PSEUDOELASTICITY IN SHAPE MEMORY ALLOYS - AN EXTREME CASE OF THERMOELASTICITY

by

Ingo Müller

FB9-Hermann Föttinger Institut
Technical University, BERLIN

1. Thermoelasticty for the Description of Phase Transformation

Traditionally thermoelastictiy has investigated stored energy functions that are convex for all deformations in the whole regime of interest. To be sure, convexity was absent in the case of the van der Waals gas, but one did not think of that phenomenon as fit for studies by elasticians; this was clearly a case for a thermodynamicist interested in phase transitions.

That neat separation of fields of research broke down with the investigation of high polymers in particular rubber - an elastic material par excellence. A piece of rubber heats up upon extension, so it is properly thermoelastic and deformations up to a thousand percent are elastically recovered. However, if the piece of rubber, stretched to a thousand percent is allowed to cool, it stays long after unloading. A phase transition has occurred, the rubber is now crystallized.

Phenomena like these motivated a closer study of the stored energy function. This function really is the free energy or Helmholtz free energy of thermodynamics and it has an energetic and an entropic part. In polymers and gases the entropy dominates, it is nice and convex, but the energy is concave and at low temperatures the energetic contribution to free energy can overwhelm the entropic one, thus producing a non-convex stored energy function, which can be interpreted as implying the possibility of a phase transition.

The significance of the non-convex stored energy function, the question of convexifying it and what that means with respect to stability, metastability and phase equilibrium is the subject of a lively field of research in mathematics to which this paper contributes nothing.

Rather this paper is concerned with describing another instance of non-convexity, the occurrence of pseudoelasticity in shape memory alloys which undergo an austenitic-martensitic phase transformation.

2. Pseudoplastic and Pseudoelastic Behavior of Shape-Memory Alloys

Memory alloys have widely different load-deformation curves at different temperatures. Figure 1 illustrates this phenomenon by presenting four load-deformation curves corresponding to increasing temperatures T_1 through T_4.

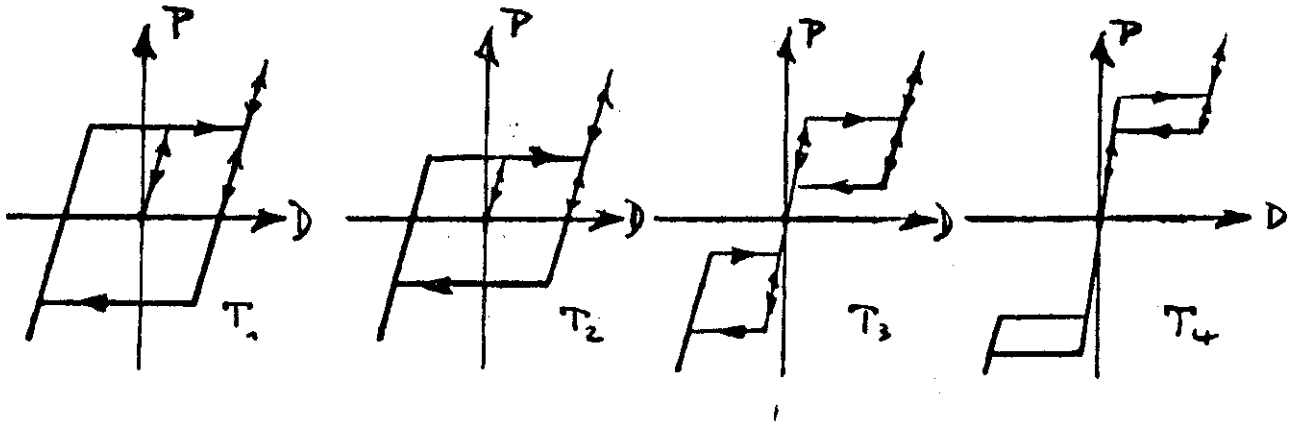


Fig. 1: Schematic load-deformation curves at different temperatures. The arrows indicate possible directions in loading and unloading. These curves have been abstracted from papers in [1] and [2].

At low temperature the curves are much like those of a plastic body with a virginal elastic curve, yield limit and residual deformation, and the yield limit decreases with increasing temperature. We might therefore call the alloys pseudoplastic in that range of temperature, even though that term is not commonly used. What distinguishes the curves from those of a truely plastic body is, of course, the existence of the elastic branches that close off the hystere-

sis region right and left.

At high temperatures the load deformation curves are called pseudoelastic. The bodies are elastic at those temperatures, in that they return to the origin after loading and unloading, but they are only pseudoleastic because there is a hysteresis. The hysteresis region is bounded above and below by the yield limit and the recovery limit respectively. Both grow with increasing temperature and the hysteresis becomes smaller.

It is obvious from Figure 1 why the materials are called memory alloys. Indeed, if at low temperature a body remains deformed after unloading, it will return to the origin upon heating, since only the deformation $D = 0$ is possible for $P = 0$ at higher temperature. We may therefore say that the materials remember their natural state.

Typically strains of up to 8% are recoverable and the temperature range of Figure 1 may be 50K.

The purpose of this paper is the formulation of a thermoelastic constitutive equation that can describe the pseuodelastic behaviour. We shall reach the objective by applying arguments from statistical mechanics to a simple one-dimensional model of a body with shape memory.

The model was described in 1981 by Müller & Wilmanski [3] who thought at the time that they had satisfactorily explained pseudoelasticity. Later that explanation was found to be an artifact of the approximation used. The present paper provides a correction of the part of [3] that concerns pseudoelasticity. The treatment of pseudoplasticity was correct in [3] and we shall therefore not discuss it here. However, the description of the model will be repeated here for self-containedness of this paper.

3. The Model

3.1 Motivation

Metallurgist have determined the cause of the load-deformation temperature behaviour of memory alloys to be a martenistic-austenistic phase transition and the formation of twins in the martensite. At high temperature the highly symmetric austenitic phase is stable while at low temperatures martensite is stable. On the lateral elastic lines in Figure 1 we have one or another martensitic twin while inside the low temperature hysteresis loops different twins coexist in different proportions. At high temperature the elastic line through the origin pertains to austenite and along both the yield line and the recovery line austenite and one martensitic twin coexist in different proportions.

3.2 Basic Element of the Model

The basic element of the model is a lattice particle, a small piece of the metallic lattice. Figure 2 shows such a lattice particle in three equilibrium positions denoted by A and $M_{\pm}$ for austenite and the martensitic twins respectively. $M_{\pm}$ are simply sheared versions of A and intermediate shear lengths Δ are also possible. Figure 2 shows the postulated form of the potential energy of a lattice particle as a function of Δ. It is characterized by three minima, two stable ones corresponding to $M_{\pm}$ and a metastable one for A. In the sequel we shall simplify the potential in the manner shown in Figure 3 in order to simplify the calculations.

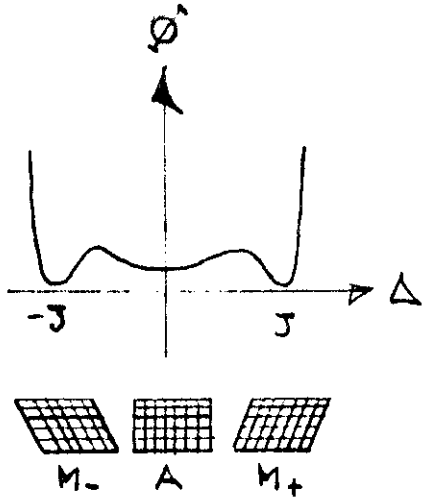


Figure 2: Lattice particles and their potential energy.

