

Engineering Aspects of Shape Memory Alloys

T W Duerig, BS, ME, PhD

Technical Director, Raychem Corporation, Menlo Park, California, USA

K N Melton, BA, PhD

Director of Research and Development, Raychem Corporation, Swindon, UK

D Stöckel, Dip Ing, PhD

Technical Marketing Manager, Raychem Corporation, Menlo Park, California, USA

C M Wayman, BS, MS, PhD

Professor of Materials Science and Engineering, University of Illinois, Urbana,
Illinois, USA

Butterworth-Heinemann

London Boston Singapore Sydney Toronto Wellington



PART OF REED INTERNATIONAL P.L.C.

All rights reserved. No part of this publication may be reproduced in any material form (including photocopying or storing it in any medium by electronic means and whether or not transiently or incidentally to some other use of this publication) without the written permission of the copyright owner except in accordance with the provisions of the Copyright, Designs and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency Ltd, 33–34 Alfred Place, London, England WC1E 7DP. Applications for the copyright owner's written permission to reproduce any part of this publication should be addressed to the Publishers.

Warning: The doing of an unauthorised act in relation to a copyright work may result in both a civil claim for damages and criminal prosecution.

This book is sold subject to the Standard Conditions of Sale of Net Books and may not be re-sold in the UK below the net price given by the Publishers in their current price list.

First published 1990

© Butterworth-Heinemann Ltd, 1990

British Library Cataloguing in Publication Data

Engineering aspects of shape memory alloys.

1. Alloys. Shape memory effects

I. Duerig, T. W.

669.94

ISBN 0-750-61009-3

Library of Congress Cataloging-in-Publication Data

Engineering aspects of shape memory alloys / edited by T.W. Duerig.

p. cm.

Includes bibliographical references and index.

ISBN 0-750-61009-3

1. Alloys--Thermomechanical properties. 2. Shape memory effect.

I. Duerig, T. W.

TN6980.E546 1990

669'.9--dc20

90-2104

CIP

Introduction

It has now been over 50 years since the first observations of shape memory, and over 20 years since people first began to find applications for the effect. Certainly many people believe that practical application has progressed much slower than expected: when inventive people first observe the effect they immediately begin to conjure ideas for its application, amazed that it has been known for so long by the scientific community and is yet nearly unknown to design engineers. Shape memory has even become famous as "a solution looking for a problem".

But this reputation is unfair if one considers that the entire technology is new. These are not simply new alloys of steel or titanium, but represent an entirely new philosophy of engineering and design. The most fundamental property descriptors are different: yield strength, modulus, and ductility are replaced by stress rate, recovery stress and M_s . Moreover, product designs using shape memory are generally not evolutionary, but revolutionary in nature. One can hardly expect large industries to immediately convert basic product designs to shape memory. In fact progress has been impressive. At the time of writing, it is estimated that the worldwide business in shape memory exceeds 30 million US dollars, and is growing at over 25% per year. Product diversity is also most impressive, including fine medical wires, electrical switches, eyeglass frames, appliance controllers, pipe couplings and electronic connectors.

Still, it is clear that the technology lags behind the science. The origin and mechanism of shape memory are now well understood, but many of the engineering aspects are not. The purpose of this book is to extend our understanding of shape memory by defining terms, properties and applications. It includes tutorials, overviews, and specific design examples - all written with the intention of minimizing the science and maximizing the engineering aspects. Although the individual chapters have been written by many different authors, each one of the best in their fields, the overall tone and intent of the book is not that of a proceedings, but that of a textbook. There has been a concerted editing effort to unify terms, avoid duplication, and fill gaps.

Shape memory applications can generally be divided into four categories: free recovery, constrained recovery, work production (actuators) and superelasticity. These groupings are made according to the primary function of the memory element, but are useful in defining common product screening criteria, pitfalls and engineering design parameters. We define the groups as follows:

1. *Free recovery* includes applications in which the sole function of the memory element is to cause motion or strain. For example, one could cool a wire into the martensitic regime, bend it to a new shape, then heat to recover the original shape.
2. *Constrained recovery* includes applications in which the memory element is prevented from changing shape and thereby generates a stress. The ideal example would be the recovery of a ring onto a solid, rigid rod. In this case there may be some free recovery before contact is made, but the primary function of the memory element is to generate a stress.
3. *Actuator, or work production* applications are those in which there is motion against a stress and thus work is being done. The ideal case would be a wire or spring which lifts a weight when heated (and perhaps drops the weight again when cooled). More often, the memory element works against a biasing spring. Actuators are generally of two types: *thermal* and *electrical*. Electrical actuators are actuated via direct current and are generally in competition with servo-motors or solenoid. Thermal actuators are driven by changes in ambient temperature and generally compete with thermostatic bimetals.
4. *Superelastic or pseudoelastic* applications are isothermal in nature and involve the storage of potential energy. Although memory elements can operate as "supersprings", the temperature range over which the effect is found is rather narrow (usually only 80°C).

The book consists of five parts. Part I deals with the mechanism of shape memory and the alloys that exhibit the effect. It also defines many essential terms that will be used in later parts. Part II deals primarily with constrained recovery, but to some extent with free recovery. There is an introductory paper which defines terms and principles, then several specific examples of products based on constrained recovery. Parts III and IV deal with actuators, with part III introducing engineering principles and part four several of the specific examples. Finally part V deals with superelasticity, again with an introductory paper and then several specific examples of product engineering.

Acknowledgements

This volume is based on contributions to the Shape Memory Alloy technology conference held at Michigan State University on 15–17 August 1988. The editors wish to acknowledge the following firms who sponsored the conference:

Raychem Corporation
300 Constitution Drive
Menlo Park, CA 94025

Memry Inc
83 Keeler Avenue
Norwalk, CT 06854

An Introduction to Martensite and Shape Memory

C.M. Wayman^{*} and T.W. Duerig[†]

^{*}Department of Materials of Science and Engineering
University of Illinois at Urbana-Champaign,
1304 West Green Street
Urbana, Illinois 61801

[†]Raychem Corporation
300 Constitution Dr.,
Menlo Park, CA 94025

Shape memory refers to the ability of certain materials to "remember" a shape, even after rather severe deformations: once deformed at low temperatures (in their martensitic phase), these materials will stay deformed until heated, whereupon they will spontaneously return to their original, pre-deformation shape. The basis for the memory effect is that the materials can easily transform to and from martensite. A detailed description of martensitic transformations¹⁻⁴ can be very complex and is beyond the scope of this book; still, one cannot obtain an understanding of even the most elementary engineering aspects of shape memory without first becoming familiar with some basic principles of martensite and its formation. In this section, the key microscopic and macroscopic aspects of martensite will be qualitatively reviewed. Although one may be tempted to pass over some of the microscopic aspects, it must be emphasized that these form a foundation upon which subsequent discussions of shape memory are based. Many of the terms used later in the book will also be defined. After an introduction to martensite, the shape memory event itself will be introduced, along with the related phenomenon called "superelasticity".

1. Martensite: A Microscopic Perspective

Solid state transformations are of two types: diffusional and displacive. *Diffusional transformations* are those in which a new phase can only be formed by moving atoms randomly over relatively long distances. Long range diffusion is required because the new phase is of a different chemical composition than the matrix from which it is formed. Since atomic migration is required, the progress of this type of transformation is dependent upon both time and temperature. In contrast, *displacive transformations* do not require such long range movements; in these cases atoms are cooperatively rearranged into a new, more stable crystal structure, but without changing the chemical nature of the matrix. Because no atomic migration is necessary, these displacive transformations generally progress in a time-independent fashion, with the motion of the interface between the two phases being limited only by the speed of sound. They are referred to as *athermal* transformations, since the amount of the new phase present is usually dependent only upon temperature, and not the length of time at temperature. *Martensitic transformations* are generally of this second type, and are formed upon cooling from a higher temperature phase called the *parent phase*, or *austenite*.

A precise definition of martensite has never been agreed upon. The terms "martensite" and "austenite" were originally intended to refer only to phases of steels; although some argue the point, the more generalized definition referring to the type of transformation product, and not the particular material, is now widely accepted. Martensitic transformations are first order transformations, meaning that heat is liberated when martensite is formed, there is a hysteresis associated with the transformation, and there is a temperature range over which austenite and martensite co-exist. Summarizing the key features of martensite, we see that it is formed upon cooling with the volume fraction of martensite increasing as the temperature is reduced but with the volume fraction being independent of time, and that it inherits the composition and atomic order of the parent phase. Most of the other debates regarding what defines martensite are in regard to crystallographic details that need not concern us in the present context.

Crystallographically, the transformation from austenite to martensite is often thought of in two parts: the *Bain strain* and the *lattice-invariant shear*. Although these can be crystallographically quite complex, a qualitative two-dimensional approach can be quite simple and perfectly adequate in the present context. The Bain strain, or *lattice deformation*, consists of all the atomic movements needed to produce the new structure from the old; in Figure 1, the austenite structure is schematically illustrated in

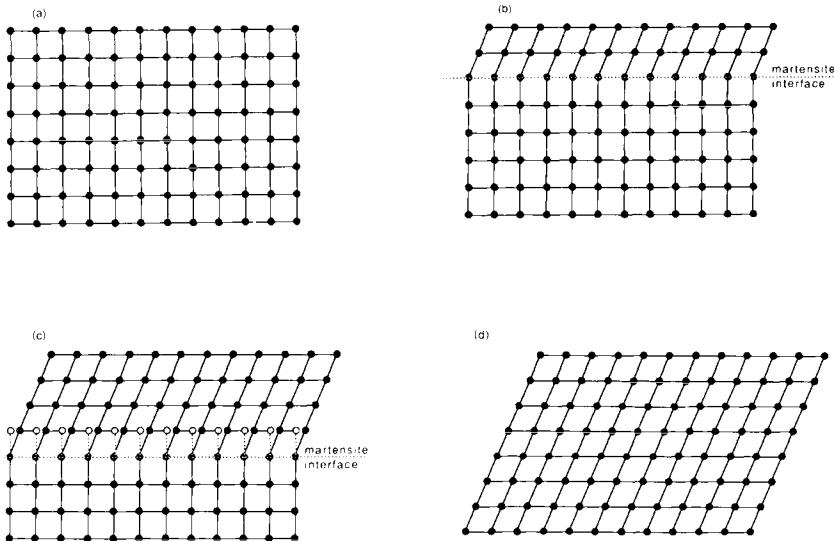


Figure 1: The transformation from austenite to martensite is shown schematically in two dimensions, (a) being completely austenitic and (d) being completely martensitic. Note in (c) that as the interface advances, each layer of atoms is displaced only a very small distance.

diagram (a), and the progression to a fully martensitic structure is schematically illustrated by (b) through (d). Note that as the interface progresses one atomic layer, each atom is required to move by only a very small amount (Figure 1(c)). The end results of all these small coordinated movements is the new martensitic structure, and

the movements needed to produce the new structure are called the Bain strain. In real materials, the Bain strains generally consists of several small atomic shuffles in addition to the type of movement illustrated in Figure 1.

The second part of a martensitic transformation, the lattice invariant shear, is an accommodation step: the martensitic structure produced by the Bain strain is of a different shape, and often volume, than the surrounding austenite (compare Figures 1a and 1d). Martensite in steel involves both a volume and a shape change, whereas shape memory alloys such as Ni-Ti undergo basically only a shape change. Either the overall shape of the new phase, or the surrounding austenite must be altered to accommodate the new structure. By way of comparison, one cannot change the shape of a single brick in the center of a brick wall - either the surrounding bricks must deform, or the new brick must accommodate its form to the space available. There are two general mechanisms by which this can happen: *slip* (Figure 2a) and *twinning* (Figure 2b). In both cases, each individual cell, or parallelogram, has the new martensitic structure, but the overall shape is that of the original austenite. Slip is a permanent process and is a common accommodation mechanism in many martensites. Twinning is unable to accommodate volume changes (should that be necessary) but can accommodate shape changes in a reversible way. For shape memory to occur to any significant extent, we require that the accommodation be fully

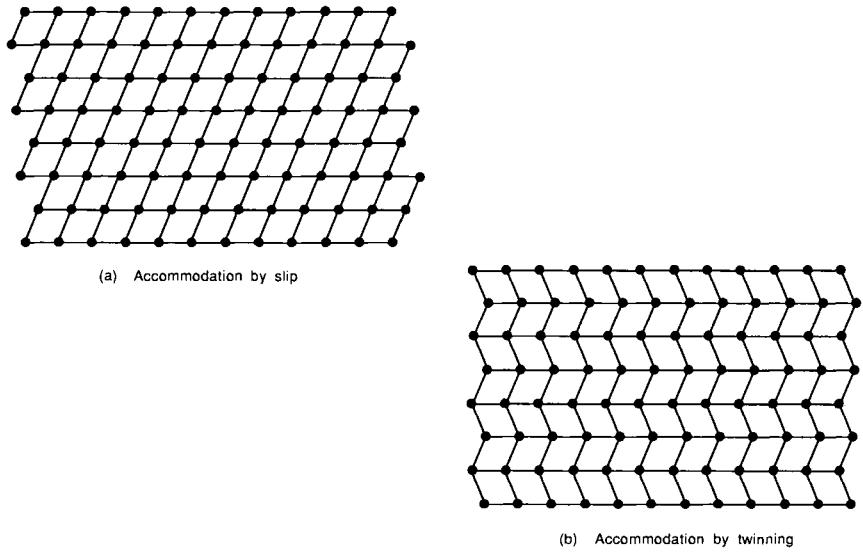


Figure 2: The two mechanisms of accommodating the shape change due to the atomic shear of a martensitic transformation. In slip (above), the microstructure is irreversibly "damaged". In the case of twinning (below) accommodation is reversible, but substantial volume changes cannot be allowed.

reversible, or in other words, that twinning be the dominant accommodation process. In the two-dimensional model of Figure 2, only two directions or *variants* of shear are required to restore the original, overall shape of the matrix; in three dimensions the situation can be complicated: Cu-Zn-Al martensites, for example, require four

6 An Introduction to Martensite and Shape Memory

martensite variants for full, three dimensional accommodation, and Ni-Ti martensites require three.

The twinning process of accommodation plays a key role in the shape memory effect and should be reviewed in more detail. As can be seen in Figure 3, the twin boundary is a mirror plane: when positioned on the boundary, the view in one direction is a mirror of the other. Atoms situated on that boundary see the same number and types of

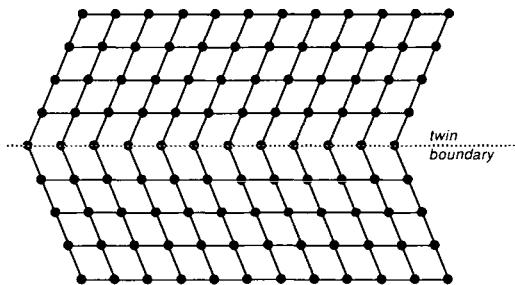


Figure 3: Schematic view of a twin boundary. An atom situated on the boundary sees mirror views left and right of the boundary. The atoms at the boundary are bonded very similarly to those not on the boundary, having the same number of nearest neighbors. This makes twin boundaries very low energy interfaces, as well as being very mobile.

bonds in both directions. Some key properties of twin boundaries are that they are of a very low energy and that they are quite mobile; thus the relative stability of a martensitic phase is not strongly affected by the number or location of these boundaries. By comparing the edges of the structures shown in Figures 2(a) and 2(b), one can see that slip accommodation requires that atomic bonds be broken, while all bonds remain intact in the twinned structure. If a stress is applied to the structure shown in Figure 2b, the twin boundaries will easily move, producing a shape which better accommodates the applied stress. An example is shown in Figure 4. The result of moving a twin boundary is thus to convert one orientation or twin variant into another. That variant will be chosen which is most favorably oriented to the applied stress. In the ideal case, a single variant of martensite can be produced by straining a sufficient amount. This process (the condensation of many twin variants into a single favored variant) is called *detwinning*. Thus far we have only considered the twins within individual martensite plates, but crystallographic analysis has also shown that the boundaries between martensite plates also behave as twin boundaries - i.e. the individual plates of martensite themselves are twins with respect to adjoining plates. Thus the term *twin boundaries*, generally refers to the boundaries between martensite plates as well as the boundaries within plates.

In Figures 1 through 4, atom types are not distinguished, but in an alloy several species of atoms are present. We need to consider then, how these atoms distribute themselves on the lattice sites. In steel, the atoms are *disordered*, meaning that different elements are randomly distributed on the lattice sites. In Ni-Ti, however, the atoms are *ordered*, meaning that the Ni and Ti atoms are found on very specific sites (see Figure 5). During a martensitic transformation, the martensite takes on the same

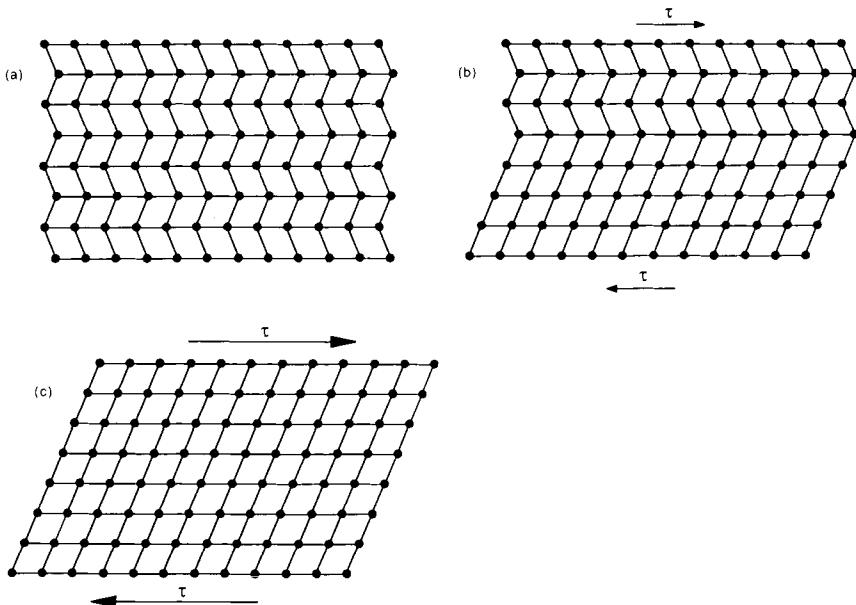


Figure 4: Twin boundaries in martensite can be readily moved by the application of a shear stress. The motion of the twins causes an imbalance in accommodation, and thus a net shape change.

ordering of the austenite. This is called *inherited ordering*. Note also the structures shown in Figures 5a and 5b have a body centered symmetry. (Figure 5a is in fact a Body Centered Cubic (BCC) structure, while Figure 5b is technically not BCC but is called a B2 structure or CsCl structure.) Shape memory alloys are generally based upon a BCC symmetry, some with the BCC structure, more often with the B2 structure, and some with an even more complex ordering called DO_3 , still based upon a BCC symmetry (Figure 5c). It is interesting to note the progression from the BCC to B2 to the DO_3 simply satisfies a need for the different atoms to stay separated from one another.

Martensite normally appears as plates, resting on complex crystallographic planes called *habit planes*. In many shape memory alloys, the martensite plates are large and easily viewed through an optical microscope. One exception, however, is Ni-Ti which exhibits very fine plates that cannot be individually resolved optically. In all systems, but especially Ni-Ti, one has to take great care in preparing samples for viewing since simple grinding and polishing can disturb the martensite, or even create martensite when in fact there was none to begin with. Figure 6 shows a typical micrograph of a martensitic material. A few of the ideas presented above can be supported by this figure. First, notice that there is a scratch running from left to right; this scratch was made on the surface while the material was austenitic, then the material was cooled to form martensite. Each martensite plate changes the direction of the scratch slightly (reflecting the shear nature of the transformation), but notice that the neighboring plate brings the scratch direction back on course. This is a direct result

of the self accommodating nature of the martensite plates: each plate causes a displacement, but when summed, the contributions of all plates cancel. Another very evident feature of a martensitic microstructure is that it is accompanied by a distinct surface relief that one can easily feel in some cases by rubbing a finger over a surface.

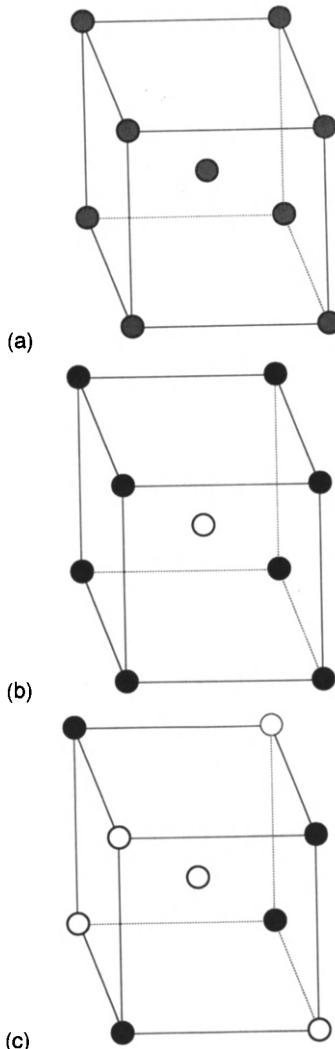


Figure 5: Ordered and disordered structures that are commonly found in shape memory alloys. (a) shows a disordered BCC structure, where different atom types are randomly distributed. (b) shows the B2 structure found, for example, in Ni-Ti, where the atom types are situated on specific lattice sites in order to separate themselves as much as possible. Even though both (a) and (b) appear to have a body centered cubic symmetry, the ordered structure (b) is not body centered cubic since the atoms at the corners are different in nature to those at the center. (c) shows a still more highly ordered state called the D₀₃, found, for example, in Cu-Al-Ni alloys.

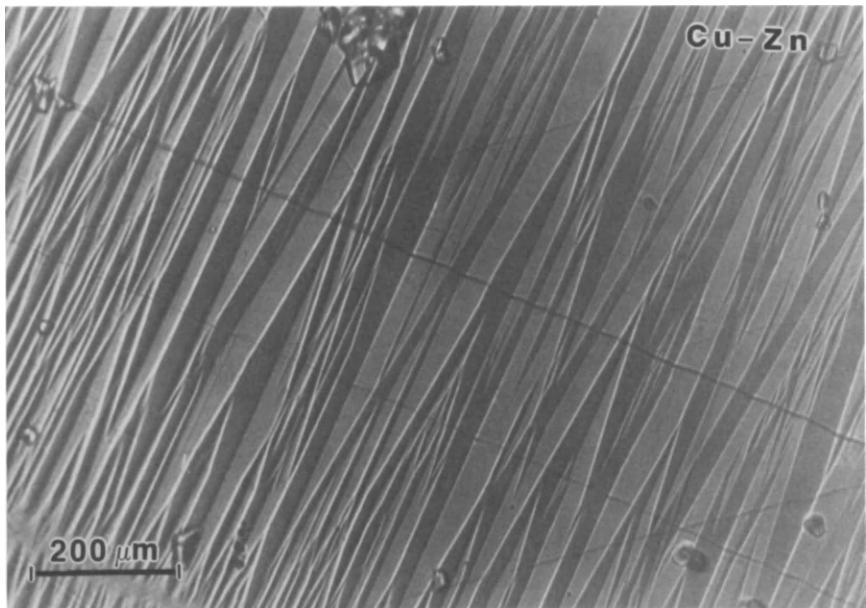


Figure 6: Optical micrograph showing surface relief due to the formation of martensite in a Cu-Zn shape memory alloy. The overall macroscopic displacement of the scratch running from upper left to lower right is nil, although on a local scale the individual martensite plates displace the scratch differently.

This is understood by considering that the plates at the surface are not constrained in three dimensions, and need not be accommodated in a direction normal to the surface.

2. Martensite: a Macroscopic Perspective

Nearly all physical properties of austenite and martensite are different, and thus as one passes through the transformation point, a variety of significant property changes occur. Any of these can be used to follow the progress of the transformation, as is illustrated in Figure 7. The temperatures, M_S , M_f , A_S and A_f , are indicated in the graph and refer to the temperatures at which the transformation to martensite starts and finishes, and the temperatures at which the reversion to austenite starts and finishes. Note that there is a *hysteresis* associated with martensitic transformations; in other words, the transformation temperatures differ upon heating and cooling. The magnitude of the hysteresis depends upon the alloy, but values of 20-40°C are typical for shape memory alloy systems. Microscopically, hysteresis can be thought of as the friction associated with the movement of twin-related martensite boundaries.

One property that changes in a most significant way is the yield strength. The martensitic structure can deform by moving twin boundaries, which are quite mobile. Thus the yield strength of the martensite is extremely low compared to that of the austenite which must deform by dislocation generation and movement. Only a certain amount of martensitic deformation can be accommodated by this twin movement process and once this is exceeded, the material will again deform elastically and eventually yield a second time, this time by irreversible processes (dislocation

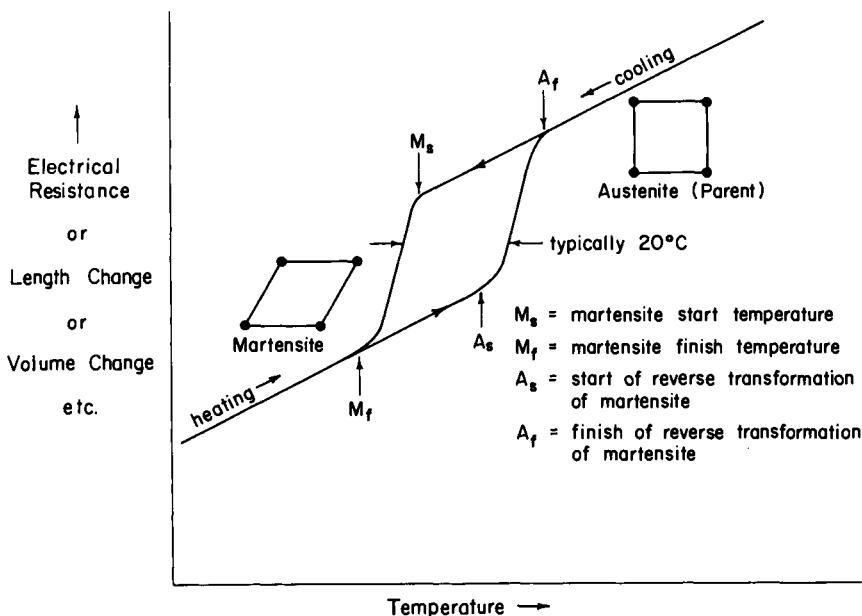


Figure 7: Hypothetical plot of property change vs. temperature for a martensitic transformation occurring in a shape memory alloy. The parent phase (austenite) is represented by the square lattice, which upon martensitic transformation is distorted into the rhombic product phase (martensite). Characteristic temperatures are defined in the inset.

movement). The resulting unusual tensile behavior is exemplified by Figure 8. Note that the *plateau stress* is related to thermal hysteresis, in that both are controlled by the frictional stress of the twin boundaries. The ratio of resistances to reversible and

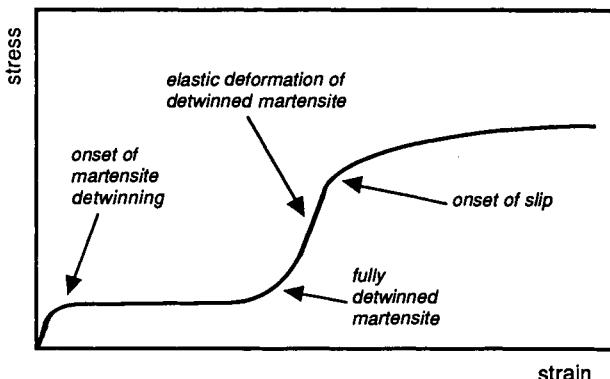


Figure 8: Typical stress-strain curve for a twinned martensitic material shows two distinct elastic regions and two distinct plasticity plateaus, the first due to twin motion, and the second due to slip.

irreversible deformation (twin movement to slip) is thus characterized by the yield strength ratio of martensite to austenite. In general one wants to maximize this ratio in a shape memory alloy, so that all deformation can be recovered upon heating. Typical ratios would be 0.1 to 0.2.

3. The Origin of Shape Memory

Although it has not been explicitly stated, it is implied from the above examples that martensite is generally a lower symmetry phase than is the austenite. The consequence of this is that there are several ways in which martensite can form from austenite, but there is only one possible route which will return the austenite structure. In our two dimensional example, two shear directions can be applied to the squares to produce two different rhombus variants (Figure 2b, for example), but there are no other possible variants of austenite so both of these rhombus shapes would have to return to the same square geometry shown in Figure 1a.

This simple geometrical concept becomes the foundation for the shape memory effect. Upon cooling from austenite (Figure 9a), the self-accommodating variants of martensite are formed (Figure 9b). The twin boundaries migrate during deformation,

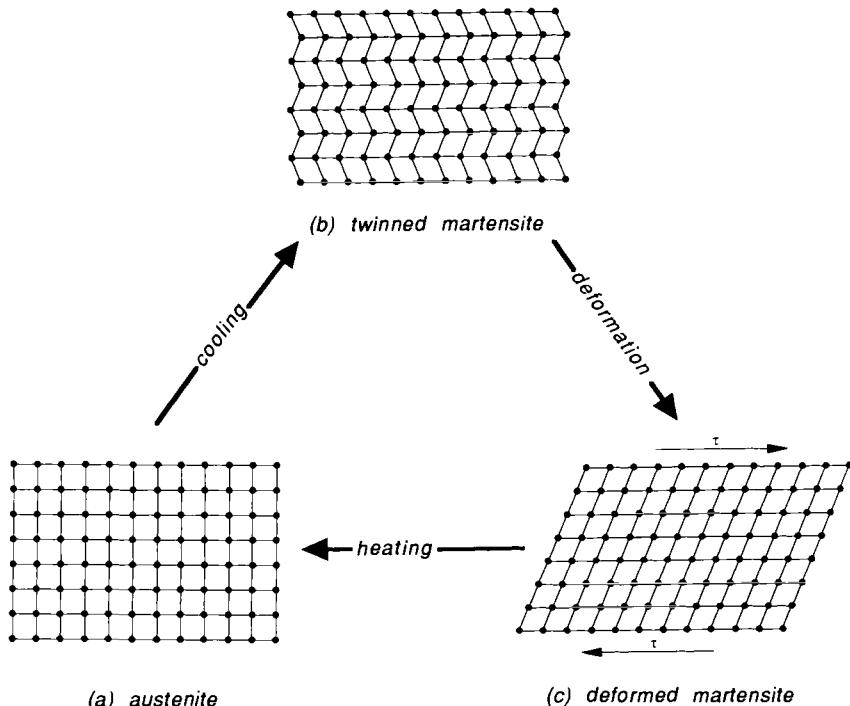


Figure 9: The shape memory process is shown microscopically: austenite (a) is cooled to form twinned martensite (b) without undergoing a shape change, then is deformed by moving twin boundaries (c). Heating either state (b) or (c) will return the originally austenitic structure and shape.

resulting in a biased distribution of martensite variants (or in the extreme case shown in Figure 9c, a single variant). But no matter what the distribution of martensite variants, there is only one possible reverted structure (that of Figure 9a), and with reversion to austenite must return the original shape. Thus the shape accommodation due to twin boundary movement can only be supported by the less symmetrical martensitic structure, and when the more symmetric austenite structure is returned, the twinning deformation must also disappear.

The shape memory effect can be described with reference to the cooling and heating curves in Figure 10. There is no change in the shape of a specimen cooled from

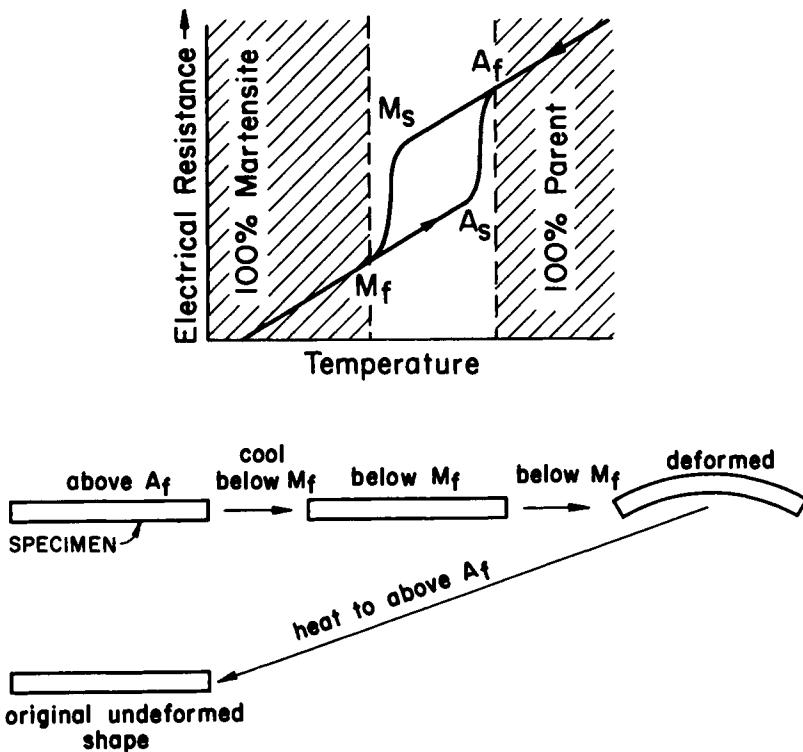


Figure 10: The shape memory effect is described with reference to a plot of electrical resistance vs. temperature from which the characteristic transformation temperatures M_s , M_f , A_s and A_f are determined.

above A_f to below M_f . When the specimen is deformed below M_f it remains so deformed until it is heated. The shape recovery begins at A_s and is completed at A_f . At the inflection point between A_s and A_f , about 50% of the original shape is recovered. Once the shape has recovered at A_f there is no change in shape when the specimen is cooled to below M_f and the shape memory can only be reactivated by deforming the martensitic specimen once again - schematically, cooling results in

Figure 9b, not 9c. In other words, the shape memory effect is a one time only occurrence and because of this is frequently referred to as *one-way shape memory*, in contrast to the two-way shape memory which will be described later. As Figure 10 indicates, recoverable strains on the order of 7% are typical of shape memory alloys, though some show recoveries as high as 10%.

Table 1 shows that many alloys exhibit the shape memory effect. Of those alloy systems shown, only the Cu-Zn-Al, Cu-Al-Ni and Ti-Ni alloys are presently of commercial importance. Comparing the shape memory alloys listed in Table 1, one

Table 1: Some Alloy Systems Exhibiting the Shape Memory Effect

Alloy	Reference	Austenite Structure
Au-Cd	1	B2
Cu-Zn	2	B2
In-Tl	3	FCC
Ni-Ti	4	B2
Cu-Zn-Al	5	B2, DO ₃
Ti-Nb	6	BCC (disordered)
Au-Cu-Zn	7	
Cu-Zn-Sn	8	B2
Cu-Zn-Si	9	B2
Cu-Al-Ni	10	DO ₃
Ag-Cd	11	
Cu-Sn	12	B2
Cu-Zn-Ga	13	B2
Ni-Al	14	B2
Fe-Pt	15	LI ₂
U-Nb	16	BCC (disordered)
Ti-Pd-Ni	17	B2
Fe-Mn-Si	18	FCC (disordered)

1. L.C. Chang and T.A. Read: *Trans. AIME* **191**, 49 (1951).
2. E.J. Suoniemi, R.M. Gennevray and M.B. Bever: *Trans. AIME* **206**, 283 (1956).
3. M.W. Burkart and T.A. Read: *Trans. AIME*, 1516 (1953).
4. W.J. Buehler, J.V. Gilfrich and R.C. Wiley: *J. Applied Physics* **34**(5), 1475 (1963).
5. K. Otsuka and K. Shimizu: *Scripta Met.* **4**, 469 (1970).
6. Baker: *Metal. Sci. J.* **5**, 92 (1971).
7. N. Nakanishi, Y. Nurakami and S. Kachi: *Scripta Met.* **5**, 433 (1971).
8. J.D. Eisenwasser and L.C. Brown: *Met. Trans.* **3**, 1359 (1972).
9. D.V. Wield and E. Billam: *Scripta Met.* **6**, 1157 (1972).
10. C.W. Chen: *J. Metals*, 1202 (October 1957).
11. R.V. Krishnan and L.C. Brown: *Met. Trans.* **4**, 423 (1973).
12. I.A. Arguzova et al.: *Fizika Met. i Metallovedenie* **35**(6), 1278 (1973).
13. T. Saburi and S. Nenno: *Scripta Met.* **8**, 1363 (1974).
14. K. Enami and S. Nenno: *Met. Trans.* **2**, 1487 (1971)..
15. C.M. Wayman: *Scripta Met.* **5**, 489 (1971).
16. J.N. Pridgeon: Private Communication, (1970).
17. A.I. Lotkov, V.N. Grishkov and V.V. Fadin: *Phys. Stat. Sol.* **70**, 513 (1982).
18. A. Sato et al.: *Acta Metall.* **32**(4), 539 (1984).,

finds that in nearly all cases the parent phase is ordered. Other requirements for shape memory can be deduced from earlier discussions:

1. The martensite-austenite transformation must involve only a very small volume change.
2. The martensite must be accommodated by twinning, and not involve slip.

4. Stress-Induced Martensite and Superelasticity

The discussion up to now shows that the shape memory effect is both thermal and mechanical. The martensite is initially formed by cooling and is then deformed below the M_f temperature then heated to above the A_f temperature to cause the shape recovery i.e., the shape memory is caused by heating. We now consider another type of shape memory which is temperature independent: *superelasticity*.

The formation of martensite is a *thermoelastic* process, meaning that an incremental decrease in temperature between M_s and M_f results in a slight growth of existing martensite plates and the nucleation of new ones, but when the temperature is incrementally raised, the newly nucleated plates disappear and those which grew slightly on incremental cooling correspondingly shrink back a little. Put another way, there is an equivalence between temperature and stress: a decrease in temperature is equivalent to an increase in stress, both stabilizing martensite. The martensite is also crystallographically reversible, which means that the reversion of a given plate upon heating is just the inverse of the formation process: the plate undergoes a "backwards shear" as it disappears. Normally on cooling, the martensite forms at M_s under no stress. But in the same material, martensite can form above M_s if a stress is applied, and the martensite so-formed is termed *stress-induced martensite* (SIM). The driving force for the transformation is now mechanical, as opposed to thermal.

Above M_s the stress required to produce SIM increases with increasing temperature, as shown in Figure 11. In fact, the variation in the stress needed to produce SIM

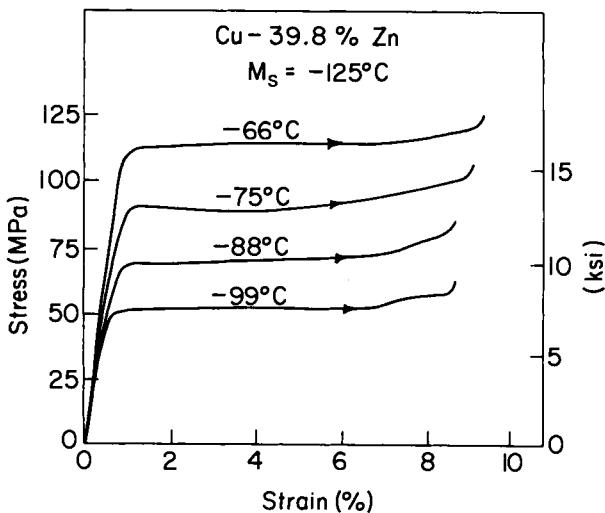


Figure 11: Stress-strain curves for a Cu-Zn single crystal loaded in tension above M_s . As the M_s temperature is approached, the stress required to induce martensite is lowered.

increases linearly with temperature above M_s as shown in Figure 12. Note that the extrapolated stress drops to zero at M_s . The linear variation in stress to induce

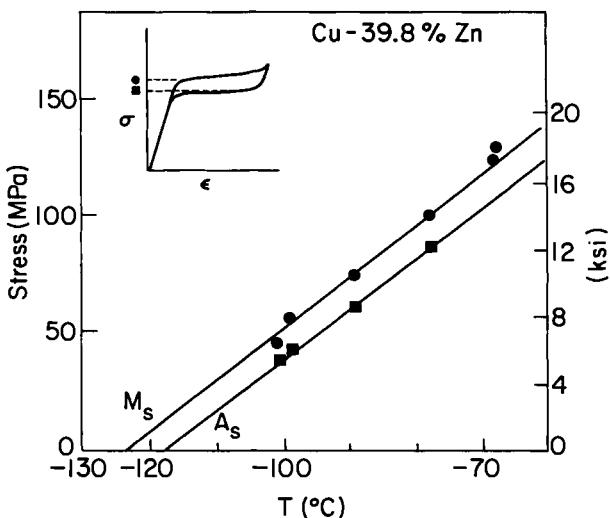


Figure 12: Plotting the plateau stresses such as shown in Figure 12 as a function of temperature gives a linear plot which obeys the Clausius-Clapeyron relationship. The alloy's zero stress A_s and M_s are marked on the ordinate.

martensite as a function of temperature obeys the Clausius-Clapeyron equation, usually written as:

$$\frac{dP}{dT} = \Delta H/T\Delta V$$

where P is the pressure, T is the temperature, ΔH is the transformation latent heat and ΔV is the transformation volume change. This equation has been traditionally used by chemists, but metallurgists, on the other hand, use the Clausius-Clapeyron equation in the form:

$$\frac{d\sigma}{dM_s} = -\Delta H/T\varepsilon_0$$

where ΔH and T have the same meanings as before, and σ , M_s and ε_0 are, respectively, the applied stress, the shifted M_s temperature and the transformational strain resolved along the direction of the applied stress. The increase in difficulty to stress induce martensite continues to increase with temperature until M_d , above which the critical stress for inducing martensite is greater than that needed to move dislocations - this makes M_d the highest temperature at which it is possible to have martensite. Thus the temperature range for SIM is from M_s to M_d .

For a number of SMA systems the agreement in the temperature dependence of the stress to form SIM according to the Clausius-Clapeyron equation is quite striking. In fact, the equation works equally well for the non-isothermal case: in the above scenario, temperature was held constant while the stress needed to form martensite was measured, but one can equally well apply a constant load and measure M_s . If one goes through such an exercise one finds that the same Clausius-Clapeyron relation is valid, and the two slopes ($d\sigma/dM_s$) are equal. Moreover, all the

transformation temperatures are affected by stress in the same way (Figure 13), thus generalizing the form as $d\sigma/dT$, a very important fundamental descriptor of shape memory alloys called the *stress rate*.

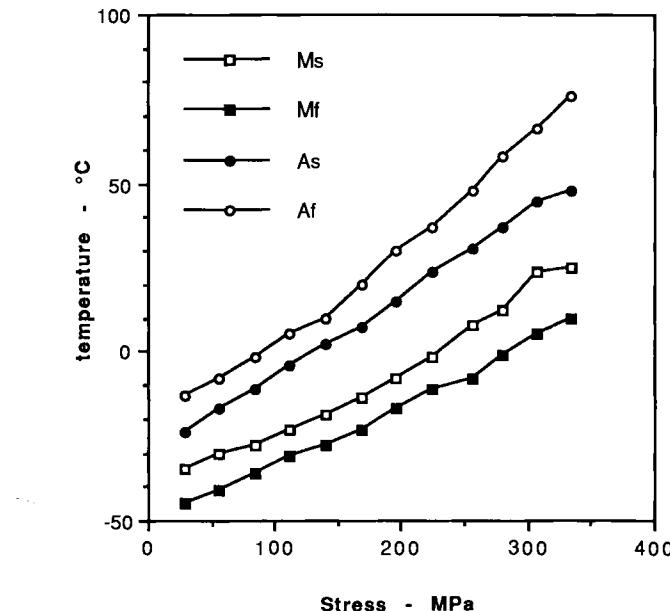


Figure 13: In an experiment that is the inverse of that shown in Figures 11 and 12, one finds the same result. Here a Cu-Zn-Al-Mn alloy is thermally cycled at different stress levels and M_s , M_f , A_s and A_f plotted as a function of stress. The stress rates for all transformation temperatures are the same.

Superelasticity occurs when a material is deformed above A_s , but still below M_d . In this range, martensite can be made stable with the application of stress, but becomes unstable again when the stress is removed. Figure 14 shows a "superelastic" stress strain curve (often referred to as a superelastic loop) for a Cu-39.8%Zn SMA. The upper plateau corresponds to the formation of martensite under stress while the lower plateau represents the reversion of the SIM when the stress is released. Note that 9% strain is fully recovered during unloading, in what can be viewed as a mechanical shape memory effect. When the SIM is formed, usually only one variant of martensite is formed, as shown in Figure 15 for a single crystal of the same Cu-Zn alloy. In Figure 15a, only a few plates of SIM are formed, with a habit plane that is sympathetic with respect to the applied stress axis. That is, the shape deformation of that particular variant produces maximum elongation of the specimen along the tensile axis. Figure 15b shows the same specimen at a higher strain level with many SIM plates of the same variant. Since only one martensite variant is formed under stress, there is a shape change (elongation) which is fully recovered upon release of the stress. This situation is unlike the case of thermal martensite, where because of self-accommodation there is no overall shape change accompanying to the formation of many variants of martensite.

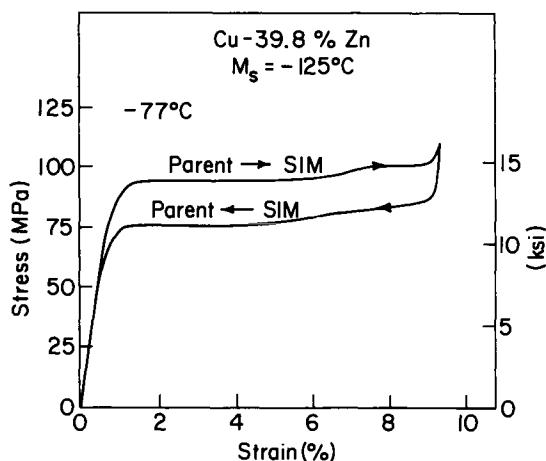


Figure 14: Stress-strain curve for a Cu-Zn shape memory alloy loaded above the A_s temperature and then unloaded shows distinct superelastic behavior. Stress-induced martensite is formed during loading, which becomes unstable and disappears upon unloading.

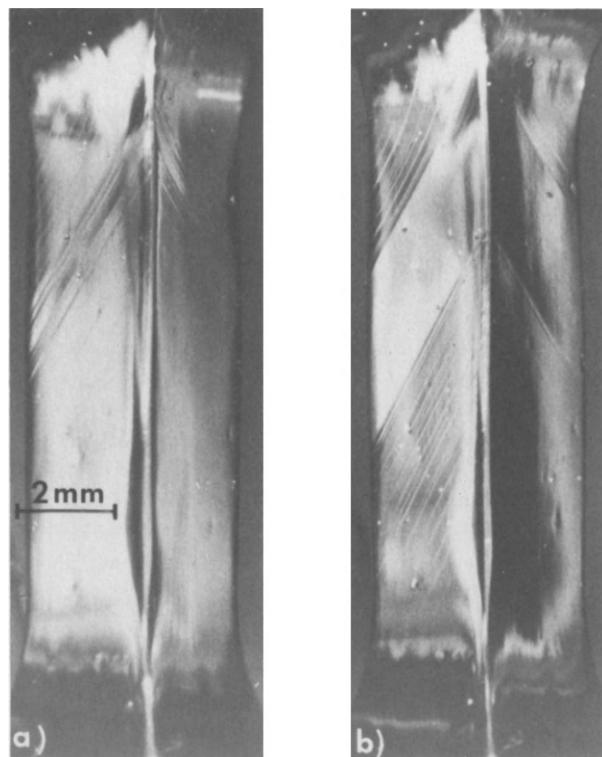


Figure 15: The gage length of a Cu-Zn single crystal showing the formation of only one variant (orientation) of stress-induced martensite is formed during tensile loading above the M_s temperature.

For purposes of comparison, Figure 16 shows a superelastic stress-strain curve for a Ti-Ni SMA spring compared to that for a typical spring material, piano wire. For the shape memory alloy the strain is completely recovered, but when the piano wire spring is extended the same amount, it undergoes permanent deformation and only a part of the strain is recovered.

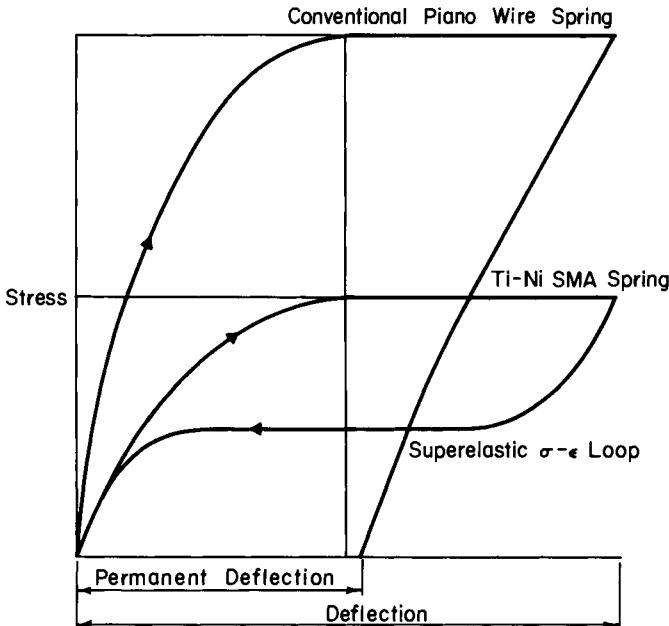


Figure 16: Comparison of springs made of superelastic shape memory alloy and conventional piano wire.

The mechanical behavior of shape memory alloys is summarized graphically in Figure 17, adapted from the behavior of a Ti-Ni alloy. At the extreme rear the stress-strain shown in the x-y plane corresponds to the deformation of martensite below M_f . The induced strain, about 4%, recovers between A_s and A_f after the applied stress has been removed and the specimen heated, as seen in the x-T plane. At a temperature above M_s but below M_d SIM is formed, leading to the usual superelastic loop with an upper and lower plateau. At a still higher temperature (above M_d), the front x-y plane, no SIM is formed. Instead, the parent phase undergoes ordinary plastic deformation.

5. The Two-Way Shape Memory

Finally, we consider the *two-way shape memory effect* (TWSM), illustrated in Figure 18. In the upper part of the figure, a collapsed SMA spring is deformed by extension below M_f . The original spring shape (contracted) is recovered following heating to above A_f . The contracted shape remains when the specimen is again cooled to below M_f . This is the one-way shape memory behavior, which, as noted before, is a one time only deployment. In contrast, the TWSM is depicted in the lower half of the figure, in

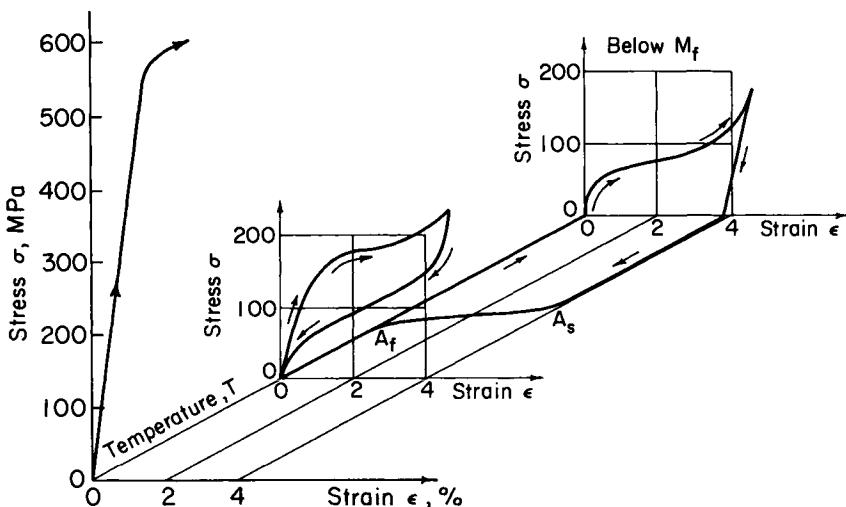


Figure 17: Three-dimensional stress-strain-temperature diagram showing the deformation and shape memory behavior of a Ti-Ni alloy deformed below M_f , above A_f and above M_d . See text for discussion.

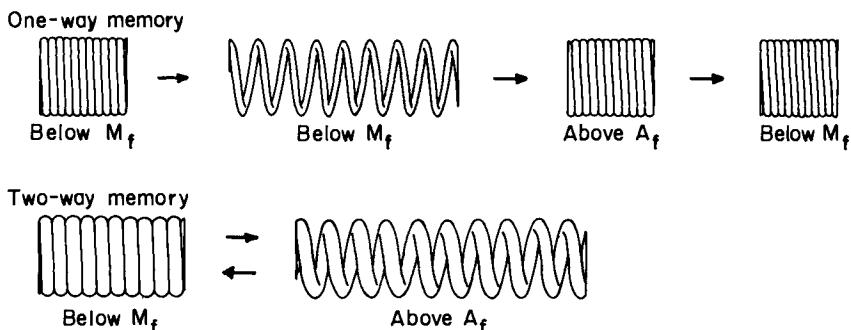


Figure 18: Comparison of the one-way (upper) and two-way (lower) shape memories using a coil spring as an example.

which case a contracted spring extends when heated to above A_f , but now spontaneously contracts when again cooled below M_f . The spring extends again when heated above A_f and contracts again when cooled below M_f , repeating indefinitely.

To produce the two-way behavior, special thermomechanical treatment is required, the details of which will be presented later in this book. Though there are several such treatments that can be used (called *training* treatments), all introduce microstresses in the material that tend to bias the nucleation and growth of martensite,

causing some variants to form preferentially. In terms of Figure 9, the effect of these microstresses is to cause the 9c structure to form instead of the structure shown in 9b.

6. Summary

The shape memory effect is a consequence of a crystallographically reversible martensitic phase transformation occurring in the solid state. Although there are many ways (orientations) to produce the martensite phase from its parent during cooling, once the lower symmetry martensite is formed it has only one unique reversion path during the reverse transformation because of crystallographic restrictions. The transformation of the parent phase into martensite is basically a deformation process, but because the individual units of martensite self-accommodate, the overall macroscopic deformation upon transformation is zero. When the shape memory martensite is deformed, a particular orientation of the various self-accommodating units - that most favorably oriented with respect to the applied stress - grows at the expense of others, eventually leading to a single orientation of martensite. As before, this orientation has only one reversion path, which is the essence of the shape memory - a deformed and reoriented martensitic phase which is thermally responsive. Strains on the order of seven to ten percent are typically recoverable in this manner.

Shape memory alloys also display superelasticity, a mechanical type of shape memory as opposed to the thermally induced (by heating) shape memory described above. In this case, when the parent phase is deformed above the martensite start temperature, the martensitic transformation occurs "prematurely" because the applied stress substitutes for the thermodynamic driving force usually obtained by cooling. But since the applied stress is basically uniaxial, only one orientation (out of many) of martensite is selectively formed, and this imparts an overall deformation to the specimen. This deformation disappears when the stress is released and the original specimen shape is restored, leading to a mechanical shape memory.

Finally, a two-way shape memory can be realized in shape memory materials, whereby a specimen is programmed by means of thermomechanical treatment producing microstresses in the parent phase which in turn program the specimen to behave as a stress-induced martensitic transformation. That is, the microstresses favor only a single orientation of martensite upon subsequent cooling, which produces a spontaneous deformation. When the specimen is heated the normal one-way shape memory process occurs and its original shape is reproduced. The two-way process can be repeated indefinitely (as with a thermostat) as opposed to the one-way memory, which is a one time only operation (as in making a mechanical connection).

Ni-Ti Based Shape Memory Alloys

K.N. Melton
Raychem Corporation
Menlo Park, California

Shape memory alloys based on nickel and titanium have to date provided the best combination of material properties for most commercial applications. In fact it was the discovery in the 1960's of shape memory properties in the Ni-Ti system which led to the rapid growth of interest in the shape memory phenomenon. This discovery took place at NOL, the Naval Ordnance Laboratory and hence the acronym NiTi-NOL or Nitinol which is commonly used when referring to Ni-Ti based SMA's. A series of extensive reports¹⁻³ documented the properties of Ni-Ti, mainly in wire form, and provided much useful information for the designer. However since that time there has been significant progress in the understanding of the alloy, particularly with regard to the effects of processing and heat treatment on the mechanical and shape memory behaviour. This paper is intended as an overview of the Ni-Ti SMA's, with the emphasis on some of the more recent developments.

1. Metallurgy of Ni-Ti Alloys

Ni-Ti SMA's are ordered intermetallic compounds based on the equiatomic composition. From the phase diagram⁴ this compound exists as the stable phase down to room temperature. Consequently, in contrast to copper based alloys, no betassising and quenching is necessary to prevent the decomposition into other phases at intermediate temperatures. However, at low temperatures the stoichiometric range of Ni-Ti is very narrow and so the alloys often contain precipitates of a second intermetallic phase. The microstructure is thus primarily single phase, with small amounts of other phases distributed in the matrix, Figure 1a. One factor which is often ignored in considering the microstructure of Ni-Ti alloys, is the presence of oxygen. Titanium is very reactive, particularly in the molten state, and some oxygen is invariably present in the alloy. From the Ni-Ti-O phase diagram⁵, oxygen decreased the stoichiometric range of the NiTi compound and can unexpectedly result in compositions within a three phase field. Thus Ni_3Ti can be present for example in a titanium rich alloy. Furthermore, the oxide $\text{Ti}_4\text{Ni}_2\text{O}$ is isostructural with the intermetallic Ti_2Ni ⁶ which can make unique phase identification difficult. If the composition of the alloy deviates from stoichiometry, then larger precipitates are present, as can be seen in Figure 1b for a Ti-rich alloy. These larger second phase particles can have a marked effect on the hot workability of Ni-Ti, particularly on the titanium rich side where they are brittle and often result in cracking.

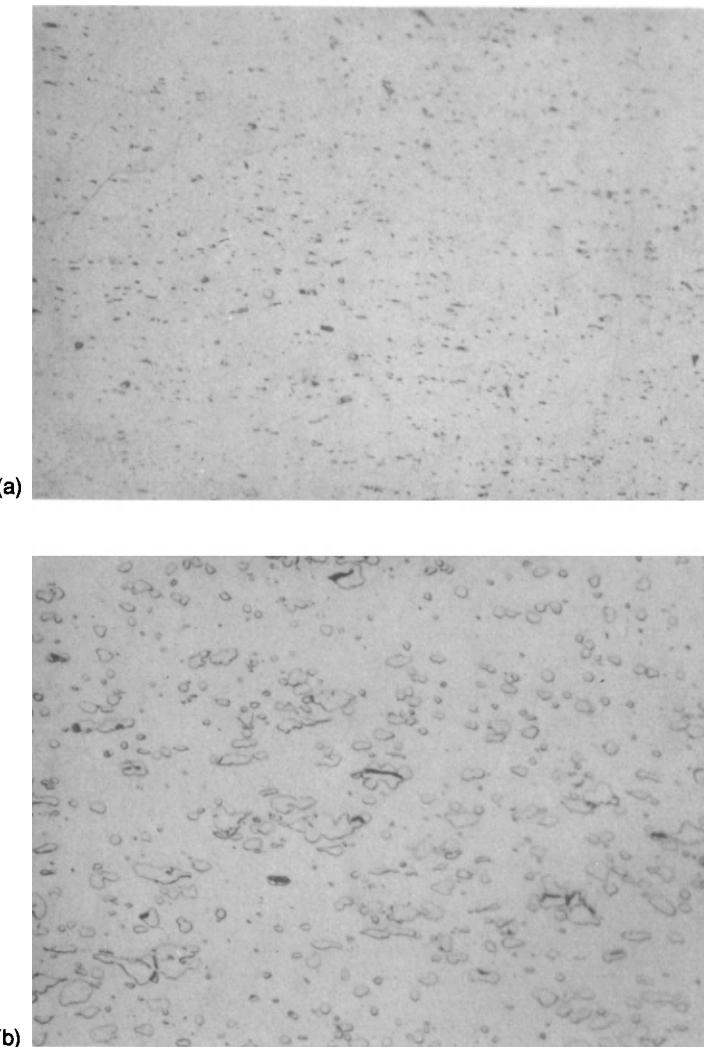


Figure 1: (a) The microstructure of a hot worked 50 at % Ni alloy showing a fine distribution of second phase particles, and (b) the microstructure of a hot worked 52 at % Ni alloy showing a higher volume fraction of coarser particles compared to 1a. Note some cracking resulting from hot working these brittle particles.

Within the composition range at which the Ni-Ti phase exists at ambient temperature, M_s depends quite strongly on composition, Figure 2, particularly on the Ni-rich side. Ti-rich alloys show less sensitivity, primarily as a result of the formation of a Ti-rich precipitate, leaving the matrix composition essentially the same. The composition dependence of M_s shown in Figure 2 has important practical consequences. Precise composition control is required when melting the alloys. Depending upon the desired

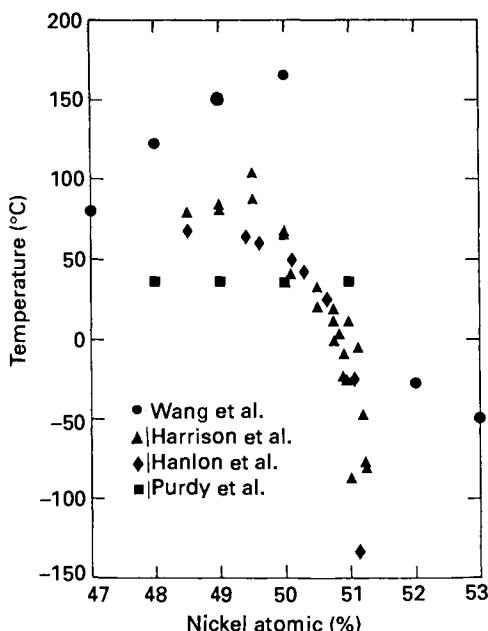


Figure 2: The dependence of the transformation temperature M_s on composition.

M_s , the necessary control is to between one tenth and one hundredth of a percent! By way of comparison, the typically allowed variation in chromium content of 18/8 stainless steel is 2%. The compositional accuracy required for Ni-Ti is mostly more precise than the errors in chemical analysis. Consequently, in most cases for quality control purposes the transformation temperature itself is measured, rather than the chemical composition. As a result of this precise composition control to achieve the desired M_s , other properties such as yield strength etc show significantly smaller lot to lot variations than in most other engineering metallic materials.

Figure 2 includes data from Wang et al⁷, which indicates a maximum transformation temperature of around 150°C. However the results were obtained from damping experiments of a not very sophisticated kind. Bars were struck with a hammer, and the transformation temperature subjectively determined from whether the sound was a leaden thud or a ringing sound similar to a bell. Quantitative measurements showed⁸ that the damping behaviour is in fact quite complex. Consequently the data from Wang et al⁷ support the general trend of transformation temperature as a function of composition, but transformation temperatures in the vicinity of 150°C should not be expected.

Ni-rich alloys are unstable in the sense that M_s shifts in annealed material can occur during prolonged exposure at temperatures which could be met during service. This instability is the consequence of the formation of precipitates, and the precipitation sequence has been studied in detail⁹. Figure 3 shows the A_s temperature for a series of alloys in the quenched and quenched and aged conditions. It can be seen that, for Ti-rich alloys there is very little difference between these heat treat conditions, whereas Ni-rich alloys show a low A_s in the as-quenched condition, but subsequent

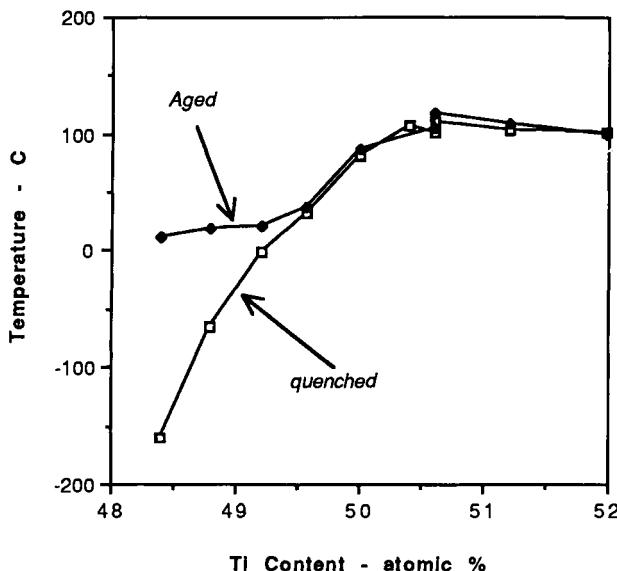


Figure 3: A_s temperatures as a function of composition for water quenched from 850°C and quenched and aged at 400°C samples.

aging can lead to a dramatic increase. On aging, the Ni-Ti decomposes into a more Ti-rich matrix composition with higher A_s , together with a Ni-rich phase finely dispersed within it.

The fact that the instability occurs on the nickel rich side only is a consequence of the solubility range of Ni-Ti extending to higher nickel contents at temperatures above around 500°C⁴. At high temperatures excess nickel goes into solution in Ni-Ti, and on aging for longer times at lower temperatures it precipitates out again. On the titanium rich side the solubility range is almost independent of temperature. For many applications of Ni-Ti alloys, this instability is of little importance, since service temperatures in the range where the M_s shift is rapid do not often occur. However for couplings where a cryogenic annealed alloy is used, service exposure in the 200 to 300°C range is likely to raise M_s above the minimum operating temperature, rendering the coupling useless.

2. Mechanical Properties

Like most SMA's, Ni-Ti alloys show marked differences in mechanical behavior depending on whether they are tested in the austenitic or the martensitic phases. The stress-strain curve of the martensite can be divided into three well defined regions¹⁰⁻¹³, see Figure 4. An initial low plateau results from the stress induced growth of one martensite orientation at the expense of an adjacent, unfavourably oriented one, this process occurring by a detwinning type of mechanism. At higher stresses there is a second region which is usually linear, although not purely elastic. It is believed¹³ that the deformation mechanism in this stage is a mixture of elastic deformation of the detwinned martensite, together with the formation of new orientations of martensite

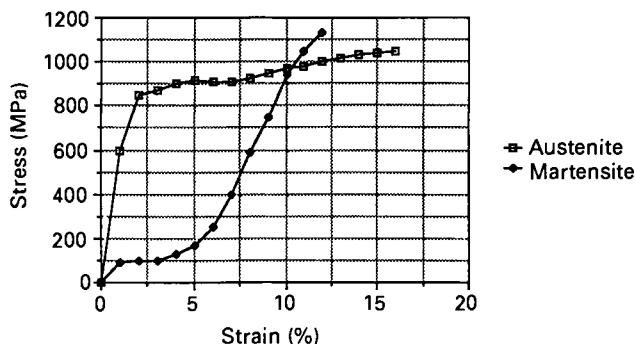


Figure 4: The stress-strain curves of a Ni-Ti-10% Cu alloy in the austenitic and martensitic conditions.

which intersect those already present, and which provide additional heat recoverable strain.

The transition to the third region is a result of the onset of irreversible plastic deformation, as in the case of the yielding of all conventional metals. Thus the maximum amount of heat recoverable or memory strain is obtained by initially deforming to the end of stage two. If larger deformation strains are used, then the reversible martensitic deformation processes and the dislocations resulting from plastic flow interact and the memory strain decreases, see below.

The length of the martensitic plateau in the stress-strain curve extends typically to around 5-6%. However, depending on the details of the alloy and particularly its prior thermo-mechanical history, this "plateau" can vary from a continuous curve with an inflection point to a clear horizontal plateau with a sharp yield point and upturn. Because of this variation in yield behavior, many have found it more useful to use a 1% strain offset than a 0.2% offset when defining a yield stress. Furthermore, for wrought material the shape of the stress strain curve depends on orientation. Figure 5.

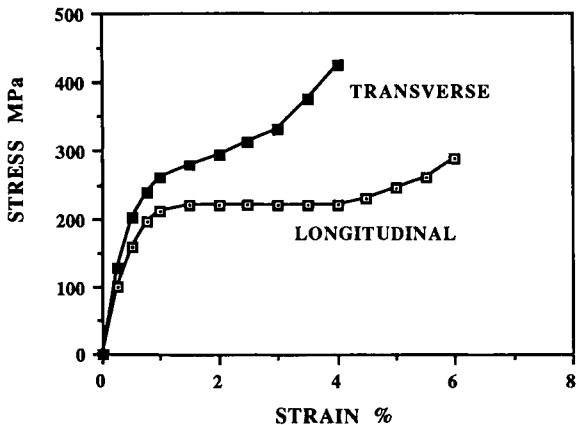


Figure 5: Martensitic stress strain curves of specimens taken longitudinally and transversely from rolled sheet.

shows curves from rolled sheet, where it can be seen that testing in the longitudinal direction results in a well defined plateau, whereas in the transverse direction none exists and the stress levels are much higher. The details of the martensite curve depend also on the mode of deformation, marked differences can be obtained in tension, compression and torsion¹⁴, see Figure 6.

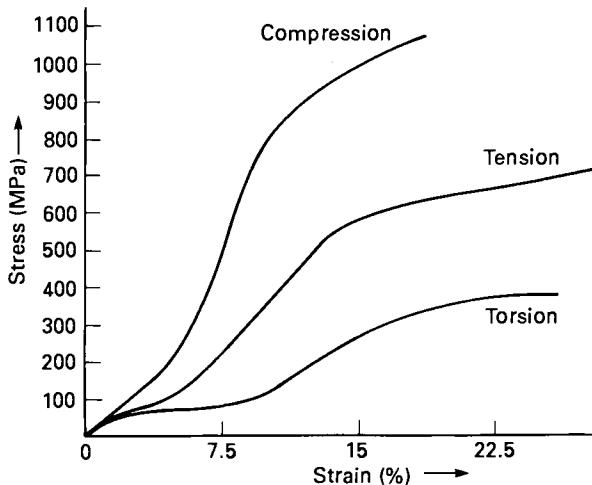


Figure 6: Stress strain curves for a Ni-Ti-10% Cu alloy measured in compression, tension and torsion.

The length of the martensite plateau also determines the strain at which the transition to region III of the stress strain curve occurs, and can therefore affect the amount of memory strain. Thus although it is often stated that up to 8% strain is heat recoverable, the actual amount depends on the alloy, its thermo-mechanical processing, testing direction and deformation mode. In many cases, a conservative design would use less than 8% strain.

For test temperatures just above M_s , and where this temperature is reached by cooling, the alloy is austenitic at the start of the test and the mechanism of yielding is the onset of stress induced martensite. As the test temperature is increased, the deformation mode stays the same, but the martensite is thermodynamically less stable and so a greater stress is required to induce its formation (as discussed in the previous chapter). Between M_s and M_d the yield stress (to induce martensite) increases linearly with temperature, the slope being given by the Clausius-Clapryon equation. A plot of initial yield stress vs temperature Figure 7, can be divided into three regions depending on the deformation temperature T_d .

$T_d < M_s$	detwinning of martensite
$M_s < T_d < M_d$	stress induced martensite
$T_d > M_d$	plastic deformation

Figure 7 shows the data for two Ni-Ti alloys, illustrating also the general trend that as the nickel content of the alloy decreases (and thus M_s increases) the austenitic yield strength decreases.

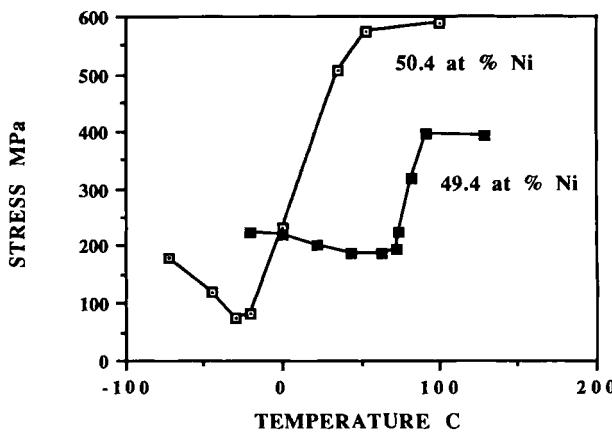


Figure 7: Yield stress measured at different temperatures for two Ni-Ti compositions. Each data point represents a separate specimen.

Methods of determining transformation temperatures will be discussed in detail in a later chapter. However one method is to thermally cycle a specimen under load, producing a curve similar to that illustrated in Figure 8. The transformation temperatures are conveniently obtained by the intersection of tangents, as shown. Ni-Ti based alloys have a hysteresis width of typically 30-50°C. Transformation

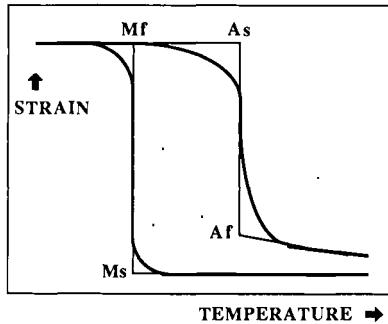


Figure 8: A schematic plot of strain versus temperature of a sample which is loaded in the austenitic condition, and then cooled and reheated through the transformation at constant load.

temperatures increase with increasing load in a linear fashion, Figure 9, the slope being thermodynamically controlled by the stress rate: $d\sigma/dT$ (the same as that of the yield stress vs temperature plot in the testing temperature range between M_s and M_d , Figure 7). This slope, as well as being of thermodynamic significance, is also important to bear in mind when choosing an alloy for a particular application. For

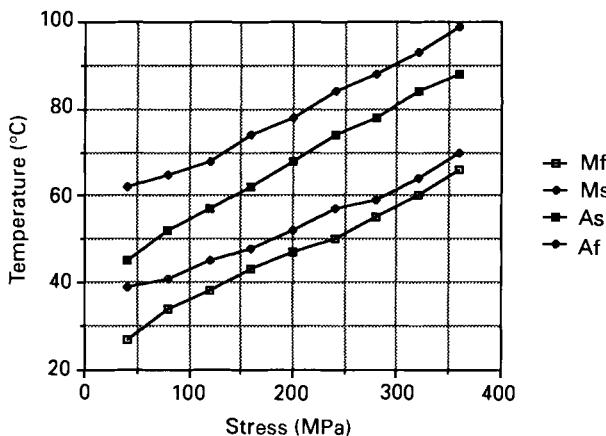


Figure 9: The transformation temperatures of a Ni-Ti-10% Cu alloy measured as a function of load by the method indicated in Figure 8.

example, both the response temperature (A_s) and reset temperature (M_s) of an SMA actuator will depend on the opposing force. Furthermore, for applications where the SMA is required to maintain its mechanical integrity (couplings, fasteners etc) the stress in the component can decrease on cooling below M_d . The stress rate of Ni-Ti can also change significantly from alloy to alloy, covering a range from 2.5 MPa/ $^{\circ}$ C to over 15 MPa/ $^{\circ}$ C. Generally alloys with lower transformation temperatures have lower stress rates, dictated by the Clausius-Clapeyron equation and a decreasing latent heat of transformation. Higher strength alloys also seem to have generally higher stress rates.

The mechanical design of a structure will often use modulus as a parameter. However for shape memory alloys in general and Ni-Ti alloys in particular, the concept of modulus is not straightforward. Figure 10 shows the Young's modulus measured dynamically¹⁵ of three Ni-Ti based alloys plotted as a function of temperature, where it can be seen that dramatic changes take place. In particular, there can be a pronounced modulus decrease in the austenitic phase on cooling, which occurs before M_s . Another complicating factor concerning modulus is that the slope of the elastic loading and unloading lines are not the same, and the unloading "line" is often not in fact linear. These points can be seen from the austenitic curve shown in Figure 11. For the case of deformation of the martensite, the unloading behaviour shows considerable departure from linearity. For deformation strains above 8% the dislocations introduced can act as pinning centers for martensite twin boundaries. When the external stress is removed the dislocations move back to a new equilibrium position and take the twin boundary with them, giving much more unloading strain than normal alloys. Figure 12 shows that for a total deformation strain of 20%, only around 15% plastic strain is introduced, in other words a springback of 5% is obtained. However for 10% total strain, the plastic value is around 9%, indicating only 1% springback.

3. Effects of Thermo-mechanical Processing

In 1965, four years after the alloy patent on binary Ni-Ti was filed¹⁶, a process patent was filed claiming cold working in the martensite as a way of increasing the yield

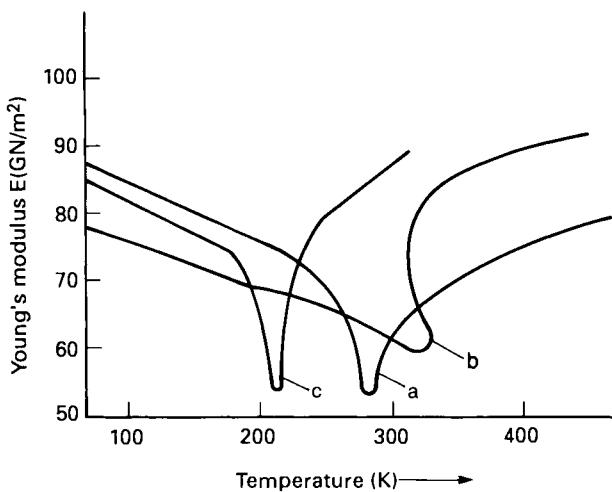


Figure 10: The dynamic Youngs modulus of three Ni-Ti based alloys as a function of temperature, from reference 15.

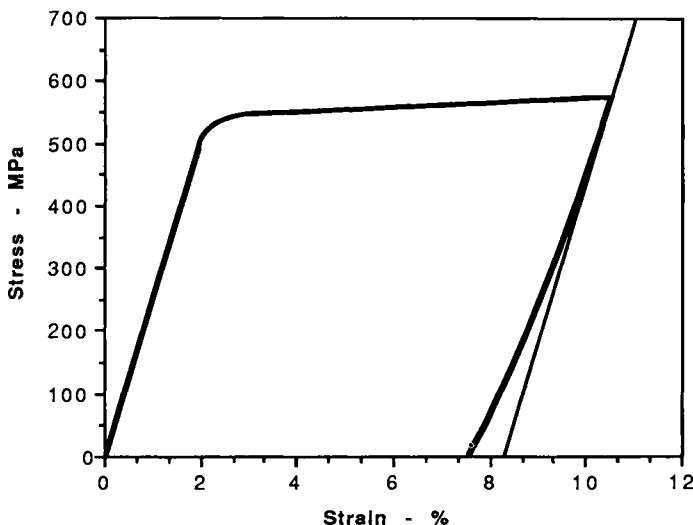


Figure 11: The loading and unloading stress strain behavior of a Ni-Ti alloy tested in the austenitic condition.

strength of the alloy¹⁷. Since that time the combination of cold work in the martensite together with a subsequent anneal has been extensively explored as a way of improving the SMA characteristics.

Cold work alone, i.e. without the annealing step, destroys the martensitic plateau on the stress strain curve. Thus a material cold worked 20% in the martensite has a very

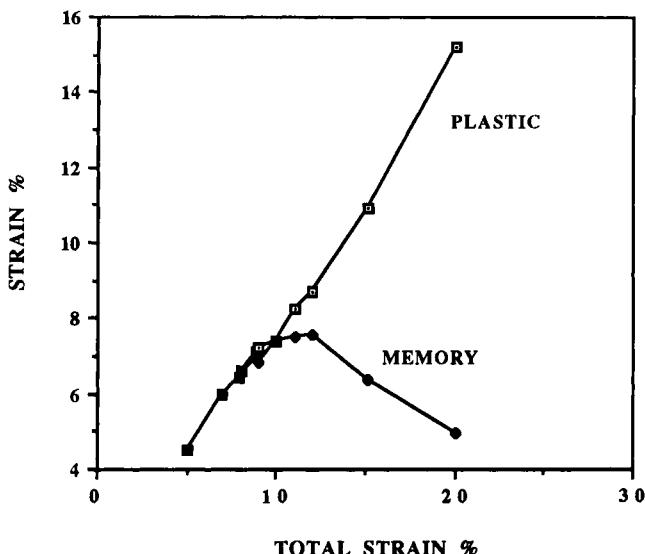


Figure 12: The plastic and shape memory strains of a Ni-Ti alloy plotted as a function of total strain.

high yield strength¹⁸, but its shape memory properties are poor in that only very low recoverable strains are possible. Annealing will restore the memory effect, but decrease the yield stress. The choice of amount of cold work and the actual annealing temperature dictate the trade off in these two properties. It is believed that cold work introduces a high density of "random" dislocations, which impede the mobility of the twin boundaries. Annealing rearranges these dislocations into cells of relatively dislocation free areas within which the martensite twins can be mobile, but surrounded by dislocation networks. It is the presence of the cells or subboundaries which gives the hardening.

Figure 13 shows the effect on austenitic yield strength of annealing temperature after 40% cold work in the martensite of a 50.6 at %Ni alloy. It can be seen that a rapid decrease is observed in the range 350 to 450°C, followed by a more gradual decrease out to 850°C. At the same time, M_s increases, Figure 14, and again the increase is rapid between 350 and 450°C. Comparing Figures 13 and 14 it is apparent that for a given alloy, although processing can increase the yield strength, it is done at the expense of a lower M_s .

