Reinhard Pippan
Peter Gumbsch
Editors

International Centre for Mechanical Sciences

Multiscale Modelling of Plasticity and Fracture by Means of Dislocation Mechanics

CISM Courses and Lectures, vol. 522

SpringerWienNewYork

SpringerWienNewYork

CISM COURSES AND LECTURES

The series presents lecture notes, monographs, edited works and proceedings in the field of Mechanics, Engineering, Computer Science and Applied Mathematics.
Purpose of the series is to make known in the international scientific and technical community results obtained in some of the activities organized by CISM, the International Centre for Mechanical Sciences.

INTERNATIONAL CENTRE FOR MECHANICAL SCIENCES

COURSES AND LECTURES - No. 522

MULTISCALE MODELLING OF PLASTICITY AND FRACTURE BY MEANS OF DISLOCATION MECHANICS

EDITED BY

REINHARD PIPPAN
AUSTRIAN ACADEMY OF SCIENCES, LEOBEN, AUSTRIA

PETER GUMBSCH
FRAUNHOFER-INSTITUT FÜR WERKSTOFFMECHANIK, IWM,
FREIBURG, GERMANY

SpringerWienNewYork

This volume contains 187 illustrations

Printed in Italy
SPIN 80016150

All contributions have been typeset by the authors.

ISBN 978-3-7091-0282-4 SpringerWienNewYork

PREFACE

The main purpose of the
CISM-BIANCHI SESSION ON MULTISCALE MODELLING OF PLASTICITY AND FRACTURE BY MEANS OF DISLOCATION MECHANICS, Udine, July 4-8, 2005
was the discussion of the latest state of the modelling of plasticity and fracture of crystalline materials on the nano- and micro-scale. The present volume offers the revised lecture notes of all who contributed the course and one contribution of the CISM-course on Multiscale Modelling and Design of New Materials, held at the same time in Udine. Plasticity and fracture are multi-scale problems. The course was devoted to regimes from atomistic to meso-scale. The central part was the discrete dislocation mechanisms and neighbouring fields above and below molecular dynamics and crystal plasticity, respectively. The contributions give an introduction to the physical phenomena, the theoretical basis, the mathematical description and the different types of simulations. Different important practical problems such as plasticity of thin films, mechanics of polycrystalline materials, fatigue crack propagation, fracture of brittle and semi-brittle materials, formation of dislocation structures, etc. were considered as examples. All contributions start with text book like introductions, describing the different simulation techniques and discussing the latest state of possibilities and limitation of the different methods. A further important part in all contributions is the interconnection of the different hierarchical levels. The contributions in the present volume describe the different phenomena following the scale from the bottom up.

The first contribution is focused to molecular dynamics methods. It takes into account the theoretical background and gives practical examples to demonstrate its capabilities and limitations. The main topic is devoted to the fundamentals of the fracture processes. The second contribution is devoted to basic dislocation properties in an elastic continuum. The basics for accurate dislocation dynamic simulation of plasticity, the dislocation-dislocation and the dislocation self interaction is introduced in detail. The method is then applied to 2-dimensional and 3-dimensional problems. The third contribution gives an introduction to the linear elastic stress and a strain field of a

crack taking into account the presence of dislocations. Then the consequence of the discrete nature of plasticity is demonstrated with the example of a moderately cyclic loaded crack. The example is used to demonstrate the importance of the discrete dislocation consideration.

The fourth contribution deals also with discrete dislocation mechanics and its impact on fracture and fatigue crack propagation. Whereas the methods introduced in chapter three are limited to special geometries, in this part a method is introduced which permits a very general application of discrete dislocation mechanics. The fifth contribution deals with the formation of a dislocation pattern. It shows the newest results, how the collective behaviour of dislocations can be treated within a statistical physical framework

The sixth contribution gives an introduction to phenomenological crystal plasticity models, to their properties and their use either in component calculations or in microstructurally modelling. The last contribution, presented in the Multiscale Modelling and Design of New Materials-course is addressed to the key components of the computational homogenization scheme, i.e. the formulation of the microstructural boundary value problem and the coupling between the different levels based on the averaging theorems. With the detailed analyses of the different scales and bridging we do hope that the presented volume will contribute useful source of information on the fundamentals and the present-day techniques to model plastic deformation and fracture on the nano- and micro-scale.

It is the pleasure of the editors to express our gratitude to the Scientific Council of CISM for supporting the course as well as permitting these publications. We thank all the authors for their contributions and the fruitful discussions with the participants.

Peter Gumbsch and Reinhard Pippan

CONTENTS

Atomistic Simulation Methods and their Application on Fracture

Bernhard Eidel [*], Alexander Hartmaier [*] and Peter Gumbsch [†]

[*] Interdisciplinary Centre for Advanced Materials Simulation (ICAMS), Ruhr-Universität Bochum, Stiepeler Strasse 129, 44801 Bochum, Germany

[†] Institut für Zuverlässigkeit von Bauteilen und Systemen, Universität Karlsruhe and Fraunhofer Institut für Werkstoffmechanik, Wöhlerstr. 11, 79108 Freiburg, Germany

Abstract *The present work on the molecular dynamics method covers the theoretical background of the method and gives practical examples to demonstrate its capabilities and limitations. The work focusses on topics which reveal fundamental mechanisms associated with fracture processes. Moreover, promising hybrid methods based on a concurrent atomistic/continuum coupling are reviewed since they combine accuracy and efficiency in a most favorable manner.*

1 Introduction

For many engineering questions connected to the mechanical properties of materials, one can of course profitably apply continuum mechanical descriptions of materials behaviour. However, when it comes to describing small specimens or when material specific questions need to be addressed, it is usually indispensable to investigate defect properties. For plastic deformation, discrete dislocation simulations (see (Kubin et al., 1992; Devincre and Roberts, 1996; Deshpande et al., 2003; Weygand et al., 2002; Weygand and Gumbsch, 2005) or the article by V. Mohles in this collection) can be applied. In these simulations the interaction of dislocation with interfaces determines the influence of the microstructures. These discrete defect based methods, however, require governing laws for the individual defect properties and more detailed descriptions for the short range interaction of defects. Such properties are difficult to obtain experimentally and are therefore usually investigated by atomistic methods.

Therefore, atomistic modelling is essential in advancing our understanding of the mechanical properties of materials. This is most obvious for the investigation of fracture processes. Materials behaviour with respect to fracture is of course ultimately determined by events on the atomic scale.

In the case of brittle fracture this connection is obvious, since the crack in a perfectly brittle material must be atomically sharp at its tip. The crack moves by breaking individual bonds between atoms and can therefore be regarded as a macroscopic probe for the atomic bonding. The transition from perfectly brittle to ductile behaviour similarly relies on atomistic processes since the multiplication or nucleation of dislocations at the crack tip is an indispensable ingredient of any modelling of the brittle to ductile transition. Atomistically the nucleation or multiplication of dislocations at the crack tip is identified with bond shearing events, which compete with the bond breaking events at crack extension.

The interaction of multiple defects, like cracks and dislocations, is still not routinely studied even with the simplest atomistic simulation methods since such studies necessarily require models that significantly extend in all three dimensions and consequently require the handling of millions of atoms. Since the application of atomistic techniques has great potential (Li et al., 2003; Kassner et al., 2005) but is still not widespread in the investigation of the mechanical properties of materials, it is probably advisable to give a short overview on the applicable methods and then to provide a few examples of successful application of atomistic modelling techniques to explain experimentally observed phenomena.

In the following section we give a general overview of the different descriptions for the atomic bonding, because they are the basis for all atomistic modelling. The next section describes the molecular dynamics method in some detail, also including the boundary conditions and visualisation of defects. This part is intended to give the reader, who is not familiar with atomistic simulation methods a feeling for the versatility of this approach. The capabilities and limitations of atomistic methods are not just related to the available description of the atomic bonding but also to the handling of boundary conditions and the analysis of the results. The subsequent section 4 deals with concurrent multiscale methods, which couple atomistic and continuum descriptions in a rather seamless manner, hence enabling the description of crystalline solids with atomistic accuracy but at smaller computational costs. The Finite Element Atomistic Method (FEAt) and variants of the Quasicontinuum (QC) method are reviewed and compared.

After this methods-oriented part, in Section 5 the phenomenology of fracture is discussed in the light of atomistic modelling. First, brittle fracture is put in the spotlight, which is the domain of the atomistic methods, as we have seen above. Second, it is shown that the analysis of the plastic zone around a crack tip is now within the scope of large scale atomistic simulations.

It is noted here that this article can neither provide a practical guide

for the use of atomistic methods nor can it claim to give an exhaustive overview of all the work where these methods provide insight into atomistic processes of fracture. It is rather meant to be a first introduction into the topic, highlighting a few illustrative examples and showing the possibilities and limitations of the most frequently used methods. Interested readers should also consult recent overviews on the subject (Li et al., 2003; Kassner et al., 2005) for more details and for additional references.

In order to limit the references to a useful number, we restrict ourselves to citations concerning models and numerical simulations to the literature dealing with fracture, where adequate. Furthermore, if the same authors published a number of articles on similar topics in fracture, the most recent work is usually given here since it will often provide guidance to previous work.

2 Description of Interatomic Bonds

In this section the different atomistic methods that have been used to model fracture processes are briefly introduced. The scope of this section is to give a general introduction into the basic idea behind the atomistic, *i.e.* non-continuum, method being used to model fracture of materials, to introduce the essential terms, and to provide the basic literature for further reading. No attempt is made to provide cooking recipes enabling the reader to implement such methods solely based on this text.

Before presenting the different methods to describe the atomic interaction, it is worth mentioning that atomistic modelling may be applied with very different intentions. In some cases, the atomistic simulations are used as testing grounds for ideas about the behaviour or energetics of defects. It may then be perfectly justified to use the simplest generic form of the interaction model. Alternatively, the goal may be to make quantitative or semiquantitaive predictions about the properties of specific materials. In this case one has to resort to materials-specific models or even quantum mechanical *ab initio* methods.

2.1 Quantum Mechanics Based Descriptions of the Atomic Bonding

The most fundamental description of interatomic bonds that may have ionic, covalent, metallic or van-der-Waals character, or any mixture of these pure bond types, is given by explicitly dealing with all the electrons involved in the formation of the atomic bonds. The quantum mechanical description of a crystalline material is a classical problem of solid state physics and

appropriate tools have been developed there, which today are available for the use in many different application areas. In all these codes, the many-body problem of atomic nuclei and electrons must be solved for different geometries. Lattice periodicity is usually imposed. The solution is generally approached in three steps:

1. Within the Born-Oppenheimer approximation the motion of the atomic nuclei is (adiabatically) decoupled from the motion of the electrons. Since the mass and the inertia of the atomic nuclei is orders of magnitude larger than the mass of the electrons, the electronic quantum gas follows the motion of the nuclei almost instantaneously. Furthermore the electrons are always assumed to remain in their ground state with respect to the momentary position of the nuclei.
2. The quantum mechanical ground state of the inhomogeneous electron system in the system of the nuclei is determined then in an external electrostatic potential, given by the charge at the instantaneous position of the nuclei. In practical calculations, the many particle state of the interacting electrons is usually constructed from single particle states of non-interacting electrons in an effective (mean) field of all the other electrons. These single particle states are self consistently determined by iteratively solving the coupled set of single particle problems for all the electrons. The most successful approach to this problem has been *density functional theory (DFT)* (Hohenberg and Kohn, 1964; Kohn and Sham, 1965). Within the *local density approximation* (LDA) it is assumed that the exchange correlation of the electrons can be calculated based purely on the local electron density. This is sometimes not enough and generalised gradient approximations (GGA) have been introduced.
3. In the last step we now reintroduce the motion of the atomic nuclei. The ground state energy of the electron gas can be regarded as the potential energy for the nuclei in their configurational space. The atomic motion can then be simulated in this *adiabatic* potential. Local minima in this adiabatic potential reflect statically stable structures for the solid while saddle points are the static barriers for structural transformations. With the knowledge of the adiabatic potential, it is possible to determine kinetic, dynamic or statistical materials properties (e.g. phonon properties, transport properties, or phase equilibria) either by classically solving Newtons equations of motion, by minimising the total energy of a structure or by solving quantum mechanical problems for the nuclei (e.g. for zero point vibrations of light elements).

The main task of electron theory is to solve for the electronic ground state, the second step above, in a numerical way. There are several successful methods available to do so, based on plane waves, localised orbitals or a mixture of them. The Cambridge (Serial) Total Energy Package CASTEP, the Vienna Ab-initio Simulation Package VASP or the Mixed Basis Pseudo-Potential Code MBPP are standard packages for this task.

Since these quantum mechanical methods do not require any adjustable parameters and are therefore often termed *ab initio* or *first priciples* techniques. Their results provide highly accurate and predictive results on materials properties.

However, the methods are computationally very demanding and often require periodic structures, which are constructed by defining a unit cell comprising the atoms to be considered and repeating this unit cell periodically in all directions. This procedure is identical to the periodic boundary conditions that are used extensively in molecular dynamics simulations, as will be described below. However, the periodic cells in molecular dynamics simulations may contain several million atoms, while *ab initio* methods, due to their mathematical complexity are restricted to hundreds of atoms within the unit cells. This limitation is rather severe for applications in mechanical problems, because the generation of defects or the driving force on a defect, like the energy release rate during crack advance, all depend on the total elastic energy stored within the volume. Hence, if the volume under consideration is extremely small, the elastic strains in the volume have to be close to their theoretical limits in order to store sufficient elastic energy to drive the defects. Therefore, *ab initio* simulations of defect behaviour must always be carefully checked for size effects and artifacts caused by the restrictions on the effective volume. Notwithstanding these words of warning, it must be stated again that *ab initio* methods are the most fundamental methods available to describe behaviour of materials and thus possess the most predictive power. There are also some interesting and potentially fruitful ideas to combine *ab initio* methods with molecular dynamics simulations which will be described in section 4.6.

Before leaving this section it is worth mentioning that *ab initio* methods are not just applied to study some specific fracture problem, they also play an important role in the verification and the adjustment of simpler models of the atomic interaction. To develop material-specific (semi-) empirical interaction models, various adjustable parameters and sometimes even the functional form of the interaction model has to be chosen to reproduce available data. However, experiment will usually only provide data close to mechanical and thermodynamic equilibrium, whereas the interaction models will be used at grain boundaries or even in the highly strained region near

crack tips. Therefore it is advisable to adjust the parameters of the interaction models so that they perform well under these circumstances. Because of the lack of experimental data, *ab initio* calculations of large deformations or structural energy differences are often used instead. It therefore turned out to be very useful to build up extensive *ab initio* data bases for the development of simpler interaction models. While the first such data bases for aluminium (Ercolessi and Adams, 1994) or the intermetallic nickelaluminide B2-NiAl (Ludwig and Gumbsch, 1995) still constituted demanding calculations, such databases can today be generated routinely and very systematically.

2.2 Atomic Interaction Models, Potentials

While it is generally possible to use *ab initio* methods to study the dynamic evolution of an atomistic system, the computational burden is usually too high and (semi-)empirical potentials are therefore applied in molecular dynamics (MD) simulations. The atomic interaction as described by semi-empirical potentials always depends on the distance of interacting pairs of atoms and, for the more elaborate potentials, also on the bond angles and the local electron density. All these distance or angle dependent functions are represented either analytically or as tabulated functions of these parameters. The most simple pair potentials, like Lennard-Jones or the Morse potential, have just two or three free parameters that are used to change the characteristic properties of the potential like the lattice parameter, the bulk elastic modulus or the cohesive energy. More sophisticated potentials have a number of free parameters that yield much better results for material specific properties like lattice constant, sublimation energy, anisotropic elastic constants, vacancy formation energy, stacking fault energy, or whatever is felt necessary for a certain investigation. The free parameters of the potentials are usually fitted to material properties by calculating exactly these properties on rather small atomic ensembles and then varying the potential parameters until a reasonable match to experimental or *ab initio* data is reached. The interatomic potentials can roughly be categorized into three classes according to their level of approximation of the "real" quantum mechanical atomic interaction:

- angularly-dependent tight-binding or bond-order potentials
- multibody or embedded atom method (EAM) potentials
- pair potentials

The properties and characteristics of the different potentials are given in some detail below. For application purposes, the choice of a potential will usually involve a trade-off between the accuracy of the description of the atomic interaction from first principles calculations over semi-empirical po-

tentials to simple pair potentials and the size of the system that can be studied, *i.e.* the number of atoms that can be explicitly treated in the sample. The decision, whether higher accuracy in the description of the interatomic forces or larger system size, has to be made for each problem under consideration. For example, during fracture of covalently bonded systems like silicon, polymers or biological matter the behaviour of the entire system depends critically on the behaviour of the individual atomic bonds immediately in front of the crack tip (see Section 5.1). Here it becomes necessary to accurately evaluate the interatomic forces in order to obtain meaningful results. In contrast, for rather generic investigations of dynamic crack stability as a function of crack driving force, where the energy release rate and the stored elastic energy in the solid are decisive, it may be essential to have a reasonably large system (Gumbsch et al., 1997; Buehler et al., 2003). Even in these cases, however, materials-specific questions must again be treated with the more accurate DFT descriptions or coupled methods (Kermode et al., 2008).

Pair potentials. Pair potentials are the simplest form of interatomic potentials. This approach limits the interaction between two atoms to a dependence on their mutual distance, thus excluding completely any information about neighbouring atoms. Pair potentials in their most general form are written as

$$U_{\mathrm{tot}} = \frac{1}{2} \sum_{i,j \neq i} V_{ij}(r_{ij}) + U(\Omega) \tag{1}$$

where V_{ij} is the pair potential, r_{ij} is the separation between atoms i and j and $U(\Omega)$ is a contribution to the cohesive energy which depends on the average volume per atom Ω. This latter term may mainly determine the cohesive energy, while the pair potential always determines the structural dependence.

Often, however, it is not attempted to mimic the properties of a particular material but more generic atomistic questions are of interest. Then, potentials like the Lennard-Jones potential or the Morse potential are applied. These have only few adjustable parameters, which are fitted to the nearest neighbour distance and the binding energy or the bulk modulus. Despite their simplicity and the lack of materials specificity, pair potentials can contribute significantly to our understanding of material behaviour and fracture in particular. Their simple structure permits investigation of generic effects, e.g. of the atomic size or of the role of elastic properties for a certain phenomenon. Other examples are when the consequences of the discrete crystal structure of solids shall be compared to predictions of

continuum models or when the influence of non-linearities in the elastic behaviour shall be mentioned that can ideally be compared with linear-elastic solids described by a harmonic pair potential (Buehler et al., 2003). In these examples use is made of the simplicity of pair potentials that allows clean studies of single effects, while excluding all other influences that render the behaviour of "real" materials so much more complex.

Of course, pair potentials cannot be used to describe situations where directional bonds or bond angles play a role. They are also not well suited to describe metals because the well established dependence of bond strength on coordination cannot be represented. Furthermore, simple pair potentials result in extremely low stacking fault energies, because they fail to distinguish the energy difference between face centred cubic (fcc) and hexagonal close-packed (hcp) structures.

Embedded atom method (EAM) potentials. The most widely used interaction models for metals are the EAM potentials. In this scheme the volume-dependent term from Equation (1) is expressed as a local density-dependent contribution to the total potential energy

$$U_{\mathrm{tot}} = \frac{1}{2} \sum_{i,j \neq i} V_{ij}(r_{ij}) + \sum_{i} F(\rho_i) \tag{2}$$

with

$$\rho_i = \sum_{j \neq i} \rho_{ij}(r_{ij}) \tag{3}$$

where ρ_{ij} can be viewed as the contribution from atom j to the total electron density at atom i and F is the embedding energy associated with placing atom i in this environment. Finally, V_{ij} is the pair potential contribution to the potential energy of atom i. Different physical interpretations of the terms are possible and consequently the functional forms and the way in which the various parameters are determined may differ. Usually the functions are adjusted to reproduce at least the lattice parameter, the cohesive energy and the anisotropic elastic constants of the metal under consideration. For alloys it is additionally necessary to adjust structural energy differences of intermetallic phases and heats of solution (Ludwig and Gumbsch, 1995). As mentioned above, it is also desirable to not only adjust to such equilibrium crystal properties, but to also compare to some *ab initio* data from atomic structures far from equilibrium or for large deformations. Reliable and well tested potentials are available for the noble metals and nickel as well as for some aluminium alloys (e.g. (Ercolessi and Adams, 1994; Ludwig and Gumbsch, 1995; Mishin et al., 2001)).

The main advantage of EAM potentials is that they can be applied to inhomogeneous systems such as surfaces or cracks, since these potentials approximately incorporate the variation of bond strength with coordination. Decreasing the coordination (density) at an atom usually increases the strength of each of its bonds and decreases the bond length. In particular for fcc materials it has been easily possible to reproduce anisotropic elastic constants, sublimation energy, vacancy formation energy, surface and stacking fault energies with EAM potentials and EAM potentials can be applied to many different types of simulations. For body centred cubic (bcc) metals the method may still give a sufficiently precise description. For some of these metals, however, it has been shown that the EAM fails to give a valid description of the core structure of dislocations (Mrovec et al., 2004) and thus the Peierls stress to move the dislocation through the lattice. Therefore, these potentials cannot be used reliably to model plasticity in these materials. This may be seen as an indication of the importance of angular bonding characteristics and one may consequently resort to tight binding or bond order descriptions (Pettifor and Oleinik, 1999; Mrovec et al., 2004, 2007a).

Tight binding and bond order potentials. The need for material specific models and for accuracy in atomistic simulations implies that a quantum mechanical description of the atomic interaction is required. The semi-empirical tight binding scheme is such an approximate quantum mechanical description of the energetics of systems of atoms (Harrison, 1980; Pettifor, 1995). The total energy of a system is given by:

$$U_{\mathrm{tot}} = U_{\mathrm{rep}} + U_{\mathrm{bond}} \tag{4}$$

where U_{rep} is a repulsive energy, generally given as a sum of pair potentials and U_{bond} is the bonding part of the energy. The latter is obtained by solving the eigenvalue problem for a given Hamiltonian, which is assumed to be fixed and not evaluated self-consistently as in the *ab initio* methods.

The Hamiltonian matrix elements are usually assumed to be rapidly decaying functions of the atomic separation, which have to be empirically adjusted to experimental data or results of *ab initio* calculations to give a material-specific model. The least clear but most important question in this context is the transferability of these matrix elements from ideal structures to highly distorted atomic environments.

The semi-empirical tight-binding scheme is in principle applicable to materials with various types of bonding but is most naturally suited to covalently bonded insulators and semiconductors as well as the transition metals and intermetallics .

The key computational aspect in solving a tight-binding model is the diagonalisation of the eigenvalue problem. This can of course be attempted in reciprocal space, where a whole set of very efficient methods is available, which however usually scale to the third power with the number of atoms in the system. Linear scaling can be achieved with real space methods of which the density matrix method for semiconductors and insulators and the bond order approach for metals and alloys are the most promising approaches. In the simplest approximations the latter can even be formulated analytically within the framework of bond order potentials (Pettifor and Oleinik, 1999) where the energy of the system is written as

$$U_{\mathrm{tot}} = \frac{1}{2} \sum_{i,j \neq i} V_{ij}(r_{ij}) + \sum_{i\alpha,j\beta} H_{i\alpha,j\beta} \Theta_{i\alpha,j\beta} \tag{5}$$

where V_{ij} is the pair potential, depending on the type of atom and r_{ij} is the separation between atoms i and j. H is the Slater Koster hopping integral, which also depends on distance r_{ij} and where the Greek indices stand for the type of orbital. Θ is the bond-order matrix which gives the difference in the number of electrons in the bonding and antibonding states. Modern bond order potentials for carbon (Mrovec et al., 2007b) and some transition metals (Mrovec et al., 2004, 2007a) are just becoming available.

In the same spirit as the bond-order potentials, several simplified angularly-dependent potentials of Tersoff and Brenner-type have been developed for silicon and the hydrocarbons (see for example (Pastewka et al., 2008)). These potentials have seen continuous improvements but were notoriously difficult to adjust to bond breaking problems. Despite recent success with an explicit screening formulation for the hydrocarbons (Pastewka et al., 2008) giving up specific bonds in a way which does not require a self-consistent formulation of the electronic configuration remains an important and difficult problem.

3 The Molecular Dynamics Method

To continue we describe the pertinent methods used in atomistic simulations with some focus on the application in fracture processes and related problems. We start out in the present section with the Molecular Dynamics (MD) method. A short overview of the main ingredients like time integration and relaxation is given, furthermore the role of boundary and initial conditions is described and illustrated for mechanical problem sets in materials science. Section 3.6 gives a brief account of the visualisation and analysis of defects in atomistic simulations, which plays an ever more important role in computational materials science. Some popular and powerful

defect indicators are described. In Section 4 we analyse and compare two concurrent multiscale methods.

The main ingredients of an MD simulation are basically threefold:

(i) A model is needed for the interaction between the system constituents (e.g. atoms or molecules).

(iia) Time integration is required to advance the particle trajectories (positions and velocities) from time t to $t + \Delta t$.

(iib) Alternatively one may want to solve a stability problem which in an atomistic system requires an algorithm to relax the atomic coordinates to positions of vanishing forces.

(iii) An ensemble has to be chosen, for which boundary conditions and thermodynamic quantities like temperature, pressure or the number of particles are controlled.

3.1 Force Calculation

Forces are derived from the potential energy U that depends on the positions of all atoms. The description for the calculation of the energy can be based on different physical approximations as described above. The force acting on an atom i is given by taking the derivative of the potential energy with respect to the position vector $\boldsymbol{x}_i$ of atom i

$$\boldsymbol{f}_i = -\frac{dU(\boldsymbol{x})}{d\boldsymbol{x}_i}\,, \tag{6}$$

where $\boldsymbol{x}$ denotes the coordinates of all atoms. Once the force vector $\boldsymbol{f}_i$ acting on all atoms is known, the Newtonian equation of motion

$$\boldsymbol{f}_i = m_i \frac{d^2\boldsymbol{x}_i}{dt^2}\,, \tag{7}$$

can be integrated in time t to yield the motion of the atoms in space. The mass of the atom is given by m_i.

3.2 Integrating the Equations of Motion

Equation (7) constitutes a set of second-order ordinary differetial equations (ODEs), which can be strongly nonlinear. By converting them to first-order ODEs in the $6N$-dimensional space of $\{\boldsymbol{x}_N, \dot{\boldsymbol{x}}_N\}$, general numerical algorithms for solving ODEs such as the Runge-Kutta method could be applied. However, these general methods are rarely used in practice,

because the existence of a Hamiltonian allows for much simpler and even more accurate integration algorithms.

To represent other thermodynamic ensembles than the *micro-canonical* ensemble for which Equation (7) can be integrated directly, requires that Equation (7) is modified to create a dynamics in phase space that has the desired distribution density of e.g. a *canonical* or a *grand-canonical* ensemble (see e.g. (Frenkel and Smit, 2002)). The time-average of a single-point operator on such a trajectory then approaches the thermodynamic average.

An integrator serves the purpose of propagating particle positions and velocities over small time increments Δt.

$$\boldsymbol{x}^{3N}(t_0) \rightarrow \boldsymbol{x}^{3N}(t_0 + \Delta t) \rightarrow ... \rightarrow \boldsymbol{x}^{3N}(t_0 + L\,\Delta t) \tag{8}$$

The time step Δt has to be chosen such that the thermal oscillations of the atoms around their equilibrium positions are resolved in time. A typical frequency of this oscillation is the Debye frequency $\nu_D = c_t/a$, where c_t is the speed of transverse sound waves and a is the lattice parameter. A typical value for metals is $\nu_D \approx 10^{13}\,\mathrm{Hz}$. This implies that the typical time step for MD simulations has to be on the order of femtoseconds ($=10^{-15}\,\mathrm{s}$), which generally limits the method to simulations of fast processes such as brittle fracture or high-strain-rate plastic deformation.

Some popular time integration algorithms are the central difference algorithms: Verlet, velocity Verlet or leap-frog. They are shortly introduced below.

• **Verlet algorithm.** Assuming that the $\boldsymbol{x}^{3N}(t)$ trajectories are smooth, one may perform a third-order Taylor expansion of the positions $\boldsymbol{x}_i(t_0)$ forward ($\boldsymbol{x}_i(t_0 + \Delta t)$) and backward ($\boldsymbol{x}_i(t_0 - \Delta t)$) in time; their sum yields

$$\boldsymbol{x}_i(t_0 + \Delta t) + \boldsymbol{x}_i(t_0 - \Delta t) = 2\boldsymbol{x}_i(t_0) + \ddot{\boldsymbol{x}}_i(t_0)(\Delta t)^2 + O((\Delta t)^4). \tag{9}$$

Since $\ddot{\boldsymbol{x}}_i(t_0) = \boldsymbol{f}_i(t_0)/m_i$ can be evaluated given the atomic positions at $t = t_0$, $\boldsymbol{x}^{3N}(t + \Delta t)$ in turn may be approximated by,

$$\boldsymbol{x}_i(t_0 + \Delta t) = -\boldsymbol{x}_i(t_0 - \Delta t) + 2\boldsymbol{x}_i(t_0) + \frac{1}{m}\boldsymbol{f}_i(t_0)(\Delta t)^2 + O((\Delta t)^4). \tag{10}$$

Neglecting the $O((\Delta t)^4)$ term, we obtain a recursion formula to compute $\boldsymbol{x}^{3N}(t_0 + \Delta t)$. Although velocities are not needed in the recursion, they are often calculated since they are required for analysis of ensemble properties. They can be approximated by

$$\boldsymbol{v}_i(t_0) \equiv \dot{\boldsymbol{x}}_i(t_0) = \frac{1}{2\Delta t}[\boldsymbol{x}_i(t_0 + \Delta t) - \boldsymbol{x}_i(t_0 - \Delta t)] + O((\Delta t)^2). \tag{11}$$

This algorithm is not only one of the simplest, but also a good choice in general. It is fast, but not particularly accurate for long time steps, such that the forces on all particles must be computed rather frequently. It requires about as little memory as is at all possible. This is useful when very large systems are simulated. Verlet's short-term energy conservation is fair but, more important, it exhibits little long-term energy drift. This is related to the fact that the Verlet algorithm is time reversible and area preserving. In fact, although the Verlet algorithm does not conserve the total energy of this system exactly, strong evidence indicates that it does conserve a pseudo-Hamiltonian approaching the true Hamiltonian in the limit of infinitely short time steps.

• **Velocity-Verlet algorithm.** It starts with $\boldsymbol{v}^{3N}(t_0)$ and $\boldsymbol{x}^{3N}(t_0)$. One then evaluates

$$\boldsymbol{x}_i(t_0+\Delta t) = \boldsymbol{x}_i(t_0) + \boldsymbol{v}_i(t_0)\Delta t + \frac{1}{2}\frac{\boldsymbol{f}_i(t_0)}{m_i}(\Delta t)^2 + O((\Delta t)^3), \tag{12}$$

with $\boldsymbol{f}^{3N}(t_0+\Delta t)$ evaluated from $\boldsymbol{x}_i(t_0+\Delta t)$ one gets

$$\boldsymbol{v}_i(t_0+\Delta t) = \boldsymbol{v}_i(t_0) + \frac{1}{2}\left[\frac{1}{m_i}\boldsymbol{f}_i(t_0) + \frac{1}{m_i}\boldsymbol{f}_i(t_0+\Delta t)\right]\Delta t + O((\Delta t)^3), \tag{13}$$

and has advanced by one step. This algorithm requires a little more computing but is very popular since it gives $\boldsymbol{x}^{3N}$ and $\boldsymbol{v}^{3N}$ simultaneously.

• **Leap-frog algorithm.** In the leap-frog algorithm, position and velocities are calculated with the same accuracies but are offset by $\Delta t/2$. It starts with $\boldsymbol{v}^{3N}(t_0-\Delta t/2)$ and $\boldsymbol{x}^{3N}(t_0)$. Time integration is then first done on $\boldsymbol{v}$

$$\boldsymbol{v}_i(t_0+\frac{1}{2}\Delta t) = \boldsymbol{v}_i(t_0-\frac{1}{2}\Delta t) + \frac{1}{m_i}\boldsymbol{f}(t_0)\Delta t + O((\Delta t)^3), \tag{14}$$

followed by integration of $\boldsymbol{x}$,

$$\boldsymbol{x}_i(t_0+\Delta t) = \boldsymbol{x}_i(t_0) + \boldsymbol{v}_i(t_0+\frac{1}{2}\Delta t) + O((\Delta t)^3). \tag{15}$$

It can be shown that the leap-frog algorithm produces identical trajectories to the Verlet algorithm besides numerical rounding errors. It therefore has similar properties than the Verlet algorithm but of course provides coordinates and velocities at once.

For the description of *Predictor-Corrector Algorithms* and *Symplectic Integrators* we refer to standard text-books (Frenkel and Smit, 2002; Schlick, 2002; Rapaport, 2004).

3.3 Relaxation Algorithms

Mechanically stable configurations or thermally activated processes are extremely difficult if not impossible to assess with MD methods. Examples for such difficult-to-study processes in the context of fracture would be void formation and coalescence or motion of dislocations in lattices with high Peierls barriers (Marian et al., 2004). For these cases it is usually advisable to map out the energy landscape using relaxation algorithms and then to perform a metadynamics on the basis of such an energy landscape.

Mechanically stable equilibrium configurations correspond to minima of the total energy. Hence it is a most common task in computational materials science and solid state physics to find local or global minima of the potential energy, where a given initial configuration is the point of departure. To solve this task a variety of well-established optimisation methods are available (see e.g. in (Nocedal and Wright, 2006; Leach, 2001; Schlick, 2002)). The optimisation methods generally can be classified according to the highest order derivative used to minimise the (energy) functional. Hence, a non-derivative minimisation method like the simplex method can be considered as zeroth-order method. Zeroth-order methods are rarely used in molecular modelling since first derivatives of the energy (i.e. forces) are usually available and these methods then do not exploit all the available information. In first-order minimisation methods, the gradient of the energy indicates the direction to a minimum, its magnitude measures the steepness of the local slope. These methods are frequently used in molecular modelling. Prominent examples are the steepest descent method and variants of the conjugate gradient (CG) method. For a nice overview we refer to (Shewchuk, 1994). Most recent MD integration-based methods that also fall into this category are described below. Second order methods additionally use the second derivatives and thus the information of the local curvature to locate a minimum. Current state-of-the-art methods like the limited-memory version of the Broyden-Fletcher-Goldfarb-Shanno scheme (l-BFGS) (Nocedal and Wright, 2006) explicitly use only first order derivatives but accumulate information to obtain some approximate representation for the Hessian matrix to determine line search directions. Second order methods are by far not as robust as first order methods and can therefore only be used in molecular modelling if no structural changes occur.

MD-based methods, FIRE. Within the realm of molecular simulations, some methods have been proposed that serve the purpose of energy minimisation, that start out with MD and proceed in removing kinetic energy from the system, termed 'quenching'. This strategy has been successfully

applied in local minimisation as well as in global minimisation. In (Bitzek et al., 2006) a simple, yet powerful MD scheme for structural relaxation was proposed which belongs to this family of minimisers. Different from existing schemes this new algorithm crucially relies on inertia, it has therefore been named the Fast Intertial Relaxation Engine (FIRE). In the original paper the method's functional principle was explained by means of a blind skier searching for the fastest way to the bottom of a valley in an unknown mountain range described by the potential energy landscape $E(\boldsymbol{x})$ with $\boldsymbol{x} = (x_1, x_2)$. Assuming that the skier is able to retard and steer, the recommended strategy for the skier is to follow an equation of motion given by:

$$\dot{\boldsymbol{v}}(t) = 1/m\, \boldsymbol{F}(t) - \gamma(t)|\boldsymbol{v}(t)|[\hat{\boldsymbol{v}}(t) - \hat{\boldsymbol{F}}(t)], \tag{16}$$

with the mass m, the velocity $\boldsymbol{v} = \dot{\boldsymbol{x}}$, the force $\boldsymbol{F} = -\nabla E(\boldsymbol{x})$, and hat denoting unit vector. The recommended strategy is that the skier introduces acceleration in a direction that is 'steeper' than the current direction of motion via the function $\gamma(t)$, if the power $P(t) = \boldsymbol{F}(t) \cdot \boldsymbol{v}(t)$ is positive, and in order to avoid uphill motion he simply stops as soon as the power becomes negative. $\gamma(t)$ must be chosen appropriately but should not be too large, because the current velocities carry information about the reasonable 'average' descent direction and energy scale. A discretised version of this equation in combination with an adaptive time step results in a minimisation scheme for multidimensional functions $E(x_1; \ldots x_M)$ which is competitive in speed with the fastest optimisers currently available (Bitzek et al., 2006), but has also other important features as we shall demonstrate.

The numerical treatment of the algorithm is simple. Any MD integrator can be used as the basis for propagation of the trajectories due to the conservative forces. The MD trajectories are continuously readjusted by two kinds of velocity modifications: (a) the above-mentioned immediate stop upon uphill motion and (b) a simple mixing of the global velocity and force vectors $\boldsymbol{v} \to (1-\alpha)\boldsymbol{v} + \alpha\hat{\boldsymbol{F}}|\boldsymbol{v}|$ resulting from an Euler-discretisation of the last term in Equation (16) with time step Δt and $\alpha = \alpha_{\text{start}}$. Both, Δt and α, are chosen adaptively on the fly.

The propagation rules for the FIRE algorithm can be summarised as follows (given: initial values for Δt, $\alpha = \alpha_{\text{start}}$ and the global vectors $\boldsymbol{x}$ and $\boldsymbol{v} = \boldsymbol{0}$):

1. MD integrator: calculate $\boldsymbol{x}$, $\boldsymbol{F} = -\nabla E(\boldsymbol{x})$ and $\boldsymbol{v}$ using any common MD integrator; check for convergence.
2. calculate $P = \boldsymbol{F} \cdot \boldsymbol{v}$.
3. set $\boldsymbol{v} \to (1-\alpha)\boldsymbol{v} + \alpha|\boldsymbol{v}|\hat{\boldsymbol{F}}$.

4. if $P > 0$ and the number of steps since P was negative is larger than N_{min}, increase the time step $\Delta t \rightarrow \mathrm{min}(\Delta t f_{\mathrm{inc}}, \Delta t_{\mathrm{max}})$ and decrease $\alpha \rightarrow \alpha f_{\alpha}$.
5. if $P \leq 0$, decrease time step $\Delta t \rightarrow \Delta t f_{\mathrm{dec}}$, freeze the system $\boldsymbol{v} \rightarrow \mathbf{0}$, and set α back to α_{start}.
6. Return to MD.

In relaxation an accurate calculation of the atomic trajectories is not necessary, and the adaptive time step allows FIRE to increase Δt until either the largest stable time step Δt_{max} is reached, or an energy minimum along the current direction of motion ($P < 0$) is encountered. In the latter case the system is instantly frozen ($\boldsymbol{v} \rightarrow \mathbf{0}$) and the time step is substantially reduced in order to have a smooth restart. A short 'latency' time of N_{min} MD steps before accelerating the dynamics is important for the stability of the algorithm.

In (Bitzek et al., 2006) it is shown in several benchmark sets that FIRE is very competitive with sophisticated algorithms like the l-BFGS scheme and easily beats even advanced versions of the CG method (e.g. Polak-Ribière). The key advantage of the FIRE algorithm, however, is its extreme robustness. It finds the (local) minima even for extreme structural rearrangements like molecular folding and rotation or for atomic reconstructions in the core of crystalline defects and therefore lends itself ideally as a general purpose minimiser.

3.4 Boundary and initial conditions

Initial conditions. The integration of Newton's equations of motion requires an integrator and initial conditions (IC), namely $\boldsymbol{x}^{3N}(t = 0)$ and $\dot{\boldsymbol{x}}^{3N}(t = 0)$, the initial particle positions and velocities.

Generating $\boldsymbol{x}^{3N}(t = 0)$ for crystalline solids is easily done by a structure generator setting up a perfect crystal or an interface between two crystalline phases. Generating suitable velocity distributions or structures for a liquid or an amorphous solid is significantly more difficult. It can, however, often be circumvented by running the system from an artificial set-up for an extended equilibration time. One can for example melt a crystal and obtain the IC for an amorphous configuration then by quenching from the liquid.

Boundary conditions. Boundary conditions can be classified into two major types: periodic boundary conditions (PBC) and isolated (or free) boundary conditions (IBC). IBC, in which surface atoms exhibit dangling bonds due to a lack of neighbours, are chosen for the analysis of surfaces, clusters and molecules. In addition, there can be extra forces or displace-

ments acting on the boundary atoms, resulting in so-called mechanical boundary conditions.

PBC are intended to mimic bulk solids or liquids, or the cores of very large systems, which are much larger than the simulated number of particles. The particles are contained within a primary simulation volume. This volume is sometimes referred to as the simulation box, unit cell or supercell.

When a particle leaves one side of this volume, it re-enters from the opposite side keeping the number of atoms in the central box constant. Atoms sitting in the vicinity of one side of the box, through periodic repetition of the box, are connected to the atoms on the other side of the box. The simulations therefore proceed as if the primary volume was surrounded in all directions by identical copies of itself to form a quasi-infinite volume. PBC for the two-dimensional case are illustrated in Figure 1. If PBC are used, the case that the cutoff radius is less than half the diameter of the periodic box is of special interest since in that case only the interaction of a given atom with the nearest periodic image of any other atom needs to be considered (*minimum image convention*). This case is displayed in Figure 1, where the dotted box comprises all nearest images of any other atom.

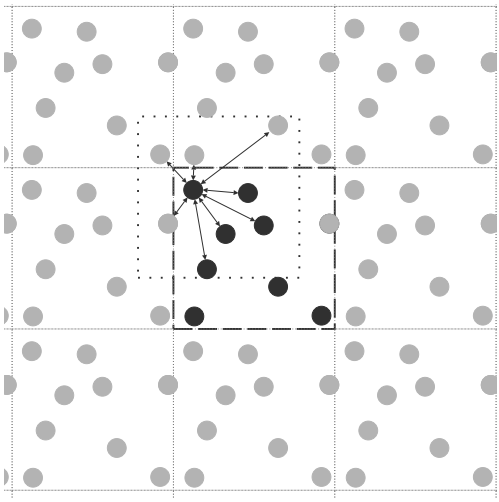


Figure 1. Periodic boundary conditions: the basic simulation volume (drawn in *black*) is repeated in all dimensions. Here, particles in neighbour volumes are drawn in *gray*. Interactions of one atom with its neighbours are indicated with arrows.

The cubic cell is the simplest geometry of an unit cell to visualise and

to program. Of course, any other shape can be chosen provided it fills all of space by translation operations. Five shapes satisfy this condition: the cube (and its close relation, the parallelepiped), the hexagonal prism, the truncated octahedron, the rhombic dodecahedron and the elongated dodecahedron, see e.g. (Leach, 2001).
Using PBC, surface effects can be eliminated from the simulation, but some inherent limitations dictate rules in applying them, see (Pöschel and Schwager, 2005).

(i) Since each particle or defect has an infinite number of images, there is always an infinite number of interactions with each other particle or defect.

(ii) If the chosen primary volume is too small, there appear correlations between opposite edges of the primary volume. If the spatial structures are of the same characteristic size as the system itself, a particle may interact directly or indirectly with itself across the primary volume. It is therefore not possible to achieve fluctuations that have a wavelength greater than the length of the cell.

The first item causes problems when long-range forces, such as electrostatic interactions or elastic defect-defect interactions are involved. Sophisticated techniques have been developed, such as Ewald-summation to simulate systems of charged or gravitating particles. For long-range electrostatic interactions it is frequently accepted that some long-range order will be imposed upon the system. Hence, the range of the interactions between particles in the system must generally be taken into account while planning a simulation. No problems will arise if the periodic box is large compared with the range over which the atomic interaction (e.g. of an EAM potential) acts. A box size greater than three times the cut-off radius of the potential is always sufficient. Specific measures have to be taken in the calculation of the forces if shorter box lengths shall be realized.

The second complication listed above is more substantial, since it is inherent to the system and cannot be solved by improved algorithms. Consequently, the basic volume size has to be chosen large enough to avoid undesired artificial correlations. Size scaling studies have to be performed to assess such effects. In favourable cases, the long range interactions of periodic arrays of defects are known and can be subtracted from the total energy of the unit cell.

Boundary conditions that mix IBC and PBC can be required in case that the simulated structure exhibits periodicity in some directions, but not in the others. Typical examples are one-dimensional structures like very long slabs or wires and systems where loads are applied in certain directions.

Mechanical loading through boundary conditions. For the study of the mechanical behaviour of solids like e.g. in fracture processes, a proper definition of boundary conditions and their adequate application are necessary. A simple approach to apply displacement boundary conditions is realized via a domain decomposition, in that a domain is added at the boundary with prescribed stress or displacements. In these domains, atoms are not subject to the dynamics of Newton's equation of motion in an MD simulation; instead, they follow prescribed displacements of the boundaries, see e.g. (Buehler, 2008). Figure 2 schematically displays this approach for the simulation of fracture. A crystalline slab is strained by means of rigid, sufficiently thick boundary layers which are subject to prescribed displacements, whereas the atoms in the interior follow Newton's equations of motion.

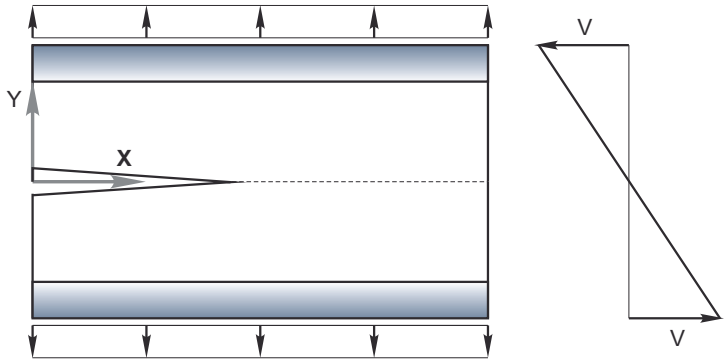


Figure 2. Application of displacement boundary conditions: Atoms in boundary domains (gray-shaded) follow prescribed boundary conditions and are not subject to the equations of motion.

The application of pressure – or generally stress – instead of displacement boundary conditions can be realized by utilizing appropriate ensemble schemes such as the Parinello-Rahman scheme (Parinello and Rahman, 1980).

Since time steps in typical MD simulations are of the order of femtoseconds to keep track of thermal vibrations, the velocity of applied displacement boundary conditions during dynamic straining is necessarily very high. To apply increased mechanical loading, a linear velocity gradient can be established prior to simulation to avoid shock wave generation from the boundaries, see Figure 2, right. However, studying processes at low loading rates is difficult. Sometimes, stress boundary conditions may allow to

simulate somewhat slower loading.

For well controlled loading situations, however, one has to resort to the application of constant displacement boundary conditions and study the system as it evolves in a given field. In the example of Figure 2, the application of a constant displacement u provides a constant (energetic) driving force for the propagation of the crack. With such boundary conditions one can reach the desired case of studying crack propagation at constant energy release rate G.

Since a propagating crack is expected to generate heat, a local temperature control (Finnis et al., 1991), resembling an electronic heat bath for the ions, is applied there. At the outer border of the model the coupling (damping) parameter is gradually increased to prevent reflections from the borders. The model is then first equilibrated at an applied strain corresponding to the Griffith load $G_0 = 2\gamma$. During the equilibration time the crack remains stationary. Thereafter the model is instantaneously strained further to a defined overload ΔG by scaling all displacements. For small overloads $\Delta G = 0.03-0.10\,G_0$ a short acceleration phase can be detected immediately after loading. This suggests that the crack has finite but very small inertia. After this acceleration phase the crack runs at constant velocity.

3.5 Stable Defects under Load

For the investigation of discrete defects like an individual crack or an individual dislocation one wants to supply the defect with the natural stress and strain fields it also would experience in a realistically large system. This is of course not possible with the rigid straight boundary conditions sketched in Figure 2. Instead a finite size model must be supplied in the border region with the forces or displacements determined from the continuum mechanical stress and strain fields of the defect. For the case of a sharp crack, the model is loaded by first applying the anisotropic linear elastic continuum solution (Sih and Liebowitz, 1968) for a fixed value of the stress intensity factor K to all atoms in the model. As a starting value for K the stress intensity factor K_G (Griffith load) is used. Then all atoms are relaxed to their equilibrium positions by a relaxation algorithm, except for the outermost atoms, which are held fixed at their initial positions. This configuration is then used for further incremental loading or unloading of the model, which is achieved by scaling all displacements and relaxing the whole model at each incremental step. Changes in the atomic configurations at the crack tip are determined by visual inspection of the relaxed configurations. The load at which one or more bonds are broken upon loading is taken as the upper critical stress intensity K_c^+. (See (Kohlhoff et al., 1991; Gumbsch,

1995) for details about the loading procedure, size scaling tests and the comparison to other methods.) Often periodic boundary conditions are applied along the crack front.

3.6 Visualisation and Analysis of Defects

A key issue in large-scale atomistic simulations is the automatic identification and visualisation of defects and microstructures.

For that purpose algorithms have been developed to identify, highlight and classify typical defect types, see e.g. (Li, 2007). Hence, pictures or movies based on these indicators realize the interface between experiments in the virtual laboratory and the scientist. They enable a transformation of vast data sets to structured information. Since visualisation criteria allow the judicious selection of regions of interest, they also serve a similar purpose as concurrent multiscale methods, namely to compress or reduce large data sets. In the following some of these visualisation criteria are put forward.

Energy method. One way to extract crystal defects from their undisturbed neighbourhood is the energy method. In this method, atoms with an energy value larger than a defined threshold – or within a specified interval – are targeted for display, which exploits the fact that defects exhibit high-energy. The energy method has been successfully applied to visualise microcracks, dislocations, nanovoids and the like. For nanoindentation into (001) fcc aluminium the energy method selects dislocation loops slipping on (111) planes as visualised in Figure 3. Moreover it can be seen that the energy method accounts for the energy difference between free surfaces and bulk material. In order to avoid free surfaces preventing analysis, a subregion of interest is chosen for visualisation. A shortcoming of the energy-method is that it cannot be applied at elevated temperatures. Furthermore, crystalline defects like stacking faults are hard to see using the energy method. In these cases it is favorable to use defect-indicators which take into account the structure of the crystal lattice and its symmetries.

Centrosymmetry parameter. The centrosymmetry parameter (CP) as introduced by (Kelchner et al., 1998) is defined for an fcc atom according to

$$P = \sum_{i=1}^{6} |\boldsymbol{r}_i + \boldsymbol{r}_{-i}|^2 \ , \tag{17}$$

where vectors $\boldsymbol{r}_i$ and $\boldsymbol{r}_{-i}$ correspond to the six pairs of next neighbours lying at opposite sites with respect to the considered atom in the lattice. By definition, the CP is zero for an atom in the bulk of a perfect material subject

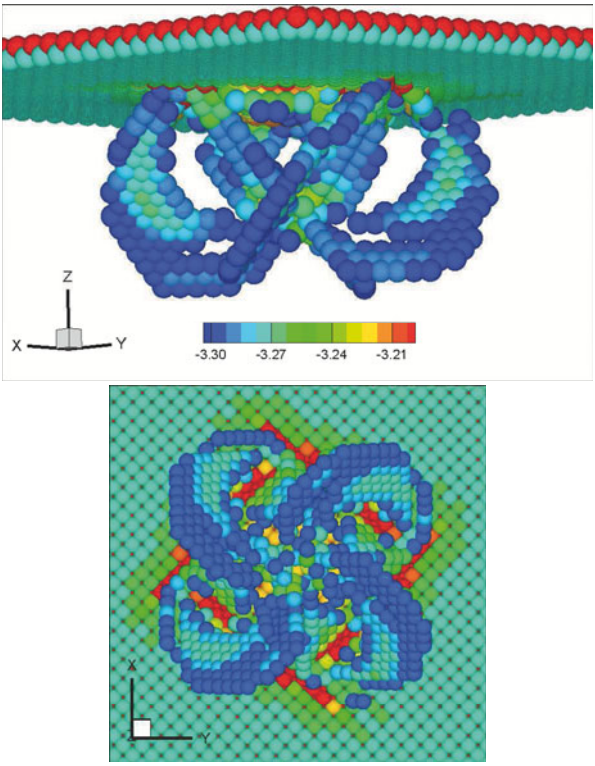


Figure 3. Simulation of nanoindentation into (001) fcc Al at zero temperature: the internal energy criterion extracts the free surface and the dislocation loops.

to purely homogeneous elastic deformations. The deviation of P from zero therefore measures the strength of disturbed centrosymmetry at a lattice site. Opposed to the Energy Method, the CP enables identification and classification of defects like free surfaces, partial dislocations and stacking faults by a certain number.

Since the CP – opposed to the Energy-Method – is related to the crystal's structure, it is invariant to thermal fluctuations and hence also applicable at finite temperatures.

Slip vector analysis. The slip vector analysis (SVA) was proposed by (Zimmerman et al., 2001) in the context of MD simulation of nanoindenta-

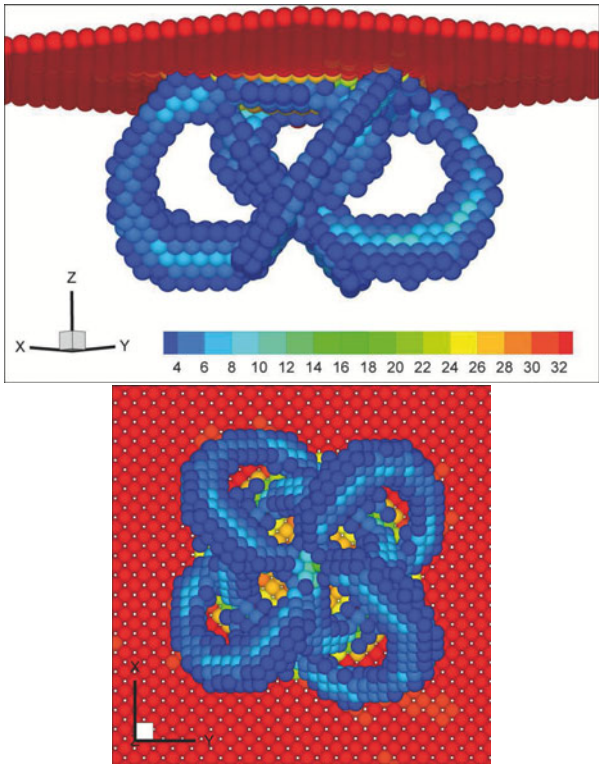


Figure 4. Simulation of nanoindentation into (001) fcc Al: the centrosymmetry parameter extracts the free surface and the dislocation loops.

tion; the slip vector of atom α is defined as

$$\boldsymbol{s}^{\alpha} = -\frac{1}{n_s} \sum_{\beta \neq \alpha}^{n_\alpha} \left\{ \boldsymbol{r}^{\alpha\beta} - \boldsymbol{R}^{\alpha\beta} \right\} \ , \tag{18}$$

where n_s is the number of slipped neighbours, n_α is the number of nearest neighbours, $\boldsymbol{r}^{\alpha\beta}$ and $\boldsymbol{R}^{\alpha\beta}$ are the vector differences between atom α and atom β in the current and the reference configuration, respectively.

The particular advantage of the SVA over CP is, that Burgers vectors are directly accessible and thus, SVA realizes the bridge to crystallography. Furthermore, the slip vector approach can be applied to very different types of defects in crystalline solids, whereas CP measures only disturbed centrosymmetry.

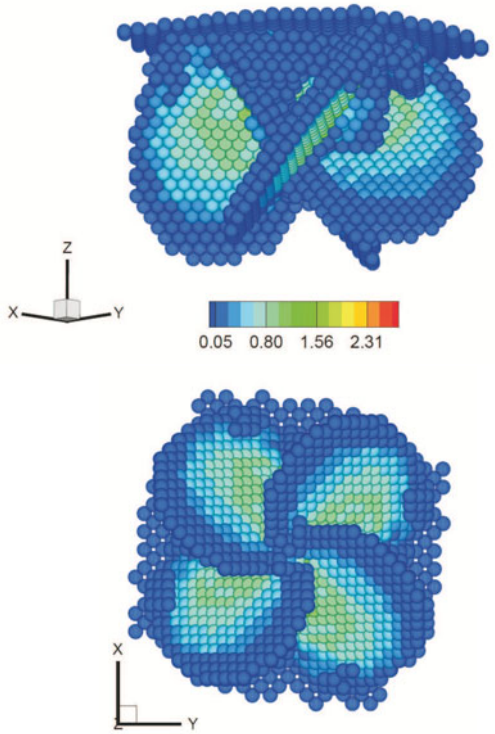


Figure 5. Simulation of nanoindentation into (001) fcc Al: The slip vector analysis, L_2 norm of $\boldsymbol{s}^\alpha$, extracts the dislocation loops.

Other criteria to identify and visualise defects in crystalline solids are the common neighbour analysis (Honeycutt and Andersen, 1987) and the Bond Angle Analysis (Ackland and Jones, 2006). The bond angle analysis is particularly useful to determine the local coordination, distinguishing fcc, hcp, bcc, and other relatively close-packed structures.

4 Concurrent Multiscale Methods

4.1 Introduction and Classification of Multiscale Methods

Despite the considerable achievements by means of molecular simulations, which have been sped up by the ever more increasing computer power, the range of applicability of atomistic models and methods is still rather lim-

ited in that they have not yet reached the typical time and length scales of engineering applications. The reason is that quite disparate time and length scales have to be considered; for MD the maximum time step is dictated by the frequency of thermal vibrations, hence in the order of femtoseconds, whereas a process like e.g. crack propagation may occur in the order of seconds. The spatial problem is not less demanding, since the length scale at the bottom is in the range of atomic spacings, hence of nanometres, whereas the world of engineering problems lives in the range of some centimetres - and beyond. For that reason many efforts have been undertaken to overcome the time scale and length scale dilemma by *coarse-graining* approaches. The accurate, at best *seamless* information passing from a bottom scale to a corresponding coarse-grained scale –and eventually backwards– is one of the key challenges in computational materials science.

In the following we will consider only the coupling of length scales and not time scales, and doing this, we will moreover restrict to the atomistic-to-continuum scale-coupling at zero temperature. Multiscale methods can be generally separated into two main categories, *hierarchical/sequential* and *concurrent*.

- *Hierarchical or sequential multiscale methods.* In this concept, material information on the atomic scale is generated and passed to a larger length scale. A simple example of this concept is the atomistic calculation of material parameters like elastic constants, thermal expansion coefficients, hardening moduli which are then used as input in continuum constitutive equations. The embedded atom method (EAM) itself can be seen as an example of a hierarchical multiscale method, since the parameters of EAM are determined in ab-initio calculations. Another instance is the identification of parameters in traction-separation laws of atomic debonding, which are employed in cohesive zone finite elements.

  This concept of information-passing from a small-scale model to a larger scale model is relatively simple and cheap. Since the material's small-scale response is parametrised for a fixed set of tests to feed the larger-scale model, the larger-scale model cannot account for the full complexity of the material's behaviour in situations far from the test bed, where the model parameters were calibrated. In order to overcome these limitations a second branch of multiscale methods has emerged.
- *Concurrent multiscale methods.* Opposed to the aforementioned class of methods, concurrent multiscale methods explicitly couple models on different length scales. Corresponding computational frameworks allow the running of simulations on disparate length scales in parallel,

i.e. *concurrently*. For that aim the simulation domain is decomposed into different regions, where fully atomistic resolution is retained in critical regions, where deformation strongly varies and where inelastic deformations occur. In regions with weakly varying, purely elastic deformations, continuum constitutive laws are typically employed along with a *coarse-graining* as realized by the finite element method. The overall goal is to achieve a result as accurate as necessary and as effective as possible.

Concurrent multiscale methods are typically used for problem sets where inelastic deformations localise in regions of confined size. These regions typically are embedded into other regions, which deform elastically and which form the largest portion of the entire simulation domain. Since crack propagation is a prominent example in this class, research in the mechanics of fracture has driven the development of concurrent multiscale models and methods, and vice versa.

A number of review articles for concurrent multiscale methods coupling different length scales is available, see e.g. (Ortiz and Phillips, 1999; Miller and Tadmor, 2002; Curtin and Miller, 2003; Miller and Tadmor, 2009), and with a special focus on fracture processes, (Abraham et al., 2000).

In the following we describe two prominent concurrent multiscale methods, the Finite Element Atomistic Method (FEAt) and the Quasicontinuum (QC) Method.

4.2 The Finite Element Atomistic (FEAt) Method

The Finite Element Atomistic (FEAt) Method as introduced in (Kohlhoff, 1990; Kohlhoff et al., 1991) is based on the decomposition of a crystal into generally three different domains as visualised in Figure 6:

A : a lattice region with fully atomistic resolution, where interatomic potentials are employed.

C : a continuum region, discretized by finite elements, where a local, but nonlinear continuum constitutive law of anisotropic elasticity is employed.

T : a transition region in between A and C, where the coupling between the local continuum and the nonlocal lattice is mediated. In this region the framework of Kröner's nonlocal elasticity theory is used. It is a continuum theory, which takes the finite range of internal forces into account and therefore can be seen as a continuation of the lattice, (Kröner, 1963).

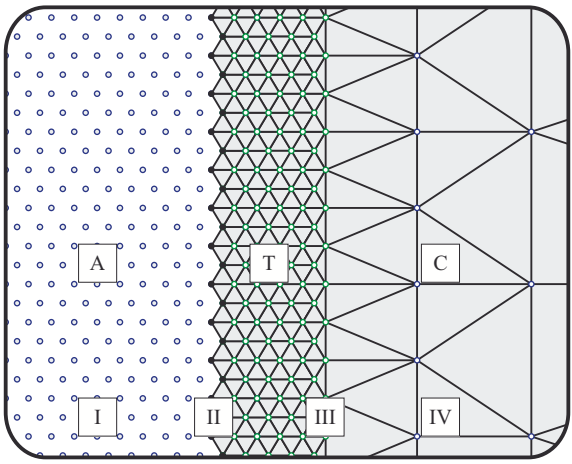


Figure 6. Scale transition scheme in FEAt.

Coupling conditions. The coupling between the atomistic and continuum regions in FEAt is in the first place based on the (i) strong kinematical compatibility and on (ii) stress/force compatibility.

Kinematical coupling. The kinematical compatibility in the transition T region is realized as follows; the transition region where the lattice and the continuum overlap is divided into two zones, II and III, Figure 6. Each zone provides the displacement boundary conditions for the other zone. Zone II, which reduces to a surface in a three-dimensional model, supplies the boundary conditions for the continuum, region IV. The FE nodes on this surface coincide with the atoms of the lattice and move with them. Conversely, in zone III the atoms, which constitute the outer shell of the lattice, are made to move in accord with the FE nodes with which they coincide. A one-to-one correspondence of nodes and atoms throughout the transition region is obtained in this way. This is why this coupling is generally termed *strong compatibility* in the classification of domain decompositions in multiscale models.

The width of zone III must be at least equal to the cut-off length of the potentials used to describe the atomistic core region. Note that the transition zone III must be $2R_{cut}$ thick when the underlying atomistic model is of the EAM type, and thus atomic forces (derivatives of energy) depend on the electron-density at an atom and at an atom's neighbour R_{cut} away, the latter of which depends on a neighbour's neighbour, up to another R_{cut} away.

Force coupling. The design of the model ensures equality of the displacement fields in the lattice and the continuum throughout the transition region, and thus equality of strains at the interface. However, since the use of forces has been explicitly avoided in this coupling scheme, any direct interaction between the stress fields of the two media is prevented. This means that force equilibrium or equality of stress between the lattice and the continuum is not established a priori. Additional conditions are necessary. In FEAt, the coupling condition is on the force level, whereas other concurrent multiscale methods perform the coupling on the energy level. To define these coupling conditions in a consistent fashion, the elastic energy $E = \hat{E}(\varepsilon)$ is expanded into a Taylor series about the state of zero strain under the assumption of zero stress,

$$E(\varepsilon) = E(0) + \left.\frac{\partial E}{\partial \varepsilon_{ij}}\right|_0 + \frac{1}{2}\left.\frac{\partial^2 E}{\partial \varepsilon_{ij}\partial \varepsilon_{kl}}\right|_0 \varepsilon_{ij}\varepsilon_{kl} + \frac{1}{6}\left.\frac{\partial^3 E}{\partial \varepsilon_{ij}\partial \varepsilon_{kl}\partial \varepsilon_{mn}}\right|_0 \varepsilon_{ij}\varepsilon_{kl}\varepsilon_{mn} + \dots . \tag{19}$$

For the equality of stresses, the strains and all coefficients in this series must be equal in the atomistic region and in the continuum. Since the strains have been made equal by means of the strong compatibility of displacements, equality of the stresses amounts to the requirement, that the elastic constants of the continuum equal those defined by the interatomic potential in the atomistic region:

$$\mathbb{C}_{ij} = \left.\frac{\partial E}{\partial \varepsilon_{ij}}\right|_0 , \quad \mathbb{C}_{ijkl} = \left.\frac{\partial^2 E}{\partial \varepsilon_{ij}\partial \varepsilon_{kl}}\right|_0 , \quad \mathbb{C}_{ijklmn} = \left.\frac{\partial^3 E}{\partial \varepsilon_{ij}\partial \varepsilon_{kl}\partial \varepsilon_{mn}}\right|_0 . \tag{20}$$

Since the reference state of the series in Equation (19) is a homogeneous deformation, $\varepsilon = 0$, it is indeed permissible to assign to the continuum the second- and higher order elastic constants as defined by Equation (20) and which are derived from the interatomic potential. The first-order elastic constants in the continuum are zero by definition, which imposes the restriction on the potential that it must provide zero stress in a perfect lattice. Within the framework of *local* and *linear* elasticity theory, equilibrium between the lattice and the FE continuum is fulfilled, if terms up to the second order are matched. FEAt accounts for elastic nonlinearity in that, additionally, elastic constants of third order, are adapted as well.

Two approximations in this coupling scheme are made. First, that the series in Equation (19) has to be cut off at some stage. Second, that there is a transition at the discrete interface from interatomic and hence finite-range forces to continuum Cauchy-type stresses. The latter approximation introduces a discontinuity in the nonlocal part of the stress tensor whenever

the strain gradient at the interface does not vanish. As a consequence, the magnitudes of the strain and of the strain gradient at the atomic-continuum influences the quality of the approximation in the model and the applicability to various situations.

Since in the transition region, II and III, finite element mesh nodes and lattice sites as well as their degrees of freedom coincide by virtue of strong compatibility, dispensable degrees of freedom are elimininated by condensation before the solution process. In (Kohlhoff, 1990) it was observed, that the global finite element stiffness matrix is not symmetric, which indicates that the forces are not conservative and hence that the governing equations are not derived from a variational principle. This is due to the fact, that the FEAt model does not start out from a well-defined total energy for the entire coupled problem but rather effects the coupling between the atomistic and continuum domains on the force/stress level as described above.

The fundamental difference between force coupling and energy coupling, its implications and consequences, are analyzed in Section 4.4 where we compare two variants of the QC method.

Application to fracture simulation. The method was used in (Kohlhoff et al., 1991) to analyse crack propagation on cleavage and non-cleavage planes in bcc crystals, using potentials for iron and tungsten as examples. The results explain why both, the {100} and {110} planes, are cleavage planes in bcc metals and why cleavage on {100} is easier than on the close-packed {110} planes.

4.3 The Quasicontinuum-Method Based on the Cauchy-Born Rule

The quasicontinuum method is a prominent example of a bottom-up, concurrent multiscale method aiming at a seamless link of atomistic with continuum length scales. This aim is achieved by three main building blocks which are common to each of the existing QC-versions, the QC based on Cauchy-Born elasticity, (Tadmor et al., 1996), (Shenoy et al., 1999), and two variants of a fully nonlocal QC method, (Knap and Ortiz, 2001), (Eidel and Stukowski, 2009):

(i) a coarse-graining of fully atomistic resolution via kinematic constraints in order to reduce the number of degrees of freedom. Full atomistic resolution is retained where necessary.
(ii) an approximation of the energy/forces in coarse-grained regions via numerical quadrature which avoids the explicit computation of the site energy of all the atoms.
(iii) adaptivity, i.e. spatially adaptive resolution, is necessary to automatically balance accuracy and efficiency. It must be directed by a suitable

refinement indicator.

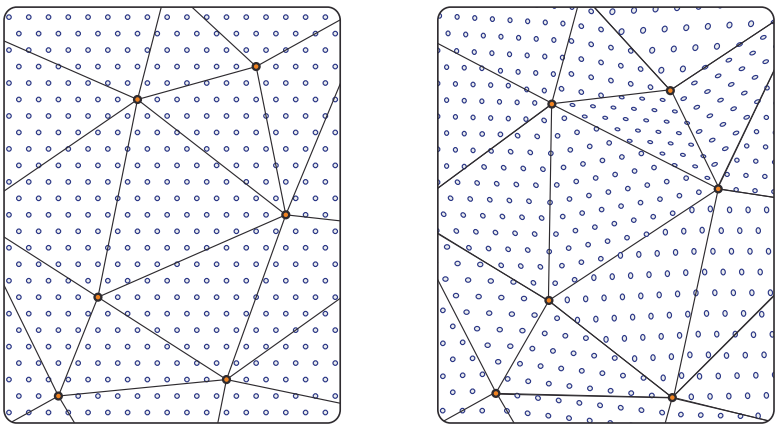


Figure 7. Finite-element discretisation of a crystal in the QC-method in the (left) undeformed configuration and (right) in the deformed configuration. Atoms within elements smoothly follow the mesh nodes by linear interpolation.

To set the stage, we consider a crystal in d-dimensional space consisting of a set $\mathscr{L} \subset \mathbb{Z}^d$ of atoms, that are initially located on a Bravais lattice spanned by lattice vectors $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_d$. Their coordinates in the initial configuration read $\boldsymbol{X}_{\boldsymbol{l}} = \sum_{i=1}^{d} l^{(i)} \boldsymbol{A}_i$, $\boldsymbol{l} \in \mathscr{L} \subset \mathbb{Z}^d$. The corresponding atomic coordinates in the current configuration are denoted by vector $\boldsymbol{x}_{\boldsymbol{l}}$.

Upscaling via coarse-graining. In regions of weakly varying elastic deformation it is sufficient to consider the movement of some judiciously selected, *representative* atoms (rep-atoms), $\mathscr{L}_h \subset \mathscr{L}$. Only these atoms keep their independent degrees of freedom, whereas all other atoms, $\overline{\mathscr{L}_h} = \mathscr{L} \backslash \mathscr{L}_h$, are forced to follow via kinematic constraints borrowed from the finite element method: $\boldsymbol{x}_{\boldsymbol{l}} = \sum_{j \in \mathscr{L}_h} \boldsymbol{x}_{\boldsymbol{j}} \varphi_{\boldsymbol{j}}(\boldsymbol{X}_{\boldsymbol{l}})$, $\boldsymbol{l} \in \overline{\mathscr{L}_h}$. FE shape functions, $\varphi_{\boldsymbol{j} \in \mathscr{L}_h}$, exhibit the properties, $\sum_{\boldsymbol{j} \in \mathscr{L}_h} \varphi_{\boldsymbol{j}}(\boldsymbol{X}_{\boldsymbol{i}}) = 1 \; \forall \; \boldsymbol{i} \in \mathscr{L}$ (partition of unity), and, $\varphi_{\boldsymbol{j}}(\boldsymbol{X}_{\boldsymbol{j}}{}') = \delta_{\boldsymbol{j}\boldsymbol{j}'} \; \forall \; \boldsymbol{j}, \boldsymbol{j}' \in \mathscr{L}_h$ (compact support). The use of (here: linear) shape functions for interpolation requires the generation of a triangulation with representative atoms as mesh nodes. Figure 7 schematically displays the discretisation of the crystal into finite elements.

The interpolation of nodal displacements implicitly introduces a continuum assumption into the QC method. Notwithstanding, this first approx-

imation is purely kinematical in nature, since no constitutive assumptions are made.

The approximation step of discretisation or coarse-graining reduces the number of arguments in the exact total potential, $E^{\text{tot}}(\{\boldsymbol{x_i}| \ \boldsymbol{i} \in \mathscr{L}\}) \rightarrow E^{\text{tot}}(\{\boldsymbol{x_i} \mid \boldsymbol{i} \in \mathscr{L}_h\}) =: E^{\text{tot},h}$, and thus reduces the number of unknowns in the computation. Both existing QC methods have this approximation step in common, but differ in the way further approximations are made.

Next, we focus on the QC version based on Cauchy-Born elasticity, in Section 4.4 the fully nonlocal QC versions are described and compared.

Efficient energy/force calculation: the local QC. After thinning-out dispensable degrees of freedom via the kinematic constraints in terms of linear finite-element shape functions, the first QC-version as proposed by (Tadmor et al., 1996) accomplishes an efficient energy/force calculation in the continuum region by recourse to the so-called *Cauchy-Born (CB) Rule* -hence QC-CBR- resulting in what is referred to as the *local* formulation of the QC.

The CBR postulates that when a monatomic crystal is subjected to a small linear displacement of its boundary, then all atoms will follow this displacement, see (Born and Huang, 1998), (Ericksen, 1983), (Zanzotto, 1996). The CBR is schematically illustrated in Figure 8 for a crystalline cantilever undergoing elastic bending deformation.

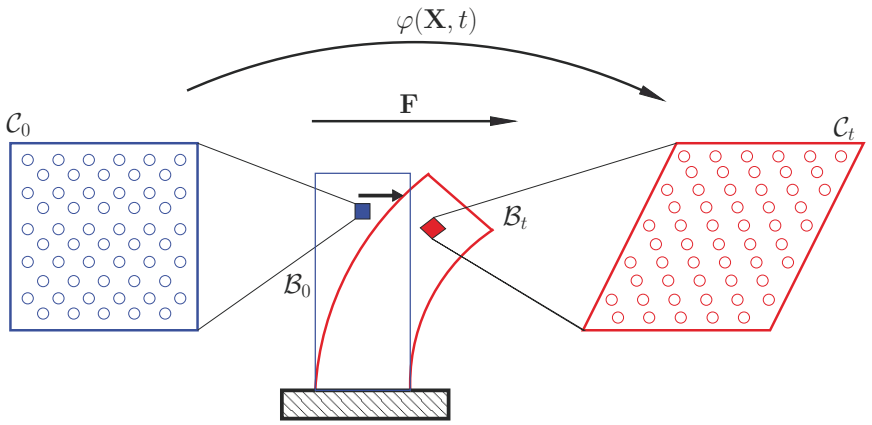


Figure 8. The Cauchy-Born-rule assuming a homogeneous deformation state in small representative volumes.

The CBR is applied in the QC method in that the continuum defor-

mation gradient as a macro-scale quantity is directly mapped to a *uniform* deformation of a small volume on the atomistic or nano-scale. For crystalline solids with a simple lattice structure the assumption of locally homogeneous deformation state implies that every atom in a region subject to a uniform deformation gradient will be energetically equivalent. As a consequence, calculating the energy within a specific finite element can be approximated by computing the energy of only one single atom in the deformed state and multiplying this figure by the number of atoms in the specific finite element. Within the QC-computational framework, the calculation of the CB energy is done separately in a subroutine; for a given deformation gradient $\boldsymbol{F}$ the lattice vectors in a unit cell with PBC is deformed according to $\boldsymbol{F}$

$$\boldsymbol{a}_i = \boldsymbol{F}\boldsymbol{A}_i\,, \tag{21}$$

where $\boldsymbol{A}_i$ and $\boldsymbol{a}_i$ are the lattice vectors in the undeformed configuration and in the deformed configuration, respectively.

The deformed lattice vectors enter the employed potential for energy calculation, such that the CBR enables the free energy of a deformed crystalline body (as a function of lattice vectors) to be expressed alternatively as a function of the deformation gradient $\boldsymbol{F}$. The corresponding strain energy density in the element is then given by

$$\mathcal{E} = \frac{E_0(\boldsymbol{F})}{\Omega_0}\,, \tag{22}$$

where Ω_0 is the unit cell volume (in the reference configuration) and E_0 is the energy of the unit cell when its lattice vectors are distorted according to $\boldsymbol{F}$. Now the total energy of a finite element is this energy density times the element volume, the total energy of the problem is simply the sum of all element energies:

$$E^{tot,h} = \sum_{i=1}^{N_{element}} \Omega_i \mathcal{E}(\boldsymbol{F}_i)\,, \tag{23}$$

where Ω_i is the volume of element i.

Linear interpolation functions in tetrahedral finite elements require only one single Gauss-point for numerical quadrature and therefore imply a constant deformation gradient per element as visualised in the right of Figure 7. Note, that the locally constant deformation gradient in the finite element matches the assumption of a locally constant deformation gradient in the CBR. As a consequence, the application of the CBR implies that in the energy calculation the summation over the number of lattice sites boils down

to the number of finite elements $N_{element}$, see Equation (23). Since the crystal is in general subject to inhomogeneous deformations, the element-wise constant deformation gradient is an approximation and so is the calculated energy via the CBR. In settings where the deformation is varying slowly and the element size is adequate with respect to the variations of the deformation, this type of energy calculation is sufficiently accurate and very effective.

For a mathematical analysis on the range of validity of the CBR we refer to (Friesecke and Theil, 2002), where it is found that the CBR fails for relatively small elastic deformations. An extension of the classical linear CBR to high order is proposed in (Sunyk and Steinmann, 2003).

Nonlocal QC. In nonlocal regions, which can be eventually refined to fully atomistic resolution, the energy E_α of an atom residing on a mesh node α is calculated by numerical quadrature. Specifically the new approximate energy takes the form

$$E^{tot,h} = \sum_{\alpha=1}^{N_{rep}} n_\alpha E_\alpha(\boldsymbol{u}_h)\,, \tag{24}$$

where $\boldsymbol{u}_h$ represents the finite element nodal displacements. The computational saving is that the summation of all the atoms is replaced by a sum over all representative atoms N_{rep}. In the line of numerical quadrature, n_α is the weight function for repatom α which requires for consistency

$$\sum_{\alpha=1}^{N_{rep}} n_\alpha = N\,. \tag{25}$$

Hence, n_α is the number of atoms represented by atom α, which implies in the limiting case of fully atomistic resolution $n_\alpha = 1$.

Mixed local-nonlocal QC. In order to combine the high accuracy of the nonlocal formulation with the efficiency of the local formulation, the former is employed in critical regions, where atomic scale accuracy is required, where the latter formulation is employed in regions where the deformation is changing relatively slowly on the atomic scale. As a result, the QC-CBR runs both formulations concurrently in a single simulation.

As in the energy-based nonlocal QC, the coupled approach is based on the ansatz that the energy can be approximated by computing only the energy of the repatoms. In the coupled approach however, each repatom is judiciously selected as being either local or nonlocal depending on its

deformation environment. Thus, the repatoms are divided into N_{loc} local repatoms and N_{nonloc} nonlocal repatoms ($N_{loc} + N_{nonloc} = N_{rep}$). Doing this, the total energy is approximated as

$$E^{tot,h} = \sum_{\alpha=1}^{N_{nonloc}} n_\alpha E_\alpha(\boldsymbol{u}_h) + \sum_{\alpha=1}^{N_{loc}} n_\alpha E_\alpha(\boldsymbol{u}_h)\,. \tag{26}$$

The weights n_α for each repatom (local or nonlocal) are determined from a tessellation that divides the body into cells around each repatom. The numerically expensive Voronoi tessellation can be replaced by an approximate Voronoi diagram. The Voronoi cell of repatom α contains a total of n_α atoms. Of these atoms, n_α^i reside in element i adjacent to repatom α. The total weighted energy contribution of repatom α is then calculated by use of the CBR within each element adjacent to α, hence

$$n_\alpha E_\alpha = \sum_{i=1}^{M} n_\alpha^i \Omega_0 \mathcal{E}(\boldsymbol{F}_i), \qquad n_\alpha = \sum_{i=1}^{M} n_\alpha^i\,, \tag{27}$$

where $\mathcal{E}$ is the energy density in element i by the CB rule, Ω_0 is the Wigner-Seitz volume of a single atom and M is the number of elements adjacent to α.

The scale transition from fully atomistic resolution to a coarse-grained description is visualised in Figure 9.

The ghost-force problem. QC-CBR inherently exhibits so-called *ghost forces*, defined as spurious forces arising at the interface between local and nonlocal regions. These forces thus follow from the fact that the motion of rep-atoms in the local region subject to the CBR will effect the energy of nonlocal rep-atoms, while the converse may not be true. Hence, this force mismatch stems from different physical assumptions on how atoms interact, which is a compatibility-problem.

There are two different concepts to reduce ghost forces in QC-CBR:

1. Correction by applying a *static correction force field*.
   This remedy against ghost forces proposed in (Shenoy et al., 1999) exhibits the drawback that static correction forces are not derivable from a 'correction potential energy', i.e., they are nonconservative. This may lead to serious problems with energy conservation during a molecular-dynamics simulation as reported in (Shimokawa et al., 2004).
2. Correction by *continuation*.
   In order to cure the ghost force problem without new shortcomings

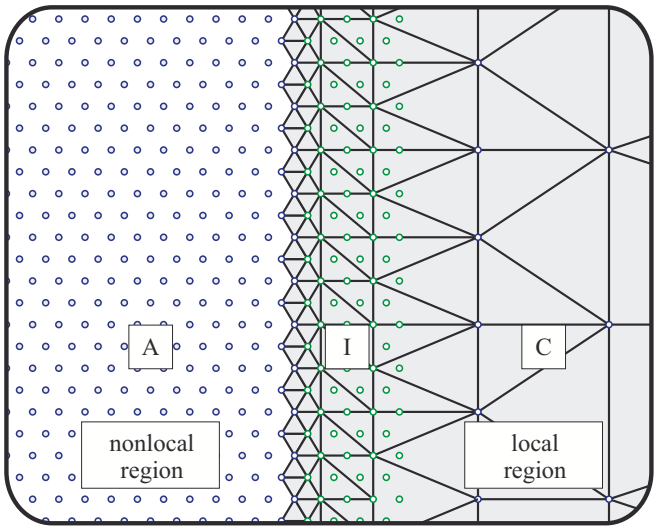


Figure 9. The scale transition scheme in the QC-CBR, where the nonlocal region (A) overlaps in the interface zone (I) with the continuum region (C), which is discretised with finite elements and subject to the CBR.

(Shimokawa et al., 2004) introduced a buffer layer between the two regions of space, where atoms are subject to specific rules concerning how they interact with their local and nonlocal neighbourhood. In a similar spirit is the contribution of (E. et al., 2006), where the approach of local reconstruction schemes is generalised.

Application to fracture. In (Miller et al., 1998a) and (Miller et al., 1998b) the QC-CBR method has been applied to crack tip deformation and is shown to account for both brittle fracture and crack tip dislocation emission. The analysis of a crack propagating into a grain boundary revealed both, migration of the boundary and that the boundary is a source for the emission of dislocations.

4.4 The Fully Nonlocal Cluster-Based Quasicontinuum-Method

A fully nonlocal QC-version (QC-FNL) as proposed by (Knap and Ortiz, 2001) aims to overcome the aforementioned force mismatch between local and nonlocal regions in QC-CBR. For that aim they replace the CBR by a unified nonlocal theory to be described in the following and thus avoid in-

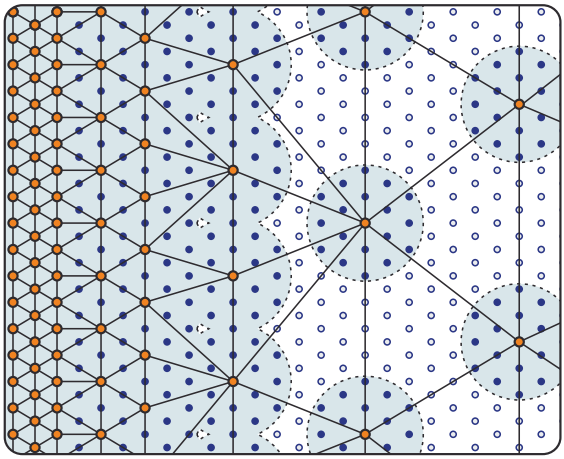


Figure 10. The scale transition scheme in fully nonlocal QC-versions based on force/energy sampling in clusters: the atomistic-continuum coupling is realized in a continuous manner by gradual coarse-graining enabling a seamless scale-transition.

compatibilities of different physical descriptions at discrete interfaces. This then enables the seamless scale transition between fully atomistic resolution and coarse-grained continuum regions. The scale transition in QC-FNL is realized in a continuous manner by gradual coarse-graining, see Figure 10, whereas in QC-CBR it is realized at the discrete interface where different physical models meet, see Figure 9. In the above structure of QC building blocks, (i)–(iii), fully nonlocal QC versions introduce for property (ii) the use of summation rules for the sampling of *forces* or *energies* in spherical clusters, which can be seen as representative crystallites.

Force versus energy sampling in clusters. Even after coarse-graining, the total energy still depends on the site energy $E_{\boldsymbol{k}}$ of each and every atom $\boldsymbol{k}$, $E^{tot,h} = \sum_{\boldsymbol{k}\in\mathscr{L}} E_{\boldsymbol{k}}$. Due to the prohibitive computational expense of this task, a second approximation becomes necessary, which is again, like discretisation, a very standard in classical finite element methods: numerical quadrature.

For that purpose Knap and Ortiz (2001) proposed to perform force evaluations no longer at each lattice site in the crystal but to restrict them to sampling clusters. These sampling clusters $\mathscr{C}_{\boldsymbol{i}}$ are spheres of radius R_c and are chosen to have a mesh node in its centre, see Figures 10 and 11. Hence,

they are defined as $\mathscr{C}_{\boldsymbol{i}} = \{\boldsymbol{k} : |\boldsymbol{X}_{\boldsymbol{k}} - \boldsymbol{X}_{\boldsymbol{i}}| \leq R_c(\boldsymbol{i})\}$. Note, that the sampling clusters may have different positions, e.g. in the interior of the finite element.

Assuming a pair potential $V = V(\boldsymbol{r}_{\boldsymbol{kl}})$ with $\boldsymbol{r}_{\boldsymbol{kl}} = \boldsymbol{x}_{\boldsymbol{k}} - \boldsymbol{x}_{\boldsymbol{l}}$ and omitting here and in the following the contribution of an external potential V^{ext}, the force acting on node $\boldsymbol{a}$ reads for **force sampling** in clusters

$$\boldsymbol{f}^h_{\boldsymbol{a}} = \sum_{\boldsymbol{i}\in\mathscr{L}_h} n_{\boldsymbol{i}} \sum_{\boldsymbol{k}\in\mathscr{C}_{\boldsymbol{i}}} \boldsymbol{f}_{\boldsymbol{k}} \varphi_{\boldsymbol{a}}(\boldsymbol{X}_{\boldsymbol{k}}) = -\sum_{\boldsymbol{i}\in\mathscr{L}_h} n_{\boldsymbol{i}} \sum_{\boldsymbol{k}\in\mathscr{C}_{\boldsymbol{i}}} \left[\sum_{\boldsymbol{l}\in\mathscr{L}} V'(|\boldsymbol{r}_{\boldsymbol{kl}}|) \frac{\boldsymbol{r}_{\boldsymbol{kl}}}{|\boldsymbol{r}_{\boldsymbol{kl}}|}\right] \varphi_{\boldsymbol{a}}(\boldsymbol{X}_{\boldsymbol{k}})\,. \tag{28}$$

The equilibrium configurations of interest are the minimisers of $E^{tot,h}$, i.e. the solutions of the variational problem:

$$\min_{\{\boldsymbol{x}_{\boldsymbol{a}}\}} E^{tot,h} \quad \rightarrow \quad \boldsymbol{f}^h_{\boldsymbol{a}} = \boldsymbol{0} \quad \forall\, \boldsymbol{a} \in \mathscr{L}_h\,. \tag{29}$$

Energy minimisation physically corresponds to solving for the configuration for which at every mesh node $\boldsymbol{a}$ the sum of forces on each degree of freedom is zero. Based on this fact, Knap and Ortiz (2001) search for the equilibrium by directly working from an approximate expression for the forces according to Equation (28) rather than working from the explicit differentiation of a total energy functional.

In Eidel and Stukowski (2009) however, the sampling is introduced at the energy level, thus

$$E^{\text{QC}} = \sum_{\boldsymbol{i}\in\mathscr{L}_h} n_{\boldsymbol{i}} \sum_{\boldsymbol{k}\in\mathcal{C}_{\boldsymbol{i}}} E_{\boldsymbol{k}} \approx E^{tot,h}\,, \tag{30}$$

which yields for **energy sampling** in clusters to the force expression

$$\boldsymbol{f}^h_{\boldsymbol{a}} = -\frac{\partial E^{\text{QC}}}{\partial \boldsymbol{x}_{\boldsymbol{a}}} = -\sum_{\boldsymbol{i}\in\mathscr{L}_h} n_{\boldsymbol{i}} \sum_{\boldsymbol{k}\in\mathscr{C}_{\boldsymbol{i}}} \frac{1}{2} \sum_{\boldsymbol{l}\in\mathscr{L}} \left[V'\left(|\boldsymbol{r}_{\boldsymbol{kl}}|\right) \frac{\boldsymbol{r}_{\boldsymbol{kl}}}{|\boldsymbol{r}_{\boldsymbol{kl}}|} \left[\varphi_{\boldsymbol{a}}\left(\boldsymbol{X}_{\boldsymbol{k}}\right) - \varphi_{\boldsymbol{a}}\left(\boldsymbol{X}_{\boldsymbol{l}}\right)\right]\right], \tag{31}$$

which is the counterpart of Equation (28).

It is worth to note that for energy sampling – as opposed to force sampling – the force expression $\boldsymbol{f}^h_{\boldsymbol{a}}$ in Equation (31) is explicitly derived from a well defined total potential E^{QC}, see Equation (30).

Calculation of the weighting factors. In both cases, i.e. for force and energy sampling, factor $n_{\boldsymbol{i}}$ is the weighting of the force /energy contribution

of cluster $\mathscr{C}_{\boldsymbol{i}}$. The cluster weights $n_{\boldsymbol{i}}$, $\boldsymbol{i} \in \mathscr{L}_h$, are calculated under the requirement that the summation over all linear interpolation functions must be exact, see Knap and Ortiz (2001), hence

$$\sum_{\boldsymbol{i}\in\mathscr{L}_h} n_{\boldsymbol{i}} \sum_{\boldsymbol{k}\in\mathcal{C}_{\boldsymbol{i}}} \varphi_{\boldsymbol{j}}(\boldsymbol{X}_{\boldsymbol{k}}) = \sum_{\boldsymbol{k}\in\mathscr{L}} \varphi_{\boldsymbol{j}}(\boldsymbol{X}_{\boldsymbol{k}}) \qquad \forall \qquad \boldsymbol{j} \in \mathscr{L}_h\,. \tag{32}$$

The calculation of the weights implies the assumption that the quantity subject to sampling can be exactly approximated if it is linear between the mesh nodes.

When the clusters shrink to the size of the rep-atoms, i.e. $\mathscr{C}_{\boldsymbol{i}} = \{\boldsymbol{i}\}\ \forall\, \boldsymbol{i} \in \mathscr{L}_h$, it holds that $\varphi_{\boldsymbol{a}}(\boldsymbol{X}_{\boldsymbol{k}}) = \delta_{\boldsymbol{a}\boldsymbol{k}}$, and the cluster summation rule boils down to a node-based summation rule $\boldsymbol{f}^h_{\boldsymbol{a}} = \sum_{\boldsymbol{k}\in\mathscr{L}_h} n_{\boldsymbol{k}} \boldsymbol{f}_{\boldsymbol{k}} \varphi_{\boldsymbol{a}}(\boldsymbol{X}_{\boldsymbol{k}}) = n_{\boldsymbol{a}} \boldsymbol{f}_{\boldsymbol{a}}$. In this case the weighting factor $n_{\boldsymbol{k}}$ is the number of atoms represented by rep-atom $\boldsymbol{k}$, thus $n_{\boldsymbol{k}} = \sum_{\boldsymbol{l}\in\mathscr{L}} \varphi_{\boldsymbol{k}}(\boldsymbol{X}_{\boldsymbol{l}})\ \forall\, \boldsymbol{k} \in \mathscr{L}_h$, which implies $n_{\boldsymbol{k}} = 1$ for fully atomistic resolution, i.e. $\boldsymbol{l} \in \mathscr{L}$, and which ensures that $\sum_{\boldsymbol{k}\in\mathscr{L}_h} n_{\boldsymbol{k}} = |\mathscr{L}|$ is fulfilled.

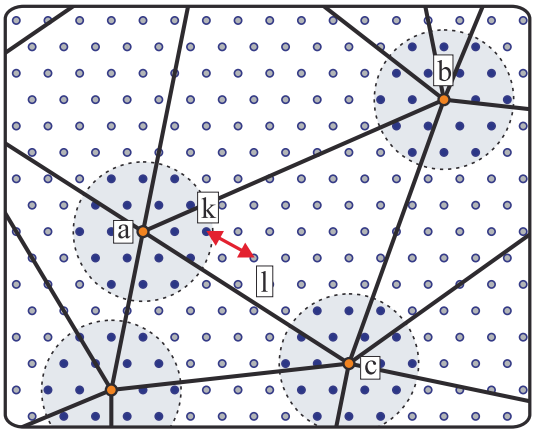


Figure 11. Spherical clusters around mesh nodes for the explicit sampling of forces or energies. The interaction of sampling atom $\boldsymbol{k}$ inside the cluster with non-sampling atom $\boldsymbol{l}$ outside the cluster must be symmetric to fulfill Newton's third law.

Is Newton's third law of motion preserved? A critical issue in concurrent multiscale modelling is the coupling of atomistics with a continuum, which is largely due to the general *locality* of continuum constitutive laws and the *nonlocality* of interatomic forces.

In the following we check both sampling schemes to ensure the symmetry of atomic interactions. For that aim we consider the atomic interaction of a sampling atom $\boldsymbol{k}$ (within the spherical cluster adjacent to mesh node $\boldsymbol{a}$) with a non-sampling atom $\boldsymbol{l}$ as illustrated in Figure 11. It is instrumental to explicitly separate all four force terms (I)-(IV) (see below) for this type of interaction and adhering therein to the lattice statics terms for pair potentials. First, we consider force terms captured by energy sampling. The nonlocal action of energy $E_{\boldsymbol{k}\leftarrow\boldsymbol{l}}$ induces a force on atom $\boldsymbol{k}$ within the cluster, expression (I), and on atom $\boldsymbol{l}$ outside the cluster, expression (II). These two forces are equal up to the opposite sign, Newton's third law holds, *actio=reactio*. The interaction of atom $\boldsymbol{k}$ with atom $\boldsymbol{l}$ is schematically illustrated in Figure 11. Both forces are distributed according to their barycentric coordinates from $\boldsymbol{k}$ and $\boldsymbol{l}$ to adjacent nodes $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$; for atom $\boldsymbol{a}$ the distribution is mediated by factor $[\varphi_{\boldsymbol{a}}(\boldsymbol{X}_{\boldsymbol{k}}) - \varphi_{\boldsymbol{a}}(\boldsymbol{X}_{\boldsymbol{l}})]$ in Equation (31). Moreover, since energy sampling is conceptually restricted to clusters, forces at site $\boldsymbol{k}$ and site $\boldsymbol{l}$ due to the energy contribution $E_{\boldsymbol{l}\leftarrow\boldsymbol{k}}$, expressions (III) and (IV), are missing in Equation (31); this explains the factor 1/2 therein. It is the function of properly defined weighting factors $n_{\boldsymbol{i}}$ to account for the energy contribution of non-sampling atoms.

(I) Force on atom $\boldsymbol{k}$ due to energy $E_{\boldsymbol{k}\leftarrow\boldsymbol{l}}$:

$$\tilde{\boldsymbol{f}}_{\boldsymbol{k}} = -\frac{\partial E_{\boldsymbol{k}\leftarrow\boldsymbol{l}}}{\partial \boldsymbol{x}_{\boldsymbol{k}}} = -\frac{1}{2}V'(|\boldsymbol{r}_{\boldsymbol{kl}}|)\frac{\boldsymbol{r}_{\boldsymbol{kl}}}{|\boldsymbol{r}_{\boldsymbol{kl}}|}$$

.

(II) Force on atom $\boldsymbol{l}$ due to energy $E_{\boldsymbol{k}\leftarrow\boldsymbol{l}}$:

$$\tilde{\boldsymbol{f}}_{\boldsymbol{l}} = -\frac{\partial E_{\boldsymbol{k}\leftarrow\boldsymbol{l}}}{\partial \boldsymbol{x}_{\boldsymbol{l}}} = +\frac{1}{2}V'(|\boldsymbol{r}_{\boldsymbol{kl}}|)\frac{\boldsymbol{r}_{\boldsymbol{kl}}}{|\boldsymbol{r}_{\boldsymbol{kl}}|}$$

.

(III) Force on atom $\boldsymbol{k}$ due to energy $E_{\boldsymbol{l}\leftarrow\boldsymbol{k}}$:

$$\tilde{\tilde{\boldsymbol{f}}}_{\boldsymbol{k}} = -\frac{\partial E_{\boldsymbol{l}\leftarrow\boldsymbol{k}}}{\partial \boldsymbol{x}_{\boldsymbol{k}}} = -\frac{1}{2}V'(|\boldsymbol{r}_{\boldsymbol{kl}}|)\frac{\boldsymbol{r}_{\boldsymbol{kl}}}{|\boldsymbol{r}_{\boldsymbol{kl}}|}$$

.

(IV) Force on atom $\boldsymbol{l}$ due to energy $E_{\boldsymbol{l}\leftarrow\boldsymbol{k}}$:

$$\tilde{\tilde{\boldsymbol{f}}}_{\boldsymbol{l}} = -\frac{\partial E_{\boldsymbol{l}\leftarrow\boldsymbol{k}}}{\partial \boldsymbol{x}_{\boldsymbol{l}}} = +\frac{1}{2}V'(|\boldsymbol{r}_{\boldsymbol{kl}}|)\frac{\boldsymbol{r}_{\boldsymbol{kl}}}{|\boldsymbol{r}_{\boldsymbol{kl}}|}$$

.

The application of the force-based ansatz, Equation (28), gives full account of forces acting on cluster atom $\boldsymbol{k}$, summing up expressions (I) and

(III). Contrary to the present ansatz however, Equation (28) does not consider the opposite force acting on atom $\boldsymbol{l}$, since $\boldsymbol{l}$ is not a sampling atom. Briefly, energy-sampling preserves symmetry in atomic interactions whereas force-sampling does not.
It can be equally shown, (Eidel and Stukowski, 2009), that this lack of symmetry in the interaction of sampling atoms with non-sampling atoms implies an asymmetry of the corresponding stiffness matrix, thus indicating nonconservative forces. By contrast, energy sampling leads to strictly symmetric stiffness matrices indicating conservative forces.
The stiffness matrix for energy sampling reads

$$\begin{aligned}\boldsymbol{k}^h_{\boldsymbol{ab}} &= \frac{\partial^2 E^{\mathrm{QC}}}{\partial \boldsymbol{x}_{\boldsymbol{a}} \partial \boldsymbol{x}_{\boldsymbol{b}}} \\ &= \sum_{\boldsymbol{i}\in\mathscr{L}_h} n_{\boldsymbol{i}} \sum_{\boldsymbol{k}\in\mathscr{C}_{\boldsymbol{i}}} \frac{1}{2} \sum_{\boldsymbol{l}\in\mathscr{L}} \Big[\varphi_{\boldsymbol{a}}(\boldsymbol{X}_{\boldsymbol{k}}) - \varphi_{\boldsymbol{a}}(\boldsymbol{X}_{\boldsymbol{l}})\Big]\Big[\varphi_{\boldsymbol{b}}(\boldsymbol{X}_{\boldsymbol{k}}) - \varphi_{\boldsymbol{b}}(\boldsymbol{X}_{\boldsymbol{l}})\Big] \cdot \\ &\qquad \left[\frac{V'(|\boldsymbol{r}_{\boldsymbol{kl}}|)}{|\boldsymbol{r}_{\boldsymbol{kl}}|}\mathbf{1} + \left(\frac{V''(|\boldsymbol{r}_{\boldsymbol{kl}}|)}{|\boldsymbol{r}_{\boldsymbol{kl}}|^2} - \frac{V'(|\boldsymbol{r}_{\boldsymbol{kl}}|)}{|\boldsymbol{r}_{\boldsymbol{kl}}|^3}\right) \boldsymbol{r}_{\boldsymbol{kl}} \otimes \boldsymbol{r}_{\boldsymbol{kl}}\right]. \end{aligned} \tag{33}$$

and for force sampling:

$$\begin{aligned}\boldsymbol{k}^h_{\boldsymbol{ab}} = -\frac{\partial \boldsymbol{f}^h_{\boldsymbol{a}}}{\partial \boldsymbol{x}_{\boldsymbol{b}}} &= \sum_{\boldsymbol{i}\in\mathscr{L}_h} n_{\boldsymbol{i}} \sum_{\boldsymbol{k}\in\mathscr{C}_{\boldsymbol{i}}} -\frac{\partial \boldsymbol{f}_{\boldsymbol{k}}}{\partial \boldsymbol{x}_{\boldsymbol{b}}} \\ &= \sum_{\boldsymbol{i}\in\mathscr{L}_h} n_{\boldsymbol{i}} \sum_{\boldsymbol{k}\in\mathscr{C}_{\boldsymbol{i}}} \varphi_{\boldsymbol{a}}(\boldsymbol{X}_{\boldsymbol{k}}) \sum_{\boldsymbol{l}\in\mathscr{L}} [\varphi_{\boldsymbol{b}}(\boldsymbol{X}_{\boldsymbol{k}}) - \varphi_{\boldsymbol{b}}(\boldsymbol{X}_{\boldsymbol{l}})] \cdot \\ &\qquad \left[\frac{V'(|\boldsymbol{r}_{\boldsymbol{kl}}|)}{|\boldsymbol{r}_{\boldsymbol{kl}}|}\mathbf{1} + \left(\frac{V''(|\boldsymbol{r}_{\boldsymbol{kl}}|)}{|\boldsymbol{r}_{\boldsymbol{kl}}|^2} - \frac{V'(|\boldsymbol{r}_{\boldsymbol{kl}}|)}{|\boldsymbol{r}_{\boldsymbol{kl}}|^3}\right) \boldsymbol{r}_{\boldsymbol{kl}} \otimes \boldsymbol{r}_{\boldsymbol{kl}}\right]. \end{aligned} \tag{34}$$

Summarising, the fully nonlocal QC formulation based on energy sampling, QC-eFNL, exhibits advantages compared to force sampling: sampling at the energy level instead of the force level preserves the variational structure of lattice statics leading to conservative forces, as indicated by symmetric stiffness matrices. More specifically, energy sampling implies the strict symmetry of atomic interactions in all regions, even across the boundary of clusters, whereas force sampling does not in general. Energy sampling also exhibits some numerical advantages. Standard algorithms for the numerical minimisation of functionals like CG methods can directly be applied, since they generally require gradients as well as evaluations of the functional (the energy) itself. Moreover, a minimiser can be found, if the energy exhibits a minimum.

For residual forces observed in QC-eFNL simulations, the following properties have been shown in (Eidel and Stukowski, 2009). Residual forces are conservative in nature; they do not follow from an asymmetry in atomic interactions as a consequence of inconsistent a priori assumptions on how atoms interact; they stem from the error in numerical quadrature and therefore can be reduced (to identically zero) by a sufficiently large cluster size. As such, the present residual forces differ from ghost forces in QC-CBR by source and property, which is the reason why we distinguish by name.

In (Eidel and Stukowski, 2009) QC-eFNL is employed to simulate nanoindentation into (001) aluminium. The simulations have shown the promising capacity of the method to reduce the prohibitive computational expense of fully atomistic resolution (lattice statics) while faithfully simulating the material's response in significant details like the force-depth curve and the load level and locus of dislocation nucleation.

4.5 Other Concurrent Multiscale Methods

There are a variety of concurrent multiscale methods based on a domain decomposition method like FEAt and QC. Here we restrict to name some of them along with key references for the interested reader.

- the Coupling of Length Scale (CLS) Method, (Rudd and Broughton, 1998), (Rudd and Broughton, 2000), (Rudd and Broughton, 2005).
- the Atomistic-to-Continuum Coupling (AtC) method, (Fish et al., 2007).
- the Bridging Scale (BSM) method, (Wagner and Liu, 2003).
- the Coupled Atomistic and Discrete Dislocation (CADD) method, (Shilkrot et al., 2002), (Shilkrot et al., 2004).

4.6 'Learn-On-The-Fly' - LOTF

FEAt and the variants of the QC method deal with the transition from continuum (or coarsed grained) regions of the solid to its fully atomistic resolution, where e.g. EAM potentials are used to describe interatomic forces. As mentioned previously, the parameters in EAM potentials are generally fitted to the results of ab-initio calculations in different settings and are kept fixed for a 'lifetime'. In this sense, EAM potentials themselves can be seen as a hierarchical multiscale method. The method of 'learning on the fly' (LOTF), (Csanyi et al., 2004), is a generalisation of this kind of hierarchical multiscale method in that the parameters of the potential are recalculated and hereby adjusted 'on the fly' for judiciously selected atoms. Doing this, the approach of LOTF ensures a maximum of *accuracy* and *transferablity* to the broad variety of simulation settings which cannot be considered in

total in a priori calculations that are performed for a conventional fitting of empirical potential.

According to the general classification, LOTF can also be seen as a concurrent multiscale method. However, since LOTF is adhering to a unified classical force model for the entire system it avoids the 'inherent boundary problem' between different physical models along with inconsistencies which plague many of the concurrent multiscale methods. It is worth mentioning that LOTF has turned out to be quite insensitive to the chosen potentials. It selects simple parametrised potentials and 'augments' it at run time with the necessary extra information, which is computed on the fly by means of quantum calculations.

According to this concept, the flow chart structure of LOTF can be described as follows:

1. *Initialisation*: start out for the physical system in its initial condition with a reasonably parametrised classical potential.
2. *MD predictor, extrapolation*: as in standard molecular dynamics (MD), the chosen potential is used with fixed parameters to predict a system trajectory for a small number of time steps.
3. *Testing*: in the latest configuration, the local validity of the classical potential is assessed on a site by site basis, and a selected subset of atoms is flagged for quantum treatment.
4. *Quantum Mechanics*: use any quantum method (DFT or TB) which provides the desired accuracy to compute the forces on only the selected subset of atoms.
5. *Force Fitting*: the parameters of the classical potential are tuned locally around the selected atoms until it reproduces the accurate force.
6. *MD corrector: interpolation*: return the state of the system to that before the extrapolation and rerun the dynamics, interpolating the potential parameters between the old and the new values.
7. *Return to 2.*

The successful application of LOTF critically depends on the proper choice of criteria that will correctly identify regions for quantum treatment.

Recent application examples of the LOTF method include studies of dynamic crack propagation (Kermode et al., 2008) which will be discussed in some more details later. A review of recent progress in the methodology of hybrid quantum/classical (QM/MM) atomistic simulations for solid-state systems can be found in (Bernstein et al., 2009). In this reference a unified terminology is defined into which the various and disparate schemes fit, based on whether the information from the QM and MM calculations is combined at the level of energies or forces. Moreover, the pertinent issues for achieving 'seamless' coupling and the advantages and disadvantages of

the proposed schemes are discussed. Finally, the applications and scientific results obtained to date are summarised.

5 Atomistic Aspects of Fracture

Fracture is a phenomenon which spans over many length scales. The macroscopic dimensions of the crack and the specimen determine the intensity of the stress concentration at the crack tip and are equally important as the microstructure of the material, which provides preferred fracture paths. Ultimately, fracture reduces to the breaking of atomic bonds, which in the case of brittle fracture occurs at an atomically sharp crack tip (Lawn, 1993; Clarke, 1992). In a perfectly brittle material, the crack moves by no other process than the breaking of individual bonds between atoms. Nevertheless, traditional theory of brittle fracture processes does not focus on individual atomic bonds but resorts to the treatment of Griffith (Griffith, 1921), which is based on continuum thermodynamics. Following Griffith, one may regard the static crack as a reversible thermodynamic system for which one seeks equilibrium. The equilibrium condition leads to the so-called Griffith criterion, which balances the crack driving force and the material resistance against fracture. The crack driving force can be expressed as the stress intensity factor K while the material resistance against fracture for the perfectly brittle case must be at least the surface energy of the two fracture surfaces. With the implication of thermodynamic equilibrium, the Griffith picture provides a reference value K_G for the analysis of the crack driving forces. It however cannot explain why and how fracture proceeds.

From an atomistic point of view, one immediately identifies the material's resistance against fracture with the forces needed to break the crack tip bonds successively. The first atomistic studies of fracture (Thomson et al., 1971) showed that the discrete bond breaking event manifests itself in a finite stability range, which was attributed to the discreteness of the lattice and called the "lattice-trapping" effect. Lattice trapping causes the crack to remain stable and not to advance/heal until loads K^+ or K^-, somewhat larger/smaller than the Griffith load are reached. Other influences of the atomic nature of a crack have recently been summarised in a series of articles in the MRS Bulletin (Selinger and Farkas, 2000). Consequences of the lattice trapping have been reviewed (Gumbsch and Cannon, 2000; Gumbsch, 2001) and only two particularly important and enlightening aspects of the lattice trapping, crack propagation anisotropy and the production of metastable fracture surfaces, are discussed below.

5.1 Lattice Trapping and the Directional Cleavage Anisotropy

Silicon is reported to have two principal cleavage planes: {111} planes, usually the easy cleavage planes, and {110} planes (Michot, 1988; George and Michot, 1993), the planes of easy cleavage in polar III-V semiconductors. The most accurate constant-K experiments (Michot, 1988) seem to show that {110} planes have a slightly lower fracture toughness than {111} planes. For both cleavage planes, the measured fracture toughness gives surface energies ($\gamma_{110} = 2.3\mathrm{J/m^2}$, $\gamma_{111} = 2.7\mathrm{J/m^2}$) which are significantly larger than the values calculated atomistically using density functional theory (DFT) based quantum mechanical methods ($\gamma_{110} = 1.7\mathrm{J/m^2}$, $\gamma_{111} = 1.4\mathrm{J/m^2}$) (Pérez and Gumbsch, 2000a).

Propagation direction anisotropy has been observed for both cleavage planes. The preferred propagation direction is along <110> on both cleavage planes (Michot, 1988; George and Michot, 1993). On the {111} fracture surface, the anisotropy with respect to propagation direction is minimal. In contrast, cleavage on the {110} plane is extremely anisotropic. Propagation along the <110> direction results in nearly perfectly flat fracture surfaces (Michot, 1988; George and Michot, 1993). Attempts to achieve propagation in the <001> direction, perpendicular to the preferred direction, have not been successful because the crack deflects onto {111} planes (George and Michot, 1993; Cramer et al., 2000). The relation of the calculated surface energies and elastic anisotropy cannot account for this deviation of the crack (Pérez and Gumbsch, 2000b) and an atomistic investigation therefore is attractive.

DFT calculations of the crack tip stability, the anisotropy in fracture behaviour with respect to the propagation direction on the {110} plane was explained as a consequence of a difference in lattice trapping for the different propagation directions (Pérez and Gumbsch, 2000b). A {110} crack propagating in the "easy" <110> direction (see Figure 12 (left)) continuously opens successive bonds at the tip of the crack. This continuous process leads to a relatively small trapping, and it can be argued that the trapping may further decrease as the size of the model (specimen) is increased. In contrast, a crack driven in the "difficult" <001> direction on the {110} plane, displayed in Figure 12 (right), shows a clearly discontinuous bond breaking. Figure 13 shows the bond distances of the crack tip bonds (labeled "B" in Figure 12) for both systems. Further analysis of this discontinuous bond breaking process shows that it is mainly a result of the relaxation of the six or eight atoms around the crack tip and connected with a significant load sharing between the crack tip bond and the one above (Pérez and Gumbsch, 2000a). Comparing both the loading and unloading processes, it is seen that the discontinuous bond breaking is also connected with a

larger lattice trapping range (Pérez and Gumbsch, 2000b). This difference in the trapping effectively destabilises the {110} crack propagation in the <001> direction against deflection onto an inclined {111} cleavage plane. Thereby lattice trapping appears to provide the only reasonable explanation for the experimentally observed cleavage anisotropy with respect to the propagation direction for the {110} cracks in silicon.

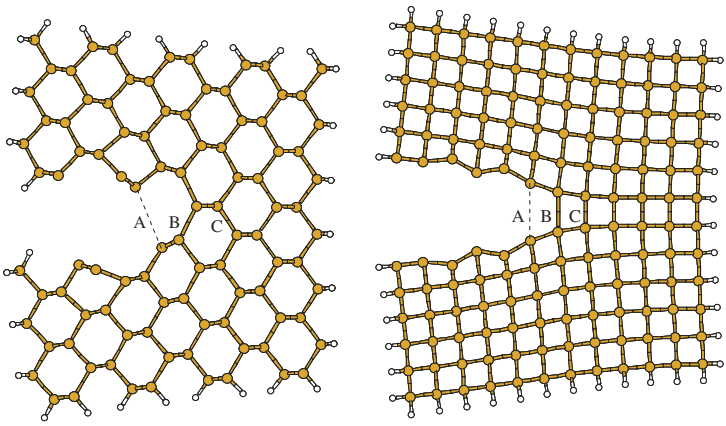


Figure 12. Sequence of bond breaking events during fracture in silicon for different orientations as given in the text.

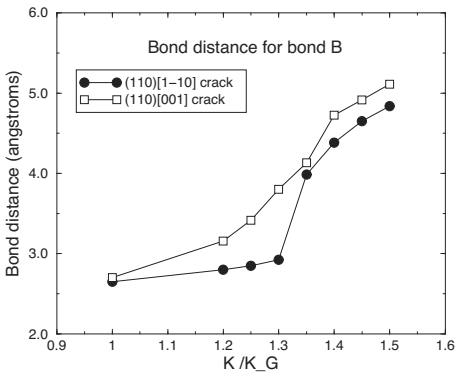


Figure 13. Bond distances of crack tip bonds "B" in Figure 12 during loading.

The same type of propagation anisotropy has recently been found in DFT

calculations of the cleavage of diamond (Pastewka et al., 2008). Similarly an anisotropy with respect to the propagation direction had also been predicted for the cleavage of tungsten single crystals (Kohlhoff et al., 1991; Riedle et al., 1996; Gumbsch et al., 1998)

5.2 Metastable Fracture Surfaces

A crack which experiences significant trapping can be loaded above the Griffith load and still does not propagate. Upon increasing the load, the crack is expected to eventually take the path that is associated with the lowest energy barrier to propagation. With trapping this path is not necessarily the one which leads to the surface of lowest energy. A very prominent example of such behaviour is the surface reconstruction after cleavage of silicon {111} planes. Low temperature cleavage ($\leq$ 600K) produces the 2x1 Pandey π-bonded chain reconstruction (Pandey, 1981), while high temperature annealing gives the so called 7x7 reconstruction. This latter structure is energetically more favourable but not directly accessible through the fracture process.

Atomistically simulating the fracture of the intermetallic alloy B2-NiAl (Ludwig and Gumbsch, 1998), using an embedded atom potential specifically developed for B2-NiAl (Ludwig and Gumbsch, 1995), it was observed that a {100} crack driven in a <001> direction did not cleanly separate the adjoining (200) Ni and Al layers. It instead took a zig-zag path which split the Al (200) plane directly in front of the crack tip and deposited every second row of atoms on the upper and the lower fracture surfaces, respectively (see Figure 14). This result would not be so remarkable if the surface energy of the half-occupied Al (100) surface (1.86 J/m^2) were not higher than the average for the Ni and Al terminated surfaces (1.76 J/m^2) (Ludwig and Gumbsch, 1998). The crack had obviously been overloaded to the extent that it could release enough elastic energy so as to create the high-energy half-occupied surfaces. It did so because the energy barrier for this process (0.18 J/m^2) is lower than the trapping barrier for the ideal surfaces.

In this example of NiAl, the crack is creating two identical half-occupied fracture surfaces instead of the asymmetric arrangement of the two surfaces being fully occupied by Ni and Al, respectively. Similar half-occupied surfaces would also be expected in equilibrium for an ionic material where the asymmetric fully occupied surfaces would be oppositely charged (Tasker, 1986). With this analogy in mind, one may search for even more pronounced trapping effects in ionic materials where the crack faces the problem of sorting out the charges.

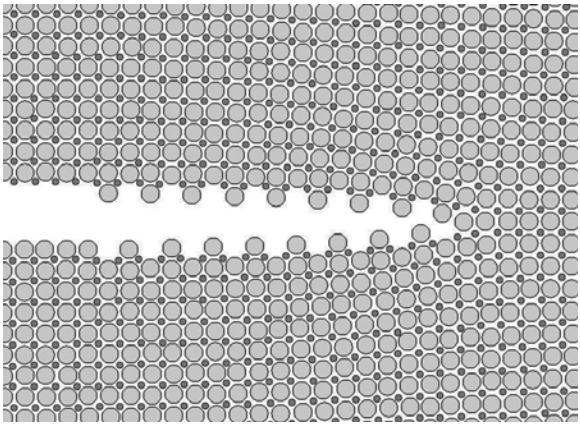


Figure 14. Rough fracture surface of B2-NiAl created during dynamic fracture.

6 Dynamics of Brittle Crack Propagation

The fracture of materials can be a dynamic process, particularly in the final stage of supercritical propagation. Although this final stage of fracture might at first seem almost irrelevant, closer consideration shows that it is precisely the dynamics of the brittle crack which competes with the rate-dependent plasticity in the near tip region to determine whether a propagating crack can ever be stopped. The dynamic crack propagation has therefore recently attracted significant attention.

The first set of atomistic investigations of dynamically moving cracks were directed towards understanding the steady state propagation, crack speed and the onset of dynamic instabilities. Analytical atomistic studies (Marder and Gross, 1995) on simplified one- and two-dimensional structures show that the dynamically propagating crack can only access a limited velocity regime. After initiation, crack tip speed immediately reaches about 20% of the Rayleigh wave velocity and approaches a branching instability at about half the Rayleigh wave velocity (Marder and Gross, 1995). The instability manifests itself in the breaking of bonds at the flanks of the crack before the breaking of the next bond in the propagation direction and is interpreted as a branching instability.

MD simulations (Gumbsch et al., 1997) of the propagation of a mode I crack with a straight crack front and a short periodic length along the crack front (quasi-two-dimensional geometry) essentially confirm the analytical results. They confirm a lower band of forbidden velocities for the straight

crack and also reveal an upper critical velocity. The upper critical velocity for the mode I crack is shown to strongly depend on the non-linearity of the atomic interaction. For harmonic snapping spring force laws (Gumbsch et al., 1997) and for open crystal structures with strong directional bonds (Hauch et al., 1999; Swadener et al., 2002) the velocities can be almost as high as the Rayleigh wave velocity, the relativistic upper limit. Only 40% of the Rayleigh wave velocity is reached for closed packed crystals and more realistic non-linear atomic interactions (Gumbsch et al., 1997). Up to 50% of the shear wave speed is reached in the more complex quasicrystalline structures (Mikulla et al., 1998).

Increasing temperature reduces this band of forbidden velocities and successively allows cracks to also propagate at lower speeds (Holland and Marder, 1999; Rudhart et al., 2003). In amorphous or quasicrystalline structures, increasing the temperature may also lead to a change in crack propagation mechanism from the propagation of a distinct crack tip to crack propagation by successive opening of pores or daughter cracks in front of the main crack and their backward propagation (Falk and Langer, 1998, 2000; Rudhart et al., 2003). This of course drastically reduces the crack propagation speed.

Above the critical velocity, the MD simulations reveal a rich set of different types of instabilities depending on the crystallographic orientation of the crack and on the crystal structure (Gumbsch et al., 1997; Hauch et al., 1999; Mikulla et al., 1998). The generation of cleavage steps and dislocation emission are usually observed at lower overloads, while crack bifurcation was only observed at the highest overloads. Dislocation emission usually leads to a pronounced change in crack propagation direction.

Surprisingly, a different dynamic instability was found in silicon on the (111) cleavage plane not at high speeds but rather at low speeds. In (Kermode et al., 2008) LOTF was applied to investigate this low-speed crack propagation instability using quantum-mechanical hybrid, multi-scale modelling and single-crystal fracture experiments. The simulations predict a crack-tip reconstruction that makes low-speed crack propagation unstable to deflect towards just one of the two crack faces on the (111) cleavage plane, which is conventionally thought of as the most stable cleavage plane. An asymmetrical crack tip reconstruction was found to be responsible for this reconstruction. A small energy barrier needs to be overcome to assess this crack tip reconstruction. Corresponding experiments confirm this instability prediction at a range of low speeds, using an experimental technique designed for the investigation of fracture under low tensile loads.

Further simulations (Kermode et al., 2008) also explain why, at moderately high speeds crack propagation on the (110) cleavage plane becomes

unstable and deflects onto (111) planes, as previously observed experimentally.

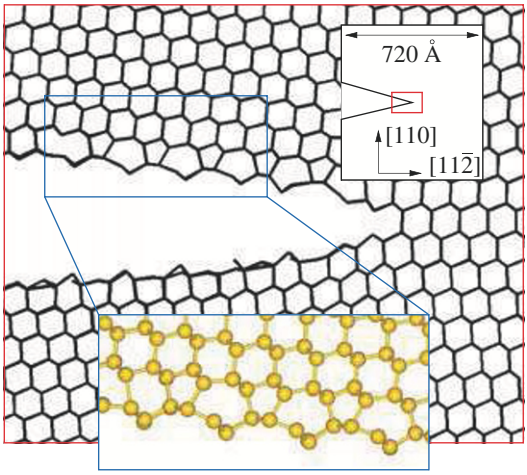


Figure 15. Propagation of the (111)[110] crack in silicon, using the LOTF scheme and the SIESTA code as quantum engine. Brittle fracture propagation on a (111) cleavage plane is correctly predicted.

Large scale MD simulations have recently confirmed continuum mechanical analysis (Gao et al., 1999) on the fact that mode II cracks just like edge dislocations (Gumbsch and Gao, 1999) are not bound by the shear wave speed as an upper limit (Gao et al., 2001). They show a transition to intersonic propagation via the nucleation of a daughter crack out of a subsonic mother crack. Large-scale MD simulations on dynamic fracture under mode I conditions have revealed that the properties of atomic bonds in the vicinity of the crack tip determine the maximum crack propagation speed (Buehler et al., 2003). By taking into account that atomic bonds under high strains, as they occur around crack tips, show non-linear elastic behaviour, it was found that the the maximum velocity may deviate significantly from that calculated by the global linear elastic properties of the material. At large strains, metallic bonds typically weaken and thus show a reduced stiffness before they break. In polymers, in contrast, the material typically becomes stiffer at large tensile strain, because the long-chained molecules are stretched. In (Buehler et al., 2003) it could be shown that the local elastic properties in a certain critical volume around the crack tip governs the dynamic fracture, rather than the global material properties.

Hence, if the material stiffens under tensile load, the crack may become supersonical compared to the global speed of sound in the material, while it is locally still subsonical.

7 Summary

The flexibility of MD methods makes them a versatile tool for studying a wide range of materials phenomena, and in particular fracture in dynamic situations. Because the size of the samples under consideration is considerably larger than in *ab initio* simulations, MD simulations can be used to compare the instability of a crack tip against propagation with its instability against shear. Since the latter process leads to the generation of dislocations, the fracture behaviour is no longer pure cleavage, but is accompanied by some plasticity. In this article the fundamentals of classical MD methods have been introduced, to demonstrate the possibilities and also the limitations of atomistic method. Furthermore, scale bridging methods have also been included which will in the future allow the community to study fracture in larger volumes and possibly on larger time scales than classical MD allows for.

Some applications of MD methods to studying and understanding fracture processes have been briefly mentioned to demonstrate the power of this method. In particular, MD simulations contributed to the understanding of brittle crack advance as a process of continuous bond breaking, which is necessarily described incompletely by any continuum method. Due to the limitations of MD methods to short time scales, dynamic fracture is a fertile field for investigations, but also the conditions for dislocation nucleation at crack tips and the interaction of propagating cracks with dislocations or pores can be studied at the atomic scale, (Zhou et al., 1996; Bitzek, 2006; Bitzek and Gumbsch, 2007).

This article thus provides an introduction to atomistic simulation methods and their application to fracture processes. Literature for in depth reading is also provided.

Bibliography

F. F. Abraham, N. Bernstein, J. Q. Broughton, and D. Hess. Dynamic fracture of silicon: Concurrent simulation of quantum electrons, classical atoms, and the continuum solid. *MRS Bull.*, 25/5:27, 2000.

G. J. Ackland and A. P. Jones. Applications of local crystal structure measures in experiment and simulation. *Phys. Rev. B*, 73(5), 2006. ISSN 1098-0121.

N. Bernstein, J. R. Kermode, and G. Csanyi. Hybrid atomistic simulation methods for materials systems. *Rep. Prog. Phys.*, 72(2), 2009. ISSN 0034-4885.

E. Bitzek. *Atomistic simulation of dislocation motion and their interaction with crack tips and voids*. PhD thesis, University of Stuttgart, Germany, 2006.