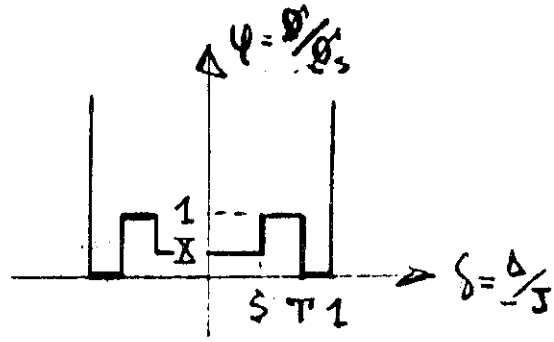


Figure 3: Simplified form of potential energy. Φ and Δ are normalized.

3.3 The Body as a Whole

We arrange the lattice particles in layers and stack the layers to form the body as shown in Figure 4a. [+)] Layers of M_+ and M_- alternate in Figure 4a and there is no austensite so that the figure refers to small temperatures. In order to become familiar with the model we discuss its comportment under a tensile load in the vertical direction.

Under the load the layers will be sheared. M_- layers become steeper, M_+ layers become flatter and each layer contributes the vertical component of its shear length to the overall length of the body which we call L (see Figure 4b). Thus we have for the deformation $D = L - L_0$

$$D = \frac{1}{\sqrt{2}} \sum_{i=1}^{N} \Delta_i \qquad (3.1)$$

+) This arrangement is obviously very special. Achenbach, Atanackovic & Müller [4] have refined the model so as to be able to discuss polycrystalline bodies in loading situations that are not necessarily uniaxial.

The summation extends over all N layers. After removal of the load the body will return to the configuration of Figure 4a, i.e. its deformation was elastic. A bigger load, however, will be able to flip the M_- layers into the M_+ configuration and the flipping is accompanied by a large growth of the shear lengths, whereby the deformation grows drastically as shown in Figure 4c. Moreover that deformation is non-elastic, because upon removal of the load all layers settle into the M_+-minimum of the potential energy so that there remains a large residual deformation (see Figure 4d).

Thus we see that the elastic deformation, the yield and the residual deformation of the body can be simulated by the shearing and flipping of the lattice layers of the model. To conclude this qualitative discussion let us consider how the model simulates the recovery.

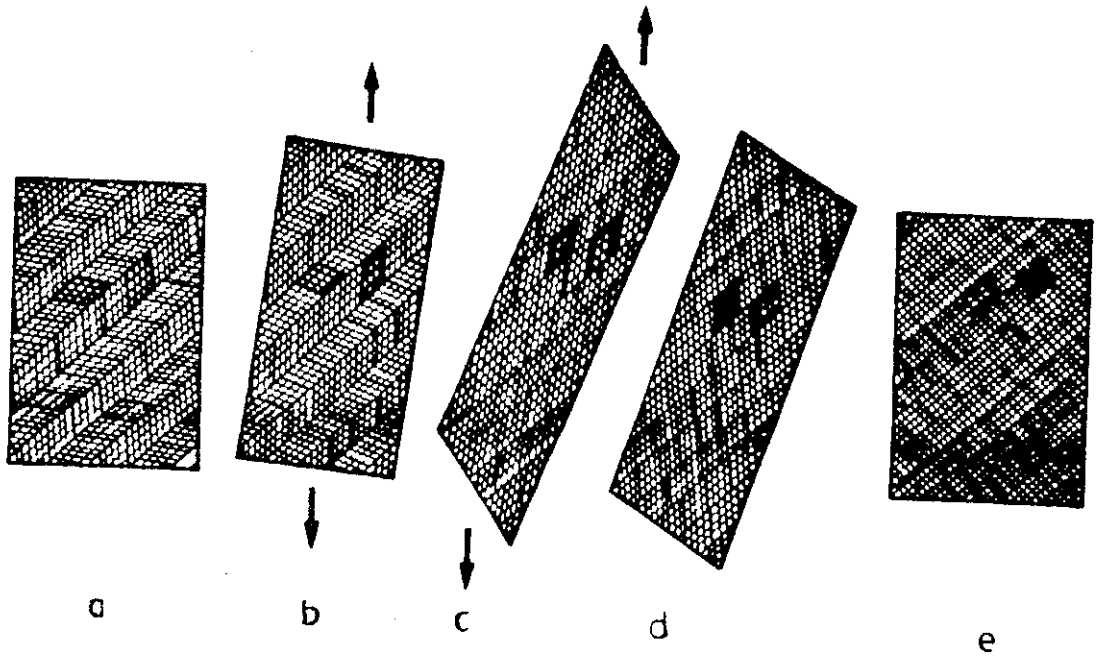


Figure 4. Model of a body built from martensitic and austenitic layers.

Recall that at high temperature the body is austenitic. Therefore upon heating the configuration of Figure 4d converts into the configuration of Figure

4e with austenitic layers. When the body is subsequently allowed to cool down, its layers will again become martensitic and in all probability there will be equal amounts of M_+ and M_-, so that we shall have returned to the configuration of Figure 4a which has the same shape to the naked eye as that of Figure 4e.

3.4 The Effect of Temperature T and Phase Fractions.

The higher the temperature is the more vigorous are the thermal fluctuations. That fact accounts for the observation of the austenitic phase at high temperature. Indeed, while at low temperature the lattice particles are caught in the deep lateral martensitic minima, at high temperature the fluctuations help them out of those minima so that they can assemble in the central austenitic minimum.

When fluctuations are present it is impractical to characterize the state of the body by the set of shear lengths Δ_i $(i = 1,2,\ldots,N)$. Rather we characterize the state by giving the numbers N_Δ of layers with a particular shear length Δ in the range available to the layers. In case of the potential of Figure 3 that range lies between $-J$ and $+J$. With this new description equation (1) assumes the form

$$D = \frac{1}{\sqrt{2}} \sum_{\Delta=-J}^{J} \Delta N_\Delta \tag{3.2}$$

Always referring to the potential of Figure 3 we shall say that a layer is in the austenitic phase if its shear length is in the range $-T < \Delta < T$. On the other hand, if $|\Delta| > T$ we call the layer martensitic. In a self explanation manner we write

$$\left.\begin{matrix}[\,] \\ ]\,[\end{matrix}\right\} \text{for the} \begin{matrix}\text{austenitic} \\ \text{martensitic}\end{matrix} \text{range of } \Delta. \tag{3.3}$$

The fraction of austenitic and martensitic layers is denoted by x_A and x_M respectively, so that we may write

$$x_A = \frac{1}{N} \sum_{[\,]} N_\Delta \qquad \text{and} \qquad x_M = \frac{1}{N} \sum_{][} N_\Delta \tag{3.4}$$

Obviously we nave

$$x_A + x_M = 1. \tag{3.5}$$

3.5 Interfacial Energies

Whenever there is an interface between two layers of phase, M_+, M_- or A there is a distortion of the metallic lattice and that carries a certain interfacial energy, so that energetically interfaces are unfavourable. This effect is less pronounced at M_+-M_- interfaces, because the twins have the same lattice structure and the distortion at the interface is minimal. Therefore we ignore the corresponding energies. In the case of a M-A interface however the lattice distortion is often considerable and we ascribe the energy e to each such interface.

