approach, the viscosity ratios in (15) are evaluated with respect to the effective medium and not the original host material. In the Fletcher's work, inclusion strain rate has been evaluated according to eqn. (4) and (5) taking into account the overall anisotropy. The effective normal viscosity normalized by the viscosity of the host $\beta_n = \mu_n^{eff} / \mu^{host}$ and the effective anisotropy factor δ^{eff} are determined by solving a system of non-linear equations

$$\begin{aligned} &\eta\beta_n^2 + \left\{\sqrt{\delta^{eff}}\left[f + (1-f)R\right] - \eta\left[(1-f) + fR\right]\right\}\beta_n - \sqrt{\delta^{eff}}\,R = 0 \\ &\beta_n^2 + \left\{\eta\sqrt{\delta^{eff}}\left[f + (1-f)R\right] - \delta^{eff}\left[(1-f) + fR\right]\right\}\beta_n - \eta\left(\sqrt{\delta^{eff}}\right)^3 R = 0 \end{aligned} \tag{16}$$

The system is solved numerically. For high concentrations and viscosity ratios, auxiliary intermediate steps are recommended when moderate parameter values are used and the obtained solution is utilized as an initial guess in the final calculation.

For circular inclusions, the SCA estimate of the effective isotropic viscosity μ^{eff} is a solution of the quadratic equation

$$\beta^2 + (1-2f)(R-1)\beta - R = 0 \tag{17}$$

where $\beta = \mu^{eff} / \mu^{host}$. We note that in the isotropic case the SCA results in the geometrical mean of the host and inclusion viscosities for composites with equal phase concentrations irrespective of which phase (strong or weak) forms the inclusions. This continuity at the concentration of fifty percent stems from the inherent equivalence of the inclusions and the host in the SCA method.

In the DEM another approach is taken to adjust the estimate at high concentration for evolving overall properties. Here, the medium is constructed in an iterative manner by placing individual inclusions into the host and reevaluating the host properties afterwards (e.g., Berryman et al. 2002). In the isotropic case, this process is described by the following differential equation

$$\frac{d\beta}{df} = \frac{1}{1-f}\left(R - \beta\right)\frac{2}{1 + R/\beta} \tag{18}$$

and the initial condition is given by 1.

For circular inclusions, we find that the normalized effective viscosity β is a solution of

$$\beta + (1-f)(R-1)\sqrt{\beta} - R = 0 \tag{19}$$

Thus, the DEM prediction can be obtained by taking the SCA estimate, eqn. (17), in a zero to fifty percent concentration range, scaling the concentration by two and squaring the result. In the DEM case, the overall properties of composites of equal phase abundances are sensitive to the mechanical character of the inclusion forming phase. Such aggregates with competent inclusions exhibit a smaller effective isotropic viscosity than cases where the strong phase is the structure supporting host.

We have modified the DEM scheme to take into account the anisotropy according to (4) and (5). The effective viscosities are obtained by integrating the two coupled ordinary differential equations

$$\begin{aligned} \frac{d\beta_n}{df} &= \frac{1}{1-f}\left(R-\beta_n\right)\frac{\sqrt{\delta^{eff}}+\nu}{\sqrt{\delta^{eff}}R/\beta_n+\nu} \\ \frac{d\beta_s}{df} &= \frac{1}{1-f}\left(R-\beta_s\right)\frac{\sqrt{\delta^{eff}}+\nu}{\sqrt{\delta^{eff}}+\nu R/\beta_s} \end{aligned} \tag{20}$$

where $\beta_s = \mu_s^{eff} / \mu^{host}$ and the initial values of both β_n and β_s are 1. This system is again solved numerically.

In the rigid inclusion limit, i.e. $R \to \infty$, we have been able to obtain a closed-form solution

$$\sqrt{\beta_n} = 1 + \frac{\eta+1}{2}\frac{f}{1-f} \tag{21}$$

$$\sqrt{\beta_s} = 1 + \frac{\eta+1}{2\eta}\frac{f}{1-f} \tag{22}$$

Both SCA and DEM schemes can be shown to respect the reciprocity relation

$$\frac{\mu_n^{eff}\left(R,f,\sigma\right)}{\mu^{host}} = \frac{\mu^{host}}{\mu_s^{eff}\left(1/R,f,\sigma\right)} \tag{23}$$

that links alternating effective viscosities of composites obtained by a phase swapping. This allows to obtain the incompressible void limit of (21) and (22).

FEM model

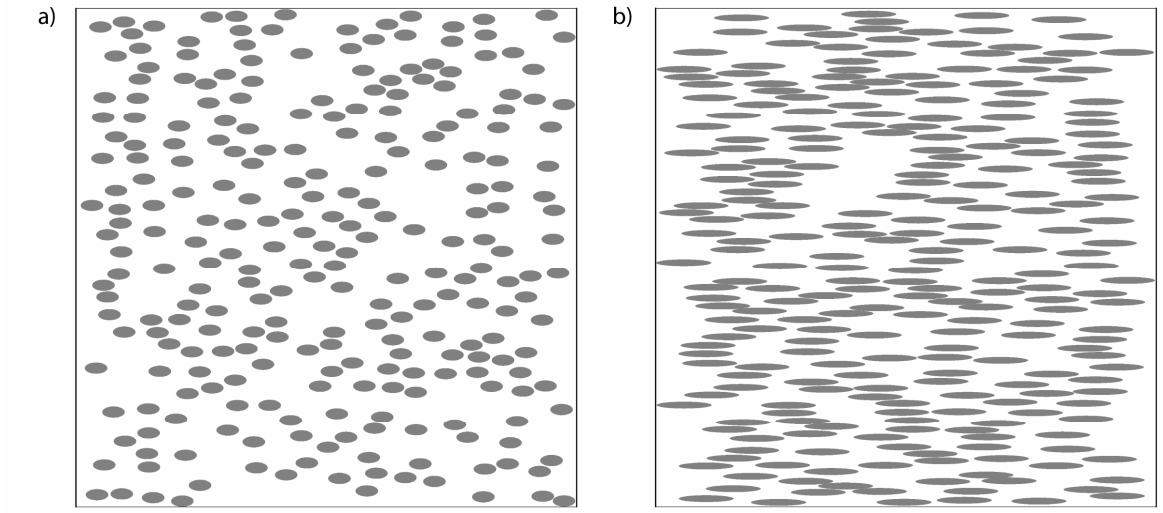


Figure 3
Examples of model setups used in effective viscosity calculations a) aspect ratio 2, concentration 20% b) aspect ratio 8, concentration 30%

To validate the SCA and DEM schemes, a finite element model (FEM) that allows us to directly resolve the mechanical response of composites consisting of numerous inclusions of a constant size and orientation has been employed (see Appendix for FEM details). We have systematically scanned through the parameter space of the inclusion concentration (up to 50%), ellipticity (up to 16) and viscosity ratio (between 1/1000 to 1000). Non-overlapping inclusions of equal size have been seeded randomly in the computational domain (see Figure 3). For a given ellipticity and concentration, ten samples of different inclusion configuration have been analyzed. The effective normal and shear viscosity are measured by applying appropriate kinematic boundary conditions.

Numerical results for models consisting of circular inclusions are presented in Figure 4. It is evident in Figure 4a that the presence of the strong host results in a higher effective viscosity for a composite of equal phase concentrations. The scatter of the overall property due to changing the composite configuration increases with the inclusion concentration. However, even in densely packed cases, the spread is rather small and the data presented at 5% concentration increments overlap minimally in terms of effective viscosity. In the circular inclusion case, the anisotropy factor has not exceeded a magnitude of 1.1. We note that the results in Figure 4a show that the reciprocity relation is satisfied only approximately.

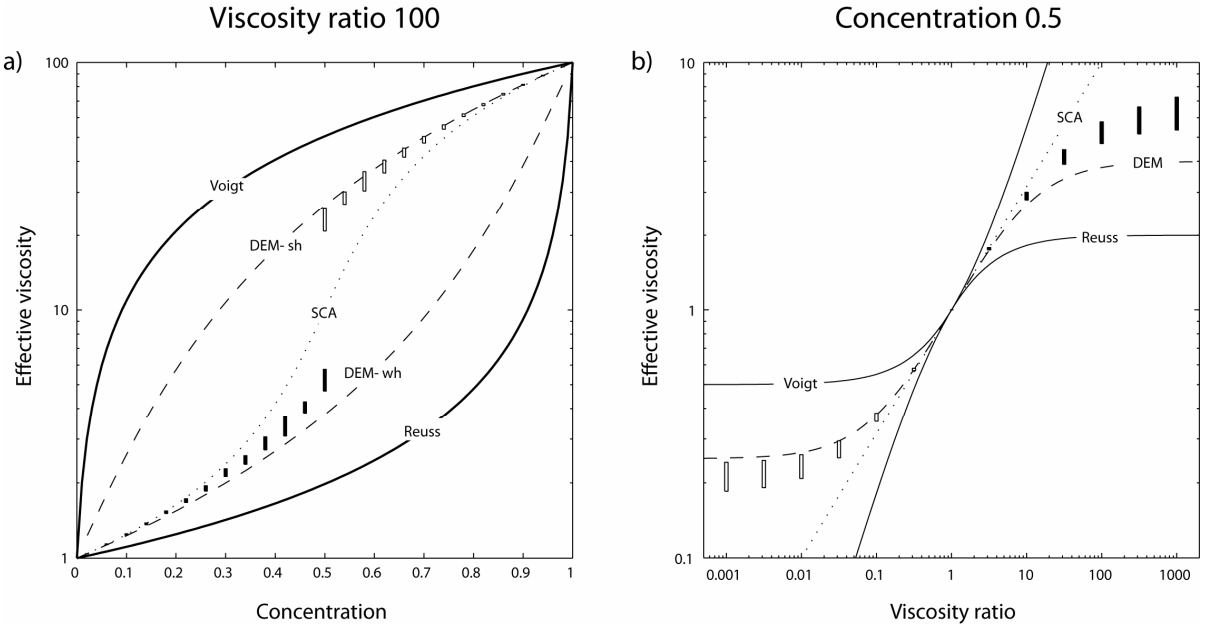


Figure 4

Effective viscosity of composites consisting of 256 circular non-overlapping inclusions. Open and filled bars correspond to models with weak and strong inclusion, respectively. Bottom and top of the bars are given by minimal and maximal values recorded for 10 samples. Upper (Voigt) and lower (Reuss) bounds, self-consistent average (SCA) and differential effective medium estimate for strong (DEM-sh) and weak host (DEM-wh) are given. a) Viscosity ratio is set to 100 and concentration refers to the strong phase b) Concentration is fixed at 50% and inclusion-host viscosity ratio is varied.

The impact of the viscosity ratio on the effective property is depicted in Figure 4b for models of equal host and inclusion concentration. The effective viscosity is virtually insensitive to inclusion or host viscosity changes once the strong to weak phase viscosity ratio exceeds several hundred. We find that the limiting values corresponding to rigid inclusions or incompressible voids are not precisely reciprocal.

In all cases, the effective viscosity falls between the theoretical bounds given by the weighted arithmetic (Voigt) and harmonic (Reuss) averages of the phase viscosities. The SCA provides a good estimate of the effective viscosity for inclusion concentrations below thirty percent. The continuity of the SCA estimate prevents it from predicting the saturation effect for high concentrations and towards large viscosity ratios. The DEM estimate is capable of capturing the effective viscosity saturation and provides a very good fit to the numerical results, especially in the weak inclusion case.

Selected results obtained for models consisting of aligned elliptical inclusions are presented in Figure 5. Numerical simulations show that the effective normal viscosity is systematically

greater than the shear viscosity (see Figure 5a). The data scatter for a given concentration is larger for the normal viscosity in the strong inclusion case. The reciprocal relation again provides a good approximation, but it is not satisfied exactly. Both overall viscosities saturate with respect to the viscosity ratio changes (see Figure 5b). In the weak inclusion case, the effective normal viscosity shows a relatively small and quickly flattening increase with the inclusion ellipticity, whereas the shear viscosity exhibits a substantial drop still ongoing at the aspect ratio of sixteen (see Figure 5c). For a composite of the equal phase content, the overall normal viscosity of the strong inclusion configuration becomes larger than the overall shear viscosity of the weak configuration for the ellipticity larger than 8. The degree of the anisotropy grows with the inclusion aspect ratio and inclusion-host viscosity ratio (see Figure 5d). Enhancing the viscosity ratio for a given ellitpicity results in a limited increase of the anisotropy factor. The anisotropy factor achieved for the aspect ratio of sixteen and the viscosity ratio of a thousand yields a magnitude of ten. The anisotropy factor is not exactly symmetric with respect to the phase swap again reflecting the approximate character of the reciprocity relation in the systems.

The overall viscosities respect the theoretical upper and lower bounds in all simulation runs. The DEM provides a good fit to the numerical data over the whole range of the concentrations, whereas a quality of the SCA estimate deteriorates for densely packed composites (see Figure 5a). The normal viscosity in the weak case (or the shear viscosity in the strong case) is particularly precisely estimated. The remaining overall viscosities are better predicted in weak inclusions models, similarly to the isotropic case. The DEM again captures a limited change of the effective viscosities for the extreme viscosity ratios (see Figure 5b). The advantage of the DEM technique is clearly visible in Figure 5c, where the SCA fails to distinguish the weak and the strong inclusion scenario. The anisotropy factors are to some extent underestimated by the DEM method (see Figure 5d), The SCA fails to estimate the overall anisotropies for high inclusion concentrations, but it still provides a relatively good prediction of the anisotropy factor. During the presentation of the numerical results, we have focused on the concentration 50% case to analyze a performance of the averaging schemes in this difficult limit. Averaging-related errors substantially drop for both schemes with decreasing the concentration. At concentrations of 35% the numerical results and the DEM estimate are virtually indistinguishable in the weak case. At 10% inclusion concentration both DEM and SCA practically coincide with the FEM results.

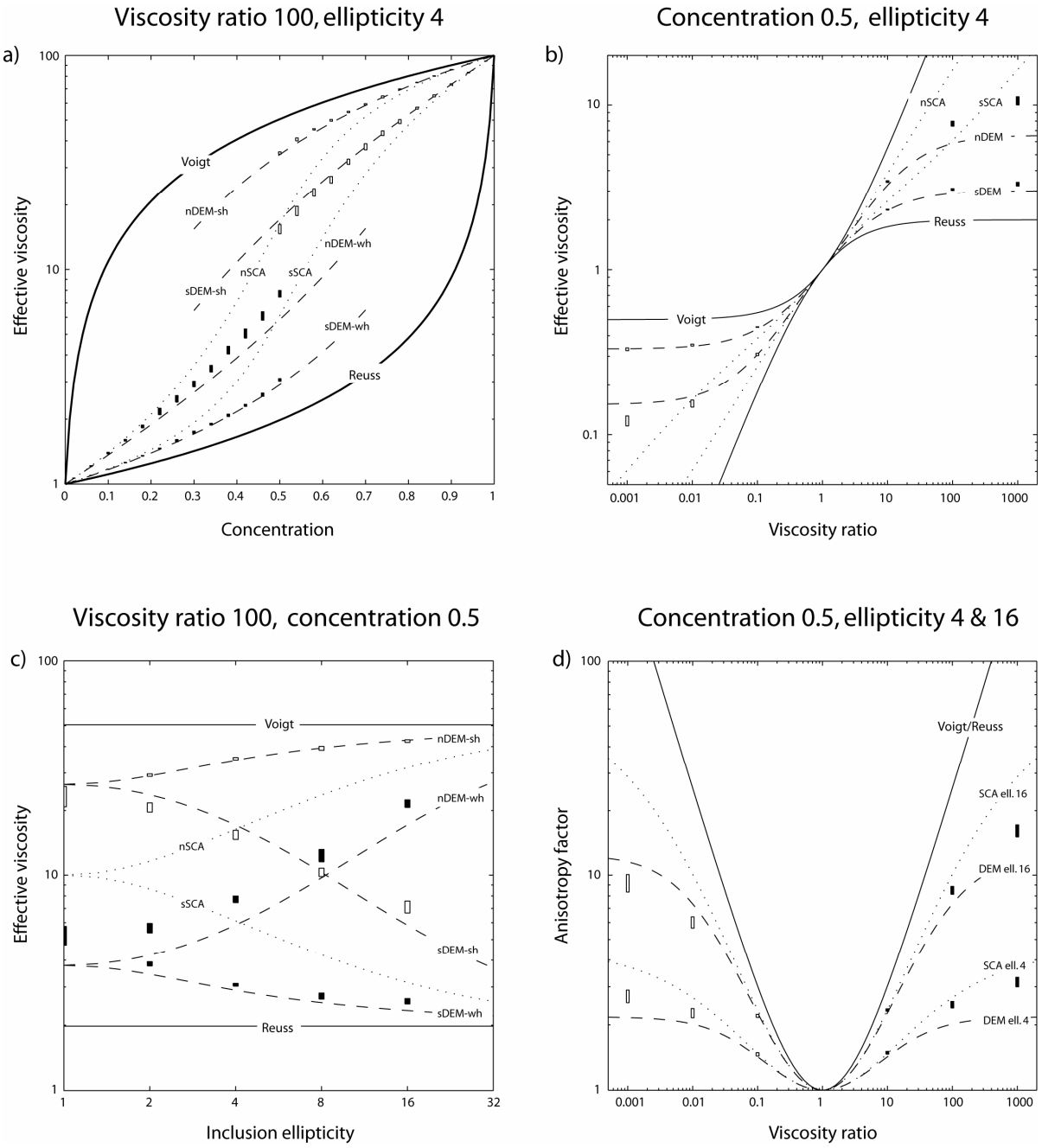


Figure 5

Effective normal and shear viscosity for composites consisting of 256 non-overlapping elliptical inclusions. Data bars are like in Figure 4 and normal viscosity is always above shear viscosity. Upper (Voigt) and lower (Reuss) bounds, self-consistent average (nSCA, sSCA) and differential effective medium estimate for strong (nDEM-sh, sDEM-sh) and weak host (nDEM-wh, sDEM-wh) are given Prefices n and s indicate normal and shear viscosity. a) Viscosity ratio is set to 100, ellipticity set to 4 and concentration refers to the strong phase b) Concentration is fixed at 50%, ellipticity set to 4 and inclusion-host viscosity ratio is varied.c) Viscosity ratio is set to 100, concentration is fixed at 50% and inclusion ellipticity is varied. d) Concentration is set to 50% and inclusion-host viscosity ratio is varied; computed anisotropy factors are plotted for aspect ratios 4 (lower values) and 16 (higher values).

Effective anisotropy of a two-phase composite subject to large deformation

We have so far derived the uncoupled shape evolution and overall anisotropy models. The shape evolution model treats an isolated elliptical inclusion coaxially embedded in a host with a constant anisotropy. The anisotropy model predicts the shear and normal viscosity of a composite consisting of aligned elliptical inclusions. Here, we combine them into a final set of equations describing the effective viscosity evolution for a two-phase composite subject to a large deformation. Finite element simulations of composites containing numerous deformable inclusions under pure and simple shear allow us to validate the analytical model.

Analytical model

We assume that a rock consisting of initially circular weak inclusions develops an overall shape fabric that evolves identically to an isolated inclusion. The SPO formation results in a change of an effective viscosity and the developing anisotropy feeds back to the shape evolution of the inclusions. The system is driven by a combination of far field pure and simple shear of rates $\dot{q}$ and $\dot{s}$, respectively. This combined far field condition is characterized by the vorticity of $-\dot{s}/2$ and instantaneous shearing rate of $\sqrt{\dot{q}^2+\left(\dot{s}/2\right)^2}$ with stretching axis inclined at $\varphi^{ISA}=\arctan\left(1\Big/\sqrt{1+\left(\dot{s}/2\dot{q}\right)^2}-1\right)$. Due to the anisotropy the overall stress is not necessarily coaxial with the overall strain rate tensor. In the bulk mechanical characterization, we use the work performing coaxial effective stress component normalized by the strain rate intensity, i.e. the effective anisotropic viscosity is resolved in the instantaneous stretching axis and the normal component for the normal load is taken. We refer to it as the ISA viscosity. By combining the previous models, a final set of equations is obtained

$$\dot{\sigma}=2\sigma\frac{\sigma^2+2\sqrt{\delta^{eff}}\sigma+1}{\sigma^2+2R/\beta_n\sqrt{\delta^{eff}}\sigma+1}\left(\dot{q}\cos\left(2\varphi\right)+\frac{\dot{s}}{2}\sin\left(2\varphi\right)\right) \tag{24}$$

$$\dot{\varepsilon}_{xy}^{incl}=\frac{\sigma^2+2\sqrt{\delta^{eff}}\sigma+1}{R/\beta_s\,\sigma^2+2\sqrt{\delta^{eff}}\sigma+R/\beta_s}\left(\dot{q}\cos\left(2\varphi\right)+\frac{\dot{s}}{2}\sin\left(2\varphi\right)\right) \tag{25}$$

$$\dot{\varphi}=-\frac{\dot{s}}{2}+\frac{\left(R/\beta_s-1\right)\left(\sigma^2-1\right)}{R/\beta_s\,\sigma^2+2\sqrt{\delta^{eff}}\sigma+R/\beta_s}\left(\dot{q}\cos\left(2\varphi\right)+\frac{\dot{s}}{2}\sin\left(2\varphi\right)\right)+\frac{\sigma^2+1}{\sigma^2-1}\dot{\varepsilon}_{xy}^{incl} \tag{26}$$

$$\beta_n=\beta_n^{DEM}\left(\sigma,R,f\right),\quad \beta_s=\beta_s^{DEM}\left(\sigma,R,f\right),\quad \delta^{eff}=\beta_n^{DEM}/\beta_s^{DEM} \tag{27}$$

$$\mu_{nn}^{ISA}=\mu_n^{eff}\cos^2 2\left(\varphi-\varphi^{ISA}\right)+\mu_s^{eff}\sin^2 2\left(\varphi-\varphi^{ISA}\right) \tag{28}$$

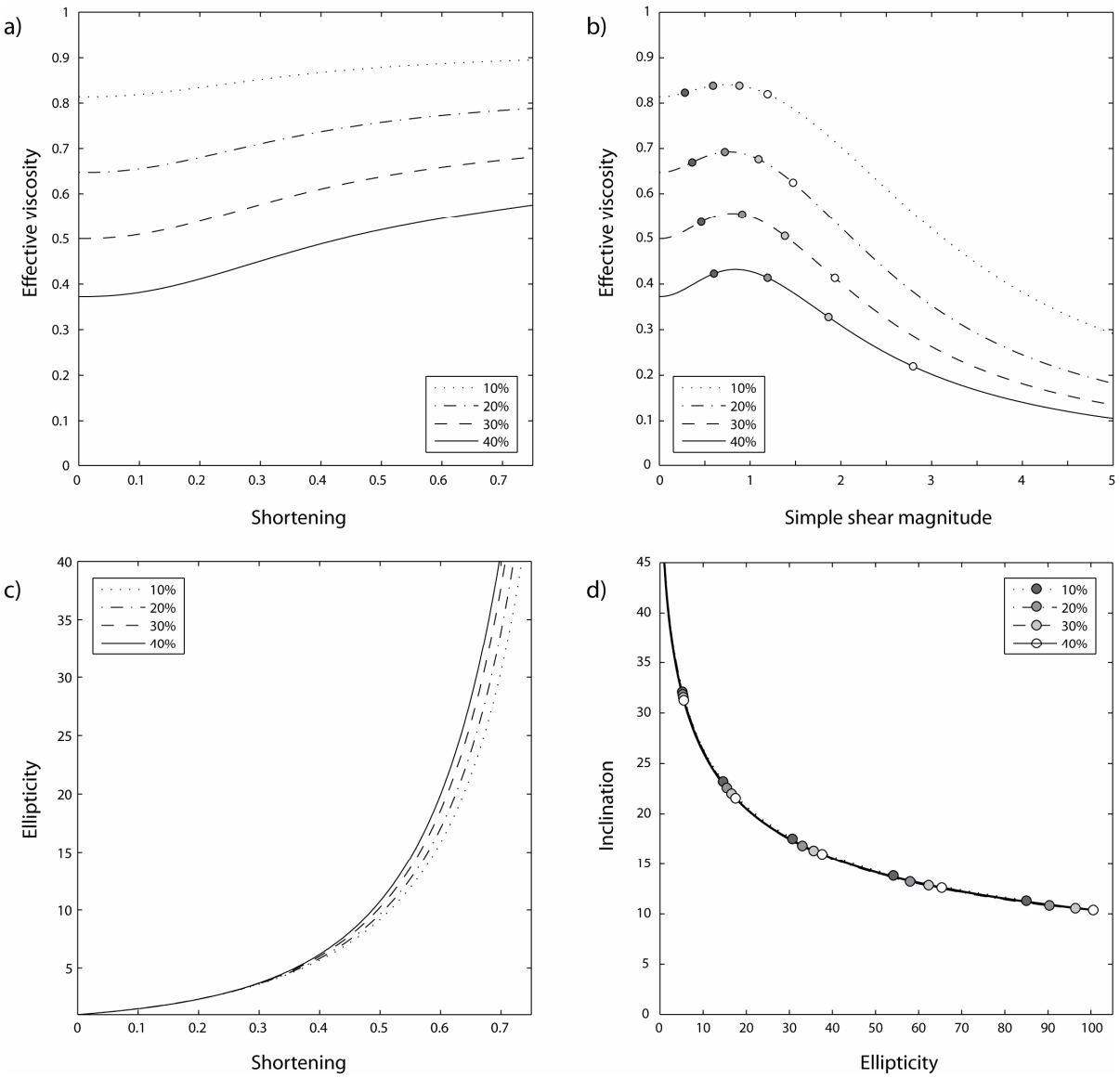


Figure 6

Composite of initially circular inclusions 100 times weaker than host under pure and simple shear. Inclusion concentrations of 10, 20, 30 and 40% are analyzed. Evolution of effective viscosity in pure shear (a) and simple shear (b). In the simple shear case, markers are plotted after a time needed to achieve shear strain of 0.25, 0.5, 0.75 and 1 in a pure host assuming a constant stress for all composites. c) Ellipticity evolution in pure shear case. d) Inclination-ellipticity paths in simple shear. Here, markers are plotted for unit increments of overall strain. Simple shear strain of magnitude 5 corresponds to overall shortening of ca. 80%.

