Symmetry and reversibility of martensitic transformations

Kaushik Bhattacharya⋆§, Sergio Conti†§, Giovanni Zanzotto‡§ and Johannes Zimmer†§

December 23, 2003

Martensitic transformations are diffusionless solid-to-solid phase transformations characterized by a rapid change of crystal structure, observed in metals, alloys, ceramics, and proteins1,2. They come in two widely different classes. In steels, the transformation microstructure induced by quenching remains essentially unchanged upon subsequent loading or heating; the transformation is not reversible. In shape-memory alloys, on the contrary, the microstructures formed on cooling are easily manipulated by loads and disappear upon reheating; the transformation is reversible. Here we show that these sharp differences are dictated by the symmetry of the energetic landscape. In weak transformations the symmetry groups of both phases are included in a common finite symmetry group3, in reconstructive transformations they are not4. We demonstrate that for reconstructive transformations the energy barrier to lattice-invariant shears (as in twinning or slip) is no higher than the barrier to the phase transition itself. A remarkable implication is that reconstructive transformations are accompanied by plastic deformation through dislocations and twinning in the parent phase, making these phase changes irreversible. In contrast, for weak transformations, the energy barrier to lattice-invariant shears is independent of that to the phase transition. Consequently, weak transformations can occur with virtually no plasticity and are potentially reversible.

Martensitic transformations are at the basis of numerous technological applications. Most notable amongst these is in steel, where the transformation induced by quenching (fast cooling) is exploited for enhancing the alloy’s strength1. Another is the fascinating shape-memory effect in alloys like Nitinol, used in medical and engineering devices5. Martensitic phase changes are also exploited to toughen structural ceramics6 such as zirconia, and observed in biological systems such as the tail sheath of the T4 bacteriophage virus7. Ideas
originating from the study of these transformations have led to improved materials for actuation (ferromagnetic shape-memory alloys\textsuperscript{8,9} and ferroelectrics\textsuperscript{10,11}) and to candidates for artificial muscles\textsuperscript{12}. Finally, the rich microstructure (distinctive patterns developed at scales ranging from a few nanometers to a few microns) that accompanies these transitions, has made this a valuable theoretical sand-box for the development of multi-scale modeling tools\textsuperscript{13}.

In some materials such as shape-memory alloys, the martensitic phase change is almost perfectly reversible (and often termed ‘thermoelastic’). As one starts at high temperature and cools a shape-memory alloy like Nitinol, the transformation proceeds by the appearance and growth of microstructure with a (twinned) plate-like morphology; the transformation from start to finish takes about 20 degrees Celsius. The microstructure is mobile and can be changed by the application of loads. The transformation can be completely reversed, with the disappearance of the microstructure upon reheating and the appearance of little or no dislocations or twinning in the parent (high temperature, high symmetry) phase.

In stark contrast, the martensitic transformation is not reversible in materials like steel and other alloys, such as CoNi. In steels, the microstructure forms in a sudden burst with a lath-like morphology upon cooling, and is immobile on loading. Moreover, the transformation is irreversible and the microstructure does not disappear upon reheating. In CoNi the phase change is also irreversible, as successive microstructures form both on heating and cooling. These materials undergoing irreversible transformations are characterized by significant dislocations and twinning in the parent phase.

We provide an explanation for this difference in (ir-)reversibility on the basis of the symmetry change during the transformation. We call weak the martensitic transformations in which the symmetry group of both the parent and product phase are included in a common finite symmetry group\textsuperscript{3} (which includes symmetry breaking), and reconstructive otherwise\textsuperscript{4} (note that this is different from the usage of the term ‘reconstructive’ to mean ‘diffusional’; all the transformations considered here are diffusionless). We show through rigorous mathematical theory and numerical simulation that irreversibility is inevitable in a reconstructive phase transformation, and not so in a weak transition.

