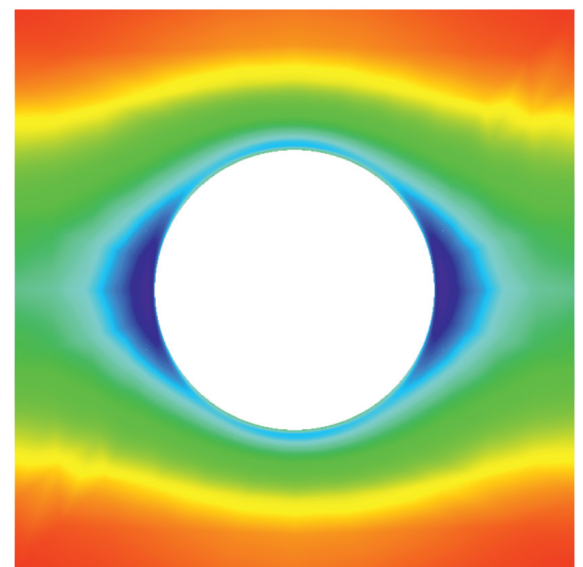
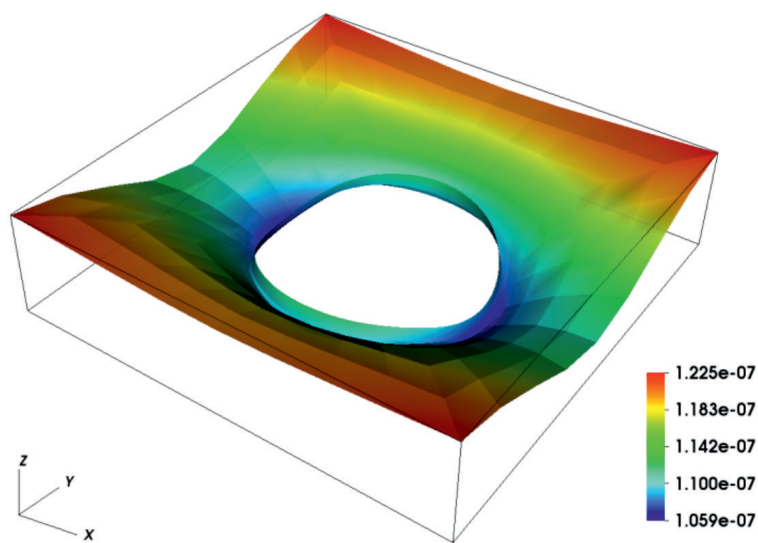


Continuum micromechanics of martensitic laminate microstructure with interface energies

Mehdi Goodarzi



**Continuum Micromechanics of
Martensitic Laminate Microstructure
with Interface Energies**

D I S S E R T A T I O N

zur

Erlangung des Grades

Doktor-Ingenieur

der

Fakultät für Maschinenbau
der Ruhr-Universität Bochum

von

Mehdi Goodarzi

aus Borojerd Iran

Bochum 2016

Dissertation eingereicht am: 18.10.2016

Tag der mündlichen Prüfung: 22.12.2016

Erster Referent: Prof. Dr. rer. nat. Klaus Hackl

Zweiter Referent: Prof. Dr. rer. nat. Wolfgang H. Müller

Vorsitzender: Prof. Dr.-Ing. Andreas Kilzer

Copyright © 2016 Mehdi Goodarzi
Lehrstuhl für Mechanik - Materialtheorie
Ruhr-Universität Bochum

Published electronically by Universitätsbibliothek Bochum 2017

Abstract

Due to their non-convex free energy and their distinctive lattice symmetries, martensitic materials typically form a laminate microstructure over a broad spectrum of length scales, to which most of their thermomechanical response and characteristics are owed. Surfaces and interfaces that are formed within this laminate microstructure contribute to almost every aspect of its behavior. These effects have been investigated either from a general perspective, such as regularity and scaling laws, or on particular aspects, such as optimal morphologies of accommodation patterns and boundaries. The present work develops an all-around model capable of reflecting micro-, meso-, and macroscopic features of the martensitic laminate microstructure employing energy methods with particular emphasis on surface and interface effects.

Our continuum micromechanical model has three major components. First, the elastic energy is formulated based on an ansatz of a first-order laminate. This ansatz enforces the coherence condition between martensite variants (almost) strictly. The stiffness tensors of the variants are oriented exactly as their lattices. Second, twin interface energies are estimated analytically using crystallographic and atomistic considerations of a sharp and coherent interface, viz an interfacial monolayer. Third, an ansatz for grain boundary energy is derived based on the microstrain energy of a self-similar construct within a thin boundary layer. The energy minimization framework, implemented as a nonlinear finite element scheme, is a competition between these three contributions. The model successfully predicts the thermomechanical response at the mesoscale, i.e. a single grain, and also the geometrical characteristics of refinement and accommodation of the microstructure.

In addition to the numerical treatment, a brief analytical study of a simplified formulation is presented, whose results comply with some of the well-known theoretical and experimental observations on scaling laws and the closely related energy estimates.

Kurzfassung

Aufgrund ihrer nicht-konvexen freien Energie und ihrer charakteristischen Gittersymmetrien bilden martensitische Materialien typischerweise eine Laminat-Mikrostruktur über ein breites Spektrum von Längenskalen, aus der die meisten ihrer thermomechanischen Merkmale abgeleitet werden können. Oberflächen und Grenzflächen, die innerhalb der martensitischen Laminat-Mikrostruktur gebildet werden, beeinflussen fast jeden Aspekt ihres Verhaltens. Diese Effekte wurden entweder aus einer allgemeinen Perspektive, wie Regularität und Skalierungsgesetze, oder hinsichtlich bestimmter Aspekte, wie zum Beispiel optimale Morphologien der Anpassungsmuster und Grenzflächen, untersucht. Die vorliegende Arbeit entwickelt ein Allround-Modell bezüglich der Grenzflächen, welches die mikro-, meso- und makroskopischen Eigenschaften der martensitischen Laminat-Mikrostruktur unter Verwendung von Energiemethoden simuliert.

Das mikromechanische Kontinuumsmodell hat drei Hauptkomponenten. Eine Komponente formuliert die elastische Energie auf Basis eines Laminat-Ansatzes erster Ordnung, der die Kohärenzbedingung zwischen Martensitvarianten nahezu exakt berücksichtigt. Dabei werden die Steifigkeitstensoren von Martensitvarianten genau nach ihren Kristallgittern orientiert. Der zweite Bestandteil schätzt die Zwillingsgrenzflächenenergien unter kristallographischer und atomistischer Berücksichtigung einer scharfen und kohärenten Grenzfläche, nämlich einer Monoschicht, analytisch ab. Ein Ansatz für die Korngrenzenergie wird durch die letzte Komponente erstellt. Dieser Ansatz basiert auf der Mikrodehnungsenergie eines selbstähnlichen Konstrukts in einer dünnen Grenzschicht.

Im Rahmen der Energieminimierung, die als nichtlineares Finite-Elemente-Schema implementiert wird, wird das thermomechanische Gleichgewicht durch eine Konkurrenz zwischen diesen drei Beiträgen erreicht. Das Modell sagt die thermomechanische Antwort auf der Mesoskala, sprich ein einzelnes Korn, sowie die geometrischen Eigenschaften der Verfeinerung und Anpassung der Mikrostruktur erfolgreich voraus.

Zusätzlich zu der numerischen Behandlung wird eine kurze analytische Untersuchung einer vereinfachten Formulierung dargestellt. Hierbei entsprechen die Ergebnisse dieser Untersuchung einigen hinlänglich bekannten theoretischen und experimentellen Beobachtungen über die Skalierungsgesetze und über die damit verbundenen Energieabschätzungen.

Acknowledgement

Without the kind presence, the support, and the generosity of many people it would not have been possible for me to start, continue, and finish this dissertation. I would like to take this opportunity to express my gratitude to them.

I am thankful to professor Klaus Hackl for taking me under his supervision, for his support, and for all that he has taught me over these years.

I thank professor Wolfgang H. Müller of TU Berlin for refereeing my dissertation and for his support and constructive comments. My thanks also go to professor Andreas Kilzer, the chairman of my doctoral examination board.

Next, I would like to thank my colleagues at the Institute of Mechanics for the joy of companionship. Some of them have become very good friends.

I am most grateful to my family and fiends for always being there for me.

Over the past ten years I have had the pleasure of knowing professor Franz Peters. I see him as my mentor, someone I have learned from and someone I have relied on. He has my sincerest gratitude.

My especial thanks go to my mother. I wouldn't have made it this far, if it weren't for her. Thank you mom!

Contents

Abstract	iv
Kurzfassung	v
Acknowledgement	vi
Nomenclature	x
1 Introduction	1
1.1 Thermomechanics of solid microstructure	1
1.2 Mathematics: convexity and related topics	5
1.3 A short survey of the literature	7
2 Fundamentals	9
2.1 Laminate microstructure in martensite	9
2.1.1 Important features of accommodation	12
2.1.2 Shape-memory and pseudoplasticity	13
2.1.3 Superelasticity	14
2.2 Remarks on continuum mechanics	15
2.2.1 Rotation, stretch, and shear	16
2.2.2 Coherent solid interfaces	17
2.3 Basic crystallography	17
2.3.1 Lattice symmetries	19
2.3.2 Types of martensitic transformation	20
2.3.3 Cauchy–Born hypothesis: lattice-continuum link	20
2.4 Helmholtz free energy and its symmetries	21
2.4.1 Frame indifference (objectivity)	21
2.4.2 Material symmetry	21
2.5 Symmetries of martensitic transformation	22
2.5.1 Ericksen–Pitteri neighborhood	23
2.5.2 Mechanical twins vs growth twins	23

3	Continuum micromechanical model	25
3.1	Kinematics of laminate microstructure	26
3.1.1	Twinning equation	27
3.1.2	Periodic structure of the laminate and its geometry	28
3.1.3	Laminate in eigen configuration	33
3.1.4	Accommodation strains	34
3.2	Kinetics	36
3.2.1	Linear hyperelastic material—an interlude	38
3.2.2	Average laminate energy	40
3.2.3	Existence of solution to mesoscopic problem	42
3.3	Crystallography of NiTi martensite	43
3.4	Interface energies	49
3.4.1	Energy of coherent crystal interfaces	53
3.4.2	Martensite twin interface energy	61
3.5	Grain boundary energy	68
3.6	Chemical energies	73
3.7	The minimization problem	74
4	Finite element formulation	75
4.1	The minimization problem	76
4.2	Functional differentiation	76
4.3	Variational formulation	77
4.4	Boundary value problem: Euler–Lagrange equations	78
4.5	The Ritz–Galerkin method	80
4.5.1	Galerkin method for elliptic problems	81
4.5.2	Galerkin method for nonlinear problems	82
4.6	The finite element method	83
4.6.1	Finite elements	84
4.6.2	Piecewise polynomial test functions and quadrature rule	85
5	Numerical and analytical results at microscale	88
5.1	The basics	88
5.2	Single grain under simple tension	93
5.2.1	Convergence rate	97
5.2.2	Mesh independence	97
5.3	The influence of loading conditions	99
5.3.1	Deformation along twinning’s shear direction	101
5.4	A one-dimensional problem, analytical treatment	102
5.5	Homogenization: cuboid RVE with cylindrical inclusion	106
5.5.1	Mesh independence	107
5.6	Energy versus inclusion size	109

6 Conclusion and outlook	113
6.1 Summary	113
6.2 Possible extensions	114
A Miscellaneous mathematical notes and formulas	115
A.1 Some tensor algebra and calculus	115
A.1.1 Rotation matrix	115
A.1.2 Dyadic products	116
A.1.3 Fourth-order identity (unit) tensors	116
A.2 Average compatibility condition	117
A.3 An interpretation of Lagrange multipliers	117
B Elastic Constants of Monocrystalline	119
C Scaling and asymptotic self-similarity	122
D Nonlinear laminate kinematics, a possible extension	124
Bibliography	129

Nomenclature

Symbols

$\bullet$	Generic mathematical placeholder
$\hat{\bullet}$	Unit vectors (hat-accented)
$\check{\bullet}$	Degenerate, reduced, or limiting case of a function or functional $\bullet$
$\underline{\bullet}$	Mesoscopic quantity $\bullet$ / Lattice point along vector $\bullet$
$\#\bullet$	Cardinality of a set
$ \bullet $	Determinant of matrix / Absolute value of real number
$\ \bullet\ $	Norm as in normed vector space (to be understood from context)
$\bullet[\bullet]$	Functionals (arguments enclosed in brackets)
∇	Nabla operator
∇	Nabla operator on surface, e.g. along laminate interfaces
$\bullet \otimes \nabla, \nabla \otimes \bullet$	Gradient acting from right or from left
$\bullet \cdot \nabla, \nabla \cdot \bullet$	Divergence acting from right or from left
$\bullet \times \nabla, \nabla \times \bullet$	Curl acting from right or from left
$\nabla^s \bullet$	Symmetric gradient
$\bullet \nabla, \nabla \bullet$	Gradient acting from right or from left
$d\bullet$	Infinitesimal/differential element
$\delta\bullet$	Functional variation
$\Delta\bullet$	Finite increment or variation
$\mathbf{a}, \mathbf{b}, \mathbf{A}, \mathbf{B}, \boldsymbol{\alpha}, \dots$	Tensors and vectors (bold upright letters)
$\mathbf{a}_i$	Lattice vectors
$\mathbf{a}_i^\circ$	Parent phase lattice
$\mathbf{a}_i^a$	Austenite lattice
$\mathbf{a}_i^m$	Martensite lattice
$\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$	Lattice vectors
$\{\mathbf{a}^1, \mathbf{a}^2, \mathbf{a}^3\}$	Reciprocal lattice vectors

$\mathfrak{A}$	Lattice unit cell
$\mathfrak{A}_o$	Primitive lattice unit cell
$\mathcal{B}_0, \mathcal{B}(t)$	Continuum body in reference and current configuration
$\mathbb{C}$	Stiffness tensor
$\mathbf{E}_i$	Elementary matrix operator corresponding to i th variant
$\mathbf{F}$	Deformation gradient, $\mathbf{F} = \partial \mathbf{x} / \partial \mathbf{X}$ (microscopic scale, often)
$\mathbf{F}^\circ$	Eigen deformation gradient
$\underline{\mathbf{F}}$	Mesoscopic deformation gradient, $\underline{\mathbf{F}} = \partial \underline{\mathbf{x}} / \partial \underline{\mathbf{X}}$
$\mathcal{G}(\mathfrak{a}_i)$	Lattice point group
$\mathcal{G}_+(\mathfrak{a}_i)$	Lattice point group excluding shears, reflection, and inversions
$l^1 l^2 l^3$	Lattice coordinates
$l_1 l_2 l_3$	Lattice reciprocal coordinates
$\ell_1 \ell_2 \ell_3$	Lattice parameters (constants)
$\mathcal{L}(\mathfrak{a}_i, \mathfrak{r}_j)$	Crystal lattice system
$\mathcal{N}_\tau$	Set of nodes of element τ
$\mathfrak{r}_j$	Atom basis
$\mathbb{R}^N$	N -dimensional Euclidean space
$\mathcal{R}(\mathfrak{a}_i)$	Lattice translational group
$\text{SO}(3)$	Special orthogonal group in $\mathbb{R}^3$ (group of all proper rotations in 3D)
$\mathcal{T}_h(\Omega)$	Triangulation of the domain Ω
$\mathcal{V}(\Omega)$	Function space of solutions to BVP, VP, or MP
$\mathcal{V}_{\Gamma_D}(\Omega)$	$\mathcal{V}(\Omega)$ with homogeneous Dirichlet boundary
$\tilde{\mathcal{V}}(\Omega)$	Finite dimensional subspace $\mathcal{V}(\Omega)$
$\mathbf{x}$	Spatial/current coordinates
$\mathbf{X}$	Material/reference coordinates
$\mathbf{x}^\circ$	Spatial coordinates in eigen configuration
$\underline{\mathbf{x}}, \underline{\mathbf{X}}$	Mesoscopic current and reference coordinates
$\mathbf{y}$	Boundary coordinates
$\partial\Omega$	Boundary of domain
γ	Specific surface or interface energy
Γ_D	Dirichlet boundary
Γ_N	Neumann boundary
$\boldsymbol{\varepsilon}$	Linearized strain tensor
$\boldsymbol{\epsilon}$	Totally antisymmetric tensor, AKA Levi–Civita or Permutation tensor
η	Transformation strain

θ	Temperature / Spherical coordinate's angle relative to z -axis
Λ, λ	Volume fraction in reference and in current configuration
ϱ	Surface or interface density with units kg.m^{-2} or mol.m^{-2}
$\boldsymbol{\sigma}$	Cauchy stress tensor
τ_i	The i th element of a triangulation (mesh)
ϕ	Generic intensive thermodynamic state variable or potential
ψ	Specific Helmholtz free energy
$\omega(\underline{\mathbf{X}})$	Microscopic domain surrounding mesoscopic material point $\underline{\mathbf{X}}$
$\boldsymbol{\omega}$	Rotation tensor
Ω	A domain (often endowed with topological and/or algebraic structure)

Acronyms

1D, 2D, 3D	One-, Two-, Three-Dimensional
EPN	Ericksen–Pitteri Neighborhood
FE FEM	Finite Element Finite Element Method
HMM	Heterogeneous Multiscale Method
IVP	Initial Value Problem
IBVP	Initial-Boundary Value Problem
MP	Minimization Problem
NNBB	Nearest Neighbor Broken Bond
PDE	Partial Differential Equation
RVE	Representative Volume Element
SM SMA	Shape-Memory Shape-Memory Alloy
VP	Variational Problem

Logos

©	Copyrighted material
cc	Licensed under creative commons

Chapter 1

Introduction

This chapter makes the connection between the fundamental concepts and the particular objective of this dissertation. We mainly borrow from axiomatic thermomechanics, crystallography, continuum theory of martensitic transformation, homogenization methods, and also partly from the usual mathematics that goes with them such as functional analysis.

1.1 Thermomechanics of solid microstructure

Macroscopic characteristics of solid materials are manifestations of the aggregate behavior of their microscopic structure. In this sense, material science tries to understand and explain the thermal, mechanical, and electromagnetic properties of solids by bridging across a spectrum of *length scales* ranging from interatomic distances, i.e. $\sim 10^{-10}$ m, up to structural dimensions, i.e. $\sim 10^0$ m, or maybe higher in geomechanical problems. Regardless of the length scales, most often the fundamental assumption is that the underlying solid system is at the *thermodynamic limit*, which is to say that we are interested in those properties of the solid that are determined by a collection of many atoms ($\sim 10^{23}$). Exceptions are the *ab initio computations* where many-particle systems are simulated with each particle having several degrees of freedom, as required. But even in those computations often the ultimate objective is to simulate the aggregate behavior of the solid in terms of *first principles*, in order to deepen our understanding in a reductionistic sense.

A system is said to have reached the thermodynamic limit if under the assumption of homogeneity the number of particles, n , in any of its subsystems is proportional to the volume, V , of that subsystem. The subtlety of this assumption is that n is a discrete quantity whereas V is a continuous one. Continuity of V means that a volume element dV can be defined. But since $n \propto V$, the smallest physically meaningful volume is equal to $\frac{V}{n}$, which bounds dV from below. This contradicts the mathematical definition of an infinitesimal element which requires that $dV \rightarrow 0$. From a practical perspective, however, $\frac{V}{n}$ is diminutive especially in liquids and solids and also for gases in most circumstances. This gives birth to the concept of *continuum*, where all physical quantities of the system are assumed to be continuous. An infinitesimal element of any *extensive macroscopic quantity*,

especially the volume element, corresponding to a continuum solid body

- is large enough so that the thermodynamic limit is reached, i.e. spatial fluctuations of macroscopic quantities remain undetected, and yet
- it is small enough so that the macroscopic quantities can be assumed homogeneous across the element.

A closely related concept is that of equilibrium. A thermodynamic system is said to have reached equilibrium if its *macroscopic state* is well-defined. That is, if a set of independent *state variables* corresponding to the system are fixed, then there will be no apparent evolution of the system over time. When a system is at equilibrium, its hypothetical subsystems must be in equilibrium with each other. Then, corresponding to the thermodynamic state variables, there shall exist *thermodynamic potentials* that determine whether two systems in *contact* with each other, such as two subsystems of a system, can be in equilibrium or not. Conceptually, for a system to reach equilibrium one has to wait long enough so that any perturbation within the system dissipates. Alternatively, the so called quasi-static conditions can be assumed, which is basically equilibrium under non-equilibrium! This requires that any disturbance within the system or at its boundaries propagates with a wave length much larger than the dimensions of the system¹.

From the above descriptions it is already clear that the equilibrium of a system is not absolute. Even in the absence of large gradients and sudden changes the temporal fluctuations in the system might become noticeable in certain observations. The same is true for homogeneity of a system—being relative. A solid medium is usually made up of various phases at the microscale. Even a monophasic solid is often made up of *grains*, which at least differ in their lattice orientations. Mathematically, the *spatial* and *temporal* averages of the microscopic degrees of freedom pertaining to these inhomogeneities and fluctuations are the thermodynamic state variables and potentials whose existence are postulated within the axiomatic framework of thermodynamics. The idea of traversing time and length scales through averaging procedures can be generalized to continuum theory of solids, where both macroscopic solid and its *microstructure*² are assumed to be continua. This encompasses a variety of methods, depending on the type of inhomogeneity and the objective of averaging.

The averaging procedure can be employed to obtain *effective macroscopic properties* of the solid. This is termed *homogenization*. Here, specific distributions (often probabilistic) are assigned to the microstructural inhomogeneities, which are seen as texture. Then a

¹In thermomechanics of solids, the thermodynamic variables are classified as external and internal. The external variables are determined and controlled by the system's surrounding. The *internal variables* are of two types. First, *internal state variables* which have no inertia and their changes cannot correspond to external work. Therefore, the power expended by them can only be dissipative, and their variations do not pertain to any flux, and the energy content of the system does not change under such variations. Second, *internal degrees of freedom* which have inertia and expend external work, therefore have flux and obey balance equations. They usually express history-dependent behavior [21]. In other words internal variables describe those degrees of freedom that can be altered without any change in the surrounding [101].

²Solid microstructures are basically small-scale inhomogeneities possessing varying degrees of regularity and order.

weighted average of the macroscopic property is obtained based on the solution of an auxiliary microscopic optimization or boundary value problem. Homogenization by itself does not yield much information on why and how a certain composition and pattern is formed at the microscale; and it does not usually answer the fundamental questions such as stability, optimality, and uniqueness of the particular configurations that emerge. Of course, through homogenization one can eventually generate enough data to be able to hone the macroscopic behavior of the material by modifying its microstructure. But this would be a rather phenomenological approach that is more suited to practical purposes [100].

To answer fundamental questions, such as why and how a specific microstructure appears, we can still use the ideas behind homogenization, but we need to augment the averaging procedure with additional features. Namely, we need to confine the phenomenological and heuristic postulates to a bare minimum or avoid them entirely if possible. This is facilitated by relying on the thermomechanics of the microstructure, which in principle should eliminate the need for heuristic assumptions. This, at the same time, will enable us to make the connection between the micro- and macroscale both ways. Notice that in homogenization the focus is almost entirely on micro-to-macro transition. Further, we have to study the optimality of the macroscopic effective properties with respect to microscopic degrees of freedom, rather than just methodically compute them.

The thermodynamic concept of a *phase* is central to understanding microstructures. Generally, a phase is a body of a substance that is homogenous with respect to all its observable properties, both chemical and physical [101]. Different phases of a pure substance are distinguished by the so called *order parameter*, which is a thermodynamic state variable. This newly introduced state variable is not necessarily an additional one, i.e. does not necessarily add to the number of degrees of freedom. An order parameter is sometimes simply one of the readily identified state variables and some other times it replaces one of them. In any case, it is crucial to understand that the choice of an order parameter is neither unique nor straightforward. Nevertheless, it is often directly or indirectly related to a first derivative of the *Gibbs free energy*.

A solid phase of a pure chemical composition appears as a crystalline lattice. A widely used order parameter for a crystal is the lattice displacement field which is defined as the displacement of the lattice points from their original position, i.e. the lattice configuration of lowest energy. All crystal lattices are translationally symmetric, that is, their energy does not alter under collective lattice translation. Therefore, the lattice displacement field is replaced by lattice deformation gradient. Consequently, the most suitable thermodynamic potential to describe the equilibrium state of such a solid system is the *Helmholtz free energy*, designated by $\Psi(\mathbf{F}, \theta)$, in which $\mathbf{F}$ denotes the deformation gradient tensor and θ is the temperature.

Remark 1. The terms “energy” and “energy density” are often used interchangeably throughout this text. The precise meaning shall be clear from the context in each particular case, at least through physical dimensions.

The free energy *density* of a stable monophase crystalline should be *convex* so that there

exists a *ground state* corresponding to a unique minimum of the free energy. This minimum energy, unambiguously called the ground energy, corresponds to $\psi_0(\theta) = \psi(\mathbf{I}, \theta)$. There are, however, crystalline solids whose free energies are non-convex depending on temperature. Fig. 1.1 schematically illustrates an instance of a non-convex energy versus the order parameter, i.e. lattice deformation gradient, for varying temperatures. At high temperatures, it has *metastable* phases at $\mathbf{F}_i^\circ$ and $\mathbf{F}_j^\circ$, and a stable phase at $\mathbf{F} = \mathbf{I}$. At low temperatures the two phases at $\mathbf{F}_i^\circ$ and $\mathbf{F}_j^\circ$ become stable, and $\mathbf{F} = \mathbf{I}$ becomes metastable instead. In between these two regimes, there is a certain temperature θ_c , called transformation temperature, at which all the the states of local minima have equal energies. However, this state is very short-lived and the transition from high temperature regime to low temperature regime is spontaneous, such that ψ is not differentiable with respect to θ , which signifies a *first order phase transformation*³.

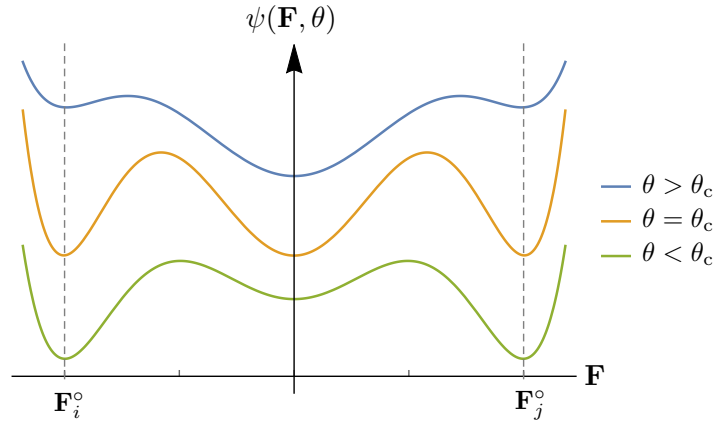


Fig. 1.1: Schematic free energy of a cubic lattice above, at, and below transformation temperature.

High temperatures are associated with intense lattice vibrations and hence disorder, because stronger thermal fluctuations make it less likely that the system becomes trapped in a subset of *microstates* that can pertain to a stable or *metastable* macrostate. In a given thermodynamic system, disorder is synonymous with symmetry. An ultimate example is a liquid phase. All liquids possess full symmetry, viz *continuous* translational and rotational symmetry, because they have no lattice structure which implies complete disorder. So, the above *solid-to-solid* phase transition accompanies development of order in the system, in the sense of reducing symmetry. Therefore, solid-to-solid phase transformations are said to be *symmetry breaking*, which is mathematically identified when the *symmetry group* of a stable low temperature phases is a strict subgroup of the symmetry group of the globally stable high temperature phase, i.e. $\mathcal{G}(\mathbf{F}_i^\circ) < \mathcal{G}(\mathbf{I})$. We summarize these concepts in the following definition. The high-symmetry phase is called *austenite* and the low-symmetry phase is called *martensite*.

³The order of a phase transition equals the order of lowest derivative of the free energy that is discontinuous at the transition state with respect to the varying state variable, or at the phase boundary with respect to order parameter. For instance, discontinuity of ψ with respect to θ means a discontinuity of entropy which is identified as the *latent heat*.

Definition 1. MARTENSITIC PHASE TRANSFORMATION is a reversible diffusionless first-order symmetry breaking solid-to-solid phase transformation.

Reversibility of the transformation will be clarified later when we lay down its crystallographic aspects. The importance of the reversibility is that it results in a symmetry connection between the low-symmetry states of energy minima, which are therefore called *martensite variants*. In Fig. 1.1 these variants correspond to $\mathbf{F}_i^\circ$ and $\mathbf{F}_j^\circ$. We will explain later how the symmetry groups of the variants are connected to each other and to that of austenite. In the present context, the most significant result of these crystallographic considerations is the possibility of the so called *twin-compatible* variants and/or phases. This condition is best formulated in terms of a kinematic constraint, called *coherence condition*, on transformation strains⁴. Two coherent phases or variants are capable of forming *twins*, which is the very basis of the *laminate microstructure* that is central to our work. As it so happens, this kinematic constraint has the form of a *rank-one connection* between the transformation strains which has some deep mathematical implications, as we will briefly address.

1.2 Mathematics: convexity and related topics

The most general mathematical formulation of a thermomechanical problem is in the form of minimization of an *energy functional* in terms of ψ . As mentioned, whenever ψ is a (strictly) convex function of the order parameter, provided some additional assumptions such as *coercivity* of ψ , the existence and uniqueness of a minimizer is guaranteed [41, pp. 106]. Further, assuming that the minimizers are smooth enough, e.g. belong to C^2 , the convexity of ψ ultimately causes the Euler–Lagrange equation(s) to be elliptic, and the unique minimizer to be their solution. In the case of non-convex energy densities, such as martensite, the existence of a minimizer to the energy functional can still be investigated but its uniqueness becomes almost intractable. Instead, the focus is turned towards regularity of minimizer(s). It turns out that the existence of a minimizer does not necessarily require strict convexity of the energy density, but rather relies on a much broader condition called *weak lower semicontinuity* [5, 41]. Consequently, the notion of convexity can be generalized to *polyconvexity*, *quasiconvexity*, and *rank-one-convexity*⁵, in order of weakness (as opposed to strictness in a mathematical sense), that is,

$$\text{Convexity} \subseteq \text{Polyconvexity} \subseteq \text{Quasiconvexity} \subseteq \text{Rank-1-Convexity}. \quad (1.1)$$

It is then immediate that for every function $f : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}$, for which the respective *convex envelopes* can be defined, we have

$$\mathcal{C}f \leq \mathcal{P}f \leq \mathcal{Q}f \leq \mathcal{R}f, \quad (1.2)$$

⁴Transformation deformation gradients $\mathbf{F}_i^\circ$, and transformation strains $\boldsymbol{\eta}_i$ are sometimes used interchangeably throughout this text. The meaning should be obvious from the context.

⁵Rank-one-convexity and ellipticity are equivalent for C^2 functions.

where for every $\xi \in \mathbb{R}^{n \times m}$, the convex envelope is defined as

$$\mathcal{C}f(\xi) = \sup \{g(\xi) : g \leq f \text{ and } g \text{ is convex}\} ,$$

and similarly for polyconvex $\mathcal{P}f$, quasiconvex $\mathcal{Q}f$, and rank-one-convex $\mathcal{R}f$ envelopes.

A well-known theorem of the calculus of variations states that if the energy functional is coercive and follows a monomial growth rate, the notions of polyconvexity and weak lower semicontinuity coincide. This guarantees the existence of a minimizer for quasiconvex energies. Furthermore, if ψ is a quadratic form in terms of the order parameter, then the notions of quasiconvexity and rank-one-convexity are equivalent, which again guarantees the existence of minimizer(s) for the energy functional, particularly if a linear hyperelastic material is adopted. We mentioned also that the martensitic variants are rank-one-connected through the kinematic coherence condition. At the same time, these variants pertain to the extreme points of the energy landscape—being at the local minima. It follows that according to Minkowski theorem⁶ the rank-one-convex envelope constructed from the coherent martensitic variants coincides with the general rank-one-convex envelope of the free energy. This construct is often termed the *relaxed energy*, and the process of constructing it is called *energy relaxation*. The relaxation procedure is usually based on either the quasiconvex envelope or the rank-one-convex envelope. So, the basic idea behind the majority of material models for martensitic phase transformation is to substitute the non-convex energy with a piecewise quadratic energy which is then replaced by its relaxed counterpart. In practical terms, relaxation of free energy can be viewed as a homogenization approach, and therefore carries some of the common traits of homogenization with it. One last remark is in order here. The relaxation theorems are mathematically stated for arbitrary domains, which renders them ineffective for applied purposes. Fortunately, however, it can be shown that if a function is quasiconvex with respect to periodic perturbations (test functions) over the unit hypercube, then it is quasiconvex for arbitrary domains and vice versa [41, pp. 173].

Although the energy relaxation makes certain that the energy functional attains a minimum, it has its own drawbacks. First of all, the relaxed energy is bound to underestimate the global energy of the solid because, as its name suggests, it poses weaker conditions on the free energy density. These weaker conditions correspond to replacing the kinematic coherence condition with a much more lenient requirement, called the *average compatibility conditions*. This, however, is not a serious shortcoming in most circumstances, since calculation of the energy *upper-bounds* show very little difference to the relaxed energy, which sets an energetic lower bound. This implies that using the lower bound, i.e. relaxed energy, is good enough an estimate to the free energy in most situations [63, 64]. Similarly, the relaxed energy fails to capture most of the small-scale characteristics of the solid, because their corresponding microscopic degrees of freedom are averaged out through the relaxation procedure. Most notably, the average compatibility conditions leave out a precise representation of the solid surfaces and interfaces, which play a crucial role at small length scales. Notice that the surface energy density scales as $\psi_\gamma \propto 1/l$, in which l is the characteristic length scale of the

⁶Minkowski theorem asserts that the convex hull of a compact set coincides with the convex hull of its extreme points [41, pp. 338].

thermomechanical solid; whereas the free energy density is independent of l . This means that eventually the surface energies become decisive as the length scale drops, so much so that the *energetic competition* between at least three contributions must be accounted for, namely

- the free energy of the bulk of solid, which is often identified as elastic strain energy in the present context,
- the energy of the coherent phase boundaries, which we call twin interfaces,
- and the boundary energy of the intended micro-domain, which itself originates in part from microstrain energies yet at a smaller length scale.

Before we proceed, let us further emphasize the significance of interface phenomena in solid materials in general. Solid interfaces and surfaces act as barriers to various dynamic and transport processes which hinder the evolution of thermomechanical solid system towards a state of global energetic minimum. Most solid systems at best march into energetic minima only locally in such a way that the overall energy of the solid body is the resultant of a *bifurcation* into any of the innumerable meso- or microstates, which are generally indistinguishable from each other for a macroscopic observer. So in this sense, we can identify the surface and interface phenomena as the major factor in formation of solid microstructures, which ultimately determines the macroscopic behavior of solid materials.

The present work is concerned with developing a continuum micromechanical model for the laminate microstructure of martensitic materials that takes the surface and interface energies into account. This is accomplished by enforcing the coherence condition *almost* strictly at the microscale, which translates to a periodic kinematic ansatz in the form of a *first-order laminate*. We assign different elastic stiffness tensors to the two twin-compatible martensitic variants in their **exact** orientations relative to each other and to the parent phase, which is unprecedented as far as we know. The twin interface energies are estimated analytically based on crystallographic theory and lattice potentials of the martensitic phase transformation based on its exact lattice correspondence. Some of the derivations relate to the particular case of NiTi (49.75 at. % Ni), but in principle they should be generalizable to other martensitic materials. The grain boundary energies are estimated based on a self-similar ansatz. This ansatz is assumed isotropic, i.e. the relative orientation of neighboring grains and the crystallographic orientation of grain boundaries are neglected.

1.3 A short survey of the literature

A brief look at the relevant literature in a semi-chronological order is presented here. This is a mere personal tale of how the author has encountered and traced some of the references back in time and is neither comprehensive nor historically accurate. The general continuum micromechanical theory of polycrystalline can be traced back to Hill 1965 [73] whose work relied mostly on the famous papers of Eshelby 1953–1961 [48, 49, 50]. These together with the mathematical theory of polycrystalline solids that was established by Hashin and Shtrik-

man 1961–1962 [68, 69, 70] in a variational framework were the basis for many of the later developments.

The pioneering works of Khachaturyan, Shatalov and Roitburd 1966–1974 [93, 92, 136, 137] set the stage for understanding the thermomechanical aspects of martensitic phase transformation and its microstructure. From a mathematical standpoint, the various notions of convexity were rigorously investigated by Ball et al 1976–1984 [5, 10, 13] which to some extent were based on the earlier work of Morrey 1952 [110]. Later on these ideas found extensive use in the rationalization of continuum micromechanics of polycrystalline, especially the martensite, and were further developed by Ball, James, Chu, and Carstensen 1987–2004 [11, 12, 9, 7, 8, 6]. Of particular relevance to our work are the extensions and developments by Kohn, Otto, and Müller 1991–1997 [96, 98, 112, 97, 99]. Their work covers a lot of ground spanning from polyconvexity, relaxation methods, scaling laws, and energy estimates all the way to coherence conditions, interface effects, and pattern formation. The reviews by Müller 1999 [113] and by Conti 2000 [38] give relatively accessible roundups of these concepts. Along the same lines, the exposition by Dacorogna 2012 [41] is much recommended for the mathematically inclined reader.

Among the above mentioned works, the more recent ones rely to a large extent on the works of Ericksen 1979–1991 [42, 43, 44, 45, 46], and Pitteri and Zanzotto 1984–1998 [125, 126, 127, 128] who took especial interest in making a precise connection between the crystallographic theory of martensitic phase transformation and the continuum theory. They thoroughly studied the symmetry implications of non-degenerate continuum mechanics of solid phase mixtures and transformations. This is particularly significant to our work since we especially focus on interface phenomena which dependent on coherence conditions and the entailing symmetry connections between the martensitic variants. The works of Bhattacharya et al 1991–2003 [22, 23, 24, 26, 25] can be also categorized in this group.

The general crystallography of martensitic phase transformation was well established as early as 1956 by Bilby and Christian [28]. In the particular case of NiTi, Otsuka et al 1999–2005 [121, 133, 134, 122] have done precise and comprehensive research.

Other than the above mentioned works of Kohn, Müller and Otto which investigate various mathematical aspects of interface effects in the martensitic microstructure, the computational work of Hackl et al 2004 [66] studies the twin interface energy and predicts general accommodation traits by branching of the microstructure near the boundaries. Also Stupkiewicz et al 2002–2012 [146, 144, 145] present a relatively versatile model on twin interface and the grain boundary effects especially the optimal morphologies of the laminate at the boundaries.

The present work sets itself apart by making various extensions and generalizations based on fewer and less restrictive assumptions. It develops an all-around model, including surface and interface effects, capable of reflecting micro-, meso-, and macroscopic features of the martensitic laminate microstructure simultaneously.

Chapter 2

Fundamentals

This chapter revisits some of the topics that will be used throughout this work. It serves merely as an entry point. It neither tries to be comprehensive, nor does it mean to give an exhaustive index of the concepts that will show up in the sequel. Some more elementary formulas and topics have been moved to the appendices so that they do not interrupt the flow of the text.

2.1 Laminate microstructure in martensite

The choice of martensitic materials for this dissertation is motivated by their extraordinary properties that originate from what we have explained so far about their microscopic features. Most conspicuously, the martensitic laminate microstructure has been identified as the origin of *shape-memory* (SM) phenomenon. We demonstrate a contrived example, in terms of a scalar displacement gradient, of how shape-memory and other closely related concepts work, in order to set the stage for a full tensorial representation that will follow. Consider the following energy minimization problem

$$\min_u \left\{ \mathcal{E}(u) = \int_0^L \int_0^1 ((\partial_x u)^2 - 1)^2 \, dx \, dy \mid u : \Omega \rightarrow \mathbb{R}, u(\partial\Omega) = 0 \right\}, \quad (2.1)$$

where $\Omega = (0, 1) \times (0, L)$. The energy density $\psi = ((\partial_x u)^2 - 1)^2$ has two minima at $\partial_x u = \pm 1$, where it attains $\psi_{\min} = 0$ (see Fig. 2.1). Since ψ is not convex, the energy functional $\mathcal{E}(u)$ does not possess an optimal solution in a classical sense. Hence, we construct a periodic ansatz of the deformation field u for all $N \in \mathbb{N}$, as

$$u_N(x, y) = \begin{cases} x - n/N & x \in [n/N, (n + 0.5)/N) \\ -x + (n + 1)/N & x \in [(n + 0.5)/N, (n + 1)/N] \end{cases} \quad (2.2)$$

where $n = 1, 2, \dots, N$.

Fig. 2.2 shows the laminate construct (2.2) for $N = 5$. The displacement gradient is equal to $+1$ or -1 almost all-over the domain except within the two jagged boundary layers

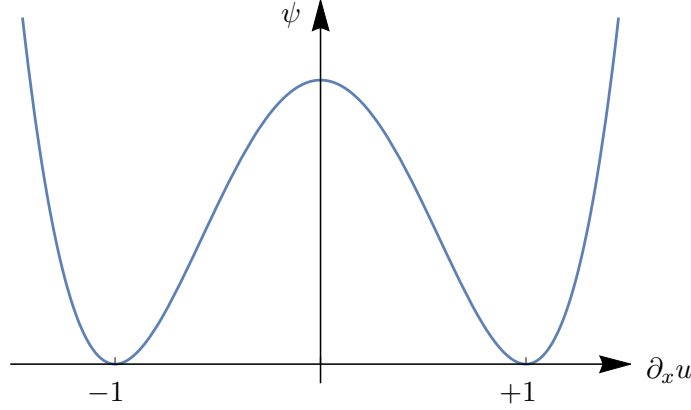


Fig. 2.1: Double-well energy corresponding to $\psi = ((\partial_x u)^2 - 1)^2$.

next to $y = 0$ and $y = L$. This means the stored energy equals zero except at the boundary layers, whose total area is proportional to $1/N$. If we assume that the energy density ψ remains finite within boundary layers, we can conclude that the total stored energy at the boundary vanishes as $N \rightarrow \infty$. Therefore,

$$\lim_{N \rightarrow \infty} \mathcal{E}(u_N) \rightarrow 0, \quad (2.3)$$

which states that the minimization problem (2.1) has a minimizing sequence in the form of the laminate ansatz (2.2).

Despite its mathematical validity, the above result gives rise to a physical problem which is when N tends to infinity the width of each twin, $1/N$, tends to zero. This is unrealistic since the *lath width* (twin width) cannot be smaller than the corresponding interplanar lattice spacing. Here, the twin interface energy comes to rescue. The energy functional in (2.1), assuming sharp twin interfaces, is rewritten as [97, 99]

$$\mathcal{E}(u) = \int_0^L \int_0^1 ((\partial_x u)^2 - 1)^2 + \gamma |\partial_x^2 u| \, dx \, dy. \quad (2.4)$$

If we substitute the ansatz (2.2) into the latter functional, the resultant interface energy will be equal to $2L\gamma N$, which grows indefinitely as $N \rightarrow \infty$. Hence, the interface energy counterbalances the elastic energy and sets a finitely small lower bound to the lath width, $1/N$.

The jagged construct at the vicinity of the boundaries, at $y = 0$ and $y = L$, is conceivably the simplest model for the boundary. Except for very particular cases where the so called *average coherence condition* between the laminate and its neighboring phase across the boundary is fulfilled, the laminate microstructure is not so uniform as in (2.2) [23, 146, 83]. It is well established, both empirically and theoretically, that the martensitic laminate microstructure *accommodates* its discrepancies with the surrounding phases by

- varying its lath width, often called *branching*, and
- by varying the volume fractions of variants, often called *needles*,

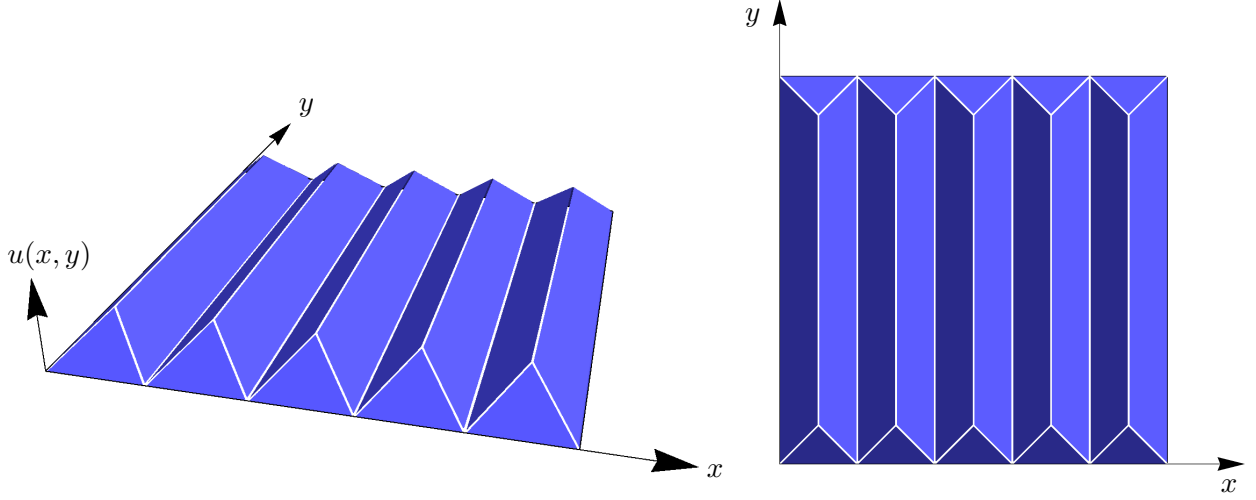


Fig. 2.2: The laminate microstructure composed of equal fractions of variants with $\partial_x u = +1$ and $\partial_x u = -1$, illustrated in three dimensions (left) and its top view (right).

as it approaches the boundaries that are not parallel to the twin interfaces. In thermomechanical terms, the accommodation is again an effort to minimize the overall energy of the laminate when the boundary energy is taken into account. Accordingly, the energy functional (2.4) shall be further modified to reflect the accommodation strains. A simple version can be given as

$$\mathcal{E}(u) = \int_0^L \int_0^1 \varphi(x, y) \, dx \, dy + \int_0^{-\infty} \int_0^1 \beta(x, y) \, dx \, dy + \int_L^{+\infty} \int_0^1 \beta(x, y) \, dx \, dy, \quad (2.5)$$

where $\varphi(x, y) = ((\partial_x u)^2 - 1)^2 + (\partial_y u)^2 + \gamma |\partial_x^2 u|$ and $\beta(x, y) = c_1 (\partial_x u)^2 + c_2 (\partial_y u)^2$. The integrals outside the laminate domain relate to strain energy propagated into the surrounding medium, which is caused by the deformation mismatch between the laminate and its surroundings. The two boundary energy terms can be simplified using dimensional considerations and Fourier analysis to (see [76, 18])

$$\mathcal{E}_\Gamma(u) = c \left(\sum_k |k| (\hat{u}_0(k))^2 + \sum_k |k| (\hat{u}_L(k))^2 \right). \quad (2.6)$$

The amplitude $\hat{u}$ of the frequency $k = 1/N$ is significantly larger than all the others, for apparent physical reasons. This makes the corresponding term to have the decisive magnitude in the boundary energy, i.e. $\mathcal{E}_\Gamma(u) \approx c/N ((\hat{u}_0(1/N))^2 + (\hat{u}_L(1/N))^2)$. Therefore, it is clear that the boundary energy favors

- smaller lath width, and
- low amplitude oscillation of the deformation field

towards the boundary, which are synonymous to laminate branching and formation of needles, respectively [38] (see Fig. 2.3 and Fig. 3.1). The question of whether these configurations are dynamically accessible from the energy well (ground state) is a subtle one ([99] pp. 280, and [5] pp. 399).

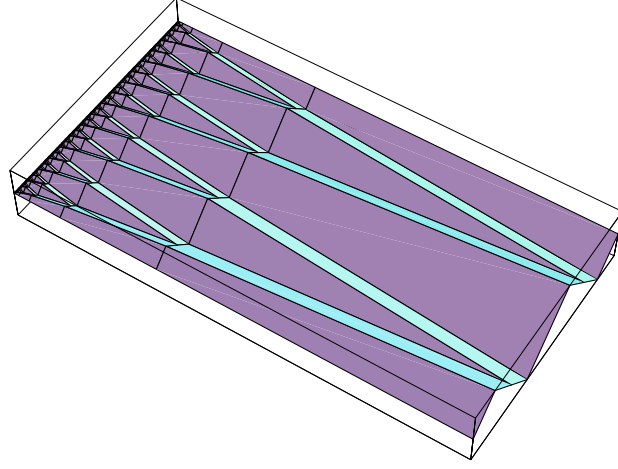


Fig. 2.3: An example construct of microstructural accommodation by branching [38, © permission obtained].

2.1.1 Important features of accommodation

The accommodation patterns are not unique and there are many equally favorable branching patterns that fulfill the above minimization problem. Additionally, the two accommodation responses, i.e. branching and needles, compete. Branching becomes the energetically favored accommodation response when [98, 38]

$$\frac{\|\mathbb{C}_r\|}{\|\mathbb{C}_m\|} \gg \left(\frac{\gamma}{\|\mathbb{C}_m\| L} \right)^{1/3}, \quad (2.7)$$

which takes place in one of the following physical situations

- The laminate is extended very long $L \gg h$.
- The surface energy is very small $\gamma \ll 1$.
- The surrounding phase is much stiffer than the martensite $\|\mathbb{C}_r\| \gg \|\mathbb{C}_m\|$.

Whenever branching is dominant, the laminate's characteristic length, i.e. the lath width, scales as

$$h(x) \sim \left(\frac{\gamma}{\|\mathbb{C}_m\|} \right)^{1/3} x^{2/3}, \quad (2.8)$$

where x is the distance from the boundary. Depending on the softness of the boundary, complex branching patterns may couple with needles near the boundary and the total stored energy scales as the minimum between the energy of the laminate with rigid boundary and the one of the elastically deformable boundary [99, pp. 278]:

$$E_\gamma^{(6)} \sim \min \left\{ \|\mathbb{C}_r\|^{1/2} \gamma^{1/2} L^{1/2}, \gamma^{2/3} L^{1/3} \right\}. \quad (2.9)$$

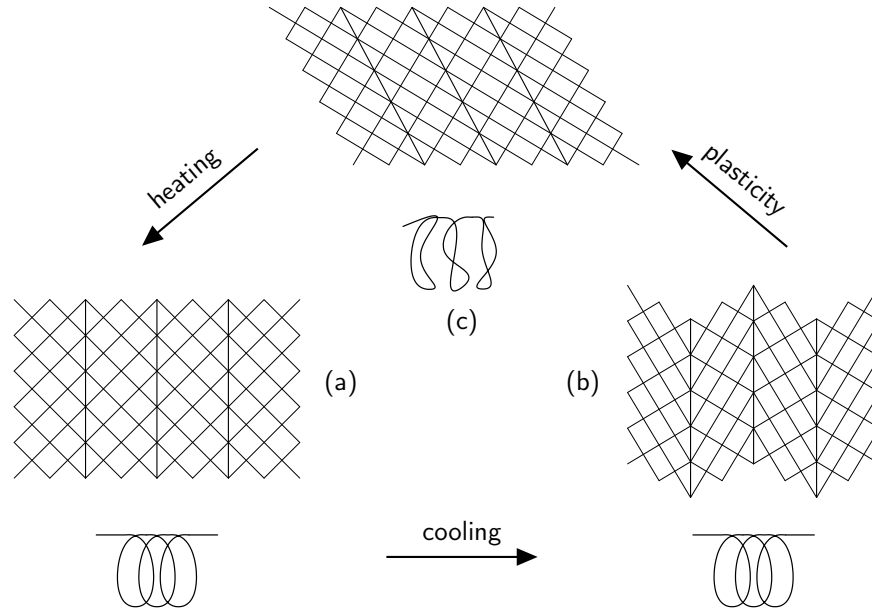


Fig. 2.4: Schematic illustration of shape-memory phenomenon and the underlying mechanisms of temperature-induced martensitic phase transformation [25, © permission obtained].

2.1.2 Shape-memory and pseudoplasticity

Fig. 2.4 is a schematic illustration of the shape-memory effect in martensitic materials and how it relates to the microstructure. The specimen goes through a cycle between three states as follows.

1. The specimen is kept at temperatures much higher than the transformation temperature, having an entirely austenitic composition. It is then shaped into a particular form, which is designated as state (a) on the figure.
2. Keeping the specimen constrained to stay in its form, it is cooled down to temperatures well below the transformation temperature. In order to fulfill the kinematic constraints of its fixed shape, the material forms laminate microstructure as it transitions from state (a) to (b). Now, if the constraints are removed, the specimen keeps its *trained* shape on its own.
3. With temperatures below the transformation, if the specimen goes through deformations large enough to endure plasticity, the microstructure reconfigures to state (c). The plastic slip systems are strongly influenced by the presence of the twin interfaces, in such a way that heating the specimen again back to temperatures well above the transformation temperature will cause the crystal structure to relapse into its original trained shape in the austenite phase.
4. In a final step, cooling down the specimen back to martensite without the presence of any constraints will result in the microstructure in state (b).

This is how the material remembers its shape through the cycle of heating $\rightarrow$ training $\rightarrow$ cooling $\rightarrow$ plastic deformation $\rightarrow$ heating.

The plastic deformation in the shape-memory cycle differs substantially from what is commonly understood as plasticity, because the material finally restores its original (trained) shape by heating. This is why the recoverable plastic deformation of martensite below transformation temperature is often called *pseudoplastic* deformation. Formulation of a material model on pseudoplasticity and the shape-memory effect requires a special treatment of *dissipation* due to its differences from the ordinary plastic behavior in solids. Ideally, a thorough investigation of the dissipation in martensite would require an exact representation of evolution of the twin interfaces. Because at the microscopic level the appearance and disappearance of twin interfaces is a major cause of dissipation. In this respect, our kinematic ansatz of the laminate will set the stage for such material models, since it enforces the coherence condition almost strictly at the twin interfaces.

2.1.3 Superelasticity

We already showed how a solid *mixture of phases* fulfills homogeneous boundary conditions and at the same time approaches a global energetic minimum. For a double-well free energy density as Fig. 2.1 this is astonishing because satisfying homogeneous boundary conditions pertains to $\partial_x u = 0$, for which the free energy of a monophase solid would be far from minimum. This can be further generalized to situations where a nonzero homogenous strain, i.e. $\langle \partial_x u \rangle \neq 0$, is prescribed. Assume that the laminate ansatz (2.2) is modified to

$$u_N(x, y) = \begin{cases} x - n/N & x \in [n/N, (n + \lambda)/N) \\ -x + (n + 2\lambda)/N & x \in [(n + \lambda)/N, (n + 1)/N] \end{cases} \quad (2.10)$$

where λ is the volume fraction of variant $\partial_x u = +1$. Then, the average displacement gradient of the laminate along the x -axis is not zero anymore, but rather

$$\langle \partial_x u \rangle = 2\lambda - 1. \quad (2.11)$$

Notice that the ansatz (2.10) still fulfills the sequential convergence to the minimum energy, as (2.3), which means that deforming the laminate in this particular direction does not cost elastic energy. This is the origin of the so called *superelastic* or *pseudo-elastic* behavior, where loading the austenite slightly above the transformation temperature indicates an almost linearly elastic behavior followed by a deformation-induced martensitic transformation. This transformation causes a plateau in the stress–strain curve (Fig. 2.5), as if the material is deforming plastically. However, unloading the specimen will result in an almost complete return to the austenitic phase manifested in the complete recovery of the original shape.

For a comprehensive narrative of the shape-memory, superelasticity, and pseudoplasticity in martensitic materials and their respective underlying mechanisms, the interested reader may refer to [88, 65, 19].

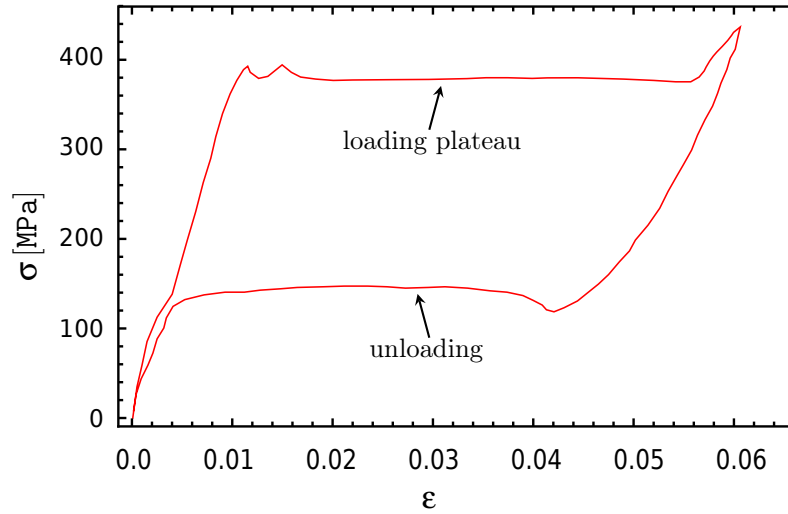


Fig. 2.5: Superelastic behavior in a tensile test of NiTi in an experiment by McNaney et al [107, © permission obtained].

2.2 Remarks on continuum mechanics

Since the Helmholtz free energy depends on the deformation field, we shortly address some of the terminology and concepts of continuum mechanics which are particularly relevant to our upcoming discussions. For further reading please refer to the textbooks by Chou and Pagano [35], Holzapfel [75], Ogden [118], Marsden and Hughes [106], Ciarlet [37], in order of sophistication and mathematical rigor.

The basic kinematic assumption of continuum mechanics is that the position of a solid body $\mathcal{B}$ at any given time t can be expressed via a differentiable map $\chi : \mathbb{R}^3 \times \mathbb{R}^+ \rightarrow \mathbb{R}^3$, which gives the space-time trajectory of all the particles of the body as a function of initial position in the so called *reference configuration*, $\mathcal{B}_0$, and time:

$$\mathbf{x} = \chi(\mathbf{X}, t), \quad \mathbf{X} \in \mathcal{B}_0, \quad \mathbf{x} \in \mathcal{B}(t). \quad (2.12)$$

The existence of a reference configuration is simply postulated. It can be the state of $\mathcal{B}$ at any time for which all the relevant thermomechanical variables are known. Accordingly, the state of the body at the present time, $\mathcal{B}(t)$, is called the *current configuration*. In kinematical contexts, the reference and current configurations are sometimes termed *undeformed* and *deformed* configurations, respectively.

An intensive thermodynamic state variable or potential, denoted by ϕ , corresponding to $\mathcal{B}$ can be expressed as a *field* variable in either of the two forms:

1. $\phi(\mathbf{X}, t)$, which is known as the *Lagrangian description* of ϕ , and $\mathbf{X}$ is called *material coordinates*.
2. $\phi(\mathbf{x}, t)$, which is known as the *Eulerian description* of ϕ , and $\mathbf{x}$ is the *spatial coordinates*.

The general motion of a continuum can be decomposed into three modes, namely translation, rotation, and deformation. The principle of *objectivity* states that translations and

rotations may not alter the free energy of the solid body. Hence, we focus on deformation. Intuitively, deformation occurs when the relative distance between any two particles of $\mathcal{B}$ changes. The pointwise measure of this change is called *strain*, which is a tensor field over $\mathcal{B}$. The following table lists some of the more common strain measures and their corresponding thermodynamic conjugate with respect to the Helmholtz free energy. Notice that $\boldsymbol{\sigma}$ is the Cauchy stress tensor, $\mathbf{F}$ is the deformation gradient, and J is the Jacobian determinant of $\mathbf{x}(\mathbf{X})$.