One of the principal improvements of Ni-Ti alloy properties obtained by processing is cyclic behaviour. Figure 15 shows the strain measured as a function of temperature of the same binary alloy annealed at 850°C, with a load corresponding to an initial stress of 150 MPa being applied. The main thing to notice is the large amnesia of around 1.5%, in other words a single memory cycle with this stress leads to an irreversible deformation of 1.5%. Doing the same test after cold working 40% and annealing at 400°C results in amnesia of around 0.5% Figure 16. The strain obtained on cooling and the amnesia on heating are plotted in Figures 17 and 18 as a function of stress after cold working and annealing at 350°C and 450°C respectively. The 350°C anneal results in a very small amnesia for stresses up to 300 MPa, whereas the 450°C anneal

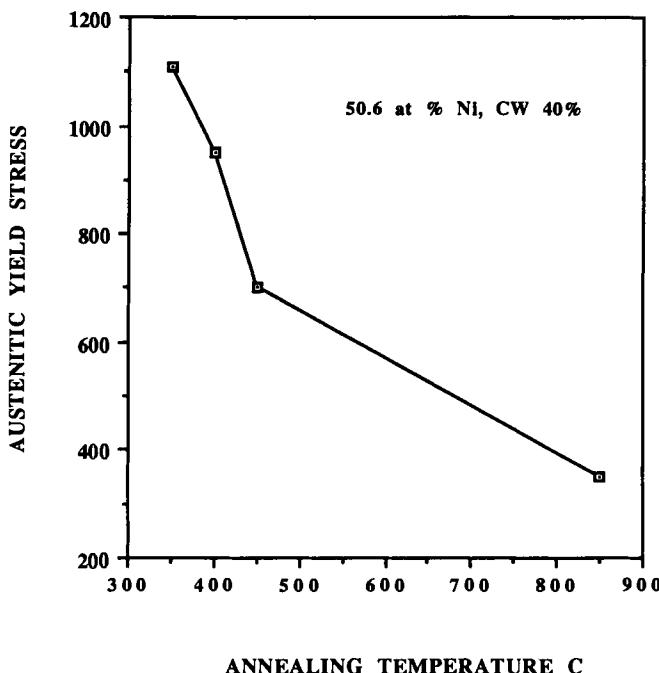


Figure 13: The austenitic yield stress of a 50.6 at.%Ni alloy, cold worked and then annealed for 30 mins. at the temperatures indicated.

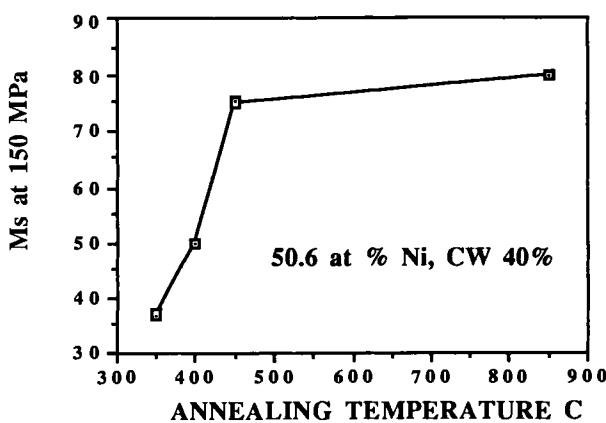


Figure 14: The transformation temperature Ms (measured at 150MPa) as a function of annealing temperature for the same alloy used in Figure 13.

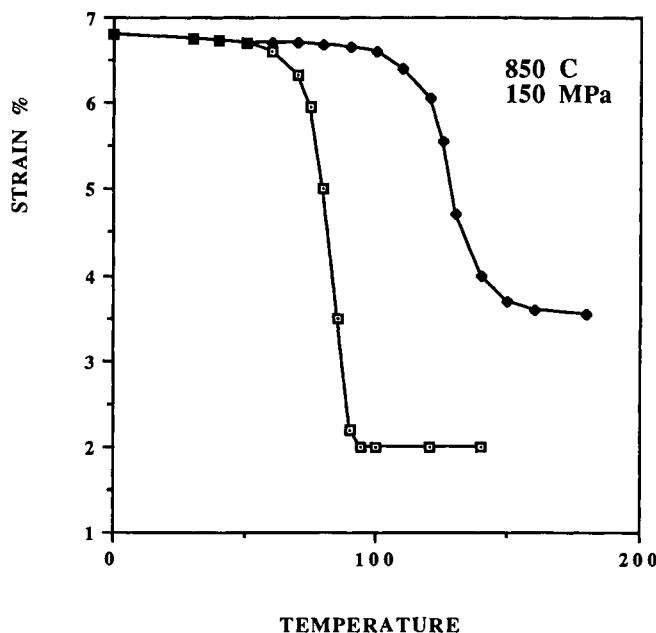


Figure 15: Strain versus temperature curves at 150 MPa for a 50.6 at.%Ni alloy annealed at 850°C.

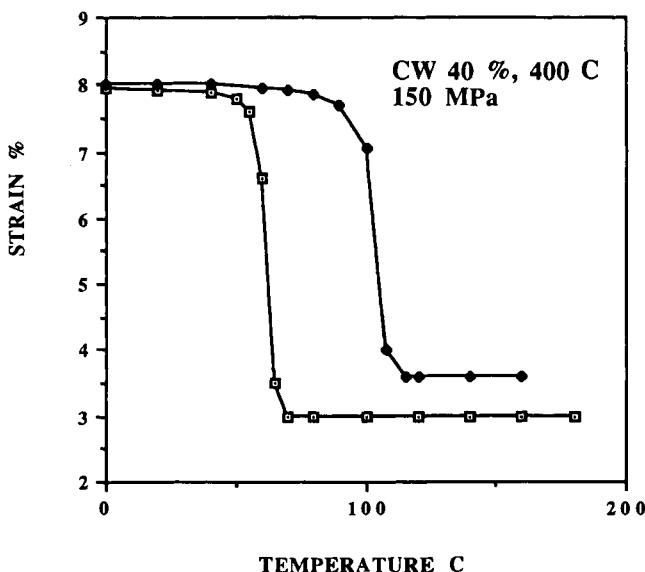


Figure 16: Strain versus temperature curves at 150 MPa for the same alloy as Figure 15, but cold worked 40% and annealed at 400°C.

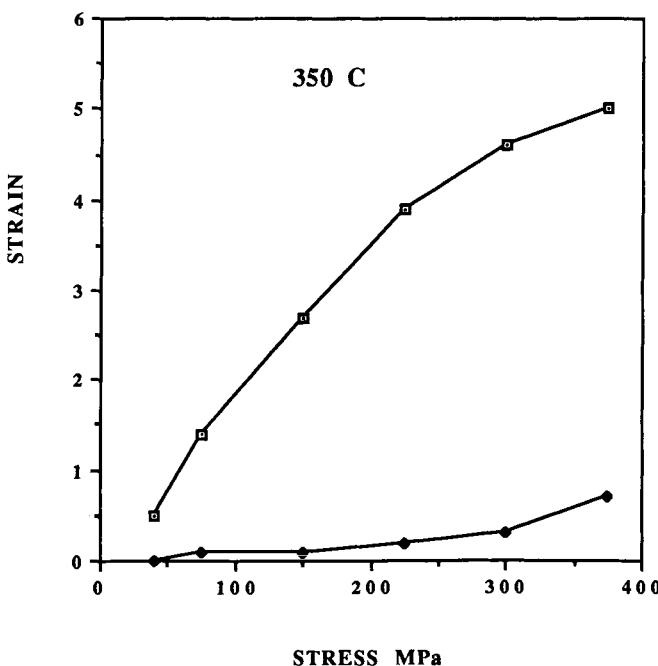


Figure 17: The strain obtained on cooling (upper curve) and the amnesia (lower curve) as a function of stress for a Ni-Ti alloy cold worked 40% and annealed at 350°C.

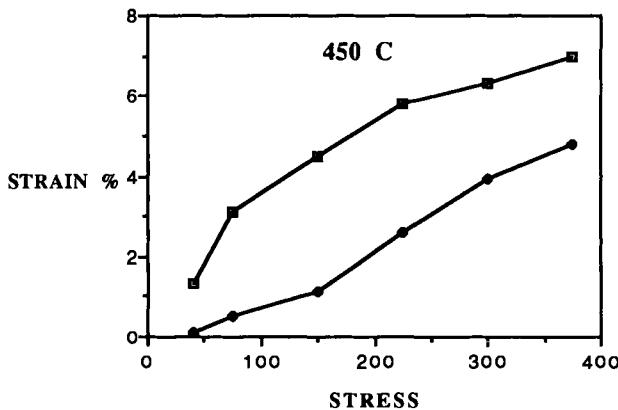


Figure 18: The data of Figure 17 after annealing at 450°C

gives significant amnesia at all useful working stress levels. For low temperature anneals where the amnesia is low, the data obtained from the first cycle, as shown here, approximates the cyclic behaviour after many cycles. However where the amnesia on the first cycle is high, the cyclic stability is so poor that ratcheting or walking occurs, in other words the strain end points shift with increasing cycles.

When choosing a Ni-Ti alloy for an actuator type of application, it must be remembered that for a given composition, a high M_s temperature is obtained by annealing out any effects of prior cold work. The cyclic stability is therefore poor. Devices triggering at or above 100°C are unlikely to have mechanical stability over tens of thousands of cycles. However, it should be remembered that this is a *mechanical* instability, where the strain per cycle varies at high working loads. There is no *metallurgical* instability leading to shifts in the response temperature as a result of prolonged exposure around 100°C, in contrast to Cu-based SMA's.

Thermo-mechanical processing is important for the optimization of pseudoelastic behaviour. Increasing the yield strength of the austenite can widen the temperature range over which martensite can be stress induced without the intervention of plastic deformation. This will be discussed in detail in a later Chapter. From a mechanistic point of view however, the binary superelastic alloy is more complicated. If superelastic behaviour at ambient temperature is required, then M_s will be lower and the alloy will tend to be nickel rich. Solution treatment followed by rapid cooling and cold working will provide additional hardening from the precipitation of nickel rich phases as a result of the instabilities discussed above.

4. Corrosion Behavior

In the galvanic series, Ni-Ti based alloys as a family are slightly more noble than 316 stainless steel, and show similar corrosion behaviour. The excellent corrosion resistance is provided by a naturally formed thin adherent oxide layer known as a passive film. This film is very stable, so the Ni-Ti alloys are resistant to many forms of corrosive attack. However in some aggressive conditions such as highly acidified chloride solutions, breakdown of this passive film can occur. Such corrosive environments are very severe for most engineering materials; nevertheless if Ni-Ti is to be used in these conditions, some form of protective coating is advisable.

Electrochemical measurements show Ni-Ti based alloys have a good resistance to pitting in a chloride environment^{19,20}. However data obtained from a scratch test indicated that the healing of the passive film may be a difficult and relatively slow process^{20,21}.

Compared with other commercially available SMA's, Ni-Ti is by far the most corrosion resistant. For most applications in the actuator, electrical connector and fastener fields, Ni-Ti has a superior corrosion behaviour than the other components of the assembly. In the case of coupling applications, the corrosion resistance is more than adequate except in severe conditions, where protection is advisable. As a general rule, if no corrosion protection of the tubes or pipes is deemed necessary, none is required of the Ni-Ti coupling used to join them.

5. Effect of Alloying Elements

The addition of third or fourth elements to Ni-Ti provide a powerful tool for controlling the properties, and can be used to:

- control transformation temperatures
- increase the stability of M_s with respect to thermal history
- control the hysteresis width
- increase austenitic strength
- reduce or increase martensitic strength

- improve corrosion resistance
- suppress the R-phase

Some of the additions giving particularly useful combinations of properties are copper, niobium, and precious metals. These form the subject of separate chapters in this book.

For applications requiring M_s to be below room temperature, binary alloys show instability or an M_s dependence on prior thermal history. Furthermore they tend to have poor ductility. Alloying elements such as Fe, Co or Cr which are known to depress M_s and substitute primarily for Ni³, 21, 22, can be used as additions to an approximately 50 at % Ti alloy. In this range, (Figure 2) the sensitivity of M_s to Ni/Ti ratio variations is relatively low, but the M_s value corresponding to this plateau is lowered by the third element.

References

1. Wang F.E. Ed, Proceedings of Symposium on TiNi and related compounds, Report NOLTR 68-16 (1988).
2. Cross W.B., Kariotis A.H., and Stimler F.J. NASA Report CR-1433 (1969).
3. Jackson C.M., Wagner H.J., and Wasilewski R.J. NASA Report SP-5110 (1972).
4. Wasilewski R.J., Butler S.R., and Hanlon J.E., *Met. Sci.* 1, 104 (1974).
5. Chattopadhyay G., and Kleykamp H., *Z Metallkd.* 74, 182 (1983).
6. Mueller M.H., and Knott H.W., *Trans. AIME* 227, 674 (1963).
7. Wang F.E., Buehler W.J., and Pickart S.J., *J. Appl. Phys.* 36, 3232 (1965).
8. Mercier O., Melton K.M., and DePreville Y., *Acta. Met.* 27, 1467 (1979).
9. Nishida M., Wayman C.M., and Honma T. *Met. Trans.* 17A, 1505 (1986).
10. Perkins J., *Scripta Met.* 8, 1469 (1974).
11. Mohamed H.A., and Washburn J., *Met. Trans.* 7A, 10, (1976).
12. Mohamed H.A., and Washburn J. J. *Mater. Sci.* 12, 467, (1977).
13. Melton K.N., and Mercier O., *Met. Trans.*, 9A, 1487 (1978).
14. Mercier O. and Melton K.N., unpublished research, Brown Boveri, Switzerland.
15. Mercier O., Melton K.N., Gottardt R. and Kulik A., Proc.Int.Conf. on Solid-Solid Phase Transformations Ed Aaronson H.I., Laughlin D.L., Sekerka R.F., and Wayman C.M., pp.1259-63 (1982).
16. Buehler W.J., and Wiley R.C., *US Patent* 3,174,851 (1965).
17. Rozner A.G., and Buehler W.J., *US Patent* 3,351,463 (1967).
18. Rozner A.G., and Buehler W.J., *Trans. ASM* 59, 350 (1966).
19. Vicentini B., Rondelli G., Cigada A., and Turisini G. Shape Memory Alloy 86, China Academic Publishers, p. 447 (1986).
20. Rondelli G., Vicentini B., and Cigada A., Corrosion Properties of NiTi Shape Memory Alloys, MRS Tokyo (1988).
21. Schwenk W. and Huber J., *SAMPE Quart* 5, 17 (1974).
22. Eckelmeyer E.K., *Scripta Met.*, 10, 667 (1976).

Introduction to the R-Phase Transition

K. Otsuka
Institute of Materials Science, University of Tsukuba

Shape memory alloys are now recognized as functional materials and may be used in various ways. One important field of application is as thermal actuators. Although these devices generally require a small thermal hysteresis, most martensitic transformations are associated with a temperature hysteresis in excess of 10°C. However there is a martensite-like transition which is associated with a temperature hysteresis as small as 1.5°C: the so-called R-phase transition. The R-phase transition appears upon cooling, prior to the martensitic transformation in Ni-Ti alloy systems under certain conditions to be described later.

Figure 1 shows a typical resistance vs. temperature curve for a solution-treated Ti-50.0 at.%Ni alloy (the characteristic temperatures M_s , M_f , A_s and A_f are indicated). We observe a very simple and usual behavior. However, in the $Ti_{50}Ni_{47}Fe_3$ alloy shown

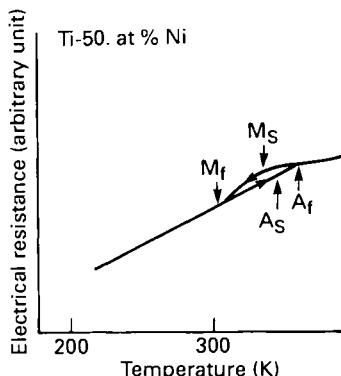


Figure 1: Electrical resistance vs. temperature curves for Ti-50.0 at%Ni alloy, which was solution treated at 1000°C for 1 hour and quenching.¹

in Figure 2², we observe a very sharp increase in resistance at T_R' , which then further increases gradually beyond T_R until M_s . The T_R temperature is a few degrees below T_R' and is close to the inflection point. Between T_R' and T_R , there is a slight crystallographic change, but the lattice remains cubic without any lattice distortion².

The lattice begins to distort at T_R . T_R is thus the onset, or start temperature R-phase tension. Elsewhere in this book the terms R_s will be used, with R_t then the temperature at which the transformation to the R-phase is complete. The transition is characterized by a very small temperature hysteresis, as small as 1.5°C . Thus, the transition is extremely useful in thermal actuator applications.

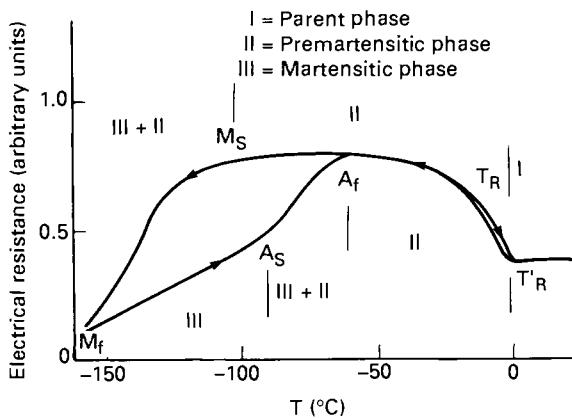


Figure 2: Electrical resistance vs. temperature curve for a $\text{Ti}_{50}\text{Ni}_{47}\text{Fe}_3$ alloy.²

Figure 3³ shows how the lattice changes with decreasing temperature. The parent phase is cubic, and the lower temperature phase is rhombohedral, thus the transition is called the R-phase transition. The cubic lattice changes as if it is elongated along

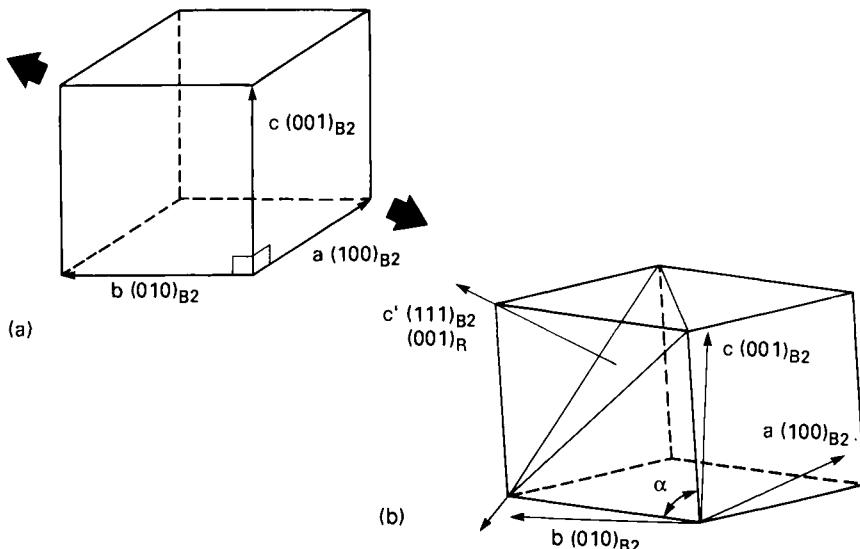


Figure 3: Lattice change associated with the R-phase transition: (a) the B_2 type parent phase and (b) the R-phase.³

one of the diagonal lines. The cube angle α decreases from 90° with decreasing temperature, as shown in Figure 4.^{4,5} Thus, the lattice distortion increases with decreasing temperature. With further cooling the martensitic transformation starts at the M_s temperature (Figure 2).

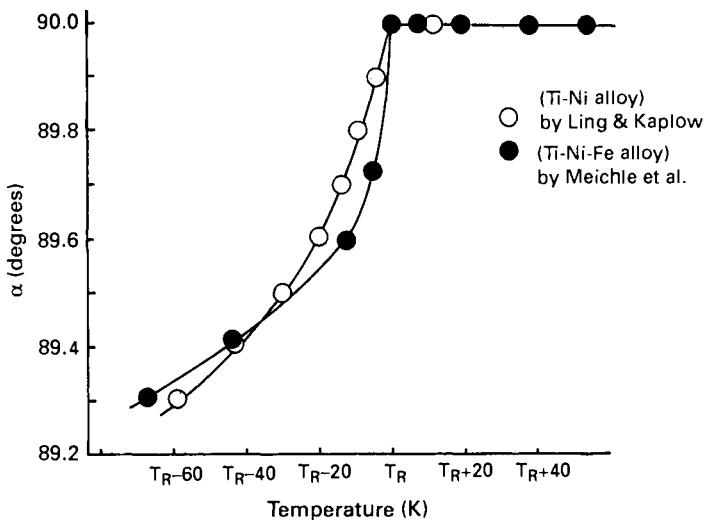


Figure 4: The change of angle α with temperature.^{4,5}

1. The Shape Memory Effect and Superelasticity

Since the R-phase transition is associated with a lattice distortion, one might expect to observe shape memory and superelasticity. Figure 5⁶ shows a series of stress-strain curves as a function of temperature. The dashed lines in the figure mean that the apparent permanent strain after loading and unloading recovers by heating to temperatures above T_R (i.e. there is a shape memory effect). In Figure 5 (h)-(j), the strain recovers just by unloading, without heating (superelasticity). Thus both shape memory and superelasticity are associated with the R-phase transition. This means that the R-phase transition is a kind of martensitic transformation with a very small thermal hysteresis. The critical stresses (indicated by arrows in Figure 5) and the recoverable strains are plotted in Figure 6.⁶ The recoverable strain increases with decreasing temperature, consistent with the fact that the lattice distortion increases with decreasing temperature, as shown in Figure 4. We also note that the maximum recoverable strain is only about 0.5%. In other alloys this could be as high as 1% (see Figure 9), but still is quite small compared to martensitic shape memory.

Figure 7³ shows the orientation and temperature dependencies of the recoverable strain in single crystals, which were calculated from Figure 4. The $[-1\ 1\ 1]$ orientation gives the largest recoverable strain, as expected from the lattice deformation mechanism described in Figure 3. The above orientation and temperature dependencies were found to agree well with experimental results in single crystals.³

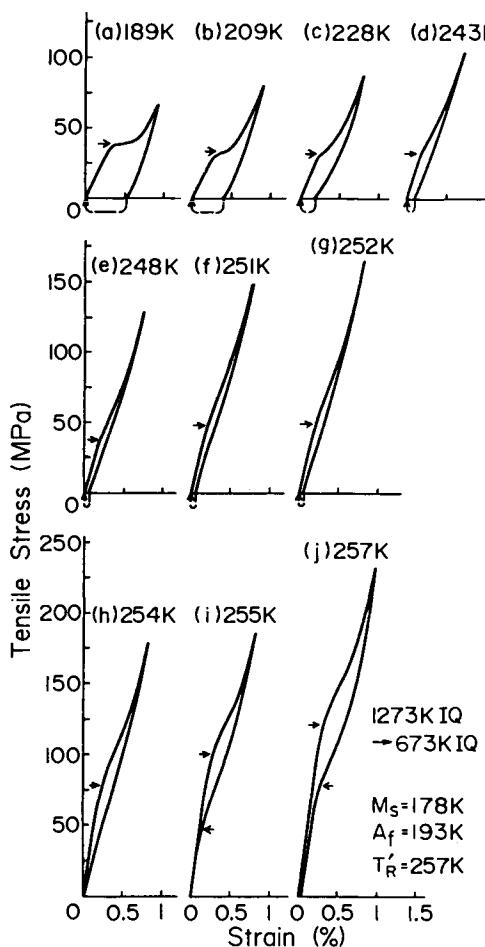


Figure 5: Stress-strain curve at various temperatures for a Ti₅₀Ni₄₇Fe₃ alloy. The dashed lines indicate the strain recovery on heating to 100°C

2. Conditions for Obtaining the R-phase Transition

The basic approach to realizing the R-phase transition is to recognize that we have two competing transformations: the R-phase transition and the martensitic transformation. Thus, the R-phase can be realized by suppressing the martensitic transformation relative to the R-phase. There are three ways to do this:¹

- (a) The introduction of rearranged dislocations, produced by cold-working then annealing at temperatures between 400°C and 500°C.
- (b) The introduction of precipitates by solution-treated and aging alloys of greater than 50.5 at % Ni at temperatures between 400°C and 500°C.
- (c) Addition of certain third elements that suppress a martensitic transformation (e.g. Fe or Al).

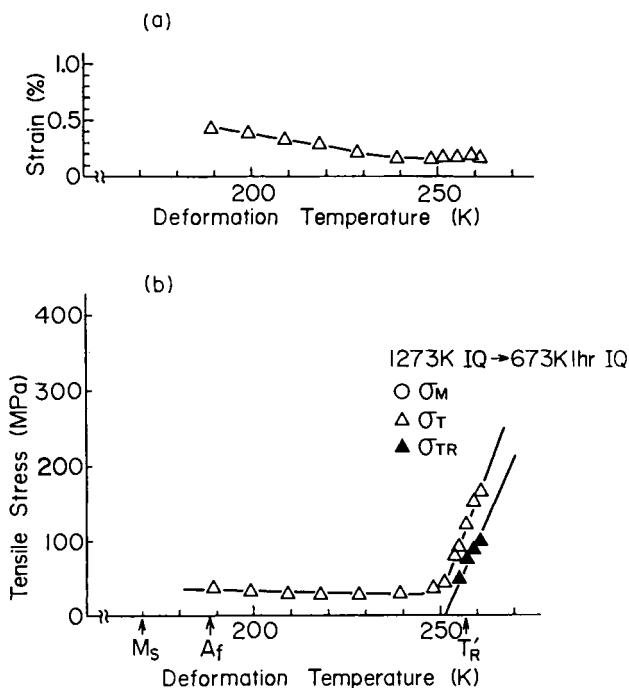


Figure 6: (a) Transformational strain of the R-phase as a function of deformation temperature, and (b) critical stresses in the first stage formation as a function of temperature.

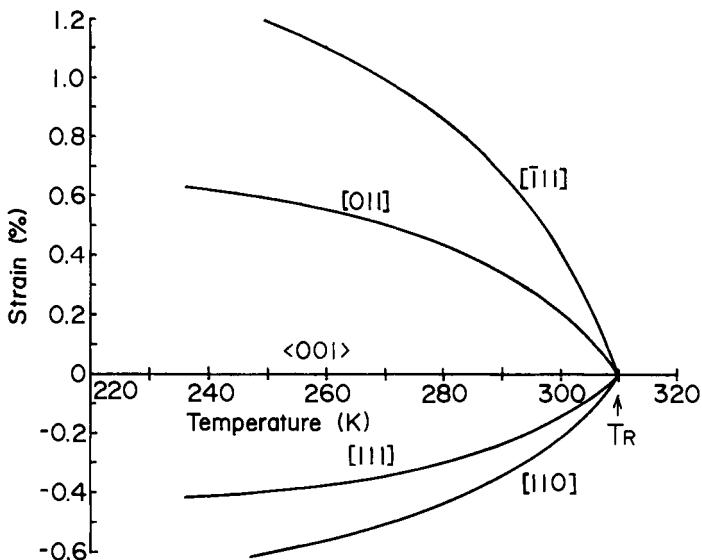


Figure 7: Calculated temperature dependence of the recoverable strains for various orientations.³

The effect of the thermo-mechanical treatments in (a) and (b) above are shown in the resistance vs. temperature curves of Figure 8.¹ When a specimen of Ti-50.6 at.%Ni is solution-treated, no R-phase transition appears, as shown in Figure 8 (a); but when it is cold-worked and followed by an anneal at 400°C, the R-phase stands out (Figure 8 (c)). Similarly, when it is solution-treated at 1000°C followed by aging at 400°C, the R-phase appears (Figure 8(b)). The effect of ternary additions was clearly shown in Figure 2.

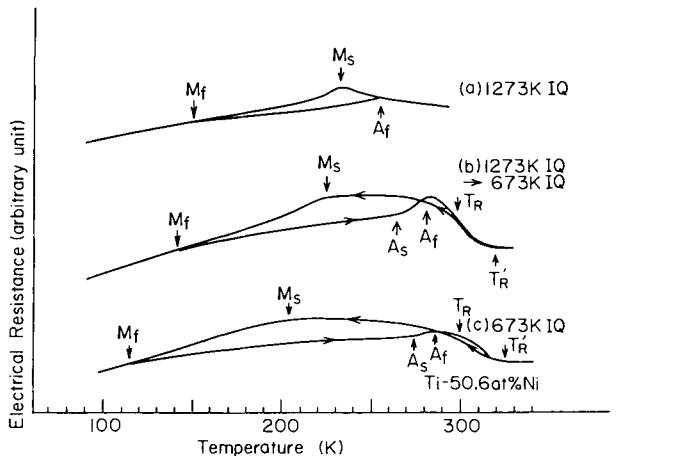
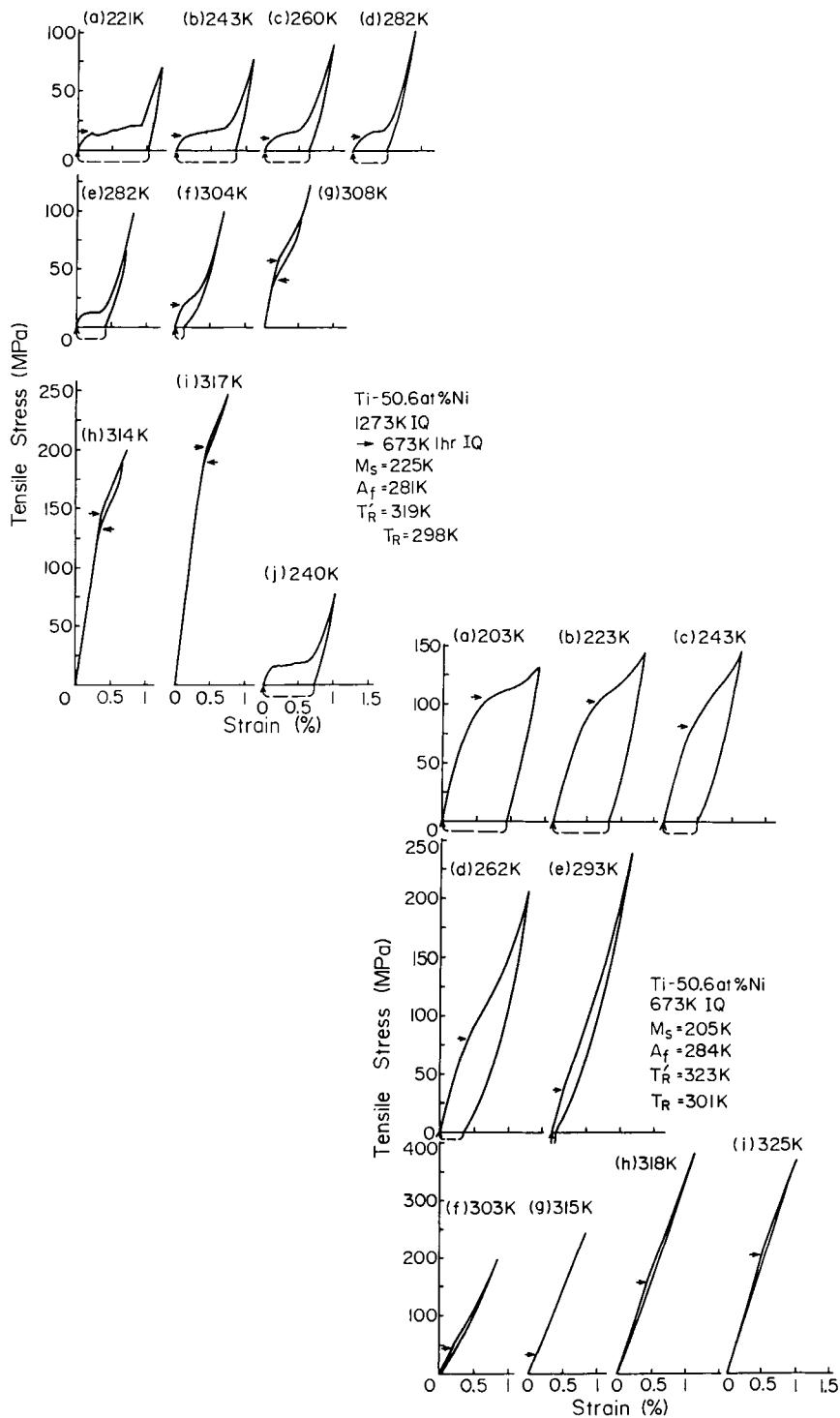


Figure 8: Electrical resistance vs. temperature curves for a Ti-50.6 at%Ni alloy which were thermo-mechanically treated as indicated.¹

The mechanical behavior associated with these heat treatment is shown in Figure 9(a) and (b)¹. Both aged, and cold-worked and annealed specimens show shape memory and superelasticity associated with the R-phase transition, but there are some differences. When the specimen is aged after solution-treatment, the critical stresses are low (Figure 9(a)), but when it is annealed after cold-working, the critical stresses are higher. This means that the dislocations have a stronger effect on the mechanical properties associated with the R-phase transition.

3. Mechanical Behavior Under Constant Load Conditions

Mechanical test results under constant temperature conditions were shown above. In this section, we will show some results under constant load conditions, which may be more useful in many applications. Figure 10 shows a strain vs. temperature curve under constant load.⁷ When the load is small, the temperature hysteresis is small (characteristic of the R-phase transition). But when the load is large, the temperature hysteresis is large, which is indicative of the induction of martensite. Thus, the application of load is limited when the R-phase transition is used. Based on a series of such experiments, Todoroki proposed the phase diagram shown in Figure 11.⁸ As can be seen, transformation occurs directly from the parent phase to martensite when the applied load is large.



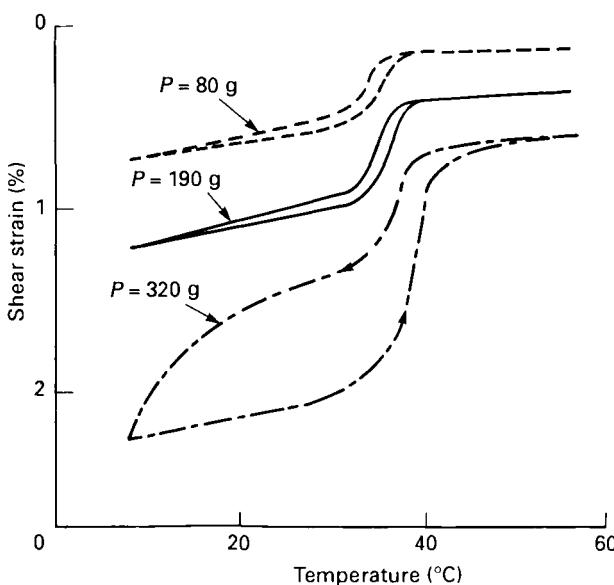


Figure 10: Shear strain vs. temperature curve under constant load for a Ni-Ti alloy.⁷

4. The Shape Memory Effect under Cyclic Conditions

When the shape memory effect is used for actual applications, the stability of the effect under cyclic loading conditions is very important. The results of an experiment aimed at characterizing this is shown in Figure 12.⁹ The deflection vs. temperature curve is essentially unchanged after 500,000 cycles. This result also indicates that the fatigue life for the R-phase is also very good compared to the martensite.²¹ Thus the shape memory effect associated with the R-phase was successfully applied to commercial air-conditioners by Matsushita Electric Co., and is quite promising for further use.

5. Historical

The R-phase transition was first observed by x-ray diffraction by Dautovich and Purdy,¹⁰ and Wang et. al.,¹¹ but the nature of the transition and the relation with the subsequent martensitic transformation was not clear. These works were followed by Sandrock et. al.¹² Vatanayon and Hehemann¹³ first proposed the structure as

Figure 9: (a) Stress-strain curves at various temperatures for a Ti-50.6 at.-%Ni alloy which was aged at 400°C for 1 hour after solution treatment. The arrows indicate the critical stress for the first stage; the dashed lines indicate the strain recovery upon heating to 100°C. (b) Stress-strain curves at various temperatures for a Ti 50.6 at.-%Ni alloy which was annealed at 400°C for 1 hour immediately after cold-working.¹

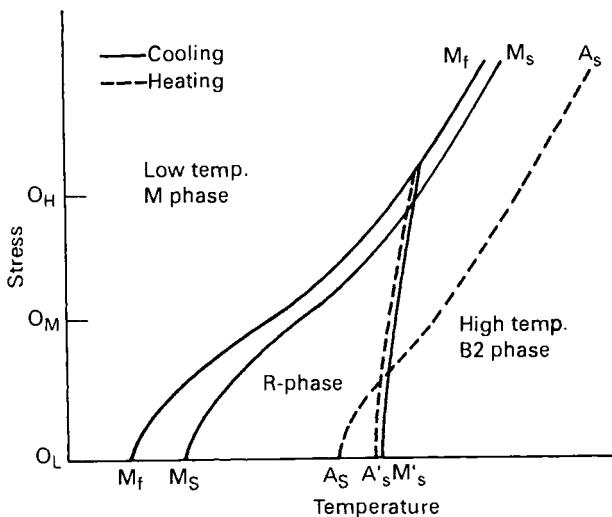


Figure 11: Schematic phase diagram in temperature-stress coordinates for Ti-Ni alloys.⁸

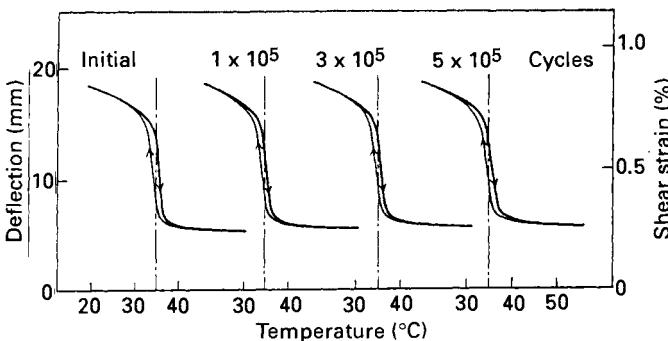


Figure 12: Effect of cycling on the deflection vs. temperature curves for a Ti-Ni alloy. Load was 62 MPa.^{2,9}

rhombohedral, and more extensive studies on the nature of the transition were made by Salamon and Wayman using neutron, x-ray and electron diffraction techniques.^{2,14,15} They interpreted the transition as originating from charge density waves. However, Yamada et. al.^{16,17} carried out detailed neutron and x-ray diffraction experiments and as a result developed a modulated lattice relaxation model.

The mechanical behavior associated with the R-phase transition was first investigated by Khachin et. al.,^{18,19} who were first to find the shape memory effect associated with the transition. Ling and Kaplow²⁰ also studied the shape memory effect in detail. Miyazaki and Otsuka made more extensive investigations on the associated mechanical behavior, and found superelasticity associated with transition.^{1,6} They also carried out a quantitative study using single crystals³.

Readers are referred to Refs. 2 and 17 for more details on the transition mechanism, and Refs. 1,3 and 6 for mechanical behavior.

6. Conclusions

Although the recoverable strain is limited to 1%, the R-phase transition can be extremely useful for accurate thermal actuators and proportional control devices since the temperature hysteresis is as small as 1.5°C and the temperature vs. strain curve is quite stable during thermal cycling.

References

1. Miyazaki, S. and Otsuka, K.: Met. Trans. A, 17, 53 (1986).
2. Hwang, C.M., Meichle, M., Salamon, M.B. and Wayman, C.M.: Phil. Mag. A, 47, 9, 31 (1983).
3. Miyazaki, S., Kimura, S. and Otsuka, K.: Phil. Mag. A, 57, 467 (1988).
4. Ling, H.C. and Kaplow, R. : Met. Trans. A, 11, 77 (1981).
5. Meichle, M.E. Ph.D. Thesis, University of Illinois (1981).
6. Miyazaki, S. and Otsuka, K. : Phil. Mag. A, 50, 393 (1984).
7. Todoroki, T.: J. Japan Inst. Met., 49, 439 (in Japanese), (1985).
8. Todoroki, T., Tamura, H. Suzuki, and Y.: Proc. ICOMAT-86 at Nara, p.748 (1986).
9. Todoroki, T.: Ph.D. Thesis, Tohoku University (1987).
10. Dautovich, D. P. and Purdy, G. R.: Can. Metall. Q., 4, 129 (1965).
11. Wang, F.E., Buehler, W.J. and Pickart, S.J.: J. Appl. Phys., 36, 3232 (1965).
12. Sandrock, G.D., Perkins, A. J. and Hehemann, R.F.: Met. Trans., 2, 2769 (1971).
13. Vatanayon, S. and Hehemann, R.F.: Shape Memory Effects in Alloys, (ed. J. Perkins), Plenum Press, N.Y., p.115 (1975).
14. Salamon, M.B., Meichle, M., Wayman, C.M. and Shapiro, S.M.: Modulated Structures, (ed. J.B. Cohen, M.B. Salamon and B.J. Buensch), A.I.P. conf., Proc. No. 53, American Institute of Phys., p.223.
15. Satija, S.K., Shapiro, S.M., Salamon, M.B. and Wayman, C.M. : Phys. Rev. B, 29, 6031 (1984).
16. Yamada, Y., Noda, Y. and Takimoto, M.: Solid State Commun., 55, 1003, (1985).
17. Yamada, Y.: Proc. ICOMAT-86 at Nara, p.89 (1986).
18. Khachin, V.N., Paskal, Y.I., Guter, V.E., Monasevich, A.A. and Sivokha, V.P.: Phys. Metals Metallogr., 46, 49 (1978).
19. Khachin, V.N., Gunter, V.E., Sivokha, V.P. and Savvinov, A.S.: Proc. ICOMAT-79 at Boston, p.474 (1979).
20. Ling, H.C. and Kaplow, R.: Met. Trans. A, 11, 77; Ibid. (1980), 12 2101 (1981).
21. Tamura, H. , Suzuki, Y. and Todoroki, T.: Proc. ICOMAT-86 at Nara, P.736 (1986).

Ni-Ti-Cu Shape Memory Alloys

Warren J. Moberly[†] and Keith N. Melton^{*}

[†] Stanford University
Palo Alto, CA

^{*} Raychem Corp.
300 Constitution Dr.
Menlo Park, CA

The shape memory properties of Ni-Ti can be readily modified by adding ternary elements which are chemically similar to Ni or Ti. Other transition metals may substitute for Ni, however small additions (1-3 at %) of elements such as Co, Fe or Cr have been shown^{1,2} to depress the martensitic transformation temperature such that the shape memory effect occurs well below ambient. In contrast, up to 30% Cu, which is a neighbor of Ni in the periodic table, may be substituted while retaining the same high temperature austenitic phase³. A minor change in the low temperature martensite structure (from monoclinic to orthorhombic) of alloys containing more than approximately 10% Cu does not appreciably affect the basic shape memory phenomenon. Certain associated property modifications, in particular a more narrow transformation hysteresis and lower martensitic yield strength, are actually beneficial for many applications. In this chapter the effects and advantages of Cu additions to Ni-Ti alloys are reviewed.

1. Effects of Copper Additions on Microstructure

As noted in Chapter 2, the Ni-Ti phase has limited solubility for either Ni or Ti. The Cu-Ti binary system has a phase diagram very similar to that of Ni-Ti⁴. However it should be noted that the austenitic phase of NiTi has a B2 cubic structure, whereas the CuTi compound is tetragonal⁵. As a consequence, more than 30% Cu will result in the presence of both tetragonal and cubic phases. When Cu is substituted for Ni in Ni-Ti SMA's, the overall Ti concentration must remain close to 50 atomic % in order to yield a single phase material desirable for shape memory applications.

Although the austenitic phase retains its cubic structure for Cu contents up to 30%, the pseudobinary phase diagram of Figure 1 does not consider the martensitic transformation occurring near room temperature. For alloys containing ~10% Cu or less, the austenite transforms to an oblique (monoclinic) martensite typical of the binary Ni-Ti alloy at low temperature. However, higher Cu - containing alloys transform to a perpendicular (orthorhombic) martensite. This orthorhombic structure has been observed using X-ray diffraction and transmission electron microscopy⁶⁻⁸. Table 1 lists the lattice parameters of the binary and different ternary alloys. The lattice parameters themselves vary little as a function of Cu concentration, with the major difference being the change from a 96 degree angle to a 90 degree angle for higher copper contents.

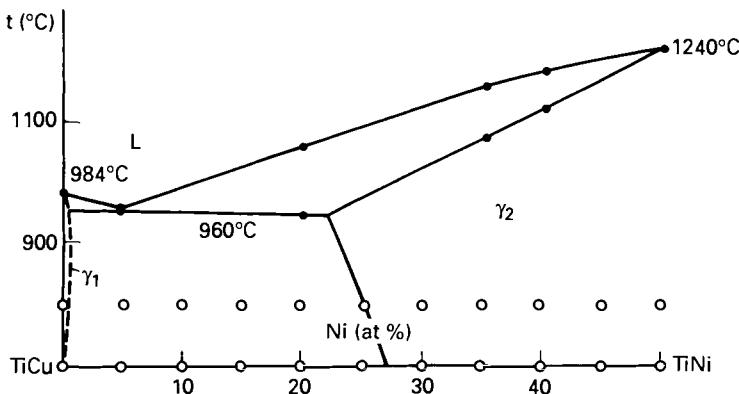


Figure 1: Pseudobinary phase diagram between the high temperature austenite phases of CuTi (tetragonal) and NiTi (cubic)⁸.

Table I: Lattice Parameters (in nm) for Different Cu Concentrations

	<u>0% Cu</u>	<u>5% Cu</u>	<u>10% Cu</u>	<u>15% Cu</u>	<u>20% Cu</u>	<u>25% Cu</u>
a_0 (austenite)	0.3020	0.3022 ^b	0.303 ^b	0.302	0.3016 ^d	0.304 ^b
a_0 (martensite)	0.2889 ^a	0.2889 ^c	0.287 ^c	0.289 ^d	0.2889 ^d	0.287 ^b
b_0 (martensite)	0.4622 ^a	0.4622 ^c	0.457 ^c	0.45 ^d	0.4514 ^d	0.453 ^b
c_0 (martensite)	0.4120 ^a	0.412 ^c	0.411 ^c	0.426 ^d	0.4265 ^d	0.426 ^b
γ (angle between a_0 and b_0)	96.8° ^a	96.8° ^c	95° ^c	90° ^d	90° ^d	90° ^d

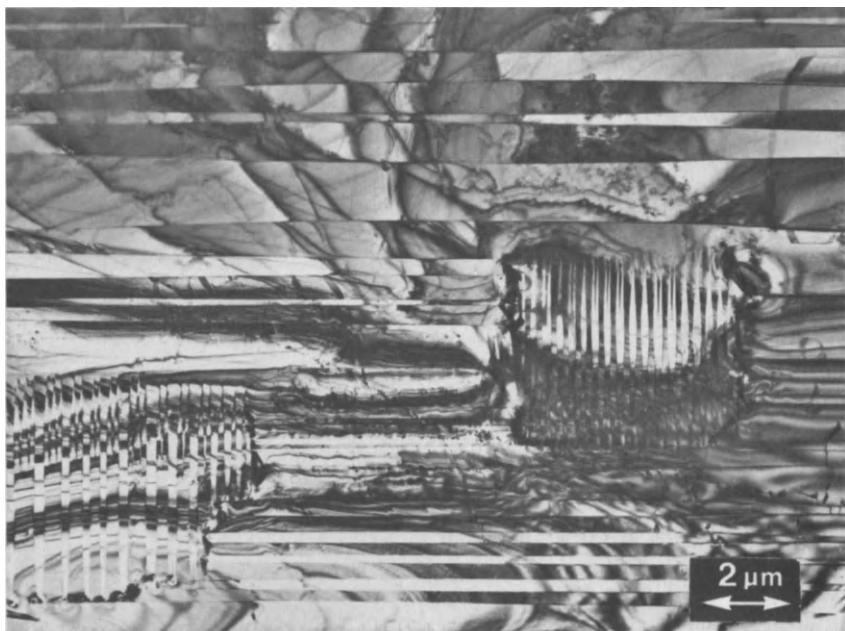
^a from reference 20

^b from reference 11

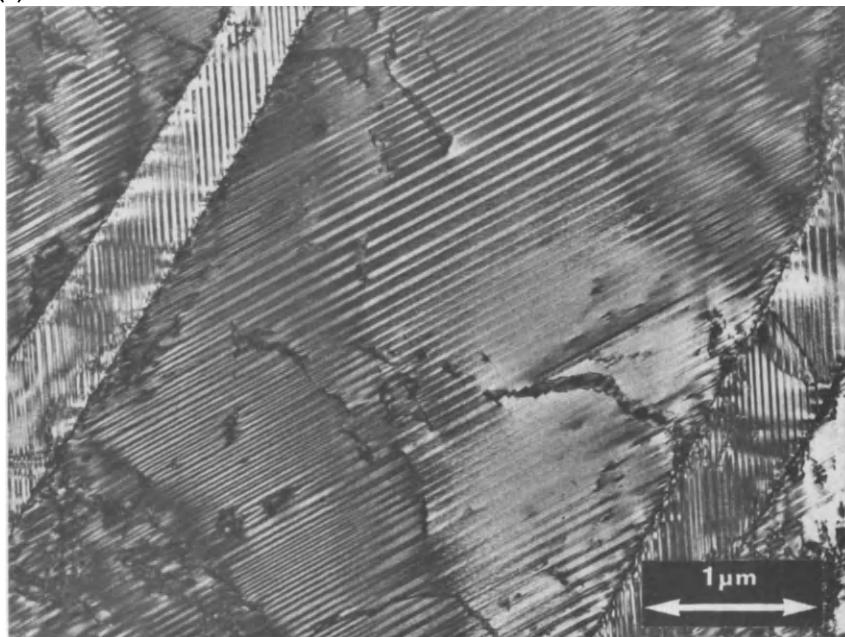
^c from reference 7

^d from reference 8

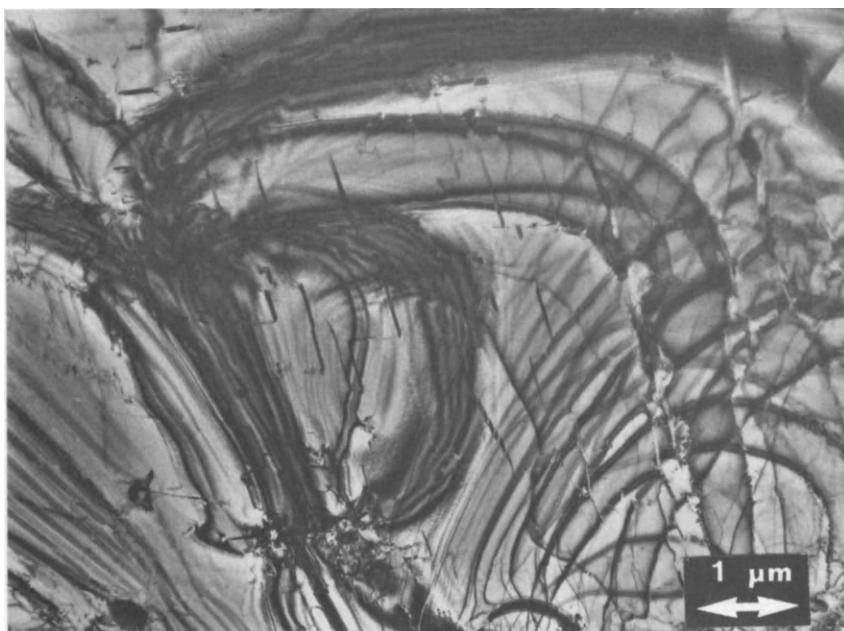
The monoclinic martensite in Ni-Ti-Cu alloys has a twinned microstructure (see Figure 2a) very similar to that of the binary alloy⁹. However, for alloys with higher Cu contents that have the orthorhombic martensite, there are significantly fewer twins. Figure 2b shows the microstructure of a 15% Cu alloy where some grains with twins can be seen. As a result of the higher symmetry of the orthorhombic structure, not all twinning planes are present in the high Cu ternary alloys so an untwinned microstructure can be obtained, see Figure 2c. The twinning differences between the two types of martensite are illustrated in Figure 3. The presence of twinless grains of martensite is unusual for such transformations, and their contribution to the shape memory effect of these high Cu alloys is yet to be established^{8,10}.



(a)



(b)



(c)

Figure 2: Transmission electron micrographs: (a) Ni-Ti-10%Cu alloy with finely twinned monoclinic martensite; (b) Ni-Ti-15%Cu alloy with twinned orthorhombic martensite; (c) Ni-Ti-15%Cu alloy with twinless orthorhombic martensite.

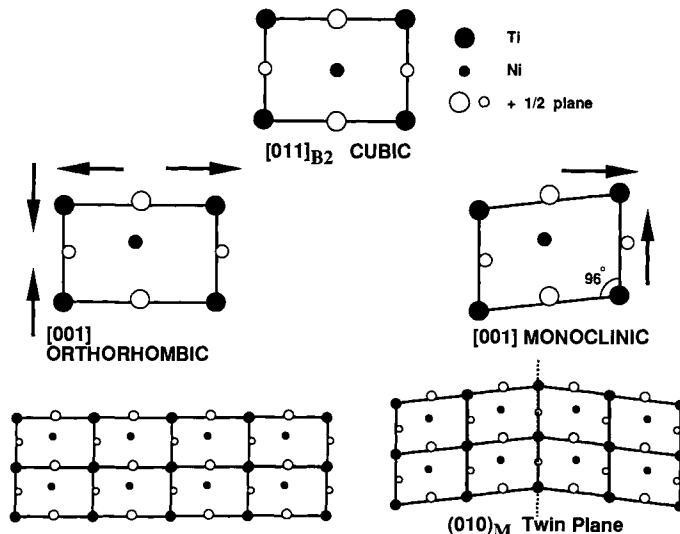


Figure 3: Two-dimensional, schematics of the unit cells for the two different martensitic structures. The monoclinic martensite requires twinning, in this case with a (010) twin plane. The (010) plane will not involve twinning for orthorhombic martensite.

Early microstructural observations of Ni-Ti-Cu martensites reported a monoclinic structure at all Cu contents¹¹. A later more detailed study⁷ observed orthorhombic martensite for high Cu alloys. However the particular structure formed can depend on the prior thermal and processing history of the specimen. Figure 4 exhibits the structure of an alloy with 25% Cu, exhibiting a monoclinic martensite in the as - cast condition. A long anneal at high temperature ($>0.8T_m$) results in the formation of an orthorhombic martensite. Consistent processing conditions are required in order to produce alloys with uniform microstructures and controllable properties.



Figure 4: TEM micrograph of Ni-Ti-25%Cu exhibiting a monoclinic martensite structure after casting. The equilibrium orthorhombic martensite is obtained after annealing at $>0.8T_m$. The microstructures (and shape memory properties) are very dependent on processing history.

Alloys with 10% Cu have been reported^{12,13} as undergoing a two-step transformation with decreasing temperature. X-ray diffraction¹² indicates that austenite transforms to an orthorhombic martensite phase at intermediate temperatures, prior to transforming to the low temperature monoclinic martensite structure. The effect of Cu concentration on this martensite to martensite transformation is not well established⁸. The presence of two stable martensite structures for 10% Cu alloys may relate to differences reported in the structure of higher Cu containing martensites, noted above.

The reactivity of titanium and the resulting formation of oxide impurities of the type $Ti_4Ni_2O_x$ has been discussed in Chapter 2. Similar behavior is found for Ni-Ti-Cu alloys, the composition of the oxide being $Ti_4(Ni,Cu)_2O_x$. Because the oxide phase is titanium rich, it is more prevalent in titanium rich alloys, and furthermore, its formation alters the Ni-Ti ratio in the matrix and consequently affects M_s . There is also a tendency for more of this oxide phase to be present in the higher Cu alloys. In extreme cases in Ni-Ti-Cu, this brittle oxide phase can form as a semi-continuous film at a grain boundaries, resulting in poor workability of the as-cast structure.

2. Effects of Copper Additions on Shape Memory Properties

The addition of even small concentrations of many third elements to Ni-Ti results in a large change in the M_s temperature such that controllable adjustments of M_s are not easily achieved^{1,2}. In contrast, substitution of even large concentrations of Cu does not change the M_s temperature significantly. Figure 5 plots M_s as a function of Cu for a

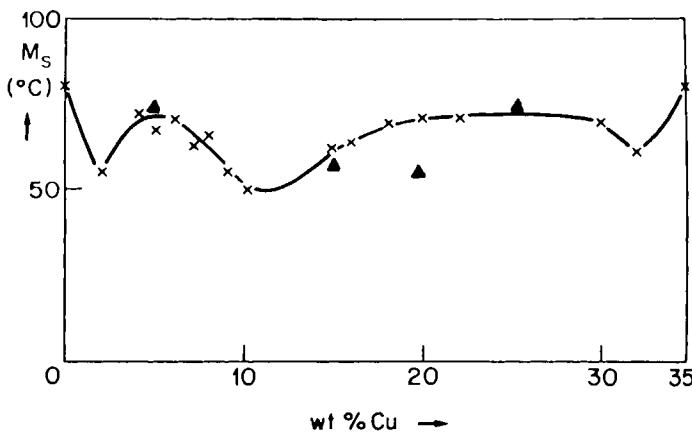


Figure 5: M_s varies with % Cu, as determined by differential scanning calorimetry (triangles) and resistivity¹⁴ (crosses).

constant titanium concentration. Data from both differential scanning calorimetry (DSC) and resistivity experiments¹⁴ are provided. Differences between the two may be attributed at least in part to different melting and processing techniques from different laboratories. The addition of copper results in various alloys all with approximately the same M_s , even though properties, such as hysteresis width may significantly change.

The presence of Cu also makes the M_s temperature less sensitive to variations in the Ni-Ti ratio. In the binary alloy, M_s drops sharply as the Ni concentration increases from 50 to 51 atomic %. This sensitivity is suppressed by the addition of Cu, see Figure 6. A less concentration dependent M_s allows for easier production of commercial quantities of material having a controlled M_s for actuator use.

Electrical resistivity measurements¹⁵ indicate that alloys with more than 20% Cu have a substantially more-narrow hysteresis than the binary alloy. Table II lists the

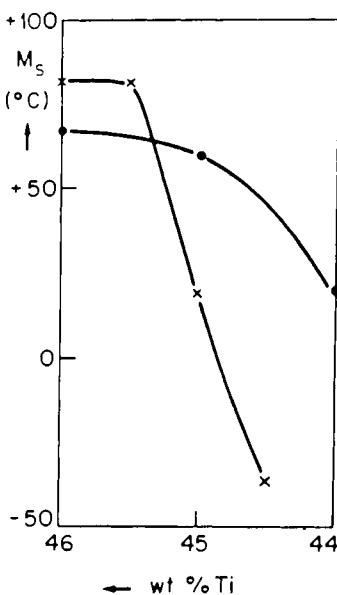


Figure 6: The M_s temperature decreases sharply as the Ni:Ti ratio increases. This concentration sensitivity of the M_s is less severe for ternary Ti-Ni-5%Cu (dots) than for the binary (crosses) alloys¹⁴.

Table II: Transformation Temperatures vs Cu Concentrations.

	<u>0% Cu</u>	<u>5% Cu</u>	<u>15% Cu</u>	<u>20% Cu</u>	<u>25% Cu</u>
M_s	38°	75°	56°	48°	76°
M_f	20°	56°	45°	29°	67°
A_s	56°	62°	50°	35°	70°
A_f	82°	78°	62°	55°	80°
ΔT_o	42°	3.4°	5.6°	4°	3.6°

Note: Temperature measurements (in °C) are based on Differential Scanning Calorimetry measurements. The hysteresis (ΔT_o) is measured as the temperature difference between the peaks on the DSC curves upon heating and cooling.

transformation temperatures and hysteresis widths, determined by DSC, for different Cu concentrations. It can be seen that this hysteresis is reduced from around 30°C for the binary alloy to less than 15°C for material with 5% Cu. Further Cu additions decrease the hysteresis only slightly, with a 25% Cu alloy having a width of only 10°C. Figure 7 displays DSC plots for a binary and a 25% Cu alloy, where these differences can clearly be observed.

The narrow transformation hysteresis that occurs during thermal cycling is also observed in a deformation cycle. When a sample is deformed at a constant

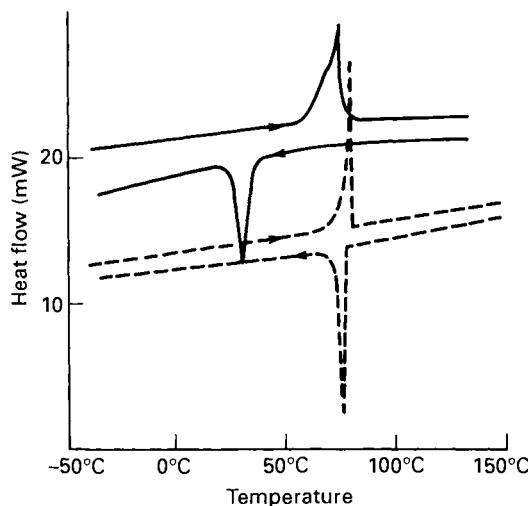


Figure 7: Differential scanning calorimetry graphs exhibiting the transformations upon heating and cooling for binary (solid line) and 25%Cu (dashed line) alloys.

temperature just above M_s , martensite can be stress induced. If this temperature is also above A_f , then the martensite will revert upon unloading, with the result being a complete transformation cycle. A stress hysteresis is associated with the transformation, with the stress-induced martensite transformation due to loading occurring at a higher stress level than the reverse transformation upon unloading. This stress hysteresis is much narrower for Ni-Ti-Cu alloys than for the binary.

The narrower hysteresis of Ni-Ti-Cu alloys has practical importance. Applications requiring a fast response time on thermal cycling are easier to realize with a narrow hysteresis alloy. The superelastic properties of Ni-Ti are also altered by the narrow hysteresis of ternary alloys. The maximum stress to induce martensite is similar for both the binary and Ni-Ti-Cu alloys and is limited by the austenitic yield strength. Thus the narrower stress hysteresis of Ni-Ti-Cu means that for a similar process condition, the unloading or reversion stress is higher for the ternary alloy. The stored energy density of Ni-Ti-Cu superelastic springs is correspondingly higher than those of the binary alloy. Superelasticity will be discussed in more detail later in this book.

Another property influenced by copper additions is the yield strength of the martensite, i.e. the stress level at which the twins will re-orient. Table III¹⁶ lists the yield strengths of binary Ni-Ti and a Ni-Ti-10%Cu alloy, tested in both the martensitic and austenitic phases. Although processed to provide similar austenite yield strengths, the martensitic strength of the Cu containing alloy is almost half that of the binary. This can be important for cyclic applications. One way of redeforming a device when it cools back to the martensitic phase is to provide a reset spring. On heating, the SMA can do useful external work, but some of the available energy must be stored in this reset spring. Subsequent cooling of the SMA will cause it to soften when transforming to martensite, and if designed correctly the reset spring will then apply sufficient force to deform it back to its original low temperature position. The softer martensite of the ternary alloy requires a lower reset force, so less "redundant" work is dissipated per

cycle. The larger the strength differential between the austenite and martensite phases, the more the work that can be done by the SMA during recovery.

Since many shape memory components are used in multiple cycle applications, the stability of the properties of SMA's during cycling is a major concern. A reliable component should exhibit a constant M_s independent of cycling. Secondly, the amount of actually recovered shape change (or strain) should be independent of the number of cycles. Any inability of a shape memory component to retain these constant properties can be defined as "fatigue". The cyclic dependency of the M_s temperature of annealed binary and ternary (10% Cu) alloys are shown in Figure 8¹⁶. As indicated, the M_s temperature of the binary increases by around 20°C during the first 10 thermal cycles. Further cycling does not significantly change M_s . The ternary alloy however exhibits an M_s which is relatively constant with cycling. A comparison of the cyclic

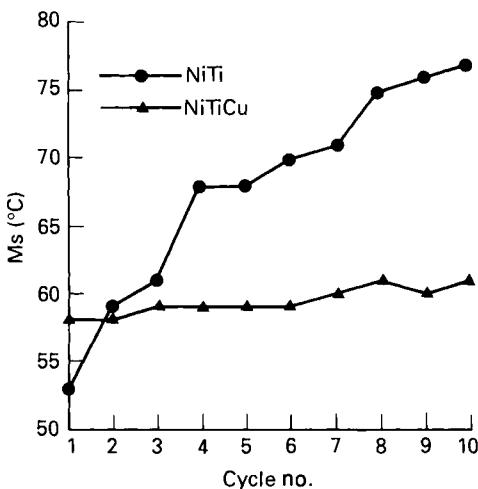


Figure 8: The M_s temperature is less sensitive to transformation cycling for the ternary (10%Cu) alloy than for the binary alloy¹⁶.

dependence of the recoverable strains of the binary and ternary alloys obtained during the transformation due to an applied load is shown in Figure 9¹⁶. Data are shown for two stress levels, with the strains of the binary alloy decreasing rapidly with number of cycles whereas the values for the Ni-Ti-Cu alloy remained relatively constant. The initial strain of the Cu alloy is lower, however. This is probably a consequence of the lower M_f , such that cooling to ambient temperature only causes partial transformation.

The reasons for these differences in cyclic behavior are not clear. One possibility is associated with incomplete cycling. Heating was terminated at 100°C, which was below A_f for the binary. Thus some martensite would be retained in this case, but not in the ternary. On the other hand, M_f for the Ni-Ti-Cu alloy is well below M_s , and cooling to room temperature does not allow the transformation to martensite to go to completion for the ternary alloy. This is one reason why the initial cyclic strain is less

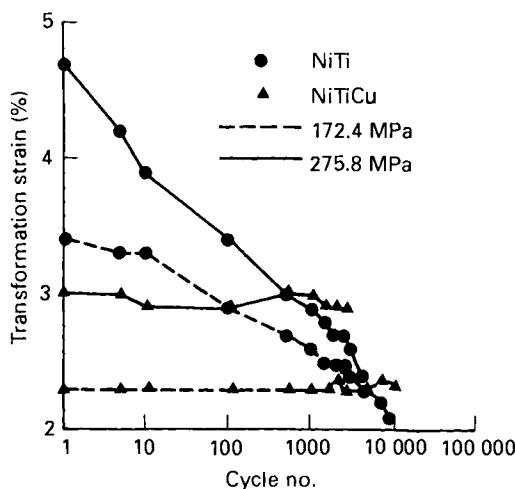


Figure 9: The recoverable strain exhibits less fatigue due to transformation cycling for the ternary (10%Cu) alloy than for the binary alloy¹⁶.

(Figure 9). Despite these uncertainties as to mechanism, it is clear that Ni-Ti-Cu offers many potential advantages as a SMA of cyclic applications.

As noted above, alloys with around 10 at.% Cu can exhibit a two-step transformation on cooling, from austenite to orthorhombic martensite to monoclinic martensite. This is also manifested in two-step property changes observed during the transformation¹⁷. Figure 10 plots the strain occurring in a 10% Cu alloy during thermal cycling under different applied loads. The transformation curves display discontinuities (steps) upon both cooling and heating. If the applied load is too great, >150MPa, the discontinuity is not observed, implying either a direct transformation to monoclinic martensite, or a

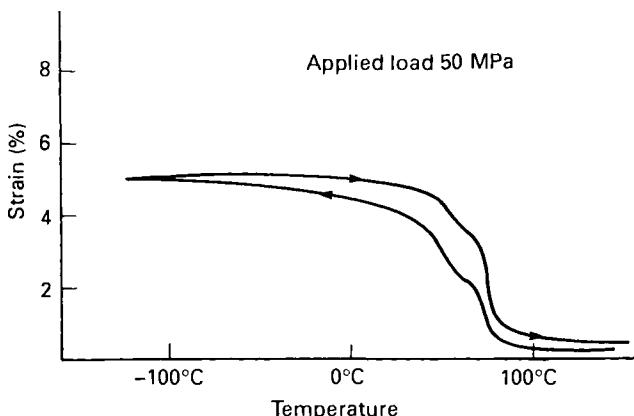


Figure 10: The strain - temperature transformation curve for the 10%Cu ternary alloy exhibits a 2-step change, indicative of the two successive martensite phases.

smearing out of the two transformations such that they overlap. For moderate loads however, the two-step change is reproducible for many cycles, and could be beneficial for actuators requiring two-step motion.

This two step transformation can also be detected using resistivity measurements. Figure 11. Both the narrow hysteresis and the discontinuity are well distinguished for

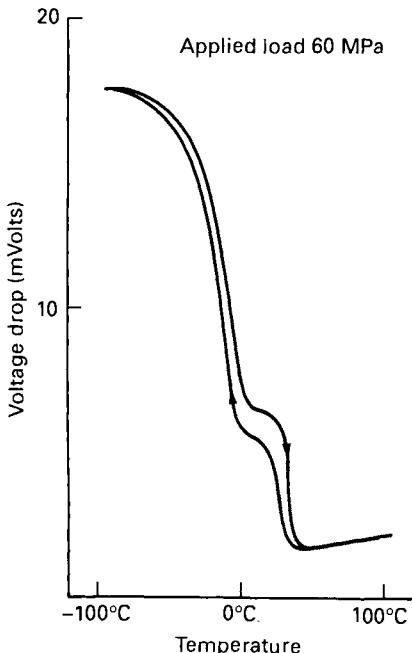


Figure 11: The resistance change with temperature for the 10%Cu alloy exhibits a 2-step transformation.

low applied loads. Resistivity measurements also display two major differences between binary and Ni-Ti-Cu alloys. Firstly, the resistivity of the martensite phase is higher than that of the austenite for the ternary alloy, the reverse being true for the binary. Secondly, resistivity measurement on the binary alloy can exhibit an

Table III: Yield Strengths of Binary and Ternary Alloys

	<u>Ni-Ti</u>	<u>Ni-Ti (10%Cu)</u>
Martensite (tested @25°C)	208 MPa	106 MPa
Austenite (tested @200°C)	1053 MPa	1177 MPa

intermediate peak indicative of the R phase¹⁸. No such peak is observed for the ternary alloys containing greater than 5% Cu¹⁵. In addition, no other property changes provide evidence of "locking-in" the R-phase^{15,19} for the Cu-containing ternary alloys.

3. Conclusions

The substitution of Cu for Ni in the Ni-Ti SMA results in minor structural changes and numerous modifications of the shape memory properties. However, the basic shape memory effect is observed in alloys containing up to around 30% Cu. Some advantages of adding Cu are a narrower transformation hysteresis and a more constant M_s temperature less dependent on concentration variations and cyclic fatigue. Furthermore, the Cu-containing ternary alloy also has a martensite phase with a lower yield strength, thereby requiring a lower reset force for some cyclic applications. Alloys with 10% Cu exhibit a two step transformation, and a corresponding two step shape change.

References:

1. D.M. Goldstein, et al.: Naval Ordnance Lab. Technical Report # 64-235, (1965).
2. T. Honma, et al.: Proc. International Conference on Martensitic Transformation ICOMAT '79, Cambridge, Mass. 259 (1979).
3. F.J.J. Van Loo: Journal of the Less Common Metals **57**, 111 (1978).
4. M. Hansen: Constitution of Binary Alloys, 2nd ed., McGraw-Hill, N.Y., 643 (1958).
5. S.P. Alisova, et al.: Izvestiya Akademii Nauk SSSR. Metally, No. 5, 210 (1986).
6. Y. Shugo, F. Hasegawa, and T. Honma: Japan Inst. Metals Proc., Tokyo, (1980).
7. T. Tadaki and C.M. Wayman: Metallography **15**, 233 (1982).
8. W.J. Moberly: PhD Thesis, Mechanical Twinning and Twinless Martensite in Ternary TiNi Alloys, Stanford University, (1989).
9. R. Sinclair: AIP Conference Proc. #53, J.M. Cowley et al., eds., 269 (1979).
10. W.J. Moberly, J.L. Proft and T.W. Duerig: ICOMAT '89 Proc., Sydney, (1989).
11. R.H. Bricknell, K.N. Melton and O. Mercier: Met. Trans. **10A**, 693 (1979).
12. V. Yerofeyev, V. A. Monasevich and Yu I. Pavskaya: Phys. Metals **4(1)**, 86 (1982).
13. Y. Shugo and T. Honma: JIM Proceedings, Fukuoka, Japan (1980).
14. O. Mercier and K.N. Melton: Met. Trans. **10A**, 387 (1979).
15. O. Mercier et al.: Proc. Int. Conf. on Solid-Solid Phase Transformations, H.L. Aaronson et al. eds., 1259 (1982).
16. J.L. Proft, K.N. Melton and T.W. Duerig: Proc. of the Inter. MRS Conf., Tokyo, Japan (1987).
17. V.N. Tokarev, A.S. Savvinov and V.N. Khachin: Physics of Metals and Metallurgy **56**, 117 (1983).
18. D.P. Dautovich and G.R. Purdy: Canadian Metallurgical Quart. **4(2)**, 129, (1965).
19. O. Mercier and E. Török: Proc. ICOMAT '82 in J. De Physique 43-C4, 267 (1982).
20. G.M. Michal and R. Sinclair: Acta Crystal B37, 1803 (1981).

Shape Memory and Transformation Behavior of Martensitic Ti-Pd-Ni and Ti-Pt-Ni Alloys

P.G. Lindquist and C.M. Wayman

Department of Materials Science and Engineering
University of Illinois at Urbana-Champaign

Metals substituted for Ni in Ni-Ti alloys affect the parent to martensite transformation temperatures by markedly differing amounts. Substantial Pd and Pt additions strongly elevate the transformation temperatures compared to other elements; but small additions of these shift M_s to below 0°C. The M_s temperature ranges from -26°C (for 10 at.% Pd) to 563°C (for 50 at.% Pd) and from -10°C (10 at.% Pt) to 1040°C (for 50 at.% Pt).

The transformation behavior after deformation of the martensite was investigated by incrementally deforming Ti-Pd-Ni and Ti-Pt-Ni alloys at room temperature and monitoring the shape recovery in a high temperature dilatometer. This confirmed the one-way shape memory effect. The 30 at.% Pd alloys could be deformed up to 4.5% while still observing a 100% shape recovery. For other compositions, the shape memory effect diminished when the alloys were deformed more than 2.5%.

High temperature X-ray diffraction of Ti-Au, Ti-Pd and Ti-Pt alloys indicates that a B2 (CsCl) to B19 (orthorhombic) martensitic transformation occurs¹. Khachin² showed that ternary alloys of Ti-Pd-Ni can be treated as pseudo-binary alloys and that substitutions of Pd for Ni shifts M_s from 50°C for the Ni-Ti binary up to 540°C for Ti-Pd. In light of the small thermal hysteresis, some have viewed these transformations as thermoelastic^{3,4}. However more information is needed to confirm the transformation behavior. This report focuses on the transformation behavior of Ti-Pd-Ni and Ti-Pt-Ni ternary alloys. First, the compositional dependence of the transformation temperatures was determined by electrical resistivity and dilatometry. Then, the transformation behavior after deformation was investigated via mechanical testing. In this paper we discuss the compositional dependence of the transformation temperatures, which compositions behave thermoelastically, and the limits of the shape memory effect.

1. Experimental

$Ti_{50}-(Ni_{(50-x)}Pd_x)$ and $Ti_{50}-(Ni_{(50-x)}Pt_x)$ alloys of composition x=5, 10, 20, 30, 40, 45 and 50 at.% were prepared by tungsten arc melting, homogenizing for 24 hours at 1000°C and hot rolling between 800 and 1000°C. Higher temperatures, intermediate annealing and smaller amounts of reduction were necessary for alloys rich in Pd or Pt. All compositions containing Pd and those containing less than 30 at.% Pt were successfully fabricated. Alloys with higher Pt content were brittle and could not be rolled. Plates could be rolled to a final thickness of 1mm, vacuum annealed at 800°C

and finally water quenched. Samples for electrical resistivity, dilatometry and mechanical testing were cut into rods about 1x1x60mm.

Electrical resistivity was measured while heating and cooling at 5°C/min in vacuum using the four-probe method and a constant current of 0.500 A. The resistivity measurements were computer controlled for automated testing and data acquisition⁵. A computer controlled dilatometer was developed with components similar to the resistivity system. For mechanical testing, specimens from each composition were cycled in the dilatometer 2 to 3 times from room temperature to 600°C. Initially each sample was deformed at room temperature in an Instron machine, then loaded and heated to temperatures above A_f at a heating rate of 5°C/min. After cooling a second dilatometer run was made to determine whether the mechanical deformation affected the transformation behavior. The sample was then deformed to the next higher strain increment and cycled through the dilatometer again. The strain increments used were: 0.5, 1.0, 2.0, 3.0, 3.5, 4.5 and 6.0%, as measured by a 13mm extensometer. Only samples with transformation temperatures above room temperature were tested: Pd = 20, 30, 40, 45 and 50, Pt = 20, 30 at.%. The sample containing 30 at.% Pt broke at strains less than 0.5% and was omitted.

2. Transformation Behavior

The temperature dependence of the electrical resistivity of Ti-Pd-Ni alloys is summarized in Table 1. Alloys containing less than 10 at.% Pd transform from the

Table 1. Transformation temperatures and electrical resistivity data for Ti-Pd-Ni alloys.

Composition Atomic %	Transformation Temperature, °C				$d\rho/dT$ ($\times 10^{-15} \Omega m/\text{°C}$)	
	A_s	A_f	M_s	M_f	Parent	Martensite
0	80	89	55	15	0.50	1.59
5	45	50	26	12	0.30	1.47
10	-26	-7	-18	-39	0.20	1.30
20	90	107	95	65	NA	1.23
30	230	241	241	208	0.29	1.03
40	419	427	403	387	0.30	0.92
45	503	509	486	467	0.31	1.00
50	580	591	563	550	0.35	1.04

parent phase (B2) to the R-phase to martensite as the temperature is decreased. The presence of the R-phase is supported by a marked increase in the resistivity at temperatures slightly above M_s , Figure 1(a). This transformation sequence agrees with low temperature X-ray diffraction results.⁶ Comparing present results to the Ni-Ti-Fe system it appears that the temperature range over which the R-phase is stable is much less. The R-phase transition is most evident for alloys containing 5 at.% Pd. For compositions between 5 and 30 at.% Pd the resistivity increases only slightly with decreasing temperature prior to the martensitic transformation. The resistivity profiles for compositions undergoing high temperature martensitic transformations are square hysteresis loops with well defined transformation temperatures, Figure 1(b). The change in slope of the resistivity curve between A_s and A_f , indicates a slight impediment in the transformation from martensite to austenite for alloys of high Pd or Pt content. All compositions studied undergo a martensitic transformation.

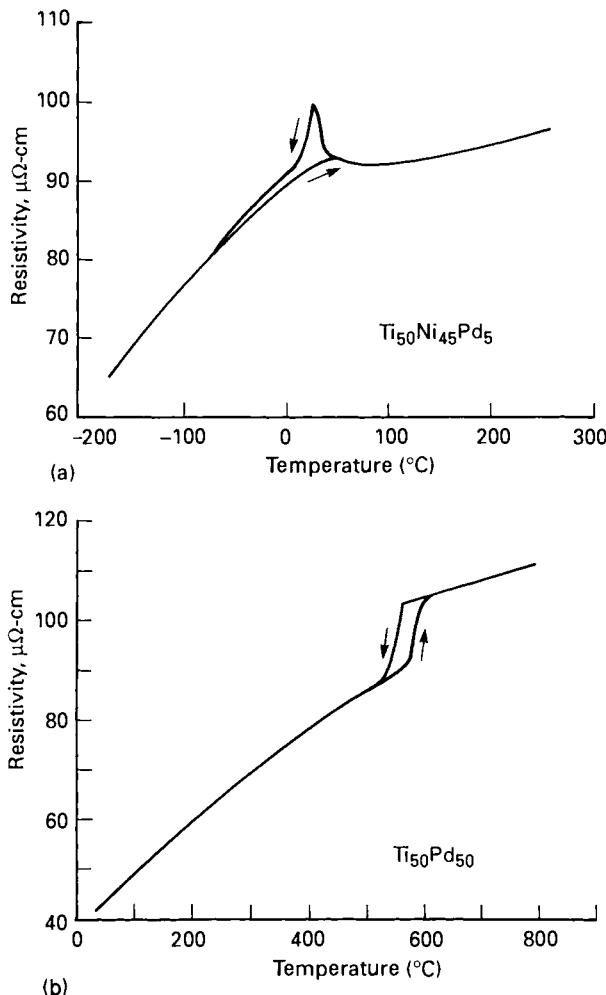
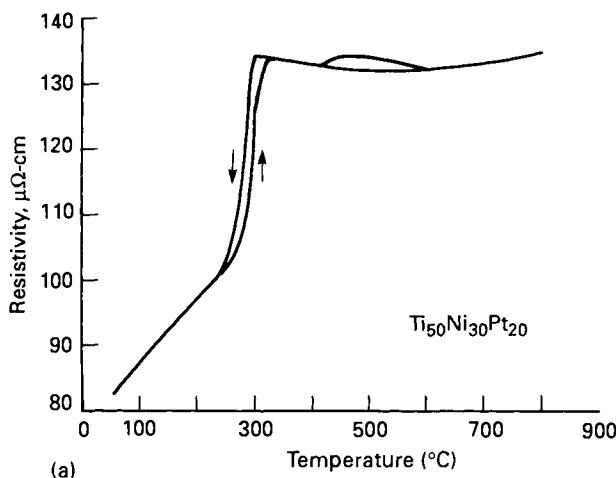


Figure 1. Electrical resistivity vs. temperature plots for Ti-Pd-Ni alloys: (a) 5 at.% Pd and (b) 50 at.% Pd. Note the marked increase in the electrical resistivity in (a) due to the R phase transition.

