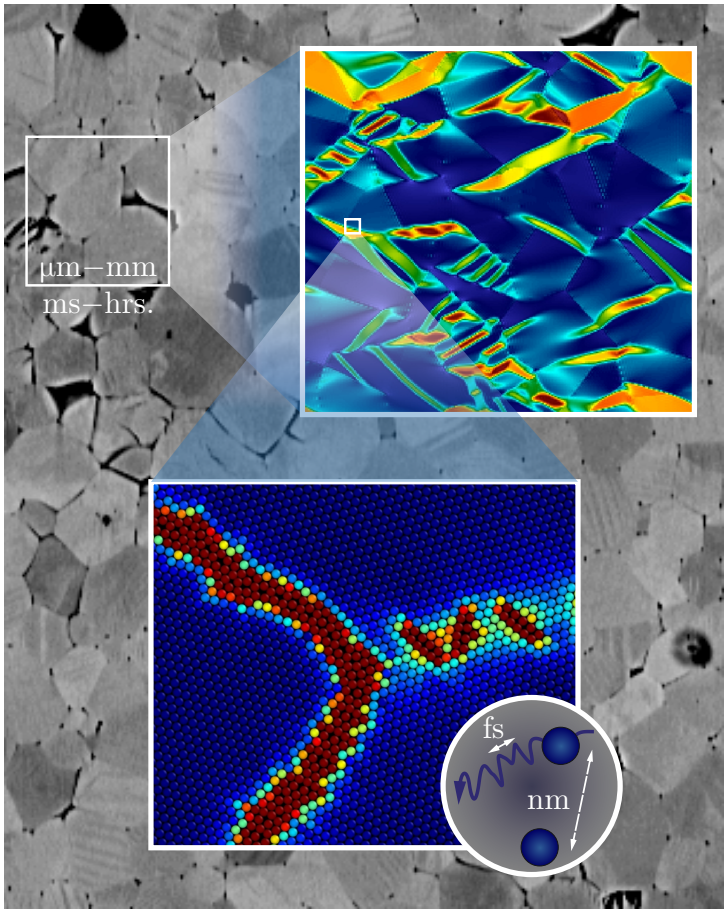
Multiscale Modeling

(Spring 2023)

Prof. Dr. Dennis M. Kochmann

Mechanics & Materials Lab
Institute of Mechanical Systems
Department of Mechanical and Process Engineering
ETH Zürich



Contents

0 Preface

Multiscale modeling is a hot topic across scientific and engineering disciplines. Roughly speaking, we need multiscale modeling techniques whenever a problem of interest can not be modeled on a single (time or length) scale. Consider, e.g., the engineering properties of classical materials, such as their stiffness, strength, toughness, damping, density, or heat capacity – to mention only a few. All such properties are *effective* properties – properties that we perceive on the macroscale with the naked eye but which are the collective manifestation of numerous features and mechanisms at significantly smaller scales: from the electronic and atomic configuration of the material to its mesoscale structure. Of course, we do have powerful computational modeling techniques at each of those scales (e.g., Density Functional Theory and Molecular Dynamics at the electronic and atomic scales, or phase field and finite element techniques at the mesoscales), yet those have a limited range of applicability. Multiscale techniques are required whenever a problem of interest bridges across those individual scales, e.g., because we are interested in the macroscale behavior of a material with complex microstructure, or because we would like to simulate atomistic systems beyond the limitations of microseconds and nanometers.

This course aims at giving an overview of multiscale modeling techniques. We start by considering the fundamental problem of extracting effective, macroscale properties from a known microstructure, which allows us to define the key concepts of *homogenization* and of a micro-to-macro transition in modeling solids. We stress that the focus of this course is on the numerical solution of multiscale problems (rather than deriving analytical expressions for those few cases, where homogenization theory admits closed-form solutions). To this end, we will combine our multiscale concepts with the finite element method as well as with spectral techniques. While the first half of the course focuses on *continuum scales*, the second half will survey all scales classically of interest to modelers in engineering: starting on the electric scale with a cartoon introduction to quantum mechanics and Density Functional Theory, climbing up in scales to Molecular Dynamics and interatomic potentials, and finally discussing avenues for upscaling atomistics to the continuum scale through concurrent and hierarchical multiscale techniques.

Of course, this course can only give an overview of the aforementioned methods, so you are encouraged to follow the recommended literature to find out more about each topic of interest. I would like to point out that, owing to my personal scientific background, the focus of many applications will be on composites and crystalline solids, while the discussed techniques are more widely applicable.

In the following, it is assumed that you have a solid *background* in continuum mechanics and in the finite element method. A brief review of continuum mechanics and an introduction to the notation used in this course is available in Appendix A, while Appendix B provides a link to finite element lecture notes, which you may want to consult as needed.

I apologize in advance for any *typos* that may have found their way into these lecture notes, and I would like to thank Roman Indergand for proofreading. This is a truly evolving set of notes, which aims to highlight the key concepts and is *not* meant as a complete textbook (in fact, it is far from a stand-alone textbook). Though I have made a great effort to ensure these notes are free of essential typos, I cannot rule out that some have remained. If you spot any mistakes, feel free to send me a highlighted PDF at the end of the semester, so I can make corrections for future years.

I hope you will find the course interesting and these notes supportive while studying.

Dennis M. Kochmann
Zürich, September 2023

Part I:
Multiscale Modeling at Continuum Scales

1 Microstructure & Unit Cells

1.1 Multiple Scales, Averages and Unit Cells

Strictly speaking, the term "*multiscale*" refers to the involvement of more than one length and/or time scale in a problem, so that the physics at play cannot be considered using a single set of governing equations and constitutive laws. For example, how do the engineering properties of stiffness, strength and toughness emerge from the atomic or molecular structure of a metal, ceramic or polymer? Concepts like stiffness, strength and toughness usually arise in the context of continuum mechanics, where we define continuous stress and strain distributions that are governed by continuum balance laws and respond to the load applied to a macroscopic sample. The atomic-level behavior of a material, by contrast, is governed by interactions between large numbers of discrete particles that move in a stochastic fashion and jointly are responsible for the material properties at larger scales. How does one connect such two distinct scales and the features and mechanisms found on both? It is questions like this that we aim to answer.

When speaking of *multiscale* problems in practice, in many cases we refer to the simple scenario of a *two-scale* problem, in which we deal with two separate (but somehow connected) scales: a large one and a small one. Using the language of the above example, the engineering device level defines what we will call the **macroscale**, whereas the atomic level in this example defines the **microscale**. Here and in the following, we will use the terms *macroscale* and *microscale* to denote the larger and smaller scale, respectively, irrespective of the particular length scales that are involved. For example, when discussing the properties of fiber-reinforced composite materials, both the macroscale (scale of an airplane hull) and the microscale (micrometer- to millimeter-scale of individual fibers) may be described by continuum mechanics and the finite element method (FEM). As a consequence, one may use similar theoretical and computational techniques on both scales in this particular two-scale problem. By contrast, when discussing the physics of grain boundaries in metals, we may need to treat the microscale (individual lattice defects) with atomistics, whereas the macroscale (networks of many grains) can be treated by continuum mechanics and/or FEM. What specific modeling tools to use on each scale and what questions to ask depends on the problem at hand – irrespective of which we will use the terms *macroscale* and *microscale* in the following.

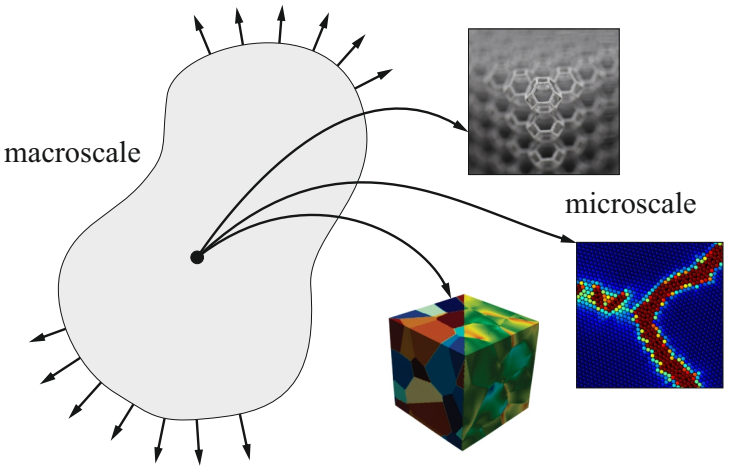


Figure 1.1: Illustration of the separation between macroscale and microscale (shown are three different examples of microscales in truss metamaterials, atomistic crystals, and polycrystalline metals, modeled by, respectively, structural models, atomistics, and FEM).

In many practical problems, e.g., we are interested in the the macroscale *response* of a body and how it depends on *features* on the microscale of the material that the body is made of. Ultimately, we often want to tackle the *inverse problem*: to achieve certain macroscale material behavior, how do we need to design the microscale architecture?

Although multiscale modeling plays a crucial role in bridging across both length and time scales, we will – within the scope of this course – focus primarily on length scales, going from the sub-atomic level of electronic structure calculations to the atomic level of molecular dynamics to discrete and continuum mesoscale models to phase field and macroscale continuum mechanics. The associated challenge of bridging across time scales – extracting the long-time behavior that emerges from many fast events – is usually connected quite naturally to the length scale problem: features on smaller scales tend to evolve much more rapidly than those on larger scales (compare, e.g., the atomic-level motion of atoms and molecules to the resulting effective engineering material behavior). We will touch upon *temporal upscaling* in the context of atomistics and coarse-grained atomistics. Until then, we will focus on *spatial upscaling*.

A key assumption in most spatial multiscale models is that a material is **statistically homogeneous**, i.e., the statistical properties of the microstructure do not change from point to point of the material at the macroscale. Statistical properties include, e.g., volume averages but also higher-order statistical data such as two-point correlation or n-point correlation data. For example, if particles in a composite are randomly distributed, we consider this material to be statistically homogeneous. If a material is not statistically homogeneous (e.g., when the particle concentration in a composite shows a gradient, as in Figure 1.2), it is called **statistically inhomogeneous**. As an important consequence, we may assume that the macroscale properties of a material with a statistically homogeneous microstructure are independent of where on the macroscale we probe the material properties.

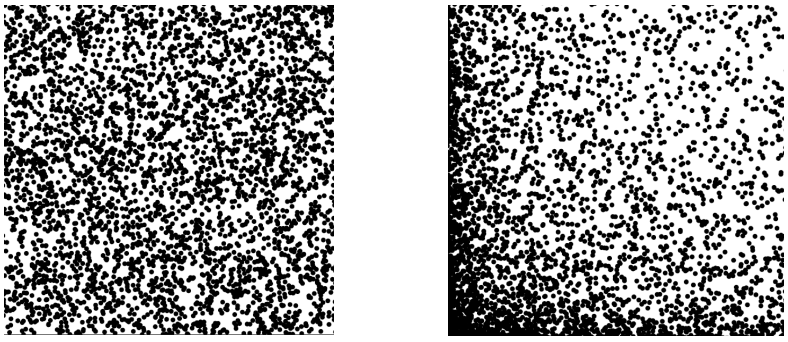


Figure 1.2: Statistically homogeneous vs. statistically inhomogeneous distributions of particles in a matrix (resulting in statistically homogeneous vs. inhomogeneous effective properties) [93].

When discussing microstructures, we often consider some notion of a **unit cell** (UC), which is in some sense representative of the overall microstructure. Without going into detail here (we will discuss the selection of a UC later), we will need average quantities, computed by averaging over the volume $V = |\Omega|$ of a unit cell Ω. Thus, we define for conciseness

$$\boxed{\langle\cdot\rangle_\Omega = \frac{1}{V}\int_\Omega (\cdot)\,\mathrm{d}V} \tag{1.1}$$

as the volume average of any quantity $(\cdot)$. For example, in a two-phase composite let $\chi(\boldsymbol{X})$ denote an **indicator function**, which is defined by

$$\chi(\boldsymbol{X}) = \begin{cases} 1, & \text{if } \boldsymbol{X} \text{ belongs to phase 1,} \\ 0, & \text{if } \boldsymbol{X} \text{ belongs to phase 2} \end{cases} \tag{1.2}$$

for any point $\boldsymbol{X} \in \Omega$. This allows us to compute the **volume fraction** v_i of either phase as

$$v_1 = \langle \chi \rangle_\Omega, \qquad v_2 = \langle (1-\chi) \rangle_\Omega = 1 - \langle \chi \rangle_\Omega = 1 - v_1. \tag{1.3}$$

For mechanical properties of interest, the unit cell generally undergoes some deformation, so it is important to point out that we define the average in (1.1) with respect to the *undeformed* configuration; i.e., we integrate over the undeformed unit cell Ω and not over the deformed unit cell $\varphi(\Omega)$ (where φ denotes the deformation mapping that links undeformed points $\boldsymbol{X} \in \Omega$ to deformed points $\boldsymbol{x} \in \varphi(\Omega)$ via $\boldsymbol{x} = \boldsymbol{\varphi}(\boldsymbol{X}, t)$ at time t). Since volume changes with deformation, those averages over the undeformed and deformed configuration are not necessarily the same:

$$\langle \cdot \rangle_{\varphi(\Omega)} = \frac{1}{v} \int_{\varphi(\Omega)} (\cdot)\, \mathrm{d}v = \frac{\displaystyle\int_{\varphi(\Omega)} (\cdot)\, \mathrm{d}v}{\displaystyle\int_{\varphi(\Omega)} \mathrm{d}v} = \frac{\displaystyle\int_\Omega (\cdot) \circ \boldsymbol{\varphi}\, J\, \mathrm{d}V}{\displaystyle\int_\Omega J\, \mathrm{d}V} = \frac{\langle \cdot \rangle_\Omega}{\langle J \rangle_\Omega} \neq \langle \cdot \rangle_\Omega, \tag{1.4}$$

with $v = |\varphi(\Omega)|$ and $J = \det \boldsymbol{F}$ being the local *Jacobian* (i.e., the determinant of the deformation gradient $\boldsymbol{F} = \operatorname{Grad} \boldsymbol{\varphi}$ which represents the local volume change, since $\mathrm{d}v = J\, \mathrm{d}V$). Unless the volume does not change (i.e., unless $J = 1$), (1.4) is an inequality, so averages with respect to undeformed and deformed unit cells do not agree. In the following, we will always *average with respect to the undeformed configuration* and therefore often omit the subscript $_\Omega$ (tacitly assuming averages over the undeformed configuration). Note that in linearized kinematics ("*small strains*") there is no difference anyways and the two averages coincide.

Microstructures often display either **randomness** or **periodicity**, or combinations thereof[1].

1.2 Periodic Microstructures

A *periodic* microstructure admits the identification of a **unit cell** as the simplest (smallest) repeating substructure, such that the complete microstructure results from periodic repetition of the unit cell in all dimensions. The size of the unit cell matches the length scale of periodicity. For periodic systems, the choice of the unit cell is never unique; i.e., its size is unique, but the exact location within the microstructure is not, since the system displays **translational invariance**. Examples of periodic microstructures are atoms stacked perfectly on a lattice, or particles placed periodically in a matrix to form an acoustic metamaterial.

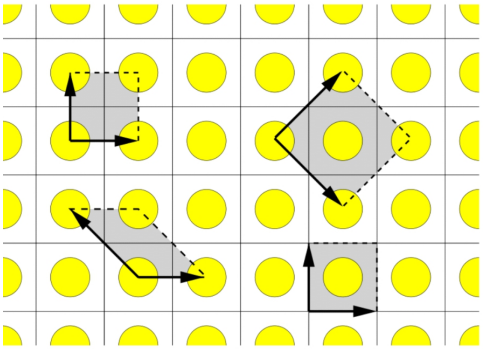


Figure 1.3: Periodic microstructure (periodic arrangement of inclusions or holes in a matrix, which leads to a non-unique unit cell definition – e.g., the four shown gray domains are all valid unit cells by translational invariance (arrows show the translation vectors of each unit cell) [93].

[1] Combinations include so-called *quasirandom* microstructures; see, e.g., low-discrepancy sequences.

1.3 Random Microstructures

By contrast, if we generate *random* microstructures, then each such random microstructure creation results in a different **configuration**, also called **realization**. Consider, e.g., throwing particles or fibers randomly into a matrix, or consider an n-phase high-entropy alloy, in which atoms of n different species are arranged randomly on a regular atomic lattice. Each random realization differs from the others. A set of multiple realizations is called an **ensemble** of realizations (see Figure 1.4).

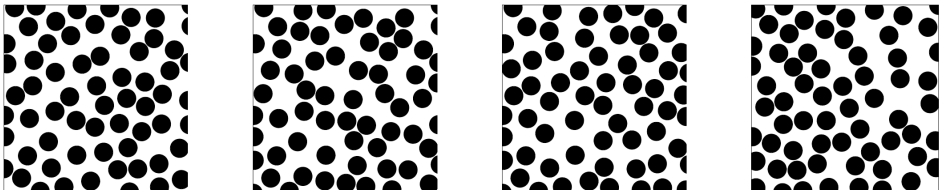


Figure 1.4: Four different *realizations* obtained from randomly placing a total of 64 particles of constant radius within a square unit cell. All realizations together form an *ensemble*. [93]

Assume that each configuration produces a **response**, which generally differs between realizations. Consider, e.g., the effective stiffness of a fiber-matrix composite, in which the fibers are randomly thrown into the matrix. If in a realization most fibers point in the same direction, the composite is clearly anisotropic (stiff in the fiber direction, soft in the transverse directions). If all fibers are truly randomly aligned, the elastic response is rather isotropic.

Let us define the **ensemble average** of a response $(\cdot)$ over a set of N realizations as

$$\boxed{\langle\langle\cdot\rangle\rangle_N = \frac{1}{N}\sum_{i=1}^{N}(\cdot)_i} \tag{1.5}$$

where we denote by $(\cdot)_i$ the response of the ith realization. For a large number N of realizations, the **central limit theorem** (CLT) from probability theory tells us that, for a random unit cell generation process, the response of the N random realizations follows approximately a Gaussian distribution. Moreover – and this is an essential finding – the sequence $\langle\langle\cdot\rangle\rangle_N$ converges with

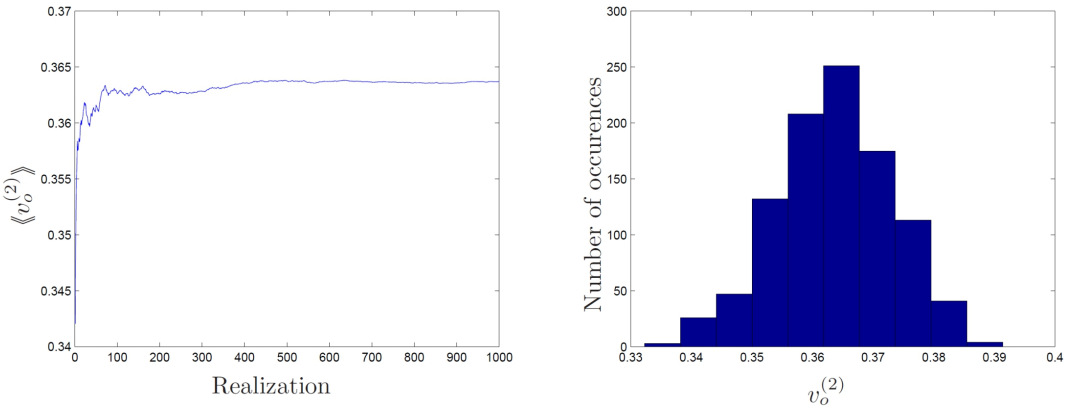


Figure 1.5: Statistics of an ensemble: average volume fraction of particles in a UC as obtained from an increasing number of realizations along with the histogram of the final distribution of 1000 realizations (64 particles of constant radius are placed in a square UC; the volume fraction varies because of particles near the boundaries) [93].

increasing N to a certain value:

$$\lim_{N\to\infty} \langle\langle\cdot\rangle\rangle_N = \langle\langle\cdot\rangle\rangle_\infty. \tag{1.6}$$

In simple terms, averaging over more and more realizations is expected to converge to a fixed value (and the distribution converges to a Gaussian distribution). This is the basis for **ensemble enlargement**, the procedure of increasing N at fixed unit cell size and feature size to obtain a converged average response (see Figure 1.5).

An alternative strategy to reaching a statistical limit is by **sample enlargement**, i.e., by increasing the size of the unit cell while keeping all microstructural feature sizes constant (e.g., increasing the size of the unit cell of a particle-matrix composite, while keeping the particle size and volume fraction constant; see Figure 1.6). With enlargement, the response of the unit cell typically also exhibits convergence. Let us denote by L the unit cell size and by l a characteristic microstructural size (e.g., the particle diameter). This lets us define another limit, viz.

$$\boxed{\langle\cdot\rangle_\infty = \lim_{L/l\to\infty} \langle\cdot\rangle_{\Omega(L)}} \tag{1.7}$$

where $\Omega(L)$ stands for the unit cell of size L. Note that there is no reason for the unit cell to have a specific shape (e.g., a cube). It can, in principle, have any shape, and the shape loses significance as the sample size increases (in the limit $L/l \to \infty$ it does not matter if one averages over a cubic, spherical, or tetrahedral unit cell).

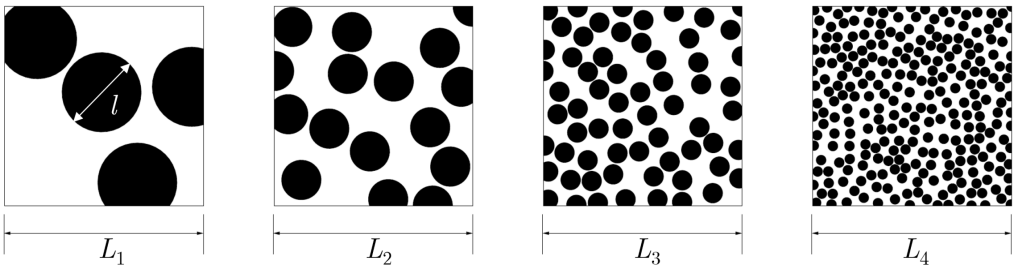


Figure 1.6: *Sample enlargement*: the unit cell size L_i is enlarged ($L_1 < L_2 < L_3 < L_4$) while the size of each particle remains constant ($l = \text{const.}$); adapted from [93].

Now, we have two approaches to reach a converged response from a random unit cell: by ensemble enlargement and by sample enlargement. In general, the convergence by sample enlargement vs. ensemble averaging is not equally uniform; ensemble averages typically lead to smoother convergence. However, for an infinitely large sample the response should agree with the response averaged over many realizations, so

$$\boxed{\langle\cdot\rangle_\infty = \langle\langle\,\langle\cdot\rangle_\infty\,\rangle\rangle_\infty} \tag{1.8}$$

This in turn implies that, if a unit cell is sufficiently large (ideally infinitely large), then a single simulation can provide the sought response, whereas smaller unit cells do show a spread in their responses, so ensemble averaging is required to obtain a converged response (see Figure 1.7).

Of course, based on the above we would ideally want to use an infinitely large sample in our simulations. Such a sample is called **statistically representative** of the statistically homogeneous microstructure, and a single analysis would be sufficient with such a sample to reveal the effective behavior of the microstructure. Such a sample or unit cell is oftentimes called a **representative volume element** (RVE), or representative area element (RAE) in 2D. As a sloppy convention, we usually use the term RVE in any dimension.

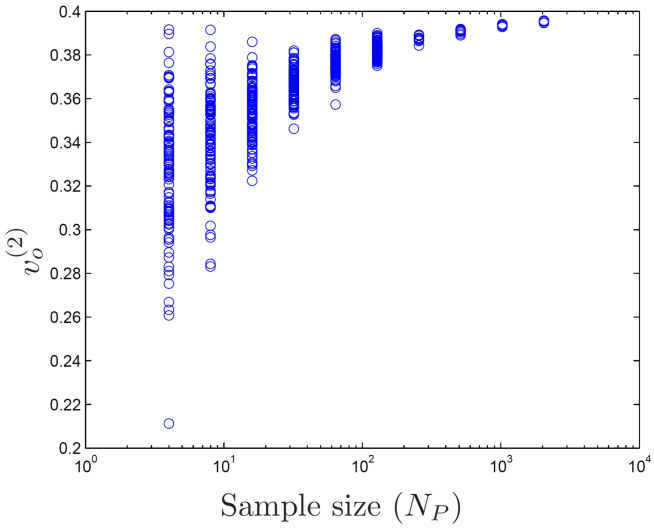


Figure 1.7: Distribution of the volume fraction of circular particles in a square unit cell for different sample sizes N_P, showing convergence with sample enlargement ($N_P \to \infty$) and a Gaussian distribution for each sample size N_P (same details as in Figure 1.5; each data point is the response of a single realization) [93].

For all practical purposes, we cannot work with infinite RVE sizes but will rather choose a *finite size* of the simulation domain, whose response is as close to the infinite-size limit as needed. Such an RVE should be as large as necessary yet as small as possible. One can identify the required number of realizations N or the size L of an RVE, e.g., by checking for a specific quantity if, respectively,

$$\frac{\langle\langle\cdot\rangle\rangle_{N+1} - \langle\langle\cdot\rangle\rangle_N}{\langle\langle\cdot\rangle\rangle_N} \leq \epsilon \qquad \text{or} \qquad \frac{\langle\cdot\rangle_{\Omega(L_{i+1})} - \langle\cdot\rangle_{\Omega(L_i)}}{\langle\cdot\rangle_{\Omega(L_i)}} \leq \epsilon, \tag{1.9}$$

i.e., if the change in the response with increasing ensemble or sample size reaches a tolerance level $\epsilon > 0$. (We here assume that L_i represents a sequence of increasing sample sizes with $\lim_{i\to\infty} L_i = \infty$.) This procedure leads to a sample size that is sufficiently large for accuracy, so that a single calculation based on the identified RVE is sufficient to extract effective material behavior. Of course, any such sample is only an *approximation* and, strictly speaking, is *not* a statistically representative volume element. Yet, it is sufficient for practical purposes and we will generally refer to such a unit cell as our **RVE**.

If the microstructure is *periodic*, then a single unit cell is sufficient in most cases to serve as RVE (unless in case of, e.g., bifurcation or instabilities as well as for vibrations, in which the deformation could show periodicity at a larger scale, e.g., by choosing an eigenmode that is periodic across $n \times n$ unit cells but not at the level of a single unit cell; in such cases the RVE may need to include $n \times n$ of the periodic unit cells). Also, for periodic microstructures the choice of the RVE is not unique.

In all of the following, we will assume that *a suitable RVE for computational purposes has been identified and is used for subsequent calculations and analysis.* For mechanical problems, we usually aim to identify the macroscopic constitutive behavior, which results from mechanisms on the microscale (e.g., finding the macroscale stress-strain response of a composite material, polycrystal, or complex polymer network). The key fields of interest here are the stresses and strains as well as the strain energy density (if available) and the stiffness tensor – and it will be our next task to understand how macroscale quantities (such as macroscopic stress and strain tensor) are linked to the nonuniform stress and strain distributions across an RVE on the microscale.

2 Averaging Theorems

When talking about the *response* of an RVE in mechanical problems, we are usually interested in the stress and strain of an RVE to be used, e.g., for constitutive modeling at larger scales. However, there is no such thing as "the stress" or "the strain" of an RVE, since every point inside the RVE, in principle, experiences different stresses and strains, based on the local deformation and based on the microstructure. It is generally accepted that we identify as "the stress" and "the strain" of an RVE the *volume averages* of, respectively, stress and strain across the RVE (and the same applies to, e.g., the stiffness and strain energy density of an RVE). We will hence need to compute average mechanical quantities from RVEs, for which the following so-called **averaging theorems** will be helpful. Because we have different formulations at small and large strains, we will consider both scenarios.

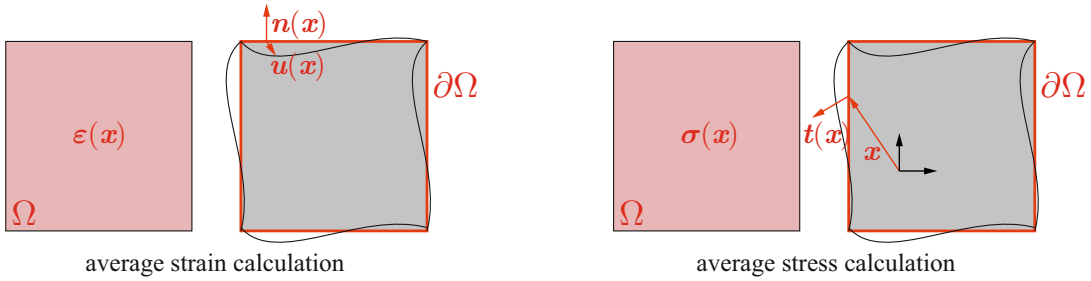


Figure 2.1: The averaging theorems translate volume (or area) integrals into surface (or boundary) integrals over an RVE, as we derive for both strains and stresses – here shown for linearized kinematics.

2.1 Averaging Theorems in Linearized Kinematics

Consider an RVE Ω with volume $V = |\Omega|$. If points $\boldsymbol{x} \in \Omega$ within the RVE are characterized by a displacement field $\boldsymbol{u}(\boldsymbol{x})$, then the infinitesimal strain tensor follows as $\boldsymbol{\varepsilon} = \frac{1}{2}\left[\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^{\mathrm{T}}\right]$ or, in components[2], $\varepsilon_{ij} = \frac{1}{2}(u_{i,j} + u_{j,i})$.

To compute the volume-averaged strain tensor across the RVE, we can use the divergence theorem[3] to arrive at

$$\langle\varepsilon_{ij}\rangle = \frac{1}{V}\int_\Omega \varepsilon_{ij}\,\mathrm{d}V = \frac{1}{2V}\int_\Omega (u_{i,j} + u_{j,i})\,\mathrm{d}V = \frac{1}{2V}\int_{\partial\Omega} (u_i n_j + u_j n_i)\,\mathrm{d}S, \tag{2.1}$$

which is equivalent to stating symbolically

$$\boxed{\langle\boldsymbol{\varepsilon}\rangle = \frac{1}{V}\int_{\partial\Omega} \mathrm{sym}\,(\boldsymbol{u}\otimes\boldsymbol{n})\,\mathrm{d}S} \tag{2.2}$$

Note that in case of discontinuous displacement fields, the boundary integral over $\partial\Omega$ must account for displacement jumps across any existing interfaces:

$$\langle\boldsymbol{\varepsilon}\rangle = \frac{1}{V}\int_{\partial\Omega} \mathrm{sym}\,(\boldsymbol{u}\otimes\boldsymbol{n})\,\mathrm{d}S + \frac{1}{V}\int_C \mathrm{sym}\,([[\boldsymbol{u}]]\otimes\boldsymbol{n})\,\mathrm{d}S, \tag{2.3}$$

where C is an interior discontinuous interface, $[[\cdot]] = (\cdot)^+ - (\cdot)^-$ denotes the jump across C, and $\boldsymbol{n}$ is a unit normal vector chosen such that it points from the $+$ to the $-$ side of the interface. If

[2] Here and in the following we use classic index notation for tensors, the comma denoting partial derivatives with respect to the spatial coordinates; i.e., $(\cdot)_{,i} \equiv \partial(\cdot)/\partial x_i$

[3] The divergence theorem states that $\int_\Omega (\cdot)_{,i}\,\mathrm{d}V = \int_{\partial\Omega}(\cdot)n_i\,\mathrm{d}S$ with $\boldsymbol{n}$ the unit outward vector on the surface, thus turning the volume integral into a surface integral.

no such discontinuities exist, we say the composite phases are **perfectly bonded**. The latter also implies that tractions $\boldsymbol{t} = \boldsymbol{\sigma n}$ (derived from the infinitesimal stress tensor $\boldsymbol{\sigma}$ and unit normal vector $\boldsymbol{n}$) are continuous across interfaces.

Analogously, the average stress tensor is derived by exploiting the relation (with $x_{j,k} = \partial x_j/\partial x_k = \delta_{jk}$, where the latter represents Kronecker's delta)

$$\sigma_{ij} = \sigma_{ik}\delta_{jk} = \sigma_{ik}x_{j,k}. \tag{2.4}$$

The local condition of linear momentum balance implies that $\operatorname{div}\boldsymbol{\sigma} + \rho\boldsymbol{b} = \rho\boldsymbol{a}$, where ρ is the mass density, $\rho\boldsymbol{b}$ are body forces, and $\boldsymbol{a} = \ddot{\boldsymbol{u}}$ is the acceleration vector. Using index notation, linear momentum balance reads $\sigma_{ij,j} + \rho b_i = \rho a_i$. Further noting that tractions are $\boldsymbol{t} = \boldsymbol{\sigma n}$ or $t_i = \sigma_{ik}n_k$, we compute the average stress tensor components as

$$\begin{aligned}\langle\sigma_{ij}\rangle &= \frac{1}{V}\int_\Omega \sigma_{ij}\,\mathrm{d}V = \frac{1}{V}\int_\Omega \sigma_{ik}x_{j,k}\,\mathrm{d}V = \frac{1}{V}\left[\int_{\partial\Omega}\sigma_{ik}x_j n_k\,\mathrm{d}S - \int_\Omega \sigma_{ik,k}x_j\,\mathrm{d}V\right]\\ &= \frac{1}{V}\left[\int_{\partial\Omega} t_i x_j\,\mathrm{d}S - \int_\Omega \rho(a_i - b_i)x_j\,\mathrm{d}V\right].\end{aligned} \tag{2.5}$$

This is equivalent to saying in symbolic form

$$\langle\boldsymbol{\sigma}\rangle = \frac{1}{V}\left[\int_{\partial\Omega}\boldsymbol{t}\otimes\boldsymbol{x}\,\mathrm{d}S - \int_\Omega \rho(\boldsymbol{a}-\boldsymbol{b})\otimes\boldsymbol{x}\,\mathrm{d}V\right]. \tag{2.6}$$

In case of *quasistatic processes* and *in the absence of body forces* (which is what is being considered in the majority of homogenization problems) we thus arrive at

$$\boxed{\langle\boldsymbol{\sigma}\rangle = \frac{1}{V}\int_{\partial\Omega}\boldsymbol{t}\otimes\boldsymbol{x}\,\mathrm{d}S = \frac{1}{V}\int_{\partial\Omega}\boldsymbol{x}\otimes\boldsymbol{t}\,\mathrm{d}S} \tag{2.7}$$

The last step exploited that that $\boldsymbol{\sigma} = \boldsymbol{\sigma}^{\mathrm{T}}$ is by definition symmetric. Also, note that $[[\boldsymbol{t}]] = \boldsymbol{0}$ at internal interfaces in equilibrium. Therefore, we here do not need to account for internal discontinuities as for the strains (in multi-phase materials like composites, the strains may jump across material interfaces, whereas the stresses must be continuous by equilibrium).

Relations (2.2) and (2.7) come handy, because they allow us to calculate volume averages merely from information on the surface. Aside from being helpful when defining RVE boundary conditions later, they are also of practical importance as they provide an efficient shortcut to computing volume averages (for very large problems, numerically integrating over the surface is significantly less expensive than integrating over the volume).

2.2 Averaging Theorems in Finite Kinematics

We can repeat the above derivations for a finite-kinematics setting, in which the deformation is characterized by the deformation mapping $\boldsymbol{x} = \boldsymbol{\varphi}(\boldsymbol{X})$, and the deformation gradient tensor is $\boldsymbol{F} = \operatorname{Grad}\boldsymbol{\varphi} = \partial\boldsymbol{x}/\partial\boldsymbol{X} = \boldsymbol{I} + \nabla\boldsymbol{u}$. Again invoking the divergence theorem, the average deformation gradient tensor for an RVE Ω is obtained as

$$\langle F_{iJ}\rangle = \frac{1}{V}\int_\Omega \varphi_{i,J}\,\mathrm{d}V = \frac{1}{V}\int_{\partial\Omega}\varphi_i N_J\,\mathrm{d}S \tag{2.8}$$

or

$$\boxed{\langle\boldsymbol{F}\rangle = \frac{1}{V}\int_{\partial\Omega}\boldsymbol{\varphi}\otimes\boldsymbol{N}\,\mathrm{d}S} \tag{2.9}$$

where $\boldsymbol{N}$ represents the undeformed surface unit normal.

Analogous to the linearized setting, we need to compute average stresses. In finite strains, we usually deal with the first Piola Kirchhoff stress tensor $\boldsymbol{P}$ (defined as actual force per undeformed area). We again exploit the relation

$$P_{iJ} = P_{iK}\delta_{JK} = P_{iK}X_{J,K}. \tag{2.10}$$

Linear momentum balance in this setting reads $\operatorname{Div}\boldsymbol{P} + R\boldsymbol{B} = R\boldsymbol{A}$, where R is the mass density per undeformed volume, $R\boldsymbol{B}$ is the body force and $\boldsymbol{A}$ the acceleration vector, both in the undeformed reference configuration. With tractions $\boldsymbol{T} = \boldsymbol{P}\boldsymbol{N}$ or $T_i = P_{iK}N_K$, this leads to the average first Piola Kirchhoff stress components

$$\begin{aligned}\langle P_{iJ}\rangle &= \frac{1}{V}\int_\Omega P_{iK}X_{J,K}\,\mathrm{d}V = \frac{1}{V}\left[\int_{\partial\Omega} P_{iK}X_J N_K\,\mathrm{d}S - \int_\Omega P_{iK,K}X_J\,\mathrm{d}V\right]\\ &= \frac{1}{V}\left[\int_{\partial\Omega} T_i X_J\,\mathrm{d}S - \int_\Omega R(A_i - B_i)X_J\,\mathrm{d}V\right].\end{aligned} \tag{2.11}$$

Using symbolic notation, the above is equivalent to

$$\langle\boldsymbol{P}\rangle = \frac{1}{V}\left[\int_{\partial\Omega}\boldsymbol{T}\otimes\boldsymbol{X}\,\mathrm{d}S - \int_\Omega R(\boldsymbol{A}-\boldsymbol{B})\otimes\boldsymbol{X}\,\mathrm{d}V\right]. \tag{2.12}$$

Under *quasistatic* conditions ($\boldsymbol{A}=\boldsymbol{0}$) and in the *absence of body forces* ($\boldsymbol{B}=\boldsymbol{0}$), we thus have

$$\boxed{\langle\boldsymbol{P}\rangle = \frac{1}{V}\int_{\partial\Omega}\boldsymbol{T}\otimes\boldsymbol{X}\,\mathrm{d}S} \tag{2.13}$$

where $\boldsymbol{X}$ represents the undeformed position of points on the surface of the RVE. Note that $\boldsymbol{P}\neq\boldsymbol{P}^{\mathrm{T}}$ (equality cannot hold for a two-point tensor anyways). Like in linearized kinematics, the averaging relations (2.9) and (2.13) provide convenient shortcuts to replace volume integrals by surface integrals.

3 Numerical Evaluation of RVE Averages

Aside from few analytically tractable problems, most multiscale simulations are to be conducted numerically. At the continuum scale this happens most frequently by the **finite element method** (FEM). When the averaging theorems derived in Section 2 are used within the FEM context, it is convenient to find their approximate analogues within the discrete finite element (FE) representation.

3.1 Average Deformation Gradient

Let us begin with the average deformation gradient in an RVE. When the continuous deformation mapping $\boldsymbol{x} = \boldsymbol{\varphi}(\boldsymbol{X})$ is approximated by an interpolation on an FE mesh, we obtain the interpolated mapping $\boldsymbol{\varphi}_e^h$ within an element e as

$$\boldsymbol{\varphi}_e^h(\boldsymbol{X}) = \sum_{a=1}^{n} \boldsymbol{\varphi}_e^a N_e^a(\boldsymbol{X}), \tag{3.1}$$

where we assumed n nodes per element, which carry the degrees of freedom $\boldsymbol{\varphi}_e^a$, while $N^a(\boldsymbol{X})$ denotes the shape function associated with node a ($a = 1, \dots, n$). The approximate deformation mapping for the entire mesh is hence the sum over all elements:

$$\boldsymbol{\varphi}^h(\boldsymbol{X}) = \sum_e \sum_{a=1}^{n} \boldsymbol{\varphi}_e^a N_e^a(\boldsymbol{X}). \tag{3.2}$$

The FE-approximated deformation gradient (which is generally discontinuous across element boundaries, e.g., in standard polynomial elements) follows as

$$\boldsymbol{F}^h = \operatorname{Grad} \boldsymbol{\varphi}^h = \sum_e \sum_{a=1}^{n} \boldsymbol{\varphi}_e^a \otimes \nabla N^a(\boldsymbol{X}). \tag{3.3}$$

Instead of averaging this approximate deformation gradient over all elements in the mesh, we refer to the averaging relation derived above:

$$\langle \boldsymbol{F} \rangle = \frac{1}{V} \int_{\partial\Omega} \boldsymbol{\varphi} \otimes \boldsymbol{N} \, \mathrm{d}S. \tag{3.4}$$

The approximate numerical counterpart reads

$$\begin{aligned} \langle \boldsymbol{F}^h \rangle &= \frac{1}{V} \sum_e \int_{\partial\Omega_e'} \sum_{a=1}^{n} \boldsymbol{\varphi}_e^a N_e^a(\boldsymbol{X}) \otimes \boldsymbol{N} \, \mathrm{d}S \\ &= \frac{1}{V} \sum_e \sum_{a=1}^{n} \boldsymbol{\varphi}_e^a \int_{\partial\Omega_e'} N_e^a(\boldsymbol{X}) \otimes \boldsymbol{N} \, \mathrm{d}S, \end{aligned} \tag{3.5}$$

where $\partial\Omega_e' \subset \partial\Omega$ denotes the external part of the boundary associated with element e. Consider, e.g., elements that interpolate linearly on the boundary (e.g., simplicial elements or (bi/tri)linear elements) and an RVE geometry with planar faces. In this case $\boldsymbol{N} = \boldsymbol{N}_e = \text{const.}$ across elements on the same face, so that

$$\langle \boldsymbol{F}^h \rangle = \frac{1}{V} \sum_e \sum_{a=1}^{n} \boldsymbol{\varphi}_e^a \otimes \boldsymbol{N}_e \int_{\partial\Omega_e'} N_e^a(\boldsymbol{X}) \, \mathrm{d}S = \frac{1}{V} \sum_{a=1}^{n} \boldsymbol{\varphi}^a \otimes \frac{1}{d} \sum_e \boldsymbol{N}_e^a A_e, \tag{3.6}$$

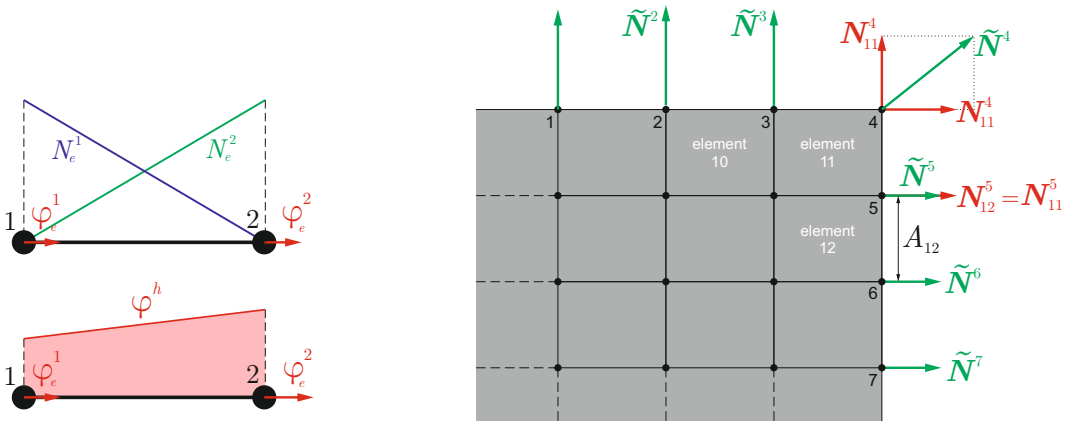


Figure 3.1: FE interpolation of the deformation mapping over an element boundary (here shown for linear shape functions) along with a schematic of the nodal normal vector conventions and definitions on the boundary.

where $\boldsymbol{N}_e^a$ denotes the outward unit normal at node a as seen from element e (for corner or edge nodes, two or three such normals exist per node, respectively, and all must be accounted for within all adjacent elements). The factor $1/d$ stems from the integration of the shape function on the boundary (for simplicial elements, d is the number of dimensions). A_e denotes the element external boundary area (in 3D) or external edge length (in 2D). Let us introduce an effective nodal normal (see Figure 3.1)

$$\tilde{\boldsymbol{N}}^a = \frac{1}{d}\sum_e \boldsymbol{N}_e^a A_e, \tag{3.7}$$

which is obtained for each node a by averaging over the normals of adjacent elements e, weighted by their respective surface areas A_e. Finally writing $\boldsymbol{x}^a = \boldsymbol{\varphi}^a$ leads to

$$\boxed{\langle \boldsymbol{F}^h \rangle = \frac{1}{V}\sum_{a=1}^{n} \boldsymbol{x}^a \otimes \tilde{\boldsymbol{N}}^a} \tag{3.8}$$

Thus, (3.4) has been transformed into a discrete sum over the deformed nodal positions $\boldsymbol{x}^a$ and the associated surface normals $\tilde{\boldsymbol{N}}^a$ of all surface nodes, which can conveniently be evaluated for a given RVE and state of deformation.

3.2 Average Stress Tensor

Analogous to the above, we derive a discrete version of the averaging theorem for the stress tensor, starting with

$$\langle \boldsymbol{P} \rangle = \frac{1}{V}\int_{\partial\Omega} \boldsymbol{T} \otimes \boldsymbol{X}\, \mathrm{d}S. \tag{3.9}$$

Within the FE context, the external force $\boldsymbol{F}^a$ applied to node a is obtained from the external tractions $\boldsymbol{T}$ acting on the element boundaries via (Figure 3.2)

$$\boldsymbol{F}^a = \int_{\partial\Omega} \boldsymbol{T} N^a\, \mathrm{d}S. \tag{3.10}$$

For isoparametric elements (which use the same interpolation of coordintes as of degrees of freedom) we may write for the undeformed position $\boldsymbol{X}$ within each element e

$$\boldsymbol{X} = \sum_{a=1}^{n} \boldsymbol{X}_e^a N_e^a(\boldsymbol{X}). \tag{3.11}$$

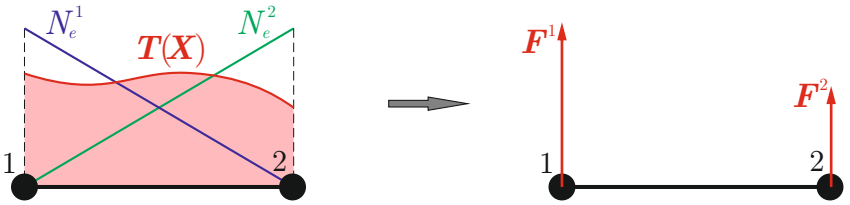


Figure 3.2: Nodal forces are obtained from continuously distributed surface tractions by weighted integrals involving the shape functions on the surface.

Applying both to (3.9) yields the approximate first Piola-Kirchhoff stress tensor

$$\langle \boldsymbol{P}^h \rangle = \frac{1}{V} \sum_e \int_{\partial\Omega_e'} \boldsymbol{T} \otimes \sum_{a=1}^{n} \boldsymbol{X}_e^a N_e^a(\boldsymbol{X}) \,\mathrm{d}S = \frac{1}{V} \sum_e \sum_{a=1}^{n} \int_{\partial\Omega_e'} \boldsymbol{T} N_e^a \,\mathrm{d}S \otimes \boldsymbol{X}^a, \tag{3.12}$$

where a runs over all (global) boundary nodes. If we now introduce the total force $\boldsymbol{F}^a$ applied to each boundary node as the sum of all element contributions, i.e.,

$$\boldsymbol{F}^a = \sum_e \int_{\partial\Omega_e'} \boldsymbol{T} N_e^a \,\mathrm{d}S, \tag{3.13}$$

then we arrive at

$$\boxed{\langle \boldsymbol{P}^h \rangle = \frac{1}{V} \sum_a \boldsymbol{F}^a \otimes \boldsymbol{X}^a} \tag{3.14}$$

Again, we have transformed the continuous averaging theorem, (3.9), into a discretized version to be evaluated in FE simulations by summing over all boundary nodes a the outer product of force $\boldsymbol{F}^a$ acting on that node and undeformed position $\boldsymbol{X}^a$ of that node.

3.3 Linearized Kinematics

The analogous relations can be derived for linearized kinematics. Without showing the full derivations (which are analogous to the finite-deformation case and are omitted here for conciseness), we present the averaging theorem for the infinitesimal strain tensor,

$$\langle \boldsymbol{\varepsilon} \rangle = \frac{1}{V} \int_{\partial\Omega} \operatorname{sym}(\boldsymbol{u} \otimes \boldsymbol{n}) \,\mathrm{d}S, \tag{3.15}$$

which can be treated analogously to (3.8), leading to the approximate averaging relation

$$\boxed{\langle \boldsymbol{\varepsilon}^h \rangle = \frac{1}{V} \sum_{a=1}^{n} \operatorname{sym}(\boldsymbol{u}^a \otimes \tilde{\boldsymbol{n}}^a)} \quad \text{with} \quad \tilde{\boldsymbol{n}}^a = \frac{1}{d} \sum_e \boldsymbol{n}_e^a A_e. \tag{3.16}$$

Similarly, starting with the averaging theorem for the stress tensor,

$$\langle \boldsymbol{\sigma} \rangle = \frac{1}{V} \int_{\partial\Omega} \boldsymbol{t} \otimes \boldsymbol{x} \,\mathrm{d}S, \tag{3.17}$$

by analogy to (3.14) we arrive at the discrete version of the average stress theorem:

$$\boxed{\langle \boldsymbol{\sigma}^h \rangle = \frac{1}{V} \sum_a \boldsymbol{F}^a \otimes \boldsymbol{x}^a} \quad \text{with} \quad \boldsymbol{F}^a = \sum_e \int_{\partial\Omega_e'} \boldsymbol{t} N_e^a \,\mathrm{d}S. \tag{3.18}$$

We have thus derived a set of handy relations that allow us to extract the average stress and strain tensors (in linearized and finite kinematics) of an RVE from information on the RVE boundary – both for the continuous case and for the FE setting.

4 Homogenization Problem

We proceed to discuss the core of all computational homogenization, the classical **homogenization problem**. For our purposes we limit ourselves to the *mechanical* homogenization problem (and we start with finite kinematics), while analogous formulations can be derived for other problems such as those in electromagnetism or thermal transport. Throughout, we consider a material with a statistically homogeneous microstructure.

Consider a *heterogeneous body* with some microstructure, so the constitutive relation changes from point to point, i.e., $\boldsymbol{P} = \boldsymbol{P}(\boldsymbol{X}, \boldsymbol{F})$. For example, in a two-phase composite each phase i has its own constitutive behavior $\boldsymbol{P} = \boldsymbol{P}_i(\boldsymbol{F})$; consequently, if we denote by $\Omega_i \subset \Omega$ the domain within a body Ω which is occupied by phase i, then we have

$$\boldsymbol{P}(\boldsymbol{X}, \boldsymbol{F}) = \begin{cases} \boldsymbol{P}_1(\boldsymbol{F}), & \text{if } \boldsymbol{X} \in \Omega_1, \\ \boldsymbol{P}_2(\boldsymbol{F}), & \text{if } \boldsymbol{X} \in \Omega_2, \end{cases} \tag{4.1}$$

and we must have $\Omega_1 \cup \Omega_2 = \Omega$ and $\Omega_1 \cap \Omega_2 = \emptyset$ for uniqueness. Hence, the constitutive law jumps when crossing a phase boundary. The same applies, e.g., to the mass density $R(\boldsymbol{X})$.

The mechanical boundary value problem on this heterogeneous body is what we refer to as the **heterogeneous problem**, composed of linear momentum balance with suitable boundary conditions and constitutive relations (assuming quasistatics):

$$\left[\begin{aligned} &\text{find } \boldsymbol{\varphi}(\boldsymbol{X}) \text{ s.t.} \\ &\quad \operatorname{Div} \boldsymbol{P}\big(\boldsymbol{X}, \boldsymbol{F}(\boldsymbol{X})\big) + R(\boldsymbol{X})\boldsymbol{B}(\boldsymbol{X}) = \boldsymbol{0} \quad \text{in } \Omega, \\ &\qquad \boldsymbol{\varphi}(\boldsymbol{X}) = \hat{\boldsymbol{x}}(\boldsymbol{X}) \quad \text{on } \partial\Omega_D, \\ &\qquad \boldsymbol{P}\big(\boldsymbol{X}, \boldsymbol{F}(\boldsymbol{X})\big)\boldsymbol{N}(\boldsymbol{X}) = \hat{\boldsymbol{T}}(\boldsymbol{X}) \quad \text{on } \partial\Omega_N, \\ &\qquad \boldsymbol{P} = \boldsymbol{P}(\boldsymbol{X}, \boldsymbol{F}), \end{aligned}\right. \tag{4.2}$$

where the second equation denotes essential boundary conditions ($\hat{\boldsymbol{x}}(\boldsymbol{X})$ is the imposed (deformed) position of point $\boldsymbol{X}$ on the Dirichlet boundary $\partial\Omega_D$ of the RVE), and the third equation enforces natural boundary conditions ($\hat{\boldsymbol{T}}(\boldsymbol{X})$ is the imposed traction vector at position $\boldsymbol{X}$ on the Neumann boundary $\partial\Omega_N$). Analogous relations can be derived for initial boundary value problems considering dynamic effects (which additionally requires initial conditions).

Because the small-scale variations in material properties from point to point in the macroscopic body are generally much smaller than the macroscopic body itself, all mechanical fields (including the deformation mapping, stresses, and strains) are highly oscillatory when seen on the macroscale (i.e., the scale of body Ω). Therefore, analytical solutions to the above problem are challenging to find in general, while numerical solutions require very fine and hence costly meshes.

Assume that the fields of interest (e.g., stresses and strains) on the microscale vary over a characteristic length scale l, whereas the macroscale features of body Ω (e.g., its dimensions, variations of boundary conditions $\hat{\boldsymbol{T}}$ and $\hat{\boldsymbol{x}}$ and body forces $\boldsymbol{B}$) are of a characteristic length scale L (see Figure 4.1). If

$$\epsilon = \frac{l}{L} \ll 1, \tag{4.3}$$

then we assume what is known as a **separation of scales**, with the goal of separating the *effective* smooth macroscopic material behavior from the highly-oscillatory behavior at smaller scales. Consider, e.g., a metallic polycrystal: from grain to grain the constitutive behavior may

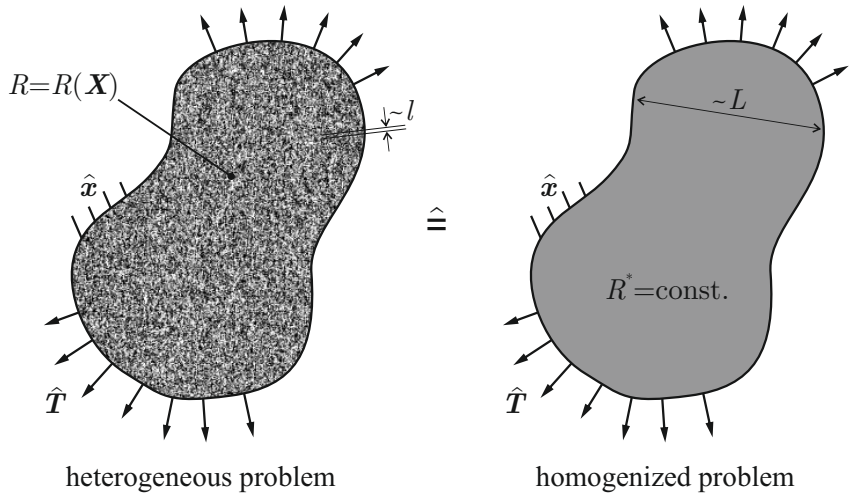


Figure 4.1: Heterogeneous problem (resolving all fluctuations caused my the heterogeneous material microstructure) vs. homogenized problem (solving the problem on a homogeneous macroscale body with effective properties). The separation of scales assumes $l \ll L$.

change significantly due to the changes in crystallographic orientation, yet on the macroscale (orders of magnitude above the characteristic size of grains), we "feel" a homogeneous, somehow averaged material constitutive behavior. This is our separation of scales. By contrast, if you perform, e.g., nano-indentation experiments into the polycrystal, so that the boundary conditions (the size of the indenter) does indeed pick up individual grains, we cannot assume a separation of scales. One may think of the homogenized problem as one in which the small-scale features are "smeared out".

If we may (approximately) assume a separation of scales, then taking the limit $\epsilon \to 0$ leads to the constitutive behavior of a homogenized or **effective material** with *constant*, homogeneous material properties on the macroscale, as long as the microstructure is *statistically homogeneous*.

For example, the homogenized material has an *effective* mass density $R(\boldsymbol{X}) = R^*$ for all $\boldsymbol{X} \in \Omega$ in a macroscopic body Ω (the small-scale variations are "smeared out"). Likewise, the effective constitutive response $\boldsymbol{P}(\boldsymbol{X}, \boldsymbol{F}) = \boldsymbol{P}^*(\boldsymbol{F})$ characterizes the averaged material behavior at the macroscale. Let us generally label all *effective* macroscale quantities by an asterisk $(\cdot)^*$. Now, we are in place to define the **homogenized problem** (again considering quasistatics):

$$\left[\begin{array}{l}\text{find } \boldsymbol{\varphi}^*(\boldsymbol{X}) \text{ s.t.}\\ \qquad \operatorname{Div} \boldsymbol{P}^*\big(\boldsymbol{F}(\boldsymbol{X})\big) + R^*\boldsymbol{B}(\boldsymbol{X}) = \boldsymbol{0} \quad \text{in } \Omega,\\ \qquad\qquad \boldsymbol{\varphi}^*(\boldsymbol{X}) = \hat{\boldsymbol{x}}(\boldsymbol{X}) \quad \text{on } \partial\Omega_D,\\ \qquad\qquad \boldsymbol{P}^*(\boldsymbol{F}(\boldsymbol{X}))\boldsymbol{N}(\boldsymbol{X}) = \hat{\boldsymbol{T}}(\boldsymbol{X}) \quad \text{on } \partial\Omega_N,\\ \qquad\qquad \boldsymbol{P}^* = \boldsymbol{P}^*(\boldsymbol{F}).\end{array}\right. \tag{4.4}$$

The solution $\boldsymbol{\varphi}^*(\boldsymbol{X})$ of (4.4) is smooth compared to the solution $\boldsymbol{\varphi}(\boldsymbol{X})$ of the heterogeneous problem (4.2) (instead of accounting for all small-scale variations in material properties, the homogeneous problem only "sees" the average material response at a larger scale). Note that we here assume a *local* form of the constitutive relations ($\boldsymbol{P}$ only depends on $\boldsymbol{F}$ and not on gradients of $\boldsymbol{F}$), which may not be possible in general due to, e.g., discontinuous stress or strain fields at material interfaces at the micro-level or when modeling damage or localization. However, when using the FE method, we solve the problem with relaxed continuity requirements anyways and we assume that the possibly discontinuous micro-fields translate into smooth macro-fields.

We point out that one can readily extend the above homogenized problem to *statistically in-*

homogeneous microstructures, in which case we have an *inhomogeneous effective material* with $\boldsymbol{P}^* = \boldsymbol{P}^*(\boldsymbol{X}, \boldsymbol{F}^*)$ and $R^* = R^*(\boldsymbol{X})$, where the variation in $\boldsymbol{X}$ is of $O(L)$ and not $O(l)$, owing to the separation of scales.

Solving the homogenized problem (4.4) can be accomplished by classical FE techniques, as long as the effective material properties are known. Therefore, in order to solve the homogenized problem, we first need to find the *effective density* R^* as well as the *effective constitutive relation* $\boldsymbol{P}^* = \boldsymbol{P}^*(\boldsymbol{F}^*)$. As discussed before, we generally aim to identify the **effective response** from an RVE that was deemed sufficiently large to find (an approximation of) the limit $\langle\cdot\rangle_\infty$ by a single test, independent of the particular microstructural realization. Precisely that is the objective of **computational homogenization** to be discussed in the following.

5 Micro-to-Macro Transition

We need to introduce a so-called **micro-to-macro transition** which uniquely defines the effective material properties of a heterogeneous solid to be used in the homogenized problem on the macroscale.

5.1 Effective Material Properties

Let us consider an RVE of volume V, which is (approximately) representative of the statistically homogeneous microstructure of a material. By conservation of mass, we can find the **effective mass density** via

$$R^*V = \int_\Omega R(\boldsymbol{X})\,\mathrm{d}V \qquad \Rightarrow \qquad \boxed{R^* = \langle R\rangle} \tag{5.1}$$

You may think of this relation as replacing the heterogeneous RVE with spatially varying density $R(\boldsymbol{X})$ by a homogenized RVE of the same volume but with the constant effective density R^*. Obviously, these two should have the same mass – this is exactly the above equality. The homogenized material in this case is also referred to as a homogeneous **reference medium**.

Analogously, we define the effective stress and strain tensors as the RVE-averaged quantities, specifically

$$\boxed{\boldsymbol{P}^* = \langle \boldsymbol{P}\rangle} \qquad \text{and} \qquad \boxed{\boldsymbol{F}^* = \langle \boldsymbol{F}\rangle} \tag{5.2}$$

Note that this is in fact a definition and not immediately justified by conservation arguments.

In order to find an effective constitutive relation we must relate $\boldsymbol{P}^* = \langle \boldsymbol{P}\rangle$ to $\boldsymbol{F}^* = \langle \boldsymbol{F}\rangle$. That is, we must apply some boundary conditions (BCs) to the RVE in order to induce, e.g., an average strain field $\boldsymbol{F}^*$. For a given set of BCs, one can then solve the equilibrium equations within the RVE and use the above averaging theorems to identify the respective other average, e.g., $\boldsymbol{P}^*$. Then, one *defines* the effective constitutive relation as that functional form that relates $\boldsymbol{P}^*$ and $\boldsymbol{F}^*$. What exact BCs to apply is a question we will need to discuss.

It is important to note this choice of relating $\boldsymbol{P}^*$ to $\boldsymbol{F}^*$ is not unique. One could alternatively use other stress or strain measures (e.g., the second Piola-Kirchhoff stress tensor $\boldsymbol{S} = \boldsymbol{F}^{-1}\boldsymbol{P}$ or the Cauchy stress tensor $\boldsymbol{\sigma} = J^{-1}\boldsymbol{P}\boldsymbol{F}^{\mathrm{T}}$, and the right Cauchy-Green strain tensor $\boldsymbol{C} = \boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}$ or the Green-Lagrange tensor $\boldsymbol{E} = \frac{1}{2}(\boldsymbol{C} - \boldsymbol{I})$). Choosing any of those alternative stress and strain measures for the effective stress and strain will result in different effective constitutive relations because generally, e.g.,

$$\langle \boldsymbol{P}\rangle = \langle \boldsymbol{F}\boldsymbol{S}\rangle \neq \langle \boldsymbol{F}\rangle\,\langle \boldsymbol{S}\rangle. \tag{5.3}$$

That is, even though $\boldsymbol{P} = \boldsymbol{F}\boldsymbol{S}$ holds pointwise in the RVE, we cannot use $\langle \boldsymbol{F}\rangle$ and $\langle \boldsymbol{P}\rangle$ to find $\langle \boldsymbol{S}\rangle$ since $\langle \boldsymbol{S}\rangle \neq \langle \boldsymbol{F}\rangle^{-1}\langle \boldsymbol{P}\rangle$ in general. In the following, we stick to our choice of $\boldsymbol{P}^*$ and $\boldsymbol{F}^*$ as stress and strain measures in order to extract the effective response. In linearized kinematics this is not an issue, as one can uniquely use $\boldsymbol{\varepsilon}^*$ and $\boldsymbol{\sigma}^*$.

In order to avoid the influence of body forces (which is legitimate if body forces vary slowly on the macroscale), we will solve the RVE problem *in the absence of body forces*. Also, we will assume that *the effective constitutive behavior is an intrinsic material property* that is not affected by the boundary value problem to be solved. This implies that, if dynamic behavior is present, the wavelength λ of any dynamic behavior is on the order of the macroscale, $\lambda \sim L$ (and not

$\lambda \sim l$). This allows us to include inertial effects on the macroscale, so that the RVE problem to be solved does not include inertial effects. Overall, this implies that *the homogenization problem to be solved is the quasistatic* one, i.e.,

$$\boxed{\operatorname{div} \boldsymbol{P}(\boldsymbol{X}, \boldsymbol{F}) = \boldsymbol{0} \quad \text{in } \Omega} \qquad \text{or} \qquad \boxed{\operatorname{div} \boldsymbol{\sigma}(\boldsymbol{x}, \boldsymbol{\varepsilon}) = \boldsymbol{0} \quad \text{in } \Omega} \tag{5.4}$$

The **general strategy** to extract an effective constitutive behavior of a material from its RVE is the following: we impose $\boldsymbol{F}^* = \langle \boldsymbol{F} \rangle$ onto an RVE through some suitable BCs and solve the equilibrium relations (5.4) within the RVE. The resulting stresses within the RVE are averaged to obtain $\boldsymbol{P}^* = \langle \boldsymbol{P} \rangle$. Of course, this procedure can also be inverted (imposing $\boldsymbol{P}^*$ and calculating $\boldsymbol{F}^*$). Either way, the effective relation between $\boldsymbol{P}^* \longleftrightarrow \boldsymbol{F}^*$ is computed from the RVE. Unfortunately, in order to identify suitable BCs that let us impose average strains or stresses onto an RVE, we need a few more concepts to be discussed next.

5.2 The Hill-Mandel Condition

To define a micro-to-macro transition for stress-strain relations, we start by applying another conservation principle: the conservation of energy. Strictly speaking, for a conservative system (without any dissipation), the strain energy stored within the microscale should match the energy stored in the material point on the macroscale. For example, for a *linear elastic* problem with a strain energy density $W = W(\boldsymbol{x}, \boldsymbol{\varepsilon})$ on the microscale and an effective energy density $W^* = W^*(\boldsymbol{\varepsilon}^*)$, we postulate

$$V\, W^*(\boldsymbol{\varepsilon}^*) = \int_\Omega W\big(\boldsymbol{x}, \boldsymbol{\varepsilon}(\boldsymbol{x})\big)\, \mathrm{d}V, \tag{5.5}$$

where $V = |\Omega|$ is again the RVE volume. This relation is analogous to (5.1). Noting that, for a linear elastic medium, $\boldsymbol{\sigma} = \mathbb{C}\boldsymbol{\varepsilon}$ and $W = \frac{1}{2}\boldsymbol{\varepsilon} \cdot \mathbb{C}\boldsymbol{\varepsilon} = \frac{1}{2}\boldsymbol{\sigma} \cdot \boldsymbol{\varepsilon}$ with the fourth-order stiffness tensor $\mathbb{C}$, (5.5) implies immediately that

$$W^* = \langle W(\boldsymbol{x}) \rangle \qquad \Rightarrow \qquad \boldsymbol{\sigma}^* \cdot \boldsymbol{\varepsilon}^* = \langle \boldsymbol{\sigma}(\boldsymbol{x}) \cdot \boldsymbol{\varepsilon}(\boldsymbol{x}) \rangle. \tag{5.6}$$

While this does not yet define the effective stress and strain, it does provide a link between macroscale and microscale quantities, which holds within linear elasticity.

A more general approach that holds for all types of material models is to invoke the principle of virtual work, to postulate that the **virtual work** at the macroscale equals the total virtual work within the RVE on the microscale – or that the virtual work of the heterogeneous RVE matches that of its homogeneous reference medium. In the finite or linearized kinematic settings, virtual work at a point is, respectively, $\boldsymbol{P} \cdot \delta\boldsymbol{F}$ or $\boldsymbol{\sigma} \cdot \delta\boldsymbol{\varepsilon}$, so we postulate (in finite strains)

$$\boxed{\boldsymbol{P}^* \cdot \delta\boldsymbol{F}^* = \langle \boldsymbol{P}(\boldsymbol{X}) \cdot \delta\boldsymbol{F}(\boldsymbol{X}) \rangle} \tag{5.7}$$

This is the called **Hill-Mandel condition** for micro-macro equivalence of a general (local) constitutive behavior.

The Hill-Mandel condition can be re-interpreted in two ways.

First, the above statement is equivalent to equating the micro- and macro-power:

$$\boxed{\boldsymbol{P}^* \cdot \dot{\boldsymbol{F}}^* = \langle \boldsymbol{P}(\boldsymbol{X}) \cdot \dot{\boldsymbol{F}}(\boldsymbol{X}) \rangle} \tag{5.8}$$

That is, the rate of energy increase or loss across the heterogeneous RVE must match that of the homogeneous reference medium (or of the material point on the macroscale). If the constitutive

behavior is conservative, then there exists an energy density $W = W(\boldsymbol{F})$ such that $\boldsymbol{P} = \partial W/\partial \boldsymbol{F}$ (analogous to the linear elastic case). Taking a time derivative to formulate the rate of change of energy as

$$\dot{W} = \frac{\partial W}{\partial \boldsymbol{F}} \cdot \dot{\boldsymbol{F}} = \boldsymbol{P} \cdot \dot{\boldsymbol{F}} \tag{5.9}$$

shows that equating the rate of change of energy on the micro- and macroscale yields

$$V\,\dot{W}^* = \frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega W(\boldsymbol{X}, \boldsymbol{F}(\boldsymbol{X}))\,\mathrm{d}V = \int_\Omega \frac{\partial W}{\partial \boldsymbol{F}}(\boldsymbol{X}, \boldsymbol{F}(\boldsymbol{X}))\cdot\dot{\boldsymbol{F}}(\boldsymbol{X})\,\mathrm{d}V = \int_\Omega \boldsymbol{P}(\boldsymbol{X})\cdot\dot{\boldsymbol{F}}(\boldsymbol{X})\,\mathrm{d}V, \tag{5.10}$$

which again arrives at (5.8), since $\dot{W}^* = \boldsymbol{P}^* \cdot \dot{\boldsymbol{F}}^*$.

Second, instead of solving the strong form, computational techniques usually prefer to solve the weak form which derives from the first variation of the associated variational problem. Since we seek an effective macroscale formulation that is equivalent to the full-resolution microscale problem, we postulate that the averaged first variation on the microscale,

$$\frac{1}{V}\,\delta\int_\Omega W\,\mathrm{d}V = \frac{1}{V}\int_\Omega \frac{\partial W}{\partial \boldsymbol{F}} \cdot \delta\boldsymbol{F}(\boldsymbol{X})\,\mathrm{d}V = \langle \boldsymbol{P}(\boldsymbol{X}) \cdot \delta\boldsymbol{F}(\boldsymbol{X})\rangle, \tag{5.11}$$

equals the effective pointwise variation on the macroscale, which is

$$\delta W^* = \boldsymbol{P}^* \cdot \delta\boldsymbol{F}^*. \tag{5.12}$$

Equating $V\,\delta W^*$ to (5.11) again leads to (5.7).

In summary, we have identified the Hill-Mandel relation (5.7) as a means to enforce energetic consistency between the micro- and macroscales, which is applicable in general. Next, we need to exploit that relation to find out which boundary conditions to apply to an RVE.

For ease of notation, we will in the following omit the variation in the Hill-Mandel condition (if we can show that the equality holds for $\boldsymbol{P}\cdot\delta\boldsymbol{F}$, then it should also hold for $\boldsymbol{P}\cdot\boldsymbol{F}$ and vice-versa). Thus, we write in finite and infinitesimal kinematics, respectively,

$$\boldsymbol{P}^* \cdot \boldsymbol{F}^* = \langle \boldsymbol{P}(\boldsymbol{X}) \cdot \boldsymbol{F}(\boldsymbol{X})\rangle \qquad \text{and} \qquad \boldsymbol{\sigma}^* \cdot \boldsymbol{\varepsilon}^* = \langle \boldsymbol{\sigma}(\boldsymbol{x}) \cdot \boldsymbol{\varepsilon}(\boldsymbol{x})\rangle. \tag{5.13}$$

Only in linear elasticity, this relation applies stricly and admits a clear energetic interpretation, as discussed above (otherwise, the equivalence must be interpreted with regards to the weak form or the first variation or virtual work).

Before we discuss specific types of boundary conditions, let us rewrite the Hill-Mandel condition to arrive at a more convenient form. First, we expand stresses and strains within the RVE into averages plus fluctuation fields:

$$\boldsymbol{P}(\boldsymbol{X}) = \langle \boldsymbol{P}\rangle + \widetilde{\boldsymbol{P}}(\boldsymbol{X}), \qquad \boldsymbol{F}(\boldsymbol{X}) = \langle \boldsymbol{F}\rangle + \widetilde{\boldsymbol{F}}(\boldsymbol{X}). \tag{5.14}$$

Taking averages of these two forms indicates that

$$\langle \widetilde{\boldsymbol{P}}(\boldsymbol{X})\rangle = \boldsymbol{0}, \qquad \langle \widetilde{\boldsymbol{F}}(\boldsymbol{X})\rangle = \boldsymbol{0}, \tag{5.15}$$

i.e., the fluctuation fields must have zero averages (this is in fact what defines a fluctuation field). Inserting these decompositions leads to (dropping the dependence on $\boldsymbol{X}$ for brevity)

$$\begin{aligned}
\langle \boldsymbol{P}\cdot\boldsymbol{F}\rangle &= \langle \Big(\langle\boldsymbol{P}\rangle + \widetilde{\boldsymbol{P}}\Big)\cdot\Big(\langle\boldsymbol{F}\rangle + \widetilde{\boldsymbol{F}}\Big)\rangle \\
&= \langle\langle\boldsymbol{P}\rangle\cdot\langle\boldsymbol{F}\rangle\rangle + \langle\widetilde{\boldsymbol{P}}\cdot\langle\boldsymbol{F}\rangle\rangle + \langle\langle\boldsymbol{P}\rangle\cdot\widetilde{\boldsymbol{F}}\rangle + \langle\widetilde{\boldsymbol{P}}\cdot\widetilde{\boldsymbol{F}}\rangle \\
&= \langle\boldsymbol{P}\rangle\cdot\langle\boldsymbol{F}\rangle + \langle\widetilde{\boldsymbol{P}}\rangle\cdot\langle\boldsymbol{F}\rangle + \langle\boldsymbol{P}\rangle\cdot\langle\widetilde{\boldsymbol{F}}\rangle + \langle\widetilde{\boldsymbol{P}}\cdot\widetilde{\boldsymbol{F}}\rangle \\
&= \langle\boldsymbol{P}\rangle\cdot\langle\boldsymbol{F}\rangle + \langle\widetilde{\boldsymbol{P}}\cdot\widetilde{\boldsymbol{F}}\rangle,
\end{aligned} \tag{5.16}$$

where we exploited that the average of a constant is that constant. The Hill-Mandel condition requires that $\langle \boldsymbol{P}\cdot\boldsymbol{F}\rangle = \langle \boldsymbol{P}\rangle\cdot\langle \boldsymbol{F}\rangle$, which implies that we can reformulate the Hill-Mandel criterion in terms of the fluctuation fields as

$$\boxed{\langle \widetilde{\boldsymbol{P}}\cdot\widetilde{\boldsymbol{F}}\rangle = 0} \tag{5.17}$$

The following relation will be helpful when analyzing the different boundary condition applied to the RVE:

$$\begin{aligned}\langle \boldsymbol{P}\cdot\boldsymbol{F}\rangle &= \frac{1}{V}\int_\Omega P_{iJ}\varphi_{i,J}\,\mathrm{d}V = \frac{1}{V}\left[\int_{\partial\Omega} P_{iJ}N_J\varphi_i\,\mathrm{d}S - \int_\Omega P_{iJ,J}\varphi_i\,\mathrm{d}V\right]\\ &= \frac{1}{V}\left[\int_{\partial\Omega}\boldsymbol{T}\cdot\boldsymbol{\varphi}\,\mathrm{d}S - \int_\Omega \mathrm{Div}\,\boldsymbol{P}\cdot\boldsymbol{\varphi}\,\mathrm{d}V\right]\\ &= \frac{1}{V}\left[\int_{\partial\Omega}\boldsymbol{T}\cdot\boldsymbol{\varphi}\,\mathrm{d}S - \int_\Omega R(\boldsymbol{A}-\boldsymbol{B})\cdot\boldsymbol{\varphi}\,\mathrm{d}V\right],\end{aligned} \tag{5.18}$$

where we first applied the divergence theorem and then inserted linear momentum balance. Let us assume that phases are perfectly bonded (otherwise, discontinuities would have to be accounted for by additional interface integrals). In addition, as discussed before, the RVE problem is solved without body forces and inertial effects, so that (5.18) simplifies to

$$\boxed{\langle \boldsymbol{P}\cdot\boldsymbol{F}\rangle = \frac{1}{V}\int_{\partial\Omega}\boldsymbol{T}\cdot\boldsymbol{\varphi}\,\mathrm{d}S} \tag{5.19}$$

Analogous to the averaging theorems, this relation allows us to evaluate the volume average of $\boldsymbol{P}\cdot\boldsymbol{F}$ by integration over the RVE boundary only.

6 RVE Boundary Conditions

Having derived those conditions we must adhere to when solving the RVE problem, we still need to define what boundary conditions are to be applied to the RVE (which must ensure that the Hill-Mandel condition is satisfied). For computational homogenization we usually aim to derive or compute effective stress-strain relations for a known RVE with microstructure. In other words, given an effective deformation gradient $\boldsymbol{F}^*$ on the macroscale, what is the corresponding and resulting effective stress $\boldsymbol{P}^* = \boldsymbol{P}^*(\boldsymbol{F}^*)$. In principle, we must take the effective $\boldsymbol{F}^* = \langle \boldsymbol{F} \rangle$ and apply it to the RVE, followed by solving the RVE boundary value problem and then extracting the effective stress as $\boldsymbol{P}^* = \langle \boldsymbol{P} \rangle$. The catch is that we need to *impose an average* $\boldsymbol{F}^* = \langle \boldsymbol{F} \rangle$, i.e., we do not know anything about the boundary of the RVE but we must apparently choose our boundary conditions such that the deformation inside the RVE *on average* equals $\boldsymbol{F}^*$. We could also do the opposite: impose an average $\boldsymbol{P}^* = \langle \boldsymbol{P} \rangle$ and extract the resulting average $\boldsymbol{F}^* = \langle \boldsymbol{F} \rangle$ – yet, we again need to impose an average. Before discussing particular boundary conditions, let us start with an introductory example.

6.1 An Introductory Example: Affine Deformation & Composite Bounds

As a naive example that imposes an average deformation gradient $\boldsymbol{F}^*$ (or strain $\boldsymbol{\varepsilon}^*$) onto an RVE, consider an RVE undergoing affine deformation (Figure 6.1); i.e., every single point inside the RVE deforms according to

$$\boldsymbol{\varphi} = \boldsymbol{F}^* \boldsymbol{X} \qquad \text{or} \qquad \boldsymbol{u} = \boldsymbol{\varepsilon}^* \boldsymbol{x}. \tag{6.1}$$

In this case it is obvious that an average $\boldsymbol{F}^*$ can be imposed easily, because, if every single point exhibits the average deformation gradient $\boldsymbol{F}^*$, then the average must also coincide with $\boldsymbol{F}^*$. That is, for each $\boldsymbol{X}$ within the RVE we have

$$\boldsymbol{F}(\boldsymbol{X}) = \frac{\partial \varphi}{\partial \boldsymbol{X}} = \boldsymbol{F}^* = \text{const.} \qquad \text{or} \qquad \boldsymbol{\varepsilon}(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{\varepsilon}^* + \boldsymbol{\varepsilon}^{*,\mathrm{T}}) = \boldsymbol{\varepsilon}^* = \text{const.}, \tag{6.2}$$

so that

$$\langle \boldsymbol{F} \rangle = \frac{1}{V}\int_\Omega \boldsymbol{F}(\boldsymbol{X})\,\mathrm{d}V = \frac{1}{V}\int_\Omega \boldsymbol{F}^*\,\mathrm{d}V = \boldsymbol{F}^* \qquad \text{or} \qquad \langle \boldsymbol{\varepsilon} \rangle = \frac{1}{V}\int_\Omega \boldsymbol{\varepsilon}(\boldsymbol{x})\,\mathrm{d}v = \boldsymbol{\varepsilon}^*, \tag{6.3}$$

in finite or linearized kinematics, respectively.

Such affine deformation applied to, e.g., a composite of i homogeneous phases (each with a constitutive law $\boldsymbol{P} = \boldsymbol{P}_i(\boldsymbol{F})$ for the ith phase) leads to an average stress response of

$$\langle \boldsymbol{P} \rangle = \frac{1}{V}\int_\Omega \boldsymbol{P}(\boldsymbol{X}, \boldsymbol{F}^*) = \frac{1}{V}\sum_i V_i \boldsymbol{P}_i(\boldsymbol{F}^*) = \sum_i v_i \boldsymbol{P}_i(\boldsymbol{F}^*), \tag{6.4}$$

where V_i is the volume and v_i the volume fraction of phase i. That is, affine deformation simply results in a volume-averaged stress over the various composite phases. Such a construction of the average response is known as the **Taylor model** (frequently used, e.g., in plasticity to extract the effective response of polycrystals with homogeneous constitutive laws assumed within each grain but different from grain to grain because of varying orientations).

The above affine deformation trivially satisfies the Hill-Mandel condition, since $\boldsymbol{F}(\boldsymbol{X}) = \boldsymbol{F}^* = \text{const.}$, so

$$\langle \boldsymbol{P} \cdot \boldsymbol{F} \rangle = \langle \boldsymbol{P} \cdot \boldsymbol{F}^* \rangle = \langle \boldsymbol{P} \rangle \cdot \boldsymbol{F}^* = \langle \boldsymbol{P} \rangle \cdot \langle \boldsymbol{F} \rangle. \tag{6.5}$$

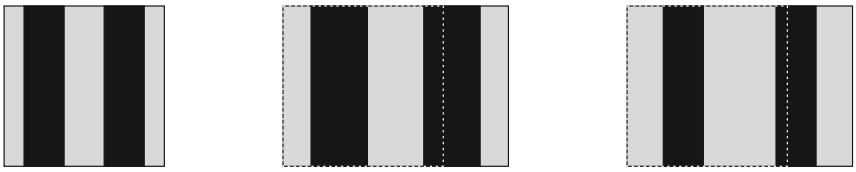


Figure 6.1: A two-phase laminate composite (consisting of a *stiff* black phase and *soft* white phase) is strained uniaxially: affine deformation assumes the same strain in each phase, whereas in reality the soft phase carries most deformation, while the stiff phase remains approximately undeformed. (Note that this is only a schematic illustration; in reality shear effects play a role.)

This hence allows us to impose an average $\boldsymbol{F}^*$ onto the RVE and to calculate the resulting $\boldsymbol{P}^* = \langle \boldsymbol{P} \rangle$. The problem of this affine construction is that it generally leads to a response that is *too stiff*. As shown in Figure 6.1, imagine, e.g., a composite made of one very compliant and one very stiff phase. When only an average deformation is to be imposed, it is unrealistic to assume that both phases are deforming in exactly the same way (it is much more likely that the compliant phase absorbs significantly more deformation than the stiff phase). Hence, saying that the RVE deforms on average with $\boldsymbol{F}^*$ does not necessarily imply that each point inside the RVE should deform with $\boldsymbol{F}^*$.