The system of eqn. (24) and (26) is integrated numerically employing the effective properties evaluated according to the DEM model (27). For a given concentration f and inclusion-host viscosity ratio R, the effective viscosities can be precomputed for an expected range of the inclusion ellipticity σ. The overall mechanical response is obtained by the tensorial transformation of the anisotropic viscosity according to eqn. (28).

We present model predictions for weak inclusions (R=0.01) of concentrations of 10, 20, 30 and 40 % under pure and simple shear in Figure 6. In the pure shear case, the effective mechanical resistance is given by the overall normal viscosity. The analytical model predicts a stiffening behavior of a magnitude that increases for higher inclusion concentrations. For a given concentration, the increase of the overall viscosity is bound by the corresponding Voigt limit representing the response of a perfect bilaminate. The strengthening saturates quicker for smaller concentrations (see Figure 6a). We find that the influence exerted by the developing host anisotropy on the inclusion shape evolution is relatively weak and only becomes noticeable for high strains. This is corroborated by the fact that the anisotropy factors achieved after a shortening of fifty percent yield values of only 1.5, 2.5, 4 and 6.5 for the analyzed concentrations, respectively. The anisotropy that is relatively low in magnitude and emerges only at the late deformation stages is unlikely to have a significant impact on the shape evolution. In addition, in the high concentrations and large strains regime interaction effects are expected to be of major importance. Therefore, we suggest that the anisotropy feedback on the inclusion motion can be neglected.

In the simple shear case, the overall mechanical resistance of a two-phase system exhibits an initial hardening stage that is followed by a pronounced softening. The level of hardening again increases with the concentration (up to fifteen percent of the initial viscosity for the densest analyzed composite) and the peak strength is reached after a strain of around one. This transitional strain increases slightly with the concentration. The Reuss bound in the system yields 1/9 1/13, 1/15 and 1/15 of the initial viscosity for the studied concentration of 10, 20, 30 and 40%, respectively. We notice that the lower bound for 50% is given by 1/13 of the initial effective viscosity and the softening potential is lower in this case. The final degree of the softening (after a simple shear of 5) relative to the initial viscosity is approximately the same for the 20, 30 and 40% concentrations. We find that that the softening rate is enhanced for higher concentrations, but it is counteracted by a stronger initial hardening in these cases.

The inclination-ellipticity paths are virtually indistinguishable for different inclusion concentrations. The anisotropy influence is also minor in terms of the inclination-and ellipticity-change with strain. Similarly to the pure shear case, we propose that the model could be simplified by assuming the shape evolution of an inclusion embedded in a homogeneous host.

Strain evolution in 1D heterogeneous column

Assuming a 1D profile of varying concentrations subject to simple shear (a constant shear stress in the profile), we observe that denser composites pass through the transition strain be-

fore the dilute portions that are still subject to the hardening (see Figure 6b). The weakest parts in the column exhibit the largest strain that promotes the effective viscosity softening. However, the strain localization is practically determined by the initial viscosity profile.

Numerical model

To verify our finite strain analytical model, we have numerically studied the structural evolution of linear viscous composites initially comprising circular inclusions under pure and simple shear. Inclusion concentrations of 10 and 30% have been analyzed and are referred as sparse and dense case, respectively. We have focused on the incompetent inclusion scenario with inclusions hundred times weaker than the host material. The system evolution has been integrated to 70% of shortening in the pure shear case and a simple shear of magnitude 3. The computational domain is square in the pure shear case. Virtual load steps are introduced during the pure shear simulations to measure the effective shear viscosity (see Appendix). The rectangular domain of the width to height ratio of two has been utilized in the simple shear case. The velocity on the lateral walls is assumed periodic. We present the effective viscosity and overall shape evolution together with the predictions of the finite strain analytical model using the DEM scheme. The second strain rate invariant maps are shown at certain deformation stages.

The results obtained for the sparsely populated aggregate in the pure share case are shown in Figure 7. We find that the computed evolution of the effective normal viscosity closely follows the analytical prediction. The composite undergoes the hardening stage of about 10% and at 60% shortening the effective response is virtually given by the upper Voigt bound for this system. The analytical model predicts a nearly two-fold decrease of the effective shear viscosity at the end of the simulation. The shear viscosity measured during the virtual steps of the simulation show a small discrepancy compared to the analytical estimate. The anisotropy factor yields approximately 2 after the shortening of 60%.

The numerical model shows that the mean ellipticity of the individual inclusions assumes smaller values than predicted even in the isotropic host case. Thus, the effect of the developing anisotropy on the shape evolution is completely obliterated by the inclusion interaction. At the final stages, the standard deviation of the long axis orientation from the horizontal direction yields 3 degrees and the variation reaches as much as 8 degrees. The dispersion of the inclusion axes results in an even smaller ellipticity measure based on the overall shape (12 compared to predicted 18). The higher effective shear viscosity observed in the numerical simulation may be explained by the difference between the shape evolution prediction and computation.

In the dense composite case analyzed in Figure 8, the effective normal and shear viscosity at the simulation onset differ by several percent. This initial weak anisotropy can be attributed to the inclusion center configuration. The overall normal viscosity is predicted to raise around 30% after 60% of shortening. The numerical result shows an increase of around 20% that is followed by stagnation after 40% of shortening and even a weak decrease afterwards. The effective shear viscosity is predicted to drop more than has been recorded in the numerical simulation. The analytical model estimate approaches the Reuss bound towards the simulation end and results in more than 10-fold drop of the property. The final value recorded in the FEM simulation is only twice lower than the initial viscosity. The computed anisotropy factor yields between two and three, while the model predicts even three times stronger anisotropy after 60% shortening.

The shape evolution results corroborate our previous finding that the interaction reduces the inclusion stretch. The overall ellipticity is twice lower than analytical prediction for the isotropic host at the simulation end. The long axis inclination of the fitting ellipses can reach as much as 14 degrees from the horizontal direction with standard deviation of the axes dispersion given by 5 degrees. Despite the strong growth reduction, the inclusion aspect ratio still increases monotonically with shortening. The effective normal viscosity stagnation or even reduction after a considerable amount of strain indicate that the composite develops a complex structure and the simple elliptical SPO measure cannot be used to calculate the effective mechanical properties.

Inspecting the strain rate intensity maps shown in Figure 8, we find that an interesting pattern of localized deformation forms at the late stage of the simulation. After 60% of shortening, conjugate zones of localized shearing become apparent. These segments are oriented at approximately 30 degrees to the horizontal direction. The strain rate intensity within the bands locally exceeds the background rate by factor 5. The formation of the bands can be traced back to the earlier stages showing the band propagation and reorientation towards the shallower inclinations.

The bands are formed by inclusion trails with the intense shearing focused into their middle segments. The inclination of the central inclusion parts is steeper then the overall inclusion orientation that is below 14 degrees. The rim portions of these inclusions are horizontal. Consequently, the inclusions involved in the band formation are characterized by the complex sigmoidal shapes. We note that the stretch measured along the inclusion is substantially larger than if obtained by the ellipse fitting.

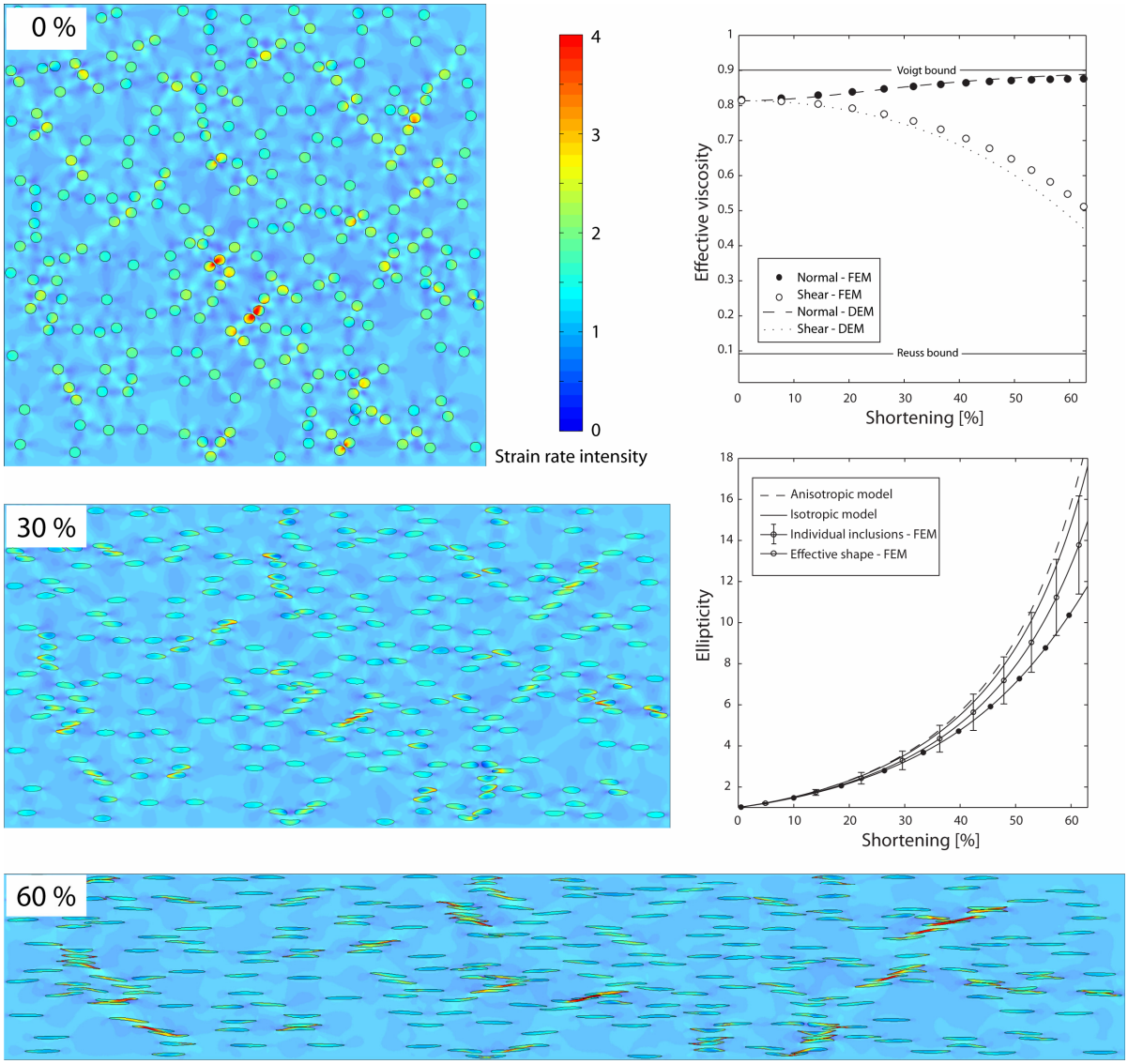


Figure 7

Finite element results for a two-phase composite with 10% of initially circular inclusions subject to pure shear. Inclusions viscosity is 0.01of host viscosity. Strain rate intensity (second strain rate invariant) map is shown at initial step and after 30% and 60% of shortening. The results are plotted with higher transparency in the host. Background strain rate intensity is equal to 1. Effective normal and shear viscosity is shown in the upper plot together with the DEM predictions and Reuss and Voigt bounds. The inclusion ellipticity evolution is shown in the lower plot. Ellipticity is given for individual inclusions (mean value and standard deviation) and for effective shape (for details see Appendix). The anisotropic and isotropic model predictions are given for comparison.

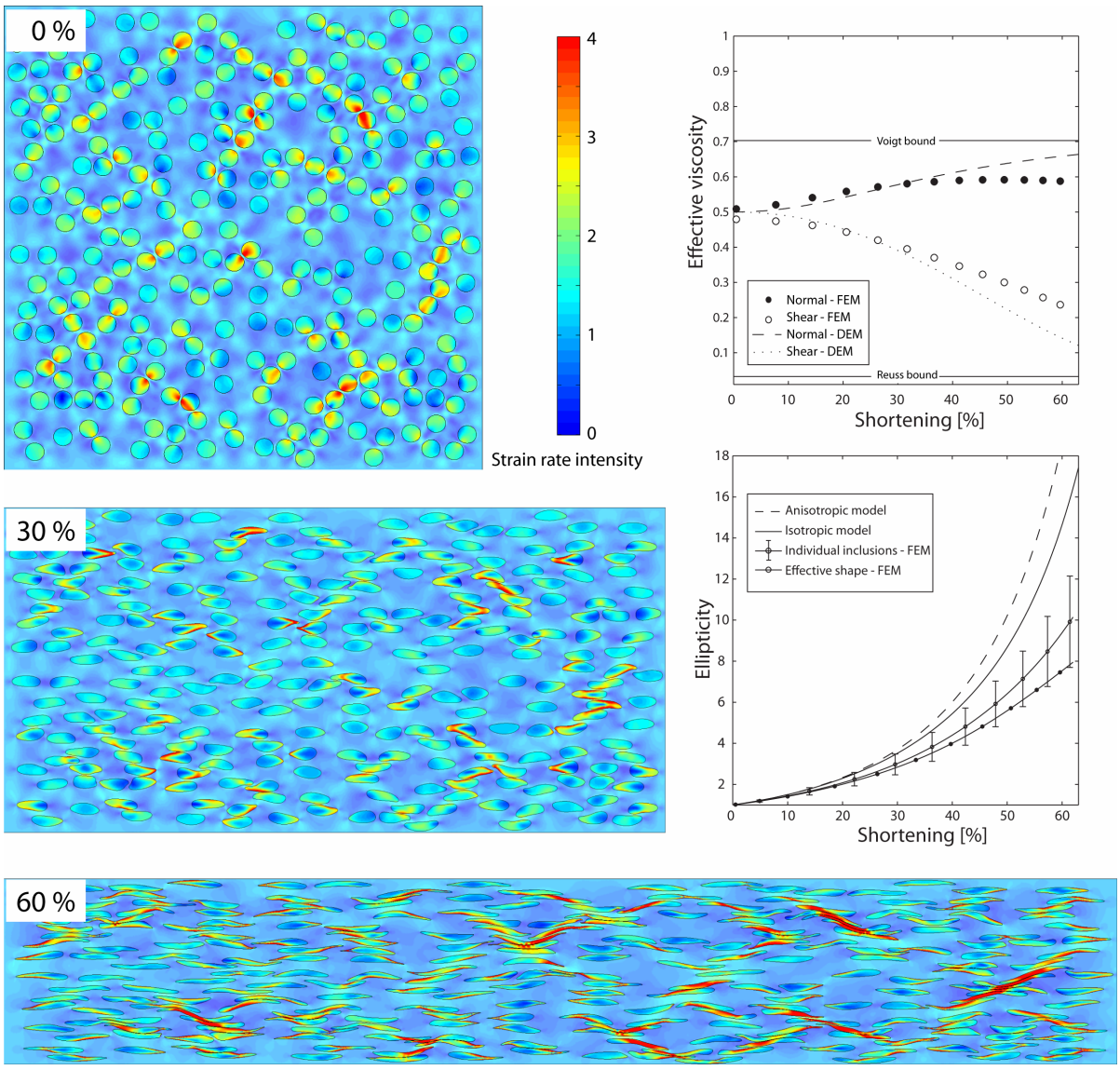


Figure 8

Finite element results for a two-phase with 30% of initially circular inclusions subject to pure shear. See caption below Figure 7 for more details.

The network of the localized deformation bands encloses rhomboidal domains that still contain the inclusions. The deformation rates within the inclusions entrapped in these domains are low and often exhibit values below the background strain rate. These inclusions, effectively detached from the far field loads, are characterized by the ellipticity as low as 5 at the end of the simulation. For comparison, the ellipticity of a passive isolated initially circular inclusion would assume more than 6 in this case.

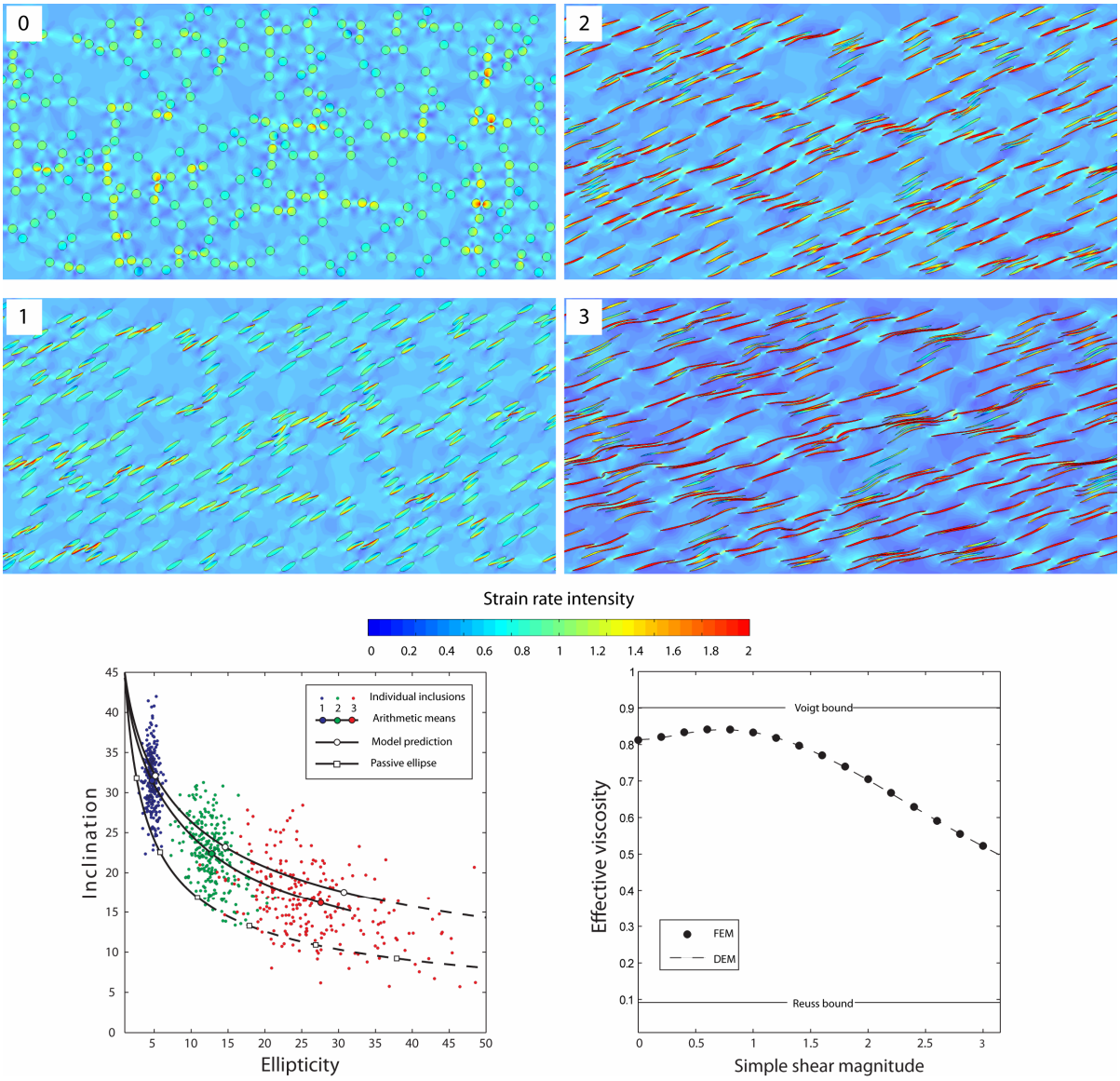


Figure 9

Finite element results for a two-phase composite consisting of 10% of weak inclusions subject to a simple shear. Inclusions viscosity is 0.01 of host viscosity. Strain rate intensity (second strain rate invariant) map is shown at initial step and after simple shear of magnitude 1, 2 and 3. Background strain rate intensity is 0.5 (half of simple shear rate). Ellipticity-inclination data are shown in the left plot (individual inclusions after unit increments of strain, effective shape, model prediction, reference passive evolution). ISA viscosity computation and model estimate are presented in the right plot.