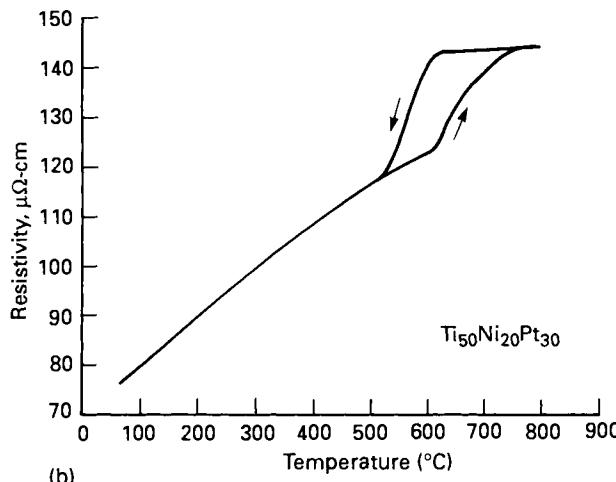
The temperature dependence of the electrical resistivity for Ti-Pt-Ni alloys is summarized in Table 2. For alloys with low transformation temperatures, 5 and 10 at.% Pt, the transformation behavior was similar to that of the 5 and 10 at.% Pd alloys. The high temperature transformation behavior of the 20 and 30 at.% Pt alloys is shown in Figure 2(a) and (b). These results are also similar to the Ti-Pd-Ni alloys, but there are two differences. For 20 at.% Pt, there is an anomalous increase in resistivity while cooling between 600 and 400 $^{\circ}\text{C}$. Upon heating this anomaly does not follow the cooling curve. If the sample is cycled between 400 and 600 $^{\circ}\text{C}$ the loop remains. Based on results from the dilatometer where no anomalies were observed, it is concluded that there are no first order transitions occurring. No confirming high temperature X-ray diffraction or transmission electron microscopy was done. A second

Table 2. Transformation temperatures and electrical resistivity data for Ti-Pt-Ni alloys

Composition Atomic %	Transformation Temperatures, °C				$d\rho/dT$ ($\times 10^{-15}$ $\Omega m/^\circ C$)	
	A_s	A_f	M_s	M_f	Parent	Martensite
0	80	89	55	15	0.50	1.59
5	36	49	29	10	0.24	1.44
10	-27	36	18	-8	NA	1.23
20	263	300	300	241	0.12	0.96
30	626	702	619	537	0.10	1.00



(a)



(b)

Figure 2. Electrical resistivity vs. temperature plots for Ti-Pt-Ni alloys: (a) 20 at.% Pt and (b) 30 at.% Pt.

difference is that the martensitic to austenite transformation begins gradually and occurs over a greater temperature range for the 30 at.% Pt alloy than observed for the other alloys.

Comparing the results for Ti-Pd-Ni and Ti-Pt-Ni alloys it is clear that Pt is more effective than Pd in increasing the transformation temperature. Figure 3 summarizes the

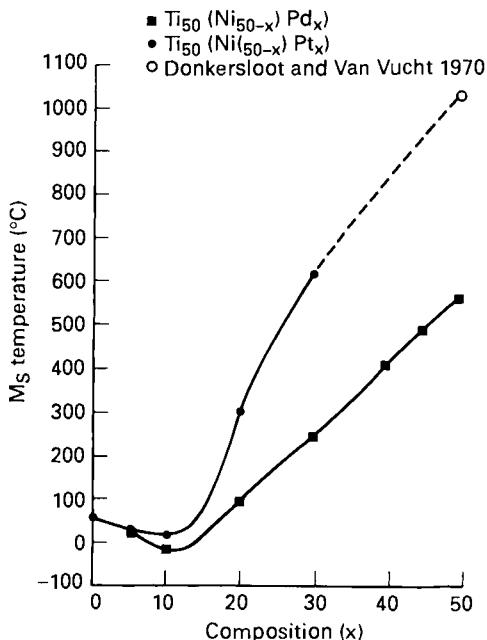


Figure 3. Plot of M_s temperatures as a function of composition for Ti-Pd-Ni and Ti-Pt-Ni alloys.

compositional dependence of the M_s temperature for Pd and Pt alloys. The compositional dependence of M_s for the Ti-Pd-Ni alloys is comparable to that found by Khachin et al.⁸.

A high temperature dilatometer was used as an alternate method to determine the transformation temperatures for samples with M_f temperatures above room temperature (Pd or Pt > 20 at.%). There is good agreement between the resistivity results and the volume change associated with the martensitic transformation⁵. Assuming that the volume change is ~3 times the length change, the volume change upon transformation is < 0.2% for 20 at.% Pd and increases to 0.75% for 50 at.% Pd. The volume changes for both 20 and 30 at.% Pt alloys were 0.65%.

3. Transformation Behavior after Deformation

Qualitative tests to determine whether the alloys exhibited shape memory behavior were performed on all alloys. First the samples were bent at room temperature and then heated. For all specimens the one way shape memory effect was observed.

Deformation and recovery experiments were devised for a more quantitative analysis of the shape memory. These experiments determined how mechanical deformation of the martensite affects the transformation behavior.

The shape recovery after deformation was measured in a dilatometer. Figure 4 shows the first two thermal cycles as a function of temperature. The 40 at.% Pd sample was deformed 1% over its total length (1.7% strain between the gage length of the

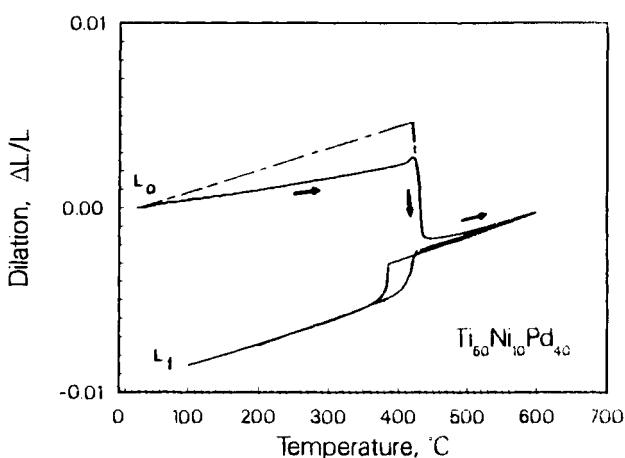


Figure 4. Plots of the dilation vs. temperature for the first two heating and cooling cycles after deformation. The composition of this specimen was 40 at.% Pd and it was deformed 1.70% over the gage length. Upon heating the specimen begins to recover. The major portion of the shape recovery begins and ends at the A_s and A_f temperatures, respectively. Transformation behavior observed in the second heating and cooling cycle is similar to that of undeformed alloys.

extensometer). During the first heating cycle a continuous shape recovery began at room temperature and continued up to the A_s temperature; this represents a 0.25% strain recovery. The amount of recovery is estimated from the differences in the slope (thermal expansion coefficient) of the deformed and unstrained martensite expansion curves, Figure 4. If the sample did not recover one would expect these slopes to be essentially the same, and the heating curve would follow the dashed line instead of the solid line. The thermal expansion coefficient, α , of the deformed martensite is $6 \times 10^{-6}/^\circ\text{C}$, not $\sim 13 \times 10^{-6}/^\circ\text{C}$ as expected. A possible explanation for this behavior is that the martensite twins are being rearranged, although no confirming microscopic observations were attempted.

The major portion of the shape recovery occurs between A_s and A_f . In this temperature range the dilation is a combination of two effects: an increased length from the volume increase of the martensitic transformation, marked by a slight upturn in the curve at A_s and a decrease in length from the one way shape memory effect, marked by a large decrease which begins slightly above A_s and finishes near A_f . Above A_f there is a slight decrease in length.

The cooling curves for the first and second thermal cycles are identical, and the second heating curve is comparable to the heating curve of an undeformed sample. Comparing the cooling curves and the second heating curve to an undeformed sample, no differences in the transformation temperatures or the volume changes are observed. Generally there was little or no shift in transformation temperatures when compared to alloys that were not deformed, regardless of the amount of deformation. The amount of deformation affected only the shape recovery. Some alloys recovered nearly 100% and others recovered only a fraction of the applied strain.

For specimens strained between 2 and 6%⁵ the dilatometer profiles had the following common features:

1. The thermal expansion coefficients of the deformed martensite were less than those observed on the 2nd heating cycle or for undeformed martensite.
2. The shape recovery between A_s and A_f appears to proceed in two steps. At the A_s temperature there is a marked decrease in sample length as temperature is increased. The slope of the thermal expansion curve is constant up to A_f . Near A_f there is a second inflection of the thermal expansion curve and the remaining shape recovery occurs within a few degrees of this region.
3. Between A_f and 600°C a small amount of shape recovery is observed.

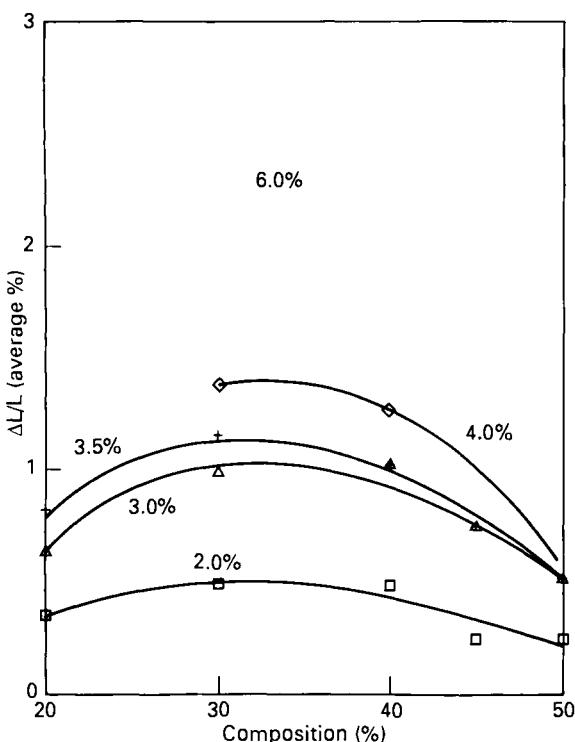


Figure 5. Plots of the average shape recovery as a function of composition and also the strain increments for Ti-Pd-Ni alloys. For each specimen the average shape recovery was determined by the differences between the initial and final length of the specimens.

Explanations for the 1st and 3rd observations above were discussed previously. A possible explanation of the 2nd observation is that the shape recovery is masked by the portion of the sample being constrained by the grips and the difference in transformation behavior of low strain vs. high strain martensite. At low strains, the shape recovery begins slightly below A_s . At increased strain increments the results indicate that the temperature at which the shape recovery begins is above A_s ; this temperature is a function of strain⁵. Due to experimental constraints there is always a strain gradient in the sample. At 2% strain the amount of high strain martensite is low and the shape recovery is observed to begin near A_s . Conversely, at 6% strain the amount of high strain martensite is high and the shape recovery temperature is elevated to some 10 to 15°C above the A_s . For highly strained martensite the shape recovery begins abruptly and completes the transformation to the parent phase within a few degrees. The shape recovery of low strain martensite is observed to proceed continuously from the A_s to the A_f temperature.

Analyzing sample lengths before and after deformation shows that those containing Pd or Pt undergo complete shape recovery for strain increments investigated, provided that the composition is less than 45 at.% Pd. For 45 and 50 at.% Pd recovery is incomplete beyond 1% strain. Figure 5 shows the percent of the average strain recovered vs. composition. The strain increments were 2.0, 3.0, 3.5, 4.0 and 6.0%.

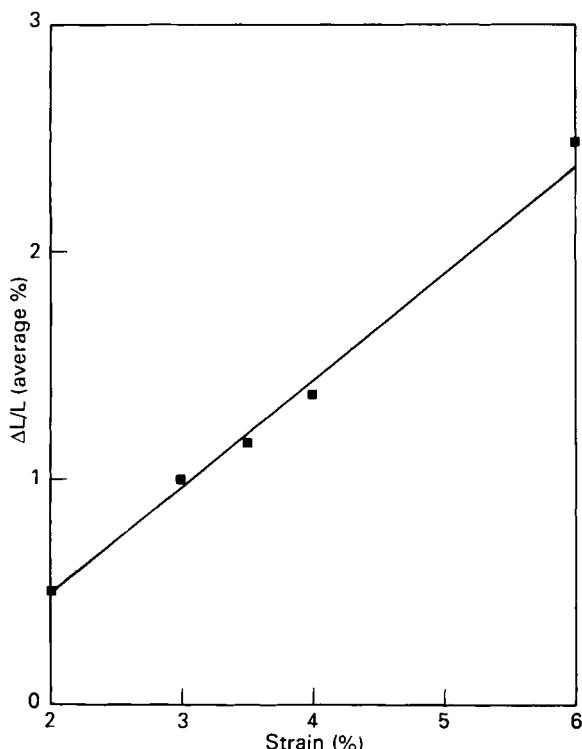


Figure 6. Plot of the average shape recovery vs. applied strain for an alloy containing 30 at.% Pd.

From the figure it is apparent that the average strain recovered is much less than the strain measured across the gage length of the extensometer. The average strain is determined from the total sample length, but in a 60mm sample only about 20 to 25mm is strained uniformly. The ends of the sample are gripped by the tensile grips and are not uniformly strained. Figure 5 shows that for alloys containing more than 40 at.% Pd the shape recovery saturates at about 0.5%. For other compositions, the shape recovery is typically near 99.5% of the total deformation applied. Figure 6 is a plot of average strain recovered vs. applied strain for an alloy containing 30 at.% Pd. Samples of this composition exhibited nearly perfect shape recovery up to an applied strain of 6.0%. All other Pd and Pt containing alloys fractured at strains > 3.5%. No experiments were undertaken to determine the source of this type of failure. The maximum strain each sample could withstand varied from sample to sample. The profile of the stress-strain curve is comparable to that of Ni-Ti alloys deformed at

Table 3. Stress, strain and shape recovery after martensite deformation for Ti-Pd-Ni and Ti-Pt-Ni alloys.

At. %	Element	σ (MPa)	ϵ (%)	ϵ_r (%)	$\Delta L/L$ (%)
20	Pd	266	2.0	0.89	0.35
30	Pd	351	2.0	1.00	0.50
40	Pd	418	2.0	0.92	0.50
45	Pd	475	2.0	0.83	0.27
50	Pd	507	2.0	0.66	0.27
20	Pt	378	2.0	1.06	0.48
30	Pt	XX	2.0	XX	XX
20	Pd	343	3.0	1.69	0.64
30	Pd	454	3.0	1.88	1.00
40	Pd	516	3.0	1.79	1.04
45	Pd	578	3.0	++	0.77
50	Pd	587	3.0	++	0.55
20	Pt	525	3.0	1.74	0.94
20	Pd	378	3.5	1.84	0.81
30	Pd	498	3.5	2.27	1.16
40	Pd	534	3.5	2.20	1.04
45	Pd	578	3.5	2.11	0.69
50	Pd	605	3.5	1.90	0.34
20	Pt	534	3.5	2.17	1.00
20	Pd	XX	4.0	XX	XX
30	Pd	547	4.0	2.70	1.38
40	Pd	583	4.0	2.66	1.29
45	Pd	XX	4.0	XX	XX
50	Pd	614	4.0	2.39	0.54
20	Pt	XX	4.0	XX	XX
30	Pd	792	6.0	4.38	2.50
40	Pd	XX	6.0	XX	XX
50	Pd	--	--	--	--

XX Sample broke

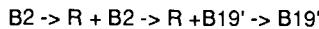
++ Could not attach extensometer

-- Omitted: did not fully recover at lower strain

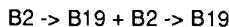
temperatures below M_f^0 . There is a linear elastic region, followed by a non-linear stress-strain relationship, and then the strain varies linearly as the stress is increased. The strain range of each region was composition dependent. The stress at 3% stain varied with composition: from 343 MPa for 20 at.% Pd to 587 MPa for the 50 at.% Pd alloy. The stress-strain curve of the 20 at.% Pt sample was similar to that of the Pd alloys. The only difference was that the stress at 3% strain was 525 MPa. The 30 at.% Pt sample broke at less than 0.5% strain because of micro-cracks from hot rolling. During unloading the stress-strain curve was non-linear and ~1.3% of the 3.0% applied strain recovered. Table 3 lists the stress, strain, permanent set, and recovery strain after thermal cycling.

4. Summary and Conclusions

The transformation and deformation behavior of Ti-Pd-Ni and Ti-Pt-Ni alloys has been investigated by electrical resistivity, dilatometry and controlled deformation and recovery methods. Prior to deformation the transformation temperatures were found to decrease when the amount of Pd or Pt was increased from 0 to 10 at.%. For additions of 10 at.% the M_s went through a minimum in both alloy systems. Further additions of these alloying elements increase the transformation temperature linearly with composition up to a maximum for the Ti-Pd and Ti-Pt binary alloys. For compositions containing less than 30 at.% Pd or Pt the transformation upon cooling from the parent phase was:



This is the same behavior as found in binary and certain ternary TiNi alloys. Likewise, upon cooling, the transformation sequence for compositions containing greater than 20 at.% Pd or Pt was:



This transformation sequence differs from the above because now the martensite is orthorhombic (B19) and not monoclinic(B19'). In other words, initial additions of Pd or Pt to TiNi depress the characteristic B19' martensitic temperatures (M_s , M_f , A_s and A_f), but once a certain enrichment is reached (~30at.% Pd or Pt) these temperatures begin rising again. But now the martensite formed is no longer B19', but instead it becomes B19. One may make the parallel statement that the addition of Ni to TiPd or TiPt causes a depression of the characteristic temperatures associated with the B2->B19 martensitic transformations, and that additions of Pd or Pt to TiNi similarly lower the characteristic B2->B19 martensitic transformations. Since the B19 and B19' martensites have intrinsically different crystal structures, identical shape memory behavior is not expected.

The shape recovery of both alloy systems deformed at temperatures below M_f begins once the sample is heated above room temperature. From room temperature up to A_s there is a measurable sample contraction, confirmed by the difference in the thermal expansion coefficient, α , between deformed and undeformed martensites. Above A_f the specimen contracts slightly; however the major shape recovery occurs between A_s and A_f as with other shape memory alloys. The maximum amount of recoverable strain varied from sample to sample. For alloys containing 30 at.% Pd, nearly 100% of the 6% applied strain was recovered upon heating to the parent phase. For alloys containing more than 45 at.% Pd, 100% recovery was limited to deformations of less

than 1%. For the other compositions the maximum elongation and the shape recovery was less than the 6% observed in the 30 at.% Pd alloy; for the 20 and 40 at.% Pd alloys it ranged between 3 and 3.5%. Based on the results of combined electrical resistivity, dilatometry and deformation and recovery experiments it is concluded that the martensitic transformations in Ti-Pd-Ni and Ti-Pt-Ni alloys are thermoelastic in nature over the composition and strain ranges studied.

Acknowledgement

This work was supported by the National Science Foundation through the Materials Research Laboratory at the University of Illinois, Grant NSF/DMR 86-12860.

References

1. Donkersloot, H.C. and Van Vucht, J.H.N., J. Less-Common Metals, **20**, 83, 1970.
2. Khachin, V.N., Grunter, V.E., Sivokha, V.P. and Savvinov, A.S., in Proc. Intl. Conf. on Martensitic Transformations (ICOMAT-79), Cambridge, MA, USA, 474, 1979.
3. Wu, S. K., Wayman and C.M., Metallography, **20**, 359, 1987.
4. Enami, K., Yoshida, T. and Nenno, S., in Proc. Int. Conf. on Martensitic Transformations (ICOMAT-86), 103, 1986.
5. Lindquist, P.G., "Structure and Transformation Behavior of Martensitic Ti-(Ni,Pd) and Ti-(Ni,Pt) Alloys", Ph.D Thesis, University of Illinois, 1988.
6. Sivokha, V.P., Savvinov, A.S., Voronin, V.P. and Khachin, V.N., Phys. Met. Metall., **56**, No. 112, 1983.
7. Hwang, C.M., Meichle, M., Salamon, M.B. and Wayman, C.M., Phil. Mag., **47**, 9, 1983.
8. Khachin, V.N., Matveeva, N.M., Sivokha, V.P., Chernov, D.B. and Kouneristy, Yu K., Doklady Akademii Nauk SSSR, **257**, 167, 1981.
9. Miyazaki, S. and Otsuka, K., Met. Trans. A, **17A**, 53, 1986.

Cu-Based Shape Memory Alloys

Ming H. Wu
Memory Metals, Inc., Norwalk, Connecticut

Both Ni-Ti and Cu-based shape memory alloys, such as Cu-Zn-Al and Cu-Al-Ni, are presently available for commercial shape memory applications. Ni-Ti alloys generally exhibit better shape memory properties and corrosion resistance than do Cu-based alloys, however Ni-Ti in finished form is very expensive because of fabrication difficulties associated with melting and forming. In many applications, Cu-based alloys provide a more economical alternative to Ni-Ti. Although early Cu-based alloys suffered from intergranular failure due to their intrinsic coarse grain structure, the recent development of fine-grain Cu-based alloys has improved mechanical properties significantly. In many cases today, Cu-based memory alloys are more competitive than Ni-Ti when comparing to bimetals and other low cost materials commonly used in conventional thermal-mechanical actuation mechanisms.

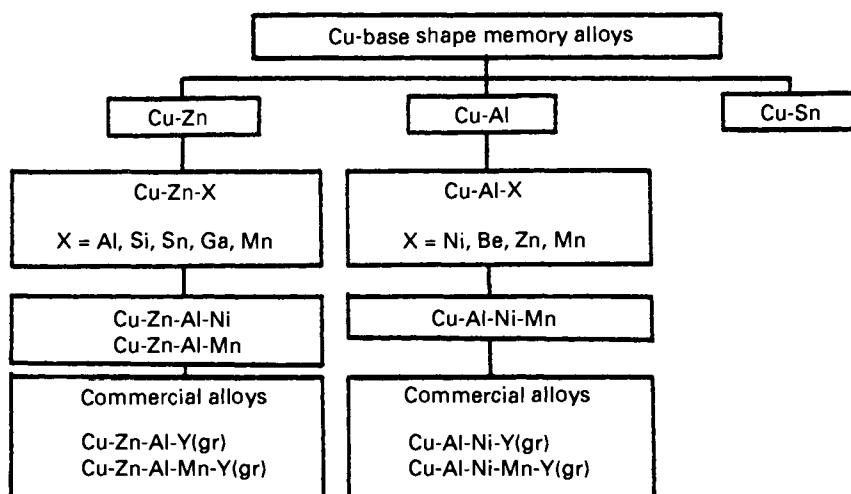
1. Cu-Based Shape Memory Alloy Systems

Current Cu-based shape memory alloys are derived from three binary alloy systems: Cu-Zn, Cu-Al and Cu-Sn. The martensitic transformation in Cu-Sn alloys is not ideally thermoelastic¹ and suffers from a rapid degradation of shape memory properties during aging at even moderate temperatures.^{1,2} These alloys therefore have been more of a laboratory interest in aging studies than as potential commercial shape memory alloys.

As listed in Figure 1, Cu-Zn-base ternary alloys containing Al³, Si^{4,5}, Sn⁶, Ga⁷, or Mn⁸ and Cu-Al-base ternary alloys with alloying elements of Ni^{9,10}, Be¹¹, Zn¹² or Mn¹³ have been explored for their shape memory properties. The studies have been extended to quaternary alloys of Cu-Zn-Al-Ni¹⁴, Cu-Zn-Al-Mn¹⁵ and Cu-Al-Ni-Mn.¹⁶ The commercially available alloys have been developed from among these, either in the ternary form of Cu-Zn-Al or Cu-Al-Ni, or in the quaternary modifications containing Mn. The commercial Cu-base shape memory alloys are also grain-refined via additives such as B, Ce, Co, Fe, Ti, V and Zr.¹⁷⁻²¹

2. Effects of Alloy Elements

The shape memory properties of Cu-base shape memory alloys are quite sensitive to alloying elements which are added to adjust the martensitic transformation



$Y(gr)$ = Grain refining elements; B, Ce, Co, Fe, Ti, V, Zr.

Figure 1. The classification of Cu-base shape memory alloy systems and their alloying elements.

temperatures and to optimize thermal stability as well as mechanical properties. The martensitic transformations in Cu-base shape memory alloys occur from an ordered b.c.c. β phase. The high temperature β phase has a disordered A2 structure, but upon cooling the structure goes through a nearest-neighbor ordering transition and develops the B2 superlattice structure. Further cooling induces next-nearest-neighbor ordering and the structure eventually becomes the D_0_3 or L_2_1 superlattice structure (depending on alloy composition and cooling rate).²²⁻²³

The structure of martensites in Cu-base shape memory alloys are far more complex than are those of the Ni-Ti system. All are of a Long Period Stacking Order (LPSO) type: 3R, 9R (or 18R depending on the order state) and 2H which correspond to α' , β' and γ' martensite respectively.²⁴ The thermally-induced martensites in both Cu-Zn-Al and Cu-Al-Ni alloys consist predominantly of β' martensites. γ' martensites are usually found in alloys with higher Al content. Mixtures of both β' and γ' martensites have been observed in Cu-21.00% Zn-5.90% Al²⁵ and Cu-13.03% Al-4.09% Ni²⁶. (All compositions in this paper will be given in weight percent). In both systems, the stability of γ' martensite increases with Al content.²⁵⁻²⁷ The stress-induced martensites from the austenite or parent β normally have the same structure as thermally-induced martensites. α' and β' martensites can also be induced upon stressing the γ' martensite.²⁸⁻³⁰ The stability of each martensite depends on both stress and temperature.³⁰

2.1 Martensitic Transformation

The martensitic transformation temperature is also very sensitive to small variations in alloy composition. To a first approximation, the transformation temperature follows a

linear relationship with alloying element content. Several empirical relationships are useful in obtaining estimated transformation temperatures:

Cu-Zn-Al

$$A_s(^{\circ}\text{C}) = 2117 - 58.79 (\text{wt.\%Zn}) - 149.64(\text{wt.\%Al})^{31} \quad (1)$$

$$M_s(^{\circ}\text{C}) = 2212 - 66.9[1.355(\text{at.\%Al}) + 1(\text{at.\%Zn})]^{32} \quad (2)$$

Cu-Al-Ni

$$M_s(^{\circ}\text{C}) = 2020 - 45(\text{wt.\%Ni}) - 134(\text{wt.\%Al})^{16} \quad (3)$$

Although the transformation temperatures of both Cu-Zn-Al and Cu-Al-Ni alloys can be manipulated over a wide range, the practical upper limits are 120°C and 200°C respectively; above these temperatures the transformations tend to be unstable due to rapid aging effects.

2.2 The Equilibrium Phases

Figure 2 shows a 6 wt.% Al section of a pseudo-binary phase diagram of Cu-Zn-Al³³ which also incorporates a plot of the reverse transformation temperature, A_s , as a

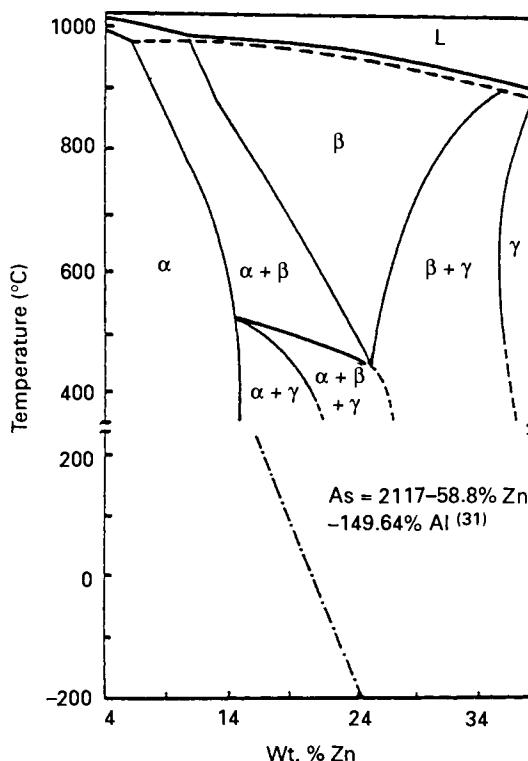


Figure 2. Cu-Zn-6wt.%Al pseudo-binary phase diagram³³ and a plot of the A_s temperature as a function of the Zn content.

function of Zn content.³¹ The composition of Cu-Zn-Al shape memory alloys with practical transformation temperatures falls in the hypoeutectoid range which contains both the α and β equilibrium phases at intermediate temperatures and, in addition, γ equilibrium phase at lower temperatures. Low Al content, 4% for example, Cu-Zn-Al alloys of $\alpha+\beta$ microstructure exhibit good low temperature workability, where cold reductions of more than 20% can be obtained. These alloys can be cold finished with intermediate annealing between each deformation pass.³³ Increasing Al content causes the composition of Cu-Zn-Al alloys to approach the eutectoid and eventually fall into the hypereutectoid range.³³ Consequently, the amount of ductile α phase gradually decreases and is replaced by the brittle γ phase in the equilibrium microstructure. Cold workability therefore deteriorates. A similar phase diagram of Cu-Al-3%Ni³⁴ is shown in Figure 3 together with a plot of the M_s as a function of Al content.¹⁶ In contrast to Cu-Zn-Al, Cu-Al-Ni alloys with practical transformation temperatures are mostly eutectoid or hypereutectoid alloys,

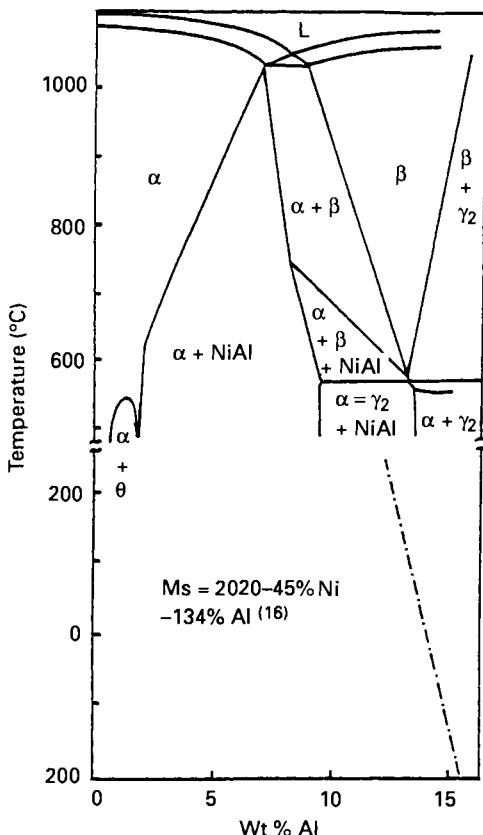


Figure 3. Cu-Al-3wt.%Ni pseudo-binary phase diagram³⁴ and a plot of the M_s temperature as a function of the Al content.

and exhibit low ductility at low temperatures due to the presence of γ phase in the microstructure. They can only be hot worked¹⁶, usually in the β phase region.

The presence of Ni in Cu-Al-Ni alloys shifts the Cu-Al eutectoid to higher Al content³⁵ and retards the hypereutectoid decomposition.²³ Ni contents up to 4 % also help to reduce the transformation hysteresis. However, in order to avoid deterioration in hot workability, the Ni content should be kept below 7%.³⁶

Commercial quaternary Cu-base shape memory alloys also contain Mn. Similar to the effects of Ni, Mn also shifts the Cu-Al eutectoid to higher Al content.¹⁵ In Cu-Zn-Al alloys, Mn allows a higher Al content while maintaining good workability.¹⁵ Similarly in Cu-Al-Ni alloys, ductility can be improved by substituting Al with Mn so that a near-eutectoid alloy is obtained rather than a hypereutectoid alloy.¹⁶

2.3 Effects of Grain-Refining Elements

Because of rapid grain growth, heat treated Cu-base shape memory alloys usually possess coarse grains with sizes in the mm range. Coarse grained Cu-base alloys are brittle and prone to premature intergranular failure.^{21,37} They also exhibit shorter thermomechanical cycling life due to rapid degradation in shape memory performance.¹⁷

Among the grain-refining elements listed in Figure 1, B, Ti and Zr are very effective in refining the grain structure of Cu-Zn-Al alloys. As-cast grain sizes less than 50 μm can be easily obtained by an addition of 0.3-1.3% Zr to Cu-Zn-Al alloys; these additions effectively suppress subsequent grain growth during betatizing heat treatments.¹⁸ Additions of 0.04% B in Cu-Zn-Al and 0.1% B in Cu-Al-Ni have remarkable effects in refining the as-solidified microstructure¹⁹, however relatively fast grain growth during subsequent heat treatment tends to reduce overall effectiveness. Ti is another effective grain refining element: Cu-Zn-Al alloys containing 0.2-0.8%Ti usually possess grains of 50-100 μm in size.¹⁸ In Cu-Al-Ni alloys, 0.5% Ti effectively limits grain sizes to 100 μm even after prolonged betatizing at 800°C.²¹

Grain refining elements usually have very limited solubility in Cu-base shape memory alloys. Refining is largely due to the formation of small insoluble particles which either assist grain nucleation or inhibit grain growth.¹⁹ These elements therefore have only very slight effects on transformation temperature.

3. Thermal Effects on the Martensitic Transformation

Because of the metastable nature of both the parent β phase and the martensite in Cu-base shape memory alloys, the stability of their shape memory properties is influenced strongly by aging. Specifically, the transformation temperatures M_s , M_f , A_s and A_f , are affected by heat treatment and subsequent thermal exposure during service.

3.1. Betatizing Heat Treatment

The effect of betatizing heat treatment parameters on the transformation temperatures of Cu-base shape memory alloys can be seen in the example in Figure 4. Here, M_s , M_f , A_s and A_f temperatures of a Cu-20.6%Zn-5.7%Al-0.1%Zr alloy after betatizing at various temperatures for 10 and 50 minutes are plotted versus betatizing temperature. All transformation temperatures are increased with increasing betatizing temperature.

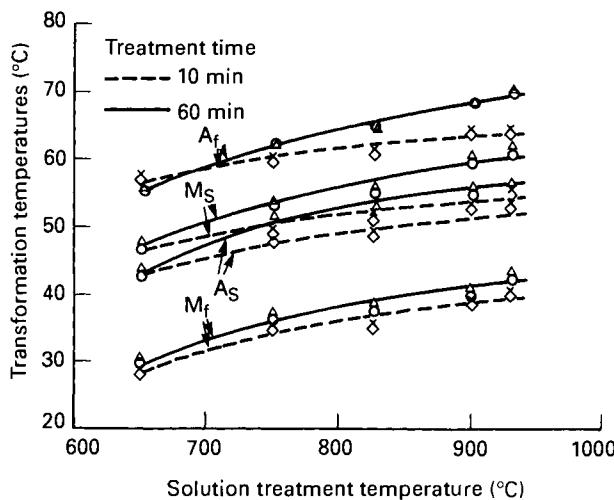


Figure 4. Transformation temperatures, M_s , M_f , A_s and A_f , of a Cu-20.6Zn-0.1Zr alloy after betatizing heat treatment at temperatures of 650-950°C for 10 and 60 minutes respectively.

Extending the betatizing time from 10 to 60 minutes also raises transformation temperatures. This effect is in fact related to grain growth during the betatizing heat treatment. The change in transformation temperature correlates well with the change in grain size and follows a Hall-Petch relationship.³¹

3.2 Quenching Rate

To obtain a shape memory effect in Cu-base alloys, the metastable β phase must be retained, meaning that a sufficiently rapid cooling from the betatizing temperature is necessary to avoid decomposition of the β phase into the equilibrium phases. For Cu-base shape memory alloys, the transformation temperatures are very sensitive to this quenching rate. For example, rapidly quenching a Cu-Zn-Al alloy into the parent β phase results in depressed and unstable initial transformation temperatures, while directly quenching into the martensitic state causes stabilized martensites with higher than nominal transformation temperatures. These effects will be discussed in more detail in section 3.3 and in the next chapter. Stable transformation temperatures can be obtained by a post-quench aging treatment or by step-quenching, however they are still very sensitive to the quenching rate.

The microstructures and the transformation temperatures of a Cu-26.0Zn-3.8Al alloy and a Cu-12.7Al-5Ni-2Mn alloy have been studied after betatizing at 800°C for 30 minutes and subsequently performing a Jominy end quench.³⁸ After the Jominy quench, the Cu-Zn-Al and Cu-Al-Ni-Mn alloys were aged at 80°C for 24 hours and 150°C for 1 hour respectively to obtain stable transformation temperatures.

The micrographs in Figure 5 show the microstructures of the Cu-Zn-Al Jominy specimen along the longitudinal direction. At locations further and further away from the quench surface (corresponding to a gradually decreasing quench rate) the

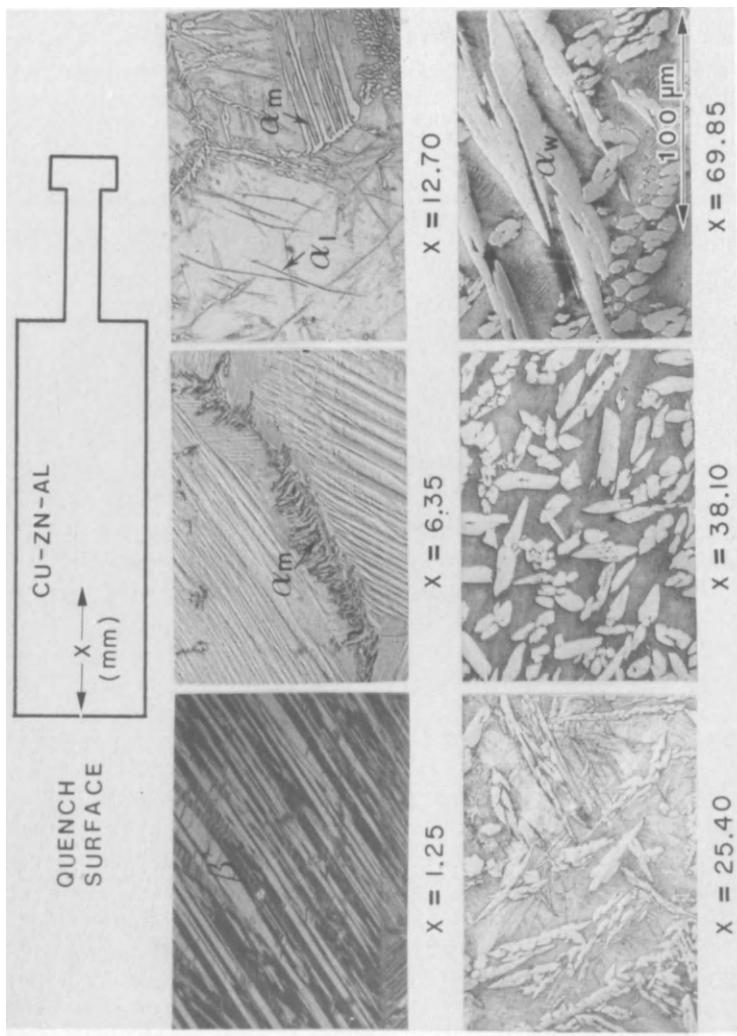


Figure 5. The microstructure of a Cu-26.0Zn-3.8Al Jominy bar at various distances from the quench surface.

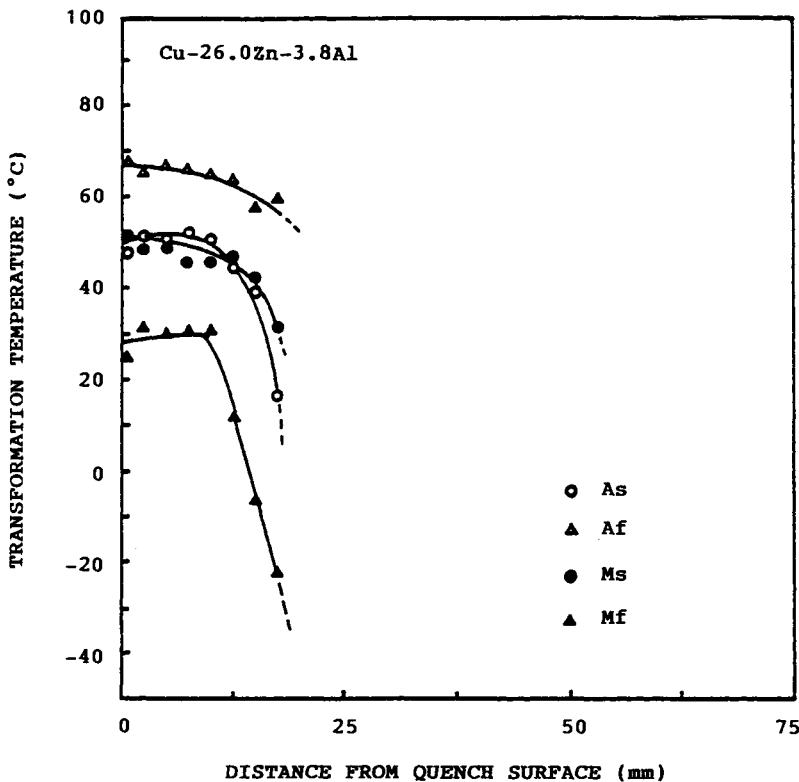


Figure 6. The transformation temperatures, M_s , A_t , A_s and F_t , of a Cu-26.0Zn-3.8Al Jominy bar at various distances from the quench surface.

In contrast to the Cu-Zn-Al alloy, a martensitic structure is obtained throughout the entire specimen of Cu-Al-Ni-Mn, as shown in Figure 7. However, the transformation temperatures exhibit a strong dependence on the quench rate, as can be seen in Figure 8. In areas close to the quench surface, the martensitic transformation occurs at temperatures close to the nominal transformation temperature. With increasing

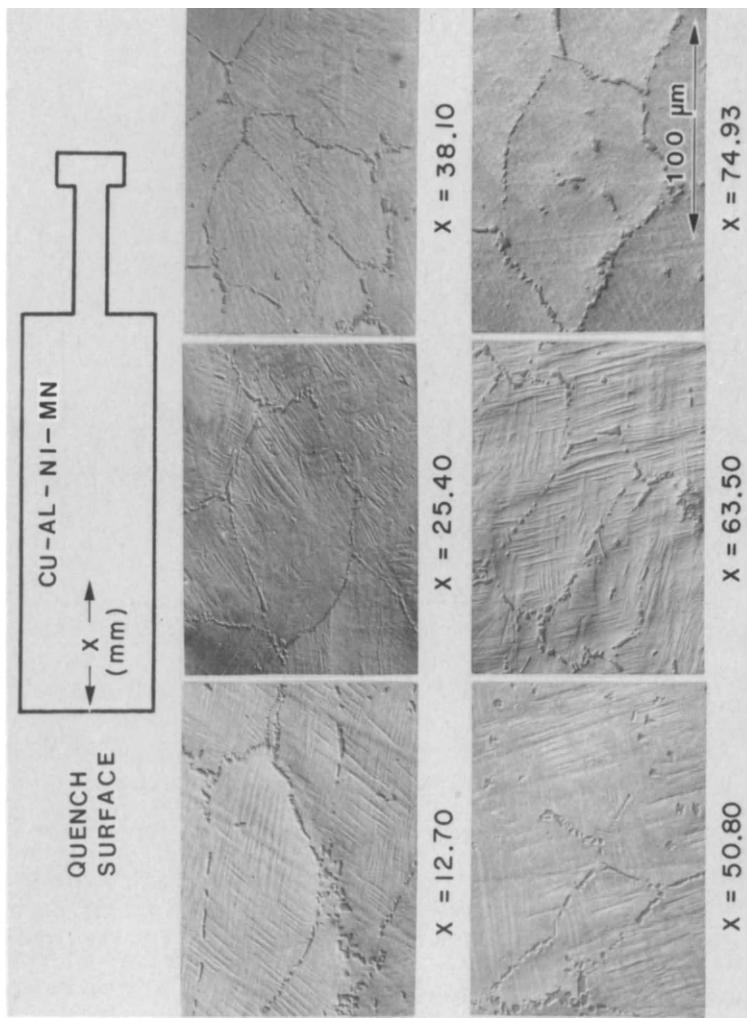


Figure 7. The microstructure of a Cu-12.7Al-5Ni-2Mn-0.5Ti Jominy bar at various distance from the quench surface.

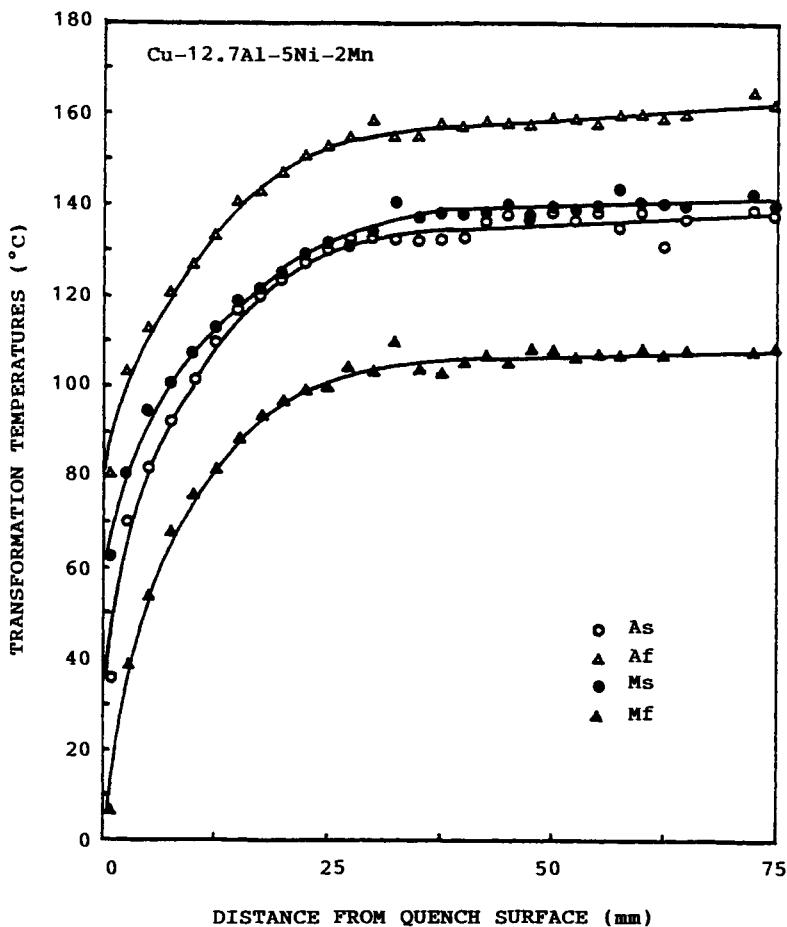


Figure 8. The transformation temperatures, M_s , A_f , A_s and A_f , of a Cu-12.7Al-5Ni-Mn-0.5Ti Jominy bar at various distances from the quench surface.

distance from the quench surface (decreasing quench rate) the transformation temperatures gradually increase and reach constant values. The effect is similar to the isothermal aging effect of as-quenched Cu-Al-Ni²³ to be discussed in the next section.

3.3 Aging Effects

Both the austenite β phase and martensite in Cu-base shape memory alloys are susceptible to aging effects even at moderate temperatures. For Cu-Zn-Al, a direct quench into the austenite β phase results in depressed transformation temperatures which increase and eventually stabilize during subsequent aging at low temperatures (<70°C).³⁹ An example of a transformation temperature change during the post-quench aging of Cu-26.67%Zn-4%Al at 25°C is shown in Figure 9.⁴⁰

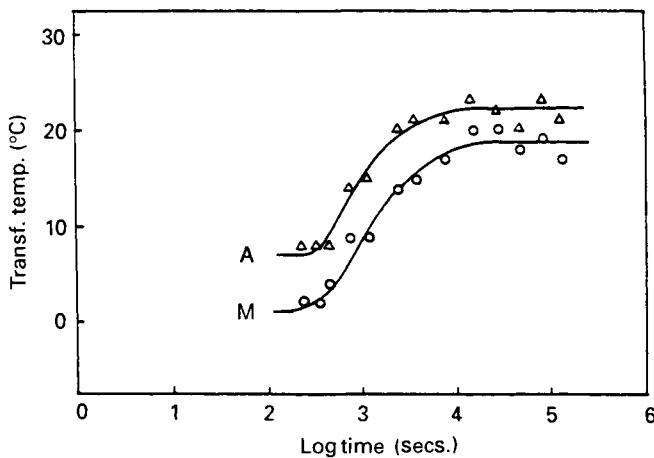


Figure 9. Changes in M and A temperatures of a Cu-26.67Zn.4Al alloy during post-quench aging at 25°C. The alloy had been betatized at 900°C for 3 minutes and subsequently quenched to a water bath at ambient temperature. The M and A temperatures correspond to the temperatures of maximum martensitic and reverse transformation rates respectively.

When Cu-Zn-Al is stabilized by post-quench aging and is then subjected to exposure at temperatures of 100-350°C, the change in transformation temperatures can be classified into three stages, as illustrated in Figure 10.⁴⁰ The first is characterized by a rapid drop of transformation temperatures which are then recoverable during subsequent aging at temperatures below 70°C.³⁹ The amount of depression, which is plotted versus aging temperature in Figure 11, increases with aging temperature up to 300°C and then decreases with further increases in the aging temperature. The transformation temperatures increase only moderately with time during the second stage. It is not until the third stage that the temperatures decrease drastically and the hysteresis is widened significantly.

Several studies^{39,40} interpreted the post-quench aging effect and the first stage behavior to be related to the ordering in Cu-Zn-Al alloys. The second stage effect has drawn less attention. It is probably related to the growth of ordered domains.⁴⁰ The third stage aging characteristics have been demonstrated to be caused by the decomposition of the β phase into either the α_1 bainite⁴² or the γ_2 phase.⁴³

The distribution of the three stages is very much dependent on the aging temperatures. At low temperatures, stages one and two tend to dominate. Decomposition occurs only after prolonged aging. As the aging temperature increases, the third stage gradually becomes dominant and occurs at a much shorter aging time.

Direct quenching of Cu-Zn-Al alloys into martensitic state causes stabilization of martensites which may not transform back to the parent phase during subsequent heating. Figure 12-I shows the resistivity curve of a Cu-25.80Zn-3.58Al alloy (nominal $M_s = 100^\circ\text{C}$) after directly quenching from the betatizing temperature to the ambient. No transformation is detected. A step-quench process with an intermediate anneal

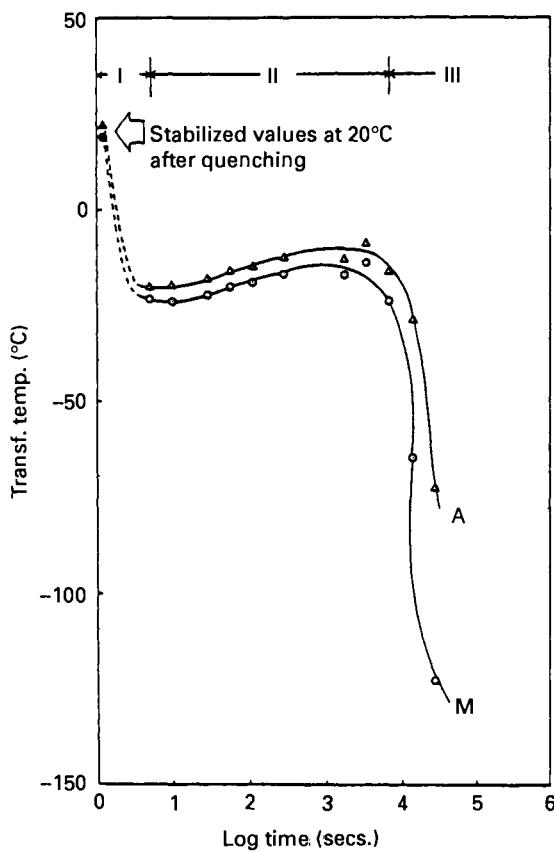


Figure 10. Changes in M and A temperatures of a Cu-26.67Zn-4Al alloy during isothermal aging at 200°C.

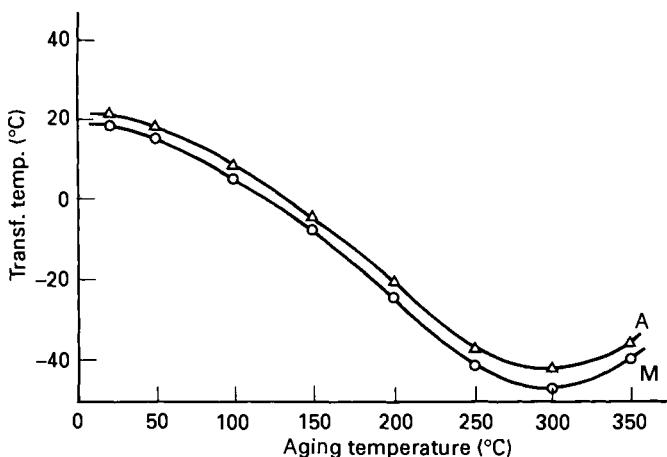


Figure 11. A plot of M and A temperatures of a Cu-26.6Zn-4Al alloy after the first stage depression as functions of the aging temperature.

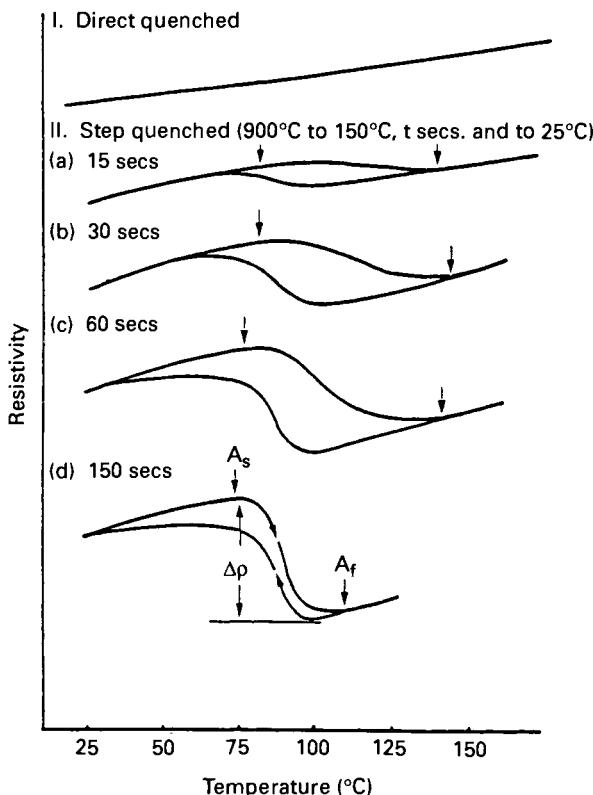


Figure 12. Effects of quenching on the transformation resistivity curve of a Cu-25.8Zn-3.6Al alloy; (I) direct-quenched specimens and (II) step-quenched specimens after intermediate holding at 150°C for t seconds.

in the austenite β phase state at 150°C gradually restores normal transformation and shape memory properties, as shown in figure 12-II.

Isothermal aging of Cu-Zn-Al martensites also causes martensite stabilization. Figure 13 shows the change of transformation resistivity curves during aging of a Cu-25.80Zn-3.58Al alloy at 80°C. Both A_s and A_f temperatures are gradually shifted to higher temperatures. The amount of transformation, as indicated by the total resistivity change, $\Delta\rho$, also decreases with martensite stabilization.

The cause of martensite stabilization is not yet well understood. It is generally thought to be related to vacancy-assisted reordering,⁴⁴⁻⁴⁷ structural modification⁴⁶ and interface pinning due to excess point defects or eventually precipitations.^{48,49} Excessive quenched-in point defects have been considered to accelerate the direct-quench effect⁵⁰.

The aging effects in Cu-Al-Ni alloys are sensitive to the Ni content.²³ For most alloys of interest containing 3-5% Ni, the transformation temperatures increase with aging time at 200-300°C and decrease at higher aging temperatures, e.g., 400°C.⁴¹ The aging characteristics are related to reordering at the early stage and later to the precipitation

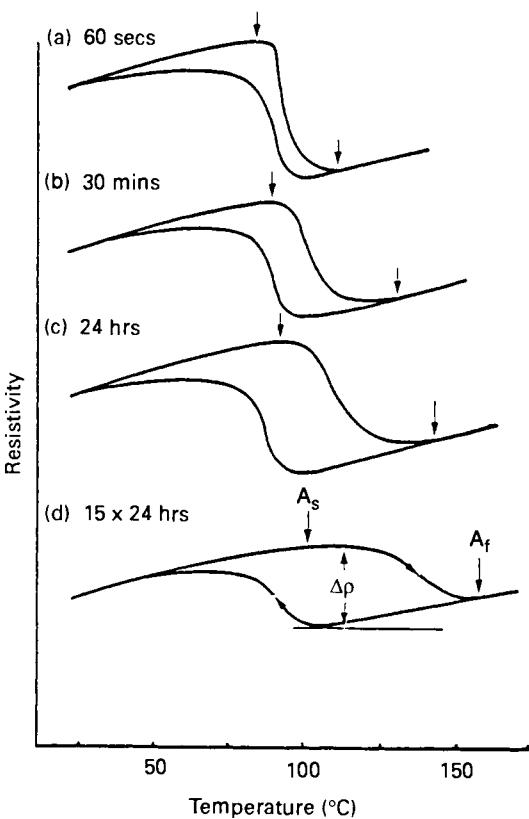


Figure 13. Changes in the reverse transformation resistivity curve of a step-quenched Cu-25.8Zn-3.6Al alloy during isothermal aging in the martensitic state at 80°C.

of γ_2 .^{41,51} Eventually, the presence of substantial amount of γ_2 causes the depression of transformation temperatures.

Cu-Al-Ni martensites are also susceptible to martensite stabilization, however, the kinetics are much slower. A study of a Cu-11.34Al-3Ni-0.04B alloy indicated that the stabilization effect is insignificant at temperatures below 110°C. The effect can be significant when the aging temperature is raised above 125°C.⁵²

The thermal stabilities of Cu-Zn-Al and Cu-Al-Ni are compared in Figure 14. The curve for the Cu-26.67Zn-4Al alloy defines the eventual degradation of shape memory due to β phase decomposition and does not consider the effects of drifts in transformation temperature. For applications which demand close control of transformation temperature, the limitation on the thermal exposure is indeed more restricted than those defined by this curve. The curve for the Cu-14.2Al-3.3Ni alloy, on the other hand, considers some arbitrary tolerance in shape memory performances such as switching temperature, the magnitude of the two-way effect and stress relaxation. It is obvious that Cu-Al-Ni alloys possess better thermal stability than Cu-Zn-Al alloys and are more suitable for higher temperature applications.

Table 1: Properties of Cu-base Shape Memory Alloys

	<u>Cu-Zn-Al</u>	<u>Cu-Al-Ni</u>
PHYSICAL PROPERTIES		
Density (g/cm ³)	7.64	7.12
Resistivity ($\mu\Omega\cdot\text{cm}$)	8.5-9.7	11-13
Thermal Conductivity (J/m·s·K)	120	30-43
Heat Capacity (J/Kg·°K)	400	373-574
MECHANICAL PROPERTIES		
Young's Modulus (GPa)		
β -phase	72	85
Martensite	70	80
Anisotropy Ratio	15	12
Yield Strength (MPa)		
Austenite	350	400
Martensite	80	130
Ultimate Tensile Strength (MPa)	600	500-800
SHAPE MEMORY PROPERTIES		
One-way Strain (%)	4	4
Two-way Strain (%)	2	2
Hysteresis (°K)	10-25	15-20

4. Mechanical Properties

Figure 15 contains stress-strain curves of a fine grained Cu-26.16%Zn-3.77%Al alloy ($M_s=30^\circ\text{C}$) which was tested in the temperature range of 0-100°C. As is typical of shape memory alloys, the stress-strain characteristics of Cu-base shape memory alloys are highly temperature dependent and exhibit multi-stage yielding at temperatures approaching the transformation temperatures. At temperatures between A_f and M_d , the alloy exhibits pseudoelasticity and the first stage deformation is associated with the stress-induced formation of martensite. The critical stress to induce martensite decreases linearly with decreasing temperature following a Clausius-Clapeyron relationship.⁵³ As the temperature decreases further approaching the M_f temperature, the initial deformation mode gradually changes from the stress-induced formation of martensite to the reorientation of the existing martensite variants, and the pseudoelasticity is gradually replaced by the shape memory effect.

Other mechanical properties of Cu-base shape memory alloys worthy of note are the anomalous elastic property of the parent β phase and the pseudoelasticity associated with the stress-induced martensite-to-martensite transformation.

It is common in β phase alloys that elastic constants soften as the temperature approaches the martensitic transformation temperature. Significant decrease of shear elastic constant, C' , in β phase Cu-Zn-Al⁵⁴ and Cu-Al-Ni⁵⁵ alloys have been observed with decreasing temperature in the neighborhood of martensitic transformation temperature. Although less significant in magnitude, other elastic constants also exhibit positive temperature dependence in this temperature range. Consequently, the β phase Cu-Zn-Al and Cu-Al-Ni alloys are highly anisotropic with anisotropy ratios as high as 15 and 12 for Cu-Zn-Al and Cu-Al-Ni respectively. The high anisotropy has been considered to be responsible for the intergranular failure of Cu-base shape memory alloys⁵⁶.

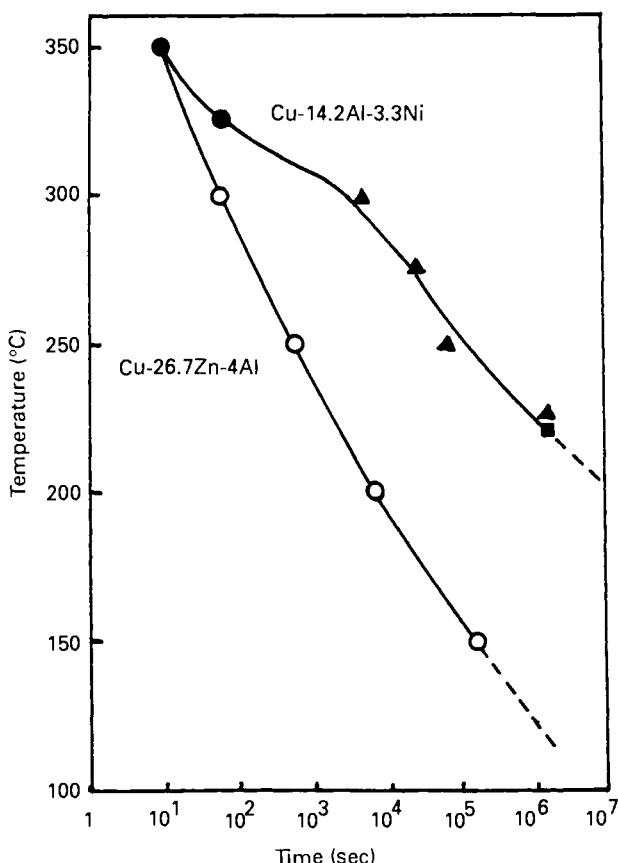


Figure 14. The decomposition curve of a Cu-25.7Zn-4Al alloy and a thermal stability curve which defines the acceptable thermal exposure of a Cu-14.2Al-3.3Ni alloy.

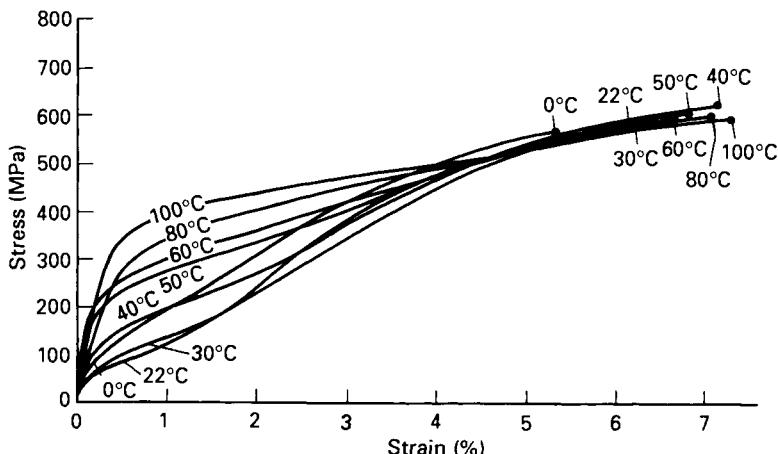


Figure 15. Isothermal stress-strain curves of a Cu-26.16Zn-3.77Al alloy with M_s temperature of 30°C. The test was performed at temperatures of 0-100°C.

The pseudoelastic strain in Cu-base shape memory alloys is typically 4-6%. However, when a successive martensite-to-martensite transformation is stress-induced, the magnitude of pseudoelastic strain can be greatly increased. In both austenitic β_1 Cu-Zn-Al⁵⁷ and Cu-Al-Ni³⁰ alloys, $\beta_1 \rightarrow \beta_1 \rightarrow \alpha_1'$ transformation can be stress-induced at low temperatures while at high temperatures, the α_1' martensite is directly induced from the parent β_1 phase. In martensitic γ_1' Cu-Al-Ni alloys, the $\gamma_1' \rightarrow \beta_1'' \rightarrow \alpha_1' (\beta_1''$ is another stacking version of 18R martensite) successive transformation can also be a dominant deformation mode.³⁰ These successive stress-induced martensite-to-martensite transformations give rise to multistage pseudoelasticity, the total recoverable strain of which can be as high as 17% in single crystals. The mechanical properties together with other physical and shape memory properties are summarized in Table 1. Young's modulus of a shape memory alloy becomes difficult to define as the temperature approaches the transformation temperature where the alloy exhibits nonlinear pseudoelasticity and the modulus exhibits both temperature and strain dependence.

5. Rapid Solidification and Powder Metallurgy

Fine grain Cu-base shape memory alloys without grain-refining element additions can be produced by powder metallurgy or rapid solidification processing. A grain size of around 30 μm has been obtained in a powder processed Cu-Zn-Al alloy⁵⁸ which was water atomized to an averaged particle size of 150 μm and subsequently cold compacted and hot extruded. Powder processed Cu-Zn-Al shows better resistance to cycling degradation in recoverable strain than conventionally processed alloys. A powder processed Cu-Al-Ni alloy with grain size of 20 μm also exhibits a significantly improved ductility (5 to 7% compared to 0.6% in cast alloys).⁹

Fine grain Cu-base shape memory alloys with grain sizes of 1-20 μm have also been obtained by rapid solidification processing techniques, such as melt spinning⁵⁶⁻⁶³ and melt extraction.⁶⁴ Because of their small grain size and large quenched-in defect population, rapidly solidified Cu-Zn-Al⁵⁹ and Cu-Al-Ni⁶³ alloys show transformation at lower temperatures than nominal. The transformation temperatures decrease with increasing cooling rate and melt temperature. The initial transformation in the as-processed alloys also tends to be sluggish and exhibit a wide hysteresis. Consistent transformation with recovered transformation temperatures can be obtained after post-quench aging⁶² or thermal cycling.⁵⁹

The shape memory properties of rapidly solidified Cu-Zn-Al alloys are usually less predictable than those of similarly processed Cu-Al-Ni alloys. Melt extracted Cu-Zn-Al alloys that show transformation temperatures significantly higher than those of bulk alloys have been reported.⁶⁴ This complication appears to come from a quench stabilization effect⁶⁰ and also the difficulty in controlling the Zn loss through vaporization in the melt.⁶¹

6. Environmental Effects

Cu-Zn-Al memory alloys are susceptible to dezincification which occurs preferentially along surface crevices, grain boundaries and martensite interfaces.⁶⁵ This depends strongly on both the phase present and the composition. At comparable Zn contents, martensitic alloys are more sensitive to dezincification than are austenitic alloys. Resistance to dezincification is also improved by increasing the Al content.

Cu-Zn-Al alloys are also susceptible to stress-corrosion cracking which causes intergranular failure.⁶⁶⁻⁶⁷ Unlike corrosion, stress-corrosion cracking of Cu-Zn-Al alloys is insensitive to either phases or Al contents of between 4-8wt.%. The NO₂⁻ anion is the most aggressive for stress-corrosion attack in Cu-Zn-Al alloys. The problem also occurs in solutions containing NO₃⁻ and SO₄²⁻ ions and ammoniacal solution with PH values of 7.9-11.

Surface coating may improve the resistance of Cu-base alloys to both dezincification and stress-corrosion cracking, however, surface coating acts as biasing a constraint to the shape memory effect. For most metallic coatings, the allowable elastic strain is typically less than 0.5% which is too small to accommodate the shape memory strain which can be as high as 6%. The interfacial bonding will deteriorate with shape memory cycling. To use surface-coated shape memory alloys, it is therefore necessary to keep the shape memory strain as small as possible.

7. Joining

Cu-base shape memory alloys can be welded, brazed or soldered to make good joints. However, thermal effects on the weld and the heat affected zone need to be evaluated. In most cases, the weld is shape memory inactive due to either the change in chemical composition or the significantly different cooling process. The heat affected zone may also suffer a loss in shape memory properties due to decomposition or an aging-induced drift in transformation temperatures. When the presence of an inactive or degraded weld and heat affected zone is unacceptable, mechanical joining techniques are to be preferred.

8. Conclusion

To select a shape memory alloy for a particular application, both the cost and the alloy performance need to be considered. With their relatively low cost, Cu-base alloys can be more cost-effective than Ni-Ti alloy in many applications though their limitations (particularly in thermal stability and corrosion resistance) should be evaluated. Cu-Al-Ni alloys are better shape memory alloys in terms of their thermal stability and corrosion resistance, however, higher processing costs are usually encountered because of their lack of appropriate low temperature ductility.

References:

1. Miura, S., Morita, Y. and Nakanishi, N., Shape Memory Effects in Alloys, ed. Perkins, J., Plenum Press, N.Y., 389 (1975).
2. Stice, J.D. and Wayman, C.M., Met. Trans A, 13A, 1687 (1982).
3. Delaey, L., Van Humbeeck, J., Chandrasekaran, M., Janssen, J. Andrade, A. and Mwamba, N., Metals Forum, 4, 164 (1981).
4. Wayman, C.M., Shape Memory Effects in Alloys, ed. Perkins, J., Plenum Press, N.Y., 1 (1975).
5. Wield, D.V. and Gallam, E., Scripta Met., 6, 1157 (1972).
6. Guzman, L. and Ruhle, M., Proceedings, International Conference on Martensitic Transformations, Cambridge, Massachusetts, 590 (1979).

7. Saburi, T. and Nenno, S., Scripta Met., **8**, 1368 (1974).
8. Chandrasekaran, L. and Miodownik, A.P., Proceedings, International Conference on Martensitic Transformations, Cambridge, Massachusetts, 584 (1979).
9. Duerig, T.W., Albrecht, J. and Gessinger, G.H., J. of Metals, **34**, 14 (1982).
10. Otsuka, K. and Shimizu, K., Scripta Met., **4**, 469 (1970).
11. Higuchi, A., Suzuki, K., Masumoto, Y., Sugimoto, K., Komatsu, S. and Nakamura, Y., Proceedings, International Conference on Martensitic Transformations, Leuven, Belgium, C4-767 (1982).
12. Nagasawa, A., Nakanishi, N., Suzuki, T., Enami, K. and Nenno, S., Proceedings, International Conference on Martensitic Transformations, Cambridge, Massachusetts, 429 (1979).
13. Kubo, H., Miyaki, A. and Shimizu, K., Trans. Jap. Inst. of Metals, **24**, 603 (1983).
14. Mwanba, N. and Delaey, L., Proceedings, International Conference on Martensitic Transformations, Leuven, Belgium, C4-639 (1982).
15. Brook, P.L., U.S. Patent, No. 4166739, September (1979).
16. Sugimoto, K., Bulletin of Jap. Inst. of Metals, **24**, 45 (1985).
17. Ikai, Y., Murakami, K. and Mishima, K., Proceedings, International Conference on Martensitic Transformations, Leuven, Belgium, C4-785 (1982),
18. Lee, J.S. and Wayman, C.M., Metallography, **19**, 401 (1986).
19. Elst, R., Van Humbeeck, J., Meeus, M. and Delaey, L., Z. Metalld, **77**, 421 (1986).
20. Sugimoto, K., Kamei, K., Matsumoto, H., Komatsu, S., Akamatsu, K. and Sugimoto, T., Proceedings, International Conference on Martensitic Transformations, Leuven, Belgium, C4-761 (1982).
21. Sure, G.N. and Brown, L.C., Met. Trans.A, **15A**, 1613 (1984).
22. Rapacioli, R. and Ahlers, M., Scripta Met., **11**, 1147 (1979).
23. Nakamura, F., Kusui, J., Shimizu, K. and Takamura, J., J. Jap. Inst. of Metals, **44**, 1302 (1980).
24. Delaey, L., Chandrasekaran, M., Andrade, M. and Van Humbeeck, J., Proceedings, International Conference on Solid-Solid Phase Transformations, Pittsburgh, TMS-AIME, 1429 (1981).
25. Dunne, D.P. and Kennon, N.F., Scripta Met., **10**, 729 (1982).
26. Agafonov, V., Naudot, P., Dubertret, A. and Dubois, B., Scripta Met., **22**, 489 (1988).
27. Melton K.N. and Mercier, O., Met. Trans. A, **10A**, *75 (1979).
28. Barcelo, G., Ahlers, M. and Rapacioli R., Z. Metallkde, **70**, 732 (1979).
29. Barcelo, G. and Ahlers, M., Scripta Met., **16**, 1 (1982).
30. Otsuka, K., Sakamoto, H. and Shimizu, K., Acta Met., **27**, 585 (1979).
31. Adnyana, D.N., Memory Metals, Inc., Internal Report, (1985).
32. Ahlers, M., Scripta Met., **8**, 213 (1974).
33. Delaey, L., Chandrasekaran, M., De Jonghe, W., Rapacioli, R. and Deruyttere, A., International Copper Research Association, Inc., Research Report, No.238.
34. Dunne, D.P. and Kennon, N.F., Metals Forum, **4**, 176 (1981).
35. Brezina, P., International Metals Reviews, **27**, 77 (1985).
36. Zhou, X., Liu, X. and Zheng, Y., Proceedings, International Conference on Shape Memory Alloys, Guilin, China, 261 (1987).
37. Han, Y.S. and Kim, Y.G., Scripta Met., **21**, 947 (1987).
38. Wu, M.H., Unpublished Research Work, Memory Metals, Inc.
39. Schofield, D. and Miodownik, A.P., Metals Technology, **7**, 167 (1980).
40. Wu, M.H. and Wayman, C.M., Unpublished Research Work, the University of Illinois.

41. Cook, J.M. and Brown, L.C., Scripta Met., **12**, 949 (1978).
42. Kennon, N.F., Dunne, D.P. and Middleton, L., Met. Trans. A, **13A**, 551 (1982).
43. Rapacioli, R. and Chandrasekaran, M., Proceedings, International Conference on Martensitic Transformations, Cambridge, Massachusetts, 586 (1979).
44. Scarsbrook, G., Cook, J. and Stobbs, W.M., Proceedings, International Conference on Martensitic Transformations, Leuven, Belgium, C4-703 (1982).
45. Jassen, J., Van Humbeeck, J., Chandrasekaran, M., Mwamba, N. and Delaey, L., Proceedings, International Conference on Martensitic Transformations, Leuven, Belgium, C4-715 (1982).
46. Delaey, L., Suzuki, T. and Van Humbeeck, J., Scripta Met., **18**, 899 (1984).
47. Arab, A.A., Chandrasekaran, M. and Ahlers, M., Scripta Met., **18**, 709 (1984).
48. Arab, A.A., and Ahlers, M., Proceedings, International Conference on Martensitic Transformations, Leuven, Belgium, C4-715 (1982).
49. De Graef, M., Van Humbeeck, J., Andrade, M. and Delaey, L., Scripta Met., **19**, 643 (1985).
50. Van Humbeeck, J., Segers, D. and Delaey, L., Scripta Met., **19**, 477 (1985).
51. Fisher, P., Dunne, D. and Kennon, N., Proceedings, International Conference on Martensitic Transformations, Nara, Japan, 946 (1986).
52. Van Humbeeck, J., Delaey, L. and Roedolf, D., Proceedings, International Conference on Martensitic Transformations, Nara, Japan, 862 (1986).
53. Warlimont, H., Delaey, L., Krishnan, R.V. and Tas, H., J. Mat. Sci., **9**, 1545 (1974).
54. Guenin, G., Morin, M., Gobin, P.F., De Jonghe, W. and Delaey, K., Scripta Met., **11**, 1071 (1977).
55. Suezaure, M. and Sumino, K., Scripta Met., **10**, 789 (1976).
56. Miyazaki, S., Otsuka, K., Sakamoto, H. and Shimizu, K., Trans. Jap. Inst. of Metals, **4**, 224 (1981).
57. Delaey, I., Janssen, J., Van Humbeeck, J., Luyten, J. and Deruyttere, A., Proceedings, International Conference on Martensitic Transformations, Cambridge, Massachusetts, 645 (1979).
58. Janssen, J., Willems, F., Verelst, B., Maertens, J. and Delaey, L., Proceedings, International Conference on Martensitic Transformations, Leuven, Belgium, C4-809 (1982).
59. Perkins, J., Raymant, J.J. and Cantor, B., Proceedings, International Conference on Solid to Solid Phase Transformations, Pittsburgh, TMS-AIME, 1481 (1981).
60. Scarsbrook, G. and Stobbs, W.M., Acta Met., **35**, 47 (1987).
61. Oshima, R., Tanimoto, M., Oka, T., Fujita, F.E., Hanadate, Y., Hamada, T. and Miyagi, M., Proceedings, International Conference on Martensitic Transformations, Leuven, Belgium, C4-749 (1982).
62. Wood, J.V. and Shingu, P.J., Met. Trans. A, **15A**, 471 (1984).
63. Eucken, S. and Honbogen, E., Proceedings, International Conference on Martensitic Transformations, Nara, Japan, 780 (1986).
64. Stobbs, W.M. and Wood, J.V., Acta Met., **27**, 575 (1979).
65. Celis, J.P., Roos, J.R. and Terwinghe, F., J. Electrochem. Soc., **130**, 2314 (1983).
66. Terwinghe, F., Celis, J.P. and Roos, J.R., Br. Corrosion J., **19**, 107 (1984).
67. Terwinghe, F., Celis, J.P. and Roos, J.R., Br. Corrosion J., **19**, 115 (1984).

Cu-Al-Ni-Mn: A New Shape Memory Alloy for High Temperature Applications

Koichi Sugimoto*, Kiyoshi Kamei* and Masayosi Nakaniwa**

*Department of Metallurgical Engineering, Faculty of Engineering
Kansai University, Japan

**Graduate student of Kansai University, Japan

Cu-Zn-Al, which is the most popular copper-based shape memory alloy, does not always exhibit sufficient mechanical strength and thermal stability. It is also well known that the maximum M_s that can be achieved in the Ti-Ni, which is the commercially leading shape memory alloy in Japan, is only 100°C. On the other hand Cu-Al-Ni can exhibit a much higher M_s than Ti-Ni or Cu-Zn-Al. It does not, however, normally exhibit as good deformability as Cu-Zn-Al alloys because of the brittle γ_2 -precipitates, and hence Cu-Al-Ni has never been used commercially in spite of superior thermal stability. The purpose of the present report is to review the research work¹⁻² for improving the deformability of Cu-Al-Ni shape memory alloys in order to develop a new type of commercial alloy, the CANTiM series, which could be successfully used, for instance, as electrical circuit-breakers operating at temperatures around 160-170°C.

1. The Principles of the Development Program

Metallographic examination of a fractured specimen of a Cu-13.5wt.%Al-4wt.%Ni alloy revealed that the fracture itself was mainly caused by brittle γ_2 -precipitates. It was therefore concluded that the suppression of the precipitation of γ_2 -phase would bring about a better deformability. This has been proven in the case of a Cu-13.5wt.%Al-4wt.%Ni alloy, when the Al content was reduced to about 12% and no precipitation of γ_2 -phase can occur. M_s , however, was increased from 40°C to about 350°C by this change, since the effect of Al on the M_s is very large. In order to decrease the M_s from 350°C to 125°C it was necessary to add other alloying elements, such as Mn and Ni, which are both soluble in the β -phase and decrease the M_s considerably.

Finally the addition of 1% Ti has led to a well-defined grain-refinement¹, probably because of the finely dispersed precipitates of the x-phase⁴, which are enriched with Al, Ni and Ti. These can be obstacles for grain boundary motion and suppress grain growth.

The above three effects worked cooperatively to result in the final composition (by weight):

11.88% Al
5.06% Ni
2.01% Mn
1.01% Ti
Balance Cu

The new alloy was named as "CANTiM-125" (the 125 indicating the A_s temperature). By adjusting the Al-content the A_s can be controlled in a range from 75°C to 175°C. These alloys are called CANTiM-75, CANTiM-100, CANTi-125, CANTiM-150, and CANTiM-175. Their chemical compositions with those of some reference alloys are given Table 1.

TABLE 1: Chemical Composition (in wt%) of Shape Memory Alloys

Alloys	Al	Ni	Mn	Ti	Zn	Cu	Ms
CANTiM-175	11.85	4.92	1.87	1.04	--	Bal.	172
CANTiM-125	11.88	5.06	2.01	1.01	--	Bal.	126
CANTiM-75	12.40	5.08	2.02	1.00	--	Bal.	74
CZA-70	5.72	--	--	--	20.59	Bal.	69
CZA-50	4.30	--	--	--	25.10	Bal.	40
Ti-Ni-60				44.6		Bal.	60

2. Shape Memory Effect

Figure 1 shows the ultimate tensile strength of a CANTiM-125 alloy specimen as a function of solution treatment temperature. The tensile strength of both the martensite (at 10°C) and the austenitic β -phase (200°C) are remarkably high, for instance 880MPa when solution-treated at 800°C. It is much higher than in the Cu-Zn-Al alloys.