Strain	Stress	Configuration
Almansi (true) $\mathbf{e} = \frac{1}{2} (\mathbf{I} - \mathbf{F}^{-T} \mathbf{F}^{-1})$	Cauchy (true) $\boldsymbol{\sigma}$	Current
Almansi (true) $\mathbf{e}$	Kirchhoff $\boldsymbol{\tau} = J \boldsymbol{\sigma}$	Reference
Deformation gradient $\mathbf{F} = \frac{\partial \mathbf{x}}{\partial \mathbf{X}}$	1 st Piola–Kirchhoff $\mathbf{P} = J \boldsymbol{\sigma} \mathbf{F}^{-T}$	Reference
Green–Lagrange $\mathbf{E} = \frac{1}{2} (\mathbf{F}^T \mathbf{F} - \mathbf{I})$	2 nd Piola–Kirchhoff $\mathbf{S} = J \mathbf{F}^{-1} \boldsymbol{\sigma} \mathbf{F}^{-T}$	Reference

Tab. 2.1: Various strain measures and the corresponding conjugate stresses.

2.2.1 Rotation, stretch, and shear

There are particular modes of deformation that are important to any continuum mechanical study of solid-to-solid phase transformations. We mention the following theorem, without proof, which tells us how the deformation gradient is uniquely decomposed into rotation and *stretch* tensors.

Theorem 1. POLAR DECOMPOSITION THEOREM. *Every non-singular $\mathbf{F} \in \mathbb{R}^{3 \times 3}$ with $\det \mathbf{F} > 0$, can be uniquely represented as*

$$\mathbf{F} = \mathbf{Q} \mathbf{U} = \mathbf{V} \mathbf{Q}, \quad (2.13)$$

with $\mathbf{Q} \in \text{SO}(3)$, and the symmetric positive-definite tensors $\mathbf{U}$ and $\mathbf{V}$ are called the right and left stretch tensors.

Another important mode of deformation is the so called *simple shear*, which is when the deformation gradient $\mathbf{F}$ can be represented as

$$\mathbf{F} = \mathbf{I} + \mathbf{a} \otimes \mathbf{b}, \text{ such that } \det \mathbf{F} = 1, \text{ and } \mathbf{a} \cdot \mathbf{b} = 0. \quad (2.14)$$

The following lemma will be utilized in our upcoming derivations.

Lemma 1. For any two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, such that $\mathbf{a} \cdot \mathbf{b} = 0$, we have

$$(\mathbf{I} + \mathbf{a} \otimes \mathbf{b})^{-1} = \mathbf{I} - \mathbf{a} \otimes \mathbf{b}.$$

Proof. Multiply $(\mathbf{I} - \mathbf{a} \otimes \mathbf{b})$ by $(\mathbf{I} + \mathbf{a} \otimes \mathbf{b})$ from left and right to obtain $\mathbf{I}$. $\square$

2.2.2 Coherent solid interfaces

Two adjacent variants of a solid phase that have evolved by transformation from a single parent phase might be able to form a coherent interface through transformation. Such a configuration is called a *twin* and the two phases are said to be *twin-compatible*. A solid interface is considered to be coherent when the displacement field suffers no discontinuity across the interface. This in turn requires that any area element of the interface is deformed equally in either of the two adjacent phases, which means

$$\text{Cof } \mathbf{F}_1 \cdot \hat{\mathbf{N}} = \text{Cof } \mathbf{F}_2 \cdot \hat{\mathbf{N}}, \quad (2.15)$$

where $\mathbf{F}_1$ and $\mathbf{F}_2$ are the deformation gradients of the two neighboring phases with respect to the common parent phase, and $\hat{\mathbf{N}}$ is the unit normal vector to the interface in the reference configuration, i.e. the parent phase. The above condition is usually restated in a more accessible form, as

$$\mathbf{F}_2 = \mathbf{F}_1 + \mathbf{a} \otimes \hat{\mathbf{N}}, \quad (2.16)$$

where $\mathbf{a}$ is a vector whose obvious physical interpretation is the jump in deformation gradient across the interface, that is,

$$\mathbf{a} = \llbracket \mathbf{F} \rrbracket_{\hat{\mathbf{N}}} = (\mathbf{F}_2 - \mathbf{F}_1) \cdot \hat{\mathbf{N}}. \quad (2.17)$$

It is crucial to keep in mind that the coherence condition does not bear information on the orientation of the parent phase. This has pros and cons, depending on the context. We will return to this topic shortly.

2.3 Basic crystallography

In this work we have adopted the conventions and notations of Hermann's textbook [72] for the most part, and also partially those of Bhattacharya's [25].

A crystal *lattice*, also known as *Bravais lattice*, is an infinite periodic array of *lattice points* (atoms), designated by $\mathcal{L}(\mathbf{a}_i, \mathbf{r}_j)$, which can be expressed as

$$\mathbf{r} = \mathbf{r}_j + l^i \mathbf{a}_i, \quad (2.18)$$

with $\mathbf{a}_i, \mathbf{r}_j \in \mathbb{R}^3$, $i \in \{1, 2, 3\}$, $j \in \{1, \dots, p\}$, and $l^i \in \mathbb{Z}$. The $\mathbf{a}_i$'s are the *lattice vectors* and $\mathbf{r}_j$'s form the *atom basis*, such that $\mathbf{r}_j = x_j^i \mathbf{a}_i$, $x_j^i \in \mathbb{R}$ and $|x_j^i| < 1$. The magnitude of the lattice vectors are called lattice parameters, or lattice constants and are shown by

$\{\ell_1, \ell_2, \ell_3\}$, i.e. $\ell_i = \|\mathbf{a}_i\|$. In the case of monatomic crystals the atom basis has only one element which can be taken zero by a proper choice of lattice vectors, so each lattice point can be represented as $\mathbf{r} = l^i \mathbf{a}_i$. Accordingly, a *unit cell* is defined as the subdomain $\mathfrak{A} \subset \mathcal{L}$ whose repetition covers the entire lattice and at least includes one lattice point. The most apparent choice of a unit cell would be $\mathfrak{A} \equiv \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$, whose volume equals

$$V_{\mathfrak{A}} = |\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)|. \quad (2.19)$$

The choice of unit cell is obviously not unique. Therefore, the so called *primitive unit cell*, $\mathfrak{A}_o$, is defined such that its volume is the minimum of all possible unit cells. A more illuminating definition of a primitive unit cell is given as follows:

$$\mathfrak{A}_o = \left\{ \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\} \mid \nexists \mathbf{r} \in \mathcal{L}(\mathbf{a}_i) : \mathbf{r} = x^i \mathbf{a}_i, x^i \in [0, 1) \right\}. \quad (2.20)$$

All these definitions are trivially generalized to lattices in $\mathbb{R}^2$. For instance, a primitive unit cell in a two-dimensional lattice has the minimum area, i.e. $\mathfrak{A}_o = \arg \min_{\mathfrak{A}} |\mathbf{a}_1 \times \mathbf{a}_2|$.

A *lattice direction* or lattice edge can be represented by a vector which can be decomposed onto the lattice vectors: $\mathbf{e} = l^1 \mathbf{a}_1 + l^2 \mathbf{a}_2 + l^3 \mathbf{a}_3$. The l^1, l^2, l^3 are known as the *Miller indices*, which are collectively denoted by $[l^1 l^2 l^3]$ to show the lattice direction. A *lattice plane*, also called a crystallographic plane or a netplane, is a plane passing through at least three non-coplanar lattice points. A stacking of lattice planes generates the original lattice. Thus, as soon as the atom basis of the lattice plane is identified, it is considered a crystal *monolayer*. A lattice plane is determined by its normal vector, $\hat{\mathbf{n}}$, which does not necessarily possess its simplest representation in the same lattice coordinates as the crystal itself. Therefore, it may be adapted via a lattice transformation $\mathbf{T} \in \mathbb{Z}^{3 \times 3}$. The *alternate lattice*, $\mathcal{L}'(\mathbf{a}'_i, \mathbf{r}'_j)$, is defined such that $\mathbf{a}'_i = \mathbf{T} \mathbf{a}_i$ and $\mathbf{r}'_j = \mathbf{T} \mathbf{r}_j$, where $\mathbf{T} \in \mathbb{Z}^{3 \times 3}$, $\det \mathbf{T} = 1$. The transformation $\mathbf{T}$ is then an apparent symmetry of the lattice. In general, a symmetry transformation of a lattice belongs to $\mathbb{R}^{3 \times 3}$. Suppose that two lattice directions $\mathbf{e}_1$ and $\mathbf{e}_2$, parallel to the lattice plane, are known. Then

$$\hat{\mathbf{n}} = \alpha(\mathbf{e}_1 \times \mathbf{e}_2) = \alpha T_1^i \mathbf{a}_i \times T_2^j \mathbf{a}_j = \alpha \{l_1 (\mathbf{a}_2 \times \mathbf{a}_3) + l_2 (\mathbf{a}_3 \times \mathbf{a}_1) + l_3 (\mathbf{a}_1 \times \mathbf{a}_2)\}. \quad (2.21)$$

where $\alpha = |\mathbf{a}_1 \times \mathbf{a}_2|^{-1}$, is a normalizing factor, and $l_1 = T_1^2 T_2^3 - T_1^3 T_2^2$, and so forth. The l_i 's are Miller indices of the lattice plane, denoted by $(l_1 l_2 l_3)$, expressed in the so called *reciprocal basis*. The reciprocal basis, indicated as $\{\mathbf{a}^1, \mathbf{a}^2, \mathbf{a}^3\}$, is defined such that

$$\mathbf{a}^i = \epsilon^{ijk} \mathbf{a}_j \times \mathbf{a}_k / V_{\mathfrak{A}}. \quad (2.22)$$

It should be readily clear that $\mathbf{a}_i \cdot \mathbf{a}^j = \delta_i^j$.

We itemize some useful assertions without proof. It should not take much effort to verify them though.

- Every reciprocal lattice point $[l_1 l_2 l_3]$ corresponds to a lattice plane whose unit normal is $\hat{\mathbf{n}} = (l_1 \mathbf{a}^1 + l_2 \mathbf{a}^2 + l_3 \mathbf{a}^3) / |l_1 \mathbf{a}^1 + l_2 \mathbf{a}^2 + l_3 \mathbf{a}^3|$.
- The reciprocal of reciprocal lattice is the lattice itself,

- The lattice plane marked by Miller indices $(l_1 l_2 l_3)$ corresponds to a plane in the lattice coordinates that passes through the points $\mathbf{a}_1/l_1$, $\mathbf{a}_2/l_2$, and $\mathbf{a}_3/l_3$. In a cubic lattice, where the lattice and its reciprocal coincide, this comes naturally. But for a lattice cell chosen arbitrarily, the reciprocal coordinates simplify the geometrical representation.
- The distance between two adjacent lattice planes equals

$$d_{(l_1 l_2 l_3)} = (l_1^2 + l_2^2 + l_3^2)^{-1/2}. \quad (2.23)$$

- The density of atoms on a monolayer equals

$$\varrho_{(l_1 l_2 l_3)} = \frac{n_{(l_1 l_2 l_3)}}{|\mathbf{e}_1 \times \mathbf{e}_2|} = n_{(l_1 l_2 l_3)} d_{(l_1 l_2 l_3)} / V_{\mathfrak{A}}. \quad (2.24)$$

2.3.1 Lattice symmetries

The symmetries of a Bravais lattice can be classified as translational and point symmetries, denoted by $\mathcal{R}(\mathbf{a}_i)$ and $\mathcal{G}(\mathbf{a}_i)$ respectively. Point symmetries have three subclasses, namely

- shears,
- reflections,
- and rotations,

that map the lattice to itself. Among these three types of transformation, we confine our attention to rotations only, and here is why. Reflections, although mathematically interesting, do not relate to any realistic deformation map. We therefore neglect them. Furthermore, the shears in $\mathcal{G}(\mathbf{a}_i)$ correspond to plastic deformations driven by *dislocation* and *slip* in the lattice, which violate the Cauchy–Born hypothesis. We know, on the other hand, that there is no plastic deformation involved in the martensitic transformation, which renders slips and dislocations irrelevant. Therefore, we only consider a subset of the lattice symmetry group which we call the *point group*, denoted by $\mathcal{G}_+(\mathbf{a}_i)$, that excludes shears and reflections from $\mathcal{G}(\mathbf{a}_i)$.

Two lattice directions or planes are *crystallographically equivalent* if there exists a transformation $\mathbf{R} \in \mathcal{G}_+(\mathbf{a}_i)$ that maps one direction to the other, or one plane to the other, i.e. $\mathbf{R} \cdot \mathbf{e} = \mathbf{e}'$ or $\mathbf{R} \cdot \hat{\mathbf{n}} = \hat{\mathbf{n}}'$. Under the action of the lattice point group $\mathcal{G}_+(\mathbf{a}_i)$, the equivalence class of all lattice directions corresponding to $[l^1 l^2 l^3]$ is denoted by $\langle l^1 l^2 l^3 \rangle$, and the equivalence class of all lattice planes corresponding to $(l_1 l_2 l_3)$ are denoted by $\{l_1 l_2 l_3\}$.

Bravais lattices can be classified based on their distinctive symmetry properties. There are five Bravais lattices in two dimensions: oblique, rectangular, centered rectangular, hexagonal, and square. In three dimension, there are fourteen Bravais lattices: cubic, body-centered cubic, face-centered cubic, tetragonal, body-centered tetragonal, orthorhombic, base-centered orthorhombic, body-centered orthorhombic, face-centered orthorhombic, monoclinic, base-centered monoclinic, triclinic, rhombohedral, and hexagonal. For illustrations and symmetry groups see [67].

2.3.2 Types of martensitic transformation

Identifying the lattice structure of the austenite (before transformation) and the martensite (after transformation) is not complicated for crystallographers. Nonetheless, determining the *lattice correspondence* between the two phases is not an easy task. Lattice correspondence means knowing exactly which crystallographic directions and planes in the austenite relate to their counterparts in the martensite after transformation. For this reason, the choice of a unit cell is just short of an art, because the unit cells that are chosen to describe a martensitic transformation are not necessarily the simplest ones when austenite and martensite are to be described irrespective of the transformation. We will explain this issue in the next chapter.

Here, we only give a short list of different types of martensitic transformations. For detailed illustrations and derivations refer to [129, 25]. The four most common types of martensitic transformation are the following:

- cubic $\rightarrow$ tetragonal,
- cubic $\rightarrow$ orthorhombic,
- cubic $\rightarrow$ monoclinic-I along $\langle 110 \rangle$ (NiTi belongs to this type),
- cubic $\rightarrow$ monoclinic-II along $\langle 100 \rangle$,

and less common types are

- cubic $\rightarrow$ rhombohedral/trigonal (NiTi R-phase),
- tetragonal $\rightarrow$ orthorhombic,
- tetragonal $\rightarrow$ monoclinic.

We will later study the martensitic transformation of NiTi (49.75 at. % Ni) and will derive the transformation matrices and their polar decompositions.

2.3.3 Cauchy–Born hypothesis: lattice-continuum link

The *Cauchy–Born hypothesis* states that the solid crystal lattice deforms in accordance with the corresponding continuum deformations. That means for a monatomic lattice $\mathcal{L}(\mathbf{a}_i)$ under the deformation $\mathbf{x} \rightarrow \mathbf{X}$:

$$\mathbf{F} = \frac{\partial \mathbf{x}}{\partial \mathbf{X}} \quad \Leftrightarrow \quad \mathbf{a}_i(\mathbf{x}) = \mathbf{F}(\mathbf{X}) \mathbf{a}_i(\mathbf{X}) . \quad (2.25)$$

This hypothesis results in a fundamental connection between lattice and continuum frameworks; namely, for any thermodynamic state variable or potential ϕ of the continuum and its lattice counterpart ϕ_* we have

$$\phi(\mathbf{F}, \theta) = \phi_*(\mathbf{F}\mathbf{a}_i(\mathbf{X}), \theta) = \phi_*(\mathbf{a}_i(\mathbf{x}), \theta) . \quad (2.26)$$

In the case of multi-lattices, i.e. crystals consisting of two or more elements, with a crystal system $\mathcal{L}(\mathbf{a}_i, \mathbf{r}_j)$ the above hypothesis has to be modified to accommodate the *lattice shifts*

during deformation. Under the assumption of equilibrium, this is accomplished by finding the configuration of the atom basis that minimizes the free energy, viz

$$\psi_*(\mathbf{a}_i(\mathbf{x}), \theta) = \min_{\mathbf{r}_j} \psi_*(\mathbf{a}_i, \mathbf{r}_j, \theta) . \quad (2.27)$$

Then for other thermodynamic variables or potentials, say ϕ_* , we have $\phi_*(\mathbf{a}_i(\mathbf{x}), \theta) = \phi_*(\mathbf{a}_i, \bar{\mathbf{r}}_j, \theta)$, where $\bar{\mathbf{r}}_j = \arg \min_{\mathbf{r}_j} \psi_*(\mathbf{a}_i, \mathbf{r}_j, \theta)$. Therefore, the lattice–continuum link remains valid for multi-lattices via the modified version of the Cauchy–Born hypothesis.

2.4 Helmholtz free energy and its symmetries

Let's denote the Helmholtz free energy of the lattice by $\psi_*(\mathbf{a}_i, \theta)$, and its continuum counterpart by $\psi(\mathbf{F}, \theta)$. Thermal expansions are negligible compared to distortions due to martensitic transformation at and around the transformation temperature. Also since the martensitic transformation is diffusionless, it can be fully described by a deformation gradient tensor, denoted by $\mathbf{F}_i$, as the identifying quantity that together with its associated transformation temperature θ_{am} can fully characterize the martensitic transformation. In continuum mechanical formulations we sometimes use transformation strains instead of transformation deformation gradients, that is, $\boldsymbol{\eta} = \text{sym}(\mathbf{F} - \mathbf{I})$.

The ground state pertains to the undeformed lattice $\psi_*(\mathbf{a}_i(\mathbf{X}), \theta)$, which we assume zero without loss of generality. The Cauchy–Born hypothesis resulted in the fundamental connection (2.26) between lattice and continuum frameworks which can be applied to the free energy, as

$$\psi(\mathbf{F}, \theta) = \psi_*(\mathbf{F}\mathbf{a}_i(\mathbf{X}), \theta) = \psi_*(\mathbf{a}_i(\mathbf{x}), \theta) . \quad (2.28)$$

2.4.1 Frame indifference (objectivity)

Since the free energy is a scalar quantity, it must be invariant under rigid rotations of the reference frame, thus

$$\psi_*(\mathbf{R}\mathbf{a}_i(\mathbf{x}), \theta) = \psi_*(\mathbf{a}_i(\mathbf{x}), \theta) , \quad \forall \mathbf{R} \in \text{SO}(3) , \quad (2.29)$$

which based on (2.28) gives

$$\psi(\mathbf{R}\mathbf{F}, \theta) = \psi(\mathbf{F}, \theta) , \quad \forall \mathbf{R} \in \text{SO}(3) . \quad (2.30)$$

$\text{SO}(3)$ is the set of proper orthogonal 3×3 matrices. The above condition is often called the *principle of objectivity*, which is one of the fundamental requirements that every well-formulated continuum thermomechanical model has to fulfill.

2.4.2 Material symmetry

As we mentioned, every Bravais lattice $\{\mathbf{a}_i\}$ is invariant under the action of its point group $\mathcal{G}_+(\mathbf{a}_i)$, therefore

$$\forall \mathbf{R} \in \mathcal{G}_+(\mathbf{a}_i) : \mathbf{R}\mathbf{a}_i = \mathbf{a}_i \quad \Rightarrow \quad \psi_*(\mathbf{R}\mathbf{a}_i, \theta) = \psi_*(\mathbf{a}_i, \theta) , \quad (2.31)$$

which, according to (2.28), yields

$$\psi(\mathbf{FR}, \theta) = \psi(\mathbf{F}, \theta) , \quad \forall \mathbf{R} \in \mathcal{G}_+(\mathbf{a}_i) . \quad (2.32)$$

It is often more convenient to use an alternative form of this formulation which simultaneously borrows from the objectivity, viz

$$\psi(\mathbf{R}^T \mathbf{FR}, \theta) = \psi(\mathbf{F}, \theta) , \quad \forall \mathbf{R} \in \mathcal{G}_+(\mathbf{a}_i) . \quad (2.33)$$

2.5 Symmetries of martensitic transformation

Since the martensitic transformation is symmetry breaking and reversible, it can be concluded that the point group of martensite is a subgroup of the point group of austenite, i.e. $\mathcal{G}_+(\mathbf{a}_i^m) \leq \mathcal{G}_+(\mathbf{a}_i^a)$. This can be proven by contradiction, because otherwise the symmetry group of the austenite will not be finite (see [129, pp. 42]), that is to say, there will be infinitely many variants of martensite. Then, according to Lagrange theorem on cardinality of subgroups⁷, we obtain the number of **distinct** martensite variants

$$N_{\text{m.v.}} = \frac{\#\mathcal{G}_+(\mathbf{a}_i^a)}{\#\mathcal{G}_+(\mathbf{a}_i^m)} . \quad (2.34)$$

It then follows that the transformation stretches, called *Bain matrices*, are symmetry-connected through the point group of the parent phase via

$$\mathbf{U}_j = \mathbf{R}^T \mathbf{U}_i \mathbf{R} , \quad \exists \mathbf{R} \in \mathcal{G}_+(\mathbf{a}_i^a) . \quad (2.35)$$

Also, from material symmetry and objectivity, it is concluded that the *energy wells* corresponding to distinct martensite variants have equal minima at any given temperature [126, 128]. Furthermore, at the transformation temperature, θ_{am} , the austenite phase and the martensite variants shall be at equilibrium. From *Clapeyron equation* for phase equilibria we can conclude that the austenite and all martensite variants have the same free energy minima. This can be verified easily by starting from free energy of the lattice and going to continuum framework by Cauchy–Born hypothesis [25, pp. 58–60].

Since the free energy is objective, according to the polar decomposition theorem, 1, we can always identify each martensite energy well by its symmetric positive definite stretch $\mathbf{U}_i$. This extends the possibility of forming twins between martensite variants, through the coherence condition (2.16), to

$$\mathbf{R}_2 \mathbf{Q}_2 \mathbf{U}_2 - \mathbf{R}_1 \mathbf{Q}_1 \mathbf{U}_1 = \mathbf{a} \otimes \hat{\mathbf{N}} , \quad (2.36)$$

where $\mathbf{R}_i \in \text{SO}(3)$ and $\mathbf{F}_i = \mathbf{Q}_i \mathbf{U}_i$ is the right polar decomposition of the transformation $\mathbf{F}_i$. This can be rewritten in the more compact form

$$\mathbf{Q} \mathbf{U}_2 - \mathbf{U}_1 = \mathbf{b} \otimes \hat{\mathbf{N}} , \quad (2.37)$$

⁷Lagrange theorem asserts that the order (cardinality) of a finite group is an integral multiple of the order of each of its subgroups.

in which $\mathbf{Q} = \mathbf{Q}_1^T \mathbf{R} \mathbf{Q}_2$, $\mathbf{R} \in \text{SO}(3)$, and $\mathbf{b} = \mathbf{Q}_1^T \mathbf{a}$. The latter form of the coherence condition is most widely used for its algebraic simplicity. One has to take great care when transforming back and forth between alternative forms of (2.37), (2.36), and (2.16) as needed.

Remark 2. In the theory of self-accommodating microstructures, two assumptions are usually made which we leave out. One is that only *isochoric* martensitic transformations are capable of forming *self-accommodating* microstructures. The second assumption is that there is a particular set of average deformation gradients (mesoscopic) whom the microstructure is able to accommodate [25, pp. 155]. We pose none of these restrictions on our kinematic ansatz and stipulate an alternative mechanism for accommodation.

2.5.1 Ericksen–Pitteri neighborhood

For the continuum theory of martensitic phase transformation to be non-degenerate, it is postulated that the Bain matrices remain within the so called *Ericksen–Pitteri neighborhood* (EPN), which is an open bounded neighborhood of the parent phase [47]. A direct consequence of this notion is that the Ericksen–Pitteri neighborhood of the lattice $\mathcal{L}(\mathbf{a}_i^a)$, denoted by $\mathcal{N}(\mathbf{a}_i^a)$, is invariant under the action of the lattice point group $\mathcal{G}_+(\mathbf{a}_i^a)$. More generally, and trivially so, it is also invariant with respect to $\text{SO}(3)$. Within an EPN, a martensitic transformation is limited to those in which the symmetry group of the child lattice is the subgroup of the symmetry group of the parent lattice. However, there exist martensitic transformations for which this is not the case. These transformations are, nevertheless, excluded from the study of shape-memory phenomena because they bring about irreversible changes in the lattice.

2.5.2 Mechanical twins vs growth twins

We mentioned that the coherence condition carries no information about the parent phase orientation. Especially the form of *twinning equation* (2.37) accepts compatible variants pertaining to energy wells up to a rotation. In certain situations this might become problematic, and here is why. Depending on their formation mechanism the martensitic twins are classified as [129, pp. 259]

- *mechanical twins*, which are formed by transformation from austenite via mechanical or thermal loading,
- and *growth twins*, which are formed via crystal growth.

The twinning equation implicitly assumes a reversible transition from a reference configuration, i.e. the parent phase, which physically does not apply to growth twins. Further, for mechanical twins the orientation of a parent phase has to be known in order for the coherence condition to be meaningfully expressed.

The majority of the available models for martensitic microstructure do not address this issue. Because they often involve a homogenization procedure based on a randomized RVE orientation which makes a precise choice of the parent phase orientation superfluous. Here

we do not have this luxury, because our model shall include precise representation of twin interfaces and accommodation strains which can only be realized via twinning equation based on a fixed orientation.

One might argue that because the transformations are within the EPN, the parent phase can undergo an arbitrary rotation without violating the basic assumptions. However, according to minimization theorems in elasticity the rotation must be homogeneous throughout the entire domain [156, 82], which violates the mesoscopic boundary conditions. Therefore, we have to make additional assumptions on the orientation of the parent phase explicitly, which we do in the next chapter.

Chapter 3

Continuum micromechanical model

We intend to develop a thermomechanically consistent continuum model for the martensitic laminate microstructure which *almost* strictly conforms to the coherence condition and evolves quasi-statically. The laminate is assumed to be made of *deformation twins* and not growth twins; which can either be

- a mixture of two martensitic variants⁸ sharing a common austenitic parent, or
- a mixture of martensitic phase with its parent austenitic phase,

as long as the coherence condition (2.16) between the two phases or variants is fulfilled.

The second case, i.e. mixing of austenite and martensite, is usually realized through the mixing of austenite with a phase which itself is a mixture of two martensitic variants whose average deformation gradient is twin-compatible with the austenite phase. Such a configuration is usually termed a *second-order laminate* or a second-order mixing of phases, which we exclude from this work. So, in short, a first-order laminate composed of deformation twins is assumed. Notice that assuming a quasi-static evolution of the microstructure dismisses any influence of dynamic constraints.

The external loading and constraints, originating from the mesoscopic state of the material, can be collectively represented either as a homogeneous deformation or as a homogeneous loading imposed on (prescribed at) the boundary of the microstructure. We adopt the first approach here, i.e. a displacement-controlled solution at the micro level, which we will explain later. This homogeneous boundary deformation does not generally coincide with the volume-averaged deformation of the phase mixtures, if the martensitic laminate remains in its *eigen configuration*. Following the customary terminology of the homogenization methods (see [114]) we refer to the unstrained state of martensite as the eigen state or eigen configuration, and to the respective deformation as eigen deformation. In Fig. 1.1 the eigen strains mark the states at which the martensite attains its minimum Helmholtz energy.

The martensitic microstructure can conceivably accommodate this homogenous boundary deformation through two mechanisms.

⁸Here we adopt the terminology popular among material scientists, in which variants are distinguished by their difference in symmetries with respect to a common parent phase, whereas phases bear their usual thermodynamical interpretation.

1. The mesoscopic homogenous deformation at the boundary can be satisfied via a kinematic construct called *interpolation layer*. This approach is particularly of a rigorous mathematical makeup and is widely used in continuum theory of martensitic transformation [25, pp 94–96].
2. An alternative approach, which we employ, allows the laminate to deviate from its eigen state and to respond to the mesoscopically induced boundary condition via deformation. We believe that this method is more capable in reflecting the morphological attributes of the laminate microstructure especially the actual accommodation patterns which are empirically observed.

Having decided for the second approach, wherever the laminate meets its boundary in parallel the boundary condition is automatically fulfilled. On the other hand, at the boundary points where this may not be the case the accommodation takes place either by

- *branching*, which changes the lath width, or by
- *needles*, which alters the volume fraction of the constituent phases.

We will show, as is readily well established, that these variations make for a more energetically favorable microstructure. At these parts of the boundary, nevertheless, there persists a mismatch between the laminate and its mesoscopic surroundings which we account for by employing the first approach, i.e. by constructing a boundary layer.

3.1 Kinematics of laminate microstructure

Let the material coordinates pertain exclusively to the parent (austenite) phase and be denoted by $\mathbf{X}$, and let the spatial coordinates describe the laminate microstructure in its fully accommodated state and be denoted by $\mathbf{x}$. The respective deformation gradient is then $\mathbf{F} = \partial\mathbf{x}/\partial\mathbf{X}$. Also, let the eigen configuration be addressed by its coordinates $\mathbf{x}^\circ$, and its respective deformation gradient $\mathbf{F}^\circ = \partial\mathbf{x}^\circ/\partial\mathbf{X}$.

The microscale domain; that is the domain in which the microstructure resides; which encloses the mesoscopic *material point* $\underline{\mathbf{X}}$, is designated as $\omega(\underline{\mathbf{X}})$. The prescribed homogeneous deformation on the microdomain requires that the average deformation gradient of the microstructure, which is given by

$$\langle \mathbf{F} \rangle_\omega = \frac{1}{|\omega(\underline{\mathbf{X}})|} \int_{\omega(\underline{\mathbf{X}})} \mathbf{F} \, dV, \quad (3.1)$$

be equal to the mesoscopic deformation gradient at the material point $\underline{\mathbf{X}}$. From a material modeling viewpoint, $\langle \mathbf{F} \rangle_\omega$ connects the microscopic length scales and the mesoscopic length scales. We might leave out the subscript ω in $\langle \bullet \rangle_\omega$ wherever the meaning is clear from the context. We may also adopt a shorter notation $\bar{\mathbf{F}} = \langle \mathbf{F} \rangle_\omega$. Accordingly, we introduce the intermediate configurations $\bar{\mathbf{x}}$ and $\bar{\mathbf{x}}^\circ$ which express the average laminate deformation in the accommodated state and in the eigen state respectively.

As mentioned before, the microstructure cannot fulfill all the conceivable mesoscopic deformations merely by microscopic mixing of phases via lamination. This means that

$$\exists \underline{\mathbf{F}} : \langle \mathbf{F}^\circ \rangle_\omega \neq \underline{\mathbf{F}}, \quad (3.2)$$

where $\underline{\mathbf{F}} = \partial \underline{\mathbf{x}} / \partial \underline{\mathbf{X}}$, is the mesoscopic deformation gradient in which the mesoscopic material coordinates are denoted by $\underline{\mathbf{X}}$ and the mesoscopic spatial coordinates by $\underline{\mathbf{x}}$. This is why we tell apart the eigen configuration $\mathbf{x}^\circ$ and the accommodated configuration $\mathbf{x}$. Nevertheless, the average accommodated deformation gradient is supposed to match the mesoscopic deformation gradient, which is perceived as a homogeneous deformation at the microscale, in order to guarantee *scale separation*.

Remark 3. THE MICRO-BOUNDARY CONDITIONS. Equation (3.2) tells us that in most actual situations the eigen configuration is hardly realized due to the constraints imposed on the microstructure by its surroundings. In order to model these constraints, we have two choices:

- *Load driven* approach in which the mesoscopic effects are imposed on the microstructure via traction boundary conditions.
- *Deformation driven* approach in which the influence of mesoscopic constraints on the microstructure is simulated via a homogeneous deformation imposed on its boundary.

We opt for the second method. For the microstructural deformation map $\mathbf{x}(\mathbf{X})$, the average micro boundary condition in the form of

$$\mathbf{x}(\mathbf{X}) = \underline{\mathbf{F}} \cdot \mathbf{X}, \quad \mathbf{X} \in \partial\omega$$

will be *almost strictly* enforced. What we mean by this is that we do our best to fulfill this condition by choosing a suitable kinematic ansatz for the microstructure. There will be parts of the boundary, however, where a mismatch between our ansatz and the homogeneous mesoscopic deformation is inevitable; no matter how well-formulated our ansatz is. At these points the above condition will be replaced by a carefully formulated boundary energy ansatz which penalizes this mismatch:

$$\psi_\gamma(\mathbf{x}(\mathbf{X}) - \underline{\mathbf{F}} \cdot \mathbf{X}) \rightarrow \min, \quad \mathbf{X} \in \partial\omega. \quad (3.3)$$

The minimization is usually carried out in a variational sense. The microscopic deformed configuration $\mathbf{x}(\mathbf{X})$ is obtained by integrating the microscopic ansatz for deformation gradient

$$\mathbf{x}(\mathbf{X}) = \int_{\mathbf{x}_0}^{\mathbf{X}} \underline{\mathbf{F}} \cdot d\mathbf{X}, \quad \forall \mathbf{X} \in \omega.$$

3.1.1 Twinning equation

In order for the laminate to fulfill the coherence condition, we naturally begin with the twinning equation (2.16) which expresses this condition for two neighboring solid phases:

$$\mathbf{F}_2 - \mathbf{F}_1 = \mathbf{a} \otimes \hat{\mathbf{N}}, \quad (2.16)$$

where the orientation of the twin interface is customarily determined by a unit normal vector $\hat{\mathbf{N}}$, expressed in material coordinates. The two phases are distinguished via their transformations, $\mathbf{F}_1$ and $\mathbf{F}_2$, with respect to their parent phase. The vector $\mathbf{a}$ denotes the jump in deformation gradient across the interface, viz

$$\mathbf{a} = \llbracket \mathbf{F} \rrbracket_{\hat{\mathbf{N}}} = (\mathbf{F}_2 - \mathbf{F}_1) \cdot \hat{\mathbf{N}}. \quad (2.17)$$

It is important to remember that $\mathbf{F} = \mathbf{I}$ expresses the undeformed state corresponding to the parent phase.

Since deformation gradient is related to displacement gradient by $\mathbf{F} = \mathbf{u} \nabla_{\mathbf{X}} + \mathbf{I}$, the twinning equation can be reformulated as

$$\mathbf{u}_2 \nabla_{\mathbf{X}} - \mathbf{u}_1 \nabla_{\mathbf{X}} = \mathbf{a} \otimes \hat{\mathbf{N}}. \quad (3.4)$$

Notice that $\hat{\mathbf{N}}$ and $\mathbf{a}$ retain their meaning in (2.17), namely $\mathbf{a} = \llbracket \mathbf{u} \nabla_{\mathbf{X}} \rrbracket \cdot \hat{\mathbf{N}} = \llbracket \mathbf{F} \rrbracket \cdot \hat{\mathbf{N}}$, and that $\nabla_{\mathbf{X}}$ expresses the material (not the spatial) derivative.

3.1.2 Periodic structure of the laminate and its geometry

The laminate orientation is determined by its unit normal vector $\hat{\mathbf{N}}$ in the reference configuration. Note that the actual normal to the twin interfaces may deviate from $\hat{\mathbf{N}}$ as the microstructure accommodates external loads and constraints; however, these patterns occur in such a way that the overall direction of the lamellae remains unaltered (Fig. 3.1). In this sense, $\hat{\mathbf{N}}$ is perpendicular to the laminate *in average*. We will elaborate on plausibility of this assumption later.

The main characteristic of such a microstructure is its periodic deformation gradient in X_3 as per

$$\mathbf{F}(\mathbf{X}) = \begin{cases} \mathbf{F}_1(\mathbf{X}_{\parallel}) , & -\Lambda H \leq X_{\perp} \leq 0 , \\ \mathbf{F}_2(\mathbf{X}_{\parallel}) , & 0 \leq X_{\perp} \leq (1 - \Lambda) H , \end{cases} \quad (3.5)$$

The period H is the width of a single twin in the undeformed configuration, which we often call *lath width*, and $\Lambda \in [0, 1]$ denotes the fraction of phases⁹ along $\hat{\mathbf{N}}$. The coordinates are naturally decomposed into parallel to the laminate $\mathbf{X}_{\parallel} = \mathbf{X} - (\mathbf{X} \cdot \hat{\mathbf{N}})\hat{\mathbf{N}}$, and normal to the laminate $X_{\perp} = \mathbf{X} \cdot \hat{\mathbf{N}}$, or equivalently $\mathbf{X}_{\perp} = (\mathbf{X} \cdot \hat{\mathbf{N}})\hat{\mathbf{N}}$. We shall keep in mind that our kinematic ansatz allows H and Λ to vary along the laminate; that is

$$H = H(\mathbf{X}_{\parallel}) , \quad \Lambda = \Lambda(\mathbf{X}_{\parallel}) . \quad (3.6)$$

In equations (3.5) and (3.6), all the kinematic quantities are allowed to vary parallel to the laminate, $\mathbf{X}_{\parallel}$, except the normal vector $\hat{\mathbf{N}}$. Variations of Λ and H clearly change the pointwise orientation of twin interfaces. This obviously does not comply with a fixed

⁹In homogenization or energy relaxation methods Λ is usually identified as volume fraction; since there they merely consider the average compatibility conditions (A.23) rather than a strict pointwise compatibility.

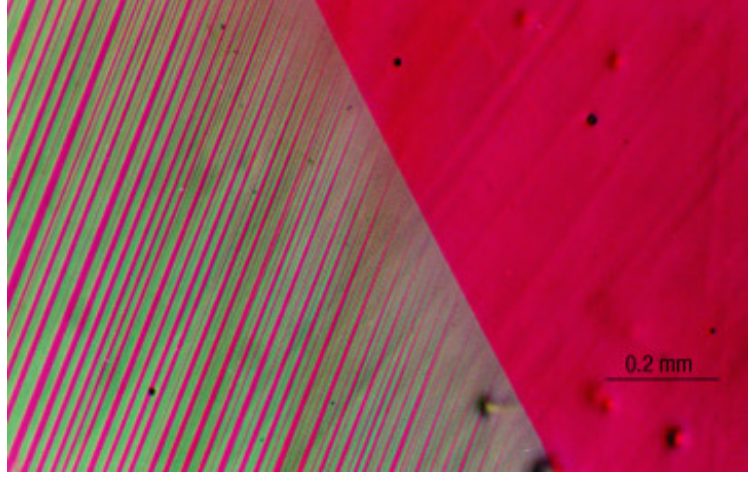


Fig. 3.1: Periodicity of laminate in the needle microstructure [84, © permission obtained].

$\hat{\mathbf{N}}$. On other hand, the geometry of the laminate allows for designation of an *average orientation*. Let's say $\hat{\mathbf{N}}$ represents this average orientation (Fig. 3.1). Now we introduce the unit normal vector $\hat{\mathbf{B}}$ which deviates locally from $\hat{\mathbf{N}}$ ever so slightly, so it can stay normal to the twin interface: $\hat{\mathbf{B}} = \hat{\mathbf{B}}(\mathbf{X}_{\parallel})$. Then, the compatibility condition has to be rewritten as $\mathbf{F}_2 - \mathbf{F}_1 = \mathbf{a} \otimes \hat{\mathbf{B}}$. However, the laminate is still periodic in $\hat{\mathbf{N}}$ direction. Eventually, this will correct our formulation up to a scalar factor of $\mathbf{a}$, which is mathematically superfluous. Therefore, we let go of $\hat{\mathbf{B}}$ entirely.

Since the laminate is periodic along $\hat{\mathbf{N}}$, and since the changes in this direction are much faster than the changes parallel to the laminate, we first deploy an averaging in the normal direction in order to capture the relevant characteristics of the laminate while eliminating any dependence on $X_{\perp}$ in the energy formulation. Taking the average in $\hat{\mathbf{N}}$ direction, follows

$$\langle \mathbf{F} \rangle_{\perp} = \frac{1}{H} \int_{-\Lambda H}^{(1-\Lambda)H} \mathbf{F} \, dX_{\perp}. \quad (3.7)$$

The $\perp$ subscript can be dropped for brevity, but the meaning should be clear from the context. Applying this to the configuration (3.5) gives

$$\langle \mathbf{F} \rangle_{\perp} = \Lambda \mathbf{F}_1 + (1 - \Lambda) \mathbf{F}_2. \quad (3.8)$$

Now we substitute the twinning equation (2.16) into (3.8) to obtain

$$\begin{aligned} \langle \mathbf{F} \rangle_{\perp} &= \Lambda \mathbf{F}_1 + (1 - \Lambda) \mathbf{F}_2 = \Lambda \mathbf{F}_1 + (1 - \Lambda) \left(\mathbf{F}_1 + \mathbf{a} \otimes \hat{\mathbf{N}} \right) = \mathbf{F}_1 + (1 - \Lambda) \mathbf{a} \otimes \hat{\mathbf{N}} \\ &= \Lambda \mathbf{F}_1 + (1 - \Lambda) \mathbf{F}_2 = \Lambda \left(\mathbf{F}_2 - \mathbf{a} \otimes \hat{\mathbf{N}} \right) + (1 - \Lambda) \mathbf{F}_2 = \mathbf{F}_2 - \Lambda \mathbf{a} \otimes \hat{\mathbf{N}}. \end{aligned} \quad (3.9)$$

This indicates that $\mathbf{F}_1$, $\mathbf{F}_2$, and $\langle \mathbf{F} \rangle$ are all rank-one-connected along the same direction $\hat{\mathbf{N}}$, which is mathematically synonymous to

$$\text{Cof} \langle \mathbf{F} \rangle_{\perp} \hat{\mathbf{N}} = \text{Cof} \mathbf{F}_2 \hat{\mathbf{N}} = \text{Cof} \mathbf{F}_1 \hat{\mathbf{N}}. \quad (3.10)$$

Next, let us find a more specific formula of the form (3.5) by substituting (3.9) back into (3.5) to arrive at

$$\mathbf{F}(\mathbf{X}) = \langle \mathbf{F} \rangle_{\perp} + \begin{cases} -(1 - \Lambda) \mathbf{a} \otimes \hat{\mathbf{N}}, & -\Lambda H \leq X_{\perp} \leq 0, \\ \Lambda \mathbf{a} \otimes \hat{\mathbf{N}}, & 0 \leq X_{\perp} \leq (1 - \Lambda) H. \end{cases} \quad (3.11)$$

Equation (3.11) will be the very basis of our upcoming laminate formulation. Before moving forward, we need to illustrate some of its mathematical properties, most importantly the relation between the orientation and the width of the laminate in material and spatial coordinates. In other words: we need to know how the laminate forms with respect to its parent phase.

The orientation of the laminate is determined by its normal vector $\hat{\mathbf{N}}$. One can apply the *Nanson's formula* to track the changes of laminate orientation and interfacial stretches. In the average laminate coordinates $\bar{\mathbf{x}}$ and in the spatial coordinates $\mathbf{x}$, we have

$$d\mathbf{a} \hat{\mathbf{n}} = dA \text{Cof } \mathbf{F}_i \hat{\mathbf{N}}, \quad d\alpha \hat{\bar{\mathbf{n}}} = dA \text{Cof } \bar{\mathbf{F}} \hat{\mathbf{N}}, \quad (3.12)$$

where dA , da , and $d\alpha$ are the area elements in the reference configuration (austenite), current configuration (martensite), and average martensite coordinates respectively. The unit normal vectors in average configuration and in final configuration are shown by $\hat{\bar{\mathbf{n}}}$ and $\hat{\mathbf{n}}$. This latter result together with equation (3.10) shows that $da \hat{\mathbf{n}} = d\alpha \hat{\bar{\mathbf{n}}}$, therefore

$$d\alpha = da, \quad \hat{\bar{\mathbf{n}}} = \hat{\mathbf{n}}. \quad (3.13)$$

Explicit formulas for the orientation and the area elements are given based on (3.12) as

$$\hat{\mathbf{n}} = \frac{\text{Cof } \mathbf{F}_i \hat{\mathbf{N}}}{\|\text{Cof } \mathbf{F}_i \hat{\mathbf{N}}\|} = \frac{\mathbf{F}_i^{-T} \hat{\mathbf{N}}}{\|\mathbf{F}_i^{-T} \hat{\mathbf{N}}\|}, \quad \hat{\bar{\mathbf{n}}} = \frac{\text{Cof } \bar{\mathbf{F}} \hat{\mathbf{N}}}{\|\text{Cof } \bar{\mathbf{F}} \hat{\mathbf{N}}\|} = \frac{\bar{\mathbf{F}}^{-T} \hat{\mathbf{N}}}{\|\bar{\mathbf{F}}^{-T} \hat{\mathbf{N}}\|}, \quad (3.14)$$

$$da = \|\text{Cof } \mathbf{F}_i \hat{\mathbf{N}}\| dA, \quad d\alpha = \|\text{Cof } \bar{\mathbf{F}} \hat{\mathbf{N}}\| dA. \quad (3.15)$$

Here we have relied on the fact that $\text{Cof } \mathbf{A} = |\mathbf{A}| \mathbf{A}^{-T}$ for any invertible matrix $\mathbf{A}$.

Equation (3.13) indicates that the multiplicative decomposition of the deformation gradient into an average and a fluctuating part conceptually simplifies the formulation to a great extent, because the fluctuating part neither rotates nor stretches the twin interface. Thus, introducing the intermediate coordinates $\bar{\mathbf{x}}$ pertaining to the average deformation marks an explicit separation between two modes of deformation:

- one that stretches and rotates the austenite, and
- one that oscillates and forms martensitic variants and their interfaces without rotating the body in average or stretching it along the twin interface.

Notice that this leaves open the possibility that the fluctuating part of the deformation might stretch the laminate along the normal direction $\hat{\bar{\mathbf{n}}}$. So let us see how the laminate stretches in this direction. For this, starting from (3.13) we are able to track the changes in thickness of the constituting layers.

Suppose a cylindrical volume element whose axis is normal to the laminate with a height equal to the lamainte thickness and with a base which is an area element that we just investigated in (3.12). The volume of such a cylindrical element changes in proportion to the Jacobian determinant of the deformation, that is, $dv = \det \mathbf{F} dV$. Constructing this element across the two variants before and after deformation, we have

$$\underbrace{\lambda h da}_{dv_1} = \det \mathbf{F}_1 \underbrace{\Lambda H dA}_{dV_1}, \quad \underbrace{(1 - \lambda) h da}_{dv_2} = \det \mathbf{F}_2 \underbrace{(1 - \Lambda) H dA}_{dV_2}. \quad (3.16)$$

Using equation (3.15) and knowing that $\text{Cof } \mathbf{A} = |\mathbf{A}| \mathbf{A}^{-T}$, we eliminate the area elements from the two equalities to come up with a relation between the thicknesses as

$$\lambda h = \frac{\det \mathbf{F}_1 \Lambda H}{\|\text{Cof } \mathbf{F}_1 \mathbf{N}\|} = \frac{\Lambda H}{\|\mathbf{F}_1^{-T} \hat{\mathbf{N}}\|}, \quad (1 - \lambda) h = \frac{\det \mathbf{F}_2 (1 - \Lambda) H}{\|\text{Cof } \mathbf{F}_2 \mathbf{N}\|} = \frac{(1 - \Lambda) H}{\|\mathbf{F}_2^{-T} \hat{\mathbf{N}}\|}. \quad (3.17)$$

In the next step we utilize the following well-known lemma from linear algebra without giving its proof.

Lemma 2. *For any full-rank matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ and vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$,*

$$\det(\mathbf{A} + \mathbf{a} \otimes \mathbf{b}) = (1 + \mathbf{b} \cdot \mathbf{A}^{-1} \cdot \mathbf{a}) \det \mathbf{A}.$$

A relation between the Jacobian determinants of the two variants will enable us to find an explicit relation between the lath widths based on (3.17). So a slight rearrangement of equation (3.9) to $\mathbf{F}_1 = \bar{\mathbf{F}} - (1 - \Lambda) \mathbf{a} \otimes \hat{\mathbf{N}}$ and $\mathbf{F}_2 = \bar{\mathbf{F}} + \Lambda \mathbf{a} \otimes \hat{\mathbf{N}}$, applying lemma 2 twice, we come upon

$$\det \mathbf{F}_1 = \left(1 - (1 - \Lambda) \hat{\mathbf{N}} \cdot \bar{\mathbf{F}}^{-1} \cdot \mathbf{a}\right) \det \bar{\mathbf{F}}, \quad \det \mathbf{F}_2 = \left(1 + \Lambda \hat{\mathbf{N}} \cdot \bar{\mathbf{F}}^{-1} \cdot \mathbf{a}\right) \det \bar{\mathbf{F}}.$$

If the first equation is multiplied by Λ and the second by $(1 - \Lambda)$, and then the results are added together, we obtain

$$\det \bar{\mathbf{F}} = \Lambda \det \mathbf{F}_1 + (1 - \Lambda) \det \mathbf{F}_2. \quad (3.18)$$

Now we add the two equations in (3.17), and then use (3.10) to find a common denominator, and finally plug in (3.18) and simplify:

$$\begin{aligned} \lambda h + (1 - \lambda) h &= \frac{\det \mathbf{F}_1 \Lambda H}{\|\text{Cof } \mathbf{F}_1 \mathbf{N}\|} + \frac{\det \mathbf{F}_2 (1 - \Lambda) H}{\|\text{Cof } \mathbf{F}_2 \mathbf{N}\|} = \\ &= \frac{(\Lambda \det \mathbf{F}_1 + (1 - \Lambda) \det \mathbf{F}_2) H}{\|\text{Cof } \bar{\mathbf{F}} \mathbf{N}\|} = \frac{\det \bar{\mathbf{F}}}{\|\text{Cof } \bar{\mathbf{F}} \mathbf{N}\|} H = \frac{H}{\|\bar{\mathbf{F}}^{-T} \hat{\mathbf{N}}\|}, \end{aligned}$$

which shortens to

$$h = \frac{H}{\|\bar{\mathbf{F}}^{-T} \hat{\mathbf{N}}\|}. \quad (3.19)$$

This tells us that the fluctuating part does not alter the lath width for arbitrary twin-compatible deformations. Equation (3.17), on the other hand, shows that individual deformations $\mathbf{F}_1$ and $\mathbf{F}_2$ stretch the laminate along the normal direction $\hat{\mathbf{N}}$. This can only be realized by a change in the volume fractions, as the overall lath width must remain unchanged with respect to the average deformation. Substituting (3.19) into (3.17) will tell us how the volume fraction changes, as

$$\lambda = \frac{\|\bar{\mathbf{F}}^{-T} \hat{\mathbf{N}}\|}{\|\mathbf{F}_1^{-T} \hat{\mathbf{N}}\|} \Lambda = 1 - \frac{\|\bar{\mathbf{F}}^{-T} \hat{\mathbf{N}}\|}{\|\mathbf{F}_2^{-T} \hat{\mathbf{N}}\|} (1 - \Lambda) . \quad (3.20)$$

Decomposition of deformation

Since the deformation gradient of the laminate microstructure is periodic, it makes sense to decompose it into an average component and a fluctuating component in order to build the connection between the microscale and the mesoscale more transparently. We do this decomposition multiplicatively as

$$\mathbf{F} = \tilde{\mathbf{F}} \bar{\mathbf{F}} , \quad (3.21)$$

which implies a hypothetical two-step deformation scheme:

$$\mathbf{X} \xrightarrow{\bar{\mathbf{F}}} \bar{\mathbf{x}} \xrightarrow{\tilde{\mathbf{F}}} \mathbf{x} . \quad (3.22)$$

It is instructive to explicitly formulate the fluctuating part of the deformation gradient. Following its definition in (3.21) we know $\tilde{\mathbf{F}} = \mathbf{F} \bar{\mathbf{F}}^{-1}$. Therefore, multiplying $\mathbf{F}$ in equation (3.11) with $\bar{\mathbf{F}}^{-1}$ from the right hand side gives

$$\tilde{\mathbf{F}}(\mathbf{X}) = \mathbf{I} + \begin{cases} -(1 - \Lambda) \mathbf{a} \otimes (\bar{\mathbf{F}}^{-T} \hat{\mathbf{N}}) , & -\Lambda H \leq X_{\perp} \leq 0 , \\ \Lambda \mathbf{a} \otimes (\bar{\mathbf{F}}^{-T} \hat{\mathbf{N}}) , & 0 \leq X_{\perp} \leq (1 - \Lambda) H . \end{cases} \quad (3.23)$$

However, $\tilde{\mathbf{F}} = \partial \mathbf{x} / \partial \bar{\mathbf{x}}$; so we rewrite $\tilde{\mathbf{F}}$ as a function of $\bar{\mathbf{x}}$. In doing so, we make use of equations (3.14) and (3.19) which give

$$\bar{\mathbf{F}}^{-T} \hat{\mathbf{N}} = H / h \hat{\mathbf{n}} . \quad (3.24)$$

Also the lath width and volume fraction in the average laminate configuration are

$$\bar{h} = h , \quad \bar{\lambda} = \Lambda , \quad (3.25)$$

which finally reformulate (3.23) into

$$\tilde{\mathbf{F}}(\bar{\mathbf{x}}) = \mathbf{I} + \frac{H}{h} \begin{cases} -(1 - \Lambda) \mathbf{a} \otimes \hat{\mathbf{n}} , & -\Lambda h \leq \bar{\mathbf{x}} \cdot \hat{\mathbf{n}} \leq 0 , \\ \Lambda \mathbf{a} \otimes \hat{\mathbf{n}} , & 0 \leq \bar{\mathbf{x}} \cdot \hat{\mathbf{n}} \leq (1 - \Lambda) h . \end{cases} \quad (3.26)$$

We now move on to studying the kinematics of the eigen deformation and how it relates to what we have developed so far.

3.1.3 Laminate in eigen configuration

A first-order laminate in its eigen configuration is composed of two variants with deformation gradients $\mathbf{F}_1^\circ$ and $\mathbf{F}_2^\circ$ which characterize martensitic transformations from a common austenitic parent. Such a laminate microstructure is expected to fulfill the coherence condition:

$$\mathbf{F}_2^\circ - \mathbf{F}_1^\circ = \mathbf{a}^\circ \otimes \hat{\mathbf{N}}, \quad (3.27)$$

where, from the crystallographic theory according (2.35), we know that the eigen deformation gradients are symmetry-connected through the point group of the parent austenitic phase via

$$\mathbf{U}_2^\circ = \mathbf{R}^T \mathbf{U}_1^\circ \mathbf{R}, \quad \exists \mathbf{R} \in \mathcal{G}_+(\mathbf{a}_i^a), \quad (3.28)$$

where $\mathbf{U}_i^\circ$ are the Bain matrices of transformation. These are in turn connected to transformation deformation gradients through polar decomposition

$$\mathbf{F}_i^\circ = \mathbf{Q}_i \mathbf{U}_i^\circ, \quad \mathbf{Q}_i \in \text{SO}(3). \quad (3.29)$$

The periodic structure of the laminate in this state resembles that of equation (3.5), namely

$$\mathbf{F}^\circ(\mathbf{X}) = \begin{cases} \mathbf{F}_1^\circ, & -\Lambda H \leq X_\perp \leq 0, \\ \mathbf{F}_2^\circ, & 0 \leq X_\perp \leq (1 - \Lambda) H. \end{cases} \quad (3.30)$$

Remember that the transformations $\mathbf{F}_1^\circ$ and $\mathbf{F}_2^\circ$ are physical properties of the martensite and are therefore predetermine with respect to a given austenite orientation.

It should be obvious that all the arguments and conclusions starting from equation (3.8) all the way to (3.26) shall identically hold for the eigen configuration, since they solely rely on the compatibility and the periodicity of the laminate. The eigen configuration, however, bears an additional restriction—the symmetry connection (3.28). In this equation since $\mathbf{R}$ belongs to the point group of the austenite, it is proper orthogonal and therefore $\det \mathbf{R} = \det \mathbf{R}^T = 1$; hence

$$\det \mathbf{F}_2^\circ = \det(\mathbf{R}^T \mathbf{F}_1^\circ \mathbf{R}) = \det \mathbf{R}^T \det \mathbf{F}_1^\circ \det \mathbf{R} = \det \mathbf{F}_1^\circ. \quad (3.31)$$

This together with (3.18) immediately gives

$$\det \langle \mathbf{F}^\circ \rangle = \det \mathbf{F}_1^\circ = \det \mathbf{F}_2^\circ, \quad (3.32)$$

which in turn can be combined with (3.10) and (3.20) to arrive at the conclusion that the volume fraction cannot change via eigen deformation:

$$\lambda(\mathbf{x}^\circ) = \Lambda. \quad (3.33)$$

Comparing this with (3.20) we realize that the eigen configuration by itself does not allow for the accommodation which we intend for our kinematic ansatz to be capable of. Further, the following lemma shows that through the eigen deformation the the laminate interfaces cannot deflect.

Lemma 3. *For every two rank one connected tensors $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$, the rank one connection of the form $\mathbf{A} - \mathbf{B} = \mathbf{a} \otimes \mathbf{n}$, such that $\mathbf{a}, \mathbf{n} \in \mathbb{R}^n$ and $|\mathbf{n}| = 1$, is unique up to a factor ± 1 .*

Proof. We can prove this assertion by contradiction. Assume that there exists two distinct rank one connections, say

$$\mathbf{A} - \mathbf{B} = \mathbf{a}_1 \otimes \mathbf{n}_1, \quad \mathbf{A} - \mathbf{B} = \mathbf{a}_2 \otimes \mathbf{n}_2.$$

We inner-multiply both identities from the right hand side by $\mathbf{n}_1$ and $\mathbf{n}_2$ to get

$$(\mathbf{A} - \mathbf{B}) \cdot \mathbf{n}_1 = \mathbf{a}_1 = (\mathbf{n}_2 \cdot \mathbf{n}_1) \mathbf{a}_2, \quad (\mathbf{A} - \mathbf{B}) \cdot \mathbf{n}_2 = \mathbf{a}_2 = (\mathbf{n}_1 \cdot \mathbf{n}_2) \mathbf{a}_1.$$

Comparing these two relations results in $\mathbf{n}_1 \cdot \mathbf{n}_2 = \pm 1$. Since $|\mathbf{n}_1| = |\mathbf{n}_2| = 1$, we then conclude that $\mathbf{n}_1 = \pm \mathbf{n}_2$, which consequently requires that $\mathbf{a}_1 = \mp \mathbf{a}_2$. $\square$

The deformation gradient $\mathbf{F}^\circ$ in (3.30) is piecewise constant, therefore, the coherence condition (3.27) requires that the twin interfaces remain flat following the recent lemma. But if we let H and Λ vary along the laminate, as we did in (3.6), the laminate interfaces deflect and bifurcate which contradicts the flatness of the interfaces. In hindsight, this clarifies the motivation behind designing a more flexible ansatz as we did in formulating (3.11) and (3.26). However, we did not formulate the un-accommodated laminate only to motivate our approach after the fact. We did so because the eigen state represents the *ground energy level* when the martensitic transformation has already taken place. We will see that there is no apparent dynamic connection between the thermomechanical ground state, i.e. the martensite in eigen state, and the fully accommodated laminate. How this ground state relates to the stored energy of the accommodated laminate is what we will explain shortly.

3.1.4 Accommodation strains

Thermomechanical behavior of solids is formulated in terms of various state functions such as thermodynamic potentials, and these functions depend on deformation gradient rather than the deformation field itself. Consequently, the information that we have on the martensitic phase transformation and its symmetry properties are directly expressed in terms of deformation gradient. That is why we started our kinematic formulation by formulating the corresponding deformation gradients and not the deformation field.