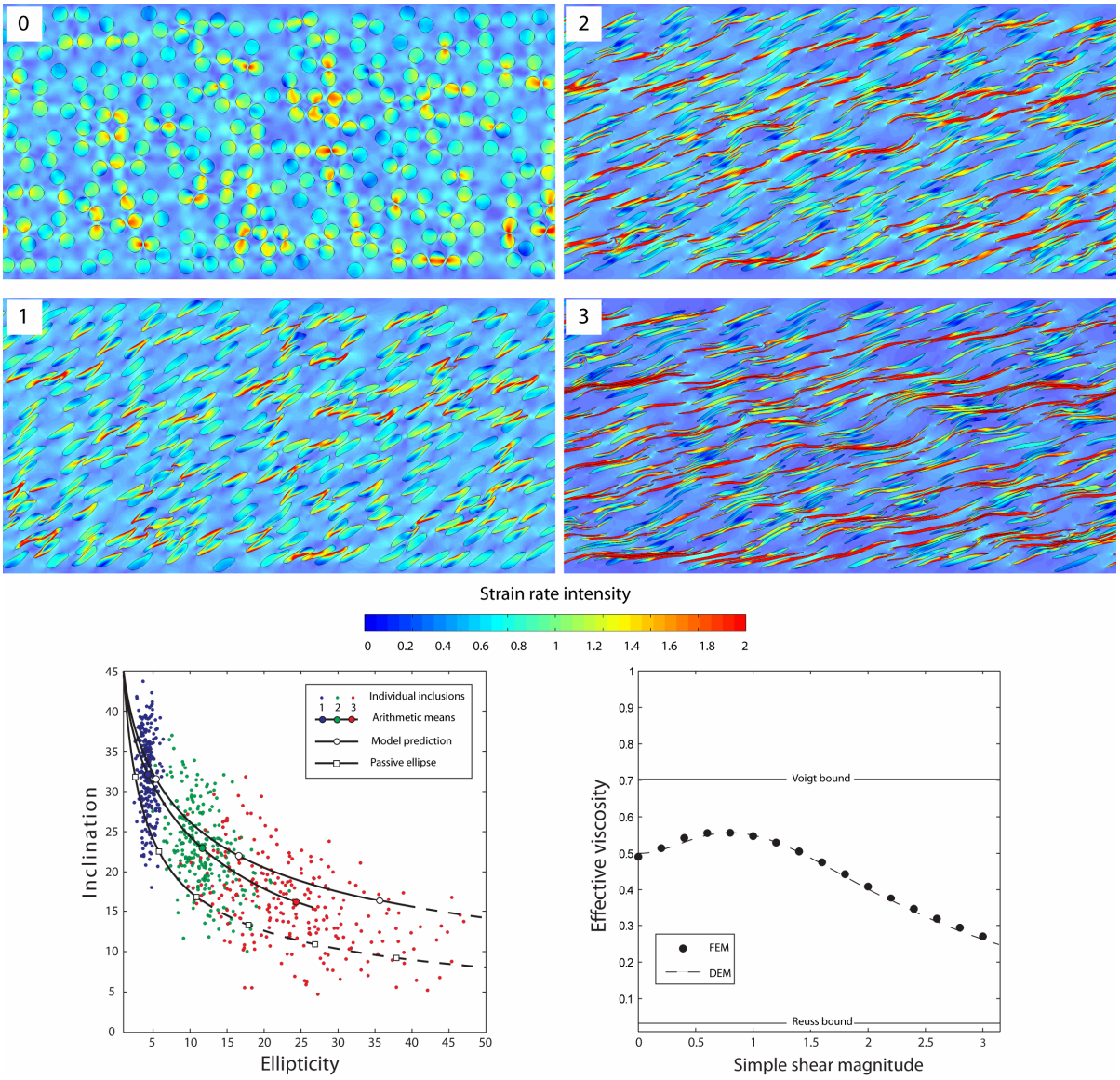


Figure 10

Finite element results for a two-phase consisting of 30% of weak inclusions subject to a simple shear. See caption below Figure 9 for more details.

In the second set of our numerical experiments, a simple shear of magnitude 3 has been reached at the end of the simulations. We provide the reference inclination-ellipticity paths of an initially circular passive object, while discussing the inclusion shape evolution. The overall deformation accumulated at the end of the simulation runs is characterized by the 70% shortening and the finite strain ellipse long axis is inclined at 15 degrees to the simple shear direction.

For the sparse composite (see Figure 9), the ISA viscosity measured during the simulation follows the analytical prediction. Thus, the system undergoes the initial hardening followed by the weakening after the strain of magnitude one. The ISA viscosity at the simulation end yields

60% of the initial value. Unlike in the pure shear case, there is no discrepancy between the numerical result for the dense aggregate (see Figure 10) and the finite strain analytical model prediction. Here, the ISA viscosity exhibits a two-fold drop over a simple shear of magnitude 3 and the magnitude of the hardening phase yields 10% of the initial value.

The inclination-ellipticity plots show that the SPO ellipticity development is suppressed in comparison with the analytical prediction, but to a lesser degree than in the pure shear case. In comparison with the analytical prediction, the average inclination is shifted towards the shallower values typical of an initially circular passive object evolution. The inclinations of the individual inclusion are strongly dispersed and the data is not even contained between the curves given by the analytical prediction curve and the passive inclusion paths. The data scatter increases with concentration, but the dilute case is already strongly affected by the interaction effects.

The strain rate intensity maps clearly show that the shearing is progressively focused into the inclusions during the deformation. The inclusions pervasively develop complex sigmoidal shapes and overall model appearance after simple shear strain of 2-3 resembles of an asymmetric S-C structure. The strongly sheared middle parts of the inclusions (C) are transposed in the horizontal directions, whereas the inclusion tips (S) are even steeper than 45% and show a tendency to the antithetical rotation.

Discussion

We have derived an analytical model of the effective viscosity evolution for a two-phase rock characterized by the inclusion-host geometry type. The model consists of coupled equations describing rates of the ellipticity and inclination for an isolated inclusion in the anisotropic host that represents a developing SPO in the composite. The anisotropy is directly related to the SPO and we have derived and numerically validated the differential effective medium approach to model the overall normal and shear viscosity of a composite consisting of aligned elliptical inclusions. Model predictions have been analyzed for the pure and simple shear deformation paths for composites with initially circular weak inclusions. Analytical considerations have been supplemented by a finite element modeling. The following discussion is divided into a three major topics: inclusion shape evolution, effective properties of a composite comprising aligned elliptical inclusions and effective viscosity evolution during a finite deformation.

Shape evolution

The motion of an isolated elliptical inclusion embedded in the anisotropic host and subject to a uniform far field strain rate can be entirely described in the ellipticity-inclination space. This is due to the fact that an elliptical object uniformly sheared remains elliptical. Assuming constant host anisotropy with the inclination determined by the inclusion orientation, we have obtained the evolution paths for an initially circular inclusion subject to a pure or simple shear. In the pure shear case, a closed-form solution has been presented. The presence of strong anisotropy in the host substantially enhances the inclusion stretching and also leads to elevated rotation rates at the initial stages of a simple shear.

The analytical finite strain model predicts moderate anisotropy factors for a composite consisting of initially circular inclusions hundred times weaker than the host even after a considerable amount of strain. In particular, the anisotropy is too weak to exert any significant effect on the inclusion motion until the late deformation stages. The FEM modeling shows that in fact the inclusion interaction strongly overprints the effect of the anisotropy. The average inclusion aspect ratio is reduced and this effect is particularly prominent in the pure shear case. In this case, the inclusions residing within the domains enclosed by the localized deformation bands may exhibit ellipticity even smaller than that of a passive marker. This indicates that caution is needed while inferring overall strain based on the aspect ratios of weak inclusions occurring in a non-dilute composite rock.

The band forming inclusions develop complex sigmoidal shapes both in pure and simple shear cases. Other stretch measures could be used for these objects, since the ellipse fitting provides only a crude approximation to the shape in this case. The inclusions involved in the band formation may ultimately split into multiple objects and consequently bias the strain analysis. We note that the inclinations of the bands in the pure shear case and S segments in the simple shear case undergo a complex evolution with strain.

The motion of a weak inclusion entrapped or grown in an intrinsically anisotropic material is prone to follow the predictions of our shape evolution model for anisotropic materials. Porphyroblast grown in a medium with a pre-existing lamination of strongly contrasting materials provides a viable setup in this respect. In this case, the anisotropy factor is given by the ratio of the arithmetic and harmonic means weighted by phase compositions and may assume much higher values than analyzed here. Another possible scenario is given by an inclusion contained in a crystal. The viscosity of quartz crystals has been reported to vary around hundred times depending on a load direction (Linker et al. 1984). The inclusions in both cases are subject to a

constant strong anisotropy from the deformation onset, but the interaction effects may be still significant. The flow perturbation around an inhomogeneity results in a local stirring of the anisotropy orientation. The corresponding effects have been shown to have a large impact on the inclusion rotation for the rigid objects subject to a far field simple shear (Dabrowski & Schmid, submitted to JSG)

Effective anisotropy of a composite of aligned elliptical inclusions

We have developed a differential effective medium scheme for a composite consisting of aligned ellipses based on an analytical solution describing strain rates in an elliptical inclusion embedded in an anisotropic host and subject to a uniform far field loads. An existing scheme relying on the self-consistent averaging technique derived by Fletcher (2004) has been analyzed for a comparison. Both schemes are free of phenomenological input parameters and the output is provided in terms of the overall normal and shear viscosity. FEM results obtained for a wide range of inclusions ellipticity, concentrations and host-inclusion viscosity ratios have been utilized to validate the schemes.

The developed DEM model provides better estimate over the SCA for higher concentrations. The discrepancies between the scheme predictions reflect a fundamental difference between the two methods: the DEM is designed for inclusion-host systems, whereas the SCA is more suitable for a poly-grain medium, where none of the phases can be considered as inclusions. This statement is corroborated by the FEM results that are predicted by the DEM scheme with a high accuracy up to large concentrations irrespective of the inclusion aspect ratio.

Our FEM results show that choosing a weak or strong inclusion forming phase leads to significantly different effective properties. This is in agreement with the notion of a load-bearing framework that has been previously suggested in the geological literature (e.g., Handy 1990). Existing theoretical models often have difficulties in predicting the overall viscosity in this transitional regime between the weak and strong phase supported structures (Ji et al. 2001). For example, the SCA scheme cannot predict the observed discontinuous transition between the two composite types, because it treats the host equivalently to the inclusions. Consequently, for a two-phase composite of the equal phase concentration the overall viscosity is the geometrical mean in the isotropic case, and the alternate viscosities have to match during the phase exchange in the anisotropic case. The DEM scheme is well capable of differentiating the weak and strong phase supported structures and we suspect it may explain experimental data both in isotropic and anisotropic case.

The other characteristic of the SCA scheme is an unbound behavior of the overall viscosity at the critical concentration of fifty percent if the inclusion phase is assumed to be rigid or infinitely weak. In this case, the DEM model predicts finite viscosities for any concentration and provides a good approximation to the corresponding FEM measurements. Extreme viscosity contrast is typical for melt-crystal systems, where the crystal phase additional exhibits the non-linear rheology (e.g., Arbaret et al. 2007). The advantage of the DEM scheme is two fold in this case. It not only captures the high concentration behavior better than the SCA, but also incorporating non-linear flow laws for the inclusions is straightforward.

We conclude that our DEM scheme provides a high-quality prediction of the overall properties for composites consisting of aligned elongated objects. We emphasize that our DEM scheme similarly to the SCA is free of unconstrained phenomenological parameters and provides an output in a suitable form of the normal and shear viscosity.

Both DEM and SCA scheme satisfy the reciprocity rule. This rule has been proven to hold exactly for the effective thermal conductive of composites with inclusions located on a periodic lattice (Keller 1987)). In general, an exact constraint on the effective viscosity dependence on the inclusion-host ratio for an arbitrary microstructure (or a particular microstructure such as inclusion-host geometry) is sought in this context. For example, is it possible to predict exactly an overall property for an arbitrary value of the viscosity ratio, if we know the answer for the infinitely weak and stiff inclusion cases? We have found that the reciprocity rule is not precisely satisfied in our simulations, but still provides a good approximation.

Finally, we note that our two-dimensional composite model consisting of cylindrical inclusions can be considered only as an approximation to a real poly-phase rock. The extension of the DEM to the three-dimensional ellipsoidal anisotropic case is rather straightforward, although in general the interior strain rates are not given in a closed form (e.g., Mura 1987). However, in the dilute limit the effect of the inclusion non-cylindricity on the effective properties can be easily evaluated by invoking the solution derived by Eshelby (1957) for an ellipsoidal inclusion embedded in the isotropic host. For instance, we find that for a spherical inclusion that is hundred times more viscous than the host, the effective viscosity increase with concentration in a dilute limit is around 40% stronger than in the cylindrical case.

Viscosity evolution due to structural development in a two-phase composite

Our finite deformation simulations show that a structural evolution in composites consisting of weak inclusions results in evolving overall viscosity both in pure and simple shear. In the pure shear case, the inclusion-host system exhibits a weak and quickly saturating hardening phase.

In the simple shear case, an initial hardening phase is followed after a unit strain by a pronounced weakening and this softening behavior is still active at a strain of three. We have attributed this rheological evolution to overall anisotropy emergence that is related to a developing shape preferred orientation. An analytical model that combines the shape evolution of an isolated elliptical inhomogeneity and the differential effective medium estimate for the overall anisotropy of a two-phase composite consisting of aligned elliptical inclusions has been derived. The FEM results can be reproduced with our analytical model to high accuracy even for densely packed composites up to large strain, despite complex inclusion shape development. The largest discrepancy is observed in the pure shear case, where the formation of localized deformation bands after a considerable amount of strain is manifested by the onset of weakening. The analytical model can be adapted for multiple phases and is suitable for any deformation paths. An ad hoc ductile strain softening rules are often employed in geodynamic simulations and our scheme provides an improvement in that respect.

A similar study employing a different numerical technique has been performed for regular arrays of square inclusions subject to a simple shear (Takeda & Griera 2006). Stress-strain curves in the weak inclusion scenario (strong phase supported structure) reported therein exhibit a similar sequence of hardening and softening behavior with a transition point located also around a shear strain of a magnitude one. We suggest that our model, although derived for circular inclusions, can provide a reasonable approximation to a finite deformation evolution of the effective mechanical property in the square inclusion case.

Multi-phase materials often exhibit a pronounced weakening after an initial transient stage of a stress build-up in simple shear laboratory experiments. The examples include experiments performed in shear apparatus (Holyoke & Tullis 2006) and or in torsion (Barnhoorn et al. 2005). Several models have been invoked to explain this behavior. Dynamic recrystallization resulting in a grain size reduction that enhances the diffusional creep mechanism provides a commonly used explanation. The related transient effects for different grain size reduction models have been analyzed by Montesi & Hirth (2003). The dynamic recrystallization model can also explain strain softenening mono-phase materials. However, the grain size reduction due to subgrain rotation and grain boundary migration is counteracted by the competing grain growth (Herwegh et al. 2005). Another typically invoked mechanism explaining the strain weakening in the ductile regime is the formation of LPO (e.g., Pieri et al. 2001). The stress evolution model proposed by Montesi (2007) for shear zones operating in the brittle-ductile regime relies on the anisotropy development due to the structural transformation. Similarly to the recrystalli-

zation model, the Montesi's structural model predicts no initial hardening phase in simple shear.

Processes resulting in strain weakening are often considered important with regard to ductile shear zone formation (for a review of mechanisms see e.g.. Bruhn & Burlini 2005, Regenauer-Lieb & Yuen 2003). The localization in a 1D column has been recently discussed by Paterson (2007) and the assessment of selected mechanisms including material non-linearity has been performed for two-dimensional setups by Kaus & Podladchikov (2006). Modeling of a shear zone development in inclusion-host systems shows a more marked impact of strain weakening than strain rate (host power-law rheology) weakening effects (Mancktelow 2002). His work shows that allowing for a 5-fold viscosity drop over a strain of 20% only results in a limited degree of localization that is far from explaining strongly focused zones occurring in natural settings. Takeda & Griera (2006) proposed a localization model in a rock column of a varying inclusion content based on a hardening and softening behavior observed for a two-phase medium in simple shear. Their key observation is that the softening phase is reached earlier in weaker portions, while deformed to a lesser degree stronger parts are still subject to hardening. We have applied our analytical model to a similar setup to quantify the degree of localization. A viable localization mechanism is expected to lead to substantial viscosity contrasts in a slightly perturbed material already after a small strain. However, the strain distribution in the profile is virtually determined by the initial concentration related viscosity variations even after a considerable strain. The initial viscosity contrasts are insufficient to explain the observed shear zone profiles in relatively homogeneous materials and we conclude that the structural softening cannot account for the ductile localization alone.

The question arises if introduction of unequally sized inclusions, intrinsic anisotropy or a non-linear rheology for the host or inclusions could enhance the strain weakening to a degree that a strong localization could occur in a perturbed zone. To avoid qualitative speculations, we focus on an example of a power-law fluid forming the inclusions. Our analytical model can be extended to include this case without a problem, since any rheology can be prescribed to inclusions subject to a constant strain rate in the model. During a simple shear the deformation is evidently focused into the inclusion phase. Hence, a decrease of the apparent viscosity results for inclusions following a power-law rheology. We have shown that the effective properties of a two-phase composite saturate with respect to the viscosity ratio variations (see e.g. Figure 5b). The shape evolution of an inclusion that is hundred times weaker than the host is virtual undistinguishable from the incompressible hole case. Consequently, for inclusions which are

significantly weaker than the host (more than a hundred) already at the deformation onset, the effects related to the power-law introduction are minor in the model. We note that this hypothesis may not hold in the FEM runs, where inclusions interact and develop complex shapes.

The FEM results show that in densely packed composites, the weakening behavior takes over after the initial hardening phase in the pure shear case. We have attributed this transition to the formation of the localized deformation bands. The effective viscosity may be significantly reduced due to the development of the pervasive rhomboidal network of weak bands effectively resulting in the change of the composite topology. Inclusion-host geometry reversal has been suggested as an efficient localization mechanism (e.g., Handy 1994). In the simple shear case, the formation of the prominent S-C structure is surprisingly not manifested in the stress-strain curves for the studied deformation magnitude. In our simulations, we have seen only the onset of the inclusion phase merging and the additionally strain required to bring the localization to the completion is difficult to assess. In the localization context, we seek a mechanism that could explain the abrupt topology reversal after a small strain. We admit that processes such as dynamic recrystalization that are prone to accelerate the weak phase coalescence are not taken into account in our simulations. Further studies are required to constrain the onset and duration of the topology reversal as a function of inclusion concentrations and other relevant factors.

Conclusions

The combined shape and mechanical evolution model for a two-phase rock consisting of inclusions subject to large deformation has been derived. We have utilized the analytical solution for an isolated elliptical inclusion embedded in an anisotropic matrix to model the shape preferred orientation development. The emergence of shape preferred orientation in a rock results in an overall mechanical anisotropy. A scheme based on the differential effective medium approach predicting mechanical properties of a composite consisting of aligned elliptical inclusions has been proposed and numerically validated. The scheme is free of phenomenological parameters and provides a good fit to the numerical data in a wide range of inclusion concentration, ellipticity and inclusion-host ratio. In particular, the weak and strong supporting phase cases are distinguished by the model leading to an improved estimate for compositions higher than 20 percent in comparison with the SCA. The anisotropic DEM scheme also removes the deficiency of an unbound prediction in the limit of extremal viscosity ratios for high concentrations. The evolving anisotropy is allowed to influence the shape development. The inclusion stretch is significantly enhanced if a strong host anisotropy is present.

In the pure shear case, flattening of weak initially circular inclusions results in an increasing normal viscosity and an effective hardening. The viscosity growth respects the Voigt bound and yields maximally fifty percent relative to an initial value for densely populated inclusion. In the simple shear case, the anisotropy emerges along the instantaneous stretching direction and is progressively transposed into a shearing direction. Consequently, a weak initial hardening phase occurs and after a strain of around unit magnitude, irrespective of inclusion concentration, a pronounced softening follows towards the Reuss bound. In both cases, feedback of the evolving anisotropy on the shape evolution is irrelevant.

Numerical simulations directly resolving a finite deformation evolution of a composite consisting of numerous weak inclusions have been performed. The numerical results show that analytical model can capture the anisotropy emergence and overall normal and shear viscosity development up to high inclusion concentrations and strains. Both in pure and simple shear, flow localizes into favorably oriented conjugate arrays of inclusions leading to the development of sigmoidal inclusion shapes. However, the shape evolution model relying on an isolated inclusion solution provides a satisfactory approximation to the shape preferred orientation development in an interacting inclusion aggregate. The localization on an inclusion scale is manifested in the overall system behavior only in the pure shear case after a considerable amount of strain.

Our model provides a viable explanation of a strain weakening observed in poly-phase materials. However, it cannot explain alone localization in a rock mass heterogeneously populated with inclusions. The model can be run for any deformation paths and constrains constitutive laws incorporating structural evolution factor that are often employed in large geodynamic simulations.

Appendix

A large number of inclusions needed for a sufficiently representative composite area results in a high discretization level exceeding million degrees of freedom present in our models. It necessitates the use of an efficient implementation of FEM. In this study we have utilized our unstructured mesh FEM code MILAMIN implemented entirely in MATLAB. We have employed the mixed formulation of FEM that allows us to tackle incompressible flow problems. In this formulation pressure field is approximated independently from the velocity field and corresponding degrees of freedom are explicitly present in the model. A particular element type that we have utilized is the seven-node Crouzeix-Raviart triangle with the pressure field

interpolated by a discontinuous linear function. The incompressible constraint is enforced through the penalty approach and performing Hestenes-Powell iterations allows us to use a moderate penalty factor. By eliminating the pressure degrees of freedom on the element level, we gain the possibility to operate on a positive-definite global matrix. The direct solvers are tailored for such systems and Cholesky factorizations are well known for their high performance. Details concerning implementation and applied optimizations have been describe in our other paper (Dabrowski et al. in press).

For the instantaneous models, numerical setup consists of a square box containing 256 inclusions of varying concentration and ellipticity that are aligned and elongated in x-direction (see Figure 3). We have studied concentrations in the range between 0.02 and 0.50 in 0.02 increments and inclusion aspect ratios of 1, 2, 4, 8 and 16. During the input model generation inclusions are sequentially seeded in the computational domain in random locations, and accepted only if the minimum distance to all other inclusions exceeds 5% of the short axis length. All studied models have been run for the viscosity ratio $r = \mu_{incl} / \mu_{host}$ 10, 100, and 1000. Concentrations higher than 50% are achieved by the phase swap (inclusions become weak in this case) and consequently the viscosity ratios of 0.1, 0.01 and 0.001 are then analyzed. For a given ellipticity, concentration and viscosity ratio 10 samples have been analyzed to assess the dependence on the inclusion network configuration. Each model instance is subject to two different modes of pure shearing given by average strain rates $\dot{\varepsilon}_{xx} = -\dot{\varepsilon}_{yy} = 1$ and $\dot{\varepsilon}_{xy} = \dot{\varepsilon}_{yx} = 1$. Both horizontal and vertical components of the velocity field on the domain boundaries are prescribed according to the average strain rates. The effective normal and shear viscosity is computed based on the mean value of the appropriate stress component.

In the finite deformation models, we have utilized a rectangular computational domain of the width to height ratio of 2 in the simple shear case and a square one in the pure shear case. In the pure shear case, we have prescribed kinematically determined boundary conditions on all walls. In the simple shear case, we have applied appropriate velocities on the bottom and top walls and enforced periodic velocities on the lateral walls. The inclusion geometry has been explicitly updated according to the current velocity field with the time step adjusted in the pure shear case to take into account the changing box geometry. We used 0.5% shortening increments in the pure shear case and 0.01 simple shear unit increments. The dilatational component of the displacement at the inclusion interfaces has been subtracted to ensure constant phase concentrations. In both cases, we have was frequently recreated the computational mesh to maintain its high quality.