Fig. 1 illustrates our main idea through a square-to-hexagonal reconstructive phase change in a two-dimensional crystal. Consider first a square lattice with a unit cell shown on the left, and suppose this cell is transformed to the unit cell of the hexagonal lattice shown in solid (middle). The symmetry of the hexagonal lattice implies that the solid and dashed unit cells shown in the middle are equivalent. If the crystal is transformed back to the square phase, the dashed hexagonal cell can go to, say, the dashed cell on the right. Crucially, the square cell on the left is then transformed to the sheared cell of the square lattice on the right. In short, upon transforming, performing a symmetry operation, and transforming back, we have deformed the crystal through a lattice-invariant shear, i.e., a shearing deformation that leaves the entire (ideal, infinite) lattice invariant. Our results hold for finite lattices if boundary effects can be neglected, as illustrated below by means of numerical simulations.

Various experimental observations confirm our predictions. Pure iron (Fe) undergoes a temperature induced reconstructive $\gamma$-$\alpha$ (face-centered-cubic to body-centered-cubic, fcc-to-bcc) transformation. As Ni and C are added to obtain steel, the martensite becomes body-centred tetragonal (bct) with an in-
Figure 1. A lattice-invariant shear can be generated by a forward and reverse square-to-hexagonal phase transition. The transformation takes the solid square on the left to the solid rhombus in the middle. The hexagonal symmetry implies the equivalence of the solid rhombic cell to the dashed rhombic one. The reverse transformation takes the latter to the dashed square on the right. In the process, the original solid square on the left has sheared by a lattice-invariant shear to the solid parallelogram on the right.

Increasing tetragonality, so that the martensitic transformation becomes weak. Correspondingly, Maki and Tamura\textsuperscript{14} found that the reversibility of the phase changes increased, the amount of plastic deformation decreased, and the morphology changed from lath to butterfly to lenticular to plate-like with increasing Ni and C. Iron also undergoes a pressure-induced reconstructive $\alpha$-$\varepsilon$ (bcc to hexagonal-close-packed, hcp) transformation, again accompanied by both twins and dislocations\textsuperscript{15,16}. Another set of observations concern materials undergoing the reconstructive fcc-to-hcp transformation. Liu \textit{et al.}\textsuperscript{17} considered the CoNi system; they hardened the parent phase to prevent any plastic deformation, but found that the transition was still irreversible, with both heating and cooling producing twins and plastic deformation. Similar observations were made in FeMnCrSiNi steels\textsuperscript{18,19}. Finally, in a block co-polymer, the (reconstructive) bcc-to-hexagonal transformation was observed to be irreversible because of orientation proliferation through twinning in both phases\textsuperscript{20}.

In spirit, our results are related to those of Otsuka and Shimizu\textsuperscript{21} who discuss the effects of ordering on crystallographic reversibility of martensitic transformation in alloys. They observe that in ordered alloys the transformation path must be such that the order is not destroyed, whereas in disordered ones only the atomic positions need to be recovered. While their results are consistent with many experiments, they also note that the fcc-to-face-centered-tetragonal (fct) transformation is an ‘exception’, which they argue is a consequence of the fact that ‘the lattice correspondence is unique in the reverse transformation because of the so simple lattice change and lower symmetry of the fct phase’. We show here that the essential difference depends on the crystal symmetry, rather than on order and disorder. This provides a general framework which includes both their theory and exception, and also generalizes to other situations.
The common explanation (e.g., Maki and Tamura\textsuperscript{[14]}) for the irreversibility in Fe (or low Ni or C steels) is based on the volume change that accompanies the transition. The idea is that a partially transformed region causes stress and plastic deformation. However, a system like the styrene block co-polymer considered by Lee \textit{et al.}\textsuperscript{[20]} should largely be unaffected by any volume change. Our explanation is \textit{independent} of such a volume effect. It is possible, however, that both mechanisms contribute to the irreversibility in Fe.

As our result is only based on symmetry, it is independent of material parameters, such as the elastic moduli. We show that lattice-invariant shears in reconstructive transformations cost only as much as the phase transition itself, making reconstructive transformations irreversible. On the other hand, weak transformations have the potential to be reversible, because the energy barriers to lattice-invariant shears and the phase transition are independent of each other. Our numerical simulations indicate that they are in fact reversible in a generic situation. However, the barriers might happen to be comparable in particular materials undergoing a weak transformation. This for example is the case in Ni-50Ti, and plastic deformation masks the reversible transformation. However, (precipitate) hardening this material makes these barriers different thereby revealing the reversible character of the underlying weak transition\textsuperscript{[5]}. In contrast, hardening does not rescue the irreversibility of CoNi\textsuperscript{[17]} or FeMnCrSiNi\textsuperscript{[18,19]}, which undergo reconstructive transformations. In summary, being weak is a necessary condition for reversibility.