4. The Free Energy of the Model

4.1 Entropy, Energy and Free Energy

As always in statistical mechanics the entropy is $H = k \ln W$, where W here is the number of arrangements $\{N_\Delta\}$ of the N layers of the model that lead to the same deformation D. Thus, we have

$$H = k \ln \frac{N!}{\prod_\Delta N_\Delta!} \qquad \text{with} \quad N = \sum_\Delta N_\Delta \quad \text{and} \quad D = \frac{1}{\sqrt{2}} \sum_\Delta \Delta N_\Delta. \tag{4.1}$$

Assuming all N_Δ to be much bigger than 1, we may use the Stirling formula

$\ell n\, a! = a\, \ell n\, a - a$ and write

$$H = -k \sum_{\Delta} (\ell n \frac{N_\Delta}{N}) N_\Delta . \tag{4.2}$$

The energy of the layers is obviously

$$E_L = \sum_{\Delta} \Phi(\Delta) N_\Delta , \tag{4.3}$$

where N_Δ is subject to the constraints $(1)_{2,3}$. There is also the contribution of the M-A interfaces and if there are K such interfaces, that contribution is

$$E_I = e \cdot K. \tag{4.4}$$

The free energy $\psi = (E_L + E_I) - TH$ of the model thus reads

$$\psi = \sum_{\Delta} (\Phi(\Delta) + kT \ell n \frac{N_\Delta}{N}) N_\Delta + eK. \tag{4.5}$$

The free energy in thermodynamic equilibrium follows from (5) by insertion of the distribution N_Δ that minimizes ψ under the constraints $(1)_{2,3}$. Unfortunately that minimization cannot be carried out before we know how the number K of interfaces depends on the distribution $\{N_\Delta\}$.

4.2 The Most Probable Number K of M - A Interfaces.

It is clear that there is a correlation, albeit a weak one, between the distribution $\{N_\Delta\}$ and the number K of interfaces. For instance, if $x_A = \frac{1}{N} \sum_{[\,]} N_\Delta$ is small, there cannot be many M-A interfaces but there may be differently many. The number K can range from 1 to $2N_A$, where $N_A = Nx_A$. +)

+) $2N_A$ is the maximal number, if the body begins and ends with a M-layer. If either one side or both are occupied by A-layers, the maximal number is $2N_A - 1$ and $2N_A - 2$ respectively. We ignore such fine points in the sequel.

However, not all these numbers are equally probable. In this section we shall show that the most probable number K is uniquely determined by the distribution $\{N_\Delta\}$.

We begin by determining the probability of the number K for fixed values of $N_A = Nx_A$ and $N_M = Mx_M$. Consider: If there are K interfaces, there are $R_A = \frac{K}{2}$ austenistic regions and $R_M = \frac{K}{2}$ martensitic ones, each generally consisting of several layers. [The factor 1/2 comes about, because each region is bounded by two interfaces.] Now we ask for the number of possibilities to split N_A layers into R_A regions. The answer is $\binom{N_A-1}{R_A-1}$. +) Similarly there are $\binom{N_M-1}{R_M-1}$ possibilities to distribute the N_M layers over the R_M regions. It follows that the number of possibilities for the realization of K interfaces among N_A and N_M austenistic and martensitic layers respectively is equal to

$$W_K = \binom{N_A - 1}{R_A - 1}\binom{N_M - 1}{R_M - 1}$$

or with $R_A = R_M = \frac{K}{2}$ and $N_M = N - N_A$

$$W_K = \binom{N_A - 1}{\frac{K}{2} - 1}\binom{N - 1 - N_A}{\frac{K}{2} - 1} .$$

If W_K is big for a given number K_1 and small for K_2, we expect to have K_1 interfaces more probably than K_2. Indeed we expect to have <u>the</u> value K <u>most</u> probably for which W_K is maximal, viz.

$$\begin{aligned} K_{m.p} &= \frac{2}{N} N_A(N - N_A), \\ K_{m.p} &= 2Nx_A(1 - x_A), \end{aligned} \tag{4.7}$$

where the Stirling formula has been used.

+) It may help the reader to realize that this combinatorial question is equivalent to the following one: How many possibilities are there for distributing N_A particles over R_A cells so that no cell remains empty?

If the maximum of W_K is steep, it is safe to assume that $K_{m.p}$ is the number of interfaces seen exclusively rather than most probably. We proceed on that assumption.

Although equation (7) was obtained by a somewhat complex argument, it is easy to interpret. Indeed obviously K must vanish in the pure phases, i.e. at $x_A = 0$ and $x_A = 1$. And we intuitively expect K to be maximal between those limits at $x_A = \frac{1}{2}$ and that expectation is born out by (7).

In order to anticipate a possible criticism of the derivation of (7) I should like to add this: The purely probabilistic determination of K, which we have seen, might not be the correct way. Indeed, one could argue that certain configurations, even though they are very probable by our count, are nevertheless seldom seen, because they are energetically unfavourable. I have tried to take such energetic effects into account, but was unable to carry the calculations through to a specific end.

4.3 Free Energy in Equilibrium

By the consideration of the previous section we now replace K in (5) by the value $K_{m.p}$ given in (7) and we obtain by use of (3.4) and (3.5)

$$\psi = \sum_{\Delta}(\Phi(\Delta) + \kappa T \ell n \frac{N_\Delta}{N})N_A + 2\frac{e}{N}(\sum_{][} N_\Delta)(\sum_{[]} N_\Delta). \qquad (4.8)$$

Thus we have ψ as a simple function of the individual N_Δ's and we can proceed to find the equilibrium distribution $\{N_\Delta\}$, i.e. the one that minimizes ψ under the constraints $(1)_{2,3}$.

The minimalization proceeds in a standard manner by use of Lagrange multipliers and we obtain

$$\frac{N_\Delta}{N} = e^{\frac{\alpha+\beta\Delta}{kT}-1}\, e^{-\frac{\Phi(\Delta)+2ex_A}{kT}} \quad \text{for} \quad \Delta \in][$$

$$\frac{N_\Delta}{N} = e^{\frac{\alpha+\beta\Delta}{kT}-1}\, e^{-\frac{\Phi(\Delta)+2ex_M}{kT}} \quad \text{for} \quad \Delta \in [] \tag{4.9}$$

α and β are the Lagrange multipliers for the constraints $(1)_2$ and $(1)_3$ respectively.

Insertion into (8) gives the equilibrium free energy

$$\psi = (\alpha - kT)N + \beta\sqrt{2}\,D - 2Nex_Ax_M. \tag{4.10}$$

The Lagrange multipliers must be calculated from $(1)_{2,3}$ by insertion of (9) and we obtain

$$\alpha - kT = -kT\,\ell n\left(e^{-\frac{2ex_A}{kT}} \sum_{][} e^{\frac{\beta\Delta-\Phi(\Delta)}{kT}} + e^{-\frac{2ex_M}{kT}} \sum_{[]} e^{\frac{\beta\Delta-\Phi(\Delta)}{kT}}\right) \tag{4.11}$$

$$\sqrt{2}\,\frac{D}{N} = \frac{e^{-\frac{2ex_A}{kT}} \sum_{[]} \Delta e^{\frac{\beta\Delta-\Phi(\Delta)}{kT}} + e^{-\frac{2ex_M}{kT}} \sum_{[]} \Delta e^{\frac{\beta\Delta-\Phi(\Delta)}{kT}}}{e^{-\frac{2ex_A}{kT}} \sum_{][} e^{\frac{\beta\Delta-\Phi(\Delta)}{kT}} + e^{-\frac{2ex_M}{kT}} \sum_{[]} e^{\frac{\beta\Delta-\Phi(\Delta)}{kT}}} \tag{4.12}$$

Note that from $(9)_1$ we obtain by summation over all $\Delta \in][$

$$x_M = \frac{e^{-\frac{2ex_A}{kT}} \sum_{[]} \Delta e^{\frac{\beta\Delta-\Phi(\Delta)}{kT}}}{e^{-\frac{2ex_A}{kT}} \sum_{][} e^{\frac{\beta\Delta-\Phi(\Delta)}{kT}} + e^{-\frac{2ex_M}{kT}} \sum_{[]} e^{\frac{\beta\Delta-\Phi(\Delta)}{kT}}} \tag{4.13}$$

In principle (13) gives $x_M = x_M(\beta,T)$, insertion of that function into (12), gives $\beta = \beta(D,T)$, again in principle. Thus we have x_M and, by (11), α as

functions of D and T. From (10) we may then calculate $\psi = \psi(D,T)$, the free energy in equilibrium.