E. Bitzek and P. Gumbsch. Atomistic simulations of dislocation - crack interaction. In W. E. Nagel, W. Jäger, and M. Resch, editors, *High Performance Computing in Science an Engineering 06*, pages 127–135, Berlin, 2007. Springer Verlag.

E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, and P.Gumbsch. Structural relaxation made simple. *Phys. Rev. Lett.*, 97(17), 2006. ISSN 0031-9007.

M. Born and K. Huang. *Dynamical Theory of Crystal Lattices*. Oxford University Press, 1998.

M. J. Buehler. *Atomistic Modeling of Materials Failure*. Springer, 2008.

M. J. Buehler, F. F. Abraham, and H. Gao. Hyperelasticity governs dynamic fracture at a critical length scale. *Nature*, 426:141–146, 2003.

D. R. Clarke. *Semiconductors and Semimetals, edited by K.T. Faber and K. Malloy, Academic Press,*, page 79., 1992.

T. Cramer, A. Wanner, and P. Gumbsch. Energy dissipation and path instabilities in dynamic fracture of silicon single crystals. *Phys. Rev. Lett.*, 85:788, 2000.

G. Csanyi, T. Albaret, MC. Payne, and A. De Vita. "Learn on the fly": A hybrid classical and quantum-mechanical molecular dynamics simulation. *Phys. Rev. Lett.*, 93(17), OCT 22 2004. ISSN 0031-9007. doi: 10.1103/PhysRevLett.93.175503.

W. A. Curtin and R. E. Miller. Atomistic/continuum coupling in computational materials science. *Model. Simul. Mater. Sci. Eng.*, 11(3):R33–R68, 2003. ISSN 0965-0393.

V. S. Deshpande, A. Needleman, and E. Van der Giessen. Discrete dislocation plasticity modeling of short cracks in single crystals. *Acta Mater.*, 51:1–15, 2003.

B. Devincre and S. G. Roberts. Three-dimensional simulation of dislocation-crack interactions in b.c.c. metals at the mesoscopic scale. *Acta Mater.*, 44:2891–2900, 1996.

W. E., J. Lu, and J. Yang. Uniform accuracy of the quasicontinuum method. *Phys. Rev. B*, 74(21), 2006. ISSN 1098-0121.

B. Eidel and A. Stukowski. A variational formulation of the quasicontinuum method based on energy sampling in clusters. *J. Mech. Phys. Solids*, 57 (1):87–108, 2009.

F. Ercolessi and J. B. Adams. Interatomic potentials from first-principles calculations: the force-matching method. *Europhys. Lett.*, 26:583–588, 1994.

J. Ericksen. *The Cauchy and Born Hypotheses for Crystals*. Defense Technical Information Center, 1983.

M. L. Falk and J. S. Langer. Dynamics of viscoplastic deformation in amorphous solids. *Phys. Rev. E*, 57:7192, 1998.

M. L. Falk and J. S. Langer. From simulation to theory in the physics of deformation and fracture. *MRS Bull.*, 25/5:40, 2000.

M. W. Finnis, P. Agnew, and A. J. E. Foreman. Thermal excitations of electrons in energetic displacement cascades. *Phys. Rev. B*, 44:567, 1991.

J. Fish, M. A. Nuggehally, M. S. Shephard, C. R. Picu, S. Badia, M. L. Parks, and M. Gunzburger. Concurrent AtC coupling based on a blend of the continuum stress and the atomistic force. *Comp. Meth. Appl. Mech. Eng.*, 196(45-48):4548–4560, 2007. ISSN 0045-7825.

D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, 2002.

G. Friesecke and F. Theil. Validity and failure of the Cauchy-Born hypothesis in a two-dimensional mass-spring lattice. *J. Nonlinear Sci.*, 12(5): 445–478, 2002. ISSN 0938-8974.

H. Gao, Y. Huang, P. Gumbsch, and A. J. Rosakis. On radiation-free transonic motion of cracks and dislocations. *J. Mech. Phys. Solids*, 47: 1941, 1999.

H. Gao, Y. Huang, and F. F. Abraham. Continuum and atomistic studies of intersonic crack propagation. *J. Mech. Phys. Solids*, 49:2113, 2001.

A. George and G. Michot. Dislocation loops at crack tips - nucleation and growth - an experimental-study in silicon. *Mater. Sci. Eng. Ag*, 164:118, 1993.

A. A. Griffith. Phenomena of rupture and flow in solids. *Philos. Trans. R. Soc. London, Ser. A*, 221:163., 1921.

P. Gumbsch. An atomistic study of brittle fracture: Towards explicit failure criteria from atomistic modelling. *J. Mater. Res.*, 10:2897–2907, 1995.

P. Gumbsch. *Materials Science for the 21st Century, vol. A*, The Society of Materials Science, JSMS, Japan, 2001.

P. Gumbsch and R. M. Cannon. Atomistic aspects at brittle fracture. *MRS Bull.*, 25/5:15., 2000.

P. Gumbsch and H. Gao. Dislocations faster than the speed of sound. *Science*, 283:965, 1999.

P. Gumbsch, S.-J. Zhou, and B. L. Holian. Molecular dynamics investigation of dynamic crack stability. *Phys. Rev. B*, 55:3445, 1997.

P. Gumbsch, J. Riedle, A. Hartmaier, and H. F. Fischmeister. Controlling factors for the brittle-to-ductile transition in tungsten single crystals. *Science*, 282:1293, 1998.

W.A. Harrison. *Solid state theory.* Courier Dover Publications, 1980.

J. A. Hauch, D. Holland, M. P. Marder, and H. L. Swinney. Dynamic fracture in single crystal silicon. *Phys. Rev. Lett.*, 82:3823, 1999.

P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136 (3B):B864–B871, 1964.

D. Holland and M. Marder. Cracks and atoms. *Adv. Mater.*, 11:793, 1999.

J. D. Honeycutt and H. C. Andersen. Molecular dynamics study of melting and freezing of small lennard- jones clusters. *J. Phys. Chem.*, 91:4950–4963, 1987.

M. E. Kassner, S. Nemat-Nasser, Z. G. Suo, G. Bao J. C. Barbour, L. C. Brinson, H. Espinosa, H. J. Gao, S. Granick, P. Gumbsch, K. S. Kim, W. Knauss, L. Kubin, J. Langer B. C. Larson, L. Mahadevan, A. Majumdar, S. Torquato, and F. van Swol. New directions in mechanics. *Mech. Mater.*, 37:231–259, 2005.

C. L. Kelchner, S. J. Plimpton, and J. C. Hamilton. Dislocation nucleation and defect structure during surface indentation. *Phys. Rev. B*, 58:11085–11088, 1998.

J. R. Kermode, T. Albaret, D. Sherman, N. Bernstein., P. Gumbsch, M.C. Payne, G. Csanyi, and A. De Vita. Low-speed fracture instabilities in a brittle crystal. *Nature*, 455(7217):1224–U41, 2008. ISSN 0028-0836.

J. Knap and M. Ortiz. An analysis of the quasicontinuum method. *J. Mech. Phys. Solids*, 49(9):1899–1923, 2001. ISSN 0022-5096.

S. Kohlhoff. *Ein Verfahren für die Rechnermodellierung eines atomistischen Gitterbereiches, umgeben von einem Finite-Elemente-Kontinuum, und seine Anwendung auf Rissfortschrittsprozesse in kubisch-raumzentrierten Metallen.* PhD thesis, MPI for Metal Research, Stuttgart, 1990.

S. Kohlhoff, P. Gumbsch, and H. F. Fischmeister. Crack propagation in b.c.c. crystals studied with a combined finite-element and atomistic model. *Phil. Mag. A*, 64:851–878, 1991.

W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140(4A):A1133–A1138, 1965.

E. Kröner. On the physical reality of torque stresses in continuum mechanics. *Int. J. Engng. Sci.*, 1:261–278, 1963.

L. P. Kubin, G. Canova, M. Condat, B. Devincre, V. Pontikis, and Y. Bréchet. Dislocation microstructures and plastic flow: A 3D simulation. *Solid State Phen.*, 23&24:455–472, 1992.

B. R. Lawn. *Fracture of Brittle Solids, Cambridge University Press, Cambridge, UK,*, 1993.

A. Leach. *Molecular modelling: principles and applications*. Pearson Education, 2001.

J. Li. *Handbook of Materials Modeling*, chapter Atomistic Visualization, pages 1051–1068. Springer Netherlands, 2007.

J. Li, A. H. W. Ngan, and P. Gumbsch. Atomistic modeling of mechanical behavior. *Acta Mater.*, 51:5711–5742, 2003.

M. Ludwig and P. Gumbsch. An empirical interatomic potential for b2 nial. *Modelling Simul. Mater. Sci. Eng.*, 3:533–542, 1995.

M. Ludwig and P. Gumbsch. Cleavage fracture and crack tip dislocation emission in B2 NiAl: an atomistic study. *Acta Mater.*, 46:3135–3143, 1998.

M. Marder and S. Gross. Origin of crack-tip instabilities. *J. Mech. Phys. Solids*, 1:43, 1995.

J. Marian, W. Cai, and V. V. Bulatov. Dynamic transitions from smooth to rough to twinning in dislocation motion. *Nat. Mater.*, 3:158–163, 2004.

G. Michot. Fundamentals of silicon fracture. *Crys. Prop. Prep.*, 55:17–18, 1988.

R. Mikulla, J. Stadler, F. Krul, H.-R. Trebin, and P. Gumbsch. Crack propagation in quasicrystals. *Phys. Rev. Lett.*, 81:3163, 1998.

R. Miller, M. Ortiz, R. Phillips, V. Shenoy, and E. B. Tadmor. Quasicontinuum models of fracture and plasticity. *Eng. Fract. Mech.*, 61(3-4): 427–444, 1998a. ISSN 0013-7944.

R. Miller, E. B. Tadmor, R. Phillips, and M. Ortiz. Quasicontinuum simulation of fracture at the atomic scale. *Model. Simul. Mater. Sci. Eng.*, 6 (5):607–638, 1998b. ISSN 0965-0393.

R. E. Miller and E. B. Tadmor. The quasicontinuum method: Overview, applications and current directions. *J. Comput-Aided Mater. Des.*, 9(3): 203–239, 2002. ISSN 0928-1045.

R. E. Miller and E. B. Tadmor. A unified framework and performance benchmark of fourteen multiscale atomistic/continuum coupling methods. *Model. Simul. Mater. Sci. Eng.*, 17(5):1–51, 2009.

Y. Mishin, M. J. Mehl, D. A. Papaconstantopoulos, A. F. Voter, and J. D. Kress. Structural stability and lattice defects in copper: ab initio, tight-binding and embedded-atom calculations. *Phys. Rev. B*, 63:224106–1 to 224106–16, 2001.

M. Mrovec, D. Nguyen-Manh, D. G. Pettifor, and V. Vitek. Bond-order potential for molybdenum: Application to dislocation behavior. *Phys. Rev. B*, 69:094115–1–16, 2004.

M. Mrovec, R. Gröger, A. G. Bailey, D. Nguyen-Manh, C. Elsässer, and V. Vitek. Bond-order potential for simulations of extended defects in tungsten. *Phys. Rev. B*, 75:104119–1–16, 2007a.

M. Mrovec, M. Moseler, C. Elsässer, and P. Gumbsch. Atomistic modeling of hydrocarbon systems using analytic bond-order potentials. *Prog. Mater. Sci.*, 52:230–254, 2007b.

J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, New York, 2006.

M. Ortiz and R. Phillips. Nanomechanics of defects in solids. In *Adv. Appl. Mech.*, volume 36, pages 1–79. Academic Press, 1999.

K. C. Pandey. New π-bonded chain model for si (111)-(2×1) surface. *Phys. Rev. Lett.*, 47:1913, 1981.

M. Parinello and A. Rahman. Crystal structure and pair potentials: a molecular dynamics study. *Phys. Rev. Lett.*, 45:1196–1199, 1980.

L. Pastewka, P. Pou, R. Pérez, P. Gumbsch, and M. Moseler. Describing bond-breaking processes by reactive potentials: Importance of an environment-dependent interaction range. *Phys. Rev. B*, 78:161402, 2008.

R. Pérez and P. Gumbsch. An ab initio study of the cleavage anisotropy in silicon. *Acta Mater.*, 48:4517, 2000a.

R. Pérez and P. Gumbsch. Directional anisotropy in the cleavage fracture of silicon. *Phys. Rev. Lett.*, 84:5347, 2000b.

D. G. Pettifor. *Bonding and Structure of Molecules and Solides*. Oxford University Press, 1995.

D. G. Pettifor and I. I. Oleinik. Analytic bond-order potentials beyond Tersoff-Brenner. I. Theory. *Phys. Rev. B*, 59:8487–8499, 1999.

T. Pöschel and T. Schwager. *Computational Granular Dynamics*. Springer, 2005.

D. C. Rapaport. *The Art of Molecular Dynamics Simulation*. Cambridge University Press, 2004.

J. Riedle, P. Gumbsch, and H. F. Fischmeister. Cleavage anisotropy in tungsten single crystals. *Phys. Rev. Lett.*, 76:3594–3597, 1996.

R. E. Rudd and J. Q. Broughton. Coarse-grained molecular dynamics and the atomic limit of finite elements. *Phys. Rev. B*, 58(10):R5893–R5896, 1998. ISSN 0163-1829.

R. E. Rudd and J. Q. Broughton. Concurrent coupling of length scales in solid state systems. *Phys. Status Solidi B-Basic Solid State Phys.*, 217 (1):251–291, 2000. ISSN 0370-1972.

R. E. Rudd and J. Q. Broughton. Coarse-grained molecular dynamics: Nonlinear finite elements and finite temperature. *Phys. Rev. B*, 72(14), 2005. ISSN 1098-0121.

C. Rudhart, P. Gumbsch, and H.-R. Trebin. *Quasicrystals, Structure and Physical Properties, H.-R. Trebin (ed.), Wiley-VCH, Weinheim*, page 484, 2003.

T. Schlick. *Molecular Modeling and Simulation: An Interdisciplinary Guide*. Springer, 2002.

R. L. B. Selinger and D. Farkas. Atomistic theory and simulation of fracture. *MRS Bull.*, 25/5:11, 2000.

V. B. Shenoy, R. Miller, E. B. Tadmor, D. Rodney, R. Phillips, and M. Ortiz. An adaptive finite element approach to atomic-scale mechanics - the quasicontinuum method. *J. Mech. Phys. Solids*, 47(3):611–642, 1999. ISSN 0022-5096.

J. R. Shewchuk. An introduction to the conjugate gradient method without the agonizing pain, 1994. URL `http://www.cs.cmu.edu/ jrs/jrspapers.html#cg`.

L. E. Shilkrot, W. A. Curtin, and R. E. Miller. A coupled atomistic/continuum model of defects in solids. *J. Mech. Phys. Solids*, 50 (10):2085–2106, 2002. ISSN 0022-5096.

L. E. Shilkrot, R. E. Miller, and W. A. Curtin. Multiscale plasticity modeling: coupled atomistics and discrete dislocation mechanics. *J. Mech. Phys. Solids*, 52(4):755–787, 2004. ISSN 0022-5096.

T. Shimokawa, J. J. Mortensen, J. Schiotz, and K. W. Jacobsen. Matching conditions in the quasicontinuum method: Removal of the error introduced at the interface between the coarse-grained and fully atomistic region. *Phys. Rev. B*, 69(21):214104, 2004. ISSN 1098-0121.

G. C. Sih and H. Liebowitz. Mathematical theories of brittle fracture. In H. Liebowitz, editor, *Fracture*, volume 2, pages 67–190, New York, 1968. Academic Press.

R. Sunyk and P. Steinmann. On higher gradients in continuum-atomistic modelling. *Int. J. Solids Struct.*, 40(24):6877–6896, 2003. ISSN 0020-7683.

J. G. Swadener, M. I. Baskes, and M. Nastasi. Molecular dynamics simulation of brittle fracture in silicon. *Phys. Rev. Lett.*, 89:085503, 2002.

E. B. Tadmor, M. Ortiz, and R. Phillips. Quasicontinuum analysis of defects in solids. *Phil. Mag. A*, 73(6):1529–1563, 1996. ISSN 0141-8610.

P. W. Tasker. Surfaces and interfaces in ionic materials. In G. Jacucci, editor, *Computer Simulation in Physical Metallurgy*, pages 21–40. ECSC, EEC, EAEC, Brussels and Luxembourg, 1986.

R. Thomson, C. Hsieh, and J. Rana, V. Lattice trapping of fracture cracks. *Appl. Phys.*, 42:3154., 1971.

G. J. Wagner and W. K. Liu. Coupling of atomistic and continuum simulations using a bridging scale decomposition. *J. Comput. Phys.*, 190(1): 249–274, 2003. ISSN 0021-9991.

D. Weygand and P. Gumbsch. Study of dislocation reactions and rearrangements under different loading conditions. *Mat. Sci. Eng. A*, 400:158–161, 2005.

D. Weygand, L. H. Freidman, E. Van der Giessen, and A Needleman. Aspects of boundary-value problem solutions with three-dimensional dislocation dynamics. *Model. Simul. Mater. Sci. Eng.*, 10:437–468, 2002.

G. Zanzotto. The Cauchy-Born hypothesis, nonlinear elasticity and mechanical twinning in crystals. *Acta Crystallogr. Sect. A*, 52(Part 6):839–849, 1996. ISSN 0108-7673.

S. J. Zhou, D. M. Beazley, P. S. Lomdahl, and B. L. Holian. Dynamic crack processes via molecular dynamics. *Phys. Rev. Lett.*, 76:2318–2321, 1996.

J. A. Zimmerman, C. L. Kelchner, P. A. Klein, J. C. Hamilton, and S. M. Foiles. Surface step effects on nanoindentation. *Phys. Rev. Lett.*, 86: 165507–1 to 165507–4, 2001.

Fundamental dislocation theory and 3D dislocation mechanics

Volker Mohles *

* Institute of Physical Metallurgy and Metals Physics, RWTH Aachen, Aachen, Germany

Abstract In chapter 1 an introduction to basic dislocation properties in an elastic continuum is given. Displacements, strains, stresses and energies of straight edge and screw dislocations are compiled as well as forces on dislocations, implications of dislocation motion and aspects of dislocations in real crystals. Chapter 2 details the models of dislocation self interaction for curved dislocations including the line tension model and linear elastic self interaction. The former is essential for basic understanding, whereas the latter is the basis of accurate dislocation dynamics simulations of plasticity. In chapter 3 these models are applied for 2-dimensional dislocation glide which allow to calculate the strengthening effect of second phase particles and solute atoms in a material. Finally, aspects of 3-dimensional dislocation motion are outlined in chapter 4.

1 Basic Dislocation Theory

1.1 Heuristic Dislocation Creation

In the literature, a number of excellent introductions to the basics of dislocation theory have been given (e.g. (Hirth and Lothe, 1992), (Hull and Bacon, 1992), (Weertman, 1992)). They cover the historical development of the dislocation concept (Volterra, 1907), (Peierls, 1940), experimental discovery of dislocations in crystals (Orowan, 1934), (Polanyi, 1934), (Taylor, 1934) and overviews of physical phenomena which can only show up in the presence of dislocations. Moreover, detailed introductions to linear elasticity theory have been given (Hirth and Lothe, 1992) because elastic distortions and stresses determine most of the dislocation properties: about 80-90% of a dislocation's energy is stored in the elastic strains. The present compilation of dislocation theory addresses students and researchers who already have a certain knowledge in dislocation theory, but who want to refresh and extend their understanding, in particular with respect to dislocation dynamics simulations. From the start, it is emphasized that, in

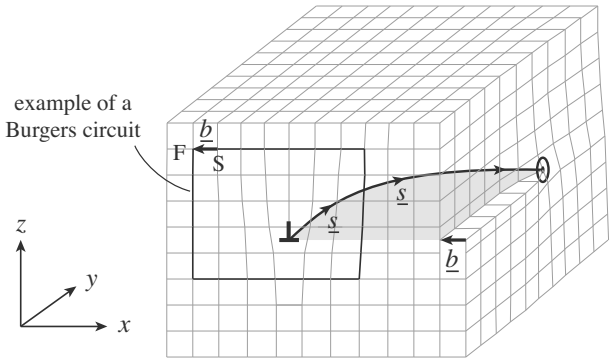


Figure 1. Dislocation in a linear elastic medium. The dislocation (black line) demarcates the area (light grey) in which the material has slipped by the Burgers vector $\underline{b}$, which is a lattice vector. The sign of $\underline{b}$ and of the line vector $\underline{s}$ can only be defined in combination.

general, dislocations are curved and flexible.

The generation of a dislocation in a linear elastic continuum can be envisaged by the following hypothetical steps: First, a cut is performed through part of a specimen along a certain plane. Then one of the surfaces which have been newly created by the cut is shifted by a vector $\underline{b}$, called the Burgers vector. Finally the surfaces are rejoined and the specimen is allowed to relax. An example for the outcome of this procedure is plotted in figure 1. The light grey area is the plane in which the cut had been performed. This area is framed by the specimen's dimensions and by the line which the tip of the hypothetical knife had moved along. This latter line (thick and black in figure 1) is the new dislocation. All material that does not include a piece of dislocation is distorted and elastically stressed, but it is still continuous. This includes the area of the hypothetical cut.

The dislocaton itself is a displacement singularity, as can be seen from the Burgers circuit. The construction of this circuit requires a lattice in the material, which is naturally given by the atomic arrangement of any real crystal. In figure 1 a simple cubic lattice has been used for simplification. The Burgers circuit is started at any lattice point S. From S, a number of i, j, and k steps along the lattice vectors in x-, y-, and z-direction, respectively, are drawn subsequently. From there, the same numbers of steps are drawn in the same order along the lattice vectors in the opposite direction. This defines the final lattice point F. In a perfect lattice or even in a distorted but continuous one, F equals S so that a closed circuit has been

described by these six lattice vectors. But if the described sequence of six vectors encloses a dislocation, an additional lattice vector from F to S is required to close the circuit. This defines the Burgers vector $\underline{b}$. Disregarding continuous lattice distortions, its magnitude and direction are independent of the choice of the numbers i, j, and k and of the starting point S as long the dislocation is enclosed by the circuit. However, a convention about the sign of $\underline{b}$ is required because an inversion of the circuit's sense would yield $-\underline{b}$ instead of $\underline{b}$, while the real geometry is not changed at all. The convention must link the sign of $\underline{b}$ to the actual shift that had been used in the above hypothetical dislocation generation. A convenient and always applicable way is to assign a local tangent vector $\underline{s}$ to the dislocation. This defines the line direction or 'line vector'. Then the Burgers circuit is performed following the right-hand-rule, where the thumb points in the line direction. $\underline{b}$ is defined as the vector from start (S) to finish (F). This is the most frequently used rule (see (Hirth and Lothe, 1992)), and it is adopted subsequently. However, there are prominent exceptions (Weertman, 1992); therefore one must be careful about which convention is taken. In any case, the sign of the Burgers vector $\underline{b}$ and the line vector $\underline{s}$ can only be defined in combination. It is important for consistency that the right-hand Burgers circuit yields the vector $\underline{b}$ indicated in the figure. This can easily be verified on the front plane (y=const., with j=0) or on the right side (x=const., with i=0) of the specimen. When $\underline{s}$ is inverted, $\underline{b}$ is reversed as well so that the same geometry is described.

From the hypothetical way in which the dislocation in figure 1 had been produced, several fundamental dislocation properties can be deduced. One is that the Burgers vector $\underline{b}$, being identical to the shift vector, is a lattice vector, which is constant by definition with respect to magnitude and direction. Obviously three cases are to be distinguished: Firstly, $\underline{b}$ and the line vector $\underline{s}$ are parallel and lie in the 'cutting plane'. This is the case for the dislocation part that leaves the crystal on the right side of figure 1. Owing to the geometry of the displacements visible on this free surface, this dislocation part is said to have screw character. Secondly, $\underline{s}$ is normal to $\underline{b}$ but both lie in the cutting plane. This case is found on the dislocation portion that leaves the crystal on the front plane. This part has, by definition, edge character. And thirdly, $\underline{b}$ is normal to the cutting plane (and to $\underline{s}$); this case is called a prismatic dislocation (see below).

Another feature becoming apparent from figure 1 is that a dislocation cannot terminate inside a crystal because it represents the fringe of a plane, the slipped area. Still it can be useful to refer to dislocation segments or parts (short straight pieces) in order to describe local properties, like the dislocation character. The possibility of a blurred fringe and hence a

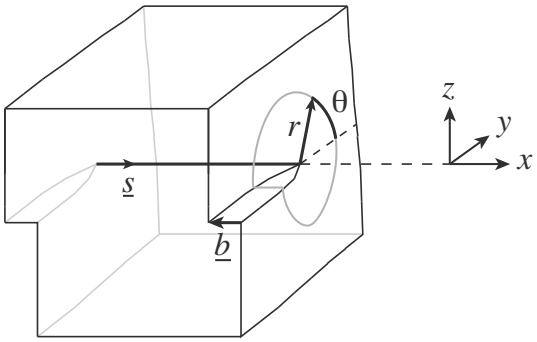


Figure 2. Straight screw dislocation. With $\underline{s}$ parallel to $\underline{b}$, the glide plane is undefined.

distributed dislocation is discussed later. When a dislocation moves, the slipped region increases or decreases with the motion. This means that the crystal is deformed, which enables crystal plasticity. In the following, dislocation features are described in more detail.

1.2 Basic Dislocation Types

According to the hypothetical generation of a dislocation described in the introduction, several basic types of dislocation can be distinguished, depending on the angular relations between Burgers vector $\underline{b}$, line vector $\underline{s}$ and the 'cutting plane'. In the following, the coordinate system is equivalently given either by $\underline{x}_1 = (x_1, 0, 0)$, $\underline{x}_2 = (0, x_2, 0)$, $\underline{x}_3 = (0, 0, x_3)$, or by $\underline{x} = (x, 0, 0)$, $\underline{y} = (0, y, 0)$, $\underline{z} = (0, 0, z)$, whichever is more convenient.

Straight screw dislocation The displacement u_x in the direction x caused by a straight screw dislocation along the x-axis can be read directly from figure 2: When the angle θ around the dislocation line is increased by 2π, a displacement of one lattice vector along the x-axis occurs. This is, by the definition of the Burgers circuit, one Burgers vector $\underline{b}$. When elastic isotropy is assumed for the material and the block is extended infinitely in all directions, the incremental displacement $\mathrm{d}u_x/\mathrm{d}\theta$ will be equally distributed (i.e., constant) over the whole range $0 \leq \theta < 2\pi$. This yields

$$u_x(r, \theta) = b\frac{\theta}{2\pi} = \frac{b}{2\pi}\tan^{-1}\left(\frac{z}{y}\right) \tag{1}$$

Note that u_x does not depend on the radius r of the circuit. In figure

2 this has been relaxed: there the displacement has been assumed to be distributed approximately over the range $0 < r < b = |\underline{b}|$; this is equivalent to a blurred fringe of the cutting plane mentioned earlier. With the general definition of the strain tensor $\boldsymbol{\varepsilon}$ or $\{\varepsilon_{ij}\}$,

$$\varepsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) \tag{2}$$

and with Hook's law

$$\sigma_{ij} = C_{ijkl}\ \varepsilon_{kl} \tag{3}$$

which connects the stess tensor $\{\sigma_{ij}\}$ with the strains $\{\varepsilon_{ij}\}$ and the stiffness tensor $\{C_{ijkl}\}$, the stresses of the straight screw dislocation can be derived from (1) by elementary mathematics. Note that in equation (3) and subsequently, Einstein's summation convention is applied: summation is to be performed over equal indices (k and l in (3)). In case of an elastically isotropic material the stiffness tensor $\{C_{ijkl}\}$ contains only two independent elastic constants (see e.g. (Hirth and Lothe, 1992)), for instance, the shear modulus μ and Poisson's ratio ν. With these, the resulting stresses for the screw dislocation along the $\underline{x}$-axis can be written as $\sigma_{yz} = \sigma_{xx} = \sigma_{yy} = \sigma_{zz} = 0$ and

$$\sigma_{xy} = -\frac{\mu b}{2\pi}\frac{z}{y^2+z^2}, \quad \sigma_{xz} = \frac{\mu b}{2\pi}\frac{y}{y^2+z^2} \tag{4a}$$

or

$$\sigma_{x\theta} = -\frac{\mu b}{2\pi r} \tag{4b}$$

with all other components vanishing in cylindrical coordinates.

The stresses and strains around the dislocation mean that a certain elastic energy is stored by the dislocation. As mentioned in the introduction, this energy accounts for 80-90% of the total energy of a dislocation and therefore controls many of its properties. In general, the elastic energy E can be calculated by integrating over the volume the elastic energy density, which is $\sigma_{ij}\varepsilon_{ij}/2$. For a dislocation it is more useful to calculate the energy per unit length, E^*, by integration over the area normal to it. Here and subsequently, the asterisk is meant to indicate a length specific unit.

$$E^* = \frac{1}{2}\int \sigma_{ij}\ \varepsilon_{ij}\ dy\ dz \tag{5}$$

For the screw dislocation it is most convenient to use cylindrical coordinates and equation (4b) for $\sigma_{x\theta}$. In case of elastic isotropy with $\varepsilon_{x\theta} = \sigma_{x\theta}/\mu$, this

yields

$$E^*(\text{screw}) = \frac{1}{2}\int_0^{2\pi}\int_{r_{\text{core}}}^{r_{\text{shield}}} \frac{\mu b^2}{4\pi^2 r^2} r \; dr \; d\theta = \frac{\mu b^2}{4\pi}\ln\frac{r_{\text{shield}}}{r_{\text{core}}} \qquad (6a)$$

Here the integration has been restricted to the range $r_{\text{core}} < r < r_{\text{shield}}$ because the integral diverges for $r \to 0$ and for $r \to \infty$. The divergence of $E^*(r \to 0)$ arises obviously because the validity of linear elasticity breaks down when r reaches down to atomic scales, that is a few times b. Hence with $r_{\text{core}} = 3b$, for example, the energy of the dislocation core is left out of E^*. Anyway, with an appropriate choice of r_{core}, the missing core energy with its atomistic nature can still be taken into account approximately: this is usually done with a smaller core radius $r_{\text{core}} \approx 1b$.

In a real specimen the divergence of $E^*(r \to \infty)$ is prevented by the finite size of the specimen. This shows that even if the core energy were known accurately, it is not possible to ascribe a certain value to the total energy of a dislocation. Usually a crystal contains a number of dislocations in each of the opposite the signs. In that case these opposite dislocations compensate for each other's stresses and strains, in effect, shielding each other. Therefore a useful shielding distance r_{shield} is half the distance between nearest dislocations. Because of the logarithmic dependence of $E^*(r_{\text{shield}}/r_{\text{core}})$ and $r_{\text{shield}} \gg r_{\text{core}}$, the exact choice of r_{shield} is not too important.

For the derivation of $\sigma_{x\theta}$ (equation (4b)) it had been assumed that the dimensions of the specimen are infinite in all directions. This condition is obviously violated when r_{shield} is chosen to be finite in equation (6a), which uses $\sigma_{x\theta}$. When $\sigma_{x\theta}$ is derived for a screw dislocation in a rod with free cylindrical boundaries, E^* comes out to be lower than in equation (6aa) by the term $\mu b^2/(4\pi)$. Hence altogether, a good approximation for the energy of a straight screw dislocation is

$$E^*(\text{screw}) = \frac{\mu b^2}{4\pi}\left(\ln\frac{r_{\text{shield}}}{b} - 1\right) \qquad (6b)$$

Straight edge dislocations The derivation of stresses and strains is more elaborate for straight edge than for screw dislocations. The reasons are that there is no radial symmetry around the dislocation, and that there are normal stresses $(\sigma_{xx}, \sigma_{yy}, \sigma_{zz})$ and normal strains $(\varepsilon_{xx}, \varepsilon_{yy}, \varepsilon_{zz})$ involved. Both can be seen in figure 3: The dislocation is the line which terminates the half-plane inserted from the top or squeezed into the rod from the right side during the hypothetical generation procedure outlined. Obviously, there is a compressive stress σ_{xx} above the light grey plane and a tension below it,

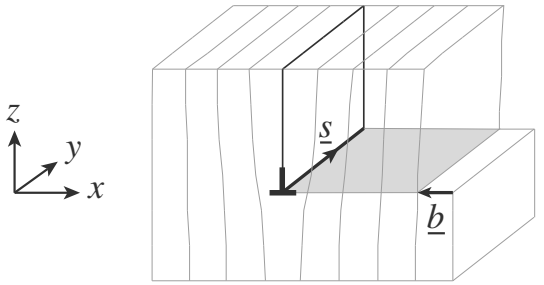


Figure 3. A straight edge dislocation is equivalent to an additional or missing half plane.

where the lattice is widened in x-direction. The derivation of all the stress components are omitted here, but the results are compiled for later reference. For an edge dislocation in an infinite medium with the line vector $\underline{s}$ equal to the y-axis and a Burgers vector $\underline{b}$ pointing in positive x-direction (inverse to the case of figure 3), the stress components are $\sigma_{xy} = \sigma_{yz} = 0$ and

$$\begin{aligned} \sigma_{xx} &= \frac{\mu b}{2\pi(1-\nu)} \frac{-z\left(3x^2+z^2\right)}{\left(x^2+z^2\right)^2}, \quad \sigma_{zz} = \frac{\mu b}{2\pi(1-\nu)} \frac{z\left(x^2-z^2\right)}{\left(x^2+z^2\right)^2}, \quad (7) \\ \sigma_{yy} &= \nu\left(\sigma_{xx}+\sigma_{zz}\right), \quad \sigma_{zx} = \frac{\mu b}{2\pi(1-\nu)} \frac{x\left(x^2-z^2\right)}{\left(x^2+z^2\right)^2} \end{aligned}$$

Again, elastic isotropy is assumed, and ν denotes the Poisson ratio. The strain energy of the edge dislocation is found in full analogy to E^*(screw) of equation (6a). This yields:

$$E^*(\text{edge}) = \frac{\mu b^2}{4\pi(1-\nu)} \left(\ln \frac{r_{\text{shield}}}{b} - 1\right) \quad (8)$$

Dislocations with mixed character Edge and screw dislocations have their line vectors $\underline{s}$ normal or parallel, respectively, to the Burgers vector $\underline{b}$. As is obvious from the hypothetical dislocation generation, a general dislocation has neither edge nor screw, but mixed character. In case of a curved dislocation, the character even varies locally along the dislocation line: $\underline{b}$ is constant, but $\underline{s}$ varies (figure 1). Still, the distinction between edge and screw character remains useful. For instance, the energy E^* of a straight dislocation with mixed character can be superimposed from equations (6ab) and (8). This is possible because parallel screw and edge dislocations do

not interact with each other (within linear elasticity): a screw dislocation along the x_2-direction (parallel to the edge dislocation in figure 3) has only σ_{12} and σ_{23} as non-vanishing stress components (and the respective strain components ε_{12} and ε_{23}); all other components equal zero. This can be seen from (4b)) with x_1- and x_2-directions exchanged. For the edge dislocation this is just vice versa (see above equation (8)): $\sigma_{12} = \sigma_{23} = 0$. With this and the Peach-Koehler formula detailed in section 1.3, it can be shown that E^* of the mixed-character dislocation can be written as:

$$E^*\left(\underline{b}^0, \underline{s}^0\right) = \left|\underline{b}^0 \times \underline{s}^0\right|^2 E^*(\text{edge}) + \left|\underline{b}^0 \cdot \underline{s}^0\right|^2 E^*(\text{screw}) \tag{9}$$

Here and subsequently, the superscript 0 in $\underline{b}^0$ and $\underline{s}^0$ indicates that the respective vectors have unit length. The fact that E^* is an energy per unit length suggests that the total energy of a curved dislocation can be found by integrating equation (9) along its curved path. However, this is only an approximation. The theory of curved dislocations is detailed in chapter 2.

Prismatic dislocation loops In case of the dislocations described in the previous sections, the shift vector $\underline{b}$ lay in the cutting plane. If $\underline{b}$ is normal to this plane, a layer of material must be inserted between the two cutting surfaces before rejoining, as sketched in figure 4, or removed. The dislocation bordering the cutting plane has edge character regardless of its course in this plane. Consequently, its energy E^* is essentially given by equation (8). If the shape of the loop is assumed circular, the shield radius r_{shield} is given by the radius of the loop: the dislocation part on the opposite site of the loop has the opposite sign, because the line vector points in the opposite direction there. Prismatic loops represent dislocation dipoles; as such they hardly contribute to plasticity.

1.3 Moving Dislocations

The most important property of dislocations is that they enable plastic deformation: the overall shape of a crystal changes permanently when dislocations move. For instance, a cube is transformed into the shape in figure 5 when a dislocation with the indicated Burgers vector $\underline{b}$ moves through it in the grey surface. There are other mechanisms which enable plastic deformation, like twinning, grain boundary sliding and phase transformations, but in general, plasticity is dominated by dislocations.

Dislocation glide and climb Two basic types of dislocation motion are to be distinguished: glide and climb. In figure 5, the effect of dislocation glide has been sketched. Note that this shape may have resulted from the

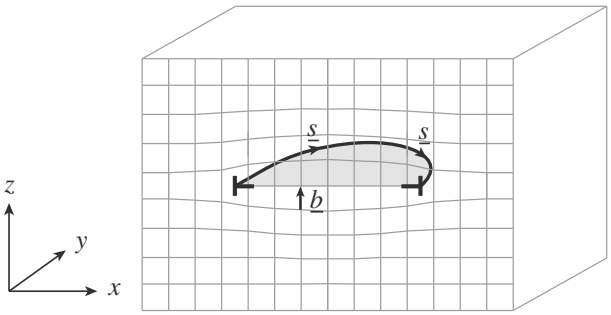


Figure 4. Half a prismatic dislocation loop.

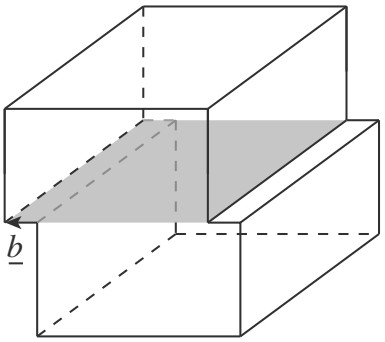


Figure 5. Sheared Cube after glide of dislocations in figures 1, 2 or 3.

motion of a screw dislocation, an edge dislocation or even a curved one: In figure 2, the screw dislocation may have moved from the front to the back plane; in figure 3, the edge dislocation may have moved from the right to the left side. Or in figure 1, the quarter dislocation loop may have expanded to the back left corner. For all three cases the grey area in figure 5 indicates the glide plane. It is described by the dislocation's line vector $\underline{s}$ and, in principle, by the direction of its motion, which is always perpendicular to $\underline{s}$. It is more convenient, though, to use $\underline{s}$ and the Burgers vector $\underline{b}$ to define the glide plane and its normal vector $\underline{n}^0$:

$$\underline{n}^0 = \frac{\underline{b} \times \underline{s}}{|\underline{b} \times \underline{s}|} \tag{10}$$

In the case of a screw dislocation, where $\underline{s}$ is parallel to $\underline{b}$, the glide plane $\underline{n}^0$ is undefined: The screw dislocation can glide in any direction — but only

in the continuum description used here. In real crystals, Burgers vectors and glide planes are restricted to certain crystallographic directions (section 1.4). But within these restrictions, several glide directions (and hence actual glide planes) are equivalent for screw dislocations. This means that screw dislocations can even switch their glide plane; this effect is called cross-slip and detailed in section 4.1.

For edge and mixed-character dislocations, though, the glide plane $\underline{n}^0$ is fixed. The motion normal to $\underline{n}^0$ is called climb. It is associated with material transport: when the edge dislocation in figure 3 moves up or downwards (normal to the glide plane), material must be moved away from or towards the dislocation line, respectively. This changes the total volume of the rod; therefore, climb is called non-conservative motion, as opposed to glide, which is conservative. The material transport can happen by diffusion. Except at very high temperatures, diffusion is a very slow process. Hence climb is extremely slow as compared to glide, which can proceed at sound velocity. Therefore in most cases, plastic deformation is dominated by glide.

In case of the prismatic loop of figure 4, the glide plane is a cylinder: the dislocation can glide up and down. To extend the loop, in contrast, reguires material to diffuse towards the dislocation. However, prismatic dislocation loops are usually of minor importance.

Macroscopic shape change Since the Burgers vector is usually very small compared to the sample's dimensions, the shape change caused by dislocation glide can be considered as continuous since the steps on a surface like in figure 3 are usually negligibly small. Still, the overall shape change is of importance, for instance its elongation. To quantify the length change in relation to dislocation glide, a rod is sketched in figure 6 before and after the glide of a dislocation with arbitrary glide plane and Burgers vector. The rod suffers the elongation $\Delta h = b \cdot \cos\lambda$, where λ is the angle between $\underline{b}$ and the z-direction, in which the length h is measured. The relative length change, $\Delta\varepsilon = \Delta h/h$, called strain, can be written as

$$\Delta\varepsilon = \frac{b\ \cos\lambda}{h} = b\ \cos\lambda \frac{\Delta x}{V}\Delta y = b\ \cos\lambda \frac{\Delta x}{V}\ \Delta y_{\mathrm{g}}\ \cos\phi \tag{11}$$

Here $V = h\Delta x\Delta y$ is the volume of the rod, Δy_{g} is the distance the dislocation moves in its glide plane, and ϕ is the angle between the glide plane vector $\underline{n}$ and the z-direction. The expression $\Delta x/V$ can be interpreted as the dislocation density in the rod, which is defined as the total dislocation length (here: Δx) per volume. More accurately, $\rho_{\mathrm{m}} = \Delta x/V$ is the mobile dislocation density because only dislocations that actually move contribute to deformation. If we divide equation (11) by a time step Δt and read

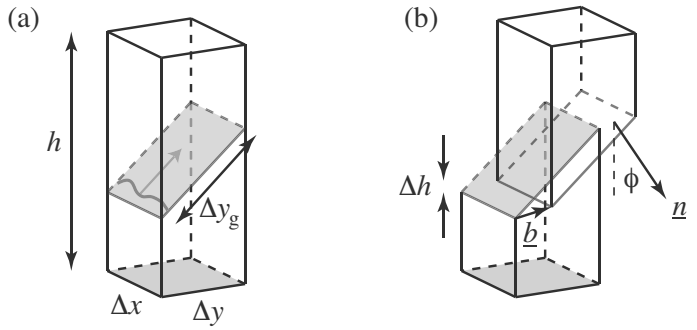


Figure 6. A rod before and after glide of a single general dislocation.

$\Delta y_{\mathrm{g}}/\Delta t$ as the dislocation's velocity v, we get the strain rate $\mathrm{d}\varepsilon/dt$:

$$\dot{\varepsilon} = b\ \rho_{\mathrm{m}}\ v\ \cos\lambda\ \cos\phi \tag{12}$$

When we introduce the Schmid factor $S = \cos\lambda\cos\phi$ and the resolved strain rate or 'shear rate' $\dot{\gamma} = \dot{\varepsilon}/S$, we obtain the Orowan equation. It relates the macroscopic amount of shear strain to the abundance and velocity of moving dislocations (regardless of any forces that drive the dislocations)

$$\dot{\gamma} = b\ \rho_{\mathrm{m}}\ v \tag{13}$$

Forces on dislocations When dislocation motion enables a shape change of a piece of material, then the corresponding external work must be 'consumed' by the moving dislocation, for instance, for overcoming obstacles. From this fact the forces on dislocations can be derived. In figure 7, a general force $\underline{F} = (F_x, F_y, F_z)$ is applied to shift the upper half of a cube to the left along the vector $\underline{b}$ (figure 5). If this force is caused by a general stress tensor $\boldsymbol{\sigma}$ acting on the top surface A_z of the cube, then $\underline{F}$ can be written as $\underline{F}_A = (A_z\sigma_{zx},\ A_z\sigma_{zy},\ A_z\sigma_{zz})$. With the shift vector $\underline{b}$, the external work is

$$\Delta W_{\mathrm{g}} = \underline{b} \cdot \underline{F}_A = b_x A_z \sigma_{zx} + b_y A_z \sigma_{zy} + b_z A_z \sigma_{zz} \tag{14}$$

The subscript g in ΔW_{g} stands for the 'global' view of the energy. The deformation is accomplished by an edge or screw dislocation, which takes a more local view: when it glides a distance d while a general force F is acting on it (F and d have the same directions), it consumes the energy $\Delta W_{\mathrm{l}} = Fd$. Here the subscript l stands for 'local'. By using $\Delta W_{\mathrm{l}} = \Delta W_{\mathrm{g}}$ (no other energies are involved) and the dislocation's length $L = A_z/d$

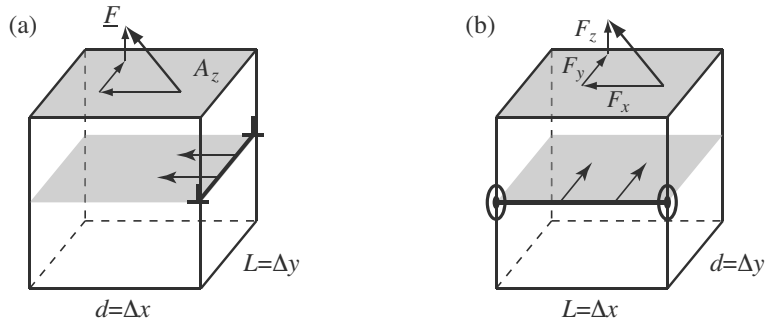


Figure 7. A force F is used to transform a cube into the shape of figure 5. The corresponding external work is equal to the energy that an edge (a) or screw (b) dislocation 'consumes' while gliding.

(figures 7(a) or 7(b)), we get

$$\frac{F}{L} = b_x\sigma_{zx} + b_y\sigma_{zy} + b_z\sigma_{zz} \tag{15a}$$

or, if we define $\tau = F/(bL)$,

$$\tau = b_x^0\sigma_{zx} = b_y^0\sigma_{zy} + b_z^0\sigma_{zz} \tag{15b}$$

τ is called the resolved shear stress. It is useful to keep in mind that by the definition just introduced, the resolved shear stress multiplied by the (constant) magnitude of the Burgers vector of a dislocation equals the force per unit length on this dislocation: $b\tau = F/L$. It should also be noted that in the derivation of τ we made no assumption about the line vector $\underline{s}$, hence τ is independent of $\underline{s}$. However, we assumed that $\boldsymbol{\sigma}$ acts only on the top surface A_z in figure 7, which is parallel to the glide plane, and underhand we prevented rotation. If we assume $\boldsymbol{\sigma}$ to act on all surfaces instead, rotation is avoided automatically and equation (15b) takes the more general form:

$$\tau = \underline{b}^0 \cdot \boldsymbol{\sigma} \cdot \underline{n}^0 \tag{16}$$

Again, n_0 is the normalized glide plane vector. If we assume $\underline{n}^0 = (0, 0, 1)$, we retain equation (15b). Since the Burgers vector $\underline{b}$ lies in the glide plane its component b_z vanishes; from this it can be seen that only shear components of a stress tensor $\boldsymbol{\sigma}$ give rise to τ and hence drive dislocation glide (σ_{zx} and σ_{zy} of equation (15b)). Equation (16) projects any stress tensor $\boldsymbol{\sigma}$ into the 'glide system', defined by the glide plane $\underline{n}$ and the Burgers vector $\underline{b}$. Hence,

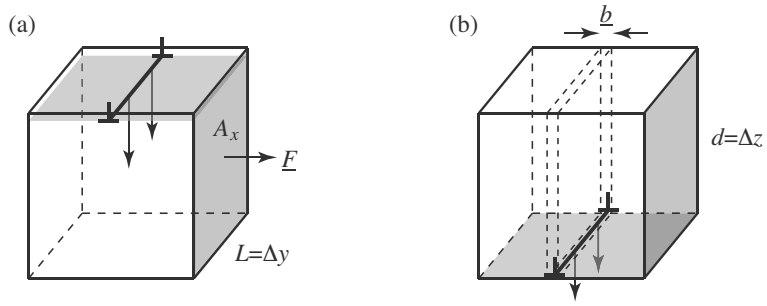


Figure 8. A force $\underline{F}$ is used to elongate a cube, which gives rise to a climb force on an edge dislocation.

it is also a generalized version of Schmid's law. The latter considers only a single stress component like in figure 6, where the sketched elongation may be caused by a force in z-direction on the upper and lower surfaces A_z. This means that all stress components of $\boldsymbol{\sigma}$ vanish except for σ_{zz}. Equation (16) can then be written as $\tau = S\sigma_{zz}$ where the Schmid factor, $S = \cos\lambda\cos\phi$, projects the normal stress σ_{zz} into the glide system.

But equation (16) is restricted to dislocation *glide* (as the 'glide system' implies). The driving force on an edge dislocation to *climb* can be derived in a similar way (figure 8): the external work $\Delta W_\mathrm{g} = \underline{F}_A \cdot \underline{b} = A_x\sigma_{xx}b_x + A_x\sigma_{xy}b_y + A_x\sigma_{xz}b_z$ is set equal to the energy $\Delta W_\mathrm{l} = Fd$ consumed by the climbing dislocation. From this, the climb force per unit length, F/L, and the respective climb stress τ_c (c stands for climb) is found to be

$$\tau_\mathrm{c} = b_x^0\sigma_{xx} + b_y^0\sigma_{xy} + b_z^0\sigma_{xz} \tag{17}$$

τ_c renders a driving force for material (interstitial atoms or vacancies) to move towards or away from the dislocation. When all surfaces A_x, A_y, and A_z are taken into account, τ and τ_c, which are both scalar values, can be combined to form the Peach-Koehler-formula (Peach and Koehler, 1950). It describes the local force vector $\underline{\Delta F}$ on a dislocation segment described by the line vector of finite length Δs:

$$\underline{\Delta F} = (\underline{b} \cdot \boldsymbol{\sigma}) \times \underline{\Delta s} \tag{18}$$

It should be emphasized that the cause for the stress $\boldsymbol{\sigma}$ plays no role here; it may result from an external force $\underline{F}$ as assumed above, from other dislocations as by equations (4) or (7), or from any other kind of stress source.

Forces on dislocations arise not only from elastic stresses $\boldsymbol{\sigma}$, but also

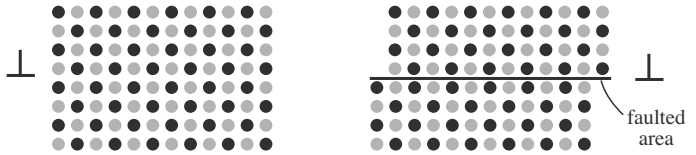


Figure 9. A dislocation creating an antiphase boundary while gliding through a long-range ordered crystal.

whenever the overall energy of a specimen is changed as a dislocation moves. For example, when the atoms in a specimen are usually ordered in some way, this order may be destroyed when a dislocation glides (or climbs) through it. A simplified case is sketched in figure 9, where a dislocation creates an antiphase boundary while gliding through a long-range ordered crystal. The boundary means an energy increment by the amount $E = A\gamma_{\text{fault}}$, where A is the faulted area and γ_{fault} is the specific fault energy. A dislocation of length L moving by a distance x sweeps the area $A = xL$, therefore, senses the force $F = -\mathrm{d}E/\mathrm{d}x = -L\gamma_{\text{fault}}$. The force per unit length F/L equals $-\gamma_{\text{fault}}$, and by the definition above the corresponding shear stress is

$$\tau_\gamma = -\gamma_{\text{fault}}/b \tag{19}$$

An energy change depending on dislocation motion can also arise when the elastic properties, the shear modulus μ in particular, vary locally inside a crystal. As can be seen from the derivation of equations (6ab) or (8), a dislocation's energy is stored in the elastic distortions around its core, and it is proportional to μ. For instance when a dislocation approaches a region with low shear modulus (a constant lattice constant is assumed here), its energy will decrease. Hence there will be a force that attracts the dislocation towards the region with low μ.

Interaction between straight dislocations With the stress tensors of screw and edge dislocations given by equations (4) and (7), respectively, and applying the Peach-Koehler equation (16), we can calculate the driving stress τ for glide that parallel dislocations impose on each other. To derive some qualitative insight in the interaction between dislocations we consider two dislocations as a stress source and a receiver, lying in parallel glide planes with normal vector $\underline{n}^0 = (0, 0, 1)$ and with parallel line vectors $\underline{s}^0 = (0, 1, 0)$, as sketched in figure 10. If we assume two screw dislocations first, their Burgers vectors are $\underline{b}_{\text{s}} = (0, b_{\text{s}}, 0)$ and $\underline{b}_{\text{r}} = (0, b_{\text{r}}, 0)$, and their mutual

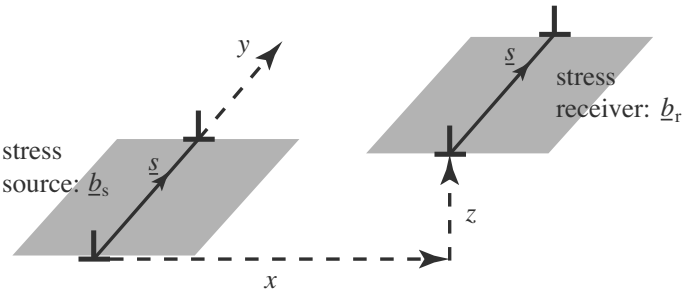


Figure 10. A stress receiving dislocation parallel to a source dislocation. The respective Burgers vectors $\underline{b}_r$ and $\underline{b}_s$ are meant to be arbitrary.

glide stress $\tau_{\text{screw}\to\text{screw}}$ is

$$\tau_{\text{screw}\to\text{screw}} = \frac{\mu b_s}{2\pi} \frac{x}{x^2 + z^2} \tag{20}$$

If the stress receiver is located right from the source ($x > 0$), it senses a positive stress $\tau_{\text{screw}\to\text{screw}}$ which drives it to more positive x-values, i.e. away from the source: parallel screw dislocations repel each other if they have the same sign. With either $\underline{b}_s$ or $\underline{b}_r$ inverted, the dislocations are called antiparallel, and they attract each other. When they meet, they annihilate each other. For instance, the two opposing half planes of antiparallel edge dislocations form a full plane of the crystal. For equation (20) it had been assumed that the dislocations are bound to the glide plane $\underline{n}^0 = (0, 0, 1)$. Pure screw dislocations in a continuous medium are not bound to a glide plane; for this case it may be derived that $\tau_{\text{screw}\to\text{screw}} = \mu b_s/(2\pi d)$, where d is the distance between the dislocations. However, in real crystals (section 1.4), screw dislocation motion is still bound to certain crystallographic planes.

If we consider two edge dislocations (as sketched in figure 10), the Burgers vectors are $\underline{b}_s = (b_s, 0, 0)$ and $\underline{b}_r = (b_r, 0, 0)$, and the mutual glide stress $\tau_{\text{edge}\to\text{edge}}$ is much more complicated. In particular, the sign of the interaction depends on x and z:

$$\tau_{\text{edge}\to\text{edge}} = \frac{\mu b_s}{2\pi(1-\nu)} \frac{x\left(x^2 - z^2\right)}{\left(x^2 + z^2\right)^2} \tag{21}$$

If the stress receiver is close to the source's glide plane ($z^2 < x^2$) the edge dislocations with equal sign repel each other, essentially like in the case of

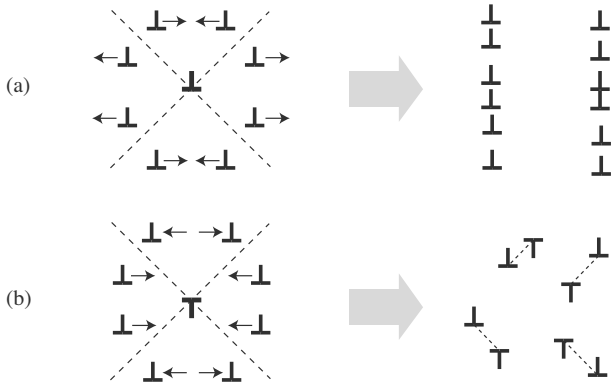


Figure 11. Interaction between edge dislocations in parallel glide planes: (a) parallel ones tend to form low angle boundaries; (b) antiparallel ones tend to form dipoles.

screw dislocations. But in the opposite case $z^2 < x^2$ the receiver is attracted, so that these dislocations tend to group above each other along the z-direction (figure 11(a)). Such groups form an interface between two parts of a crystal with slightly tilted crystallographic orientation: They are low angle tilt grain boundaries.

When edge dislocations of opposite sign are considered, $\tau_{\text{edge}\to\text{edge}}$ just switches its sign. The effect of this can be seen in figure 11(b): the stress receiving dislocation is attracted to the 45°-lines where $z^2 = x^2$; there, $\tau_{\text{edge}\to\text{edge}} = 0$. This means that edge dislocations of opposite sign tend to form dislocation dipoles, with a 45° angle between the glide plane and the shortest vector connecting both.

Dislocation dipoles have the important property that they cannot be driven by a homogeneous stress $\boldsymbol{\sigma}$: the two edge dislocations are driven in opposite directions. This may break up the dipole, but it will not drive it. In this sense, dipoles can be considered as immobile, in contrast to the mobile dislocations which actually enable plasticity. Even if they move, the bits of deformation the two edge dislocations cause cancel out each other for the most part.

Another important feature of dipoles is that the dislocations involved can move towards each other by a mix of glide and climb and finally annihilate. Climb requires material transport by diffusion, and hence this annihilation is kinetically inhibited. But since the edge dislocations attract each other, the dipole renders a driving force which attracts or emits interstitial atoms

or vacancies.

Between parallel edge and screw dislocations there is no elastic interaction: the combinations of $\underline{b}_\mathrm{s} = (0, b_\mathrm{s}, 0)$, $\underline{b}_\mathrm{r} = (b_\mathrm{r}, 0, 0)$, and $\underline{b}_\mathrm{s} = (b_\mathrm{s}, 0, 0)$, $\underline{b}_\mathrm{r} = (0, b_\mathrm{r}, 0)$, both yield $\tau_{\mathrm{screw}\to\mathrm{edge}} = 0$ and $\tau_{\mathrm{edge}\to\mathrm{screw}} = 0$, respectively. Hence without energy change, parallel screw and edge dislocations can be moved to fully coincide. By this procedure a straight dislocation with mixed character is generated. Conversely, this means that any mixed character dislocation with Burgers vector $\underline{b}$ may be considered as being composed of an edge and a screw dislocation with Burgers vector components $b_\mathrm{edge} = |\underline{b} \cdot \underline{s}^0|$ and $b_\mathrm{screw} = |\underline{b} \times \underline{s}^0|$, respectively. With these, the stress tensor $\boldsymbol{\sigma}$ of a mixed character dislocation may be written as

$$\boldsymbol{\sigma}^{\mathrm{mixed}}(\underline{b}) = \boldsymbol{\sigma}^{\mathrm{screw}}\left(b_\mathrm{screw}\right) + \boldsymbol{\sigma}^{\mathrm{edge}}\left(b_\mathrm{edge}\right) \tag{22}$$

where the components of $\boldsymbol{\sigma}^{\mathrm{screw}}$ and $\boldsymbol{\sigma}^{\mathrm{edge}}$ are, in principle, given by equations (4a) or (7), respectively. But before that, equation (4a) must be transformed such that the line vector $\underline{s}$, which had been assumed to point in x-direction for equation (4a) (figure 2), equals that for equation (7) (figure 3). This geometry has for instance been used in (Hirth and Lothe, 1992). However, in the present description, it has been chosen to use the same Burgers vector $\underline{b}$ for the cases in figures 1 to 3 to emphasize that $\underline{b}$ of a given dislocation is constant, so that a dislocation's character is exclusively determined by (variations of) the line vector $\underline{s}$ since, a variation of $\underline{b}$ is impossible for a given dislocation.

Image stresses So far it has been assumed that the dislocations reside in a specimen with infinite dimensions. However, often boundary conditions like free surfaces need to be considered. As is usual in other physical boundary condition problems, this can be done by assuming infinite dimensions, but adding entities which enforce the conditions involved with a given boundary. In case of elasticity and hence dislocations, a free boundary means that all forces on the boundary surface must vanish. In the case of figure 12, in which a free surface is assumed normal to the x-axis, this means $\sigma_{xx} = \sigma_{xy} = \sigma_{xz} = 0$. For a screw dislocation, this can easily be achieved by adding an additional dislocation of opposite sign mirrored to the opposite side of the surface: from equation (4b) it can be seen that the only non-vanishing stress component in cylindrical coordinates of a dislocation along the y-axis, $\sigma_{y\theta}$, depends only on the distance $r = \left(x^2 + z^2\right)^{1/2}$, which is always positive. Hence, all stresses of the dislocation (with $\underline{b}^{(1)}$, at $x^{(1)} = -L$, figure 12(a)) and its image counterpart (with $\underline{b}^{(2)} = -\underline{b}^{(1)}$, at $x^{(2)} = +L$) combined vanish in the whole plane $(0, y, z)$. The original

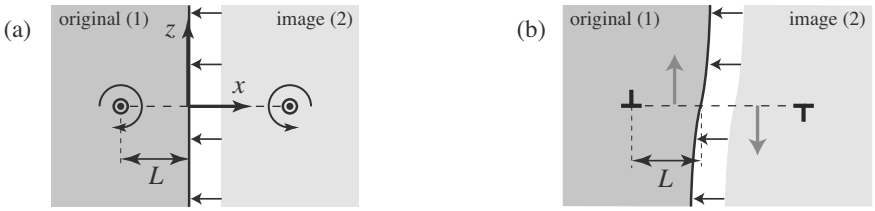


Figure 12. Material with image dislocations used to account for a free surface; (a) exact solution for a screw dislocation; (b) approximation for an edge dislocation.

dislocation is attracted towards the image dislocation and hence towards the free surface as described by equation (20) with $z = 0$ and $x = 2L$, i.e. with $\tau_{(2)\to(1)} = \mu b/(4\pi L)$.

The case of an edge dislocation parallel to a free surface is more complex. For the geometry used in figure 12(b), the stress components σ_{xy} vanish for both the original and the mirrored dislocation (section 1.2), and the componens σ_{xx} of the two compensate for each other (equation (7)). But in case of σ_{xz}, the components for the two dislocations add up because the sign of σ_{xz} depends not only on that of $\underline{b}$ but also on that of x (equation (7)). In figure 12(b) the non-vanishing stress components are indicated by grey arrows. In case of straight dislocations, the interaction between the dislocation and its image is still given by equation (21), with $z = 0$ and $x = 2L$, which results to $\tau_{(2)\to(1)} = \mu b/(4\pi L(1-\nu))$ (Hirth and Lothe, 1992). The stress components σ_{xz} are simply removed by a minor translational shift along the directions of the grey arrows.

The case of a free surface may be considered as an interface inside a specimen where the shear modulus switches from μ to zero, or where the material is infinitely soft. The other extreme is an interface to an infinitely rigid material ($\mu \to \infty$), which allows no elastic (or plastic) strains: $\varepsilon_{xx} = \varepsilon_{xy} = \varepsilon_{xz} = 0$. Similar to the case discussed above, this can be achieved by adding material with an image dislocation with the same Burgers vector: $\underline{b}^{(2)} = +\underline{b}^{(1)}$. In that case the dislocation experiences the same amount of stresses $\tau_{(2)\to(1)}$, but with inverted sign: The dislocation is repelled from the rigid part.

1.4 Dislocations In Real Crystals

In real crystals, the Burgers vector $\underline{b}$ and the glide plane vector $\underline{n}$ are bound to crystallographic directions: the Burgers circuit (section 1.1) is to

be performed along atomic positions, leaving an interatomic vector for $\underline{b}$. Since the energy of dislocations scales with b^2, the Burgers vector must be short; this leaves only a few possibilities. For a dislocation to glide, a glide plane must be 'smooth'. This is the case only when its normal vector points in certain crystallographic directions.

Perfect dislocations In a face centred cubic (fcc) lattice the elementary cell consists of a cube with one eighth of an atom in all eight corners and half an atom on all six faces. The shortest interatomic distance in this lattice is along the diagonals of the sides; hence in terms of lattice vectors, the Burgers vector has the type $\langle 110 \rangle$ and the length $a_0/2^{1/2}$, where a_0 is the lattice constant. The closest packed planes are normal to the space diagonal, hence the glide planes have Miller indices of the type $\{111\}$. In each of the four independent glide planes (111), $(\bar{1}11)$, $(1\bar{1}1)$, and $(11\bar{1})$, there are three possible Burgers vector directions such that $\underline{b} \cdot \underline{n} = 0$ (for instance $[\bar{1}10]$, $[\bar{1}01]$ and $[01\bar{1}]$ for (111)), hence there are 12 glide systems of this kind. One example is sketched in figure 13. An edge dislocation in the depicted $(1\bar{1}1)$-plane with the indicated Burgers vector would have the line vector $s^0 = 6^{-1/2}[121]$ along the dark atoms. These latter atoms can be seen as the first ones of the missing half plane inside the volume that has not been plotted. There are actually two rows of dark atoms along the dashed lines; this will be detailed later on in the context of partial (i.e. non-perfect) dislocations.

The elementary cell of a body centred cubic (bcc) lattice is a cube with one eighth of an atom in all eight corners and one atom in its centre. The shortest interatomic vector and hence the Burgers vector is of the type $\langle 111 \rangle$ and has the length $\left(3^{1/2}/2\right) a_0$. There are three possible types of glide planes which are experimentally observed: $\{110\}$, $\{112\}$ and $\{123\}$. Among these, the $\{110\}$-planes sketched in figure 14 have the closest packing. There are six non-parallel planes of this type, and each of them can contain two Burgers vectors, for instance, $[111]$ and $[11\bar{1}]$ for the $(\bar{1}10)$-plane. So there are 12 glide systems of this type. Of the $\{112\}$-planes there are 12. Each can contain only one Burgers vector, for instance $[11\bar{1}]$ for the (112)-plane. This gives another 12 glide systems. Finally there are 24 $\{123\}$-planes, which contain one Burgers vector each. Hence there are 48 glide systems altogether in body centred cubic crystals.

In principle, any plane qualifies as a glide plane as long as it contains the Burgers vector. For instance, the (134)-plane contains the direction $[11\bar{1}]$, but glide on $\{134\}$-planes is not observed: the glide is prevented by the Peierls stress τ_{Peierls}. This stress describes all resistance against glide that a dislocation experiences due to the atomistic nature of its core in a

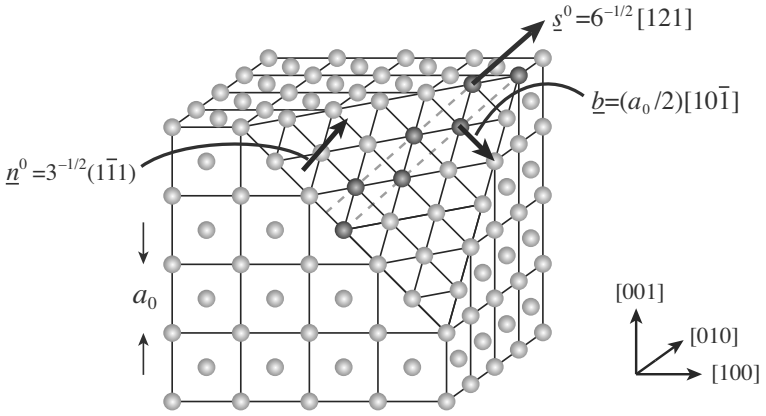


Figure 13. Glide plane $\underline{n}$ and Burgers vector $\underline{b}$ in a face centred cubic crystal. The dark atoms indicate the direction of an edge dislocation, for instance the first atoms of the missing half plane.

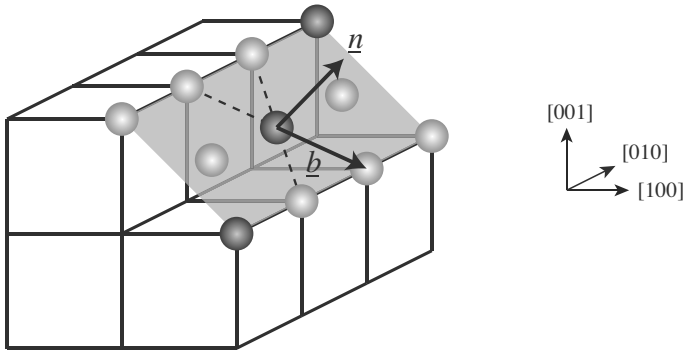


Figure 14. A glide plane $\underline{n} = (101)$ and Burgers vector $\underline{b} = [11\bar{1}]$ in a body centred cubic crystal. The dark atoms indicate the direction $[\bar{1}21]$ of an edge dislocation, for instance the first atoms of the missing half plane.

given plane. This resistance has a short-ranged periodic nature inherited from the crystal structure, the glide plane and the glide direction. Brownian motion of the atoms, i.e. thermal pulses on them, helps a dislocation to overcome this resistance, as if the glide plane were smoothed out by temperature. Therefore, τ_{Peierls} is a function of temperature. In case of face centred cubic crystals the glide planes are closely packed (figure 13) and therefore 'smooth'; the Peierls stress in these crystals is negligible even at room temperature. But in body centred cubic materials, only the $\{110\}$-planes are rather smooth. The other glide planes $\{112\}$ and $\{123\}$ need to be 'smoothed' by temperature, so that they are active only at higher temperatures.

Among metals the hexagonal crystal structure is also of importance. Here the basal planes $\{0001\}$ enable three glide directions of the type $\langle 11\bar{2}0\rangle$. Furthermore, three prism planes of the type $\{10\bar{1}0\}$ with one glide direction each (type $\langle 11\bar{2}0\rangle$) can be active as well as the pyramidal $\{10\bar{1}1\}$-planes, also with a $\langle 11\bar{2}0\rangle$ glide direction. As long as the axes of the hexagonal crystal have a ratio $c/a \geq 1.63$, the basal plane is closest packed and always active. However, in case of an axis ratio $c/a \leq 1.63$, the density in the plane can get as low as that in the other planes. Again, temperature decides which glide systems get activated first.

Partial dislocations As outlined before, a Burgers vector $\underline{b}$ must be short because a dislocation's energy scales with b^2. So far we have assumed that $\underline{b}$ is a lattice vector, but this is not always required. A face centred cubic crystal, for instance, can be seen as a periodic stack of three closely packed atom layers, as indicated in figure 15(a). The layers A, B, and C differ by small translational vectors for valid atom positions in the plane, as can be seen in figure 15(b). In a perfect fcc crystal, which has the closest possible packing density, a C layer must follow a sequence of A and B to form an $ABCABC$ stack. The same packing density can be found with an $ABABAB$ stacking sequence, but this would mean a slight increase of the free energy. An $ABABAB$ stack would form a hexagonal crystal with closest packing and an axis ratio of $c/a = 1.63$. However, the quoted fcc materials obviously prefer the fcc crystal and, therefore, the $ABCABC$ sequence.

In figure 15(b), part of the atoms in layer B, which lies on top of a C layer, have been shifted such that they reach another valid B-position. This corresponds to the Burgers vector $\underline{b}$ of a perfect dislocation. But instead, these atoms can also shift so that they reach an A-position. This creates a partial dislocation (here: Shockley partial dislocation) and a stacking fault on one side of it. The stacking fault entails an increase of the free energy,

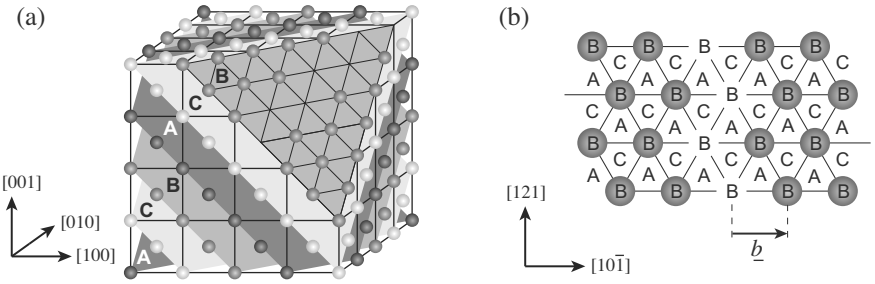


Figure 15. Atom layers in a face centred cubic lattice. A row of atoms is omitted to indicate the first missing atoms of the missing half plane of an edge dislocation.

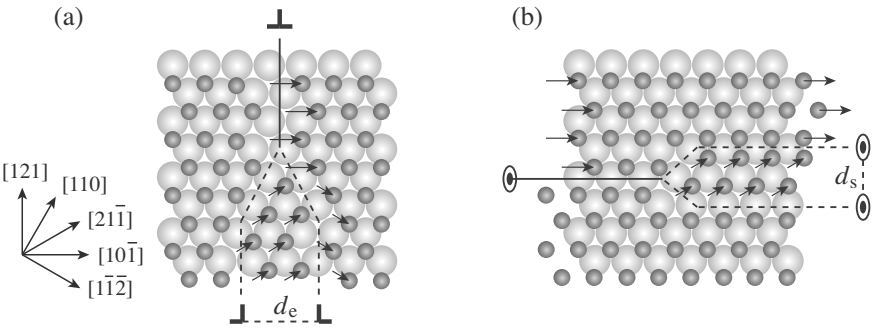


Figure 16. Atom layers adjacent to the glide plane of a perfect/dissociated dislocation of (a) edge and (b) screw character.

but the total energy is lowered because the Burgers vector $\underline{b}_p$ of the partial dislocation (short: partial) is smaller than that of a perfect one. Partials are usually found in pairs of two; this keeps the energy penalty from the stacking fault low. Together such a pair is called a dissociated dislocation. Shockley partial Burgers vectors are of the type $\langle 112 \rangle$ and have the length $b_p = a_0/6^{1/2} = b/3^{1/2}$. The sum of the partial Burgers vectors of a pair equals that of the dissociated perfect dislocation: $\underline{b}_1 + \underline{b}_2 = \underline{b}$. In figure 16 an example is given for the glide plane $\underline{n} = (1\bar{1}1)$ and a screw or edge dislocation with the Burgers vector $\underline{b} = \left(a_0/2^{1/2}\right)(10\bar{1})$ dissociated into $\underline{b}_1 = \left(a_0/6^{1/2}\right)(21\bar{1})$ and $\underline{b}_2 = \left(a_0/6^{1/2}\right)(1\bar{1}\bar{2})$.

It should be noted that the Burgers circuit, which is used to define the Burgers vector, cannot be performed for single partial dislocations unless

the circuit is started and finished in a point in the stacking fault. The region of stacking sequence $ABAB$ around the stacking fault may be considered as a small volume with hexagonal crystal structure. But besides that, partials can be considered as normal dislocations, in particular with respect to all properties deriving from elasticity. In section 1.3 it has been shown that a dislocation with a faulted area on one side of it in the glide plane senses the constant stress $\tau_\gamma = -\gamma_{\text{fault}}/b$ (equation (19)), where γ_{fault} is the energy density of the faulted area. Accordingly, in case of partials with a stacking fault energy γ_{sf}, the stress is $\tau_{\text{sf}} = \pm\gamma_{\text{sf}}/b_{\text{p}}$. The sign of it is such that it drives the partials towards each other. On the other hand, the partials repel each other elastically reciprocal with their distance d, as can be seen from equations (20) or (21) with $z = 0$. The equilibrium distance between partials d_{eq} can be found by equating the repulsive stress with τ_{sf}. Here we have to consider that the partial Burgers vectors are neither normal nor parallel to each other, but they enclose an angle of 60°. As described by equation (22), the dislocations can be decomposed in their screw and edge components. Then the superimposed stress tensors $\boldsymbol{\sigma}^{\text{mixed}}\,(x = d_{\text{eq}}, z = 0)$ for a distance d_{eq} can be subjected to the Peach-Koehler formula (16) to calculate the glide component τ. Equating this with τ_{sf} yields for d_{eq}:

$$d_{\text{eq}} = \frac{\mu b_{\text{p}}^2}{2\pi\gamma_{\text{sf}}} \left[\left(\underline{b}_1^0 \cdot \underline{s}^0\right)\left(\underline{b}_2^0 \cdot \underline{s}^0\right) + \frac{\left(\underline{b}_1^0 \times \underline{s}^0\right) \cdot \left(\underline{b}_2^0 \times \underline{s}^0\right)}{1-\nu} \right]^{-1} \tag{23}$$

When a material with $\nu = 1/3$ is considered with equation (23), edge dislocations are found to dissociate stronger than screw ones by a factor 7/3. This strong difference is caused by the fact that the partials of a pair are in part parallel and antiparallel when decomposed in edge and screw components. The parallel parts repel each other, whereas the antiparallel ones attract. Hence the edge components of the partials of a dissociated edge dislocation repel rather strongly (factor $3/4/(1-\nu)$, for $b_{\text{edge}}/b = 3^{1/2}/2$ with $\alpha = 60°$ for Shockley partials) and attract rather weakly (factor 1/4, for $b_{\text{screw}}/b = 1/2$ with $\alpha = 60°$), whereas in the case of a dissociated screw dislocation, repulsion (factor 3/4 with $\alpha = 30°$) and attraction (factor $1/4/(1-\nu)$) are more balanced and d_{eq} is rather small. The dislocation width is important for the probability of dislocations to cross-slip, as is detailed in section 4.1. Cross-slip means that a screw dislocation leaves its primary glide plane. Perfect screw dislocations have no fixed glide plane, but dissociated ones have a primary glide plane which is defined by the plane of the stacking fault regardless of the dissociation width d_{eq}.

In hexagonal crystals an equivalent stacking fault may occur in the base plane, forming a region with an $ABCABC$ stacking sequence enclosed by

Shockley partials. However, this is only possible when the axis ratio c/a is close to 1.63 because otherwise the stacking fault energy is likely to be high. In body centred cubic crystals dislocation dissociation can even extend in three dimensions.

2 Curved Dislocations

In section 1 on basic dislocation theory it has been emphasized from the start that in general, dislocations are curved and flexible. While several basic dislocation features are understood best by considering straight dislocations, the flexibility should always be kept on mind. As a result of stresses, dislocations can bend locally and thereby change their local line vector $\underline{s}$. The latter thereby changes the local dislocation character (edge/screw), because the Burgers vector $\underline{b}$ is constant.

The degree of flexibility, that is, the stress sensitivity of bending is quite important for the course of events that may happen during plastic deformation. The flexibility is defined by the dislocation's self interaction: a dislocation causes and at the same time senses elastic stresses, including those from itself. Various models have been derived in the past to account for this, with differing complexity and accuracy. Today even the rather complex model for the elastic (self-) interaction of curved dislocations can be utilized in computer programs to simulate dislocation glide in various problems of plasticity. This self interaction model is described in section 2.2. But before that, the simple line tension model is described in detail because it is required for understanding and proper interpretations of simulation results. The description of the line tension model given here differs from other approaches. Unlike other descriptions it is fully consistent and therefore believed to provide a better understanding.

2.1 Line Tension Model

In equation (9) the length specific energy E^* of a straight dislocation with mixed character had been given. The character dependence has been expressed by the directions of the Burgers vector $\underline{b}$ and the line vector $\underline{s}$. Using the angle α between $\underline{b}$ and $\underline{s}$, and $R = r_{\text{shield}}/e$ for simplification, we can write:

$$E^*(\alpha) = \left[\frac{\mu b^2}{4\pi(1-\nu)}\,\sin^2\alpha + \frac{\mu b^2}{4\pi}\,\cos^2\alpha\right] \ln\left(\frac{R}{b}\right) \tag{24}$$

For later reference, the square bracket in equation (24) is called the prelogarithmic energy factor $K_E(\alpha)$. It contains the shear modulus μ and Poisson's ratio ν as elastic constants; elastic isotropy has been assumed in

section 1. In general, $K_E(\alpha)$ is a more complex function (see e.g. (Nembach, 1996) containing more elastic constants, like C_{ijkl}. But for the time being, K_E is considered constant.

Constant line energy With a constant energy per unit length E^*, any object and in particular a dislocation tries to contract itself in order to reduce the free energy. The energy of a segment described by $\underline{\Delta s}$ is $E^*|\underline{\Delta s}|$. This causes the force $\underline{F} = -\underline{\nabla}\,(E^*\Delta s)$, which obviously points along $\underline{\Delta s}$ and has the magnitude $F = E^*$. This means that the dislocation behaves like a string under constant tension in this model. This knowledge can be used to evaluate the force equilibrium in the middle point of figure 17, which holds the force caused by both adjacent dislocation arcs in their given circular shape. The x-components of the forces from the arcs, $F_x = \pm E^* \cos\varphi$, cancel each other; the superimposed y-component is $F_y = 2E^* \sin\varphi$. Assuming small angles φ for simplicity (this is not a necessary condition) such that $\sin\varphi \approx \varphi$, and expressing φ by the arc length L and the curvature radius R_c $(\varphi = L/R_\mathrm{c})$, we find $F_y = E^* L/R_\mathrm{c}$. This is the force that the arcs cause in the middle point. But as mentioned above, the dislocation arcs would try to contract; the same force F_y is needed for each point or arc to keep the dislocation in its shape. Over the length L, the required stress is known to be $\tau_\mathrm{ext} = F_y/(bL)$ (stress definition in section 1.3). Since a stress equilibrium must be fulfilled and no other stresses are involved, this external stress must equal the stress $\tau_{\mathrm{self}-E}$ that the dislocation produces on itself in every point. Altogether this yields:

$$\tau_{\mathrm{self}-E} = \frac{E^*}{b\ R_\mathrm{c}} \tag{25}$$

According to equation (25) the dislocation's self-interaction depends only on the curvature $1/R_\mathrm{c}$, which is a very local property. This is only a rough approximation, as is seen later. In addition, it must be emphasized here that equation (25) holds only for the rather hypothetical case of a constant line energy E^*. The effects of $E^*(\alpha)$ are described later on; before that, we take a look at the importance of a dislocation's flexibility or rigidity in strengthening models. The rigidity is represented here by E^*, and later more accurately by the line tension S.

Simple Strengthening Models Strengthening models relate properties of obstacles, which impede dislocation glide, to the critical stress required to enforce dislocation glide to continue. The simplest but still important strengthening model is that of Orowan (Orowan, 1934). It assumes that dislocation glide in a plane is impeded by obstacles which, for any reasons,

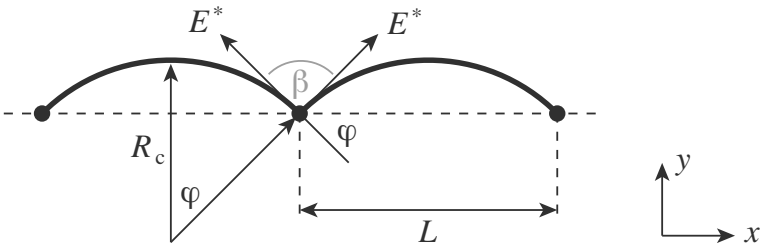


Figure 17. Force on the middle point: the x-components $\pm E^*\cos\varphi$ of the left and right dislocation arcs cancel each other, the y-components $E^*\sin\varphi$ add up. Often the cusp angle $\beta = \pi - 2\varphi$ is used instead of φ.

cannot be cut through by dislocations: the obstacles are impenetrable. Dislocations can still overcome these obstacles provided that an external stress τ_{ext} exceeds some critical stress τ_{Orowan}, as sketched in figure 18(a). Between obstacles, in a static configuration (no motion) τ_{ext} equals $\tau_{\text{self}-E}$ because no other stresses are involved. Hence according to equation (25) τ_{ext} will cause a curvature with the radius $R_{\text{c}} = E^*/(b\tau_{\text{ext}}))$. When τ_{ext} is high enough such that R_{c} is equal or just below the free space $L_{\text{cc}} - 2R$ between the obstacles, the angle φ (see figure 17) will reach $\pi/2$, (or, $\beta = 0$). This defines the Orowan stress

$$\tau_{\text{Orowan}} = \frac{E^*}{b\,(L_{\text{cc}} - 2R)} \tag{26}$$

where R and L_{cc} are the obstacle radius and their centre-to-centre spacings, respectively. At $\tau_{\text{ext}} = \tau_{\text{Orowan}}$, the dislocation arcs are driven forwards by the force $(L_{\text{cc}} - 2R)b\tau_{\text{ext}}$. This force equals $2E^*$ and therefore is just enough to keep the length of the two dislocation arcs that touch one obstacle. When $\tau_{\text{ext}} > \tau_{\text{Orowan}}$ the dislocation length is increased, and the dislocation overcomes the obstacles by circumventing them. Thereby parts of the dislocation with opposite line vectors annihilate each other, and Orowan loops around the particles are left behind, as indicated in figure 18(b). Equation (26) describes the strengthening effect that the impenetrable obstacles give to the material in terms of b, of the obstacle's geometrical parameters L_{cc} and R, and of E^*. For the case $R \ll L_{\text{cc}}$, only the obstacle spacing L_{cc} is of importance, but no obstacle parameters. L_{cc} may be expressed in more useful terms, like a volume density of obstacles. But here only the importance of E^* is emphasized: apart from geometrical factors, the strengthening effect of the obstacles depends on the rigidity E^*, which is a pure dislocation

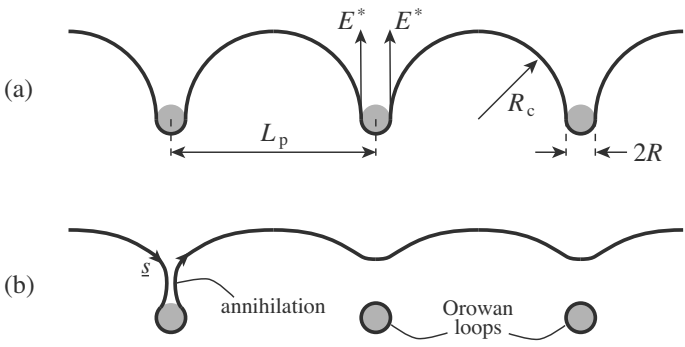


Figure 18. A dislocation overcoming an equidistant row of impenetrable obstacles.

property.

Another simple but important strengthening model is that of Friedel (Friedel, 1964), and derivations thereof. Here the obstacles are not impenetrable, so that they can only hold a finite maximum force $F_{\max}$. However, we assume the obstacle to be point-like ($R = 0$) for simplification. In the terms of figure 17, the finite force $F_{\max}$ means that the range of possible angles is reduced to $\sin\varphi < F_{\max}/(2E^*)$, or $\cos(\beta/2) < F_{\max}/(2E^*)$. This influences the effective number of obstacles that prevent dislocation glide: in a random field of point obstacles, an extremely rigid, i.e. straight dislocation touches only very few obstacles, whereas a rather flexible dislocation ($E^* \to F_{\max}/2$) will touch many. To derive the effect of this, we consider the configurations given in figure 19, where a dislocation first lies behind a row of three equidistant obstacles and then breaks free from the obstacle in the middle. It then bows out until it touches another obstacle in a random obstacle arrangement, thereby sweeping the area A. Before and after the break-through the curvature radius R_c is the same because the external stress τ_{ext} is constant.

Assuming a parabolic dislocation arc, for simplicity, instead of a circular one (restricting the model to low angles φ) before and after break-through, the following purely geometrical relation can be derived:

$$A = 2\frac{2}{3}\,(1 - 2/8)\;\; L_F \cdot h = L_F \cdot h \quad \text{with} \quad h = \frac{1}{2R_c} L_F^2 \tag{27}$$

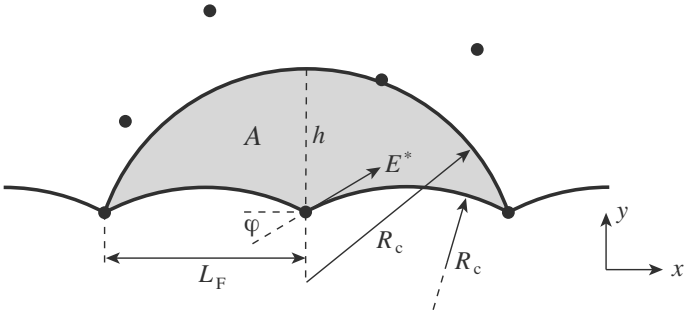


Figure 19. Geometry for Friedel's strengthening model. The grey area A is swept when the dislocation breaks through the obstacle in the middle, until the next obstacle is found.

Here so far, none of the variables A, L_F, and h are known. But it is known that on average during dislocation glide, one new obstacle must be found by the dislocation after it has overcome another. This means that $1/A$ equals the area density of obstacles, c, which is fixed and can be measured, or $A = L_\mathrm{square}^2$, where L_square denotes the square lattice spacing. This is the shortest distance between obstacles if they were arranged in a square lattice. Next, the force equilibrium in the obstacles is applied: at the critical stress τ_Friedel, the force $bL_\mathrm{F}\tau_\mathrm{Friedel}$ acts on one dislocation arc of the length L_F and hence on the obstacles, so we get $F_\mathrm{max} = bL_\mathrm{F}\tau_\mathrm{Friedel}$. Finally, from the stress equilibrium between the obstacles we know that, like in the Orowan model above, $\tau_\mathrm{Friedel} = \tau_{\mathrm{self}-E} = E^*/(bR_\mathrm{c})$. With these considerations, equation (27) can be resolved for the length L_F, called the Friedel length:

$$L_\mathrm{F} = \frac{b}{\sqrt{ck}} = \frac{L_\mathrm{square}}{\sqrt{k}} \quad \text{with} \quad k = \frac{F_\mathrm{max}}{2E^*} \tag{28}$$

Here k denotes a relative obstacle strength. If the obstacles are weak ($\mathrm{F}_\mathrm{max} \ll 2E^*$, or $k \ll 1$), L_F is large compared to the square lattice spacing L_square, which is a constant parameter given by the obstacle array. This means that the dislocation touches only very few obstacles when they are weak, or, as indicated above, when the dislocation is very rigid and remains straight. This emphasizes the importance of the line tension model, because the line energy is needed to rate the obstacle force F_max, even in simulations in which the line tension model is not used at all. However, it must be stressed again that it had been assumed that E^* is constant. If E^* is a function of the dislocation character, often (but not always) the line tension $S(\alpha)$ introduced in section 2.1 is to be used instead of $E^*(\alpha)$. With L_F and

k from equation (28), the strengthening effect τ_{Friedel} of the obstacles can finally be written as:

$$\tau_{\text{Friedel}} = \sqrt{k}\ \tau_{\max} \quad \text{with} \quad \tau_{\max} = \frac{F_{\max}}{bL_{\text{square}}} \tag{29}$$

At a given obstacle spacing L_{square}, the strengthening effect of obstacles obviously increases with $F_{\max}$ with a power of 3/2, but only up to $k = 1$. When $F_{\max} \geq E^*$, only the obstacle spacing is of importance, as seen in the Orowan model.

Line Torque So far it has been shown that a dislocation with the length specific energy E^* is under tension with the force $F = E^*$ along its line direction. The stress $\tau_{\text{self}-E}$ (equation (25)), and the force $\Delta F_E = b\tau_{\text{self}-E}\Delta s$ normal to the segment vector $\underline{\Delta s}$ resulted from this assumption. This holds for a constant line energy E^*. But usually E^* is a function of the dislocation character, indicated by the angle α between Burgers- and line vector. From $E^*(\alpha)$, a dislocation segment of length Δs senses a torque ΔM that attempts to rotate the segment such that its energy $(E^*\Delta s)$ is lowered:

$$M(\alpha) = -\left.\frac{\partial(E^*\Delta s)}{\partial\alpha}\right|_{\Delta s} \quad \text{or} \quad M^*(\alpha) = -\frac{\partial E^*}{\partial\alpha} \tag{30}$$

where $M^* = \Delta M/\Delta s$ is a length specific torque, called line torque in analogy to the line energy. Here Δs is assumed to be constant because the effect of a length change has already been covered by $\tau_{\text{self}-E}$ in section 2.1. The torque tries to rotate the segment, but for the most part this is prevented by the neighbouring segments since they want to rotate in the same direction when they are close and sense the same torque. In figure 20 three segments with the same lengths Δs are plotted. They sense slightly varying torques $\Delta M^{(-1)}$, $\Delta M^{(0)}$, and $\Delta M^{(+1)}$ resulting from a varying M^*. The torques ΔM result in forces in the segments' end points. Between the segments (0) and (+1), the sum of forces in y-direction is $\Delta F_M = \left(\Delta M^{(0)}/\Delta s\right) - \left(\Delta M^{(+1)}/\Delta s\right) = M^{*(0)} - M^{*(+1)}$. With the distance between the connections being the segment lengths Δs, the corresponding stress is

$$\tau_M = \frac{M^{*(0)} - M^{*(+1)}}{b\ \Delta s} = -\frac{1}{b}\frac{\mathrm{d}M^*}{\mathrm{d}s} \tag{31}$$

The second part of equation (31) is the transition to infinitesimally short. For equation (31) no assumption has been used, so far, about the cause of the difference between $M^{*(0)}$ and $M^{*(1)}$, or for $\mathrm{d}M^*/\mathrm{d}s$. The line torque M^* may have an explicit dependence on the location $\underline{x}$, like in the case of

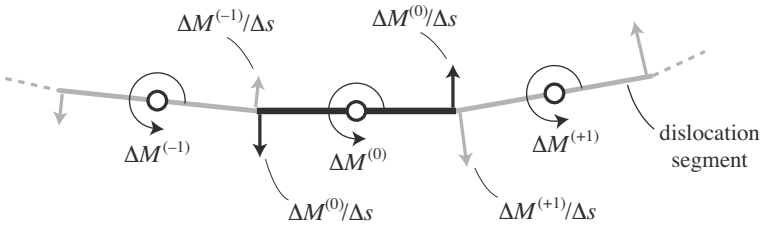


Figure 20. A dislocation cut into short segments of length Δs. When a line torque M^* with a gradient acts on the dislocation, a non-vanishing force results in the segments' connections.

elastic interaction between a dislocation and a size mismatched particle with a fixed location $\underline{x}_\mathrm{p}$ (Mohles, 1997). Another possibility is the dependence of M^* on α:

$$\frac{\mathrm{d}M^*}{\mathrm{d}s} = \frac{\partial M^*}{\partial s} + \frac{\partial M^*}{\partial \alpha}\frac{\mathrm{d}\alpha}{\mathrm{d}s} \tag{32}$$

When the dislocation has the curvature $1/R_\mathrm{c}$, then going from s to $s + \Delta s$ along the dislocation rotates the local line vector direction $\underline{s}^0$ by the angle $\Delta s/R_\mathrm{c}$, and α changes by the same amount. In figure 20 a slight curvature is indicated, but R_c and $\Delta\alpha$ are not indicated because $R_\mathrm{c} \gg \Delta s$. With $\Delta\alpha = \Delta s/R_\mathrm{c}$, M^* from equation (30), and $\partial M^*/\partial s = 0$ (this is not used here) we get for τ_M:

$$\tau_{\mathrm{self}-M} = +\frac{1}{b}\frac{\partial^2 E^*(\alpha)}{\partial^2 \alpha}\frac{1}{R_\mathrm{c}} \tag{33}$$

This ist the stress that a dislocation poses upon itself only because its line energy $E^*(\alpha)$ is character dependent, and because this results in a torque. Usually, $\tau_{\mathrm{self}-M}$ has the same order of magnitude as $\tau_{\mathrm{self}-E}$. But unlike $E^*(\alpha)$ itself, its derivative $\partial^2 E^*/\partial\alpha^2$ can be negative.

Line Tension Both the stresses $\tau_{\mathrm{self}-E}$ and $\tau_{\mathrm{self}-M}$ arise from the line energy E^*. But they consider different, independent forces: $\tau_{\mathrm{self}-E}$ is derived from forces along the line vector $\underline{s}$, $\tau_{\mathrm{self}-M}$ from those normal to $\underline{s}$. The total self interaction stress τ_line — in the line tension model — is, therefore, superimposed of both $\tau_{\mathrm{self}-E}$ and $\tau_{\mathrm{self}-M}$:

$$\tau_\mathrm{self} = \frac{S(\alpha)}{b\,R_\mathrm{c}} \quad \text{with} \quad S(\alpha) = E^*(\alpha) + \frac{\partial^2 E^*(\alpha)}{\partial^2 \alpha} \tag{34}$$

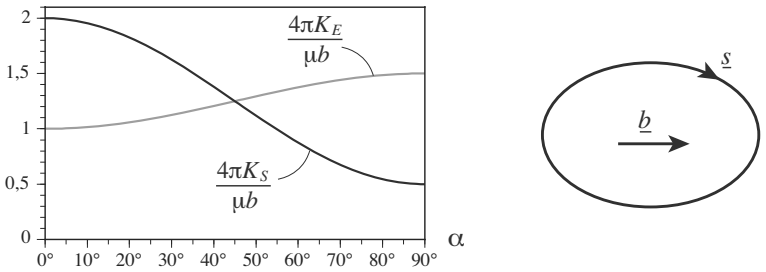


Figure 21. Pre-logarithmic factors K_E and K_S as functions of the dislocation character for $\nu = 1/3$, and the corresponding equilibrium shape of a dislocation loop.

$S(\alpha)$ is called the line tension and is actually a force, like $E^*(\alpha)$. The relation (34) between S and E^* holds in general, including elastically anisotropic materials. S is the true rigidity in the line tension model, not E^*. Consider for instance the pre-logarithmic factor $K_E(\alpha)$ of E^* for an isotropic material as given in equation (24), and its counterpart $K_S(\alpha)$ for the line tension S:

$$K_E = \frac{\mu b^2}{4\pi}\left(\cos^2\alpha + \frac{\sin^2\alpha}{1-\nu}\right) \tag{35a}$$

and

$$K_S = \frac{\mu b^2}{4\pi}\left(\frac{1+\nu}{1-\nu}\cos^2\alpha + \frac{1-2\nu}{1-\nu}\sin^2\alpha\right) \tag{35b}$$

As can be seen in figure 21, K_E and hence E^* are larger for edge than for screw dislocations, but for K_S and S, the opposite holds. This means that screw dislocations are stiffer than edge ones: when a screw dislocation is bent, edge components are generated, which needs much energy. On the other hand, when an edge dislocation is bent, only the lower energy screw components are generated. Accordingly, an equilibrium dislocation loop with a constant self-stress τ_{line} is elongated in the direction of the Burgers vector.

It is frequently claimed or assumed that S were the actual force along the dislocation line, but this perception is misleading and involves inconsistencies, as demonstrated below. The line tension model may be seen as inconsistent, anyway, in light of the more accurate self interaction model (section 2.2). But since the line tension is used for interpretations of simulations using the dislocation self-interaction, inconsistencies in the line tension model may lead to misinterpretations.

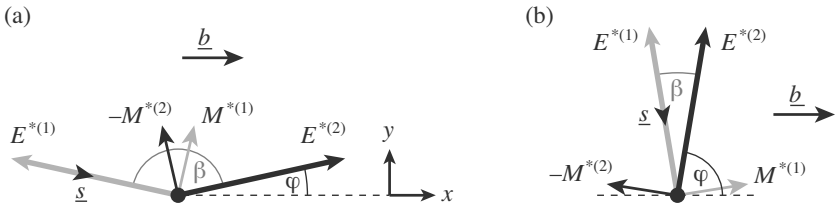


Figure 22. Forces from a dislocation on a point obstacle.

Forces on Obstacles In figure 22 all forces from a dislocation exerted on a point obstacle are plotted for two cases: large and small cusp angles β. The forces are E^* along the dislocation and M^* normal to it. The opposite forces of the line torque M^* are not plotted; they are distributed along the dislocation line in $b\tau_{\text{self}-M}$ (see above). Equation (36) describes this sum of forces in terms of $\varphi = (\pi - \beta)/2$.

$$F_y = \sin\,\varphi \left(E^{*(1)} + E^{*(2)}\right) + \cos\,\varphi \left(M^{*(1)} - M^{*(2)}\right) \tag{36}$$

Note that $M^{*(2)}$ in equation (36) and in figure 22 has a negative sign; this is because this force acts on the left side of a dislocation segment, and not on the right as $M^{*(1)}$. This is consistent with equation (31). Yet in figure 22, both force vectors $M^{*(1)}$ and $-M^{*(2)}$ point in positive y-direction against the point obstacle. This is caused by the fact that $M^*(\alpha)$ changes its sign as the angle α between Burgers and line vector changes from $-\varphi$ to $+\varphi$ in the obstacle; the sign of M^* is always so that M^* tries to rotate the dislocation into screw character, which has the lowest energy E^*.

The only difference between figures 22(a) and 22(b) lies in the line vector $\underline{s}$, such that $\varphi \to 0$ in figure 22(a) as for weak obstacles, and $\varphi \to \pi/2$ in 22(b) as for strong obstacles. In the latter case, the force contributions $M^{*(1)}$ and $-M^{*(2)}$ cancel each other (figure 22(b)). Hence the total force F_y pulling on the obstacle equals $2E^*$: in case of strong obstacles, like for Orowan's strengthening model (see above), the line energy is the appropriate force.

In contrast, for weak obstacles $M^{*(1)}$ and $-M^{*(2)}$ add up, and they add to the force $2E^*\sin\varphi$. With $\varphi \to 0$, $\cos\varphi = 1$, and $E^{*(1)} = E^{*(2)} = E^*$, equation (36) can be written as

$$F_y = 2E^*\,\sin\,\varphi - \frac{\Delta M^*}{\Delta\alpha}\Delta\alpha \tag{37}$$

Here $-\Delta M^*$ replaces $M^{*(1)} - M^{*(2)}$, and $\Delta\alpha$ is just inserted without damage. Then $(\Delta M^*/\Delta\alpha)$ can be read as the first derivative of $M^*(\alpha)$. While crossing the particle, the line vector $\underline{s}$ changes by $\Delta\alpha = 2\varphi$ (figure 22(a)). Now inserting this and replacing φ by $\sin\varphi$ ($\varphi \to 0$), and using $M^* = -\partial E^*/\partial\alpha$ (equation (30)), we get:

$$F_y = 2 \sin \varphi \left(E^* + \frac{\partial^2 E^*}{\partial \alpha^2} \right) \tag{38}$$

The bracket in equation (38) is obviously identical with the line tension S as introduced in equation (34). Hence equation (38) can be read as if in figure 22(a), two forces $S^{(1)}$ and $S^{(2)}$ acted on the obstacle instead of $E^{*(1)}$ and $E^{*(2)}$, and no forces M^*. Hence, for weak obstacles, the line tension S may be considered as the appropriate force acting on an obstacle. But this perception is not consistent because for strong obstacles, E^* is the appropriate force (see above). This means that at some point in the transition from weak to strong obstacles, a 'switch' from S to E^* would be needed if the line torque were to be disregarded. In case of obstacles of medium strength, around $\varphi = \pi/4$, either force S or E^* would only be an approximation. It seems more useful to keep a fully consistent view in mind, in which the true forces, namely the line energy E^* and the line torque M^*, are considered as in equation (36). With these, even the force components F_x, which act on the obstacle in figure 22 from the sides, can be treated correctly with an equation equivalent to (36). Such side forces F_x may be larger than F_y and hence be critical for overcoming an obstacle.

While the line tension model as described here is consistent and exact for a linear object with a line energy E^*, it is still only an approximation for dislocations because they have a long-ranged self-interaction.

2.2 Dislocation Self-Interaction

Each infinitesimal piece of a dislocation somehow interacts elastically with every other piece. This can be stated, as seen later, in spite of the fact that a single piece of dislocation does not exist. In principle the interaction goes through every point in space; however, when an elastically homogeneous material is assumed, the latter complication is dropped. Since elasticity is linear, the stress and strain contributions caused by all dislocation pieces can be calculated separately for an arbitrary point in space and then superimposed later. This is Green's principle, which is usually well known and used in electrodynamics or heat conduction. In case of elasticity, a Green's displacement tensor is required instead of a Green's function.

Green's displacement tensor The general elastic stress equilibrium can be written as $\underline{\nabla} \cdot \boldsymbol{\sigma} + \underline{f} = 0$, where $\boldsymbol{\sigma}$ is a general stress tensor, $\underline{f}$ is a volume force (force per unit volume) and $\underline{\nabla}$ denotes the gradient operator. With the definition of a strain $\boldsymbol{\sigma}$ in terms of a displacement $\underline{u}$, $\varepsilon_{kl} = \left(\partial u_k/\partial x_l + \partial u_l/\partial x_k\right)/2$, and Hooke's law $\sigma_{ij} = C_{ijkl}\varepsilon_{kl}$, the stress equilibrium can be written as a relation between between a volume force $\underline{f}$ and the corresponding displacements $\underline{u}$:

$$C_{ijkl}\frac{\partial^2 u_k}{\partial x_j \partial x_l} + f_i = 0 \tag{39}$$

Here C_{ijkl} denotes the 4th grade tensor of elastic constants, and Einstein's summation convention over equal indices is applied. The subsequent derivation is restricted to an elastically isotropic medium. For this, the tensor of elastic constants can be written in terms of the shear modulus μ and the Poisson ratio ν: $C_{ijkl} = \left[\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk} + (2\nu/(1-2\nu))\,\delta_{ij}\delta_{kl}\right]\mu$, where δ_{ij} is the Kronecker symbol with $\delta_{ij} = 1$ if $i = j$, and $\delta_{ij} = 0$ if $i \neq j$. Using this in equation (39) and considering a volume force $\underline{f} = (\delta(\underline{r}), 0, 0)$ pointing only to the x_1-direction, we can write:

$$\begin{aligned}
\frac{\mu}{1-2\nu}\frac{\partial}{\partial x_1}(\underline{\nabla}\cdot\underline{u}) + \mu\,\underline{\nabla}^2 u_1 + \delta(\underline{r}) &= 0 \\
\frac{\mu}{1-2\nu}\frac{\partial}{\partial x_2}(\underline{\nabla}\cdot\underline{u}) + \mu\,\underline{\nabla}^2 u_2 &= 0 \\
\frac{\mu}{1-2\nu}\frac{\partial}{\partial x_3}(\underline{\nabla}\cdot\underline{u}) + \mu\,\underline{\nabla}^2 u_3 &= 0
\end{aligned} \tag{40}$$

This is a set of coupled differential equations for the displacement $\underline{u}$ caused by the 'force' $\delta(\underline{r})$, which acts only in x_1-direction and only in the origin $r = 0$ due to the delta function $\delta(\underline{r})$. This function has the dimension m^{-3} instead of Nm^{-3} for a volume force because by definition, a Green's function is the reaction of a system to the inhomogeneity of the δ-function. With some help from potential theory, equation (41) can be solved analytically (e.g. (Hirth and Lothe, 1992)). The solution is $\underline{u}_1$, with the subindex 1 indicating the direction of the force used. Repeating this for forces in x_2- and x_3-directions, we find the matrix u_{ij}, where i is the index for the displacement component and j that for the force component. This is already Greens displacement tensor. Because of the choice of the δ-function as the inhomogeneity in equation (40), the dimension of u_{ij} is not that of a displacement (m) but mN^{-1}. Therefore, we rename the u_{ij} into the usual name for a Greens function or tensor, G_{ij}.

$$G_{ij} = \frac{1}{8\pi\mu}\left(\delta_{ij}\underline{\nabla}^2 r - \frac{1}{2(1-\nu)}\frac{\partial^2 r}{\partial x_i \partial x_j}\right) \tag{41}$$

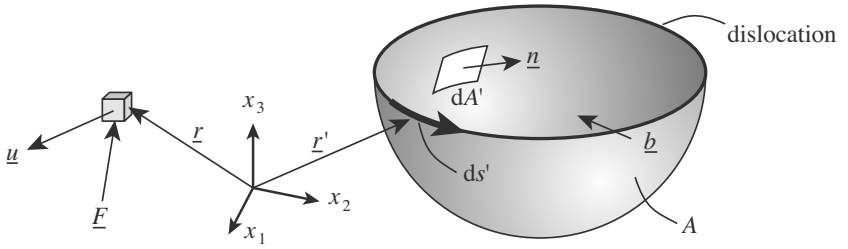


Figure 23. A dislocation is generated by a cutting procedure along the area A in the presence of a test force $\underline{F}$ in the point $\underline{r}$ to derive the dislocation's displacements there.

As is the idea for any Green's function, the displacement caused by a given distribution of volume forces $\underline{f}(\underline{r})$, can then be found by integration over the volume:

$$u_i(\underline{r}) = \int G_{ij}\,(\underline{r} - \underline{r}')\,f_j\,(\underline{r}')\;\mathrm{d}V' \tag{42}$$

Dislocation stress due to Peach and Koehler To derive the stress caused by an arbitrarily curved dislocation, the energy required to generate this dislocation by the cutting procedure described in section 1.1 is considered in the presence of a test force $\underline{F}$ acting in an arbitrary location $\underline{r}$. The cutting area A can be arbitrarily curved, as depicted in figure 23. The test force contributes the energy

$$W = \underline{F} \cdot \underline{u}(\underline{r}) = F_m u_m(\underline{r}) \tag{43}$$

to the dislocation's creation, where $\underline{u}$ is the displacement caused by the dislocation. On the other hand, this energy can also be expressed by the stress $\boldsymbol{\sigma}^F$ caused by $\underline{F}$ and the displacement in the area A during the dislocation's generation. In A, the displacements are obviously given by the Burgers vector b.

$$W = -\int_A b_i\;\sigma_{ij}^F\,(\underline{r}' - \underline{r})\;\mathrm{d}A'_j \tag{44}$$

The sign in equation (44) is negative because F helps in the dislocation creation. With Hooke's law, $\sigma_{ij} = C_{ijkl}\varepsilon_{kl}$, and the strain definition $\varepsilon_{kl} = (\partial u_k/\partial x_l + \partial u_l/\partial x_k)/2$, the stress $\boldsymbol{\sigma}^F$ can be expressed by the displacement $\underline{u}^F$ that F causes. Due to the symmetry of C_{ijkl} ($C_{ijkl} = C_{ijlk}$), $\boldsymbol{\sigma}$ is simplified to $\sigma_{ij} = C_{ijkl}\partial u_k/\partial x_l$. Since $\underline{F}$ acts in the point $\underline{r}$, the displacements $\underline{u}^F$ are known from Green's displacement tensor: $\underline{u}^F = \underline{F} \cdot \boldsymbol{G}$.

Hence equation (44) can be rewritten as

$$W = -F_m \int_A b_i \, C_{ijkl} \frac{\partial G_{km}\left(\underline{r}' - \underline{r}\right)}{\partial x_l'} \, \mathrm{d}A_j' \tag{45}$$

Now by comparing equation (45) with (43), the displacement components u_m caused by the dislocation in the point $\underline{r}$ is found. Subsequently, $\underline{r}$ is chosen as the origin. To yield the strain tensor, $\underline{u}$ is differentiated with respect to a general direction x_s:

$$\frac{\partial u_m}{\partial x_s} = \int_A b_i \left(C_{ijkl} \frac{\partial^2 G_{km}\left(\underline{r}'\right)}{\partial x_s' \partial x_l'} \right) \mathrm{d}A_j' \tag{46}$$

Here it has been used that $\partial/\partial x_s = -\partial/\partial x_s'$. Multiplying equation (46) with C_{ijms} yields the stress tensor σ_{ij} of the dislocation, but this would be an expression of the cutting area A, which is unhandy. Moreover, A has no physical meaning; a line integral along the dislocation line is needed. In the point of strain observation (the origin), which is not on the dislocation itself, there are no volume forces. Therefore, by using the general stress equilibrium of equation (39) with $f_i = 0$, the bracket in equation (46) can be rewritten such that

$$\frac{\partial u_m}{\partial x_s} = b_i \, C_{ijkl} \int_A \left(\frac{\partial^2 G_{km}\left(\underline{r}'\right)}{\partial x_s' \partial x_l'} \mathrm{d}A_j' - \frac{\partial^2 G_{km}\left(\underline{r}'\right)}{\partial x_j' \partial x_l'} \mathrm{d}A_s' \right) \tag{47}$$

This expression can be subjected to Stoke's theorem to be converted into a line integral along the dislocation:

$$\frac{\partial u_m}{\partial x_s} = \epsilon_{jsn} \, b_i \, C_{ijkl} \oint \frac{\partial G_{km}\left(\underline{r}'\right)}{\partial x_l'} \mathrm{d}x_n' \tag{48}$$

Here ϵ_{ijk} is the Einstein permutation operator with $\epsilon_{ijk} = 0$ except for $\epsilon_{123} = \epsilon_{231} = \epsilon_{312} = 1$ and $\epsilon_{321} = \epsilon_{213} = \epsilon_{132} = -1$. Equation (48) is generally valid for anisotropic materials. However, here Green' displacement tensor G_{ij} for isotropic materials of equation (41) and the isotropic constants $C_{ijkl} = \left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk} + (2\nu/(1-2\nu))\delta_{ij}\delta_{kl}\right)\mu$ are inserted, and Hooke's law is applied to finally get the stress $\boldsymbol{\sigma}$ of an arbitrary dislocation loop:

$$\begin{aligned} \sigma_{\mathrm{disl},\alpha\beta} = &-\frac{\mu}{8\pi} \oint_C b_m \frac{\partial}{\partial x_i'} \underline{\nabla}'^2 r' \left(\epsilon_{im\alpha} \mathrm{d}x_\beta' + \epsilon_{im\beta} \mathrm{d}x_\alpha' \right) - \\ &- \frac{\mu}{4\pi(1-\nu)} \oint_C b_m \, \epsilon_{imk} \frac{\partial}{\partial x_i'} \left(\frac{\partial^2 r'}{\partial x_\alpha' \partial x_\beta'} - \delta_{\alpha\beta} \underline{\nabla}'^2 r' \right) \mathrm{d}x_k' \end{aligned} \tag{49}$$

where $r' = \left(x_1'^2 + x_2'^2 + x_3'^2\right)^{1/2}$ is the length of the vector from the origin to a point on the dislocation curve C. Equation (49) is a useful result, as is seen later on, and it seems as if the integrand defined the stress contribution of a dislocation segment $\underline{\mathrm{d}x}$. But the integrand is not unequivocal; one alternative is Brown' stress formula (Brown, 1964), (Brown, 1967), (Hirth and Lothe, 1992).

Dislocation stress due to Brown For the derivation of Brown's stress formula of a curved dislocation we start with equation (46), and the point of stress observation in the origin, $\underline{r} = 0$. When we insert Green's tensor G_{km} of equation (41) and remove all primes for simplicity, the result can be written as

$$\frac{\partial u_m}{\partial x_s} = \int_A \frac{a_{ms}\left(x_1/r, x_2/r\right)}{r^3} \mathrm{d}x_1 \mathrm{d}x_2 \tag{50}$$

where the term a_{ms} is just used as an abbreviation; it has been formulated such that it depends on the dimensionless terms x_i/r. The term a_{ms}/r^3 satisfies Euler's identity of -3rd degree, which means that:

$$\frac{a_{ms}}{r^3} = -\frac{\partial}{\partial x_1}\left(\frac{x_1}{r^3} a_{ms}\right) - \frac{\partial}{\partial x_2}\left(\frac{x_2}{r^3} a_{ms}\right) \tag{51}$$

This can be verified with a_{ms} defined by the comparison of equation (50) with the combination of equations (46) and (41). Subsequently, this particular function a_{ms} is not used. Instead, it is only assumed that $\partial u_m/\partial x_s$ can be expressed as in equation (50) with a function a_{ms} that is yet unknown, but that satisfies Euler's identity of equation (51). The advantage of this procedure is that the assumption of elastic isotropy used for G_{km} in equation (41) is dropped. This allows for the consideration of elastically anisotropic materials. Using equation (51) and Green's theorem, equation (50) can be rewritten as an integral along the dislocation line C:

$$\frac{\partial u_m}{\partial x_s} = \oint_C \frac{1}{r^2}\left(\frac{x_1}{r}\frac{\mathrm{d}x_2}{\mathrm{d}s} - \frac{x_2}{r}\frac{\mathrm{d}x_1}{\mathrm{d}s}\right) a_{ms}\left(x_1/r, x_2/r\right) \mathrm{d}s \tag{52}$$

The terms x_1/r, x_2/r, $\mathrm{d}x_1/\mathrm{d}s$, and $\mathrm{d}x_2/\mathrm{d}s$ can be interpreted as $\cos(\theta)$, $\sin(\theta)$, $\cos(\varphi)$ and $\sin(\varphi)$, respectively, with the meanings of the angles φ and θ given in figure 24(a). Using these angles, equation (52) can be rewritten as:

$$\frac{\partial u_m}{\partial x_s} = \oint_C \frac{1}{r}\frac{\sin(\theta-\varphi)}{r}\, a_{ms}(\theta)\mathrm{d}s \tag{53}$$

As mentioned above, the function $a_{ms}(\varphi)$ is still considered as unknown. Subsequently, the known stress of a straight dislocation (section 1.2) is used

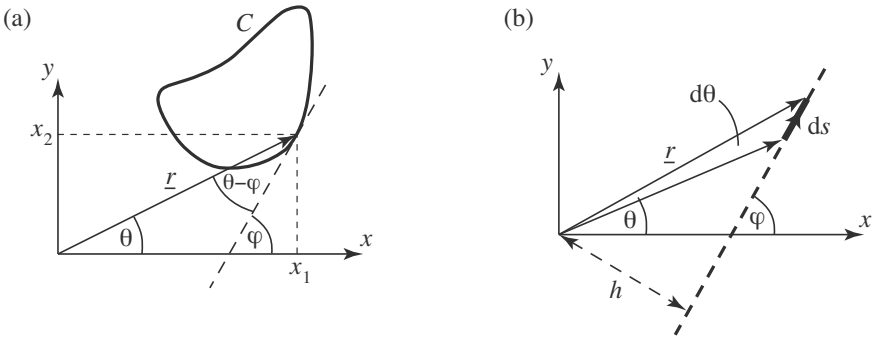


Figure 24. Geometrical relations: (a) angles, (b) tangential segment ds.

to find $a_{ms}(\theta)$. When we consider a straight dislocation, as in figure 24(b), the distance $h = r \cdot \sin(\theta - \varphi)$ is constant, and $1/h$ can be taken out of the integral. Further using the relation $\mathrm{d}s = r/\sin(\theta - \varphi)) \cdot \mathrm{d}\theta$ we can write:

$$\frac{\partial u_m}{\partial x_s} = \frac{A_{ms}(\varphi)}{h} = \frac{1}{h}\int_{\varphi}^{\varphi+\pi} \sin(\theta - \varphi)\, a_{ms}(\theta)\mathrm{d}s \tag{54}$$

The defined integral $A_{ms}(\varphi)$ is still unknown; however, its second derivative with respect to φ can be written as:

$$\frac{\partial^2 A_{ms}}{\partial \varphi^2} = 2a_{ms}(\varphi) - \int_{\varphi-\pi}^{\varphi} \sin(\theta - \varphi)\, a_{ms}(\theta)\mathrm{d}s \tag{55}$$

Here Leibniz' formula has been used because the limits of integration depend on φ. Combining equations (54) and (55) yields a simple relation between a_{ms} and A_{ms}:

$$a_{ms}(\varphi) = \frac{1}{2}\left(A_{ms} + \frac{\partial^2 A_{ms}}{\partial \varphi^2}\right) \tag{56}$$

This relation can be inserted for in equation (53), so that the strain $\varepsilon_{ms} = \partial u_m/\partial x_s$ of a generally curved dislocation is expressed by the still unknown expression $A(\theta) + \partial^2 A/\partial\theta^2$ of piecewise straight segments ds. Applying Hooke's law leads to

$$\sigma_{ij} = \oint_C \frac{\sin(\theta - \varphi)}{r^2}\left[B_{ij}(\theta) + \frac{\partial^2 B_{ij}}{\varphi^2 \theta}\right]\mathrm{d}s \tag{57}$$

with the tensor $B_{ij} = C_{ijms}A_{ms}/2$. Finally, considering the shear stresses in a glide system with Burgers vector $\underline{b}$ and glide plane $\underline{n}$ by applying

the Peach-Koehler-formula (41) and comparing these stresses with those of straight dislocations given in section 1.2, the tensor B_{ij} is identified, and the shear stress of the curved dislocation is found to be

$$\tau_{\mathrm{disl}} = \underline{b}^0 \, \boldsymbol{\sigma} \, \underline{n}^0 = \oint_C \frac{\sin(\theta - \varphi)}{r^2} K_S(\theta) \mathrm{d}s \tag{58}$$

where $K_S(\theta)$ is the pre-logarithmic factor for the line tension of a straight dislocation: $K_S = K_E + \partial^2 K_E / \partial \theta^2$. For the case of elastic isotropy, the function $K_S(\theta)$ is given by equation (35b), but equation (58) is also valid in case of anisotropy: isotropy has not been assumed, not even by using the isotropic Green's tensor G_{km} of equation (41).

Dislocation segments To utilize a stress formula like (49) or (58) in computer simulations of dislocation motion, a dislocation must be cut into segments, like in case of a polygon. Hence the line integrals are decomposed into contributions from segments k connecting the points A and B at $\underline{r}^A$ and $\underline{r}^B$:

$$\sigma_{ij} = \sum_k \sigma_{ij}^{\mathrm{seg}} \left(\underline{r}_k^A, \underline{r}_k^B \right) \quad \text{with}$$

$$\sigma_{ij}^{\mathrm{seg}} = \int_A^B I_{ij} \mathrm{d}s = \sigma_{ij}^{\mathrm{semi}} \left(\underline{r}^B, \underline{s}^0 \right) - \sigma_{ij}^{\mathrm{semi}} \left(\underline{r}^A, \underline{s}^0 \right) \tag{59}$$

where I_{ij} is the integrand of (49) or (58). The integral is expressed as the contributions from the two integration limits; these may be seen as the stress contributions $\boldsymbol{\sigma}^{\mathrm{semi}}$ from semi-segments. Besides on $\underline{r}$, $\boldsymbol{\sigma}^{\mathrm{semi}}$ depends on the direction $\underline{s}^0$. When equation (49) is used for integration this dependence results from the direction of $\underline{\mathrm{d}x}'$, in case of equation (58) from the angle φ of the tangent. For $\boldsymbol{\sigma}^{\mathrm{semi}}$ of a straight segment with fixed line vector $\underline{s}^0$, Devincre (Devincre, 1995) derived:

$$\sigma_{ij}^{\mathrm{semi}} =$$

$$\frac{\mu}{2\pi Y^2} \left[(\underline{b} \times \underline{Y})_i \left(\underline{s}^0 \right)_j + (\underline{b} \times \underline{Y})_j \left(\underline{s}^0 \right)_i - \frac{\left(\underline{b} \times \underline{s}^0 \right)_i Y_j + \left(\underline{b} \times \underline{s}^0 \right)_j Y_i}{1 - \nu} \right]$$

$$- \frac{\mu}{2\pi Y^2} \frac{\underline{b} \times \underline{Y} \cdot \underline{s}^0}{1 - \nu} \left[\delta_{ij} + \left(\underline{s}^0 \right)_i \left(\underline{s}^0 \right)_j - \frac{2}{Y^2} \left(h_i Y_j + h_j Y_i + Y_i Y_j L / R \right) \right] \tag{60}$$

with the definitions $L = \underline{r} \cdot \underline{s}^0$, $\underline{h} = \underline{r} - L\underline{s}^0$, and $\underline{Y} = (r - L)\underline{s}^0 - \underline{h}$, and the point of stress observation being the origin. Some of these definitions are sketched in figure 25(a). The summation of stress contributions of semi-segments in equation (59) is equivalent to considering single segments with

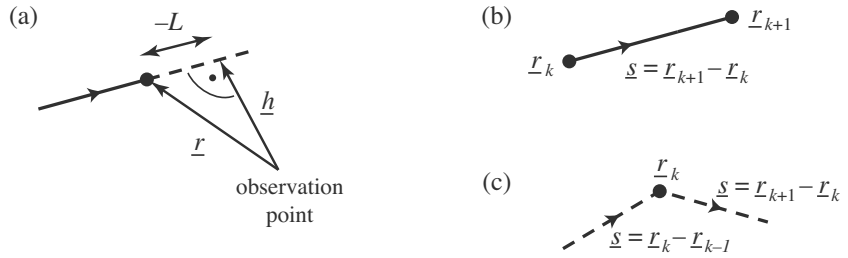


Figure 25. (a) Geometrical meanings for equation (60); (b) segment summation; (c) angular summation.

two different ending points $\underline{r}$ and the same line vector $\underline{s}^0$, as indicated in figure 25(b). This type of summation may be inconvenient, depending on the type of segmentation in a simulation code. An alternative approach is to evaluate equation (60) for the connection points $\underline{r}$ between segments with different line vectors $\underline{s}^0$, as indicated in figure 25(c).