Another consequence of our remarks is that materials like FeNi should be comparatively softer at compositions that are closer to a reconstructive transformation. This effect, however, may be obscured by the fact that the very softness of the ideal lattice produces plastic deformation and the entanglement of dislocations in the crystal, leading eventually to hardening. None of these phenomena are present in crystals undergoing weak martensitic transformations.

We now present the general theory followed by a concrete example, and numerical simulations. We state the theory for the simple case of Bravais lattices, which can be readily extended to crystals whose translational symmetry is not discontinuous across the phase change. A Bravais lattice \(L(e_i)\) is given by the linear combinations, with integral coefficients, of three independent vectors \(\{e_1, e_2, e_3\}\) forming the lattice basis. Another basis \(\{f_1, f_2, f_3\}\) generates the same lattice, \(L(e_i) = L(f_i)\), if and only if \(f = \sum_{j=1}^{3} m_{ij}^j e_j\) with a matrix \(m\) belonging to the group\textsuperscript{[22,3,23]}

\[
\mathcal{G} := \{m : m_{ij}^j \text{ integers and } \det(m) = \pm 1\},
\]

which is the \textit{global symmetry group} of Bravais lattices. Applied to any given lattice, \(\mathcal{G}\) consists of rotations, reflections, lattice-invariant shears, and combinations thereof; the restriction to rotations and reflections gives the \textit{lattice group} (a matrix representation of the point group) of that lattice.

The free energy density \(\Phi\) of the crystal is a function of the lattice basis at a fixed temperature. It is invariant under the global symmetry group \(\mathcal{G}\), as it cannot distinguish among bases generating the same lattice\textsuperscript{[24,25,13]}:

\[
\Phi(e_i) = \Phi(m_{ij}^j e_j) \text{ for every } m \in \mathcal{G}.
\]

This global framework takes all possible deformations into account, including large shearing distortions. Due to (2) the energy landscape of the crystal has
infinitely many wells, which are not contained in any bounded region in strain space (see Fig. 2).

For weak martensitic transformations the invariance of the energy (2) can be limited to a finite subgroup of $G$ due to the group-subgroup relation; correspondingly, the domain of the energy can be restricted to a neighbourhood of the reference configuration. This neighbourhood does not contain any lattice-invariant shears and only contains a finite number of energy wells. Such domains are called *Ericksen-Pitteri neighbourhoods* (EPNs)\(^{24,25}\), see Fig. 2. Therefore the classical framework of Landau theory and nonlinear elasticity\(^{24,26,4,23,13}\) applies in these cases.

For reconstructive martensitic transformations, on the other hand, the symmetry decreases along the transformation path, but increases again at the final state, and there is no reference configuration whose lattice group contains those of the two given phases.

We establish the following mathematical fact: the lattice groups of the two phases in any reconstructive transformation necessarily generate unbounded distortions. This implies that in this case the barrier between the infinitely many energy wells of the crystal is at most equal to that of the underlying transformation. Consequently no reduction to a finite number of wells in a bounded region is possible (no EPN can be extracted). In particular, since one can find two states of the crystal related through an arbitrarily large distortion which are only separated by a barrier as high (low) as the transition’s, such material cannot resist certain arbitrarily large distortions; this creates defects in the lattice and makes the reconstructive phase change necessarily non-thermoelastic and irreversible. Further, no classical approach of the Landau type is possible. (Some authors have extended the Landau framework to encompass reconstructive transformations by introducing a ‘transcendental order parameter’\(^4\), which gives a partial description of the lattice periodicity characterized by the global group $G$.)

We show in the methods section that in a reconstructive transformation an unbounded element of $G$ is always generated. Precisely, one necessarily obtains an element with an infinite period (i.e., whose powers can become arbitrarily large), akin to slip and twinning in the parent phase. We further notice that the most common reconstructive transformations involve phases with maximal point symmetry, i.e., the primitive-cubic, the fcc, the bcc, or the hexagonal subgroups of $G$. One can show that, when suitable lattice bases are considered, the phase changes arising in these cases generate the entire group $G$, with its full set of lattice-invariant shears. We elaborate on the fcc-to-bcc transformation below.