The inclusion shape and effective viscosity evolution have been monitored during simulation runs. Several different measures of the inclusions shape have been tested. We have found that computing the second polygon moments is superior to the least-square ellipse fitting, because it is not biased by a non-uniform distribution of the inclusion boundary nodes. Eigenvalues and eigenvectors have been computed for all shape matrices to find ellipticity and inclination of the individual inclusions. In addition, the overall shape has been estimated by performing eigenvector analysis on the arithmetic mean of the individual shape tensors. The relevant components of the effective viscosity have been computed by averaging adequate stress components over the whole computational domain. For the pure shear case, the effective normal viscosity component is supplemented by the effective shear component. This requires an additional virtual step of calculations when boundary conditions are changed to a pure shear oriented at 45 degrees. The results obtained in this step have been used to compute the overall shear viscosity, but the model geometry has not been updated according to them.

References

Arbaret, L., Bystricky, M. & Champallier, R. 2007. Microstructures and rheology of hydrous synthetic magmatic suspensions deformed in torsion at high pressure. *Journal of Geophysical Research-Solid Earth* **112**(B10), -.

Barnhoorn, A., Bystricky, M., Kunze, K., Burlini, L. & Burg, J. P. 2005. Strain localisation in bimineralic rocks: Experimental deformation of synthetic calcite-anhydrite aggregates. *Earth and Planetary Science Letters* **240**(3-4), 748-763.

Berryman, J. G., Pride, S. R. & Wang, H. F. 2002. A differential scheme for elastic properties of rocks with dry or saturated cracks. *Geophysical Journal International* **151**(2), 597-611.

Bilby, B. A., Eshelby, J. D. & Kundu, A. K. 1975. Change of Shape of a Viscous Ellipsoidal Region Embedded in a Slowly Deforming Matrix Having a Different Viscosity. *Tectonophysics* **28**(4), 265-274.

Bilby, B. A. & Kolbuszewski, M. L. 1977. Finite Deformation of an Inhomogeneity in 2-Dimensional Slow Viscous Incompressible-Flow. *Proceedings of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences* **355**(1682), 335-353.

Biot, M. A. 1965. *Mechanics of incremental deformations : theory of elasticity and viscoelasticity of initially stressed solids and fluids, including thermodynamic foundations and applications to finite strain*. Wiley, New York.

Bruhn, D. & Burlini, L. 2005. *High-strain zones : structure and physical properties*. Geological Society, London.

Christensen, R. M. 2005. *Mechanics of composite materials*. Dover Publications, Mineola, N.Y.

Dabrowski, M., Krotkiewski, M. & Schmid, D. W. in press. MILAMIN: MATLAB-based FEM solver for large problems. *Geochemistry, Geophysics, and Geosystems* **in press**.

Dawson, P. R. & Wenk, H. R. 2000. Texturing of the upper mantle during convection. *Philosophical Magazine a-Physics of Condensed Matter Structure Defects and Mechanical Properties* **80**(3), 573-598.

Eshelby, J. D. 1957. The Determination of the Elastic Field of an Ellipsoidal Inclusion, and Related Problems. *Proceedings of the Royal Society of London Series a-Mathematical and Physical Sciences* **241**(1226), 376-396.

Fletcher, R. C. 2004. Anisotropic viscosity of a dispersion of aligned elliptical cylindrical clasts in viscous matrix. *Journal of Structural Geology* **26**(11), 1977-1987.

Handy, M. R. 1990. The Solid-State Flow of Polymineralic Rocks. *Journal of Geophysical Research-Solid Earth and Planets* **95**(B6), 8647-8661.

Handy, M. R. 1994. Flow Laws for Rocks Containing 2 Nonlinear Viscous Phases - a Phenomenological Approach. *Journal of Structural Geology* **16**(3), 287-301.

Herwegh, M., de Bresser, J. H. P. & ter Heege, J. H. 2005. Combining natural microstructures with composite flow laws: an improved approach for the extrapolation of lab data to nature. *Journal of Structural Geology* **27**(3), 503-521.

Holyoke, C. W. & Tullis, J. 2006. Mechanisms of weak phase interconnection and the effects of phase strength contrast on fabric development. *Journal of Structural Geology* **28**(4), 621-640.

Ji, S. C., Wang, Z. C. & Wirth, R. 2001. Bulk flow strength of forsterite-enstatite composites as a function of forsterite content. *Tectonophysics* **341**(1-4), 69-93.

Kaus, B. J. P. & Podladchikov, Y. Y. 2006. Initiation of localized shear zones in viscoelastoplastic rocks. *Journal of Geophysical Research-Solid Earth* **111**(B4), -.

Keller, J. B. 1987. Effective Conductivity of Periodic Composites Composed of 2 Very Unequal Conductors. *Journal of Mathematical Physics* **28**(10), 2516-2520.

Linker, M. F., Kirby, S. H., Ord, A. & Christie, J. M. 1984. Effects of Compression Direction on the Plasticity and Rheology of Hydrolytically Weakened Synthetic Quartz Crystals at Atmospheric-Pressure. *Journal of Geophysical Research* **89**(Nb6), 4241-4255.

Madi, K., Forest, S., Cordier, P. & Boussuge, M. 2005. Numerical study of creep in two-phase aggregates with a large rheology contrast: Implications for the lower mantle. *Earth and Planetary Science Letters* **237**(1-2), 223-238.

Mancktelow, N. S. 2002. Finite-element modelling of shear zone development in viscoelastic materials and its implications for localisation of partial melting. *Journal of Structural Geology* **24**(6-7), 1045-1053.

McNamara, A. K., van Keken, P. E. & Karato, S. I. 2002. Development of anisotropic structure in the Earth's lower mantle by solid-state convection. *Nature* **416**(6878), 310-314.

Montesi, L. G. J. 2007. A constitutive model for layer development in shear zones near the brittle-ductile transition. *Geophysical Research Letters* **34**(8), -.

Montesi, L. G. J. & Hirth, G. 2003. Grain size evolution and the rheology of ductile shear zones: from laboratory experiments to postseismic creep. *Earth and Planetary Science Letters* **211**(1-2), 97-110.

Mura, T. 1987. *Micromechanics of defects in solids*. Kluwer Academic Publ., Dordrecht.

Nemat-Nasser, S. & Hori, M. 1993. *Micromechanics : overall properties of heterogeneous materials*. North-Holland, Amsterdam.

Paterson, M. S. 2007. Localization in rate-dependent shearing deformation, with application to torsion testing. *Tectonophysics* **445**(3-4), 273-280.

Pieri, M., Burlini, L., Kunze, K., Stretton, I. & Olgaard, D. L. 2001. Rheological and microstructural evolution of Carrara marble with high shear strain: results from high temperature torsion experiments. *Journal of Structural Geology* **23**(9), 1393-1413.

Regenauer-Lieb, K. & Yuen, D. A. 2003. Modeling shear zones in geological and planetary sciences: solid- and fluid-thermal-mechanical approaches. *Earth-Science Reviews* **63**(3-4), 295-349.

Spengler, D., van Roermund, H. L. M., Drury, M. R., Ottolini, L., Mason, P. R. D. & Davies, G. R. 2006. Deep origin and hot melting of an Archaean orogenic peridotite massif in Norway. *Nature* **440**(7086), 913-917.

Takeda, Y. T. & Griera, A. 2006. Rheological and kinematical responses to flow of two-phase rocks. *Tectonophysics* **427**(1-4), 95-113.

Treagus, S. H. 2003. Viscous anisotropy of two-phase composites, and applications to rocks and structures. *Tectonophysics* **372**(3-4), 121-133.

Treagus, S. H. & Treagus, J. E. 2002. Studies of strain and rheology of conglomerates. *Journal of Structural Geology* **24**(10), 1541-1567.

Willis, J. R. 1964. Anisotropic Elastic Inclusion Problems. *Quarterly Journal of Mechanics and Applied Mathematics* **17**(2), 157-&.

Paper 4: Evolution of Large Amplitude 3D Fold Patterns: a FEM Study

Submitted to Physics of the Earth and Planetary Interiors by

M. Dabrowski, D.W. Schmid, and M. Krotkiewski

Abstract

The study of three dimensional (3D) fold patterns requires numerical models with large numbers of degrees of freedom (≥100'000'000). We have developed BILAMIN, an unstructured (geometry fitted) mesh implementation of the finite element method for incompressible Stokes flow that is capable of solving such systems. All repetitive and computationally intensive steps are fully parallelized. One of the main components is the iterative solver. We chose the minimum residual method (MINRES) because it allows operating directly on the indefinite systems resulting from the incompressibility condition. We use BILAMIN in a case study of fold pattern evolution. Folds are ubiquitous in nature, and contain both mechanical and kinematic information that can be deciphered with appropriate tools. Our results show that there is a relationship between fold aspect ratio and in-plane loading conditions. We propose that this finding can be used to determine the complete parameter set potentially contained in the geometry of three dimensional folds: mechanical properties of natural rocks, maximum strain, and relative strength of the in-plane far-field load components. Furthermore, we show how folds in 3D amplify and that there is a second deformation mode, besides continuous amplification, where compression leads to a lateral rearrangement of blocks of folds. Finally, we demonstrate that the textbook prediction of dome and basin (egg carton) structures resulting from folding instabilities in constriction is incorrect. The fold patterns resulting in this setting are curved, elongated folds with random orientation.

Introduction

Folds on all scales form millimeters to kilometers may be the result of the mechanical instability that arises when a mechanically stratified system is subjected to layer-parallel compression. While the resulting fold patterns are three dimensional, their geometries are often simplified by assuming that there is no shape variation in the third dimension. This facilitates the analysis and has resulted in a large number of studies that investigate the folding instability for a variety of rheologies: viscous (e.g., Biot, 1961; Fletcher, 1977; Smith, 1975), visco-elastic (Biot, 1961;

Muhlhaus et al., 2002b; Schmalholz and Podladchikov, 1999), visco-elasto-plastic (Gerbault et al., 1999), power-law (Fletcher, 1974; Smith, 1977), and anisotropic (Biot, 1965; Fletcher, 2005; Kocher et al., 2006; Muhlhaus et al., 2002a). Studies of three dimensional folding have mostly focused on kinematic models and resulted in the classification of fold interference patterns caused by the superposition of folds of different generations (e.g., Grasemann et al., 2004; Ramsay, 1967; Thiessen and Means, 1980). Mechanical studies of viscous single layer 3D folding were performed by Fletcher (1991; 1995) who derived the dispersion relationships for the 3D folding instability in viscous materials using the thick-plate analysis. These small strain results were elaborated on by Kaus and Schmalholz (2006) who used the finite element method to achieve large strains and developed an expression for the finite amplitude development in analogy with the results that were obtained by Schmalholz and Podladchikov (2000). While their 3D folding study nicely illustrates the finite strain evolution of a few folds, it cannot illustrate the natural patterns that emerge out of randomly perturbed layers where many folds interact with each other. Furthermore, in these relatively small models the boundary conditions seriously affect the results. The employed free slip boundary conditions allow for large strain fold development even at the boundaries. However, the free slip condition corresponds to vertical planes of symmetry, which are either axial planes or cross-sections through quasi-cylindrical folds. Hence, these boundaries, although relatively "soft", force the locations of the folds. If fold patterns out of random noise are to be studied, this effect must be considered and the modeled area must be large relative to the one that is affected by this type of boundary effect, which is what we do in this study. Our focus in this paper is to show and quantify the pattern evolution of 3D folding as a function of different loading conditions. Furthermore we propose that the aspect ratio of folds in map view may be used to infer the relative strength of the two principal in-plane loads.

Numerical Code: Model, Implementation, Performance, and Benchmark

The mechanical model we solve for is incompressible Stokes flow for Newtonian fluids in the absence of gravity

$$\begin{aligned} \dot{\varepsilon}_{ij} &= \frac{1}{2}\left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right) \\ \dot{\varepsilon}_{ii} &= 0 \\ \sigma_{ij} &= -p + 2\eta\dot{\varepsilon}_{ij} \\ \frac{\partial \sigma_{ij}}{\partial x_j} &= 0 \end{aligned} \tag{1}$$

where $i = 1,2,3$ represent the three spatial directions, x_i are the coordinates, v_i are the velocities, $\dot{\varepsilon}_{ij}$ is the strain rate tensor, σ_{ij} is the stress tensor, $p = -\sigma_{ii}/3$ is the pressure, and η is the viscosity. The governing equations are solved with BILAMIN, a self developed Lagrangian finite element code that employs a mixed velocity-pressure approach (Brezzi and Fortin, 1991). The element used is a 15-node Crouzeix-Raviart tetrahedron with discontinues linear shape functions for pressure and pseudo cubic continuous velocities, see e.g. Bertrand et al. (1992). In order to solve large systems of equations that arise in modeling of 3D problems, it is crucial to employ an iterative method, as memory and time requirements of direct solvers become unrealistic for large systems. The penalty method approach, although robust and accurate for large enough penalty parameter values (e.g., Powell and Hestenes iterations, Cuvelier et al., 1986), is too costly if used together with iterative methods, as the condition number of the global system of equations deteriorates. In order to alleviate this limitation, incomplete Uzawa type of the methods are usually utilized; for discussion see e.g. Bertrand and Tanguy (2002). The idea is to perform conjugate gradient (CG) iterations that operate on the symmetric positive-definite pressure Schur complement of the global system. The pressure Schur complement involves the inverse of the velocity part of the global matrix. However, the action of this operator is applied only inexactly through several inner CG iterations. In our code we follow a different approach and employ the minimum residual iterative method (MINRES, Paige and Saunders, 1975). Thus, we directly operate on the symmetric indefinite system resulting from the discretization of incompressible Stokes flow equations (Elman et al., 2006; Wathen et al., 1995). In order to accelerate the convergence we use a block preconditioning approach for this saddle point problem (Elman, 2002; Silvester and Wathen, 1994).

The unstructured, geometry-adapted meshes are generated by T3D, a tetrahedron mesh generator developed by D. Rypl (2007). Mesh generation and node reordering are both sequential, and are treated as a pre-processing stage of the solver that must only be performed every time a new mesh is used. The matrix assembly is implemented using a symbolic approach to sparse matrices, where a symbolic matrix representation (the non-zero pattern) resulting from the mesh topology is created. Our implementation is memory efficient, as we do not explicitly create the connectivity arrays, but directly the non-zero structure. Symbolic structure preparation, matrix computation, and matrix assembly are fully parallel in terms of computations and memory usage. Communication between the processes is implemented over the message passing interface (MPI). The system of linear equations is explicitly constructed in order to speed up the iterative solver. The basic parallel building blocks of the solver are dot product and sparse matrix by vector multiplication, the latter being a challenge from the point of view of per-CPU performance. Reverse Cuthill-McKee (RCM) reordering is used in order to gain best per-CPU cache reuse and at the same time decent inter-CPU data exchange during sparse matrix - vector product.

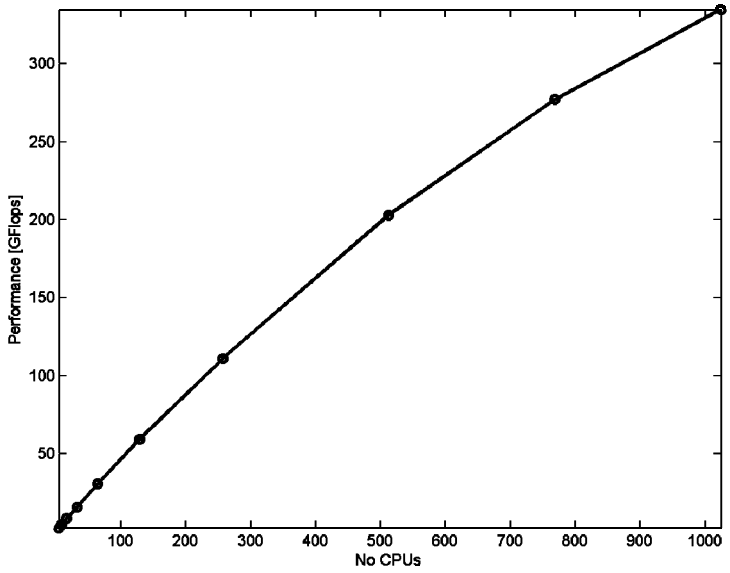


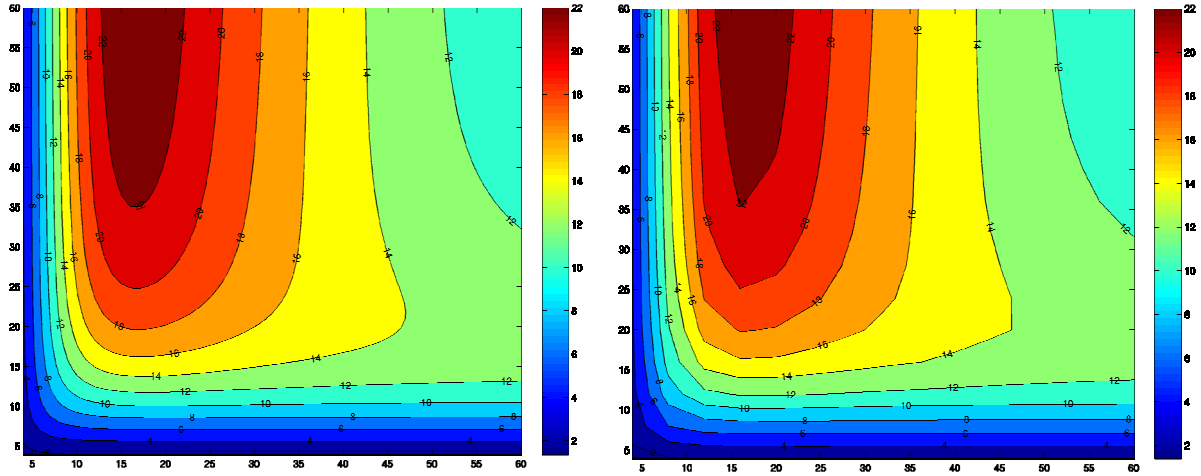
Figure 1

Performance results in GFLOPS obtained by the BILAMIN implementation of the MINRES solver for systems with approximately 3'000'000 degrees of freedom per CPU. Code benchmarks were performed on an InfiniBand connected Opteron cluster located at the University of Minnesota, Minneapolis (1192 CPUs). Each node of the cluster consists of two dual-core CPUs and 8 GB of DDR2/666MHz RAM.

We have tested BILAMIN on large clusters and successfully solved systems with more than 100'000'000 degrees of freedom and meshes consisting of 26'000'000 nodes (6'000'000 elements). See Figure 1 for performance scaling. The efficiency on 1000 CPUs is around 65% starting from a close to optimal sequential implementation.

The model studied in this paper consists of a viscous layer overlying a less viscous matrix (viscosity ratio 50:1). The top boundary is a free surface; all other boundaries are free slip boundaries, which are specified in terms of normal velocities that result in constant strain rates. The vertical velocity of the bottom is zero, while the normal velocities of the lateral walls are varied to systematically test the influence of different compressive regimes. The thickness of the matrix relative to the layer thickness is twice the analytically predicted dominant wavelength and the perturbation velocities do not feel the restriction due to the bottom boundary (infinite half space, cf Ramberg, 1961). The accuracy of the numerical code and the chosen model configuration are shown in Figure 2. With BILAMIN it is possible to reproduce the analytical predictions with less than 2% error even at low resolutions with 19x7x7 elements.

While the results in Figure 2 are obtained by looping through single sinusoid configurations of varying wavelength to thickness ratio, the actual pattern study runs were performed with randomly perturbed layers. The red noise used is scaled so that the amplitude is maximally 1/50 of the layer thickness (routine modified after Yearsley, 2004).



a) b)

Figure 2

Comparison of analytical (a) versus numerical (b) growth rate spectrum for an embedded layer with viscosity ratio 100:1 between layer and matrix.

Results

An illustration of the 3D folding results is shown in Figure 3. The mechanical instability results in finite amplitude folds of limited length that interact with each other. A detailed comparison of fold patterns resulting from different loading conditions is shown in Figure 4. a) shows the initial random perturbation of the model with the corresponding spectra. These spectra are averages of the spectra in the x and y direction. b) to e) show the results of different loading conditions, symbolized by arrows, at an identical shortening in the x direction of 30%.

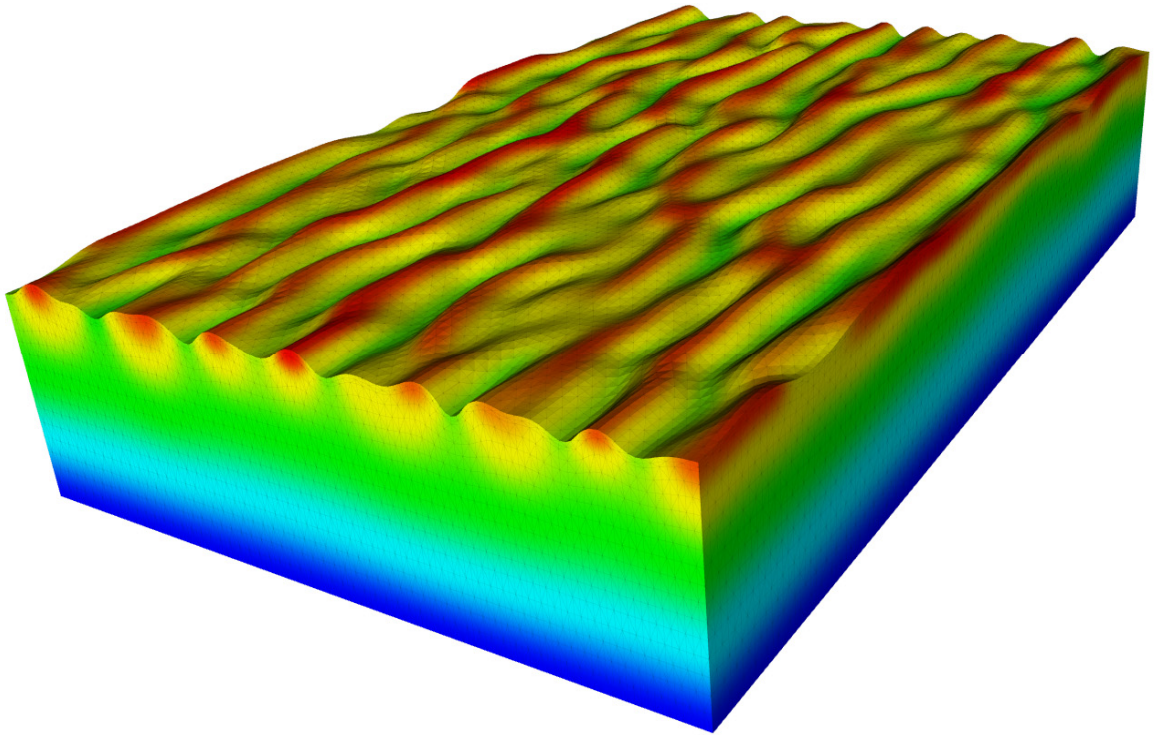


Figure 3

3D illustration of a $\dot{\varepsilon}_{xx} = -1.00, \dot{\varepsilon}_{yy} = 0.00$ *run.*