For the simple example of *linear elasticity*, one can even extract the effective material model from the above construction in closed analytical form. Recall that in linear elasticity

$$\boldsymbol{\sigma}(\boldsymbol{x}, \boldsymbol{\varepsilon}(\boldsymbol{x})) = \mathbb{C}(\boldsymbol{x})\boldsymbol{\varepsilon}(\boldsymbol{x}), \tag{6.6}$$

at every point $\boldsymbol{x}$ inside the RVE (where $\mathbb{C}(\boldsymbol{x})$ is the stiffness tensor, which varies locally). Consequently, for a composite made of phases i with constant moduli $\mathbb{C}_i$ within each phase,

$$\langle \boldsymbol{\sigma}(\boldsymbol{x}) \rangle = \frac{1}{V} \sum_i V_i \mathbb{C}_i \boldsymbol{\varepsilon}^* = \mathbb{C}^*_{\text{Voigt}} \boldsymbol{\varepsilon}^* \qquad \text{with} \qquad \boxed{\mathbb{C}^*_{\text{Voigt}} = \frac{1}{V} \sum_i V_i \mathbb{C}_i} \tag{6.7}$$

That is, the Taylor construction results in the so-called **Voigt bound** on the elastic moduli of the effective, homogenized linear elastic medium. The Voigt bound presents an *upper bound* on the linear elastic effective composite response.

More generally, the **Voigt bound** (6.7) for a general heterogeneous RVE (not an n-phase composite) is

$$\boxed{\mathbb{C}^*_{\text{Voigt}} = \langle \mathbb{C}(\boldsymbol{x}) \rangle = \frac{1}{V} \int_\Omega \mathbb{C}(\boldsymbol{x}) \, \mathrm{d}V} \tag{6.8}$$

which for the case of a composite with constant moduli in each phase reduces to (6.7).

To understand the upper-bound property of the Voigt bound, recall that the solution (e.g., the displacement field $\boldsymbol{u}(\boldsymbol{x})$) of a linear elastic boundary value problem can be obtained from a variational principle, viz. by minimizing the total potential energy $I[\boldsymbol{u}]$ with respect to the displacements (which must match the applied average strains $\boldsymbol{\varepsilon}^*$). That is, we have

$$\boldsymbol{u}(\boldsymbol{x}) = \arg\min I[\boldsymbol{u}] = \arg\min \frac{1}{2} \int_\Omega \boldsymbol{\varepsilon} \cdot \mathbb{C} \, \boldsymbol{\varepsilon} \, \mathrm{d}V \quad \text{s.t.} \quad \langle \boldsymbol{\varepsilon} \rangle = \langle \frac{1}{2} \left(\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^{\mathrm{T}} \right) \rangle = \boldsymbol{\varepsilon}^*. \tag{6.9}$$

Now, let us expand the displacement field into affine deformation plus perturbation, i.e.,

$$\boldsymbol{u}(\boldsymbol{x}) = \boldsymbol{\varepsilon}^* \boldsymbol{x} + \tilde{\boldsymbol{u}}(\boldsymbol{x}) \qquad \Rightarrow \qquad \boldsymbol{\varepsilon}(\boldsymbol{x}) = \boldsymbol{\varepsilon}^* + \tilde{\boldsymbol{\varepsilon}}(\boldsymbol{x}) \qquad \text{and} \qquad \langle \tilde{\boldsymbol{\varepsilon}}(\boldsymbol{x}) \rangle = \boldsymbol{0} \tag{6.10}$$

with $\tilde{\boldsymbol{\varepsilon}} = \frac{1}{2}\left(\nabla\tilde{\boldsymbol{u}} + (\nabla\tilde{\boldsymbol{u}})^{\mathrm{T}}\right)$. Exploiting symmetry of the strain tensor lets us rewrite the total potential energy by inserting $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^* + \tilde{\boldsymbol{\varepsilon}}$ as

$$\begin{aligned} I[\boldsymbol{u}] &= \frac{1}{2}\int_\Omega \boldsymbol{\varepsilon}^* \cdot \mathbb{C}\,\boldsymbol{\varepsilon}^*\,\mathrm{d}V + \int_\Omega \boldsymbol{\varepsilon}^* \cdot \mathbb{C}\,\tilde{\boldsymbol{\varepsilon}}\,\mathrm{d}V + \frac{1}{2}\int_\Omega \tilde{\boldsymbol{\varepsilon}} \cdot \mathbb{C}\,\tilde{\boldsymbol{\varepsilon}}\,\mathrm{d}V \\ &= \frac{1}{2}\boldsymbol{\varepsilon}^* \cdot \int_\Omega \mathbb{C}\,\mathrm{d}V\,\boldsymbol{\varepsilon}^* + \boldsymbol{\varepsilon}^* \cdot \int_\Omega \mathbb{C}\,\tilde{\boldsymbol{\varepsilon}}\,\mathrm{d}V + \frac{1}{2}\int_\Omega \tilde{\boldsymbol{\varepsilon}} \cdot \mathbb{C}\,\tilde{\boldsymbol{\varepsilon}}\,\mathrm{d}V, \end{aligned} \tag{6.11}$$

and equilibrium now corresponds to the minimizer

$$\tilde{\boldsymbol{u}}(\boldsymbol{x}) = \arg\min\left[\frac{1}{2}\boldsymbol{\varepsilon}^* \cdot V\langle\mathbb{C}\rangle\boldsymbol{\varepsilon}^* + \boldsymbol{\varepsilon}^* \cdot \int_\Omega \mathbb{C}\,\tilde{\boldsymbol{\varepsilon}}\,\mathrm{d}V + \frac{1}{2}\int_\Omega \tilde{\boldsymbol{\varepsilon}} \cdot \mathbb{C}\,\tilde{\boldsymbol{\varepsilon}}\,\mathrm{d}V\right] \quad \text{s.t.} \quad \langle\tilde{\boldsymbol{\varepsilon}}\rangle = \mathbf{0}. \tag{6.12}$$

In compliance with the Hill-Mandel condition, the energy of the *equilibrated* RVE ($I[\boldsymbol{u}]$ minimized with respect to $\tilde{\boldsymbol{u}}$) must match that of the effective, homogenized reference medium, so

$$\begin{aligned} \min_{\boldsymbol{u}(\boldsymbol{x})} I[\boldsymbol{u}(\boldsymbol{x})] &= \min_{\tilde{\boldsymbol{u}}(\boldsymbol{x})} I[\boldsymbol{\varepsilon}^*\boldsymbol{x} + \tilde{\boldsymbol{u}}(\boldsymbol{x})] = \min_{\tilde{\boldsymbol{u}}(\boldsymbol{x})}\left[\frac{V}{2}\boldsymbol{\varepsilon}^* \cdot \langle\mathbb{C}\rangle\boldsymbol{\varepsilon}^* + \boldsymbol{\varepsilon}^* \cdot \int_\Omega \mathbb{C}\,\tilde{\boldsymbol{\varepsilon}}\,\mathrm{d}V + \frac{1}{2}\int_\Omega \tilde{\boldsymbol{\varepsilon}} \cdot \mathbb{C}\,\tilde{\boldsymbol{\varepsilon}}\,\mathrm{d}V\right] \\ &= \frac{V}{2}\boldsymbol{\varepsilon}^* \cdot \langle\mathbb{C}\rangle\boldsymbol{\varepsilon}^* + \min_{\tilde{\boldsymbol{u}}(\boldsymbol{x})}\left[\boldsymbol{\varepsilon}^* \cdot \int_\Omega \mathbb{C}\,\tilde{\boldsymbol{\varepsilon}}\,\mathrm{d}V + \frac{1}{2}\int_\Omega \tilde{\boldsymbol{\varepsilon}} \cdot \mathbb{C}\,\tilde{\boldsymbol{\varepsilon}}\,\mathrm{d}V\right] \\ &\overset{!}{=} \frac{V}{2}\boldsymbol{\varepsilon}^* \cdot \mathbb{C}^*\boldsymbol{\varepsilon}^*. \end{aligned} \tag{6.13}$$

Note that $\tilde{\boldsymbol{\varepsilon}}(\boldsymbol{x})$ could be any field as long as it satisfies $\langle\tilde{\boldsymbol{\varepsilon}}(\boldsymbol{x})\rangle = \mathbf{0}$. If we pick one particular such perturbation field, $\tilde{\boldsymbol{\varepsilon}}(\boldsymbol{x}) = \mathbf{0} = \text{const.}$ (which clearly satisfies the constraint), then we must have

$$\min_{\tilde{\boldsymbol{u}}(\boldsymbol{x})}\left[\boldsymbol{\varepsilon}^* \cdot \int_\Omega \mathbb{C}\,\tilde{\boldsymbol{\varepsilon}}\,\mathrm{d}V + \frac{1}{2}\int_\Omega \tilde{\boldsymbol{\varepsilon}} \cdot \mathbb{C}\,\tilde{\boldsymbol{\varepsilon}}\,\mathrm{d}V\right] \leq \left[\boldsymbol{\varepsilon}^* \cdot \int_\Omega \mathbb{C}\,\tilde{\boldsymbol{\varepsilon}}\,\mathrm{d}V + \frac{1}{2}\int_\Omega \tilde{\boldsymbol{\varepsilon}} \cdot \mathbb{C}\,\tilde{\boldsymbol{\varepsilon}}\,\mathrm{d}V\right]_{\tilde{\boldsymbol{\varepsilon}}=\mathbf{0}} = 0. \tag{6.14}$$

Insertion into (6.13) shows that

$$\frac{V}{2}\boldsymbol{\varepsilon}^* \cdot \mathbb{C}^*\boldsymbol{\varepsilon}^* = \frac{V}{2}\boldsymbol{\varepsilon}^* \cdot \langle\mathbb{C}\rangle\boldsymbol{\varepsilon}^* + \min_{\tilde{\boldsymbol{u}}(\boldsymbol{x})}\left[\boldsymbol{\varepsilon}^* \cdot \int_\Omega \mathbb{C}\,\tilde{\boldsymbol{\varepsilon}}\,\mathrm{d}V + \frac{1}{2}\int_\Omega \tilde{\boldsymbol{\varepsilon}} \cdot \mathbb{C}\,\tilde{\boldsymbol{\varepsilon}}\,\mathrm{d}V\right] \leq \frac{V}{2}\boldsymbol{\varepsilon}^* \cdot \langle\mathbb{C}\rangle\boldsymbol{\varepsilon}^* \tag{6.15}$$

and hence

$$\boldsymbol{\varepsilon}^* \cdot \mathbb{C}^*\boldsymbol{\varepsilon}^* \leq \boldsymbol{\varepsilon}^* \cdot \langle\mathbb{C}\rangle\boldsymbol{\varepsilon}^* \qquad \text{for all symmetric strain tensors } \boldsymbol{\varepsilon}^*. \tag{6.16}$$

Therefore, (6.16) shows that the Voigt construction $\langle\mathbb{C}\rangle$ provides an upper bound to the effective stiffness tensor $\mathbb{C}^*$. (By inserting specific cases, one can show that (6.16) translates into the analogous relations, e.g., for the Young and shear moduli of isotropic composites.)

Alternatively, we could also *impose a uniform constant stress* $\boldsymbol{\sigma}(\boldsymbol{x}) = \boldsymbol{\sigma}^*$ (the so-called **Sachs model**) and use the inversion $\boldsymbol{\varepsilon}(\boldsymbol{x}) = \mathbb{C}^{-1}(\boldsymbol{x})\boldsymbol{\sigma}(\boldsymbol{x})$ to arrive at

$$\langle\boldsymbol{\varepsilon}(\boldsymbol{x})\rangle = \frac{1}{V}\int_\Omega \mathbb{C}^{-1}\boldsymbol{\sigma}^*\,\mathrm{d}V = \left(\mathbb{C}^*_{\text{Reuss}}\right)^{-1}\boldsymbol{\sigma}^* \qquad \text{with} \qquad \boxed{\mathbb{C}_{\text{Reuss}} = \left(\frac{1}{V}\int_\Omega \mathbb{C}^{-1}\right)^{-1} = \langle\mathbb{C}^{-1}\rangle^{-1}} \tag{6.17}$$

and the composite version (for a composite of homogeneous phases)

$$\langle \boldsymbol{\varepsilon}(\boldsymbol{x}) \rangle = \frac{1}{V} \sum_i V_i \mathbb{C}_i^{-1} \boldsymbol{\sigma}^* = (\mathbb{C}^*_{\text{Reuss}})^{-1} \, \boldsymbol{\sigma}^* \qquad \text{with} \qquad \boxed{\mathbb{C}_{\text{Reuss}} = \left(\frac{1}{V} \sum_i V_i \mathbb{C}_i^{-1} \right)^{-1}} \tag{6.18}$$

This results in the **Reuss bound** of the effective linear elastic moduli. The Reuss bound presents a **lower bound** on the linear elastic effective composite response, which is analogous to the above upper-bound property of the Voigt bound (and we mention it here without proof; see, e.g., reference [48]).

Overall, we have thus seen that the stiffness of a heterogeneous, linear elastic medium is bounded by

$$\boxed{\mathbb{C}_{\text{Reuss}} = \langle \mathbb{C}^{-1} \rangle^{-1} \;\leq\; \mathbb{C}^* \;\leq\; \mathbb{C}_{\text{Voigt}} = \langle \mathbb{C} \rangle} \tag{6.19}$$

For example, for an isotropic two-phase composite we conclude that its effective Young's modulus E^* is bounded by (see Figure 6.2 for an example)

$$\left(\frac{v_1}{E_1} + \frac{v_2}{E_2} \right)^{-1} \;\leq\; E^* \;\leq\; v_1 E_1 + v_2 E_2. \tag{6.20}$$

We have thus seen that imposing uniform strains $\boldsymbol{F}^*$ (or $\boldsymbol{\varepsilon}^*$) within the RVE is a viable option to enforce an average $\boldsymbol{F}^* = \langle \boldsymbol{F} \rangle$, and it also allows us to extract an effective stress $\boldsymbol{P}^* = \langle \boldsymbol{P} \rangle$. This could already be used as a computational surrogate for a material model: whenever a stress $\boldsymbol{P}^*$ is sought for a given $\boldsymbol{F}^*$, we apply a uniform deformation across the RVE, compute the resulting stresses and average those to obtain $\boldsymbol{P}^*$. This gives the sought relation $\boldsymbol{P}^* = \boldsymbol{P}^*(\boldsymbol{F}^*)$. For linear elasticity, one can even obtain a closed-form solution for the resulting material model in terms of the effective elastic stiffness tensor: $W(\boldsymbol{\varepsilon}^*) = \frac{1}{2} \boldsymbol{\varepsilon}^* \cdot \mathbb{C}^* \boldsymbol{\varepsilon}$ and $\boldsymbol{\sigma}^* = \mathbb{C}^* \boldsymbol{\varepsilon}^*$. Unfortunately, we also discovered that imposing affine deformation results in a generally too stiff response (while imposing an average stress results in a response that is generally too soft). Therefore, we need improved, more accurate ways to estimate the effective constitutive response – and this boils down to defining proper boundary conditions to be applied to the RVE.

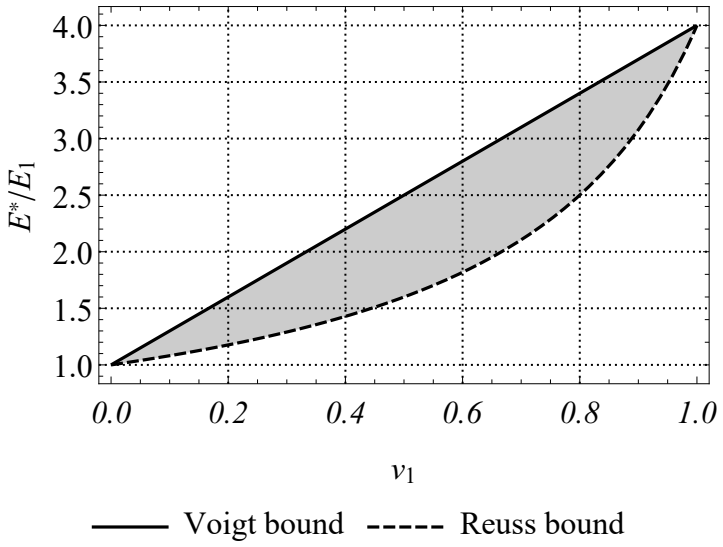


Figure 6.2: Reuss (lower) and Voigt (upper) bound on the effective stiffness of a two-phase composite (phases having moduli E_1 and $E_2 = 4E_1$ and volume fractions v_1 and $v_2 = 1 - v_1$).

6.2 RVE Boundary Conditions

In the above construction, we simply imposed a uniform strain across the entire RVE. It is obvious that this process will result in too stiff of a response: no point inside the RVE is allowed to deform on top of the applied average (e.g., in a composite having soft and stiff phase, it is unreasonable to assume that each phase will deform in the same fashion). Our approach is to not impose the deformation across the entire RVE Ω but only to impose some constraints to its boundary $\partial\Omega$. This will still allow us to impose averages, as explained below, while it allows points inside the RVE and on the boundary to relax – allowing the RVE to accommodate heterogeneous strain and stress distributions and letting the microstructure take control over where high and low strains are found.

There are (at least) three types of boundary conditions that allow us to impose averages while satisfying the Hill-Mandel condition (see Figure 6.3):

(i) **affine displacement BCs**, which enforce Dirichlet boundary conditions on the entire boundary $\partial\Omega$ of the RVE with a given constant deformation gradient $\boldsymbol{F}_0$ (or strains $\boldsymbol{\varepsilon}_0$):

$$\boldsymbol{x} = \boldsymbol{F}_0\boldsymbol{X} \qquad \text{or} \qquad \boldsymbol{u} = \boldsymbol{\varepsilon}_0\boldsymbol{x} \qquad \text{on } \partial\Omega. \tag{6.21}$$

(ii) **uniform traction BCs**, which enforce Neumann boundary conditions on the entire boundary of the RVE with a given constant stress tensor $\boldsymbol{P}_0$ (or $\boldsymbol{\sigma}_0$):

$$\boldsymbol{T} = \boldsymbol{P}_0\boldsymbol{N} \qquad \text{or} \qquad \boldsymbol{t} = \boldsymbol{\sigma}_0\boldsymbol{n} \qquad \text{on } \partial\Omega. \tag{6.22}$$

(iii) **periodic BCs** which enforce on the entire boundary of the RVE:

$$\begin{aligned} &\boldsymbol{x}^+ = \boldsymbol{x}^- + \boldsymbol{F}_0(\boldsymbol{X}^+ - \boldsymbol{X}^-) \qquad \text{and} \qquad \boldsymbol{T}^+ = -\boldsymbol{T}^- \qquad \text{on } \partial\Omega, \\ \text{or} \quad &\boldsymbol{u}^+ = \boldsymbol{u}^- + \boldsymbol{\varepsilon}_0(\boldsymbol{x}^+ - \boldsymbol{x}^-) \qquad \text{and} \qquad \boldsymbol{t}^+ = -\boldsymbol{t}^- \qquad \text{on } \partial\Omega, \end{aligned} \tag{6.23}$$

where superscripts $^\pm$ refer to pairs of periodic points on opposite boundaries of the RVE, as we will discuss below.

Note that we deliberately write $(\cdot)_0$ instad of $(\cdot)^*$ for the imposed strain and stress tensor, since we have not yet demonstrated that any of the above boundary conditions imposes certain average stresses or strains within the RVE. In the following, let us verify that each of these three types of BCs indeed satisfies the Hill-Mandel condition (and is hence suitable for computational homogenization) and let us find out what averages they impose.

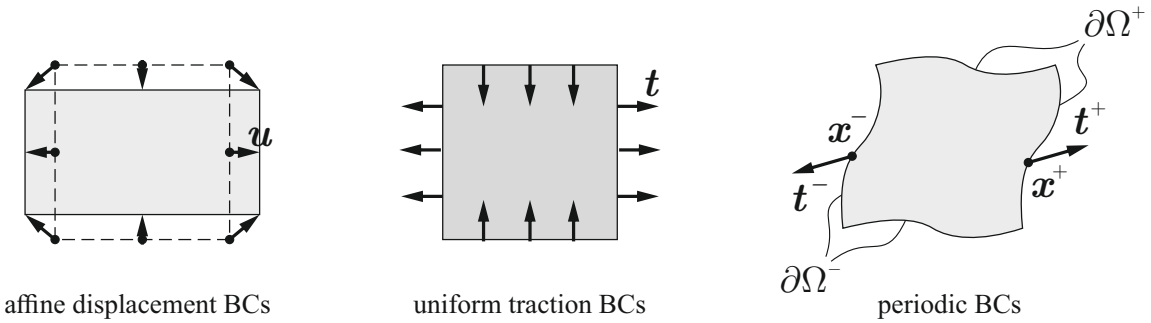


Figure 6.3: Schematic illustration of affine displacement boundary conditions, uniform traction boundary conditions, and periodic boundary conditions applied to an RVE in 2D.

6.3 Affine Displacement BCs

Let us first consider the case of affine displacements imposed across the entire RVE boundary (and let us stick to finite deformations; the case of small strains follows analogously). Here, we

have

$$\boxed{\boldsymbol{\varphi} = \boldsymbol{F}_0 \boldsymbol{X} \quad \text{on } \partial\Omega} \tag{6.24}$$

with some constant tensor $\boldsymbol{F}_0$. Using the average strain theorem (2.9) shows that

$$\langle \boldsymbol{F} \rangle = \frac{1}{V}\int_{\partial\Omega} \boldsymbol{\varphi} \otimes \boldsymbol{N}\,\mathrm{d}S = \frac{1}{V}\int_{\partial\Omega} \boldsymbol{F}_0\boldsymbol{X} \otimes \boldsymbol{N}\,\mathrm{d}S = \frac{\boldsymbol{F}_0}{V}\int_{\partial\Omega} \boldsymbol{X} \otimes \boldsymbol{N}\,\mathrm{d}S = \boldsymbol{F}_0, \tag{6.25}$$

where we used that (using index notation and the divergence theorem)

$$\frac{1}{V}\int_{\partial\Omega} X_I N_J\,\mathrm{d}S = \frac{1}{V}\int_{\Omega} X_{I,J}\,\mathrm{d}V = \delta_{IJ}\frac{1}{V}\int_{\Omega}\mathrm{d}V \qquad \Leftrightarrow \qquad \frac{1}{V}\int_{\partial\Omega} \boldsymbol{X}\otimes\boldsymbol{N}\,\mathrm{d}S = \boldsymbol{I}. \tag{6.26}$$

This shows that $\boldsymbol{F}_0$ is the average deformation gradient inside the RVE, which offers a convenient way to *impose average RVE deformation gradients* via affine displacement BCs, viz. by choosing $\boldsymbol{F}_0 = \boldsymbol{F}^*$.

Next, let us verify that the Hill-Mandel condition is satisfied by affine BCs (where we exploit the relation (5.19)):

$$\begin{aligned}\langle \boldsymbol{P}\cdot\boldsymbol{F}\rangle &= \frac{1}{V}\int_{\partial\Omega} \boldsymbol{T}\cdot\boldsymbol{\varphi}\,\mathrm{d}S = \frac{1}{V}\int_{\partial\Omega} \boldsymbol{T}\cdot\boldsymbol{F}_0\boldsymbol{X}\,\mathrm{d}S = \frac{1}{V}\int_{\partial\Omega} T_i F^0_{iJ} X_J\,\mathrm{d}S \\ &= F^0_{iJ}\frac{1}{V}\int_{\partial\Omega} T_i X_J\,\mathrm{d}S = \boldsymbol{F}_0\cdot\frac{1}{V}\int_{\partial\Omega}\boldsymbol{T}\otimes\boldsymbol{X}\,\mathrm{d}S = \boldsymbol{F}_0\cdot\langle\boldsymbol{P}\rangle,\end{aligned} \tag{6.27}$$

where the last step resulted from the average stress theorem (2.12) (for quasistatic problems and in the absence of body forces).

Therefore, affine displacement BCs with $\boldsymbol{F}_0 = \langle\boldsymbol{F}\rangle$ guarantee that

$$\langle \boldsymbol{P}\cdot\boldsymbol{F}\rangle = \boldsymbol{F}_0\cdot\langle\boldsymbol{P}\rangle = \langle\boldsymbol{F}\rangle\cdot\langle\boldsymbol{P}\rangle, \tag{6.28}$$

which confirms that the Hill-Mandel condition is satisfied. In summary, affine boundary conditions (6.24) satisfy the Hill-Mandel condition, and they allow us to impose an average deformation gradient $\boldsymbol{F}^*$ by choosing $\boldsymbol{F}_0 = \boldsymbol{F}^*$.

6.4 Uniform Traction BCs

Next, consider uniform tractions applied to the entire boundary of the RVE, i.e.,

$$\boxed{\boldsymbol{T} = \boldsymbol{P}_0\boldsymbol{N} \quad \text{on } \partial\Omega} \tag{6.29}$$

where $\boldsymbol{P}_0$ is a constant first Piola-Kirchhoff stress tensor (again we show the derivation for finite strains here, which can analogously by applied to linearized kinematics). Using the average stress theorem for quasistatic problems and in the absence of body forces gives

$$\langle\boldsymbol{P}\rangle = \frac{1}{V}\int_{\partial\Omega}\boldsymbol{T}\otimes\boldsymbol{X}\,\mathrm{d}S = \frac{1}{V}\int_{\partial\Omega}\boldsymbol{P}_0\boldsymbol{N}\otimes\boldsymbol{X}\,\mathrm{d}S = \frac{\boldsymbol{P}_0}{V}\int_{\partial\Omega}\boldsymbol{N}\otimes\boldsymbol{X}\,\mathrm{d}S = \frac{\boldsymbol{P}_0}{V}\int_{\Omega}\mathrm{Grad}\,\boldsymbol{X}\,\mathrm{d}V = \boldsymbol{P}_0,$$

where we used that

$$\frac{1}{V}\int_{\Omega} X_{I,J}\,\mathrm{d}V = \delta_{IJ}\frac{1}{V}\int_{\Omega}\mathrm{d}V \qquad \Leftrightarrow \qquad \frac{1}{V}\int_{\Omega}\mathrm{Grad}\,\boldsymbol{X}\,\mathrm{d}V = \boldsymbol{I}. \tag{6.30}$$

This indicates that $\boldsymbol{P}_0$ is nothing but the average stress inside the RVE, which offers a convenient way to *impose average RVE stresses* via uniform traction BCs, i.e., by choosing $\boldsymbol{P}_0 = \boldsymbol{P}^*$.

Now, let us turn to the Hill-Mandel condition and observe that, using (5.19),

$$\begin{aligned}\langle \boldsymbol{P}\cdot\boldsymbol{F}\rangle &= \frac{1}{V}\int_{\partial\Omega}\boldsymbol{T}\cdot\boldsymbol{\varphi}\,\mathrm{d}S = \frac{1}{V}\int_{\partial\Omega}\boldsymbol{P}_0\boldsymbol{N}\cdot\boldsymbol{\varphi}\,\mathrm{d}S = \frac{1}{V}\int_{\partial\Omega}P^0_{iJ}N_J\varphi_i\,\mathrm{d}S\\ &= \frac{P^0_{iJ}}{V}\int_{\partial\Omega}N_J\varphi_i\,\mathrm{d}S = \boldsymbol{P}_0\cdot\frac{1}{V}\int_{\partial\Omega}\boldsymbol{\varphi}\otimes\boldsymbol{N}\,\mathrm{d}S = \boldsymbol{P}_0\cdot\langle\boldsymbol{F}\rangle,\end{aligned} \tag{6.31}$$

where we re-used identity (2.9) to arrive at the last expression. This implies that uniform traction BCs with $\boldsymbol{P}_0 = \langle\boldsymbol{P}\rangle$ guarantee that

$$\langle \boldsymbol{P}\cdot\boldsymbol{F}\rangle = \boldsymbol{P}_0\cdot\langle\boldsymbol{F}\rangle = \langle\boldsymbol{P}\rangle\cdot\langle\boldsymbol{F}\rangle, \tag{6.32}$$

which confirms that the Hill-Mandel condition is satisfied. In summary, we have thus found a convenient way to impose an average stress to the RVE (viz. by imposing uniform traction boundary conditions), while satisfying the Hill-Mandel condition.

6.5 Periodic BCs

To impose **periodic BCs**, we decompose the entire boundary of the RVE into two opposite parts $\partial\Omega^+$ and $\partial\Omega^-$, so that $\partial\Omega = \partial\Omega^+\cup\partial\Omega^-$ (see Figure 6.3). Each point $\boldsymbol{X}^+\in\partial\Omega^+$ is linked by periodicity to a point $\boldsymbol{X}^-\in\partial\Omega^-$, and the outward normal vectors at those points satisfy $\boldsymbol{N}^- = -\boldsymbol{N}^+$. (Note that this is a strong condition for the RVE mesh, which must be structured such that there are always pairs of symmetrically arranged opposite nodes on the boundaries of the RVE[4].) Periodic BCs at finite strains are then defined by

$$\boxed{\boldsymbol{\varphi}(\boldsymbol{X}^+) - \boldsymbol{\varphi}(\boldsymbol{X}^-) = \boldsymbol{F}_0(\boldsymbol{X}^+ - \boldsymbol{X}^-) \qquad \text{and} \qquad \boldsymbol{T}^+ = -\boldsymbol{T}^- \quad \text{on } \partial\Omega} \tag{6.33}$$

That is, deformed positions $\boldsymbol{x}^\pm = \boldsymbol{\varphi}(\boldsymbol{X}^\pm)$ on the surface (and hence also surface displacements) are forced to be periodic, while tractions are forced to be anti-periodic. By comparison, affine displacement BCs also enforce a periodic deformation (the first condition in (6.33) is satisfied) but they do not necessarily satisfy the second condition in (6.33) (tractions on opposite surfaces may differ). By contrast, uniform traction BCs do enforce the anti-periodic tractions, but they impose no constraint on the deformation, so the first periodicity condition in (6.33) is generally not satisfied there. Notice that for practical purposes we may also rewrite the above (using $\boldsymbol{x} = \boldsymbol{X} + \boldsymbol{u}$) as

$$\boldsymbol{u}^+ - \boldsymbol{u}^- = (\boldsymbol{F}_0 - \boldsymbol{I})(\boldsymbol{X}^+ - \boldsymbol{X}^-). \tag{6.34}$$

The analogous version in linearized kinematics reads

$$\boxed{\boldsymbol{u}^+ - \boldsymbol{u}^- = \boldsymbol{\varepsilon}(\boldsymbol{x}^+ - \boldsymbol{x}^-) \qquad \text{and} \qquad \boldsymbol{t}^+ = -\boldsymbol{t}^- \quad \text{on } \partial\Omega} \tag{6.35}$$

Periodic BCs ensure that, if the RVE is in equilibrium with BCs (6.33), the "tiled" solution of periodically repeated RVEs is also in equilibrium (since tractions are anti-periodic, the sum of all forces at any point on the boundary of the RVE vanishes when periodically tiled). Moreover, the tiled assembly is geometrically matching (since displacements are periodic on opposite boundaries). This altogether implies that stress and strain fields are continuous and periodic,

[4] We here tacitly glance over the complication that arises from the corners of the RVE, since each corner is, in principle, linked by periodicity to multiple other points – essentially all corners are linked to each other by periodicity. We will discuss the treatment of individual points and, in particular, of corners of the RVE in detail in Section 7 and here assume that to each point $\boldsymbol{X}^+$ there exists a unique $\boldsymbol{X}^-$.

which reveals that periodic BCs are ideally suited to simulate bodies with *periodic microstructure*, whose unit cell coincides with the RVE. In practice, one often uses periodic BCs also for problems whose microstructure is not periodic as a good approximation.

In order to find the average strains, let us use the average strain theorem (2.9) and make use of the boundary decomposition and the periodic BCs:

$$\begin{aligned} V\langle \boldsymbol{F}\rangle &= \int_{\partial\Omega} \boldsymbol{\varphi}\otimes\boldsymbol{N}\,\mathrm{d}S = \int_{\partial\Omega^+} \boldsymbol{\varphi}^+\otimes\boldsymbol{N}^+\,\mathrm{d}S + \int_{\partial\Omega^-} \boldsymbol{\varphi}^-\otimes\boldsymbol{N}^-\,\mathrm{d}S \\ &= \int_{\partial\Omega^+} \left[\boldsymbol{\varphi}^- + \boldsymbol{F}_0(\boldsymbol{X}^+ - \boldsymbol{X}^-)\right]\otimes\boldsymbol{N}^+\,\mathrm{d}S + \int_{\partial\Omega^-} \boldsymbol{\varphi}^-\otimes\boldsymbol{N}^-\,\mathrm{d}S \end{aligned} \tag{6.36}$$

We note that for each $\pm$-pair of nodes, there exists a unique $\boldsymbol{\varphi}^\pm$ and outward normals $\boldsymbol{N}^- = -\boldsymbol{N}^+$. This implies that we can replace an integral over $\partial\Omega^-$ by an integral over $\partial\Omega^+$ by flipping the normal, so that continuing the above derivation leads to

$$\begin{aligned} V\langle \boldsymbol{F}\rangle &= \int_{\partial\Omega^+} \left[\boldsymbol{\varphi}^- + \boldsymbol{F}_0(\boldsymbol{X}^+ - \boldsymbol{X}^-)\right]\otimes\boldsymbol{N}^+\,\mathrm{d}S - \int_{\partial\Omega^+} \boldsymbol{\varphi}^-\otimes\boldsymbol{N}^+\,\mathrm{d}S \\ &= \int_{\partial\Omega^+} \boldsymbol{F}_0(\boldsymbol{X}^+ - \boldsymbol{X}^-)\otimes\boldsymbol{N}^+\,\mathrm{d}S \\ &= \int_{\partial\Omega^+} \boldsymbol{F}_0\boldsymbol{X}^+\otimes\boldsymbol{N}^+\,\mathrm{d}S - \int_{\partial\Omega^+} \boldsymbol{F}_0\boldsymbol{X}^-\otimes\boldsymbol{N}^+\,\mathrm{d}S \\ &= \int_{\partial\Omega^+} \boldsymbol{F}_0\boldsymbol{X}^+\otimes\boldsymbol{N}^+\,\mathrm{d}S + \int_{\partial\Omega^-} \boldsymbol{F}_0\boldsymbol{X}^-\otimes\boldsymbol{N}^-\,\mathrm{d}S \\ &= \int_{\partial\Omega} \boldsymbol{F}_0\boldsymbol{X}\otimes\boldsymbol{N}\,\mathrm{d}S = \boldsymbol{F}_0\int_{\partial\Omega} \boldsymbol{X}\otimes\boldsymbol{N}\,\mathrm{d}S = V\boldsymbol{F}_0, \end{aligned} \tag{6.37}$$

where we again exploited the relation (6.26). Overall, this proves that $\langle\boldsymbol{F}\rangle = \boldsymbol{F}_0$ so that we have identified another way to *impose an average deformation gradient* $\boldsymbol{F}^*$ by choosing $\boldsymbol{F}_0 = \boldsymbol{F}^*$: by imposing the periodic BCs (6.33)

Next, we need to verify that the periodic BCs (6.33) satisfy the Hill-Mandel condition. By using analogous steps as in the above calculation of the average deformation gradient, we see that

$$\begin{aligned} V\langle \boldsymbol{P}\cdot\boldsymbol{F}\rangle &= \int_{\partial\Omega} \boldsymbol{T}\cdot\boldsymbol{\varphi}\,\mathrm{d}S = \int_{\partial\Omega^+} \boldsymbol{T}^+\cdot\boldsymbol{\varphi}^+\,\mathrm{d}S + \int_{\partial\Omega^-} \boldsymbol{T}^-\cdot\boldsymbol{\varphi}^-\,\mathrm{d}S \\ &= \int_{\partial\Omega^+} \boldsymbol{T}^+\cdot\left[\boldsymbol{\varphi}^- + \boldsymbol{F}_0(\boldsymbol{X}^+ - \boldsymbol{X}^-)\right]\,\mathrm{d}S - \int_{\partial\Omega^+} \boldsymbol{T}^+\cdot\boldsymbol{\varphi}^-\,\mathrm{d}S \\ &= \int_{\partial\Omega^+} \boldsymbol{T}^+\cdot\boldsymbol{F}_0(\boldsymbol{X}^+ - \boldsymbol{X}^-)\,\mathrm{d}S \\ &= \int_{\partial\Omega^+} \boldsymbol{T}^+\cdot\boldsymbol{F}_0\boldsymbol{X}^+\,\mathrm{d}S - \int_{\partial\Omega^+} \boldsymbol{T}^+\cdot\boldsymbol{F}_0\boldsymbol{X}^-\,\mathrm{d}S \\ &= \int_{\partial\Omega^+} \boldsymbol{T}^+\cdot\boldsymbol{F}_0\boldsymbol{X}^+\,\mathrm{d}S + \int_{\partial\Omega^-} \boldsymbol{T}^-\cdot\boldsymbol{F}_0\boldsymbol{X}^-\,\mathrm{d}S \\ &= \int_{\partial\Omega} \boldsymbol{T}\cdot\boldsymbol{F}_0\boldsymbol{X}\,\mathrm{d}S = \int_{\partial\Omega} T_i F^0_{iJ} X_J\,\mathrm{d}S = F^0_{iJ}\int_{\partial\Omega} T_i X_J\,\mathrm{d}S \\ &= \boldsymbol{F}_0\cdot\int_{\partial\Omega} \boldsymbol{T}\otimes\boldsymbol{X}\,\mathrm{d}S = \boldsymbol{F}_0\cdot V\langle\boldsymbol{P}\rangle, \end{aligned} \tag{6.38}$$

where we used the average stress theorem (2.13) to arrive at the final form (for quasistatic problems and in the absence of body forces). Therefore, since $\boldsymbol{F}_0 = \langle\boldsymbol{F}\rangle$ (as shown above), we conclude that

$$\langle \boldsymbol{P}\cdot\boldsymbol{F}\rangle = \boldsymbol{F}_0\cdot\langle\boldsymbol{P}\rangle = \langle\boldsymbol{F}\rangle\cdot\langle\boldsymbol{P}\rangle, \tag{6.39}$$

which proves that the Hill-Mandel condition is satisfied for periodic boundary conditions. We have thus found in periodic BCs an alternative to affine BCs (Section 6.3), which also imposes an average deformation gradient and satisfies the Hill-Mandel condition.

In summary, we have seen that *uniform traction* BCs offer a convenient way to impose average RVE stresses $\boldsymbol{P}^* = \langle \boldsymbol{P} \rangle$, whereas *affine displacement* BCs can conveniently be used to enforce average deformation gradients $\boldsymbol{F}^* = \langle \boldsymbol{F} \rangle$. Both types of BCs have shortcomings when considering the physics of the problem:

- **Uniform traction BCs** result in a response that is generally *too compliant*: the RVE is in equilibrium but it deforms in a fashion that is not periodically "tile-able", so, loosely speaking, the RVE has more freedom to deform than it should have. Therefore, *uniform tractions generally provide a lower bound* for the obtained constitutive response.

- **Affine displacement BCs**, by contrast, result in a response that is *too stiff*: the RVE must deform affinely, resulting in a "tile-able" geometry but at the cost that the boundary is rigidly constrained to deform in a fixed manner (so the RVE has no freedom to deform on its boundary). Therefore, *uniform displacements generally provide an upper bound.*

- **Periodic BCs** offer a compromise: the RVE is in equilibrium and opposite tractions cancel pairwise, while the RVE is also geometrically periodic and hence "tile-able". This means the *periodically tiled RVE is geometrically compatible and statically in equilibrium.* Its response lies between the upper and lower bounds of affine displacement and uniform traction BCs and in practice provides a reasonable approximation of the real, homogenized constitutive response (even for non-periodic microstructures).

Recall that affine displacements across the entire RVE (as discussed in Section 6.1) provided the Voigt upper bound (that construction enforces an average deformation but leaves no point, not even within the RVE, the freedom to deform other than affinely). By contrast, the Sachs construction of uniform stresses across the entire RVE provided the Reuss lower bound. Therefore, wen computing the effective RVE response, one generally observes

$$\boxed{\text{Reuss} \;\leq\; \text{uniform traction BC} \;\leq\; \text{periodic BCs} \;\leq\; \text{affine displacement BCs} \;\leq\; \text{Voigt}}$$

where the inequality implies softer/stiffer.

6.6 Enforcing RVE Averages

So far we have only discussed different types of boundary conditions but not under what circumstances they should be used and, especially, how to numerically enforce them. We generally differentiate between three types of RVE formulations:

(i) **average strain-driven BCs** enforce $\langle \boldsymbol{F} \rangle = \boldsymbol{F}^*$. As discussed above, it is natural to use either *affine displacement BCs* or *periodic BCs* for this case.

(ii) **average stress-driven BCs** enforce $\langle \boldsymbol{P} \rangle = \boldsymbol{P}^*$. As discussed above, it is natural to use *uniform traction BCs* for this case.

(iii) **mixed stress/strain-driven BCs** enforce some components of $\langle \boldsymbol{F} \rangle = \boldsymbol{F}^*$ and some components of $\langle \boldsymbol{P} \rangle = \boldsymbol{P}^*$ (i.e., each component enforces either a stress or a strain).

Note that the above examples are not exclusive; e.g., one can also use periodic BCs to enforce average stresses – the formulation is just not as simple as for the cases mentioned above.

The three most popular choices in computational homogenization are the following:

- From an implementation viewpoint, *strain-driven affine displacement BCs* are simplest to enforce, since all they require is to enforce (6.24) for a given average $\boldsymbol{F}_0 = \langle \boldsymbol{F} \rangle = \boldsymbol{F}^*$ everywhere on the RVE boundary $\partial\Omega$.

- *Stress-driven uniform traction BCs* are, in principle, simple to enforce since they require the application of tractions (6.29) with known $\boldsymbol{P}_0 = \langle \boldsymbol{P} \rangle = \boldsymbol{P}^*$ everywhere on the RVE boundary $\partial\Omega$. The devil is in the detail though, since any numerical implementation requires the suppression of rigid body motion, but fixing individual nodes on $\partial\Omega$ leads to uncontrollable reaction forces that may break the applied tractions. Therefore, special caution is required.

- *Strain-driven periodic BCs* can be implemented in a number of ways. Section 7 will introduce a particular numerical formulation to implement those in an elegant practical fashion. Alternatives are, e.g., penalty methods which add to the total potential energy to be minimized for a given the boundary value problem a **constraint potential** of the form

$$I_C[\boldsymbol{u}] = \frac{1}{2\epsilon} \int_{\partial\Omega^+} \left\| (\boldsymbol{F}_0 - \boldsymbol{I})(\boldsymbol{X}^+ - \boldsymbol{X}^-) - (\boldsymbol{u}^+ - \boldsymbol{u}^-) \right\|^2 \mathrm{d}S, \tag{6.40}$$

where $0 < \epsilon \ll 1$. Taking variations yields

$$\begin{aligned} \delta I_C[\boldsymbol{u}] = -\frac{1}{\epsilon} \int_{\partial\Omega^+} \left[(\boldsymbol{F}_0 - \boldsymbol{I})(\boldsymbol{X}^+ - \boldsymbol{X}^-) - (\boldsymbol{u}^+ - \boldsymbol{u}^-) \right] \cdot \delta\boldsymbol{u}^+ \,\mathrm{d}S \\ + \frac{1}{\epsilon} \int_{\partial\Omega^-} \left[(\boldsymbol{F}_0 - \boldsymbol{I})(\boldsymbol{X}^+ - \boldsymbol{X}^-) - (\boldsymbol{u}^+ - \boldsymbol{u}^-) \right] \cdot \delta\boldsymbol{u}^- \,\mathrm{d}S = 0. \end{aligned} \tag{6.41}$$

Therefore, if one considers the total potential energy

$$I[\boldsymbol{u}] = \int_\Omega W(\boldsymbol{F}) \,\mathrm{d}V + I_C[\boldsymbol{u}], \tag{6.42}$$

then the first variation becomes

$$\begin{aligned} \delta I[\boldsymbol{u}] = 0 = \int_\Omega \boldsymbol{P} \cdot \delta\boldsymbol{F} \,\mathrm{d}V - \frac{1}{\epsilon} \int_{\partial\Omega^+} \left[(\boldsymbol{F}_0 - \boldsymbol{I})(\boldsymbol{X}^+ - \boldsymbol{X}^-) - (\boldsymbol{u}^+ - \boldsymbol{u}^-) \right] \cdot \delta\boldsymbol{u}^+ \,\mathrm{d}S \\ + \frac{1}{\epsilon} \int_{\partial\Omega^-} \left[(\boldsymbol{F}_0 - \boldsymbol{I})(\boldsymbol{X}^+ - \boldsymbol{X}^-) - (\boldsymbol{u}^+ - \boldsymbol{u}^-) \right] \cdot \delta\boldsymbol{u}^- \,\mathrm{d}S. \end{aligned} \tag{6.43}$$

Here, we tacitly assumed that the problem is variational (i.e., an energy density and associated potential energy exist). Yet, even for non-variational material models, relation (6.43) can be interpreted as the principle of virtual work, so the penalty term arising in (6.43) can also be added to the virtual work in a more general setting.

Note that, since the work performed by boundary tractions typically enters the virtual work in the form $-\int_{\partial\Omega} \boldsymbol{T} \cdot \delta\boldsymbol{u} \,\mathrm{d}V$, we can identify those two terms arising from the constraint potential as tractions applied on the boundary, viz.

$$\begin{aligned} \boldsymbol{T}^+ &= \frac{1}{\epsilon} \left[(\boldsymbol{F}_0 - \boldsymbol{I})(\boldsymbol{X}^+ - \boldsymbol{X}^-) - (\boldsymbol{u}^+ - \boldsymbol{u}^-) \right] \qquad \text{on } \partial\Omega^+, \\ \boldsymbol{T}^- &= -\frac{1}{\epsilon} \left[(\boldsymbol{F}_0 - \boldsymbol{I})(\boldsymbol{X}^+ - \boldsymbol{X}^-) - (\boldsymbol{u}^+ - \boldsymbol{u}^-) \right] \qquad \text{on } \partial\Omega^-. \end{aligned} \tag{6.44}$$

This shows that tractions are indeed anti-periodic, and parameter ϵ controls the enforcement of periodicity.

The drawback of this method is that for finite $\epsilon < \infty$, the solution is only approximately periodic (and for $\epsilon \ll 1$ conditioning issues can affect the numerical solution process). Similar constraint potentials can be introduced to enforce, e.g., *average strain-driven uniform traction BCs*, but this will not be discussed here.

7 Numerical Aspects of Periodic Boundary Conditions

There are many different ways of implementing periodic BCs numerically. In this section, we will outline one possible, practical solution that avoids penalty constraints and instead enforces the periodicity directly by modifying the system of equations to be solved (note that there are various alternatives, we here focus on one example).

To begin, let us review the general concept of imposing constraints within the FE context. The solution of quasistatic boundary value problems (BVPs) is in general obtained by solving a (non)linear system of equations, here abbreviated as

$$\boldsymbol{f}(\boldsymbol{U}) = \boldsymbol{F}_{\text{int}}(\boldsymbol{U}) - \boldsymbol{F}_{\text{ext}} = \boldsymbol{0}, \tag{7.1}$$

which enforces equilibrium by equating the internal forces $\boldsymbol{F}_{\text{int}}(\boldsymbol{U})$, which depend on the vector $\boldsymbol{U} = \big(\boldsymbol{u}^1, \dots, \boldsymbol{u}^n\big)$ of all nodal displacements, to the externally applied force $\boldsymbol{F}_{\text{ext}}$ vector; both are global vectors containing contributions from all nodes. In general, $\boldsymbol{f}(\boldsymbol{U}) \in \mathbb{R}^{dn}$ for n nodes in d dimensions. Any constraint to be imposed onto the system can be formulated as

$$f_c(\boldsymbol{U}) = 0. \tag{7.2}$$

The simplest case to consider is that of essential BCs. For example, imagine that the ith dof of node a is to be set to δ; this is equivalent to

$$f_c(\boldsymbol{U}) = u_i^a - \delta = 0, \tag{7.3}$$

where $u_i^a = (\boldsymbol{u}^a)_i$ is the ith component of the displacement vector $\boldsymbol{u}^a$ associated with node a. More generally, we may write any scalar linear constraint as

$$f_c(\boldsymbol{U}) = \sum_{a=1}^{n} \sum_{i=1}^{d} \alpha_i^a u_i^a - \delta = 0 \tag{7.4}$$

with some coefficients $\alpha_i^a \in \mathbb{R}$ (most of which will be zero in practice).

Now, consider periodic boundary conditions as a special case, which are applied to a pair of $(\pm)$-nodes, enforcing the following d-dimensional constraint:

$$\boldsymbol{f}_c(\boldsymbol{U}) = \boldsymbol{u}^+ - \boldsymbol{u}^- - (\boldsymbol{F}^* - \boldsymbol{I})(\boldsymbol{X}^+ - \boldsymbol{X}^-) = \boldsymbol{0}. \tag{7.5}$$

Note that (7.5) is vector-valued, so it presents d scalar constraint equations in d dimensions.

As an illustrative example, we consider a general four-sided RVE in 2D (see Figure 7.1); the generalization to 3D is straight-forward but more cumbersome to formulate. Let the corners be nodes 1 through 4 (starting in the bottom left corner), and the two lattice vectors (defining the periodic tiling directions of the undeformed RVE) be defined as

$$\boldsymbol{L}_{21} = \boldsymbol{X}_2 - \boldsymbol{X}_1 = \boldsymbol{X}_3 - \boldsymbol{X}_4, \qquad \boldsymbol{L}_{41} = \boldsymbol{X}_4 - \boldsymbol{X}_1 = \boldsymbol{X}_3 - \boldsymbol{X}_2. \tag{7.6}$$

To prevent rigid body motion, we fix the bottom left node, i.e.,

$$\boldsymbol{\varphi}^1 = \boldsymbol{X}_1 \qquad \text{or} \qquad \boldsymbol{u}^1 = \boldsymbol{0}. \tag{7.7}$$

As a consequence, the displacements of the remaining corner nodes must satisfy

$$\begin{aligned}
\boldsymbol{u}^2 &= \boldsymbol{u}^1 + (\boldsymbol{F}^* - \boldsymbol{I})(\boldsymbol{X}^2 - \boldsymbol{X}^1) = (\boldsymbol{F}^* - \boldsymbol{I})\boldsymbol{L}_{21}, \\
\boldsymbol{u}^4 &= \boldsymbol{u}^1 + (\boldsymbol{F}^* - \boldsymbol{I})(\boldsymbol{X}^4 - \boldsymbol{X}^1) = (\boldsymbol{F}^* - \boldsymbol{I})\boldsymbol{L}_{41}, \\
\boldsymbol{u}^3 &= \boldsymbol{u}^1 + (\boldsymbol{F}^* - \boldsymbol{I})(\boldsymbol{X}^3 - \boldsymbol{X}^1) = (\boldsymbol{F}^* - \boldsymbol{I})(\boldsymbol{L}_{21} + \boldsymbol{L}_{41}) = \boldsymbol{u}^2 + \boldsymbol{u}^4.
\end{aligned} \tag{7.8}$$

This shows that *the displacements of all four corner nodes are uniquely known* in terms of $\boldsymbol{F}^*$.

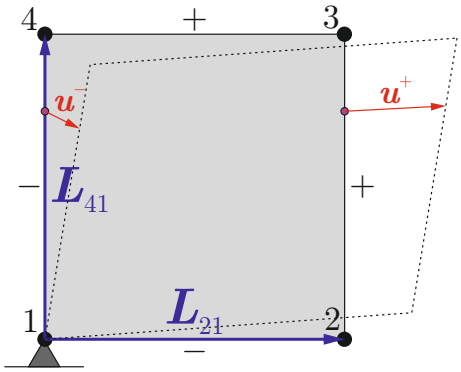


Figure 7.1: Periodic RVE fixed at the bottom left node.

7.1 Enforcing Average Deformation Gradients

In order to implement strain-driven periodic BCs, we consider a pair of $(\pm)$-points located on the left and right boundaries of the RVE. By the periodicity of the unit cell we have

$$\boldsymbol{u}^+ = \boldsymbol{u}^- + (\boldsymbol{F}^* - \boldsymbol{I})(\boldsymbol{X}^+ - \boldsymbol{X}^-) = \boldsymbol{u}^- + (\boldsymbol{F}^* - \boldsymbol{I})\boldsymbol{L}_{21} = \boldsymbol{u}^- + \boldsymbol{u}^2. \tag{7.9}$$

Analogously, a node pair on the top and bottom edges is constrained by

$$\boldsymbol{u}^+ = \boldsymbol{u}^- + (\boldsymbol{F}^* - \boldsymbol{I})(\boldsymbol{X}^+ - \boldsymbol{X}^-) = \boldsymbol{u}^- + (\boldsymbol{F}^* - \boldsymbol{I})\boldsymbol{L}_{41} = \boldsymbol{u}^- + \boldsymbol{u}^4. \tag{7.10}$$

This demonstrates that the displacements of all boundary nodes can be traced back to $\boldsymbol{u}^2$ and $\boldsymbol{u}^4$. In summary, the constraints to be imposed are the following:

$$\begin{aligned}
&\boldsymbol{f}_1(\boldsymbol{U}) = \boldsymbol{u}^1 = \boldsymbol{0}, \\
&\boldsymbol{f}_3(\boldsymbol{U}) = \boldsymbol{u}^3 - \boldsymbol{u}^2 - \boldsymbol{u}^4 = \boldsymbol{0}, \\
&\boldsymbol{f}_+(\boldsymbol{U}) = \boldsymbol{u}^+ - \boldsymbol{u}^- - \boldsymbol{u}^2 = \boldsymbol{0} \qquad \text{for each left/right pair of nodes (not including corners)}, \\
&\boldsymbol{f}_+(\boldsymbol{U}) = \boldsymbol{u}^+ - \boldsymbol{u}^- - \boldsymbol{u}^4 = \boldsymbol{0} \qquad \text{for each top/bottom pair of nodes (not including corners)}
\end{aligned} \tag{7.11}$$

along with

$$\begin{aligned}
\boldsymbol{u}^2 &= (\boldsymbol{F}^* - \boldsymbol{I})\boldsymbol{L}_{21}, \\
\boldsymbol{u}^4 &= (\boldsymbol{F}^* - \boldsymbol{I})\boldsymbol{L}_{41}.
\end{aligned} \tag{7.12}$$

Notice that we formulated all relations in (7.11) free of the average deformation gradient, which will become helpful below. If the average $\boldsymbol{F}^*$ to be enforced is known, all we need to do is implement constraints (7.11) and (7.12), so the complete set of constraints becomes

$$\begin{aligned}
&\boldsymbol{f}_1(\boldsymbol{U}) = \boldsymbol{u}^1 = \boldsymbol{0}, \\
&\boldsymbol{f}_2(\boldsymbol{U}) = \boldsymbol{u}^2 - (\boldsymbol{F}^* - \boldsymbol{I})\boldsymbol{L}_{21} = \boldsymbol{0}, \\
&\boldsymbol{f}_3(\boldsymbol{U}) = \boldsymbol{u}^3 - (\boldsymbol{F}^* - \boldsymbol{I})(\boldsymbol{L}_{21} + \boldsymbol{L}_{41}) = \boldsymbol{0}, \\
&\boldsymbol{f}_4(\boldsymbol{U}) = \boldsymbol{u}^4 - (\boldsymbol{F}^* - \boldsymbol{I})\boldsymbol{L}_{41} = \boldsymbol{0}, \\
&\boldsymbol{f}_+(\boldsymbol{U}) = \boldsymbol{u}^+ - \boldsymbol{u}^- - \boldsymbol{u}^2 = \boldsymbol{0} \qquad \text{for each left/right pair of nodes (not including corners)}, \\
&\boldsymbol{f}_+(\boldsymbol{U}) = \boldsymbol{u}^+ - \boldsymbol{u}^- - \boldsymbol{u}^4 = \boldsymbol{0} \qquad \text{for each top/bottom pair of nodes (not including corners)}.
\end{aligned} \tag{7.13}$$

When it comes to imposing constraints onto the global system of equations, one usually refers to one of two approaches: matrix *condensation* (which is computationally efficient for large systems but harder to implement) and matrix row/column *substitution* (which is easy to implement but can corrupt the conditioning of the system of equations); see, e.g., Section 21.2 of the lecture notes on Computational Mechanics I – Intro to FEA. While it is not the method of choice in performance codes, we here choose substitution for simplicity, and we will also use it in our computational project (condensation is discussed below). Following this approach, one simply replaces rows in the system of equations by constraints (the row to be replaced must be associated with a constrained node). For example, assume $\boldsymbol{KU} = \boldsymbol{F}$ is the linearized system to be solved. In order to set $u_i^a = \delta$, one zeroes out in the matrix that row corresponding to that dof i of node, places 1 on the diagonal in that zeroed out row of the matrix, and sets δ on the right-hand side of the system of equations in the same row. This effectively replaces the original equilibrium equation associated with that node by $1 \cdot u_i^a = \delta$, is required. See Section 21.2 of Computational Mechanics I – Intro to FEA for further details and examples.

It is important to note the following **warning** when implementing periodic boundary conditions following the above strategy. For each pair of $\pm$-nodes, there are two (vector-valued) equations in the global system of equations to be solved (one for each node). However, for each pair of $\pm$-nodes there is only one (vector-valued) constraint equation (see, e.g., (7.13)), so only one of the original two equations in the global system can be replaced by the respective constraint. The other equation, however, cannot remain as is, because it would imply that the force on one of the $\pm$-nodes is set to zero (as on a free boundary). Instead, we must enforce that the sum of the two forces acting on the pair of opposite $\pm$-nodes vanishes (recall that $\boldsymbol{T}^+ = -\boldsymbol{T}^-$, which also guarantees that the tiled RVE is in equilibrium). Therefore, when replacing one equation by a constraint, the respective *other equation must be replaced by the sum of the original two equations* for that pair of $\pm$=nodes. This ensures proper equilibrium on the boundary of the RVE.

An alternative and more general approach is to reformulate the constraint problem by the aid of a so-called *embedding* function. To this end, notice that the full vector of degrees of freedom of the RVE problem, $\boldsymbol{U} \in \mathbb{R}^{dn}$, can be traced back to a vector $\boldsymbol{U}_{\text{red}} \in \mathbb{R}^{dm}$ (with $m < n$) containing all unique or *reduced* degrees of freedom. For example, for each pair of $\pm$-nodes there is only one unique node (the **master node**), while the other node (the **slave node**) follows through the constraint. This allows us to reduce the number of (actual) degrees of freedom significantly by retaining only the master nodes in $\boldsymbol{U}_{\text{red}}$. Figure 7.2 illustrates a simple example of master vs. slave nodes in 2D. Given the above linear constraints, we can formulate the **embedding** as

$$\boldsymbol{U} = \boldsymbol{T}\boldsymbol{U}_{\text{red}} + \boldsymbol{c}, \tag{7.14}$$

with a matrix $\boldsymbol{T} \in \mathbb{R}^{(dn)\times(dm)}$ (containing 1 and 0 entries) and a vector $\boldsymbol{c} \in \mathbb{R}^{dn}$ (containing the differences between the displacements of each $\pm$-pair of nodes). The inverse of the embedding is known as the **submerging** operation.

Recall that we overall seek to minimize the total energy functional I which, in the FE context, is approximated to depend on the degrees of freedom, so $I^h(\boldsymbol{U}) \to \min$. Following the Rayleigh-Ritz approach, the system of (generally nonlinear) equations to be solved is $\partial I/\partial \boldsymbol{U} = \boldsymbol{0}$ subject to constraints. Here, we may introduce the embedding to arrive at

$$I[\boldsymbol{U}] = I[\boldsymbol{T}\boldsymbol{U}_{\text{red}} + \boldsymbol{c}] \to \min_{\boldsymbol{U}_{\text{red}}} \qquad \Rightarrow \qquad \frac{\partial I}{\partial \boldsymbol{U}_{\text{red}}} = \frac{\partial I}{\partial \boldsymbol{U}}\boldsymbol{T} = (\boldsymbol{F}_{\text{int}} - \boldsymbol{F}_{\text{ext}})\boldsymbol{T} = \boldsymbol{0}, \tag{7.15}$$

where $\boldsymbol{F}_{\text{int}} - \boldsymbol{F}_{\text{ext}} \in \mathbb{R}^{dn}$ is the global force vector (conjugate to *all* degrees of freedom).

Note that $\boldsymbol{T}^{\text{T}}$ also has an interesting meaning: it sums all those forces of $\pm$-nodes to result in

$$\boldsymbol{F}_{\text{red}} = \boldsymbol{T}^{\text{T}}\boldsymbol{F}_{\text{int}} \in \mathbb{R}^{dm}, \tag{7.16}$$

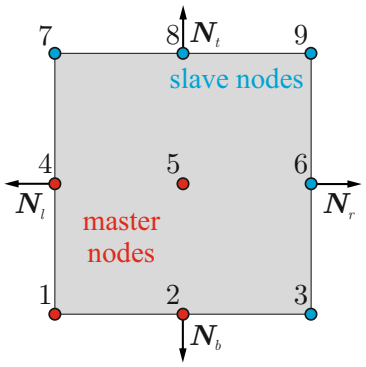


Figure 7.2: Schematic illustration of master nodes (having unique dofs) and slave node (use dofs are linked to those of the master nodes) for a square RVE in 2D.

the global (reduced) net force vector acting on all master nodes only. Therefore, the system of equations to be solved for the reduced degrees of freedom and its linearized version (starting from the full version $\boldsymbol{F}_{\text{int}}(\boldsymbol{U}) \approx \boldsymbol{K}\boldsymbol{U} = \boldsymbol{F}_{\text{ext}}$) are, respectively

$$\boldsymbol{T}^{\text{T}}(\boldsymbol{F}_{\text{int}} - \boldsymbol{F}_{\text{ext}}) = \boldsymbol{0}, \qquad \text{and} \qquad \boldsymbol{T}^{\text{T}}\boldsymbol{K}\boldsymbol{T}\boldsymbol{U}_{\text{red}} = \boldsymbol{T}^{\text{T}}\left(\boldsymbol{F}_{\text{ext}} - \boldsymbol{K}\boldsymbol{c}\right), \tag{7.17}$$

where the latter is the effective equilibrium equation on only the reduced degrees of freedom.

Let us consider the **example** of a square-shaped RVE in 2D of side length L with nine equally spaced nodes (Figure 7.2). The reduced set of degrees of freedom involves only the master nodes 1, 2, 4 and 5, while all other nodes are slave to those master nodes. The embedding is defined (for simplicity written out for a single degree of freedom u per node) by

$$\underbrace{\begin{pmatrix} u^1 \\ u^2 \\ u^3 \\ u^4 \\ u^5 \\ u^6 \\ u^7 \\ u^8 \\ u^9 \end{pmatrix}}_{=\boldsymbol{U}} = \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}}_{=\boldsymbol{T}} \underbrace{\begin{pmatrix} u^1 \\ u^2 \\ u^4 \\ u^5 \end{pmatrix}}_{=\boldsymbol{U}_{\text{red}}} + \underbrace{\begin{pmatrix} 0 \\ 0 \\ c^3 \\ 0 \\ 0 \\ c^6 \\ c^7 \\ c^8 \\ c^9 \end{pmatrix}}_{=\boldsymbol{c}}, \tag{7.18}$$

where the non-zero components of $\boldsymbol{c}$ are to be chosen according to the periodic constraints (involving those $(\boldsymbol{F}^* - \boldsymbol{I})\boldsymbol{L}$-terms). The corresponding nodal forces on all master nodes follow from

$$\underbrace{\begin{pmatrix} F^1_{\text{total}} \\ F^2_{\text{total}} \\ F^4_{\text{total}} \\ F^5_{\text{total}} \end{pmatrix}}_{=\boldsymbol{F}_{\text{red}}} = \underbrace{\begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}}_{=\boldsymbol{T}^{\text{T}}} \underbrace{\begin{pmatrix} F^1 \\ F^2 \\ F^3 \\ F^4 \\ F^5 \\ F^6 \\ F^7 \\ F^8 \\ F^9 \end{pmatrix}}_{=\boldsymbol{F}_{\text{int}} - \boldsymbol{F}_{\text{ext}}} = \begin{pmatrix} F^1 + F^3 + F^7 + F^9 \\ F^2 + F^8 \\ F^4 + F^6 \\ F^5 \end{pmatrix}. \tag{7.19}$$

Also, note that the diagonal entries of matrix $\boldsymbol{T}^{\mathrm{T}}\boldsymbol{T}$ reflect the number of nodal $\pm$-linkages per node. Overall, we have thus *condensed* the system of equations, which is now to be solved only for the dofs of the master nodes.

We have hence discussed two approaches – one based on substitution in the system of equations and one based on a condensation of the system of equations – which both allow us to implement periodic boundary conditions within the FE setting. If the average deformation gradient $\boldsymbol{F}^*$ is known, then those can be used directly to impose the constraints (7.13) and to solve the RVE problem numerically for the displacements of all nodes.

7.2 Enforcing Average Stress Components

In some problems, we may want to impose average stresses. If average stresses $\boldsymbol{P}^*$ are to be imposed in conjunction with periodic BCs, we must impose tractions on the boundary $\partial\Omega$ of the RVE such that $\hat{\boldsymbol{T}} = \boldsymbol{P}^*\boldsymbol{N}$, as shown in Section 6.4. Thus, the total potential energy functional of the RVE (including both internal energy and work done by constant external forces) becomes

$$\begin{aligned} I[\boldsymbol{\varphi}] &= \int_\Omega W(\boldsymbol{F})\,\mathrm{d}V - \int_{\partial\Omega} \hat{\boldsymbol{T}}\cdot\boldsymbol{\varphi}\,\mathrm{d}S \\ &= \int_\Omega W(\boldsymbol{F})\,\mathrm{d}V - \int_{\partial\Omega} \boldsymbol{P}^*\boldsymbol{N}\cdot\boldsymbol{\varphi}\,\mathrm{d}S \\ &= \int_\Omega W(\boldsymbol{F})\,\mathrm{d}V - \boldsymbol{P}^*\cdot\int_{\partial\Omega} \boldsymbol{\varphi}\otimes\boldsymbol{N}\,\mathrm{d}S, \end{aligned} \tag{7.20}$$

which by the averaging theorem (2.9) is equivalent to (V being the undeformed RVE volume)

$$I[\boldsymbol{\varphi}] = \int_\Omega W(\boldsymbol{F})\,\mathrm{d}V - V\,\boldsymbol{P}^*\cdot\langle\boldsymbol{F}\rangle. \tag{7.21}$$

Note that the boundary integral in (7.20) can be decomposed into its bottom (b), top (t), left (l) and right (r) edges in 2D (see Figure 7.2). Furthermore, using $\boldsymbol{\varphi} = \boldsymbol{x} = \boldsymbol{X} + \boldsymbol{u}$ and using the periodicity of top/bottom and left/right edges, we obtain

$$\begin{aligned} \int_{\partial\Omega} \boldsymbol{\varphi}\otimes\boldsymbol{N}\,\mathrm{d}S &= V\boldsymbol{I} + \int_{\partial\Omega_b} \boldsymbol{u}^b\otimes\boldsymbol{N}_b\,\mathrm{d}S + \int_{\partial\Omega_t} \boldsymbol{u}^t\otimes\boldsymbol{N}_t\,\mathrm{d}S + \int_{\partial\Omega_l} \boldsymbol{u}^l\otimes\boldsymbol{N}_l\,\mathrm{d}S + \int_{\partial\Omega_r} \boldsymbol{u}^r\otimes\boldsymbol{N}_r\,\mathrm{d}S \\ &= V\boldsymbol{I} - \int_{\partial\Omega_t} \boldsymbol{u}^b\otimes\boldsymbol{N}_t\,\mathrm{d}S + \int_{\partial\Omega_t} (\boldsymbol{u}^b+\boldsymbol{u}^4)\otimes\boldsymbol{N}_t\,\mathrm{d}S \\ &\quad - \int_{\partial\Omega_r} \boldsymbol{u}^l\otimes\boldsymbol{N}_r\,\mathrm{d}S + \int_{\partial\Omega_r} (\boldsymbol{u}^l+\boldsymbol{u}^2)\otimes\boldsymbol{N}_r\,\mathrm{d}S \\ &= V\boldsymbol{I} + \boldsymbol{u}^4\otimes\int_{\partial\Omega_t} \boldsymbol{N}_t\,\mathrm{d}S + \boldsymbol{u}^2\otimes\int_{\partial\Omega_r} \boldsymbol{N}_r\,\mathrm{d}S, \end{aligned} \tag{7.22}$$

where we re-used relation (6.26) and exploited that the boundary of the RVE can be decomposed into the four edges of the 2D RVE. Be aware that the *node numbering is that of Figure 7.1*. Altogether, we have obtained

$$\begin{aligned} \boldsymbol{P}^*\cdot\int_{\partial\Omega} \boldsymbol{\varphi}\otimes\boldsymbol{N}\,\mathrm{d}S &= \boldsymbol{P}^*\cdot\left[V\boldsymbol{I} + \boldsymbol{u}^4\otimes\int_{\partial\Omega_t} \boldsymbol{N}_t\,\mathrm{d}S + \boldsymbol{u}^2\otimes\int_{\partial\Omega_r} \boldsymbol{N}_r\,\mathrm{d}S\right] \\ &= V\boldsymbol{P}^*\cdot\boldsymbol{I} + \boldsymbol{u}^4\cdot\boldsymbol{P}^*\int_{\partial\Omega_t} \boldsymbol{N}_t\,\mathrm{d}S + \boldsymbol{u}^2\cdot\boldsymbol{P}^*\int_{\partial\Omega_r} \boldsymbol{N}_r\,\mathrm{d}S. \end{aligned} \tag{7.23}$$

In summary, the total potential energy becomes (dropping the term $V\boldsymbol{P}^*\cdot\boldsymbol{I}$, since it is independent of displacements and hence plays no role in the optimization problem)

$$I[\boldsymbol{\varphi}] = \int_\Omega W(\boldsymbol{F})\,\mathrm{d}V - \boldsymbol{u}^4\cdot\boldsymbol{P}^*\int_{\partial\Omega_t}\boldsymbol{N}_t\,\mathrm{d}S - \boldsymbol{u}^2\cdot\boldsymbol{P}^*\int_{\partial\Omega_r}\boldsymbol{N}_r\,\mathrm{d}S. \tag{7.24}$$

If the average stress tensor $\boldsymbol{P}^*$ is known and to be imposed, than the above potential energy in (7.24) can be reinterpreted as that of a BVP which applies constant external forces

$$\hat{\boldsymbol{F}}_4 = \boldsymbol{P}^*\int_{\partial\Omega_t}\boldsymbol{N}_t\,\mathrm{d}S \qquad\text{and}\qquad \hat{\boldsymbol{F}}_2 = \boldsymbol{P}^*\int_{\partial\Omega_r}\boldsymbol{N}_r\,\mathrm{d}S \tag{7.25}$$

to nodes 4 and 2, respectively. Thus, we can enforce an average stress tensor $\boldsymbol{P}^*$ by imposing constraints (7.11) (which impose periodicity without imposing any strain components) along with applying external forces (7.25) to nodes 4 and 2. As we do *not* impose constraints (7.12) in this case, nodes 2 and 4 are free to displace, while the above external forces are being applied to those two nodes.

7.3 Enforcing Mixed Averages

Next, assume that some of the components of $\boldsymbol{F}^*$ and some of $\boldsymbol{P}^*$ are known. For example, consider a uniaxial extension test enforcing a known stretch λ in the X_1-direction, while leaving the transverse surfaces with X_2-normal stress-free. This means

$$\boldsymbol{F}^* = \begin{pmatrix}\lambda & F_{12}\\ F_{21} & F_{22}\end{pmatrix}, \qquad \boldsymbol{P}^* = \begin{pmatrix}P_{11} & 0\\ 0 & 0\end{pmatrix}, \tag{7.26}$$

where the four unknowns are F_{12}, F_{21}, F_{22}, P_{11}.

Enforcing such mixed averages is challenging in general. However, if the RVE has a rectangular undeformed shape, then note that (7.25) reduces to

$$\hat{\boldsymbol{F}}_2 = \boldsymbol{P}^*\boldsymbol{N}_r L_{r/l} = L_{r/l}\begin{pmatrix}P^*_{11}\\ P^*_{21}\end{pmatrix}, \qquad \hat{\boldsymbol{F}}_4 = \boldsymbol{P}^*\boldsymbol{N}_t L_{t/b} = L_{t/b}\begin{pmatrix}P^*_{12}\\ P^*_{22}\end{pmatrix}, \tag{7.27}$$

where $L_{r/l}$ is the side length of the vertical (right/left) edges and $\boldsymbol{N}_r$ is the normal on the right edge, whereas $L_{t/b}$ is the side length of the horizontal (top/bottom) edges and and $\boldsymbol{N}_t$ is the normal on the top edge. Also, (7.12) becomes

$$\boldsymbol{u}^2 = (\boldsymbol{F}^*-\boldsymbol{I})\boldsymbol{N}_r L_{t/b} = \begin{pmatrix}F^*_{11}-1\\ F^*_{21}\end{pmatrix}L_{t/b}, \qquad \boldsymbol{u}^4 = (\boldsymbol{F}^*-\boldsymbol{I})\boldsymbol{N}_t L_{r/l} = \begin{pmatrix}F^*_{12}\\ F^*_{22}-1\end{pmatrix}L_{r/l}. \tag{7.28}$$

Conveniently, the equations have decoupled, so that we may enforce individual components of $\boldsymbol{P}^*$ and $\boldsymbol{F}^*$ simultaneously by enforcing selected components from those two sets of equations. That is, we need to impose periodic BCs (7.11) complemented by whichever components of (7.27) and (7.28) apply – enforcing components of $\boldsymbol{F}^*$ translates into enforcing associated displacement components, while enforcing components of $\boldsymbol{P}^*$ translates into enforcing associated nodal force components.

For example, for the uniaxial extension test described by (7.26), we must enforce (7.11) along with

$$u^2_1 = (\lambda-1)L_{r/l}, \qquad F^2_2 = 0 \qquad\text{and}\qquad \hat{\boldsymbol{F}}_4 = \begin{pmatrix}0\\ 0\end{pmatrix}, \tag{7.29}$$

so that u^2_2 and $\boldsymbol{u}^4$ are to be determined as part of the BVP (node numbering, again, referring to Figure 7.1).

7.4 Three Dimensions

Periodic BCs can be enforced analogously in 3D, where instead of $\boldsymbol{u}^2$ and $\boldsymbol{u}^4$ we must formulate BCs by working with three corner nodes to enforce average stresses and strains. All relations for enforcing average deformation gradients and stress tensors apply analogously, but omit the details here for conciseness (the approach is analogous to the one described above in 2D).

7.5 Linearized Kinematics

Even though we formulated the problem in terms of the deformation gradient and first Piola-Kirchhoff stresses, the above concepts apply analogously in linearized kinematics. As a key difference, periodicity involves the infinitesimal strain tensor $\boldsymbol{\varepsilon}$ instead of deformation gradient $\boldsymbol{F}$. That is, the corresponding linearized expressions are obtained by replacing $(\boldsymbol{F}-\boldsymbol{I})$ by $\boldsymbol{\varepsilon}$ (as well as their averages and effective values) and using normal vectors $\boldsymbol{n}_i$, positions $\boldsymbol{x}_i$, etc. in the linearized framework, while the first Piola-Kirchhoff stress tensor $\boldsymbol{P}$ is replaced by the infinitesimal (Cauchy) stress tensor $\boldsymbol{\sigma}$.

8 The FE^2 Method

Since the above computational homogenization procedure provided a methodology to compute effective stress-strain relations $\boldsymbol{P}^* = \boldsymbol{P}^*(\boldsymbol{F}^*)$, or analogously $\boldsymbol{\sigma}^* = \boldsymbol{\sigma}^*(\boldsymbol{\varepsilon}^*)$, it has become convenient to use such effective constitutive laws within macroscale FE calculations. That is, the effective constitutive law to be used within each element and at each quadrature point on the macroscale is the one obtained from homogenization based on an RVE on the microscale. For simple problems in linear elasticity, the effective response can be derived analytically in closed from (cf. the effective stiffness tensor estimates for composites in Section 6.1). However, in general no such analytical homogenization is possible. Instead, one can compute the effective stress-strain relation numerically *on-the-fly* by solving the BVP at the RVE-level, whenever a stress-strain relation must be evaluated. As this approach couples two FE BVPs (one on the macroscale, and one on the microscale), this approach is known as the **FE^2** method (see Figure 8.1).

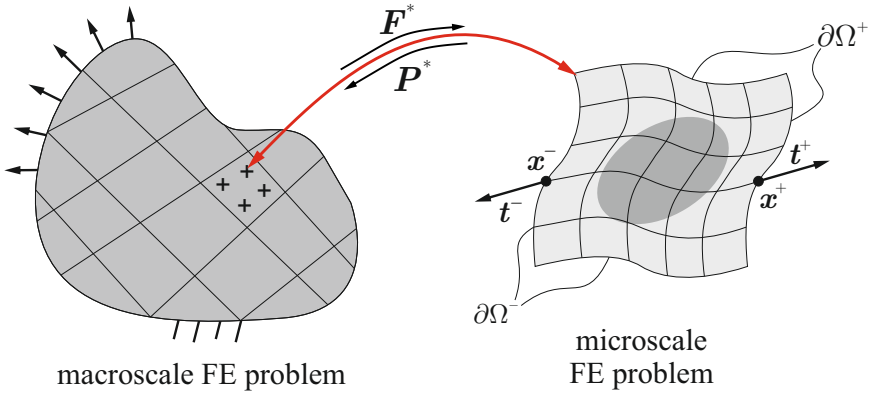


Figure 8.1: Schematic illustration of the FE^2 method, which solves a finite element BVP on the macroscale, whose local constitutive response is obtained from solving a nested finite element BVP on the microscale.

This approach is in general *computationally expensive*, since it requires the solution not only of the macroscale FE problem, but it also requires to solve a nested microscale FE problem every single time that a stress-strain relation is required on the macroscale (and for nonlinear problems with iterative solvers, this leads to a very large number of nested FE calculations). One additional complication arises, since solving the macro-problem iteratively may require the *tangent matrix* of the material model to be used at the quadrature point-level, i.e., we need

$$\boxed{\mathbb{C}^* = \frac{\partial \boldsymbol{\sigma}^*}{\partial \boldsymbol{\varepsilon}^*} \qquad \text{or} \qquad \mathbb{C}^* = \frac{\partial \boldsymbol{P}^*}{\partial \boldsymbol{F}^*}} \tag{8.1}$$

One possibility is to compute $\mathbb{C}^*$ as the **numerical tangent** (i.e., perturb each component of $\boldsymbol{\varepsilon}^*$ or $\boldsymbol{F}^*$ and recompute the stress, to then apply a finite-difference approximation of the derivative). This is unfortunately prone to causing numerical errors and computationally slow and costly, but it is generally easy to implement and (almost) always applicable.

As an alternative, a general strategy is to decompose all nodes into interior (i) and boundary (b) nodes of the RVE, so that displacements and corresponding nodal forces and tangent matrix can be rearranged by grouping interior and boundary nodes according to

$$\boldsymbol{U} = \begin{pmatrix} \boldsymbol{U}_\mathrm{i} \\ \boldsymbol{U}_\mathrm{b} \end{pmatrix}, \quad \boldsymbol{F} = \begin{pmatrix} \boldsymbol{F}_\mathrm{i} \\ \boldsymbol{F}_\mathrm{b} \end{pmatrix}, \qquad \Rightarrow \qquad \boldsymbol{T} = \frac{\partial \boldsymbol{F}_\mathrm{int}}{\partial \boldsymbol{U}^h} = \begin{pmatrix} \boldsymbol{K}_\mathrm{ii} & \boldsymbol{K}_\mathrm{ib} \\ \boldsymbol{K}_\mathrm{bi} & \boldsymbol{K}_\mathrm{bb} \end{pmatrix}, \tag{8.2}$$

where $\boldsymbol{T}$ denotes the global consistent stiffness matrix of the FE problem at the RVE level. Since no interior forces are being applied, we have $\boldsymbol{F}_\mathrm{i} = \boldsymbol{0}$. Also, by definition we must have

symmetry so that $\boldsymbol{K}_{\text{ib}} = \boldsymbol{K}_{\text{bi}}$. As an example of how this works in practice, let us discuss the case of affine BCs below.