For 2-dimensional simulations using only one glide plane $\underline{n}^0$, it is useful to apply a specialized and, therefore, simpler expression. With $\underline{n}^0 = (0, 0, 1)$ and the Peach-Koehler equation (16) applied to equation (49), the resolved stress of a segment from $\underline{r}_A$ to $\underline{r}_B$ is found as (Mohles, 2001c):

$$\tau_{\mathrm{seg}} = \frac{\mu}{4\pi}\frac{1}{D}\left[(\underline{b}_{\mathrm{s}} \cdot \underline{k})\,(\underline{b}_{\mathrm{r}}^0 \cdot \underline{s}) + \frac{(\underline{b}_{\mathrm{s}} \cdot \underline{k}_\perp)\,(\underline{b}_{\mathrm{r}}^0 \cdot \underline{s}_\perp)}{1-\nu}\right] \tag{61}$$

with $D = (\underline{r}_A \times \underline{r}_B)_z$, $\underline{s} = (s_x, s_y, 0) = \underline{r}_B - \underline{r}_A$, $\underline{k} = (k_x, k_y, 0) = \underline{r}_B^0 - \underline{r}_A^0$, $\underline{s}_\perp = (-s_y, s_x, 0)$, $\underline{k}_\perp = (-k_y, k_x, 0)$, and $\underline{b}_{\mathrm{s}}$ and $\underline{b}_{\mathrm{r}}$ denoting the Burgers vectors of the stress source and the stress receiving dislocation, respectively.

Self interaction Equations (49), (58), (60), and (61) describe the stresses a dislocation imposes on another one. In principle, the stress receiver can be the same as the source. But on the dislocation line, elasticity theory predicts a stress singularity as can already be seen from straight dislocations (equation (4) or (7)). This is an artifact of the dislocation being assumed to exist in a continuous medium, which is a conflict. In real materials this problem does not exist due to the atomistic nature of a real crystal. In applications of the quoted stress formulae, a bit of material around the point of stress observation must be cut out, in analogy to the calculation of the energy of straight dislocations.

One option for this is Brown's approach (Brown, 1964), (Brown, 1967)

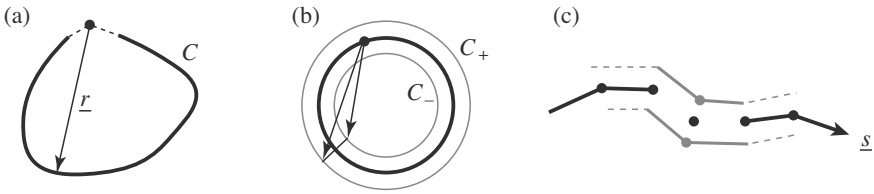


Figure 26. Cut-out procedures for the dislocation core introduced by (a) Bacon (Bacon, 1967); (b), (c) Brown (Brown, 1964), (Brown, 1967).

sketched in figure 26: instead of evaluating a stress formula on the dislocation line C itself (figure 26(a)), two lines C_+ and C_- parallel to C at given distances of $\pm r_{\mathrm{cut}}$ are considered (figure 26(c)), and the mean value of the corresponding stresses is used. By this procedure the dislocation core with radius r_{cut} is left out of the calculation. At large distances $|\underline{r}|$ this calculation yields the same stress contribution as using the original line C because the vectors to C_+ and C_- are very similar to $\underline{r}$ (figure 26(b)). But in the vincinity of the node (figure 26(c)), the vectors to C_+ and C_- have essentially opposite directions, and the respective stress contributions cancel each other. The additional computational effort of two, instead of one, stress calculations can be mostly avoided by using C_+ and C_- only in the vincinity of each node.

Another option to cut out the dislocation core is to disregard the stress contributions of the dislocation segments connected to the observation point, as sketched in figure 26(a). However, this procedure must be used with caution for two reasons: Firstly, the stress of a dislocation segment of finite length is not unique. As has been shown (Mohles, 2001c), the stress $\underline{b}^0 \boldsymbol{\sigma}_{\mathrm{disl}} \underline{n}^0$ with $\boldsymbol{\sigma}_{\mathrm{disl}}$ of equation (49) equals τ_{disl} of equation (58) if the same closed dislocation loop is considered. But if a dislocation segment is left out of the integration, the stresses do not match. After all, such a segment of finite length does not exist anyway (section 1.1). But in the calculations, errors of unknown magnitude can arise from this stress equivocality for non-closed dislocations. Such errors can be avoided by a special segmentation procedure described in section 3.1.

The second implication of disregarding the neighbour segments of a node is that the cut-out length L_{cut} has a meaning equivalent to that of r_{cut}: it defines the energy associated with the dislocation core (see equation (6aa), $r_{\mathrm{cut}} \approx r_{\mathrm{core}}$) and hence influences a dislocation's flexibility, as can be seen from the line tension model. In simulations, the lengths of the neighbour segments of every node will vary; but the core energy and the corresponding

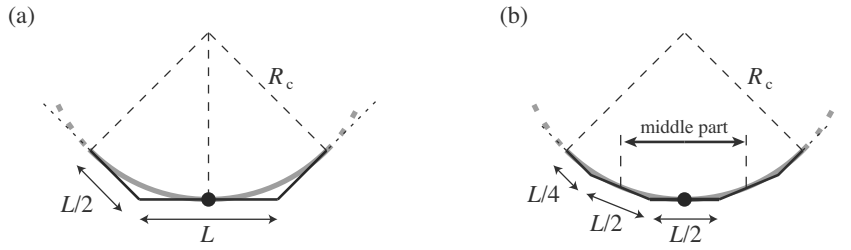


Figure 27. A curved dislocation segment (grey) of length $2L$ in different approximations (black) with (a) three and (b) five segments.

length L_{core} should be constant. In order to find a way to achieve this, we consider a curved dislocation segment (arc) of constant length $2L$ and curvature radius R_{c}. This arc is approximated by two different sets of straight segments, as sketched in figure 27. Subsequently, $L \ll R_{\mathrm{c}}$ is assumed; in figure 27 this condition is disregarded in order to improve the visibility of the angles. The dislocation arc is assumed to have a parabolic shape. With these assumptions and equation (61) the stress of the arc on its centre point can be calculated for the approximations with (a) three and (b) five segments. In both cases, the stress contribution of the segment in the centre vanishes for symmetry reasons. This centre segment may be seen as being cut out in the approximations. It gives no stress contribution, regardless of the stress formula used.

$$\tau_{(\mathrm{a})} = \frac{\mu b}{24 R_{\mathrm{c}}} \left(\frac{9 + 3\nu}{1 - \nu} \cos^2\alpha + \frac{9 - 12\nu}{1 - \nu} \sin^2\alpha \right) \tag{62}$$

$$\tau_{(\mathrm{b})} = \frac{\mu b}{24 R_{\mathrm{c}}} \left(\frac{13 + 7\nu}{1 - \nu} \cos^2\alpha + \frac{13 - 20\nu}{1 - \nu} \sin^2\alpha \right) \tag{63}$$

Here α is the angle between Burgers and line vector. Note that $\tau_{(\mathrm{a})}$ and $\tau_{(\mathrm{b})}$ depend on the curvature radius R_{c}, but they are independent of the length L. This means that when approximation (a) had been constructed with the same R_{c} but with an overall length of L instead of $2L$, the same stress $\tau_{(\mathrm{a})}$ would have been obtained. This is exactly the case for the middle part of the approximation in figure 27(b). Hence, the outer parts of (b) generate the stress $\tau_{(\mathrm{b})} - \tau_{(\mathrm{a})}$. This difference is the *additional* stress caused by the transition from approximation (a) to (b). In this transition, the length of the straight segment in the centre has been reduced by a *factor* two. Hence if the length L_{cut} is actually cut out of the calculations in a simulation, but L_{core} is the aspired cut-out length defined by the core energy, the *additional*

stress $\tau_{\rm arc}$ must be considered:

$$\tau_{arc} = \left(\tau_{(\rm b)} - \tau_{(\rm a)}\right)\ \log_2 \frac{L_{\rm cut}}{L_{\rm core}} = \left[\frac{2}{3\ln 2}\right] \frac{K_S(\alpha)}{bR_{\rm c}}\ \ln\ \frac{L_{\rm cut}}{L_{\rm core}} \qquad (64)$$

In the second part of equation (64), the pre-logarithmic stress factor $K_S(\alpha)$ of equation (35b) has been used to replace $\left(\tau_{(\rm b)} - \tau_{(\rm a)}\right)$. The similarity of equation (64) with the self-stress $\tau_{\rm self}$ in the line tension model (equation (34)) is not accidental: the line tension model is very local, just like the stress of a dislocation arc of finite length. But in applications of the line tension model, a more global, radial cutting procedure is used with the factor $\ln\left(r_{\rm shield}/r_{\rm core}\right)$ (equation (6b)) instead of the more local factor $\ln\left(L_{\rm cut}/L_{\rm core}\right)$.

The square bracket of equation (64) equals about 0.961. This resulted from the comparison of approximations (a) and (b). When a higher order approximation with halved segments and the cut-out length $L/4$ had been compared with approximation (b), the factor $[24/(35ln2)] \approx 0.989$ would have been found. Obviously the square bracket can be set to equal 1.0, like in the line tension model. Altogether the total self interaction of a dislocation in a point $\underline{r}_k$ including stress contributions from distant segments can be summarized by

$$\tau_{\rm self}\left(\underline{r}_k\right) = \sum_{i \neq k} \tau_{\rm seg}\left(\underline{r}_i - \underline{r}_k\right) + \frac{K_S\left[\alpha\left(\underline{r}_k\right)\right]}{bR_{\rm c}\left(\underline{r}_k\right)}\ \ln\ \frac{L_{\rm seg}\left(\underline{r}_k\right)}{L_{\rm core}} \qquad (65)$$

where $L_{\rm seg}\left(\underline{r}_k\right)$ is the local cut-out length and $L_{\rm core} \approx 2b$ is the length determining the core energy.

With equation (65) and a segment fomula like that of equation (60) or (61), the mathematics for dislocation dynamics simulations is complete. But the numerical effort for such simulations using the self interaction concept is very high because, in principle, the interaction of every dislocation segment with every other one must be calculated in every time step. This makes it inevitable to apply certain optimisations in a simulation code; some successful ones are described in section 4.2.

3 2-D Applications

Many plasticity related properties of materials depend on the flexibility of dislocations because the latter determines the actual number of obstacles a dislocation encounters simultaneously while moving. This can be seen from Friedel's basic strengthening model (section 2.1) for 2-dimensional dislocation glide, but it holds as well in three dimensions. The flexibility is mainly

determined by the dislocation's elastic self interaction (section 2.2) and approximately described by the line tension (section 2.1). In the present section, simulations of 2-D dislocation glide are described using both the line tension model as well as true self interaction.

3.1 Simulation Technique

The dislocation is described as a flexible line in the glide plane; the equilibrium of resolved stresses is considered along this line. Regardless of the dislocation model, the local equilibrium of resolved shear stresses τ along the dislocation line can be used as the basis for simulations:

$$\tau_{\text{ext}} + \tau_{\text{disl}} + \tau_{\text{allobst}} = 0 \tag{66}$$

Here τ_{ext} is an external stress driving the dislocation forwards; it can be derived from a global stress tensor via the Peach-Koehler formula (16). τ_{disl} denotes the dislocation's self interaction, either as given by τ_{self} of equation (65) or approximated by τ_{self} of equation (34). The obstacle stress τ_{allobst} can be defined by solute foreign atoms, particles of secondary phases, other dislocations, grain boundaries or any other kind of obstacles. The equilibrium (66) holds generally for static dislocation configurations like those considered in section 2.1; equation (66) has been used there implicitly.

If the equilibrium is violated the sum of resolved stresses renders a driving force on the line. For the simulations a non-static stress equilibrium is introduced by adding a viscous drag stress $\tau_{\text{drag}} = -(B/b)v_\perp$ to the left hand side of equation (66), where $v_\perp$ is the local velocity normal to the dislocation line and B is a drag coefficient. Inertial effects can be considered as well (see below) but are disregarded here so that the motion is assumed to be overdamped. The viscous drag term τ_{drag} may be seen to represent the phonon drag; Jassby and Vreeland (Jassby and Vreeland, 1973) have measured coefficients B of real specimens. However, the physical basis for τ_{drag} is irrelevant here because we are only looking for a static equilibrium here ($v_\perp \to 0$). When completed by τ_{drag}, equation (66) can be written as

$$v_\perp(\underline{x}) = (b/B)\,(\tau_{\text{ext}} + \tau_{\text{disl}} + \tau_{\text{allobst}}) \tag{67}$$

where $v_\perp$ defines the magnitude and the sign of the velocity, and it depends on the observation point $\underline{x}$ on the dislocation. The direction of the motion is defined by the unit vector $\underline{s}^0_\perp$ normal to the local line vector $\underline{s}$. Since only one glide plane is considered in 2-D, only planar vectors are noted subsequently. The directions $\underline{s}^0$ and $\underline{s}^0_\perp$ can be expressed by the derivatives $x' = \partial x/\partial s$ and $y' = \partial y/\partial s$: $\underline{s}^0 = (x', y')$ (see figure 28), where s denotes the arc length, and by definition $\underline{s}^0_\perp = (-y', x')$. The velocity vector $\mathrm{d}\underline{x}/\mathrm{d}t = (\mathrm{d}x/\mathrm{d}t, \mathrm{d}y/\mathrm{d}t)$ is

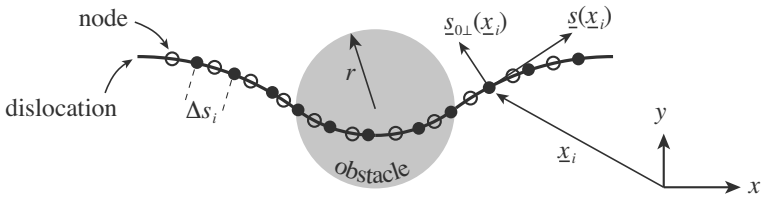


Figure 28. Local line vector $\underline{s}(x, y)$, and two possible representations of the same dislocation by different sets of nodes (empty and filled circles) of equal mean distance Δs (segmentation coarseness). Individual distances Δs vary. There are either 4 (empty) or 5 (filled) nodes inside the obstacle.

defined as $\mathrm{d}\underline{x}/\mathrm{d}t = \underline{s}^0_\perp v_\perp$ so that

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -\partial y/\partial s \\ \partial x/\partial s \end{pmatrix} \frac{b}{B}\left(\tau_{\mathrm{ext}} + \tau_{\mathrm{disl}} + \tau_{\mathrm{allobst}}\right) \tag{68}$$

τ_{disl} depends on the local curvature with the radius $R_{\mathrm{c}} = (y''x' - x''y')^{-1}$, where x'' and y'' denote the second derivatives $\partial^2 x/\partial s^2$ and $\partial^2 y/\partial s^2$, respectively. The dependence on R_{c} is obvious in the line tension model ($\tau_{\mathrm{disl}} = \tau_{\mathrm{line}}$, equation (35) but also holds for the self-interaction concept (equation (65)). In a simulation the derivatives x' and y', and R_{c} must be calculated for all points $\underline{x}_i$ and in each time step from the present configuration. This can be done for instance by constructing a circle through three neighbouring nodes $\underline{x}_{i-1}$, $\underline{x}_i$, and $\underline{x}_{i+1}$ and using the tangent in $\underline{x}_i$ for $\underline{s}_0$ and the radius for R_{c}. Altogether equation (68) is a partial differential equation which can be solved numerically, for instance using a Runge-Kutta type method (Press et al., 1992).

Dislocation discretisation For the simulations a discretisation is required. This means that a finite number of points $\underline{x}_i$ must be chosen to represent the dislocation line; equation (68) is then solved in these points. In figure 28, two different but equivalent discretisations of the same dislocation are plotted. Both representations have the same coarseness (average node distance), but either four or five nodes are located inside an obstacle. Individual distances Δs_i vary during a simulation run as the nodes move. But they have to be kept close to a mean value Δs by removing nodes or inserting new ones (by interpolation) during a simulation as required locally. This 'mean value' Δs may be forced to depend on the local curvature in order to keep the total number of nodes and hence the computational effort

low. However, the present author advises to use a constant value Δs within a simulation because a curvature dependent value Δs has sometimes been found to cause minor but systematic errors in simulation results.

For the subsequent considerations all individual distances Δs_i are assumed to equal Δs. In the example of figure 28, five of the filled nodes, but only four of the empty ones are inside the obstacle, where the dislocation is supposed to experience a constant obstacle stress $\tau_{\mathrm{obst}} = -\gamma/b$ (equation (19)) as in the case of a long-range ordered particle as an obstacle. This means that in y-direction (forward direction), the simulated dislocation in its critical configuration senses either the force $4\gamma\Delta s$ or $5\gamma\Delta s$, depending on which set of nodes happens to represent the dislocation. Hence, in principle, the simulated dislocation over- or underestimates the true maximum obstacle force $F_{\max} = 2r\gamma$ by the amount $\gamma\Delta s/2$, which causes a relative error of $\pm\Delta s/(4r)$. For the case of figure 28 this would mean an error of more than $\pm 10\%$ ($\pm 0.5\gamma\Delta s$ of $\approx 4.5\gamma\Delta s$). This statistical error can easily be reduced by choosing a higher node density Δs^{-1}. But this must be avoided because the calculation effort in the computer is roughly proportional to about Δs^{-3}: the number of nodes is proportional to Δs^{-1}. Likewise, in the self-interaction concept, the number of interaction partners for each node is proportional to Δs^{-1}. Moreover, the number of integration steps the integrator (any one) must take for a simulation run increases as Δs decreases. Therefore, in order to keep the calculation times in the computer low, the obstacle stress τ_{allobst} should be smoothed over the length Δs if it contains a discontinuity, like on the surface of the obstacle in figure 28. In that case $\Delta s = r/3$ can be chosen as the mean node distance for good simulation results, where r is the obstacle radius. However, this is only an upper limit. If two dislocations are simulated which get very close, like Shockley partial dislocations (section 1.4), Δs must not be larger than the distance between these dislocations (equation (23)). Otherwise, when Δs is too large, numerical instabilities can occur.

A lower limit for Δs is given by the length L_{core} (section 2.2), which is basically equivalent to the inner cut-off radius r_{core} in the line tension model. With a chosen length $r_{\mathrm{core}} = 2b$ it has been found that the simulations are only stable if $\Delta s > b/2$. For the upper limit $\Delta s = r/3$ quoted above this means that only obstacles with $r > 1.5b$ can be treated, for instance, particles with a diameter larger than 1 nm. Obviously this restriction arises from the linear elastic continuum model itself, which fails to cover atomistic properties well. While Δs affects the accuracy of the simulation results and the calculation speed, it does not affect the simulation results systematically as long as it is chosen constant.

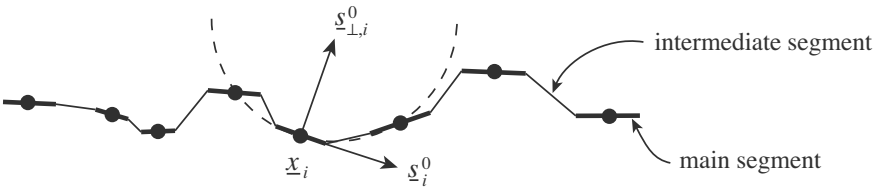


Figure 29. Segmentation procedure with main and intermediate segments used to ensure that the nodes are in the centres of straight (main) segments.

Dislocation segmentation When the dislocation self interaction model is used for τ_{disl}, a segmentation of the dislocation must be defined in addition to the discretisation in order to utilize a segment stress formula like equation (61). A frequently applied segmentation method is to use the node connecting vectors $\underline{x}_{i+1} - \underline{x}_i$ as segments. However, for the dislocation self interaction, a piece of dislocation around the point of stress observation must be cut out because of the stress singularity on the dislocation line. This may cause errors of unknown magnitude because the stress of a non-closed dislocation is undefined, as outlined in section 2.2.

However, the subsequent segmentation procedure fully avoids such errors becauses it enforces that every node is placed in the centre of a segment, where the stress of this segment vanishes unequivocally. This is accomplished in three steps. At first in every time step, a circle (dashed in figure 29) through the points $\underline{x}_{i-1}$, $\underline{x}_i$, and $\underline{x}_{i+1}$ is constructed for each node i. The local line direction $\underline{s}^0_i = \underline{s}^0(\underline{x}_i)$ and the local curvature radius $R_{\mathrm{c},i} = R_{\mathrm{c}}(\underline{x}_i)$ are derived from this circle. Then a straight dislocation segment with the direction $\underline{s}^0_i$ is assigned to node i so that this node lies in the centre of the segment. This is called a main segment. In figure 29 the main segments are drawn as thick lines. The length of main segments is chosen to be $\Delta s_i/2$, where $\Delta s_i = 1/2\,|\underline{x}_i - \underline{x}_{i-1}| + 1/2\,|x_{i+1} - \underline{x}_i|$ is the mean distance to the adjacent nodes. Finally, the end points of the main segments are connected by intermediate segments so that a closed polygon results. In figure 29 the intermediate segments are drawn as thin lines for distinction; they have no nodes in their centres. With this segmentation there are twice as many segments as nodes; to each node belong one main segment and two half intermediate segments on the left and the right. This segmentation may seem quite laborious, but possible errors arising from the stress ambiguity of segment stresses are avoided. Moreover, numerical instabilities have not been encountered with the present segmentation, even in cases which are prone to such instabilities (Duesbery et al., 1992). For the cut-out length

L_{cut} in the arc stress τ_{arc} of equation (64), the quoted value of $\Delta s_i/2$ is to be used.

Quasistatic simulation procedure A simulation run is started with any dislocation configuration, for instance a straight dislocation (or several ones) near the lower side of a rectangular obstacle field with the obstacle stress $\tau_{\mathrm{allobst}}(\underline{x})$. Usually this field contains many obstacles k at the locations $\underline{x}_k$; their stress contributions are superimposed locally:

$$\tau_{\mathrm{allobst}}(\underline{x}) = \sum_k \tau_{\mathrm{obst},k}\,(\underline{x} - \underline{x}_k, P_k) \tag{69}$$

In general the obstacles can have individual stress functions $\tau_{\mathrm{obst},k}(\underline{\Delta x})$ and parameters P_k. By solving the differential equation (68) for subsequent time steps, new configurations are found, hence dislocation glide is simulated. Various approaches can be used to derive quantitative results from this glide. One approach is to start with a low external stress τ_{ext}. This drives the dislocation forwards against the obstacles; usually there are many. The dislocation bows out between them until a static equilibrium configuration is found ($v_\perp = 0$ in all points). Then τ_{ext} is increased by a small step $\Delta\tau$ so that the dislocation bows out a bit more. Thereby the dislocation may overcome some obstacles by shearing or circumventing them before it finds the next equilibrium. Then τ_{ext} is increased again; this is repeated until the dislocation touches the upper side of the obstacle field. Then the simulation run is ended, and the last value of τ_{ext}, called $\mathrm{Max}[\tau_{\mathrm{ext}}]$, obviously suffices to keep the dislocation running continuously in this obstacle field. The simulated critical resolved shear stress τ_{sim} is defined as

$$\tau_{\mathrm{sim}} = \mathrm{Max}\,[\tau_{\mathrm{ext}}] - \Delta\tau \tag{70}$$

because this value is the highest one that yields a static dislocation configuration. $\Delta\tau$ should be chosen to be about 2% or 3% of an estimation of the end result τ_{sim}, which may, for instance, be the strengthening contribution τ_{p} of particles. In figure 30 several dislocation configurations are plotted while a single obstacle is overcome, either by shearing or by circumventing it.

The present definition of τ_{sim} may appear a bit arbitrary when an (irregular) array of obstacles is considered: if the array is particularly long so that the dislocation must overcome many obstacles successively, the probability will increase that the dislocation finds an obstacle arrangement which is particularly hard to overcome. This would increase τ_{sim}. On the other

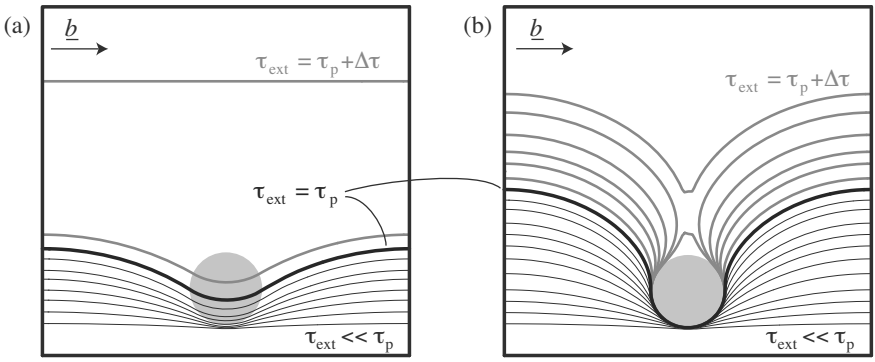


Figure 30. A dislocation overcoming an obstacle by (a) shearing or (b) circumventing it. Black lines are static equilibrium configurations, the grey lines are snapshots of the dislocation moving at a constant stress larger than the critical one, τ_p. In (b) an Orowan loop is left behind around the obstacle.

hand, if the obstacle array is particularly wide, so that a longer dislocation sweeps out in forward direction, this longer dislocation will have an increased chance to find an easy way to move on. This would lower τ_sim. These effects have been found to be rather weak and mostly compensate for each other. In test simulations no systematic dependence of τ_sim on the size of the obstacle field has been detected when the length and the width of the field were varied by equal factors. The typical scatter of τ_sim involved with individual obstacle fields of (mostly) equal statistical properties and size has been found to be $\pm 6\%$ for arrays with 500 to 1000 obstacles (Bacon, 1967), (Foreman and Makin, 1966), (Foreman and Makin, 1967).

The outlined approach to derive τ_sim is quasistatic: The dislocation's velocity tends to vanish in the static equilibrium positions, the last of which defines τ_sim. This means that the derived result is independent of the strain rate. Moreover this means that the drag coefficient B has no meaning and is arbitrary in these simulations: According to equation (68), B defines the time scale in the simulations, but the time required in reality to accomplish the stress increments $\Delta\tau$ is disregarded.

Dynamic simulation procedures In experiments the measured stress to overcome obstacles is often found to depend on the strain rate $\dot{\varepsilon}$. Such a dependence of $\tau_\mathrm{sim}(\dot{\varepsilon})$ can be introduced in simulations by assuming a mobile dislocation density ρ_m and enforcing a predefined dislocation velocity:

$v = \dot{\gamma}/(b\rho_\mathrm{m})$ (Orowan equation (13)). This can be done by continuously adjusting the external stress τ_ext accordingly (Monnet, 2006); in that case τ_sim is defined as the temporal average of τ_ext. By this method the tendency $\partial\tau_\mathrm{sim}/\partial\dot{\varepsilon} > 0$ is found, in agreement with experiments. But this strain rate dependence is only linked to the viscous drag factor B, which accounts for the phonon drag. The much stronger strain rate dependence resulting from thermally activated dislocation motion is not considered in this approach.

In a truly dynamic simulation the thermal stress pulses τ_therm on a dislocation are accounted for. They introduce probabilities for dislocations to overcome obstacles by the help of these pulses at a given temperature. In addition, effects of a dislocation's inertia can be accounted for by a stress term τ_inert. Hence the stress equilibrium of equation (66) is extended to

$$\tau_\mathrm{inert} + \tau_\mathrm{drag} + \tau_\mathrm{ext} + \tau_\mathrm{therm} + \tau_\mathrm{disl} + \tau_\mathrm{allobst} = 0 \tag{71}$$

where again $\tau_\mathrm{drag} = -(B/b)v_\perp$ and equivalently, the inertial stress is defined as $\tau_\mathrm{inert} = -(m^*/b)a_\perp$ with a length specific mass m^* and the acceleration $a_\perp$ normal to the dislocation line vector $\underline{s}^0$. It must be emphasized here that the introduction of m^* is as vague as the introduction of a line tension because the dislocation's inertia is not in the line but in the motion and mass of all atoms surrounding it. Still, a mass of about one atom weight is usually attributed to the length of one Burgers vector (Isaac and Granato, 1988), (Mohles, 1997). With $a_\perp$ being the second derivative in time, the equivalent of equation (68) can be written as

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\underline{x} = \underline{s}^0_\perp \frac{b}{m^*}\left(-(B/b)v_\perp + \tau_\mathrm{ext} + \tau_\mathrm{allobst} + \tau_\mathrm{disl} + \tau_\mathrm{therm}\right) \tag{72}$$

The thermal stress pulses, τ_therm, are essentially random numbers without correlation in time and space, which impose Brownian motion on the dislocation. The pulses add or substract energy to the dislocation in random amounts; on a temporal average energy is added by the term τ_therm. But as the random motion of the dislocation increases, so does the dissipation via τ_drag. Hence an equilibrium will be found in which the energy added to the dislocation by τ_therm is compensated on a temporal average by τ_drag, such that an average amount of random motion resides in the dislocation: this is the temperature. This fluctuation-dissipation-theorem for dislocations defines the amplitude of the random pulses τ_therm:

$$\tau_\mathrm{therm}\left(\underline{r}_i, \Delta t_j\right) = \frac{1}{b}\sqrt{\frac{2Bk_\mathrm{B}T}{\Delta s_i\ \Delta t_j}}\ \Lambda_{i,j} \tag{73}$$

$$\text{with} \quad f(\Lambda) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\Lambda^2}{2}\right)$$

Here k_B is the Boltzmann constant, T is the temperature, Δs_i is the length attributed to the dislocation segment located at $\underline{r}_i$, and Δt_j is the time interval during which the particular stress $\tau_\mathrm{therm}\left(\underline{r}_i, \Delta t_j\right)$ acts. $\Lambda_{i,j}$ are random numbers picked individually from the indicated Gaussian distribution $f(\Lambda)$ for $\underline{r}_i$ and Δt_j. The dependence of τ_therm on the drag coefficient B reflects the energy equilibrium quoted above. Equation (74) ensures for instance that each dislocation node in a fixed glide plane, which represents one degree of freedom, has the average kinetic energy $\left\langle 1/2\left(m^* \Delta s_i\right) v^2 \right\rangle = 1/2 k_\mathrm{B} T$.

A simulation using the stochastic equation (72) never finds an equilibrium configuration due to the thermal stresses. Instead, the dislocation will move through the obstacle field in a jerky manner, as is also observed by in-situ transmission electron microscopy. Instead of a critical stress the mean glide velocity can be evaluated as a function of temperature and τ_ext; this is discussed in section 3.4.

3.2 Static Simulations Using the Line Tension Model

The present section summarizes some simulations using the simple line tension model to calculate the strengthening effect of simple obstacles which give some general insights. This means that equation (68) is used, with τ_line of equation (34) used for τ_disl, and τ_ext is imcremented by steps $\Delta\tau$ as described in section 3.1. Hence only τ_allobst is to be defined via equation (69). For the simulations of the present section, the following simplifying assumptions are used in order to allow for a direct comparison with analytical models (section 2.1) and the more refined simulations from literature:

(i) The obstacles are circular. Inside the obstacles $\tau_\mathrm{obst} = -\gamma/b = \text{const.}$, and outside $\tau_\mathrm{obst} = 0$. This corresponds to energy storing obstacles: When the obstacle area q is swept by a dislocation, the energy $q\gamma$ is employed (by τ_ext) and stored in form of a faulted boundary. This case is similar to order strengthened materials (section 3.3, $\gamma = \pm\gamma_{APB}$). (ii) The radius r of all obstacles in the glide plane is the same. Accordingly the maximum obstacle force $F_\mathrm{max} = 2r\gamma$ and energy $q\gamma = \pi r^2 \gamma$ are well-defined. (iii) The obstacle arrangement in the glide plane is purely random in section 3.2; later on, this condition is relaxed to some extent. (iv) The line tension model is used with the line energy E^* and hence the line tension S being constant, hence $S = E^*$ according to equation (34). (v) Only one dislocation is simulated. All simulated data are presented in the common scales described subsequently. These scales have been chosen because they are convenient for weak and strong obstacles of any extension and arrangement.

An important length scale in a real crystal is the obstacle radius r. Another length scale is the shortest distance L_cc between the obstacles (centre

to centre) which would hold if they were arranged in a square grid. L_{cc}^{-2} is the number of obstacles per unit area. Each obstacle occupies the area $q = \pi r^2$, hence

$$L_{cc} = r\sqrt{\pi/f} \tag{74}$$

where f is the area fraction covered by the obstacles. The obstacle strength $F_{max} = 2r\gamma$ is related to the line tension S; the relative strength is denoted by k.

$$k = \frac{F_{max}}{2S} = r\frac{\gamma}{S} = r\frac{b\tau_{obst}}{S} \tag{75}$$

Note that the fraction S/γ is another length scale in addition to r and L_{cc}. One of these three scales may be taken as the unit length and be kept constant in simulations. For analytical considerations, L_{cc} may be useful as a unit length; however, in the present simulations it is more convenient to keep (S/γ) constant. This latter parameter depends on material parameters alone, and not on the geometry of the obstacles and their arrangement. This geometry is subject to variations in the subsequent simulations.

All stresses are described in units of τ_{unit}, which is defined as:

$$\tau_{unit} = F_{max}/\left(bL_{cc}\right) = \sqrt{(4/\pi)f}\,(\gamma/b) \tag{76}$$

By definition, τ_{unit} is independent of S. This is helpful since S is not known exactly, neither in evaluations of experiments nor in simulations using the self-interaction concept (section 3.3). But τ_{unit} does already include the basic stress dependence on f : $\tau_{unit} \sim f^{1/2}$. For instance, in these units the critical stress predicted by Friedel's model (section 2.1) simply reads as $\tau_{Friedel} = k^{1/2}\tau_{unit}$ with *apparently* no dependence on f.

The magnitude b of the Burgers vector in equation (76) is, in principle, a fourth length scale. But it appears only in τ_{unit} and, therefore, does not interfere with the other length scales.

The normalized simulation results τ_{crit}/τ_{unit} are basically a function of the parameters k and f alone. This gets obvious when all terms in the stress equilibrium (equation (68) with $\tau_{disl} = \tau_{line}$) are divided by τ_{unit}. The ratio τ_{obst}/τ_{unit} depends only on f (τ_{obst} of assumption (i)). The ratio τ_{line}/τ_{unit} can be written as a function of k, f, and the local curvature R_c (in units of the chosen unit length S/γ). The latter stands for the dislocation configuration (on the scale of S/γ). This configuration *adjusts itself* to the parameters (τ_{ext}/τ_{unit}), k, and f in the given obstacle arrangement. The last stable equilibrium configuration for given values of k and f, which defines $\tau_{crit} = Max\,[\tau_{ext}]$, depends on the individual spatial arrangement of the obstacle field used. Hence τ_{crit}/τ_{unit} is a function of k and f and the geometry of the individual obstacle arrangement. The latter merely

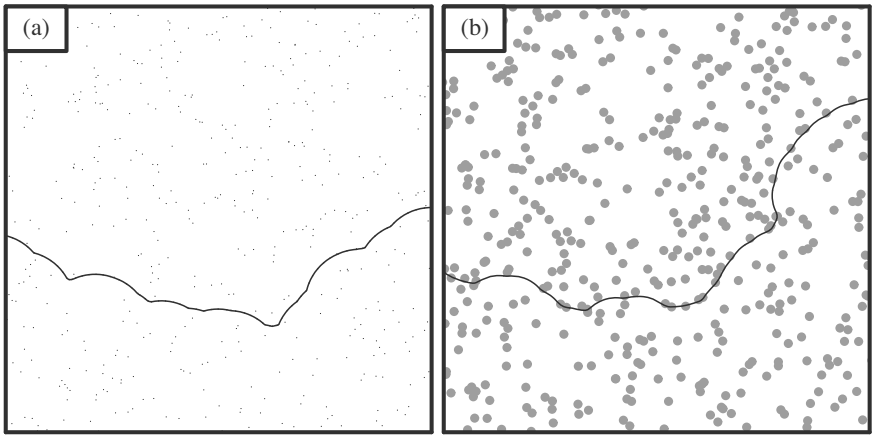


Figure 31. Critical dislocation configurations simulated with the line tension model for randomly distributed overlapping obstacles. Only parts of the obstacle arrays are shown. The number of obstacles ($\approx$ 440 of 8000 total) and the relative obstacle strength $k = 0.304$ are the same. The area fractions of the obstacles and the critical resolved shear stress are (a) $f = 0.0025$ ($\approx$ point obstacles), $\tau_{\rm crit}/\tau_{\rm unit} = 0.60$; (b) $f = 0.16$ (extended obstacles), $\tau_{\rm crit}/\tau_{\rm unit} = 0.91$.

produces a statistical scatter in $\tau_{\rm crit}$. It has been found that this scatter is about $\pm 6\%$ for obstacle arrays containing 500 to 1000 obstacles (see section 3.1). For the present simulations with 8000 obstacles in the square-shaped obstacle field, the scatter is estimated less than $\pm 4\%$.

Effects of the obstacle strength and size When obstacles are vanishingly small ($r \ll L_{\rm cc}$, or $f \to 0$) they can be considered as point obstacles. The effect of the relative obstacle strength k on dislocation configurations and on $\tau_{\rm crit}$ in arrays of point obstacles have been worked out in literature by analytical models like Friedel's and also by computer simulations (Foreman and Makin, 1966), (Foreman and Makin, 1967). However, the condition $r \ll L_{\rm cc}$ (or $f \ll 1$) is not sufficient to rate the applicability of the results (Mott and Nabarro, 1948), (Schwarz and Labusch, 1978) because another independent length scale beside r and $L_{\rm cc}$ is involved, namely the material dependent constant S/γ.

If obstacles have a finite extension, a distinction must be made whether the obstacles are allowed to overlap each other spatially or not. Of course,

objects like second phase particles cannot overlap in reality, but their obstacle stresses can. Subsequently the respective obstacles are called overlapping and discrete, respectively. The term 'non-overlapping' as the opposite of 'overlapping' is avoided because it is prone to cause confusion. An example for overlapping obstacles are particles with a lattice mismatch (see e.g. section 3.3): their coherence stresses superimpose (overlap) linearly. In contrast, in dispersion strengthening and order strengthening (section 3.3) the stresses do not overlap spatially; these particles are discrete obstacles.

Figure 31 shows two examples of simulated critical dislocation configurations for a very low ($f = 0.0025$) and a rather high ($f = 0.16$) area fraction. The relative obstacle strength $k = 0.304$ is the same in both cases. With $f = 0.0025$ the dislocation configuration is very similar to those found by Foreman and Makin and successors by means of the circle rolling technique (Foreman and Makin, 1966), (Foreman and Makin, 1967), (Hanson and Morris, 1975). Even with $f = 0.16$ the configuration looks rather similar. This may surprise because with less free space between the obstacles at $f = 0.16$, weaker bow-outs are to be expected. But in the obstacle array this is 'compensated' for by a higher external stress: $\tau_{\mathrm{crit}}/\tau_{\mathrm{unit}}$ is found to be higher for $f = 0.16$ than for $f = 0.0025$ by about 50% although the obstacle strength k and density (their number) are the same. Hence this must be an effect of f, or the relative obstacle extension r/L_{cc}.

In figure 32, τ_{crit} is plotted (empty symbols) as a function of the square root of the normalised obstacle strength k, equation (75), for four area fractions f. For comparison results from literature are added. As to be expected, $\tau_{\mathrm{crit}}/\tau_{\mathrm{unit}}$ for $f = 0.0025$ agrees well with results from literature for point obstacles in the full range of k: with $L_{\mathrm{cc}} \gg r$ the obstacles are point-like. At low k, the simulations agree with Friedel's result τ_{Friedel} of equation (71) very well. Near $k \approx 1$ the strength reaches a maximum, albeit τ_{crit} is reduced by a factor 0.8 relative to τ_{Friedel}; this is attributed to the randomness of the obstacle array (Foreman and Makin, 1966), (Foreman and Makin, 1967). In the range $k > 1$, $\tau_{\mathrm{crit}}/\tau_{\mathrm{unit}}$ decreases because the Orowan process operates here.

At higher volume fractions $f = 0.01$, which might still be considered as small, systematic deviations with increasing f are found: $\tau_{\mathrm{crit}}/\tau_{\mathrm{unit}}$ increases with increasing f, for weak and for strong obstacles. This means that for reasonably strong particle strengthening effects, which usually require $f > 0.01$, point obstacles are rather inaccurate as a model. In the following, generally, the data with $f = 0.16$ are discussed because they show the strongest effects. But the statements and conclusions also hold for lower area (or volume) fractions.

For weak obstacles an increase of τ_{crit} with increasing obstacle extension

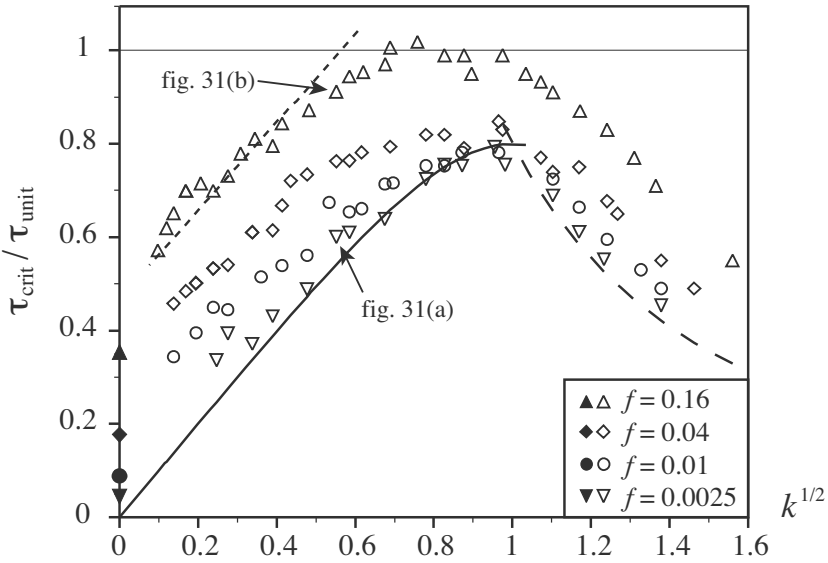


Figure 32. Normalized critical stress as a function of k and f. Empty symbols represent simulations, filled symbols indicate the energy storing effect: $\tau_{\text{store}} = f\gamma/b$; —— $\tau_{\text{FM}} = (1 - 0.2k^2)k^{1/2}\tau_{\text{unit}}$; — — $\tau_{\text{OFM}}(f = 0) = 0.8k^{-1}\tau_{\text{unit}}$; - - - - $\tau_{\text{LS}}(f = 0.16) = (0.94k^{1/2} + 1.17f^{1/2})\tau_{\text{unit}}$.

has already been found by Labusch and Schwarz (Labusch and Schwarz, 1992). The computer simulations of these authors were originally designed for solid solution hardening, but the results can also be applied to particle strengthening. Labusch and Schwarz used obstacles extended only in the direction normal to the dislocation line. Similar to point obstacles, such obstacles cannot overlap because they are one-dimensional (1-D); their areas vanish. This may lead to the perception that these obstacles are non-overlapping and hence discrete, but physically these 1-D obstacles overlap since a non-vanishing obstacle area (area fraction $f > 0$) is assigned to the 1-D obstacles in order to apply a simulation result, hence they inevitably have a finite width $w_x > 0$. The obstacle arrangement of Labusch and Schwarz was purely random (a pre-requisite for the scaling used by these authors); therefore, the obstacles can actually overlap, and they increasingly do so with increasing w_x or f.

In the units used here, the simulation results τ_{SL} of Labusch and Schwarz for energy storing obstacles reads: $\tau_{\text{LS}} = \left(0.94k^{1/2} + 1.17f^{1/2}\right)\tau_{\text{unit}}$. As in-

dicated for $f = 0.16$ in figure 32, this result also represents the present simulations well. Essentially it appears that in the range $0 < k < 1$, $\tau_{\mathrm{crit}}/\tau_{\mathrm{unit}}$ is increased relative to $\tau_{\mathrm{Friedel}}/\tau_{\mathrm{unit}} = k^{1/2}$ by the amount $f^{1/2}$. This effect is caused by the energy storing character of the obstacles: when a dislocation is straight as in the case $k = 0$, the fraction f of it resides inside obstacles and senses the obstacle stress $\tau_{\mathrm{obst}} = -\gamma/b$. This leads to the critical stress $\tau_{\mathrm{crit}} = f\gamma/b = (\pi f/4)^{1/2}\tau_{\mathrm{unit}}$ for straight dislocations. If the obstacles involved negative and positive obstacle stresses τ_{obst} in equal amounts, this contribution to τ_{crit} would vanish. This latter case means that the dislocation stores no energy when it overcomes obstacles, unlike in the present case. More accurate statistical investigations have shown (Arsenault et al., 1989a), (Arsenault et al., 1989b) that stress fluctuations also give rise to a stress $\tau_{\mathrm{crit}}(k = 0)$; however, this is beyond the scope of the present contribution.

In the case of very strong obstacles with $k > 1$, which are circumvented by the Orowan process, τ_{crit} decreases with increasing k because the free space between the obstacles, $L_{\mathrm{cc}} - 2r$, scales with k (equation (75)). In the present units, Orowan's stress prediction (26) reads $\tau_{\mathrm{crit}} = \left(1 - (4f/\pi)^{1/2}\right)^{-1} k^{-1}\tau_{\mathrm{unit}}$. This also predicts an increase of the relative stress $\tau_{\mathrm{crit}}/\tau_{\mathrm{unit}}$ with increasing volume fraction f, in full agreement with figure 32.

Around $k = 1$, τ_{crit} has a maximum. At $f = 0.16$ it can be seen that $\tau_{\mathrm{crit}}/\tau_{\mathrm{unit}} \leq 1$ defines a maximum for τ_{crit}. It is reached when (i) the obstacles exert the maximum force (i.e. $2S$) and (ii) the distance between obstacles is small. The smallest distance that can be reached everywhere in the glideplane for a given number density is L_{cc}, the centre-to-centre square lattice spacing. This had been used for the definition of τ_{unit} (equation (76), so that $\tau_{\mathrm{crit}} \leq \tau_{\mathrm{unit}}$ holds for any other obstacle arrangement, regardless of f.

Effects of the obstacle arrangement The geometry of the obstacle arrangement can affect the resulting critical stress, as can be seen in figure 33 by comparing the overlapping obstacles of figures 31 and 32 with discrete obstacles as defined above. The arrangement of discrete obstacles is generated in the computer by assigning random locations to all obstacles first, disregarding overlappings like above. Then all obstacles which are found to overlap with others are assigned new random locations. The latter step is repeated until no more overlaps are found. The prevention of overlaps means a restriction for the randomness of the obstacle arrangement: at very large area fractions f, the arrangement of rigid obstacles will be close to a closest packing, which is a very regular arrangement. The restriction of

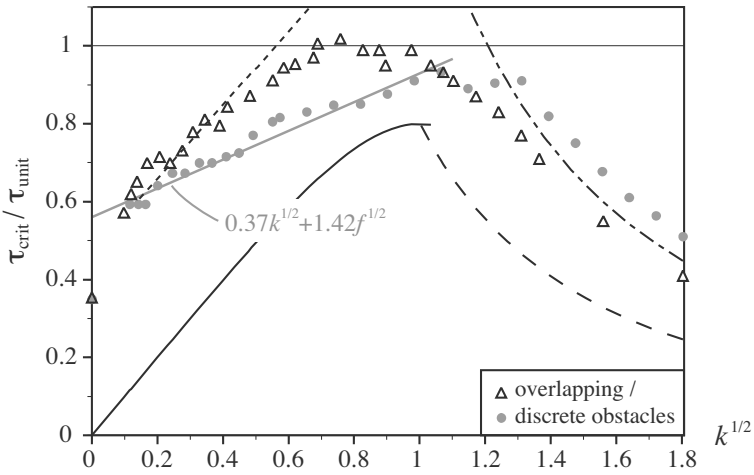


Figure 33. Normalized critical stress as a function of k at $f = 0.16$ for overlapping (black) and discrete (grey) obstacles; lines as in figure 32 except for — - — $\tau_{\mathrm{OFM}}(f = 0.16) = 0.8k^{-1}\left(1 - (4f/\pi)^{1/2}\right)^{-1}\tau_{\mathrm{unit}}$.

randomness is obviously most pronounced at high area fractions; therefore only the case $f = 0.16$ is considered subsequently.

In the range of strong obstacles ($k > 1$), the discrete obstacles yield a higher critical stress than the overlapping ones (figure 33). This has two reasons. Firstly, in the case of overlapping obstacles the fraction f of f, i.e. f^2, does not contribute to strengthening because it overlaps. This reduces the effective volume fraction to $f - f^2$ and lowers τ_{crit} correspondingly. Secondly, in case of a regular obstacle arrangement as is the tendency for the discrete obstacles, the probability for the dislocation to find weak spots (large individual obstacle distances) is low; this also increases τ_{crit} .

For weak obstacles ($k < 1$), the overlapping obstacles yield a higher stress than the discrete ones. This is most obvious for $k^{1/2} \approx 0.7$. A possible explanation is that the overlapping areas can be seen as separate obstacles with double strength ($\tau_{\mathrm{obst}} = -2\gamma/b$). The area fraction of these obstacles of double strength equals $f^2 = 0.0256$. The total area fraction covered by obstacles is reduced from f to $f - f^2$ due to overlapping. This would lower τ_{crit} like in the case $k > 1$. But this is overcompensated by the obstacles with double strength because they have a stronger influence than the normal ones: $\tau_{\mathrm{crit}} \sim k^{1/2}\tau_{\mathrm{unit}}$ according to Friedel's model, where both τ_{unit} and k increase linearly with γ.

The simulated data of discrete obstacles for $f = 0.16$ and $k < 1$ are well represented by a linear function of $k^{1/2}$ (grey line in figure 33), similar to τ_{LS} (see figure 32), but with different coefficients: $\tau_{\mathrm{crit}}/\tau_{\mathrm{unit}} = 0.37k^{1/2} + 1.42f^{1/2}$ instead of $\tau_{\mathrm{LS}}/\tau_{\mathrm{unit}} = 0.94k^{1/2} + 1.17f^{1/2}$. While $\tau_{\mathrm{LS}}/\tau_{\mathrm{unit}}$ is essentially in line with τ_{Friedel} and τ_{FM} (parallel lines in figure 32), which had also been derived for random obstacle arrangements, the results for the more regular arrangement of the discrete obstacles deviates strongly. This emphasizes the importance of using realistic obstacle arrangements in simulations.

3.3 Simulations Using Dislocation Self Interaction: Particle Strengthening

The present section summarizes some simulations using the accurate dislocation self interaction model and the quasi-static procedure to calculate the strengthening effect of spherical second phase particles in a material. This means that equation (68) is used, with τ_{self} of equation (65) for τ_{disl}. τ_{allobst} is defined by equation (69) with a realistic distribution of particles as obstacles (locations $\underline{x}_k$, radii r_k). The interaction functions $\tau_{\mathrm{obst}}(\underline{\Delta x})$ depend on the particle types.

Realistic particle arrays As has been shown in section 3.2, the arrangement of obstacles has a significant impact on the critical stress τ_{crit}; the same holds for other geometrical properties. For instance, the impact of the size distribution of particles on their strengthening effect has been worked out statistically by Nembach (Nembach, 1996). Realistic particle arrangements may be obtained from atom-probe tomographic reconstructions or phase field or Monte-Carlo-simulations of precipitation. But in most cases, the obtainable particle arrangements are too small, meaning that they contain too few particles to keep the statistical scatter involved with individual arrangements low.

For the subsequent simulations, an arrangement is used which has been generated artificially and compared to real particle arrangements (Mohles and Fruhstorfer, 2002). The particles are assumed to be spherical and to have the radius distribution $g_{\mathrm{LSW}}(\rho)$ which has been derived analytically by Lifshitz, Slyozov (Lifshitz and Slyozov, 1961) and Wagner (Wagner, 1961) for the case of Ostwald ripened particles of low volume fractions f. In the

range $\rho = r/r_0 < 1.5$,

$$g_{\mathrm{LSW}} = \frac{4}{9}\rho^2 \left(\frac{3}{3+\rho}\right)^{7/3} \left(\frac{1.5}{1.5-\rho}\right)^{11/3} \exp\left(\frac{\rho}{\rho - 1.5}\right) \tag{77}$$

and $g_{\mathrm{LSW}}\,(\rho \geq 1.5) = 0$, where r are individual radii and r_0 is the mean particle radius. With this size distribution a rather closely packed arrangement of spheres is generated. Each of these spheres is meant to represent the corona of volume from which one particle has obtained its material during the Ostwald ripening process. By scaling appropriately all distances and individual corona radii down to the particle radii, the mean particle radius r_0 and the volume fraction f is selected. Finally, these particles are shifted by small additional random vectors to adjust the arrangement's randomness to experimental findings. The procedure is detailed and the resulting arrangements are discussed in (Mohles and Fruhstorfer, 2002).

Dispersion strengthening In dispersion strengthened crystals the glide system in the matrix is not continued inside the particles. Hence the particles are impenetrable; they must be circumvented by the Orowan process. In simulations this is enforced with a negative obstacle stress of large amount ($\tau_{\mathrm{obst}} \to -\infty$) inside the obstacles and no stress outside. For reasons of numerical stability it may be useful or even necessary (depending on the numerical integration method) to assume a finite but sufficiently high stress instead, and to smooth out the stress transition in $\tau_{\mathrm{obst}}(\underline{\Delta x})$ on the particle's interface in order to the limit slope of the step function. Details on the chosen stress value inside the particles and on the smoothing procedure are given in (Mohles, 2001c).

To give an impression, two dislocation configurations of a simulation using this obstacle stress are plotted in figure 34. $\tau_{\mathrm{obst}}(\underline{x})$ is represented by the dark grey disks: these are the intersections of the particles (arranged in 3-D) with the glide plane. The size distribution of these disks result from the particle size distribution $g_{\mathrm{LSW}}\,(r/r_0)$ and from the distribution of heights at which the particles intersect the plane. The starting configuration was a straight screw dislocation (dashed; $\underline{b}$ is indicated in figure 34); however, the dislocation bends strongly while gliding such that all dislocation characters are present. The local bending between the particles is stronger in x- than in y-direction because edge dislocations are more flexible than screw dislocations. This is predicted by the line tension model (section 2.1) and obviously remains true in the present case, where each dislocation piece interacts with every other one, including the Orowan loops. Quantitative results of dispersion strengthening are presented and discussed in (Mohles

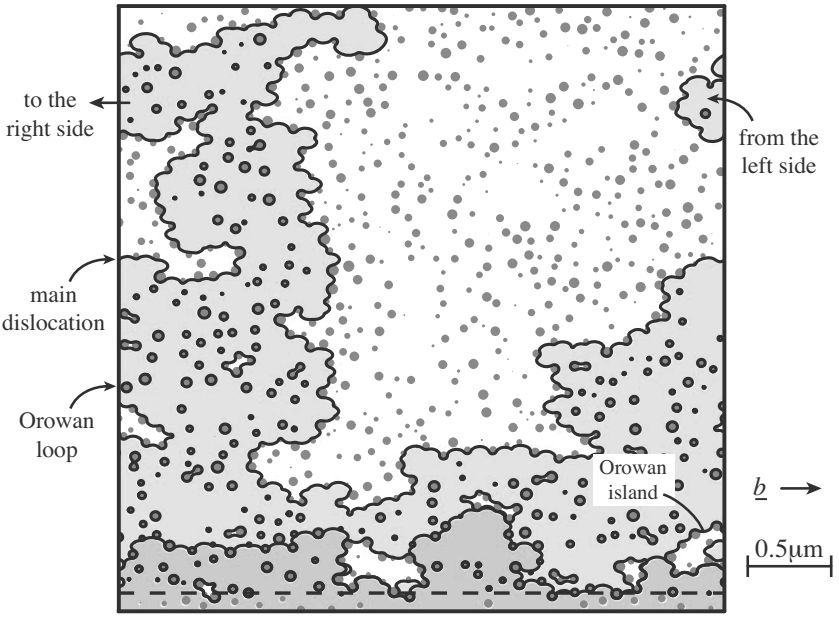


Figure 34. Two dislocation configurations in a dispersion strengthened crystal with $r_0 = 100b$, $f = 0.1$, Poisson ratio $\nu = 0.4$. The dashed line indicates the starting position; on the left and right periodic boundary conditions are applied. Dark grey: particle intersections with the glide plane; medium / light grey: area swept out at $\tau_{\text{ext}} = 226\text{MPa}$ / 230MPa. Representative configurations for other parameters can be seen in (Mohles and Fruhstorfer, 2002). Note that every piece of dislocation interacts elastically with all other pieces.

and Nembach, 2001).

Order strengthening In order strengthened materials, the crystal structure is continued inside particles (they are coherent), and they usually have long-range order whereas the matrix is disordered. This causes the dislocations to glide in pairs of two: the first one of a pair is pushed backward inside the particles because it destroys the order there, which takes energy, and creates an antiphase boundary with the energy γ_{APB}. A second dislocation in the same glide plane may restore this order and is, therefore, pushed forward. Hence according to equation (19) the obstacle stress for

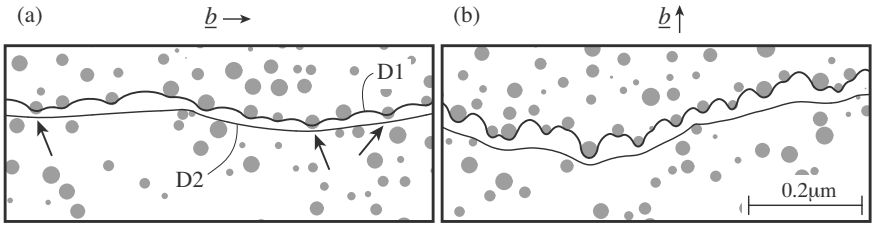


Figure 35. Critical dislocation configurations in a peak-aged order strengthened crystal for a pair of (a) screw and (b) edge dislocations. In the screw configuration the leading dislocation D1 often penetrates the particles (arrows) so that D2 is likely to follow immediately; edge dislocations do not show this effect.

these dislocations is described by

$$\tau_{\text{obst}}\,(\underline{\underline{\Delta x}}) = \begin{cases} -\gamma_{\text{APB}}/b \text{ for 1st dislocation inside particle} \\ +\gamma_{\text{APB}}/b \text{ for 2nd dislocation inside particle} \end{cases} \tag{78}$$

and $\tau_{\text{obst}} = 0$ outside the particles. Like in dispersion strengthening a smooth step function should be used for $\tau_{\text{obst}}(\underline{x})$ on the particle interfaces.

Figures 35(a) and (b) show sections of a pair of simulated screw and edge dislocation pairs. All parameters are chosen close to those of the Nickel based superalloy Nimonic PE16: $b = 0.25$nm, $\mu = 65$GPa, $\nu = 0.4$, $\gamma_{\text{APB}} = 0.25\text{Jm}^{-2}$. The volume fraction is $f = 0.1$, and the mean particle radius is $r_0 = 49b$. These latter parameters define the aging state of the material: a thermal treatment (prior to our plastic deformation) will define the amount (f) and the size (r_0) of the particles that precipitate from the matrix due to classical nucleation, growth and ripening.

If only one dislocation is considered, equation (78) is similar to the obstacle stress $-\gamma/b$ assumed for the obstacles in section 3.1. Accordingly the critical stress of order strengthed materials can be discussed in terms of the relative obstacle strength k introduced in equation (75): for a given material, which defines defines (S/γ_{APB}), the parameter k is essentially equivalent to the mean particle radius r_0. The line tension S is not defined in the present simulations because the dislocation self interaction model is used, but S is known to be rather constant as it depends on the surrounding dislocation geometry only through a logarithm function (equation 24). Similar to the simulation results of section 3.2, the strengthening of particles reaches a maximum at some mean radius r_0 (where $k \approx 1$). The corresponding material is denoted as peak-aged. With smaller or larger radii, the material is

in the underaged or overaged aging state, respectively.

The dislocation configurations in figure 35 belong to the peak-aged state. The leading edge dislocation D1 in figure 35(b) bows out strongly between the particles, almost as strong as in the case of dispersion strengthening in figure 34: D1 almost circumvents the particles. Actually, in case of large radii r_0, the configuration of D1 looks just like in figure 34; D2 is stuck behind Orowan loops left behind by D1. On the other hand, the leading screw dislocation D1 in figure 35(a) shears the particles, as is indicated by arrows. The trailing dislocation D2 is always rather straight because it touches almost no particles; and if it does so, it is pushed forwards as modelled. In the underaged state (small radii) both dislocations are rather straight, and D1 and D2 move as a pair. Altogether the dependence $\tau_{\mathrm{crit}}\left(f, r_0\right)$ for order strengthening is rather complex because of the combined shearing and circumventing of the particles, as well as by D1 and D2 moving either as a pair or individually. This is explained in more detail in (Mohles, 2004).

Lattice mismatch strengthening In the previous subsections of section 3.3 the particles were discrete obstacles because only in their interior, the dislocations interacted with them. This is different when the particles have a lattice mismatch: inside the particles there is hydrostatic stress, and outside there are shear stresses that cause a long-ranged interaction with dislocations. Hence these particles are overlapping obstacles, and moreover, their arrangement in 3-D space is important because the obstacle stresses depend on the height Δz of the particle centre over the glide plane. In the former cases, only the particle intersections with the glide plane mattered. The stress tensor of a spherical inclusion in an isotropic medium had been derived by Eshelby (Eshelby, 1956); applying the Peach-Koehler formula (16) on this yields the obstacle stress:

$$\tau_{\mathrm{obst}}\left(\underline{\Delta x}, \underline{b}^0\right) = \begin{cases} 6\varepsilon\mu\Delta z r^3 |\underline{\Delta x}|^{-4} \left(\underline{b}^0 \cdot \underline{\Delta x}^0\right) & \text{outside the particle} \\ 0 & \text{inside} : |\underline{\Delta x}| < r \end{cases} \tag{79}$$

Here ε denotes the constrained (relaxed) relative lattice mismatch. Inside the particle τ_{obst} vanishes because a driving force for dislocation glide results only from shear stresses. Outside, the stress depends on the direction of the Burgers vector $\underline{b}$ (but not on the line vector $\underline{s}$). This means that a static simulation procedure with τ_{obst} and τ_{allobst} changes as $\underline{b}$ is rotated, but for a given dislocation, $\underline{b}$ is constant and hence the obstacle stress is unique for it. Still, edge and screw dislocations experience different overall stresses because for a given $\underline{b}$, screw and edge dislocations experience τ_{obst} in different locations (as $\underline{s}$ must differ).

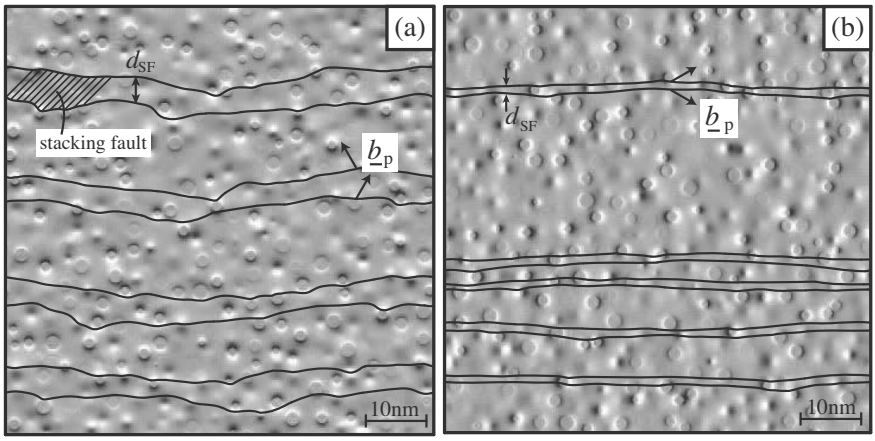


Figure 36. Configurations of one pair of Shockley partial dislocations in an underaged lattice mismatch strengthened crystal with $f = 0.1$. The greyscale plot of τ_{allobst} (dark: > 0, light: < 0) is that of the corresponding perfect dislocation. (a) Dissociated edge, (b) screw dislocation.

The partials of a dissociated dislocation sense different obstacle stresses when their Burgers vectors $\underline{b}_{p1}$ and $\underline{b}_{p2}$ differ, like in the case of Shockley partials (section 1.4). In figure 36 several dislocation configurations of such a case are plotted for a copper matrix with cobalt particles ($\mu = 42\text{GPa}$, $\nu = 0.43$, $\underline{b}_p = 0.148\text{nm}$, $\gamma_{SF} = 0.036\text{Jm}^{-2}$, $\varepsilon = 0.015$). The edge dislocation in figure 36(a) is dissociated more strongly than the screw dislocation in 36(b). With the elastic dislocation repulsion and the attraction stress $\pm\gamma_{SF}$ from the stacking fault considered in the simulation, the distance d_{SF} between the partials adjusts automatically in a simulation, and the mean value of d_{SF} agrees with d_{eq} of equation (23). The background of figures 36(a) and (b) is a greyscale plot of the obstacle stress $\tau_{\text{allobst}}\left(\underline{\Delta x}, \underline{b}^0\right)$, which is the combination of equation (79) with (69) for a given arrangement of particles k located at $\underline{x}_k$. Since τ_{allobst} depends on $\underline{b}_0$, the vector of the corresponding perfect dislocation has been used for the greyscale plot.

The dissociation width d_{SF} is another length scale in addition to those described in section 3.2. This makes analytic predictions of τ_{crit} very difficult, in particular when d_{SF} is in the range of the particle radius or the free space between them, like in the case of figure 36. Another difficulty for analytic considerations is that τ_{allobst} represents a wide spectrum of obstacle strengths; the effective number and strength of obstacles cannot be defined

unequivocally as can be seen from the greyscale plots. This makes it hard to apply analytic strengthening models as those in section 2.1 (Nembach, 1996). However, in dislocation simulations all effects can be considered concurrently. More details on such simulations can be found in (Mohles and Nembach, 2001), (Mohles, 2001a), (Mohles, 2001b), (Mohles, 2002), (Mohles, 2003).

3.4 Simulations of Thermally Activated Dislocation Glide

In section 3.1, the procedure for dynamic simulations considering random thermal stress pulses has been described. The stochastic nature of the differential equation (72) prevents the dislocations from ever finding a local equilibrium; they always stay in motion. This makes simulations particularly slow when the self interaction concept is used because some of its optimizations (section 4.2) remain useless. Therefore the line tension approach is used for the present simulations: $\tau_{\mathrm{disl}} = \tau_{\mathrm{line}}$ of equation (34).

Thermal activation helps dislocations to overcome obstacles. This means that the stress required to overcome the obstacles is lowered by a dimensionless factor $\tau_{\mathrm{rel}} < 1$ which depends on the temperature T, the imposed shear rate $\dot{\gamma}$ and possibly on other parameters:

$$\tau_{\mathrm{ext}} = \tau_{\mathrm{rel}}\,(T,\ \dot{\gamma},\ ...)\,\tau_{\mathrm{0K}} \tag{80}$$

where τ_{0K} is the stress required to overcome the obstacles by mechanical stress alone, that is at $T = 0\mathrm{K}$. In the terms used so far for the static simulations, τ_{0K} is the same as τ_{crit} with all its dependences on r_0, on f or on other parameters. When the obstacles are large, the thermal pulses are not strong enough to push a dislocation over obstacles. In that case the stress $\tau_{\mathrm{ext}} = \tau_{\mathrm{0K}}$ is required, or $\tau_{\mathrm{rel}} \to 1$ virtually independent of T. In contrast, small obstacles like solved foreign atoms are overcome by thermal activation. This causes solid solution strengthening to be temperature dependent.

Jerky dislocation glide In order to simulate solid solution strengthening, the obstacle stress $\tau_{\mathrm{obst}}(\underline{\underline{\Delta x}})$ of single foreign atoms must be defined. Here only an approximating function can be used as long as the foreign atom's interaction with the atomistic nature of the dislocation core is not known accurately. A possible approach is to consider a foreign atom as a sphere and calculate its size mismatch by equation (79) with the radius $r = b/2$. The locations $\underline{x}_k$ can be chosen to represent lattice or interstitial sites of the crystal lattice considered. In the greyscale plots of τ_{obst} in figure 37 used later on, the foreign atoms are assumed to be substitutive in a face centred cubic crystal. The stress of equation (79) is altered slightly

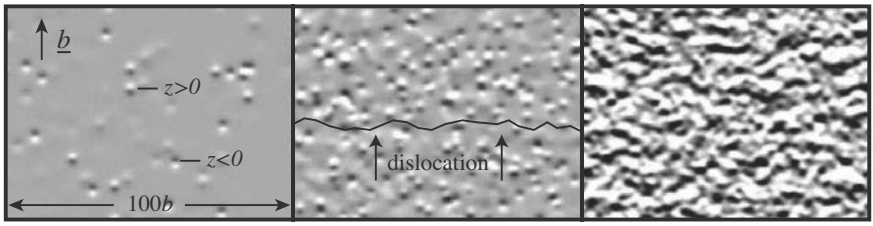


Figure 37. Greyscale plot of the assumed atomic mismatch obstacle stress at $c_{\mathrm{at}} = 10^{-3}/10^{-2}/10^{-1}$.

by ignoring the 'hole' (vanishing stress) inside the atoms. Furthermore, the stress is convoluted with a Gauss function $\exp\left(-|\underline{\Delta x}|^2/b^2/2\right)$ in order to account for the non-discrete nature of a real dislocation core; details are not important here. Figure 38 shows dynamic configurations of thermally activated dislocation glide in such obstacle fields in equidistant time steps. In some areas the dislocation moves fast (low line density) and in other places its overall velocity vanishes; here the dislocation only fluctuates (black areas), waiting for activation. Accordingly, the mean velocity v_{mean} of this jerky motion through the whole field can be written as

$$v_{\mathrm{mean}} = \frac{\lambda}{t_{\mathrm{wait}} + t_{\mathrm{run}}} \approx \frac{\lambda}{t_{\mathrm{wait}}} \tag{81}$$

where λ is the mean free path between waiting positions, and t_{run} and t_{wait} are the mean time intervals of running and waiting, respectively. Usually t_{run} is fully negligible compared to t_{wait}. The free path λ may be considered as a constant defined by the obstacle density, for instance by $\lambda^2 \approx b^2/c_{\mathrm{at}}$. This is not always the case, especially when inertial effects become important and the dislocation overruns obstacles with its kinetic energy (Mohles, 1997). Even in figure 38, an estimated λ would be larger for $c_{\mathrm{at}} = 10\%$ than for $c_{\mathrm{at}} = 1\%$, which is in contradiction to the aforementioned definition of λ. Still, the mean velocity is mostly determined by t_{wait} and its dependence on the temperature T and the present stress τ_{ext}, both of which assist activation (transition from waiting to running).

In figures 38(b) and (c), the same parameters have been used except for the sets of random numbers for τ_{therm}. The resulting mean velocities v_{mean} for this obstacle field differ by a factor two in these simulation runs. Hence in order to derive a function $v_{\mathrm{mean}}(T, \tau_{\mathrm{ext}})$ from simulations, many single simulation results must be averaged. This of course means a high numerical effort, which is another argument against using the dislocation self interaction concept for thermally activated dislocation glide. On the other hand,

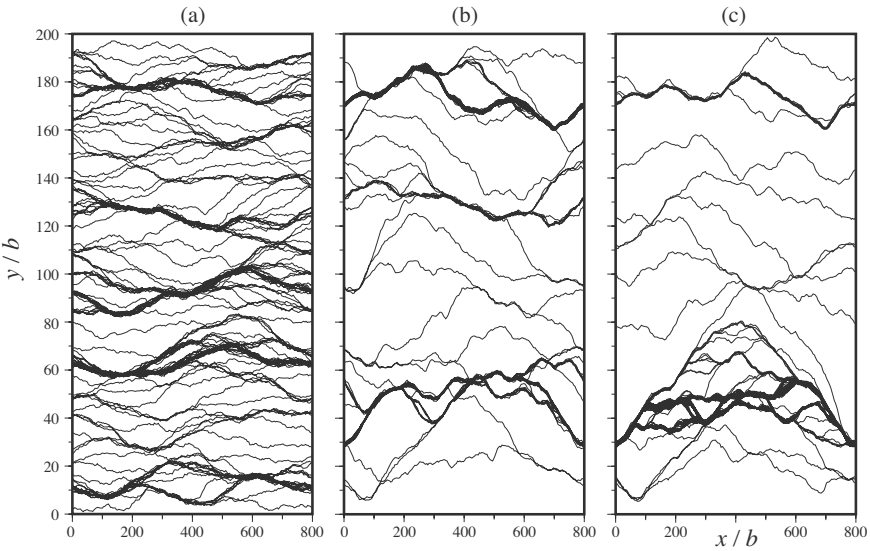


Figure 38. Dislocation configurations of thermally activated dislocation glide after equidistant time intervals. (a) $c_{\mathrm{at}} = 1\%$, $\tau_{\mathrm{ext}} = 7\mathrm{MPa} \rightarrow v_{\mathrm{mean}} = 10\mathrm{m/s}$; (b) $c_{\mathrm{at}} = 10\%$, $\tau_{\mathrm{ext}} = 27\mathrm{MPa} \rightarrow v_{\mathrm{mean}} = 6.1\mathrm{m/s}$; (c) $c_{\mathrm{at}} = 10\%$, $\tau_{\mathrm{ext}} = 27\mathrm{MPa} \rightarrow v_{\mathrm{mean}} = 3.3\mathrm{m/s}$. In (b) and (c) all parameters are equal except for the sets of random numbers for τ_{therm}, causing different configurations and mean velocities v_{mean}. Note that the y-direction has been streched by a factor 8 for clarity.

v_{mean} has essentially been broken down to the mean activation event with its mean waiting time t_{wait}. Individual waiting times vary much stronger than by a factor of two, but the activation event can be treated analytically and also by simulation.

Simulated activation event In figure 39, three simulated activation events are plotted. The obstacles are small size mismatched spherical particles which are not intersected by the glide plane; details are not important here. In figure 39(a), the obstacles are arranged at random. They are all equal in strength, size and shape, but there are attractive and repulsive obstacles. Starting from the static equilibrium position found at $T = 0\mathrm{K}$ (dashed line), the dislocation fluctuates for a while at $T = 300\mathrm{K}$ (grey lines around the dashed one) and eventually reaches a critical configuration (thick black line). When such a configuration has been reached, the dislocation

moves on (grey lines beyond the critical configuration): it has overcome the obstacle arrangement. The fact that the activation takes place on the boundary is physically meaningless because the boundary is periodic. The critical configuration has been estimated from observation of many simulated activation events; a different one has been published in (Mohles and Rönnpagel, 1996) for the same obstacle and dislocation parameters. Always in a critical configuration, the dislocation has moved forwards from its static equilibrium by Δy equal to about one obstacle diameter. The length of dislocation which has moved forward is much larger than Δy, by roughly two orders of magnitude. Note that in figure 39, the y-direction has been stretched by a factor ten. The estimated number of obstacles that are overcome concurrently in the critical configuration is two or three.

The activation event for the equidistant obstacle row in figure 39(b) can be described in exactly the same way. Even the same estimated activation length L_{est} is found from the critical configuration. It has often been assumed in literature (e.g. (Kocks et al., 1975), (Nadgornyi, 1988)) that this length is related to the distance L_{cc} between obstacles, but figure 39(c) proves that this assumption is fundamentally wrong. There, a continuous obstacle wall with the stress $\tau_{\mathrm{wall}}(y)$ replaces the obstacles such that

$$\tau_{\mathrm{wall}}(y) = \frac{1}{L_{\mathrm{cc}}} \int_{-\infty}^{\infty} \tau_{\mathrm{row}}(x, y)\mathrm{d}x \tag{82}$$

where $\tau_{\mathrm{row}}(x, y)$ is the stress function of the obstacles in figure 39(b). The obstacle wall stands for an obstacle row with a high density ($L_{\mathrm{cc}} \to 0$) of weak obstacles ($\mathrm{d}F_{\max} = b\tau_{\mathrm{obst}}\mathrm{d}x \to 0$). Since no length L_{cc} is involved in the activation event of figure 39(c), the activation length L_{act} (of which L_{est} is an estimation) cannot be related to the obstacle density. Instead, L_{act} depends on the dislocation's flexibility (or, the line tension S) and the present stresses τ_{wall} and τ_{ext}. This is similar to section 3.2, where the term $(b\tau_{\mathrm{obst}}/S)$ defined an important length scale. In the subsequent section the activation event and the related length L_{act} are considered analytically.

Analytic description of the activation event In the simulations of section 4.2, a critical dislocation configuration has been estimated from simulations. The criticality of this configuration lies in the dislocation's decision to either move back and continue waiting or to move forward and leave the waiting position: this is an unstable state close to an unstable equilibrium (saddlepoint configuration). The static stress equilibrium of a dislocation is given by equation (71) with $\tau_{\mathrm{therm}} = \tau_{\mathrm{drag}} = \tau_{\mathrm{inert}} = 0$. For the present case, $\tau_{\mathrm{allobst}}(x, y)$ is replaced by $\tau_{\mathrm{wall}}(y)$, and $\tau_{\mathrm{line}}(R_{\mathrm{c}}$ of equation (34)) is used for τ_{disl}. The dislocation's shape can be described as a function

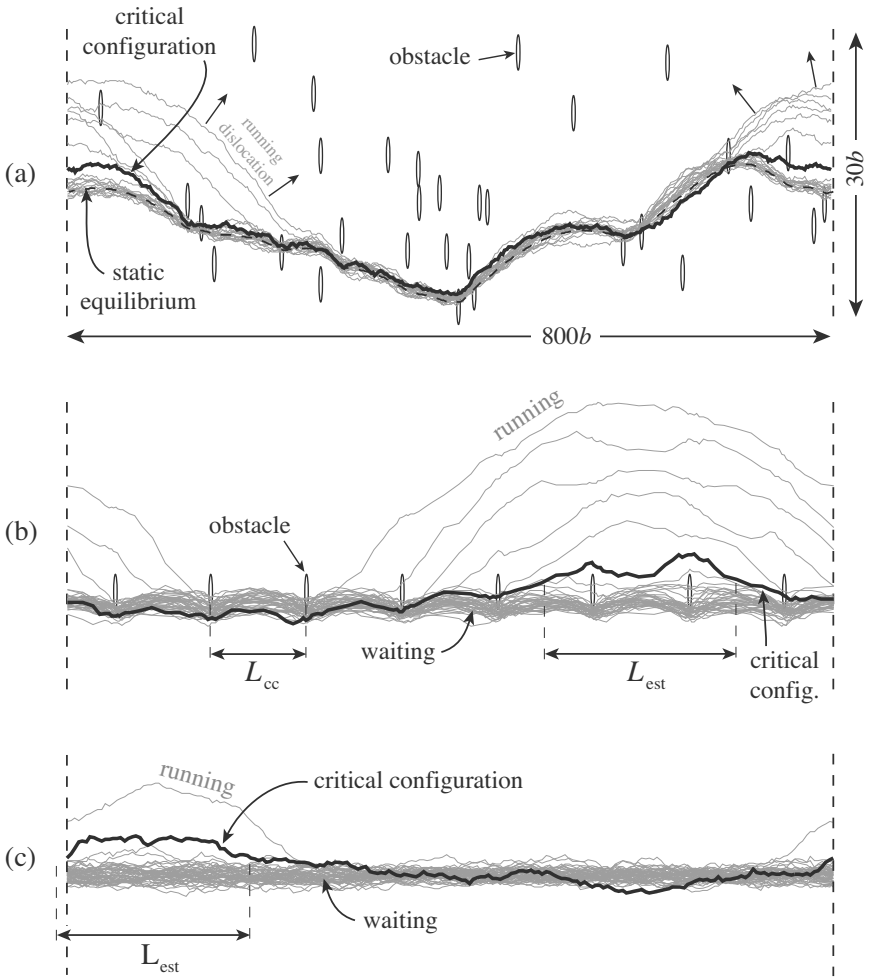


Figure 39. Dislocation configurations of activation events at different obstacle arrangements: (a) random arrangement, (b) equidistant obstacle array, (c) continuous obstacle wall. In all cases $\tau_{0K} = 25.3$MPa, $\tau_{ext} = 23$MPa, $T = 300$K. The obstacle wall stands for an infinite density of infinitely weak obstacles ($L_{cc} \to 0$, $F_{max} \to 0$). Note that the y-direction is stretched by a factor 10 for improved visibility.

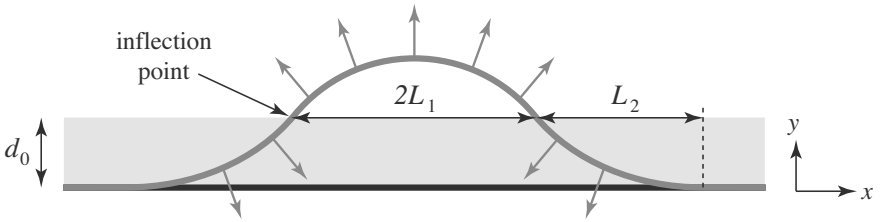


Figure 40. Unstable (grey) and stable (black) equilibrium configuration (T=0) of a dislocation at an obstacle wall (light grey).

$y(x)$. Approximating the curvature $(1/R_c)$ by the derivative $\mathrm{d}^2y/\mathrm{d}x^2$ yields the differential equation

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = -\frac{b}{S}\left(\tau_{\mathrm{wall}}(y) + \tau_{\mathrm{ext}}\right) \tag{83}$$

for $y(x)$. The quoted approximation is justified because the critical configurations are known to be flat (figure 39). The function $\tau_{\mathrm{wall}}(y)$ is described by the wall's maximum height, $\tau_{0\mathrm{K}}$, and by a function describing its 'shape' in y-direction. The latter contains a depth parameter d_0 which is related to the obstacle diameter. For simplicity a rectangular wall shape is assumed here; in (Mohles, 1997) a parabolic shape has also been examined with essentially the same results.

$$\tau_{\mathrm{wall}}(y) = \begin{cases} -\tau_{0\mathrm{K}} & \text{if } 0 < y \le d_0 \\ 0 & \text{otherwise} \end{cases} \tag{84}$$

With the simplified wall of equation (84), the solution $y(x)$ of differential equation (83) can be found easily by assembling dislocation pieces of constant curvature, as sketched in figure 40. A trivial solution is $y = 0$, where $|\tau_{\mathrm{wall}}| = \tau_{\mathrm{ext}}$ by definition. This describes the stable configuration. In the unstable configuration of figure 40, the curvature over the length $2L_1$ is $K_1 = b\tau_{\mathrm{ext}}/S$ (i.e., $K_1 < 0$), and in the two pieces of L_2, the curvature is $K_2 = b\left(\tau_{0\mathrm{K}} - \tau_{\mathrm{ext}}\right)/S$ (i.e., $K_2 > 0$). The length L_2 is defined geometrically by K_2 and d_0; with the quoted parabolic approximation $L_2 = \left(2d_0/K_2\right)^{1/2}$. In the inflection points the slopes have the magnitude $|m| = K_2L_2 = K_1L_1$, which finally defines L_1. Altogether this yields

$$L_1 = \sqrt{\frac{2d_0S\left(\tau_{0\mathrm{K}} - \tau_{\mathrm{ext}}\right)}{b\ \tau_{\mathrm{ext}}^2}} \quad \text{and} \quad L_2 = \sqrt{\frac{2d_0S}{b\ \left(\tau_{0\mathrm{K}} - \tau_{\mathrm{ext}}\right)}} \tag{85}$$

With the geometry $y(x)$ of the unstable configuration known, the energy ΔG can be calculated. ΔG is required to bring the dislocation from the stable to the unstable configuration. ΔG is comprised of two contributions. One is given by the obstacle stress between the configurations, reduced by the energy that is expended by the external stress. This part is found by integrating $(b\tau_{\mathrm{wall}} - b\tau_{\mathrm{ext}})$ in the area between the unstable and the stable configuration; it is large if L_1 and L_2 are large. The other contribution is given by the increment of dislocation length $\left[\int \left(\mathrm{d}x^2 + \mathrm{d}y^2\right)^{1/2}\right]$ from the stable to the unstable configuration; this part is large when L_1 and L_2 are small. In the equilibrium configuration $y(x)$, however, these contributions have equal magnitudes, as can be verified. Together this yields

$$\Delta G = \frac{4\sqrt{2}}{3}\sqrt{b\ d_0^3\ S\ \tau_{0\mathrm{K}}}\frac{1}{\tau_{\mathrm{rel}}}\left(1-\tau_{\mathrm{rel}}\right)^{1/2} \tag{86}$$

where τ_{ext} has been expressed by the relative stress τ_{rel} introduced in equation (80). ΔG is the energy that must be expended by the thermal pulses for activation. The energy contribution expended for this by the external stress has already been accounted for in the first contribution outlined above. The first square root in equation (86) is a universal feature of the activation energy, independent of the shape of the obstacle wall. It emphasizes that the activation event is governed by both the integrated obstacle stress $\tau_{0\mathrm{K}}$, which equals the maximum of τ_{wall} from equation (82), and the line tension S. The dependence of ΔG on τ_{rel} and the pre-factor is defined by the wall shape. Other wall shapes have been discussed in (Mohles, 1997). Instead of a stress wall, a potential trough can also be considered. When the trough is V-shaped with a slope equivalent to τ_{0K}, equation (86) is recovered (Kocks, 1985).

With equation (86) and the general Boltzmann equation, the probability of one attempted activation event (one attack) to be successful is known. The attack frequency, ν_0, is known to be in the range 10^{10} to 10^{11}Hz (Mohles, 1997), (Granato et al., 1964). Hence the waiting time for activation is

$$t_{\mathrm{wait}} = \frac{1}{\nu_0}\ \exp\frac{\Delta G}{k_{\mathrm{B}}T} \tag{87}$$

with ΔG and its dependence on $\tau_{0\mathrm{K}}$, S, and d_0 defined by equation (86). The dependence on τ_{rel} remains to be refined. In particular, close to the top of the wall, i.e. for $\tau_{\mathrm{rel}} \to 1$, the rectangular shape is not a good approximation; however, for small obstacles like foreign atoms this is not so critical.

Solid solution strengthening In deformation experiments often a constant shear rate $\dot{\gamma}$ is enforced. With a density of mobile dislocations ρ_m defined by the specimen's deformation state this enforces a mean dislocation velocity, as described by Orowan's equation (13). When the mean free running path λ of dislocations is assumed to be constant, this means that t_wait and hence ΔG is given (equations (81), (87), ν_0=const., T=const.). Since τ_0K, S and d_0 are material parameters, this defines τ_rel (equation (86)) and τ_ext (equation (80)), which here is called the solute strengthening contribution τ_sol here. Resolving for $\tau_\mathrm{ext} = \tau_\mathrm{sol}$ yields

$$\tau_\mathrm{sol} = \tau_{0K} \frac{2}{1+\sqrt{1+T^2/T_0^2}} \quad \text{with} \quad T_0 = \frac{\sqrt{8bd_0^3 S\tau_\mathrm{0K}}}{3k_\mathrm{B}\ln\left(\lambda\nu_0 b\rho_\mathrm{m}/\dot{\gamma}\right)} \tag{88}$$

where T_0 has no particular meaning as a special temperature; it is just an abbreviation for the temperature scale. Altogether τ_sol depends on the solute concentration c_at, but this dependence only comes through $\tau_\mathrm{0K}\,(c_\mathrm{at})$. Calculation of the latter dependence is an independent problem.

When reasonable values for solute atoms are inserted in equation (88) (e.g. $d_0 = b$, b=0.25nm, $S = \mu b^2/2$, μ=50GPa, τ_0K=50MPa, $\ln(\lambda\nu_0 b\rho_\mathrm{m}/\dot{\gamma}) = 20$), τ_rel comes out to be about 1/3 at room temperature: these small obstacles are indeed overcome by thermal activation. On the other hand, small particles with a diameter of only $d_0 = 8b = 2$nm yield $\tau_\mathrm{rel} = 0.97$. This means that thermal pulses do not help (much) to overcome particles. As noted above, the assumption of a rectangular wall shape is not a good one for $\tau_\mathrm{rel} \to 1$, so this result is not accurate. However, it can be shown with a parabolic shape of $\tau_\mathrm{wall}(y)$ (Mohles, 1997), which is more appropriate for $\tau_\mathrm{rel} \to 1$, that the statement remains valid that the strengthening contribution even of rather small particles ($d_0 \geq 4$nm) is independent of temperature (no thermal activation).

Equation (88) shows that τ_sol depends on the strain rate only weakly, but that it decreases with increasing temperature considerably. However, even at high temperatures the strengthening effect of solutes does not vanish entirely. This is all in agreement with experimental findings (e.g. (Kocks, 1985)).

4 3-D aspects

When dislocations glide in their glide planes they will encounter other ones in other glide planes. These dislocations interact mutually as described by elasticity theory, but they are also subjected to a number of reactions and effects that are intrinsically 3-dimensional. They lead to specific patterns of dislocation arrangements. For instance in face centred cubic crystals during

tensile or compression tests, the dislocations form cells with high dislocation density in the cell walls and low density in the interiors. In contrast, fatigue tests with thousands of loading and unloading cycles tend to lead to alternating layers of high and low dislocation density. One aim of 3-dimensional dislocation dynamics simulations is to investigate the corresponding pattern formation and the associated work hardening effects. However, such investigations are still in their infancy due to the enormous computational effort of dislocation dynamics simulations. Here some fundamental aspects of 3-D dislocation motion are compiled, in particular those with implications on the implementation of dislocation dynamics simulations.

4.1 Non-elastic 3-D effects

The elastic interaction between dislocations has already been described in section 1.3 for straight dislocations and in section 2.2 for arbitrary configurations. The present section outlines additional effects which in particular include irreversible dislocation reactions.

Intersecting dislocations When dislocations on different glide planes cross each other they impose a shape change on each other. Several cases are to be distinguished depending on the directions of the line and Burgers vectors. In figure 41, four cases are plotted in which two initially straight dislocations cross each other at right angles. In each of these cases and in general, every dislocation acquires a piece of dislocation length (added local line vector) of the magnitude and direction of the other dislocation's Burgers vector. This is simply a consequence of the Burgers vector being a shift vector. In the cases (a) and (d), one of the two dislocations only gains or loses length of the other dislocation's Burgers vector, whereas in the other cases both dislocations acquire a jog or kink.

In figure 41(b), each edge dislocation gets a short piece of screw dislocation lying in the same glide plane. Such local character changes in the same glide plane are called kinks; they can accompany the dislocations' glide without resistance (planar Peierls stresses are disregarded here). The edge dislocation in figure 41(d) has also acquired a kink, and the screw dislocation in 41(c) also has a kink: the line vector of this kink lies in the screw dislocation's glide plane, which in the present case is defined by the line vector and the direction of motion (see section 1.1).

The upper dislocation in figure 41(a) and the edge dislocation in 41(c) have acquired a jog, i.e. a segment with a line vector $\underline{s}$ normal to its glide plane (see 41(a)). With their own line vectors but unchanged Burgers vectors, these jogs define their own glide planes $\underline{n}_{\mathrm{jog}}$. The glide directions of

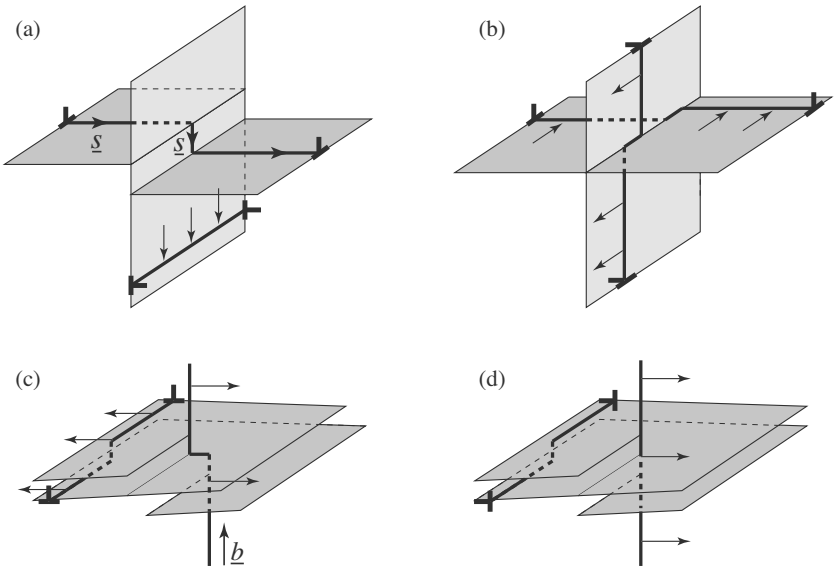


Figure 41. Straight dislocations after intersecting with another one normal to it; (a) two edges with normal Burgers vectors, (b) two edges with parallel Burgers vectors, (c) edge and screw with normal Burgers vectors, (d) edge and screw with parallel Burgers vectors.

the dislocations containing these jogs lies in the respective jog's glide planes $\underline{n}_{\text{jog}}$. This means that these jogs can, in principle, glide along with their parent dislocation's motion. If the glide plane $\underline{n}_{\text{jog}}$ is as "smooth" (section 1.4) as the dislocations' glide planes, the jogs will glide without additional resistance. However, this is not necessarily the case; additional Peierls stresses may apply for the jogs.

In figure 42 the intersection of two screw dislocations is considered; their line vectors (and consequently the Burgers vectors) are normal to each other. Both screws acquire jogs during the intersection. Unlike in the previous cases of figure 41, the jogs cannot glide with the rest of the dislocation because the direction of this motion does not lie in the glide plane of the jogs, as can be seen in figure 42(b). There the jogs may wait for approaching vacancies in order to climb in the indicated direction. This allows only for slow overall motion because the vacancy density is usually very low. On the other hand, at high stresses driving the dislocation forwards, the jogs may be dragged along, thereby emitting interstitial atoms. This can happen at

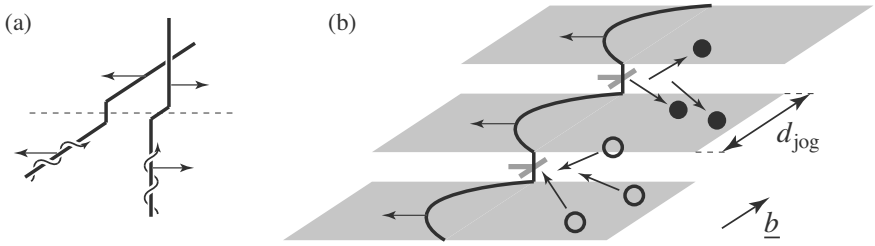


Figure 42. Straight screw dislocations after intersecting each other; (a) both screws acquire jogs; (b) the jogs must climb in order to follow the dislocation's motion by emitting or attracting interstitial atoms or vacancies. When the distance between the grey planes equals several atom layers, the jogs between them are called superjogs.

higher rates because the number of interstitials that can be generated and diffuse away is virtually unlimited. For motion in the opposite direction, or for jogs of opposite sign, the roles of vacancies and interstitials are exchanged; however, the energetics for vacancy and interstitial creation and mobility differ.

According to section 1.3 the critical stress τ_{jog} required for the dislocation to drag the jogs is $\tau_{\mathrm{jog}} = F_{\mathrm{jog}}/\left(bd_{\mathrm{jog}}\right)$, where d_{jog} is the average distance between them (figure 42(b)) and F_{jog} is the force to drag one jog. The latter is related to the energy $E_{\mathrm{v/i}}$ to create a vacancy/interstitial, which is expended while the jog climbs by the distance $d_{\mathrm{plane}} \approx b$ to the next plane. Assuming that the climb force is homogeneously distributed over the length d_{plane} such that $E_{\mathrm{v/i}} = d_{\mathrm{plane}} F_{\mathrm{jog}}$ yields

$$\tau_{\mathrm{jog}} = E_{\mathrm{v/i}}/\left(bd_{\mathrm{plane}} d_{\mathrm{jog}}\right) \tag{89}$$

For aluminium, for instance, the energy E_{v} to create a vacancy is about 0.67eV (Gottstein, 2004); the energy E_{i} of interstitial atoms is about two to four times higher. Even higher energies can occur when a screw dislocation is intersected by several other screws in the same glide plane. In that case, the length of the jog amounts to several times the Burgers vector. Such larger jogs are called superjogs; in order to move them with the gliding screw dislocation, several interstitials or vacancies must be created at once for every step of motion. Accordingly, the respective energy and the force of the superjog are larger.

In figure 42(b) it can be seen that the screw dislocation bows out between the jogs, like between all obstacles. The curvature of bending is limited by the obstacle force in relation to the line tension (see section 2.1), or by

the normalized obstacle strength $k = F_{\text{jog}}/(2S)$ of equation (74). With $S \approx \mu b^2/2$ and the values $\mu = 27\text{GPa}$, $b = 0.286\text{nm}$ and $E_{\text{v}} = 0.67\text{eV}$ for aluminium, k for vacancy creating jogs is found to be about 0.17; this means that the bending is weaker than in figure 31(a). Interstitials creating jogs or superjogs will cause stronger bending. In case of superjogs with $k \geq 1$ the dislocation will bend forwards like in the case of the Orowan process, but annihilation like in figure 18 will not occur (immediately) in the present case because the dislocation bows involved lie in parallel but different planes. Further details on jogs and superjogs can be found in (Hull and Bacon, 1992).

A possibly important effect of climbing jogs on plasticity is that this motion is thermally activated. This can introduce a strain rate and temperature dependence of plasticity in a similar way as solutes do (section 3.4). But the activation events are different and involve different activation energies; moreover the vacancy density and the jog density play important roles. Since these densities inevitably increase with continued plastic deformation as dislocations intersect each other and move on, work hardening can be affected in particular.

Dislocation reactions Dislocations can be subjected to exothermal reactions with other dislocations. A trivial example for this is when two dislocations with opposite sign meet in the same glide plane: they will annihilate each other irreversibly, like in the case of the Orowan process. Even when they are not in the same glide plane but come very close (neighbouring parallel planes, for instance), they can annihilate spontaneously and thereby release a row of vacancies or interstitial atoms. These reactions lower the free energy of the system and have a softening effect because the dislocation density is lowered.

A dislocation reaction can also occur when two dislocations in different non-parallel glide planes meet and partially annihilate, thereby lowering their combined energy. A prominent example for this is the formation of a Lomer-Cottrell lock, in which two partial dislocations lying in different glide planes react, as outlined in figure 43: The reacting partials P2 and P3 have the Burgers vectors $a/6\,[\bar{1}2\bar{1}]$ and $a/6\,[1\bar{1}2]$, respectively. They have the same magnitude $b_{\text{p}} = b/3^{1/2}$ and, for the most part, point in opposite directions. The resulting dislocation after the reaction has the Burgers vector $a/6\,[011]$ with the length $b_{\text{LC}} = b/3$. The energy of this dislocation is much lower than the combined energy of the reacting partials ($E^* \sim b^2$, equations (6) and (8)), therefore, this reaction is favourable and almost irreversible. In a realistic network of curved dislocations the reaction will only occur over a finite length, as indicated in figure 43. High stresses on the partials P2 and

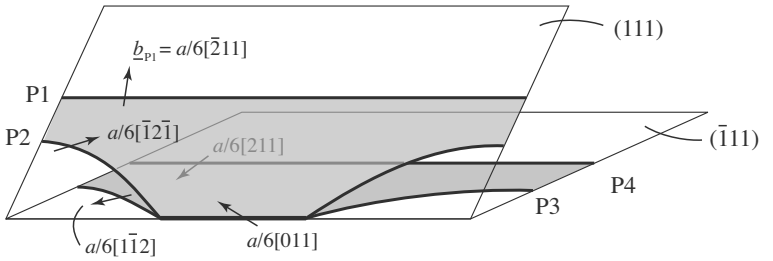


Figure 43. A Lomer-Cottrell-Lock formed locally by two Shockley partial dislocations from different planes. The length of the combined Burgers vector $a/6\,[011]$ is particularly small.

P3 may unzip this lock, but this is unlikely to happen because usually, the initial dissociated dislocations have been driven by an external stress such that they form the lock; hence essentially, a stress inversion is required to unzip this lock. The Burgers vector of the combined lock dislocation does not lie in either of the two initial glide planes. Therefore, this dislocation cannot glide with the original dislocations and hence locks them.

Another possible dislocation reaction can occur when dislocations intersect, like in figure 41(b). By the indicated motion of the two dislocations in this figure, a cube of material will be transformed into the shape of figure 44(a). Two alternatives to create the latter shape are indicated in figures 44(b) and (c). In figure 44(b), two straight edge dislocations with parallel Burgers vectors and normal line vectors move in the same direction; in 44(c) two angular dislocations move. The particular feature to be noticed here is that when the angular dislocations meet in their angles, this arrangement is indistinguishable from the case when the straight dislocations of figure 44(b) intersect. In this moment the configuration may switch from (b) to (c) or vice versa; this switch is called a collinear reaction. This reaction can be favourable for instance under certain stress conditions, or driven by dislocation length reductions which are possible in real cases where dislocations are actually curved.

A collinear reaction may also happen to the dislocations sketched in figure 41(b), which move in opposite directions and are therefore certain to intersect. The dislocations are driven in the indicated directions by given stresses, the origin of which plays no role. After the reaction occurs, the result is that of figure 44(d): for each of the angular dislocations created, the driving stresses on the two arms are opposing each other. At least in part this impedes further glide of the angular dislocations. Moreover, their

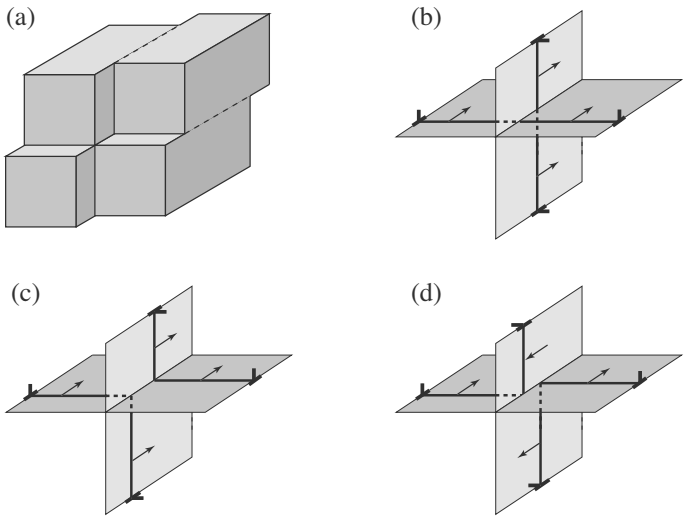


Figure 44. The shape in (a) is created from a cube by the dislocation motion in figure 41(b), or by the combined motion of the straight or angular dislocations in (b) or (c). A switch between the configurations (b) and (c) is possible when the two dislocations intersect. Likewise, (d) can be the result of a switch when the dislocations of figure 41(b) intersect. The resulting angular dislocations are essentially immobile.

glide can no longer lead to the shape of figure 44(a). Similar to the case of dislocation dipoles (section 1.3), angular dislocations are immobile or at least less mobile, so that new mobile dislocations must be created for continued plasticity. From this, a rise in material strength is to be expected, increasingly so with the number of dislocation intersections. These, in turn, increase with plastic strain. Dislocation dynamics simulations (Devincre et al., 2005) have approved that the collinear reaction in fact makes a significant contribution to work hardening.

Non-planar dislocations The angular dislocations of the previous section generated during dislocation intersections are one example of non-planar dislocations: each dislocation arm is curved in its own glide plane, and two planes are involved. Another case of a non-planar dislocation is given when a dislocation leaves its primary glide plane over a certain length. Here, like in section 1.3, a distinction between climb and cross-slip must be made: a piece of dislocation with screw character may glide into another

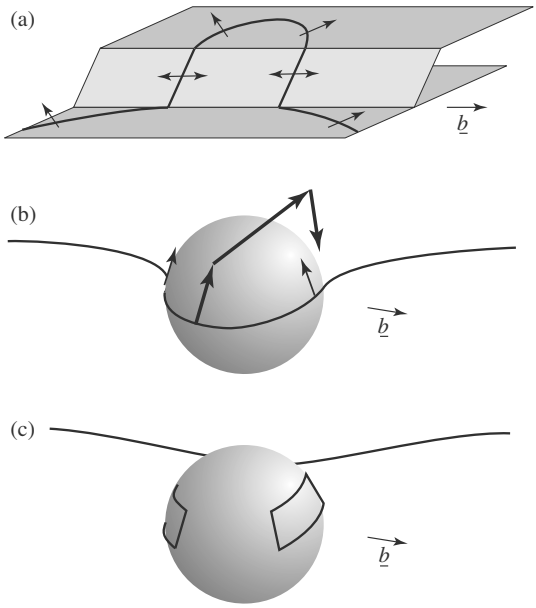


Figure 45. (a) Double cross slip of a piece of screw dislocation, arrows indicate glide directions; (b) a dislocation overcoming a particle by a combination of cross slip (big arrows) and climb (smaller arrows); (c) when a particle is overcome by cross slip alone, two prismatic dislocation loops of opposite signs are left behind, which may vanish later on by diffusion along the particle interface.

glide plane; however, a piece of edge dislocation can only climb. The latter involves material transport (diffusion) and is, therefore, slow.

In figure 45(a) a case of double cross slip is sketched: a piece of dislocation changes its glide plane twice. Changing the glide plane involves an angle to be created on the dislocation, which is energetically unfavourable. Hence there must a driving force, like from an external stress, for cross slip to occur. In contrast, climb may occur when there is an abundance of vacancies: they are attracted by edge dislocations, thereby making them climb.

Since dislocations are curved and have mixed character in general, cross slip and climb will have to occur concurrently. This means for instance that a driving force for cross slip will at the same time act as a driving force for the generation or absorption of vacancies or interstitials: the line ten-

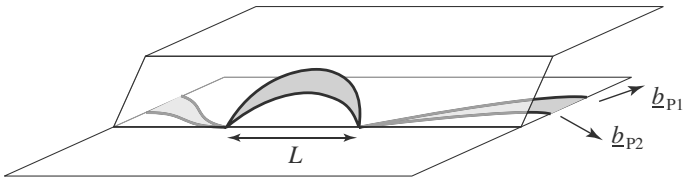


Figure 46. Cross slip of a dissociated screw dislocation. The grey area indicates the stacking fault.

sion will pull on the edge parts of the dislocation. In figure 45(b) a case is sketched where a particle is about to be overcome by a dislocation. This particle cannot be sheared. Hence it may be overcome by the Orowan process (section 2.1), or by a combination of cross slip and climb (large/smaller arrows in figure 45(b)). When cross slip and climb can take place at equal velocities, the dislocation can locally leave its primary glide plane near the particle and move back to this plane beyond it. On the other hand, when the edge components of the dislocation are unable to climb (for instance due to low temperature), the dislocation may overcome the particle by creating two prismatic dislocation loops, as sketched in figure 45(c). The prismatic loops can vanish later on by diffusing vacancies or interstitials, while the dislocation moves on in its primary glide plane. The details of what actually happens depends on a number of parameters like all stress components and the energetics and densities of vacancies and possibly interstitial atoms. Here the term "energetics" summarizes all activation energies for the creation and migration of vacancies and interstitials.

A dissociated screw dislocation cannot cross slip directly. One model for cross slip in this case is that the corresponding partial dislocations recombine to form a perfect dislocation over a certain length L before this piece of dislocation glides into the secondary plane. There it may dissociate again, as indicated in figure 46. The recombination of the two partials is a thermally activated event similar to that described in section 3.4 for overcoming obstacle walls. This explains the fact that cross slip is observed especially at high temperatures. In materials with a high stacking fault energy, like aluminium, the distance between the partials is very low (equation (23)). This means that the corresponding activation enthalpy in these materials is low, and that cross slip is very frequent even at low temperatures. This is in fact observed in aluminium alloys. However, it should also be noted that molecular dynamics simulations in literature have demonstrated the possibillity that a dissociated screw dislocation can cross slip without the recombination of the partials. In that case, cross slip must be accompanied

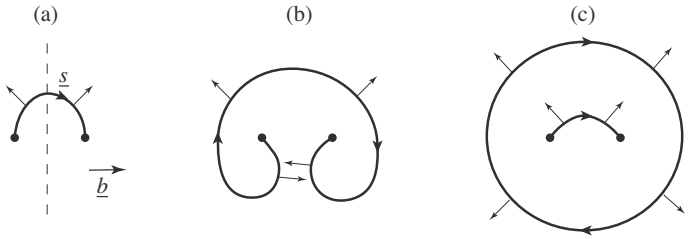


Figure 47. Operation of a Frank-Read dislocation source. The points indicate where the dislocation enters from another plane. In (b), two dislocation parts with edge character and opposite sign are about to annihilate each other; in (c) an expanding loop has been generated, and the new source is reestablished. The dashed line in (a) indicates a potential surface of the material.

by climb of the edge components of the partials.

Dislocation sources A very important aspect of cross slip is that it generates a Frank-Read dislocation source that can generate a virtually infinite number of new dislocations. Figure 47 represents the upper plane of figure 45(a), where double cross slip had just occurred. Under a given stress the dislocation segment bows out forwards between the points in which it enters this plane. When a critical stress is reached, the dislocation can move on to form the configuration of figure 47(b). Then, two dislocation parts with edge character and opposite sign annihilate each other (two half planes of atoms combine to form a full plane). The result is that of figure 47(c): an expanding dislocation loop enabling plastic deformation and a new segment, which can act as a dislocation source again.

The aforementioned critical stress is similar to that for the Orowan process: it depends on the distance between the fixed points, like τ_{Orowan} of equation (26) depends on the free space $L_{\mathrm{cc}} - 2R$ between obstacles. However, this is only an approximation because the dislocation parts outside the plane of figure 47 also cause significant stresses. Moreover, once the source has created a loop or several ones, these loops shield the external stress to some degree: the upper half of the loop in figure 47(c) repels the segment in its centre, pushing it downwards, and the lower part of the loop attracts the segment, also pulling it downwards. This means that only a finite number of loops can be generated at a given external stress, before the source "dries up". In dislocation dynamics simulations (as of yet, unpublished) it has been shown that this shielding effect is essentially independent of the size

of the (expanding) loops. For continued plastic deformation, the external stress must be increased further. Obviously this shielding effect contributes to work hardening.

When a dislocation changes its glide plane in a single point and ends on a free surface, this point can act as a spiral dislocation source. This is indicated in figure 47(a) by the dashed line: if this line describes a free surface and only the right half, for instance, of the figure is actually present, then the remaining dislocation arm can spiral around the fixed point and create half dislocation loops extending to the right into the material, just as if a mirrored counterpart were present on the left side (see section 1.3). The single fixed point may stem from half a double cross slip event (for instance, only the right half of figure 45(a)) or it may be one ending point of a superjog (see figure 42(b)).

4.2 Computational Aspects for 3-D Simulations

Some aspects of 2-D dislocation dynamics simulations, like segmentation, have been covered in section 3.1. They remain valid for 3-D simulations, but additional considerations are necessary in 3-D. For instance, decisions must be made on which glide planes are active, how cross slip is activated, and if or how climb is enabled. Generally speaking, the local direction of motion must be chosen. The subsequent considerations originate from the present author. However, there are a number of other successful approaches in literature, all with specific characteristics, see e.g. (Zbib et al., 2004), (Zbib et al., 2000a), (Zbib et al., 2000b), (Schwarz, 1999), (Devincre et al., 2001), (Devincre and Condat, 1992), (von Blanckenhagen et al., 2004), (von Blanckenhagen and Gumbsch, 2004), (Hartmaier and Gumbsch, 2004) (Ghoniem, 2000).

Direction of motion The direction of motion is unequivocal in 2-D (normal to the line vector, figure 28), but in 3-D, the glide plane must be chosen. This can be done by using predefined glide planes, depending on the crystal structure, and switching between these planes randomly with certain probabilities. These probabilities depend on temperature and on the resolved shear stresses in the glide planes of choice. But they also depend on the local dislocation configuration: a certain length of screw dislocation is required for cross-slip. But since dislocations are curved in general, it remains difficult to define such a critical length. In case of dissociated dislocations, the partial dislocations need to be merged into a perfect one over a certain length L in order or cross slip, as indicated in figure 46. This is known to

occur in a thermally activated manner, similar to the case of a dislocation overcoming an obstacle wall in figure 39. The merged length L may be used as a parameter to formulate the cross slip probability. But as seen later on, the merged length is not well-defined either.

An alternative to the approach above with predefined glide planes is to allow the simulated dislocations to move in any direction normal to its line, in principle, and to confine the motion to the preferred directions by other means. For instance, Xiang and Srolovitz (Xiang and Srolovitz, 2006) have applied the level-set method to simulate a dislocation overcoming a second phase particle. The authors used a strongly reduced dislocation mobility (a large drag factor B) for climb directions in order to account for the slowness of climb relative to slip.

Another concept using dislocations allowed to move in any direction is suggested subsequently. In this concept, a 3-D Peierls potential is used to make the dislocations prefer certain directions. It involves the symmetry and periodicity of the crystal structure under consideration. A dislocation tends to remain in low energy states within this potential; minor deviations cause stresses that drive the dislocation back. But larger deviations, caused for instance by thermal pulses, can make the dislocation leave its primary glide plane. In figure 48(a) this concept is sketched. In every point equivalent to that indicated by "0", the dislocation is free to move into the directions indicated by A, B, or C, and it will do so with the probabilities $P_A > P_B > P_C$ when the arrow in figure 48 indicates the Peach-Koehler force of equation (18).

In general, a dislocation would be curved in its glide plane, and it may even have kinks as indicated in figure 48(b). There, a piece of the dislocation (grey) has left its primary glide plane and moved into a parallel one. This motion involved double cross slip and, at the same time, climb. This means that a number of vacancies (or possibly interstitial atoms) have been created or consumed. In the concept of the 3-D Peierls potential, this means that also the energy for vacancy formation and consumption is to be considered for every geometrically necessary vacancy. The strength of the Peierls potential defines the degree of confinement.

Dislocation merging An important difference between 2-D and 3-D dislocation dynamics simulations is that there are non-trivial dislocation reactions possible in 3-D. Examples are the lock formation in figure 43 and the switch between the configurations of figures 44(b) and (c). In such reactions, dislocations can touch or intersect each other, which implicates numerical problems due to the stress singularity in the dislocations' cores. In order to avoid such singularities, all stress components defined by stress formulas

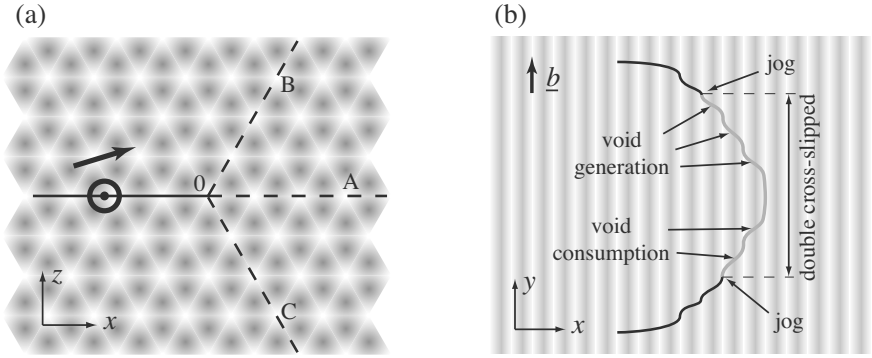


Figure 48. Concept of a 3-D Peierls potential to confine a dislocation capable of moving in any direction to glide planes. Darker shades of grey indicate a higher potential energy. (a) Straight screw dislocation along the y-axis; (b) a curved dislocation (with kinks), a piece of which (grey) has moved into a parallel glide plane.

like (49) or (58) can be multiplied by a function $f_{\mathrm{core}}(r)$ like

$$f_{\mathrm{core}}(r) = \frac{r^2}{r^2 + r_{\mathrm{core}}^2} \tag{90}$$

which accounts for the atomistic nature of the dislocation. Here, r is the distance from the dislocation and $r_{\mathrm{core}} \approx b$ is an adjustable parameter for the core radius. Equation (90) ensures that all stress components vanish in the dislocation core. Applying this function is essentially equivalent to using Peierls' dislocation model (Peierls, 1940) instead of Volterra's purely elastic approach (Volterra, 1907), on which all previous functions are based.

Besides avoiding the stress singularity in the code, a decision must be made for dislocation dynamics simulations on how to handle the joined dislocation part, like the Lomer-Cottrell lock (figure 43). One possibility is to ignore the joining, such that the dislocations coincide as indicated in figure 49(a). As a result of $f_{\mathrm{core}} \sim r^2$ for $r \ll r_{\mathrm{core}}$, and $\tau \sim r^{-1}$ for all elastic stresses, actually a linearly increasing interaction stress proportional to r would act inside the cores, which keeps attracting dislocations close together in a numerically stable way. This approach allows the simulations to automatically decompose the lock back into the two partial dislocations, provided that strong stresses on the side arms P2 and P3 in figure 43 tear them apart. This approach is convenient and useful if the dislocations that have merged are distinguishable in general. In the case of the Lomer-Cottrell

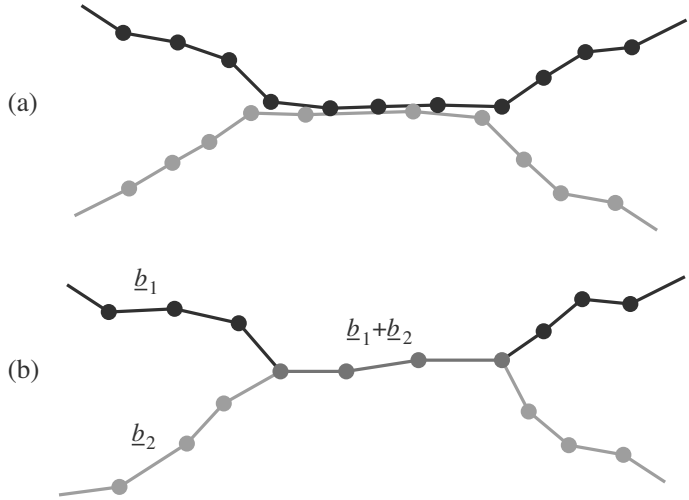


Figure 49. Merging of two dislocations (a), possibly in different glide planes, by the introduction of triple junctions (b). For the reverse effect from (b) to (a), the triple junctions may move towards each other until a quadruple junction is formed. If the Burgers vectors $\underline{b}_1$ and $\underline{b}_2$ differ, the respective dislocations are separated.

lock the distinction can always be made because the Burgers vectors of the respective partials differ.

On the other hand, dislocation reactions may also occur between dislocations with equal Burgers vectors, like in the case discussed in section 4.1 with regard to figures 44(b) and (c). There may even exist yet unidentified kinds of dislocation reactions. In such cases, ignoring the ambiguity of the dislocation configurations after intersections or after merging can lead to possibly severe errors in continued simulated plasticity. Therefore, in general, any possible cross-linking of dislocations should be considered. This can be done for instance by allowing dislocations to fully merge in one or several calculation nodes, i.e. in dislocation pieces with combined Burgers vector, as indicated in figure 49(b). Merging may start in a single node when two nodes from different dislocations come very close. This works for attracting as well as for repelling dislocations which are driven towards each other by external stresses. Starting from the first merged node (quadruple junction), its neighbour nodes will (or will not) attract each other and extend the merging to a piece of dislocation with two triple junctions at its ends. The reverse effect will be accomplished by the triple junctions mov-

ing towards each other until a quadruple junction is formed. When this happens, the junction can be resolved, and the four adjacent dislocation segments can be rejoined to two full dislocations as the Burgers vectors allow. In the case that all four Burgers vectors are equal ($\underline{b}_1 = \underline{b}_2$) in figure 4.9(b)), a decision is required on which dislocation segments are to be joined, depending on the local stresses.

Calculation effort In dislocation dynamics simulations (2-D or 3-D), the largest numerical effort lies in the calculation of the elastic interaction stresses between dislocation segments. Roughly spoken, every segment interacts with every other segment, which leads to an effort proportional to N^2. In simulations, the number of interaction partners must be reduced to those within a certain cut-off distance d_{cut} in order to keep the effort low. However, d_{cut} must be pretty large because the dislocation stress decreases only as $1/d_{\mathrm{cut}}$. For N segments with M interaction partners each, this leads to a calculation effort per integration step proportional to $N \cdot M$, where N and M can easily exceed 10000. Assuming that one segment-segment interaction of a single time step takes one microsecond to compute, it becomes obvious that the simulations are slow, given that at least tens of thousands of time steps are required for a meaningful simulation. Therefore, when aiming for dislocation dynamics simulations, measures to reduce the numerical effort need to be planned from the start.

Three measures used by Mohles (Mohles, 2001c) have proved to be quite successful. Firstly, it should be kept on mind that the calculations will implicate several small error contributions. Examples are the error which is introduced by a finite cut-off radius d_{cut}, or the errors involved with the spatial resolution Δs_i and with the finite stress increments $\Delta\tau$ as discussed in section 3.1. All error contributions have to be negligible. Reducing the error limits involved with one special respect often means dramatically increased calculation times. The overall error may still remain large as a result of other error influences. Hence calculation time would be wasted; it is, therefore, extremely useful to have equal error limits for all error influences involved.

Another measure to reduce the calculation effort is to introduce the possibility of local equilibria. Often, only parts of the dislocations move fast while the rest is hardly moving. This latter part can be put to sleep until something happens that might destroy the local equilibrium. The sleeping parts still act as interaction partners, but they do no stress evaluations on their own. When the external stress is increased, or when the stresses from other dislocation parts in the vicinity changes due to motion, the sleeping parts should be "awakened" to newly evaluate whether the state of local

equilibrium is still valid. If so, these parts can be put back to sleep, and otherwise they start moving. The definition of such local equilibria introduces another error contribution, the amount of which should again be kept as small or large as the other ones.

Thirdly, it can be exploited that the interaction stress of distant dislocation segments is low. Like in molecular dynamics simulations, the numerical effort can be reduced to $N \cdot M(d_{\mathrm{cut}})$ where $M(d_{\mathrm{cut}})$ is the number of interaction partners within a cut-off distance d_{cut}, beyond which, all contributions from interaction partners can be neglected. To find the relevant interaction partners fast, the cell method can be applied. This consists of two steps both the efforts of which scale with N^1: Firstly all N objects (dislocation segments or atoms in the case of molecular dynamics) tell "the space" where they are located; this space memorizes these locations in a discrete way, for instance by means of a 2-D or 3-D array of pointers or indices representing subspaces (cells). The array covers the whole simulated space, and the pointers or indices point to the corresponding objects. Secondly, all objects can ask "the space" which other objects are close ($d < d_{\mathrm{cut}}$) by examining the neighbouring subspaces. Hence the neighbour objects are found quickly with an effort that scales optimally with N^1.

In the case of molecular dynamics, $M(d_{\mathrm{cut}} \ll N)$, so that the numerical effort is lowered strongly. But in the case of dislocation segments, the condition $M \ll N$ it is not quite sufficient because the interaction is long-ranged: $\tau \sim d^{-1}$. This means that d_{cut} must be chosen pretty large so that the number $M(d_{\mathrm{cut}})$ also gets large, $M \approx N$. Hence the efficiency gain from the cell method alone is limited.

However, the cell method can be refined by taking basic physical principles into account. For instance, the interaction between very distant segments can change only slowly in time, depending on their velocities. Hence the interaction stress of the more distant segments needs to be calculated less often. This principle can be utilized most efficiently by introducing multiple shells of interaction distances around segments. Each segment should remember the interaction stresses of all segments which are located in certain shells (distance intervals) around it, separately for each shell. Hence, every segment can re-use the stress values it remembers instead of recalculating them. Since the outer, larger shells containing many interaction partners collects only very distant partners, the high calculation effort for this shell is required only infrequently.

With all the above measures to reduce the numerical effort combined, discrete dislocation dynamics simulations are speeded up by several orders of magnitude, increasingly so for larger simulated volumes.

Bibliography

R.J. Arsenault, S. Patu, and D.M. Esterling. Computer simulation of solid solution strengthening in fcc alloys: Part i. friedel and mott limits. *Met. Trans. A.*, 20A:1411–1418, 1989a.

R.J. Arsenault, S. Patu, and D.M. Esterling. Computer simulation of solid solution strengthening in fcc alloys: Part ii. at absolute zero temperature. *Met. Trans. A.*, 20A:1419–1428, 1989b.

D.J. Bacon. A method for describing a flexible dislocation. *Phys. Stat. Sol.*, 23:527–538, 1967.

L.M. Brown. The self-stress of dislocations and the shape of extended nodes. *Phil. Mag.*, 10:441–466, 1964.

L.M. Brown. A proof of lothes theorem. *Phil. Mag.*, 15:363, 1967.

B. Devincre. Three dimensional stress field expressions for straight dislocation segments. *Solid State Comm.*, 93:875–878, 1995.

B. Devincre and M. Condat. Model validation of a 3d simulation of dislocation dynamics: discretization and line tension effects. *Acta metall. mater.*, 40:2629–2637, 1992.

B. Devincre, L.P. Kubin, C. Lemarchand, and R. Madec. Mesoscopic simulations of plastic deformation. *Mat. Sci. Eng.*, 309-310:211–219, 2001.

B. Devincre, T. Hocb, and L P. Kubin. Collinear interactions of dislocations and slip systems. *Mat. Sci. Eng. A*, 400-401:182–185, 2005.

M.S. Duesbery, N.P. Louat, and K. Sadananda. The numerical simulation of continuum dislocations. *Phil. Mag.*, A65:311–325, 1992.

J.D. Eshelby. The continuum theory of lattice defects. In F. Seitz and D. Turnbull, editors, *Solid State Physics*, page 79. Academic Press Inc., New York, Vol.3, 1956.

A.J.E. Foreman and M.J. Makin. Dislocation movement through random arrays of obstacles. *Phil. Mag.*, 14:911–924, 1966.