4.4 Reformulation of Problem in Dimensionless Variables.

We replace the sums in (11) through (13) by integrals with the assumption that the number of possible shear lengths between Δ and $\Delta + d\Delta$ be equal to $zd\Delta$, where z is some constant of no interest in this paper. Also we introduce dimensionless variables by the definitions

$$d = \frac{\sqrt{2D}}{NJ}\ ,\ \delta = \frac{\Delta}{J}\ ,\ a = \frac{kT}{\Phi_s}\ ,\ \phi = \frac{\Phi}{\Phi_s}\ ,\ \varepsilon = \frac{2e}{\Phi_s}\ ,\ p = \frac{\beta J}{\Phi_s}\ ,\ \psi = \frac{\psi + NkT\ \ell n(zJ)}{N\Phi_s}\ . \quad (4.14)$$

By use of these definitions and by elimination of α between (10) and (11) we derive a fairly concise form of our equations (10), (12) and (13), viz.

$$\psi = -a\ \ell n\left[e^{-\frac{\varepsilon}{a}x_A}\int_{][} e^{\frac{1}{a}(p\delta-\phi(\delta))}d\delta + e^{-\frac{\varepsilon}{a}x_M}\int_{[]} e^{\frac{1}{a}(p\delta-\phi\delta))}d\delta\right] + pd - \varepsilon x_A x_M, \quad (4.15)$$

$$d = \frac{e^{-\frac{\varepsilon}{a}x_A}\int_{][} \delta e^{\frac{1}{a}(p\delta-\phi(\delta))}d\delta + e^{-\frac{\varepsilon}{a}x_M}\int_{[]} \delta e^{\frac{1}{a}(p\delta-\phi(\delta))}d\delta}{e^{-\frac{\varepsilon}{a}x_A}\int_{][} e^{\frac{1}{a}(p\delta-\phi(\delta))}d\delta + e^{-\frac{\varepsilon}{a}x_M}\int_{[]} e^{\frac{1}{a}(p\delta-\phi(\delta))}d\delta} \quad (4.16)$$

$$x_M = \frac{e^{-\frac{\varepsilon}{a}x_A}\int_{][} \delta e^{\frac{1}{a}(p\delta-\phi(\delta))}d\delta}{e^{-\frac{\varepsilon}{a}x_A}\int_{][} e^{\frac{1}{a}(p\delta-\phi(\delta))}d\delta + e^{-\frac{\varepsilon}{a}x_M}\int_{[]} e^{\frac{1}{a}(p\delta-\phi(\delta))}d\delta} \quad (4.17)$$

It is with these equations that we shall proceed to work. In particular, from these equations we shall derive the form of the free energy $\psi = \psi(d,a)$ for the potential shown in Figure 3. Note that ψ,d, and a are the dimensionless

expressions for free energy, deformation and temperature respectively.

4.5 Load and Load-Deformation Curves

We recall from thermodynamics that $\left(\frac{\partial \psi}{\partial d}\right)_a$ is the load on the body; and from (15) through (17) we obtain

$$\left(\frac{\partial \psi}{\partial d}\right)_a = p \tag{4.18}$$

as a brief calculation shows. Thus the quantity p, which sofar was introduced as the dimensionless Lagrange multiplier p, is infact the load. It is therefore possible to use the equations (16) and (17) to derive load-deformation curves. Indeed, (17) gives $x_M = x_M(p,a)$ and, if that function is introduced we obtain the functions $p = p(d,a)$ which are of particular interest, since we wish to understand the curves of Figure 1.

4.6 Equilibria and Frozen Equilibria

Note that we let the distribution $\{N_\Delta\}$ be determined by deformation and temperature (see (9) with (11) through (13)). In particular, by (17), x_M is determined by p and a, or by d and a. This is the consequence of admitting only such distributions that minimize the free energy and this in turn means that we consider fairly high temperatures. Indeed, such temperatures and the corresponding fluctuations will permit the lattice particles to overcome the energetic barriers between the phases so that states of different free energy do compete and the body can assume the state of minimum free energy, i.e. equilibrium.

Consider the alternative of low temperature: In that case the lattice particles are "frozen" into the deepest potential wells which are the martensitic ones. The phase fractions x_{M_+} and x_{M_-} are then determined by the initial

distribution rather than by the trend to assume a small free energy. States of different free energy cannot compete, and we can change the fractions $x_{M_\pm}$ only by brute force. This case was adequately described by Müller & Wilmanski in [3], it is the case of frozen equilibrium.

The question is at what temperature the dividing line between equilibrium and frozen equilibrium should be drawn. A natural choice for this temperature is the one that allows an unloaded body to be austenitic, because that phenomenon proves that the lattice particles can fluctuate out of their martensitic potential wells. This temperature will be called T_K and the corresponding dimensionless temperature is a_K.

Since we are going to use the relations derived earlier in this chapter, we are limiting the attention to the range $a > a_K$. And as the condition is that the unloaded body be austenitic, we shall be in the pseudoelastic range (see chapter 2).

5. Isotherms in the (p,d)-Diagram and the (ψ,d)-Diagram

5.1. The Necessity of Interfacial Energies for Pseudoelastic Behaviour.

We do not expect to derive load-deformation curves of the form shown in Figure 1c. and 1d. which exhibit a hysteresis loop. What we do expect, if indeed we are successful, are non-monotone load-deformation curves which will imply hysteretic behaviour with a simple additional argument (see Section 5.3). This expectation is motivated by experience with other statistical models of phase transitions and with phenomenological models of the Landau-Devonshire type (see [4]). The prototype of a non-monotone load deformation curve in statistical mechanics is, of course furnished by the van der Waals equations.

Let us investigate the possibility of having non-monotone load-deformation curves: We define the function

$$\Phi(\rho,d;a) = e^{-\frac{\varepsilon}{a}x_A}\int_{]L}(\delta - d)e^{\frac{1}{a}(\rho\delta-\phi(\delta))}d\delta + e^{-\frac{\varepsilon}{a}x_M}\int_{[]}(\delta - d)e^{\frac{1}{a}(\rho\delta-\phi(\delta))}d\delta. \quad (5.1)$$

This is a function of ρ, d and a as indicated and by (4.16) the vanishing of this function defines the (ρ,d)-isotherm. Keeping a fixed we form the differential

$$d\Phi = -(e^{-\frac{\varepsilon}{a}x_A}\int_{]L}e^{\frac{1}{a}(\rho\delta-\phi(\delta))}d\delta + e^{-\frac{\varepsilon}{a}x_M}\int_{[]}e^{\frac{1}{a}(\rho\delta-\phi(\delta))}d\delta)dd +$$

$$+ [\frac{1}{a}(e^{-\frac{\varepsilon}{a}x_A}\int_{]L}(\delta-d)^2e^{\frac{1}{a}(\rho\delta-\phi(\delta))}d\delta + e^{-\frac{\varepsilon}{a}x_M}\int_{[]}(\delta-d)^2e^{\frac{1}{a}(\rho\delta-\phi(\delta))}d\delta) +$$

$$+ \frac{d}{a}(e^{-\frac{2}{a}x_A}\int_{]L}(\delta-d)e^{\frac{1}{a}(\rho\delta-\phi(\delta))}d\delta + e^{-\frac{\varepsilon}{a}x_M}\int_{[]}(\delta-d)e^{\frac{1}{a}(\rho\delta-\phi(\delta))}d\delta) +$$

$$+ \frac{\varepsilon}{a}(e^{-\frac{\varepsilon}{a}x_A}\int_{]L}(\delta-d)e^{\frac{1}{a}(\rho\delta-\phi(\delta))}d\delta - e^{-\frac{\varepsilon}{a}x_M}\int_{[]}(\delta-d)e^{\frac{1}{a}(\rho\delta-\phi(\delta))}d\delta)(\frac{\partial x_M}{\partial\rho})_a]d\rho.$$