8.1 Affine BCs

Consider affine BCs applied to an RVE with an average deformation gradient $\boldsymbol{F}^*$. Assume that there are n_b nodes on the boundary in d dimensions, so that the boundary displacement vector $\boldsymbol{U}_{\text{b}} \in \mathbb{R}^{dn_b}$ can be formulated as

$$\boldsymbol{U}_{\text{b}} = \boldsymbol{D}\boldsymbol{E}^*, \tag{8.3}$$

where we introduced $\boldsymbol{E}^* \in \mathbb{R}^{d^2}$ as the vector containing all components of $(\boldsymbol{F}^* - \boldsymbol{I})$ (simply listing all components, i.e., $[\boldsymbol{E}^*] = (F^*_{11} - 1, F^*_{12}, F^*_{21}, F^*_{22} - 1, \ldots)$). $\boldsymbol{D}$ is a matrix that depends on the boundary node locations $\boldsymbol{X}$ such that (8.3) implements the BC $\boldsymbol{u} = (\boldsymbol{F}^* - \boldsymbol{I})\boldsymbol{X}$ for all boundary nodes. The governing equations are thus summarized as

$$\left[\begin{aligned} \boldsymbol{F}_{\text{i}}(\boldsymbol{U}^h) &= \boldsymbol{0}, \\ \boldsymbol{F}_{\text{b}}(\boldsymbol{U}^h) - \boldsymbol{\Xi} &= \boldsymbol{0}, \\ \boldsymbol{U}_{\text{b}} - \boldsymbol{D}\boldsymbol{E}^* &= \boldsymbol{0} \end{aligned}\right. \tag{8.4}$$

with some unknown boundary forces summarized in vector $\boldsymbol{\Xi}$. Let us linearize this system of equations with respect to the unknowns $\{\boldsymbol{U}^h, \boldsymbol{\Xi}\}$ about an equilibrium solution $\{\boldsymbol{U}^h_0, \boldsymbol{\Xi}_0\}$ (e.g., the vectors of nodal displacements and forces obtained from solving for equilibrium inside the RVE for a given $\boldsymbol{F}^*$). A small perturbation $\Delta\boldsymbol{F}^*$ of the deformation gradient should result in small perturbations $\Delta\boldsymbol{U}_{\text{i}}$ and $\Delta\boldsymbol{U}_{\text{b}}$ of the interior and boundary node displacements, respectively. Applying Taylor expansions up to linear order to all three of the above equations with those perturbations about $\{\boldsymbol{U}^h_0, \boldsymbol{\Xi}_0\}$ yields

$$\left[\begin{aligned} \boldsymbol{F}_{\text{i}}(\boldsymbol{U}^h_0) + \boldsymbol{K}_{\text{ii}}(\boldsymbol{U}^h_0)\Delta\boldsymbol{U}_{\text{i}} + \boldsymbol{K}_{\text{ib}}(\boldsymbol{U}^h_0)\Delta\boldsymbol{U}_{\text{b}} &= \boldsymbol{0}, \\ \boldsymbol{F}_{\text{b}}(\boldsymbol{U}^h_0) - \boldsymbol{\Xi}_0 + \boldsymbol{K}_{\text{bi}}(\boldsymbol{U}^h_0)\Delta\boldsymbol{U}_{\text{i}} + \boldsymbol{K}_{\text{bb}}(\boldsymbol{U}^h_0)\Delta\boldsymbol{U}_{\text{b}} - \Delta\boldsymbol{\Xi} &= \boldsymbol{0}, \\ \boldsymbol{U}_{\text{b},0} - \boldsymbol{D}\boldsymbol{E}^* + \Delta\boldsymbol{U}_{\text{b}} - \boldsymbol{D}\Delta\boldsymbol{E}^* &= \boldsymbol{0}. \end{aligned}\right. \tag{8.5}$$

As we start with an equilibrium solution $\{\boldsymbol{U}^h_0, \boldsymbol{\Xi}_0\}$ satisfying (8.4), perturbations must satisfy

$$\Delta\boldsymbol{U}_{\text{b}} - \boldsymbol{D}\,\Delta\boldsymbol{E}^* = \boldsymbol{0}. \tag{8.6}$$

Moreover, in equilibrium we know that $\boldsymbol{F}_{\text{i}}(\boldsymbol{U}^h_0) = \boldsymbol{0}$ and $\boldsymbol{F}_{\text{b}}(\boldsymbol{U}^h_0) = \boldsymbol{\Xi}_0$, so that the first equation yields

$$\Delta\boldsymbol{U}_{\text{i}} = -\boldsymbol{K}^{-1}_{\text{ii}}(\boldsymbol{U}^h_0)\boldsymbol{K}_{\text{ib}}(\boldsymbol{U}^h_0)\Delta\boldsymbol{U}_{\text{b}}, \tag{8.7}$$

while the second equation leads to

$$\boldsymbol{K}_{\text{bi}}(\boldsymbol{U}^h_0)\Delta\boldsymbol{U}_{\text{i}} + \boldsymbol{K}_{\text{bb}}(\boldsymbol{U}^h_0)\Delta\boldsymbol{U}_{\text{b}} - \Delta\boldsymbol{\Xi} = \boldsymbol{0}. \tag{8.8}$$

Altogether, inserting (8.7) and (8.6) into (8.8) gives

$$\begin{aligned} \Delta\boldsymbol{\Xi} &= \left[\boldsymbol{K}_{\text{bb}}(\boldsymbol{U}^h_0) - \boldsymbol{K}_{\text{bi}}(\boldsymbol{U}^h_0)\boldsymbol{K}^{-1}_{\text{ii}}(\boldsymbol{U}^h_0)\boldsymbol{K}_{\text{ib}}(\boldsymbol{U}^h_0)\right]\Delta\boldsymbol{U}_{\text{b}} \\ &= \left[\boldsymbol{K}_{\text{bb}}(\boldsymbol{U}^h_0) - \boldsymbol{K}_{\text{bi}}(\boldsymbol{U}^h_0)\boldsymbol{K}^{-1}_{\text{ii}}(\boldsymbol{U}^h_0)\boldsymbol{K}_{\text{ib}}(\boldsymbol{U}^h_0)\right]\boldsymbol{D}\Delta\boldsymbol{E}^*. \end{aligned} \tag{8.9}$$

Finally, recall that, following (3.14), the average RVE stress tensor in the FE setting can be obtained from (where the sum over a is over the boundary nodes only)

$$\boldsymbol{P}^* = \frac{1}{V}\sum_a \boldsymbol{F}^a_{\text{b}} \otimes \boldsymbol{X}^a_{\text{b}} = \frac{1}{V}\boldsymbol{D}^{\text{T}}\boldsymbol{F}_{\text{b}}, \tag{8.10}$$

where we reused the transformation matrix $\boldsymbol{D}$ introduced above (which depends on the undeformed nodal locations on the RVE boundary) and, with an abuse of notation, we write the stress tensor in analogous form to $\boldsymbol{E}^*$ in vector form as $\boldsymbol{P}^* = (P_{11}, P_{12}, \dots, P_{33})$. Linearization of (8.10) for small perturbations $\Delta\boldsymbol{F}^*$ (and hence $\Delta\boldsymbol{E}^*$ as above) and the resulting small perturbations in the nodal forces results in

$$\boldsymbol{P}^* + \mathbb{C}^*\Delta\boldsymbol{E}^* = \frac{1}{V}\left[\boldsymbol{D}^{\mathrm{T}}\boldsymbol{F}_{\mathrm{b}}(\boldsymbol{U}_0^h) + \boldsymbol{D}^{\mathrm{T}}\Delta\boldsymbol{\Xi}\right], \tag{8.11}$$

so that the perturbations are linked via

$$\mathbb{C}^*\Delta\boldsymbol{E}^* = \frac{1}{V}\boldsymbol{D}^{\mathrm{T}}\Delta\boldsymbol{\Xi}. \tag{8.12}$$

Insertion of (8.9) finally admits the conclusion

$$\boxed{\mathbb{C}^* = \frac{\partial\boldsymbol{P}^*}{\partial\boldsymbol{E}^*} = \frac{1}{V}\boldsymbol{D}^{\mathrm{T}}\left[\boldsymbol{K}_{\mathrm{bb}}(\boldsymbol{U}_0^h) - \boldsymbol{K}_{\mathrm{bi}}(\boldsymbol{U}_0^h)\boldsymbol{K}_{\mathrm{ii}}^{-1}(\boldsymbol{U}_0^h)\boldsymbol{K}_{\mathrm{ib}}(\boldsymbol{U}_0^h)\right]\boldsymbol{D}} \tag{8.13}$$

That is, the consistent tangent matrix $\mathbb{C}^* = \partial\boldsymbol{P}^*/\partial\boldsymbol{E}^*$ can be obtained directly from the element stiffness matrices, once one load step has been equilibrated. That is, once a solution is found for the microscale BVP in the RVE, the (tangent) stiffness matrices in that equilibrium state uniquely define the effective tangent stiffness, and there is no need for numerical derivatives. Note that because $\boldsymbol{K}_{\mathrm{ii}}$ is constructed from only the inner nodes, it is invertible by definition. Further, note that we wrote both $\boldsymbol{P}^*$ and $\boldsymbol{E}^*$ in vector form (Voigt notation), so the resulting $\mathbb{C}^*$ in (8.13) is the consistent tangent based on those two quantities, which can easily be converted into the classical matrix or tensor form.

8.2 Uniform Tractions, Periodic BCs

Similar procedures can be applied to uniform traction BCs and periodic BCs. For *uniform traction BCs*, the procedure is analogous but may require iterations (for nonlinear problems). For *periodic BCs*, the procedure is more involved but in principle analogous. We skip the details here, which can be found, e.g., in reference [63] for the setting of linearized kinematics.

9 Analytical Techniques: Cauchy-Born Kinematics

The above multiscale techniques involved solving a BVP on the microscale to extract the effective material behavior on the macroscale – with the only exception of applying affine displacements across the entire RVE (Section 6.1). Even though the latter provided an effective response that was shown to be overly stiff, under certain conditions one may want to formulate such simple microscale problems that do not necessarily require solving a microscale BVP but admit analytical solutions for average microscale quantities (yet not as simple as linear elasticity). Let us here discuss the family of models based on so-called **Cauchy-Born** kinematics as one such example (which will become helpful later when discussing coarse-grained atomistics).

9.1 Local Cauchy-Born Kinematics

As before, we assume a spatial separation of scales between micro- and macroscale features. On the macroscale, body Ω is deforming according to a deformation mapping $\varphi : \Omega \to \mathbb{R}^d$. Within an infinitesimally small neighborhood of a point $\boldsymbol{X}_0 \in \Omega$, we may expand the deformation (using index notation) as

$$\begin{aligned}\varphi_i(\boldsymbol{X}) &= \varphi_i(\boldsymbol{X}_0) + \frac{\partial\varphi_i}{\partial X_J}(\boldsymbol{X}_0)\,\mathrm{d}X_J + \frac{1}{2}\frac{\partial^2\varphi_i}{\partial X_J\,\partial X_K}(\boldsymbol{X}_0)\,\mathrm{d}X_J\,\mathrm{d}X_K + \text{h.o.t.}\\ &= \varphi_i(\boldsymbol{X}_0) + F_{iJ}(\boldsymbol{X}_0)\,\mathrm{d}X_J + \frac{1}{2}F_{iJ,K}(\boldsymbol{X}_0)\,\mathrm{d}X_J\,\mathrm{d}X_K + \text{h.o.t.},\end{aligned} \tag{9.1}$$

with deformation gradient components F_{iJ}, second deformation gradient $(\mathrm{Grad}\,\boldsymbol{F})_{iJK} = F_{iJ,K}$ (a third-order tensor), and higher-order terms continued analogously. Hence, if the deformation at a point $\boldsymbol{X}_0 \in \Omega$ – characterized by $\boldsymbol{\varphi}(\boldsymbol{X}_0)$, $\boldsymbol{F}(\boldsymbol{X}_0)$, $\mathrm{Grad}\,\boldsymbol{F}(\boldsymbol{X}_0)$, etc. – is known, then the deformation in its infinitesimal neighborhood can be obtained from the above Taylor expansion (see Figure 9.1). Dropping higher-order terms by assuming that the neighborhood is infinitesimally small ($|\mathrm{d}\boldsymbol{X}| \ll 1$) results in an approximation.

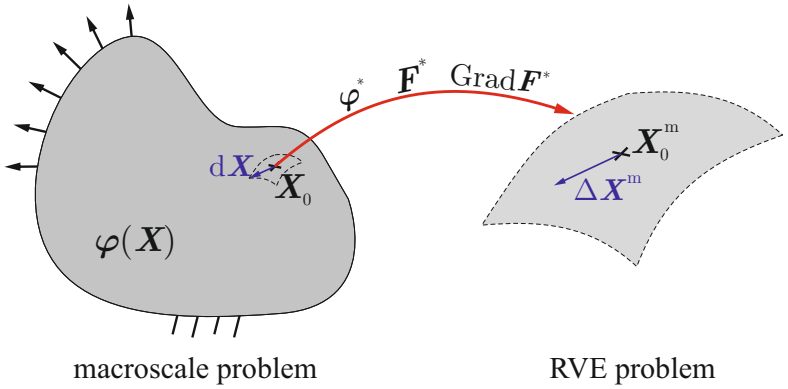


Figure 9.1: Schematic illustration of the Taylor expansion applied to the deformation within an infinitesimal neighborhood of a point $\boldsymbol{X}_0$ on the macroscale and to a (finite-size but sufficiently small) RVE on the microscale.

Now, consider a microstructure that admits a separation of scales from the macroscale. In this case, one may regard (9.1) as an approximation of the deformation across an RVE centered at $\boldsymbol{X}_0$, so that infinitesimal distances $\mathrm{d}\boldsymbol{X}$ are replaced by finite but small microscale distances $\Delta\boldsymbol{X}^{\mathrm{m}} = \boldsymbol{X}^{\mathrm{m}} - \boldsymbol{X}_0$ for any point $\boldsymbol{X}^{\mathrm{m}}$ inside the RVE. Here and in the following, we will denote microscale quantities with superscript $^{\mathrm{m}}$ to differentiate those from the macroscale ones. We further denote the effective deformation metrics passed between micro- and macroscales

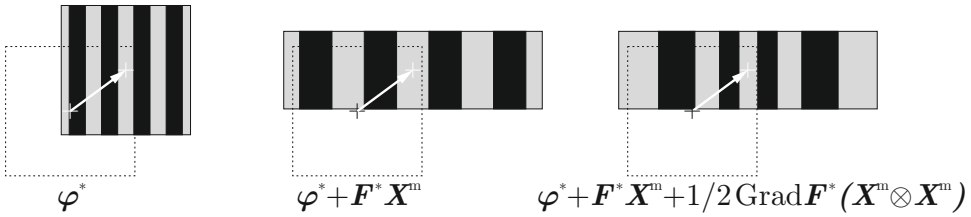


Figure 9.2: Illustration of the individual terms in the Taylor expansion (9.2): the first term represents pure translation, the second term affine straining (constant deformation gradient), the third term linear straining (linearly varying deformation gradient), etc.

as $\boldsymbol{\varphi}^* = \boldsymbol{\varphi}(\boldsymbol{X}_0)$ and $\boldsymbol{F}^* = \boldsymbol{F}(\boldsymbol{X}_0)$. Consequently, if $\boldsymbol{X}_0 = \boldsymbol{X}_0^{\mathrm{m}}$ is identified as the center of the RVE, the deformation of any point $\boldsymbol{X}^{\mathrm{m}}$ in the RVE (see Figures 9.1 and 9.2) can be approximated by

$$\varphi_i(\boldsymbol{X}^{\mathrm{m}}) \approx \varphi_i^* + F_{iJ}^* \Delta X_J^{\mathrm{m}} + \frac{1}{2} F_{iJ,K}^* \Delta X_J^{\mathrm{m}} \, \Delta X_k^{\mathrm{m}} + \text{h.o.t.} \tag{9.2}$$

The simplest approximation is obtained by dropping all but the linear term, which results in the so-called (historically not quite correct, as will discuss in Section 9.4) **local Cauchy-Born rule**:

$$\boxed{\boldsymbol{\varphi}(\boldsymbol{X}^{\mathrm{m}}) = \boldsymbol{\varphi}^* + \boldsymbol{F}^* \Delta \boldsymbol{X}^{\mathrm{m}}, \qquad \Delta \boldsymbol{X}^{\mathrm{m}} = \boldsymbol{X}^{\mathrm{m}} - \boldsymbol{X}_0^{\mathrm{m}}} \tag{9.3}$$

where $\boldsymbol{X}_0^{\mathrm{m}}$ is a reference point on the RVE-level (e.g., the center of mass of the RVE). Obviously, (9.3) prescribes an *affine deformation* with constant deformation gradient $\boldsymbol{F}^* = \boldsymbol{F}(\boldsymbol{X}_0)$ across the entire RVE, as previously discussed in Section 6.1. Note that for *quasistatic* problems the constant term $\varphi_i^* = \varphi_i(\boldsymbol{X}_0)$ constitutes rigid body motion of the RVE, so that omitting this term (or, equivalently, arbitrarily choosing $\boldsymbol{X}_0$ at the RVE-level) does not affect the deformation of the RVE. In such a case, we may as well drop the leading-order term and arbitrarily choose the RVE-level coordinate origin such that $\boldsymbol{X}_0^{\mathrm{m}} = \boldsymbol{0}$, which leads to the simplified version

$$\boxed{\boldsymbol{\varphi}(\boldsymbol{X}^{\mathrm{m}}) = \boldsymbol{F}^* \boldsymbol{X}^{\mathrm{m}}} \tag{9.4}$$

Assume now that the material in the RVE has a strain energy density $W(\boldsymbol{F}, \boldsymbol{X})$, then the above approximation results in the average energy density of the RVE being (for simplicity choosing $\boldsymbol{X}_0 = \boldsymbol{0}$)

$$W^*(\boldsymbol{F}^*) = \frac{1}{V} \int_\Omega W(\boldsymbol{F}^*, \boldsymbol{X}) \, \mathrm{d}V = \langle W(\boldsymbol{F}^*, \boldsymbol{X}) \rangle. \tag{9.5}$$

This is equivalent to the Taylor construction of Section 6.1. The key point here is that this energy density allows for an analytical calculation of the stress-strain response. Specifically, the average first Piola-Kirchhoff stress tensor can be obtained in closed form by differentiation:

$$\boldsymbol{P}^* = \frac{\partial W^*}{\partial \boldsymbol{F}^*}(\boldsymbol{F}^*) = \langle \frac{\partial W}{\partial \boldsymbol{F}}(\boldsymbol{F}^*, \boldsymbol{X}) \rangle = \langle \boldsymbol{P}(\boldsymbol{F}^*, \boldsymbol{X}) \rangle. \tag{9.6}$$

Analogously, the incremental stiffness tensor $\mathbb{C}^*$ follows from averaging:

$$\mathbb{C}^*(\boldsymbol{F}^*) = \frac{\partial \boldsymbol{P}^*}{\partial \boldsymbol{F}^*}(\boldsymbol{F}^*) = \frac{\partial}{\partial \boldsymbol{F}^*} \langle \boldsymbol{P}(\boldsymbol{F}^*, \boldsymbol{X}) \rangle = \langle \frac{\partial \boldsymbol{P}}{\partial \boldsymbol{F}}(\boldsymbol{F}^*, \boldsymbol{X}) \rangle = \langle \mathbb{C}(\boldsymbol{F}^*, \boldsymbol{X}) \rangle. \tag{9.7}$$

Of course, the obtained behavior agrees with the *Taylor construction* discussed in Section 6.1. Consequently, one may expect the effective response to be overly stiff. Yet, one can derive a closed-form analytical expression for the effective stresses and stiffness. This is numerically beneficial compared to, e.g., FE^2 (see Section 8), which is computationally expensive due to the on-the-fly numerical calculation of stresses and stiffness. In terms of accuracy, of course, one can do better, which is why we next discuss higher-order expansions.

9.2 Nonlocal Cauchy-Born Kinematics

Let us take the above concepts a step further. The so-called **extended Cauchy-Born rule** (also known as **nonlocal Cauchy-Born rule**) retains one more term in the Taylor expansion, resulting in the approximate deformation of a point $\boldsymbol{X}^\mathrm{m}$ within the RVE on the microscale (and again dropping the leading-order term relating to rigid body motion and choosing the center of the RVE as the coordinate origin)

$$\boxed{\varphi_i(\boldsymbol{X}^\mathrm{m}) = F^*_{iJ} X^\mathrm{m}_J + \frac{1}{2} F^*_{iJ,K} X^\mathrm{m}_J X^\mathrm{m}_K} \tag{9.8}$$

Here, the RVE deformation is no longer affine but experiences a linearly-varying deformation gradient (see Figure 9.2) since

$$F_{iJ}(\boldsymbol{X}^\mathrm{m}) = \frac{\partial \varphi_i}{\partial X_J}(\boldsymbol{X}^\mathrm{m}) = F^*_{iJ} + F^*_{iJ,K} X^\mathrm{m}_K, \tag{9.9}$$

where we exploited that $F_{iJ,K} = \varphi_{i,JK} = \varphi_{i,KJ} = F_{iK,J} = (\mathrm{Grad}\, \boldsymbol{F})_{iJK}$.

If the material inside the RVE has a strain energy density $W(\boldsymbol{F}, \boldsymbol{X})$, then the effective energy density becomes

$$W^*(\boldsymbol{F}^*, \mathrm{Grad}\, \boldsymbol{F}^*) = \frac{1}{V} \int_\Omega W(\boldsymbol{F}^* + (\mathrm{Grad}\, \boldsymbol{F}^*)\boldsymbol{X}^\mathrm{m}, \boldsymbol{X}^\mathrm{m})\, \mathrm{d}V = \langle W(\boldsymbol{F}^* + \mathrm{Grad}\, \boldsymbol{F}^* \boldsymbol{X}^\mathrm{m}, \boldsymbol{X}^\mathrm{m}) \rangle. \tag{9.10}$$

The result is an effective **nonlocal model**, whose energy density depends on both $\boldsymbol{F}^*$ and $\mathrm{Grad}\, \boldsymbol{F}^*$ on the macroscale. This has important consequences for the effective material behavior, since the constitutive law no longer depends on local information only (i.e., knowing the strains at a point is insufficient to define the resulting stress – required in addition is the gradient of strains). A further consequence is the emergence of size effects when using this model on the macroscale (important for instabilities and localization or damage), but this is a story for another course.

The resulting first Piola-Kirchhoff stress tensor is by definition

$$P^*_{iJ} = \frac{\partial W^*}{\partial F^*_{iJ}}(\boldsymbol{F}^*) = \langle P_{iJ}(\boldsymbol{F}^* + (\mathrm{Grad}\, \boldsymbol{F}^*)\boldsymbol{X}^\mathrm{m}, \boldsymbol{X}^\mathrm{m}) \rangle\,, \tag{9.11}$$

and there is a third-order couple stress tensor

$$\begin{aligned} Y^*_{iJK} &= \frac{\partial W^*}{\partial F^*_{iJ,K}}(\boldsymbol{F}^*) = \frac{1}{V} \int_\Omega \frac{\partial}{\partial F_{kL}} W(\boldsymbol{F}^* + (\mathrm{Grad}\, \boldsymbol{F}^*)\boldsymbol{X}^\mathrm{m}, \boldsymbol{X}^\mathrm{m}) \frac{\partial F^*_{kL} + F^*_{kL,M} X^\mathrm{m}_M}{\partial F^*_{iJ,K}}\, \mathrm{d}V \\ &= \langle P_{iJ}(\boldsymbol{F}^* + (\mathrm{Grad}\, \boldsymbol{F}^*)\boldsymbol{X}^\mathrm{m}, \boldsymbol{X}^\mathrm{m}) X^\mathrm{m}_K \rangle. \end{aligned} \tag{9.12}$$

Note that also the macroscale variational problem must be revised, since the energy density is no longer local. This results in added complexity and requires reformulation in particular of boundary conditions and variations. For example, consider a total potential energy functional of a BVP on the macroscale (dropping the asterisks for convenience and omitting work done by external forces for conciseness) in its general form

$$I[\boldsymbol{\varphi}] = \int_\Omega W(\boldsymbol{\varphi}, \boldsymbol{F}, \mathrm{Grad}\, \boldsymbol{F}, \mathrm{Grad}\, \mathrm{Grad}\, \boldsymbol{F}, \ldots)\, \mathrm{d}V. \tag{9.13}$$

Its first variation reads

$$\delta I[\boldsymbol{\varphi}] = \int_\Omega \left(\frac{\partial W}{\partial \varphi_i}\delta\varphi_i + \frac{\partial W}{\partial F_{iJ}}\delta\varphi_{i,J} + \frac{\partial W}{\partial F_{iJ,K}}\delta\varphi_{i,JK} + \dots \right) \mathrm{d}V, \tag{9.14}$$

where the higher-order stresses arise naturally. Using the usual discretization by a Bubnov-Galerkin FE scheme, the internal force vector on node a (ith component) becomes

$$F^a_{\mathrm{int},i} = \int_\Omega \left(\frac{\partial W}{\partial \varphi_i}N^a + P_{iJ}N^a_{,J} + Y_{iJK}N^a_{,JK} + \dots \right) \mathrm{d}V. \tag{9.15}$$

Here we see the higher-order stress tensors enter again. This further requires that we must use higher-order FE interpolation. For example, using the above formulation with $W = W(\boldsymbol{F}, \mathrm{Grad}\,\boldsymbol{F})$, we need to compute second derivatives of shape functions (cf. $N^a_{,JK}$), so those must be defined. This is generally a critical issue, since simply using quadratic shape functions does not solve the problem, as long as the interpolated strains are discontinuous across element boundaries. Instead, one needs an interpolation that ensures that $\boldsymbol{\varphi} \in H^2(\Omega)$, so that the second derivatives of shape functions are defined, non-zero in general, and square-integrable. Alternatively, one may apply the divergence theorem in order to successively remove the higher-order variational terms. This, however, introduces cumbersome boundary terms to account for. We will not discuss this problem here in detail (see, e.g., the rich literature on **gradient elasticity** theories).

Of course, the above construction can be extended to arbitrary orders of gradients by retaining more and more terms in the Taylor expansion. This leads to more and more accurate (but also more and more complex) material models. Ultimately, the choice of terms to retain results from the competition between accuracy and efficiency and depends on the problem to be solved on the macroscale. For many problems, local kinematics (Section 9.1) are sufficient. If nonlocality is required (for problems involving localization, e.g.), then retaining second gradients is usually sufficient. Of course, it all hinges upon a reasonable separation of scales between RVE and macroscale BVP and a suitable choice of the RVE.

9.3 Application to Discrete Systems

Among others, the above Cauchy-Born-based strategy can be applied to discrete systems, where it has become a powerful tool to connect discrete and continuum length scales (the Arruda-Boyce model being one example, in which we linked discrete polymer chains to a continuum deformation gradient; see Section 9.1). Imagine, e.g., an RVE filled with a number of atomic particles, or an RVE filled with discrete beams forming a truss network. In such cases, the RVE is best described by a finite set of discrete nodes that are mechanically connected, whereas the macroscale may require a continuum constitutive model to be used, e.g., within an FE simulation. In such a scenario, Cauchy-Born kinematics can be beneficial, as discussed here.

For a discrete RVE, the expansion (9.2) is applied to the position of each discrete point that is tracked inside the RVE (e.g., the position of each atom or the position of junctions in a truss). In this case, the unit cell is defined by a discrete point set at undeformed positions $\{\boldsymbol{X}_1, \dots, \boldsymbol{X}_n\}$ and a suitable connectivity in between these n points. For example, in case of stretching-dominated trusses (where each strut behaves like a geometrically linear or nonlinear spring) the energy of the RVE may be formulated via a pair potential $V(r)$ depending on the distance r between any two nodes (or atoms). If we approximate the RVE deformation by the local Cauchy-Born approximation, then the energy density of the RVE becomes

$$W^*(\boldsymbol{F}^*) = \frac{1}{2V}\sum_{i=1}^{n}\sum_{j\neq i} V\left(r_{ij}\right) \qquad \text{with} \qquad r_{ij} = \left\|\boldsymbol{x}^{\mathrm{m}}_i - \boldsymbol{x}^{\mathrm{m}}_j\right\|, \quad \boldsymbol{x}^{\mathrm{m}}_i = \boldsymbol{F}^*\boldsymbol{X}^{\mathrm{m}}_i \tag{9.16}$$

or, altogether,

$$W^*(\boldsymbol{F}^*) = \frac{1}{2V}\sum_{i=1}^{n}\sum_{j\neq i} V\left(\left\|\boldsymbol{F}^*\boldsymbol{X}_i^{\mathrm{m}} - \boldsymbol{F}^*\boldsymbol{X}_j^{\mathrm{m}}\right\|\right). \tag{9.17}$$

We here sum the energy of each connection, which in turn depends on the (change of) length of the connection. Note that a factor of $1/2$ enters to avoid double counting of each connection. For generality, we here assume that each node i is connected to all $n-1$ other nodes j. In practice, interactions are usually *short-range* (i.e., only involving some selection of nearest neighbor), which can easily be dealt with by assuming that potential $V(r)$ has a **cut-off distance** (i.e., it goes to zero beyond a certain distance). An illustration of such a scenario for discrete particles is shown in Figure 9.3.

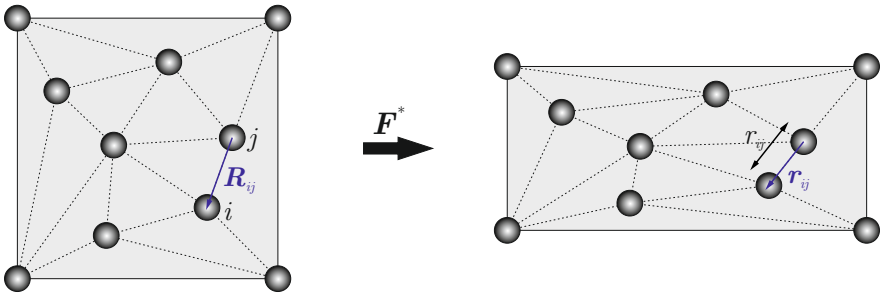


Figure 9.3: Discete particles at deformed distances r_{ij} within an RVE that is deformed affinely with an average deformation gradient $\boldsymbol{F}^*$.

From (9.17), the first Piola-Kirchhoff stress tensor follows as

$$\begin{aligned}
\boldsymbol{P}^*(\boldsymbol{F}^*) &= \frac{\partial W^*}{\partial \boldsymbol{F}^*}(\boldsymbol{F}^*) = \frac{1}{2V}\sum_{i=1}^{n}\sum_{j\neq i} V'(r_{ij})\,\frac{\partial\left\|\boldsymbol{F}^*\boldsymbol{X}_i - \boldsymbol{F}^*\boldsymbol{X}_j\right\|}{\partial \boldsymbol{F}^*} \\
&= \frac{1}{2V}\sum_{i=1}^{n}\sum_{j\neq i} V'(r_{ij})\,\frac{\boldsymbol{x}_i - \boldsymbol{x}_j}{\left\|\boldsymbol{F}^*\boldsymbol{X}_i - \boldsymbol{F}^*\boldsymbol{X}_j\right\|}\otimes(\boldsymbol{X}_i - \boldsymbol{X}_j) \\
&= \frac{1}{2V}\sum_{i=1}^{n}\sum_{j\neq i} V'(r_{ij})\,\hat{\boldsymbol{r}}_{ij}\otimes\boldsymbol{R}_{ij} = \frac{1}{2V}\sum_{i=1}^{n}\sum_{j\neq i}\boldsymbol{f}_{ij}\otimes\boldsymbol{R}_{ij},
\end{aligned} \tag{9.18}$$

where the deformed unit direction of a connection (or *bond*) ij and the undeformed distance vector were introduced as, respectively

$$\hat{\boldsymbol{r}}_{ij} = \frac{\boldsymbol{x}_i - \boldsymbol{x}_j}{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|} = \frac{\boldsymbol{x}_i - \boldsymbol{x}_j}{r_{ij}}, \qquad \boldsymbol{R}_{ij} = \boldsymbol{X}_i - \boldsymbol{X}_j \tag{9.19}$$

and

$$\boldsymbol{f}_{ij} = V'(r_{ij})\,\hat{\boldsymbol{r}}_{ij} \tag{9.20}$$

is the force acting between nodes i and j in the direction of the link in between the nodes. This allows us to compute the effective constitutive relation $\boldsymbol{P}^* = \boldsymbol{P}^*(\boldsymbol{F}^*)$ analytically in a straight-forward fashion:

$$\boxed{\boldsymbol{P}^*(\boldsymbol{F}^*) = \frac{1}{2V}\sum_{i=1}^{n}\sum_{j\neq i}\boldsymbol{f}_{ij}\otimes\boldsymbol{R}_{ij}} \tag{9.21}$$

The effective stiffness tensor can be derived analogously (but is omitted here for conciseness).

Of course, the assumed affine deformation of the RVE is a strong simplification leading to an overly stiff response in general, yet the simple structure and analytical derivatives make for a powerful tool to extract an effective response from a discrete RVE. This approach forms the basis, e.g., for the so-called *quasicontinuum method* to bridge from atomistics to the continuum scales (which we will discuss later).

A further example of this approach is the so-called **Arruda-Boyce** model named after its inventors, see reference [6], and also known as the **eight-chain model**. This is a hyperelastic constitutive model commonly used to describe the mechanical response of rubbers and polymers. Leaving all statistical-mechanics complications aside, the model is based on a cubic RVE containing eight polymer chains oriented along each of the room diagonals in the cubic RVE, including a total of eight connections ij, which connect the center of the cube to each of the eight corners (see Figure 9.4). The mechanical stretch of each chain is related to its stored energy density $V(r_{ij})$, and the energy V stored in a given chain as a function of its elongation can be obtained from statistical mechanics. Application of the affine deformation from (9.21) is exploited to extract the effective material response, which yields an effective energy density W^* of the homogenized hyperelastic material as (assuming the coordinate origin is at the center of the cubic RVE with side length L)

$$W^*(\boldsymbol{F}^*) = \frac{1}{V}\sum_{i=1}^{8} V\left(\|\boldsymbol{F}^*\boldsymbol{X}_i^{\mathrm{m}}\|\right) = \frac{1}{L^3}\sum_{i=1}^{8} V\left(\sqrt{\boldsymbol{X}_i^{\mathrm{m}}\cdot(\boldsymbol{F}^*)^{\mathrm{T}}\boldsymbol{F}^*\boldsymbol{X}_i^{\mathrm{m}}}\right). \tag{9.22}$$

Inserting the appropriate form for V and simplification leads to

$$W(\boldsymbol{F}^*) = C_1\sum_{i=1}^{5}\alpha_i\beta^{2(i-1)}(I_1^i - 3^i), \tag{9.23}$$

where C_1 is a calibration parameter, α_i are known constants, $I_1 = \mathrm{tr}[(\boldsymbol{F}^*)^{\mathrm{T}}\boldsymbol{F}^*]$ is the first invariant, and $\beta = 1/N$ is related to the number N of chains in the network of a cross-linked polymer (we omit the derivation here for brevity). Of course, the approach does not always simplify to such neat hyperelastic expressions. Nevertheless, its analytical structure is oftentimes preferable over more accurate albeit more expensive numerical approaches.

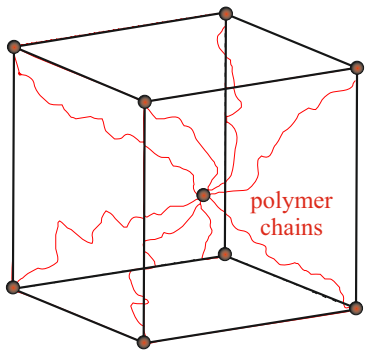


Figure 9.4: The Arruda-Boyce model is based on eight polymer chains connecting the corners of a cubic RVE to its center, assumed to undergoing affine deformation for the extraction of the effective response.

Note that this strategy can also be applied to a discrete **granular** RVE, where the interaction between static particles is described by the potential V (as written here in the special case without rotational/frictional contact forces).

When using this framework for **truss** lattices, one complication arises as one may need to introduce *rotational degrees of freedom* to describe the rotation of individual nodes (e.g., when using a beam description for the connections between nodes, whose bending requires the use

of rotational degrees of freedom, see Figure 9.5). In such a scenario, the RVE must introduce rotational degrees of freedom $\boldsymbol{\theta}$ at all nodes (with one component in 2D, and three in 3D). While we can still couple the *positions* of nodes to the macroscale via Cauchy-Born kinematics, e.g., assuming

$$\boldsymbol{\varphi}(\boldsymbol{X}^{\mathrm{m}}) = \boldsymbol{\varphi}^* + \boldsymbol{F}^*\boldsymbol{X}^{\mathrm{m}} + \text{h.o.t.}, \tag{9.24}$$

we must define how to solve for the rotations of RVE nodes. There are two commonly used approaches. The first is to not couple the RVE rotations to the macroscale and to instead solve for those rotations by solving for equilibrium inside the RVE and assuming periodic rotations on the RVE boundary (thereby effectively imposing a zero average rotation). The second approach is to consider a macroscale problem described by a deformation mapping $\boldsymbol{\varphi}^* = \boldsymbol{\varphi}(\boldsymbol{X}_0)$ and a rotational field $\boldsymbol{\theta}^* = \boldsymbol{\theta}(\boldsymbol{X}_0)$, whose gradients are the deformation gradient $\boldsymbol{F}^* = \boldsymbol{F}(\boldsymbol{X}_0)$ and the curvature tensor $\boldsymbol{\kappa}^* = \boldsymbol{\kappa}(\boldsymbol{X}_0)$ with $\boldsymbol{\kappa} = \operatorname{Grad}\boldsymbol{\theta}$. Even though there is no clear physical interpretation of the rotation field on the macroscopic continuum level, its meaning becomes clear from the microscale, where the application of the Cauchy-Born rule to both fields yields

$$\boldsymbol{\varphi}(\boldsymbol{X}^{\mathrm{m}}) = \boldsymbol{\varphi}^* + \boldsymbol{F}^*\boldsymbol{X}^{\mathrm{m}} + \text{h.o.t.}, \qquad \boldsymbol{\theta}(\boldsymbol{X}^{\mathrm{m}}) = \boldsymbol{\theta}^* + \boldsymbol{\kappa}^*\boldsymbol{X}^{\mathrm{m}} + \text{h.o.t.}. \tag{9.25}$$

By again defining the effective energy density as the sum over all beams inside the RVE, we arrive at an effective, homogenized macroscale problem with the energy functional

$$I[\boldsymbol{\varphi}, \boldsymbol{\theta}] = \int_\Omega W^*(\boldsymbol{\varphi}^*, \boldsymbol{F}^*, \boldsymbol{\theta}^*, \boldsymbol{\kappa}^*)\,\mathrm{d}V - (\text{work done by external forces and torques}). \tag{9.26}$$

This also admits defining the effective "stresses"

$$\tau_i^* = \frac{\partial W^*}{\partial \varphi_i^*}, \qquad P_{iJ}^* = \frac{\partial W^*}{\partial F_{iJ}^*}, \qquad m_i^* = \frac{\partial W^*}{\partial \theta_i^*}, \qquad M_{iJ}^* = \frac{\partial W^*}{\partial \kappa_{iJ}^*}. \tag{9.27}$$

Notice that the dependence on $\boldsymbol{\varphi}^*$ can be dropped since it implies rigid-body translation of the RVE, which should not contribute to the stored energy for physical reasons (unless dynamics are to be considered). By contrast, $\boldsymbol{\theta}^*$ cannot be dropped, since $\boldsymbol{\theta}^*$ implies a uniform rotation of all nodes inside the RVE, which does *not* correspond to rigid body rotation. Also, special caution is required when defining the boundary terms in (9.26), as their physical interpretation may be intricate (and they also require specialized higher-order FE interpolation, as discussed in Section 9.2). This is an area of ongoing research, especially in the field of metamaterials and architected materials made of periodic truss architectures (see, e.g., references [29, 30]).

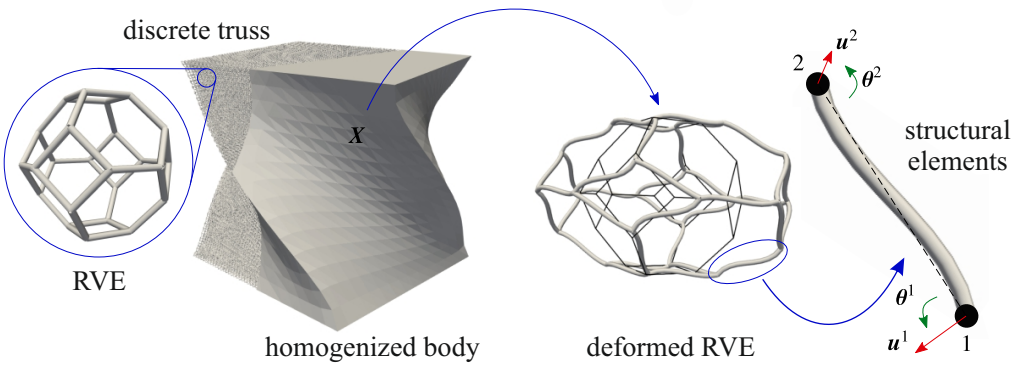


Figure 9.5: Homogenized response of periodic trusses from linking deformation $\boldsymbol{\varphi}$ and rotations $\boldsymbol{\theta}$ between the homogenized macroscale body and the structural RVE made of discrete beams.

9.4 On the Cauchy-Born Rule

The **Cauchy-Born rule** dates back to Augustin-Louis Cauchy, who developed the theory for the eponymous Cauchy solids, which was much later extended by Max Born. Strictly speaking, the above concept of applying (to leading order) a uniform strain to the entire RVE, so that all atoms or nodes inside the unit cell follow the macroscopic/average straining, only applies to a so-called **simple lattice**. If a truss RVE in 2D consists of, e.g., a simple square with four nodes at its corners and connections in between (the *square* unit cell), then the nodes form a **simple Bravais lattice** described by two Bravais basis vectors (see Section 10.1). The same applies to a *triangular* unit cell, composed of three interconnected nodes. In such cases of simple Bravais lattices, the assumption of a uniform deformation gradient is reasonable. Consider by contrast a *hexagonal* unit cell, composed of six nodes that are connected to form a hexagon. If such a truss unit cell made of slender beams is strained uniaxially as an example, deforming all nodes in an affine manner is an overly stiff assumption, since slender beams have a much higher stretching than bending stiffness. Here, a realistic unit cell response to on-average uniaxial straining is to engage beams in bending to the extent possible, while minimizing the stretching of beams. The result is clearly a non-uniform deformation gradient within the unit cell. The hexagon unit cell is a **complex Bravais lattice** or a **multi-lattice** (see Figure 9.7): its nodes do not lie on a simple Bravais lattice (the nodal locations cannot be described by two basis vectors only). Instead, one can find all nodes from the composition of two simple Bravais lattices (therefore, a *multi-lattice*).

For a multi-lattice, imposing an average deformation gradient $\boldsymbol{F}^*$ can be interpreted as imposing the average $\boldsymbol{F}^*$ for each of the multiple lattices (i.e., points on each lattice are stretched relative to each other by $\boldsymbol{F}^*$), while the lattices can show constant non-zero displacements (so called **shifts**) relative to each other. This translates into

$$\boldsymbol{\varphi}(\boldsymbol{X}^{\mathrm{m}}) = \boldsymbol{\varphi}^* + \boldsymbol{F}^*\boldsymbol{X}^{\mathrm{m}} + \sum_a \delta\boldsymbol{\varphi}_a \psi_a(\boldsymbol{X}^{\mathrm{m}}) + \text{h.o.t.}, \tag{9.28}$$

where $\delta\boldsymbol{\varphi}_a$ is the *shift* vector of lattice a with respect to a static frame of reference, and we introduced the indicator function

$$\psi_a(\boldsymbol{X}^{\mathrm{m}}) = \begin{cases} 1 & \text{if node } \boldsymbol{X}^{\mathrm{m}} \text{ lies in lattice } a, \\ 0 & \text{else.} \end{cases} \tag{9.29}$$

In this case, we still satisfy the average deformation, since the deformation at any node i is characterized by

$$\frac{\partial}{\partial \boldsymbol{X}} \left(\boldsymbol{\varphi}^* + \boldsymbol{F}^*\boldsymbol{X}_i + \delta\boldsymbol{\varphi}_i\right) = \boldsymbol{F}^*, \tag{9.30}$$

where $\delta\boldsymbol{\varphi}_i$ denotes the shifts of node i. Note that if one additionally requires $\langle\boldsymbol{\varphi}\rangle = \boldsymbol{\varphi}^*$ (e.g., in case of dynamic problems), then additional constraints on the shifts are required.

Since the shifts are independent of the macroscopic deformation, they must be solved at the RVE-level. For quasistatic problems it is reasonable that the shifts minimize the RVE energy, so they are obtained from an additional equilibrium relation. For example, re-considering the pairwise interaction formulation in (9.17), we here need to expand the latter to

$$W^*(\boldsymbol{F}^*) = \min_{\delta\boldsymbol{\varphi}_1,\dots,\delta\boldsymbol{\varphi}_N} \frac{1}{2V} \sum_{i=1}^{N} \sum_{j\neq i} V\left(\left\|\boldsymbol{F}^*\boldsymbol{X}_i^{\mathrm{m}} + \delta\boldsymbol{\varphi}_i - \boldsymbol{F}^*\boldsymbol{X}_j^{\mathrm{m}} - \delta\boldsymbol{\varphi}_j\right\|\right), \tag{9.31}$$

where we minimize the energy with respect to the N shifts associated with the N Bravais lattices. In order to remove rigid body motion introduced through the shifts (e.g., $\delta\boldsymbol{\varphi}_1 = \dots = \delta\boldsymbol{\varphi}_N$

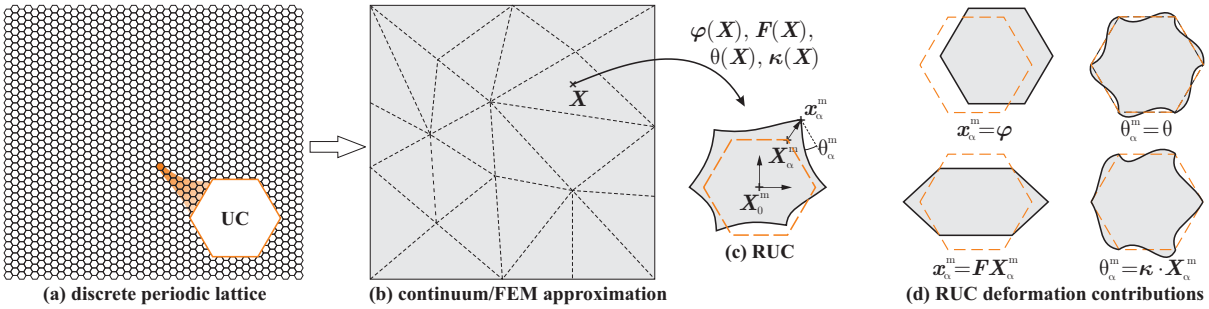


Figure 9.6: Truss modeling based on FE^2-type homogenization [29], where a macroscale finite element simulation derives the constitutive model at each quadrature point from an RVE-calculation (applying the multi-lattice Cauchy-Born rule to both nodal translations and nodal rotations and computing shifts through on-the-fly energy minimization).

implies rigid body translation), the shifts must be constrained, e.g., by requiring that $\langle \boldsymbol{\varphi} \rangle = \boldsymbol{\varphi}^*$. This approach finds application, e.g., in the homogenization of truss lattices but also in atomic crystals. Note that the calculation of the shifts by energy minimization cannot always be carried out analytically, especially if the underling problem is nonlinear. In such cases, the shifts can be computed numerically on the fly like in the FE^2 method (see Figures 9.6 and 9.7 and, e.g., references [29, 30] for the numerical FE^2-type treatment).

In summary, the Cauchy-Born rule, strictly speaking, requires the introduction of relative motion between the various simple Bravais lattices in a multi-lattice – even though the term *Cauchy-Born rule* is often, incorrectly, used to refer to affine deformation (strictly speaking, that is only correct in case of a simple Bravais lattice).

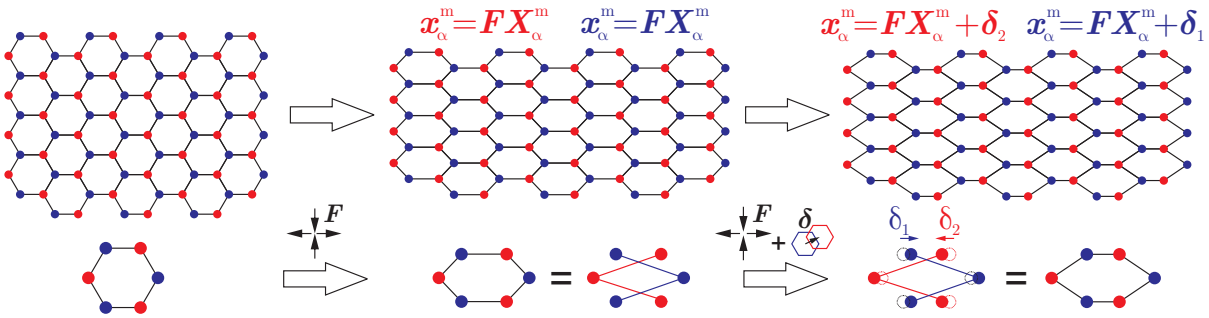


Figure 9.7: Shifts are required in multi-lattices, where affine deformation is overly stiff (here, the example of a hexagonal lattice is shown, which is the composition of two simple Bravais lattices that may shift relative to each other).

10 Spectral/FFT-Based Homogenization Techniques

Applying periodic boundary conditions to an RVE using the finite element method (as discussed in previous sections) is a popular technique, but it can be technically cumbersome: creating meshes that are periodic, identifying opposite pairs of nodes, destroying the sparse/banded structure of the stiffness matrix by periodic boundary conditions, etc. An alternative is the use of so-called *spectral* techniques, which aim to build the periodicity into the ansatz functions that are used to approximate the fields of interest (see Figure 10.1). Along with the use of a regular grid instead of an FE mesh, this ensures periodicity by definition – and it comes with certain numerical advantages (but it, of course, also introduces some complications). This section will give an introduction to such techniques based on the **fast Fourier transform** (FFT) – a concept that dates back to the late 90s (see reference [66]).

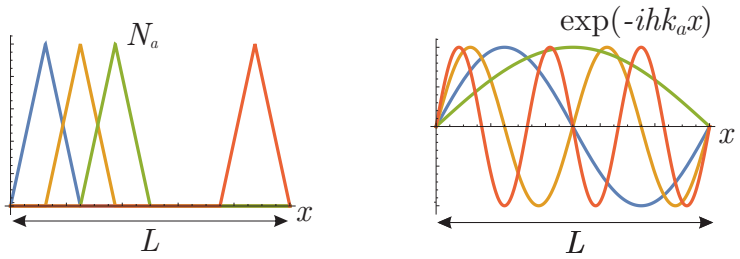


Figure 10.1: Schematic illustration of local (linear) shape functions within the FEM context on the left vs. global Fourier-based shape functions that satisfy the periodicity of the RVE problem on the right (for an RVE of length L in 1D).

10.1 Reciprocal Bravais Lattices

We consider an RVE Ω discretized by a regular grid of nodes located at lattice sites $\boldsymbol{X}_j$ defined by a Bravais basis $\mathcal{B} = \{\boldsymbol{A}_1, \dots, \boldsymbol{A}_d\}$, which consists of d basis vectors $\boldsymbol{A}_i$, such that every grid point $\boldsymbol{X}_j \in \Omega$ falls onto a linear combination of the Bravais basis vectors (see Figure 10.2):

$$\boldsymbol{X}_j = \sum_{i=1}^{d} c_i \boldsymbol{A}_i, \qquad \text{with some unique } c_1, \dots, c_d \in \mathbb{Z}. \tag{10.1}$$

The resulting infinite point set spanned by $\mathcal{B}$ is called a **Bravais lattice**.

Now consider an RVE having straight edges and flat surfaces (a parallelpiped, see Figure 10.2). We choose the corners of the RVE to align with the above grid, so that the RVE contains n grid points. Note that, owing to the periodicity of the RVE, we can tessellate all of $\mathbb{R}^d$ by periodically repeating the RVE. Thus, we have a second set of Bravais lattice vectors $\mathcal{L} = \{\boldsymbol{L}_1, \dots, \boldsymbol{L}_d\}$ that define the corners of all RVEs in $\mathbb{R}^d$, and we choose the vectors such that

$$\boldsymbol{L}_i = L_i \boldsymbol{A}_i \qquad \text{and} \qquad L_1 \cdot \ldots \cdot L_d = n, \tag{10.2}$$

where L_i defines the number of grid points within the RVE in the $\boldsymbol{A}_i$-direction (note that to avoid double-counting of $\pm$-nodes on the boundary, only the unique master nodes are counted). The locations of all unique grid points contained within the RVE is $\boldsymbol{X} = \{\boldsymbol{X}_1, \dots, \boldsymbol{X}_n\}$.

Periodicity implies that all field quantities u (such as the stress and strain components in a mechanical problem) must also be periodic across all RVEs, i.e., we must ensure that

$$u(\boldsymbol{X}) = u\left(\boldsymbol{X} + \sum_{i=1}^{d} c_i \boldsymbol{L}_i\right) \qquad \forall\ c_i \in \mathbb{Z}. \tag{10.3}$$

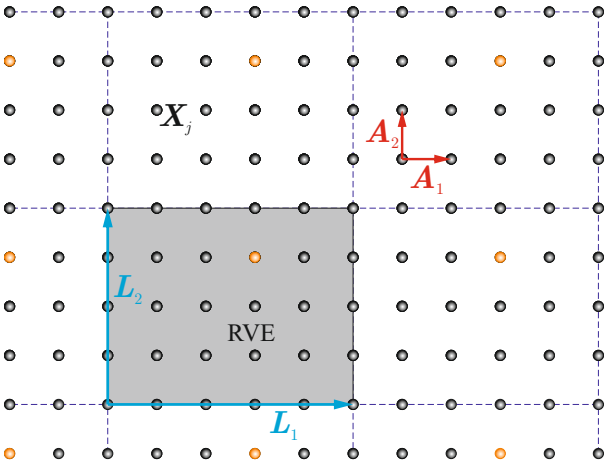


Figure 10.2: A Bravais lattice in 2D with basis $\mathcal{B} = \{\boldsymbol{A}_1, \boldsymbol{A}_2\}$ and a rectangular-shaped RVE with translation vectors/basis $\mathcal{L} = \{\boldsymbol{L}_1, \boldsymbol{L}_2\}$. As an example, all orange nodes are linked by periodicity.

To ensure this periodicity constraint, we may express the field $u(\boldsymbol{X})$ in its *discrete Fourier representation*, where we interpolate the field u by using the inverse **discrete Fourier transform**. Specifically, we write

$$u(\boldsymbol{X}) = \sum_{\boldsymbol{K}\in\mathcal{T}} \hat{u}(\boldsymbol{K}) \exp\left(ih\boldsymbol{K}\cdot\boldsymbol{X}\right), \qquad h = \frac{2\pi}{L}, \tag{10.4}$$

where $L = |\boldsymbol{L}_i|$ is the side length of the RVE. Note that, for simplicity, we here and in the following assume that $|\boldsymbol{L}_1| = |\boldsymbol{L}_2| = \ldots = |\boldsymbol{L}_d|$, i.e., all side lengths are equal. For the more general case of different RVE side lengths, the above equation must be extended accordingly by using different values of L_i in each of the d directions.

We introduced the variable h for convenience and for simple redefinition of the transform (each numerical library or code uses its own definition of the (inverse) discrete Fourier transform, unfortunately). $\mathcal{T}$ is the set of all *unit wave vectors* $\boldsymbol{K} \in \mathbb{Z}^d$, and $\hat{u}(\boldsymbol{K})$ is the *amplitude* associated with wave vector $h\boldsymbol{K}$. Unlike in the FE context, where the nodal degrees of freedom are multiplied by the nodal shape functions to arrive at the FE interpolation, the solution is here decomposed into Fourier modes and the unknowns are the amplitudes of all those modes.

To build the periodicity into the above ansatz (10.4) for $u(\boldsymbol{X})$, the particular choice of vectors $\boldsymbol{K}$ must ensure (10.3). Since the coefficients $\hat{u}(\boldsymbol{K})$ are independent, this means that we must satisfy (10.3) for each term in the Fourier series. This in turn implies that

$$\begin{aligned}\exp\left(ih\boldsymbol{K}\cdot\boldsymbol{X}\right) &= \exp\left[ih\boldsymbol{K}\cdot\left(\boldsymbol{X} + \sum_{i=1}^{d} c_i\boldsymbol{L}_i\right)\right] \qquad \forall\, c_i\in\mathbb{Z},\ \boldsymbol{K}\in\mathcal{T} \\ &= \exp\left(ih\boldsymbol{K}\cdot\boldsymbol{X}\right)\cdot\exp\left(ih\boldsymbol{K}\cdot\sum_{i=1}^{d} c_i\boldsymbol{L}_i\right) \qquad \forall\, c_i\in\mathbb{Z},\ \boldsymbol{K}\in\mathcal{T}.\end{aligned} \tag{10.5}$$

Recalling Euler's identity, the above reduces to (for all $c_i \in \mathbb{Z}$ and $\boldsymbol{K} \in \mathcal{T}$)

$$\exp\left(ih\boldsymbol{K}\cdot\sum_{i=1}^{d} c_i\boldsymbol{L}_i\right) = 1 \qquad \Leftrightarrow \qquad h\boldsymbol{K}\cdot\sum_{i=1}^{d} c_i\boldsymbol{L}_i \overset{!}{\in} 2\pi\,\mathbb{Z}, \tag{10.6}$$

which, after inserting $h = 2\pi/L$, yields

$$\frac{1}{L}\sum_{i=1}^{d} c_i\boldsymbol{K}\cdot\boldsymbol{L}_i \in \mathbb{Z} \quad \forall\, c_i\in\mathbb{Z},\ \boldsymbol{K}\in\mathcal{T}. \tag{10.7}$$

Now, we may choose, e.g., all $c_i = 0$ except for one $c_j = 1$, which illustrates that each $\boldsymbol{K}$-vector must satisfy

$$\boldsymbol{K} \cdot \frac{\boldsymbol{L}_i}{L} \in \mathbb{Z} \quad \forall\, i = 1, \dots, d. \tag{10.8}$$

This in turn defines an infinite set of $\boldsymbol{K}$-vectors or, re-interpreted, the so-called **reciprocal lattice** of grid points $\boldsymbol{K}$, which is also a Bravais lattice (each $\boldsymbol{K}$-vector composed of integer components).

To illustrate the reciprocal lattice, consider, e.g., the 2D case and draw rays perpendicular to vectors $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ at distances in $\mathbb{Z}$; the intersections of all those lines form a new periodic lattice (see Figure 10.3) with a new basis $\mathcal{K} = \{\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_d\}$. Therefore, we may write

$$\boldsymbol{K} = \sum_{i=1}^{d} k_i \boldsymbol{Z}_i \quad \text{with } k_i \in \mathbb{Z}. \tag{10.9}$$

From (10.8), we conclude that the new basis (the **reciprocal basis**) must be related to the original basis via

$$\boxed{\boldsymbol{Z}_i \cdot \frac{\boldsymbol{L}_j}{L} = \delta_{ij}} \tag{10.10}$$

using Kronecker's delta δ_{ij}. (As discussed above, this can analogously be extended to non-uniform $L_1 \neq L_2 \neq L_3$.)

As an important feature, the reciprocal lattice of the reciprocal lattice is the original lattice itself with basis vectors $\boldsymbol{L}_i/L$, which is known as the **Pontryagin duality** of the associated vector spaces. (If you do the geometric construction of drawing parallel rays and identifying intersections in the reciprocal lattice analogous to Figure 10.3), you will notice that you arrive back at the original lattice.)

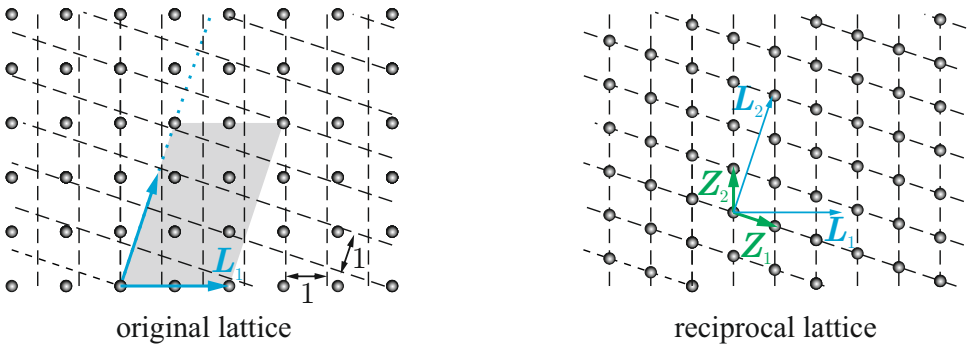


Figure 10.3: From an original lattice in 2D with Bravais basis $\mathcal{L} = \{\boldsymbol{L}_1, \boldsymbol{L}_2\}$ we construct the reciprocal lattice with basis $\mathcal{K} = \{\boldsymbol{Z}_1, \boldsymbol{Z}_2\}$ such that $\boldsymbol{Z}_i \cdot \boldsymbol{L}_j / L = \delta_{ij}$.

Examples:

Consider a 2D lattice with Bravais basis $\mathcal{L} = \{\boldsymbol{L}_1, \boldsymbol{L}_2\}$. The reciprocal lattice has Bravais basis vectors (with $\boldsymbol{R} \in \mathrm{SO}(2)$ denoting a rotation tensor performing a rotation by 90° about the out-of-plane axis)

$$\boldsymbol{Z}_1 = L \frac{\boldsymbol{R}\boldsymbol{L}_2}{\boldsymbol{L}_1 \cdot \boldsymbol{R}\boldsymbol{L}_2}, \qquad \boldsymbol{Z}_2 = L \frac{\boldsymbol{R}\boldsymbol{L}_1}{\boldsymbol{L}_1 \cdot \boldsymbol{R}\boldsymbol{L}_2}. \tag{10.11}$$

We can easily verify the reciprocity relation (10.10) from

$$\begin{aligned} \boldsymbol{Z}_1 \cdot \boldsymbol{L}_1 &= L\frac{\boldsymbol{R}\boldsymbol{L}_2}{\boldsymbol{L}_1 \cdot \boldsymbol{R}\boldsymbol{L}_2} \cdot \boldsymbol{L}_1 = L, & \boldsymbol{Z}_2 \cdot \boldsymbol{L}_1 &= L\frac{\boldsymbol{R}\boldsymbol{L}_1}{\boldsymbol{L}_1 \cdot \boldsymbol{R}\boldsymbol{L}_2} \cdot \boldsymbol{L}_1 = 0, \\ \boldsymbol{Z}_1 \cdot \boldsymbol{L}_2 &= L\frac{\boldsymbol{R}\boldsymbol{L}_2}{\boldsymbol{L}_1 \cdot \boldsymbol{R}\boldsymbol{L}_2} \cdot \boldsymbol{L}_2 = 0, & \boldsymbol{Z}_2 \cdot \boldsymbol{L}_2 &= L\frac{\boldsymbol{R}\boldsymbol{L}_1}{\boldsymbol{L}_1 \cdot \boldsymbol{R}\boldsymbol{L}_2} \cdot \boldsymbol{L}_2 = L, \end{aligned} \tag{10.12}$$

so that indeed $\boldsymbol{Z}_i \cdot \boldsymbol{L}_j / L = \delta_{ij}$.

Similarly, a 3D Bravais basis $\mathcal{L} = \{\boldsymbol{L}_1, \boldsymbol{L}_2, \boldsymbol{L}_3\}$ has reciprocal Bravais basis vectors

$$\boldsymbol{Z}_1 = L\frac{\boldsymbol{L}_2 \times \boldsymbol{L}_3}{\boldsymbol{L}_1 \cdot (\boldsymbol{L}_2 \times \boldsymbol{L}_3)}, \qquad \boldsymbol{Z}_2 = L\frac{\boldsymbol{L}_3 \times \boldsymbol{L}_1}{\boldsymbol{L}_1 \cdot (\boldsymbol{L}_2 \times \boldsymbol{L}_3)}, \qquad \boldsymbol{Z}_3 = L\frac{\boldsymbol{L}_1 \times \boldsymbol{L}_2}{\boldsymbol{L}_1 \cdot (\boldsymbol{L}_2 \times \boldsymbol{L}_3)}, \tag{10.13}$$

which satisfy the reciprocity relation (10.10) since $\boldsymbol{Z}_i \cdot \boldsymbol{L}_j / L = \delta_{ij}$.

10.2 Fourier Spectral Solution Techniques

For a field variable u, let us use the above discrete Fourier approximation (for a 1D field $u(\boldsymbol{X})$ for simplicity)

$$u(\boldsymbol{X}) = \mathcal{F}^{-1}(\hat{u}) = \sum_{\boldsymbol{K}\in\mathcal{T}} \hat{u}(\boldsymbol{K}) \exp(ih\boldsymbol{K}\cdot\boldsymbol{X}), \qquad h = \frac{2\pi}{L}. \tag{10.14}$$

Unlike in the FE method, where the degrees of freedom are the nodal values u^a, the complete discretized information about u is now contained either in its values at all grid points in real space, $\{\boldsymbol{u}\} = \{u(\boldsymbol{X}_1), \dots, u(\boldsymbol{X}_n)\}$, or in the amplitudes in Fourier space, $\{\hat{\boldsymbol{u}}\} = \{\hat{u}(\boldsymbol{K}_1), \dots, \hat{u}(\boldsymbol{K}_n)\}$. By using the transform (10.14) (and its inverse), both sets of data contain equivalent information if we choose as many wave vectors $\boldsymbol{K}$ in reciprocal space, as we have grid points in real space (i.e., if the wave vector set $\mathcal{T}$ is chosen accordingly).

Aside from introducing periodicity, the above Fourier parametrization has further advantages, which are exploited by FFT-based homogenization and so-called **spectral techniques**. Their fundamental idea is to *solve all governing equations in Fourier space* (such as linear momentum balance for a mechanical problem, or the heat equation for thermal problems). This not only ensures a high level of accuracy in computing derivatives of all fields (unlike low-order polynomial FE interpolation, the Fourier-based ansatz functions live in Sobolev spaces of infinite order). It also provides a simple way to enforce average field quantities.

Notice that out of all Fourier amplitudes $\hat{u}(\boldsymbol{K})$ one is special, as the integral over the RVE reveals:

$$\int_\Omega u(\boldsymbol{X})\,\mathrm{d}V = \sum_{\boldsymbol{K}\in\mathcal{T}} \hat{u}(\boldsymbol{K}) \int_\Omega \exp(ih\boldsymbol{K}\cdot\boldsymbol{X})\,\mathrm{d}V = V\,\hat{u}(\boldsymbol{0}). \tag{10.15}$$

Here, we exploited that all integrals in the sum vanish except for the case $\boldsymbol{K} = \boldsymbol{0}$, because (written in $d = 3$ dimensions for a cuboid RVE of side lengths $L_1 = L_2 = L_3 = L$ and volume

$V = L_1L_2L_3 = L^3$, and rescaling the Cartesian coordinates as $X_i = L_i\xi_i$)

$$\begin{aligned}
\int_\Omega \exp(ih\boldsymbol{K}\cdot\boldsymbol{X})\,\mathrm{d}V &= \int_0^L\int_0^L\int_0^L \exp\left(i\frac{2\pi}{L}\sum_{i=1}^3 K_i\boldsymbol{Z}_i\cdot\boldsymbol{X}\right)\mathrm{d}X_1\,\mathrm{d}X_2\,\mathrm{d}X_3\\
&= \int_0^1\int_0^1\int_0^1 \exp\left(i\frac{2\pi}{L}\sum_{i=1}^3 K_i\boldsymbol{Z}_i\cdot\sum_{a=1}^3\xi_a\boldsymbol{L}_a\right)L\,\mathrm{d}\xi_1\,L\,\mathrm{d}\xi_2\,L\,\mathrm{d}\xi_3\\
&= V\int_0^1\int_0^1\int_0^1 \exp\left(i2\pi\sum_{i=1}^3 K_i\xi_i\right)\mathrm{d}\xi_1\,\mathrm{d}\xi_2\,\mathrm{d}\xi_3\\
&= V\prod_{i=1}^3\int_0^1 \exp\left(i2\pi K_i\xi_i\right)\mathrm{d}\xi_i \qquad \text{where} \quad K_i\in\mathbb{Z},\\
&= V\begin{cases}1, & \text{if } K_1=K_2=K_3=0 \text{ (i.e., } \boldsymbol{K}=\boldsymbol{0}),\\ 0, & \text{else.}\end{cases}
\end{aligned}$$

The last equality can easily be verified by expanding the exponential into cosines and sines using Euler's identity.

Overall, by rewriting (10.15) we have thus arrived at the useful identity

$$\boxed{\langle u\rangle = \hat{u}(\boldsymbol{0})} \tag{10.16}$$

This could have been expected indeed: the lowest mode with $\boldsymbol{K}=\boldsymbol{0}$ is a constant mode, whereas all modes with non-zero $\boldsymbol{K}\neq\boldsymbol{0}$ contribute harmonic fluctuations around 0. As those periodic fluctuations vanish on average across the RVE, the volume average is given by $\hat{u}(\boldsymbol{0})$. This will allow us to conveniently impose a volume average by setting the value of $\hat{u}(\boldsymbol{0})$.

Recall that all governing equations (mechanical, thermal, electromagnetic, etc.) usually involve computing *derivatives* of the primary fields, e.g., to obtain strains from displacements, heat flux from temperature, etc. Here, notice that

$$\frac{\partial}{\partial\boldsymbol{X}}u(\boldsymbol{X}) = \sum_{\boldsymbol{K}\in\mathcal{T}} ih\boldsymbol{K}\hat{u}(\boldsymbol{K})\exp\left(ih\boldsymbol{K}\cdot\boldsymbol{X}\right), \tag{10.17}$$

so that

$$\boxed{\widehat{\frac{\partial u}{\partial X_J}}(\boldsymbol{K}) = ihK_J\hat{u}(\boldsymbol{K})} \tag{10.18}$$

This shows that taking spatial derivatives in Fourier space is simple and implies multiplication of the Fourier amplitudes by $ih\boldsymbol{K}$.

Our strategy will now be to transform all governing ODEs/PDEs into Fourier space and to solve the resulting equations in Fourier space for the Fourier amplitudes, where possible. Since spatial derivatives in Fourier space transform according to (10.18), ODEs in real space become algebraic equations in Fourier space (as we will show below).

10.3 Thermal Boundary Value Problem

Before studying the mechanical problem, we review a simple thermal problem (where the field of interest, the temperature field $T(\boldsymbol{X})$, is scalar and hence simpler than in the mechanical problem, where the displacement field $\boldsymbol{u}(\boldsymbol{X})$ is vector-valued). To illustrate the concepts at play, we will subdivide this problem into two versions of increasing complexity.

10.3.1 Uniform conductivity: direct solver

Consider a thermal problem that involves equilibrating an RVE by solving the static heat equation with a known heat source distribution $S(\boldsymbol{X})$ and with periodic boundary conditions, while enforcing an average temperature $\langle T\rangle$. The static heat equation with a *spatially uniform* conductivity $\lambda(\boldsymbol{X}) = \lambda = \text{const.}$ reads

$$\lambda\nabla^2 T + S = 0 \qquad \Leftrightarrow \qquad \lambda\, T_{,II} + S = 0 \qquad \forall\ \boldsymbol{X} \in \Omega, \tag{10.19}$$

which in Fourier space becomes

$$\lambda \sum_{\boldsymbol{K}\in\mathcal{T}} (ihK_I)(ihK_I)\hat{T}(\boldsymbol{K})\, \exp\left(ih\boldsymbol{K}\cdot\boldsymbol{X}\right) + \sum_{\boldsymbol{K}\in\mathcal{T}} \hat{S}(\boldsymbol{K})\, \exp\left(ih\boldsymbol{K}\cdot\boldsymbol{X}\right) = 0. \tag{10.20}$$

For this to hold at all points $\boldsymbol{X} \in \Omega$, the above must hold for each $\boldsymbol{K}$-mode independently. Therefore, we conclude that

$$-\lambda h^2 K_I K_I \hat{T}(\boldsymbol{K}) + \hat{S}(\boldsymbol{K}) = 0 \qquad \forall\ \boldsymbol{K} \in \mathcal{T}, \tag{10.21}$$

which can be solved, unless $\boldsymbol{K} = \boldsymbol{0}$, to yield

$$\hat{T}(\boldsymbol{K}) = \frac{\hat{S}(\boldsymbol{K})}{\lambda h^2 \left\|\boldsymbol{K}\right\|^2} \qquad \forall\ \boldsymbol{K} \in \mathcal{T} \text{ except for } \boldsymbol{K} = \boldsymbol{0}. \tag{10.22}$$

Therefore, we can solve the problem by utilizing Fourier transforms as follows:

(i) We first transform the known source term $S(\boldsymbol{X})$ from real space into Fourier space to obtain $\hat{S}(\boldsymbol{K})$ for all $\boldsymbol{K} \in \mathcal{T}$.

(ii) Next, we solve for all $\hat{T}(\boldsymbol{K})$ in Fourier space:

$$\boxed{\hat{T}(\boldsymbol{K}) = \begin{cases} \langle T\rangle & \text{if } \boldsymbol{K} = \boldsymbol{0} \\ \dfrac{\hat{S}(\boldsymbol{K})}{h^2\lambda\left\|\boldsymbol{K}\right\|^2} & \text{else} \end{cases}} \tag{10.23}$$

where $\langle T\rangle$ is a known average temperature to be imposed.

(iii) Finally, we apply the inverse discrete Fourier transform to obtain $T(\boldsymbol{X})$ from $\hat{T}(\boldsymbol{K})$ for all grid points $\boldsymbol{X}_i$.

The **advantages** of this method are obvious:

- *No approximating shape functions* must be introduced (and we can compute approximated derivatives to far higher order than with usual polynomial interpolation).
- *No iterative solver nor linear-system solver* is required for the above problem; the only numerical steps required are the (standard and inverse) Fourier transforms, which are numerically highly inexpensive in state-of-the-art codes.
- *No global stiffness matrix* is required for the solution, which is a big memory advantage and makes the approach applicable to very large problems.
- *RVE-averages* of the primary fields (or their derivatives) can be trivially computed and hence enforced.

10.3.2 Non-uniform conductivity: iterative solver

Next, let us reconsider the thermal problem of Section 10.3.1 – this time with a known but *spatially non-uniform* conductivity $\lambda(\boldsymbol{X})$, which may be the result, e.g., of material heterogeneity such as within a composite whose constituents have different thermal conductivity values. The governing equation in this case becomes

$$\text{Div}(\lambda\,\nabla T) + S = 0 \qquad \Leftrightarrow \qquad (\lambda\,T_{,I})_{,I} + S = 0 \qquad \forall\ \boldsymbol{X} \in \Omega. \tag{10.24}$$

When introducing the heat flux vector $\boldsymbol{Q}$, the heat equation may be rewritten as

$$\boldsymbol{Q} = -\lambda\,\nabla T \qquad \Leftrightarrow \qquad Q_I = -\lambda T_{,I} \qquad \Rightarrow \qquad -Q_{I,I} + S = 0. \tag{10.25}$$

Notice that applying a Fourier transform to (10.24) would turn the first term into a convolution, which cannot be treated in the same simple fashion as above. To this end, we introduce a *constant* reference conductivity $\lambda_0 > 0$ (with arbitrary value) along with a correction (or perturbation) field $\boldsymbol{P}$ such that

$$Q_I = -\lambda_0 T_{,I} + P_I \qquad \Leftrightarrow \qquad P_I = \lambda_0 T_{,I} + Q_I. \tag{10.26}$$

Of course, for a constant thermal conductivity $\lambda = \text{const.}$ we may choose $\lambda_0 = \lambda$, so that $\boldsymbol{P} = \boldsymbol{0}$. Insertion into (10.25) yields

$$(\lambda_0 T_{,I} - P_I)_{,I} + S = 0 \qquad \Leftrightarrow \qquad \lambda_0 T_{,II} - P_{I,I} + S = 0, \tag{10.27}$$

which now can be Fourier-transformed without convolution, resulting in

$$-\lambda_0 h^2 K_I K_I \hat{T}(\boldsymbol{K}) - ihK_I \hat{P}_I(\boldsymbol{K}) + \hat{S}(\boldsymbol{K}) = 0 \qquad \forall\ \boldsymbol{K} \in \mathcal{T} \text{ except for } \boldsymbol{K} = \boldsymbol{0}. \tag{10.28}$$

Unless $\boldsymbol{K} = \boldsymbol{0}$, we now obtain the solution

$$\hat{T}(\boldsymbol{K}) = \frac{\hat{S}(\boldsymbol{K}) - ih\boldsymbol{K}\cdot\hat{\boldsymbol{P}}(\boldsymbol{K})}{\lambda_0 h^2\,\|\boldsymbol{K}\|^2} \qquad \forall\ \boldsymbol{K} \in \mathcal{T}. \tag{10.29}$$

This form looks analogous to (10.22). Notice, however, that $\boldsymbol{P} = \boldsymbol{P}(\nabla T)$ since $\boldsymbol{P} = \lambda_0\,\nabla T + \boldsymbol{Q}$ by our definition, so that the term $\boldsymbol{K}\cdot\hat{\boldsymbol{P}}(\boldsymbol{K})$ depends on temperature T. Therefore, one cannot expect the above equation to yield the correct solution. Yet, it provides the basis for an iterative solution scheme, e.g., for the following *fixed-point iteration* scheme:

1. Choose a reference conductivity $\lambda_0 > 0$ (e.g., $\lambda_0 = \langle\lambda\rangle$).
2. Pick an initial guess for $T(\boldsymbol{X})$ at all grid points $\boldsymbol{X}_i$ in real space.
3. Compute $\boldsymbol{Q}(\nabla\boldsymbol{T})$ and $\boldsymbol{P}$ at all grid points in real space. Note that computing gradients like ∇T can also be accomplished conveniently in Fourier space without having to introduce interpolations in real space.
4. Transform $\boldsymbol{Q}(\boldsymbol{X})$ to Fourier space to obtain $\hat{\boldsymbol{P}}(\boldsymbol{K})$ at all $\boldsymbol{K}$-points.
5. Solve (10.29) for all $\hat{T}(\boldsymbol{K})$ for a given average $\langle T\rangle$ or $\langle\nabla T\rangle$ (same treatment as before).
6. Transform $\hat{T}(\boldsymbol{K})$ back into physical space to obtain $T(\boldsymbol{X})$ at all grid points.
7. Return to step 3. until the solution has converged for all grid points.

This scheme is computationally more expensive than the simple solution scheme discussed in Section 10.3.1 for a constant conductivity λ. While, there, a simple Fourier transform back and forth was sufficient, we here require an iterative solution procedure. Yet, we again do not need to form a stiffness matrix nor to solve a (non)linear system, which are significant advantages. The above fixed-point iteration can be replaced by more efficient (Newton- or Krylov-based) techniques, but we will not go into detail here (see, e.g., reference [42]).

10.4 Mechanical Boundary Value Problem

The quasistatic mechanical BVP is, in principle, analogous to the heat conduction scenario discussed above. There are two main complications though: first, the unknown field (the displacement field $\boldsymbol{u}(\boldsymbol{X})$) is vector-valued and not scalar as the temperature field before. Second, the stress–strain relation is in general *nonlinear*, so that the ODE to be solved is not necessarily linear (unless one limits the analysis to linear elasticity and linearized kinematics). In the following, let us first investigate the most general finite-kinematics setting and then turn to linearized kinematics as a special case.

10.4.1 Finite Kinematics

Consider a quasistatic, purely mechanical problem in the absence of body forces, whose linear momentum balance reads

$$\operatorname{Div} \boldsymbol{P} = \boldsymbol{0}. \tag{10.30}$$

Since $\boldsymbol{P} = \boldsymbol{P}(\boldsymbol{F})$ is a generally nonlinear stress–strain relation, transforming the above ODE into Fourier space does not present a viable strategy. Therefore, let us introduce a **perturbation stress tensor** $\boldsymbol{\tau}(\boldsymbol{x})$ such that

$$\boldsymbol{P}(\boldsymbol{X}) = \mathbb{C}^0\, \boldsymbol{F}(\boldsymbol{X}) - \boldsymbol{\tau}(\boldsymbol{X}) \tag{10.31}$$

with some constant reference modulus tensor $\mathbb{C}^0$ of fourth order (e.g., the RVE average incremental stiffness tensor $\langle \mathbb{C} \rangle$; any strongly elliptic choice of $\mathbb{C}$ is legitimate).

Inserting the strain–displacement relation

$$\boldsymbol{F}(\boldsymbol{X}) = \operatorname{Grad} \boldsymbol{\varphi}(\boldsymbol{X}) \tag{10.32}$$

into (10.30) gives, using index notation,

$$\left[\mathbb{C}^0_{iJkL}\, \varphi_{k,L}(\boldsymbol{X}) - \tau_{iJ}(\boldsymbol{X})\right]_{,J} = 0 \tag{10.33}$$

or (since $\mathbb{C}^0 = \text{const.}$ across the RVE)

$$\mathbb{C}^0_{iJkL}\, \varphi_{k,LJ}(\boldsymbol{X}) = \tau_{iJ,J}(\boldsymbol{X}). \tag{10.34}$$

This is a linear equation in the deformation mapping $\boldsymbol{\varphi}$. As before, let us introduce the (inverse) Fourier transform to approximate the deformation mapping[5] by

$$\boldsymbol{\varphi}(\boldsymbol{X}) = \sum_{\boldsymbol{K} \in \mathcal{T}} \hat{\boldsymbol{\varphi}}(\boldsymbol{K})\, \exp\left(ih \boldsymbol{K} \cdot \boldsymbol{X}\right) \tag{10.35}$$

with a suitably chosen set of wave vectors $\mathcal{T}$. Analogously transforming $\boldsymbol{\tau}(\boldsymbol{X})$ into Fourier space turns (10.34) in Fourier space into

$$-h^2 \mathbb{C}^0_{iJkL}\, \hat{\varphi}_k(\boldsymbol{K}) K_L K_J = ih \hat{\tau}_{iJ}(\boldsymbol{K})\, K_J. \tag{10.36}$$

[5] We acknowledge that applying the Fourier transform to the deformation mapping (or later to the displacement field) may not be the best or most obvious choice, since displacements are not periodic in general, whereas strains are. However, this has no effect on our general solution approach, as we will not need to evaluate the displacements in the following. Therefore, we stick to this formulation – which is analogous to the heat conduction problem. One could easily re-write the derivation by applying the Fourier transform to deformation gradient instead of to the deformation mapping without any effect on the resulting equations.

We may identify the **acoustic tensor** (which owes its name to the fact that its eigenvalues yield the elastic wave speeds)

$$A_{ik}(\boldsymbol{K}) = \mathbb{C}^0_{iJkL} K_J K_L, \tag{10.37}$$

so that (10.36) can be transformed into

$$-hA_{ik}(\boldsymbol{K})\hat{\varphi}_k(\boldsymbol{K}) = i\hat{\tau}_{iJ}(\boldsymbol{K})K_J, \tag{10.38}$$

which can be solved for the deformation mapping in Fourier space:

$$\boxed{\hat{\varphi}_k(\boldsymbol{K}) = -\frac{i}{h} A^{-1}_{ik}(\boldsymbol{K})\hat{\tau}_{iJ}(\boldsymbol{K})K_J} \tag{10.39}$$

Note that $\boldsymbol{A}(\boldsymbol{K})$ is invertible only if $\boldsymbol{K} \neq \boldsymbol{0}$ (and if $\mathbb{C}_0$ is strongly elliptic).

Next, let us transform the strain-deformation relation (10.32) into Fourier space, which gives

$$\hat{F}_{kL}(\boldsymbol{K}) = ih\hat{\varphi}_k(\boldsymbol{K})K_L. \tag{10.40}$$

By comparing (10.40) to (10.39), we can turn the Fourier-space law (10.39) into one that depends on deformation gradients only (without the need to involve the deformation mapping $\boldsymbol{\varphi}$). To this end, we pre-multiply (10.39) by ihK_L to arrive at

$$\boxed{\hat{F}_{kL}(\boldsymbol{K}) = \begin{cases} \langle \boldsymbol{F} \rangle & \text{if } \boldsymbol{K} = \boldsymbol{0}, \\ A^{-1}_{ik}(\boldsymbol{K})\hat{\tau}_{iJ}(\boldsymbol{K})K_J K_L & \text{else,} \end{cases}} \tag{10.41}$$

where we again exploited the fact that $\langle \boldsymbol{F} \rangle = \hat{\boldsymbol{F}}(\boldsymbol{0})$. It is important to recall that $\boldsymbol{\tau}$ depends on $\boldsymbol{F}$, so that solving (10.41) for $\boldsymbol{F}$ can only be part of an *iterative* solution scheme (similar to the non-uniform thermal conductivity example discussed in Section 10.3.2). Like the thermal problem in Section 10.3.2, (10.41) can be solved by *fixed-point iteration* as an example (starting with initial guesses of $\boldsymbol{F}$ at the grid points and iterating until convergence is achieved).

As an example, the *fixed-point solution algorithm* is as follows:

1. Choose a (strongly elliptic) reference stiffness tensor $\mathbb{C}^0$ (e.g., the RVE average $\mathbb{C}^0 = \langle \mathbb{C} \rangle$).
2. Pick an initial guess for $\boldsymbol{F}(\boldsymbol{X})$ at all grid points $\boldsymbol{X}_i$ in physical space.
3. Compute $\boldsymbol{P}(\boldsymbol{X}) = \boldsymbol{P}(\boldsymbol{F}(\boldsymbol{X}))$ from the constitutive law and $\boldsymbol{\tau}(\boldsymbol{X}) = \mathbb{C}^0\boldsymbol{F}(\boldsymbol{X}) - \boldsymbol{P}(\boldsymbol{X})$ at all grid points in physical space.
4. Transform $\boldsymbol{\tau}(\boldsymbol{X})$ into Fourier space to obtain $\hat{\boldsymbol{\tau}}(\boldsymbol{K})$ at all $\boldsymbol{K}$-points.
5. Solve (10.41) for all $\hat{\boldsymbol{F}}(\boldsymbol{K})$ within a given (imposed) average $\langle \boldsymbol{F} \rangle$.
6. Transform $\hat{\boldsymbol{F}}(\boldsymbol{K})$ back into physical space to obtain $\boldsymbol{F}(\boldsymbol{X})$ at all grid points.
7. Return to step 3., until the solution has converged for all grid points.
8. After convergence is reached, (10.39) can be used for computing the deformation mapping in a postprocessing step.

We note that, even though (10.41) is formulated in terms of the deformation gradient (and the deformation mapping is not even required), the solution is guaranteed to be *compatible* (i.e., a

deformation mapping $\boldsymbol{\varphi}$ is guaranteed to exist). Mechanical compatibility requires (using the permutation symbol ϵ_{IJK})

$$\operatorname{Curl} \boldsymbol{F} = \boldsymbol{0} \qquad \Leftrightarrow \qquad \epsilon_{KLI} F_{kL,K} = 0 \tag{10.42}$$

or, in Fourier space,

$$ih\epsilon_{KLI} \hat{F}_{kL} K_K = 0. \tag{10.43}$$

Insertion of (10.41) shows that

$$\begin{aligned} ih\epsilon_{KLI} \hat{F}_{kL} K_K &= ih\epsilon_{KLI} A^{-1}_{ik}(\boldsymbol{K}) \hat{\tau}_{iJ}(\boldsymbol{K}) K_J K_K K_L \\ &= ih \frac{1}{2} (\epsilon_{KLI} + \epsilon_{LKI}) A^{-1}_{ik}(\boldsymbol{K}) \hat{\tau}_{iJ}(\boldsymbol{K}) K_J K_K K_L = 0. \end{aligned} \tag{10.44}$$

Thus, the deformation gradient field resulting from (10.41) is per definition compatible.