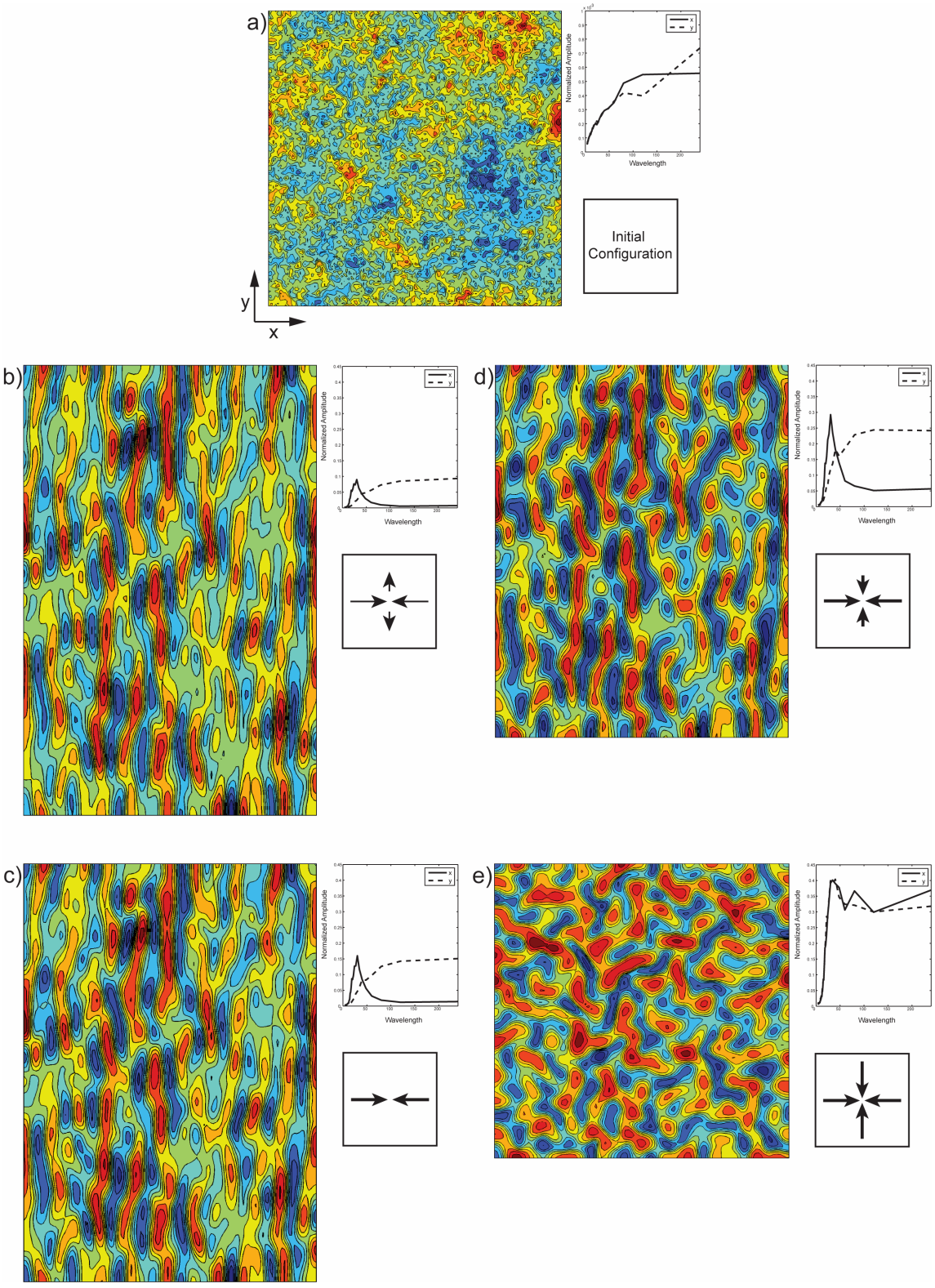


Figure 4

Fold patterns developed out of the same initial perturbation (a) for different loading conditions (b-d) given for the same shortening in x-direction (30%). Spectra show averaged 1D Fourier analyses in x and y direction. b) $\dot{\varepsilon}_{xx} = -1.25, \dot{\varepsilon}_{yy} = +0.25$, *c)* $\dot{\varepsilon}_{xx} = -1.00, \dot{\varepsilon}_{yy} = 0.00$, *d)* $\dot{\varepsilon}_{xx} = -0.75, \dot{\varepsilon}_{yy} = -0.25$, *e)* $\dot{\varepsilon}_{xx} = -0.50, \dot{\varepsilon}_{yy} = -0.50$.

Analysis of the patterns confirms the finding of Fletcher (1995) that most folds ought to be cylindrical. As long as there is a preferential compression direction (b-c) the resulting folds have elongated shapes that are oriented approximately orthogonal to the maximum compression direction. However, using the plane strain experiment (c), corresponding to the typically employed simplified 2D models, as an example, it is obvious that the folds are clearly elongated, but are far from being perfectly cylindrical and seem to have a specific aspect ratio (fold length/width). The only exception to the finding that folds are elongated and oriented orthogonal to the maximum compression direction is shown in e) where actually no unique maximum compression direction exists, as the normal loading in all in-plane directions is identical. The textbook (e.g., Ramsay, 1967; Twiss and Moores, 1992) prediction for this loading is the development of dome and basins, resembling an egg carton, that often would be interpreted as the superposition of two sequential folding events. No distinction of deformation events is required to interpret the results in Figure 4e, as it certainly is one event. However, a regular egg carton pattern cannot be observed; the folds show clear elongations, but no preferred orientation. The fold patterns in all the presented finite strain models resemble each other, especially the ones where a principal compression direction exists. The pattern relation with respect to the initial perturbation is less obvious and demonstrates the selectivity of the folding instability, which amplifies out of the initial perturbation only the components that exhibit the fastest growth rate. Note though that the top and bottom interface have different initial perturbations. This should be taken into account when analyzing the fold pattern evolution with respect to the initial perturbation.

The comparison of the initial and the finite strain spectra also shows how the folding instability amplifies the fastest growing components (folds). The finite strain spectra are plotted with identical vertical axes and reveal that for the same given shortening in the x-direction the fold amplitudes are furthest developed in e), which is because here the shortening in the y-direction is strongest and contributes most to the fold development. A comparison of the amplitude evolution is shown in Figure 5, which reproduces the findings of Schmalholz and Podladchikov (2000). Initially the folds amplify exponentially, as predicted by the classical theories. However, with increasing strain the growth rates slow down and are controlled by the layer length. For the model in Figure 4e this transition is reached at a shortening of 20%, for the other models the transition occurs later in terms of maximum shortening because the relative contribution from the y-direction is less. This finding is also a confirmation of the finite amplitude theory of Kaus and Schmalholz (2006) for large random systems.

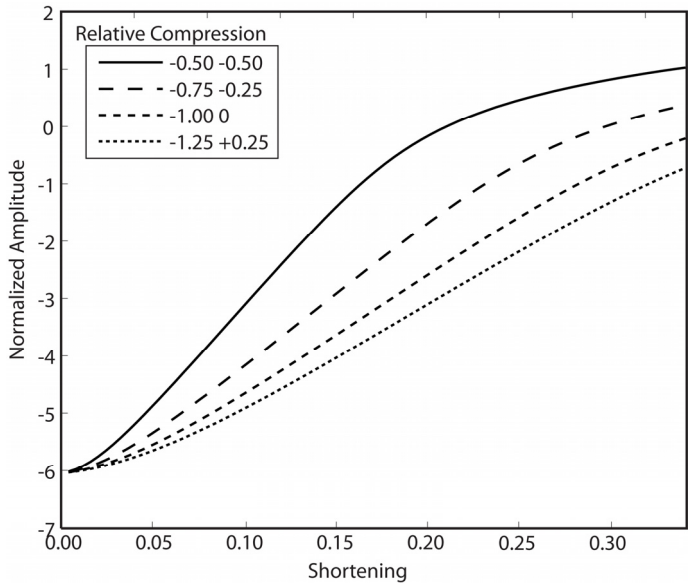


Figure 5

Normalized amplitude evolution for the different loading conditions.

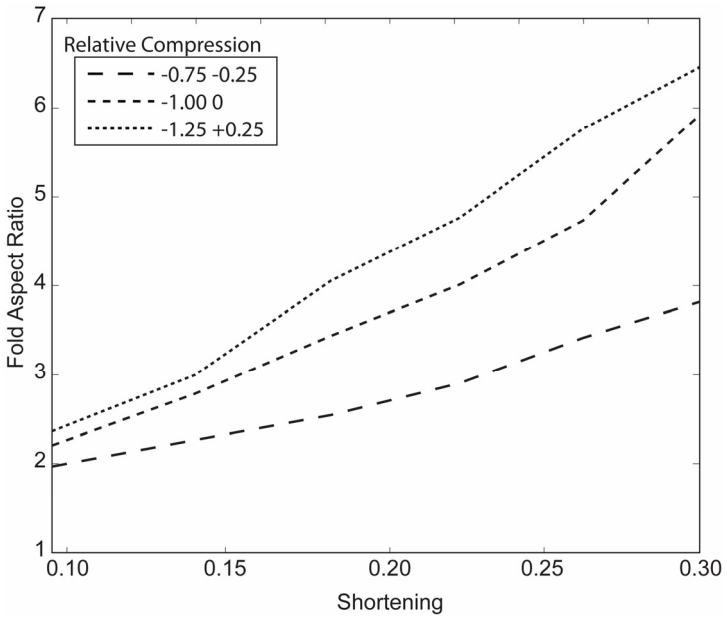


Figure 6

Fold aspect ratio as a function of maximum shortening and loading condition.

The averaged spectra in the y-direction presented in Figure 4b)-Figure 4d) show no selectivity within the size of the modeled system. This was predicted by the thick plate analysis of Fletcher (1995). Yet, the folds in this runs show distinct aspect ratios. These aspect ratios are a function of the maximum amount of shortening, as well as the loading condition, see Figure 6. They are determined by finding the absolute deviation of the top model surface from its aver-

age value, contouring this field at the intermediate level, and calculating the average aspect ratio of the resulting closed contours.

The finding of a relationship between loading and folds aspect ratio is potentially very useful. 2D, cross-section based analyses with all possible corrections such as layer thickening (Sherwin and Chapple, 1968) can be used to infer the mechanical rock properties, as well as the strain in the maximum shortening direction (e.g. Schmalholz and Podladchikov, 2001; Schmalholz et al., 2002). For a complete 3D analysis only the relative strength of the maximum to the minor in-plane compression remains to be determined, which can be achieved with a diagram such as shown in Figure 6. Since the total maximum strain as well as the viscosity ratio are known, only the fold aspect ratio must be determined and a unique result for the in-plane loading conditions can be obtained. Since the relationships shown in Figure 6 can, at the moment, not be predicted analytically, a large number of numerical runs is required in order to produce these relationships for different mechanical properties and ranges. Furthermore, the robustness of the relationship must be investigated, especially with respect to the initial perturbation of the strong layer and possibly with respect to numerical parameters such as resolution.

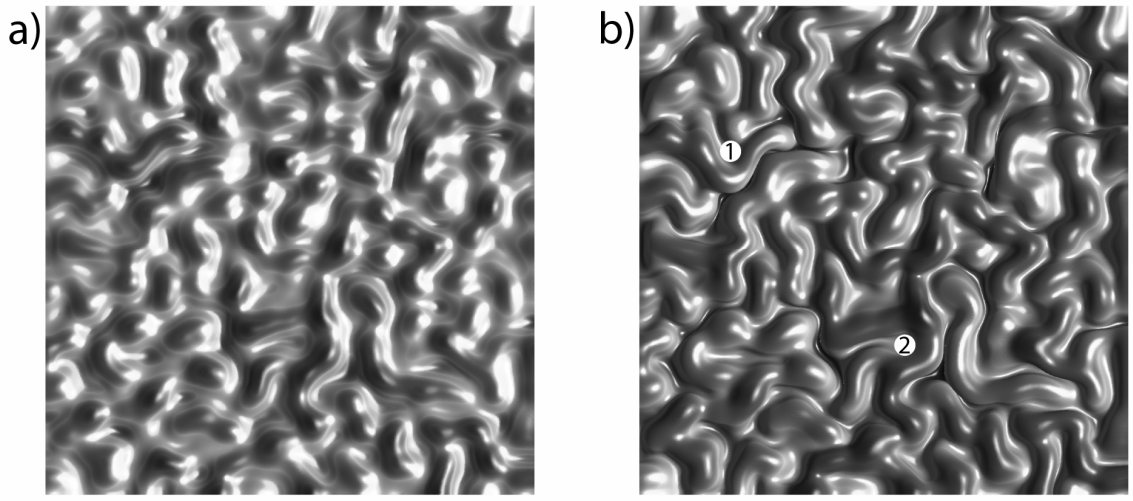


Figure 7
Fold pattern evolution in the -0.50 -0.50 run at 25% (a) and 40% (b) shortening. Domains are scaled to have the same size in order to facilitate comparison. Remeshing was required for b).

The fold amplitude evolutions shown in this paper bear many of the characteristics of 2D models and the results from 2D analysis are mostly applicable to 3D, as discussed above. However, in 2D models folds become mechanically ineffective once the fold limbs are close to isoclinal. In such large strain cases the folded layer will start to behave as an effective layer with the fold height as the layer thickness and a second generation of larger (absolute) wavelength folds will

grow. In large strain 3D models the folded layer has more possibilities, as shown in Figure 7. Note that the shown 40% shortening would not be considered extreme in a 2D model. However, for the given 3D model even overhanging fold shapes result (fan folds, interlimb angle <0º). A comparison of Figure 7a) and b) shows that the progressive strain leads to fold amplification and merging of neighboring fold hinges. However, the folds also have the possibility to migrate laterally. This effect can be seen in the regions marked 1 and 2 in Figure 7b but also elsewhere in this model. Kaus and Schmalholz (2006) discuss how the folding instability leads to a softening of the system. While this conclusion can be drawn for the entire system, it may be that the observation of migrating folds is an indication that at large strains folds become structurally stiff and rather move with respect to each other than amplify further. Finally we would like to point out that also at the largest strains we have calculated (Figure 7b) the fold patterns have no resemblance with egg carton structures, but are densely packed, curved, elongated folds with random orientations.

Conclusions

We have used our implementation of the finite element method for incompressible Stokes flow to study the evolution of fold patterns resulting from large, randomly perturbed domains. Large physical model sizes are required in order to obtain patterns with many (dominant) fold wavelengths that allow comparison with natural fold patterns. This causes the use of numerical models with resolutions of approximately 100’000’000 degrees of freedom, which can only be solved with massively parallel iterative solution strategies. MINRES is a particularly useful choice for this because it allows operating directly on indefinite symmetric systems resulting from the discretization of eqn. (1). Furthermore, the implementation must be efficient so that a large number of time steps can be calculated in order to achieve large strains. The latter also requires the possibility of performing remeshing steps. All of these criteria are met by BILAMIN, the code we used in this paper.

The study of the fold patterns confirm the theory developed by Fletcher (1995), which states that there is no wavelength selectivity in the direction orthogonal to the maximum compression direction. This can be seen in the spectral analysis of the fold patterns. Despite this the folds can be characterized in the physical domain with an aspect ratio that depends on strain and relative strength of the in-plane far-field loading. This observation, combined with previously developed theories (e.g., Kaus and Schmalholz, 2006; Schmalholz and Podladchikov, 2001) could be used to determine complete parameter set potentially contained in the geometry of three dimensional folds: mechanical properties of natural rocks, maximum strain, and rela-

tive strength of the in-plane far-field loading conditions. While the finiteness of the folds is confirmed by patterns in natural rocks, it is unclear how robust the relationship between fold aspect ratio and maximum strain and loading conditions is. Further investigations should focus on analyzing a broad class of mechanical stratifications and verify the influence of the initial perturbation.

The model runs show that the same initial perturbation will be selectively amplified in a similar fashion by a variety of different loading conditions. As long as there is a unique maximum compression direction the fold patterns are similar. Given that in 3D models both principal in-plane directions may contribute to the fold amplification, large amplitude folds can be developed substantially faster in constrictional settings compared to the 2D plane strain simplifications. This also causes a faster transition from active (exponential) folding to layer controlled growth (Schmalholz and Podladchikov, 2000). Another difference between 2D and 3D models is that in 3D folds may react to large compression not only by fold amplification, but also by lateral rearrangement of blocks of folds. The fold patterns resulting from constrictional experiments with randomly perturbed layers do not result in egg carton shaped dome and basins such as proposed in textbooks; curved, elongated fold shapes are formed instead. The interpretation of natural fold patterns, and the identification of interference patterns and several stages of deformation must be performed with care. The complexity of fold patterns resulting from our single deformation event models should be kept in mind.

Acknowledgements

We would like to thank Daniel Rypl for making T3D, the mesh generator used in this paper, available. We would also like to thank the Minnesota Supercomputing Institute (MSI) and the Norwegian High Performance Computing (NOTUR) network to grant us machine access for development and production.

References

Bertrand, F. and Tanguy, P.A., 2002. Krylov-based Uzawa algorithms for the solution of the Stokes equations using discontinuous-pressure tetrahedral finite elements. Journal of Computational Physics, 181(2): 617-638.

Bertrand, F.H., Gadbois, M.R. and Tanguy, P.A., 1992. Tetrahedral Elements for Fluid-Flow. International Journal for Numerical Methods in Engineering, 33(6): 1251-1267.

Biot, M.A., 1961. Theory of folding of stratified viscoelastic media and its implications in tectonics and orogenesis. Geological Society of America Bulletin, 72(11): 1595-1620.

Biot, M.A., 1965. Mechanics of Incremental Deformations. John Wiley & Sons, Inc., New York.

Brezzi, F. and Fortin, M., 1991. Mixed and hybrid finite element methods. Springer, New York etc., IX, 350 pp.

Cuvelier, C., Segal, A. and Steenhoven, A.A., 1986. Finite element methods and Navier-Stokes equations. Reidel, Dordrecht a. o., XVI, 483 pp.

Elman, H.C., 2002. Preconditioners for saddle point problems arising in computational fluid dynamics. Applied Numerical Mathematics, 43(1-2): 75-89.

Elman, H.C., Silvester, D.J. and Wathen, A.J., 2006. Finite elements and fast iterative solvers with applications in incompressible fluid dynamics. Oxford University Press, New York, 400 pp.

Fletcher, R.C., 1974. Wavelength selection in the folding of a single layer with power-law rheology. Am. J. Sci., 274(11): 1029-1043.

Fletcher, R.C., 1977. Folding of a Single Viscous Layer - Exact Infinitesimal Amplitude Solution. Tectonophysics, 39(4): 593-606.

Fletcher, R.C., 1991. 3-Dimensional Folding of an Embedded Viscous Layer in Pure Shear. Journal of Structural Geology, 13(1): 87-96.

Fletcher, R.C., 1995. 3-Dimensional Folding and Necking of a Power-Law Layer - Are Folds Cylindrical, and, If So, Do We Understand Why. Tectonophysics, 247(1-4): 65-83.

Fletcher, R.C., 2005. Instability of an anisotropic power-law fluid in a basic state of plane flow. Journal of Structural Geology, 27(7): 1155-1167.

Gerbault, M., Burov, E.B., Poliakov, A.N.B. and Daignieres, M., 1999. Do faults trigger folding of the lithosphere? Geo. Res. Lett., 26(2): 271-274.

Grasemann, B., Wiesmayr, G., Draganits, E. and Fusseis, F., 2004. Classification of refold structures. Journal of Geology, 112(1): 119-125.

Kaus, B.J.P. and Schmalholz, S.M., 2006. 3D finite amplitude folding: Implications for stress evolution during crustal and lithospheric deformation. Geophysical Research Letters, 33(14): -.

Kocher, T., Schmalholz, S.M. and Mancktelow, N.S., 2006. Impact of mechanical anisotropy and power-law rheology on single layer folding. Tectonophysics, 421(1-2): 71-87.

Muhlhaus, H.B., Dufour, F., Moresi, L. and Hobbs, B., 2002a. A director theory for visco-elastic folding instabilities in multilayered rock. International Journal Of Solids And Structures, 39(13-14): 3675-3691.

Muhlhaus, H.B., Moresi, L., Hobbs, B. and Dufour, F., 2002b. Large amplitude folding in finely layered viscoelastic rock structures. Pure And Applied Geophysics, 159(10): 2311-2333.

Paige, C.C. and Saunders, M.A., 1975. Solution of Sparse Indefinite Systems of Linear Equations. Siam Journal on Numerical Analysis, 12(4): 617-629.

Ramberg, H., 1961. Contact strain and folding instability of a multilayered body under compression. Geol. Rdsch., 51: 405-439.

Ramsay, J.G., 1967. Folding and fracturing of rocks. International series in the earth and planetary sciences. McGraw-Hill, New York, 568 pp.

Rypl, D., 2007. T3D - Tetrahedron mesh generator.

Schmalholz, S.M. and Podladchikov, Y., 1999. Buckling versus folding: Importance of viscoelasticity. Geophysical Research Letters, 26(17): 2641-2644.

Schmalholz, S.M. and Podladchikov, Y.Y., 2000. Finite amplitude folding: transition from exponential to layer length controlled growth (vol 179, pg 363, 2000). Earth and Planetary Science Letters, 181(4): 619-633.

Schmalholz, S.M. and Podladchikov, Y.Y., 2001. Strain and competence contrast estimation from fold shape. Tectonophysics, 340(3-4): 195-213.

Schmalholz, S.M., Podladchikov, Y.Y. and Burg, J.-P., 2002. Control of folding by gravity and matrix thickness: Implications for large-scale folding. Journal of Geophysical Research B: Solid Earth, 107(B1): 10.1029/2001JB000355.

Sherwin, J.A. and Chapple, W.M., 1968. Wavelengths of Single Layer Folds - a Comparison between Theory and Observation. American Journal of Science, 266(3): 167-179.

Silvester, D. and Wathen, A., 1994. Fast Iterative Solution of Stabilized Stokes Systems Part .2. Using General Block Preconditioners. Siam Journal on Numerical Analysis, 31(5): 1352-1367.

Smith, R.B., 1975. Unified Theory of Onset of Folding, Boudinage, and Mullion Structure. Geological Society of America Bulletin, 86(11): 1601-1609.

Smith, R.B., 1977. Formation of Folds, Boudinage, and Mullions in Non-Newtonian Materials. Geological Society of America Bulletin, 88(2): 312-320.

Thiessen, R.L. and Means, W.D., 1980. Classification of Fold Interference Patterns - a Reexamination. Journal of Structural Geology, 2(3): 311-316.

Twiss, R.J. and Moores, E.M., 1992. Structural Geology. W. H. Freeman and Company, New York.

Wathen, A., Fischer, B. and Silvester, D., 1995. The Convergence Rate of the Minimal Residual Method for the Stokes Problem. Numerische Mathematik, 71(1): 121-134.

Yearsley, J., 2004. spatialPattern. MATLAB File Exchange.

Paper 5: Fractional Steps Methods for Transient Problems on Commodity Computer Architectures

Submitted to Physics of the Earth and Planetary Interiors by

M. Krotkiewski, M. Dabrowski, and Yu.Yu.Podladchikov

Abstract

Fractional Steps methods are suitable for modeling transient processes that are central to many geological applications. Low memory requirements and modest computational complexity facilitates calculations on high-resolution three dimensional models. An efficient implementation of Alternating Direction Implicit/Locally One Dimensional schemes for an Opteron-based shared memory system is presented. The memory bandwidth usage, the main bottleneck on modern computer architectures, is specially addressed. High efficiency of above 2 GFlops per CPU is sustained for problems of 1 billion degrees of freedom. The optimized sequential implementation of 1D sweeps is comparable in execution time to copying the data in the memory. Scalability of the parallel implementation on up to 8 CPUs is close to perfect. Performing one timestep of the Locally One Dimensional scheme on a system of 1000^3 unknowns takes only 11 seconds. We present an application of the methods to parabolic and hyperbolic problems using high-resolution, real-world oil reservoir data.

