BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 00, Number 0, Pages 000–000
S 0273-0979(XX)0000-0

MATERIALS FROM MATHEMATICS

RICHARD D. JAMES

Abstract. I survey some examples of materials whose recent discovery was based in an essential way on mathematical ideas. The main idea concerns "compatibility" – the fitting together of the phases of a material. Some of the emerging materials have the ability to change heat directly into electricity, without the need of a separate electrical generator.

Contents

1. Mathematics and materials science

Metallurgy in the early 20^{th} century was mainly concerned with the understanding of the phase diagram of steel and its use in designing processing treatments. It blossomed into materials science at mid century. From the beginning mathematicians have been interested. A touchstone for the aficionados is von Neumann's one page discussion in 1952 of Cyril Stanley Smith's paper, "Grain shapes and other metallurgical applications of topology"[57], where he discovered the $n-6$ law for grain growth[1]. C. S. Smith replies, "The discussion of Dr. von Neumann is much appreciated, and his conclusions are as remarkable as they are nonobvious on first consideration of the problem."[2]. But the fascination of using mathematical reasoning to understand materials goes back much further. In his 1745 paper, "Physical

2010 *Mathematics Subject Classification.* Primary 74N05; Secondary 74N30, 74N20, 82B26.
This work was supported by ONR (N00014-14-1-0714), AFOSR (FA9550-15-1-0207), NSF (DMREF-1629026), and the MURI program (FA9550-18-0095, FA9550-16-1-0566).

[1]See also Mullins' reinterpretation [46].

[2]To which C. S. Smith adds, "It is greatly to be hoped that he, or some other mathematician, will be able to deduce similar relations in three dimensions...", a hope that would have to wait 55 years for fulfillment [43].

investigations on the smallest parts of matter" Euler [27] reasons about how, owing to the presence of elemental molecules, bodies of different material can exhibit different masses and differing responses to gravity.

Today, largely due to the extreme forms of nonlinearity encountered in the behavior of materials, mathematics and materials science enjoy a healthy interaction. The mutual respect for certainty is pleasing. Like some theorems, the discovery of a spectacular new material represents an unmistakable advance, not clouded by shades of meaning.

In this paper we survey some recent developments and open problems in a central subfield of materials science: phase transformations. More precisely, our discussion concerns the mathematical theory that underlies the synthesis of materials that undergo phase transformations. What elements does one use, in what proportion, and with what processing, to achieve unprecedented behavior? Our behavior of interest will concern the hysteresis and reversibility of phase transformations.

The line of research surveyed here draws on, and owes much to, the work of many mathematicians and materials scientists. The author would particularly like to acknowledge the critical contributions of J. Ball, K. Bhattacharya, X. Chen, S. Conti, J. Cui, I. Fonseca, G. Friesecke, D. Kinderlehrer, R. Kohn, A. Ludwig, M. Luskin, S. Müller, F. Otto, E. Quandt, N. Schryvers, H. Seiner, Y. Song, V. Srivastava, V. Šverák, I. Takeuchi, M. Wuttig, G. Zanzotto, J. Zhang and B. Zwicknagl.

2. Phase transformations, hysteresis and reversibility

The types of phase transformations we consider are called structural or martensitic transformations. These are solid-to-solid phase transformations in which there is a change of crystal structure. The simplest example is a cubic-to-tetragonal phase transformation. In this case the unit cell of the high temperature cubic phase spontaneously elongates (or shrinks) along one of the four-fold axes upon cooling to the transformation temperature θ_c, changing the cubic unit cell to a tetragonal one (Figure 2 below). By symmetry, there are three four-fold axes, and so three *variants* of the tetragonal phase. The high temperature, often high symmetry, phase is called *austenite* and the low temperature, low symmetry, phase *martensite*. The change from the cubic cell to one of the three tetragonal cells involves a deformation, but no diffusion, i.e., no switching of atom positions, so these transformations can happen quite fast. Also – and this is probably the feature of greatest interest in materials science – the electronic structure or bonding pattern can change drastically during the transformation, because these aspects are sensitive to the geometry of the unit cell. For this reason the two phases can have very different properties. For example, one phase can be a strong magnet while the other phase is nonmagnetic, a feature we will exploit below in Section 8.

For a given material the identification of a particular transformation temperature θ_c is an oversimplification. In fact, one has to cool the material to a temperature $\theta^- < \theta_c$ before transformation occurs, and similarly, upon heating, one has to heat the material to $\theta^+ > \theta_c$. (The meaning of θ_c then becomes unclear, and we shall return to this later.) This phenomenon is called *hysteresis* and is one of the main features we will discuss here. The difference $\theta^+ - \theta^-$ is a measure of the hysteresis. Even in transformations that have a big distortion, it can range from 100s of degrees C to 1 degree C. In cases that the transformation is not so abrupt people measure a property such as electrical resistance vs. temperature by steadily heating, then

steadily cooling, the material and they fit the resulting graph by a parallelogram (Figure 1). The width of the parallelogram is then a measure of the hysteresis.

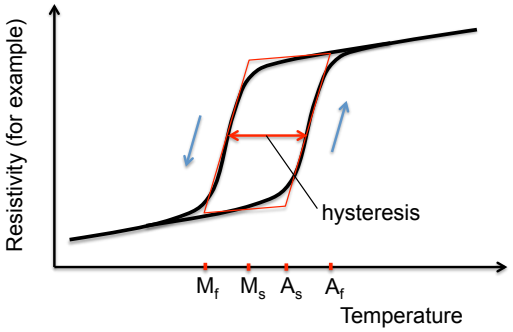


FIGURE 1. Hysteresis loop. The four temperatures M_f, M_s, A_s, A_f define an approximating parallelogram; the hysteresis is quantified by the value of $(A_s + A_f - M_s - M_f)/2$.

At first sight it is difficult to notice anything very different about a material with 100° C hysteresis vs. one with 1° C hysteresis. The first can be a bigger or smaller transformation than the second by any obvious criterion: size of the distortion (measured in various norms), size of the latent heat, stiffness. There is also no obvious correlation with the elements used in making the alloys, or their proportions. One might think that it would correlate with the volume change – bigger volume change means bigger hysteresis – but this is also not true in general ([22], Figure 4b).

Hysteresis is interesting partly because in applications, such as the one described in Section 8, it is synonymous with loss, and thus it is desirable to make it as small as possible. It is also fascinating from the viewpoint of mathematical theory. Usually, loss would be associated with the "damping terms", i.e., processes of viscoelasticity or viscosity, as embodied say in the Navier-Stokes equations. The expectation from scaling the energy equation of the Navier-Stokes equations is that, if one shears a fluid back and forth and measures some overall displacement vs. some overall force, the resulting hysteresis loop will shrink to zero as the rate of shearing (at fixed amplitude) gets lower and lower[3]. But the hysteresis loops in phase transformations, as far as we can tell, do not shrink to zero as the rate of change of temperature or force tends to zero. Rather, there is a limiting loop at zero rate. This is called *rate-independent hysteresis*. So, studying the effect of damping terms is evidently not the right idea.

Reversibility is a general term that is usually quantified experimentally by passing back and forth through the phase transformation many times, by say periodically changing the temperature, and measuring some property each cycle. A good property to measure is latent heat, since it is proportional to the amount of material that actually transforms. Solid-solid phase transformations have a latent heat that is absorbed on heating through the phase transformation (just like water boiling on the stove) and released on cooling. It is measured by calorimetry. A phase transformation is seen to be lacking reversibility if the latent heat decreases each cycle. Often in these cases nonreversibility is also seen more dramatically as a complete failure of the material after a certain number of cycles. A highly nonreversible phase transformation is the β to α transformation in the element tin, that occurs a little below room temperature. Transform a shiny bar of β-tin by cooling a little

[3]Here, the analogy to phase transformations is closer than it may seem. Phase transformations can often also be induced by applying a cyclic force, leading again to a hysteresis loop as in Figure 1, but with "temperature" replaced by force and "resistivity" by displacement. See Section 7.

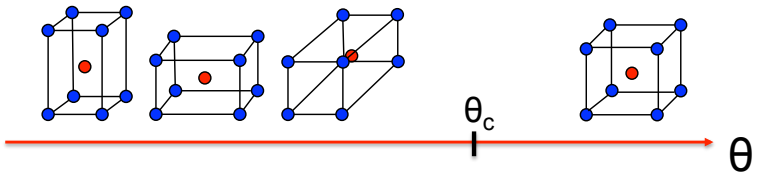


FIGURE 2. Cubic to tetragonal phase transformation. The red axis represents temperature θ. There is no diffusion, only distortion.

below room temperature. As it transforms, it turns into a pile of gray powder of α-tin.

3. Theory of phase transformations

We start with the simple cubic-to-tetragonal phase transformation mentioned above. A crystal (such as $BaTiO_3$, Fe_7Pd_3, In_4Tl, Mn_9Cu, $Ni_{65}Al_{35}$, Ni_2MnGa) transforms upon cooling somewhat below θ_c (because of hysteresis) by elongating along one of the four-fold cubic axes, to yield the three *variants of martensite*. To obtain the variants of martensite from the cubic structure, one applies linear transformations $U_1 = diag(\eta_2, \eta_1, \eta_1)$, $U_2 = diag(\eta_1, \eta_2, \eta_1)$, $U_3 = diag(\eta_1, \eta_1, \eta_2)$ to the cubic lattice, here written in an orthonormal cubic basis.

Two elementary points should be emphasized. First, the red and blue balls in Figure 2 indicate this is an ordered alloy, but, generally speaking, and consistent with the nonstoichiometric composition of some of the alloys given above, there can be a certain probability of finding an atom on a certain site, for example, a body center. For example, the nonstoichiometric alloys Mn_xCu_{1-x} $(0.8 < x < 0.95)$ all undergo cubic-to-tetragonal phase transformations, and all the compositions listed at the beginning of this section can be perturbed within limits. Second, let us number, left to right, the variants 1, 2, 3 in Figure 2. All the variants are exactly the same up to rigid rotation. For example, variant 1 can be rotated to look exactly like variant 2 by a rigid rotation $R_{\pi/2}$ of angle $\pi/2$. That does not mean that they are the same: what matters crucially here is the deformation, and these are of course different, $R_{\pi/2}U_1 \neq U_2$. However, for an appropriate rigid rotation $R \in \mathrm{SO}(3)$, the two deformations $y_1(x) = U_1 x$ and $y_2(x) = RU_2 x$ do agree on a lower dimensional set, and this observation will be relevant to our study of hysteresis.

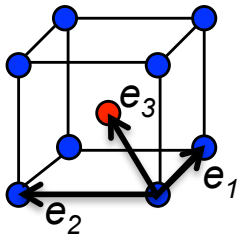


FIGURE 3. Lattice vectors for the BCC lattice (with so-called B2 ordering).

We begin with a lattice model of the phase transformation. In the simplest case we consider a Bravais lattice such as any one of the lattices shown in Figure 2. This is the set of points $\mathcal{L}(e_1, e_2, e_3) = \{\nu^1 e_1 + \nu^2 e_2 + \nu^3 e_3 : (\nu^1, \nu^2, \nu^3) \in \mathbb{Z}^3\}$ where e_1, e_2, e_3 are given linearly independent vectors in $\mathbb{R}^3$, called *lattice vectors*. In Figure 3 the lattice vectors can be chosen as the vectors from a blue atom to two nearest blue atoms, together with a vector to the red atom (Figure 3). Let $e_1 = \alpha\hat{e}_1, e_2 = \alpha\hat{e}_2$ where $\hat{e}_1, \hat{e}_2, \hat{e}_3 = \hat{e}_1 \times \hat{e}_2$ are orthonormal and $\alpha > 0$. We can consider various tetragonal Bravais lattices defined by lattice vectors $\alpha\hat{e}_1, \alpha\hat{e}_2, \alpha(\hat{e}_1 + \hat{e}_2 + \gamma\hat{e}_3)/2$, with $\gamma > 0$. (The constants α, γ that quantify the distances between atoms are called *lattice parameters*). The value $\gamma = 1$ gives

the BCC lattice of Figure 3. A famous observation of Bain is that there is exactly one other choice of $\gamma > 0$ in which the associated Bravais lattice is cubic, that being $\gamma = \sqrt{2}$, which gives the face-centered cubic lattice (FCC). In fact quite a few phase transformations can be viewed as perturbations of the BCC to FCC (or the reverse) transformation, obtained by moving γ from 1 to $\sqrt{2}$ and changing α a bit.