A.J.E. Foreman and M.J. Makin. Dislocation movement through random arrays of obstacles. *Can. J. Phys.*, 45:511–517, 1967.

J. Friedel. *Dislocations*. Pergamon Press, 1964.

N.M. Ghoniem. Computational methods for mesoscopic, inhomogeneous plastic deformation. In D. Walgraef, C. Worner, and J. Martinez-Mardones, editors, *Materials Instabilities*, pages 76–158. World Scientific, Singapore, 2000.

G. Gottstein. *Physical Foundations of Masterials Science*. Springer-Verlag, 2004.

A.G. Granato, K. Luecke, J. Schlipf, and L.J. Teutonico. Entropy factors for thermally activated unpinning of dislocations. *J. Appl. Phys.*, 35: 2732–2745, 1964.

K. Hanson and J.W. Morris. Limiting configuration in dislocation glide through a random array of point obstacles. *J. Appl. Phys.*, 46:983, 1975.

A. Hartmaier and P. Gumbsch. Discrete dislocation dynamics simulation of crack-tip plasticity. In D. Raabe, F. Roters, F. Barlat, and L.-Q. Chen, editors, *Continuum Scale Simulation of Engineering Materials*, pages 413–427. Wiley-VCH, Weinheim, 2004.

J.P. Hirth and J. Lothe. *Theory of Dislocations*. Krieger Publishing Company, 1992.

D. Hull and D.J. Bacon. *Introduction to Dislocations*. Pergamon Press, 1992.

R.D. Isaac and V. Granato. Rate theory of dislocation motion: Thermal activation and inertial effects. *Phys. Rev. B.*, 37:9278, 1988.

K.M. Jassby and T. Vreeland. Dislocation mobility in pure copper at 4.2°k. *Phys. Rev.*, 8B:3537–3541, 1973.

U.F. Kocks. Kinetics of solution hardening. *Metall. Trans. A*, 16:2109–2129, 1985.

U.F. Kocks, A.S. Argon, and M.F. Ashby. Thermodynamics and kinetics of slip. *Prog. Mat. Sci.*, 19:1, 1975.

R. Labusch and R.B. Schwarz. Simulation of thermally activated dislocation motion in alloys. In *Proc. of ICSMA 9*, pages 47–69. 1992.

I.M. Lifshitz and V.V. Slyozov. The kinetics of precipitation from supersaturated solid solutions. *Phys. Chem. Solids*, 19:35–50, 1961.

V. Mohles. *Thermisch aktivierte Versetzungsbewegung in Kristallen auf der Grundlage von Simulationsrechnungen*. Shaker Verlag, 1997.

V. Mohles. The critical resolved shear stress of single crystals with long-range ordered precipitates calculated by dislocation dynamics simulations. *Mat. Sci. Eng. A*, 365:144–150, 2004.

V. Mohles. Simulation of particle strengthening: Lattice mismatch strengthening. *Mat. Sci. Eng. A*, 319-321:201–205, 2001a.

V. Mohles. Simulation of particle strengthening: the effects of the dislocation dissociation on lattice mismatch strengthening. *Mat. Sci. Eng. A*, 319-321:206–210, 2001b.

V. Mohles. Computer simulations of the glide of dissociated dislocations in lattice mismatch strengthened materials. *Mat. Sci. Eng. A*, 324:190–195, 2002.

V. Mohles. Superposition of dispersion strengthening and size mismatch stengthening: computer simulations. *Phil. Mag.*, 83:9–19, 2003.

V. Mohles. Simulation of dislocation glide in overaged precipitation-hardened crystals. *Phil. Mag. A*, 81:971–990, 2001c.

V. Mohles and B. Fruhstorfer. Computer simulations of orowan process controlled dislocation glide in particle arrangements of various randomness. *Acta. Mat.*, 50:2503–2516, 2002.

V. Mohles and E. Nembach. The peak- and overstaged states of particle strengthened materials: computer simulations. *Acta. Mat.*, 49:2405–2417, 2001.

V. Mohles and D. Rönnpagel. Thermal activation analysis of dislocations in obstacle fields. *Comp. Mat. Sci.*, 7:98–102, 1996.

G. Monnet. Investigation of precipitation hardening by dislocation dynamics simulations. *Phil. Mag.*, 86:5927–5941, 2006.

N.F. Mott and F.R.N. Nabarro. *Creep and plastic flow.* The Physical Society, London, 1948.

E. Nadgornyi. Dislocation dynamics and mechanical properties of crystals. *Progr. Mat. Sci.*, 31:1, 1988.

E. Nembach. *Particle Strengthening of Metals and Alloys.* John Wiley, 1996.

E. Orowan. über den mechanismus des gleitvorganges. *Z. Phys.*, 89:634, 1934.

M. Peach and J.S. Koehler. The forces exerted on dislocations and the stress fields produced by them. *Phys. Rev.*, 80:436, 1950.

R. Peierls. The size of a dislocation. *Proc. Phys. Soc., London*, 52:34, 1940.

M. Polanyi. über eine art gitterstörung, die einen kristall plastisch machen könnte. *Z. Phys.*, 89:660, 1934.

W.H. Press, S.A. Teukolsky, V.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C.* University Press, 1992.

K.W. Schwarz. Simulation of dislocations on the mesoscopic scale. i. methods and examples. *J. Appl. Phys.*, 85:108–119, 1999.

R.B. Schwarz and R. Labusch. Dynamic simulation of solution hardening. *J. Appl. Phys.*, 49:5174–5187, 1978.

G.I. Taylor. *Proc. R. Soc., London*, A145:362, 1934.

V. Volterra. Sur lequilibre des corps elastiques multiplement connexes. *Ann. Ecole Norm. Super.*, 24:401–517, 1907.

B. von Blanckenhagen and P. Gumbsch. Discrete dislocation dynamics simulation of thin film plastcity. In D. Raabe, F. Roters, F. Barlat, and L.-Q. Chen, editors, *Continuum Scale Simulation of Engineering Materials*, pages 397–412. Wiley-VCH, Weinheim, 2004.

B. von Blanckenhagen, E. Arzt, and P. Gumbsch. Discrete dislocation simulation of plastic deformation in metal thin films. *Acta Metallurgica*, 52: 773–784, 2004.

C. Wagner. Theorie der alterung von niederschlägen durch umlösen. *Elektrochem.*, 65:581, 1961.

J. Weertman. *Elementary Dislocation Theory.* Oxford University Press, New York, 1992.

Y. Xiang and D.J. Srolovitz. Dislocation climb effects on particle bypass simulations. *Phil. Mag.*, 86:3937–3957, 2006.

H.M. Zbib, T.D. de La Rubia, M. Rhee, and J.P. Hirth. 3d dislocation dynamics: stress-strain behaviour and hardening mechanisms in fcc and bcc metals. *J. Nucl. Mater.*, 276:154–165, 2000a.

H.M. Zbib, M. Rhee, and J.P. Hirth. On plastic deformation and the dynamics of 3d dislocations. *Int. J. Mech. Sci.*, 40:113–127, 2000b.

H.M. Zbib, M. Hiratani, and M. Shehadeh. Multiscale discrete dislocation dynamics plasticity. In D. Raabe, F. Roters, F. Barlat, and L.-Q. Chen, editors, *Continuum Scale Simulation of Engineering Materials*, pages 201–229. Wiley-VCH, Weinheim, 2004.

Plasticity of moderately loaded cracks and the consequence of the discrete nature of plasticity to fatigue and fracture

R. Pippan † and H. Weinhandl † and H.G.M. Kreuzer †
† Erich Schmid Institute of Materials Science,
Austrian Academy of Sciences, Leoben, Austria

1 Introduction

Ideal brittle fracture dominates the failure of ceramics and semi-conductors at low and medium temperatures. The only irreversible process, in this case, is the breaking of atomic bonds.

During the fracture of metals, even in the most brittle ones, a certain amount of plasticity is involved. This plasticity is responsible for the broad variations in the fracture resistance of these types of materials. The discrete nature of plasticity, i.e. the dislocations, their movement and arrangement, plays a key role for many phenomena in this field. For example: the cleavage fracture of single crystals and polycrystals, the failure of interfaces and grain boundaries, the plasticity induced fracture of brittle particles, the formation of nano- and micro-cracks, the initiation of pores, the fatigue crack propagation, etc. These phenomena can be usually separated into two sub-problems – the evolution of the local stress field on the nanometer scale, where the fracture processes take place and the intrinsic fracture process itself, which causes the separation of the material. The local stresses on the nanometer scale are determined by the external loading and the internal sources, for example, the stresses induced by the thermal mismatch or the stresses induced by dislocations.

Hence, the main components for solving these problems are the stress fields of a crack and the stress field of dislocations in the presence of a crack. Real cracks can have a complex three-dimensional (3D) shape and a complex 3D dislocation arrangement. For this case the tools are available, however the calculation is very cumbersome and usually difficult to interpret. However, for the explanation of many phenomena the much simpler 2D solution is sufficient or can be used to estimate what may happen in the more complex 3D case.

In the following, a short introduction to and a discussion of the stress fields of a crack and the dislocations in the vicinity of the crack is presented. The second

part of this chapter is devoted to the deformation processes at the crack tip under cyclic loading, under moderate loading and their consequences to the fatigue crack propagation behaviour of metals. In fracture mechanics textbooks or text books dealing with fatigue and fracture, the phenomena are described by classical elasto-plastic continuum mechanics. Such description of the deformation in front of a crack is suitable, if the plastic deformed region is large compared to the characteristic dimensions of the dislocation substructure and when the considered volume elements are also larger than this characteristic dimensions. However, when the size of the plastic zone and the fracture controlling volume elements are smaller than this size the discrete nature of the dislocation has to be taken into account. In semi-brittle metals and alloys the size of the plastic zone, when the crack starts to propagate, is often smaller than these characteristic dimensions and the fracture process is often controlled by the stress fields at even smaller scales. Therefore the knowledge of the stress field of a crack as well as the stress field of the dislocation in the vicinity of the crack and the evolution of the dislocation structure are essential for the understanding of these processes.

For very low crack growth rates, the situation is very similar in all materials under fatigue, the size of the cyclically plastic deformed zone and the size of the dislocation substructure becomes smaller than the characteristic dimension of the microstructure . Furthermore the zone where the separation of the material occurs is even smaller, it is in the order of atomic spacing. Hence the stress field induced by the applied load and the dislocations in the imidiate vicinity of the crack tip are essential.

2 Stress field of a crack in a linear elastic material

There are many different possibilities to derive the stress field of a crack in a linear elastic body, for details see, for example (1–3). In this subsection, only the solution will be shortly discussed. The different types of loading a crack can be described by three basic modes. Those are tension perpendicular to the crack plane, shear in the crack plane perpendicular to the crack front and shear in the crack plane parallel to the crack front, they are schematically illustrated in Fig.1 and called mode I, mode II and mode III loading, respectively. Any arbitrary loading of a crack can then be ascribed by a combination of these basic loading modes. First, let us consider the complete stress fields for the three types of loading for a crack in an infinite plate, where the crack is located at the x_1 axis in the x_3-x_1 plane with the crack front parallel to the x_3 axis. In the coordinate system given in Fig.2, with the subscript L referring to the left hand crack tip, R to the right hand crack tip and C to the center of the crack, the stress field of a mode I crack in a linear

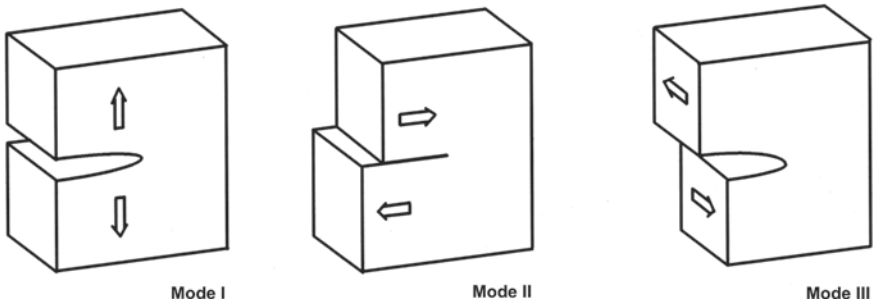


Figure 1: Schematic illustration of the three modes of loading a crack. In mode I, the opening mode, the loading is normal to the crack plane. In mode II, in-plane shear, the shear loading is in the plane of the crack normal to the crack front. In mode III, antiplane shear, the shear loading is in the plane of the crack along the crack front.

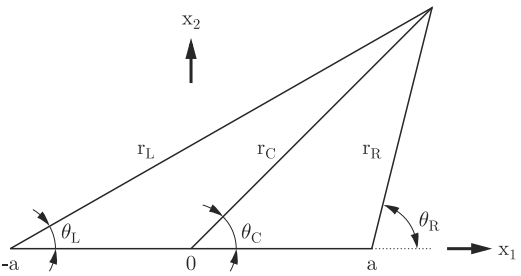


Figure 2: Coordinates for the description of the general stress field of a crack with length $2a$.

elastic isotropic solid is

$$
\begin{aligned}
\sigma_{11} &= -\sigma_A + \sigma_A \frac{r_C}{2r^*}\Big[2\cos(\Theta^* - \Theta_C) + 2\sin\Theta_C \sin\Theta^* \\
&- \sin\Theta_R \sin(\Theta^* + \Theta_R - \Theta_C)\} - \sin\Theta_L \sin(\Theta^* + \Theta_L - \Theta_C)\Big] \\
\sigma_{22} &= \sigma_A \frac{r_C}{2r^*}\Big[2\cos(\Theta^* - \Theta_C) - 2\sin\Theta_C \sin\Theta^* \qquad (1) \\
&+ \sin\Theta_R \sin(\Theta^* + \Theta_R - \Theta_C) + \sin\Theta_L \sin(\Theta^* + \Theta_L - \Theta_C)\Big] \\
\sigma_{12} &= \sigma_A \frac{r_C}{2r^*}\Big[\sin\Theta_R \cos(\Theta^* + \Theta_R - \Theta_C) \\
&+ \sin\Theta_L \cos(\Theta^* + \Theta_L - \Theta_C) - 2\sin\Theta_C \cos\Theta^*\Big].
\end{aligned}
$$

All other stresses are equal to zero, except

$$\sigma_{33} = \nu(\sigma_{11} + \sigma_{22} - \sigma_A). \tag{2}$$

The applied stress σ_A is the tensile stress σ_{22} at large distances, where

$$\sigma_{22} \to \sigma_A, \sigma_{33}, \sigma_{11}, \sigma_{12} \to 0 \tag{3}$$

as $r \to \infty$.

For the mode II loading

$$\begin{aligned}
\sigma_{11} &= -\sigma_{22} - 2\sigma_A \frac{r_C}{r^*} \sin(\Theta^* - \Theta_C) \\
\sigma_{22} &= \sigma_A \frac{r_C}{2r^*} \Big[\sin\Theta_R \cos(\Theta^* + \Theta_R - \Theta_C) \\
&+ \sin\Theta_L \cos(\Theta^* + \Theta_L - \Theta_C) - 2\sin\Theta_C \cos\Theta^* \Big] \\
\sigma_{12} &= \sigma_A \frac{r_C}{2r^*} \Big[2\cos(\Theta^* - \Theta_C) + 2\sin\Theta_C \sin\Theta^* \\
&- \sin\Theta_R \sin(\Theta^* + \Theta_R - \Theta_C) - \sin\Theta_L \sin(\Theta^* + \Theta_L - \Theta_C) \Big]
\end{aligned} \tag{4}$$

and all other stress components are equal to zero except the stress in the x_3 direction

$$\sigma_{33} = \nu\left(\sigma_{11} + \sigma_{22}\right). \tag{5}$$

At large distances from the crack $\sigma_{12} = \sigma_A$ and σ_{11} and $\sigma_{22} \to 0$.

For mode III loading the stress field is

$$\begin{aligned}
\sigma_{31} &= -\sigma_A \frac{r_C}{2r^*} \sin(\Theta^* - \Theta_C) \\
\sigma_{23} &= \sigma_A \frac{r_C}{2r^*} \cos(\Theta^* - \Theta_C).
\end{aligned} \tag{6}$$

In eq. (2), (5) and (7)

$$\begin{aligned}
r_c &= \sqrt{x_1^2 + x_2^2} \\
r_L &= \sqrt{(x_1 + a)^2 + x_2^2} \\
r_R &= \sqrt{(x_1 - a)^2 + x_2^2} \\
r^* &= \sqrt{r_L r_R}
\end{aligned} \tag{7}$$

$$\begin{aligned}
\Theta_C &= \arctan\left(\frac{x_2}{x_1}\right) \\
\Theta_L &= \arctan\left(\frac{x_2}{x_1+a}\right) \\
\Theta_R &= \arctan\left(\frac{x_2}{x_1-a}\right) \\
\Theta^* &= \frac{1}{2}(\Theta_L+\Theta_R)
\end{aligned} \tag{8}$$

and

$$-\pi \leq \{\Theta_C, \Theta_L, \Theta_R, \Theta^*\} \leq \pi. \tag{9}$$

These complete analytic expressions of the general stress fields are usually not given in textbooks related to fracture mechanics. They are presented in this comprehensive form in "Dislocation based fracture mechanics" (3). For the understanding of the effect which the size of the crack has on the crack affected region, these general expressions are very helpful.

The contour plots of the normalized maximum shear stress

$$\frac{\tau}{\sigma_A} = \sqrt{\sigma_{12}^2 + \frac{1}{4}\left(\sigma_{11}-\sigma_{22}\right)^2} \tag{10}$$

for the mode I, mode II and mode III crack are shown in Figs.3, 4, and 5, respectively. The plots are presented using two different scales, the first one gives an overview and the second one presents the stress field in the near crack tip region. In Fig.6 the contour plot of the normalized "pressure" term of the stress field

$$\frac{1}{2\sigma_A}(\sigma_{11}+\sigma_{22}) \tag{11}$$

is plotted for the mode I and mode II loading in order to give an impression of the difference in values of tension to shear stress components of the stress field. For the mode III loading this term is zero. As mentioned the applied stress σ_A is equal to the σ_{22}, σ_{12} and σ_{23} at large distances from the crack for the mode I, mode II and mode III loading, respectively. From the contour plots it is evident that stresses at distances in front of the crack larger than the length are not significantly affected by the presence of the crack. The maximum increase of the stress at a distance equal to the crack length is about 10% of the applied stress. Only at distances smaller than about 1/3 of the crack length the increase in the maximum stresses is larger than $2\sigma_A$. Near the crack tip the stress field can be expressed in a much simpler form, which one can expect from the contour plots in Figs.3-5. For

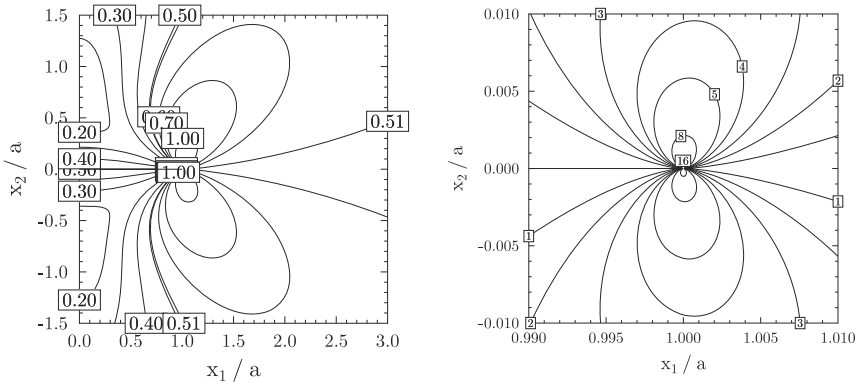


Figure 3: Contours τ/σ_A at large and small normalized distances x_1/a, x_2/a of a mode I crack.

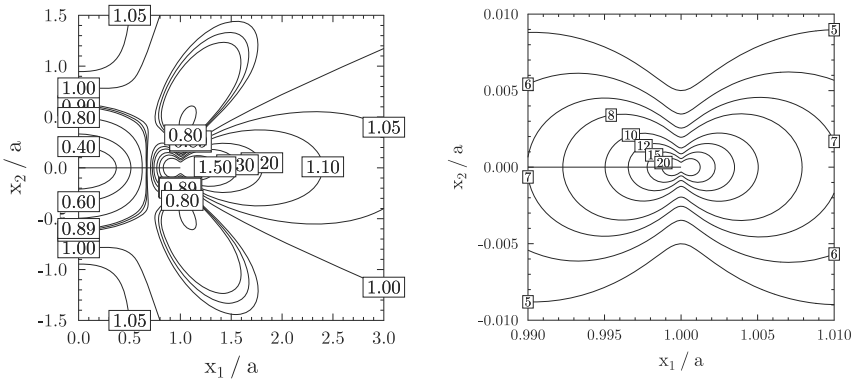


Figure 4: Contours τ/σ_A at large and small normalized distances x_1/a, x_2/a of a mode II crack.

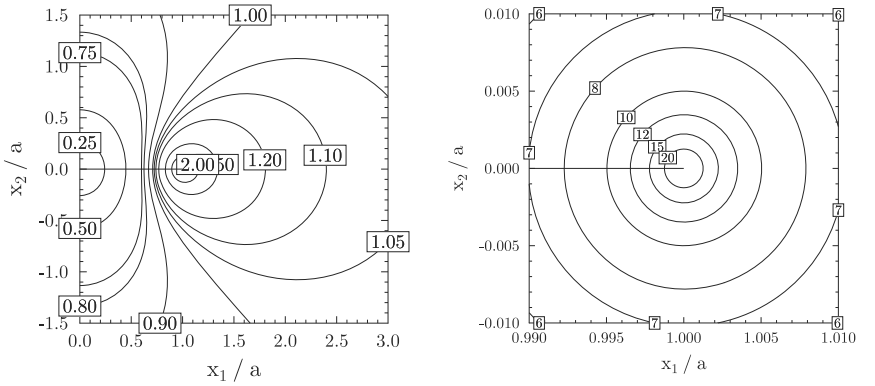


Figure 5: Contours τ/σ_A at large and small normalized distances x_1/a, x_2/a of a mode III crack.

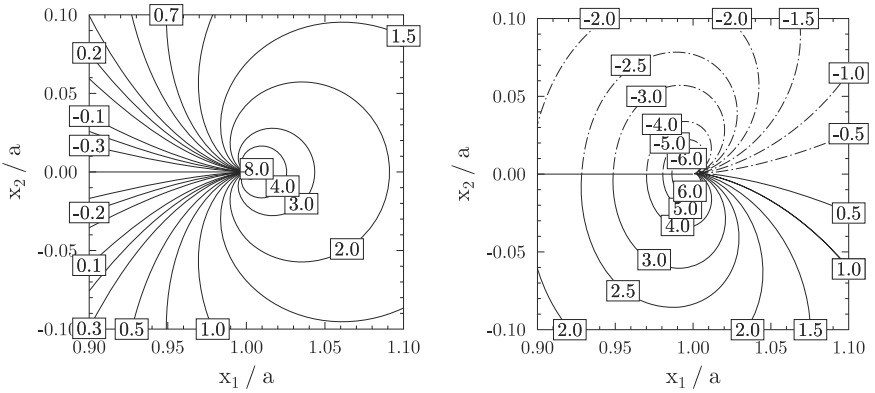


Figure 6: $\frac{1}{2}(\sigma_{11}+\sigma_{22})/\sigma_A$ contours for the near crack tip region of a mode I and mode II crack in (a) and (b), respectively.

the mode I loading

$$\begin{aligned}
\sigma_{11} &= \frac{K_I}{\sqrt{2\pi r}} \cos\frac{\Theta}{2}\left(1 - \sin\frac{\Theta}{2}\sin\frac{3\Theta}{2}\right) \\
\sigma_{22} &= \frac{K_I}{\sqrt{2\pi r}} \cos\frac{\Theta}{2}\left(1 + \sin\frac{\Theta}{2}\sin\frac{3\Theta}{2}\right) \\
\sigma_{12} &= \frac{K_I}{\sqrt{2\pi r}} \sin\frac{\Theta}{2}\cos\frac{\Theta}{2}\cos\frac{3\Theta}{2}
\end{aligned} \tag{12}$$

for mode II loading

$$\begin{aligned}
\sigma_{11} &= -\frac{K_{II}}{\sqrt{2\pi r}} \sin\frac{\Theta}{2}\left(2 + \cos\frac{\Theta}{2}\cos\frac{3\Theta}{2}\right) \\
\sigma_{22} &= \frac{K_{II}}{\sqrt{2\pi r}} \sin\frac{\Theta}{2}\cos\frac{\Theta}{2}\cos\frac{3\Theta}{2} \\
\sigma_{12} &= \frac{K_{II}}{\sqrt{2\pi r}} \cos\frac{\Theta}{2}\left(1 - \sin\frac{\Theta}{2}\sin\frac{3\Theta}{2}\right)
\end{aligned} \tag{13}$$

and for the mode III loading

$$\begin{aligned}
\sigma_{31} &= -\frac{K_{III}}{\sqrt{2\pi r}} \sin\frac{\Theta}{2} \\
\sigma_{23} &= \frac{K_{III}}{\sqrt{2\pi r}} \cos\frac{\Theta}{2}
\end{aligned} \tag{14}$$

where $r = r_R$ and $\Theta = \Theta_R$ in Fig.2. The terms K_I, K_{II} and K_{III} are given by $\sigma_A\sqrt{\pi a}$, where the applied stress σ_A is the far field σ_{22}, σ_{12} and σ_{23} for the mode I, mode II and mode III loading, respectively. The stress field near the tip is therefore characterized by a single loading parameter, K, which is called the stress intensity factor, a radius $(1/\sqrt{2\pi r})$ and an angular dependent term. A numerical comparison of the near tip expression with the general stress fields eq.(2, 5) and eq.(7) as well as a closer look to the contour plots show that the eqs.(13, 14) and (15) are a very good approximation for the crack tip stress fields at $r < 0.1a$. These equations for the near tip stress field are not only valid for a crack in an infinite sheet, they describe the near crack tip stress field in any elastic material, only K becomes geometry dependent. Usually K can be calculated by

$$K = \sigma_A\sqrt{\pi a}Y(a/W) \tag{15}$$

where $Y(a/W)$ is a dimensionless geometry parameter, which is expressed as a function of ratio, crack length devided by a characteristic size parameter of the considered geometry.

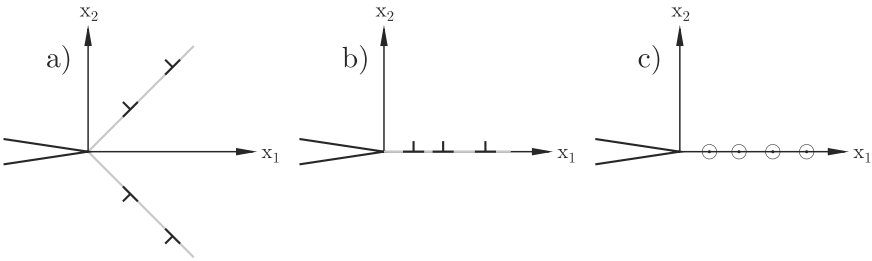


Figure 7: Schematic illustration of dislocation arrangement to realize a mode I (a), mode II (b) and mode III deformation of the crack tip (c)

From the scaling point of view for the consideration of cracks in a material it is important to repeat that the size of the K-dominated region in front of a crack scales with crack lenght, i.e., if the crack length is $1\mu m$ the K-dominated region is 100 nm or 200 nm, depending on the accuracy, if the crack length is 1 m, the K-dominated region is about 100 mm.

3 Dislocation crack tip interaction

Due to the very high stresses in the vicinity of a crack tip one expects a plastic relaxation even at relatively small loads. Such plastic deformations have to be realized by the generation and the movement of dislocations. The generation of dislocations is governed by the local stresses and the available sources. When the density of activiable sources are scare, the crack tip can also act as dislocation source. The movement of the dislocations is controlled by the Peach-Koehler force, which is determined by the character of the dislocation and the local stresses. Hence, the local elastic stresses play a central role in the plastic deformation. They are given by the superposition of the stresses induced by the applied loading and the stresses induced by the other dislocations. The stress fields caused by the applied loading in the presence of a crack are introduced in the previous section. Since the crack flanks are traction free, the stress field of dislocations in the vicinity of a crack differs from the stress field of a dislocation in an infinite body. Similar as in the case of the stress fields of a crack, we limit our consideration to the 2D-problem solely. In such 2D considerations the plastic deformation of the crack tip under mode I and mode II loading are realized by edge dislocations, whereas the mode III crack tip deformation are realized by screw dislocations as illustrated in Fig.7. In the following, this derivation for the stress field of an edge dislocation near a crack tip will be shortly introduced. For more details and more general solutions (not only the near crack tip expression) the reader is referred to the excellent book (3) and overview articles (2) and (6).

3.1 Linear elastic analysis of the stresses and deformations induced by an edge dislocation near a crack tip

Let us assume an isotropic, linear elastic, and unbounded body. The deformations u and stresses σ are desribed by two complex potentials $\phi(z)$ and $\omega(z)$ according to Kolosov as follows:

$$\begin{aligned} \sigma_{11}+\sigma_{22} &= 2\,(\phi'(z)+\overline{\phi'(z)}) \\ \sigma_{22}-i\sigma_{12} &= \phi'(z)+\overline{\omega'(z)}+(z-\bar{z})\phi''(z) \\ u=u_1+i\,u_2 &= \frac{1}{2\mu}(\kappa\phi(z)-(z-\bar{z})\overline{\phi'(z)}-\overline{\omega(z)})\;. \end{aligned} \tag{16}$$

Here $i=\sqrt{-1}$, $z=x_1+i\,x_2$ is an arbitrary point in the complex plane. A prime denotes a differentiation with respect to z and a bar lables the complex conjugate function. The elastic constants are the shear modulus, μ, and $\kappa=3-4\,\nu$ for plane strain conditions, where ν denotes the Poisson ratio.

The complex potentials of an edge dislocation with the Burgers vector $\mathbf{b}=(b_{x_1},b_{x_2})$ in its complex form $b=b_{x_1}+i\,b_{x_2}$ at the point $z_0=x_{1_0}+i\,x_{1_0}$ are given by Eq. 17.

$$\begin{aligned} \phi_0'(z) &= \frac{2A}{z-z_0} \\ \omega_0'(z) &= \frac{2\bar{A}}{z-z_0}-\frac{2(z_0-\overline{z_0})}{(z-z_0)^2}A\;. \end{aligned} \tag{17}$$

The constant $A=\frac{\mu b}{2\pi i(\kappa+1)}$ characterizes the strength of the dislocation. If now a crack is introduced along the negative x_1-axis image stresses develop. This image stress field can be calculated according to Muskhelishvili (8) with Eq. 18.

$$\begin{aligned} \phi_1'(z) &= \frac{1}{2\pi i\sqrt{z}}\int_{-\infty}^{0}\frac{\sqrt{t}}{t-z}p(t)\;dt \\ \omega_1'(z) &= \overline{\phi'(\bar{z})}\;, \end{aligned} \tag{18}$$

where $p(t)$ are the tractions at the negative x_1-axis before the crack was introduced. The function $\sqrt{z}$ which appears in Eq. 18 is a multi-valued function. In the analysis one chooses the branch cut, that is purely imaginary at the negative x_1-axis.

The total stress field in the body is now simply the sum of the original stress field defined by Eq. 17 and the image fields. The result, which is obtained by

applying Cauchy's integral theorem, is:

$$\begin{aligned}
\phi'(z) &= \phi_0'(z) + \phi_1'(z) = \frac{A}{z - z_0}\left[\sqrt{\frac{z_0}{z}} + 1\right] + \frac{A}{z - \overline{z_0}}\left[\sqrt{\frac{\overline{z_0}}{z}} - 1\right] \\
&+ \frac{\bar{A}(z_0 - \overline{z_0})}{2(z - \overline{z_0})^2}\left[\sqrt{\frac{\overline{z_0}}{z}} + \sqrt{\frac{z}{\overline{z_0}}} - 2\right] \\
\omega'(z) &= \omega_0'(z) + \omega_1'(z) = \frac{A}{z - z_0}\left[\sqrt{\frac{z_0}{z}} + 1\right] + \frac{A}{z - \overline{z_0}}\left[\sqrt{\frac{\overline{z_0}}{z}} - 1\right] \\
&- \frac{A(z_0 - \overline{z_0})}{2(z - \overline{z_0})^2}\left[\sqrt{\frac{z_0}{z}} + \sqrt{\frac{z}{z_0}} + 2\right] . \qquad (19)
\end{aligned}$$

The same equations were derived by Lin and Thomson (9). A simple integration with respect to z gives the final solution. Choosing the integration constant, so that the crack tip is in the origin of the coordinate system, one obtains:

$$\begin{aligned}
\phi = \phi_0 + \phi_1 = 2A\,\mathrm{L}og\left[\frac{z_0 - z}{z_0}\right] &+ \bar{A}\,\frac{z_0 - \overline{z_0}}{\overline{z_0}} \cdot \frac{\sqrt{z}}{\sqrt{z} + \sqrt{\overline{z_0}}} - \\
2A\,\mathrm{L}og\left[\frac{(\sqrt{z} + \sqrt{z_0})(\sqrt{z} + \sqrt{\overline{z_0}})}{\sqrt{z_0 \cdot \overline{z_0}}}\right] & \qquad (20) \\
\omega = \omega_0 + \omega_1 = 2\bar{A}\,\mathrm{L}og\left[\frac{z_0 - z}{z_0}\right] &+ 2A\,\frac{z \cdot (z_0 - \overline{z_0})}{z_0 \cdot (z - z_0)} - \\
A\,\frac{(z_0 - \overline{z_0})}{z_0} \cdot \frac{\sqrt{z}}{\sqrt{z} + \sqrt{z_0}} &- 2\bar{A}\,\mathrm{L}og\left[\frac{(\sqrt{z} + \sqrt{\overline{z_0}})(\sqrt{z} + \sqrt{z_0})}{\sqrt{z_0 \cdot \overline{z_0}}}\right] .
\end{aligned}$$

As the branch cut of the logarithm the negative x_1-axis has to be used. Eq. 21 in conjunction with Eq. 17 describe the stresses and deformations induced by dislocations near cracks. Fig.8 shows the contour plot of the σ_{22} stress for an edge dislocation in front of a crack. In the vicinity of the dislocation, the stress field shows the typical shape and concentration of the stress field of a dislocation. However, also in the vicinity of the crack tip, a stress concentration is visible, which is typical for a combined mode I and II loading. This dislocation induced stress concentration is called dislocation shielding or anti-shielding depending on the sign of the stresses. In order to visualize the effect of a crack on a stress field of a dislocation a screw dislocation is located at the x_1 axis at a distance $x_1^{'}$ ahead of the crack tip. The stress caused by the dislocation σ_{23}^D in the plane which is co-planar to the crack is given as

$$\sigma_{23}^D\Big|_{x_1, x_2 = 0} = -\frac{\mu b}{2\pi}\left(\frac{x_1^{'}}{x_1}\right)^{1/2} \frac{1}{x_1^{'} - x_1}. \qquad (21)$$

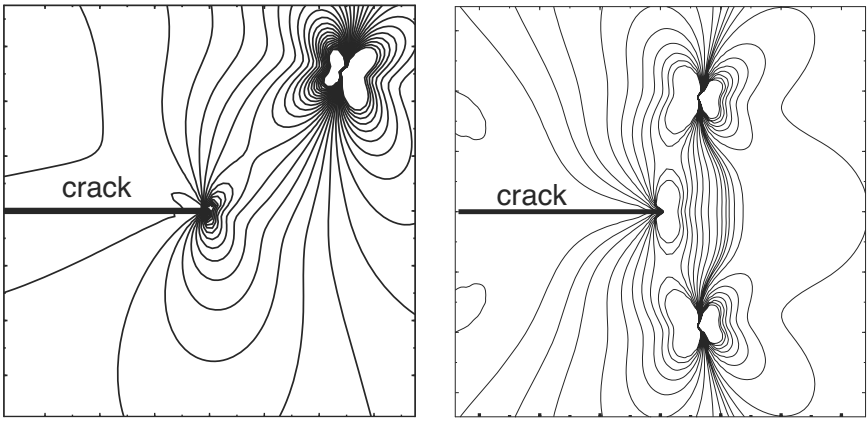


Figure 8: Contour plot of the σ_{22} stresses for an edge dislocation in front of a crack, (a) shows the asymmetric stress field, caused by a single dislocation, and (b) then symmetric stress field of a symmetric arrangement of two dislocations

For $x_1 << x_1^{'}$ eq.21 can be written as

$$\sigma_{23}^{D} = -\frac{1}{\sqrt{2\pi x_1}} \cdot \frac{\mu b}{\sqrt{2\pi x_1^{'}}} \tag{22}$$

Since the first term in eq.22 is equal to the radius dependent term of the near tip stress field (Eq.15), the stress field near the crack tip caused by the dislocation can be characterized by a stress intensity factor, which is termed the shielding stress intensity factor, k_D. A screw dislocation parallel to the crack front induces always a local mode III stress intensity factor, which is given by

$$k_{IIID} = \lim_{x_1 \to 0} \left\{ \sqrt{2\pi x_1} \sigma_{23}^{D} \right\} \tag{23}$$

for the considered case where the dislocation is located at $x_1^{'}$ and $x_2^{'} = 0$

$$k_{IIID} = -\frac{\mu b}{\sqrt{2\pi x_1^{'}}}. \tag{24}$$

The σ_{12} stress of an edge dislocation on x_1 axis at $x_1^{'}$ with a Burgers vector parallel to the x_1 axis differs only by a factor $1/(1+\nu)$ from Eq.(21). The shielding in this case is a pure mode II shielding and is given by

$$k_{II,D} = -\frac{\mu b}{\sqrt{2\pi x'}(1+\nu)} \ . \tag{25}$$

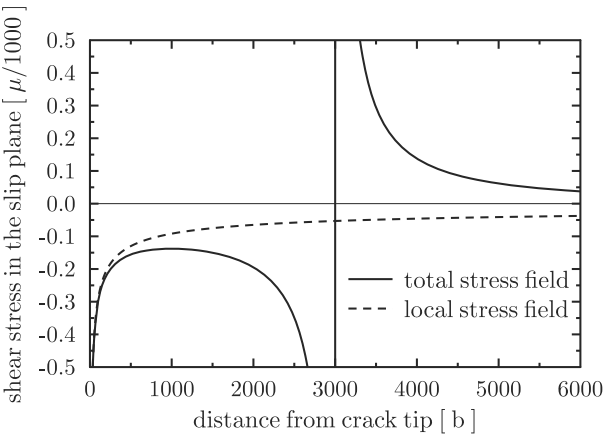


Figure 9: Comparison of the shear stress on the x_1 axis for a screw dislocation, which is located at $x_1^{'} = 3000b$ with the shear stress caused by external mode III loading with a stress intensity equal to the shielding stress intensity caused by the dislocation.

It is evident from this figure 10 and Eq.(21) that in the vicinity of the dislocation and in the vicinity of the crack tip the stress field is characterized by an expression typical for dislocation and typical for crack tips, respectively. The size of the region in front of the crack, where the shielding or antishielding stress intensity factor characterizes the stress field induced by the dislocation, is about 1/10 of distance between the crack tip and the dislocation. Since in our linear elastic consideration the stresses can be superimposed, the stress field very near the crack tip can be always described by a local stress intensity factor

$$k = K + k_D, \tag{26}$$

where K is the applied stress intensity factor and k_D is the stress intensity factor induced by dislocations, called shielding, when its value is negative and antishielding, when its value is positive. When plastic deformation takes place at a crack tip, one has to distinguish between different regions. Very near to the crack tip there is a zone, which can be characterized by a local stress intenisty factor or a combination of local stress intensity factors from the different types of loading. Then there are the zones, where the dislocations are, which are called plasic zones. In this region the applied stress and dislocation stress fields control the local stress field. If this plastic zone is small compared to the K dominated zone, one call this type of loading small scale yielding, i.e. the dislocations are in a K dominated stress field. The description is then easier because the near tip expressions of the crack tip stress fields are sufficient to describe the plastic deformation, the

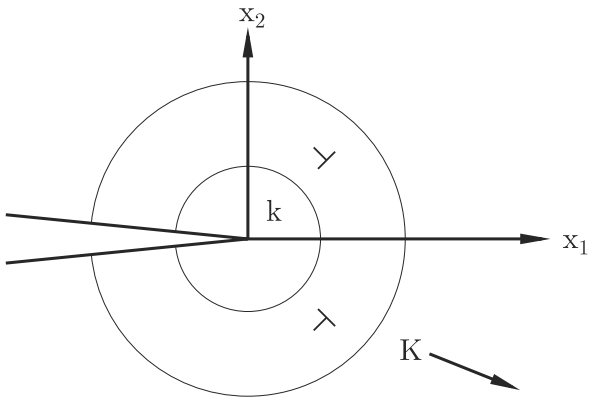


Figure 10: Schematic illustration of the zones, which are controlled by the local stress intenisty factor k and the applied stress intensity factor in the presence by a of a arrangement of dislocation in the vicinity of the crack tip.

generation and movement of dislocation. If the plastic zone is larger than the K dominated zone, the far field stress has to be also taken into account. This type of loading is called large scale yielding. Independent of large or small scale yielding, the local stress intensity is the applied stress intensity plus the sum of shielding or antishielding contributions from the dislocations. Eqs.(24) and (25) give only the shielding stress intensity for a dislocation located at the x_1 axis. For a somewhat more general case, where dislocations are located at a slip plane through the crack tip, for edge dislocations (6).

$$k_{ID} = -\frac{3\mu b_e}{2(1+\nu)\sqrt{2\pi r'}} \sin\Theta \cos\frac{\Theta}{2} \tag{27}$$

$$k_{IID} = -\frac{\mu b_e}{2(1+\nu)\sqrt{2\pi r'}} \left(2\cos\Theta\cos\frac{\Theta}{2} - \sin\Theta\sin\frac{\Theta}{2}\right) \tag{28}$$

and for screw dislocations :

$$k_{IIID} = -\frac{\mu b_e}{\sqrt{2\pi r'}} \cos\frac{\Theta}{2} \tag{29}$$

r' is the distance from the crack tip to the dislocation and Θ is the angle between the crack plane (x_1 axis) and the slip plane. For more general cases, where the slip plane does not intersect the crack tip or where r is in the order or longer than the crack length, i.e. the case of large scale yielding or often also called short crack case - the reader is referred to (2) or (3).

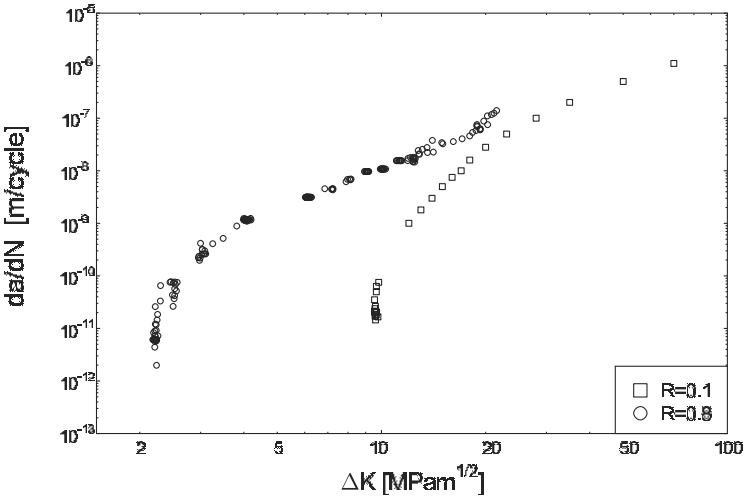


Figure 11: Fatigue crack propagation rate da/dN as a function of the stress intensity factor range ΔK at a stress ratio $R = 0.1$ and 0.8 in an austenitic steel A220 (12).

3.2 Moderate cyclic loading of a crack

Paris (10) recognized that fatigue crack propagation is controlled by the applied stress intensity range $\Delta K = K_{\max} - K_{\min}$, where $K_{\max}$ and $K_{\min}$ are the maximum and the minimum of the stress intensity factor in a loading cycle. This opened the way for the analyses of fatigue in terms of fracture mechanics. Elber (11) realized that only a certain portion of the applied loading applitude is actually transferred to the crack tip, i.e. the crack flanks come into contact during the unloading sequence, even under cyclic tension loading, which leads to a reduction of the active stress intensity range at the crack tip. This contact of the crack flanks – which are usually termed crack closure – is mainly responsible for the mean stress or stress ratio ($R = K_{\min}/K_{\max}$) effect on fatigue crack propagation in ductile metals at low and medium crack growth rates. Fig. 11 shows the fatigue crack propagation rate da/dN as a function of the stress intensity range at a stress ratio, $R = 0.1$ and 0.8 for an austenitic steel, A220 (delivered from Böhler Edelstahl, it is a modified 316L steel). The presented da/dN vs ΔK curves are typical for a ductile material. At medium ΔK values the crack propagation rate is proportional to ΔK^m, this part of the crack propagation curve is called Paris regime. At smaller ΔK values the crack propagation rate decreases progressively untill about 1Å/cycle. Then the propagation rate drops very fast to zero or far below 1Å/cycle. The value, where ΔK drops to zero, is called the threshold of stress intensity range, or short the threshold. The difference in the crack propagation rate in the present example is

mainly caused by crack closure. At low R-ratio and low ΔK the crack is closed over a relatively large part of the load amplitude. With increasing ΔK the percentage of the load amplitude, where the crack is closed, decreases. Hence, the effect of stress ratio is not as pronounced as near the threshold. At higher R-ratios the crack tip is open over nearly the full amplitude of ΔK and the crack growth rate becomes independent of R. The change of the fatigue crack propagation rate versus the stress intensity range is called fatigue crack propagation curve and characterizes the resistance against fatigue crack propagation. In the discussion of the fatigue crack propagation resistance one has to distinguish between the effect induced by crack closure and processes, which takes place at the crack tip and causes crack propagation during the load amplitude, where the crack is open. There are different mechanisms causing crack closure, they can be classified into groups. The most important are the plasticity induced crack closure, roughness induced crack closure and the oxide induced crack closure. The crack closure reduces the cyclic deformation at the crack tip and can therefore affect the crack propagation rate significantly as mentioned above. Ritchie (13) introduced the term extrinsic resistance for the contribution of crack closure to the fatigue crack propagation resistance. The process at the crack tip controlling the real extension of the crack during fatigue loading is denoted as an intrinsic mechanism, and it is responsible for the intrinsic resistance. Hence, the real resistance can be expressed as a sum of the extrinsic (contribution of crack closure) and intrinsic resistance against fatigue crack propagation.

The fatigue crack propagation of materials has been experimentally investigated very intensively within the last 4 decades. The differences between the different materials have been discussed mainly in respect to the different extrinsic mechanisms. The "intrinsic data" are usually obtained from the total fatigue crack propagation resistance of a material reduced by the extrinsic contribution obtained via crack closure measurements. Uncertainties and imprecisions during crack closure measurements therefore reflected themselves in the intrinsic fatigue crack growth curve. This is one of the main reasons, why the intrinsic fatigue resistance is not as well analysed as the extrinsic one.

The simulation of the fatigue crack propagation processes require also a separation of the simulation of the extrinsic processes as well as of the intrinsic phenomena. Despite the increase of computer power and the large number of new approaches with methods ranging from Finite Element simulations (14–16), mesoscale methods (17–19) and discrete dislocation modelling (20–36) down to molecular dynamic simulations (38), we are far away from a complete simulation of all phenomena influencing the fatigue crack propagation resistance of materials.

In the next a summary of improvements in the understanding of near threshold fatigue crack propagation behaviour by the means of discrete dislocation simulations will be presented. Before this we discuss why the diffeent simulation tech-

nuques are needed to understand the different aspects of fatigue crack propagation.

3.3 The involved length scales

To understand why we need different simulation techniques to model fatigue crack propagation and why discrete dislocation models are appropriate for the description of the near threshold regime, we first consider the crack growth curve in Fig. 11 at $R = 0.8$ of the austenitic steel. At this large R-ratio the effect of crack closure can be neglected, hence, this curve can be considered as the intrinsic (or effective) fatigue crack growth curve. The effective threshold is about 2.5 MPa$\sqrt{m}$, The growth rate shows a sharp increase after the threshold and reaches a moderate slope with an exponent of about 2 in the log-log plot in the upper part of the crack propagation curve. At large stress intensity ranges the quasi-static fracture limit is approached, however this part of the curve is not plotted in the Figure 11. A simple estimation of the size of the cyclic plastic zone $\Delta\omega$ based on elasto-plastic continuum mechanics ($\Delta\omega \approx 0.1\Delta K^2/4\sigma_y^2$, where σ_y is the yield stress) shows that $\Delta\omega$ is about 3 orders of magnitude larger than the crack propagation rate. A comparison of the characteristic lengths, the grain size of material which is between few μm to few 100 μm, the lattice spacing, the crack propagation rate and the cyclic plastic zone size indicates clearly:

- At moderate stress intensity ranges the cyclic plastic zone size $\Delta\omega$ is smaller than or in the same order as the grain size. $\Delta\omega$ becomes larger than microstructural features only for very large stress intensity ranges.
- The crack growth increment per loading cycle shrinks to lattice dimensions in the near threshold regime.

All length scales from atomistic to macroscopic are met along a crack growth curve, which indicates, that different simulation methods are needed to describe the fatigue crack propagation processes. The physical length scale of the near threshold regime is the Burgers vector of the dislocations, therefore discrete dislocation models must be used to describe the near threshold plasticity. The finite element methods with classical elastoplastic continuum mechanics can be used to describe the macroscopic plastic deformation in the upper Paris regime, mesoscale methods are appropriate for the lower Paris regime and the molecular dynamic simulations are useful to answer the question, where the plasticity comes from, i.e., where and when dislocations are generated. Furthermore molecular dynamic simulations are usually required to describe the atomistic separation processes. In the following it will be shown that in the case of fatigue the irreversible generation of new surfaces by dislocation generation or annihilation at the crack tip are sufficient to cause a crack propagation. This process should give a lower limit for the intrinsic fatigue resistance. Additional atomistic fracture processes may increase the crack propagation rate, for such processes a molecular dynamic simulation would be necessary.

The dislocation model used by our group to describe the plastic deformation at cracks under near threshold conditions will be briefly introduced in the next section.

4 Modelling of plasticity, crack propagation and fracture surface contact

In order to illustrate the different phenomena occurring during fatigue loading a mode I crack under cyclic loading will be considered. Discrete dislocation modelling is in principle a linear elastic description of stress and strain fields, where the nonlinearity is taking into account by the motion of discrete dislocations on pre-defined slip planes. Our computer algorithm for the simulation of plastic deformation induced by cyclic loading consists of

1. an incremental increase in applied load (or decrease during unloading),
2. inspection of dislocation generation,
3. inspection of fracture surface contact and determination of contact stresses, and
4. seeking the equilibrium positions of the dislocations (in the present simulation the static equilibrium configuration are determined).

A two-dimensional mode I crack is used that means the dislocations are parallel to the crack front.

The crack propagation mechanism is schematically depicted in figure 12. The crack tip is assumed to be the dislocation source. Dislocations are generated, when the stress intensity is larger than a critical value k_e. A symmetric emission of dislocations is assumed. The first two generated dislocations (one in the upper half-space and one in the lower half-space) form a V-shaped notch. The next two dislocations are generated at the tip of the notch; hence, the spacing between the slip planes of the first- and the second-generated dislocation is equal to the lattice spacing. This opening of the V-shaped notch continues until the maximum load is reached. During unloading, the last-generated dislocations return to the tip of the V-shaped notch. It is assumed that the resharpened crack does not reweld. About the same number of dislocations, which returned to the crack tip are generated at the new crack tip and again form a V-shaped notch. During unloading, most of these dislocations return and the crack grows by this blunting and resharpening process over a distance that is proportional to the cyclic crack-tip-opening displacement ($\Delta CTOD$). Contact of the fracture surface at positive stress ratios R is only possible on the newly created fracture surfaces. In the first few cycles this zone is very small; hence, this contact does not affect the cyclic plastic deformation. With an increasing number of cycles this contact zone increases and the contact stresses have to be taken into account. A procedure, which enables us to calculate the contact stresses and their influence on the dislocation motion has been

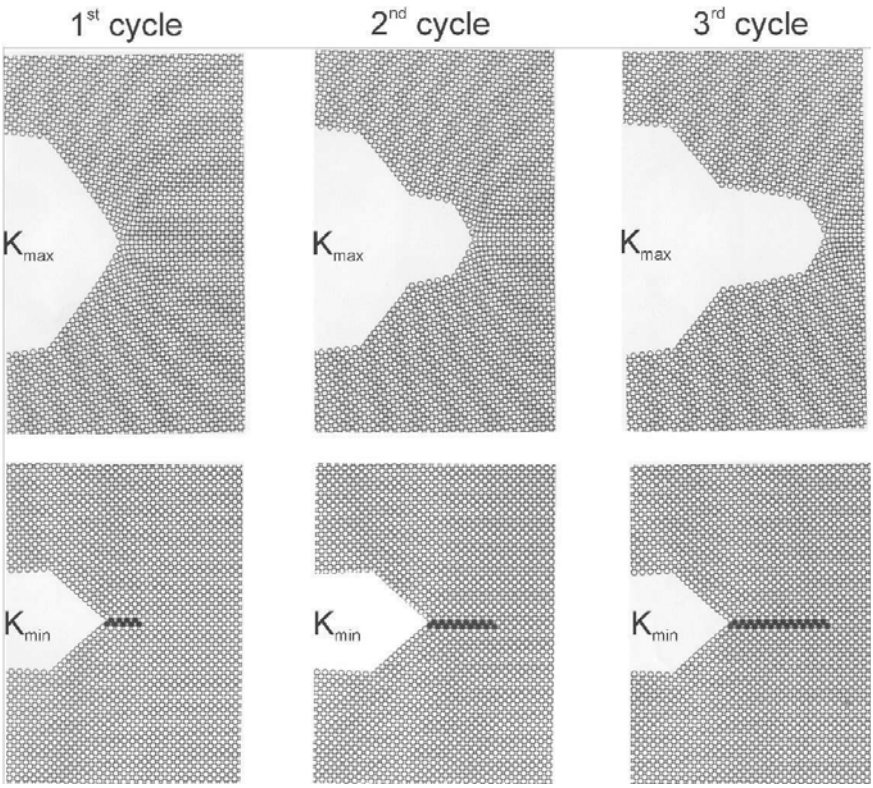


Figure 12: Schematic representation of the crack propagation mechanism by blunting and resharpening of the crack tip

described in detail by Riemelmoser and Pippan (20). The computer algorithm is based on a collocation method. It starts with calculation of the crack contour and inspection of overlapping regions, which is divided into small subdomains. The contact stress in a subdomain j displaces the midpoint of the element j, the so-called collocation point, and also the collocation points of all the other elements i by an amount g_{ij}. This g_{ij} was derived by Tada et al. (40). The collocation point j is displaced not only by the stress in the element j but also by the stresses in the other elements. Now a stress distribution is sought such that the displacement of each collocation point is zero, in other words the displacements at the collocation points produced by the contact stresses must be equal to the displacement in the overlapping state. The solution of the linear algebraic equation system provides a first approximation. Physically, contact stresses must be compressive. The

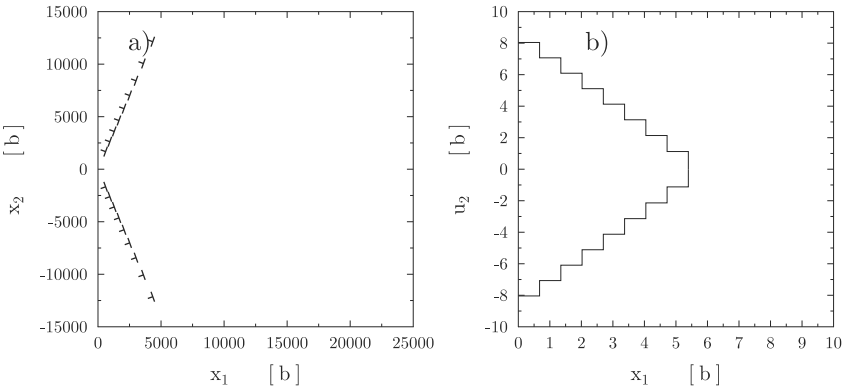


Figure 13: Arrangement of dislocations at $K_{\max} = 1.4k_e$ (a). Since the repulsion force between dislocations are too small, the dislocations during the unloading to $K_{\min} = 0.14k_e$ do not return to the crack tip. The indicated plastic opening of the crack tip (b) does not change during further cycling.

described method, however, provides tension stresses in a few elements. These elements are cancelled in the second iteration. The iteration is continued until the stress in each element is compressive. The next step is the evaluation of the force on a dislocation caused by the contact stresses. This is given by the Eshelby integral (41), which has been transferred in the complex potential notation by Budiansky and Rice (42). The total force on the dislocation due to the contact stresses then is simply the linear sum over all elements.

4.1 The cyclic plastic deformation as a function of load amplitude and number of cycles

In the present simulations the stress ratio $R = K_{\min}/K_{\max} = 0.1$. The material parameters used are the shear modulus $\mu = 80000MPa$, Poisson's ratio $\nu = 0.3$, a lattice friction stress of $\mu/1000$, a critical stress intensity to generate a dislocation at the crack tip $k_e = 0.4\mu\sqrt{b}$ and an angle α between the crack propagation direction and the slip plane of 70.3^o. At the beginning of our simulation we start with a crack in a perfect single crystal without dislocations. If $K_{\max}$ is smaller than k_e, no dislocations will be generated at the crack tip. During cyclic loading at such small load amplitude a pure elastic loading and unloading and therefore no crack extension will take place. Since no dislocations are generated, the local stress field at the crack tip is determined by the applied stress intensity factor K_{ap} solely or in other words the local stress intensity is $k = K_{ap}$. If K_{ap} is larger than k_e during the first loading, a dislocation will be generated at k_e. It will move away

from the crack tip till it reaches its equilibrium position, for details see (21; 29). These dislocations – due to our symmetric arrangement we have two dislocations in front of the crack – reduce the stress field at the crack tip. They shield the crack tip. In order to generate further dislocations, one has to increase the applied stress intensity furthermore. When the local stress intensity overcomes again the critical value to generate a dislocation, the next dislocations are emitted. They push the existing dislocations away untill they reach their equilibrium positions. This process continues untill $K_{\max}$. If the applied ΔK is only somewhat larger than the critical stress intensity to generate a dislocation at or near the crack tip, the repulsive forces between the dislocations are not sufficiently large to push a few of the dislocations back to the crack tip during unloading.

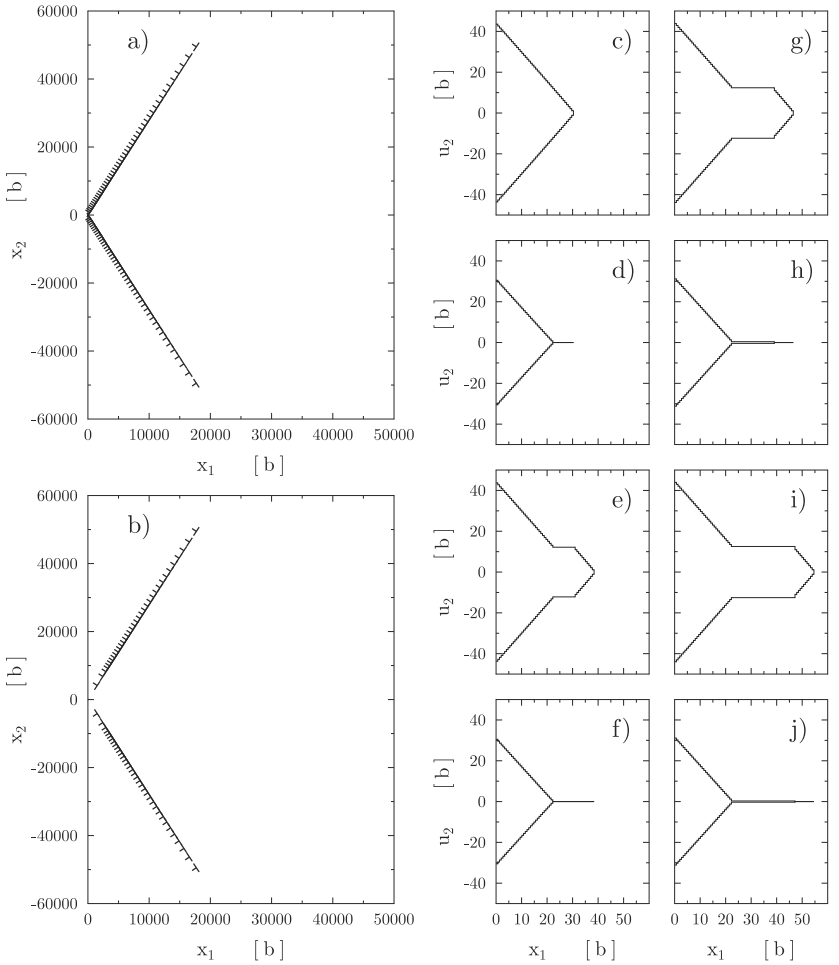


Figure 14: Arrangement of dislocations at $K_{\max} = 2.5k_e$ (a) and $K_{\min} = 0.25k_e$ (b) in the fourth load cycle. The shape of the crack tip in the 1st, 2nd, 3rd and 4th cycle at $K_{\max}$ and $K_{\min}$ are represent in (c)-(j).

For such loading case the dislocation arrangement is shown in Fig. 13. At the applied $K_{\max} = 1.4k_e$ the dislocations forming the plastic zone of about few microns, and the crack tip opening displacement is about 4.5 nm. In the immediate vicinity of the crack tip there are no dislocations. This region is called the dislocation free zone, which is a consequence of the assumed dislocation source – the crack tip. During further loading between $1.4k_e$ and $0.14k_e$ the dislocations remain more or less at their equilibrium position, which they take up at the maximum load in the first cycle. In the further loading cycles with the same load amplitude a linear elastic loading occurs, in the vicinity of the crack tip the stress field can be described by a local k, which varies in this case, linear with the applied K_{ap}, $k = k_e - (K_{ap} - K_{\max})$. For a propagation of a crack one needs a cyclic plastic deformation at the crack tip, i.e. one needs a cyclic generation of surface. Hence a somewhat larger load amplitude is required. The arrangement of the dislocations and the shape of the crack tip are shown in Fig. 14 at $K_{\max}$ and $K_{\min}$. The applied $K_{\max} = 2.5k_e$. In the first cycle during loading a large number of dislocations is generated and one forms a plastic zone of about 10 μm and a relatively large crack tip opening displacement of about 20 nm. During unloading, the former generated dislocations return to the crack tip and reshapen the blunted crack tip as schematically depicted in Fig. 12. In the next loading cycle about the same number of dislocations – which returned before – are generated during loading and return again during unloading. This blunting and resharpening causes the crack propagation. The process is depicted in Fig. 14 for the fourth cycle. The new generated fracture surface came into contact at $K_{\min}$. This contact does not effect the cyclic plastic deformation of the crack tip at such small crack extension. However, after greater crack extension the crack flank contact at a stress intensity factor significantly larger than $K_{\min}$, this reduces the cyclic plastic deformation. At small ΔK values this contact can induce a disappearance of the cyclic plastic deformation and which causes therefore, a stopping of the crack propagation. In Fig. 15 the dislocation arrangement and the shape of the crack flanks are shown for the same load amplitude as in Fig. 14. The crack is grown over about 20 μm, and the cyclic plastic deformation is reduced. Only one dislocation is generated during loading and it returns to the crack tip during unloading. This reduction of the cyclic plastic deformation is caused mainly by crack closure as mentioned above. At ΔK the crack is closed over a distance of about few μm behind the crack tip. After further small crack extension at this load amplitude the crack stops propagating.

At larger load amplitudes the decrease of the crack propagation rate occurs similarly, however a stopping of the crack does not take place. The cyclic plastic deformation and as a consequence, the fatigue crack growth rate approaches a nearly constant value. The dislocation arrangement and the crack flank contour for such a "steady state" growing fatigue crack is shown in Fig. 16. In Fig. 17 the calculated $\Delta CTOD$ values as a function of the crack extension are plotted for

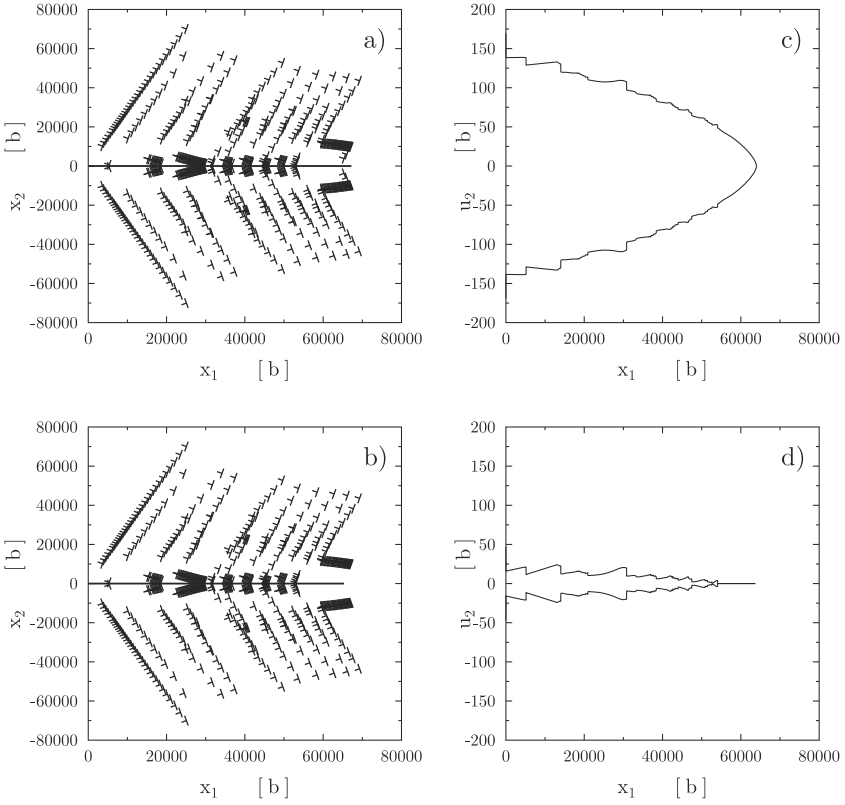


Figure 15: Arrangement of dislocations at $K_{\max} = 2.5k_e$ (a) and $K_{\min} = 0.25k_e$ (b) in the 10000th cycle. The correspinding shape of the crack flanks at $K_{\max}$ and $K_{\min}$ are dipicted in (c) and (d).

different ΔK values; the stress ratio is always 0.1. In this diagram the mentioned decay and the final vanish of cyclic $CTOD$ at small ΔK is evident. At lower ΔK values, the disappearance of $\Delta CTOD$ takes place at crack extensions in the order of micrometers.

The last time crack propagation stops at a crack extension of few 10 μm, which is in the order of magnitude of the plastic zone size, as can be seen in Fig. 15. At ΔK values larger than a certain critical value the cyclic $CTOD$ reaches a nearly constant value at a crack extension somewhat larger than the size of the plastic zone. In Fig.17 CTOD in the first cycle at maximum load, the initial cyclic $CTOD$, $\Delta CTOD_i$; and the steady state cyclic $CTOD$, $\Delta CTOD_s$ are plotted as

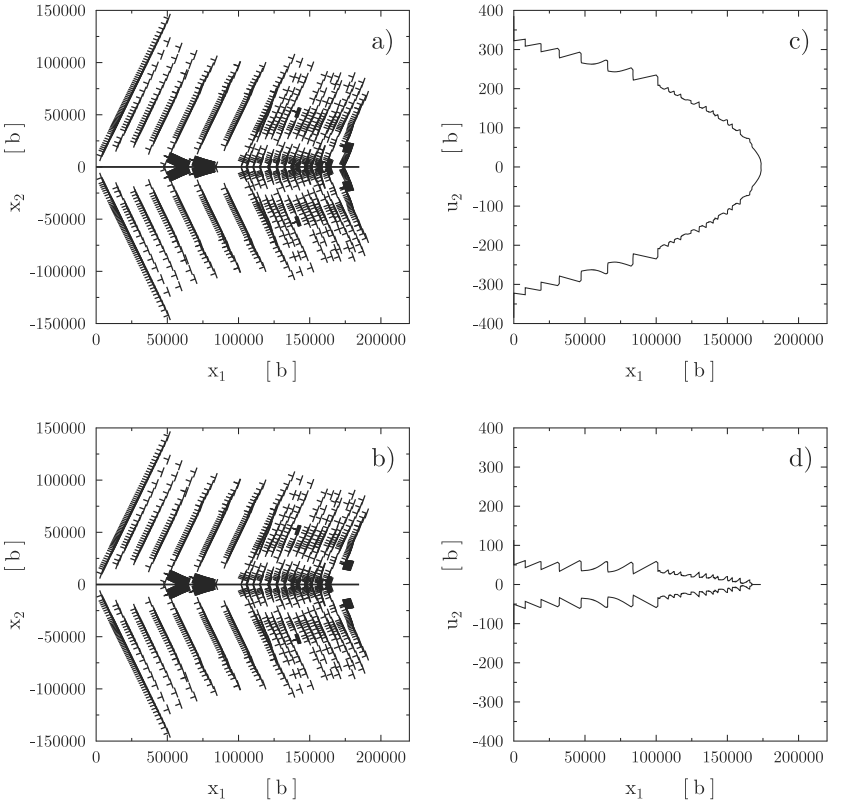


Figure 16: Arrangement of dislocations at $K_{\max} = 3.5k_e$ (a) and $K_{\min} = 0.35k_e$ (b) in the 16000th cycle. The correspinding shape of the crack flanks at $K_{\max}$ and $K_{\min}$ are shown in (c) and (d).

a function of ΔK. Since the fatigue crack propagation rate should be proportional to $\Delta CTOD$, the $\Delta CTOD_i$ vs ΔK and the $\Delta CTOD_s$ vs ΔK can be interpreted as fatigue crack growth curves. As already mentioned, Ritchie (44) introduced the term intrinsic and extrinsic crack propagation resistance. The intrinsic mechanism responsible for the propagation of the crack in ductile metals is the cyclic plastic deformation at the crack tip, similar as presented in Fig. 12. The extrinsic mechanisms reduce or increase the local crack driving force, in our idealized case it is the reduction of cyclic plastic deformation caused by the contact of the crack flanks. By using the terminology of Ritchie and taking the described crack propagation mechanism, $\Delta CTOD_i$ vs ΔK and $\Delta CTOD_s$ vs ΔK can be interpreted as the

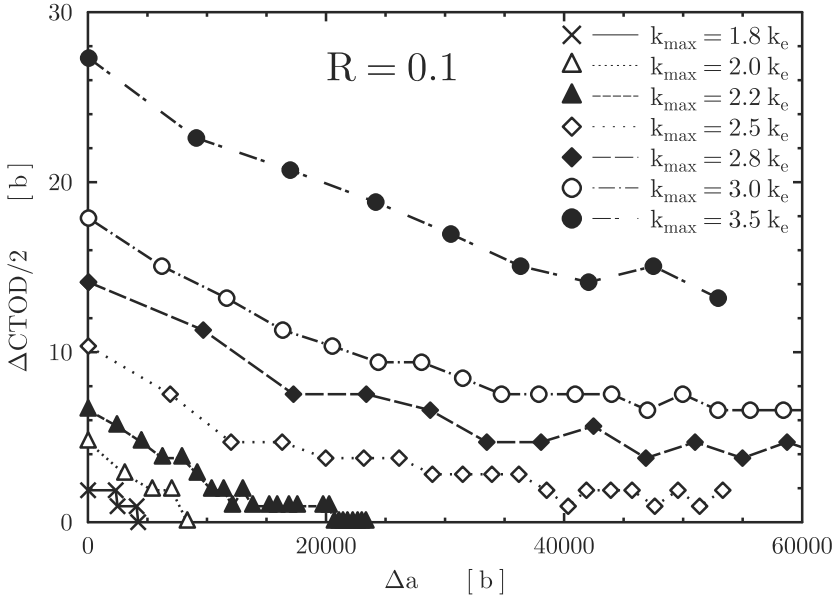


Figure 17: Determined cyclic plastic crack tip opening dispolacement $\Delta CTOD$ for different ΔK values as a function of crack extension, Δa, from the performed discrete dislocation simulation (43).

intrinsic crack growth curve and the long crack growth curve, which is given by the sum of the extrinsic and intrinsic resistances, respectively.

Different mechanisms can be responsible for crack closure. The three most important are: plasticity induced crack closure, roughness induced crack closure and oxide debris induced crack closure. In our considered idealized case only the plasticity induced crack closure is considered, because a plane crack extension without oxidation is considered in the simulations. From the continuum mechanics point of view under constant – amplitude loading and steady state condition, the ratio of closure stress intensity factor to maximum stress intensity factor, $K_{\mathrm{cl}}/K_{\mathrm{max}}$, is independent of ΔK and it is a function only of the stress ratio R. The value $K_{\mathrm{cl}}/K_{\mathrm{max}}$ is a measure of the relative contribution of the effect of crack closure. This is clearly evident from the plane-stress analysis made by Budiansky and Hutchinson (48) and Führing and Seeger (49). In the Paris regime – the higher ΔK regime in Fig. 18 – it seems that the contribution of crack closure reaches a constant value, as expected from continuum plasticity. However, in the near-threshold regime the discrete nature of plasticity causes an increase in the effect of crack closure. It is surprising that in the near threshold, when the plasticity is

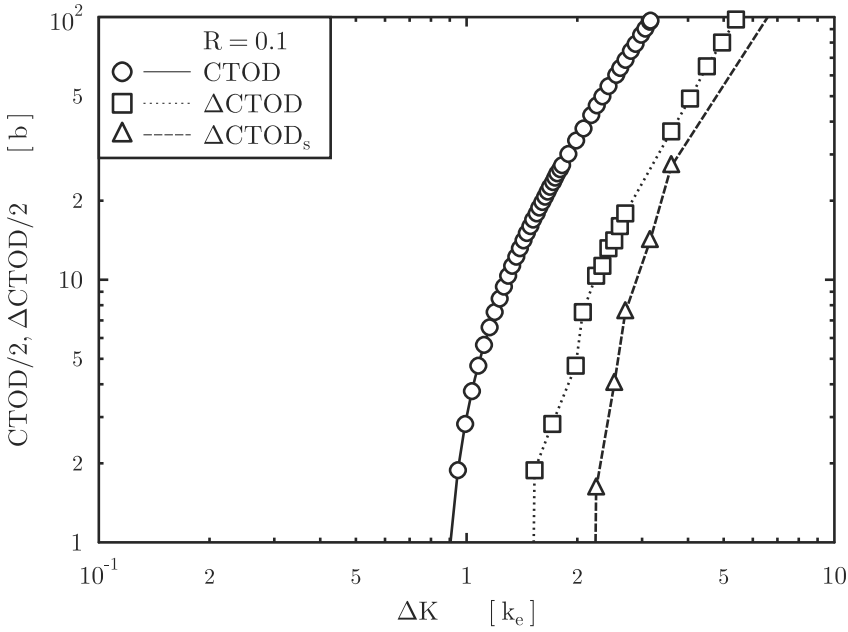


Figure 18: Crack tip opening displacement at $K_{\max}$ in the first cycle, $CTOD$, the cyclic crack tip opening displacement in the first cycle, $\Delta CTOD$ and the steady state cyclic crack tip opening displacement in the first cycle as a function of the applied stress intensity range ΔK.

constrained, the plasticity-induced crack closure increases. A closer look from the dislocation point of view can explain this phenomenon. Crack closure is caused under plane-strain conditions – which is considered here – by wake dislocations. The number of dislocations in the wake of a growing fatigue crack is given by the number of dislocations generated during loading minus the number of dislocations returned to the crack tip or annihilated. Near the threshold of stress intensity range, the number of dislocations returning to the crack tip goes to zero; therefore nearly all generated dislocations during propagation can contribute to closure. This effect is visible, when we compare $CTOD$ and $\Delta CTOD$ in Fig. 18, which characterizes the number of dislocations generated and the number of dislocations returned to the crack tip, respectively. In other words, due to the decrease of ΔK the reduction in the monotonic deformation, which generates shielding dislocations, is not as pronounced as the reduction of the number of the returning dislocations.

More important than this effect of the discrete nature of plasticity on the increased crack closure effect near the threshold is its impact on the intrinsic fa-

tigue crack growth behaviour or the $\Delta CTOD$ vs ΔK curve. $\Delta CTOD$ as well as $CTOD$ in terms of continuums plasticity should be proportional to ΔK^2 and K^2, respectively, as long as small scale yielding is fulfilled. Fig. 18 indicates that for larger K and ΔK values or for $CTOD$ and $\Delta CTOD$ larger than 100 b the discrete dislocation simulation agrees well with the continuum plasticity (1). From the physical point of view both plasticity models are equivalent in the limit of large plastic deformation. For smaller ΔK-values or when $CTOD$ or $\Delta CTOD$ shrinks to some 10 Burgers vectors, the plasticity is constrained by its discrete nature. Such phenomena are naturally lost by the smoothing procedure of continuum plasticity. This constrain leads to a sharp decrease in plastic deformation at small K and small ΔK, near the threshold of stress intensity range. Near the threshold the difference between the discrete plastic deformation and continuum plasticity is roughly one or two order of magnitudes (23; 24; 27; 28).

So far our discussion has focused on the effect of the discrete nature of plasticity on the cyclic plastic deformation and the development of crack closure. Another interesting point is the developed dislocation structure. At small loading amplitudes the wake disloctions arrange in slip bands as can be seen in Fig. 15-17. The formation of these slip bands can be explained as follows:

> In the first loading sequence many dislocations are generated on the two inclined slip bands. During unloading few dislocations return to the crack tip, where they annihilate. About the same number of dislocations, which returned during the first cycle are generated and returned at the crack tip during the second and many further loading cycles. By this blunting and "resharpening" process the crack propagates over a short distance and leaves the two slip bands formed in the first cycle behind the crack tip. As the crack tip moves away from the first two slip bands, their repulsive force on the newly generated dislocations decreases. After a certain crack extension, the interaction force between the pre-existing slip bands and the new generated dislocations is small enough to let pass the two pre-existing slip bands. A second pair of slip bands is formed. Similar processes in the later cycles lead to the dislocation arrangement depicted in Figs. 15 to 17. The distance between the slip bands is about few 1000 Burgers vectors, but the crack growth rate is only few Burgers vectors per cycle. Smaller distances between the slip bands do not occur because the elastic interaction stresses of the dislocations in the last band and the new generated dislocations are too large. The distance between the slip bands does not depend significantly on the crack growth increment per load cycle. Each slip band leaves a step on the fracture surface parallel to the crack front with a characteristic distance in the order of some tenths of a micron. In (51) it was noted that this distance

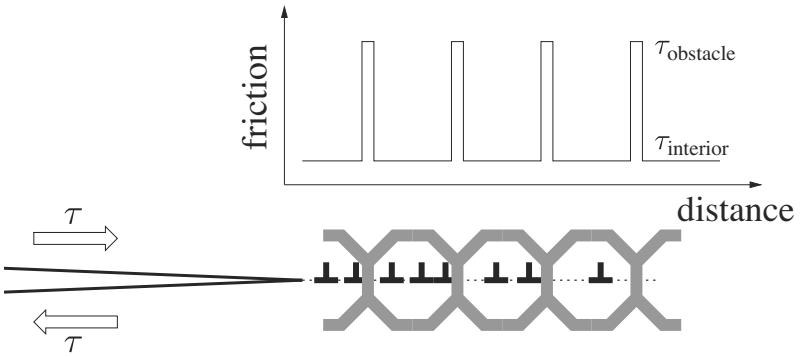


Figure 19: Schematic representation of a 2D mode II crack to simulate the effect of grain boundaries on cyclic plastic deformation near the threshold of crack propagation.

agrees with the striation spacings observed in metals.

A symmetric dislocation arrangement as shown in Fig. 14-17 is likely to occur at intermediate stress ranges. At small stress intensity ranges asymmetric crack tip plasticity should occur more often like in the models proposed by Neumann (45) and Pelloux (46). It is, however, obvious that the dislocation arrangement of the asymmetric crack is also governed by the dislocation-dislocation interaction forces, which should lead also to the typical distance between two large slip bands of the order of some tenths of a micron. Following this argument, the striations observed at small stress intensity ranges are traces of slip bands on the fracture surfaces.

4.2 The effect of boundaries

The analyses of monotonic or cyclic deformation of a mode II and mode III crack by means of discrete dislocations with a slip plane in the $x_1 - x_3$ plane with the crack along the negative x_1-axis and the slip plane along the positive x_1-axis are much simpler to simulate than the previous considered mode I cracks. The general features in the cyclic plastic deformation, however, remain the same, except crack closure does not occur at such idealized mode II and mode III cracks because no dislocations remain in the wake of the crack tip. For the consideration of the effect of microstructure on the cyclic plastic deformation we will analyse the periodical arrangement of grain boundaries ahead of a mode II crack, as schematically depicted in Fig.19. A fine grained material with a grain size $d = 3000b$ (20) and a coarse grained material with $d = 10.000b$ have been studied. In the case of the fine grained material the grain size is comparable to the size of the dislocation

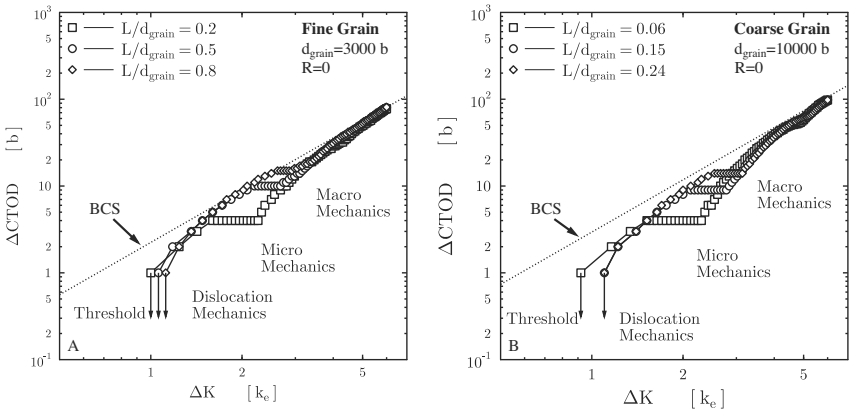


Figure 20: Cyclic crack tip opening displacement of a mode II crack as a function of ΔK for different distances to the first grain boundary. In (a) and (b) $\Delta CTOD$ vs ΔK for a fine and coarse grained material is shown, respectivel in both figures the distance between the crack tip and the nearest grain boundary is equal. Material data: $\tau_{\text{friction}} = 0.002\mu$, $\tau_{\text{barrier}} = 5\tau_{\text{friction}}$, $k_{\text{e}} = 0.2\mu\sqrt{b}$, $\nu = 0.3$, $R = 0$.

free zone, whereas in the coarse grained case d is significantly larger than the dislocation free zone. It is instructive to consider first ΔK values larger than $5k_{\text{e}}$. In this regime the cyclic plastic deformation is insensitive to the actual location of the crack tip with respect to the grain boundaries. The curves for the three simulations in both materials collapse. The cyclic plastic deformation in this regime is in good agreement with calculated values by the BCS model (Bilby, Cottrell and Swinden model (52)) with the assumption of a homogeneous macroscopic yield stress. This BCS model describes plasticity by distribution of infinitesimal small Burgers vectors. Plasticity is smeared out along the shear plane. Such type of description corresponds to the continuum mechanics model. Hence, the cyclic plastic deformation for the different positions of the crack in the grain is in this loading regime well predicted by continuum mechanics. The microstructure is reflected only in the different macroscopic yield stress. Following the Hall-Petch model the macroscopic yield stress is 250 and 190 MPa for the fine and coarse grained materials, respectively.

At intermediate stress intensity ranges the cyclic plastic deformation is significantly affected by the microstructure, i.e., the distance between the crack tip and the grain boundary as well as the size of the grains. At the threshold the various curves approach each other again. This example indicates that one can distinguish between three regimes: the macro mechanics, the micromechanics and the dislocation mechanics. In the macro mechanics regime the cyclic plastic deformation

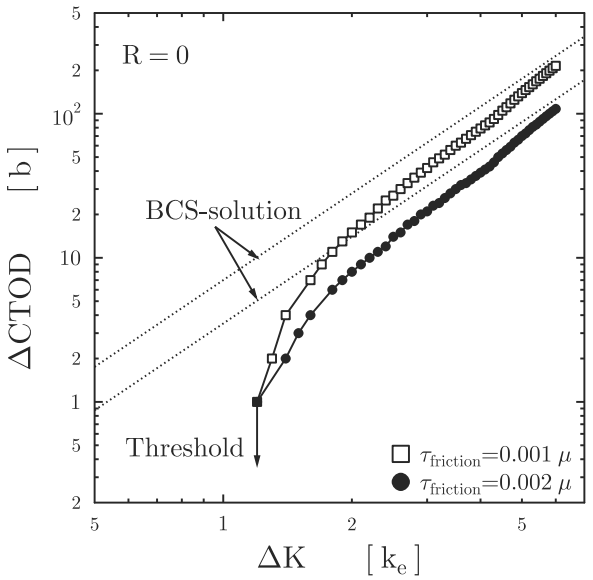


Figure 21: The crack tip opening displacement as a function of ΔK. Comparison of two different critical shear stresses. $k_e = 0.2\mu\sqrt{b}$.

at the crack tip is almost independent from the distance between the crack tip and the next grain boundary. This is different in the micromechanic regime at intermediate stress intensity ranges, when the cyclic plastic zone is in the order of the distance between the crack tip and the grain boundary. In this regime the grain boundary acts as barrier for the cyclic plastic deformation, and reduces $\Delta CTOD$. Such effects has been observed also by analyses based on distributed dislocations (53). In this micromechanical regime continuum mechanics analyses are still useful, however, the local variation of the yield stress has to be taken into account explicitly.

Finally, close to the threshold of stress intensity range a different behavior is observed. For certain cases $\Delta CTOD$ even increases for smaller distances between the crack tip and the next grain boundary. This behavior as well as the occurance of a threshold is a consequence of the discrete nature of plasticity and can not be explained by continuum plasticity. In summary this example has shown that at large stress intensity ranges the grain size determines the cyclic plastic deformation at the crack tip. At smaller stress intensity ranges, $\Delta CTOD$ is controlled by the distance between the crack tip and the next grain boundary and near the threshold the activation of the dislocation source is the dominant effect. This is a very general

behavior, which one can obtain for many different loading cases of a fatigue crack, see for details (54), however, one can observe such phenomena in many other localized deformation processes, for example in the case of nanoindentation (55).

4.3 The threshold of cyclic plastic deformation and the effective threshold of stress intensity range

Fig.21 shows the calculated cyclic crack tip opening for an ideal mode II crack in a single crystal with a slip plane in the x_1x_3 plane for different friction stresses. This Figure shows again the strong effect of the discrete nature in the near threshold regime and the continuous approach to the continuum mechanic BCS prediction at larger ΔK where $\Delta CTOD = \frac{\Delta K_{II}^2(1-\nu)}{4\mu\tau_{\text{friction}}}$. At large ΔK values $\Delta CTOD$ follows the BCS model, $\Delta CTOD$ is therfore inversly proportional to the friction stress, however the onset of cyclic plasticity is independent of the assumed friction stress. This behavior has been also observed in the studies of the effect of grain size, the effect of obstancle distance and obstacle strength (29; 50) and analysis of a lamellar structure. All these analyses indicate that the threshold stress intensity range to generate a cyclic plastic deformation is mainly controlled by the source stress or more accurately the stress intensity to generate a dislocation at the crack tip or in the immediate vicinity of the crack tip. Assuming the fatigue crack propagation mechanism by blunting and re-sharpening as depicted in Fig.12 this threshold value to generate a cyclic plastic deformation should be the effective threshold of stress intensity range, $\Delta K_{\text{eff, th}}$. The simulations have shown that $\Delta K_{\text{eff, th}}$ is about $1.3k_{\text{e}}$. The Rice-Thompson model (1; 2) for dislocation generation at the crack tip leads to an estimate of k_{e} and hence, to an effective threshold equal to $\Delta K_{\text{eff, th}} = f(\Theta)\,\mu \cdot \sqrt{b}$, where Θ is the angle between the slip planes for the dislocation emission and the crack plane, and $f(\Theta)$ is about 1 or somewhat smaller for reasonable Θ values. The absolute values are compared in Table 1. The experimental results are from Ref.(53) and the estimated values are from Refs.(55; 56). It has to be noted that same experimentally determined $\Delta K_{\text{eff, th}}$ in the literature are significantly larger. In the author's opinion this deviation is mainly caused by uncertainties in the mesurements of crack closure or for the cases, where the effective value of ΔK_{th} is determined from experiments at large R values, this cracks may not be completely crack closure free. The measured threshold values are systematically somewhat larger than the values predicted by the discrete dislocation model. Reasons for this deviation might be :

- even these presented experimental values may be not closure free,
- the crack deflection and crack branching can reduce the local stress intensity range in real cracks,
- the oxide layer may change the resistance against dislocation generation,
- the assumed dislocation generation criterion is somewhat too simple.

Table 1: Comparison of measured and predicted threshold stress intensity ranges

Alloy system	measured effective treshold $\Delta K_{\text{eff, th}}$ in $MPa\sqrt{m}$	Hertzberg's estimation $E\sqrt{b}$ $MPa\sqrt{m}$	Threshold from dislocation model $\Delta K_{\text{eff, th}}$ in $MPa\sqrt{m}$
Fe	2.5 – 2.9	3.3	1.6
Al	0.8 – 1.0	1.2	0.6
Cu	1.5 – 2.5	1.9	0.9

Taking into account the mentioned uncertainties, the agreement between the prediction and the measured effective ΔK is relatively good. The relative insensitivity of $\Delta K_{\text{eff, th}}$ in respect of the microstructure has been reported by different groups. Herztberg (55; 56) proposed an empirical relation, based on the experimental findings, with $\Delta K_{\text{eff, th}} = E\sqrt{b}$. This value is also listed in Table 1. This relation reflects the outcome from the discrete dislocation analyses well.

4.4 Other discrete dislocation simulations of fatigue crack propagation

Also other groups have performed similar discrete dislocation simulations of cyclically loaded cracks under small scale yielding conditions. The results are in good agreement with the above mentioned behaviour, see, for example (30). Beside those, two somewhat different types of simulations should be mentioned here shortly (32–36). Despande, Needleman and van der Giessen (35) and (36) analysed a cyclically loaded mode I crack under small scale yielding condition by using discrete dislocation dynamics. At each stage of loading, the stress field of the dislocations and the solution of boundary condition obtained in these simulations by finite element solution are superimposed. The dislocation motion, dislocation nucleation (the sources are statistically distributed), dislocation interaction with obstacles and annihilation are taken into account in a similar way as in the previous described simulations. The essential difference in their simulation is the assumed crack propagation mechanism based on a cohesive zone model. It is surprising that the behavior regarding the onset of fatigue crack propagation is relatively similar, despite the different growth process. However, the growth rate is very sensitive to cohesive energy, source and obstacle density. In the author's opinion such propagation mechanisms should govern the fatigue crack growth in intermetallics and other semibrittle materials whereas in ductile crystalline materials only the deformation controlled growth mechanism by blunting and re-sharpening is dominant. More details regarding this type of simulations are presented in one of the following chapters (see the contribution of Van der Giessen).

Finally, a series of discrete dislocation simulations is mentioned, which were

performed by the group of Melin (32–34; 57). In the previously discussed simulation long cracks are considered, i.e., the crack length was large in relation to the size of the plastic zone and the microstructural features. The group of Melin analysed the crack propagation of short cracks, where the size of the plastic zone and the microstructural features are in the order of the size of the crack length. The geometry of the boundary and the crack is described using dislocation dipole elements, whereas the plasticity is described by the movement of discrete dislocations. Crack angle, crack shape, distance to the grain boundary can be taken into in such simulation. This type of simulations show clearly that relatively small plastic deformation of a crack can be simulated in a relatively complex microstructure, which helps to prove the physical ideas behind the phenomena in the early stage of fatigue crack propagation.

In summary, the described important consequences of the discrete dislocation mechanics for fatigue crack propagation can be obtained despite the simplification due to the assumption. This should be the case also for the other phenomena, especially for fracture, however one has to analyse when the 3D-effect might be important.

Bibliography

[1] J.R. Rice, 1968, *Mathematical analysis in the mechanics of Fracture*, in: *Fracture*, edited by H. Liebowitz, Academic Press, Vol. 2, pp. 191-311.

[2] R. Thomson *Physics of Fracture*, *Solid state physics*, Vol 39, 1-127,

[3] J. Weertman, 1996 *Dislocation based fracture mechanics*, *World Scientific Publications, Singapore*

[4] T.L. Anderson, 1995, *Fracture Mechanics, Fundamentals and Application*, CRC Press

[5] J.O. Eshelby, 1973 *The continuum theory of lattice defects*, *Sol. Stat. Phys.*, Vol.3, p.79.

[6] S.M. Ohr, 1985 *Electron microscope study of crack tip deformation and its impact on the dislocation*, *Materials Science and Engineering* **72**, 1-35.

[7] N. Louat, K. Sadananda, M. Duesbery and A.K. Vasudevan, 1993, *A theoretical evaluation of crack closure*, *Metall. Trans. A*, Vol.24A, pp.2225-2232.

[8] N.I. Muskhelishvili *Some basic problems of the theory of elasticity*, Noordhof, Groningen, The Nederlands (1963).

[9] L.H. Lin and R. Thomson, 1986, *Cleavage, dislocation emission and shielding for cracks under general loading*, *Acta Met.*, vol.34, pp.187-206.

[10] P.C. Paris, F. Erdogan, 1963, *J. Basic Eng* **85**, 528-534.

[11] W. Elber, 1970, *Engineering Fracture Mechanics* **2**, 37-45.

[12] C. Bichler, R. Pippan, 2007, *Engineering Fracture Mechanics* , 39

[13] R.O. Ritchie, S. Suresh, 1982, *Metall. Trans.* **13A**, 937-940.

[14] R.C. McClung, H. Sehitoglu, 1989, *Engineering Fracture Mechanics* **33**, 237-272.

[15] R.C. McClung, D.L. Davidson, 1991, *Engineering Fracture Mechanics* **39**, 113-130.

[16] J.C. Newman Jr., C.A. Bigelow, K.N. Shivakumar, 1993, *Engineering Fracture Mechanics*, **46**, 1-13.

[17] K. Tanaka, T. Mura, 1984, *Acta Metall.* **32**, 1731-1740.

[18] K.S. Ravichandran, E.S. Dwarakadasa, 1991, *Acta Metall. Mater.* **39**, 1343-1357.

[19] L. Xu-Dong, L. Edwards ,1996, *Engineering Fracture Mechanics* **54**, 35-48.

[20] F.O. Riemelmoser, R. Pippan, 2000, *Proc. of Fatigue Crack Growth Thresholds, Endurance Limits and Design*, ASTM STP 1372, (ed. J.C. Newman and R.S. Piascik), 252-265.

[21] T. Yokobori, A.T. Yokobori Jr., A. Kamei, 1974, *Phil. Mag.* **30**, 367-378.

[22] J.C.M. Li, 1986, *Scr. Metall* **20**, 1477-1482.

[23] R. Pippan, 1991, *Acta Metall. Mater.* **39**, 255-262.

[24] F.O. Riemelmoser, R. Pippan, H.P. Stüwe, 1997, *Int. J. Fract.* **85**, 157-168.

[25] A.J. Wilkinson, S.G. Roberts, P.B. Hirsch, 1998, *Acta Metall. Mater.* **46**, 379-390.

[26] V. Doquet, 1998, *Fatigue & Fracture of Engineering Materials & Structures*, **21**, 661-672.

[27] R. Pippan, F.O. Riemelmoser, 1998, *Kovove Materialy-Metallic Materials* **36**, 183-192.

[28] A.J. Wilkinson, S.G. Roberts, 1999, *Proc. of the 6th International Fatigue Congress, Fatigue '99*, eds. X.R. Wu and Z.G. Wang, p.678.

[29] R. Pippan, 1992, *Int. J. Fracture* **58**, 305-318.

[30] A.J. Wilkinson, S.G. Roberts, 1996, *Scripta Mater.* **35**, 1365-1371.

[31] F.O. Riemelmoser, R. Pippan, H.P. Stüwe, 1998, *Acta Mater.* **46**, 1793-1799.

[32] C. Bjerkén, S. Melin, 2003, *A tool to model short crack fatigue growth using a discrete dislocation formulation*, *Int. J. Fatigue* **25**, 559-566.

[33] C. Bjerkén, S. Melin, 2004, *A study of the influence of grain boundaries on short crack growth during varying load using a dislocation technique*, *Eng. Fract. Mech.* **71**, 2215-2227.

[34] P. Hansson, S. Melin, 2005, *Dislocation-based modeling of the growth of a microstrucurally short crack by single shear due to fatigue loading*, *Int. J. Fatigue* **27**, 347-356.

[35] E. Van der Giessen, V.S. Deshpande, Cleveringa, A. Needleman, 2001, *J. Mech. Phys. Solids* **49**, 2133-2153.

[36] V.S. Deshpande, A. Needleman, E. Van der Giessen, 2001, *Acta Mater.* **49**, 3189-3203

[37] R. Pippan, F.O. Riemelmoser, H. Weinhandl, H. Kreuzer, 2002, *Plasticity-induced crack closure under plane strain conditions in the near-threshold regime*, *Phil. Mag.* **82**, 3299-3309.
[38] H. Inoue, Y. Akahosi, S. Harada, 1995, *Mater. Sci. Res. Int.* **1**, 95-99.
[39] C. Bichler, R. Pippan, 2007, *Effect of single overloads in ductile metals: A reconsideration*, *Engineering Fracture Mechanics* **74**, 1344-1350.
[40] H. Tada, P.C. Paris, G.R. Irwin, 1985 *The Stress Analysis of Cracks Handbook*, St. Louis, Missouri: Del Research Corporation.
[41] J.O. Eshelby, 1956, *The continuum theory of lattice defects*, *Solid Stat. Phys.* **3**, 56.
[42] B. Budiansky, J.R. Rice, 1973, *Conservation laws and energy-release rates*, *J. Appl. Mech.* **40**, 201.
[43] R. Pippan, F.O. Riemelmoser, 1995, *Dislocation shielding of fatigue cracks*, *Z. Metallkunde* **86**, 823-826.
[44] R.O. Ritchie, 1988, *Mater. Sci. Engng* **103**, 15
[45] P. Neumann, 1969, *Acta Metall.* **17**, 1219-1225.
[46] R.M.N. Pelloux, 1998, *Eng. Fract. Mech.* **1**, 697-704.
[47] Ch. Bichler, R. Pippan, 1999, *Direct observation of the formation of striations*, in: *Engineering against fatigue*. Ed. J.H. Beynon et al. A.A. Balkema, Rotterdam, 211-218.
[48] B. Budiansky, W. Hutchinson, 1978, *J. Appl. Mech.* **45**, 267.
[49] H. Führing, W. Seeger, 1979, *Engng Fract. Mech.* **11**, 99.
[50] F.O. Riemelmoser, P. Gumbsch, R. Pippan, 2001, *Mater. Trans.* **42**, 2.
[51] F.O. Riemelmoser, R. Pippan, H.P. Stüwe, 1998 *Acta mater.* **46**, 1793.
[52] B.A. Bilby, A.H. Cottrell, K.H. Swinden, 1963, *Proc. Roy. Soc. London* **A272**, 304.
[53] P.K. Liaw, T.R. Leax, W.A. Logsdon, 1983, *Acta Metall.* **31**, 1581-1587.
[54] R. Pippan, 1998, *Phil. Mag.* **77**, 861-873.
[55] R.W. Hertzberg, 1993, *Int. J. Fracture* **64**, R53-58.
[56] R.W. Hertzberg, 1995, *Mater. Sci. Eng.* **A190**, 25-32.
[57] P. Hansson, S. Melin, 2008, *Engng. Fract. Mech.*, **75**, 1400-1411

Discrete Dislocation Plasticity Analysis of Cracks and Fracture

Erik Van der Giessen
Dept. of Applied Physics, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

1 Introduction

Fracture in plastically deforming crystals involves several length scales, as illustrated in Fig. 1 for cleavage-like crack growth. The relevant length scales range from that of the macroscale object to the atomic scale, including the various microstructural length scales in between that are associated with, for example, particles, grains, and defect structures.

At the large length scale of the macroscopic world, Fig. 1a, plastic deformation is conveniently described by a phenomenological continuum theory. The stress field near the tip of a mathematically sharp crack tip then is singular at the tip, Fig. 1b. At the scale where the polycrystalline nature is revealed, Fig. 1c, plastic deformation is a physical process that is inherently inhomogeneous and anisotropic. This is caused by the fact that each grain is anisotropic with a finite number of slip systems on which glide can take place. When zooming in further, one will see that plastic deformation within each grain involves the collective motion of many dislocations, Fig. 1d. Finally, the finest scale shown in Fig. 1e governs where atomic bonds are broken upon crack propagation.

The challenge in understanding fracture lies in the fact that all scales are connected and all may contribute to the total fracture energy. It is worth emphasizing that although the atomistics of the separation of surfaces may only contribute a small fraction of the total energy release rate, it can still be controlling. This is because dissipative mechanisms can only operate if fracture is delayed sufficiently to allow them to come into play. Indeed, as pointed out by Rice and Wang [28], the surface energy can play a valve-like role. Surface energies are typically of the order of 1 J/m^2, while fracture energies for ductile crystalline metals are often an order of magnitude higher. The difference between the fracture energy and the surface energy is the plastic dissipation in the vicinity of the crack tip. Many details are left out in the above discussion, but it emphasizes that fracture, i.e. the creation of new surface, is highly localized at the atomic scale, but is driven by

the macroscopic applied load communicated to the atomic scale via stress fields on smaller and smaller length scales. It is the precise communication down these scales which determines whether or not crack growth occurs and how much energy is dissipated.

Much is known about the near crack-tip fields at the continuum scale. At the polycrystalline scale, and if the average response is isotropic, the smooth HRR plastic fields developed by Hutchinson [18] and by Rice and Rosengren [26] are dominant, while the fields change to a completely different nature, characterized by piece-wise uniform sectors [27, 29, 30], when the crack tip is contained in a single crystal. At the latter scale of observation, i.e. in between Fig. 1c and d, the theories predict near-tip stresses that are a few times the yield strength. This is far below the atomic bond strength, thus raising the 'paradox' that crack growth

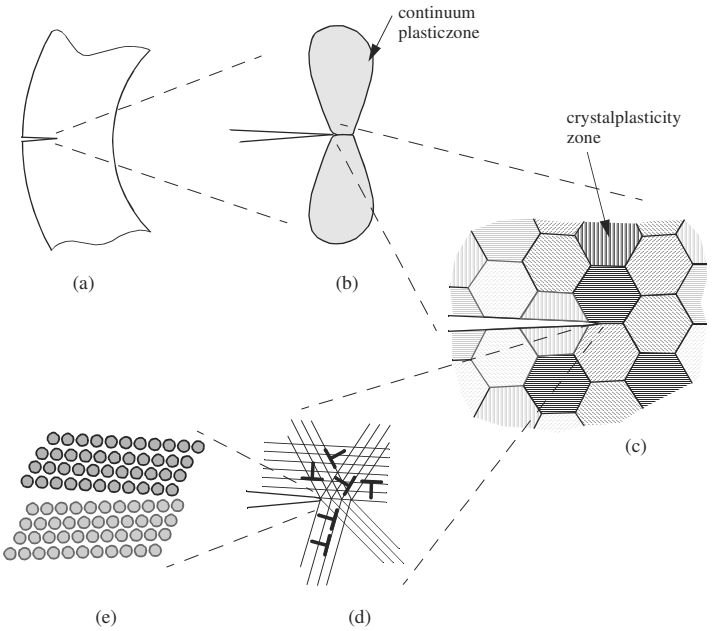


Figure 1. The various relevant scales that may determine the response of a crack in a macroscopic component. (a) The component scale. (b) The plastic zone governed by macroscopic continuum plastic flow. (c) The grain scale in a polycrystalline metal. (d) The scale of discrete slip planes and of individual dislocations. (e) The atomic scale.

could not occur in plastically deforming materials. Experience proves us wrong. This chapter intends to present insight in resolving this, principally by describing plastic deformation by the generation and motion of discrete dislocations, i.e. the size scale of Fig. 1d. An introduction to *discrete dislocation plasticity* may be found in [5]; a short summary is included in the first two sections.

2 Elastic Models of Dislocations

2.1 General Idea

A dislocation is a line defect in an atomic lattice ([17, 16]). It is a line on an atomic plane that separates those regions of the plane that are intact from regions where the lattice has undergone slip. The relative shift of atomic planes is in the direction denoted by the Burgers vector $\boldsymbol{b}$ and is essentially uniform in the region enclosed by the dislocation. The dislocated material can be constructed in a thought experiment from a perfect crystal by a cut-displace by $\boldsymbol{b}$-and-reweld procedure. Because of conservation of mass, a dislocation is a closed loop.

The geometry of a dislocation is governed by a number of variables:

- the slip plane, denoted with its unit normal vector $\boldsymbol{m}$;
- the dislocation line as a parameterized line on this plane and with a local tangent vector $\boldsymbol{t}$;
- the Burgers vector $\boldsymbol{b}$.

There are a few special parts of a generic loop, namely

$$\text{edge:} \quad \boldsymbol{b}\cdot\boldsymbol{t} = 0\,; \tag{1}$$

$$\text{screw:} \quad \boldsymbol{b}\cdot\boldsymbol{t} = \pm b\,, \tag{2}$$

b being the length of $\boldsymbol{b}$: $b = |\boldsymbol{b}|$. Edge and screw dislocations are the central notions in two-dimensional studies, see Fig. 2. As the crack problems studied later on in this chapter are concerned with mode I, plane strain conditions, attention will be focused on edge dislocations.

Although a dislocation is a lattice defect, it has proved very useful to describe it in the framework of continuum theory in which the atomic positions are averaged out. This reduces the total number of degrees of freedom enormously: from all atom positions to a mathematical, functional form of the geometric variables $\boldsymbol{m}$, $\boldsymbol{t}$ and $\boldsymbol{b}$. In a continuum framework the definition of the Burgers vector becomes

$$\boldsymbol{b} = \oint_C \frac{\partial \boldsymbol{u}}{\partial c} dc \tag{3}$$

where C is a closed circuit around the dislocation, traversed by local coordinate c and $\boldsymbol{u}$ is the displacement field away from the perfect crystal. The real key to

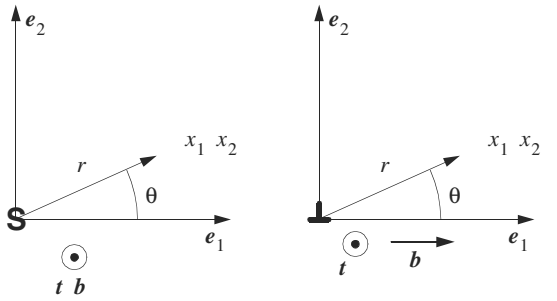


Figure 2. Definitions of (a) screw and (b) edge dislocation configurations discussed in the text.

dislocations in a continuum description is that it involves a *displacement discontinuity* inside the dislocation loop. It is the expansion of dislocation loops that creates what we observe on a larger scale as *permanent*, that is, *plastic deformation*. Apart from the dislocation motion, the distortion of the lattice is entirely elastic. Thus, the picture of plasticity emerges of dislocation loops sweeping through an otherwise elastic continuum.

The current continuum theory of discrete dislocations employs linear elasticity. Clearly, it will break down inside the core region, where the strains will be too large for the linear approximation to hold. Away from the core by about 5 to $6b$, comparison with atomistics has shown that the linear elastic solution is very accurate. Hence, discrete dislocation plasticity holds the view that the fields in a dislocated body can be described by linear elasticity, excluding the core regions. The use of linear elasticity has the enormous advantage that many solutions for the governing equations are known and that one can use superposition. The latter will be exploited in full power in Sec. 3.

For completeness, we recall that the governing equations for linear elasticity are

$$\text{equilibrium} \quad : \sigma_{ij,j} = 0 \tag{4}$$

$$\text{elasticity} \quad : \sigma_{ij} = \mathcal{L}_{ijkl}\varepsilon_{kl} \tag{5}$$

$$\text{strains} \quad : \varepsilon_{ij} = \frac{1}{2}(u_{i,j} + u_{j,i}) \tag{6}$$

with $\mathcal{L}_{ijkl}$ the elastic moduli. Since the crystals we shall consider here are cubic, we should be using the cubic elastic moduli expressed in terms of the usual C_{11}, C_{12} and C_{44}. However, for simplicity, we will assume isotropic elasticity, with the

moduli expressed either in terms of Young's modulus E and Poisson's ratio ν,

$$\mathcal{L}_{ijkl} = \frac{E}{1+\nu}\left[\frac{1}{2}\left(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}\right)+\frac{\nu}{1-2\nu}\delta_{ij}\delta_{kl}\right] \tag{7}$$

or as

$$\mathcal{L}_{ijkl} = \mu\left(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}\right)+(k-\frac{2}{3}\mu)\delta_{ij}\delta_{kl}\,. \tag{8}$$

in terms of the shear modulus $\mu = E/2(1+\nu)$ and the bulk (or compression) modulus $k = E/3(1-2\nu)$.

2.2 Edge Dislocations

An *edge dislocation* poses a plane strain problem. If the dislocation line is arranged with its line direction perpendicular to the x_1–x_2 plane of consideration, i.e. $\boldsymbol{t} = \boldsymbol{e}_3$, the Burgers vector lies in the x_1–x_2 plane, Fig. 2b. Specifically, we shall assume again that the x_1–x_3 plane is the slip plane and that $\boldsymbol{b}$ points in the positive x_1 direction, see Fig. 2b. This problem is then conveniently solved by application of the Airy stress function approach in a manner that is similar to that leading to the asymptotic singular field near a sharp crack tip. Leaving the details to [16], the solution reads

$$\sigma_{11} = -\frac{\mu b}{2\pi(1-\nu)}\frac{x_2(3{x_1}^2+{x_2}^2)}{({x_1}^2+{x_2}^2)^2} \tag{9}$$

$$\sigma_{22} = \frac{\mu b}{2\pi(1-\nu)}\frac{x_2({x_1}^2-{x_2}^2)}{({x_1}^2+{x_2}^2)^2} \tag{10}$$

$$\sigma_{12} = \frac{\mu b}{2\pi(1-\nu)}\frac{x_1({x_1}^2-{x_2}^2)}{({x_1}^2+{x_2}^2)^2} \tag{11}$$

for the in-plane stress components, while $\sigma_{33} = \nu(\sigma_{11}+\sigma_{22})$. As for a screw dislocation, the stress field is singular at the dislocation and decays with the distance r as $1/r$. The displacement field is given by

$$u_1 = \frac{b}{2\pi(1-\nu)}\left[\frac{1}{2}\frac{x_1x_2}{{x_1}^2+{x_2}^2}-(1-\nu)\arctan\left(\frac{x_1}{x_2}\right)\right] \tag{12}$$

$$u_2 = \frac{b}{2\pi(1-\nu)}\left[\frac{1}{2}\frac{{x_2}^2}{{x_1}^2+{x_2}^2}-\frac{1}{4}(1-2\nu)\ln\frac{{x_1}^2+{x_2}^2}{b^2}\right] \tag{13}$$

The u_1 field, i.e. parallel to the slip plane, is shown in Fig. 3.

Note that the solution for the edge dislocation field assumes the body to be infinitely large – no boundary conditions have been incorporated. Since the solution is singular, as the solution for a screw dislocation, it is to be expected that a correction due to boundary conditions has no significant effect close to the dislocation, but it will do so at larger distances.

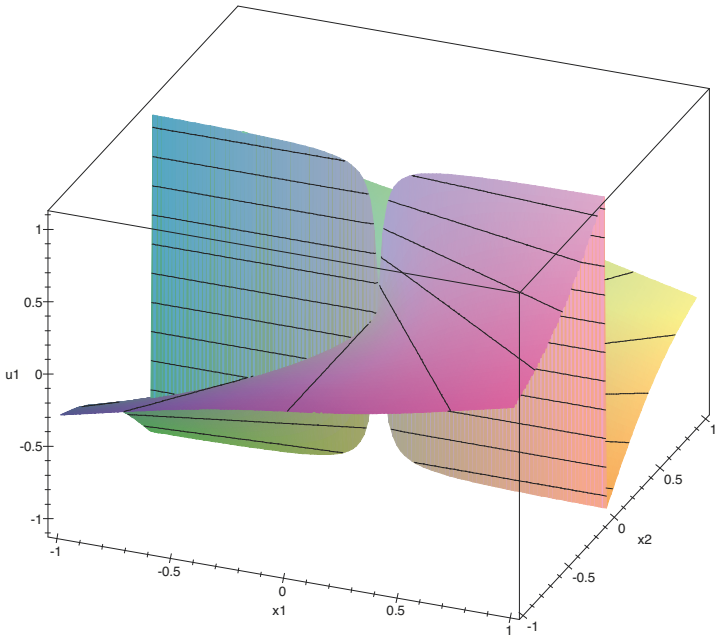


Figure 3. Displacement field $u_1(x_1,x_2)$, normalized by $b/2\pi(1-\nu)$, of an edge dislocation with $\boldsymbol{b} = b\boldsymbol{e}_1$ at $(x_1,x_2) = (0,0)$.

3 Boundary Value Problems

Even for straight two-dimensional dislocations, as discussed in the previous section, known closed-form solutions do not incorporate any boundaries of the crystal. The solutions are, strictly speaking, for dislocations in infinite space. Interactions with the boundaries of course do exist, and their are commonly [16] referred to as *image effects*. Rather clever image constructions have been developed but they remain limited to particular configurations.

Several years ago, Van der Giessen and Needleman [33] proposed a versatile approach based on superposition. The idea is to make use of the known solutions in infinite space and to superpose an 'image' solution to correct for the boundary conditions. To this end, the displacement, strain and stress fields are decomposed as

$$\boldsymbol{u} = \tilde{\boldsymbol{u}} + \hat{\boldsymbol{u}}, \quad \boldsymbol{\varepsilon} = \tilde{\boldsymbol{\varepsilon}} + \hat{\boldsymbol{\varepsilon}}, \quad \boldsymbol{\sigma} = \tilde{\boldsymbol{\sigma}} + \hat{\boldsymbol{\sigma}}. \tag{14}$$

The $\tilde{(\,)}$ fields are the superposition of the singular fields of the individual disloca-

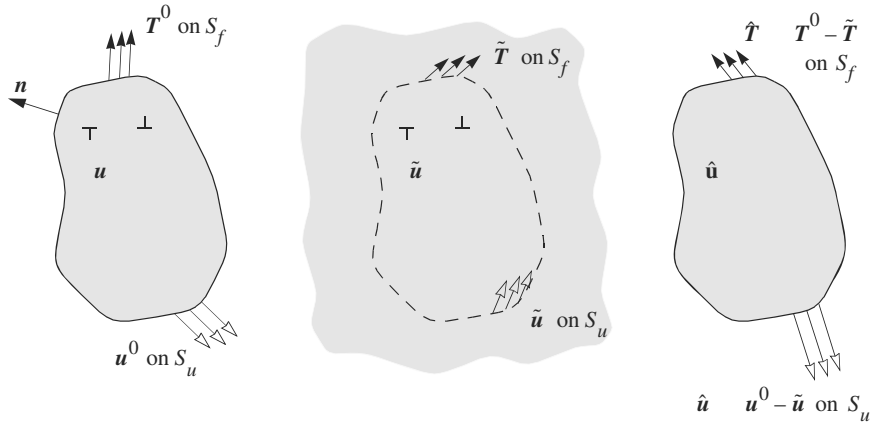


Figure 4. Decomposition of the problem for the dislocated body into the problem of interacting dislocations in an infinite solid ($\tilde{}$ fields) and the complementary problem for the body without dislocations ($\hat{}$ fields).

tions, as discussed above, in infinite space. Identifying the fields for dislocation I by a superscript (I), the $(\tilde{})$ stress field, for example, is obtained as

$$\tilde{\boldsymbol{\sigma}} = \sum_I \boldsymbol{\sigma}^{(I)}$$

The $(\tilde{})$ fields will in general not meet the boundary conditions in terms of tractions on part S_f of the boundary nor the prescribed displacements on part S_u. Instead they will give rise to displacements $\tilde{\boldsymbol{u}}$ on S_u and tractions $\tilde{\boldsymbol{T}} = \tilde{\boldsymbol{\sigma}} \cdot \boldsymbol{n}$ on S_f (with normal $\boldsymbol{n}$). The actual boundary conditions, $\boldsymbol{u}^0$ on S_u and $\boldsymbol{T}^0$ on S_f, are imposed through the $(\hat{})$ fields, in such a way that the sum of the $(\tilde{})$ and the $(\hat{})$ fields in (14) gives the solution that satisfies all boundary conditions. Since the $(\tilde{})$ fields satisfy the governing equations (4)–(6), the $(\hat{})$ fields also have to satisfy the elasticity equations, i.e.

$$\begin{aligned} \text{equilibrium:} \quad & \operatorname{div} \hat{\boldsymbol{\sigma}} = 0 \\ \text{elasticity:} \quad & \hat{\boldsymbol{\sigma}} = \boldsymbol{\mathcal{L}} : \hat{\boldsymbol{\varepsilon}} \\ \text{strains:} \quad & \hat{\boldsymbol{\varepsilon}} = \frac{1}{2}[\operatorname{grad} \hat{\boldsymbol{u}} + (\operatorname{grad} \hat{\boldsymbol{u}})^T] \end{aligned}$$

supplemented with the boundary conditions

$$\hat{\boldsymbol{T}} = \boldsymbol{T}^0 - \tilde{\boldsymbol{T}} \text{ on } S_f \qquad \hat{\boldsymbol{u}} = \boldsymbol{u}^0 - \tilde{\boldsymbol{u}} \text{ on } S_u$$

It is important to note that the solution of the $(\hat{})$ problem does not involve any dislocations. Therefore, the $(\hat{})$ fields (often called 'image' fields) are smooth and the boundary value problem for them can conveniently be solved using a finite element method.

A crucial concept for the evolution of dislocations, to be discussed in the next section, is the *Peach–Koehler* force. It is defined as the configurational force associated to motion of the dislocations: the work of these forces as the dislocations move is the change in *potential energy* Π, i.e.

$$\delta\Pi = -\sum_I \oint_{\mathcal{L}^{(I)}} \boldsymbol{f}^{(I)} \cdot \delta \boldsymbol{s}^{(I)} dl\,.$$

It is seen that $\boldsymbol{f}^{(I)}$ has the dimension of force per unit (dislocation line) length and that it generally changes along the loop. The Peach–Koehle force can be made more explicit, in this superposition approach, by calculating the potential energy [33]. The final result is that the component of the Peach–Koehler force in the glide plane and in the direction $\boldsymbol{t}^{(I)} \times \boldsymbol{m}^{(I)}$ normal to the dislocation can be expressed as

$$f^{(I)} = \boldsymbol{m}^{(I)} \cdot \left(\hat{\boldsymbol{\sigma}} + \sum_{J \neq I} \boldsymbol{\sigma}^{(J)} \right) \cdot \boldsymbol{b}^{(I)}\,. \tag{15}$$

4 Dislocation Dynamics

So far, we have discussed the state of the material in the presence of dislocations, but they have to move in order to produce plastic deformation. In this section, we give a brief summary of various physical phenomena that govern the motion of dislocations, focusing on those aspects that are connected to *dislocation glide* of *straight edge dislocations*. Climb, i.e. motion perpendicular to the plane, also occurs under certain circumstances but will not be treated, nor will cross slip (dislocation motion out of the original glide plane).

The driving force for glide is the Peach–Koehler force component given by (15). During glide, however, the dislocation may be subjected to various sources of resistance against motion. Denoting their collective force by $f^{(I)}_{\text{resist}}$ and ignoring inertia of dislocations, the motion of dislocation I is be governed by the force balance

$$f^{(I)} = f^{(I)}_{\text{resist}}\,. \tag{16}$$

The two main contributions to the resistance of edge dislocations are:

- *Peierls–Nabarro stress.* As a dislocation moves, existing atomic bonds have to be broken and new ones formed. Because of the periodic nature of the lattice, the energy landscape that the dislocation moves across is periodic.

This acts as a *friction stress*, referred to as the Peierls–Nabarro stress, against the initiation of dislocation motion, corresponding to a resistance force of $b\tau_{\mathrm{f}}$ in the direction opposing the dislocation velocity $v^{(I)}$. The value of τ_{f} depends sensitively on the crystal structure. At 0 K, the values for FCC and BCC crystals typically are $10^{-6}\mu$ and $10^{-3}\mu$, respectively, while for intermetallics its value can be as large as 0.1μ.

- *Drag.* As a dislocation glides, it experiences drag originating from: (i) phonon drag; (ii) electron drag; (iii) impurity effects. Of these, phonon drag is dominating in many materials and gives a net, viscous force $Bv^{(I)}$. It is possible to make estimates of each of the contributions to phonon drag, but in practice B is measured experimentally or from molecular dynamics simulations. Values of B show quite some scatter; a typical value for aluminum is on the order of $B = 10^{-4}$ Pa s. When drag is the only source of resistance, the balance (16) states

$$f^{(I)} = Bv^{(I)}$$

so that the velocity, for a given Peach-Koehler force, can be calculated from

$$v^{(I)} = f^{(I)}/B \tag{17}$$

A linear relationship between velocity and driving force has been confirmed experimentally. Data for various materials over various temperatures and stress levels suggest a relation of the type

$$v^{(I)} \propto (f^{(I)})^m e^{-E/k_{\mathrm{B}}T}$$

but for relatively low stress levels, the exponent m is close to unity.

4.1 Annihilation

Attraction of opposite-signed dislocations can lead to *annihilation*. This is easily seen for straight edge or screw dislocations of opposite sign coming close together: the defect is eliminated by coalescence of the dislocations once they are sufficiently close together. The critical annihilation distance is typically taken to be $6b$.

4.2 Frank–Read sources

One possible mechanism for the generation of new dislocations is through the so-called Frank–Read mechanism illustrated in Fig. 5a. The source of this mechanism is a dislocation segment that is being pinned between two hard points. This segment bows out under the influence of a Peach-Koehler force, where the self-interaction helps to create a closed loop surrounding the initial pinned segment that finally closes onto itself. Ultimately, a smooth new dislocation loop has emerged

as well as a copy of the initial segment. This copy can operate as a source again, etc., so as to generate more dislocations.

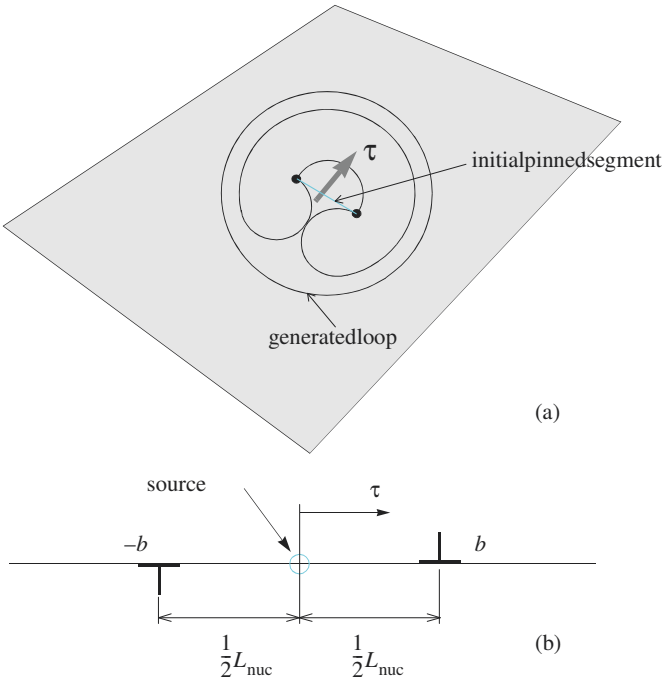


Figure 5. (a) Frank–Read mechanism of generating a new loop from an initial pinned segment. (b) Two-dimensional simplification (from [33]) as a mechanism for nucleation of edge dislocations.