Introduction

Geological systems are usually heterogeneous and exhibit large material property contrasts. They are often formed by multi-physics processes interacting on many temporal and spatial scales. In order to understand these systems numerical models are frequently employed. Appropriate resolution of the behavior of these heterogeneous systems, without the (over-) simplifications of a priori applied homogenization techniques, requires numerical models capable of efficiently and accurately dealing with high resolution models.

A popular technique is the finite element method (FEM) combined with unstructured meshes capable of dealing with the geometrical complexities of geological problems. In these methods a linear system of equations is assembled and solved. While it can be successfully done for two dimensional models with high resolution even on a modern desktop computer, three dimensional problems require supercomputers and sophisticated numerical methods. Direct solvers

are unfeasible due to enormous memory requirements and large computational times. Iterative solvers (like Conjugate Gradients) are an available alternative, but require good preconditioners that can be efficiently parallelized.

Methods based on structured meshes, although less accurate in terms of geometry representation, are often employed. The known structure of the mesh makes them cheaper in terms of memory requirements, and can significantly decrease the computational cost. Powerful methods like geometrical multigrid exist for structured meshes, however available implementations are far from theoretical CPU performance (Adams et al., 2003).

Classical explicit methods are usually impractical for high-resolution problems due to severe timestep restriction. Operator splitting techniques try to overcome this limitation. The general idea is to divide a single timestep into a sequence one-dimensional implicit sweeps through the domain. The computational cost of such schemes is comparable to explicit methods, but the timestep restriction can be avoided.

We first present the class of *Fractional Steps* methods for transient parabolic and hyperbolic problems. The suitability of contemporary shared memory, Opteron-based commodity architecture for this approach is investigated. We focus on high resolution problems with up to 1000^3 degrees of freedom and heterogeneous material properties. An optimized algorithm for efficient computation and solution of tridiagonal systems of linear equations on an 8 Dual-core Opteron machine is presented.

Fractional step method

Consider an initial-boundary value parabolic problem of the heat conduction

$$\begin{aligned} \rho c_p \frac{\partial \varphi}{\partial t} &= \mathbf{div}(k \cdot \mathbf{grad}\,\varphi) + f \quad \text{in } \Omega \times T \\ \varphi &= \varphi_\Gamma \quad \text{on } \partial\Omega \times T \\ \varphi &= \varphi_0 \quad \text{in } \Omega \quad \text{for } t = 0 \end{aligned} \tag{1}$$

where φ denotes temperature, k is thermal conductivity, f is heat generation term, and ρc_p is a product of density and specific heat capacity. Thermal field φ_Γ is prescribed on the boundary $\partial\Omega \times T$ and initial conditions φ_0 are given in the whole spatial domain Ω.

In the case of three-dimensional homogeneous media, the finite difference (FD) discretization A of the operator $(\rho c_p)^{-1}\mathbf{div}(k \cdot \mathbf{grad}\ldots)$ on the uniform Cartesian grid yields

$$A = A_x + A_y + A_z$$
$$A_x = \frac{\kappa}{(\Delta h)^2}(\Delta_x \nabla_x) \tag{2}$$

where $\kappa = k / \rho c_p$ is thermal diffusivity, $\Delta h = \Delta x = \Delta y = \Delta z$ is a grid spacing and $(\Delta_x T)_i = T_i - T_{i-1}, (\nabla_x T)_i = T_{i+1} - T_i$ denote backward and forward difference operator in the *x* direction, respectively. The operators A_y, A_z are analogous to A_x and A is the standard 7-point stencil for the Laplacian scaled by $\kappa/(\Delta h)^2$. For heterogeneous materials and non-uniform grids we use the finite volume spatial discretization that in one dimensional case is stated as

$$\frac{\Delta x_{i+1/2} + \Delta x_{i-1/2}}{2}(\rho c_p)_i \frac{\partial \varphi_i}{\partial t} =$$
$$= \frac{k_{i-1/2}}{\Delta x_{i-1/2}}\varphi_{i-1} - \frac{(\Delta x_{i-1/2} k_{i+1/2} + \Delta x_{i+1/2} k_{i-1/2})}{\Delta x_{i-1/2}\Delta x_{i+1/2}}\varphi_i + \frac{k_{i+1/2}}{\Delta x_{i+1/2}}\varphi_{i+1} + \frac{\Delta x_{i+1/2} + \Delta x_{i-1/2}}{2} f_i \tag{3}$$

where fractional indices correspond to the center of the edge between two neighboring points in the *x* direction. Since in our approach the conductivity *k* is defined in the centers of the 3D cells, it has to be averaged in that mid-edge point. The simplest approach is the arithmetical average

$$k_{i-1/2} = \frac{1}{4(\Delta y_{j-1/2} + \Delta y_{j+1/2})(\Delta z_{k-1/2} + \Delta z_{k+1/2})}\begin{pmatrix} k_{i-1/2,j-1/2,k-1/2}\Delta y_{j-1/2}\Delta z_{k-1/2} + k_{i-1/2,j+1/2,k-1/2}\Delta y_{j+1/2}\Delta z_{k-1/2} + \\ + k_{i-1/2,j-1/2,k+1/2}\Delta y_{j-1/2}\Delta z_{k+1/2} + k_{i-1/2,j+1/2,k+1/2}\Delta y_{j+1/2}\Delta z_{k+1/2} \end{pmatrix} \tag{4}$$

and $k_{i+1/2}$ is computed analogously.

The classical explicit scheme used to integrate (1)

$$\frac{\varphi^{i+1} - \varphi^i}{\tau} = A\varphi^i + f \tag{5}$$

is second order accurate in spatial coordinates, first order in time and is stable under the restriction $\tau \le \frac{(\Delta h)^2}{6\kappa}$ (Courant et al., 1967). The maximum admissible integration step τ^{cr} becomes very small for refined grids, and in the case of non-uniform grid spacing it is determined by the size of the smallest cell. Moreover, in heterogeneous materials τ^{cr} is restricted by the strongest heterogeneity, even if it is insignificant in size. The computational complexity of explicit methods per integration step is small, but due to the timestep restriction they may require a large number of iterations to integrate the evolution of the system.

In order to alleviate the timestep restriction an implicit approach can be used. The second order accurate in time Crank-Nicholson scheme

$$\frac{\varphi^{i+1}-\varphi^{i}}{\tau}=\frac{1}{2}A(\varphi^{i+1}+\varphi^{i})+f \tag{6}$$

is unconditionally stable (no restriction on τ). It requires solving a system of linear equations

$$(\mathrm{I}-\frac{\tau}{2}A)\varphi^{i+1}=(\mathrm{I}+\frac{\tau}{2}A)\varphi^{i}+\tau f \tag{7}$$

which is symmetric and sparse, i.e. most of the matrix entries are zeros. For 1D problems the resulting system matrix is tridiagonal and can be easily solved using the *Thomas Algorithm*. In the 2D case it is practical to use variants of the well known Guassian elimination, e.g. the Cholesky factorization for symmetric positive definite systems, LU factorization with pivoting for non-symmetric systems, or a generalized Thomas Algorithm for block tridiagonal matrices. Sparse direct solvers are easy to use, robust, and very efficient on modern computers. Unfortunately, they can not be applied to large 3D problems because of extreme memory requirements and computational complexity. The number of new non-zero entries introduced during the factorization is much higher than in the 2D case. Various iterative methods, like Conjugate Gradients usually used with some preconditioner, can be used instead. However, for strongly heterogeneous problems the convergence rate deteriorates. Moreover, preconditioners are problem dependent and finding a good, parallelizable one is often difficult.

The fractional step methods described in this paper combine the advantages of the two mentioned strategies for time integration of system (1): low computational cost of the explicit scheme, and the stability of the implicit approach. The general idea is to replace a complex operator A by simpler ones that are used in sequence (fractional steps) during integration of the parabolic system like (1). In the context of the heat diffusion operator the split is naturally dictated by the spatial components A_x, A_y, A_z. During every fractional step many one dimensional, and thus tridiagonal systems of equations have to be solved.

Alternating Direction Implicit schemes

The Alternating Direction Implicit (ADI) scheme for two dimensional parabolic problems was proposed by Peaceman and Rachford (1955) and Douglas (1955)

$$\begin{aligned}\frac{\varphi^{j+1/2}-\varphi^{j}}{\tau}&=\frac{1}{2}(A_x\varphi^{j+1/2}+A_y\varphi^{j})\\ \frac{\varphi^{j+1}-\varphi^{j+1/2}}{\tau}&=\frac{1}{2}(A_x\varphi^{j+1/2}+A_y\varphi^{j+1})\end{aligned} \tag{8}$$

In order to define the stability, we introduce the difference step operator $C(\tau,\Delta h)$ that is defined as the action of the scheme in the whole step

$$\varphi^{i+1}=C(\tau,\Delta h)\varphi^{i} \tag{9}$$

The stability of the scheme requires that $\|C(\tau,\Delta h)\|\le 1$. The algorithm (8) is unconditionally stable, i.e. it is stable for any $\tau\ge 0$. In addition to the truncation error of the Crank-Nicholson implicit scheme (6), the error term related to the splitting is $O(\tau^2)$. Thus, ADI approximates the original problem with the same order of accuracy. However, the additional error term may be large and the improvements to the original scheme were suggested, e.g. (Douglas and Kim, 2001).

The 2D ADI scheme is also applicable as an iterative solver for the steady-state variant of the problem (1). In whole steps, using two-layer difference scheme notation (Janenko, 1971) it is given by

$$\begin{aligned}&\frac{\varphi^{i+1}-\varphi^{i}}{\tau}=\Omega_1\varphi^{i+1}+\Omega_2\varphi^{i}\\ &\Omega_1=\frac{A_x+A_y}{2}-\frac{\tau}{4}A_xA_y\\ &\Omega_2=\frac{A_x+A_y}{2}+\frac{\tau}{4}A_xA_y\end{aligned} \tag{10}$$

It is clearly seen that the relation $A=\Omega_1+\Omega_2$ is satisfied for any τ, which assures that the scheme converges to the steady-state solution independently of the value of the pseudo-timestep. This condition is referred to as the complete consistency. The choice of an optimal parameter sequence $\tau_1\dots\tau_n$ as well as other techniques accelerating the convergence are summarized e.g. in (Marchuk, 1990).

The operators A_x, A_y obtained for the homogeneous media on uniform grids commute, i.e.

$$A_xA_y=A_yA_x \tag{11}$$

In practice the commutativity condition (11) proves to be important in deriving properties of the fractional step schemes like stability and consistency. The convergence of the two-

dimensional ADI scheme without the requirement of the commutativity of the operators A_x, A_y was discussed by Birkhoff and Varga (1959). Further considerations related to the rate of convergence and parameter choice can be found in (Pearcy, 1962; Widlund, 1966).

The simple extension of the ADI scheme to three dimensional cases results in loss of unconditional stability (Janenko, 1971). The stable version was proposed by Douglas and Rachford (1956)

$$\begin{aligned}
\frac{\varphi^{j+1/3}-\varphi^{j}}{\tau} &= A_x\varphi^{j+1/3} + A_y\varphi^{j} + A_z\varphi^{j} \\
\frac{\varphi^{j+2/3}-\varphi^{j+1/3}}{\tau} &= A_y(\varphi^{j+2/3}-\varphi^{j}) \\
\frac{\varphi^{j+1}-\varphi^{j+2/3}}{\tau} &= A_z(\varphi^{j+1}-\varphi^{j})
\end{aligned} \tag{12}$$

The scheme is proven to be unconditionally stable for homogeneous media and is completely consistent. The second order accurate in time version was suggested later by Douglas (1962).

The pair-wise commutativity is required for the stability of the classical ADI schemes in 3D, thus they are not stable for heterogeneous materials. This limits their use as iterative solvers for steady state problems. For non-commuting positive-definite operators multistage alternating direction method was suggested (Douglas et al., 1966). The absolutely stable scheme (multicomponent method of alternating directions) preserving the complete approximation property was presented in (Abrashin et al., 2001; Abrashin and Mukha, 1992).

Locally One Dimensional schemes

The splitting algorithms belong to the other class of fractional step techniques.

$$\begin{aligned}
\frac{\varphi^{j+1/3}-\varphi^{j}}{\tau} &= \frac{1}{2}A_x(\varphi^{j}+\varphi^{j+1/3}) \\
\frac{\varphi^{j+2/3}-\varphi^{j+1/3}}{\tau} &= \frac{1}{2}A_y(\varphi^{j+1/3}+\varphi^{j+2/3}) \\
\frac{\varphi^{j+1}-\varphi^{j+2/3}}{\tau} &= \frac{1}{2}A_z(\varphi^{j+2/3}+\varphi^{j+1})
\end{aligned} \tag{13}$$

In the above, each of the equations involves only one-dimensional difference operators and the scheme is therefore categorized as *locally one dimensional* (LOD). A similar, fully implicit variant of (13) is possible. Both algorithms are unconditionally stable (Janenko, 1971).

The two-dimensional variant of the scheme (13) is second order accurate in time for homogeneous media, but this property is lost for non-commutative operators $\mathrm{A}_x, \mathrm{A}_y$. It can be re-

stored by introducing two-cycle splitting (xy sweeps followed by reversed order yx sweeps) (Marchuk, 1990).

The two dimensional version of the LOD scheme (13) for homogeneous media is identical in whole steps with ADI. However, this equivalence holds only unless boundary conditions are considered. Imposing boundary conditions onto one-dimensional sweeps in (13), leads to the finite approximation error at points located next to the boundaries. The resulting scheme in whole step yields

$$(I-\frac{1}{2}\tau\mathrm{A}_x)(I-\frac{1}{2}\tau\mathrm{A}_y)u^{n+1}-(I+\frac{1}{2}\tau\mathrm{A}_x)(I+\frac{1}{2}\tau\mathrm{A}_y)u^{n}=R_n \tag{14}$$

where $R_n=0$ everywhere inside the computational domain, expect in the points near the boundaries. The method of undetermined functions is an elegant and efficient way of imposing boundary conditions. An auxiliary right-hand side vector is introduced that is determined by the requirement of the vanishing R_n. Similar considerations apply to the heat generation term that needs to be modified before it enters the LOD scheme.

The two dimensional LOD scheme is strongly consistent (Janenko, 1971). This property is lost for three-dimensional case and the scheme (13) cannot be directly used in order to obtain steady-state solution.

Similar approaches have been proposed for hyperbolic type of equations in (Marchuk, 1990). The problem of acoustic wave propagation and locally one dimensional scheme in this case are given by

$$\frac{\partial^2\varphi}{\partial t^2}-\sum_{i=1}^{3}\frac{\partial}{\partial x_i}\left(k_i(x,t)\frac{\partial\varphi}{\partial x_i}\right)=0$$

$$\begin{aligned}
\frac{\varphi_x^{j+1}-\varphi_y^{j}-\varphi_z^{j}+\varphi_x^{j}}{\tau^2}&=\frac{1}{3}A_x(\varphi_x^{j+1}+\varphi_x^{j}) \qquad \varphi_x^{j}=\varphi^{j+1/3}\\
\frac{\varphi_y^{j+1}-\varphi_x^{j+1}-\varphi_z^{j}+\varphi_y^{j}}{\tau^2}&=\frac{1}{3}A_y(\varphi_y^{j+1}+\varphi_y^{j}) \qquad \varphi_y^{j}=\varphi^{j+2/3}\\
\frac{\varphi_z^{j+1}-\varphi_x^{j+1}-\varphi_y^{j+1}+\varphi_z^{j}}{\tau^2}&=\frac{1}{3}A_z(\varphi_z^{j+1}+\varphi_z^{j}) \qquad \varphi_z^{j}=\varphi^{j+1}
\end{aligned} \tag{15}$$

Tridiagonal systems of equations

Fractional step methods consist of a sequence of implicit one dimensional sweeps through the domain. Forming a system of equations for a discretized one dimensional PDE results in a tridiagonal matrix A:

$$A = \begin{pmatrix} b_1 & c_1 & & & 0 \\ a_2 & b_2 & c_2 & & \\ & a_3 & b_3 & \ddots & \\ & & \ddots & \ddots & c_{n-1} \\ 0 & & & a_n & b_n \end{pmatrix} \tag{16}$$

where n denotes the number of points in the discretization. The simplest algorithm to solve such a system is the *Thomas Algorithm (TA),* which is in fact the simplest case of the Gaussian elimination, or LU decomposition (Conte and Boor, 1980).

Thomas Algorithm

The Thomas algorithm consists of two phases: forward elimination and backward substitution, described by first order linear and non-linear recurrences presented in equations (17) and (18), e.g. $p_i = f(p_{i-1})$. Basic implementation of TA requires $8n$ floating point operations, two vectors of length n (x, the result and *rhs*, the right-hand side), two auxiliary vectors of length n (p and q) used during the factorization process, and the tridiagonal matrix A of size $3n$. TA can not be parallelized due to the existence of first order recurrences. The forward elimination phase is described by the following

$$\begin{aligned} p_i &= \frac{c_i}{b_i - a_i \cdot p_{i-1}} & p_1 &= \frac{c_1}{b_1} \\ q_i &= \frac{rhs_i - a_i \cdot q_{i-1}}{b_i - a_i \cdot p_{i-1}} & q_1 &= \frac{rhs_1}{b_1} \end{aligned} \tag{17}$$

and the backward substitution phase can be written as

$$x_i = q_i - x_{i+1} \cdot p_i \quad x_n = q_n \tag{18}$$

The above formulas directly reflect the steps taken during the Gaussian elimination applied to this kind of matrices.

Parallel Tridiagonal Solvers

The drawback of TA is that it is strictly sequential. Many parallel algorithms for solution of tridiagonal systems of equations have been developed. Cyclic reduction was proposed by

Hockney (1965) and has since been widely used on distributed machines (Allmann et al., 2001). The idea is to recursively reduce the number of equations by 2 until a system of only 2 equations is obtained. Even in sequential case it is often preferred to the TA because of its natural ability to handle periodic boundary conditions. Divide and conquer algorithm is a similar approach (Gander and Golub, 1997). Recursive doubling was proposed by Stone (1973) and its suitability for hypercube architectures was investigated in (Egecioglu et al., 1989). This algorithm is based on recursive doubling solutions of linear recurrence relations, which in the case of tridiagonal systems allow to compute LU decomposition of the matrix in $O(\log_2(p))$ parallel steps, where p is the number of processors. Solving tridiagonal systems on distributed architectures in the context of ADI-type of methods has been studied in (Wakatani, 2004), where a pre-propagation scheme for solving first order recurrences has been proposed. As in the case of all the previously mentioned approaches, parallelization of the solver has been done at the expense of at least doubling the computational complexity.

Other known methods include iterative relaxation schemes, like Gauss-Seidel, Jacobi, Red-black line relaxation and segment relaxation.

In the cases of where many tridiagonal systems of equations have to be solved (like ADI-type of methods), there have been attempts to parallelize TA for distributed machines through pipelining, e.g. (Povitsky, 1999). For a comparison of the performance of many of the above algorithms, see (Hofhaus and VandeVelde, 1996).

Since our interest are ADI-type of methods for 3D problems on shared memory architectures, for our parallel implementation we chose to use the TA because of its optimal computational complexity and modest memory requirements. As shown later in the performance analysis section, we were able to obtain a scalable parallel implementation without the need to use the pipelining approach.

Implementation

Large class of numerical methods discretizing PDEs on structured meshes, such as *Finite Differences* implemented as stencil operations, only require a small number of floating point operations per a memory access. During the recent years a growing discrepancy between the memory and the CPU speeds is observed (McCalpin, 1991-2007), which results in the memory bandwidth being a bottleneck, and not the CPU speed. The theoretical peak performances of modern CPUs are only achievable for algorithms that are computationally heavy compared to the memory bandwidth requirements (e.g. BLAS Level3 operations (Whaley et al., 2001)).

In the following section we demonstrate how to efficiently implement a wide class of *ADI/LOD* methods for three dimensional problems under these constraints. For all our performance tests we use an 8-way Opteron system with 2.4GHz DualCore CPUs and DDR333 memory. The code is written in C and compiled using the Intel C compiler.

Performance measurements

A common measure of computational efficiency is *flops* (Floating Point Operations per Second). Different approaches can be used to estimate this value. The number of floating point operations required by a best known algorithm can be computed, but since the operations are often not the bottleneck, this can be meaningless. Another way is to add all the operations performed explicitly in the code. Unfortunately, because of heavy optimizations performed by the compiler this number may differ from what the CPU actually executes. Hardware counters capable of tracking performance statistics, like cache misses and the actual number of flops performed by a CPU recently became popular (Browne et al., 2000). They are very useful when identifying performance problems in the code, but this analysis can be complex. In our paper we chose to present values estimated directly from the fastest obtained sequential implementation. In particular, we include all the compulsory operations that can not be removed by the compiler (e.g. $8n$ in the Thomas Algorithm), and we include operations that could potentially be pre-computed before the execution and stored in auxiliary arrays, but it is not done since it would either significantly increase the total memory requirements (threefold in the case of the conductivity parameter k), or even increase the execution time due to higher memory bandwidth requirements. As we show later, averaging the conductivity on the fly does not increase execution time much. Since our numbers are probably slightly higher than the real ones due to compiler optimizations, we also provide the CPU time spent on the computations and compare it to simplest unimprovable cases, such as making a copy of a memory area.

Operations on multidimensional data

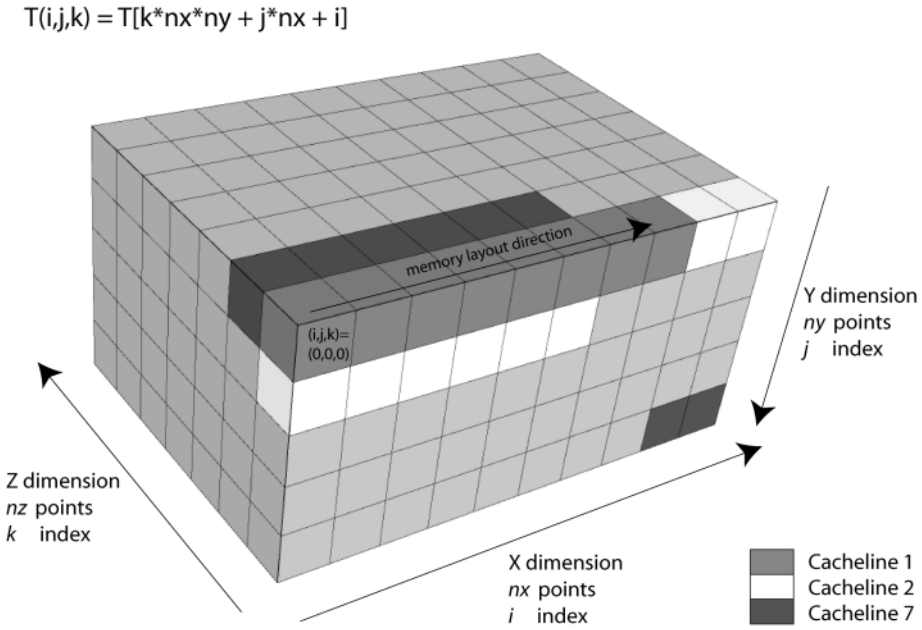


Figure 1

Three dimensional array placement in computer's memory.