Many lattices of interest in phase transformations are not simply Bravais lattices. Rather, they are general periodic structures, i.e., the periodic extension of a finite number of atomic positions. These can be viewed as the union of a finite number of Bravais lattices, all made with the same lattice vectors, that is,

$$\mathcal{L}(e_1, e_2, e_3; x_1, \dots, x_M) = \{x_k + \mathcal{L}(e_1, e_2, e_3) : k = 1, \dots, M\}, \tag{3.1}$$

where the *base points* $x_1, \dots x_M$ are given points in $\mathbb{R}^3$. Conventionally, a description is chosen with smallest M, in which case we speak of e_1, e_2, e_3 as a set of *primitive* lattice vectors. A completely ordered lattice assigns a certain species to all positions with the same value of k. The lattice vectors e_1, e_2, e_3 can now also be interpreted as defining the periodicity. During a phase transformation in a complex lattice, the lattice vectors can change, the base points can change, and typically even the number M of base points changes. In its low temperature phase an important alloy discussed later $Zn_{45}Au_{30}Cu_{25}$ consists of $M = 18$ Bravais lattices.

Of course at positive temperature the atoms are vibrating about average positions, and phenomena such as the release of latent heat are intimately related to these vibrations. Nevertheless, it is convenient to use a kinematics of periodic lattices based say on these averaged positions.

It would seem to be an easy matter to decide which atom goes where during a phase transformation. Already, that decision has been made tacitly by drawing Figure 2 and assigning U_1, U_2, U_3. But in fact, especially with complex lattices, this is a nontrivial problem. Currently, there is no experimental method that can follow individual atoms during a phase transformation, i.e., the pathway, though there are some possibilities on the horizon. To determine the pathway, Bain [1] favored minimizing a measure of "strain" defined from mappings between the parent and transformed lattices, but he did not specify a norm. Mathematically, the difficulty can be appreciated by noticing that $\mathcal{L}(e_i) = \mathcal{L}(\mu_i^j e_j)$, where $(\mu_i^j) \in GL(3, \mathbb{Z})$. (Here, equality of two Bravais lattices means that they consist of the same infinite set of points, and $GL(3, \mathbb{Z})$ is the set of real 3×3 matrices of integers with determinant ±1. Also, we use the summation convention: $\mu_i^j e_j = \sum_{j=1}^3 \mu_i^j e_j$.) In fact, a classical theorem of crystallography (easily proved) says that $\mathcal{L}(e_1, e_2, e_3) = \mathcal{L}(f_1, f_2, f_3)$ for linearly independent vectors f_1, f_2, f_3 if and only if $f_i = \sum_{j=1}^3 \mu_i^j e_j$ for some $\mu \in GL(3, \mathbb{Z})$. So, many different choices of lattice vectors implies many possible pathways[4].

Many transformations between phases involve complex lattices. Empirically, in the notation of (3.1), the often accepted mechanism of transformation is that a Bravais sublattice of austenite with lattice vectors $\nu_i^j e_j$, with $\nu_i^j \in \mathbb{Z}^{3\times3}$ but with $\det \nu > 1$, is transformed to a primitive lattice of martensite[5]. As in [39], this again

[4]Recently, using a certain measure of strain, Muehlemann and Koumatos [39] prove that the Bain mechanism for BCC to FCC gives the smallest strain.

[5]Always, in this case $\det \nu$ has a value such that the volume of the unit cell associated to $\nu_i^j e_j$ is about the same as a primitive unit cell of the martensite phase.

gives rise to integer minimization problems for which rigorous algorithms can be devised that converge to a minimizer in a finite number of steps. An example using a particular measure of strain (different than [39]) is given in [17], and software can be found at `http://www.structrans.org`.

Really, determination of the pathway should be the province of first principles calculations, and many examples are being explored in this context [53, 68]. To describe a typical approach, consider the cubic to tetragonal transformation of Figure 2, with cubic lattice vectors e_1^c, e_2^c, e_3^c, such as those shown in Figure 3 (relabeled). Schematically, a typical procedure is the postulation of a one-parameter family of unit cells, say defined by linear transformations of the cubic unit cell $F(\xi), 0 \leq \xi \leq 1$, with $F(0) = I,\ F(1) = U_1$. Then, density functional theory calculations of periodic lattices defined by lattice vectors Fe_1^c, Fe_2^c, Fe_3^c are carried out, and optimized using the nudged elastic band method (see [25] for a mathematical description of this method in a different context). This method in principle gives the lowest saddle on a pathway $F(\xi)$ between cubic and tetragonal phases. But, on further contemplation, it misses a critical aspect of phase transformations which is central to this article: microstructure and compatibility! In fact, even in a near perfect single crystal, transformation never proceeds by a homogeneous deformation. Rather, the new phase nucleates and then grows. Inhomogeneous pathways[6] must have lower – likely much lower – saddles. The experimental evidence is that in many cases the highly inhomogeneous austenite/martensite interface (whose energy cannot be represented by a one-parameter ansatz of the type described here) represents the lowest saddle. It is a big challenge to have a first principles method that could cope with even the simplest microstructures, but well worth investigating. What is a few-parameter first principles anzatz that captures the austenite/martensite interface?

With this atomistic background we give a brief summary of a continuum theory of phase transformations [4, 9] we will use, while pointing out some deficiencies along the way. We do this first in the simplest case of transformations between Bravais lattices.

We assume that lattice vectors for the austenite are e_1^a, e_2^a, e_3^a and, for martensite, e_1^m, e_2^m, e_3^m. We wish to encompass also elastic deformations of both lattices, so we use the notation e_1, e_2, e_3 for generic lattice vectors. Eventually we will have to specify a domain, but for now we just assume preservation of orientation, $(e_1 \times e_2) \cdot e_3 > 0$. A general atomistic model will generate a free energy per unit volume once the lattice and temperature are prescribed, so we assume such free energy $\hat{\varphi}(e_1, e_2, e_3, \theta)$ defined for $(e_1 \times e_2) \cdot e_3 > 0$ and temperature $\theta > 0$.

This free energy $\hat{\varphi}(e_1, e_2, e_3, \theta)$ is subject to basic symmetries. We have frame-indifference, $\hat{\varphi}(Re_1, Re_2, Re_3, \theta) = \hat{\varphi}(e_1, e_2, e_3, \theta)$ for all $R \in \mathrm{SO}(3)$ and $\theta > 0$, and the condition that the free energy should only depend on the lattice $\mathcal{L}(e_1, e_2, e_3)$ and not otherwise on the lattice vectors: $\hat{\varphi}(\mu_1^j e_j, \mu_2^j e_j, \mu_3^j e_j, \theta) = \hat{\varphi}(e_1, e_2, e_3, \theta)$ for $(\mu_i^j) \in GL(3, \mathbb{Z})$ and $\theta > 0$.

The function $\hat{\varphi}(e_1, e_2, e_3, \theta)$ assigns a value of free energy to a perfect Bravais lattice $\mathcal{L}(e_1, e_2, e_3)$ at temperature θ. But we would like to treat more complex structures than perfect lattices, such as the microstructures shown in Figures 4, 9,

[6]Given a smooth $F(\xi)$, $F(0) = I, F(1) = U_1$, an interesting relevant mathematical problem that relies on the Cauchy-Born rule described below is to solve, under weak conditions of regularity, $\nabla y(x) = R(x)F(\xi(x))$ for $y : \Omega \to \mathbb{R}^3,\ R : \Omega \to \mathrm{SO}(3),\ \xi : \Omega \to (0, 1)$, where Ω is a domain in $\mathbb{R}^3$.

10. Locally, near most points, they look almost like perfect lattices. This suggests that we think of the austenite lattice as a reference lattice and define, for linear transformations F from $\mathbb{R}^3$ to $\mathbb{R}^3$ with $\det F > 0$,

$$\varphi(F,\theta) = (\det F)\,\hat{\varphi}(Fe_1^a, Fe_2^a, Fe_3^a, \theta). \tag{3.2}$$

(The presence of $\det F$ converts the free energy per volume of $\mathcal{L}(e_1,e_2,e_3)$ to a free energy per volume of the reference lattice $\mathcal{L}(e_1^a,e_2^a,e_3^a)$.) For a smooth mapping $y : \Omega \to \mathbb{R}^3$ the gradient ∇y is the local linear transformation, and therefore suggests a passage to continuum theory

$$\inf_{y\in\mathcal{A}} \int_\Omega \varphi(\nabla y(x),\theta)\,dx. \tag{3.3}$$

This cornerstone of the theory (3.2), (3.3) is called the Cauchy-Born rule [26]. It can be approached in a simple but rigorous way [13] via the "large body limit", e.g., the asymptotics as $\varepsilon \to 0$ of $y_\varepsilon : \Omega/\varepsilon \to \mathbb{R}^3$ given by $y_\varepsilon(x) = (1/\varepsilon)y(\varepsilon x)$, which has the feature of making $\nabla y_\varepsilon(x/\varepsilon)$ more and more constant, and therefore representing a more perfect lattice, on a bigger and bigger collection of atoms near x/ε as $\varepsilon \to 0$. A complete understanding of the Cauchy-Born rule likely involves difficulties beyond the already insanely difficult "crystallization problem" [62, 28]. Nevertheless, by making clever but realistic assumptions, many interesting studies shed light on its successes and failures, e.g., [30, 14, 67].

The unspecified $\mathcal{A}$ in (3.3) raises another issue. Normally in the calculus of variations, $\mathcal{A}$ would be chosen to match the growth conditions on φ. But the symmetry $\hat{\varphi}(\mu_1^j e_j, \mu_2^j e_j, \mu_3^j e_j, \theta) = \hat{\varphi}(e_1,e_2,e_3,\theta)$ rules out any of the usual growth conditions that would put finite energy deformations y in a reasonable Sobolev space. Note that a relevant choice of (μ_i^j) in $GL(3,\mathbb{Z})$ is, for arbitrary large $m \in \mathbb{Z}$,

$$(\mu_i^j) = \begin{pmatrix} 1 & m & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{3.4}$$

Of course, we could brutally assign strong growth conditions for φ, but then we would violate the symmetry, which is supposed to *preserve* the energy. Another disturbing fact is that a deformation gradient F satisfying $Fe_i^a = \mu_i^j e_i^a$ with μ_i^j as in (3.4), and say $m = 1$, if imposed on a crystal, is likely to cause either failure of the crystal or massive plastic deformation, phenomena that are not so relevant to the study of phase transformations.

These observations suggest a resolution due to Ericksen [26] and Pitteri [51]: cut down the symmetry $\hat{\varphi}(\mu_1^j e_j, \mu_2^j e_j, \mu_3^j e_j, \theta) = \hat{\varphi}(e_1,e_2,e_3,\theta)$ to a subgroup, excluding exactly those μ_i^j in $GL(3,\mathbb{Z})$ that would correspond to massive plastic deformation, and at the same time cut down the domain of φ to be invariant under exactly these symmetries (and frame-indifferent), and, with luck, include in the domain the tetragonal or other lower symmetry phases that are of interest.

Such a domain $\mathcal{D}$ can be found [4, 9, 52]. In the simplest case the answer is the following. The energy density $\varphi : \mathcal{D} \times (0,\infty)$ satisfies

$$R\mathcal{D}H = \mathcal{D} \quad \text{and} \quad \varphi(RFH,\theta) = \varphi(F,\theta), \quad \text{for } F \in \mathcal{D},\ R \in \mathrm{SO}(3),\ H \in G^a, \tag{3.5}$$

where

$$G^a = \{Q \in \mathrm{SO}(3) : Qe_i^a = \mu_i^j e_j^a,\ i = 1,2,3, \text{ and for some } \mu_i^j \in GL(3,\mathbb{Z})\}. \tag{3.6}$$

Hence, as indicated by the notation, the austenite phase dominates the symmetry. From its definition G^a is a finite, crystallographic group of rotations, i.e., one of the 11 Laue groups.