Now in order to account for accommodation strains that are caused by changes in volume fraction and lath width we need to calculate the corresponding deformation gradients and strains along the laminate. For this we first have to integrate $\mathbf{F}$ in (3.11) to obtain the deformation, and then calculate the gradients that originate from the changes in Λ and H .

$$\begin{aligned} \mathbf{x} &= \int_{\mathbf{x}_0}^{\mathbf{x}} \mathbf{F} \cdot d\mathbf{X} = \int_{\mathbf{x}_0}^{\mathbf{x}} \langle \mathbf{F} \rangle_\perp \cdot d\mathbf{X} + \int_0^{X_\perp} \left\{ \begin{array}{c} -(1-\Lambda) \mathbf{a} \otimes \hat{\mathbf{N}} \\ \Lambda \mathbf{a} \otimes \hat{\mathbf{N}} \end{array} \right\} \cdot \hat{\mathbf{N}} dX_\perp \\ &= \int_{\mathbf{x}_0}^{\mathbf{x}} \langle \mathbf{F} \rangle_\perp \cdot d\mathbf{X} + \left\{ \begin{array}{c} -\Lambda(1-\Lambda)H - (1-\Lambda)X_\perp \\ -\Lambda(1-\Lambda)H + \Lambda X_\perp \end{array} \right\} \mathbf{a}, \end{aligned} \quad (3.34)$$

where a compact notation is adopted here, in which the variant thickness intervals are implied in braces $\{\bullet\}$, for the first argument as $X_\perp \in [0, \Lambda H]$ and for the second argument as $X_\perp \in [\Lambda H, H]$. In principle, the normal average deformation gradient $\langle \mathbf{F} \rangle_\perp$ can vary along $\mathbf{X}_\parallel$. Nevertheless, one might assume $\langle \mathbf{F} \rangle_\perp = \langle \mathbf{F} \rangle_\omega$, since the laminate ansatz cannot match the mesoscopic homogeneous deformation at the boundary anyway (see remark 3). So letting $\langle \mathbf{F} \rangle_\perp$ vary parallel to the laminate would basically make the formulation more complicated without producing any significant information. Therefore, equation (3.34) can be written as

$$\mathbf{x} = \langle \mathbf{F} \rangle_\omega \cdot \mathbf{X} + \begin{Bmatrix} -\Lambda(1-\Lambda)H - (1-\Lambda)X_\perp \\ -\Lambda(1-\Lambda)H + \Lambda X_\perp \end{Bmatrix} \mathbf{a}. \quad (3.35)$$

We separate the variables as much as possible in order to simplify the consequent calculation of derivatives:

$$\mathbf{x} = \langle \mathbf{F} \rangle_\omega \cdot \mathbf{X} + \left(-\Lambda(1-\Lambda)H \begin{Bmatrix} 1 \\ 1 \end{Bmatrix} + X_\perp \begin{Bmatrix} -(1-\Lambda) \\ \Lambda \end{Bmatrix} \right) \mathbf{a}. \quad (3.36)$$

We can further compact the formulation by introducing

$$\bar{\mathbf{a}} = -\Lambda(1-\Lambda)\mathbf{a}, \quad (3.37)$$

which turns (3.36) into

$$\mathbf{x} = \langle \mathbf{F} \rangle_\omega \cdot \mathbf{X} + \left(H \begin{Bmatrix} 1 \\ 1 \end{Bmatrix} + X_\perp \begin{Bmatrix} 1/\Lambda \\ -1/(1-\Lambda) \end{Bmatrix} \right) \bar{\mathbf{a}}. \quad (3.38)$$

In calculating the deformation gradient, most of quantities are independent of the $X_\perp$. Therefore, their gradients reduce to components parallel to the laminate. It is, therefore, helpful to introduce the reduced gradient as

$$\nabla \equiv (\mathbf{I} - \hat{\mathbf{N}} \otimes \hat{\mathbf{N}}) \cdot \nabla_{\mathbf{X}}. \quad (3.39)$$

Using this notation, the deformation gradient is

$$\begin{aligned} \mathbf{x} \otimes \nabla_{\mathbf{x}} &= \langle \mathbf{F} \rangle_\omega \cdot \frac{\partial \mathbf{X}}{\partial \mathbf{x}} + \left(H \begin{Bmatrix} 1 \\ 1 \end{Bmatrix} + X_\perp \begin{Bmatrix} 1/\Lambda \\ -1/(1-\Lambda) \end{Bmatrix} \right) \bar{\mathbf{a}} \otimes \nabla + \\ &\quad \bar{\mathbf{a}} \otimes \left((H\nabla) \begin{Bmatrix} 1 \\ 1 \end{Bmatrix} + \hat{\mathbf{N}} \begin{Bmatrix} 1/\Lambda \\ -1/(1-\Lambda) \end{Bmatrix} + X_\perp (\Lambda\nabla) \begin{Bmatrix} -1/\Lambda^2 \\ -1/(1-\Lambda)^2 \end{Bmatrix} \right). \end{aligned} \quad (3.40)$$

The first term on the right hand side can be simplified using $\partial \mathbf{X} / \partial \mathbf{x} = \mathbf{I}$. Factoring and gathering similar terms, the above equation becomes

$$\mathbf{x} \otimes \nabla_{\mathbf{x}} = \langle \mathbf{F} \rangle_\omega + \begin{Bmatrix} H + \frac{1}{\Lambda} X_\perp \\ H - \frac{1}{1-\Lambda} X_\perp \end{Bmatrix} \bar{\mathbf{a}} \otimes \nabla + \bar{\mathbf{a}} \otimes \begin{Bmatrix} H\nabla + \frac{1}{\Lambda} \hat{\mathbf{N}} - \frac{1}{\Lambda^2} X_\perp (\Lambda\nabla) \\ H\nabla - \frac{1}{1-\Lambda} \hat{\mathbf{N}} - \frac{1}{(1-\Lambda)^2} X_\perp (\Lambda\nabla) \end{Bmatrix}. \quad (3.41)$$

Remark 4. ON MORPHOLOGY OF ACCOMMODATION. It can be shown, through purely geometrical arguments, that both the accommodation by branching and the needles require that

$$\hat{\mathbf{N}} - (\hat{\mathbf{N}} \cdot \hat{\mathbf{B}})\hat{\mathbf{B}} = [\Lambda H/2] \nabla. \quad (3.42)$$

But this is not an additional equation. Since it can re-parameterize two quantities at best; namely, Λ and H , while at the same time it introduces two new unknowns corresponding to the *unit* vector $\hat{\mathbf{B}}$. Nevertheless, the geometrical implications of this identity can be the basis for a kinematic ansatz of the laminate in which $\mathbf{a}$ does not need to vary, while retaining all the important characteristics, such as accommodation by both branches and needles, of the seemingly less restrictive ansatz we just established (see Appendix D).

We are now ready to proceed to calculation of the average laminate energy.

3.2 Kinetics

A thermodynamic potential is a *state function* whose value is determined relative to a fixed reference value corresponding to which exists a reference state. The choice of this reference state is immaterial from a mathematical viewpoint, since its changes will shift the underlying thermodynamic potential merely by a constant. However, from a practical perspective, the reference state is often assigned to a state that is empirically easy to identify, reproduce, and track. The state of minimum potential is often the most appropriate choice, since the system reaches a stable equilibrium at that state; meaning: the system has the tendency to reach its minimum potential energy and tends to remain there as long as no significant perturbation is present.

A laminate microstructure attains its state of minimum energy if its constituent phases reach their respective energetic minima. Should this be the case, the laminate is in eigen configuration or eigen state whose kinematics we already studied. In a deformation-controlled framework the state of the material is determined by its deformation field $\mathbf{F}(\mathbf{X})$ and its temperature field $\theta(\mathbf{X})$. The corresponding thermodynamic potential will, therefore, be the Helmholtz free energy density $\psi(\mathbf{F}, \theta)$. Since we excluded the thermal effects from our model, at a fixed temperature the Helmholtz free energy will be a function of deformation only: $\psi = \psi(\mathbf{F})$. Having the eigen state fixed as the ground energy state, we then obtain

$$\psi(\mathbf{F}) = \psi(\mathbf{F}^\circ) + \int_{\mathbf{F}^\circ}^{\mathbf{F}} d\psi = \psi(\mathbf{F}^\circ) + \int_{\mathbf{F}^\circ}^{\mathbf{F}} \frac{\partial \psi}{\partial \mathbf{F}} : d\mathbf{F}. \quad (3.43)$$

The differential $d\mathbf{F}$ is meaningful only if there exists a continuous deformation that takes the material from the eigen state $\mathbf{F}^\circ$ to the accommodated state $\mathbf{F}$; which is sometimes called a *dynamic connection* between the two states. However, the accommodated state depends pointwise on the volume fraction $\Lambda(\mathbf{X}_\parallel)$ and the lath width $H(\mathbf{X}_\parallel)$ along the laminate; that is $\mathbf{F} = \mathbf{F}(\Lambda(\mathbf{X}_\parallel), H(\mathbf{X}_\parallel); \mathbf{X}_\perp)$. While, on the other hand, the eigen state can be fully identified pointwise with a volume fraction and a lath width which turn out to be constant,

i.e. $\mathbf{F}^\circ = \mathbf{F}^\circ(\Lambda, H, \mathbf{X}_\perp)$. Since the accommodated state has apparently more *internal degrees of freedom*, there cannot exist a one-to-one correspondence between the accommodated state and the eigen state—let alone a dynamic connection. Intuitively, the accommodated state permits the laminate interfaces to bifurcate, whereas the eigen state requires flatness of the interfaces; which are two entirely different topologies with no seeming *morphism*. As dire as the situation may look, there is a way out.

We mentioned that a non degenerate continuum theory of martensitic phase transformation is based on the assumption that the transformation strains remain within the Ericksen–Pitteri neighborhood to eliminate the possibility of dislocation and slip. This readily limits the possibility of large deformations to a great extent. Furthermore, in any variational formulation of the martensitic microstructure, the strict existence of a minimizer is replaced with a weaker condition, namely the existence of minimizing sequences. Hence, instead of knowing exactly how the system arrived at a certain state, it is enough to be able to uniquely determine the free energy as a function of the state. In these minimizing constructs, it is well established that deviations from the energy well (eigen state) shall be small compared with the eigen deformations [83]. That is to say

$$\|\mathbf{U}_i - \mathbf{I}\| = \mathcal{O}(\delta) \quad \Rightarrow \quad \text{dist}(\mathbf{F}, \text{SO}(3) \mathbf{U}_i) = \mathcal{O}(\delta^2) , \quad (3.44)$$

and the eigen deformations themselves are limited to Ericksen–Pitteri neighborhoods, which means $\|\mathbf{F}^\circ - \mathbf{I}\| \ll 1$. So we can safely assume that the Helmholtz free energy of the laminate can be asymptotically expanded around the eigen state:

$$\psi(\mathbf{F}) = \underbrace{\psi(\mathbf{F}^\circ) + \frac{\partial\psi}{\partial\mathbf{F}}\Big|_{\mathbf{F}^\circ} : (\mathbf{F} - \mathbf{F}^\circ)}_{=0} + \frac{1}{2} (\mathbf{F} - \mathbf{F}^\circ) : \underbrace{\frac{\partial^2\psi}{\partial\mathbf{F} \otimes \partial\mathbf{F}}\Big|_{\mathbf{F}^\circ}}_{=\mathbb{C}} : (\mathbf{F} - \mathbf{F}^\circ) + \mathcal{O}(\|\mathbf{F} - \mathbf{F}^\circ\|^3) , \quad (3.45)$$

in which the second term vanishes, since $\mathbf{F} = \mathbf{F}^\circ$ marks the martensitic energy well. The third term is nothing but the linearized elastic energy, which we designate as ψ_ℓ . So we finally arrive at $\psi(\mathbf{F}) \approx \psi(\mathbf{F}^\circ) + \psi_\ell(\mathbf{F} - \mathbf{F}^\circ)$, in which the term $\psi(\mathbf{F}^\circ)$ denotes the ground energy level and is usually attributed to the material's stored chemical energy, denoted as c , which is assumed to be a known material parameter depending on the temperature, $c = c(\theta)$. For practical purposes the approximation is precise enough to be exchanged for equality, so we finally write

$$\psi(\mathbf{F}) = \psi_\ell(\mathbf{F} - \mathbf{F}^\circ) + c(\theta) , \quad (3.46)$$

where

$$\psi_\ell(\mathbf{F} - \mathbf{F}^\circ) = \frac{1}{2} (\mathbf{F} - \mathbf{F}^\circ) : \mathbb{C} : (\mathbf{F} - \mathbf{F}^\circ) , \quad (3.47)$$

and

$$\mathbb{C} = \frac{\partial^2\psi}{\partial\mathbf{F} \otimes \partial\mathbf{F}}\Big|_{\mathbf{F}^\circ} . \quad (3.48)$$

In the above equations, while subtracting $\mathbf{F}^\circ$ from $\mathbf{F}$, we need to clarify two issues:

- Orientation of the accommodated state and the eigen state must be given with respect to the parent phase.
- The two configurations, i.e. $\mathbf{F}$ and $\mathbf{F}^\circ$, must be defined on a common support.

Regarding the relative orientation, as far as we know, all the minimizing sequences are constructed based on the assumption that the two configurations share the same orientation $\hat{\mathbf{N}}$ in the reference configuration (parent austenite). These minimizing constructs also allow the eigen configuration to vary the lath width and the volume fraction in the same manner as the accommodated deformation gradient. It will be obvious, then, that the eigen deformation gradient in (3.30) will not fulfill the kinematic compatibility condition; meaning: there is no admissible deformation field corresponding to the eigen configuration. Notice, however, that the formation of a fully accommodated laminate takes place through phase transformation, which is spontaneous and does not necessarily pass through any intermediate state—including the eigen configuration. There is extensive and rigorous mathematical treatment of these minimizing sequences, using weak differentiability and compactness arguments, which deem this approach favorable (see [97, 99, 98]). The bottom line is that subtraction of deformation gradients in (3.46) can be carried out in material coordinates without ambiguity.

Remark 5. Neglecting the thermal effects requires that the temperature θ is kept constant throughout the body, which means that the body is in contact with a heat bath, and that the temperature variations within the body take place at time scales much smaller than those of mechanical loading.

3.2.1 Linear hyperelastic material—an interlude

According to (3.46), the kinetics of the martensitic laminate is restricted to small deformations at the microscale. A solid material whose Helmholtz free energy takes the form of (3.46) is called a *linear elastic material*. The argument of the linearized term, $\psi_\ell(\mathbf{F} - \mathbf{F}^\circ)$, can be expressed in terms of displacement field instead of deformation field, using $\mathbf{F} = \mathbf{I} + \mathbf{u}\nabla$, and $\mathbf{F}^\circ = \mathbf{I} + \mathbf{u}^\circ\nabla$, which yield $\mathbf{F} - \mathbf{F}^\circ = \mathbf{u}\nabla - \mathbf{u}^\circ\nabla$. Then, the linearized energy ψ_ℓ is explicitly written as

$$\psi_\ell(\mathbf{u}\nabla - \mathbf{u}^\circ\nabla) = \frac{1}{2} (\mathbf{u}\nabla - \mathbf{u}^\circ\nabla) : \mathbb{C} : (\mathbf{u}\nabla - \mathbf{u}^\circ\nabla) . \quad (3.49)$$

The stiffness tensor, defined in (3.48), is a fourth-order tensor of third rank, i.e. $\mathbb{C} \in \mathbb{R}^{3 \times 3 \times 3 \times 3}$. Since $\partial\mathbf{F}/\partial\mathbf{u}\nabla = \mathbf{I}$, the stiffness tensor can also be written as

$$\mathbb{C} = \left. \frac{\partial^2 \psi}{\partial \mathbf{u}\nabla \otimes \partial \mathbf{u}\nabla} \right|_{\mathbf{u}^\circ\nabla} , \quad (3.50)$$

In addition to inheriting the crystal symmetries of its respective material, the stiffness tensor possesses the so called *major and minor symmetry*.

The major symmetry expressed as

$$C_{ijkl} = C_{klij} , \quad (3.51)$$

is an immediate consequence of the definition (3.50). Expressed in index notation, this definition becomes:

$$C_{iJkL} = \frac{\partial^2 \psi}{\partial u_{i,J} \partial u_{k,L}} \Big|_{\mathbf{u} \circ \nabla} = \frac{\partial^2 \psi}{\partial u_{k,L} \partial u_{i,J}} \Big|_{\mathbf{u} \circ \nabla} = C_{kLiJ}.$$

Here, we have employed the property that the sequence of differentiation is interchangeable. Also, notice that the distinction between the *Eulerian* indices i and k , and the *Lagrangian* indices J and L can be neglected due to the small deformations; as in (3.51).

The minor symmetry, given by

$$C_{ijkl} = C_{jikl} = C_{ijlk}, \quad (3.52)$$

comes from the principle of *frame indifference*, or *objectivity*, applied to ψ_ℓ . Writing (3.49) in a more tractable form, namely $\psi_\ell(\mathbf{u}\nabla) = \frac{1}{2} \mathbf{u}\nabla : \mathbb{C} : \mathbf{u}\nabla$, and keeping in mind that the deformations are small, we decompose the displacement gradient *additively* into symmetric part and antisymmetric part

$$\nabla^s \mathbf{u} = \frac{1}{2} (\mathbf{u}\nabla + \nabla \mathbf{u}), \quad \nabla^a \mathbf{u} = \frac{1}{2} (\mathbf{u}\nabla - \nabla \mathbf{u}). \quad (3.53)$$

Next, putting the decomposed form back into ψ_ℓ , using the major symmetry of $\mathbb{C}$, we obtain

$$\begin{aligned} \psi_\ell(\mathbf{u}\nabla) &= \psi_\ell(\nabla^s \mathbf{u} + \nabla^a \mathbf{u}) \\ &= \frac{1}{2} \nabla^s \mathbf{u} : \mathbb{C} : \nabla^s \mathbf{u} + \frac{1}{2} \nabla^a \mathbf{u} : \mathbb{C} : \nabla^a \mathbf{u} + \nabla^a \mathbf{u} : \mathbb{C} : \nabla^s \mathbf{u} \\ &= \psi_\ell(\nabla^s \mathbf{u}) + \psi_\ell(\nabla^a \mathbf{u}) + \nabla^a \mathbf{u} : \mathbb{C} : \nabla^s \mathbf{u}. \end{aligned} \quad (3.54)$$

The second term vanishes due to objectivity of ψ_ℓ , because $\nabla^a \mathbf{u}$ geometrically represents a rotation. This term can be written in index notation as $1/2 u_{i,j}^a C_{ijkl} u_{k,l}^a$. Since $\nabla^a \mathbf{u}$ is antisymmetric, we have $u_{i,j}^a = -u_{j,i}^a$; therefore

$$0 = 1/2 u_{i,j}^a C_{ijkl} u_{k,l}^a = 1/2 u_{j,i}^a C_{jikl} u_{k,l}^a = -1/2 u_{i,j}^a C_{jikl} u_{k,l}^a,$$

which leads to

$$\forall u_{i,j}^a; \quad u_{i,j}^a (C_{ijkl} - C_{jikl}) u_{k,l}^a = 0 \quad \Rightarrow \quad C_{ijkl} - C_{jikl} = 0,$$

since $C_{ijkl} - C_{jikl}$ is itself antisymmetric with respect to ij . This result together with major symmetry finally yields the *left and right minor symmetry*:

$$C_{ijkl} = C_{jikl} = C_{ijlk}, \quad (3.52)$$

As an immediate consequence of minor symmetries, the last term in (3.54) vanishes too, because double-contraction of a symmetric tensor with an antisymmetric tensor equals zero. Therefore, we finally obtain

$$\psi_\ell(\boldsymbol{\varepsilon}) = \frac{1}{2} \boldsymbol{\varepsilon} : \mathbb{C} : \boldsymbol{\varepsilon}, \quad (3.55)$$

in which $\boldsymbol{\varepsilon} = \nabla^s \mathbf{u}$ and is called *Cauchy strain*, or sometimes small strain, or linear strain. Accordingly, the stiffness tensor can be reformulated as

$$\mathbb{C} = \left. \frac{\partial^2 \psi}{\partial \boldsymbol{\varepsilon} \otimes \partial \boldsymbol{\varepsilon}} \right|_{\boldsymbol{\varepsilon}=\mathbf{0}}. \quad (3.56)$$

This tensor has another obvious yet important property, being *positive definite* over the set of small strains, which means

$$\boldsymbol{\varepsilon} : \mathbb{C} : \boldsymbol{\varepsilon} \geq 0, \quad (3.57)$$

for all symmetric second-order tensors $\boldsymbol{\varepsilon}$ that fulfill the kinematic compatibility condition as $\nabla \times \boldsymbol{\varepsilon} \times \nabla = \mathbf{0}$. For future reference, let us also introduce the small rotation tensor $\boldsymbol{\omega} = \nabla^a \mathbf{u}$. Remember that both strain $\boldsymbol{\varepsilon}$ and rotation $\boldsymbol{\omega}$ must remain small for linear hyperelasticity to hold.

Restated in terms of *thermodynamic driving forces*, the energetic formulation (3.55) is equivalent to saying that the stress–strain response of the material is linear at the vicinity of equilibrium. Then, the slope of this linear stress–strain response is nothing but the stiffness tensor

$$\boldsymbol{\sigma} = \mathbb{C} : \boldsymbol{\varepsilon}, \quad (3.58)$$

where the *Cauchy stress* tensor, i.e. the thermodynamic driving force, is given by

$$\boldsymbol{\sigma} = \left. \frac{\partial \psi}{\partial \boldsymbol{\varepsilon}} \right|_{\boldsymbol{\varepsilon}=\mathbf{0}}. \quad (3.59)$$

Most of the foregoing arguments are basically a direct generalizations of the Hooke's law. That is why a linear hyperelastic material is often called a *Hookean material*.

Remark 6. Mathematically, since the stiffness tensor is left and right minor symmetric, we have not gained anything in deriving (3.55) from (3.49), because inner multiplication of a symmetric tensor with an arbitrary tensor is not influenced by the antisymmetric part of the arbitrary tensor. From a computational standpoint, however, putting the symmetric gradients directly into the formulation will help avoiding the computational round-off errors that will eventually render the strains non-symmetric.

3.2.2 Average laminate energy

The following calculations have been partially carried out symbolically using **Mathematica**TM. We start out by substituting $\mathbf{F}$ from (3.41) into (3.47) and averaging the free energy across a single martensite twin as

$$\bar{\psi} = \frac{1}{2H} \int_{-AH}^{(1-A)H} (\mathbf{F} - \mathbf{F}^\circ) : \mathbb{C} : (\mathbf{F} - \mathbf{F}^\circ) \, dX_\perp. \quad (3.60)$$

The free energy ψ is expressed in the reference configuration; that is, the free energy per unit volume in material coordinates. The terms $(\mathbf{F} - \mathbf{F}^\circ)$ will be symmetrized after the

simplifications, to avoid unnecessary calculations. The above integral is split over the two variants as $\int_{-\Lambda H}^{(1-\Lambda)H} \bullet = \int_{-\Lambda H}^0 \bullet + \int_0^{(1-\Lambda)H} \bullet$. For the first variant we have

$$\begin{aligned}
\bar{\psi}_1 &= \frac{1}{2H} \int_{-\Lambda H}^0 [\mathbf{F}_1 - \mathbf{F}_1^\circ] : \mathbb{C}_1 : [\mathbf{F}_1 - \mathbf{F}_1^\circ] \, dX_\perp = \\
&\quad \frac{1}{2} [\bar{\mathbf{a}} \otimes H\nabla] : \mathbb{C}_1 : \left[\Lambda H(\bar{\mathbf{a}} \otimes \nabla) + H(\bar{\mathbf{a}} \otimes \Lambda \nabla) + 2(\bar{\mathbf{a}} \otimes \hat{\mathbf{N}}) + 2\Lambda (\bar{\mathbf{F}} - \mathbf{F}_1^\circ) \right] \\
&\quad + \frac{H}{6\Lambda} [\bar{\mathbf{a}} \otimes \Lambda \nabla] : \mathbb{C}_1 : \left[\Lambda H(\bar{\mathbf{a}} \otimes \nabla) + 3(\bar{\mathbf{a}} \otimes \hat{\mathbf{N}}) + 3\Lambda (\bar{\mathbf{F}} - \mathbf{F}_1^\circ) \right] \\
&\quad + \frac{1}{2\Lambda} [\bar{\mathbf{a}} \otimes \hat{\mathbf{N}}] : \mathbb{C}_1 : \left[\Lambda H(\bar{\mathbf{a}} \otimes \nabla) + 2\Lambda (\bar{\mathbf{F}} - \mathbf{F}_1^\circ) \right] - \frac{\Lambda H}{2} [\bar{\mathbf{a}} \otimes \nabla] : \mathbb{C}_1 : [\bar{\mathbf{F}} - \mathbf{F}_1^\circ] \\
&\quad + \frac{\Lambda H^2}{6} [\bar{\mathbf{a}} \otimes \nabla] : \mathbb{C}_1 : [\bar{\mathbf{a}} \otimes \nabla] + \frac{\Lambda}{2} [\bar{\mathbf{a}} \otimes H\nabla] : \mathbb{C}_1 : [\bar{\mathbf{a}} \otimes H\nabla] + \frac{H^2}{6\Lambda} [\bar{\mathbf{a}} \otimes \Lambda \nabla] : \mathbb{C}_1 : [\bar{\mathbf{a}} \otimes \Lambda \nabla] \\
&\quad + \frac{1}{2\Lambda} [\bar{\mathbf{a}} \otimes \hat{\mathbf{N}}] : \mathbb{C}_1 : [\bar{\mathbf{a}} \otimes \hat{\mathbf{N}}] + \frac{\Lambda}{2} [\bar{\mathbf{F}} - \mathbf{F}_1^\circ] : \mathbb{C}_1 : [\bar{\mathbf{F}} - \mathbf{F}_1^\circ] , \tag{3.61}
\end{aligned}$$

and for the second variant

$$\begin{aligned}
\bar{\psi}_2 &= \frac{1}{2H} \int_0^{(1-\Lambda)H} [\mathbf{F}_2 - \mathbf{F}_2^\circ] : \mathbb{C}_2 : [\mathbf{F}_2 - \mathbf{F}_2^\circ] \, dX_\perp = \\
&\quad \frac{1}{2} [\bar{\mathbf{a}} \otimes H\nabla] : \mathbb{C}_2 : \left[(1-\Lambda)H(\bar{\mathbf{a}} \otimes \nabla) - H(\bar{\mathbf{a}} \otimes \Lambda \nabla) - 2(\bar{\mathbf{a}} \otimes \hat{\mathbf{N}}) + 2(1-\Lambda) (\bar{\mathbf{F}} - \mathbf{F}_2^\circ) \right] \\
&\quad + \frac{H}{6(1-\Lambda)} [\bar{\mathbf{a}} \otimes \Lambda \nabla] : \mathbb{C}_2 : \left[-(1-\Lambda)H(\bar{\mathbf{a}} \otimes \nabla) + 3(\bar{\mathbf{a}} \otimes \hat{\mathbf{N}}) - 3(1-\Lambda) (\bar{\mathbf{F}} - \mathbf{F}_2^\circ) \right] \\
&\quad + \frac{1}{2(1-\Lambda)} [\bar{\mathbf{a}} \otimes \hat{\mathbf{N}}] : \mathbb{C}_2 : \left[-(1-\Lambda)H(\bar{\mathbf{a}} \otimes \nabla) - 2(1-\Lambda) (\bar{\mathbf{F}} - \mathbf{F}_2^\circ) \right] \\
&\quad + \frac{(1-\Lambda)H}{2} [\bar{\mathbf{a}} \otimes \nabla] : \mathbb{C}_2 : [(\bar{\mathbf{F}} - \mathbf{F}_2^\circ)] + \frac{(1-\Lambda)H^2}{6} [\bar{\mathbf{a}} \otimes \nabla] : \mathbb{C}_2 : [\bar{\mathbf{a}} \otimes \nabla] \\
&\quad + \frac{1-\Lambda}{2} [\bar{\mathbf{a}} \otimes H\nabla] : \mathbb{C}_2 : [\bar{\mathbf{a}} \otimes H\nabla] + \frac{H^2}{6(1-\Lambda)} [\bar{\mathbf{a}} \otimes \Lambda \nabla] : \mathbb{C}_2 : [\bar{\mathbf{a}} \otimes \Lambda \nabla] \\
&\quad + \frac{1}{2(1-\Lambda)} [\bar{\mathbf{a}} \otimes \hat{\mathbf{N}}] : \mathbb{C}_2 : [\bar{\mathbf{a}} \otimes \hat{\mathbf{N}}] + \frac{1-\Lambda}{2} [\bar{\mathbf{F}} - \mathbf{F}_2^\circ] : \mathbb{C}_2 : [\bar{\mathbf{F}} - \mathbf{F}_2^\circ] . \tag{3.62}
\end{aligned}$$

Introducing the energy inner product and its corresponding norm

$$\langle \alpha, \beta \rangle_{\mathbb{C}_i} = \int_{\Omega_i} \alpha : \mathbb{C}_i : \beta \, dV , \quad \|\bullet\|_{\mathbb{C}_i}^2 = \langle \bullet, \bullet \rangle_{\mathbb{C}_i} , \tag{3.63}$$

we can summarize the above formulas for the two variants, $i = 1, 2$, in the form of

$$\begin{aligned}
\bar{\psi}_i &= \frac{1}{2H} \int_{\Omega_i} (\mathbf{F}_i - \mathbf{F}_i^\circ) : \mathbb{C}_i : (\mathbf{F}_i - \mathbf{F}_i^\circ) \, dX_\perp = \\
&\quad + \frac{1}{6\Lambda_i} \left\| \Lambda_i H(\bar{\mathbf{a}} \otimes \nabla) + H\bar{\mathbf{a}} \otimes (\Lambda_i \nabla) \right\|_{\mathbb{C}_i}^2 - \frac{1}{6\Lambda_i} \langle \Lambda_i H(\bar{\mathbf{a}} \otimes \nabla), H\bar{\mathbf{a}} \otimes (\Lambda_i \nabla) \rangle_{\mathbb{C}_i} , \\
&\quad + \frac{1}{2\Lambda_i} \langle (\Lambda_i H\bar{\mathbf{a}}) \otimes \nabla \pm \bar{\mathbf{a}} \otimes \hat{\mathbf{N}} + \Lambda_i (\bar{\mathbf{F}} - \mathbf{F}_i^\circ), \Lambda_i \bar{\mathbf{a}} \otimes (H\nabla) \pm \bar{\mathbf{a}} \otimes \hat{\mathbf{N}} + \Lambda_i (\bar{\mathbf{F}} - \mathbf{F}_i^\circ) \rangle_{\mathbb{C}_i} ,
\end{aligned} \tag{3.64}$$

in which $\Lambda_1 = \Lambda$, $\Lambda_2 = 1 - \Lambda$, $\Omega_1 = \{x \in \mathbb{R} : -\Lambda H < x < 0\}$, and $\Omega_2 = \{x \in \mathbb{R} : 0 < x < (1 - \Lambda)H\}$. In terms with the $\pm$ sign, the plus sign corresponds to the variant from which $\hat{\mathbf{N}}$

points *outward*, and the minus sign to the variant to which $\hat{\mathbf{N}}$ points *inward*. For instance when $\mathbf{F}_2 - \mathbf{F}_1 = \mathbf{a} \otimes \hat{\mathbf{N}}$, the plus sign holds for variant 1 and the minus sign for variant 2. Alternatively, instead of $\pm$ we could have put a factor $(-1)^{i+1}$ in front of those terms. We refrain from presenting the laborious derivations required to arrive at the compact form of average laminate energy in equation (3.64). When in doubt, one can verify their identity by reverse derivation, namely by expanding (3.64) for each variant and subtracting the result from (3.61) and (3.62) respectively.

One thing to notice here is that we have not used the energy formulation in terms of $\boldsymbol{\varepsilon}$, but rather the original formulation in terms of $\mathbf{F}$ is directly employed which may be undesirable in numerical implementation. We could, of course, have used displacement gradients or even their symmetrized counterpart instead of deformation gradients which would have cost additional calculation with no apparent gain at this point. Nevertheless, in the upcoming numerical solutions we will symmetrize the strains behind the scene to avoid round-off errors, as remarked 6. This comes down to symmetrizing the terms within $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$, in equation (3.64).

3.2.3 Existence of solution to mesoscopic problem

According to an existence theorem by J.M. Ball in 1976, it is required that the energy density be quasiconvex [5, 10, 13]. This makes it necessary that subtraction of average deformation gradient (mesoscopic deformation) from its pointwise value (microscopic deformation) is equal to a *perturbation* that fulfills homogeneous boundary conditions over the microscopic domain. Stated mathematically, over a microscopic domain ω enclosing a *material point* it should hold that

$$\mathbf{F} = \langle \mathbf{F} \rangle_\omega + \tilde{\mathbf{u}} \nabla : \quad \tilde{\mathbf{u}}|_{\partial\omega} = \mathbf{0}. \quad (3.65)$$

The infimum of the microscopic energy over all acceptable perturbations is called the quasi-convexified energy, give by

$$\mathcal{Q}\Psi(\mathbf{F}) = \inf_{\tilde{\mathbf{u}}} \left\{ \frac{1}{|\omega|} \int_{\omega} \Psi(\mathbf{F} + \tilde{\mathbf{u}} \nabla) \, dV \mid \tilde{\mathbf{u}}(\partial\omega) = \mathbf{0} \right\}, \quad (3.66)$$

which is realized through minimizing sequences. The quasiconvexified energy has a minimum for certain. An obvious necessary condition for existence of the infimum in (3.66) is that the perturbation part in (3.41) has to fulfill the the zero boundary condition. To see this, we distinguish between those parts of $\partial\omega$ where the laminate is parallel to the boundary, and the parts where this is not the case.

- Wherever the laminate is parallel to $\partial\omega$, the deformation $\mathbf{u} = \mathbf{x} - \mathbf{X}$, in (3.34) vanishes at $X_\perp = -\Lambda H$ and at $X_\perp = (1 - \Lambda)H$.
- At those parts where the laminate is not parallel to $\partial\omega$, any mismatch between the laminate ansatz and the homogeneous mesoscopic deformation is penalized via boundary energy, according to remark 3. In the theoretical limit when $\Lambda \rightarrow 0$, or $\Lambda \rightarrow 1$, or alternatively $H \rightarrow 0$, this boundary energy will make sure that $\tilde{\mathbf{u}} = \mathbf{0}$.

Still, this is merely a necessary condition. A mathematical investigation of the sufficient conditions for existence of the quasiconvex energy is beyond the scope of this work.

3.3 Crystallography of NiTi martensite

To a complete crystallographic study of a solid-to-solid phase transformation, there are three ingredients [122]:

1. Lattice parameters of the parent and child phases.
2. Lattice correspondence between the parent and child phases.
3. Lattice transformation map from the parent to the child phase.

Figure 3.2 shows the primitive unit cells of NiTi (49.75 at. % Ni) for B2 austenite and B19' martensite phases. The B2 austenite has a body-centered cubic structure in which the Ni atoms are located at the center of the unit cell. The B19' has an *almost* face-centered monoclinic structure; that is, the Ni planes are slightly shifted from their supposed positions. This unit cell is in *lattice correspondence* with a B2 face-centered orthorhombic non-primitive unit cell which is twice the size of the cubic B2 primitive unit cell. The lattice parameters B2 and B19' phases of NiTi are summarized in Tab. 3.1.

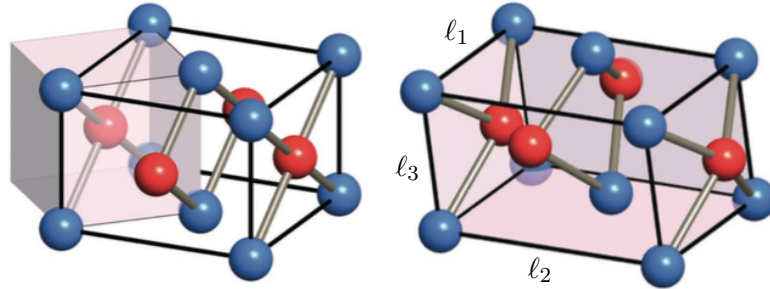


Fig. 3.2: Lattice structures of NiTi (49.75 at. % Ni) for: **(left)** B2 austenite with body-centered cubic primitive unit cell and the face-centered orthorhombic *non-primitive* unit cell; **(right)** B19' martensite with and almost face-centered monoclinic unit cell [78, © permission obtained].

Remark 7. It has been hypothesized theoretically and investigated experimentally that the B19' martensite at its ground energy state is not entirely stable and has the tendency towards a stable martensite phase called B33 [17, 149]. There is consensus, nevertheless, that the ground energy of B19' is stabilized at the presence of shear stresses not larger than 1 GPa [158, 149]. Since we consider deformation twins only, the instability of B19' should not be of concern, because the phase transformations in deformation twins take place under stress anyway.

Fig. 3.3 indicates the corresponding unit cells of the parent B2 austenite phase (left) and the B19' martensite phase (right) before and after the transformation. The cubic unit cell has an edge length ℓ_0 . The Ti atoms are at the corners of the cubic B2 unit cell and the

Phase	Unit cell	ℓ_1 [Å]	ℓ_2 [Å]	ℓ_3 [Å]	θ [deg]	$V_{\mathfrak{A}}$ [Å ³]
Austenite B2	Cubic (BC)	3.015	3.015	3.015	90	27.41
Austenite B2	Tetragonal (FC)	4.264	4.264	3.015	90	54.82
Martensite B19'	monoclinic (almost FC)	4.120	4.622	2.889	96.8	54.63

Tab. 3.1: Lattice parameters of NiTi (49.75 at. % Ni) for B2 (austenite) and B19' (martensite) phases based on experimental results [122, 78, 149, 121]. The non-primitive tetragonal unit cell for B2 (second row) is in correspondence with the monoclinic B19' primitive unit cell.

Ni atoms at their centers. To describe the lattice correspondence between B2 and B19', we start out by choosing an orthorhombic B2 cell whose edges are along $[1\bar{1}0]$, $[110]$, and $[001]$ directions, in cubic austenite coordinates, and consequently have the lengths $\sqrt{2}\ell_0$, $\sqrt{2}\ell_0$, and ℓ_0 respectively. This cell is shared among four adjacent cubic cells and has a volume twice a cubic unit cell. The Ti atoms are at the corners of the orthorhombic unit cell and also at the centers of upper and lower faces. The four Ni atoms at the center of each cubic unit cell now lie at the centers of the four vertical faces of the orthorhombic unit cell, $V_{\mathfrak{A}} = 2\ell_0^3$.

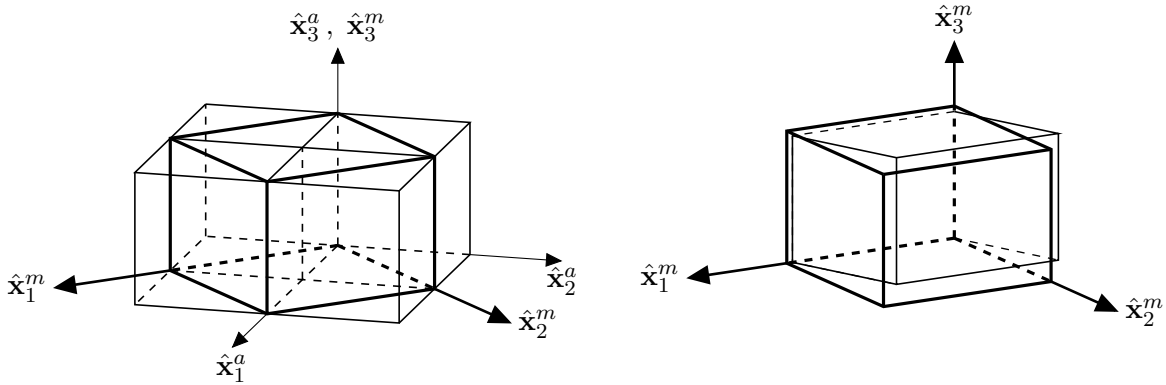


Fig. 3.3: Lattice correspondence between B2 face-centered orthorhombic non-primitive cell (left) and B19' almost-face-centered monoclinic primitive unit cell (right) [122, © permission obtained].

To carry out the geometric mapping from the parent B2 phase to the child B19' phase, we first elongate the orthorhombic cell along each edge to arrive at the lattice constants of the martensite as

$$\ell_1 = \kappa_1 \sqrt{2} \ell_0, \quad \ell_2 = \kappa_2 \sqrt{2} \ell_0, \quad \ell_3 = \kappa_3 \ell_0. \quad (3.67)$$

Finally, we allow the lattice direction $[110]$ to tilt towards $[001]$ by an angle equal to $90^\circ - \theta$, which concludes the formation of the monoclinic B19' primitive unit cell¹⁰. Expressed in the original cubic austenite lattice coordinates the B2 $\rightarrow$ B19' transformation consists of the

¹⁰The complete name for this transformation is “cubic to monoclinic-I”. Alternatively, one could tilt $[110]$ lattice direction towards $[1\bar{1}0]$, which is called “cubic to monoclinic-II transformation”. This type of transformation does not occur in NiTi, so we exclude it from our discussion.

following mappings

$$\begin{aligned}\mathbf{a}_1^a = \ell_0 \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} &\rightarrow \frac{\ell_1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = \mathbf{a}_1^m, & \mathbf{a}_2^a = \ell_0 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} &\rightarrow \frac{\ell_2}{\sqrt{2}} \begin{pmatrix} \sin \theta \\ \sin \theta \\ \sqrt{2} \cos \theta \end{pmatrix} = \mathbf{a}_2^m, \\ \mathbf{a}_3^a = \ell_0 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} &\rightarrow \ell_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \mathbf{a}_3^m.\end{aligned}\tag{3.68}$$

The transformation should then take the following form

$$\mathbf{F} = \{\mathbf{a}^m\} \{\mathbf{a}^a\}^{-1} = \mathbf{a}_1^m \otimes \frac{\mathbf{a}_1^a}{\|\mathbf{a}_1^a\|^2} + \mathbf{a}_2^m \otimes \frac{\mathbf{a}_2^a}{\|\mathbf{a}_2^a\|^2} + \mathbf{a}_3^m \otimes \frac{\mathbf{a}_3^a}{\|\mathbf{a}_3^a\|^2},\tag{3.69}$$

where $\{\mathbf{a}\} = (\mathbf{a}_1^T, \mathbf{a}_2^T, \mathbf{a}_3^T)$ is the matrix whose columns are the lattice vectors. The above formula gives

$$\mathbf{F} = \begin{pmatrix} \frac{1}{2}(\kappa_2 \sin \theta + \kappa_1) & \frac{1}{2}(\kappa_2 \sin \theta - \kappa_1) & 0 \\ \frac{1}{2}(\kappa_2 \sin \theta - \kappa_1) & \frac{1}{2}(\kappa_2 \sin \theta + \kappa_1) & 0 \\ \frac{1}{\sqrt{2}}\kappa_2 \cos \theta & \frac{1}{\sqrt{2}}\kappa_2 \cos \theta & \kappa_3 \end{pmatrix}.\tag{3.70}$$

The right polar decomposition of the above deformation, i.e. $\mathbf{F} = \mathbf{Q}\mathbf{U}$, yields the so called *Bain matrix* or the *transformation matrix*, and its corresponding rotation:

$$\mathbf{U} = \begin{pmatrix} \alpha & \delta & \epsilon \\ \delta & \alpha & \epsilon \\ \epsilon & \epsilon & \gamma \end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix} (\gamma/\kappa_3 + 1)/2 & (\gamma/\kappa_3 - 1)/2 & -\epsilon/\kappa_3 \\ (\gamma/\kappa_3 - 1)/2 & (\gamma/\kappa_3 + 1)/2 & -\epsilon/\kappa_3 \\ \epsilon/\kappa_3 & \epsilon/\kappa_3 & \gamma/\kappa_3 \end{pmatrix}.\tag{3.71}$$

where

$$\begin{aligned}\gamma &= \frac{\kappa_3(\kappa_2 \sin \theta + \kappa_1)}{\sqrt{\kappa_2^2 + \kappa_3^2 + 2\kappa_3\kappa_2 \sin \theta}}, & \alpha &= \frac{1}{2} \left(\frac{\kappa_2(\kappa_3 \sin \theta + \kappa_2)}{\sqrt{\kappa_2^2 + \kappa_3^2 + 2\kappa_3\kappa_2 \sin \theta}} + \kappa_1 \right), \\ \epsilon &= \frac{\kappa_2\kappa_3 \cos(\theta)}{\sqrt{2}\sqrt{\kappa_2^2 + \kappa_3^2 + 2\kappa_3\kappa_2 \sin \theta}}, & \delta &= \frac{1}{2} \left(\frac{\kappa_2(\kappa_3 \sin \theta + \kappa_2)}{\sqrt{\kappa_2^2 + \kappa_3^2 + 2\kappa_3\kappa_2 \sin \theta}} - \kappa_1 \right),\end{aligned}\tag{3.72}$$

We then substitute the lattice parameters from Tab. 3.1 into (3.67) and then (3.72) to obtain

$$\begin{aligned}\kappa_1 &= 0.9663, & \kappa_2 &= 1.084, & \kappa_3 &= 0.9582, \\ \gamma &= 0.9563, & \epsilon &= -0.0427, & \alpha &= 1.0243, & \delta &= 0.058.\end{aligned}\tag{3.73}$$

Putting these values into (3.71) finally gives

$$\mathbf{U} = \begin{pmatrix} 1.0243 & 0.058 & -0.0427 \\ 0.058 & 1.0243 & -0.0427 \\ -0.0427 & -0.0427 & 0.9563 \end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix} 0.999 & -0.001 & 0.0445 \\ -0.001 & 0.999 & 0.0445 \\ -0.0445 & -0.0445 & 0.998 \end{pmatrix}.\tag{3.74}$$

During the construction of the monoclinic martensite from the cubic austenite, we had the choice to decide for alternative orthorhombic cells, due to cubic symmetry. Also, while tilting the orthorhombic cell we could have chosen alternative edges and alternative inclinations. All these choices, which are determined by the symmetries of the parent and the child phase, add up to the existence of different variants of martensite.

We mentioned that since martensitic phase transformation is reversible, the point group of martensite must be a subgroup of the point group of austenite. Otherwise, doing consecutive forward and backward transformations, one could produce infinitely many variants for both. So it was shown in (2.34) that $N_{\text{m.v.}} = \#\mathcal{G}_+(\mathbf{a}_i^a)/\#\mathcal{G}_+(\mathbf{a}_i^m)$. In the case of cubic to monoclinic transformation, $N_{\text{m.v.}} = 24/2 = 12$.

Basically, all the distinct martensite variants are obtained through the same geometrical procedure, except that the lattice directions along which the elongations and the inclination occur are permuted. So in principle we should be able to express the transformation matrices of all variants by applying suitable *elementary row and column operations*. Tab. 3.2 lists all the possible variants with their Bain matrix and their respective elementary operators such that

$$\mathbf{U}_i = \mathbf{E}_i \mathbf{U}_1 \mathbf{E}_i^T, \quad \mathbf{Q}_i = \mathbf{E}_i \mathbf{Q}_1 \mathbf{E}_i^T, \quad (3.75)$$

where $\mathbf{U}_i$ is the Bain matrix of the i th variant, and $\mathbf{E}_i$ is the corresponding elementary row operator. Hence, $\mathbf{E}_i^T$ acts as an elementary column operator. Notice that (3.75)₂ is an immediate consequence of the fact that $\mathbf{E}_i \mathbf{E}_i^T = \mathbf{I}$, for every i .

The compatibility condition of the form

$$\mathbf{R} \mathbf{Q}_j \mathbf{U}_j - \mathbf{Q}_i \mathbf{U}_i = \mathbf{a} \otimes \hat{\mathbf{N}}, \quad (3.76)$$

in which $\mathbf{R} \in \text{SO}(3)$, and $\mathbf{Q}_i \mathbf{U}_i$ is the polar decomposition of the transformation $\mathbf{F}_i$ of the variant i with respect to the parent phase. This determines whether any two variants i and j are capable of forming a twin microstructure and is often rewritten as

$$\mathbf{Q} \mathbf{U}_j - \mathbf{U}_i = \mathbf{b} \otimes \hat{\mathbf{N}}, \quad (3.77)$$

in which $\mathbf{b} = \mathbf{Q}_i^T \mathbf{a}$ and $\mathbf{Q} = \mathbf{Q}_i^T \mathbf{R} \mathbf{Q}_j$. Notice that since $\mathbf{Q}_i$ and $\mathbf{Q}_j$ are in the point group of austenite, they belong to $\text{SO}(3)$, and therefore the tensor $\mathbf{Q}$ is also a rotation, i.e. $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}$.

The requirements for existence of the solution as well as the general solution itself are stated in the following theorem by Ball and James 1987 (see [11, pp. 32] for the proof).

Theorem 2. *The compatibility condition $\mathbf{Q} \mathbf{U}_j - \mathbf{U}_i = \mathbf{b} \otimes \hat{\mathbf{N}}$, with $\mathbf{Q} \in \text{SO}(3)$, $\det \mathbf{U}_i > 0$, and $\det \mathbf{U}_j > 0$, has a solution if and only if $\mathbf{C} := \mathbf{U}_i^{-T} \mathbf{U}_j^T \mathbf{U}_j \mathbf{U}_i^{-1} \neq \mathbf{I}$, and the second largest eigenvalue of $\mathbf{C}$ equals one, i.e. $\lambda_2 = 1$, where $0 < \lambda_1 \leq \lambda_2 \leq \lambda_3$. Then, there exist*

i	$\mathbf{U}_i$	$\mathbf{E}_i$	i	$\mathbf{U}_i$	$\mathbf{E}_i$
1	$\begin{pmatrix} \alpha & \delta & \epsilon \\ \delta & \alpha & \epsilon \\ \epsilon & \epsilon & \gamma \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	2	$\begin{pmatrix} \alpha & \delta & -\epsilon \\ \delta & \alpha & -\epsilon \\ -\epsilon & -\epsilon & \gamma \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$
3	$\begin{pmatrix} \alpha & -\delta & \epsilon \\ -\delta & \alpha & -\epsilon \\ \epsilon & -\epsilon & \gamma \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	4	$\begin{pmatrix} \alpha & -\delta & -\epsilon \\ -\delta & \alpha & \epsilon \\ -\epsilon & \epsilon & \gamma \end{pmatrix}$	$\begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$
5	$\begin{pmatrix} \alpha & \epsilon & \delta \\ \epsilon & \gamma & \epsilon \\ \delta & \epsilon & \alpha \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$	6	$\begin{pmatrix} \alpha & \epsilon & -\delta \\ \epsilon & \gamma & -\epsilon \\ -\delta & -\epsilon & \alpha \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}$
7	$\begin{pmatrix} \alpha & -\epsilon & \delta \\ -\epsilon & \gamma & -\epsilon \\ \delta & -\epsilon & \alpha \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$	8	$\begin{pmatrix} \alpha & -\epsilon & -\delta \\ -\epsilon & \gamma & \epsilon \\ -\delta & \epsilon & \alpha \end{pmatrix}$	$\begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$
9	$\begin{pmatrix} \gamma & \epsilon & \epsilon \\ \epsilon & \alpha & \delta \\ \epsilon & \delta & \alpha \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$	10	$\begin{pmatrix} \gamma & \epsilon & -\epsilon \\ \epsilon & \alpha & -\delta \\ -\epsilon & -\delta & \alpha \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix}$
11	$\begin{pmatrix} \gamma & -\epsilon & \epsilon \\ -\epsilon & \alpha & -\delta \\ \epsilon & -\delta & \alpha \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$	12	$\begin{pmatrix} \gamma & -\epsilon & -\epsilon \\ -\epsilon & \alpha & \delta \\ -\epsilon & \delta & \alpha \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$

Tab. 3.2: Cubic to monoclinic-I transformation matrices, in which $\mathbf{U}_i$ is the Bain matrix, $\mathbf{E}_i$ is the elementary row operator such that $\mathbf{U}_i = \mathbf{E}_i \mathbf{U}_1 \mathbf{E}_i^T$.

two solutions of the form

$$\begin{aligned}
 \hat{\mathbf{N}} &= \frac{\sqrt{\lambda_3} - \sqrt{\lambda_1}}{\nu \sqrt{\lambda_3 - \lambda_1}} \left(-\sqrt{1 - \lambda_1} \mathbf{U}_i^T \mathbf{e}_1 + \bar{\kappa} \sqrt{\lambda_3 - 1} \mathbf{U}_i^T \mathbf{e}_3 \right), \\
 \mathbf{b} &= \nu \left(\sqrt{\frac{\lambda_3(1 - \lambda_1)}{\lambda_3 - \lambda_1}} \mathbf{e}_1 + \bar{\kappa} \sqrt{\frac{\lambda_1(\lambda_3 - 1)}{\lambda_3 - \lambda_1}} \mathbf{e}_3 \right),
 \end{aligned} \tag{3.78}$$

where $\bar{\kappa} = \pm 1$, $\nu \neq 0$ is a suitable normalizing factor for $\hat{\mathbf{N}}$, and $\mathbf{e}_i$'s are the normalized eigenvectors of $\mathbf{C}$ corresponding to eigenvalues λ_i 's.

Here, the tensor $\mathbf{C}$ expresses the relative deformation of variant j with respect to variant i . Therefore, the condition $\lambda_2 = 1$ guarantees that there is a flat plane shared by the two variants that remains undistorted. This ensures compatibility of the two variants. If the existence conditions are fulfilled, there are two solutions of the twinning equation corresponding to values $\bar{\kappa} = \pm 1$. Each of the two solutions has a particular geometrical interpretation,

which calls for a precise geometrical interpretation of the twinning itself.

Remark 8. When two compatible solid phases form a twin, the lattice on each side can be obtained by either

- a simple shear, or
- a rotation

of the lattice on the other side. For details and proofs see [127, pp. 676–679] and [25, pp. 68]. The simple shear here can be represented as

$$s = \|\mathbf{b}\| \left\| \mathbf{U}_i^{-1} \hat{\mathbf{N}} \right\|, \quad \hat{\mathbf{s}} = \frac{\mathbf{b}}{\|\mathbf{b}\|}, \quad \hat{\mathbf{n}} = \frac{\mathbf{U}_i^{-1} \hat{\mathbf{N}}}{\left\| \mathbf{U}_i^{-1} \hat{\mathbf{N}} \right\|}, \quad (3.79)$$

where s denotes shear, $\hat{\mathbf{s}}$ shows the shear direction, and $\mathbf{n}$ is the shear plane. These three quantities are often collectively called *twinning elements*. The above mentioned rotation originates from the symmetry connection between the two variants, as mentioned before: $\mathbf{U}_j = \mathbf{R}^T \mathbf{U}_i \mathbf{R}$ where $\mathbf{R} \in \mathcal{G}_+(\mathfrak{a}_i^\circ)$. The rotation that maps the lattice of variant i to the lattice of its twin variant j is of the form $\mathbf{Q}_{ji} = \mathbf{R} \mathbf{Q}_j \mathbf{R}^T$. For every two compatible variants i and j , the two solutions of the twinning equation (3.77) given by (3.78) fall under one of the following classes:

- **Type I twins.** The twinning plane $\mathbf{n}$ is rational in the austenite lattice coordinates and is therefore a symmetry plane of the parent phase.
- **Type II twins.** The shear direction $\hat{\mathbf{s}}$ is rational in the austenite lattice coordinates and is therefore a symmetry direction of the parent phase.
- **Compound twins.** Both the shear direction $\hat{\mathbf{s}}$ and the twinning plane $\mathbf{n}$ are rational in the austenite lattice coordinates, and both are symmetric in the parent phase. This happens when the transformations are related to each other by a 180° rotation $\mathbf{R}$ through $\mathbf{F}_j^T \mathbf{F}_j = \mathbf{R}^T \mathbf{F}_i^T \mathbf{F}_i \mathbf{R}$. This is known as Mallard's law [129, pp. 285].

If $\mathbf{n}$ and $\mathbf{a}$ of two distinct twin systems are related through a symmetry of the parent phase, e.g. cubic symmetry for NiTi, they belong to the same *twinning mode*. Each twinning mode comprises either one type I twin together with one type II twin or two compound twins.

There are listed solutions of all twin-compatible martensitic variants of NiTi in various references [129, 122, 25], which are given directly based on (3.77) instead of (3.76). This does not influence $\hat{\mathbf{N}}$, as can be seen from (3.77), but we have to take into account that $\mathbf{a} = \mathbf{Q}_i^T \mathbf{b}$ with respect to the lattice coordinates of the j th variant. We emphasize this because in computing the interface energy (3.108) we need to follow the exact correspondence between lattices in order to determine the lattice interplanar spacing and the shear direction. In order to avoid the hassle of calculating planar atom density and interplanar spacing for non-rational lattice planes, we calculate the interface energy for type I twins only, and assume that the corresponding type II twin has roughly the same interface energy. Although this is circumventing the problem rather than solving it, this assumption is not entirely unintuitive

Mode	Type	$\mathbf{b}$	$\hat{\mathbf{N}}$	s	$D_{\mathbf{n}}^a [\text{\AA}]$	$\varrho_{\mathbf{n}}^m [\text{\AA}^{-2}]$
A	comp.	$(-0.1602, -0.1602, 0.01263)$	$(0, 0, -1)$	0.2385	3.015	0.1050
	comp.	$(-0.01143, -0.01143, 0.2562)$	$\frac{1}{\sqrt{2}}(1, 1, 0)$	0.2385		
B	I	$(0.2274, 0.01969, -0.1709)$	$(0, -1, 0)$	0.2804	3.015	0.1122
	II	$(0.01637, 0.2890, -0.01204)$	$(-0.8113, 0, 0.5846)$	0.2804		
C	I	$(0.2835, 0.1160, 0.08138)$	$\frac{1}{\sqrt{2}}(0, -1, 1)$	0.3096	2.132	0.1598
	II	$(0.02148, 0.2276, -0.2131)$	$(-0.9072, -0.2975, -0.2975)$	0.3096		
D	I	$(0.03755, 0.09447, -0.08812)$	$\frac{1}{\sqrt{2}}(0, -1, -1)$	0.1422	2.132	0.1477
	II	$(0.001744, 0.1114, 0.1037)$	$(-0.3306, -0.6674, 0.6674)$	0.1422		

Tab. 3.3: Solutions of twinning equation for all twinning modes of cubic to monoclinic-I transformation of NiTi. Austenite lattice coordinates are used.

Variant	1	2	3	4	5	6	7	8	9	10	11	12
1	–	A	B	B	C	–	–	D	C	–	D	–
2	A	–	B	B	–	C	D	–	–	C	–	D
3	B	B	–	A	–	D	C	–	D	–	C	–
4	B	B	A	–	D	–	–	C	–	D	–	C
5	C	–	–	D	–	B	A	B	C	D	–	–
6	–	C	D	–	B	–	B	A	D	C	–	–
7	–	D	C	–	A	B	–	B	–	–	C	D
8	D	–	–	C	B	A	B	–	–	–	D	C
9	C	–	D	–	C	D	–	–	–	B	B	A
10	–	C	–	D	D	C	–	–	B	–	A	B
11	D	–	C	–	–	–	C	D	B	A	–	B
12	–	D	–	C	–	–	D	C	A	B	B	–

Tab. 3.4: Twin compatibility of NiTi martensite variants identified by their twinning mode.

because type I and type II twins in each twinning mode have equal shears. This shall roughly translate to equivalent change of interatomic distances across the interface. On the other hand, this simplification is justified because the exact crystallography of type II twins is still open [104, 105]. Therefore, a methodical calculation of their interface energy goes beyond the scope of our work.

3.4 Interface energies

A solid interface is defined either as a *sharp boundary* or as a *transition layer* that joins two bulks of solid which are distinguished by their different physical properties. The two solids can be made of different materials, or different phases of the same material in which case the interface is often called a phase boundary, or they can pertain to variants of the same phase which again can be seen as a phase boundary of sorts. A closely related concept

is that of a *solid surface*, which differs from a solid interface in that it separates a solid bulk from its surroundings—be it vacuum or a fluid phase. In what follows, the two terms might be used interchangeably since the discussion applies equivalently to both interfaces and surfaces, unless stated otherwise.

The two neighboring solid phases, sharing an interface, may or may not be in equilibrium. Nonequilibrium interfaces might move, exchange energy or matter, or they may even be chemically reactive. These may be accompanied by interface *formation*, interface *deformation*, or a combination of both. Equilibrium interfaces, on the other hand, are static and do not exchange energy or matter. This is the case when the two mediums that share the interface are in thermodynamic equilibrium with each other and with the interface, which in turn implies that all the pertinent thermodynamic potentials are equal among the two enclosing mediums and the interface itself. Here in this work, since it is postulated that the microstructure evolves quasi-statically, all martensite twin interfaces are assumed to be at equilibrium.