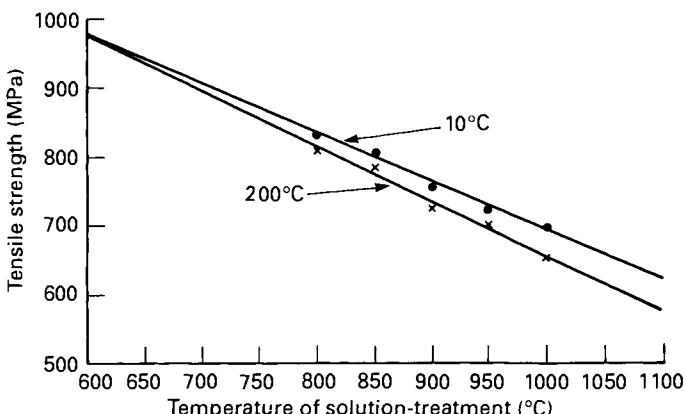


Figure 1 Ultimate tensile strength of martensite (at 10°C) and austenite (at 200°C) measured with a sheet material of CANTiM-125 alloy as a function of solution-treatment temperature.

Figure 2 shows the result of the displacement measurements in a shape memory coil-spring made of CANTiM-150 wire materials as a function of temperature under a constant load of 1N. It is seen from Figure 2 that the shape memory effect in this alloy is almost perfectly reversible even in a very high operating temperature range (100°C-180°C).

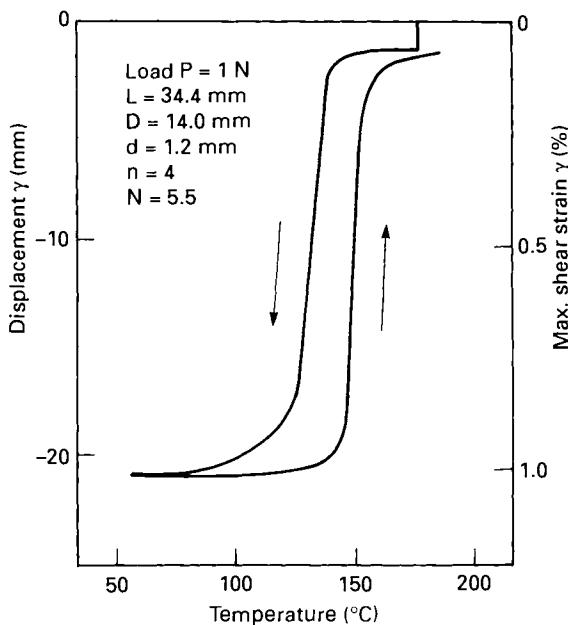


Figure 2 Actuator characteristics of a coil spring made of CANTiM-150 alloy, solution-treated for 10 min. at 900°C and then water quenched.

3. Thermal Stability

Figure 3 shows the DSC curves measured with CANTiM-75 and CZA-70. They are solution-treated and quenched, then exposed at 200°C, while in the CZA-70 alloy, which is a typical Cu-Zn-Al alloy, A_s was considerably decreased after a 7 hour exposure at 200°C, and completely eliminated by a 1 day exposure at 200°C.

Figure 4 summarizes the behavior of the transformation temperatures with exposure time at 200°C for both alloys shown in Figure 3. In CANTiM-75 (Figure 4 top) the transformation temperatures remained rather stable up to an exposure time of 36 hours and then increase gradually with exposure time. On the other hand the transformation temperatures in CZA-70 (Figure 4 bottom) show a drastic change with exposure time, when exposed longer than 3 hours. A much better thermal stability of the CANTiM75 alloy was thus confirmed experimentally by comparing the result with that of the Cu-Zn-Al alloy CZA-70 at the same operating temperature.³

The difference in the direction of the change in M_s , i.e., increase of M_s in CANTiM-75 and decrease in CZA-70, can be understood presumably by considering the difference in the order structure of the beta-phase.

4. Recovery Force

Figure 5 shows a comparison of recovery force in three different types of shape memory alloys on constrained recovery under torsion of a straight wire. Note that the CANTiM-175 alloy wire can exhaust the torsional force quite suddenly. The thermal

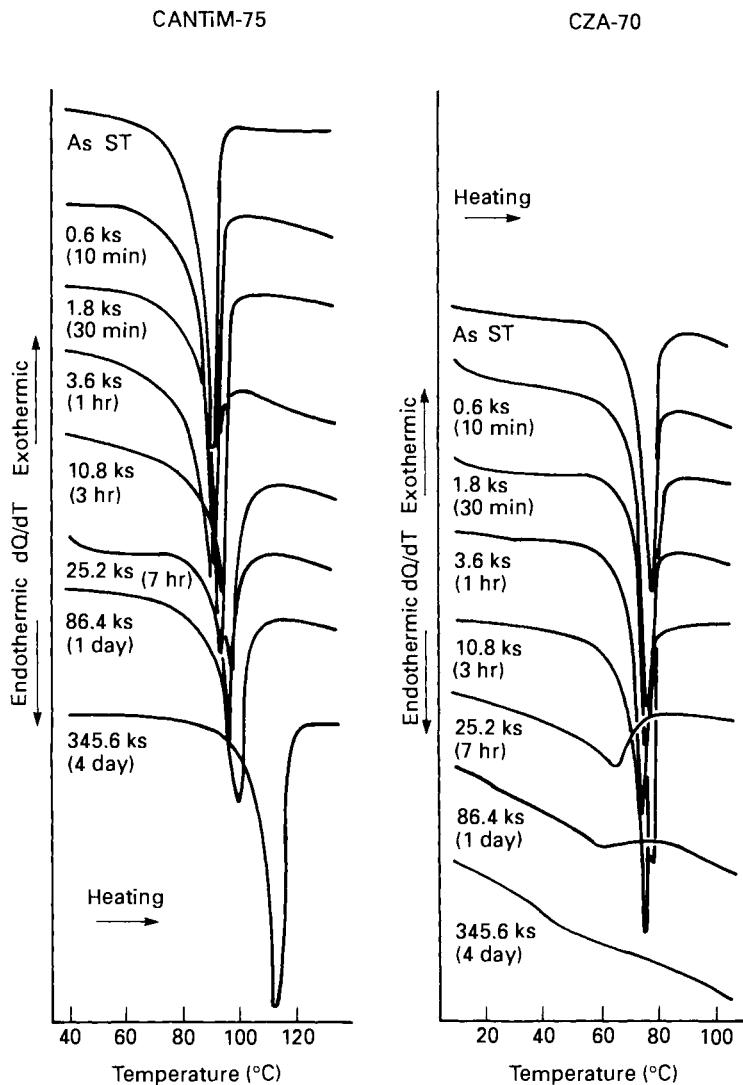


Figure 3 Change in the DSC-curves of both CANTiM-75 (left) and CZA-70 (right) alloys for various times of exposure to 200°C. CANTiM-75 did not change much except for one exposed for 4 days, while those of CZA-70 totally changed after a 7-hr exposure to 200°C. Solution treatment was for 10 min at 900°C (CANTiM-75) and 800°C (CZA-70), respectively, and then quenched into water.

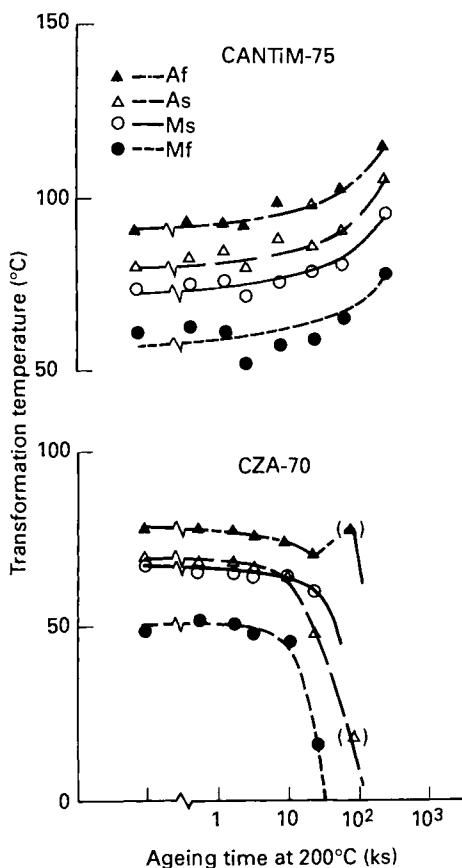


Figure 4 Change in martensitic (or reverse) transformation temperatures with time of exposure to 200°C in CANTiM-75 (top) and CZA-70 (bottom) alloys. The transformation temperatures in CANTiM-75 are stable even after a 4-day exposure to 200°C in contrast to the drastic change in those of CZA-70 exposed to the same temperature.

hysteresis of the curves in CANTiM-175 was much smaller than in the TiNi-60 or CZA-50 alloys.

5. Conclusions

A series of new Cu-based shape memory alloys, the CANTiM Series, has been developed for applications at elevated temperatures, especially electrical applications around 160-170°C. The following results were obtained:

- Cu-Al-Ni shape memory alloys can be used for high temperature applications requiring high strength, when the alloys are properly tuned by alloying and heat-treatment.
- Reducing Al content to 12% suppresses brittle fracture due to the γ_2 -phase and therefore dramatically improves ductility.

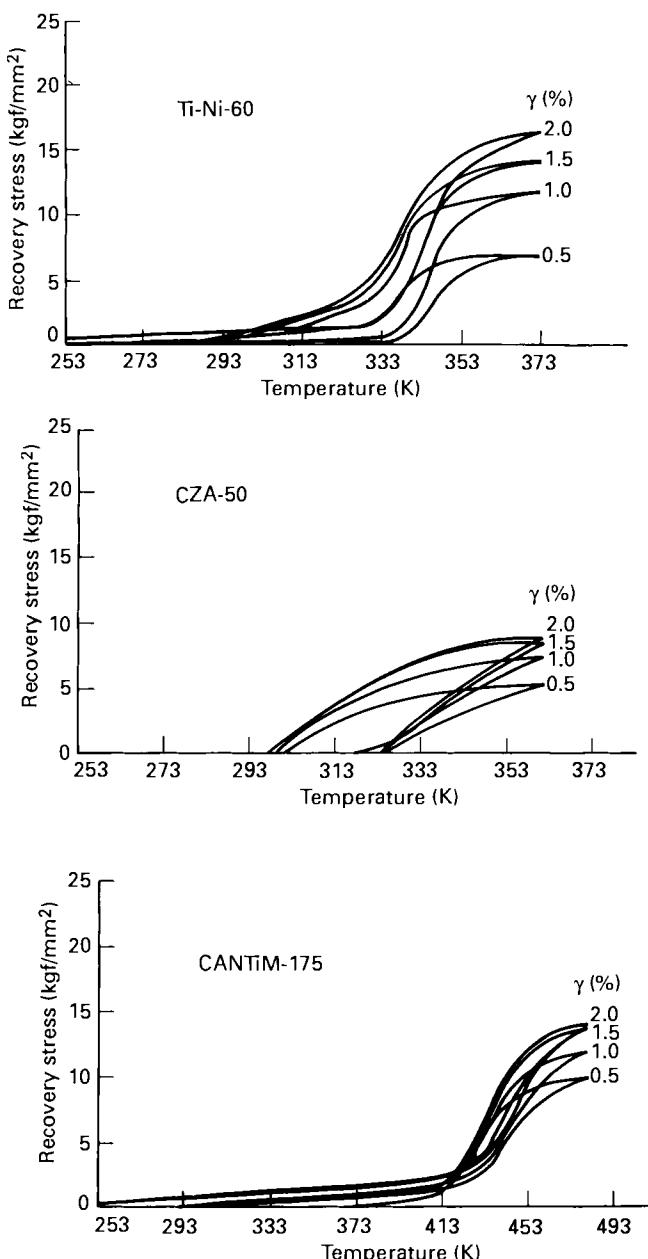


Figure 5 Comparison of the recovery-stress versus temperature curves on constrained recovery of a straight wire specimen under torsion in three types of shape memory alloy, TiNi (top), CZA-50 (middle), and CANTiM-175 (bottom). Note the steep stress increase with temperature showing less hysteresis in CANTiM-175 alloy than others.

- c) Additions of 5%Ni stabilize the β -phase and decrease A_s to a reasonable value. Ni can also increase the tensile strength.
- d) Additions of 2%Mn suppress the eutectoid reaction ($\beta \Rightarrow \alpha + \gamma_2$) and hence promote the martensite formation at lower quenching rates.
- e) Additions of 1%Ti result in a grain refinement and improve the mechanical properties.

6. Acknowledgements

This work has been partly supported by a Grant-in-Aid for "University-to-University Cooperative Research Program", Japanese Ministry of Education, contract No. 62045041. The authors wish to express their hearty thanks to Professors Y. Murakami, S. Komatsu, and K. Akamatsu, and to full-time lecturers Mr. Y. Nakamura, Dr. K. Nakao, and Dr. T. Sugimoto, Kansai University.

They would like to thank Mr. T. Takashima, Mr. A. Miyamoto, and Mr. Y. Itsumi, Kobe Steel Ltd., Kobe, for their technical support in preparation and analysis of the alloy materials. The scientific support from overseas to this research project by Prof. L. Delaey, Prof. E. Aernoudt, Dr. Jan Van Humbeeck, and Mr. R. Stalmans, Katholieke Universiteit Leuven, Belgium, is deeply appreciated by the authors.

References

1. Sugimoto, K. et al., J. de Phys., Tome 43, C4-761-766 (1982).
2. Kamei, K. et al., Shindou Gijutsu Kenkyuu Kaishi, 17, 92-99, (1978), 19, 201-211, (1980), 20, xx-26 (1981), 22, 156-160 (1983), 26, 137-143 (1987), (in Japanese).
3. Komatsu, S. et al., Bull. Jpn. Inst. Met., 51, 599-607, (in Japanese), (1987).
4. Adachi, K. et al., Scripta Met., 21, 453-458, (1987).

On The Stability of Shape Memory Alloys

J. Van Humbeeck, R. Stalmans, M. Chandrasekaran and
L. Delaey

Department of Metallurgy and Materials Engineering,
Katholieke Universiteit Leuven
De Croylaan 2, 3030 Heverlee (Belgium)

The unique properties of shape memory alloys are controlled by and are dependent on four external parameters: temperature (T), stress (σ), strain (ϵ) and time (t). These parameters cannot be changed independently. The complete mechanical behavior of shape memory alloys has to be determined from a (T, σ, ϵ) -diagram in which the temperature axis covers the general temperature range from about 50°C below to about 100°C above M_s . The value of a point along the (T, σ, ϵ) -surface is not always constant and can move in any direction in that space. This time-dependency can result from creep, stress relaxation and changes due to variations in the chemical free energy of martensite and/or parent phases.

Since in most shape memory alloys we are dealing with metastable phases and a diffusionless martensitic transformation, the generated phases are not necessarily in their lowest possible energy state immediately after the thermomechanical processing. In nature all systems tend towards lowest free energy, and this process, which is at the origin of the time dependency of the SME behavior, is called stabilization. As a consequence of this change in free energy, a time dependent, thermally activated change in transformation temperature will be observed, shifting the (T, σ, ϵ) -surface along the temperature axis. Fortunately, creep, stress-relaxation and stabilization can be easily controlled and the effects of stabilization can be partially or completely annihilated by simply heating slightly above the martensitic reverse transformation temperature, A_f . The shape memory effect, which occurs upon heating above A_s , causes a reversing of the σ and ϵ changes. The original transformation temperatures are restored because the parent phase very rapidly regains its original atomic configuration.

In most shape memory applications, the changes in properties or dimensions due to instabilities are small and can be neglected. In some cases, however, they have to be taken into consideration with the design of the memory device and in a few rare cases they will really limit or make impossible the application of a shape memory element. Moreover, in a few cases instabilities may be desirable (as will be shown later in this paper).

In the present paper the most important instabilities encountered in shape memory alloys will be described, omitting detailed discussions on physical origin; those interested in more details of the physical origin of stabilization are urged to consult the literature given at the end of this paper.

The most important instabilities of which one must be aware are those due to:

- (a) the hysteresis of the martensitic transformation
- (b) overheating in the beta phase
- (c) stabilization of the martensitic phase
- (d) shape memory degradation
- (e) differences between design and actual behavior

Before proceeding with the description of these phenomena, it should be made clear that they are only valid for alloys which have undergone the normal "reference" heat-treatment. For Ni-Ti alloys, this would mean a solution treatment possibly followed by a special aging treatment. For Cu-Zn-Al alloys, we consider a heat-treatment in which excess vacancies are annihilated and the material is fully ordered. This condition is best obtained by a so-called step-quench, which means quenching after betatizing to an intermediate temperature before quenching to room temperature. For Cu-Al-Ni alloys, the reference treatment is a solution treatment at high temperature followed by quenching in cold water.

1. The Hysteresis Effect

The martensitic transformation in Ni-Ti and Cu-based alloys is thermoelastic. During the transformation, the free energies (including the chemical, surface and elastic contributions) of both martensite and parent phase are thus equal as long as both phases coexist. The progress of the transformation is thus not only controlled by the chemical driving force due to a decrease or increase in temperature, but also to the elastic strain energy stored in the material during the forward transformation and released during the reverse transformation, and by the frictional force due to the movement of interfaces and creation of defects. The latter two contributions are the main controlling factors for the hysteresis of the transformation. The frictional energy means an energy loss, and the greater the frictional energy, the broader the hysteresis. The elastic energy that is stored or released during transformation will hinder the forward transformation but will assist the reverse transformation. The difference between A_f and M_s (or A_s and M_f) will thus be proportional to the amount of the stored elastic energy. The influence of both factors on the transformation hysteresis curve is schematically shown in Figure 1. Stable and reproducible shape memory performances will be achieved by cycling between the fully martensitic and fully austenitic conditions (meaning between A_f and M_f). If, during cycling, only partial transformation is possible, instabilities may arise. Such an instability causes a gradual shift of the SME device towards the shape it has in the fully β or fully martensite condition. The hysteretic behavior of thermoelastic martensitic alloys on partial cycling shows two remarkable features:

- (a) the hysteresis is static, i.e. within some limits it is independent of the variation rate of the parameter controlling the transformation(temperature or stress).
- (b) the hysteresis exhibits global memory (as opposed to local memory), i.e. the transformed volume for a given value of temperature or stress can only be determined from the path that has been followed in the T- σ - ϵ space.

As a consequence, regulating mechanisms relying on partial cycling with low amplitude temperature fluctuations (2 to 5°C) are almost impossible unless the

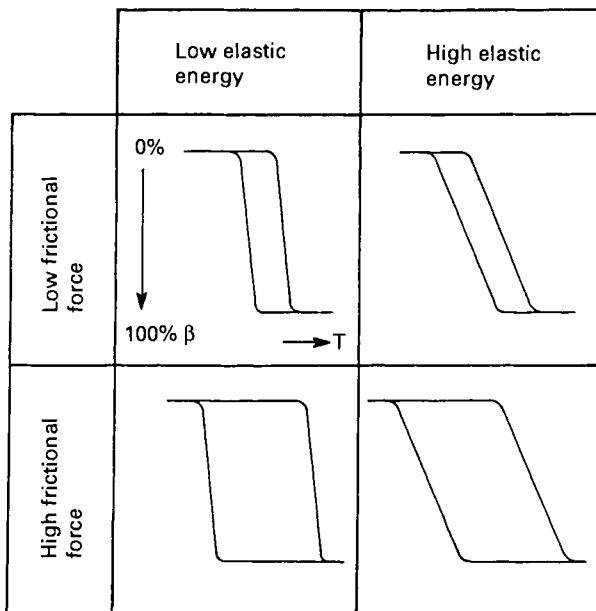


Figure 1 The influence of the elastic strain energy and the frictional force on the hysteresis curve of a thermoelastic martensite transformation.

material can be regularly transformed completely into one of the two phases, either beta or martensite. It can be shown that the device can then be brought back in its original position.

2. Overheating in the Beta-Phase above A_f

Most Cu-based alloys and some Ni-Ti alloys have to be rapidly cooled after the betatizing treatment at high temperatures, the reason being that the beta-phase is thermodynamically not the most stable phase. For example, Cu-based alloys will tend to form the equilibrium α and/or γ -phases during cooling. However only the beta-phase should be retained in order to obtain a good reliable shape memory effect. At sufficiently low temperatures and above a critical cooling rate the beta-phase can be obtained in a metastable condition. Around room-temperature, the driving force for decomposition into phases with a lower free energy is so small that the beta-phase can be treated as a stable single phase. At higher temperatures, diffusion is possible and the incubation time for decomposition becomes very small. This time is, in the present context, called the "overheating time". It is the time that a shape memory alloy can be stored at a temperature before other phases precipitate which deteriorate the memory behavior (such as bainite, α , γ or others, depending on the system). Overheating times can be important in some devices: for example circuit breakers should withstand several minutes at say 200°C or even 300°C without losing the memory effect (and this for a certain number of cycles). It is then important to know the cumulative time of all the overheating events since this time should be less than the total allowed overheating time. It should also be noted that the precipitation of the

equilibrium phases will deplete or enrich the original beta-matrix with solute elements which may shift M_s . Cu-Al-Ni alloys have overheating times of about 10 to 60 minutes at 300°C and 2 to 14 days at 200°C, though this is very much dependent on the exact composition. Overheating times for Ni-Ti can be several hours even at 400°C.

3. Stabilization of the Martensite Phase

Martensite stabilization is the result of aging in the martensitic phase and has been observed in several Cu-based alloys, but especially in Cu-Zn-Al. The most important macroscopic effect is the increase of the reverse transformation temperatures A_s and A_f . The process is thermally activated, which means that the increase in A_s and A_f occurs faster with higher aging temperatures. However the shift is not dramatic when the material has previously received an appropriate heat treatment. For example, aging at room temperature shifts A_s and A_f only 2 to 4 degrees per year. Aging at 60°C, however, gives a shift of about 15 degrees per year. Of course, this means that the material is stored without interruption or shape memory action at that particular temperature. Fortunately, this shift in A_s and A_f will immediately disappear as soon as the material is heated above the shifted A_s temperature. Thus, the effect becomes relevant only when the material is kept for a very long time just below A_s and the shape memory action must occur at a critical predefined temperature. Many applications where the material is stored for long times are situated in the area of fire protection. In those cases, the reaction time of the element will not be significantly influenced by a shift in A_s of even 10 to 20 degrees relative to a typical nominal value of 80°C. In control devices, however, the nominal A_s can be critical so that regular testing and thus annealing-out the A_s -shift may be required (or other more expensive shape memory alloys less prone to stabilization such as Cu-Al-Ni or Ni-Ti must be used).

On other occasions stabilization can provide an interesting opportunity for postponing the reverse transformation. For example, a predeformed coupling with a nominal A_s below room temperature can be stabilized, then stored at room temperature and thus economically transported. Once put into use by heating above the shifted A_f the stabilization effect disappears and the coupling remains in the stronger austenitic state even after cooling to room temperature. This will be discussed in more detail in a later paper.

Concurrent with an increase in A_s , other physical properties of the martensite change: hardness increases, damping capacity decreases and electrical resistance decreases. The kinetics as well as the degree of change of all these aspects is strongly influenced by the concentration of quenched-in vacancies: changes are greater and occur faster as the number of vacancies increases. Therefore, the thermal treatment should be chosen to minimize the concentration of vacancies in the martensite if one wishes to minimize stabilization effects. A maximum number of vacancies is obtained by quenching the material from a high temperature in the beta-phase directly to below M_f ; for alloys transforming above room temperature, quenching in ice-water will have this effect. As a consequence of the high quenched-in concentration of vacancies, the material will stabilize so dramatically that no reverse transformation will be observed after aging for several hours at room temperature. Moreover, the hardness and material strength will be doubled (though the material shows significant spring-back during unloading (see Figures 2 and 3). This property could provide an interesting aspect for energy storage or superelastic applications.

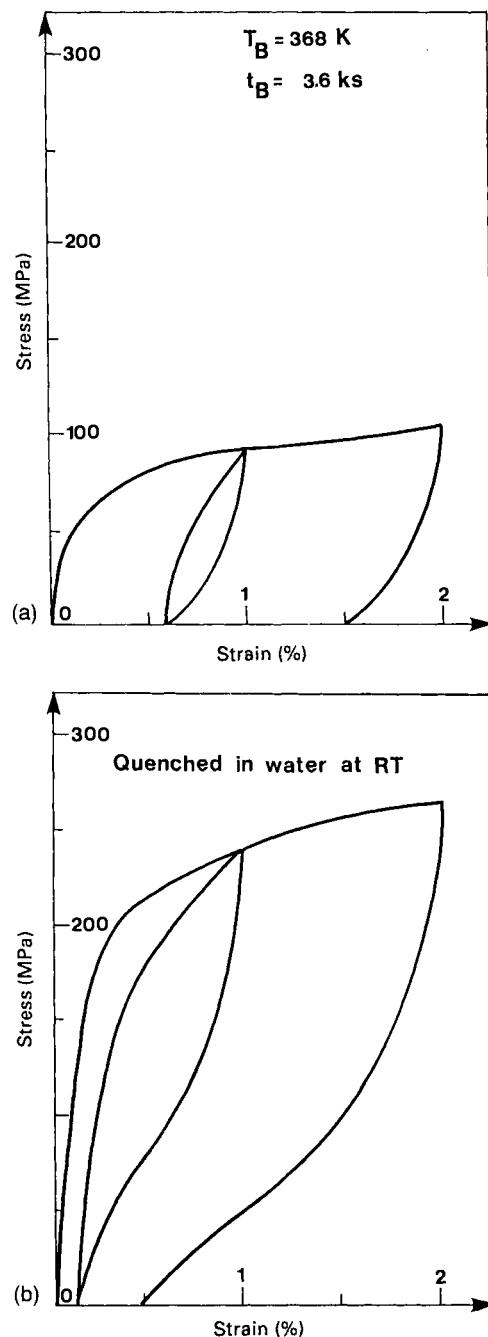


Figure 2 Stress-strain behavior: (a) after quenching in water at room temperature and annealing for one hour at 95°C and (b) after quenching in water at room temperature.

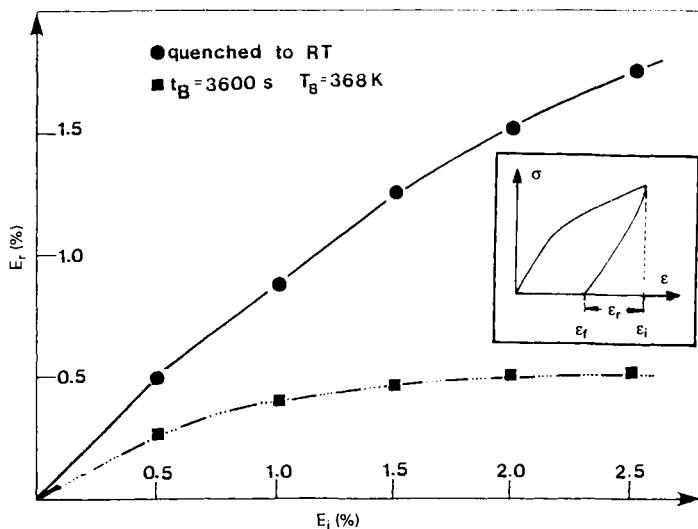


Figure 3 Springback as a function of prestrain.

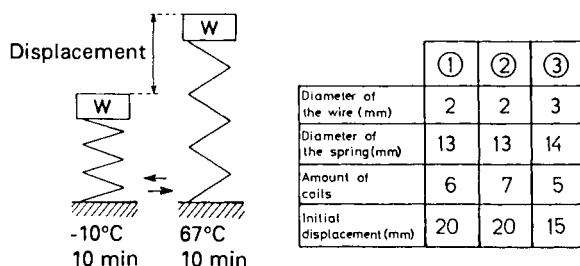
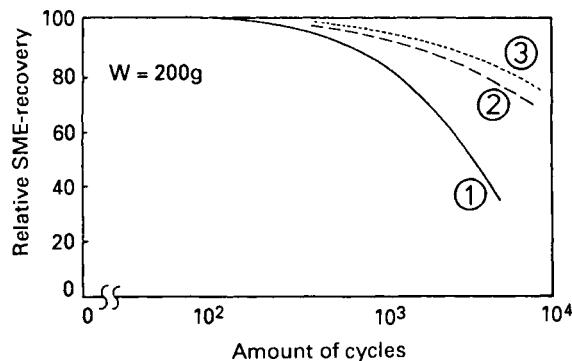


Figure 4: Degradation of the two-way shape memory effect.

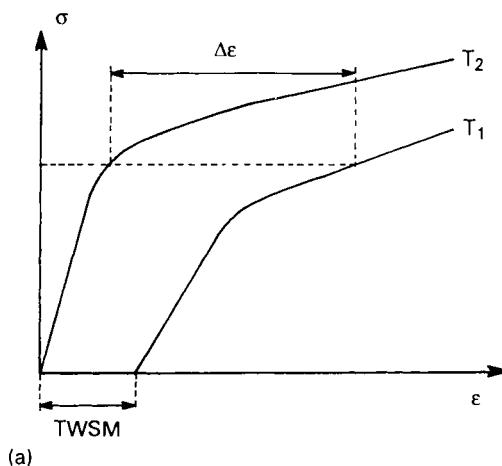
4. Degradation of the Two-Way Shape Memory Effect

Once a material is trained for two-way shape memory the question of its lifetime, generally expressed by the number of cycles, is put forward. It is, however, one of the least explored properties, probably because of the influence of the many parameters involved and their interactions (preload, temperature, shape memory deformation, alloy composition and alloy system).

The memory effect can degrade in both directions: on the spontaneous shape change during the reverse transformation and on the trained shape change during the forward transformation. Generally, degradation has been observed during heating because of incomplete transformation due to the stabilization of specific variants. Since stabilization is enhanced by stress, the degradation will occur faster when the memory element has to supply work during heating. Also the greater the applied shape memory strain, the faster the degradation will occur. These factors are summarized in Figure 4, demonstrating the effect on a spring. In sequence of increasing sensitivity to degradation, shape memory alloy system can be classified as: Ni-Ti, Cu-Al-Ni and Cu-Zn-Al.

5. Differences between Design and Actual Behavior

A set of isothermal loading curves (Figure 5a) is often used for the design of SME elements. The corresponding thermomechanical design path (Figure 5b) means loading at temperature T_1 , unloading, then heating to T_2 and again loading and unloading. The displacement $\Delta\epsilon$ is used as one of the design parameters.



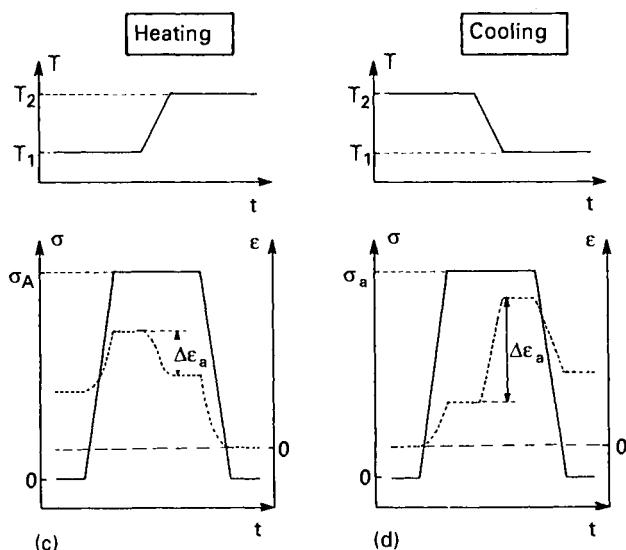
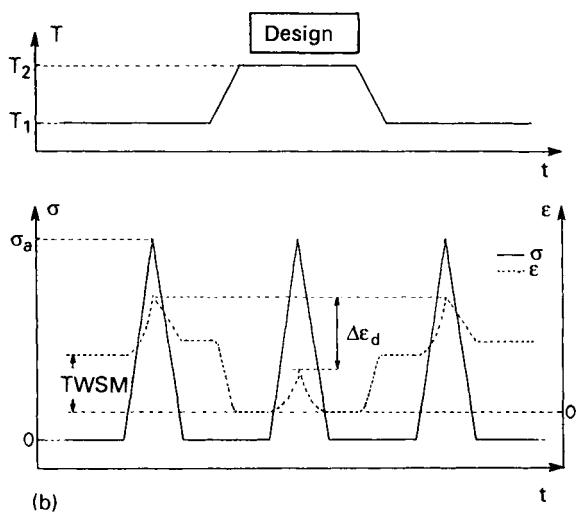


Figure 5: Differences between design and actual behavior, e.g. dead load: (a) design methodology, (b) "design" thermomechanical path, (c) thermomechanical path followed during heating, and (d) thermomechanical path followed during cooling.

Deviations between the actual ($\Delta\epsilon_a$) and the "design" displacement ($\Delta\epsilon_d$) are due to:

- (a) the difference between the actual and designed thermomechanical paths followed during heating and cooling
- (b) training effects
- (c) time dependent phenomena such as transformational creep
- (d) cycle dependent phenomena such as degradation due to thermal cycling
- (e) hysteresis effects

The "design" stress-strain-temperature path followed during heating is quite different from the path followed during use of the element (e.g. dead load, Figures 5c and 5d). Since the variables (stress, strain and temperature) are not state functions, deviations between the designed and actual behaviors can be expected. In the case of a dead load, the restraining stress retards the progress of shape recovery during heating. This results in small actual displacements ($\Delta\epsilon_a < \Delta\epsilon_d$). The deviation ($\Delta\epsilon_d - \Delta\epsilon_a$) becomes larger the lower the temperature difference between T_2 and A_f and the greater the dead load.

The $\sigma-\epsilon$ path followed during cooling from T_2 to T_1 (Figure 5d) results in even larger deviations, especially when using elements without two-way memory. The actual displacement is larger, due to the preferential growth of thermoelastic martensite under stress. The larger the dead load and the lower the two-way memory, the larger the deviation. The isothermal loading curves result from tensile or torsion tests on untrained samples. The two-way memory training causes a softening of the austenitic sample. Until now, the relationship between softening and two-way memory is unknown.

The large transformation creep rates within the transformation region are linked to the ongoing transformation and occurs even at very low loads. The effect is strongly dependent on previous thermomechanical history and can even lead to negative creep strain rates.

6. Conclusions

Designing with shape memory alloys is presently receiving increased interest from industry and many successful products have already been presented. However very often the use of shape memory alloys has been rejected due to the ignorance of the instabilities. For successful application and design using shape memory alloys, complete information has to be supplied regarding the environmental and working conditions of the device as well as regarding the behavior of the memory element under those conditions. In this way, proper design can be optimized by adjusting the memory element to the requirements as well as by adjusting the device to the memory element.

Acknowledgements

J. Van Humbeeck and R. Stalmans acknowledge the Belgian National Science Foundation for financial support.

References

1. Duerig, T.W., Albrecht, J., Gessinger, G.H.: *J. of Metals*, 12, 14 (1982).
2. Van Humbeeck, J., Janssen, J., Mwamba-Ngoie, Delaey, L.: *Scripta. Met.* 18, 893 (1984).

3. Verguts, H., Aernoudt, E., "Monitoring the response of shape memory alloys along thermomechanical paths in transformation space", Proc. of the 7th Int. Conf. on the Strength of Metals and Alloys, Pergamon Press, Montreal 12-16 August, 563-568 (1985).
4. Elst, R., Van Humbeeck J., Delaey, L.: Mat. Science and Technology, 4, 644 (1988).
5. Ortin, J., Planes, A.: Acta Metal., 36, 1873-1889 (1988).

Measurable Changes Concomitant with the Shape Memory Effect Transformation

J.D. Harrison, Raychem Corporation, Menlo Park, CA

Recovery of deformation on heating and pseudoelasticity are the most noted occurrences associated with shape memory alloys, but a host of other changes occur simultaneously. Before the recognition of SME in gold-cadmium, Ölander¹ reported a rubbery feel in the alloy. This turns out to be an apt description of SME alloys within the transformation range. Buehler's^{2,3} curiosity was aroused by the observation that warm nickel-titanium gave a sharp ring when struck whereas the cooled alloy gave a dull thud. For a time the change in the tone of the ring was used to measure the transformation range in Ni-Ti. A few years later Muzzey⁴ heated deformed samples of nickel-titanium and revealed the SME. Such techniques still serve as the rudimentary tools for the exploration of SME. If the sample has a dull thud, deform it, then heat to see whether the deformation is recovered over some temperature range. If the sample rings when struck, cool the sample under an elastic flexial load. If the sample collapses into a bent configuration, heat and note the temperature of recovery.

1. Property Changes

Table I lists a number of characteristics which change in SME alloys. To date, the best developed measuring techniques are those based on the SME itself: observation of elongation during cooling under a tensile load and recovery during subsequent heating.⁵ The direct applicability of the resulting data assures this will always be a major test method. However, an unavoidable question is how deformation modifies the transformation itself in the course of the test cycle. A sublime early study skirted

Table I: Measurable Changes Accompanying the SME Transformation

- Feel
- Ring
- Bend
- Color
- Roughness
- Hardness
- Expansion
- Yield Strength
- Young's Modulus
- Damping
- Internal Friction
- Velocity of Sound
- Acoustic Emission
- Electrical Resistance
- Magnetic Properties
- Electromotive Force
- Thermoelectric Power
- Hall Coefficient
- Heat Capacity
- Latent Heat of Transformation
- Thermal Conductivity
- Lattice Spacing
- Electron Density Waves

this problem handily by measuring Rockwell hardness at successive temperature increments on a single piece of Ni-Ti during cooling after annealing.⁶ Sub-annealing-temperature heating and cooling cycles followed. Hardness showed subtle changes in the transformation as a function of repeated thermal cycles. Hardness has also been an advantage for quality control.

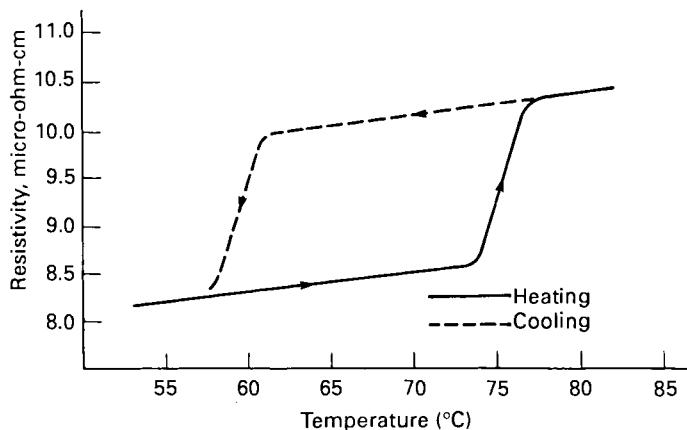


Figure 1: Resistivity- temperature curve for single crystal Au-Cd alloy containing 47.5 atomic pct Cd⁶.

Changes of electrical resistance versus temperature⁷⁻¹⁰ have proven to be a major tool for monitoring the transformation either in the absence or presence of deformation. The classic resistivity measurement on the cadmium-gold transformation was made by Chang and Read, Figure 1.11 Similar results are found from resistivity testing of copper-base SME alloys, thus resistivity is the usual transformation measuring technique for these alloys. With nickel-titanium, resistivity is useful but the results are a more complex function of both alloy composition and processing. Figure 2 shows frequently recurring patterns for various nickel-titanium base alloys and how the patterns can change with post-anneal heat treatment. These changes in electrical resistance have been put to practical use by providing feedback to control the motion of robotic arms.

Measurement of the latent heat of transformation¹²⁻¹⁶ made by Differential Scanning Calorimetry, DSC, share the advantage of electrical resistance in that stress and deformation need not perturb the transformation during the test cycle. Elongation and resistivity record transformation in cumulative and semi-cumulative fashion, respectively, Figure 3a&b, whereas the thermal data from the DSC highlights the discrete events within the transformation range as shown in Figure 3c. Samples for DSC can be small (10 milligrams) and require minimal sample preparation. Residual stress from cold work, etc. cause the DSC peaks to be diffuse so samples often require annealing after cutting and before testing.

The other changes provide opportunities for products, for feedback signals, for quality control measurements, as well as for understanding of the SME transformation itself. Some copper-base shape memory alloys show a noticeable color change with the transformation;¹⁷ a concept for a computer memory disc uses this color change. Macro- or microscopic surface roughening or platelet formation accompany many of the SME transformation permitting direct microscopic observation. Thermal expansion

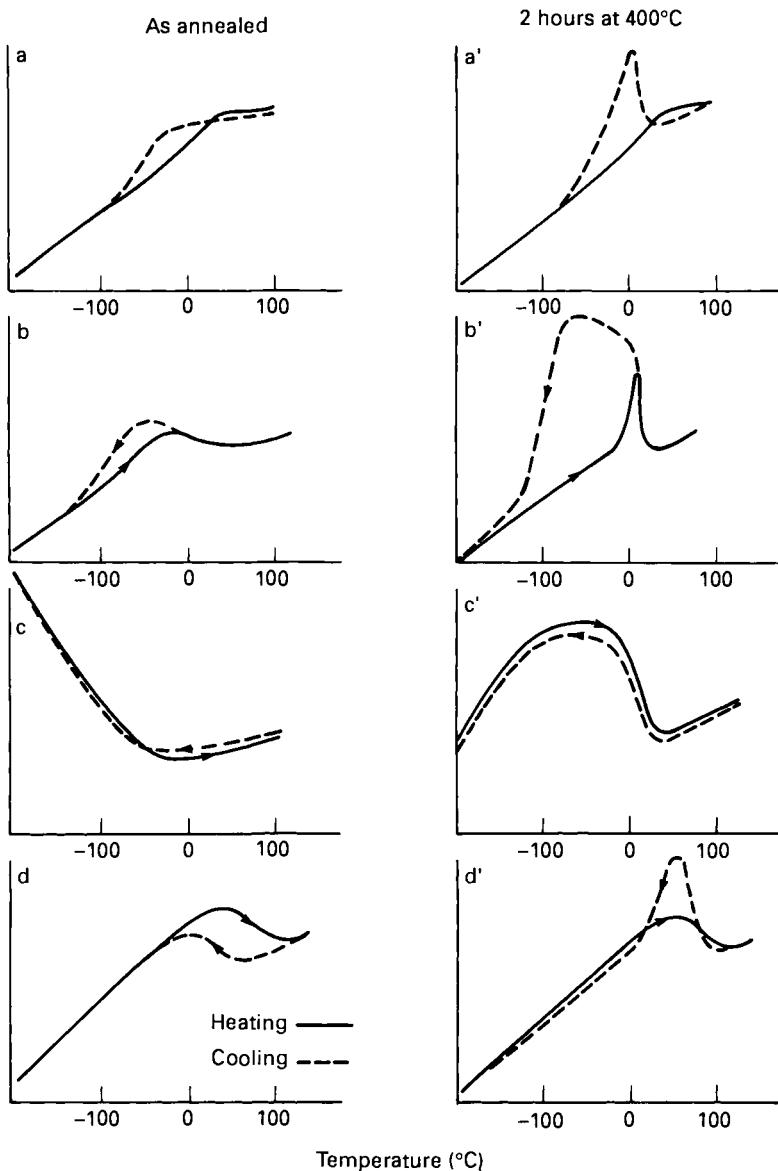


Figure 2: Characteristic resistance versus temperature curves for unstable Ni-Ti alloys: a-d, as annealed; a'-d', after holding at 400°C for two hours.

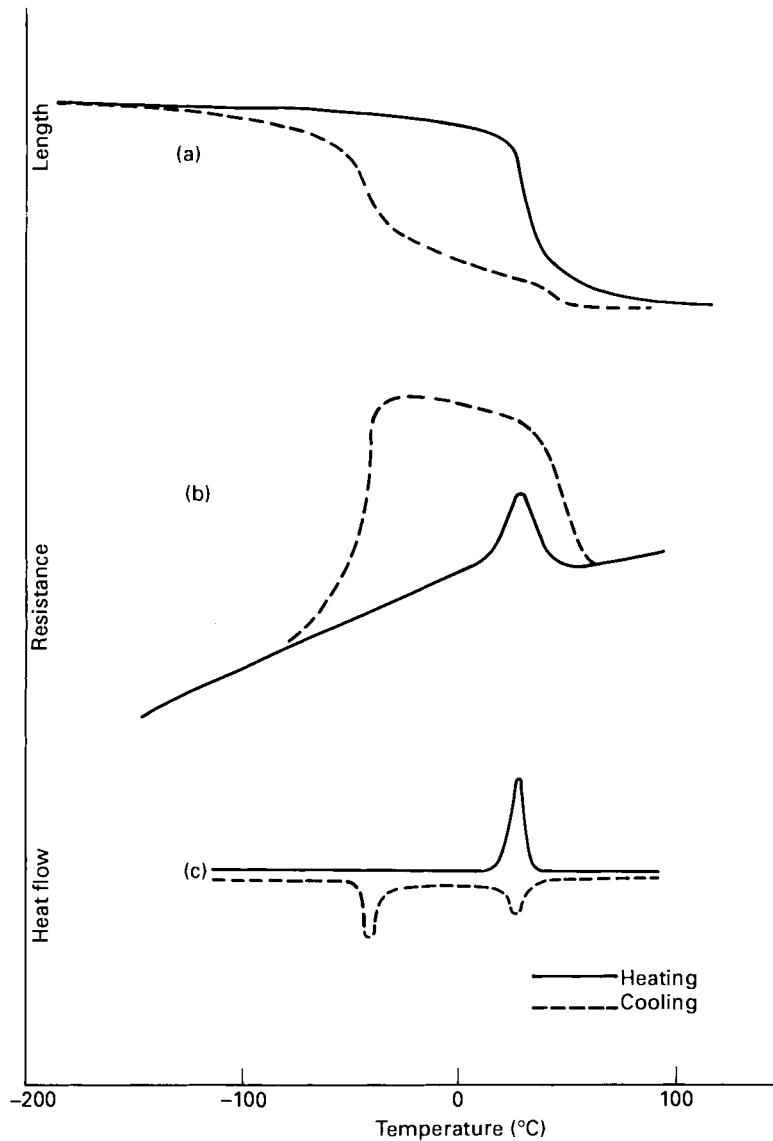


Figure 3: Mechanical, electrical, and DSC measurements versus temperature for Ni-Ti sample: (a) length vs. T under a 70 MPa stress; (b) resistance vs. T under no load; and (c) heat flow vs. T under no load.

traces the martensitic change in some alloys.^{2,12,18,19} Yield strength and Young's modulus both go through minima near M_s^{20,21}. Damping^{2,21}, the acoustic ring^{2,3}, and internal friction²² all track the SME, as does the velocity of sound^{20,23}. Acoustic emission²⁴ accurately monitors the transformation in cadmium-gold, but not at all in nickel-titanium. Magnetic property changes^{7,12,25} parallel the SME in some alloys. Electromotive force changes^{26,27} have been measured during the transformation of nickel-titanium. Thermoelectric power^{7,12} and changes in the Hall Coefficient^{7,25,28,29} have also been used to monitor the transformation range. The SME can be traced by changes in thermal conductivity.^{30,31} Crystallographic measurements ranging from simple lattice spacing via X-ray³²⁻³⁷, electron or neutron diffraction^{38,39} on to electron density wave observation³⁸ all give valuable information about SME.

3. Conclusion

The obverse of the opportunities provided by the host of changes accompanying the SME is the frustration of simple goals caused by complex interactions. For both product development and for ongoing quality control, prudence calls for the use of test methods which parallel the actual function of the product in addition to the primary testing techniques mentioned above.

References

1. Ölander, A., J. Am. Chem. Soc. **54**, 3819, 1932
2. Buehler, W.J.; and Wiley, R.C., Rept. NOLTR 61-75 (AD 266607), U.S. Naval Ordnance Laboratory, 1961
3. Buehler, W.J.; Gilfrich, J.V.; and Wiley, R.C., J. Appl. Phys., **34**, 1475, 1963
4. Hartmann, G.K., Symposium on Ni-Ti and Associated Compounds. NOLTR 68-16, U.S. Naval Ordnance Laboratory, 1968
5. Eckelmeyer, K.H., Scripta Met., **10**, No.8, 667, 1976
6. Mukherjee, K.; Milillo, F.; and Chandrasekaran, M., Materials Science and Engineering, **14**, 143, 1974
7. Hanlon, J.E.; Butler, S.R.; and Wasilewski, R.J., Transactions of the Metallurgical Society of AIME, **239**, 1323, 1967
8. Cross, W.B.; Kariotis, A.H.; and Stimler, F.J., NASA Cr-1433, 1969
9. Cooper, J.E.; Bowker, D.E.; and Cross, W.B., 'Materials and Processes for the 1970's, Proceedings of the 15th Annual Symposium of the Society of Aerospace Material and Process Engineers, 1969
10. Marques, M.C.; Sousa, J.B.; Pinheiro, M.F.; and Braga, M.E., Scr. Metall., **11**, 197, 1977
11. Chang, L.C.; and Read, T.A., Transactions of AIME, **191**, 47, 1951
12. Wasilewski, R.J.; Butler, S.R.; and Hanlon, J.E., Metal Sci. J., **1**, 104, 1967
13. Cheng, C.H.; Wei, C.T.; and Beck, P.A., Phys. Rev., **120**, 426, 1960
14. Wang, F.E.; Buehler, W.J.; and Pickart, S.J., J. Appl. Phys., **36**, 3232, 1965
15. Dautovich, D.P.; Melkvi, Z.; Purdy, G.R.; and Stager, C.V., J. Appl. Phys., **37**, 2513, 1966
16. Berman, H.A.; West, E.F.; and Rozner, A.G., J. Appl. Phys., **38**, 4473, 1967
17. Busch, R.E.; Luedeman, R.T.; and Gross, P.M., Final Report-Part 1. ARMA CR65-02/1, U.S. Army Material Research Agency, 1966
18. Spinner, S.; and Rozner, A.G., J. Acoust. Soc. Am., **40**, No.5, 1009, 1966
19. Buehler, W.J., Proceedings of the Seventh Navy Science Symposium, Rept. ONR-16, Vol.1 (AD 421708), 1-30, 1963
20. Rozner, A.G.; and Spinner, S., J. Acoust. Soc. Am., **40**, 1000, 1966
21. Wasilewski, R.J., Transactions of AIME, **233**, 1691, 1965

22. Hasiguti, R.R.; and Iwasaki, K., *J. Appl. Phys.*, **39** No.5, 2182, 1968
23. Bradley, D., *J. Acoust. Soc. Am.*, **37**, No.4, 700, 1965
24. Mints, R.I.; Memekhin, V.P.; Ievler, I.U.; and Bukhalenkov, V.V., *Fiz. Tverd. Tela*, **14**, No.5, 1582, 1972
25. Wang, F.E.; DeSavage, B.F.; Buehler, W.J.; and Hosler, W.R., *J. Appl. Phys.*, **39**, 2166, 1968
26. Wang, F.E.; and Buehler, W.J., *Appl. Phys. Lett.*, **21**, No.3, 105, 1972
27. Johnson, A.D.; and Katz, P.I., *J. Appl. Phys.*, **48**, No.1, 73, 1977
28. Wang, F.E.; DeSavage, B.F.; Bernard, F.; Buehler, W.J.; and Hosler, W.R., *J. Appl. Phys.*, **39**, No.5, 2166, 1968
29. Allgaier, R.S., *J. Phys. chem. Solids*, **28**, 1293, 1967
30. Goff, J.F., *J. Appl. Phys.*, **35**, 2929, 1964
31. Goff, J.F., *J. Appl. Phys.*, **39**, No.5, 2208, 1968
32. Duwez, P.; and Taylor, J.L., *Transactions of AIME*, **188**, 1173, 1950
33. Poole, D.M.; and Hume-Rothery, W., *J. Inst. Metals*, **83**, 473, 1954
34. Purdy, G.R.; and Parr, J.G., *Transaction of AIME*, **221**, 636, 1961
35. Taylor, A.; and Floyd, R.W., *J. Inst. Metals*, **80**, 577, 1952
36. Pietrokowsky, P.; and Youngkin, F.G., *J. Appl. Phys.*, **31**, 1763, 1960
37. Marcinkowski, M.J.; Sastri, A.S.; and Koskimaki, D., *Phil. Mag.*, **18**, No.155, 945, 1968
38. Salamon, M.B.; Shapiro, S.M.; Meichle, M.; and Wayman, C.M., *Internation Conference on Modulated Structures*, Hawaii, 1979
39. Golestaneh, A.A.; and Carpenter, J.M, 'Study of the Martensitic Transformation in Shape-Memory Ni-Ti Alloy by Time-of-Flight Neutron Diffraction Techniques', submitted to *Acta Met.*, 1988

The Mechanical Aspects of Constrained Recovery

J.L. Proft and T.W. Duerig
Raychem Corporation
300 Constitution Drive
Menlo Park, CA 94025 USA

The most widely exploited uses of shape memory alloys are those in which an external constraint prevents the alloy from returning to its original shape on heating. This event is commonly referred to as *constrained recovery*, and can result in the generation of high recovery forces. It is the foundation for some of the oldest and most successful applications of materials exhibiting the shape memory effect.

If a shape memory alloy is deformed in its martensitic state, all or a portion of the imparted strain can be recovered by heating. If nothing interferes with this process, the event is referred to as *free recovery*. Though there are few engineering applications of free recovery, the event provides useful insight into constrained recovery. The process of isothermal deformation of martensite and subsequent heating to form austenite will therefore first be explored in the absence of any constraint. Two profiles (stress-strain and strain-temperature) are useful in understanding and describing the free recovery event.

In order to completely analyze and apply constrained recovery, a third profile, stress-temperature, must be introduced. The parameters of all three profiles can then be examined in terms of their relationships to one another. Of particular importance are the dependencies of recovery stress on total deformation strain, contact strain, and compliance of the external constraint. In conjunction, the recovery stress turns out not to be a state function, but depends critically on the path taken during the recovery process.

The implications of the above relationships on the design of useful components will be emphasized in this chapter, along with data for both Ni-Ti based and Cu based alloys. A brief overview of the experimental techniques and considerations for obtaining such data, as well as assessing the capabilities of all of these materials will be included.

1. Measurement Methods

Except where otherwise noted, data presented here were obtained using specimens of a 47%Ni, 50%Ti and 3%Fe (all by atomic percent) alloy machined from hot swaged bar and annealed at 850°C. The M_s temperature of the alloy was measured using mechanical techniques to be -125°C at 280 MPa. Deformation and recovery was conducted using a closed-loop servo-hydraulic test machine equipped with a chamber capable of controlling temperatures between -196 to 300°C. All Ni-Ti-Fe specimens were deformed at -196°C.

Tensile experiments are useful to emulate and quantify the constrained recovery event. In the case of free recovery experiments, specimens were deformed in strain control using an extensometer at a strain rate of 0.001 sec^{-1} and then unloaded in load control. Zero load was maintained during heating while continuously monitoring strain and temperature. For constrained recovery, specimens were deformed and unloaded in the same way, allowed to freely recover to a selected strain, then constrained by shifting to a strain control mode and maintaining a "constant" strain. The term "constant", however bears further comment. In fact, the computer was constantly correcting the "fixed" strain to remove apparent stress changes due to differential thermal expansion ($\Delta\alpha$) and temperature strain errors in the extensometer (ζ). The apparent stress build up ($\Delta\sigma$) that would be created in a static, non-correcting constraint system (i.e. simply fixing the cross-heads of an Instron machine) can be extremely large (equation 1) making simple static constraint very inaccurate.

$$\Delta\sigma = E\Delta\varepsilon = E(\Delta\alpha + \zeta)(\Delta T) \quad (1)$$

2. Free Recovery

For the free recovery event, the initial deformation of the martensite is adequately described by the stress-strain profile as shown in Figure 1. Recovery is better

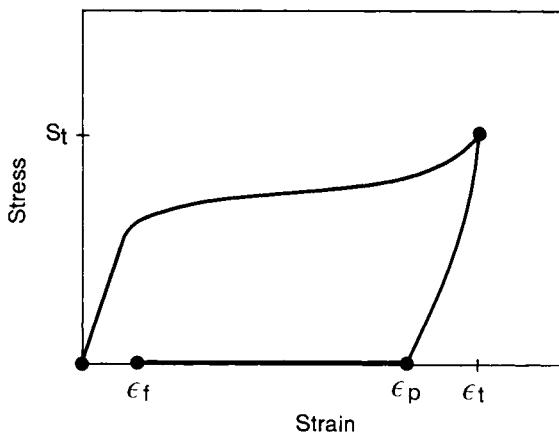


Figure 1: The stress-strain perspective of the free recovery event, showing deformation to ε_t , unloading to ε_p , and recovery to ε_t .

understood by examining the associated strain-temperature profile shown in Figure 2. Important parameters of the stress-strain perspective include the *total strain* imparted to the specimen (ε_t), the apparent *plastic strain* upon unloading (ε_p), and the *final strain* after heating (ε_t). Additional parameters introduced in the strain-temperature profile are the *deformation temperature* (T_d) and the austenite start and finish temperatures (A_s' and A_f'). Note that a distinction is made here for recovery temperatures after deformation as opposed to a thermal transformation in the absence of imparted strain (i.e. A_s and A_f). Note that the shift is not caused by stress but by the plastic strain.

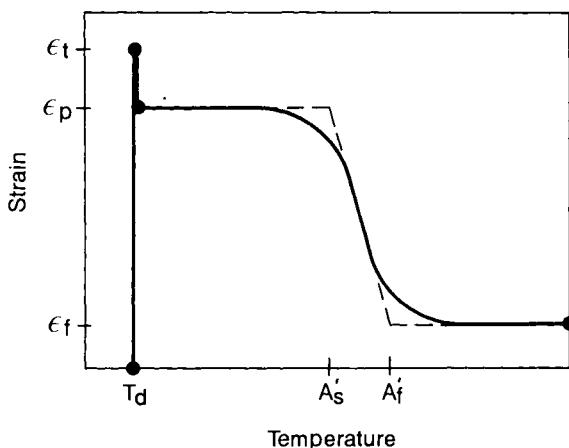


Figure 2: The strain-temperature perspective of the free recovery event, showing deformation, unloading, and then heating, with recovery beginning at A'_s and ending at A'_f .

The free recovery behavior of an alloy can be characterized by looking at the influence of ϵ_t on both the amount of recoverable strain (ϵ_r), defined as the difference ($\epsilon_p - \epsilon_f$), and on the recovery temperature (A'_s). The data is obtained from a series of specimens deformed to various total strains at T_d ($T_d < M_s$), unloading, and then heating while maintaining zero load. Stress-strain and strain-temperature are simultaneously recorded for each test. The recoverable strain and the recovery temperatures are extracted and plotted as a function of total strain. The results of such an experiment for the Ni-Ti-Fe alloy described in the previous section are shown in Figures 3 and 4 respectively.

It is evident from Figure 3, that as the total strain is increased, ϵ_p and ϵ_r also increase. The recoverable strain- ϵ_r , increases up to a maximum and then decreases as permanent slip begins to impede the reversible deformation accommodated by twin boundary movement. This incompleteness of recovery is often appropriately called the amnesia strain. For the designer, the maximum motion available and the deformation to achieve that maximum are usually of primary interest. The Ni-Ti-Fe alloy displays a maximum of approximately 8% motion (typical of most Ni-Ti based alloys). In other systems such as Cu-Al-Ni or Cu-Zn-Al, fracture occurs before the strain to achieve maximum recovery can be imparted due to their inherently low ductilities. The recovery strain in these materials is therefore generally limited to only 3 to 5%.

The recovery temperature (A'_s) may also be a heavily weighed alloy selection parameter and its dependence can also be quantified by the above tests (Figure 4). Obviously if one were more concerned with high recovery temperatures than motion, some gains could be made by using very large deformation strains. However, for reasons which will become evident later, the use of deformation strains to produce specific A'_s temperatures may not always be desirable if trying to optimize other responses.

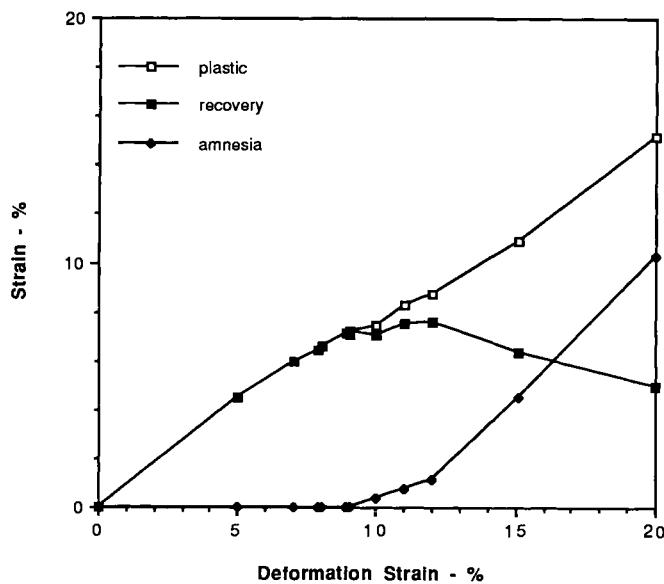


Figure 3: As shape memory alloys are deformed beyond the end of the stress-strain plateau, amnesia increases. In the Ni-Ti-Fe alloy shown above, perfect recovery occurs for deformations less than 8.5%, and the recovery strain reaches a maximum of over 7% after some 11% total deformation.

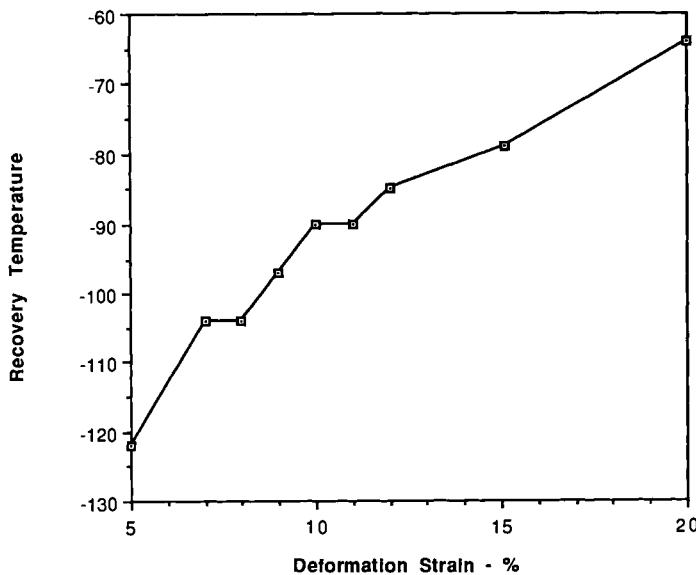


Figure 4: Again as one goes beyond the end of the stress-strain plateau, plastic deformation, or amnesia, begins to occur. This is accompanied by an increase in the recovery temperature, in this case defined as the average between A_s' and A_f' .

3. Constrained Recovery

Constrained recovery is best visualized by imagining a ring, expanded in the martensitic state then shrunk onto a solid shaft preventing complete recovery. The ring will freely recover until contact is made, then be rigidly constrained and generate a stress. The substrate may be deformed elastically or plastically depending on its own mechanical properties and the magnitude of the internal stresses developed in the shape memory element.

To quantify the constrained recovery event, the stress-temperature profile must be introduced in conjunction with the stress-strain and strain-temperature perspectives (shown in Figures 5 through 7). Four new parameters are introduced including the

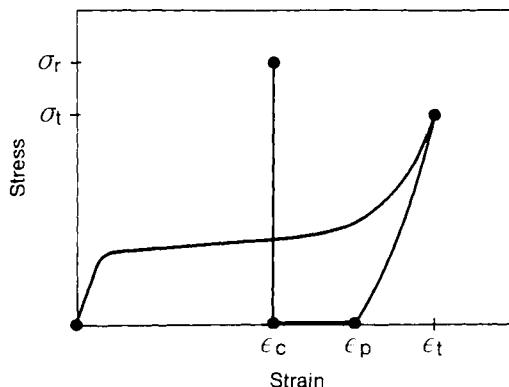


Figure 5: Similar to the case of free recovery (Figure 1), the constrained recovery event begins with deformation of the martensite, unloading and free recovery, but this time recovery is externally prevented and a stress- σ_r is generated.

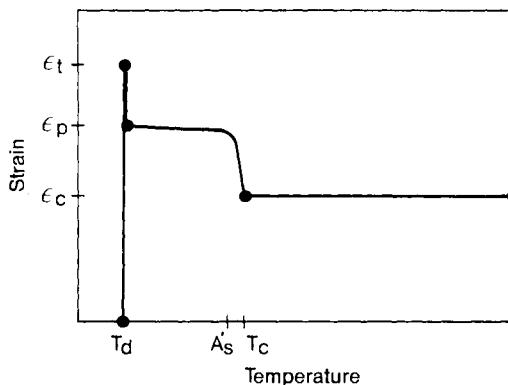


Figure 6: The strain-temperature perspective of the event shown in Figure 5 simply shows that recovery is prevented at the contact strain- ϵ_c .

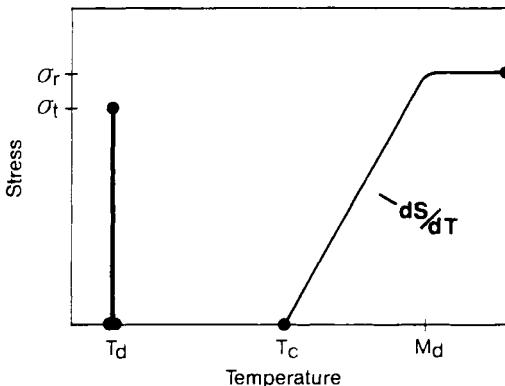


Figure 7: The stress-temperature perspective of Figure 5 shows that stress begins to build up at the contact temperature T_c , and then increases linearly from there until M_d is reached.

contact strain (ϵ_c), the contact temperature (T_c), the recovery stress (σ_r), and the stress rate ($d\sigma/dT$). The contact strain is the point at which the ring mentioned above contacts the shaft. Note that until this point, the process is no different than the free recovery event. *Unresolved recovery* is implicitly defined as $\epsilon_c - \epsilon_f$: the additional strain that could have been recovered had recovery been free. The contact temperature T_c is the temperature at which contact with the substrate is made. It is most easily viewed in the stress-temperature profile as the point at which a stress is first evident. This stress continues to increase as recovery proceeds up to the alloy's M_d . The recovery stress reaches a maximum value σ_r , at M_d . The rate at which the stress increases with temperature, $d\sigma/dT$, is also obtained from the stress-temperature profile. It is a constant for a given alloy in a given condition and can be derived theoretically from the Clausius-Clapeyron equation. Note that this is a third definition of the stress rate - in Chapter 1, the stress rate was shown in constant temperature and in constant stress environments. The constant strain definition described here is in fact the same as the other two. The martensitic transformation in Ni-Ti usually shows a $d\sigma/dT$ of between 4 and 20 MPa/ $^{\circ}$ C, with $d\sigma/dT$ decreasing as the austenitic strength is increased. The R-phase transformation typically has a much higher $d\sigma/dT$: between 30 and 70 MPa/ $^{\circ}$ C. Values for the Cu-based alloys tend to be lower, normally between 2 and 5 MPa/ $^{\circ}$ C.

3.1 Quantification of Constrained Recovery

The characterization of an alloy to be used in the constrained recovery mode usually includes determining the dependence of σ_r on ϵ_c and ϵ_r . For example, if a series of specimens are deformed to a fixed total strain and then allowed to recover to various contact strains, a plot of σ_r versus ϵ_c similar to that shown in Figure 8 would result. In this particular case, a Ni-Ti-Fe alloy was deformed to 9% total strain at -196 $^{\circ}$ C and then allowed to recover to different strains before fixing the strain at the levels shown. In general, the recovery stress increases with increasing contact strain and becomes fairly constant beyond a certain level of contact strain. In the example shown, designing with unresolved strains of less than 1.5% should be avoided to assure

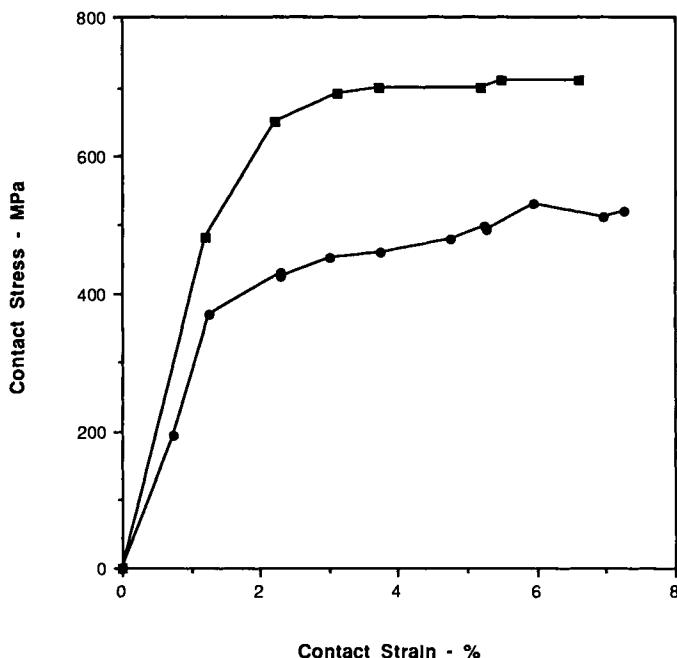


Figure 8: The stress generated during constrained recovery is quite constant if the contact strain, or unresolved recovery, exceeds 1.5%. Two Ni-Ti-Fe alloys are shown in this case: a fully annealed alloy (circles) and a hardened alloy (squares). The total deformation strain in this case was 9%.

consistent stress. This implies that although one can measure a certain amount of available motion from free recovery experiments (i.e. 8% for Ni-Ti), not all of that motion is useable if predictable or constant stress levels are required. This provides, still, huge advantages over many other joining techniques in that if one designs with at least 1.5% unresolved recovery, the stress is well defined - this sharply contrasts to torqued bolts where large variations in frictional coefficients are found, or thermally shrunk rings based on differential expansion where even very exact tolerances can allow large stress variations. Typical ranges of recovery stress for Ni-Ti based alloys as well as other alloy systems are listed in Table 1.

Table 1: Typical Values for Recovery Stress

Alloy	M _s (°C)	T _d (°C)	ε _r (%)	σ _r (MPa)
Ni-Ti	-200	-200	8.0	500-900
Cu-Al-Ni	+70	+25	2.0	300-400
Cu-Zn-Al-Mn	-40	-90	3.5	550-650

The second relationship which must be understood when designing with the constrained recovery mode is the dependence of recovery stress on total deformation strain. As was shown in the case of free recovery, there is an optimum deformation strain to achieve a maximum response from the alloy. For the case of constrained recovery, this is determined by varying the total strain and fixing the contact strain relative to the apparent plastic strain for a series of specimens. The results of such an experiment are shown in Figure 9 for the same Ni-Ti-Fe alloy deformed at -196°C. Here the contact strain was held constant at 2% less than the unloaded strain- ϵ_p . In this case, choosing a total deformation strain between 8 and 10% would maximize the recovery stress at approximately 500 MPa.

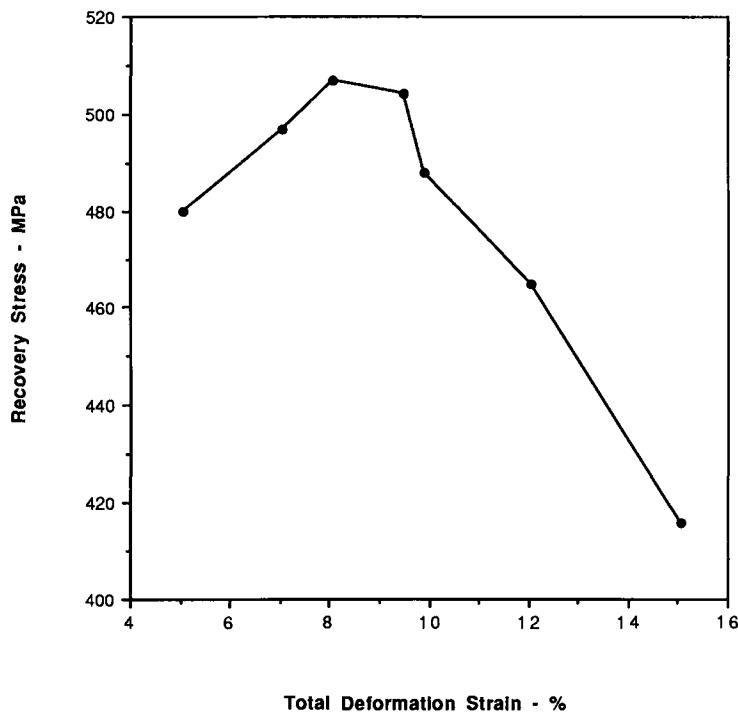


Figure 9: The recovery stress decreases as the total strain is increased (and the relative contact strain kept constant).

The above schematic representation of constrained recovery and actual results pertain to the case of a perfectly rigid substrate with the same coefficient of thermal expansion as the shape memory alloy. In reality, the substrate deforms during the process, either elastically or plastically and there is often a significant differential thermal expansion which alters the final stress obtained by an amount $\Delta\sigma = E' \Delta\alpha \Delta T$, where E' is a system compliance which accounts for the compliance of both the substrate and memory material. In order to predict recovery stresses in real applications, these effects must be taken into consideration. This can be achieved experimentally by imposing a stress-strain path which simulates substrate deformation and a strain-temperature relationship which accounts for the α mismatch.

In Figure 10, the recovery stress obtained from various linear elastic substrate paths is shown with compensation for any thermal expansion strains. It is evident that substrate compliance plays an important role in determining the final recovery stress with higher compliances resulting in lower recovery stresses. A comparison of these results to those obtained by using an the infinitely rigid substrate is shown in Figure

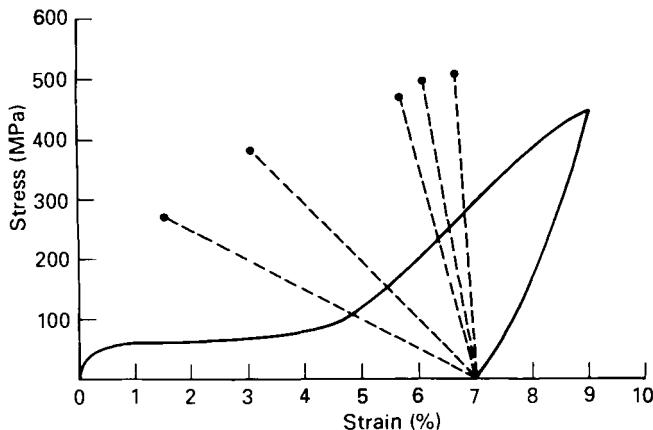


Figure 10: Most substrates are compliant, and thus some work is done during constrained recovery. For the same contact strain, lower compliances lead to lower recovery stresses.

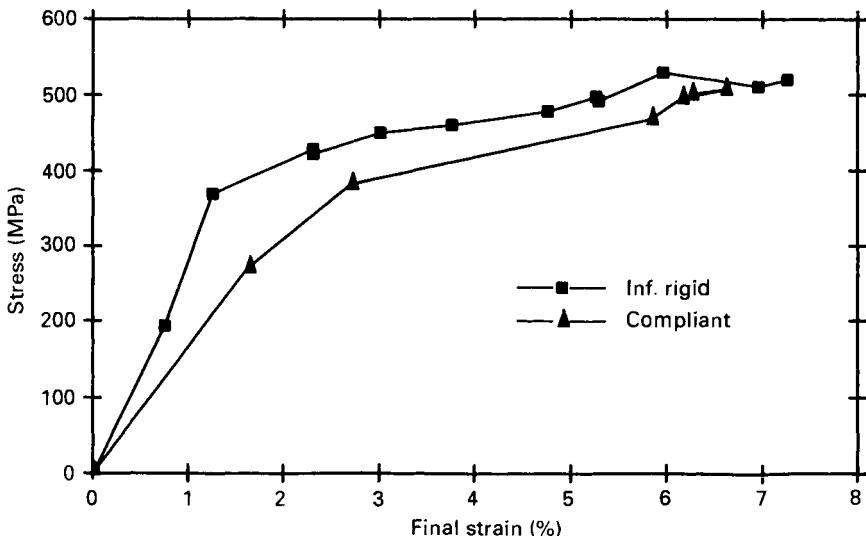


Figure 11: In comparing Figure 10 and Figure 8, one sees that just specifying the end strain (ϵ_f) is insufficient to determine the end stress (σ_r). By doing work in getting to the end point (recovery against a compliant substrate), the end stress is lower than if one freely recovers to that strain and then rigidly constrains.

11. It is clear from this comparison that recovery stress is a path dependent property: knowing only the final strain is not sufficient to predict the final stress.

It is interesting to note that the data points of either curve in Figure 11 resemble an austenite stress-strain curve. Several studies have shown that recovery stresses are typically 10 to 15% lower than the austenite flow stress values obtained from a traditional isothermal tensile test^{1,2}. A comparison of recovery stresses for typical contact strains with an austenite stress-strain curve obtained at -20°C for the Ni-Ti-Fe alloy is shown in Figure 12. It is interesting to contemplate the origin of this 10-15%

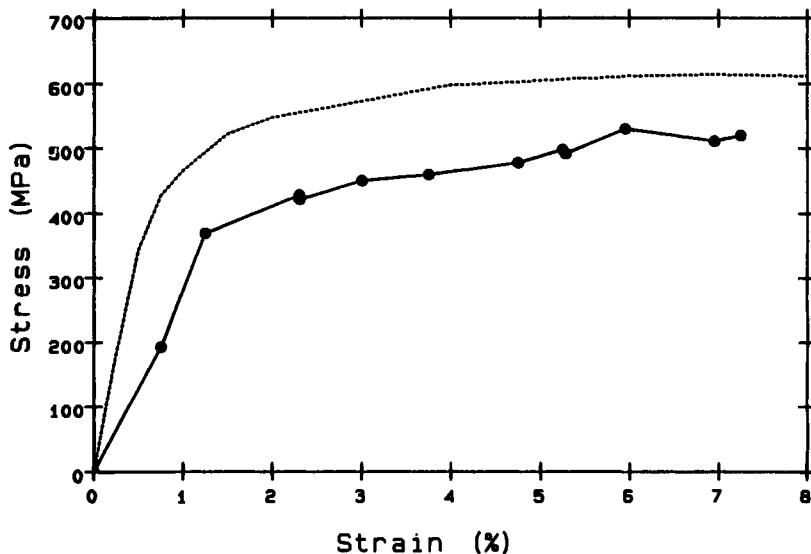


Figure 12: The envelope of recovery values plotted against the contact strain falls some 10-15% below a stress-strain curve of the same material.

difference. When reverting to austenite in a constrained situation, the strain originally accommodated by the martensite twins must be exchanged, and the recovery stress should be the stress needed to invoke irreversible plasticity. The mechanism could be any of the following:

- (1) The martensite deformation can be exchanged for slipped austenite, in which case we would expect coincidence in the curves of Figure 12.
- (2) The martensite could itself slip and then revert to austenite, in which case the constrained recovery curve should be coincident with the flow stress curve for martensite slip (which cannot be practically determined).
- (3) The martensite could revert to different variants of austenite, for which there is some evidence in that mechanical twins are found in the austenite after constrained recovery.

3.2 Characteristics of Joints Made by Constrained Recovery:

One very important aspect of any mechanical joint is the compliance of the interference stress at the completed joint: if the compliance is low, the joint is very

susceptible to stress loss from vibrations, corrosion, differential thermal expansion effects, etc. - the joint must be able to accommodate small geometrical shifts without losing stress. One advantage of Ni-Ti is its very low unloading modulus after constrained recovery (Figures 13 and 14). These moduli are less than 1/3 those of

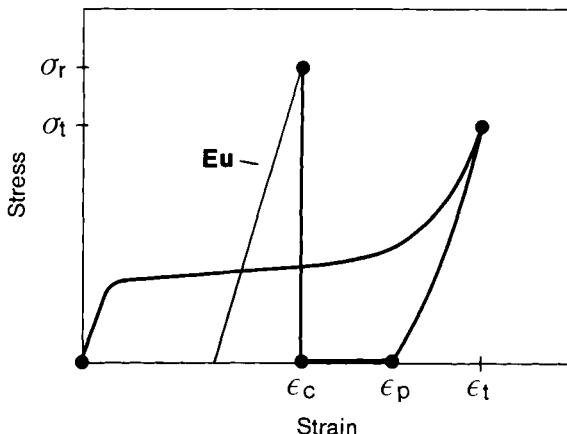


Figure 13: The *unloading modulus* measured after a constrained recovery is important in that it determines the compliance or the susceptibility of the joint to stress decay due to differential thermal expansion and small shifts in geometry.

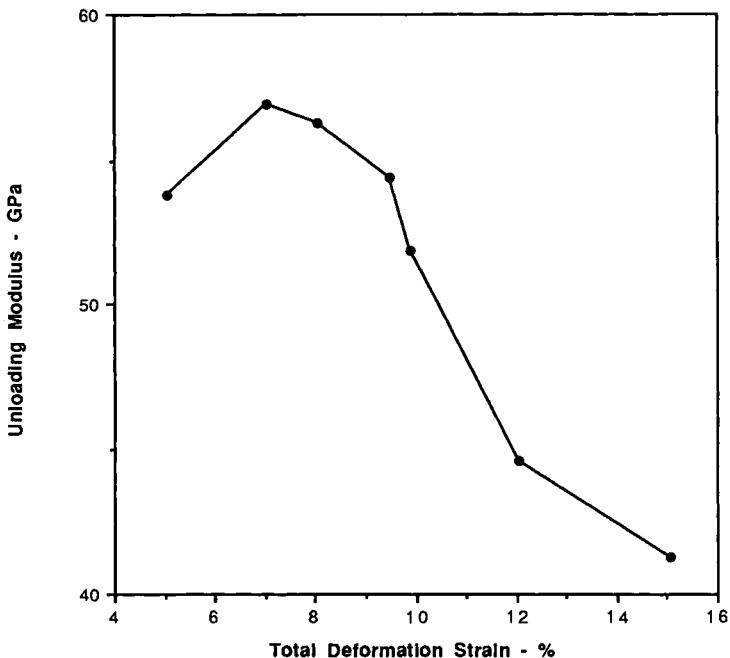


Figure 14: The unloading modulus of Ni-Ti-Fe is extremely low compared to most structural materials, and decreases as the total deformation strain increases.

steel. This provides a *live joint*, able to maintain stress over a wide range of temperatures and conditions.