These discrete and continuous symmetries imply an energy-well structure, that also is assumed to change with temperature in a way that is consistent with the phase transformation. A symmetric, positive-definite *transformation stretch matrix* U_1 is given, and its orbit under the symmetries (3.5), (3.6) is $\mathrm{SO}(3)U_1 \cup \cdots \cup \mathrm{SO}(3)U_n$. The free energy φ is assumed to be smooth on $\mathcal{D} \times (0, \infty)$ and a typical energy-well structure is

$$\begin{aligned} \theta &> \theta_c, \quad \varphi(\cdot, \theta) \text{ minimized on } \mathrm{SO}(3) \\ \theta &< \theta_c, \quad \varphi(\cdot, \theta) \text{ minimized on } \mathrm{SO}(3)U_1 \cup \cdots \cup \mathrm{SO}(3)U_n, \end{aligned} \tag{3.7}$$

where $\{U_1, \ldots, U_n\} = \{QU_1Q^T : Q \in G^a\}$. In fact, it can be seen that the Laue group of the martensite is $G^m = \{Q \in G^a : QU_1Q^T = U_1\}$, which also shows, by Lagrange's theorem, that $n =$ order G^a/order G^m. From a practical viewpoint the restriction $QU_1Q^T = U_1$ for all $Q \in G^m$ is very useful, since these groups are often easily known from an X-ray measurement, but direct determination of U_1 is hampered by the issue raised above about knowing where atoms go.

We shall use the theory in the form above, but in some cases applying it to lattices that are more complex than Bravais lattices. In the case of a general lattice of the form (3.1) one expects an atomistic free energy[7] $\hat{\varphi}(e_1, e_2, e_3, x_2 - x_1, x_3 - x_1, \ldots, x_m - x_1, \theta)$, together with a Cauchy-Born rule of the form

$$\varphi(F, s_1, \ldots, s_{m-1}, \theta) = \hat{\varphi}(Fe_1^a, Fe_2^a, Fe_3^a, s_1, s_2, \ldots, s_{m-1}, \theta), \tag{3.8}$$

with again F replaced at continuum level with $\nabla y(x)$, and $s_1(x), \ldots, s_{m-1}(x)$, $x \in \Omega$, as unconstrained functions. Under suitable hypotheses for the resulting problem in the calculus of variations, we could minimize out $s_1(x), \ldots, s_{m-1}(x)$, leading back to a theory somewhat like that given above, with potentially a significant lack of smoothness due to intersections of branches of minimizers. These possibilities are interesting, but it is disgraceful that, as of 2018, we do not have a complete theory of symmetry for the more general multilattice case[8] of (3.1).

For definiteness we will put $\varphi(I, \theta_c) = 0$ in this paper.

4. Microstructure and nonattainment

As can been seen by the many examples presented in the article by C. S. Smith [57] mentioned at the beginning of this paper, small bubbles of a soap froth disappear and big ones grow, and the grains of a polycrystalline metal coarsen over time. As in the simplest linear elliptic and parabolic equations, there is a strong tendency to simplify and smooth. Exactly the opposite often happens in a martensitic phase transformation. One begins with a uniform crystal of austenite and, upon cooling through the transformation, one gets a plethora of fine microstructures of martensite. The mathematical origins of the spontaneous formation of fine structure comprise a fascinating and ongoing chapter of nonlinear analysis that began with the work of L. C. Young [66, 65, 64]. For a broad overview see the article of S. Müller [45].

[7] The presence of the differences $x_i - x_1$ arises from the translation invariance of $\hat{\varphi}$.

[8] The state-of-the-art is the last chapter of Pitteri and Zanzotto [52], especially Section 11.7.

We should clarify the distinction between grains (which collectively are often also called "microstructure") and the microstructure resulting from a phase transformation. The theory presented above is for a single crystal, when in its austenite phase, modeled at atomic scale by the uniform lattice $\mathcal{L}(e_1^a, e_2^a, e_3^a)$. When it transforms, say by cooling, it forms microstructure, due to the tendency arising from energy minimization of the deformation $y : \Omega \to \mathbb{R}^3$ to have a gradient ∇y near the energy wells $\mathrm{SO}(3)U_1, \dots, \mathrm{SO}(3)U_n$. But often the materials (both metals and oxides) that undergo phase transformations are studied as polycrystals, with differently oriented grains. Then one has to rewrite the theory presented in Section 3 for a polycrystal, which is easy to do [11]. The "fighting between the grains" during transformation has consequences: single crystals and polycrystals of the same material do exhibit somewhat different macroscopic response. Usually, phase transformations in polycrystals are studied at lower temperatures, in which case the grains do not coarsen during normal time scales.

The simplest relevant example of the emergence of microstructure is the following. Let $A \neq B \in \mathbb{R}^{3\times 3}$, with $\mathrm{rank}(B - A) = 1$, i.e., $B - A = a \otimes n$. Make the drastic and unphysical simplification that the free energy density is smooth on $\mathbb{R}^{3\times 3}$, independent of temperature, and satisfies

$$\varphi(A) = \varphi(B) = 0 < \varphi(F), \quad F \notin \{A, B\}. \tag{4.1}$$

We have brutally omitted all the symmetries and retained only the structure of having energy wells, and then only two. However, we have made them "rank-1 connected", a feature which is shared by the tetragonal phase in the model described above: there exists $R \in \mathrm{SO}(3)$ such that $\mathrm{rank}(RU_2 - U_1) = 1$ (see Lemma 5.1 below). Assume φ satisfies the mild growth conditions of being bigger than a positive constant outside a sufficiently large ball $|F| > \rho > 0$, and also assume that $\Omega \subset \mathbb{R}^3$ is open, bounded and has a Lipschitz boundary. Let $0 < \lambda < 1$ and consider

$$\inf_{\substack{y \in W^{1,1}(\Omega,\mathbb{R}^3),\\ y(x) = (\lambda B + (1-\lambda)A)x,\ x \in \partial\Omega}} \int_\Omega \varphi(\nabla y(x))\, dx. \tag{4.2}$$

This example exhibits nonattainment of the infimum by a mechanism that is common in martensitic crystals, and the proof is simple but typical. We first show that the infimum is 0. We define the 1-periodic function $\chi_\lambda : \mathbb{R} \to \mathbb{R}$ by

$$\chi_\lambda(s) = \begin{cases} 1-\lambda, & i \le s < i+\lambda, \\ -\lambda & i+\lambda \le s < i+1, \end{cases} \qquad i \in \mathbb{Z}. \tag{4.3}$$

Note that the integral of χ_λ over one period is zero. Then, for $k = 1, 2, \dots$ we define

$$y^{(k)}(x) = (\lambda B + (1-\lambda)A)x + \left(\frac{1}{k}\int_0^{k(x\cdot n)} \chi_\lambda(s)ds\right)a \tag{4.4}$$

By differentiation $\nabla y^{(k)}(x)$ takes the values B and A on alternating layers of width λ and $1-\lambda$, and $y^{(k)}$ converges uniformly to the linear map $(\lambda B+(1-\lambda)A)x$ on $\mathbb{R}^3$ as $k \to \infty$. Thus, $y^{(k)}$ is uniformly close to satisfying the boundary conditions. It can be made to satisfy them exactly by introducing a smooth function $\psi_\varepsilon : \mathbb{R}^3 \to [0, 1]$,

depending on the small parameter $\varepsilon > 0$, and satisfying

$$\psi_\varepsilon(x) = \begin{cases} 1, & \Omega_\varepsilon, \\ 0, & \mathbb{R}^3 \setminus \Omega, \end{cases} \qquad |\nabla \psi_\varepsilon| < 2/\varepsilon, \tag{4.5}$$

where $\Omega_\varepsilon = \{x \in \Omega : \mathrm{dist}(x, \partial\Omega) > \varepsilon\}$. (One can replace Ω by a slightly smaller interior domain, choose ψ_ε to be proportional to the distance to the boundary of that domain (or 1), and then mollify, choosing the various scales appropriately.) Then, the function

$$y_{k,\varepsilon}(x) = \psi_\varepsilon(x) y^{(k)}(x) + (1 - \psi_\varepsilon(x))(\lambda B + (1-\lambda)A)x), \quad x \in \Omega, \tag{4.6}$$

satisfies the boundary conditions $y_{k,\varepsilon}(x) = (\lambda B + (1-\lambda)A)x$, $x \in \partial\Omega$. Also, $\nabla y_{k,\varepsilon}(x)$ takes the values A and B on Ω_ε and

$$\begin{aligned} |\nabla y_{k,\varepsilon}(x)| &= |\psi_\varepsilon\big(\nabla y^{(k)} - (\lambda B + (1-\lambda)A)\big) \\ &+ \big(y^{(k)}(x) - (\lambda B + (1-\lambda)A)x\big) \otimes \nabla\psi_\varepsilon| \\ &\leq |B - A| + \frac{2|a|}{k\varepsilon} \end{aligned} \tag{4.7}$$

Choosing, say, $\varepsilon = 1/k$ we have sequence with bounded gradient that satisfies the boundary conditions, whose gradient takes the values A and B on larger and larger fractions of Ω as $k \to \infty$. Thus, the infimum in (4.2) is zero.

The term "microstructure" refers to the the fact that the infimum of (4.2) is not attained. To show nonattainment, we note that by the growth conditions, any minimizer $\hat{y}$ would have to lie in $W^{1,\infty}(\Omega, \mathbb{R}^3)$ and give zero energy density,

$$\nabla \hat{y} \in \{A, B\} \quad a.e.\ \Omega. \tag{4.8}$$

Extend $\hat{y}$ to all of $\mathbb{R}^3$ by making $\hat{y}(x) = (\lambda B + (1-\lambda)A)x$ on $\mathbb{R}^3 \setminus \Omega$, and note that $\hat{y}$ is in $W^{1,\infty}(\mathbb{R}^3, \mathbb{R}^3)$. On Ω we can write

$$\nabla \hat{y}(x) = \eta(x)B + (1 - \eta(x))A = A + \eta(x) a \otimes n, \quad \text{where } \eta(x) \in \{\pm 1\}. \tag{4.9}$$

Thus $z(x) = \hat{y}(x) - Ax$ satisfies $\nabla z(x) n^\perp = 0$ on Ω and, in fact, on all of $\mathbb{R}^3$, and for all $n^\perp \cdot n = 0$. This we can integrate on $\mathbb{R}^3$ to get $z(x) = f(x \cdot n)$. The function f is completely determined by its value *outside* Ω, and therefore it must be $f(x) = \lambda(n \cdot x)a$. Hence,

$$\hat{y}(x) = Ax + z(x) = (A + \lambda a \otimes n)x = (\lambda B + (1-\lambda)A)x \tag{4.10}$$

Since $0 < \lambda < 1$, $\lambda B + (1-\lambda)A$ is in no case equal to A or B, and therefore we have reached a contradiction with (4.8), and therefore we have proved nonattainment.

This kind of example can be generalized to cases that satisfy the underlying symmetries [4, 12, 8, 42] but still, all known results of this type are quite special. In fact the main important examples of this article, and the concept of supercompatibility, illustrate the subtlety in trying to prove some kind of general result on non-attainment. Much more can be said, and in particular the Young measure is a beautifully simple tool to learn more about the structure of minimizing sequences without explicitly calculating them.

One is led by these results to study minimizing sequences, as well as minimizers, and to understand their relation to observed microstructures. The most important microstructure in martensite becomes an immediate target: the classic austenite/martensite interface. This is the microstructure shown in Figure 4. It is widely seen as the interface between austenite and martensite at the finest level. When first observed in detail in the 1950s, it was puzzling because, when

people measured the normal m to the interface (by viewing one such interface on two inclined free surfaces, such as near a corner), they found that it was, in the language of materials science, "irrational". That is, when m was expressed in the orthonormal cubic[9] basis of austenite, the components were not small integers. How could such a beautifully planar interface be noncrystallographic, i.e., not apparently have any relation to the adjacent crystal lattice of austenite?