We gave two distinct geometrical interpretations of an interface, namely a sharp interface versus a transition layer. A sharp interface is a two dimensional entity and so has no volume. Whereas a transition layer is extended in three dimensions but it has a thickness considerably smaller than the overall characteristic span of the enclosing mediums along the normal to the interface. Although equivalent in most circumstances, these two views have at least one substantial difference when it comes to establishing the governing equations in terms of *intensive* and *extensive* thermodynamic properties of the system.

If the interface is treated as a thin transition layer, the governing equations keep their usual form, since every extensive property of the overall system equals the sum of that property over the three mediums: the two adjacent bulks plus the transition layer joining them. Most prominently, the volume and the free energy of the system are

$$V = V_- + V_0 + V_+, \quad \Psi = \Psi_- + \Psi_0 + \Psi_+, \quad (3.80)$$

where plus, minus, and zero subscripts designate the two mediums and the interface respectively. Assuming a homogeneous distribution, it follows that the energy densities are related to the total free energy by

$$\Psi = V_- \psi_- + V_0 \psi_0 + V_+ \psi_+. \quad (3.81)$$

But in reality the situation is not nearly as ideal. In general, the free energy density as well as any other intensive quantity corresponding to an extensive quantity is a field function. If the curvature of the interfacial transition layer is large compared to its thickness, it is possible to approximate the interface by a tangent plane at its close vicinity (see [52, 111]). Having done so, we can express the interfacial free energy as the *excess* free energy stored at the interface:

$$\gamma = \frac{1}{A_0} \left(\int_{z_-}^{z_+} \psi \, dV - V_- \psi_- - V_+ \psi_+ \right), \quad (3.82)$$

which, most noticeably, is defined as free energy per unit area of the interface. This way of defining the interfacial energy facilitates a seamless transition to thermomechanical formula-

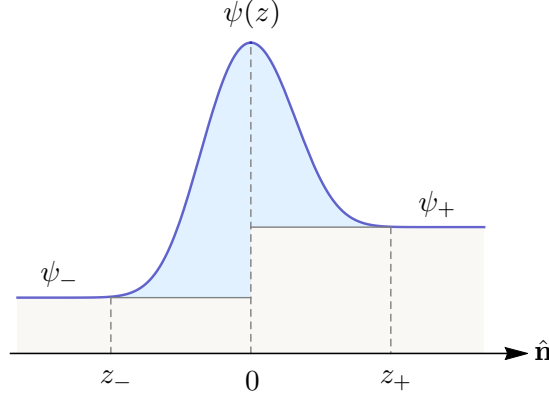


Fig. 3.4: Schematic distribution of an extensive quantity Ψ with volume density ψ across a thin interface layer. The light blue filling shows the excess.

tion of sharp interfaces; although, it conceals the variations in the thickness of the interface layer, which is largely irrelevant to formulation of the interfaces anyway, as will be seen.

For a sharp interface, the equation (3.81) reduces to $\Psi = V_- \psi_- + V_+ \psi_+$, because $V_0 = 0$. By itself, this means the interface does not carry energy which cannot be true. Otherwise, the bulk would form as many interfaces as possible, rather than maintaining continuity, in an effort to maximize its entropy; which is contrary to what happens in reality where a single phase free from external interactions tends to minimize its surface and interface area. From a purely theoretical standpoint, an argument can be made based on the fact that a thermodynamic system at equilibrium attains minimum energy while simultaneously its entropy is maximized. Among the configurations of a system that have equal energy, the one with more microscopic degrees of freedom possesses a higher entropy. To represent a solid body that forms an interface, more degrees of freedom are needed at the microscale; namely, at least the locations of the interfaces. Thus, a solid body in a continuous configuration has lower entropy than the same body with the same energy, but with interfaces. It follows that the formation of interfaces has to cost energy.

Remark 9. From a thermomechanical viewpoint, defining the interface energy as the excess energy means that the interface energy alone is at a minimum because the overall potential attains its minimum due to equilibrium condition: $\Psi = \Psi_- + \Gamma + \Psi_+ \rightarrow \min$, while simultaneously each of the two solid phases, being at equilibrium, are also at their minimum state of energy: $\Psi_- \rightarrow \min$, and $\Psi_+ \rightarrow \min$. It follows that the interface energy must attain its minimum: $\Gamma \rightarrow \min$; which guarantees that the interface is thermodynamically stable, and the interface deformation can be modeled as a hyperelastic medium the same way as the bulk.

The disappearance of the interface term from (3.81) is an apparent case of singularity of the formalism. However, we are readily prepared to resolve the situation. Based on the above arguments and particularly the equation (3.82), the interface energy is identified as the excess energy with respect to the sum of energies stored in the bulk. This immediately

reformulates (3.81) to

$$\Psi = V_- \psi_- + A_0 \gamma + V_+ \psi_+ . \quad (3.83)$$

Analogous to ψ , the interfacial free energy density is a field function of temperature, deformation gradient, and amount of substance:

$$\gamma = \gamma(\theta_0, \mathbf{F}_0, N_0; \mathbf{X}) . \quad (3.84)$$

Its changes at a fixed position $\mathbf{X}$ are therefore obtained from the usual thermodynamic derivatives, as

$$d\gamma = \frac{\partial \gamma}{\partial \theta_0} d\theta_0 + \frac{\partial \gamma}{\partial \mathbf{F}_0} : d\mathbf{F}_0 + \frac{\partial \gamma}{\partial N_0} dN_0 = -s_0 d\theta_0 + \mathbf{P}_0 : d\mathbf{F}_0 + \mu_0 dN_0 , \quad (3.85)$$

where s_0 is the interfacial entropy density, $\mathbf{P}_0$ is the first Piola–Kirchhoff *surface stress* (force per unit length), and μ_0 is the interfacial chemical potential density—all in the reference configuration. Then, the question comes to mind on how to rewrite (3.85) in the current configuration, viz in terms of $\gamma = \gamma(\theta_0, \mathbf{F}_0, N_0; \mathbf{x})$. The apparent connection between the energy densities in the two configurations is $\gamma(\cdot; \mathbf{X}) = J\gamma(\cdot; \mathbf{x})$, in which $J = \det \mathbf{F}_0$. A term by term reformulation of the differentials in (3.85) gives $\partial \gamma(\cdot; \mathbf{X}) / \partial \theta_0 = J \partial \gamma(\cdot; \mathbf{x}) / \partial \theta_0$ and $\partial \gamma(\cdot; \mathbf{X}) / \partial N_0 = J \partial \gamma(\cdot; \mathbf{x}) / \partial N_0$. In the case of deformation energy we obtain a somewhat different result, namely

$$\mathbf{P}_0 = J \boldsymbol{\tau} \mathbf{F}_0^{-T} = J \gamma(\cdot; \mathbf{x}) \mathbf{F}_0^{-T} + J \frac{\partial \gamma(\cdot; \mathbf{x})}{\partial \mathbf{F}_0} , \quad (3.86)$$

where $\boldsymbol{\tau}$ is the interfacial Cauchy stress (real stress) and has the dimensions of force per unit length. Multiplying both sides of the above equation with $J^{-1} \mathbf{F}_0^T$ from the right, simplifies it further to

$$\boldsymbol{\tau} = \gamma(\cdot; \mathbf{x}) \mathbf{I} + \frac{\partial \gamma(\cdot; \mathbf{x})}{\partial \mathbf{F}_0} \mathbf{F}_0^T . \quad (3.87)$$

Considerable body of research has been dedicated to equation (3.87), often called the *Shuttleworth* equation. Most references interpret the first term on the right hand side as the energy required to **form** an interface, and the second term as the energy required to **de-form** the already created interface. This interpretation seems intuitively plausible from a physical standpoint; however, it may violate the assumption that the free energy density was initially calculated per unit reference volume. As simple as this equation might seem, there is hardly any consensus on validity of the assumptions on which it is established (see for instance [52]). We confine our formulation to the reference configuration based on (3.85) which, in any case, shall remove any shortcomings that might arise. Further, regardless of these issues, a fully extended formulation would have little relevance to the present work, mostly because we neglect the thermal and deformation effects on all interfaces.

Remark 10. In continuum mechanical study of fluids, γ is often referred to as *surface tension*; because fluids cannot statically undergo shear stress and also interface cohesion in fluids is not strong enough to allow for any significant interface deformation without interface formation, which means the first term in (3.87) vanishes and $\boldsymbol{\tau} = \gamma(\cdot; \mathbf{x}) \mathbf{I}$. In

solids, on the other hand, *surface stress* is distinguished from surface energy; because neither of the two reasons apply to solid interfaces. Thus, γ is referred to as interface energy and τ as interface stress, in order to emphasize that the interfacial energy in solids comprises both formation and deformation energies.

A brief look at the literature

There is extensive body of research dedicated to the study of solid interfaces, in general, and martensitic twin interfaces, in particular. We briefly mention those that we have encountered and have also benefited from.

Fischer et al [52], and Muller et al [111] give a relatively comprehensive and recent review of continuum thermomechanics of solid interfaces. Sonderegger and Kozeschnik [142] gives a general algorithm for counting the effective number of broken bonds of nearest neighbors of any order. For a common atomistic model based on two- and multi-body interatomic potentials see [56]. Hackl et al [66] study the effect of interface energy in the energy of martensite laminate based on a heuristic approach and homogenization. We also have benefited from the works of Wen et al [154], Vitos et al [148], Wang et al [153], Zhang et al [157], and Galanakis et al [55]. Last but not least, we also have extended and built upon our previous works [61, 60, 62].

The various methods of estimating or computing solid interface energy roughly belong to either of three categories, namely

- Method of “nearest neighbor broken bond”: is based on lattice structure and bond energies.
- “Ab initio” (first principles) computations: relying on particular interatomic potentials, these methods simulate the lattice kinetic from which one can obtain the thermodynamic quantities of interest.
- Continuum models: which are loosely based on lattice structure and mostly based on fundamental thermodynamic principles.

While all the above approaches somehow introduce material parameters that account for various aspects of the underlying physical system, it is generally preferred that these parameters be simple and directly measurable, if possible. In other words, we prefer simple quantities like lattice constants over complex phenomenological attributes like heuristically introduced coefficients. Because simple quantities have dimensions that we can easily make sense of and measure, whereas the phenomenological coefficients have to be estimated indirectly by fitting the computational results to the experimental data.

3.4.1 Energy of coherent crystal interfaces

Crystallographically, solid interfaces can be classified as coherent, semi-coherent, and incoherent, depending on how closely they resemble a perfect crystal *monolayer* (Fig. 3.5). In this classification, a coherent interface is the sharpest physically possible interface, because

by its definition a coherent interface is a perfect crystal monolayer that is shared by two adjacent monocrystalline solid phases. The sharpness of a coherent interface is only limited by lattice vibrations which at high temperatures become so significant that the interface leans toward a *diffuse interface*. This, however, is not pertinent to our model, because shape memory materials manifest their interesting behaviors at around transformation temperature and below, which is not high enough to make a coherent twin interface markedly diffuse. We therefore assume that the martensitic twin interfaces are perfectly sharp.

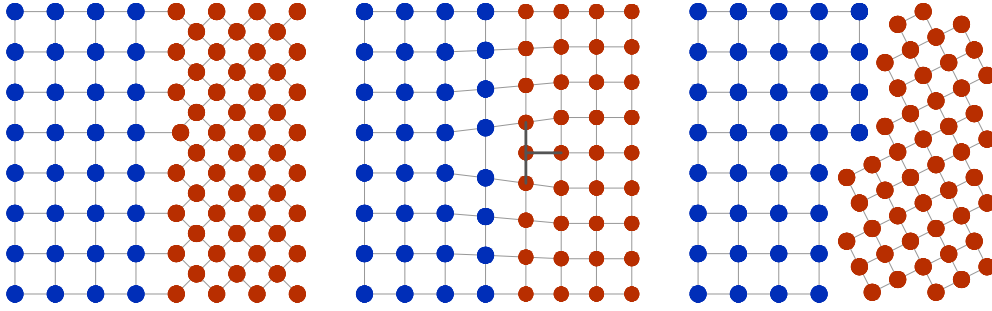


Fig. 3.5: An illustration of a coherent (left), semicoherent (middle), and incoherent interface.

We mentioned before that the accommodation strains are small compared with the transformation strains. Hence, the interface deformation energy is negligible in comparison to the interface formation energy, and consequently the twin interface energy is independent of interface strains. Also, the thermal effects are excluded from our model, so the changes in γ caused by temperature variations are left out. Putting all these assumptions together, we conclude that the twin interface energy is fully determined by crystallography of the two adjacent martensite variants, which itself is fully expressed in terms of transformation strains. So we will once again rely on the crystallographic theory of martensitic transformation, seeking to understand how it determines the twin interface energy as a thermodynamic state function.

As usual, based on the Cauchy–Born hypothesis we make the connection between the variations of interatomic distances and the corresponding thermodynamic state variable, i.e. the transformation strains. For illustration purposes only, Fig. 3.6 shows a cubic to tetragonal transformation that is capable of forming martensitic twins. However, the following arguments are general and remain equally valid for other types of twin-compatible martensitic transformations with any possible twinning plane. We again postulate that all martensitic transformations must remain within the Ericksen–Pitteri neighborhood to avoid plastic deformation and slip, in order to eliminate any possible degeneracy of the outcoming model.

Atomistic modeling of crystalline materials is based on a stored energy in the form of the sum of interatomic potentials over contributions of all the atomic bonds in a representative element under periodic boundary conditions. Depending on the type of chemical bonds between the constituent atoms, the interatomic potentials take the form of two-, three-, or four-body-potentials, or a combination of these. For instance in a crystal system with ionic bonds the major contribution to the interatomic potential comes from electrostatic

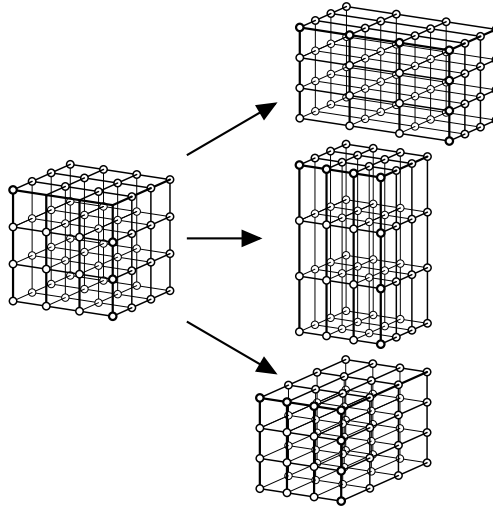


Fig. 3.6: The three distinct variants of the cubic to tetragonal martensitic transformation [161, © permission obtained].

interactions which are hence two-body potentials that decay as $1/r$, where r is the interatomic distance. This translates to interatomic forces proportional to $1/r^2$ which are therefore long-range. Metallic crystalline, on the other hand, form metallic bonds which are much the same as ionic bonds, but since the outermost (valence) electrons are shared among the neighboring atoms, the interatomic forces act over much shorter distances than the ionic bonds. So, while the interatomic potentials corresponding to metallic crystalline are often considered to be two-body potentials, they decay much quicker than $1/r$. A common example is the so called *Lennard-Jones* potential, given by

$$\psi(r) = \varepsilon \left(\left(\frac{r_m}{r} \right)^{12} - 2 \left(\frac{r_m}{r} \right)^6 \right). \quad (3.88)$$

Despite the fact that more precise representations of the metallic bond are available, the Lennard-Jones potential is widely employed in lattice computations, mainly due to its simplicity which lowers the computation cost while still yielding satisfactory results in most circumstances.

Calculation of solid interface energy based on crystal structure and interatomic potentials is certainly not a recent finding [20] and spans a variety of methods. The basic idea behind many of these methods comes from the Gibbs interpretation of extensive surface quantities as the *excess*. Namely, one usually starts from homogeneous solid phases and conceives a set of geometric operations on their lattices which lead to formation of the intended interface. In each of these operations, one keeps track of the variations of the interatomic energies. Finally, subtracting the energy of the initial lattice from that of the final lattice yields the excess energy, which by definition equals the interface energy.

Many of the above techniques follow the method of “nearest neighbor broken bond”, NNBB. The idea is to calculate the energy required to break all the bonds along a given lattice plane in two different solid phases to create two solid surfaces which will later join to create a coherent interface. In joining the two solid surfaces, the released bonding energy is subtracted

from the bond breaking energies to yield the interface energy: $\Gamma = \Psi_b - (\Psi_{s1} + \Psi_{s2})$. For a coherent interface between two solid phases, the interface energy density can be then written as

$$\gamma = n_0 z_0 \left(E_{12} - \frac{1}{2} (E_{11} + E_{22}) \right), \quad (3.89)$$

where n_0 denotes the number of atoms per unit interface area, z_0 is the number of broken bonds per interface atom, and E_{ij} is the bond energy between two atoms in phases i and j . This can be generalized to account for neighboring atoms farther than the nearest neighbor, in the case of a crystalline whose interatomic potentials act over larger distances, such as ionic lattices [142].

Another broad family of methods is directly based on atomic interactions. Here, the overall morphology of the lattice as well as proper interatomic potentials are the key ingredients. The sum of interatomic potentials is computationally minimized over a set of predesignated atomic degrees of freedom within a representative volume element under periodic boundary conditions. The lattice energy is usually decomposed into a sum of short-range and long-range potentials [51, 147, 56, 29, 102]. Concerning the degrees of freedom, only the atoms belonging to a few monolayers parallel to the interface are allowed to *relax*, i.e. deviate from their original lattice position, and the monolayers farther from the interface are assumed to be fixed, because they remain unaffected from the asymmetry caused by the interface. Thus, the lattice energy of a solid with an interface is formulated as

$$E = \min_{\mathbf{r} \in \mathcal{I}} \sum_i \left(\frac{1}{2} \sum_{j \neq i}^{\mathcal{I}} \psi(|\mathbf{r}_{ij}|^2) + \sum_k^{\mathcal{B}} \psi(|\mathbf{r}_{ik}|^2) \right), \quad (3.90)$$

where $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ denotes the relative position of the i th atom with respect to the j th atom; $\mathcal{I}$ denotes the set of atoms within a thin layer containing the interface, and $\mathcal{B}$ is the set of atoms within the bulk comprising two layers which enclose the interface layer. For solid interfaces (not surfaces) when the temperature is fixed the kinetic energy of the lattice can be left out, since it does not yield any excess energy. Notice in (3.90) that the potential accounts for the interface excess energy only, since the contribution of the bulk $E_{\mathcal{B}} = \frac{1}{2} \sum_i^{\mathcal{B}} \sum_{j \neq i}^{\mathcal{B}} \psi(|\mathbf{r}_i - \mathbf{r}_j|^2)$, has been already subtracted from the overall potential.

Although most authors express the general pairwise interatomic potential as $\psi(|\mathbf{r}_i - \mathbf{r}_j|)$, we adopt the alternative form $\psi(|\mathbf{r}_i - \mathbf{r}_j|^2)$, following Born and Huang (1967) [31, pp. 129–133]. While the two forms are mathematically equivalent, we find the Born formalism to be more convenient in reflecting the lattice symmetries (see Appendix A).

Twin interface energy: the atomistic–continuum link

In order to incorporate the twin interface energy into our continuum micromechanical model, a link has to be found between the atomistic mechanisms of interface formation and the thermomechanical variables that describe the microstructure, namely Λ , H , $\hat{\mathbf{N}}$, and $\mathbf{a}$. To this end, we first establish a hypothetical geometric transformation that leads to formation of a coherent martensitic interface (Fig. 3.7). We do so by starting from a homogeneous

solid phase of a single martensite variant with transformation $\mathbf{F}_1$, which occupies a domain large enough to be initially free from boundary effects. For this variant, there exist possible twinning planes whose orientation can be denoted by their respective unit normal vector $\hat{\mathbf{N}}$. We partition the domain along one of the twinning planes into two subdomains, and then deform one subdomain under the deformation gradient $\mathbf{F}_2\mathbf{F}_1^{-1}$; given that $\mathbf{F}_2$ is twin-compatible with $\mathbf{F}_1$ as in (2.16), viz $\mathbf{F}_2 - \mathbf{F}_1 = \mathbf{a} \otimes \hat{\mathbf{N}}$. This sequence of geometric operations forms the desired configuration with a coherent interface. We now look into the atomistic potentials corresponding to this configuration.

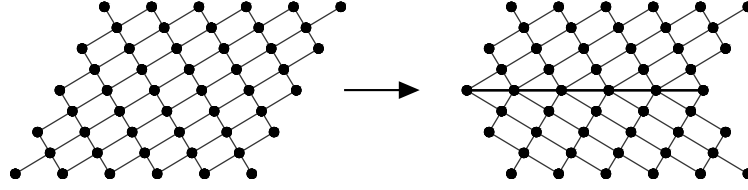


Fig. 3.7: A hypothetical mapping from one tetragonal variant of martensite to another that leads to formation of a compatible interface.

Remark 11. Contrary to the calculation of strain energy in the bulk where the transformation strains had to be expressed with respect to the lattice orientation of the parent austenite, here only the orientation of the martensitic variants relative to each other is important because the overall orientation of twin interface does not alter its formation energy due to objectivity. Instead, we may use the twinning equation in terms of transformation stretches, which are symmetric and positive-definite and are obtained from polar decomposition of the deformation gradients corresponding to the martensitic transformations according (2.37), as $\mathbf{Q}\mathbf{U}_2 - \mathbf{U}_1 = \mathbf{a} \otimes \hat{\mathbf{N}}$.

Following the aforementioned approach, we now confine our attention to the two lattice monolayers adjacent to the interface because the atoms farther from the interface remain almost uninfluenced by its presence. The reason for this is the following: we exclusively deal with metallic substances for which the interatomic forces are typically short-range. For instance based on the Lennard–Jones potential, here is how interatomic forces and potentials compare between the immediate neighbors of an atom and its non-immediate neighbors:

$$\frac{\psi(2r_m)}{\psi(r_m)} = 0.031, \quad \frac{\partial\psi/\partial r|_{1.9r_m}}{\partial\psi/\partial r|_{0.9r_m}} = -0.006. \quad (3.91)$$

Now that we have limited the lattice potential to the interface monolayer and the two monolayers adjacent to it, we could allow the adjacent monolayers to relax, i.e. deviate from their original lattice positions. However, this would require additional thermomechanical variables in the continuum formulation which would inevitably make for a more complex mathematical formulation. We therefore neglect the relaxation effects, which means that the lattice points are fixed at their respective lattice sites before and after transformation. Let us designate the i th lattice site corresponding to variants $\mathbf{F}_1$ and $\mathbf{F}_2$ respectively by $\mathbf{r}_{i[1]}$

and $\mathbf{r}_{i[2]}$. The lattice energy can be then expressed as

$$\begin{aligned}
 E_0 = & \sum_i^{N_I} \sum_j^{N_{II}} \psi(|\mathbf{r}_{i[1]} - \mathbf{r}_{j[2]}|^2) + \frac{1}{2} \sum_i^{N_I} \sum_{j \neq i}^{N_I} \psi(|\mathbf{r}_{i[1]} - \mathbf{r}_{j[1]}|^2) + \frac{1}{2} \sum_i^{N_{II}} \sum_{j \neq i}^{N_{II}} \psi(|\mathbf{r}_{i[2]} - \mathbf{r}_{j[2]}|^2) \\
 & - \sum_i^{N_I} \sum_j^{N_{II}} \psi(|\mathbf{r}_{i[1]} - \mathbf{r}_{j[1]}|^2) - \frac{1}{2} \sum_i^{N_I} \sum_{j \neq i}^{N_I} \psi(|\mathbf{r}_{i[1]} - \mathbf{r}_{j[1]}|^2) - \frac{1}{2} \sum_i^{N_{II}} \sum_{j \neq i}^{N_{II}} \psi(|\mathbf{r}_{i[1]} - \mathbf{r}_{j[1]}|^2)
 \end{aligned} \quad (3.92)$$

where N_I denotes the number of atoms in the subdomain I including those on the interface, and N_{II} is the number of atoms in subdomain II adjacent to the interface. Since the interface monolayer is not subjected to deformation under $\mathbf{F}_2 \mathbf{F}_1^{-1}$, there is no need to count its atoms separately, which is why we may include the interface atoms in N_I . The above equation is directly simplified to

$$\begin{aligned}
 E_0 = & \sum_i^{N_I} \sum_j^{N_{II}} \left\{ \psi(|\mathbf{r}_{i[1]} - \mathbf{r}_{j[2]}|^2) - \psi(|\mathbf{r}_{i[1]} - \mathbf{r}_{j[1]}|^2) \right\} + \\
 & + \frac{1}{2} \sum_i^{N_{II}} \sum_j^{N_{II}} \left\{ \psi(|\mathbf{r}_{i[2]} - \mathbf{r}_{j[2]}|^2) - \psi(|\mathbf{r}_{i[1]} - \mathbf{r}_{j[1]}|^2) \right\}. \quad (3.93)
 \end{aligned}$$

We can simplify this equation even further, knowing that the two martensite variants have equal ground energies. This makes the second sum in (3.93) vanish. Thus, we obtain

$$E_0 = \sum_i^{N_I} \sum_j^{N_{II}} \left\{ \psi(|\mathbf{r}_{i[1]} - \mathbf{r}_{j[2]}|^2) - \psi(|\mathbf{r}_{i[1]} - \mathbf{r}_{j[1]}|^2) \right\}. \quad (3.94)$$

In this equation we distinguish two types of interactions, namely

1. between the atoms at the interface and the adjacent monolayer in II, and
2. between the atoms at the two monolayers adjacent to the interface in I and II.

This way, we can rewrite the summation in (3.94) as $E_0 = \sum_i^{N_0} \sum_j^{N_+} \bullet + \sum_i^{N_-} \sum_j^{N_+} \bullet$. We remember, however, that $\mathbf{F}_1$ and $\mathbf{F}_2$ are symmetry connected via $\mathbf{F}_2 = \mathbf{R}^T \mathbf{F}_1 \mathbf{R}$, where $\mathbf{R}$ belongs to the point group of the parent austenite. Therefore, the interatomic distances in the two variants are equal and their crystallographic difference comes down to a mere rotation. It can be then concluded that the first sum must be identically zero. So in evaluating the sum in (3.94), we only consider the interactions between the atoms residing on the two monolayers adjacent to the interface:

$$E_0 = \sum_i^{N_-} \sum_j^{N_+} \left\{ \psi(|\mathbf{r}_{i[1]} - \mathbf{r}_{j[2]}|^2) - \psi(|\mathbf{r}_{i[1]} - \mathbf{r}_{j[1]}|^2) \right\}. \quad (3.95)$$

In order to carry out the summation in the above equation, we need to calculate

- the number of nearest neighbors of each atom,

- the atom density of each monolayer, and
- the changes of the interatomic distances between the two monolayers adjacent to the interface under the transformation $\mathbf{F}_2\mathbf{F}_1^{-1}$.

Depending on the number of nonzero Miller indices of the monolayers, each atom at the interface may have 1, 2, or 3 immediate neighbors on each adjacent monolayer. This means that each atom on the monolayer adjacent to the interface in subdomain I can have 1, 3, or 6 second-nearest neighbors¹¹ on the monolayer in subdomain II adjacent to the interface. To explain this, we take on the most difficult case; namely, when the twinning plane has 3 nonzero Miller indices, say $(l_1 l_2 l_3)$, which is to say that each interface atom has 3 immediate neighbors on each of its two adjacent monolayers. For the given lattice vectors $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$, these atoms can be listed as

$$\begin{array}{lll} -\mathbf{a}_1/l_1, & -\mathbf{a}_2/l_2, & -\mathbf{a}_3/l_3, & \text{on monolayer } \mathbf{F}_1, \\ \mathbf{a}_1/l_1, & \mathbf{a}_2/l_2, & \mathbf{a}_3/l_3, & \text{on monolayer } \mathbf{F}_2. \end{array} \quad (3.96)$$

For each atom residing on the monolayer in I, we seek to list all its *distinct* second-nearest neighbors located on the monolayer in II across the interface. These are:

$$\begin{array}{lll} \mathbf{a}_1/l_1 + \mathbf{a}_1/l_1, & \mathbf{a}_1/l_1 + \mathbf{a}_2/l_2, & \mathbf{a}_1/l_1 + \mathbf{a}_3/l_3 \\ \circlearrowleft & \mathbf{a}_2/l_2 + \mathbf{a}_2/l_2, & \mathbf{a}_2/l_2 + \mathbf{a}_3/l_3, \\ \circlearrowleft, & \circlearrowleft & \mathbf{a}_3/l_3 + \mathbf{a}_3/l_3. \end{array} \quad (3.97)$$

The case of 2 and 1 nonzero Miller indices should be obvious by now.

As the coherence condition demands, the *atom density* on all crystal monolayers parallel to the interface is the same and is equal to

$$\varrho_{(l_1 l_2 l_3)} = n_{(l_1 l_2 l_3)} d_{(l_1 l_2 l_3)} / V_{\mathfrak{A}}, \quad (3.98)$$

where $n_{(l_1 l_2 l_3)}$ is the number of coplanar atoms in the unit cell of the monolayer $(l_1 l_2 l_3)$, and

$$d_{(l_1 l_2 l_3)} = 1 / |l_1 \mathbf{a}^1 + l_2 \mathbf{a}^2 + l_3 \mathbf{a}^3|, \quad (3.99)$$

is the distance between two adjacent parallel monolayers, in which $\{\mathbf{a}^1, \mathbf{a}^2, \mathbf{a}^3\}$ are the reciprocal lattice vectors. The volume of the unit cell

$$V_{\mathfrak{A}} = \mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3), \quad (3.100)$$

is the same for the two variants since $\det \mathbf{F}_1 = \det \mathbf{F}_2$. Along the same line of reasoning, due to the lattice correspondence between each martensite variant and the common parent austenite, based on the coherence condition we can conclude that the atomic densities of the two monolayers parallel to the interface are equal, and so are the distances between the

¹¹See [142] and [72] for a recent account of interface crystallography, particularly a general method for calculating the number of interface atoms. Also, see the original work of [116] for a precise view of crystallographic neighborhoods.

interface monolayer and the adjacent monolayer in each martensite variant. It follows that the interfacial lattice potential in (3.95) can be written as

$$E_0 = \varrho_{(l_1 l_2 l_3)} A_0 \sum_{i=1}^3 \sum_{j=i}^3 \left\{ \psi(|\underline{\mathbf{a}}_i + \mathbf{T}\underline{\mathbf{a}}_j|^2) - \psi(|\underline{\mathbf{a}}_i + \underline{\mathbf{a}}_j|^2) \right\}, \quad (3.101)$$

where A_0 is the area of the interface monolayer; $\mathbf{T} = \mathbf{F}_2 \mathbf{F}_1^{-1}$ is the deformation gradient that we formerly introduced; and $\underline{\mathbf{a}}_i = \mathbf{a}_i / l_i$, is a mere abbreviation of the vectors in equation (3.97). Notice that the index j of the inner sum starts from i , to avoid duplicate counts of the second neighbors, as also shown in (3.97).

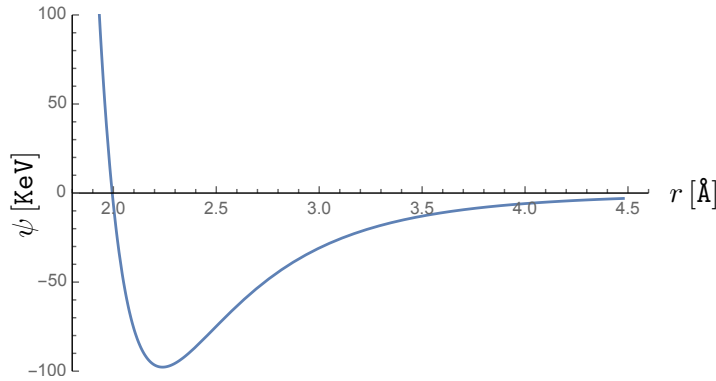


Fig. 3.8: An example of Lennard–Jones potential based on computed constants of a metal.

In a final step, since the martensitic transformations are assumed to remain within the Ericksen–Pitteri neighborhood, a Taylor’s expansion of individual terms of the sum in equation (3.101) is in order. The only concern which might arise here is that the two variants belong to two different energetic wells, which means that their transformation strains are too far apart for a Taylor’s expansion to be meaningful. While this argument seems to be genuine at first, it is not relevant since this we consider the interactions between second neighbors only. At typical distances between second neighbors the interaction potential, e.g. a Lennard–Jones type (see Fig. 3.8), behaves smooth enough in order to be estimated by a Taylor expansion. This expansion can be given in powers of

$$\begin{aligned} |\underline{\mathbf{a}}_i + \mathbf{T}\underline{\mathbf{a}}_j|^2 - |\underline{\mathbf{a}}_i + \underline{\mathbf{a}}_j|^2 &= \underline{\mathbf{a}}_j \cdot (\mathbf{T}^T \mathbf{T} - \mathbf{I}) \cdot \underline{\mathbf{a}}_j + 2\underline{\mathbf{a}}_i (\mathbf{T} - \mathbf{I}) \cdot \underline{\mathbf{a}}_j \\ &= \underline{\mathbf{a}}_j \cdot (\underline{\mathbf{u}}\nabla + \nabla\underline{\mathbf{u}} + \nabla\underline{\mathbf{u}} \cdot \underline{\mathbf{u}}\nabla) \cdot \underline{\mathbf{a}}_j + 2\underline{\mathbf{a}}_i \cdot (\underline{\mathbf{u}}\nabla) \cdot \underline{\mathbf{a}}_j \\ &= \underline{\mathbf{a}}_j \cdot (\nabla\underline{\mathbf{u}} \cdot \underline{\mathbf{u}}\nabla) \cdot \underline{\mathbf{a}}_j + 2(\underline{\mathbf{a}}_i + \underline{\mathbf{a}}_j) \cdot (\underline{\mathbf{u}}\nabla) \cdot \underline{\mathbf{a}}_j. \end{aligned} \quad (3.102)$$

where $\underline{\mathbf{u}}\nabla = \mathbf{T} - \mathbf{I} = \mathbf{a} \otimes (\mathbf{F}_1^{-T} \hat{\mathbf{N}}) = \mathbf{a} \otimes \mathbf{b}$, for which we showed that $\mathbf{a} \cdot \mathbf{b} = 0$. Therefore, $\nabla\underline{\mathbf{u}} \cdot \underline{\mathbf{u}}\nabla = \mathbf{0}$, and the above equation is simplified to

$$|\underline{\mathbf{a}}_i + \mathbf{T}\underline{\mathbf{a}}_j|^2 - |\underline{\mathbf{a}}_i + \underline{\mathbf{a}}_j|^2 = 2(\underline{\mathbf{a}}_i + \underline{\mathbf{a}}_j) \cdot (\underline{\mathbf{u}}\nabla) \cdot \underline{\mathbf{a}}_j. \quad (3.103)$$

Expanding (3.101) around $\mathbf{T} = \mathbf{I}$ up to $\mathcal{O}(|\underline{\mathbf{u}}\nabla|)$ yields

$$E_0 \approx \varrho_{(l_1 l_2 l_3)} A_0 \sum_{i=1}^3 \sum_{j=i}^3 \left\{ \psi'(|\underline{\mathbf{a}}_i + \underline{\mathbf{a}}_j|^2) [2(\underline{\mathbf{a}}_i + \underline{\mathbf{a}}_j) \cdot (\underline{\mathbf{u}}\nabla) \cdot \underline{\mathbf{a}}_j] \right\}. \quad (3.104)$$

Due to the lattice correspondence, for every pair of vectors $\mathbf{a}_i$ and $\mathbf{a}_j$ it holds that $\mathbf{a}_i + \mathbf{T}\mathbf{a}_j = \mathbf{a}_j + \mathbf{T}\mathbf{a}_i$, which shows $\mathbf{T}\mathbf{a}_j - \mathbf{a}_j = \mathbf{T}\mathbf{a}_i - \mathbf{a}_i$. But we know that $\mathbf{T} - \mathbf{I} = \underline{\mathbf{u}}\nabla$, therefore

$$\underline{\mathbf{u}}\nabla \cdot \mathbf{a}_1 = \underline{\mathbf{u}}\nabla \cdot \mathbf{a}_2 = \underline{\mathbf{u}}\nabla \cdot \mathbf{a}_3. \quad (3.105)$$

These equalities have an interesting geometrical interpretation. We mentioned that $\underline{\mathbf{u}}\nabla = \mathbf{a} \otimes (\mathbf{F}_1^{-T} \hat{\mathbf{N}})$, which can be reformulated, using equations (3.12) to (3.15), to

$$\underline{\mathbf{u}}\nabla = \frac{1}{|\mathbf{F}_1|} \mathbf{a} \otimes (\text{Cof } \mathbf{F}_1 \hat{\mathbf{N}}) = \frac{dV}{dv} \frac{da}{dA} \mathbf{a} \otimes \hat{\mathbf{n}}, \quad (3.106)$$

where dV and dv are the volume elements in austenitic and martensitic phase, dA and da are the area elements on the twin interface before and after martensitic transformation, and $\hat{\mathbf{n}}$ is the unit vector perpendicular to twin interfaces in the current configuration, i.e. after transformation. We know, however, that the ratios $\frac{dV}{dA}$ and $\frac{dv}{da}$ are the distances between the two adjacent monolayers parallel to the twinning plane before and after transformation respectively. Furthermore, multiplication of the unit normal with any of the lattice vectors equals the distance between the two adjacent monolayers, $\hat{\mathbf{n}} \cdot \mathbf{a}_i = d_{(l_1 l_2 l_3)}$, which leaves us with

$$\underline{\mathbf{u}}\nabla \cdot \mathbf{a}_j = \frac{D_{(l_1 l_2 l_3)}}{d_{(l_1 l_2 l_3)}} d_{(l_1 l_2 l_3)} \mathbf{a} = D_{(l_1 l_2 l_3)} \mathbf{a}. \quad (3.107)$$

Eventually the lattice interface energy per unit area from (3.104) equals

$$\gamma \approx \left[2 \varrho_{(l_1 l_2 l_3)} D_{(l_1 l_2 l_3)} \sum_{i=1}^3 \sum_{j=i}^3 \psi'(|\mathbf{a}_i + \mathbf{a}_j|^2) [\mathbf{a}_i + \mathbf{a}_j] \right] \cdot \mathbf{a}. \quad (3.108)$$

The vector within the large brackets can be thought as a *material constant*. Notice that its variations due to accommodation of the microstructure are of higher-order and can be therefore left out. We will show that this material constant has a direct connection with the elastic constants of the martensite and to the martensitic transformation itself. But just by looking at its terms and factors, calculation of (3.108) requires the following information

- $\varrho_{(l_1 l_2 l_3)}$, $D_{(l_1 l_2 l_3)}$, and $\mathbf{a}_i$ require crystallographic data and the lattice parameters.
- $\mathbf{a}$ is the jump across twin interface and so requires a study of all possible twinning modes among distinct martensite variants.
- $\psi'(\bullet)$ involves explicit derivative of lattice potential, therefore requires the type and the parameters of the potential, e.g. Lennard–Jones.

3.4.2 Martensite twin interface energy

Tab. 3.3 lists the solution (3.78) for various possible twinning modes for the cubic to monoclinic-I transformation of NiTi variants. This table is complemented by Tab. 3.4 which shows all the twin-compatible variants together with their twinning mode. This table will be used in our general computation of the laminate microstructure. It is crucial to calculate interplanar spacings $D_{(l_1 l_2 l_3)}$ and the planar atomic densities $\varrho_{(l_1 l_2 l_3)}$ of the twinning

plane based on the B2 orthorhombic unit cell and not the cubic unit cell to keep the lattice correspondence fulfilled. These values are obtained from the following calculations

$$D_A = \ell_0, \quad D_B = \ell_0, \quad D_C = \frac{\ell_0}{\sqrt{2}}, \quad D_D = \frac{\ell_0}{\sqrt{2}}, \quad (3.109a)$$

$$\varrho_A = \frac{4 \times \frac{1}{4} + 1}{\|\mathbf{a}_1^m \times \mathbf{a}_2^m\|}, \quad \varrho_B = \frac{4 \times \frac{1}{4} + 2 \times \frac{1}{2}}{\|(\mathbf{a}_1^m + \mathbf{a}_2^m) \times \mathbf{a}_3^m\|}, \quad (3.109b)$$

$$\varrho_C = \frac{3 \times \frac{1}{6} + 3 \times \frac{1}{2}}{\|(\mathbf{a}_1^m - \mathbf{a}_3^m) \times (\mathbf{a}_1^m + \mathbf{a}_2^m)\|}, \quad \varrho_D = \frac{3 \times \frac{1}{6} + 3 \times \frac{1}{2}}{\|(\mathbf{a}_2^m - \mathbf{a}_3^m) \times (\mathbf{a}_1^m + \mathbf{a}_2^m)\|}. \quad (3.109c)$$

The above calculations also imply the lattice directions and planes in the martensite coordinates that correspond to those given in Tab. 3.3 in austenite lattice coordinates.

Calculation of the first derivatives $\psi'(\bullet)$ in (3.107) involves explicit derivative of the lattice potential. In the case of Lennard–Jones potential (3.88), for instance, we need to determine two parameters ε and r_m to be able to calculate the derivative. An apparent method for determining these parameters would be to solve analytical expressions of macroscopic properties, such as elastic constants, that explicitly depend on the lattice potential and/or its derivatives [159, 131]. This approach takes the following steps

1. Deriving an explicit formulation of nonzero elements of the elastic stiffness tensor in terms of lattice parameters and an interatomic potential.
2. Assuming appropriate form(s) of interatomic potential(s) whose constants are unknowns which are to be determined.
3. Substituting the known crystallographic values together with interatomic potentials into the formulation of elastic constants from step 1 and setting these constants to their known experimental values.
4. Solving the system of equations, established in step 3, in terms of the unknown constants of the interatomic potential(s). Obviously, we must setup the same number of equations in step 3 as there are unknowns.

Before tending to the above procedure, let us visit some of the available results on NiTi elastic constants. Ren et al (2001) present empirical data on elastic constants of B2 phase in various NiTi alloys within a wide temperature range, prior to martensitic transformation [134]. The B2 phase, having a cubic lattice, has 3 independent elastic constants. To date, We have not been able find empirical data on the elastic constants of B19', which are 13 due to its monoclinic lattice structure. For B19' phase, therefore, we consult the works of Wagner et al (2008) and the more recent work of Šesták et al (2010) which report computational results on the elastic constants of B2 and stress-stabilized B19' based on first principles [149, 140]. Both assume thermodynamic ground states, i.e. temperatures at 0K. Yet we know that deformation-driven martensitic transformation takes place at temperatures higher than the *austenite finish temperature*, $\theta_{Af} = 397.5\text{K}$ up to almost $\theta_{Af} + 40\text{K}$ [54, pp. 3450, 3454] (see Fig. 3.9). Generally, elastic constants are not much influenced by temperature. In B2 phase, however, the elastic constants related to lattice directions that are involved in phase transformation show strong temperature dependence [139].

For B2 we take the empirical results in [134] as close as possible to 400 K and 49.75 at. % Ni.

$$C_{ijkl} = \begin{cases} 175 & i = j = k = l, \\ 149 & i = j \wedge k = l, \\ 40 & (i = k \wedge j = l) \vee (i = l \wedge j = k). \end{cases} \quad (3.110)$$

All values are in GPa. For B19' we find the results of [149] to be more reliable because their prediction for B2 is closer to the empirical data, and also because of their more robust approach towards stability and their stress-stabilized results for B19':

$$\begin{aligned} C_{1111} &= 223, & C_{2222} &= 241, & C_{3333} &= 200, \\ C_{1122} &= 129, & C_{2233} &= 125, & C_{3311} &= 99, \\ C_{1113} &= 27, & C_{2213} &= -9, & C_{3313} &= 4, \\ C_{1212} &= 77, & C_{2323} &= 76, & C_{3131} &= 21, & C_{2312} &= -4. \end{aligned} \quad (3.111)$$

Again, the values are in GPa. The major and the minor symmetries will produce the remaining nonzero elements. The above stiffness tensor is expressed in the martensite rectangular coordinates for the 1st variant (Fig. 3.3). For the n th martensite variant, the stiffness tensor has to be transformed under the corresponding elementary operator listed in Tab. 3.2, as

$$\mathbb{C}_n = \mathbf{E}_n \otimes \mathbf{E}_n : \mathbb{C}_1 : \mathbf{E}_n^T \otimes \mathbf{E}_n^T \quad \text{or} \quad C_{ijkl}^n = E_{ia}^n E_{jb}^n E_{kc}^n E_{ld}^n C_{abcd}^1, \quad (3.112)$$

where superscripts stand for variants in the index notation.

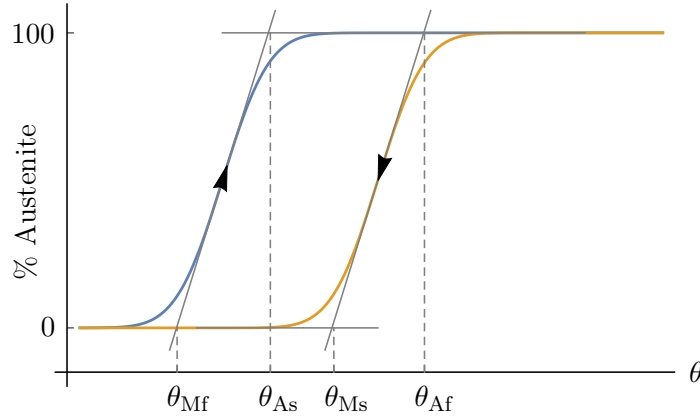


Fig. 3.9: Martensitic transformation temperatures; martensite start θ_{Ms} and finish θ_{Mf} temperatures, and austenite start θ_{As} and finish θ_{Af} temperatures.

As compared to a simple monatomic monoclinic lattice, an atomistic energy formulation of martensitic NiTi faces at least three difficulties. One is that the martensitic NiTi has a face-centered monoclinic structure whose lattice energy formulation is more involved than a simple monoclinic lattice. The second difficulty is that there are two constituent elements, i.e. Nickel and Titanium, which would require at least three types of interatomic potentials in the lattice energy formulation. These are Ni–Ni, Ti–Ti, and Ni–Ti interactions. This adds

to the complexity of the problem. There is also a third twist, namely the stability of the empirically established unit cell. It is well known that the B19' unit cell given in Tab. 3.1 does not have a stable ground energy. Our formulation of the elastic constants confirms this too. Due to the analytical nature of our method we have to face these three issues, but we do so only by approximation.

As for the three distinct types of atomic interactions, it should be safe to assume an average common potential because the two elements, Ni and Ti, are present in almost equal proportions. Also because the atoms lying at face centers of the unit cell, although slightly shifted due to dissimilarity of the face symmetries, are of both kinds which roughly shows that the interactions between similar and dissimilar atoms should not differ substantially. So we are going to adopt a single Lennard–Jones potential (3.88) for all three types of interaction. Following this simplification we can treat the lattice as that of a monatomic substance which will pave the way for resolving the other difficulty: the face-centeredness of the NiTi martensite.

Following the technique in appendix A the energy density of a face-centered monoclinic lattice is expressed by

$$e = \frac{1}{V_{\mathfrak{A}}} \sum_i \left\{ 4 \times \frac{1}{4} \psi(|\mathbf{a}_i|^2) + 4 \times \frac{1}{2} \sum_{j>i} \left\{ \psi(1/4 |\mathbf{a}_i + \mathbf{a}_j|^2) + \psi(1/4 |\mathbf{a}_i - \mathbf{a}_j|^2) \right\} \right\}. \quad (3.113)$$

This is the energy stored in a unit cell, as the factor $1/V_{\mathfrak{A}}$ suggests. The first term of the outer sum relates to the interactions between corner atoms whose interatomic distances are perceived as edges of the unit cell. There are 12 edges. Every 4 edges are parallel, which makes for 3 different interatomic distances, and each edge belongs to 4 unit cells. This explains the factor $4 \times 1/4$. The inner sum represents the interaction between an atom at a corner of a face and the one at its center, whose interatomic distance equals either $(\mathbf{a}_i + \mathbf{a}_j)/2$ or $(\mathbf{a}_i - \mathbf{a}_j)/2$. Each of these interactions is shared by two neighboring unit cells and each has a multiplicity of 4 (2 for opposite faces and 2 for each diagonal on every face). This will explain the factor $4 \times 1/2$. According to approximation (B.8) we can write the elastic stiffness tensor as

$$\mathbb{C} = \frac{1}{V_{\mathfrak{A}}} \sum_i \left\{ \mathbb{K}(|\mathbf{a}_i|^2) + 2 \sum_{j>i} \left\{ \mathbb{K}(1/4 |\mathbf{a}_i + \mathbf{a}_j|^2) + \mathbb{K}(1/4 |\mathbf{a}_i - \mathbf{a}_j|^2) \right\} \right\}, \quad (3.114)$$

in which $\mathbb{K}(\bullet)$ is given by (B.13).

Since the Lennard–Jones potential is determined by two parameters, we need to derive mathematical expressions for two non-zero elements of the elastic stiffness tensor whose values are known from experiments or from *ab initio* computations. Expanding the above equation for C_{1111} and C_{2222} gives:

$$\begin{aligned} C_{1111} = \frac{1}{V_{\mathfrak{A}}} & \left\{ \ell_1^4 \left[4\psi''(\ell_1^2) + \psi''\left(\frac{\ell_1^2 + \ell_2^2}{4}\right) + \psi''\left(\frac{\ell_1^2 + \ell_3^2}{4}\right) \right] \right. \\ & \left. + 2\ell_1^2 \left[\psi'(\ell_1^2) + \psi'\left(\frac{\ell_1^2 + \ell_2^2}{4}\right) + \psi'\left(\frac{\ell_1^2 + \ell_3^2}{4}\right) \right] \right\} \end{aligned} \quad (3.115)$$

$$\begin{aligned}
& + 9\ell_2^2 \left[2\psi'(\ell_2^2) + 2\psi'\left(\frac{\ell_1^2 + \ell_2^2}{4}\right) + \psi'\left(\frac{\ell_2^2 + 2\ell_2\ell_3 \cos \theta + \ell_3^2}{4}\right) + \psi'\left(\frac{\ell_2^2 - 2\ell_2\ell_3 \cos \theta + \ell_3^2}{4}\right) \right] \\
& + 18\ell_2\ell_3 \cos \theta \left[\psi'\left(\frac{\ell_2^2 + 2\ell_2\ell_3 \cos \theta + \ell_3^2}{4}\right) - \psi'\left(\frac{\ell_2^2 - 2\ell_2\ell_3 \cos \theta + \ell_3^2}{4}\right) \right] \\
& + 9\ell_3^2 \left[2\psi'(\ell_3^2) + 2\psi'\left(\frac{\ell_1^2 + \ell_3^2}{4}\right) + \psi'\left(\frac{\ell_2^2 + 2\ell_2\ell_3 \cos \theta + \ell_3^2}{4}\right) + \psi'\left(\frac{\ell_2^2 - 2\ell_2\ell_3 \cos \theta + \ell_3^2}{4}\right) \right] \Big\} \\
C_{2222} = & \frac{1}{V_{\mathfrak{A}}} \left\{ 18\ell_1^2 \left[\psi'(\ell_1^2) + \psi'\left(\frac{\ell_1^2 + \ell_2^2}{4}\right) + \psi'\left(\frac{\ell_1^2 + \ell_3^2}{4}\right) \right] \right. \\
& + \frac{1}{2}\ell_2^4 \sin^4(\theta) \left[8\psi''(\ell_2^2) + 2\psi''\left(\frac{\ell_1^2 + \ell_2^2}{4}\right) + \psi''\left(\frac{\ell_2^2 + \ell_3^2 + 2\ell_2\ell_3 \cos \theta}{4}\right) + \right. \\
& \left. \left. \psi''\left(\frac{\ell_2^2 + \ell_3^2 - 2\ell_2\ell_3 \cos \theta}{4}\right) \right] + \ell_2^2(4 \cos(2\theta) + 5) \left[2\psi'(\ell_2^2) + 2\psi'\left(\frac{\ell_1^2 + \ell_2^2}{4}\right) + \right. \right. \\
& \left. \left. \psi''\left(\frac{\ell_2^2 + \ell_3^2 + 2\ell_2\ell_3 \cos \theta}{4}\right) + \psi''\left(\frac{\ell_2^2 + \ell_3^2 - 2\ell_2\ell_3 \cos \theta}{4}\right) \right] \right. \\
& + 18\ell_2\ell_3 \cos \theta \left[\psi'\left(\frac{\ell_2^2 + 2\ell_2\ell_3 \cos \theta + \ell_3^2}{4}\right) - \psi'\left(\frac{\ell_2^2 - 2\ell_2\ell_3 \cos \theta + \ell_3^2}{4}\right) \right] \\
& \left. + 9\ell_3^2 \left[2\psi'(\ell_3^2) + 2\psi'\left(\frac{\ell_1^2 + \ell_3^2}{4}\right) + \psi'\left(\frac{\ell_2^2 + 2\ell_2\ell_3 \cos \theta + \ell_3^2}{4}\right) + \psi'\left(\frac{\ell_2^2 - 2\ell_2\ell_3 \cos \theta + \ell_3^2}{4}\right) \right] \right\}. \tag{3.116}
\end{aligned}$$

Now we take the first two derivatives of the Lennard–Jones potential (3.88) with respect to r^2 , which are

$$\psi'(r^2) = 6\varepsilon \left(\frac{r_m^6}{r^8} - \frac{r_m^{12}}{r^{14}} \right), \quad \psi''(r^2) = 6\varepsilon \left(-4\frac{r_m^6}{r^{10}} + 7\frac{r_m^{12}}{r^{16}} \right), \tag{3.117}$$

and the lattice parameters given in Tab. 3.1, and the elastic constants in [149]:

$$\begin{aligned}
\ell_1 &= 4.120 \text{ \AA}, \quad \ell_2 = 4.622 \text{ \AA}, \quad \ell_3 = 2.889 \text{ \AA}, \quad \theta = 96.8^\circ, \\
C_{1111} &= 223 \text{ GPa}, \quad C_{2222} = 241 \text{ GPa},
\end{aligned} \tag{3.118}$$

into the (3.115) and (3.116) to get to the following system of equations in terms of Lennard–Jones parameters:

$$\begin{aligned}
\varepsilon (0.0279663 r_m^6 - 0.00003168 r_m^{12}) &= 223, \\
\varepsilon (0.0338999 r_m^6 - 0.00009489 r_m^{12}) &= 241,
\end{aligned} \tag{3.119}$$

Solving this system under $\varepsilon, r_m \in \mathbb{R}^+$, yields

$$r_m = 1.982 \text{ \AA}, \quad \varepsilon = 1.413 \times 10^{-19} \text{ J}. \tag{3.120}$$

Remark 12. The results we just presented are of estimative nature. First of all, a more reliable calculation of elastic constants should have been based on the *relaxed base-centered orthorhombic* unit cell since the empirical face-centered monoclinic cell is not energetically stable [133]. We nevertheless opted for the empirical cell because inevitably we have calculated the martensitic transformations (Tab. 3.2) based on the experimental data. Hence, one should not be tempted to produce all the other elastic constants based on the com-

puted Lennard–Jones parameters (3.120) whatsoever. Yet, among the 13 ab initio elastic constants (3.111) we took the two largest values, C_{1111} and C_{2222} , which also happen to correspond to lattice directions that are energetically most stable.

For computing the twin interface energy per unit area expressed by (3.108), we need one more ingredient, namely the Miller indices of the twinning plane in each twinning mode, which are

$$\text{A: } (00\bar{1}), \quad \text{B: } (1\bar{1}0), \quad \text{C: } (1\bar{1}1), \quad \text{D: } (1\bar{1}\bar{1}). \quad (3.121)$$

When equation (3.108) is expanded for twinning modes B, C, and D the face-centered atoms contribute to the interface energy. This why we see lattice distances in the form of $\frac{1}{2}\mathbf{a}_i + \frac{2}{3}\mathbf{a}_j$ in the following expressions. Introducing $w(\mathbf{u}) := \psi'(|\mathbf{u}|^2) \mathbf{u}$, the results are:

$$\begin{aligned} \gamma_A &= 2\rho_A D_A w(-\mathbf{a}_3 - \mathbf{a}_3) \cdot \mathbf{a}_A, \\ \gamma_B &= 2\rho_B D_B \left[w(\mathbf{a}_1 + \mathbf{a}_1) + w(\mathbf{a}_1 - \mathbf{a}_2) + w(-\mathbf{a}_2 - \mathbf{a}_2) \right. \\ &\quad \left. + w\left(\frac{3}{2}\mathbf{a}_1 - \frac{1}{2}\mathbf{a}_2\right) + w\left(\frac{1}{2}\mathbf{a}_1 - \frac{3}{2}\mathbf{a}_2\right) \right] \cdot \mathbf{a}_B, \\ \gamma_C &= 2\rho_C D_C \left[w(\mathbf{a}_1 + \mathbf{a}_1) + w(\mathbf{a}_1 - \mathbf{a}_2) + w(\mathbf{a}_1 + \mathbf{a}_3) \right. \\ &\quad \left. + w(-\mathbf{a}_2 - \mathbf{a}_2) + w(-\mathbf{a}_2 + \mathbf{a}_3) + w(\mathbf{a}_3 + \mathbf{a}_3) \right. \\ &\quad \left. + w\left(\frac{3}{2}\mathbf{a}_1 - \frac{1}{2}\mathbf{a}_2\right) + w\left(\frac{1}{2}\mathbf{a}_1 - \frac{3}{2}\mathbf{a}_2\right) + w\left(-\frac{3}{2}\mathbf{a}_2 + \frac{1}{2}\mathbf{a}_3\right) \right. \\ &\quad \left. + w\left(-\frac{1}{2}\mathbf{a}_2 + \frac{3}{2}\mathbf{a}_3\right) + w\left(\frac{3}{2}\mathbf{a}_3 + \frac{1}{2}\mathbf{a}_1\right) + w\left(\frac{1}{2}\mathbf{a}_3 + \frac{3}{2}\mathbf{a}_1\right) \right] \cdot \mathbf{a}_C, \\ \gamma_D &= 2\rho_D D_D \left[w(\mathbf{a}_1 + \mathbf{a}_1) + w(\mathbf{a}_1 - \mathbf{a}_2) + w(\mathbf{a}_1 - \mathbf{a}_3) \right. \\ &\quad \left. + w(-\mathbf{a}_2 - \mathbf{a}_2) + w(-\mathbf{a}_2 - \mathbf{a}_3) + w(-\mathbf{a}_3 - \mathbf{a}_3) \right. \\ &\quad \left. + w\left(\frac{3}{2}\mathbf{a}_1 - \frac{1}{2}\mathbf{a}_2\right) + w\left(\frac{1}{2}\mathbf{a}_1 - \frac{3}{2}\mathbf{a}_2\right) + w\left(-\frac{3}{2}\mathbf{a}_2 - \frac{1}{2}\mathbf{a}_3\right) \right. \\ &\quad \left. + w\left(-\frac{1}{2}\mathbf{a}_2 - \frac{3}{2}\mathbf{a}_3\right) + w\left(-\frac{3}{2}\mathbf{a}_3 + \frac{1}{2}\mathbf{a}_1\right) + w\left(-\frac{1}{2}\mathbf{a}_3 + \frac{3}{2}\mathbf{a}_1\right) \right] \cdot \mathbf{a}_D, \end{aligned} \quad (3.122)$$

which after substitution of the numerical values of the physical quantities give

$$\gamma_A = -0.004059 \text{ J.m}^{-2}, \quad \gamma_B = -0.002917 \text{ J.m}^{-2}, \quad \gamma_C = 0.05531 \text{ J.m}^{-2}, \quad \gamma_D = 0.02296 \text{ J.m}^{-2}. \quad (3.123)$$

Data on coherent interface energies are scarce for obvious experimental limitations. We are therefore content to check whether our numerical results have decent orders of magnitude compared to the values in the literature. Waitz et al (2005) suggest values of the order 10^{-2} J.m^{-2} for a sharp twin interface between B19' compatible variants [151]. Olmsted et al (2009) compute $\sim 0.06 \text{ J.m}^{-2}$ for coherent interfaces in FCC Ni [119]. Kaptay (2012) obtains values between 10^{-1} and 10^{-2} J.m^{-2} for various alloys and at various temperatures [89]. Our numerical results should hence be reasonable enough for use in the subsequent computations.