We can conveniently use the Fourier transform to also compute other RVE-averages such that, e.g., the average stress tensor needed for computational homogenization since

$$\langle \boldsymbol{P}(\boldsymbol{X}) \rangle = \hat{\boldsymbol{P}}(\boldsymbol{0}) = \frac{1}{V} \int_\Omega \boldsymbol{P}(\boldsymbol{X}) \, \mathrm{d}V = \frac{1}{n} \sum_{i=1}^n \boldsymbol{P}(\boldsymbol{X}_i). \tag{10.45}$$

Unfortunately, the effective incremental stiffness tensor

$$\begin{aligned} \mathbb{C}^* = \frac{\partial \boldsymbol{P}^*}{\partial \boldsymbol{F}^*} &= \frac{1}{n} \frac{\partial}{\partial \boldsymbol{F}^*} \sum_{i=1}^n \boldsymbol{P}(\boldsymbol{F}(\boldsymbol{X}_i)) = \frac{1}{n} \sum_{i=1}^n \frac{\partial \boldsymbol{P}}{\partial \boldsymbol{F}}(\boldsymbol{X}_i) \cdot \frac{\partial \boldsymbol{F}(\boldsymbol{X}_i)}{\partial \boldsymbol{F}^*} \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{C}(\boldsymbol{X}_i) \cdot \frac{\partial}{\partial \boldsymbol{F}^*} \left[\hat{\boldsymbol{F}}(\boldsymbol{0}) + \sum_{\boldsymbol{K} \neq \boldsymbol{0}} \hat{\boldsymbol{F}}(\boldsymbol{K}) \, \exp\left(ih \boldsymbol{K} \cdot \boldsymbol{X}_i\right) \right] \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{C}(\boldsymbol{X}_i) \cdot \left[\mathbb{I} + \sum_{\boldsymbol{K} \neq \boldsymbol{0}} \exp\left(ih \boldsymbol{K} \cdot \boldsymbol{X}_i\right) \boldsymbol{A}^{-\mathrm{T}}(\boldsymbol{K}) \frac{\partial \hat{\boldsymbol{\tau}}(\boldsymbol{K})}{\partial \boldsymbol{F}^*} \boldsymbol{K} \otimes \boldsymbol{K} \right] \end{aligned} \tag{10.46}$$

cannot be evaluated in closed form as easily from the Fourier representation, so that using the above technique within an FE^2 setting requires sophisticated approaches to computing for incremental tangent (which we do not discuss here for brevity).

10.4.2 Linearized Kinematics

In linearized kinematics, the methodology is analogous to the above setup in finite kinematics. Now, we have $\boldsymbol{\varepsilon}(\boldsymbol{x}) = \operatorname{sym}(\nabla \boldsymbol{u}(\boldsymbol{x}))$ and $\boldsymbol{\sigma}(\boldsymbol{x}) = \mathbb{C}(\boldsymbol{x}) \boldsymbol{\varepsilon}(\boldsymbol{x})$. We start by introducing the perturbation stress tensor

$$\boldsymbol{\tau}(\boldsymbol{x}) = \mathbb{C}^0 \boldsymbol{\varepsilon}(\boldsymbol{x}) - \boldsymbol{\sigma}(\boldsymbol{x}). \tag{10.47}$$

Applying the same steps as in Section 10.4.1, we now aim at solving for the displacement field $\boldsymbol{u}(\boldsymbol{x})$. In Fourier space we obtain the analogous relation to (10.39), viz.

$$\hat{u}_k(\boldsymbol{k}) = \frac{i}{h} A^{-1}_{ik}(\boldsymbol{k}) \hat{\tau}_{ij}(\boldsymbol{k}) k_j. \tag{10.48}$$

The strain–displacement relation $\boldsymbol{\varepsilon} = \operatorname{sym}(\nabla \boldsymbol{u})$ becomes in Fourier space

$$\hat{\varepsilon}_{kl}(\boldsymbol{k}) = -\frac{ih}{2} \left[\hat{u}_k k_l + \hat{u}_l k_k \right]. \tag{10.49}$$

Insertion of (10.48) yields an analogous form to (10.41), viz.

$$\boxed{\hat{\varepsilon}_{kl}(\boldsymbol{k}) = \begin{cases} \langle \boldsymbol{\varepsilon} \rangle & \text{if } \boldsymbol{k} = \boldsymbol{0}, \\ \frac{1}{2}\big[A_{ik}^{-1}(\boldsymbol{k})k_l + A_{il}^{-1}(\boldsymbol{k})k_k\big]\tau_{ij}(\boldsymbol{k})k_j & \text{else.} \end{cases}} \tag{10.50}$$

The displacement field in Fourier space follows from (10.48). The numerical algorithm is analogous to the finite-deformation formulation (see, e.g., Section 10.4.1 for a fixed-point iteration scheme).

As a special case, consider **linear elasticity**. It is important to realize that, even though the stress strain relation in linear elasticity is pointwise linear (i.e., $\sigma_{ij} = \mathbb{C}_{ijkl}\varepsilon_{kl}$), the Fourier transform can only result in a linear system – like in the heat conduction – if $\mathbb{C}$ is spatially uniform. The latter is a rather uninteresting case because, without body forces or dynamic effects, it will result in uniform deformation within the RVE. To see this, we apply linear momentum balance to the case of linear elasticity with spatially constant $\mathbb{C}(\boldsymbol{x}) = \mathbb{C}$:

$$0 = \sigma_{ij,j} = \left(\mathbb{C}_{ijkl}u_{k,l}\right)_{,j} = \mathbb{C}_{ijkl}u_{k,lj} \qquad \Rightarrow \qquad -h^2\mathbb{C}_{ijkl}\hat{u}_k(\boldsymbol{k})k_lk_j = 0, \tag{10.51}$$

so that we conclude $\hat{\boldsymbol{u}}(\boldsymbol{k}) = \boldsymbol{0}$ for all $\boldsymbol{k} \in \mathcal{T}$ except for $\boldsymbol{k} = \boldsymbol{0}$ (which implements the average strain in the RVE). The result is hence a uniform strain inside the RVE.

By contrast, if the stiffness is not spatially uniform (such as in a composite or in a polycrystal with anisotropic local stiffness), then

$$\sigma_{ij}(\boldsymbol{x}) = \mathbb{C}_{ijkl}(\boldsymbol{x})\varepsilon_{kl}(\boldsymbol{x}), \tag{10.52}$$

so that the Fourier transform involves a convolution and the equation in Fourier space is not linear (which is why the perturbation stress tensor was introduced above).

10.5 Fourier-Related Stability and Accuracy Issues

All of the above Fourier spectral formulations used a regular discrete grid with a finite set of grid points to approximate a function $f(\boldsymbol{x})$ (such as the temperature or displacement field) by

$$f_h(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\mathcal{T}} \hat{f}(\boldsymbol{k}) \exp(ih\boldsymbol{k}\cdot\boldsymbol{x}), \tag{10.53}$$

where $\mathcal{T}$ denotes the *finite* set of grid points in reciprocal space. To find the exact solution, we would need to work with $\mathcal{T}_\infty$, the countably infinite set of the exact Fourier representation. Obviously, we make an approximation by truncating the infinite Fourier series to its lowest-order terms, by neglecting those modes corresponding to $\mathcal{T}^* = \mathcal{T}_\infty \setminus \mathcal{T}$. In general, Fourier coefficients have pointwise convergence; however, the error of a truncated Fourier series, due to dropping the higher-frequency terms, depends on the smoothness of the function.

To see this, one may bound the **truncation error** at any point $\boldsymbol{x} \in \Omega$ (here written in linearized kinematics for short) by

$$\begin{aligned} |f_h - f| &= \left| \sum_{\boldsymbol{k}\in\mathcal{T}^*} \hat{f}(\boldsymbol{k}) \exp(ih\boldsymbol{k}\cdot\boldsymbol{x}) \right| = \left| \sum_{\boldsymbol{k}\in\mathcal{T}^*} \hat{a}(\boldsymbol{k}) \sin(h\boldsymbol{k}\cdot\boldsymbol{x}) + \hat{b}(\boldsymbol{k}) \cos(h\boldsymbol{k}\cdot\boldsymbol{x}) \right| \\ &\leq \sum_{\boldsymbol{k}\in\mathcal{T}^*} \left| \hat{a}(\boldsymbol{k}) + \hat{b}(\boldsymbol{k}) \right| \leq \sum_{\boldsymbol{k}\in\mathcal{T}^*} |\hat{a}(\boldsymbol{k})| + |\hat{b}(\boldsymbol{k})|, \end{aligned} \tag{10.54}$$

where we used the triangle inequality for the last step, and we defined $\hat{a} = i\hat{f}$, $\hat{b} = \hat{f}$.

By applying Poincaré's inequality to the transform and using Parseval's identity, one obtains

$$\begin{aligned}\sum_{\boldsymbol{k}\in\mathcal{T}^*} |\hat{a}(\boldsymbol{k})| + |\hat{b}(\boldsymbol{k})| &\leq \sum_{\boldsymbol{k}\in\mathcal{T}^*} \frac{1}{|\boldsymbol{k}|}\left(|\hat{a}'(\boldsymbol{k})| + |\hat{b}'(\boldsymbol{k})|\right) \\ &\leq \sqrt{2}\left[\sum_{\boldsymbol{k}\in\mathcal{T}^*} |\hat{a}'(\boldsymbol{k})|^2 + |\hat{b}'(\boldsymbol{k})|^2\right]^{1/2} \leq \sqrt{\frac{2}{\pi n}}\left\|f'\right\|_{L_2(\Omega)}^{1/2}.\end{aligned} \tag{10.55}$$

Thus, the following upper bound holds for the truncation error incurred by working with only the lowest-order terms in our Fourier representation:

$$\boxed{\sup_{\boldsymbol{x}\in\Omega} |f_h - f| \leq \sqrt{\frac{2}{\pi n}}\left\|f'\right\|_{L_2(\Omega)}^{1/2}} \tag{10.56}$$

Simply put, this error depends on the smoothness of the function f (smoother functions yielding smaller errors). Importantly, the decay rate of the Fourier coefficients results in a non-uniform convergence. The consequence is a problem known as **ringing artifacts** in case of sharp contrasts in local properties (such as when dealing with jumps in the thermal conductivity, elastic constants, or crystal orientation), which manifest through *high-frequency oscillations* in the solution near such discontinuities (see the examples in Figures 10.7, 10.5, and 10.4 below).

One can reduce such artifacts by ensuring that all spatial derivatives within the RVE are bounded (thus enforcing a certain level of smoothness). This can be accomplished, e.g., by approximating the differential operator by a discrete, finite-difference operator before the Fourier transform. While this ensures asymptotic consistency, the spectral accuracy of the method is reduced to that of the finite-difference operator (as shown, e.g., in reference [95]). The approach works as follows. Instead of applying the Fourier transform to a spatial derivative of f directly, i.e.

$$\widehat{\frac{\partial f}{\partial x_i}} = ihk_i\widehat{f}, \tag{10.57}$$

we first apply a finite-difference approximation to the partial derivatives. For example, by using a symmetric central-difference approximation with a uniform grid spacing $|\Delta x_i| \ll 1$, we approximate a spatial derivative of f in the x_i-direction (with Cartesian unit vector $\boldsymbol{e}_i$, no summations over i implied) by

$$\frac{\partial f}{\partial x_i}(\boldsymbol{x}) = \frac{f(\boldsymbol{x} + \Delta x_i\boldsymbol{e}_i) - f(\boldsymbol{x} - \Delta x_i\boldsymbol{e}_i)}{2\Delta x_i} + O(\Delta x_i^2). \tag{10.58}$$

Neglecting all higher-order terms and applying the discrete inverse Fourier transform yields

$$\begin{aligned}\frac{\partial f}{\partial x_i}(\boldsymbol{x}) &= \frac{\partial}{\partial x_i}\sum_{\boldsymbol{k}\in\mathcal{T}} \hat{f}(\boldsymbol{k})\,\exp\left(ih\boldsymbol{k}\cdot\boldsymbol{x}\right) \\ &\approx \sum_{\boldsymbol{k}\in\mathcal{T}} \hat{f}(\boldsymbol{k})\frac{\exp\left[ih\boldsymbol{k}\cdot(\boldsymbol{x} + \Delta x_i\boldsymbol{e}_i)\right] - \exp\left[ih\boldsymbol{k}\cdot(\boldsymbol{x} - \Delta x_i\boldsymbol{e}_i)\right]}{2\Delta x_i}.\end{aligned} \tag{10.59}$$

Using the relation

$$\exp(ih\boldsymbol{k}\cdot(\boldsymbol{x}\pm\Delta\boldsymbol{x})) = \exp(ih\boldsymbol{k}\cdot\boldsymbol{x})\exp(\pm ih\boldsymbol{k}\cdot\Delta\boldsymbol{x}), \tag{10.60}$$

as well as Euler's identity

$$\frac{\exp(ih\boldsymbol{k}\cdot\Delta\boldsymbol{x})) - \exp(-ih\boldsymbol{k}\cdot\Delta\boldsymbol{x})}{2i} = \sin(h\boldsymbol{k}\cdot\Delta\boldsymbol{x}), \tag{10.61}$$

we may transform (10.59) into (again, no summations over i implied)

$$\frac{\partial f}{\partial x_i}(\boldsymbol{x}) \approx \sum_{\boldsymbol{k}\in\mathcal{T}} \hat{f}(\boldsymbol{k})\, \exp\left(ih\boldsymbol{k}\cdot\boldsymbol{x}\right) \frac{i\sin(hk_i\Delta x_i)}{\Delta x_i}. \tag{10.62}$$

For a regular grid with equal spacings $\Delta x_i = \Delta x$, we thus obtain

$$\frac{\partial f}{\partial x_i} \approx \sum_{\boldsymbol{k}\in\mathcal{T}} \frac{i\sin(hk_i\Delta x)}{\Delta x} \hat{f}(\boldsymbol{k})\, \exp\left(ih\boldsymbol{k}\cdot\boldsymbol{x}\right). \tag{10.63}$$

Finally, taking the Fourier transform leads to an approximate form of (10.57), viz.

$$\boxed{\widehat{\frac{\partial f}{\partial x_i}} \approx i\frac{\sin(hk_i\Delta x)}{\Delta x}\widehat{f}} \tag{10.64}$$

where the fractional term is related to the so-called *Lanczos-σ factor*.

In the limit of infinitely fine grids, notice that the approximation (10.64) converges to the exact relation (10.57), since

$$\lim_{\Delta x\to 0} i\frac{\sin(hk_i\Delta x)}{\Delta x} = ihk_i. \tag{10.65}$$

While the above finite-difference approximation smooths derivatives and hence reduces ringing artifacts, the associated error introduced by the central difference scheme creates numerical errors in the solution. These can generally be reduced by starting with higher-order finite difference stencils in (10.58). For example, the *fourth-order* central difference approximation

$$\frac{\partial f}{\partial x_i}(\boldsymbol{x}) = \frac{-f(\boldsymbol{x}+2\Delta x\,\boldsymbol{e}_i) + 8f(\boldsymbol{x}+\Delta x\,\boldsymbol{e}_i) - 8f(\boldsymbol{x}-\Delta x\,\boldsymbol{e}_i) + f(\boldsymbol{x}-2\Delta x\,\boldsymbol{e}_i)}{12\Delta x} + O(\Delta x^4), \tag{10.66}$$

yields, proceeding with the same approach outlined above,

$$\widehat{\frac{\partial f}{\partial x_i}} \approx -i\left[\frac{8\sin(hk_i\Delta x)}{6\Delta x} - \frac{\sin(2hk_i\Delta x)}{6\Delta x}\right]\widehat{f}. \tag{10.67}$$

Again, we recover the exact solution as the grid spacing tends to zero, because

$$\lim_{\Delta x\to 0} -i\left[\frac{8\sin(hk_i\Delta x)}{6\Delta x} - \frac{\sin(2hk_i\Delta x)}{6\Delta x}\right] = ihk_i. \tag{10.68}$$

The same procedure can be applied to finite-difference approximations of arbitrary order to result in increased accuracy.

The effectiveness of the discrete derivative approximation is demonstrated in Fig. 10.4, which illustrates the numerically-computed derivatives of two 1D examples: a Heaviside step function and a non-periodic half-sine wave. In both cases, we compare the exact (discontinuous) solution, the solution based on the standard Fourier transform, and the modified Fourier transform described above. Although they do not disappear entirely, high-frequency artifacts are considerably reduced by the finite-difference approximation as compared to the classical Fourier-transform. Besides numerical errors in the solution scheme, oscillations in the shown results obtained from classical and modified Fourier transforms also stem from the Fourier interpolation after solving for function values at grid points (cf. the discrete solution at grid points also

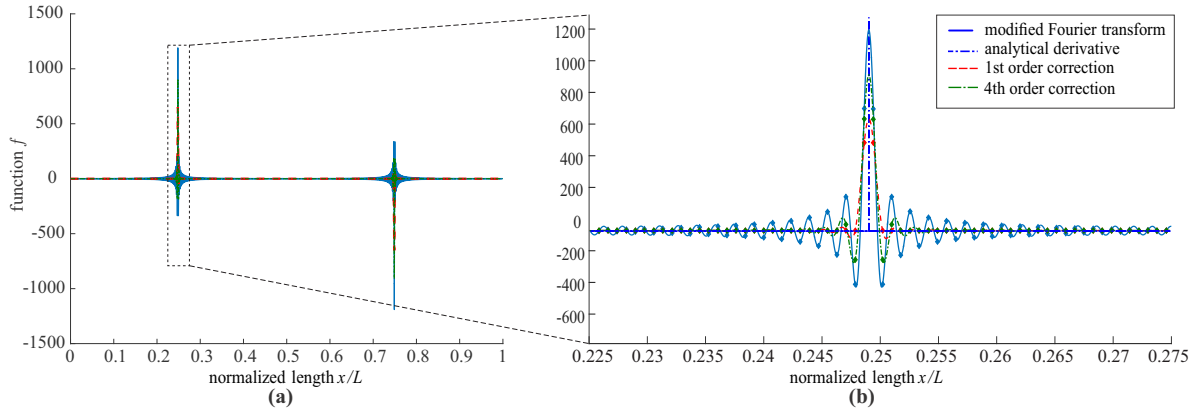


Figure 10.4: Illustration of the effectiveness of the first- and fourth-order finite-difference correction on the spectral derivative (using FFT) of the double step function $f(x) = \delta(x - \frac{1}{4} + 2^{-10}) + \delta(x - \frac{3}{4} - 2^{-10})$ periodically continued with period $x \in [0, 1)$. Shown are solutions obtained from the classical Fourier spectral method, the modified Fourier transform (first-order and fourth-order correction), and the exact analytical solution [95].

shown in Figure 10.4). Similarly, Figure 10.5 illustrates the ringing artifacts and their remedy for the example of a half-sine function.

To illustrate the effectiveness of higher-order corrections, Figure 10.6 presents the convergence order of the Euclidean error norm for corrections of varying order applied to the smooth example function

$$f(x) = \sum_{n=1}^{n=9} \sin(2n\pi x). \tag{10.69}$$

As can be expected for smooth functions, the spectral accuracy is degraded when using the modified Fourier transform with finite-difference approximation (see Figure 10.6a). Using a higher-order correction is shown to be advantageous, as the error decreases significantly while resulting in lower convergence order (see Figure 10.6b). Thus, the correction order can be chosen to control the competition between accuracy and convergence rate.

Next, we demonstrate the modified spectral scheme for a simple elasticity problem. Consider a composite bar of length L having piecewise-constant cross-sections and Young's moduli, as

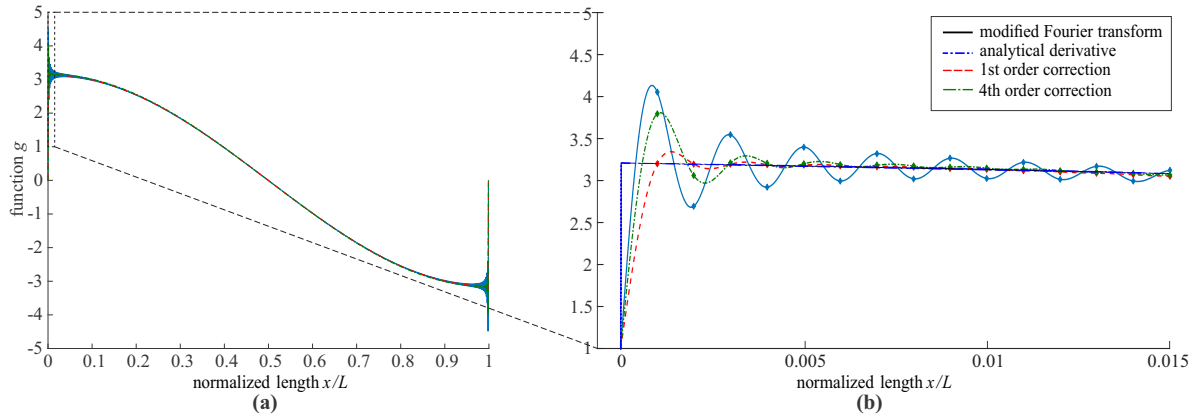


Figure 10.5: Illustration of the effectiveness of the first- and fourth-order finite-difference correction on the spectral derivative (using FFT) of the half-sine function $f(x) = \pi \cos(\pi x)$ periodically continued with period $x \in [0, 1)$. Shown are solutions obtained from the classical Fourier spectral method, the modified Fourier transform (first-order and fourth-order correction), and the exact analytical solution [95].

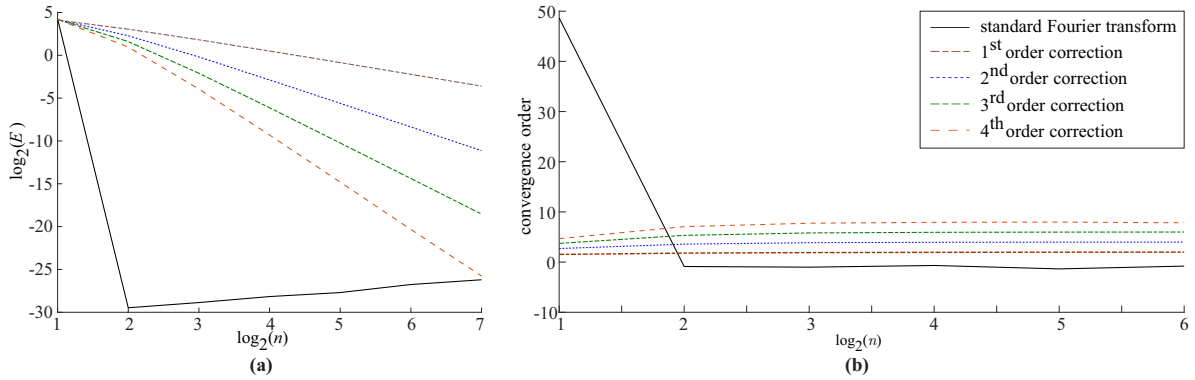


Figure 10.6: (a) Mesh convergence indicating the loss of accuracy of the spectral method; shown is the error $E = ||u_h - u||_{L_2}$ vs. number of grid points n. (b) Comparison of the convergence order of the Fourier spectral scheme for different orders of the finite-difference correction [95].

shown in Figure 10.7, and subjected to uniaxial tension such that the average strain is 0.1. The outer two sections have Young's modulus $2E$ and lengths $0.25L$, while the inner section has Young's modulus E and length $0.5L$. The resulting axial strain distribution in the bar as obtained from the classical and modified Fourier spectral scheme with first- and fourth-order corrections is plotted in Figure 10.7, demonstrating the effectiveness of the modified scheme in suppressing ringing artifacts near the sharp transitions.

As a representative 2D example of a high-contrast RVE, we compute the response of a linear elastic composite made of a circular inclusion (radius $0.25L$; normalized Lamé moduli $\lambda = 1$, $\mu = 1$) embedded in a matrix (normalized Lamé moduli $\lambda = 0.6$, $\mu = 0.6$, outer side length L). Plotted in Figure 10.8 is the normal stress distribution within the RVE for a biaxial tension test (for simplicity, shown is the affine interpolation of the 256×256 grid point values, i.e., the shown solution is not affected by errors in the Fourier interpolation). Obviously, the classical Fourier spectral method results in strong artifacts near the interface between inclusion and matrix

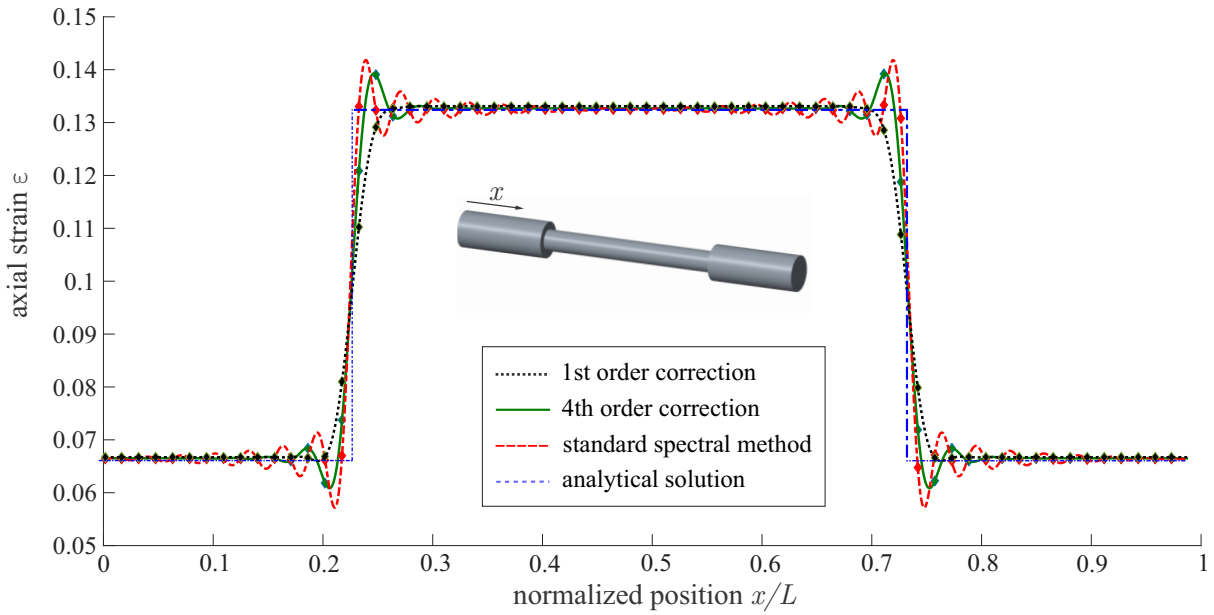


Figure 10.7: Distribution of the axial strain in a composite bar obtained from solving linear momentum balance by the iterative spectral method with first-order, fourth-order and without correction, compared to the exact piecewise-constant solution. Dots denote the converged grid point values, whereas lines follow from Fourier interpolation [95].

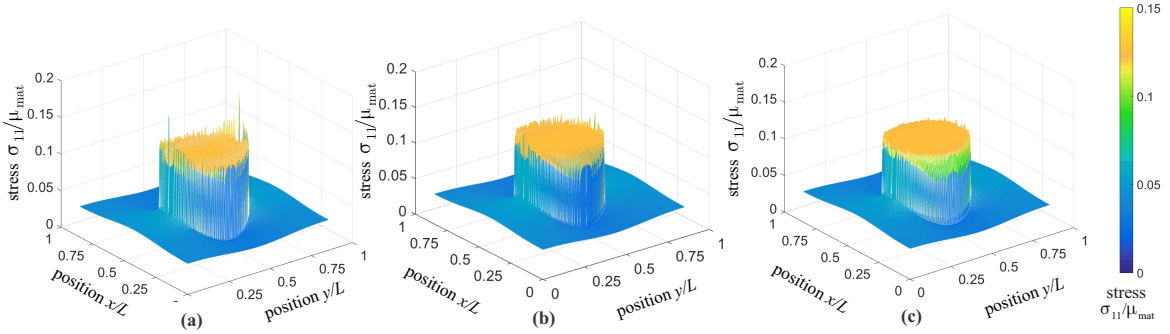


Figure 10.8: Uniaxial stress distribution in an elastic square-RVE with a circular inclusion using (a) the standard iterative spectral method and the modified spectral method with (b) fourth-order and (c) first-order correction. The first-order correction over-smooths, the fourth-order correction still produces small oscillations, but note that no Fourier interpolation has been performed here (shown are the grid point values) [95].

(Figure 10.8a), which is reduced significantly by using the first-and fourth-order corrections.

It is important to point out that several previous approaches circumvented ringing artifacts near high-contrast interfaces by aborting the iterative solution scheme before oscillations appeared in the solution. Unlike the corrections presented above, such an approach does not guarantee convergence nor bounded errors and, especially, is not suitable for a staggered, time-stepping solution procedure as that used for the dissipative processes in ferroelectrics. The numerical error of each truncated iteration can propagate and escalate over time, which is why the method presented here is superior especially in time/load-stepping scenarios.

Figures 10.9 and 10.10 show an example application of the finite-difference-corrected FFT-based homogenization scheme (taken from reference [96]). The stress-strain response of polycrystalline Mg is simulated, using a crystal plasticity model that includes slip and twinning. For the texture shown in the pole figure of Figures 10.9(b), RVEs with different numbers of grains are generated (see Figures 10.10). Since the response depends on which orientation is assigned to which grain, we repeat each calculation for ten different RVEs of the same size and number of grains but with differently sampled grain orientations from the pole figure's orientation distribution (hence performing ensemble averaging). The results is shown in Figures 10.9: for low numbers of grains, the stress-strain response varies significantly between different realizations (shown are the mean response as well as the standard deviation for each RVE size). With increasing RVE size and hence with increasing numbers of grains within the RVE (i.e., with sample enlargement), the mean converges and the standard deviation decreases as well, until converge is achieved towards an effective stress-strain response.

Such simulations with the shown number of grains reaching 1000 (and each grain being considerably resolved to capture heterogeneous stress and strain distributions within grains) would not be feasible with standard FE techniques. The FFT-based formulation enables the high resolution at reasonable computational costs, which is why it was chosen here.

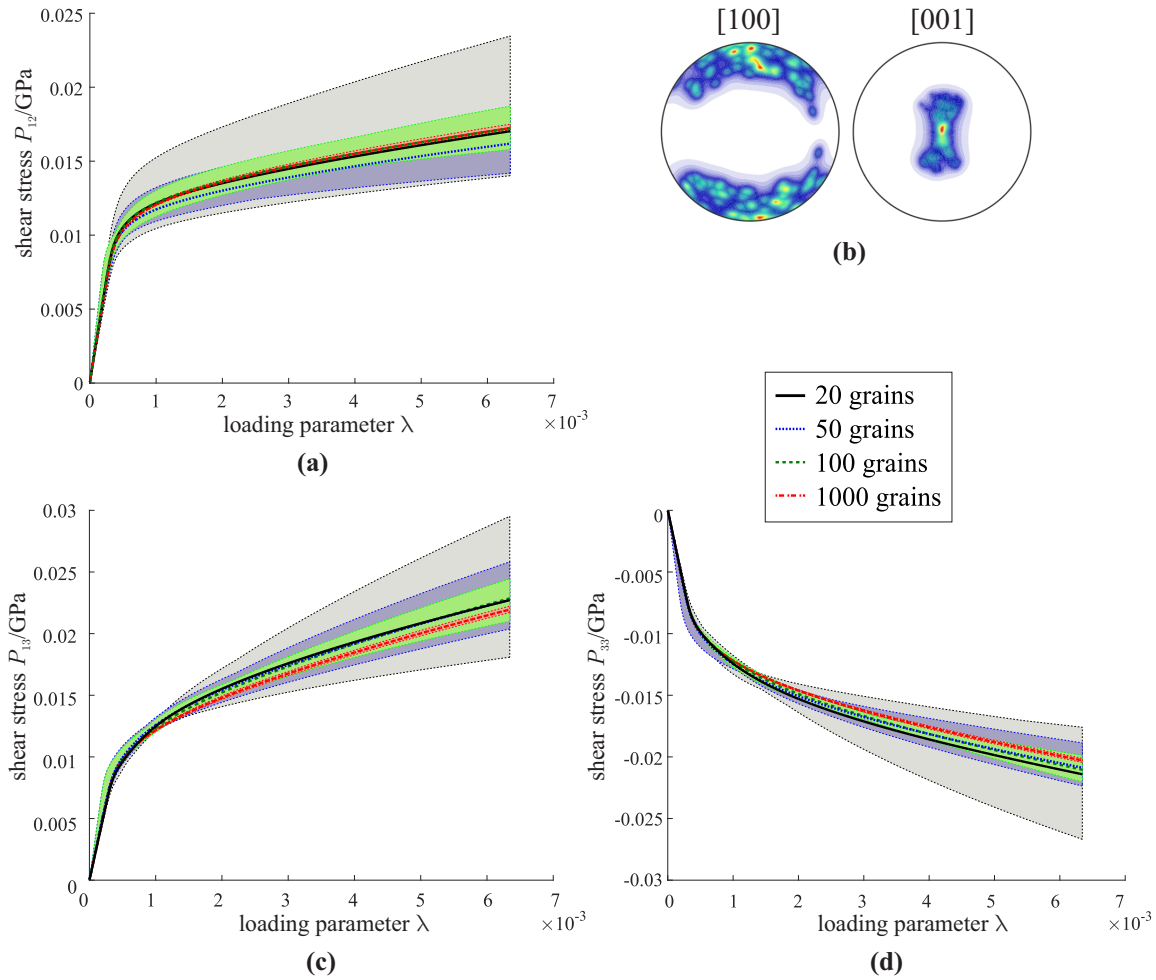


Figure 10.9: Simulated stress-strain behavior of polycrystalline magnesium shown in terms of mean and standard deviation of ten RVE realizations with different permutations of grain-orientation assignments for RVEs containing 20, 50, 100, and 1000 grains. Grain orientations are taken from the pole shown in (b); the given components of the first Piola-Kirchhoff stress tensor include (a) tensile, (c) shear, and (d) compressive stresses. Mean stresses are shown as thick lines, standard deviations as shaded color regions (from reference [96]).

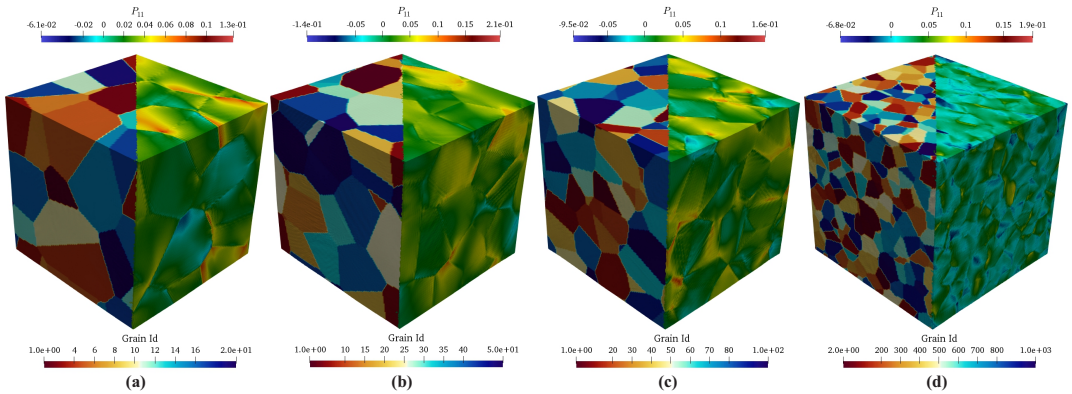


Figure 10.10: Polycrystalline RVEs with (a) 20, (b) 50, (c) 100, and (d) 1000 grains, used to produce the results in Figure 10.9. The left half of each graphic illustrates the grain size and arrangement, whereas the right half shows the tensile/compressive first Piola-Kirchhoff stress distribution at a representative load of 1% (from reference [96]).

11 Bloch Wave Analysis

So far, we have restricted our attention to quasistatic homogenization problems, in which inertial effects played a negligible role. When inertial effects do play a role, it is important to consider their impact on the effective dynamic response of a microstructure. Unlike the dynamic response of a continuum, the heterogeneous microstructure can give rise to a phenomenon known as **scattering** of waves: simply put, waves can interact with the microstructure, which leads to waves of different frequencies leading to different behavior – from transmission to complete attenuation. While we cannot give a thorough account of the topic in its entirety (that is a research area of its own), we here give a brief summary of the key concepts. Detailed discussions can be found in textbooks, starting with the seminal of *Léon Brillouin* (see reference [13]). For simplicity, we also limit our discussion to linear elastic media (while the extension to a more general setting can be technically demanding, it is conceptually analogous).

11.1 Wave Motion in an Elastic Medium

11.1.1 Homogeneous Bar

Consider a homogeneous, linear elastic bar in the absence of body forces, whose dynamics is governed by the equation of motion in 1D:

$$\operatorname{div}\sigma = \rho a \qquad \Rightarrow \qquad E u_{,xx}(x,t) = \rho \ddot{u}(x,t). \tag{11.1}$$

We know that the above wave equation supports wave solutions of the type

$$u(x,t) = \hat{u}(kx - \omega t), \tag{11.2}$$

where k is the **wave number**, ω the wave's (angular) **frequency**, and $\hat{u}(\cdot)$ is a smooth, differentiable function determined by the initial and boundary conditions of an initial boundary value problem. For an infinite harmonic wave travelling through the medium, one usually assumes $\hat{u}(kx-\omega t) = \hat{u}\exp\left[i(kx-\omega t)\right]$ with some constant amplitude $\hat{u} \in \mathbb{R}$. Note that in this case the **wave number** $k = 2\pi/\lambda$ is inversely proportional to the **wavelength** λ of the wave. Insertion into 11.1 yields

$$Ek^2\hat{u}''(kx-\omega t) = \rho\omega^2\hat{u}''(kx-\omega t) \quad \forall\, x,t \qquad \Leftrightarrow \qquad Ek^2 = \rho\omega^2, \tag{11.3}$$

so that

$$|\omega| = c|k| \qquad \text{with} \qquad c = \sqrt{\frac{E}{\rho}}. \tag{11.4}$$

where c is the wave speed. This is the classical solution for waves propagating in a homogeneous infinite bar. To each wave number k, there exists a unique frequency ω at which the wave propagates – and the relation between those is linear.

11.1.2 Heterogeneous Medium: the Monatomic Chain

Next, let us consider a toy example of an elastic medium in 1D with a discrete microstructure: imagine an infinite chain of identical particles, each of mass m and connected to its nearest neighbors through elastic springs of stiffness k_S (see Figure 11.1). This scenario is known as

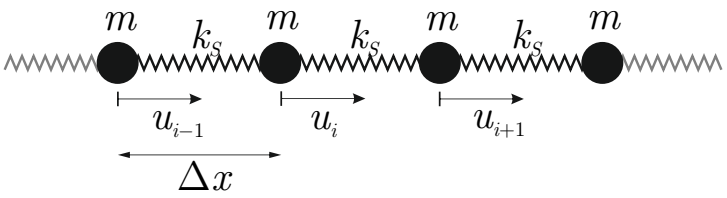


Figure 11.1: Monatomic chain of identical particles of masses m, connected by elastic springs of stiffness k_S and spaced at equal distances Δx in 1D.

the *mono-atomic* or **monatomic chain**. Assume that the ith particle is located at $x_i = i \cdot \Delta x$, where Δx is the initial, equilibrium spacing between particles, and the displacement of the ith particle is denoted by u_i. The equation of motion of the ith particle is

$$m\ddot{u}_i = k_S \left(u_{i+1} - 2u_i + u_{i-1}\right). \tag{11.5}$$

When we insert the ansatz for a travelling wave,

$$u(x,t) = \hat{u} \exp\left[i(kx - \omega t)\right] \tag{11.6}$$

with some amplitude $\hat{u} \in \mathbb{R}$, we arrive at

$$-m\omega^2 \hat{u} e^{i(kx-\omega t)} = k_S \left(e^{ik\,\Delta x} - 2 + e^{-ik\,\Delta x}\right) \hat{u} e^{i(kx-\omega t)} \quad \forall\ x,t. \tag{11.7}$$

Exploiting Euler's identity, the above arrive may be transformed into

$$-m\omega^2 = 2k_S \left[\cos(k\,\Delta x) - 1\right] = -4k_S \sin^2\left(\frac{k\,\Delta x}{2}\right), \tag{11.8}$$

so that

$$|\omega| = \sqrt{\frac{4k_S}{m}} \left|\sin\left(\frac{k\,\Delta x}{2}\right)\right|. \tag{11.9}$$

This is the so-called **dispersion relation** of the discrete elastic chain, which links the wave number k to the frequency ω and is illustrated in Figure 11.2. We observe that the relation is no longer linear as in the case of a continuous, homogeneous elastic bar (cf. (11.4)).

In the so-called **long-wavelength limit** $k \to 0$ (which is equivalent to the wavelength $\lambda \to \infty$), we recover a linear relation as in (11.4), since

$$\lim_{k\to 0} |\omega| = \sqrt{\frac{4k_S}{m}} \lim_{k\to 0} \left|\sin\left(\frac{k\,\Delta x}{2}\right)\right| = \sqrt{\frac{4k_S}{m}} \left|\frac{k\,\Delta x}{2}\right| = \sqrt{\frac{k_S}{m}} \Delta x |k| = c_{\text{eff}} |k| \tag{11.10}$$

with $c_{\text{eff}} = \Delta x \sqrt{k_S/m}$ acting as an effective wave speed in the discrete chain. As q increases, the dispersion relation (11.9) deviates increasingly from the linear behavior of the elastic bar, given by (11.4). In particular, ω will saturate at $k = \pi/\Delta x$, where it reaches a maximum frequency of

$$\omega_{\max} = \sqrt{\frac{4k_S}{m}}. \tag{11.11}$$

That is, the chain does not support the propagation of any signals at frequencies above $\omega_{\max}$: if any particle in the chain is excited at a frequency $\omega > \omega_{\max}$, then this signal cannot propagate due to **dispersion** in the discrete chain (within a nonlinear setting, it could excite lower frequencies, which may be allowed to propagate).

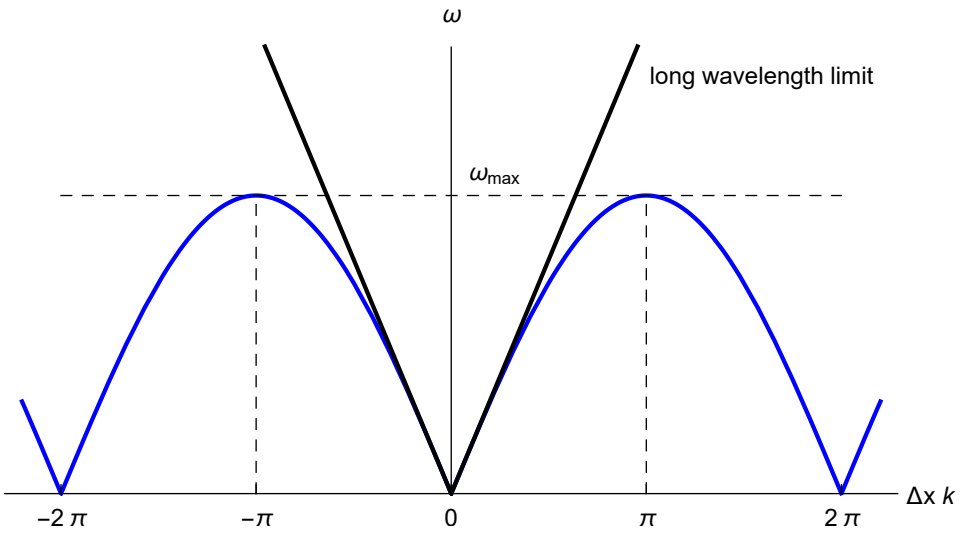


Figure 11.2: Dispersion relation of the monatomic chain, illustrating the relation (11.9) between frequency ω and wavenumber k.

By inspecting the dispersion relation (11.9), which holds for each particle in the 1D chain, we notice that ω is periodic in k. Moreover, all values of ω are contained in the range

$$-\frac{\pi}{\Delta x} \leq k \leq \frac{\pi}{\Delta x}, \tag{11.12}$$

which defines the so-called **first Brillouin zone**. Note that, by double-symmetry, the interval $[0, \pi/\Delta x]$ contains all the required information, which is known as the **Irreducible Brillouin Zone** (IBZ).

Let us also introduce two important velocities. The **group velocity** defines the velocity of a *wave pulse* moving with average frequency ω and average wave vector k and is defined by

$$v_\mathrm{g} = \frac{\partial \omega}{\partial k}. \tag{11.13}$$

Loosely speaking, v_g represents the speed at which vibrational energy flows for a bundle of (nearby) frequencies. At $k = \pi/\Delta x$, the group velocity vanishes, so the solution corresponds to a standing wave.

In addition, we define the **phase velocity** as the velocity of a *pure wave* of frequency ω and wave vector k, defined by

$$v_\mathrm{p} = \frac{\omega}{k}. \tag{11.14}$$

In the long-wavelength limit ($k \to 0$)

$$v_\mathrm{g} \to v_\mathrm{p} \qquad \text{as} \qquad k \to 0, \tag{11.15}$$

whereas both velocities differ for non-zero values of the wavenumber k.

As an illustration of *group vs. phase velocities*, let us consider a harmonic wave moving with a frequency ω and wave vector k, defined by

$$u(x,t) = \cos(\omega t - kx) = \cos\left[-k\left(x - \frac{\omega}{k}t\right)\right]. \tag{11.16}$$

From the above rearrangement of the equation, it becomes clear that every point on the wave travels with a velocity $v_\mathrm{p} = \frac{\omega}{k}$; this is the **phase velocity**. Next, imagine that we superimpose two such waves with distinct values of ω and k, i.e., we consider

$$\begin{aligned} u(x,t) &= \cos(\omega_1 t - k_1 x) + \cos(\omega_2 t - k_2 x) \\ &= 2\cos\left(\frac{\omega_1-\omega_2}{2}t - \frac{k_1-k_2}{2}x\right)\cos\left(\frac{\omega_1+\omega_2}{2}t - \frac{k_1+k_2}{2}x\right), \end{aligned} \tag{11.17}$$

where we invoked the trigonometric addition theorem. Now, let us rewrite the two frequencies and wave vectors in terms of their average and deviations. That is, we define

$$\omega_1 = \omega_\mathrm{ave} + \Delta\omega, \quad \omega_2 = \omega_\mathrm{ave} - \Delta\omega \qquad \text{and} \qquad k_1 = k_\mathrm{ave} + \Delta k, \quad k_2 = k_\mathrm{ave} - \Delta k. \tag{11.18}$$

Insertion into (11.17) yields

$$u(x,t) = 2\cos\left(\Delta\omega\, t - \Delta k\, x\right)\cos\left(\omega_\mathrm{ave} t - k_\mathrm{ave} x\right). \tag{11.19}$$

This is a wave moving with $v_\mathrm{g} = \omega_\mathrm{ave}/k_\mathrm{ave}$, modulated by the first term. Especially if the two waves are quite similar, i.e., if assume $|\Delta\omega|, |\Delta k| \ll 1$, the first term above implies much longer wavelengths than the second (see Fig. 11.3). This *breathing* behavior is characterized by the **group velocity**, and – from inspecting $\cos\left(\Delta\omega\, t - \Delta k\, x\right)$ in analogy to (11.16) – is given by

$$v_\mathrm{g} = \frac{\Delta\omega}{\Delta k} \to \frac{\mathrm{d}\omega}{\mathrm{d}k}. \tag{11.20}$$

This is illustrated in a simplified manner in Fig. 11.3. Evaluating the wave profile (11.16) at time $t = 0$ yields $u(x,t) = \cos(-kx)$. Choosing, e.g., $k_1 = -1$ and $k_2 = -1.1$ leads to the two waves shown in Fig. 11.3a, whose addition $\cos(x) + \cos(1.1x) = 2\cos(0.05x)\cos(1.05x)$ is a wave with a large "breathing" profile, i.e., with a modulating amplitude of frequency 0.05 (indicated by the gray dashed lines in Fig. 11.3b). The closer the two k-values are, the longer the modulation wavelength. Fig. 11.3c shows the analogous outcome for $k_1 = 1$ and $k_2 = 1.2$.

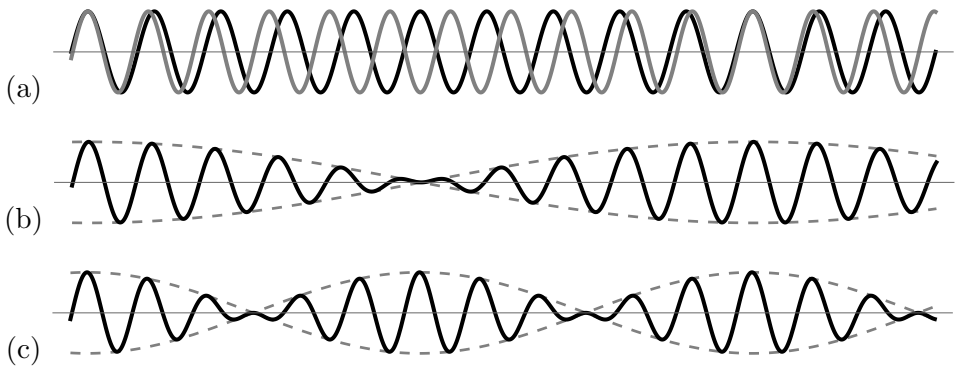


Figure 11.3: (a) Two waves of similar frequencies: shown are the functions $\cos(x)$ and $\cos(1.1x)$ in black and gray, respectively. (b) Their addition leads to a "breathing" wave with the average wavelength of the two waves and a modulation amplitude $\pm 2\cos(0.05x)$, which is indicated by the gray dashed lines. (c) The superposition of $\cos(x)$ and $\cos(1.2x)$ with a modulation amplitude $\pm 2\cos(0.1x)$.

11.1.3 Heterogeneous Medium: the Diatomic Chain

As a simple extension, let us consider the so-called **diatomic chain** made of particles of different, alternating masses m_1 and m_2, still connected by identical springs of stiffness k_S and at equal distances Δx (see Figure 11.1). We repeat the derivation of the dispersion relations for this example of a heterogeneous discrete 1D system.

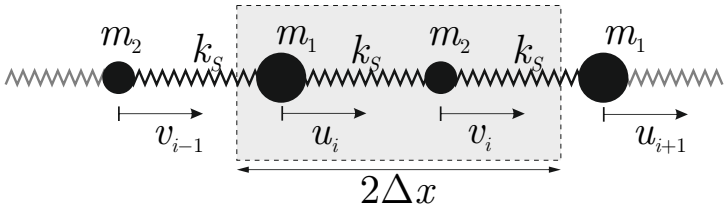


Figure 11.4: Diatomic chain of alternating particles of masses m_1 and m_2, connected by elastic springs of equal stiffness k and spaced at equal distances Δx in 1D.

In this setting, we must account for the fact that each *unit cell* of the chain is composed of two atoms (the smallest repeating unit), so we have $2\Delta x$-periodicity. Let us denote by u_i the displacements of all masses m_1 and by v_i the displacements of all masses m_2, so that $\{u_i, v_i\}$ are the degrees of freedom of the ith unit cell. The respective equations of motion of masses m_1 and m_2 inside that unit cell are given by

$$\begin{aligned} m_1\ddot{u}_i &= k_S\left(v_i - 2u_i + v_{i-1}\right), \\ m_2\ddot{v}_i &= k_S\left(u_{i+1} - 2v_i + u_i\right). \end{aligned} \tag{11.21}$$

The ansatz for a propagating wave must also account for the $2\Delta x$-periodicity, which is why we assume a traveling wave solution of the type

$$u(x,t) = \hat{u}\exp\left[i(kx-\omega t)\right] \qquad \text{and} \qquad v(x,t) = \hat{v}\exp\left[i(kx-\omega t)\right]. \tag{11.22}$$

Insertion into the equations of motion (and division by the common exponential term) yields

$$\begin{aligned} -\omega^2 m_1\hat{u} &= -2k_S\hat{u} + k\left(\hat{v} + \hat{v}e^{-2ik\,\Delta x}\right), \\ -\omega^2 m_2\hat{v} &= -2k_S\hat{v} + k\left(\hat{u}_i e^{2ik\,\Delta x} + \hat{u}\right). \end{aligned} \tag{11.23}$$

The non-trivial solution to this system of equations,

$$\begin{pmatrix} 2k_S - \omega^2 m_1 & -k_S\left(1+e^{-2ik\,\Delta x}\right) \\ -k_S\left(1+e^{2ik\,\Delta x}\right) & 2k_S - \omega^2 m_2 \end{pmatrix}\begin{pmatrix}\hat{u}\\ \hat{v}\end{pmatrix} = \mathbf{0}, \tag{11.24}$$

is found from the characteristic equation:

$$\det\begin{pmatrix} 2k_S - \omega^2 m_1 & -k_S\left(1+e^{-2ik\,\Delta x}\right) \\ -k_S\left(1+e^{2ik\,\Delta x}\right) & 2k_S - \omega^2 m_2 \end{pmatrix} = 0. \tag{11.25}$$

Simplification of the determinant leads to

$$m_2 m_2\omega^4 - 2k_S(m_1+m_2)\omega^2 + 2k_S^2\left[1-\cos(2k\Delta x)\right] = 0, \tag{11.26}$$

which has solutions (assuming the ordering $\omega_1 \leq \omega_2$ in the following)

$$\omega_{1/2}^2 = k_S\left(\frac{1}{m_1}+\frac{1}{m_2}\right) \pm k_S\sqrt{\left(\frac{1}{m_1}+\frac{1}{m_2}\right)^2 - \frac{4\sin^2(k\Delta x)}{m_1 m_2}}. \tag{11.27}$$

Eq. (11.27) serves as the *dispersion relation for the diatomic chain*. We notice that both solutions of ω are periodic on the interval $[-\pi/(2\Delta x), \pi/(2\Delta x)]$, so all values of ω are covered by this *first Brillouin zone* (the size of the RVE doubled, so the size of the Brillouin zone decreases by a factor of 2). The maximum supported frequency here is found for $k = 0$ at

$$\omega_{\max} = \sqrt{2k_S\left(\frac{1}{m_1}+\frac{1}{m_2}\right)}, \tag{11.28}$$

while the long-wavelength limit is again recovered as $k \to 0$ for the lower of the two solutions $\omega_{1/2}$, while the higher solution attains $\omega_{\max}$ at $k = 0$.

Since we have two solutions $\omega_{1/2}$, the dispersion plot in Figure 11.5 reveals interesting features. Those frequency ranges for which no frequency ω corresponds to any value of k are called **stop bands** or **band gaps**, while those frequency ranges covered by the solutions $\omega_{1/2}$ are known as **pass bands**. In this example, we identify a band gap across the range of frequencies bounded by ω_1 and ω_2 at $k = \pi/(2\Delta x)$, which evaluates to

$$\left[\sqrt{2k_S/m_1}, \sqrt{2k_S/m_2}\right] \qquad \text{for} \qquad m_2 > m_1. \tag{11.29}$$

This band gap divides the dispersion plot into two *branches*, identified by the two solutions $\omega_{1/2}$ (assuming $m_2 > m_1$) : the lower band, covering frequencies within $[0, \sqrt{2k_S/m_1}]$ is known as the **acoustic branch**. The upper band, covering frequencies within $[\sqrt{2k_S/m_2}, \omega_{\max}]$ is the so-called **optical branch**. The terminology stems from phonons in crystal lattices, where this concept was first introduced. In fact, a 1D atomic lattice, in which each atom vibrates at small amplitude about its ground state, is nothing but an elastic chain of particles (possibly with more complex inter-particle interactions not necessarily confined to nearest neighbors). In such a crystalline setting, the elastic wave modes investigated here govern the transport of energy through the crystal. A **phonon** is a *quasi-particle* of vibrational mechanical energy in such a lattice. *Acoustic* phonons are defined through linear relation in the long-wavelength limit (as observed here for the lower branch in the dispersion plot). *Optical* phonons have a non-zero frequency at the Brillouin zone's center ($\omega_{\max}$ in our case), and they exhibit no dispersion close to the long-wavelength limit (the slope near $\omega_{\max}$ is zero, so all wave vectors propagate at the same frequency close to $k = 0$ on the optical branch). While the acoustic branch's group velocity approaches the speed of sound in the long-wavelength limit, the group velocity of the optical branch vanishes in the long-wavelength limit (implying a *standing wave*).

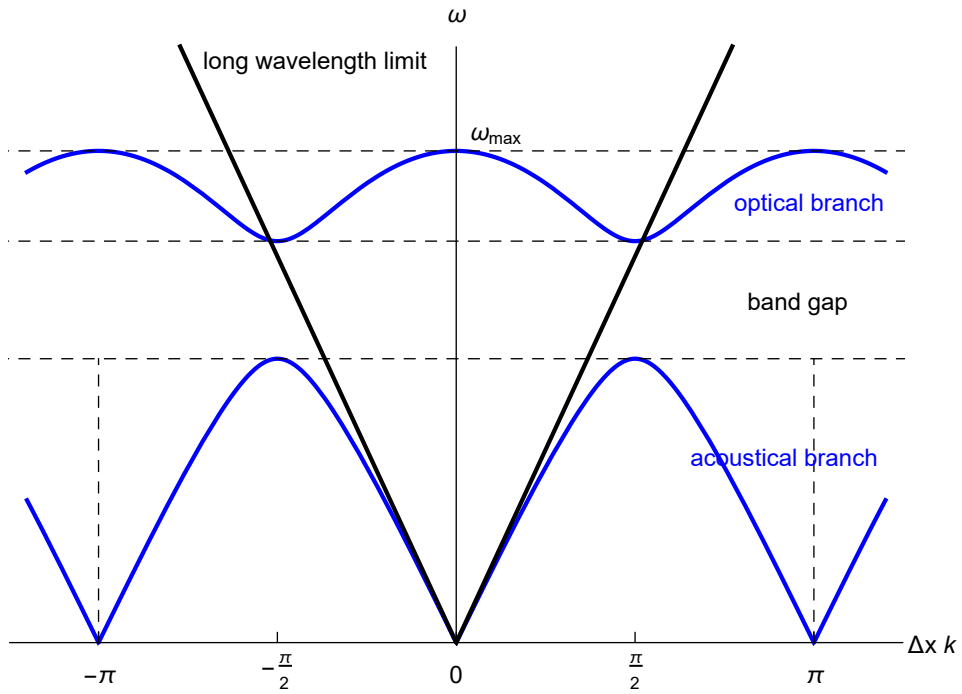


Figure 11.5: Dispersion relation of the diatomic chain, illustrating the relation (11.27) between ω and k (plotted for $m_2 = 2m_1$. The acoustic branch and optical branch are identified by, respectively $\omega_1(k)$ and $\omega_2(k)$.

The two branches can further be distinguished by the eigenmodes associated with the two eigenfrequencies $\omega_{1/2}$. Solving the eigenvalue problem (11.24) yields the following pairs of eigenvalues/-vectors (assuming the ordering $\omega_1 \leq \omega_2$):

$$\omega_1 \;:\; \begin{pmatrix}\hat{u}\\ \hat{v}\end{pmatrix} = \begin{pmatrix}1\\ 1\end{pmatrix}, \qquad \omega_2 \;:\; \begin{pmatrix}\hat{u}\\ \hat{v}\end{pmatrix} = \begin{pmatrix}1\\ -1\end{pmatrix}. \tag{11.30}$$

That is, in the acoustic branch the two particles within a unit cell vibrate *in phase*, whereas in the optical branch the two particles vibrate *out of phase*.

11.1.4 Wave Dispersion in Higher Dimensions

When going to higher dimensions, the same concepts apply as discussed above for 1D wave motion, yet the complexity increases.[6] Computing dispersion relations is an important tool to understand wave motion in heterogeneous media, which is commonly achieved by so-called **Bloch-Floquet analysis** (which was named after its two prominent inventors and originated from studying harmonic atomic crystals). We here provide only a brief introduction to this complex topic.

Let us consider a heterogeneous RVE (composed of either a discrete structure such as in a truss or particles connected by springs, or a continuous medium such as a composite or any other microstructured solid). To obtain the dispersion relations, we study linear waves, which is why we assume a time-dependent displacement field of the general form

$$\boldsymbol{u}(\boldsymbol{x},t) = \hat{\boldsymbol{u}}(\boldsymbol{x})\exp(-i\omega t) \tag{11.31}$$

with a propagation frequency $\omega \in \mathbb{R}$. For a *periodic microstructure*, one typically studies so-called **Bloch waves**, which are defined by (see Figure 11.6)

$$\hat{\boldsymbol{u}}(\boldsymbol{x}) = \hat{\boldsymbol{u}}_0(\boldsymbol{x}_0)\exp\left[-i\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{x}_0)\right], \tag{11.32}$$

where $\boldsymbol{k}$ denotes the wave vector. $\boldsymbol{x}_0 \in \Omega_0$ is a material point located within a reference RVE Ω_0, which is linked to point $\boldsymbol{x}$ by periodicity, i.e., we assume

$$\boldsymbol{x} = \boldsymbol{x}_0 + \sum_{a=1}^{d} c_a \boldsymbol{L}_a \qquad \text{with} \qquad \boldsymbol{x}_0 \in \Omega_0. \tag{11.33}$$

In the long-wavelength limit ($|\boldsymbol{k}| \to 0$), the Bloch wave ansatz (11.32) implies periodicity of the solution across all periodic unit cells (i.e., the same displacements in every unit cell in the

[6] For illustrative examples, try the online dispersion relation calculator for 2D material structures available here: https://henriquemiranda.github.io/phononwebsite/phonon.html.

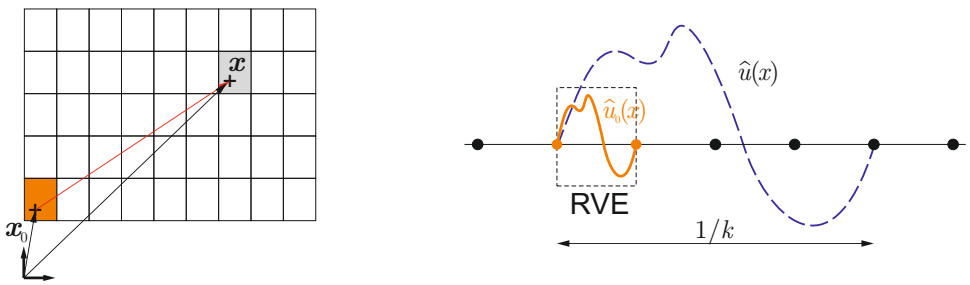


Figure 11.6: Schematic illustration of the link between points in the reference RVE and all other RVEs, and of the modulated wave profile.

periodic medium). For wave vectors $\boldsymbol{k} \neq \boldsymbol{0}$, (11.32) implies a modulated solution – i.e., the displacements within the reference RVE Ω_0 are multiplied by $\boldsymbol{k}$-dependent harmonic functions to produce the complete displacement field in the periodic body. Either way, displacements $\hat{\boldsymbol{u}}_0(\boldsymbol{x})$ within Ω_0 along with $\boldsymbol{k}$ uniquely determine the complete displacement field within the periodic body, which is why in the following we will need to solve for $\hat{\boldsymbol{u}}_0(\boldsymbol{x}_0)$ inside the reference RVE Ω_0. Since this is rarely possible analytically (unless in simple toy problems such as the 1D elastic chains discussed above), we turn to a finite element description: the reference mesh Ω_0 is discretized by an FE mesh of n nodes, whose nodal displacements $\boldsymbol{u}^a$ $(a = 1, \dots, n)$ are the unknowns to be determined for any frequency ω.

For a linear(ized) elastic medium, the governing equations are the linear equations of motion:

$$\boldsymbol{M}\ddot{\boldsymbol{U}} + \boldsymbol{K}\boldsymbol{U} = \boldsymbol{F}, \tag{11.34}$$

with $\boldsymbol{M}$ and $\boldsymbol{K}$ denoting the global mass and stiffness matrices, respectively, $\boldsymbol{F}$ is the global external force vector, and $\boldsymbol{U} = \{\boldsymbol{u}^1, \dots, \boldsymbol{u}^n\}$ is the global displacement vector of all nodal degrees of freedom. Note that we here assume the absence of any damping or other sources of dissipation (otherwise, any wave will decay with time). In a nonlinear setting, (11.34) is obtained by linearizing the nonlinear equations of motion about the equilibrium configuration.

Application of

$$\boldsymbol{U}(\boldsymbol{x}, t) = \hat{\boldsymbol{U}}(\boldsymbol{x}) \exp(-i\omega t) \qquad \text{and} \qquad \boldsymbol{F}(\boldsymbol{x}, t) = \hat{\boldsymbol{F}} \exp(-i\omega t) \tag{11.35}$$

yields

$$(\boldsymbol{K} - \omega^2 \boldsymbol{M})\hat{\boldsymbol{U}} = \hat{\boldsymbol{F}}. \tag{11.36}$$

For a free-standing body, the absence of external forces $\hat{\boldsymbol{F}} = \boldsymbol{0}$ turns the above into an eigenvalue problem, whose solution yields the eigenfrequencies and associated eigenmodes for a free vibration.

Here, by contrast, we apply the Bloch wave ansatz, which assumes a periodic tesselation of RVEs, whose boundary conditions are dictated by the assumed wave form (11.32). By application to (11.36), the RVE problem now becomes

$$(\boldsymbol{K} - \omega^2 \boldsymbol{M})\hat{\boldsymbol{U}}_0 = \hat{\boldsymbol{F}}_0 \tag{11.37}$$

with

$$\hat{\boldsymbol{U}}_0^+ = \hat{\boldsymbol{U}}_0^- \exp\left[-i\boldsymbol{k} \cdot (\boldsymbol{x}^+ - \boldsymbol{x}^-)\right] \qquad \text{and} \qquad \hat{\boldsymbol{F}}_0^+ = -\hat{\boldsymbol{F}}_0^- \exp\left[i\boldsymbol{k} \cdot (\boldsymbol{x}^+ - \boldsymbol{x}^-)\right]. \tag{11.38}$$

This problem can be solved in multiple ways. Notice the similarity between the above relations and the RVE problem with periodic boundary conditions, discussed in Section 7. Therefore, displacements were linked by periodicity and forces at opposite nodes were of equal magnitude and opposite sign. Here, displacements are linked through the Bloch wave ansatz. Because of inertial contributions to the moving wave, forces are no longer the same on opposite surfaces but linked through the second relation in (11.38).

The so-called **transformation matrix** approach to solving the above system, separates the displacements of internal nodes $\boldsymbol{U}_\mathrm{i}$, master boundary nodes $\boldsymbol{U}_\mathrm{m}$, and slave boundary nodes $\boldsymbol{U}_\mathrm{s}$. Based on those categories, we introduce a permutation matrix $\boldsymbol{A}$ (with $\boldsymbol{A}^\mathrm{T}\boldsymbol{A} = \boldsymbol{I}$), such that the global vector of all displacements, $\hat{\boldsymbol{U}}_0$ in the reference RVE Ω_0, is decomposed according to

$$\boldsymbol{A}\hat{\boldsymbol{U}}_0 = \begin{bmatrix} \boldsymbol{U}_\mathrm{i} \\ \boldsymbol{U}_\mathrm{m} \\ \boldsymbol{U}_\mathrm{s} \end{bmatrix}. \tag{11.39}$$

In order to separate the real from the imaginary parts of displacements (most FE codes do not operate on complex numbers), we further introduce a boundary-condition matrix $\boldsymbol{Q}$, which depends on wave vector $\boldsymbol{k}$ and relates the displacements of master and slave boundary nodes according to the Bloch wave assumption:

$$\begin{bmatrix} \boldsymbol{U}_\mathrm{i}^\mathrm{Re} \\ \boldsymbol{U}_\mathrm{m}^\mathrm{Re} \\ \boldsymbol{U}_\mathrm{s}^\mathrm{Re} \\ \boldsymbol{U}_\mathrm{i}^\mathrm{Im} \\ \boldsymbol{U}_\mathrm{m}^\mathrm{Im} \\ \boldsymbol{U}_\mathrm{s}^\mathrm{Im} \end{bmatrix} = \boldsymbol{Q}(\boldsymbol{k}) \begin{bmatrix} \boldsymbol{U}_\mathrm{i}^\mathrm{Re} \\ \boldsymbol{U}_\mathrm{m}^\mathrm{Re} \\ \boldsymbol{U}_\mathrm{i}^\mathrm{Im} \\ \boldsymbol{U}_\mathrm{m}^\mathrm{Im} \end{bmatrix}. \tag{11.40}$$

The analogous relations apply for the external force vectors $\hat{\boldsymbol{F}}_0$. Therefore, problem (11.37) is transformed into

$$\boldsymbol{Q}^\mathrm{T}(\boldsymbol{k}) \left(\begin{bmatrix} \boldsymbol{K} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{K} \end{bmatrix} - \omega^2 \begin{bmatrix} \boldsymbol{M} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{M} \end{bmatrix} \right) \begin{bmatrix} \boldsymbol{A}^\mathrm{T} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}^\mathrm{T} \end{bmatrix} \boldsymbol{Q}(\boldsymbol{k}) \begin{bmatrix} \hat{\boldsymbol{U}}_\mathrm{i}^\mathrm{Re} \\ \hat{\boldsymbol{U}}_\mathrm{m}^\mathrm{Re} \\ \hat{\boldsymbol{U}}_\mathrm{i}^\mathrm{Im} \\ \hat{\boldsymbol{U}}_\mathrm{m}^\mathrm{Im} \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{F}}_\mathrm{i}^\mathrm{Re} \\ \hat{\boldsymbol{F}}_\mathrm{m}^\mathrm{Re} \\ \hat{\boldsymbol{F}}_\mathrm{i}^\mathrm{Im} \\ \hat{\boldsymbol{F}}_\mathrm{m}^\mathrm{Im} \end{bmatrix}. \tag{11.41}$$

For wave propagation starting from the undeformed ground state (which we usually assume), the right-hand side vanishes. Note that the real and imaginary systems of equations are coupled through $\boldsymbol{Q}$, which can be enforced in practice by introducing kinematic constraints between the two separate (real and imaginary) eigenvalue problems to be solved simultaneously for the set of frequencies ω for a given $\boldsymbol{k}$.

As in any eigenvalue problem, we can identify as many frequencies ω for a given $\boldsymbol{k}$ as there are degrees of freedom in the RVE. Hence, the strategy is as follows: pick a wave vector $\boldsymbol{k}$ and solve the eigenvalue problem(s) (11.41) for the set of eigenfrequencies $\boldsymbol{\omega} = \{\omega_1, \dots, \omega_n\}$. Repeating this step for all wave vectors $\boldsymbol{k}$ yields the complete dispersion relation between $\boldsymbol{k}$ and $\boldsymbol{\omega}$. Of course, as in the 1D examples we do not have to check all possible wave vectors since

$$\exp\left[i(\boldsymbol{k}\cdot\boldsymbol{x} + 2\pi n)\right] = \exp\left[i\boldsymbol{k}\cdot\boldsymbol{x}\right] \qquad \forall\, n \in \mathbb{Z}. \tag{11.42}$$

Therefore, we can restrict the choice of $\boldsymbol{k}$ to vectors lying in the **first Brillouin Zone**, i.e., the primitive cell in reciprocal space. By exploiting symmetry, one can often reduce the space of $\boldsymbol{k}$ further to the **Irreducible Brillouin Zone**.

11.2 Example: Wave Dispersion in 2D Truss Lattices

As an example application of the above concepts, let us consider a discrete periodic truss in 2D, made by periodically repeating a 2D unit cell composed of linear elastic slender struts. In this case, the RVE is identified as the unit cell of the periodic truss, and the eigenvalue problem(s) (11.41) are solved, based on an FE discretization using Euler-Bernoulli beam elements to describe individual struts. For ease of comparison, the following examples will report the *normalized frequency* and *wave vector*, defined as, respectively,

$$\bar{f} = \frac{\omega L}{2\pi c}, \qquad \text{and} \qquad \bar{\boldsymbol{k}} = L\boldsymbol{k}, \tag{11.43}$$

where L is a characteristic length of the unit cell (e.g., the strut length), and $c = \sqrt{E/\rho}$ is the wave speed in each linear elastic bar.

During vibrational wave motion, struts in 2D lattices undergo longitudinal and flexural motion both within and out of the plane spanned by the lattice, leading to *in-plane* and *out-of-plane wave modes*. To differentiate between those two vibrational modes in the dispersion relations, we categorize mode shapes $\hat{\boldsymbol{U}}_0$ as *in-plane* vs. *out-of-plane* as follows. Starting with an eigenvector $\hat{\boldsymbol{U}}_0$ (defined up to a constant), we apply mass normalization, i.e., we transform an eigenvector $\hat{\boldsymbol{U}}_0$ into (realizing that mass matrix $\boldsymbol{M}$ is positive-definite)

$$\tilde{\boldsymbol{U}}_0 = \frac{\hat{\boldsymbol{U}}_0}{\hat{\boldsymbol{U}}_0 \cdot \boldsymbol{M}\hat{\boldsymbol{U}}_0}. \tag{11.44}$$

We further introduce projection matrices $\boldsymbol{P}_1$, $\boldsymbol{P}_2$ and $\boldsymbol{P}_3$ such that

$$\boldsymbol{\phi}_i = \boldsymbol{P}_i\tilde{\boldsymbol{U}}_0 \tag{11.45}$$

contains only the displacement components in the x_i-coordinate direction. The coordinate system is chosen such that the lattice lies in the (x_1, x_2)-plane, while x_3 is perpendicular to the lattice plane. Based on the above, we define mass-weighted directional contributions p_i related to (twice) the associated kinetic energy via (no summation over i implied)

$$p_i = \boldsymbol{\phi}_{(i)} \cdot \boldsymbol{M}\boldsymbol{\phi}_{(i)}. \tag{11.46}$$

To differentiate between in-plane and out-of-plane modes, we calculate a scalar *in-plane ratio*

$$\beta = \frac{\sqrt{p_1^2 + p_2^2}}{\sqrt{p_1^2 + p_2^2 + p_3^2}} \in [0,1] \qquad \Rightarrow \qquad \beta \approx \begin{cases} 1 & \text{for in-plane modes,} \\ 0 & \text{for out-of-plane modes.} \end{cases} \tag{11.47}$$

Therefore, $\beta \in [0,1]$ represents the ratio of the wave vector projected into the lattice plane to the length of the wave vector (so that $\beta = 1$ implies the wave vector lies in the plane, whereas $\beta = 0$ implies that wave motion occurs out-of-plane).

Figure 11.7 shows the simple example of a periodic truss made of rectangular unit cells of aspect ratio γ. The unit cell is bounded by 2×2 trusses, which translates into a rectangular first Brillouin zone in reciprocal space. By symmetry, the Irreducible Brillouin Zone (IBZ) is given by one quarter of the first Brillouin zone, so that only wave vectors $\boldsymbol{k}$ within the IBZ must be investigated – the solution for any other wave vectors is contained within those solutions for $\boldsymbol{k}$-vectors from within the IBZ.

Bloch-Floquet analysis, as outlined above, leads to the dispersion relations, which are illustrated in Figure 11.8: to each (normalized) wave vector $\boldsymbol{k}$, there corresponds a set of (normalized)

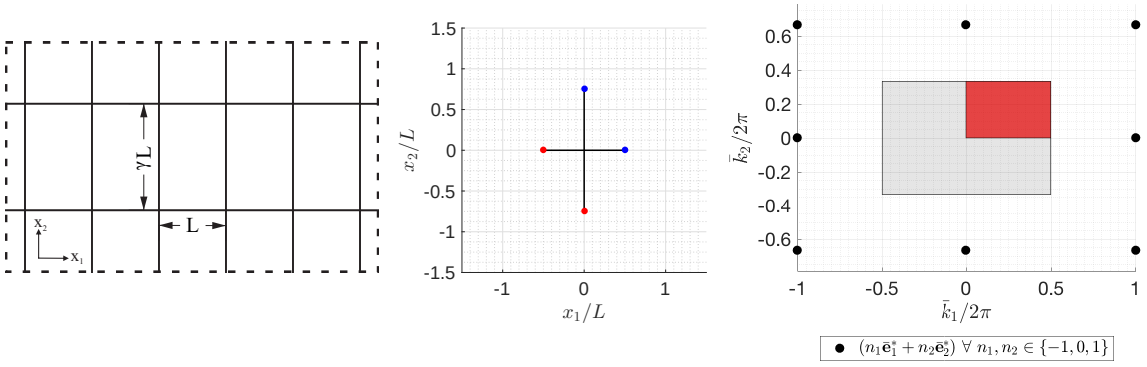


Figure 11.7: A periodic truss composed of elastic beams arranged in rectangular unit cells of aspect ratio $\gamma > 0$, along with its first Brillouin zone (a single rectangular unit cell) and the reciprocal space of 2D wave vectors $\boldsymbol{k}$ to be studied (shown in gray, with the IBZ highlighted in red); adopted from ref. [104]

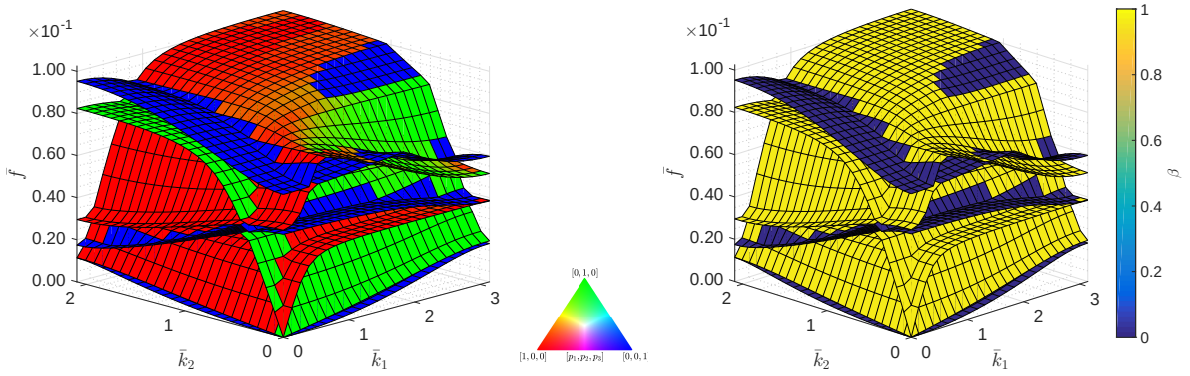


Figure 11.8: Dispersion relations for a 2D rectangular truss lattice: the lowest few dispersion surfaces are color-coded by the directional contributions p_i; the same surfaces on the right are color-coded by the in-plane ratio β, distinguishing in-plane from out-of-plane vibrational modes; adopted from ref. [104]

frequencies f, which solve (11.41). Only the lowest few dispersion surfaces are shown. By calculating the in-plane ratio β (also shown in Figure 11.8), we identify surfaces associated with in-plane and out-of-plane vibrational modes. The mode shapes of the lowest four dispersion surfaces are summarized in Figure 11.9, where the in-plane and out-of-plane modes are clearly identifiable. Note that for linear elastic beams the out-of-plane bending stiffness is typically lower than any in-plane stiffness, which is why the lowest mode corresponds to an out-of-plane mode. Also, recall the normalization $\bar{f} = \frac{\omega L}{2\pi c}$, which implies that the frequencies ω increase with decreasing strut size L (the smaller the unit cell, the higher the frequencies – this is of importance, e.g., when designing acoustic metamaterials to be exploited for their wave dispersion). The dispersion surface plots of the type in Figure 11.8 can be interpreted as the 2D extension of the 1D dispersion plots of Figure 11.5.

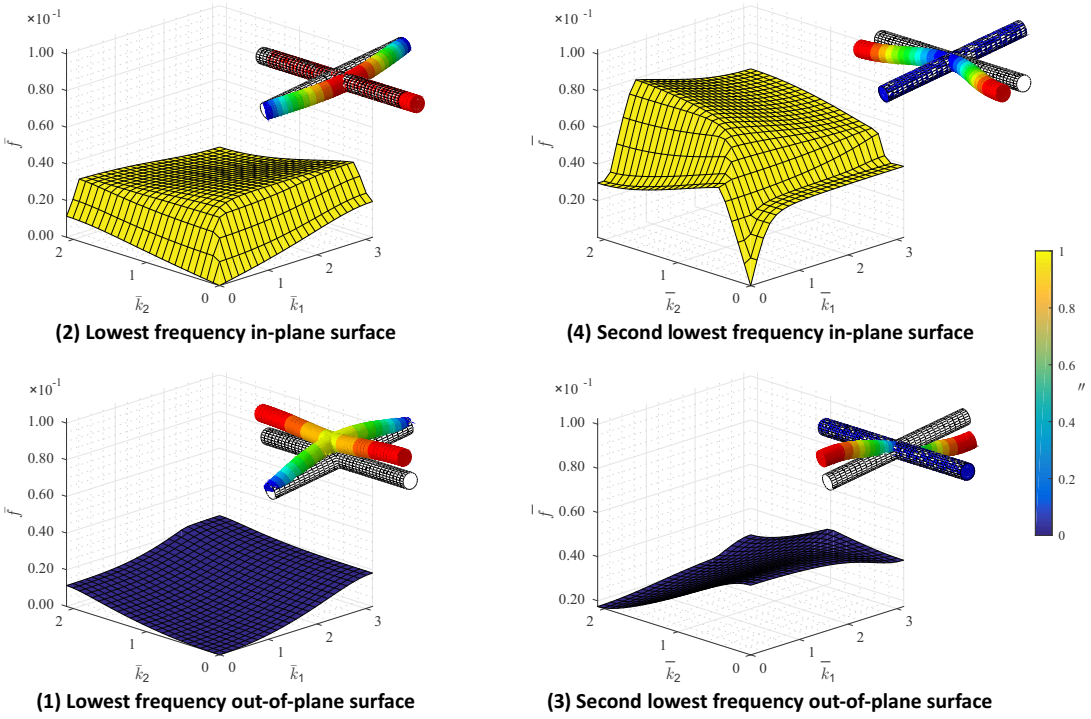


Figure 11.9: Vibrational modes of the lowest four dispersion surfaces of the rectangular truss lattice, showing in-plane and out-of-plane modes (displacements are of arbitrary magnitude; the dispersion plots are colored by parameter $\beta \in [0, 1]$); from ref. [104].

To understand the impact of the heterogeneous RVE (here, of the discrete structure) on the propagation of waves, one may favor another representation of the dispersion relations. Due to the overlapping and intersecting dispersion surfaces in Figure 11.8, it is hard to identify, e.g., band gaps, which are of frequent practical interest. Here, we briefly introduce two types of visualizations to *identify band gaps*. One common approach is to introduce a so-called $\boldsymbol{k}$-path, i.e., a piecewise-linear path through the IBZ that covers all directions. The rationale is that, if there is a band gap for certain $\boldsymbol{k}$-vectors within the IBZ, then it should appear when plotting the dispersion relations along the path in the IBZ. However, there is no proof; in fact, the opposite has been demonstrated for (a few constructed) examples, in which band gaps did not surface when plotting the dispersion relations along the $\boldsymbol{k}$-path only. An example from reference [50] is shown in Figure 11.10, which illustrates band gaps arising in a 2D metamaterial with the shown unit cell.