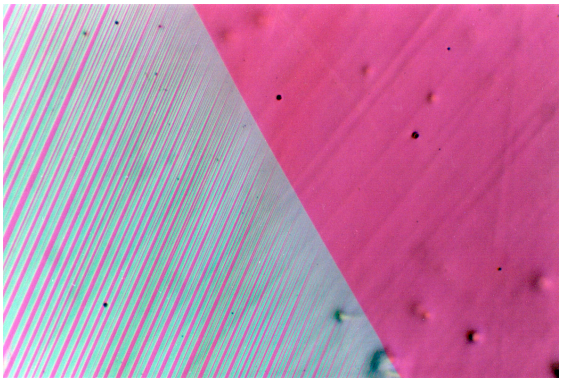


FIGURE 4. Austenite-martensite interface in $Cu_{69.5}Al_{27}Ni_{3.5}$. The banded structure on the left is martensite, and the uniform phase on the right is austenite (Courtesy of C. Chu).

The study of minimizing sequences associated to the austenite/martensite interface explains this irrationality and, quantitatively, the observed normal m, as well as other features of Figure 4. Given the minimizing sequence constructed above, it is very easy to find a related one modeling the austenite/martensite interface. Consider an energy density of the type given in (3.5)-(3.7), and choose A, B from the martensite wells, e.g. $A = \hat{R}U_2$ and $B = U_1$, with $\hat{R} \in \mathrm{SO}(3)$. Here we are using the freedom of an overall rotation of, say, Figure 4 to omit a possible rotation matrix in front of U_1. In fact, as already mentioned, it will be seen from results presented in Section 6 that there is often a choice of $\hat{R}$ such that

$$\hat{R}U_2 - U_1 = a \otimes n, \tag{4.11}$$

$a, n \in \mathbb{R}^3$. Equation (4.11) is called the *twinning equation*. We assume (4.11), choose $A = U_1$ and $B = \hat{R}U_2$ and repeat the construction (4.3)-(4.4) verbatim. The steps (4.5) to (4.6) can also be repeated, except now, say,

$$\psi_\varepsilon(x) = \begin{cases} 1, & x \cdot m < -\varepsilon, \\ 0, & x \cdot m > 0, \end{cases} \qquad |\nabla \psi_\varepsilon| < 2/\varepsilon, \tag{4.12}$$

where, without loss of generality, we have put the origin on the austenite-martensite interface. On the austenite $x \cdot m > 0$ (red in Figure 4)) we choose an arbitrary deformation gradient from the austenite well, $R \in \mathrm{SO}(3)$, so (4.6), (4.7) are replaced by

$$\begin{aligned} y_{k,\varepsilon}(x) &= \psi_\varepsilon(x) y^{(k)}(x) + (1 - \psi_\varepsilon(x)) Rx, \\ |\nabla y_{k,\varepsilon}(x)| &= |\psi_\varepsilon(\nabla y^{(k)} - R) + ((\lambda B + (1-\lambda)A)x - Rx) \otimes \nabla \psi_\varepsilon|, \end{aligned} \tag{4.13}$$

but now, in the crucial last term, Rx has only to agree approximately with $(\lambda B + (1-\lambda)A)x$ when $-\varepsilon < x \cdot m < 0$. This gives the sufficient (and necessary, for any reasonable choice of ψ_ε) condition

$$R^T(\lambda B + (1-\lambda)A) = I + b \otimes m. \tag{4.14}$$

[9]In many cases, including that of Figure 4, the austenite is cubic

Rewriting (4.14) and (4.11) in a common notation and, without loss of generality, replacing $R^T \in \mathrm{SO}(3)$, by R, we obtain the equations of the *crystallographic theory of martensite*:

$$\hat{R}U_2 - U_1 = a \otimes n, \quad R(\lambda \hat{R}U_2 + (1-\lambda)U_1) = I + b \otimes m \tag{4.15}$$

Since (4.11) will be solved below, it is easiest to think of $\hat{R}, U_1, U_2$ as given, consistent with the first of (4.15), and so the unknowns are the volume fraction λ, $R \in \mathrm{SO}(3)$ and $b, m \in \mathbb{R}^3$. As noted, m usually comes out irrational in the sense of materials science. Now the reason for the "irrationality" is clear: (4.15) embodies a geometric as well as a crystallographic restriction. Using (4.11), (4.15) can also be written

$$\hat{R}U_2 - U_1 = a \otimes n, \quad R(U_1 + \lambda a \otimes n) = I + b \otimes m \tag{4.16}$$

The solutions $(R, \lambda, b \otimes m)$ of (4.16) have been checked against measured values of these quantities many times, with amazing success [63].

This is one kind of minimizing sequence, but there are many more. If one looks at almost any martensitic microstructure, one can guess a minimizing sequence or, in some cases, a minimizer, and one can learn about this microstructure. One can get very good at this kind of guessing, then filling in many details by rather simple calculations, and it is certainly very informative. However, it is not really predictive of the microstructure that will result under such-and-such conditions.

The austenite-martensite interface, as for example shown in Figure 4, is produced by cooling a free crystal (no loads). Thus, nominally, the associated minimization problem is (3.3) with no imposed boundary conditions and $\theta = \theta_c$. But this minimization problem has a much simpler minimizer: $y(x) = x$, i.e., all austenite. Thus, one can say that there exists a minimizer, but *Nature prefers a minimizing sequence*, Figure 4. Of course, somehow, during cooling, the material has to go from austenite to martensite, and the crystallographic theory of martensite provides an obvious low energy pathway ("a zero limiting energy pathway"). But, at least in the normal pursuit of solutions of the calculus of variations, if one finds an absolute minimizer, one usually quits and does not then seek a minimizing sequence, especially one whose weak limit is not a minimizer!

There are several features of Figure 4 that are not predicted by the crystallographic theory of martensite. Obviously, the bands on the left are not infinitely fine. There are also more bands near the interface than far from the interface – a branching phenomenon can be noticed, and this is quite common in martensites. By careful examination, 4-5 generations of branching can be seen in Figure 4. However, the measured the volume fraction λ predicted by the crystallographic theory[10] agrees closely (e.g., within 2-3%) with that measured on Figure 4 either close or far from the austenite interface. These observations are believed to be a consequence of a small regularization which has been studied mathematically in simpler models (see [36] for a survey, and [54, 16, 15]). The consequences of this regularization will be critical for our understanding of hysteresis. We return to this in Section 5.

[10] suitably modified to represent the volume fraction of bands on the image $y(\Omega)$ and measured on a window that contains at least about 10 bands.

5. Hysteresis

What causes hysteresis? There are diverse thoughts about hysteresis in a large literature[11], some of which are in fact inconsistent with the observation of rate-independent hysteresis. Ideas also include the pinning of interfaces by defects [61] and spin flips in an Ising model mediated by disorder [56]. There is also a deep mathematical literature on the *modeling* of hysteresis, which is not so much concerned with its origins, but rather with the accurate simulation of hysteresis loops [44]. These kinds of simulations have been used in industry, for example to quantify whether the hysteresis of a new batch of material is the same as that of the last batch, but they are not so useful for the discovery of materials.

The concept we will explore is related to the ubiquitous austenite/martensite interface Figure 4, the crystallographic theory of martensite, and regularized models. As explained in Section 4, the austenite/martensite interface is a low energy structure modeled by a minimizing sequence, leading to the algebraic problem (4.16). We also noted some discrepancies between theory and experiment, notably the nonzero scale of the twin bands on the left of Figure 4. This is believed to be related to a small interfacial energy on the boundaries of these bands. Finer and finer bands (at fixed volume fraction λ) reduce the elastic energy in the transition layer near the austenite/martensite interface, but have more and more interfacial energy. Coarse bands, on the other hand, have little interfacial energy but unacceptable elastic energy in the transition layer. People believe that what you see in Figure 4 is a compromise between these two tendencies, that in fact is captured by regularized models [36, 38, 37]. The story is subtle: many regularized models assert that the transition layer is delocalized, and is accompanied by branching of the bands as seen in Figure 4. "Delocalized"' means that the elastic energy arises not from a layer parallel to the austenite/martensite interface, as in (4.12), but rather arises from the martensite bands having normals slightly perturbed away from n, due to the branching. In martensites in hard materials, one pays a big energetic penalty for perturbing n. More on this below.

Figure 5. Schematic nucleus of martensite, bounded by two austenite/martensite interfaces.

So, there are two sources of energy, both positive, that are missed by the minimizing sequence constructed above: the interfacial energy on the boundaries of the martensite bands, and the elastic energy in the transition layer. We can imagine that this sets up an *energy barrier*. On cooling from the austenite phase, as soon as a nucleus of martensite forms (such as that seen in Figure 5), it is accompanied by extra interfacial and transition-layer energy. For it to grow, a lowering of the temperature below θ_c is required, in order that the lowering of the martensite well with temperature compensates for the barrier. A similar process could happen upon heating. As soon as austenite forms, it must be accompanied by transition-layer and interfacial energy, and another barrier is set up, requiring a increase of temperature to above θ_c. Hence, hysteresis.

[11] see Section 1.4 of [10] for a broad summary

Mathematically, the identification of energy barriers is not part of the usual calculus of variations. Linear stability analysis, e.g., the study of the loss of positivity of the second variation as a parameter θ is varied, identifies energy barriers, but clearly the study of the second variation would miss the barrier identified here. The "nucleus" shown in Figure 5 is missed by linearization about the undistorted austenite phase: it is a large, localized disturbance. More precisely, at the measured temperature $\theta < \theta_c$, on the shoulder of the hysteresis loop (Figure 1) where the martensite begins to grow, the second variation at the ambient austenite phase is strongly positive-definite. Promising emerging methods [34], [35] not based on linearization exploit the fact that the nucleus has small support, together with the energy well structure and compatibility. Otherwise, the only known method is: make a clever ansatz. Numerically, one can collect a zoo of ansatzes, but, without insight from experiment, this does not seem hopeful. Fortunately, in the present case, one can make a reasonable ansatz, Figure 5.

This ansatz has been studied in some detail [70, 71]. First, one notes that there are no singularities at the ends of the needle and the curvature of the interface plays an insignificant role: one may as well study the energy per unit area of two parallel austenite/martensite interfaces, as the distance w between them is varied. Of course, one needs to include interfacial energy per unit area on the boundaries of the bands as well as the transition-layer energy, both of which can be included in the context of an ansatz. Sure enough, there is a barrier when $\theta < \theta_c$. When w is small, the interfacial/transition-layer energies dominate leading to an increase of energy with w. When w is large, the bulk energy dominates, due to $\varphi(U_1, \theta) = \varphi(\hat{R}U_2, \theta) < \varphi(R, \theta)$ in the notation of (4.11)-(4.16), leading to a linear decrease of energy with w. The balancing of elastic and interfacial energy is quite subtle, see [71].

If this barrier is indeed responsible for hysteresis, there is obvious experimental test. It is related to an elementary lemma.

Lemma 5.1. [3] *Let the symmetric* 3×3 *matrix have ordered eigenvalues* $0 < \lambda_1 \leq \lambda_2 \leq \lambda_3$ *and corresponding orthonormal eigenvectors* e_1, e_2, e_3. *A necessary and sufficient condition that there exists* $Q \in SO(3)$ *such that* $QU_1 - I = b \otimes m$ *is* $\lambda_2 = 1$. *The solutions are expressible in the form*

$$\begin{aligned} b &= \rho \left(\sqrt{\frac{\lambda_3(1-\lambda_1)}{\lambda_3 - \lambda_1}}\, e_1 + \sigma \sqrt{\frac{\lambda_1(\lambda_3 - 1)}{\lambda_3 - \lambda_1}}\, e_3 \right), \\ m &= \frac{1}{\rho} \left(\frac{\sqrt{\lambda_3} - \sqrt{\lambda_1}}{\sqrt{\lambda_3 - \lambda_1}} \right) \left(-\sqrt{1-\lambda_1}\, e_1 + \sigma \sqrt{\lambda_3 - 1}\, e_3 \right), \end{aligned} \tag{5.1}$$

where $\sigma \in \pm 1$ *and* $\rho \neq 0$.

Proof. Operate $(QU_1)^T(QU_1) = U_1^2 = (I + m \otimes b)(I + b \otimes m)$ on $b \times m$ to see that one eigenvalue of U_1 is 1, then write $U_1^2 = (I + m \otimes b)(I + b \otimes m)$ in the eigenvector basis of U_1 to show it is the middle eigenvalue, and also that $\lambda_2 = 1$ is sufficient.