For $\Phi \equiv 0$ we obtain an expression for the derivative of the (ρ,d) isotherms, viz.

$$(\frac{\partial\rho}{\partial d})_a = \frac{e^{\frac{\varepsilon}{a}x_A}\int_{]L}e^{\frac{1}{a}(\,)}d\delta + e^{-\frac{\varepsilon}{a}x_M}\int_{[]}e^{\frac{1}{a}(\,)}d\delta}{\frac{1}{a}(e^{-\frac{\varepsilon}{a}x_A}\int_{]L}(\delta-d)^2e^{\frac{1}{a}(\,)}d\delta + e^{-\frac{\varepsilon}{a}x_M}\int_{[]}(\delta-d)^2e^{\frac{1}{a}(\,)}d\delta) + \frac{\varepsilon}{a}(e^{-\frac{\varepsilon}{a}x_a}\int_{]L}(\delta-d)e^{\frac{1}{a}(\,)}d\delta + e^{-\frac{\varepsilon}{a}x_M}\int_{[]}(\delta-d)e^{\frac{1}{a}(\,)}d\delta)(\frac{\partial x_M}{\partial\rho})_a} \quad (5.2)$$

Inspection shows that for $\varepsilon = 0$ the right hand side is always positive, so that the load-deformation curve is monotonically increasing. This is not what we need, and we conclude that a non-zero interfacial energy is essential for pseudoelasticity. +)

+) Müller & Wilmanski [3] did obtain non-monotone load-deformation curves without interfacial energies. The present agrument shows that their result was an artifact created by the approximate evaluation of the particular function which involved a kind of virial expansion.

For a non-zero interfacial energy $\left(\frac{\partial p}{\partial d}\right)_a$ may vanish, provided that $\left(\frac{\partial x_M}{\partial p}\right)_a$ has singularities. This is indeed the case in a certain range of temperature as we shall presently see. The equation that determines the dependence of x_M on p and a is (4.17) which, in a neater form may be written as

$$\frac{\int_{[]} e^{\frac{1}{a}(\,)}\, d\delta}{\int_{[]} e^{a(\,)}\, d\delta} = \frac{x_M}{1-x_M}\, e^{\frac{\varepsilon}{a}(1 - 2x_M)} \tag{5.3}$$

where the p-dependent terms are nicely separated from the x_M-dependent ones. Obviously for $\left(\frac{\partial x_M}{\partial p}\right)_a$ to become singular the right hand side of (3) must have extrema for $0 \leq x_M \leq 1$. And indeed, it has such extrema at

$$x_M^{\pm} = \frac{1}{2}\left(1 \pm \sqrt{1 - \frac{2a}{\varepsilon}}\right) - \begin{matrix}\text{minimum} \\ \text{maximum}\end{matrix}, \tag{5.4}$$

provided that we have

$$\varepsilon \geq 2a. \tag{5.5}$$

We conclude from the definitions (4.14) that for non-monotone load-deformation isotherms we must have interfacial energies e greater than kT. Or else, if T grows above $\frac{e}{k}$, the non-monotone isotherms disappear and consequently there is no longer pseudoelasticity.

We proceed to look into this question in more detail.

5.2 Construction of Load-Deformation Isotherms

As described above, for the determination of the functions $x_M = x_M(p,a)$ we rely on (4.17) or on its more elegant version (3). Indeed (3) lends itself for an easy graphical evaluation that is indicated in Figure 5.

On the right hand side of Figure 5 we have drawn.

$$\frac{x_M}{1 - x_m} e^{\frac{\varepsilon}{a}(1-x_M)} \quad \text{as a function of } x_M \tag{5.6}$$

for $\varepsilon = 0.75$ and for various values of a, all of them satisfying (5). On the left hand side of Figure 5 we have drawn

$$\frac{\int_{[} e^{\frac{1}{a}(\,)} d\delta}{\int_{[} e^{a(\,)} d\delta} \quad \text{as a function of } p. \tag{5.7}$$

For the calculation of the integrals in (7) we have relied on the simple box potential of Figure 3 and chosen

$$S = 0.2, \quad T = 0.9, \quad X = 0.15 \tag{5.8}$$

in that potential.

The manner of exploitation of Figure 5 is indicated by the construction of the lines with arrows: We choose a value of p and find the value of the quotient (7) by going vertically upwards until we intersect the curve pertaining to the selected a. From that point we move horizontally to the right. That horizontal line will generally intersect the curves of the function (6) in three points, whose absissae are the values x_M pertaining to the choosen value of p. Figure 6 gives a plot of x_M as funtion of p for different temperatures.

Once $x_M(p,a)$ is known we can insert it into (4.16) and obtain $d(p,a)$, or indeed $p = p(d,a)$, the load displacement isotherms. Figure 7 shows the resulting curves. Of course it is just as simple to introduce $x_M(p,a)$ into (4.15) and obtain $\psi = \psi(d,a)$, the free energy. These curves are plotted in Figure 8, they have been moved vertically so that all of them start at the origin.