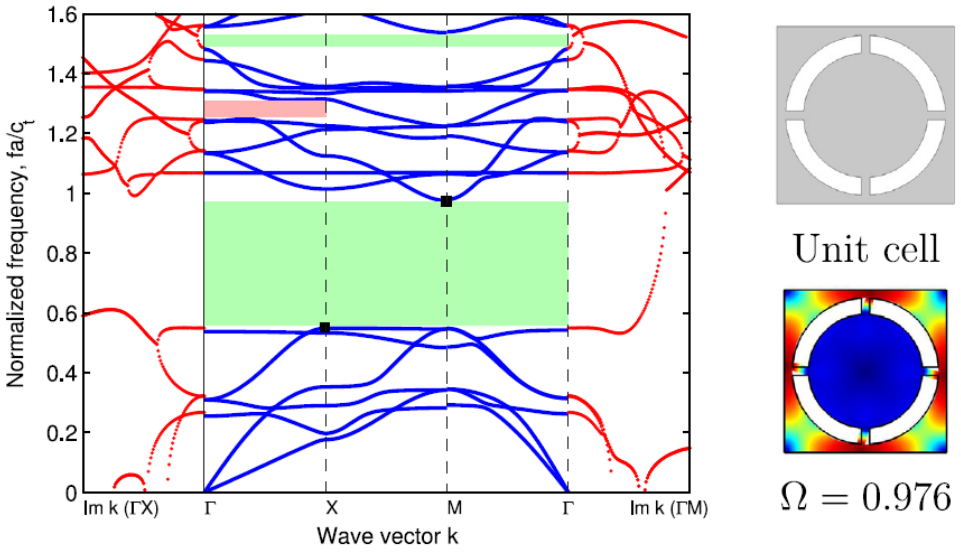
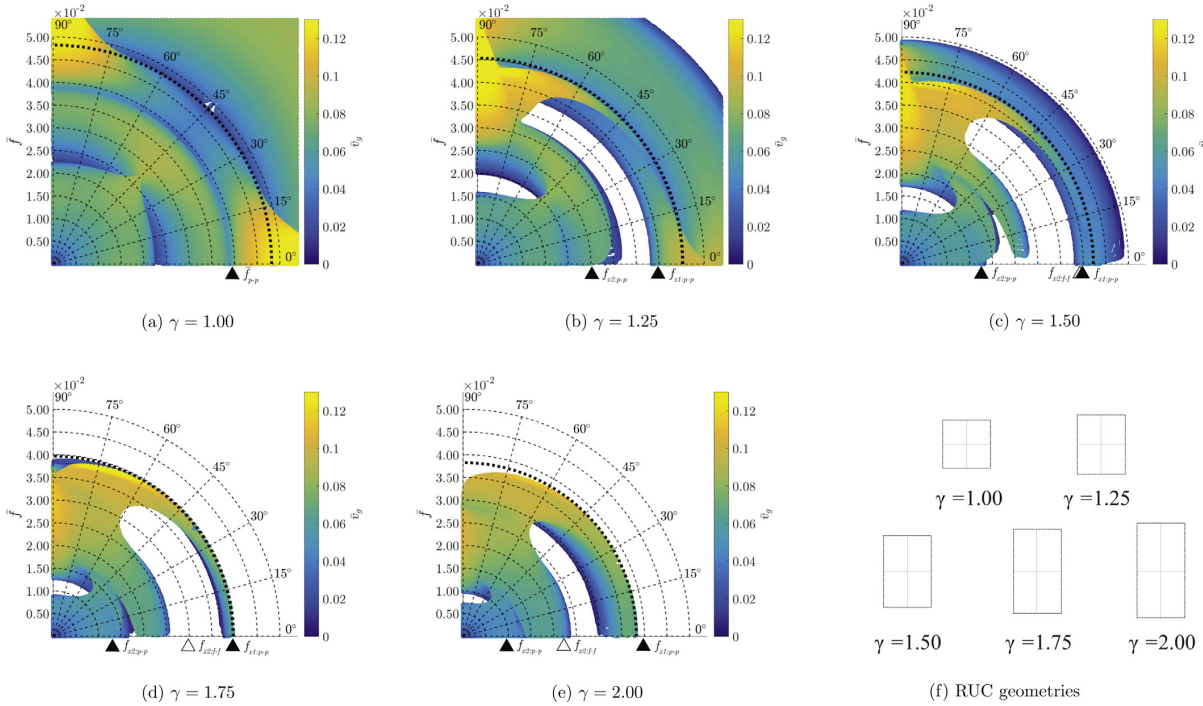


Figure 11.10: Dispersion relations and band gaps for a periodic 2D plate-type structure, as obtained from computing $\omega(\boldsymbol{k})$ along a $\boldsymbol{k}$-path throughout the IBZ. Letters Γ, X, M, stand for, respectively, the center of the RVE, the mid-point of the right edge of the RVE, and the top-right corner node of the RVE. That path $\Gamma - X - M - \Gamma$ thus parametrizes a triangular path through the IBZ; from reference [50].

Another approach is to study the propagation of energy in the periodic medium (which was introduced in reference [104]). From the dispersion surfaces in Figure 11.8, we can calculate the group velocity for each point $\boldsymbol{k}$ on a dispersion surface as

$$\boldsymbol{v}_\mathrm{g} = \frac{\partial \omega}{\partial \boldsymbol{k}}(\boldsymbol{k}) \tag{11.48}$$

with magnitude $v_\mathrm{g} = |\boldsymbol{v}_\mathrm{g}|$ and direction $\boldsymbol{v}_\mathrm{g}/|\boldsymbol{v}_\mathrm{g}|$. This direction represents the direction of energy flow. If we compute $\boldsymbol{v}_\mathrm{g}$ for every point on each dispersion surface and plot the resulting distribution of group velocities, we obtain the characteristic plots in Figure 11.11. Each group velocity obtained from Figure 11.8 corresponds to a point in the plot of Figure 11.11. The angle indicates the direction of $\boldsymbol{v}_\mathrm{g}$ and hence of energy flow, the radial coordinate represents the frequency at each the group velocity was computed. The result are the radial plots in Figure 11.11, whose color point illustrates the magnitude of the group velocity: dark blue implies no flow of energy (waves do not propagate at those frequencies into those directions), whereas light yellow denotes a high group velocity (and hence the fast propagation of waves in those directions at those frequencies). Band gaps appear as white regions, in which no wave propagates (no group velocity exists for those frequencies and directions in the white zones).



(a) $\gamma = 1.00$ (b) $\gamma = 1.25$ (c) $\gamma = 1.50$

(d) $\gamma = 1.75$ (e) $\gamma = 2.00$ (f) RUC geometries

Figure 11.11: Illustration of the (normalized) group velocity $v_\mathrm{g} = |\boldsymbol{v}_\mathrm{g}|$ (shown by the color code) as a function of the direction of energy flow (i.e., the direction of $\boldsymbol{v}_\mathrm{g}$) and of frequency ω. Shown is only the top right quadrant, the others follow by symmetry. Each plot corresponds to a different aspect ratio γ of the truss unit cell, showing the two lowest out-of-plane frequency surfaces. The black dashed line illustrates the lowest frequency of the third surface. For reference, solid and open triangles denote resonance frequencies of free-standing beams in their pinned-pinned and fixed-fixed beams, respectively. Where applicable, resonant frequencies of beams aligned with the x_1- and x_2-directions are shown as $f_{x_i:p-p}$ or $f_{x_i:f-f}$; from ref. [104].

While the square unit cell ($\gamma = 1$) displays directionality but no band gaps, these start to appear as soon as γ is increased, indicating that dispersion leads to directional-dependent wave motion. Figure 11.12 shows a validation by an explicit dynamic simulation of a rectangular truss lattice, which is excited at its center at small amplitude and at a fixed frequency. The obtained displacement field confirms the strong directional dependence of the wave speed in the 2D truss lattice, in compliance with the group velocity map.

Aside from its original application in atomic crystal lattices, Bloch-Floquet analysis has gained increasing interest over the past decades in the field of photonic, phononic, and acoustic **metamaterials**, which are designed to have tailored wave dispersion for use as wave guides, wave filters, signal processors, sensors, energy harvesters or energy absorbers, and more; see, e.g., reference [38] for a review of phononic metamaterials.

Band gaps usually arise due to **local resonance** (in which individual members of the periodic RVE are excited to resonance) or due to **Bragg scattering** (which arises when the wavelength and the path length in certain directions in the periodic medium are such that the recombination of waves leads to attenuation). Local resonance is commonly exploited in phononic crystals and many acoustic metamaterials: e.g., heavy particles embedded in a soft matrix can be excited to resonance, which absorbs kinetic energy (which may be dissipated through a surrounding viscoelastic matrix). This leads to metamaterials with band gap bands for wave attenuation at dedicated frequencies. Bragg scattering is more intricate to exploit, and a detailed discussion goes beyond the scope of this course.

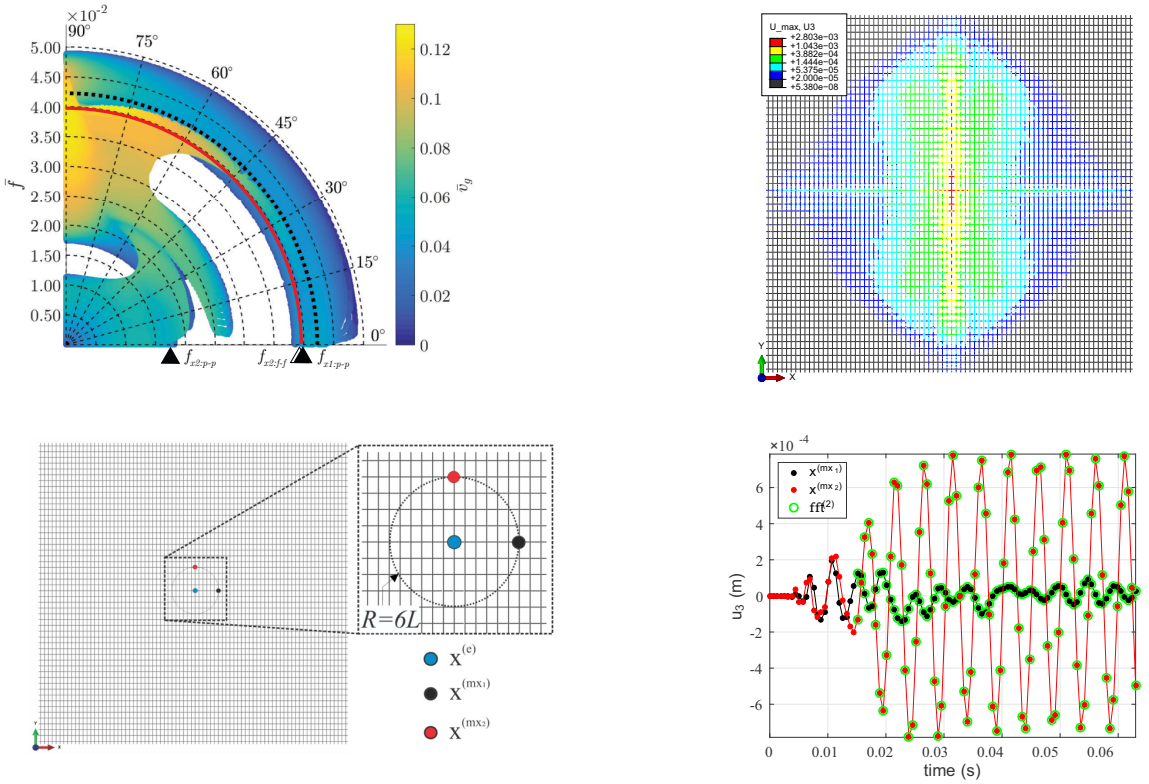


Figure 11.12: Comparison of the group velocity map with an explicit dynamic simulation of the rectangular truss lattice. The center-point of the truss (blue circle) is excited at the frequency indicated by the red circle in the group velocity map. The displacement profile (top right) clearly shows directional wave motion with significantly faster speed in the vertical than horizontal direction (the bottom right image shows the Fourier transform of the displacement signal at the black and red points in the truss); from ref. [104].

Part II:
Material Modeling from Particles to the Macroscale

12 From Quantum Mechanics to Atomistics

In the following sections of our course on *Multiscale Modeling*, we would like to give a brief overview of material modeling techniques across length and time scales – starting at the fundamental level of quantum mechanics and continuing to atomistics and molecular dynamics to finally arrive at methods that bridge from atomistics to the continuum. Given the time limitations of this course, we unfortunately cannot go into detail[7], which is why you are encouraged to follow up on topics of interest in the pertinent literature; see, e.g., reference [89] for a textbook that covers much of the contents discussed in the following.

12.1 A Cartoon Introduction to Quantum Mechanics

At the level of elementary particles, matter is commonly described as waves rather than particles, owing to the famous **wave-particle duality** (dating back to the work of Max Planck around 1900). This key concept in quantum mechanics assumes that every particle (such as electrons, protons and neutrons, but also photons and other elementary particles) may be described as either a *particle* or a *wave*. To visualize this concept (without going into details), we may think of standing waves such as the vibrations of elastic rods with fixed or free ends, whose eigenmodes exhibit distinct maxima and nodes (Figure 12.1). If one interprets the wave solution as a probability distribution, then the maxima correspond to those spots where the "particle" is most likely to be found. Thus we may alternatively interpret standing waves as particles.

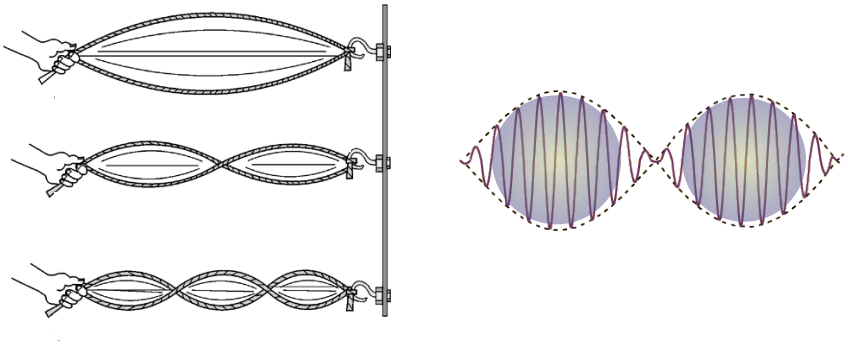


Figure 12.1: Standings waves in a vibrating rod or string and their interpretation as particles.

Consider a vibrating elastic rod, whose 1D equation of motion governing the displacement field $u(x,t)$ is the wave equation

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2} \tag{12.1}$$

with wave speed $c = \sqrt{E/\rho}$. This is the approach of classical mechanics. At the quantum level, the governing equations can be considered as follows. (Note that what follows is only a motivation and *not* a rigorous derivation of the governing equations of quantum mechanics.)

Consider a general form of a wave moving in 1D at constant speed, described by

$$u(x,t) = \hat{u} \exp\big(2\pi i(kx - \nu t)\big) \tag{12.2}$$

[7] An apology to all those who have taken Quantum Mechanics classes and will be bored or laughing at the level of explanations provided here. I simply try to convey the basic concepts within the frame of one lecture, without assuming a background in quantum mechanics.

with an amplitude $\hat{u}$. $2\pi\nu$ serves as the frequency, and k is the wave number. For a given set of boundary conditions, (12.2) inserted into (12.1) yields the eigenfrequencies and associated eigenmodes of the system.

Notice that we may conclude from (12.2) that

$$\frac{\partial^2 u}{\partial x^2} = -4\pi^2 k^2 u, \tag{12.3}$$

and

$$\frac{\partial u}{\partial t} = -i\, 2\pi\nu\, u. \tag{12.4}$$

Next, we need to consider the impulse and energy of a free particle, which behaves like a wave – unlike in classical mechanics. The great contributions of a series of physicists of the 20th century were given in the form of postulates.

De Broglie postulated in 1924 that (for electrons) the wavelength λ (and wave number k) and the particle's impulse p (for a classical particle $p = m\dot{u}$) are related via Planck's constant $h = 6.62607004 \cdot 10^{-34}\mathrm{kg\,m^2/s}$ according to

$$\lambda = \frac{h}{p} \qquad \text{and} \qquad \lambda = 1/k \qquad \Rightarrow \qquad p = kh. \tag{12.5}$$

λ is also known as the **de Broglie wavelength**. Notice that by introducing for convenience the constant

$$\hbar = \frac{h}{2\pi} \tag{12.6}$$

and multiplying (12.3) by $\hbar^2/2m$, we arrive at

$$\frac{\hbar^2}{2m}\frac{\partial^2 u}{\partial x^2} = -\frac{h^2}{2m\,4\pi^2}4\pi^2 k^2 u = -\frac{(hk)^2}{2m}u = -\frac{p^2}{2m}u. \tag{12.7}$$

This allows us to reinterpret the relation between the impulse and kinetic energy of a discrete particle vs. its wave representation as

$$\frac{p^2}{2m} \to -\frac{\hbar^2}{2m}\nabla^2. \tag{12.8}$$

Furthermore, the **Einstein-Planck relation** postulates that a particle's energy is given by the product of frequency and Planck's constant, i.e.,

$$E = h\,\nu. \tag{12.9}$$

Notice that multiplication by ih transforms (12.4) into

$$ih\frac{\partial u}{\partial t} = h\,2\pi\nu\,u = 2\pi(h\nu)u = 2\pi E u \qquad \Leftrightarrow \qquad i\hbar\frac{\partial u}{\partial t} = Eu. \tag{12.10}$$

Finally, for a single particle of mass m and impulse p, exposed to an external potential V (e.g., due to an electric field), the total energy of a particle at position x and with momentum p is given by

$$E(x,p) = \frac{p^2}{2m} + V(x). \tag{12.11}$$

In summary, we have

$$\frac{\hbar^2}{2m}\frac{\partial^2 u}{\partial x^2} = -\frac{p^2}{2m}u \qquad \text{and} \qquad i\hbar\frac{\partial u}{\partial t} = Eu = \left(\frac{p^2}{2m} + V(x)\right)u. \tag{12.12}$$

Rearrangement and combination of the two equations as well as replacing the displacement $u(x,t)$ by a general wave form $\Psi(x,t)$ yields

$$\left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x)\right]\Psi(x,t) = i\hbar\frac{\partial}{\partial t}\Psi(x,t), \tag{12.13}$$

which we may equivalently write as **Schrödinger's equation**

$$\boxed{\hat{\mathcal{H}}\Psi = i\hbar\frac{\partial\Psi}{\partial t}} \qquad \text{with} \qquad \boxed{\hat{\mathcal{H}} = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x)} \tag{12.14}$$

with the differential operator $\mathcal{H}$.

Like the wave equation, the solution to Schrödinger's equation can be written in a separable form

$$\Psi(x,t) = \sum_{n=1}^{\infty}\Psi_n(x)\,\exp(-iE_nt/\hbar), \tag{12.15}$$

whose insertion into Schrödinger's equation yields an eigenvalue problem:

$$\boxed{\hat{\mathcal{H}}\,\Psi_n = E_n\,\Psi_n} \tag{12.16}$$

with eigenvalues E_n (the energy levels) and associated (standing) wave forms Ψ_n. This implies that discrete, quantized energy levels E_n correspond to specific steady wave forms Ψ_n. (12.16) is commonly referred to as the **time-independent Schrödinger equation**.

When considering elementary atomic particles, the dominant interactions are typically of Coulombic type. Recall that the interaction potential between two charged particles at a distance r having charges q_1 and q_2 (see Figure 12.2) is given by the **Coulomb potential**

$$V(r) = \frac{1}{4\pi\varepsilon_0}\frac{q_1q_2}{r} \tag{12.17}$$

with ε_0 denoting the **electric permittivity** of free space. Charges of the same sign repel each other by minimizing the energy in the limit $r\to\infty$, while opposite charges attract each other by minimizing the energy as $r\to 0$.

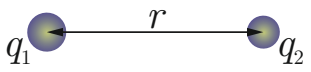


Figure 12.2: Coulombic interaction of two charged particles with charges q_i at a distance r.

Let us consider a general system of multiple charged particles: **electrons** with charge $-e$ and nuclei made up of Z **protons**, each with charge $+e$, and **neutrons** without a charge (see Figure 12.3). The total potential energy follows as

$$V = \underbrace{\sum_{\alpha\neq\beta}\frac{(Z_\alpha e)(Z_\beta e)}{4\pi\varepsilon_0\,|r_{\alpha\beta}|}}_{\text{nuclei interactions}} + \underbrace{\sum_{i\neq j}\frac{e^2}{4\pi\varepsilon_0\,|r_{ij}|}}_{\text{electron interact.}} - \underbrace{\sum_{i,\alpha}\frac{(Z_\alpha e)e}{4\pi\varepsilon_0\,|r_{\alpha i}|}}_{\text{electron/nuclei interact.}} \tag{12.18}$$

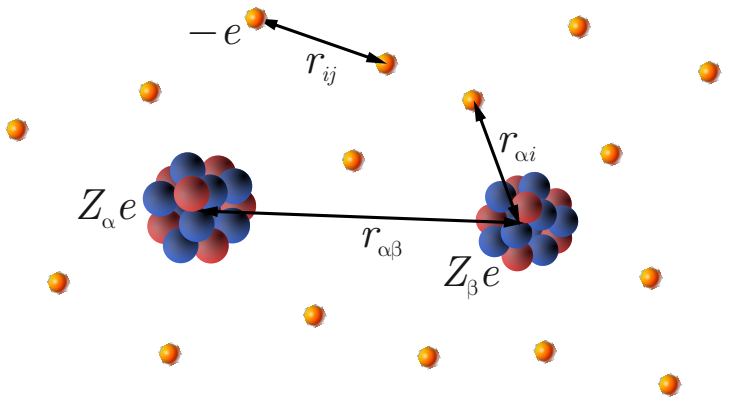


Figure 12.3: Schematic view of electrons (charge $-e$) and nuclei composed of Z protons (charge $+e$) and neutrons along with their distances.

where Z_i denotes the number of protons in the ith nucleus, and we sum over all nucleus-nucleus, electron-electron, and nucleus-electron interactions (avoiding double-counting of interactions).

Most important for the development of the concepts of atomistics is the **Born-Oppenheimer approximation**: the relatively light *electrons are assumed to be moving orders of magnitude faster than the heavy nuclei* (protons/neutrons are about $2,000$ times heavier than electrons). Therefore, one may solve for the wave function of all electrons while assuming the nuclei remain (quasi)static at known positions, and solve for the nuclei positions at much slower time scales.

12.2 Example: the Hydrogen Atom

Hydrogen is the simplest atom possible, having only one proton in its nucleus and one electron revolving around the proton (see Figure 12.4). Consequently, Schrödinger's equation for the single electron – assuming the proton is fixed at $\boldsymbol{x} = \boldsymbol{0}$ in 3D by the Born-Oppenheimer approximation – simplifies dramatically. The $\mathcal{H}$-operator reduces to

$$\hat{\mathcal{H}} = -\frac{\hbar^2}{2m_e}\nabla^2 - \frac{e^2}{4\pi\varepsilon_0\,\|\boldsymbol{x}\|}, \tag{12.19}$$

so that the resulting eigenvalue problem of the time-independent Schrödinger equation becomes

$$\left(-\frac{\hbar^2}{2m_e}\boldsymbol{\nabla}^2 - \frac{e^2}{4\pi\varepsilon_0\,\|\boldsymbol{x}\|}\right)\Psi_n(\boldsymbol{x}) = E_n\,\Psi_n(\boldsymbol{x}). \tag{12.20}$$

This eigenvalue problem can be solved analytically to yield the discrete energy levels and the eigen-waveforms in spherical coordinates:

$$E_n = -\frac{m_e e^4}{8\varepsilon_0^2 h^2 n^2}, \qquad \Psi_{(n,l,m)}(r,\theta,\phi) = c_{n,l,m} R_{n,l}(r)\,P_l^m(\cos\theta)\,\exp(im\phi). \tag{12.21}$$

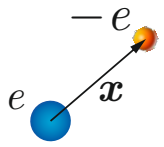


Figure 12.4: Schematic view of a hydrogen atom consisting of one proton and one electron.

The **stationary states** $\Psi_{(n,l,m)}$ involve the Legendre polynomials P_m^l and Laguerre polynomials $R_{n,l}$. $c_{n,l,m}$ are known, closed-form constants (omitted here for conciseness). The discrete energy levels are sometimes rewritten as

$$E_n = -\frac{e^2}{8\pi\varepsilon_0 a_0 n^2}, \tag{12.22}$$

and the energy of the **ground state** is obtained for $n = 1$, etc.

The three *principal quantum numbers* n, m and l are not independent for the above solution to hold: for $1 \leq n$ one needs to have $0 \leq l \leq n-1$ and $-l \leq m \leq l$. For example, the lowest-energy **ground state** corresponds to $(n, l, m) = (1, 0, 0)$ and

$$\Psi_{(1,0,0)}(r) = \frac{1}{a_0^{3/2}} \exp\left(-r/a_0\right) \qquad \text{with} \qquad a_0 = \frac{4\pi\varepsilon_0\hbar^2}{m_e e^2} \tag{12.23}$$

being **Bohr's radius** (i.e., the most probable distance between proton and electron in the hydrogen atom in its ground state). ε_0 is the permittivity of free space.

Every wave function $\Psi_{(n,l,m)}$ defines an atomic **orbital** for a single electron (and E_n is the corresponding ground-state energy). To this end, we may interpret $|\Psi_{(n,l,m)}|^2$ as the charge distribution or the electron **probability density** in an atom. In other words, $|\Psi_{(n,l,m)}|^2$ defines the probability of finding an electron at a distance r from the proton in the hydrogen atom. Multiplication by $4\pi r^2$ (the surface area of a sphere of radius r) yields the **radial distribution function**, which defines the probability of finding an electron anywhere on the spherical surface located at distance r from the proton.

For example, $\Psi_{(1,0,0)}$ from (12.23) corresponds to the *1s orbital*. Principle quantum numbers $(2, 0, 0)$ yield to the *2s-orbital*. In general, $l = 0$ implies *s-orbitals*. Both cases are shown in Figure 12.5. Combinations of quantum numbers with $l = 1$ correspond to *p-orbitals*, $l = 2$ are d-orbitals, $l = 3$ are *f-orbitals*. A complete summary of common orbitals and their quantum numbers is provided in Figure 12.6).

We point out that there is a *fourth quantum number*: the particle's **spin** (also known as *spin quantum number*) m_s can only be $+1/2$ or $-1/2$. **Pauli's exclusion principle** states that

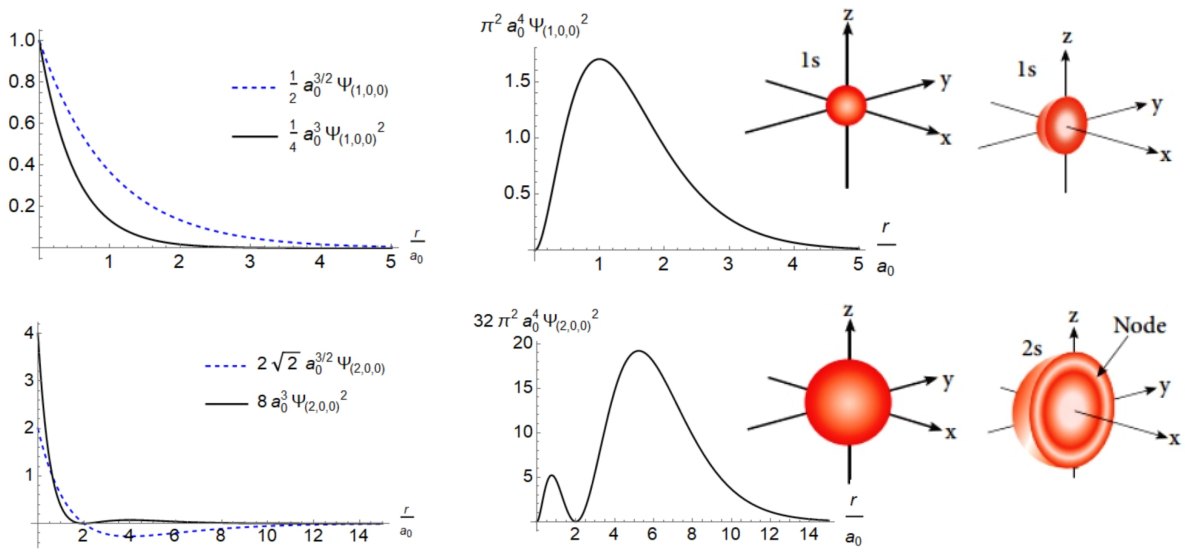


Figure 12.5: Wave function Ψ, probability density Ψ^2, and radial distribution function $4\pi r^2\Psi^2$ of the 1s-orbital ($\Psi_{(1,0,0)}$) and the 2s-orbital ($\Psi_{(1,0,0)}$) of the hydrogen atom (all normalized). Analogous to classical waves, a *node* is a location with a zero probability.

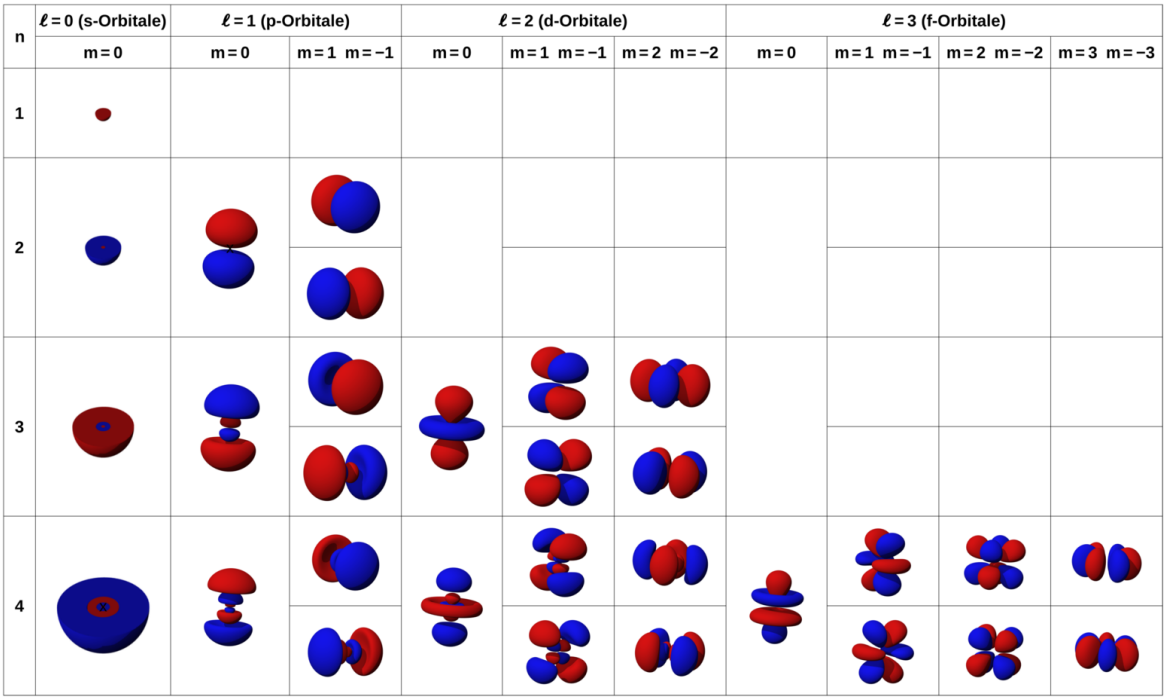


Figure 12.6: Overview of common orbitals and their quantum numbers n, m, l. The hydrogen atom's 1s-orbital is included for $n = 1$, $l = m = 0$. Graphic by MikeRun, adapted from Wikipedia.

there can be at most two electrons for every orientation, and those two electrons must be of opposite spin. In addition, **Hund's rule** states that an orbital is first filled by electrons with $m_s = +1/2$; only after all orbitals have been filled in this way, the orbitals are filled additionally with spin $m_s = -1/2$.

12.3 More Complex Systems

Let us consider more complex atomic systems composed of *electrons* at positions $\boldsymbol{x}^e = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ and *nuclei* at positions $\boldsymbol{q}^n = \{\boldsymbol{q}_1, \dots, \boldsymbol{q}_n\}$. The time-independent Schrödinger equation for this scenario is

$$\hat{\mathcal{H}}(\boldsymbol{x}, \boldsymbol{q})\, \Psi(\boldsymbol{x}, \boldsymbol{q}) = E\, \Psi(\boldsymbol{x}, \boldsymbol{q}) \tag{12.24}$$

with the combined operator (involving electron and nuclei momenta $\boldsymbol{p}^e$ and $\boldsymbol{p}^n$, respectively)

$$\hat{\mathcal{H}}(\boldsymbol{x}, \boldsymbol{q}) = \sum_{i=1}^{N_e} \frac{|\boldsymbol{p}_i^e|^2}{2m_e} + \sum_{i=1}^{N_n} \frac{|\boldsymbol{p}_i^n|^2}{2m_n} + \sum_{\alpha \neq \beta} \frac{(Z_\alpha e)(Z_\beta e)}{4\pi\epsilon_0\, |\boldsymbol{q}_\alpha - \boldsymbol{q}_\beta|} + \sum_{i \neq j} \frac{e^2}{4\pi\epsilon_0\, |\boldsymbol{x}_i - \boldsymbol{x}_j|} - \sum_{i,\alpha} \frac{(Z_\alpha e) e}{4\pi\epsilon_0\, |\boldsymbol{q}_\alpha - \boldsymbol{x}_i|}, \tag{12.25}$$

and recall that we may replace

$$\sum_{i=1}^{N_e} \frac{|\boldsymbol{p}_i^e|^2}{2m_e} \to -\frac{\hbar^2}{2m_e} \sum_{i=1}^{N_e} \nabla^2_{\boldsymbol{x}_i}. \tag{12.26}$$

Using the Born-Oppenheimer approximation, one separates the (fast) electronic from the (slow) nuclei contributions by assuming a solution of the form

$$\boxed{\Psi(\boldsymbol{x}, \boldsymbol{q}) = \Psi_{\text{el}}(\boldsymbol{x}; \boldsymbol{q})\, \Psi_{\text{nuc}}(\boldsymbol{q})} \tag{12.27}$$

That is, the wave function of the electrons, $\Psi_{\text{el}}(\boldsymbol{x};\boldsymbol{q})$, depends on the positions $\boldsymbol{q}$ of the nuclei and of the electrons, $\boldsymbol{x}$, whereas the wave function for the nuclei, $\Psi_{\text{nuc}}(\boldsymbol{q})$, is independent of the electron positions. The rationale is the following, based on a *two-scale* separation: for a given set of nuclei positions $\boldsymbol{q}$, we will solve for the electron positions $\boldsymbol{x}$ while assuming that the nuclei positions are fixed ($\boldsymbol{q}=$ const.). In a higher-order second step, we will solve for the nuclei positions $\boldsymbol{q}$.

This is similar in spirit to the FE^2 method (see Section 8). There, we solved a macroscale FE problem and, every time we needed a constitutive law, we had to solve a nested microscale FE problem for the given macroscale average strain (or stress). Here, we solve for the nuclei positions on the "*macroscale*", and, every time we need the energy of the configuration of nuclei, we turn to the "*microscale*" problem of solving for the electron configuration for a given set of nuclei positions.

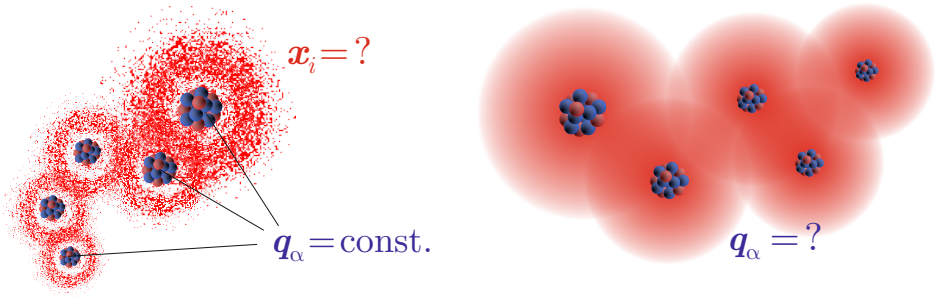


Figure 12.7: Illustration of the two-scale scheme resulting from the Born-Oppenheimer approximation: for fixed nuclei positions $\boldsymbol{q}_\alpha$, we solve for the electron configuration with electrons at positions $\boldsymbol{x}_i$. Once the electron configuration and the associated energy of the system is known for given $\boldsymbol{q}_\alpha$, we solve for $\boldsymbol{q}_\alpha$.

As we are concerned with eigenstates of electrons relative to nuclei positions, (12.27) must satisfy the time-independent Schrödinger equation for the electrons interacting with an approximately *static* array of nuclei. After applying leading-order expansion (the details of which we omit here, since we do not want to dive deep into quantum mechanics), the result is a *decoupled* system of equations for the electrons and nuclei.

The electron wave function Ψ_{el} (for given nuclei positions $\boldsymbol{q}$) is found from solving

$$\hat{\mathcal{H}}_{\text{el}}(\boldsymbol{x},\boldsymbol{q})\,\Psi_{\text{el}}(\boldsymbol{x}) = V(\boldsymbol{q})\,\Psi_{\text{el}}(\boldsymbol{x}), \tag{12.28}$$

with

$$\hat{\mathcal{H}}_{\text{el}} = -\sum_{i=1}^{N_e}\frac{\hbar^2}{2m_e}\nabla^2_{\boldsymbol{x}_i} + \sum_{\alpha\neq\beta}\frac{(Z_\alpha e)(Z_\beta e)}{4\pi\epsilon_0\,|\boldsymbol{q}_\alpha-\boldsymbol{q}_\beta|} + \sum_{i\neq j}\frac{e^2}{4\pi\epsilon_0\,|\boldsymbol{x}_i-\boldsymbol{x}_j|} - \sum_{i,\alpha}\frac{(Z_\alpha e)e}{4\pi\epsilon_0\,|\boldsymbol{q}_\alpha-\boldsymbol{x}_i|}. \tag{12.29}$$

For given nuclei positions, one solves for the electron eigenstates $\Psi_{\text{el}}(\boldsymbol{x})$ and the associated energy $V(\boldsymbol{q})$, which is a function of the nuclei positions only[8].

In a next, higher-level step, one then finds the solutions for the nuclei positions from

$$\hat{\mathcal{H}}_{\text{nuc}}(\boldsymbol{q})\,\Psi_{\text{nuc}}(\boldsymbol{q}) = E\,\Psi_{\text{nuc}}(\boldsymbol{q}), \tag{12.30}$$

with

$$\hat{\mathcal{H}}_{\text{nuc}} = \sum_{i=1}^{N}\frac{|\boldsymbol{p}_i|^2}{2m_i} + V(\boldsymbol{q}). \tag{12.31}$$

[8] $V(\boldsymbol{q})$ is known as the *adiabatic* contribution of electrons to the energy of the system; while there are other terms to consider below, those other terms are negligibly small, which is where the leading-order assumptions come in – which we do not discuss here.

Note that it is relatively arbitrary for the solution whether or not the second term in (12.29) is retained there or taken into (12.31). At this point, all electron contributions have been 'condensed out'; i.e., if $V(\boldsymbol{q})$ is known, we can solve the above problem for the nuclei positions only. This is the approach taken by **atomistics**, which treats the nuclei as classical particles in the Hamiltonian sense (to be discussed in Section 12.5). Varying the nuclear positions creates a high-dimensional energy landscape of the 'ground-state potential' $V(\boldsymbol{q})$, so that we seek to find the motion of the nuclei in this potential energy landscape.

One frequent assumption being made in solving the above equations is that of *non-interacting particles*. As a simple example consider a hydrogen molecule, H_2, which contains two electrons. A closed-form solution is feasible if these electrons are treated as non-interacting, so that the wave function for each electron is obtained individually.

12.4 Density Functional Theory in a Nutshell

The problem with the equations of quantum mechanics is the extreme complexity involved when treating systems with many particles (recall that the specifics of each electron and nucleus are to be solved for). Therefore, alternatives have been proposed to solve large systems in more efficient ways, one such example being the famous **Density Functional Theory** (DFT).

Consider Schrödinger's equation for a system of N_e electrons with known nuclei positions:

$$\hat{\mathcal{H}}(\boldsymbol{x};\boldsymbol{q})\Psi_{\text{el}}(\boldsymbol{x}) = E\,\Psi_{\text{el}}(\boldsymbol{x}), \qquad \hat{\mathcal{H}}(\boldsymbol{x};\boldsymbol{q}) = -\sum_{i=1}^{N_e}\frac{\hbar^2}{2m_e}\nabla^2_{\boldsymbol{x}_i} + \sum_{i=1}^{N_e} V(\boldsymbol{x}_i;\boldsymbol{q}) + \sum_{i\neq j}\frac{e^2}{4\pi\epsilon_0\,|\boldsymbol{x}_i-\boldsymbol{x}_j|}, \tag{12.32}$$

where we abbreviated the interaction potential of each electron i with the surrounding protons by $V(\boldsymbol{x}_i;\boldsymbol{q})$. Note that for a given wave function $\Psi_{\text{el}}(\boldsymbol{x})$ the associated energy can be computed as

$$E = \frac{\langle\Psi_{\text{el}}|\mathcal{H}|\Psi_{\text{el}}\rangle}{\langle\Psi_{\text{el}}|\Psi_{\text{el}}\rangle}, \tag{12.33}$$

where we turned to classical quantum-mechanics notion with the definitions

$$\langle a|b\rangle = \int_{\mathbb{R}^3}\cdots\int_{\mathbb{R}^3} a^*(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{N_e})\,b(\boldsymbol{x}_1,\dots,\boldsymbol{x}_N)\,\mathrm{d}\boldsymbol{x}_1\cdots\mathrm{d}\boldsymbol{x}_{N_e} \tag{12.34}$$

and

$$\langle a|\mathcal{A}|b\rangle = \int_{\mathbb{R}^3}\cdots\int_{\mathbb{R}^3} a^*(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{N_e})\mathcal{A}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{N_e})b(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{N_e})\,\mathrm{d}\boldsymbol{x}_1\cdots\mathrm{d}\boldsymbol{x}_{N_e} \tag{12.35}$$

The basic idea of density functional theory is to *minimize* E with respect to the wave functions to find the electron configurations, which allows for the introduction of approximations (similar in spirit to finding approximations in energy minimization problems in FE, e.g.). Note that the last term in the operator (12.32) complicates calculations, as it involves correlations between different electron wave functions.

Therefore, the most important step of density functional theory is to avoid considering each of the N_e electrons individually by introducing the **electron density** $\rho(\boldsymbol{r})$ as the probability of finding an electron at a location $\boldsymbol{r}\in\mathbb{R}^3$ (with all electrons being interchangeable), i.e.,

$$\rho(\boldsymbol{r}) = \sum_{i=1}^{N_e}\int_{\mathbb{R}^3}\cdots\int_{\mathbb{R}^3}|\Psi(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{N_e})|^2\,\delta(\boldsymbol{r}-\boldsymbol{x}_i)\,\mathrm{d}\boldsymbol{x}_1\dots\mathrm{d}\boldsymbol{x}_{N_e} = \langle\Psi|\sum_{i=1}^{N_e}\delta(\boldsymbol{r}-\boldsymbol{x}_i)|\Psi\rangle \tag{12.36}$$

such that

$$\int_{\mathbb{R}^3} \rho(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r} = N_e. \tag{12.37}$$

After insertion into the governing equations, certain approximations are applied to make the variational problem solvable for the electron density, resulting in the different flavors of DFT (which, again, we cannot discuss here for brevity).

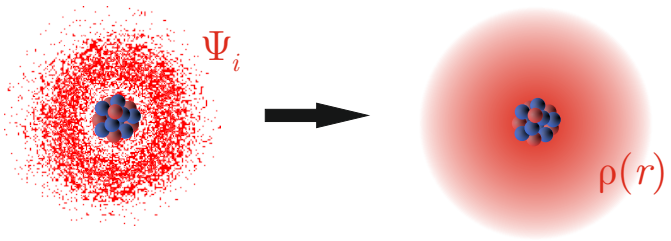


Figure 12.8: Density Functional Theory solves, instead of the wave functions of all electrons, for the electron density function $\rho(r)$.

As a result, the energy functional is generally expressed in the form

$$E[\rho] = E_{\mathrm{ZZ}}[\rho] + V_{\mathrm{ext}}[\rho] + V_{ee}[\rho] + E_{\mathrm{kin}}[\rho]. \tag{12.38}$$

The individual contributions include the potential of the nuclei interactions,

$$E_{\mathrm{ZZ}}[\rho] = \sum_{\alpha\neq\beta} \frac{(Z_\alpha e)(Z_\beta e)}{4\pi\epsilon_0\, |\boldsymbol{q}_\alpha - \boldsymbol{q}_\beta|}, \tag{12.39}$$

which is independent of the electron wave functions. The interaction of electrons with nuclei is captured by the *external interaction energy functional*

$$V_{\mathrm{ext}}[\rho] = \int_{\mathbb{R}^3} V_{\boldsymbol{q}^n}(\boldsymbol{r})\rho(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r} \qquad \text{with} \qquad V_{\boldsymbol{q}^n}(\boldsymbol{r}) = -\sum_\beta \frac{(Z^\beta e)e}{|\boldsymbol{r} - \boldsymbol{q}_\beta|}. \tag{12.40}$$

The *electron interaction energy* is usually approximated by the so-called **Hartree energy**,

$$V_{ee}[\rho] \approx V_{\mathrm{H}}[\rho] = \frac{1}{2}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \frac{e^2\rho(\boldsymbol{r})\,\rho(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|}\, \mathrm{d}\boldsymbol{r}\, \mathrm{d}\boldsymbol{r}', \tag{12.41}$$

which is not exact because it includes purely Coulombic interactions and neglects quantum-mechanical effects. Since this introduces approximation errors, one usually adds a correction term, the so-called **exchange-correlation functional** $E_{\mathrm{xc}}[\rho]$, so that

$$V_{ee}[\rho] = V_{\mathrm{H}}[\rho] + E_{\mathrm{xc}}[\rho]. \tag{12.42}$$

The split of the energy into that of non-interacting electrons and a correction goes back to Kohn and Sham (see reference [49]), which is why this approach is usually referred to as **Kohn-Sham theory**. We should further note that this method is fundamentally based on the theorems of Hohenberg and Kohn (see reference [36])

Finally, the kinetic energy is another critical contribution that can only be evaluated as a function of the electron density under certain assumptions. For example, the **Thomas-Fermi model** approximates the kinetic energy by

$$E_{\mathrm{kin}}[\rho] = c_F \int_{\mathbb{R}^3} \rho^{5/3}(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r} \qquad \text{with} \qquad c_F = \frac{3h^2}{10m_e}\left(\frac{3}{8\pi}\right)^{2/3}. \tag{12.43}$$

Irrespective of the approximations made for the different energetic contributions, the goal of DFT is to find $\rho(\boldsymbol{r})$ for a given set of nuclei positions (instead of solving for the positions of all electrons). Since electrons are interchangeable (for our purposes), $\rho(\boldsymbol{r})$ contains equivalent information – but it has the significant advantage that we may postulate approximations for $\rho(\boldsymbol{r})$ as a global field without having to consider individual electrons. In order to solve for the electron density, one introduces approximations via basis functions. For example, it is convenient to write

$$\rho(\boldsymbol{r}) = \sum_{i=1}^{N} |\varphi_i(\boldsymbol{r})|^2 \tag{12.44}$$

with to be determined orbitals $\varphi_i(\boldsymbol{r})$. The key assumption is that *any ground-state observable is a function of* $\rho(\boldsymbol{r})$. We unfortunately cannot go deeper into this topic, but a long list of available textbooks and literature on Quantum Mechanics and Density Functional Theory are available (see, e.g., reference [89] and references therein).

The primary reason DFT has become so popular (and was worthy the Nobel Prize in Chemistry in 1998) is its *tremedous efficiency* as opposed to solving Schrödinger's equation for large systems. Unlike solving Schrödinger's equation exactly, DFT can handle systems of hundreds to a few thousands of atoms (using high-performance computing). It is also highly accurate for most systems of interest. It also be combined with atomistic techniques, and it can be extended to account for, e.g., chemical reactions.

12.5 Hamiltonian Description

Atomistic techniques such as molecular dynamics (MD) or molecular statics (MS) make effective use of the Born-Oppenheimer assumption in that they separate the fast electronic configurational changes from the slow motion of the atomic nuclei. In particular, if the spacing between particles is sufficiently large, then their wave functions are spatially localized and they can be treated as *classical particles*. In general, one may introduce a characteristic **de Broglie wavelength** (of a single particle at a temperature T with kinetic energy $E \propto k_B T$)

$$\Lambda = \sqrt{\frac{h^2}{2\pi m k_B T}} \tag{12.45}$$

and treats atoms as *classical particles*, if the average interatomic spacing $a \gg \Lambda$. By contrast, if Λ is on the order of interatomic distances, then the wave-like behavior of atoms is non-negligible. Especially at high temperatures all atoms can be treated as classical particles.

The starting point of atomistics is (12.31) where, however, the nuclei wave functions are replaced by classical particle positions and momenta. To this end, an atomistic ensemble containing N atoms is uniquely described by how their nuclei *positions*

$$\boldsymbol{q}(t) = \{\boldsymbol{q}_1(t), \dots, \boldsymbol{q}_N(t)\} \tag{12.46}$$

and the derived nuclei *momenta*

$$\boldsymbol{p}(t) = \{\boldsymbol{p}_1(t), \dots, \boldsymbol{p}_N(t)\} \tag{12.47}$$

with $\boldsymbol{p}_i = m_i \dot{\boldsymbol{q}}_i$ evolve over time t. m_i represents the mass of atom i, and dots denote time derivatives. The ensemble's **total Hamiltonian** is given by

$$\boxed{\mathcal{H}(\boldsymbol{q}, \boldsymbol{p}) = \sum_{i=1}^{N} \frac{|\boldsymbol{p}_i|^2}{2m_i} + V(\boldsymbol{q})} \tag{12.48}$$

where the first term accounts for the *kinetic energy* and the second term represents the *potential energy* of the ensemble with V denoting a suitable **atomic interaction potential** (which includes all information about electron interactions in between the atomic nuclei; it is thus a condensed form depending only on the nucleus or atomic positions, see Section 12.3).

Compared to the electron wave functions discussed above, atomic nuclei are moving relatively slowly, which is why one assumes that the time evolution of an atomistic system is governed by **Hamilton's equations**,

$$\boxed{\dot{\boldsymbol{q}} = \frac{\partial \mathcal{H}}{\partial \boldsymbol{p}}, \qquad \dot{\boldsymbol{p}} = -\frac{\partial \mathcal{H}}{\partial \boldsymbol{q}}} \tag{12.49}$$

The second equation yields Newton's equations of motion for all atoms $i = 1, \dots, N$:

$$\boxed{m_i \ddot{\boldsymbol{q}}_i = \boldsymbol{f}_i(\boldsymbol{q}) = -\frac{\partial V}{\partial \boldsymbol{q}_i}(\boldsymbol{q})} \tag{12.50}$$

where $\boldsymbol{f}_i(\boldsymbol{q})$ represents the total (net) force acting on atom i. For **molecular statics**, these equations reduce to

$$\boldsymbol{f}_i(\boldsymbol{q}) = -\frac{\partial V}{\partial \boldsymbol{q}_i}(\boldsymbol{q}) = \boldsymbol{0}, \tag{12.51}$$

and they aim to minimize the total potential energy (this scenario only applies at zero temperature, as we will discuss later).

The challenge is now to identify the interatomic potential V from either first principles (see Section 12.3) or to introduce empirical potentials whose parameters are fitted to reproduce known material properties (such as elastic constants, thermal expansion coefficients, etc.).

In most materials (including metals, ceramics and organic materials), the interaction potential allows for an additive decomposition, i.e.,

$$V(\boldsymbol{q}) = \sum_{i=1}^{N} V_i(\boldsymbol{q}) \tag{12.52}$$

with V_i being the energy of atom i. Some authors in the scientific literature have preferred to distinguish between external and internal forces by introducing external forces $\boldsymbol{f}_{i,\text{ext}}$ on all atoms $i = 1, \dots, N$, so that

$$\boldsymbol{f}_i(\boldsymbol{q}) = \boldsymbol{f}_{i,\text{ext}}(\boldsymbol{q}_i) - \frac{\partial V}{\partial \boldsymbol{q}_i}(\boldsymbol{q}). \tag{12.53}$$

Ideally (and in any reasonable physical system), such external forces are conservative and derive from an external potential $V_{\text{ext}}(\boldsymbol{q})$ and we may combine internal and external forces into a single potential, which is tacitly assumed in the following. Examples of conservative external forces include gravitation, long-range Coulombic interactions, or multi-body interactions such as during contact.

13 Interatomic Potentials

As we have seen above, the constitutive description of atomistic ensembles is provided by interatomic potentials. An **interatomic potential** describes the effective binding energy of an atom embedded into an ensemble of atoms whose nuclei are located at $\boldsymbol{q} = \{\boldsymbol{q}_1, \dots, \boldsymbol{q}_N\}$:

$$V = V(\boldsymbol{q}). \tag{13.1}$$

In the following, we only consider potentials that depend on atomic (nuclei) positions (which can be generalized to potentials that also depend on atomic momenta, if needed). The potential accounts for both repulsive and attractive forces.

In principle, every potential (written for an ensemble of N atoms) can be expanded as

$$\boxed{V(\boldsymbol{q}) = \sum_{i=1}^{N} V_1(\boldsymbol{q}_i) + \frac{1}{2}\sum_{i\neq j} V_2(\boldsymbol{q}_i, \boldsymbol{q}_j) + \frac{1}{3!}\sum_{i\neq j, i\neq k} V_3(\boldsymbol{q}_i, \boldsymbol{q}_j, \boldsymbol{q}_k) + \dots} \tag{13.2}$$

where one usually avoids summation over identical indices (i.e., when writing $i \neq j$ we imply summation over $i = 1, \dots, N$ and $j = 1, \dots, N$ with the exclusion of the case where $i = j$).

The term involving V_1 can be neglected, unless atoms are in an external field (e.g., ions in an electric field). Otherwise, the potential should be invariant under rigid-body motion so $V_1 \equiv 0$. Retaining only V_2 and neglecting higher-order dependencies results in *pair potentials* (discussed in Section 13.1), whereas retaining V_2 and V_3 results in *three-body potentials* (discussed in Section 13.3), etc.

13.1 Pair Potentials

The specific class of **two-body potentials** or **pair potentials** assumes the form

$$V(\boldsymbol{q}) = \sum_{i\neq j} f(\boldsymbol{q}_i, \boldsymbol{q}_j) = \sum_{i\neq j} f(r_{ij}) \qquad \text{with} \qquad r_{ij} = \|\boldsymbol{q}_i - \boldsymbol{q}_j\|\,, \tag{13.3}$$

where we made use of the fact that invariance under rigid body motion requires the potential to only depend on relative distances between between pairs of atoms (lowest-order correlation).

Example I: Coulombic crystals

As a simple example, consider ions with charges $Q = \{Q_1, \dots, Q_N\}$ interacting through **Coulombic interaction**:

$$\boxed{V(\boldsymbol{q}) = \frac{1}{4\pi\varepsilon_0}\sum_{i\neq j}\frac{Q_i Q_j}{r_{ij}}} \qquad \text{with} \qquad r_{ij} = \|\boldsymbol{q}_i - \boldsymbol{q}_j\|\,. \tag{13.4}$$

The force between two isolated ions ($N = 2$) is thus obtained as

$$\boldsymbol{f}_1 = -\boldsymbol{f}_2 = -\frac{\partial V}{\partial \boldsymbol{q}_1} = -\frac{\partial}{\partial \boldsymbol{q}_1}\frac{1}{4\pi\varepsilon_0}\frac{Q_1 Q_2}{r} = \frac{1}{4\pi\varepsilon_0}\frac{Q_1 Q_2}{r^2}\hat{\boldsymbol{r}} \qquad \text{with} \qquad \hat{\boldsymbol{r}} = \frac{\boldsymbol{q}_1 - \boldsymbol{q}_2}{\|\boldsymbol{q}_1 - \boldsymbol{q}_2\|}. \tag{13.5}$$

Consequently, equally-signed charges with $\text{sign}(Q_1) = \text{sign}(Q_2)$ repel each other, and they produce a positive **binding energy** $V > 0$. In contrast, charges of opposite signs, i.e.,

$\text{sign}(Q_1) = -\text{sign}(Q_2)$, attract each other, resulting in a negative binding energy $V < 0$. The force on an atom i in an ensemble of atoms is obtained by superposition:

$$\boldsymbol{f}_i = -\frac{\partial V}{\partial \boldsymbol{q}_i} = \frac{1}{4\pi\varepsilon_0}\sum_{j\neq i}\frac{Q_iQ_j}{r_{ij}^2}\hat{\boldsymbol{r}}_{ij}. \tag{13.6}$$

Coulombic interactions are **long-range** because the $1/r$-dependence decays slowly with distance.

As an example, consider an *infinite 1D chain* of ions of opposite charges $Q_1 = -Q_2 = Q$ (so that nearest-neighbors are attracting each other, whereas next-to-nearest neighbors are repelling each other). This is the prototype of an **ionic crystal**. Assume that the chain has equal interatomic spacings a between each pair of atoms. The potential energy of every atom in the chain, depending on distance a, then becomes

$$V(a) = \frac{2}{4\pi\varepsilon_0}\sum_{i=1}^{\infty}\left[\frac{Q^2}{2ia} - \frac{Q^2}{(2i-1)a}\right] = -\frac{2Q^2}{4\pi\varepsilon_0 a}\left(1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\dots\right), \tag{13.7}$$

where we exploited symmetry (the factor of 2 results from symmetry of neighbors). Using the Taylor expansion

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots \tag{13.8}$$

allows us to write the potential as

$$V(a) = -\frac{Q^2}{2\pi\varepsilon_0 a}\ln 2. \tag{13.9}$$

We notice that this ionic crystal does not have an **equilibrium spacing** $a_0 = \arg\min V(a)$, since the minimum V is attained for $a \to 0$, i.e., $a_0 = 0$ in this example (which could have been expected since the $1/r$-scaling attracts nearest neighbors more strongly than next-to-nearest neighbors repel one another).

Example II: Lennard-Jones potential

Next, consider one of the simplest two-body interatomic potentials with an equilibrium distance, the **Lennard-Jones** potential (named after *Sir John Lennard-Jones*) which has the general form

$$\boxed{V(r) = -\frac{A}{r^n} + \frac{B}{r^m}} \tag{13.10}$$

with integer exponents $n, m > 0$ and real-valued strength coefficients $A, B > 0$. A represents the strength of the attractive contribution, whereas B dominates the repulsive contribution. For an equilibrium to exist, atoms should attract over large distances but repel over short distances (unlike in the 1D chain example above); this implies that we must have $m > n$.

Note that two atoms at a distance r now have the binding energy (13.10), which has an extremum if

$$0 = \frac{\partial V}{\partial r} = n\frac{A}{r^{n+1}} - m\frac{B}{r^{m+1}} = \frac{1}{r^{n+1}}\left(nA - \frac{mB}{r^{m-n}}\right) \quad \Leftrightarrow \quad r_0 = \left(\frac{mB}{nA}\right)^{1/(m-n)}. \tag{13.11}$$

Furthermore, notice that

$$\left.\frac{\partial^2 V}{\partial r^2}\right|_{r=r_0} = -n(n+1)\frac{A}{r_0^{n+2}} + m(m+1)\frac{B}{r_0^{m+2}} = \frac{1}{r_0^{n+2}}\left[m(m+1)\frac{B}{r_0^{m-n}} - n(n+1)A\right]$$
$$= \frac{1}{r_0^{n+2}}\left[m(m+1)\frac{BnA}{mB} - n(n+1)A\right] = \frac{nA}{r_0^{n+2}}\left[(m+1)-(n+1)\right] = \frac{n(m-n)A}{r_0^{n+2}}. \tag{13.12}$$

Consequently, the potential has a minimum at $r = r_0$ if $m > n$ (and $A, B > 0$), as expected. $V(r_0)$ is the **maximum binding energy** (see Figure 13.1.

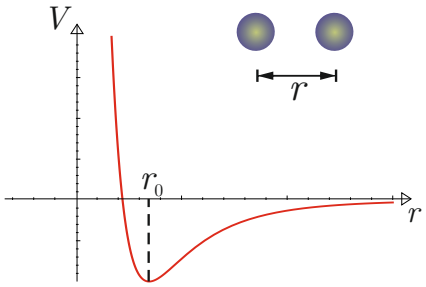


Figure 13.1: Lennard-Jones potential for two particles at a distance r.

As an example, let us compute the equilibrium spacing a_0 of an *infinite chain of identical atoms* (see Figure 13.2). To this end, consider the energy of a single atom in an infinite chain of atoms at equal distances a, which is:

$$V = 2\sum_{i=1}^{\infty}\left(-\frac{A}{(ia)^n} + \frac{B}{(ia)^m}\right). \tag{13.13}$$

It is convenient to introduce the so-called *Riemann zeta-function*

$$\zeta(n) = \sum_{i=1}^{\infty}\frac{1}{i^n}, \tag{13.14}$$

so that

$$V = 2\left(-A\frac{\zeta(n)}{a^n} + B\frac{\zeta(m)}{a^m}\right). \tag{13.15}$$

The minimizer is obtained from

$$0 = \frac{\partial V}{\partial a} = 2\left(nA\frac{\zeta(n)}{a^{n+1}} - mB\frac{\zeta(m)}{a^{m+1}}\right) = \frac{2}{a^{n+1}}\left(nA\zeta(n) - mB\frac{\zeta(m)}{a^{m-n}}\right), \tag{13.16}$$

so that

$$a_0 = \left(\frac{mB\zeta(m)}{nA\zeta(n)}\right)^{1/(m-n)}. \tag{13.17}$$

We may take, e.g., the most common form (known as the **6-12 potential**) with $n = 6$ and $m = 12$. In this case

$$a_0 = \left(\frac{1382\pi^6}{675675}\frac{B}{A}\right)^{1/6} \approx \left(1.96639\frac{B}{A}\right)^{1/6}. \tag{13.18}$$

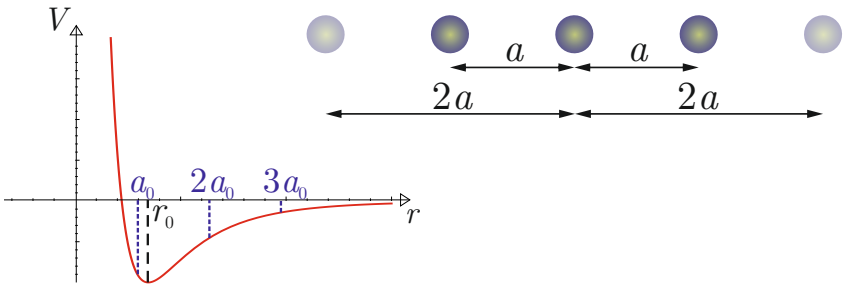


Figure 13.2: Lennard-Jones potential for an atom in an infinite chain of equi-distanced particles at a distance a.

Note that for the same choice of $n = 6$ and $m = 12$, (13.11) yields

$$r_0 = \left(\frac{2B}{A}\right)^{1/6}. \tag{13.19}$$

That is, the equilibrium distance r_0 between *two isolated atoms* is different from the equilibrium distance a_0 between atoms in an *infinite chain* (and yet another distance would be obtained when considering only a finite number of neighboring atoms or 2D arrays of atoms). This makes sense, since atoms in the infinite chain interact also with all other atoms in the chain, not only with their nearest neighbor(s). Therefore, the total energy of an atom takes the form

$$2\left[V(a) + V(2a) + V(3a) + \text{etc.}\right], \tag{13.20}$$

whose minimizer is distinct from $V(a)$.

Finally, note that choices such as $m = 12$ and $n = 6$ result in quickly-decaying interaction forces, which is why such potentials are called **short-range** (in contrast to the Coulombic interactions discussed above, which are **long-range**).

If the equilibrium spacing a_0 and the maximum binding energy $V(a_0)$ are known (which can both be determined experimentally), then the potential parameters can be identified by fitting.

Binding energies are typically on the order of **electron-volts**:

$$1\text{eV} = |q_e| \cdot 1\text{V} = 1.60217662 \cdot 10^{-19}\text{J}, \tag{13.21}$$

where $q_e = 1.60217662 \cdot 10^{-19}$C is the charge of a single electron. Therefore, 1eV corresponds to the energy of a single electron that has passed through a potential of 1V.

Equilibrium spacings are on the order of a few **Ångström** (named after Swedish physicist *Anders Jonas Ångström* who studied the wavelengths of electromagnetic radiation, where this characteristic length emerges):

$$1\text{Å} = 10^{-10}\text{m}. \tag{13.22}$$

Example III: Morse potential

A further example of a two-body potential of practical relevance is the **Morse potential** (named after American physicist *Philip McCord Morse*):

$$V(r) = D\big[1 - \exp\left(-a(r - r_0)\right)\big]^2 \tag{13.23}$$

with equilibrium distance r_0 (between two isolated atoms) and well depth

$$D = V_\infty - V(r_0), \tag{13.24}$$

i.e., the energetic difference between the equilibrium and the energy at $r \to \infty$. The limit

$$V_\infty = \lim_{r\to\infty} V(r) \tag{13.25}$$

is known as the **dissociation energy**.

13.2 (An)Harmonicity and the Quasiharmonic Approximation

All of the above potentials have one feature in common: they are **anharmonic**, i.e., the potential is not symmetric around the equilibrium point. Of importance later, one can always construct a **harmonic potential** through a Taylor expansion:

$$V(r) = V(r_0) + \left.\frac{\partial V}{\partial r}\right|_{r_0} (r - r_0) + \frac{1}{2}\left.\frac{\partial^2 V}{\partial r^2}\right|_{r_0} (r - r_0)^2 + \text{h.o.t.} \tag{13.26}$$

Note that, if the linear term vanishes by the definition of equilibrium, then – dropping higher-order terms – we arrive at

$$\boxed{V_{\text{qh}}(r) \approx V(r_0) + \frac{C}{2}(r - r_0)^2} \tag{13.27}$$

This expansion is known as the **quasiharmonic approximation** with a **force constant**

$$C = \left.\frac{\partial^2 V}{\partial r^2}\right|_{r_0} > 0. \tag{13.28}$$

This approximation is of particular importance when studying atomic vibrations, which are of small amplitude, so that the quasiharmonic approximation can be used to compute the lowest-order vibrational frequencies (e.g., computing the **phonon modes** via dispersion relations). Note that this assumes that each atom is vibrating individually. Therefore, the quasiharmonic approximation generally yields accurate predictions at low absolute temperature, whereas it leads to *significant errors at elevated temperatures* where vibrational amplitudes become so large that the assumption no longer holds.

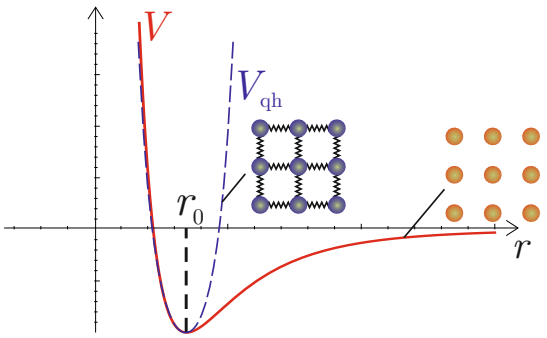


Figure 13.3: The quasiharmonic approximation replaces the anharmonic potential V in the vicinity of its minimum at r_0 by a harmonic potential V_{qh}.

13.3 Multi-Body Potentials

Pair potentials are simple and efficient but not suitable to describe most material behavior, especially that of solid bodies (while it may be suitable, e.g., for gases). Here, multi-body potentials gain importance.

Example I: Stilinger-Weber potential

The **Stilinger-Weber** potential for silicon is an example for a three-body potential. In addition to pair interactions, the third-order potential term $V_3(\boldsymbol{q}_i, \boldsymbol{q}_j, \boldsymbol{q}_k)$ is conveniently expressed in terms of the two distances and the **bond angle** between triplets of atoms. The Stilinger-Weber potential thus has the general form

$$V(\boldsymbol{q}) = \sum_{i,j\in\mathcal{L}} V_2(r_{ij}) + \sum_{i,j,k\in\mathcal{L}} V_3(r_{ij}, r_{ik}, \theta_{ijk}), \tag{13.29}$$

where θ_{ijk} denotes the angle between $\boldsymbol{r}_{ij}$ and $\boldsymbol{r}_{ik}$.

Example II: Embedded Atom Method

The most common type of potential for metals is based on the so-called **Embedded Atom Method** (**EAM**), whose potential for an atom i surrounded by a set $\mathcal{L}$ of atoms is defined by

$$\boxed{V_i(\boldsymbol{q}) = \frac{1}{2}\sum_{j\in\mathcal{L}} \Phi(r_{ij}) + U(\rho_i)} \qquad \text{with} \qquad r_{ij} = \|\boldsymbol{q}_i - \boldsymbol{q}_j\|\,. \tag{13.30}$$

$\Phi(r)$ is a pair potential representing nucleus-nucleus interactions. $U(\rho)$ is called the **embedding function**, which denotes the excess energy due to embedding atom i within the electron density ρ_i present at the location of atom i due to all other atoms. For simplicity, one commonly assumes that ρ_i depends on the distances r_{ij} to all other atoms j:

$$\rho_i = \sum_{j\in\mathcal{L}} f(r_{ij}). \tag{13.31}$$

In order to compute the force acting on an atom k, we must consider the total potential energy

$$V(\boldsymbol{q}) = \sum_{i\in\mathcal{L}} V_i(\boldsymbol{q}). \tag{13.32}$$

The force acting on an atom k is derived as

$$\boldsymbol{f}_k = -\frac{\partial V}{\partial \boldsymbol{q}_k} = -\left[\frac{1}{2}\sum_{i\in\mathcal{L}}\sum_{j\in\mathcal{L}} \Phi'(r_{ij}) + \sum_{i\in\mathcal{L}} U'(\rho_i)\sum_{j\in\mathcal{L}} f'(r_{ij})\right]\frac{\partial r_{ij}}{\partial \boldsymbol{q}_k}. \tag{13.33}$$

For atomic ensembles, the partial derivative can be computed by using $r_{ij} = \|\boldsymbol{q}_i - \boldsymbol{q}_j\|$, so that

$$\sum_{i,j}(\cdot)_{ij}\frac{\partial r_{ij}}{\partial \boldsymbol{q}_k} = \sum_{i,j}(\cdot)_{ij}\frac{\boldsymbol{r}_{ij}}{r_{ij}}\left(\delta_{ik} - \delta_{jk}\right) = \sum_{i,j}(\cdot)_{ij}\hat{\boldsymbol{r}}_{ij}\left(\delta_{ik} - \delta_{jk}\right). \tag{13.34}$$

Note that $r_{ij} = r_{ji}$ and $\boldsymbol{r}_{ij} = -\boldsymbol{r}_{ji}$, so that the forces simplify to

$$\begin{aligned}
\boldsymbol{f}_k &= -\left[\frac{1}{2}\sum_{i\in\mathcal{L}}\sum_{j\in\mathcal{L}} \Phi'(r_{ij}) + \sum_{i\in\mathcal{L}} U'(\rho_i)\sum_{j\in\mathcal{L}} f'(r_{ij})\right]\frac{\boldsymbol{r}_{ij}}{r_{ij}}\left(\delta_{ik} - \delta_{jk}\right) \\
&= -\left[\frac{1}{2}\sum_{j\in\mathcal{L}} \Phi'(r_{jk}) + U'(\rho_k)\sum_{j\in\mathcal{L}} f'(r_{jk})\right]\frac{\boldsymbol{r}_{kj}}{r_{jk}} + \left[\frac{1}{2}\sum_{i\in\mathcal{L}} \Phi'(r_{ik}) + \sum_{i\in\mathcal{L}} U'(\rho_i) f'(r_{ik})\right]\frac{\boldsymbol{r}_{ik}}{r_{ik}} \\
&= \sum_{j\in\mathcal{L}}\left[\Phi'(r_{jk}) + \left(U'(\rho_j) + U'(\rho_k)\right) f'(r_{jk})\right]\hat{\boldsymbol{r}}_{jk},
\end{aligned}$$

where $\hat{\boldsymbol{r}}_{jk} = \boldsymbol{r}_{jk}/r_{jk}$ is the unit vector between atoms j and k, $\boldsymbol{r}_{jk} = \boldsymbol{q}_j - \boldsymbol{q}_k$, and the amplitude of the force between atoms j and k is

$$f_{jk} = \Phi'(r_{jk}) + \big(U'(\rho_j) + U'(\rho_k)\big)\, f'(r_{jk}). \tag{13.35}$$

Overall, this leads to the force on atom k as

$$\boldsymbol{f}_k = \sum_{j\in\mathcal{L}} f_{jk}\hat{\boldsymbol{r}}_{jk}. \tag{13.36}$$

For most potentials, pair interactions and electron densities are **short-range**, so that summations over $\mathcal{L}$ can usually be restricted to summations over a small number of neighboring atoms. To this end, one introduces a **cut-off radius** r_{cutoff}, which defines the cut-off neighborhood around an atom i as

$$\mathcal{C}_i = \big\{j \in \mathcal{L} \;:\; \|\boldsymbol{q}_j - \boldsymbol{q}_i\| \leq r_{\text{cutoff}}\big\}, \tag{13.37}$$

resulting in

$$\boldsymbol{f}_k = \sum_{j\in\mathcal{L}} f_{jk}\hat{\boldsymbol{r}}_{jk} \approx \sum_{j\in\mathcal{C}_k} f_{jk}\hat{\boldsymbol{r}}_{jk}. \tag{13.38}$$

Example III: Force Fields

Especially for organic materials such as, e.g., polymers, **force fields** have gained popularity. In a nutshell, a molecule is approximated by a collection of charged atoms linked by elastic connections. Consequently, the potential decomposes into contributions from bonds, angles, dihedral angles, and long-range non-bond angles with the general form:

$$\begin{aligned} V(\boldsymbol{q}) = &\sum_{i,j\in\text{bonds}} k_{ij}^{\text{bond}}(r_{ij} - r_{\text{eq.}})^2 + \sum_{i\in\text{angles}} k_i^{\text{angle}}(\theta_i - \theta_{\text{eq.}})^2 + \sum_{i,j\in\text{bonds}} k_i^{\text{dihedral}} f(\phi_i, \theta_i) \\ &+ \sum_{i,j\in\mathcal{L}} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{i,j\in\mathcal{L}} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}. \end{aligned} \tag{13.39}$$

Figure 13.4 illustrates the different types of atomic interactions. The Lennard-Jones term represents **van der Waals** interactions, while the Coulombic term stands for ionic interactions.

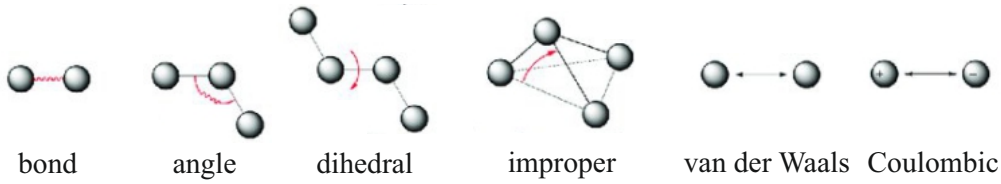


Figure 13.4: Illustration of the different types of interactions considered in force-field potentials.

13.4 Centrosymmetry

Before we end this section, let us introduce one convenient concept for interpreting atomistic data. Many crystals in their equilibrium ground state exhibit what is known as **centrosymmetry**: if any atom has a neighbor located at the relative position $+\boldsymbol{q}$, then it also has a neighboring atom at $-\boldsymbol{q}$. Let us define the centrosymmetry parameter

$$\boxed{\text{CS} = \sum_i \|\boldsymbol{q}_{i,+} + \boldsymbol{q}_{i,-}\|^2} \tag{13.40}$$

where $(\boldsymbol{q}_{i,+}, \boldsymbol{q}_{i,-})$ are the relative positions of the ith pair of opposite neighbors of an atom (relative with respect to the atom's location), and the sum is to carried out over all pairs of neighbors in the centrosymmetric crystal (see Figure 13.5). If the crystal is in its equilibrium ground state, then obviously CS $= 0$. However, if the crystal structure is perturbed, e.g., in the vicinity of defects or near free surfaces, then CS > 0. Moreover, the order of magnitude of CS can be used to identify specific types of defects (e.g., much higher on free surfaces than near defects in the bulk). This is of particular interest to identify those atoms of interest in simulations involving very large numbers of atoms. There are a number of related techniques to identify defects in crystals, which we cannot discuss in detail (see, e.g., reference [82]).

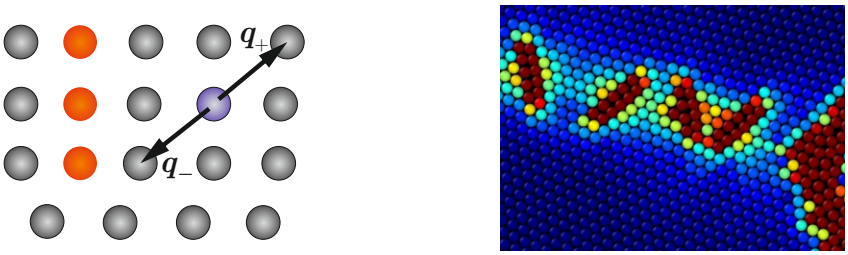


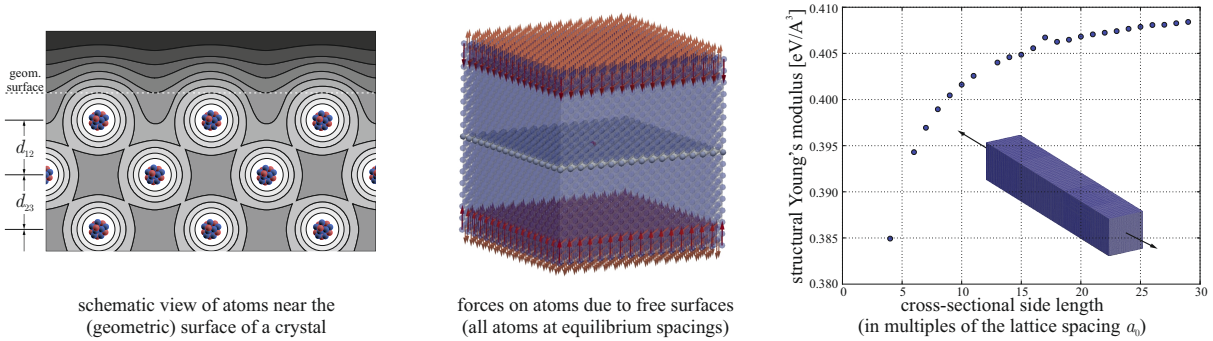
Figure 13.5: The centrosymmetry parameter quantifies deviations of neighboring atoms of a particle atom (here, e.g., of the blue atom) from their symmetric positions. While CS $= 0$ in a perfect single-crystal, CS > 0 in the vicinity of defects (such as the edge dislocation sketched by the red atoms). Shown on the right is an atomic configuration color-coded by CS, indicating in dark blue CS $= 0$, whereas atoms participating in defects (such as the dislocations forming grain boundaries in this example) have CS > 0. This is an effective way to identify atoms of interest out of large ensembles of atoms in a simulation.

13.5 Infinite vs. finite-size samples

Finally, we point that the aforementioned construction of an equilibrium spacing a_0 (irrespective of the potential being used) applies only to *infinite* crystals. It is important to realize that, in an infinite centrosymmetric lattice, the net force on every atom always vanishes for equally spaced atoms, irrespective of the chosen distance between atoms. That is, even when moving atoms to larger or smaller distances than a_0, they will not experience any net force because of symmetrically cancelling interatomic forces. The equilibrium spacing is therefore generally defined as the *minimizer of the energy*, $a_0 = \arg\min V(a)$ (and *not* via solving for zero net forces).

By contrast, in finite-size samples with free surfaces, those atoms near the surface "feel" their missing neighbors and experience a non-zero net force in case of equally spaced atoms. Thus, energy minimization here leads to **surface relaxation**, i.e., changes in atomic spacing or even faceting (i.e., the formation of more complex surface facets) near the free surface. Note that, since the free surfaces break the symmetry of the lattice, in this case the equilibrium configuration can be obtained by energy minimization as well as by solving for zero net forces. For a sufficiently large sample, the interior atoms (which do not "feel" the surface) will obey the same equilibrium spacing a_0 as in an infinite crystal.

Figure 13.6 illustrates the behavior of an atomic crystal near free surfaces. Atoms near the free surface of a crystal relax to the minimum-energy configuration, resulting in layer spacings d_{ij} that vary with depth into the crystal, reaching the bulk lattice spacing far away from the surface. Shown in the middle is a crystal with two free surfaces at the top and bottom which, if all atoms are in the bulk equilibrium spacing a_0 produces inward forces due to missing neighbors near the surface; these forces are responsible for the surface relaxation effect (atoms near the surface accommodating different equilibrium spacings than in the bulk). Finally, shown on the



schematic view of atoms near the (geometric) surface of a crystal

forces on atoms due to free surfaces (all atoms at equilibrium spacings)

Figure 13.6: Illustration of atoms near the free surface of a crystal, which leads to surface relaxation. Shown in the middle are forces on atoms on the top and bottom surfaces of a crystal, whose atoms are all at equal equilibrium spacings a_0. Shown on the right is the impact of the cross-sectional size of a nano-rod (made of pure copper) on the measured effective Young's modulus of the rod (force per cross-sectional area); adapted from reference [4].

right is an example of the effect of free surfaces on mechanical properties: when a nano-rod is pulled uniaxially, one can extract the effective Young's modulus as the ratio of applied force and undeformed cross-sectional area in the linear regime. For large samples the surface plays no significant role and the measured modulus is independent of the size of the sample. However, at very small scales (here, the cross-section includes only up to 30 atoms per side) the surface has an appreciable effect on the measured Young's modulus, since atoms close to the surface behave differently than those in the bulk.

14 Cauchy-Born Approximation & Effective Elastic Properties at Zero Temperature

Interatomic potentials provide the basis for atomistic modeling techniques such as Molecular Dynamics (MD). It is therefore imperative to ensure that a chosen interatomic potential $V(\boldsymbol{q})$, as introduced in Section 13, accurately approximates the behavior of atoms in a chosen material system (unlike ab-initio methods like DFT with higher predictive power, the accuracy of MD results depends strongly on the chosen interatomic potential; a badly calibrated interatomic potential can lead to nonsensical predictions). A common approach to calibrating interatomic potentials is to fit its parameters to known material equilibrium properties, such as the elastic constants, thermal expansion coefficient, etc. This does not ensure that the potential has predictive accuracy, but it at least ensures that a range of material properties are reproduced. Recall how in Section 9 we introduced the **Cauchy-Born rule** as an analytical approach to extracting approximate material models based on applying affine or higher-order Taylor expansions to a unit cell on the microscale. Here, we revisit that strategy to estimate the zero-temperature elastic constants of crystalline solids, which provides a convenient means to compute the elastic constants of any given potential. (It also forms the basis for the quasicontinuum method discussed later.)

Consider an atomic ensemble of N atoms at their undeformed positions $\boldsymbol{Q} = \{\boldsymbol{Q}_1, \dots, \boldsymbol{Q}_N\}$. For crystals, these positions are usually the sites of a Bravais lattice with basis $\mathcal{B} = \{\boldsymbol{A}_1, \dots, \boldsymbol{A}_d\}$, viz.

$$\boldsymbol{Q}_i = \sum_{j=1}^{d} c_j \boldsymbol{A}_j \qquad \text{with} \qquad c_j \in \mathbb{Z}. \tag{14.1}$$

When an average deformation gradient $\boldsymbol{F}$ is applied to a simple Bravais lattice of identical atoms in a pure crystalline material (such as in face-centered cubic (fcc) or body-centered cubic (bcc) metals and ceramics), then it is reasonable to assume that, under small deformations, the response is an *affine deformation* of the crystal lattice. Thus, we may apply the local Cauchy-Born rule with average deformation gradient $\boldsymbol{F}$ to find the deformed atomic positions as

$$\boldsymbol{q}_a = \boldsymbol{F}\boldsymbol{Q}_a, \qquad a = 1, \dots, N. \tag{14.2}$$

This allows us to define an *effective energy density*

$$W(\boldsymbol{F}) = \frac{1}{|\Omega|} V(\{\boldsymbol{q}_1, \dots, \boldsymbol{q}_N\}), \tag{14.3}$$

where $V(\{\boldsymbol{q}_1, \dots, \boldsymbol{q}_N\})$ stands for the energy of atoms in an atomic unit cell, and $|\Omega|$ denotes its volume. In the simplest case of a perfect infinite crystal, each atom experiences the exact same neighborhood changes (all atoms are equal). Then, $V(\{\boldsymbol{q}_1, \dots, \boldsymbol{q}_N\})$ can be taken as the *energy of a single atom* only. In this case, we need to consider only the potential energy of a single atom and apply (14.2) to the atom and all of its neighbors within the cut-off radius. (Since the choice of the coordinate origin is arbitrary, we may choose to place the origin at the atom whose energy is computed, so that $\boldsymbol{Q}_a$ in (14.2) is simply the distance vector of every neighbor with respect to the atom of interest.