Three dimensional arrays are stored linearly in computers memory, as shown in Figure 1. The data is transferred from the main memory to the CPU cache in blocks called *cache lines* (usually 64 bytes), thus the cost of accessing a whole cache line is the same as that of accessing a single value. Using all the values from a given cache line at least once before it is removed from the cache back to the main memory significantly decreases the cost per value. The following pseudo-code fragments show two possible implementations of the difference operator in the Z dimension. The only difference is the order of the loops:

```
for i=0:nx
 for j=0:ny
  for k=0:nz-1
    diffz[k*nx*ny + j*nx + i] = T[(k+1)*nx*ny + j*nx + i] - T[(k+0)*nx*ny + j*nx + i];
  done
 done
done

for k=0:nz-1
 for j=0:ny
  for i=0:nx
    diffz[k*nx*ny + j*nx + i] = T[(k+1)*nx*ny + j*nx + i] - T[(k+0)*nx*ny + j*nx + i];
  done
 done
done
```

Because of the loop order, first implementation will be referred to as *ijk,* the second one – *kji*. Traversal of the 3D array for both of these implementations is presented in Figure 2 and Figure 3 respectively. The innermost loop in the *kji* version is consistent with the linear memory layout, thus it is made sure that all the values from every loaded cache line are used at least once.

The importance of this practice is best reflected in the following performance comparison. For a 3D array of size nx=ny=nz=1000 and single precision floating point numbers, the *kji* implementation of the difference operator in the Z direction takes 6.9 seconds, while the *ijk* takes 95 seconds. For a difference operator in the X direction, the execution time is 4.9 seconds, the same as just copying an array of this size from one place to another in the memory.

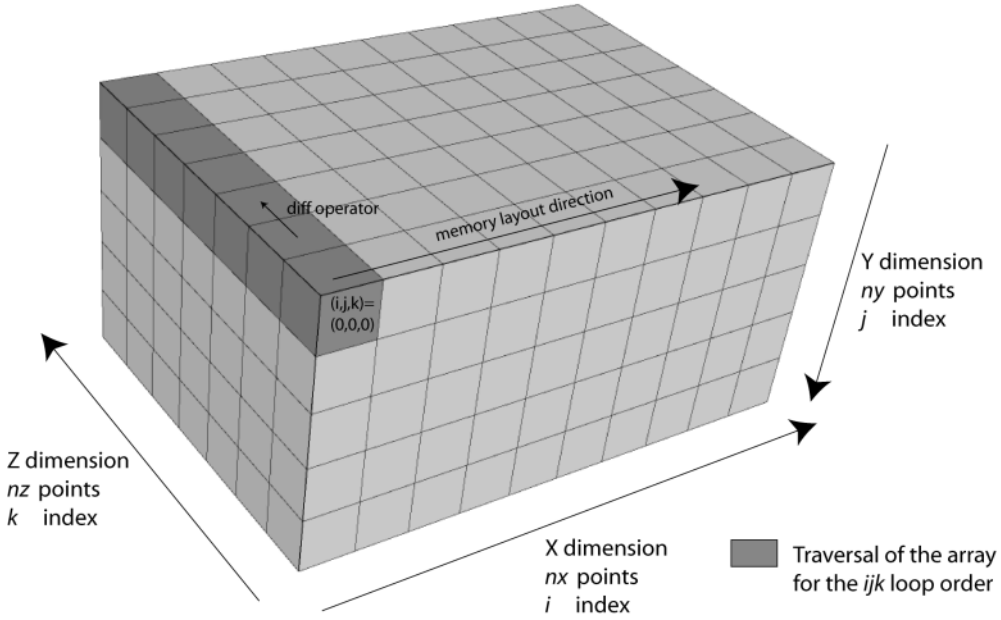


Figure 2

Direction of memory access during the Z difference operator, ijk *loop order.*

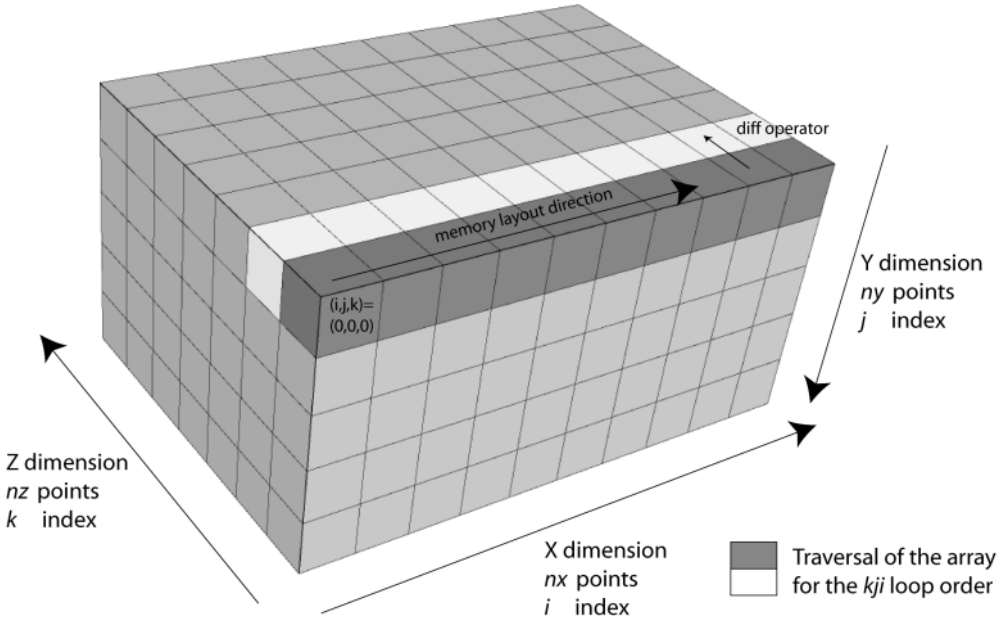


Figure 3

Direction of memory access during the Z difference operator, kij loop order.

REMARK. In our performance tests we use single precision floating point numbers because they require twice less memory than double precision, and the computations are faster. In the case of our implementation and the inspected architecture this results in more than two-fold speedup. Using single precision for computations wherever possible grows more and more popular, especially with the introduction of high performance GPUs, FPGAs, or the Cell processor, on which the performance difference reaches one order of magnitude (Langou et al., 2006). Although single precision floats may not provide enough accuracy for certain problems, it is recently widely studied how to combine them with double precision computations only for accuracy critical parts of the code, and obtain a full double precision result (Kurzak and Dongarra, 2007; Strzodka and Göddeke, 2006). With some care this approach can also be applied to ADI-type of methods. It is beyond the scope of this paper. For now, our implementation can run in full double precision mode when required. All our performance considerations also hold for this case.

On modern architectures the cost of accessing RAM is further decreased by assuring that the data is constantly read while the CPU is busy performing computations. This technique is known as prefetching, and is automatically activated by the CPU or a compiler, provided that subsequent cache lines are processed in order. We refer to this approach as the *linear memory access*. In the *kji* implementation we traverse the array along the direction of the memory (*i* index), therefore prefetching will be used during execution.

Most modern CPUs are capable of limited vectorization of the floating point operations as long as the vector elements are located next to each other in the memory. On x86 architecture it is realized through the SSE2 instruction set, which makes it possible to perform 4 single precision or 2 double precision operations at the same time. This feature usually has to be activated using special compiler flags. With icc we use the '-xW –O3' flags. In a common vector notation, the second loop would be executed by the CPU as:

```
for k=0:nz-1
 for j=0:ny
  for i=0:4:nx
    diffz(i:i+3,j,k) = V(i:i+3,j,k+1) - V(i:i+3,j,k);
  done
 done
done
```

To make vectorization possible, the vectors operated on need to be stored linearly in the memory. Although the performance improvement is negligible for a simple difference operator, it becomes much more pronounced in a more computationally heavy code presented later.

Cache reuse

As noted in the previous section, one should always use all the data from a cache line at least once. However, this simple approach does not assure that a given cache line will not be loaded from RAM again during later computations. A common efficient programming paradigm is based on maximizing the cache reuse, i.e. using every cache line in all or most of the required computations. In the case of stencil operations this leads to so called *blocking* or *tiling* (Rivera and Tseng, 2000). In short, a 3D grid is divided into smaller cubes that fit completely into the CPUs cache. All required operations on a given block are performed before moving to the next one.

During the past years lots of work has been done on designing cache-friendly algorithms. As described in (Kamil et al., 2005), many of them have become ineffective on modern architectures due to the increased importance of data prefetching and linear memory access on the overall performance. Our results agree with these findings. In the simple example of the difference operator, with optimal cache reuse for the Z direction one could only hope to decrease the execution time from 6.9 seconds to 4.9 seconds. For problems with relatively more computations per data access the possible gains are even smaller. While Kamil and others concentrate on local stencil operations, we show how to efficiently handle *Fractional Steps/ADI* type of algorithms, where 1D implicit sweeps have to be completed in sequential order and need to access the data along a whole dimension, which is in general not consistent with the linear memory layout.

LOD/ADI

A general engine of any *ADI/LOD* type of algorithms looks as follows:

```
for X, Y, Z dimensions in turns
   for all equations in given dimension
      -  create tridiagonal matrix A
      -  create right-hand side, possibly with a source term and (depending on the
         scheme used) explicit differences in other dimensions
      -  call Thomas Algorithm to solve the equations
      -  store the result in the three dimensional array
```

A simple implementation of every 1D sweep consists of three nested loops over i, j and k indices for X, Y and Z dimensions respectively, where the innermost loop is iterated along the direction of the sweep. Here we study the simplest LOD scheme presented in equation (13). In our implementation, during the tridiagonal matrix assembly material parameters K assigned to cells are averaged on the fly. rhoCp is assigned directly to the nodes. The right-hand side consists of the old temperature values (a fully implicit scheme) and the source/sink term G applied only during the X sweep. Effectively during the X sweep we operate on four 3D arrays of size

nx*ny*nz (where temperatures are both read and written), and during Y/Z sweeps three arrays are used. We perform ca. 50 operations per point, where most of the operations are related to averaging of the K properties. While this step could be performed before running the solver, this would increase the total memory requirements for K array threefold, and increase the per-sweep memory requirements three times for the schemes requiring derivatives in the other dimensions. Moreover, we show that computing these values during the sweeps does not affect the execution times much.

For the X sweep the best order of the loops is *kji*. For the Y sweep we can choose between *ikj* and *kij*. Similarly, for the Z sweep we can use either *ijk*, or *jik* implementation. Only in the case of the X sweep the linear memory access is obtained. Performance results of all these implementations for a model of size nx=ny=nz=1000 are presented in Table 1.

Table 1 Performance of the naive *implementation of the implicit LOD method.*

	X sweep, kji	*Y sweep, ikj*	*Y sweep, kij*	*Z sweep, ijk*	*Z sweep, jik*
Time (s)	29	288	85	337	212
MFlops	1750	169	572	144	229

Compared to the difference operator, we use more than twice the memory amount (not taking into account the matrix A, the right-hand side and the TA workspace), and perform much more computations. Assuming that the computations take no time, best achievable performance estimated based on the time required to read and write the data in memory is around 10 seconds for X sweep, and around 8 seconds for Y/Z sweeps.

Improving the Y and Z sweeps requires obtaining linear memory access, i.e. implementing the loop order as *kji* and *jki* respectively. The idea is presented in Figure 4. Instead of building and solving a single tridiagonal system, we build and solve a whole plane of systems of equations. The pseudocode for the Z sweep is presented below:

```
for j=1:ny-1
  for k=1:nz-1
    for i=1:nx-1
      compute row k in tridiagonal matrix A_{i,k}=f(T_{i,j,k},K)
      compute entry k in right-hand side vector rhs_{i,k}=f(T_{i,j,k},K)
    end
  end
  call Thomas Algorithm (A, x, rhs) for nx systems of equations simultaneously
  store results x_{i=1:nx-1,k=1:nz-1} in the three-dimensional array T_{i=1:nx-1,j,k=1:nz-1}
end
```

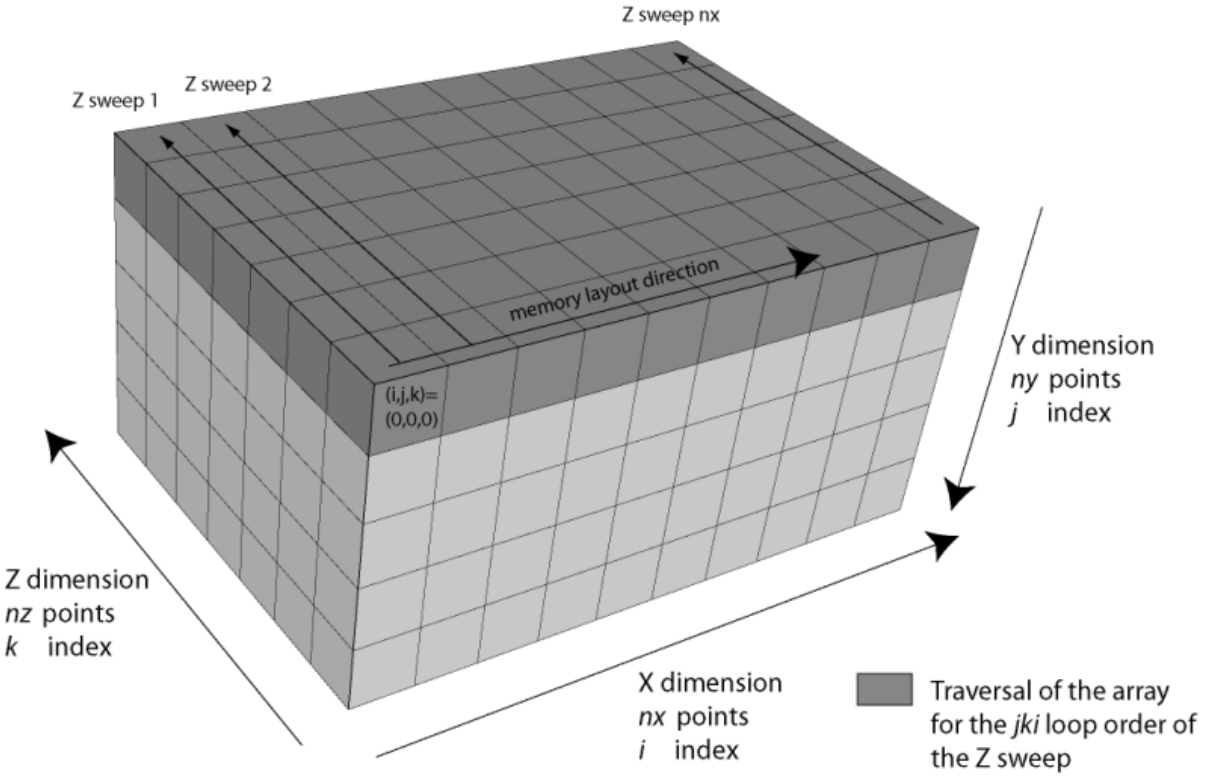


Figure 4

In order to obtain linear memory access during the Z sweep a whole plane of systems of equations has to be simultaneously created and solved.

Described loop transformation assures linear memory access during the Y and Z sweeps, and allows the use of SSE vectorization for the matrix computation. Moreover, as a natural consequence of solving a plane of systems of equations, SSE can also be used to implement a vectorized *Thomas Algorithm*. On the other hand, in the unchanged X sweep vectorization is only used during the matrix computation. Using a vectorized TA also in this case would require transposing the A matrices and *rhs* vectors after they have been computed, which is memory-wise inefficient and brings no speedup. The importance of using vectorization wherever possible can be verified by compiling and running the code with and without SSE2 support (see Table 2).

Table 2 The impact of SSE vectorization on execution time.

	X sweep not vectorized	*X sweep SSE vectorized*	*Y sweep, kji not vectorized*	*Y sweep, kji SSE vectorized*	*Z sweep, jki not vectorized*	*Z sweep, jki SSE vectorized*
Time (s)	48	29	45	32	47	33
MFlops	1039	1750	1073	1491	1028	1452

Presented approach requires additional storage for *nx* tridiagonal matrices, right-hand side vectors and the factor. Since these arrays are of considerable size and do not fit into CPUs cache for large problems, they decrease the overall performance of the algorithm. It can be noticed that every entry of the A matrix is only used once during the forward elimination stage of TA. Including this stage into the matrix building loop and factorizing the matrix A on the fly eliminates the need to store and later read its entries. Doing so also allows us to use the *rhs* vector in place of the *q* vector from equation (17). Finally, in the fully implicit case the temperature values themselves can be used as the *rhs* vector. These improvements can also be applied in the Crank-Nicolson schemes, and the schemes that require explicit derivatives in the other dimensions by using an auxiliary vector of size *nx* instead of *nx*nz*.

Table 3 Performance of SSE vectorized code with integrated TA solver.

	memcpy	*explicit scheme*	*X sweep, kji*	*vectorized Y sweep, kji*	*vectorized Y sweep, kji integrated TA*	*vectorized Z sweep, jki*	*vectorized Z sweep, jki integrated TA*
Time (s)	~8-10	37	29	32	20	33	22
MFlops	0		1750	1491	2428	1452	2229

The performance results for the methods described above are summarized in Table 3. It is worth noticing that best Y and Z sweeps implementations are faster than the X sweep. This is due to the SSE vectorization of the TA, which is a natural consequence of presented approach, but can not be efficiently implemented for the X sweep. For comparison, the table also includes the time needed to perform a single step of our implementation of the explicit finite difference scheme with variable material coefficients, SSE vectorization and proper memory access.

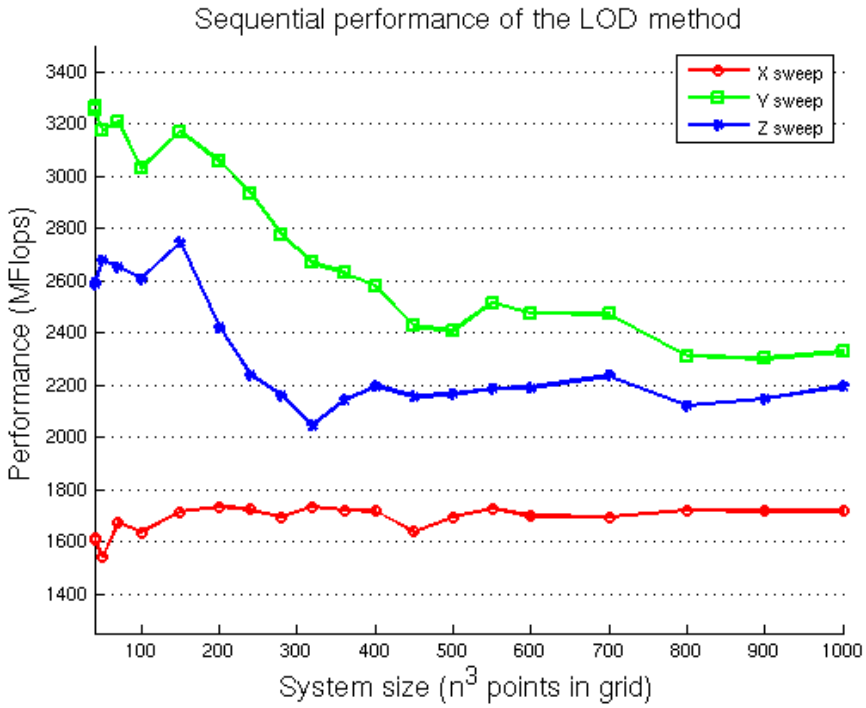


Figure 5

Flops performance of the optimized sequential implementation of 1D sweeps in all special directions depending on the problem size.

Figure 5 presents the flops performance of the optimized code depending on the system size. High efficiency is sustained for large problems and for sweeps in all spatial directions. Note that the flops performance is up to 30% higher for small systems, which exhibit some cache reuse of the K properties array between the solves done on subsequent planes, and the factor and *rhs* during forward and backward stages of the TA. This also indicates that for large systems the code is relatively slower due to the heavier memory requirements and not the CPU speed.

Parallel implementation

In the simplest approach *ADI/LOD* methods are parallelized by dividing the outermost loop between the threads. This assures locality of most of the data used during the sweeps in 2 dimensions, the exception being the material properties K, for which neighboring CPUs share the border. On the other hand, the sweep in the third dimension accesses the data across all the CPUs, which involves communication and a performance hit, which on a shared memory system is considerably lower than on distributed memory architectures. Starting from a close to optimal sequential performance we show that this simple approach yields very good results for the studied computer architecture.

On shared memory machines the whole memory of the system can be allocated and addressed directly as a single array. This way there is no need to explicitly program the communication, which is performed automatically by the CPU while accessing the required data in the global memory space. On NUMA (*Non-Uniform Memory Access*) capable architectures every CPU has its private memory bank, which assures parallelization of not only the computations, but also the memory bandwidth. On these systems in order to obtain scalable code it is important to assure that the data accessed by a thread belongs to the private memory bank of the CPU the thread is executed on. For scientific applications it is commonly achieved by binding threads to specific CPUs and using a technique called *first touch* for data allocation. Basically, every thread initializes (e.g. sets to 0) only this part of an array, which it will later use during the computations. The operating system then assigns all parts of the array to the proper memory banks.

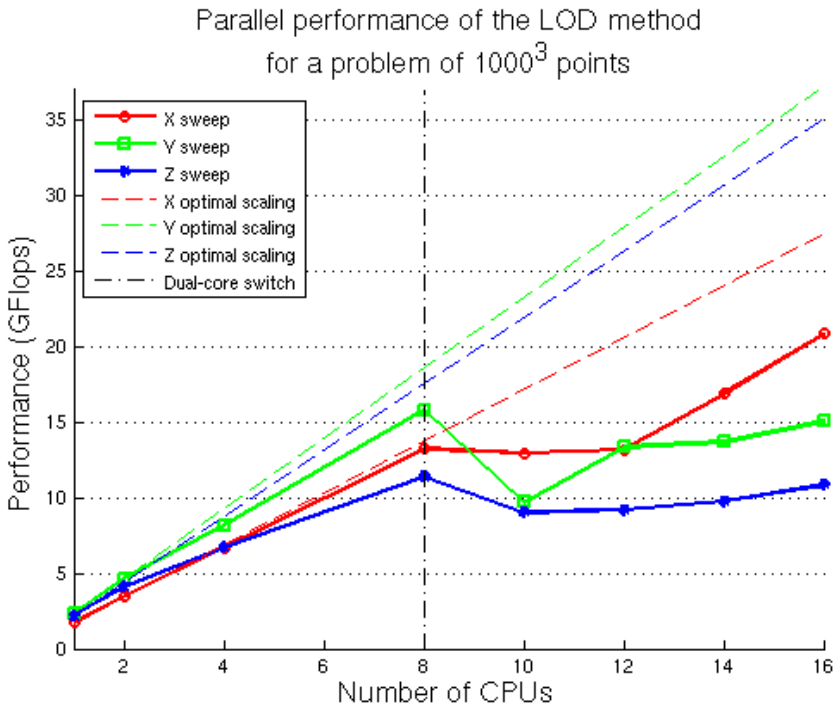


Figure 6

Flops performance of parallel execution of the optimized implementation for a constant problem size of nx=ny=nz=1000.

Figure 6 presents scaling of the optimized code on model of size nx=ny=nz=1000. Speedup of the X and Y sweeps on up to 8 CPUs is linear and close to perfect. The Z sweep suffers a small but acceptable performance penalty due to the communication between CPUs. Using more CPUs requires switching to the second core, in which case limited speedup can only be observed for the X sweep. On a multi-core NUMA system all cores of a CPU share the same

memory bus, which results in the threads competing for the memory bandwidth. Since in the case of Y and Z sweeps most of it was already used by a single thread (for smaller problems computations are done faster, which indicates that the implementation is memory bounded for larger problems), no speedup is to be expected. On the other hand, the TA solver in the sequential X sweep was not vectorized, and the computations took relatively longer time, leaving some memory bandwidth to be utilized by the second core. For studied problem size, the X, Y and Z sweeps on 8 CPUs take 3.8, 3.0 and 4.2 seconds respectively.