Most conspicuously, for twinning modes A and B we have obtained negative interface energies; or better to say, the twin formation reduces the overall stored energy of the lattice. This can be intuitively explained as follows. Fig. 3.10 is an illustrative comparison of the overall span of a single variant martensite (left) and a laminate composed of two compatible variants (right). The two pieces have the same amount of material and since the two variants are compatible, their crystallographic difference comes down to a symmetry connection within the point group of the parent phase. This means that the distances between the immediate neighboring atoms do not change when transforming from the left configuration to the right configuration. On the other hand, although the volume of the specimen does not change, the overall span of the body reduces by forming the laminate microstructure. This should mean that the interatomic distances between non-immediate neighbors decrease. Therefore, the lattice energy of the laminated configuration (right) is inevitably lower than the single-variant configuration (left). Of course, whether formation of laminate results in such a phenomenon depends on the lattice structure, as well as the orientation of the twinning plane and the direction of the transformation shear. So, in those cases where $\mathbf{F}_i \mathbf{F}_j^{-1}$ results in the reduction of the overall span of the martensitic body $\mathbf{F}_j$, the interface energy *can be* negative, in the sense we explained.

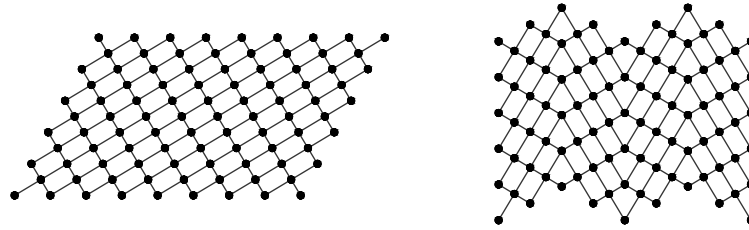


Fig. 3.10: Comparison of overall lattice span of a single martensite variant (left) with a laminate composed of two compatible variants (right).

However, the twinned structure having a lower energy than a single variant structure inescapably presents the possibility that martensite should have the tendency to form laminated microstructure as fine as possible, in an effort to lower its energy, which does not seem plausible. The answer to this is multifaceted. First, between every two compatible variants of martensite there is a dynamic energy barrier due the parent phase energy well. This barrier prohibits the formation of overly fine twins at least in the case of deformation twins. Second, in our calculation of the twin interface energy we neglected the interatomic potentials between the atoms farther apart than second neighbors. If the laminated microstructure becomes extremely fine, say down to length scale of lattice constants of the martensite, then the effect of non-immediate neighbors can become significant and our first-order crystallographic calculation shall fail. Last but not least, our energetic formulation of the laminate microstructure has one more term which we have yet to investigate: the grain boundary energy. Forming a finer and finer twin microstructure will form a more jagged boundary which has to results in a higher energy stored in a boundary layer enclosing the microstructure. However, we again will present a lowest-order model of the boundary energy that does not counter-balance the laminate refinement. Therefore, for all twinning modes

and types we use the largest estimated value

$$\gamma = 0.05531 \text{ J.m}^{-2}, \quad (3.124)$$

in order to avoid the just explained irregularity of the model, and to dilute the ill-conditionedness of the tangent stiffness in the upcoming finite element computations.

Finally, to include the twin interface energies in the overall energy formulation of the laminate microstructure, we calculate the pointwise interface energy per unit volume of the *current configuration* which is given by

$$\frac{dE_0}{dv} = 2 \frac{\gamma dx_1 dx_2}{h(x_1, x_2) dx_1 dx_2} = 2 \frac{\gamma}{h}. \quad (3.125)$$

The volume element whose size appears here in the denominator has its particular form due to periodicity of the laminate along its normal direction x_3 . It is therefore sufficient to consider the volumetric density of interface energy for a single twin. The factor 2 reflects that each twin has one interface within and two interface at its sides which are shared by two neighboring twins. This sums up to $1 + 2 \times \frac{1}{2} = 2$. As a final comment, it can be easily verified that the above formula remains valid in both material and spacial coordinates using (3.15), (3.16), and (3.19).

Remark 13. In addition to remark 12, there can be other sources of inaccuracy in the presented method of estimating the interface energy (3.124). The input arguments in (3.122), namely ϱ , D , $\mathbf{a}_i$, $\mathbf{a}$, are expressed in terms of lattice parameters either directly or indirectly. Lattice parameters are commonly measured via X-ray diffraction. The accuracy of these measurements can be estimated using Bragg's law augmented with probabilistic analyses [71]. These suggest a relatively conservative estimate of the uncertainty of lattice parameters as $\delta\ell \approx 10^{-4}\ell$, which for B19' phase of NiTi roughly gives $\delta\ell \sim 10^{-4} [\text{\AA}]$. Now using elementary geometry and first-order approximations, the uncertainty of the remaining input arguments can be estimated up to orders of magnitude as $\delta\varrho \sim 10^{-5} [\text{\AA}^{-2}]$ and $\delta D \sim \delta \|\mathbf{a}_i\| \sim \delta \|\mathbf{a}\| \sim 10^{-4} [\text{\AA}]$. Substituting these into, once again, a first-order estimate of (3.122) will leave us with $\delta\gamma_C \sim 10^{-3}\gamma_C$, which does not seem significant.

It has been shown that the 12-6 version of the Lennard-Jones potential is not optimal for solid metals. When used in computing bulk or shear modulus, it might lead to significant discrepancies [53, pp. 49], which is why we carried out our computations based on Young's moduli only. Even so, our estimates might still require a correction up to 30% [53, pp. 52]. To improve, we could have used modified forms of the L-J potential, for instance those worked out by Zhen and Davies 1983 [159]. Let us emphasize that we opted for the standard Lennard-Jones potential to simplify the calculations, and there should basically be no problem in using more sophisticated forms.

3.5 Grain boundary energy

The third and last part of our energy formulation of the martensitic laminate microstructure comes from the energy stored at the boundaries. We term this the “grain boundary energy”,

which turns out to be the most heuristic piece of our model due to its inherent complexity. An actual grain in a martensitic material capable of manifesting shape memory effects, such as NiTi, is a cluster of laminates with different orientations and a wide spectrum of length scales [36, 113, 3, 99] (Fig. 3.11). Some patches of laminate might be composed of lamellae that are made of twin microstructures at a smaller length scale, which is called a second order laminate. If the length scale allows, this could even proceed further to formation of higher-order laminates; that is, laminate within laminate within laminate and so on. In this picture, the size, shape, morphology, and composition of a grain is overwhelmingly complex and its analysis will go far beyond the scope of this work. There is a vast body of literature dedicated to modeling different aspects of grain formation and evolution, and its influence on precipitation, transformation temperature, mobility, and so on. Each work, however, focuses on very specific characteristics of a grain, as the entirety of the problem is inhibitingly complicated.

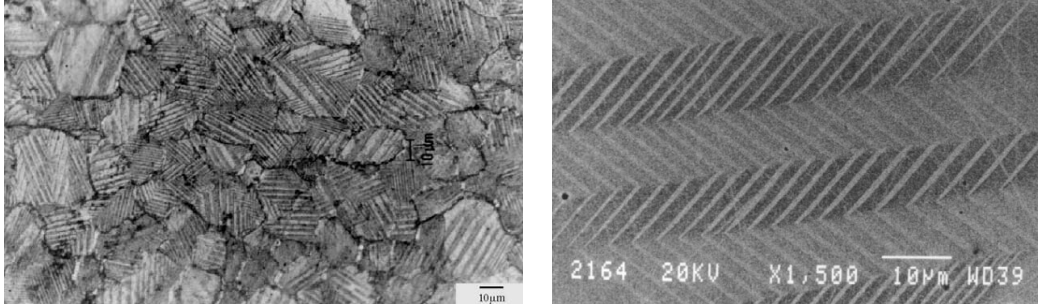


Fig. 3.11: **Left:** a cluster of laminates distributed across numerous grains in PtAl microstructure; **Right:** PtAl laminate at various orders [27, ©].

Our ultimate objective is to formulate a microscopic energy formulation incorporating interface effects to model the mesoscopic and eventually the macroscopic behavior of shape memory materials. Therefore, at some point we have to settle for a tractable level of sophistication that will allow for a homogenization scheme to be numerically implemented. That is why we chose not to consider laminates of orders higher than one. In the same way, we pick up the simplest possible grain construct, namely a single patch of laminate. Then again, choosing more complex constructs would contradict our initial premise of a first-order laminate. Because two adjacent patches of laminate often fulfill compatibility condition at their common interface in an average sense [11], which is mathematically equivalent to a first-order coherence condition for which the compatibility of the laminate microstructure would stand as a second-order coherence condition.

Our kinematic ansatz (3.11) of the laminate fulfills homogeneous boundary conditions, to wit, external constraints are collectively simulated as Dirichlet boundary condition $\mathbf{x}(\mathbf{X}) = \mathbf{F} \cdot \mathbf{X}$ for all $\mathbf{X} \in \partial\omega$. This will be satisfied almost strictly at the boundary perpendicular to the normal direction $\hat{\mathbf{N}}$. These are the points at which the boundary is parallel to the lamellae or nearly so. However, there will always be discrepancies between the laminate and its boundary where the oscillating displacement field of the laminate meets the homogeneous displacement of the boundary head-on. In short, the grain boundary will store energy in two

forms: formation energy which is the energy stored at the atomic scale, and deformation energy due to microstrains propagated by the jaggedness of the laminate into a thin boundary layer around it. Wherever the laminate is parallel to the boundary the microstrain energy will be absent.

The boundary formation energy is conceptually similar to the twin interface energy, except that there is no simple way of determining whether at each point the grain boundary is coherent, semicoherent, or incoherent because of the randomness of relative lattice orientations among neighboring grains. The degree of incoherence of an interface is determined by the misorientation between lattices that form the interface. As the lattice misorientation increases the boundary energy grows as dislocations populate, although the energy per dislocation decreases. We take the grain boundary formation energy as the ground thermochemical energy of the boundary and assume it constant. Olmsted et al give grain boundary formation energies from 0.06 up to 1 J.m⁻² for FCC Ni, depending on lattice misorientation [119, 74]. We use values whose orders range from 10⁻¹ to 10⁰ in various examples, especially

$$\gamma_{\Gamma}^a = 0.41 \text{ J.m}^{-2}. \quad (3.126)$$

The elastic microstrain energy of the grain boundary, among other things, depends on the precise morphology of the laminate tips. Stupkiewicz et al have investigated some of the possible geometries that locally attain energetic minima and have formulated their results in the form of particular characterizing shape parameters [144, 145]. Their boundary, however, lies at the *habit plane*. Since we impose Dirichlet boundary conditions on the laminate, it is reasonable to use their results. Nevertheless, since we do not intend to add to the number of thermomechanical variables of our model, we employ some of their average values only as a reference point to make a proper first guess at the factors that we will introduce in the sequel.

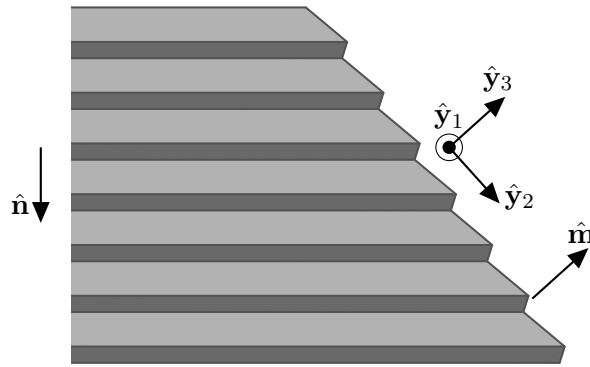


Fig. 3.12: The laminate meets the grain boundary at an inclination angle. The relative orientation of the unit vectors $\hat{\mathbf{n}}$ and $\hat{\mathbf{m}}$, normal to laminate and to boundary, is crucial to our formulation.

Fig. 3.12 illustrates a typical cross section of the laminated boundary. The unit vectors $\hat{\mathbf{n}}$ and $\hat{\mathbf{m}}$ are normal to the laminate and the grain boundary in the current configuration. The boundary coordinates are denoted by $\mathbf{y}$ and its orthonormal basis is defined locally as

$$\hat{\mathbf{y}}_1 = \frac{\hat{\mathbf{n}} \times \hat{\mathbf{m}}}{|\hat{\mathbf{n}} \times \hat{\mathbf{m}}|}, \quad \hat{\mathbf{y}}_2 = \hat{\mathbf{m}} \times \hat{\mathbf{y}}_1, \quad \hat{\mathbf{y}}_3 = \hat{\mathbf{m}}. \quad (3.127)$$

Our basic assumption is that the micro displacement field caused by the laminated has a self-similar structure of the form

$$\mathbf{u}(\mathbf{y}) = \Phi\left(\frac{\mathbf{y}}{\underline{h}}\right) \cdot \mathbf{v}, \quad (3.128)$$

where $\Phi : \mathbb{R}^3 \rightarrow \mathbb{R}^{3 \times 3}$ is a self-similarity tensor which is a direct generalization of its scalar counterpart, $\underline{h}$ is the projection of the lath with onto the boundary, and $\mathbf{v}$ is the maximum laminate displacement at the boundary which are given by

$$\underline{h} = \frac{h}{|\hat{\mathbf{n}} \times \hat{\mathbf{m}}|}, \quad \mathbf{v} = [\Lambda(1 - \Lambda) H \mathbf{a}]_{\partial\Omega}. \quad (3.129)$$

Notice that we have left $\mathbf{v}$ in material coordinates. The reason for this will become clear shortly.

Remark 14. The laminate elastic energy (3.64) was formulated based on the assumption that the gradients along $\hat{\mathbf{N}}$ vanish. If the laminate and its boundary meet obliquely, a case which we did not exclude, one might argue that this assumption is violated. It can be counter-argued that our calculation of the laminate elastic energy was expressed in the reference configuration where the parent phase resides; whereas the grain boundary energy is formulated in the current configuration where the martensitic transformation might have deformed the initial cuboid domain into a slanted parallelepiped, as in Fig. 3.12. After completing the formulation, we will *pull back* all the quantities to the reference configuration in which our fundamental premise is intact.

The micro displacement gradient is the directly obtained from (3.128):

$$\frac{\partial \mathbf{u}}{\partial \mathbf{y}} = \frac{1}{\underline{h}} \frac{\partial \Phi_{ij}}{\partial \xi_k} v_j \hat{\mathbf{y}}_i \otimes \hat{\mathbf{y}}_k = \frac{\mathbf{v}}{\underline{h}} \cdot (\Phi^T \otimes \nabla_\xi), \quad (3.130)$$

in which $\xi = \mathbf{y}/\underline{h}$ is the nondimensional coordinates.

We assume that the matrix surrounding the grain is homogeneous and isotropic at least within the boundary layer in which the microstrain energy is calculated. As mentioned before, this is a sound assumption since the relative orientation of the neighboring grains and their distribution is random. The elastic microstrain energy density is then obtained as

$$\psi_\Gamma = \frac{1}{2} \frac{\mathbf{v}}{\underline{h}} \cdot (\Phi^T \otimes \nabla_\xi) : \mathbb{C} : (\nabla_\xi \otimes \Phi) \cdot \frac{\mathbf{v}}{\underline{h}}. \quad (3.131)$$

Since the boundary is a two dimensional entity, we need to calculate the microstrain energy per uni area of the boundary. Since the laminate is periodic along $\mathbf{n}$, the microstrain field should be also periodic along $\hat{\mathbf{y}}_2$, with the same period $\underline{h}$. We therefore integrate (3.131) along a single period to obtain:

$$\gamma_\Gamma^e = \frac{1}{\underline{h} \, dy_1} \underline{h}^2 \int_0^{e/\underline{h}} \int_0^1 \frac{1}{2} \frac{\mathbf{v}}{\underline{h}} \cdot (\Phi^T \otimes \nabla_\xi) : \mathbb{C} : (\nabla_\xi \otimes \Phi) \cdot \frac{\mathbf{v}}{\underline{h}} \, d\xi_2 \, d\xi_3 \, dy_1. \quad (3.132)$$

where $\xi = \mathbf{y}/\underline{h}$ is the non-dimensional coordinates, $\underline{h} \, d\underline{x}_1$ is an infinitesimal area element on the boundary, factor $\underline{h}^2$ comes from Nondimensionalization of $d\underline{x}_1$ and $d\underline{x}_3$. We also assumed that the thickness of the boundary layer, e , is proportional to $\underline{h}$, along the lines of self-similarity postulate (3.128) [95, 34]. Notice that there is no integration along y_1 , therefore the final result should depend on it. After some elementary derivations we arrive at

$$\gamma_R^e = \frac{1}{\underline{h}} \mathbf{v} \cdot \overline{\Phi} \cdot \mathbf{v}, \quad (3.133)$$

in which

$$\overline{\Phi} = \int_0^{e/\underline{h}} \int_0^1 \frac{1}{2} (\Phi^T \otimes \nabla_\xi) : \mathbb{C} : (\nabla_\xi \otimes \Phi) \, d\xi_2 \, d\xi_3, \quad (3.134)$$

is a 3×3 tensor which generally depends on y_1 only. If we substitute $\mathbf{v}$ and $\underline{h}$ from (3.129) the final result will be

$$\gamma_R^e = (1 - (\hat{\mathbf{n}} \cdot \hat{\mathbf{m}})^2)^{1/2} \frac{H^2}{h} \Lambda^2 (1 - \Lambda)^2 \mathbf{a} \cdot \overline{\Phi} \cdot \mathbf{a}. \quad (3.135)$$

The term in the first parentheses comes from the trigonometric identity

$$|\hat{\mathbf{n}} \times \hat{\mathbf{m}}| = |\sin \theta| = \sqrt{1 - \cos^2 \theta} = (1 - (\hat{\mathbf{n}} \cdot \hat{\mathbf{m}})^2)^{1/2}.$$

For every combination of martensite and surrounding matrix the tensor $\overline{\Phi}$ can be determined at least numerically. With just one additional assumption, namely uniform microstrain distribution along y_2 , one can even obtain explicit analytical solutions. This is because the problem of an isotropic elastic half-space is among the most widely investigated classical problems of elasticity, especially due its vast uses in the field of contact mechanics. Most such solutions are based on classical works of Mindlin and Cerruti and using Green function and Fourier analysis [108, 87, 152]. We, however, intend not to add to the complexity of our upcoming numerical solution, because the homogenization of what we have formulated so far will be computationally expensive as it is. So let us simplify the grain boundary energy by assuming that the tensor $\overline{\Phi}$ is independent of y_1 and is isotropic, none of which are uncommon assumptions [113, 38, 98]. Thus, $\overline{\Phi}$ in its principal coordinates takes the form $\overline{\Phi} = \chi \mathbf{I}$, and simplifies (3.135) to

$$\gamma_R^e = \chi (1 - (\hat{\mathbf{n}} \cdot \hat{\mathbf{m}})^2)^{1/2} \frac{H^2}{h} \Lambda^2 (1 - \Lambda)^2 |\mathbf{a}|^2, \quad (3.136)$$

where χ is an additional material constant. Notice that the above formulation is expressed partly in material coordinates and partly in spatial coordinates. We have to reformulate this into material coordinates solely, so that we can directly implement it in the upcoming finite element computation. Based on (3.14) and (3.19) we have

$$\hat{\mathbf{n}} = \frac{\overline{\mathbf{F}}^{-T} \hat{\mathbf{N}}}{\|\overline{\mathbf{F}}^{-T} \hat{\mathbf{N}}\|}, \quad \hat{\mathbf{m}} = \frac{\overline{\mathbf{F}}^{-T} \hat{\mathbf{M}}}{\|\overline{\mathbf{F}}^{-T} \hat{\mathbf{M}}\|}, \quad h = \frac{H}{\|\overline{\mathbf{F}}^{-T} \hat{\mathbf{N}}\|}, \quad (3.137)$$

which after substitution in (3.136) give

$$\gamma_R^e = \chi \sin(\beta) H \Lambda^2 (1 - \Lambda)^2 |\mathbf{a}|^2, \quad (3.138)$$

with

$$\sin(\beta) = \left[\|\bar{\mathbf{F}}^{-T} \hat{\mathbf{N}}\|^2 - \frac{\|\bar{\mathbf{F}}^{-T} \hat{\mathbf{N}}\|}{\|\bar{\mathbf{F}}^{-T} \hat{\mathbf{M}}\|} \left(\bar{\mathbf{F}}^{-T} \hat{\mathbf{N}} \cdot \bar{\mathbf{F}}^{-T} \hat{\mathbf{M}} \right)^2 \right]^{1/2}. \quad (3.139)$$

We estimate the newly introduced material parameter χ based on the results by Stupkiewicz et al [145]. For an interface along the *habit plane* of the laminate formed by type I twinning between variants 9 and 5 in mode C, they compute the following values:

$$\gamma_F^e = 0.22 \text{ J.m}^{-2}, \quad \hat{\mathbf{n}} \cdot \hat{\mathbf{m}} = 0.604641, \quad h = 20 \text{ nm}, \quad \Lambda = 0.3101, \quad \mathbf{a} = (0.0814, 0.1161, 0.2835). \quad (3.140)$$

Pulling h back to H using (3.137)₃, we obtain $H = 20.511 \text{ nm}$ which after substitution into (3.136) leaves us with

$$\chi = 2.928 \times 10^9 \text{ J.m}^{-3}. \quad (3.141)$$

We have chosen the values in (3.140) such that χ is overestimated, because in general there is no guarantee that the laminate meets the boundary along the habit plan or even nearly so. Therefore, in our model we need to overestimate χ so it may compensate for the possibility of misorientation between the laminate and the boundary. A more precise model should take the dislocation energy into account; something we avoid to stay away from an overly sophisticated model. We also chose the boundary formation energy (3.126) to have the highest suggested value for such unforeseen effects.

3.6 Chemical energies

The ground energy of individual phases, which marks the minimum of each energy well, is often referred to as the chemical energy. As far as the material modeling is concerned, all we will need is the relative value of chemical energy of the martensite phase with respect to its parent austenite, because all the martensitic energy wells (designating variants) share the same ground energy at any given temperature [65, 66, 80].

Since we will not carry out any mesoscopic and macroscopic computation, the chemical energy is not of much use to the finite element implementation. Nevertheless, for the sake of completeness we present the relative chemical energy of NiTi martensite with respect to austenite as a function of temperature:

$$c_{\text{ma}}(\theta) = c_{\text{m}}(\theta) - c_{\text{a}}(\theta) = 0.775501 (\theta - 268.470) \text{ [MJ/m}^3], \quad (3.142)$$

where temperatures are in Kelvin. Clearly, for any given temperature this function determines the threshold for both deformation- and temperature-induced martensitic phase transformation. We will use the above formula in the next chapter to determine such a threshold.

At this point our energy formulation of the microstructure has all its contributions accounted for, and we can move to putting the model into test. Our example problems will be of two types. First, we try to reproduce the general predictions of the mathematical theory of martensitic phase transformation, which is computationally trivial but conceptually

important. This part has to do with general characteristics of the laminate energetics and morphology, such as scaling laws. Next, we implement a single microscopic grain which lies at the heart of any homogenization that may be implemented in the future.

3.7 The minimization problem

Putting all the energy contributions together, our task amounts to the following minimization problem. For given twin-compatible variants i and j and for prescribed average deformation gradient $\bar{\mathbf{F}}$ over the domain Ω , find Λ , H , and $\mathbf{a}$ such that

$$\int_{\Omega} \psi_i^{\text{el}} + \psi_j^{\text{el}} + \psi^{\text{tw}} \, dV + \int_{\Gamma} \gamma_{\Gamma}^e \, dA \rightarrow \min, \quad (3.143)$$

where

$$\begin{aligned} \psi_i^{\text{el}} = & \frac{1}{6\Lambda_i} \left\| \Lambda_i H(\bar{\mathbf{a}} \otimes \nabla) + H\bar{\mathbf{a}} \otimes (\Lambda_i \nabla) \right\|_{\mathbb{C}_i}^2 - \frac{1}{6\Lambda_i} \langle \Lambda_i H(\bar{\mathbf{a}} \otimes \nabla), H\bar{\mathbf{a}} \otimes (\Lambda_i \nabla) \rangle_{\mathbb{C}_i}, \\ & + \frac{1}{2\Lambda_i} \langle (\Lambda_i H\bar{\mathbf{a}}) \otimes \nabla \pm \bar{\mathbf{a}} \otimes \hat{\mathbf{N}} + \Lambda_i (\bar{\mathbf{F}} - \mathbf{F}_i^{\circ}), \Lambda_i \bar{\mathbf{a}} \otimes (H\nabla) \pm \bar{\mathbf{a}} \otimes \hat{\mathbf{N}} + \Lambda_i (\bar{\mathbf{F}} - \mathbf{F}_i^{\circ}) \rangle_{\mathbb{C}_i}, \end{aligned} \quad (3.64)$$

and

$$\psi^{\text{tw}} = \frac{2\gamma}{H}, \quad (3.125)$$

$$\gamma_{\Gamma}^e = \chi \sin(\beta) H \Lambda^2 (1 - \Lambda)^2 |\mathbf{a}|^2, \quad (3.138)$$

subject to central symmetry with respect to the grain's midpoint. Notice that this symmetry assumption might result in slightly overestimated energies but at the same time makes for considerable computational performance gain due to its inherent numerical stability. Regardless, there are compelling mathematical arguments that support this supposition [98, pp. 700].

The grain boundary formation energy γ_{Γ}^a , and the chemical energies of the martensite variants have been left out because they are independent of the minimization variables Λ , H and $\mathbf{a}$, and can be added to the resultant energy afterwards, whenever needed.

Chapter 4

Finite element formulation

Mathematical models of physical phenomena are almost always expressed in either of the following forms:

- a boundary value problem (BVP),
- a minimization problem (MP), or
- a variational problem (VP).

In almost all cases, especially as far as the mechanics of solids is concerned, one can show that these formulations are equivalent under certain *regularity requirements*, not particularly restrictive of the class of physical problems we are considering. In order to demonstrate the equivalence of these formulations one usually goes by a cyclic scheme; namely, starting from the initial-boundary value problem, conclude the minimization formulation. Next, from the minimization formulation arrive at the variational formulation. And finally, from the variational formulation conclude the differential formulation. The last step poses additional regularity conditions on the solutions. It is in this vein that the variational formulation is also called the *weak formulation* since the solution of the variational problem poses *less restrictive* regularity conditions on the solution as compared to the initial-boundary value problem which relies on the expression of the problem in terms of differential equations. For an accessible exposition refer to Johnson's textbook [86, pp. 15].

Formulation of a finite element scheme for any particular problem is constructed upon variational (weak) formulation of the problem. As we just mentioned, whether or not the problem is originally expressed as a variational problem, we can always establish one. As it happens, our continuum model of the laminate microstructure appears in the form of a minimization problem. Since its full representation would obscure the underlying ideas in the forthcoming derivations, we opt for a compact general form for clarity's sake. An extensive derivation is committed to the sequel. This chapter borrows to various extents from the books by Daya Reddy [132], Wriggers [155], Zienkiewicz [160], Dacorogna [40, 41], and Johnson [86].

4.1 The minimization problem

We seek the minimizer(s) of the free energy

$$\Psi[\mathbf{u}] = \int_{\Omega} \psi(\mathbf{u}, \mathbf{u}\nabla; \mathbf{X}) \, dV + \int_{\partial\Omega} \gamma(\mathbf{u}; \mathbf{X}) \, dA, \quad (4.1)$$

in the vector field $\mathbf{u} = [\bar{\mathbf{a}}(\mathbf{X})^T, H(\mathbf{X}), \Lambda(\mathbf{X})]^T \in \mathcal{V}(\Omega)$ over the elastic domain $\Omega \subset \mathbb{R}^d$, where

$$\mathcal{V}(\Omega) = \left\{ \mathbf{u} \in W^{1,2}(\Omega) \left| \begin{array}{l} h \in (0, H_{\Omega}] , \quad \mathbf{X}|_{\partial\Omega} = \langle \mathbf{F} \rangle_{\Omega} \cdot \mathbf{X}, \\ \lambda \in [0, 1] , \quad \bar{\mathbf{a}} = -\Lambda(1 - \Lambda)\mathbf{a} \end{array} \right. \right\}, \quad (4.2)$$

for which the average deformation gradient $\langle \mathbf{F} \rangle_{\Omega}$ is given and the coordinates $\mathbf{X}$ are chosen such that $\hat{\mathbf{X}}_3$ is normal to the twinning plane among all possible twin-compatible (coherent) variants. $W^{1,2}(\Omega)$ is the Sobolev space augmented by a suitable energy norm $\|\cdot\|_{\mathcal{C}}$.

Notice that the problem at hand is a continuous minimization problem with respect to $\mathbf{u}(\mathbf{X})$, and at the same time a discrete minimization problem with respect to the choice of twin-compatible variants. We remember that in mechanical twinning the parent austenite phase has a fixed (possibly random) orientation. We also recall that the orientation of the possible twinning planes, designated as $\hat{\mathbf{X}}_3$, is determined merely with respect to the parent phase. Hence, in order to solve the above *micro minimization problem*, one has to repeat solving a continuous minimization problem for each discrete choice of twin-compatible variants, and then choose the minimal solution among the discrete outcomes. In summary we intend to solve the following problem

$$\mathbf{u} = \arg \min_{\mathbf{u}, \mathbf{F}_1^{\circ}, \mathbf{F}_2^{\circ}} \Psi(\mathbf{u}), \quad \mathbf{u} \in \mathcal{V}(\Omega), \quad \mathbf{F}_2^{\circ} \parallel \mathbf{F}_1^{\circ}, \quad (4.3)$$

where we have introduced the compatibility relation $\mathbf{F}_2^{\circ} \parallel \mathbf{F}_1^{\circ}$ with the usual meaning that there exist $\mathbf{a}^{\circ}$ and the unit vector $\hat{\mathbf{N}}$ such that $\mathbf{F}_2^{\circ} = \mathbf{F}_1^{\circ} + \mathbf{a}^{\circ} \otimes \hat{\mathbf{N}}$. Reformulating the minimization problem into a variational one involves functional differentiation which generalizes the concept of derivative from functions to functionals.

4.2 Functional differentiation

The directional derivative of functions can be generalized to a real functional $F[\mathbf{u}] : \mathcal{V}(\Omega) \mapsto \mathbb{R}$, by

$$\delta_{\mathbf{w}} F[\mathbf{u}] := \lim_{\epsilon \rightarrow 0} \frac{F[\mathbf{u} + \epsilon \mathbf{w}] - F[\mathbf{u}]}{\epsilon}, \quad \mathbf{w} \in \mathcal{V}(\Omega). \quad (4.4)$$

In general, $X(\Omega)$ is a properly chosen function space and the above functional differentiation goes by the name *Gâteaux derivative*. A formal definition of Gâteaux derivative is well beyond the scope of this work due to the requirements of mathematical rigor. Nevertheless,

our representation here shall be sufficient for most practical purposes. The above definition can be restated as an ordinary derivative, viz

$$\delta_{\mathbf{w}} F[\mathbf{u}] = \left. \frac{d}{d\epsilon} F[\mathbf{u} + \epsilon \mathbf{w}] \right|_{\epsilon=0}. \quad (4.5)$$

If this derivative exists for all directions $\mathbf{w}$, then we can proceed further to generalize differentials from functions to functionals. Let us define the so called *variation* of the functional $F[\mathbf{u}]$ as

$$\delta F[\mathbf{u}] = \lim_{\epsilon \rightarrow 0} F[\mathbf{u} + \epsilon \mathbf{w}] - F[\mathbf{u}], \quad \mathbf{w} \in V. \quad (4.6)$$

It then makes sense to identify $\delta \mathbf{u} = \lim_{\epsilon \rightarrow 0} \epsilon \mathbf{w}$, as the variation of $\mathbf{u} \in V$, and write

$$\delta F[\mathbf{u}] = \frac{\delta F}{\delta \mathbf{u}} \cdot \delta \mathbf{u}. \quad (4.7)$$

$\delta F/\delta \mathbf{u}$ is called the functional or variational derivative of F . It is important to bear in mind that if the function space $\mathcal{V}(\Omega)$ poses Dirichlet boundary conditions on its members, then the variation $\delta \mathbf{u}$ must vanish at the Dirichlet boundary in order for $\mathbf{u} + \delta \mathbf{u}$ to belong to $\mathcal{V}(\Omega)$. Hence, it must hold that $\delta \mathbf{u} \in \mathcal{V}_{\Gamma_D}(\Omega)$, where the newly introduced space $\mathcal{V}_{\Gamma_D}$ is the subspace of V whose members vanish at the Dirichlet boundary $\Gamma_D \subseteq \partial\Omega$.

4.3 Variational formulation

A variational (weak) formulation of our minimization problem (4.3) is based on the idea that if $\mathbf{u}$ is a solution, then any variation of the total energy around $\mathbf{u}$ must vanish, which is to say

$$\delta \Psi[\mathbf{u}] = \Psi[\mathbf{u} + \delta \mathbf{u}] - \Psi[\mathbf{u}] = 0, \quad \forall \delta \mathbf{u} \in W_{\Gamma_D}^{1,2}(\Omega). \quad (4.8)$$

Let us apply this condition to (4.1):

$$\delta \Psi[\mathbf{u}] = \int_{\Omega} \frac{\delta \psi(\mathbf{u}, \mathbf{u} \nabla; \mathbf{X})}{\delta \mathbf{u}} \cdot \delta \mathbf{u} \, dV + \int_{\partial\Omega} \frac{\delta \gamma(\mathbf{u}; \mathbf{X})}{\delta \mathbf{u}} \cdot \delta \mathbf{u} \, dA = 0. \quad (4.9)$$

We have taken the liberty to move the variational derivative $\delta \bullet / \delta \mathbf{u}$ into the integral, since the integration domains Ω and $\partial\Omega$ are independent of the variation $\delta \mathbf{u}$. Following through with the derivation, we have

$$\delta \Psi[\mathbf{u}] = \int_{\Omega} \left(\frac{\partial \psi}{\partial \mathbf{u}} \cdot \frac{\delta \mathbf{u}}{\delta \mathbf{u}} + \frac{\partial \psi}{\partial (\mathbf{u} \nabla)} : \frac{\delta (\mathbf{u} \nabla)}{\delta \mathbf{u}} \right) \cdot \delta \mathbf{u} \, dV + \int_{\partial\Omega} \frac{\partial \gamma}{\partial \mathbf{u}} \cdot \frac{\delta \mathbf{u}}{\delta \mathbf{u}} \cdot \delta \mathbf{u} \, dA \quad (4.10a)$$

$$= \int_{\Omega} \frac{\partial \psi}{\partial \mathbf{u}} \cdot \delta \mathbf{u} + \underbrace{\frac{\partial \psi}{\partial (\mathbf{u} \nabla)} : ([\delta \mathbf{u}] \nabla)}_{(\delta \mathbf{u} \cdot \frac{\partial \psi}{\partial (\mathbf{u} \nabla)}) \cdot \nabla - (\frac{\partial \psi}{\partial (\mathbf{u} \nabla)} \cdot \nabla) \cdot \delta \mathbf{u}} \, dV + \int_{\partial\Omega} \frac{\partial \gamma}{\partial \mathbf{u}} \cdot \delta \mathbf{u} \, dA \quad (4.10b)$$

$$= \int_{\Omega} \left(\frac{\partial \psi}{\partial \mathbf{u}} - \frac{\partial \psi}{\partial (\mathbf{u} \nabla)} \cdot \nabla \right) \cdot \delta \mathbf{u} \, dV + \underbrace{\int_{\partial\Omega} \left(\delta \mathbf{u} \cdot \frac{\partial \psi}{\partial (\mathbf{u} \nabla)} \right) \cdot \nabla \, dV}_{\int_{\partial\Omega} (\delta \mathbf{u} \cdot \frac{\partial \psi}{\partial (\mathbf{u} \nabla)}) \cdot \mathbf{n} \, dA} + \int_{\partial\Omega} \frac{\partial \gamma}{\partial \mathbf{u}} \cdot \delta \mathbf{u} \, dA \quad (4.10c)$$

$$= \int_{\Omega} \left(\frac{\partial \psi}{\partial \mathbf{u}} - \frac{\partial \psi}{\partial (\mathbf{u} \nabla)} \cdot \nabla \right) \cdot \delta \mathbf{u} \, dV + \underbrace{\int_{\partial \Omega} \left(\frac{\partial \psi}{\partial (\mathbf{u} \nabla)} \cdot \mathbf{n} + \frac{\partial \gamma}{\partial \mathbf{u}} \right) \cdot \delta \mathbf{u} \, dA}_{\int_{\Gamma_N} (\bullet) \cdot \delta \mathbf{u} \, dA + \int_{\Gamma_D} (\bullet) \cdot \delta \mathbf{u} \, dA} = 0. \quad (4.10d)$$

In the second step (4.10b) some vector calculus wizardry is performed. In the third step (4.10c) the volume integral is transformed to surface integral using the Gauß theorem. In the last step (4.10d) the surface integral is first split over Dirichlet and Neumann boundaries, with the implicit necessary assumption that

$$\Gamma_D \cup \Gamma_N = \partial \Omega, \quad \Gamma_D \cap \Gamma_N = \emptyset. \quad (4.11)$$

Next, the surface integral over the Dirichlet boundary vanishes due to the basic requirement of the variational formulation (4.8). Thus, the variational form of the problem can be summarized as

$$\delta \Psi[\mathbf{u}] = \int_{\Omega} \delta \mathbf{u} \cdot \left(\frac{\partial \psi}{\partial \mathbf{u}} - \frac{\partial \psi}{\partial (\mathbf{u} \nabla)} \cdot \nabla \right) \, dV + \int_{\Gamma_N} \delta \mathbf{u} \cdot \left(\frac{\partial \psi}{\partial (\mathbf{u} \nabla)} \cdot \mathbf{n} + \frac{\partial \gamma}{\partial \mathbf{u}} \right) \, dA = 0. \quad (4.12)$$

We have permuted the variation $\delta \mathbf{u}$ to the left, which is allowed inner product of vectors is symmetric, although there is no a priori justification for doing so. This is really a mere provision, for the subsequent formulation of finite element, that will simplify the matrix operations down the road.

4.4 Boundary value problem: Euler–Lagrange equations

Let us digress and briefly address a notable implication of the variational formulation, namely the Euler–Lagrange equations. In derivation of the variational formulation (4.12) we assumed no a priori restriction on the nabla operator; and the fact that we have used ∇ instead of ∇ is solely because of particularities of the laminate microstructure. In fact the above equations remain true for a large class of problems as long as a few fundamental thermodynamic assumptions are not violated. So we may change the coordinates from laminate coordinates to general coordinates, and the planar nabla operator ∇ to its general counterpart ∇ , and the same formulation will then remain valid for such general problems:

$$\delta \Psi[\mathbf{u}] = \int_{\Omega} \delta \mathbf{u} \cdot \left(\frac{\partial \psi}{\partial \mathbf{u}} - \frac{\partial \psi}{\partial (\mathbf{u} \nabla)} \cdot \nabla \right) \, dV + \int_{\Gamma_N} \delta \mathbf{u} \cdot \left(\frac{\partial \psi}{\partial (\mathbf{u} \nabla)} \cdot \mathbf{n} + \frac{\partial \gamma}{\partial \mathbf{u}} \right) \, dA = 0, \quad (4.13)$$

$$\mathbf{u} \in \{ W^{1,2}(\Omega) \mid \mathbf{u}(\mathbf{x}) = \mathbf{u}_0, \mathbf{x} \in \Gamma_D \}, \quad \forall \delta \mathbf{u} \in L^2_{\Gamma_D}(\Omega),$$

wehre $\psi = \psi(\mathbf{u}, \mathbf{u} \nabla; \mathbf{x})$ and $\gamma = \gamma(\mathbf{u}; \mathbf{x})$.

According to the *fundamental lemma of variational calculus*, this can be true only if the co-operand of the variation is identically zero, which is to say

$$\frac{\partial \psi}{\partial \mathbf{u}} - \frac{\partial \psi}{\partial (\mathbf{u} \nabla)} \cdot \nabla = \mathbf{0}, \quad \forall \mathbf{x} \in \Omega, \quad (4.14a)$$

$$\frac{\partial \psi}{\partial(\mathbf{u}\nabla)} \cdot \mathbf{n} + \frac{\partial \gamma}{\partial \mathbf{u}} = \mathbf{0}, \quad \forall \mathbf{x} \in \Gamma_N. \quad (4.14b)$$

From a thermomechanical viewpoint, the above equations express the balance of thermodynamic forces over the body Ω and at the Neumann (natural) boundary Γ_N respectively. From a mathematical viewpoint, on the other hand, the above equations are the statement of the underlying boundary value problem up to the Dirichlet (essential) boundary condition(s).

Remark 15. NATURAL BOUNDARY CONDITIONS. The equation (4.14b) has the apparent form of a boundary condition; one that is not imposed a priori, but rather deduced directly from the variational formulation, which in turn is the direct consequence of the minimization principle. Such boundary conditions are called *natural boundary conditions*.

One has to take great care when handling the natural boundary conditions in a finite element scheme. Within the present context, boundary conditions can be classified as

- natural boundary conditions, and
- essential boundary conditions.

When finite element implementation is concerned, essential boundary conditions are imposed explicitly and are, therefore, strictly fulfilled. Natural boundary conditions are, on the other hand, only approximately satisfied depending on the order of finite element scheme and on the quadrature rule used in numerical integration.

Any part of the boundary on which no explicit condition is imposed becomes natural boundary automatically. For a solid body this will be a traction-free boundary—also known as the homogeneous Neumann boundary condition.

Deducing the Euler–Lagrange equations (4.14) from the variational formulation places an additional mathematical restriction on the problem. Namely, in moving from (4.10b) to (4.10c) the order of differentiation increases from one to two when the following identity is applied

$$\frac{\partial \psi}{\partial(\mathbf{u}\nabla)} : ([\delta \mathbf{u}] \nabla) = \left(\delta \mathbf{u} \cdot \frac{\partial \psi}{\partial(\mathbf{u}\nabla)} \right) \cdot \nabla - \frac{\partial \psi}{\partial(\mathbf{u}\nabla)} \cdot \nabla$$

Up to this additional differentiability requirement and the Dirichlet boundary conditions, we have implicitly shown the equivalence of the variational formulation (4.9) with the system of differential equations (4.14). Since we already have demonstrated the equivalence of the minimization problem and the variational problem, one can summarize the results schematically as

$$[\text{Boundary Value Problem}] \rightleftarrows [\text{Variational Problem}] \Leftrightarrow [\text{Minimization Problem}]$$

Remark 16. The fact that the BVP is more restrictive in terms of regularity of the solution can be a severe setback for some mathematical problems. This is because the function space in which the solution of (4.14) is sought, i.e., $C^2(\Omega)$, is much smaller than that

of (4.9), namely the Sobolev space. A Sobolev space is defined based on the Lebesgue norms and weak derivatives, which extends the realm of regularity well beyond continuous differentiability. This implies that the variational formulation can cover a substantially larger class of problems than its corresponding boundary value problem. The solutions to VP are often called *weak solutions* as opposed to *strong solutions* of PDE's, because the *weak formulation* (VP) poses less restrictive (weaker) regularity requirements on the set of solutions.

4.5 The Ritz–Galerkin method

The function spaces in which the solutions to either the minimization problem, the boundary value problem, or the variational problem are sought are infinite dimensional. For some problems we might be able to find analytical solutions which are often expressed in terms of infinite series over a particular orthonormal basis of the underlying function space, such as Fourier, Legendre, or Bessel functions. However, these situations are rare and amount to a very limited subset of the problems we face in any branch of physical sciences. For practical purposes in the end, what we look for is numerical solution to a given problem that can be tuned to yield results with required precisions.

The basic idea behind most numerical schemes is the same; namely, we substitute the function space of solutions $\mathcal{V}(\Omega)$ with a suitably chosen finite dimensional subspace $\tilde{\mathcal{V}}(\Omega)$, with dimension N and basis functions $\mathcal{B} = \{\phi_1, \phi_2, \dots, \phi_N\}$. This reduces the original problem to an algebraic one of finding the finite decomposition of a possible solution

$$\tilde{\mathbf{u}} = \sum_{i=1}^{\#\mathcal{B}} u_i \phi_i, \quad (4.15)$$

in terms of u_i 's. For instance, difference schemes and collocation methods encompass two broad family of methods based on application of this approach directly to boundary value problems, or minimization problems.

If we intend to employ the same technique to numerically solve the variational problem (4.13), we notice that in addition to the solutions' function space there is the variations' function space which also has to be substituted by one of its suitably chosen finite dimensional subspaces. In general, there are problems in which the variation space has to be chosen differently than the solution space. This is called the *Galerkin method*. In most circumstances, however, the two spaces coincide except that the variation space has to vanish at Dirichlet boundary. This will be then called the *Ritz method*. Obviously, the Ritz method is a subset of the Galerkin's. Nevertheless, since the underlying ideas are only subtly different, they are unified under the common class of “Ritz–Galerkin methods”.

The discrete Galerkin formulation of the variational problem (4.13) takes on the form

$$\arg_{\tilde{\mathbf{u}}} \left\{ \tilde{\mathbf{u}} \in \tilde{\mathcal{V}}(\Omega) \mid \delta\Psi[\tilde{\mathbf{u}}] = 0, \forall \delta\tilde{\mathbf{u}} \in \tilde{\mathcal{V}}_{\Gamma_D}(\Omega) \right\}. \quad (4.16)$$

The functional variation $\delta\tilde{\Psi}[\tilde{\mathbf{u}}]$ vanishes along every function variation $\delta\tilde{\mathbf{u}} \in \tilde{\mathcal{V}}_{\Gamma_D}$, if and only if it vanishes along all basis functions of $\tilde{\mathcal{V}}_{\Gamma_D}$, denoted by $\{\bar{\phi}_1, \dots, \bar{\phi}_N\}$. So we can rewrite the above problem in a more accessible form, namely

$$\arg \left\{ \delta_{\bar{\phi}_i} \Psi[u_j \phi_j] = 0, \forall \bar{\phi}_i \right\}, \quad (4.17)$$

which is a system of N algebraic equations, each corresponding to one of $\bar{\phi}_i$'s, to be solved for the vector of unknowns $\mathbf{U} = (u_1, u_2, \dots, u_N)^T$.

In deriving the Euler–Lagrange equations from the variational problem (4.10) we passed through two intermediate steps. Due to algebraic considerations, which will become clear later, the Galerkin method is best formulated in terms of (4.10b), that is,

$$\delta\Psi[\mathbf{u}] = \int_{\Omega} \delta\mathbf{u} \cdot \frac{\partial\psi}{\partial\mathbf{u}} + (\delta\mathbf{u})\nabla : \frac{\partial\psi}{\partial(\mathbf{u}\nabla)} dV + \int_{\Gamma_N} \delta\mathbf{u} \cdot \frac{\partial\gamma}{\partial\mathbf{u}} dA. \quad (4.18)$$

4.5.1 Galerkin method for elliptic problems

If the underlying PDE is *elliptic*, as is the case with linear elasticity, it appears as

$$\begin{aligned} -(\mathbb{A} : \mathbf{u}\nabla) \cdot \nabla + \mathbf{f} \cdot \mathbf{u} &= \mathbf{0}, & \mathbf{x} \in \Omega, \\ \mathbf{u}\nabla \cdot \mathbf{n} &= \mathbf{g}, & \mathbf{x} \in \Gamma_N, \\ \mathbf{u} &= \mathbf{u}_0, & \mathbf{x} \in \Gamma_D, \end{aligned} \quad (4.19)$$

where $\mathbb{A}$ is a symmetric positive definite fourth-order tensor.

Remark 17. Elliptic operators are linear. Hence, if an elliptic differential equation must satisfy non-homogeneous Dirichlet boundary conditions, $\mathbf{u}|_{\Gamma_D} = \mathbf{u}_0$, one can re-parameterize $\mathbf{u}$ as $\mathbf{u}^* = \mathbf{u} - \mathbf{u}_0$, and reformulate the problem into a homogenous one; in which case the variation space $\mathcal{V}_{\Gamma_D}(\Omega)$ and the solution space $\mathcal{V}(\Omega)$ coincide. This is why most references do not distinguish between the two spaces and call them under a common name, as either *test functions* or *shape functions*. So in writing the Galerkin formulation (4.17) we assume that $\bar{\phi}_i = \phi_i$ for all i , granted the above re-parameterization.

Based on (4.18) the system of equations (4.17) assumes the following form:

$$\int_{\Omega} -\phi_i \cdot \mathbf{f} + (\phi_i \nabla : \mathbb{A} : \phi_j \nabla) u_j dV + \int_{\Gamma_N} \phi_i \cdot \mathbf{g} dA = 0, \quad (4.20)$$

which is a linear system of equations of the form

$$\mathbf{K} \mathbf{U} = \mathbf{R}, \quad (4.21)$$

in which $\mathbf{K}$ is historically called the stiffness matrix and is given by

$$K_{ij} = \int_{\Omega} \phi_i \nabla : \mathbb{A} : \phi_j \nabla dV, \quad (4.22)$$

and $\mathbf{U}$, the vector of unknowns, has been already introduced. The so called *load vector* is given by

$$R_i = \int_{\Omega} \phi_i \cdot \mathbf{f} \, dV - \int_{\Gamma_N} \phi_i \cdot \mathbf{g} \, dA. \quad (4.23)$$

Solving (4.21) for large systems, with many *degrees of freedom*, is usually carried out via numerically efficient algorithms called *linear solvers*, or more precisely sparse linear solvers in the case of finite elements, which we will present shortly.

Comparing (4.19) with the individual terms of the Euler–Lagrange equations (4.14), we find the following physical interpretations

$$-\frac{\partial \psi}{\partial \mathbf{u}} = \mathbf{f} : \text{body force}, \quad \frac{\partial \psi}{\partial (\mathbf{u} \nabla)} = \boldsymbol{\sigma} : \text{Cauchy stress}, \quad -\frac{\partial \gamma}{\partial \mathbf{u}} = \mathbf{t} : \text{boundary traction}. \quad (4.24)$$

Note that within the framework of linear elasticity, since the deformations are infinitesimal, the domain Ω itself does *not* evolve in time.

4.5.2 Galerkin method for nonlinear problems

In problems where $\delta \Psi$ is a nonlinear functional, the system of equations (4.17) becomes nonlinear. Numerical solution of general nonlinear algebraic equations is most often realized through *Newton–Raphson method* or one of its customized variants. Focusing on the vector of unknowns $\mathbf{U}$, we have to solve a system of equations of the form $\mathbf{F}(\mathbf{U}) = \mathbf{0}$, which is handled recursively as follows.

1. Start with a suitably chosen *initial guess* $\mathbf{U}^0$.
2. Having the solution $\mathbf{U}^n$ for the n th step, we have $\mathbf{U}^{n+1} = \mathbf{U}^n + \Delta \mathbf{U}^n$, where the increment $\Delta \mathbf{U}^n$ is obtained by solving the linear system of equations

$$\left. \frac{\partial \mathbf{F}}{\partial \mathbf{U}} \right|_{\mathbf{U}^n} \cdot \Delta \mathbf{U}^n = -\mathbf{F}(\mathbf{U}^n). \quad (4.25)$$

3. If $\|\Delta \mathbf{U}^n\| < \epsilon$, then stop, otherwise repeat the step 2. Here, ϵ is a properly chosen tolerance.

Apparently, the most important piece in the above algorithm is $\partial \mathbf{F} / \partial \mathbf{U}$, the so called *tangent stiffness*. The convergence of the method and its rate depends very much on *condition number* of the tangent stiffness. A neatly modified version of the N–R algorithm (see Goldstein 1962 [39]) tries to damp or accelerate the step size in order to ensure convergence with near-optimal rate. So in step 2, the increment assumes the following form

$$\mathbf{U}^{n+1} = \mathbf{U}^n + \varrho^n \Delta \mathbf{U}^n. \quad (4.26)$$

There are various approaches to calculating ϱ^n . We do not give further details but the numerical library we use to implement finite element has built-in procedures for this.

The tangent stiffness based on (4.18) is calculated via

$$\begin{aligned} K_{ij}^n = & \int_{\Omega} \left\{ \phi_i \cdot \left[\frac{\partial^2 \psi}{\partial \mathbf{u} \otimes \partial \mathbf{u}} \right]_{\mathbf{U}^n} \cdot \phi_j + \phi_i \nabla : \left[\frac{\partial^2 \psi}{\partial (\mathbf{u} \nabla) \otimes \partial (\mathbf{u} \nabla)} \right]_{\mathbf{U}^n} : \phi_j \nabla \right. \\ & \left. + \phi_i \cdot \left[\frac{\partial^2 \psi}{\partial \mathbf{u} \otimes \partial (\mathbf{u} \nabla)} \right]_{\mathbf{U}^n} : \phi_j \nabla + \phi_i \nabla : \left[\frac{\partial^2 \psi}{\partial (\mathbf{u} \nabla) \otimes \partial \mathbf{u}} \right]_{\mathbf{U}^n} \cdot \phi_j \right\} dV \\ & + \int_{\Gamma_N} \phi_i \cdot \left[\frac{\partial^2 \gamma}{\partial \mathbf{u} \otimes \partial \mathbf{u}} \right]_{\mathbf{U}^n} \cdot \phi_j dA, \end{aligned} \quad (4.27)$$

and the right hand side based on

$$R_i^n = - \int_{\Omega} \left\{ \phi_i \cdot \left[\frac{\partial \psi}{\partial \mathbf{u}} \right]_{\mathbf{U}^n} + \phi_i \nabla : \left[\frac{\partial \psi}{\partial (\mathbf{u} \nabla)} \right]_{\mathbf{U}^n} \right\} dV - \int_{\Gamma_N} \phi_i \cdot \left[\frac{\partial \gamma}{\partial \mathbf{u}} \right]_{\mathbf{U}^n} dA. \quad (4.28)$$

Most notably, the tangent stiffness is symmetric for the particular form of energy (4.9) we are dealing with. It is important to know, however, that this may not be the case in general. For the problem at hand, the symmetry of the tangent stiffness will enable us to use particular linear solvers which would not be possible otherwise.

Once again we have allowed $\mathcal{V}_{\Gamma_D}(\Omega)$ to be the same as $\mathcal{V}(\Omega)$, which might seem implausible for problems with non-homogeneous Dirichlet boundary conditions as the variational form is not linear anymore. However, since the nonlinear problem is solved by stepwise linearization of the residual (4.25) via N-R algorithm, we can homogenize the Dirichlet boundary conditions, if any, as we did for the elliptic boundary value problem.

4.6 The finite element method

The computation cost of solving a linear system of equations, like (4.21) or (4.25), is determined by the size of $\mathbf{U}$ which equals the number of unknowns, also referred to as the number of degrees of freedom (DOFs), which also equals the dimension of $\tilde{\mathcal{V}}(\Omega)$. Assuming that the solution to the weak formulation exists and is unique, we can expect that

$$\lim_{n \rightarrow \infty} \sum_{i=1}^{\#\mathcal{B}} u_i \phi_i \rightharpoonup \mathbf{u}. \quad (4.29)$$

Then we can roughly say that increasing the number of DOFs will yield a more precise estimate to the actual solution. Clearly when the underlying phenomenon that is being modeled becomes more complex, we have to increase the number of DOFs in order to find a reasonable estimate to the supposed solution. In real-world applications the required number of DOFs can grow as high as millions, and handling linear systems of that size can be extremely demanding if not impossible. This is where a clever choice of test functions can change the game entirely, and that is what finite element method (FEM) is all about.

The finite element method chooses the test functions ϕ_i 's in a way that the stiffness matrix becomes sparse, i.e., to have many zero elements. In linear algebra we often represent the solution to a linear system like (4.21) symbolically as $\mathbf{U} = \mathbf{K}^{-1}\mathbf{R}$. In reality, however,

we do not invert the stiffness matrix $\mathbf{K}$ and then multiply it by $\mathbf{R}$ to obtain the solution, except for illustrative purposes in small-sized system of equations. There are highly efficient linear solvers which take advantage of particular structures of the stiffness matrix. The most efficient of these algorithms are the sparse linear solvers for obvious reasons. Now let us see how FEM manages to make $\mathbf{K}$ sparse.

4.6.1 Finite elements

Let $\mathcal{T}_h(\Omega)$ be a partitioning of the domain the domain Ω into M subdomains as

$$\mathcal{T}_h(\Omega) = \left\{ \tau_i \subset \Omega \mid \bigcup_{i=1}^{\#\mathcal{T}_h} \tau_i \approx \Omega; \dim(\tau_i \cap \tau_j) < \dim(\Omega), i \neq j \right\}, \quad (4.30)$$

such that every τ_i is often a *simplex*, that is, a line segment if $\Omega \subseteq \mathbb{R}$, a triangle if $\Omega \subseteq \mathbb{R}^2$, or a tetrahedron if $\Omega \subseteq \mathbb{R}^3$. $\mathcal{T}_h(\Omega)$ is often called a *triangulation* or a *mesh*. Each τ_i is called an *element* and its corners are called *nodes*. Although, additional nodes can be added to elements without changing their shape in order to assign more test functions or to increase their order. The nodes that appear on the edges or faces of an element, due to refinement of the neighboring elements via an *adaptive scheme*, and the element has not assigned test functions to them are called *hanging nodes*.

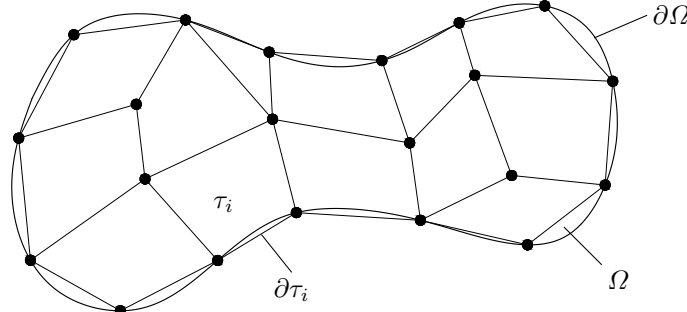


Fig. 4.1: A finite element mesh of linear polygonal elements.

In general, the union of all τ_i 's covers Ω only approximately. This is because the boundary $\partial\Omega$ is neither necessarily polygonal nor a parameterized curve (Fig. 4.1). Surely, if that were the case, then the mesh would cover the domain completely. Nevertheless, this approximative coverage of Ω should not be a concern, because in principle for a well-defined and regular domain Ω one should be able to improve the approximation by choosing a finer mesh, that is,

$$\lim_{N_e \rightarrow \infty} \bigcup_{i=1}^{\#\mathcal{T}_h} \tau_i \rightarrow \Omega. \quad (4.31)$$

The choice of a simplex as the building block of $\mathcal{T}_h$ is not mandatory. Another common option is quadrilateral element in 2D and its 3D counterpart the hexahedral element. Simplex elements, however, are commonly preferred in mathematical theory of finite elements because

they are inherently convex. They are, however, harder to implement in adaptive finite element schemes, where the mesh has to be refined at places of high gradient, which is why quadrilateral elements are sometimes preferred for adaptivity. One can use interpolants of orders greater than one to generate elements with curved boundaries, which is rarely done because the computation cost of having a finer mesh of standard polygonal or polyhedral elements is almost equal, while being easier to implement.