Figure 5 shows a two-dimensional version of this mechanism, which one can imagine as a cross-section of a three-dimensional source and subsequent projection onto the plane of observation normal to the slip plane. The pinned segment now appears as a point, and the Frank-Read mechanism has been translated into two dimensions by [33] as follows. When the shear stress on the source is sufficiently high for a sufficiently long time, an edge dipole is generated. The strength of the source τ_{nuc} is, in principle, determined by the three-dimensional dislocation configuration (initial segment length, Burgers vector, etc.); in the two-dimensional model it becomes a parameter. The same holds for the nucleation time t_{nuc}. The polarity of the dipole is determined by the direction of the shear stress. The width of the dipole, L_{nuc}, is the two-dimensional cross-section of the loop at the moment

that the initial pinned segment has been re-formed and the source is ready to operate again. The value of L_{nuc}, again, can be determined from three-dimensional simulations. However, a different and practical criterion has been proposed by [33] based on the observation that the dislocations in a dipole feel strong attractive forces that are inversely proportional to their distance. If this interactive force is larger than $\tau_{\rm nuc} b$, the dislocations will move towards each other and annihilate so that effectively no dislocation has been generated. Hence, there is a minimum distance which is given by [33]

$$L_{nuc} = \frac{\mu b}{2\pi(1-\nu)\tau_{nuc}} .$$

5 Methodology for Crack Problems

The analysis of cracks and fracture within the framework of discrete dislocation plasticity employs the generic mode I problem sketched in Fig. 6. The calculations are carried out for small-scale yielding, with plasticity being confined to a window around the initial crack tip. Single-crystal studies (Secs. 6, 7) are performed by defining a set of two or three slip systems inside this window at an angle of $\phi^{(\alpha)}$ with respect to the crack plane. Two slip systems are necessary to allow for any mode of plastic deformation, while three slip systems mimic the excess of available slip systems in a real three-dimensional FCC crystal. The process window is filled with a number of grains for the study of polycrystals in Sec. 8

Because of the assumed symmetry, there is a mirror dislocation for each dislocation in the region analyzed numerically. This mirror dislocation does not need to be accounted for explicitly when superimposing the fields of all dislocations, for example as in the sum in (15). Rather, its presence is accounted for through the symmetry boundary conditions. What does need to be accounted for in the dislocation analysis is that when a dislocation crosses the closed crack plane, it leaves the region analyzed; but, due to symmetry, a mirror dislocation enters into the system along the mirror slip plane.

The crack is initially sharp and a cohesive surface is laid out in front of it. At the scale of interest here, the cohesive surface is taken to mimic atomic debonding. Therefore, the constitutive response of the cohesive surface is taken from the universal binding law Rose *et al.* [31] and is specified by the following relation between the traction normal to the cohesive surface, T_n, and the separation Δ_n:

$$T_n(\Delta_n) = \sigma_{\max} \frac{\Delta_n}{\delta_n} \exp\left(1 - \frac{\Delta_n}{\delta_n}\right) , \tag{18}$$

As the cohesive surface separates, the magnitude of the traction increases, reaches a maximum and then approaches zero to represent the formation of a traction-free

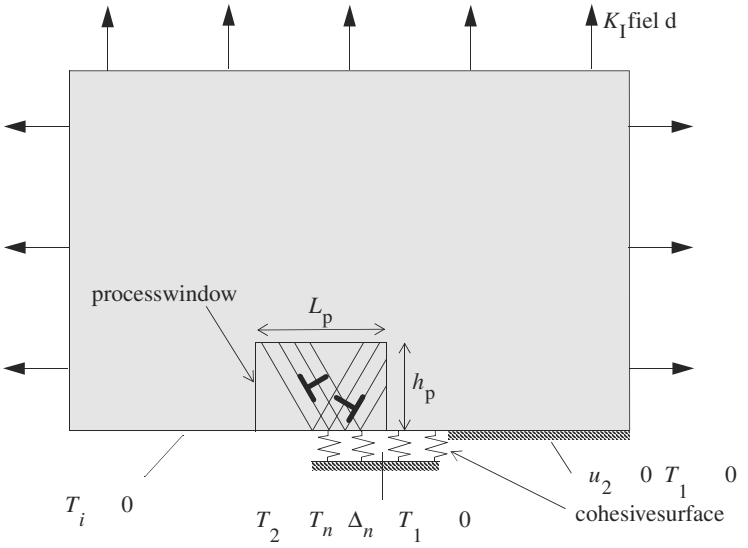


Figure 6. Small-scale yielding analysis under mode I conditions with discrete dislocations moving inside a process window. Because of symmetry, only half the problem needs to be analyzed. The cohesive surface ahead of the initial crack is used to describe crack growth.

crack, see Fig. 7. The strength $\sigma_{\max}$ and the corresponding separation δ_n characterize the fracture process, implying a work of separation $\phi_n = \exp(1)\sigma_{\max}\delta_n$. It is essential to note that the use of a cohesive surface eliminates the need of a fracture criterion: whether or not crack growth occurs is an outcome of the solution of the problem.

When studying a stationary crack, we take the value of $\sigma_{\max}$ to be very large so as to avoid significant opening (yet, the purely elastic stress singularity of a mathematically sharp crack is always removed). Studies of crack growth by a cleavage mechanism are carried out by taking values $\sigma_{\max} = 0.6\,\text{GPa}$ and $\delta_n = 4b$, giving a work of separation $\phi_n = 1.63\text{J/m}^2$. This value of the cohesive strength is about a factor of four smaller than the expected theoretical strength of aluminum and is used for numerical reasons because: (i) the length scale over which large gradients occur is inversely proportional to the cohesive strength, so that a finer mesh is required for higher values of the cohesive strength; and (ii) the number of dislocations increases with increasing cohesive strength, so that more dislocation interactions have to be computed and a larger process window is needed.

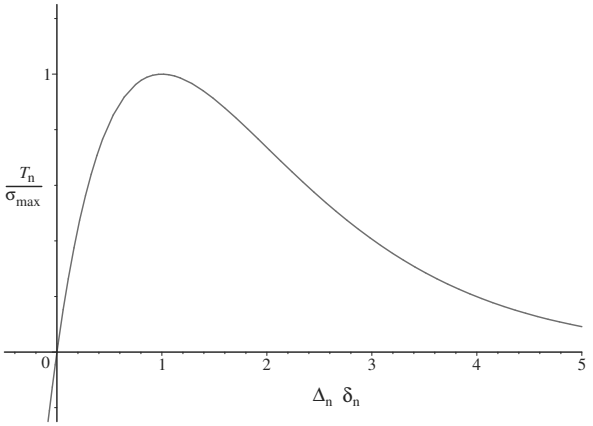


Figure 7. Traction–separation law corresponding to the universal binding law, according to (18).

The boundary conditions for the problem sketched in Fig. 6 are: (i) the crack faces remain traction free, $T_i = 0$; (ii) the displacements on the remote boundary are specified according to the well-known elastic singular field; and (iii) for the symmetric mode I loading cases discussed here (18) is satisfied together with $T_1 = 0$ on the crack plane ahead of the initial crack tip. The load level is thus characterized by the remote stress intensity factor K_I. In the absence of dislocation motion, the critical energy release rate is equal to ϕ_n, from which we define the reference intensity factor K_0 by

$$K_0 = \sqrt{\frac{E\phi_n}{1-\nu^2}}\,.$$

Because of the cohesive law (18), the problem sketched in Fig. 6 is nonlinear, and is solved in an incremental manner by phrasing the governing equations in rate form. Assuming that at time t the stress field and the current positions of all dislocations are known, the incremental (denoted by a superposed dot) ($\hat{}$)–fields are governed by the virtual work statement

$$\int_V \hat{\sigma}_{ij}\delta\varepsilon_{ij}\,dV + \frac{1}{2}\int_{S_{\rm coh}} k_n\left(\tilde{\Delta}_n^{(t+\Delta t)} + \hat{\Delta}_n^{(t)}\right)\hat{\Delta}_n\delta\Delta_n\,dS =$$
$$-\frac{1}{\Delta t}\left[\int_V \hat{\sigma}_{ij}^{(t)}\delta\varepsilon_{ij}\,dV - \frac{1}{2}\int_{S_{\rm coh}} T_n\left(\tilde{\Delta}_n^{(t+\Delta t)} + \hat{\Delta}_n^{(t)}\right)\delta\Delta_n\,dS\right] \qquad (19)$$

as shown by Cleveringa *et al.* [8]. Here, the instantaneous cohesive stiffness, $k_n = -\partial T_n/\partial \Delta_n$, and the instantaneous normal traction are evaluated at the fictitious opening $\tilde{\Delta}_n^{(t+\Delta t)} + \hat{\Delta}_n^{(t)}$. The factor $1/2$ in (19) stems from the fact that, by virtue of symmetry, only half of the work in the cohesive surface contributes to the work in the region analyzed. The second integral in the left-hand side provides a contribution to the overall finite-element stiffness matrix that changes every time step; but computational costs can be reduced by factorizing the matrix except the part pertaining to the degrees of freedom connected to the cohesive elements.

6 Cracks in Single Crystals

A set of two or three slip systems is defined inside the window at an angle of $\phi^{(\alpha)}$ with respect to the crack plane. Two orientations are considered, which are an approximation of the projections of three-dimensional orientations of FCC and BCC crystals respectively that lead to plane-strain plastic deformations [27]:

- FCC: $\phi^{(1,2)} = \pm 60^\circ$ (or, more exactly $\pm 54.6^\circ$), and $\phi^{(3)} = 0$ when considering three slip systems;
- BCC: $\phi^{(1,2)} = \pm 30^\circ$ (or, more exactly $\pm 35.3^\circ$), and $\phi^{(3)} = 90^\circ$ when considering three slip systems;

The calculations to be presented in subsequent sections do not aim at modeling a specific material, but properties representative of aluminum are used. The elastic properties are taken to be isotropic, with Young's modulus $E = 70\,\mathrm{GPa}$ and Poisson ratio $\nu = 0.33$. A representative value for the drag coefficient in (17) is $B = 10^{-4}\,\mathrm{Pa\,s}$ [19]. Unless otherwise noted, the strength of the dislocation sources is randomly chosen from a Gaussian distribution with mean strength $\bar{\tau}_{\mathrm{nuc}} = 50\,\mathrm{MPa}$ and standard deviation $0.2\bar{\tau}_{\mathrm{nuc}}$. The nucleation time for all sources is taken as $t_{\mathrm{nuc}} = 10$ ns. All obstacles are taken to have the same strength $\tau_{\mathrm{obs}} = 150$ MPa. Attention is focused on metal crystals with a relatively high density of initial defects, which are modeled by a random distribution of point sources and obstacles in the process window; there are no initial dislocations on the active slip systems in the simulations. Also, there is no special nucleation of dislocations from the crack tip.

6.1 Stationary Crack-Tip Fields

Figure 8 shows the stress distribution for an FCC crystal with three slip systems, and with such a high cohesive strength that the crack does not propagate. All three stress distributions exhibit large fluctuations, which are due to the singularities of the individual dislocations. In fact, the fluctuations shown are damped because of the way the contours are plotted on the finite element mesh that was used for the computation (80 by 80 elements in the process window).

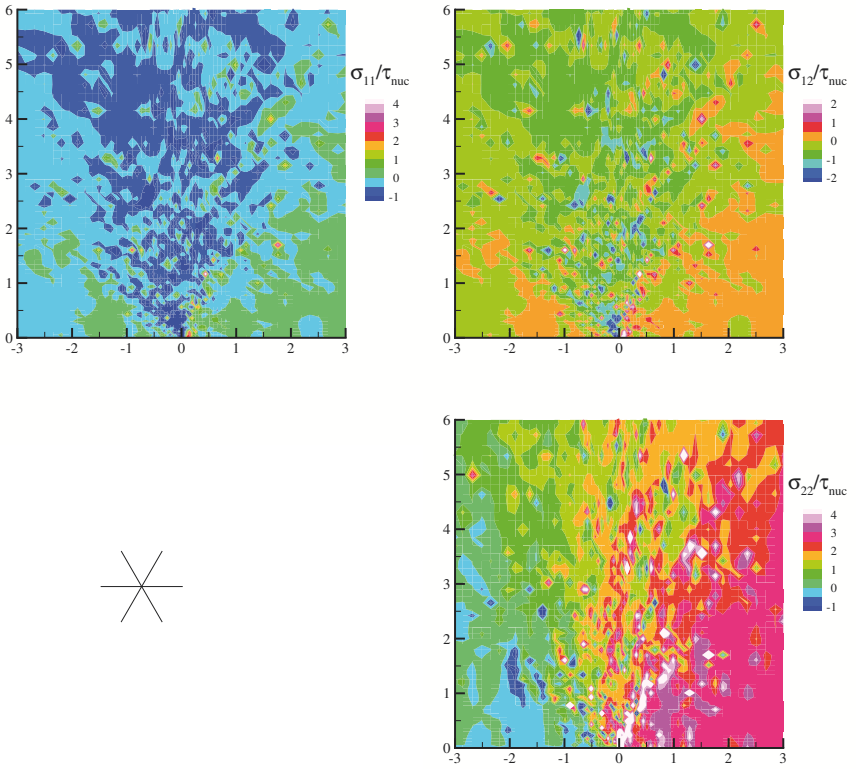


Figure 8. Stress distributions, normalized by the nucleation strength τ_{nuc}, in an FCC crystal with three slip systems as indicted in the inset at $K_I = 0.6\text{MPa}\sqrt{\text{m}}$. From [34].

It is remarkable, however, that three sectors appear around the crack tip in which the stresses, on average, look different from one another. This stress distribution is reminiscent of the analytical near-tip stress field obtained by Rice [27] on the basis of a continuum plasticity theory for non-hardening crystals. When his analysis for the true FCC crystal geometry is modified to account for the set of slip systems used, four uniform stress sectors are obtained. The boundaries between these sectors are predicted to be 60°, 90° and 120°, the first and last of which are consistent with the fields in Fig. 8. Van der Giessen *et al.* [34] carried out a quantitative comparison by actually averaging the stresses inside the four mentioned

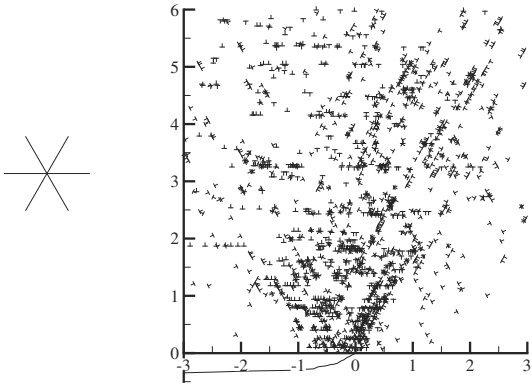


Figure 9. Dislocation distribution in the same crystal and at the same load level as in Fig. 8. From [34].

sectors, and it was found that these average stresses agreed quite well with Rice's continuum solution.

However, the discrete dislocation solution did not appear to agree with another element of Rice's [27] solution, namely that slip activity on the 0°-slip planes would concentrate in a kink band at 90°. The discrete dislocation results showed no evidence of this, as illustrated by the dislocation distribution shown in Fig. 9. The explanation for the absence of kink bands is that they would require an abundance of sources. Subsequently, Drugan [30] carried out an analysis similar to Rice's [27], but without requiring a kink band. He found several solution families, including a family of solutions which involve only a slip band at $\theta = 60°$; consistent with our discrete dislocation simulations. The solution that is closest to the discrete dislocation results is one where there are three sectors with boundaries at 60° and 105°. This solution is illustrated in Fig. 10. Even though the 105° sector boundary is not obvious from Fig. 8, averaging of the stress fields over these sectors showed very good agreement with this continuum prediction. Small differences in the exact average stress values are attributed to the fact that the continuum solution assumes no hardening, whereas some degree of hardening may occur in the discrete dislocation results.

6.2 Crack Propagation under Monotonic Loading

When averaging the discrete dislocation results over sectors to compare with the continuum plasticity predictions [34], the very near tip region with a radius of

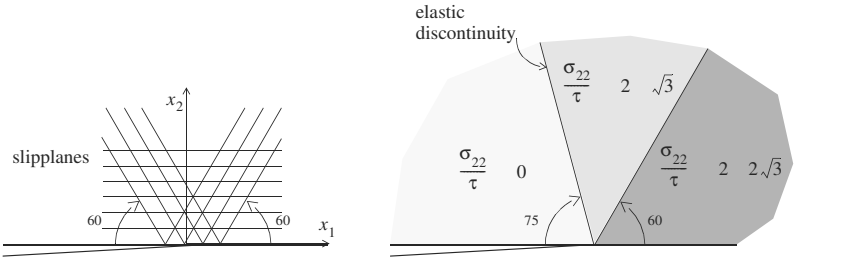


Figure 10. Opening stress states (σ_{22} normalized by the critical resolved shear stress τ) in the three sectors of Drugan's [30] continuum solution.

0.5μm was excluded. The reason is that the stresses in this region are much higher than the sector averages, see Fig. 8. In fact, the results of Cleveringa *et al.* [8] suggest that the stresses in this region can become high enough for crack advance by cleavage (although, for numerical reasons, the cohesive strengths in the calculations are smaller than representative of actual metal cohesive strengths). This is illustrated in Fig. 11 for a case with two slip systems ($\pm 60^\circ$) and with a cohesive surface characterized by the values $\sigma_{\max} = 0.6\,\mathrm{GPa}$ and $\delta_n = 4b$. For these parameter values, the stationary crack tip blunts because of dislocation activity, Fig. 11a. The sector-average stresses at this instant are quite low, but the opening stress σ_{22} in a small region ahead of the crack reaches the cohesive strength. The crack then propagates until the crack tip arrives at a location where the near-tip opening stress is below the cohesive strength. Then, more dislocations are generated near the current tip, until the opening stress again reaches the strength, Fig. 11b, and the crack jumps forward again. This process of blunting and crack jumping continues as the load increases, giving rise to a distinct *R*-curve behavior.

It is worth emphasizing that in the calculations by Cleveringa *et al.* [8] there is no emission of dislocations from the crack tip. This is in contrast to simulations, e.g. by Hirsch and Roberts [13] and by Nitzsche and Hsia [14], where it is assumed that dislocations can be emitted only by the crack tip. The same assumption has been made in the analyses of mode III cracks by Zacharopoulos *et al.* [15]. This class of calculations aim at initially dislocation-free materials, such as silicon, where crack-tip emission is the key parameter in the transition from brittle fracture (no dislocations) to ductile fracture accompanied by dislocation motion, see also [12]. However, the model considered here is intended to mimic a metal in which there is an initial distribution of dislocations that act as Frank-Read sources or as forest dislocation obstacles.

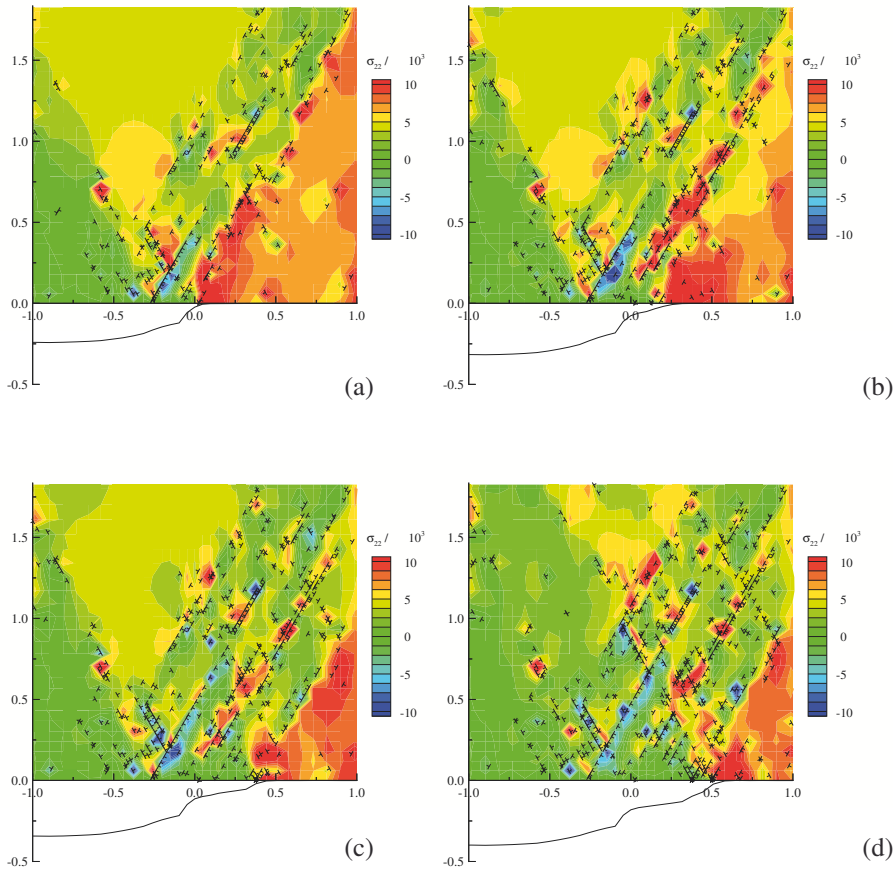


Figure 11. Distribution of dislocations and the opening stress σ_{22} in the immediate neighborhood (2μm $\times$ 2μm) of the crack tip for the FCC crystal with $\rho_{\text{nuc}} = 49/\mu\text{m}^2$ and $\rho_{\text{obs}} = 98/\mu\text{m}^2$ at four different stages of loading. The corresponding crack opening profiles (displacements magnified by a factor of 10) are plotted below the x_1-axis. From [8].

The main conclusion from the studies of Cleveringa *et al.* [8] is that dislocations play a dual role in fracture. On the one hand, dislocations are the vehicle for plastic deformation, and this reduces, on average, the stresses near the crack and provides a way to dissipate the energy flowing to the crack. On the other hand, dislocations can arrange themselves in structures, which lead to locally enhanced

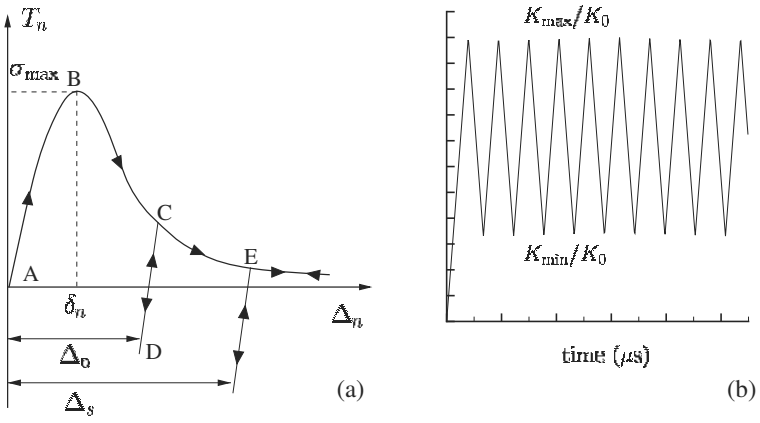


Figure 12. (a) Irreversible cohesive law used for fatigue calculations. (b) Schematic of the applied stress intensity factor as a function of time. From [10].

stress levels that can trigger crack growth. This stress enhancement is not modeled by conventional continuum plasticity.

7 Fatigue Crack Growth

Deshpande *et al.* [10] have extended the computations reported in the previous section to cyclic loading, specified by a remote stress intensity factor that zig-zags between K_{max} and K_{min}, Fig. 12b. The only change in the model is that the cohesive law is extended to be irreversible, as indicated in Fig. 12a, and which models the effect of complete oxidation of newly formed surface. This is a second source of irreversibility, next to the discrete dislocation plasticity, which is necessary for fatigue to occur.

Figure 13 summarizes the salient findings by Deshpande *et al.* [9, 10]:

1. The maximum stress intensity factor K_{max} needs to be high enough, i.e. above a critical value K^*_{max}, in order to provide a minimum of dissipation. For $K_{max} \gg K^*_{max}$, interactions within the now dense dislocation structure act to retard dislocation motion. Accordingly, a minimum cyclic stress intensity factor range ΔK_I is needed to induce dislocation motion during unloading and reloading. Thus, in this regime, ΔK_I below a critical *fatigue threshold* value ΔK^*_{th} precludes crack growth.

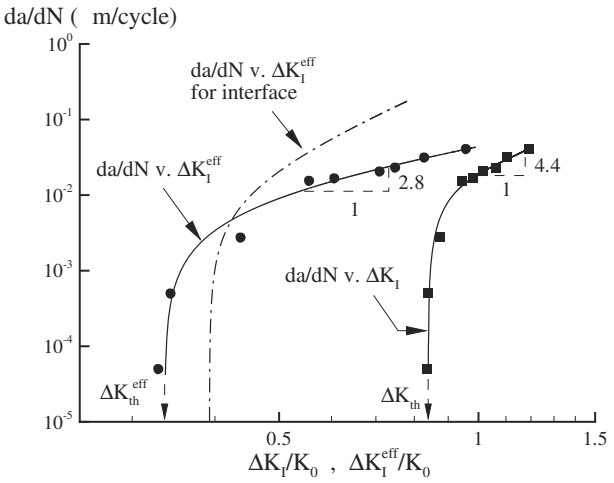


Figure 13. The cyclic crack growth rate da/dN versus $\Delta K_I/K_0$ and $\Delta K_I^{\text{eff}}/K_0$ for the mode I cyclic loading of a single crystal ($R = K_{\min}/K_{\max} = 0.3$). The slopes of the curves marked correspond to the Paris law exponents for the curves fitted through the numerical results. From [10].

2. The curves of crack advance per cycle, da/dN, as a function of ΔK_I, Fig. 13, show two distinct regimes of behavior: a steeply rising $\log(da/dN)$ versus $\log(\Delta K_I/K_0)$ curve in the threshold regime followed by a more gradual slope in the so-called *Paris regime*. The exponent m in the Paris relation

$$\frac{da}{dN} \propto (\Delta K_I)^m$$

for this case is ≈ 4.4. When the same data is expressed in terms of the effective stress intensity range

$$\Delta K^{\text{eff}} = \begin{cases} K_{\max} - K_{\text{op}} & \text{for } K_{\min} < K_{\text{op}} \\ \Delta K_I & \text{for } K_{\min} \geq K_{\text{op}}, \end{cases}$$

where K_{op} is the stress intensity at which the crack faces first separate [9], the Paris exponent is only around 2.8.

A fit da/dN versus $\Delta K_{\text{I}}^{\text{eff}}/K_0$ curve for an interface crack is also plotted in Fig. 13. The effect of the mode mixity at the interface is to increase the fatigue threshold of the interface crack but to reduce its resistance to cyclic crack growth at higher values of applied ΔK_I. This behavior is expected to be dependent on

the degree of mode mixity and hence affected by the cohesive properties and the applied loading.

Based on the experimental work of Laird and Smith [20] and Neumann [21], fatigue crack growth in ductile metals is often presumed to occur by an alternating slip mechanism which is a deformation-controlled phenomenon that does not require high stresses. On the basis of this, Pippan and co-workers [23, 24] and Wilkinson *et al*. [25] developed discrete dislocation models to represent this deformation-controlled fatigue crack growth mechanism. These models incorporate the crack growth mechanism as an ingredient of the model rather than have it emerge as a prediction of the analysis. By contrast, in the framework presented here [9, 10] fracture is both a deformation and stress-governed phenomenon and takes place by a mechanism that is possible under both monotonic and cyclic loading conditions. Striations are also predicted by the model, as shown in Fig. 14.

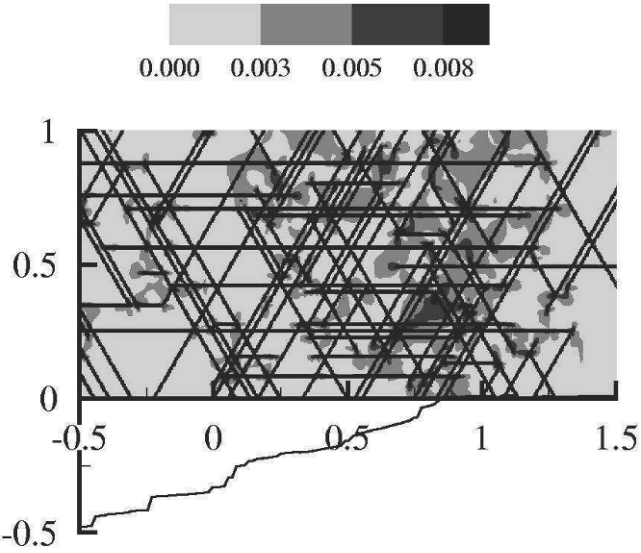


Figure 14. Contours of total slip showing the localized deformation pattern in the crack tip vicinity. All distances are in μm. The crack opening profile (displacements magnified by a factor of 20) is plotted below the x_1-axis. From [10].

8 Cracks in Polycrystals

Finally, we look at the crack tips fields in polycrystals, cf. Fig. 1c. For this purpose, the problem in Fig. 15b is analyzed, where the process window now contains a number of square grains. A mixture of FCC and BCC grains are considered, arranged in a checker-board pattern, Fig. 16; results will be presented here mainly for the case where the crack tip is in an FCC grain, Fig. 16a. All grains have the same density of sources, $\rho_{\text{nuc}} = 20\mu\text{m}^{-2}$ with average strength $\bar{\tau}_{\text{nuc}} = 46\,\text{MPa}$ and standard deviation 9.2 MPa; $\rho_{\text{obs}} = 40\mu\text{m}^{-2}$ in all grains. The grain size d is taken to vary between 0.2μm and $d = 5.0\mu$m.

In a continuum view, and assuming overall isotropic behaviour, the HRR solution for the opening stress is of the type

$$\sigma_{22} \propto \left(\frac{1}{r}\right)^{\frac{1}{s+1}} \tag{20}$$

where s is the strain hardening exponent in a shear stress τ versus shear strain γ relation of the form $\gamma \propto \tau^s$. The overall stress–strain behaviour of the polycrystals is extracted from pure shear computations on the grains inside the process window, as illustrated in Fig. 15a. The results in Fig. 17 show that the response is strongly dependent on the grain size. The shear stress between $\gamma = 0.15\%$ and 0.25%, $\bar{\tau}$, as a measure of yield stress, can be fit to the Hall-Petch type relation

$$\bar{\tau} - \tau_0 = \beta \left(\frac{d}{d_0}\right)^{-q}, \tag{21}$$

with an exponent $q = 0.415$ when using $\tau_0 = 20\,\text{MPa}$, the flow strength of a single crystal, and taking the reference grain size $d_0 = 1\mu$m. The Hall–Petch effect in these computations arises from the fact that grain boundaries act to stop the motion of dislocations.

The differences in deformation fields in polycrystalline aggregates with different grain size are represented in Fig. 18 in terms of the total slip

$$\Gamma = \sum_{\alpha=1}^{3} |\gamma^{(\alpha)}|, \quad \gamma^{(\alpha)} = s_i^{(\alpha)} \varepsilon_{ij} m_j^{(\alpha)}.$$

While the total density of sources and obstacles is independent of grain size and the same as for the single crystal, Fig. 18a, the slip distribution becomes smoother as the grain size decreases (note that the process window has the same dimensions). To show two extreme cases, Fig. 19 gives the stress distributions in a single crystal and in a polycrystal with small grains, $d = 0.6\mu$m. In contrast to the single crystal response (as dicussed previously), the stress distribution inside the polycrystal is

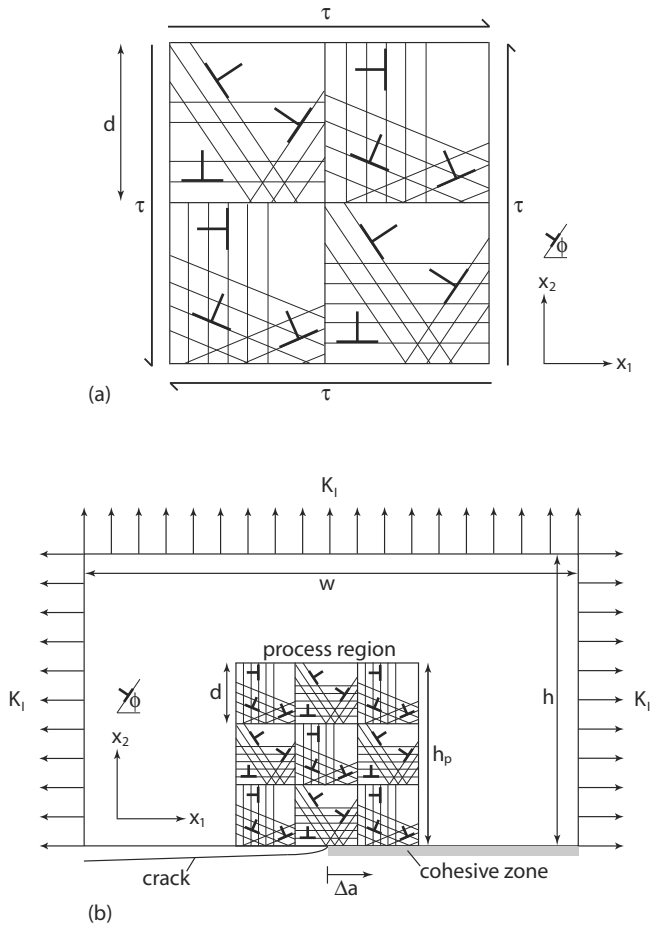


Figure 15. Sketch of (a) the pure shear problem with doubly-periodic boundary conditions and (b) the small scale yielding crack problem with imposed mode I loading. From [3].

seen to more closely resemble the isotropic HRR predictions. Thus, both the stress and strain fields around the crack tips in our polycrystal analyses indicate that the grain boundaries successfully block the formation of slip bands and tend to diffuse plastic deformation which results in more isotropic distributions of stress and strain.

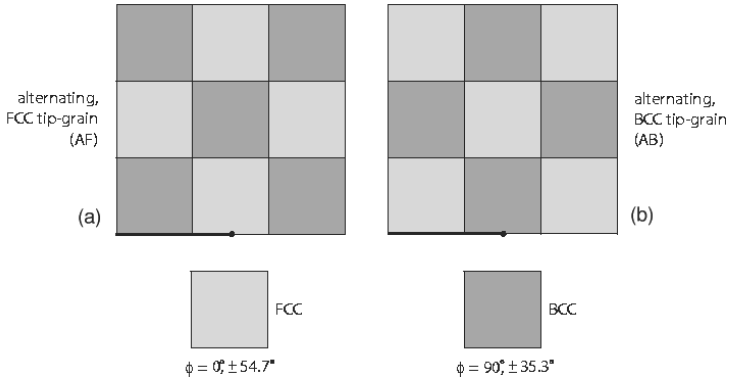


Figure 16. Two polycrystalline grain arrangements used in the mode I crack simulations. The arrangements are illustrated for a 3×3 array of grains with the crack plane indicated by the dark line. From [3].

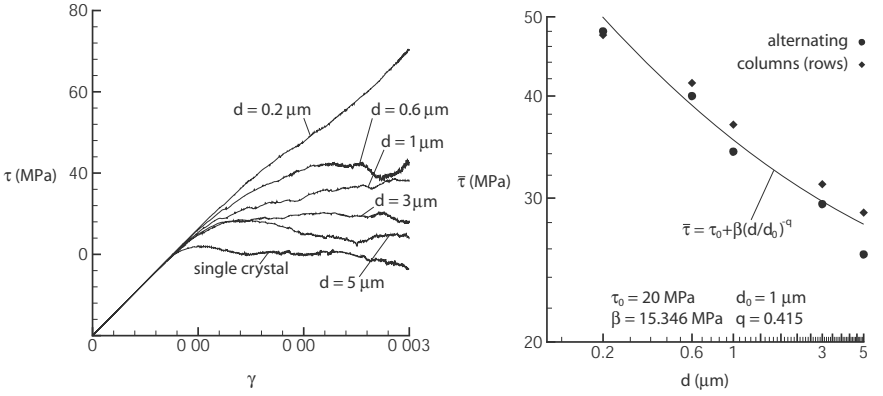


Figure 17. (a) Applied shear stress τ versus shear strain γ response of the polycrystalline and single crystal materials analyzed in this study. (b) Average shear stress $\bar{\tau}$ between $\gamma = 0.15\%$ and 0.25% for both grain arrangements as a function of grain size d. The Hall–Petch type relation eq. (21) is fit to the data. From [3].

The distribution of the opening stress ahead of the crack tip, Fig. 20, reveals three distinct regimes. Sufficiently far from the tip, $\sigma_{22} \propto 1/\sqrt{r}$ according to the

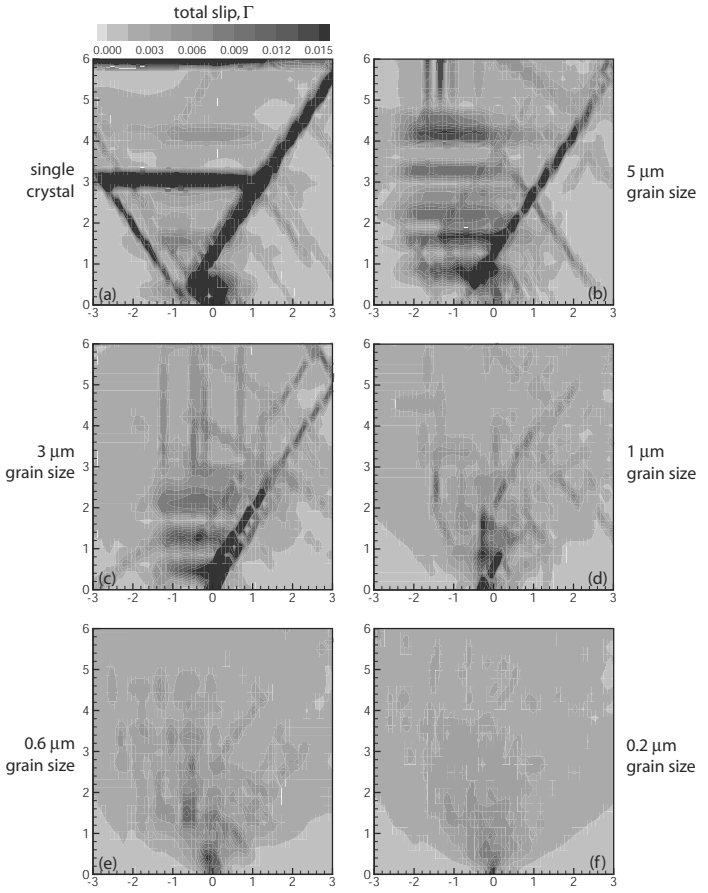


Figure 18. Distributions of total slip Γ around the stationary crack tip of polycrystals at an applied $K_I/K_0 = 1.75$: (a) pertains to the FCC single crystal and (b) through (f) are for the $d = 5\mu$m to $d = 0.2\mu$m polycrystals, in descending order. (All distances are in μm.) From [3].

elastic K-field, while very close to the tip the discreteness of the dislocations governs the field. In between there is a regime which has been fitted to a power-law consistent with the HRR field (20).

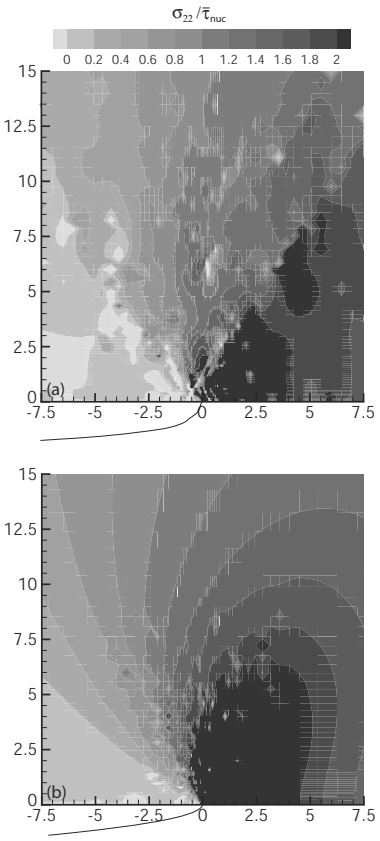


Figure 19. Distributions of normalized stress, $\sigma_{22}/\bar{\tau}_{\rm nuc}$, around the stationary crack tip at $K_I/K_0 = 1.75$: (a) the FCC single crystal (b) the polycrystal with $d = 0.6\,\mu$m. The crack-tip profile with displacement magnified by a factor of 50 is included in both cases. (All distances are in μm.) From [3].

Bibliography

[1] R.J. Asaro, *Adv. Appl. Mech.* **23**, 1–115 (1983).

[2] M.F. Ashby, *Phil. Mag.*, **21**, 399 (1970).

[3] D.S. Balint, V.S. Deshpande, A. Needleman and E. Van der Giessen, *Philos. Mag.* **85** (2005).

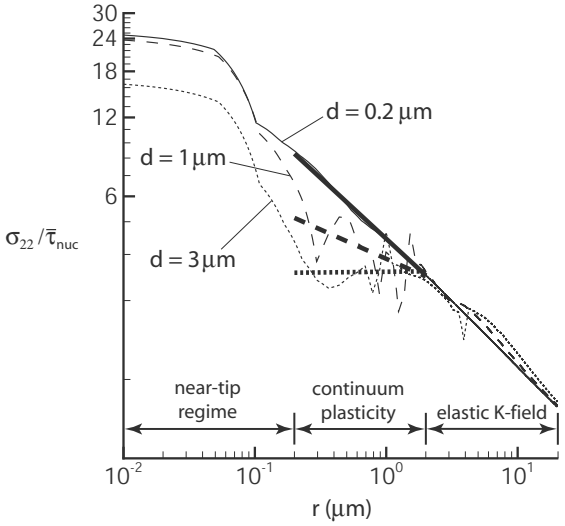


Figure 20. Normalized crack opening stress, $\sigma_{22}/\bar{\tau}_{nuc}$, on the crack plane ahead of the stationary crack tip at an applied $K_I/K_0 = 1.75$ for the polycrystal. Thick lines are fits to an HRR field for the continuum plasticity regime. From [3].

[4] W.D. Callister, *Materials Science and Engineering: An Introduction*. Fifth Edition. John Wiley & Sons, New York, 2000.
[5] E. Van der Giessen, *Mechanics of Microstructured Materials*, ed. by H.J. Bhm, CISM Courses and Lectures No. 464, 259–282 (2004).
[6] H.H.M. Cleveringa, E. Van der Giessen and A. Needleman, *Acta Mat.* **45**, 3163–3179 (1997).
[7] H.H.M. Cleveringa, E. Van der Giessen and A. Needleman, *Int. J. Plast.* **15**, 837–868 (1999).
[8] H.H.M. Cleveringa, E. Van der Giessen and A. Needleman, *J. Mech. Phys. Solids* **48**, 1133–1157 (2000).
[9] V.S. Deshpande, A. Needleman and E. Van der Giessen, *Acta Mat.* **49**, 3189–3203 (2001).
[10] V.S. Deshpande, A. Needleman and E. Van der Giessen, *Acta Mat.* **50**, 831–846 (2002).
[11] V.S. Deshpande, A. Needleman and E. Van der Giessen, *Acta Mat.* **51**, 1–15 (2003).
[12] P. Gumbsch, J. Riedle, A. Hartmaier, H.F. Fischmeister, *Science* **282**, 1293–1295 (1998).

[13] P.B. Hirsch, S.G. Roberts, *Scr. Metall.* **23**, 925–930 (1989).
[14] V.R. Nitzsche, K.J. Hsia, *Mat. Sci. Engng.* **A176**, 155–164 (1994).
[15] N. Zacharopoulos, D.J. Srolovitz, R. LeSar, *Acta Mat.* **45**, 3745–3763 (1997).
[16] J.P. Hirth and J. Lothe, *Theory of Dislocations*. McGraw–Hill, USA, 1968.
[17] D. Hull and D.J. Bacon, *Introduction to Dislocations*. Fourth Edition, Butterworh–Heinemann, UK, 2001.
[18] J.W. Hutchinson, *J. Mech. Phys. Solids* **16**, 13–31 (1968).
[19] L.P. Kubin, G. Canova, M. Condat, B. Devincre, V. Pontikis, Y. Bréchet, *Solid State Phenomena* **23 & 24**, 455-472 (1992).
[20] C. Laird and G.C. Smith, *Phil. Mag.* **7** 847–857 (1962).
[21] P. Neumann, *Acta Metall.* **17**, 1219–1225 (1969).
[22] J.F. Nye, *Acta. Metall.* **1**, 153 (1953).
[23] R. Pippan, *Acta Metall. Mat.* **39**, 255–262 (1991).
[24] F.O. Riemelmoser, P. Gumbsch, R. Pippan, *Mat. Trans. JIM* **42**, 2–13 (2001).
[25] A.J. Wilkinson, S.G. Roberts, P.B. Hirsch, *Acta Mat.* **46**, 379–390 (1998).
[26] J.R. Rice and G. Rosengren, *J. Mech. Phys. Solids* **16**, 1–12 (1968).
[27] J.R. Rice, *Mech. Mater.* **6**, 317–335 (1987).
[28] J.R. Rice and J.-S. Wang, *Mat. Sci. Engin.* **A107**, 23–40 (1989).
[29] M. Saeedvafa, J.R. Rice, *J. Mech. Phys. Solids* **37**, 673–691 (1989).
[30] W. Drugan, *J. Mech. Phys. Solids* **49**, 2155–2176 (2001).
[31] J.H. Rose, J. Ferrante, J.R. Smith, *Phys. Rev. Lett.* **47**, 675-678 (1981).
[32] J.Y. Shu, N.A. Fleck, E. Van der Giessen and A. Needleman, *J. Mech. Phys. Solids* **49**, 1361–1395 (2001).
[33] E. Van der Giessen and A. Needleman, *Model. Simul. Mater. Sci. Eng.* **3**, 689–735 (1995).
[34] E. Van der Giessen, V.S. Deshpande, H.H.M. Cleveringa and A. Needleman, *J. Mech. Phys. Solids* **49**, 2133-2153 (2001).
[35] E. Van der Giessen and A. Needleman, *Scripta Mat.* **48**, 127–132 (2003).

Statistical physical approach to describe the collective properties of dislocations

István Groma

Department of General Physics, Eötvös University Budapest, 1117 Budapest
Pázmány P. sétány 1/A

1 Introduction

The concept of dislocation was introduced by Polanyi, Orovan and Taylor in 1934 to explain the almost three orders of magnitude difference between the measured and theoretically estimated flow stress of crystalline materials. During the next 20-30 years, thanks to the contribution of a vast number of scientists, the theory of dislocations was successfully applied to explain several properties of the plastic deformation observed experimentally. Among other things the basic phenomena leading to work and precipitation hardening were understood [1, 2, 3].

On the other hand TEM investigations revealed that dislocations formed during plastic deformation tend to form different dislocation patterns, like the cell structure (see Figure 1.) developing at unidirectional load [4], or the so called ladder structure (see Figure 2.) developing under cyclic loading [5].

In spite of several attempts (Kuhlmann-Wilsdorf et.al. [6, 7, 8], Holt [9], Walgraef and Aifantis[10, 11, 12], Kratochvil et.al. [13, 14, 15]) proposed to model the pattern formation, we are far from the complete understanding of this typically self organizing phenomena. One of the most striking features of the dislocation patterning, which is a great challenge to model, is the large variety of the patterns observed. The variety manifests itself not just in the "geometry" of the dense dislocation regions but also in several statistical properties of the different dislocation ensembles. It is known for example that cyclic loading can lead to periodic structures with well defined self selected length scale [5], while unidirectional loading often results in fractal like structures which do not have any length scales (Hähner et.al. [16, 17]). Another interesting feature of the patterning process observed recently by X-ray diffraction (Székely et.al.[18, 19]) is that in case of unidirectional

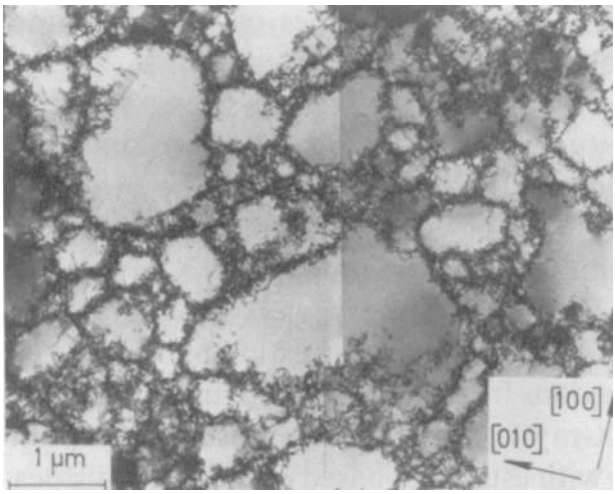


Figure 1. Dislocation cell structure obtained on Cu single crystal oriented for multiple slip [4].

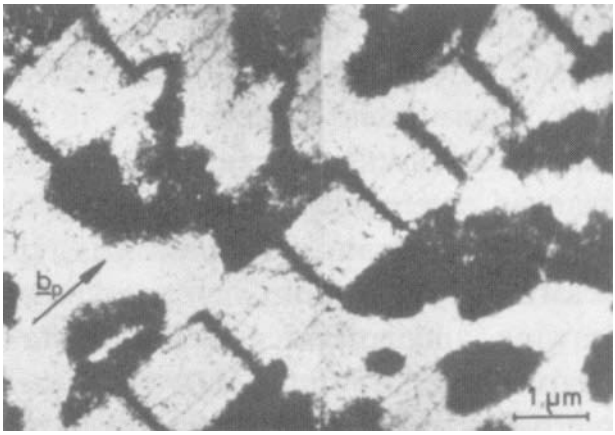


Figure 2. Ladder structure obtained on Cu single crystal deformed cyclically [5].

loading the relative dislocation density fluctuation σ^2 defined as

$$\sigma^2 = V \frac{\int \rho^2(\vec{r}) d^3r}{\left[\int \rho(\vec{r}) d^3r\right]^2}, \tag{1}$$

where $\rho(\vec{r})$ is the dislocation density and V is the crystal volumes, undergoes a sharp maximum as deformation proceeds (see Figure 3). The result

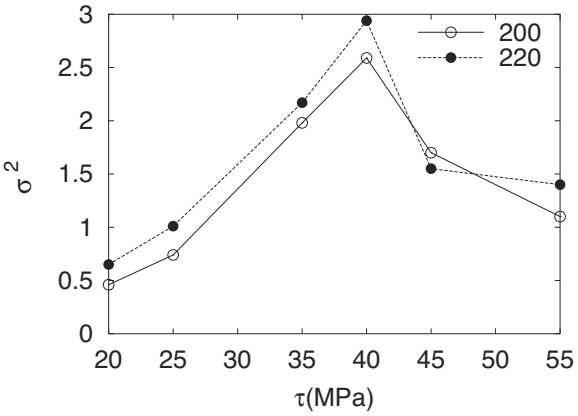


Figure 3. Relative dislocation density fluctuation versus applied stress, obtained on Cu single crystal oriented for multiple slip. The two curves correspond to relative dislocation density fluctuations determined from the broadening of (200) and (220) Bragg peaks [18, 19].

indicates that during plastic deformation the dislocation system tends to become more and more inhomogeneous, but after a certain deformation level (depending on the crystal orientation and temperature) this separation process cannot continue any longer and the system becomes more and more homogeneous.

Another challenging problem intensively studied nowadays (Flack et.al.[20], McElhaney et.al. [21]) is that resent experimental investigations revealed (see Figures 4 and 5), if the characteristic size of a specimen is less than about 10μm the plastic response of crystalline materials depend on the size. The phenomenon is commonly called "size effect". One can easily explain this size dependence by assuming that the crystalline materials have an internal degree of freedom which "feel" the sample surface. This immediately indicates that a theory able to account for the size effects has to be non-local, since the sample surface is "seen" from the bulk. The simplest

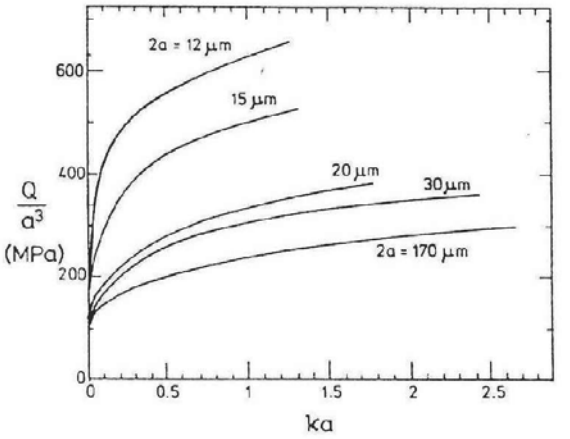


Figure 4. Normalised torque versus shear deformation obtained on torsionally deformed wires with different diameters a. The curves indicates, if a is smaller than 50μm, the hardening of the wires increase with decreasing diameter [20].

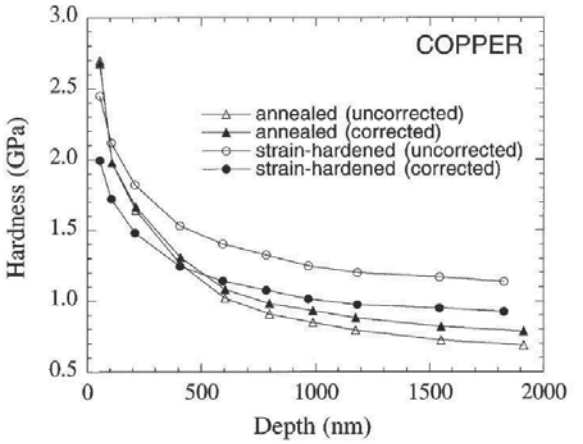


Figure 5. Microhardness versus indentation depth obtained on cold rolled Cu. It can be seen, if the indentation depth is less than 1μm the microhardness increases with decreasing indentation depth [21].

possible way to introduce nonlocality is to add gradient terms to the "local" ones. There are several different phenomenological ways to do this. As a possibility one may introduce an effective shear stress τ_{eff} as

$$\tau_{eff} = \tau(\gamma, \dot{\gamma}, ...) + \mu l^2 \Delta\gamma, \tag{2}$$

where $\tau(\gamma, \dot{\gamma}, ...)$ is the "classical" local stress-strain relation, γ is the shear deformation, μ is the shear modulus and l is a parameter with length dimension. It is important to stress that l has to be introduced for the appropriate dimension of the second term. Since size effects appears at micron scale, the value of l needs to be in the order of μm.

During the past 10 years several non-local plasticity theories based on similar arguments explained above have been proposed (Aifantis [22, 23, 24], Flack and Hutchinson [25], Gurtin [26], Svendsen [27]) and successfully applied to explain experimental results. However, these phenomenological non-local theories suffer from the common drawback, namely, the physical origin of the gradient terms are not clear. This is especially pronounced if one tries to explain the origin of a material parameter (l) with a value around μm.

On the other hand, we know that in crystalline materials plastic deformation is carried out by dislocations. They definitely act as internal degrees of freedom. Due to the long range stress field of dislocations they "feel" the sample surface. This indicates that size effect is the result of the collective motion of dislocations. The results of discrete dislocation dynamics simulations clearly demonstrate this (see Nicola et.al. [28]).

The aim of the present paper is to explain how the collective behaviour of dislocations can be treated within a statistical physics framework. Since, however, the dislocations form a strongly dissipative system, the classical theory of statistical physics developed for Hamiltonian systems cannot be directly applied. In the first part of the paper the field theory of dislocations developed by Kröner and Kosevich is summarised [29, 30, 31, 32]. Although, the Kröner and Kosevich theory is not able to capture all the statistical features of dislocation systems, it plays an important role in our further considerations.

In the second part of the paper it is shown that for a simple 2D dislocation configuration a continuum theory of dislocations can be rigorously derived. According to numerical investigations it is able to account for certain type of size effects. At the end of this chapter, attempts to generalise the results for 3D are shortly explained.

Finally, in the last part the properties of the internal stress distribution generated by dislocations are discussed.

2 Kröner-Kosevich field theory of dislocations

2.1 Nye's dislocation density tensor

Shortly after the concept of dislocation was introduced, the elastic field generated by dislocation loops was determined (see for example [1]). In the early 1950s Kröner and Kosevich reconsidered the problem and proposed an extremely elegant formalism to describe the elastic properties of dislocated crystals. In this section the dislocation field theory they have developed is shortly summarised. For more details the reader is refereed to [29, 30, 31, 32].

Let us consider a body subject to shape change. Its deformation is uniquely described by the displacement field $\vec{u}(\vec{r})$. In the small deformation limit the change of atomic distances is measured by the total deformation tensor

$$\epsilon^t_{ij} = \frac{1}{2}\left[\frac{\partial u_i}{\partial r_j} + \frac{\partial u_j}{\partial r_i}\right] \qquad \left(\hat{\epsilon}^t = \left[\frac{d\vec{u}}{d\vec{r}}\right]_{sym}\right). \tag{3}$$

(For better understanding expressions are given both with component and tensor notations. Tensors are denoted by hats.)

In the common theory of elasticity it is assumed that the stress tensor $\hat{\sigma}$ is a unique function of the deformation tensor $\hat{\epsilon}^t$. In linear elasticity considered in this paper

$$\sigma_{ij} = L_{ijkl}\epsilon^t_{lk} \qquad \left(\hat{\sigma} = \hat{L} : \hat{\epsilon}^t\right) \tag{4}$$

where $\hat{L}$ is the tensor of elastic moduli.

Deformation of a body, however, does not necessarily leads to the development of internal stress. If we cut a body into two pieces along a plane, slide the two parts with respect to each other and glue them together, the deformation is obviously nonzero along the plane, but the final stage is stress free. This means, the deformation can have a part which generates stress and another one which does not. According to this, the starting point of the plasticity theory explained below is that the gradient of the displacement field commonly referred as total distortion is the sum of two terms, the so called plastic ($\hat{\beta}^p$) and elastic ($\hat{\beta}$) distortion i.e.

$$\frac{\partial u_j}{\partial r_i} = \beta_{ij} + \beta^p_{ij} \qquad \left(\frac{d\vec{u}}{d\vec{r}} = \hat{\beta} + \hat{\beta}^p\right). \tag{5}$$

It needs to be mentioned, that the above expression is valid only in small deformation limit. For large deformations, as a constitutive rule, it is commonly assumed that the final deformation stage of the body is reached by

two successive steps, a plastic and an elastic deformation. For small deformations this is equivalent with Eq. (5). For the sake of simplicity we consider only this case.

Splitting the total distortion into two terms is rather formal so far. The plastic and elastic distortions need to be defined. It has to be mentioned at this point that, as it is explained in details below, the separation is not unique.

As it is indicated by its name the elastic distortion is that part of the total one, which is related to the stress developed in the body due to the deformation. In linear elasticity

$$\sigma_{ij} = L_{ijkl}\beta_{lk} \qquad \left(\hat{\sigma} = \hat{L} : \hat{\beta}\right). \tag{6}$$

However, due to the symmetric properties of $\hat{L}$, $\hat{\sigma}$ determines only the symmetric part of $\hat{\beta}$. With other words, the stress determines only the elastic deformation $\hat{\epsilon}$ defined as

$$\epsilon_{ij} = \frac{1}{2}(\beta_{ij} + \beta_{ji}) \qquad \left(\hat{\epsilon} = \left[\hat{\beta}\right]_{sym}\right). \tag{7}$$

From Eqs. (6,7) the elastic deformation reads as

$$\epsilon_{ij} = L^{-1}_{ijkl}\sigma_{lk} \qquad \left(\hat{\epsilon} = \hat{L}^{-1} : \hat{\sigma}\right), . \tag{8}$$

where $\hat{L}^{-1}$ denotes the inverse of the tensor of elastic moduli.

The next step is to define the plastic distortion. Since, by definition, the total distortion is the gradient of the displacement field, it is curl free, i.e.

$$e_{ikl}\frac{\partial}{\partial r_i}\frac{\partial u_j}{\partial r_i} \equiv 0 \qquad \left(\nabla \times \frac{d\vec{u}}{d\vec{r}} \equiv 0\right), \tag{9}$$

where e_{ikl} is the antisymmetric tensor. The plastic distortion, however, is not necessarily curl free. After Nye [33] the curl of the plastic distortion is called dislocation density tensor $\hat{\alpha}$:

$$\alpha_{ij} = e_{ikl}\frac{\partial}{\partial r_k}\beta^p_{lj}. \qquad \left(\hat{\alpha} = \nabla \times \hat{\beta}^p\right) \tag{10}$$

(In the literature $\hat{\alpha}$ is often refereed to Nye's dislocation density tensor.) From Eqs. (5,9,10) one easily obtains that

$$\alpha_{ij} = -e_{ikl}\frac{\partial}{\partial r_k}\beta_{lj} \qquad \left(\hat{\alpha} = -\nabla \times \hat{\beta}\right). \tag{11}$$

In order to see the physical meaning of $\hat{\alpha}$ introduced above, let us consider its integral for an arbitrary surface A

$$b_j = \int_A \alpha_{ij} dA_i. \tag{12}$$

From the (11) expression of $\hat{\alpha}$, with the help of the Stockes' theorem, one can find that

$$b_j = -\int_A e_{ikl} \frac{\partial}{\partial r_k} \beta_{lj} dA_i = -\oint \beta_{ij} ds_i = -\oint du_i^e. \tag{13}$$

Equation

$$b_j = -\oint du_i^e \tag{14}$$

obtained is the one Burgers originally used [34] to define a dislocation as a singularity of the elastic displacement field $\vec{u}^e$. Therefore, $\hat{\alpha}$ defined by Eq. (10) is the net Burgers vector of line singularities crossing a unit area. From Eq. (12) one can easily find that for a single straight dislocation

$$\alpha_{ij} = l_i b_j \delta^{(2)}(\xi) \qquad \left(\hat{\alpha} = \vec{l} \circ \vec{b}\ \delta^{(2)}(\xi)\right), \tag{15}$$

where $\vec{l}$ is the tangential vector of the dislocation line and ξ is the distance from the line.

Before we proceed further, it should be mentioned that the dislocation density tensor does not define uniquely the plastic distortion. Knowing $\hat{\alpha}$ leaves $\hat{\beta}^p$ uncertain up to a gradient of an arbitrary vector field. However, as it is explained below, this does not affect the stress field created by the dislocation system.

2.2 Internal stress generated by the dislocation system

In order to derive the field equations, let us consider the symmetric part of the total distortion. According to Eq. (5)

$$\frac{1}{2}\left(\frac{\partial u_j}{\partial r_i} + \frac{\partial u_i}{\partial r_j}\right) = \epsilon_{ij} + \epsilon_{ij}^p \qquad \left(\left[\frac{d\vec{u}}{d\vec{r}}\right]_{sym} = \hat{\epsilon} + \hat{\epsilon}^p\right), \tag{16}$$

where $\hat{\epsilon}$ and $\hat{\epsilon}^p$ are the symmetric parts of the elastic and plastic distortions, respectively. Using the curl grad $\equiv 0$ identity one can find that

$$-e_{ikm} e_{jln} \frac{\partial}{\partial r_k} \frac{\partial}{\partial r_l} \left(\frac{\partial u_j}{\partial r_i} + \frac{\partial u_i}{\partial r_j}\right) \equiv 0 \qquad \left(\nabla \times \left[\frac{d\vec{u}}{d\vec{r}}\right]_{sym} \times \nabla \equiv 0\right). \tag{17}$$

It is useful to introduce the tensor operator inc ("incompatibility") defined as

$$\text{inc} = -e_{ikm}e_{jln}\frac{\partial}{\partial r_k}\frac{\partial}{\partial r_l}. \tag{18}$$

With this definition Eq. (17) reads as

$$\text{inc}\left[\frac{d\vec{u}}{d\vec{r}}\right]_{sym} \equiv 0. \tag{19}$$

For further considerations it is useful to introduce the symmetric tensor field

$$\eta_{ij} = -e_{ikm}e_{jln}\frac{\partial}{\partial r_k}\frac{\partial}{\partial r_l}\epsilon_{mn} \qquad (\text{ inc } \hat{\epsilon} = \hat{\eta})\,. \tag{20}$$

After a long but straightforward calculation one can obtain from Eqs. (10, 16, 19, 20) that

$$\eta_{ij} = \frac{1}{2}\left(e_{jln}\frac{\partial}{\partial r_l}\alpha_{im} + e_{iln}\frac{\partial}{\partial r_l}\alpha_{jm}\right) \qquad \left(\hat{\eta} = [\hat{\alpha}\times\nabla]_{sym}\right). \tag{21}$$

By substituting Eq. (8) into (20) we arrive at the equation that the stress created by the dislocation has to fulfill:

$$\eta_{ij} = -e_{ikm}e_{jln}\frac{\partial}{\partial r_k}\frac{\partial}{\partial r_l}L^{-1}_{mnop}\sigma_{op} \qquad \left(\text{ inc}\left(\hat{L}^{-1}\hat{\sigma}\right) = \hat{\eta}\right). \tag{22}$$

Since, for an arbitrary vector field $\vec{f}$

$$\text{inc}\left[\frac{d\vec{f}}{d\vec{r}}\right]_{sym} \equiv 0. \tag{23}$$

Eq. (22) itself does not determine $\hat{\sigma}$ completely. However, supplementing it with the equilibrium condition

$$\frac{\partial}{\partial r_i}\sigma_{ij} = 0 \qquad (\text{ div } \hat{\sigma} = 0) \tag{24}$$

we already have enough equations to determine the stress field generated by the dislocations.

2.3 Second order stress function tensor

Like in electrodynamics it is useful to reformulate Eqs. (22,24) into a potential theory. Let us introduce a second order stress function tensor $\hat{\chi}$ defined with the relation

$$\sigma_{ij} = -e_{ikm}e_{jln}\frac{\partial}{\partial r_k}\frac{\partial}{\partial r_l}\chi_{mn} \qquad (\hat{\sigma} = \text{ inc } \hat{\chi})\,. \tag{25}$$

Due to the identity

$$\text{div } \text{ inc} \equiv 0 \tag{26}$$

the (22) form of $\hat{\sigma}$ guarantees that the equilibrium condition (24) is fulfilled. With the stress function tensor introduced above Eq. (21) reads as

$$\begin{aligned} \eta_{ij} &= e_{ikm}e_{jln}e_{oqv}e_{puw}L^{-1}_{mnop}\frac{\partial}{\partial r_k}\frac{\partial}{\partial r_l}\frac{\partial}{\partial r_q}\frac{\partial}{\partial r_u}\chi_{vw} \\ (\hat{\eta} &= \text{inc}\left(\hat{L}^{-1}\text{ inc }\hat{\chi}\right)\Big)\,. \end{aligned} \tag{27}$$

For an anisotropic medium the above equation is rather difficult to solve, but for isotropic materials, with shear modulus μ and Poisson's ratio ν, a general solution can be obtained. It is expedient to introduce another tensor potential $\hat{\chi}'$ defined as

$$\chi'_{ij} = \frac{1}{2\mu}\left(\chi_{ij} - \frac{\nu}{1+2\nu}\chi_{kk}\delta_{ij}\right) \tag{28}$$

$$\chi_{ij} = 2\mu\left(\chi'_{ij} + \frac{\nu}{1-\nu}\chi'_{kk}\delta_{ij}\right). \tag{29}$$

By inserting Eq. (29) into Eq. (27) one can find, if $\hat{\chi}'$ fulfills the gauge condition

$$\frac{\partial}{\partial r_i}\chi'_{ij} = 0 \qquad (\text{ div } \hat{\chi}' = 0) \tag{30}$$

Eq. (25) simplifies to the biharmonic equation

$$\nabla^4\chi'_{ij} = \eta_{ij} \qquad \left(\nabla^4\hat{\chi}' = \hat{\eta}\right). \tag{31}$$

A remarkable feature of this equation is that the different components of $\hat{\chi}'$ obey separate equations making the problem much easier to solve. For an infinite medium the general solution of Eq. (31) is

$$\chi'_{ij}(\vec{r}) = -\frac{1}{8\pi}\int\int\int |\vec{r}-\vec{r}'|\eta_{ij}(\vec{r}')dV' \tag{32}$$

2.4 2D problems

In the next section the statistical properties of an ensemble of parallel edge dislocations are discussed. In this case the stress and the strain do not vary along the dislocation line direction $\vec{l}$. Taking $\vec{l}$ parallel to the z axis (with $\vec{l} = (0, 0, -1)$ in the above expressions the derivatives with respect to z vanish ($\partial/\partial z \equiv 0$). One can find that Eq. (25) simplifies to [30]:

$$\sigma_{11} = -\frac{\partial^2 \chi}{\partial y^2}, \qquad \sigma_{22} = -\frac{\partial^2 \chi}{\partial x^2}, \qquad \sigma_{12} = \frac{\partial^2 \chi}{\partial x \partial y}, \qquad \chi \equiv \chi_{33} \tag{33}$$

$$\sigma_{23} = -\frac{\partial \phi}{\partial x}, \qquad \sigma_{13} = \frac{\partial \phi}{\partial y}, \qquad \phi = -\frac{\partial \chi_{23}}{\partial x} + \frac{\partial \chi_{31}}{\partial y}. \tag{34}$$

Furthermore, from Eqs. (21,27) one obtains that the two scalar fields χ and ϕ introduced above obey the equations

$$\nabla^4 \chi = \frac{2\mu}{1-\nu} \left(b_1 \frac{\partial}{\partial y} - b_2 \frac{\partial}{\partial x} \right) (\rho_{d+} - \rho_{d-}) \tag{35}$$

$$\nabla^2 \phi = \mu b_3 (\rho_{d+} - \rho_{d-}), \tag{36}$$

where b_1, b_2 and b_2 are the x, y, and z directional components of the Burgers vector, respectively. The notations ρ_{d+} and ρ_{d-} stand for the dislocation densities with positive and negative signs, respectively, They are the sum of $\delta(\vec{r} - \vec{r}_i)$ Dirac delta functions, were $\vec{r}_i$ denotes the position of a dislocation. For the sake of simplicity we assumed that all dislocations belong to the same slip system (single slip), but the expressions can be easily generalised for multiple slip.

For an infinite medium the solutions of Eqs. (35,36) read as

$$\chi(\vec{r}) = \frac{2\mu}{2\pi(1-\nu)} \int \left(b_1 \frac{\partial}{\partial y'} - b_2 \frac{\partial}{\partial x'} \right) [\rho_{d+}(\vec{r}') - \rho_{d-}(\vec{r}')] R^2 \ln R \; d^2 \vec{r}' \tag{37}$$

and

$$\phi(\vec{r}) = -\frac{\mu b_3}{2\pi} \int [\rho_{d+}(\vec{r}') - \rho_{d-}(\vec{r}')] \ln(R) \; d^2 \vec{r}' \tag{38}$$

where $R = |\vec{r} - \vec{r}'|$.

2.5 Time evolution of the dislocation density tensor

As it is explained above, if we know the dislocation density tensor (i.e. we know the dislocation line geometry) the internal stress field can be determined from Eq. (27). This is however, only a "static" description. In order to be able to describe the response of the dislocation system to external signals, the governing equations of the time evolution of the dislocation density tensor should be determined.

If we take the partial time derivative (denoted by " $\dot{}$ ") of Eq. (10), we find that

$$\dot{\alpha}_{ij} + e_{ikl}\frac{\partial}{\partial r_k}j_{lj} = 0 \qquad \left(\dot{\hat{\alpha}} + \nabla \times \hat{j} = 0\right) \tag{39}$$

where

$$\hat{j} = -\dot{\hat{\beta}}^p \tag{40}$$

is called dislocation current density [31]. The above equation is the "conservation law of the Burgers vector" in differential form. Indeed, if we integrate both sides of Eq. (39) for an arbitrary area contoured by the closed curve L, according to Eq. (12), we obtain that

$$\frac{db_j}{dt} = -\oint_L j_{ij} ds_i \tag{41}$$

It is obvious from this relation that $\hat{j}$ is the Burgers vector carried by the dislocations crossing a unit length part of the contour line L per unit time.

For an individual dislocation one can find that

$$j_{ik} = e_{ilm} l_l v_m b_k \delta^{(2)}(\xi) \qquad \left(\hat{j} = (\vec{l} \times \vec{v}) \circ \vec{b}\, \delta^{(2)}(\xi)\right), \tag{42}$$

where $\vec{v}$ is the velocity of the dislocation line at a given point. It is important to note that if we added the gradient of an arbitrary vector field to $\hat{j}$ given above, this would also satisfy the conservation law (39). The problem is obviously related to the non-uniqueness of the plastic distortion discussed earlier. However, expression (42) is the only one which is physically meaningful. One expects that there is no plastic current anywhere else but at the dislocation line. Nevertheless, strictly speaking we have to postulate this.

The above results clearly show that $\hat{j}$ has to be considered as an independent quantity. In order to be able to describe the time evolution of the dislocation system we have to set up a constitutive relation which gives how $\hat{j}$ depends on the dislocation density tensor and the external stress. Due to the long range nature of the dislocation-dislocation interaction, the constitutive relation is obviously non-local in $\hat{\alpha}$. Beside this, the constitutive

relation has to be able to account for several different "local" phenomena (self loop interaction, junction formation, annihilation etc.) making even more difficult to determine its form.

One possible approach to handle this problem is to set up the constitutive relation from phenomenological considerations. During the past years several phenomenological expressions were proposed and successfully applied for modelling certain phenomena [26] but the problem is far not completely solved.

Another widely used approach to study the time evolution of dislocation systems is discrete dislocation dynamics (DDD) simulation in which the dislocation loops are considered individually. After setting up velocity laws for the dislocation segments the dislocation loop geometry is updated numerically. Describing the actual numerical techniques used in DDD simulations is out of the scope of this paper. The reader can find the details e.g. in [35, 36, 37, 38, 39, 40, 42, 43, 44, 45]. Although DDD simulations are extremely important for the better understanding of the collective properties of dislocations, due to their high computational demand their applicability in engineering practice is limited.

2.6 Time evolution of the displacement field

In the previous part we have discussed how the stress field generated by the dislocations can be determined and what can be said in general about the time evolution of the dislocation density tensor. However, in many applications it is important to determine the displacement field $\vec{u}(\vec{r})$, too.

Let us go back to our starting equation (5), multiply it with the tensor of elastic moduli $\hat{L}$, and take the div of the equation. Using Eqs. (6,24) one obtains that

$$\frac{\partial}{\partial r_i} L_{ijkl} \frac{\partial u_k}{\partial r_l} = \frac{\partial}{\partial r_i} L_{ijkl} \beta^p_{kl} \qquad \left(\text{ div } \hat{L} \frac{d\vec{u}}{d\vec{r}} = \text{ div } \hat{L}\hat{\beta}^p \right). \tag{43}$$

This is formally equivalent with the common equilibrium equation of elasticity with body force density

$$f_j = -\frac{\partial}{\partial r_i} L_{ijkl} \beta^p_{kl} \qquad \left(\vec{f} = - \text{ div } \hat{L}\hat{\beta}^p \right) \tag{44}$$

Since, as it is explained earlier, the dislocation density tensor does not determine the plastic distortion uniquely, the above equation is not enough to determine the displacement field. Taking, however, the time derivative

of Eq. (43), with the (40) definition of $\hat{j}$, one finds that

$$\frac{\partial}{\partial r_i} L_{ijkl} \frac{\partial \dot{u}_k}{\partial r_l} = \frac{\partial}{\partial r_i} L_{ijkl} j_{kl} \qquad \left(\text{div } \hat{L} \frac{d\dot{\vec{u}}}{d\vec{r}} = \text{ div } \hat{L}\hat{j} \right). \tag{45}$$

As it is discussed above, based on physical arguments, $\hat{j}$ can be uniquely defined, so the deformation velocity field $\dot{\vec{u}}$ can already be determined if $\hat{j}$ is known. Integrating it with respect to time gives the change of the displacement field which is the quantity one can really measure.

2.7 Problems related to coarse graining

The dislocation density tensor introduced above is a highly singular quantity. It is infinite along the dislocation lines and vanishes elsewhere. More precisely, it is proportional to a delta function along the dislocation lines. The same holds for the dislocation current density. The conservation law (39) guarantees that during the evolution of the dislocation system this delta function does not "spread out", only the shape of the loops can change. This is certainly what we expect physically. This means, however, if we want to follow the evolution of the system we have to follow the track of each dislocation loop as it is done in DDD simulations.

We may hope, like for many other physical systems, to predict the macroscopic response of the dislocation system, we do not need this detailed knowledge of the evolution of the dislocation configuration. One should try to operate with locally averaged quantities. This means, the different quantities, like dislocation density tensor, stress, dislocation current density, etc., are convolved with a window function. In the physics literature the procedure is commonly called as "coarse graining", while in the engineering community the term "homogenisation" is more frequently used. One can immediately raise the question what is the appropriate function we should use for the shape of the window function, and what determines its half width. One cannot have a general recipe how to resolve these problems. Nevertheless, we can hope that within certain limits the result obtained by the coarse graining is not sensitive to the actual window function shape and its width. If this is not the case, this clearly indicates that all the microscopical details are important. So, the coarse graining procedure always requires extra care. Beside this, it is important to stress that, before the equations obtained by coarse graining are applied, for predicting the response of the material in a given situation, one always has to study the relevance of the homogenisation.

In order to indicate the difficulties, as a simple example, let us consider a set of parallel edge dislocations with $\pm\vec{b}$ Burgers vectors parallel to the x

axis. For this case Eq. (35) simplifies to

$$\nabla^4 \chi = \frac{2\mu}{1-\nu} \left(b \frac{\partial}{\partial y} \right) \kappa_d, \tag{46}$$

where $\kappa_d = \rho_{d+} - \rho_{d-}$ is the signed dislocation density that is a sum of delta functions. If we take the convolution of Eq. (46) with a window function $w(\vec{r})$ we obtain that

$$\int w(\vec{r} - \vec{r}') \nabla'^4 \chi(\vec{r}') d\vec{r}' = \frac{2\mu}{1-\nu} \int w(\vec{r} - \vec{r}') \left(b \frac{\partial}{\partial y'} \right) \kappa_d(\vec{r}') d\vec{r}'. \tag{47}$$

After partial integrations we get that

$$\nabla^4 \int w(\vec{r} - \vec{r}') \chi(\vec{r}') d\vec{r}' = \frac{2\mu}{1-\nu} \left(b \frac{\partial}{\partial y} \right) \int w(\vec{r} - \vec{r}') \kappa_d(\vec{r}') d\vec{r}'. \tag{48}$$

As it can be seen, the coarse grained fields denoted by

$$< \chi >= \int w(\vec{r} - \vec{r}') \chi(\vec{r}') d\vec{r}' \qquad < \kappa >= \int w(\vec{r} - \vec{r}') \kappa_d(\vec{r}') d\vec{r}' \tag{49}$$

are related to each other as

$$\nabla^4 < \chi >= \frac{\mu}{1-\nu} \left(b \frac{\partial}{\partial y} \right) < \kappa > \tag{50}$$

that is formally equivalent with Eq. (46). With a similar argument, from Eq. (33) one can find that

$$\begin{aligned} < \sigma >_{11} &= -\frac{\partial^2 < \chi >}{\partial y^2}, \\ < \sigma >_{22} &= -\frac{\partial^2 < \chi >}{\partial x^2}, \\ < \sigma >_{12} &= \frac{\partial^2 < \chi >}{\partial x \partial y}. \end{aligned} \tag{51}$$

This means, the coarse grained fields are related to each other as the "discrete" ones.

However, important information is lost during coarse graining. If we consider two dislocation configurations indicated in Figure 6 and coarse grain them for the same square areas indicated by the boxes, we get the same signed dislocation density value. On the other hand, it is obvious that the response of the two configurations are strongly different, if one applies

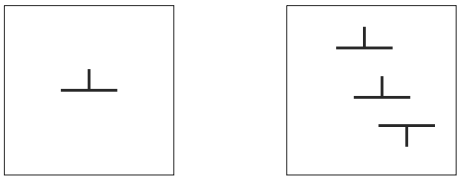


Figure 6. Two strongly different dislocation configurations giving the same κ if they are coarse grained for the areas indicated by the boxes.

an external shear. So, in a continuum theory of dislocations, in which we operate with smooth fields, the coarse grained dislocation density tensor is not enough to characterise the state of the system. In the next section we discuss how a continuum theory can be derived from the equation of motion of straight parallel dislocations and what relevant quantities are needed to have an appropriate description of this simple dislocation system on the mesoscopic scale.

3 Linking micro- to mesoscale for a 2D dislocation system

3.1 Discrete dislocation dynamics simulations in 2D

In the early 90s due to the fast increase of the available computer power it became possible to investigate the collective properties of dislocations by computer simulations. In DDD simulations the equations of motion of the dislocations (dislocation segments) are integrated numerically. During the past 15 years a vast amount of DDD simulations were performed both in 2D [46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56] and 3D [35, 36, 37, 38, 39, 40, 42, 43, 44, 45]. The detailed descriptions of the numerical methods applied and the results obtained are out of the scope of the paper, but to demonstrate the potentials and the limitations of the numerical simulations we will shortly review simulation results obtained by the author [50, 51] on the same system that is studied below in the statistical physics considerations.

In the simulations, the evolution of a system of parallel edge dislocations was studied in a square area with periodic boundary conditions. (This means, the originally 3D problem was simplified to 2D.) The system was subject of unidirectional deformation with a constant external deformation rate $\dot{\epsilon}$. The stress required to keep $\dot{\epsilon}$ constant was calculated. Dislocation

motion was allowed in two slip systems (denoted by dotted lines in Figures 7) with 60^0 between the two slip directions. The system was oriented for single slip, i.e. in one of the two slip systems the shear stress generated by the external load was much larger than in the other one.

In the simulations reported here only dislocation glide was allowed (i.e. the velocity of the dislocations were parallel to their Burgers vectors). In the equation of motion of the dislocations, the inertia term (ma) was neglected beside the friction force $\vec{F}_f$ accounting for the energy dissipation during dislocation motion. This is called overdamped dynamics. For simplicity, we assumed that $\vec{F}_f$ was proportional to the dislocation velocity, $\vec{F}_f = -B\vec{v}$, where the coefficient B is called as dislocation mobility.

Beside the friction force, if the stress is nonzero at the dislocation, the Peach-Koehler force [1]

$$\vec{F}_{PK} = \vec{l} \times (\hat{\sigma}\vec{b}) \tag{52}$$

also acts on a dislocation, where $\hat{\sigma}$ is the sum of the stress created by the dislocations and the external stress. With these, the equation of motion of the ith dislocation is

$$\vec{v}_i = B^{-1}\vec{s}_i \left[\vec{m}_i \left(\sum_{j=1}^{N} \hat{\sigma}^{ind}(\vec{r}_j - \vec{r}_i, \vec{b}_j) + \hat{\sigma}^{ext} \right) \vec{b}_i \right] \tag{53}$$

where $\vec{s}_i$ and $\vec{m}_i$ are unite vectors parallel and perpendicular to the slip direction of the ith dislocation, respectively, $\hat{\sigma}_{ext}$ is the external stress, and $\hat{\sigma}^{ind}(\vec{r}-\vec{r}_i, \vec{b})$ is the stress field created by the ith dislocation (with Burgers vector $\vec{b}$). According to Eqs. (33,35), $\hat{\sigma}^{ind}$ needs to be determined from the equations

$$\nabla^4 \chi^{ind} = \frac{2\mu}{1-\nu} \left(b_1 \frac{\partial}{\partial y} - b_2 \frac{\partial}{\partial x} \right) \delta(\vec{r} - \vec{r}_i) \tag{54}$$

$$\sigma_{11}^{ind} = -\frac{\partial^2 \chi^{ind}}{\partial y^2}, \qquad \sigma_{22}^{ind} = -\frac{\partial^2 \chi^{ind}}{\partial x^2}, \qquad \sigma_{12}^{ind} = \frac{\partial^2 \chi^{ind}}{\partial x \partial y}. \tag{55}$$

For periodic boundary conditions used in the simulations $\hat{\sigma}^{ind}$ does not depend on the absolute position of the dislocation position. $\hat{\sigma}^{ind}$ can be determined numerically either by the Fourier transformation of the above equation or by adding the $\hat{\sigma}_{\infty}^{ind}$ stress of the appropriate periodic mirror dislocations, where $\hat{\sigma}_{\infty}^{ind}$ is the stress in an infinite medium. It has to be

mentioned here that for other boundary conditions Needleman and coworkers developed an efficient method in which the stress is the sum of the stress the dislocation would create in an infinite medium plus a "compensating" component $\hat{\sigma}^{comp}$ needed to fulfill the boundary conditions. Since $\hat{\sigma}^{comp}$ is nonsingular at the dislocations it can be calculated with finite element methods.

During the system evolution dislocation multiplication was allowed with a global and a local conditions. The global condition was set up to reflect the experimental fact that a certain amount of external work is stored in the self-energy of dislocations. To mimic this, if the external work increased a certain amount, a new dislocation dipole was added. Since the new dislocations are generated by the stress as a "local" rule, the new dipole was placed to a random position chosen with a probability proportional to the local stress. For the sake of simplicity dislocation annihilation was not taken into account.

A typical simulation result can be seen in Figure 7. As the stress-time

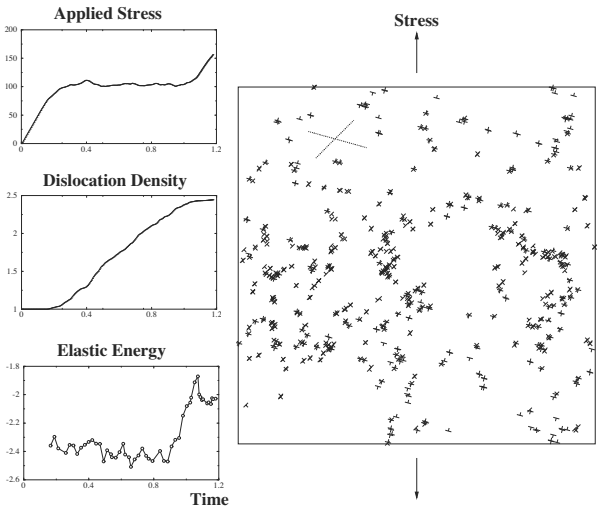


Figure 7. The time variation of stress, dislocation density and elastic energy together with the "final" dislocation configuration obtained on a dislocation system subject of a constant external deformation rate.

relation obtained indicates, the system has a finite flow stress (the initial part of the curve is linear and reversible). This is the consequence of the relatively narrow random dipole initial configuration used. While the external stress is less than the debounding stress of the dipoles there is no "free"

dislocation motion, so there is no plastic deformation. The deformation is elastic. After this, the stress-strain curve is practically horizontal with a few dislocations moving in the primary slip system. This stage shows several similarities with the easy glide stage of plastic deformation of single crystals. This stage is followed by a much steeper part of the stress-strain curve. The transition is coupled with a quite sudden rearrangement of the dislocations. A cell like structure starts to form.This stage is similar the to stage II. deformation regime.

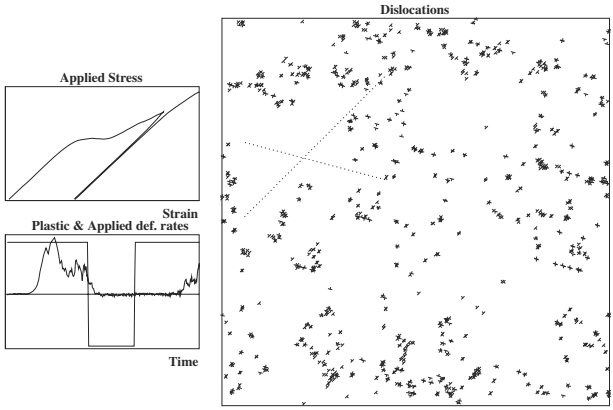


Figure 8. Hardening obtained on the 2D dislocation system.

We have investigated if the stress increment needed to keep the deformation rate constant results the enhancement of the flow stress. In stage II at a certain stress level the direction of the deformation was reversed and the system was unloaded. After this the system was reloaded again with the same rate applied earlier. As it can be seen in Figure 8 during unloading after a short relaxation period the plastic deformation rate dropped to zero (left bottom curve) and the system continued to deform elastically. During the reloading period macroscopic plastic deformation did not occur until the stress reached the level at which the system was unloaded. This means, that this simple model system shows hardening. Similar results were observed by several other authors ([54, 57]).

The above results demonstrate that already a strongly oversimplified 2D dislocation system is able to reproduce several properties of plastic deformation. Due to the huge computational demand needed for the numerical integration of the equations of motion of the dislocations, one can afford only a couple of thousand dislocations in the simulations. The problem is even

more complex in 3D. At the moment about $10^{12}\mathrm{m}^{-2}$ dislocation density can be reached in a 10x10x10 μm box [45] applying the world largest supercomputers. Although, these numerical investigations are extremely important, at the moment their applicability is very limited in the engineering practice.

3.2 Continuum theories developed for other systems, analogies and differences

At the end of the 19th century Boltzmann developed the statistical theory of fluids and gases. The key quantity he introduced is the density function $f(t,\vec{r},\vec{p})$ giving the probability density of finding an atom at the $(\vec{r},\vec{p})$ point of the phase space. From the conservation of the phase space volume (Liouville theorem) he obtained that the evolution of $f(t,\vec{r},\vec{p})$ is described by the relation

$$\frac{\partial}{\partial t}f(t,\vec{p},\vec{r})+\frac{\vec{p}}{m}\frac{\partial}{\partial\vec{r}}f(t,\vec{p},\vec{r})+\vec{F}(\vec{r})\frac{\partial}{\partial\vec{p}}f(t,\vec{p},\vec{r})=\frac{\delta f_c}{\delta t}, \tag{56}$$

where m is the mass of the atoms, $\vec{F}(\vec{r})$ is the external force and $\delta f_c/\delta t$ is the so called collision term accounting for the momentum change occurring at the collision of two atoms. Its actual form is difficult to determine. Since, however, the interaction between atoms is short ranged the collision is a local short event. This means that the collision time τ_c is much shorter than the mean collision free travelling time. As a consequence of this the collision term can be well approximated with a relaxation term leading to the equation

$$\frac{\partial}{\partial t}f(t,\vec{p},\vec{r})+\frac{\vec{p}}{m}\frac{\partial}{\partial\vec{r}}f(t,\vec{p},\vec{r})+\vec{F}(\vec{r})\frac{\partial}{\partial\vec{p}}f(t,\vec{p},\vec{r})=-\frac{1}{\tau_c}(f-f_\infty), \tag{57}$$

where f_∞ is the equilibrium Boltzmann distribution. An outstanding feature of Eq. (57) is that the Navier-Stokes equation of fluid dynamics can be derived from Eq. (57).

On the other hand it is obvious that Eq. (57) cannot be used for systems where the interaction between the particles is long ranged. Plasma is a typical example for this. The Coulomb interaction between the charged particles is long ranged. The same holds for dislocations. The interaction between straight dislocations is proportional to $1/r$.

For plasma, Vlasov obtained that the collision term $\delta f_c/\delta t$ is the sum of two terms

$$\frac{\delta f_c}{\delta t}=-\vec{F}_{sc}(\vec{r})\frac{\partial}{\partial\vec{p}}f(t,\vec{p},\vec{r})+S(\vec{r}), \tag{58}$$

where $\vec{F}_{sc}(\vec{r})$ is the so called self-consistent field generated by the charge density $\rho_c(\vec{r})$ and the electric current density $\vec{j}_c(\vec{r})$. The relations between $\rho_c(\vec{r})$, $\vec{j}_c(\vec{r})$, and $\vec{F}_{sc}(\vec{r})$ determined by the Maxwell equations, are strongly non-local. However, due to the Debye screening appearing in charged systems, the second term denoted by $S(\vec{r})$ is already local. According to these, for plasma the Boltzmann equation (56) reads as

$$\frac{\partial}{\partial t} f(t,\vec{p},\vec{r}) + \frac{\vec{p}}{m}\frac{\partial}{\partial \vec{r}} f(t,\vec{p},\vec{r}) + \left[\vec{F}(\vec{r}) + \vec{F}_{sc}(\vec{r})\right]\frac{\partial}{\partial \vec{p}} f(t,\vec{p},\vec{r}) = S(\vec{r}). \quad (59)$$

As it is explained below, a similar equation can be derived for straight dislocations. Certainly, the special properties of the dislocation-dislocation interaction and the dissipative nature of dislocation motion have to be taken into account.

3.3 Hierarchy of the different order density functions

Let us consider N straight parallel edge dislocations. As a first step let us assume that each dislocation has the same Burgers vector $\vec{b}$ parallel to the x axis. This simplification is needed only to have shorter equations. The results obtained can be easily generalised. In order to get results that are physically relevant some generalisation is definitely needed.

For this case the equation of motion of the dislocations given by the general form (53) simplifies to

$$\vec{v}_i \quad = \quad B^{-1}\vec{b}\left(\sum_{j\neq i}^{N} \tau_{ind}(\vec{r}_i - \vec{r}_j) + \tau_{ext}\right), \quad (60)$$

where τ_{ext} is the external shear and $\tau_{ind}(\vec{r})$ is the shear stress created by a dislocation. In an infinite isotropic medium

$$\tau_{ind}(\vec{r}) = \frac{b\mu}{2\pi(1-\nu)}\frac{x(x^2-y^2)}{(x^2+y^2)^2}. \quad (61)$$

As it is well known from statistical physics, instead of giving the time dependence of the coordinates of the N particles one can describe the state and the evolution of the system with the N particle probability density function f_N varying in the 6N dimensional phase space. Although, the dislocations form a nonconservative system, some of the results of statistical mechanics can be applied. Since the equations of motion of dislocations are only first order differential equations (assuming overdamped motion) for the problem considered, f_N is a 2N dimensional function of the dislocation coordinates.

By definition $f_N(t,\vec{r}_1,\vec{r}_2...\vec{r}_N)d^2\vec{r}_1d^2\vec{r}_2,..d^2\vec{r}_N$ is the probability of finding the N dislocations in the $d^2\vec{r}_1d^2\vec{r}_2..d^2\vec{r}_N$ vicinity of the $\vec{r}_1,\vec{r}_2...\vec{r}_N$ points at time t.

If we assume that the number of dislocations is conserved (later this restriction will be lifted), f_N has to fulfill the conservation law [58]

$$\begin{aligned} f_N(t,\vec{r}_1,\vec{r}_2,...,\vec{r}_N)d^2\vec{r}_1d^2\vec{r}_2,...,d^2\vec{r}_N = \\ f_N(t+\Delta t,\vec{r}_1+\vec{v}_1\Delta t,\vec{r}_2+\vec{v}_2\Delta t...\vec{r}_N+\vec{v}_N\Delta t) \\ \times d^2(\vec{r}_1+\vec{v}_1\Delta t)d^2(\vec{r}_2+\vec{v}_2\Delta t)...d^2(\vec{r}_N+\vec{v}_N\Delta t). \end{aligned} \tag{62}$$

The above relation reflects the simple fact that the probability of finding a dislocation at a certain point can change only if the dislocation moves from one point to another one. It is interesting to mention that in contrast with the conservative systems

$$\begin{aligned} f_N(t,\vec{r}_1,\vec{r}_2,...,\vec{r}_N) \neq \\ f_N(t+\Delta t,\vec{r}_1+\vec{v}_1\Delta t,\vec{r}_2+\vec{v}_2\Delta t,...,\vec{r}_N+\vec{v}_N\Delta t), \end{aligned} \tag{63}$$

This is the consequence of that the $d\vec{r}_1d\vec{r}_2...d\vec{r}_N$ volume is not conserved during the evolution of the system.

After some simple algebraic manipulations Eq. (62) can be rewritten into a partial differentiation equation

$$\frac{\partial f_N}{\partial t}+\sum_{i=1}^{N}\frac{\partial}{\partial\vec{r}_i}\left\{f_N(t,\vec{r}_1,\vec{r}_2,...,\vec{r}_N)\vec{v}_i\right\} = 0. \tag{64}$$

By substituting the left hand side of Eq. (60) into $\vec{v}_i$ we get that

$$\frac{\partial f_N}{\partial t}+\sum_{i\neq j}^{N}\frac{\partial}{\partial\vec{r}_i}\left\{f_N\vec{F}(\vec{r}_i-\vec{r}_j)\right\} = 0 \tag{65}$$

where $\vec{F}(\vec{r})=\vec{b}\tau_{ind}(\vec{r})$. ($B$ is dropped out from Eq. (65). With the appropriate selection of the time unit one can always take $B=1$.) For the sake of simplicity in the above equation the external shear was not taken into account. It is important to note that Eq. (65) is mathematically equivalent with the original equations of motion of the dislocations (60). To find a solution of the two equations are equally difficult.

For many applications, however, we do not need that detailed description represented by the N particle probability density function. A less detailed description of the system is the k-th order probability density function defined as

$$f_k(\vec{r}_1,\vec{r}_2,..,\vec{r}_k)=\int\int..\int f_N(t,\vec{r}_1,\vec{r}_2...\vec{r}_N)d^2\vec{r}_{k+1}d^2\vec{r}_{k+2}...d^2\vec{r}_N. \tag{66}$$

After integrating Eq. (65) with respect to the variables $\vec{r}_{k+1}, \vec{r}_{k+2}, ..\vec{r}_N$, from the above definition of f_k (66) we obtain that

$$\frac{\partial f_k}{\partial t} = -\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} \int \frac{\partial}{\partial \vec{r}_i}\left\{f_N \vec{F}(\vec{r}_i - \vec{r}_j)\right\} d^2\vec{r}_{k+1} d^2\vec{r}_{k+2} ... d^2\vec{r}_N \quad (67)$$

The double sum at the right hand side of the equation can be split into three parts

$$\begin{aligned}
&\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} \int \frac{\partial}{\partial \vec{r}_i}\left\{f_N \vec{F}(\vec{r}_i - \vec{r}_j)\right\} d^2\vec{r}_{k+1} d^2\vec{r}_{k+2} ... d^2\vec{r}_N = \\
&\quad \sum_{i=1}^{k}\sum_{j=1,j\neq i}^{k} \frac{\partial}{\partial \vec{r}_i}\left\{f_k \vec{F}(\vec{r}_i - \vec{r}_j)\right\} \\
&\quad + \sum_{i=1}^{k}\sum_{j=k+1}^{N} \int \frac{\partial}{\partial \vec{r}_i}\left\{f_N \vec{F}(\vec{r}_i - \vec{r}_j)\right\} d^2\vec{r}_{k+1} d^2\vec{r}_{k+2} ... d^2\vec{r}_N \\
&\quad + \sum_{i=k+1}^{N}\sum_{j=1,j\neq i}^{N} \int \frac{\partial}{\partial \vec{r}_i}\left\{f_N \vec{F}(\vec{r}_i - \vec{r}_j)\right\} d^2\vec{r}_{k+1} d^2\vec{r}_{k+2} ... d^2\vec{r}_N .
\end{aligned} \quad (68)$$

The last term is the integral of a div, so it can be transformed into a contour integral along the border of the system. Assuming that the distribution functions tend to zero fast enough at infinity, this term vanishes. Taking into account that f_N needs to be invariant with respects to swapping the coordinates of two dislocations we get that

$$\begin{aligned}
&\frac{\partial f_k}{\partial t} + \sum_{i=1}^{k}\sum_{j=1,j\neq i}^{k} \frac{\partial}{\partial \vec{r}_i}\left\{f_k \vec{F}(\vec{r}_i - \vec{r}_j)\right\} \\
&\qquad + (N-k)\int \frac{\partial}{\partial \vec{r}_i}\left\{f_{k+1} \vec{F}(\vec{r}_i - \vec{r}_{k+1})\right\} d^2\vec{r}_{k+1} = 0.
\end{aligned} \quad (69)$$

As it can be seen the equation for the k-th order probability distribution function depends on the $k+1$-th order one. So, the reduction procedure applied results a hierarchy of the equations. In fluid dynamics and plasma physics this is called as BBGKY hierarchy.

For our further consideration the equations for f_1 and f_2 play an important role, so we give their explicit forms [58]:

$$\frac{\partial \rho_1(\vec{r}_1, t)}{\partial t} + \int \frac{\partial}{\partial \vec{r}_1}\left\{\rho_2(\vec{r}_1, \vec{r}_2, t)\vec{F}(\vec{r}_1 - \vec{r}_2)\right\} d^2\vec{r}_2 \quad = \quad 0 \quad (70)$$

and

$$\frac{\partial \rho_2(\vec{r}_1,\vec{r}_2,t)}{\partial t}+\left(\frac{\partial}{\partial \vec{r}_1}-\frac{\partial}{\partial \vec{r}_2}\right)\rho_2(\vec{r}_1,\vec{r}_2,t)\vec{F}(\vec{r}_1-\vec{r}_2)$$
$$+\frac{\partial}{\partial \vec{r}_1}\int \rho_3(\vec{r}_1,\vec{r}_2,\vec{r}_3,t)\vec{F}(\vec{r}_1-\vec{r}_3)d^2\vec{r}_3+1\leftrightarrow 2=0, \qquad (71)$$

where the notations $\rho_1=Nf_1, \rho_2=N(N-1)f_2, \rho_3=N(N-1)(N-2)f_3$ were introduced. The advantage of using these quantities is that, in contrast with the probability densities f_1, f_2 and f_3 normalised to 1, they are system size independent. They are commonly referred to as one, two and three particle density functions, respectively.

It is useful to show that Eq. (70) can be derived with another method, too [59]. This can help to have a deeper understanding of the physical meaning of the equation obtained. As a first step let us multiply (60) with $\delta(\vec{r}-\vec{r}_i)$ and take its derivative with respect to $\vec{r}$:

$$\frac{d}{d\vec{r}}\left\{\frac{d\vec{r}_i}{dt}\delta(\vec{r}-\vec{r}_i)\right\} = \frac{d}{d\vec{r}}\left\{\left(\sum_{j\neq i}^{N}\vec{F}(\vec{r}_i-\vec{r}_j)\right)\delta(\vec{r}-\vec{r}_i)\right\}. \qquad (72)$$

It is useful to introduce the "discrete" dislocation density

$$\rho_d(\vec{r}) = \sum_{i=1}^{N}\delta(\vec{r}-\vec{r}_i) \qquad (73)$$

that is the same as ρ_{d+} defined in subsection 2.4, but since in the present analysis only one type of dislocation was considered, the subscript + was dropped. With this, the left hand side of Eq. (72) can be rewritten into a weighted integral. Furthermore, taking into account that

$$\frac{d}{d\vec{r}}\left\{\frac{d\vec{r}_i}{dt}\delta(\vec{r}-\vec{r}_i)\right\}=-\frac{d\vec{r}_i}{dt}\frac{d}{d\vec{r}_i}\delta(\vec{r}-\vec{r}_i)=-\frac{d}{dt}\delta(\vec{r}-\vec{r}_i), \qquad (74)$$

from Eq. (72) we get that

$$-\frac{d}{dt}\delta(\vec{r}-\vec{r}_i) \qquad (75)$$
$$= \frac{d}{d\vec{r}}\left\{\left(\int \vec{F}(\vec{r}-\vec{r}')[\rho_d(\vec{r}')-\delta(\vec{r}-\vec{r}')]d^2\vec{r}'\right)\delta(\vec{r}-\vec{r}_i)\right\}$$

(where $\delta(\vec{r}-\vec{r}')$ beside $\rho_d(\vec{r}')$ is needed to avoid self dislocation interaction.) By summing up with respect to i we conclude

$$-\frac{d}{dt}\rho_d(\vec{r}) = \frac{d}{d\vec{r}}\left\{\left(\int \vec{F}(\vec{r}-\vec{r}')[\rho_d(\vec{r}')-\delta(\vec{r}-\vec{r}')]d^2\vec{r}'\right)\rho_d(\vec{r})\right\}, \qquad (76)$$

which is a nonlinear strongly non-local equation for the "discrete" dislocation density $\rho_d(\vec{r})$. Like it was done with the field equation (46), to get rid of the singular character of $\rho_d(\vec{r})$ we can coarse grain Eq. (76). By introducing the coarse grained quantities

$$\rho_1(\vec{r}) =< \rho_{disc}(\vec{r}) > \tag{77}$$

$$\rho_2(\vec{r}_1, \vec{r}_2) =< \rho_{disc}(\vec{r}_1)\rho_{disc}(\vec{r}_2) - \rho_{disc}(\vec{r}_1)\delta(\vec{r}_1 - \vec{r}_2 >, \tag{78}$$

we get back Eq. (70) derived earlier. The procedure applied above clearly shows that the form of Eq. (70) does not depend on the actual form of the window function applied for the coarse graining. However, $\rho_1(\vec{r})$ and $\rho_2(\vec{r}_1, \vec{r}_2)$ can depend on $w(\vec{r})$ chosen. Certainly, this is not a problem until we do not assume some relation between $\rho_1(\vec{r})$ and $\rho_2(\vec{r}_1, \vec{r}_2)$. We can say that Eq. (70) is exact but it is not enough to describe the time evolution of the dislocation density.

Before we discuss how a closed theory can be obtained, the above results have to be generalised for the case where Burgers vector of the dislocations are not the same. The simplest generalisation is if we allow that the Burgers vectors of the dislocations can differ in sign. This is still a strong simplification of a real dislocation ensemble but an important step forward. Without going into the details with a similar procedure explained above one can find that

$$\frac{\partial \rho_+(\vec{r}_1, t)}{\partial t} + \vec{b}\frac{\partial}{\partial \vec{r}_1}\left[\rho_+(\vec{r}_1, t)\tau_{ext} + \int \left\{\rho_{++}(\vec{r}_1, \vec{r}_2, t) - \rho_{+-}(\vec{r}_1, \vec{r}_2, t)\right\} \tau_{ind}(\vec{r}_1 - \vec{r}_2)d\vec{r}_2\right] = 0 \tag{79}$$

$$\frac{\partial \rho_-(\vec{r}_1, t)}{\partial t} + \vec{b}\frac{\partial}{\partial \vec{r}_1}\left[-\rho_-(\vec{r}_1, t)\tau_{ext} + \int \left\{\rho_{--}(\vec{r}_1, \vec{r}_2, t) - \rho_{-+}(\vec{r}_1, \vec{r}_2, t)\right\} \tau_{ind}(\vec{r}_1 - \vec{r}_2)d\vec{r}_2\right] = 0 \tag{80}$$

where $\vec{b}$ is the Burgers vector of the positive signed dislocations. The subscripts "+" and "-" indicate the sign of the Burgers vector the different density functions are corresponding to. We mention here that the negative signs in front of ρ_{+-} and ρ_{-+} in Eqs. (79) and (80) come from the simple fact that the interaction force acting between dislocations with opposite signs is $-F_{ind}$.

By adding and substituting the two equations we obtain:

$$\frac{\partial\rho(\vec{r}_1,t)}{\partial t}+\vec{b}\frac{\partial}{\partial\vec{r}_1}[\kappa(\vec{r_1},t)\tau_{ext}+\int\{\rho_{++}(\vec{r}_1,\vec{r}_2,t)+\rho_{--}(\vec{r}_1,\vec{r}_2,t) -\rho_{+-}(\vec{r}_1,\vec{r}_2,t)-\rho_{-+}(\vec{r}_1,\vec{r}_2,t)\}\tau_{ind}(\vec{r}_1-\vec{r}_2)d\vec{r}_2]=0, \tag{81}$$

$$\frac{\partial\kappa(\vec{r}_1,t)}{\partial t}+\vec{b}\frac{\partial}{\partial\vec{r}_1}[\rho(\vec{r_1},t)\tau_{ext}+\int\{\rho_{++}(\vec{r}_1,\vec{r}_2,t)-\rho_{--}(\vec{r}_1,\vec{r}_2,t) -\rho_{+-}(\vec{r}_1,\vec{r}_2,t)+\rho_{-+}(\vec{r}_1,\vec{r}_2,t)\}\tau_{ind}(\vec{r}_1-\vec{r}_2)d\vec{r}_2]=0 \tag{82}$$

where $\rho(\vec{r},t)=\rho_+(\vec{r},t)+\rho_-(\vec{r},t)$ is the total and $\kappa(\vec{r},t)=\rho_+(\vec{r},t)-\rho_-(\vec{r},t)$ is the signed dislocation density. (κ is the same as $<\kappa>$ introduced in Eq. (49) but to have shorter equations the brackets $<..>$ were omitted .)

3.4 Evolution of the plastic shear

Before we discuss how a closed theory can be obtained for the evolution of ρ and κ it is useful to analyse the evolution of plastic shear. For the dislocation geometry considered the only non-vanishing component of the dislocation density tensor is

$$\alpha_{31}=b\kappa. \tag{83}$$

According to the definition of $\hat{\alpha}$ given by Eq. (10) for the plane problem considered the only component of the plastic distortion contributing to α_{31} is β^p_{21} and

$$b\kappa=-\frac{\partial\beta^p_{21}}{\partial x}. \tag{84}$$

With the notation $\gamma=\beta^p_{21}$ commonly used, the above equation can be rewritten as

$$\kappa=-\frac{\vec{b}}{b^2}\frac{d\gamma}{d\vec{r}}, \tag{85}$$

i.e. κ is proportional to the gradient of the plastic shear. With other words, this means, to get spatially varying plastic shear one has to introduce dislocations. This is why κ is often called geometrically necessary dislocation (GND) density.

Taking the time derivative of Eq. (85) we get that

$$\frac{\partial\kappa}{\partial t}=-\frac{\vec{b}}{b^2}\frac{d\dot{\gamma}}{d\vec{r}}. \tag{86}$$

By comparing this with Eq. (82) we obtain an explicit expression for the plastic shear rate $\dot{\gamma}$:

$$\dot{\gamma} = b^2[\rho(\vec{r_1}, t)\tau_{ext} + \int \{\rho_{++}(\vec{r}_1, \vec{r}_2, t) - \rho_{--}(\vec{r}_1, \vec{r}_2, t) - \rho_{+-}(\vec{r}_1, \vec{r}_2, t) + \rho_{-+}(\vec{r}_1, \vec{r}_2, t)\}\tau_{ind}(\vec{r}_1 - \vec{r}_2)d\vec{r}_2]. \tag{87}$$

3.5 Self-consistent field approximation

In order to have a closed continuum theory describing the evolution of the dislocation system, the (69) hierarchy of equations has to be cut at some order. In order to do this, from some considerations independent from the Eq. (69) we have to give how the density functions with order higher than a given one can be built from the lower order ones. The simplest possible assumption is that the two particle density functions are the products of the one particle density functions [59], i.e.

$$\rho_{ss'}(\vec{r}_1, \vec{r}_2, t) = \rho_s(\vec{r}_1)\rho_{s'}(\vec{r}_2), \;\; s, s' \in \{+, -\}. \tag{88}$$

This means, that the short range correlations are neglected. As it is explained below this leads to a self-consistent field theory. Similar approximation is often used in plasma physics.

By substituting Eq. (88) into Eqs. (81,82) we arrive at

$$\frac{\partial \rho(\vec{r}, t)}{\partial t} + \vec{b}\frac{\partial}{\partial \vec{r}}[\kappa(\vec{r}, t)\{\tau_{sc}(\vec{r}, t) + \tau_{ext}\}] = 0, \tag{89}$$

$$\frac{\partial \kappa(\vec{r}, t)}{\partial t} + \vec{b}\frac{\partial}{\partial \vec{r}}[\rho(\vec{r}, t)\{\tau_{sc}(\vec{r}, t) + \tau_{ext}\}] = 0, \tag{90}$$

where

$$\tau_{sc}(\vec{r}) = \int \kappa(\vec{r}_1, t)\tau_{ind}(\vec{r} - \vec{r}_1)d\vec{r}_1 \tag{91}$$

is a field (with stress dimension) created by the coarse grained signed dislocation density. τ_{sc} is often called as self-consistent stress field. However, τ_{sc} is not a "new" quantity. From Eq. (61) one can see that τ_{sc} fulfill the field equations

$$\Delta^2\chi = \frac{2b\mu}{(1-\nu)}\frac{\partial}{\partial y}\kappa(\vec{r}), \qquad \tau_{sc} = \frac{\partial^2}{\partial x \partial y}\chi \tag{92}$$

If we compare Eq. (92) with Eqs. (50, 52) we can see that τ_{sc} is nothing but the coarse grained shear stress $< \sigma >_{12}$.

It is important to note that dislocation multiplication and annihilation can also be taken into account by adding an $f(\rho, \tau_{ext} + \tau_{sc}, ...)$ source term to the right hand side of Eq. (89):

$$\frac{\partial \rho(\vec{r},t)}{\partial t} + \vec{b}\frac{\partial}{\partial \vec{r}}[\kappa(\vec{r},t)\{\tau_{sc}(\vec{r},t) + \tau_{ext}\}] \quad = \quad f(\rho, \tau_{ext} + \tau_{sc}, ...). \quad (93)$$

Determining the actual form of the source term is a difficult issue. We will come back to this problem later on, but it has to be stressed at this point that Eq. (90) has to remain unchanged because it expresses that the total net Burgers vector of the dislocation system cannot change during deformation.

3.6 Stability analysis

Since dislocation multiplication is a "local" event, it is plausible to assume that in a given point the source term $f(\rho, \tau_{ext} + \tau_{sc}, ...)$ depends on only the local values of the dislocation density and the total shear stress $\tau = \tau_{ext} + \tau_{sc}$ [59]. It is easy to see that in this case Eqs. (90, 93) have a trivial solution that is $\kappa(\vec{r},t) = 0$,and $\rho(\vec{r},t) = \rho_0(t)$ where $\rho_0(t)$ is the solution of equation

$$\frac{d\rho_0}{dt} = f(\rho_0, \tau_{ext}). \quad (94)$$

Since, however, the equations are strongly nonlinear it is important to analyse the stability of this homogeneous solution [59]. For this, let us linearise Eqs. (89,90,92) around the trivial solution

$$\frac{d}{dt}\rho' + \frac{\partial}{\partial x}\{b\tau_{ext}\kappa'\} \quad = \quad \left.\frac{\partial f}{\partial \rho}\right|_{\rho=\rho_0}\rho' + \left.\frac{\partial f}{\partial \tau}\right|_{\tau=\tau_{ext}}\tau', \quad (95)$$

$$\frac{d}{dt}\kappa' + \frac{\partial}{\partial x}\{b(\tau_{ext}\rho' + \rho_0\tau')\} \quad = \quad 0, \quad (96)$$

$$\Delta^2\chi' \quad = \quad \frac{2b\mu}{(1-\nu)}\frac{\partial}{\partial y}\kappa', \qquad \tau' = \frac{\partial^2}{\partial x \partial y}\chi', \quad (97)$$

where κ', $\rho' = \rho - \rho_0$, and τ' are small perturbations. The solution of the linearised equations can be found in the form (assuming that $\rho_0(t)$ varies

slowly in time)

$$\begin{bmatrix} \tau'(\vec{r},t) \\ \kappa'(\vec{r},t) \\ \rho'(\vec{r},t) \\ \chi'(\vec{r},t) \end{bmatrix} = \begin{bmatrix} \overline{\tau'} \\ \overline{\kappa} \\ \overline{\rho'} \\ \overline{\chi'} \end{bmatrix} exp\{\lambda t + i(q_x x + q_y y)\}. \tag{98}$$

Substituting this form into the Eqs. (95-96) one can find that λ and the wave vector (q_x, q_y) have to fulfill the characteristic equation

$$\begin{vmatrix} \lambda - \frac{\partial f}{\partial \rho}\Big|_{\rho=\rho_0} & iq_x b\tau_{ext} + i\frac{T(\Phi)}{q_x}\frac{\partial f}{\partial \tau}\Big|_{\tau=\tau_{ext}} \\ iq_x b\tau_{ext} & \lambda + \rho_0 T(\Phi) \end{vmatrix} = 0, \tag{99}$$

where

$$T(\Phi) = \frac{2b^2\mu}{(1-\nu)}\frac{q_x^2 q_y^2}{(q_x^2+q_y^2)^2} = \frac{b^2\mu}{2(1-\nu)}\sin^2(2\Phi), \tag{100}$$

in which Φ is the angle between the x axis and the wave vector.

The homogeneous solution is stable only if the real part of λ is non-positive for any wave vector (q_x, q_y). This guarantees that there is no growing perturbation.

As a first step let us analyse the stability of the homogeneous solution if the total number of dislocations is conserved, i.e. if the source term $f(\rho,\tau)$ is zero. In this case the solution of the characteristic equation (99) reads as

$$\lambda_{1,2} = \frac{-T(\Phi)\rho_0 \pm \sqrt{T(\Phi)^2\rho_0^2 - 4(b\tau_{ext})^2 q_x^2}}{2}. \tag{101}$$

Since, $T(\Phi)$ is non-negative the real part of both λ_1 and λ_2 are non-positive, so in the absence of source term the homogeneous solution is stable. However, an important feature of the linearised equations is that if the wave vector is parallel either to the x or the y axes the real parts of λ_1 and λ_2 vanish. This means, periodic perturbations which are either parallel or perpendicular to the Burgers vector neither growth nor die out, they are marginally stable.

To see the influence of the source term $f(\rho,\tau)$ it is enough to study the sum of the two roots λ_1 and λ_2. From Eq. (99) we get that

$$\lambda_1 + \lambda_2 = -T(\Phi)\rho_0 + \frac{\partial f}{\partial \rho}\Big|_{\rho=\rho_0}. \tag{102}$$

Since $T(\Phi)$ vanishes for $\Phi = 0$ and $\Phi = 90^0$, if $\partial f/\partial\rho$ is positive, there is a wave vector domain where at least one of the two λ-s is positive. This means, if $\partial f/\partial\rho > 0$ the homogeneous solution is not stable any more.

3.7 Numerical studies

Since the self-consistent field equations (90,93) are complicated nonlinear equations, studying the properties of their solutions requires numerical investigations. The numerical results presented in this paper [59] were obtained on a 128x128 grid defined in a square simulation area with periodic boundary conditions. The time integration of Eqs. (90,93) was carried out by the Newton method. The internal stress was determined from Eq. (92) by fast Fourier transformation. In each calculation a constant dislocation density and a random κ distribution was used as initial configuration. For the source term the form

$$f(\rho,\tau) = C(\dot{\gamma}\tau - 0.09b^2\mu^2\rho^2) = C(\rho\tau^2 - 0.09b^2\mu^2\rho^2) \tag{103}$$

was used. In expression (103) the first, dislocation creation term mimics the experimental observation that a certain amount of plastic work is stored as the self energy of dislocations. The second, annihilation term simply expresses that annihilation requires to have two dislocations close to each other. The constants are determined according to the Taylor relation

$$\tau_{ext} = 0.3b\mu\sqrt{\rho_0}, \tag{104}$$

which we expect to hold at steady state.

Since, apart from C, the size of the simulation area and the material parameters can be scaled out from the equations the input parameters ($\rho(t = 0,\vec{r})$, $\kappa(t = 0,\vec{r})$, τ_{ext}, C) and the results of the numerical calculations are given in arbitrary units.

A typical snapshot obtained at periodic external stress can be seen in Figure 9. Both the dislocation density map (left box) and its autocorrelation function (right box) indicate that the dislocations tend to form a more or less periodic arrangement of dense regions. The dislocation pattern developed is very similar to the so called "matrix" structure experimentally observed on fatigued fcc single crystals [60] (see Figure 10).

It should be mentioned, that if the gradients in the dislocation pattern developing during the deformation become large the numerical solution blows up indicating that the self-consistent field theory can become unstable. In order to resolve this problem the influence of dislocation-dislocation correlation has to be analysed.

3.8 The role of dislocation-dislocation correlation

The self-consistent field theory explained above was obtained by assuming that the two particle density functions are the product of the corresponding one particle densities. Without restricting generality, the two particle

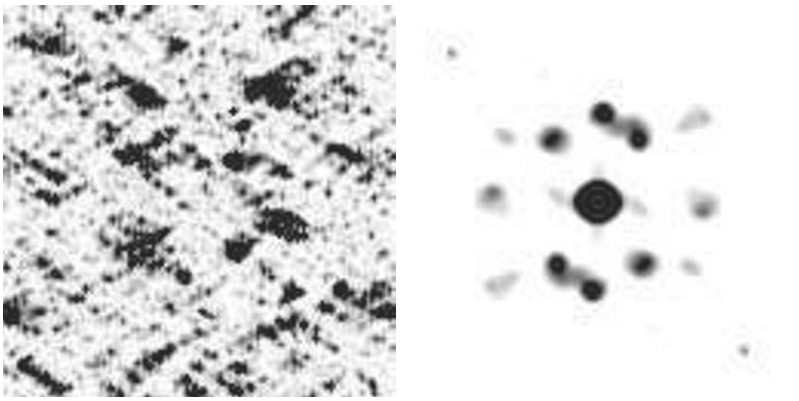


Figure 9. Dislocation density map (left box) and its autocorrelation function $A(\vec{r}) = \int \rho(\vec{r} - \vec{r}')\rho(\vec{r}')d\vec{r}'$ (right box) obtained at periodic external load. (The glide direction is horizontal.)

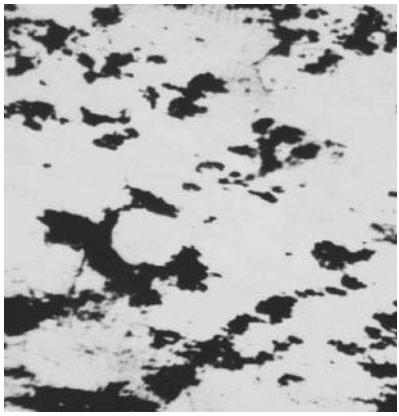


Figure 10. Dislocation "matrix" structure obtained on fatigued Cu single crystal oriented for single slip [60].

density functions can be given in the form:

$$\rho_{ss'}(\vec{r}_1,\vec{r}_2,t)=\rho_s(\vec{r}_1)\rho_{s'}(\vec{r}_2)(1+d_{ss'}(\vec{r}_1,\vec{r}_2)) \qquad s,s'\in\{+,-\} \quad (105)$$

where $d_{ss'}$ is called dislocation-dislocation correlation function. In order to be able to say something about the correlation function as a first step it is useful to analyse the properties of dislocation-dislocation correlations in an originally homogeneous relaxed dislocation system [61, 62]. Although the BBGKY hierarchy explained earlier gives the possibility to investigate the properties of $d_{ss'}$ analytically (assuming something about the three particle density functions), but due to the complicated nonlinear character of the equations, apart from some simple general statements, it is rather difficult to say anything about $d_{ss'}$.

For initially homogeneous, relaxed dislocation systems $d_{ss'}$ can be determined by DDD simulations. For this we do not have to study large systems (a few 100 dislocations is already enough), but we need several (around 1000) relaxed configurations to have the necessary statistics. Knowing the relaxed positions of dislocations $d_{ss'}$ can be determined by simply counting the number of dislocation pairs at different relative positions.

Figure 11 shows the correlation function $d=d_{++}+d_{--}+d_{+-}+d_{-+}$ obtained numerically. In the simulations parallel straight edge dislocations

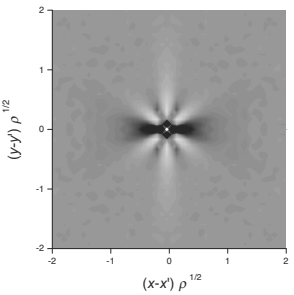


Figure 11. The total $d=d_{++}+d_{--}+d_{+-}+d_{-+}$ dislocation-dislocation correlation function determined numerically at single slip geometry [61].

were considered at single slip geometry ($\vec{b}$ is parallel to the x axis). The number of dislocations was kept constant. Initially the dislocations were

randomly distributed. By the numerical integration of Eq. (60) the relaxed dislocation configuration was determined at zero external stress.

The two most important properties of the correlation function are that next to the origin it is inversely proportional to the distance from the origin (analytical investigations revealed the same), and it decays to zero exponentially within a couple of average dislocation spacing. So, for originally random relaxed dislocation configurations the dislocation-dislocation correlation is short-ranged. With other words, if the distance of two dislocations is larger than a couple of times the average dislocation spacing the correlation between them is negligible [61, 62]. We have to keep in mind, however, that this is valid only if the relaxed configuration is obtained from an initially random dislocation distribution. One can obviously set up initial configuration that relaxes to a strongly correlated state like for example a Taylor lattice. The problem is related to the constrained motion of dislocations. Since in the simulations only dislocation glide is allowed and dislocation multiplication is excluded, the number of dislocations in any narrow strip parallel to the Burgers vector cannot change during the relaxation of the system. It is determined by the initial configuration. The system does not "forget" fully the initial configuration. In reality, of course, the number of dislocations in a strip is determined by the dislocation multiplication. Since, however in 2D there is no "natural" law for dislocation multiplication one should investigate the correlation properties in 3D. Investigations are under their way, but at the moment we do not have conclusive results. Based on the results of 2D simulations in the following we assume that the correlation function is short range.

According to the results explained above it is plausible to assume that the correlation functions $d_{ss'}(\vec{r}_1, \vec{r}_2)$ defined by the Eq. (105) can be approximated with the correlation function corresponding to a homogeneous system with dislocation density $\rho(\vec{r}_1)$. It follows that $d_{ss'}(\vec{r}_1, \vec{r}_2)$ practically depends only on $(\vec{r}_1 - \vec{r}_2)$, the direct $\vec{r}_1$ or $\vec{r}_2$ dependence is weak, it appears only through the spatial variation of the dislocation density, i.e.

$$\rho_{ss'}(\vec{r}_1, \vec{r}_2, t) = \rho_s(\vec{r}_1)\rho_{s'}(\vec{r}_2)(1 + d_{ss'}(\vec{r}_1 - \vec{r}_2)) \qquad s, s' \in \{+, -\} \quad (106)$$

Similar approximation is used successfully for many other systems like for example in first principle quantum mechanics calculations to estimate the exchange energy. It is called "local density approximation".