A second interesting feature of Ni-Ti alloys in constrained recovery applications is that they are generally self-inspecting, meaning that any ring or coupling that survives expansion and recovery will survive in service. Figure 15 shows notch sensitivity as a function of temperature; there is a sharp minimum in notched ductility at approximately the M_d temperature. If we then compare the notch ductility with the recovery stress, and the expansion stress (Figure 16), we see that the critical temperature is reached upon heating through M_d . Thus material defects that may be harmful to service life will instead cause failure in expansion or recovery.

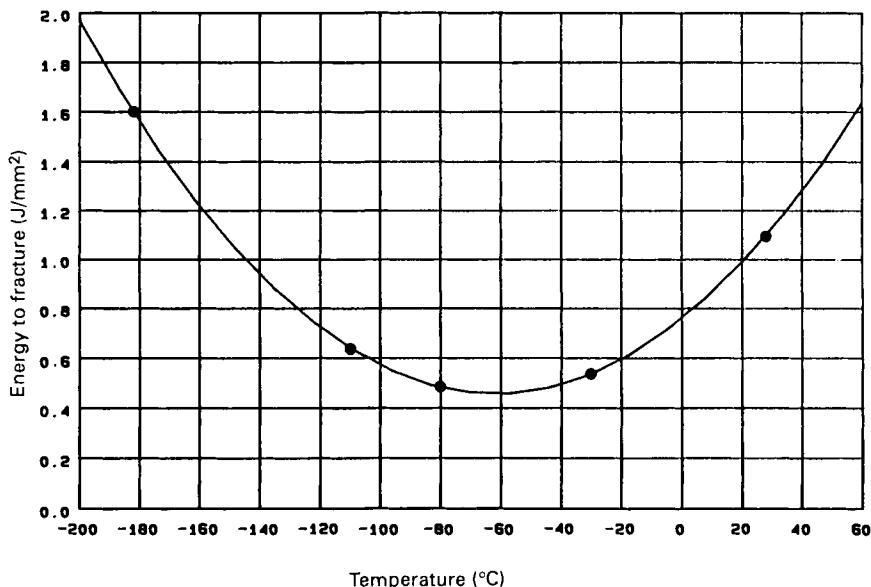


Figure 15: The toughness of Ni-Ti-Fe goes through a minimum with respect to temperature.

Finally, we look at the effect of ambient temperature on the performance of a joint made by shape memory. Two aspects must be considered: overheating and cooling. Overheating results in stress relaxation due to conventional creep processes, just as would occur in any other metal. Dynamically controlled stress relaxation tests have been conducted on a Ni-Ti-Fe alloy in order to quantify the upper service temperature limit; these tests are conducted in a creep rig equipped with an extensometer, and using a microprocessor to automatically shed the load as creep occurs, thus keeping strain constant. An example is shown in Figure 17. Conducting tests such as these over a range of temperature leads to the envelope behavior shown in Figure 18. Stress relaxation was found to follow the following behavioral law:

$$\frac{\sigma}{\sigma_0} = 1 - K \log\left(\frac{t}{t_0}\right) e^{-Q/RT} \quad (2)$$

where σ_0 and t_0 are the initial values stress and time, and σ and t are the values after exposure to temperature-T. With t and t_0 expressed in hours, the constant K was

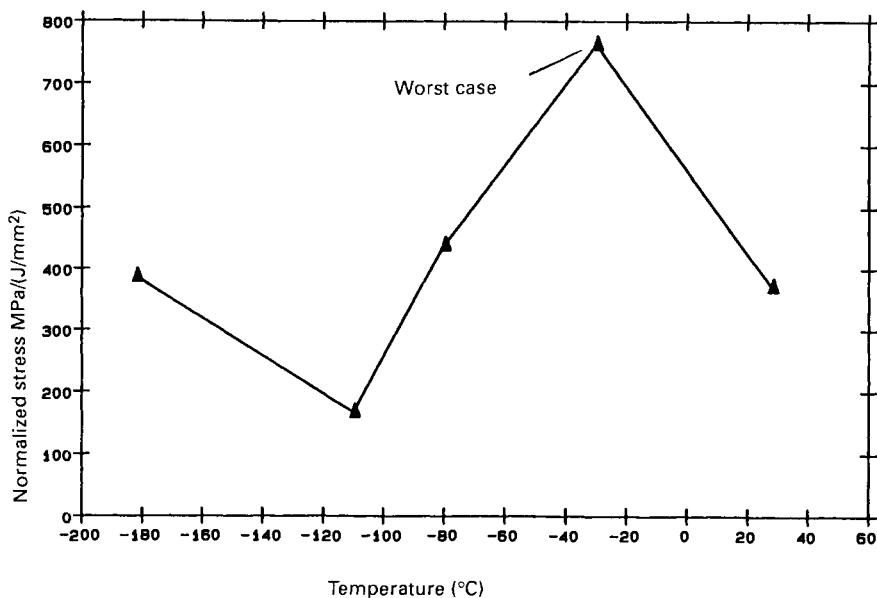


Figure 16: If one normalizes the recovery stress (which increases with temperature) to toughness, one sees that the critical temperature is reached at M_d : during recovery, not in service.

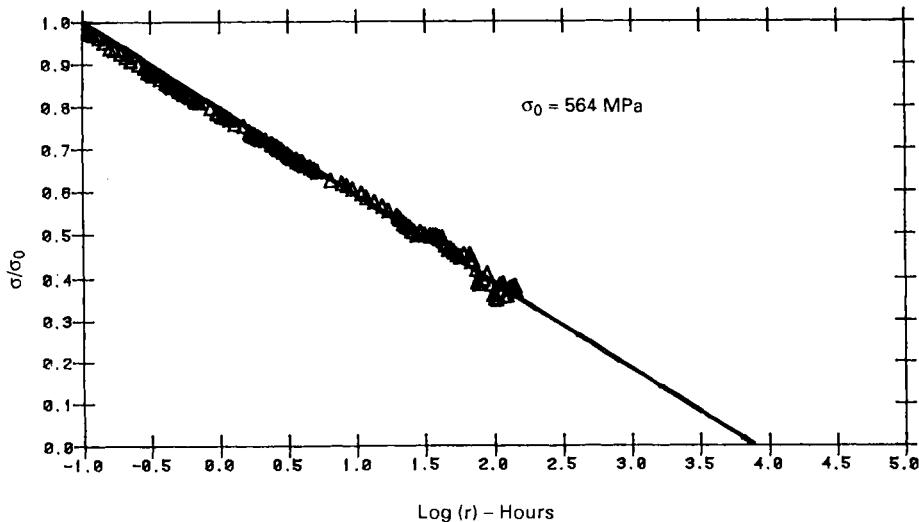


Figure 17: Stress decays logarithmically at 375°C from the initial recovery stress of 564 MPa.

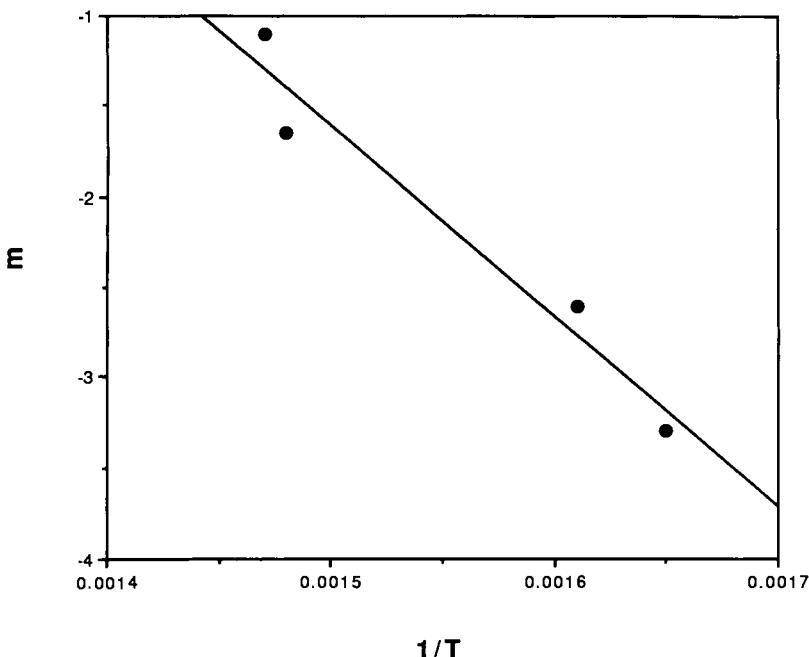


Figure 18: The rate of stress decay, $m = \Delta\sigma/\Delta\log(t)$, as a function of inverse temperature.

empirically determined to be 161,000 for the Ni-Ti-Fe alloy measured and Q was found to be 18.1 kCal/mole (or 9045.R °K). This indicates that this particular alloy is not susceptible to relaxation below 350°C.

Stress can decay during cooling for a variety of reasons:

- As one cools below M_d , martensite will be stress induced, and stresses will relax according to: $\Delta\sigma = (M_d - T)(d\sigma/dT)^{-1}$.
- Modulus is a strong function of temperature, generally reaching a minimum near M_d . Stress would decay according to: $\Delta\sigma = \Delta E \epsilon_c$.
- The R-phase often precedes martensite formation. Although the strain is small, it is large enough to significantly reduce the constraint stress.

For a shape memory alloy to be successfully used in constrained recovery M_s must be below $T_o - \sigma_r (d\sigma/dT)^{-1}$, where T_o is the lowest expected service temperature and σ_r is the minimum required recovery stress, and R_s must be below T_o .

4.0 The Application SMA's in Constrained Recovery

For the most part, the constrained recovery of shape memory alloys is applied in couplings and fasteners, which will be described in detail in subsequent papers of this section. In these cases, applications are single cycle and parts are heated beyond M_d .

during recovery. After installation, the SMA becomes a structural member of the system, and its shape memory properties are no longer considered (except in so far as they govern the usable temperature range and provide the live joint properties discussed earlier). As structural materials, corrosion, strength, fatigue resistance, ductility, toughness, etc. are important performance characteristics, and thus Ni-Ti alloys are far better candidates than Cu-based alloys - especially when creep or fatigue are key issues. The new Fe-based alloys may also be useful, but at the time of writing, there remain concerns about their corrosion and low temperature performance.

A second group of applications which are usually considered as constrained recovery are those in which multiple applications are required, e.g. the electrical connectors. In these cases M_d must not be exceeded, or if it is, a very high compliance must be built into the system. In many senses they can be considered actuators, though in terms of material selection and function they fit more into the constrained recovery category.

5.0 Summary

Constraining the recovery of a shape memory alloy allows one to build-up large stresses - in many alloys, over 700 MPa. These stresses have been used with great success for a wide range of fastener and coupling applications. The advantages that shape memory provides over competitive mechanical joining techniques typically include highly reliable and predictable recovery stresses, highly compliant joints that maintain their stresses despite small dimensional variations, and excellent corrosion resistance (in the case of Ni-Ti).

Conversely, there remain some pitfalls of which one must be aware. Most importantly, one cannot allow the temperature of a joint drop to near M_s , or the joint will spontaneously relax or even open. The tolerances of the items to be joined can also be critical if one needs to achieve a constant stress. Finally, it must be recognized that shape memory materials are substantially more expensive than steel or even titanium. In many cases this is compensated by the lower installation costs, but it still must be considered.

References

1. J. Perkins: Metals Forum 4(3), 153 (1981).
2. T. Duerig and J. Proft: unpublished research (1985).

Wide Hysteresis Shape Memory Alloys

T.W. Duerig, K.N. Melton and J.L. Proft
Raychem Corporation, Menlo Park, CA

The first commercial applications of shape memory (and the most successful to date) are as pipe couplings and mechanical fasteners, where shape memory offers advantages in installation time, reliability, weight and versatility. But in order to maintain the integrity of a shape memory joint, be it mechanical or hermetic, it is essential that the installed parts remain austenitic throughout their service life. If they become martensitic, the recovery stresses will be reduced at least to the martensitic yield stress, and more likely, the two-way effect will cause them to decay completely to zero. In the past, this has mandated the use of cryogenic alloys which remain austenitic at all normal service temperature, but must be expanded, stored, inventoried and assembled in liquid nitrogen to prevent premature recovery. Although many users of these products do not consider cryogenics to be a drawback, there is no question that the widespread commercial use of constrained recovery devices will only occur through the elimination of cryogenics.

All solutions to this problem involve increasing the hysteresis of the alloys - or more precisely, separating A_s from M_s as far as possible. The reasoning is that M_s controls the stress decay process mentioned above, while A_s dictates the maximum storage temperature. This principle is shown in Figure 1. The extent to which the hysteresis must be expanded depends, of course, upon the application. Shipping temperatures can reach 45°C in extreme cases, so the unstressed A_s of a successful wide hysteresis alloy must be at least that. How low M_s must be, however, is rather complex since the installed part is under stress and M_s is a strong function of applied stress. In general, one can calculate this from the following equation:

$$M_s < T_s - \sigma_r (d\sigma/dT)^{-1}$$

where T_s is the lowest expected service temperature, $d\sigma/dT$ is the stress rate of the alloy (typically 5 MPa/°C), and σ_r is the minimum recovery stress needed to maintain product integrity. For commercial couplings that are likely to carry only water and therefore cannot freeze, it may be sufficient to have an M_s temperature of -40°C; in military applications, where systems must be proofed at -55°C, the zero-stress M_s must clearly be below -90°C.

Although there are shape memory alloys which show large hysteresis loops (β -Ti¹ and iron based alloys², for example) the strains and recovery stresses of these alloys are

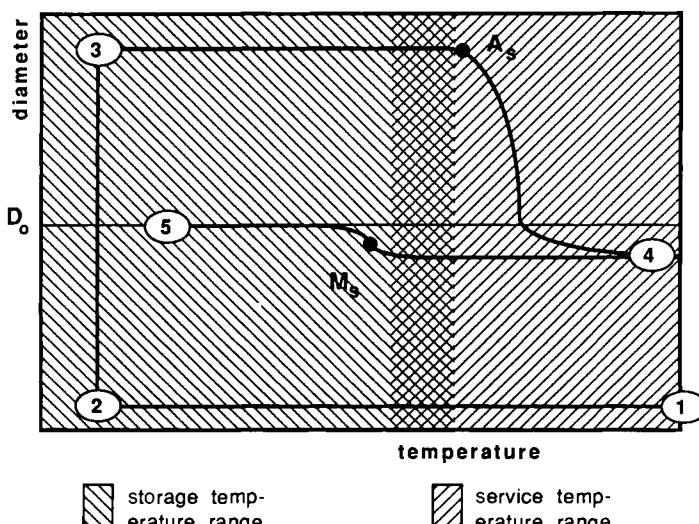


Figure 1: A schematic graph describing how a coupling is made: the alloy is cooled to below M_s (1-2), the diameter is expanded so that it is larger than that of the tubes (D_o) to be joined (2-3), and is then heated until it recovers to make contact with and subsequently crush the tube (4). If subsequent cooling takes place below M_s , the tube opens up and loses integrity (5). Thus the *service temperature range* bounded by M_s and the *storage temperature range* limited by A_s (above which premature recovery would make the part useless).

insufficient to allow one to design an effective device. A_s is essentially the same as M_s in Cu-Zn-Al and exceeds M_s by only 20 degrees in Ni-Ti. In order to provide what has been dubbed *heat-to-shrink* capabilities in these alloys, the hysteresis loops must be significantly expanded from their natural widths. In fact practical methods to expand the hysteresis have been found in both systems. The mechanisms of the hysteresis expansion are quite different in the two cases, but there are similarities in that it is A_s that is shifted to higher temperatures while M_s remains unaffected. Both approaches will be individually treated and then compared.

2. Cu-Based Alloys

The first alloy to be used in the above described heat-to-shrink mode was Cu-based: more precisely, a Cu-Zn-Al alloy with substantial additions of manganese (roughly 5.0 wt. %) and a greater than normal aluminum concentration (10.0 wt.%). The high aluminum content was chosen to increase the austenitic strength, retard stress relaxation at elevated temperatures, and improve stress corrosion resistance. The manganese addition was included to shift the eutectoid composition to higher aluminum levels, effectively preventing the formation of the brittle γ_2 phase³. The further addition of grain refining agents and adjustments in Zn concentration to tune the M_s to -40°C led to the so-called "Betalloy®" composition which has been in continuous use in constrained recovery applications for over 15 years.

Although the recoverable strain in this alloy is somewhat low (~3%), the alloy is capable of generating high recovery stresses (>400 MPa), is more resistant to stress relaxation than conventional Cu-Zn-Al (with extended service capabilities up to 125°C), is of relatively low cost and most importantly, can be produced with a large hysteresis. The process by which the hysteresis is expanded is called *preconditioning*⁵ and takes advantage of the martensite stabilization phenomenon discussed earlier. When the material is aged in the martensitic phase, that phase is stabilized and the reversion to austenite delayed to higher temperatures - in other words, M_s remains that of the unaged alloy, while A_s is increased. The shifted A_s value is often referred to as A'_s . It is important to note, however, that the martensite stabilization, or shifting of A_s , is temporary; once reversion is complete, A'_s shifts back to the original A_s . Figure 2 summarized this preconditioning process.

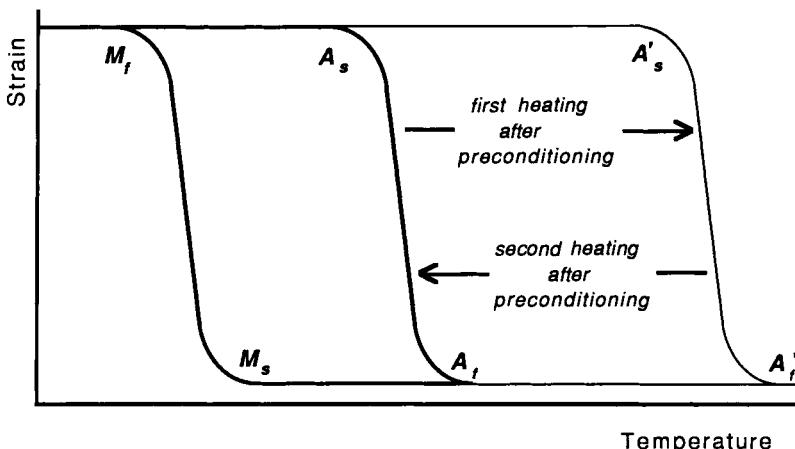


Figure 2: The preconditioning process is a one-time displacement of A_s and A_f . Once recovery is complete, martensite can be reformed, after which A_s and A_f are restored to their original values.

There are three approaches to practical preconditioning:

1. *Isothermal preconditioning* has already been described: one simply quenches to martensite and holds until A'_s is shifted to the desired 45°C. Unfortunately, the activation energy for the aging process is such that this can require several weeks below the -40°C M_s of the alloy being discussed and thus the process is largely impractical.
2. *Non-isothermal preconditioning* accelerates the above process by slowly increasing the aging temperature but assuring that it always remains below the current A'_s . While A_s continues to increase, the aging temperature can be increased and the entire process significantly accelerated, reducing the total preconditioning time to under 4 days.

3. Constrained preconditioning further shortens the process by taking advantage of the martensite thermoelasticity. After deformation in the martensitic phase, the parts can be constrained and prevented from recovering, thereby shifting M_s upwards - in principle allowing aging at temperatures up to M_d . In the subject alloy this allows aging well above room temperature, reducing the required times to under one hour. All three of the above processes can be used to shift A_s' to 45°C or above. For completeness one should note that there are two undesirable side effects of preconditioning: A_f is shifted even more than A_s so that the recovery response becomes sluggish, and the achievable recovery stresses are decreased to some extent. Even with these drawbacks, preconditioning has led to the development of two of the oldest and most successful shape memory products: the Permacouple® (Figure 3a) and the backshell ring fastener (Figure 3b).

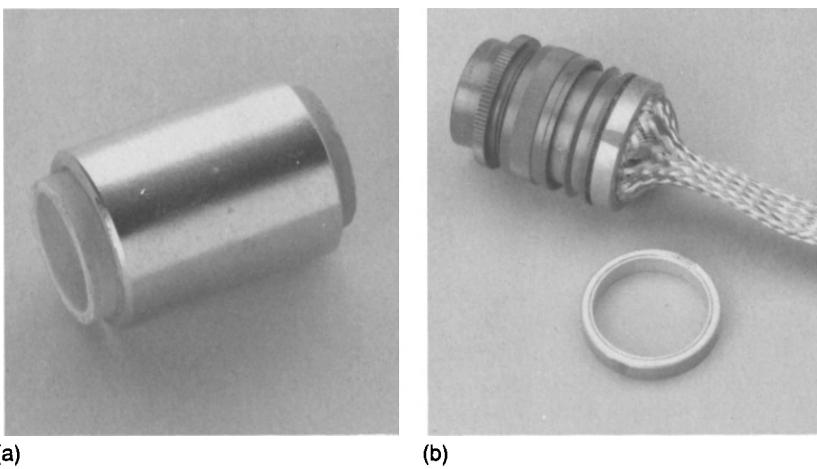


Figure 3: Two products made by preconditioned Cu-Zn-Al-Mn are (a) the Permacouple® and (b) the backshell braid ring.

3. Hysteresis Expansion in Ni-Ti Based Systems

The mechanical advantages of Ni-Ti over the Cu-based systems (greater ductility, fatigue resistance, strength, creep resistance, corrosion resistance, etc.) have always made this the preferred system for critical couplings and fasteners even though heat-to-shrink capabilities have not been available until relatively recently. The development of this capability, in fact, has long been considered one of the keys to the widespread deployment of shape memory. Although the martensite in Ni-Ti is not susceptible to stabilization like the Cu-base alloys, a similar sort of effect has been achieved by introducing an inhomogeneous microstructure. This is made possible by the addition of some 9 at. % niobium, and then overdeforming in the martensitic phase. The niobium is essentially insoluble in the matrix, forming large volume fractions of essentially pure niobium particles. The hysteresis of these alloys is somewhat larger than that of the binary (80°C versus 30°C) when measured without applying any external stress (such as would be measured by DSC or resistivity); but when the transformation is accompanied by a shape change, the hysteresis becomes much larger (in excess of 120°C).

The microstructural mechanism responsible for the hysteresis expansion has not been entirely defined, but can be understood in general terms by considering the schematic representation of the microstructure shown in Figure 4. X-ray and TEM observations

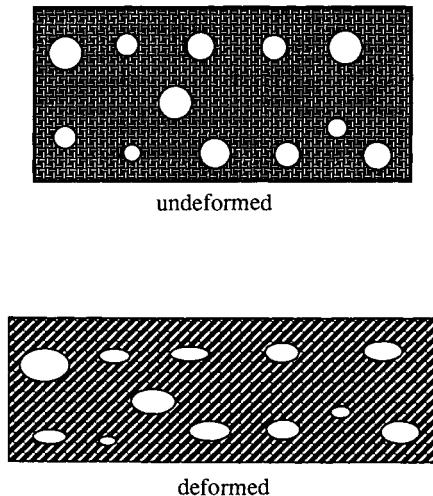


Figure 4: Schematic representation of the undeformed (above) and deformed (below) microstructures of martensitic Ni-Ti-Nb. The light spherical phase represents nearly pure, very soft niobium, and the matrix represents the essentially niobium-free Ni-Ti matrix. Both phases participate in the deformation process, providing irreversible and reversible components, with the irreversible component increasing hysteresis.

have shown that both the martensitic Ni-Ti matrix and the niobium particles undergo deformation at relatively low stresses.⁶ The martensite deforms reversibly by the usual twin boundary motion while the niobium deforms irreversibly via slip. This partitioning of the total strain into reversible and irreversible components appears to be unique to the Ni-Ti-Nb system since we know of no other systems in which a large volume fraction of a soft second phase is formed. When the deformed composite structure is heated above A_s , the matrix phase begins to recover but can only proceed significantly by plastically reforming the niobium particles back to their original shape. Thus the irreversible deformation delays recovery, increasing A_s to A_s' . One would expect the shift in A_s to increase as the total strain increases since the irreversible strain component increases (Figure 5). One also expects and observes that the second cycle A_s (after recovery and reforming of martensite) is again the original value, just as in Cu-based alloys. A more complete description of this process can be found in reference 7.

Applications made from the above described process have been commercially available since 1986. The most successful to date has been a backshell braid ring which is a welded wire ring coated on the ID with a polymer to allow electrical heating without passing the current through the substrate (Figure 6).

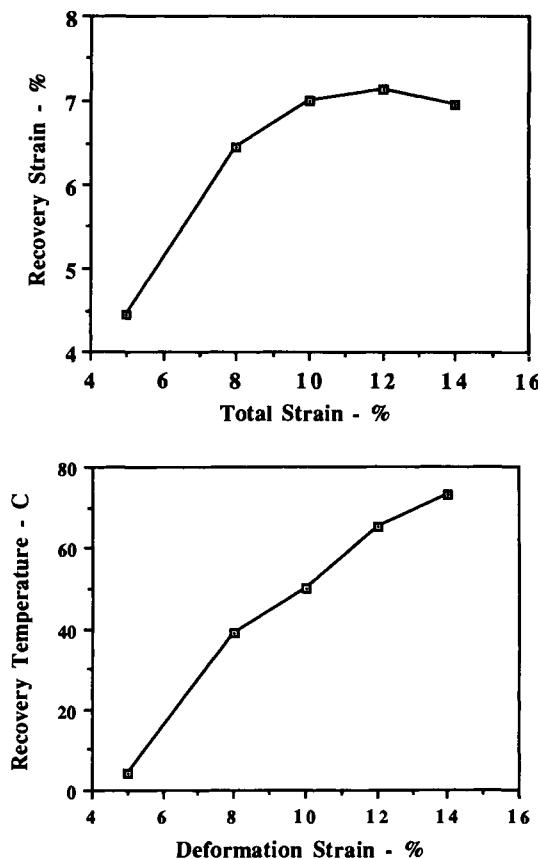


Figure 5: The effects of overdeformation on the recovery strain and the recovery temperature (A_s') of Ni-Ti-Nb.

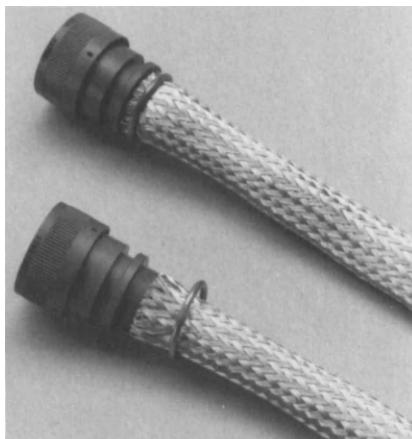


Figure 6: The first product to be made from the heat-to-shrink Ni-Ti-Nb alloy was a backshell braid ring to perform similar tasks as the Cu-Zn-Al-Mn version (Figure 3b).

3. Comparison of Ni-Ti-Nb with Cu Based Alloys

Table I compares the properties of Ni-Ti-Nb with Cu based heat-to-shrink alloys. Although it appears at first glance that the Ni-Ti based alloys are far superior, the Cu-based alloys can be made at a far lesser cost. To date, applications for the Cu-based alloys have not been wide-spread enough to allow us to take full advantage of lower raw material costs, but the potential is there.

Table I: Typical Properties of High Hysteresis Alloys

Property	Cu-Zn-Al	Ni-Ti-Nb
M_s (zero load)	-40°C	-140°C
maximum A'_s	+50°C	+50°C
ϵ_r	3.5 %	8.0 %
σ_r	400 MPa	700 MPa
$d\sigma/dT$	2 MPa/°C	3.5 MPa/°C

References

1. T.W. Duerig, D. Richter and J. Albrecht: Scripta Met., 16, 957 (1982).
2. A. Nagasawa: J. Phys. Soc. Jap., 30, 1505 (1979).
3. P.L. Brooks: U.S. Patent 4,166,739 (1979).
4. K.N. Melton, J. Simpson, and T.W. Duerig: ICOMAT 1986, Jap. Inst. Metals, 1053 (1986).
5. G.B. Brook, P.L. Brooks and R.F. Iles: U.S. Patent 4,067,752 (1978)
6. L.C. Zhao, T.W. Duerig and C.M. Wayman: MRS Conf., Tokyo, (1988).
7. T.W. Duerig and K.N. Melton: The Martensite Transformation in Science and Technology (E. Hornbogen and N. Jost, eds.), DGM, Oberursel, Germany, 191 (1988).

Shape Memory Alloy Tube and Pipe Couplings

Michael Kagan and Keith N. Melton
Raychem Corporation
300 Constitution Drive
Menlo Park, CA 94025

It is generally recognized that the application of shape memory alloys to tube or pipe couplings has been the most successful to date. The initial use was on the Grumman F-14 fighter, which was one of the first aircraft to use titanium tubing for hydraulic lines. At the time there was no convenient way of making permanent joints on the airplane, and a shape memory coupling was designed and qualified in very short time. This has been a consistent theme of shape memory couplings, they provide a unique technical solution to fitting problems. Over the years, the technology has also been applied to join pipe for marine or industrial use. In this paper, the application will be described in detail: how SMA couplings work, how they compare with other competitive methods, how they are installed, and finally some of the reasons for their success, which may be relevant to other applications.

1. Principles of Shape Memory Alloys Couplings

SMA couplings are a classical example of restrained recovery,^{1,2} where the forces generated by preventing the shape memory effect from going to completion are used to effect the join. Typically it is a single cycle event, the couplings provide a permanent connection.

In the first step of making a coupling, the part is machined in the austenitic condition, such that its inner diameter (ID) is smaller than that of the tube or pipe outer diameter (OD) to be joined. To aid in creating a seal, sealing lands or teeth are provided on the inner surface. The part is then cooled to its martensitic state, and mechanically expanded while martensitic such that its ID now exceeds the outer diameter of the tube or pipe to be joined. Once expanded, the part must be stored in the martensite until it is ready to be installed. In the case of a cryogenic alloy this requires maintaining the coupling at a low temperature, and this is conveniently done using liquid nitrogen. Since cryogenic alloys were for many years the only ones available for couplings, much of the discussion here will concentrate on this technology. The heat-to-shrink alternative was discussed in the previous chapter. Installation is done by removing the coupling from liquid nitrogen, inserting the pre-marked tubes, and then allowing the alloy to warm up and transform to austenite. On doing so it tries to go back to its as machined dimension but is constrained from doing so by the tube. The high stress then generated creates the joint. This process is illustrated in Figure 1.

The tooling required to do an installation is very simple, Figure 2. It comprises a check

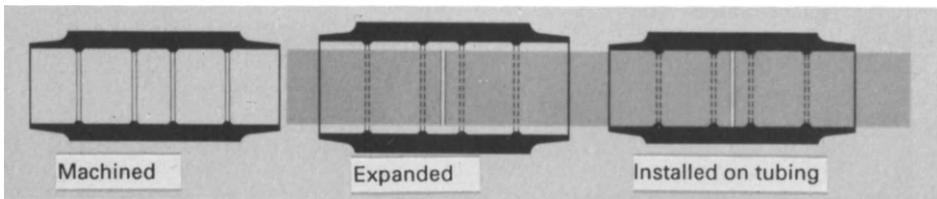


Figure 1 The principle of the SMA coupling

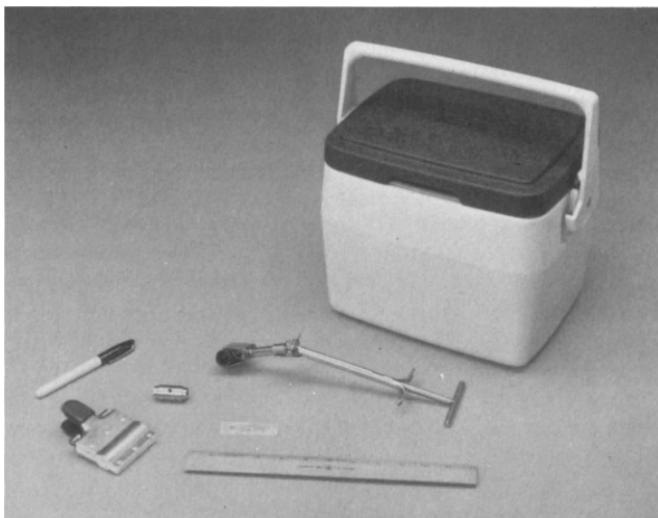


Figure 2: The installation tooling for a Cryofit®, SMA coupling installation.

gage and marker pen to mark the insertion depth on the tubing, a gage to perform a post-installation inspection, a tube chiller, a workbox which is a small container capable of carrying liquid nitrogen, into which the coupling is transferred from the transport dewar, and a spring loaded tool to remove the coupling and position it on the tubes. By comparison the tooling for externally swaged fittings is much heavier and more expensive. It typically consists of a hydraulic pump, and a swaging head. Because the forces necessary to create the joint have to be externally supplied, the pump and head have to provide these. The swaging tool is therefore of robust construction, and heavy to use. Furthermore the high forces can create a safety problem if the tooling is not meticulously maintained, and can lead to costly wear of the swaging dies and head. Also, as a consequence of the size of the swaging tool, it cannot be used in tight spaces or with tubing packed closely together.

Figure 3 shows the sequence of events for an SMA coupling installation. The test coupling, machined from a conventional alloy to have the same dimensions as the expanded SMA coupling, is inserted over the tubes, the ends of which must be visible in the hole in the side of the test coupling, Figure 3a, and the tube is marked. In order to provide more working time, the ends of the tubes are cooled using the chiller. The

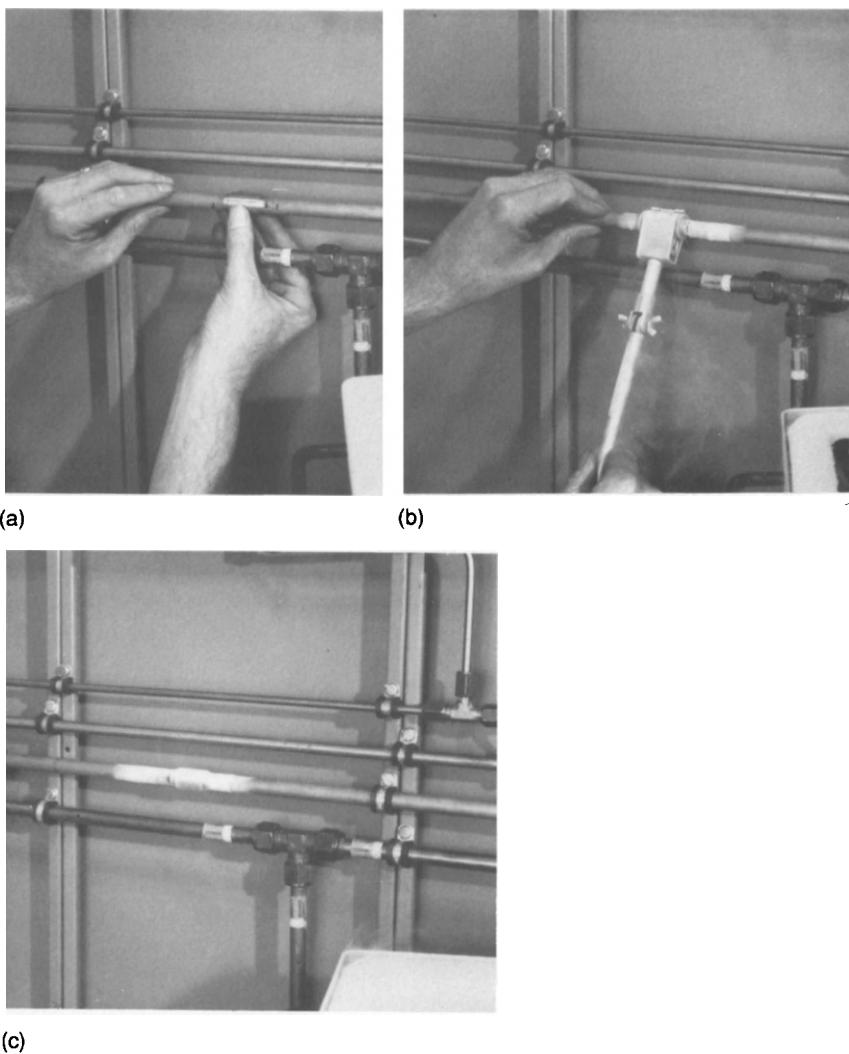


Figure 3: The tubing being marked ready for installation using a test coupling. (a) It is then positioned on the tubing, the ends of which have been previously chilled. (b) The coupling is located inside the tool used to remove it from the liquid nitrogen in the work box. The installed coupling is then warmed up to room temperature (c).

tool is then removed and the coupling allowed to warm up to room temperature (Figure 3c) and the joining is complete. This whole operation is simple, fast and requires very little operator training to perform. A minimum amount of post installation inspection is necessary - basically all that is required is to confirm that the two tubes are approximately centrally located such that each tube is grabbed by two sealing lands. The craft insensitivity of the whole procedure is one of the big advantages of shape memory couplings.

The small size of the SMA coupling installation tooling results in easy installation, even in difficult to access situations. In particular it gives the hydraulic systems designer the freedom of laying out his tubes how he would like, without the joining method providing constraints because of tool access. As can be seen in Figure 4, SMA couplings can be successfully installed when the tubes are very close together. If a swaged fitting were to be used, the tubing would need to be locally bent to allow tool access. As well as being an additional operation, this adds both tubing weight and hydraulic fluid weight to the system. Such a weight penalty is unacceptable in a flying vehicle.

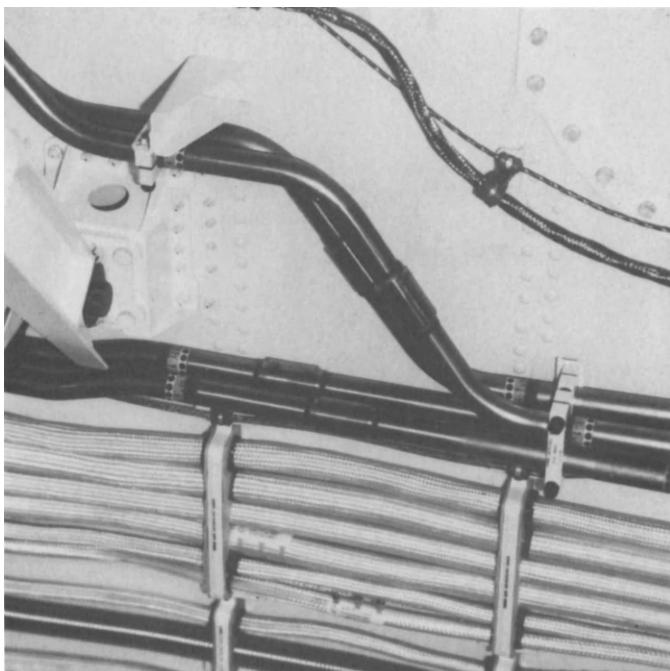


Figure 4: An on-aircraft installation of Cryofit showing how closely spaced the tubes can be.

Figure 5 shows a cutaway of an installed Cryofit® coupling, (Cryofit is a Registered Trademark of Raychem Corporation), where it can clearly be seen that the sealing lands have been compressed into the tube, causing local plastic deformation and creating a metal-to-metal seal. The effectiveness of the seal depends on the geometry of the tooth, and on the mechanical properties of the tubing material and the SMA itself. Using the types of data generated by constrained recovery testing, and Finite Element Method (FEM) analysis, the interaction between the tooth and the tube can be modelled. An example is shown in Figure 6, where the local plastic strains can clearly be seen, as well as the "bite" where the tooth has penetrated into the tube. The different shadings correspond to different stress levels. Computer programs capable

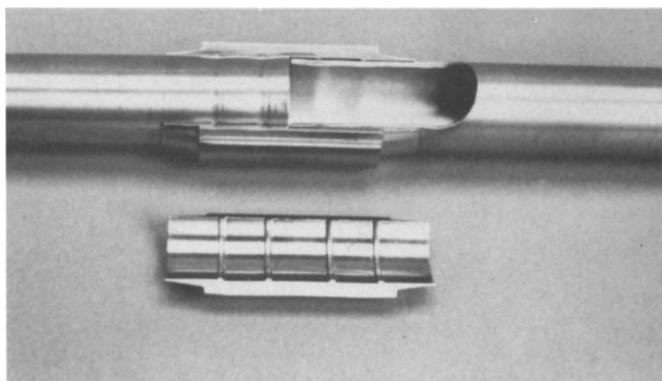


Figure 5: A cut-away of an installed SMA coupling.

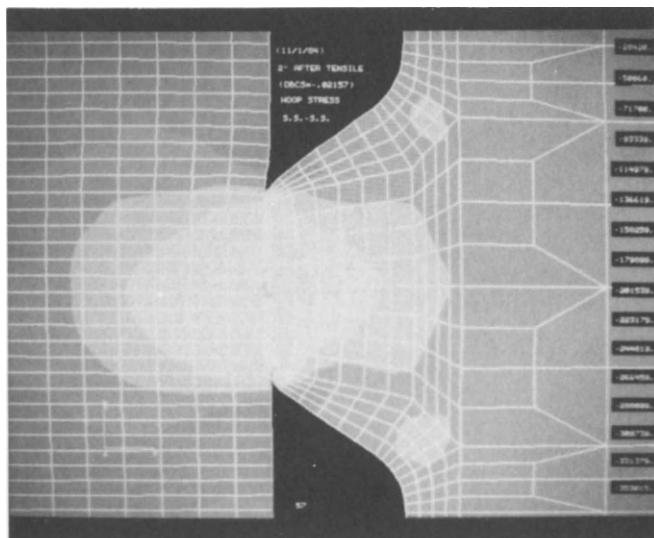


Figure 6: Finite Element Analysis of the interaction between the sealing land of an SMA coupling and the tube.

of handling a finite element mesh in three dimensions and also non-linear plastic deformations can be used to model the stress in a fitting as a whole, as can be seen in Figure 7. A quantitative understanding of the stresses in a coupling allows its optimization with respect to performance, and also a minimization of its weight. SMA's allow the control of stress in different parts of the coupling, which is of particular importance in achieving good performance in flex (see below).

2. Testing of Couplings

SMA couplings undergo exactly the same type of qualification tests as any other mechanically attached or welded fitting. Typically these include:

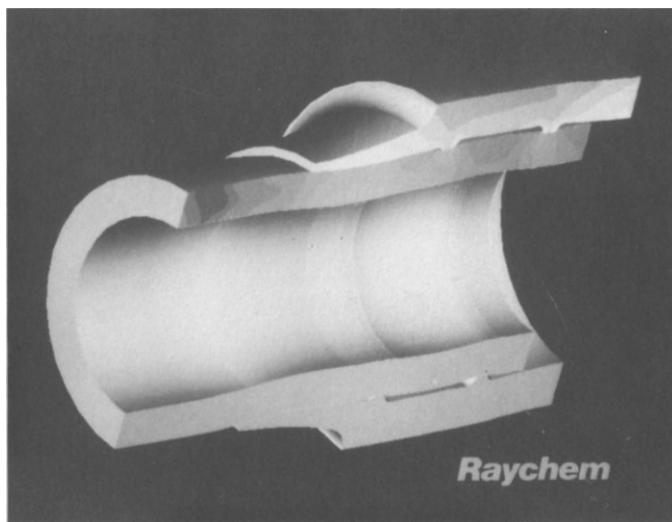


Figure 7: A section of a Finite Element Analysis of the stress distribution in an installed coupling.

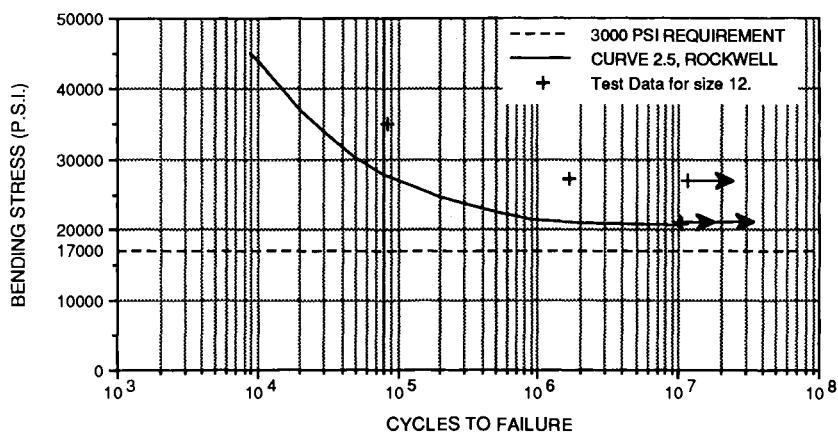


Figure 8: The cyclically applied bending stress (S) as a function of the number of cycles to failure (N). Failure should be above one of the curves shown.

- a) Gas leak, where a fitting assembly is internally pressurized with gas, normally at low pressures up to 7 MPa (1000 psi).
- b) Hydraulic proof, where the fitting assembly is pressurized with hydraulic fluid at a pressure of twice the nominal operating pressure. In both these cases the pressure is maintained for a period of time, and the requirement is that there shall be no leaks.
- c) Burst test, where after the hydraulic proof test the pressure is continually increased until failure occurs. The requirement is that failure shall not occur below four times the nominal pressure of the system.
- d) Flex test, in which a fitting assembly is internally pressurized to the nominal operating pressure and simultaneously subjected to reversed bending loads. The requirement is that failure (as indicated by a loss of pressure) will be above a particular curve when the applied bending stress is plotted against the number of cycles (bending reversals), see Figure 8. This figure shows two curves highlighted - the lower one was a previously used standard introduced for 3000 psi systems and the upper one recently introduced by Rockwell on the B1 program at 4000 psi pressure. The work done at that time redefined the standards for modern airplanes. Also shown in Figure 8 are data for Cryofit couplings which clearly fall well above this increased requirement.

3. Marine and Industrial Couplings

So far the discussion has focussed on couplings for aircraft hydraulic tubing. The same technology can however also be used to join pipes, e.g. for marine^{3,4} or industrial⁵ application. Pipe differs from tubing in that it is normally thicker in wall and it is dimensioned and tolerated differently. Furthermore it is not usually exposed to the same high frequency vibration experienced by an airplane. Consequently a different style of coupling can be used, and a cutaway is shown in Figure 9. All the

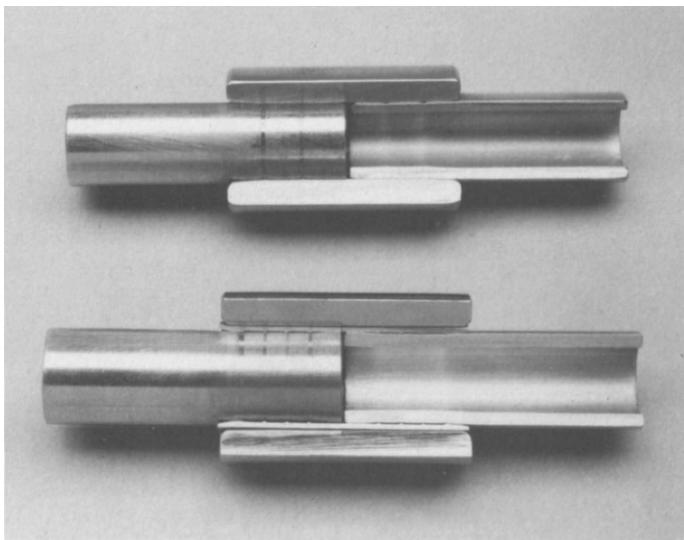


Figure 9: Cut-away sections of two types of pipe coupling installations.

installation advantages of the aircraft product also apply here. For example, Figure 10 shows that these couplings can also be installed with the pipework very close together. Because of the larger mass of these pipe couplings compared to those used for aircraft tubing, chilling of the pipe ends is not normally needed prior to installation. A design of fitting which offers some advantages is that shown in Figure 11, where the sealing teeth are machined on the inside of a cylindrical tube or liner of conventional

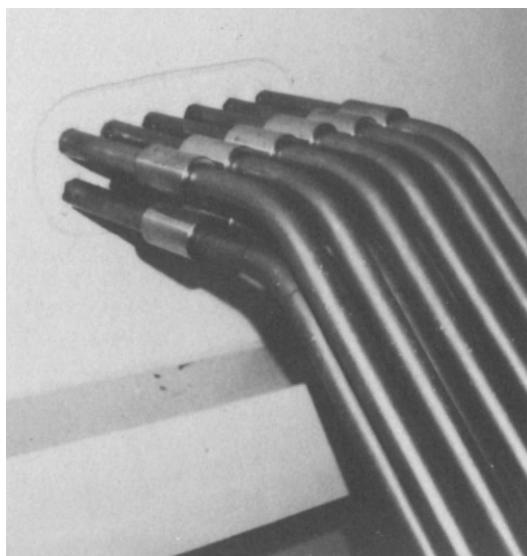


Figure 10: Illustrates the high packing density of pipes which can be joined using SMA couplings. This photograph is taken on board a ship, and other joining techniques e.g.welding would not be possible.

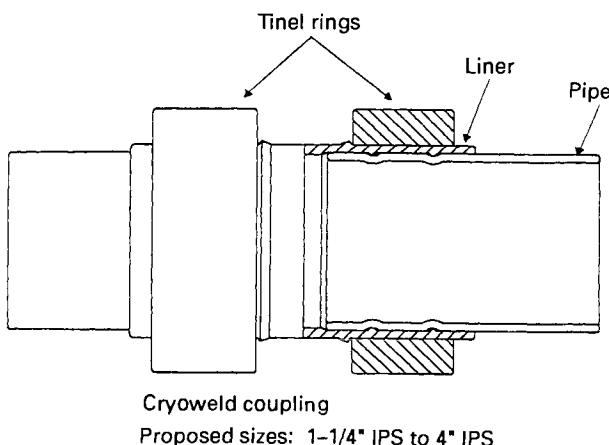


Figure 11: A schematic illustration of the ring-liner concept of SMA coupling design.

material, chosen to be compatible with the pipe. Around the teeth are shrunk rings of SMA, which initially crush the liner and subsequently drive the teeth into the pipe (Figure 12). This configuration is particularly beneficial for tee or elbow shapes, Figure 13 where the weight of SMA used is much less than if three (in the case of a tee) conventional couplings were used. Consequently the fitting can be more cost effective. A range of such fittings are available for low pressure copper and copper-nickel pipe.

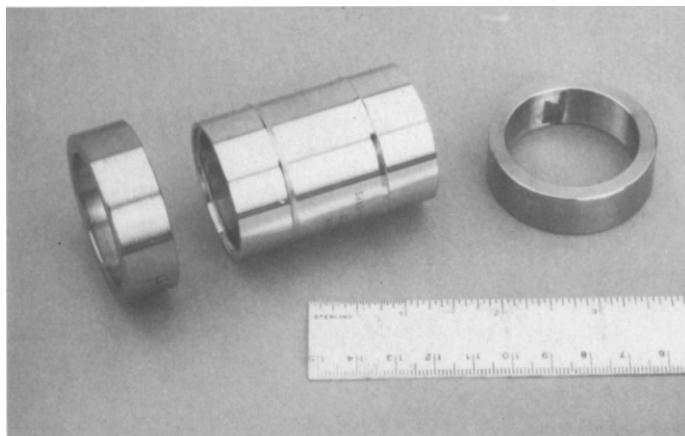


Figure 12: The ring-liner concept to fittings.

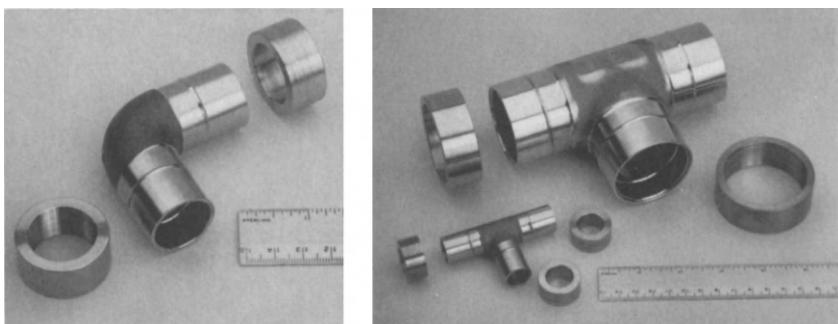


Figure 13: A range of fittings for joining marine pipes which use the ring-liner concept.

The installation tooling and procedures for the marine/industrial couplings are very similar to those already described for aircraft use. However, one difference is that pipe is not always handled as carefully as aircraft tubing, and can become damaged. Longitudinal gouges such as can be seen in Figure 14a do not lend themselves readily to creating a seal, and so for such cases the ends of the pipe must be prepared by lightly abrading, Figure 14b. An installation sequence for such a SMA pipe fitting is

shown in Figure 15. The procedure is very rapid and requires significantly less training than for example welding, resulting in a lower total installed cost. Once installed, the metal to metal seal is much more temperature resistant than a brazed fitting. Consequently the fire survivability of a SMA coupling is much superior to a brazed joint.

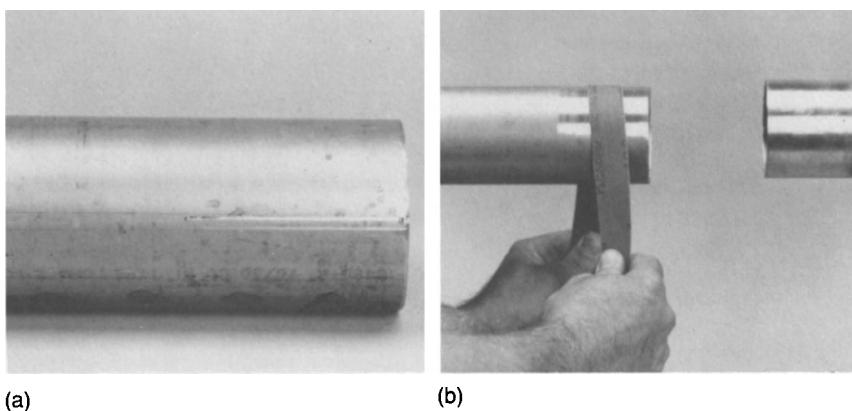


Figure 14: If gross defects such as gouges (a) are present, then the pipe ends should be lightly abraded (b) to remove them.

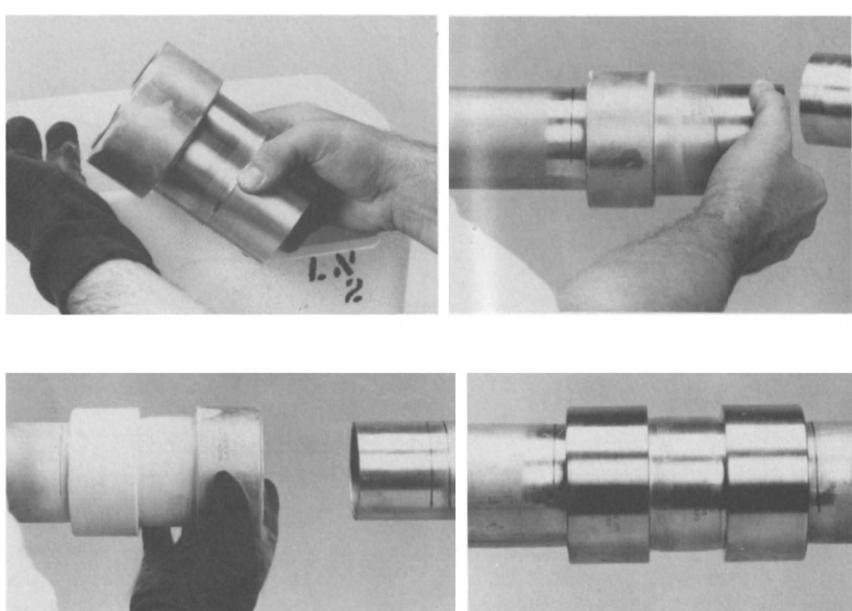


Figure 15: The installation sequence for a ring-liner pipe coupling assembly.

4. Tolerances

When assessing the feasibility of using SMA couplings for a pipe joining application, the dimensional tolerances of the pipe must be taken in consideration. This is in contrast to aircraft hydraulic tubing where the tolerances are much tighter. For example, 1/2 inch tube has an OD of 0.500 ins. with a tolerance of plus 0.003 ins. minus zero. In contrast 1/2" pipe has an OD of 0.840 ins. with a tolerance of plus 0.015 ins. minus 0.030 ins. The reason why tolerances have to be considered is that a fixed maximum amount of shape memory motion is available. Care must be taken that after provision for adverse tolerance stack-up, enough motion is left in the worst case to create the desired metal to metal seal. As previously noted the recovery stress in a constrained recovery test starts to fall off rapidly below a contact strain value of around 2%, in other words, the high stress capability of SMA is only optimally used if there is more than 2% unresolved recovery (i.e. a contact strain of over 2%).

Approximating the amount of available memory effect as 8% for Ni-Ti, then this leaves 6% which must be split between:

- a) coupling ID tolerance i.e. machining tolerance
- b) assembly clearance, i.e. the amount of initial free recovery before constraining
- c) the amount of crush on the pipe
- d) the OD tolerance of the pipe

Exactly how to make the decision as to where to partition this depends on the details of the particular application. For instance, a coupling intended for repair use in the field might have a larger assembly clearance than one installed in a controlled factory environment. In order to illustrate the point regarding tolerances we will make here the arbitrary decision that of the 6% available after taking unresolved recovery into account, half of this i.e. 3% will be used to accommodate the pipe OD tolerance. We will further assume a symmetrical tolerance i.e. $\pm 1.5\%$. Figure 16 then shows the

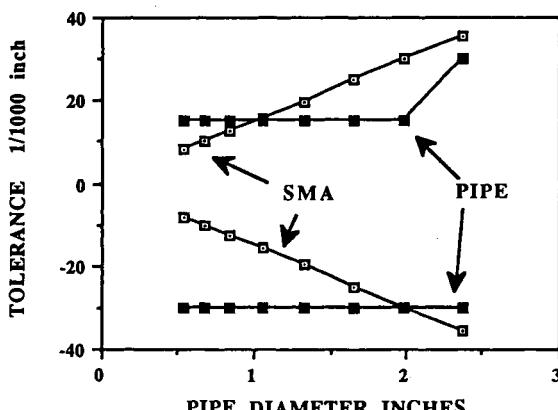


Figure 16: The pipe tolerances (ref. 6) as a function of pipe diameter, compared with the tolerances permitted by the SMA making the assumptions outlined in the text.

resulting actual tolerance dimension in inches, as a function of pipe size. Also shown is the allowable tolerance according to the pipe specification⁶. It can be seen that an SMA coupling designed in this manner will not accept the full range of pipe tolerance except at the very largest sizes. It should be noted also that other mechanically attached fittings suffer the same tolerance problem. In practice this is less of a problem, since the pipe specifications were written many years ago, and current manufacturing procedures provide much more consistent product. In other words, the full allowable range is rarely met in practice. To guarantee that the SMA coupling will work on a particular pipe, a simple "go/no go" gage is used to verify that indeed the pipe is within the SMA allowable dimensional tolerance.

5. Success of Coupling Application

Finally let us examine why SMA couplings have been so successful, and see what can be learned from this which is applicable to other SMA uses.

- The coupling solved a customer problem, and provided a technically superior product.
- A lot of attention was paid to quality. The couplings work every time, and soon established an outstanding reliability record.
- The product is price competitive on an installed cost basis.
- A fully engineered product is provided - the customer does not need to understand SMA in order to make it work.

References

1. Duerig, T.W. and Melton, K.N., Shape Memory Alloy 86, Proceedings of International Symposium on Shape Memory Alloys, Sept 6-9, Guilin, China, Academic Publishers P.397.
2. Duerig, T.W., Melton, K.N., "Designing with the Shape Memory Effect", MRS Conference, Tokyo, (1988).
3. Caskey, M.R. and Embury, G.D., Naval Engineers J. April, 45, (1979).
4. Liberatore, D.J. and Baskerville, J.E., Naval Engineers J., Dec., 63, (1982).
5. Bryant, N.J. and Platshon, M.C., American Society of Sanitary Eng., Vol.63, Part I, (1986).

The Design of Electrical Interconnection Systems with Shape Memory Alloys

Ed Cydzik
Raychem Corporation
Menlo Park, CA.

For over 15 years, SMA connectors have been used in critical applications, where reliability and high stress environments have precluded the use of any interconnection scheme other than hard wiring. When properly designed, SMA connectors are not susceptible to environmentally induced degradation such as fretting and fretting corrosion, or damage from chemical environments which affect the integrity of the contact zone.

1. Shape Memory Alloy Connector Performance

Typical SMA connectors perform extremely well in adverse environments because they are zero insertion force (ZIF) when open, and provide high retention forces during operation. The zero insertion force feature allows one to mate the connector with little if any frictional force. This is true for single lines or multiple lines. Contrast this to a typical connector system in which every line has a frictional force during insertion. As the pin count increases, the mating force remains very low for SMA connectors, but increases in a linear fashion for standard connectors. Therefore, very little if any of the mating force is transferred to the supporting structure in SMA connectors.

In addition, because most SMA connectors are always internally in a force balance, no actuation forces are transferred to any external structures during connector opening. This is not the case for mechanical ZIF connectors which tend to transfer forces and apply moments to their mounting structures. For SMA connectors then, designers do not need to stiffen any of the connector mounting structure due to excessive actuation or mating forces.

High retention forces mean that once in the mated position, SMA connectors exert extremely high normal forces which inhibits any relative movement between connector members. This prevents fretting and associated fretting corrosion from occurring. Since normal forces are very high relative to standard contact systems, the contact zones are "gas tight"; i.e. they prevent the penetration of corrosive gases.

In practice, Cryotact™ DIP (Dual In-line Package) connectors and PGAP (Pin Grid Array Package) connectors have been tested in extreme dynamic environments. The following table summarizes their performance and illustrates the superiority of the SMA connector in shock and vibration.

Table I: Performance summary in shock and vibration for various SME connector systems.

Contact/connector type	Vibration performance	Shock performance
Cryotact	80 G Sine 4 hrs/axis 3 axes No discontinuity >20 nanoseconds	1500 G .5 milliseconds 3 axes No discontinuity >20 nanoseconds
	51.5 GRMS Random 4 hrs/axis 3 axes No discontinuity >20 nanoseconds	MIL-STD-910C lightweight hammer 3 axes No discontinuity >20 nanoseconds
PGAP	30 G Sine 4 hrs/axis 3 axes No discontinuity >20 nanoseconds	500 G 1 millisecond 6 axes No discontinuity >20 nanoseconds
	32.3 GRMS Random 4 hrs/axis 3 axes No discontinuity >20 nanoseconds	

2. Designing Connectors with Shape Memory Effect Alloys

To properly design a connector using SMA, it is imperative to define the operating and storage environments in which the connector is expected to perform. It is also important to assess the components which are to be interconnected; i.e. is the connection between I.C. chips and Circuit Boards or between two PCB's such as in a Mother/Daughter board arrangement, or is the connection a production break in a harness?

One must arrive at the following answers prior to proceeding with the design:

- a) What current levels is the contact system expected to carry?
- b) What is the maximum allowable contact resistance?
- c) What are the mating part masses?
- d) To what acceleration levels will the system be driven?
- e) What are the anticipated amplification factors which may increase the actual acceleration at the connector interface?
- f) What are the operating temperature extremes that the system will be exposed to?

This gives the designer some preliminary answers which allow him to start pencilng in required contact cross sections and required normal forces. The results of the temperature survey will start narrowing the designer's alloy selection options. It also brings up the question of how to employ the SMA. Does one choose to use the alloy in its austenitic state throughout the operating temperature range, or does one pick a different alloy which is martensitic in the connector operating temperature range.

An example of the latter case is the Betaflex® Connector, which employs a single SMA driver as an actuator to open the connector. More insight can be gained into this method of employing an SMA by reviewing Reference 2. For the purposes of further discussion, only connectors with alloys which are austenitic in the operating temperature regime shall be covered. Once this concept is well understood, the opposite scenario is rather obvious.

It is of some benefit to digress now and look at a very basic but extremely effective SMA contact. Figure 1 describes a Cryocon™ contact, comprised of a split heavy

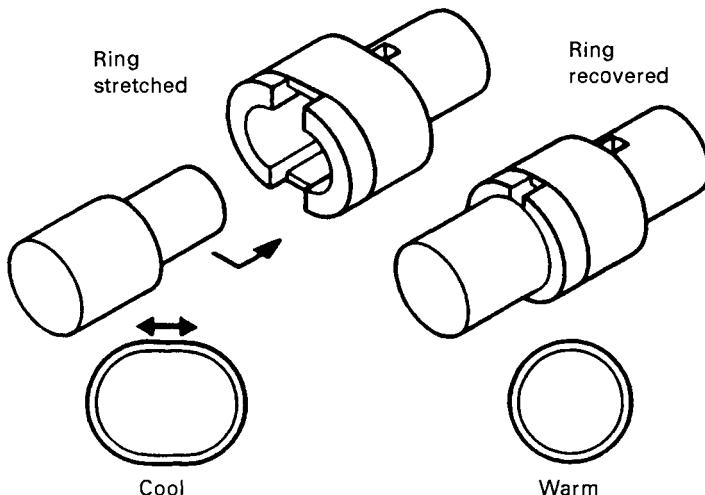


Figure 1: Cryocon contact.

walled tube called the *biasing spring* (typically Cu-Be or other conductive spring material), and a ring of Ni-Ti, called the *driver*. Prior to assembly, the tube is set in a splayed orientation. Once the SMA ring is installed, this system will reach an equilibrium position. This equilibrium position may be altered by changing the temperature of the connector.

Typically, the design of this system is such that the SMA ring is austenitic in the operating temperature range. This means that during operation, the ring will force the split-walled tube closed, decreasing the tube diameter well below the no load set dimension. Upon cooling of the system below M_f , the split-walled tube will elongate the Ni-Ti ring and increase the relative tube diameter. These two positions will allow a mating pin to be held firmly in the tube when the Ni-Ti is austenitic, and will allow the pin to be withdrawn when martensitic.

Usually, it is necessary to cool the SME alloy below its M_f temperature by using tooling which convectively cools the alloy with a Liquid Nitrogen/Nitrogen gas mixture. This process may be further understood by examining a stress-strain diagram for a SMA (Figure 2), and the corresponding force-stroke diagram for the system (Figure 3). Figure 2 shows two stress-strain curves for Ni-Ti: one above the A_f (solid line), and the other far below M_f (shown as a dashed line). As an alloy transforms between austenite and martensite, the biasing spring and driver system will follow the dotted

line between the austenite and martensite conditions, which is known as the *operating path*. This is shown in Figure 3. The difference between the stroke position in martensite and the stroke position in austenite is the total stroke which can be achieved by the system.

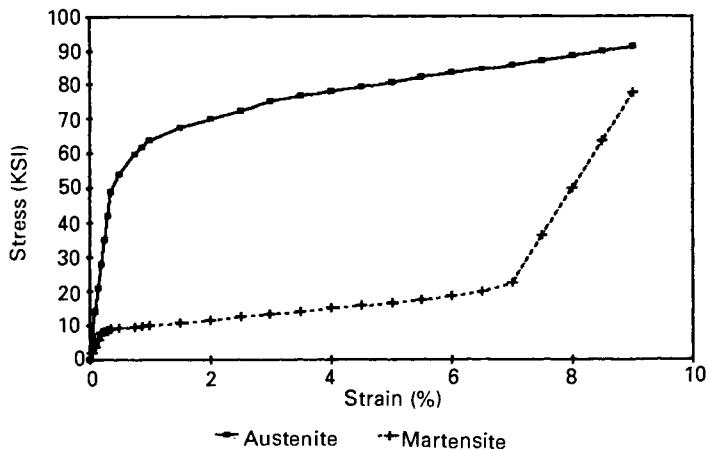


Figure 2: SMA stress vs. strain diagram.

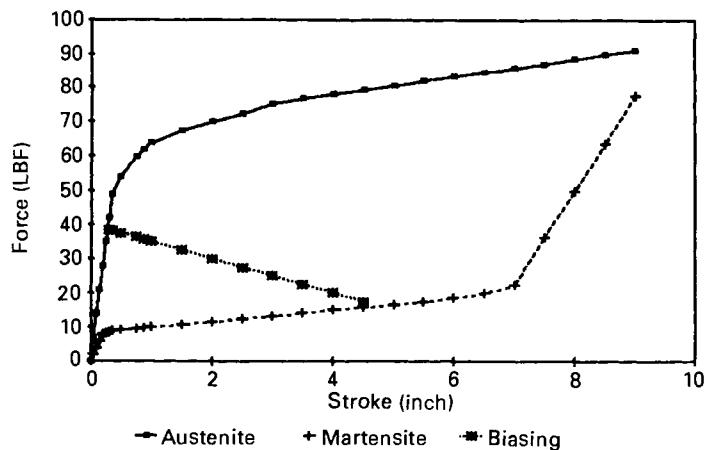


Figure 3: Force vs. stroke curve with biasing for a SME alloy connector.

When a mating pin is inserted into the Cryocon® contact, an additional non-compliant member is introduced into the system. This changes the operating path of the system,

as shown in Figure 4, which is the force-deflection diagram of the system with a non-compliant member installed in the contact zone. Note the change in slope of the operating path. The all important normal force is the difference between the force on the operating path curve where the change in slope takes place (the tube encounters the pin), and the force at the final equilibrium condition where the operating path intersects the austenite curve.

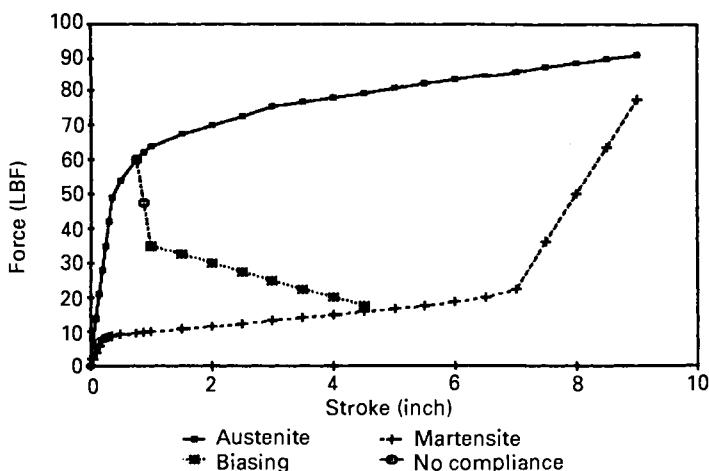


Figure 4: Force vs. stroke with a pin for a SME alloy connector.

The driver stress rises dramatically with no additional system stroke generated. Care must be taken to ensure that the operating stresses stay below a predetermined maximum for satisfactory cyclic performance. In practice, this means that the mating pin diameters for this type of a contact system must be extremely precise. In addition, the operating stress in the mated condition must not drive the alloy M_s temperature into the operating temperature range of the connector. It is important to remember that the alloy M_s temperature and alloy motion are functions of applied stress.

For connector designs, one is now in the position to start answering some additional questions such as:

- (g) How much stroke (motion) is needed between the open and closed condition?
- (h) How many actuation cycles one requires?

Typically, the total stroke required depends on the mating pin tolerances and the positional location of the matrix of pins for a multiple pin connector. The number of actuation cycles required will determine the maximum allowable system operating stress.

To create a more flexible contact system, some compliance must be designed in the connector. Figure 5 shows the familiar force-stroke diagram with a compliant member added. Compare the slopes of the operating paths between Figure 3, 4 and 5.

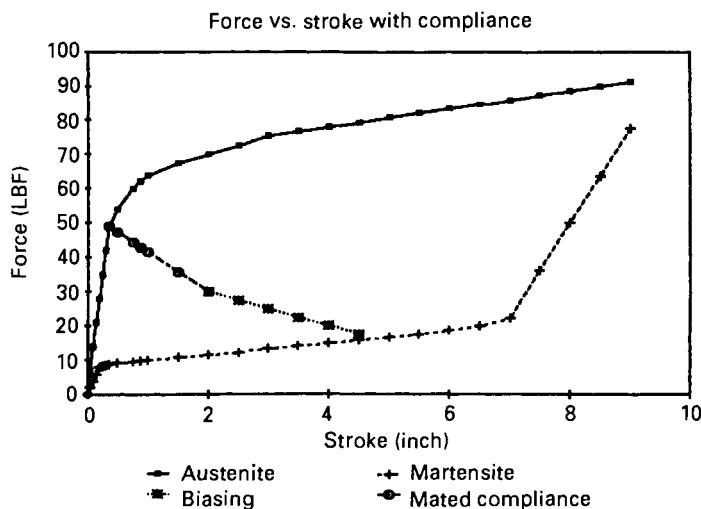


Figure 5: Force vs. stroke with compliance for a SME alloy connector.

Compliance can be added to the system by designing the biasing portion of the system so that when a rigid constraint is inserted into the contact zone, the biasing system can continue to deflect and therefore absorb driver motion without overstressing the driver. In Engineering terms, one adds compliance by reducing the biasing system's spring constant. Figure 6 shows the Cryotact contact which has compliance designed into it.

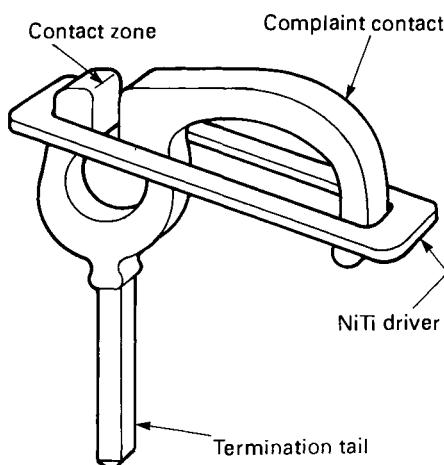


Figure 6: Cryotact contact assembly.

As one can clearly see, the benefits of adding compliance are the reduction of the driver operating stress and the increase in stroke. The designer is now not nearly as constrained in the area of mating pin tolerances. One may also have to adjust the normal force to compensate for the added compliance.

3. SME Alloy Interconnection Systems in Use

Quite a number of SMA connector systems are in use today in addition to the Cryocon® connector introduced in the preceding section.

The DIP connector is comprised of multiple lines of Cryotact® contacts, one of which is shown in Figure 6. This contact system is similar to a Cryocon®, but has a long, compliant member built in to prevent driver overstress and accommodate large pin tolerances. It is designed to interconnect Dual-in-line packages to circuit boards. Figure 7 shows a photograph of a DIP connector.

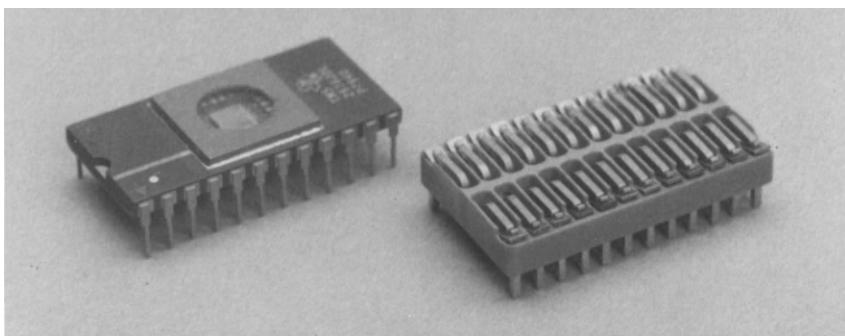


Figure 7: Photo of Cryotact DIP connector assembly.

The PGAP® connector is designed to provide an electrical path between large pin count Pin Grid Array packages and multilayer printed circuit boards. It utilizes a single piece of Ni-Ti to actuate many contacts simultaneously. Compliance is built into the contacts with the same resultant benefits as discussed before. Figure 8 shows an exploded view of a PGAP® connector, with an accompanying photograph shown in Figure 9.

A photograph of the Betaflex Connector, which was discussed briefly in Section 2, is shown in Figure 10.

The Cryocon®, Cryotact®, and PGAP® connectors and other types of SMA connectors are covered thoroughly in References 1 and 2. Readers interested in understanding specific performance parameters are urged to refer to these publications.

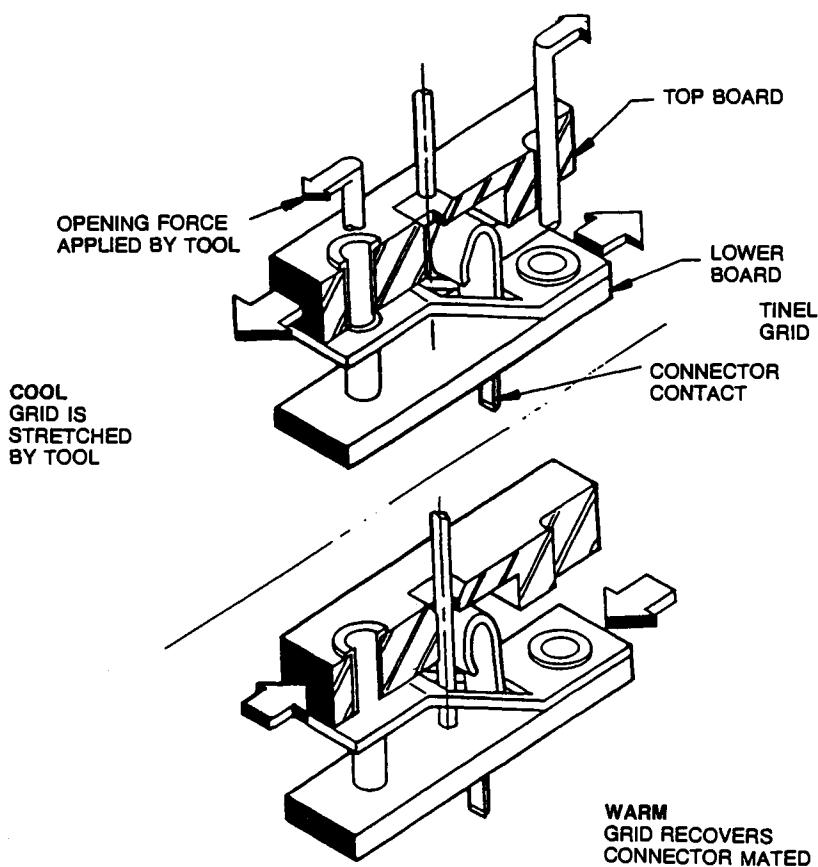


Figure 8: Exploded view of a PGAP connector.

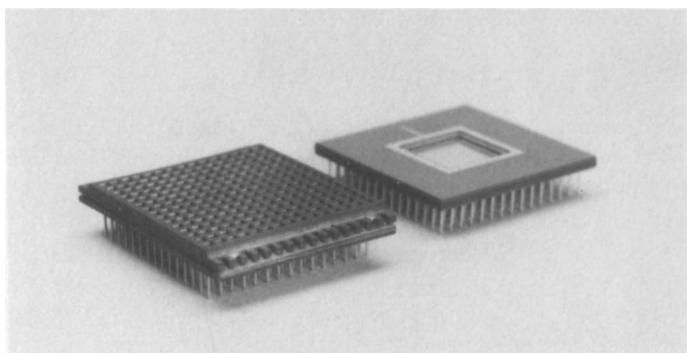


Figure 9: Photo of a PGAP connector.

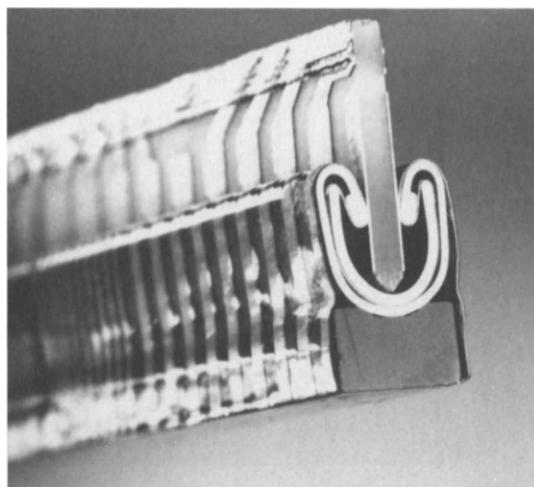


Figure 10: Photo of a Betaflex connector.

4. Design Constraints

The designer is faced with some typical design choices when deciding on the use of SME alloys for connectors. Some of these are:

- Is there enough volume to accommodate an SME driver?
- What is the projected cost of the system being built?
- What are the environments in which the system will operate? Do these warrant the use of an SME Alloy Connector?
- Are the total number of actuation cycles required above the practical limits of an SMA, in terms of net stroke required?

These are but a few of the many constraints which a designer must continuously address in incorporating a shape memory alloy into the design.

5. Conclusions

The proper design of an interconnection system with shape memory alloys boils down to the correct design of the actuator and biasing members. To do this properly, one must have a thorough understanding of the alloy being considered in terms of stress, strain and temperature. Once these are mastered, one can design connectors which greatly enhance the performance of the interconnection systems.

References

1. Flot, R., McCullough, J. and McGaffigan, T.: Shape-Memory Effect Alloys as an Interconnection Technology for High Density IC Packages, Electronic Components Conference, New Orleans, May (1984).
2. Krumme, J.F.: Electrically Actuated ZIF Connectors Use Shape Memory Alloys, Connection Technology, April, (1987).