The experimental test is at hand, when one realizes that $QU_1 - I = b \otimes m$ is necessary and sufficient that there is a continuous function $y : \Omega \to \mathbb{R}^3$ taking the values QU_1 (martensite) and I (austenite). (Note also that the matrices $U_1, \ldots, U_n$ all have the same eigenvalues.) For examples, see the pictures at $f = 0, 1$ in Figure 7. So, if $\lambda_2 = 1$, there is no need of the fine bands, no need of the transition layer.

We can transform by passing a single plane separating austenite and martensite through the material, at least for a single crystal. In principle, there could be a small interfacial energy on this plane but, essentially, the barrier is gone. This is also seen by specialization of the results in [70, 71] to the case $\lambda_2 = 1$.

How can this be an experimental test? The matrix U_1 (and therefore $U_2, \dots U_n$) are properties of the material. While it is difficult to change the symmetries (3.5) and (3.6) of a material, the transformation stretch matrix U_1 does change with composition. So, start with a material having U_1 with a middle eigenvalue reasonably close to 1, and tune the composition to make λ_2 exactly 1. When this was done[12], the results were astonishing. Later, people who do combinatorial synthesis[13] tried this and, in fact, these studies very much highlighted the importance of "combi-methods". A collection of measurements done with both combi and bulk methods is shown in Figure 6. Each marker is a different alloy. One can see that the hysteresis can be reduced to near zero by tuning the composition to make $\lambda_2 = 1$, some of these alloys being exceptional. The perfect interfaces of "$\lambda_2 = 1$ alloys" have been seen in transmission electron microscopy [24], and in fact the angle in Q and b, m can be measured from the micrographs, showing nice agreement with Lemma 5.1. This is not at all restricted to NiTiX alloys. Today, there is a lively ongoing effort to make new low hysteresis alloys this way.

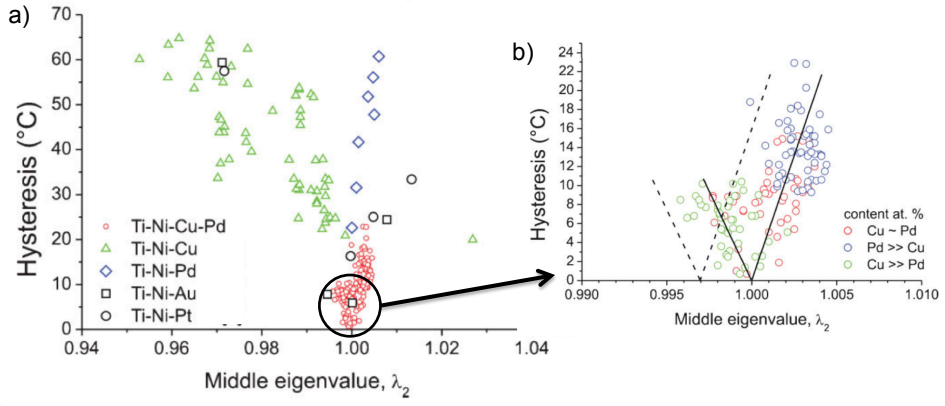


FIGURE 6. Measured width of the hysteresis of alloys in the NiTiX system ([22, 70, 69]). Each marker is a different alloy. Panel b) is a close-up of a) near $\lambda_2 = 1$. Note that the width of the hysteresis loop can be reduced to near zero.

There is another interesting consequence that is revealed by combi methods [22]. The X-ray measurement that gives values of λ_2 for the many alloys, also gives measurements of λ_1, λ_3. The product $\lambda_1\lambda_2\lambda_3$ is the volume ratio of the two phases. If $\lambda_1\lambda_2\lambda_3 \neq 1$ and the new phase nucleates on the interior, it grows up in a hole of the wrong volume. Changing the volume of the hard material requires a lot of energy, and one would think this would set up an energy barrier. However, when hysteresis is plotted vs. $\lambda_1\lambda_2\lambda_3$ for combinatorial library, unlike in Figure 6, there is no clear correlation. In fact, some of the alloys with the biggest hysteresis and some with the smallest hysteresis have $\lambda_1\lambda_2\lambda_3 = 1$. Perhaps nucleation always

[12]first, primitively, in the lab of the author

[13]notably I. Takeuchi and A. Ludwig and their groups

occurs from the boundary? This is not so easy to check, as few martensites are transparent.

The NiTiX system is the most intensely studied transforming material system. In recent years it has been the most intensely studied system among all metal alloys [49]. People have made many thousands of NiTiX alloys and characterized them. How did they miss the obvious sharp drop of Figure 6? The reason is that λ_2 is extremely sensitive to composition (see Fig. 1 of [18]). When people previously made their alloy series, they always jumped over the composition at which $\lambda_2 = 1$. Without a theory, it can be hard to find a singularity.

6. Supercompatibility

We can appreciate from these results on hysteresis, that a) for some aspects of material behavior, there really is a difference between minimizing sequences and minimizers in the problem (3.3), b) regularized models reveal this difference quite clearly, and c) degeneracies like $\lambda_2 = 1$ have an important effect on hysteretic behavior. By satisfying $\lambda_2 = 1$ we also disrupt the balance between bulk and interfacial energy. From a practical viewpoint, in a "$\lambda_2 = 1$ material" we might still expect to see a compatible interface in a 100nm (or even 10nm) crystal, but if λ_2 is not extremely close to 1 we are unlikely to see an austenite/martensite interface at these scales.

It seems unlikely that there is any further lowering of the energy barrier between austenite and martensite that is possible than by having a perfect, untwinned austenite/martensite interface implied by $\lambda_2 = 1$. But there is the potential to find degeneracies that allow many possible low energy ways to mix austenite and martensite. Such conditions could remove barriers that form when, say, several $\lambda_2 = 1$ interfaces are forced to meet, such as at a grain boundary. One such degeneracy is embodied in the *cofactor conditions* [32].

The cofactor conditions are degeneracy conditions of the crystallographic theory of martensite, which we have reduced to (4.16). We suppose as above that the twinning equation $\hat{R}U_2 - U_1 = a \otimes n$ has been satisfied. To solve this for $\hat{R} \in \mathrm{SO}(3)$, $a, n \in \mathbb{R}^3$ one can recast it in the form of Lemma 5.1 or use the Proposition 12 from [18]. This proposition states, under the conditions that the 3×3 matrix U_1 is symmetric and positive-definite and $U_2 = QU_1Q^T$ for some $Q \in \mathrm{SO}(3)$ (all of which are assumed above), there is a solution $\hat{R} \in \mathrm{SO}(3)$, $a, n \in \mathbb{R}^3$ of $\hat{R}U_2 - U_1 = a \otimes n$ if and only if there is $\hat{e} \in \mathbb{R}^3, |\hat{e}| = 1$, such that

$$U_2 = (-I + 2\,\hat{e} \otimes \hat{e})U_1(-I + 2\,\hat{e} \otimes \hat{e}). \tag{6.1}$$

Formulas for $\hat{e}, a, n$ are given in [18] (see (10) and (A.1)-(A.6) there). We note that there are usually lots of 180 degree rotations $(-I + 2\,\hat{e} \otimes \hat{e})$ in the Laue group G^a, so lots of pairs of matrices U_i and U_j satisfy the twinning equation, and can be the basis of constructing austenite/martensite interfaces.

Given $\hat{R}, a, n$ we now turn to the second condition $R(U_1+\lambda a\otimes n) = I+b\otimes m$ of the crystallographic theory of martensite (4.16), to be solved for $R \in \mathrm{SO}(3)$, $0 \le \lambda \le 1$, and $b, m \in \mathbb{R}^3$. Following [3] we eliminate R by calculating $(I + b \otimes m)^T(I + b \otimes m)$ to get the necessary condition

$$G(\lambda) := (U_1 + \lambda n \otimes a)(U_1 + \lambda a \otimes n) = (I + m \otimes b)(I + b \otimes m). \tag{6.2}$$

This condition is sufficient for the existence of $R \in \mathrm{SO}(3)$ if $\det(I + b \otimes m) = 1 + b \cdot m > 0$. This follows from the polar decomposition theorem and the fact that $\det(\cdot)$ is rank-1 affine, and so $\det(U_1 + \lambda a \otimes n) = \det(\lambda \hat{R} U_2 + (1-\lambda) U_1) = \det U_1 > 0$. One can further notice that (6.2) is related to Lemma 5.1, and therefore is solvable with $1 + b \cdot m > 0$ if and only if the $0 \leq \lambda \leq 1$ can be chosen so that the middle eigenvalue of $G(\lambda)$ is 1.

Hence, we seek $0 \leq \lambda \leq 1$ such that the middle eigenvalue of $G(\lambda)$ is 1. If so, necessarily, $\det(G(\lambda) - I) = 0$. This looks like a 6th order polynomial in λ, but, due again to the fact that $\det(\cdot)$ is rank-one affine, it is actually quadratic and symmetric about 1/2. Thus, aside from the issue of whether the resulting eigenvalue = 1 of $G(\lambda)$ is the middle one (an inequality given below), the crystallographic theory of martensite reduces to the question of whether a particular symmetric, quadratic function has roots $0 \leq \lambda^* \leq 1$ and $1 - \lambda^*$. In fact, if one uses the matrices U_1 and U_2 (appropriately selected!) for the alloy $Cu_{69.5}Al_{27}Ni_{3.5}$ shown in Figure 4, one does in fact have such roots and the resulting solution agrees nicely with Figure 4, with the qualifications mentioned above.

The quantity $\det(G(\lambda) - I)$ can of course be evaluated for any material that undergoes a phase transformation and has an energy well structure with the symmetries of Section 3. Materials (such as tin) with an appreciable value of $|U_1 - I|$ for which $\det(G(\lambda) - I)$ has no roots are usually not reversible.

For a "$\lambda_2 = 1$ material" as discussed in Section 5, necessarily we have solutions of the crystallographic theory because the second of (4.16) is satisfied at $\lambda = 0, 1$. That is, the symmetric, quadratic function $\det(G(\lambda) - I)$ satisfies $\det(G(0) - I) = \det(G(1) - I) = 0$, and, of course, the roots $0, 1$ give middle eigenvalues of $G(\lambda)$.

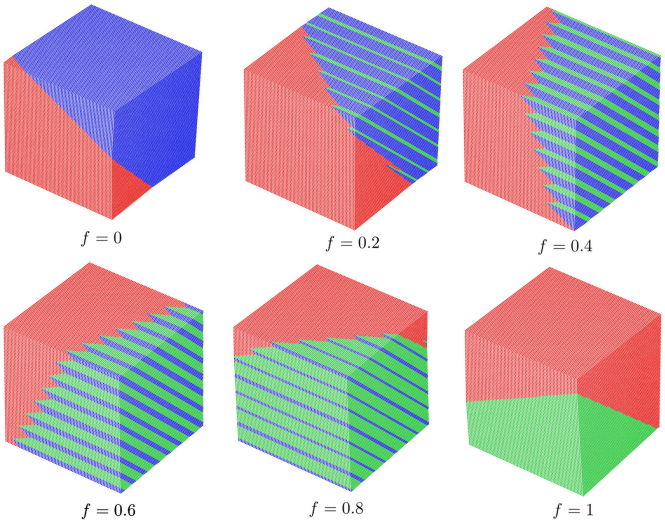
In this framework an obvious degeneracy presents itself: the function $\det(G(\lambda) - I)$ could be *identically zero*. Then, assuming the roots obtained give *middle* eigenvalues of $G(\lambda)$, we would have solutions of the crystallographic theory for every $0 \leq \lambda \leq 1$. A quadratic function, symmetric about 1/2, is identically zero if and only if its value at 0 and its derivative at 0 are zero. These two conditions, together with an inequality that is necessary and sufficient that the eigenvalues = 1 obtained are the middle ones, are the *cofactor conditions* [32, 18]:

$$\lambda_2 = 1, \quad a \cdot U_1 cof(U_1^2 - I) n = 0, \quad \mathrm{tr} U_1^2 - \det U_1^2 - \frac{|a|^2 |n|^2}{4} \geq 2. \tag{6.3}$$

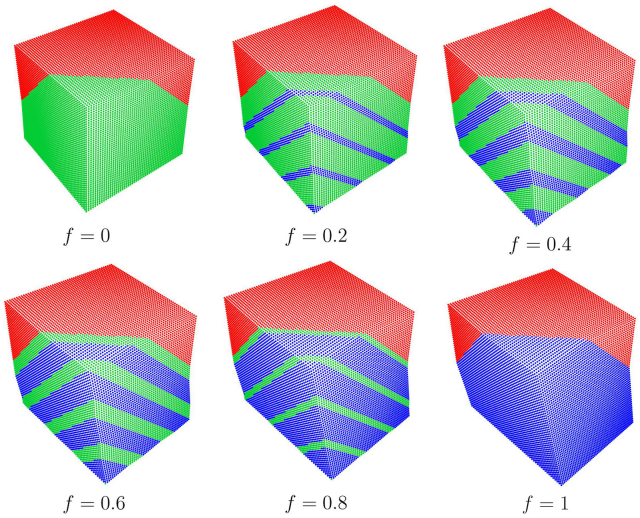
We review the known alloys that have been tuned to satisfy the cofactor conditions in Section 7.