The **first Piola-Kirchhoff stress** tensor components now follow as (writing q_i^a for the ith

component of position vector $\boldsymbol{q}_a$)

$$\begin{aligned} P_{iJ}(\boldsymbol{F}) &= \frac{1}{|\Omega|}\frac{\partial}{\partial F_{iJ}}W(\boldsymbol{F}) = \frac{1}{|\Omega|}\sum_{a=1}^{N}\frac{\partial V}{\partial q_k^a}(\boldsymbol{q})\frac{\partial q_k^a}{\partial F_{iJ}} \\ &= -\frac{1}{|\Omega|}\sum_{a=1}^{N} f_k^a(\boldsymbol{q})\frac{\partial}{\partial F_{iJ}}F_{kL}Q_L^a = -\frac{1}{|\Omega|}\sum_{a=1}^{N} f_i^a(\boldsymbol{q})Q_J^a \end{aligned} \tag{14.4}$$

or

$$\boldsymbol{P}(\boldsymbol{F}) = -\frac{1}{|\Omega|}\sum_{a=1}^{N}\boldsymbol{f}_a(\boldsymbol{q})\otimes\boldsymbol{Q}_a, \tag{14.5}$$

with the interatomic force contribution

$$\boldsymbol{f}_a(\boldsymbol{q}) = -\frac{\partial V}{\partial \boldsymbol{q}_a}(\boldsymbol{q}). \tag{14.6}$$

Analogously, we derive the incremental stiffness tensor components as

$$\mathbb{C}_{iJkL} = \frac{\partial P_{iJ}}{\partial F_{kL}} = -\frac{1}{|\Omega|}\frac{\partial}{\partial F_{kL}}\sum_{a=1}^{N} f_i^a(\boldsymbol{q})Q_J^a = -\frac{1}{|\Omega|}\sum_{a,b=1}^{N}\frac{\partial f_i^a}{\partial q_k^b}(\boldsymbol{q})Q_J^aQ_L^b. \tag{14.7}$$

If the potential is known as an analytical expression, then the derivative f_i^a/q_k^b can be computed in closed-form (but it depends on the specific form of a chosen potential, which is why we stop the derivation here and show specific examples below). The usual mapping relations apply to transform tensors between the current and reference configurations. For example, to obtain the components of the elastic stiffness tensor in the current configuration (which are the ones used in linear elasticity theory), we may exploit

$$c_{ijkl} = \frac{1}{\det\boldsymbol{F}}F_{jJ}F_{lL}\mathbb{C}_{iJkL}. \tag{14.8}$$

For simple crystals (those for which the affine Cauchy-Born rule provides a reasonable approximation), the above gives a convenient recipe to compute the *elastic constants* of a crystalline material at zero temperature (i.e., from **Molecular Statics**). When going to non-zero temperature, we must account for thermal vibrations of atoms, which will be discussed later.

Example: Embedded Atom Method

As an example, we revisit the **Embedded Atom Method** (EAM), for which the interatomic potential provides the energy of a single atom as

$$V(\boldsymbol{q}) = \frac{1}{2}\sum_{\alpha\in\mathcal{L}}\Phi(r^\alpha) + U(\rho) \qquad \text{with} \qquad \rho = \sum_{\alpha\in\mathcal{L}} f(r^\alpha), \tag{14.9}$$

where we sum over all neighboring atoms $\alpha\in\mathcal{L}$ of an atom at $\boldsymbol{q}_0 = \boldsymbol{F}\boldsymbol{Q}_0$, and we defined

$$\boldsymbol{r}^\alpha = \boldsymbol{q}_\alpha - \boldsymbol{q}_0 = \boldsymbol{F}\boldsymbol{Q}_\alpha - \boldsymbol{F}\boldsymbol{Q}_0 \tag{14.10}$$

and $r^\alpha = |\boldsymbol{r}^\alpha|$. ρ is the effective electron density at the location of the atom. For convenience we may choose $\boldsymbol{q}_0 = \boldsymbol{0}$ without loss of generality (rigid-body motion does not affect the energy of the atomic ensemble).

Without showing the full derivation (feel free to verify it yourself), the first Piola-Kirchhoff stress tensor is obtained from (14.5) by insertion of (14.9) as

$$P_{iJ} = \frac{1}{|\Omega|} \sum_{\alpha\in\mathcal{L}} \left(\left[U'(\rho) f'(r^\alpha) + \frac{1}{2}\Phi'(r^\alpha) \right] \frac{r_i^\alpha r_j^\alpha}{r^\alpha} \right) F_{Jj}^{-1}. \tag{14.11}$$

The incremental modulus tensor in the current configuration follows as

$$\begin{aligned} c_{ijkl} = \frac{1}{|\Omega|} \Bigg\{ & U''(\rho) \left[\sum_{\alpha\in\mathcal{L}} f'(r^\alpha) \frac{r_i^\alpha r_j^\alpha}{r^\alpha} \right] \left[\sum_{\beta\in\mathcal{L}} f'(r^\beta) \frac{r_k^\beta r_l^\beta}{r^\beta} \right] \\ & + \sum_{\alpha\in\mathcal{L}} \left[\left(\left(U'(\rho) f''(r^\alpha) + \frac{1}{2}\Phi''(r^\alpha) \right) - \frac{1}{r^\alpha} \left(U'(\rho) f'(r^\alpha) + \frac{1}{2}\Phi'(r^\alpha) \right) \right) \frac{r_i^\alpha r_j^\alpha r_k^\alpha r_l^\alpha}{(r^\alpha)^2} \right] \Bigg\}. \end{aligned} \tag{14.12}$$

Thus, for given atomic reference positions $\boldsymbol{Q}$ in the crystal lattice and a given potential with known functions Φ, U and f, the above can be used to compute the (zero-temperature) elastic moduli of an infinite, perfect single-crystal. Note that for fcc and bcc lattices, the resulting moduli will not be isotropic but exhibit *cubic symmetry*. The usual isotropic elastic moduli used in engineering applications must be obtained from the above orthotropic moduli by averaging over many grains in a polycrystal to obtain the *effective elastic moduli of a polycrystal*.

Specific elastic constants, such as effective Young, shear or bulk moduli, can be obtained from the above stiffness tensor. For example, for a crystal with cubic symmetry, the shear modulus is

$$\mu = c_{2323} = \frac{c_{1111} - c_{1122}}{2}, \tag{14.13}$$

while the bulk modulus is

$$K = \frac{c_{1111} + 2c_{1122}}{3}, \tag{14.14}$$

while Young's modulus depends on the direction of loading. Young's modulus in a direction with unit vector $\boldsymbol{n}$ is computed, e.g., from the compliance tensor $s = c^{-1}$ as

$$E(\boldsymbol{n}) = \left(s_{ijkl} n_i n_j n_k n_l \right)^{-1}. \tag{14.15}$$

One may also compute the direction-dependent Poisson's ratio, when stretching in the $\boldsymbol{n}$-direction and being interested in the transverse contraction in the $\boldsymbol{t}$-direction (both vectors being of unit length):

$$\nu(\boldsymbol{n}, \boldsymbol{t}) = -E(\boldsymbol{n}) s_{ijkl} n_i n_j t_k t_l. \tag{14.16}$$

Alternatively, the above elastic moduli can also be obtained directly from the potential energy in a *numerical fashion* – without the need to perform any analytical derivatives (which is convenient, e.g., if the potential cannot be expressed in closed form). The rationale is that the elastic constants are related to the curvature of the potential energy about its equilibrium (after all, we defined $\mathbb{C}_{iJkL} = \partial P_{iJ}/\partial F_{kL} = \partial^2 V/\partial F_{iJ}\partial F_{kL}$). Therefore, one may conveniently deform the crystal at small strains about its equilibrium configuration and extract the effective elastic constants from the curvature of the computed potential energy $V(\boldsymbol{q})$.

For example, the effective bulk modulus of a single crystal can be obtained from volumetrically deforming the crystal:

$$K = \frac{1}{9\Omega}\frac{\partial^2 V(\boldsymbol{F}_{\mathrm{K}}(\gamma)\boldsymbol{Q})}{\partial\gamma^2}\bigg|_{\gamma=0} \qquad \text{with} \qquad \boldsymbol{F}_{\mathrm{K}}(\gamma) = \begin{bmatrix} 1+\gamma & 0 & 0 \\ 0 & 1+\gamma & 0 \\ 0 & 0 & 1+\gamma \end{bmatrix}. \tag{14.17}$$

Analogously, the effective shear modulus may be defined as

$$\mu = C_{44} = \frac{1}{4\Omega}\frac{\partial^2 V(\boldsymbol{F}_{\mu}(\gamma)\boldsymbol{Q})}{\partial\gamma^2}\bigg|_{\gamma=0} \qquad \text{with} \qquad \boldsymbol{F}_{\mu}(\gamma) = \begin{bmatrix} 1 & \gamma & 0 \\ \gamma & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{14.18}$$

while the uniaxial elastic constant can be extracted via

$$c_{1111} = \frac{1}{\Omega}\frac{\partial^2 V(\boldsymbol{F}_{11}(\gamma)\boldsymbol{Q})}{\partial\gamma^2}\bigg|_{\gamma=0} \qquad \text{with} \qquad \boldsymbol{F}_{11}(\gamma) = \begin{bmatrix} 1+\gamma & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{14.19}$$

In all three cases, Ω denotes the volume associated with a single atom, and V is the potential energy of a single atom, whose neighbors deform by the Cauchy-Born rule (14.2) with the respective $\boldsymbol{F}$ for each of the above cases.

In practice, one sets up a perfect infinite crystal in its equilibrium configuration and then applies one of the above deformation gradients $\boldsymbol{F}$ to all atoms many times with varying $\gamma \in [-\epsilon, \epsilon]$, where $\epsilon > 0$ is a small number (thereby making sure that only small strains are imposed). For each choice of γ one computes the energy of a single atom as $V(\gamma) = V(\boldsymbol{F}_{\mu}(\gamma)\boldsymbol{Q})$. This yields a function $V(\gamma)$, which can be fitted to a polynomial (because of the anharmonicity of the potential, it will not be perfectly symmetric – cf. Section 13.2 – but it can be approximated well by a polynomial). Differentiating the resulting polynomial approximations of $V(\gamma) = V(\boldsymbol{F}_{\mu}(\gamma)\boldsymbol{Q})$ can be inserted into the above definitions of K, μ, and c_{1111}

Note that the affine deformation mappings discussed here only apply in simple Bravais lattices and in infinite crystals. They are thus well-suited to describe elastic constants or crystalline behavior under infinitesimal strains. They fail, however, when localization, pattern formation, or inhomogeneous deformation emerges.

Further note that multi-species crystals may form **multi-lattices** (cf. Section 9.3), so that even infinitesimal deformation may not be affine but a multi-lattice description with relative shifts between the various lattices must be accounted for, as discussed in Section 9.3. This is especially relevant in case of phase transformations.

15 Molecular Dynamics

Molecular Dynamics or **MD** is the most popular method of simulating atomic system, promoted by the availability of state-of-the-art computational libraries for high-performance MD calculations such as the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) dating back to reference [72] (LAMMPS is available freely online). To understand the basis of MD, we need to first introduce a number of basic principles from statistical mechanics.

15.1 Some Basics of Statistical Mechanics

Recall that we describe the state of an atomistic system of N atoms by positions $\boldsymbol{q} = \{\boldsymbol{q}_1, \dots, \boldsymbol{q}_N\}$ and momenta $\boldsymbol{p} = \{\boldsymbol{p}_1, \dots, \boldsymbol{p}_N\}$ with $\boldsymbol{p}_i = m_i \dot{\boldsymbol{q}}_i$. The total Hamiltonian yields Hamilton's equations of motion:

$$\mathcal{H}(\boldsymbol{q}, \boldsymbol{p}) = \sum_{i=1}^{N} \frac{|\boldsymbol{p}_i|^2}{2\, m_i} + V(\boldsymbol{q}), \qquad \Rightarrow \qquad \dot{\boldsymbol{q}}_i = \frac{\partial \mathcal{H}}{\partial \boldsymbol{p}_i}, \qquad \dot{\boldsymbol{p}}_i = -\frac{\partial \mathcal{H}}{\partial \boldsymbol{q}_i}. \tag{15.1}$$

Every state of the system is thus described by a point in $6N$-dimensional space (per atom 3 DOF for its position and 3 DOF for its momentum), the so-called **phase space** which is often denoted by Γ. Motion of atoms can be reinterpreted as a **trajectory** $(\boldsymbol{q}(t), \boldsymbol{p}(t))$ through phase space with some initial condition $(\boldsymbol{q}(0), \boldsymbol{p}(0))$.

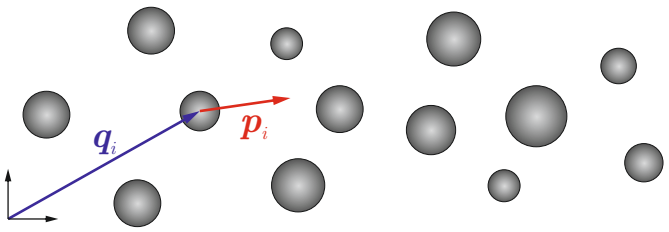


Figure 15.1: Hamiltonian mechanics treats an atomistic system as a collection of N classical particles at positions $\boldsymbol{q}_i$ and with momenta $\boldsymbol{p}_i$.

15.1.1 Example: The Harmonic Oscillator

As a simple example, consider a classical **harmonic oscillator**, i.e., a mass m attached to a linear spring with stiffness k (see Figure 15.2). Its equation of motion $m\ddot{q} + kq = 0$ has the harmonic solution

$$q(t) = A \sin(\omega t) + B \cos(\omega t) \qquad \text{with} \qquad \omega = \sqrt{k/m}. \tag{15.2}$$

By differentiation with respect to time we obtain

$$p(t) = m\dot{q}(t) = m\omega \left[A \cos(\omega t) - B \sin(\omega t)\right]. \tag{15.3}$$

Applying initial conditions $q(0) = q_0$ and $p(0) = p_0$ yields $B = q_0$ and $A = p_0/(m\omega)$. Take, e.g., the simple case $p_0 = 0$ (i.e., starting from rest and from a stretched spring), so that

$$q(t) = q_0 \cos(\omega t), \qquad p(t) = -q_0 m\omega \sin(\omega t). \tag{15.4}$$

By comparing $p(t)$ and $q(t)$, we notice that the mass' motion in phase space describes the ellipse

$$\left(\frac{q}{q_0}\right)^2 + \left(\frac{p}{q_0 m\omega}\right)^2 = 1 \qquad \Leftrightarrow \qquad q^2 + \left(\frac{p}{m\omega}\right)^2 = q_0^2, \tag{15.5}$$

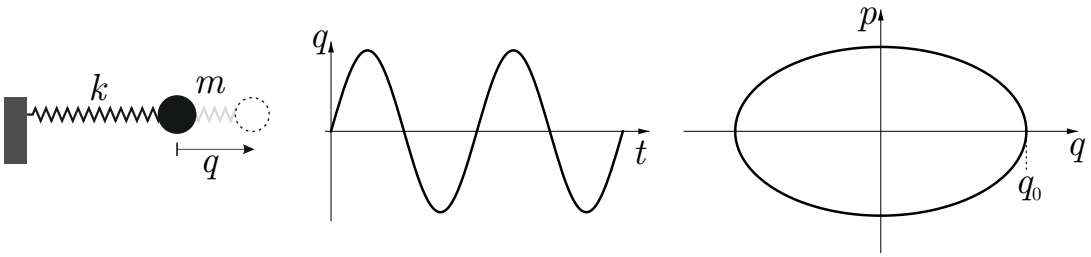


Figure 15.2: The classical harmonic oscillator (particle of mass m attached to a spring of stiffness k along with the state-space description $q(t)$ of its motion and its phase-space representation (q,p).

with a unique ellipse for each initial condition $q_0 = q(0)$. Note that the Hamiltonian of the particle evaluates to

$$\begin{aligned}\mathcal{H} &= \frac{p^2}{2m} + \frac{k}{2}q^2 = \frac{1}{2m}q_0^2 m^2\omega^2 \sin^2(\omega t) + \frac{k}{2}q_0^2\cos^2(\omega t)\\ &= \frac{k}{2}q_0^2\left[\sin^2(\omega t) + \cos^2(\omega t)\right] = \frac{k}{2}q_0^2.\end{aligned} \tag{15.6}$$

That is, without external forcing, the Hamiltonian $\mathcal{H} = \frac{k}{2}q_0^2$ is constant along the particle's trajectory in phase space and, moreover, the area enclosed by the ellipse (which is proportional to q_0^2, see (15.5)) grows with increasing $\mathcal{H}$. Specifically, the ellipse's major principal axis is

$$q_0 = \sqrt{2\mathcal{H}/k}. \tag{15.7}$$

15.1.2 Statistical Averages & Ensembles

The above treatment can be generalized for systems in classical mechanics. To this end, notice that for the total Hamiltonian of a system of N particles, $\mathcal{H} = \mathcal{H}(\boldsymbol{q},\boldsymbol{p})$, we have

$$\frac{\mathrm{d}\mathcal{H}}{\mathrm{d}t} = \sum_{i=1}^{N}\left(\frac{\partial\mathcal{H}}{\partial\boldsymbol{q}_i}\cdot\dot{\boldsymbol{q}}_i + \frac{\partial\mathcal{H}}{\partial\boldsymbol{p}_i}\cdot\dot{\boldsymbol{p}}_i\right) = \sum_{i=1}^{N}\left(-\dot{\boldsymbol{p}}_i\cdot\dot{\boldsymbol{q}}_i + \dot{\boldsymbol{q}}_i\cdot\dot{\boldsymbol{p}}_i\right) = 0, \tag{15.8}$$

i.e., the total energy is conserved, and the Hamiltonian is constant along trajectories in space. Moreover, since motion is unique, trajectories in space cannot cross themselves. In other words, given the current position and momentum of each and every particle in the system, the system's time evolution (for an autonomous, not externally driven system) is unique. Starting from the same configuration hence cannot evolve in more than one direction, which is why trajectories cannot cross.

Especially at finite temperature, where atomic vibrations about their equilibrium positions are important contributions to changes in particle positions and momentum, it is convenient to compute time-averaged quantities.

Similar in spirit to homogenization, we therefore define macroscopic, effective observables like the state variables in continuum mechanics (e.g., energy, temperature, stress, etc.). Since the atomistic system is uniquely defined by $(\boldsymbol{q},\boldsymbol{p})$, we may assume that each **observable** A can be expressed as a function $A = A(\boldsymbol{q},\boldsymbol{p})$. Unlike in computational homogenization discussed before, the key challenge here is to bridge across *time scales*, which is why one may define the effective quantity $\mathcal{A}(t)$ at time t as the **time average**

$$\mathcal{A}(t) = \frac{1}{\Delta t}\int_t^{t+\Delta t} A\left(\boldsymbol{q}(\tau),\boldsymbol{p}(\tau)\right)\,\mathrm{d}\tau, \tag{15.9}$$

where the time interval Δt must be chosen large compared to the atomic vibrations.

Having introduced effective quantities, we define **equilibrium** in this context as $\mathcal{A}(t) = \text{const.}$, in which case $\mathcal{A}$ is independent of time t. For almost all initial conditions and without providing external power, the limit $\overline{\mathcal{A}}$ obtained as $\Delta t \to \infty$ exists and represents the true **macroscopic quantity**:

$$\overline{\mathcal{A}} = \lim_{\Delta t \to \infty} \frac{1}{\Delta t} \int_t^{t+\Delta t} A\left(\boldsymbol{q}(\tau), \boldsymbol{p}(\tau)\right) \, \mathrm{d}\tau. \tag{15.10}$$

Numerically, such quantities can – more or less – easily be computed during numerical simulations (which calculate the particle positions and momenta at discrete times t_i) by making the approximation

$$\overline{\mathcal{A}} \approx \frac{1}{n_\infty} \sum_{i=1}^{n_\infty} A\left(\boldsymbol{q}(t_i), \boldsymbol{p}(t_i)\right). \tag{15.11}$$

The above time averaging is numerically simple but computationally expensive and theoretically inconvenient. Therefore, one alternatively uses statistical approaches based on probability densities.

In statistical mechanics, one distinguishes between **macrostates** (described by the macroscopic observables) and the associated **microstates** (describing all realizations of a system in agreement with a particular macrostate). For example, we may fix the total energy of a system, i.e.,

$$\sum_{i=1}^{N} \frac{|\boldsymbol{p}_i|^2}{2m_i} + V(\boldsymbol{q}) = E = \text{const.} \tag{15.12}$$

Here, energy E is a macroscopic observable, which defines a macrostate (viz. that the energy is E), whereas there are many microstates (i.e., particular atomic positions and momenta) that satisfy the constraint (15.12). All microstates in agreement with a given macrostate are collectively referred to as an **ensemble**.

Depending on the macro-constraint, one differentiates between the classical ensembles:

- the **microcanonical ensemble** (NVE-ensemble) which describes an *isolated* system with constant energy E (besides constant particle number N and constant volume V, therefore we refer to this as the NVE-ensemble),
- the **canonical ensemble** (NVT-ensemble) which describes a system in contact with a heat bath, i.e., at constant temperature (and constant particle number and volume),
- the **grand canonical ensemble** (μVT-ensemble) which is an extension of the canonical ensemble and describes a system in contact with both heat and particle baths (constant temperature, volume and chemical potential μ).

There are further ensembles of practical relevance such as, e.g., the **isothermal-isobaric ensemble** (NpT), which maintains both a constant temperature and a constant pressure.

If only a macrostate such as E is prescribed, there will be large numbers of potential microstates (e.g., for the harmonic oscillator the entire ellipse presents admissible microstates that are compatible with a fixed energy E). Note that we can draw an analogy to the computational homogenization problem discussed in Section 6. There, we imposed average strains or stresses

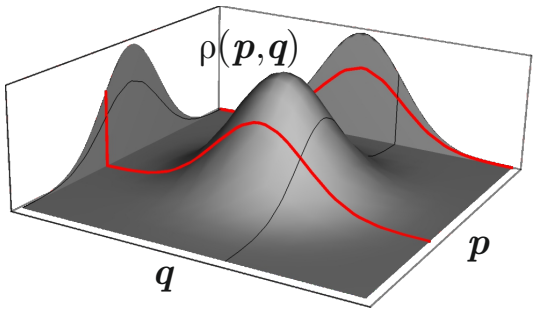


Figure 15.3: Schematic illustration of the probability distribution $\rho(\boldsymbol{q},\boldsymbol{p})$, defining the probability of finding the system at $(\boldsymbol{q},\boldsymbol{p})$ in phase space.

(the macrostates) onto the RVE, while points inside the RVE could undergo arbitrary displacements (the microstates) as long as their average would comply with the given average strains or stresses.

If one takes an ensemble with a given macro-constraint, the full system will follow trajectories that comply with the imposed macrostate. For such a constrained system, we can define other macroscopic quantities $\mathcal{A}$. For example, while fixing the energy E of an atomistic system, the atoms are still free to move through space – compliant with the given energy E – so that we could, e.g., compute the average velocity of all atoms, which is not necessarily constant, even though E is fixed. In practice, this implies that we should sample over all possible configurations that have energy E and compute the average atom velocity over all such configurations. In practice, it is impossible to study every possible configuration, but one can make approximations.

Recall that a trajectory through space depends on the initial conditions, which means that in order to obtain an approximate average $\mathcal{A}$, one should ideally run very many, n to be specific, parallel simulations with the same macrostate but starting with different initial conditions so as to sample over as many points in phase space as possible that comply with the given macrostate. This yields the approximate *ensemble average*

$$\langle\mathcal{A}\rangle \approx \frac{1}{n}\sum_{i=1}^{n}\mathcal{A}(\boldsymbol{q}_n,\boldsymbol{p}_n), \tag{15.13}$$

where each $(\boldsymbol{q}_n,\boldsymbol{p}_n)$ is a configuration that complies with the macrostate. If we consider the limit $n\to\infty$, we obtain the macroscopic observable $\mathcal{A}$ as the **ensemble average** or **phase average**

$$\boxed{\langle\mathcal{A}\rangle = \int_\Gamma A(\boldsymbol{q},\boldsymbol{p})\,\rho(\boldsymbol{q},\boldsymbol{p})\,\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{q}} \tag{15.14}$$

where $\rho(\boldsymbol{q},\boldsymbol{p})$ is a **probability distribution function** or **probability density**, which satisfies

$$\boxed{\int_\Gamma \rho(\boldsymbol{q},\boldsymbol{p})\,\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{q} = 1} \tag{15.15}$$

Roughly speaking, the probability density $\rho(\boldsymbol{q},\boldsymbol{p})$ defines the likelihood of finding an ensemble in the configuration $(\boldsymbol{q},\boldsymbol{p})$.

For the phase average to make physical sense, Ludwig Boltzmann introduced the so-called hypothesis of **ergodicity**: *if one waits sufficiently long, any physical system will visit* ***all*** *possible points in phase space consistent with a given macrostate.* This admits replacing the discrete

sum over a countable number n of measurements by the integral in phase space. That is, it implies that

$$\langle \mathcal{A} \rangle \to \overline{\mathcal{A}} \quad \text{as} \quad n \to \infty, \tag{15.16}$$

and we will assume

$$\boxed{\langle \mathcal{A} \rangle \approx \overline{\mathcal{A}}}. \tag{15.17}$$

15.2 Center of mass coordinates

In the following sections, we will link thermal properties to atomic vibrations. Therefore, it is convenient to reformulate atomic positions and momenta as those relative to the center of mass and the mean momentum of all N atoms, which we define as

$$\boldsymbol{Q} = \frac{\sum_{i=1}^N m_i \boldsymbol{q}_i}{\sum_{i=1}^N m_i}, \qquad \boldsymbol{P} = \sum_{i=1}^N m_i \dot{\boldsymbol{q}}_i = M\dot{\boldsymbol{Q}}, \tag{15.18}$$

with the total mass $M = \sum_{i=1}^N m_i$. Balance of linear momentum for the complete system of particles becomes

$$M\ddot{\boldsymbol{Q}} = \sum_{i=1}^N \boldsymbol{f}_{\text{ext}} = \boldsymbol{F}_{\text{ext}}. \tag{15.19}$$

Next, we express all atomic positions and momentum with respect to the average values:

$$\boldsymbol{q}_i = \boldsymbol{Q} + \delta\boldsymbol{q}_i, \qquad \boldsymbol{p}_i = m_i\dot{\boldsymbol{Q}} + \delta\boldsymbol{p}_i \qquad \text{such that} \qquad \delta\boldsymbol{p}_i = m_i\delta\dot{\boldsymbol{q}}_i. \tag{15.20}$$

The total Hamiltonian becomes

$$\begin{aligned}
\mathcal{H} &= \sum_{i=1}^N \frac{\left\| m_i\dot{\boldsymbol{Q}} + \delta\boldsymbol{p}_i \right\|^2}{2m_i} + V(\boldsymbol{Q} + \delta\boldsymbol{q}_1, \dots, \boldsymbol{Q} + \delta\boldsymbol{q}_N) \\
&= \sum_{i=1}^N \left[\frac{m_i^2 \left\| \dot{\boldsymbol{Q}} \right\|^2}{2m_i} + \frac{\|\delta\boldsymbol{p}_i\|^2}{2m_i} + \frac{m_i\dot{\boldsymbol{Q}} \cdot \delta\boldsymbol{p}_i}{m_i} \right] + V(\boldsymbol{Q} + \delta\boldsymbol{q}_1, \dots, \boldsymbol{Q} + \delta\boldsymbol{q}_N) \\
&= \frac{1}{2}\left(\sum_{i=1}^N m_i \right) \left\| \dot{\boldsymbol{Q}} \right\|^2 + \sum_{i=1}^N \frac{\|\delta\boldsymbol{p}_i\|^2}{2m_i} + \dot{\boldsymbol{Q}} \cdot \sum_{i=1}^N \delta\boldsymbol{p}_i + V(\boldsymbol{Q} + \delta\boldsymbol{q}_1, \dots, \boldsymbol{Q} + \delta\boldsymbol{q}_N)
\end{aligned} \tag{15.21}$$

Note that

$$\sum_{i=1}^N \delta\boldsymbol{p}_i = \sum_{i=1}^N m_i\,\delta\dot{\boldsymbol{q}}_i = \frac{\mathrm{d}}{\mathrm{d}t}\sum_{i=1}^N m_i\,\delta\boldsymbol{q}_i = \frac{\mathrm{d}}{\mathrm{d}t}\sum_{i=1}^N m_i\,(\boldsymbol{q}_i - \boldsymbol{Q}) = \frac{\mathrm{d}}{\mathrm{d}t}\,(M\boldsymbol{Q} - M\boldsymbol{Q}) = \boldsymbol{0}, \tag{15.22}$$

so that the total Hamiltonian reduces to

$$\mathcal{H} = \frac{\|\boldsymbol{P}\|^2}{2M} + \sum_{i=1}^N \frac{\|\delta\boldsymbol{p}_i\|^2}{2m_i} + V(\boldsymbol{Q} + \delta\boldsymbol{q}_1, \dots, \boldsymbol{Q} + \delta\boldsymbol{q}_N), \tag{15.23}$$

and we conclude the following Hamiltonian equations for the mean positions and momenta:

$$\dot{\boldsymbol{Q}} = \frac{\partial\mathcal{H}}{\partial\boldsymbol{P}}, \qquad \dot{\boldsymbol{P}} = -\frac{\partial\mathcal{H}}{\partial\boldsymbol{Q}} = -\sum_{i=1}^N \frac{\partial V}{\partial\boldsymbol{q}_i} = -\sum_{i=1}^N(-\boldsymbol{f}_i) = \boldsymbol{F}_{\text{ext}}. \tag{15.24}$$

Furthermore, notice that

$$\delta\dot{\boldsymbol{q}}_i = \frac{\partial\mathcal{H}}{\partial\delta\boldsymbol{p}_i}, \qquad \delta\dot{\boldsymbol{p}}_i = \dot{\boldsymbol{p}}_i - \dot{\boldsymbol{P}} = -\frac{\partial\mathcal{H}}{\partial\delta\boldsymbol{q}_i} - \frac{m_i}{M}\dot{\boldsymbol{P}} = \boldsymbol{f}_i - \frac{m_i}{M}\boldsymbol{F}_{\text{ext}}. \tag{15.25}$$

Therefore, if no external forces act on the ensemble ($\boldsymbol{F}_{\text{ext}} = \boldsymbol{0}$), then the average motion decouples from the atomic fluctuations, so we may work with the total Hamiltonian

$$\mathcal{H} = \sum_{i=1}^{N} \frac{\|\delta\boldsymbol{p}_i\|^2}{2m_i} + V(\boldsymbol{Q} + \delta\boldsymbol{q}_1, \dots, \boldsymbol{Q} + \delta\boldsymbol{q}_N) \tag{15.26}$$

for which

$$\delta\dot{\boldsymbol{q}}_i = \frac{\partial\mathcal{H}}{\partial\delta\boldsymbol{p}_i}, \qquad \delta\dot{\boldsymbol{p}}_i = -\frac{\partial\mathcal{H}}{\partial\delta\boldsymbol{q}_i}. \tag{15.27}$$

Since the mean motion of an ensemble is often of minor interest (unless material behavior is rate dependent and the mean motion is so fast that it affects the material behavior), we will – in the following – work only with the perturbations $(\delta\boldsymbol{q}, \delta\boldsymbol{p})$ to describe atomic motion. For convenience, we will drop the δ in all perturbations and simply write $(\boldsymbol{q}, \boldsymbol{p})$ (while assuming that these are the positions and momenta with respect to the mean position and momentum of an atomic ensemble). In the following we discuss particular ensembles defined by the chosen constant macrostates.

15.3 The microcanonical ensemble

The **microcanonical ensemble** considers an isolated system containing a constant number of N atoms, contained in a constant volume V and having a constant energy E (this is the NVE-ensemble). Because the Hamiltonian of an isolated system is constant along trajectories in phase space, all microstates satisfying $\mathcal{H} = E$ form a closed *hypersurface* (for the harmonic oscillator this was simply an ellipse, see Section 15.1.1), which encloses a volume $V_R(E)$:

$$V_R(E) = \int_\Gamma H\big(E - \mathcal{H}(\boldsymbol{q},\boldsymbol{p})\big)\,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p} = \int_{\mathcal{H}<E} \mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p} \tag{15.28}$$

with the Heaviside jump function $H(\cdot)$. This allows us to define the **density of states** as

$$D(E) = \frac{\mathrm{d}V_R}{\mathrm{d}E}(E). \tag{15.29}$$

Loosely speaking, $D(E)$ indicates how quickly V_R changes with energy E. Alternatively, it is the density of states contained between the hypersurfaces defined by E and $E + \mathrm{d}E$, which defines the *hypershell*

$$\Sigma(E, \Delta E) = \big\{(\boldsymbol{q},\boldsymbol{p}) \;:\; E \le \mathcal{H}(\boldsymbol{q},\boldsymbol{p}) \le E + \Delta E\big\}. \tag{15.30}$$

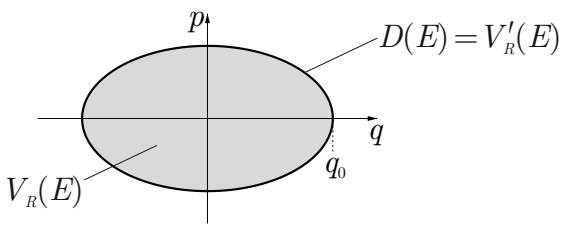


Figure 15.4: The hypersurface $D(E)$ bounds the volume V_R for a given energy E.

The distribution function (or probability density) ρ of the microcanonical ensemble can now be defined as follows. Let us define $\rho(E)$ such that it is constant within $\Sigma(E,\Delta E)$ and zero everywhere else. Note that we must have

$$\int_\Gamma \rho(\boldsymbol{q},\boldsymbol{p};E,\Delta E)\,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p} = 1, \tag{15.31}$$

which results in

$$\rho(\boldsymbol{q},\boldsymbol{p};E,\Delta E) = \begin{cases} \dfrac{1}{V_R(E+\Delta E)-V_R(E)}, & \text{if } (\boldsymbol{q},\boldsymbol{p})\in\Sigma(E,\Delta E),\\ 0, & \text{else.}\end{cases} \tag{15.32}$$

If we now take the limit $\Delta E\to 0$, the phase average becomes

$$\langle\mathcal{A}\rangle = \lim_{\Delta E\to 0}\int_\Gamma \mathcal{A}(\boldsymbol{q},\boldsymbol{p})\rho(\boldsymbol{q},\boldsymbol{p};E,\Delta E)\,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p} = \lim_{\Delta E\to 0}\frac{\int_{\Sigma(E,\Delta E)}\mathcal{A}(\boldsymbol{q},\boldsymbol{p})\,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p}}{V_R(E+\Delta E)-V_R(E)}. \tag{15.33}$$

Next, we exploit that $\Sigma(E,\Delta E) = V_R(E+\Delta E)\setminus V_R(E)$ and we expand numerator and denominator by ΔE to arrive at

$$\begin{aligned}\langle\mathcal{A}\rangle &= \lim_{\Delta E\to 0}\frac{\frac{1}{\Delta E}\left[\int_{V_R(E+\Delta E)}\mathcal{A}(\boldsymbol{q},\boldsymbol{p})\,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p} - \int_{V_R(E)}\mathcal{A}(\boldsymbol{q},\boldsymbol{p})\,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p}\right]}{\frac{V_R(E+\Delta E)-V_R(E)}{\Delta E}}\\ &= \frac{1}{D(E)}\frac{\partial}{\partial E}\int_{V_R(E)}\mathcal{A}(\boldsymbol{q},\boldsymbol{p})\,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p}\\ &= \frac{1}{D(E)}\frac{\partial}{\partial E}\int_\Gamma \mathcal{A}(\boldsymbol{q},\boldsymbol{p})\,H\big(E-\mathcal{H}(\boldsymbol{q},\boldsymbol{p})\big)\,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p},\end{aligned} \tag{15.34}$$

where we first replaced the definition of the partial derivative and then used the Heaviside function in order to expand the integral to all of phase space in the last line. Finally, we use that $H'(\cdot)=\delta(\cdot)$ with the Dirac delta function $\delta(\cdot)$, so that we obtain the phase average of $\mathcal{A}$ as

$$\langle\mathcal{A}\rangle = \frac{1}{D(E)}\int_\Gamma \mathcal{A}(\boldsymbol{q},\boldsymbol{p})\,\delta\left(E-\mathcal{H}(\boldsymbol{q},\boldsymbol{p})\right)\,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p} = \int_\Gamma \mathcal{A}(\boldsymbol{q},\boldsymbol{p})\,\rho(\boldsymbol{q},\boldsymbol{p},E)\,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p} \tag{15.35}$$

with the **microcanonical distribution function**

$$\rho(\boldsymbol{q},\boldsymbol{p},E) = \frac{\delta\left(E-\mathcal{H}(\boldsymbol{q},\boldsymbol{p})\right)}{D(E)}. \tag{15.36}$$

Notice that the normalization constraint (15.31) implies that

$$\int_\Gamma \delta\left(E-\mathcal{H}(\boldsymbol{q},\boldsymbol{p})\right)\,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p} = D(E). \tag{15.37}$$

Now that we have the probability density, we can look at macroscopic state variables. For example, we may define the **internal energy** as

$$U = \langle\mathcal{H}\rangle, \tag{15.38}$$

so that for an isolated system $U=E$. From thermodynamics, recall that the **entropy** S, **temperature** T, and the internal energy U are related via

$$\frac{\partial S}{\partial U} = \frac{1}{T} \qquad\Rightarrow\qquad T = \frac{1}{\partial S/\partial E}. \tag{15.39}$$

Boltzmann postulated the entropy as $S = k_B \ln \Omega$, where k_B is Boltzmann's constant and Ω is a total number of microstates compliant with a given macrostate. In our case, this can be restated as

$$S(E) = k_B \ln D(E), \tag{15.40}$$

which in the *thermodynamic limit* $N \to \infty$ can be shown to approach (without providing the detailed derivation here)

$$S(E) = k_B \ln V_R(E). \tag{15.41}$$

Using (15.34) and the fact that $E = \text{const.}$ lets us compute (with x_i denoting any component of the $(\boldsymbol{q}, \boldsymbol{p})$-vector in phase space)

$$\begin{aligned}
\left\langle x_i \frac{\partial \mathcal{H}}{\partial x_j} \right\rangle &= \frac{1}{D(E)} \frac{\partial}{\partial E} \int_{V_R(E)} x_i \frac{\partial \mathcal{H}}{\partial x_j} \,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p} \\
&= \frac{1}{D(E)} \frac{\partial}{\partial E} \int_{V_R(E)} x_i \frac{\partial (\mathcal{H}-E)}{\partial x_j} \,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p} \\
&= \frac{1}{D(E)} \frac{\partial}{\partial E} \left[\int_{V_R(E)} \frac{\partial x_i (\mathcal{H}-E)}{\partial x_j} \,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p} - \int_{V_R(E)} \delta_{ij} (\mathcal{H}-E) \,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p} \right] \\
&= \frac{1}{D(E)} \frac{\partial}{\partial E} \left[\int_{S(E)} x_i (\mathcal{H}-E) n_j \,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p} - \int_{V_R(E)} \delta_{ij} (\mathcal{H}-E) \,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p} \right]
\end{aligned} \tag{15.42}$$

with $\boldsymbol{n}$ denoting the normal vector on the hypersurface $S(E)$. Note that $\mathcal{H} = E$ on $S(E)$, so that the first integral vanishes and we are left with

$$\begin{aligned}
\left\langle x_i \frac{\partial \mathcal{H}}{\partial x_j} \right\rangle &= \frac{1}{D(E)} \frac{\partial}{\partial E} \int_{V_R(E)} \delta_{ij} (E-\mathcal{H}) \,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p} \\
&= \frac{\delta_{ij}}{D(E)} \frac{\partial}{\partial E} \int_{\Gamma} (E-\mathcal{H})\, H(E-\mathcal{H}) \,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p} \\
&= \frac{\delta_{ij}}{D(E)} \left[\int_{\Gamma} H(E-\mathcal{H}) \,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p} + \int_{\Gamma} (E-\mathcal{H})\, \delta(E-\mathcal{H}) \,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p} \right].
\end{aligned} \tag{15.43}$$

Recall that the first integral is by definition $V_R(E)$, cf. (15.28), and notice that the second integral vanishes. Therefore, exploiting (15.41) yields

$$\left\langle x_i \frac{\partial \mathcal{H}}{\partial x_j} \right\rangle = \delta_{ij} \frac{V_R(E)}{D(E)} = \delta_{ij} \frac{V_R(E)}{\frac{\partial V_R}{\partial E}(E)} = \delta_{ij} \left[\frac{\partial}{\partial E} \ln V_R(E) \right]^{-1} = \delta_{ij} \frac{k_B}{\partial S / \partial E} = \delta_{ij} k_B T. \tag{15.44}$$

The relation

$$\left\langle x_i \frac{\partial \mathcal{H}}{\partial x_j} \right\rangle = \delta_{ij} k_B T. \tag{15.45}$$

is known as the (most general form of the) **equipartition theorem**.

Finally, let us take $x_i = p_i$ (i.e., we consider the ith component of the momentum of one of the N atoms), so that $\partial \mathcal{H}/\partial x_j = \partial \mathcal{H}/\partial p_j = \dot{q}_j$. Then,

$$\delta_{ij} k_B T = \left\langle p_i \frac{\partial \mathcal{H}}{\partial p_j} \right\rangle = \langle p_i \dot{q}_j \rangle = \left\langle p_i \frac{p_j}{m_j} \right\rangle. \tag{15.46}$$

Considering $i = j$ yields an interesting reinterpretation of temperature:

$$\boxed{\left\langle \frac{p_i^2}{2m_i} \right\rangle = \frac{1}{2} k_B T} \tag{15.47}$$

Summing over all $3N$ degrees of freedom thus gives the relation

$$\sum_{i=1}^{3N} \left\langle \frac{p_i^2}{2m_i} \right\rangle = \sum_{i=1}^{3N} \left\langle \frac{1}{2} m_i \dot{q}_i^2 \right\rangle = \langle \mathcal{K}_{\text{vibr.}} \rangle = \frac{3N}{2} k_B T \qquad \Leftrightarrow \qquad \boxed{T = \frac{2\langle \mathcal{K}_{\text{vibr.}} \rangle}{3N k_B}} \tag{15.48}$$

which links the average vibrational energy to the temperature of the ensemble. In other words, the average kinetic energy of all atoms in the microcanonical ensemble is proportional to the temperature of the ensemble. This also offers us a way to compute the energy of the atomic ensemble from the atomic motion. It further is a restatement of the *equipartition theorem*, which implies that all degrees of freedom contribute equally to the kinetic energy of the system.

Alternatively, using $x_i = q_i$ in (15.45) (i.e., we now pick the ith component of the position vector of one of the atoms in the ensemble) gives

$$\delta_{ij} k_B T = \left\langle q_i \frac{\partial \mathcal{H}}{\partial q_j} \right\rangle . \tag{15.49}$$

Let us pick $i = j$ and, according to Einstein's summation convention in (15.49), sum over all 3 degrees of freedom of an atom α (using $\delta_{ii} = 3$). This leads to

$$3 k_B T = \sum_{i=1}^{3} \left\langle q_i \frac{\partial \mathcal{H}}{\partial q_i} \right\rangle = \sum_{i=1}^{3} \left\langle q_i \frac{\partial V}{\partial q_i} \right\rangle = -\langle \boldsymbol{q}_\alpha \cdot \boldsymbol{f}_\alpha \rangle \tag{15.50}$$

with force $\boldsymbol{f}_a = -\partial V / \partial \boldsymbol{q}_a$. If we sum over all N atoms in the system and call the resulting quantity the **virial** $\mathcal{W}$ of the system, we arrive at

$$-\sum_{\alpha=1}^{N} \langle \boldsymbol{q}_\alpha \cdot \boldsymbol{f}_\alpha \rangle = -\langle \mathcal{W} \rangle = 3 N k_B T. \tag{15.51}$$

By comparing to (15.48), note that

$$\langle \mathcal{W} \rangle = -2 \langle \mathcal{K}_{\text{vibr.}} \rangle, \tag{15.52}$$

which is referred to as the **virial theorem**.

Without derivation, we stress that other macrostate quantities can be derived similarly, including the **virial stress tensor** (which lives, like the Cauchy stress tensor, in the current configuration)

$$\boxed{\boldsymbol{\sigma} = -\frac{1}{|\Omega|} \sum_{a=1}^{N} \left\langle \frac{\boldsymbol{p}_a \otimes \boldsymbol{p}_a}{2m_a} + \boldsymbol{f}_a \otimes \boldsymbol{q}_a \right\rangle} \qquad \text{with} \qquad \boldsymbol{f}_a = -\frac{\partial V}{\partial \boldsymbol{q}_a}. \tag{15.53}$$

This defines the effective, average stress tensor of an atomistic system. Notice that the above definition includes both kinetic terms and terms relating to the interaction potential). In static equilibrium ($\boldsymbol{p}_\alpha \equiv \boldsymbol{0}$ for all α), the above coincides with the *Cauchy-Born-based stress tensor* derived in Section 14.

15.4 Brief Summary

In summary, we resort to a *phase-space* description to describe the configurations of an atomistic system, where – assuming *ergodicity* – the *probability distribution* reveals the likelihood of all *microstates* (i.e., of all points in phase space). An atomistic *ensemble* fixes certain *macrostates* (such as the microcanonical ensemble keeps N, V, and E constant), and we may compute other quantities of an ensemble, such as the average *temperature* or the *virial stress tensor* of an ensemble. Any such averages imply phase-space averaging over all those microstates compliant with the given macrostate, conveniently evaluated using the probability distribution. We note that, although the above derivation focused on the microcanonical ensemble, analogous derivations are available for *other ensembles* (such as for the *grand canonical* and *canonical ensembles*). For brevity, we omit detailed discussions here and refer to, e.g., reference [89].

16 Atomistic Solution Algorithms

So far, we have discussed the theoretical basis for atomistic modeling: from interatomic potentials that encode the constitutive behavior through the forces acting on atoms, to the Hamiltonian treatment of atoms as classical particles, to the statistical-mechanical basis for atomistic ensembles and the definition of effective quantities such as temperature and stresses. What we have not discussed so far is how **Molecular Statics** (**MS**) and **Molecular Dynamics** (**MD**) work in practice.

We recapitulate that the equations of motion for every atom i, treated as classical Hamiltonian particles, are given by

$$m_i \ddot{\boldsymbol{q}}_i = \boldsymbol{f}_i = -\frac{\partial \mathcal{V}}{\partial \boldsymbol{q}_i} \qquad \text{for} \qquad i = 1, \dots, N, \tag{16.1}$$

where $\boldsymbol{f}_i$ represents the net force acting on atom i. Unless for simple toy examples, the above equations are to be solved numerically, oftentimes for very large numbers of particles (consider the discrepancy between typical interatomic spacings and the characteristic sizes of samples of interest in applications). In the following, we will separate between a *zero-temperature* treatment (in which thermal vibrations play no role) and *finite-temperature* simulations (in which thermal vibrations are significant and the above discussion of atomistic ensembles gains importance).

16.1 Zero-Temperature Molecular Statics

The simplest case, molecular statics at zero temperature, seeks to solve the system of equations given by

$$\boldsymbol{0} = \boldsymbol{f}_i = -\frac{\partial \mathcal{V}}{\partial \boldsymbol{q}_i} \qquad \text{for} \qquad i = 1, \dots, N. \tag{16.2}$$

For a given interatomic potential $\mathcal{V}$, the above is, in principle, straightforward. However, owing to the symmetry of atomic lattices, many local energy minima exist (and oftentimes many global maxima), which is why the solution algorithm can be numerically demanding. In addition, for large numbers N the assembly and storage of system-wide matrices becomes computationally intractable, which is why *matrix-free iterative solvers* are popular. Typical solution algorithms include solvers of conjugate gradient and steepest descent type; the so-called **Fast Inertial Relaxation Engine** (FIRE) is a related popular scheme (see reference [12]).

Special attention must be given to the application of **boundary conditions**. Essential BCs to individual atoms are generally to be avoided since large displacements applied to atoms may result in atomic distances that exceed the cut-off distance, so atoms simply do not 'feel' each other anymore. Instead, one rather introduces *padding regions* to apply essential BCs over large numbers of atoms. *Periodic BCs* are a common way to simulate RVEs.

A general complication of particle methods such as atomistics (when compared to, e.g., FEM) is the necessity for frequent **neighborhood searches** and the associated updates of neighbor lists of all N atoms. At every load or time step in a simulation, it must be ensured that the calculation of $\boldsymbol{f}_i$ in (16.1) and (16.2) uses the correct and complete list of interacting neighbors for each atom (and these interactions may change as a consequence of atomic motion).

16.2 Zero-Temperature Molecular Dynamics

Molecular Dynamics (MD) integrates the dynamic equations of motion (16.1) over time t, which is generally done in an explicit fashion to avoid the assembly and solution of a global system for computational reasons. (Especially in massively-parallel implementations, the explicit scheme is the method of choice for computational ease.)

The most common time integration scheme is called **velocity-Verlet** and can be summarized as follows. Using constant time increments $\Delta t = t^\alpha - t^{\alpha-1}$ so that $t^\alpha = \alpha \cdot \Delta t$ and abbreviating $\boldsymbol{q}^\alpha = \boldsymbol{q}(t^\alpha)$, we use the expansion

$$\boldsymbol{q}_i(t^\alpha \pm \Delta t) = \boldsymbol{q}_i^{\alpha\pm1} = \boldsymbol{q}_i^\alpha \pm \dot{\boldsymbol{q}}_i^\alpha \Delta t + \frac{1}{2}\ddot{\boldsymbol{q}}_i^\alpha \Delta t^2 + O(\Delta t^3). \tag{16.3}$$

From (16.1), we know the acceleration of atom i as

$$\boldsymbol{a}_i^\alpha = \ddot{\boldsymbol{q}}_i^\alpha = \boldsymbol{f}_i^\alpha / m_i. \tag{16.4}$$

Thus, for known initial conditions $\boldsymbol{q}^0 = \{\boldsymbol{q}_1(0), \dots, \boldsymbol{q}_N(0)\}$ and $\boldsymbol{p}_0 = \{\boldsymbol{p}_1(0), \dots, \boldsymbol{p}_N(0)\}$ we can integrate the above equations of motion. Note that, in order to ensure time reversibility (which comes with improved accuracy alongside quadratic convergence), one usually uses a so-called *average acceleration* (*leapfrog*) scheme, which assumes

$$\boldsymbol{v}_i^{\alpha+1} = \boldsymbol{v}_i^\alpha + \frac{\boldsymbol{a}_i^\alpha + \boldsymbol{a}_i^{\alpha+1}}{2}\Delta t \tag{16.5}$$

for the velocity of atom i. Dropping the higher-order terms in (16.3) yields

$$\boldsymbol{q}_i(t^\alpha - \Delta t) = \boldsymbol{q}^{\alpha-1} \approx \boldsymbol{q}_i^\alpha - \dot{\boldsymbol{q}}_i^\alpha \Delta t + \frac{1}{2}\ddot{\boldsymbol{q}}_i^\alpha \Delta t^2 = \boldsymbol{q}_i^\alpha - \boldsymbol{v}_i^\alpha \Delta t + \frac{1}{2}\boldsymbol{a}_i^\alpha \Delta t^2, \tag{16.6}$$

which we may re-arrange, inserting (16.5), to obtain

$$\begin{aligned}
\boldsymbol{q}_i^\alpha &= \boldsymbol{q}_i^{\alpha-1} + \boldsymbol{v}_i^\alpha \Delta t - \frac{1}{2}\boldsymbol{a}_i^\alpha \Delta t^2 \\
&= \boldsymbol{q}_i^{\alpha-1} + \boldsymbol{v}_i^{\alpha-1}\Delta t + \frac{\boldsymbol{a}_i^{\alpha-1} + \boldsymbol{a}_i^\alpha}{2}\Delta t^2 - \frac{1}{2}\boldsymbol{a}_i^\alpha \Delta t^2 \\
&= \boldsymbol{q}_i^{\alpha-1} + \boldsymbol{v}_i^{\alpha-1}\Delta t + \frac{\Delta t^2}{2}\boldsymbol{a}_i^{\alpha-1}.
\end{aligned} \tag{16.7}$$

This allows us to compute the new atomic position $\boldsymbol{q}_i^\alpha$ of atom i as a function of its position $\boldsymbol{q}_i^{\alpha-1}$, velocity $\boldsymbol{v}_i^{\alpha-1}$, and acceleration $\boldsymbol{a}_i^{\alpha-1}$ at the previous time step (the latter being linked to the net force on atom i at the previous time step). For practical reasons, one typically stores $\boldsymbol{q}_i(t)$ and $\boldsymbol{v}_i(t) = \boldsymbol{p}_i(t)/m$ for each atom for the next time step.

The general algorithm for **zero-temperature MD** is as follows (for each atom $i = 1, \dots, N$):

(i) start with $t^0 = 0$ and known $\boldsymbol{q}_i^0$ and $\boldsymbol{v}_i^0 = \boldsymbol{p}_i^0/m$
(ii) $\boldsymbol{f}_i^0 = -\partial\mathcal{V}/\partial\boldsymbol{q}_i(\boldsymbol{q}^0)$
(iii) $\boldsymbol{a}_i^0 = \boldsymbol{f}_i^0/m_i$
(iv) while $t \le t_{\text{end}}$:
$\boldsymbol{q}_i^{\alpha+1} = \boldsymbol{q}_i^\alpha + \boldsymbol{v}_i^\alpha \Delta t + \boldsymbol{f}_i^\alpha \Delta t^2/2m_i$
$\boldsymbol{f}_i^{\alpha+1} = -\partial\mathcal{V}/\partial\boldsymbol{q}_i(\boldsymbol{q}^{\alpha+1})$
$\boldsymbol{a}_i^{\alpha+1} = \boldsymbol{f}_i^{\alpha+1}/m_i$
$\boldsymbol{v}_i^{\alpha+1} = \boldsymbol{v}_i^\alpha + \Delta t(\boldsymbol{a}_i^\alpha + \boldsymbol{a}_i^{\alpha+1})/2$

$$t^\alpha = t^\alpha + \Delta t$$

Note that the force calculation $\boldsymbol{f}_i^\alpha = -\partial\mathcal{V}/\partial\boldsymbol{q}_i(\boldsymbol{q}^\alpha)$ in general involves neighborhood updates; i.e., for each atom i whose force is to be computed, one finds, stores and frequently updates the neighboring atoms within the **Verlet radius** r_v, which is typically chosen larger than the cut-off radius r_cut to reduce the frequency of neighborhood updates (see Figure 16.1). If the Verlet radius is chosen as the cut-off radius, then neighborhood searches are required at each time step. If the Verlet radius is considerably larger than the cut-off radius, then neighborhood updates can be performed only every nth time step (using numerical experiments to determine n for a particular scenario). Atoms within the Verlet radius are summarized in the **Verlet list** of each atom.

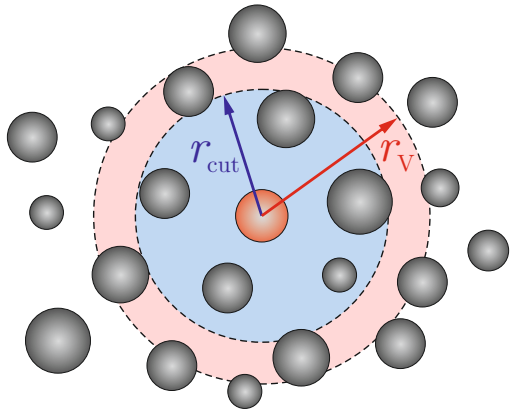


Figure 16.1: The Verlet radius r_V is chosen larger than the cut-off radius r_cut of the interatomic potential to necessitate less frequent neighborhood updates.

Without proof, we mention that the velocity-Verlet scheme is (approximately) *energy conserving*; i.e., it conserves the total Hamiltonian in an approximate sense (more specifically, it conserves a quantity known as *shadow Hamiltonian*, which converges to the exact Hamiltonian as $\Delta t \to 0$; the error of the Hamiltonian scales as $O(\Delta t^2)$). This is an important feature, which makes the velocity-Verlet scheme directly applicable to the canonical (NVE) ensemble. For example, for the simple harmonic oscillator discussed in Section 15.1.1, the shadow Hamiltonian can be derived exactly as

$$\mathcal{H}^*(p,q) = \mathcal{H}(p,q) - \frac{k^2\Delta t^2}{8m}q^2, \tag{16.8}$$

which obviously agrees with the exact Hamiltonian $\mathcal{H}$ in the limit $\Delta t \to 0$.

16.3 Molecular Dynamics at Finite Temperature

By contrast to the above zero-temperature treatment, finite-temperature simulations usually require monitoring and/or controlling of the system temperature T. Temperature is not an explicit variable in the MD scheme, yet one can compute the **instantaneous temperature** of the system at a given time step t^α in a postprocessing step. Specifically, the definition of temperature in (15.48) yields

$$T(t^\alpha) = \frac{2}{3Nk_B}\sum_{i=1}^{N}\frac{m_i}{2}\left\|\delta\boldsymbol{v}_i^\alpha\right\|^2, \tag{16.9}$$

where $\delta\boldsymbol{v}_i^\alpha = \boldsymbol{v}_i(t^\alpha) - \boldsymbol{P}(t^\alpha)/m_i$ is the particle velocity relative to the center of mass of the system, as discussed in Section 15.2. Note that the temperature $T(t)$ fluctuates, and it makes more sense to compute the long-term average $\overline{T}$ as an approximation of the phase average.

Irrespective of the ensemble being used, (16.9) can be exploited to set the **initial temperature** T_{ini} of an atomic ensemble as follows. First, one initializes all particles with a random velocity and, in a subsequent step, ensures that the average momentum $\boldsymbol{P} = \sum_i m_i \boldsymbol{v}_i$ vanishes. This is accomplished, e.g., by adjusting all particle velocities as

$$\boldsymbol{v}_i^0 \leftarrow \boldsymbol{v}_i^0 - \frac{\boldsymbol{P}^0}{M} \qquad \text{with} \qquad M = \sum_{i=1}^N m_i. \tag{16.10}$$

Next, using (16.9) we compute the instantaneous temperature T^0 and rescale all atomic velocities according to

$$\boldsymbol{v}_i^0 \leftarrow \sqrt{\frac{T_{\text{ini}}}{T^0}} \boldsymbol{v}_i^0, \tag{16.11}$$

such that we automatically ensure that the temperature of the system is T_{ini}, since

$$\frac{2}{3Nk_B} \sum_{i=1}^N \frac{m_i}{2} \left\| \delta \boldsymbol{v}_i^0 \right\|^2 = \frac{2}{3Nk_B} \sum_{i=1}^N \frac{m_i}{2} \frac{T_{\text{ini}}}{T^0} \left\| \boldsymbol{v}_i^0 \right\|^2 = \frac{T_{\text{ini}}}{T^0} \frac{2}{3Nk_B} \sum_{i=1}^N \frac{m_i}{2} \left\| \boldsymbol{v}_i^0 \right\|^2 = T_{\text{ini}}. \tag{16.12}$$

Maintaining a **constant temperature** during a simulation, such as when using the microcanonical (NVT) ensemble, is a separate challenge, which is commonly met by using special time-integration schemes that modify atomic velocities at each time step (as discussed in Section 16.4).

As a step in this direction, we point out that fixing the total kinetic energy can be accomplished, e.g., *by Gauss' principle of least constraint*. To this end, notice that the equations of motion can be restated as minimizers of

$$\mathcal{C} = \sum_{i=1}^N \frac{m_i}{2} \left(\ddot{\boldsymbol{q}}_i - \frac{\boldsymbol{f}_i}{m_i} \right)^2 \qquad \text{because} \qquad \boldsymbol{0} = \frac{\partial}{\partial \ddot{\boldsymbol{q}}} \mathcal{C} = m_i \left(\ddot{\boldsymbol{q}}_i - \frac{\boldsymbol{f}_i}{m_i} \right) = m_i \ddot{\boldsymbol{q}}_i - \boldsymbol{f}_i. \tag{16.13}$$

This allows us to add constraints via Lagrange multipliers. For example, to maintain a constant temperature T, we introduce the constraint

$$c = \sum_{i=1}^N \frac{m_i}{2} \left\| \dot{\boldsymbol{q}}_i \right\|^2 - \frac{3}{2} N k_B T = 0 \qquad \text{with} \qquad T = \text{const.}, \tag{16.14}$$

which can be differentiated to obtain the restated constraint

$$\frac{\partial c}{\partial t} = \sum_{i=1}^N m_i \ddot{\boldsymbol{q}}_i \cdot \dot{\boldsymbol{q}}_i = 0. \tag{16.15}$$

This motivates the definition

$$\mathcal{C}^* = \mathcal{C} - \lambda \sum_{i=1}^N m_i \ddot{\boldsymbol{q}}_i \cdot \dot{\boldsymbol{q}}_i \tag{16.16}$$

with a Lagrange multiplier λ to enforce the constant temperature. The resulting modified equations of motion are

$$\frac{\partial}{\partial \ddot{\boldsymbol{q}}} \mathcal{C}^* = m_i \ddot{\boldsymbol{q}}_i - \boldsymbol{f}_i - \lambda m_i \dot{\boldsymbol{q}}_i = \boldsymbol{0} \qquad \Leftrightarrow \qquad m_i \ddot{\boldsymbol{q}}_i = \boldsymbol{f}_i + \lambda m_i \dot{\boldsymbol{q}}_i. \tag{16.17}$$

The Lagrange multiplier λ can be obtained from inserting $\ddot{\boldsymbol{q}}_i = \boldsymbol{f}_i/m_i + \lambda\dot{\boldsymbol{q}}_i$ into the constraint equation (16.15), which yields

$$\sum_{i=1}^{N} m_i \left(\boldsymbol{f}_i/m_i + \lambda\dot{\boldsymbol{q}}_i\right)\cdot\dot{\boldsymbol{q}}_i = 0 \qquad \Leftrightarrow \qquad \lambda = -\frac{\sum_{i=1}^{N}\boldsymbol{f}_i\cdot\dot{\boldsymbol{q}}_i}{\sum_{a=1}^{N} m_a\left\|\dot{\boldsymbol{q}}_a\right\|^2}. \tag{16.18}$$

Overall, we have thus arrived at the modified equations of motion

$$m_i\ddot{\boldsymbol{q}}_i = \boldsymbol{f}_i - \frac{\sum_{b=1}^{N}\boldsymbol{f}_b\cdot\dot{\boldsymbol{q}}_b}{\sum_{a=1}^{N} m_a\left\|\dot{\boldsymbol{q}}_a\right\|^2} m_i\dot{\boldsymbol{q}}_i. \tag{16.19}$$

The extra term on the right-hand side appears as a nonlinear damping force, which aims to maintain the constant kinetic energy. Note that this formulation fixes the average *total* kinetic energy, but it does not allow any control of the atomic fluctuations with respect to the mean motion (as required in the temperature calculation). This is commonly achieved by so-called *thermostats*.

16.4 Thermostats

As discussed above, MD at constant temperature (using the NVT-ensemble) requires fixing the mean kinetic energy with respect to the mean motion of the system. As an example, the **Langevin thermostat** modifies the equations of motion into

$$m_i\ddot{\boldsymbol{q}}_i = \boldsymbol{f}_i - \gamma_i m_i\dot{\boldsymbol{q}}_i + \boldsymbol{g}_i(t), \tag{16.20}$$

where γ_i is a damping/viscosity constant and $\boldsymbol{g}_i$ denotes a random, time-varying force that may be interpreted as noise. The damping term may be interpreted as **Brownian motion** (at a significantly higher collision rate than that of atomic interactions). Loosely speaking, the added terms in the Langevin thermostat introduce random forcing to maintain the constant temperature, while the damping term removes access kinetic energy.

For the noise term to not introduce any bias, $\boldsymbol{g}_i(t)$ must be *uncorrelated* in both space and time; i.e., it must satisfy

$$\langle\boldsymbol{g}_i\rangle = 0 \qquad \Leftrightarrow \qquad \langle g_{i,k}(t)g_{j,l}(t')\rangle = 2\gamma_i m_i k_B T\delta_{kl}\delta_{ij}\delta(t-t'), \tag{16.21}$$

where we denote by $g_{i,k}$ the kth component of vector $\boldsymbol{g}_i$. This can be achieved in practice, while enforcing a temperature T, if the force is a randomly chosen sample from a normal distribution with variance

$$\sigma_i^2 = 2\gamma_i m_i k_B T/\Delta t \tag{16.22}$$

across the atomistic system at each time step of size Δt.

A similar scheme is the **Andersen thermostat** which also introduces random forces as in the Langevin approach but does so not at every time step.

The probably most common thermostat is the **Nosé-Hoover thermostat**, which starts with a fictitious Hamiltonian

$$\mathcal{H}^* = \frac{P^2}{2M} + \sum_{i=1}^{N}\frac{\|\tilde{\boldsymbol{p}}_i\|^2}{2m_iQ^2} + V(\boldsymbol{q}) + 3Nk_BT\ln Q, \tag{16.23}$$

which may be interpreted as follows. $\mathcal{H}^*$ includes the 1-D motion of a *fictitious atom* having mass M, position Q and momentum P; and it uses rescaled atomic momenta $\tilde{\boldsymbol{p}}_i = Q\,m_i\dot{\boldsymbol{q}}_i$.

Applying Hamilton's equations of motion to all particles $i = 1, \dots, N$ as well as to the fictitious atom results in

$$\begin{aligned} \frac{\mathrm{d}\boldsymbol{q}_i}{\mathrm{d}t} &= \frac{\partial \mathcal{H}^*}{\partial \tilde{\boldsymbol{p}}_i} = \frac{\tilde{\boldsymbol{p}}_i}{m_i Q^2}, \qquad \frac{\mathrm{d}\tilde{\boldsymbol{p}}_i}{\mathrm{d}t} = -\frac{\partial \mathcal{H}^*}{\partial \boldsymbol{q}_i} = \boldsymbol{f}_i, \\ \frac{\mathrm{d}Q}{\mathrm{d}t} &= \frac{\partial \mathcal{H}^*}{\partial P} = \frac{P}{M}, \qquad \frac{\mathrm{d}P}{\mathrm{d}t} = -\frac{\partial \mathcal{H}^*}{\partial Q} = \frac{1}{Q}\left(\sum_{i=1}^{N} \frac{\|\tilde{\boldsymbol{p}}_i\|^2}{m_i Q^2} + 3Nk_BT\right). \end{aligned} \tag{16.24}$$

If we introduce a rescaled time t' defined by $\mathrm{d}t = Q\ \mathrm{d}t'$ (and abusing the dot for derivatives with respect to t'), the above equations reduce to

$$\frac{\mathrm{d}\boldsymbol{q}_i}{\mathrm{d}t} = \frac{\dot{\boldsymbol{q}}_i}{Q} = \frac{\tilde{\boldsymbol{p}}_i}{m_i Q^2}, \qquad \frac{\mathrm{d}\tilde{\boldsymbol{p}}_i}{\mathrm{d}t} = \frac{\dot{\tilde{\boldsymbol{p}}}_i}{Q} = \boldsymbol{f}_i, \qquad \frac{\mathrm{d}Q}{\mathrm{d}t} = \frac{\dot{Q}}{Q} = \frac{P}{M} \tag{16.25}$$

and

$$Q\frac{\mathrm{d}P}{\mathrm{d}t} = \dot{P} = \sum_{i=1}^{N} \frac{\|\tilde{\boldsymbol{p}}_i\|^2}{m_i Q^2} + 3Nk_BT = \sum_{i=1}^{N} \frac{\|\boldsymbol{p}_i\|^2}{m_i} + 3Nk_BT. \tag{16.26}$$

Multiplying the first equation in (16.25) by $m_i Q^2$ and differentiating with respect to t' leads to (inserting the second equation from (16.25))

$$m_i Q \dot{\boldsymbol{q}}_i = \tilde{\boldsymbol{p}}_i \qquad \Rightarrow \qquad m_i \dot{Q} \dot{\boldsymbol{q}}_i + m_i Q \ddot{\boldsymbol{q}}_i = \dot{\tilde{\boldsymbol{p}}}_i = Q\, \boldsymbol{f}_i. \tag{16.27}$$

Next, division by Q and solving for the acceleration term (while inserting the third equation from (16.25)) produces

$$m_i \ddot{\boldsymbol{q}}_i = \boldsymbol{f}_i - m_i \frac{\dot{Q}}{Q} \dot{\boldsymbol{q}}_i = \boldsymbol{f}_i - m_i \frac{P}{M} \dot{\boldsymbol{q}}_i. \tag{16.28}$$

Finally, by exploiting (16.26) we turn (16.28) into at a system of evolution equations (defining $\dot{\gamma} = \dot{P}/M$ with $M = \text{const.}$):

$$\begin{aligned} m_i \ddot{\boldsymbol{q}}_i &= \boldsymbol{f}_i - \gamma m_i \dot{\boldsymbol{q}}_i, \\ \dot{\gamma} &= \frac{1}{M}\left(\sum_{i=1}^{N} \frac{\|\boldsymbol{p}_i\|^2}{m_i} - 3Nk_BT\right). \end{aligned} \tag{16.29}$$

Thus, the Nosé-Hoover thermostat is similar in nature to the Langevin/Andersen thermostats. Here, the viscous drag coefficient γ is not constant but evolves with time and presents an intrinsic force that aims to drive the system to the enforced temperature T. The only parameter to be adjusted is the fictitious mass M, which my be interpreted as the 'inertia' of the viscous coefficient.

We note that the same MD algorithm introduced in Section 16.2 for zero-temperature conditions can be adopted for finite temperature simulations, if we modify the atomic forces according to the chosen thermostat. For example, for the *Langevin thermostat* we must re-define the atomic forces as

$$\boldsymbol{f}_i^*(t^\alpha) = \boldsymbol{f}_i(t^\alpha) - \gamma m_i \dot{\boldsymbol{q}}_i(t^\alpha) + \boldsymbol{g}_i(t^\alpha). \tag{16.30}$$

Analogously, the Nosé-Hoover thermostat re-defines the atomic forces as

$$\boldsymbol{f}_i^*(t^\alpha) = \boldsymbol{f}_i(t^\alpha) - \gamma(t^\alpha) \dot{\boldsymbol{q}}_i(t^\alpha) \tag{16.31}$$

and additionally updates the viscous coefficient γ at each time step in the average-acceleration fashion according to

$$\gamma^{\alpha+1} = \gamma^{\alpha} + \frac{\Delta t}{2}\left(\dot{\gamma}^{\alpha} + \dot{\gamma}^{\alpha+1}\right), \tag{16.32}$$

by application of the second equation in (16.29).

Finally, without going into details, one can analogously impose an average pressure in the system (using the NpT-ensemble) through a so-called **barostat**. In general, pressure is defined as $p = \frac{1}{3}\operatorname{tr}\boldsymbol{\sigma}$, so that the virial stress tensor, defined in (15.53), formulates the constant-pressure constraint as

$$p = \frac{1}{3}\operatorname{tr}\boldsymbol{\sigma} = -\frac{1}{3|\Omega|}\sum_{i=1}^{N}\left\langle \frac{\boldsymbol{p}_i \cdot \boldsymbol{p}_i}{2m_i} + \boldsymbol{f}_i \cdot \boldsymbol{q}_i \right\rangle = \text{const.} \tag{16.33}$$

The most frequent barostats are those named after **Parrinello-Rahman** and **Berendsen** (see, e.g., reference [89] for further information).

17 Multiscale Modeling Techniques

17.1 Why do we need multiscale techniques?

Atomistic modeling is a prime example for why multiscale techniques are required. Molecular dynamics and statics, discussed in previous sections, are popular and convenient simulation methodologies with state-of-the-art software available open-source. However, computer architectures limit the length and time scales accessible by those techniques. Length scales are limited by computer *memory* to the total number of atoms that can be simulated, whereas time scales are limited by the computer *speed* (clock rate) due to the finite time step size Δt used in atomistic simulations.

Let us perform a simple back-of-the-envelope calculation: consider an atomistic ensemble of N atoms in 3D, which requires at least $6N$ degrees of freedom to handle (viz. three components for each atomic position and momentum vector). If each degree of freedom is stored as a float, then it requires 4 bytes of memory, which leads to a total of $24N$ bytes of memory for the entire atomistic ensemble. Now, consider, e.g., an fcc crystal which has four atoms in one of its cubic unit cells with equilibrium side length a_0 (recall that for copper, e.g., $a_0 \approx 3.6\mathring{A} = 3.6 \cdot 10^{-10}$m). Hence, a nano-cube of side length L contains $N \approx 4(L/a_0)^3$ atoms. Altogether, we conclude that the nano-cube with N atoms requires a total amount of memory M of

$$M = 24 \cdot 4 \left(\frac{L}{a_0} \right)^3 \text{ bytes.} \tag{17.1}$$

Note that this back-of-the-envelope calculation only considers the memory requirement for the degrees of freedom and entirely neglects all other memory needed for auxiliary variables and functionality required to run a simulation. For a nano-cube made of copper, Figure 17.1 plots M vs. side length L on a log-log-scale.

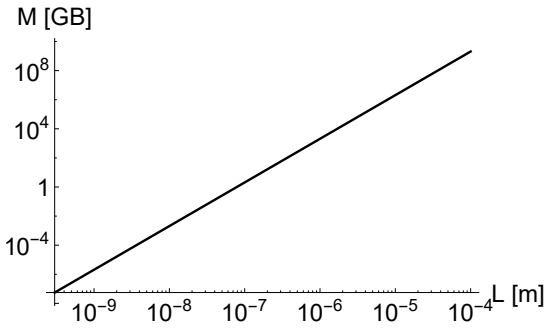


Figure 17.1: Memory requirement M (in GB) as a function of the side length L of a single-crstalline cube of pure Cu.

To run a simulation efficiently, the above memory D must fit into the RAM of a machine. Therefore, it becomes apparent that simulating length scales above the nanometer scale requires high-performance computing infrastructure with serious RAM sizes. For comparison, ETH's *Euler VII* cluster has 292 compute nodes, each equipped with 256 GB of RAM, resulting in a total of ca. 75,000 GB. Figure 17.1 implies that this allows for nano-cubes of at most 3.3μm side length. (Note that this considerably overestimates the size, because only the storage of degrees of freedom is considered. It also assumes that 100% of the cluster can be used by a single user to run nothing but this MD simulations.)

Summarized in the following are benchmark simulations from the past decade, which indicate the trends in feasible numbers of simulated particles over the years, limited by the size of

computer memory (not limited to atomistics):

- 2003: 10^8 atoms (LANL Q) *(Kadau et al., 2004)*
- 2007: 10^{10} atoms (BlueGene/L) *(Kadau et al., 2007)*
- 2012: 10^{12} atoms (BlueGene/Q) *(Habib et al., 2012)*
- 2015: 10^{13} particles (Titan) *(Koumoutsakos, 2015)*
- 2018-21: 10^{15} atoms *(prediction of exascale computing initiatives)*

Although computer power has significantly increased in the recent past and the so-called *exascale* initatives have launched tremendous new high-performance machines, current capabilities are still behind 10^{15} particles (and *Moore's law* tells us that we cannot expect this trend of increasing numbers to continue much longer). Therefore, current computing architecture severely limits the length scales accessible by atomistic techniques.

Even though the number of particles that can be simulated has increased over time due to increasingly large memory on supercomputers, the same cannot be said about time scales, because communication times and CPU speeds will not advance at the same pace as memory. Moreover, atomistic simulations rely on explicit dynamic time integration, whose characteristic time step size Δt is on the order of one femtosecond (1fs $= 10^{-15}$s). Considering typical clock rates and speeds of current super-computers, this limits achievable times in atomistic simulations to pico- or nanoseconds (at most microseconds). LAMMPS provides a list of (by now outdated) benchmarks. Their '*billion-atom Lennard-Jones benchmark*' ran 100 time steps of an MD simulation with one billion atoms on various state-of-the-art machines. The best performance (at that time) was achieved on a Cray XT5, which took $1.49 \cdot 10^{-6}$ seconds per atom and per time step. Considering realistic values of the number of atoms N in a simulation (see the discussion around Figure 17.1), this quickly becomes prohibitive. For example, a system of one billion atoms takes about 1490 s of computing time per time step. Simulating a process over a total time $T = 1$s requires $T/\Delta t = 10^{15}$ time steps, which on the above machine amounts to millions of years of computing time. Therefore, even though high-performance computing infrastructure is continuously increasing (and the number of particles to be simulated has increased significantly over time), time scales will remain a bottleneck.

17.2 Types of Multiscale Techniques

The state of the art in material modeling offers highly-accurate methods for each individual scale, from density functional theory (DFT) and molecular dynamics (MD) at the lower scales all the way up to continuum theories and associated computational tools for the macroscale and structural applications (see Figure 17.2). Unfortunately, a wide gap exists due to a lack of models applicable at the intermediate scales (sometimes referred to as *mesoscales*). Here, the continuum hypothesis fails because the discreteness of the atomic crystal becomes apparent, e.g., through the emergence of size effects. At the same time, atomistic techniques tend to incur prohibitively high computational expenses when reaching the scales of hundreds of nanometers or microns (as discussed in Section 17.1). Mastering this gap between atomistics and the continuum is the key to understanding a long list of diverse open problems. These include the mechanical response of nanoporous or nanostructured (i.e., nanocrystalline or nanotwinned) metals, the effective mechanical properties of nanometer- and micron-sized structures, devices and engineered (meta-)materials, further the underlying mechanisms leading to inelasticity and material failure, or heat and mass transfer in nanoscale materials systems. Overall, there is

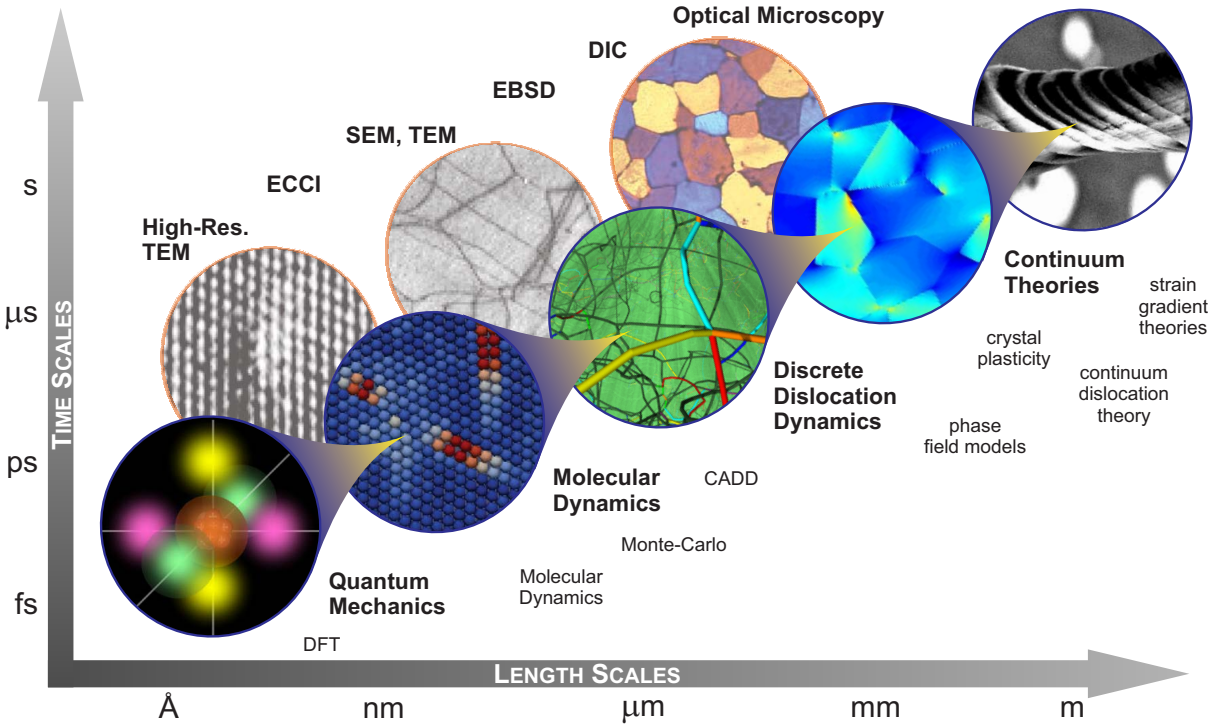


Figure 17.2: Bridging across scales in crystalline solids: from the electronic structure all the way up to the macroscale (including some of the prominent modeling and experimental techniques).

urgent need for techniques that bridge across length and time scales in order to accurately describe, to thoroughly understand, and to reliably predict the mechanics and physics of solids.

Bridging across length scales in material modeling usually relies on one of following two strategies, examples of which are illustrated in Figure 17.3:

- **hierarchical scale-bridging**, also known as **vertical scale-bridging**, assumes a separation of scales, so that a hierarchy of scales can be established. Information from the lower scale is passed on to the larger scale through the definition of *effective measures*. Computational homogenization and FE^2, discussed before, are prime examples: RVE-information is passed to the macroscale, based on the spatial separation between micro- and macroscales. Difficulties arise when a clear separation of scales is not warranted (e.g., when macroscale features are on the same scale as microscale specifics) or when localization occurs on the microscale (e.g., when a crack emerges in the RVE, which then grows into a crack on the macroscale).

- **concurrent scale-bridging**, also known as **horizontal scale-bridging**, does not assume a separation of scales but combines multiple simulation techniques concurrently on the same level. The idea behind concurrent techniques is to limit lower-scale resolution to where it is indeed needed, while using an efficient large-scale description in the remaining simulation domain. For example, when running atomistic simulations, interesting features occur in the vicinity of defects, interfaces and surfaces. This motivates to use atomistic techniques near those features while using a more efficient description in those regions deforming more homogeneously far away from defects, interfaces and surfaces. Therefore, one may use atomistics in those interesting regions of the simulation domain while using more efficient and less accurate descriptions in the remaining domain. Ideally, such domain decomposition should be performed adaptively, so that no knowledge is required a-priori about where the expensive lower-scale resolution will be required during a simulation.