The core assumption of the finite element method is that the i th test function ϕ_i is defined corresponding to the i th node such that it is nonzero within the elements that share the node and zero otherwise. As the result of this particular choice, many of the elements of the stiffness matrix $\mathbf{K}$ will become zero. This can be understood by revisiting either (4.22) or (4.27). In both cases each K_{ij} is a bilinear form of ϕ_i and ϕ_j . If the supports of ϕ_i and ϕ_j have no intersection, then the integrals evaluate to zero. Thus, for every two test functions ϕ_i and ϕ_j whose respective nodes i and j do not belong to a common element the corresponding K_{ij} vanishes. This makes for a sparse stiffness matrix which radically reduces the computation cost, especially for problems with a large number of DOFs. Despite the substantial simplification of the discrete problem, one can show that under certain conditions, not particularly restrictive, the finite element method is indeed convergent, viz, (4.29) holds (see Lax–Milgram theorem [32]).

4.6.2 Piecewise polynomial test functions and quadrature rule

The test functions in FEM are almost exclusively chosen as polynomials, and since their support is confined to elements neighboring to a node, they are defined as piecewise polynomials. The form and degree of these polynomials are dictated by the regularity requirements, such as belonging to H^1 or L_2 , which is easy to accomplish since polynomials are the simplest class of functions. They also have to be tuned to be capable of simulating the physical constraints of the underlying phenomenon, such as incompressibility of a continuum. Of course, one should always seek a proper compromise between precision and computation cost.

Other than solving large linear systems, computing the individual elements of the stiffness matrix and then storing and assembling them into $\mathbf{K}$ is among the most expensive tasks in any serious finite element computation. Computing K_{ij} 's involves integration which is usually carried out using *quadrature rules*. A quadrature rule is a weighted sum of discrete values of the integrand at optimally chosen points within the integration domain, that is, for a function $f : \Omega \rightarrow \mathbb{R}$

$$\int_{\Omega} f(x) \, dx \approx \sum_{i=1}^Q \omega_i f(x_i) , \quad x_i \in \Omega , \, \omega_i \in \mathbb{R} , \quad (4.32)$$

in which P is called the order of the quadrature.

The so called *Gauss–Legendre quadrature* is the standard choice in FEM, because for a polynomial of given degree we can choose Q to achieve exact results. The idea is that a polynomial of degree n is determined by $n + 1$ coefficients and since its integral is explicit, a quadrature rule of order Q such that $2Q \geq n + 1$ shall be able to evaluate its integral

exactly, if the weights and the quadrature points (often called Gauss points in FEM) are chosen properly. Calculating the weights and the quadrature points using orthogonality of Legendre polynomials is relatively straightforward [143, see pp. 177]. We give a rough sketch of the underlying ideas in the following.

Remark 18. Let us denote the Legendre polynomial of degree m by $\ell_m(x)$. It can be shown that every polynomial $p(x) \in \mathcal{P}_{2n+1}$ can be written as $p(x) = q(x) \ell_{n+1}(x) + r(x)$, where $q(x), r(x) \in \mathcal{P}_n$ and $r(x) = \sum_{i=0}^n r(x_i) \ell_i(x)$ such that $\ell_{n+1}(x_i) = 0$ for all x_i . Then, for every x_i we have $f(x_i) = r(x_i)$. Consequently, $r(x) = \sum_{i=0}^n f(x_i) \ell_i(x)$, and therefore

$$\int_a^b p(x) dx = \underbrace{\int_a^b q(x) \ell_{n+1}(x) dx}_{=0} + \int_a^b r(x) dx = \sum_{i=0}^n f(x_i) \int_a^b \ell_i(x) dx.$$

which expresses the Gauss–Legendre quadrature (4.32), with the weight factors given by $\omega_i = \int_a^b \ell_i(x) dx$, and shows that of G–L quadrature of order n is exact for polynomials up to degree $2n + 1$.

The same ideas can be generalized to domains of higher dimensions $\Omega \subseteq \mathbb{R}^d$. We do not give further details except for one remark. The Gaussian quadrature is based on the field values at the Gauss points whereas the finite element solution computes the nodal values of the field variables. Therefore, interpolation procedures have to be put in place in order to convert the nodal values to Gauss point values.

Physical versus natural coordinates

Reusability of formulation and implementation of the finite element method is crucial, because it is generally a time consuming procedure. In this regard, one tries to make the formulation coordinate-independent wherever possible. The first step taken in this direction is to use fixed dimensions and elements within the so called *natural coordinates*. In natural coordinates the elements are defined such that their primary nodes, i.e. their corners, are positioned at coordinate values -1 , 0 , or $+1$. Then of course, we have to apply mappings from natural coordinates to the *physical coordinates* and vice versa. This mapping is done via interpolation, which can be given by the general form:

$$\mathbf{x}(\xi) = \sum_{i=1}^{\#\mathcal{N}_\tau} N_i(\xi) \mathbf{x}_i, \quad (4.33)$$

where $\mathbf{x}$ denotes the physical coordinates, ξ the natural coordinates, $N_i(\xi)$ the interpolant corresponding to the i th node, and $\mathbf{x}_i$ the position of the i th node in physical coordinates. $\mathcal{N}_\tau$ is the set of nodes belonging to the element τ . If the shape functions corresponding to each node are also used as interpolant, then the finite elements are called *isoparametric*.

Fig. 4.2 shows a quadratic triangular element, which is also an example of an element with curved boundaries. The edge nodes are used to increase the order of the shape functions, since a parabola is uniquely determined by three points. Having formulated the FEM

for the element in natural coordinates, the same formulation is applied to any triangular element which is an economic approach. The mapping of line, area, and volume elements

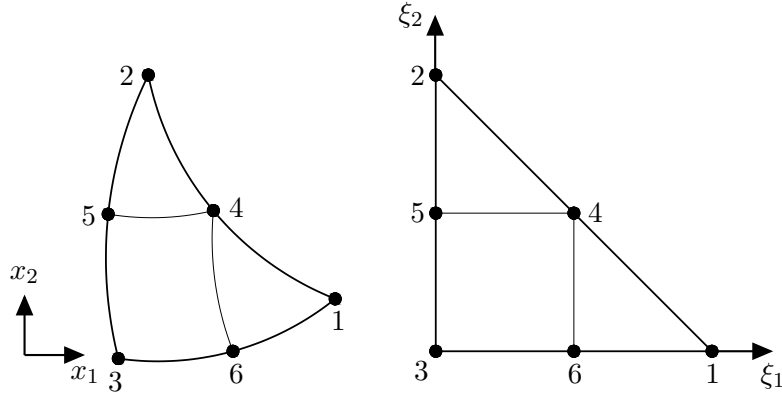


Fig. 4.2: A quadratic triangular element in physical (left) and natural (right) coordinates.

between natural and physical coordinates is performed using Jacobian $\mathbf{J} = \partial \mathbf{x} / \partial \boldsymbol{\xi}$. When the interpolants are linear, or multilinear, the Jacobian is homogeneous across the element. But in the case of curved elements, such as quadratic elements, the Jacobian varies and therefore in order to map the element size (area for instance) one has to integrate the Jacobian determinant over the element, that is, $A_e = \int_0^1 \int_0^{1-\xi_2} |\mathbf{J}(\boldsymbol{\xi})| d\xi_1 d\xi_2$. Also, integrating any field function over the element using a quadrature formula must take the value of the Jacobian determinant at the quadrature points into account:

$$\int_{\Omega_e} f dV \approx \sum_i^{\#Q} \omega_i |\mathbf{J}(\boldsymbol{\xi}_i)| f(\boldsymbol{\xi}_i) . \quad (4.34)$$

Chapter 5

Numerical and analytical results at microscale

This chapter presents some numerical and analytical results at the microscale. We first put together a general computation of the laminate microstructure for NiTi (49.75 at. % Ni). A comparison between the predictions of our model and the available experimental data would require us to bridge the length scales to the macro level, using *multiscale* finite element methods. This would require parallelization of computations due to the inherent complexity of our micromechanical model. We leave this task to the future.

We present a simplified one-dimensional problem that reflects some of the major mathematical characteristics of the model. These will be compared with well-known theoretical and empirical results mostly concerning the scaling laws that the martensitic microstructure manifests.

5.1 The basics

Before we proceed to the actual computations, let us lay down the basic assumptions on which we establish the subsequent implementation.

Choice of reference

The deformation ansatz (3.35) expresses a finite deformation. We have therefore formulated all the energy terms consistently in material coordinates so far. Since we are dealing with a geometrically nonlinear formulation, the average (mesoscopic) deformation gradient $\bar{\mathbf{F}}$ will be prescribed stepwise and the field variables Λ , H , and $\mathbf{a}$ will computationally evolve as $\bar{\mathbf{F}}$ changes. Despite this geometrical nonlinearity, the material will be linearly elastic, since the deviation of the finite strains (3.41) from the transformation strains are assumed to be infinitesimal, as given by (3.49).

Grain geometry and local martensite coordinates

We confined the laminate ansatz to first-order twins only, which lead to assuming that each grain includes one single patch of the laminate microstructure. It is therefore reasonable to assume that each grain is a *cuboid*—a rectangular parallelogram in the reference configuration. Choosing a parallelogram as the laminate domain is not uncommon, due to both empirical and theoretical evidence. But assuming a rectangular parallelogram has a particular reason, which is as follows. We divided the grain boundary energy to formation (atomistic) energy (3.126) and microstrain energy (3.138): $\gamma_T = \gamma_T^a + \gamma_T^e$. For the sake of simplicity, we also took γ_T^a to be independent of the relative orientation of the boundary and the lattice. This relative orientation is thermomechanically determined by minimizing γ_T (see for instance [144]). When γ_T^a is independent of orientation, this minimization becomes superfluous, and $\hat{\mathbf{m}} \cdot \hat{\mathbf{n}}$ is solely determined by laminate deformation. So, we assume that in the reference configuration the grain has the simplest possible geometry, which translates to $\hat{\mathbf{M}} \cdot \hat{\mathbf{N}} = 0$, and a cuboid domain.

However, this cuboid is still indeterminate up to a rotation about $\hat{\mathbf{N}}$. To fix this, we heuristically make an additional assumption; namely, the edges of the cuboid coincide with the following unit vectors:

$$\hat{\mathbf{e}}_1 = \widehat{\mathbf{F}_I^{-1} \mathbf{a}}, \quad \hat{\mathbf{e}}_2 = \hat{\mathbf{N}} \times \widehat{\mathbf{F}_I^{-1} \mathbf{a}}, \quad \hat{\mathbf{e}}_3 = \hat{\mathbf{N}}, \quad (5.1)$$

where $\mathbf{F}_J = \mathbf{F}_I + \mathbf{a} \otimes \hat{\mathbf{N}}$. Notice that $\hat{\mathbf{e}}_1$ denotes the shear direction introduced in remark 8, when we take into account the orientation of martensitic variants relative to the parent phase according (3.71) and (3.76). This choice will simplify the implementation of the laminate deformation ansatz. Using $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3)$ coordinates to express the reference configuration also makes sure that the zero gradients along $\hat{\mathbf{N}}$ keep their simple expressions, since $\hat{\mathbf{e}}_3$ coincides with $\hat{\mathbf{N}}$.

Since all the kinematic variables as well as stiffness tensors are originally expressed with respect to the parent (austenite) phase coordinates and with respect to martensite coordinates, for the sake of consistency we transform all of them first to austenite coordinates and then to grain coordinates (5.1), before substituting them into the minimization problem (3.143).

- The austenite lattice coordinates are taken as the material coordinates. If the austenite is rotated by a rotation $\mathbf{R}_o$, all vector and tensor quantities shall be transformed under $\mathbf{R}_o^T$. At the microscale we assume the austenite to be fixed. That is to say $\mathbf{R}_o = \mathbf{I}$.
- The martensite stiffness tensor is originally given with respect to the martensite lattice coordinates. These coordinates have a relative orientation to the austenite give by the rotation tensor $\mathbf{Q}_i$ obtained from polar decomposition of the transformation tensor $\mathbf{F}_i$ as expressed in equation (3.71). Furthermore, for every twin-compatible pair of martensite variants i and j , there is a small rotation $\mathbf{R}$ originating from the twinning equation (3.76). Therefore, the stiffness tensors of variants i and j must be transformed by $\mathbf{Q}_i^T$ and $\mathbf{R}^T \mathbf{Q}_j^T$, respectively.

- A nonlinear finite element scheme is often based on Newton method which requires suitable initial guess values for the field variables. Among the field variables of our problem, Λ and H are scalars and are not influenced by the choice of coordinates. The vector $\mathbf{a}$, on the other hand, shall be transformed before being set. A suitable guess value for $\mathbf{a}$ is its original value obtained from the solution of twinning equation, given in Table 3.3. However, these values have to be transformed by $\mathbf{Q}_i^T$, because they are based on (3.77), as we explained before.
- All the vector- and tensor-valued quantities must also be transformed by $(\hat{\mathbf{e}}_1^T, \hat{\mathbf{e}}_2^T, \hat{\mathbf{e}}_3^T)^T$ to the grain coordinates.

A two dimensional finite element setting

Expressed in local grain coordinates $(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3)$, the field variables are independent of X_3 . We therefore need a two dimensional finite element implementation. We take the domain to be a square whose edge length will be referred to as *grain length* or *grain size*. This square domain is partitioned using quadrilateral elements that become finer towards the boundary (Fig. 5.1). On each element, a bilinear Lagrange polynomial is used and integration is carried

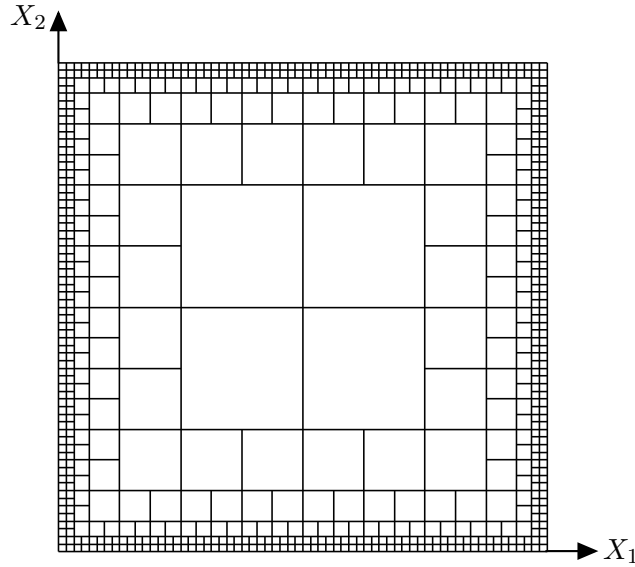


Fig. 5.1: The 2D mesh composed of quadrilateral elements refined towards boundaries.

out using two Gauss quadrature point in each coordinate direction, i.e. four Gauss points for each element.

The particular choice of element is dictated by the finite element library `deal.ii` on which we have based our implementation [14]. This leads to the appearance of the so called *hanging nodes* which is handled by enforcing the continuity condition explicitly across the neighboring elements at these nodes.

Finding energy-minimizing variants: a two-step algorithm

As soon as the austenite lattice is fixed and the average deformation gradient $\bar{\mathbf{F}}$ is prescribed, we need to determine which pair(s) of compatible variants make for the lowest energy. Although indispensable, solving the problem for each possible pair is computationally expensive. To reduce the computation time, we first run solve the problem for all the possible variants on a coarser mesh, with a larger tolerance for convergence criterion, and using fewer load steps. The computed energy densities for all the compatible pairs of martensite variants are stored and compared with each other to find the least value(s). Then, for the low-energy pairs of variants we repeat the computation on a finer mesh, with a stricter convergence criterion, and using more load steps. This two-step scheme save considerable computation time and has proved to yield reproducible results in out computations.

The physical constrains on Λ and H

The volume fraction Λ and the lath width H must fulfill

$$0 \leq \Lambda \leq 1, \quad 0 < H, \quad (5.2)$$

due to obvious physical requirements. Solving the energy minimization problem (3.143) subject to these constrains can be accomplished via the so called *Karush–Kuhn–Tucker conditions* (KKT). However, since the above constraints are in the form of simple upper and lower bounds on the nodal values of field variables, we can use a simplified version of the KKT as follows [117, pp. 358–360].

In each Newton update we check whether the updated value violates the respective constraint. If so, we rescale the increment such that the constraint becomes at most active. That is, for instance if at the n th Newton update at the i th node, we have $\Lambda_i^n + \Delta^n \Lambda_i < 0$, then we rescale $\Delta^n \Lambda_i$ by a factor

$$\alpha_i^n = \varrho \left| \frac{\Lambda_i^n}{\Delta^n \Lambda_i} \right|, \quad (5.3)$$

in which ϱ is a constant which either equals one or is slightly smaller than one. If a lower bound (upper bound) constraint becomes active, the corresponding element of the right-hand-side vector of the Newton scheme will be excluded from the convergence check and instead will be checked for negativity (positivity). For H the constraint is a strict inequality. In such cases it is common to replace the strict inequality by an inequality, adding a small residual e.g.

$$10^{-10} \leq H.$$

The grain size and orientation

A thorough study of grain size effects is not a part of this work. Nevertheless, we carried out a series of tests to calculate the energy density of the laminate, under displacement-controlled simple tension $\varepsilon = 0.0167$ in austenite lattice coordinates, over grain sizes of various orders

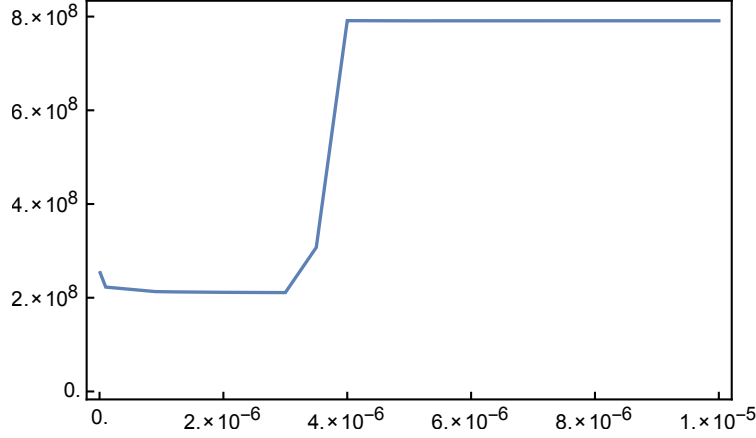


Fig. 5.2: Energy density $[\text{J.m}^{-3}]$ versus grain size in $[\text{m}]$.

of magnitude. This particular value is obtained by solving

$$\frac{1}{2} (\varepsilon \hat{\mathbf{e}}_1^a \otimes \hat{\mathbf{e}}_1^a) : \mathbb{C}_a : (\varepsilon \hat{\mathbf{e}}_1^a \otimes \hat{\mathbf{e}}_1^a) = c_{\text{ma}}(300) = 2.44513 \times 10^7 [\text{J.m}^{-3}], \quad (5.4)$$

in which $c_{\text{ma}}(300)$ is the relative ground energy of martensite with respect to austenite at 300 K, according (3.142). In theory, this should be the strain at which the overall energy (chemical plus elastic plus interface energies) of the NiTi austenite is equal to the ground (chemical) energy of the NiTi martensite, and therefore the threshold for deformation-induced martensitic transformation. Fig. 5.2 shows energy density versus grain size. Our criterion is to choose the grain size for which the energy density is minimum, which suggests an approximate value of

$$L \approx 3.0 \times 10^{-6} \text{ m}, \quad (5.5)$$

which we will be using throughout the upcoming computations.

A proper investigation of the grain size effect would at least require us to include grain orientation. There are two approaches to include the orientation. One is to take orientation as a continuous field variable within the energy minimization, which is extremely costly except in very simple cases. Another method, which is very common, is to repeat the microscale (grain level) computation over a predesignated discrete set of grain orientations and compute a weighted average of their outcomes.

Since our microscale computations are very time-consuming, including the grain orientation would require parallelization, which is beyond the time limitations of the present work. Therefore, we skip the grain orientation entirely, and confine our numerical examples to a single grain under simple loading conditions.

Grain boundary formation (atomistic) energy

In the above calculations of the grain energy density, the atomistic share of the boundary energy was left out and here is why. For a cubic grain of side length l , the boundary formation energy equals $6 \times l^2 \gamma_\Gamma^a$, which after division by grain volume gives the energy density as

$$\psi_\Gamma^a = \frac{6\gamma_\Gamma^a}{l}. \quad (5.6)$$

Using (3.126), for a grain size of $l = 3.0 \times 10^{-6} \text{ m}$ we have $\psi_F^a = 8.2 \times 10^5 \text{ J.m}^{-3}$, which is insignificant compared to the overall value of the energy density $\psi = 2.1 \times 10^8 \text{ J.m}^{-3}$; and for larger grain sizes even less significant. On the other hand, the smaller grains have higher energy and adding ψ_F^a does not influence our choice of the optimal grain size. Hence, in all the following computations, the boundary formation energy γ_F^a will be ignored.

5.2 Single grain under simple tension

Suppose that the laminate microstructure is formed under a displacement-controlled loading along the X_1 axis with respect to the austenite lattice coordinates:

$$\bar{\mathbf{F}}_{\text{fin}} = \mathbf{I} + \varepsilon \hat{\mathbf{e}}_1^a \otimes \hat{\mathbf{e}}_1^a. \quad (5.7)$$

As explained before, we compute the overall energy for all pairs of twin-compatible variants according Tab. 3.3 in order to find those pairs with lowest energy. Then, we repeat a more precise computation for these least energy twins. The initial state of twins is taken to be a mixture of the two martensite variants with equal volume fraction $\Lambda = 0.5$, that is,

$$\bar{\mathbf{F}}_{\text{init}} = 0.5\mathbf{F}_i^\circ + 0.5\mathbf{F}_j^\circ. \quad (5.8)$$

Then the loading proceeds stepwise towards the final state using a simple linear interpolation between the initial and final deformation states, such that the m th deformation step is obtained from

$$\bar{\mathbf{F}}_m = \left(1 - \frac{m}{M}\right) \bar{\mathbf{F}}_{\text{init}} + \frac{m}{M} \bar{\mathbf{F}}_{\text{fin}}, \quad (5.9)$$

where $m = 0, 1, \dots, M$. The initial computation yields four possibilities as the least energy twins, listed below, for the final tensile strain $\varepsilon = 0.0334$, which is twice the nominal transformation strain calculated in (5.4).

Type	Variant i	Variant j
I	1	3
I	2	4
I	5	6
II	7	8
$\psi = 2.1948075 \times 10^8 \text{ J.m}^{-3}$		

Fig. 5.3 displays H and Λ across the laminate parallel to the grain. Variations of H clearly show accommodation by branching towards the boundary. On the other hand, accommodation by needles at the vicinity of the boundary takes place only slightly. The dominance of one type of accommodation over the other is determined by material constants and loading conditions. In general, however, needles often form at those boundaries where an average compatibility condition has to be fulfilled, e.g. habit planes [83]. Using a larger boundary

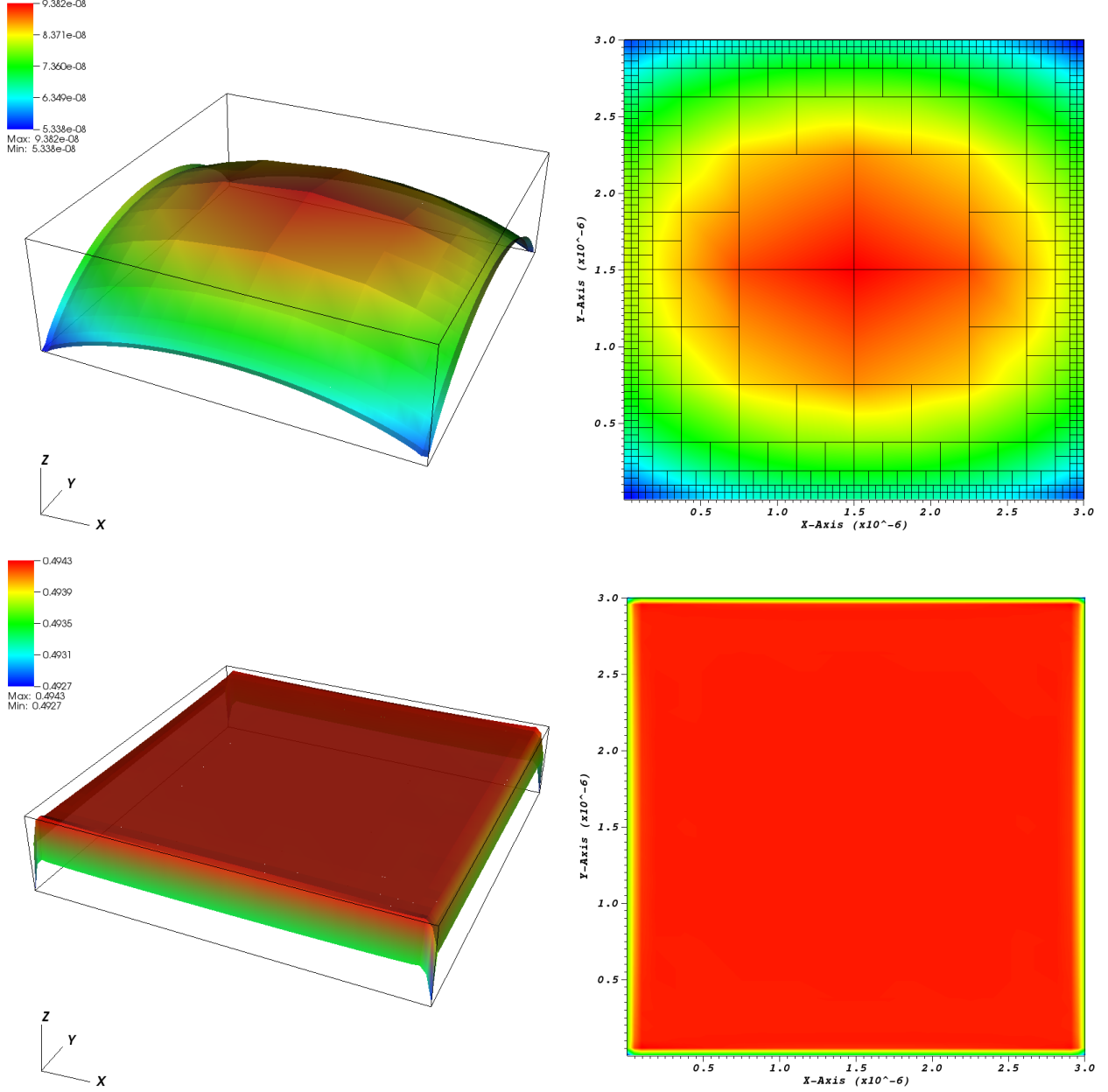


Fig. 5.3: Distribution of lath width H (top row) and volume fraction A (bottom row) across the grain parallel to the laminate, under prescribed deformation (5.8), for type I twins between variants 1 and 3. Both accommodation mechanisms, i.e. branches and needles, can be seen. Although, needles are inconspicuous and branching is dominant due to the particular choice of material properties.

energy coefficient χ in (3.138) would result in more noticeable needles, but at the same time causes the condition number of the tangent stiffness (4.27) to become larger, which makes the computation more expensive at best and less stable at worst. In summary, if one is particularly interested in investigating accommodation by needles, it is best to endow the minimization problem (3.143) with average compatibility boundary conditions at the corresponding boundary.

A sudden increase in the lath width H can be seen at the vicinity of the boundary,

which is interpreted a reverse branching of the laminate. A somewhat similar effect is visible in Λ right before the boundary, where it slightly increases before decreasing at the vicinity of the boundary. This can be intuitively explained by considering the orders of individual field variables in the boundary energy (3.138). The volume fraction Λ and the deformation jump $\mathbf{a}$ appear as fourth- and second-order, whereas the lath width H is of the first order. In an attempt to minimize the boundary energy, the field variables decrease towards the boundary, but the terms of higher order drop faster, which at the same time causes a fast increase of the accommodation strains $2(\Lambda_i^2 - \Lambda_i)\mathbf{a} \otimes H\nabla + H(3\Lambda_i - 2)\mathbf{a} \otimes \Lambda_i\nabla + H(\Lambda_i^2 - \Lambda_i)\mathbf{a} \otimes \nabla$, and their corresponding energy according (3.64). However, this can be partially compensated by reversing the direction of $H\nabla$, which results in the reverse branching at the close neighborhood of the grain boundary. Of course, this is an artifact of mesh size which shall be alleviated as a finer mesh next to the grain boundary is used.

The distribution of elastic and twin interface energy densities are illustrated in Fig. 5.4. Elastic energy density is orders of magnitude larger than the twin interface energy density. This is because in the elastic energy ansatz (3.64) the terms involving $\bar{\mathbf{F}} \pm (\Lambda_i - 1)\mathbf{a} \otimes \mathbf{N} - \mathbf{F}_i^\circ$ are independent of H , and depending on loading conditions may grow very quickly. These terms reflect the part of elastic energy that is independent of accommodation and solely comes from the mismatch between the prescribed deformation and the average transformation strains. To see the effect of loading on distribution of energy, let us show the same

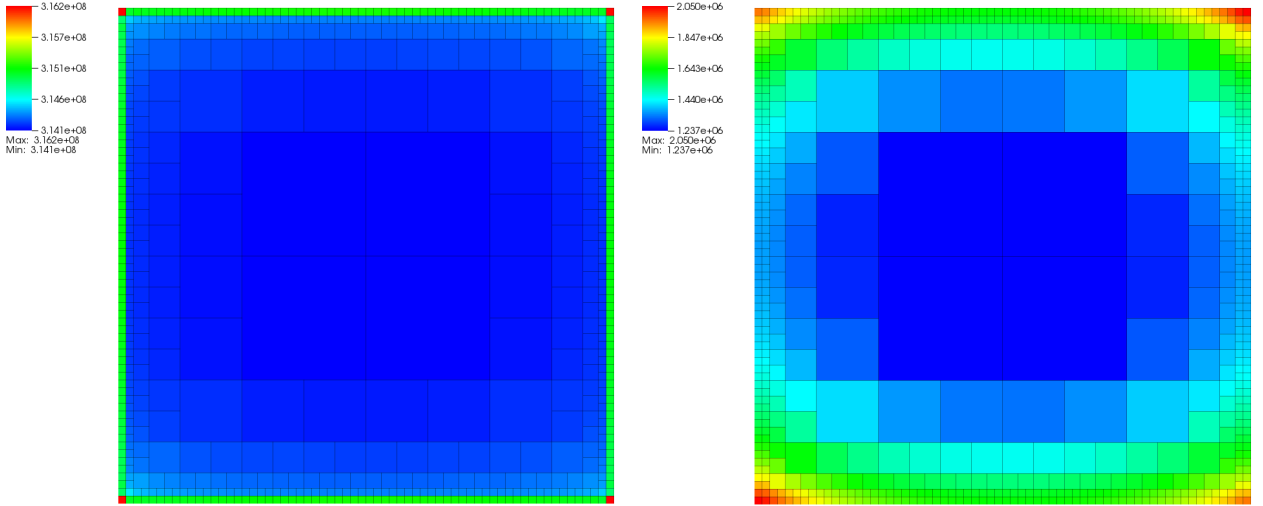


Fig. 5.4: Distribution of elastic energy density (left) and twin interface energy density (right) across the grain parallel to the laminate, under prescribed deformation (5.8), for type I twins between variants 1 and 3.

laminate under a different loading condition; namely, that of

$$\bar{\mathbf{F}} = \frac{1}{20} (0.5\mathbf{F}_1^\circ + 0.5\mathbf{F}_2^\circ) + \frac{19}{20} (\mathbf{I} + 0.0167\hat{\mathbf{e}}_1^a \otimes \hat{\mathbf{e}}_1^a), \quad (5.10)$$

in Fig. 5.5. Here, the difference between interface and elastic energy densities is small and the overall energy is partitioned almost equally between the twin interface energy and the

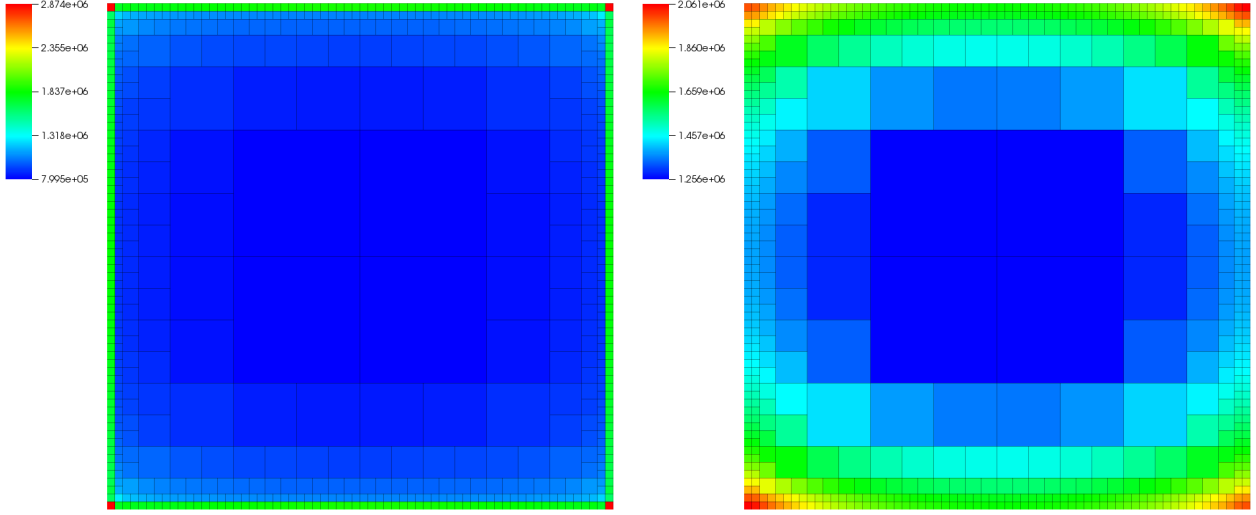


Fig. 5.5: Same as Fig. 5.4, except that the prescribed deformation is at its first step from its initial state, i.e. (5.10).

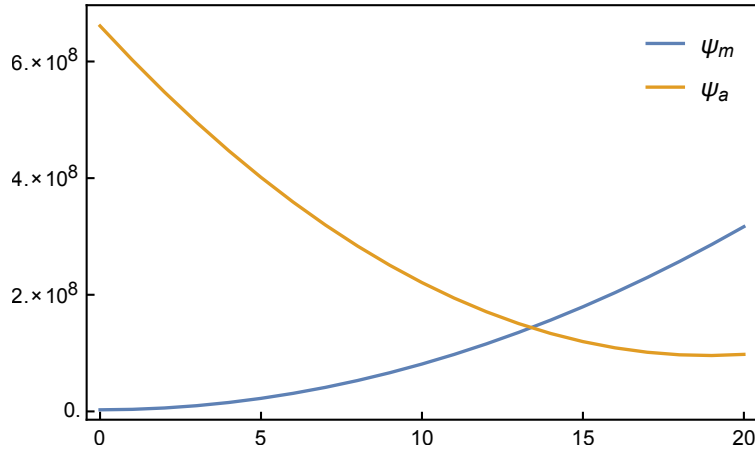


Fig. 5.6: Energy densities [J.m^{-3}] of austenite, ψ_a , and martensite laminate, ψ_m , of type I twins between variants 1 and 3, plotted against deformation step starting from the initial state (5.8) and terminating in the final state in equation (5.7) with $\varepsilon = 0.0167$ after 20 steps according (5.9).

elastic energy, because the discrepancy between the prescribed average deformation (5.10) and the average eigen deformation, caused by transformation strains, is not large.

The evolution of the energy density for the austenite phase and for the martensite laminate microstructure, under the above mentioned conditions, is illustrated in Fig. 5.6. The energy of the austenite phase does not change considerably, due to the particular deformation path that has been chosen. The specific step at which the two curves intersect is coincidental. Now, if we unload the martensite laminate towards zero average strain, that is, if we go through the following deformation route in reverse direction:

$$\bar{\mathbf{F}}_m = \frac{m}{20} (\mathbf{I} + \varepsilon \hat{\mathbf{e}}_1^a \otimes \hat{\mathbf{e}}_1^a), \quad (5.11)$$

it produces the energy densities in Fig. 5.7. At first, it seems that the energy density of the laminate has been severely overestimated. However, we have to remember that the

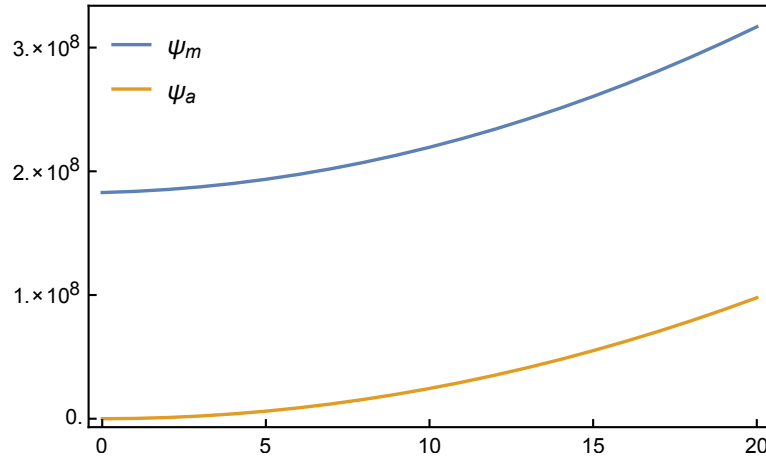


Fig. 5.7: Energy densities [J.m^{-3}] of austenite, ψ_a , and martensite laminate, ψ_m , of type I twins between variants 1 and 3, plotted against deformation path given by (5.11).

grain orientation with respect to the prescribed deformation is by no means optimal. To see how significant the loading conditions can be, we can take a look back on Fig. 5.6, where despite the sophisticated deformation path, it sufficiently clarifies the importance of grain orientation with respect to loading.

Under the same conditions as above, Fig. 5.8 illustrates the vector $\mathbf{a}$ distributed parallel to the laminate. The normal component $\mathbf{a} \cdot \hat{\mathbf{e}}_3$ has been shown separately as a color map. As the laminate accommodates towards the grain boundaries, the discontinuity of deformation gradient in normal direction across the twin interface, expressed by $\mathbf{a} \cdot \hat{\mathbf{e}}_3$, grows slightly. However, the components of $\mathbf{a}$ parallel to the laminate become smaller. This is because the grain boundary energy favors a flat laminate microstructure over a gaged one, as understood from (3.138).

5.2.1 Convergence rate

The convergence rate of the finite element computation can be seen in Fig. 5.9, which is a semi-log plot of ℓ^2 -norm of the Newton iteration's residual versus accumulative step length, and shows a linear convergence. We can also see the log-log plot of the computed overall stored energy of the grain versus accumulative step length of the Newton scheme in Fig. 5.10. Notice that the precision of the computed energy does not improve any further already when the residual has dropped by two orders of magnitude. This translates to almost 50% of the computation time, and can be especially valuable in computing macroscopic problems to save unnecessary computation time.

5.2.2 Mesh independence

In a stable finite element scheme the results shall be ideally mesh independent. To show that our computations are mesh independent, we redo the above example on a coarser grid and with a less restrictive convergence criterion, which stops the computation if the ℓ^2 -norm of

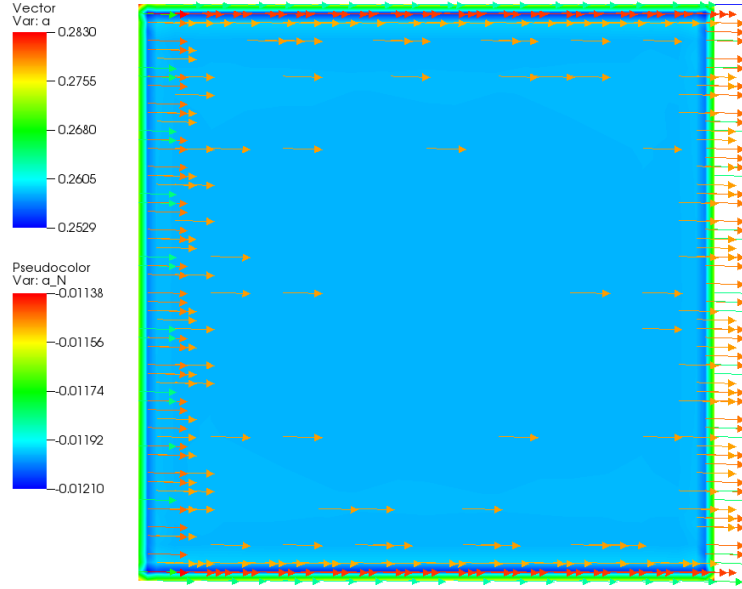


Fig. 5.8: Distribution of vector $\mathbf{a}$ over the grain parallel to the laminate of type I twins between variants 1 and 3, under loading (5.7) with $\varepsilon = 0.0167$. The normal component $\mathbf{a} \cdot \mathbf{e}_3$ has been shown as a color map, and the components parallel to the laminate as a vector field.

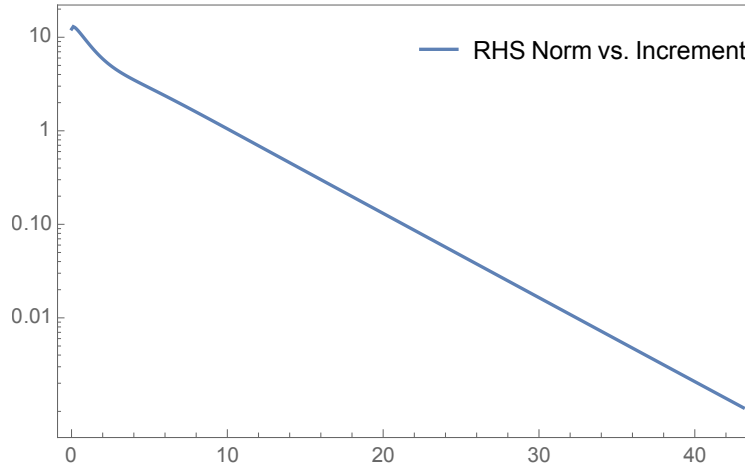


Fig. 5.9: Semi-log plot of ℓ^2 -norm of Newton iterations' residual versus update's increment. Linear convergence is seen.

the residual becomes smaller than 10^{-2} times the ℓ^2 -norm of the initial residual. The results are shown in Fig. 5.11. The only apparent mesh dependency occurs within the single layer of cells next to the boundary, where the presence of the boundary is propagated deeper into the grain due to a coarser mesh. To some extent, this is an expected artifact of our particular energy formulation which does not include any explicit boundary conditions, and instead involves grain boundary energy. The slight difference between the computed minimum and maximum in Λ and H are mostly due to a less strict convergence. We showed, however, that this does not significantly influence the computed energy density, which is the main purpose of microscale finite element computations of this kind (Fig. 5.12). Nevertheless, we

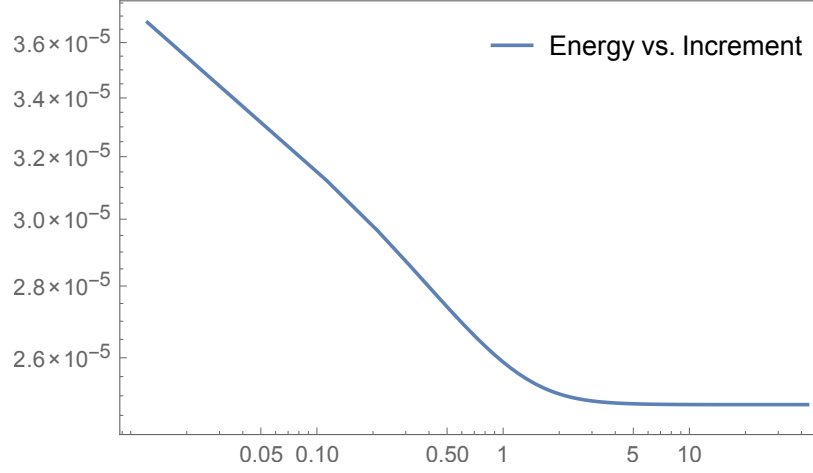


Fig. 5.10: Log-log plot of the laminate stored energy versus Newton update's increment.

will illustrate, for a different example problem, that the *apparent boundary layer* will become mesh independent if the mesh had been refined rather than coarsened, while the convergence criterion remained unaltered.

5.3 The influence of loading conditions

We replace the specific form of prescribed deformation (5.7) by each of the following deformation gradients:

$$\bar{\mathbf{F}}(\varepsilon) = \mathbf{I} + \varepsilon (\hat{\mathbf{e}}_1^a \otimes \hat{\mathbf{e}}_1^a) , \quad (5.12a)$$

$$\bar{\mathbf{F}}(\varepsilon) = \mathbf{I} + \varepsilon (\hat{\mathbf{e}}_1^a \otimes \hat{\mathbf{e}}_1^a + \hat{\mathbf{e}}_2^a \otimes \hat{\mathbf{e}}_2^a) , \quad (5.12b)$$

$$\bar{\mathbf{F}}(\varepsilon) = \mathbf{I} + \varepsilon (\hat{\mathbf{e}}_1^a \otimes \hat{\mathbf{e}}_1^a + \hat{\mathbf{e}}_2^a \otimes \hat{\mathbf{e}}_2^a + \hat{\mathbf{e}}_3^a \otimes \hat{\mathbf{e}}_3^a) , \quad (5.12c)$$

$$\bar{\mathbf{F}}(\varepsilon) = \mathbf{I} + \varepsilon (\hat{\mathbf{e}}_1^a \otimes \hat{\mathbf{e}}_2^a) , \quad (5.12d)$$

$$\bar{\mathbf{F}}(\varepsilon) = \mathbf{I} + \varepsilon (\hat{\mathbf{e}}_1^a \otimes \hat{\mathbf{e}}_2^a + \hat{\mathbf{e}}_1^a \otimes \hat{\mathbf{e}}_3^a) , \quad (5.12e)$$

$$\bar{\mathbf{F}}(\varepsilon) = \mathbf{I} + \varepsilon (\hat{\mathbf{e}}_1^a \otimes \hat{\mathbf{e}}_2^a + \hat{\mathbf{e}}_2^a \otimes \hat{\mathbf{e}}_3^a) , \quad (5.12f)$$

$$\bar{\mathbf{F}}(\varepsilon) = \mathbf{I} + \varepsilon (\hat{\mathbf{e}}_1^a \otimes \hat{\mathbf{e}}_2^a + \hat{\mathbf{e}}_1^a \otimes \hat{\mathbf{e}}_3^a + \hat{\mathbf{e}}_2^a \otimes \hat{\mathbf{e}}_3^a) , \quad (5.12g)$$

and then, similar to (5.4), we solve

$$\frac{1}{2} (\bar{\mathbf{F}}(\varepsilon) - \mathbf{I}) : \mathbb{C}_a : (\bar{\mathbf{F}}(\varepsilon) - \mathbf{I}) = c_{\text{ma}}(\theta) , \quad (5.13)$$

for ε , which designates the nominal transformation strain along the respective directions at the given temperature $\theta = 300$ K. Next, deform the laminates twice the nominal transformation strain following the deformation route (5.9) starting from (5.8) and then back to zero average deformation gradient as we did for (5.7). Note that the first of equations (5.12a) is a mere repetition of (5.7), for completeness.

Tab. 5.1 lists the least energy twinning modes and variants for each of the above prescribed deformations, and Fig. 5.13 displays the evolution of energy densities with respect to

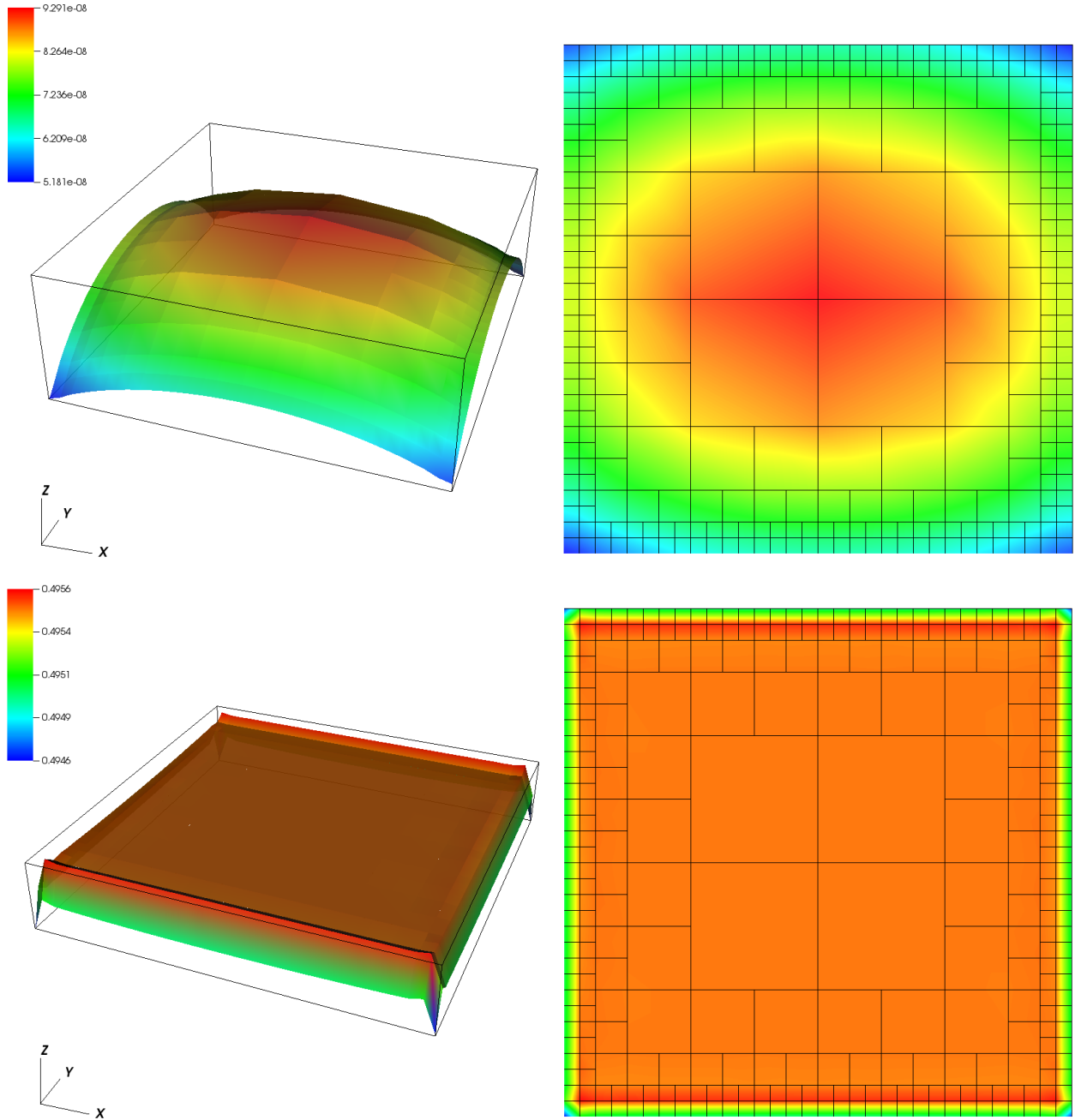


Fig. 5.11: Lath width H (top row) and volume fraction Λ (bottom row), under the same conditions as Fig. 5.3, with coarser mesh and less strict convergence criterion, to indicate mesh independence.

time step according (5.11). The simple uniaxial tension along $\hat{\mathbf{e}}_1^a$ has the overall lowest energy, followed by triaxial shear simultaneously along $\hat{\mathbf{e}}_1^a$, $\hat{\mathbf{e}}_2^a$, and $\hat{\mathbf{e}}_3^a$. This means that, among all seven tested forms of deformation, the uniaxial loading is closest to the least energy conditions. As mentioned, we do not intend to find the globally optimal loading conditions for the NiTi austenite, and these computations are only meant to put our micromechanical model to the test.

Let us also look at the twinning and accommodation patterns for each of the above prescribed deformations. Fig. 5.14 and 5.15 illustrate the distributions of lath width H and

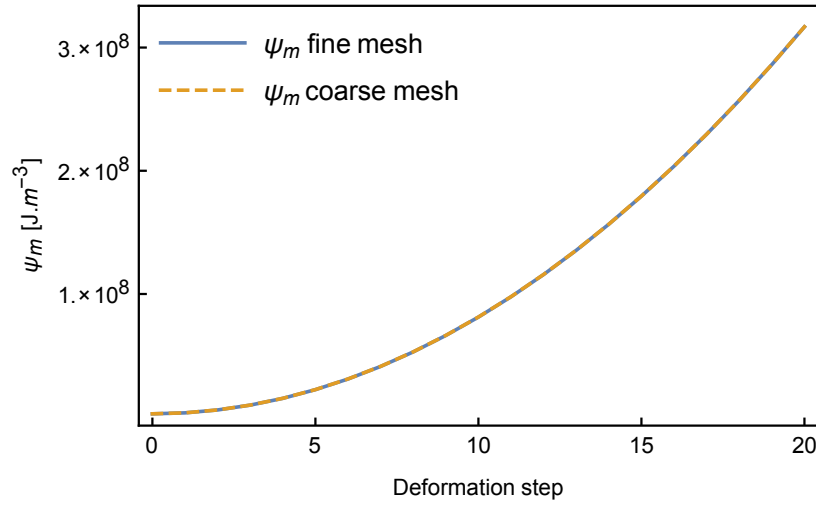


Fig. 5.12: Comparison of computed stored energy for fine and coarse grids to show mesh independence. At the same time on the coarse mesh, a less strict convergence criterion has been used to save computation time. The two curves are almost an exact match, and we had to manually shift a single data point at the 10th deformation step to show the curve hidden beneath.

$\overline{\mathbf{F}}(\varepsilon)$	Twin type and variants	ε	ψ [J.m ⁻³]
(5.12a)	I, 1, 3/I, 2, 4/I, 5, 6/II, 7, 8	0.0167165	2.19439×10^8
(5.12b)	I, 2, 10/I, 4, 12/II, 1, 9/II, 3, 11	0.0118205	9.87066×10^8
(5.12c)	I, 2, 10/I, 4, 12/II, 1, 9/II, 3, 11	0.0096515	1.00338×10^9
(5.12d)	I, 7, 11/I, 8, 12	0.0349652	8.35494×10^8
(5.12e)	I, 2, 10	0.0247241	1.08495×10^9
(5.12f)	I, 8, 12	0.0247241	6.95376×10^8
(5.12g)	I, 4, 12	0.0201872	6.14481×10^8

Tab. 5.1: Least energy twinning modes and variants for various loading conditions according (5.12).

volume fraction Λ across the grain parallel to laminates of least energy (Tab. 5.1) under prescribed deformations (5.12b–5.12g). The very small values of Λ and very large values (larger than the grain size itself) of H essentially suggest that no laminate microstructure forms under these loading conditions, and a monophasic grain of a single variant of martensite is energetically favorable over a laminate microstructure.

5.3.1 Deformation along twinning's shear direction

If the prescribed deformation on the laminate varies along the twinning shear direction, according (3.11) the changes in deformation gradient will be a factor of $\mathbf{a} \otimes \mathbf{N}$. Then, the prescribed average deformation gradient takes the following incremental path:

$$\overline{\mathbf{F}}_m = \left(1 - \frac{m}{M}\right) \mathbf{F}_i^\circ + \frac{m}{M} \mathbf{F}_j^\circ. \quad (5.14)$$

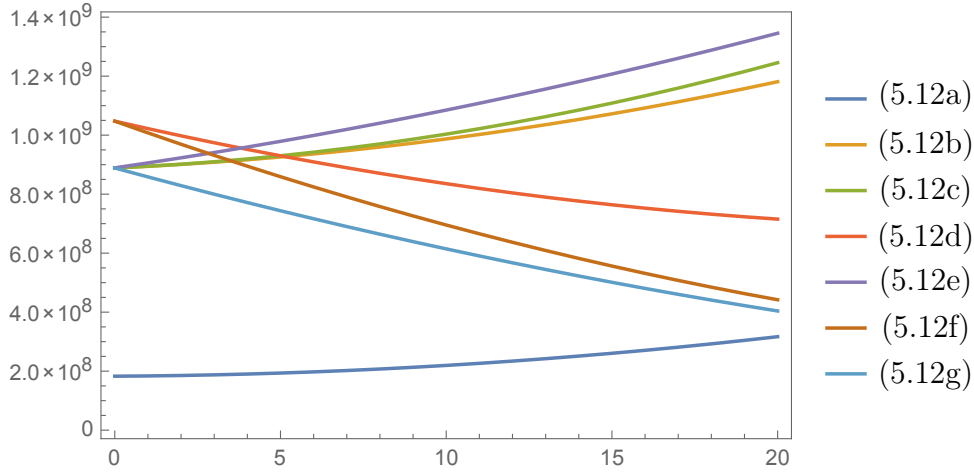


Fig. 5.13: Energy density versus deformation step for various prescribed deformations according equations (5.12) and Tab. 5.1.

Fig. 5.16 illustrates the laminate energy and its constituents, i.e. boundary, twin interface, and elastic energies, for varying mixture percentage of type I twin between variants 1 and 3 of NiTi martensite. Expectedly, the share of the elastic energy pertaining to the mismatch between $\bar{\mathbf{F}}$ and the transformations $\mathbf{F}_i^\circ$, $\mathbf{F}_j^\circ$ vanishes, and the remaining elastic energy is confined to that of accommodation only which should not be significant due to near-ideal match between the prescribed and the laminate average deformation gradients. The twin interface energy (almost) vanishes for $m = 0$ and $m = M$, because only a single variant is present. The same is true for boundary energy. As the percentage of the phase mixture changes, the relative boundary–laminate orientation changes in deformed configuration, which is also reflected in the boundary energy.

5.4 A one-dimensional problem, analytical treatment

If the general minimization problem (3.143) is simplified under the following assumptions:

- the gradients are nonzero only along one dimension along $x \in \Omega = [0, L] \subset \mathbb{R}$,
- H and $\mathbf{a}$ are constant,
- boundary energy is replaced by boundary conditions $\Lambda(0) = \Lambda(L) = 0$,
- the two martensitic variants have equal stiffness,

then we end up with

$$\min_{\Lambda} \left\{ \mathcal{E} = \int_0^L k_1(\mathbf{a}, \Lambda) H \Lambda' + k_2(\mathbf{a}, \Lambda) H^2 \Lambda'^2 + \frac{2\gamma}{H} dx \mid \mathbf{a}, H = \text{const.}, \Lambda(0) = \Lambda(L) = 0 \right\}, \quad (5.15)$$

in which $\Lambda(x)$ and consequently k_1 and k_2 are even functions around $x = L/2$, because of the inherent symmetry of the problem. It is clear then that the integral of the first term vanishes, because Λ' will be an odd function around $x = L/2$. Therefore, the above problem

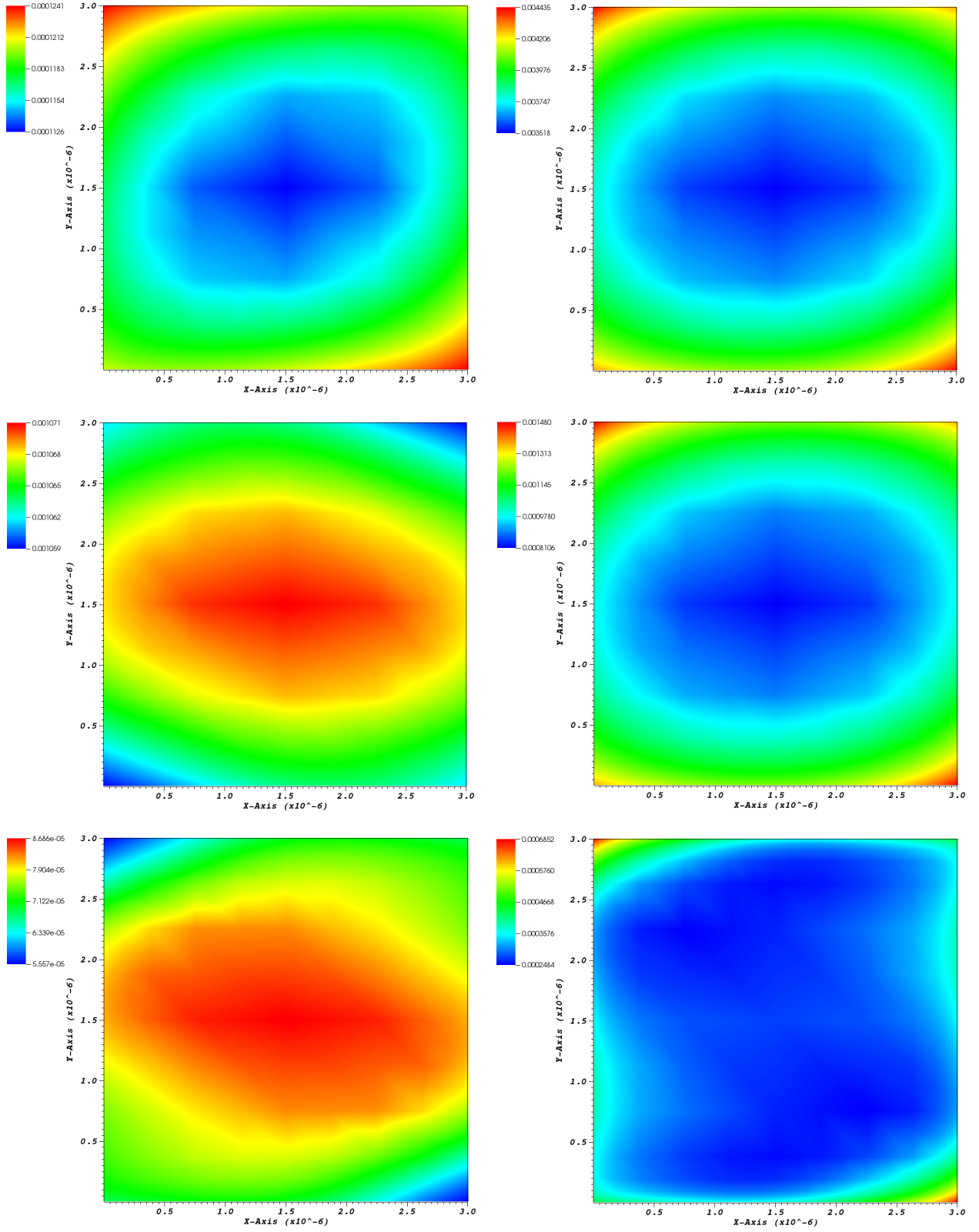


Fig. 5.14: Distribution of lath width H (left) and volume fraction A (right) across the grain parallel to laminates of least energy under prescribed deformations **(top)** 5.12b, **(middle)** 5.12c, **(bottom)** 5.12d.