The cofactor conditions depend on the "twin system" a, n. It is easily seen by operating $Q \ldots Q^T$ on (6.2), $Q \in G^a$, that its satisfaction for one twin system implies its satisfaction for other twin systems, and there can be further multiplicities of this type [18], depending on the symmetries.

The cofactor conditions imply a plethora of additional austenite/martensite interfaces modeled by minimizing sequences, but it is not obvious that they provide additional zero elastic energy structures beyond those guaranteed by $\lambda_2 = 1$. It seems from the results of Section 6 that elimination of the elastic energy altogether might be most important. But, as degeneracy gets piled on degeneracy, there can be other unexpected accidents, and that is the case here. Some but not all of these are collected in Figure 7. Briefly, to understand this figure, one needs to know that there is a classification of solutions of the twinning equation into Type I, Type II and Compound twins. (In microstructures with austenite and compound twins it



(a) austenite/martensite with Type I twins



(b) austenite/martensite with Type II twins

FIGURE 7. Zero elastic energy austenite/martensite interfaces possible under the cofactor conditions, from [18]. Red is austenite and blue/green are two variants of martensite. These pictures exhibit large deformations, zero elastic energy and perfect fitting of the phases, under continuous variation of the volume fraction f. For Type II twins the cofactor conditions imply that the twin boundaries are parallel to the austenite/martensite interface, which clearly makes for easy construction of zero elastic energy microstructures.

is not known if one can eliminate the elastic energy.) The details can be found in [18]. Not pictured here are zero elastic energy curved austenite/martensite interfaces, and zero elastic energy mechanisms for nucleation. There may well be other

families of zero energy microstructures, and a compete understanding is currently missing. It will be shown in Section 7 that unprecedented reversibility is seen in the two known alloys that closely satisfy the cofactor conditions.

We should mention that there is potentially a completely different interpretation of the cofactor conditions. This concerns the relaxed energy [40]. The relaxed energy is the limiting energy of the lowest energy minimizing sequence having a given weak limit. From a materials science viewpoint: fix the average deformation, find a (possibly complex) microstructure that minimizes the energy and has this given average deformation. That is, assume $1 < p < \infty$ and Ω bounded and open, with a Lipschitz boundary. Let $y \in W^{1,p}$ be given, and for sequences

$$y^{(j)} \rightharpoonup y \ \text{ in } L^p \quad \textit{and} \quad \nabla y^{(j)} \rightharpoonup \nabla y \ \text{ in } L^p \tag{6.4}$$

minimize the energy:

$$E_{macro}(y) = \inf_{\{y^{(j)}\}} \left\{ \liminf_{j\to\infty} \int_\Omega \varphi(\nabla y^{(j)}, \theta_c)\, dx \right\}. \tag{6.5}$$

It is known [40] (see also [29]) that if $\varphi(F, \theta_c) > c|F|^p$, $c > 0$, when $|F|$ is large, then E_{macro} is representable in terms of a macroscopic energy density $\tilde{\varphi}$ by

$$E_{macro}(y) = \int_\Omega \tilde{\varphi}(\nabla y(x))\, dx. \tag{6.6}$$

Properties of $\tilde{\varphi}$ are given by Kristensen [40]. Of interest here is its zero level set. Does it suddenly get larger when the cofactor conditions are satisfied? Naively, one would think "yes", since the new austenite/martensite interfaces, exhibited for example in Figure 7, should enlarge the flat region present on $\tilde{\varphi}$.

Our vague sense is rather that the presence of the zero elastic energy minimizers, which disrupts the delicate balance between elastic and interfacial energy, is perhaps most important consequence of the cofactor conditions. Therefore, by *supercompatibilty* we shall mean degeneracy conditions like $\lambda_2 = 1$ or the cofactor conditions that permit the phases to fit together with finite interfacial area and without stressed transition layers. Another interesting recent example is [21]. With a good collection of examples one can hope for a general theory of supercompatibility.

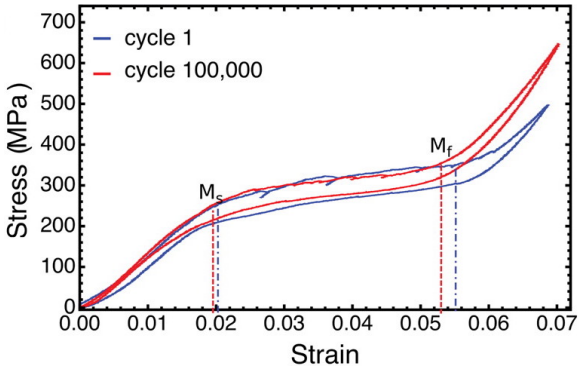
7. Reversibility

Two alloys have been found that accurately satisfy the cofactor conditions: $Zn_{45}Au_{30}Cu_{25}$ [58] and $Ti_{54.7}Ni_{30.7}Cu_{12.3}Co_{2.3}$ [19]. Briefly, they both exhibit exceptional reversibility of the phase transformation. They are quite different alloys both chemically and structurally: the ZnAuCu alloy undergoes a cubic to monoclinic transformation while the TiNiCuCo alloy undergoes a cubic to orthorhombic transformation. The ZnAuCu alloy was found by systematic alloy development: make a specimen starting from high purity elemental Zn, Au and Cu in the right proportions, check for changes of composition that may have occurred by loss to the environment during melting, measure accurately the lattice parameters of both phases by X-ray methods, calculate the quantities in (6.3) and repeat. After several specimens, one develops the relation between (6.3) and composition, from which satisfaction of (6.3) to high accuracy is possible. The TiNiCuCo alloy was made by thin[14] film sputtering methods, removal the substrate and polishing the boundary of the specimen. Both specimens were made to undergo standard heat treatments

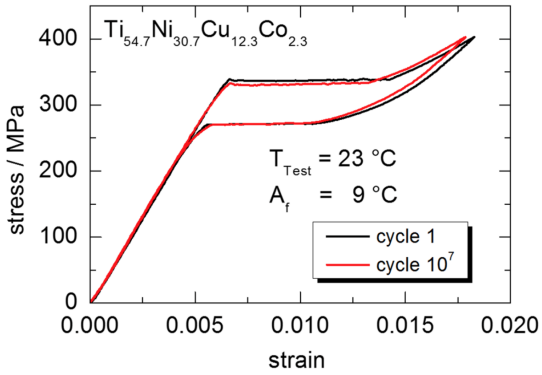
[14]actually quite thick, so it would be considered more like bulk material

after synthesis. Further information on synthesis and processing can be found in [19, 31].

To understand reversibility, a nice test to do is the *shape memory effect*: cool the material from austenite, deform the material in the martensite phase (which rearranges the variants) leaving the material with a large overall deformation, heat it up and it returns to its starting shape as the martensite transforms back to austenite. But this heating and cooling would take too long for these highly reversible alloys, i.e., for these materials such tests would take years.



(a) $Zn_{45}Au_{30}Cu_{25}$ [48, 58]



(b) $Ti_{54.7}Ni_{30.7}Cu_{12.3}Co_{2.3}$ [19]

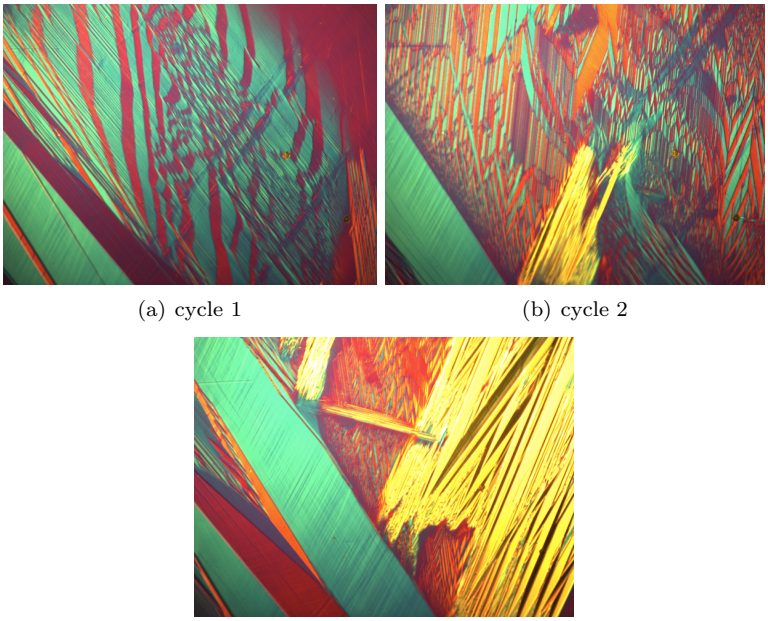
FIGURE 8. Stress-induced transformation in ZnAuCu and TiNiCuCo. See text. (a) is reprinted with permission from [48], copyright 2016, American Chemical Society. (b) is from [19] and is reprinted with permission from AAAS.

Another test, which in many ways is even more demanding of the material but can be done in a matter of weeks, is *stress-induced transformation*. In the simplest case – say the stress is a uniaxial tension in the direction $e, |e| = 1$, with stress $\sigma > 0$ – this corresponds mathematically to the study of the a modified energy minimization problem

$$\inf_{y \in \mathcal{A}} \int_\Omega (\varphi(\nabla y(x), \theta) - \sigma\, e \cdot \nabla y e)\, dx. \tag{7.1}$$

In the case of these two alloys the main effect of this particular stress is to raise the transformation temperature, which can be easily understood by solving[15] (7.1). Thus, one can do stress-induced transformation by fixing the temperature above θ_c and increasing the stress σ until the material transforms. This is seen as a flat region on the measured stress-strain curves in Figure 8.

Basic information on reversibility is seen in Figure 8. The ZnAuCu alloy shows nearly the same response after 100,000 cycles, under demanding conditions of almost 7% strain each cycle and peak (compressive) stresses of more than 500 MPa. The TiNiCuCo alloys is in many ways more impressive: even though the strain is lower (almost 2%) these tests were done in tension (more demanding) and the stress-strain curve at cycle 1 is extremely close to that at cycle 10 million. In both cases the material was made to undergo nearly complete transformation to martensite each cycle. For mathematicians unfamiliar with these units, a typical value of the yield (i.e., failure) stress of the steel beams that hold up a department of mathematics is 300 MPa. Several other measurements test reversibility in other ways. For example, $Zn_{45}Au_{30}Cu_{25}$ has a remarkably low thermal hysteresis (e.g., Figure 1), as low as 0.2°C.



(a) cycle 1 (b) cycle 2

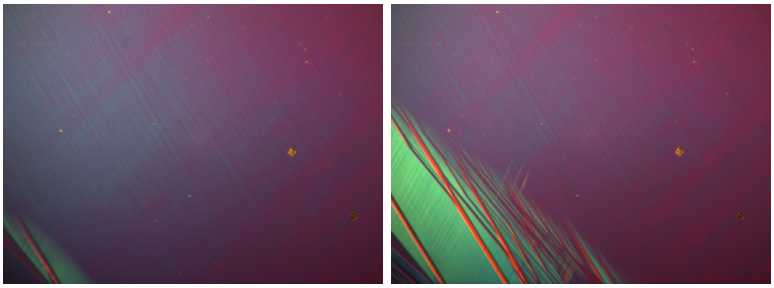
(c) cycle 3

FIGURE 9. Three successive transformation cycles at the same location in a grain showing the non repeatability of microstructure in $Zn_{45}Au_{30}Cu_{25}$.