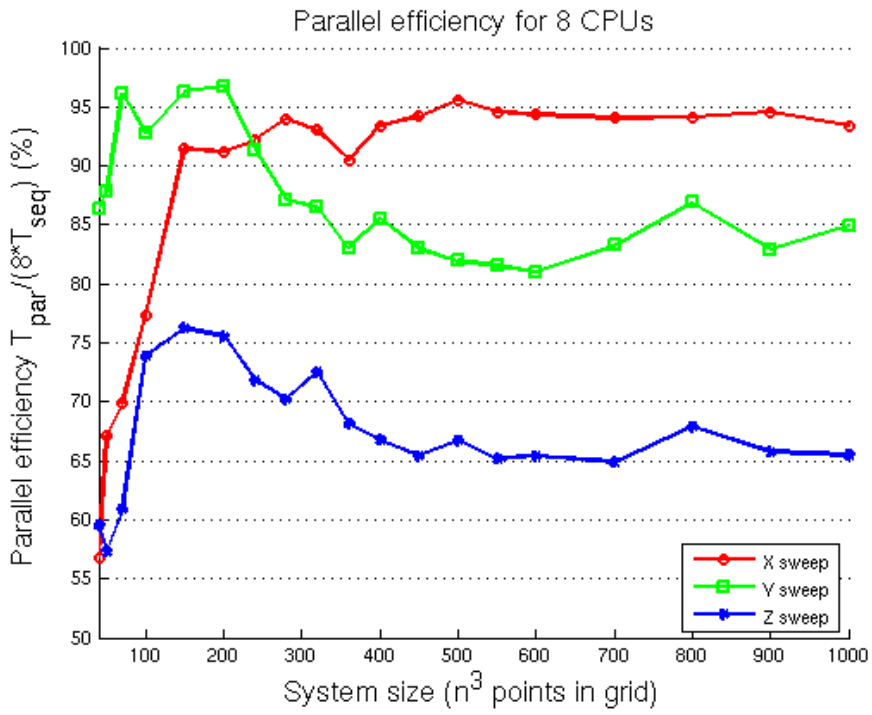


Figure 7

Parallel efficiency of the method on 8 Opteron CPUs depending on problem size

Figure 7 shows parallel efficiency of the code for 8 CPUs. An interesting observation can be made for small problems (nx=ny=nz=~200), for which not only the sequential performance of Y and Z sweeps is higher (see Figure 5), but also the parallel efficiency. This means that the cache reuse also limits the data exchange between the CPUs. It indicates possible improvements to the parallel code involving duplication of the material parameters matrix K at the CPU boundaries. This however would require allocation of a separate K matrix for the Z dimension. For small enough problems some speedup is also observed for all the sweeps when using the second core. In this case the code spends relatively more time on computations, leaving some memory bandwidth unused. It brings us to a conclusion that although cache reuse plays much smaller role for sequential codes, it may yet prove to be important for utilizing the

parallel computational power of multi-core systems. A more in-depth analysis is beyond the scope of this paper.

Numerical examples

The LOD scheme for parabolic problems shown in equation (13) has been used to compute pressure evolution in an oil reservoir. We have used real-world data with resolution around 800x800x30. Figure 8 presents the three dimensional model and the porosity data. In our models we have assumed zero permeability in the matrix, i.e. around the reservoir. Figure 9 shows a horizontal cut through the model. Figures 10 and 11 present the results of our computation of the pressure inside the reservoir. Two sources have been used, one with positive and one with negative pressure values. The LOD scheme for hyperbolic problems shown in equation (15) has been used to model P-wave hitting the Gullfaks reservoir. The wave has been initiated through a Gaussian initial pressure distribution. Non-reflecting boundary conditions are used. A frame from the simulation is presented in Figure 12.

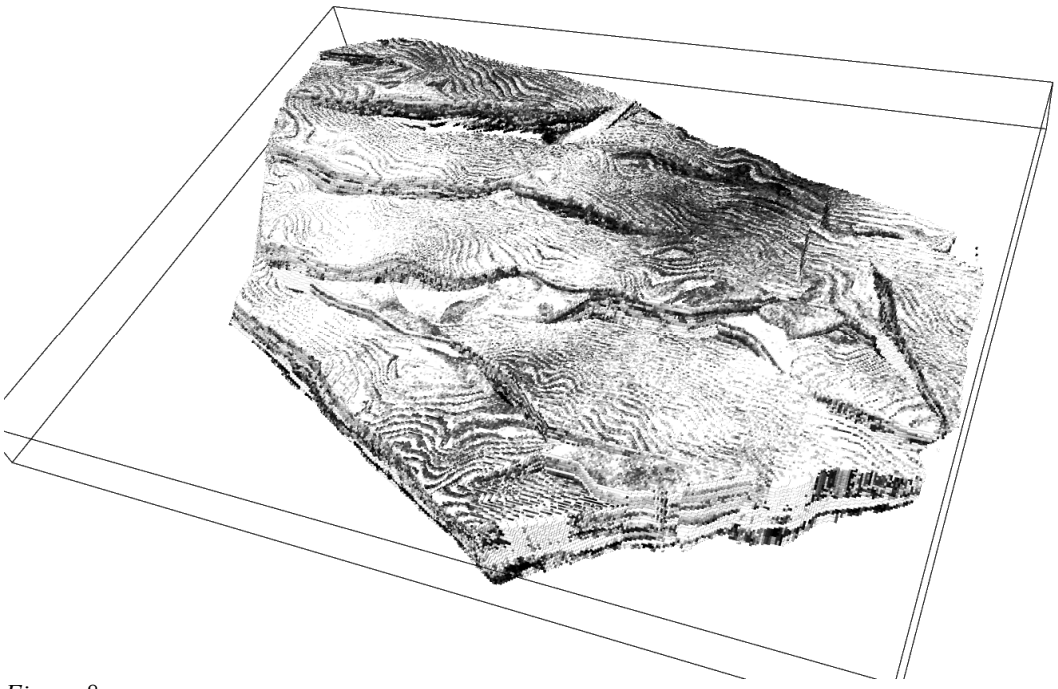


Figure 8

The Gullfaks oil reservoir. Three dimensional model, spatial 800x1000x70 resolution. Different shades show porosity data (darker shade = higher porosity value).

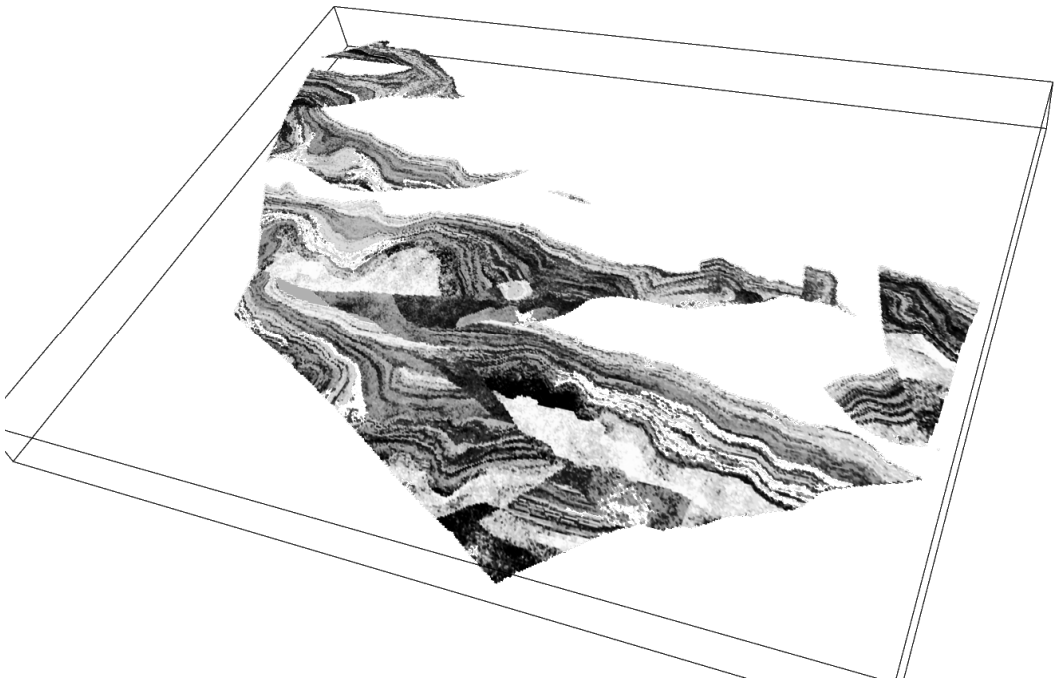


Figure 9

A horizontal cut through the Gullfaks reservoir model at half depth. Different shades show porosity data (darker shade = higher porosity value).

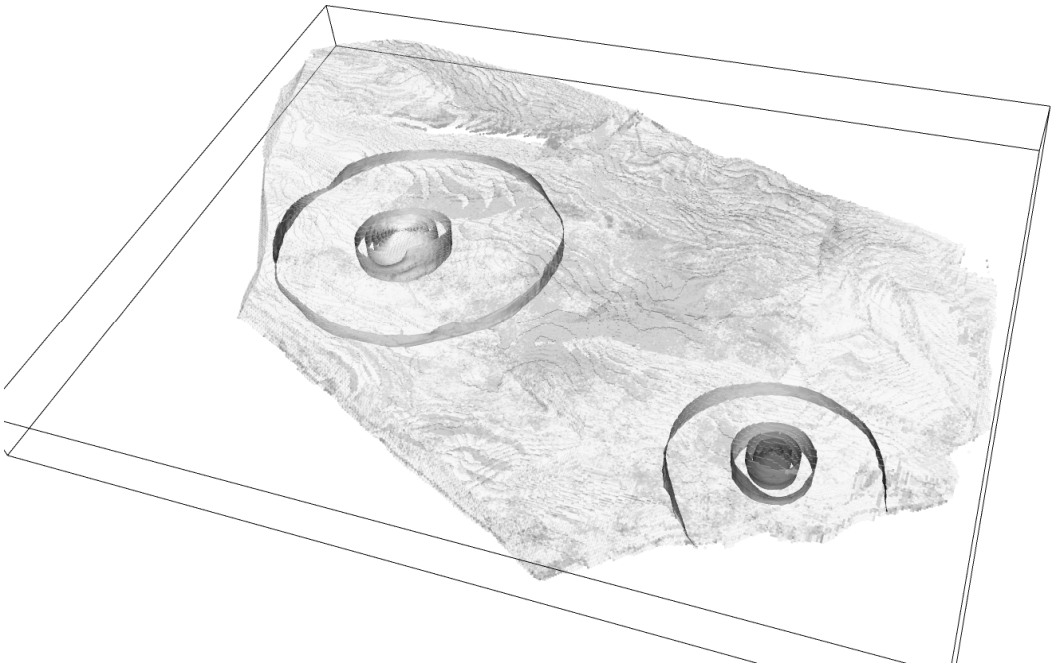


Figure 10

Pressure evolution inside the reservoir, initial stages. Figure present iso-contours of chosen pressure values. Darker contours denote negative pressure values, brighter – positive pressure values.

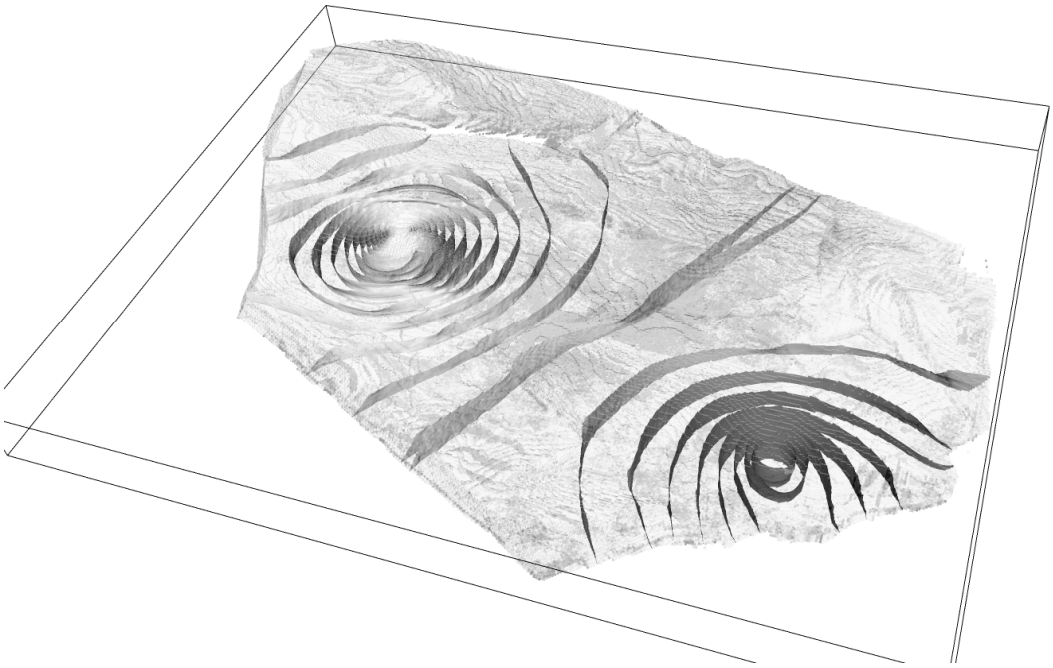


Figure 11

Pressure evolution inside the reservoir, final stage. Figure present iso-contours of chosen pressure values. Darker contours denote negative pressure values, brighter – positive pressure values.

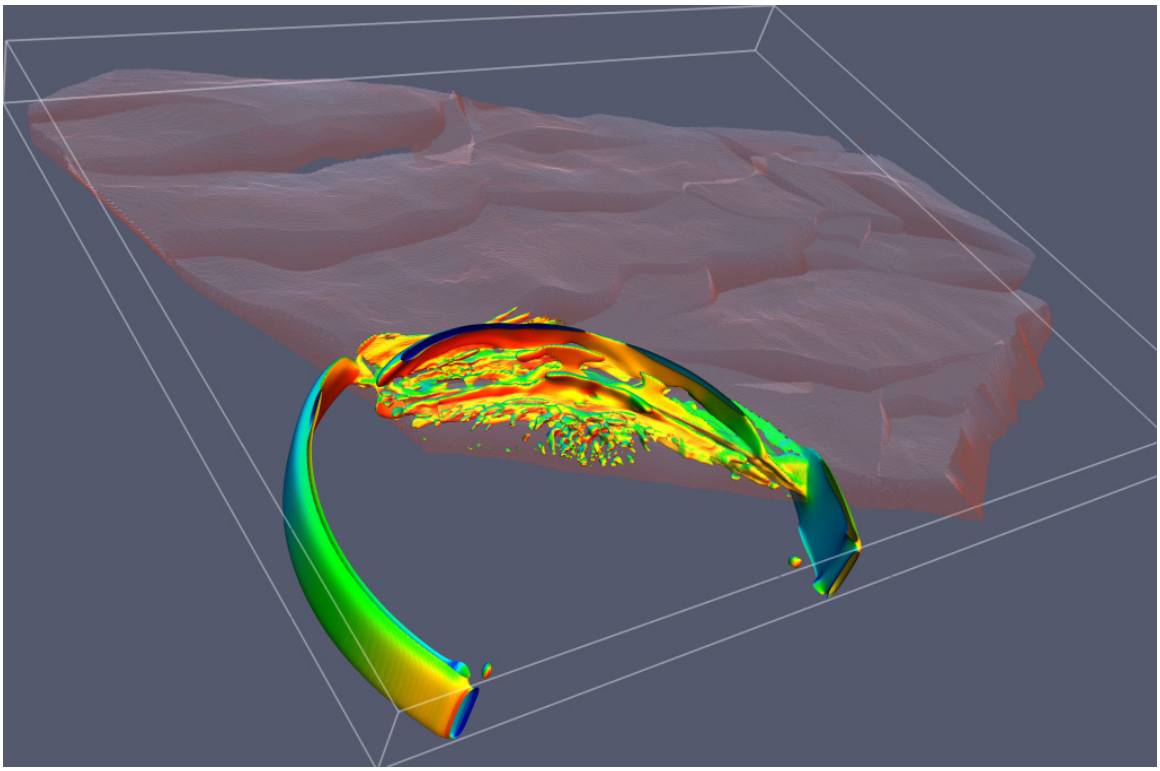


Figure 12

P-wave hitting a reservoir. 800x1000x70 resolution, non-reflecting boundaries, initial conditions implemented as Gaussian pressure distribution.

Conclusions

We have presented an efficient implementation of *ADI/LOD* type of methods for three dimensional problems on modern commodity architectures. Special attention has been paid to optimization of the memory bandwidth usage. High efficiency is sustained for problems of 1 billion and more degrees of freedom. Optimized sequential implementation of the Y and Z sweeps is comparable in execution time to just copying the data in the memory. The time needed to perform one LOD iteration is approximately twice the time of an explicit scheme. Efficiency of the parallel implementation on a NUMAcc shared memory system with 8 CPUs is close to perfect. Scalability when using the second core is limited because of memory bandwidth starvation. Computing one timestep of the LOD scheme on a system of 1000^3 unknowns on 8 CPUs takes 11 seconds. Applications of the methods to parabolic and hyperbolic problems have been presented using high-resolution, real-world oil reservoir data.

References

Abrashin, V.N., Egorov, A.A. and Zhadaeva, N.G., 2001. On a class of additive iterative methods. Differential Equations, 37(12): 1751-1760.

Abrashin, V.N. and Mukha, V.A., 1992. A Class of Economical Difference-Schemes for Solving Multidimensional Problems of Mathematical Physics. Differential Equations, 28(10): 1458-1469.

Adams, M., Brezina, M., Hu, J. and Tuminaro, R., 2003. Parallel multigrid smoothing: polynomial versus Gauss--Seidel. J. Comput. Phys., 188(2): 593-610.

Allmann, S., Rauber, T. and Runger, G., 2001. Cyclic Reduction on Distributed Shared Memory Machines, Proceedings of the Ninth Euromicro Workshop on Parallel and Distributed Processing, Mantova, Italy, pp. 290-297.

Birkhoff, G. and Varga, R.S., 1959. Implicit Alternating Direction Methods. Transactions of the American Mathematical Society, 92(1): 13.

Browne, S., Dongarra, J., Garner, N., Ho, G. and Mucci, P., 2000. A portable programming interface for performance evaluation on modern processors. International Journal of High Performance Computing Applications, 14(3): 189-204.

Conte, S.D. and Boor, C.W.D., 1980. Elementary Numerical Analysis: An Algorithmic Approach. McGraw-Hill Higher Education, 408 pp.

Courant, R., Friedric.K and Lewy, H., 1967. On Partial Difference Equations of Mathematical Physics. Ibm Journal of Research and Development, 11(2): 215-&.

Douglas, J., 1955. On the Numerical Integration of D2U-Dx2+D2U-Dy2=Du-D+ Implicit Methods. Journal of the Society for Industrial and Applied Mathematics, 3(1): 42-65.

Douglas, J., 1962. Alternating direction methods for three space variables. Numerische Mathematik, 4(1): 41.

Douglas, J., Jr., Garder, A.O. and Pearcy, C., 1966. Multistage Alternating Direction Methods. SIAM Journal on Numerical Analysis, 3(4): 570.

Douglas, J., Jr. and Rachford, H.H., Jr., 1956. On the Numerical Solution of Heat Conduction Problems in Two and Three Space Variables. Transactions of the American Mathematical Society, 82(2): 421.

Douglas, J. and Kim, S., 2001. Improved accuracy for locally one-dimensional methods for parabolic equations. Mathematical Models & Methods in Applied Sciences, 11(9): 1563-1579.

Egecioglu, O., Koc, C.K. and Laub, A.J., 1989. Recursive Doubling Algorithm for Solution of Tridiagonal Systems on Hypercube Multiprocessors. Journal of Computational and Applied Mathematics, 27: 95-108.

Gander, W. and Golub, G.H., 1997. Cyclic Reduction – History and Applications, Proceedings of the Workshop on Scientific Computing. Springer Verlag, Hong Kong.

Hockney, R.W., 1965. A Fast Direct Solution of Poissons Equation Using Fourier Analysis. Journal of the Acm, 12(1): 95-&.

Hofhaus, J. and VandeVelde, E.F., 1996. Alternating-direction line-relaxation methods on multicomputers. Siam Journal on Scientific Computing, 17(2): 454-478.

Janenko, N.N., 1971. <<The>> method of fractional steps the solution of problems of mathematical physics in several variables. Springer, Berlin etc., VIII, 160 pp.

Kamil, S., Husbands, P., Oliker, L., Shalf, J. and Yelick, K., 2005. Impact of modern memory subsystems on cache optimizations for stencil computations, Proceedings of the 2005 workshop on Memory system performance. ACM, Chicago, Illinois.

Kurzak, J. and Dongarra, J., 2007. Implementation of mixed precision in solving systems of linear equations on the cell processor. Concurrency and Computation-Practice & Experience, 19(10): 1371-1385.

Langou, J. et al., 2006. Exploiting the performance of 32 bit floating point arithmetic in obtaining 64 bit accuracy (revisiting iterative refinement for linear systems), Proceedings of the 2006 ACM/IEEE conference on Supercomputing. ACM, Tampa, Florida.

Marchuk, G.I., 1990. Splitting and alternating direction methods. Handbook of Numerical Analysis, 1, Amsterdam, 197-462 pp.

Mark, F.A., Harun, H.B., Tony, M.K. and Panayiotis, P., 2003. Applications of Algebraic Multigrid to Large-Scale Finite Element Analysis of Whole Bone Micro-Mechanics on the IBM SP, Proceedings of the 2003 ACM/IEEE conference on Supercomputing. IEEE Computer Society.

McCalpin, J.D., 1991-2007. STREAM: Sustainable Memory Bandwidth in High Performance Computers, a continually updated technical report.

Peaceman, D.W. and Rachford, H.H., 1955. The Numerical Solution of Parabolic and Elliptic Differential Equations. Journal of the Society for Industrial and Applied Mathematics, 3(1): 28-41.

Pearcy, C., 1962. On convergence of alternating direction procedures. Numerische Mathematik, 4(1): 172.

Povitsky, A., 1999. Parallelization of pipelined algorithms for sets of linear banded systems. Journal of Parallel and Distributed Computing, 59(1): 68-97.

Rivera, G. and Tseng, C.-W., 2000. Tiling optimizations for 3D scientific computations, Proceedings of the 2000 ACM/IEEE conference on Supercomputing (CDROM). IEEE Computer Society, Dallas, Texas, United States.

Stone, H.S., 1973. An Efficient Parallel Algorithm for the Solution of a Tridiagonal Linear System of Equations. J. ACM, 20(1): 27-38.

Strzodka, R. and Göddeke, D., 2006. Mixed Precision Methods for Convergent Iterative Schemes, Proceedings of the 2006 Workshop on Edge Computing Using New Commodity Architectures.

Wakatani, A., 2004. A parallel and scalable algorithm for ADI method with pre-propagation and message vectorization. Parallel Computing, 30(12): 1345-1359.

Whaley, R.C., Petitet, A. and Dongarra, J.J., 2001. Automated Empirical Optimization of Software and the ATLAS Project. Parallel Computing, 27(1-2): 3-35.

Widlund, O.B., 1966. On the Rate of Convergence of an Alternating Direction Implicit Method in a Noncommutative Case. Mathematics of Computation, 20(96): 500.