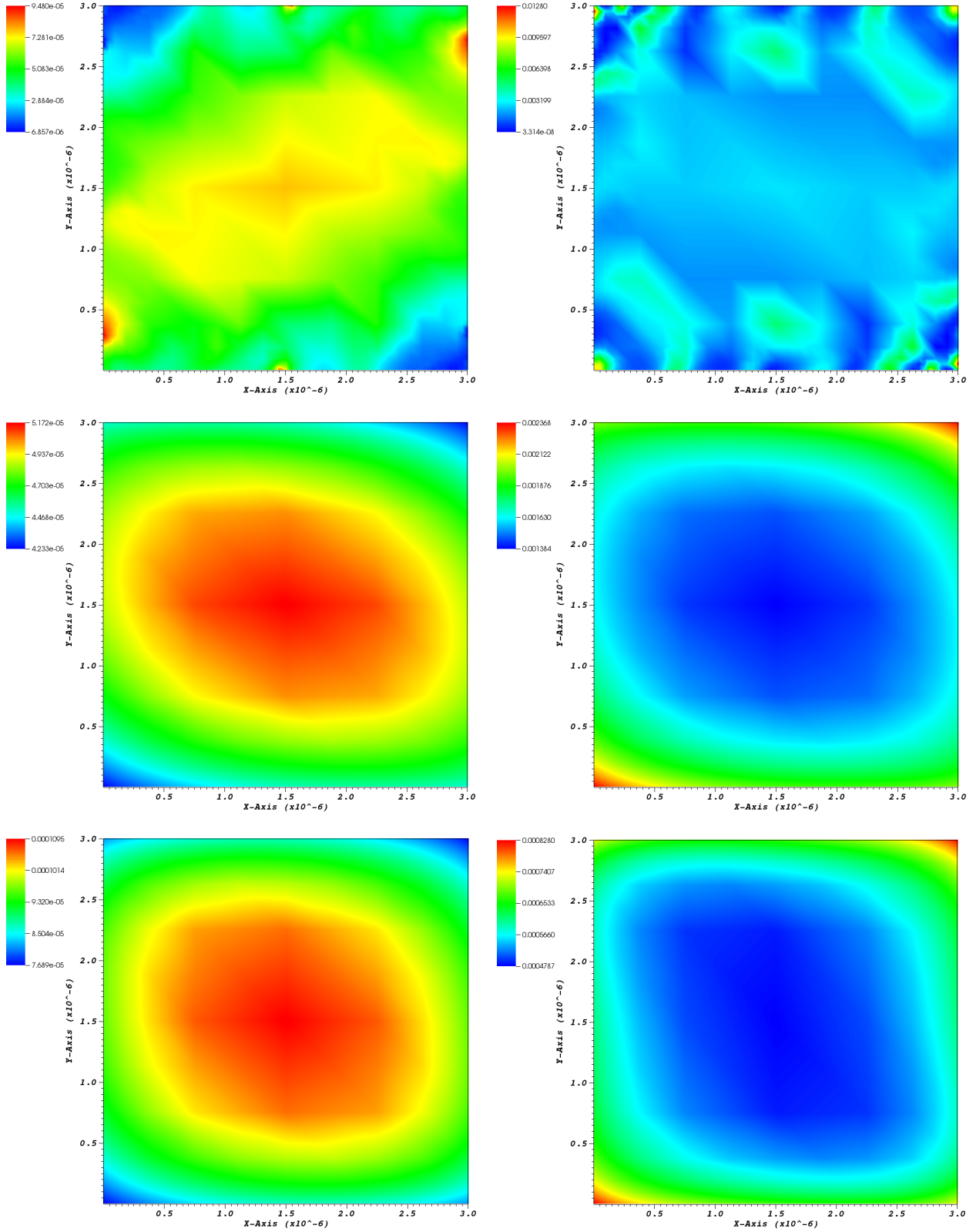


Fig. 5.15: Distribution of lath width H (left) and volume fraction Λ (right) across the grain parallel to laminates of least energy under prescribed deformations **(top)** 5.12e, **(middle)** 5.12f, **(bottom)** 5.12g.

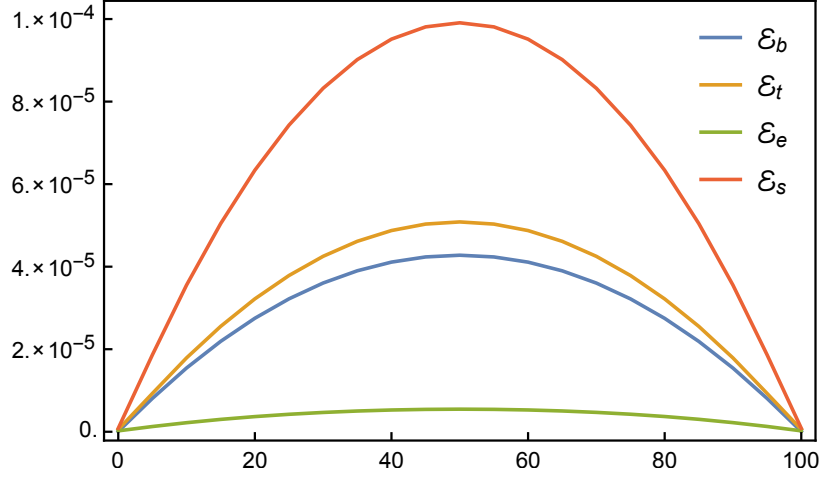


Fig. 5.16: Laminate energy [J] versus percentage of volume fraction $\Lambda\%$, and its contributions from (1) boundary energy, (2) twin interface energy, (3) elastic energy. The laminate is composed of type I twins between variants 1 and 3 of NiTi martensite.

simplifies further to

$$\min_{\Lambda} \left\{ \mathcal{E} = 2\gamma \frac{L}{H} + H^2 \int_0^L k_2(\mathbf{a}, \Lambda) \Lambda'^2 dx \mid \mathbf{a}, H = \text{const.}, \Lambda(0) = \Lambda(L) = 0 \right\}. \quad (5.16)$$

Next, we nondimensionalize the spatial coordinate as $\xi = x/L$, which gives $dx = Ld\xi$, and $\Lambda'(x) = \bar{\Lambda}'(\xi)/L$. Substitution of these into (5.16) yields

$$\min_{\bar{\Lambda}} \left\{ \mathcal{E} = 2\gamma \frac{L}{H} + \frac{H^2}{L} \int_0^1 k_2(\mathbf{a}, \bar{\Lambda}) \bar{\Lambda}'^2 d\xi \mid \mathbf{a}, H = \text{const.}, \bar{\Lambda}(0) = \bar{\Lambda}(1) = 0 \right\}. \quad (5.17)$$

Now, if $\bar{\Lambda}_m$ is a minimizer of this problem, then the energy takes the form of

$$\mathcal{E} = 2\gamma \frac{L}{H} + \kappa \frac{H^2}{L}, \quad (5.18)$$

in which $\kappa = \int_0^1 k_2(\mathbf{a}, \bar{\Lambda}_m) \bar{\Lambda}_m'^2 d\xi$, is independent of H and L . The optimality of $\mathcal{E}$, requires that

$$\frac{\partial \mathcal{E}}{\partial H} = 0 \Rightarrow H^3 = \frac{\gamma}{\kappa} L^2,$$

which gives

$$H = \left(\frac{\gamma}{\kappa} \right)^{1/3} L^{2/3}. \quad (5.19)$$

This scaling relation is a confirmation of the experimental result by Chu and James 1993 [36]. Their experiment showed the distribution of laminate patch span versus average lath width follows a scaling law of the form $H \propto L^{2/3}$, which we just derived as a special case of our micromechanical model.

Let us now put the above value of H back into (5.18) to obtain

$$\mathcal{E} = 3 \left(\gamma^2 \kappa \right)^{1/3} L^{1/3}. \quad (5.20)$$

This is another important result which has been derived by Kohn, Otto, Müller [99, 98, 113, 97], both as a scaling law and as an upper and lower bound for the laminate energy in certain regimes (see the neat review by Conti [38]).

In its own right, it is very interesting that how two seemingly independent results, i.e. a theoretical result on energy bounds and an empirical observation on length scale distribution, are connected through a single minimization problem like (5.15), which is simple enough to be treated analytically using elementary derivations.

5.5 Homogenization: cuboid RVE with cylindrical inclusion

The microstructural features and, as a result, the structural and functional properties of alloys can be modified via thermomechanical treatment and by adjusting the constituent elements and their proportions. The NiTi is no exception. The nickel content in the so called nickel-rich NiTi has a significant influence on transformation properties and on intermediate metastable phases that emerge as *precipitates* [123, pp. 50–53]. A particularly interesting example is Ni_4Ti_3 because it supports the R-phase and therefore encourages the shape memory and pseudoelasticity [94, 85]. Ni_4Ti_3 commonly appears in the form of lenticular precipitates and its presence propagates into its surrounding matrix through strain and concentration gradients [138]. These effects as well as some aspects of anisotropy of Ni_4Ti_3 have been modeled via phase field and ab initio methods to some extent [150, 91, 90].

In order to simulate the strain field surrounding such a precipitate, in this section we assume a simplified model of an inclusion. Namely, we replace the formerly presented idealized geometry of the cuboid grain with a cuboid *representative volume element* (RVE) of side length $L_{\text{RVE}} = 3.0 \times 10^{-6}$ m that has a cylindrical inclusion of diameter $D_{\text{incl}} = L_{\text{RVE}}/2$ across its center and normal to its faces. Given that the gradients normal to the twin interfaces are neglected in our model, we again substitute the three-dimensional domain with its two-dimensional counterpart, namely a quadratic RVE with a circular inclusion, where apparently the cylindrical inclusion is assumed perpendicular to twin interfaces (Fig. 5.17). This geometry mimics the situation where an island of austenite is trapped within the transformed martensite.

The minimization problem (3.143) is slightly modified such that the RVE's outer boundary possesses no energy and fulfills *periodic boundary conditions*, and its inner boundary, i.e. that of the circular inclusion, is endowed with boundary energy (3.138). This setting corresponds to a completely standard homogenization problem. Since our method is energy based, there is no need for verifying the *Hill–Mandel* condition [115, 114]. This condition ensures that homogenization procedures do not violate the conservation of energy.

The loading condition follows (5.9) with a prescribed deformation in the form

$$\bar{\mathbf{F}}_{\text{fin}} = \mathbf{I} + \varepsilon (\hat{\mathbf{e}}_1^a \otimes \hat{\mathbf{e}}_1^a + \hat{\mathbf{e}}_1^a \otimes \hat{\mathbf{e}}_3^a), \quad (5.21)$$

with $\varepsilon = 0.01508158$ which equals the nominal transformation strain calculated analogous

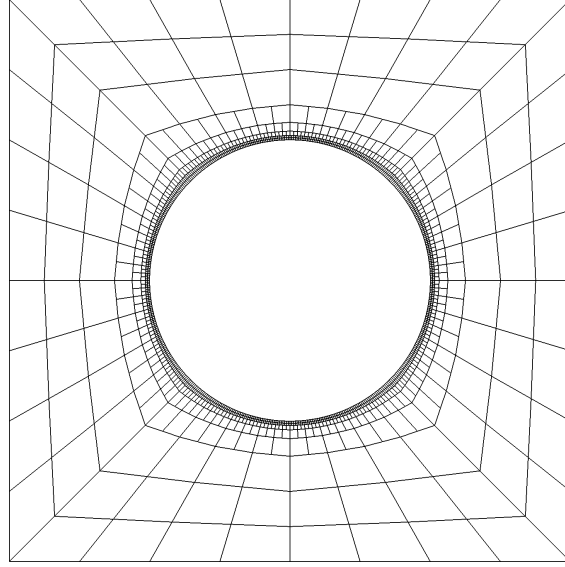


Fig. 5.17: The 2D mesh of the quadratic RVE of side length $L_{\text{RVE}} = 3.0 \times 10^{-6}$ m with circular inclusion of diameter $D_{\text{incl}} = L_{\text{RVE}}/2$.

to (5.4) by

$$\frac{\varepsilon^2}{2} \|\hat{\mathbf{e}}_1^a \otimes \hat{\mathbf{e}}_1^a + \hat{\mathbf{e}}_1^a \otimes \hat{\mathbf{e}}_3^a\|_{\mathbb{C}_a} = c_{\text{ma}}(300) = 2.44513 \times 10^7. \quad (5.22)$$

The deformation is applied in five steps, $M = 5$. The rest of the implementation, especially the choice of martensite coordinates (5.1), remains the same. Fig. 5.18 illustrates the distribution of twin width and volume fraction throughout the RVE for a type I twin between variants 1 and 3 of the NiTi martensite. Both accommodation mechanisms, i.e. branches and needles, can be seen and once again needles are inconspicuous and branching is dominant due to the particular choice of material properties.

The distribution of the vector $\mathbf{a}$ and the various energy densities are shown in Fig. 5.19. The vector $\mathbf{a}$ is again decomposed into normal and tangential components, which have been overlaid as a scalar color map, $a_{\perp}$, and a vector field, $\mathbf{a}_{\parallel}$, respectively. Since the elastic energy density is almost uniform across the domain, a magnified figure around the inclusion has been illustrated. The same is true for the boundary energy trivially, because it is absent throughout the domain except for the cells adjacent to the boundary. Similar to the case of quadratic grain, the boundary energy density has almost the same order of magnitude as the elastic energy density. The twin interface energy density, on the other hand, is considerably smaller. Further down, the overall energy contributions will be compared as a function of inclusion size and for varying surface energies.

5.5.1 Mesh independence

Let us once again evaluate the reliability of the finite element implementation in terms of mesh independence, this time for the RVE problem. We refine the mesh once at the layer next to the inclusion and then repeat the computations to compare the results with the previous ones. To obtain a more convincing visual confirmation, we zoom into an arbitrarily

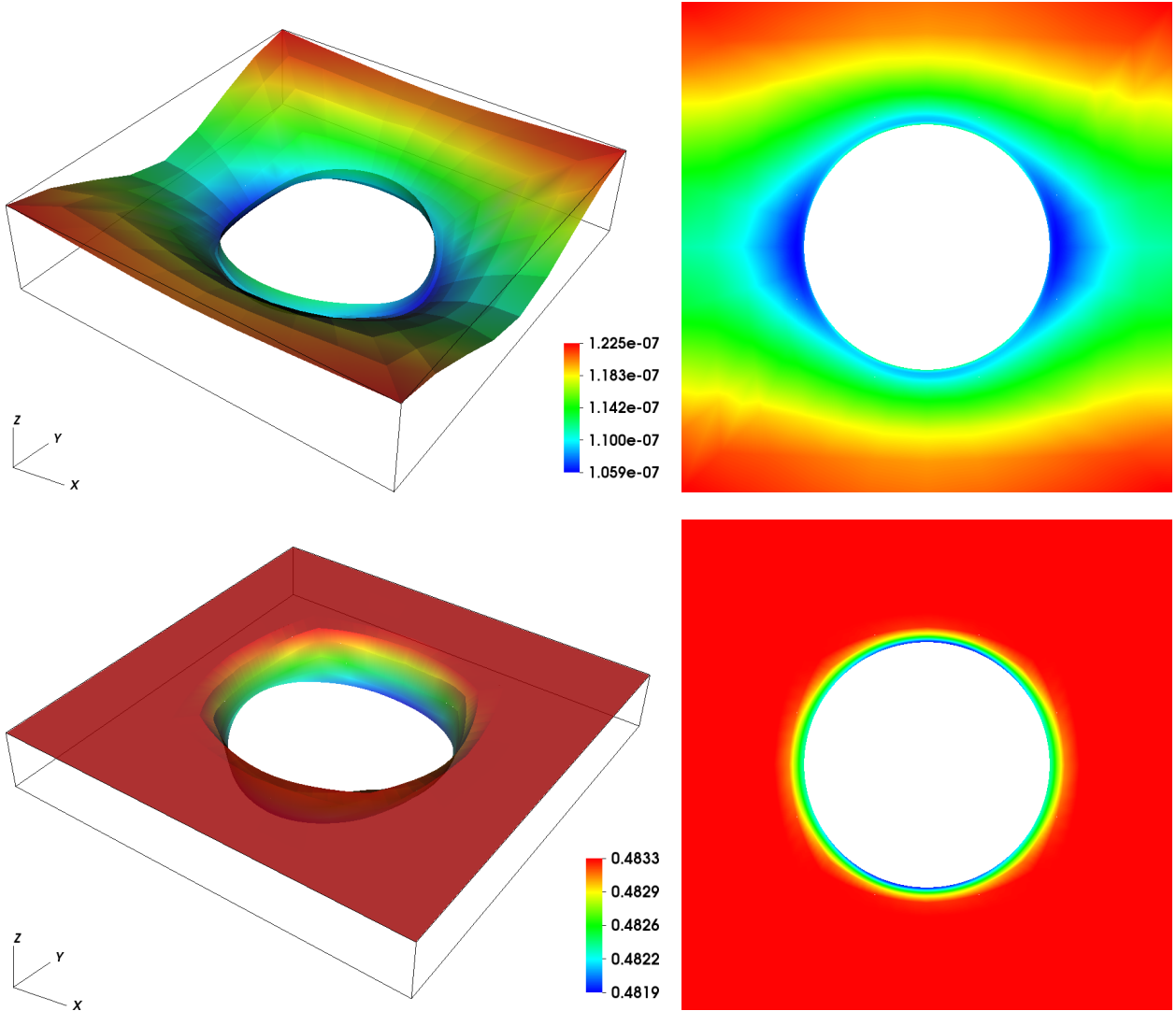


Fig. 5.18: Lath width H (top row) and volume fraction A (bottom row) across the RVE parallel to the laminate, under deformation (5.21), in type I twins of variants 1 and 3.

chosen subdomain adjacent to the inclusion. Fig. 5.20 compares the resulting color maps for H , A , and $a_{\perp}$ for the initial and the refined meshes. The mesh independence is almost self evident here both in terms of visual similarity as well as the range of values. One has to be aware that some degree of mesh dependence is inherent in any convergent numerical scheme and is therefore expected. This is hypothesized in equation (4.31) for the finite element method. The slight apparent mesh dependence here can also be seen as an artifact of the hanging nodes. The presence of the hanging nodes poses additional equality constraints on the numerical scheme which ensure continuity of the fields across those cell boundaries where the hanging nodes reside.

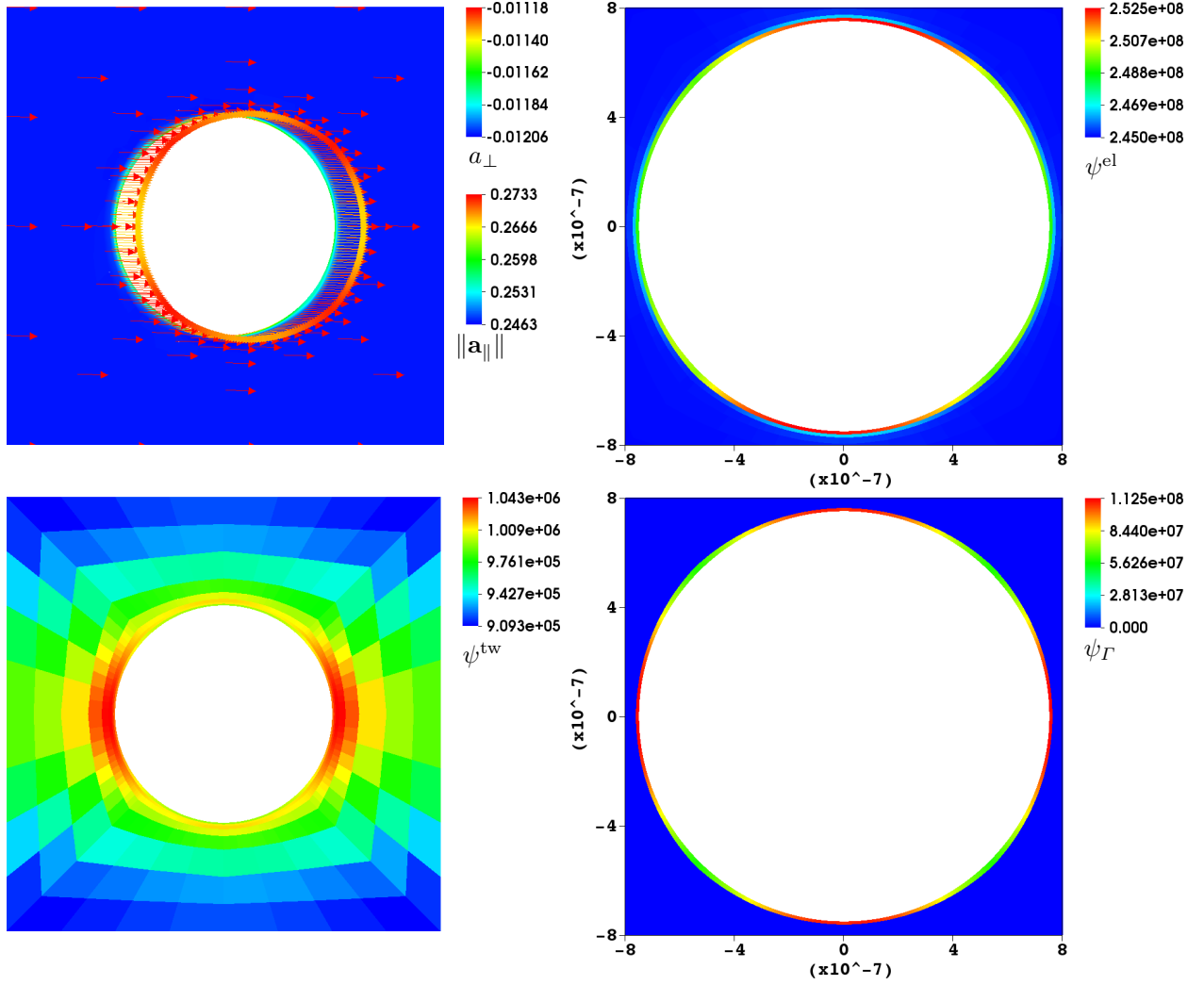


Fig. 5.19: Distribution of deformation gradient jump $\mathbf{a}$ (top left), twin interface energy density (top right), elastic interface energy density (bottom left), and boundary energy density (bottom right) across the RVE for the same example as Fig. 5.18.

5.6 Energy versus inclusion size

The characteristic geometrical parameter for the chosen RVE is the ratio of the inclusion diameter to the RVE's side length, $D_{\text{incl}}/L_{\text{RVE}}$. Alternatively, one might obtain a similar non-dimensional parameter from the ratio of cross-section areas, $\sqrt{A_{\text{incl}}/A_{\text{RVE}}} = \sqrt{\pi}/2 (D_{\text{incl}}/L_{\text{RVE}})$, which differs by a constant factor only.

We vary this ratio in the range $0.10 \leq D_{\text{incl}}/L_{\text{RVE}} \leq 0.85$ over evenly spaced discrete values and compute the stored energy density of the grain corresponding to elastic deformation, twin interface, and the inclusion boundary. The laminate is again composed of type I twins of variants 1 and 3. The prescribed deformation has the form of (5.12a), i.e. $\bar{\mathbf{F}}_{\text{fin}} = \mathbf{I} + 2\varepsilon \hat{\mathbf{e}}_1^a \otimes \hat{\mathbf{e}}_1^a$, in which ε is the nominal transformation strain at $\theta = 300$ K, which according to Tab. 5.1 is $\varepsilon = 0.0167165$. The mesh is generated exactly as in Fig. 5.17, and the loading is carried out following (5.9) with $M = 5$. Convergence was assumed when the

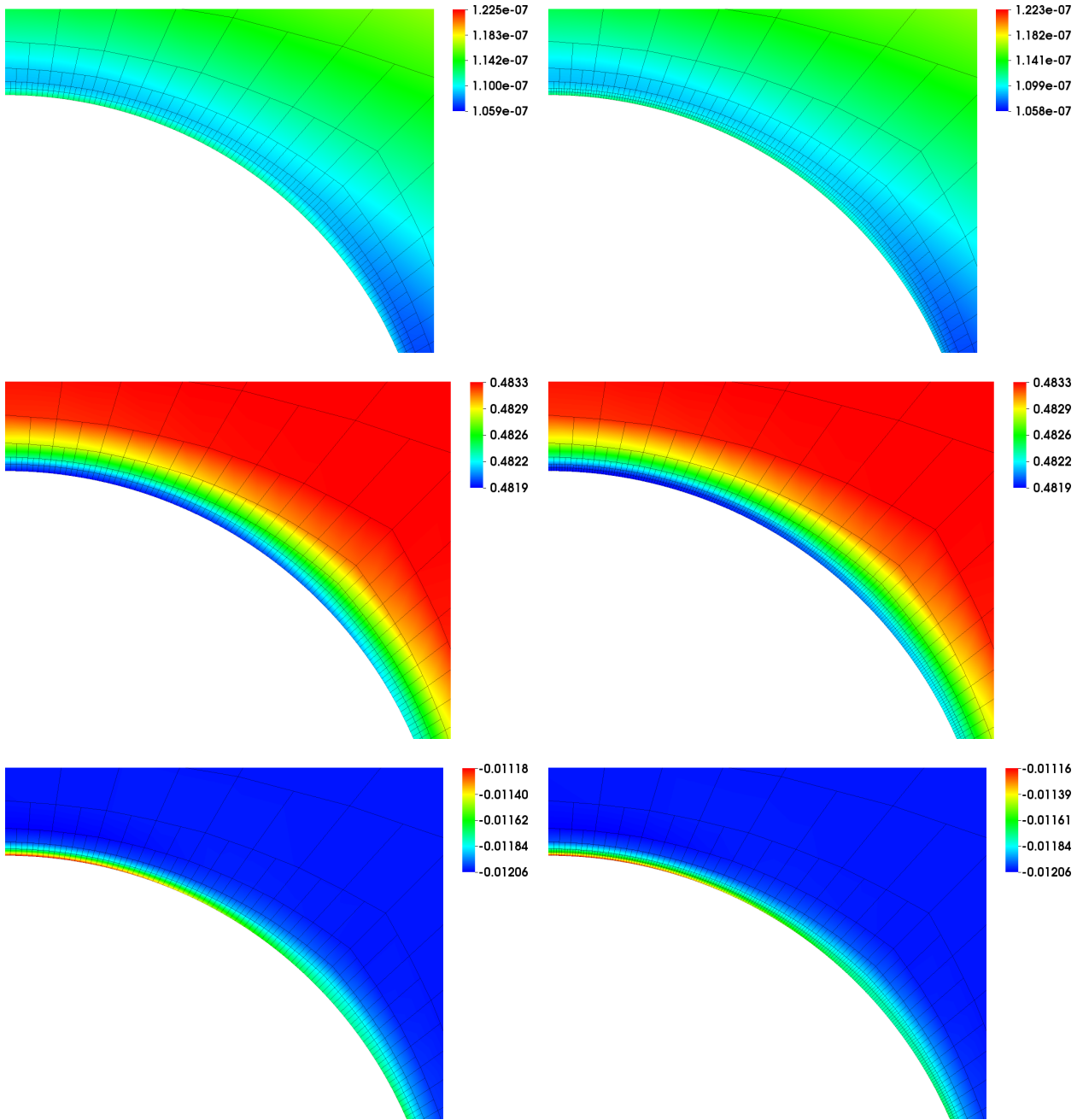


Fig. 5.20: Magnified distribution of H (top), Δ (middle), and $a_{\perp}$ (bottom) at the vicinity of the inclusion for the results in Fig. 5.18 (left) compared with the refined mesh (right).

Frobenius norm of the residual reached 10^{-2} times its initial value.

Fig. 5.21 shows the outcome. Since the interface and boundary energies are much smaller than the elastic energy, they have been re-illustrated separately for clarity. Notice that more sampling points were added for values $D_{\text{incl}}/L_{\text{RVE}} > 0.8$, where sudden changes and jumps appear. Despite the higher number of sampling points, however, the computed energies show apparent instability which signals a failure of the model. This can be understood since the neighboring inclusions become too close to each other and their interactions become too significant to be ignored. The presented model has obviously no provision of such interac-

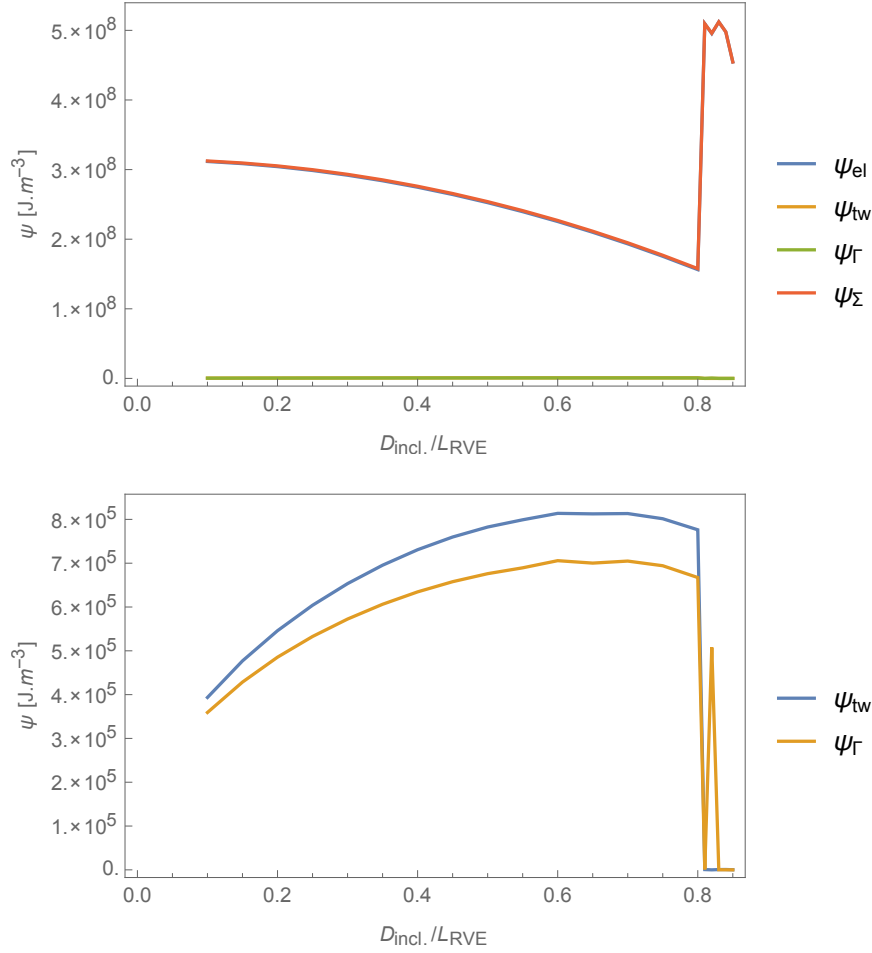


Fig. 5.21: Stored energy densities, ψ , versus inclusion ratio, $D_{\text{incl.}}/L_{\text{RVE}}$, for the cubic RVE with cylindrical inclusion.

tions. It is noticeable that the twin interface and boundary energies stay close to each other and grow with the inclusion size, which makes sense. On the other hand, the elastic energy decreases with inclusion size, which again is no surprise. A rather conspicuous observation is that the elastic energy is orders of magnitude larger than the other two contributions, which seemingly does not speak to some well-known mathematical results [99, pp. 277]. We can give two speculative reasons for this. First, our definition of elastic energy includes two contributions, namely that of the accommodation strains and the pointwise deviation of the strain with respect transformation strains. On the contrary, the aforementioned mathematical derivations identify the second contribution, i.e. deviation of the pointwise strain from transformation strains, as the “surface energy due to twinning”. Second, their results are based on certain simplifying assumptions on material parameters and boundary conditions which barely have any analogy to their counterparts in the present model.

To see the effect of twin interface energy, γ , and boundary energy coefficient, χ , we redo the above computations for three variations of these constants, where either one of them or both of them are 10 times their original values. The results are summarized in Fig. 5.22.

It can be seen that higher values of γ have a stabilizing effect at larger inclusion size,

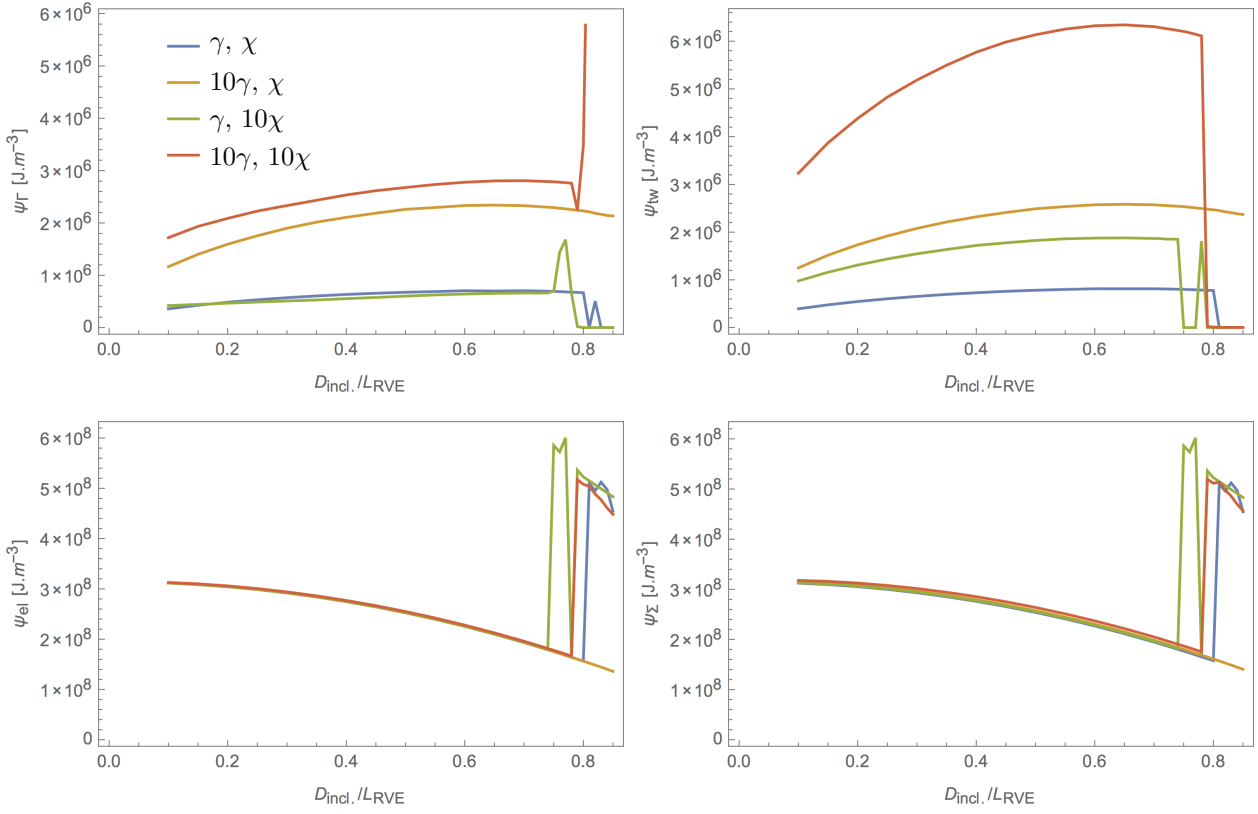


Fig. 5.22: Stored energy densities versus inclusion ratio corresponding to boundary (top left), twin interfaces (top right), elastic strains (bottom left), and total stored energy (bottom right) for RVE.

whereas a larger boundary energy is destabilizing. This is physically plausible because larger boundary energy coefficients amplify the interaction between neighboring inclusions. On the other hand, larger twin interface energy hinders the propagation of the boundary effects into the medium, hence the interaction between neighboring inclusions is weakened. The general trend of increasing twin interface and boundary energies and decreasing elastic energy with inclusion size is maintained in all cases.

Chapter 6

Conclusion and outlook

6.1 Summary

A continuum micromechanical model for a first-order laminate microstructure of martensite based on energy minimization has been presented with three contributions as follows.

- Elastic energy has been calculated based on a strictly coherent ansatz of the laminate with martensitic variants whose fully anisotropic stiffness tensors are oriented in accordance with their exact lattice orientation. Having different stiffnesses for the two variants in their exact orientation turns out to be challenging.
- Twin interface energies were analytically estimated based on crystallographic and atomistic arguments. The assumption is that the interfaces are sharp and correspond to crystal monolayers.
- An ansatz of the grain boundary energy was derived based on a self-similar construct.

The energy minimization has been implemented as a nonlinear finite element scheme. The model has been applied to predict geometrical characteristics of refinement and accommodation of microstructure, as well as its mesoscopic mechanical response. The numerical results are mesh independent and stable, with an almost linear convergence rate which is acceptable due to the highly nonlinear nature of the problem. A basic one-dimensional analytical treatment of the model confirmed some of the well-established theoretical and experimental observations on the scaling features of the laminate.

We would like to emphasize that some parts of this work are particularly new, to the best of our knowledge. These are the following. First of all, we enforce compatibility condition (twinning equation) strictly in the sense of mechanical twins and not growth twins. Second, we extended the usual energy minimization framework to include martensite variants having different elastic stiffness tensors with exact orientations compliant to the twinning equation, which is novel (see [63, pp. 444]). Lastly, all material constants came from experiments directly or indirectly. When not available, they were calculated based on crystallographic and atomistic values. Therefore, none of the material parameters have regularizing purposes or were obtained by data fitting.

6.2 Possible extensions

Since a precise representation of interfaces has been developed, the dissipation can be modeled as a function of interface formation and annihilation. The computations turned out to be quite expensive for single-threaded algorithms. Therefore, we confined our numerics to the mesoscale and to a few discrete grain orientations. A parallelized finite element implementation would make it possible to bridge the scale up to macroscopic length scales, and also to include continuous optimization of grain orientation. We excluded anisotropy from the grain boundary energy ansatz, which resulted in a fixed relative orientation of the laminate with respect to its boundary. An anisotropic ansatz would correct this shortcoming. However, this would not be significant from a material modeling perspective, and would only help to better understand the boundary effects at the microscale. The presented numerical framework can be extended to investigate optimal grain size with regard to minimum energy density and for arbitrary grain orientations, and also the effect of grain size on transformation temperature and dissipation.

Appendix A

Miscellaneous mathematical notes and formulas

In this appendix, we briefly mention some mathematical formulas and notations that are used throughout the text.

A.1 Some tensor algebra and calculus

A.1.1 Rotation matrix

There are various forms of rotation matrix. We use the one which is determined based on a rotation axis identified by unit vector $\hat{\mathbf{e}}$, and a rotation angle denoted by θ , in the following form

$$\mathbf{R}(\hat{\mathbf{e}}, \theta) = \cos \theta \mathbf{I} - \sin \theta \boldsymbol{\epsilon} \cdot \hat{\mathbf{e}} + (1 - \cos \theta) \hat{\mathbf{e}} \otimes \hat{\mathbf{e}}, \quad (\text{A.1})$$

where $\boldsymbol{\epsilon}$, called *permutation* tensor, or the third-order *Levi-Civita* tensor, or sometimes the *totally antisymmetric* tensor, is defined as

$$\epsilon_{ijk} = \begin{cases} +1 & (i, j, k) \text{ be an } \textit{even} \text{ permutation of } (1, 2, 3) , \\ -1 & (i, j, k) \text{ be an } \textit{odd} \text{ permutation of } (1, 2, 3) , \\ 0 & i = j, \text{ or } j = k, \text{ or } k = i . \end{cases} \quad (\text{A.2})$$

It is worth noting that the term $-\boldsymbol{\epsilon} \cdot \hat{\mathbf{e}}$ has an interesting interpretation, namely for every vector $\mathbf{v}$

$$[\mathbf{v}]_{\times} := -\boldsymbol{\epsilon} \cdot \mathbf{v} = -\mathbf{v} \cdot \boldsymbol{\epsilon} = \epsilon_{ikj} v_k \hat{\mathbf{e}}_i \hat{\mathbf{e}}_j , \quad (\text{A.3})$$

is a tensor such that the cross product of $\mathbf{v}$ with any vector can be calculated by

$$\mathbf{v} \times \mathbf{u} = [\mathbf{v}]_{\times} \cdot \mathbf{u} , \quad \mathbf{u} \times \mathbf{v} = \mathbf{u} \cdot [\mathbf{v}]_{\times} . \quad (\text{A.4})$$

A.1.2 Dyadic products

Dyadic product of two tensors gives a tensor whose order is the sum of orders of its operands. When two first-order tensors (vectors) are multiplied we have

$$\mathbf{u} \otimes \mathbf{v} = u_i v_j \hat{\mathbf{e}}_i \otimes \hat{\mathbf{e}}_j . \quad (\text{A.5})$$

For higher-order tensors the indices can contract in various ways. Most interesting to us are the three following cases where $\mathbf{A} = A_{ij} \mathbf{e}_i \otimes \mathbf{e}_j$ and $\mathbf{B} = B_{kl} \mathbf{e}_k \otimes \mathbf{e}_l$,

$$\mathbf{A} \otimes \mathbf{B} = A_{ij} B_{kl} \mathbf{e}_i \otimes \mathbf{e}_j \otimes \mathbf{e}_k \otimes \mathbf{e}_l , \quad (\text{A.6})$$

$$\mathbf{A} \overline{\otimes} \mathbf{B} = A_{ik} B_{jl} \mathbf{e}_i \otimes \mathbf{e}_j \otimes \mathbf{e}_k \otimes \mathbf{e}_l , \quad (\text{A.7})$$

$$\mathbf{A} \underline{\otimes} \mathbf{B} = A_{il} B_{jk} \mathbf{e}_i \otimes \mathbf{e}_j \otimes \mathbf{e}_k \otimes \mathbf{e}_l . \quad (\text{A.8})$$

$$(\text{A.9})$$

A.1.3 Fourth-order identity (unit) tensors

A common definition of the fourth-order identity tensor is based on differential calculus; namely, for a differentiable second-order tensor field $\mathbf{A}$ we introduce

$$\overline{\mathbb{I}} = \frac{\partial \mathbf{A}}{\partial \mathbf{A}} = \frac{\partial A_{ij}}{\partial A_{kl}} \mathbf{e}_i \otimes \mathbf{e}_j \otimes \mathbf{e}_k \otimes \mathbf{e}_l = \delta_{ik} \delta_{jl} \mathbf{e}_i \otimes \mathbf{e}_j \otimes \mathbf{e}_k \otimes \mathbf{e}_l = \mathbf{I} \overline{\otimes} \mathbf{I} , \quad (\text{A.10})$$

$$\underline{\mathbb{I}} = \frac{\partial \mathbf{A}}{\partial \mathbf{A}^T} = \frac{\partial A_{ij}}{\partial A_{lk}} \mathbf{e}_i \otimes \mathbf{e}_j \otimes \mathbf{e}_k \otimes \mathbf{e}_l = \delta_{il} \delta_{jk} \mathbf{e}_i \otimes \mathbf{e}_j \otimes \mathbf{e}_k \otimes \mathbf{e}_l = \mathbf{I} \underline{\otimes} \mathbf{I} . \quad (\text{A.11})$$

Based on the above definition, we have

$$\overline{\mathbb{I}} : \mathbf{A} = \mathbf{A} : \overline{\mathbb{I}} = \mathbf{A} , \quad \underline{\mathbb{I}} : \mathbf{A} = \mathbf{A} : \underline{\mathbb{I}} = \mathbf{A}^T , \quad (\text{A.12})$$

where $\mathbf{A}$ is any second-order tensor. Also, it can be shown that for any two vectors $\mathbf{a}$ and $\mathbf{b}$

$$\mathbf{a} \cdot \overline{\mathbb{I}} \cdot \mathbf{b} = (\overline{\mathbb{I}} \cdot \mathbf{a}) \cdot \mathbf{b} = \mathbf{a} \cdot (\mathbf{b} \cdot \overline{\mathbb{I}}) = \mathbf{b} \otimes \mathbf{a} , \quad (\text{A.13})$$

$$\mathbf{a} \cdot \underline{\mathbb{I}} \cdot \mathbf{b} = (\mathbf{a} \cdot \mathbf{b}) \mathbf{I} , \quad \mathbf{a} \cdot (\mathbf{b} \cdot \underline{\mathbb{I}}) = (\underline{\mathbb{I}} \cdot \mathbf{a}) \cdot \mathbf{b} = \mathbf{a} \otimes \mathbf{b} . \quad (\text{A.14})$$

and

$$\overline{\mathbb{I}} \cdot \mathbf{a} = \mathbf{a} \cdot \overline{\mathbb{I}} = \mathbf{I} \overline{\otimes} \mathbf{a} = \mathbf{I} \underline{\otimes} \mathbf{a} , \quad \underline{\mathbb{I}} \cdot \mathbf{a} = \mathbf{a} \otimes \mathbf{I} , \quad \mathbf{a} \cdot \underline{\mathbb{I}} = \mathbf{I} \otimes \mathbf{a} . \quad (\text{A.15})$$

The symmetric and skew-symmetric fourth-order identity tensors are defined as

$$\mathbb{I}^{\text{sym}} = \frac{1}{2} (\overline{\mathbb{I}} + \underline{\mathbb{I}}) , \quad \mathbb{I}^{\text{skw}} = \frac{1}{2} (\overline{\mathbb{I}} - \underline{\mathbb{I}}) . \quad (\text{A.16})$$

These two tensors can act as operators that symmetrize and skew-symmetrize second-order tensors, for instance

$$\mathbb{I}^{\text{sym}} : \mathbf{u} \nabla = \frac{1}{2} (\mathbf{u} \nabla + \nabla \mathbf{u}) = \boldsymbol{\varepsilon} , \quad \mathbb{I}^{\text{skw}} : \mathbf{u} \nabla = \frac{1}{2} (\mathbf{u} \nabla - \nabla \mathbf{u}) = \boldsymbol{\omega} . \quad (\text{A.17})$$

where $\boldsymbol{\varepsilon}$ and $\boldsymbol{\omega}$ denote the linearized strain tensor and the rotation tensor respectively. For particular purposes we go on to define the following fourth-order operators

$$\mathbb{I}^{\text{vol}} = \frac{1}{3} (\mathbf{I} \otimes \mathbf{I}) , \quad \mathbb{I}^{\text{dev}} = \mathbb{I}^{\text{sym}} - \mathbb{I}^{\text{vol}} , \quad (\text{A.18})$$

which, as the notation suggests, extract volumetric and deviatoric parts of strain as

$$\mathbb{I}^{\text{vol}} : \mathbf{u} \nabla = \frac{1}{3} \text{tr}(\boldsymbol{\varepsilon}) \mathbf{I} = \boldsymbol{\varepsilon}^{\text{vol}} , \quad \mathbb{I}^{\text{dev}} : \mathbf{u} \nabla = \boldsymbol{\varepsilon} - \frac{1}{3} \text{tr}(\boldsymbol{\varepsilon}) \mathbf{I} = \boldsymbol{\varepsilon}^{\text{dev}} . \quad (\text{A.19})$$

Some formulas useful for elasticity theory

$$\frac{\partial \mathbf{A}^T \mathbf{A}}{\partial \mathbf{A}} = \mathbf{I} \bar{\otimes} \mathbf{A}^T + \mathbf{A}^T \bar{\otimes} \mathbf{I} , \quad \frac{\partial \mathbf{A} \mathbf{A}^T}{\partial \mathbf{A}} = \mathbf{I} \bar{\otimes} \mathbf{A} + \mathbf{A} \bar{\otimes} \mathbf{I} . \quad (\text{A.20})$$

$$\frac{\partial \mathbf{A}}{\partial \mathbf{A}_{\text{sym}}} = 2 \bar{\mathbb{I}} + 2 \underline{\mathbb{I}} - 3 \mathbf{I} \otimes \mathbf{I} = \frac{\partial \mathbf{A}^T}{\partial \mathbf{A}_{\text{sym}}} , \quad (\text{A.21a})$$

$$\frac{\partial \mathbf{A} \mathbf{A}^T}{\partial \mathbf{A}_{\text{sym}}} = 2 (\mathbf{I} \bar{\otimes} \mathbf{A} + \mathbf{A} \bar{\otimes} \mathbf{I} + \mathbf{I} \underline{\otimes} \mathbf{A} + \mathbf{A} \underline{\otimes} \mathbf{I}) - 3 (\mathbf{A}^T \otimes \mathbf{I} - \mathbf{A} \otimes \mathbf{I}) , \quad (\text{A.21b})$$

$$\frac{\partial \mathbf{A}^T \mathbf{A}}{\partial \mathbf{A}_{\text{sym}}} = 2 (\mathbf{I} \bar{\otimes} \mathbf{A}^T + \mathbf{A}^T \bar{\otimes} \mathbf{I} + \mathbf{I} \underline{\otimes} \mathbf{A}^T + \mathbf{A}^T \underline{\otimes} \mathbf{I}) - 3 (\mathbf{A} \otimes \mathbf{I} - \mathbf{A}^T \otimes \mathbf{I}) . \quad (\text{A.21c})$$

$$\frac{\partial \det \mathbf{F}}{\partial \mathbf{F}} = \det \mathbf{F} \mathbf{F}^{-T} = \text{Cof } \mathbf{F} . \quad (\text{A.22})$$

A.2 Average compatibility condition

For a solid microstructure composed of N phases with order parameters (deformation gradients) $\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_N$ and volume fractions $\lambda_1, \lambda_2, \dots, \lambda_N$, the average compatibility conditions are given by (see [25, 109, pp. 132])

$$\begin{aligned} \langle \mathbf{F} \rangle &= \sum_{i=1}^N \lambda_i \mathbf{F}_i \\ \text{Cof} \langle \mathbf{F} \rangle &= \sum_{i=1}^N \lambda_i \text{Cof } \mathbf{F}_i \\ \det \langle \mathbf{F} \rangle &= \sum_{i=1}^N \lambda_i \det \mathbf{F}_i . \end{aligned} \quad (\text{A.23})$$

A.3 An interpretation of Lagrange multipliers

A constrained optimization problem is typically formulated as

$$f(x_i) \rightarrow \min , \quad \text{under } g_j(x_i) = u_j , \quad (\text{A.24})$$

where $i \in \{1, \dots, N\}$ and $j \in \{1, \dots, M\}$. Lagrange formulation of the problem reads

$$\min_{\mathbf{x}, \boldsymbol{\lambda}} \phi(\mathbf{x}, \boldsymbol{\lambda}) \ , \quad \text{where} \quad \phi(\mathbf{x}, \boldsymbol{\lambda}) := f(\mathbf{x}) - \boldsymbol{\lambda} \cdot [\mathbf{g}(\mathbf{x}) - \mathbf{u}] \ . \quad (\text{A.25})$$

Now suppose that

$$\arg \min_{\mathbf{x}, \boldsymbol{\lambda}} \phi(\mathbf{x}, \boldsymbol{\lambda}) = (\mathbf{x}^*, \boldsymbol{\lambda}^*) \ , \quad \text{and} \quad \phi(\mathbf{x}^*, \boldsymbol{\lambda}^*) = \phi^* \ . \quad (\text{A.26})$$

It should be clear that ϕ^* is a function of $\mathbf{u}$, since $\phi(\mathbf{x}^*, \boldsymbol{\lambda}^*) = f(\mathbf{x}^*) - \boldsymbol{\lambda}^* \cdot [\mathbf{g}(\mathbf{x}^*) - \mathbf{u}]$, therefore

$$\frac{\partial \phi^*}{\partial \mathbf{u}} = \boldsymbol{\lambda}^* \ . \quad (\text{A.27})$$

In most situations $\mathbf{u}$ and ϕ , and therefore also ϕ^* , have specific and clear physical meanings, and so does the Lagrange multiplier, $\boldsymbol{\lambda}$, based in the above identity.

Appendix B

Elastic Constants of Monocrystalline

Suppose a homogeneous solid body of monocrystalline with lattice vectors $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ is at equilibrium and is large enough so that the thermomechanical effects of its boundary are negligible. Such a solid body represents an ideal case of anisotropy. If the crystal is composed of metallic substances, be it monatomic or an alloy¹², the corresponding lattice energy can be expressed as the sum of *pairwise interaction* potentials of all atoms [31, 79]:

$$E = \frac{1}{2} \sum_{i \neq j} \psi(|\mathbf{r}_{ij}|^2) , \quad (\text{B.1})$$

where $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ is the relative position of any two atoms within the lattice.

Remark 19. Many authors base their lattice energy formulations upon a pairwise potential in the form $\psi(|\mathbf{r}_{ij}|)$, which may seem a more appropriate choice, since it can represent arbitrary potentials more directly. For instance, the *Coulomb–Buckingham* potential

$$\psi(r) = \beta \left(e^{r/r_0} - \left(\frac{r_0}{r} \right)^6 \right) + \frac{q_1 q_2}{4\pi\epsilon_0} \frac{1}{r} ,$$

would require additional square-root operations when expressed in the form $\psi(|\mathbf{r}_{ij}|^2)$, which we prefer to use. While both formulations must eventually produce mathematically equivalent results, our particular form of potential in (B.1) has the advantage that the asymptotic expansion of the lattice energy will explicitly depend on the Green–Lagrange strain which in turn guarantees the *frame-indifference*. If we were to go with the widely employed form $\psi(|\mathbf{r}_{ij}|)$, we would need to enforce the symmetry requirements a posteriori. Hence, we prefer the original formulation proposed by Born and Huang 1967 [31] and by Huntington 1958 [79].

Let the entire lattice deform homogeneously under the deformation gradient $\mathbf{F}$ so slightly so that no plastic deformation or slip may take place. The energy of the lattice, in the absence

¹²We previously mentioned that in the case of alloys, the lattice *shift* can be eliminated through energy minimization. Therefore, only the simple case of a monatomic lattice needs to be considered.

of thermal effects, can be then written as

$$E(\mathbf{F}) = \frac{1}{2} \sum_{i \neq j} \psi(|\mathbf{F}\mathbf{r}_{ij}|^2), \quad (\text{B.2})$$

Since the deformation is supposed to remain within the elastic limit, it shall be small enough so that the energy can be approximated asymptotically around the equilibrium configuration, $\mathbf{F} = \mathbf{I}$. The argument of the potential in (B.1) varies according

$$|\mathbf{F}\mathbf{r}_{ij}|^2 - |\mathbf{r}_{ij}|^2 = \mathbf{r}_{ij} \cdot (\mathbf{F}^T \mathbf{F} - \mathbf{I}) \cdot \mathbf{r}_{ij} = \mathbf{r}_{ij} \cdot (\nabla \mathbf{u} + \mathbf{u} \nabla + \nabla \mathbf{u} \cdot \mathbf{u} \nabla) \cdot \mathbf{r}_{ij}. \quad (\text{B.3})$$

The Taylor's expansion of (B.2) around $\mathbf{F} = \mathbf{I}$, keeping up the terms up to $\mathcal{O}(|\mathbf{u} \nabla|^2)$, is given by

$$E \approx \frac{1}{2} \sum_{i \neq j} \left\{ \psi(\mathbf{r}_{ij}^2) + \psi'(\mathbf{r}_{ij}^2) [\mathbf{r}_{ij} \cdot (\nabla \mathbf{u} + \mathbf{u} \nabla + \nabla \mathbf{u} \cdot \mathbf{u} \nabla) \cdot \mathbf{r}_{ij}] + \frac{1}{2!} \psi''(\mathbf{r}_{ij}^2) [\mathbf{r}_{ij} \cdot (\nabla \mathbf{u} + \mathbf{u} \nabla) \cdot \mathbf{r}_{ij}]^2 \right\}. \quad (\text{B.4})$$

The first-order term in the expansion must vanish due to equilibrium, therefore $\partial E / \partial \boldsymbol{\varepsilon}|_0 = \mathbf{0}$; that is

$$\sum_{i \neq j} \psi'(\mathbf{r}_{ij}^2) [\mathbf{r}_{ij} \otimes \mathbf{r}_{ij}] = \mathbf{0}, \quad (\text{B.5})$$

which subsequently turns (B.4) into

$$E \approx \frac{1}{2} \sum_{i \neq j} \left\{ \psi(\mathbf{r}_{ij}^2) + \psi'(\mathbf{r}_{ij}^2) [\mathbf{r}_{ij} \cdot \nabla \mathbf{u} \cdot \mathbf{u} \nabla \cdot \mathbf{r}_{ij}] + \frac{1}{2} \psi''(\mathbf{r}_{ij}^2) [\mathbf{r}_{ij} \cdot (\nabla \mathbf{u} + \mathbf{u} \nabla) \cdot \mathbf{r}_{ij}]^2 \right\}. \quad (\text{B.6})$$

The forth order tensor of the elastic constants is defined as $\mathbb{C} = \partial^2 e / \partial \boldsymbol{\varepsilon} \otimes \partial \boldsymbol{\varepsilon}$, with $\boldsymbol{\varepsilon} = 1/2 (\mathbf{u} \nabla + \nabla \mathbf{u})$, being the linearized strain and e the energy density. In the following derivations we use the tensor identities given in equation (A.21).