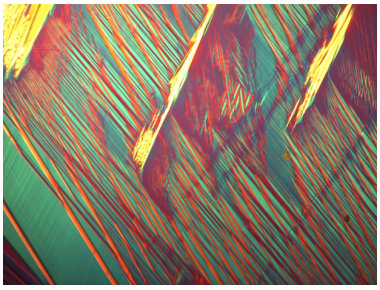
[15]This is easily done under mild growth conditions on φ. To see the simplest asymptotic result quickly, let $K = \mathrm{SO}(3)U_1 \cup \cdots \cup \mathrm{SO}(3)U_n$, assume that $\varphi(\cdot, \theta)$ rises steeply from the energy wells, note that one can minimize the integral by minimizing the integrand, and therefore reduce the problem to $\min_{F \in \mathrm{SO}(3) \cup K}(f_a(\theta)\chi_{\mathrm{SO}(3)} + f_m(\theta)\chi_K - \sigma e \cdot Fe)$. The energy wells for the ZnCuAu alloy can be found in the supplement of [58].

Typically, martensites exhibit a high degree of repeatability of the pattern of microstructure on heating and cooling [56]. Transform a typical martensitic material by cooling and one sees a pattern of microstructure. Heat to austenite (which wipes out the microstructure) and cool again: the pattern is very nearly the same. Often, in a polycrystal, plates of martensite appear by growing out of a defect or triple junction in the same way during each cycle.

Thus, it comes as a striking observation that, when $Zn_{45}Au_{30}Cu_{25}$ is heated and cooled back and forth through the transformation, the microstructure is completely different each cycle[16]. See Figure 9 or, for the full video, the supplement of [58], or, at a bit lower resolution, `http://www.aem.umn.edu/~james/research/`. How can such a highly reversible alloy behave in such a highly nonrepeatable way? Is it that, by satisfying the cofactor conditions, we have so flattened its infinite-dimensional energy landscape that the material can take any path? Sounds good at first, until



(a) frame 1 (b) frame 2



(c) frame 3

FIGURE 10. The process of transformation in $Zn_{45}Au_{30}Cu_{25}$ seen in three consecutive frames. Austenite is dark gray. Notice that once the microstructure appears there is very little further relaxation.

one recalls that, in an ordinary martensitic material, the plates of martensite tend to emerge from triple junctions and defects. Shouldn't the defects then even more strongly bias the microstructure with an otherwise flat energy landscape?

These and related questions have inspired mathematicians to look more critically at this video. We collect two observations that are particularly interesting. One is due to J. M. Ball and F. Della Porta [23] and is illustrated by Figure 10. It is as if the microstructure is already set behind a blanket, and one moves the blanket aside to reveal it. That is, there is very little further relaxation once the microstructure

[16]It should be mentioned that every effort was made in the heating/cooling device to give a periodic temperature profile vs. time.

appears. Perhaps this also can be rationalized by a very flat energy landscape – or, anthropomorphically, wherever you are, there is little driving force to push you elsewhere. Upon reflection, it is clear that this is a very strong restriction on the microstructure: previously, it was compatible with the austenite across a (possibly irregular) interface [23] and, after that, it did not change. In general, this kind of restriction, which embodies the idea that there must have been a low energy pathway to an observed energy minimizing state, has not been studied much, and also relates to the study of barriers mentioned above. For another interesting barrier see the wonderful experiment of H. Seiner [6].

A fascinating observation on $Zn_{45}Au_{30}Cu_{25}$ due to Noemi Barrera and Giovanni Zanzotto [7] relates to power-law behavior [47, 20] and the theory of self-organized criticality [41]. It is known that the martensitic phase transformation often takes

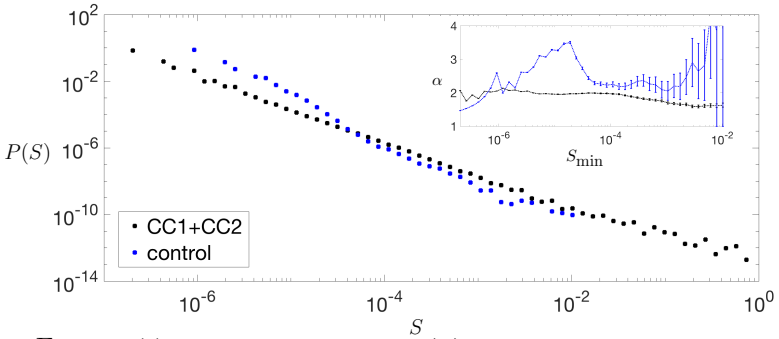


FIGURE 11. Empirical frequency $P(S)$ for the size S of transformation avalanches during the video (supplement, [58]), where S is the fraction of pixels in connected domains where there is an austenite-martensite switch of color. The plot (black) shows power-law behavior with exponent near 2 over a remarkable range of 6 orders of magnitude for $Zn_{45}Au_{30}Cu_{25}$ (labeled CC1+CC2). The control plot (blue) refers to transformation avalanches in a generic alloy. The inset shows the values of the exponent α of $P(S)$ for $Zn_{45}Au_{30}Cu_{25}$ (black) vs. the generic control (blue), determined by the maximum likelihood method [20] as a function of the lower cutoff imposed on the data.

place through abrupt strain events ("avalanches") even when the temperature or loading is smoothly changing [50, 2]. Figure 11 shows a plot of the empirical frequency of avalanche sizes, where size S refers to the fraction of pixels in a connected domain that undergoes an austenite-to-martensite switch of color between successive frames. Barrera and Zanzotto note [7] that the empirical avalanche statistics of $Zn_{45}Au_{30}Cu_{25}$ (labelled CC1 + CC2) have an exceedingly good power-law character, to a degree that is rare in materials science. Qualitatively, this lack of one or more characteristic scales indicates that $Zn_{45}Au_{30}Cu_{25}$ can perform a much wider and more efficient collection of adjustments of microstructure to environmental changes. It is a striking example of this type, that includes sand piles, earthquakes, stick-slip friction, the firing of neurons, and fluctuations in financial markets. Of course, it would be good to relate this behavior to the theory given above; see [5] for a model in this direction and [55] for recent diverse perspectives on the origins of behavior such as that shown in Figure 11.

8. The direct conversion of heat to electricity

We finish this article with a few brief remarks about one of the most interesting[17] applications of reversible transforming materials. It concerns the use of these materials for the direct conversion of heat to electricity in the small temperature difference regime. Here, "direct" means that the material itself creates the electricity without a separate electrical generator. The "small temperature difference regime" is the regime $10-200^\circ$ C, for which there does not currently exist a reasonable energy conversion device. Sources in this regime are ubiquitous: concentrated solar-thermal sources, data centers (which now consume $\sim 3\%$ of the energy budget in the US), waste heat from industrial sources, desktop and laptop computers, air conditioning systems, power plants, and even hand-held electronic devices[18].

There are at several ways that transforming materials can be used for the direct conversion of heat to electricity, and we will briefly mention two of them. They are enabled by the abrupt change of magnetoelectric properties that can occur in materials with big first order phase transformations, like those discussed here. The two cases are based on magnetism and ferroelectricity. In the first case one uses a material for which the low temperature phase is non-magnetic and the high temperature phase is strongly magnetic. For an example of such a material, which also has λ_2 quite close to 1, see [60].

One heats the material through the phase transformation. If left alone it would demagnetize itself[19] by forming domains, so we place it on top of a permanent magnet to bias it. This biasing can be understood from the theory of micromagnetics. As it transforms to the strongly magnetic phase as we heat it up, it magnetizes. We can think of magnetization as a time-dependent vector field $M : \Omega \times [0, t_1) \to \mathbb{R}^3$ satisfying $|M(x,t)| = \chi_{\Omega_1(t)}(x) M_s$ on the region Ω occupied by the material. Here, $\Omega_1(t) \subset \Omega$, $\Omega_1(0) = \emptyset$, $\Omega_1(t_1) = \Omega$, much like the frames in Figure 10 run backwards. By heating we increase $\fint_\Omega M(x,t)dx$. Ferromagnetism is governed by a well-known dipolar relation[20]

$$B = \mu_0(H + M) \tag{8.1}$$

as well as the Maxwell equations

$$\mathrm{div} B = 0, \quad \mathrm{curl}\, E = -\frac{\partial B}{\partial t}. \tag{8.2}$$

Interpreted physically, the latter means that, if the material is surrounded by a coil, a current will be generated, or, briefly, induction. In fact, one gets a current of opposite sign on cooling back through the phase transformation.

This is the barest explanation, but there are many subtle aspects. For example, the presence of the field of the permanent magnet changes the transformation temperature, since there is an effect of magnetic field on transformation temperature. This can be understood from energy minimization, analogously to the way

[17] in the opinion of the author

[18] for which, of course, the generated electricity would be used to help recharge the battery. In some of the computer examples mentioned, it is in fact a significant technological obstacle to *get rid of the heat*.

[19] since it is a soft magnet, which means, from the point of view of energy minimization, that its static behavior is well predicted by absolute energy minimization based on the theory of micromagnetics.

[20] For a mathematical perspective on this relation see [33].

stress was treated in (7.1): there is a contribution to the energy from the applied field. Moreover, the current induced in the coil modifies this field in an important way. The effect of this current on heating and cooling is different, and one gets two transformation temperatures. This splitting of the transformation temperature turns out to be terribly important to the efficiency and power output of such a device. Of course, minimizing the hysteresis is also critical. See [59] for a basic model.

As one can see from (8.1) and (8.2) the rapidly changing M is partitioned between B and H, but it is $\partial B/\partial t$ that creates the electricity in the surrounding coil. This partitioning is well-known to mathematicians who work in micromagnetics. A study of this partitioning also reveals a deficiency of this method: the good shape of Ω for a favorable $\partial B/\partial t$ seems to be a bad shape for the also crucial aspect of heat transfer. This dilemma suggests that, in fact, the ferroelectric case (with capacitance, instead of induction) is preferred, as will be explained in forthcoming work.

References

1. Edgar C. Bain, *The nature of martensite*, Trans. AIME **70** (1924), no. 1, 25.
2. Xavier Balandraud, Noemi Barrera, Paolo Biscari, Michel Grédiac, and Giovanni Zanzotto, *Strain intermittency in shape-memory alloys*, Physical Review B **91** (2015), no. 17, 174111.
3. J. M. Ball and R. D. James, *Fine phase mixtures as minimizers of energy*, Arch. Ration. Mech. Anal. **100** (1987), no. 1, 13.
4. J. M. Ball and R. D. James, *Proposed experimental tests of a theory of fine microstructure and the two-well problem*, Phil. Trans.: Phys. Sci. Eng. **338** (1992), no. 1650, 389.
5. John M. Ball, Pierluigi Cesana, and Ben Hambly, *A probabilistic model for martensitic avalanches*, MATEC Web of Conferences, vol. 33, EDP Sciences, 2015.
6. John M. Ball, Konstantinos Koumatos, and Hanuš Seiner, *Nucleation of austenite in mechanically stabilized martensite by localized heating*, Journal of Alloys and Compounds **577** (2013), S37–S42.
7. Noemi Barrera and Giovanni Zanzotto, *Power-law behavior and avalanches in phase transformations*, private communication, 2017.
8. Pavel Belik and Mitchell Luskin, *Stability of microstructure for tetragonal to monoclinic martensitic transformations*, ESAIM: Mathematical Modelling and Numerical Analysis **34** (2000), no. 3, 663–685.
9. Kaushik Bhattacharya, *Microstructure of martensite: why it forms and how it gives rise to the shape-memory effect*, Oxford series on materials modeling, Oxford University Press, 2003.
10. Kaushik Bhattacharya, Sergio Conti, Giovanni Zanzotto, and Johannes Zimmer, *Crystal symmetry and the reversibility of martensitic transformations*, Nature **428** (2004), no. 6978, 55–59.
11. Kaushik Bhattacharya and Robert V. Kohn, *Symmetry, texture and the recoverable strain of shape-memory polycrystals*, Acta Mater. **44** (1996), no. 2, 529–542.
12. Kaushik Bhattacharya, Bo Li, and Mitchell Luskin, *The simply laminated microstructure in martensitic crystals that undergo a cubic-to-orthorhombic phase transformation*, Archive for rational mechanics and analysis **149** (1999), no. 2, 123–154.
13. Xavier Blanc, Claude Le Bris, and P.-L. Lions, *Convergence de modèles moléculaires vers des modèles de mécanique des milieux continus*, C. R. Acad. Sci. Paris Sér. I Math. **332** (2001), 949–956.
14. ———, *From molecular models to continuum mechanics*, Archive for Rational Mechanics and Analysis **164** (2002), no. 4, 341–381.
15. Antonio Capella and Felix Otto, *A quantitative rigidity result for the cubic-to-tetragonal phase transition in the geometrically linear theory with interfacial energy*, Proceedings of the Royal Society of Edinburgh Section A: Mathematics **142** (2012), no. 2, 273–327.
16. Allan Chan and Sergio Conti, *Energy scaling and domain branching in solid-solid phase transitions*, pp. 243–260, Springer International Publishing, Cham, 2014.