By substituting Eq. (106) into Eqs. (81,82) after a long, but straightforward calculation we arrive at

$$\frac{\partial \rho(\vec{r}, t)}{\partial t} + \vec{b}\frac{\partial}{\partial \vec{r}}[\kappa(\vec{r}, t)\{\tau_{sc}(\vec{r}) + \tau_{ext} - \tau_f(\vec{r}) + \tau_b(\vec{r})\}] \quad = \quad 0, \quad (107)$$

$$\frac{\partial\kappa(\vec{r},t)}{\partial t}+\vec{b}\frac{\partial}{\partial\vec{r}}[\rho(\vec{r},t)\left\{\tau_{sc}(\vec{r})+\tau_{ext}-\tau_f(\vec{r})+\tau_b(\vec{r})\right\}] = 0, \quad (108)$$

where

$$\tau_f(\vec{r})=\frac{1}{2}\int\rho(\vec{r}_1)d^a(\vec{r}-\vec{r}_1)\tau_{ind}(\vec{r}-\vec{r}_1)d\vec{r}_1 \quad (109)$$

$$\tau_b(\vec{r})=\int\kappa(\vec{r}_1)d(\vec{r}-\vec{r}_1)\tau_{ind}(\vec{r}-\vec{r}_1)d\vec{r}_1, \quad (110)$$

in which the notations

$$d(\vec{r}) = 1/4[d_{++}(\vec{r})+d_{--}(\vec{r})+d_{+-}(\vec{r})+d_{-+}(\vec{r})] \quad (111)$$
$$d^a(\vec{r}) = 1/2[d_{+-}(\vec{r})-d_{-+}(\vec{r})] \quad (112)$$

are introduced. Due to the following obvious symmetry properties of the correlation functions

$$d_{+-}(\vec{r})=d_{-+}(-\vec{r}), \qquad d_{++}(\vec{r})=d_{++}(-\vec{r}), \qquad d_{--}(\vec{r})=d_{--}(-\vec{r}) \quad (113)$$

$d(\vec{r})$ is an even, while $d^a(\vec{r})$ is an odd function of $\vec{r}$. Furthermore, since the correlation functions correspond to a homogeneous system there is no other internal length scale but the average dislocation spacing $1/\sqrt{\rho}$. It is obvious from simple dimensional analysis that the correlation functions depend only on the dimensionless quantity $\vec{r}\sqrt{\rho}$, i.e. $d(\sqrt{\rho}\vec{r})$ and $d^a(\sqrt{\rho}\vec{r})$.

Taking into account that the corelation functions decay to zero within a few dislocation spacing the fields $\kappa(\vec{r}_1)$ and $\rho(\vec{r}_1)$ appearing in Eqs. (109,110) can be approximated by their Taylor expansion around the point $\vec{r}$. Keeping only the first nonvanishing terms we get that [61, 62]

$$\tau_f(\vec{r})=\frac{\rho(\vec{r})}{2}\int d^a(\vec{r})\tau_{ind}(\vec{r})d\vec{r}, \quad (114)$$

$$\tau_b(\vec{r})=-\frac{\partial\kappa(\vec{r})}{\partial\vec{r}}\int\vec{r}d(\vec{r})\tau_{ind}(\vec{r})d\vec{r}. \quad (115)$$

(To obtain expressions (114,115) one has to take into account the symmetry properties of $d(\vec{r})$ and $d^a(\vec{r})$ explained above, and the relation $\tau_{ind}(\vec{r}) = -\tau_{ind}(-\vec{r})$.)

With the variable substitution $\vec{\eta}=\sqrt{\rho}\vec{r}$, $\tau_f(\vec{r})$ reads as

$$\tau_f(\vec{r})=\frac{\sqrt{\rho(\vec{r})}}{2}\int d^a(\vec{\eta})\tau_{ind}(\vec{\eta})d^2\vec{\eta}. \quad (116)$$

where we took into account that τ_{ind} is proportional to $1/|\vec{r}|$. By substituting the actual form of τ_{ind} given by Eq. (61) into Eq. (116) we get that

$$\tau_f(\vec{r}) = \frac{AC}{2} b \sqrt{\rho(\vec{r})}, \tag{117}$$

where $A = \mu/[2\pi(1-\nu)]$ and

$$C = \int \frac{\eta_x(\eta_x^2 - \eta_y^2)}{(\eta_x^2 + \eta_y^2)^2} d'^a(\eta_x, \eta_y) d\eta_x d\eta_y. \tag{118}$$

in which the prime in $d'^a(\eta_x, \eta_y)$ indicates that $\vec{\eta}$ has to be measured in unit of average dislocation spacing. In order to see the physical meaning of the above expression, the external stress dependence of the parameter C has to be analysed. The correlation function d_{+-} obviously varies with external stress (the equilibrium configuration of dislocation dipoles varies if stress is applied). Let us assume that the change of d_{+-} resulted by the external stress increases C. Since the change in d_{-+} is the opposite of the change of d_{+-} this causes also the increase of C because of the minus sign in front of d_{-+} in the definition of d^a. As we see, the parameter C depends on the external stress. Beside this, τ_f scales with $\sqrt{\rho}$. These support association of τ_f with the flow stress. Certainly, we have to be careful with this statement. In real dislocation systems hardening is caused by the forest dislocations which are not included into our model in any sense. Nevertheless, a stress like term showing similar properties as the flow stress appears naturally in the theory. The actual form of the stress dependence of C is difficult to determine, but one can speculate that τ_f acts as static friction. It prevents dislocation motion, but it has a maximum scaling with $\sqrt{\rho}$.

After this let us analyse τ_b in more details. With the same variable substitution applied above we get that

$$\tau_b(\vec{r}) = -\frac{\partial \kappa(\vec{r})}{\partial \vec{r}} \frac{1}{\rho(\vec{r})} \int \vec{\eta} d'(\vec{\eta}) \tau_{ind}(\vec{\eta}) d^2\vec{\tau}. \tag{119}$$

i.e.

$$\tau_b(\vec{r}) = -\frac{AD\vec{b}}{\rho} \frac{\partial \kappa(\vec{r})}{\partial \vec{r}}, \tag{120}$$

where

$$D = \int \frac{\eta_x^2(\eta_x^2 - \eta_y^2)}{(\eta_x^2 + \eta_y^2)^2} d'(\eta_x, \eta_y) d\eta_x d\eta_y \tag{121}$$

is a dimensionless constant. In contrast with C, D has only a weak external stress dependence because $d(\vec{r})$ contains the sum of the two correlation functions d_{+-} and d_{-+} changing oppositely. The actual value of D can only be determined numerically. According to the numerical studies explained below it is in the order of magnitude of 1.

With the results obtained above the evolution equations (107,108) read as

$$\frac{\partial \rho(\vec{r},t)}{\partial t} + \vec{b}\frac{\partial}{\partial \vec{r}}\left[\kappa(\vec{r},t)\left\{\tau(\vec{r},t) - \tau_f - AD\frac{\vec{b}}{\rho(\vec{r},t)}\frac{\partial \kappa(\vec{r},t)}{\partial \vec{r}})\right\}\right] = f(\rho,\tau), \tag{122}$$

$$\frac{\partial \kappa(\vec{r},t)}{\partial t} + \vec{b}\frac{\partial}{\partial \vec{r}}\left[\rho(\vec{r},t)\left\{\tau(\vec{r},t) - \tau_f - AD\frac{\vec{b}}{\rho(\vec{r},t)}\frac{\partial \kappa(\vec{r},t)}{\partial \vec{r}}\right\}\right] = 0 \tag{123}$$

where $\tau = \tau_{sc} + \tau_{ext}$ is the total macroscopic stress.

From Eq. (87) the constitutive equation of the plastic shear rate can also be given as

$$\dot{\gamma} = b^2\rho(\vec{r},t)\left\{\tau(\vec{r},t) - \tau_f - AD\frac{\vec{b}}{\rho(\vec{r},t)}\frac{\partial \kappa(\vec{r},t)}{\partial \vec{r}}\right\}. \tag{124}$$

With Eq. (85)

$$\dot{\gamma} = b^2\rho(\vec{r},t)\left\{\tau(\vec{r},t) - \tau_f - AD\frac{1}{b^2\rho(\vec{r},t)}\left(\vec{b}\frac{\partial}{\partial \vec{r}}\right)^2\gamma(\vec{r},t)\right\}. \tag{125}$$

If we introduce the effective stress

$$\tau_{eff}(\vec{r}) = \tau(\vec{r}) - AD\frac{1}{b^2\rho(\vec{r})}\left(\vec{b}\frac{\partial}{\partial \vec{r}}\right)^2\gamma(\vec{r}) \tag{126}$$

it looks similar to the effective stress (2) suggested by E. Aifantis from phenomenological considerations. An important difference, however, is that in Eq. (126) the length scale $l = 1/\sqrt{\rho}$ appearing in front of the gradient term is a natural one, it is not a material parameter suggested in the phenomenological non-local continuum theories. The length scale $1/\sqrt{\rho}$ obeys an evolution equation (Eq. (refgroma.rhof).

3.9 Deformation of a constrained channel

To illustrate some implications of the evolution equations derived in the previous subsection we study a very simple example, namely a constrained

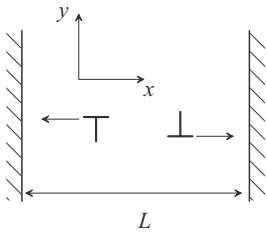


Figure 12. Geometry of a constrained channel

channel deforming in simple shear as shown in Figure 12 [62]. A channel of width L in the x direction and infinite extension in the y direction is bounded by walls that are impenetrable for dislocations (i.e., the plastic deformation in the walls is zero). The slip direction corresponds to the x direction, and the layer is sheared by a constant shear stress τ_{ext}. The whole assembly is embedded in an infinite crystal.

The system envisaged is particularly simple because it is homogeneous in the y direction (the dislocation densities depend on the coordinate x in the slip direction only). It follows from Eq. (92) that in this case the long-range self-consistent stress field is zero for an arbitrary function $\kappa(x)$, i.e., any dislocation interactions in the system are of short-range nature and hence described by the flow stress τ_f and the gradient-dependent stress τ_b.

Before investigating the behaviour resulting from Eqs. (122) and (123) and comparing it with the results obtained from discrete simulations, it is instructive to have a look at the results we get from the mean-field model defined by Eqs. (89) and (90). Since the self-consistent stress is zero, the mean-field model becomes trivial: Whatever the initial conditions, for an arbitrarily small, positive value of the external stress all positive dislocations 'condense' at the right wall and all negative dislocations at the left one. For an initially homogeneous dislocation distribution with density ρ_0, the strain achieved by this condensation is $\gamma_\infty = \rho_0 bL/2$. Hence, the system exhibits a trivial size effect (the achievable strain is proportional to the size of the system, which determines the mean dislocation path). However, as demonstrated in the following, the prediction that this strain is achieved at arbitrarily small external stress is grossly unrealistic.

We now revert to the gradient-dependent model derived in the previous subsection. We assume an initially homogeneous dislocation distribution of density ρ_0. To facilitate comparison with discrete simulations, it is convenient to introduce scaled stress, space and dislocation density variables through $\tau = Ab\sqrt{\rho_0}\tilde{\tau}$, $x = D\tilde{x}/\sqrt{\rho}$, $\rho = \rho_0\tilde{\rho}$, and $\kappa = \rho_0\tilde{\kappa}$. In scaled

variables and after corresponding re-scaling of time, Eqs. (122) and (123) read

$$\partial_t \tilde{\rho}(\tilde{x},t) = -\partial_{\tilde{x}} \left[\tilde{\kappa}(\tilde{x},t) \left\{ \tilde{\tau} - [1/\tilde{\rho}(\tilde{x},t)] \partial_{\tilde{x}} \tilde{\kappa}(\tilde{x},t) \right\} \right], \tag{127}$$

$$\partial_t \tilde{\kappa}(\tilde{x},t) = -\partial_{\tilde{x}} \left[\tilde{\rho}(\tilde{x},t) \left\{ \tilde{\tau} - [1/\tilde{\rho}(\tilde{x},t)] \partial_{\tilde{x}} \tilde{\kappa}(\tilde{x},t) \right\} \right]. \tag{128}$$

To formulate the boundary conditions at the walls located at $\tilde{x} = \pm\tilde{L}/2$, we note that no dislocations can enter the system through the walls. Hence, the density of positive dislocations (moving to the right) at the left wall and the density of negative dislocations at the right wall are zero, i.e. $\tilde{\kappa}(-\tilde{L}/2) = -\tilde{\rho}(-\tilde{L}/2)$, $\tilde{\kappa}(\tilde{L}/2) = \tilde{\rho}(\tilde{L}/2)$. Furthermore, the dislocation fluxes at the walls must be zero, which requires that $[\tilde{\rho}\tilde{\tau} - \partial_{\tilde{x}}\tilde{\kappa}] = 0$ at $\tilde{x} = \pm\tilde{L}/2$.

The initial conditions are $\tilde{\rho}(\tilde{x},0) = 1$ and $\tilde{\kappa}(\tilde{x},0) = 0$ everywhere except directly at the walls where we assume non-zero values of κ in a narrow boundary layer to satisfy the boundary conditions. We make the simplifying assumption that the effective stress can be represented as the external stress diminished by the (spatially homogeneous) flow stress of an infinite system, and perform a 'deformation experiment' as follows: we increase the effective stress from zero in an adiabatically slow manner, i.e., after each small stress increment the system is allowed to relax until it reaches a stationary configuration. After this relaxation, the scaled strain is calculated as $\tilde{\gamma} = -\int \tilde{\kappa} d\tilde{x}$ (see Eq. (85)), then the stress is increased again, etc. The resulting stress-strain curves for different values of L are compiled in Figure 13 [62]. It

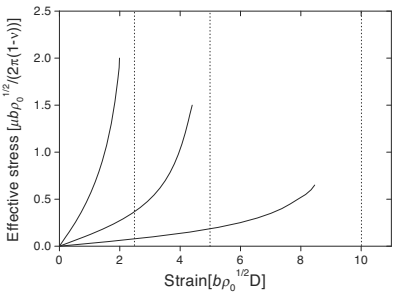


Figure 13. Stress-strain curves obtained at different channel size L.

can be seen that the behaviour is very different from the prediction of the mean-field model: the strain increases gradually with stress and reaches the

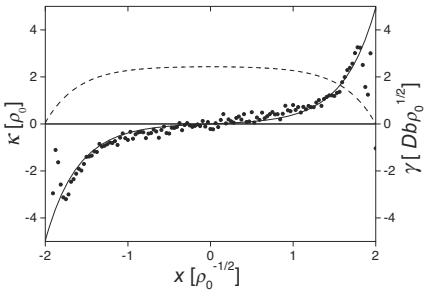


Figure 14. Signed dislocation density and strain profiles for a system of width $L = 4/\sqrt{\rho_0}$ at scaled stress $\tilde{\tau} = 2$; $\kappa(x)$ from discrete simulation (data points), $\kappa(x)$ from continuum model (full line), strain profile from continuum model (dashed line).

limit strain γ_∞ ($\tilde{\gamma}_\infty = \tilde{L}/2$ in scaled units) only asymptotically. In physical terms this behaviour stems from the fact that there is a short-range repulsion between individual dislocations of the same sign as they pile up against the walls (see Figure 14). To increase the strain towards the asymptotic strain, this repulsion must be overcome, which requires an increasing stress that diverges as $\gamma \to \gamma_\infty$.

Looking at the distribution of dislocation densities and strains within the channel, we find that at high stresses two boundary layers emerge near the walls (Figure 14). Its properties can be analysed by the equilibrium condition

$$\tau_{ext} - AD\frac{\vec{b}}{\rho(\vec{r},t)}\frac{\partial\kappa(\vec{r},t)}{\partial\vec{r}} = 0 \tag{129}$$

required to hold at steady state. (Since, for the geometry considered the self-consistent field is zero, $\tau = \tau_{ext}$). Near the boundaries most of the dislocations have the same sign, $\rho \approx |\kappa|$. According to this, near the left boundary Eq. (129) reads as

$$\tau_{ext}\kappa = -ADb\frac{d\kappa(x)}{dx} \tag{130}$$

with solution

$$\kappa(x) = \kappa_0 \exp\left\{-\frac{\tau_{ext}}{ADb}x\right\}. \tag{131}$$

(A similar expression obviously holds near the right side, too.) As it is seen, the width of the boundary layers decreases with increasing external stress.

It is interesting to compare this result with the prediction of the phenomenological gradient approach. According to Eq. (2) for the problem considered the equilibrium condition is

$$\tau_{ext} - \frac{\mu}{l^2}\frac{\kappa(x)}{dx} = 0 \tag{132}$$

with solution

$$\kappa(x) = \kappa_0 + \frac{l^2}{\mu}\tau_{ext}x \tag{133}$$

In the centre part of the channel, where ρ is nearly constant, the linear relation predicted by the phenomenological gradient approach describes well the observed variation of κ, but it is not able to account for the boundary layers.

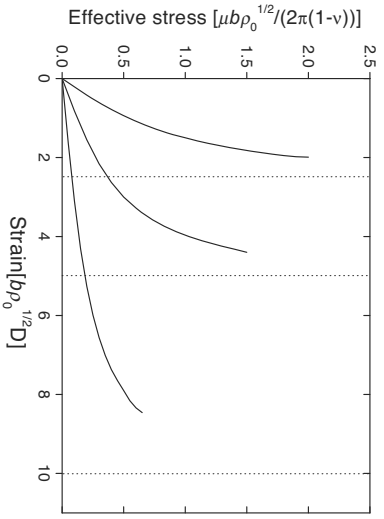


Figure 15. Comparison of stress-strain graphs; full line continuum model; data points: discrete simulation. Parameters are the same as in Figure 14.

The results obtained are compared with DDD simulations performed on the same system. To get reliable statistics, stress-strain graphs and

the corresponding dislocation density profiles were averaged over a huge ensemble (typically several thousands of simulations). An example of a dislocation density profile obtained from this procedure is illustrated in Figure 14 which shows a $\tilde{\kappa}(\tilde{x})$ profile averaged over 2000 simulations of systems with length $L = 4/\sqrt{\rho_0}$ and (periodically repeated) height $16/\sqrt{\rho_0}$. The profile shown in the figure has been taken at a scaled effective stress $\tilde{\tau} = 2$. It is seen that indeed two boundary layers emerge. From the width of these boundary layers we can directly determine the constant D for the present type of simulation, which turns out to be $D = 0.8$. Using the continuum model with this value of D yields the full line in Figure 14, which shows that the density profile obtained from the continuum model matches well the discrete simulation except in the immediate vicinity of the walls.

As seen from Figure 15, also the stress-strain curves for the discrete and continuum models exhibit almost perfect agreement. By varying the system size and initial dislocation density, we find that, for sufficiently high stresses the width of the boundary layers at fixed stress τ_{ext} is within the error margins indeed independent on the system size and the dislocation density. If the applied stress is increased, the boundary layer width is found to decrease. Again all these findings are in line with the predictions of the continuum model.

3.10 Application to metal-matrix composite

In order to demonstrate the capability of the continuum theory of dislocations explained above we shortly summarise the simulation results obtained on a 2D model system of a metal-matrix composite [54, 63]. It contains rigid rectangular particles arranged in a hexagonal packing, as illustrated in Figure 16. The cell is subjected to plane strain, simple shear, which is prescribed through the boundary conditions

$$u_1 = \pm h\Gamma \quad , \quad u_2 = 0 \quad \text{along } x_2 = \pm h, \tag{134}$$

where Γ is the applied shear. Periodic boundary conditions are imposed along the lateral sides $x = \pm w$. The slip plane normal $\vec{n}$ is in the y-direction and the Burgers vector is parallel to the x-direction. Two reinforcement morphologies were analysed having the same area fraction of 20% but different geometric arrangements of the reinforcing phase. In one morphology, material (i), the particles are square and are separated by unreinforced veins of matrix material while in the other, material (iii), the particles are rectangular and do not leave any unreinforced veins of matrix material.

The problem was studied in details by Cleveringa et al. [54, 63] with DDD simulation. Here we compare their DDD simulation results with the

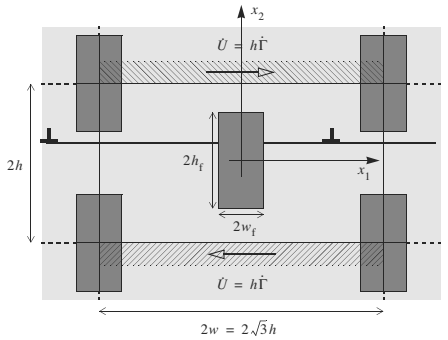


Figure 16. Unit cell in a doubly-periodic array of elastic particles, subjected to simple shear. The slip planes are taken to be parallel to the shear direction (x).

results obtained by solving Eqs. (122,123) with a finite element method (for the details of the numerical technique used see [64, 65]). For the dislocation density field equations, boundary conditions need to be specified at the boundary of the cell as well as along the interface with the elastic particles. Along the cell sides $x = \pm w$, periodic boundary conditions are applied, while along $y = \pm h$ we have the natural condition that there is no flux of dislocations across these boundaries. Similar conditions apply along the top and bottom interfaces with the particles. Along the vertical sides of the particles, we impose that the slip rate vanishes.

As it is seen in Figure 17 the stress-strain curves obtained by DDD simulation and from the continuum theory match extremely well for both reinforcement morphologies investigated [64, 65]). Figure 18 shows the ρ and κ maps obtained for the (iii) morphology. It can be seen that, like at the shear of the channel discussed above, a boundary layer of geometrically necessary dislocations develops at the vertical unpenetrable surface of the composite particles.

In conclusion it can be stated that for single slip the continuum theory developed predicts the same behaviour as the DDD simulations. How to generalise the continuum model to multiple slip is being studied intensively [66, 67]. At the moment the theory is far from being fully developed. It is even more difficult to extend the statistical approach to 3D. In order to show the possible approaches, in the next part we shortly summarise two

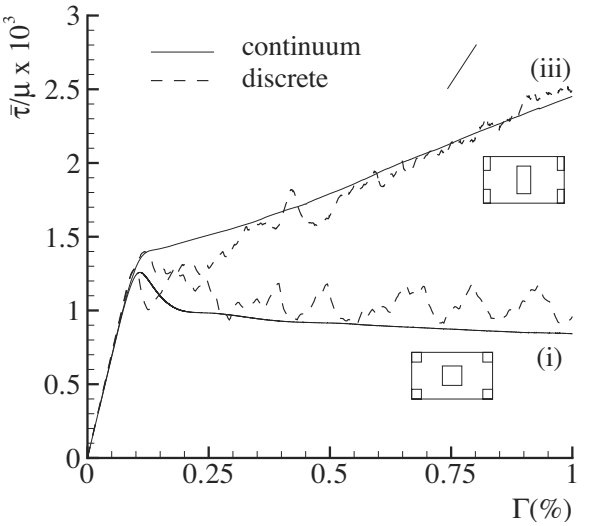


Figure 17. Stress-stain curves obtained by DDD simulations (dashed lines) and by solving the continuum Eqs. (122,123) (full line).

3D continuum models proposed recently.

3.11 Boltzmann theory of dislocations

To describe the evolution of dislocation loops El-Azab has recently proposed a new theoretical framework [68], which is the generalisation of the Boltzmann equation for line type objects. In his model only planar loops are considered, i.e. dislocation climb is excluded.

In contrast with point like particles (considered in Boltzmann's original theory), beside its position a dislocation segment has another degree of freedom, its line direction $\vec{l}$. In a plane the line direction $\vec{l}$ can be represented with a scalar parameter Θ which is the angle between $\vec{l}$ and a given direction in the plane. According to this, the probability density function f is the function of $\vec{r}$, $\vec{v}$ and Θ. (To avoid the problems related to define the mass of a dislocation, instead of the moment the velocity of the dislocation segment is used to give the state of the segment.) Since, as it is explained earlier, the probability density function is system size dependent (f is normalised

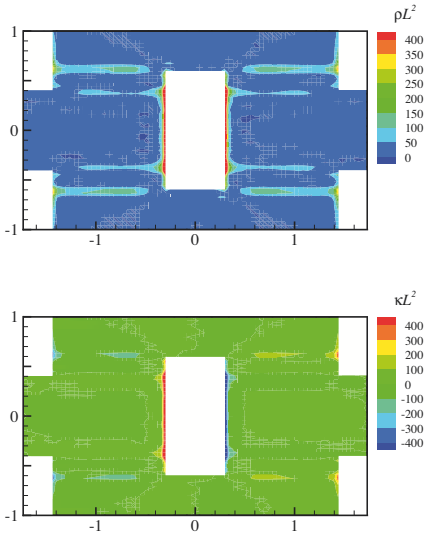


Figure 18. ρ (top box) and κ (bottom box) maps at 1$ deformation for the (iii) morphology.

to 1) it is more convenient to operate with dislocation line length density ϕ. Hence, $\phi^{(i)}(\vec{r},\vec{v},\Theta,t)$ can be given the definition: $\phi^{(i)}(\vec{r},\vec{v},\Theta,t)d^3\vec{r}d^3\vec{v}d\Theta$ is the dislocation line length with Burgers vector $\vec{b}_i$ $(i=1..N)$ in the $d^3\vec{r}d^3\vec{v}d\Theta$ phase space at time t. It has to be noted that since $\vec{v}$ is perpendicular to $\vec{l}$, $\dot{\Theta}$ is not an "independent" variable, that is why it does not appear among the arguments of $\phi^{(i)}$.

The conventional field variables are now derived from $\phi^{(i)}$. The dislocation line density is defined as follows:

$$\rho(\vec{r},t)=\sum_{i=1}^{N}\rho^{(i)}(\vec{r},t) \tag{135}$$

where

$$\rho^{(i)}(\vec{r},t)=\int\int\phi^{(i)}(\vec{r},\vec{v},\Theta,t)d^3\vec{v}d\Theta \tag{136}$$

According to Eq. (15) the dislocation density tensor $\hat{\alpha}$ is given by

$$\alpha_{mn}(\vec{r},t) = \sum_{i=1}^{N} \alpha_{mn}^{(i)}(\vec{r},t) \tag{137}$$

where

$$\alpha_{mn}^{(i)}(\vec{r},t) = \int\int l_m(\Theta) b_n^{(i)} \phi^{(i)}(\vec{r},\vec{v},\Theta,t) d^3\vec{v} d\Theta, \tag{138}$$

while from Eq. (42) the dislocation current density $\hat{j}$ reads as

$$j_{mn}(\vec{r},t) = \sum_{i=1}^{N} j_{mn}^{(i)}(\vec{r},t) \tag{139}$$

where

$$j_{nk}^{(i)}(\vec{r},t) = \int\int e_{nmp} l_m(\Theta) v_p b_k^{(i)} \phi^{(i)}(\vec{r},\vec{v},\Theta,t) d^3\vec{v} d\Theta. \tag{140}$$

Taking into account that the velocity of a dislocation segment is perpendicular to the line direction, we can conclude that the evolution equation of $\phi^{(i)}$ has the form (for the details of the derivation see [68]):

$$\left(\frac{\partial}{\partial t} + v_j \frac{\partial}{\partial x_j} + \dot{v}_j(\hat{\alpha},\hat{\sigma},..) \frac{\partial}{\partial v_j} \right) \phi^{(i)}(\vec{r},\vec{v},\Theta,t) = s^{(i)}(\vec{r},\vec{v},\Theta), \quad i = 1,..,N \tag{141}$$

where $s^{(i)}(\vec{r},\vec{v},\Theta)$ represents all the possible source terms, and $\dot{v}_j(\hat{\alpha},\hat{\sigma},..)$ stands for the acceleration of the dislocation. which is a function of the different fields like $\hat{\alpha}$ and $\hat{\sigma}$. A remarkable feature of the above equation is that the derivative with respect to Θ does not appear in the left hand side.

If we multiply the above equation with $l_m(\Theta) b_n^{(i)}$ and integrate it with respect to the velocity and Θ, after long but straightforward algebraic manipulations we conclude:

$$\dot{\alpha}_{mn}^{(i)} + e_{mkl} \frac{\partial}{\partial r_k} j_{ln}^{(i)} = S_{nm}^{(i)}, \tag{142}$$

where

$$S_{nm}^{(i)} = \int\int l_n(\Theta) b_m^{(i)} s^{(i)}(\vec{r},\vec{v},\Theta) d^3\vec{v} d\Theta. \tag{143}$$

By summing up Eq. (142) over all Burgers vectors and comparing the result with Eq. (39) we obtain an important condition the source terms have to fulfill

$$\sum_{i=1}^{N} S_{nm}^{(i)} = 0. \tag{144}$$

To get a closed equation for the evolution of $\phi^{(i)}(\vec{r}, \vec{v}, \Theta, t)$, the actual forms of $s^{(i)}(\vec{r}, \vec{v}, \Theta)$ and $\dot{v}_j(\hat{\alpha}, \hat{\sigma}, ..)$ should be given. At the present stage of the theory these are not well established.

We mention that it has been proposed recently by Hochrainer and Zaiser [69] that beside Θ the curvature of the dislocation line should also be considered as an independent variable. This may help setting up the constitutive relations.

3.12 Hydrodynamics approach proposed by Kratochvíl and Sedláček

In hydrodynamics the state of the material is described by the density $\rho(\vec{r}, t)$ and the velocity $\vec{u}(\vec{r}, t)$ fields. They are the appropriate mean values of the probability density function $f(\vec{r}, \vec{p}, t)$. One may operate with the same quantities in dislocation theory. However, to account for the line direction degree of freedom an additional field denoted by $\beta(\vec{r}, t)$ has to be introduced. The three fields can be formally defined as

$$\rho(\vec{r}, t) = \int\int \phi(\vec{r}, \vec{v}, \Theta, t) d\vec{v} d\Theta \tag{145}$$

$$\vec{u}(\vec{r}, t) = \frac{1}{\rho(\vec{r}, t)} \int\int \vec{v} \phi(\vec{r}, \vec{v}, \Theta, t) d\vec{v} d\Theta \tag{146}$$

$$\beta(\vec{r}, t) = \frac{1}{\rho(\vec{r}, t)} \int\int \Theta \phi(\vec{r}, \vec{v}, \Theta, t) d\vec{v} d\Theta, \tag{147}$$

where for the sake of simplicity we assumed that all dislocation loops have the same Burgers vector (the superscript (i) is dropped out) and their slip planes are parallel.

Taking the z axis perpendicular to the slip plane of the loops, the two nonvanishing components of the dislocation density tensor are

$$\alpha_{11} = b\rho \cos(\beta), \qquad \alpha_{21} = b\rho \sin(\beta). \tag{148}$$

Since the dislocation density tensor is the curl of the plastic distortion, it has to be div free:

$$\frac{\partial \alpha_{ij}}{\partial r_i} = 0. \tag{149}$$

One can find from Eqs(148,149) that ρ and β have to satisfy the conservation law

$$\frac{\partial \rho \cos(\beta)}{\partial x} + \frac{\partial \rho \sin(\beta)}{\partial y} = 0. \tag{150}$$

On the other hand, from the general expression of the evolution of the dislocation density tensor given by Eq. (39),the following evolution equations can be deduced for α and β fields(for details see [70])

$$\begin{aligned} \rho\dot{\beta} &= \cos(\beta)\frac{\partial \rho v}{\partial x} + \sin(\beta)\frac{\partial \rho v}{\partial y}, \\ \dot{\rho} &= \sin(\beta)\frac{\partial \rho v}{\partial x} - \cos(\beta)\frac{\partial \rho v}{\partial y}, \end{aligned} \tag{151}$$

where $v = |\vec{v}|$.

To have a closed theory a constitutive relation is needed between the three fields $\rho(\vec{r},t)$, $\beta(\vec{r},t)$ and $v(\vec{r},t)$. Kratochvíl and Sedláček [70] suggested the following constitutive relation for the velocity field :

$$Bv = \begin{cases} b\sigma_{13} + C\kappa_s - b\tau_0 - b\tau & if \quad b\sigma_{13} + C\kappa_s > b\tau_0 + b\tau \\ 0 & if \quad |b\sigma_{13} + C\kappa_s| < b\tau_0 + b\tau \\ b\sigma_{13} + C\kappa_s + b\tau_0 + b\tau & if \quad b\sigma_{13} + C\kappa_s < -b\tau_0 - b\tau \end{cases} \tag{152}$$

where $b\sigma_{13}$ is the Peach-Koehler force due to the local shear stress, $C\kappa_s$ is the self-force, $b\tau_0$ is the friction force, and $b\tau$ represents the interaction between the gliding dislocations an the dislocation loops.

The self-force $C\kappa_s$ is considered in the line tension approximation, where κ_s is the dislocation line tension. The curvature of a dislocation segment $C(\vec{r},\beta,t) = -\operatorname{div}\vec{n}$, where $\vec{n}$ is the unit normal to the dislocation segment. As it is explained in details in [70] C can be approximated by the expression

$$C = \cos(\beta)\frac{\partial \beta}{\partial x} + \sin(\beta)\frac{\partial \beta}{\partial y} \tag{153}$$

The most difficult problem is to set up an appropriate expression for $b\tau$. For this Kratochvíl and Sedláček [70] suggested that

$$b\tau = Fc^{1/3} \tag{154}$$

were c is the loop density, and F is a constant.

According to detailed analytical and numerical investigations [70] the model explained above is able to predict both dislocation patterning and size effect. Nevertheless, the justification of the assumptions used requires further investigation.

4 Internal stress distribution generated by the dislocations

In order to describe the statistical properties of a dislocation ensemble, a key quantity is the internal stress distribution. In this section the general form of the probability distribution of the internal shear stress $P(\tau)$ is determined for a system of parallel edge dislocations. As it is discussed later the results can be generalised to 3D systems.

4.1 General considerations

Like in the previous section, let us consider a system of N parallel straight edge dislocations positioned at the points $\vec{r}_i$, $i = \overline{1,N}$ in the xy plane perpendicular to the dislocation lines. For the sake of simplicity, we assume that each dislocation has the same Burgers vector $\vec{b}$. As it is shown later, the generalisation of the results for systems consisting of dislocations with different Burgers vectors is straightforward [71].

The internal shear stress at the point $\vec{r}$ is the sum of the stress fields of the individual dislocations

$$\tau(\vec{r}) = \sum_{i=1}^{N} \tau_{ind}(\vec{r} - \vec{r}_i). \tag{155}$$

In the first part of our analysis we assume that there is no dislocation at the point $\vec{r}$. Later on, the stress distribution at the dislocations is discussed, too.

The precise definition of the problem addressed in this section is to determine the $P(\tau_0)d\tau_0$ probability of occurrence of τ in the range

$$\tau_0 - \frac{d\tau_0}{2} \leq \tau(\vec{r}) \leq \tau_0 + \frac{d\tau_0}{2} \tag{156}$$

where τ_0 is a preassigned value for τ. $P(\tau_0)$ can be obtained as a direct application of Markoff's method [72] applied for several problems, like the problem of random flights, or for the determination of the distribution of forces in gravitationally interacting random systems. In contrast with the two problem mentioned, for dislocations the N particle distribution function cannot be built up from the one particle distribution functions since as it will be shown later it would lead to system size dependent internal stress distribution function $P(\tau_0)$. To avoid this, dislocation-dislocation correlation must be taken into account.

Denoting the N particle dislocation density function by $f_N(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N)$

the internal stress distribution can be expressed as

$$P(\tau_0)d\tau_0 = \int \dots \int f_N(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N)d\vec{r}_1 d\vec{r}_2 \dots d\vec{r}_N, \tag{157}$$

where the integration is effected only over that part of configuration space for which the inequalities (156) are satisfied. By the introduction of the factor

$$\Delta(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N) = \begin{cases} 1 & \text{whenever} \quad \tau_0 - \frac{d\tau_0}{2} \leq \tau \leq \tau_0 + \frac{d\tau_0}{2} \\ 0 & \text{otherwise.} \end{cases} \tag{158}$$

the integral in Eq. (157) can be extended over $2N$ dimensional space $\Re^{2N}$:

$$P(\tau_0)d\tau_0 = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \Delta(\vec{r}_1, \dots, \vec{r}_N) f_N(\vec{r}_1, \dots, \vec{r}_N)d\vec{r}_1 d \dots d\vec{r}_N. \tag{159}$$

For the determination on the structure of expression (159) one has to consider the integral

$$\delta = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\sin(\alpha q)}{q} \exp[i\gamma q]dq, \tag{160}$$

which is the well-known discontinuous integral of the Dirichlet function with the properties

$$\delta = \begin{cases} 1 & \text{whenever} \quad -\alpha < \gamma < \alpha \\ 0 & \text{otherwise.} \end{cases} \tag{161}$$

By taking

$$\alpha = \frac{d\tau_0}{2} \quad \text{and} \quad \gamma = \sum_{i=1}^{N} \tau_{ind}(\vec{r} - \vec{r}_i) - \tau_0 \tag{162}$$

from Eq. (158) one gets that $\Delta = \delta$. With the substitution of the (160) form of δ into Eq. (159), we obtain that

$$P(\tau_0)d\tau_0 = \frac{1}{\pi} \int_{\Re} dq \int_{\Re^{2N}} d\vec{r}_1 d\vec{r}_2 \dots d\vec{r}_N f_N(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N) \times \frac{\sin(\frac{1}{2}d\tau_0 q)}{q} \exp\left\{ i \left[\sum_{i=1}^{N} \tau_{ind}(\vec{r} - \vec{r}_i)q - \tau_0 q \right] \right\}. \tag{163}$$

It can be seen from the structure of the above expression that the Fourier transform of the internal stress distribution

$$A_N(\vec{r}, q) = \overline{P}(\tau_0) \tag{164}$$

has the form

$$A_N(\vec{r},q)=\int f_N(\vec{r}_1,\dots,\vec{r}_N)\prod_{j=1}^{N}\exp\left\{iq\tau(\vec{r}-\vec{r}_j)\right\}d^2\vec{r}_1d^2\dots d^2\vec{r}_N. \quad (165)$$

If we introduce the function

$$B(\vec{r},q)=1-\exp\{i\tau_{ind}(\vec{r})q\} \quad (166)$$

expression (165) can be rewritten into a power series of $B(\vec{r},q)$

$$\begin{aligned} A_N(\vec{r},q) &= \int d\vec{r}_1d\vec{r}_2\dots d\vec{r}_N f_N(\vec{r}_1,\vec{r}_2,\dots,\vec{r}_N)\prod_{j=1}^{N}[1-B(\vec{r}-\vec{r}_j,q)] \\ &= 1-\int\rho_1(\vec{r}_1)B(\vec{r}-\vec{r}_1,q)d^2\vec{r}_1 \\ &+ \frac{1}{2}\int\rho_2(\vec{r}_1,\vec{r}_2)B(\vec{r}-\vec{r}_1,q)B(\vec{r}-\vec{r}_2,q)d^2\vec{r}_1d^2\vec{r}_2+\dots \quad (167)\end{aligned}$$

where $\rho_k(\vec{r}_1,\vec{r}_2,\dots,\vec{r}_k)$ is the k-th order dislocation-density function defined by Eq. (66). Eq. (167) can be transformed into an exponential form

$$A_N(\vec{r},q) = \exp\{E(q,\vec{r})\} \quad (168)$$

where

$$\begin{aligned} E(q,\vec{r}) &= -\int\rho_1(\vec{r}_1)B(\vec{r}-\vec{r}_1,q)d^2\vec{r}_1 \\ &+ \frac{1}{2}\int D_2(\vec{r}_1,\vec{r}_2)B(\vec{r}-\vec{r}_1,q)B(\vec{r}-\vec{r}_2,q)d^2\vec{r}_1d^2\vec{r}_2+\dots \quad (169)\end{aligned}$$

in which

$$\begin{aligned} D_2(\vec{r}_1,\vec{r}_2) &= \rho_2(\vec{r}_1,\vec{r}_2)-\rho_1(\vec{r}_1)\rho_1(\vec{r}_2) \\ &= \rho_1(\vec{r}_1)\rho_1(\vec{r}_2)(1+d(\vec{r}_1,\vec{r}_2)) \quad (170)\end{aligned}$$

is the dislocation-dislocation correlation function.

Eq. (169) is straightforward to generalise for the case where dislocations

with Burgers vectors $\pm\vec{b}$ are allowed. One can find that

$$\begin{aligned} E(q,\vec{r}) &= -\int \left[\rho_+(\vec{r}_1)B(\vec{r}-\vec{r}_1,q)+\rho_-(\vec{r}_1)B^*(\vec{r}-\vec{r}_1,q)\right] d^2\vec{r}_1 \\ &+ \frac{1}{2}\int D_{++}(\vec{r}_1,\vec{r}_2)B(\vec{r}-\vec{r}_1,q)B(\vec{r}-\vec{r}_2,q)d^2\vec{r}_1 d^2\vec{r}_2 \\ &+ \frac{1}{2}\int D_{+-}(\vec{r}_1,\vec{r}_2)B(\vec{r}-\vec{r}_1,q)B^*(\vec{r}-\vec{r}_2,q)d^2\vec{r}_1 d^2\vec{r}_2 \\ &+ \frac{1}{2}\int D_{--}(\vec{r}_1,\vec{r}_2)B^*(\vec{r}-\vec{r}_1,q)B^*(\vec{r}-\vec{r}_2,q)d^2\vec{r}_1 d^2\vec{r}_2 \\ &+ \frac{1}{2}\int D_{-+}(\vec{r}_1,\vec{r}_2)B^*(\vec{r}-\vec{r}_1,q)B(\vec{r}-\vec{r}_2,q)d^2\vec{r}_1 d^2\vec{r}_2 + \ldots \end{aligned} \tag{171}$$

where "*" denotes complex conjugate.

4.2 Stress distribution at the dislocations

In the previous subsection we assumed that there is no dislocation at point $\vec{r}$ at which the probability of the internal stress is determined. One can also ask the probability distribution of the shear stress at the dislocations $P^c(\tau)$.

Let assume that the dislocation at which we ask the probability distribution is at the point $\vec{r}_1$. The stress at the dislocation is obviously

$$\tau(\vec{r}_1) = \sum_{i=2}^{N} \tau_{ind}(\vec{r}_1 - \vec{r}_i). \tag{172}$$

The derivation procedure explained above can be repeated for this case, but $f_N(\vec{r}_1,\vec{r}_2,\ldots,\vec{r}_N)$ needs to be replaced with the conditional probability density $f^c(\vec{r}_2,\ldots,\vec{r}_N|\vec{r}_1)$.
$f^c(\vec{r}_2,\ldots,\vec{r}_N|\vec{r}_1)d\vec{r}_2 d\vec{r}_3..d\vec{r}_N$ is the the probability of finding the 2nd, 3th, .. Nth dislocation in the $d\vec{r}_2 d\vec{r}_3..d\vec{r}_N$ vicinity of points $\vec{r}_2,\vec{r}_3..\vec{r}_N$ assuming there is a dislocation at point $\vec{r}_1$. One gets that the Fourier transform of $P^c(\tau)$ reads as

$$\begin{aligned} A^c(\vec{r},q) &= \int d^2\vec{r}_2 d^2\vec{r}_3 \ldots d^2\vec{r}_N f^c(\vec{r}_2,\ldots,\vec{r}_N|\vec{r}_1) \\ &\prod_{j=2}^{N}[1-B(\vec{r}_1-\vec{r}_j,q)] \end{aligned} \tag{173}$$

The conditional probability $f^c(\vec{r}_2,\ldots,\vec{r}_N|\vec{r}_1)$ can be given as

$$f^c(\vec{r}_2,\ldots,\vec{r}_N|\vec{r}_1) = \frac{f_N(\vec{r}_1,\vec{r}_2,\ldots,\vec{r}_N)}{f_1(\vec{r}_1)}. \tag{174}$$

Substituting this into Eq. (173) and rewriting it into a power series of $B(\vec{r},q)$ we obtain

$$A^c(\vec{r}_1,q) = 1-\int\left[\frac{\rho_{++}(\vec{r}_1,\vec{r}_2)}{\rho_+(\vec{r}_1)}B(\vec{r}_1-\vec{r}_2,q) + \frac{\rho_{+-}(\vec{r}_1,\vec{r}_2)}{\rho_+(\vec{r}_1)}B^*(\vec{r}_1-\vec{r}_2,q)\right]d^2\vec{r}_2+... \quad (175)$$

where we assumed that the dislocation at point $\vec{r}_1$ has positive Burgers vector. (A similar expression holds for dislocations with negative Burgers vector.)

4.3 The mean values of distributions $P(\tau)$ and $P^c(\tau)$

Since dislocations form strongly inhomogeneous distributions the explicit form of $P(\tau)$ and $P^c(\tau)$ cannot be determined analytically. Nevertheless, analytical results can be obtained for some of their properties [71, 73].

An important characteristic value of a distribution function $P(x)$ is its mean value $< x >= \int xP(x)dx$. As it is well known the mean value can be determined from the $A(q)$ Fourier transform of the distribution function by the relation

$$< x > = \frac{i}{A(0)}\left.\frac{dA(q)}{dq}\right|_{q=0} \quad (176)$$

Applying the above expression for Eqs. (172,175) we find that the mean value of the internal stress at a point where there is no dislocations is

$$< \tau(\vec{r}) > = -i\int\left[\rho_+(\vec{r})\left.\frac{dB}{dq}\right|_{q=0}+\rho_-(\vec{r})\left.\frac{dB^*}{dq}\right|_{q=0}\right]d\vec{r}_1 = \int\kappa(\vec{r})\tau_{ind}(\vec{r}-\vec{r}_1)d\vec{r}_1. \quad (177)$$

If we compare this with expression (91) we can see that $< \tau(\vec{r}) >$ is the self-consistent field introduced earlier.

On the other hand, from Eq. (175) the mean value of shear stress at a dislocation is

$$< \tau_+(\vec{r}_1) > = -i\int\left[\frac{\rho_{++}(\vec{r}_1,\vec{r}_2)}{\rho_+(\vec{r}_1)}\left.\frac{dB}{dq}\right|_{q=0}+\frac{\rho_{+-}(\vec{r}_1,\vec{r}_2)}{\rho_+(\vec{r}_1)}\left.\frac{dB^*}{dq}\right|_{q=0}\right]d\vec{r}_2 = \frac{1}{\rho(\vec{r}_1)}\int\left[\rho_{++}(\vec{r}_1,\vec{r}_2)-\rho_{+-}(\vec{r}_1,\vec{r}_2)\right]\tau_{ind}(\vec{r}-\vec{r}_1)d\vec{r}_2 \quad (178)$$

(The subscript "+" in $< \tau_+(\vec{r}_1) >$ indicates that it corresponds to a dislocation with positive sign.)

If we compare the above expression with the third terms in Eqs. (79,80) we can see that Eqs. (79,80) can be rewritten as

$$\frac{\partial \rho_+(\vec{r}_1,t)}{\partial t} + \vec{b}\frac{\partial}{\partial \vec{r}_1}\left\{\rho_+(\vec{r}_1,t)[\tau_{ext} + < \tau_+(\vec{r}_1) >]\right\} = 0 \tag{179}$$

$$\frac{\partial \rho_-(\vec{r}_1,t)}{\partial t} - \vec{b}\frac{\partial}{\partial \vec{r}_1}\left\{\rho_-(\vec{r}_1,t)[\tau_{ext} + < \tau_-(\vec{r}_1) >]\right\} = 0 \tag{180}$$

giving a new physical interpretation of the form of the dislocation evolution equations obtained.

4.4 Asymptotic properties of the stress distribution function

According to the investigations of Groma and Bakó [71] for small enough Fourier parameter q the Fourier transform of $P(\tau)$ reads as (for the derivation see [71]):

$$A_N(\vec{r},q) = \exp\left\{i < \tau(\vec{r}) > q + C_d\rho(\vec{r})q^2 \ln\frac{q}{R_{eff}} + \ldots\right\}, \tag{181}$$

where R_{eff} is an effective correlation length determined by the correlation functions d_{++}, d_{--}, d_{+-}, d_{-+}, and

$$C_d = \frac{b^2}{2}\int_0^{2\pi} r^2\tau_{ind}^2(r,\varphi)d\varphi. \tag{182}$$

For edge dislocations in isotropic medium

$$C_d = \frac{(\mu b)^2}{8\pi^2(1-\nu)^2}\int_0^{2\pi}\cos^2(\varphi)\cos^2(2\varphi)d\varphi = \frac{(\mu b)^2}{16\pi(1-\nu)^2}. \tag{183}$$

It has to be mentioned that if the dislocation system consists of more than one type of dislocations, Eq. (181) remains valid. The only difference is that in this case R_{eff} and C_d are the appropriate weighted average of the R_{eff}^i and C_d^i values corresponding to the dislocations with Burgers vector $\vec{b}^i$.

One can obtain (see [71]) that Eq. (181) follows that the tail of probability distribution function decays as

$$P(\tau) = C_d\rho(\vec{r})\frac{1}{\tau^3}. \tag{184}$$

A remarkable feature of the above expression is that it is independent from the dislocation- dislocation correlations. In other words it does not depend on the microscopic arrangement of the dislocations. Certainly the stress value from which the distribution function follows the above asymptotic expression depends on dislocation-dislocation correlation. To demonstrate this, the internal stress distribution was numerically determined on a system of 1000 randomly distributed dislocation dipoles (seen Figure 19). As it can

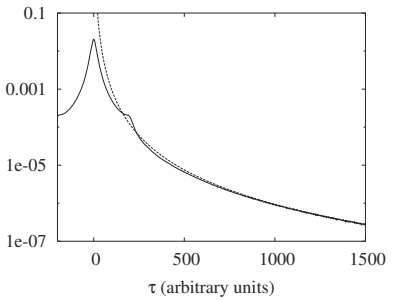


Figure 19. $P(\tau)$ obtained numerically on a randomly distributed dipole system (full line) with the asymptotic curve given by Eq. (184) (dotted line).

be seen the asymptotic curve describes extremely well the tail part of the internal stress distribution obtained numerically. It has to be mentioned that the asymptotic curve is not fitted, C_d is determined according to Eq. (183).

Recent investigations revealed [73] that for a dipole system the centre part of the distribution is Lorentzian:

$$P(\tau) = \frac{D_d \rho}{\pi} \frac{1}{\tau^2 + (D\rho)^2} \tag{185}$$

where the actual value of D can only be determined numerically. For edge dislocation dipoles with 45^0 between the polarisation and the Burgers vectors

$$D_d = \frac{b\mu}{2\pi(1-\nu)} 3.55|\vec{p}| \tag{186}$$

As it can be seen in Figure 20 the centre part of $P(\tau)$ is well described by the Lorentzian given by Eq. (185).

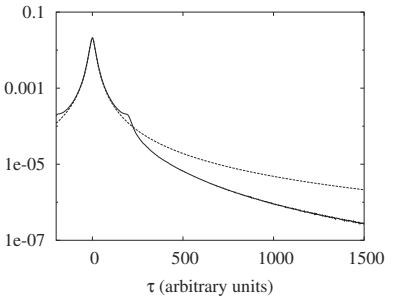


Figure 20. $P(\tau)$ obtained numerically on a randomly distributed dipole system (full line) with the Lorentzian given by Eq. (185) (dotted line).

Acknowledgement

During the past years the author had the pleasure to work with Botond Bakó, Ferenc Csikor, Erik van der Giessen, Géza Györgyi, Péter Dusán Ispánovity, Ferenc Székely, Serge Yefimov and Michael Zaiser. Most of the results presented in the paper were obtained together with them. Their outstanding scientific contributions and friendships are gratefully acknowledged.

Financial supports of the Hungarian Scientific Research Fund (OTKA) under contract number K 67778 and the European Community's Human Potential Programme under contracts HPRN-CT-2002-00198 [DEFINO] and MRTN-CT-2003-504634 [SizeDepEn] are also acknowledged.

Bibliography

[1] I. Kovács and L. Zsoldos, Dislocations and Plastic Deformation, Pergamon Press, London (1973).

[2] Basinski S.J. and Basinski Z.S., Dislocations in Solids, ed. by F.R.N. Nabarro, North-Holland, Amsterdam, 1979

[3] Hirsh JP. Lothe J. Theory of dislocations, John Wiley and Sons, 1982

[4] Mughrabi H., Ungár T., Kienle W., M. Wilkens, Phil. Mag. 53, 793 (1986)

[5] Mughrabi H., Ackerman F., Herz K., in J. Fang (editor), Fatigue Mechanisms, Proc. of ASTM-NBS-NSF Symposium, ASTM-STP 675, 1979, p. 69.

[6] Kuhlmann-Wilsdorf D., van der Merwe J., Mat. Sci. Eng. 35, 79 (1982)

[7] Bassim M., Jesser W., Kuhlmann-Wilsdorf D., Wilsdorf H., Mat. Sci. Eng. 81, 122 (1986)

[8] Kuhlmann-Wilsdorf D., Mat. Sci. Eng. A 113, 1 (1989)

[9] Holt D., Journal Appl. Phys. 41, 3179 (1970)

[10] Walgraef D., Aifantis E., J. Appl. Phys. 15, 688 (1985)

[11] Aifantis E., Mat. Sci. Eng. 81, 563 (1986)

[12] Walgraef D., Solid State Phenomena, 3-4, 77 (1988)

[13] Kratochvil J., Scripta Metall. Mater. 24, 891 (1990)

[14] Kratochvil J., Scripta Metall. Mater. 24, 1225 (1990)

[15] Kratochvil J., and Saxlova M., Scripta Met. Mater. 26, 113 (1993)

[16] Hähner P., Bay K., Zaiser M., Phys. Rev. Lett. 81, 2470 (1998)

[17] Zaiser M., Bay K., Hähner P., Acta mater. 47, 2463 (1999)

[18] Székely F., Groma I., Lendvai J., Mat. Sci. Eng. A 309, 352 (2001)

[19] Székely F., Groma I., Lendvai J., Scripta Mater 45, 55 (2001)

[20] Fleck N.A., Mller G.M., Ashby M.F. and Hutchinson J.W., Acta Mater. 42, 475 (1994)

[21] McElhaney K.W., Vlassah JJ. and Nix W.D., J. Math. Res. 13, 1300 (1998)

[22] Aifantis E.C., J. Eng. Mater. Technol. 106, 326 (1984)

[23] Aifantis E.C., Int. J. Plast. 2, 211 (1987)

[24] Aifantis E.C., J. Mech. Behav. Mater. 5, 335 (1994)

[25] Fleck N.A. and Hutchinson J.W., J. Mech. Phys. Solids, 49, 2245 (2001)

[26] Gurtin M.E., J. Mech. Phys. Solids, 50, 5 (2002)

[27] Svendsen B., J. Mech. Phys. Solids, 50, 1297 (2002)

[28] Nicola L., van der Giessen E. and Needleman A., Journal of Applied Physics 93, 5920 (2003)

[29] Kröner E., Z. Phys. 151, 504 (1958)

[30] E. Kröner, Continuum Theory of Defects, Physics of Defects, Eds. R. Balian et.all, North Holland, Amsterdam (1981).

[31] L.D. Landau, E.M. Lifsitz, Theory of Elasticity, Course of Theoretical Physics, vol 7, Pergamon Press, Oxford (1986)

[32] Kosevich A.M. in Dislocations in Solids, ed. by F.R.N. Nabarro, North-Holland, Amsterdam, 1979

[33] Nye J.F. Acta. Mat. 1, 153 (1953)

[34] Burgers J.M. Proc. Kon. Nederl. Akad. Wetensch. 42, 293 and 378 (1939)

[35] Devincre B., and Condat M., Acta Metall. Mater. 40, 2629 (1992)

[36] Fivel M., Gosling T.,and Canova , Modell. Simul. Mater. Sci. Eng. 4, 581 (1996)

[37] Huang J.M., and Ghoniem N.M., Model. Simul. Mater. Sci. 11, 21 (2003)

[38] Singh B.N., Ghoniem N.M., and Trinkaus H. J. Nucl. Mater. 307, 159 (2002)

[39] Huang H.M., and Ghoniem N.M. Comp. Mater. Sci. 23, 225 (2002)

[40] Walgraef D., and Ghoniem N.M. Phys. Rev. B 65, art. no. 155304 (2002)

[41] Madec R., Devincre B., and Kubin L.P., Phys. Rev. Lett. 89, 255508 (2002)

[42] Madec R., Devincre B., and Kubin L.P. Scripta Mater 47, 689 (2002)

[43] Lemarchand C., Chaboche J.L., and Devincre B,. J. Phys IV 9, 271 (1999)

[44] Bulatov V., Abraham F.F., Kubin L., Devincre B., and Yip S., Nature 391, 669 (1998)

[45] Cai W., Bulatov V.V., Chang J., Li J. and Yip S., Phil. Mag. A, 81, 1257 (2001).

[46] Lepinoux J., and Kubin L., Scripta Met. Mat. 21, 833 (1987)

[47] Gulluoglu A., and Hartly C., Modell. Simul. Mater. Sci. Eng. 1, 383 (1993)

[48] Gullouglu A., Srolovity D., LeSar R., and Lomdahl P., Scripta Met. 23, 1347 (1989)

[49] Ghoniem N., and Amodeo R., Phys. Rev. B 41, 6968 (1989)

[50] Groma I. and Pawley G.S., Philos. Mag. A 67 1459 (1993)

[51] Groma I. and Pawley G.S., Mat. Sci. Eng. A 164, 306 (1993)

[52] Van der Giessen E., and Needleman A., Model. Simul. Mater. Sci. Eng. 3, 689 (1995)

[53] Cleveringa H. H. M., V. der Giessen E.,and Needleman A., Acta mater. 45, 3163 (1997)

[54] Cleveringa, H.H.M., Van der Giessen, E., and Needleman, Acta Mater. 45, 3163 (1997)

[55] Cleveringa, H.H.M., Van der Giessen, E. and Needleman, A., International Journal of Plasticity 15, 837 (1999)

[56] Cleveringa, H.H.M., Van der Giessen, E. and Needleman, A., Phil. Mag. A79, 863 (1999)

[57] A. Benzerga, Y. Brechet, A. Needleman and E. Van der Giessen Model. Simul. Mat. Sci. Eng. 12, 159 (2004)

[58] Groma I., Phys. Rev. B 56, 5807 (1997)

[59] Groma I. and Balogh P., Acta Mater. 47, 3647 (1999)

[60] Buchinger L. and Stanzl S., Phil. Mag. A50, 275 (1984)

[61] Zaiser M., Miguel M.C., Groma I., Phys Rev B 64, 224102 (2001)

[62] Groma I., Csikor FF., Zaiser M., Acta Mater. 51, 1271 (2003)

[63] Cleveringa, H.H.M., Van der Giessen, E. and Needleman, A., J. Physique IV, 83, 92 (1998)

[64] Yefimov S., Groma I. and van der Giessen E. J. de Physique IV 11, 103 (2001)

[65] S. Yefimov, I. Groma and E. van der Giessen J. Mech. and Phys. Solids 52, 279 (2004)

[66] Yefimov S. and van der Giessen E European Journal of Mechanics A/Solids 24, 183 (2005)

[67] Yefimov S. and van der Giessen E. Int. J. of Solids Struct. 42, 3375 (2005)

[68] El-Azab A., Phys. Rev. B, 61, 11956, (2000)

[69] Hochrainer T. and Zaiser M., private communication

[70] Kratochvíl J. and Sedláček R., Phys. Rev. B, 67, 094105 (2003)

[71] Groma I. and Bak B., Phys. Rev. B 58, 2969 (1998)

[72] Chandrasekhar S., Rev. Modern Phys. 15, 8 (1943)

[73] Csikor F.F. Groma I. Phys. Rev. B 70, 064106 (2004)

[74] Groma I.,Phys. Rev. B 57, 7535 (1998)

Basic ingredients, development of phenomenological models and practical use of crystal plasticity

Georges Cailletaud
MINES ParisTech
Centre des Matériaux, CNRS UMR 7633

Abstract This part is an introduction to phenomenological crystal plasticity models, to their properties and their use either in component calculations or in microstructure modelling. Time dependent or time independent versions of the models are presented. The applications deal with single crystal or polycrystalline materials. In this last case, the aggregate is modelled either by a finite element mesh, in order to represent the local microstructure, or by using so called uniform fields models, which allow to introduce the average effect of each crystalline phase in the global behaviour.

1 Introduction

The theoretical framework needed to develop single crystal models has been built in the seventies (Mandel, 1971; Hill, 1966; Hill and Rice, 1972; Rice, 1970, 1971; Asaro and Rice, 1977), and the first practical applications a few years later (Asaro, 1983a,b). The history of crystal plasticity is then well known, and can be reviewed in classical books (Havner, 1992; Teodosiu, 1997). The purpose of this section is then not to add a new description of rather classical approaches, but to enter the topic according to an alternative route, namely the framework of thermodynamics of irreversible processes, and to present a mechanical approach, seen as a multipotential theory, and to illustrate its capabilities by a series of examples.

The interest of this thermodynamical approach is to provide natural schemes for the choice of the evolution equations and the type of the hardening variables. The deformation mechanisms are restricted to slip, on given planes and along given directions. This class of model can be applied to single crystal or polycrystal description. Single crystal slip is illustrated in Fig.1, which shows the development of plasticity on a Ni-base superalloy specimen, after 1, 2 and 11 cycles: a first plane is activated after the initial

tension; more and more slip activity can be observed after 2 and 11 cycles. The crystal was oriented for single slip, so that only one line can be seen. Multiple slip can also be found, depending on the orientation of the specimen. In polycrystals, various grain orientations are present, so that single and multiple slips are both present in the deformation process. Figure 2 shows the case of a FCC polycrystal (Fig.2a) and the case of a HCP polycrystal (Fig.2b). For the second one, the number of slip planes is reduced, if compared to FCC material, so that other mechanisms may become active, like twinning, or intergranular damage.

Volume element as well as finite element simulations will be shown. A series of models have been built in the last twenty years, coming from two communities of research: the literature shows either "purely phenomenological" models, or "dislocation based" models. Due to the reduced space, it has been decided to promote the "operational" aspect of the models, so that the presentation will be focused on the phenomenological models. Nevertheless, a link will be made between the two approaches. The outline of the part is as follows:

- to be complete, a short introduction to thermodynamical approach is first given, and followed by the presentation of a class of crystal plasticity models;
- some applications of these models in finite element computations are then shown: this will concern first material identification and finite element computations of components or specimens, then some "microstructure computations", where single crystal models are used to represent individual grains in polycrystals;
- the last part is devoted to the applications of these models in uniform fields approaches, that may offer sometimes an interesting compromise between the macroscopic models and the full microstructure computations.

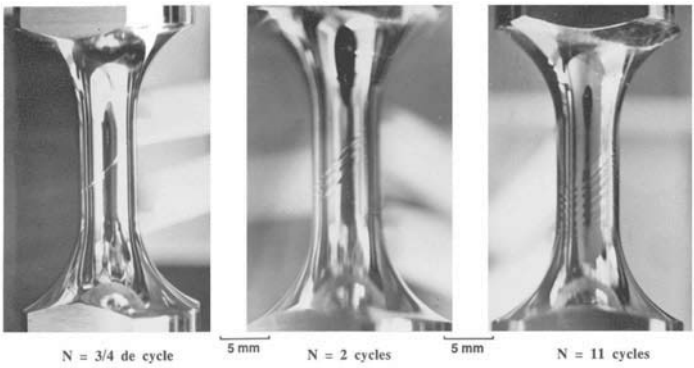
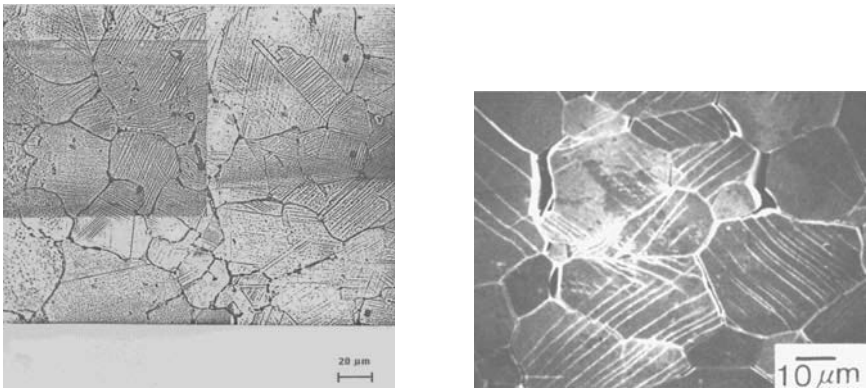


Figure 1. Slip system activity on a single crystalline specimen (Hanriot, 1993)



(a) (b)

Figure 2. Slip system activity in polycrystalline specimens: (a) waspaloy (Clavel, 1980), (b) zircaloy, with intergranular damage (Kubo et al., 1985)

2 A thermodynamical approach to single crystal plasticity

The thermodynamical framework Coleman (1964); Halphen and Nguyen (1975); Germain et al. (1983) is recalled for the general case of inelastic constitutive equations. The presentation is restricted to the case of small strains, since this is the typical deformation range used in the present document.

2.1 General framework

First and second principle. The first principle is the expression of the conservation of the energy. On a domain $\mathcal{D}$, the internal energy E, that is obtained by integrating the specific internal energy, e, is the sum of the power of the external forces and of the heat exchanges:

$$\frac{dE}{dt} = \int_{\mathcal{D}} \rho \frac{de}{dt} dV = \mathcal{P}^{(e)} + \dot{Q} \tag{1}$$

with the following definitions:

- Power of external forces:

$$\mathcal{P}^{(e)} = \int_{\mathcal{D}} \underset{\sim}{\boldsymbol{\sigma}} : \dot{\underset{\sim}{\boldsymbol{\varepsilon}}} dV \tag{2}$$

- Heat exchanged, using the rate of captured heat, $\underline{\boldsymbol{q}}$, and $\underline{\boldsymbol{n}}$, outside normal to the surface $\partial\mathcal{D}$, and r, volumetric heat:

$$\dot{Q} = \int_{\mathcal{D}} r dV - \int_{\partial\mathcal{D}} \underline{\boldsymbol{q}}.\underline{\boldsymbol{n}} dS = \int_{\mathcal{D}} (r - div\underline{\boldsymbol{q}})\, dV \tag{3}$$

This provides the following expression for the first principle:

$$\rho \frac{de}{dt} = \underset{\sim}{\boldsymbol{\sigma}} : \dot{\underset{\sim}{\boldsymbol{\varepsilon}}} + r - div\underline{\boldsymbol{q}} \tag{4}$$

The second principle provides an upper bound of the heat rate received by the volume $\mathcal{D}$ at a temperature T, and can be expressed as a function of the entropy S or of the specific entropy s :

$$\frac{dS}{dt} \geqslant \int_{\mathcal{D}} \frac{r}{T} dV - \int_{\partial\mathcal{D}} \frac{\underline{\boldsymbol{q}}.\underline{\boldsymbol{n}}}{T} dS \tag{5}$$

$$\text{then:} \qquad \int_{\mathcal{D}} \left(\rho \frac{ds}{dt} - \frac{r}{T} + div\left(\frac{\underline{\boldsymbol{q}}}{T}\right) \right) dV \geqslant 0 \tag{6}$$

Using Helmoltz free energy Ψ, such as $e = \Psi + Ts$, one get the so called *Clausius-Duhem* inequality:

$$\underset{\sim}{\boldsymbol{\sigma}} : \dot{\underset{\sim}{\boldsymbol{\varepsilon}}} - \rho\frac{d\Psi}{dt} - \rho s\dot{T} - \frac{1}{T}\underline{\boldsymbol{q}}.\underline{\boldsymbol{grad}}(T) \geqslant 0 \tag{7}$$

The *method of local state* assumes that the whole history of the material can be represented by the temperature T and by a set of *state variables* α_i, that are involved in Ψ, and which will be used in the classical Coleman–Noll argument. One successively get:

$$\frac{d\Psi}{dt} = \frac{\partial\Psi}{\partial T}\dot{T} + \frac{\partial\Psi}{\partial\alpha_i}\dot{\alpha}_i \tag{8}$$

$$s = -\frac{\partial\Psi}{\partial T} \tag{9}$$

$$\sigma_{ij}\dot{\varepsilon}_{ij} - \rho\,\frac{\partial\Psi}{\partial\alpha_i}\dot{\alpha}_i - \frac{1}{T}\,\underline{\boldsymbol{q}}\,.\,\underline{\boldsymbol{grad}}(T) \geqslant 0 \qquad (4) \tag{10}$$

The intrinsic dissipation Φ_1 and the thermal dissipation Φ_2 are considered:

$$\Phi_1 = \sigma_{ij}\dot{\varepsilon}_{ij} - \rho\frac{\partial\Psi}{\partial\alpha_i}\dot{\alpha}_i \qquad \Phi_2 = -\frac{1}{T}\,\underline{\boldsymbol{q}}\,.\,\underline{\boldsymbol{grad}}(T) \tag{11}$$

One assumes then that both of them must be positive. This is directly given by the Fourier's law, which provides the heat equation in presence of mechanical strain

$$\underline{\boldsymbol{q}} = -k(T,\alpha_i)\,\underline{\boldsymbol{grad}}(T) \tag{12}$$

$$div\left(k\,\underline{\boldsymbol{grad}}(T)\right) = \rho C_\varepsilon\dot{T} - r - \sigma_{ij}\dot{\varepsilon}_{ij} + \rho\left(\frac{\partial\Psi}{\partial\alpha_i} - T\frac{\partial^2\Psi}{\partial T\partial\alpha_i}\right)\dot{\alpha}_i \qquad (5) \tag{13}$$

(with $C_\varepsilon = T\partial s/\partial T$, specific heat at a constant strain) The shape of the mechanical part may take various forms, according to the type of material.

Thermoelasticity. The only internal variable is the elastic strain. Since elasticity is a non dissipative process, Φ_1 remains equal to zero. This provides a definition for the stress tensor:

$$\Phi_1 = \underset{\sim}{\boldsymbol{\sigma}} : \dot{\underset{\sim}{\boldsymbol{\varepsilon}}}^e - \rho\frac{\partial\Psi}{\partial\underset{\sim}{\boldsymbol{\varepsilon}}^e} : \dot{\underset{\sim}{\boldsymbol{\varepsilon}}}^e = 0 \tag{14}$$

These constitutive equations involve then two state variables, T and $\underset{\sim}{\boldsymbol{\varepsilon}}^e$, and two conjugated variables, the entropy s and the stress tensor $\underset{\sim}{\boldsymbol{\sigma}}$. Ψ can be seen as a *thermodynamic potential* which characterises reversible processes.

Dissipative processes. A series of hardening variables is now related to the list of the state variables α_I, so that the model is formulated according to the following scheme:

State variables		Conjugated variables	
temperature	T	$s = -\dfrac{\partial \Psi}{\partial T}$	entropy
elastic strain	$\underset{\sim}{\varepsilon}^e$	$\underset{\sim}{\sigma} = \rho \dfrac{\partial \Psi}{\partial \underset{\sim}{\varepsilon}^e}$	stress
state variables	α_I	$A_I = \rho \dfrac{\partial \Psi}{\partial \alpha_I}$	hardening variables

The intrinsic dissipation can be rewritten:

$$\Phi_1 = \underset{\sim}{\sigma} : \dot{\underset{\sim}{\varepsilon}}^p - A_I \dot{\alpha}_I = \mathrm{Z}\,\dot{\mathrm{z}} \tag{15}$$

with : $\mathrm{Z} = \{\underset{\sim}{\sigma}, A_I\}$; $\mathrm{z} = \left\{\underset{\sim}{\varepsilon}^p, -\alpha_I\right\}$

A model is *standard* (Halphen and Nguyen, 1975) if one can find a potential $\Omega \equiv \Omega(\mathrm{Z})$ such as:

$$\dot{\mathrm{z}} = \frac{\partial \Omega}{\partial \mathrm{Z}} \tag{16}$$

If Ω is a convex function of Z which includes the origin, the dissipation is then automatically positive, since:

$$\phi_1 = \mathrm{Z}\,\frac{\partial \Omega}{\partial \mathrm{Z}} \tag{17}$$

One can also define (through the Legendre-Fenchel transform) a companion potential in terms of $\dot{\mathrm{z}}$:

$$\Omega^*(\dot{\mathrm{z}}) = \max_{\mathrm{Z}} \left(\mathrm{Z}\,\dot{\mathrm{z}} - \Omega(\mathrm{Z})\right) \tag{18}$$

As a consequence, either $\Omega^*(\dot{\mathrm{z}})$ or $\Omega(\mathrm{Z})$ can be introduced to characterize the dissipative processes. A dissipative model can then be fully characterised by the definition of two potentials Ψ and Ω. The relation between state variables and hardening variables will then be derived from Ψ, and the nature of the hardening variables and their evolution rules from Ω. As an example, isotropic and nonlinear kinematic hardenings come from the following choice for the sets $(A_I,\ \alpha_I)$ is:

Type of hardening	State variable	Conjugated variable
Isotropic hardening	r	R
Kinematic hardening	$\underset{\sim}{\boldsymbol{\alpha}}$	$\underset{\sim}{\boldsymbol{X}}$

$$\dot{\underset{\sim}{\varepsilon}}^p = \frac{\partial \Omega}{\partial \underset{\sim}{\sigma}} \qquad \dot{r} = -\frac{\partial \Omega}{\partial R} \qquad \dot{\underset{\sim}{\alpha}} = -\frac{\partial \Omega}{\partial \underset{\sim}{X}} \tag{19}$$

Assuming that J is a von Mises norm, such as $J(\underset{\sim}{x}) = ((3/2)\underset{\sim}{x} : \underset{\sim}{x})^{1/2}$, a viscoplastic model is simply obtained by means of a potential Ω, according to the following steps:

- Definition of a viscoplastic yield:

$$f(\underset{\sim}{\sigma}, \underset{\sim}{X}, R) = J(\underset{\sim}{\sigma} - \underset{\sim}{X}) - R \tag{20}$$

- Definition of a potential, function of f:

$$\Omega = \frac{K}{n+1} \left(\frac{f}{K}\right)^{n+1} \tag{21}$$

- Viscoplastic flow:

$$\dot{\underset{\sim}{\varepsilon}}^p = \frac{\partial \Omega}{\partial \underset{\sim}{\sigma}} = \frac{\partial \Omega}{\partial f} \frac{\partial f}{\partial \underset{\sim}{\sigma}} \tag{22}$$

- Intensity of the flow (here, Norton with internal stress):

$$\dot{v} = \frac{\partial \Omega}{\partial f} = \left(\frac{f}{K}\right)^n \tag{23}$$

- Flow direction (normal to the yield surface):

$$\underset{\sim}{n} = \frac{\partial f}{\partial \underset{\sim}{\sigma}} \tag{24}$$

- Hardening:

$$\dot{r} = -\frac{\partial \Omega}{\partial R} \qquad \dot{\underset{\sim}{\alpha}} = -\frac{\partial \Omega}{\partial \underset{\sim}{X}} \tag{25}$$

One can switch from a viscoplastic to a time independent plastic model by replacing the viscoplastic potential by a plastic pseudo-potential, as shown in Fig.3, where Ind(f) jumps from 0 (for $f < 0$) to infinity (for $f = 0$). The strain rate is then no longer given by the partial derivative of Ω with respect to f, but by the consistency condition. The direction of the flow is recovered by introduction of *Hill's principle*, which assumes that the real stress field provides a maximum of the intrinsic dissipated power $\Phi_1 = \underset{\sim}{\sigma}^* : \dot{\underset{\sim}{\varepsilon}}^p$ of any admissible stress field, when there is no hardening:

$$\forall \underset{\sim}{\sigma}^* \text{ admissible } (\underset{\sim}{\sigma} - \underset{\sim}{\sigma}^*)\dot{\underset{\sim}{\varepsilon}}^p \geqslant 0 \tag{26}$$

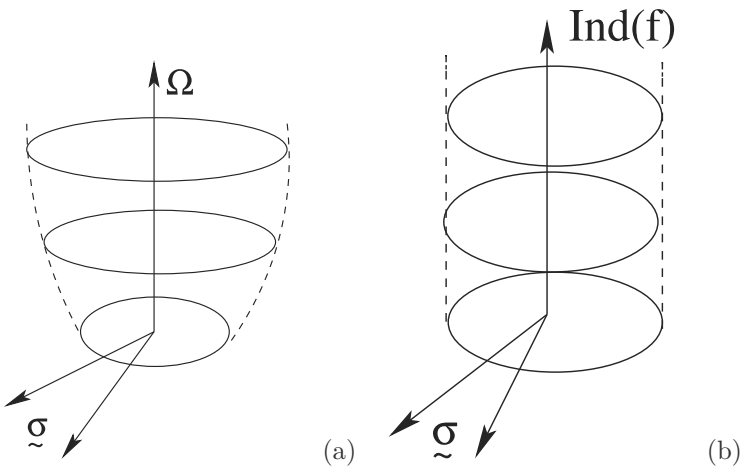


Figure 3. Illustration of (a) the viscoplastic potential, (b) the plastic pseudo-potential

For the case of a material with hardening, the preceding equation has to be reformulated as:

$$\Phi_1 = \underset{\sim}{\boldsymbol{\sigma}}^* : \dot{\underset{\sim}{\varepsilon}}^p - A_I^* \dot{\alpha}_I = \underset{\sim}{\boldsymbol{\sigma}} : \dot{\underset{\sim}{\varepsilon}}^p - \dot{\Psi}_p = \mathrm{Z}\ \dot{z} \qquad \text{maximum} \tag{27}$$

where Z includes stress and the hardening variables A_I, and z includes plastic strain and the state variables ($-\alpha_I$):

$$(\mathrm{Z} - \mathrm{Z}^*)\, \dot{z} \geqslant 0 \tag{28}$$

The maximization of Φ_1 *under the constraint* $f \leqslant 0$ can be seen as an extension of Hill's principle. Let us define $\mathbb{F}(\mathrm{Z}) = \mathrm{Z}\dot{z} - \dot{\lambda} f$ and search for the zero of $\partial\mathbb{F}/\partial\mathrm{Z}$

$$\dot{z} = \dot{\lambda} \frac{\partial f}{\partial \mathrm{Z}} \qquad \text{then:} \quad \dot{\underset{\sim}{\varepsilon}}^p = \dot{\lambda} \frac{\partial f}{\partial \underset{\sim}{\boldsymbol{\sigma}}} = \dot{\lambda}\, \underset{\sim}{\boldsymbol{n}} \qquad \dot{\alpha}_I = -\dot{\lambda} \frac{\partial f}{\partial A_I} \tag{29}$$

$\dot{\lambda}$ (at first unknown) plays in plasticity the rôle of the equivalent strain rate in viscoplasticity.

This type of approach offers the opportunity to introduce two types of coupling:

- *State coupling*, in the free energy (note the symmetry of the interactions):

$$\Psi(\alpha_1, \alpha_2) = \frac{1}{2} c_{11} \alpha_1^2 + \frac{1}{2} c_{22} \alpha_2^2 + \frac{1}{2} c_{12} \alpha_1 \alpha_2 \tag{30}$$

α_1 et α_2 :

$$\frac{\partial A_1}{\partial \alpha_2} = \frac{\partial A_2}{\partial \alpha_1} = \frac{\partial^2 \phi}{\partial \alpha_1 \partial \alpha_2} \tag{31}$$

- *Dissipative coupling*, when Ω is the sum of several potential functions, Ω_K:

$$\dot{z} = \sum_K \frac{\partial \Omega_K}{\partial Z} \tag{32}$$

More information about the development of specific models can be found in (Germain et al., 1983; Lemaitre and Chaboche, 1990; Besson et al., 1998).

2.2 Derivation of single crystal models

General framework The single crystal is seen as a collection of N slip systems, defined by their slip planes $\underline{\boldsymbol{n}}^s$ and slip direction $\underline{\boldsymbol{l}}^s$. The orientation tensor is:

$$\underset{\sim}{\boldsymbol{m}}^s = \frac{1}{2}(\underline{\boldsymbol{n}}^s \otimes \underline{\boldsymbol{l}}^s + \underline{\boldsymbol{l}}^s \otimes \underline{\boldsymbol{n}}^s) \tag{33}$$

In small perturbation, the resolved shear stress is computed as

$$\tau^s = \underset{\sim}{\boldsymbol{\sigma}} : \underset{\sim}{\boldsymbol{m}}^s \tag{34}$$

and the strain partition involves an elastic and an inelastic strain

$$\dot{\underset{\sim}{\varepsilon}} = \dot{\underset{\sim}{\varepsilon}}^e + \dot{\underset{\sim}{\varepsilon}}^p \tag{35}$$

The elasticity is characterized as usual, meanwhile dissipative behaviour must be expressed for each slip system. As a consequence, isotropic and kinematic variables have to be defined on each slip system. The resulting set of variables is then as follows:

Phenomenon	State variable	Associated variable
Elasticity	$\underset{\sim}{\boldsymbol{\varepsilon}}^e$	$\underset{\sim}{\boldsymbol{\sigma}}$
Isotropic hardening	ρ^s, $s = 1..N$	r^s, $s = 1..N$
Kinematic hardening	α^s, $s = 1..N$	x^s, $s = 1..N$

The relation between state and associated variables comes from free energy, where the elastic and the inelastic part are assumed to be disconnected:

$$\rho\psi(\underset{\sim}{\boldsymbol{\varepsilon}}^e, \rho^s, \alpha^s) = \rho\psi^e(\underset{\sim}{\boldsymbol{\varepsilon}}^e) + \rho\psi^p(\rho^s, \alpha^s) \tag{36}$$

The elasticity (non–dissipative process) is fully defined by

$$\rho\psi^e = \frac{1}{2}\,\underset{\sim}{\varepsilon}^e : \underset{\approx}{\mathbf{\Lambda}} : \underset{\sim}{\varepsilon}^e \tag{37}$$

The state variables α^s and ρ^s are present in the free energy, and the hardening variables x^s and r^s are defined as their partial derivatives:

$$\rho\psi^p = \frac{1}{2}\,c\sum_s (\alpha^s)^2 + \frac{1}{2}\,Q\sum_r\sum_s h_{rs}\rho^r\rho^s \tag{38}$$

$$x^r = c\alpha^r \quad ; \quad r^r = b\,Q\sum_s h_{rs}\rho^s \tag{39}$$

Note the *interaction matrix*, whose components h_{rs} characterize self–hardening (if $r = s$) and cross–hardening (if $r \neq s$). The multi-mechanism effect has been studied in detail in the past (Koiter, 1960; Mandel, 1965) from a theoritical point of view. Experimental aspects will be discussed later.

Viscoplasticity needs the definition of a viscoplastic potential, plasticity needs the definition of a plastic pseudopotential. They are built using the expression of the yield criterion on each slip system

$$f^s = |\tau^s - x^s| - r^s - \tau_0 \tag{40}$$

Viscoplastic formulation A viscoplastic potential is introduced. As an example, a power function of the yield is chosen here, but any other increasing function would be a good candidate:

$$\Omega = \sum_r \Omega_r(f^r) = \frac{K}{n+1}\sum_r \left\langle \frac{f^r}{K} \right\rangle^{n+1} \tag{41}$$

The viscoplastic flow can then be derived as

$$\dot{\underset{\sim}{\varepsilon}}^p = \frac{\partial\Omega}{\partial\underset{\sim}{\boldsymbol{\sigma}}} = \sum_r \frac{\partial\Omega_r}{\partial\underset{\sim}{\boldsymbol{\sigma}}} = \sum_r \frac{\partial\Omega_r}{\partial f^r}\,\frac{\partial f^r}{\partial\underset{\sim}{\boldsymbol{\sigma}}} = \sum_r \dot{v}^r\underset{\sim}{\boldsymbol{m}}^r\eta^r = \sum_r \dot{\gamma}^r\underset{\sim}{\boldsymbol{m}}^r \tag{42}$$

In the preceding equation, η^r stand for $sign(\tau^r - x^r)$; the scalar part of the expression characterises the viscoplastic shear rate:

$$\frac{\partial\Omega}{\partial f^r} = \left\langle \frac{f^r}{K} \right\rangle^n = \dot{v}^r \qquad \dot{\gamma}^r = \dot{v}^r sign(\tau^r - x^r) = \dot{v}^r\eta^r \tag{43}$$

and the tensorial part the flow direction, given by the normality rule:

$$\frac{\partial f^r}{\partial\underset{\sim}{\boldsymbol{\sigma}}} = \frac{\partial(|\underset{\sim}{\boldsymbol{m}}^r : \underset{\sim}{\boldsymbol{\sigma}} - x^r| - r^r - \tau_0)}{\partial\underset{\sim}{\boldsymbol{\sigma}}} = \underset{\sim}{\boldsymbol{m}}^r\eta^r \tag{44}$$

For a standard model, the variables attached to isotropic and kinematic hardenings are imposed:

- Kinematic hardening on system s is driven by the actual slip on this system, γ^s

$$\dot{\alpha}^s = -\frac{\partial \Omega}{\partial x^s} = -\sum_r \frac{\partial \Omega_r}{\partial x^s} = -\frac{\partial \Omega_s}{\partial f^s}\frac{\partial f^s}{\partial x^s} = \dot{v}^s \eta^s = \dot{\gamma}^s \tag{45}$$

- Isotropic hardening on system s is driven by the *accumulated slip* on this system, v^s:

$$\dot{\rho}^s = -\frac{\partial \Omega}{\partial r^s} = -\sum_r \frac{\partial \Omega_r}{\partial r^s} = -\frac{\partial \Omega_s}{\partial f^s}\frac{\partial f^s}{\partial r^s} = \dot{v}^s \tag{46}$$

Non standard models can also be defined:

$$\dot{\alpha}^s = (\eta^s - d\alpha^s)\dot{v}^s \tag{47}$$

$$\dot{\rho}^s = (1 - b\rho^s)\dot{v}^s \tag{48}$$

Since $\underset{\sim}{\boldsymbol{\sigma}} : \dot{\underset{\sim}{\varepsilon}}^p = \underset{\sim}{\boldsymbol{\sigma}} : \sum_s \underset{\sim}{\boldsymbol{m}}^s \dot{\gamma}^s = \sum_s \underset{\sim}{\boldsymbol{\sigma}} : \underset{\sim}{\boldsymbol{m}}^s \dot{\gamma}^s = \sum_s \tau^s \dot{\gamma}^s = \sum_s \tau^s \eta^s \dot{v}^s$ the intrinsic dissipation writes:

$$\phi_1 = \underset{\sim}{\boldsymbol{\sigma}} : \dot{\underset{\sim}{\varepsilon}}^p - \sum_s x^s \dot{\alpha}^s - \sum_s r^s \dot{\rho}^s \tag{49}$$

$$= \sum_s (\tau^s \eta^s \dot{v}^s - x^s(\eta^s - d\alpha^s)\dot{v}^s - r^s(1 - b\rho^s)\dot{v}^s) \tag{50}$$

$$= \sum_s (f^s + \tau_0 + \frac{d}{c}\,(x^s)^2 + br^s\rho^s)\dot{v}^s \tag{51}$$

The following contributions are then exhibited:

- Viscous dissipation : $\sum_s f^s \dot{v}^s$
- Friction dissipation : $\tau_0 \sum_s \dot{v}^s$
- Dissipation due to nonlinear hardening: $\left(\frac{d}{c}\,(x^s)^2 + br^s\rho^s\right)\dot{v}^s$

Plastic formulation. Figure 4 illustrates the difference between viscoplastic and plastic formulation. Like for the classical case (Fig.3), the current state in the stress space must stay on the yield surface instead of being on an equipotential. The plastic strain rate for each slip system is no

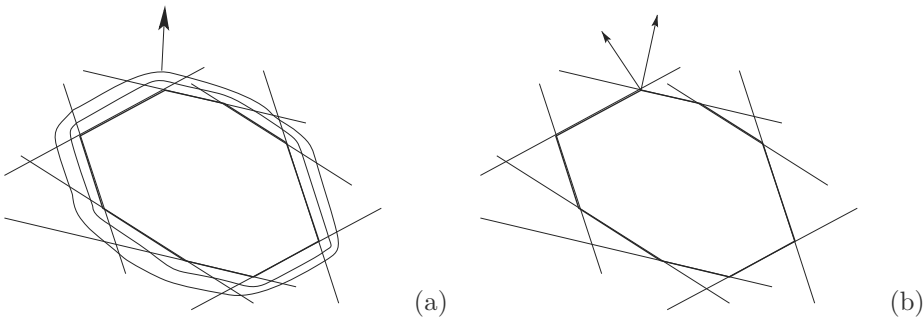


Figure 4. Illustration of (a) the viscoplastic flow, (b) the plastic flow

longer defined by the distance to the relevant yield, but must be deduced from the consistency condition. The regularisation provided by the viscous effect is no longer present, and the direction of the flow has to be defined by additional rules when the current point is on a corner of the domain, that is in situation of multiple slip. The crucial point is then to determine the set of active slip systems. One plastic multiplyier is determined for each slip system. The non zero plastic multiplyiers are solutions of the linear system formed by the consistency conditions on the active slip systems, $\dot{f}^s = 0$. Several sets of slip systems can produce the same viscoplastic strain rate tensor. An additional condition must be used to select the relevant set of slips, as discussed in section (2.5).

$$\underset{\sim}{\dot{\varepsilon}}^p = \sum_{s=1}^{N} \dot{\lambda}^s \frac{\partial f^s}{\partial \underset{\sim}{\boldsymbol{\sigma}}} = \sum_{s=1}^{N} \dot{\lambda}^s \underset{\sim}{\boldsymbol{m}}^s \tag{52}$$

A rather simple computation allows to obtain the relevant system, starting from:

- the strain partition

$$\dot{\underset{\sim}{\boldsymbol{\sigma}}} = \underset{\approx}{\boldsymbol{\Lambda}} : \left(\dot{\underset{\sim}{\boldsymbol{\varepsilon}}} - \sum_r \underset{\sim}{\boldsymbol{m}}^r \dot{\gamma}^r \right) \tag{53}$$

- the consistency condition applied to active slip systems:

$$f^s = |\underset{\sim}{\boldsymbol{m}}^s : \underset{\sim}{\boldsymbol{\sigma}} - x^s| - r^s - \tau_0 = 0 \qquad 0 = \underset{\sim}{\boldsymbol{m}}^s : \dot{\underset{\sim}{\boldsymbol{\sigma}}} - \dot{x}^s - \eta^s \dot{r}^s \tag{54}$$

using the notation $\eta^s = sign(\tau^s - x^s)$

The following step consists in computing $\underset{\sim}{\boldsymbol{m}}^s : \dot{\underset{\sim}{\boldsymbol{\sigma}}}$ in equation (54), introducing H_{sr}, as specified below, and keeping the notation $\dot{v}^s$ for the plastic shear strain rate on system r:

$$\underset{\sim}{\boldsymbol{m}}^s : \dot{\underset{\sim}{\boldsymbol{\sigma}}} = \dot{x}^s + \eta^s \dot{r}^s = \sum_r H_{sr} \dot{v}^r \tag{55}$$

In the next step, one replaces $\dot{\underset{\sim}{\boldsymbol{\sigma}}}$ by its expression in equation (53):

$$\underset{\sim}{\boldsymbol{m}}^s : \underset{\approx}{\boldsymbol{\Lambda}} : \dot{\underset{\sim}{\boldsymbol{\varepsilon}}} - \sum_r \underset{\sim}{\boldsymbol{m}}^s : \underset{\approx}{\boldsymbol{\Lambda}} : \underset{\sim}{\boldsymbol{m}}^r \eta^r \dot{v}^r = \sum_r H_{sr} \dot{v}^r \tag{56}$$

Under prescribed strain rate, N equations are then defined (N = number of active slip systems) to compute the plastic multiplyiers:

$$\sum_r \left(\underset{\sim}{\boldsymbol{m}}^s : \underset{\approx}{\boldsymbol{\Lambda}} : \underset{\sim}{\boldsymbol{m}}^r + H_{sr} \right) \dot{v}^r = \underset{\sim}{\boldsymbol{m}}^s : \underset{\approx}{\boldsymbol{\Lambda}} : \dot{\underset{\approx}{\boldsymbol{\varepsilon}}} \tag{57}$$

- The set of possible slip systems able to respect these equations is not unique. Several combinations of slip can provide the same macroscopic strain rate. This point will be examined later in section (2.5).
- Note that, in the present model, the matrix formed by the coefficients H_{sr} is non symmetric:

$$\dot{x}^s = c\dot{\gamma}^s - dx^s \dot{v}^s = (c\eta^s - dx^s)\dot{v}^s = \sum_r (c\eta^r - dx^r)\delta_{rs}\dot{v}^r \tag{58}$$

$$\dot{r}^s = Q \sum_r bh_{sr} \exp(-bv^r)\dot{v}^r \tag{59}$$

$$H_{sr} = (c\eta^r - dx^r)\delta_{sr} + Qbh_{sr} \exp(-bv^r) \tag{60}$$

Nevertheless, it becomes symmetric with no kinematic hardening and linear isotropic hardening, $r^s = \sum_r h_{sr} v^r$

$$H_{sr} = h_{sr} \tag{61}$$

- The following double dot products may be useful in the computations:

$$\underset{\sim}{\boldsymbol{m}}^s : \underset{\sim}{\boldsymbol{I}} = m^s_{ij}\delta_{ij} = \mathrm{trace}(\underset{\sim}{\boldsymbol{m}}^{\boldsymbol{s}}) = m^s_{ii} = n^s_i l^s_i = 0 \tag{62}$$

$$\underset{\sim}{\boldsymbol{m}}^s : \underset{\sim}{\boldsymbol{m}}^s = m^s_{ij} m^s_{ij} = \frac{1}{4}(n^s_i l^s_j + l^s_i n^s_j)(n^s_i l^s_j + l^s_i n^s_j) = \frac{1}{2} \tag{63}$$

$$\underset{\sim}{\boldsymbol{m}}^s : \underset{\sim}{\boldsymbol{m}}^r = m^s_{ij} m^r_{ij} = \frac{1}{4}(n^r_i l^r_j + l^r_i n^r_j)(n^s_i l^s_j + l^s_i n^s_j) \tag{64}$$

Table 1. The components of the 12 orientation tensors for octahedral slip systems

num syst	1	2	3	4	5	6	7	8	9	10	11	12
$\sqrt{6}m_{11}$	-1	0	-1	-1	0	1	0	-1	-1	-1	1	0
$\sqrt{6}m_{22}$	0	-1	1	0	-1	-1	-1	1	0	1	0	1
$\sqrt{6}m_{33}$	1	1	0	1	1	0	1	0	1	0	-1	-1
$2\sqrt{6}m_{12}$	-1	-1	0	1	1	0	1	0	1	0	1	1
$2\sqrt{6}m_{23}$	1	0	1	-1	0	1	0	1	1	-1	1	0
$2\sqrt{6}m_{31}$	0	1	-1	0	1	1	-1	1	0	1	0	1

- For the case of isotropic elasticity, the system (57) can be simplified, since

$$\Lambda_{ijkl} = \lambda\delta_{ij}\delta_{kl} + \mu(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) \tag{65}$$

$$\Lambda_{ijkl}m^r_{kl} = \mu m^r_{ij} \tag{66}$$

$$\underset{\sim}{\boldsymbol{m}}^s : \underset{\approx}{\boldsymbol{\Lambda}} : \underset{\sim}{\boldsymbol{m}}^r = \mu m^s_{ij} m^r_{ij} \tag{67}$$

$$\tag{68}$$

2.3 Yield surfaces

Yield surfaces provide a synthetic information on the behaviour of the materials. This is illustrated here for the simple case of FCC materials, where slip occurs on the four octahedral planes $\{111\}$, in the directions $\langle 110\rangle$. The case of Ni-Base superalloys, where cubic planes $\{001\}$ may also be involved, with the same slip directions $\langle 110\rangle$, will be also mentioned.

The yield surfaces can be seen as a collection of hyperplans which equations are:

$$|\tau^s| - \tau_c = 0 \quad \text{or} \quad \underset{\sim}{\boldsymbol{\sigma}} : \underset{\sim}{\boldsymbol{m}}^s - \tau_c = 0 \tag{69}$$

Plastic deformation of FCC materials classically involves the 12 octahedral slip systems only. The components of the 12 orientation tensors are shown in Table 1.

The shape of the yield loci can be illustrated on planes like biaxial or tension–shear loading:

- If the only non zero terms of the stress tensor are σ_{11} and σ_{22}, the criterion writes:

$$|\sigma_{11}m_{11} + \sigma_{22}m_{22}| - \tau_c = 0 \tag{70}$$

Table 2. Values of τ^s for the 12 slip systems of a FCC single crystal

(a) *case of a biaxial tension loading* σ_{11}–σ_{22}

num syst	1	2	3	4	5	6
τ^s	$-\sigma_{11}+\sigma_{22}$	σ_{22}	$-\sigma_{11}$	$-\sigma_{11}+\sigma_{22}$	σ_{22}	σ_{11}
num syst	7	8	9	10	11	12
τ^s	σ_{22}	$-\sigma_{11}$	$-\sigma_{11}+\sigma_{22}$	$-\sigma_{11}$	$\sigma_{11}-\sigma_{22}$	$-\sigma_{22}$

(b) *case of a tension–shear loading* σ_{11}–σ_{12}

num syst	1	2	3	4	5	6
τ^s	$-\sigma_{11}-\sigma_{12}$	$-\sigma_{12}$	$-\sigma_{11}$	$-\sigma_{11}+\sigma_{12}$	σ_{12}	σ_{11}
num syst	7	8	9	10	11	12
τ^s	σ_{12}	$-\sigma_{11}$	$-\sigma_{11}+\sigma_{12}$	$-\sigma_{11}$	$\sigma_{11}+\sigma_{12}$	σ_{12}

The resulting equations for the 12 slip systems are shown in Table 2a.

- If the only non zero terms of the stress tensor are σ_{11}, σ_{12} and σ_{21}, the criterion writes:

$$|\sigma_{11}m_{11}+2\sigma_{12}m_{12}|-\tau_c=0 \tag{71}$$

The resulting equations for the 12 slip systems are shown in Table 2b.

As a result, for biaxial tension loading, the domain is then defined by three types of systems :

- systems 1, 4, 9 and 11 : $|\sigma_{11}-\sigma_{22}|=\tau_c\sqrt{6}$
- systems 3, 6, 8, 10 : $|\sigma_{11}|=\tau_c\sqrt{6}$
- systems 2, 5, 7, 13 : $|\sigma_{22}|=\tau_c\sqrt{6}$

meanwhile two types of systems are present for tension–shear loading, as illustrated in Fig.5:

- systems 1 and 11 give : $|\sigma_{11}+\sigma_{12}|=\tau_c\sqrt{6}$
- systems 4 and 9 give : $|\sigma_{11}-\sigma_{12}|=\tau_c\sqrt{6}$

Figure 5 illustrates also the influence of the hardening matrix. It is assumed that the current loading point is at the corner located located on the σ_{12} axis, for a positive value of the resolved shear stress. A stress increment is applied, with an angle θ with the horizontal at this point. Numbers in Fig.5a indicate the number of active slip systems if the interaction matrix is diagonal. On the other hand, if cross hardening (the terms outside of the diagonal) is high, slip on systems like 1 or 11 may dramatically increase the

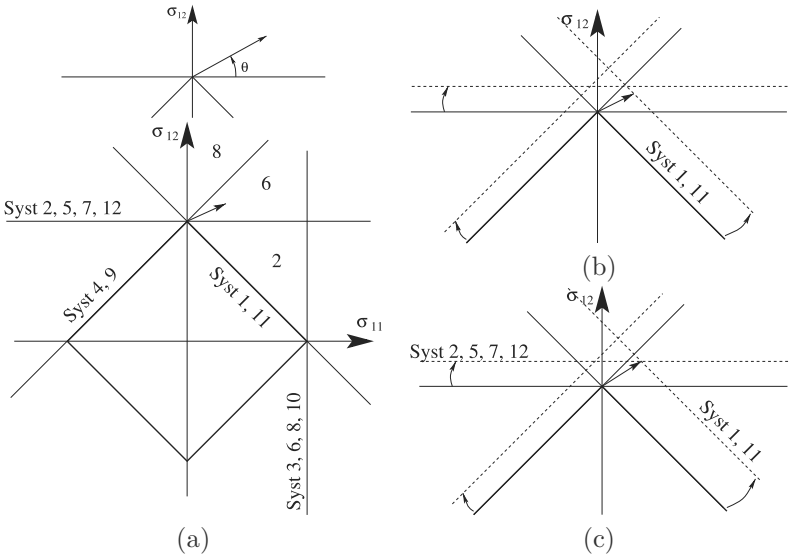


Figure 5. Graphical illustration of the slip system activation: (a) number of potentially active slip systems for various load directions, (b) high latent hardening, a few active systems, (c) low latent hardening, many active systems

resolved shear stress on systems like 2, 5,... so that they will not become active, even if the stress increment is in the sector indicated (6), as demonstrated by Fig.5b. An intermediate solution is recovered for intermediate values of the cross hardening terms (Fig.5c).

Multiple slip is often present in single crystal. This might be due to the rotation of the slip planes during the deformation process. On the other hand, this is also a character which is related to the balance between elastic and plastic flow direction. For instance, in Fig.6a, two loading paths under prescribed strain are considered: for the upper path, the ratio $\varepsilon_{12}^{max}/\varepsilon_{11}^{max}$ is equal to 0.525, for the lower path, the same ratio is equal to 0.475. Since the ratio $\varepsilon_{12}^{pmax}/\varepsilon_{11}^{pmax}$ remains always equal to 0.5, there is a continuous deviation for the stress rate which explains the shape of the stress path. The same argument is valid for explaining the stress path in Fig.6b. Note that in Fig.6a, systems like 2, 5, 7, 12 start being active due to the loading deviation for the upper path, and systems 3, 6, 8, 10 for the lower path.

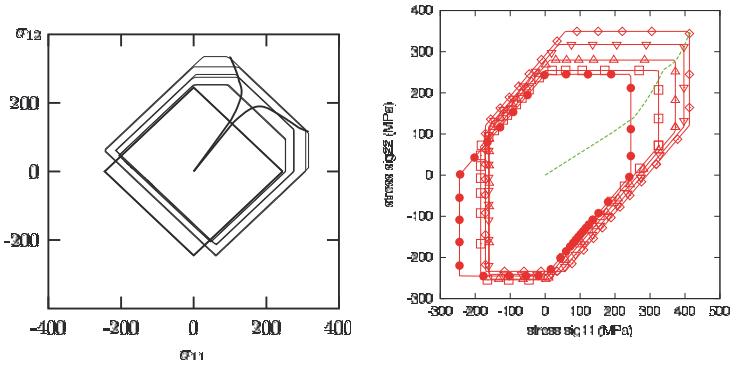


Figure 6. Initial and subsequent yield surfaces: (a) for tension–shear stress states, (b) for biaxial stress states

2.4 Identification under tension and tension–shear loadings

For a given crystal orientation, the present model can be compared with a macroscopic model. Its identification can then be made either by curve fitting using tensile curves, or by a direct use of the material parameters of a macroscopic model. Examples of such an approach can be found in (Méric et al., 1991; Hanriot et al., 1991) for superalloy single crystals. For the case of multiple slip, with N equivalent slip systems, and a Schmid factor m, one can transform the general expression giving $\dot{\underset{\sim}{\varepsilon}}^p$:

$$\dot{\underset{\sim}{\varepsilon}}^p = \sum_s \underset{\sim}{\boldsymbol{m}}^s \dot{v}^s = \sum_s \underset{\sim}{\boldsymbol{m}}^s \left\langle \frac{|\tau^s - x^s| - r^s}{k} \right\rangle^n \tag{72}$$

$$= Nm \left\langle \frac{m(\sigma - x) - r}{k} \right\rangle^n \tag{73}$$

which can be compared to a macroscopic model

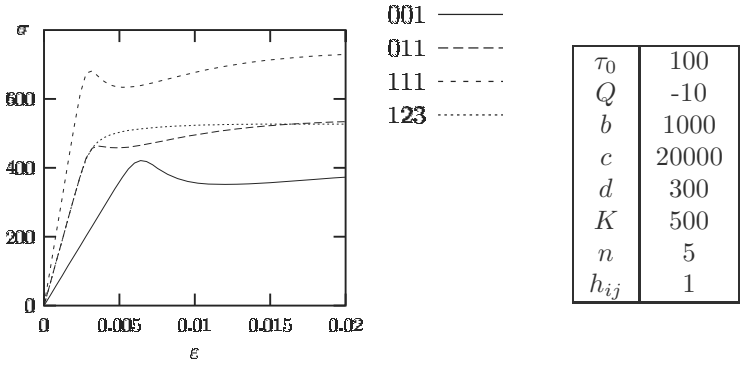
$$\dot{\varepsilon}^p = \left\langle \frac{(\sigma - X) - R}{K} \right\rangle^n \text{ with } K = \frac{k}{m} \frac{1}{mN}^{1/n} \quad X = \frac{x}{m} \quad R = \frac{r}{m} \tag{74}$$

The relations between the macroscopic material parameters and the parameters of the crystalline models are given in Table 3 for the case of octahedral slip and cube slip.

Two examples are given, in order to illustrate the various types of mechanical response of the mode. Figure 7 simply shows the tensile curves

Table 3. Relation between the material parameters of macroscopic models and of crystal plassticity models for octahedral slip and cube slip.

Coefficient	Value for multiple slip (m,M)	Value for 001 tension $N = 8$, $m = 1/\sqrt{6}$	Value for 111 tension $N = 6$, $m = \sqrt{2}/3$
K	$\frac{k}{m(Mm)^{1/n}}$	$\frac{\sqrt{6}k}{(8/\sqrt{6})^{1/n}}$	$\frac{3k}{2^{(n+1)/2n}}$
R_0	$\frac{r_o}{m}$	$\sqrt{6}r_o$	$\frac{3r_o}{\sqrt{2}}$
Q	$\frac{Q}{m}$	$\sqrt{6}Q$	$\frac{3Q}{\sqrt{2}}$
b	$\frac{b}{mM}$	$\frac{\sqrt{6}b}{8}$	$\frac{b}{\sqrt{2}}$
C	$\frac{c}{Mm^2}$	$\frac{3c}{4}$	$\frac{3c}{2}$
D	$\frac{d}{Mm}$	$\frac{\sqrt{6}d}{8}$	$\frac{d}{\sqrt{2}}$



τ_0	100
Q	-10
b	1000
c	20000
d	300
K	500
n	5
h_{ij}	1

Figure 7. Simulation of the tensile behaviour for various crystallographic orientations

obtained with a negative value of the material parameter Q for isotropic hardening. Figure 8 is the result of the identification of the model on an experimental data base obtained on AM1 for various crystal orientations (Hanriot et al., 1991).

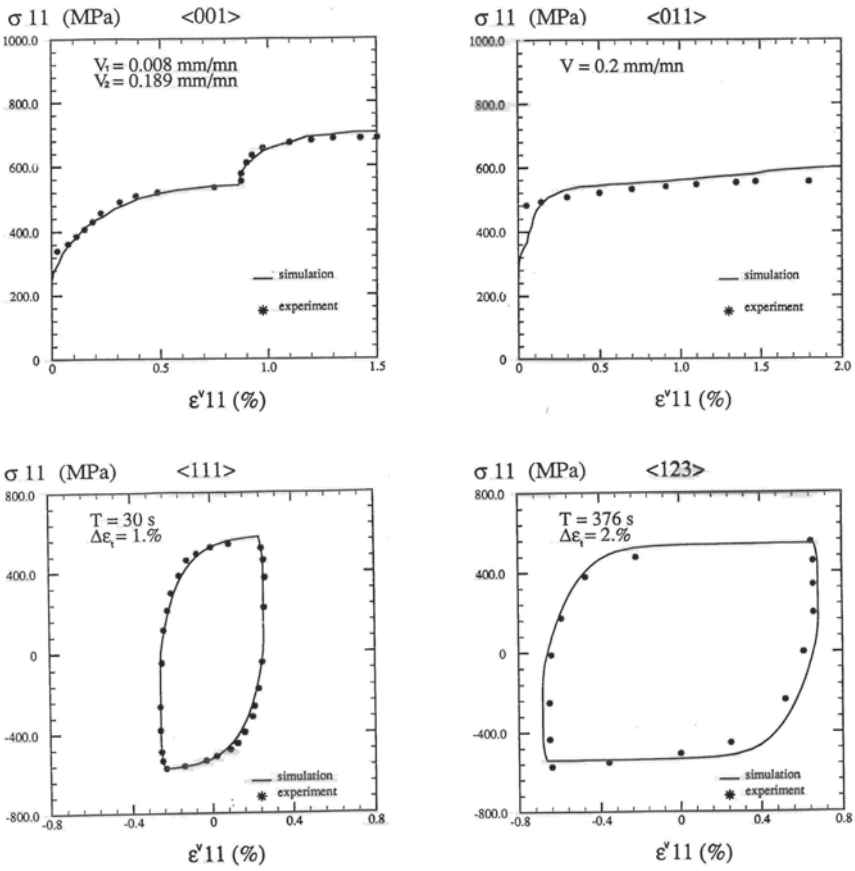


Figure 8. Simulation of the cyclic behaviour for various crystallographic orientations on AM1 specimens (Hanriot et al., 1991)

2.5 Slip system selection

As previously pointed out (section 2.2), there is no unique set of slip systems that provides a given plastic strain rate tensor. The condition to fullfil for each set of slip systems to be admissible can be precised by means of the variational approach, using either internal or external power.

The first problem is to select a set of admissible shear rates $\dot{v}^s$ for a prescribed plastic strain rate $\dot{\underset{\sim}{\varepsilon}}^p$ (assuming $\underset{\sim}{\varepsilon} \equiv \underset{\sim}{\varepsilon}^p$). This has to be done by minimising the internal power of the material element (Taylor, 1938):

$$\mathcal{P}_i = \sum_s \tau_c^s \dot{v}^s \tag{75}$$

the shear strain rates being submitted to the constraint:

$$\underset{\sim}{\boldsymbol{g}}(\dot{v}^s) = \dot{\underset{\sim}{\varepsilon}}^p - \sum_s \underset{\sim}{\boldsymbol{m}}^s \dot{v}^s \eta^s = 0 \tag{76}$$

Let us define the lagrangian $\mathcal{F}_i$, and search for the saddle point

$$\mathcal{F}_i\left(\dot{v}^s, \underset{\sim}{\boldsymbol{\lambda}}\right) = \mathcal{P}_i + \underset{\sim}{\boldsymbol{\lambda}} : \underset{\sim}{\boldsymbol{g}} \tag{77}$$

$$\frac{\partial \mathcal{F}_i}{\partial \underset{\sim}{\boldsymbol{\lambda}}} = \dot{\underset{\sim}{\varepsilon}}^p - \sum_s \underset{\sim}{\boldsymbol{m}}^s \dot{v}^s = 0 \qquad \frac{\partial \mathcal{F}_i}{\partial \dot{v}^s} = \tau_c^s - \underset{\sim}{\boldsymbol{\lambda}} : \underset{\sim}{\boldsymbol{m}}^s \eta^s = 0 \tag{78}$$

The tensor $\underset{\sim}{\boldsymbol{\lambda}}$ is nothing but the stress tensor. To find the set of shear strain rates, which minimizes internal power, it is then necessary to find a stress tensor which obeys the yield conditions, *id est* which allows to build resolved shear stresses that reach τ_c^s on the active slip systems and that are smaller than τ_c^s on the inactive ones.

The problem can also be posed in terms of unknown stress, and rephrased as "find the stress state which maximise the external power" (Bishop and Hill, 1951), written as

$$\mathcal{P}_e = \underset{\sim}{\boldsymbol{\sigma}} : \dot{\underset{\sim}{\varepsilon}}^p \tag{79}$$

under the constraint

$$g^s(\underset{\sim}{\boldsymbol{\sigma}}) = \tau_c^s - \underset{\sim}{\boldsymbol{\sigma}} : \underset{\sim}{\boldsymbol{m}}^s \geqslant 0 \tag{80}$$

The lagrangian function is then now:

$$\mathcal{F}_e(\underset{\sim}{\boldsymbol{\sigma}}, \dot{\mu}^s) = \mathcal{P}_e + \sum_s \dot{\mu}^s g^s \tag{81}$$

$$\frac{\partial \mathcal{F}_e}{\partial \underset{\sim}{\boldsymbol{\sigma}}} = \dot{\underset{\sim}{\varepsilon}}^p - \sum_s \underset{\sim}{\boldsymbol{m}}^s \dot{\mu}^s = 0 \qquad \frac{\partial \mathcal{F}_e}{\partial \dot{\mu}^s} = \tau_c^s - \underset{\sim}{\boldsymbol{\sigma}} : \underset{\sim}{\boldsymbol{m}}^s = 0 \tag{82}$$

The multiplyiers $\dot{\mu}^s$ are nothing but plastic strain rates. To find a stress tensor, one has to find a set of plastic shear strain rates which are zero on the inactive slip systems, and positive on the active slip systems.

The two approaches can be summarised by the double inequality

$$\underset{\sim}{\boldsymbol{\sigma}}^* : \dot{\underset{\sim}{\varepsilon}}^p \leqslant \underset{\sim}{\boldsymbol{\sigma}} : \dot{\underset{\sim}{\varepsilon}}^p = \sum_s \tau_c^s \dot{v}^s \leqslant \sum_s \tau_c^s \dot{v}'^s \tag{83}$$

This discussion can be found in (Chin and Mammel, 1969). It means that:

- between all the admissible stress states conjugated to the real plastic strain rate, the real stress tensor maximises the plastic power. For regular points of the yield surface, this ensures that the plastic strain rate is normal to the yield surface. Nevertheless, this does not provide the full information for a corner of the yield surface, in conditions of multislip;
- between all the possible set of viscoplastic shear strain rates, the real one minimizes the internal power.

The dissipation computed in the thermodynamical approach for such a simple model has two terms only:

$$\Phi_1 = \underset{\sim}{\boldsymbol{\sigma}} : \dot{\underset{\sim}{\varepsilon}}^p - \sum_s r^s \dot{v}^s \tag{84}$$

where the variables r^s, defined in equation (59), denote the *increase* of critical resolved shear stress. According to equation (40) (with $x^s = 0$, since there is no kinematic hardening here), if all the systems are equivalent, the equation can be reduced:

$$\Phi_1 = \sum_s \tau_0^s \dot{v}^s = \tau_0 \sum_s \dot{v}^s \tag{85}$$

2.6 Other crystal plasticity models

The model which has been considered up to now represents the branch of the phenomenological single crystal models, inspired from the macroscopic formulations, like also for instance (Jordan and Walker, 1985). An other class of models is represented by the so called *dislocation based models*, like for instance (Busso and McClintock, 1996). On the other hand, crystal plasticity is a versatile solution to express various phenomena acting on the level of the crystal network, like twinning. These two topics are shown in the present section. Additional capabilities will be shown later, namely developments needed to represent volume changes and damage.

Phenomenological and dislocation based models. We refer to the first denomination when the hardening variables are plastic slip, and to the second one when the dislocation mechanisms are explicitely introduced. One has to get back to Taylor (Taylor, 1938) to read the expression of the first phenomenological model, which assumes that the critical resolved shear stress on the slip systems linearly depend on the slips on all the systems, by means of the interaction matrix. The so called *Taylor interaction* is obtained when all the terms of the matrix are equal to 1: self-hardening and cross-hardening are exactly equivalent. Experimental measurement of cross hardening have then allowed to reach better estimations of the hardening matrix (Kocks and Brown, 1966; Kocks, 1970; Franciosi, 1985a,b). This matrix is now estimated by means of Dislocation Dynamics computations (Devincre et al., 2006). The most widely used phenomenological model is written in a viscoplastic framework without threshold (Asaro, 1983b), so that all the systems are allways active. A multiplicative isotropic hardening is introduced; it depends on accumulated slip:

$$\dot{\gamma}^s = \dot{\gamma}_0 \left(\frac{\tau^s}{\tau_c^s} \right)^{1/m} \tag{86}$$

$$\tau_c^s = \sum_{r=1}^{N} h_{sr} \gamma^r \tag{87}$$

$$\tag{88}$$

The phenomenological models previously presented in this paper and the dislocations based models have the same nature: their variables globally represent averages of the local dislocation densities. To fully illustrate this point, one can take (equations 89–91) the example of a dislocation based model (Tabourot et al., 1997), where b is the Burgers vector, and ρ^s the dislocation density on slip system s:

$$\dot{\gamma}^s = \dot{\gamma}_0 \left(\frac{\tau^s}{\tau_\mu^s} \right)^{1/m} \tag{89}$$

$$\tau_\mu^s = \alpha \mu b \left(\sum_{r=1}^{N} h_{sr} \rho^r \right)^{1/2} \tag{90}$$

$$\dot{\rho}^s = \frac{1}{b} \left(\frac{\left(\sum_{r=1}^{N} a^{sr} \rho^r \right)^{1/2}}{K} - 2 y_c \rho^s \right) \dot{\gamma}^s \tag{91}$$

After choosing $r^s = b^2\rho^s$ and $y_c = bk_c$, the last two equations can be rewritten:

$$\tau_\mu^s = \alpha\mu \left(\sum_{r=1}^{N} h_{sr} r^r \right)^{1/2} \tag{92}$$

$$\dot{r}^s = \left(\frac{\left(\sum_{r=1}^{N} a^{sr} r^r \right)^{1/2}}{K} - 2k_c r^s \right) \dot{\gamma}^s \tag{93}$$

The variable r^s which is defined here plays exactly the role of ρ^s in equations (38) and (48).

Twinning and phase transformation Twinning and martensitic transformation can be represented in a local frame $\underline{\boldsymbol{n}}^k$, $\underline{\boldsymbol{t}}^k$,$\underline{\boldsymbol{u}}^k$ by a transformation gradient F^k

$$F = I + \underline{\boldsymbol{t}^k} \otimes \underline{\boldsymbol{n}^k} \tag{94}$$

The stress-free transformation tensor is then $\frac{1}{2}\left(F^T F - I\right)$ in large strain, or $\frac{1}{2}\left(F^T + F\right) - I$ in a small perturbation framework. In this last case, the transformation of the crystal lattice produced by the variant k can be seen as a parametric strain

$$\underset{\sim}{\boldsymbol{\varepsilon}}^{tp} := \begin{pmatrix} \varepsilon & \gamma & 0 \\ \gamma & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{95}$$

Various theories are available to determine vectors $\underline{\boldsymbol{n}}$ and $\underline{\boldsymbol{t}}$, depending on the material crystallography (for instance Wechsler-Lieberman-Read theory, 24 variants)

3 Finite element computations of single crystalline components

The purpose of this section is to illustrate the behaviour of single crystal laboratory specimens and of industrial components. After a quick look to the algorithmic expression of the model, one shows the case of cylindrical tension specimens, then tubular specimens. A few elements are given to figure the type of calculations that can be made on turbine blades.

3.1 Algorithm for the numerical integration

The integration of these equations in a finite element code can be made either by an explicit or an implicit method. If compared with classical inelastic equations, single crystal models must solve the additional problem of choosing the active slip systems. This is why, after classical viscoplastic solutions where all the slip systems are active (Pierce et al., 1985; Asaro and Needleman, 1985), many authors have proposed their original solutions (Méric and Cailletaud, 1991; Cuitino and Ortiz, 1992; Anand and Kothari, 1996; Schröder and Miehe, 1997; Simo and Hughes, 1997; McGinty and McDowell, 2006). The interest of the implicit approach is that it provides the jacobian matrix needed for computing the consistent tangent matrix. Classically, the system of ordinary differential equations is solved by a Newton technique, which is applied to the incremental form coming from the application of a θ–method. For the sake of simplicity, the present model is restricted to small strain, and linear isotropic hardening. In such a case, the integration variables are the increment of elastic strain $\Delta\varepsilon^e$ and the increments of (visco)plastic shear strain, Δp^s, which are positive for all the slip systems. For the case of time independent plasticity, these two variables will be deduced from two types of equations:

- The strain partition:

$$\Delta\underset{\sim}{\varepsilon} = \Delta\underset{\sim}{\varepsilon}^e + \sum_r \underset{\sim}{m}^r \Delta p^r \tag{96}$$

- The expression of the criteria for the active slip systems:

$$f^s = |\tau^s| - r^s = 0 \tag{97}$$

$$\text{with } \tau^s = \underset{\sim}{m}^s : \underset{\approx}{\Lambda} : \underset{\sim}{\varepsilon}^e \qquad r^s = \sum_r h_{sr} p^r labeleq : hardx \tag{98}$$

As a result, the system to solve is formed of equations (99) and (100):

$$\mathcal{F}_e = -\Delta\underset{\sim}{\varepsilon} + \Delta\underset{\sim}{\varepsilon}^e + \sum_r \underset{\sim}{m}^r \Delta p^r \tag{99}$$

$$\mathcal{F}_p^s = \left| \underset{\sim}{m}^s : \underset{\approx}{\Lambda} : (\underset{\sim}{\varepsilon}^e + \Delta\underset{\sim}{\varepsilon}^e) \right| - \tau_0 - \sum_r h_{sr}(p^r + \Delta p^r) \tag{100}$$

If the model is viscoplastic instead of being plastic, equations (100) must be replaced by (101), the equation coming from the strain partition being unchanged:

$$\mathcal{F}_p^s = \left| \underset{\sim}{m}^s : \underset{\approx}{\Lambda} : (\underset{\sim}{\varepsilon}^e + \Delta\underset{\sim}{\varepsilon}^e) \right| - \tau_0 - \sum_r h_{sr}(p^r + \Delta p^r) - K \left(\frac{\Delta p^s}{\Delta t} \right)^{1/n} \tag{101}$$

This algorithm is implemented in the `Zmat` library of the `ZéBuLoN` code (Foerch et al., 1997) which has been used for all the simulations of this chapter.

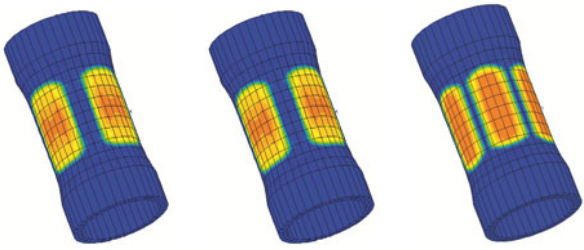
3.2 Laboratory specimens

It is well known that the stress and strain state is not homogeneous in "tensile" specimens which are not oriented along high symmetry directions like {001} or {111}. If the rotations are blocked, the strain heterogeneity can even be very high, as illustrated in Fig.9a (Méric and Cailletaud, 1991). Other cases are still more remarkable: torsions of a thin wall tube. According to the orientation of the tube axis, the number of "soft" zones around the circumference will change. This can be easily understood by computing the resolved shear stress along the circumference. For the case of a {001} oriented specimen, there is four soft zones (Fig.9b), and for a {111} oriented specimen, six soft zones (Fig.9c). Note that a series of criteria proposed in the literature for single crystal modelling, which are based on Hill's quadratic criterion do not capture this effect, since all the points of the circumference have the same shear stress. Experiments have been made to confirm the prediction given by the model. This is illustrated by Fig.10, which shows the location of the jauges (Fig.10a), meanwhile Fig.10b and c respectively show the loops measured by the jauges, respectively in the $< 110 >$ and in the $< 100 >$ areas: the strain measured by the jauge is on the horizontal axis, and the torsion couple is on the vertical axis. The Schmid factor is rather small in the second case, thus the behaviour is not far from purely elastic. On the contrary, $< 110 >$ is a soft zone, with a large hysteresis loop. The reader is invited to go back to (Nouailhas and Cailletaud, 1995) for a more detailed view on the tests available. Yield surfaces have been determined in tension–torsion. It is worth noting that, according to the ratio between tensile and shear loadings, the softer zones can be either $< 110 >$ or $< 100 >$!

Other specimens have also be considered in the past, for instance bicrystalline specimens (Méric et al., 1994).

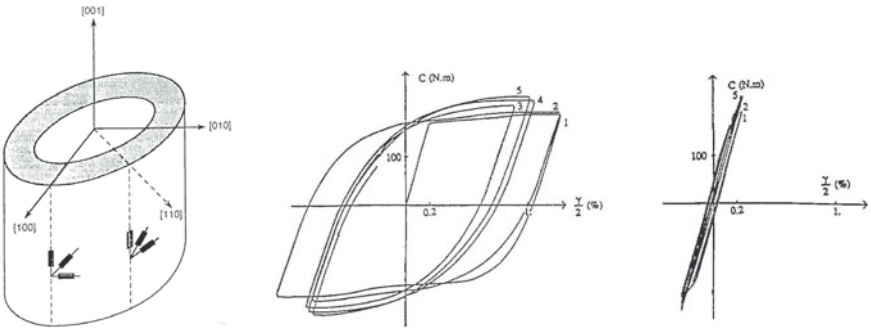
3.3 Turbine blades

Single crystals are used in industry, specially in the hotest sections of the turbines (for planes or power plants), where they accept higher temperatures than classical alloys (Cailletaud et al., 2001; Busso et al., 2003). Figure 11 shows a typical computation result performed in cooperation with Snecma (Safran group): after stress–strain computations, specific post-processing are used to estimate the creep–fatigue life of the component. The most re-



(a) (b) (c)

Figure 9. Various loading cases producing strain heterogeneties on laboratory specimens: (a) {123} oriented specimen loaded in tension; (b) {001} oriented specimen loaded in torsion; {111} oriented specimen loaded in torsion;



(a) (b) (c)

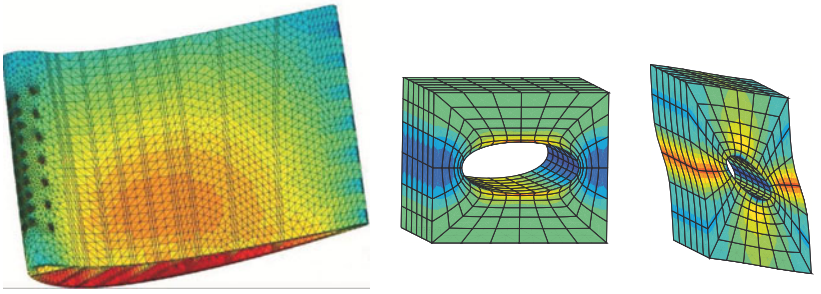
Figure 10. (a) Location of the jauges on the tube; (b) soft response of the jauge located in $< 110 >$ area; (c) allmost elastic response of the jauge located in $< 100 >$ area

Figure 11. (a) Life prediction on a turbine blade submitted to a thermo-mechanical loading

cent advances in blade design are related to more and more complex shapes, in relation with aerodynamics which make 3D computations mandatory, and also with more and more clever cooling systems. Modern blades are hollow, with a complex network of holes to have cool gas coming from inside the blade and form a protective film against the combustion gases. This is the reason why the critical zones are no longer predictable by inspired guessing from the shape of the blade.