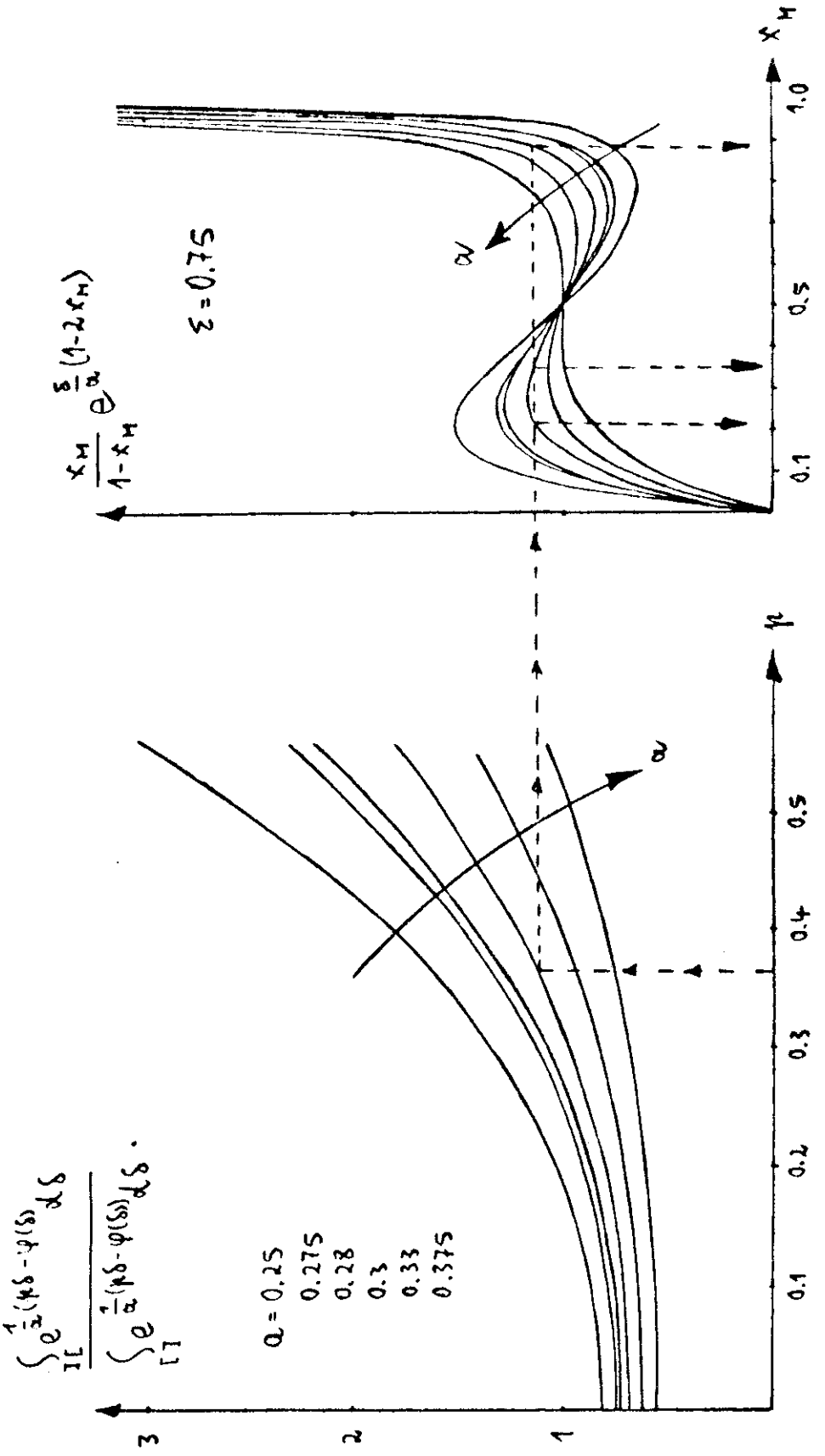
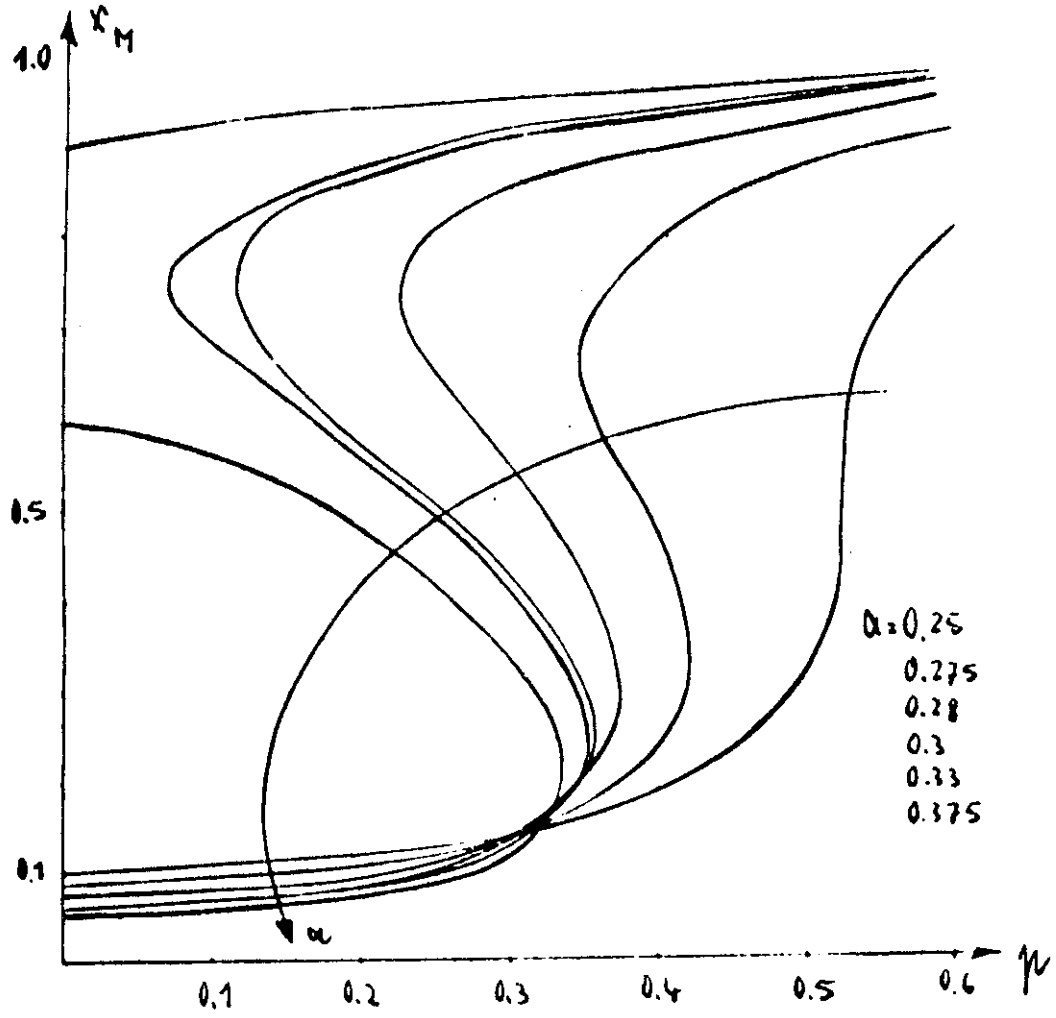


Figure 5: On the construction of (p,d)-isotherms



Figuer 6: x_M as a funtion of ρ and α.