Shape Memory Alloy Fastener Rings

Tom Borden
Raychem Corporation
300 Constitution Drive
Menlo Park, Ca. 94025

Shape memory alloys possess a unique combination of properties which afford them distinct advantages over conventional materials when utilized in fastening applications. These properties are particularly useful when the shape memory alloy is configured as a ring. Shape memory fastener rings are typically used for making permanent, non-separable joints. The applications are diverse. For example, shape memory rings may be used to:

- Terminate the electromagnetic shielding braid on a cable to a connector, providing excellent mechanical and electrical properties.
- Fix the location of a bearing or gear at any point on a shaft, if desired, locking in a controlled axial preload force.
- Assemble a cluster of radially disposed elements by compressing them with controlled uniform radial pressure.
- Seal the end of a thin walled metal tube or cup to a metal, ceramic, or plastic substrate.

This paper describes the properties of shape memory alloys as they relate to fasteners and goes on to show how these properties translate into features which are critical in fastening applications. These features are compared to those of conventional fastening techniques. The discussions focus on fastener rings, although most of the concepts are equally valid for fasteners which utilize shape memory in bending or torsion. For the sake of simplicity, discussions will be based on Ni-Ti alloys. The same concepts would apply for Cu-Zn-Al and Cu-Al-Ni alloys, though Ni-Ti is generally preferred due to its high recovery strains and excellent structural properties.

1. Intrinsic Advantages of Shape Memory Alloys for Fastening Applications

Shape memory alloys possess a unique combination of properties which afford them a variety of advantages over conventional fastening techniques. Typically, these advantages include; larger tolerance on mating parts, operator insensitive assembly, low temperature installation, immunity to vibration, shock, and thermal cycling, repeatability and lower installed cost.

1.1 Motion

The large recovery strains available in shape memory alloys make it possible to accommodate large tolerances and clearances on mating parts. The 6% to 8% shape memory strain is almost an order of magnitude larger than the strain typically available in thermal shrink fit or elastic installations. For comparison, the total differential thermal expansion between an aluminum shaft cooled to liquid nitrogen temperature (-196°) and an aluminum ring heated almost to melting (370°C) is only 1.2%. Elastic strains are limited to a similar number. Beta titanium, with its high yield strength and low modulus, can accept an elastic strain of 1.4%. The looser tolerances and more generous clearances allowed with shape memory, lead to cheaper, more producible parts that are easier to assemble.

1.2 Consistent Stress Generation

The stress generated by a ring in constrained recovery is a function of its unresolved recovery, or contact strain. Figure 1 exemplifies this. The stress rises rapidly up to a

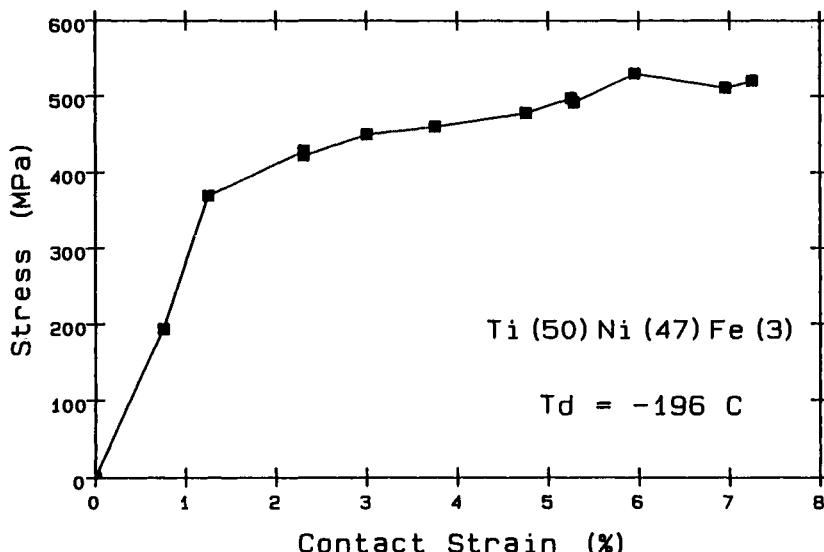


FIGURE 1

The hoop stress generated in a ring increases with increasing unresolved recovery. The pressure that a ring exerts on a substrate thus increases with increasing substrate diameter. However, at unresolved recoveries above 1.5%, the dependence of stress on substrate diameter is very small.

contact strain (unresolved recovery) of approximately 1.5%, at which point a plateau is reached where stress increases only slightly with increasing strain. Fastener rings would typically be designed to operate in the range above 1.5% unresolved recovery to take advantage of the flat region of the curve. The stress generated by such a ring is relatively independent of the tolerances of the parts it is recovered onto.

As installed, fasteners based on crimping, elastic deformation, and differential thermal expansion almost always operate on the elastic portion of their stress-strain curves. Because of the steepness of the curve, a small change in strain produces a large variation in stress. Consider a 0.1% variation in installed strain due to tolerances of the mating parts. In a steel part, the stress could vary by 207 MPa (30,000 psi). In a shape memory part, such a variation in strain would cause a variation in stress of less than 13.8 MPa (2000 psi).

1.3 Uniform Radial Pressure

Because shape memory is present throughout the material in a ring, the stress generated by constrained recovery is uniform around the circumference of a ring. Press fit and thermal shrink fit rings offer the same benefit. However, the amount of radial motion required or the logistics of installation often preclude the use of these techniques. When compared with multipoint crimping, split clamps, hose clamps, etc. the stress applied by a NiTi ring is much more uniform. This is why the sealing of cylindrical parts has been such a successful application of this technology.

1.4 Elastic Interference

Because of the low modulus and high recovery stress of Ni-Ti, installed shape memory rings have a large residual elastic interference with their substrate. This means that the rings will be highly resistant to the effects of thermal cycling and vibration. As the substrate changes dimensions under these conditions, the ring's excess of stored elastic energy allows it to follow the excursions while maintaining a relatively consistent level of stress.

A Ni-Ti ring with an installed stress of 413 MPa (60,000 psi) and a modulus of 55 GPa (8×10^6 Psi) would have 0.75% elastic interference with its substrate. This compares with 0.13% strain for 316 stainless steel at its yield stress and 0.4% for 6061-T6 aluminum at its yield stress (See Figure 2). A 0.05% decrease in the effective diameter of the substrate would decrease the stress in a steel ring by 40% and a Ni-Ti ring by 7%.

It should be noted that in a typical crimped, elastic, or shrink fit application, the conventional fastener will be operating at a stress well below yield as installed. Likewise, the initial elastic interference will be sub-optimum.

1.5 Installation Temperature

Ni-Ti can be installed at low or cryogenic temperatures. Various candidate alloys for fasteners are installed either by warming from cryogenic temperatures or by heating to 150°C - 200°C. When compared with welding, brazing, or even soldering, this can be a considerable advantage (Figure 3). Many applications of fastener rings involve heat sensitive materials such as electronic circuitry, polymeric materials, pyrotechnics, etc.

1.6 Operator Insensitive Installation

The installation of shape memory rings is extremely operator insensitive. Installation requirements are that the ring be positioned properly prior to installation and then be warmed to installation temperature. Typically, a thermochromic paint indicator is used on heat-to-shrink parts to signal when heating is complete. The performance of a

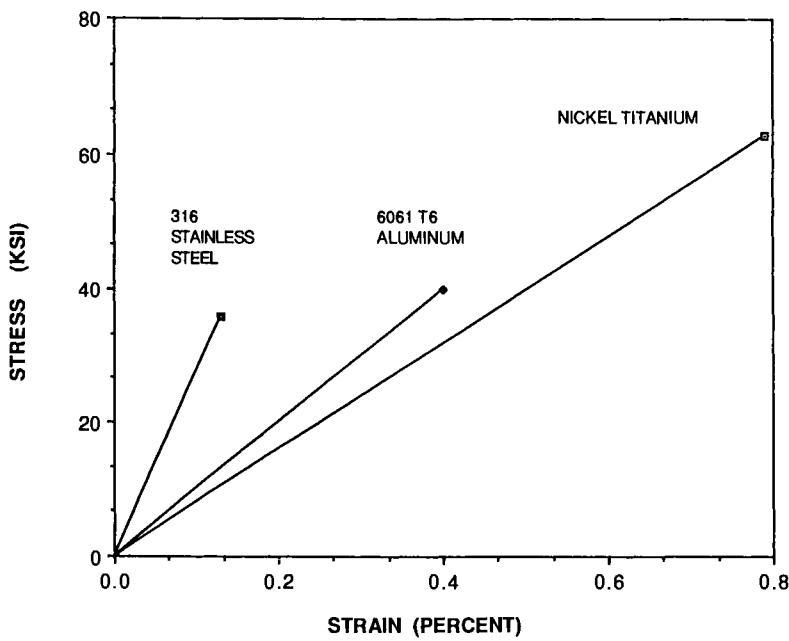


FIGURE 2

As a result of the high recovery stress and low elastic modulus, an installed ring will have a large elastic interference with its substrate. This strain is compared with elastic strains of common engineering materials.

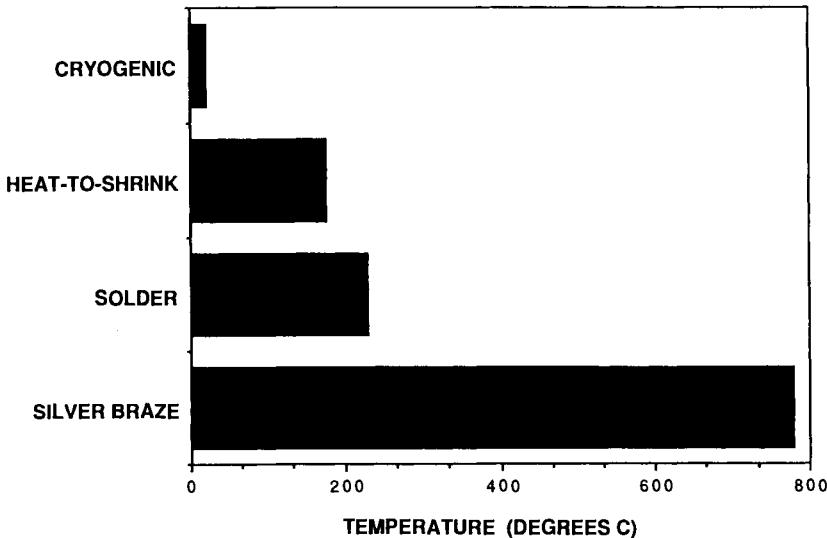


FIGURE 3

Shape memory rings offer low installation temperatures in comparison with soldering and brazing.

shape memory fastener is determined when it and its mating parts are manufactured, not when they are assembled.

1.7 Wide Operating Temperature Range

Shape memory alloys have a wide operating temperature range. This can run from -65°C to over 300°C. The low end is limited by stress decay due to the elastic modulus decline or due to the transformation to martensite. The top end is limited by thermally activated stress relaxation of the shape memory alloy. This range is much wider than that provided by soldering, adhesives or many mechanical systems.

1.8 Summary of Advantages

To summarize, shape memory alloy rings are attractive for fastening applications because they offer the following benefits when compared with conventional fastening techniques:

- Provide consistent stress generation
- Apply uniform radial pressure
- Produce large residual elastic interference
- Low installation temperature
- Operator insensitive installation
- Wide operating temperature range

2. Designing Fastener Rings

Shape memory fastener rings do not require sophisticated design techniques. The fundamental analyses are that of the radial motion required and then of the radial force to be exerted.

2.1 Strain Analysis

When examining a potential shape memory application, a motion analysis is usually the first step. The recovery strain available from a shape memory ring is typically split over five elements (Figure 4). Tolerances and clearances must be held tight enough that the sum of them does not exceed the motion available.

Given tensile strain relationships for a particular alloy, ring inside diameter motion may be roughly calculated with the equations given in Figure 5. The tensile strain data required would be the springback strain versus expansion strain and the recovery strain vs expansion strain.

2.2 Stress Analysis

Because shape memory fasteners are designed to operate on the plateau of the stress versus unresolved recovery curve, solid mechanics relationships derived for fully plastically yielded metals can be applied. For fully yielded cylinders these equations are quite simple. The radial pressure, P_r , between a ring and its substrate is given by:

Thin Walled Cylinders

$$P_r = \sigma_r t / r$$

Thick walled Cylinders

$$P_r = \sigma_r \ln(D/d)$$

where t is the ring thickness, σ_r is the recovery stress, r is the ring radius, and D and d are the outside and inside diameters respectively. Given the plateau stress for a given

MOTION ALLOCATION

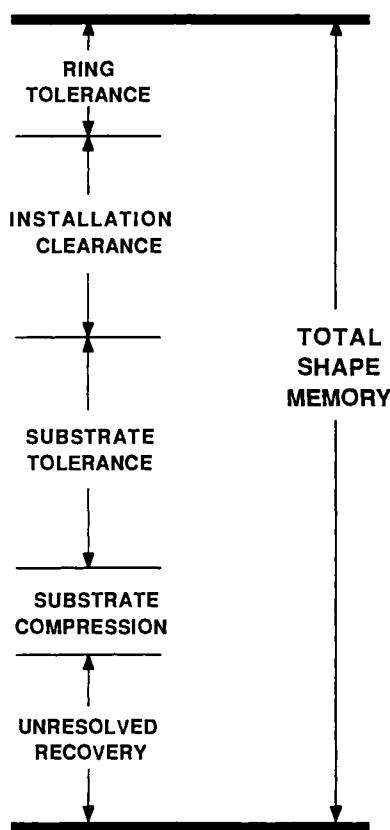


FIGURE 4

The limited amount of motion available from a shape memory alloy ring must be judiciously allocated to insure ease of assembly and that sufficient unresolved recovery is left when the largest ring is installed on the smallest substrate.

alloy, and the geometry of the shape memory ring, the radial pressure under the ring can be calculated. This interface pressure between the ring and substrate can then be used in other analyses to predict the behavior of the particular assembly.

2.3 Testing

The simple stress analysis described above can give a good idea of performance as installed at room temperature. However, for most applications the worst case occurs at

$$\text{EXPANSION STRAIN} = \frac{(\text{MANDREL DIAMETER} - \text{MACHINED ID})}{(\text{MACHINED ID} + \text{WALL THICKNESS})}$$

$$\text{SPRINGBACK STRAIN} = \frac{(\text{EXPANDED ID} - \text{SPRINGBACK ID})}{(\text{RECOVERED ID} + \text{WALL THICKNESS})}$$

$$\text{RECOVERY STRAIN} = \frac{(\text{EXPANDED ID} - \text{RECOVERED ID})}{(\text{RECOVERED ID} + \text{WALL THICKNESS})}$$

$$\text{UNRESOLVED RECOVERY} = \frac{(\text{INSTALLED ID} - \text{RECOVERED ID})}{(\text{RECOVERED ID} + \text{WALL THICKNESS})}$$

FIGURE 5

These conventions for calculating ring strains, while being rather simplistic, have proven adequate for designing rings and correlating data.

the minimum operating temperature. This is because the stress in a shape memory alloy declines on cooling towards its zero stress M_s temperature. It is difficult to analytically predict the combined effects of stress induced martensite, modulus decay, differential thermal expansion, and substrate compliance. For this reason, assemblies incorporating shape memory fasteners should be functionally tested at their minimum operating temperature.

3. Application Examples

A wide variety of fastener applications for Ni-Ti have been introduced, and it is worth reviewing a few.

3.1 Braid Termination

Heat-to-shrink rings are used to permanently attach cable EMI/EMP shielding braids to backshells of connectors, as in Figure 6. A ring can be recovered in 5 seconds using a resistance heater which drives a large current at low voltage through the ring. This braid termination provides a joint with a DC resistance of less than 1 milli-ohm and a tensile strength of 890 newtons (200 pounds). High frequency shielding is optimum due to the 360 degree peripheral contact between the braid and adapter. The joint is unaffected by ambient temperatures from -65°C to 150°C, mechanical shock, and vibration.

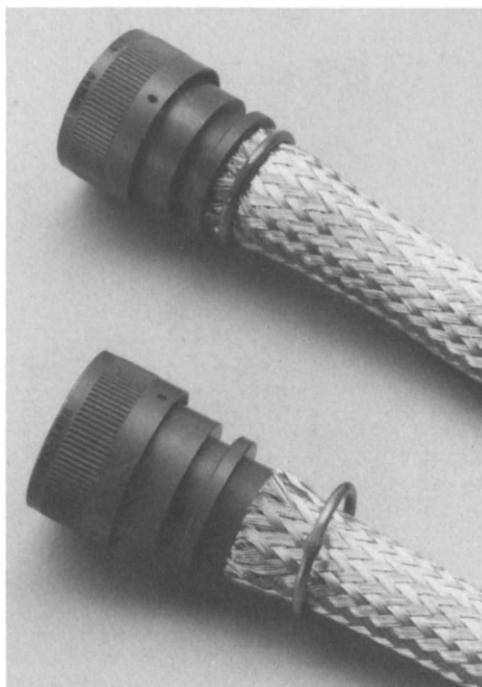


FIGURE 6

A shape memory alloy ring terminates tin plated copper braid to a cylindrical adapter by compressing the braid against a controlled surface on the adapter.

The primary advantages of this system are:

- Operator insensitive installation vs soldering, hose clamps, and other mechanical fasteners.
- Speed of installation.
- Small size and weight.
- Long term reliability and resistance to severe environments.

3.2 Bearing Preload

When assemblies are made up on shafts it is often necessary to have an axially adjustable locating collar to take up the accumulated tolerances of the components on the shaft, Figure 7. In addition, it is often necessary to apply a controlled axial preload to the stack of components for the proper functioning of bearings, etc. A shape memory collar can fill these needs. The installation procedure is to slip the collar onto the shaft, preload the collar and components, and heat the collar to lock it in place. This problem has traditionally been addressed with threaded fasteners or thermal shrink fit rings. Threaded fasteners have several limitations. They are bulky. They often have an eccentric mass that makes them unsuitable for use on high speed rotating equipment. Also, they must be torqued precisely on installation in order to

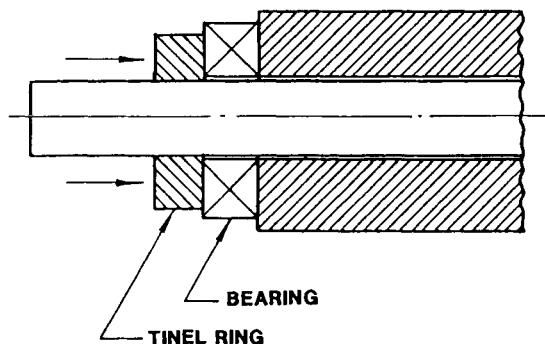


FIGURE 7

A Heat-To-Recover ring is pushed against a stack of components on a shaft and then recovered. This allows the accurate axial preloading of bearings, independent of the accumulated tolerances of the components on the shaft.

function properly, and even then are susceptible to loosening. Thermal shrink fit collars require close tolerances on the shaft and the collar in order to insure a consistent grip on the shaft. The assembly sequence is restricted by the short time available to install the collar before it cools down. In very small assemblies it becomes virtually impossible to use this technique.

This is exemplified by the problem encountered in the assembly of a gyroscope. With a 1.93 mm (0.076 inch) diameter spindle, the installation time for a heated aluminum ring was a fraction of a second. The use of a heat-to-shrink shape memory ring made this an easy installation.

3.3 Missile Retaining Ring

This ring is used to control the axial location of components within a missile body. See Figure 8. The installation sequence is as follows. An aluminum ring is dropped into the missile body on top of the components. It is then axially preloaded with a press, up to the desired force. A cryogenic shape memory ring is then dropped over the body and centered over the locating ring. As the Tinel ring recovers, it swages the body against the inner ring, locking it in place. This approach offers advantages over mechanical methods in that:

- It allows for infinite adjustability of the locating ring position.
- Axial preload can be precisely controlled.
- The missile body need not be thickened to accept threads or grooves.

3.4 Radial Assembly

Another area in which shape memory alloys offer an advantage over a thermal shrink fit is in the radial gripping of angularly segmented assemblies. An example of this is the accelerometer pictured in Figure 9. This was originally designed with a thermal shrink fit ring. However, the tolerance stack-up of the components made it difficult to get a consistent clamping force. The switch to a heat-to-shrink shape memory ring

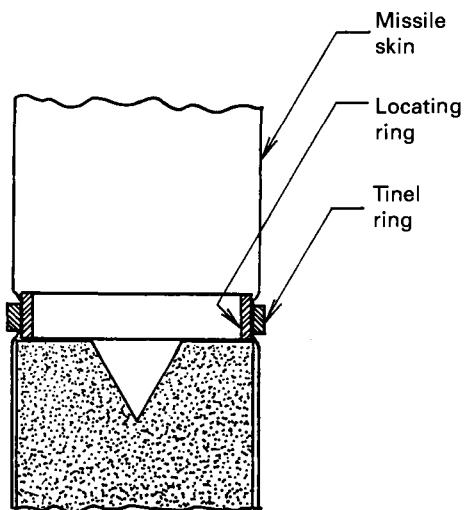


FIGURE 8

An aluminum ferrule is pressed against the shape charge of a missile. While the preload is maintained, a shape memory alloy ring is recovered on the skin of the missile, swaging the skin against the ferrule, thus locking the ferrule and shape charge in place.

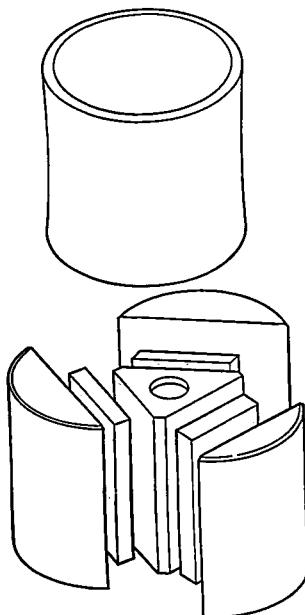


FIGURE 9

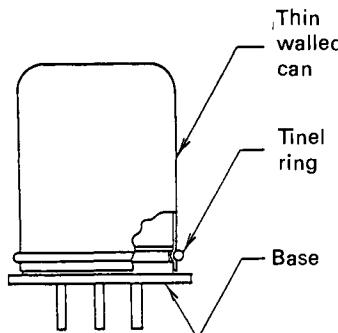
The components of a shear type piezoelectric accelerometer are compressed together by a shape memory alloy band.

resulted in consistent product performance with reduced rework rates. This, combined with the looser tolerances now permissible, resulted in significantly reduced manufacturing costs.

3.5 Hermetic Seal

It is often difficult to seal thin walled metal cylinders to metal, ceramic, and plastic substrates. This type of construction is widely used in the packaging of gyroscopes, fuses, detonators, IR detectors, and electronics packages. Organic sealing methods such as O-rings, gaskets, adhesives, and sealants do not produce true hermetic seals. Soldering, brazing, electron beam welding, and other sophisticated welding techniques are currently used for metal to metal seals. However, these techniques are expensive and often have yield problems. The risk of overheating the package contents is a major concern.

Shape memory alloys offer an alternative sealing technique. A shape memory ring is used to swage the cylinder into a groove cut in the base, as shown in Figure 10. The detail shows how two seals, one redundant, are achieved at the corners of the groove



Seal detail

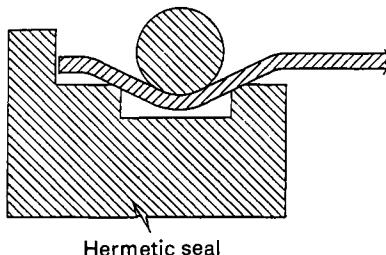


FIGURE 10

A package is hermetically sealed by a shape memory alloy ring which swages a deep drawn aluminum cup against an aluminum header.

where the material of the cylinder and the base are plastically deformed into one another. As an example of the effectiveness of the sealing; 3.6 cm (1.4 inch) diameter aluminum canisters were sealed to aluminum bases. Before and after thermocycling between -65°C and 150°C, the vacuum leak rates were less than 1.8×10^{-8} std cc/sec He (the resolution of the test equipment).

Some advantages of this sealing method are:

- Simple installation tooling.
- Low temperature installation.
- Loose tolerances on mating parts.
- Low leakage rate.
- Allows sealing of dissimilar materials, such as an aluminum canister to a non-metallized ceramic base.

4. Conclusion

Shape memory alloy rings offer significant advantages over conventional fastening technologies. The use of shape memory rings to assemble a particular product can result in reduced manufacturing costs, while at the same time providing superior product performance.

Shape Memory Alloy Applications In Space Systems

L. McDonald Schetky
Memory Metals, Inc.
83 Keeler Avenue
Norwalk, CT 06854

Space technology encompasses an incredibly broad spectrum of engineering and scientific disciplines. The prime movers for placing payloads into orbit, or to send missions into deep space, use liquid and solid rockets which require controls, seals, joints, couplings and actuators of a broad variety. Many of these are candidates for shape memory devices. The satellites themselves, whose purpose may be commercial, scientific or military, also involve passive and active components which could exploit the unique properties of shape memory alloys. As we move into the final decade of this century it is clear that large space stations will be placed into orbit as platforms for scientific and engineering studies. These large truss structures, with relatively low rigidity, will encounter various vibration problems which may also be amenable to shape memory controls. It is the purpose of this paper to describe briefly some of the more promising applications for shape memory alloys in this demanding environment.

The shape memory alloys which at this point in time can be considered commercial are Ni-Ti and variations on the Ni-Ti binary, Cu-Zn-Al and Cu-Al-Ni. The Ni-Ti alloys offer high strength, good corrosion resistance and large recoverable strain. They are usually used in one-way memory applications such as joints and couplings. As a consequence of the difficulties attending the processing of these alloys they are the most expensive of the shape memory alloys available. Though possessing a lower recoverable strain, the copper based shape memory alloys have the advantage of lower costs, making them attractive for actuators. Where higher temperature actuation is required, Cu-Al-Ni alloys are preferred since they can be formulated to give a recovery temperature of up to 190°C. A newly announced alloy from Japan would seem to offer another option for coupling and fastener use, the Fe-Mn-Si system modified with Ni, Co and Cr to provide properties similar to the 18-8 grades of stainless steel. The selection of a particular alloy system will depend on the recoverable strain required, fabricability, temperature regime, recovery force, corrosion conditions, thermal stability, resistivity, and cost.

One of the limitations to the installation of large platforms in outer space is the payload of the shuttle, requiring that sub-structures be transported to the required orbit for assembly by workers in zero gravity, with some limit on the duration of their time outside of the protective environment of the satellite. As such, means for assembling structures with minimum time and effort are required. Shape memory connectors are being developed under a NASA program to facilitate the joining of composite tube

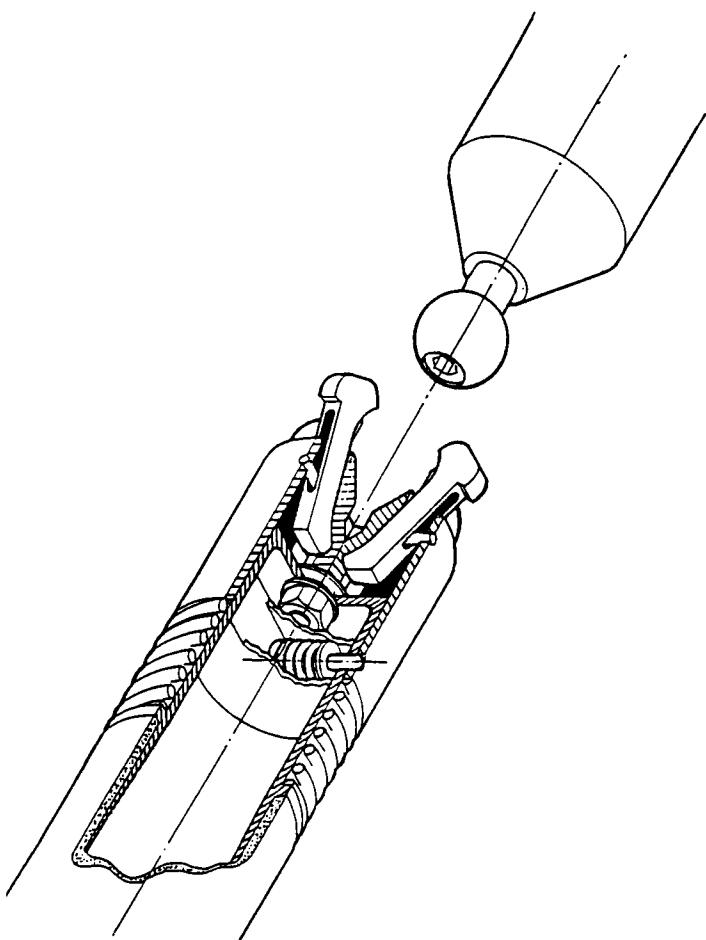


Figure 1: A ball and socket design for assembling composite tubing.

structural members. The requirements are ease of assembly and disassembly, minimum energy requirement for the actuator function, and compatibility with composite materials. Fiber reinforced epoxy tubing is one of the likely construction materials. Joining these materials with penetrating fasteners such as rivets or bolts seriously degrades their properties and the use of shape memory sleeve joints offers a possible solution. Heating the sleeve to cause shape recovery would pose a problem in view of the limited availability of electric power in the satellite. An alternative is to attach the structural tube ends to a joining device which can be used in space to couple the tube nodes to create the required truss form. The shape memory actuated tube ends would be assembled to the tube on the ground using shape memory sleeves. In space these couples could be assembled with automatic catches and then, if required, be released by a shape memory latching device. Figure 1 shows a ball and socket design which automatically latches on contact and can be released by electrically heating the shape memory actuator. Figure 2 illustrates this same principal

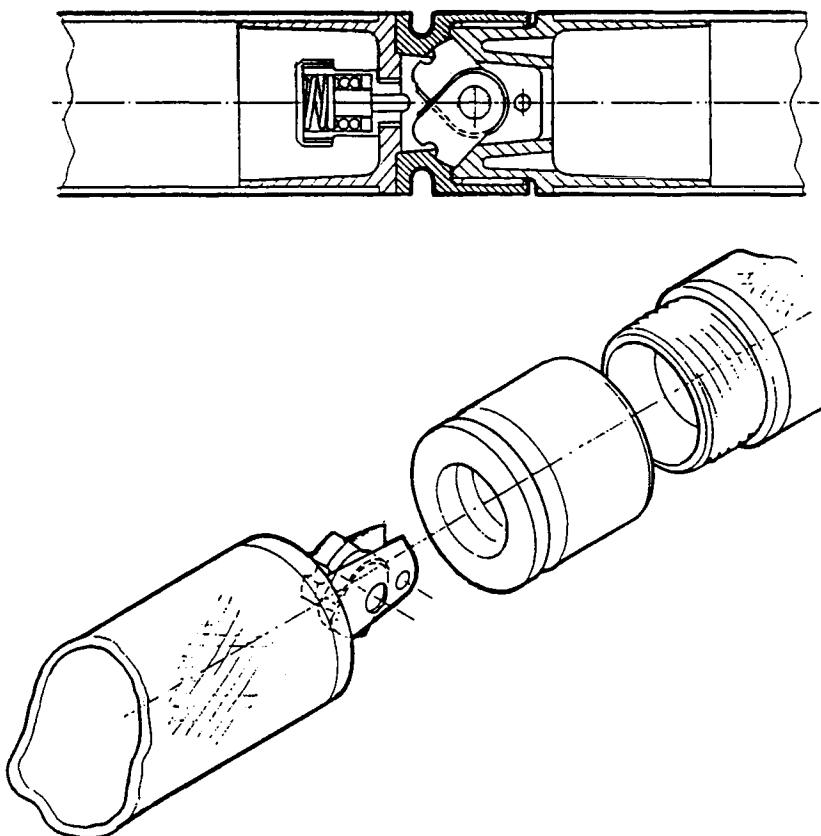


Figure 2: A sprag-type of coupling for assembling composite tubing.

using a sprag type of coupling, and Figure 3 shows a shape memory operated latching system for cross-tubular geometries. In order to test these units in a simulated zero gravity environment they have been designed with seals to permit their operation in the NASA neutral buoyancy tank.

In the absence of sunlight, outerspace temperatures are in the deep cryogenic region. An earth orbiting object exposed to direct sunlight will reach temperatures in excess of 100°C. In both cases the temperature will, of course, be cyclic as the object exposes alternate faces to the sun as it moves in its orbit. By using surface coatings of varying emissivity and reflectivity the bulk temperature of an orbiting space satellite or station can be controlled to a reasonable degree. Heat balance will be affected by conduction of heat from one component to another, by the use of radiator systems, and by heat generated by equipment or personnel in a space station. For purposes of design, the temperature limit employed in military systems seems a good approximation: -50° to 70°C. A coupling which requires activation for either latching or unlatching would therefore be electrically heated to a temperature in excess of the 70°C ambient.

As much as the present despoiling of our beaches by waste may disturb us, we seem intent on creating a similar dump in the space occupied by orbiting satellites. French

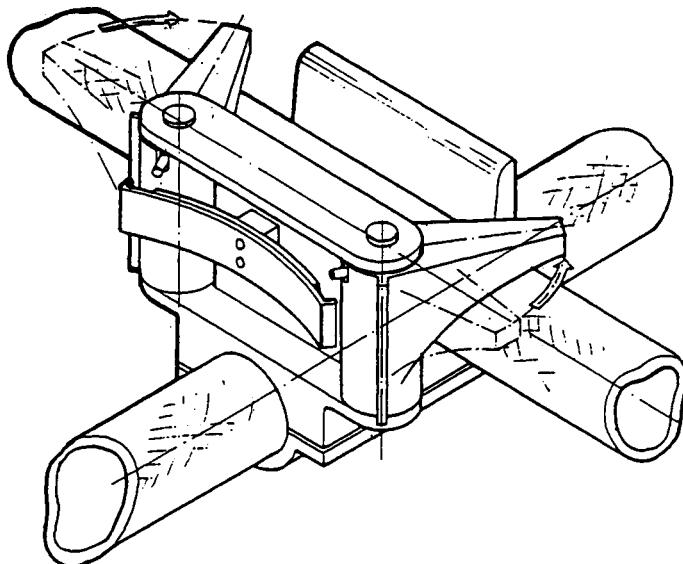


Figure 3: A latching system for assembling cross-type of structure in composite tubing.

and Russian launching systems are characterized by the eventual destruction of their launch rockets in apogee, leading to a virtual man made asteroid belt of junk. Although the circling debris is comprised of relatively small particles, their high velocity can cause damage to satellites in that same orbit. As a result of this condition some form of protective shroud will increasingly be required to shield satellites and space laboratories from these man made mini-meteorites. One approach is to make use of the flexibility of shape memory actuators to deploy shelters of various geometries. One version, shown in Figure 4, uses a folding box design with a shape memory torsion rod indicated by "A" for erecting or folding the stacked box, depending on the requirement. The volumetric packing of this structure, the ratio of folded to open volume, is 72 as shown in figure 5. An even more attractive ratio is possible with the circular geometry shown in Figure 6.

For lighter couplings than the monolithic SME sleeve type, an alternative is being examined involving the in situ winding of pre-strained SME wire around the end of the composite tube, as shown in Figure 7. When the wire is heated by the passage of electric current it recovers its original length, putting the I.D. of the tube end in compression. Techniques for overcoming the radial compliance of the tube, wire to wire friction and the wire-epoxy friction are being explored.

Large space structures are being designed for use as manned laboratories, as space manufacturing stations, and as platforms for Strategic Defense Initiative (SDI) type activities. These very large truss-like structures will have problems of stability, as a result of induced vibrations from the manned activities, motions induced by docking of shuttles or other vehicles, or vibrations involving a shift in the platform position required in tactical maneuvering.

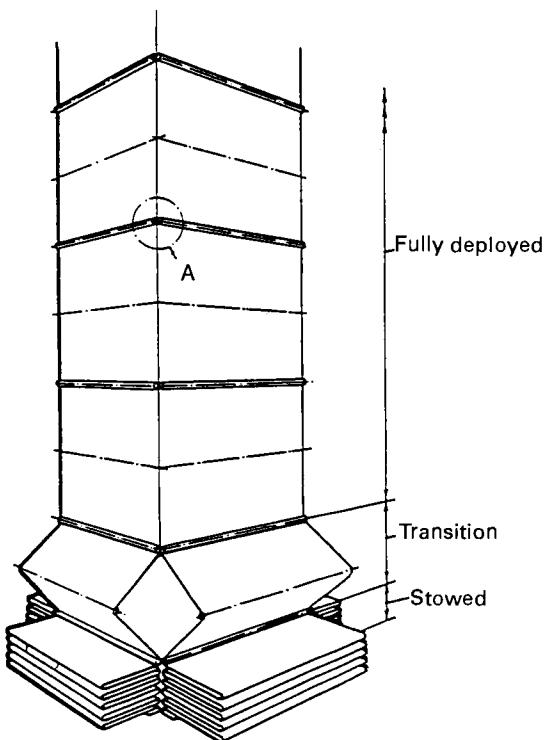


Figure 4: A folding box type of protective shell erected using a shape memory alloy actuator.

Studies to date on very large space structure designs have shown that feed back controls using small vectoring rockets will not be sufficient to provide positional stability, and that the structure must have some sort of inherent passive damping. The passive damping may be provided by some form of shape memory coupling or static member which will exploit the unique high damping properties of the martensite phase in shape memory alloys (discussed later in this book). Since a large flexible structure will have many modes of vibration at low frequency, techniques will be required for sensing the vibration and then cancelling or reducing its amplitude by some passive or dynamic means.

It has been shown that the introduction of an out of phase vibration to cancel existing vibrations can be accomplished using feed back control of hydraulic motor actuators. This concept has recently been applied to the isolation of vibration of the engine, transmission and rotor from the passenger space in a British designed helicopter. For the application of this concept to space structures some alternative to the heavy hydraulics would have to be employed, arguing for the use of SME "motors" for the generation of the anti-phase vibration.

The analysis of a large, complex space structure for possible modes of vibration and their interaction is a difficult and often intractable problem. Using methods of finite element analysis on models with a simpler structure should establish the feasibility of

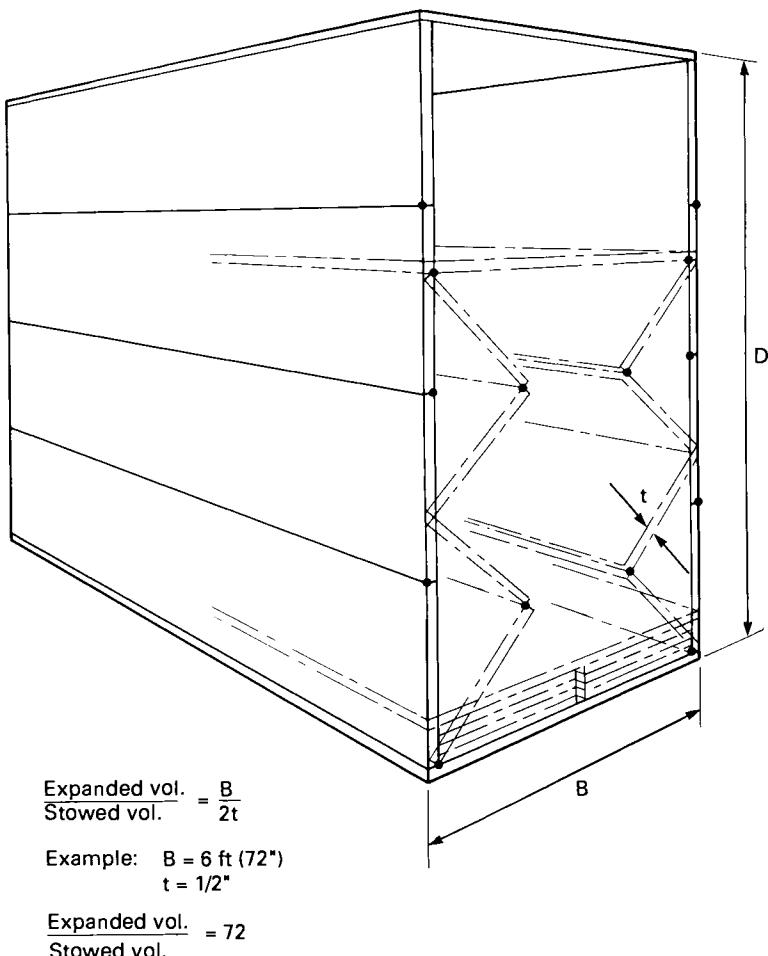


Figure 5: The volumetric packing ratio of the structure shown in Figure 4.

this concept. Two ways of countering vibrations can be envisioned: active feed back of anti-phase vibrations, and the use of variable stiffness control. In the former, sensors placed strategically throughout the structure would provide signals to a control amplifier which would energize an appropriate SME motor to provide a vibration 180° out of phase with the sensed vibration. The energy required in this type of control has proven to be modest; well within the energy available on a large platform.

The modulus of elasticity of shape memory alloys changes by a remarkable 40% when transforming from the fully martensitic phase to the fully beta BCC phase. This would represent a very large change in the stiffness of a structural member made of these materials. Since these alloys have a relatively high density they would be

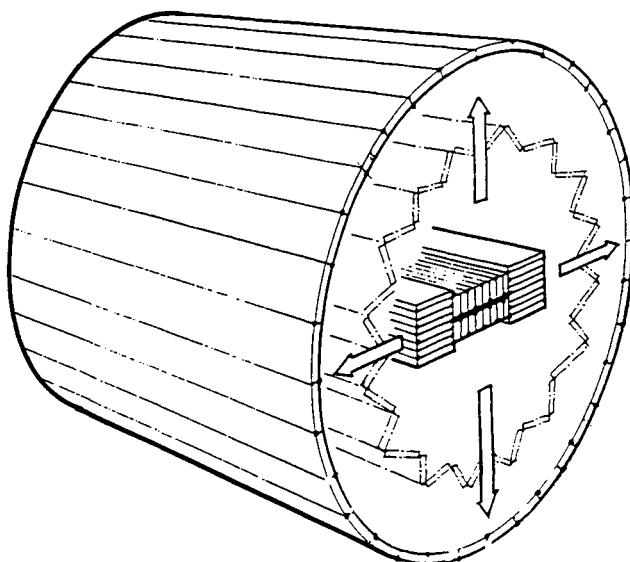


Figure 6: A circular geometry of protective shell.

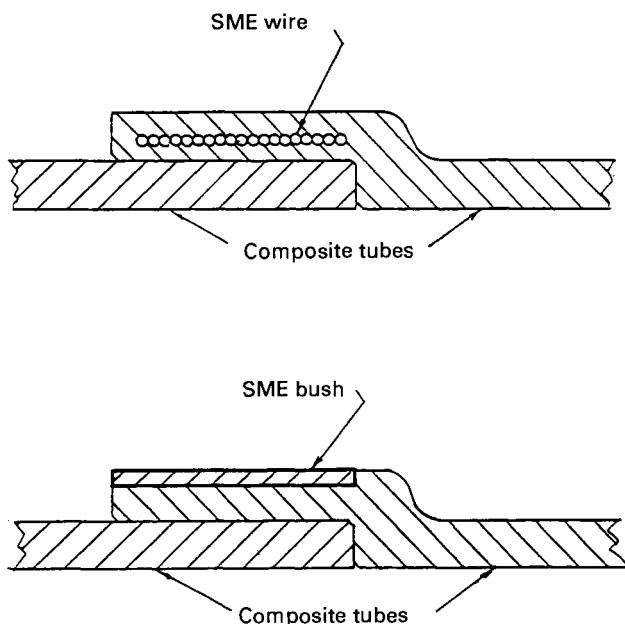


Figure 7: Lightweight shape memory alloy couplings for composite tubes.

incorporated at some volume percent in a composite material tube, and upon heating of, for example, imbedded wires, the stiffness or compliance of the structure could be changed, and as a result, the vibrational modes which are excited. In use, the variable stiffness member would be controlled by the same system of vibration sensors and control amplifiers envisioned for the active control hydraulic system mentioned above. When a damping alloy is incorporated in a structure to provide passive damping, it is worth nothing that the amount of damping material added to the structure does not have to be excessive. It has been shown in other studies of high damping materials¹ that the insertion of 10% by volume of a damping material in the structural component will yield about 80% of the benefit that a total replacement with the damping alloy would provide. This is important since the martensitic structure of the SME damping alloys has both a low modulus and yield strength, and as such, might not be suitable as a stress bearing member of the system in which damping is required. Insertion can be by adhesive bonding, brazing, interference fit or mechanical fastening.

Damping in shape memory alloys arises from the high mobility of the martensite variant plate boundaries which results in their motion when subjected to an external stress. If the stress is oscillating, i.e. vibratory, there is a resultant energy absorption which produces damping. Other high damping alloys have been developed which owe their damping to the motion of ferro-magnetic domain walls, for example the Fe-Cr-Al system known as Silentmetal, and damping due to anti-ferromagnetic domain twins in the Cu-Mn alloy known as Incramute. The latter has good damping at high stress, and the Silentmetal family of alloys has good damping at low stress. The shape memory alloys, Cu-Al-Ni, Cu-Zn-Al and Ni-Ti, although not ideal, offer high damping performance over an acceptable range of temperature and stress.

Recent problems with the shuttle booster rockets has disclosed a potential use for high integrity SME seals. The very high recovery forces possible in shape memory alloys and the retention of this force to high temperatures, argues for their consideration as protective seals in the booster rocket ring segment joints. A Belleville washer type configuration and a chevron type seals geometry have been demonstrated. Interest has been expressed in evaluating SME seal as a replacement for the currently used elastomeric materials for the next generation of large solid fuel booster rockets.

Explosive bolts are used for many of the release devices currently used to detach various stages of a launch vehicle, such as release of the rocket on lift off, the casting loose of boosters, the deployment of drogues and parachutes, and the detaching of rocket nozzles at apogee. There is a constant danger of accidental discharge of these devices due to electrical storm activity, and, as such, there are several design proposals for replacing at least a portion of these bolt systems with SME actuated release mechanisms.

Considering that one of the first examples of a shape memory device for a space application was an antenna release actuator for a British satellite system in the late 60's the use of SME alloys in this environment has had a very slow acceptance. It is certain, however, as we have seen in the various presentations in the work shop, that the role of SME alloys as a solution to many complex engineering problems is now in a rapid growth phase, and that applications to space engineering will be a part of that growth.

Reference:

1. Study of Noise and Vibration Control Applications of Incramute High Damping Alloy: INCRA Report 220 Wyle Laboratories, El Segundo, CA 90245.

Actuator and Work Production Devices

T.W. Duerig, D. Stöckel and A. Keeley
Raychem Corp.
300 Constitution Drive
Menlo Park, CA 94025

Thus far we have discussed free recovery, in which there was a recovery strain but no stress, and constrained recovery, in which there was a recovery stress but no strain. *Work* is defined as $\int \sigma d\epsilon$, so in both cases no work was done. Here we consider the case where the shape memory material recovers against a stress, doing work: in the ideal case, a shape memory wire or spring lifting a weight (Figure 1). If the stress

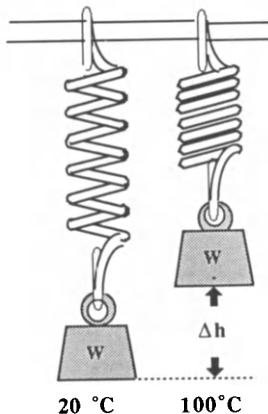


Figure 1: The classic and most simple SMA actuator consists simply of a shape memory spring with a weight on the end: as the spring is heated, the weight is lifted and when the spring is cooled, the weight stretches again.

applied by the weight (σ_0) is less than the recovery stress of the alloy, the weight will be lifted and the amount of work done will be $\sigma_0 \Delta\epsilon$. This of course describes only the work production during heating. If the applied stress is greater than the martensitic yield strength of the material σ_y^m , the stress will restretch the spring upon cooling through M_s causing a repeatable two-way motion. In fact most actuators are multiple cycle applications and require such a *biasing* stress for resetting.

Actuator devices are certainly the most complex type of shape memory application, but may well have the greatest commercial potential. In Part III and IV of this book, several

papers will detail design issues and specific examples. The purpose of this chapter is to define terms and to introduce some generalities of actuators, especially as they relate to work production and efficiency.

1. Schematic Description of Work Production:

Two σ - ϵ -T perspectives are needed to examine the work production event: stress-strain (Figure 2) and strain-temperature (Figure 3). In Figure 2, an SMA is deformed

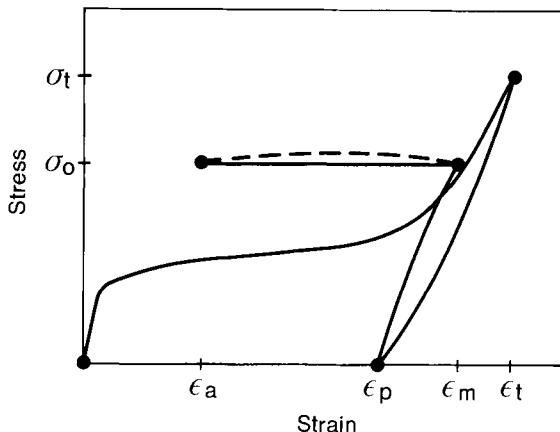


Figure 2: A stress-strain perspective of the actuator or work production event showing deformation, unloading, loading to the applied stress level- σ_o , and heating to recover to ϵ_A . Subsequent cooling (dashed line) returns the deformed shape, ϵ_M .

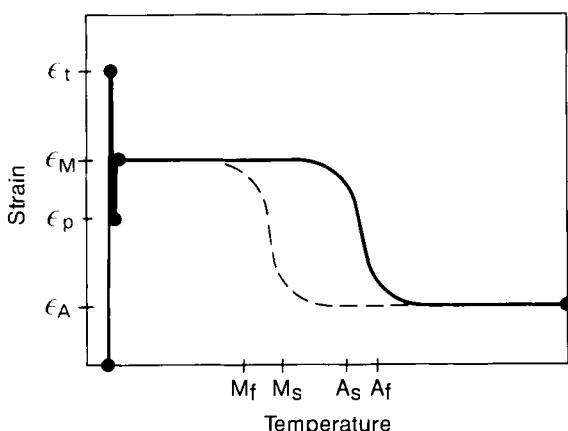


Figure 3: The strain-temperature perspective of the event described in Figure 1, showing the four transformation temperatures.

just as in the case of free recovery, but now a stress is applied after deformation. The SMA recovers upon heating to the austenitic strain- ϵ_A and is deformed again to the martensitic strain- ϵ_M upon cooling. Figure 3 shows the same event, this time with the transformation temperatures highlighted. Note that the initial deformation step could have been eliminated in Figures 2 and 3; one could equally well have just directly applied the load and reached the coordinate (ϵ_M, σ_0) . In some cases, however, there are advantages to *prestraining* beyond the desired starting point. Specifically one can often achieve larger net recovery strains, reset with lower stresses, improve fatigue resistance, or cause a two-way effect.

A second very important simplification in these figures is that the stress upon cooling is the same as that upon heating. Although work is done by the SMA upon heating, the same amount of work is being done by the weight to the SMA upon cooling - the net work output per cycle is zero. In many cases (particularly with thermal actuators) one only wishes to do work during the heating cycle, but often one wishes to have a net work output, meaning that the applied stress upon heating should be greater than that upon cooling. By analogy, one can consider a mine elevator that lifts coal, but needs some of the coal's load as ballast to return to the bottom for the next load. What we have described in the above paragraphs is analogous to an elevator that carries coal up and down a mine shaft without unloading at the top, when in fact we want to carry more coal up than down. To maximize the efficiency of the elevator, one would like to unload as much as possible and minimize the ballast necessary for returning to the bottom (*the resetting or biasing stress*). In terms of a shape memory alloy, this means minimizing σ_m^m . This concept is shown in Figure 4, with σ_H being the stress resisting recovery and σ_L being the stress used for resetting. (In Figures 2 and 3, both σ_H and σ_L were equal to σ_0 .)

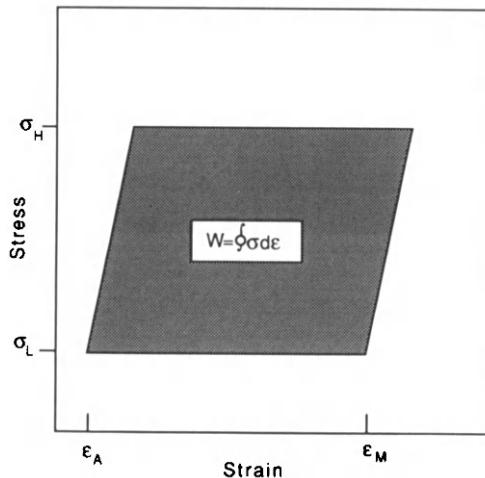


Figure 4: A typical actuator recovers against a higher load (σ_H) than is applied during cooling for resetting (σ_L), resulting in a net work output (shaded area).

One should be aware that σ_L can in fact be below σ_m^m but still be sufficient to reset in many cases. For reasons which are not entirely clear, plastic deformation will occur below the martensitic yield strength in many materials if one applies the load while

cooling through M_s : A 100 MPa stress may be insufficient to reset a spring if it is applied below M_s , but sufficient if applied while cooling through M_s . For example, a tensile stress of 75 MPa will generally not significantly deform Ni-Ti below its M_s temperature, but will stretch if the load is applied while cooling through M_s .

2. Design Data:

The complexity and variety of actuator applications make it very difficult to generalize design properties; different applications require quite different properties. Here we will treat only the most general cases. Examples of how to treat complex, "real life" situations will be given in later chapters.

2.1 Work Output:

Clearly one of the most important parameters characterizing a potential actuator alloy is the allowable applied stress- σ_H . Figure 5 shows how σ_H affects recovery strain in a fully annealed Ni-Ti-Fe alloy. In this particular alloy, substantial amnesia sets in at stresses of only 150 MPa. Clearly the work output of the heating cycle increases with

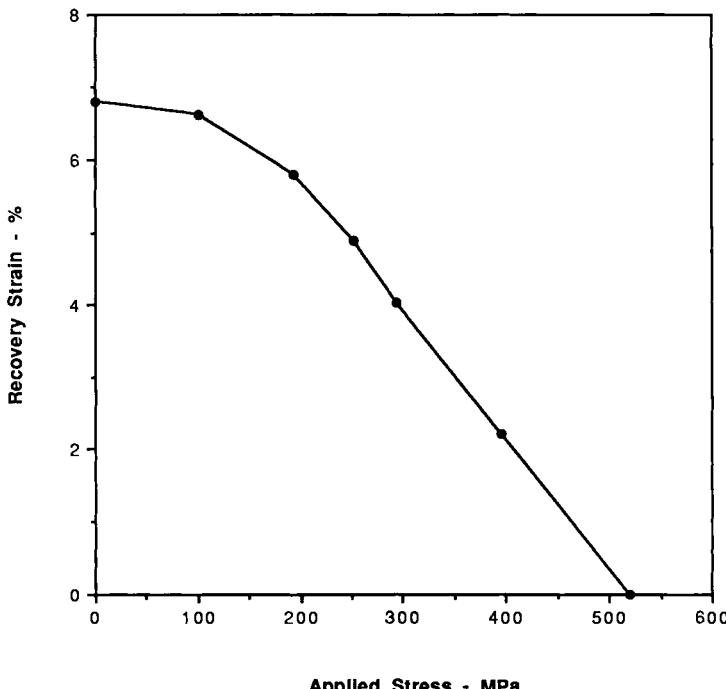


Figure 5: The recovery strain of a Ni-Ti-Fe alloy after a total deformation of 8% shows that the recovery strain ($\epsilon_M - \epsilon_A$) is reduced by an applied load. This particular alloy is fully annealed; stronger alloys would show full recovery against far greater stresses.

stress until the amnesia becomes too large, and thus a maximum in work output is found at about 300 MPa (Figure 6). In this case, if one wanted to maximize the work output of the heating cycle, one would design using a stress of 300 MPa. The alloy shown here is a low strength alloy, used because it very clearly demonstrates the principles of overloading. Ni-Ti alloys are available that can deliver work outputs well over 4 Joules/gram (upon heating, not net, as will be discussed later). The maximum work output of Cu-based alloys is substantially less, at some 1 Joule/gram. One must be aware that the work outputs reported in Figure 6 are single cycle measurements. As will be reported in later chapters, fatigue degrades SMA performance: transformation temperatures can shift, amnesia will accumulate with every cycle (a phenomenon called *walking*) and the SMA may even break. If one expects to operate an actuator for hundreds or thousands of cycles, one cannot operate at the maximum work output conditions.

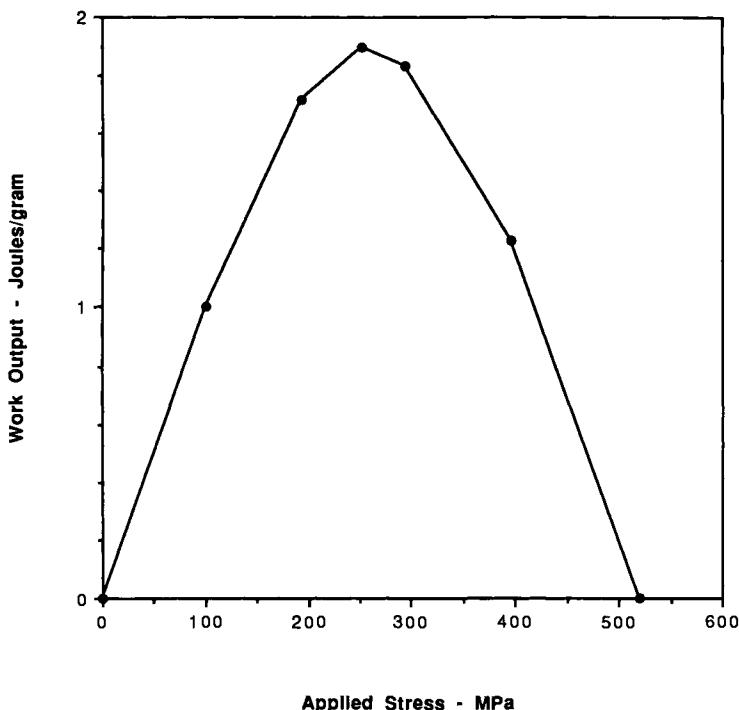


Figure 6: The work output of the same Ni-Ti-Fe alloy shown in Figure 4 is plotted against the opposing stress level, showing that the highest work output is obtained when working against a stress of 250 MPa.

2.2 Stress Rate:

A second very important design parameter of which one must be aware is the shifting of transformation temperatures with stress (the stress rate) as has already been discussed in some detail (see Figure 13 of Chapter 1). This comes into play in several

ways. First it allows one to tune an actuator's transformation temperature by controlling the opposing stress. Second, it increases the effective hysteresis of an actuator since the resetting stress is generally much lower than that of the work production, or heating stroke. Finally, it can mean that the actuation temperature may shift depending upon variations in friction, due to wear, etc. In general, it is desirable for actuator alloys to have very large stress rates.

2.3 Biasing Springs:

All of the above examples assume that the resetting forces are constant. In fact this is seldom if ever the case. More often the resetting force is provided by a *biasing spring* with a certain compliance, normally (though not always, as will be seen in later chapters) exerting a force that increases linearly with strain as ϵ_A is approached. Together with the martensite and austenite stress strain curves, the biasing spring characteristics allow one to determine the characteristics of an actuator. Figure 7 shows the general approach. The load-displacement characteristics of the biasing spring are superimposed upon that of the SMA in its martensite and austenite. The intersection points of the biasing spring with the martensite and austenite curves give the end positions of the actuator, marked B and C in Figure 7.

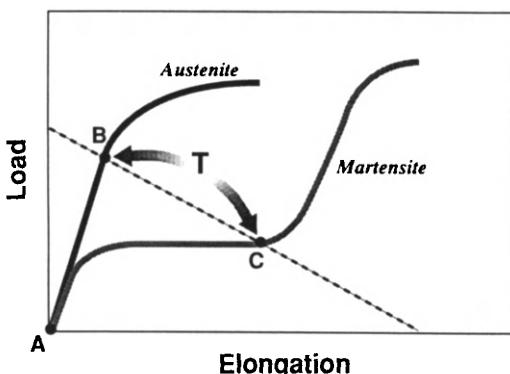


Figure 7: Figure 2 showed the actuation of a constant load, where as Figure 7 demonstrates recovery against a spring. In this case, motion ideally occurs between points B and C, the intersection points with the austenitic and martensitic stress-strain curves.

Note that the cycle shown in figure 7 provides no net work output: it corresponds simply to an SMA spring working at all times against a conventional spring. As discussed before, the heating cycle is more often expected to do external work, such as to move a switch or turn a valve. The superposition of an external stress to the biasing stress during only the heating cycle complicates matters, just as in Figure 4.

A second complication is that the stress-strain curves in Figure 7 are isothermally derived and we have already shown that shape memory alloys can behave quite differently when thermally cycled. The stress-strain curves that should be used in Figure 7 should be the cyclic curves. These are normally derived by applying various

fixed loads and thermally cycling while measuring strain; one then obtains two stress-strain curves corresponding to the two temperature extremes. Examples of these will be shown in a later paper by D. Yaeger.

3. Actuator Efficiency:

The *efficiency* of an actuator is defined as the net work provided by the actuator divided by the heat absorbed by the actuator. In many situations, especially heat engines and electrical actuators, efficiency can be of paramount importance though it is often difficult to quantitatively determine actual values.

To illustrate some basic concepts, it is useful to consider an idealized actuator alloy with the following characteristics:

1. A sharp recovery profile, approximated by $M_s = M_f = M$ and $A_s = A_f = A$.
2. A recovery strain that is independent of the applied stress.
3. No plastic deformation during transformation (i.e no walking).
4. A thermal hysteresis that is independent of the applied stress.
5. Transformational characteristics that are not affected by fatigue.

Once such an idealized material is understood, it is not too difficult to qualitatively extend the model to less ideal alloys.

Thermal cycling leads to the idealized strain-temperature profiles shown in Figure 8 (shown at two different stress levels: a high stress, σ_H , and a low stress, σ_L). Figure 9 is a crossplot of stress-temperature, showing shifting of the transformation temperatures **A** and **M** with the applied stress. Also shown in Figures 8 and 9 are the actuator cycles, corresponding to heating slightly above A_{σ_H} (A_s or A_f with an applied stress of σ_H) to slightly below M_{σ_L} (M_s or M_f with an applied stress of σ_L).

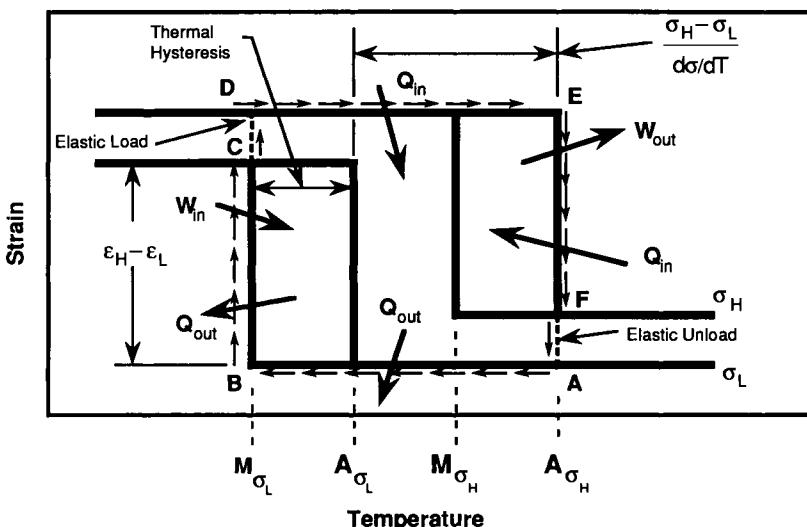


Figure 8: Two idealized strain-temperature recovery profiles, one at σ_L and the other at σ_H . A typical actuator would follow the path marked with arrows, from 'A' to 'F'. See text for details.

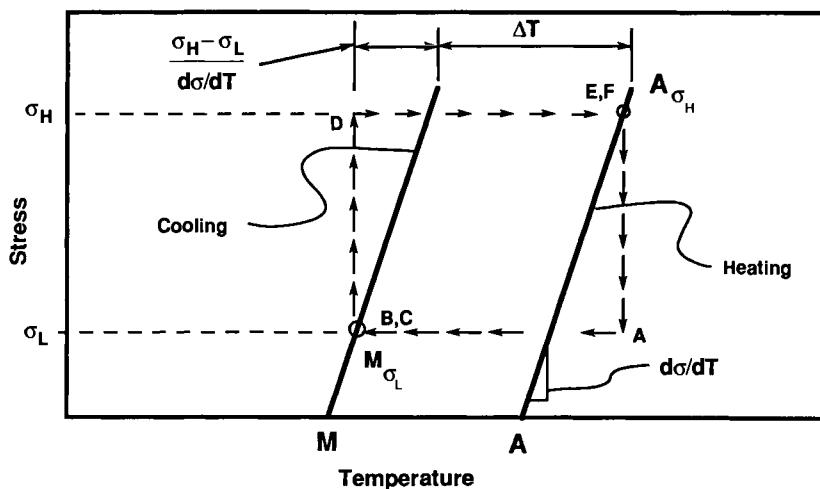


Figure 9: By "crossplotting" curves such as that shown in Figure 8, one arrives at the stress-temperature curve shown in Figure 9. Again the path of an actuator is shown by the arrows from 'A' to 'F'. See text for details.

In examining the cycle of such an actuator, start with warm austenite at a stress- σ_L slightly above the martensitic yield stress (point 'A' in figures 8 and 9). The austenite is then cooled at constant stress until it reaches M_{σ_L} (denoted 'B'). During this cooling sensible heat is removed from the austenite. As soon as the temperature reaches M_{σ_L} , the alloy begins to give up its latent heat of transformation and transforms to martensite at a rate determined by its ability to transfer heat to its environment. When fully transformed, the actuator is in state 'C'. It has required work to deform it, and its heat of transformation had to be removed from it during the process 'B'→'C'.

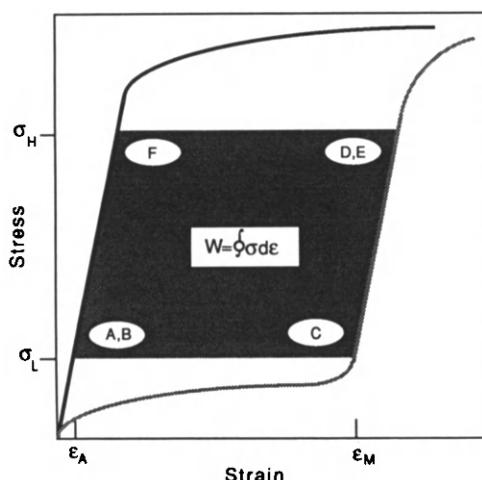


Figure 10: The actuator path described in Figures 8 and 9 are shown in terms of martensitic and austenitic stress-strain curves.

As the load applied to the martensite is increased from σ_H to σ_L , there is some elastic deformation occurring, as depicted by the trace 'C'→'D' in Figures 8 and 9. (In the elevator analogy, this corresponds to loading the car with coal at the bottom of the shaft, causing some stretching in the cables.) Heat is now added to bring the alloy to a temperature where martensite begins to transform to austenite at the higher stress level. The temperature of the alloy is increased from M_{σ_L} to A_{σ_H} by adding sensible heat to the process. The state of the alloy moves from state 'D' to state 'E'. As more heat is added, the alloy absorbs the latent heat of transformation and transforms from martensite to austenite. The metal recovers its original shape during this process and does work against the load σ_H as it moves the load from ε_M to ε_A , state F in Figures 8 and 9. Finally the stress is reduced to σ_L and the actuator cycle has completed a cycle, returning to state 'A'.

The thermal efficiency of an actuator can be written:

$$\eta = (W_{out} - W_{in}) / Q_{in} \quad (1)$$

Net work output is written as:

$$\oint \sigma d\varepsilon = (\varepsilon_M - \varepsilon_A)(\sigma_H - \sigma_L) = \Delta\varepsilon \Delta\sigma \quad (2)$$

The heat input, Q_{in} , is the sum of the sensible heat needed to warm the metal from M_{σ_L} to A_{σ_H} plus the latent heat required to supply the heat of transformation at the higher stress level, Δh_{σ_H} . The sensible heat required to warm the alloy from M_{σ_L} to A_{σ_H} is then:

$$Q_s = \rho C (A_{\sigma_H} - M_{\sigma_L}) = \rho C \left\{ \Delta T + \frac{\sigma_H - \sigma_L}{d\sigma/dT} \right\} \quad (3)$$

where ρ is the density, C is the specific heat, and ΔT is the hysteresis of the alloy. The latent heat of transformation is calculated from the Clausius-Clapeyron equation as:

$$\Delta h_{\sigma_H} = \frac{d\sigma}{dT} \Delta\varepsilon A_{\sigma_H} = \frac{d\sigma}{dT} \Delta\varepsilon (M_s + \Delta T + \frac{\sigma_H}{d\sigma/dT}) \quad (4)$$

and then finally the thermal efficiency of an actuator becomes:

$$\eta = \frac{(\varepsilon_A - \varepsilon_M)(\sigma_H - \sigma_L)}{\rho C \left(\Delta T + \frac{\sigma_H - \sigma_L}{d\sigma/dT} \right) + d\sigma/dT (\varepsilon_A - \varepsilon_M) \left(M_s + \Delta T + \frac{\sigma_H}{d\sigma/dT} \right)} \quad (5)$$

As an example of how these relationships can be used to predict the thermal efficiency of idealized alloys, consider an alloy with the following properties:

- High stress: 414 MPa
- Low stress: 207 MPa
- Transformational strain: 5%
- Stress rate: 6.55 MPa/°C
- Thermal Hysteresis: 50°C
- Specific Heat: 460 J/kg °C

- Density: 6.54 g/cc
- M_s (at zero stress): 0°C

These values are typical of Ni-Ti alloys expected to perform for 1,000 cycles or so. Putting the above values into the thermal efficiency equation gives a thermal efficiency for the alloy of 2.8%. Using this alloy as a baseline, it is possible to vary the above properties systematically to examine their effects on thermal efficiency. The results of such a variation are given in Figures 11,12,13 and 14.

Careful examination of equation 5 and related figures shows the following:

- Increasing the transformational strain ($\epsilon_A - \epsilon_M$) improves efficiency (Figure 11), though clearly fatigue damage increases with increases in $\epsilon_A - \epsilon_M$.
- Increasing the value of the stress applied during heating (σ_H) increases the thermal efficiency of the alloy, as shown in Figure 12. Although the efficiency equation can be solved parametrically to determine if an optimum value exists, in practice, these values are usually set by the martensite yield stress and by the alloy's maximum allowable fatigue stress.
- For maximum efficiency the thermal hysteresis should be minimized (Figure 13).
- The effect of $d\sigma/dT$, while decreasing the sensible heat requirements, increases the latent heat needed to transform the alloy. There is a weak optimum value of $d\sigma/dT$, as shown in Figure 14.

The alloy described in the above analysis is ideal. There are several complications that a designer would face in analyzing a "real" alloy: that transformations are in fact not isothermal increases hysteresis and complicates the calculation of Δh ; plastic

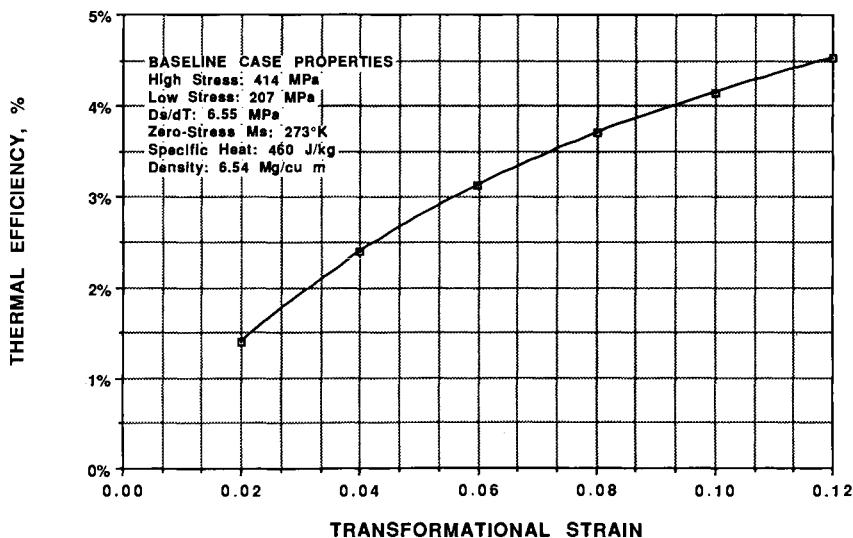


Figure 11: The thermal efficiency of an actuator increases dramatically with transformational strain. Unfortunately, fatigue considerations limit the practical strain values to under 4%.

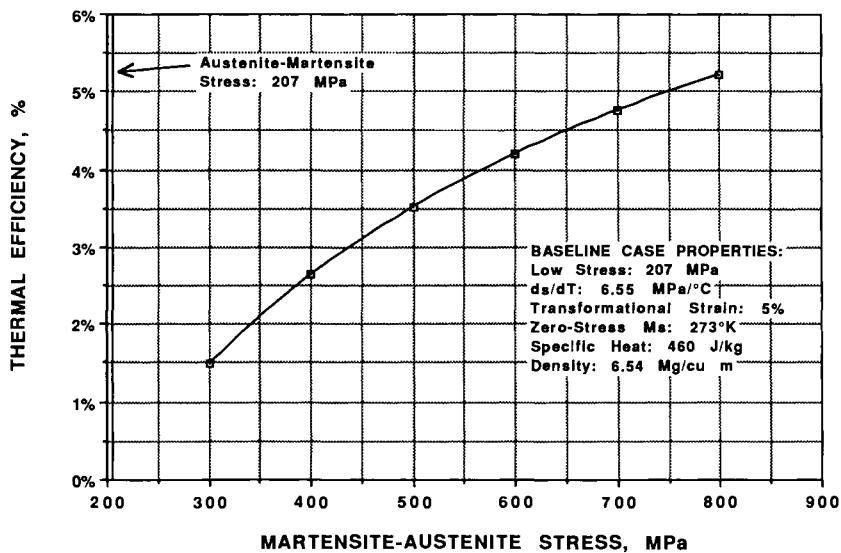


Figure 12: The importance of using alloys with high austenitic yield strengths is shown. Unfortunately there are limits that have to be placed on σ_H due to the effects shown in Figure 5, and fatigue limitations. 400 MPa appears now to be a very optimistic value when coupled with strains of 4%.

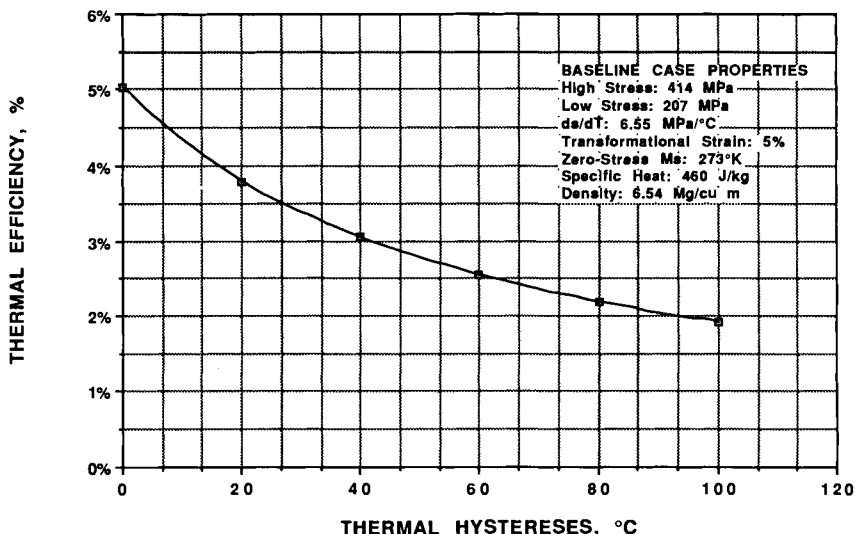


Figure 13: Thermal hysteresis decreases efficiency, providing hope for more efficient heat engines and actuators using Ni-Ti-Cu alloys, or by using only the R-phase,

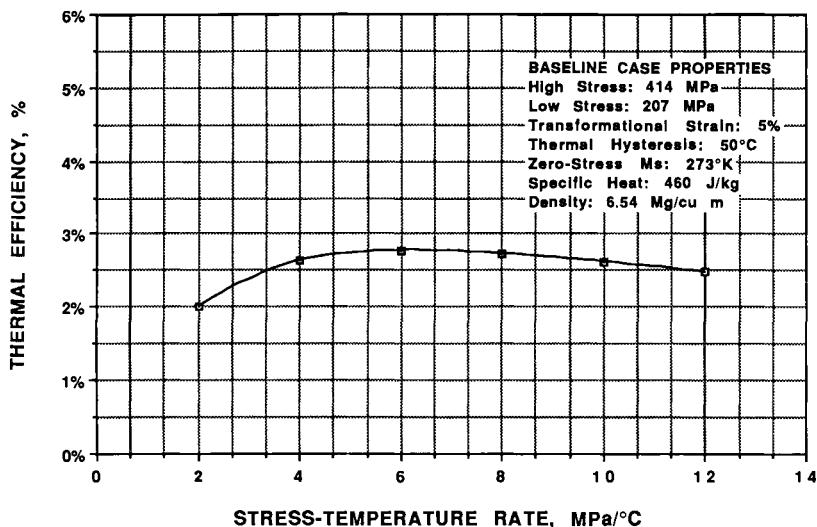


Figure 14: The effects of stress rate, $d\sigma/dT$, are unclear, depending upon what other properties are.

yielding during deformation causes irreversible increases in the transformational entropy; fatigue causes the transformation temperatures to shift; thermal hysteresis is often stress dependent, and is certainly dependent upon whether the alloy undergoes complete transformation or is only partially cycled; transformational strain is stress dependent, making the simplified equation 2 invalid. Many of these complications can be dealt with, but the above simplified analysis is qualitatively useful.

One should also keep in mind that the simplifications of the above example tend to be optimistic from an efficiency point of view (i.e. the calculated values are higher than those one would measure). In all, the efficiency of cyclic SMA devices is rather disappointing. Although this is often not a design factor, one must always be aware that energy conversion itself is probably not a legitimate reason to use an SMA device, especially considering that the relatively high price of SMA's results in a relatively high cost-per-Watt output as well as poor efficiency.

4. Types of Actuators:

In general we consider two types of application: *electrical actuators* and *thermal actuators*. Both types are expected to do work upon heating, the difference being that electrical actuators are heated by passing current directly through the SMA, while thermal actuators are heated by changes in ambient temperature.

4.1 Electrical Actuators:

Electrical actuators are generally used strictly to do work, replacing solenoids, servomotors, hydraulics, pneumatic devices, etc. Ni-Ti alloys are generally preferred in these cases because of their high electrical resistivities, work outputs and fatigue

lifetimes. Compared to other actuation methods, SMA's are typically simpler in design, quieter, more compact and often less expensive.

There are three primary things that one should consider when using SMA's as electrical actuators:

1. One can actuate (heat) quite quickly, but cooling can be very slow (a particular problem for fast acting actuators and some robotic applications).
2. Since one is normally trying to optimize the work output per cycle, fatigue is often critical.
- 3 If the actuator is to operate at high ambient temperatures, there may be a danger of self-actuation, or failure to reset.

4.2 Thermal Actuators:

Thermal actuators have two functions: to detect a temperature change and to actuate. They are generally in competition with bimetals, wax motors, etc. Both Ni-Ti and Cu-based alloys are used as SMA thermal actuators depending upon the exact requirements: Ni-Ti alloys being better in fatigue and work output, the Cu-based alloys having higher transformation temperatures and being less expensive. Compared to other actuation methods, SMA's generally are simpler, less expensive, more compact, and provide very large, sudden motions. Some factors that one should consider in determining if SMA is appropriate to a specific application are:

1. SMA's have a hysteresis, and are often not suitable for temperature control, where a unique identity of position and temperature is required.
2. Again one is limited by the low transformation temperatures of Ni-Ti, and the thermal stability of the Cu-based alloys.
3. Fatigue is usually less of an issue with thermal actuators than with electrical actuators, but it cannot be neglected.

4.3 Heat engines:

Still a third type of actuator are the heat engines, used to convert thermal energy to mechanical energy. There are many ingenious designs of heat engines in the literature, but unfortunately, it does not appear as if any of their efficiencies are sufficient to warrant commercial exploitation. The two main issues are that fatigue effects require one to run at very low stresses and strains, and that the cost of the material is so high in the first place. There are many estimates¹⁻⁵ of energy outputs and costs per kg of material, some using very over-optimistic assumptions. One can get some idea of issues involved by reviewing Figures 11 through 14. The baseline assumptions upon which these are based are quite optimistic in that an engine operating under these conditions would last for less than 1000 cycles, and still highly idealized efficiencies are under 3%. Engines based on the R-phase would have an entirely different set of baseline values, but would nevertheless fail to provide commercially interesting efficiencies.

5. Summary:

Without question, actuators are the most promising and yet the most complex of shape memory applications. In the papers that will follow, various design schemes will be presented that will highlight the complexity of shape memory actuator design, and

many examples will be given that will show variety and the benefits of using shape memory.

References:

1. P. Wollants, J.R. Roos and L. Delaey: Scripta Met. 14, 1217 (1980).
2. K. Mukherjee: Scripta Met. 14, 405 (1979).
3. R. Salzbrenner: J. Mat. Sci. 19(6), 129 (1984).
4. J.L. McNichols and J.S. Cory: J. Appl. Phys. 61(3), 972 (1987).
5. A.A. Golestaneh: J. Applied Physics 49(3), 1241 (1978).