The impossibility to restrict the symmetry to a finite subgroup of $G$, and the energy domain to a suitable EPN, has dramatic implications for the variational treatment of reconstructive phase transformations. Indeed, if one assumes that the deformation at each time is determined by minimizing the free energy subject to external forces and boundary conditions, the invariance of the energy under the whole group $G$ implies that the solid cannot resist any shear\(^{27}\). In practice, dynamics and defects, including dislocations, moderate this phenomenon, which we revisit through our numerical simulations.

We now focus on a concrete case-study, demonstrating that the composition of an fcc-to-bcc (forward) transformation and a bcc-to-fcc (reverse) transformation can in fact result in a lattice-invariant shear. We start from an fcc lattice
Figure 2. Schematic representation of weak vs. reconstructive transformations in the space of lattices.  

**a** Weak: a uniaxial deformation (dashed black arrows) of the fcc lattice (represented by squares) can give three equivalent bct lattices (circles). The reverse transformation (dashed green arrows) returns to the original fcc configuration. All the transformation strains are confined within a single Ericksen-Pitteri neighbourhood (EPN), such as the dashed circle. 

**b** Reconstructive fcc-to-bcc: the transformation leads from an fcc to three equivalent bcc lattices, and the reverse transformation from each bcc can proceed to three distinct but equivalent fcc lattices, and so on. No EPN can be singled out.
aligned with the coordinate axes, and subject it to a uniaxial stretch \( U_z \) along the \( z \) axis:

\[
U_z^{(\lambda)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/\lambda \end{pmatrix},
\]

neglecting volumetric changes. For \( \lambda \) between 1 and \( \sqrt{2} \), the product lattice has tetragonal symmetry, and we have a weak transformation. When \( \lambda \) is equal to \( \sqrt{2} \), we obtain the Bain stretch \( U_z^{(\sqrt{2})} = B_z \) which produces the bcc lattice in a reconstructive transformation, see Fig. 3. Now consider the reverse bcc-to-fcc transformation: it can occur in three possible ways, by the symmetry of the bcc phase. A crucial point is that the symmetry axes of the bcc lattice are different from those of the initial fcc crystal. One of the three possibilities for the reverse stretch corresponds to \( B_z^{-1} \), and leads back to the original lattice. The other two are stretches along the \( (1, \pm 1, 0) \) directions (still using the original coordinate system). For instance, let us consider the reverse stretch \( V \) along the \((1, 1, 0)\) direction, which can be expressed as

\[
V = QB_z^{-1}Q^T,
\]

where \( Q \in SO(3) \) is the 90-degree rotation with axis \((1, -1, 0)\), which belongs to the symmetry group of the bcc phase. After applying \( V \), the final lattice is again fcc; however, the total transformation gives

\[
VB_z = QB_z^{-1}Q^TB_z = RS,
\]

where \( R \) is a rotation and \( S \) is a shear on a \{111\}\textsubscript{fcc} plane along a \( \langle 112 \rangle \)\textsubscript{fcc} direction. One can check that \( S \) brings the original fcc basis to a \( G \)-equivalent one, that is, the transformation \( VB_z \) restores the fcc lattice to itself up to the inessential rigid-body rotation \( R \) (see Fig. 3). Successive iterations of \( VB_z \) then generate larger and larger lattice-invariant deformations, and similar strategies can generate the entire group \( G \). The barrier between the \( G \)-related sheared configurations originated in this way is exactly equal to the barrier of the phase transition. One can check that if only part of the crystal undergoes the \( VB_z \) transformation, while the rest goes back to the original state, a Shockley partial dislocation with Burgers vector \( \frac{1}{6} \langle 112 \rangle \) is formed\textsuperscript{28}. The double transformation \((VB_z)^2\) then gives the Burgers vector \( \frac{1}{2} \langle 110 \rangle \). Both kinds are typical of fcc crystals. When starting from a bcc crystal, this same mechanism generates dislocations with Burgers vector \( \langle 111 \rangle \)\textsubscript{bcc} on \{112\}\textsubscript{bcc} planes, which are among the most common ones in bcc lattices.