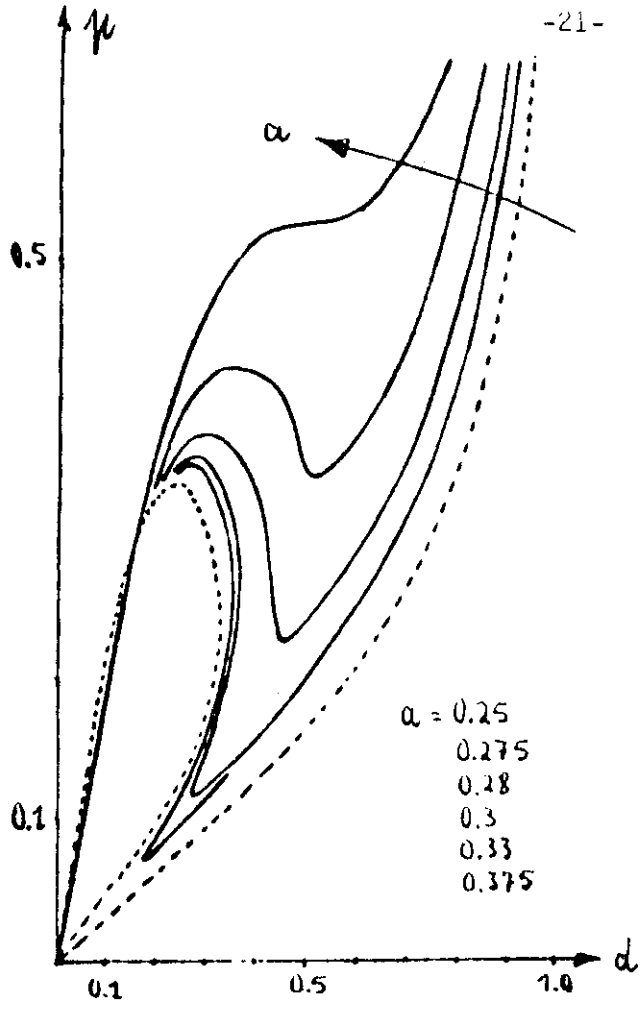


Figure 7: (ρ,d)-isotnerms

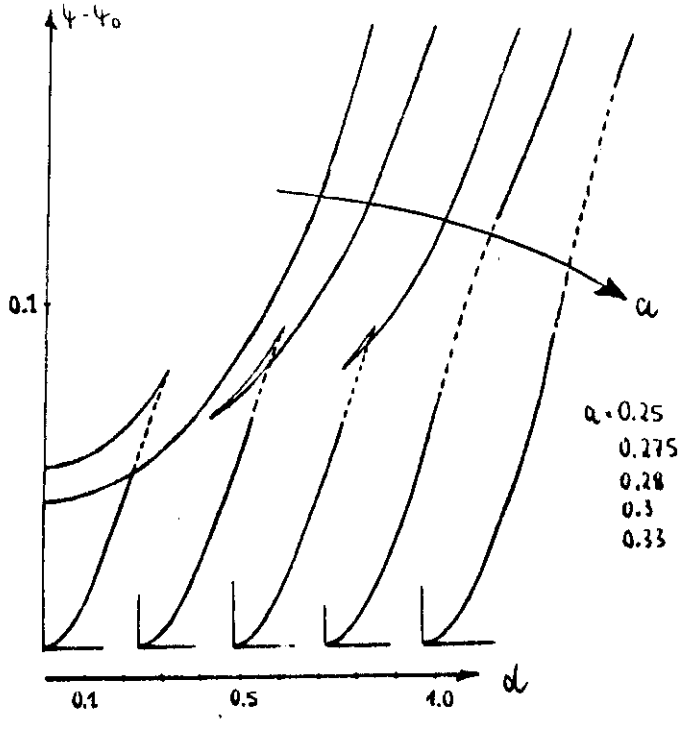


Figure 8: ψ as a function of d and a

5.3 Discussion of Load-Deformation Isotherms in Load Controlled Experiments

The solid curves of Figure 7 are the load-deformation isotherms of interest because they exhibit pseudoelastic behaviour in an experiment where the load is prescribed. Indeed, as we start to load we move up from the origin of the (p,d)-diagram along one of the solid curves. The body here has but little martensite, as Figure 6 shows, it is virtually all austenitic. Increasing the load we eventually reach the maximum and, upon a slight further increase, there will be a breakthrough to the right branch of the isotherm. On that branch we may either move up or down until we reach the minimum. Everywhere on the right branch the body is predominantly martensitic as Figure 6 shows so that the breakthrough has been accompanied by a change of phase. If we are at the minimum, a slight further decrease of load will cause another breakthrough back to the austenitic phase.

Figure 9 illustrates this behaviour by presenting the isotherms for $a = 0.28$, $a = 0.3$ and $a = 0.33$, complete with the breakthrough lines. Thus we see hysteretic behaviour just as shown schematically in Figure 1. Indeed the hysteresis loop is lifted up at higher temperature and it becomes smaller. According to the theory of course the hysteresis should vanish at $a = 0.375$ but I believe that this has not been observed.

If instead of prescribing the load, we prescribe the deformation it is much less clear from this theory what we expect to see in the load-deformation diagram. One thing, however, is clear, namely that the body cannot assume states on the branches of the isotherms that have negative slopes, because those branches are thermodynamically unstable. We shall discuss the possible states inside the hysteresis loops of Figure 9 in some detail in the section 5.5.

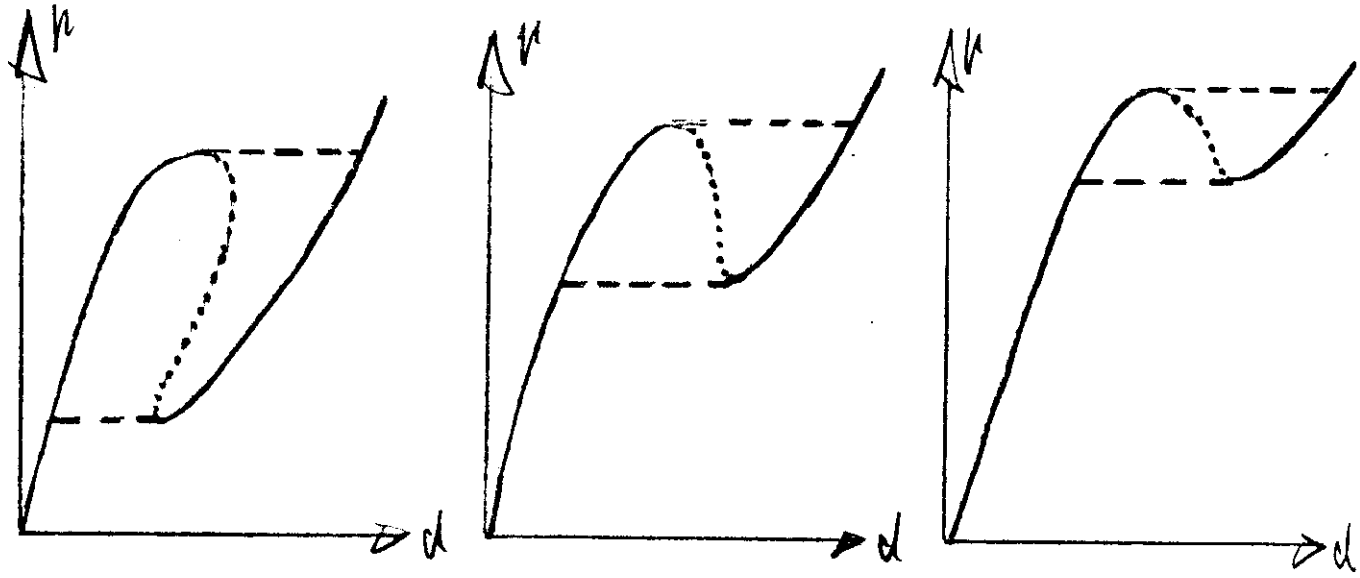


Figure 9: Pseudoelastic behaviour of the model.

The dashed lines in Figure 7 correspond to the low temperature $a = 0.25$. We see that the right branch of that curve, representing a nearly fully martensitic body, extends all the way down to the origin. I.e. at that temperature the body does not form austenite in the unloaded state. We concluded before in Section 4.6 that under those circumstances we would not consider the theory applicable any more. We decided there that the theory should be good for all $a > a_K$ for which austenite forms upon unloading. We must conclude from Figure 7 that

$$0.25 < a_K < 0.28.$$

We can do better than that, of course, in the determination of a_K. Indeed, from the construction of the Figure 5 through 7 it is clear that a_K is the value of a which satisfies (3), when in that equation p is set equal to zero and x_M is set equal to the x_M^+ of equation (4). In that case the curve on

the left hand side of Figure 5 starts out at the same height as the minimum on the right hand side of that Figure. The condition for a_k thus reads

$$\frac{\int_{[} e^{-\frac{1}{a}\phi(\delta)} d\delta}{\int_{[} e^{\frac{1}{a}\phi(\delta)} d\delta} = \frac{1 + \sqrt{1 - \frac{2\varepsilon}{a}}}{1 - \sqrt{1 - \frac{2\varepsilon}{a}}} e^{-\frac{\varepsilon}{a}\sqrt{1 - \frac{2a}{\varepsilon}}} \tag{5.9}$$

For $\varepsilon = 0.75$ and for the box potential of Figure 3 with the parameter choices (8) one can find a_k by numerically solving (9). We obtain

$$a_k = 0.2716 \ldots \tag{5.10}$$

Although curves for $a < a_k$ are useless for the study of pseudoelasticity, I should like to point out one satisfactory feature of the outer branch of the dashed curve of Figure 7. This branch runs into the origin even though the body is nearly all martensitic. This means that it must neatly fall apart into half M_+ and half M_-. Indeed this is what is to be expected, since at low temperatures the lattice particles are caught in the lateral potential wells and, as they settle down to this, in all probability half will be going into the right well and the other half to the left.