On the other hand, this complex geometry introduces sometime so many details that they cannot be meshed with a reasonable precision. This is illustrated in Fig.12a, where a series of 1 mm diameter holes are present on the leading edge of the blade. The computation with the full geometry can be made for a reference computation, but the relevant mesh to capture all the details (several millions of nodes) cannot be used for parametric studies. In such a case, the weak zone can be globally modelled by means of homogenised models taking into account the holes. Drilling a hole in a single crystal will produce a material that does not possess the cubic symmetry. The axis of the hole is a preferential axis: along this axis, the material behaviour is unchanged, except that a correction of effective surface has to be made. For a tension or a compression along the two perpendicular axes, the behaviour becomes pressure sensitive. The resulting model must then be an orthotropic pressure sensitive single crystal. An homogenized model can be calibrated by means of cell calculations, like in Fig.12b and Fig.12c. The cells are loaded with periodic boundary conditions.

Assuming that the axis of the hole is $\underline{\boldsymbol{e}}_2$, a pressure effect is created for



(a) (b) (c)

Figure 12. Development of a model to account for hole distribution (Cardona, 2000): (a) industrial component; (b) cell with a hole in tension; (c) cell with a hole in shear

the other two directions, $\underline{\boldsymbol{e}}_1$ and $\underline{\boldsymbol{e}}_3$ by introducing two new mechanisms in addition to the normal slip: this is an opening/closure of the hole, driven by the normal stress in the relevant direction, namely:

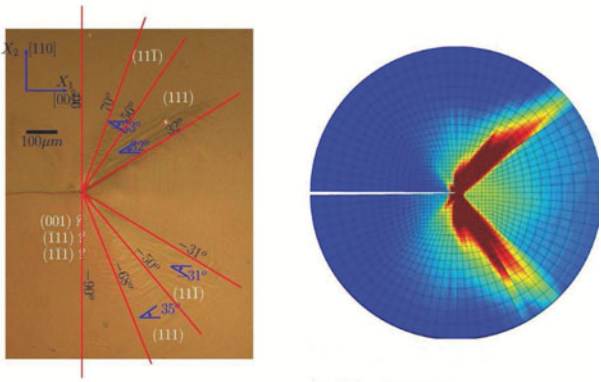
- two normal stresses are computed, $\sigma_1 = \underline{\boldsymbol{e}}_1.\underset{\sim}{\boldsymbol{\sigma}}.\underline{\boldsymbol{e}}_1$ and $\sigma_3 = \underline{\boldsymbol{e}}_3.\underset{\sim}{\boldsymbol{\sigma}}.\underline{\boldsymbol{e}}_3$
- two variables δ^1 and δ^3 are associated to these normal stresses:

$$\dot{\delta}^i = \left\langle \frac{\sigma_i - r^i}{K} \right\rangle^n$$

- δ^1 and δ^3 have their own contribution to the strain rate tensor, evaluated by means of $\underset{\sim}{\boldsymbol{n}}^1 = \underline{e}_1 \otimes \underline{e}_1$ and $\underset{\sim}{\boldsymbol{n}}^3 = \underline{e}_3 \otimes \underline{e}_3$ through the terms:

$$\dot{\delta}^1 \underset{\sim}{\boldsymbol{n}}^1 + \dot{\delta}^3 \underset{\sim}{\boldsymbol{n}}^3$$

The holes in the blades generate stress concentrations which lead to crack initiation. It is then useful to study cracks in single crystals. Due to the specificity of the yield criterion, the plastic zones ahead of the crack tip have a particular aspect. This was predicted by early studies made by Rice (Rice, 1987). Experimental veerifications and the related numerical simulations have been recently made(Flouriot et al., 2003). Figure 13, taken from this last paper, shows for instance that there is no plasticity in the direction of the crack ahead of the crack tip in an angle of about $\pm 30°$, for the case of a $\{001\}$ oriented specimen of FCC material.



(a) (b)

Figure 13. View of the plastic activity ahead of a crack tip. (a) Experimental result; (b) Numerical simulation

4 Finite Element Crystal Plasticity

In the litterature, internal stresses in materials are classified into level I, II and III. Level III is the *microscopic level*, where intragranular stress present fluctuations inside each grain, with different values in the core of the grain and near grain boundaries, for instance. Level II denotes the result obtained at the grain level, after averaging the intragranular stresses. This can be made grain by grain; In this case, the resulting tensor is defined in each in individual grain, and takes into account the real neighborhood. Most of the time, measurements are rather able to capture stress on a given set of grains having the same crystallographic nature. This leads to an alternative definintion for level II, that is now obtained after an additional averaging operation. Level I refers to macroscopical stress, the stress tensor resulting from the averaging on a representative material element, large enough to include all the significant phases and grains.

Numerical models are attached to each of these scales. Macroscopic models naturally relates to level I. Uniform field models represent level II. The finite element method with embedded crystal plasticity models (Finite Element Crystal Plasticity, FECP) can be used to capture stress and strain fields on level III. It has been applied to calibrate the other levels too. This is illustrated in the present section.

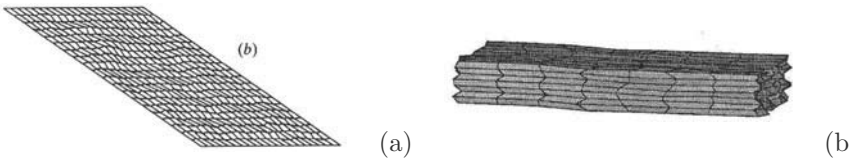


Figure 14. Use of regular meshes for studying large deformations of polycrystals and the related texture evolution: (a) 2D (Kalidindi et al., 1992), (b) 3D (Miehe et al., 1999)

Polycrystal, simplified microstructures In the past, finite element models have been used either to evaluate the homogenized behaviour of an aggregate with a predefined microstructure, or to make comparisons with local measurements in specimens with large grains.

Attempts to derive the global behaviour of an aggregate from the properties of its elementary constituents by means of 3D finite element computations can be found in the seventies (Engel, 1978). The material was represented by tetrahedric grains assembled in a cube, with a new crystal orientation for each finite element. The severe limitation at that time was the power of the computer, so that researchers had to wait for new generation of machines. Figure 14 shows further typical studies, made on 2D meshes ((Kalidindi et al., 1992)) or 3D meshes ((Miehe et al., 1999)). In these cases, authors wanted to check the global behaviour of the aggregates, and predict the texture evolution in large deformation. The aggregate has nothing to do with a realistic microstructure, since each finite element has its own crystallographic orientation (*one element–one grain*), figuring a set of cubic grains with regularly distributed neighbours. The model is rather seen as an evolution of Taylor's model, where the kinematic conditions (uniform plastic strain) is known to be too restrictive. In fact, the results obtained with such crude meshes are not far from Taylor solution: this is not surprinsing, since there is not enough degrees of freedom to reproduce the high heterogeneity present in real aggregates. On the other hand, the prediction of the texture evolution is often satisfactory. It remains that the main reliable information that can be taken from these computations are the so called first order stress and strain fields, that is the homogenized fields.

Some more complex meshes have then been used, to represent more precisely the local behaviour. A series of computations are illustrated in Fig.15. Figure 15a shows the finite element model used to compute the av-

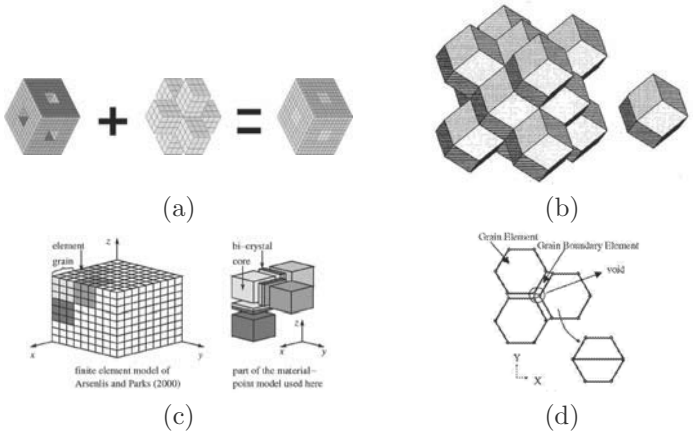


Figure 15. Examples of meshes used in the literature to represent various microstructures: (a) two phase material (Bugat et al., 1999), (b) attempt to capture a realistic grain microstructure (Mika and Dawson, 1999), (c) grain boundaries on a 3D cubic mesh (Evers et al., 2002), (d) grain boundaries on 2D hexagonal grains (Kim et al., 2002)

erage stress tensors in a two phase material (Bugat et al., 1999). The first phase (left) is represented by 20 elementary cubes, the second by 7 elementary cubes. Introducing symmetries, each phase forms a coninuous periodic network. The respective volume fraction can be controlled by the size of the internal cubes. The purpose of the computation is to evaluate a stress level in each phase, to be used as an input in brittle failure models, that is level II stress. The next example in figure (15b) is between level II and III, since a more realistic shape is introduced (Mika and Dawson, 1999). The last two examples (Fig.15c and d) try to separate the interior of the grain and a grain boundary area. Again, more realistic meshes would provide more reliable computation of the rspective influence of each zone (Evers et al., 2002; Kim et al., 2002).

Multicrystal, real microstructures The other traditional type of study consists in computing real specimens with a relatively low number of grains. Contrary to the previous case, there is a large number of elements in each grain, so that the intragranular gradients can be obtained. This allows authors to illustrate multislip effect (Delaire et al., 2000) inside the grain or

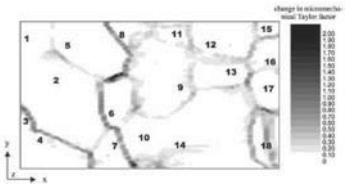
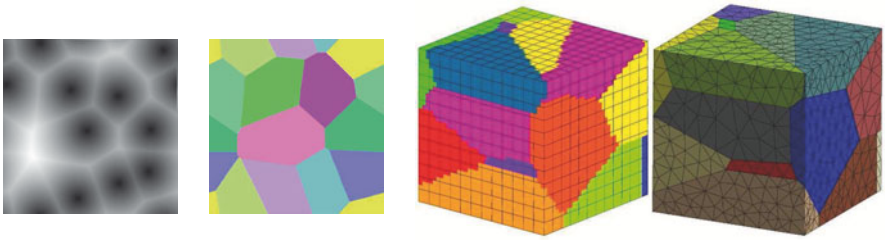


Figure 16. Examples of experiments and simulation using multicrystalline specimens: (a) experimental results on a copper specimen (Delaire et al., 2000), (b) illustration of the perturbation due to grain boundaries on Taylor factor (Raabe et al., 1981)

at the grain boundaries (Raabe et al., 1981). The state of the art for such a class of calculations if as follows:

- there is a good agreement between experiments and numerical simulations, specially for large grains;
- grain boundaries are still not well characterised; secondary slip might be present in rather large zones;
- the behavior of a grain in its environment differs from the behaviour of the corresponding single crystal;
- the unknown grains under the surface do affect the behaviour of the surface grains.

Polycrystal, realistic microstructures An other trend in FECP is the modeling of polycrystalline aggregates, involving a large number of grains (several hundreds) to capture the polycrystal effect, using realistic synthetic microstructures to capture the effects of the local morphology and of the neighborhood, and involving a reasonnable number of elements in order not to introduce artefacts due to the poor discretization. The interested reader will consult papers like (Barbe et al., 2001a,b; Diard et al., 2002; Cailletaud et al., 2004). The most popular solution to generate the numerical model starts from a Voronoï tesselation. As shown in Fig.17a and b, this construction provides domains around points that are initially dropped in the material element. The distribution can be random. Alternatively, a repulsion distance can be introduced, anisotropic distributions can be considered... Other types of distributions are also in development: they consider ellipsoids growing at various rates, or other types of assumptions based on models for grain growth. Having a microstructure in hand, the next step of a microstructure computation is to generate the finite element



(a) (b) (c) (d)

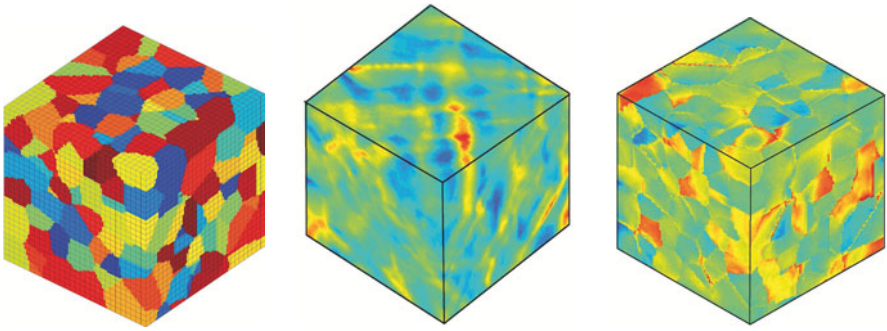
Figure 17. Mesh generation using Voronoï tesselation: (a) Distance function of a set of point sources, (b) Tesselation result after construction and labelling, (c) multiphase element technique, (d) 3D mesh respecting grain boundaries

mesh. In the past, authors have used either multiphase element technique (Fig.17c) or free mesh generation (Fig.17d). According to the first solution, material properties are affected to the Gauss points according to their position in a cube which defines the material element. This is a quite flexible approach, nevertheless, grain boundaries are cumbersome, so that it is not applicable for a good characterization of the grain boundaries. On the other hand, free mesh generation do respect grain boundaries, and is a good starting point to look for real intragranular fields (level III).

Figure 18 presents a typical result on a regular 28×28×28 mesh. Multiphase element technique has been used. The general behaviour or the strain fields is to form strain localization lines, as shown in Fig.18b. This has nothing to do with instabilities, since the material presents hardening everywhere, on a local and a global scale. This is just the fact that deformation propagates from one grain to the other, sometime without consideration of the local orientation. On the other hand, stress levels do respect grain boundaries. The von Mises equivalent stress field shown in Fig.18c is clearly sensitive to the crystallographic morphology.

Convergence studies have been performed in (Barbe et al., 2001a,b). The most remarkable points are the following:

- For moderate strains (a few percent), quadratic meshes with full integration provide more precise results than linear meshes, for the same number of nodes. Nevertheless, an oscillation due to the spherical part of the stress tensor may appear when classical element are used. Mixed elements introducing the trace (and eventually the volume change) as a degree of freedom provide more stable results. Alternatively, a post-



(a) (b) (c)

Figure 18. Typical results on a polycrystalline aggregate: (a) a 28×28×28 mesh with 200 grains, (b) local field of total axial strain, (c) local field of von Mises stress

treatment of the result can be made, to replace the trace on each Gauss point by the trace of the element.

- Reaching a stable curve for level I stress (macroscopic level) is rather easy. This can even be achieved with a one element–one grain mesh ! The variations observed on a tensile curve reaches only a few percent with two hundred grains for various mesh size.
- Level III stresses and strains are much more difficult to capture. For coarse meshes, the localization areas are not well defined, they are too large, but the maximum values are too small: the finer the mesh, the higher the maximum stress (or strain), the lower the minimum stress (or strain). This is illustrated in Fig.19, on a slice made in several 3D meshes, with an increasing number of elements.
- Fully constraint boundary conditions (imposed strain) provide the most homogeneous fields, and the highest resulting stress (at level I). Figure 20 shows the resulting axial strain–axial stress curves for each of the 200 grains of a polycrystalline aggregate (level II). In Fig.20a, the stress–strain state of the grains follows the Berveiller–Zaoui approach, which will be explained later (section 5.2): for a given time step, they are along a line whose slope decreases with strain. According to this uniform field model, each grain is surrounded by all the other grains. This is not the case in the finite element method, so that the distribution is not so smooth. Tension under strain control (Fig.20b) produces m ore scatter in stress (low scatter on strain),

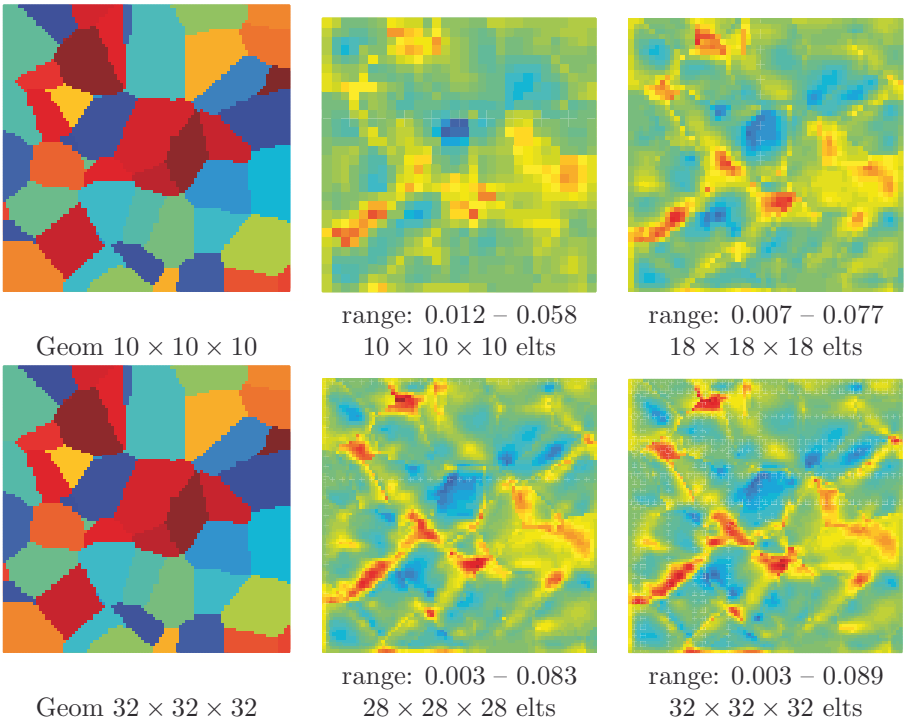


Figure 19. Evolution of the local strain field (von Mises equivalent) for various mesh sizes, for an axial tension (0.2% in the vertical direction, 200 grains) on a quadratic mesh

meanwhile tension under stress control (Fig.20c) produces more scatter in strain. The graphs in Fig.20d, e, f show the results for the lateral stresses, whose average is equal to zero. This plot shows that the stress state is triaxial inside the aggregate, even for a tensile strain.

- As far as level III is concerned, free faces promote scatter: for a given mesh, the computation performed with a free face will produce lower values *and* smaller values than the computation made under prescribed strain. The same observation is made inside the grains. Grain boundaries promote also scatter: in its vicinity, maximum stress or strain increase, minimum stress or strain decrease (see Fig.21).

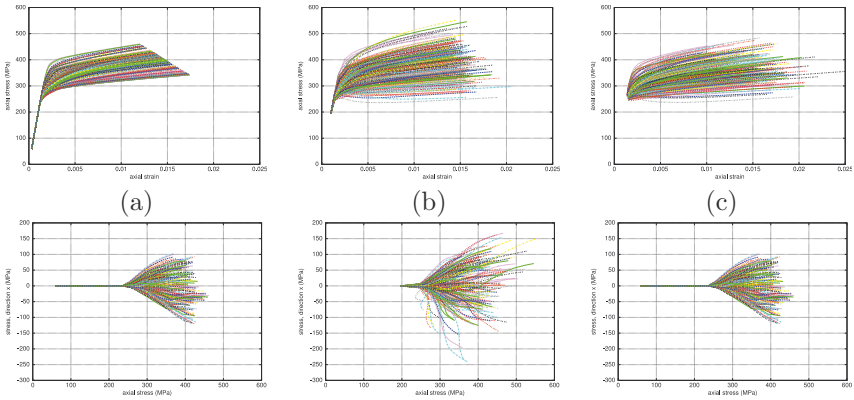


Figure 20. Local stress-strain curves for 200 grains for different boundary conditions, compared to the self-consistent approach: axial components (a) for the self-consistent model, (b) for homogeneous boundary conditions (c) for 4 free faces; lateral components (d) for the self-consistent model, (e) for homogeneous boundary conditions (f) for 4 free faces

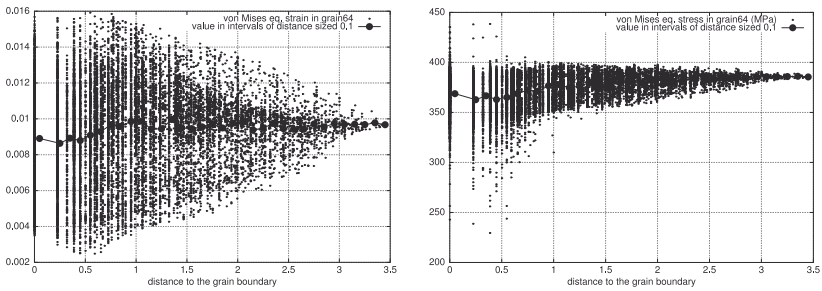


Figure 21. (a) Equivalent plastic strain (von Mises) *versus* dist to the GB; (b) Equivalent stress (von Mises) *versus* dist to the GB

Recent developments

In the preceding examples, Finite Element Crystal Plasticity has been used either directly to compute specimens or components made of a single crystal, or in representative material elements. For this last case, the output can be the macroscopic stress and strain fields, the stress and strain fields in the grains or the local intragranular fields (level I, II, III respectively). Results at level I and II are devoted to a calibration of macroscopic or uniform field models. Level III is more and more interesting, since the meshes are now large enough to capture details near critical zones like grain boundaries. On the other hand, this is the relevant scale to open dialogs with other scales and other types of plasticity models, like the dynamic dislocation simulations. Level II or III can be the relevant scale to introduce failure prediction models, like cleavage or integranular damage. Some example of recent trends are now mentioned in this section.

Figure 22 shows the modeling of a test on a OFHC copper specimen. Serial cuts were done to get the microstructure geometry below the surface, so that the grain shape is known on a parallelipipedic box whose depth is 100 μm, and the surface 200 μm$\times$150 μm. Crystal orientations were characterized by OIM. The material is modelled by the constitutive equations shown in section (2.2). The material parameters are fitted on the macroscopic tension curve. Simulated results (strain and rotation fields) are in good agreement with the measurements made by image comparison technique (Musienko et al., 2007).

As stated previously, grain boundaries are the place where specific phenomena may arise. This is the case when SCC mechanisms are active, for instance in fuel assemblies of PWR. The expansion of the uranium dioxide pellets may produce an interaction with Zy4 (a zirconium alloy) cladding, producing the so called *Pellet–Cladding Interaction*, that generates intergranular cracks, whose propagation is promoted by iodine due to the nuclear reaction, then cleavage and failure of the tube. The mechanical effect is related to an indentation of the tube by pellets: as a consequence, the affected zone is very small, and important gradients can be observed on the first few grains under the surface. An appropriate model has then to take into account the crystallographic character. The numerical model developed (Cailletaud et al., 2004) introduces a coupled FE calculation: iodine adsorption at the grain boundaries is represented by a diffusion calculation (where the diffusivity strongly depend on damage of the grain boundary, so that the high concentration profile follows the crack propagation) and a mechanical calculation (where *grain boundary affected zones* are represented by specific constitutive equations). A catastrophic process is produced by

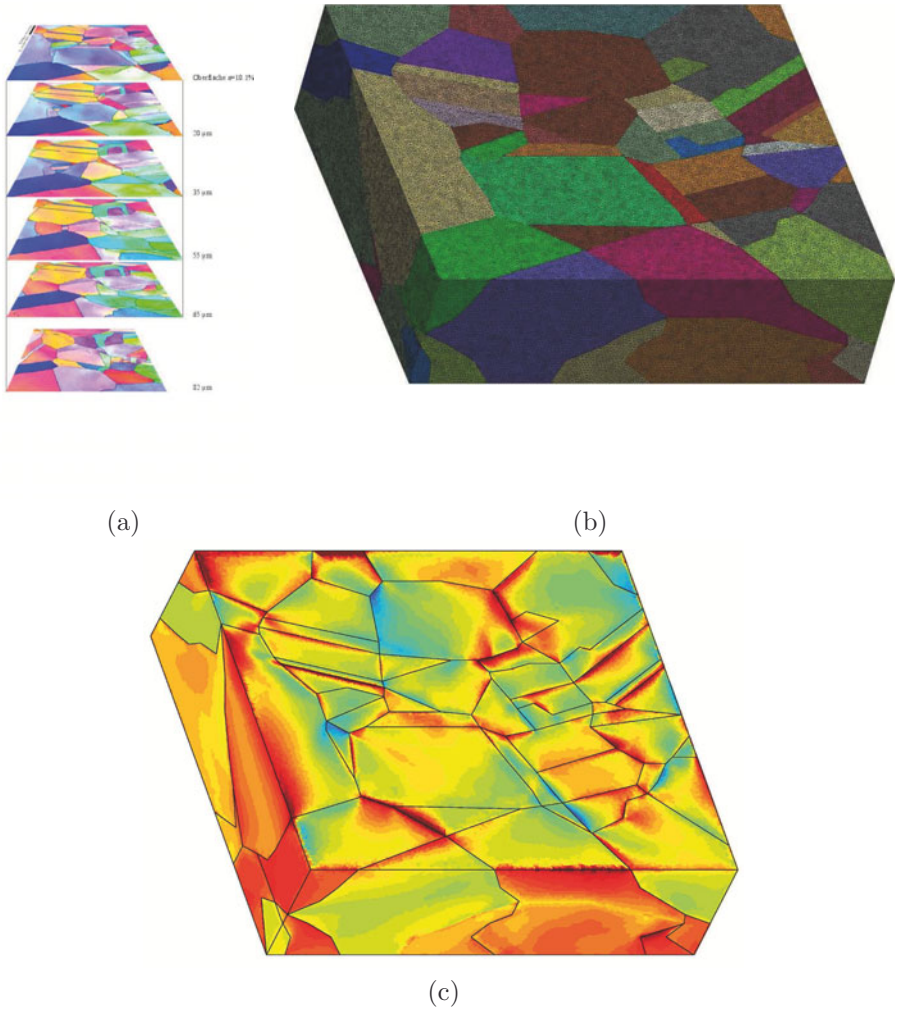
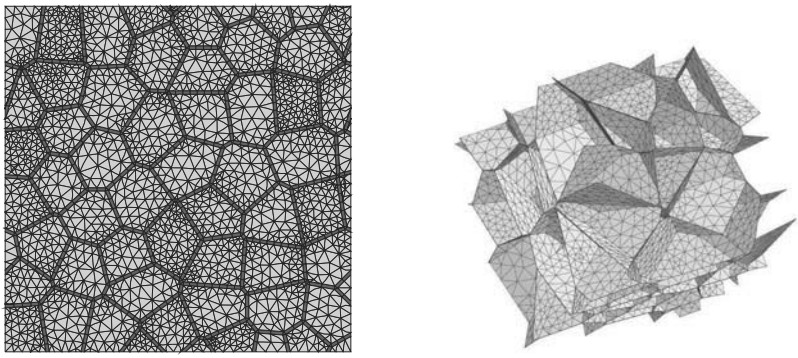


Figure 22. FE computation of a copper specimen. (a) OIM analysis on serial cuts in the specimen, (b) associated mesh, (c) contour of equivalent plastic strain for a macroscopic strain of 5% (range for local strain is 0–0.214)



(a) (b)

Figure 23. Example of mesh generation with grain boundaries : (a) 2D, (b) 3D

the fact that iodine concentration promotes damage, and damage promotes iodine advance in the grain boundaries. 2D and 3D meshes are available (Fig.23).

The model introduces an affected zone in *each* grain, so that the *grain boundary* is made of two elements. In these area, the model collects first the slip systems of the grain it belongs to: basal, (predominant) prismatic and pyramidal slip planes are present, since the material has a HCP microstructure. Three other scalar variables are present, to represent Damage, Opening and Sliding of the grain boundary (DOS model). Grain boundary opening mechanisms is built with the tensorial product of the normal vector $\underline{\boldsymbol{n}}$ to the grain boundary by itself, meanwhile grain boundary sliding is built with the tensorial product of $\underline{\boldsymbol{n}}$ by the tangent vector in the grain boundary plane, $\underline{\boldsymbol{t}}$. Damage is introduced in the model to increase the rate of opening and sliding. Its evolution depends on a critical variable which account for the normal and the shear stresses on the grain boundary. The resulting equations are given below:

- Elasticity and plastic flow

$$\text{Strain decomposition:} \quad \dot{\underset{\sim}{\varepsilon}} = \dot{\underset{\sim}{\varepsilon}}^e + \dot{\underset{\sim}{\varepsilon}}^p \tag{102}$$

$$\text{Elastic law with isotropic damage:} \quad \underset{\sim}{\boldsymbol{\sigma}} = (1-D)\underset{\approx}{\boldsymbol{L}} : \underset{\sim}{\boldsymbol{\varepsilon}}^e \tag{103}$$

$$\text{Opening and sliding:} \quad \dot{\underset{\sim}{\varepsilon}}^p = \dot{\delta}\underline{\boldsymbol{n}} \otimes \underline{\boldsymbol{n}} + \dot{\gamma}^{\{}\underline{\boldsymbol{n}} \otimes \underline{\boldsymbol{t}}^{\}} \tag{104}$$

- Flow rules for opening and sliding

$$\text{Opening:} \qquad \dot{\delta} = \left\langle \frac{< \sigma_{11} > /(1-D) - R_n}{K_n} \right\rangle^{n_n} \tag{105}$$

$$\text{Sliding:} \qquad \dot{\gamma} = \left\langle \frac{|\tau|/(1-D) - R_t}{K_t} \right\rangle^{n_t} sign(\tau) \tag{106}$$

- Damage evolution

$$\text{Critical variable:} \qquad \sigma_D = \sqrt{\sigma^2 + \beta\tau^2} \tag{107}$$

$$\text{Damage evolution:} \qquad \dot{D} = \left\langle \frac{\sigma_D - R_D}{A} \right\rangle^{r} (1-D)^{-k} \tag{108}$$

Figure 24 illustrates typical results obtained with the 2D aggregate of Fig.23. Maps (a-c) illustrate the intergranular crack propagation which is obtained by solving the mechanical problem only. In this case, cracks initiate everywhere in the material element, and the propagation is a mixture of crack growth and coalescence. For maps (d-f), iodine is introduced at the left edge of the mesh, so that the grain boundaries are more critical at this point. One main crack propagates. It is worth noting that crack branching can be seen for each triple point. The crack propagation rate is in good agreement with the experimental results. 3D simulations are still necessary, in order to capture the real aspect of the cracks in space and avoid the artefacts related to 2D computations (a crack is too critical, since it crosses the whole mesh in the third direction !).

Fatigue–fretting is an other case where the description of the material by a crystal plasticity model can be very helpful. Figure 25 shows the numerical model developed to simulate a disk–plan test, both of them being made of a titanium alloy. Grain size is around 50 μm, so that the stress gradient under the surface corresponds to a few grains in depth (Dick and Cailletaud, 2006).

A typical effect related to the use of crystal plasticity instead of a traditional J_2 model is shown in Fig.26: after some fretting cycles, local heterogeneities propagate much more in crystal plasticity than with a von Mises material; eventually, local ratchetting can be observed, so that, starting from the analytical value of the stress profile computed in elasticity (Hertz solution), von Mises model leads to a smoothly modified curve, but crystal plasticity produces a highly heterogeneous field, in which the maximum stress can locally increase. This may have important consequence for the response of subsequent damage models, trying to predict either wear of crack initiation.

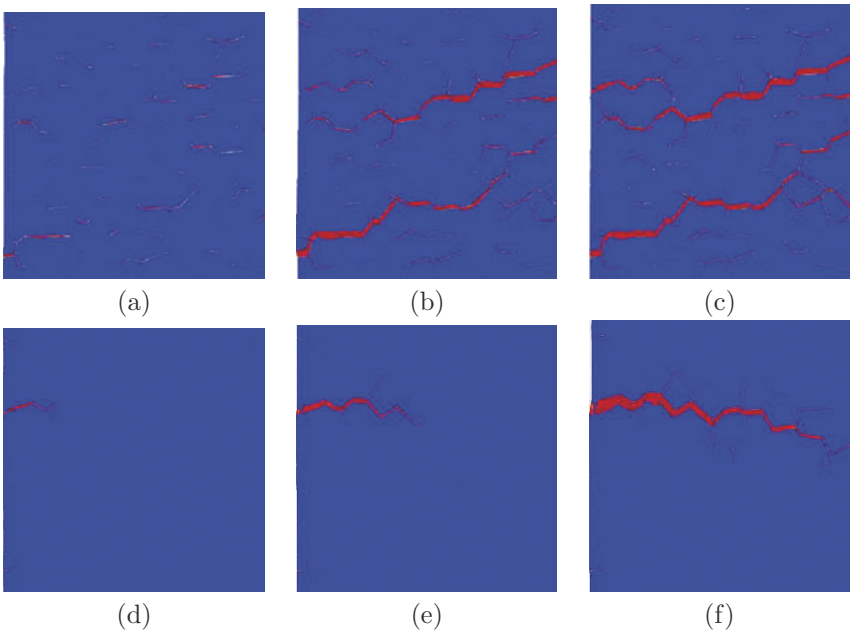


Figure 24. Crack propagation predicted by intergranular damage evolution : (a,b,c) non-coupled case; (d,e,f) coupled case

5 Uniform field models

Developing models at level II can be considered as an interesting alternative, that allows the user to represent the constitutive equations of each phase, and avoids unjustified mixture rules. The resulting models are more complex than the models on the macroscale, but they are still manageable, either in drivers of constitutive equations or even in finite element codes.

5.1 Yield surfaces

This section illustrates the transition from a single crystal surface, as shown in Fig.6, that results from a collection of hyperplanes in the stress space, to surfaces for directionnally solidified (DS) or polycrystalline materials. An uniform elasticity is assumed for each grain, so that the local stresses are equal to the macroscopic stress in the elastic regime. In DS material, grains are supposed to have the same (001) axis, and random ori-

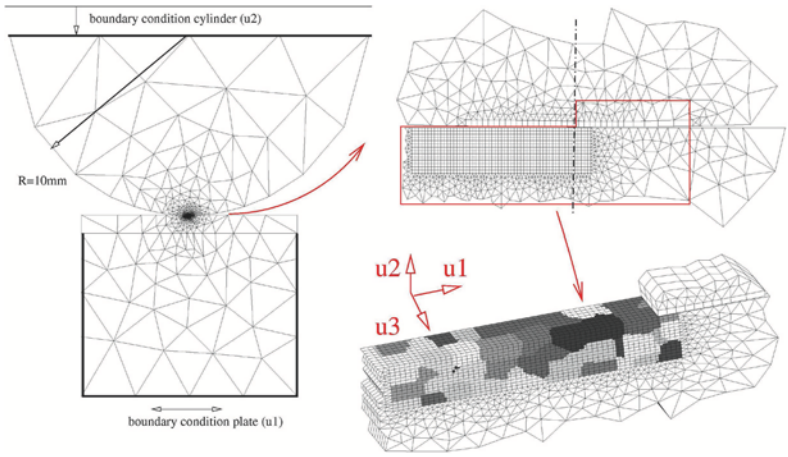


Figure 25. Geometry of the mesh used for the numerical simulation and grain morphology

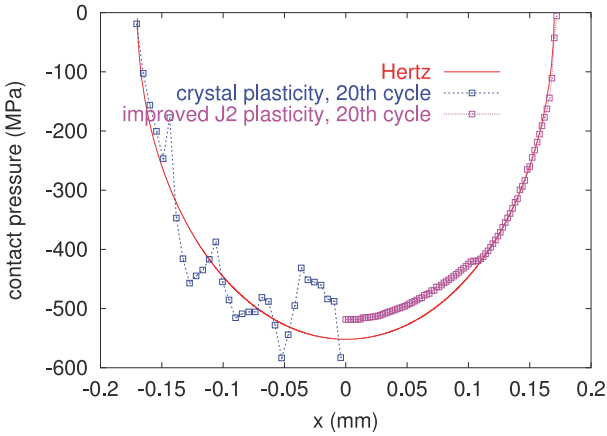
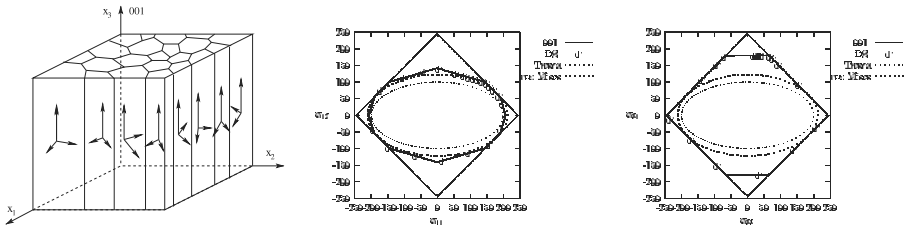


Figure 26. Comparison if the evolution of the contact pressure for J2 plasticity and for a crystallographic model



(a) (b) (c)

Figure 27. Yield surface for a DS material: (a) geometry of the aggregate, (b) yield surface in the σ_{11}–σ_{12} plane, (c) yield surface in the σ_{33}–σ_{31} plane

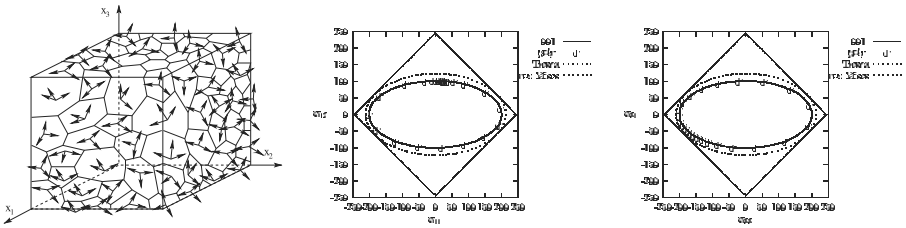
entation in the plane perpendicular to (001). The loading axes (x_1,x_2,x_3) coincide with the crystallographic axes. The resulting model for such an aggregate still presents crystal characteristics in the plane σ_{33}–σ_{31} (Fig.27c), due to the reduced number of slip systems available for shear. On the other hand, the shape of the criterion in σ_{11}–σ_{12} plane (Fig.27b) looks like a macroscopic isotropic model. The present figure was plotted with only three orientations, whose axis (100) makes an angle of 0°, 30°, 60° with x_1. The references to Tresca and von Mises are built by supposing that the critical shear stress for Schmid's law in each single crystal is taken as the critical shear for Tresca (the difference between the two extreme eignestresses), and as the octahedral shear for von Mises (the square root of the half of the sum of the squares of the three differences between the eigenstresses). A full characterization of the elasto-plastic behaviour of this material can be found elsewhere (Sai et al., 2006).

On the other hand, if the construction is made for a polycrystal with random orientations (Fig.28), the resulting criterion is isotropic, and reaches exactly Tresca criterion (the present plot uses 1000 grains).

5.2 Scale transition rules

As observed previously in FE computations, plastic or viscoplastic flows are heterogeneous in the aggregates. The plots of Fig.20 demonstrate that neither stresses nor strains are uniform from one grain to the other. The rôle of the scale transition rule is to provide an estimation of the stress tensor in each grain. The purpose of this section is to make a quick overview of the various rules and of their physical meaning. The following notations will be used:

- Stress in phase g, macroscopic stress: $\underset{\sim}{\boldsymbol{\sigma}}^g$, $\underset{\sim}{\boldsymbol{\sigma}} = \sum_g f^g \underset{\sim}{\boldsymbol{\sigma}}^g$



(a) (b) (c)

Figure 28. Yield surface for an isotropic polycrystalline material: (a) geometry of the aggregate, (b) yield surface in the σ_{11}–σ_{12} plane, (c) yield surface in the σ_{33}–σ_{31} plane

- Strain in phase g, macroscopic strain: $\underset{\sim}{\varepsilon}^g$, $\underset{\sim}{E} = \sum_g f^g \underset{\sim}{\varepsilon}^g$
- For an uniform local elasticity, the macroscopic plastic strain is also the average of the local plastic strains: $\underset{\sim}{\varepsilon}^p = \sum_g f^g \underset{\sim}{\varepsilon}^{pg}$

More detailed comments are given elsewhere Besson et al. (2001). The most popular models correspond to the following assumptions:

- Static, *uniform stress*, $\underset{\sim}{\sigma}^g = \underset{\sim}{\Sigma}$
- From Taylor to Kröner
  - Taylor (Taylor, 1938), *uniform plastic strain*, $\underset{\sim}{\varepsilon}^{pg} = \underset{\sim}{E}^p$
  - Lin–Taylor (Lin, 1957), *uniform total strain*, $\underset{\sim}{\varepsilon}^g = \underset{\sim}{E}$
  - Kröner (Kröner, 1971), *elastic accommodation*

$$\underset{\sim}{\sigma}^g = \underset{\sim}{\Sigma} + \underset{\approx}{\Lambda} : (\underset{\approx}{I} - \underset{\approx}{S})(\underset{\sim}{\varepsilon}^p - \underset{\sim}{\varepsilon}^{pg})$$

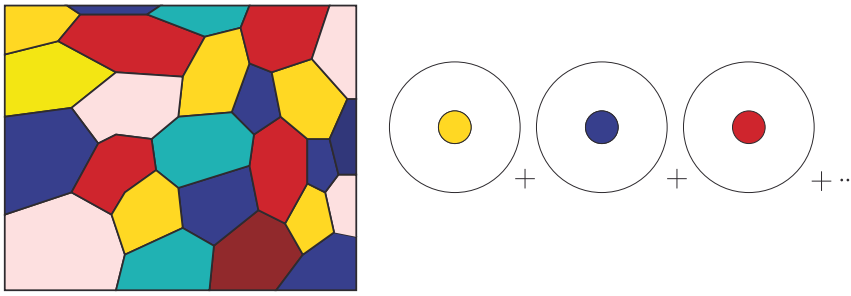
    where $\underset{\approx}{\Lambda}$ is the elastic tensor and $\underset{\approx}{S}$ the Eshelby tensor (Mura, 1987).
  - Kröner for an isotropic material (elastic and plastic) and pressure insensitive plasticity

$$\underset{\sim}{\sigma}^g = \underset{\sim}{\Sigma} + \mu\alpha(\underset{\sim}{\varepsilon}^p - \underset{\sim}{\varepsilon}^{pg}) \quad \text{with} \quad \alpha = \frac{2(7-5\nu)}{15(1-\nu)}$$

- Tangent and secant approximations (self-consistent framework)
  - Hill (Hill, 1965), *elastoplastic accommodation*

$$\dot{\underset{\sim}{\sigma}}^g = \dot{\underset{\sim}{\Sigma}} + \underset{\approx}{L}^\star : (\dot{\underset{\sim}{E}} - \dot{\underset{\sim}{\varepsilon}}^g)$$

    where $\underset{\approx}{L}^\star$ is a tangent accommodation tensor.



(a) (b)

Figure 29. An illustration of polycrystal modelling: (a) actual aggregate, (b) collection of auxiliary problems

– Berveiller–Zaoui (Berveiller and Zaoui, 1979) estimation, with α varying typically between 1 and 0.001 (valid for the particular case of isotropic elasticity, pressure insensitive plasticity and spherical inclusions), as shown later in equation (113)

$$\underset{\sim}{\boldsymbol{\sigma}}^g = \underset{\sim}{\boldsymbol{\sigma}} + \mu\alpha(\underset{\sim}{\boldsymbol{\varepsilon}}^p - \underset{\sim}{\boldsymbol{\varepsilon}}^{pg}) \quad \text{with} \quad \alpha =$$

- Viscous and viscoplastic scheme
  - Budianski, Hutchinson, Molinari... (Hutchinson, 1966; Molinari et al., 1987)
  - Translated fields (Sabar et al., 2002),

$$\dot{\underset{\sim}{\boldsymbol{\sigma}}}^g = \dot{\underset{\sim}{\boldsymbol{\sigma}}} + 2\mu(1-\beta)(\frac{5\eta}{3\eta + 2\eta^g}\dot{\underset{\sim}{\boldsymbol{\varepsilon}}}^v - \dot{\underset{\sim}{\boldsymbol{\varepsilon}}}^{vg})$$

- Parametric scale transition rule
  - Cailletaud, Pilvin, *β–model* , that will be developed below.

The way the models are obtained in the self-consistent framework is illustrated in Fig.29, which shows that all the realizations of a given grain orientation in the aggregate are put together into a single sphere, that is embedded in the homogeneous medium. The solution of the problem is then obtained as a collection of auxiliary problems, leading to an implicit scheme, since the behaviour of the homogeneous medium is not known. The calculation result will have the shape:

$$\dot{\underset{\sim}{\boldsymbol{\sigma}}} = \dot{\underset{\sim}{\boldsymbol{\Sigma}}} + \underset{\approx}{\boldsymbol{L}}^\star : (\dot{\underset{\sim}{\boldsymbol{E}}} - \dot{\underset{\sim}{\boldsymbol{\varepsilon}}}) \tag{109}$$

with $\underset{\sim}{\underset{\sim}{\boldsymbol{L}}}^{\star}$, accommodation tensor:

$$\underset{\sim}{\underset{\sim}{\boldsymbol{L}}}^{\star} = \underset{\sim}{\underset{\sim}{\boldsymbol{L}}}^{eff} : (\underset{\sim}{\underset{\sim}{\boldsymbol{S}}}^{-1} - \underset{\sim}{\underset{\sim}{\boldsymbol{I}}}) \tag{110}$$

The value of $\underset{\sim}{\underset{\sim}{\boldsymbol{L}}}^{eff}$ is obtained by solving the following implicit equation (Hill, 1965):

$$\underset{\sim}{\underset{\sim}{\boldsymbol{L}}}^{eff} = < \underset{\sim}{\underset{\sim}{\boldsymbol{L}}} : (\underset{\sim}{\underset{\sim}{\boldsymbol{L}}} + \underset{\sim}{\underset{\sim}{\boldsymbol{L}}}^{\star})^{-1} : (\underset{\sim}{\underset{\sim}{\boldsymbol{L}}}^{eff} + \underset{\sim}{\underset{\sim}{\boldsymbol{L}}}^{\star}) > \tag{111}$$

For the classical case of uniform elasticity, incompressible plastic flow, radial loading path, a simplified expression can be written, with μ', actual shear modulus, and $\beta' = 2(4-5\nu')/15(1-\nu')$

$$\dot{\underset{\sim}{\boldsymbol{\sigma}}}^{g} = \dot{\underset{\sim}{\boldsymbol{\sigma}}} + 2\mu \frac{\mu'(1-\beta')}{\beta'\mu + (1-\beta')\mu'} (\dot{\underset{\sim}{\boldsymbol{\varepsilon}}}^{p} - \dot{\underset{\sim}{\boldsymbol{\varepsilon}}}^{pg}) \tag{112}$$

For pure tension, assuming $\nu = 1/2$ and introducing the *secant* modulus $H = \sigma/\varepsilon^p$, the model can be simplified:

$$\dot{\sigma}_g = \dot{\sigma} + \frac{\mu H}{H + 2\mu} (\dot{\varepsilon}^p - \dot{\varepsilon}^{pg}) \tag{113}$$

At the onset of plastic flow, H tends to infinity, so that Kröner's rule is recovered. The accommodation factor $C = (\sigma^g - \sigma)/(\varepsilon^p - \varepsilon^{pg})$ decreases when plastic strain increases. This expression is easy to manage, nevertheless its domain of application is limited. This is why the β–rule has been proposed (Cailletaud, 1987; Pilvin and Cailletaud, 1990; Cailletaud and Pilvin, 1994).

The goal of this model is to represent the plastic accommodation, like Hill's approach, but in an explicit formulation. Instead of having a non linear multiplicative term and a linear dependency from the difference between local and global plastic strain, like in Berveiller–Zaoui's approach, it was decided to keep the linear multiplicative term (like in Kröner's formulation), and to introduce a new phenomenological variable instead of plastic strain to account for non linear accommodation.

The local stress decreases when the grain becomes more plastic than the matrix, so that a typical shape of the model is:

$$\underset{\sim}{\boldsymbol{\sigma}}^{g} = \underset{\sim}{\boldsymbol{\sigma}} + C \left(\underset{\sim}{\boldsymbol{\beta}} - \underset{\sim}{\boldsymbol{\beta}}^{g} \right) \tag{114}$$

with

$$\underset{\sim}{\boldsymbol{\beta}} = \sum_{g} f_g \underset{\sim}{\boldsymbol{\beta}}^{g} \tag{115}$$

f_g is the volume fraction of phase g, $\underset{\sim}{\boldsymbol{\beta}}^g$ characterizes the state of redistribution.

The new interphase accommodation variables $\boldsymbol{\beta}^g$ follow a kinematic evolution rule. Two evolution laws have been tested in the past.

$$\text{- rule 1: } \dot{\underset{\sim}{\boldsymbol{\beta}}}^g = \dot{\underset{\approx}{\boldsymbol{\varepsilon}}}^g - D\dot{\varepsilon}^g_{eq}\underset{\sim}{\boldsymbol{\beta}}^g \tag{116}$$

$$\text{- rule 2: } \dot{\underset{\sim}{\boldsymbol{\beta}}}^g = \dot{\underset{\approx}{\boldsymbol{\varepsilon}}}^g - D\dot{\varepsilon}^g_{eq}(\underset{\sim}{\boldsymbol{\beta}}^g - \delta\underset{\approx}{\boldsymbol{\varepsilon}}^g) \tag{117}$$

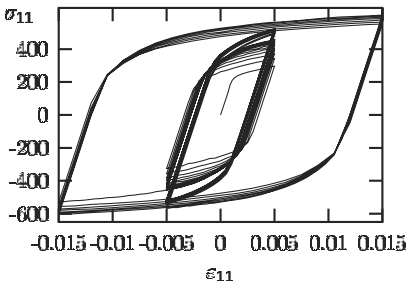
In the two rules, the equivalent strain rate in the fading memory term can be replaced by the sum of the slip rates on all the slip systems. Rule 2 is nothing but a combination of a linear an a non linear term; a quick manipulation shows that the expression can be recovered by assuming $\underset{\sim}{\boldsymbol{\beta}} = \delta\underset{\sim}{\boldsymbol{\beta}}^1 + (1-\delta)\underset{\sim}{\boldsymbol{\beta}}^2$ and:

$$\dot{\underset{\sim}{\boldsymbol{\beta}}}^1 = \dot{\underset{\approx}{\boldsymbol{\varepsilon}}}^p \quad \text{and} \quad \dot{\underset{\sim}{\boldsymbol{\beta}}}^2 = \dot{\underset{\approx}{\boldsymbol{\varepsilon}}}^p - D\underset{\sim}{\boldsymbol{\beta}}^2\ \dot{\varepsilon}^g_{eq}$$

The parameters C, D and δ are not free material coefficients, but *scale transition parameters*. Kröner's model can be found as a particular case by assuming that $D = \delta = 0$ and $C = \mu$. This is consistent with the fact that, at the onset of plasticity, the matrix behaviour deviates only a little from elastic behaviour, and an elastic accommodation is the proper response of the material. On the other hand, the rest of the parameters can be calibrated by means of more complex approaches, namely finite elements. For this purpose, a series of realistic aggregates (in terms of morphology and crystal orientations) must be selected. A large number of aggregates must be computed, to have a chance to place each grain orientation in an environment which is representative of all the grains (this is not the case for one unique computation, since, statistically, each real grain has between 15 and 25 neighbours only). A post-processing allows then to compute the macroscopic stress–strain curve, and the average values in each crystallographic phase. The identification of the coefficients can then be made by solving an inverse problem. The goal is to fit, not only the macroscopic curve, but also the local stresses and strains with the β-model. Note that this information is rather rich, so that it may be difficult to reproduce the local stresses (even if is it rather easy to get the global behaviour).

5.3 Complex paths

When coupled with a crystal plasticity model in each grain, the β-rule has been found able to provide interesting effects, which were observed in the litterature, but that cannot be modelled by macroscopic approaches, except by adding specific variables. Two classical examples are shown here: figure 30 illustrates the example of the so called *memory effect*, that is

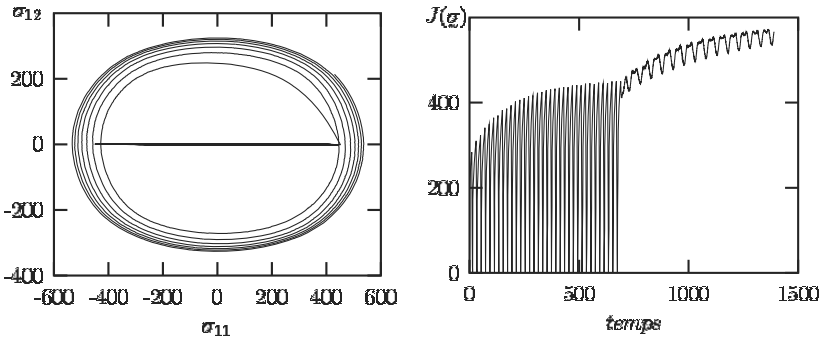


parameter	value
R_0	100
Q	50
b	10
c	5000
d	100
h_i	1,i

Figure 30. Illustration of the strain memory effect coming from the slip system activation

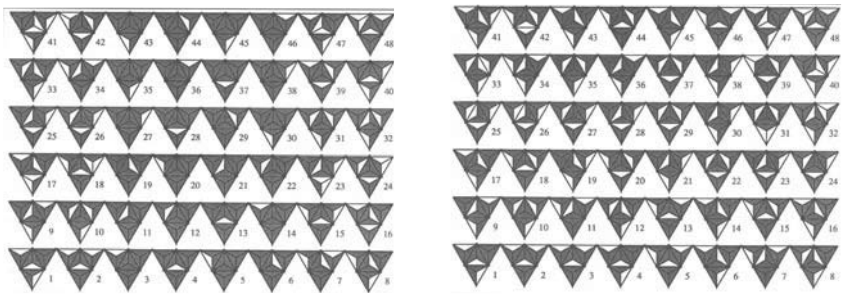
present for instance in austenitic stainless steels (Chaboche et al., 1979). Three successive loading levels are applied to a tension–compression specimen, the first amplitude is 0.5%, the second one is 1.5%, and the third one 0.5% again. The maximum stress at the third level is much bigger than the stress at the first one, even if the range is the same. The model reproduces correctly this effect due to a more intense slip system activation at the second level.

An extra slip system activation is also present for non proportional loading. This type of load (sinusoidal imposed tension and shear with a phase lag of 90°) is know to produce maximum stresses on each components much larger than the corresponding load type for a proportional loading. Figure 31 shows the response in the tension–shear stress plane, and the history of the von Mises stress as a function of time. The illustration of the slip system activity for the same test is shown in Fig.32, by comparison with the result of a tensile test. The material has a FCC microstructure, and the slip systems are represented on unfolded Thomson tetrahedra (four planes represented by triangles, with three subtriangles in each of them figuring the slip systems). The number of active slip system is 2.17 in average for a tension at 1.5%, meanwhile the number is 3.25 for the non proportional loading at the same equivalent strain. All these effects are discussed elsewhere Cailletaud (1992).



(a) (b)

Figure 31. Extra-hardening due to non-proportional loading: (a) response in the stress plane, (b) evolution of the equivalent stress



(a) Average number per grain: 2.17 (b) Average number per grain: 3.25

Figure 32. Illustration of the local behaviour of a FCC aggregate: active systems are in white (a) case of a proportional loading, (b) case of a non proportional loading

6 Conclusion and perspectives

Classical crystal plasticity is now a mature theory. The material parameters have been identified for many materials, so that the models can be used for a large number of academic or industrial applications. The goal of the present section was then to summarize a few possibilities of the models, with a special view toward the *operational* aspect, that is their practical use. A series of related subjects are not considered here, but would be of interest for continuing the discussion.

One of the main extension of the classical framework is proposed by the theory of generalized continua. There is an profuse litterature on the subject. The reader can check for instance a recent synthesis in (Papenfuss and Forest, 2006). These extended theories have already been applied to single crystal (see for instance Forest et al. (2000, 2002)). This new type of approach allows the user to take into account size effect, that is specially meaningfull for materials that present small grains or specific heterogeneous microstructures (like γ–γ').

The second type of connection is to be made with other classes of approaches. For collecting information at the scale of the grain, or, generally speaking, on submicronic sizes, the relevant theories do not refer to continuous, but to discrete modeling frameworks. This is the case of discrete dynamics of dislocations (DDD), and of molecular dynamics (MD), which have both their own space and time scales. In the future, the progress of these approaches will allow to provide inputs for the higher scales: one will *bridge the lenghtscales* by taking information for DDD from MD, and for crystal plasticity from DDD (see the rest of the present course).

Finally, a previously unattained chance of progress is now given to the theories devoted to material modeling by the advance of experimental tools and methods. Field measurements are now possible, with a very good definition (Musienko et al., 2007; Kempf et al., 2007). Using powerfull beams, experimentalists will be able to provide data like 3D-microstructure shape and crack population inside the material. Having in hand local data and local material modeling, one will be able to develop new classes of damage models, based on real mechanisms at the microstructure level.

Bibliography

L. Anand and M. Kothari. A computational procedure for rate–independent crystal plasticity. *J. Mech. Phys. Sol.*, 44:525–558, 1996.

R.J. Asaro. Micromechanics of crystals and polycrystals. *Advances in Appl. Mech.*, 23:1–115, 1983a.

R.J. Asaro. Crystal plasticity. *J. of Applied Mechanics*, 50:921–934, 1983b.

R.J. Asaro and A. Needleman. Texture development and strain hardening in rate dependent polycrystals. *Acta Metall.*, 33:923–953, 1985.

R.J. Asaro and J.R. Rice. Strain localization in ductile single crystals. *J. Mech. Phys. Sol.*, 25:309–338, 1977.

F. Barbe, L. Decker, D. Jeulin, and G. Cailletaud. Intergranular and intragranular behavior of polycrystalline aggregates. Part I: FE model. *Int. J. of Plasticity*, 17(4):513–536, 2001a.

F. Barbe, S. Forest, and G. Cailletaud. Intergranular and intragranular behavior of polycrystalline aggregates. Part II: Results. *Int. J. of Plasticity*, 17(4):537–566, 2001b.

M. Berveiller and A. Zaoui. An extension of the self–consistent scheme to plastically flowing polycrystal. *J. Mech. Phys. Sol.*, 26:325–344, 1979.

J. Besson, R. Le Riche, R. Foerch, and G. Cailletaud. Object–oriented programming applied to the finite element method. Part II: Application to material behaviors. *Revue Européenne des Éléments Finis*, 7(5):567–588, 1998.

J. Besson, G. Cailletaud, J.-L. Chaboche, and S. Forest. *Mécanique non–linéaire des matériaux*. Hermès, 2001.

J.F.W. Bishop and R. Hill. A theoretical derivation of the plastic properties of a polycrystalline face–centered metal. *Philosophical Magasine*, 42:414–427, 1951.

S. Bugat, J. Besson, and A. Pineau. Micromechanical modeling of the behavior of duplex stainless steels. *Computational Materials Science*, 16:158–166, 1999.

E. Busso, M. Toulios, and G. Cailletaud. Constitutive material formulations and advanced life assessment methods for single crystal gas turbine blades. In J. Lecomte-Beckers, M. Carton, F. Schubert, and Schriften des Forschung Zentrum Jülich Ennis, P.J., editors, *7th Conf. Materials for advanced power engineering, Liège, 30 sept.–2 oct. 2002*, pages 1175–1185, 2003.

E.P. Busso and F.A. McClintock. A dislocation mechanics-based crystallographic model of a B2-type intermetallic alloy. *Int. J. of Plasticity*, 12: 1–28, 1996.

G. Cailletaud. A micromechanical approach to inelastic behaviour of metals. *Int. J. of Plasticity*, 8:55–73, 1992.

G. Cailletaud. *Une approche micromécanique phénoménologique du comportement inélastique des métaux*. PhD thesis, Université Pierre et Marie Curie, Paris 6, 1987.

G. Cailletaud and P. Pilvin. Utilisation de modèles polycristallins pour le calcul par éléments finis. *Revue Européenne des Éléments Finis*, 3(4): 515–541, 1994.

G. Cailletaud, J.-L. Chaboche, S. Forest, and L. Rémy. On the design of single crystal turbine blades. *Revue de Métallurgie*, février 2003:165–172, 2001.

G. Cailletaud, O. Diard, and A. Musienko. Damage, opening and sliding of grain boundaries. In S. Ahzi, M.A. Khaleel, H.M. Zbib, M.A. Zikry, B. LaMatina, and M. Cherkaoui, editors, *Multiscale Modelling and Characterisation of Elastic-Inelastic Behaviour of Engineering Materials (Proc.of the IUTAM Symposium on Multiscale Modelling of Engng. Materials. Marrakech, Oct. 2002)*, pages 149–156. Kluwer, 2004.

J.-M. Cardona. *Comportement et durée de vie des pièces multiperforées : application aux aubes de turbine*. PhD thesis, École Nationale Supérieure des Mines de Paris, 2000.

J.-L. Chaboche, K. Dang Van, and G. Cordier. Modelisation of the strain memory effect on the cyclic hardening of 316 stainless steel. In *Proc. SMIRT 5*, volume L, page L11/3, Berlin, 1979.

G.Y. Chin and W.L. Mammel. Generalization and equivalence of the minimum work (Taylor) and maximum work (Bishop–Hill) principles for crystal plasticity. *Trans. of the Met. Soc. of AIME*, 245:1211–1214, 1969.

M. Clavel. *Fatigue plastique et fissuration de deux alliages durcis par des précipités cohérents. Etude comparative des mécanismes*. PhD thesis, ENSMA, 1980.

B.D. Coleman. Thermodynamics of materials with memory. *Arch. Rat. Mech. Anal.*, 17:1–46, 1964.

A.M. Cuitino and M. Ortiz. Computational modelling of single crystals. *Modelling Simul. Mater. Sci. Eng.*, 1:225–263, 1992.

F. Delaire, J.L. Raphanel, and C. Rey. Plastic heterogeneities of a copper multicrystal deformed in uniaxial tension: experimental study and finite element simulations. *Acta Metall.*, 48:1075–1087, 2000.

B. Devincre, L. Kubin, and T. Hoc. Physical analyses of crystal plasticity by dd simulations. *Scripta Metallurgica and Materiala*, 54:741–746, 2006.

O. Diard, S. Leclercq, G. Rousselier, and G. Cailletaud. Distribution of normal stress at grain boundaries in multicrystals: application to an intergranular damage modeling. *Computational Materials Science*, 25: 73–84, 2002.

T. Dick and G. Cailletaud. Fretting modelling with a crystal plasticity model of Ti6Al4V. *Computational Materials Science*, 38:113–125, 2006.

J.-J. Engel. *Modélisation du comportement global des métaux en plasticité et en viscoplasticité*. PhD thesis, École Nationale Supérieure des Mines de Paris, 1978.

L.P. Evers, D.M. Parks, W.A.M. Brekelmans, and M.G.D Geers. Crystal plasticity model with enhanced hardening by geometrically necessary dislocation accumulation. *J. Mech. Phys. Sol.*, 50:2403–2424, 2002.

S. Flouriot, S. Forest, G. Cailletaud, A. Köster, L. Rémy, B. Burgardt, V. Gros, and J. Mosset, S. Delautre. Strain localization at the crack tip in single crstal ct specimens under monotonous loading: 3d finite element analyses and application to Nickel-base superalloys. *Int. J. Frac*, 124: 43–77, 2003.

R. Foerch, J. Besson, G. Cailletaud, and P. Pilvin. Polymorphic constitutive equations in finite element codes. *Comp. Meth. Appl. Mech. Engng*, 141: 355–372, 1997.

S. Forest, F. Barbe, and G. Cailletaud. Cosserat modelling of size effects in the mechanical behaviour of polycrystals and multiphase materials. *Int. J. Sol. Structures*, 37:7105–7126, 2000.

S. Forest, R. Sievert, and E.C. Aifantis. Strain gradient crystal plasticity: thermomechanical formulations and applications. *J. of the Mechanical Behaviour of Materials*, 13:219–232, 2002.

P. Franciosi. The concepts of latent hardening and strain hardening in metallic single crystals. *Acta Metall.*, 33:1601–1612, 1985a.

P. Franciosi. On flow and work hardening expression correlations in metallic single crystal plasticity. *Revue de Physique Appliquée*, 33:1601–1612, 1985b.

P. Germain, Q.S. Nguyen, and P. Suquet. Continuum thermodynamics. *J. of Applied Mechanics*, 50:1010–1020, 1983.

B. Halphen and Q.S. Nguyen. Sur les matériaux standards généralisés. *Journal de Mécanique*, 14:39–63, 1975.

F. Hanriot. *Étude du comportement du superalliage AM1 sous sollicitations cycliques*. PhD thesis, École Nationale Supérieure des Mines de Paris, 1993.

F. Hanriot, G. Cailletaud, and L. Rémy. Mechanical behavior of a nickel-base superallot single crystal. In A.D. Freed and K.P. Walker, editors, *High Temperature Constitutive Modeling – Theory and Application*. ASME, 1991.

K.S. Havner. *Finite Plastic Deformation of Crystalline Solids*. Cambridge University Press, 1992.

R. Hill. A self–consistent mechanics of composite materials. *J. Mech. Phys. Sol.*, 13:213–222, 1965.

R. Hill. Generalized constitutive relations for incremental deformation of metal crystals by multislip. *J. Mech. Phys. Sol.*, 14:95–102, 1966.

R. Hill and J.R. Rice. Constitutive analysis of elastic–plastic crystals at arbitrary strains. *J. Mech. Phys. Sol.*, 20:401–413, 1972.

J.W. Hutchinson. Bounds ans self-consistent estimates for creep of polycrystalline materials. *Proc. Royal Soc. London*, A348:101–127, 1966.

E.H. Jordan and K.P. Walker. Biaxial constitutive modelling and testing of a single crystal superalloy at elevated temperature. In K. Miller, editor, *Conf. on Multiaxial Fatigue, Sheffield*, 1985.

S.R. Kalidindi, C.A. Bronkhorst, and L. Anand. Crystallographic texture evolution in bulk deformation processing of FCC metals. *J. Mech. Phys. Sol.*, 40:536–569, 1992.

D. Kempf, V. Vignal, G. Cailletaud, R. Oltra, J.C. Weeber, and E. Finot. High spatial resolution strain measurements at the surface of duplex stainless steels. *Philosophical Magasine*, 87:1379–1399, 2007.

D.J. Kim, T.W. Ku, and B.S. Kang. Finite element analysis of micro-rolling using grain and grain boundary elements. *J. of Materials Processing Technology*, 130–131:456–461, 2002.

U.F. Kocks. The relation between polycrystal deformation and single-crystal deformation. *Metallurgical Transaction*, 1:1121–1143, 1970.

U.F. Kocks and T.J. Brown. Latent hardening in aluminium. *Acta Metall.*, 14:87–98, 1966.

W.T. Koiter. *General theorems for elastic–plastic solids*, volume 6, pages 167–221. North–Holland Publishing Company, 1960.

E. Kröner, editor. *Statistical continuum mechanics*, volume 92 of *CISM Courses and lectures*. Springer Verlag, 1971.

T. Kubo, Y. Wakashima, K. Amano, , and M. Nagai. Effects of crystallographic orientation on plastic deformation and initiation of zirconium alloys. *J. Nucl. Mat.*, 132(1):1–9, 1985.

J. Lemaitre and J.-L. Chaboche. *Mechanics of Solid Materials*. Cambridge University Press, Cambridge, U.K., 1990.

T.H. Lin. Analysis of elastic and plastic strain in FCC crystal. *J. Mech. Phys. Sol.*, 5:143, 1957.

J. Mandel. Une généralisation de la théorie de la plasticité de W.T. Koiter. *Int. J. Sol. Structures*, 1:273–295, 1965.

J. Mandel. *Plasticité classique et viscoplasticité*, volume 97. Cours du CISM, 1971.

R.D. McGinty and D.L. McDowell. A semi-implicit integration scheme for rate independent finite crystal plasticity. *Int. J. of Plasticity*, 22:996–1025, 2006.

L. Méric and G. Cailletaud. Single crystal modeling for structural calculations. Part 2: Finite element implementation. *J. of Engng. Mat. Technol.*, 113:171–182, 1991.

L. Méric, P. Poubanne, and G. Cailletaud. Single crystal modeling for structural calculations. Part 1: Model presentation. *J. of Engng. Mat. Technol.*, 113:162–170, 1991.

L. Méric, G. Cailletaud, and M. Gaspérini. FE calculations of copper bicrystal specimens submitted to tension–compression tests. *Acta Metall.*, 42 (3):921–935, 1994.

C. Miehe, J. Schröder, and J. Schotte. Computational homogenization analysis in finite plasticity simulation of texture development in polycrystalline materials. *Comp. Meth. Appl. Mech. Engng*, 171:387–418, 1999.

D.P. Mika and P.R Dawson. Polycrystal plasticity modeling of intracrystalline boundary textures. *Acta Metall.*, 47:1355–1369, 1999.

A. Molinari, G.R. Canova, and S. Ahzi. A self–consistent approach to the large deformation polycrystal viscoplasticity. *Acta Metall.*, 35:2983–2994, 1987.

T. Mura. *Micromechanics of defects in solids.* Martinus Nijhoff, 1987.

A. Musienko, A. Tatschl, K. Schmidegg, O. Kolednik, R. Pippan, and G. Cailletaud. 3D finite element simulation of a polycrystalline copper specimen. *Acta Mat.*, 55:4121–4136, 2007.

D. Nouailhas and G. Cailletaud. Tension-torsion behavior of single-crystal superalloys: experiment and finite element analysis. *Int. J. of Plasticity*, 11(4):451–470, 1995.

C. Papenfuss and S. Forest. Thermodynamical frameworks for higher grade material theories with internal variables or additional degrees of freedom. *J. of Non-Equilibrium Thermodynamics*, 31:319–353, 2006.

D. Pierce, R.J. Asaro, and A. Needleman. Material rate dependence and localized deformation in crystalline solids. *Acta Metall.*, 31:1951, 1985.

P. Pilvin and G. Cailletaud. Intergranular and transgranular hardening in viscoplasticity. In M. Zyczkowski, editor, *Creep in Structures*, volume 4, pages 171–178, 1990.

D. Raabe, M. Sachtleber, Z. Zhao, F. Roters, and S. Zaefferer. Micromechanical and macromechanical effects in grain scale polycrystal plasticity experimentation and simulation. *Acta Metall.*, 49:3433–3441, 1981.

J. Rice. Tensile crack tip fields in elastic-ideally plastic crystals. *Mech. of Materials*, 6:317–335, 1987.

J.R. Rice. On the structure of stress-strain relations for time–dependent plastic deformation in metals. *J. of Applied Mechanics*, 37:728, 1970.

J.R. Rice. Inelastic constitutive relations for solids: An internal variable theory and its application to metal plasticity. *J. Mech. Phys. Sol.*, 19: 433–455, 1971.

H. Sabar, M. Berveiller, V. Favier, and S. Berbenni. A new class of micro–macro models for elastic–viscoplastic heterogeneous materials. *Int. J. Sol. Structures*, 39:3257–3276, 2002.

K. Sai, G. Cailletaud, and S. Forest. Micro-mechanical modeling of the inelastic behavior of directionally solidified materials. *Mech. of Materials*, 38:203–217, 2006.

J. Schröder and C. Miehe. Aspects of computational rate-independent crystal plasticity. *Computational Materials Science*, 9:168–176, 1997.

J.C. Simo and T.R.J. Hughes. *Computational Inelasticity*. Springer Verlag, 1997.

L. Tabourot, M. Fivel, and E. Rauch. Generalised constitutive laws for FCC single crystals. *Material Science and Engineering*, A234-236:639–642, 1997.

G.I. Taylor. Plastic strain in metals. *J. Inst. Metals.*, 62:307–324, 1938.

C. Teodosiu, editor. *Constitutive Modeling of Polycrystalline Metals at Large Strains*, volume 376 of *CISM Courses and lectures*. Springer Verlag, 1997.

Computational homogenization

M.G.D. Geers [*], V.G. Kouznetsova [*†], W.A.M. Brekelmans [*]
[*] Eindhoven University of Technology
Department of Mechanical Engineering Eindhoven
[†] Netherlands Institute for Metals Research

Abstract This part of the CISM course addresses basics and advanced topics on the computational homogenization of the mechanics of highly non-linear solids with (possibly evolving) microstructure under complex non-linear loading conditions. The key components of the computational homogenization scheme, i.e. the formulation of the microstructural boundary value problem and the coupling between the micro and macrolevel based on the averaging theorems, are addressed. The numerical implementation of the framework, particularly the computation of the macroscopic stress tensor and extraction of the macroscopic consistent tangent operator based on the total microstructural stiffness, are treated in detail. The application of the method is illustrated by the simulation of pure bending of porous aluminum. The classical notion of a representative volume element is introduced and the influence of the spatial distribution of heterogeneities on the overall macroscopic behaviour is investigated by comparing the results of multi-scale modelling for regular and random structures. Finally, an extension of the classical computational homogenization scheme to a framework suitable for multi-scale modelling of macroscopic localization and size effects is briefly discussed.

1 Introduction

The past years have been marked by a significant interest in the various length scales that govern the mechanics of materials. The main issue consists in identifying the relationships that bridge those various scales, i.e. multi-scale mechanics. The multi-scale methodology aims to predict, describe, quantify or qualify the 'macroscopic' behaviour of engineering materials through the consistent modelling of the mechanics and physics of the heterogeneous, multi-phase, anisotropic, discrete microstructure. Various techniques have been proposed to contribute to this challenging task. Among them, a large class of homogenization techniques exists, also called

coarse graining in the physics community (Ridderbos, 2002; Nguyen and Ortiz, 2002).

Homogenization techniques were first developed within the framework of elasticity, as an excellent tool to predict the effective or apparent linear elastic properties of heterogeneous materials. Several closed-form homogenization techniques have been proposed in this context, e.g. the Voigt-Reuss-Hill bounds, the Hashin-Shtrikman variational principle, the self-consistent method, etc., see (Nemat-Nasser and Hori, 1993) for an overview. Asymptotic or mathematical homogenization schemes have been used frequently to assess effective properties of elastic heterogeneous materials (Chung et al., 2001; Fish and Chen, 2001). Extensions towards higher-order and nonlocal constitutive equations have been considered as well, e.g. developments including Cosserat media (Forest et al., 2001), couple stress theory (Smyshlyaev and Fleck, 1994), nonlocal effective continua (Drugan and Willis, 1996) or higher-order gradient homogenized elastic materials (Triantafyllidis and Bardenhagen, 1996; Smyshlyaev and Cherednichenko, 2000; Peerlings and Fleck, 2001).

Other interesting approaches towards the analysis of random (physically nonlinear) microstructures (Ponte Castañeda, 1992; Suquet, 1993; Ponte Castañeda, 2002) are the Taylor-Bishop-Hill estimates, several generalizations of self-consistent schemes and asymptotic procedures (Fish et al., 1997). Homogenization of solids in a geometrically and physically nonlinear regime is clearly more cumbersome. Several analyses have been performed on unit cells, from which the parameters in a priori assumed macroscopic constitutive equations can be assessed. Some of them also include higher-order continuum formulations, e.g. Cosserat (van der Sluis et al., 1999) and couple stress media (M. Ostoja-Starzewski, 1999). The added value of these multi-scale methods depends on the accuracy (geometrical, physical, mechanical) with which the microstructure is modelled, as well as the technique that is used to perform the homogenization towards the macroscopic level. Closed-form homogenization towards constitutive material frameworks or effective (or rather apparent) material properties of composites turns out to be really cumbersome if one wishes to take into account more complex physics, geometrical nonlinearities or damage and localization.

Another class of hierarchical techniques are generally known as variational multi-scale methods (Hughes et al., 1998; Garikipati and Hughes, 2000). In here, the weak form of the governing equations is the point of departure, which can be separated in a coarse and a fine scale part on the basis of suitable assumptions on the fine scale field. The key issue resides in the elimination of the fine scale from the obtained formulation. Though promising, the method relies strongly on the assumptions made on

the fine scale and the restrictions that apply to enable the elimination in practice. Well-known fine scale patterns, e.g. displacement discontinuities modelled by Heaviside functions, can be easily implemented. The obtained method then shows considerable similarities with the extended finite element method (Sukumar et al., 2000; Moës and Belytschko, 2002).

Since a few years, substantial progress has been made in the two-scale computational homogenization of complex multi-phase solids. This technique is essentially based on the solution of nested boundary value problems, one for each scale. If attention is focused on the nonlinear characteristics of the material behaviour, this technique proves to be a valuable tool in retrieving the constitutive response. First-order (i.e. including first-order gradients of the macroscopic displacement field only) computational homogenization schemes fit entirely in a standard continuum mechanics framework (principle of local action) and are now readily available in literature (Suquet, 1985; Ghosh et al., 1996; Smit et al., 1998; Miehe et al., 1999b,a; Feyel and Chaboche, 2000; Terada et al., 2000; Ghosh et al., 2001; Kouznetsova et al., 2001; Terada and Kikuchi, 2001; Miehe and Koch, 2002). Main characteristics of this solution strategy are

- The constitutive response at the macro scale is a priori undetermined. No explicit assumptions are required at that level, since the macroscopic constitutive behaviour ensues from the solution of the micro scale boundary value problem.
- The method deals with large displacements (large strains and rotations) in a trivial way under the condition that the microstructural constituents are modelled adequately.
- The different phases in the microstructure can be modelled with arbitrary nonlinear and time-dependent constitutive models.
- The influence of the evolution of the microstructure (as described on the micro-scale) can be assessed directly on the macro-scale.
- The micro scale problem is a classical boundary value problem, for which any appropriate solution strategy can be used, e.g. Finite Element Method (Smit et al., 1998; Feyel and Chaboche, 2000; Terada et al., 2000; Kouznetsova et al., 2001), the Voronoi cell method (Ghosh et al., 1995, 1996), a crystal plasticity framework (Miehe et al., 1999b,a) or numerical methods based on Fast Fourier Transforms (Michel et al., 1999; Moulinec and Suquet, 1998). Galerkin, etc.
- Macroscopic constitutive tangent operators can be obtained from the microscopic overall stiffness tensor through static condensation. Consistency is preserved through this scale transition.

In spite of the huge computational cost of a nested two-scale solution problem, efficiency can be achieved by solving the problem through parallel

computations (Feyel and Chaboche (2000); Kouznetsova (2002)). Another option is selective usage, where non-critical regions are modelled by continuum closed-form homogenized constitutive relations or by the constitutive tangents obtained from the microstructural analysis but kept constant in the elastic domain, while in the critical regions the multi-scale analysis of the microstructure is fully performed (Ghosh et al. (2001)). Despite the required computational efforts the computational homogenization technique has proven to be a valuable tool to establish non-linear micro-macro structure-property relations, especially in the cases where the complexity of the mechanical and geometrical microstructural properties and the evolving character prohibit the use of other homogenization methods.

Cartesian tensors and associated tensor products will be used throughout this chapter, making use of a Cartesian vector basis $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$. Using the Einstein summation rule for repeated indices, the following conventions are used in the notations of well-known tensor products

$$\begin{aligned}
\mathbf{C} &= \vec{a}\,\vec{b} = a_i b_j \,\vec{e}_i\,\vec{e}_j \\
\mathbf{C} &= \mathbf{A}\cdot\mathbf{B} = A_{ij}B_{jk}\,\vec{e}_i\,\vec{e}_k \\
\mathbf{C} &= {}^4\mathbf{A} : \mathbf{B} = A_{ijkl}B_{lk}\,\vec{e}_i\,\vec{e}_j \\
\mathbf{C} &= {}^4\mathbf{A} \vdots {}^4\mathbf{B} = A_{iklm}B_{mlkj}\,\vec{e}_i\,\vec{e}_j
\end{aligned}$$

2 Underlying principles and assumptions

2.1 Scale separation

At the macro-scale, the material is considered as a homogeneous continuum, whereas at the micro level it is generally heterogeneous (the morphology consists of distinguishable components or phases, i.e. inclusions, grains, interfaces, cavities, etc.). This is schematically illustrated in figure 1. The microscopic length scale is much larger than the molecular dimensions

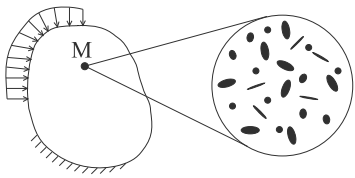


Figure 1. Macroscopic continuum point representation (M) in relation to its underlying heterogeneous microstructure.

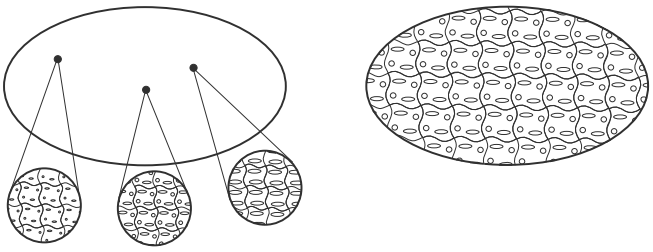
$\ell_{discrete}$, so that a continuum approach is justified for every constituent. At the same time, in the context of the *principle of separation of scales*, the microscopic length scale ℓ_{micro} is assumed to be much smaller than the characteristic length ℓ_{macro} over which the size of the macroscopic loading varies in space, i.e.

$$\ell_{discrete} \lll \ell_{micro} \lll \ell_{macro} \tag{1}$$

Note that it is not the size of the macroscopic domain which is important, but rather the spatial variation of the kinematic fields and stress fields within that domain.

2.2 Local periodicity

Most of the homogenization approaches rely on the assumption of *global periodicity* of the microstructure, implying that the whole macroscopic domain consists of spatially repeated unit cells. In a computational homogenization approach, a more realistic assumption is made, which is commonly denoted by *local periodicity*. According to this assumption, the microstructure can have different morphologies corresponding to different macroscopic points, whereas it repeats itself only in a small vicinity of each individual macroscopic point. The concepts of local and global periodicity are schematically illustrated in figure 2. The assumption of local periodicity adopted in the computational homogenization allows to incorporate a non-uniform distribution of the microstructure at the macroscopic level (e.g. in functionally graded materials). Note that the local periodicity assumption is directly linked to the principle of separation of scales.



(a) local periodicity (b) global periodicity

Figure 2. Local periodicity (a) versus global periodicity (b).

2.3 Homogenization principles

The basic principles of computational homogenization have gradually evolved from the concepts employed in other homogenization methods and well fit into the four-step homogenization scheme established by Suquet Suquet (1985):

1. definition of a microstructural representative volume element (RVE), of which the constitutive behaviour of individual constituents is assumed to be known;
2. formulation of the microscopic boundary conditions from the macroscopic input variables and their application on the RVE (macro-to-micro transition);
3. calculation of the macroscopic output variables from the analysis of the deformed microstructural RVE (micro-to-macro transition);
4. obtaining the (numerical) relation between the macroscopic input and output variables.

The main ideas of the first-order computational homogenization have been established in Suquet (1985); Guedes and Kikuchi. (1990); Terada and Kikuchi (1995); Ghosh et al. (1995, 1996) and further developed and improved in more recent works Smit et al. (1998); Miehe et al. (1999b); Miehe and Koch (2002); Michel et al. (1999); Feyel and Chaboche (2000); Terada and Kikuchi (2001); Ghosh et al. (2001); Kouznetsova et al. (2001); Kouznetsova (2002).

2.4 Computational homogenization scheme

A computational homogenization generally departs from the computation of a macroscopic deformation (gradient) tensor $\mathbf{F}_{\mathrm{M}}$, which is calculated for every material point of the macrostructure (e.g. the integration points within a macroscopic finite element domain). Here and in the following the subscript "M" refers to a macroscopic quantity, while the subscript "m" will denote a microscopic quantity. The deformation tensor $\mathbf{F}_{\mathrm{M}}$ for a macroscopic point is next used to formulate the boundary conditions to be imposed on the RVE that is assigned to this point. Upon the solution of the boundary value problem for the RVE, the macroscopic stress tensor $\mathbf{P}_{\mathrm{M}}$ is obtained by averaging the resulting RVE stress field over the volume of the RVE. As a result, the (numerical) stress-deformation relationship at the macroscopic point is readily available. Additionally, the local macroscopic consistent tangent is extracted from the microstructural stiffness. The entire framework is schematically illustrated in figure 3. The computational homogenization technique defined in this sense, is entirely consistent with the principle of local action in continuum mechanics. Therefore, the response at

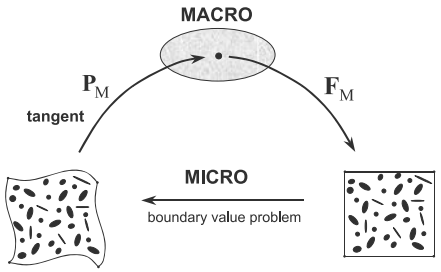


Figure 3. Computational homogenization scheme.

a (macroscopic) material point depends only on the first gradient of the displacement field. This macroscopically local computational homogenization framework may therefore be categorized as a "first-order" approach.

2.5 Kinematically driven multi-scale scheme

The multi-scale procedure outlined in this chapter is "deformation driven". The point of departure is thereby the macroscopic deformation gradient tensor $\mathbf{F}_M$, which is used to determine the stress $\mathbf{P}_M$ and the constitutive tangent, based on the response of the underlying microstructure. A "stress driven" procedure (given a local macroscopic stress, obtain the deformation) is also possible. However, such a procedure does not directly fit into a standard displacement-based finite element framework, which will be here employed to solve the macroscopic boundary value problem. Moreover, in case of large deformations the macroscopic rotational effects have to be added to the stress tensor in order to uniquely determine the deformation gradient tensor, thus complicating the implementation. Therefore, the "stress driven" approach, which is often used in the analysis of single unit cells, is generally not adopted in coupled multi-scale computational homogenization strategies of the type described here.

3 The micro-scale problem

3.1 The representative volume element

The physical and geometrical properties of the microstructure are identified by a representative volume element (RVE) (Hill, 1963; Drugan and Willis, 1996). An example of a typical two-dimensional RVE is depicted in figure 4. The actual choice of the RVE is a rather delicate task. The RVE should be large enough to represent the microstructure, without introducing

non-existing properties (e.g. undesired anisotropy) and at the same time it should be small enough to allow efficient computational modelling. Some issues related to the concept of a representative cell are discussed furtheron, in section 7. Here it is supposed that an appropriate RVE has been already selected. Then the problem on the RVE level can be formulated as a standard problem in quasi-static continuum solid mechanics.

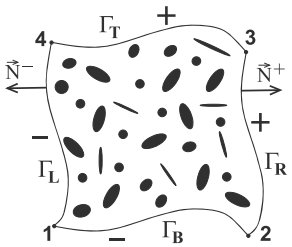


Figure 4. Schematic picture of a typical two-dimensional representative volume element (RVE).

3.2 Micro-scale characterization & equilibrium

The RVE deformation field in a point with the initial position vector $\vec{X}$ (in the reference domain V_0) and the actual position vector $\vec{x}$ (in the current domain V) is described by the microstructural deformation gradient tensor $Fm = (\vec{\nabla}_{0,\mathrm{m}}\vec{x})^c$, where the gradient operator $\vec{\nabla}_{0,\mathrm{m}}$ is taken with respect to the reference microstructural configuration; the superscript "c" indicates conjugation.

The RVE is in a state of equilibrium. This is mathematically expressed through the standard equilibrium equation in terms of the Cauchy stress tensor $\boldsymbol{\sigma}_\mathrm{m}$ or, alternatively, in terms of the first Piola-Kirchhoff stress tensor $\mathbf{P}_\mathrm{m} = \det(\mathbf{F}_\mathrm{m})\boldsymbol{\sigma}_\mathrm{m}{\cdot}(\mathbf{F}_\mathrm{m}^c)^{-1}$ according to (in the absence of body forces)

$$\vec{\nabla}_\mathrm{m}{\cdot}\boldsymbol{\sigma}_\mathrm{m} = \vec{0} \quad \text{in} \quad V, \qquad \text{or} \qquad \vec{\nabla}_{0,\mathrm{m}}{\cdot}\mathbf{P}_\mathrm{m}^c = \vec{0} \quad \text{in} \quad V_0, \tag{2}$$

where $\vec{\nabla}_\mathrm{m}$ is the the gradient operator with respect to the current configuration at the micro-scale.

The mechanical characterization of the microstructural components are described by their constitutive laws, specifying a time and history dependent stress-deformation relationship for every microstructural constituent

$$\boldsymbol{\sigma}_\mathrm{m}^{(\alpha)}(t) = \boldsymbol{\mathcal{F}}_\sigma^{(\alpha)}\{\mathbf{F}_\mathrm{m}^{(\alpha)}(\tau),\ \tau \in [0,t]\}, \ \text{ or } \ \mathbf{P}_\mathrm{m}^{(\alpha)}(t) = \boldsymbol{\mathcal{F}}_P^{(\alpha)}\{\mathbf{F}_\mathrm{m}^{(\alpha)}(\tau),\ \tau \in [0,t]\}, \tag{3}$$

where t denotes the current time; $\alpha = \overline{1, N}$, with N the number of microstructural constituents to be distinguished (e.g. matrix, inclusions, etc.). Note that the knowledge of the separate constitutive laws for each of the individual phases is essential.

3.3 The macro-micro scale transition

The macro-micro scale transition requires a method to impose the macroscopic deformation gradient tensor $\mathbf{F}_{\mathrm{M}}$ or stress tensor $\mathbf{P}_{\mathrm{M}}$ on the microstructural RVE. Classical simplified methods to do this are:

- by imposing that all the microstructural constituents undergo a constant deformation identical to the macroscopic one (the Taylor or Voigt assumption).
- by imposing an identical constant stress (and additionally identical rotation) to all the components (the Sachs or Reuss) assumption).
- by intermediate procedures, where the Taylor and Sachs assumptions are applied only to certain components of the deformation and stress tensors.

These simplified procedures do not satisfy all local static equilibrium and compatibility conditions and generally provide only rough estimates of the overall material properties. They are therefore not well-suited in complex non-linear deformation regimes. The Taylor assumption usually overestimates the overall stiffness, while the Sachs assumption leads to an underestimation of the stiffness. A computational homogenization scheme does enforce local equilibrium and compatibility between phases, and therefore necessitates a different macro-micro scale transition method.

The first-order scheme naturally departs from the classical linearization of the macroscopic nonlinear deformation map, $\vec{x} = \phi(\vec{X})$, applied to a finite material vector $\Delta\vec{x}$ in the deformed state:

$$\Delta\vec{x} \;=\; \mathbf{F}_{\mathrm{M}}{\cdot}\Delta\vec{X} + \mathcal{O}\left(\Delta\vec{X}^2\right), \tag{4}$$

with $\vec{x}$ and $\vec{X}$ associated position vectors in the deformed and reference state, respectively, and in which $\mathbf{F}_{\mathrm{M}} = (\vec{\nabla}_{0,\mathrm{M}}\vec{x})^T$ is the macroscopic deformation gradient tensor. Considering an undeformed volume V_0 of material with its centre positioned at $\vec{X}_c$, permits to write the deformed position of any point of this volume (with respect to the centre of that volume) as the sum of a macroscopic (or coarse scale) and a microscopic (or fine scale) contribution:

$$\vec{x} - \vec{x}_c \;=\; \mathbf{F}_{\mathrm{M}}{\cdot}(\vec{X} - \vec{X}_c) + \vec{w} \tag{5}$$

The fine scale contribution is here represented by the microfluctuation field $\vec{w}$. The vector $\vec{x}_c$ is the actual position of the reference RVE center $\vec{X}_c$.

Obviously, rigid body displacements have to be eliminated to uniquely determine $\vec{x}$. An arbitrary boundary point may be fixed to this purpose, e.g. for a point with label 1 (see figure 4) by imposing $\vec{x}_1 = \vec{X}_1$. Substituting this in (5) leads to

$$\vec{x} \;=\; \vec{c} + \mathbf{F}_{\mathrm{M}} \cdot (\vec{X} - \vec{X}_c) + (\vec{w} - \vec{w}_1) \tag{6}$$

where $\vec{w}_1$ is the microfluctuation in point 1 and where vector $\vec{c}$ is determined from $\vec{X}_1$, being independent of the fine scale field

$$\vec{c} \;=\; \vec{X}_1 - \mathbf{F}_{\mathrm{M}} \cdot (\vec{X}_1 - \vec{X}_c) \tag{7}$$

The deformed position $\vec{x}_c$ of the reference centre $\vec{X}_c$ is then (using the trivial relation $\vec{w}_c = \vec{0}$) given by

$$\vec{x}_c \;=\; \vec{c} - \vec{w}_1 \tag{8}$$

Note that this deformed position is unknown and implicitly depends on the fine scale field. The scale transition between the kinematics at the fine and the coarse scale typically involves the volume average $\bar{\mathbf{F}}_{\mathrm{m}}$ of the fine scale deformations tensors, i.e.

$$\bar{\mathbf{F}}_{\mathrm{m}} \;=\; \frac{1}{V_0} \int_{V_0} \mathbf{F}_{\mathrm{m}} dV_0 \tag{9}$$

This volume integral can be rewritten to the boundary Γ_0 of the RVE by making use of the divergence theorem

$$\begin{aligned} \bar{\mathbf{F}}_{\mathrm{m}} \;&=\; \frac{1}{V_0} \int_{V_0} \mathbf{F}_{\mathrm{m}} dV_0 \;=\; \frac{1}{V_0} \int_{V_0} \left(\vec{\nabla}_{0,\mathrm{m}} \cdot (\mathbf{I}\vec{x}) \right)^c dV_0 \\ &=\; \frac{1}{V_0} \int_{\Gamma_0} \left(\vec{N} \cdot (\mathbf{I}\vec{x}) \right)^c d\Gamma_0 \\ &=\; \frac{1}{V_0} \int_{\Gamma_0} \vec{x}\vec{N} d\Gamma_0 \end{aligned} \tag{10}$$

where $\mathbf{I}$ represents the second-order unit tensor, Γ_0 the external boundary of the undeformed RVE V_0 and $\vec{N}$ the outward unit normal on that boundary.

Computing the fine scale deformation gradient tensor $\mathbf{F}_{\mathrm{m}}$ by taking the fine scale spatial gradient of the position vector given in equation (6) results in

$$\mathbf{F}_{\mathrm{m}} = (\vec{\nabla}_{0,\mathrm{m}} \vec{x})^c \;=\; \mathbf{F}_{\mathrm{M}} + (\vec{\nabla}_{0,\mathrm{m}} (\vec{w} - \vec{w}_1))^c \;=\; \mathbf{F}_{\mathrm{M}} + (\vec{\nabla}_{0,\mathrm{m}} \vec{w})^c \tag{11}$$

Alternatively, making use of the micro-fluctuation field, $\bar{\mathbf{F}}_{\mathrm{m}}$ can be expanded to

$$\begin{aligned}\bar{\mathbf{F}}_{\mathrm{m}} &= \mathbf{F}_{\mathrm{M}} + \frac{1}{V_0}\int_{V_0}[\vec{\nabla}_{0,\mathrm{m}}(\vec{w}-\vec{w}_1)]^c \mathrm{d}V_0 = \mathbf{F}_{\mathrm{M}} + \frac{1}{V_0}\int_{V_0}[\vec{\nabla}_{0,\mathrm{m}}\vec{w}]^c \mathrm{d}V_0 \\ &= \mathbf{F}_{\mathrm{M}} + \frac{1}{\Gamma_0}\int_{\Gamma_0}(\vec{w}-\vec{w}_1)\vec{N}\mathrm{d}\Gamma_0 = \mathbf{F}_{\mathrm{M}} + \frac{1}{\Gamma_0}\int_{\Gamma_0}\vec{w}\vec{N}\mathrm{d}\Gamma_0\end{aligned} \tag{12}$$

In the case where $\mathbf{F}_{\mathrm{M}}$ is known and displacements at the RVE boundary are to be prescribed are constrained, use is made of a scale transition relation that enforces the macroscopic deformation gradient $\mathbf{F}_{\mathrm{M}}$ to equal the volume average of its microscopic counterparts $\bar{\mathbf{F}}_{\mathrm{m}}$,

$$\mathbf{F}_{\mathrm{M}} = \bar{\mathbf{F}}_{\mathrm{m}} \tag{13}$$

Enforcing the scale transition relation (13) clearly leads to a constraint in the form of a boundary integral

$$\int_{\Gamma_0}(\vec{w}-\vec{w}_1)\vec{N}d\Gamma_0 = \mathbf{0} = \int_{\Gamma_0}\vec{w}\vec{N}d\Gamma_0 \tag{14}$$

The boundary integral (14) is the necessary condition that enforces the averaging theorem (13), which will be used in the scale transition, see also (Miehe et al., 2002). Stronger conditions are obtained by making specific choices for $\vec{w}$ that enforce this boundary integral to vanish. A few possible choices for these boundary conditions are discussed further on.

The following remarks can be made with respect to the macro-micro scale transition:

- From equation (6) and (14) it appears that the microfluctuation field only enters the kinematics relative to $\vec{w}_1$ in point 1, i.e. through $\vec{w}-\vec{w}_1$. Taking the microfluctuation field in this point $\vec{w}_1$ equal to zero will not influence the obtained solution, since the averaging theorem remains valid. The only difference resides in the resulting vector $\vec{x}_c$, which is entirely determined from the coarse scale, i.e. $\vec{x}_c = \vec{c}$ see equation (8). Clearly, $\vec{x}_c$ no longer represents the deformed position of the original RVE centre $\vec{X}_c$, since it is translated with respect to this position. The choice $\vec{w}_1 = \vec{0}$ is often made in practical implementations of the first-order homogenization scheme, since it leads to the correct solution in a practical way.
- For the first-order case, any base point could have been taken to expand $\vec{x}$ according to (4) into the RVE, leading to the same solution as the specific choice made here (the RVE center $\vec{X}_c$).

- Logically, the solution does not depend on the point that was fixed at the boundary. A point inside the volume V_0 can be taken as well to eliminate rigid body displacements in (6). Again the deformed shape of the RVE and the stress state extracted from it, remain the same.

3.4 Micro-scale RVE boundary conditions

As emphasized in the previous section, possible RVE boundary conditions naturally result from the constraint (14) imposed by the scale transition. Among the various choices possible, only three particular cases will be considered hereafter in more detail. Note that the Taylor assumption trivially satisfies (14) since the microfluctuation field is then zero in the entire volume V_0 and hence also at its boundary Γ_0.

Displacement boundary conditions The first case considered is defined by constraining each point at the RVE boundary through the macroscopic deformation by

$$\vec{x} = \mathbf{F}_{\mathrm{M}} \cdot \vec{X} \quad \text{with} \quad \vec{X} \quad \text{on} \quad \Gamma_0, \tag{15}$$

This simply implies that the micro-fluctuation field $\vec{w}$ is zero at the boundary Γ_0, which trivially satisfies (14). The position of all points at the boundary are determined through the macroscopic deformation only, leading to a linear mapping of the RVE boundary. The boundary will therefore reproduce typical stretch (tension/compression) and shear modes only.

Traction boundary conditions This case departs from the assumption that $\mathbf{P}_{\mathrm{M}}$ is to be prescribed to the RVE. The boundary conditions are then defined by constraining all tractions at the RVE boundary to the macroscopic stress tensor by

$$\vec{t} = \vec{n} \cdot \boldsymbol{\sigma}_{\mathrm{M}} \quad \text{on} \quad \Gamma, \qquad \text{or} \qquad \vec{p} = \vec{N} \cdot \mathbf{P}^c_{\mathrm{M}} \quad \text{on} \quad \Gamma_0, \tag{16}$$

with $\vec{n}$ the normal to the current (Γ) RVE boundary. Note that the traction boundary conditions (16) do not completely define the microstructural boundary value problem, since rotations are yet undetermined. As emphasized earlier, these boundary conditions are a priori not appropriate in a deformation driven procedure as pursued in the present computational homogenization scheme. The interested reader is referred to the work of Miehe (2002, 2003), where it is shown that the traction boundary condition is the weakest condition to enforce (14). From a practical point of view, these boundary conditions generally yield unsatisfactory results. Therefore, the RVE traction boundary conditions will not be explored further in this chapter.

Periodic boundary conditions Making use of the earlier introduced concept of local periodicity, periodic boundary conditions are next introduced. The periodicity conditions for the microstructural RVE are written in a general format as

$$\vec{x}^{\,+} - \vec{x}^{\,-} = \mathbf{F}_{\mathrm{M}} \cdot (\vec{X}^{+} - \vec{X}^{-}), \tag{17}$$

or formulated in terms of the micro-fluctuation fields

$$\vec{w}^{\,+} \;=\; \vec{w}^{\,-} \tag{18}$$

Deformations are periodic since micro-fluctuations on opposite sides are identical. Here the (opposite) parts of the RVE boundary Γ_0^- and Γ_0^+ are defined such that $\vec{N}^- = -\vec{N}^+$ at corresponding points on Γ_0^- and Γ_0^+, see figure 4. The periodicity condition (17), being prescribed on an initially periodic RVE, preserves the periodicity of the RVE in the deformed state.

The periodic boundary conditions (17) clearly satisfy the constraint (14). This is easily observed by splitting the RVE boundary into the parts Γ_0^+ and Γ_0^-

$$\begin{aligned} \int_{\Gamma_0} \vec{w} \vec{N} d\Gamma_0 \;&=\; \int_{\Gamma_0^+} \vec{w}^{+} \vec{N}^{+} d\Gamma_0^+ + \int_{\Gamma_0^-} \vec{w}^{-} \vec{N}^{-} d\Gamma_0^- \\ &= \int_{\Gamma_0^+} \vec{w}^{+} \vec{N}^{+} d\Gamma_0^+ - \int_{\Gamma_0^-} \vec{w}^{+} \vec{N}^{+} d\Gamma_0^- \\ &= \mathbf{0} \end{aligned} \tag{19}$$

Note that as a result of microstructural equilibrium, tractions will be anti-periodic on opposite sides:

$$\vec{p}^{\,+} \;=\; -\vec{p}^{\,-}, \tag{20}$$

Note that, as has been observed by several authors (e.g. van der Sluis et al. (2000); Terada et al. (2000)), periodic boundary conditions provide a better estimation of the overall properties, than the prescribed displacement or prescribed traction boundary conditions.

4 The macro-scale problem

4.1 The micro-macro scale transition

Once the micro-scale problem has been solved, macroscopic quantities have to be extracted from the obtained results. Whereas deformation averaging was the key assumption for the macro-micro transition, energy averaging constitutes the key assumption for the reverse transition. This energy

averaging theorem, known in the literature as the Hill-Mandel condition or macro-homogeneity condition Hill (1963); Suquet (1985), requires that the macroscopic volume average of the variation of work performed on the RVE is equal to the local variation of the work on the macro-scale, i.e.

$$\delta W_{0\mathrm{M}} \;=\; \delta W_{0\mathrm{m}} \tag{21}$$

Formulated in terms of a work conjugated set, i.e. the deformation gradient tensor and the first Piola-Kirchhoff stress tensor, the Hill-Mandel condition reads

$$\underbrace{\mathbf{P}_{\mathrm{M}} : \delta\mathbf{F}^{c}_{\mathrm{M}}}_{\delta W_{0\mathrm{M}}} \;=\; \underbrace{\frac{1}{V_0}\int_{V_0} \mathbf{P}_{\mathrm{m}} : \delta\mathbf{F}^{c}_{\mathrm{m}}\,\mathrm{d}V_0}_{\delta W_{0\mathrm{m}}} \tag{22}$$

The averaged microstructural work in the right-hand side of (22) may be expressed in terms of RVE boundary quantities

$$\delta W_{0\mathrm{m}} = \frac{1}{V_0}\int_{V_0} \mathbf{P}_{\mathrm{m}} : \delta\mathbf{F}^{c}_{\mathrm{m}}\,\mathrm{d}V_0 = \frac{1}{V_0}\int_{\Gamma_0} \vec{p}\cdot\delta\vec{x}\;\mathrm{d}\Gamma_0, \tag{23}$$

where the relation (with account for microstructural equilibrium)

$$\mathbf{P}_{\mathrm{m}} : \nabla_{0\mathrm{m}}\delta\vec{x} = \nabla_{0\mathrm{m}}\cdot(\mathbf{P}^{c}_{\mathrm{m}}\cdot\delta\vec{x}) - (\nabla_{0\mathrm{m}}\cdot\mathbf{P}^{c}_{\mathrm{m}})\cdot\delta\vec{x} = \nabla_{0\mathrm{m}}\cdot(\mathbf{P}^{c}_{\mathrm{m}}\cdot\delta\vec{x}),$$

and the divergence theorem have been used.

As will be shown next, an important result of postulating the Hill-Mandel condition for an RVE with kinematic boundary conditions (fully prescribed or periodically tied), is the fact that the macroscopic stress tensor $\mathbf{P}_{\mathrm{M}}$ equals the volume average $\bar{\mathbf{P}}_{\mathrm{m}}$ of the microscopic stress tensors. To this purpose, it is convenient to establish the boundary relation for the mean RVE stress $\bar{\mathbf{P}}_{\mathrm{m}}$, i.e.

$$\begin{aligned}
\bar{\mathbf{P}}_{\mathrm{m}} \;&=\; \frac{1}{V_0}\int_{V_0} \mathbf{P}_{\mathrm{m}}\mathrm{d}V_0 \\
&=\; \frac{1}{V_0}\int_{V_0} \vec{\nabla}_0\cdot(\mathbf{P}^{c}_{\mathrm{m}}\,\vec{X})\,\mathrm{d}V_0 \\
&=\; \frac{1}{V_0}\int_{\Gamma_0} \vec{N}\cdot(\mathbf{P}^{c}_{\mathrm{m}}\,\vec{X})\,\mathrm{d}\Gamma_0 \\
&=\; \frac{1}{V_0}\int_{\Gamma_0} \vec{p}\,\vec{X}\,\mathrm{d}\Gamma_0
\end{aligned} \tag{24}$$

Displacement boundary conditions In case of fully prescribed boundary displacements (15), substitution of the variation of the boundary position vectors $\delta\vec{x} = \delta\mathbf{F}_{\mathrm{M}}{\cdot}\vec{X}$ into the expression for the averaged microwork (23) with incorporation of (33) gives

$$\delta W_{0\mathrm{m}} = \frac{1}{V_0}\int_{\Gamma_0} \vec{p}{\cdot}(\delta\mathbf{F}_{\mathrm{M}}{\cdot}\vec{X})\,\mathrm{d}\Gamma_0 = \left[\frac{1}{V_0}\int_{\Gamma_0} \vec{p}\vec{X}\,\mathrm{d}\Gamma_0\right] : \delta\mathbf{F}^c_{\mathrm{M}} = \bar{\mathbf{P}}_{\mathrm{m}} : \delta\mathbf{F}^c_{\mathrm{M}} \tag{25}$$

Enforcing the Hill-Mandel condition (22) thus implies that

$$\mathbf{P}_{\mathrm{M}} \;=\; \bar{\mathbf{P}}_{\mathrm{m}} \tag{26}$$

Traction boundary conditions Substitution of the traction boundary condition (16) into (23), with account for the variation of the average of the microscopic deformation gradient tensor obtained by varying relation (9), leads to

$$\delta W_{0\mathrm{m}} = \frac{1}{V_0}\int_{\Gamma_0} (\vec{N}{\cdot}\mathbf{P}^c_{\mathrm{M}}){\cdot}\delta\vec{x}\,\mathrm{d}\Gamma_0 = \mathbf{P}_{\mathrm{M}} : \left[\frac{1}{V_0}\int_{\Gamma_0} \vec{N}\delta\vec{x}\,\mathrm{d}\Gamma_0\right] = \mathbf{P}_{\mathrm{M}} : \delta\bar{\mathbf{F}}^c_{\mathrm{m}}. \tag{27}$$

In this case, the Hill-Mandel condition (22) enforces the resulting macroscopic deformation gradient to be taken as the volume average of the microscopic deformation gradients, i.e..

$$\mathbf{F}_{\mathrm{M}} \;=\; \bar{\mathbf{F}}_{\mathrm{m}} \;=\; \frac{1}{V_0}\int_{V_0} \mathbf{F}_{\mathrm{m}}\mathrm{d}V_0 \tag{28}$$

This implies that the traction boundary conditions, complemented by the Hill-Mandel conditions, constitute the weakest kinematic constraint for the boundary displacements, i.e. equation (14). Computing the volume average

of the micro-scale RVE stresses from equation (24) now yields

$$\begin{aligned}
\bar{\mathbf{P}}_{\mathrm{m}} &= \frac{1}{V_0}\int_{\Gamma_0} \vec{p}\,\vec{X}\,\mathrm{d}\Gamma_0 \\
&= \frac{1}{V_0}\int_{\Gamma_0} (\vec{N}\cdot\mathbf{P}^c_{\mathrm{M}})\,\vec{X}\,\mathrm{d}\Gamma_0 \\
&= \mathbf{P}_{\mathrm{M}}\cdot\left[\frac{1}{V_0}\int_{\Gamma_0} \vec{N}\vec{X}\,\mathrm{d}\Gamma_0\right] \\
&= \mathbf{P}_{\mathrm{M}}\cdot\left[\frac{1}{V_0}V_0\,\mathbf{I}\right] \\
&= \mathbf{P}_{\mathrm{M}}
\end{aligned} \tag{29}$$

Again, the macroscopic stress equals the volume average of the microscopic stress, but this time this conclusion does not result from the Hill-Mandel condition.

Periodic boundary conditions For the periodic boundary conditions (17) and the resulting anti-periodic tractions (20)

$$\begin{aligned}
\delta W_{0\mathrm{m}} &= \frac{1}{V_0}\left\{\int_{\Gamma_0^+} \vec{p}^{\,+}\cdot\delta\vec{x}^{\,+}\,\mathrm{d}\Gamma_0 + \int_{\Gamma_0^-} \vec{p}^{\,-}\cdot\delta\vec{x}^{\,-}\,\mathrm{d}\Gamma_0\right\} \\
&= \frac{1}{V_0}\int_{\Gamma_0^+} \vec{p}^{\,+}\cdot(\delta\vec{x}^{\,+} - \delta\vec{x}^{\,-})\,\mathrm{d}\Gamma_0^+ \\
&= \left[\frac{1}{V_0}\int_{\Gamma_0^+} \vec{p}^{\,+}(\vec{X}^{+} - \vec{X}^{-})\,\mathrm{d}\Gamma_0^+\right] : \delta\mathbf{F}^c_{\mathrm{M}} = \left[\frac{1}{V_0}\int_{\Gamma_0} \vec{p}\vec{X}\,\mathrm{d}\Gamma_0\right] : \delta\mathbf{F}^c_{\mathrm{M}} \\
&= \bar{\mathbf{P}}_{\mathrm{m}} : \delta\mathbf{F}^c_{\mathrm{M}}
\end{aligned} \tag{30}$$

Enforcing the Hill-Mandel condition (22) again implies that

$$\mathbf{P}_{\mathrm{M}} = \bar{\mathbf{P}}_{\mathrm{m}} \tag{31}$$

4.2 Macroscopic stress tensors

Since the scale transition implies stress averaging for all considered boundary conditions, the macroscopic stress tensor is given by

$$\mathbf{P}_{\mathrm{M}} = \frac{1}{V_0}\int_{V_0} \mathbf{P}_{\mathrm{m}}\mathrm{d}V_0 \tag{32}$$

$$= \frac{1}{V_0}\int_{\Gamma_0} \vec{p}\,\vec{X}\,\mathrm{d}\Gamma_0 \tag{33}$$

The volume average of the microscopic Cauchy stress tensor $\boldsymbol{\sigma}_{\mathrm{m}}$ over the current RVE volume V can be elaborated similarly to (33)

$$\boldsymbol{\sigma}^*_{\mathrm{M}} = \frac{1}{V}\int_{V} \boldsymbol{\sigma}_{\mathrm{m}}\,\mathrm{d}V = \frac{1}{V}\int_{\Gamma} \vec{t}\,\vec{x}\,\mathrm{d}\Gamma. \tag{34}$$

Just as it is the case for kinematic quantities, the usual pull-back push-forward relations between stress measures (e.g. the Cauchy and the first Piola-Kirchhoff stress tensors) are, in general, not valid for the volume averages of the microstructural counterparts $\boldsymbol{\sigma}^*_{\mathrm{M}} \neq \mathbf{P}_{\mathrm{M}}\cdot\mathbf{F}^c_{\mathrm{M}}/\det(\mathbf{F}_{\mathrm{M}})$. If the averaging is based on $\mathbf{P}_{\mathrm{M}}$, the Cauchy stress tensor on the macrolevel should be defined as

$$\boldsymbol{\sigma}_{\mathrm{M}} = \frac{1}{\det(\mathbf{F}_{\mathrm{M}})}\mathbf{P}_{\mathrm{M}}\cdot\mathbf{F}^c_{\mathrm{M}}. \tag{35}$$

Clearly, there is some arbitrariness in the choice of the governing deformation and stress tensors, whose macroscopic measures are equal to the volume average of their microscopic counterparts (through the imposed scale transition relations). Macroscopic measures defined on another configuration are then expressed in terms of the governing averaged quantities using the standard pull-back push-forward relations. The specific selection made here is mainly based on its ease of implementation. The actual choice of the "primary" averaging measures used here, i.e. the deformation gradient tensor $\mathbf{F}$ and the first Piola-Kirchhoff stress tensor $\mathbf{P}$ (and their rates), has been advocated in Miehe et al. (1999b); Hill (1984); Nemat-Nasser (1999) (in the last two references the nominal stress $\mathbf{S}_N = \det(\mathbf{F})\mathbf{F}^{-1}\cdot\boldsymbol{\sigma} = \mathbf{P}^c$ has been used). This particular choice is motivated by the fact that these two measures are work conjugated, combined with the observation that their volume averages can exclusively be defined in terms of the microstructural quantities of the undeformed RVE boundary.

5 Two-scale numerical solution strategy

Once the boundary conditions have been properly defined through one of the methods outline above and once all phases in the microstructure have been characterized, a standard boundary value problem (BVP) has been obtained. The solution of this BVP follows standard procedures. In the present computational homogenization method, it will be assumed that the finite element method has been used to this purpose. The solution of this BVP problem automatically leads to the proper determination of all position vectors in the RVE and all tractions along its boundary. The analysis is further restricted to kinematic RVE boundary conditions only.

5.1 RVE boundary value problem

The RVE problem to be solved is a standard non-linear quasi-static boundary value problem with kinematic boundary conditions. Following the standard finite element procedure for the microlevel RVE, after discretization, the weak form of equilibrium (2) with account for the constitutive relations (3) leads to a system of non-linear algebraic equations in the unknown nodal displacements $\underset{\sim}{u}$

$$\underset{\sim}{f}_{int}(\underset{\sim}{u}) = \underset{\sim}{f}_{ext}, \tag{36}$$

expressing the balance of internal and external nodal forces. This system has to be completed by the governing boundary conditions. To this purpose, the earlier introduced kinematic boundary conditions (15) or (17) will be elaborated in more detail.

Fully prescribed boundary displacements In the case of the fully prescribed displacement boundary conditions (15), the displacements of all nodes on the boundary is simply given by

$$\vec{u}_p = (\mathbf{F}_{\mathrm{M}} - \mathbf{I})\cdot\vec{X}_p, \qquad p = \overline{1, N_p} \tag{37}$$

where N_p is the number of prescribed nodes, which in this case simply equals to the number of boundary nodes. The boundary conditions (37) are simply added to the system (36) in a standard manner by static condensation, Lagrange multipliers or penalty functions.

Periodic boundary conditions Prior to the incorporation of the periodic boundary conditions (17), they have to be rewritten into a format that is more suitable for a finite element framework. Consider a two-dimensional periodic RVE schematically depicted in figure 4. The boundary of this RVE

can be split into four parts, here denoted as "T" top, "B" bottom, "R" right and "L" left. To ease application of the periodicity constraint, a finite element discretization is next considered which has a periodic distribution of nodes on opposite edges. Exploiting the initial periodicity of the RVE (in its reference configuration) allows to write for every respective pair of nodes on the top-bottom and right-left boundaries:

$$\begin{aligned} \vec{X}_{\mathrm{T}} - \vec{X}_{\mathrm{B}} &= \vec{X}_4 - \vec{X}_1, \\ \vec{X}_{\mathrm{R}} - \vec{X}_{\mathrm{L}} &= \vec{X}_2 - \vec{X}_1, \end{aligned} \tag{38}$$

where $\vec{X}_p,\ p = 1, 2, 4$ are the position vectors of the corner nodes 1, 2 and 4 in the undeformed state. Considering pairs of periodic nodes on opposite boundaries, allows to express (17) as

$$\begin{aligned} \vec{x}_{\mathrm{T}} - \vec{x}_{\mathrm{B}} &= \mathbf{F}_{\mathrm{M}} \cdot (\vec{X}_4 - \vec{X}_1), \\ \vec{x}_{\mathrm{R}} - \vec{x}_{\mathrm{L}} &= \mathbf{F}_{\mathrm{M}} \cdot (\vec{X}_2 - \vec{X}_1). \end{aligned} \tag{39}$$

Applying these relations to the four corner nodes, permits to conclude that the position vectors of the corner nodes in the deformed state are in fact prescribed according to

$$\vec{x}_p = \mathbf{F}_{\mathrm{M}} \cdot \vec{X}_p, \qquad p = 1, 2, 4 \tag{40}$$

The periodic boundary conditions may finally be rewritten as

$$\begin{aligned} \vec{x}_{\mathrm{T}} &= \vec{x}_{\mathrm{B}} + \vec{x}_4 - \vec{x}_1, \\ \vec{x}_{\mathrm{R}} &= \vec{x}_{\mathrm{L}} + \vec{x}_2 - \vec{x}_1. \end{aligned} \tag{41}$$

Since these conditions are trivially satisfied in the undeformed configuration, they may be formulated in terms of displacements

$$\begin{aligned} \vec{u}_{\mathrm{T}} &= \vec{u}_{\mathrm{B}} + \vec{u}_4 - \vec{u}_1, \\ \vec{u}_{\mathrm{R}} &= \vec{u}_{\mathrm{L}} + \vec{u}_2 - \vec{u}_1, \end{aligned} \tag{42}$$

whereby

$$\vec{u}_p = (\mathbf{F}_{\mathrm{M}} - \mathbf{I}) \cdot \vec{X}_p, \qquad p = 1, 2, 4 \tag{43}$$

In a discretized format the relations (42) lead to a set of homogeneous constraints of the type

$$\underline{\underline{C}}_a \, \underline{u}_a = \underline{0}, \tag{44}$$

with $\underline{\underline{C}}_a$ a matrix containing coefficients in the constraint relations and $\underline{u}_a$ a column with the degrees of freedom involved in the constraints. Procedures

for imposing constraints (44) include the direct elimination of the dependent degrees of freedom from the system of equations, or the use of Lagrange multipliers or penalty functions. In the following, constraints (44) are enforced by elimination of the dependent degrees of freedom. Although such a procedure may be found in many textbooks on finite elements (e.g. Cook et al. (1989)), it is here summarized for the sake of clarity and completeness, since it will be applied in section 5.3 for the derivation of the macroscopic tangent stiffness.

To this purpose, the homogeneous constraint relations (44) are partitioned according to

$$[\underline{C}_i \quad \underline{C}_d] \left[\begin{array}{c} \underset{\sim}{u}_i \\ \underset{\sim}{u}_d \end{array} \right] = \underset{\sim}{0}, \tag{45}$$

where $\underset{\sim}{u}_i$ are the independent degrees of freedom (to be retained in the system) and $\underset{\sim}{u}_d$ are the dependent degrees of freedom (to be eliminated from the system). Because there are as many dependent degrees of freedom $\underset{\sim}{u}_d$ as there are independent constraint equations in (45), matrix $\underline{C}_d$ is square and non-singular. Solution for $\underset{\sim}{u}_d$ yields

$$\underset{\sim}{u}_d = \underline{C}_{di}\, \underset{\sim}{u}_i, \qquad \text{with} \quad \underline{C}_{di} = -\underline{C}_d^{-1} \underline{C}_i. \tag{46}$$

This relation may be further rewritten as

$$\left[\begin{array}{c} \underset{\sim}{u}_i \\ \underset{\sim}{u}_d \end{array} \right] = \underline{T}\, \underset{\sim}{u}_i, \qquad \text{with} \quad \underline{T} = \left[\begin{array}{c} \underline{I} \\ \underline{C}_{di} \end{array} \right], \tag{47}$$

where $\underline{I}$ is a unit matrix of size $[N_i \times N_i]$, with N_i the number of the independent degrees of freedom.

With the transformation matrix $\underline{T}$ defined such that $\underset{\sim}{d} = \underline{T}\, \underset{\sim}{d}'$, the common transformations $\underset{\sim}{r}' = \underline{T}^T \underset{\sim}{r}$ and $\underline{K}' = \underline{T}^T \underline{K} \underline{T}$ can be applied to a linear system of equations of the form $\underline{K}\, \underset{\sim}{d} = \underset{\sim}{r}$, leading to a new system $\underline{K}'\, \underset{\sim}{d}' = \underset{\sim}{r}'$.

The standard linearization of the non-linear system of equations (36) leads to a linear system in the iterative corrections $\delta \underset{\sim}{u}$ to the current estimate $\underset{\sim}{u}$. This system may be partitioned as

$$\left[\begin{array}{cc} \underline{K}_{ii} & \underline{K}_{id} \\ \underline{K}_{di} & \underline{K}_{dd} \end{array} \right] \left[\begin{array}{c} \delta \underset{\sim}{u}_i \\ \delta \underset{\sim}{u}_d \end{array} \right] = \left[\begin{array}{c} \delta \underset{\sim}{r}_i \\ \delta \underset{\sim}{r}_d \end{array} \right], \tag{48}$$

with the residual nodal forces at the right-hand side. Noting that all the constraint equations considered above are linear, and thus their linearization is straightforward, application of the transformation (47) to the system (48) gives

$$\left[\underline{K}_{ii} + \underline{K}_{id} \underline{C}_{di} + \underline{C}_{di}^T \underline{K}_{di} + \underline{C}_{di}^T \underline{K}_{dd} \underline{C}_{di} \right] \delta \underset{\sim}{u}_i = \left[\delta \underset{\sim}{r}_i + \underline{C}_{di}^T \delta \underset{\sim}{r}_d \right]. \tag{49}$$

Note that the boundary conditions (43) prescribing displacements of the corner nodes have not yet been applied. The column of "independent" degrees of freedom $\underset{\sim}{u}_i$ includes the prescribed corner nodes $\underset{\sim}{u}_p$ among other nodes. The boundary conditions (43) should be applied to the system (49) is a standard manner.

The condition of antiperiodic tractions (20) will be addressed in section 5.2.

5.2 Extraction of the macroscopic stress

After the analysis of a microstructural RVE is completed, the RVE averaged stress have to be extracted. Of course, the macroscopic stress tensor can be calculated by numerically evaluating the volume integral (32). However it is computationally more efficient to compute the boundary integral (33), which can be further simplified for the case of the periodic boundary conditions.

Fully prescribed boundary displacements For the case of prescribed displacement boundary conditions the boundary integral (33) simply leads to

$$\mathbf{P}_{\mathrm{M}} = \frac{1}{V_0} \sum_{p=1}^{N_p} \vec{f}_p \vec{X}_p, \tag{50}$$

where $\vec{f}_p$ are the resulting external forces at the boundary nodes and $\vec{X}_p$ the position vectors of these nodes in the undeformed state; N_p is the number of the nodes on the boundary.

Periodic boundary conditions In order to simplify the boundary integral (33) for the case of periodic boundary conditions, consider all the forces acting on the RVE boundary subjected to the boundary conditions according to (42)–(43). At the three prescribed corner nodes the resulting external forces $\vec{f}^{\mathrm{e}}_p$, $p = 1, 2, 4$ act. Additionally, there are forces involved in every constraint (tying) relation (42). For example, for each constraint relation between pairs of the nodes on the bottom-top boundaries there is a tying force at the node on the bottom boundary $\vec{p}^{\,\mathrm{t}}_{\mathrm{B}}$, a tying force at the node on the top boundary $\vec{p}^{\,\mathrm{t}}_{\mathrm{T}}$ and tying forces at the corner nodes 1 and 4, $\vec{p}^{\,\mathrm{t_B}}_1$ and $\vec{p}^{\,\mathrm{t_B}}_4$, respectively. Similarly there are forces $\vec{p}^{\,\mathrm{t}}_{\mathrm{L}}$, $\vec{p}^{\,\mathrm{t}}_{\mathrm{R}}$, $\vec{p}^{\,\mathrm{t_L}}_1$ and $\vec{p}^{\,\mathrm{t_L}}_2$ corresponding to the left-right constraints. All these forces are schematically shown in figure 5.