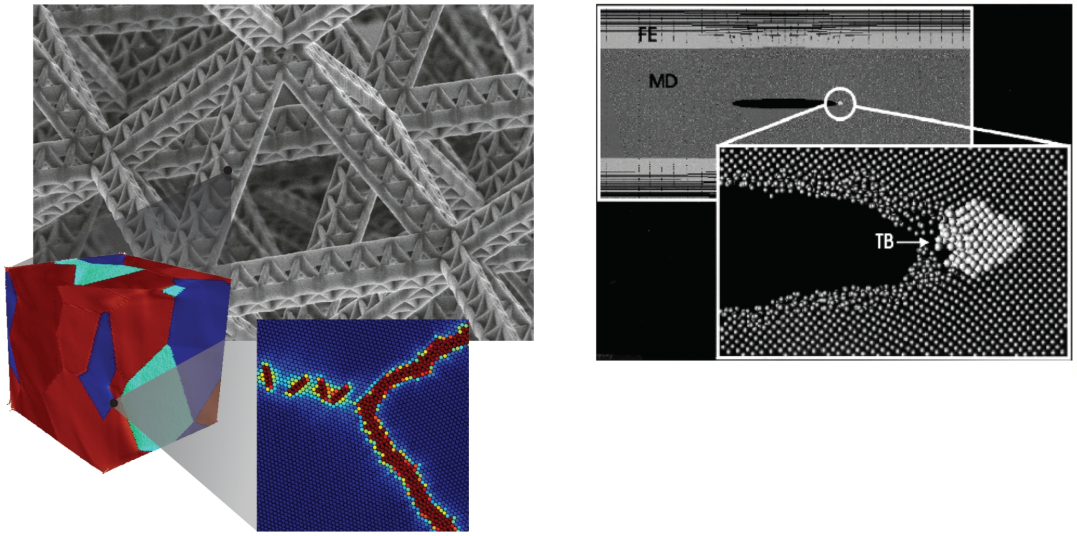


Figure 17.3: Examples of hierarchical *(left)* and concurrent *(right)* scale-bridging techniques. Hierarchical scale-bridging assumes a separation of scales, whereas concurrent scale-bridging combines different techniques at the same level. Shown here is the MAAD technique [2], which combines tight-binding (TB), atomistics (MD), and finite element (FE) regions to efficiently and accurately model a large crack.

Challenges generally arise from identifying compatible constitutive descriptions on both sides of the interface separating the two domains, and from establishing an interface that is compatible with both descriptions and allows information to flow from one domain to the other. Prominent examples comprise Coupled Atomistic/Discrete-Dislocation (CADD) models [16, 67, 79] and AtoDis [14], furthermore the Bridging Domain Method (BDM) [11] and Bridging Scale Decomposition [56], as well as Macroscopic, Atomistic, Ab Initio Dynamics (MAAD) [2, 15], which couples several scales. In such methods, passing information appropriately across interfaces is key. To pick out one example, CADD (see references [16, 67, 79]) is an elegant methodology which embeds a small MD domain into a larger region treated by a Discrete Dislocation (DD) description. Here, the passing of lattice defects across the interface between the two domains is a major challenge. In order to circumvent such difficulties, **coarse-graining** techniques apply the same lower-scale constitutive description to the entire model but scale up in space and/or in time (see Figure 17.4).

Classical examples include Coarse-Grained MD (CGMD) [74, 75] as well as the quasicontinuum (QC) method [87], which we will discuss in more detail. Spatial coarse-graining reduces the number of degrees of freedom by introducing geometric constraints, thereby making the lower-scale accuracy efficiently available for larger-scale simulations (see references [40, 41, 83, 87] for examples). Coarse-graining strategies offer a number of advantages over domain-coupling methods: (i) the model is solely based on the lower-scale constitutive laws and hence comes with superior accuracy (in contrast to coupling methods, there is no need for a separate and oftentimes empirical constitutive law in the continuum region); (ii) the transition from fully-resolved to coarse-grained regions can be seamless (no approximate hand-shake region is required between different domains in general); (iii) depending on the chosen formulation, model adaptation techniques can efficiently reduce computational complexity by tying full resolution to those regions where it is indeed required (such as in the vicinity of lattice defects or cracks and voids).

Up-scaling of atomistic simulations in the time domain is a separate challenge – going from femtoseconds to significantly larger time scales of interest. Although this challenge has been

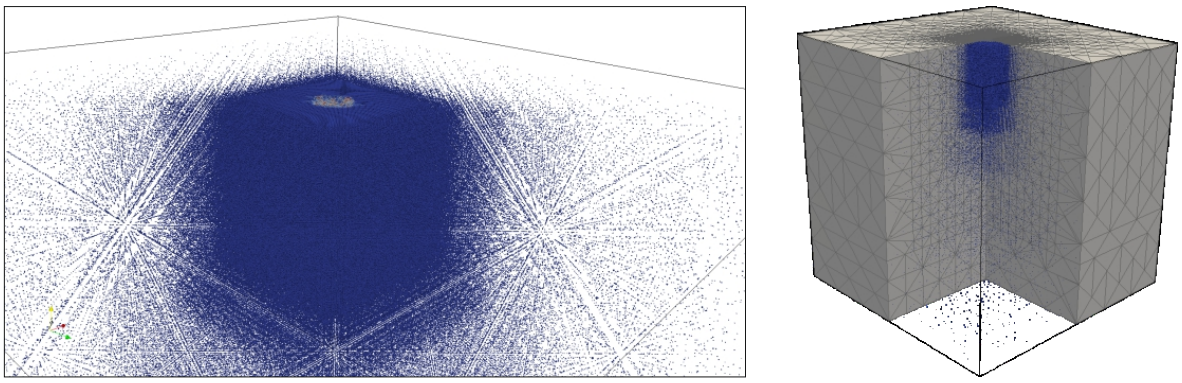


Figure 17.4: Coarse-graining is achieved, e.g., by the quasicontinuum (QC) method: full atomistic resolution is retained in regions of interest (such as, e.g., the section underneath an indenter in a nano-indentation simulation), while regions far away from those regions of interest (here, in the bulk of the crystal) are simulated efficiently by a coarse-grained description based on the same constitutive description. Coarse-graining here is performed gradually, i.e., the resolution decreases from a full atomistic representation in the mid-section to a significantly coarser description with increasing distance (adapted from reference [92]). Details will be discussed in the following.

addressed by a number of techniques, it is still a key open challenge in multiscale modeling for many problems of interest. While upscaling in time and in space are equally important, the primary focus in the following will be on *spatial upscaling*.

18 The Quasicontinuum Method

So far, we have discussed the fundamentals of multiscale modeling at continuum levels and the basics of atomistic modeling. In Section 17 we further argued that atomistic techniques cannot be applied to large system sizes, as needed for many practical applications, which motivates the use of multiscale methods. Here, we turn to a particular *coarse-graining* technique, known as the **quasicontinuum** method, which links atomistic modeling to continuum-level modeling in a *concurrent* fashion (i.e., not by passing information between scales but by integrating atomistic and continuum modeling domains on the scale of a simulation).

The *quasicontinuum method*[9] was introduced to bridge from atomistics to the continuum by applying finite element interpolation schemes to atoms in a crystal lattice [64, 70, 87] – similar in spirit to the Cauchy-Born discussed in Section 14. Unlike there, we here aim to use Cauchy-Born-type approximations only in those regions, where atomistic resolution is not required, while we retain full atomistic resolution where needed. This is achieved by three integral components:

- *geometric constraints* (which interpolate lattice site positions from the positions of a reduced set of representative atoms),
- *summation rules* or *sampling rules* (which avoid the computation of thermodynamic quantities from the full atomistic ensemble in the spirit of quadrature rules in the finite element method),
- *model adaptation* schemes (which localize atomistic resolution and thereby efficiently minimize the total number of degrees of freedom).

To date, numerous QC flavors have been developed which mainly differ in the choice of how the aforementioned three aspects are realized. Despite their many differences, all QC-based techniques share as a common basis the interpolation of atomic positions from a set of representative atoms as the primary coarse-graining tool.

The family of QC methods has been applied to a wide range of problems of scientific and technological interest such as studies of the interactions of lattice defects in fcc and bcc metals. Typical problems include nanoindentation [53, 57, 86], interactions of lattice defects [33, 70] or of defects with nanosized cracks [37, 73, 84] and nanovoids [5, 103], or with individual interfaces [54, 102]. Beyond metals, the developed modeling techniques can be extended, e.g., to ferroelectric materials and ionic crystals via electrostatic interactions [19] and multi-lattice approaches [1, 21], further to non-equilibrium thermodynamics [51, 61, 74, 76], also to structural mechanics [10, 71], and in principle to any material system with crystalline order as long as a suitable position-dependent interatomic potential is available.

In this section, we review the theoretical basis of the family of QC methods, followed by the description of particular QC schemes and available techniques with applications and extensions. Note that focus on important QC advances and contributions and cannot give an exhaustive account of the literature in this field. For further information, you are encouraged to visit the *qcmethod.org* website (see reference [90]).

18.1 Representative atoms and the quasicontinuum approximation

The rich family of QC methods is united by the underlying concepts of (i) significantly reducing the number of degrees of freedom by suitable interpolation schemes, and (ii) approximating the

[9] This section dedicated to the quasicontinuum method is adapted in parts from reference [47].

thermodynamic quantities of interest by summation or sampling rules. Let us briefly outline these concepts.

In classical mechanics, an atomistic ensemble containing N atoms is uniquely described by how their positions $\boldsymbol{q} = \{\boldsymbol{q}_1, \dots, \boldsymbol{q}_N\}$ and momenta $\boldsymbol{p} = \{\boldsymbol{p}_1, \dots, \boldsymbol{p}_N\}$ with $\boldsymbol{p}_i = m_i \dot{\boldsymbol{q}}_i$ evolve with time t (as before, m_i represents the mass of atom i, and dots denote material time derivatives). The ensemble's total Hamiltonian $\mathcal{H}$ is given by

$$\mathcal{H}(\boldsymbol{q}, \boldsymbol{p}) = \sum_{i=1}^{N} \frac{|\boldsymbol{p}_i|^2}{2\, m_i} + V(\boldsymbol{q}) \tag{18.1}$$

with V denoting a suitable atomic interaction potential. The time evolution of the system is governed by Hamilton's equations for all atoms $i = 1, \dots, N$:

$$m_i \ddot{\boldsymbol{q}}_i = \boldsymbol{f}_i(\boldsymbol{q}) = -\frac{\partial V}{\partial \boldsymbol{q}_i}(\boldsymbol{q}), \tag{18.2}$$

where $\boldsymbol{f}_i(\boldsymbol{q})$ represents the total (net) force acting on atom i. We note that V can, in principle, depend on both positions $\boldsymbol{q}$ and momenta $\boldsymbol{p}$ (the latter may be necessary, e.g., when using the quasiharmonic and other approximations for quasistatic finite-temperature formulations without time discretization, as mentioned later). For conciseness, we here restrict ourselves to position-dependent potentials. The QC method was designed for *crystalline* solids, in which the ground state atomic positions coincide with sites of a regular lattice described by a set of Bravais vectors that span the discrete periodic array of lattice sites. In most materials (including metals, ceramics and organic materials), the interaction potential usually allows for an additive decomposition, i.e.

$$V(\boldsymbol{q}) = \sum_{i=1}^{N} E_i(\boldsymbol{q}) \tag{18.3}$$

with E_i being the energy of atom i.

Due to limitations of computational resources, it is generally not feasible to apply the above framework to systems that are sufficiently large to simulate long-range elastic effects, even for short-range interatomic potentials. However, except in the vicinity of flaws and lattice defects such as cracks and dislocations, respectively, the local environment of each atom in a crystal lattice is almost identical up to rigid body motion. Therefore, the **QC approximation** replaces the full atomistic ensemble by a reduced set of $N_h \ll N$ **representative atoms** (often referred to as **repatoms** for short). This process is shown schematically in Figure 18.1.

Let us denote the repatom positions by $\boldsymbol{x}(t) = \{\boldsymbol{x}_1(t), \dots, \boldsymbol{x}_{N_h}(t)\}$. The approximate current positions $\boldsymbol{q}_i^h$ and momenta $\boldsymbol{p}_i^h$ of all atoms $i = 1, \dots, N$ are now obtained from interpolation, i.e. we have

$$\boxed{\boldsymbol{q}_i \approx \boldsymbol{q}_i^h = \sum_{a=1}^{N_h} N_a(\boldsymbol{X}_i)\, \boldsymbol{x}_a} \tag{18.4}$$

and consequently

$$\boldsymbol{p}_i \approx \boldsymbol{p}_i^h = m_i\, \dot{\boldsymbol{q}}_i^h = m_i \sum_{a=1}^{N_h} N_a(\boldsymbol{X}_i)\, \dot{\boldsymbol{x}}_a. \tag{18.5}$$

$N_a(\boldsymbol{X}_i)$ is the shape function of repatom a evaluated at the position $\boldsymbol{X}_i$ of lattice site i in the undeformed (reference) configuration. As an essential feature, one usually requires this

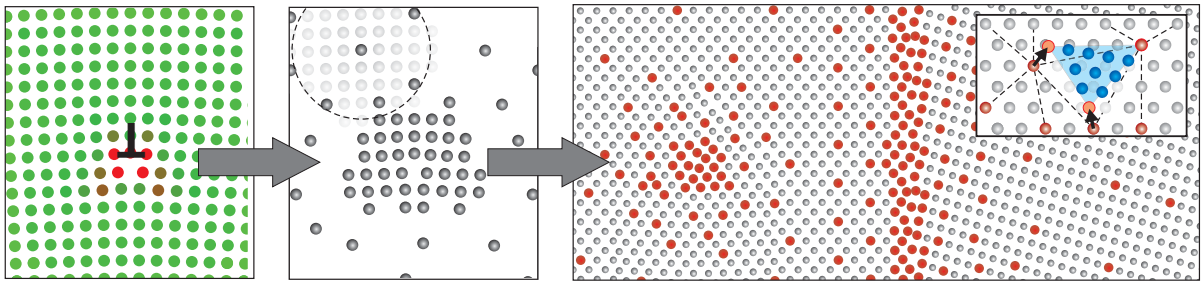


Figure 18.1: Illustration of the QC methodology: identification of atoms of interest (e.g., high-centrosymmetry lattice sites near a dislocation core), reduction to the set of repatoms (coarsening the full atomic lattice by choosing a small set of representative atomic sites), and the interpolation of atomic positions (blue in the small inset) from repatom positions (red) with an example repatom distribution around a dislocation and a grain boundary in two dimensions.

coarse-graining scheme to locally recover the exact atomic ensemble when all atoms are turned into repatoms. Therefore, shape functions should be chosen to satisfy the Kronecker property $N_a(\boldsymbol{X}_b) = \delta_{ab}$ for all $1 \leq a, b \leq N_h$ with δ_{ij} denoting Kronecker's delta ($\delta_{ij} = 1$ if $i = j$ and 0 otherwise). This is automatically satisfied when choosing shape functions from the finite element method (FEM).

By borrowing FEM concepts, the above geometric constraints within the original QC method [45, 77, 87] made use of an affine interpolation on a Delaunay-triangulated mesh. More recent versions have explored higher-order polynomial shape functions [52, 69] as well as meshless interpolations using smoothed-particle approaches [98, 101] or local maximum-entropy shape functions [46].

The introduction of the geometric constraints in (18.4) has reduced the total number of independent degrees of freedom from $d \times N$ in d dimensions to $d \times N_h$. Therefore, the approximate Hamiltonian $\mathcal{H}^h$ of the coarse-grained system, now involving approximate atomic positions $\boldsymbol{q}^h = \{\boldsymbol{q}_1^h, \dots, \boldsymbol{q}_N^h\}$ and momenta $\boldsymbol{p}^h = \{\boldsymbol{p}_1^h, \dots, \boldsymbol{p}_N^h\}$, only depends on the positions and momenta of the repatoms through (18.4):

$$\mathcal{H}^h(\boldsymbol{x}, \dot{\boldsymbol{x}}) = \sum_{i=1}^{N} \frac{|\boldsymbol{p}_i^h|^2}{2m_i} + V(\boldsymbol{q}^h). \tag{18.6}$$

Instead of solving for the positions and momenta of all N lattice sites, the QC approximation allows us to update only the positions and momenta of the N_h repatoms, which requires to compute forces on repatoms. These are obtained from the potential energy by differentiation, which yields the net force on repatom k:

$$\boldsymbol{F}_k(\boldsymbol{x}) = -\frac{\partial V(\boldsymbol{q}^h)}{\partial \boldsymbol{x}_k} = \sum_{j=1}^{N} \boldsymbol{f}_j^h(\boldsymbol{q}^h)\, N_k(\boldsymbol{X}_j) \tag{18.7}$$

with

$$\boldsymbol{f}_j^h(\boldsymbol{q}^h) = -\frac{\partial V(\boldsymbol{q}^h)}{\partial \boldsymbol{q}_j^h} = -\sum_{i=1}^{N} \frac{\partial E_i(\boldsymbol{q}^h)}{\partial \boldsymbol{q}_j^h}, \tag{18.8}$$

the total force acting on atom j. It is important to note that, in an infinite Bravais lattice (i.e., in a defect-free single-crystal in the absence of external loading or free surfaces), the forces on all atoms vanish, i.e. we have $\boldsymbol{f}_i^h(\boldsymbol{q}^h) = \boldsymbol{0}$, so that the net forces on all repatoms vanish as well ($\boldsymbol{F}_k = \boldsymbol{0}$).

18.2 Summation rules and spurious force artifacts

Summation rules have become an integral ingredient of all QC methods (even though they are sometimes not referred to as such). Although the above introduction of repatoms has reduced the total number of degrees of freedom significantly from $d \times N$ to $d \times N^h$, the calculation of repatom forces still requires computing the forces between all N atoms and their on average N_b neighbors located within each atom's radius of interaction (see the sums over N in (18.7) and (18.8)). These $\mathcal{O}(N \times N_b)$ operations become computationally not feasible for realistically-sized systems. We note that, in principle, summations in (18.7) can be reduced to the support of the shape functions (i.e., to regions in which $N_k(\boldsymbol{X}_j) \neq 0$), but even this is prohibitively expensive in regions of dilute repatom concentrations. Therefore, **summation rules** or **sampling rules** have been introduced to approximate the thermodynamic quantities of interest of the full atomistic ensemble by those of a small set of carefully-chosen lattice sites (in the following referred to as **sampling atoms**), comparable to quadrature rules commonly found in the finite element method. Note that sampling atoms do not need to coincide with repatoms (in fact, it is wise to not choose the same sets of atoms for both, as we will discuss).

QC summation rules have either approximated the energy of the system (so-called *energy-based QC*) [3, 26] or the forces experienced by the repatoms (*force-based QC*) [45]. Within each of those two categories, multiple versions have been proposed. Here, we will first describe what is known as *fully-nonlocal* energy-based and force-based QC formulations and then extract the original *local/nonlocal* QC method as a special case of energy-based QC.

18.2.1 Energy-based QC

In energy-based nonlocal QC formulations, the total Hamiltonian is approximated by a weighted sum over a carefully selected set of **sampling atoms**. To this end, we replace the sum over index i in (18.3) or (18.7) by a weighted sum over N_s carefully-chosen sampling atoms, see e.g. [3, 26, 27]. Consequently, the total potential energy is approximated by

$$\boxed{V(\boldsymbol{q}^h) = \sum_{i=1}^{N} E_i(\boldsymbol{q}^h) \quad \approx \quad \widetilde{V}(\boldsymbol{q}^h) = \sum_{\alpha=1}^{N_s} w_\alpha \, E_\alpha(\boldsymbol{q}^h)} \tag{18.9}$$

where w_α is the weight of sampling atom α. Physically, w_α denotes the number of lattice sites represented by sampling atom α. As an example, we may consider large elements, in which an affine QC interpolation leads to approximately uniform neighborhood deformation within elements. In this case one may pick one sampling atom per element, which is representative for all the lattice sites contained within the element (similar to the Cauchy-Born rule discussed before).

The force experienced by repatom k is obtained by differentiation in analogy to (18.7):

$$\widetilde{\boldsymbol{F}}_k(\boldsymbol{x}) = -\frac{\partial \widetilde{V}(\boldsymbol{q}^h)}{\partial \boldsymbol{x}_k} = -\sum_{\alpha=1}^{N_s} w_\alpha \sum_{j=1}^{N} \frac{\partial E_\alpha(\boldsymbol{q}^h)}{\partial \boldsymbol{q}_j^h} N_k(\boldsymbol{X}_j). \tag{18.10}$$

Now, the repatom force calculation has been reduced to $\mathcal{O}(N_s \times N_b)$ operations (accounting for the fact that the sum over j above effectively only involves N_b neighbors within the potential's cutoff radius). The selection of sampling atoms and the calculation of sampling atom weights aims for a compromise between maximum accuracy (ideally $N_s = N$ and $w_\alpha = 1$) and maximum efficiency (requiring $N_s \ll N$ and $w_\alpha \gg 1$).

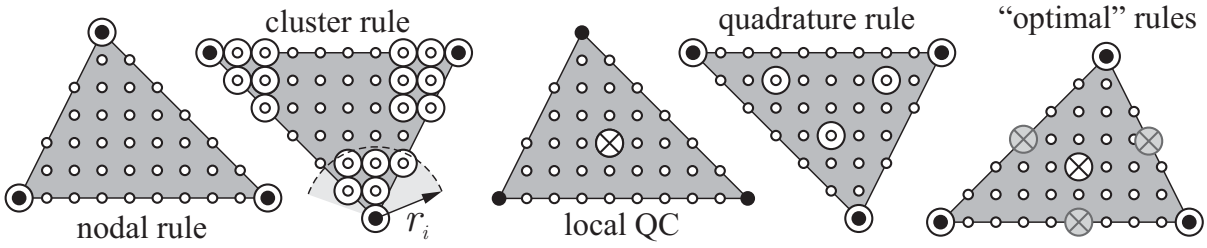


Figure 18.2: Illustration of popular summation rules: small open circles denote lattice sites, solid circles represent repatoms, and large open circles are sampling atoms (crossed circles denote sampling atoms whose neighborhoods are approximated to undergo Cauchy-Born-type affine deformations). Optimal summation rules [3] are shown of first and second order (the gray sampling atoms only exist in the second-order rule).

In order to seamlessly bridge from atomistics to the continuum without differentiating between atomistic and coarse-grained regions, the above summation rules can be interpreted as a fully **nonlocal QC** approximation, which treats the entire simulation domain in the same fashion (i.e., even the fully resolved atomistic region is represented by sampling atoms and repatoms: each atom becomes a repatom and a sampling atom with a weight of $w_\alpha = 1$). Summation rules now differ by the choice of (i) *sampling atom locations* and (ii) *sampling atom weights*. Successful examples of summation rules introduced previously include *node-based cluster summation* [45] with sampling atoms located in clusters around repatoms, and *quadrature-type summation* [31, 99] with sampling atoms chosen nearest to Gaussian quadrature points with or without the repatoms included as sampling atoms. Furthermore, *element-based summation rules* have been introduced in the nonlocal QC context [31, 99] and have demonstrated superior accuracy over traditional cluster-based summation schemes [40]. Recently, a *central summation rule* was proposed in [8] as an ad-hoc compromise between local and quadrature summation rules, which is similar in spirit and contained as a special case within the optimal summation rules recently introduced in [3].

Fig. 18.2 illustrates some of the most popular summation rules, many of which were introduced in an ad-hoc manner to mitigate particular QC deficiencies (e.g., cluster rules were introduced to remove zero-energy modes in force-based QC [45]) or by borrowing schemes from related models (e.g., finite element quadrature rules [31]).

18.2.2 Force artifacts

Summation rules on the energy level not only reduce the computational complexity but they also give rise to **force artifacts**, see e.g. [3, 26] for reviews within the energy-based context. To understand their origin, we rewrite (18.10) without neighborhood truncations as

$$\widetilde{\boldsymbol{F}}_k(\boldsymbol{x}) = -\sum_{\alpha=1}^{N_s} w_\alpha \sum_{j=1}^{N} \frac{\partial E_\alpha(\boldsymbol{q}^h)}{\partial \boldsymbol{q}_j^h} N_k(\boldsymbol{X}_j) = \sum_{j=1}^{N} \left(-\sum_{\alpha=1}^{N_s} w_\alpha \frac{\partial E_\alpha(\boldsymbol{q}^h)}{\partial \boldsymbol{q}_j^h} \right) N_k(\boldsymbol{X}_j). \tag{18.11}$$

Comparing with (18.7), we see that the term in parentheses in (18.11) is not equal to force $\boldsymbol{f}_j^h$ and hence does not vanish in general, even if $\boldsymbol{f}_j^h = \boldsymbol{0}$ for all atoms (unless we choose all atoms to be sampling atoms, i.e. unless $N_s = N$ and $w_\alpha = 1$ in the atomistic limit). As a consequence, the QC representation with energy-based summation rules shows what is known as **residual forces** in the undeformed ground state, which are non-physical.

In uniform QC meshes (i.e. uniform repatom spacings, uniform sampling atom distribution, and a regular mesh), these residual forces indeed disappear due to symmetry; specifically, the sum

in parentheses in (18.11) cancels pairwise when carrying out the full sum. Residual forces hence appear only in spatially non-uniform meshes (see also [40] for a discussion). In general, one should differentiate between residual and spurious forces [3, 26]. Here, we use the term *residual force* to denote force artifacts arising in the *undeformed* configuration. In an infinite crystal, these can easily be identified by computing all forces in the undeformed ground state. However, force errors vary nonlinearly with repatom positions. Therefore, we speak of **spurious forces** when referring to force artifacts in the deformed configuration. As shown e.g. in [3], correcting for residual forces by subtracting those as dead loads does not correct for spurious forces and can lead to even larger errors in simulations. Indeed, in many scenarios such a dead-load correction is not easily possible since the undeformed ground state may contain, e.g., free surfaces. In this case, the computed forces in the undeformed ground state contain both force artifacts and physical forces. Thus, without running a fully-atomistic calculation for comparison, it is impossible to differentiate between residual force artifacts and physical forces arising, e.g., due to free surfaces (see Section 13.5). Even though residual and spurious forces are conservative in the energy-based scheme, they are non-physical and can drive the coarse-grained system into incorrect equilibrium states. Special correction schemes have been devised, which effectively reduce or remove the impact of force artifacts, which includes so-called **ghost force correction** techniques [77]. However, such schemes require a special algorithmic treatment at the interface between full resolution and coarse-grained regions, which may create computational difficulties and, most importantly, requires the notion of an interface.

Based on mathematical analyses, further energy-based schemes have been developed, which blend atomistics and coarse-grained descriptions by a special formulation of the potential energy or the repatom forces. Here, an interfacial domain between full resolution and coarse regions is introduced, in which the thermodynamic quantities of interest are taken as weighted averages of the exact atomistic and the approximate coarse-grained ones [55, 59]. Such a formulation offers clear advantages, including the avoidance of force artifacts. It also requires the definition of an interface region, which may be disadvantageous in a fully-nonlocal QC formulation with automatic model adaption since no notion of such interfaces exists strictly [3].

18.2.3 Force-based QC

In order to avoid force artifacts, *force-based summation rules* were introduced in [45]. These do not approximate the Hamiltonian but the repatom forces explicitly. Ergo, the summation rule is applied directly to the repatom forces in (18.7), viz.

$$\boxed{\boldsymbol{F}_k(\boldsymbol{x}) = \sum_{i=1}^{N} \boldsymbol{f}_i^h(\boldsymbol{q}^h)\, N_k(\boldsymbol{X}_i) \quad \approx \quad \widetilde{\boldsymbol{F}}_k(\boldsymbol{x}) = \sum_{\alpha=1}^{N_s} w_\alpha\, \boldsymbol{f}_\alpha^h(\boldsymbol{q}^h)\, N_k(\boldsymbol{X}_\alpha)} \tag{18.12}$$

As a consequence, this force-based formulation does not produce residual forces because $\boldsymbol{f}_\alpha^h(\boldsymbol{q}^h) = \boldsymbol{0}$ in an undeformed infinite crystal. The same holds true for an infinite crystal that is affinely deformed: for reasons of symmetry the repatom forces cancel, and spurious force artifacts are effectively suppressed.

Yet, this approximation gives rise to a new problem: the force-based method is *non-conservative*; hence, there is no potential from which repatom forces derive. This has a number of drawbacks [26, 40, 65]. First, in quasistatic problems, the non-conservative framework may lead to slow numerical convergence, cause numerical instability, or converge to non-physical equilibrium states [22–24, 58, 68]. Second, for dynamic or finite-temperature scenarios, a QC approximation using force-based summation rules cannot be used to simulate systems in the microcanonical ensemble (where the system's energy is to be conserved; other ensembles may, of course, be used).

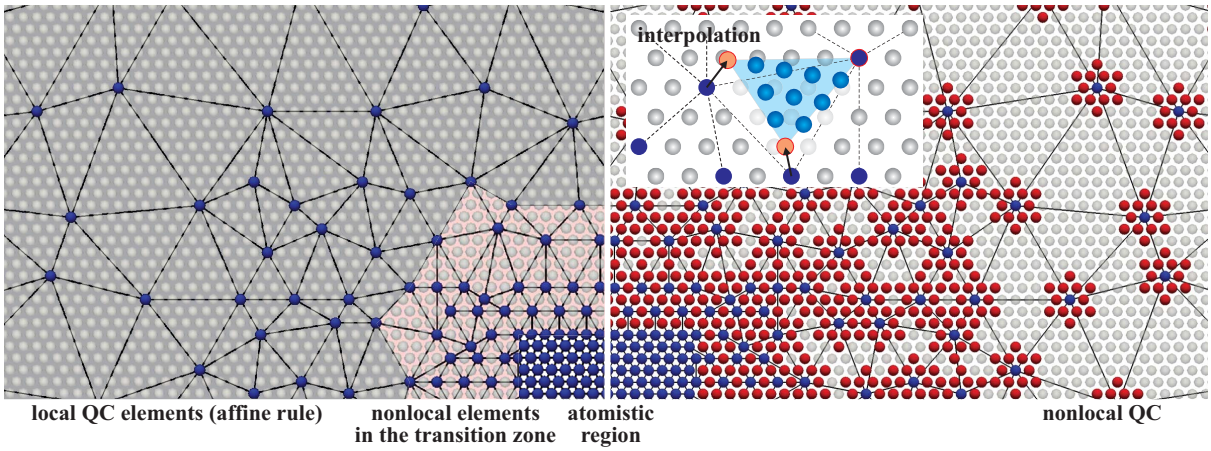


Figure 18.3: Schematic view of the local/nonlocal (left) and fully-nonlocal (right) QC formulations. The local scheme on the left treats differently atoms in the fully-resolved region (where (16.2) is solved) from those in the coarse-grained region (where FEM with the constitutive model obtained from the Cauchy-Born rule in Section 14 is used). The shown nonlocal scheme on the right uses a node-based cluster summation rule and blends smoothly from atomistic to coarsened regions. The inset illustrates the affine interpolation of atomic positions from repatoms, which is the same in both formulations.

Third, repatom masses, required for dynamic simulations, are not uniquely defined, because there is no effective kinetic energy potential when using force-based summation rules (see also Section 18.2.5 below). By contrast, energy-based summation rules lead to conservative forces and to strictly-symmetric stiffness matrices with only the six admissible zero eigenvalues [26].

18.2.4 Local/nonlocal QC

The original QC method [87] uses affine interpolation within elements and may be regarded as a special energy-based QC scheme, in which atomistic and coarse-grained regions are spatially separated. In the atomistic region and its immediate vicinity, the full atomistic description is applied and one solves the discrete (static) equilibrium equations (16.2) for each atom and approximations thereof near the interface (**nonlocal QC**). In the coarse-grained region, a particular element-based summation rule is employed (**local QC**): for each element, the energy is approximated by assuming an affine deformation of all atomic neighborhoods within the element, so that one Cauchy Born-type sampling atom per element is sufficient and the total potential energy is the weighted sum over all such element energies. This concept has been applied successfully to a myriad of examples, primarily in two dimensions (or in 2.5D by assuming periodicity along the third dimension). Since the element-based summation rule in the nonlocal coarse region produces no force artifacts [3], force artifacts here only appear at the interface between atomistic and coarse-grained regions and have traditionally been called **ghost forces** [64, 77]. For a comprehensive comparison of this technique with other atomistic-to-continuum-coupling techniques see reference [65]. Fig. 18.3 schematically illustrates the concepts of local/nonlocal and fully-nonlocal QC.

18.2.5 Repatom masses

Besides repatom forces, energy-based summation rules provide consistent repatom masses, as mentioned above. When approximated by the QC interpolation scheme, the total kinetic energy

of an atomic ensemble becomes

$$\begin{aligned}\frac{1}{2}\sum_{i=1}^{N} m_i\,(\dot{\boldsymbol{q}}_i^h)^2 &= \frac{1}{2}\sum_{i=1}^{N} m_i \left|\sum_{a=1}^{N_h} N_a(X_i)\,\dot{\boldsymbol{x}}_a\right|^2 \\ &= \frac{1}{2}\sum_{a=1}^{N_h}\sum_{c=1}^{N_h} \dot{\boldsymbol{x}}_a \cdot \left(\underbrace{\sum_{i=1}^{N} m_i\,N_a(X_i)\,N_c(X_i)}_{=M_{ac}^h}\right)\dot{\boldsymbol{x}}_c,\end{aligned} \tag{18.13}$$

where the term in parentheses may be interpreted as the components M_{ac}^h of a consistent mass matrix $\boldsymbol{M}^h$. In avoidance of solving a global system during time stepping, one often resorts to *mass lumping* in order to diagonalize $\boldsymbol{M}^h$.

Applying the energy-based summation rule to the total Hamiltonian (and thus to the kinetic energy) yields a consistent approximation of the kinetic energy with mass matrix components

$$M_{ac}^h = \sum_{i=1}^{N} m_i\,N_a(X_i)\,N_c(X_i) \quad\approx\quad \widetilde{M}_{ac}^h = \sum_{b=1}^{N_s} w_b\,m_b\,N_a(X_b)\,N_c(X_b). \tag{18.14}$$

In case of mass matrix lumping, the equations of motion of all repatoms now become

$$\widetilde{m}_k^h\,\ddot{\boldsymbol{x}}_k = \widetilde{\boldsymbol{F}}_k(\boldsymbol{x}), \tag{18.15}$$

which can be solved by (explicit or implicit) finite difference schemes. The calculation of repatom masses is of importance not only for dynamic QC simulations, but it is also relevant in quasistatic finite-temperature QC formulations.

18.2.6 Example: Embedded Atom Method

The QC approximation outlined above is sufficiently general to model the performance of most types of crystalline solids, as long as their potential energy can be represented as a function of atomic positions. Let us exemplify the general concept by a frequently-used family of interatomic potentials for metals (which has also been used most frequently in conjunction with the QC method). As discussed in Section 13.3, the Embedded Atom Method (EAM) [18] defines the interatomic potential energy of a collection of N atoms by

$$E_i(\boldsymbol{q}) = \frac{1}{2}\sum_{j\neq i}\Phi(r_{ij}) + \mathcal{F}(\rho_i), \qquad \rho_i = \sum_{j\neq i} f(r_{ij}). \tag{18.16}$$

From (18.2), the exact force $\boldsymbol{f}_k$ acting on atom k is obtained by differentiation, viz.

$$\boldsymbol{f}_k(\boldsymbol{q}) = -\sum_{i=1}^{N}\frac{\partial E_i(\boldsymbol{q})}{\partial \boldsymbol{q}_k} = -\sum_{j\neq k}\left[\Phi'(r_{kj}) + \{\mathcal{F}'(\rho_k) + \mathcal{F}'(\rho_j)\}\,f'(r_{kj})\right]\frac{\boldsymbol{r}_{kj}}{r_{kj}}. \tag{18.17}$$

Since most (non-ionic) potentials are short-range (in particular those for metals), one can efficiently truncate the above summations to include only neighboring atoms within the radius of interaction ($\mathcal{C}(i)$ denotes the set of neighbors within the sphere of interaction of atom i). Introducing the QC approximation along with an energy-based summation rule now specifies the approximate repatom force as

$$\widetilde{\boldsymbol{F}}_k(\boldsymbol{x}) = -\sum_{\alpha=1}^{N_s} w_\alpha \sum_{j\in\mathcal{C}(\alpha)}\left[\frac{1}{2}\,\Phi'(r_{\alpha j}^h) + \mathcal{F}'(\rho_\alpha^h)\,f'(r_{\alpha j}^h)\right]\frac{\boldsymbol{r}_{\alpha j}^h}{r_{\alpha j}^h}\left[N_k(\boldsymbol{X}_\alpha) - N_k(\boldsymbol{X}_j)\right], \tag{18.18}$$

which completes the coarse-grained description required for QC simulations. Note that, as discussed above, repatom forces (18.18) are prone to produce residual and spurious force artifacts in non-uniform QC meshes.

18.3 Adaptive refinement

One of the general strengths of the QC method is its suitability for adaptive model refinement. This is particularly useful when there is no a-priori knowledge about where within a simulation domain atomistic resolution will be required during the course of a simulation. For example, when studying defect mechanisms near a crack tip or around a pre-existing lattice defect, it may be sufficient to restrict full atomistic resolution to regions in the immediate vicinity of those microstructural features and efficiently coarsen away from those (recall Fig. 17.3, or the indenter problem in Fig. 17.4). However, when investigating many-defect interactions or when the exact crack path during ductile failure is unknown, it is beneficial to make use of *automatic model adaptation*. By locally refining elements down to the full atomistic limit, full resolution can effectively be tied to evolving defects in an efficient manner. In the fully-nonlocal QC method [3], this transition is truly seamless as there is no conceptual differentiation between atomistic and coarse domains. In the local/nonlocal QC method as well as in blended QC formulations the transition is equally possible but may require a special algorithmic treatment.

Mesh refinement requires two central ingredients: a *criterion for refinement* and a geometric *refinement algorithm*. The former can be realized on the element level, e.g., by checking invariants of the deformation gradient within each element and comparing those to a threshold for refinement like in classical finite element techniques (see, e.g., refernce [45]). Alternatively, one can introduce criteria based on repatoms or sampling atoms, e.g., by checking the energy or the centrosymmetry of repatoms or sampling atoms and comparing those to refinement thresholds (cf. Section 13.4). The most common geometric tool for refinement is element bisection, although a large variety of tools exist and these are generally tied to the specifically-chosen interpolation scheme.

Unlike in the finite element method, mesh adaptation within the QC method is challenging [92], since element vertices, at least in all common implementations, are restricted to sites of the underlying atomistic lattice (this ensures that full atomistics is recovered exactly upon ultimate refinement). While mesh refinement is technically challenging but conceptually straight-forward, mesh coarsening presents an even bigger challenge (how to turn small elements back into large elements, if fine resolution is no longer required but the atomic positions have changed significantly). Meshless formulations [46, 98, 101] appear promising but have not reached sufficient maturity for large-scale simulations.

18.4 Features and Extensions

The QC methods described above have been applied to a variety of scenarios which go beyond the traditional problem of quasistatic equilibration at zero-temperature. In particular, QC extensions have been proposed in order to describe phase transformations and deformation twinning, finite temperature, atomic-level heat and mass transfer, or ionic interactions such as in ferroelectrics. In addition, the basic concept of the QC method has been applied to coarse-grain discrete mechanical systems beyond atomistics. Here, we give a brief (non-exhaustive) summary of such special QC features and extensions:

- **Multilattices**: The original (local) QC approximation applies an affine interpolation to all lattice site positions within each element in the coarse domain, which results in

affine neighborhood changes within elements (the same applies to the optimal summation rules within fully-nonlocal QC). By using Cauchy-Born (CB) kinematics (as discussed in Section 9), one can account for relative shifts within the crystalline unit cells such as those occurring during domain switching in ferroelectrics or during solid-state phase transitions and deformation twinning. The CB rule is thus used in coarse-grained regions to relate atomic motion to continuum deformation gradients. In order to avoid failure of the CB kinematics, such a multilattice QC formulation has been augmented with a phonon stability analysis which detects instability and identifies the minimum required periodic cell size for subsequent simulations. This augmented approach has been referred to as *Cascading Cauchy-Born Kinematics* [20, 81] and has been applied, among others, to ferroelectrics [88] and shape-memory alloys [20].

- **Finite temperature and dynamics:** Going from zero to finite temperature within the QC framework is a challenge that has resulted in various approximate descriptions but has not been resolved entirely. In contrast to full atomistics, the physical behavior of continua is generally governed by thermodynamic state variables such as temperature and by thermodynamic quantities such as the free energy. Therefore, many finite-temperature QC frameworks have constructed an *effective, temperature-dependent free energy potential*; examples are refernces [25, 32, 51, 61, 76, 85, 91]. Dynamic simulations of finite-temperature and non-equilibrium processes raise additional questions, especially when heat and mass transfer (i.e., heat conduction and diffusion of species) is involved.

  Within the local/nonlocal QC framework, finite-temperature has often been enforced in different ways in the atomistic and the continuum regions, see e.g. [85]. While in the atomistic region a thermostat can be used to maintain a constant temperature, an effective, temperature-dependent free energy is formulated in the coarse domain, e.g., based on the *quasiharmonic approximation* [85], according to which atoms vibrate harmonically around their equilibrium state (using the quasiharmonic potential approximation; see Section 13.2). For quasistatic equilibrium calculations, the dynamic atomistic ensemble is then embedded into the static continuum description of the coarse-grained model. Alternatively, in dynamic *hotQC* simulations of this type, all repatoms (including those in both regions) evolve dynamically. The quasiharmonic approximation was shown to be well-suited to describe, e.g., the volumetric lattice expansion at moderate temperatures. Drawbacks include the inaccuracy observed at elevated temperature levels.

  An alternative approach is to use techniques originally developed for the acceleration of MD now for the QC framework. One such example, *hyperQC* [44] borrows concepts of the original *hyperdynamics* method [97] and accelerates atomistic calculations by energetically favoring rare events through the modification of the potential energy landscape.

  As a further example, finite-temperature QC has been realized by the aid of *Langevin thermostats* [61] (introduced in Section 16.4 for atomistic ensembles). Here, the QC approximation is applied within the framework of dissipative Lagrangian mechanics with a viscous term that expends the thermal energy introduced by a Langevin thermostat through a random force at the repatom level. In a nutshell, the repatoms can be pictured as an ensemble of nodes suspended in a viscous medium which represents the neglected degrees of freedom. The effect of this medium is approximated by frictional drag on the repatoms as well as random fluctuations from the thermal motion of solvent particles. As a consequence, high-frequency modes (i.e., phonons) not transmitted across mesh interfaces are dampened out by the imposed thermostat in order to sample stable canonical ensemble trajectories. This method is anharmonic and was used to study non-equilibrium, thermally-activated processes. Unfortunately, the reduced phonon spectra in the coarse-grained domains result in an underestimation of thermal properties such as

thermal expansion.

In contrast to the above finite-temperature QC formulations which interpret continuum thermodynamic quantities as atomistic time averages, one may alternatively assume ergodicity and instead take advantage of averages in phase space. This forms the basis of the *maximum-entropy* hotQC formulation [51, 94]. By recourse to mean-field theory and statistical mechanics, this approach is based on the assumption of maximizing the atomistic ensemble's entropy. To this end, repatoms are equipped with additional degrees of freedom that describe the statistics of their vibrational motion and which must be solved for in addition to the repatom positions at a given constant temperature of the system. This approach seamlessly bridges across scales and does not conceptually differentiate between atomistic and continuum domains. This concept has also been extended to account for heat and mass transfer [94].

- **Ionic crystals:** Ionic interactions, such as those arising, e.g., in solid electrolytes or in complex oxide ferroelectric crystals, present an additional challenge for the QC method because atomic interactions cease to be short-range. The QC summation or sampling rules take advantage of short-range interatomic potentials which admit the local evaluation of thermodynamic quantities of interest, ideally at the element level. When long-range interactions gain importance, that concept no longer applies and new effective summation schemes must be developed such as those reported in [62].

- **Beyond atomistics:** The QC approximation is a powerful tool whose application is not necessarily restricted to atomistic lattices. In general, any periodic array of interacting nodes can be coarsened by a continuum description which interpolates the full set of nodal positions from a small set of representative nodes. Therefore, since its original development for coarse-grained atomistics, the QC method has been extended and applied to various other fields. As an example at even lower length scales, coarse-grained formulations of *Density Functional Theory* have been reported [28], which considerably speed up electronic calculations. While the key concepts here appertain to the particular formulation of quantum mechanics based on electron densities, the coarse-graining of the set of atomic nuclei may be achieved by recourse to the QC method. As a further example at larger scales, the QC method can also be applied at the meso- and macroscales in order to coarse-grain periodic structures such as truss or fiber networks. Here, interatomic potentials are replaced by the discrete interactions of truss members or fibers, and the collective response of large networks is again approximated by the selection of representative nodes and suitable interpolation schemes. This methodology was applied to truss lattices [10, 71] and textiles [9]. As a complication, nodal interactions in those examples are not necessarily elastic, so that internal variables must be introduced to account for path and history dependence in the nodal interactions and dissipative potentials can be introduced. Fig. 18.4 shows a schematic of the coarse-graining of truss lattices as well as the simulated failure of a coarse-grained periodic truss structure loaded in three-point bending.

18.5 Applications

It is a difficult task to summarize all applications of the QC method to date; for a thorough overview we refer to the QC references found on the qcmethod.org website (see reference [90]). Without claiming completeness, the following topics have attracted significant interest in the QC community:

- size effects and deformation mechanisms during nanoindentation,

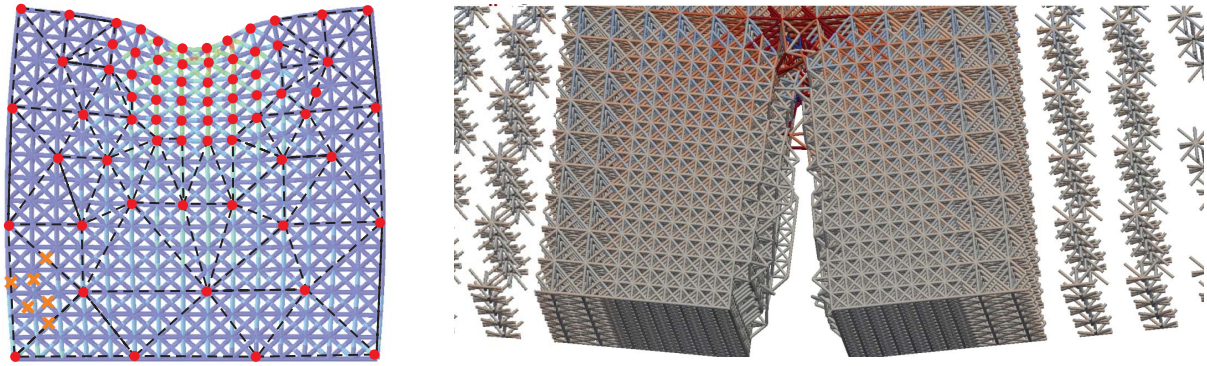


Figure 18.4: Schematic view of a coarse-grained truss network undergoing indentation (left; shown are truss members, nodes, representative nodes, and example sampling truss members within one element) and the simulation result of a truss lattice fracturing during three-point bending (right, color-coded by the axial stress within truss members which are modeled as elastic-plastic).

- nanoscale contact and friction, nanoscale scratching and cutting processes,
- interactions between dislocations and grain boundaries (GBs), GB mechanisms,
- phase transformations and deformation twinning, shape-memory alloys,
- fracture and damage in fcc/bcc metals and the brittle-to-ductile transition,
- void and cavity growth at zero and finite temperature,
- deformation and failure mechanisms in single-, bi-, and polycrystals, in particular the mechanics of nanocrystalline solids,
- crystalline sheets and rods, including graphene and carbon nanotubes,
- ferroelectrics and polarization switching,
- electronic textile, fiber networks and truss structures,
- 2D-materials such as graphene.

In the following, we will highlight a few specific applications of the QC method applied to metals and truss structures. These benchmark examples were simulated with our in-house fully-nonlocal, massively-parallel 3D QC code.

18.5.1 Nanoindentation

Nanoindentation is one of the most classical examples that has been studied by the QC method since its inception due to the relatively simple problem setup and the rich inelastic deformation mechanisms that can be observed (it hence also serves as a prime validation benchmark). Oftentimes, the indenter is conveniently modeled by an external potential [43], which avoids the handling of contact and the presence of two materials.

As an illustrative example of the nonlocal QC method, Fig. 18.5 shows the results of a virtual indentation test into single-crystalline pure copper (modeled by an extended Finnis-Sinclair potential [17]) with a spherical indenter of radius 5 nm up to a maximum indentation depth of 3 nm. For ease of visualization, this example is restricted to two dimensions. Of course, this scenario could easily be simulated by MD, but we deliberately present simple scenarios first for

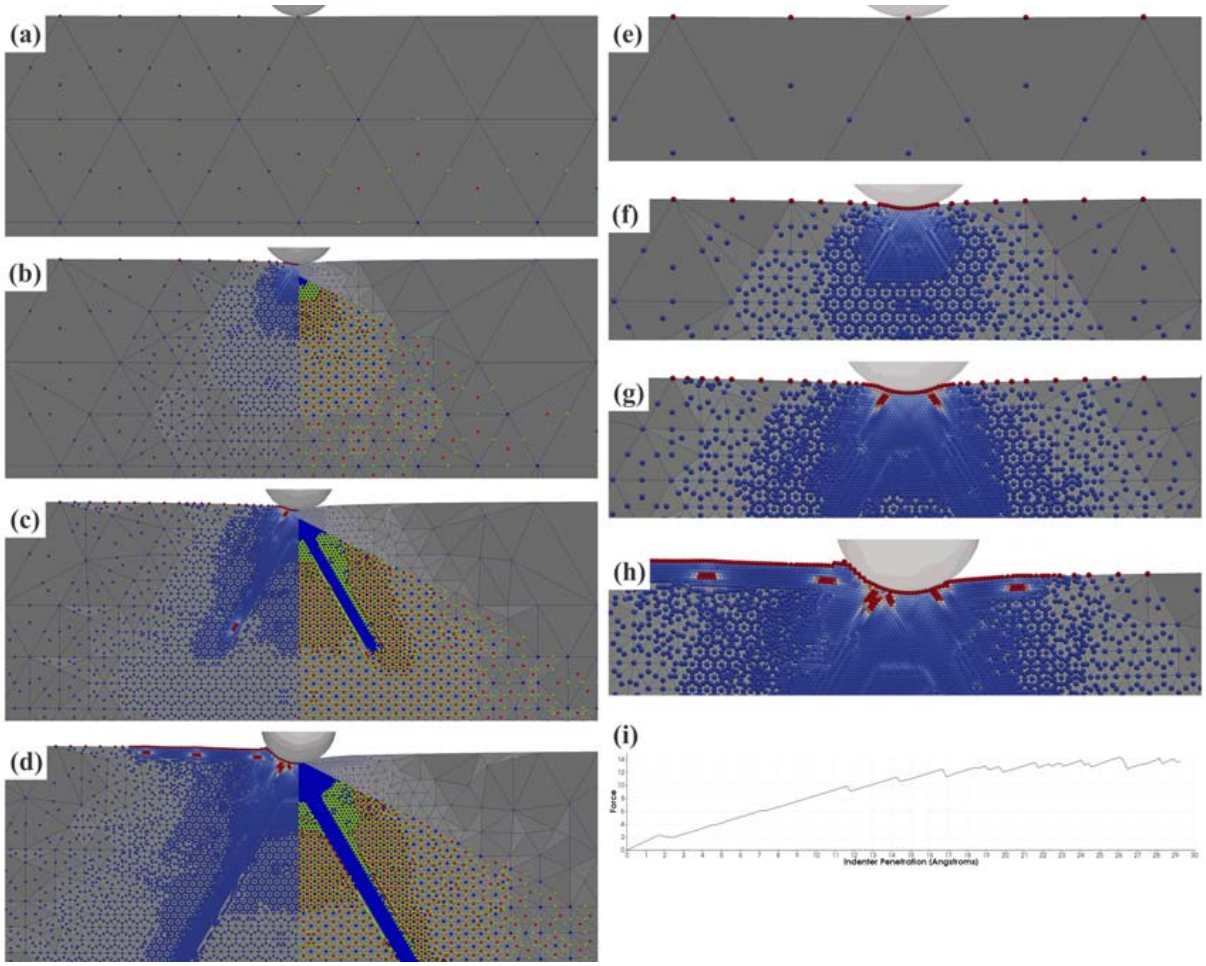


Figure 18.5: Nanoindentation into a Cu single-crystal: (a)-(d) show the distribution of repatoms in blue (bottom left panels) as well as color-coded by centrosymmetry (top left panels), the sampling atoms of the first- and second-order summation rules of [3] (shown in red and green, respectively, in the bottom right panels), and the mesh (top right panels). (e)-(h) are zooms of (a)-(d) illustrating the sampling atom centrosymmetry. (i) shows the curve of indenter force vs. indentation depth (force in eV/A). Images adapted from reference [47].

purposes of detail visualization. The four panels in Figs. 18.5 (a) to (d) illustrate the distribution of repatoms and sampling atoms along with the QC mesh. Automatic mesh adaptation leads to local refinement underneath the indenter and around lattice defects (we use the second invariant $I_2(\boldsymbol{F})$ for the refinement criterion, and longest-edge bisection for refinement). Lattice defects are visualized by plotting the centrosymmetry parameter (Section 13.4) of sampling atoms (which agree with lattice sites in the fully-resolved regions).

Fig. 18.6 shows analogous results obtained from a pyramidal indenter penetrating into the same Cu single-crystal. As before, full atomistic resolution is restricted to those regions where it is indeed required: underneath the indenter as well as in the vicinity of lattice defects. Figs. 18.6 (c) and (d) show the spreading of dislocations into the crystal after emission from the indenter, resulting in full resolution in the wake of the dislocations. The remainder of the simulation domain remains coarse-grained, thus allowing for efficient simulations having significantly fewer degrees of freedom than full atomistic calculations. As before, the distribution of repatoms and sampling atoms is shown along with the QC mesh.

The curves of load vs. indentation depth for both cases of spherical and pyramidal indenters in two dimensions are summarized in Fig. 18.7. As can be expected from experiments, results

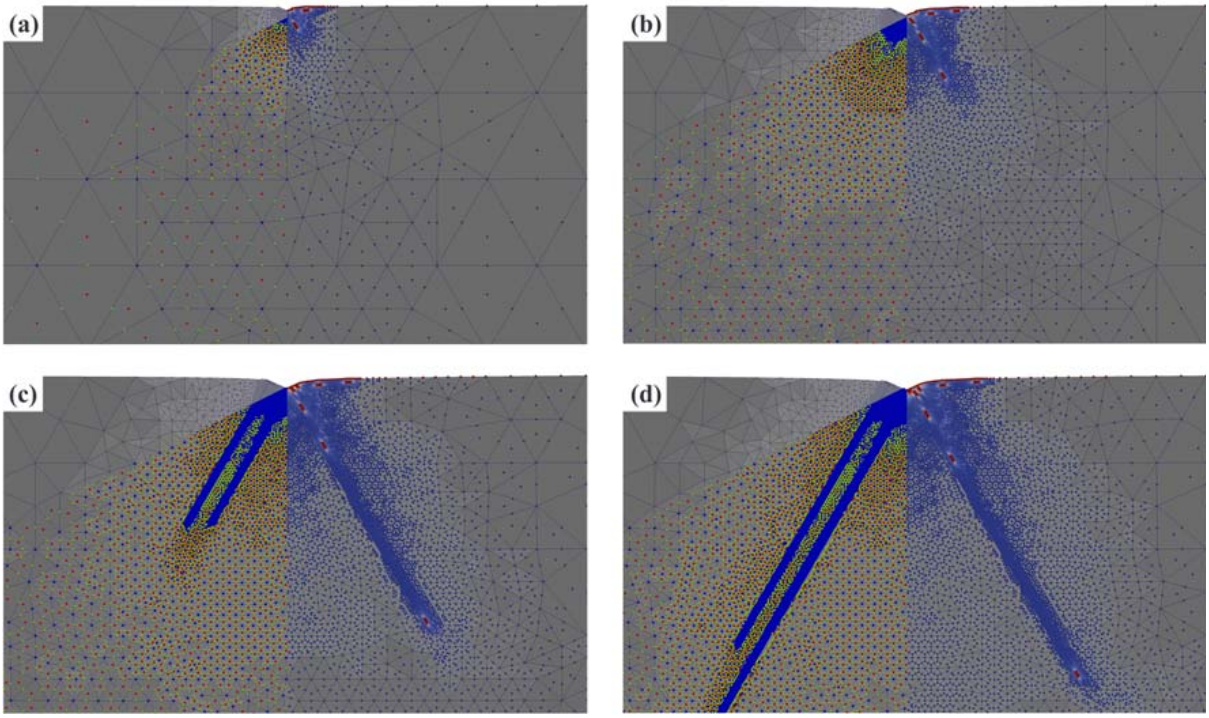


Figure 18.6: Nanoindentation into a Cu single-crystal with a conical indenter, showing the distribution of repatoms, sampling atoms, the QC mesh, and the local centrosymmetry. Colors and panel partitioning are identical to those of Fig. 18.5 (a) through (d). Images adapted from reference [47].

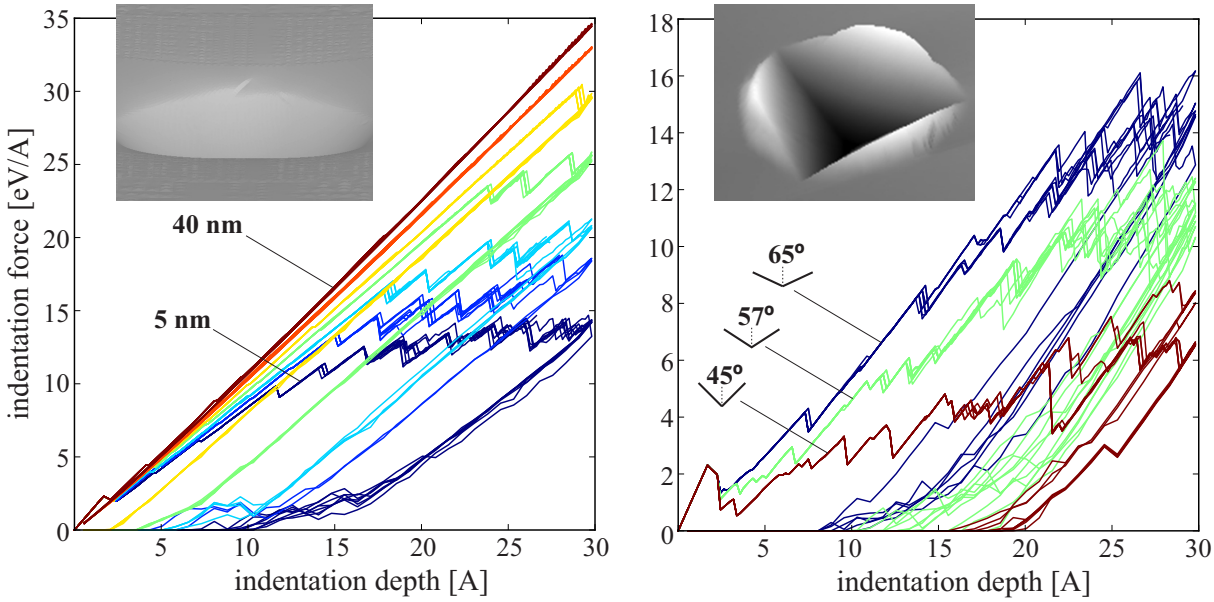


Figure 18.7: Load vs. indentation depth for spherical indenters (of radii 5, 7.5, 10, 15, 20, 30 and 40 nm) and pyramidal indenters (of different pyramidal angles) in 2D for a (100) single-crystalline Cu. Images adapted from reference [47].

demonstrate broad hysteresis loops stemming from incipient plasticity underneath the indenters. Data also show pronounced size effects in case of the spherical indenter as well as clear geometrical effects for the (self-similar) pyramidal indenters.

Figs. 18.8 and 18.9 show 3D nanoindentation simulations with spherical and pyramidal indenters, respectively. Both graphics show results for Cu single-crystals modeled by the extended

Figure 18.8: QC Simulation of 3D nanoindentation with a spherical indenter (radius 40 nm) into single-crystalline (100) Cu up to a penetration depth of about 3 nm, for details see [3]. Images adapted from reference [47].

EAM potential [17]. The spherical indenter has a radius of 40 nm, and results are shown up to an indentation depth of 3 nm; the pyramidal indenter has an angle of 65.3° against the vertical axis and a maximum indentation depth of 5 nm. Via the centrosymmetry parameter, lattice defects have been identified and highlighted (for better visibility, only those atoms are shown which contribute to lattice defects as identified by a higher centrosymmetry parameter).

As shown in these examples, the QC method efficiently reduces computational complexity by assigning full atomistic resolution where it is indeed required (underneath the indenter and near defects). While full resolution provides locally the same accuracy of molecular statics or dynamics, the QC approximation efficiently coarse-grains the remainder of the model domain, thereby allowing for simulations of significantly larger sample sizes. Similarly, by placing grain or twin boundaries as well as cavities and pre-existing defects in the crystal underneath the indenter, the QC method has been used to study defect interactions (see reference [90] for an extended list).

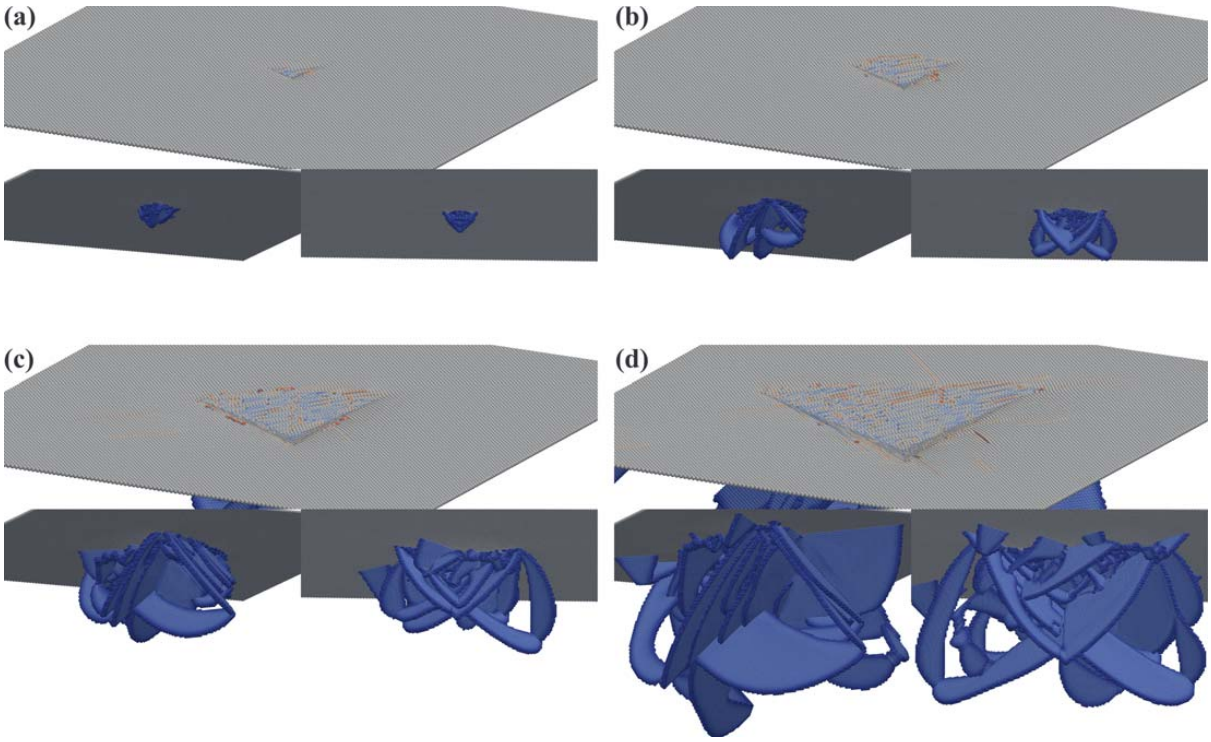


Figure 18.9: QC simulation of 3D nanoindentation with a pyramidal indenter (angle of 65.3° against the vertical axis) into single-crystalline (100) Cu up to a penetration depth of about 5 nm. Images adapted from reference [47].

18.5.2 Surface effects

Surface effects play an important role in the mechanics of nanoscale structures and devices (see Section 13.5). At those small scales, the abundance of free surfaces and the associated high surface-to-volume ratios give rise to – both elastic and plastic – size effects. As an example, consider a thin single-crystalline plate (thickness 12 nm) with a cylindrical hole, which is being pulled uniaxially at remote locations far away from the hole. Near the hole, full atomistic resolution is required to capture defect mechanisms, but MD simulations are too expensive to model large plates. Common MD solutions focus on a small plate instead and either apply periodic boundary conditions (which introduces artifacts due to void interactions) or apply the remote boundary conditions directly (which also introduces artifacts due to edge effects). Here, we use full atomistic resolution near the hole and efficiently coarsen the remaining simulation domain. By using a summation rule that places sampling atoms on element faces (the second-order optimal summation rule from reference [4]), the coarsening has no noticeable impact on the representation of the free surface. Fig. 18.10 illustrates the microstructural evolution for hole radii of 1.5 and 4.5 nm. Dislocations are emitted from the hole at a critical strain level (which increases as the hole radius decreases); the particular dislocation structures are constrained by the free surfaces at the top and bottom.

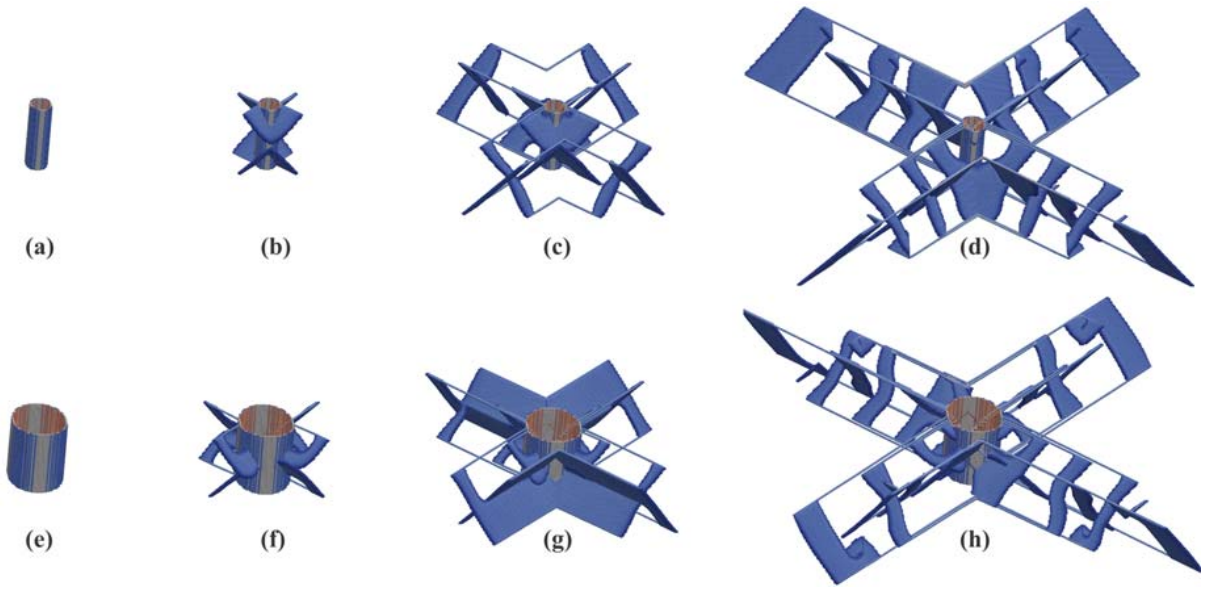


Figure 18.10: QC simulation of dislocations emitted from cylindrical holes in thin films of single-crystalline Cu (hole radii are 1.5 and 4.5 nm in the top and bottom graphics, respectively) at strains of (a) 1.40%, (b) 1.63%, (c) 1.88%, (d) 2.13% and (e) 1.33%, (f) 1.50%, (g) 1.73%, (h) 1.93%. Images adapted from reference [47].

18.5.3 Truss networks

As discussed above, the QC method applies not only to atomic lattices but can also form the basis for the coarse-graining of discrete structural lattices, as long as a thermodynamic potential is available which depends on nodal degrees of freedom. For elastic loading, the potential follows from the strain energy stored within structural components. When individual truss members are loaded beyond the elastic limit, extensions of classical QC are required in order to introduce internal variables and account for such effects as plastic flow, or truss member contact and friction. For such scenarios, effective potentials can be defined, e.g., by using the virtual power or variational constitutive updates theorem [10, 47]. Alternatively, variational constitutive

updates [60] be exploited to introduce effective incremental potentials, as will be discussed below for a truss network [71]. For simplicity, we restrict our study to elastic-plastic bars which only undergo nonlinear stretching deformation and large rotations (but do not deform in bending). Reference [71] contains a detailed description of the extension to stretching- and bending-dominated trusses.

Consider a truss consisting of two-node bars having undeformed lengths L_{ij} and deformed lengths $r_{ij} = |\boldsymbol{r}_{ij}| = |\boldsymbol{q}_i - \boldsymbol{q}_j|$ with nodal positions $\boldsymbol{q}_i$ $(i = 1, \dots, N)$. The total strain of each bar is given by $\varepsilon = (r_{ij} - L_{ij})/L_{ij}$. By adopting the concept of variational constitutive updates [60], we introduce an effective potential for elastic-plastic bars. To this end, we update history variables ε_p^n (the plastic strain) and e_p^n (the accumulated plastic strain) for each bar at each converged load step n. Assume the elastic-plastic effective energy density is given by

$$W(\varepsilon^{n+1}, \Delta\varepsilon_p; \varepsilon_p^n, e_p^n) = \frac{E}{2}\left[\varepsilon^{n+1} - (\varepsilon_p^n + \Delta\varepsilon_p)\right]^2 + H\left(e_p^n + |\Delta\varepsilon_p|\right)^2 + \Delta t\,\tau_0\left|\frac{\Delta\varepsilon_p}{\Delta t}\right|, \tag{18.19}$$

where the first term represents the elastic energy at the new load step $n+1$ (with Young's modulus E), the second term represents linear plastic hardening with hardening modulus H, and the final term defines dissipation in a rate-independent manner (τ_0 is the critical resolved shear stress, and $\Delta t = t_{n+1} - t_n > 0$ is a constant time increment). The effective potential energy of each bar (with cross-sectional area A) is thus given by

$$V\left(\boldsymbol{q}_i^{n+1}, \boldsymbol{q}_j^{n+1}; \varepsilon_p^n, e_p^n\right) = A \cdot \inf_{\Delta\varepsilon_p}\left\{W\left(\frac{r_{ij}^{n+1} - L_{ij}}{L_{ij}}, \Delta\varepsilon_p; \varepsilon_p^n, e_p^n\right)\right\} \tag{18.20}$$

and the new internal variables are

$$\varepsilon_p^{n+1} = \varepsilon_p^n + \Delta\varepsilon_p, \qquad e_p^{n+1} = e_p^n + |\Delta\varepsilon_p|. \tag{18.21}$$

Minimization can be carried out analytically for this simple case. Stationarity yields the kinetic rule of plastic flow:

$$\frac{\partial W}{\partial \Delta\varepsilon_p} = -E\left(\varepsilon^{n+1} - \varepsilon_p^n - \Delta\varepsilon_p\right) + 2H\left(e_p^n + |\Delta\varepsilon_p|\right)\operatorname{sign}(\Delta\varepsilon_p) + \tau_0\,\operatorname{sign}\Delta\varepsilon_p = 0. \tag{18.22}$$

Consider first the case of $\Delta\varepsilon_p > 0$, which leads to

$$\Delta\varepsilon_p = \frac{E\left(\varepsilon^{n+1} - \varepsilon_p^n\right) - (\tau_0 + 2He_p^n)}{E + 2H} > 0. \tag{18.23}$$

Analogously, for $\Delta\varepsilon_p < 0$ we have

$$\Delta\varepsilon_p = \frac{E\left(\varepsilon^{n+1} - \varepsilon_p^n\right) + (\tau_0 + 2He_p^n)}{E + 2H} < 0. \tag{18.24}$$

If none of the two inequalities is satisfied, the bar deforms elastically and $\Delta\varepsilon_p = 0$. Once the new values of the internal variables have been determined, the stress in the bar connecting nodes i and j is

$$\sigma_{ij}^{n+1} = E\left(\frac{r_{ij}^{n+1} - L_{ij}}{L_{ij}} - \left[\varepsilon_{p,ij}^n + \Delta\varepsilon_{p,ij}\right]\right) \tag{18.25}$$

and the corresponding force is $f_{ij} = A\sigma_{ij}$. This completes the description of a simple 1D elastic-plastic constitutive law for each bar. Finally, a failure criterion can be included by defining a critical stress σ_{cr} or ε_{cr} and removing bars whose stress or strain reaches the maximum allowable value.

The total Hamiltonian of the truss structure with nodal positions $\boldsymbol{q} = \{\boldsymbol{q}_1, \dots, \boldsymbol{q}_N\}$ and momenta $\boldsymbol{p} = \{\boldsymbol{p}_1, \dots, \boldsymbol{p}_N\}$ becomes

$$\mathcal{H}(\boldsymbol{q}, \boldsymbol{p}) = \sum_{i=1}^{N} \frac{|\boldsymbol{p}_i|^2}{2\, m_i} + \sum_{i=1}^{N} \sum_{j \in \mathcal{C}(i)} V\left(\boldsymbol{q}_i^{n+1}, \boldsymbol{q}_j^{n+1}; \varepsilon_{p,ij}^n, e_{p,ij}^n\right), \tag{18.26}$$

where we assumed that the mass of each bar is lumped to its nodes, so m_i represents the total mass of node i. $\mathcal{C}(i)$ denotes the *star* of each node, i.e. the set of all adjacent nodes connected to node i through a bar. As the structure of (18.26) is identical to that of atomistics, cf. (12.48), the QC method can be applied in the very same manner as described above for atomistic ensembles, including summation/sampling rules as well as adaptive remeshing. We note that the sums in (18.26) only involve nearest-neighbor interactions, which is why no force artifacts are expected from most classical summation rules even in non-uniform meshes (as long as elements are sufficiently large).

As an example of this approach, we show simulation results for a three-point bending test of a periodic truss lattice with an initial notch. For small numbers of truss members, such simulations can efficiently be run with full resolution (i.e., modeling each truss member as a bar element). Advances in additive manufacturing over the past decade have continuously pushed the frontiers of fabricating micron- and nanometer-sized truss structures. This has resulted in periodic and hierarchical truss lattices with an unprecedented architectural design space of (meta)material properties. As a consequence, modeling techniques are required that can efficiently predict deformation and failure mechanisms in truss structures containing millions or billions of individual truss members. The 3D three-point bending scenario of Fig. 18.11 is one such example, where the QC method reduces computational complexity by efficient coarse-graining of the bar network (bars are assumed to be monolithic and slender, undergoing the stretching-dominated elastic-plastic model described above with a failure strain of 10%). Loads are applied by three cylindrical indenter potentials. To account for manufacturing imperfections and to induce local failure, we vary the elastic modulus of all truss members by a Gaussian distribution (with a standard deviation of 50% of the nominal truss stiffness). Fig. 18.11 compares the simulated truss deformation and stress distribution before and after crack propagation. (a) and (b) demonstrate schematically the coarse-grained nodes, whereas (c) through (f) illustrate the advancing crack front in the fully-resolved region. Like in the atomistic examples in the previous sections, the QC method allows us to efficiently model large periodic systems with locally full resolution where needed.

Of course, this is only an instructive example. Truss members may be subject to flexural deformation, nodes may contribute to deformation mechanisms, and during truss compression one commonly observes buckling and densification accompanied by truss member contact and friction. Model extensions for such effects have been proposed recently and are subject to ongoing research. Fig. 18.12 illustrates an example of a 3D indentation test into a truss lattice, taking into account large rotations of truss members that undergo stretching, bending, and torsion [71]. Also highlighted is the gain in efficiency in terms of the number of "repatoms" (here representative truss unit cells).

18.6 Summary and open challenges

Although the QC method was established almost two decades ago, many challenges have remained and new ones have arisen. These include, among others:

- *Adaptive coarsening*: adaptive model refinement may be geometrically challenging in practice but it is conceptually sound, ultimately turning all lattice sites into repatoms

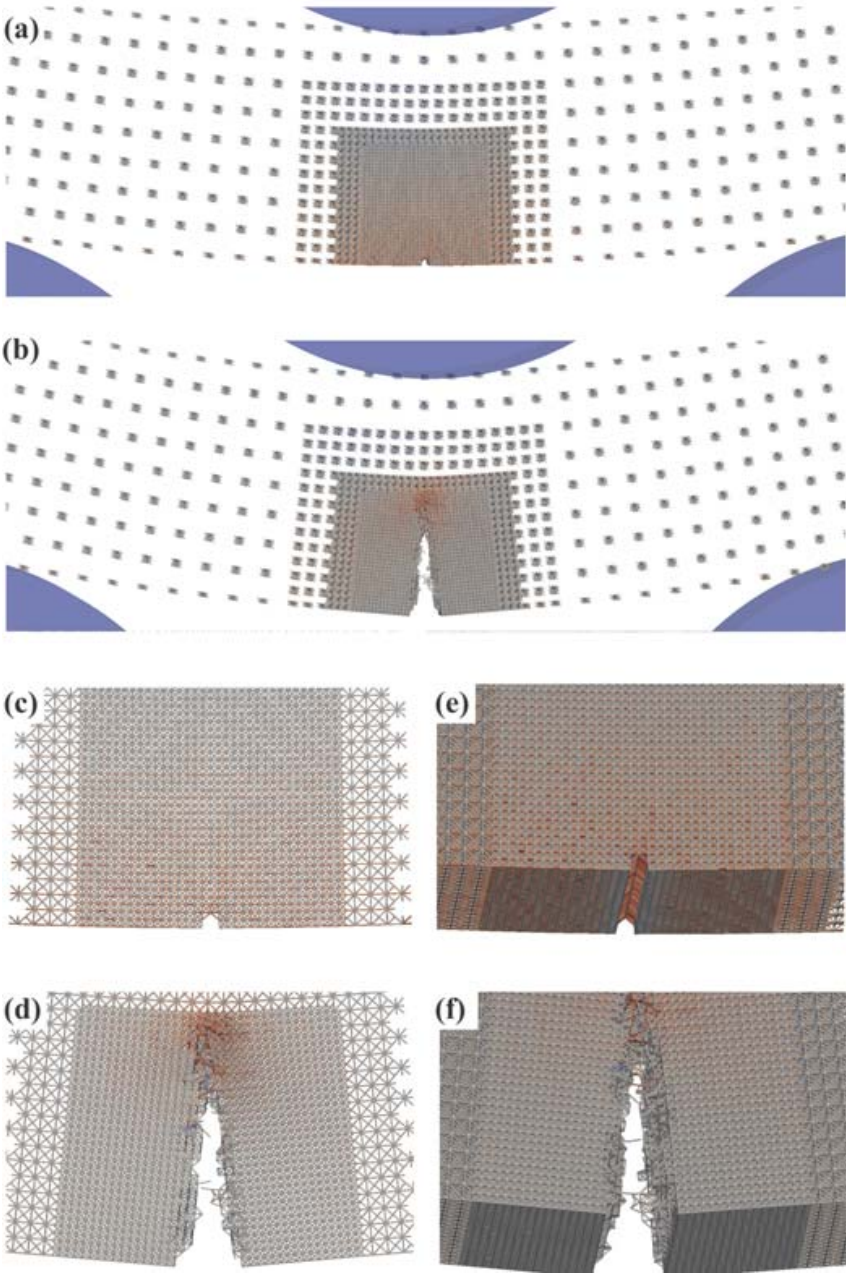


Figure 18.11: Failure of a periodic 3D truss network under three-point bending; color-code indicates stresses within truss members (in order to account for imperfections and stress concentrations; stresses scaled to maximum). (a) and (b) visualize the representative nodes in the coarse-grained QC description (full resolution around the notch and coarsening away from the notch; the three indenters are shown as blue balls). (c/e) and (d/f) show, respectively, the deformed structure with its stress distribution before and after the crack has advanced. Images adapted from reference [47].

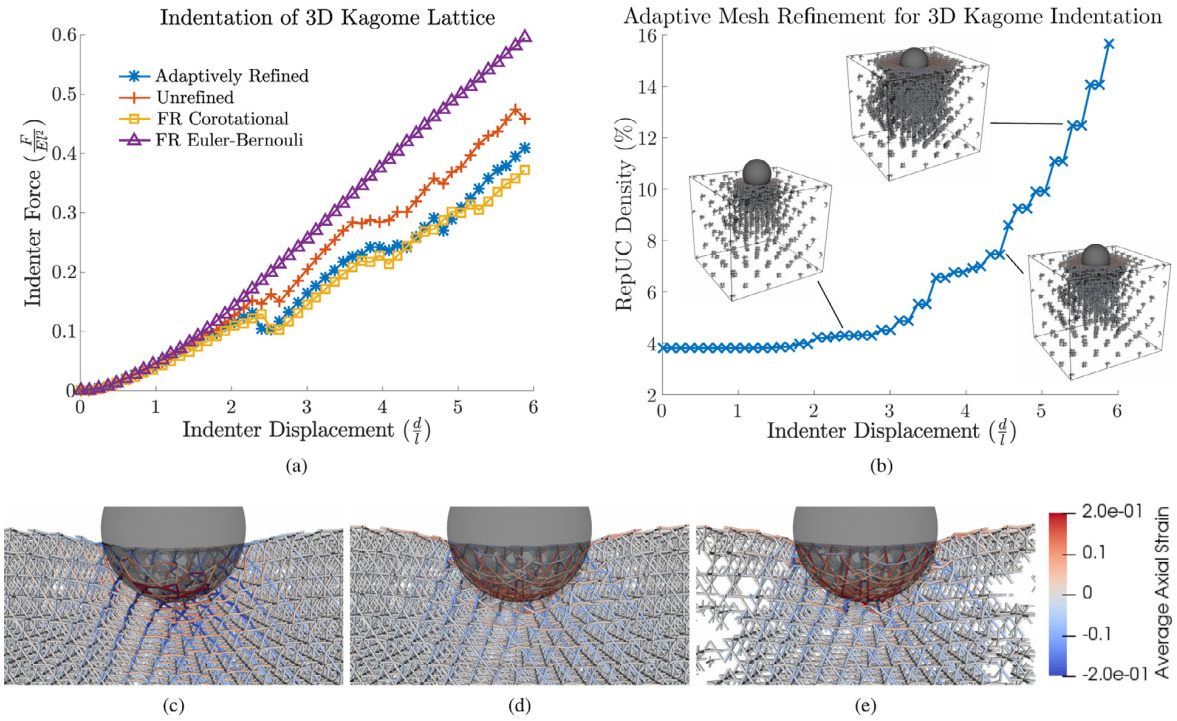


Figure 18.12: Normalized load-displacement curve for the indentation of a 3D kagome lattice (top left), the representative-UC density as the simulation progresses (top right), and slices of the final deformed configurations of the fully-resolved (bottom left) and adaptively refined (bottom right) simulations with average axial strains within bars shown by the color code. Images are adapted from reference [71].

and sampling atoms and thus recovering (locally) molecular dynamics or statics. Model coarsening, in contrast, is challenging conceptually and a key open problem in many QC research codes due to the Lagrangian formulation. Overcoming this difficulty will allow to efficiently coarsen the model by removing full resolution (e.g., in the wake of a traveling lattice defect which leaves behind a perfect crystal that is now fully resolved). Without adaptive model coarsening, simulations will accumulate atomistic resolution and ultimately turn large portions of the model domain into full MD, thus producing prohibitive computational expenses.

- *Large-scale simulations*: the number of existing massively-parallel, distributed-memory QC implementations is small for a variety of reasons, especially in three dimensions. This limits the size of domains that can be simulated. Most existing codes are not open-source but restricted to specific research groups.

- *Dynamics* is a perpetual problem of the QC method. Non-uniform meshes result in wave reflections and wave refraction, which affects phonon motion and thereby corrupts heat propagation within the solid. For these reasons, dynamic QC simulations call for new approaches. Several finite-temperature QC formulations have aimed at overcoming some of the associated problems, but those generally come with phenomenological or simplifying assumptions.

- *Mass and heat transfer* is intimately tied to the long-term dynamics of a (coarse-grained) atomistic ensemble. Recent progress [94] has enabled the computational treatment via statistical mechanics combined with mean-field theory, yet such approaches require constitutive relations to model heat and mass transfer at the atomic scale.