The first derivative of the energy density with respect to the linearized strain equals the Cauchy stress tensor:

$$\boldsymbol{\sigma} = \frac{1}{2V} \sum_{i \neq j} \left\{ \psi'(\mathbf{r}_{ij}^2) \left[2(\mathbf{r}_{ij} \otimes (\mathbf{u} \nabla \cdot \mathbf{r}_{ij}) + (\mathbf{u} \nabla \cdot \mathbf{r}_{ij}) \otimes \mathbf{r}_{ij} + (\mathbf{r}_{ij} \cdot \nabla \mathbf{u}) \otimes \mathbf{r}_{ij} + \mathbf{r}_{ij} \otimes (\mathbf{r}_{ij} \cdot \nabla \mathbf{u})) - 3\mathbf{r}_{ij} \cdot (\mathbf{u} \nabla + \nabla \mathbf{u}) \cdot \mathbf{r}_{ij} \otimes \mathbf{I} \right] + 2\psi''(\mathbf{r}_{ij}^2) [(\mathbf{r}_{ij} \cdot (\nabla \mathbf{u} + \mathbf{u} \nabla) \cdot \mathbf{r}_{ij}) \mathbf{r}_{ij} \otimes \mathbf{r}_{ij}] \right\}. \quad (\text{B.7})$$

The second derivative, we mentioned, equals the elastic stiffness tensor and is given by

$$\begin{aligned} \mathbb{C} = \frac{1}{2V} \sum_{i \neq j} \left\{ \psi'(\mathbf{r}_{ij}^2) \left[8(\mathbf{I} \otimes (\mathbf{r}_{ij} \otimes \mathbf{r}_{ij}) + \mathbf{I} \otimes (\mathbf{r}_{ij} \otimes \mathbf{r}_{ij}) + (\mathbf{r}_{ij} \otimes \mathbf{r}_{ij}) \otimes \mathbf{I} + (\mathbf{r}_{ij} \otimes \mathbf{r}_{ij}) \otimes \mathbf{I} \right) \right. \\ \left. - 24(\mathbf{I} \otimes (\mathbf{r}_{ij} \otimes \mathbf{r}_{ij}) + (\mathbf{r}_{ij} \otimes \mathbf{r}_{ij}) \otimes \mathbf{I}) + 18(\mathbf{r}_{ij} \cdot \mathbf{r}_{ij}) \mathbf{I} \otimes \mathbf{I} \right] \\ \left. + 4\psi''(\mathbf{r}_{ij}^2) [\mathbf{r}_{ij} \otimes \mathbf{r}_{ij} \otimes \mathbf{r}_{ij} \otimes \mathbf{r}_{ij}] \right\}. \quad (\text{B.8}) \end{aligned}$$

For metallic crystals the atomic interactions are relatively short-range. So the summation over all pairwise atomic interactions within the lattice can be confined to interactions between the nearest neighbors plus the interactions between the second-nearest neighbors. For each atom in the lattice $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$, there are 6 nearest neighbors located at

$$\pm \mathbf{a}_1, \quad \pm \mathbf{a}_2, \quad \pm \mathbf{a}_3, \quad (\text{B.9})$$

and there are 12 second-nearest neighbors at

$$\pm \mathbf{a}_1 \pm \mathbf{a}_2, \quad \pm \mathbf{a}_2 \pm \mathbf{a}_3, \quad \pm \mathbf{a}_3 \pm \mathbf{a}_1. \quad (\text{B.10})$$

But we do not actually need 6 + 12 terms in the energy sum for each atom, because the pairwise potential depends on the distance between atoms only and not on their actual relative positions. Hence, the lattice energy (B.1) can be reduced to

$$E = \varrho V \sum_{i=1}^3 \left\{ \psi(|\mathbf{a}_i|^2) + \sum_{j>i}^3 \{ \psi(|\mathbf{a}_i + \mathbf{a}_j|^2) + \psi(|\mathbf{a}_i - \mathbf{a}_j|^2) \} \right\}. \quad (\text{B.11})$$

Notice the factor 1/2 has been cancelled out with the factor 2 in front each pairwise potential term due to energetically equivalent lattice directions. Putting this back to (B.6) will result in

$$\mathbb{C} = \varrho \sum_{i=1}^3 \left\{ \mathbb{K}(\mathbf{a}_i) + \sum_{j>i}^3 \left\{ \mathbb{K}(\mathbf{a}_i + \mathbf{a}_j) + \mathbb{K}(\mathbf{a}_i - \mathbf{a}_j) \right\} \right\}, \quad (\text{B.12})$$

where

$$\begin{aligned} \mathbb{K}(\mathbf{a}) = \psi'(\mathbf{a}^2) & \left[18(\mathbf{a} \cdot \mathbf{a}) \mathbf{I} \otimes \mathbf{I} + 8 \left(\mathbf{I} \overline{\otimes} (\mathbf{a} \otimes \mathbf{a}) + \mathbf{I} \underline{\otimes} (\mathbf{a} \otimes \mathbf{a}) + (\mathbf{a} \otimes \mathbf{a}) \overline{\otimes} \mathbf{I} + (\mathbf{a} \otimes \mathbf{a}) \underline{\otimes} \mathbf{I} \right) \right. \\ & \left. - 24 \left(\mathbf{I} \otimes (\mathbf{a} \otimes \mathbf{a}) + (\mathbf{a} \otimes \mathbf{a}) \otimes \mathbf{I} \right) \right] + 4\psi''(\mathbf{a}^2) \left[\mathbf{a} \otimes \mathbf{a} \otimes \mathbf{a} \otimes \mathbf{a} \right] \end{aligned} \quad (\text{B.13})$$

Example: cubic lattice

As a test case, we calculate the stiffness tensor for the simplest lattice structure, namely a monatomic cubic crystal. The primitive cell of a simple cubic crystal is written as $\{a\mathbf{e}_1, a\mathbf{e}_2, a\mathbf{e}_3\}$, in which a is the lattice parameter. Based on (B.12), the elastic constants of this lattice will be obtained as

$$C_{ijkl} = \begin{cases} 38c_1 + c_2 + 2c_3 & i = j = k = l \\ 6c_1 + c_3 & i = j, k = l \\ 16c_1 + c_3 & i = k, j = l \text{ or } i = l, j = k \\ 0 & \text{otherwise.} \end{cases} \quad (\text{B.14})$$

where

$$c_1 = \varrho a^2 \psi'(a^2) + 4\varrho a^2 \psi'(2a^2), \quad c_2 = 4\varrho a^4 \psi''(a^2), \quad c_3 = 8\varrho a^4 \psi''(2a^2). \quad (\text{B.15})$$

Most noticeably, there are only three independent elastic constants which is a well known property of simple cubic crystals. Further, as we mentioned, there is no need for enforcing the lattice symmetries explicitly, since the formulation reflects those symmetries inherently.

Appendix C

Scaling and asymptotic self-similarity

In the eyes of workers in applied branches of sciences, seeking scaling equations and self-similarities in a physical problem might merely look like an effort to find a rough first estimate between the physical quantities involved. From a mathematical viewpoint, however, asymptotically self-similar solutions indicate *stability* of the actual solution which is an invaluable piece of information especially in a computational treatment of the problem [15, 58]. They also enable us to obtain *estimates* (upper and lower bounds) to solution or to certain aspects of them which in turn serve as a powerful tool in studying existence and uniqueness.

Scaling

In simplest terms, *scaling* means that the dependence between two quantities is in the form of a *power-law*, such as $E \propto L^{1/3}$. In the study of macroscopic or mesoscopic physical phenomena, the scaling regime usually represents a state in which the overall behavior of the system is independent of the details of initial and boundary conditions, while the system has not yet reached its final state [59, 16, 58].

Self-similarity

According Π -theorem, if a relation

$$a = f(a_1, a_2, \dots, b_1, b_2) ,$$

between $k + 3$ physical quantities $a, a_1, a_2, \dots, a_k, b_1, b_2$ is sought, such that only k of them, namely $a_1, a_2, \dots, a_k$, have independent dimensions, then the function $f(\bullet)$ can be rewritten as

$$\Pi = \Phi(\Pi_1, \Pi_2) , \tag{C.1}$$

where

$$\Pi = \frac{a}{a_1^{\alpha_1} a_2^{\alpha_2} \dots a_k^{\alpha_k}} , \quad \Pi_1 = \frac{b_1}{a_1^{\beta_1} a_2^{\beta_2} \dots a_k^{\beta_k}} , \quad \Pi_2 = \frac{b_2}{a_1^{\gamma_1} a_2^{\gamma_2} \dots a_k^{\gamma_k}} . \tag{C.2}$$

The exponents α_i and β_i can be determined by mere dimensional analysis. Assume a *limiting case* where one of the independent variables, say Π_2 , tends towards zero or infinity. This typically corresponds to an idealized situation in which a characteristic length or a characteristic time or any other characteristic quantity with the same dimension as b_2 becomes very small or grows very large. Due to this idealization, as $\Pi_2 \rightarrow 0$ or $\Pi_2 \rightarrow \infty$ one of the following three outcomes will arise.

1. NON-SINGULAR SELF-SIMILARITY: The function tends to a well-defined nonzero limit, say $\Phi \rightarrow \check{\Phi}_1(\Pi_1)$. This limit can be obtained by direct substitution of $\check{\Phi}_1(\Pi_1) = \Phi|_{\Pi_2=0}$, or $\check{\Phi}_1(\Pi_1) = \Phi|_{1/\Pi_2=0}$, and the functionality $\Pi = \Phi(\Pi_1)$ can apparently be written in dimensional form

$$a = a_1^{\alpha_1} a_2^{\alpha_2} \cdots a_k^{\alpha_k} f_1 \left(\frac{b_1}{a_1^{\beta_1} a_2^{\beta_2} \cdots a_k^{\beta_k}} \right), \quad (\text{C.3})$$

which has the property of *generalized homogeneity*. This does not immediately exhibit an obvious case of self-similar behavior with respect to different values of b_1 ; nevertheless, one can see the similarity structure with respect to any transformation $a_i \mapsto c_i a_i$ which in turn transforms $a \mapsto c_i^{-\alpha_i/\beta_i} a$, which ultimately implies a self-similar behavior.

2. ASYMPTOTIC SELF-SIMILARITY: The function does not have a well-defined nonzero limit, instead it tends toward a *power-law asymptotic* function with respect to Π_2 , which again possesses generalized homogeneity:

$$\Phi \rightarrow \Pi_2^{p_1} \check{\Phi}_1 \left(\frac{\Pi_1}{\Pi_2^{p_2}} \right). \quad (\text{C.4})$$

This is where we say that the solution is *asymptotically self-similar*, since substituting $\Pi_2 = 0$ into Φ , i.e. $\Phi|_{\Pi_2=0}$, does not give a finite nonzero function. Here, Φ might be weakly convergent. Hypothetically, we should be able to determine the exponents p_1 and p_2 from the governing equations, often PDEs.

3. OTHERWISE: The function has neither a well-defined limit, nor does it have an asymptotic power-law limit. This goes well beyond the scope of our work. Suffice it to mention that more sophisticated formulations are to be conceived than what Π -theorem provides.

Example: asymptotically self-similar branching of the laminate

There are branching constructs of continuous deformation fields composed of eigen deformations, in which the volume fraction does not necessarily change along the laminate [38]. The minimizing sequence, built upon these constructs, is asymptotically self-similar in the sense that the sequence

$$u^j(x, y) = \varepsilon^{-2j/3} u(\varepsilon^j x, \varepsilon^{2j/3} y), \quad 0 < \varepsilon < 1, \quad (\text{C.5})$$

has a strongly convergent subsequence in $W^{1,2}$.

Appendix D

Nonlinear laminate kinematics, a possible extension

Without loss of generality we define a deformation map from the eigen configuration $\hat{\mathbf{x}}$ to the accommodated configuration $\mathbf{x}$:

$$\mathbf{X} \xrightarrow{\mathbf{F}^\circ} \mathbf{x}^\circ \xrightarrow{\Phi} \mathbf{x}. \quad (\text{D.1})$$

Since $\mathbf{F} = \partial \mathbf{x} / \partial \mathbf{X}$, we have $\mathbf{F} = \Phi \mathbf{F}^\circ$, and therefore

$$\Phi = \mathbf{F} \mathbf{F}^{\circ -1} = \begin{cases} \Phi_1(x_1^\circ, x_2^\circ), & -\lambda^\circ h^\circ \leq x_3^\circ \leq 0, \\ \Phi_2(x_1^\circ, x_2^\circ), & 0 \leq x_3^\circ \leq (1 - \lambda^\circ) h^\circ, \end{cases} \quad (\text{D.2})$$

where $\Phi_i = \mathbf{F}_i \cdot (\mathbf{F}_i^\circ)^{-1}$, and

$$h^\circ = h^\circ(x_1^\circ, x_2^\circ), \quad \lambda^\circ = \lambda^\circ(x_1^\circ, x_2^\circ). \quad (\text{D.3})$$

In the deformation route (D.1) the austenitic parent $\mathbf{X}$ is first deformed via martensitic transformation $\mathbf{F}_i^\circ$ to form martensitic laminate in its eigen configuration $\mathbf{x}^\circ$. It is then deformed by Φ into its accommodated state $\mathbf{x}$. It therefore matches the homogeneous mesoscopic deformation according (3.22).

At this point there are various questions to be answered. The most conspicuous one is whether the overall (average) orientation of the laminate remains unchanged throughout the deformation steps in (D.1); and if not, how it changes in each step. We will pose and answer these important questions in the next section.

In the symmetry relation (3.28), since $\mathbf{Q}$ belongs to the point group of the austenite, $\det \mathbf{Q} = \det \mathbf{Q}^T = 1$; therefore

$$\det \mathbf{F}_2^\circ = \det(\mathbf{Q}^T \mathbf{F}_1^\circ \mathbf{Q}) = \det \mathbf{Q}^T \det \mathbf{F}_1^\circ \det \mathbf{Q} = \det \mathbf{F}_1^\circ. \quad (\text{3.31})$$

Let us now take the determinant of both sides of (3.9) and apply the above lemma:

$$\det \langle \mathbf{F}^\circ \rangle = \det \left(\mathbf{F}_1^\circ + (1 - \Lambda) \mathbf{a}^\circ \otimes \hat{\mathbf{X}}_3 \right) = \left(1 + (1 - \Lambda) \hat{\mathbf{X}}_3 \cdot (\mathbf{F}_1^\circ)^{-1} \cdot \mathbf{a}^\circ \right) \det \mathbf{F}_1^\circ. \quad (\text{D.4})$$

In order to evaluate the rightmost term, we take the determinant of a slightly altered version of (3.27), using the above lemma to arrive at

$$\mathbf{F}_2^\circ = \mathbf{F}_1^\circ + \mathbf{a}^\circ \otimes \hat{\mathbf{X}}_3 \Rightarrow \det \mathbf{F}_2^\circ = \det \left(\mathbf{F}_1^\circ + \mathbf{a}^\circ \otimes \hat{\mathbf{X}}_3 \right) = \left(1 + \hat{\mathbf{X}}_3 \cdot (\mathbf{F}_1^\circ)^{-1} \cdot \mathbf{a}^\circ \right) \det \mathbf{F}_1^\circ.$$

However, we just showed in (3.31) that $\det \mathbf{F}_1^\circ = \det \mathbf{F}_2^\circ$, which results in

$$\left(1 + \hat{\mathbf{X}}_3 \cdot (\mathbf{F}_1^\circ)^{-1} \cdot \mathbf{a}^\circ \right) = 1 \Rightarrow \hat{\mathbf{X}}_3 \cdot (\mathbf{F}_1^\circ)^{-1} \cdot \mathbf{a}^\circ = 0. \quad (\text{D.5})$$

Applying this to (D.4) combined with (3.31) will prove that

$$\det \langle \mathbf{F}^\circ \rangle = \det \mathbf{F}_1^\circ = \det \mathbf{F}_2^\circ. \quad (\text{D.6})$$

Having this identity, one can repeat the same procedure that proved (D.5) to show

$$\hat{\mathbf{X}}_3 \cdot (\mathbf{F}_1^\circ)^{-1} \cdot \mathbf{a}^\circ = \hat{\mathbf{X}}_3 \cdot (\mathbf{F}_2^\circ)^{-1} \cdot \mathbf{a}^\circ = \hat{\mathbf{X}}_3 \cdot \langle \mathbf{F}^\circ \rangle^{-1} \cdot \mathbf{a}^\circ = 0. \quad (\text{D.7})$$

Recall that the cofactor of a matrix $\text{Cof } \mathbf{A}$ can be written as $\det \mathbf{A} \mathbf{A}^{-T}$; hence, equations (3.10) and (D.6) put together, keeping in mind that $\det \mathbf{F}_i > 0$, yield

$$\langle \mathbf{F}^\circ \rangle^{-T} \hat{\mathbf{X}}_3 = (\mathbf{F}_1^\circ)^{-T} \hat{\mathbf{X}}_3 = (\mathbf{F}_2^\circ)^{-T} \hat{\mathbf{X}}_3. \quad (\text{D.8})$$

$$d\alpha = da, \quad \hat{\xi}_3^\circ = \hat{\mathbf{x}}_3^\circ. \quad (\text{D.9})$$

Let us step back for a moment and study these two equations. Basically what they say is that the multiplicative decomposition of the eigen deformation gradient into an average and a fluctuating part simplifies the formulation to a great extent; because the fluctuation part $\widetilde{\mathbf{F}}^\circ$ neither moves nor stretches the twin interface. Thus, introducing the intermediate coordinates ξ° pertaining to the average deformation properly marks the separation between two modes of deformation: one that stretches and rotates the austenite, and one that oscillates and forms martensitic variants and their interfaces without stretching or rotating the body in average.

Equation (D.8), however, simplifies these two equations further into two very interesting results, which are

$$\lambda^\circ = A, \quad h^\circ = \frac{H}{\left\| \langle \mathbf{F}^\circ \rangle^{-T} \hat{\mathbf{X}}_3 \right\|}. \quad (\text{D.10})$$

Remark 20. Based on equations (D.2) and (3.31), the determinants of accommodated deformation gradients depend on Φ_i 's:

$$\left. \begin{aligned} \det \mathbf{F}_1 &= \det(\Phi_1 \mathbf{F}_1^\circ) = \det \Phi_1 \det \mathbf{F}_1^\circ \\ \det \mathbf{F}_2 &= \det(\Phi_2 \mathbf{F}_2^\circ) = \det \Phi_2 \det \mathbf{F}_2^\circ \end{aligned} \right\} \Rightarrow \det \mathbf{F}_2 = \frac{\det \Phi_2}{\det \Phi_1} \det \mathbf{F}_1. \quad (\text{D.11})$$

Since there is no a priori assumption on $\det \Phi_i$, we **cannot** assume that $\det \mathbf{F}_1 = \det \mathbf{F}_2$, as we did for eigen deformations. This causes the equations (D.6), (D.7), and (D.8) to fail for the accommodated configuration. Notice that the results (3.8), (3.9), and (3.10) are mere consequences of the coherence condition, and therefore hold for $\mathbf{F}_i$'s. In general for

the twin-compatible phases $\mathbf{F}_1$ and $\mathbf{F}_2$, not necessarily symmetry-connected as in (3.28), the following identities hold:

Next, let us find a more specific formula of the form (3.5) by substituting (3.9), for the accommodated configuration, back in (3.5) to arrive at

$$\mathbf{F}^\circ(\mathbf{X}) = \langle \mathbf{F}^\circ \rangle + \begin{cases} -(1 - \Lambda) \mathbf{a}^\circ \otimes \hat{\mathbf{X}}_3, & -\Lambda H \leq X_3 \leq 0, \\ \Lambda \mathbf{a}^\circ \otimes \hat{\mathbf{X}}_3, & 0 \leq X_3 \leq (1 - \Lambda) H. \end{cases} \quad (\text{D.12})$$

Note that the remark 20 tells us that equation (3.9) also holds for the accommodated configuration.

Using an adaptation of (3.21) in the form $\tilde{\mathbf{F}}^\circ = \mathbf{F}^\circ \langle \mathbf{F}^\circ \rangle^{-1}$, we can rewrite the above formula in terms of the fluctuating part of deformation as

$$\tilde{\mathbf{F}}^\circ(\xi^\circ) = \mathbf{I} + \begin{cases} -(1 - \Lambda) \mathbf{a}^\circ \otimes \left(\langle \mathbf{F}^\circ \rangle^{-T} \hat{\mathbf{X}}_3 \right), & -\lambda^\circ h^\circ \leq \xi_3^\circ \leq 0, \\ \Lambda \mathbf{a}^\circ \otimes \left(\langle \mathbf{F}^\circ \rangle^{-T} \hat{\mathbf{X}}_3 \right), & 0 \leq \xi_3^\circ \leq (1 - \lambda^\circ) h^\circ. \end{cases} \quad (\text{D.13})$$

The term in parentheses $\left(\langle \mathbf{F}^\circ \rangle^{-T} \hat{\mathbf{X}}_3 \right)$ can be slightly modified using (D.9) and (D.10) to get

$$\langle \mathbf{F}^\circ \rangle^{-T} \hat{\mathbf{X}}_3 = \frac{H}{h^\circ} \hat{\xi}_3^\circ. \quad (\text{D.14})$$

By substituting this and (D.10) into (D.13) we have

$$\tilde{\mathbf{F}}^\circ(\xi^\circ) = \mathbf{I} + \frac{H}{h^\circ} \begin{cases} -(1 - \lambda^\circ) \mathbf{a}^\circ \otimes \hat{\xi}_3^\circ, & -\lambda^\circ h^\circ \leq \xi_3^\circ \leq 0, \\ \lambda^\circ \mathbf{a}^\circ \otimes \hat{\xi}_3^\circ, & 0 \leq \xi_3^\circ \leq (1 - \lambda^\circ) h^\circ. \end{cases} \quad (\text{D.15})$$

There is one finishing touch to this formulation which has to do with $\mathbf{a}^\circ$, the jump across the twin interface. Looking back on (D.1) we have

$$\mathbf{a} = \Phi \mathbf{a}^\circ = \mathbf{F} (\mathbf{F}^\circ)^{-1} \mathbf{a}^\circ. \quad (\text{D.16})$$

The grain average deformation gradient $\langle \mathbf{F} \rangle$ is predetermined by mesoscopic deformation and is therefore constant. The average eigen deformation gradient $\langle \mathbf{F}^\circ \rangle$, however, depends on the volume fraction according (3.9) and (3.27) as

$$\langle \mathbf{F}^\circ \rangle = \mathbf{F}_1^\circ + (1 - \Lambda) \mathbf{a}^\circ \otimes \hat{\mathbf{X}}_3 = \mathbf{F}_2^\circ - \Lambda \mathbf{a}^\circ \otimes \hat{\mathbf{X}}_3. \quad (\text{D.17})$$

In order to find the inverse, we factor the $\mathbf{F}_i^\circ$'s out, and then inverse to obtain

$$\langle \mathbf{F}^\circ \rangle^{-1} = \left(\mathbf{I} + (1 - \Lambda) (\mathbf{F}_1^\circ)^{-1} \mathbf{a}^\circ \otimes \hat{\mathbf{X}}_3 \right)^{-1} (\mathbf{F}_1^\circ)^{-1} = \left(\mathbf{I} - \Lambda (\mathbf{F}_2^\circ)^{-1} \mathbf{a}^\circ \otimes \hat{\mathbf{X}}_3 \right)^{-1} (\mathbf{F}_2^\circ)^{-1}. \quad (\text{D.18})$$

Lemma 4. For any two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, such that $\mathbf{a} \cdot \mathbf{b} = 0$, we have

$$(\mathbf{I} + \mathbf{a} \otimes \mathbf{b})^{-1} = \mathbf{I} - \mathbf{a} \otimes \mathbf{b}.$$

The proof is straightforward. Multiply $(\mathbf{I} - \mathbf{a} \otimes \mathbf{b})$ by $(\mathbf{I} + \mathbf{a} \otimes \mathbf{b})$ from left and right to obtain the identity matrix.

In order to apply this lemma to (D.18) the orthogonality of the two vectors $\hat{\mathbf{X}}_3$ and $(\mathbf{F}_i^\circ)^{-1} \mathbf{a}^\circ$ is required. The result in (D.7) holds for the eigen state a fortiori, viz

$$\hat{\mathbf{X}}_3 \cdot (\mathbf{F}_1^\circ)^{-1} \cdot \mathbf{a}^\circ = \hat{\mathbf{X}}_3 \cdot (\mathbf{F}_2^\circ)^{-1} \cdot \mathbf{a}^\circ = 0.$$

Therefore, the equation (D.18) becomes

$$\begin{aligned} \langle \mathbf{F}^\circ \rangle^{-1} &= \left(\mathbf{I} + (1 - \Lambda) (\mathbf{F}_1^\circ)^{-1} \mathbf{a}^\circ \otimes \hat{\mathbf{X}}_3 \right)^{-1} (\mathbf{F}_1^\circ)^{-1} \\ &= (\mathbf{F}_1^\circ)^{-1} - (1 - \Lambda) \left((\mathbf{F}_1^\circ)^{-1} \mathbf{a}^\circ \right) \otimes \left((\mathbf{F}_1^\circ)^{-T} \hat{\mathbf{X}}_3 \right) \\ &= (\mathbf{F}_1^\circ)^{-1} \left(\mathbf{I} - (1 - \Lambda) \mathbf{a}^\circ \otimes \left((\mathbf{F}_1^\circ)^{-T} \hat{\mathbf{X}}_3 \right) \right) \end{aligned} \quad (\text{D.19})$$

$$= (\mathbf{F}_1^\circ)^{-1} \left(\mathbf{I} - (1 - \Lambda) \frac{H}{h} \mathbf{a}^\circ \otimes \hat{\xi}_3 \right), \quad (\text{D.20})$$

and

$$\begin{aligned} \langle \mathbf{F}^\circ \rangle^{-1} &= \left(\mathbf{I} - \Lambda (\mathbf{F}_2^\circ)^{-1} \mathbf{a}^\circ \otimes \hat{\mathbf{X}}_3 \right)^{-1} (\mathbf{F}_2^\circ)^{-1} \\ &= (\mathbf{F}_2^\circ)^{-1} + \Lambda \left((\mathbf{F}_2^\circ)^{-1} \mathbf{a}^\circ \right) \otimes \left((\mathbf{F}_2^\circ)^{-T} \hat{\mathbf{X}}_3 \right) \\ &= (\mathbf{F}_2^\circ)^{-1} \left(\mathbf{I} + \Lambda \mathbf{a}^\circ \otimes \left((\mathbf{F}_2^\circ)^{-T} \hat{\mathbf{X}}_3 \right) \right) \end{aligned} \quad (\text{D.21})$$

$$= (\mathbf{F}_2^\circ)^{-1} \left(\mathbf{I} + \Lambda \frac{H}{h} \mathbf{a}^\circ \otimes \hat{\xi}_3 \right). \quad (\text{D.22})$$

If we substitute (D.19) and (D.21) back into (D.16) and put the result in (3.11), then it will take on the form

$$\mathbf{F}(\mathbf{X}) = \langle \mathbf{F} \rangle + \langle \mathbf{F} \rangle \begin{cases} -(1 - \Lambda) (\mathbf{F}_1^\circ)^{-1} \mathbf{a}^\circ \otimes \hat{\mathbf{X}}_3, & -\Lambda H \leq X_3 \leq 0, \\ \Lambda (\mathbf{F}_2^\circ)^{-1} \mathbf{a}^\circ \otimes \hat{\mathbf{X}}_3, & 0 \leq X_3 \leq (1 - \Lambda) H. \end{cases} \quad (\text{D.23})$$

In deriving this equation we have made use of the fact that $\hat{\mathbf{X}}_3 (\mathbf{F}_i^\circ)^{-1} \mathbf{a}^\circ = 0$, based on the trivial assertion that the result (D.7) is also true in eigen configuration.

Once again, give that $\tilde{\mathbf{F}} = \mathbf{F} \langle \mathbf{F} \rangle^{-1}$, we can reformulate this in terms of the fluctuating part of deformation gradient, in line with (D.15), to obtain

$$\tilde{\mathbf{F}}(\xi) = \mathbf{I} + \frac{H}{h} \langle \mathbf{F} \rangle \begin{cases} -(1 - \lambda) (\mathbf{F}_1^\circ)^{-1} \mathbf{a}^\circ \otimes \hat{\xi}_3, & -\lambda h \leq \xi_3 \leq 0, \\ \lambda (\mathbf{F}_2^\circ)^{-1} \mathbf{a}^\circ \otimes \hat{\xi}_3, & 0 \leq \xi_3 \leq (1 - \lambda) h. \end{cases} \quad (\text{D.24})$$

Before concluding this section we would like to emphasize that all the results obtained here are the consequences of three assumptions only, which are

1. the coherence condition between the eigen deformations given by (3.27),
2. the symmetry connection between the variants given by (3.28),
3. and finally our fundamental kinematic assumption formulated in (D.1).

We shall also mention that all the assertions proven for the accommodated configuration, $\mathbf{F}$, hold for the eigen configuration, $\mathbf{F}^\circ$, even by a stronger argument since the third of the above three assumptions will be extraneous.

Bibliography

- [1] H. I. Aaronson, M. Enomoto, and J. K. Lee. *Mechanisms of Diffusional Phase Transformations in Metals and Alloys*. CRC Press, 2010.
- [2] A. Abdulle and A. Nonnenmacher. A short and versatile finite element multiscale code for homogenization problems. *Computer Methods in Applied Mechanics and Engineering*, 198(37-40):2839–2859, Aug. 2009.
- [3] R. Abeyaratne, C. Chu, and R. D. James. Kinetics of materials with wiggly energies: Theory and application to the evolution of twinning microstructures in a Cu-Al-Ni shape memory alloy. *Philosophical Magazine A*, 73(2):457–497, Feb. 1996.
- [4] A. W. Adamson. *Physical chemistry of surfaces*. A Wiley-Interscience publication. Wiley, 1990.
- [5] J. M. Ball. Convexity conditions and existence theorems in nonlinear elasticity. *Archive for Rational Mechanics and Analysis*, 63(4):337–403, Dec. 1976.
- [6] J. M. Ball. Mathematical models of martensitic microstructure. *Materials Science and Engineering: A*, 378(1-2):61–69, July 2004.
- [7] J. M. Ball and C. Carstensen. Nonclassical Austenite-Martensite Interfaces. *Le Journal de Physique IV*, 07(C5):C5–35–C5–40, Nov. 1997.
- [8] J. M. Ball and C. Carstensen. Compatibility conditions for microstructures and the austenite–martensite transition. *Materials Science and Engineering: A*, 273-275:231–236, Dec. 1999.
- [9] J. M. Ball, C. Chu, and R. D. James. Hysteresis During Stress-Induced Variant Rearrangement. *Le Journal de Physique IV*, 05(C8):C8–245–C8–251, Dec. 1995.
- [10] J. M. Ball, J. C. Currie, and P. J. Olver. Null Lagrangians, weak continuity, and variational problems of arbitrary order. *Journal of Functional Analysis*, 41(2):135–174, 1981.
- [11] J. M. Ball and R. D. James. Fine phase mixtures as minimizers of energy. *Archive for Rational Mechanics and Analysis*, 100(1):13–52, 1987.

- [12] J. M. Ball and R. D. James. Proposed Experimental Tests of a Theory of Fine Microstructure and the Two-Well Problem. *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences*, 338(1650):389–450, Feb. 1992.
- [13] J. M. Ball and F. Murat. $W^{1,p}$ -quasiconvexity and variational problems for multiple integrals. *Journal of Functional Analysis*, 58(3):225–253, Oct. 1984.
- [14] W. Bangerth, T. Heister, L. Heltai, G. Kanschat, M. Kronbichler, M. Maier, B. Turcksin, and T. Young. The dealii library, version 8.2. *Archive of Numerical Software*, 3(100), 2015.
- [15] G. Barenblatt. *Scaling, Self-similarity, and Intermediate Asymptotics: Dimensional Analysis and Intermediate Asymptotics*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 1996.
- [16] G. I. Barenblatt and Y. B. Zel'Dovich. Self-similar solutions as intermediate asymptotics. *Annual Review of Fluid ...*, 1972.
- [17] S. Baroni, S. de Gironcoli, A. D. Corso, and P. Giannozzi. Phonons and related crystal properties from density-functional perturbation theory. *Reviews of Modern Physics*, 73(2):515–562, July 2001.
- [18] G. Barsch, B. Horovitz, and J. Krumhansl. Dynamics of twin boundaries in martensites. *Physical Review Letters*, 59(11):1251–1254, Sept. 1987.
- [19] T. Bartel. *Multiskalenmodellierung martensitischer Phasentransformationen in Formgedächtnislegierungen unter Verwendung relaxierter Energiepotenziale*. PhD thesis, google.de.
- [20] R. Becker. Die Keimbildung bei der Ausscheidung in metallischen Mischkristallen. *Annalen der Physik*, 424(1-2):128–140, Jan. 1938.
- [21] A. Berezovski, J. Engelbrecht, and G. A. Maugin. Internal variables and generalized continuum theories. ... *on Progress in the Theory and ...*, 2009.
- [22] K. Bhattacharya. Wedge-Like Microstructure in Martensites. *Acta Metallurgica Et Materialia*, 39(10):2431–2444, Oct. 1991.
- [23] K. Bhattacharya. Self-accommodation in martensite. *Archive for Rational Mechanics and Analysis*, 120(3):201–244, 1992.
- [24] K. Bhattacharya. Comparison of the geometrically nonlinear and linear theories of martensitic transformation. *Continuum Mechanics and Thermodynamics*, 5(3):205–242, 1993.

- [25] K. Bhattacharya. *Microstructure of Martensite*. Why it forms and how it gives rise to the shape-memory effect. Oxford University Press, Nov. 2003.
- [26] K. Bhattacharya, B. Li, and M. Luskin. The Simply Laminated Microstructure in Martensitic Crystals that Undergo a Cubic-to-Orthorhombic Phase Transformation. *Archive for Rational Mechanics and Analysis*, 149(2):123–154, Oct. 1999.
- [27] T. Biggs, M. Cortie, M. Witcomb, and L. Cornish. Platinum alloys for shape memory applications. *Platinum Metals Review*, 47(4):142–156, 2003.
- [28] B. A. Bilby and J. W. Christian. *Martensitic transformations*. The mechanism of phase, 1956.
- [29] J. M. Blakely. *Introduction to the Properties of Crystal Surfaces*. International Series on Materials Science and Technology. Elsevier, Oct. 2013.
- [30] J. M. Blakely and W. S. Owen. *Introduction to the Properties of Crystal Surfaces: International Series on Materials Science and Technology*. International series on materials science and technology. Elsevier Science, 2013.
- [31] M. Born and K. Huang. *Dynamical Theory of Crystal Lattices*. International series of monographs on physics. Clarendon Press, 1998.
- [32] S. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Methods*. Texts in Applied Mathematics. Springer New York, 2013.
- [33] H. J. Butt, K. Graf, and M. Kappl. *Physics and Chemistry of Interfaces*. Physics textbook. Wiley, 2006.
- [34] M. Chipot and D. Kinderlehrer. Equilibrium configurations of crystals. *Archive for Rational Mechanics and Analysis*, 103(3):237–277, 1988.
- [35] P. Chou and N. Pagano. *Elasticity: Tensor, Dyadic, and Engineering Approaches*. Dover Civil and Mechanical Engineering. Dover Publications, 2013.
- [36] C. H. Chu. *Hysteresis and Microstructures: A Study of Biaxial Loading on Compound Twins of Copper-aluminum-nickel Single Crystals*. PhD thesis, University of Minnesota, 1993.
- [37] P. Ciarlet. *Mathematical Elasticity: Three-dimensional elasticity*. Number v. 1 in Mathematical Elasticity. North-Holland, 1993.
- [38] S. Conti. Branched microstructures: Scaling and asymptotic self-similarity. *Communications on Pure and Applied Mathematics*, 53(11):1448–1474, Nov. 2000.
- [39] M. A. Crisfield. Accelerating and damping the modified Newton-Raphson method. *Computers & Structures*, 18(3):395–407, Jan. 1984.

- [40] B. Dacorogna. *Introduction to the Calculus of Variations*. Imperial College Press, 2004.
- [41] B. Dacorogna. *Direct Methods in the Calculus of Variations*. Applied Mathematical Sciences. Springer Berlin Heidelberg, 2012.
- [42] J. L. Ericksen. On the symmetry of deformable crystals. *Archive for Rational Mechanics and Analysis*, 1979.
- [43] J. L. Ericksen. Some phase transitions in crystals. *Archive for Rational Mechanics and Analysis*, 1980.
- [44] J. L. Ericksen. Changes in Symmetry in Elastic Crystals. In *Proceedings of the IUTAM Symposium on Finite Elasticity*, pages 167–177. Springer Netherlands, Dordrecht, Jan. 1982.
- [45] J. L. Ericksen. Twinning of Crystals (I). *Metastability and Incompletely Posed Problems*, 3(Chapter 6):77–93, 1987.
- [46] J. L. Ericksen. Weak martensitic transformations in Bravais lattices. *Mechanics and Thermodynamics of Continua*, 1991.
- [47] J. L. Ericksen. Twinning theory for some Pitteri neighborhoods. *Continuum Mechanics and Thermodynamics*, 14(3):249–262, June 2002.
- [48] J. D. Eshelby. A Tentative Theory of Metallic Whisker Growth. *Physical Review*, 91(3):755–756, Aug. 1953.
- [49] J. D. Eshelby. *Scope and limitations of the continuum approach*. Internal Stresses and Fatigue in Metals, 1959.
- [50] J. D. Eshelby. *Elastic inclusions and inhomogeneities*. Progress in solid mechanics, 1961.
- [51] P. P. Ewald. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik*, 369(3):253–287, Jan. 1921.
- [52] F. D. Fischer, T. WAITZ, D. Vollath, and N. K. Simha. On the role of surface energy and surface stress in phase-transforming nanoparticles. *Progress in Materials Science*, 53(3):481–527, Mar. 2008.
- [53] B. Flowers and E. Mendoza. *Properties of matter*. Manchester physics series. J. Wiley and Sons, 1970.
- [54] J. Frenzel, E. P. George, A. Dlouhy, C. Somsen, M. F. X. Wagner, and G. Eggeler. Influence of Ni on martensitic phase transformations in NiTi shape memory alloys. *Acta materialia*, 58(9):3444–3458, May 2010.

- [55] I. Galanakis, N. Papanikolaou, and P. H. Dederichs. Applicability of the broken-bond rule to the surface energy of the fcc metals. *Surface Science*, 511(1-3):1–12, 2002.
- [56] D. H. Gay and A. L. Rohl. MARVIN: a new computer code for studying surfaces and interfaces and its application to calculating the crystal morphologies of corundum and zircon. *Journal of the Chemical Society, Faraday Transactions*, 91(5):925, 1995.
- [57] F. Giustino. *Materials Modelling Using Density Functional Theory: Properties and Predictions*. Oxford University Press, 2014.
- [58] N. Goldenfeld. *Lectures on phase transitions and the renormalization group*. Frontiers in physics. Addison-Wesley, Advanced Book Program, 1992.
- [59] N. Goldenfeld, O. Martin, and Y. Oono. Intermediate asymptotics and renormalization group theory. *Journal of Scientific Computing*, 4(4):355–372, Dec. 1989.
- [60] M. Goodarzi and K. Hackl. Numerical simulation of interface effects in martensitic phase transformation. *PAMM*, 11(1):371–372, Dec. 2011.
- [61] M. Goodarzi and K. Hackl. Surface energies in microstructure of martensite. *PAMM*, 11(1):369–370, Dec. 2011.
- [62] M. Goodarzi and K. Hackl. A model for martensitic microstructure, its geometry and interface effects. *PAMM*, 12(1):345–346, Dec. 2012.
- [63] S. Govindjee, K. Hackl, and R. Heinen. An upper bound to the free energy of mixing by twin-compatible lamination for n-variant martensitic phase transformations. *Continuum Mechanics and Thermodynamics*, 18(7-8):443–453, Jan. 2007.
- [64] K. Hackl and R. Heinen. An upper bound to the free energy of n-variant polycrystalline shape memory alloys. *Journal of the Mechanics and Physics of Solids*, 56(9):2832–2843, Sept. 2008.
- [65] K. Hackl and P. Junker. A thermo-mechanically coupled field model for shape memory alloys. *Continuum Mechanics and Thermodynamics*, 26(6):859–877, Mar. 2014.
- [66] K. Hackl, M. Schmidt-Baldassari, W. Zhang, and G. Eggeler. Surface energies and size-effects in shape-memory-alloys. *Materials Science and Engineering: A*, 378(1-2):499–502, July 2004.
- [67] T. Hahn. *International Tables for Crystallography, Volume A: Space Group Symmetry*. International Tables for Crystallography. Springer Netherlands, 2005.
- [68] Z. Hashin and S. Shtrikman. Note on a variational approach to the theory of composite elastic materials. *Journal of the Franklin Institute*, 271(4):336–341, Apr. 1961.

- [69] Z. Hashin and S. Shtrikman. A variational approach to the theory of the elastic behaviour of polycrystals. *Journal of the Mechanics and Physics of Solids*, 10(4):343–352, Oct. 1962.
- [70] Z. Hashin and S. Shtrikman. On Some Variational Principles in Anisotropic and Nonhomogeneous Elasticity. *Journal of the Mechanics and Physics of Solids*, 10(4):335–342, 1962.
- [71] F. H. Herbstein. How precise are measurements of unit-cell dimensions from single crystals? *Acta Crystallographica Section B Structural Science*, 56(4):547–557, Aug. 2000.
- [72] K. Hermann. *Crystallography and Surface Structure: An Introduction for Surface Scientists and Nanoscientists*. Wiley, 2011.
- [73] R. Hill. Continuum micro-mechanics of elastoplastic polycrystals. *Journal of the Mechanics and Physics of Solids*, 13(2):89–101, Apr. 1965.
- [74] E. A. Holm, D. L. Olmsted, and S. M. Foiles. Comparing grain boundary energies in face-centered cubic metals: Al, Au, Cu and Ni. *Scripta Materialia*, 63(9):905–908, Nov. 2010.
- [75] G. Holzapfel. *Nonlinear Solid Mechanics: A Continuum Approach for Engineering*. Wiley, 2000.
- [76] B. Horovitz, G. R. Barsch, and J. A. Krumhansl. Twin bands in martensites: statics and dynamics. *Physical Review B*, 1991.
- [77] J. M. Howe. *Interfaces in materials: atomic structure, thermodynamics and kinetics of solid-vapor, solid-liquid and solid-solid interfaces*. A Wiley-Interscience publication. Wiley, 1997.
- [78] X. Huang, G. J. Ackland, and K. M. Rabe. Crystal structures and shape-memory behaviour of NiTi. *Nature Materials*, 2(5):307–311, Apr. 2003.
- [79] H. Huntington. *The elastic constants of crystals*. Number v. 7; v. 1958 in Solid state reprints. Academic Press, 1958.
- [80] D. Y. Huo and P. I. Müller. Nonequilibrium thermodynamics of pseudoelasticity. *Continuum Mechanics and Thermodynamics*, 5(3):163–204, 1993.
- [81] J. N. Israelachvili. *Intermolecular and Surface Forces: Revised Third Edition*. Intermolecular and Surface Forces. Elsevier Science, 2011.
- [82] R. D. James. Finite deformation by mechanical twinning. *Archive for Rational Mechanics and Analysis*, 77(2):143–176, 1981.

- [83] R. D. James, R. V. Kohn, and T. W. Shield. Modeling of Branched Needle Microstructures at the Edge of a Martensite Laminate. *Le Journal de Physique IV*, 05(C8):C8–253–C8–259, Dec. 1995.
- [84] R. D. James and Z. Zhang. A Way to Search for Multiferroic Materials with “Unlikely” Combinations of Physical Properties. In *Magnetism and Structure in Functional Materials*, pages 159–175. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.
- [85] S.-y. JIANG, Y.-q. ZHANG, Y.-n. ZHAO, S.-w. LIU, L. HU, and C.-z. ZHAO. Influence of Ni₄Ti₃ precipitates on phase transformation of NiTi shape memory alloy. *Transactions of Nonferrous Metals Society of China*, 25(12):4063–4071, Dec. 2015.
- [86] C. Johnson. *Numerical Solution of Partial Differential Equations by the Finite Element Method*. Dover Books on Mathematics Series. Dover Publications, Incorporated, 2012.
- [87] K. L. Johnson, K. Kendall, and A. D. Roberts. Surface Energy and the Contact of Elastic Solids. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 324(1558):301–313, Sept. 1971.
- [88] P. Junker. *Simulation of shape memory alloys : material modeling using the principle of maximum dissipation*. PhD thesis, Ruhr-Universität Bochum, 2011.
- [89] G. Kaptay. On the interfacial energy of coherent interfaces. *Acta materialia*, 60(19):6804–6813, Nov. 2012.
- [90] C. B. Ke, S. Cao, and X. P. Zhang. Phase field modeling of Ni-concentration distribution behavior around Ni₄Ti₃ precipitates in NiTi alloys. *Computational materials science*, 105(C):55–65, July 2015.
- [91] C.-b. KE, S.-s. CAO, X. MA, and X.-p. ZHANG. Modeling of Ni₄Ti₃ precipitation during stress-free and stress-assisted aging of bi-crystalline NiTi shape memory alloys. *Transactions of Nonferrous Metals Society of China*, 22(10):2578–2585, Oct. 2012.
- [92] A. Khachaturyan and G. Shatalov. Theory of macroscopic periodicity for a phase transition in the solid state. *Soviet Phys JETP*, 1969.
- [93] A. G. Khachaturyan. *Some questions concerning the theory of phase transformations in solids*. Soviet Phys. Solid State, 1967.
- [94] J. Khalil-Allafi, A. Dlouhy, and G. Eggeler. Ni₄Ti₃-precipitation during aging of NiTi shape memory alloys and its influence on martensitic phase transformations. *Acta materialia*, 50(17):4255–4274, Oct. 2002.
- [95] D. Kinderlehrer and G. Vergara-Caffarelli. The relaxation of functionals with surface energies. *Asymptotic Analysis*, 1989.

- [96] R. V. Kohn. The relaxation of a double-well energy. *Continuum Mechanics and Thermodynamics*, 3(3):193–236, 1991.
- [97] R. V. Kohn and S. Müller. Surface energy and microstructure in coherent phase transitions. *Communications on Pure and Applied Mathematics*, 47(4):405–435, 1994.
- [98] R. V. Kohn and S. Müller. Branching of twins near an austenite—twinned-martensite interface. *Philosophical Magazine A*, 66(5):697–715, Sept. 2006.
- [99] R. V. Kohn and F. Otto. Small surface energy, coarse-graining, and selection of microstructure. *Physica D: Nonlinear Phenomena*, 107(2-4):272–289, Sept. 1997.
- [100] R. V. Kohn and G. Strang. Structural design optimization, homogenization and relaxation of variational problems. In *Macroscopic properties of disordered media*, pages 131–147. Springer, 1982.
- [101] R. Kubo, H. Ichimura, T. Usui, and N. Hashitsume. *Thermodynamics: An Advanced Course with Problems and Solutions*. Ryōgo Kubo. North Holland, 1968.
- [102] N. D. Lang and W. Kohn. Theory of Metal Surfaces - Charge Density and Surface Energy. *Physical Review B*, 1(12):4555–&, 1970.
- [103] J. Li, S. J. Dillon, and G. S. Rohrer. Relative grain boundary area and energy distributions in nickel. *Acta materialia*, 57(14):4304–4311, Aug. 2009.
- [104] Y. Liu and Z. L. Xie. Twinning and detwinning of $\langle 0\ 1\ 1 \rangle$ type II twin in shape memory alloy. *Acta materialia*, 51(18):5529–5543, 2003.
- [105] Y. Liu and Z. L. Xie. The Rational Nature of Type II Twin in NiTi Shape Memory Alloy. *Journal of Intelligent Material Systems and Structures*, 17(12):1083–1090, Dec. 2006.
- [106] J. Marsden and T. Hughes. *Mathematical Foundations of Elasticity*. Dover Civil and Mechanical Engineering. Dover Publications, 2012.
- [107] J. M. McNaney, V. Imbeni, Y. Jung, P. Papadopoulos, and R. O. Ritchie. An experimental study of the superelastic effect in a shape-memory Nitinol alloy under biaxial loading. *Mechanics of Materials*, 35(10):969–986, Oct. 2003.
- [108] R. D. Mindlin. Force at a Point in the Interior of a Semi-Infinite Solid. *Physics*, 7(5):195, 1936.
- [109] C. Morrey. *Multiple Integrals in the Calculus of Variations*. Classics in Mathematics. Springer Berlin Heidelberg, 2009.
- [110] C. B. Morrey. Quasi-convexity and the lower semicontinuity of multiple integrals. *Pacific J Math*, 1952.

- [111] P. Muller. Elastic effects on surface physics. *Surface Science Reports*, 54(5-8):157–258, Aug. 2004.
- [112] S. Müller. Singular perturbations as a selection criterion for periodic minimizing sequences. *Calculus of Variations and Partial Differential Equations*, 1(2):169–204, 1993.
- [113] S. Müller. Variational models for microstructure and phase transitions. *Calculus of variations and geometric evolution ...*, 1999.
- [114] T. Mura. *Micromechanics of Defects in Solids*. Mechanics of Elastic and Inelastic Solids. Springer Netherlands, 1987.
- [115] V. D. Nguyen, E. Bechet, C. Geuzaine, and L. Noels. Imposing periodic boundary condition on arbitrary meshes by polynomial interpolation. *Computational materials science*, 55:390–406, Apr. 2012.
- [116] J. F. Nicholas. Calculation of Surface Energy as a Function of Orientation for Cubic Crystals. *Australian Journal of Physics*, 21(1):21–&, 1968.
- [117] J. Nocedal and S. Wright. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer New York, 2006.
- [118] R. Ogden. *Non-Linear Elastic Deformations*. Dover Civil and Mechanical Engineering. Dover Publications, 2013.
- [119] D. L. Olmsted, S. M. Foiles, and E. A. Holm. Survey of computed grain boundary properties in face-centered cubic metals: I. Grain boundary energy. *Acta materialia*, 57(13):3694–3703, Aug. 2009.
- [120] D. L. Olmsted, E. A. Holm, and S. M. Foiles. Survey of computed grain boundary properties in face-centered cubic metals—II: Grain boundary mobility. *Acta materialia*, 57(13):3704–3713, Aug. 2009.
- [121] K. Otsuka and X. Ren. Recent developments in the research of shape memory alloys. *Intermetallics*, 7(5):511–528, May 1999.
- [122] K. Otsuka and X. Ren. Physical metallurgy of Ti–Ni-based shape memory alloys. *Progress in Materials Science*, 50(5):511–678, July 2005.
- [123] K. Otsuka and C. M. Wayman, editors. *Shape Memory Materials*. Cambridge University Press, Oct. 1999.
- [124] D. Perić, E. A. de Souza Neto, R. A. Feijóo, M. Partovi, and A. J. C. Molina. On micro-to-macro transitions for multi-scale analysis of non-linear heterogeneous materials: unified variational basis and finite element implementation. *International Journal for Numerical Methods in Engineering*, 87(1-5):149–170, July 2011.

- [125] M. Pitteri. Reconciliation of local and global symmetries of crystals. *Journal of elasticity and the physical science of solids*, 1984.
- [126] M. Pitteri. On the Kinematics of Mechanical Twinning in Crystals. In *The Breadth and Depth of Continuum Mechanics*, pages 671–703. Springer Berlin Heidelberg, Berlin, Heidelberg, 1986.
- [127] M. Pitteri. On the kinematics of mechanical twinning in crystals. *The Breadth and Depth of Continuum Mechanics*, 1986.
- [128] M. Pitteri and G. Zanzotto. Beyond Space Groups: the Arithmetic Symmetry of Deformable Multilattices. *Acta Crystallographica Section A Foundations of Crystallography*, 54(4):359–373, July 1998.
- [129] M. Pitteri and G. Zanzotto. *Continuum Models for Phase Transitions and Twinning in Crystals*. Applied Mathematics. CRC Press, 2002.
- [130] D. A. Porter, K. E. Easterling, and M. Sherif. *Phase Transformations in Metals and Alloys, Third Edition (Revised Reprint)*. CRC Press, 2009.
- [131] D. J. Quesnel, D. S. Rimai, and L. P. DeMejo. Elastic compliances and stiffnesses of the fcc Lennard-Jones solid. *Physical Review B*, 48(10):6795–6807, Sept. 1993.
- [132] B. D. Reddy. *Introductory Functional Analysis: With Applications to Boundary Value Problems and Finite Elements*. Texts in Applied Mathematics. Springer New York, 2013.
- [133] X. Ren, N. Miura, K. Taniwaki, K. Otsuka, T. Suzuki, K. Tanaka, Y. I. Chumlyakov, and M. Asai. Understanding the martensitic transformations in TiNi-based alloys by elastic constants measurement. *Materials Science and Engineering: A*, 273-275:190–194, Dec. 1999.
- [134] X. Ren, N. Miura, J. Zhang, K. Otsuka, K. Tanaka, M. Koiwa, T. Suzuki, Y. I. Chumlyakov, and M. Asai. A comparative study of elastic constants of Ti–Ni-based alloys prior to martensitic transformation. *Materials Science and Engineering: A*, 312(1-2):196–206, Aug. 2001.
- [135] G. S. Rohrer. Grain boundary energy anisotropy: a review. *Journal of Materials Science*, 46(18):5881–5895, 2011.
- [136] A. L. Roĭtburd. *Domain structure of crystals formed in solid phase*. SOVIET PHYSICS SOLID STATE, 1969.
- [137] A. L. Roĭtburd. The theory of the formation of a heterophase structure in phase transformations in solids. *Soviet Physics Uspekhi*, 17(3):326–344, 1974.

- [138] D. Schryvers, W. Tirry, and Z. Q. Yang. Measuring strain fields and concentration gradients around Ni₄Ti₃ precipitates. *Materials Science and Engineering: A*, 438–440:485–488, Nov. 2006.
- [139] H. Sehitoglu, J. Jun, X. Zhang, I. Karaman, and Y. Chumlyakov. Shape memory and pseudoelastic behavior of 51.5% Ni–Ti single crystals in solutionized and overaged state. *Acta materialia*, 49(17):3609–3620, 2001.
- [140] P. Šesták, M. Černý, and J. Pokluda. Elastic Constants of Austenitic and Martensitic Phases of NiTi Shape Memory Alloy. In *Recent Advances in Mechatronics*, pages 1–6. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.
- [141] J. R. Solana. *Perturbation Theories for the Thermodynamic Properties of Fluids and Solids*. Taylor & Francis, 2013.
- [142] B. Sonderegger and E. Kozeschnik. Generalized Nearest-Neighbor Broken-Bond Analysis of Randomly Oriented Coherent Interfaces in Multicomponent Fcc and Bcc Structures. *Metallurgical and Materials Transactions a-Physical Metallurgy and Materials Science*, 40A(3):499–510, Mar. 2009.
- [143] J. Stoer, R. Bartels, W. Gautschi, R. Bulirsch, and C. Witzgall. *Introduction to Numerical Analysis*. Texts in Applied Mathematics. Springer New York, 2002.
- [144] S. Stupkiewicz, G. Maciejewski, and H. Petryk. Low-energy morphology of the interface layer between austenite and twinned martensite. *Acta materialia*, 55(18):6292–6306, Oct. 2007.
- [145] S. Stupkiewicz, G. Maciejewski, and H. Petryk. Elastic micro-strain energy of austenite–martensite interface in NiTi. *Modelling and Simulation in Materials Science and Engineering*, 20(3):035001, Apr. 2012.
- [146] S. Stupkiewicz and H. Petryk. Modelling of laminated microstructures in stress-induced martensitic transformations. *Journal of the Mechanics and Physics of Solids*, 2002.
- [147] P. W. Tasker. The surface energies, surface tensions and surface structure of the alkali halide crystals. *Philosophical Magazine A*, 39(2):119–136, Feb. 1979.
- [148] L. Vitos, A. V. Ruban, H. L. Skriver, and J. Kollar. The surface energy of metals. *Surface Science*, 411(1-2):186–202, 1998.
- [149] M. F. X. Wagner and W. Windl. Lattice stability, elastic constants and macroscopic moduli of NiTi martensites from first principles. *Acta materialia*, 56(20):6232–6245, Dec. 2008.
- [150] M. F. X. Wagner and W. Windl. Elastic anisotropy of Ni₄Ti₃ from first principles. *Scripta Materialia*, 60(4):207–210, Feb. 2009.

- [151] T. Waitz, D. Spišák, J. Hafner, and H. P. Karnthaler. Size-dependent martensitic transformation path causing atomic-scale twinning of nanocrystalline NiTi shape memory alloys. *EPL (Europhysics Letters)*, 71(1):98–103, July 2005.
- [152] K. P. Walker. Fourier Integral Representation of the Green Function for An Anisotropic Elastic Half-Space. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 443(1918):367–389, Nov. 1993.
- [153] S. G. Wang, E. K. Tian, and C. W. Lung. Surface energy of arbitrary crystal plane of bcc and fcc metals. *Journal of Physics and Chemistry of Solids*, 61(8):1295–1300, Aug. 2000.
- [154] Y.-N. Wen and J.-M. Zhang. Surface energy calculation of the fcc metals by using the MAEAM. *Solid State Communications*, 144(3-4):163–167, Oct. 2007.
- [155] P. Wriggers. *Nonlinear Finite Element Methods*. Springer Berlin Heidelberg, 2008.
- [156] G. Zanzotto. On the material symmetry group of elastic crystals and the Born Rule. *Archive for Rational Mechanics and Analysis*, 121(1):1–36, 1992.
- [157] J.-M. Zhang, F. Ma, and K.-W. Xu. Calculation of the surface energy of FCC metals with modified embedded-atom method. *Applied Surface Science*, 229(1-4):34–42, May 2004.
- [158] X. Zhang and H. Sehitoglu. Crystallography of the $B2 \rightarrow R \rightarrow B19'$ phase transformations in NiTi. *Materials Science and Engineering: A*, 374(1-2):292–302, June 2004.
- [159] S. Zhen and G. J. Davies. Calculation of the Lennard-Jones n - m potential energy parameters for metals. *physica status solidi (a)*, 78(2):595–605, Aug. 1983.
- [160] O. C. Zienkiewicz, R. L. Taylor, and J. Z. Zhu. *The Finite Element Method: Its Basis and Fundamentals*. Elsevier Science, 2013.
- [161] J. Zimmer. Jog my shape memory: dynamics as a challenge in mathematical materials science. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 364(1849):3285–3300, 2006.

LEBENS LAUF

PERSÖNLICHE ANGABEN

Name: Mehdi Goodarzi
Geburtsdatum: 16.09.1979
Geburtsort: Borojerd, Iran
Nationalität: iranisch

AUSBILDUNG

01.2009–12.2016 Promotion Maschinenbau, Ruhr-Universität Bochum
10.2006–11.2008 M.Sc. Computational Engineering, Ruhr-Universität Bochum
09.1999–02.2004 B.Sc. Maschinenbau, TU Sharif, Teheran, Iran
09.1997–06.1999 Vordiplom Architektur, Teheran Akademie der Künste, Iran
09.1993–06.1997 Gymnasium, Bahrol-Olum Gymnasium, Teheran, Iran

VERÖFFENTLICHUNGEN

- 2012** PROCEEDINGS *A model for martensitic microstructure, its geometry and interface effects*, Goodarzi M & Hackl K, PAMM, 12(1), 345-346, doi:10.1002/pamm.201210161
- 2011** PROCEEDINGS *Numerical simulation of interface effects in martensitic phase transformation*, Goodarzi M & Hackl K, PAMM, 11(1), 371-372. doi:10.1002/pamm.201110177
- 2010** PROCEEDINGS *Surface energies in microstructure of martensite*, Goodarzi M & Hackl K, PAMM, 11(1), 369-370. doi:10.1002/pamm.201110176
- 2010** GRADUATE LECTURE NOTES *A small compendium on vector and tensor algebra and calculus*, Hackl & Goodarzi, Ruhr University Bochum
- 2010** GRADUATE LECTURE NOTES *An introduction to linear continuum mechanics*, Hackl & Goodarzi, Ruhr University Bochum
- 2008** MASTER THESIS *Curve Modeling and Image Processing of A Convex Fluid Interface*, supervision by Prof. Franz Peters, Institute of Fluid Mechanics, Ruhr University Bochum
- 2004** BACHELOR THESIS *Turbulence modeling in open channel flow, a field study*, supervision by Prof. M. Rad, Sharif University of Technology, Iran

**Mitteilungen aus dem Institut für Mechanik
RUHR-UNIVERSITÄT BOCHUM
Nr. 173**

ISBN 978-3-935892-51-3