The Two-Way Shape Memory Effect

Jeff Perkins^{*} and Darel Hodgson[†]

^{*} Department of Mechanical Engineering,
Naval Postgraduate School,
Monterey, California 93943, USA

[†] Shape Memory Applications, Inc.,
10201 Torre Avenue, Suite 310,
Cupertino, California 95014, USA

The phenomenon of two-way shape memory (for which the acronym TWSM will be used throughout this paper) may be simply described as behavior in which there is a spontaneous shape change of a sample or component on both heating and cooling. In effect, the material remembers both a high temperature shape and a low temperature shape (as well as the shapes at all temperatures between these extremes), and may be cycled fairly reproducibly between different shapes by simply changing the temperature. This is something extra when compared to the usual one-way shape memory effect. In one-way shape memory the material is initially deformed

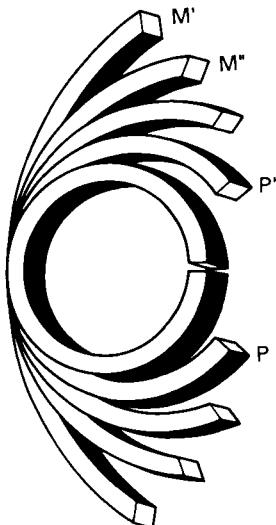


Figure 1: Schematic representation of one-way and two-way shape memory behavior, using the conference logo as a basis. For one-way shape memory behavior, the sample is cooled, then deformed to shape M', then heated to recover shape P. In two-way shape memory, the sample, after appropriate training, will spontaneously cycle between two intermediate shapes, such as M'' on cooling and P'' on heating.

to take on a new shape, and recovers its original shape on heating; the higher temperature shape is then retained on subsequent cooling and must be re-deformed to start another shape memory cycle. This is depicted schematically in Figure 1a. In TWSM, on the other hand, once the material has been conditioned to display the two-way effect, no external force is required to achieve any part of the shape change, and a component can be made to bend one way on heating and the other way when cooled, as shown in Figure 1b. However, as we will see in later discussions, the strains that may be achieved in TWSM are always significantly less than those involved for one-way shape memory in the same alloy, and special processing procedures are necessary to set the material up for TWSM.

Both the one-way and two-way shape memory effects may be described graphically by reference to Figure 2, which is a presentation format originally invented at Raychem Corporation, and which has been borrowed by the present author several times in the past for reviews of the subject¹⁻³. This figure is a schematic stress-strain-temperature diagram, which can be generated experimentally by conducting stress-strain tests on the shape memory alloy of interest at various temperatures. In this diagram, TWSM behavior is represented by the hysteresis loop (1)(2)(3)(4) in the strain-temperature

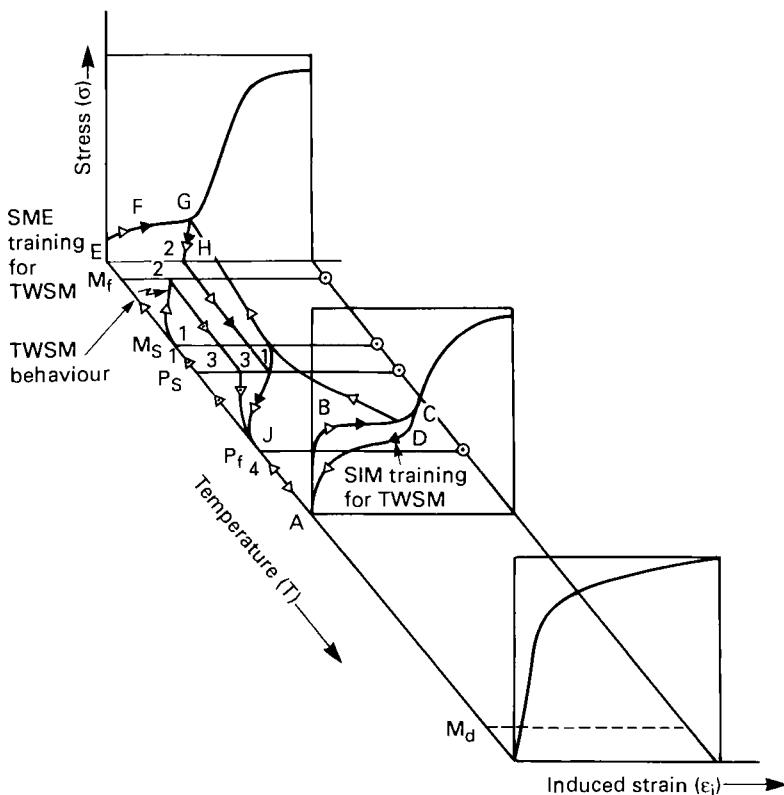


Figure 2: Schematic stress-strain-temperature diagram, used to illustrate various training routines which produce two-way shape memory behavior (see text for descriptions of the details of the training routines).

plane of the diagram. The fact that the loop is in this plane means that the material will spontaneously change shape whenever the temperature is changed within a particular range (in this case, anywhere between (4) and (2), without the application of external stress. This loop, however, which represents TWSM, is not a natural feature of the alloy, but rather is a "learned" behavior of a shape memory alloy. In other words, the alloy must be subjected to special thermomechanical treatment in order to be *trained* to display TWSM (although in some cases the training occurs naturally in the course of operation of a device which is experiencing the equivalent of one of these training routines). These training procedures will be described in a later section. For the moment, let us assume that we have a material which will display the strain-temperature behavior of TWSM. What can we do with it?

1. Applications of Two-Way Shape Memory

At present there are three general markets in which TWSM is used: engineering (industrial), toys, and medical. Each will be reviewed below.

1.1 Engineering Applications

Speaking generally, there are several broad classes of applications in which TWSM may be used, including:

- reversible fastening and coupling devices
- temperature-sensitive actuating and control devices
- deployment and retrieval assistance for components
- cyclic work generation devices (shape memory "engines")
- toys and novelty items

Within these categories, the specific applications are limited only by the bounds of the designers imagination, and already a wide variety of devices have been developed, some to the point of commercial application. The engineering applications of TWSM are in many cases similar to those for one-way shape memory, but of course are designed to take advantage of the added feature of spontaneous shape change on cooling. Briefly summarized, the main feature of TWSM that may be exploited in an application is the fact the component will deflect in one direction upon heating, and in the opposite direction on cooling. Some simple devices are illustrated schematically in Figure 3. Also, the material is able to exert a certain degree of "push" (or pull) as it deflects in one or both directions and the deflections occur at specific temperatures. The amount of memory strain one can obtain on cycling is much greater when using a small biasing stress to assist the deformation upon cooling.

With regard to coupling or fastener applications, the phenomenon of TWSM offers the opportunity of reusable fittings. For example, if a ring or cylinder is properly processed to display TWSM and is then expanded in diameter and recovered onto a substrate by heating, it will loosen and possibly even fall away from the substrate if recooled to a sufficiently low temperature. This feature could be made to apply to the well-known coupling geometry made by Raychem Corporation, for instance. It should be noted that in order to achieve this feature, the component must be properly trained for TWSM, and must be cooled below M_f in order to realize the maximum deflection upon cooling.

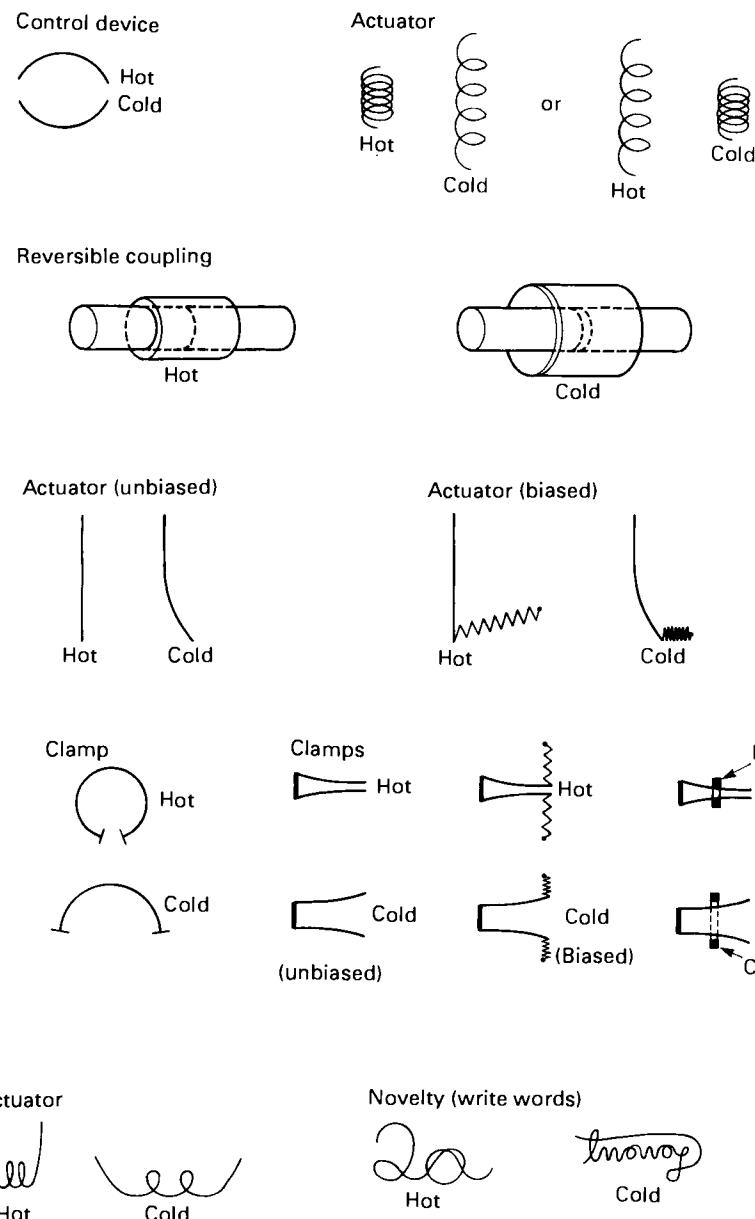


Figure 3: Applications of two-way shape memory illustrated with sketches of a variety of simple devices and configurations.

A number of researchers have encountered TWSM in the course of developing and running various shape memory heat engines. It has been observed that these engines typically need to be cycled or "run-in" for a while before the SMA elements perform at their optimum. This is due to the fact that the initial cycling develops TWSM tendencies in the elements; this dramatically reduces the stress required to deform martensite, since the TWSM behavior amounts to a spontaneous, stress-free, self-deformation of the martensite. Therefore, as time goes on, the engine runs with improved efficiency. This improvement is due to what might be called *insitu* TWSM training, that is, an unavoidable consequence of the repetitive nature of the shape memory application. The particular case of shape memory engines assisted by TWSM tendencies is perhaps not a pure case of TWSM; that is, the low temperature deformation is not occurring at absolutely zero applied stress, because there is an inherent push given by the engine operation. However, the TWSM tendency, when it develops, decreases the resistance to this push, and so allows the engine to operate with greater efficiency.

1.2 Novelty Applications

To date, the largest application of TWSM has been in demonstration and novelty items. The ability of a piece of metal to deflect without assistance on both heating and cooling is one of the most dramatic manifestations of shape memory behavior. For example, several firms and groups have developed TWSM "flowers", with petals which open and close upon heating and cooling, thus blooming and unblooming. A commercial mechanical butterfly has been marketed by Mondotronics; this device is actuated by a small loop of TWSM-trained Ni-Ti wire. When the wire is electrically heated, it contracts and causes the butterfly wings to flap in one direction; when the electrical heating is turned off and the SMA element cools down, the flexure of the plastic body of the butterfly provides a small stress by which to deform the martensite, but in time, most of the shape change upon cooling will be accomplished by spontaneous TWSM action. The wire for this particular application is produced by Toki Corporation, and has probably the greatest intentionally developed TWSM of any material currently on the market. Other commercial TWSM-bearing forms of shape memory alloys will perhaps become available as demand warrants their development.

1.3 Biomedical Applications

Special mention will be given to this category of application for TWSM. A variety of biomedical applications have already been proposed and developed, and a surprising number are in use⁴. A few of these will be briefly described here. One example is a cotter-pin-like fastener, that adopts a closed configuration at a higher temperature, an open configuration at low temperature. Such a component may be deployed at the desired position in the body with a cold probe, then allowed to close and anchor something (say a prosthesis) as it is heated in the body. Subsequent removal could be accomplished with the cold probe. The design might also include surface asperites to act as anchors, trained in such a way that they lie flat during installation, and only pop up when heated.

The use of TWSM-trained wires as heat-pulsed artificial heart muscles has been investigated. In this case the wires were embedded in an elastomeric heart wall and actuated by a programmed cyclic electrical current which caused them to bend back and forth as heated and cooled, thus flexing the heart wall.

Intrauterine contraceptive devices (IUDs), aneurysm clips and vena cava filters (to trap blood clots) are other examples of implanted devices in which the initial deployment and subsequent removal may be assisted by the ability to display TWSM. The device is caused, when cooled, to adopt a relatively simple shape for deployment or removal, but exists in the more complex functional shape at the higher temperature of the body.

In addition to the strictly TWSM-related biomedical applications mentioned here, there are a variety of other applications of shape memory behavior in the biomedical field. These will be discussed elsewhere in these proceedings.

2. Microstructural Basis For Two-Way Shape Memory

Before going into the description of training routines for TWSM, a bit of microstructural background may be useful. At the time of the 1975 International Symposium on Shape Memory Effects and Applications⁵ (held in Toronto, Canada), two-way shape memory had only just been reported for the first time, by Delaey and coworkers⁶, who presented a vivid color cine-film⁷ which revealed the microstructural aspects of the behavior. Since that time, the role of TWSM behavior has come to be understood quite well on the microstructural level, and this has allowed the behavior to be exploited in a variety of applications. Although it is not the intent of this paper to get the reader bogged down in detailed descriptions of microstructural mechanisms, a few basic ideas along those lines may serve us well.

Before discussing the origin of two-way memory it is appropriate to first quickly review some of the ideas presented in the first chapter of this book in a more visual way through the schematic sketches of Figure 4. There are three key microstructural forms for the shape memory alloy. The first of these is austenite, which exists as a normal polycrystalline grain structure (Figure 4a), and as mentioned above, represents the high temperature shape. The second key microstructure is martensite which is formed simply by cooling to below M_f (Figure 4b); when this forms from the austenite, we see that it in effect subdivides the austenite into a mosaic of wedge-shaped crystals. If there is no applied stress or prior training, the sample retains the same shape as the high temperature austenite when this martensite forms. The third key microstructure is stress-biased martensite (Figure 4c); as mentioned, this is created either by stressing martensite below M_f , or by stressing austenite near but above A_f . As we can see in Figure 4c, this causes an alteration in the mosaic pattern of the wedge-shaped martensite crystals; it is this preferred microstructural pattern which becomes "learned" by the alloy during TWSM "training", and so eventually corresponds to the remembered low temperature shape; this learned low temperature microstructure may appropriately be referred to as "stress-biased martensite".

The usual one-way shape memory effect (shape recovery on heating) is based on the fact that the alloy inherently remembers the original high temperature shape (corresponding to the microstructure of Figure 4a). Therefore, if the material is deformed while in the low temperature range, that is, below M_f (thus creating stress-biased martensite depicted in Figure 4c), heating will lead to a transformation back to the original parent phase and shape. After one sequence of this nature (cooling below M_f , deforming and heating above A_f), the parent phase has little if any memory of the stress-biased martensite and thus little if any memory of the low temperature shape; we must train it to have this additional bit of memory. In the training routines that will be described below, the general microscopic principle which is involved is that the low temperature shape (stress-biased martensite) will only be remembered if it is

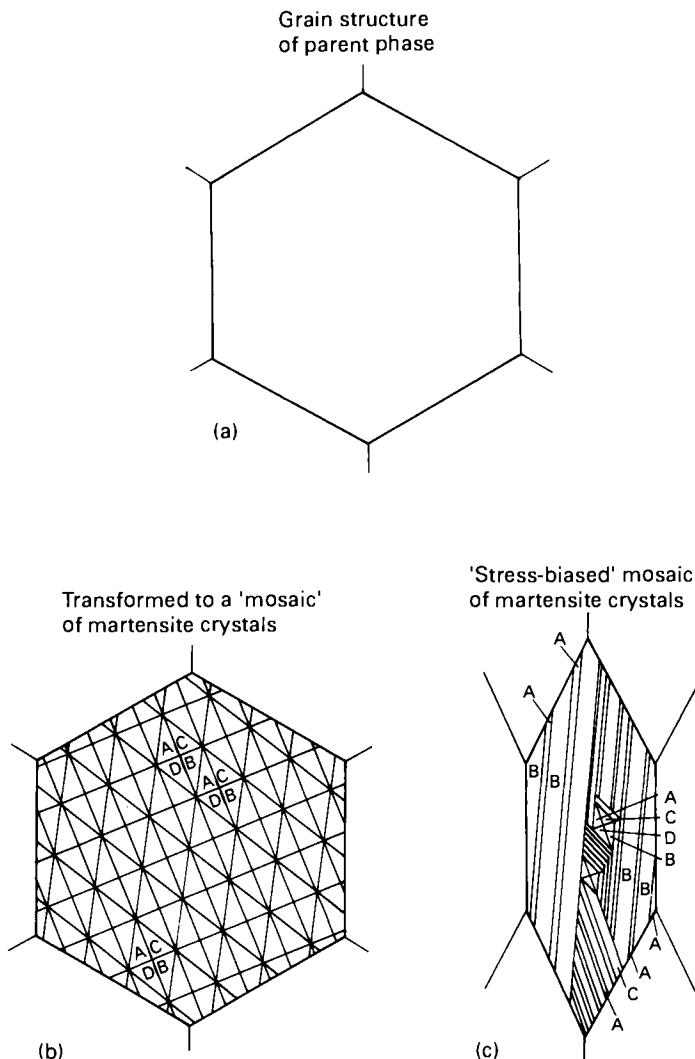


Figure 4: The microstructural basis of two-way shape memory behavior: (a) the simple polycrystalline grain structure of the high temperature parent phase, (b) each parent phase grain becomes "sub-divided" as it transforms to a mosaic of wedge-shaped martensite crystals upon cooling, (c) upon application of stress to the array of martensite crystals, the boundaries between them move and change the distribution, creating a "stress-biased martensite". It is this last sort of microstructure which the sample learns to remember during two-way shape memory training.

repeatedly introduced into the sample (in keeping with the mental analogies sometimes used in the shape memory field, this is sort of like "learning by rote repetition"), or if it is introduced to an excessive degree (which is like "learning by overdosing the memory bank").

3. How to Get Two-Way Shape Memory: Training

As mentioned earlier, TWSM is a learned behavior for a shape memory alloy. Under, normal circumstances, a shape memory alloy remembers its high temperature shape, but upon heating to recover the high temperature shape, immediately forgets the low temperature (deformed) shape. However, it can be "trained" to remember the low temperature shape as well. This is accomplished essentially by managing to leave some "reminders" of the deformed low temperature condition (stress-biased martensite) in the high temperature phase. There are several ways of doing this, which are described below.

3.1 TWSM Training by Overdeformation while in the Martensitic Condition

This sequence of events is presented in Figure 5a. The alloy is cooled below M_f , and while in the martensitic state, is severely bent, to well beyond the usual strain limit for completely recoverable shape memory. When reheated to the parent phase range, the alloy will not completely recover the original shape, due to the excessive deformation of the martensite (by exceeding the Shape Memory (SM) strain limit, a partial loss of memory results). However, if cooled again to the martensite range, the alloy will spontaneously move part of the way back toward the overdeformed shape. This training routine is represented in Figure 2 by cooling to below M_f and then deforming along the path E-F-G to a point well beyond G.

3.2 Training by Shape Memory Cycling (Cool-Deform-Heat etc):

This procedure is shown in Figure 5b. It consists simply of repeatedly carrying out shape memory cycles until the two-way behavior begins to be demonstrated. One SM training cycle would consist of the component being cooled to below M_f , deformed to a level below the SM strain limit, then heated to recover the original high temperature shape. After a number (perhaps 5 to 10) of these SM cycles have been carried out (note that the direction of the strain must be the same during each training cycle), the component will begin to spontaneously change shape on cooling, moving in the direction in which it has been consistently deformed during the training cycles. The amount of spontaneous shape change on cooling will be significantly less than that which was being induced in the SM deformation step (typically the spontaneous shape change will be perhaps 1/5 to 1/4 of the training strain; for example, if the strain induced during training was 6%, the spontaneous TWSM strain is likely to be no more than 1 or 2%). This particular training routine can also be seen by reference to Figure 2, where it is represented by repeatedly following the path A-E-F-G-H-I-J-A.

3.3 Training by Pseudoelastic (PE) Cycling (Load-Unload-etc)

This method is illustrated in Figure 5c. It consists of repeatedly stress-inducing martensite by loading and unloading the parent phase above the A_f temperature but below M_d where pseudoelastic (or superelastic) behavior is expected. This training

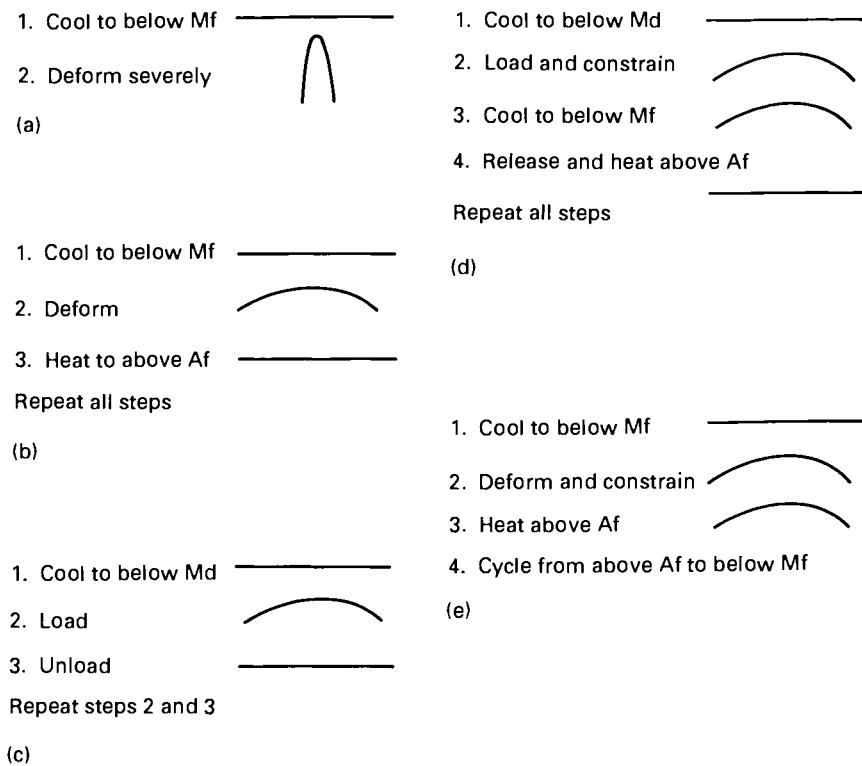


Figure 5: Illustration of several different training routines that can be used to produce two-way shape memory behavior: (a) over-deformation while in the low-temperature martensitic condition, (b) repetitive shape memory cycling, involving, on each cycle: cooling to form martensite, deformation of that martensite, reheating to recover the high temperature shape, (c) repetitive pseudoelastic cycling, involving on each cycle: stress-induction of martensite from the parent phase, reversion on unloading, (d) so-called "combined" shape memory-plus-pseudoelastic cycling, whereby on each cycle: martensite is first stress-induced from the parent phase, then the sample is constrained and cooled below M_f , then allowed to revert freely upon reheating, (e) constrained temperature cycling of deformed martensite: cool to form martensite, then deform the martensite below M_f , then constrain and repeatedly cycle from below M_f to above A_f .

routine is represented in Figure 2 by repeating the path A-B-C-D-A. As for training method 2, the number of training cycles required is typically on the order of 5 to 10, and the subsequent spontaneous shape change on cooling and heating is a fraction of the training strain.

3.4 TWSM Training by Combined SME/PE Cycling

A particularly effective training routine for TWSM has been found that combines some of the features of methods 2 and 3. This is illustrated in Figure 5d, and is represented in Figure 2 by repeatedly following the path A-B-C-G-H-I-J-A. The component is first deformed in the parent phase condition (to stress-induce a certain amount of stress-biased martensite), then cooled to below M_f while holding the induced strain in the sample (including the elastic strain), then heating up to recover the original shape. If this routine is repeated a number of times, TWSM behavior will be obtained on subsequent cooling and heating. Although effective, this method is a little more elaborate than the others that have been described.

3.5 TWSM Training by Constrained Temperature Cycling of Deformed Martensite

This is a variation on the previous method which is perhaps a bit easier to carry out in terms of temperature control, and is probably the most commonly used training method at present. As seen in Figure 5e, the sample is deformed below M_f , thus producing a stress-biased martensitic microstructure. The sample is then constrained in the deformed condition and heated to above A_f . The sample is typically cycled from below M_f to above A_f a number of times, with the sample all the while constrained in the original deformed shape, to complete the training routine. This training method proves to be particularly effective and is relatively straightforward to carry out.

4. Limitations on the Use of Two-Way Shape Memory

There are at least four inherent limitations on the application of TWSM:

- Strain limit: There is a limit to the amount of reversible strain which can be realized; typically, this is in the neighborhood of 2%.
- Hysteresis: The inherent temperature hysteresis between the heating and cooling transformations is present.
- Low transformation forces on cooling: In practice this means that you can push with the alloy much better on heating than on cooling; but you can help the alloy on cooling with a bias spring or other such device.
- Upper temperature limit: If you go too high in temperature, you can anneal out the training.

In addition to these points, there are a number of other things to keep in mind. For example, the initial state of annealing should be consistent in order to get consistent TWSM behavior. Beware of texture in worked products; for example, in sheet, the behavior along the rolling direction will tend to be different than that across the rolling direction. The exact details of the training method are very important: how much strain and how many cycles are the major variables; the optimum must be determined by trial and error and there must be consistency in applying whatever training routine is selected.

5. "All-Around" Shape Memory: A Unique Form Of Two-Way Shape Memory

For the sake of completeness, and because of its similarity to the TWSM discussed above, brief mention will be made here of a unique shape memory behavior which has been dubbed "All-Around" Shape Memory Effect, or *ARSME*. This behavior, discovered and named by Honma and coworkers⁸ in Japan, is phenomenologically the same as the TWSM we have been talking about in this paper; that is, the sample deflects in one direction when heated, in another direction when cooled. However, the metallurgical basis for this effect is not the same as has been described here for TWSM, and in fact, is a matter of some controversy at present. Most importantly, the training method for the "all-around" shape memory is quite different than that for TWSM.

The "all-around" shape memory is only able to be achieved in certain Ni-Ti alloys (those with greater than about 50.5 at.% Ni), and only after an aging treatment is conducted on the alloy in which a sample is constrained in what is intended to be the high temperature shape. A typical aging treatment would be on the order of 50 hours at 400°C. During this aging treatment, it is known that a precipitation reaction occurs⁸. It is believed that the precipitate particles effectively create a back-stress which causes a deflection away from the constrained shape when the sample is released and cooled.

The whole sequence of events during training and the manifestation of the effect is summarized schematically in Figure 6. After the constrained-aging training treatment, the sample may be cycled between the high and low temperature shapes simply by changing the temperature, just as in the case of the TWSM which has been the main subject of this paper.

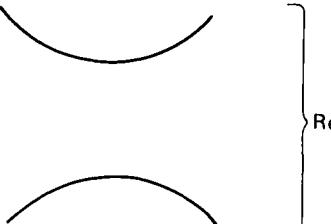
1. Ti-Ni alloy with more than 50.5 at % Ni

2. Deform and constrain



3. Age at about 400°C for about 50 hours

4. Cool



5. Heat

Figure 6: Illustration of "all-around" shape memory behavior. To introduce this effect by training, an appropriate Ni-Ti alloy must be constrained in the high temperature shape and aged. Upon cooling, the sample will spontaneously deflect away from the constrained shape and the sample will subsequently cycle between the two shapes during temperature cycles.

6. Summary:

- TWSM amounts to a spontaneous cyclic strain achieved by simply changing temperature.
- Training is required to achieve TWSM.
- The strains available in TWSM are relatively small.
- Various strain paths may be selected, and these can involve complex stress states if need be.
- Approximately the usual temperature hysteresis of transformation will be displayed during the TWSM behavior.
- Forces able to be exerted against a load will be of the usual level on heating, but relatively weak on cooling.
- A bias spring or other assistance may be employed to assist the action on cooling, giving more strain and/or more force.
- There is a tendency for the TWSM behavior to decay with continued cycling, especially if working against a bias spring.
- There is an upper temperature limit on applications, above which the training for TWSM may be annealed out.

References:

1. Perkins, J.: Shape Memory Alloy '86 (Guilin, China, September 1986) (Chu Youyi, T. S. Hsu and T. Ko, eds.), China Academic Publishers, Beijing, pp. 280-285, (1986).
2. Perkins, J.: Materials Research Society Symposium Proceedings 21, 669-674, (1984).
3. Perkins, J. and Sponholz, R. O.: Met. Trans. 15A, 313-321, (1984).
4. Mukherjee, K.: Encyclopedia of Materials Science (M.B. Bever, editor) Pergamon Press, Oxford, pp. 4368-4371, (1986).
5. Shape Memory Effects in Alloys (Jeff Perkins, editor), (Proceedings of the International Symposium on Shape Memory Effects and Applications, Toronto, May 1975) Plenum, New York, (1975).
6. Delaey, L. and Thienel, J.: Shape Memory Effects in Alloys (Jeff Perkins, editor), Plenum, New York, pp. 341-350, (1975).
7. Delaey, L., Hummel, G. and Thienel, J.: Encyclopedia Cinematographica (G. Wolf, editor), Institute furden Wissenschaftlichen Film, Gottingen, Film E 2251, (1977).
8. Nishida, M. and Honma, T.: Scripta Metallurgica 18, 1293-1298, (1984).

Thermal Actuators: A Comparison of Shape Memory Alloys with Thermostatic Bimetals and Wax Actuators

Peter Tautzenberger

G. Rau GmbH & Co
Pforzheim, West Germany

Thermal actuators, by definition, are devices which convert thermal energy into mechanical energy, utilizing effects such as thermal expansion, solid/liquid or liquid/gaseous phase transformations or solid state transformations like the shape memory effect. These devices sense changes in ambient temperature and react to these changes by changing their shape. Although focus of this book is on thermal shape memory actuators, this chapter also describes thermostatic bimetals and wax actuators, comparing their properties and benefits.

1. Shape Memory Effect:

As described earlier, the shape memory effect is based on a thermoelastic martensitic transformation¹. This transformation can be utilized for thermal actuators in different modes, as shown in the following paragraphs. The most important shape memory

Table 1: Typical Properties of Shape Memory Alloys

Property	NiTi ^a	Cu-Zn-Al ^b	Cu-Al-Ni ^c
Density (g/cm ³)	6.4-6.5	7.8-8.0	7.1-7.2
Electric Conductivity (10 ⁶ S/m)	1-1.5	8-13	7-9
Tensile Strength (MPa)	800-1000	400-700	700-800
Elongation (%)	40-50	10-15	5-6
Maximum A _s -Temperature (°C)	120	150	200
Maximum One-Way-Effect ε _{1max} (%)	8	6	5
Maximum Two-Way-Effect ε _{2max} (%)	5	2	2
Maximum Exposure (Short time) (°C)	400	160	300

a) NiTi with nearly equiatomic composition

b) Cu-Zn-Al with 15-25 wt. % Zn and 6-8 wt. % Al (fine grained)

c) Cu-Al-Ni with 13-14 wt. % Al and 3-4 wt. % Ni (fine grained)

alloys used for thermal actuators are Ni-Ti and Cu-Zn-Al with transformation temperatures between -200°C and +100°C. Cu-Al-Ni alloys show higher transformation temperatures but are brittle, difficult to deform and are not yet really commercially available. Transformation temperatures and other relevant properties of some actuator alloys are listed in Table 1. Other alloys potentially useable for thermal actuators are Fe-Mn-Si and Ti-Pd-Ni for higher operating temperatures.

1.1 The One Way-Effect:

After shape-setting the high temperature shape of the actuator and cooling back to room temperature, the one-way-effect comprises the following steps:

- deform at room temperature (or temperatures < M_f)
- heat to $T > A_f$ to recover high temperature shape
- cool again to room temperature

Cooling from high temperatures does not cause a macroscopic shape change. A deformation is necessary to create the low temperature shape. On heating transformation starts at A_s and is completed at A_f , typically some 2 to 20°C hotter depending on the alloy, loading conditions etc. A_s is determined by the alloy type and composition. It can be varied between -150°C and approximately 200°C. The one-way-effect is schematically shown in Figure 1, using a wire in bending².

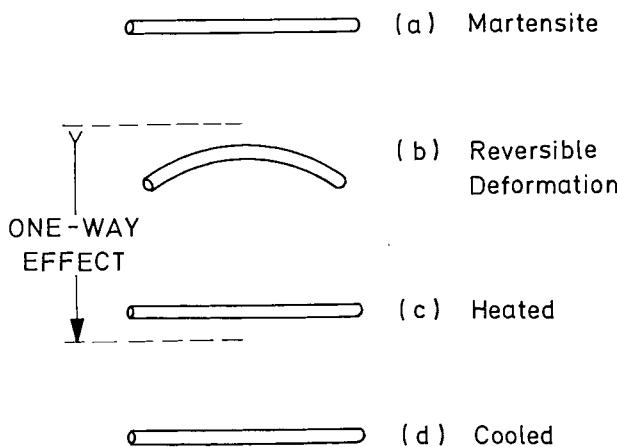


Figure 1: One-way-effect (schematic).

1.2 The Extrinsic Two-way Effect:

The one-way-effect can be repeated many times. For each cycle a deforming force is necessary. If this force is applied constantly, e.g. as a constant load attached to a spring or a bias spring working against the shape memory spring, a two-way behavior can be achieved. The applied force must be high enough to stretch the shape memory spring in the martensitic condition. On the other hand it must be sufficiently small not to

cause excessive deformation of the austenite. Since this two-way motion is caused by an externally applied stress, it is often referred to as the *extrinsic two-way effect*.

1.3 The Intrinsic Two-Way Effect:

Shape Memory Alloys can, under certain conditions, show a true two-way effect, which makes them remember two different shapes, a low and a high temperature shape, even without the application of an external force. This is the so-called *intrinsic two-way effect*. Figure 2 shows this behavior, again using a wire in bending. The

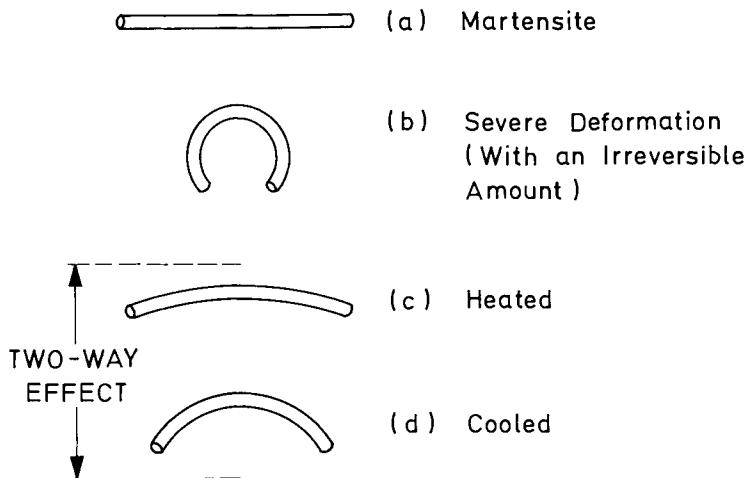


Figure 2: Two-way-effect (schematic).

microstructural reasons for the two-way effect as well as methods to induce this effect in shape memory alloys, were described in detail by Perkins³ in the previous chapter. This paper also covers the so-called all-around Shape Memory Effect, a very pronounced two-way-effect found in specific Ni-Ti alloys and shown schematically in Figure 3⁴. Two preferred shapes of two-way memory actuator elements made out of Cu-Zn-Al shape memory alloys are shown in Figure 4.⁵

1.4 Shape Memory Based on R-Phase:

The R-phase phenomenon, found in certain Ni-Ti alloys, has been described earlier in this book. This effect can also be used in thermal actuator designs. Its major advantages are an extremely narrow hysteresis and very good cyclic stability⁶. On the other hand, the available motion is only about 0.5-1%, and the transformation temperature cannot be easily shifted by alloy formulation.

1.5 Properties of Thermal Shape Memory Actuators:

The following papers in this book will describe a number of applications for thermal shape memory actuators. As the motion of shape memory actuators is not limited to

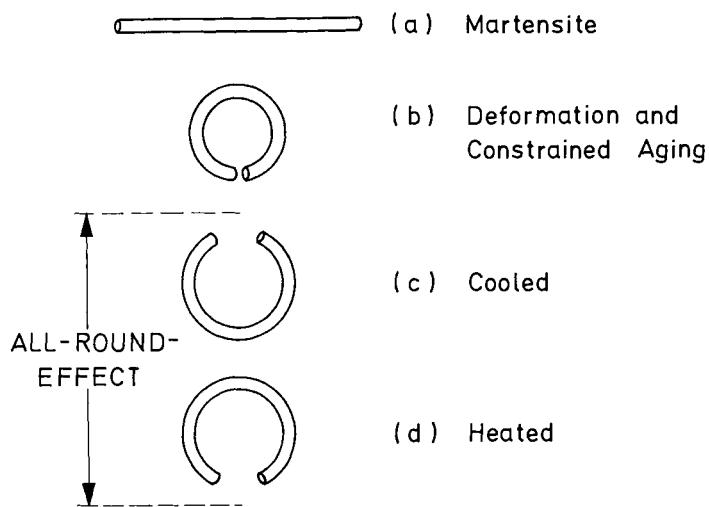


Figure 3: All-round-effect (schematic).

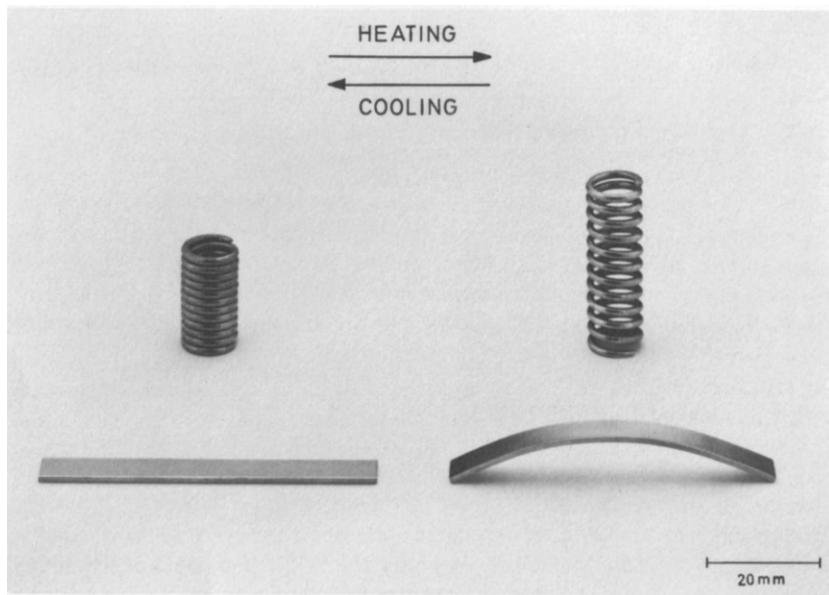


Figure 4: Two-way-effect in cu-zn-al actuators.

bending or any other single mode, one can choose from a variety of shapes and configurations (Figure 5). There are many areas in the automotive, appliance, building, plumbing industries etc., where shape memory thermal actuators can and will

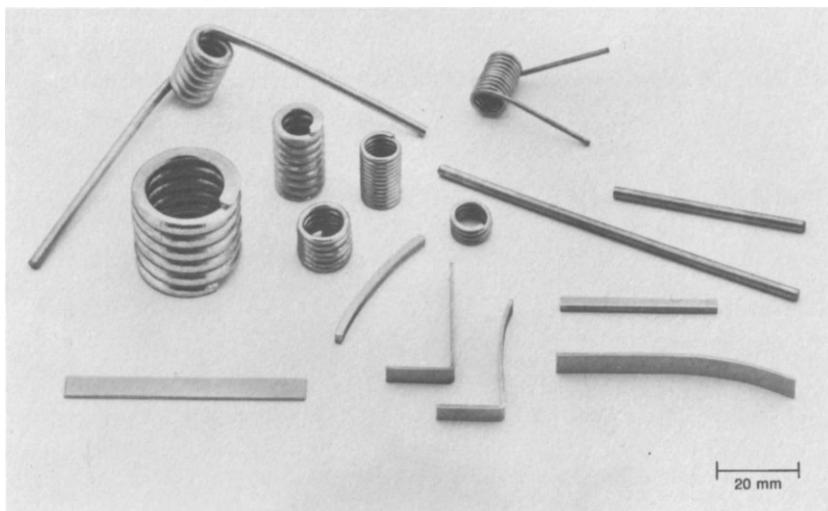


Figure 5: Different shapes of thermal actuators.

provide significant advantages over competing technologies. Some of these benefits are listed below:

- small size
- high force
- large motion
- high work output
- few mechanical parts
- non-linear characteristic

2. Thermostatic Bimetals:

Thermostatic bimetals are laminated composites using materials with a pronounced difference in thermal expansion⁷. Therefore, thermostatic bimetals change their shape whenever ambient temperature changes. The mode of the shape change is always in bending.

The components of thermostatic bimetals are chosen with regard to their thermal expansion and their elastic modulus. The most common materials for thermostatic bimetals are Fe-Ni alloys, Mn-Cu-Ni alloys and some stainless steels. Table 2 lists a few popular combinations. About 100 different thermostatic bimetals are produced worldwide. The amount of deflection or curvature of a thermostatic bimetal can be easily calculated using standard formulas. The most important basic trends are:

- thermal sensitivity of the thermostatic bimetals increases with increasing difference in thermal expansion of the components

- force output increases, motion (deflection) decreases with increasing thickness of the thermostatic bimets

The temperature-deflection characteristic of a thermostatic bimetal is schematically shown in Figure 6. Most thermostatic bimets deflect linearly with temperature

Table 2: Typical Combinations and Properties of Thermostatic Bimets

Properties of Thermostatic Bimets

High α Layer	Low α Layer	Flexivity $10^{-6}/^{\circ}\text{C}$	Range of Linear Deflection $^{\circ}\text{C}$	Operating Limit $^{\circ}\text{C}$
Fe-20%Ni-6%Mn	Fe-36%Ni	15.5	-20 to 200	450
Mn-18%Cu-10%Ni	Fe-36%Ni	20.8	-20 to 200	350
Mn-15%Ni-10%Cu	Fe-32%Ni-6%Co	23.0	-20 to 230	350
18/8 SS	Fe-32%Ni-14%Co	9.1	-20 to 400	650
18/8 SS	17%Cr SS	5.0	-20 to 600	550

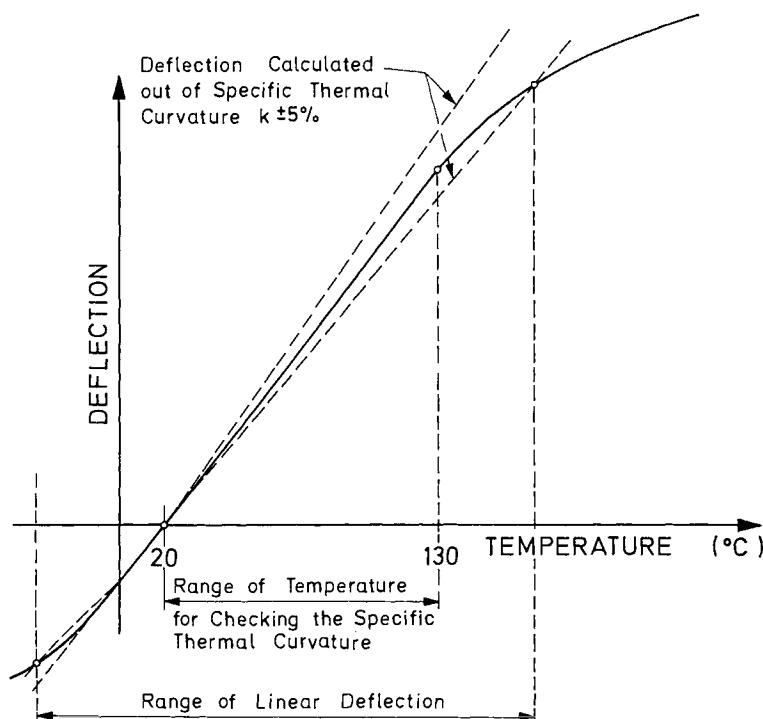


Figure 6: Temperature/deflection characteristic of a thermobimetal.

between -20°C and approximately 200°C . Above this temperature, deflection is non-linear, but still sufficient for many applications. Maximum operating temperature is given by the recrystallization temperature of the component materials and is about 350°C to 650°C , depending on the combination.

By prestressing certain shapes of thermostatic bimetallic elements, like domed (beveled) discs, Belleville washers or certain roof-shaped configurations, a snap-action performance can be achieved. The temperature-displacement curve of such an element is shown in Figure 7. It resembles the hysteresis curve of a shape memory

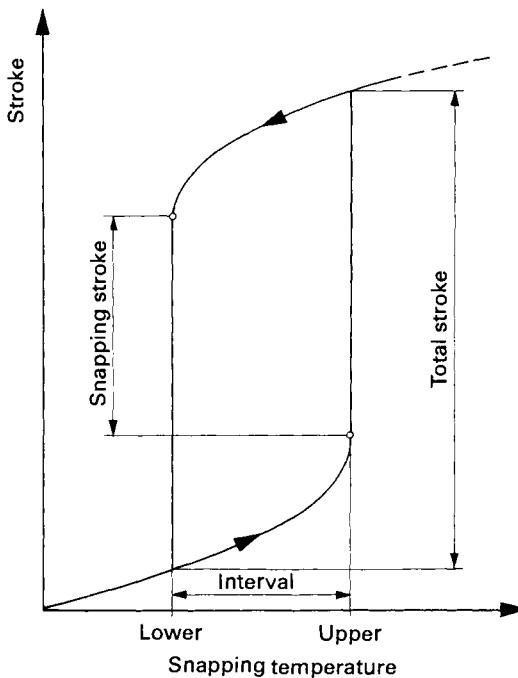


Figure 7: Temperature/deflection characteristic of a thermobimetal snap action disc.

alloy. However, the achievable motion is usually very small (typically below 1 mm) and only certain geometrical configurations can be realized. The "switching" temperature of snap action elements can be between -30°C and 350°C and the hysteresis from 6°C to 250°C , depending on the material and the geometrical configuration⁸.

Thermostatic bimetals have been mass-produced for more than 100 years. They are easy to work with and are not expensive. Millions of parts are produced annually for the use as thermal actuators and overtemperature protectors in the automotive and appliance industry, electrical engineering, chemical processing as well as heating and plumbing industry. Figure 8 shows examples of different shapes⁹.

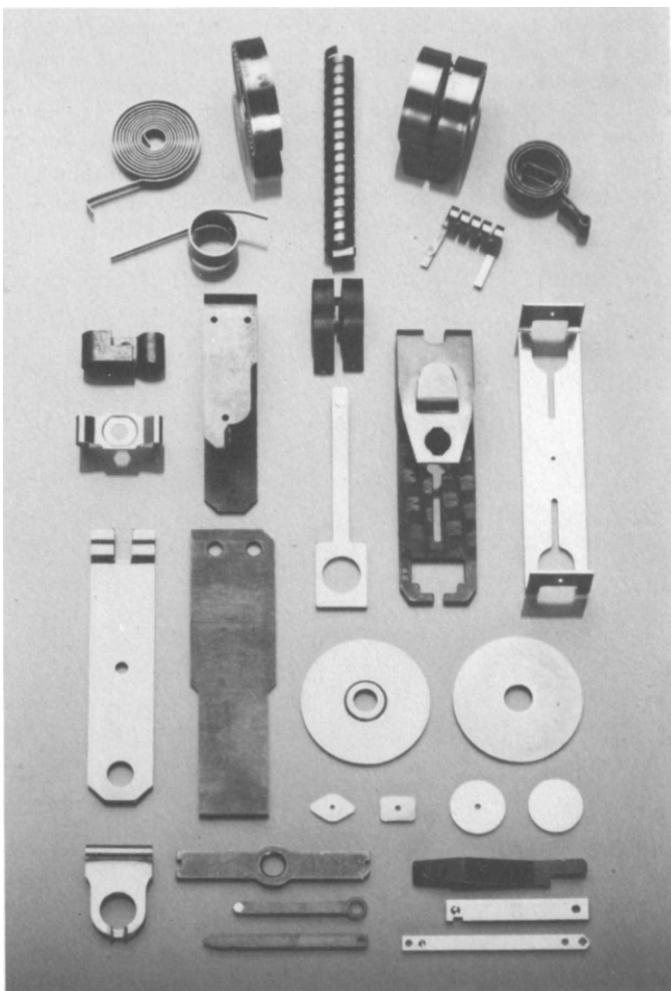


Figure 8: Different shapes of thermobimetal actuators.

3. Wax Actuators:

In wax actuators the volume change of wax during the solid/liquid transformation is used to get a linear motion and or a force¹⁰. The wax is encapsulated in a container with a moving piston transmitting the volume increase on heating as linear motion. On cooling the reverse transformation takes place with a small hysteresis (2 to 5°C), but a biasing force is needed to reset the piston. The biasing force is typically about 20 to 30% of the operating force.

Depending on the formulation of the wax, thermal actuators for proportional and non-proportional control can be made. Again depending on the type of wax used, the proportional control range can be between 15°C and 150°C. However, the total available stroke is the same. Figures 9a and b show typical stroke/temperature characteristics of wax actuators.

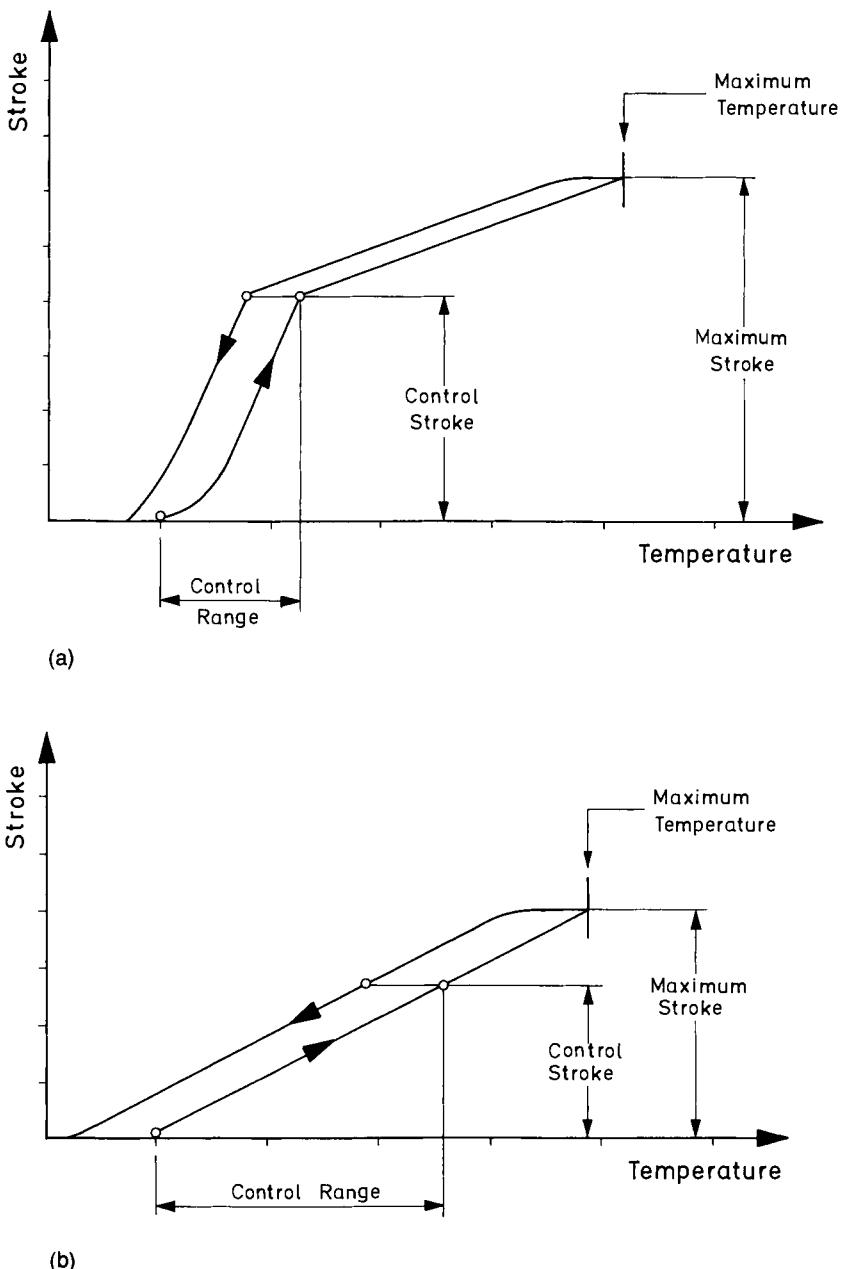


Figure 9: Temperature/stroke characteristic of wax actuators:

- small range (e.g. about 15K)
- wide range (e.g. about 150K)

Wax actuators may also generate their motion at a predetermined temperature in a very narrow temperature range with almost instantaneous reaction when the "switching" temperature is reached. The available stroke and force of the wax actuator depends on the volume of the container. It can be varied in a very wide range. Two typical configurations are listed in Table 3. It is obvious that the work output can be

Table 3: Typical Data of Wax Actuators

Outer Diameter	Length	Maximum Stroke	Maximum Load
12 mm	30 mm	7 mm	120 N
20 mm	60 mm	25 mm	300 N

very high. The operating temperature range for wax actuators is between -40°C and approximately 180°C. Because of the rather large volume, thermal response time can be quite long.

The most important application of wax actuators is thermostatic valves for automotive water and oil cooling systems. Thermostatic valves using wax actuators are also used in warm water central heating systems.

4. Conclusion:

In most cases technical and economic considerations will determine which actuator is best suited for a particular application. Only in a very limited number of applications it will be possible to use one or the other equally economically. Therefore it is very important to understand the specific benefits of either kind of actuator.

Shape memory thermal actuators combine large motion, rather high forces and small size, thus providing high work output. They usually consist of only a single piece of metal, e.g. a helical spring, and do not require sophisticated mechanical systems. As they generate the motion and/or force in a narrow temperature range, which can be predetermined by the alloy composition, work output can be more than 100 times higher than with thermostatic bimetals. In Figure 10 the temperature/deflection curves of a shape memory (Cu-Zn-Al) cantilever beam and a thermostatic bimetallic strip of the same size are plotted. The deflection of the shape memory element is based on a 1% two-way effect and measured at the free end of the beam⁹.

The stability of the shape memory effect is mainly depending on the operating stress and strain. For multi-cycle applications the operating stress should not exceed 250 MPa in the case of Ni-Ti, and about 75 MPa in the case of Cu-Zn-Al. The strain should be limited to 1 to 2%. Under these circumstances 100,000 or more thermal cycles can be achieved without significant deterioration.

Unless the R-phase is used, the motion/temperature curve of a shape memory actuator is characterized by a hysteresis of approximately 15 to 50°C (depending on the alloy). On the other hand thermostatic bimetals actuators are hysteresis-free. Their motion/temperature characteristic is linear over a wide temperature range, which

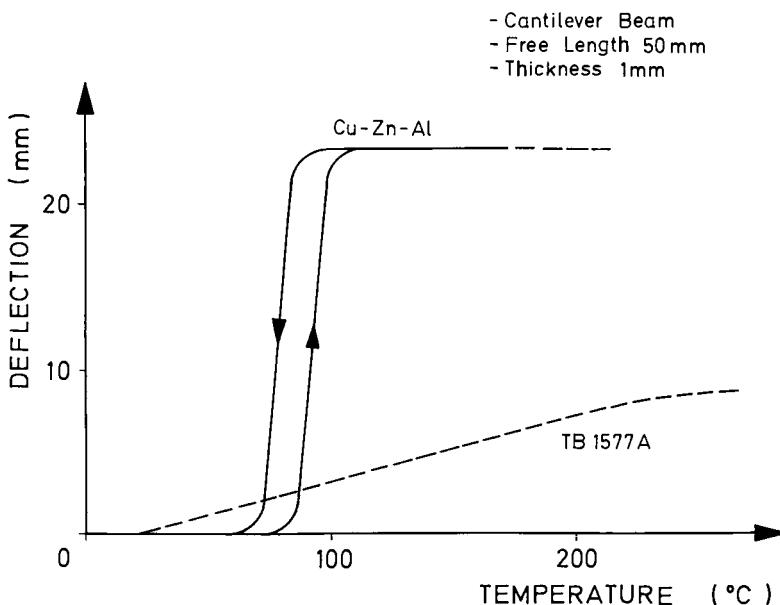


Figure 10: Temperature/deflection curves for a shape memory actuator and a thermobimetal actuator.

makes them particularly suited for proportional control. However, snap action elements can be made by prestressing certain shapes. As the thermostatic bimetallic effect is a geometrical effect, it is rather small and very dependent on the dimensions of the element. There is a trade-off between force and motion (the higher the force, the smaller the deflection). The mode of operation is always bending. Thus, if linear motion is required, mechanical conversion systems or complicated geometrical configurations are needed. Thermostatic bimetallic actuators provide excellent stability up to many million cycles. Unlike shape memory actuators, they also can do work on cooling.

Wax actuators provide very high forces and reasonable motion. However, they can be bulky and slow in response to temperature changes. Like shape memory actuators they cannot do work on cooling and need a biasing force to reset. Wax actuators are devices, not parts, consisting of a container and a moveable piston. Thus, sealing can be an issue.

References:

1. Otsuka, K.; Shimizu, K.: Proc. Int'l Summer Course Mart. Trans. 81.(1982).
2. Tautzenberger, P.; Stoeckel, D.: Z.w.F. 78, 486 (1983).
3. Perkins, J.: J. Mat. Sci. Eng., 51, 181 (1981).
4. Nishida, M.; Honma, T.: Scripta Met., 18, 1293 (1984).
5. Tautzenberger, P.; Stoeckel, D.: Z.w.F. 81, 703 (1986).
6. Miyasaki, S.; Otsuka, K.: Phil. Mag. A50 (1984) 393

218 Thermal Actuators: A Comparison with Thermostatic Bimetals and Wax Actuators

7. G. Rau GmbH & Co: Thermobimetalle, company brochure (1979)
8. Schneider, F.E.: Thermobimetalle (F.E. Schneider editor), expert verlag 93 (1982).
9. Tautzenberger, P.; Stoeckel, D.: Metall_41 , 26 (1987).
10. Janocha, H.: Conf. Actuator_88, Bremen (1988).

Electrical Actuators Alloy Selection, Processing and Evaluation

J.R. Yaeger
Seagate Technology
920 Disc Drive
Scotts Valley, CA. 95066

The uninitiated shape memory actuator designer is confronted with a seeming nightmare of alternative choices and design tradeoffs. This complexity is further compounded when he decides to excite the shape memory elements by direct electrical heating. All of the following are sufficiently interdependent that a modification of one will surely result in a change in performance of the others; for example:

- load and stress
- stroke and strain
- electrical current and power
- ambient temperature
- actuation and reset response time
- fatigue , both failure and degradation

Major progress can often be made, however, when two sets of design issues causing much of the confusion are understood:

- selection, processing and evaluation of alloys
- understanding the effects of ambient temperature

These two issues will be the primary subject of this paper. A secondary goal is to give the would-be designer an initial approach for evaluation and design.

1. Alloy Selection:

Unfortunately, there are no standards for shape-memory alloys, so it is not possible to go to a handbook and browse through properties like one might do for steel, aluminum or plastics. The best and only source of information is the vendor. Vendors often have a large portfolio of alloys, but little design data is collated, written down and generally available to public. They are eager to help and will select a specific alloy if they know the details of the application, but that's not always either possible or practical.

For electrical actuators, an A_s above room temperature is required and one quickly learns that the highest temperature, readily available Ni-Ti alloy has an A_s around 100°C (there are some experimental higher temperature alloys, but they are not readily available). So material was purchased for evaluation from several sources to the following very simple "nominal"specification:

- highest possible transformation temperature
- as-drawn condition
- 0.15mm diameter
- nickel-titanium based

Note that alloy chemistry was not specified, but left to the vendor's good judgement to supply their best material based on the sketchy (but sufficient) requirement. Modest differences in the performance of the alloys had to be expected. Using as-drawn material (typically with 20-30% cold work) permits greater flexibility in processing, and it seemed impractical to initially ask for a specific amount of cold work. As received, all had a layer of oxide which varied from very thin to thick to "chippable". Many had drawing lubricants still on them. Measured diameters (note that the tolerance was not specified) varied from 0.13mm to 0.157mm (+3/-12%) including any oxide or drawing lubricants. And finally, the evaluation excluded copper-based or exotic (e.g. palladium-rich) alloys for simplicity.

Table 1 shows the alloys delivered and evaluated. Alloy F323 was originally manufactured by Furakawa Electric. It's believed that 1032 is of similar origin but

Table 1: Alloys evaluated

<u>Alloys</u>	<u>Vendor</u>	<u>Comments</u>
B266	Raychem	
K12	Raychem	Copper doped
F323	SMA	
1032	Toki	Vendor annealed
B162	U.S. Nitinol	

undergoes subsequent processing by another vendor. 1032 was the only material evaluated "as-delivered". The copper-doped K12 is a variant from strictly nickel-titanium, but an important one for electrically excited devices. The evaluation described herein is neither a recommendation nor indictment of any specific material. All of these materials have demonstrated good electrically-excited actuators of one kind or another.

Table 2 shows independent chemical analysis of each sample. Percentages are by weight and each component part was individually measured (as opposed, for

Table 2: Alloy composition

Weight Percent

<u>Alloys</u>	<u>Nickel</u>	<u>Titanium</u>	<u>Copper</u>
B266	54.24	45.87	
K12	45.52	45.98	10.99
F323	55.53	46.12	
1032	54.97	45.65	
B162	54.96	45.12	

example, to measuring only nickel and assuming the remainder is titanium) so the totals do not add exactly to 100%. This gives a good check and measure on the accuracy of the analysis.

Samples were nominally one inch long and operated only in a tensile mode with a fixed (dead-weight) load. So we are dealing with an actuator design that will lift 130-320g (70-175 MPa) a distance of 0.25-1.25mm (1-5%).

2.0 Processing:

Processing was limited to a simple 20 minute air anneal and cool between 400°C and 525°C depending on the amount of cold work in the as-delivered materials. Multiple annealing temperatures were tested on each material, the goal being to leave some amount of cold work to minimize cyclic creep or *walking* as it's often called.

3. Stress-Strain Evaluation:

Figures 1-5 show stress-strain curves for each material. Curves are shown as a collection of unconnected points because they are derived by heating and cooling (current on/off) at a fixed load (stress) instead of pulling the sample isothermally. Since the material is path-dependent, very different results will result between this and the isothermal approach. It is also more consistent with the way the material will be used. High stress data (above 200MPa typically) incurs some yielding on successive cycles, but actuator designs seldom, if ever, work at these levels.

Alloy F323 has the characteristic martensitic plateau which is absent in all other samples. It's believed this is a compositional difference (see Table 2), not a

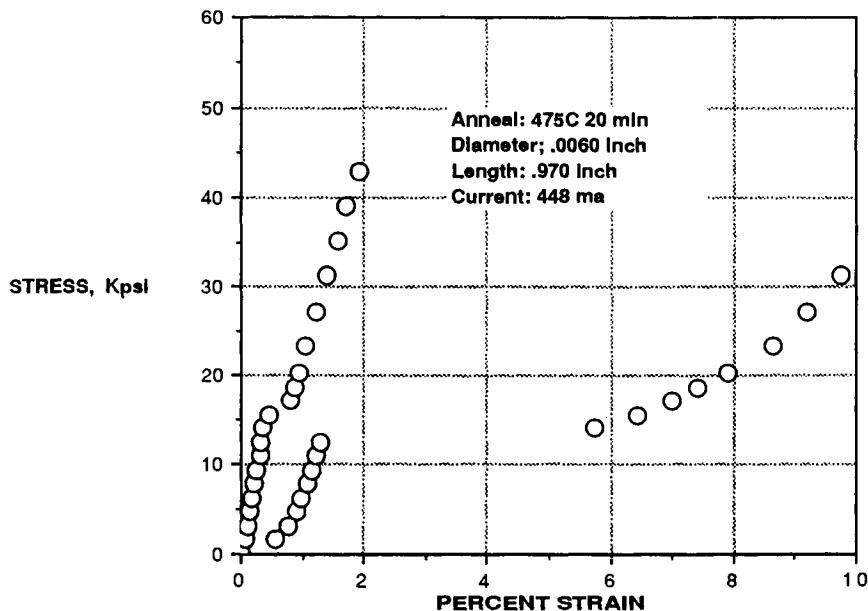


Figure 1: Alloy F323: stress-strain behavior.

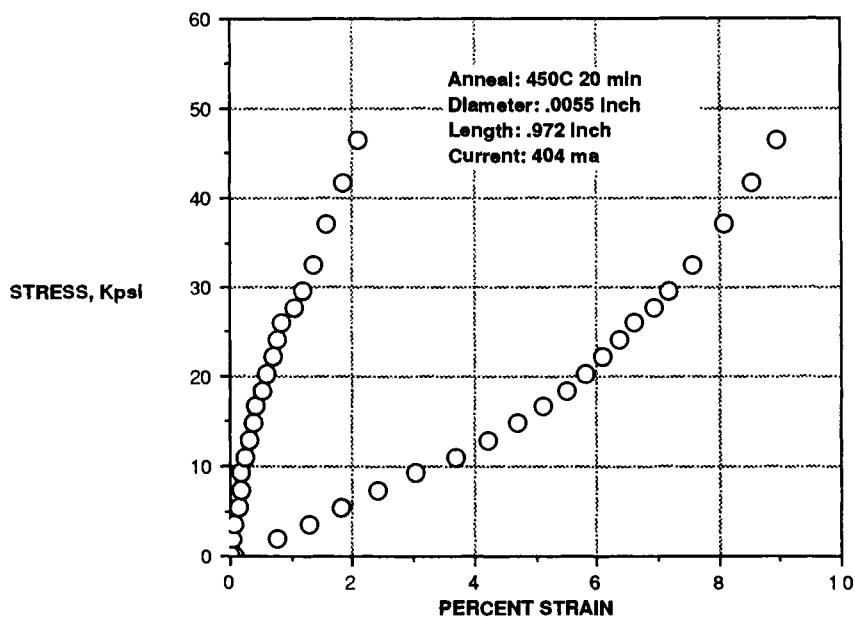


Figure 2: Alloy B266: stress-strain behavior.

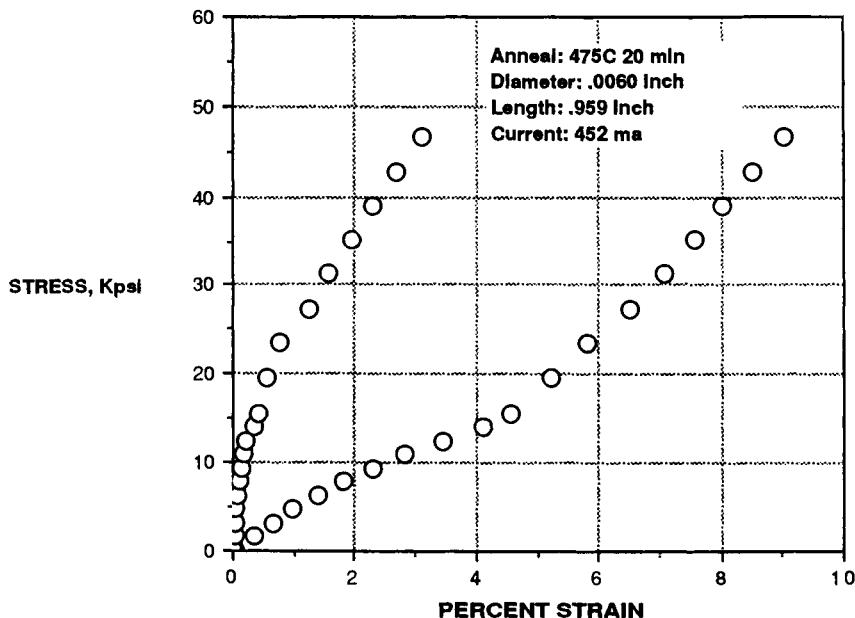


Figure 3: Alloy B162: stress-strain behavior.

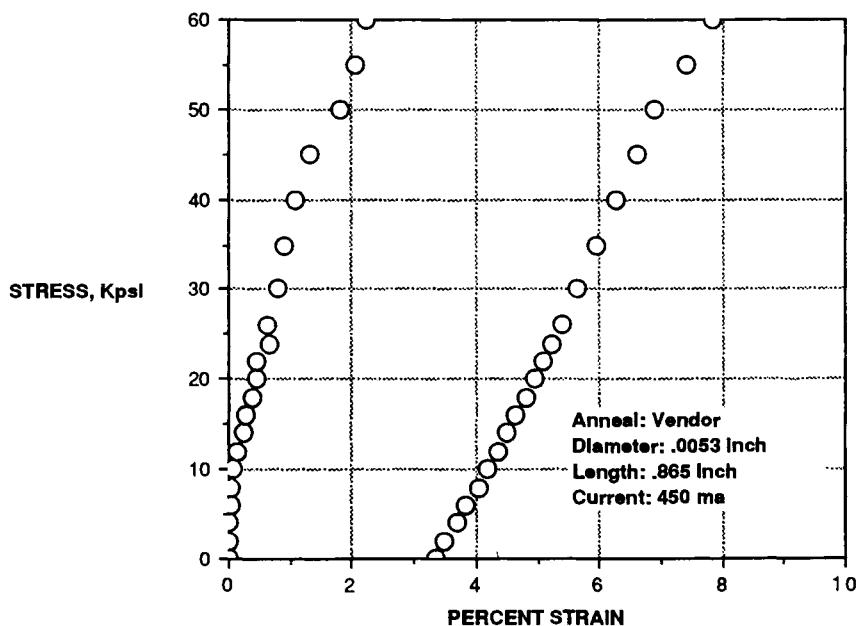


Figure 4: Alloy 1032 stress-strain behavior.

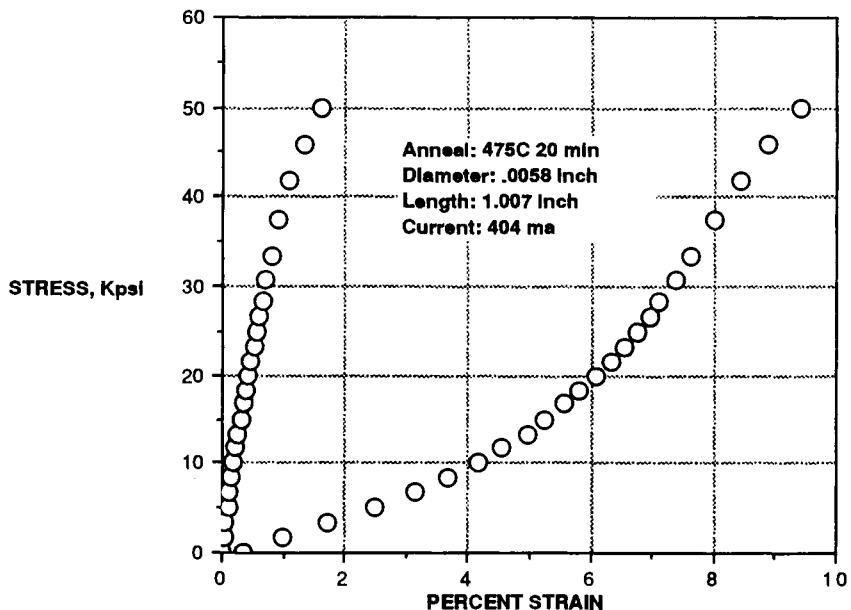


Figure 5: Alloy K12 stress-strain behavior.

processing difference. Conversely, the characteristic two-way effect at zero stress shown by alloy 1032 in Figure 4 (whose origin is believed similar to F323) is clearly due to special processing. Based on stress-strain performance alone, alloy 1032 would be the choice for a design requiring a zero-stress two-way effect. F323 might be the choice for a high-stress, high-strain combination but would probably not be used at stresses below about 100 MPa. Alloys B266 and B162, on the other hand, are useful at lower stresses and might be expected to give good fatigue performance because their strain is limited, while K12 is a nice compromise for several criteria.

These differences are all confusing since the design engineer may not know what he will get when ordering material. The dilemma is whether to specify a known stress-strain characteristic or become sufficiently knowledgeable to specify composition and processing. The former is most desirable, but the latter probably most economical in production since it requires less testing by the material supplier. Perhaps it's fortunate that different "performances" may be obtained from nominally similar materials. In the long run these materials collectively offer a rather nice portfolio of mechanical design choices. Also, as the demand for more varied wire performance increases and design engineering questions more demanding, the vendors will clearly add more data, more "options" and generally make their portfolios more accessible.

4. Hysteresis Evaluation:

The stress-strain characteristic is only one of the parameters to be examined during the evaluation of shape memory alloys. Figure 6 shows current-driven hysteresis curves for the different alloys. Displacement is plotted while current is slowly increased and decreased (the dots are one second apart) at constant load and temperature. This approach (as opposed to plotting displacement versus temperature) is more appropriate to the electrically excited device and permits accurate quick comparison of material performance. The vertical scale on Figure 6 and subsequent figures is the height above a reference plane at which the tensile wire "lifts" a load with the applied current. Allowing for some minor differences in diameter, ambient temperature and perhaps resistivity, Figure 6 indicates all alloys have nominally the same transformation current (temperature) with 1032 being the highest and K12 lowest. An appropriate safe exciting current would be 350-450 mA, somewhat dependent on alloy if the speed of actuation is not an overriding concern. Hysteresis width varies from 60 mA (K12) to 160 mA (1032), which is particularly important with electrically excited devices. It affects reset/cycle time and, along with the actuation temperature (A_s), the upper operating temperature limit. B266 has the squarest hysteresis and would most likely give the "snappiest" action for the lowest current if that's important. Hysteresis squareness is affected by most parameters (anneal, load, etc.) and will improve ease of operation over extended ambients. Over-closure of the hysteresis loop at zero current (1032) is a first indication that cyclic creep may be present for the conditions tested. For proportionally controlled devices, the "two-stage" hysteresis characteristic (K12) of the copper-doped material may be desirable.

5. Effects of Ambient Temperature:

To obtain another piece of the jigsaw puzzle, it's necessary to repeat the above current-driven hysteresis curves at different ambient temperatures as shown in Figure 7. As expected, high ambient temperatures require less current to begin actuation at A_s . Somewhere near 55°C the part fails to reset to its original zero current position.

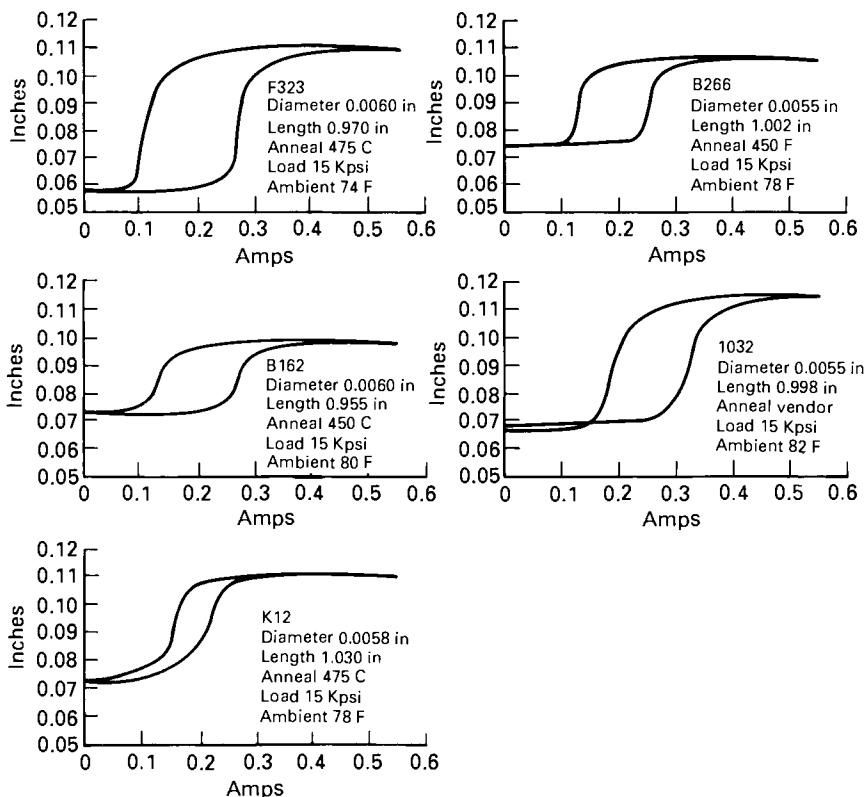


Figure 6: Current driven hysteresis curves.

It's left partially actuated and will remain there until the ambient temperature falls below the M_f temperature. The hysteresis widening with higher ambient temperatures is real and a result of the \sqrt{T} dependence of the actuation current. For any cyclic electrically excited device, therefore, the practical upper limit of ambient temperature is limited to less than the M_f temperature (about 55°C in this case). This is a rather severe limitation for broad application of electrically excited devices, particularly when the prospect for economical higher- M_f temperature alloys is not bright.

Most of the evaluated alloys are similar to Figure 7 except K12 which is shown in Figure 8. The two-stage transformation makes it difficult to assign meaningful unique A_s and M_f values and the zero current position is much more sensitive to ambient temperature. For a temperature sensitive or proportional control, this latter effect is advantageous, while a binary actuator between two fixed points might be difficult to compensate. This sensitivity at zero current is illustrated in Figure 9 for B162 and K12. In the temperature range measured, both have an "effective" negative coefficient of expansion (Figure 9 is shown principally for comparison and is not a rigorous measurement of thermal coefficients). That is, the tensile wires get shorter with increasing temperature over the range measured.

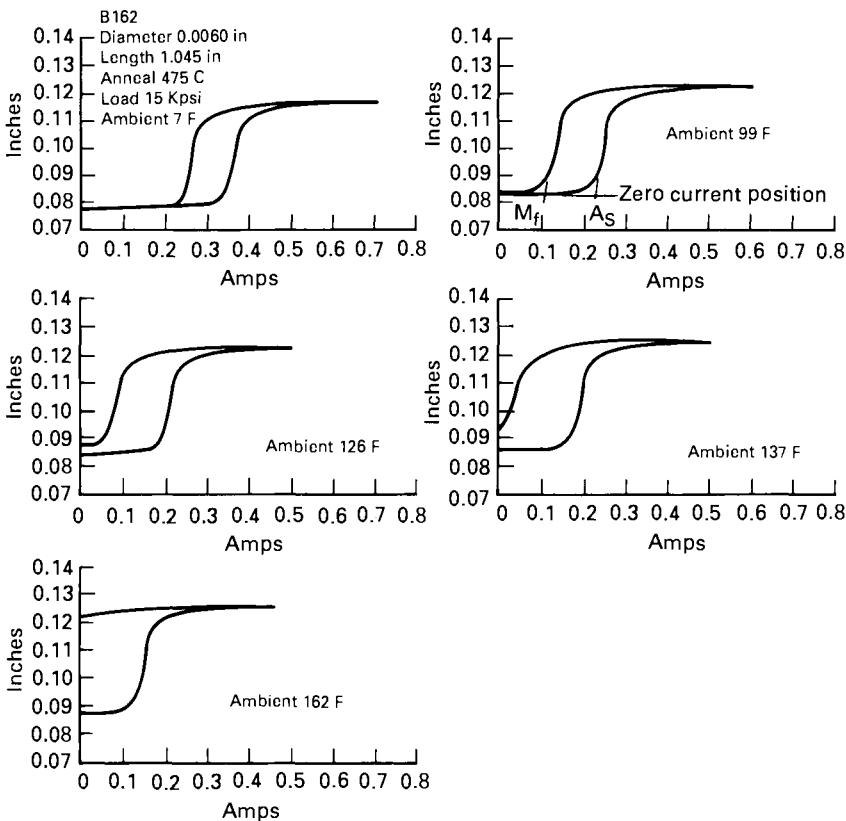


Figure 7: Alloy B162: hysteresis versus ambient temperature.

Figures 10-14 show M_f and A_s (the two most important transformation temperatures for electrically excited elements) plotted versus ambient temperature. In contrast to the varied mechanical data, these curves are remarkably similar. One obvious difference in Figures 10-14 is the amount of current required for seemingly similar samples. This is largely a result of different resistivities as listed in Table 3. Note specifically the difference between 1032 and K12. The increased resistance of K12 can be a benefit when designing a driver for an actuator which is nearly a short-circuit.