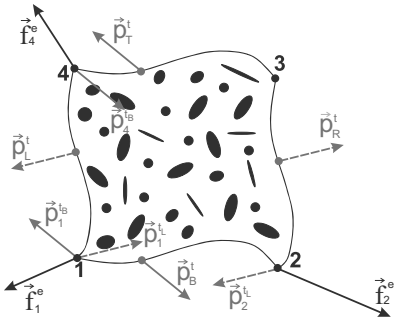


Figure 5. Tractions acting on the boundary of a two-dimensional RVE subjected to periodic boundary conditions.

Each constraint relation satisfies the condition of zero virtual work, i.e.

$$\begin{aligned}\vec{p}^{\,\mathrm{t}}_{\mathrm{B}}\cdot\delta\vec{x}_{\mathrm{B}} + \vec{p}^{\,\mathrm{t}}_{\mathrm{T}}\cdot\delta\vec{x}_{\mathrm{T}} + \vec{p}^{\,\mathrm{t}_{\mathrm{B}}}_{1}\cdot\delta\vec{x}_1 + \vec{p}^{\,\mathrm{t}_{\mathrm{B}}}_{4}\cdot\delta\vec{x}_4 = 0,\\ \vec{p}^{\,\mathrm{t}}_{\mathrm{L}}\cdot\delta\vec{x}_{\mathrm{L}} + \vec{p}^{\,\mathrm{t}}_{\mathrm{R}}\cdot\delta\vec{x}_{\mathrm{R}} + \vec{p}^{\,\mathrm{t}_{\mathrm{L}}}_{1}\cdot\delta\vec{x}_1 + \vec{p}^{\,\mathrm{t}_{\mathrm{L}}}_{2}\cdot\delta\vec{x}_2 = 0.\end{aligned} \tag{51}$$

Substitution of the variation of the constraints (41) into (51) gives

$$\begin{aligned}(\vec{p}^{\,\mathrm{t}}_{\mathrm{B}} + \vec{p}^{\,\mathrm{t}}_{\mathrm{T}})\cdot\delta\vec{x}_{\mathrm{B}} + (\vec{p}^{\,\mathrm{t}_{\mathrm{B}}}_{1} - \vec{p}^{\,\mathrm{t}}_{\mathrm{T}})\cdot\delta\vec{x}_1 + (\vec{p}^{\,\mathrm{t}}_{\mathrm{T}} + \vec{p}^{\,\mathrm{t}_{\mathrm{B}}}_{4})\cdot\delta\vec{x}_4 = 0,\\ (\vec{p}^{\,\mathrm{t}}_{\mathrm{L}} + \vec{p}^{\,\mathrm{t}}_{\mathrm{R}})\cdot\delta\vec{x}_{\mathrm{L}} + (\vec{p}^{\,\mathrm{t}_{\mathrm{L}}}_{1} - \vec{p}^{\,\mathrm{t}}_{\mathrm{R}})\cdot\delta\vec{x}_1 + (\vec{p}^{\,\mathrm{t}}_{\mathrm{R}} + \vec{p}^{\,\mathrm{t}_{\mathrm{L}}}_{2})\cdot\delta\vec{x}_2 = 0.\end{aligned} \tag{52}$$

These relations should hold for any $\delta\vec{x}_{\mathrm{B}}$, $\delta\vec{x}_{\mathrm{L}}$, $\delta\vec{x}_1$, $\delta\vec{x}_2$, $\delta\vec{x}_4$, therefore

$$\begin{aligned}\vec{p}^{\,\mathrm{t}}_{\mathrm{B}} = -\vec{p}^{\,\mathrm{t}}_{\mathrm{T}} = -\vec{p}^{\,\mathrm{t}_{\mathrm{B}}}_{1} = \vec{p}^{\,\mathrm{t}_{\mathrm{B}}}_{4},\\ \vec{p}^{\,\mathrm{t}}_{\mathrm{L}} = -\vec{p}^{\,\mathrm{t}}_{\mathrm{R}} = -\vec{p}^{\,\mathrm{t}_{\mathrm{L}}}_{1} = \vec{p}^{\,\mathrm{t}_{\mathrm{L}}}_{2}.\end{aligned} \tag{53}$$

Equation (53) reflects the antiperiodicity of tying forces on the opposite boundaries, which has been introduced previously in equation (20).

With account for all forces acting on the RVE boundary, the boundary integral (33) is written as

$$\begin{aligned}\mathbf{P}_{\mathrm{M}} = \frac{1}{V_0}\Bigg(&\vec{f}^{\,\mathrm{e}}_1\vec{X}_1 + \vec{f}^{\,\mathrm{e}}_2\vec{X}_2 + \vec{f}^{\,\mathrm{e}}_4\vec{X}_4 + \int_{\Gamma_{0\mathrm{B}}} \vec{p}^{\,\mathrm{t}}_{\mathrm{B}}\vec{X}_{\mathrm{B}}\,\mathrm{d}\Gamma_0 + \int_{\Gamma_{0\mathrm{T}}} \vec{p}^{\,\mathrm{t}}_{\mathrm{T}}\vec{X}_{\mathrm{T}}\,\mathrm{d}\Gamma_0 +\\ &\int_{\Gamma_{0\mathrm{L}}} \vec{p}^{\,\mathrm{t}}_{\mathrm{L}}\vec{X}_{\mathrm{L}}\,\mathrm{d}\Gamma_0 + \int_{\Gamma_{0\mathrm{R}}} \vec{p}^{\,\mathrm{t}}_{\mathrm{R}}\vec{X}_{\mathrm{R}}\,\mathrm{d}\Gamma_0 + \Big(\int_{\Gamma_{0\mathrm{B}}} \vec{p}^{\,\mathrm{t}_{\mathrm{B}}}_{1}\,\mathrm{d}\Gamma_0\Big)\vec{X}_1 +\\ &\Big(\int_{\Gamma_{0\mathrm{L}}} \vec{p}^{\,\mathrm{t}_{\mathrm{L}}}_{1}\,\mathrm{d}\Gamma_0\Big)\vec{X}_1 + \Big(\int_{\Gamma_{0\mathrm{B}}} \vec{p}^{\,\mathrm{t}_{\mathrm{B}}}_{4}\,\mathrm{d}\Gamma_0\Big)\vec{X}_4 + \Big(\int_{\Gamma_{0\mathrm{L}}} \vec{p}^{\,\mathrm{t}_{\mathrm{L}}}_{2}\,\mathrm{d}\Gamma_0\Big)\vec{X}_2\Bigg).\end{aligned} \tag{54}$$

Making use of the relation between tying forces (53) gives

$$\mathbf{P}_{\mathrm{M}} = \frac{1}{V_0}\Big(\sum_{p=1,2,4} \vec{f}^{\,\mathrm{e}}_p \vec{X}_p + \int_{\Gamma_{0\mathrm{B}}} \vec{p}^{\,\mathrm{t}}_{\mathrm{B}} (\vec{X}_{\mathrm{B}} - \vec{X}_{\mathrm{T}})\, \mathrm{d}\Gamma_0 + \int_{\Gamma_{0\mathrm{L}}} \vec{p}^{\,\mathrm{t}}_{\mathrm{L}} (\vec{X}_{\mathrm{L}} - \vec{X}_{\mathrm{R}})\, \mathrm{d}\Gamma_0 + \big(\int_{\Gamma_{0\mathrm{B}}} \vec{p}^{\,\mathrm{t}_{\mathrm{B}}}_1\, \mathrm{d}\Gamma_0 \big) \vec{X}_1 + \big(\int_{\Gamma_{0\mathrm{L}}} \vec{p}^{\,\mathrm{t}_{\mathrm{L}}}_1\, \mathrm{d}\Gamma_0 \big) \vec{X}_1 + \big(\int_{\Gamma_{0\mathrm{B}}} \vec{p}^{\,\mathrm{t}_{\mathrm{B}}}_4\, \mathrm{d}\Gamma_0 \big) \vec{X}_4 + \big(\int_{\Gamma_{0\mathrm{L}}} \vec{p}^{\,\mathrm{t}_{\mathrm{L}}}_2\, \mathrm{d}\Gamma_0 \big) \vec{X}_2 \Big). \tag{55}$$

Inserting the conditions of the initial periodicity of the RVE (38) results in

$$\mathbf{P}_{\mathrm{M}} = \frac{1}{V_0}\Big(\sum_{p=1,2,4} \vec{f}^{\,\mathrm{e}}_p \vec{X}_p + \int_{\Gamma_{0\mathrm{B}}} (\vec{p}^{\,\mathrm{t}}_{\mathrm{B}} + \vec{p}^{\,\mathrm{t}_{\mathrm{B}}}_1) \vec{X}_1\, \mathrm{d}\Gamma_0 + \int_{\Gamma_{0\mathrm{L}}} (\vec{p}^{\,\mathrm{t}}_{\mathrm{L}} + \vec{p}^{\,\mathrm{t}_{\mathrm{L}}}_1) \vec{X}_1\, \mathrm{d}\Gamma_0 + \int_{\Gamma_{0\mathrm{B}}} (\vec{p}^{\,\mathrm{t}_{\mathrm{B}}}_4 - \vec{p}^{\,\mathrm{t}}_{\mathrm{B}}) \vec{X}_4\, \mathrm{d}\Gamma_0 + \int_{\Gamma_{0\mathrm{L}}} (\vec{p}^{\,\mathrm{t}_{\mathrm{L}}}_2 - \vec{p}^{\,\mathrm{t}}_{\mathrm{L}}) \vec{X}_2\, \mathrm{d}\Gamma_0 \Big), \tag{56}$$

which after substitution of the remaining relations between tying forces (53) gives

$$\mathbf{P}_{\mathrm{M}} = \frac{1}{V_0} \sum_{p=1,2,4} \vec{f}^{\,\mathrm{e}}_p \vec{X}_p. \tag{57}$$

Therefore, when the periodic boundary conditions are used, all the terms with forces involved into the periodicity constraints cancel out from the boundary integral (33) and the only contribution left is by the external forces at the three prescribed corner nodes.

5.3 Extraction of the macroscopic tangent operator

When the micro-macro approach is implemented within the framework of a non-linear finite element code, the stiffness matrix at every macroscopic integration point is required. Because in the computational homogenization approach there is no explicit form of the constitutive behaviour on the macrolevel assumed a priori, the stiffness matrix has to be determined numerically from the relation between variations of the macroscopic stress and variations of the macroscopic deformation at such a point. This may be realized by numerical differentiation of the numerical macroscopic stress-strain relation, for example using a forward difference approximation as has been suggested in Miehe (1996). Another approach is to condense the microstructural stiffness to the local macroscopic stiffness. This is achieved by reducing

the total RVE system of equations to the relation between the forces acting on the RVE boundary and the associated boundary displacements. Such a procedure in combination with the Lagrange multiplier method to impose boundary constraints has been elaborated in Miehe and Koch (2002). Here an alternative scheme, which employs the direct condensation of the constrained degrees of freedom, as has been presented in Kouznetsova et al. (2001); Kouznetsova (2002) will be considered. After the condensed microscopic stiffness relating the prescribed displacement and force variations is obtained, it needs to be transformed to arrive at an expression relating variations of the macroscopic stress and deformation tensors, typically used in the finite element codes. These two steps are elaborated in the following.

Condensation of the microscopic stiffness matrix: fully prescribed boundary displacements First the total microstructural system of equations (in its linearized form) is partitioned as

$$\begin{bmatrix} \underline{K}_{pp} & \underline{K}_{pf} \\ \underline{K}_{fp} & \underline{K}_{ff} \end{bmatrix} \begin{bmatrix} \delta \underset{\sim}{u}_p \\ \delta \underset{\sim}{u}_f \end{bmatrix} = \begin{bmatrix} \delta \underset{\sim}{f}_p \\ \underset{\sim}{0} \end{bmatrix}, \tag{58}$$

where $\delta \underset{\sim}{u}_p$ and $\delta \underset{\sim}{f}_p$ are the columns with iterative displacements and external forces of the boundary nodes, respectively, and $\delta \underset{\sim}{u}_f$ the column with the iterative displacements of the remaining (interior) nodes; $\underline{K}_{pp}$, $\underline{K}_{pf}$, $\underline{K}_{fp}$ and $\underline{K}_{ff}$ are the corresponding partitions of the total RVE stiffness matrix. The stiffness matrix in the formulation (58) is taken at the end of a microstructural increment, where a converged state is reached. Elimination of $\delta \underset{\sim}{u}_f$ from (58) leads to the reduced stiffness matrix $\underline{K}_{\mathrm{M}}$ relating boundary displacement variations to boundary force variations

$$\underline{K}_{\mathrm{M}} \delta \underset{\sim}{u}_p = \delta \underset{\sim}{f}_p, \qquad \text{with} \quad \underline{K}_{\mathrm{M}} = \underline{K}_{pp} - \underline{K}_{pf} (\underline{K}_{ff})^{-1} \underline{K}_{fp}. \tag{59}$$

Condensation of the microscopic stiffness matrix: periodic boundary conditions In the case of the periodic boundary conditions the point of departure is the microscopic system of equations (49) from which the dependent degrees of freedom have been eliminated (as described in section 5.1)

$$\underline{K}^{\star} \delta \underset{\sim}{u}_i = \delta \underset{\sim}{r}^{\star}, \tag{60}$$

$$\text{with} \qquad \underline{K}^{\star} = \underline{K}_{ii} + \underline{K}_{id} \underline{C}_{di} + \underline{C}_{di}^T \underline{K}_{di} + \underline{C}_{di}^T \underline{K}_{dd} \underline{C}_{di},$$

$$\delta \underset{\sim}{r}^{\star} = \delta \underset{\sim}{r}_i + \underline{C}_{di}^T \delta \underset{\sim}{r}_d.$$

Next, system (60) is further split, similarly to (58), into the parts corresponding to the variations of the prescribed degrees of freedom $\delta \underset{\sim}{u}_p$ (which

in this case are the varied positions of the three corner nodes prescribed according to (43)), variations of the external forces at these prescribed nodes denoted by $\delta \underset{\sim}{f}^{\star}_{p}$, and the remaining (free) displacement variations $\delta \underset{\sim}{u}_{f}$:

$$\begin{bmatrix} \underline{K}^{\star}_{pp} & \underline{K}^{\star}_{pf} \\ \underline{K}^{\star}_{fp} & \underline{K}^{\star}_{ff} \end{bmatrix} \begin{bmatrix} \delta \underset{\sim}{u}_{p} \\ \delta \underset{\sim}{u}_{f} \end{bmatrix} = \begin{bmatrix} \delta \underset{\sim}{f}^{\star}_{p} \\ \underset{\sim}{0} \end{bmatrix}. \tag{61}$$

Then the reduced stiffness matrix $\underline{K}^{\star}_{\mathrm{M}}$ in case of periodic boundary conditions is obtained as

$$\underline{K}^{\star}_{\mathrm{M}} \delta \underset{\sim}{u}_{p} = \delta \underset{\sim}{f}^{\star}_{p}, \qquad \text{with} \quad \underline{K}^{\star}_{\mathrm{M}} = \underline{K}^{\star}_{pp} - \underline{K}^{\star}_{pf} (\underline{K}^{\star}_{ff})^{-1} \underline{K}^{\star}_{fp}. \tag{62}$$

Note that $\underline{K}^{\star}_{\mathrm{M}}$ is $[6 \times 6]$ matrix only (in the two-dimensional case).

Final macroscopic tangent Finally, the resulting relation between displacement and force variations (relation (59) if prescribed displacement boundary conditions are used, or relation (62) if periodicity conditions are employed) needs to be transformed to arrive at an expression relating variations of the macroscopic stress and deformation tensors

$$\delta \mathbf{P}_{\mathrm{M}} = {}^{4}\mathbf{C}^{\mathrm{P}}_{\mathrm{M}} : \delta \mathbf{F}^{c}_{\mathrm{M}}, \tag{63}$$

where the fourth order tensor ${}^{4}\mathbf{C}^{\mathrm{P}}_{\mathrm{M}}$ represents the required consistent tangent stiffness at the macroscopic integration point level.

In order to obtain this constitutive tangent from the reduced stiffness matrix $\underline{K}_{\mathrm{M}}$ (or $\underline{K}^{\star}_{\mathrm{M}}$), first relations (59) and (62) are rewritten in a specific vector/tensor format

$$\sum_{j} \mathbf{K}^{(ij)}_{\mathrm{M}} \cdot \delta \vec{u}_{(j)} = \delta \vec{f}_{(i)}, \tag{64}$$

where indices i and j take the values $i, j = \overline{1, N_p}$ for prescribed displacement boundary conditions (N_p is the number of boundary nodes) and $i, j = 1, 2, 4$ for the periodic boundary conditions. In (64) the components of the tensors $\mathbf{K}^{(ij)}_{\mathrm{M}}$ are simply found in the tangent matrix $\underline{K}_{\mathrm{M}}$ (for displacement boundary conditions) or in the matrix $\underline{K}^{\star}_{\mathrm{M}}$ (for periodic boundary conditions) at the rows and columns of the degrees of freedom in the nodes i and j. For example, for the case of the periodic boundary conditions the total matrix

$\underline{K}^{\star}_{\mathrm{M}}$ has the format

$$\underline{K}^{\star}_{\mathrm{M}} = \begin{bmatrix} \begin{bmatrix} K_{11}^{(11)} & K_{12}^{(11)} \\ K_{21}^{(11)} & K_{22}^{(11)} \end{bmatrix} & \begin{bmatrix} K_{11}^{(12)} & K_{12}^{(12)} \\ K_{21}^{(12)} & K_{22}^{(12)} \end{bmatrix} & \begin{bmatrix} K_{11}^{(14)} & K_{12}^{(14)} \\ K_{21}^{(14)} & K_{22}^{(14)} \end{bmatrix} \\ \begin{bmatrix} K_{11}^{(21)} & K_{12}^{(21)} \\ K_{21}^{(21)} & K_{22}^{(21)} \end{bmatrix} & \begin{bmatrix} K_{11}^{(22)} & K_{12}^{(22)} \\ K_{21}^{(22)} & K_{22}^{(22)} \end{bmatrix} & \begin{bmatrix} K_{11}^{(24)} & K_{12}^{(24)} \\ K_{21}^{(24)} & K_{22}^{(24)} \end{bmatrix} \\ \begin{bmatrix} K_{11}^{(41)} & K_{12}^{(41)} \\ K_{21}^{(41)} & K_{22}^{(41)} \end{bmatrix} & \begin{bmatrix} K_{11}^{(42)} & K_{12}^{(42)} \\ K_{21}^{(42)} & K_{22}^{(42)} \end{bmatrix} & \begin{bmatrix} K_{11}^{(44)} & K_{12}^{(44)} \\ K_{21}^{(44)} & K_{22}^{(44)} \end{bmatrix} \end{bmatrix}, \tag{65}$$

where the superscripts in round brackets refer to the nodes and the subscripts to the degrees of freedom at those nodes. Then each submatrix in (65) may be considered as the representation of a second-order tensor $\mathbf{K}_{\mathrm{M}}^{(ij)}$.

Next, the expression for the variation of the nodal forces (64) is substituted into the relation for the variation of the macroscopic stress following from (50) or (57)

$$\delta\mathbf{P}_{\mathrm{M}} = \frac{1}{V_0}\sum_i\sum_j (\mathbf{K}_{\mathrm{M}}^{(ij)}\cdot\delta\vec{u}_{(j)})\vec{X}_{(i)}. \tag{66}$$

Substitution of the equation $\delta\vec{u}_{(j)} = \vec{X}_{(j)}\cdot\delta\mathbf{F}^{c}_{\mathrm{M}}$ into (66) gives

$$\delta\mathbf{P}_{\mathrm{M}} = \frac{1}{V_0}\sum_i\sum_j (\vec{X}_{(i)}\mathbf{K}_{\mathrm{M}}^{(ij)}\vec{X}_{(j)})^{LC} : \delta\mathbf{F}^{c}_{\mathrm{M}}, \tag{67}$$

where the superscript LC denotes left conjugation, which for a fourth-order tensor ${}^{4}\mathbf{T}$ is defined as $T^{LC}_{ijkl} = T_{jikl}$. Finally, by comparing (67) with (63) the consistent constitutive tangent is identified as

$${}^{4}\mathbf{C}^{\mathrm{P}}_{\mathrm{M}} = \frac{1}{V_0}\sum_i\sum_j (\vec{X}_{(i)}\mathbf{K}_{\mathrm{M}}^{(ij)}\vec{X}_{(j)})^{LC}. \tag{68}$$

If the macroscopic finite element scheme requires the constitutive tangent relating the variation of the macroscopic Cauchy stress to the variation of the macroscopic deformation gradient tensor according to

$$\delta\boldsymbol{\sigma}_{\mathrm{M}} = {}^{4}\mathbf{C}^{\sigma}_{\mathrm{M}} : \delta\mathbf{F}^{c}_{\mathrm{M}}, \tag{69}$$

this tangent may be obtained by varying the definition equation of the macroscopic Cauchy stress tensor (35), followed by substitution of (50) (or

(57)) and (67). This gives

$$\delta\boldsymbol{\sigma}_{\mathrm{M}} = \left[\frac{1}{V}\sum_i\sum_j (\vec{x}_{(i)}\mathbf{K}_{\mathrm{M}}^{(ij)}\vec{X}_{(j)})^{LC} + \frac{1}{V}\sum_i \vec{f}_{(i)}\mathbf{I}\,\vec{X}_{(i)} - \boldsymbol{\sigma}_{\mathrm{M}}\mathbf{F}_{\mathrm{M}}^{-c}\right] : \delta\mathbf{F}_{\mathrm{M}}^{c}, \tag{70}$$

where the expression in square brackets is identified as the required tangent stiffness tensor ${}^4\mathbf{C}_{\mathrm{M}}^{\sigma}$. In the derivation of (70) it has been used that in case of prescribed displacements of the RVE boundary (15) or of periodic boundary conditions (17), the initial and current volumes of an RVE are related according to $J_{\mathrm{M}} = \det(\mathbf{F}_{\mathrm{M}}) = V/V_0$.

5.4 Nested solution strategy

Based on the above developments the actual implementation of the computational homogenization strategy may be described by the following subsequent steps.

The macroscopic structure to be analyzed is discretized by finite elements. The external load is applied by an incremental procedure. Increments can be associated with discrete time steps. The solution of the macroscopic non-linear system of equations is performed in a standard iterative manner. To each macroscopic integration point a discretized RVE is assigned. The geometry of the RVE is based on the microstructural morphology of the material under consideration.

For each macroscopic integration point the local macroscopic deformation gradient tensor $\mathbf{F}_{\mathrm{M}}$ is computed from the iterative macroscopic nodal displacements (during the initialization step, zero deformation is assumed throughout the macroscopic structure, i.e. $\mathbf{F}_{\mathrm{M}} = \mathbf{I}$, which allows to obtain the initial macroscopic constitutive tangent). The macroscopic deformation gradient tensor is used to formulate the boundary conditions according to (37) or (42)–(43) to be applied on the corresponding representative cell.

The solution of the RVE boundary value problem employing a fine scale finite element procedure, provides the resulting stress and strain distributions in the microstructural cell. Using the resulting forces at the prescribed nodes, the RVE averaged first Piola-Kirchhoff stress tensor $\mathbf{P}_{\mathrm{M}}$ is computed according to (50) or (57) and returned to the macroscopic integration point as a local macroscopic stress. From the global RVE stiffness matrix the local macroscopic consistent tangent ${}^4\mathbf{C}_{\mathrm{M}}^{\mathrm{P}}$ is obtained according to (68).

When the analysis of all microstructural RVEs is finished, the stress tensor is available at every macroscopic integration point. Thus, the internal macroscopic forces can be calculated. If these forces are in balance with the external load, incremental convergence has been achieved and the next time increment can be evaluated. If there is no convergence, the pro-

cedure is continued to achieve an updated estimation of the macroscopic nodal displacements. The macroscopic stiffness matrix is assembled using the constitutive tangents available at every macroscopic integration point from the RVE analysis. The solution of the macroscopic system of equations leads to an updated estimation of the macroscopic displacement field. The solution scheme is summarized in Table 1. It is remarked that the two-level scheme outlined above can be used selectively depending on the macroscopic deformation, e.g. in the elastic domain the macroscopic constitutive tangents do not have to be updated at every macroscopic loading step.

Table 1. Incremental-iterative nested multi-scale solution scheme for the computational homogenization.

MACRO		**MICRO**
1. Initialization		
▷ initialize the macroscopic model		
▷ assign an RVE to every integration point		
▷ loop over all integration points		
set $\mathbf{F}_{\mathrm{M}} = \mathbf{I}$		Initialization RVE analysis
	$\mathbf{F}_{\mathrm{M}} \longrightarrow$	▷ prescribe boundary conditions
		▷ assemble the RVE stiffness
store the tangent	$\longleftarrow$ tangent	▷ calculate the tangent ${}^{4}\mathbf{C}_{\mathrm{M}}^{\mathrm{P}}$
▷ end integration point loop		
2. Next increment		
▷ apply increment of the macro load		
3. Next iteration		
▷ assemble the macroscopic tangent stiffness		
▷ solve the macroscopic system		
▷ loop over all integration points		
calculate $\mathbf{F}_{\mathrm{M}}$		RVE analysis
	$\mathbf{F}_{\mathrm{M}} \longrightarrow$	▷ prescribe boundary conditions
		▷ assemble the RVE stiffness
		▷ solve the RVE problem
store $\mathbf{P}_{\mathrm{M}}$	$\longleftarrow \mathbf{P}_{\mathrm{M}}$	▷ calculate $\mathbf{P}_{\mathrm{M}}$
store the tangent	$\longleftarrow$ tangent	▷ calculate the tangent ${}^{4}\mathbf{C}_{\mathrm{M}}^{\mathrm{P}}$
▷ end integration point loop		
▷ assemble the macroscopic internal forces		
4. Check for convergence		
▷ if not converged ⇒ step 3		
▷ else ⇒ step 2		

6 Example: two-scale coupled analysis in bending

As an example, the computational homogenization approach is applied to pure bending of a rectangular strip under plane strain conditions. Both the length and the height of the sample equal 0.2 m, the thickness is taken 1 m. The macromesh is composed of 5 quadrilateral 8 node plane strain reduced integration elements. The undeformed and deformed geometries of the macromesh are schematically depicted in figure 6. At the left side the strip is fixed in axial (horizontal) direction, the displacement in transverse (vertical) direction is left free. At the right side the rotation of the cross section is prescribed. As pure bending is considered the behaviour of the strip is uniform in axial direction and, therefore, a single layer of elements on the macrolevel suffices to simulate the situation.

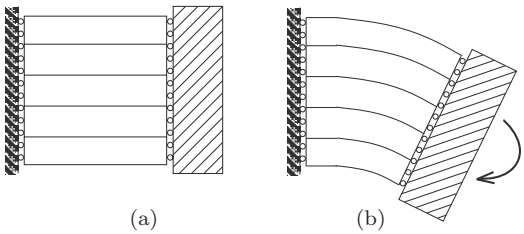
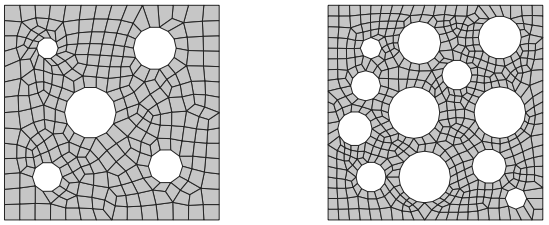


Figure 6. Schematic representation of the undeformed (a) and deformed (b) configurations of the macroscopically bended specimen.

In this example two heterogeneous microstructures consisting of a homogeneous matrix material with initially 12% and 30% volume fractions of voids are studied. To generate a random distribution of cavities in the matrix with a prescribed volume fraction, maximum diameter of holes and minimum distance between two neighbouring holes, for a two-dimensional RVE, the procedure from Hall (1991) and Smit (1998) has been adopted. The microstructural cells used in the calculations are presented in figure 7. It is worth mentioning that the absolute size of the microstructure is irrelevant for the first-order computational homogenization analysis (see also discussion in section 8).

The matrix material behaviour has been described by a modified elasto-visco-plastic Bodner-Partom model van der Aa et al. (2000). This choice is motivated by the intention to demonstrate that the method is well-suited for complex microstructural material behaviour, e.g. non-linear history and strain rate dependent at large strains. The material parameters for annealed aluminum AA 1050 determined in van der Aa et al. (2000) have been used; elastic parameters: shear modulus $G = 2.6 \times 10^4$ MPa, bulk modulus $K =$



(a) (b)

Figure 7. Microstructural cells used in the calculations with 12% voids (a) and 30% voids (b).

7.8×10^4 MPa and viscosity parameters: $\Gamma_0 = 10^8\,\mathrm{s}^{-2}$, $m = 13.8$, $n = 3.4$, $Z_0 = 81.4$ MPa, $Z_1 = 170$ MPa.

Micro-macro calculations for the heterogeneous structure, represented by the RVEs shown in figure 7 have been carried out, simulating pure bending at a prescribed moment rate equal to $5 \times 10^5\,\mathrm{N\,m\,s}^{-1}$. Figure 8 shows the distribution plots of the effective plastic strain for the case of the RVE with 12% volume fraction voids at an applied moment equal to $6.8 \times 10^5\,\mathrm{N\,m}$ in the deformed macrostructure and in three deformed, initially identical RVEs at different locations in the macrostructure. Each hole acts as a plastic strain concentrator and causes higher strains in the RVE than those occurring in the homogenized macrostructure. In the present calculations the maximum effective plastic strain in the macrostructure is about 25%, whereas at RVE level this strain reaches 50%. It is obvious from the deformed geometry of the holes in figure 8 that the RVE in the upper part of the bended strip is subjected to tension and the RVE in the lower part to compression, while the RVE in the vicinity of the neutral axis is loaded considerably milder than the other RVEs. This confirms the conclusion that the method realistically describes the deformation modes of the microstructure.

In figure 9 the moment-curvature (curvature defined for the bottom edge of the specimen) diagram resulting from the computational homogenization approach is presented. To give an impression of the influence of the holes also the response of a homogeneous configuration (without cavities) is shown. It can be concluded that even the presence of 12% voids induces a reduction of the bending moment (at a certain curvature) of more than 25% in the plastic regime. This significant reduction in the bending moment may be attributed to the formation of microstructural shear bands, which are clearly observed in figure 8. This indicates that in order to capture such an effect a detailed microstructural analysis is required. A straightforward

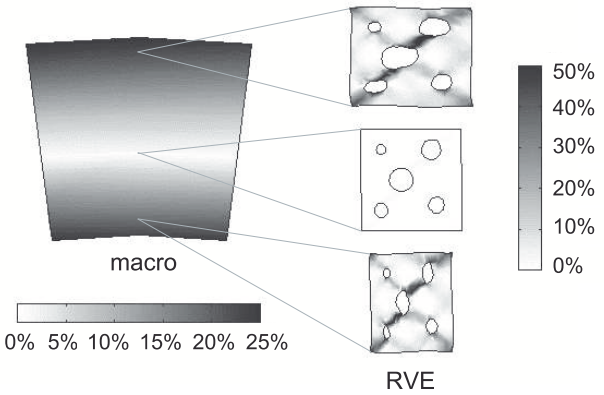


Figure 8. Distribution of the effective plastic strain in the deformed macrostructure and in three deformed RVEs, corresponding to different points of the macrostructure.

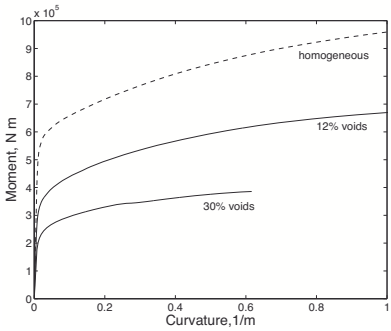


Figure 9. Moment-curvature diagram resulting from the first-order computational homogenization analysis.

application of, for example, the rule of mixtures would lead to erroneous results.

7 The RVE in first-order computational homogenization

7.1 General concept of an RVE

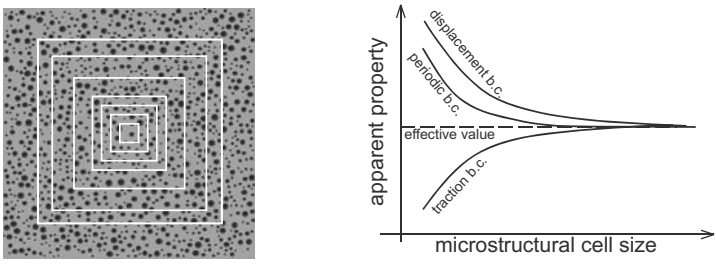
The computational homogenization approach, as well as most of other homogenization techniques, are based on the concept of a representative volume element (RVE). An RVE is a model of a material microstructure to be used to obtain the response of the corresponding homogenized macroscopic continuum in a macroscopic material point. Thus, the proper choice of the RVE largely determines the accuracy of the modelling of a heterogeneous material.

There appear to be two significantly different ways to define a representative volume element Drugan and Willis (1996). The first definition requires an RVE to be a statistically representative sample of the microstructure, i.e. to include virtually a sampling of all possible microstructural configurations that occur in the composite. Clearly, in the case of a non-regular and non-uniform microstructure such a definition leads to a considerably large RVE. Therefore, RVEs that rigorously satisfy this definition are rarely used in actual homogenization analyses. This concept is usually employed when a computer model of the microstructure is being constructed based on experimentally obtained statistical information (see e.g. Shan and Gokhale (2002)).

Another definition characterizes an RVE as the smallest microstructural volume that sufficiently accurately represents the overall macroscopic properties of interest. This usually leads to much smaller RVE sizes than the statistical definition described above. However, in this case the minimum required RVE size also depends on the type of material behaviour (e.g. for elastic behaviour usually much smaller RVEs suffice than for plastic behaviour), macroscopic loading path and difference of properties between heterogeneities. Moreover, the minimum RVE size, that results in a good approximation of the overall material properties, does not always lead to adequate distributions of the microfields within the RVE. This may be important if, for example, microstructural damage initiation or evolving microstructures are of interest.

The latter definition of an RVE is closely related to the one established by Hill Hill (1963), who argued that an RVE is well-defined if it reflects the material microstructure and if the responses under uniform displacement and traction boundary conditions coincide. If a microstructural cell does not contain sufficient microstructural information, its overall responses under uniform displacement and traction boundary conditions will differ. The homogenized properties determined in this way are called “apparent”, a no-

tion introduced by Huet Huet (1990). The apparent properties obtained by application of uniform displacement boundary conditions on a microstructural cell usually overestimate the real effective properties, while the uniform traction boundary conditions lead to underestimation. As has been verified by a number of authors van der Sluis et al. (2000); Terada et al. (2000), for a given microstructural cell size, the periodic boundary conditions provide a better estimation of the overall properties, than the uniform displacement and uniform traction boundary conditions. This conclusion also holds if the microstructure does not really possess geometrical periodicity Terada et al. (2000). Increasing the size of the microstructural cell leads to a better estimation of the overall properties, and, finally, to a "convergence" of the results obtained with the different boundary conditions to the real effective properties of the composite material, as schematically illustrated in figure 10. The convergence of the apparent properties towards the effective ones at increasing size of the microstructural cell has been investigated in Huet (1990, 1999); Ostoja-Starzewski (1998, 1999); Pecullan et al. (1999); Terada et al. (2000).



(a) (b)

Figure 10. (a) Several microstructural cells of different sizes. (b) Convergence of the apparent properties to the effective values with increasing microstructural cell size for different types of boundary conditions.

7.2 Unit cells versus RVEs

In practice, instead of a representative volume element, a unit cell is often used as a microstructural model, since it requires substantially less computational effort. This section examines the possible error, which is made in the obtained overall response of a multi-phase material, if the analysis is performed on a unit cell instead of an RVE.

As the simplest unit cell, a piece (for example a square or cube) of the matrix material containing a single heterogeneity (e.g. inclusion or void)

could be suggested. The use of such a unit cell implicitly assumes a regular arrangement of the heterogeneities in the matrix, which contradicts the observations that almost all materials have a non-periodic or even spatially random microstructural composition. Examples are precipitates in metal alloys arranged randomly by their nature and artificial fiber reinforced composites, possessing a non-regular distribution of the fibers due to the production process. At the same time, several experimental evidences exist showing that the spatial variability in the microstructure significantly influences the overall behaviour and particularly the fracture characteristics of composites, as reported in Mackay (1990); Barsoum et al. (1992).

Different authors, e.g. Brockenbrough et al. (1991); Nakamura and Suresh (1993); Ghosh et al. (1996); Moulinec and Suquet (1998), have performed a comparison of the overall composite responses resulting from the modelling of regular and random structures. They have found a significant response difference in the plastic regime, while there is almost no deviation in elastic regime. Also it has been shown Smit et al. (1999), that softening behaviour of a regularly composed structure may change to hardening in the case of a random composition. Most of these considerations, except for the latter, have been performed for small deformations, very simple elasto-plastic behaviour and relatively stiff inclusions (fibers). In this section the overall behaviour of regular and random structures is compared at large deformations, non-linear history dependent material behaviour, for voided material (an appropriate approximation for material with soft inclusions). Apart from the calculations on the microstructural cell (tensile configuration), also a full multi-scale analysis (pure bending) of both regular and random structures is presented.

A material with a 12% volume fraction of voids is considered. The regularly stacked structure is modelled by a square unit cell containing a single hole (figure 11a). For the modelling of a random structure 10 different unit cells with non-regular arrangements of voids with a distribution of void sizes have been generated (figure 11b). The averaged behaviour of these 10 unit cells is expected to be representative for the real random structure with a given volume fraction of heterogeneities. Using several small non-regular unit cells instead of one larger RVE also allows to estimate the amount of deviation of the apparent properties obtained by the unit cell modelling, from the effective values for different types of material models and loading histories.

In the subsequent sections a comparison is performed for three different constitutive models of the matrix material: hyper-elastic, elasto-visco-plastic with hardening and elasto-visco-plastic with intrinsic softening. First uniaxial extension (under plane strain conditions) of a macroscopic sample

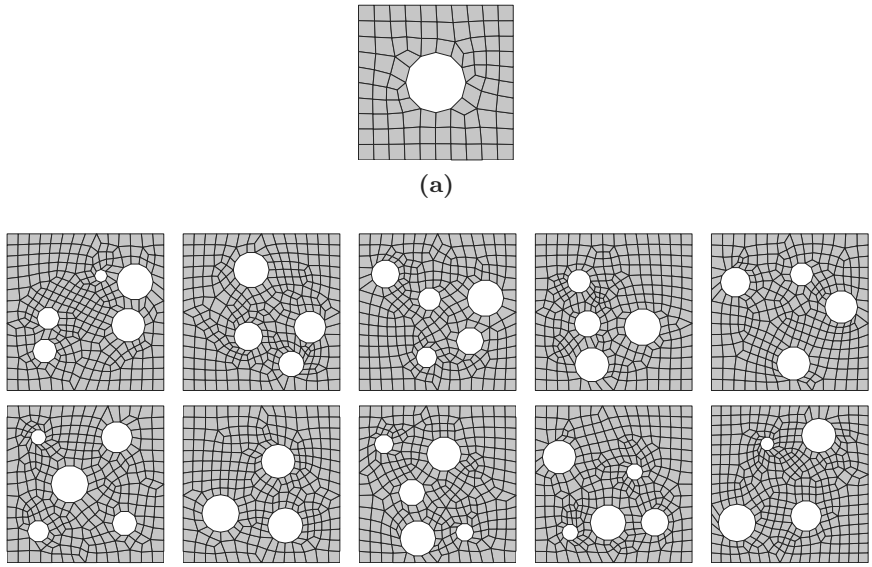


Figure 11. Unit cell with one hole (a), representing a regular structure, and 10 randomly composed unit cells (b).

is considered. Because in this case the macroscopic deformation field is homogeneous a full micro-macro modelling is not necessary and an analysis of an isolated unit cell with adequate boundary conditions (periodic) suffices. In the last section the results of a micro-macro simulation of bending using random and regular microstructures are compared.

Elastic behaviour, tension First, a comparison of the overall behaviour of regular and random structures is carried out for the case of hyperelastic behaviour of the matrix material, modelled as a compressible Neo-Hookean material. The material parameters used in the calculations are $K = 2667\,\mathrm{MPa}$, $G = 889\,\mathrm{MPa}$.

Figure 12 shows the stress-strain curves for the unit cells with regular and random void stacking. For small deformations there is almost no difference in the responses originating from the regular and random void distributions. This result is in agreement with the experiences reported in the literature for small deformations, see, e.g. Brockenbrough et al. (1991);

Nakamura and Suresh (1993); Moulinec and Suquet (1998). For large deformations the stiffer behaviour of the regular structure becomes a little bit more pronounced, however, the deviations remain small. The difference between the response of the regular structure and the response averaged over the random unit cells does not exceed 2%. This small deviation is explained by figure 13, presenting the distribution of the equivalent von Mises stress in the regular unit cell and in a random unit cell for 20% macroscopic strain. The stress field around any hole of the random structure is almost the same as around the hole of the regular structure, which indicates little interaction between voids. If only the averaged elastic constants are of interest, it is concluded that calculations performed on the simplest regular unit cell usually provide an answer within an acceptable tolerance.

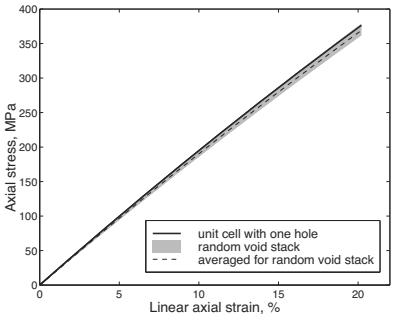


Figure 12. Tensile stress-strain responses (unit cell averages) of the regular and random structures in a voided hyper-elastic matrix material.

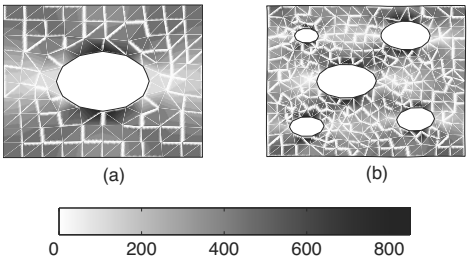


Figure 13. Distribution of the equivalent von Mises stress (MPa) in the deformed regular (a) and random (b) structures in a voided hyper-elastic matrix material.

Elasto-visco-plastic behaviour with hardening, tension The influence of the randomness of the microstructure on the macroscopic response becomes more significant when plastic yielding of one or more constituents occurs. This section investigates the responses of the regular and random unit cells under tensile loading when the matrix material exhibits elasto-visco-plastic behaviour with hardening. The constitutive description is given by the Bodner-Partom model van der Aa et al. (2000). The material parameters are the same as those used in section 6. The unit cells are subjected to uniaxial tension at a constant strain rate of $0.5\,\mathrm{s}^{-1}$.

In figure 14 the stress-strain curves are presented. In this case the difference between the overall response of the regular structure and the averaged response of the random structures reaches 10%. The rather large scattering in the responses of different random cells is due to the small number of voids included. As has been demonstrated in Smit (1998), the scattering is significantly reduced if microstructural cells contain more heterogeneities. The averaged response is, however, hardly affected, provided that a sufficient number of random realizations has been considered.

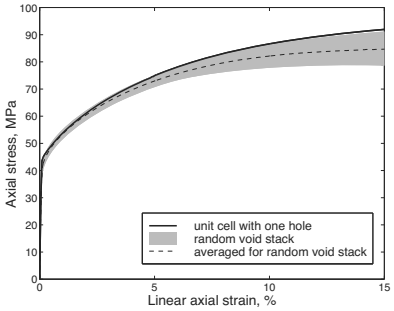


Figure 14. Tensile stress-strain responses (unit cell averages) of the regular and random structures for an elasto-visco-plastic matrix material with hardening.

The fundamental mechanism that governs the difference between the response of the regular structure and the averaged response of the random structures is illustrated in figure 15, where the distribution of the effective plastic strain in the deformed regular and random unit cells at 15% applied macroscopic strain is presented. In the regular unit cell the ligaments yield simultaneously rather than sequentially with increasing macroscopic strain, which is the case for the random unit cell. As a result, at the same value of the macroscopic strain the regular unit cell is deformed relatively smoothly,

while some ligaments in the random unit cell have already accumulated a significant amount of plastic strain. Consequently, the regular unit cell (in fact a structure with a periodic stacking of heterogeneities) has a larger overall stiffness than a random configuration.

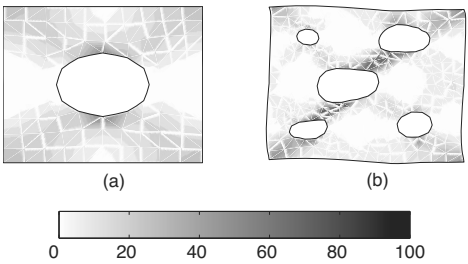


Figure 15. Distribution of the effective plastic strain in the deformed regular (a) and random (b) structures for an elasto-visco-plastic matrix material with hardening.

Elasto-visco-plastic behaviour with softening, tension The difference in yielding mechanisms for regular and random microstructures outlined in the previous section causes not only a quantitative deviation in the responses of these structures (as illustrated by figure 14), but in some cases also the qualitative character changes, as has been shown in Smit et al. (1999). For example, such a phenomenon can be observed when the matrix material is described by a generalized compressible Leonov model with intrinsic softening and subsequent hardening. The model is designed for the plastic deformation of polymers and incorporates a stress dependent Eyring viscosity extended by pressure dependence and intrinsic softening effects. Details of this model can be found in Baaijens (1991); Tervoort (1996); Govaert et al. (2000).

The resulting stress-strain curves for uniaxial tension of polycarbonate at a constant strain rate of $0.01\,\mathrm{s}^{-1}$ are given in figure 16. The overall behaviour of the regular structure in the plastic regime exhibits some initial softening followed by hardening. The response of the regular structure is, in fact, similar to the response of one single ligament, that softens according to the intrinsic material behaviour. A completely different response can be observed for the random configurations. Although some of the random unit cells also demonstrate some softening behaviour, originating from the relatively simple composition of the unit cells used in the calculations, the average response of the random unit cells does not show any softening but

exhibits continuous hardening. This is caused by the sequential appearance of elastic, softening and hardening zones within the random microstructure.

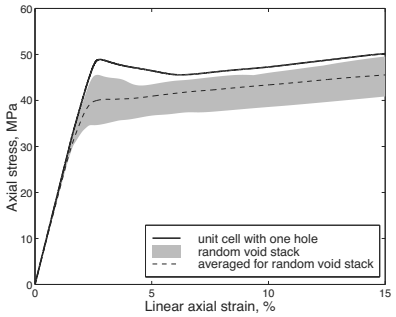


Figure 16. Tensile stress-strain responses (RVE averages) of the regular and random structures for an elasto-visco-plastic matrix material with intrinsic softening and subsequent hardening.

This example illustrates that the overall response of heterogeneous materials, when detersi3joda!mined from a modelling by a regular structure, should be interpreted with great care, particularly in the case of complex material behaviour (e.g. in case of softening followed by hardening or vice versa).

Elasto-visco-plastic behaviour with hardening, bending The comparison of the overall behaviour of the regular and random microstructures performed above has been based on the averaged behaviour of a single unit cell subjected to a particular loading history (uniaxial tension). The question remains how the randomness of the microstructure does influence the overall behaviour when a macroscopic sample is deformed heterogeneously, so that potentially every material point of the sample is subjected to a different loading history. In order to investigate this item the computational homogenization approach is a helpful tool.

As an example the influence of the spatial composition of the microstructure on the overall moment-curvature response of the voided material under pure bending is studied. The behaviour of the matrix material is described by the Bodner-Partom elasto-visco-plastic model with hardening. The macrogeometry and the material parameters are the same as these used in section 6.

Figure 17 shows the moment-curvature diagram resulting from the full

micro-macro analysis of pure bending of the material using the regular and the random microstructures. Again, the regular structure exhibits a stiffer response than the averaged random result, while the maximum deviation is only about 5%, which is considerably less than for the tensile test with the same material behaviour (figure 14). This smaller deviation originates from the fact that in case of bending all the unit cells assigned to the various macroscopic points over the height of the bended strip are loaded differently, see figure 8. The unit cell at the top of the bended strip experiences tension, so that the observations dealt with in the previous examples apply. At the same time, there are also unit cells that are stretched less or still are in elastic regime, like for example the one in the vicinity of the neutral line, so that in average for the whole bending process the influence of randomness can be expected to be smaller than for uniaxial extension.

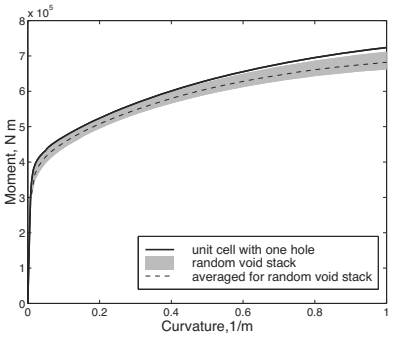


Figure 17. Moment-curvature responses of the regular and random structures for an elasto-visco-plastic matrix material with hardening.

8 Second-order computational homogenization

In spite of the attractive characteristics listed above, there are a few important limitations of the first-order framework, which can be summarized as follows

- The principle of separation of scales must be respected. Hence, the characteristic length that characterizes the spatial variations of the macroscopic loading must be very large with respect to the size of the microstructure. As a consequence, only simple first-order deformation modes (tension, compression, shear or combinations thereof) of the microstructure can be retrieved. The case shown in figure 8, which is

a typical bending mode, which from a physical point of view should appear for small, but finite, microstructural cells, cannot be found.
- The framework is completely insensitive to the absolute size of the microstructural constituents (scale independent). Size effects emanating from the absolute size at the micro scale cannot be dealt with properly.
- Macroscopic gradients must remain very small with respect to the micro scale. Localization problems, where non-uniform macroscopic deformations arise, cannot be solved properly.

Whenever strong gradients appear at the macro-level (localization, size effects) care must be taken in using a first-order scheme. In all other cases, one should continue using it and not jump to a second-order scheme for which an additional price in complexity and computational costs is to be paid.

In order to overcome these shortcomings, the computational homogenization methodology has been extended recently to higher-order continua (Geers et al., 2001; Kouznetsova et al., 2002; Kouznetsova, 2002; Geers et al., 2003; Kouznetsova et al., 2004b,a). In this course, the methodology and the essential parts of the multi-scale kinematics and statics will be outlined briefly, whereas more details can be found in the cited references. The method is next applied to heterogeneous multi-phase microstructures, as typically the case in most metals, polymer blends and composites. Some comments on the parallel implementation of the multi-scale technique are given and an illustrative example is used to scrutinize the added value of the second-order framework in relation to the more standard first-order scheme.

8.1 Principles

The second-order case, which may be considered as a generalization of the classical first-order scheme, departs from a Taylor series expansion of the classical nonlinear deformation map, $\vec{x} = \phi(\vec{X})$, applied to a finite material vector $\Delta\vec{x}$ in the deformed state:

$$\Delta\vec{x} \;=\; \mathbf{F}_{\mathrm{M}}{\cdot}\Delta\vec{X} + \tfrac{1}{2}\Delta\vec{X}{\cdot}{}^{3}\mathbf{G}_{\mathrm{M}}{\cdot}\Delta\vec{X} + \mathcal{O}\left(\Delta\vec{X}^{3}\right) \tag{71}$$

Using this Taylor series expansion, the macroscopic (coarse scale) kinematics is determined through the deformation gradient tensor $\mathbf{F}_{\mathrm{M}}$ and its Lagrangian gradient ${}^{3}\mathbf{G}_{\mathrm{M}} = \vec{\nabla}_{0,\mathrm{M}}\mathbf{F}_{\mathrm{M}}$. The key point in the second-order two-scale framework, resides in applying relation (71) to a representative part of the microstructure, such that a classical boundary value problem is obtained at the micro scale (or fine scale). The scale bridging is then realized through the application of averaging theorems. This is schematically depicted in figure 18. Note that the tensor ${}^{3}\mathbf{G}_{\mathrm{M}}$ has a minor symmetry,

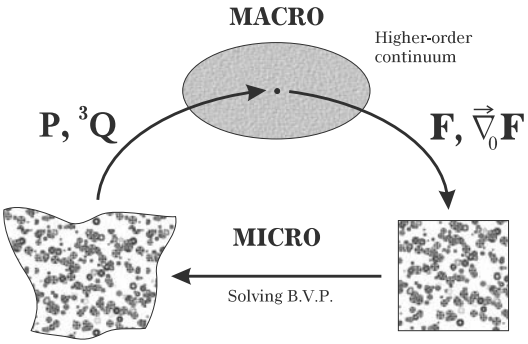


Figure 18. Second-order computational homogenization

${}^3\mathbf{G}_{\rm M} = {}^3\mathbf{G}_{\rm M}^T$ (or $G_{Mijk} = G_{Mkji}$ in index notation), which is used throughout this chapter.

8.2 Two-scale higher-order kinematics

In order to apply equation (71) to the fine scale, all higher-order terms (represented by $\mathcal{O}(\Delta\vec{X}^3)$) are condensed into an unknown microfluctuation field $\vec{w}$, which represents the fine scale contribution in the kinematics. Hence,

$$\Delta\vec{x} \;=\; \mathbf{F}_{\rm M}{\cdot}\Delta\vec{X} + \tfrac{1}{2}\Delta\vec{X}{\cdot}{}^3\mathbf{G}_{\rm M}{\cdot}\Delta\vec{X} + \vec{w} \tag{72}$$

Applying this to an undeformed volume V_0 (the RVE) with a geometrical center $\vec{X}_c$ that is located in $\vec{x}_c$ after deformation gives (notice the similarity and differences with the elaboration in the previous section, equation (5)).

$$\vec{x} - \vec{x}_c \;=\; \mathbf{F}_{\rm M}{\cdot}(\vec{X} - \vec{X}_c) + \tfrac{1}{2}(\vec{X} - \vec{X}_c){\cdot}{}^3\mathbf{G}_{\rm M}{\cdot}(\vec{X} - \vec{X}_c) + \vec{w} \tag{73}$$

Eliminating rigid body displacements like for the first-order case (e.g. by fixing a boundary point 1) then leads to

$$\vec{x} \;=\; \vec{c} + \mathbf{F}_{\rm M}{\cdot}(\vec{X} - \vec{X}_c) + \tfrac{1}{2}(\vec{X} - \vec{X}_c){\cdot}{}^3\mathbf{G}_{\rm M}{\cdot}(\vec{X} - \vec{X}_c) + (\vec{w} - \vec{w}_1) \tag{74}$$

with

$$\vec{c} \;= \vec{X}_1 + \mathbf{F}_{\rm M}{\cdot}(\vec{X}_1 - \vec{X}_c) + \tfrac{1}{2}(\vec{X}_1 - \vec{X}_c){\cdot}{}^3\mathbf{G}_{\rm M}{\cdot}(\vec{X}_1 - \vec{X}_c) \tag{75}$$

$$\vec{x}_c \;= \vec{c} - \vec{w}_1 \tag{76}$$

The microscopic deformation gradient tensor $\mathbf{F}_{\rm m}$ is easily reconstructed as

$$\begin{aligned} \mathbf{F}_{\rm m} \;&= (\vec{\nabla}_{0,{\rm m}}\vec{x})^T \\ &= \mathbf{F}_{\rm M} + (\vec{X} - \vec{X}_c){\cdot}{}^3\mathbf{G}_{\rm M} + (\vec{\nabla}_{0,{\rm m}}(\vec{w} - \vec{w}_1))^T \end{aligned} \tag{77}$$

Applying the earlier introduced scale transition relation (13) with respect to equation (77) then leads to two kinematical constraints (to be imposed on the RVE)

$$\frac{1}{V_0}\int_{V_0}(\vec{X}-\vec{X}_c)dV_0 \;=\; \vec{0} \tag{78}$$

$$\frac{1}{V_0}\int_{V_0}(\vec{\nabla}_{0,\mathrm{m}}(\vec{w}-\vec{w}_1))^T\,dV_0 \;=\; \frac{1}{V_0}\int_{\Gamma_0}(\vec{w}-\vec{w}_1)\vec{N}\,d\Gamma_0 = \mathbf{0} \tag{79}$$

where the divergence theorem was used to derive the latter relation. Equation (78) is clearly satisfied here, since the Taylor series has been expanded with respect to the geometrical centre $\vec{X}_c$ in equation (73). This appears to be a necessary condition in the second-order case, which deviates from the first-order scheme where any point to develop (4) around (instead of $\vec{X}_c$) gives the same result. The second constraint (79) applies to the unknown fluctuation field. Logically, the integral involves $(\vec{w}-\vec{w}_1)$, which implies a constraint on the boundary position vectors $\vec{x}$ through (74). There are various ways to make this boundary integral zero, e.g. by constraining $(\vec{w}-\vec{w}_1) = \vec{0}$ for all points of the RVE (Taylor/Voigt), or by constraining $(\vec{w}-\vec{w}_1) = \vec{0}$ at the boundary of the RVE only (displacement or kinematic boundary condition), or through the application of periodic boundary conditions on the microfluctuation field (the macroscopic field is generally not periodic in the second-order case!). The latter conditions are used here, leading to the following microperiodicity equations valid between the left(L)-right(R) and bottom(B)-top(T) boundaries of a two-dimensional rectangular RVE as shown in figure 19.

$$\vec{w}_L \;=\; \vec{w}_R \qquad\qquad \vec{w}_B \;=\; \vec{w}_T \tag{80}$$

Note that again all equations involve the microfluctuation field with respect to $\vec{w}_1$. Any choice for $\vec{w}_1$ will then lead to the same solution (except for $\vec{x}_c$). This is also obvious from the constraint relation (79), which can be easily elaborated to a format in which the contribution of $\vec{w}_1$ vanishes, i.e.

$$\int_{\Gamma_0}(\vec{w}-\vec{w}_1)\vec{N}\,d\Gamma_0 \;=\; \int_{\Gamma_0}\vec{w}\vec{N}\,d\Gamma_0 \;=\; \mathbf{0} \tag{81}$$

It is easy to show that the micro scale problem defined by the equations (72), (78), (80), applied to the rectangular 2D RVE depicted in figure 19 with periodic microfluctuations, fully determines the kinematics of the four corner points (Geers et al., 2001; Kouznetsova et al., 2002). This

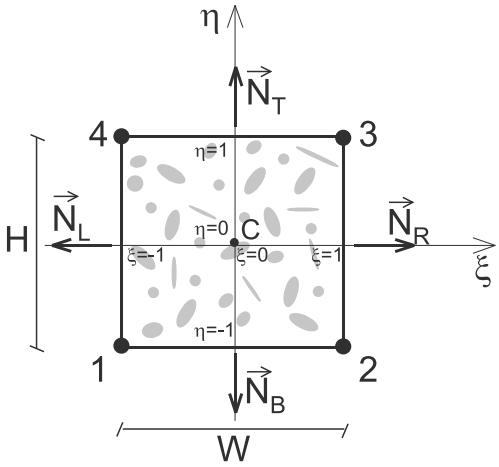


Figure 19. Lagrangian undeformed 2D reference RVE

set of equations imposes 8 macroscopic degrees-of-freedom to the 2D microstructure, whereas the full macroscopic kinematics consists of 12 degrees-of-freedom (2 rigid body displacement, 4 degrees-of-freedom in $\mathbf{F}_{\mathrm{M}}$ and 6 degrees-of-freedom in the minor-symmetric ${}^{3}\mathbf{G}_{\mathrm{M}}$). The missing kinematical quantities appear to be the stretch gradients (Geers et al., 2001; Kouznetsova et al., 2002), i.e. so far an RVE with 8 macroscopic degrees-of-freedom has been established, where the displacements are prescribed through the four corner nodes. This is a typical example of couple stress homogenization.

In order to incorporate the entire gradient field, the set of averaging relations needs to be completed in order to account for the missing stretch gradient degrees-of-freedom of ${}^{3}\mathbf{G}_{\mathrm{M}}$. On the basis of the Taylor series expansion (72), it is easy to show that the following averaging theorem can be derived (by means of some manipulations of the equations given in Kouznetsova et al. (2002)), relating ${}^{3}\mathbf{G}_{\mathrm{M}}$ to the position vectors $\vec{x}$ (implicitly incorporating the fine scale contribution through (74)) of all material

points in the the square RVE with initial size W and volume V_0:

$$\frac{W^2}{12}\left[{}^3\mathbf{G}_{\mathrm{M}} + \tfrac{1}{2}\left(\mathbf{II} : {}^3\mathbf{G}_{\mathrm{M}}^{RT}\right)^{RT}\right] + (\mathbf{I}\vec{c})^{RT} = \frac{1}{2V_0}\int_{V_0}\left(\vec{\nabla}_{0,\mathrm{m}}(\vec{x}\vec{X}) + [\vec{\nabla}_{0,\mathrm{m}}(\vec{x}\vec{X})]^T\right)dV_0 \quad (82)$$

In here, $\mathbf{I}$ is the second-order unit tensor, and ^{RT} stands for the right transpose, i.e. $T_{ijk}^{RT} = T_{ikj}$ in index notation. The third term in the left-hand side of this equation is present to account for the deformed position $\vec{x}_c$ of the center of the undeformed RVE, which is generally no longer the center of the deformed RVE. Computing the integral in the right-hand side of the latter equation through substitution of equation (74), reveals

$$\frac{1}{2V_0}\int_{V_0}\left(\vec{\nabla}_{0,\mathrm{m}}(\vec{x}\vec{X}) + [\vec{\nabla}_{0,\mathrm{m}}(\vec{x}\vec{X})]^T\right)dV_0 = \frac{W^2}{12}{}^3\mathbf{G}_{\mathrm{M}} + \frac{W^2}{24}\left(\mathbf{II} : {}^3\mathbf{G}_{\mathrm{M}}^{RT}\right)^{RT} + [\mathbf{I}\vec{c}]^{RT} + \frac{1}{2V_0}\int_{\Gamma_0}\left(\vec{X}(\vec{w}-\vec{w}_1)\vec{N} + \vec{N}(\vec{w}-\vec{w}_1)\vec{X}\right)d\Gamma_0 \quad (83)$$

Enforcing the averaging relation (82) requires that the last integral in (83) should vanish, which leads to a new constraint on the microfluctuation field.

$$\int_{\Gamma_0}\left(\vec{X}(\vec{w}-\vec{w}_1)\vec{N} + \vec{N}(\vec{w}-\vec{w}_1)\vec{X}\right)d\Gamma_0 = {}^3\mathbf{0} \quad (84)$$

This boundary integral clearly incorporates $(\vec{w}-\vec{w}_1)$, which constrains the position vectors $\vec{x}$ of the boundary points through (74). The microfluctuation $\vec{w}_1$ (in the fixed boundary point) cannot be eliminated in general as done in the previously introduced boundary integral (81). If constraint (84) is enforced, it is easy to rewrite the averaging equation (82) as a boundary integral

$${}^3\mathbf{G}_{\mathrm{M}} + \tfrac{1}{2}\left(\mathbf{II} : {}^3\mathbf{G}_{\mathrm{M}}^{RT}\right)^{RT} + \frac{12}{W^2}[\mathbf{I}\vec{c}]^{RT} = \frac{6}{V_0 W^2}\int_{\Gamma_0}\left(\vec{X}\vec{x}\vec{N} + \vec{N}\vec{x}\vec{X}\right)d\Gamma_0 \quad (85)$$

Equation (85) typically illustrates that $^3\mathbf{G}_{\mathrm{M}}$ is imposed on the RVE boundary, which is necessary to construct a classical boundary value problem at the micro scale.

For an initially square RVE ($H = W$ in figure 19), on which the microperiodicity equations (80) for the microfluctuation field hold, the constraint (84) can be simplified to

$$\begin{aligned} \int_{\Gamma_{0L}} (\vec{w}_L - \vec{w}_1) d\Gamma_0 = \vec{0} \\ \int_{\Gamma_{0B}} (\vec{w}_B - \vec{w}_1) d\Gamma_0 = \vec{0} \end{aligned} \tag{86}$$

Clearly, these two conditions enforce the shape of a part of the boundary to be equal in average to the shape ensuing from the macroscopic field. Prescribing (86) at the boundaries can be done by generalized displacement constraints (non-homogeneous tying relations), see (Kouznetsova et al., 2002; Kouznetsova, 2002) for more details on this topic. Again, it is obvious that imposing $\vec{w}_1 = \vec{0}$ does not influence the solution for the two-scale homogenization.

Note that the macroscopic ${}^3\mathbf{G}_{\mathrm{M}}$ is not the volume average of the microscopic ${}^3\mathbf{G}_{\mathrm{m}} = \vec{\nabla}_{0,\mathrm{m}}\mathbf{F}_{\mathrm{m}}$. This not possible if one wants to construct a classical boundary value problem at the micro scale. The scale transition is here driven by boundary integrals involving displacements of boundary points of the RVE only. Enforcing ${}^3\mathbf{G}_{\mathrm{M}}$ to be the volume average of ${}^3\mathbf{G}_{\mathrm{m}}$ would lead to higher-order boundary conditions on the microstructural fluctuation field, which would make the fine scale problem second-order as well.

8.3 Extracting stress tensors

The macroscopic stress quantities are next extracted from the analysis of the deformed RVE by equating the macroscopic work per unit of volume to the average work performed on the RVE (Hill-Mandel or macrohomogeneity condition). For the second-order case, this condition reads

$$\frac{1}{V_0}\int_{V_0} \mathbf{P}_{\mathrm{m}} : \delta\mathbf{F}_{\mathrm{m}}^T dV_0 \; = \mathbf{P}_{\mathrm{M}} : \delta\mathbf{F}_{\mathrm{M}}^T + {}^3\mathbf{Q}_{\mathrm{M}} \vdots \delta{}^3\mathbf{G}_{\mathrm{M}} \tag{87}$$

In here, $\mathbf{P}_{\mathrm{M}}$ is the macroscopic first Piola-Kirchhoff stress tensor, $\mathbf{P}_{\mathrm{m}}$ its microstructural counterpart and ${}^3\mathbf{Q}_{\mathrm{M}}$ the higher-order stress tensor which is work-conjugated to ${}^3\mathbf{G}_{\mathrm{M}}$. Note that equation (87) in fact defines the two macroscopic stress tensors $\mathbf{P}_{\mathrm{M}}$ and ${}^3\mathbf{Q}_{\mathrm{M}}$.

The microstructural work (per unit of volume in the reference state) can

be written as

$$\delta W_{0\mathrm{M}} = \frac{1}{V_0}\int\limits_{V_0} \mathbf{P}_{\mathrm{m}} : \delta\mathbf{F}^{c}_{\mathrm{m}}\mathrm{d}V_0 = \frac{1}{V_0}\int\limits_{\Gamma_0} \vec{p}\cdot\delta\Delta\vec{x}\ \mathrm{d}\Gamma_0, \tag{88}$$

where use has been made of the divergence theorem and the static equilibrium equation in the microstructure (2). Taking the variation of the position vector $\delta\Delta\vec{x}$ according to (72) leads to

$$\delta\Delta\vec{x} = \delta\mathbf{F}_{\mathrm{M}}\cdot\vec{X} + \tfrac{1}{2}\vec{X}\cdot\delta^3\mathbf{G}_{\mathrm{M}}\cdot\vec{X} + \delta\Delta\vec{w}, \tag{89}$$

which after substitution in equation (88) yields

$$\delta W_{0\mathrm{M}} = \frac{1}{V_0}\int\limits_{\Gamma_0} \vec{p}\vec{X}\ \mathrm{d}\Gamma_0 : \delta\mathbf{F}^{c}_{\mathrm{M}} + \frac{1}{2V_0}\int\limits_{\Gamma_0} \vec{X}\vec{p}\vec{X}\ \mathrm{d}\Gamma_0\ \vdots\ \delta^3\mathbf{G}_{\mathrm{M}} + \frac{1}{V_0}\int\limits_{\Gamma_0} \vec{p}\cdot\delta\Delta\vec{w}\ \mathrm{d}\Gamma_0. \tag{90}$$

Since the boundary constraints (80) do not contribute to the total work and accounting for (86), the last term in (90) can be proven to disappear

$$\int\limits_{\Gamma_0} \vec{p}\cdot\delta\Delta\vec{w}\ \mathrm{d}\Gamma_0 = 0, \tag{91}$$

manifesting the fact that the microstructural fluctuation field does not affect the average variation of the microscopic work.

Elaboration of this equation leads to two boundary integrals that permit to compute the stress tensors $\mathbf{P}_{\mathrm{M}}$ and ${}^3\mathbf{Q}_{\mathrm{M}}$:

$$\mathbf{P}_{\mathrm{M}} = \frac{1}{V_0}\int\limits_{\Gamma_0} \vec{p}\vec{X}\ d\Gamma_0 \tag{92}$$

$${}^3\mathbf{Q}_{\mathrm{M}} = \frac{1}{2V_0}\int\limits_{\Gamma_0} \vec{X}\vec{p}\vec{X}\ \mathrm{d}\Gamma_0 \tag{93}$$

Both stress tensors can be easily computed once the boundary value problem on the micro scale has been solved.

The above formulas relate the macroscopic stress tensor and the macroscopic higher-order stress tensor to microstructural variables defined on the RVE boundary. The relations (92) and (93) can also be transformed into volume integrals, allowing the macroscopic stress measures to be expressed in terms of volume averages of microstructural quantities. The macroscopic

stress tensor $\mathbf{P}_{\mathrm{M}}$ again equals the volume average of the microscopic stress tensor $\mathbf{P}_{\mathrm{m}}$

$$\mathbf{P}_{\mathrm{M}} = \frac{1}{V_0}\int_{V_0} \mathbf{P}_{\mathrm{m}} dV_0 \tag{94}$$

The proof of this equation is identical to that for the first-order framework (for the derivation see (32)-(33)).

The derivation for the higher-order stress tensor ${}^3\mathbf{Q}_{\mathrm{M}}$ follows the same procedure. Applying the divergence theorem to transform the boundary integral in (93) to a volume integral gives

$$\begin{aligned} {}^3\mathbf{Q}_{\mathrm{M}} &= \frac{1}{2V_0}\int_{\Gamma_0} \vec{X}\vec{p}\vec{X}\mathrm{d}\Gamma_0 = \frac{1}{2V_0}\int_{\Gamma_0} ((\vec{N}\cdot\mathbf{P}^c_{\mathrm{m}})\vec{X}\vec{X})^{LC}\mathrm{d}\Gamma_0 \\ &= \frac{1}{2V_0}\int_{V_0} (\nabla_{0\mathrm{m}}\cdot(\mathbf{P}^c_{\mathrm{m}}\vec{X}\vec{X}))^{LC}\mathrm{d}V_0, \end{aligned} \tag{95}$$

where the superscript LC denotes left conjugation, $T^{LC}_{ijk} = T_{jik}$. Finally using the equality

$$\begin{aligned} \nabla_{0\mathrm{m}}\cdot(\mathbf{P}^c_{\mathrm{m}}\vec{X}\vec{X}) &= (\nabla_{0\mathrm{m}}\cdot\mathbf{P}^c_{\mathrm{m}})\vec{X}\vec{X} + \mathbf{P}_{\mathrm{m}}\cdot(\nabla_{0\mathrm{m}}\vec{X})\vec{X} + (\vec{X}\mathbf{P}_{\mathrm{m}}\cdot(\nabla_{0\mathrm{m}}\vec{X}))^{LC} \\ &= \mathbf{P}_{\mathrm{m}}\vec{X} + (\vec{X}\mathbf{P}_{\mathrm{m}})^{LC}, \end{aligned} \tag{96}$$

where equilibrium has been exploited, the relation between the macroscopic higher-order stress tensor and microstructural quantities is obtained

$${}^3\mathbf{Q}_{\mathrm{M}} = \frac{1}{2V_0}\int_{V_0} \left(\mathbf{P}^T_{\mathrm{m}}\vec{X} + \vec{X}\mathbf{P}_{\mathrm{m}}\right) dV_0 \tag{97}$$

Note that the macroscopic higher-order stress tensor ${}^3\mathbf{Q}_{\mathrm{M}}$ does not equal the volume average of its microscopic counterpart $\vec{\nabla}_{0,\mathrm{m}}\mathbf{P}_{\mathrm{m}}$. Like for ${}^3\mathbf{G}_{\mathrm{M}}$ this is due to the fact that the micro scale problem is formulated as a classical boundary value problem. It is clear from (97) that ${}^3\mathbf{Q}_{\mathrm{M}}$ can be interpreted as the first moment (with respect to the RVE center) of the microscopic first Piola-Kirchhoff stress tensor $\mathbf{P}_{\mathrm{m}}$ over the initial RVE volume V_0.

8.4 Two-scale computational solution strategy

The boundary constraints (86) can be explicitly written in terms of the displacement vectors of the boundary points in the form

$$\int_{\Gamma_{0\mathrm{L}}} \vec{u}_{\mathrm{L}}\ \mathrm{d}\Gamma_0 = \vec{u}_{L^*}(\mathbf{F}_{\mathrm{M}}, {}^3\mathbf{G}_{\mathrm{M}}), \qquad \int_{\Gamma_{0\mathrm{B}}} \vec{u}_{\mathrm{B}}\ \mathrm{d}\Gamma_0 = \vec{u}_{B^*}(\mathbf{F}_{\mathrm{M}}, {}^3\mathbf{G}_{\mathrm{M}}) \tag{98}$$

where $\vec{u}_{L^*}$ and $\vec{u}_{B^*}$ solely depend on the given $\mathbf{F}_\mathrm{M}$ and ${}^3\mathbf{G}_\mathrm{M}$ and RVE geometry as apparent from their definitions

$$\begin{aligned}\vec{u}_{L^*} =&(\mathbf{F}_\mathrm{M}-\mathbf{I})\cdot\int_{\Gamma_{0\mathrm{L}}}(\vec{X}_\mathrm{L}-\vec{X}_1)\,\mathrm{d}\Gamma_0+\tfrac{1}{2}{}^3\mathbf{G}_\mathrm{M}^{LC}:\int_{\Gamma_{0\mathrm{L}}}(\vec{X}_\mathrm{L}\vec{X}_\mathrm{L}-\vec{X}_1\vec{X}_1)\,\mathrm{d}\Gamma_0\\ \vec{u}_{B^*} =&(\mathbf{F}_\mathrm{M}-\mathbf{I})\cdot\int_{\Gamma_{0\mathrm{B}}}(\vec{X}_\mathrm{B}-\vec{X}_1)\,\mathrm{d}\Gamma_0+\tfrac{1}{2}{}^3\mathbf{G}_\mathrm{M}^{LC}:\int_{\Gamma_{0\mathrm{B}}}(\vec{X}_\mathrm{B}\vec{X}_\mathrm{B}-\vec{X}_1\vec{X}_1)\,\mathrm{d}\Gamma_0\end{aligned} \tag{99}$$

Once the BVP associated to the microstructural RVE problem is defined (boundary conditions, constitutive equations) the micro-problem can be solved with a standard finite element method. On the basis of the resulting boundary tractions, the RVE averaged stress tensors are extracted (see equations (94), (97))) and transported to the corresponding macroscopic material point.

For the finite element solution of the macroscopic problem a stiffness matrix at every macroscopic integration point is required. As emphasized earlier, in computational homogenization schemes there is no explicit form of the macroscopic constitutive behaviour assumed a priori. Like for the first-order case, the tangent operator is determined numerically by condensation of the microscopic stiffness matrix. For this, first the elaborated constraint relations between boundary nodes (equations (80), (86)) are applied to the total assembled stiffness matrix of the RVE following a similar procedure as presented for the first-order case. Details for the second-order case are given in Kouznetsova (2002); Kouznetsova et al. (2004b,a). This results in the elimination of the dependent degrees of freedom from the system of equations. The next step is to partition the remaining system of equations as

$$\begin{bmatrix}\underline{K}_{pp} & \underline{K}_{pf}\\ \underline{K}_{fp} & \underline{K}_{ff}\end{bmatrix}\begin{bmatrix}\delta\underset{\sim}{u}_p\\ \delta\underset{\sim}{u}_f\end{bmatrix}=\begin{bmatrix}\delta\underset{\sim}{f}_p\\ \underset{\sim}{0}\end{bmatrix} \tag{100}$$

where the subscript p refers to "prescribed" degrees of freedom (degrees of freedom through which the macroscopic tensors $\mathbf{F}_\mathrm{M}$ and ${}^3\mathbf{G}_\mathrm{M}$ are imposed on the RVE). In the present framework these are the degrees of freedom corresponding to the four corner nodes of the RVE ($\vec{u}_i, i=\overline{1,4}$) and to the degrees of freedom entering the RVE system of equations through the boundary constraints (98). The subscript f in (100) refers to all remaining "free" nodes. Elimination of $\delta\underline{u}_f$ from the system (100) then leads to the reduced stiffness matrix $\underline{K}_\mathrm{M}$ that relates the variations of the prescribed degrees of freedom to the variations of the associated forces

$$\underline{K}_\mathrm{M}\delta\underset{\sim}{u}_p=\delta\underset{\sim}{f}_p,\qquad\text{with}\quad\underline{K}_\mathrm{M}=\underline{K}_{pp}-\underline{K}_{pf}(\underline{K}_{ff})^{-1}\underline{K}_{fp} \tag{101}$$

The linearized constitutive relations for the second gradient continuum can be written as

$$\delta \mathbf{P}_{\mathrm{M}} = {}^{4}\mathbf{C}_{\mathrm{M}}^{(1)} : \delta \mathbf{F}_{\mathrm{M}}^{c} + {}^{5}\mathbf{C}_{\mathrm{M}}^{(2)} \vdots \delta^{3}\mathbf{G}_{\mathrm{M}}^{RC} \tag{102}$$

$$\delta^{3}\mathbf{Q}_{\mathrm{M}} = {}^{5}\mathbf{C}_{\mathrm{M}}^{(3)} : \delta \mathbf{F}_{\mathrm{M}}^{c} + {}^{6}\mathbf{C}_{\mathrm{M}}^{(4)} \vdots \delta^{3}\mathbf{G}_{\mathrm{M}}^{RC} \tag{103}$$

where the fourth-order tensor ${}^{4}\mathbf{C}_{\mathrm{M}}^{(1)}$, the fifth-order tensors ${}^{5}\mathbf{C}_{\mathrm{M}}^{(2)}$ and ${}^{5}\mathbf{C}_{\mathrm{M}}^{(3)}$ and the sixth-order tensor ${}^{6}\mathbf{C}_{\mathrm{M}}^{(4)}$ are the macroscopic consistent tangents. Using the RVE reduced stiffness matrix $\underline{\underline{K}}_{\mathrm{M}}$ rewritten in a tensor format such that

$$\sum_{j} \mathbf{K}_{\mathrm{M}}^{(ij)} \cdot \delta \vec{u}_{(j)} = \delta \vec{f}_{(i)}, \qquad i,j = 1,2,3,4,L^{*},B^{*} \tag{104}$$

permits to extract the macroscopic consistent tangents in the following format (see Kouznetsova et al. (2004a) for the derivation)

$$\begin{aligned}
{}^{4}\mathbf{C}_{\mathrm{M}}^{(1)} &= \frac{1}{V_0} \sum_{i} \sum_{j} (\vec{X}_{(i)}^{\star} \mathbf{K}_{\mathrm{M}}^{(ij)} \vec{X}_{(j)}^{\star})^{LC} \\
{}^{5}\mathbf{C}_{\mathrm{M}}^{(2)} &= \frac{1}{2V_0} \sum_{i} \sum_{j} (\vec{X}_{(i)}^{\star} \mathbf{K}_{\mathrm{M}}^{(ij)} \mathbf{Y}_{(j)}^{\star})^{LC} \\
{}^{5}\mathbf{C}_{\mathrm{M}}^{(3)} &= \frac{1}{2V_0} \sum_{i} \sum_{j} (\mathbf{Y}_{(i)}^{\star} \mathbf{K}_{\mathrm{M}}^{(ij)} \vec{X}_{(j)}^{\star})^{C_{23}} \\
{}^{6}\mathbf{C}_{\mathrm{M}}^{(4)} &= \frac{1}{4V_0} \sum_{i} \sum_{j} (\mathbf{Y}_{(i)}^{\star} \mathbf{K}_{\mathrm{M}}^{(ij)} \mathbf{Y}_{(j)}^{\star})^{C_{23}}
\end{aligned} \tag{105}$$

with the superscript C_{23} indicating conjugation on the second and third indices and

$$\vec{X}_{(i)}^{\star} = \begin{cases} \vec{X}_{(i)} - \vec{X}_{(1)}, & \text{for} \quad i = 1,2,3,4, \\ \int\limits_{\Gamma_{0\mathrm{L}}} (\vec{X}_L - \vec{X}_{(1)}) \, \mathrm{d}\Gamma_0, & \text{for} \quad i = L^{*}, \\ \int\limits_{\Gamma_{0\mathrm{B}}} (\vec{X}_B - \vec{X}_{(1)}) \, \mathrm{d}\Gamma_0, & \text{for} \quad i = B^{*} \end{cases} \tag{106}$$

$$\mathbf{Y}_{(i)}^{\star} = \begin{cases} \vec{X}_{(i)}\vec{X}_{(i)} - \vec{X}_{(1)}\vec{X}_{(1)}, & \text{for} \quad i = 1,2,3,4, \\ \int\limits_{\Gamma_{0\mathrm{L}}} (\vec{X}_L\vec{X}_L - \vec{X}_{(1)}\vec{X}_{(1)}) \, \mathrm{d}\Gamma_0, & \text{for} \quad i = L^{*}, \\ \int\limits_{\Gamma_{0\mathrm{B}}} (\vec{X}_B\vec{X}_B - \vec{X}_{(1)}\vec{X}_{(1)}) \, \mathrm{d}\Gamma_0, & \text{for} \quad i = B^{*} \end{cases} \tag{107}$$

In the second-order computational homogenization framework the macroscopic problem represents a full second gradient continuum (Mindlin, 1964; Toupin, 1964; Fleck and Hutchinson, 1997). For such a second gradient continuum the local equilibrium equation (in the absence of body forces and body moments) is written as

$$\nabla_{0\mathrm{M}} \cdot \left(\mathbf{P}^c_{\mathrm{M}} - (\nabla_{0\mathrm{M}} \cdot {}^3\mathbf{Q}_{\mathrm{M}})^c \right) = \vec{0} \tag{108}$$

The natural boundary conditions associated with this system of partial differential equations are expressed in (i) the surface traction $\vec{t}_{\mathrm{M}}$

$$\begin{aligned} \vec{t}_{\mathrm{M}} &= \vec{N}_{\mathrm{M}} \cdot \left(\mathbf{P}^c_{\mathrm{M}} - (\nabla_{0\mathrm{M}} \cdot {}^3\mathbf{Q}_{\mathrm{M}})^c \right) + (\nabla^s_{0\mathrm{M}} \cdot \vec{N}_{\mathrm{M}}) \vec{N}_{\mathrm{M}} \cdot (\vec{N}_{\mathrm{M}} \cdot {}^3\mathbf{Q}_{\mathrm{M}})^c \\ &- \nabla^s_{0\mathrm{M}} \cdot (\vec{N}_{\mathrm{M}} \cdot {}^3\mathbf{Q}_{\mathrm{M}})^c \end{aligned} \tag{109}$$

where the surface gradient operator is defined as $\nabla^s_{0\mathrm{M}} = (\mathbf{I} - \vec{N}_{\mathrm{M}} \vec{N}_{\mathrm{M}}) \cdot \nabla_{0\mathrm{M}}$ with $\vec{N}_{\mathrm{M}}$ the unit outward normal on the surface of the macroscopic body in the undeformed configuration and (ii) the double stress traction $\vec{r}_{\mathrm{M}}$

$$\vec{r}_{\mathrm{M}} = \vec{N}_{\mathrm{M}} \cdot {}^3\mathbf{Q}_{\mathrm{M}} \cdot \vec{N}_{\mathrm{M}} \tag{110}$$

In the case of a non-smooth surface of the body (with edges) also an additional line load appears. The kinematic boundary conditions for the second gradient continuum include prescribed displacements $\vec{u}_{\mathrm{M}}$ and normal gradients of displacements $D_{0\mathrm{M}} \vec{u}_{\mathrm{M}}$ with $D_{0\mathrm{M}} = \vec{N}_{\mathrm{M}} \cdot \nabla_{0\mathrm{M}}$.

The constitutive equations relating the macroscopic first Piola-Kirchhoff stress tensor $\mathbf{P}_{\mathrm{M}}$ and the higher-order stress tensor ${}^3\mathbf{Q}_{\mathrm{M}}$ to the history of the macroscopic deformation tensor $\mathbf{F}_{\mathrm{M}}$ and its gradient ${}^3\mathbf{G}_{\mathrm{M}}$ are thus obtained numerically, whereas their variations are obtained in the linearized form (102)–(103) with the macroscopic consistent tangents calculated from the condensed microscopic stiffness matrix according to (105).

8.5 Parallel solution of the multi-scale nested boundary value problems

In spite of the large computational effort required by a computational homogenization scheme, it is well possible to make an efficient analysis if optimal use is made of the inherent parallel nature of this multi-scale framework. Whenever microstructural constitutive information is needed in a macroscopic (integration) point, a separate subroutine can be started on the RVE-level that solves the requested boundary value problem. This can be done in parallel in as many integration points as available processors. Using PVM (Parallel Virtual Machine) or MPI (Message Passing Interface), it

is relatively easy to construct such a parallel implementation for this type of problems, as schematically depicted in figure 20. Evidently, this procedure drastically reduces the total calculation time.

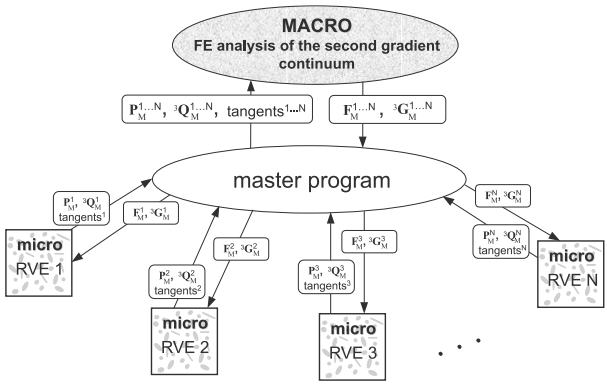


Figure 20. Schematic overview of the parallel solution of the multi-scale nested BVPs

9 Higher-order issues

9.1 First-order versus second-order

The first example concerns the comparison of the mechanical and kinematical response of a heterogeneous microstructure for the first-order and the second-order scale transition. To this purpose, an RVE is considered, which is depicted in its undeformed state in figure 21. The material consid-

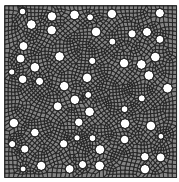


Figure 21. Undeformed two-dimensional RVE of a voided metal

ered is a metal with very weak inclusions, which have a negligible mechanical contribution (e.g. voids). The matrix material is elasto-viscoplastic, constitutively prescribed by a Bodner-Partom viscosity function. Following the

conventional multiplicative split of $\mathbf{F}$, the elastic response is modelled by a classical (isotropic) Neo-Hookean relationship, where the Kirchhoff stress tensor is given by

$$\tau \;=\; K(J-1)\mathbf{I} + G\bar{\mathbf{b}}_e^d \tag{111}$$

In here, J is the volume change ratio, K is the bulk modulus, G is the shear modulus, $\bar{\mathbf{b}}_e^d$ is the deviatoric part of the isochoric elastic left Cauchy-Green deformation tensor. The plastic part is determined through the plastic deformation rate tensor $\mathbf{D}_p$

$$\mathbf{D}_p \;=\; \frac{\sigma^d}{2\eta} \tag{112}$$

where the viscosity η is related to the von Mises equivalent stress σ_e and the effective plastic strain ϵ_p by

$$\begin{aligned} \eta \;&=\; \frac{\sigma_e}{\sqrt{12\Gamma_0}} \exp\left(\frac{1}{2}\left[\frac{Z}{\sigma_e}\right]^{2n}\right) \\ Z \;&=\; Z_1 + (Z_0 - Z_1)\mathrm{e}^{-m\epsilon_p} \end{aligned} \tag{113}$$

with Γ_0, n, Z_0, Z_1 and m material constants. In the present analysis, an aluminum matrix (AA 1050) has been considered for which the material parameters are given by $G = 2.6{\cdot}10^4$ MPa, $K = 7.8{\cdot}10^4$ MPa, $\Gamma_0 = 10^8\,\mathrm{s}^{-2}$, $m = 13.8$, $n = 3.4$, $Z_0 = 81.4$ MPa, $Z_1 = 170$ MPa.

The comparison between the first and second-order formulation is next made for a microstructure with a second phase (12% volume fraction of voids in this case) with an average size of about 6.6 μm. The macroscopic deformation history of a specific material point representing bending with superimposed tension is extracted. This history is imposed to the RVE, after which the micro scale BVP can be solved. The deformed microstructures shown in figure 22 are then obtained for the considered point (with the same macroscopic deformation history!). The deformation modes obtained and the small scale strain fields are obviously different, which reflects the kinematical enrichment of the second-order approximation. Note that the RVE is clearly bending in the second-order case, which is the result of the presence of the higher-order deformation modes that properly account for the size of the microstructure. The periodicity of the microfluctuation field can also be noticed. The macro field however, is no longer periodic for the second-order case.

9.2 Full gradient versus couple stress

Considerable attention has been devoted in the literature to the use of a couple stress theory (Toupin, 1962; Koiter, 1964), in which only the

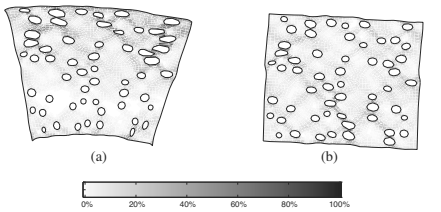


Figure 22. Deformed RVEs for the second-order (a) and first-order (b) RVE with their micro fields of the effective plastic strain

rotational gradient field is taken into account (the curl of $\mathbf{F}_{\mathrm{M}}$). In order to illustrate the contribution of the stretch gradients, the deformation history of a material point in the vicinity of the notch of a notched tensile specimen has been considered. This deformation history is used to construct the full gradient micro scale RVE and the couple stress micro scale RVE, the latter involving the antisymmetric part ${}^3\mathbf{G}_{\mathrm{M}}^A$ of ${}^3\mathbf{G}_{\mathrm{M}}$ only (i.e. switching to index notation, ${}^3\mathbf{G}_{\mathrm{M}}^A = \frac{2}{3}G_{Mijk} - \frac{1}{3}(G_{Mjki} + G_{Mkij})$, which is a third-order representation of the second-order curvature tensor that is normally used in couple stress theories). The analysis has been performed on the RVE shown in figure 21, with an average void size of $0.13\,\mu$m. The comparison between the full gradient RVE and the couple stress RVE is shown in figure 23. This

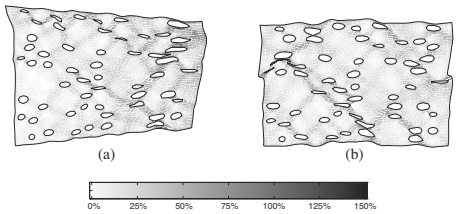


Figure 23. Full gradient, full ${}^3\mathbf{G}_{\mathrm{M}}$ (left) deformed RVE versus the couple stress, antisymmetric part of ${}^3\mathbf{G}_{\mathrm{M}}$ deformed RVE (right) in the same macroscopic point in the vicinity of the notch of a notched tensile specimen. Equivalent plastic strain fields are depicted inside the RVEs

example illustrates the difference between the full gradient and couple stress case and particularly emphasizes the relevance of the stretch gradients in the scale transition.

9.3 Geometrical size effects

The intrinsic role of the size of the microstructure becomes obvious through the complexity of the deformed RVE obtained and the constitutive response at the microscale which is triggered through the macroscopic deformation. Clearly, large microstructures will show a more pronounced gradient effect (e.g. the bending mode). Performing such a microstructural size analysis in a single macroscopic material point for a given constant loading history but with different underlying microstructures is straightforward. The extracted stress tensors are characteristic for the size effect that has been obtained. The scalar norm of the macroscopic first Piola-Kirchhoff stress tensor $\mathbf{P}_{\mathrm{M}}$ (i.e. defined as $(P_{Mij}P_{Mij})^{1/2}$) and the higher-order tensor ${}^{3}\mathbf{Q}_{\mathrm{M}}$ (i.e. by taking $(Q_{Mijk}Q_{Mijk})^{1/2}$) are good measures to illustrate this, see figure 24. Deviations from the first-order theory are increasingly

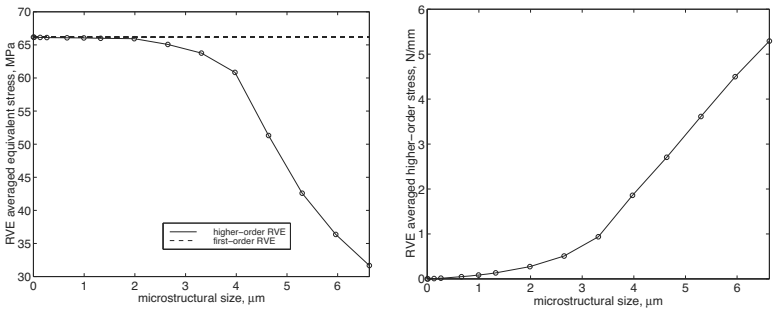


Figure 24. Scalar stress norms of the macroscopic first Piola-Kirchhoff stress tensor (top) and the higher-order stress tensor (bottom) as a function of the microstructural size in a given macroscopic material point

important for larger microstructures. In the limit of an infinitesimal RVE, the first-order solution is always recovered.

9.4 Large macroscopic gradients

For a given microstructure with fixed intrinsic sizes, the second-order framework turns out to be relevant again if local macroscopic deformations tend to be highly non-uniform, i.e. if the gradient ${}^{3}\mathbf{G}_{\mathrm{M}}$ becomes non-negligible with respect to the microstructural size. This is typically the case upon localization of the deformation at the macro scale, where deformations vary strongly in narrow zones. Localization leads to increasing values of ${}^{3}\mathbf{G}_{\mathrm{M}}$, which strongly interacts with the constitutive behaviour of the underlying microstructure. This is shown for the heterogeneous two-phase mate-

rial considered with an average size of the weak phase of 13 μm. The gradient ${}^3\mathbf{G}_{\mathrm{M}}$ is increased proportionally from $G_M^{eq} = ({}^3\mathbf{G}_{\mathrm{M}} \vdots {}^3\mathbf{G}_{\mathrm{M}})^{1/2} = 0\,\mathrm{mm}^{-1}$ to $0.19\,\mathrm{mm}^{-1}$, $0.39\,\mathrm{mm}^{-1}$, $0.78\,\mathrm{mm}^{-1}$ and $0.98\,\mathrm{mm}^{-1}$. The deformed RVEs with their effective plastic strain fields are shown in figure 25. The most left RVE has a zero macroscopic gradient, which reflects a first-order loading mode. For larger gradients, the deformed shape of the RVE becomes clearly more complex (and more representative compared to a first-order result for the real physical geometry of the microstructure).

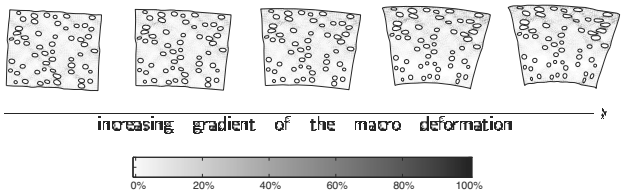


Figure 25. Higher-order RVE response as the result of an increasing macroscopic gradient in a material point with a given microstructure. Equivalent plastic strain fields are depicted inside the RVEs

9.5 Macroscopic localization

In order to scrutinize the added value of the second-order method on the macro scale, an academic benchmark problem was set up in which large macroscopic gradients appear and in which the material softens moderately on the global RVE-scale. The example consists in a periodic micro-voided plate, made of a commercial steel (T67CA), for which the matrix material can be modelled with a (hypo)elasto-perfectly plastic constitutive model ($E = 210$ GPa, $\nu = 0.3$, $\sigma_{y0} = 507$ MPa). The voids in the plate have a diameter of 4 microns, whereas the periodic cell itself measures $d = 10\mu$m, see figure 26. An imperfection (i.e. a reduction) of 20 % is applied to the yield stress in the left bottom cell, in order to trigger the appearance of macroscopic gradients, that may lead to localization of deformations. The second-order equilibrium scheme used requires higher-order boundary conditions that have to be prescribed at the edges of domain. In this case, the bottom and left edge in figure 26 are symmetry axes where normal displacements and tangential stress tractions are taken zero. The top edge undergoes a uniform vertical displacement, whereas the right edge is free. The normal derivatives of the tangential displacement components at the bottom,

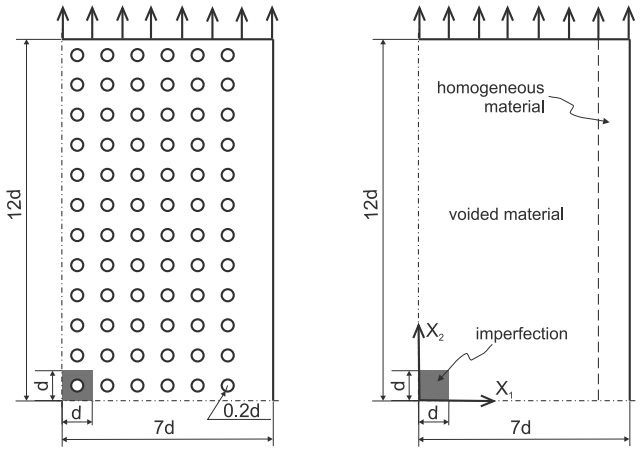
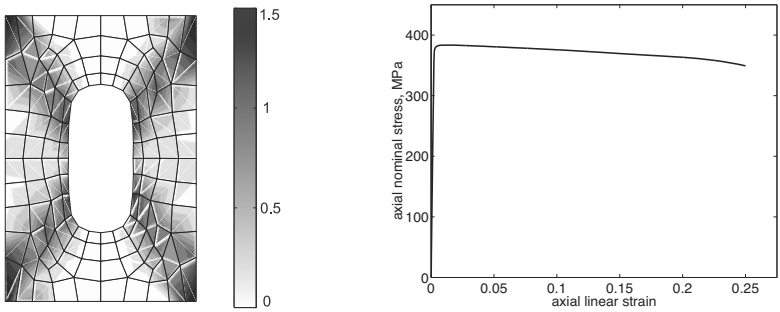


Figure 26. Benchmark problem for computational homogenization analysis

left and top edge are constrained as well. Double stress vectors (Fleck and Hutchinson, 1997) $\vec{N}\cdot{}^{3}\mathbf{Q}_{\mathrm{M}}\cdot\vec{N}$ are zero at all edges.

Strong shear bands occur inside the unit cells, which leads to moderate (geometrical) softening on the global RVE-level, see figure 27. The inade-



(a) (b)

Figure 27. Single unit cell under uniaxial tension. (a) Deformed geometry and distribution of the equivalent Green-Lagrange strain within the unit cell. (b) Stress-strain response.

quacy of the first-order scheme to deal with this type of behaviour becomes apparent in figure 28, where the prescribed macroscopic displacement measures $0.9\mu m$. In here, the solution for two different mesh sizes is depicted,

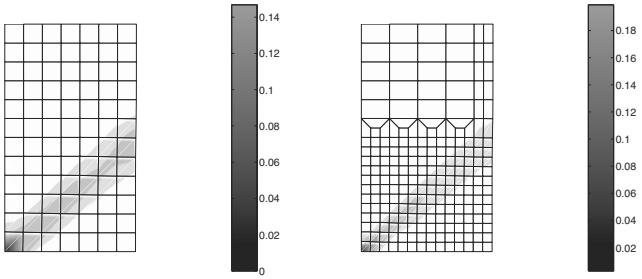


Figure 28. Localized (equivalent) Green-Lagrange strain field for two mesh sizes on the basis of the first-order scheme

where the solution fully localizes according to the size of the elements used. It is not surprising, that this homogenization method suffers from exactly the same shortcomings as classical (local) constitutive softening models as extensively investigated by many authors in the past decade (Schreyer and Chen, 1986; Bažant and Pijaudier-Cabot, 1988; Aifantis, 1992; de Borst and Mühlhaus, 1992; de Borst and Pamin, 1996; Peerlings et al., 1996; Svedberg and Runesson, 1997; Geers et al., 1998; Engelen et al., 2003). This property is inherently linked to the principle of local action, which associates for each macroscopic point a volume with infinitesimal size at the RVE level. Upon further refinement of the macroscopic mesh, the energy dissipated in the softening RVE on the micro-scale is in fact dissipated in a shrinking volume at the macroscopic scale, which is one of the main manifestations of the ill-posedness of the boundary value problem (at the macro scale) to be solved.

The second-order computational homogenization method leads to a higher-order boundary value problem, for which the regularizing effects are known to exist. The size of the microstructural volume element implicitly sets the length scale in the macro scale analysis, which makes the numerical solution independent of the mesh size (the localization band converges to a finite width), see figure 29.

A two-scale overview of the deformed state of the perforated plate is shown in figure 30. Note that the behaviour ensuing from a regularized continuum theory for failure at the micro scale or a well-posed discrete failure model at the micro scale cannot be upscaled with a first-order homogenization method. In fact, the well-posedness at the micro scale basically implies that the corresponding constitutive response at the macro scale does not depend on the discretization at the micro scale. This is of course a necessary condition, though not sufficient. The volume in which the energy was

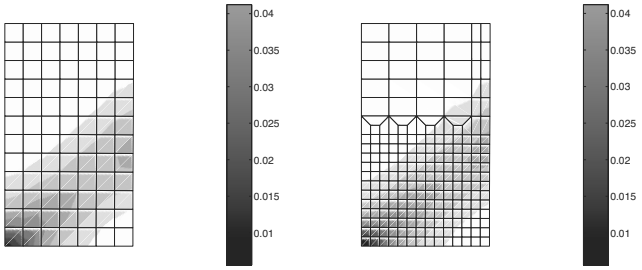


Figure 29. Localized (equivalent) Green-Lagrange strain field for two mesh sizes on the basis of the second-order scheme

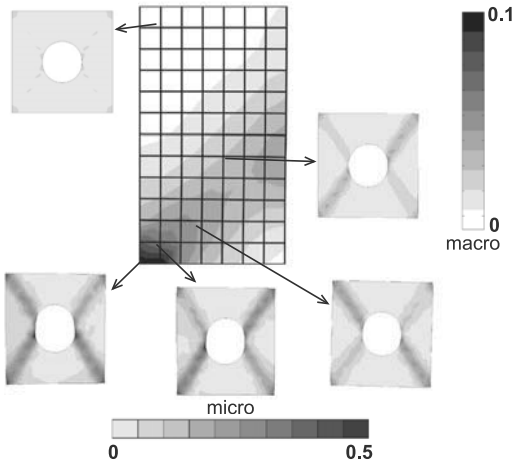


Figure 30. Deformed unit cell patterns and their equivalent Green-Lagrange strain fields embedded in the corresponding macroscopic solution field

dissipated at the micro scale has to be transported correctly to the macro scale to prevent loss of well-posedness at the macro scale. This is impossible within the standard local continuum mechanics framework (first-order scheme), which again underlines the need for higher-order homogenization methods for the upscaling of failure processes accompanied by macroscopic softening.

9.6 The higher-order RVE

It has already been emphasized that the requirement of statistical representativeness constituted an important aspect in the definition of an RVE for a classical first-order homogenization approach. As a result, there was no restriction on the (maximum) size of a representative cell (on the contrary, taking the first-order RVE as large as possible, allows to represent given statistical characteristics more accurately). This is related to the fact that the first-order computational homogenization scheme (as well as most other conventional homogenization methods) deals with an ordinary local continuum on the macroscopic level. Such a continuum does not possess a material length scale and accordingly the size of a microstructural cell does not play a role. For the second-order case, size does play a role and the definition of an RVE is therefore not trivial. A detailed analysis on this subject has been performed in Kouznetsova et al. (2004b). In here, it has been shown that the size of the microstructural RVE used in a second-order computational homogenization scheme is intrinsically related to the length scale of the resulting macroscopic homogenized higher-order continuum. Furthermore, material and geometrical non-linearities significantly contribute to the relation between the RVE size and the obtained macroscopic response. In a second-order computational homogenization two conflicting requirements on the microstructural representative volume element have to be accommodated. On the one hand, the accurate determination of the overall behaviour of a multi-phase material requires a large representative cell with many (interacting) heterogeneities. On the other hand, the size of a representative cell used in the second-order computational homogenization scheme implicitly sets the length scale of the macroscopic homogenized higher-order continuum.

In most cases, it is possible to give a reasonable estimate of the required size of the representative cell based on the qualitative assessment of the basic mechanisms of the underlying microstructural evolution, interaction and the statistics of the considered microstructure. The lower limit for the size of the RVE should be selected as the minimum size that contains enough microstructural features to allow the development of the governing microstructural physical mechanisms that are relevant for the problem under consideration. The upper limit for the RVE size directly results from the underlying assumption that the macroscopic deformations vary linearly over the microstructural cell. If this assumption does not apply (i.e. if the macroscopic deformations vary too strongly on the scale of the microstructural constituents) a computational homogenization scheme can never provide accurate results, since a separation of scales is not applicable. In such cases the analysis should be performed by detailed microstructural mod-

elling. This, in fact, sets an upper limit on the RVE size, which may be used for the second-order homogenization analysis.

10 Conclusions

This contribution presented an overview of two computational homogenization techniques for the multi-scale analysis of the mechanical behaviour of (physically and geometrically) complex microstructures. It has been shown that length scales can be bridged more accurately by transferring more information between the two scales considered. Rigourous scale transitions have been established by making use of averaging theorems and the Hill-Mandel energy condition. The first-order computational homogenization strategy fully complies with the principle of local action and the principle of separation of scales. The kinematics are essentially based on the linearization of the macroscopic nonlinear deformation map. The second-order case was based on the proper incorporation of the macroscopic gradient of the deformation tensor into the kinematical micro-macro framework. Work-conjugated stress and higher-order stress tensors are naturally retrieved and a full gradient continuum is obtained on the macro scale. The main advantage of the performed scale transition resides in the fact that the constitutive response (either first- or second-order) is obtained directly from the collective behaviour of all constituent phases at the micro scale. No assumptions need to be made on the format of the macroscopic constitutive relationship, which makes the proposed scale transition a versatile tool to assess macroscopic constitutive relations. Furthermore, the methodology can be fully implemented in a hierarchical solution scheme, where two nested boundary value problems have to be solved. Consistency of the tangent operator is preserved by the scale transition. The presented two-scale framework is parallel in its nature, which makes the implementation of the numerical solution method on a multi-processor cluster clearly beneficial.

Whether or not a second-order model should be used has to be decided considering the governing scales, loading and the presence of failure/softening. Whenever the principle of separation of scales continues to hold, macroscopic gradients remain small and failure does not occur, it is certainly recommended to use the first-order computational homogenization method. However, if damage and failure are of interest and need to be linked to microstructural events, it is obvious that a higher-order technique will be necessary. Even if a well-posed regularized solution strategy (either continuum or discrete) is used on the micro level, a second-order scheme will remain necessary. The length scale, which may be well defined on the microstructural scale is not preserved through a first-order scale transition.

Evidently, this becomes most apparent for failure analyses with softening.

Likewise, geometrical size effects, for which the size of the microstructure is not negligible with respect to the geometry of the macro-specimen, have to be captured by such a technique as well. Note that microstructural size effects, which emerge from size-dependent small scale deformation mechanisms (e.g. dislocation plasticity leading to the Hall-Petch effect, the Friedel effect, etc.) can still be upscaled appropriately with a first-order scheme. In here, the microstructural size effect becomes apparent in the constitutive response of the microstructural phases. Most important property of a second-order method is the fact that it implicitly incorporates a length scale, which depends on the size of the microstructure. This size becomes apparent on the macro-scale as the length scale that sets the width of localization zones, or that governs geometrical size effects. The role of the RVE size in a second-order scheme is crucial and has been discussed extensively in Kouznetsova et al. (2004b).

Computational homogenization seems to make constitutive modelling considerably easier. The first-order or second-order constitutive response (which is difficult to capture in a closed format with its constitutive tensors), is retrieved directly from the microstructure. This is particularly useful to assess the homogenized 'continuum' response of microstructural discrete systems, in which e.g. atomistics, molecular dynamics or discrete dislocation dynamics are used to obtain the fine scale response. Undoubtedly, many issues are still to be explored: assessing higher-order boundary conditions from the microstructure, upscaling various failure mechanisms within the microstructure that lead to macroscopic degradation, the appearance of geometrical size effects in miniaturization, etc.

Bibliography

E. C. Aifantis. On the role of gradients in the localization of deformation and fracture. *International Journal of Engineering Science*, 30:1279–1299, 1992.

F. P. T. Baaijens. Calculation of residual stresses in injection-molded products. *Rheol. Acta*, 30:284–299, 1991.

M. W. Barsoum, P. Kangutkar, and A. S. D. Wang. Matrix crack initiation in ceramic matrix composites. Part I: Experiments and test results. *Composites Science and Technology*, 44:257–269, 1992.

Z. P. Bažant and G. Pijaudier-Cabot. Nonlocal continuum damage, localization instability and convergence. *Journal of Applied Mechanics*, 55: 287–293, 1988.

J. R. Brockenbrough, S. Suresh, and H. A. Wienecke. Deformation of metal-

matrix composites with continuous fibers: geometrical effect of fiber distribution and shape. *Acta Metall. Mater.*, 39(5):735–752, 1991.

P. W. Chung, K. K. Tamma, and R. R. Namburu. Asymptotic expansion homogenization for heterogeneous media: computational issues and applications. *Composites Part A: Applied Science and Manufacturing*, 32 (9):1291–1301, 2001.

R. D. Cook, D. S. Malkus, and M. E. Plesha. *Concepts and applications of finite element analysis*. Wiley, Chichester, 1989.

R. de Borst and H. B. Mühlhaus. Gradient-dependent plasticity: Formulation and algorithmic aspects. *International Journal for Numerical Methods in Engineering*, 35:521–539, 1992.

R. de Borst and J. Pamin. Some novel developments in finite elemet procedures for gradient-dependent plasticity and finite elements. *International Journal for Numerical Methods in Engineering*, 39:2477–2505, 1996.

W. J. Drugan and J. R. Willis. A micromechanics-based nonlocal constitutive equation and estimates of representative volume element size for elastic composites. *Journal of the Mechanics and Physics of Solids*, 44 (4):497–524, 1996.

R. A. B. Engelen, M. G. D. Geers, and F. P. T. Baaijens. Nonlocal implicit gradient-enhanced softening plasticity. *International Journal of Plasticity*, 2003. In press.

F. Feyel and J. L. Chaboche. FE^2 multiscale approach for modelling the elasto-viscoplastic behaviour of long fibre SiC/Ti composite materials. *Computer Methods in Applied Mechanics and Engineering.*, 183:309–330, 2000.

J. Fish and W. Chen. Higher-order homogenization of initial/boundary-value problem. *Journal of Engineering Mechanics*, 127(12):1223–1230, 2001.

J. Fish, Kamlun Shek, Muralidharan Pandheeradi, and Mark S. Shephard. Computational plasticity for composite structures based on mathematical homogenization: Theory and practice. *Computer Methods in Applied Mechanics and Engineering*, 148(1-2):53–73, 1997.

N. A. Fleck and J. W. Hutchinson. Strain gradient plasticity. *Advances in Applied Mechanics*, 33:295–361, 1997.

S. Forest, F. Pradel, and K. Sab. Asymptotic analysis of heterogeneous Cosserat media. *International Journal of Solids and Structures*, 38:4585–4608, 2001.

K. Garikipati and T. J. R. Hughes. A variational multiscale approach to strain localization - formulation for multidimensional problems. *Computer Methods in Applied Mechanics and Engineering*, 188(1-3):39–60, 2000.

M. G. D. Geers, R. de Borst, W. A. M. Brekelmans, and R. H. J. Peerlings. Strain-based transient-gradient damage model for failure analyses. *Computer Methods in Applied Mechanics and Engineering*, 160(1-2):133–154, 1998.

M. G. D. Geers, V. G. Kouznetsova, and W. A. M. Brekelmans. Gradient-enhanced computational homogenization for the micro-macro scale transition. *Journal de Physique IV*, 11(5):5145–5152, 2001.

M.G.D. Geers, V.G. Kouznetsova, and W.A.M. Brekelmans. Multi-scale second-order computational homogenization of microstructures towards continua. *International Journal for Multiscale Computational Engineering*, 1(4):371–386, 2003.

S. Ghosh, K. Lee, and S. Moorthy. Multiple scale analysis of heterogeneous elastic structures using homogenisation theory and Voronoi cell finite element method. *International Journal of Solids and Structures*, 32(1): 27–62, 1995.

S. Ghosh, K. Lee, and S. Moorthy. Two scale analysis of heterogeneous elastic-plastic materials with asymptotic homogenization and Voronoi cell finite element model. *Computer Methods in Applied Mechanics and Engineering*, 132:63–116, 1996.

S. Ghosh, K. Lee, and P. Raghavan. A multi-level computational model for multi-scale damage analysis in composite and porous materials. *International Journal of Solids and Structures*, 38(14):2335–2385, 2001.

L. E. Govaert, P. H. M. Timmermans, and W. A. M. Brekelmans. The influence of intrinsic strain softening on strain localization in polycarbonate: modeling and experimental validation. *J. Engrg. Mat. Technol.*, 122:177–185, 2000.

J. M. Guedes and N. Kikuchi. Preprocessing and postprocessing for materials based on the homogenization method with adaptive finite element methods. *Computer Methods in Applied Mechanics and Engineering*, 83: 143–198, 1990.

R. A. Hall. Computer modelling of rubber-toughened plastics: random placement of monosized core-shell particles in a polymer matrix and interparticle distance calculations. *J. Mater. Sci.*, 26:5631–5636, 1991.

R. Hill. Elastic properties of reinforced solids: some theoretical principles. *JJournal of the Mechanics and Physics of Solids*, 11:357–372, 1963.

R. Hill. On macroscopic effects of heterogeneity in elastoplastic media at finite strain. *Math. Proc. Camb. Phil. Soc.*, 95:481–494, 1984.

C. Huet. Application of variational concepts to size effects in elastic heterogeneous bodies. *J. Mech. Phys. Solids*, 38(6):813–841, 1990.

C. Huet. Coupled size and boundary-condition effects in viscoelastic heterogeneous and composite bodies. *Mechanics of Materials*, 31:787–829, 1999.

T. J. R. Hughes, G. R. Feijóo, L. Mazzei, and J. Quincy. The variational multiscale method - a paradigm for computational mechanics. *Computer Methods in Applied Mechanics and Engineering*, 166:3–24, 1998.

W. T. Koiter. Couple-stresses in the theory of elasticity. In *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen. Series B*, volume 67, pages 17–44, 1964.

V. G. Kouznetsova. *Computational homogenization for the multi-scale analysis of multi-phase materials.* PhD thesis, Eindhoven University of Technology, Mechanical Engineering Department, December 2002.

V. G. Kouznetsova, W. A. M. Brekelmans, and F. P. T. Baaijens. An approach to micro-macro modeling of heterogeneous materials. *Computational Mechanics*, 27:37–48, 2001.

V. G. Kouznetsova, M. G. D. Geers, and W. A. M. Brekelmans. Advanced constitutive modeling of heterogeneous materials with a gradient-enhanced computational homogenization scheme. *International Journal for Numerical Methods in Engineering*, 54:1235–1260, 2002.

V.G. Kouznetsova, M.G.D. Geers, and W.A.M. Brekelmans. Multi-scale second-order computational homogenization of multi-phase materials: a nested finite element solution strategy. *Computer Methods in Applied Mechanics and Engineering*, 193:5525–5550, 2004a.

V.G. Kouznetsova, M.G.D. Geers, and W.A.M. Brekelmans. Size of a representative volume element in a second-order computational homogenization framework. *International Journal for Multiscale Computational Engineering*, 2(4):575–598, 2004b.

I. Jasiuk M. Ostoja-Starzewski, S. D. Boccara. Couple-stress moduli and characteristic length of a two-phase composite. *Mechanics Research Communications*, 26(4):387–396, 1999.

R. A. Mackay. Effect of fiber spacing on interfacial damage in a metal matrix composite. *Scripta Metall. Mater.*, 24:167–172, 1990.

J. C. Michel, H. Moulinec, and P. Suquet. Effective properties of composite materials with periodic microstructure: a computational approach. *Computer Methods in Applied Mechanics and Engineering*, 172:109–143, 1999.

C. Miehe. Numerical computation of algorithmic (consistent) tangent moduli in large-strain computational inelasticity. *Comput. Methods Appl. Mech. Engrg.*, 134:223–240, 1996.

C. Miehe. Strain-driven homogenization of inelastic microstructures and composites based on an incremental variational formulation. *International Journal for Numerical Methods in Engineering*, 55:1285–1322, 2002.

C. Miehe. Computational micro-to-macro transitions for discretized microstructures of heterogeneous materials at finite strains based on the minimization of averaged incremental energy. *Computer Methods in Applied Mechanics and Engineering*, 192:559–591, 2003.

C. Miehe and A. Koch. Computational micro-to-macro transition of discretized microstructures undergoing small strain. *Archives in Applied Mechanics*, 72:300–317, 2002.

C. Miehe, J. Schotte, and M. Lambrecht. Homogenization of inelastic solid materials at finite strains based on incremental minimization principles. application to the texture analysis of polycrystals. *Journal of the Mechanics and Physics of Solids*, 50(10):2123–2167, 2002.

C. Miehe, J. Schotte, and J. Schröder. Computational micro-macro transitions and overall moduli in the analysis of polycrystals at large strains. *Computational Materials Science*, 16(1-4):372–382, 1999a.

C. Miehe, J. Schröder, and J. Schotte. Computational homogenization analysis in finite plasticity. Simulation of texture development in polycrystalline materials. *Computer Methods in Applied Mechanics and Engineering*, 171:387–418, 1999b.

R. D. Mindlin. Microstructure in linear elasticity. *Arch. Ration. Mech. Anal.*, 16:51–78, 1964.

N. Moës and T. Belytschko. Extended finite element method for cohesive crack growth. *Engineering Fracture Mechanics*, 69(7):813–833, 2002.

H. Moulinec and P. Suquet. A numerical method for computing the overall response of non-linear composites with complex microstructure. *Computer Methods in Applied Mechanics and Engineering*, 157:69–94, 1998.

T. Nakamura and S. Suresh. Effect of thermal residual stress and fiber packing on deformation of metal-matrix composites. *Acta Metall. Mater.*, 41(6):1665–1681, 1993.

S. Nemat-Nasser. Averaging theorems in finite deformation plasticity. *Mechanics of Materials*, 31:493–523, 1999.

S. Nemat-Nasser and M. Hori. *Micromechanics: overall properties of heterogeneous materials*. Elsevier, Amsterdam, 1993.

O. Nguyen and M. Ortiz. Coarse-graining and renormalization of atomistic binding relations and universal macroscopic cohesive behavior. *Journal of the Mechanics and Physics of Solids*, 2002. In Press, Available online.

M. Ostoja-Starzewski. Random field models of heterogeneous materials. *Int. J. Solids Structures*, 35(19):2429–2455, 1998.

M. Ostoja-Starzewski. Scale effects in materials with random distributions of needles and cracks. *Mechanics of Materials*, 31:883–893, 1999.

S. Pecullan, L. V. Gibiansky, and S. Torquato. Scale effects on the elastic behavior of periodic and hierarchical two-dimentional composites. *J. Mech. Phys. Solids*, 47:1509–1542, 1999.

R. H. J. Peerlings, R. de Borst, W. A. M. Brekelmans, and J. H. P. de Vree. Gradient-enhanced damage for quasi-brittle materials. *International Journal for Numerical Methods in Engineering*, 39:3391–3403, 1996.

R. H. J. Peerlings and N. A. Fleck. Numerical analysis of strain gradient effects in periodic media. *Journal de Physique IV*, 11:153–160, 2001.

P. Ponte Castañeda. New variational principles in plasticity and their application to composite materials. *Journal of the Mechanics and Physics of Solids*, 40:1757–1788, 1992.

P. Ponte Castañeda. Second-order homogenization estimates for nonlinear composites incorporating field fluctuations: I theory. *Journal of the Mechanics and Physics of Solids*, 50(4):737–757, 2002.

K. Ridderbos. The coarse-graining approach to statistical mechanics: how blissful is our ignorance? *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, 33(1): 65–67, 2002.

H. L. Schreyer and Z. Chen. One-dimensional softening with localization. *Journal of Applied Mechanics*, 53:791–979, 1986.

Z. Shan and A. M. Gokhale. Representative volume element for non-uniform micro-structure. *Comp. Mat. Sci.*, 24:361–379, 2002.

R. J. M. Smit. *Toughness of heterogeneous polymeric systems*. PhD thesis, Eindhoven University of Technology, Eindhoven, The Netherlands, 1998.

R. J. M. Smit, W. A. M. Brekelmans, and H. E. H. Meijer. Prediction of the mechanical behaviour of non-linear systems by multi-level finite element modeling. *Computer Methods in Applied Mechanics and Engineering*, 155:181–192, 1998.

R. J. M. Smit, W. A. M. Brekelmans, and H. E. H. Meijer. Prediction of the large-strain mechanical response of heterogeneous polymer systems: local and global deformation behaviour of a representative volume element of voided polycarbonate. *J. Mech. Phys. Solids*, 47:201–221, 1999.

V. P. Smyshlyaev and K. D. Cherednichenko. On rigorous derivation of strain gradient effects in the overall behaviour of periodic heterogeneous media. *Journal of the Mechanics and Physics of Solids*, 48:1325–1357, 2000.

V. P. Smyshlyaev and N. A. Fleck. Bounds and estimates for linear composites with strain gradient effects. *Journal of the Mechanics and Physics of Solids*, 42(12):1851–1882, 1994.

N. Sukumar, N. Moës, Moran, and T. Belytschko. Extended finite element method for three-dimensional crack modelling. *International Journal for Numerical Methods in Engineering*, 48(11):1549–1570, 2000.

P. M. Suquet. *Plasticity today: modelling, methods and applications*, chapter Local and global aspects in the mathematical theory of plasticity, pages 279–310. Elsevier Applied Science Publishers, London, 1985.

P. M. Suquet. Overall potentials and extremal surfaces of power law or ideally plastic composites. *Journal of the Mechanics and Physics of Solids*, 41:981–1002, 1993.

T. Svedberg and K. Runesson. A thermodynamically consistent theory of gradient-regularized plasticity coupled to damage. *International Journal of Plasticity*, 13(6-7):669–696, 1997.

K. Terada, M. Hori, T. Kyoya, and N. Kikuchi. Simulation of the multi-scale convergence in computational homogenization approaches. *International Journal of Solids and Structures*, 37(16):2285–2311, 2000.

K. Terada and N. Kikuchi. Nonlinear homogenization method for practical applications. In S. Ghosh and M. Ostaja-Starzewski, editors, *Computational Methods in Micromechanics*, volume AMD-Vol.212, MD-Vol.62, pages 1–16. ASME, 1995.

K. Terada and N. Kikuchi. A class of general algorithms for multi-scale analyses of heterogeneous media. *Computer Methods in Applied Mechanics and Engineering*, 190(40-41):5247–5464, 2001.

T. Tervoort. *Constitutive modelling of polymer glasses: finite, nonlinear viscoelastic behaviour of polycarbonate*. PhD thesis, Eindhoven University of Technology, Eindhoven, The Netherlands, 1996.

R. A. Toupin. Elastic materials with couple-stress. *Archive for Rational Mechanics and Analysis*, 11:385–414, 1962.

R. A. Toupin. Theories of elasticity with couple-stress. *Arch. Ration. Mech. Anal.*, 17:85–112, 1964.

N. Triantafyllidis and S. Bardenhagen. The influence of scale size on the stability of periodic solids and the role of associated higher order gradient continuum models. *Journal of the Mechanics and Physics of Solids*, 44 (11):1891–1928, 1996.

H. C. E. van der Aa, M. A. H. van der Aa, P. J. G. Schreurs, F. P. T. Baaijens, and W. J. van Veenen. An experimental and numerical study of the wall ironing process of polymer coated sheet metal. *Mechanics of Materials*, 32:423–443, 2000.

O. van der Sluis, P. J. G. Schreurs, W. A. M. Brekelmans, and H. E. H. Meijer. Overall behaviour of heterogeneous elastoviscoplastic materials: effect of microstructural modelling. *Mechanics of Materials*, 32:449–462, 2000.

O. van der Sluis, P. H. J. Vosbeek, P. J. G. Schreurs, and H. E. H. Meijer. Homogenization of heterogeneous polymers. *International Journal of Solids and Structures*, 36:3193–3214, 1999.