We now illustrate the phenomena discussed above for finite domains, also in the presence of boundary conditions, with numerical simulations done for simplicity for a square-to-hexagonal (s-h) transformation in two dimensions. We consider a 50 × 50 grid of atoms interacting with a three-body nearest-neighbor potential\textsuperscript{29}, which produces for the crystal a two-dimensional version of the \( G \)-invariant energy in (2).

The simulation of a shearing experiment of a crystal close to a reconstructive phase transition is shown in the left panel of Fig. 4. As the left and the right boundaries are progressively displaced, dislocations form in the lattice, which lead to irreversible plastic deformations and defects in the crystal.

In contrast, the right panel of Fig. 4 shows the same simulation for a crystal close to a square-to-rhombic (s-r) symmetry-breaking (weak) transition. In this
Figure 3. Shear generated for a face-centered-cubic (fcc) to body-centered-cubic (bcc) transition, shown in a fcc-bcc-fcc cycle. 

a An fcc lattice, with body-centred-tetragonal (bct) cell in red.

b The uniaxial Bain stretch $B_z$ transforms the bct cell to the bcc one.

c The inverse Bain stretch $V = Q B_z^{-1} Q^T$ transforms the crystal back to fcc (blue fcc cell highlighted), but sheared relative to the original. The green arrows show a basis undergoing the lattice-invariant shear.
Figure 4. Reconstructive transformations (left panel) generate dislocations, weak ones (right panel) do not. 

a. Section of the energy profile of a crystal close to a square-to-hexagonal (s-h) transformation, along a s-h-s line. The energy has the full invariance (2), with minima s and h plotted respectively at positions 0, 1, and 0.5, 1.5, etc.

b. An incremental shear test with boundary conditions on the left and right side results in dislocations.

c. Section of the energy profile of a crystal close to a square-to-rhombic (s-r) transformation, along a s-r-h line. Here the square states s in 0, 1, etc. are metastable, and additional minima are present at intermediate rhombic configurations r, with hexagonal maxima at 0.5, 1.5, etc.

d. The same incremental shear test results in no dislocations, but only reversible transformation twins, because starting from the square minimum in 0, the ‘far-away’ square in 1 is not reachable by overcoming a small energy barrier.
case, no dislocations arise in the lattice. Instead, typical layered martensitic twins mixing with the higher symmetry parent phase can be observed, which accommodate the imposed boundary condition in a reversible way (this is at the origin of the memory effect). These observations confirm the importance of the symmetry of the stable states of a crystal for determining the macroscopic reversibility properties of the phase transition.

Methods

To prove that in a reconstructive transformation an aperiodic element of $G$ (called $\text{GL}(3, \mathbb{Z})$ in algebra) is generated, we first notice that any lattice group is finite, and conversely any finite subgroup of $G$ is included in the lattice group of some lattice (see e.g. Proposition 3.5 in Pitteri and Zanzotto\(^{23}\)). Thus a transformation is weak if and only if the lattice groups of the two crystal phases generate a finite group. Therefore a reconstructive transformation produces an infinite subgroup of $G$ with a finite number of generators. Such a group necessarily contains an element with no finite period as a consequence of the Burnside-Schur Theorem on periodic groups.

We finally establish that for suitable lattice bases, any reconstructive transformation involving Bravais lattices with maximal symmetry will generate the entire group $G$. Indeed, it is readily verified that for suitable pairs of subgroups in $G$ belonging to the four arithmetic classes with maximal point symmetry (i.e. the hexagonal and the primitive, face-centered, and body-centered cubic classes) one can indeed produce all the generators of $G$, i.e. a suitable reflection, permutation, and simple shear\(^{30}\).

References


**Acknowledgements**

The work of S. C. was partially supported by the DFG Schwerpunktprogramm *Analysis, Modeling and Simulation of Multiscale Problems*. G. Z. acknowledges the partial support of the Italian Cofin. MURST2002 *Modelli Matematici per i Materiali*. S. C. and G. Z. acknowledge the partial support of the EU TMR network *Phase Transitions in Crystalline Solids*. K. B. and J. Z. acknowledge the partial financial support of US AFOSR and the US ONR.

**Competing interests statement**

The authors declare that they have no competing financial interests. Correspondence and requests for materials should be addressed to K. B. (bhatta@caltech.edu).