17. Xian Chen, Yintao Song, Nobumichi Tamura, and Richard D. James, *Determination of the stretch tensor for structural transformations*, Journal of the Mechanics and Physics of Solids **93** (2016), 34 – 43, Special Issue in honor of Michael Ortiz.
18. Xian Chen, Vijay Srivastava, Vivekanand Dabade, and Richard D. James, *Study of the cofactor conditions: conditions of supercompatibility between phases*, J. Mech. Phys. Solids **61** (2013), no. 12, 2566.
19. Christoph Chluba, Wenwei Ge, Rodrigo Lima de Miranda, Julian Strobel, Lorenz Kienle, Eckhard Quandt, and Manfred Wuttig, *Ultralow-fatigue shape memory alloy films*, Science **348** (2015), no. 6238, 1004–1007.
20. Aaron Clauset, Cosma Rohilla Shalizi, and Mark E. J. Newman, *Power-law distributions in empirical data*, SIAM review **51** (2009), no. 4, 661–703.
21. S. Conti, M. Klar, and B. Zwicknagl, *Piecewise affine stress-free martensitic inclusions in planar nonlinear elasticity*, Proc. R. Soc. A **473** (2017), no. 2203, 20170235.
22. J. Cui, Y. S. Chu, O. Famodu, Y. Furuya, J. Hattrick-Simpers, R. D. James, A. Ludwig, S. Thienhaus, M. Wuttig, Z. Zhang, and I. Takeuchi, *Combinatorial search of thermoelastic shape memory alloys with extremely small hysteresis width*, Nature Materials (2006), 286–290.
23. Francesco Della Porta, *Modeling moving interfaces in reversible martensitic transformations*, Preprint, 2017.
24. R. Delville, S. Kasinathan, Z. Zhang, V. Humbeeck, R. D. James, and D. Schryvers, *A transmission electron microscopy study of phase compatibility in low hysteresis shape memory alloys*, Philosophical Magazine (2010), 177–195.
25. Weinan E, Weiqing Ren, and Eric Vanden-Eijnden, *String method for the study of rare events*, Physical Review B **66** (2002), no. 5, 052301.
26. J. L. Ericksen, *On the Cauchy-Born Rule*, Math. Mech. Solids **13** (2008), 199.
27. Leonhard Euler, *Recherches physiques sur la nature des moindres parties de la matière*, Berlin Academy Memoires **1** (1745), 28–32, (*Opera omnia*, III. 1: 615).
28. L. C. Flatley and Florian Theil, *Face-centered cubic crystallization of atomistic configurations*, Archive for Rational Mechanics and Analysis **218** (2015), no. 1, 363–416.
29. Irene Fonseca and Jan Malỳ, *Relaxation of multiple integrals below the growth exponent*, Annales de l'Institut Henri Poincare (C) Non Linear Analysis, vol. 14, Elsevier, 1997, pp. 309–338.
30. Gero Friesecke and Florian Theil, *Validity and failure of the Cauchy-Born hypothesis in a two-dimensional mass-spring lattice*, Journal of Nonlinear Science **12** (2002), no. 5, 445–478.
31. Hanlin Gu, Lars Bumke, Christoph Chluba, Eckhard Quandt, and Richard D. James, *Phase engineering and supercompatibility of shape memory alloys*, Materials Today **21** (2018), 265–277.
32. R. D. James and Z. Zhang, *A way to search for multiferroic materials with unlikely combinations of physical properties*, Magnetism and Structure in Functional Materials, Springer Series in Materials Science (A. Planes, L Manõsa, and A. Saxena, eds.), vol. 9, Springer-Verlag, Berlin, 2005, pp. 159–175.
33. Richard D James and Stefan Müller, *Internal variables and fine-scale oscillations in micromagnetics*, Continuum Mechanics and Thermodynamics **6** (1994), no. 4, 291–336.
34. Hans Knüpfer and Robert V Kohn, *Minimal energy for elastic inclusions*, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, The Royal Society, 2010, p. rspa20100316.
35. Hans Knüpfer, Robert V. Kohn, and Felix Otto, *Nucleation barriers for the cubic-to-tetragonal phase transformation*, Communications on pure and applied mathematics **66** (2013), no. 6, 867–904.
36. Robert V. Kohn, *Energy-driven pattern formation*, International Congress of Mathematicians, vol. 1, 2006, pp. 359–383.
37. Robert V. Kohn and Stefan Müller, *Branching of twins near an austenitetwinned-martensite interface*, Philosophical Magazine A **66** (1992), no. 5, 697–715.
38. ———, *Surface energy and microstructure in coherent phase transitions*, Communications on Pure and Applied Mathematics **47** (1994), no. 4, 405–435.
39. K. Koumatos and A. Muehlemann, *Optimality of general lattice transformations with applications to the bain strain in steel*, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **472** (2016), no. 2188, 20150865.

40. Jan Kristensen, *A necessary and sufficient condition for lower semicontinuity*, Nonlinear Analysis: Theory, Methods & Applications **120** (2015), 43–56.
41. Anna Levina, J. Michael Herrmann, and Theo Geisel, *Dynamical synapses causing self-organized criticality in neural networks*, Nature physics **3** (2007), no. 12, 857–860.
42. Mitchell Luskin, *On the computation of crystalline microstructure*, Acta Numerica **5** (1996), 191–258.
43. Robert D. MacPherson and David J. Srolovitz, *The von Neumann relation generalized to coarsening of three-dimensional microstructures*, Nature **446** (2007), no. 7139, 1053–5.
44. Isaak D Mayergoyz, *Mathematical models of hysteresis and their applications*, Academic Press, 2003.
45. Stefan Müller, *Variational models for microstructure and phase transitions*, Calculus of variations and geometric evolution problems, Springer, 1999, pp. 85–210.
46. W. W. Mullins, *Two-dimensional motion of idealized grain boundaries*, Journal of Applied Physics **27** (1956), 900–904.
47. Mark E. J. Newman, *Power laws, Pareto distributions and Zipf's law*, Contemporary physics **46** (2005), no. 5, 323–351.
48. Xiaoyue Ni, Julia R. Greer, Kaushik Bhattacharya, Richard D. James, and Xian Chen, *Exceptional resilience of small-scale $Au_{30}Cu_{25}Zn_{45}$ under cyclic stress-induced phase transformation*, Nano letters **16** (2016), no. 12, 7621–7625.
49. Patent iNSIGHT Pro, *Shape Memory Material Technology Insight Report*, Tech. report, Gridlogics Technologies Pvt. Ltd., 2015.
50. Francisco-José Pérez-Reche, Marcelo Stipcich, Eduard Vives, Lluís Mañosa, Antoni Planes, and Michel Morin, *Kinetics of martensitic transitions in cu-al-mn under thermal cycling: Analysis at multiple length scales*, Physical Review B **69** (2004), no. 6, 064101.
51. M. Pitteri, *Reconciliation of local and global symmetries of crystals*, Journal of Elasticity **14** (1984), no. 2, 175–190.
52. M. Pitteri and G. Zanzotto, *Continuum models for phase transitions and twinning in crystals*, Chapman & Hall/CRC, 2003.
53. Guang-Rui Qian, Xiao Dong, Xiang-Feng Zhou, Yongjun Tian, Artem R Oganov, and Hui-Tian Wang, *Variable cell nudged elastic band method for studying solid–solid structural phase transitions*, Computer Physics Communications **184** (2013), no. 9, 2111–2118.
54. Angkana Rüland, *The cubic-to-orthorhombic phase transition: rigidity and non-rigidity properties in the linear theory of elasticity*, Archive for Rational Mechanics and Analysis **221** (2016), no. 1, 23–106.
55. Ekhard K. H. Salje, Avadh Saxena, and Antoni Planes, *Avalanches in functional materials and geophysics*, Springer, 2017.
56. James P. Sethna, Karin Dahmen, Sivan Kartha, James A. Krumhansl, Bruce W. Roberts, and Joel D. Shore, *Hysteresis and hierarchies: Dynamics of disorder-driven first-order phase transformations*, Physical Review Letters **70** (1993), no. 21, 3347.
57. Cyril Stanley Smith, *Grain shapes and other metallurgical applications of topology*, Metal interfaces: a seminar on metal interfaces held during the 33$^{\text{rd}}$ National Metal Congress and Exposition, October 13-19, Detroit (Cleveland), American Society for Metals, 1951, pp. 65–108.
58. Y. Song, X. Chen, V. Dabade, T. W. Shield, and R. D. James, *Enhanced reversibility and unusual microstructure of a phase-transforming material*, Nature **502** (2013), 85–88.
59. Yintao Song, Chris Leighton, and Richard D. James, *Thermodynamics and energy conversion in heusler alloys*, Heusler Alloys, Springer, 2016, pp. 269–291.
60. Vijay Srivastava, Xian Chen, and Richard D. James, *Hysteresis and unusual magnetic properties in the singular heusler alloy $Ni_{45}Co_5Mn_{40}Sn_{10}$*, Appl. Phys. Lett. **97** (2010), no. 1, 014101.
61. Likun Tan and Kaushik Bhattacharya, *Length scales and pinning of interfaces*, Phil. Trans. R. Soc. A **374** (2016), no. 2066, 20150167.
62. Florian Theil, *A proof of crystallization in two dimensions*, Communications in Mathematical Physics **262** (2006), no. 1, 209–236.
63. Clarence Marvin Wayman, *Introduction to the crystallography of martensitic transformations*, Macmillan, 1964.

64. L. C. Young, *Generalized curves and the existence of an attained absolute minimum in the calculus of variations*, Comptes Rendus de la Société des Sci. et des Lettres de Varsovie **30** (1937), 212–234.
65. ———, *Generalized surfaces in the calculus of variations*, Annals of mathematics (1942), 84–103.
66. L.C. Young, *Lectures on the calculus of variations and optimal control theory*, Saunders, 1969, Reprinted by A.M.S. Chelsea 1969.
67. Giovanni Zanzotto, *The Cauchy–Born hypothesis, nonlinear elasticity and mechanical twinning in crystals*, Acta Crystallographica Section A: Foundations of Crystallography **52** (1996), no. 6, 839–849.
68. Nikolai A Zarkevich and Duane D Johnson, *Shape-memory transformations of NiTi: minimum-energy pathways between austenite, martensites, and kinetically limited intermediate states*, Physical review letters **113** (2014), no. 26, 265701.
69. R. Zarnetta, R. Takahashi, M. L. Young, A. Savan, Y. Furuya, S. Thienhaus, B. Maaß, M. Rahim, J. Frenzel, H. Brunken, Y. S. Chu, V. Srivastava, R. D. James, I. Takeuchi, G. Eggeler, and A. Ludwig, *Identification of quaternary shape memory alloys with near zero thermal hysteresis and unprecedented functional stability*, Advanced Functional Materials (2010), 1917–1923.
70. Zhiyong Zhang, R. D. James, and Stefan Müller, *Energy barriers and hysteresis in martensitic phase transformations*, Acta Materialia (Invited Overview) **57** (2009), 2332–4352.
71. Barbara Zwicknagl, *Microstructures in low-hysteresis shape memory alloys: Scaling regimes and optimal needle shapes*, Archive for Rational Mechanics and Analysis **213** (2014), no. 2, 355–421.

Department of Aerospace Engineering and Mechanics, University of Minnesota, Minneapolis, MN 55455

E-mail address: james@umn.edu