5.4 Discussion of Free Energy-Deformation Isotherms

The free energy curves of Figure 8 deserve a comment, since three of them are non-standard. One of them is the one for $a = 0.25$; this one does not really belong here, since $a < a_k$ on it and we shall not discuss it.

The free energy curves corresponding to $a = 0.275$ and $n = 0.28$ are odd in that they contain a loop. To my knowledge curves like these have not recieved any attention in the literature and maybe we should study them a little closer with the objective in mind of saying something about phase equilibrium in that case. Of course, the derivatives of these functions are the (p,d)-curves that "double back".

The free energy curves for $a = 0.3$ and $a = 0.33$ are typical for first order phase transitions. The dashed parts are ranges of negative curvature, a phenomenon that was first detected for the van der Waals gas and is now best known for free energy functions of the Landau-Devonshire type (e.g. see [5]). Curves of this type lend themselves to a suggestive interpretation of phase equilibrium because the body can assume a lower free energy by moving its state along the convex envelope of the function rather than along the function itself. Along the straight part of that envelope the deformation would then change by changing the fractions of two phases in equilibrium at the fixed deformations d_1 and d_2 (see Figure 10). The construction of the convex envelope corresponds to the construction of the Maxwell line in the (p,d)-diagram which may be known more widely.

While the construction of the convex envelope, or of the Maxwell line, is very suggestive, it is also not relevant in the case of pseudoelasticity, because otherwise we should not see hystereses. Indeed it seems that in load controlled experiments - the body traces out fully the lower convex part of the (ψ,d)-curve and thus jumps to a point on the upper convex branch where the slope is equal to the maximum slope of the lower branch.

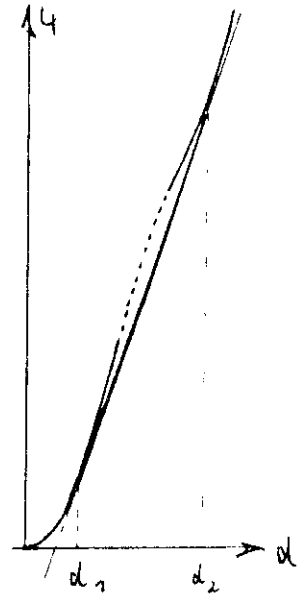


Figure 10. Free energy-deformation isotherm and its convex envelope.

5.5 Metastable States

I have already stated that the states on the downward sloping branches of the (p,d)-curves are thermodynamically unstable. So the question arises of what we see in a deformation controlled experiment as we pass that range of d values.

We may get an inkling of the answer to that question, when we realize that the (p,d)-diagram is full of curves with metastable states. Figure 11 illustrates that by showing some such curves in the (p,d)-diagram for a = 0.3. These curves have been constructed by ignoring the fact that (4.17) determines an equilibrium value of x_M, once p and a are given. Indeed, I have just taken a fixed value of x_M and calculated $d = d(p,a;x_M)$ from (4.16).

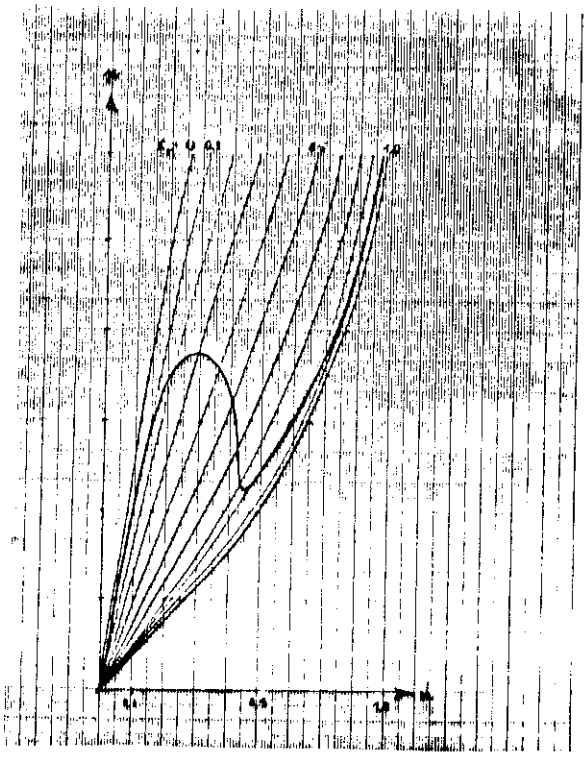


Figure 11: Stable, unstable and metastable states.

Thus we see that for a given value of d in the unstable range the body can still sit on a curve with a positive slope. Infact it can choose from a variety of such curves that differ only by the phase factor x_M. Since phase equilibrium along the Maxwell line is not established in load controlled experiments, it is unlikely to be established in deformation controlled ones. Thus the most likely behaviour is that the body will grate along a horizontal line through maximum or minimum respectively as the deformation is increased or lowered. A plot of the (p,d)-curve taken by a "hard-device" testing machine would then look as schematically shown in Figure 12. Indeed I have seen such behaviour described in the literature. The serration of the "horizontal" parts is due to the fact, that when one or several layers flip over, the body does relax a little and, before the next flip can occur, the body must be reloaded along the metastable lines of Figure 11.

It is not clear to me that a deformation-controlled experiment must always show a (p,d) curve like the one of Figure 12. It seems conceivable that some softening occurs along the upper serrated branch. Such a softening should be more pronounced on the right hand side of that branch, because that is the range where the interfacial energy decreases. However, I am unaware of any experimental results that show softening of pseudoelastic bodies.

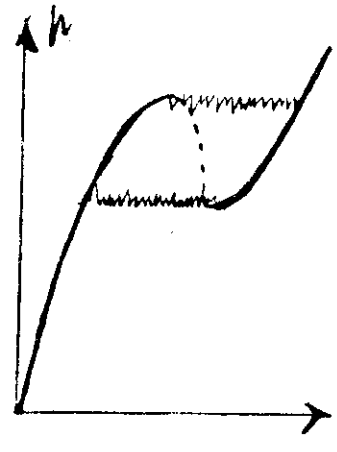


Figure 12: Load-deformation isotherm under deformation control.

REFERENCES

[1] Perkins, J. (ed.) - Shape Memory Effects in Alloys. New York, London; Plenum Press, 1975.

[2] Delaey, L., Chandrasekaran, L. (eds.): Proceedings Int. Conf. on Martensitic Transformation, Leuven 1982; J. de Physique 43 (1982).

[3] Müller, I., Wilmanski, K.: Memory Alloys, Phenomenology and Ersatz model. In O. Brulin, RKT Hsieh (eds) Continuum Models of Discrete Systems 4, North Holland Publ. Co. (1981).

[4] Achenbach, M., Atanackovic, T., Müller, I. A Model for Memory Alloys in Plane Strain. Int. J. of Solids and Structures (in press).

[5] Devonshire, A.F. Adv. Phys 3 (1954).

Acknowledgements: I gratefully acknowledge useful discussions of some ideas of this paper with Mr. M. Achenbach and Mr. Rebstock. The paper was conceived and written when I was a research fellow of the Institute for Mathematics and its Applications, University of Minnesota. The support of the Institute is gratefully acknowledged.