- *Potentials*: every atomistic simulation stands and falls by the accuracy and reliability of its interatomic potentials. Unlike MD, QC may require conceptual extensions when it comes to, e.g., long-range interactions or interaction potentials that not only depend on the positions of the atomic nuclei.
- *Surfaces* play a crucial role in small-scale structures and devices and most QC variants do not properly account for free surfaces (misrepresenting or not at all accounting for surface relaxation). Improved surface representations are hence an open area of research (see e.g. reference [4]).
- *Structures*: the QC approximation has been applied to truss and fiber networks, where new challenges arise due to the complex interaction mechanisms (including, e.g., plasticity, failure, or contact). This branch of the QC family is still in its early stages with many promising applications.
- *Imperfections*: The QC method assumes perfect periodicity of the underlying (crystal or truss) lattice. When geometric imperfections play a dominant role (such as in many nano- and microscale truss networks), new extensions may be required to account for random or systematic variations within the arrangement of representative nodes.

Of course, this can only serve as a short excerpt of the long list of open challenges associated with the family of QC methods and as an open playground for those interested in scale-bridging simulations.

19 Coarse-Grained Atomistics at Finite Temperature

Having discussed various multiscale modeling techniques – both at continuum and discrete scales – we close by discussing a few recent approaches to address the coarse-graining of atomistics at finite temperature. We discuss the extension of the quasicontinuum method (see Section 18) to finite temperature as well as an alternative approach based on Irving-Kirkwood theory. This chapter does by no means attempt to provide a complete overview but simply summarizes a few representative example techniques.

19.1 Finite-Temperature Quasicontinuum

The QC method was introduced (both historically and in Section 18) for simulations at zero temperature, where thermal fluctuations played no role. Therefore, an extension of the methodology is required when dealing with problems at finite temperature (which includes almost all applications of practical relevance). There have been multiple approaches to including temperature within the QC context, three examples of which are briefly summarized in the following.

19.1.1 Maximum-Entropy HotQC

As in classical MD, we consider a system of N particles whose positions and momenta are denoted by, respectively,

$$\boldsymbol{q} = (\boldsymbol{q}_1, \boldsymbol{q}_2, ..., \boldsymbol{q}_N) \tag{19.1}$$

$$\boldsymbol{p} = (\boldsymbol{p}_1, \boldsymbol{p}_2, ..., \boldsymbol{p}_N) \tag{19.2}$$

with $\boldsymbol{p}_i = m_i \dot{\boldsymbol{q}}_i$. Following statistical mechanics (see Section 15), we define the *phase space average* for any quantity A as

$$\langle A \rangle = \frac{1}{N! h^{3n}} \int_\Gamma A(\boldsymbol{q}, \boldsymbol{p})\, \rho(\boldsymbol{q}, \boldsymbol{p})\, \mathrm{d}\boldsymbol{q}\, \mathrm{d}\boldsymbol{p} \tag{19.3}$$

with $\rho(\boldsymbol{q}, \boldsymbol{p})$ denoting the probability distribution and h being Planck's constant

Thermal vibrations motivate the separation of atomic degrees of freedom into **mean motion** and deviations from the latter. To this end, the average (mean) positions and momenta of all particles are defined as, respectively,

$$\overline{\boldsymbol{q}}_a = \langle \boldsymbol{q}_a \rangle, \qquad \overline{\boldsymbol{p}}_a = \langle \boldsymbol{p}_a \rangle. \tag{19.4}$$

We may think of these as the expectation values of particle position and momentum (i.e., the position and momentum in which to find the particles most likely). At zero temperature, we have $\overline{\boldsymbol{q}}_a = \boldsymbol{q}_a$ and $\overline{\boldsymbol{p}}_a = \boldsymbol{p}_a$ without thermal fluctuations. At finite temperature, one defines the standard deviations from the above means (see Figure 19.1) as

$$\langle |\boldsymbol{q}_a - \overline{\boldsymbol{q}}_a|^2 \rangle = 3\tau_a^2, \qquad \langle |\boldsymbol{p}_a - \overline{\boldsymbol{p}}_a|^2 \rangle = 3\sigma_a^2, \tag{19.5}$$

whose ratio we denote by

$$\omega_a = \frac{\sigma_a}{\tau_a}. \tag{19.6}$$

Pre-multiplying $\langle |\boldsymbol{q}_a - \overline{\boldsymbol{q}}_a|^2 \rangle$ by ω^2 and adding the two equations in (19.5) leads to a single constraint for each atom a:

$$\langle |\boldsymbol{p}_a - \overline{\boldsymbol{p}}_a|^2 \rangle + \omega_a^2 \langle |\boldsymbol{q}_a - \overline{\boldsymbol{q}}_a|^2 \rangle = 6\sigma_a^2. \tag{19.7}$$

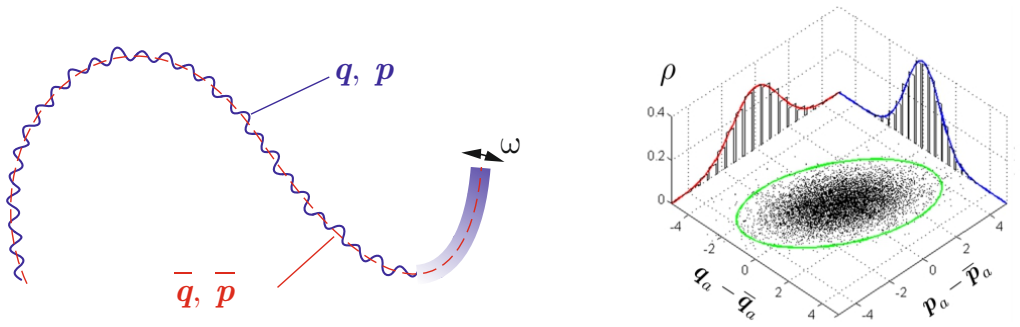


Figure 19.1: Schematic illustration of the separation of atomic motion into mean motion plus statistical variations (on the left), along with a probability distribution ρ for a single atom a, having the typical Gaussian shape centered around the expectation values of position and momentum (on the right). The spread of the Gaussian in the $\boldsymbol{q}$- and $\boldsymbol{p}$-axes is related to τ_a and σ_a, respectively.

Next, one seeks more information about the newly introduced variables by recourse to statistical mechanics. After all, all essential information about the mean position, mean momentum, and the statistical variations are encoded in the probability distribution $\rho(\boldsymbol{p},\boldsymbol{q})$, which we are to find for a given scenario. Inspired by the concepts of classical equilibrium thermodynamics, one has postulated the **principle of maximum entropy** for the probability distribution, as classically known in information theory. In information theory this principle states that the probability distribution ρ that best represents the current state of knowledge is the one with the largest entropy, where entropy is defined as

$$S\left[\rho\right] = -k_B\langle\log\rho\rangle. \tag{19.8}$$

In our context, notice that the probability distribution ρ satisfies the requirements for a probability density, as we require $0 \leq \rho(\boldsymbol{p},\boldsymbol{q})$ and $\langle 1\rangle = \int_\Gamma \rho(\boldsymbol{p},\boldsymbol{q})\,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p} = 1$. This admits evaluating the entropy of our probability distribution via (19.8).

Altogether, to find the probability distribution, one may thus formulate the following maximization problem:

$$\text{find } \rho(\boldsymbol{p},\boldsymbol{q}) = \arg\max S[\rho] \quad \text{s.t.} \quad \langle 1\rangle = 1 \text{ and } \langle|\boldsymbol{p}_a - \overline{\boldsymbol{p}_a}|^2\rangle + \omega_a^2\langle|\boldsymbol{q}_a - \overline{\boldsymbol{q}_a}|^2\rangle = 6\sigma_a^2. \tag{19.9}$$

By introducing Lagrange multipliers λ and β_a $(a = 1,\dots,N)$, this may be written conveniently by introducing the following functional to be maximized with respect to the probability density $\rho(\boldsymbol{q},\boldsymbol{p})$:

$$\mathcal{L}\left[\rho\right] = S\left[\rho\right] - \lambda\langle 1\rangle - \int_\Gamma \sum_{a=1}^{N} \beta_a\rho\left[|\boldsymbol{p}_a - \overline{\boldsymbol{p}_a}|^2 + \omega_a^2|\boldsymbol{q}_a - \overline{\boldsymbol{q}_a}|^2\right]\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{q}, \tag{19.10}$$

This maximization can be carried out analytically, leading to a probability density similar in nature to that of the canonical ensemble:

$$\rho\left(\boldsymbol{q},\boldsymbol{p}\right) = \frac{1}{Z}\exp\left(-\sum_{a=1}^{N}\frac{|p_a - \overline{p}_a|^2 + \omega_a^2|q_a - \overline{q}_a|^2}{2\sigma_a^2}\right) \tag{19.11}$$

with partition function

$$Z = \frac{1}{N!h^{3N}}\prod_{i=1}^{N}\left[\left(\frac{2\pi\sigma_i^2}{\omega_i}\right)^3\right]. \tag{19.12}$$

This probability distribution may be interpreted as a Gaussian distribution, centered at $(\overline{\boldsymbol{q}}_q, \overline{\boldsymbol{p}}_a)$ in phase space, whose standard deviations are defined in terms of ω_a and σ_a.

Insertion of $\rho(\boldsymbol{q}, \boldsymbol{p})$ into the entropy (19.8) and algebraic rearrangements in the limit of large N yields the total entropy as

$$S = \sum_{a=1}^{N} S_a \qquad \text{where} \qquad S_a = 3k_B \log \frac{\sigma_a^2}{\hbar\omega_a} + 4k_B - k_B \log N \tag{19.13}$$

as well as the standard deviations of the particle momenta as

$$\sigma_a = \sqrt{\hbar\omega_a} \exp\left[\frac{1}{6}\left(\frac{S_a}{k_B} - 4 + \log N\right)\right]. \tag{19.14}$$

Loosely speaking, this indicates that the spread in the atomic motion increases with increasing entropy, as may be expected.

In order to proceed to a finite-temperature QC version, we reinterpret the mean Hamiltonian as the *internal energy* E of the system, i.e.,

$$E(\overline{\boldsymbol{p}}, \overline{\boldsymbol{q}}.\boldsymbol{\omega}) = \langle H \rangle, \tag{19.15}$$

where we indicated that the Hamiltonian depends on the mean positions and momenta as well as on $\boldsymbol{\omega} = \{\omega_1, \dots, \omega_N\}$ through the probability distribution ((19.15) requires taking a phase-space average involving the latter). A Legendre transform provides the Helmholtz free energy as

$$F(\overline{\boldsymbol{q}}, \overline{\boldsymbol{p}}, \boldsymbol{\omega}, \boldsymbol{\theta}) = \inf_{\{S\}} \left\{ E - \sum_a \theta_a S_a \right\}, \tag{19.16}$$

where θ_a denotes the "*absolute temperature*" of atom a (here, assuming that one can assign a unique temperature to each atom in the ensemble). The temperature of atom a hence follows from thermodynamics as

$$\theta_a = \frac{\partial E}{\partial S_a} = \frac{\hbar\omega_a}{2m_a k_B} \exp\left(\frac{S_a}{3k_B} - \frac{4}{3} + \frac{\log N}{3}\right), \tag{19.17}$$

from which we obtain the relation between entropy and temperature of an atom a (by solving for S_a) as

$$S_a = 3k_B \log\left(\frac{2k_B m_a \theta_a}{\hbar\omega_a}\right) + 4k_b - k_B \log N. \tag{19.18}$$

As may be expected, the entropy of an atom grows with its temperature.

Finally, inserting everything into the Helmholtz free energy energy (19.16) and evaluating the phase-space average in (19.15) via the probability distribution (19.11) leads to

$$F(\overline{\boldsymbol{q}}, \overline{\boldsymbol{p}}, \boldsymbol{\omega}, \theta) = \sum_{a=1}^{N} \left[\frac{1}{2}\frac{|\overline{p_a}|^2}{m_a} + 3k_B\theta_a \log \frac{\hbar\omega_a}{2k_b m_a \theta_a} + \langle V_a \rangle(\overline{\boldsymbol{q}}, \overline{\boldsymbol{p}}, \boldsymbol{\omega}, \boldsymbol{\theta}) + k_B\theta_a(\log N - 1)\right], \tag{19.19}$$

which depends on the mean motion of atoms, characterized by $(\overline{\boldsymbol{q}}, \overline{\boldsymbol{p}})$, the atomic temperature the temperature θ, as well as the statistical information about the thermal fluctuations, encoded in $\boldsymbol{\omega} = \{\omega_1, \dots, \omega_N\}$. In practice, one may use this new energy to replace the original Hamiltonian.

As the main benefit, this *effective* interatomic potential contains both kinetic and potential energy contributions dependent only on the mean motion (as well as contributions from the thermal fluctuations' kinetic energy and an explicit temperature contribution in the final term). It may thus be evolved at the time scale of the mean motion, instead of resolving the details of thermal vibrations over time – it is hence a technique to *upscale in time* rather than in space (like with the original QC method).

In conclusion, we have replaced the Hamiltonian of the system of N atoms by a new, effective Hamiltonian whose associated "equations of motion" for each atom a at a given temperature θ read

$$m_a \ddot{\overline{\boldsymbol{q}}}_a = -\frac{\partial}{\partial \overline{\boldsymbol{q}}_a} F(\overline{\boldsymbol{q}}, \overline{\boldsymbol{p}}, \boldsymbol{\omega}, \theta), \tag{19.20}$$

$$0 = \frac{\partial}{\partial \omega_a} F(\overline{\boldsymbol{q}}, \overline{\boldsymbol{p}}, \boldsymbol{\omega}, \theta). \tag{19.21}$$

The first equation may be interpreted as a *thermalized* equation of motion, while the second equation is new and required to solve for the statistical fluctuations $\boldsymbol{\omega}$ of all atoms, assuming that the atomic temperatures are fixed (e.g., assuming a uniform temperature across the ensemble for fixed-temperature simulations). This completes the finite-temperature coarse-grained formulation at constant temperature; see, e.g., [51, 94] for details. We note that the extension to a *non-uniform temperature distribution* and the resulting heat flux due to thermal gradients are an open modeling challenge, which has been met with a number of approximations (see, e.g., reference [94]).

While the above concepts have been formulated for an atomistic ensemble containing N atoms, the same formulation can, of course, be applied to a spatially coarse-grained ensemble based on, e.g., the QC method. The resulting technique is commonly referred to as **hotQC** (short for the finite-temperature quasicontinuum method). In this case, the representative atoms carry as degrees of freedom their *mean* positions and momenta $(\boldsymbol{q}_a, \boldsymbol{p}_a)$ as well as ω_a, all of which are coarse-grained by the QC approximations introduced in Section 18. That is, here $(\boldsymbol{q}_a, \boldsymbol{p}_a, \omega_a)$ are interpolated across elements in the coarse-grained regions, resulting in a finite-temperature variant of QC.

19.1.2 Quasiharmonic Approximation

As an alternative approach to deriving a method for finite-temperature coarse-grained atomistics, one may start with the classical probabilty density of the canonical ensemble at finite temperature and apply phase-space averages to the total Hamiltonian of a system including the atomic mean motion and statistical deviations. The phase average of any quantity A, dependent on the mean positions $\overline{\boldsymbol{q}}$ of all atoms at a given (absolute) temperature T, thus becomes

$$A(\overline{\boldsymbol{q}}, T) = \frac{1}{h^{3N} Z} \int_\Gamma A\left(\overline{\boldsymbol{q}} + \delta\boldsymbol{q}, \boldsymbol{p}\right) \exp\left[-\frac{\mathcal{H}(\overline{\boldsymbol{q}} + \delta\boldsymbol{q}, \boldsymbol{p})}{k_B T}\right] \chi(\delta\boldsymbol{q}) \, \mathrm{d}\delta\boldsymbol{q} \, \mathrm{d}\boldsymbol{p}, \tag{19.22}$$

where we integrate over all possible positional fluctuations $\delta\boldsymbol{q} = \boldsymbol{q} - \overline{\boldsymbol{q}} = \{\delta\boldsymbol{q}_1, \dots, \delta\boldsymbol{q}_N\}$. The exponential term represents the probability distribution of the canonical ensemble, and χ is a restrictive indicator function ensuring the admissibility of $\delta\boldsymbol{q}$ (see, e.g., reference [89]). We may use the above, e.g., to calculate an effective kinetic energy – by computing the phase-space-averaged kinetic energy (carrying out all phase-space integrals and rearranging terms). This results in the *effective kinetic energy*

$$\mathcal{T}(\overline{\boldsymbol{p}}, T) = \sum_{a=1}^{N} \frac{\|\overline{\boldsymbol{p}}_a\|^2}{2m_a} + \frac{3k_B T}{2} \sum_{a=1}^{N} \log \frac{2\pi k_B T m_a}{h_p^2} \tag{19.23}$$

with an adjustable parameter h_p. The first term represents the kinetic energy due to mean motion, while the second term captures thermal fluctuations.

The *potential energy* is approximated for small perturbations $\delta\boldsymbol{q}$ by a Taylor expansion, i.e.,

$$\mathcal{V}(\overline{\boldsymbol{q}}+\delta\boldsymbol{q}) = \mathcal{V}(\overline{\boldsymbol{q}}) + \sum_{a=1}^{N} \frac{\partial\mathcal{V}}{\partial\boldsymbol{q}_a}\cdot\delta\boldsymbol{q}_a + \frac{1}{2}\sum_{a,b=1}^{N} \delta\boldsymbol{q}_a\cdot\frac{\partial^2\mathcal{V}}{\partial\boldsymbol{q}_a\,\partial\boldsymbol{q}_b}(\overline{\boldsymbol{q}})\,\delta\boldsymbol{q}_b + \text{h.o.t.} \tag{19.24}$$

Dropping the higher-order terms (and realizing that the linear term vanishes in equilibrium, since it represents the net forces on atoms) results in the **quaharmonic approximation** of the potential energy (cf. Section 13.2):

$$\mathcal{V}(\overline{\boldsymbol{q}}+\delta\boldsymbol{q}) \approx \mathcal{V}(\overline{\boldsymbol{q}}) + \frac{1}{2}\sum_{a,b=1}^{N} \delta\boldsymbol{q}_a\cdot\boldsymbol{\Phi}_{ab}(\overline{\boldsymbol{q}})\,\delta\boldsymbol{q}_b \tag{19.25}$$

with the *force-constant matrix*

$$\boldsymbol{\Phi}_{ab}(\overline{\boldsymbol{q}}) = \frac{\partial^2\mathcal{V}}{\partial\boldsymbol{q}_a\,\partial\boldsymbol{q}_b}(\overline{\boldsymbol{q}}). \tag{19.26}$$

A lengthy calculation (with $h_p\,h_q = h$ and $\boldsymbol{\Phi}_{ab}(\overline{\boldsymbol{q}})$ denoting the force-constant matrix associated with atom a) finally leads to the time-dependent *effective interatomic potential*

$$\mathcal{V}(\overline{\boldsymbol{q}},T) = \mathcal{V}(\overline{\boldsymbol{q}}) - \sum_{a=1}^{N}\frac{k_BT}{2}\log\left[\frac{h_q^{6N}\det\boldsymbol{\Phi}_a(\overline{\boldsymbol{q}})}{(2\pi k_BT)^{3N}}\right]. \tag{19.27}$$

This effective potential may be interpreted as the zero-temperature interatomic potential enriched by the effects of thermal fluctuations, which are assumed to be harmonic vibrations about the atomic equilibrium positions (and whose frequencies are linked to the elastic, quasi-harmonic potential approximation about those equilibria).

Combining the effective kinetic energy (19.23) and the effective potential energy (19.27) (both dependent on temperature and the mean motion of atoms) allows us to derive the effective equations of motion as

$$m_a\ddot{\overline{\boldsymbol{q}}}_q = -\frac{\partial\mathcal{V}}{\partial\overline{\boldsymbol{q}}_a}(\overline{\boldsymbol{q}},T), \tag{19.28}$$

which is, like (19.20) in the hotQC context, formulated in terms of the mean atomic motion – thus condensing out the thermal fluctuations. This makes the method applicable to quasistatic finite-temperature calculations.

While the above method was originally established for atomistic ensembles, it has been extended and applied to the QC formulation, resulting in another **finite-temperature QC** variant. For details see, e.g., references [85, 89].

19.2 Irving-Kirkwood Theory

As an independent finite-temperature coarse-graining theory (which is not related to QC), let us mention the approach that goes back to reference [39] and is named after its inventors the **Irving-Kirkwood theory**. In a nutshell, the idea is to produce *continuum* fields $\mathcal{A}(\boldsymbol{x},t)$ (defined at a time t and at a point $\boldsymbol{x}$ at large scales relative to the atomistics details) from a discrete atomistic ensemble by defining

$$\mathcal{A}(\boldsymbol{x},t) = \int_\Gamma A(\boldsymbol{q},\boldsymbol{p};\boldsymbol{x})\,f(\boldsymbol{q},\boldsymbol{p},t)\,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p} = \langle A,f\rangle, \tag{19.29}$$

where $f(\boldsymbol{q},\boldsymbol{p},t)$ is some appropriate distribution function that is generally time-dependent[10]. The goal is to construct functions $A(\boldsymbol{q},\boldsymbol{p};\boldsymbol{x})$ such that the continuum balance equations have the form

$$\frac{\partial \mathcal{A}}{\partial t}(\boldsymbol{x},t) = \frac{\partial}{\partial t}\langle A, f\rangle. \tag{19.30}$$

As an example, one can define an effective, pointwise **mass density** as

$$\rho^*(\boldsymbol{x},t) = \sum_{a=1}^{N} m_a \langle \delta(\boldsymbol{q}_a - \boldsymbol{x}), f\rangle, \tag{19.31}$$

where the Dirac delta effects that each of the N atoms enters with its mass m_a according the the probability of finding that particular atom a at location $\boldsymbol{x}$ at time t.

Likewise, the pointwise **velocity field** is defined as

$$\boldsymbol{v}^*(\boldsymbol{x},t) = \frac{1}{\rho(\boldsymbol{x},t)} \sum_{a=1}^{N} m_a \langle v_a \delta(\boldsymbol{q}_a - \boldsymbol{x}), f\rangle. \tag{19.32}$$

The analogous procedure may be carried out, e.g., for the energy density. Having defined the continuum field quantities, the classical thermodynamic continuum governing equations can be applied, including the *conservation of mass*, *linear momentum*, and *energy*. This leads to upscaled, continuum-level governing equations based on an underlying discrete atomistic ensemble.

In practice, one does not use the pure ("sharp") Dirac delta function as in the above definitions, but the delta function may be replaced by a "smeared-out" version based on spatial averaging through some *weighting function* w (a so-called *mollifier*). This approach dates back to Hardy in the 1980s (see reference [34]) and introduces, e.g., the continuum-level mass density as

$$\rho(\boldsymbol{x},t) = \int_{\mathbb{R}^d} w(\boldsymbol{y}-\boldsymbol{x})\rho^*(\boldsymbol{y},t)\,\mathrm{d}\boldsymbol{y} \tag{19.33}$$

and the continuum-level velocity field as

$$\boldsymbol{v}(\boldsymbol{x},t) = \int_{\mathbb{R}^d} w(\boldsymbol{y}-\boldsymbol{x})\boldsymbol{v}^*(\boldsymbol{y},t)\,\mathrm{d}\boldsymbol{y}. \tag{19.34}$$

The analogous procedure can be applied to all other field variables of interest. This admits the extraction of continuum-level field quantities from discrete atomistics ensembles, which has been exploited, e.g., in the **Concurrent Atomistic-Continuum** (**CAC**) method to couple atomistics and continuum simulations in a concurrent fashion similar in spirit to the QC method (see reference [100]).

[10] Note that we here break with our naming convention and use f to denote the probability distribution, since ρ will be needed further down for the continuum-level mass density.

20 Discrete Dislocation Dynamics

When discussing multiscale modeling techniques in crystalline solids, there are various other techniques that aim to bridge from discrete features to the continuum. Let us briefly highlight one such technique here, which has become popular in computational mechanics and materials science.

20.1 Discrete Dislocation Dynamics

The method known as **discrete dislocation dynamics** (**DDD**) aims to retain discrete information about dislocations in crystals while providing simulation capabilities for significantly larger length and time scales than MD. The key concept is the description of dislocation motion within a linear elastic continuum, thus coupling the discrete nature of dislocation lines with the continuum formulation of an elastic medium.

Consider a dislocation line segment s that is moving with a velocity v, such that the motion of the dislocation is typical described by

$$m_s \dot{\boldsymbol{v}}_s + \frac{1}{M_s(T,p)} \boldsymbol{v}_s = \boldsymbol{F}_s, \tag{20.1}$$

where m_s is an effective mass density of the dislocation segment, M_s denotes the dislocation mobility (which generally depends on both the local temperature T and pressure p), and $\boldsymbol{F}_s$ is the glide force vector per unit length of the dislocation. The glide force includes various force contributions acting onto the dislocation segment. In practice, one defines

$$\boldsymbol{F}_s = \boldsymbol{F}_{\text{Peierls}} + \boldsymbol{F}_{\text{Self}} + \boldsymbol{F}_{\text{Interaction}} + \boldsymbol{F}_{\text{Obstacle}} + \boldsymbol{F}_{\text{Image}} + \boldsymbol{F}_{\text{Osmotic}} + \boldsymbol{F}_{\text{Thermal}} + \boldsymbol{F}_{\text{External}}, \tag{20.2}$$

which include (in the listed order) the Peierls stress to drive a dislocation in its glide plane, a dislocation self-force (from interactions of this dislocation segment with the remainder of this dislocation), an interaction force stemming from all other dislocations, an interaction forces with obstacles (such as second-phase particles or interfaces and surfaces), image forces, osmotic forces associated with dislocation climb, thermal forces arising from thermal fluctuations, and externally applied loads. The individual contributions are calculated based on the classical continuum theory of dislocations (and extensions), based on the stress field produced a dislocation segment (see reference [35]). The above equation is solved by implicit/explicit time integration. In case of free surfaces, a special treatment is required to deal with the latter, which is

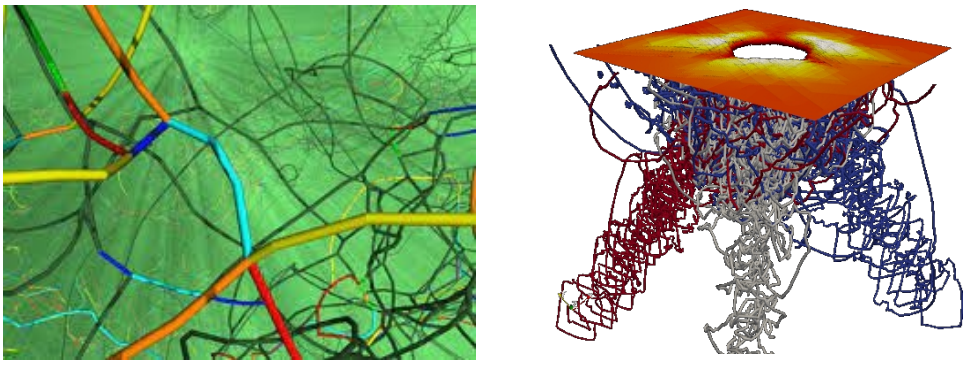


Figure 20.1: Two examples of Discrete Dislocation Dynamics, highlighting a collection of discrete dislocation line segments within a crystal treated as an elastic continuum (the left image is adapted from ParaDiS [7], the right image from TRIDIS [80]).

usually accomplished by the superposition of the associated problems in an infinite body with dislocations and in a finite-size body without dislocations but representative force fields.

Example open-source codes available to perform DDD simulations are **ParaDiS** developed at Stanford University (see reference [7]), and **TRIDIS** developed at SIMaP in Grenoble (see reference [80]).

20.2 Coupled Atomistic Discrete Dislocations

The above DDD framework provides the efficiency of a continuum calculation, while tracking individual dislocations with increased accuracy. Unfortunately, the representation of each dislocation within DDD relies on continuum elasticity solutions (specifically, one calculates the linear elastic solutions for stress and strain fields around a dislocation in an infinite medium and uses those to compute the interaction forces in (20.2)). While this is well suited to simulate the interaction of large systems of dislocations towards extracting their effective mechanical response, it does not provide the level of accuracy of, e.g., an atomistic model.

As a compromise, the method known as **Coupled Atomistic Discrete Dislocations** (**CADD**) was introduced as a *concurrent scale-bridging technique*, which couples two spatially separate simulation domains: one of which uses discrete dislocations (DD), while the other uses MD (or MS) to evolve the system in time. While both DDD and MD (or MS) are well-established, the challenge lies in coupling those two techniques at the interface(s) between the two domains. A dislocation nucleating in the atomistic region and traveling towards the MD-DD interface must be identified and "translated" into a discrete dislocation representation, while dislocations in the DD-domain require the opposite treatment when reaching the interface. For further details see, e.g., references [16, 78, 79]

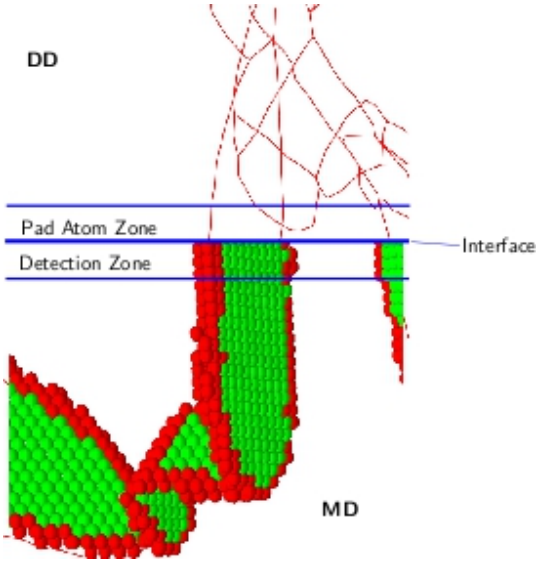


Figure 20.2: Illustration of the CADD technique, coupling atomistics (MD) to a discrete dislocation (DD) representation through a specially designed interface region (adapted from EPFL's LSMS).

References

[1] A. Abdulle, P. Lin, and A. V. Shapeev. Numerical methods for multilattices. *Multiscale Modeling & Simulation*, 10(3):696–726, 2012.

[2] F. F. Abraham, J. Q. Broughton, N. Bernstein, and E. Kaxiras. Spanning the continuum to quantum length scales in a dynamic simulation of brittle fracture. *EPL (Europhysics Letters)*, 44(6):783, 1998.

[3] J. Amelang, G. Venturini, and D. Kochmann. Summation rules for a fully nonlocal energy-based quasicontinuum method. *Journal of the Mechanics and Physics of Solids*, 82:378–413, 2015.

[4] J. S. Amelang and D. M. Kochmann. Surface effects in nanoscale structures investigated by a fully-nonlocal energy-based quasicontinuum method. *Mechanics of Materials*, 90:166–184, 2015.

[5] M. Ariza, I. Romero, M. Ponga, and M. Ortiz. Hotqc simulation of nanovoid growth under tension in copper. *International Journal of Fracture*, 174:75–85, 2012.

[6] E. M. Arruda and M. C. Boyce. A three-dimensional constitutive model for the large stretch behavior of rubber elastic materials. *Journal of the Mechanics and Physics of Solids*, 41(2):389–412, 1993.

[7] T. Arsenlis, V. Bulatov, W. Cai, G. Hommes, M. Rhee, and M. Tang. Paradis. http://paradis.stanford.edu/, 2021.

[8] L. Beex, R. Peerlings, and M. Geers. Central summation in the quasicontinuum method. *Journal of the Mechanics and Physics of Solids*, 70(0):242 – 261, 2014.

[9] L. Beex, C. Verberne, and R. Peerlings. Experimental identification of a lattice model for woven fabrics: Application to electronic textile. *Composites Part A: Applied Science and Manufacturing*, 48:82–92, 2013.

[10] L. A. A. Beex, R. H. J. Peerlings, and M. G. D. Geers. A quasicontinuum methodology for multiscale analyses of discrete microstructural models. *Int. J. Num. Meth. Eng.*, 87(7):701–718, 2011.

[11] T. Belytschko and S. P. Xiao. Coupling methods for continuum model with molecular model. *International Journal for Multiscale Computational Engineering*, 1(1):115–126, 2003.

[12] E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, and P. Gumbsch. Structural relaxation made simple. *Phys. Rev. Lett.*, 97:170201, 2006.

[13] L. Brillouin. *Wave Propagation in Periodic Structures*. Dover Publications, 1953.

[14] S. Brinckmann, D. K. Mahajan, and A. Hartmaier. A scheme to combine molecular dynamics and dislocation dynamics. *Modelling and Simulation in Materials Science and Engineering*, 20(4):045001, 2012.

[15] J. Q. Broughton, F. F. Abraham, N. Bernstein, and E. Kaxiras. Concurrent coupling of length scales: Methodology and application. *Phys. Rev. B*, 60:2391–2403, 1999.

[16] W. A. Curtin and R. E. Miller. Atomistic/continuum coupling in computational materials science. *Modelling and Simulation in Materials Science and Engineering*, 11(3):R33, 2003.

[17] X. D. Dai, Y. Kong, J. H. Li, and B. X. Liu. Extended finnin-sinclair potential for bcc and fcc metals and alloys. *Journal of Physics: Condensed Matter*, 18:4527–4542, 2006.

[18] M. S. Daw and M. I. Baskes. Embedded-atom method: Derivation and application to impurities, surfaces, and other defects in metals. *Phys. Rev. B*, 29:6443–6453, 1984.

[19] K. Dayal and J. Marshall. A multiscale atomistic-to-continuum method for charged defects in electronic materials. In *Presented as the Society of Engineering Science 2011 Annual Technical Conference, Evanston, IL*, 2011.

[20] M. Dobson, R. Elliott, M. Luskin, and E. Tadmor. A multilattice quasicontinuum for phase transforming materials: Cascading cauchy born kinematics. *Journal of Computer-Aided Materials Design*, 14(1):219–237, 2007.

[21] M. Dobson, R. S. Elliott, M. Luskin, and E. B. Tadmor. A multilattice quasicontinuum for phase transforming materials: Cascading cauchy born kinematics. *Journal of Computer-Aided Materials Design*, 14:219–237, 2007.

[22] M. Dobson, M. Luskin, and C. Ortner. Accuracy of quasicontinuum approximations near instabilities. *J. Mech. Phys. Solids*, 58:1741–1757, 2010.

[23] M. Dobson, M. Luskin, and C. Ortner. Sharp stability estimates for the force-based quasicontinuum approximation of homogeneous tensile deformation. *Multiscale Model. Simul.*, 8:782–802, 2010.

[24] M. Dobson, M. Luskin, and C. Ortner. Stability, instability, and error of the force-based quasicontinuum approximation. *Arch. Ration. Mech. Anal.*, 197:179–202, 2010.

[25] L. M. Dupuy, E. B. Tadmor, R. E. Miller, and R. Phillips. Finite-temperature quasicontinuum: Molecular dynamics without all the atoms. *Physical Review Letters*, 95:060202, 2005.

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

[27] M. I. Espanol, D. M. Kochmann, S. Conti, and M. Ortiz. A gamma-convergence analysis of the quasicontinuum method. *SIAM Multiscale Modeling and Simulation*, 11:766–794, 2013.

[28] V. Gavini, K. Bhattacharya, and M. Ortiz. Quasi-continuum orbital-free density-functional theory: A route to multi-million atom non-periodic dft calculation. *Journal of the Mechanics and Physics of Solids*, 55(4):697–718, 2007.

[29] R. N. Glaesener, C. Lestringant, B. Telgen, and D. M. Kochmann. Continuum models for stretching- and bending-dominated periodic trusses undergoing finite deformations. *International Journal of Solids and Structures*, 171:117 – 134, 2019.

[30] R. N. Glaesener, E. A. Träff, B. Telgen, R. M. Canonica, and D. M. Kochmann. Continuum representation of nonlinear three-dimensional periodic truss networks by on-the-fly homogenization. *International Journal of Solids and Structures*, 206:101 – 113, 2020.

[31] M. Gunzburger and Y. Zhang. A quadrature-rule type approximation to the quasicontinuum method. *Multiscale Model. Simul.*, 8(2):571–590, 2010.

[32] S. Hai and E. B. Tadmor. Deformation twinning at aluminum crack tips. *Acta Materialia*, 51:117–131, 2003.

[33] K. Hardikar, V. Shenoy, and R. Phillips. Reconciliation of atomic-level and continuum notions concerning the interaction of dislocations and obstacles. *Journal of the Mechanics and Physics of Solids*, 49(9):1951–1967, 2001.

[34] R. J. Hardy. Formulas for determining local properties in molecular-dynamics simulations: Shock waves. *The Journal of Chemical Physics*, 76(1):622–628, 1982.

[35] J. P. Hirth and J. Lothe. *Theory of Dislocations*. Materials Science and Engineering Series. McGraw-Hill, New York, first edition, 1968.

[36] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864–B871, Nov 1964.

[37] L. Huai-Bao, L. Jun-Wan, N. Yu-Shan, M. Ji-Fa, and W. Hong-Sheng. Multiscale analysis of defect initiation on the atomistic crack tip in body-centered-cubic metal ta. *Acta Physica Sinica*, 60(10):106101, 2011.

[38] M. I. Hussein, M. J. Leamy, and M. Ruzzene. Dynamics of Phononic Materials and Structures: Historical Origins, Recent Progress, and Future Outlook. *Applied Mechanics Reviews*, 66(4), 05 2014. 040802.

[39] J. H. Irving and J. G. Kirkwood. The statistical mechanical theory of transport processes. iv. the equations of hydrodynamics. *The Journal of Chemical Physics*, 18(6):817–829, 1950.

[40] M. Iyer and V. Gavini. A field theoretical approach to the quasi-continuum method. *Journal of the Mechanics and Physics of Solids*, 59(8):1506 – 1535, 2011.

[41] S. Izvekov and G. A. Voth. A multiscale coarse-graining method for biomolecular systems. *The Journal of Physical Chemistry B*, 109(7):2469–2473, 2005.

[42] M. Kabel, T. Böhlke, and M. Schneider. Efficient fixed point and newton–krylov solvers for fft-based homogenization of elasticity at large deformations. *Computational Mechanics*, 54(6):1497–1514, Dec 2014.

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

[44] W. K. Kim, M. Luskin, D. Perez, A. F. Voter, and E. B. Tadmor. Hyper-qc: An accelerated finite-temperature quasicontinuum method using hyperdynamics. *J. Mech. Phys. Solids*, 63:94–112, 2014.

[45] J. Knap and M. Ortiz. An analysis of the quasicontinuum method. *J. Mech. Phys. Solids*, 49(9):1899–1923, September 2001.

[46] D. Kochmann and G. Venturini. A meshless quasicontinuum method based on local maximum-entropy interpolation. *Model. Simul. Mater. Sci. Eng.*, 22:034007, 2014.

[47] D. M. Kochmann and J. S. Amelang. The quasicontinuum method: Theory and applications. In C. R. Weinberger and G. J. Tucker, editors, *Multiscale Materials Modeling for Nanomechanics*, chapter 10, pages 159–193. Springer, New York, 2016.

[48] D. M. Kochmann and G. W. Milton. Rigorous bounds on the effective moduli of composites and inhomogeneous bodies with negative-stiffness phases. *Journal of the Mechanics and Physics of Solids*, 71:46–63, 2014.

[49] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, Nov 1965.

[50] A. Krushynska, M. Miniaci, F. Bosia, and N. Pugno. Coupling local resonance with bragg band gaps in single-phase mechanical metamaterials. *Extreme Mechanics Letters*, 12:30–36, 2017. Frontiers in Mechanical Metamaterials.

[51] Y. Kulkarni, J. Knap, and M. Ortiz. A variational approach to coarse graining of equilibrium and non-equilibrium atomistic description at finite temperature. *Journal of the Mechanics and Physics of Solids*, 56:1417–1449, 2008.

[52] S. Kwon, Y. Lee, J. Y. Park, D. Sohn, J. H. Lim, and S. Im. An efficient three-dimensional adaptive quasicontinuum method using variable-node elements. *Journal of Computational Physics*, 228(13):4789–4810, 2009.

[53] J. Li, H. Lu, Y. Ni, and J. Mei. Quasicontinuum study the influence of misfit dislocation interactions on nanoindentation. *Computational Materials Science*, 50(11):3162–3170, 2011.

[54] J. Li, J. Mei, Y. Ni, H. Lu, and W. Jiang. Two-dimensional quasicontinuum analysis of the strengthening and weakening effect of cu/ag interface on nanoindentation. *Journal of Applied Physics*, 108(5):054309, 2010.

[55] X. H. Li, M. Luskin, C. Ortner, and A. V. Shapeev. Theory-based benchmarking of the blended force-based quasicontinuum method. *Computer Methods in Applied Mechanics and Engineering*, 268(0):763 – 781, 2014.

[56] W. K. Liu, H. S. Park, D. Q., E. G. Karpov, H. Kadowaki, and G. J. Wagner. Bridging scale methods for nanomechanics and materials. *Computer Methods in Applied Mechanics and Engineering*, 195(13-16):1407–1421, 2006.

[57] H. Lu, Y. Ni, J. Mei, J. Li, and H. Wang. Anisotropic plastic deformation beneath surface step during nanoindentation of fcc al by multiscale analysis. *Computational Materials Science*, 58:192–200, 2012.

[58] M. Luskin and C. Ortner. An analysis of node-based cluster summation rules in the quasicontinuum method. *SIAM J. Numer. Anal.*, 47(4):3070–3086, 2009.

[59] M. Luskin, C. Ortner, and B. V. Koten. Formulation and optimization of the energy-based blended quasicontinuum method. *Computer Methods in Applied Mechanics and Engineering*, 253(0):160 – 168, 2013.

[60] O. M. and S. L. The variational formulation of viscoplastic constitutive updates. *Computer Methods in Applied Mechanics and Engineering*, 171(3):419–444, 1999.

[61] J. Marian, G. Venturini, B. L. Hansen, J. Knap, M. Ortiz, and G. H. Campbell. Finite-temperature extension of the quasicontinuum method using langevin dynamics: entropy losses and analysis of errors. *Modelling and Simulation in Materials Science and Engineering*, 18(1):015003, 2010.

[62] J. Marshall and K. Dayal. Atomistic-to-continuum multiscale modeling with long-range electrostatic interactions in ionic solids. *Journal of the Mechanics and Physics of Solids*, 62(0):137 – 162, 2014. Sixtieth anniversary issue in honor of Professor Rodney Hill.

[63] C. Miehe and A. Koch. Computational micro-to-macro transitions of discretized microstructures undergoing small strains. *Archive of Applied Mechanics*, 72(4):300–317, Jul 2002.

[64] R. Miller, M. Ortiz, R. Phillips, V. Shenoy, and E. Tadmor. Quasicontinuum models of fracture and plasticity. *Engineering Fracture Mechanics*, 61:427–444, 1998.

[65] R. E. Miller and E. B. Tadmor. A unified framework and performance benchmark of fourteen multiscale atomistic/continuum coupling methods. *Modelling Simul. Mater. Sci. Eng.*, 17:053001, 2009.

[66] H. Moulinec and P. Suquet. A numerical method for computing the overall response of nonlinear composites with complex microstructure. *Computer Methods in Applied Mechanics and Engineering*, 157(1):69–94, 1998.

[67] A. Nair, D. Warner, R. Hennig, and W. Curtin. Coupling quantum and continuum scales to predict crack tip dislocation nucleation. *Scripta Materialia*, 63(12):1212 – 1215, 2010.

[68] C. Ortner. A priori and a posteriori analysis of the quasinonlocal quasicontinuum method in 1d. *Math. Comp.*, 80(275):1265–1285, 2011.

[69] J. Y. Park and S. Im. Adaptive nonlocal quasicontinuum for deformations of curved crystalline structures. *Phys. Rev. B*, 77:184109, 2008.

[70] R. Phillips, D. Rodney, V. Shenoy, E. Tadmor, and M. Ortiz. Hierarchical models of plasticity: dislocation nucleation and interaction. *Modelling Simul. Mater. Sci. Eng.*, 7:769–780, 1999.

[71] G. P. Phlipot and D. M. Kochmann. A quasicontinuum theory for the nonlinear mechanical response of general periodic truss lattices. *Journal of the Mechanics and Physics of Solids*, 124:758–780, 2019.

[72] S. Plimpton. Fast parallel algorithms for short-range molecular dynamics. *J. Comp. Phys.*, 117:1 – 19, 1995.

[73] I. Ringdalen Vatne, E. Ostby, and C. Thaulow. Multiscale simulations of mixed-mode fracture in bcc-fe. *Modelling and Simulation in Materials Science and Engineering*, 19(8):085006, 2011.

[74] R. E. Rudd and J. Q. Broughton. Coarse-grained molecular dynamics: Nonlinear finite elements and finite temperature. *Physical Review B*, 72:144104, 2005.

[75] R. E. Rudd and J. Q. Broughton. Coarse-grained molecular dynamics: Nonlinear finite elements and finite temperature. *Phys. Rev. B*, 72:144104, Oct 2005.

[76] V. Shenoy, V. Shenoy, and R. Phillips. Finite temperature quasicontinuum methods. *Materials Research Society Symposium Proceedings*, 538:465–471, 1999.

[77] V. B. Shenoy, R. Miller, E. B. Tadmor, R. Phillips, and M. Ortiz. Quasicontinuum models of interfacial structure and deformation. *Phys. Rev. Lett.*, 80:742–745, 1998.

[78] L. Shilkrot, R. E. Miller, and W. A. Curtin. Coupled atomistic and discrete dislocation plasticity. *Phys. Rev. Lett.*, 89:025501, 2002.

[79] L. Shilkrot, R. E. Miller, and W. A. Curtin. Multiscale plasticity modeling: coupled atomistics and discrete dislocation mechanics. *Journal of the Mechanics and Physics of Solids*, 52:755 – 787, 2004.

[80] SIMaP. Tridis. http://www.numodis.fr/tridis/, 2021.

[81] V. Sorkin, R. S. Elliott, and E. B. Tadmor. A local quasicontinuum method for 3d multilattice crystalline materials: Application to shape-memory alloys. *Modelling and Simulation in Materials Science and Engineering*, 22(5):055001, 2014.

[82] A. Stukowski. Structure identification methods for atomistic simulations of crystalline materials. *arxiv*, 2012.

[83] P. Suryanarayana. *Coarse-graining Kohn-Sham density functional theory*. PhD thesis, California Institute of Technology, 2011.

[84] E. Tadmor and S. Hai. A peierls criterion for the onset of deformation twinning at a crack tip. *Journal of the Mechanics and Physics of Solids*, 51(5):765 – 793, 2003.

[85] E. Tadmor, F. Legoll, W. Kim, L. Dupuy, and R. Miller. Finite-temperature quasi-continuum. *Applied Mechanics Reviews*, 65(1):010803, 2013.

[86] E. Tadmor, R. Miller, R. Philipps, and M. Ortiz. Nanoindentation and incipient plasticity. *Journal of Materials Research*, 14:2233–2250, 1999.

[87] E. Tadmor, M. Ortiz, and R. Phillips. Quasicontinuum analysis of defects in solids. *Philos. Mag. A*, 73(6):1529–1563, June 1996.

[88] E. Tadmor, U. Waghmare, G. Smith, and E. Kaxiras. Polarization switching in pbtio3: an ab initio finite element simulation. *Acta Materialia*, 50(11):2989 – 3002, 2002.

[89] E. B. Tadmor and R. E. Miller. *Modeling Materials: Continuum, Atomistic and Multiscale Techniques*. Cambridge University Press, 2011.

[90] E. B. Tadmor and R. E. Miller. http://www.qcmethod.org, 2021.

[91] Z. Tang, H. Zhao, G. Li, and N. R. Aluru. Finite-temperature quasicontinuum method for multiscale analysis of silicon nanostructures. *Physical Review B*, 74:064110, 2006.

[92] I. Tembhekar, J. S. Amelang, L. Munk, and D. M. Kochmann. Automatic adaptivity in the fully nonlocal quasicontinuum method for coarse-grained atomistic simulations. *International Journal for Numerical Methods in Engineering*, 110(9):878–900, 2017.

[93] I. Temizer. Micromechanics. https://sourceforge.net/projects/multiscale/, 2020.

[94] G. Venturini, K. Wang, I. Romero, M. Ariza, and M. Ortiz. Atomistic long-term simulation of heat and mass transport. *Journal of the Mechanics and Physics of Solids*, 73:242–268, 2014.

[95] A. Vidyasagar, W. Tan, and D. Kochmann. Predicting the effective response of bulk polycrystalline ferroelectric ceramics via improved spectral phase field methods. *Journal of the Mechanics and Physics of Solids*, 106:133 – 151, 2017.

[96] A. Vidyasagar, A. D. Tutcuoglu, and D. M. Kochmann. Deformation patterning in finite-strain crystal plasticity by spectral homogenization with application to magnesium. *Computer Methods in Applied Mechanics and Engineering*, 335:584–609, 2018.

[97] A. Voter. A method for accelerating the molecular dynamics simulation of infrequent events. *The Journal of Chemical Physics*, 106(11):4665–4677, 1997.

[98] S. Xiao and W. Yang. A temperature-related homogenization technique and its implementation in the meshfree particle method for nanoscale simulations. *International Journal for Numerical Methods in Engineering*, 69(10):2099–2125, 2007.

[99] Q. Yang, B. E., and A. To. Multiresolution molecular mechanics: Statics. *Computer Methods in Applied Mechanics and Engineering*, 258(0):26 – 38, 2013.

[100] S. Yang and Y. Chen. Concurrent atomistic and continuum simulation of bi-crystal strontium titanate with tilt grain boundary. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2175):20140758, 2015.

[101] W. Yang and S. Xiao. The applications of meshfree particle methods at the nanoscale. In V. Sunderam, G. Albada, P. Sloot, and J. Dongarra, editors, *Computational Science – ICCS 2005*, volume 3516 of *Lecture Notes in Computer Science*, pages 284–291. Springer Berlin Heidelberg, 2005.

[102] H. Yoshihiko, Y. Nobuhiro, S. Dmitriev, K. Masanori, and T. Shingo. Large scale atomistic simulation of cu/al2o3 interface via quasicontinuum analysis. *Journal of the Japan Institute of Metals*, 69(1):90–95, 2005.

[103] W. Yu and S. Shen. Initial dislocation topologies of nanoindentation into copper film with a nanocavity. *Engineering Fracture Mechanics*, 77(16):3329 – 3340, 2010.

[104] A. J. Zelhofer and D. M. Kochmann. On acoustic wave beaming in two-dimensional structural lattices. *International Journal of Solids and Structures*, 115-116:248–269, 2017.

Appendix

A Review of Continuum Mechanics and Notation

A.1 An introductory example: heat conduction

We describe a body $\Omega \subset \mathbb{R}^d$ with boundary $\partial\Omega$ as a collection of **material points**. Each point has a position $\boldsymbol{x}$ in a Cartesian coordinate system $(x_1, \dots, x_d)$ in d dimensions with origin $\boldsymbol{O}$.

Points are described by vectors defined by components in the Cartesian reference frame:

$$\boldsymbol{x} = \sum_{i=1}^{d} x_i \boldsymbol{g}_i = x_i \boldsymbol{g}_i. \tag{A.1}$$

Here and in the following we use Einstein's *summation convention* which implies summation over repeated indices. The usual index notation rules apply; e.g., the inner product is written as

$$\boldsymbol{a} \cdot \boldsymbol{b} = a_i \boldsymbol{g}_i \cdot b_j \boldsymbol{g}_j = a_i b_j \delta_{ij} = a_i b_i \qquad \text{with Kronecker's delta} \quad \delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{else.} \end{cases} \tag{A.2}$$

This is used to define the length of a vector as

$$\|\boldsymbol{a}\| = \sqrt{\boldsymbol{a} \cdot \boldsymbol{a}} = \sqrt{a_i a_i} = \sqrt{\sum_{i=1}^{d} a_i a_i}. \tag{A.3}$$

Matrix-vector multiplication becomes

$$\boldsymbol{a} = \boldsymbol{M}\boldsymbol{b} \qquad \Leftrightarrow \qquad a_i = M_{ij} b_j \qquad \text{and} \qquad (\boldsymbol{M}^{\mathrm{T}})_{ij} = M_{ji}. \tag{A.4}$$

We use **mappings** to denote fields. For example, the temperature field in a static problem is described by a mapping

$$T(\boldsymbol{x}) : \Omega \to \mathbb{R}, \tag{A.5}$$

which assigns to each point $\boldsymbol{x} \in \Omega$ a real number, the temperature. If the field is differentiable, one often introduces **kinematic variables** such as the temperature *gradient* field:

$$\boldsymbol{\beta} : \Omega \to \mathbb{R}^d \qquad \text{and} \qquad \boldsymbol{\beta} = \operatorname{grad} T = \nabla T \qquad \Leftrightarrow \qquad \beta_i = \frac{\partial T}{\partial x_i} = T_{,i}. \tag{A.6}$$

Here and in the following, we use *comma indices* to denote partial derivatives.

For every kinematic variable, there is conjugate field (often called **flux**) like the heat flux $\boldsymbol{q}$ in this thermal problem, which is also a mapping:

$$\boldsymbol{q} : \Omega \to \mathbb{R}^d. \tag{A.7}$$

The heat flux vector $\boldsymbol{q}$ assigns to each point in Ω a heat flux direction and magnitude. If we are interested, e.g., in the loss of heat through a point $\boldsymbol{x} \in \partial\Omega$ on the surface of Ω with outward unit normal $\boldsymbol{n}(\boldsymbol{x})$, then that amount of heat leaving Ω is the projection $\boldsymbol{q}(\boldsymbol{x}) \cdot \boldsymbol{n}(\boldsymbol{x})$. Clearly, the components of $\boldsymbol{q} = (q_1, \dots, q_d)^{\mathrm{T}}$ imply the heat flux through surfaces perpendicular to each of the d Cartesian coordinate directions.

Next, **constitutive relations** link kinematic quantities to fluxes. For example, define the heat flux vector as $\boldsymbol{q} = \boldsymbol{q}(\boldsymbol{\beta})$. Fourier's law of heat conduction states, e.g.,

$$\boldsymbol{q} = -\boldsymbol{K}\boldsymbol{\beta} \qquad \Leftrightarrow \qquad q_i = -K_{ij}\beta_{,j}, \tag{A.8}$$

where $\boldsymbol{K}$ denotes a conductivity *tensor*. K_{ij} are the components of the conductivity tensor; they form a $d \times d$ matrix. Such a *second-order tensor* is a convenient way to store the conductivity properties in any arbitrary orientation in the form of a matrix, along with the coordinate basis in which it is defined. $\boldsymbol{K}$ provides for each direction of the temperature gradient $\boldsymbol{\beta} = \alpha\boldsymbol{n}$ the resulting normalized heat flux $\boldsymbol{q}$. To this end, one defines

$$\boldsymbol{K} = K_{ij}\boldsymbol{g}_i \otimes \boldsymbol{g}_j \qquad \Rightarrow \qquad \boldsymbol{K}\boldsymbol{n} = (K_{ij}\boldsymbol{g}_i \otimes \boldsymbol{g}_j)\boldsymbol{n} = K_{ij}\boldsymbol{g}_i(\boldsymbol{g}_j \cdot \boldsymbol{n}) = K_{ij}n_j\boldsymbol{g}_i. \tag{A.9}$$

Such a second-order tensor is hence a linear mapping of vectors onto vectors. Here, we defined the *dyadic product* (or *tensor product*), which produces a tensor according to

$$\boldsymbol{M} = \boldsymbol{a} \otimes \boldsymbol{b} \qquad \Leftrightarrow \qquad M_{ij} = a_i b_j. \tag{A.10}$$

A special tensor is the identity, which maps vectors onto themselves:

$$\boldsymbol{I} = \delta_{ij}\,\boldsymbol{g}_i \otimes \boldsymbol{g}_j. \tag{A.11}$$

To solve a thermal problem, we also need **balance laws**. Here, *conservation of energy* may be used, which states the change of **internal energy** E in a body over time t balances the inward and outward flux of energy and the energy being produced inside the body (by some heat source density ρs). Mathematically, this implies

$$\frac{\mathrm{d}}{\mathrm{d}t}E = \int_\Omega \rho s \,\mathrm{d}V - \int_{\partial\Omega} \boldsymbol{q}\cdot\boldsymbol{n}\,\mathrm{d}S. \tag{A.12}$$

We can use the **divergence theorem** to write

$$\int_{\partial\Omega} \boldsymbol{q}\cdot\boldsymbol{n}\,\mathrm{d}S = \int_{\partial\Omega} q_i n_i\,\mathrm{d}S = \int_\Omega q_{i,i}\,\mathrm{d}S = \int_\Omega \operatorname{div}\boldsymbol{q}\,\mathrm{d}S, \tag{A.13}$$

which defines the *divergence of a vector* as the scalar quantity

$$\operatorname{div}(\cdot) = (\cdot)_{i,i}. \tag{A.14}$$

Energy is an **extensive** variable, i.e., the total energy doubles when adding to bodies of the same energy. This is in contrast to **intensive** variables, such as temperature or pressure, when adding to bosides having the same, e.g., temperature. Since energy is an extensive variable, we can introduce an energy density e and write

$$E = \int_\Omega e\,\mathrm{d}V \qquad \Rightarrow \qquad \frac{\mathrm{d}}{\mathrm{d}t}E = \int_\Omega \dot{e}\,\mathrm{d}V. \tag{A.15}$$

Here and in the following, we use dots to denote *rates*, i.e., *time derivatives*. Note that we use a *Lagrangian* description and that, so far, there is no motion or deformation involved in our discussion, so the time derivative does not affect the volume integral. Rewriting the conservation of energy now yields

$$\int_\Omega \dot{e}\,\mathrm{d}V = \int_\Omega \rho s\,\mathrm{d}V - \int_\Omega \operatorname{div}\boldsymbol{q}\,\mathrm{d}V. \tag{A.16}$$

This can be rewritten as

$$\int_\Omega (\dot{e} - \rho s + \operatorname{div} \boldsymbol{q}) \; \mathrm{d}V = 0. \tag{A.17}$$

Since conservation of energy does not only have to hold for Ω but for any subbody $\omega \subset \Omega$, we may conclude that the **local energy balance** equation is

$$\boxed{\dot{e} = \rho s - \operatorname{div} \boldsymbol{q}} \tag{A.18}$$

This is the local (i.e., pointwise) counterpart to the macroscopic energy balance and states that at each point $\boldsymbol{x} \in \Omega$ the rate of energy change ($\dot{e}$) is given by the local production of heat (ρs) minus the heat lost by outward fluxes $\boldsymbol{q}$ away from the point.

Finally, exploiting that thermally stored energy gives $e = \rho c_v T$ (with constant mass density ρ and specific heat capacity c_v), and we insert Fourier's law of heat conduction to overall arrive at

$$\int_\Omega \rho c_v \dot{T} \; \mathrm{d}V = \int_\Omega \rho s \; \mathrm{d}V - \int_\Omega \operatorname{div}(-\boldsymbol{K}\boldsymbol{\beta}) \; \mathrm{d}S = \int_\Omega \rho s \; \mathrm{d}V - \int_\Omega (-K_{ij} T_{,j})_{,i} \; \mathrm{d}S. \tag{A.19}$$

Note that the final term requires the use of the product rule since

$$(-K_{ij} T_{,j})_{,i} = -K_{ij,i} T_{,j} - K_{ij} T_{,ij}. \tag{A.20}$$

Let us assume a **homogeneous** body with $\boldsymbol{K}(\boldsymbol{x}) = \boldsymbol{K} = \text{const}$. Further, let us rewrite the above under a single integral:

$$\int_\Omega \left(\rho c_v \dot{T} - \rho s - K_{ij} T_{,ij} \right) \; \mathrm{d}V = 0. \tag{A.21}$$

Again, by extension of energy conservation to arbitrary subbodies, we conclude the local energy balance equation

$$\rho c_v \dot{T} = K_{ij} T_{,ij} + \rho s \tag{A.22}$$

This is the heat equation in its *anisotropic form*. In the special case of **isotropy** (i.e., the conductivity is the same in all directions), we obtain the *Laplacian* since

$$K_{ij} = \kappa \delta_{ij} \qquad \Rightarrow \qquad K_{ij} T_{,ij} = \kappa \delta_{ij} T_{,ij} = \kappa T_{,ii} = \kappa \, \Delta T \qquad \text{with} \qquad \Delta(\cdot) = (\cdot)_{,ii}, \tag{A.23}$$

and we arrive at the well-known **heat equation** with sources:

$$\boxed{\rho c_v \dot{T} = \kappa \, \Delta T + \rho s} \tag{A.24}$$

Whenever we consider a *static* problem, we assume that the body is in equilibrium and the temperature field is constant, which reduces the above to *Poisson's equation*, viz.

$$\kappa \, \Delta T = -\rho s. \tag{A.25}$$

A.2 Mechanical equilibrium

The mechanics of solids (and fluids) generally describes deformable bodies. To this end, we label each material point by its position $\boldsymbol{X}$ in a **reference configuration** (e.g., the configuration at time $t = 0$). The current position $\boldsymbol{x}$, by contrast, is a function of $\boldsymbol{X}$ and of time t: $\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{X}, t)$. Fields in the reference and **current configuration** are generally referred to by upper- and lower-case characters (and the same applies to indices), e.g.,

$$\boldsymbol{x} = x_i \boldsymbol{g}_i, \qquad \boldsymbol{X} = X_I \boldsymbol{G}_I. \tag{A.26}$$

Note that, to avoid complication, we will work with Cartesian coordinate systems only (see the tensor notes, e.g., for curvilinear coordinates).

The mechanics of a deformable body undergoing finite deformations is generally described by a **deformation mapping**

$$\boldsymbol{\varphi}(\boldsymbol{X}, t) : \Omega \times \mathbb{R} \to \mathbb{R}^d \qquad \text{such that} \qquad \boldsymbol{x} = \boldsymbol{\varphi}(\boldsymbol{X}, t). \tag{A.27}$$

Since it depends on time, we can take time derivatives to arrive at the **velocity** and **acceleration** fields, respectively:

$$\boldsymbol{V}(\boldsymbol{X}, t) = \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{x}(\boldsymbol{X}, t) = \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\varphi}(\boldsymbol{X}, t), \qquad \boldsymbol{A}(\boldsymbol{X}, t) = \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{V}(\boldsymbol{X}, t) = \frac{\mathrm{d}^2}{\mathrm{d}t^2}\boldsymbol{\varphi}(\boldsymbol{X}, t). \tag{A.28}$$

Note that those are **Lagrangian** fields (one could also write those as functions of the current position $\boldsymbol{x}$, which results in the **Eulerian** counterparts; this is usually done in fluid mechanics).

Like in the thermal problem, we introduce *kinematics* by defining the **deformation gradient**

$$\boldsymbol{F} = \operatorname{Grad} \boldsymbol{\varphi} \qquad \Leftrightarrow \qquad F_{iJ} = \frac{\partial \varphi_i}{\partial X_J} = \varphi_{i,J}. \tag{A.29}$$

Note that this is a second-order, *two-point tensor* defined across both configurations. If one uses the same coordinate frame for the undeformed and deformed configurations, one may alternative introduce the **displacement field**

$$\boldsymbol{u}(\boldsymbol{X}) = \boldsymbol{x} - \boldsymbol{X} \qquad \text{and} \qquad \boldsymbol{x} = \boldsymbol{X} + \boldsymbol{u}(\boldsymbol{X}), \tag{A.30}$$

so that

$$\boldsymbol{F} = \operatorname{Grad} \boldsymbol{\varphi} = \operatorname{Grad}(\boldsymbol{X} + \boldsymbol{u}) = \boldsymbol{I} + \operatorname{Grad} \boldsymbol{u} \qquad \Leftrightarrow \qquad F_{iJ} = \delta_{iJ} + u_{i,J}. \tag{A.31}$$

Note that in case of *no deformation*, we have $\boldsymbol{x} = \boldsymbol{X}$ so that $\boldsymbol{F} = \boldsymbol{I}$ (and *not* $\boldsymbol{F} = \boldsymbol{0}$).

$\boldsymbol{F}$ contains plenty of information about the local deformation. For example, the **volume change** at a point is given by

$$\frac{\mathrm{d}v}{\mathrm{d}V} = J = \det \boldsymbol{F}, \tag{A.32}$$

and for physical reasons we must have $J > 0$ (this ensures that the deformation mapping is injective; i.e., no two material points are mapped onto the same point in the current configuration). No volume change implies $J = 1$.

Similarly, the **stretch** in a direction defined by a unit vector $\boldsymbol{N}$ is determined by

$$\lambda(\boldsymbol{N}) = \frac{\mathrm{d}s}{\mathrm{d}S} = \sqrt{\boldsymbol{N} \cdot \boldsymbol{C}\boldsymbol{N}} \qquad \text{with} \qquad \boldsymbol{C} = \boldsymbol{F}^{\mathrm{T}}\boldsymbol{F} \tag{A.33}$$

the **right Cauchy-Green tensor**. As for $\boldsymbol{F}$, an undeformed point has $\boldsymbol{C} = \boldsymbol{I}$.

Next, we need a constitutive law that links the deformation gradient to a "flux", which in mechanical problems we refer to as the **stress**. Heat flux may be defined as energy flow per oriented area,

$$q_i = \frac{\mathrm{d}Q}{\mathrm{d}A_i} \qquad \text{with} \qquad \mathrm{d}\boldsymbol{A} = \boldsymbol{N}\,\mathrm{d}A \quad \text{and} \quad \|\boldsymbol{N}\| = 1. \tag{A.34}$$

Analogously, stresses are force vector per oriented area:

$$P_{iJ} = \frac{\mathrm{d}F_i}{\mathrm{d}A_J}, \tag{A.35}$$

which defines the so-called **First Piola-Kirchhoff** (1st PK) **stress tensor**. This is a second-order tensor that captures the force components on any oriented area into all d coordinate directions. The resulting **traction** vector on a particular *infinitesimal* area with unit normal vector $\boldsymbol{N}$ is

$$\boldsymbol{T} = \boldsymbol{P}\boldsymbol{N} \qquad \Leftrightarrow \qquad T_i = P_{iJ}N_J. \tag{A.36}$$

Notice that the thus defined tractions satisfy Newton's law of action and reaction since

$$\boldsymbol{T}(\boldsymbol{N}) = \boldsymbol{P}\boldsymbol{N} = -\boldsymbol{P}(-\boldsymbol{N}) = -\boldsymbol{T}(-\boldsymbol{N}), \tag{A.37}$$

which we know well from inner forces in undergraduate mechanics. The total force acting on a surface A is hence

$$\boldsymbol{F}_{\text{tot}} = \int_A \boldsymbol{T}\,\mathrm{d}S. \tag{A.38}$$

For a mechanical problem, the relevant balance laws are *conservation of linear momentum* and of *anguluar momentum*. As before, one can formulate those as macroscopic balance laws. For example, macroscopic linear momentum balance is nothing but the well-known equation $\boldsymbol{F}_{\text{tot}} = M\boldsymbol{A}$ (*sum of all forces equals mass times mean acceleration*). To derive the local balance law of a continuous body, note that external forces include both surface tractions $\boldsymbol{T}$ and *body forces* $\rho_0\boldsymbol{B}$ – using ρ_0 to denote the reference mass density. Overall, we obtain

$$\int_{\partial\Omega} \boldsymbol{T}\,\mathrm{d}S + \int_\Omega \rho_0\boldsymbol{B}\,\mathrm{d}V = \int_\Omega \rho_0\boldsymbol{A}\,\mathrm{d}V \qquad \Leftrightarrow \qquad \int_{\partial\Omega} \boldsymbol{P}\boldsymbol{N}\,\mathrm{d}S + \int_\Omega \rho_0\boldsymbol{B}\,\mathrm{d}V = \int_\Omega \rho_0\boldsymbol{A}\,\mathrm{d}V. \tag{A.39}$$

Practicing the divergence theorem once more, we see that

$$\int_{\partial\Omega} P_{iJ}N_J\,\mathrm{d}S = \int_\Omega P_{iJ,J}\,\mathrm{d}V \qquad \Rightarrow \qquad (\mathrm{Div}\,\boldsymbol{P})_i = P_{iJ,J}, \tag{A.40}$$

which defines the *divergence of a second-order tensor*, which is a vector. Note that we use a *capitol operator* "Div" as opposed to "div" to indicate differentiation with respect to the undeformed coordinates.

When we again exploit that the above balance law must hold for all subbodies $\omega \subset \Omega$, we arrive at the *local statement of linear momentum balance*:

$$\boxed{\mathrm{Div}\,\boldsymbol{P} + \rho_0\boldsymbol{B} = \rho_0\boldsymbol{A}} \qquad \Leftrightarrow \qquad \boxed{P_{iJ,J} + \rho_0 B_i = \rho_0 A_i} \tag{A.41}$$

Note that the special case of **quasistatics** assumes that inertial effects are negligible, so one solves the quasistatic linear momentum balance $\mathrm{Div}\,\boldsymbol{P} + \rho_0\boldsymbol{B} = \boldsymbol{0}$. Except for gravity or electro/magnetomechanics, body forces also vanish in most cases, so that one simply arrives at $\mathrm{Div}\,\boldsymbol{P} = \boldsymbol{0}$.

It is important to recall that stresses in finite deformations are not unique but we generally have different types of stress tensors. Above, we introduced the **first Piola-Kirchhoff stress tensor** $\boldsymbol{P}$, which implies *actual force per undeformed area*. Similarly, one can define the **Cauchy stress tensor** $\boldsymbol{\sigma}$, which denotes *actual force per deformed area*. The definition and link are given by

$$\sigma_{ij} = \frac{\mathrm{d}F_i}{\mathrm{d}a_j}, \qquad \boldsymbol{\sigma} = \frac{1}{J}\boldsymbol{P}\boldsymbol{F}^{\mathrm{T}}. \tag{A.42}$$

Later on, it will be helpful to link stresses and deformation to energy. If the stored mechanical energy is characterized by the **strain energy density** $W = W(\boldsymbol{F})$, then one can show that

$$\boldsymbol{P} = \frac{\partial W}{\partial \boldsymbol{F}} \qquad \Leftrightarrow \qquad P_{iJ} = \frac{\partial W}{\partial F_{iJ}}. \tag{A.43}$$

Without knowing much about tensor analysis, we may interpret the above as a derivative of the energy density with respect to each component of $\boldsymbol{F}$, yielding the corresponding component of $\boldsymbol{P}$. This concept can also be extended to introduce a *fourth-order tensor*, the **incremental tangent modulus tensor**

$$\mathbb{C} = \frac{\partial \boldsymbol{P}}{\partial \boldsymbol{F}} \qquad \Leftrightarrow \qquad \mathbb{C}_{iJkL} = \frac{\partial P_{iJ}}{\partial F_{kL}}, \tag{A.44}$$

for which each component of $\boldsymbol{P}$ is differentiated with respect to each component of $\boldsymbol{F}$. For further information on tensor analysis, see the tensor notes. Without further discussion, notice that a *stress-free* state implies that the energy attains an extremum.

Note that the dependence of W on $\boldsymbol{F}$ (and consequently the consitutive law between $\boldsymbol{P}$ and $\boldsymbol{F}$) is generally strongly nonlinear, which limits opportunities for closed-form analytical solutions. Also, for *material frame indifference* we must in fact have $W = W(\boldsymbol{C})$, but that is a technical detail of minor importance here.

A.3 Linearized kinematics

Whenever only small deformation is expected, the above framework can be significantly simplified by using **linearized kinematics**. To this end, we assume that $\|\mathrm{Grad}\,\boldsymbol{u}\| \ll 1$ ("*small strains*"). Note that in this case it does not make a significant difference if we differentiate with respect to x_i or X_I, so that one generally uses only lower-case indices for simplicity.

In small strains, the *displacement field* is the key field to be determined (rather than the deformation mapping), i.e., we seek $\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{x}, t)$.

Recall that

$$\boldsymbol{C} = \boldsymbol{F}\boldsymbol{F}^{\mathrm{T}} = (\boldsymbol{I}+\mathrm{Grad}\,\boldsymbol{u})(\boldsymbol{I}+\mathrm{Grad}\,\boldsymbol{u}^{\mathrm{T}}) = \boldsymbol{I}+\mathrm{Grad}\,\boldsymbol{u}+(\mathrm{Grad}\,\boldsymbol{u})^{\mathrm{T}}+(\mathrm{Grad}\,\boldsymbol{u})(\mathrm{Grad}\,\boldsymbol{u})^{\mathrm{T}}. \tag{A.45}$$

Now, the final term is dropped by a scaling argument ($\|\mathrm{Grad}\,\boldsymbol{u}\| \ll 1$). Therefore, we may introduce a kinematic relation like in the thermal problem:

$$\boldsymbol{\beta} = \mathrm{Grad}\,\boldsymbol{u}, \tag{A.46}$$

and all important local deformation information is encoded in $\boldsymbol{\beta}$. Like a temperature gradient causes heat flux, a displacement gradient causes stresses (if displacements are constant everywhere, the body is undergoing **rigid body translation** and does not produce any stresses).

To make sure we also do not pick up **rigid body rotation**, one introduces the infinitesimal **strain tensor**

$$\boldsymbol{\varepsilon} = \frac{1}{2}\left[\operatorname{Grad}\boldsymbol{u} + (\operatorname{Grad}\boldsymbol{u})^{\mathrm{T}}\right] = \frac{1}{2}\left(\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}^{\mathrm{T}}\right) \qquad \Leftrightarrow \qquad \varepsilon_{ij} = \frac{1}{2}\left(u_{i,j} + u_{j,i}\right). \tag{A.47}$$

Notice that, unlike in finite deformations, *no deformation* implies $\boldsymbol{\varepsilon} = \boldsymbol{0}$. Furthermore, by definition $\boldsymbol{\varepsilon}$ is symmetric since $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^{\mathrm{T}}$ (not like $\boldsymbol{F}$ which is asymmetric). The same applies for $\boldsymbol{\sigma}$ and $\boldsymbol{P}$ which are, respectively, symmetric and asymmetric.

As before, local deformation metrics are encoded into $\boldsymbol{\varepsilon}$. For example, volumetric deformation is characterized by the *trace* of $\boldsymbol{\varepsilon}$, viz.

$$\frac{\mathrm{d}v}{\mathrm{d}V} = 1 + \operatorname{tr}\boldsymbol{\varepsilon} = 1 + \varepsilon_{ii}, \tag{A.48}$$

while stretches in the three coordinate directions are given by $\varepsilon_{(ii)}$ (parentheses implying no summation over i) and angle changes are identified as $\gamma_{ij} = 2\varepsilon_{ij}$ with $i \neq j$.

In linearized kinematics, all three stress tensors coincide and one commonly uses only the Cauchy stress tensor $\boldsymbol{\sigma}$ to define the constitutive relation $\boldsymbol{\sigma} = \boldsymbol{\sigma}(\boldsymbol{\varepsilon})$. In the simplest case of **linear elasticity**, those are linearly linked via

$$\boldsymbol{\sigma} = \mathbb{C}\,\boldsymbol{\varepsilon} \qquad \Leftrightarrow \qquad \sigma_{ij} = \mathbb{C}_{ijkl}\varepsilon_{kl} \tag{A.49}$$

with a fourth-order **elasticity tensor** $\mathbb{C}$ linking each component of $\boldsymbol{\sigma}$ to those of $\boldsymbol{\varepsilon}$. Alternatively, we can again encode the constitutive response in a *strain energy density* $W = W(\boldsymbol{\varepsilon})$, which, e.g., for the case of linear elasticity reads

$$W = \frac{1}{2}\boldsymbol{\varepsilon}\cdot\mathbb{C}\boldsymbol{\varepsilon} \qquad \Leftrightarrow \qquad W = \frac{1}{2}\varepsilon_{ij}\mathbb{C}_{ijkl}\varepsilon_{kl}, \tag{A.50}$$

so that

$$\boldsymbol{\sigma} = \frac{\partial W}{\partial \boldsymbol{\varepsilon}}, \qquad \mathbb{C} = \frac{\partial \boldsymbol{\sigma}}{\partial \boldsymbol{\varepsilon}} = \frac{\partial^2 W}{\partial \boldsymbol{\varepsilon}\,\partial \boldsymbol{\varepsilon}}. \tag{A.51}$$

When taking derivatives with respect to $\boldsymbol{\varepsilon}$, caution is required since $\boldsymbol{\varepsilon}$ is symmetric, so $\varepsilon_{ij} = \varepsilon_{kl}$ and, consequently, derivatives with respect to ε_{ij} must also take into account those terms containing ε_{ji} (for $i \neq j$). Therefore, the derivative should always be computed as

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}\varepsilon_{ij}} = \frac{1}{2}\left(\frac{\partial}{\partial\varepsilon_{ij}} + \frac{\partial}{\partial\varepsilon_{ji}}\right).} \tag{A.52}$$

Alternatively, one may also use $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^{\mathrm{T}}$ and simply replace $\boldsymbol{\varepsilon} = \frac{1}{2}(\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}^{\mathrm{T}})$ before differentiating.

The traction vector on a surface with unit normal $\boldsymbol{n}$ now becomes $\boldsymbol{t} = \boldsymbol{\sigma}\boldsymbol{n}$.

The local statement of linear momentum balance in linearized kinematics is

$$\boxed{\operatorname{div}\boldsymbol{\sigma} + \rho\boldsymbol{b} = \rho\boldsymbol{a}} \qquad \Leftrightarrow \qquad \boxed{\sigma_{ij,j} + \rho b_i = \rho a_i} \tag{A.53}$$

where the small-strain versions of density, body force density and acceleration field were introduced as ρ, $\boldsymbol{b}$ and $\boldsymbol{a} = \ddot{\boldsymbol{u}}$, respectively.

In small strains, we can insert the kinematic and linear elastic constitutive relations as well as the definition of the acceleration field into linear momentum balance to obtain

$$(\mathbb{C}_{ijkl}\varepsilon_{kl})_{,j} + \rho b_i = \rho\ddot{u}_i \qquad \Leftrightarrow \qquad (\mathbb{C}_{ijkl}u_{k,l})_{,j} + \rho b_i = \rho\ddot{u}_i \tag{A.54}$$

and in case of a *homogeneous* body with $\mathbb{C}(\boldsymbol{x}) = \mathbb{C} = \text{const.}$ we finally arrive at

$$\mathbb{C}_{ijkl}u_{k,lj} + \rho b_i = \rho\ddot{u}_i, \tag{A.55}$$

which is known as **Navier's equation** to be solved for the unknown field $\boldsymbol{u}(\boldsymbol{x}, t)$.

Finally, the following will be helpful when implementing material models in our code. Note that we may use the relations

$$\varepsilon_{ij} = \frac{1}{2}(u_{i,j} + u_{j,i}) \qquad \text{and} \qquad F_{iJ} = \delta_{iJ} + u_{i,J} \tag{A.56}$$

to write (using the chain rule)

$$\frac{\partial W}{\partial u_{i,J}} = \frac{\partial W}{\partial F_{kL}}\frac{\partial F_{kL}}{\partial u_{i,J}} = P_{kL}\frac{\partial}{\partial u_{i,J}}(\delta_{kL} + u_{k,L}) = P_{kL}\delta_{ik}\delta_{JL} = P_{iJ} \tag{A.57}$$

and (exploiting the symmetry of $\boldsymbol{\sigma}$)

$$\frac{\partial W}{\partial u_{i,j}} = \frac{\partial W}{\partial \varepsilon_{kl}}\frac{\partial \varepsilon_{kl}}{\partial u_{i,j}} = \sigma_{kl}\frac{\partial}{\partial u_{i,j}}\frac{1}{2}(u_{k,l}+u_{l,k}) = \frac{1}{2}\sigma_{kl}(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}) = \frac{1}{2}(\sigma_{ij}+\sigma_{ji}) = \sigma_{ij}. \tag{A.58}$$

Therefore, we can use the alternative relations for our stresses

$$\boxed{P_{iJ} = \frac{\partial W}{\partial u_{i,J}}} \qquad \text{and} \qquad \boxed{\sigma_{ij} = \frac{\partial W}{\partial u_{i,j}}}. \tag{A.59}$$

The beauty in those relations is that the stress tensor definition is now *identical irrespective of whether we are working in linearized or finite kinematics.*

A.4 Initial boundary value problems (IBVPs)

So far, we have seen how partial differential equations govern the thermal and mechanical behavior of solid bodies (and, of course, those two can be coupled as well to describe the thermo-mechanical behavior of deformable bodies). In order to solve a problem, we need an **initial boundary value problem** (IBVP), which furnishes the above equations with appropriate **boundary conditions** (BCs) and **initial conditions** (ICs).

To this end, we subdivide the boundary $\partial\Omega$ of a body Ω into

- $\partial\Omega_D$ $\equiv$ **Dirichlet boundary**, prescribing the *primary field* ($\boldsymbol{\varphi}$, T, etc.):

$$\text{e.g.} \qquad \boldsymbol{u}(\boldsymbol{x}, t) = \hat{\boldsymbol{u}}(\boldsymbol{x}, t) \text{ on } \partial\Omega_D \qquad \text{or} \qquad T(\boldsymbol{x}, t) = \hat{T}(\boldsymbol{x}, t) \text{ on } \partial\Omega_D. \tag{A.60}$$

- $\partial\Omega_N$ $\equiv$ **Neumann boundary**, prescribing *derivatives of the primary field* ($\boldsymbol{F}$, $\boldsymbol{\beta}$, etc.):

$$\text{e.g.} \qquad \boldsymbol{t}(\boldsymbol{x}, t) = \boldsymbol{\sigma}(\boldsymbol{x}, t)\boldsymbol{n}(\boldsymbol{x}, t) = \hat{\boldsymbol{t}}(\boldsymbol{x}, t) \text{ on } \partial\Omega_N \qquad \text{or} \qquad q(\boldsymbol{x}, t) = \hat{q}(\boldsymbol{x}, t) \text{ on } \partial\Omega_N, \tag{A.61}$$

Note that we may generally assume that

$$\partial\Omega_D \cup \partial\Omega_N = \partial\Omega \qquad \text{and in most problems also} \qquad \partial\Omega_D \cap \partial\Omega_N = \emptyset. \tag{A.62}$$

In addition, all time-dependent problems require **initial conditions**, e.g.,

$$\begin{aligned} & T(\boldsymbol{x},0) = T_0(\boldsymbol{x}) \quad \forall\ \boldsymbol{X} \in \Omega, \\ \text{or} \qquad & \boldsymbol{\varphi}(\boldsymbol{X},0) = \boldsymbol{x}_0(\boldsymbol{X}) \quad \text{and} \quad \boldsymbol{V}(\boldsymbol{X},0) = \boldsymbol{V}_0(\boldsymbol{X}) \quad \forall\ \boldsymbol{X} \in \Omega. \end{aligned} \tag{A.63}$$

The number of required BCs/ICs depends on the **order of a PDE**, e.g.,

$$\rho c_v \dot{T} = \operatorname{div}(\boldsymbol{K} \operatorname{grad} T) + \rho s \tag{A.64}$$

is first-order in time and therefore requires one IC, e.g., $T(\boldsymbol{x},0) = T_0(\boldsymbol{x})$. It is second-order in space and hence requires BCs along all $\partial\Omega$ (e.g., two conditions per x and y coordinates).

In summary, we have governing PDEs supplemented by ICs and BCs, as required (e.g., quasistatic problems, of course, do not require any initial conditions). Those are to be solved for the primary fields (e.g., temperature T or displacements $\boldsymbol{u}$ or the deformation mapping $\boldsymbol{\varphi}$). Unfortunately, analytical solutions are hardly ever available – except for relatively simple problems involving simple geometries, simple material behavior, simple ICs/BCs. Therefore, realistic geometries, material models and/or ICs/BCs call for numerical techniques to obtain approximate solutions such as those obtained from the finite element method.

B Review of Finite Elements

For a comprehensive reference summarizing the concepts of finite elements (in small strains), see the lecture notes on *Computational Mechanics I – Intro to FEA*, which are available online.

Index