6. Temporal Evaluation:

Up to this point, current has been applied slowly to evaluate thermal-mechanical effects, but the electrically excited actuator always has an on/off time consideration. Figure 15 shows the time response of a typical 0.15 mm diameter element of different ambient temperatures. As expected, actuation time decreases, stroke increases and reset time increases as ambient temperature increases for the set of conditions chosen. The degree of reset time variation, however, may be surprising and is most troublesome. This is shown graphically in Figure 16 where actuation/reset times are

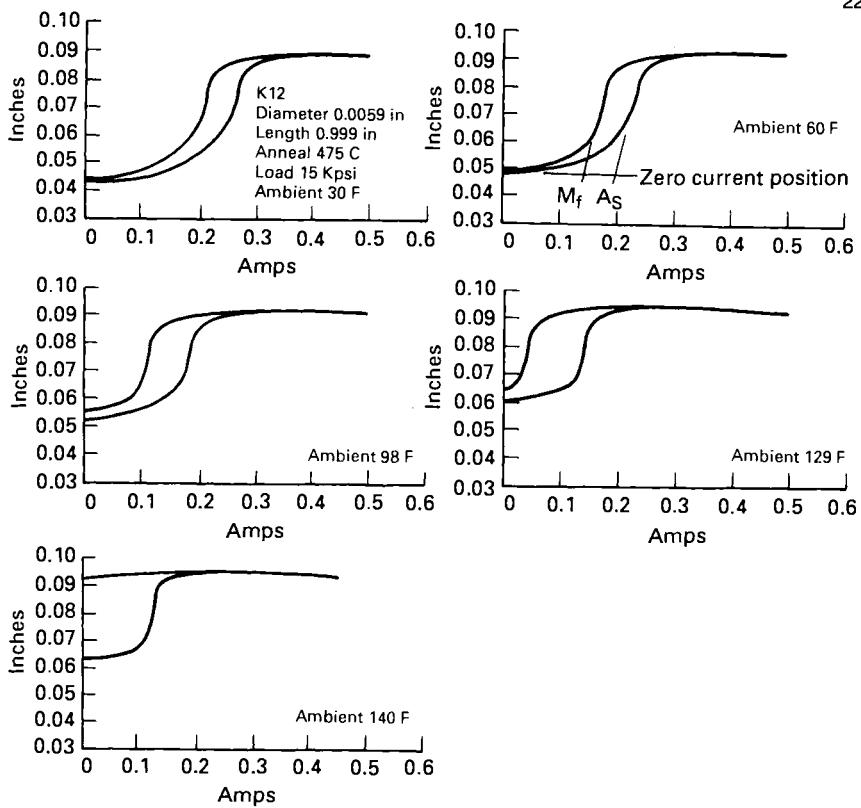


Figure 8: Alloy K12: hysteresis versus ambient temperature.

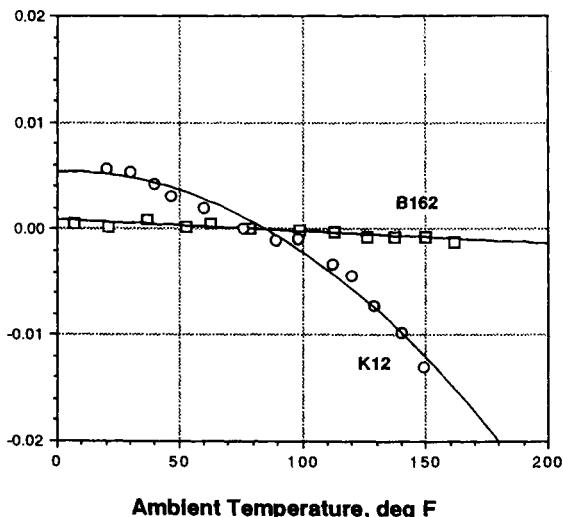


Figure 9: Relative length change versus ambient temperature, normalized to room ambient.

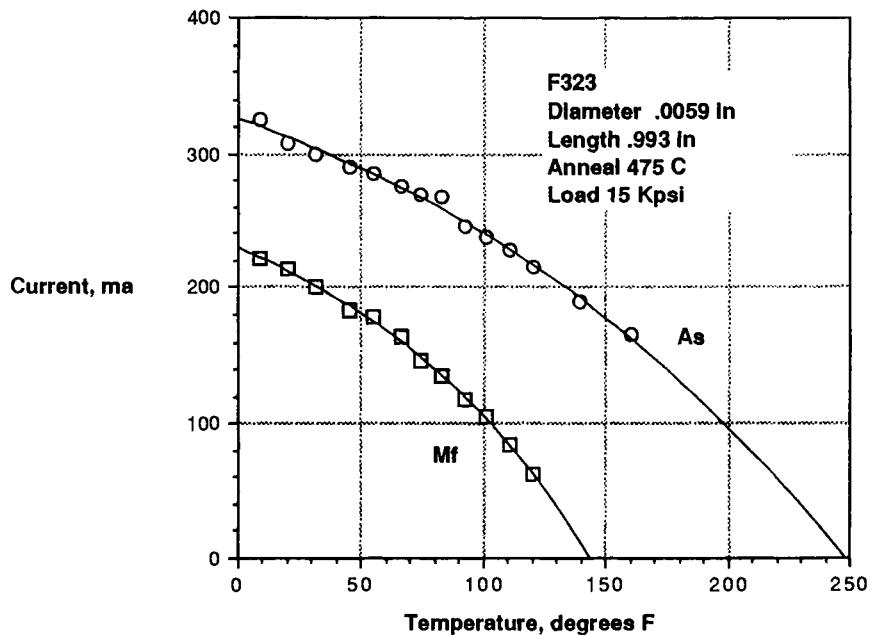


Figure 10: Alloy F323: A_s and M_f versus ambient temperature.

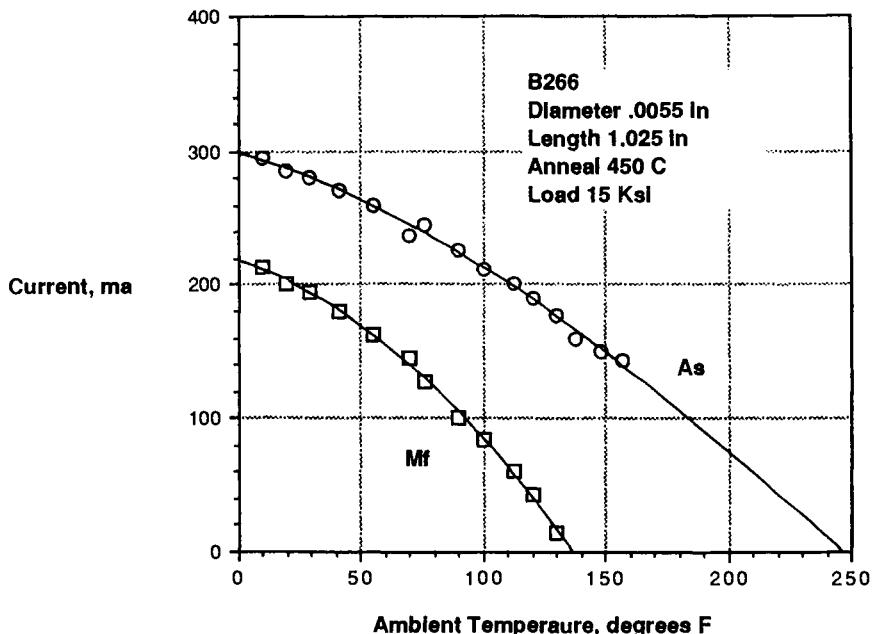


Figure 11: Alloy B266: A_s and M_f versus ambient temperature.

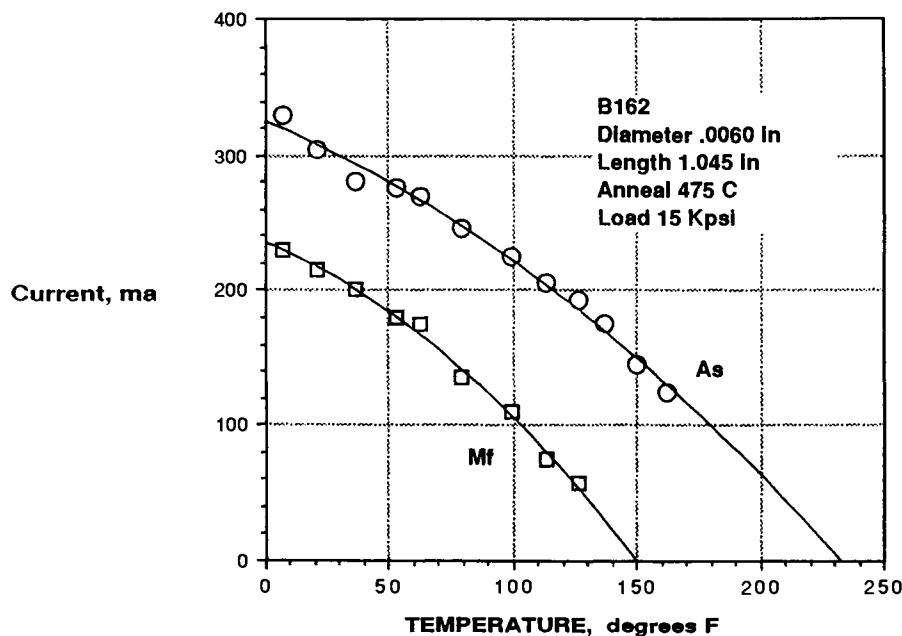


Figure 12: Alloy B162: A_s and M_f versus ambient temperature.

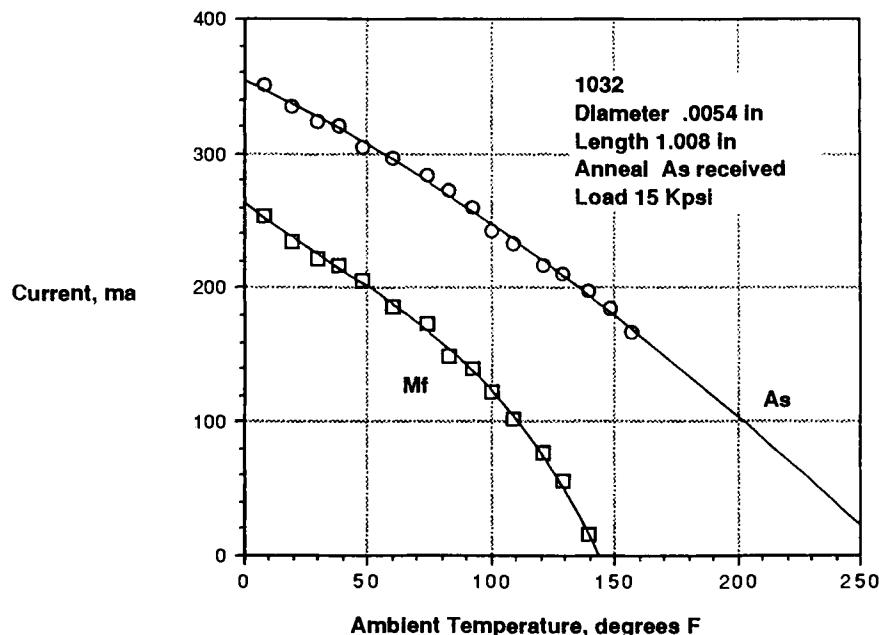
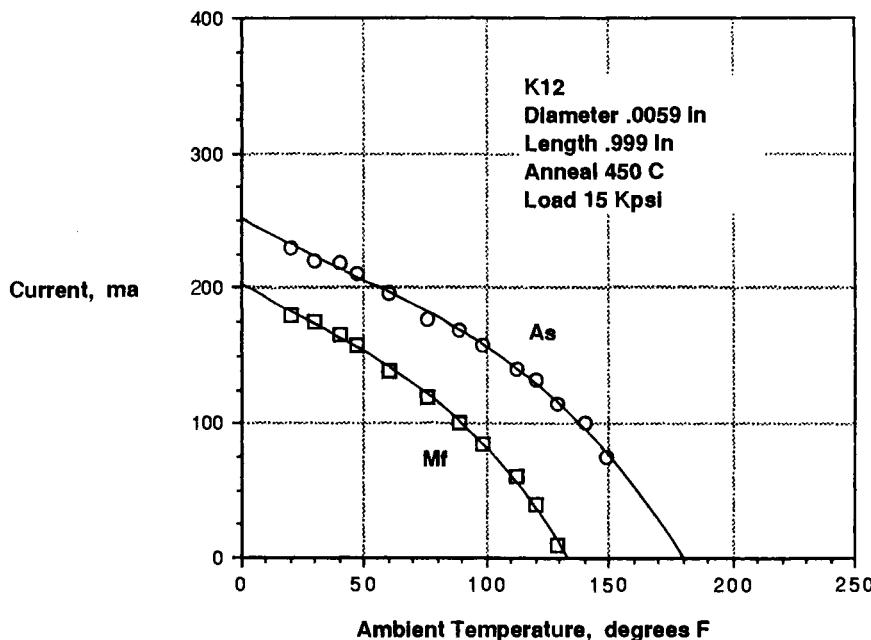


Figure 13: Alloy 1032: A_s and M_f versus ambient temperature.

Figure 14: Alloy K12: A_s and M_f versus ambient temperature.**Table 3: Resistivity****Resistivity (micro-ohm-inch)**

Alloys	A_s	A_f
B266	36.6	31.1
K12	34.9	31.4
F323	38.0	34.6
1032	24.8	22.6
B162	40.9	35.5

plotted for different 0.14 mm diameter elements. The dramatic effect and the importance of the M_f temperature is evident in Figure 16.

7. Energy Evaluation:

Data like Figure 16 can vary widely with alternative currents, pulse widths, repetition frequencies and ambient conditions other than temperature. Consider, for example, the data and theory of Figure 17 where the actuation energy is shown versus the applied current. Above about 0.5 Amps, the element reacts quicker than convection, the required energy is nearly constant and heating of the wire should be nearly equal over its length.

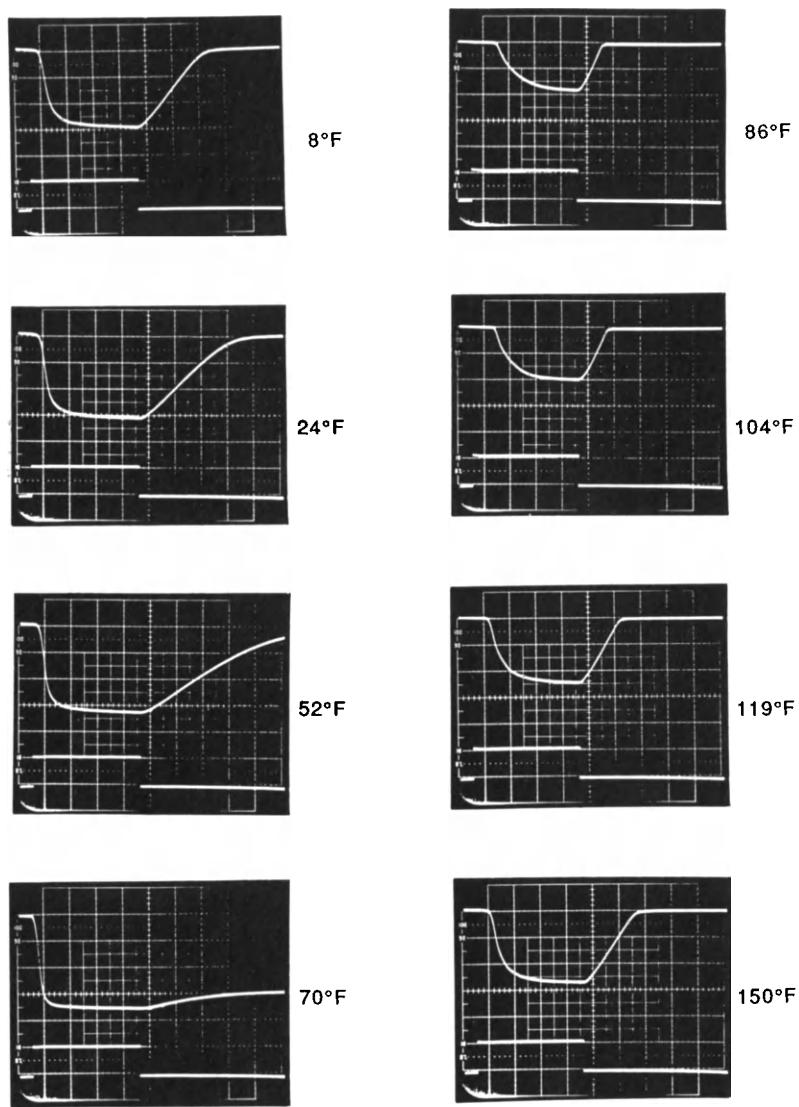


Figure 15: Actuation, reset response versus ambient temperature (diameter 0.15 mm, current 550 mA, and 1.5 seconds/division).

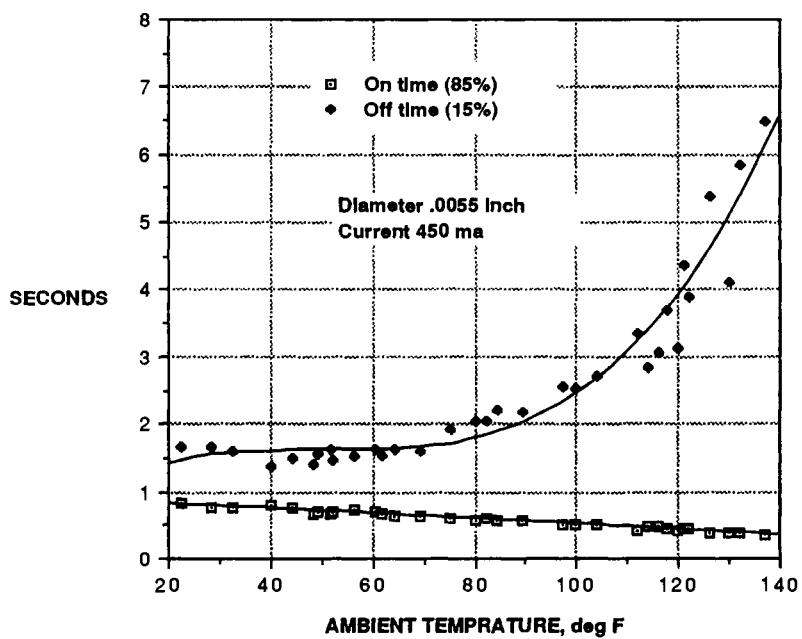


Figure 16: Typical response time versus ambient temperature.

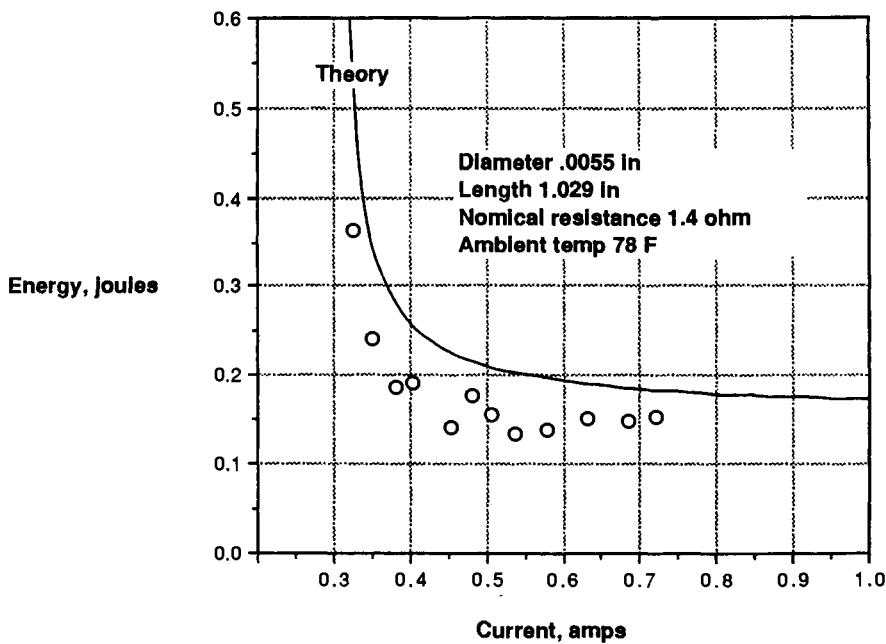


Figure 17: Energy to actuate versus applied current.

8. Conclusion:

Electrically excited shape-memory actuators can offer unique advantages, but the design scenario requires detailed understanding and often complicated trade-offs. A good approach for collecting data includes:

- Measuring mechanical data (stroke vs. load) for different material processes.
- Measuring hysteresis data (stroke vs. current) at various ambient temperatures.
- Measuring temporal data (stroke vs. time) for different ambient temperatures.
- Repeating the above as required for different materials and dimensions.

With this portfolio of data, the first-time-design has a high probability of success and any limitations of using shape-memory for the specific application will usually be obvious and quantifiable. Data presented herein should also give the would-be designer an added head start in this exciting, fun and often frustrating technology.

Design Principles For Ni-Ti Actuators

Tom Waram
Raychem Corporation
300 Constitution Drive
Menlo Park, Ca 94025

The aim of conventional helical spring design is to produce a mechanical element that will store energy by generating the desired forces at given deflections. The designer uses standard formulae based on linear elastic theory to determine wire diameter, spring diameter and number of turns. With a material such as steel, the force deflection behavior of a helical spring obeys Hooke's law, that is, there is a linear relationship between the force and the deflection of the spring, with the constant of proportionality known as the spring rate (Figure 1).

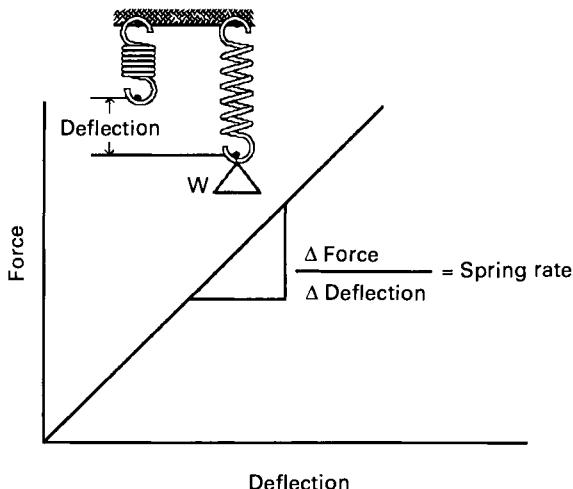


Figure 1: Steel spring behavior

Springs made from Ni-Ti for actuator applications are not used to store energy, but rather to perform useful work. They are ideally suited for use as thermal actuators, behaving as both a temperature sensor and work generating element.

From a design standpoint, the behavior of a shape memory spring is best described in terms of material rigidity. Ni-Ti exhibits a large change in shear modulus over a

relatively narrow temperature range, increasing from low to high temperature. This produces a concomitant increase in spring rate, since spring rate is directly proportional to shear modulus (equation 6 of the Appendix). The change in modulus with temperature is the result of a reversible martensite to austenite solid state phase transformation and related pretransformational phenomena, which have been described earlier. As shown in Figure 2, at low temperature a spring made from Ni-Ti is

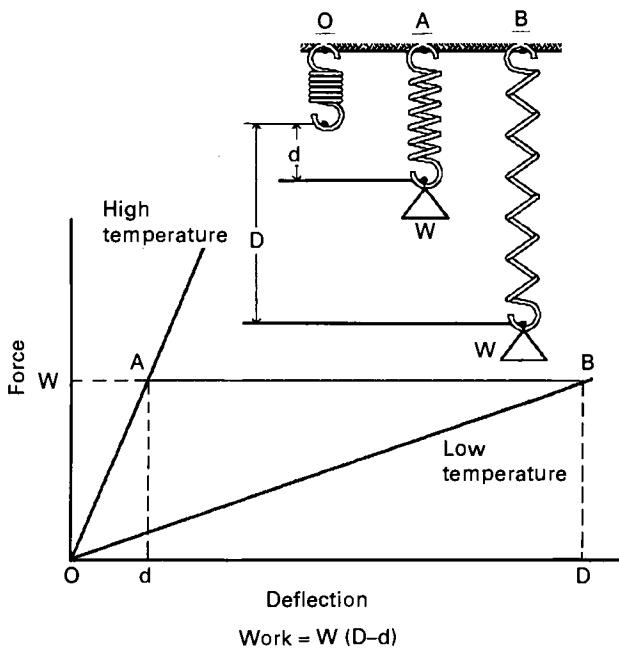


Figure 2: Shape memory spring behavior

loaded along line OB by weight W to point B, with a deflection D. When the temperature is raised the spring contracts to point A with a deflection d and a net motion of $D - d$, thus performing work equal to $W(D - d)$. When the material is again martensitic the weight extends the spring to point B, resetting it for another contraction.

1. Hysteresis Behavior

An important characteristic of the heating and cooling behavior of shape memory springs is the hysteresis that occurs. As shown in Figure 3 at a temperature below M_f the spring is 100 % martensite and fully extended. During heating the the spring begins contracting at A_s and completes its motion at A_f when the spring is 100 % austenite. During cooling the spring does not begin to extend (reset) until M_s is reached. Extension is complete at M_f .

Transformation temperatures are alloy dependent and can be altered by changing alloy composition. Hysteresis width is also alloy dependent. Transformation temperatures also depend on stress level, increasing with shear stress (Figure 4), according to the Clausius-Clapeyron relation discussed earlier.

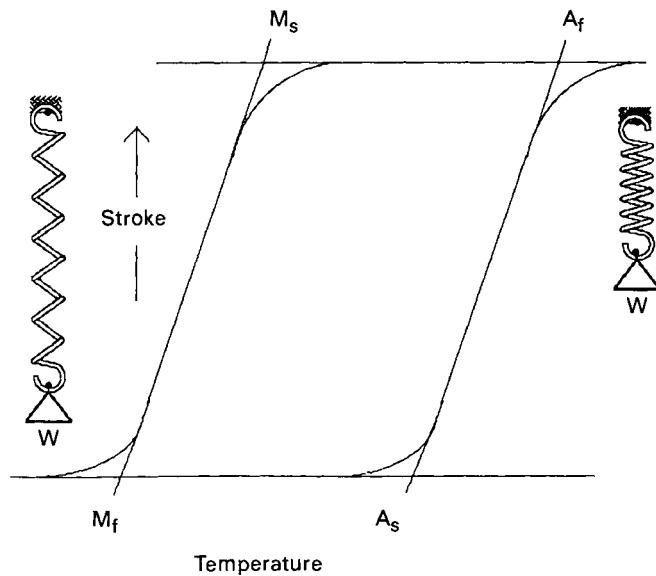


Figure 3: Temperature hysteresis

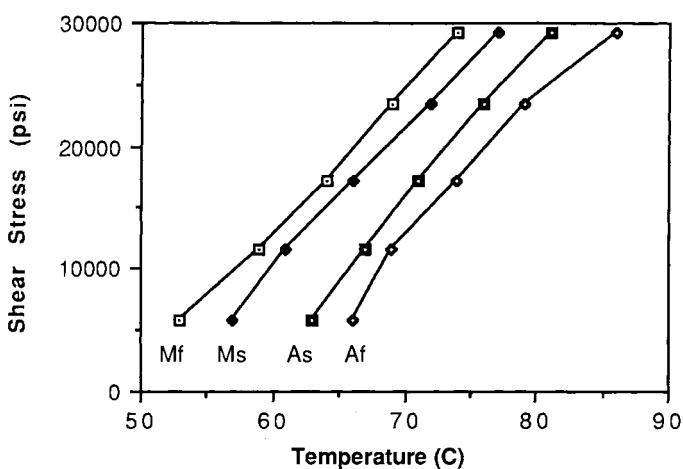


Figure 4: Effect of stress on transformation temperatures

2. Basic Design Concepts

Both extension springs and compression springs can be made from Ni-Ti. As illustrated in Figure 2, an extension spring provides a pulling force by contracting when transformed to the high temperature austenitic state. The close coiled, zero load configuration of the extension spring is the shape which is "remembered" at high temperature. This shape is set by constraining a close wound spring on a mandrel and giving it a special heat treatment. It is important for the designer to understand that all deflections, in both the austenitic and martensitic states, are measured with respect to this closed configuration (Figure 5). One should also be aware that there can be a

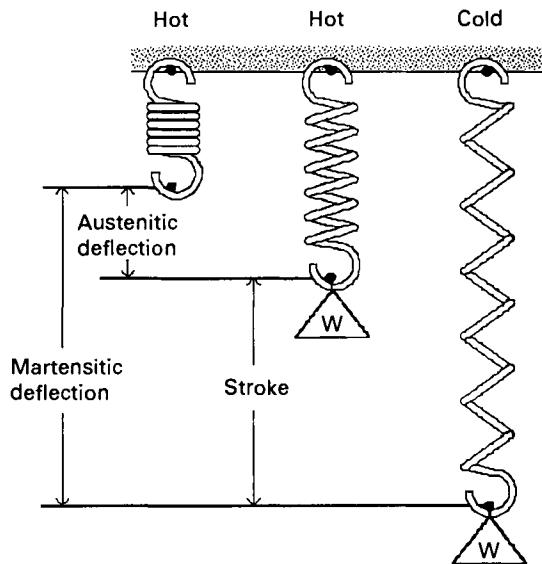


Figure 5: Extension spring deflections

pre-load in the close wound configuration that can cause shifts in transformation temperature (this is discussed in more detail in the next chapter).

A compression spring provides a pushing force by extending at high temperature. The length to which a compression spring is shape set is analogous to the free length of a steel spring and all deflections are determined with respect to it (Figure 6).

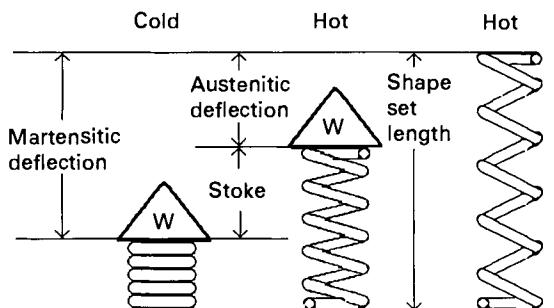


Figure 6: Compression spring deflections

Although load-deflection curves of Ni-Ti springs in the austenitic state tend to exhibit curvature at higher deflections (decreasing slope), a characteristic linear region exists at lower deformations that can be utilized in a large number of actuator designs. Figures 7 and 8 show the high temperature (100°C) load-deflection curves for several

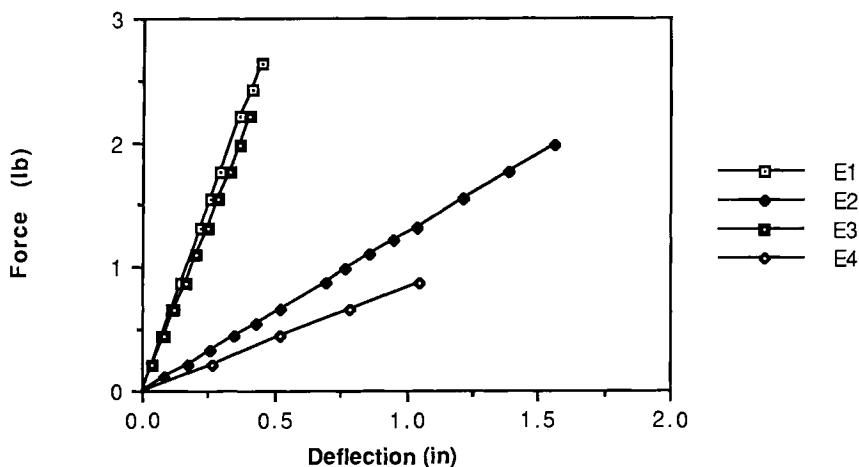


Figure 7: Extension springs (100°C)

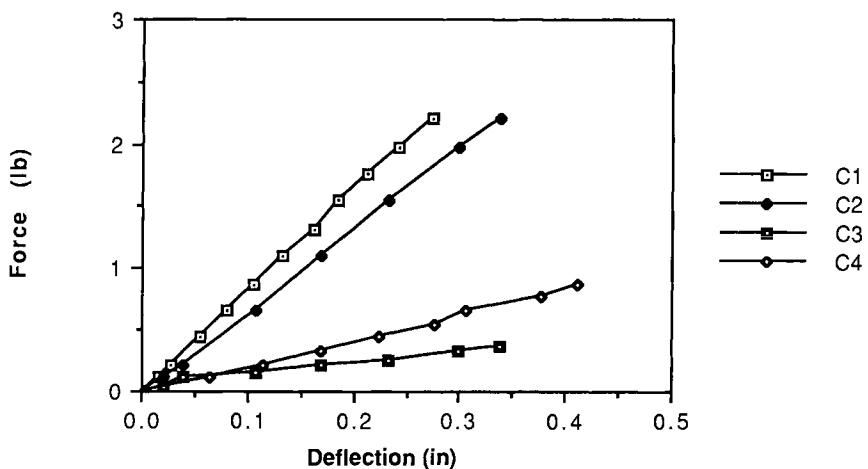


Figure 8: Compression Springs (100°C)

different configurations of extension and compression springs made from a Ni-Ti alloy with a nominal A_f temperature of 55°C . The dimensions of the various springs are given in Table 1.

Table 1: Spring Dimensions

<u>Spring</u>	<u>Wire Diameter (mm)</u>	<u>Outer Diameter (mm)</u>	<u>Active Coils</u>
E1	0.7	4.8	12.0
E2	0.7	6.6	18.5
E3	1.2	9.1	10.5
E4	1.2	13.7	20.0
C1	0.7	4.3	12.0
C2	1.2	9.1	10.5
C3	0.5	5.3	12.0
C4	0.7	6.5	14.5

Figure 9 shows the load-deflection curves of these springs converted to stress-strain using the ordinary equations for shear stress and strain (Equation 3 and 7 respectively, of the Appendix).

Although some scatter is evident, the stress-strain curves tend to overlay and produce a characteristic straight line. Regression analysis of the data in Figure 9 produced a

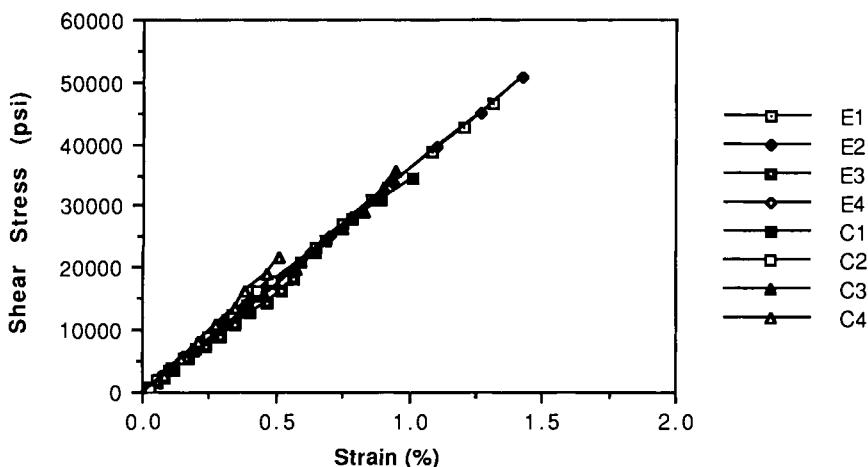


Figure 9: Load-deflection data converted to stress-strain

representative line with a slope of 24500 MPa. This value can be used as the effective linear shear modulus at 100°C in design calculations for this alloy. A similar treatment can be used to determine an effective shear modulus for springs in the martensitic state. This is because the load-deflection behavior of martensitic springs can be closely approximated using a linear relationship. This differs from the idealized "plateau" behavior shown in Figure 7 of the previous section.

3. Design Example

Assuming linear force-deflection behavior, the designer must only have a knowledge of the effective shear modulus at high and low temperature to construct force-deflection curves for any given spring geometry. The design procedure can be broken into two stages: (1) determination of the wire diameter and average spring diameter by constraining the maximum shear stress and the spring index, and (2) determination of the number of active coils in the spring through constraint of the shear strain. The following example illustrates this procedure for an extension spring required to provide a constant force of $P = 4\text{N}$ over a stroke length of 10mm. All equations referenced are given in the Appendix.

- (1) Diameter determination: Assume a spring index of $c = 6$ and a maximum shear stress of $\tau_c = 100 \text{ MPa}$. The wire diameter d is calculated by rearranging equation 5:

$$d = \sqrt{\frac{8Wpc}{\pi\tau_c}}$$

where W is the Wahl correction factor, which has a value of 1.2525 for a spring index of 6.

Solving:

$$d = \sqrt{\frac{8 \times 1.2525 \times 4 \times 6}{(3.14159 \times 100)}}$$

$$= 0.875 \text{ mm}$$

The average diameter D can be determined by rearranging equation 2:

$$D = cd = 6 \times 0.875 = 5.25 \text{ mm}$$

- (2) Determination of number of turns: The following values for the moduli are assumed:

High Temperature Modulus = $G_H = 24500 \text{ MPa}$

Low Temperature Modulus = $G_L = 2700 \text{ MPa}$

The high temperature shear strain γ_H is given by:

$$\gamma_H = \tau_c/G_H = 100/24500 = 0.00408 \text{ or } 0.408\%$$

The low temperature shear strain, γ_L , is limited to 0.015 or 1.5%.

The shear strain associated with the stroke, γ_S , is simply the difference between the high and low temperature shear strains:

$$\begin{aligned} \gamma_S - \gamma_L - \gamma_H &= 0.015 - 0.00408 \\ &= 0.0109 \text{ or } 1.09\% \end{aligned}$$

Using equation 7 in a difference form and solving for n, the number of active coils, yields:

$$\begin{aligned} n &= (d \times \text{stroke}) / (\pi \times D^2 \times \gamma_S) \\ &= (0.875 \times 10) / (3.14159 \times 5.25^2 \times 0.0109) \\ &= 9.27, \text{ rounding to 9.} \end{aligned}$$

Using equation 6, the high and low temperature spring rates, K_H and K_L can be calculated:

$$\begin{aligned} K_H &= G_H d^4 / 8nD^3 \\ &= (24500 \times 0.875^4) / (8 \times 9 \times 5.25^3) \\ &= 1.38 \text{ N/mm} \\ K_L &= G_L d^4 / 8nD^3 \\ &= (2700 \times 0.875^4) / (8 \times 9 \times 5.25^3) \\ &= 0.15 \text{ N/mm} \end{aligned}$$

The high temperature deflection, δ_H , is thus

$$\begin{aligned} \delta_H &= \text{force}/K_H \\ &= 4\text{N}/1.38 \\ &= 2.9\text{mm} \end{aligned}$$

and the low temperature deflection, δ_L , is

$$\begin{aligned} \delta_L &= \delta_H + \text{stroke} \\ &= 2.9 + 10 \\ &= 12.9\text{mm} \end{aligned}$$

The reset load is calculated by:

$$\begin{aligned} \text{minimum reset force} &= \delta_L \times K_L \\ &= 12.9 \times 0.15 \\ &= 1.94 \text{ N} \end{aligned}$$

A mechanical stop will therefore be required to keep the weight from over extending the spring on the return stroke.

Summarizing the results, the spring will have the following specifications:

wire diameter	= 0.875mm
average diameter	= 5.25mm
active turns	= 9

4. Designs Using Biasing Springs

The system shown in Figure 2 is self resetting, that is, the constant load applied by the weight deforms the shape memory spring at lower temperature. In most practical situations however, a conventional steel spring is used to provide a biasing force to reset the shape memory spring at lower temperature (Figure 10). The shape memory

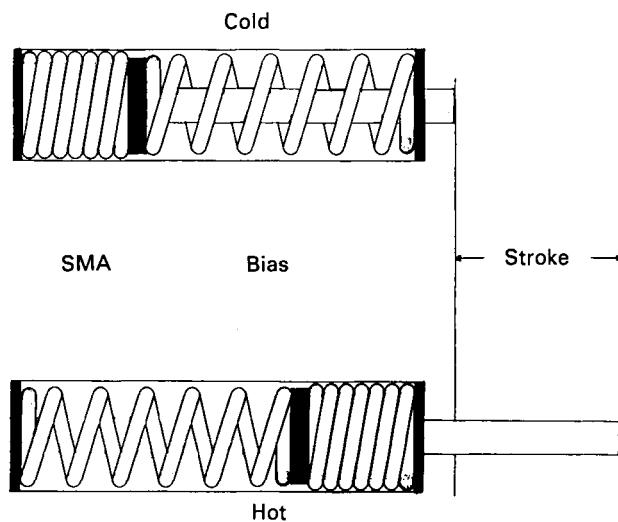


Figure 10: Shape memory spring with biasing spring

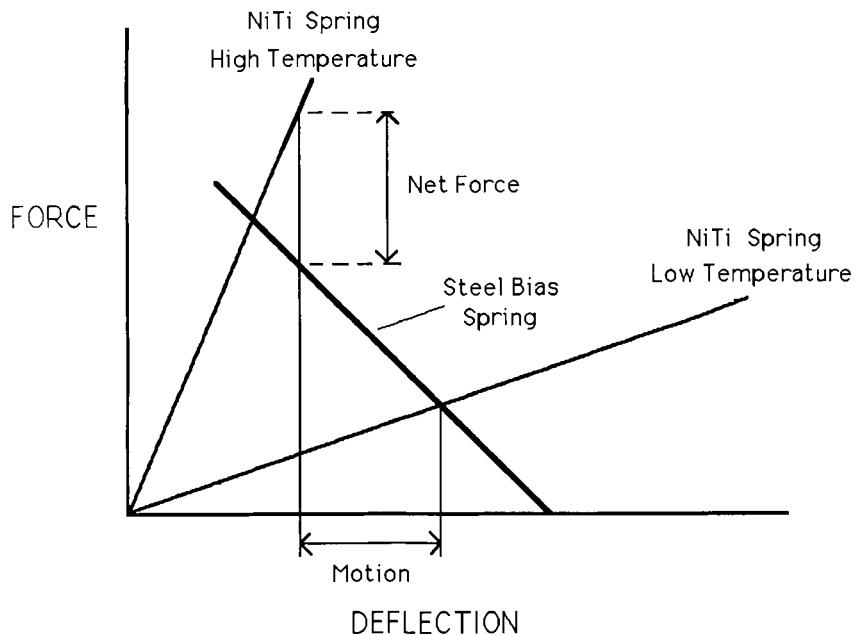


Figure 11: Generalized design using biasing spring

spring has to work against an ever increasing force from the biasing spring and the net force output of the spring diminishes as the temperature increases (Figure 11).

For a design requiring a biasing spring such as that depicted in Figure 10, the design procedure is similar to that outlined above. If the biasing force at high temperature is known, it is simply added to the required net output force and the design proceeds as above. When the low temperature reset force is determined, the biasing spring rate is found using:

$$\text{biasing rate} = \frac{\text{high temperature bias force} - \text{reset force}}{\text{motion}}$$

The biasing force at high temperature must be chosen somewhat arbitrarily and several design calculations may be required to determine optimum geometries for the shape memory and bias springs. Simple computer programs can be written, especially using spreadsheet type software, that can greatly simplify design procedures. These programs use a "what if" approach, i.e. rapid variation of design parameters to determine optimum design configuration.

5. Summary

The behavior of Ni-Ti shape memory springs can be described using a linear stress-strain model. The design equations developed for linear elastic materials can be applied with good results. This technique utilizes effective shear moduli derived from experimental load-deflection data. Although this technique is not appropriate for designs where large austenitic strains occur (leading to nonlinearity), it has proven to be a useful design model for many actuator applications.

6. Reference

- Wahl, A.M.: Mechanical Springs, 2nd Ed., McGraw-Hill Book Company, 235 (1963).

Appendix

(a) Terms:

δ	deflection
d	wire diameter
n	active turns
c	spring index
O.D.	spring outer diameter
D	spring average diameter
τ	shear stress
γ	shear strain
τ_c	corrected shear stress
p	applied force
W	Wahl correction factor

K spring rate
G shear modulus

(b) Design Equations:

- (1) $D = O.D. - d$
- (2) $c = D/d$
- (3) $\tau = 8pD/\pi d^3 = 8pc/\pi d^2$
- (4) $W = (4c - 1)/(4c - 4) + 0.615/c$
- (5) $\tau_c = \tau W = 8Wpc/\pi d^2$
- (6) $K = Gd^4/8nD^3$
- (7) $\gamma = d\delta/\pi n D^2$

Design Principles for Cu-Zn-Al Actuators

Richard F. Gordon, P.E.

Memory Metals, Inc.

83 Keeler Avenue

Norwalk, Connecticut, 06854

Two-way shape memory material creates an opportunity for developing a wide variety of thermally sensitive actuators. With proper design, a shape memory element can be used as a thermo-mechanical actuator to lift loads, to proportionally control a thermal system, or to operate thermal switches. In many cases a shape memory actuator can be designed to outperform conventional wax, gas, and bimetallic actuators. As costs are reduced more shape memory actuators will compete for thermo-mechanical applications in valve, venting, and electrical industries. The cost of a shape memory element is ultimately determined during the design process. To quickly assess cost it is convenient to have a methodology for determining the size and complexity of the shape memory element being designed. The approach taken in this paper is to introduce important design considerations and show a methodology for quickly sizing shape memory components. The focus will be on Cu-Zn-Al two-way shape memory actuators.

1. Basic Two-Way Shape Memory Properties

Figure 1 shows a variety of two-way shape memory elements that are available to the designer. Each shape memory element has its own set of unique operating characteristics. The axial expansion of a shape memory compression spring makes it a convenient element for use as a thermo-mechanical actuator. In this paper we concentrate on the design process of shape memory compression springs.

Shape memory material is first processed into wire of appropriate diameter and then wound into springs. Following winding, the coil passes through a number of heat treatment procedures to establish the two-way memory characteristic. The finished coil will contract to its closed coil bound state at temperatures below M_f and expand axially to its open coiled state at temperatures above A_f .

As designers we are concerned with the temperature at which macroscopic motion begins. We define this temperature as A_s , the temperature axis intersection of the tangent to the deflection/temperature curve on heating. The transformation temperature (A_s) is a function of the alloy composition and process heat treatment.

A common design misconception is the notion that shape memory material can be used to resist stress when the temperature is below M_f . Martensitic material is physically very weak, rubbery, and will stress relieve. On heating, a shape memory

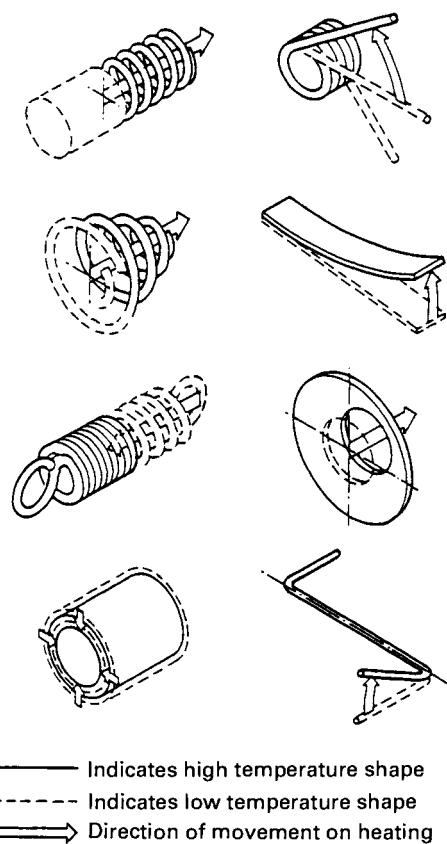


Figure 1: Various forms of Shape memory elements

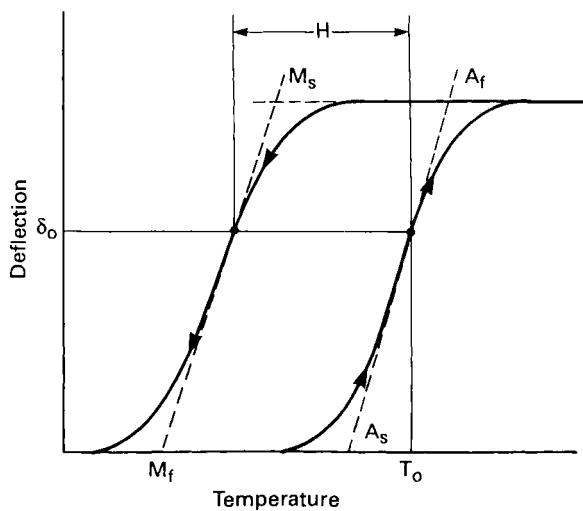


Figure 2: Typical deflection-temperature characteristics

spring expands and performs work against a load. On cooling a shape memory spring will return to its original coil bound condition but will not perform useful work.

It is well known that the force-deflection and stress-strain characteristics of shape memory are nonlinear and temperature dependent. The force-deflection characteristics of a two-way shape memory spring, measured at a series of isothermal conditions, are shown in Figure 3. The coil bound spring at temperature A_s (shown pictorially) has 2% strain in the outer fiber. During unconstrained heating the shape memory spring deflection can be predicted from the abscissa of Figure 3 for any temperature. Conversely, the force output of a constrained shape memory spring being heated can be predicted from the ordinate of Figure 3.

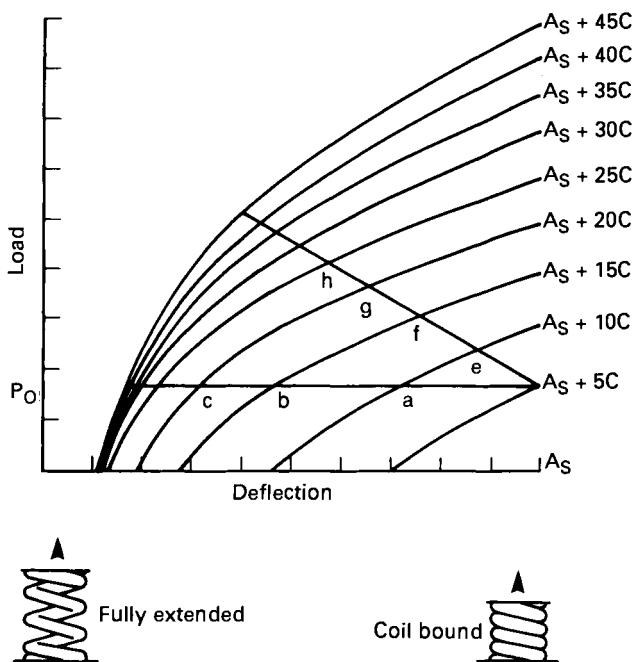


Figure 3: Load-deflection-Temperature Chart for a Shape Memory Compression Spring

2. Shape Memory Compression Springs as Actuators

When a load resists the changing length of a shape memory element, it performs work. A shape memory compression spring can be designed to operate against a dead load (P_o) or a bias spring (rate K_b). This combination of shape memory spring and working load is referred to as an actuator.

Figure 3 shows the superposition of two load characteristics of a force-deflection-temperature chart for a shape memory spring. The horizontal load line is the dead load characteristic and the diagonal load line is the bias spring characteristic.

The "start to move" temperature is defined as the temperature at which the actuator begins to move. Figure 3 shows at temperature A_s the shape memory spring is coil bound. Heating to $A_s+5^\circ\text{C}$ corresponds to the "start to move" temperature, or the temperature at which the shape memory can overcome the pre-load (P_o). The "start to move" temperature can be increased from $A_s+5^\circ\text{C}$ to $A_s+10^\circ\text{C}$ or $A_s+15^\circ\text{C}$ by increasing the initial pre-load subjected to the shape memory spring in its coil bound condition. This adjustable "start to move" property is useful for achieving actuator calibration and sensitivity control.

If the working load characteristic is known, the actuator's deflection as a function of temperature can be predicted. For a shape memory spring working against a dead load (P_o) the load-deflection curves shown in Figure 3 will be traversed through points (a), (b), (c), as temperature increases. For the case of a shape memory spring working against a biasing spring the load-deflection curves will be traversed at the biasing spring rate (K_b). Figure 3 shows the path (e), (f), (g), (h), as temperature increases.

3. Basic Control Principles

Control and calibration techniques developed over many years for other thermo-mechanical actuators are applicable to shape memory actuators. There are two calibration techniques, *variable pre-load* and *variable datum*, illustrated in Figure 4. Both techniques are used to control the "start to move" temperature of the actuator. As the name implies the variable pre-load features an adjustment for varying the pre-load (P_o) of the bias spring. The variable datum technique achieves control of the "start to move" temperature by adjusting the air gap (X) between the output displacer and the shape memory spring. The choice of control technique influences the actuator sensitivity.

Figure 5 shows deflection-temperature curves for an actuator under several different pre-loads. As the pre-load is increased the "start to move" temperature increases and the deflection of the shape memory actuator is reduced. The deflection-temperature sensitivity is also reduced as the pre-load is increased.

The deflection temperature characteristics for a variable datum actuator are shown in Figure 6. In this case the "start to move" temperature is determined by the biasing spring pre-load and adjustments of the variable air gap (X_a), (X_b), (X_c), etc. The total deflection capable by the shape memory actuator is quickly diminished as the air gap is increased. The deflection-temperature sensitivity is also reduced as the air gap increases.

A comparison of Figure 6 with Figure 5 shows the variable pre-load technique to have a higher sensitivity than the variable datum technique when calibrated for the same temperature.

The choice of calibration technique for a particular application may be dependent upon sensitivity requirements, geometric constraints, travel, and maximum allowable loads. It is sometimes beneficial for an application to combine both calibration techniques. For instance, a design might call for actuators to have a factory calibrated variable pre-load "start of move" temperature and a user controlled variable datum adjustment.

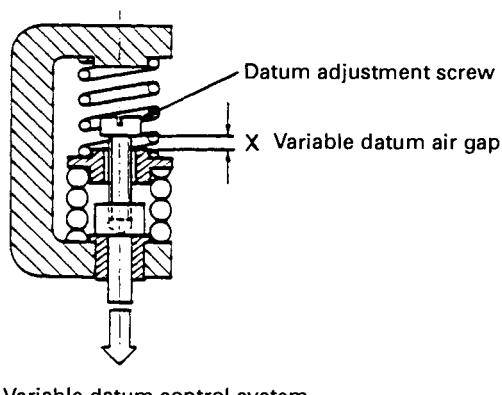
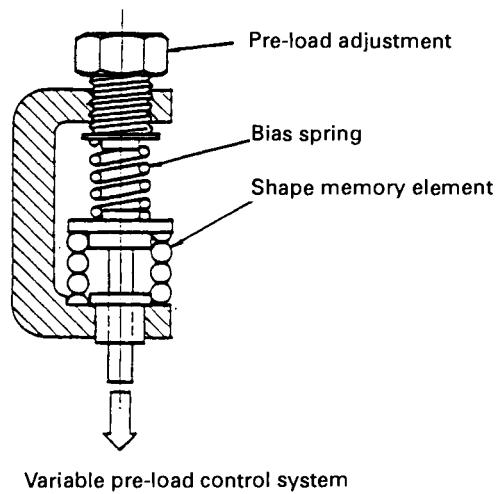


Figure 4: Two types of control systems: Variable pre-load and variable datum

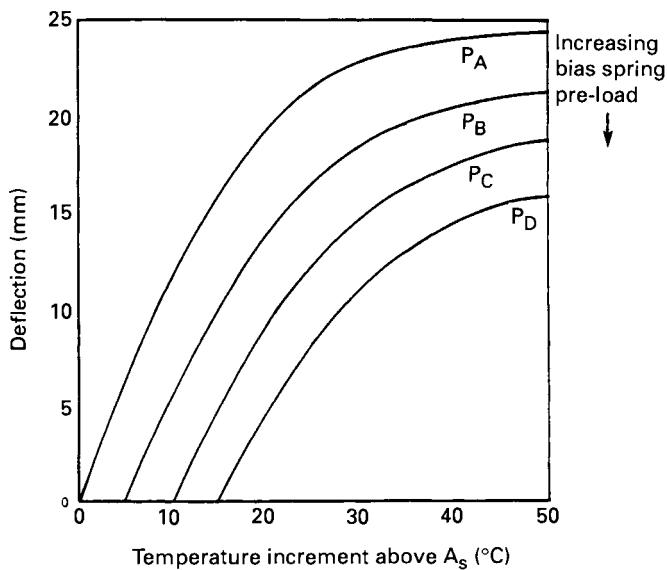


Figure 5: Variable pre-load characteristics

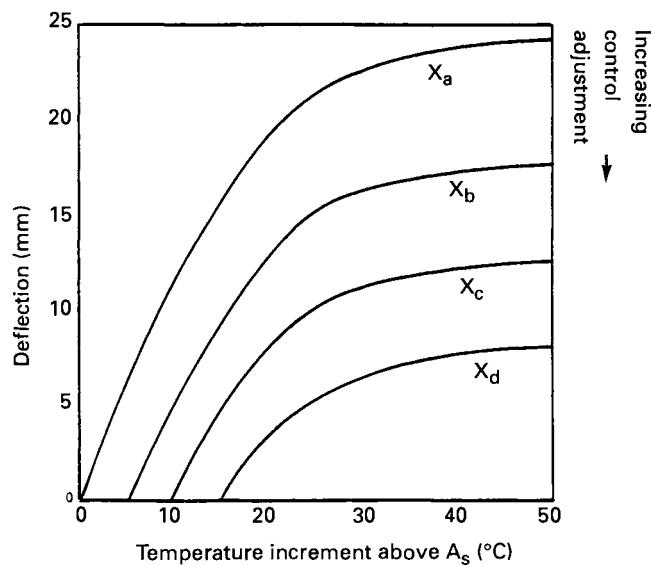


Figure 6: Variable dataum characteristics

4. Design Chart for Shape Memory Compression Spring Actuators

A design chart is a useful tool for predicting the dimensions of shape memory elements required to meet a specification. The design chart presented in Figure 7 is applicable to shape memory compression springs wound to an initial pitch that results in a strain of 2% when compressed solid. A design chart is constructed by collecting

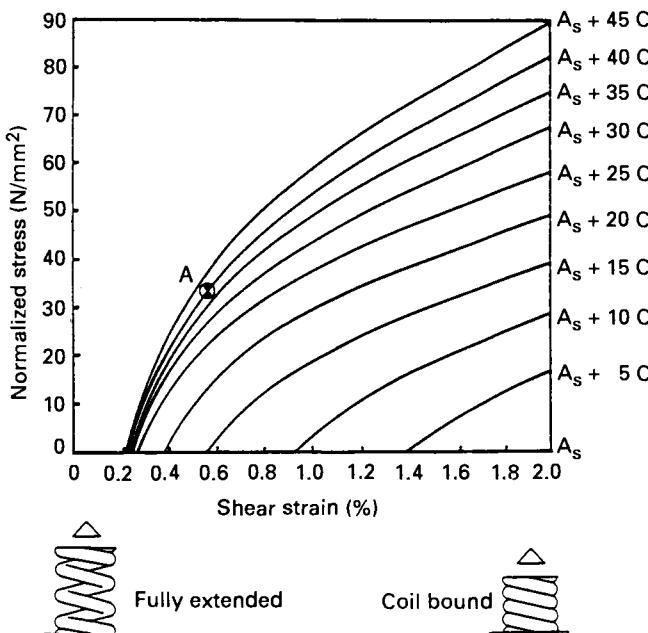


Figure 7: Design example one

experimental stress strain data for shape memory compression springs. It has been shown by Vergut³ et al that the stress strain behavior of a shape memory compression spring does not change appreciably with the number of turns and spring index. The design chart presented is applicable for shape memory springs having at least 3 active coils, a spring index greater than 4 and wire diameter greater than 1.0mm. The stress data is normalized to yield an effective scaling parameter for the design chart. The shear strain is based upon the elastic model presented by Michael and Hart¹.

The following two examples illustrate basic design methodology.

Example 1: Suppose we want to design a shape memory actuator capable of supporting a load of 50 N at a temperature 40°C above the transformation temperature. The required deflection of the spring should be 15mm from the original coil bound length. What are the spring dimensions?

Based on experience, we choose point (A), (Figure 7) on the $A_s + 40^\circ\text{C}$ isothermal as a condition that will satisfy our objective. At Point A the normalized load parameter is:

$$P \frac{D}{d^3} = 34.2 \text{ N/mm}^2 \quad (1)$$

and the shear strain is:

$$\Delta\gamma = \frac{\Delta\delta}{\pi \frac{D^2}{d} n} = 0.0058 \quad (2)$$

Using the given load of 50N we solve Equation 1 in terms of the spring index and wire diameter:

$$\frac{C}{d^2} = \frac{D}{d^3} = \frac{34.2}{50} = 0.684 \text{ mm}^{-2} \quad (3)$$

Substituting the given deflection into Equation 2 yields:

$$0.021 - 0.01 = \frac{15}{\pi C D n} \quad (4)$$

Equations 3 and 4 can now be solved in terms of the spring index (C) to give the required dimensions of the shape memory spring. A few possible results are given below:

C	d (mm)	D (mm)	n
4.0	2.49	9.94	8.5
4.5	2.64	11.88	6.3
5.0	2.78	13.90	4.8
5.5	2.91	16.00	3.8

This simplified example shows this actuator design requires the choice of an operating point (A) with specification of load and deflection at that point. The final choice of spring dimensions may be determined by economic or geometric requirements.

Example 2:

Suppose it is required to design a shape memory actuator capable of moving a load through a distance of 10mm for a temperature change of 20°C . Let the rate of the bias spring be 20 N/mm with a pre-load of 60 N. The actuator "start to move" temperature is required to be 10°C . Specify the spring dimensions and the material A_s .

Let P_a equal the initial pre-load of 60 N acting on the actuator before heating occurs. The final load acting on the actuator can be determined from the given information as:

$$P_b = 60 + 20(10) = 260 \text{ N} \quad (5)$$

Since the final and initial loads (P_a and P_b) are known a ratio of initial and final normalized stresses can be written as:

$$\frac{P_b D/d^3}{P_a D/d^3} = \frac{P_b}{P_a} = 4.33 \quad (6)$$

This is the beginning of a trial and error procedure for finding the basic spring dimensions (D), (d) and (n). If a choice is made for the initial normalized stress point (A), we immediately know the final normalized stress at point (B). From the information given we know that point (B) must lie on an isotherm 20°C above the chosen "start to move" temperature.

Clearly there are numerous possible choices for the operating load line that will satisfy the specifications. The best choice may depend upon sensitivity performance, geometric and cost requirements. A possible solution for the operating load line is shown in Figure 8.

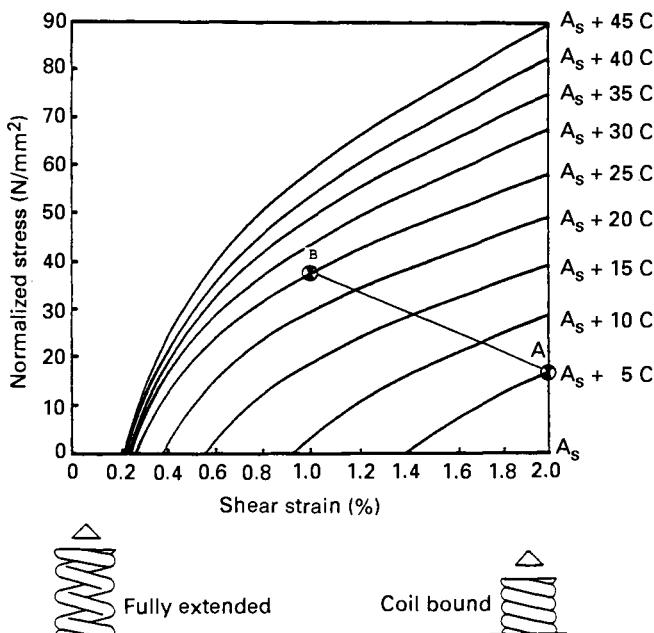


Figure 8: Design example two

The calculation proceeds to find a set of shape memory spring dimensions that satisfy the operating load line. From the design chart the initial stress is:

$$P_a \frac{D}{d^3} = 16 \text{ N/mm}^2 \quad (7)$$

and immediately we have an expression for the spring index and the wire diameter,

$$\frac{C}{d^2} = 0.267 \text{ mm}^{-2} \quad (8)$$

The operating load line corresponds to an actuator deflection of 10mm. From the design chart the change of strain when the actuator moves from point A to point B is:

$$\Delta\gamma = 0.02-0.01 = \frac{\Delta\delta}{\pi CDn} \quad (9)$$

A table of possible shape memory spring dimensions (C), (d), (D), and (n) can be obtained from Equation (8) and (9).

C	d (mm)	D (mm)	n
4.0	3.87	15.48	5.12
4.5	4.01	18.48	3.80
5.0	4.54	24.96	2.30

All the shape memory spring dimensions shown above satisfy the initial conditions given Example 2. The final spring dimensions are the designer's choice.

The material's A_s can be specified as $+5^\circ\text{C}$ since the "start to move" temperature was given as 10°C and (point A) was chosen as $A_s + 10^\circ\text{C}$.

Example 2 shows an operating load line can be established from a specified load characteristic and temperature change. Knowledge of the total deflection permits an iterative solution to be found from the normalized stress and strain equations (8) and (9). The solution yields a set of shape memory spring dimensions that satisfy the operating load line.

5. Hysteresis

Use of the design chart to develop primary shape memory spring characteristics gives no information on the hysteresis. The hysteresis originates from the material transformation and mechanical friction of the actuator. Figure 2 shows a typical performance bandwidth (H) existing between the heating and cooling curves. For a specified temperature T_o a deflection bandwidth exists and for a specified deflection δ_i a temperature bandwidth exists. The magnitude of the bandwidth can be adjusted and is important in the design process.

From a design standpoint the transformation hysteresis is a function of the cyclic temperature amplitude and the working stress. Schetky and Sims² show a 15°C hysteresis loop for an unstressed Cu-Zn-Al shape memory spring undergoing a full transformation (temperature amplitude of 40°C). Schetky and Sims² also show a 0.5°C hysteresis loop for an unstressed shape memory spring undergoing a partial transformation (temperature amplitude of 10°C). However a shape memory spring working under a relatively large stress and undergoing a full transformation (temperature amplitude of 40°C) is shown by Schetky and Sims² to have a hysteresis loop of 1.75°C .

This transformation hysteresis is unique to shape memory alloys and offers design possibilities not possible with wax, gas or bimetallic actuators. Conventional thermo-mechanical actuators have a fixed hysteresis independent of the temperature amplitude and working stress.

For a given actuator, mechanical friction affects the magnitude of the hysteresis. The friction vector of any actuator device reverses direction when cooling follows heating. This reversal of friction gives partial rise to the hysteresis bandwidth shown in Figure 2. Some increment of cooling will be required before the actuator can generate enough force to overcome the reversed frictional forces, laboratory experience has shown Cu

based shape memory actuators to have hysteresis loops of 4°C to 15°C depending upon mechanical tolerances, working stress, and temperature amplitude. Greater hysteresis precision is possible at greater cost. The only true evaluation of mechanical friction is construction of a prototype and experimental measurement of the frictional component.

6. Conclusion

The design methodology presented here for sizing shape memory elements depends heavily on two-way stress strain data. The design chart described in this paper is applicable to shape memory compression springs wound to a specified initial pitch. The design process is iterative and allows the designer to quickly determine the size of a shape memory element required. Once the actuator components and geometric requirements are known it is possible to evaluate the cost and feasibility.

The designer of shape memory actuators must balance several interdependent variables. For instance, a change in mean coil diameter will effect the free length, sensitivity, and hysteresis loop, initial winding pitch, maximum load output, and "start to move" temperature. There is no substitute for experience when trying to resolve these interrelated aspects of shape memory design. The final step, of course, to prove any shape memory actuator design is construction of a prototype.

References

1. Michael, A.D., and Hart, W.B.: The Metallurgist and Materials Technologist, 434, (1980).
2. Schetky, L.Mcd., Sims, R.B.: Proc. Int. Conf. on Martensitic Transformations, ICOMAT 1979, Cambridge, Mass., USA, 693, (1979).
3. Verguts, H., Aernoudt, E., and Vermeersch, W.: Journal De Physique C4-824, (1982).

Fatigue Properties of Ni-Ti Shape Memory Alloys

Yuichi Suzuki and Hirokazu Tamura

Yokohama R&D Laboratories, The Furukawa Electric Co. Ltd.,
2-4-3, Okano, Nishi-ku, Yokohama, 220 JAPAN

The life of a material can be one of the most important properties in product design. It essentially depends on the mechanical or thermal circumstances and independently defined according to service condition. The meaning of the term "life" is listed in Table 1 for some materials. In the case of a usual structural material such as steel, life under repetitive loading is called "fatigue life", and that under a constant load is "creep life". In the case of a Shape Memory Alloy (SMA), it is natural to evaluate the life with respect to its repetitive operation, which involves a deformation and a subsequent recovery by heating.

Table I: Various Definitions of "Life"

Material	Function	Measure of Life
Conventional Material (Structural)	Dimensional Stability	Fracture Deformation
SMA (Functional)	Shape Recovery	Loss of Recoverability
LED (Device)	Light Emission	Decay of Output Power

A usual definition of the fatigue life is "number of deformation cycles until a specimen fractures". Figure 1 shows a schematic S-N curve. The ordinate and the abscissa indicate the stress amplitude and the fatigue life respectively. The judgement of life is based on the fracture of a specimen. Since the function of a structural material is dimensional stability, a measure for the life should be a dimensional change. The fracture is most obvious and drastic loss of the function and easy to observe experimentally. Although the fatigue usually means the fracture of a material, a measure for a life should be generally based on the deterioration of a functional characteristic and not restricted to the dimensional stability. A Ni-Ti alloy, the most

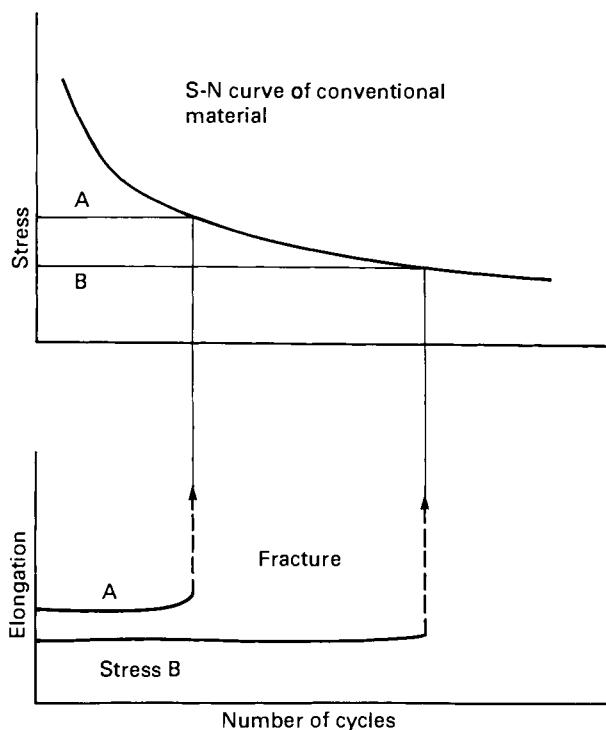


Figure 1: Relation between an S-N curve and the dimensional change in a conventional material (schematic).

popular SMA, does not fracture under normal conditions. Therefore, fracture is not a valid measure for the fatigue life of an SMA component.

The deterioration of a material's function is generally continuous. A continuous change in the shape memory characteristics of an SMA component takes place with operational cycles. A measure for the life should thus be shape recoverability. There are two parameters, recovery stress and recovery strain, which quantify the shape recoverability, however the former seems to be the more practical because the performance of an SMA component is usually specified by the recovery stress at a fixed position. Figure 2 shows a typical two-way SMA unit: positions of the SMA and the bias springs at high and low temperatures are determined at the first stage of the design. Because the change in the recovery stress of the SMA spring at such positions is to be evaluated as a fatigue property, the recovery stress becomes a more practical parameter.

Figure 3 shows schematic S-N and deterioration curves of an SMA component. In order to construct the S-N curve, we must first define N as, for example, the number of cycles at which the recovery force becomes 70% of the initial value. However, the S-N curve is less informative because each "point" thereon corresponds to one deterioration "curve". Since thermal cycling tests for an SMA component are time consuming, and the change in the performance (recovery stress) can be continuously

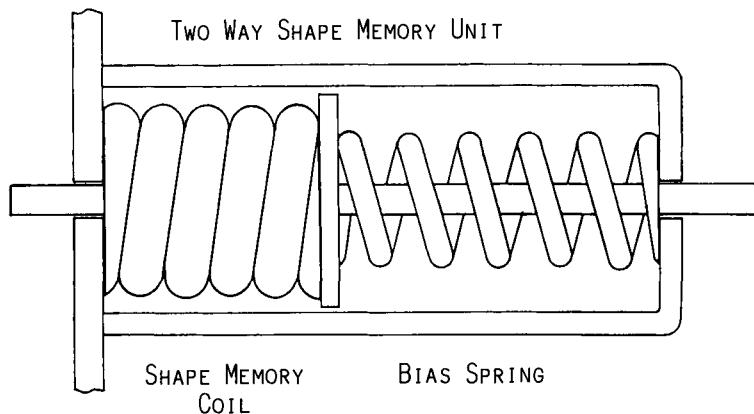


Figure 2: Typical two-way unit composed of SMA and bias springs.

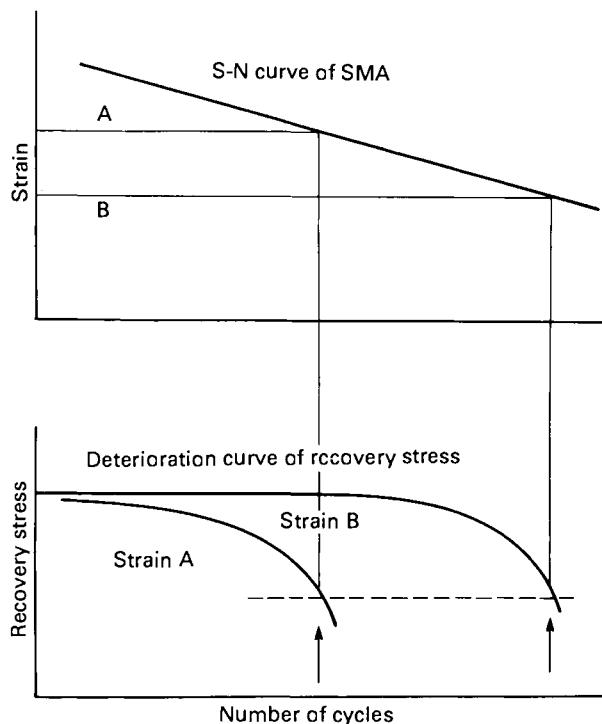


Figure 3: Schematic S-N and deterioration curves of an SMA component.

evaluated with the number of cycles, we should refer to the deterioration curve to investigate the fatigue process of the component. The deterioration curve is similar to a creep curve, where the abscissa indicates a duration time, it should be noted that the recovery stress decreases definitely only after high operation cycles.

1 Fatigue Properties of Ni-Ti and Ni-Ti-Cu Alloys

Table 2 shows the relation between the crystal structure of a low temperature phase and the fatigue properties of Ni-Ti and Ni-Ti-Cu alloys. The high temperature phases in both have the same B2 structure. Although recoverable strain is largest for the monoclinic martensite in the binary alloy, the fatigue properties are not as good. On the other hand, the R-phase provides an excellent fatigue property with smaller recoverable strain. The orthorhombic martensite in the ternary alloy is positioned between the two. Among the above three cases, fatigue tests for the R-phase transformation are most time-consuming. Since the deterioration of an SMA component appears only in a high cycle region, the estimation of life from low-cycle data is difficult. The issue is then to develop a method to estimate the fatigue life of long life components

Table 2: Summary of the Three Low Temperature Phases in the Ni-Ti System

Structure of Low Temperature Phase	Recoverable Strain	Fatigue Life
Monoclinic (Binary Ni-Ti)	High	Low
Orthorhombic (Ni-Ti-Cu)	Medium	Medium
Rhombohedral (Work Hardened)	Low	High

We analysed the fatigue properties of Ni-Ti shape memory coil springs carefully, and have found a method to estimate the life from the initial spring properties. We can now design and fabricate a long life coil component based on this knowledge. Such a component exhibited an excellent fatigue lifetimes beyond one million cycles.

2 Fatigue Properties of Ni-Ti Coil Springs

Figure 4 shows a thermal cycling machine. SMA coil components are immersed alternately in hot and cold baths. Three pairs of the baths were installed, where the temperature can be set independently. Thermal cycling was conducted with the components constrained at a constant strain, or combined with a bias spring to compose a two-way unit. Figure 5 shows a test cell. The shape memory properties of the component can be measured without detaching it from the unit. The bias spring is compressed during a measurement so as not to affect the SMA's property. Eliminating the possibility of an unexpected deformation of the SMA component. The units are

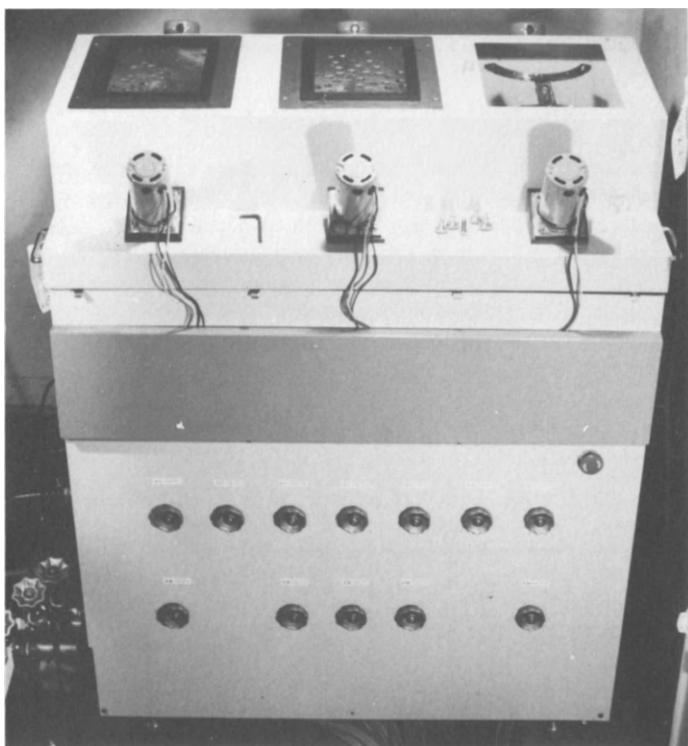


Figure 4: Thermal cycling machine.

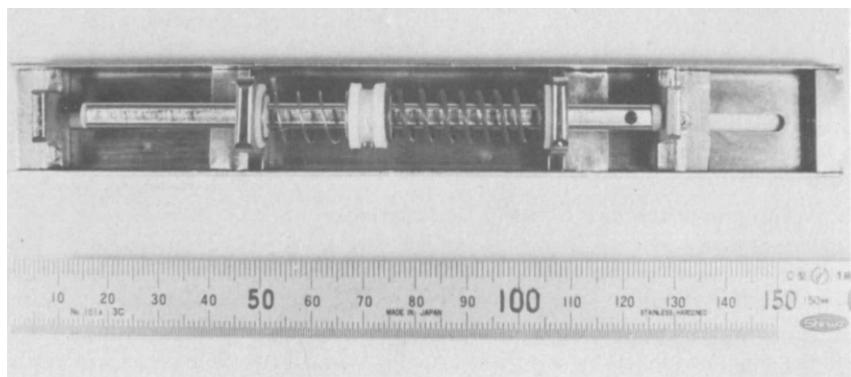


Figure 5: Setting unit for a thermal cycling test.

detached from the thermal cycling machine at predetermined numbers of cycles in order to measure the temperature-stress ($T-\tau$) curve of the SMA component.

Figure 6 shows equipment for the measurement of the shape memory properties: both the stress change under a constant deflection and the deflection change under a constant stress can be measured. All the procedures are sequentially programmed and data are filed in a floppy disk, and simultaneously recorded on a pen recorder.

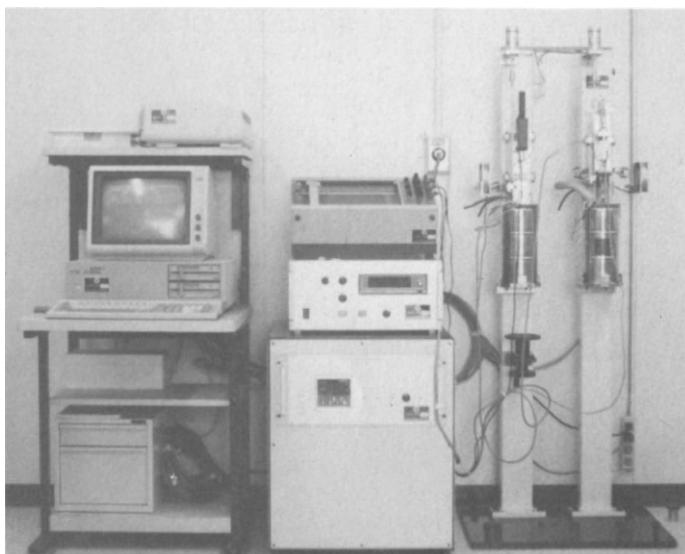


Figure 6: Equipment for the measurement of the shape memory properties.

Figure 7 shows an example of the hard copy of filed data ($T-P$ curve).

The change in the $T-\tau$ curves is compared in Figure 8, where the upper curves corresponds to the case where only the R-phase transformation occurred during the cycling and the lower to the case where the martensitic transformation did so in addition to the former. The change in the temperature strain ($T-\gamma$) curves of the two-way unit is shown in Figure 9 for the above two cases. The martensitic transformation gives rise to a large hysteresis in the $T-\gamma$ curve. When the martensitic transformation took place in the cycling, more characteristic change occurs after 10^4 cycles. Thus, it is important to utilize the R-phase transformation alone in order to achieve the excellent fatigue lifetimes. The change in the recovery force with the number of cycles is shown in Figure 10 as a deterioration curve with the force normalized to the initial value. It should be noted that the recovery force in Test II decreases steeply at the high cycle region, characteristic of the case where both the transformations occur.

Minimum ambient temperature in service is important in that the fatigue property of an SMA component is significantly affected by whether the martensitic transformation occurs or not. The composition and heat treatment of any alloy should be determined so that M_s is above the minimum expected ambient temperature. However, the effects

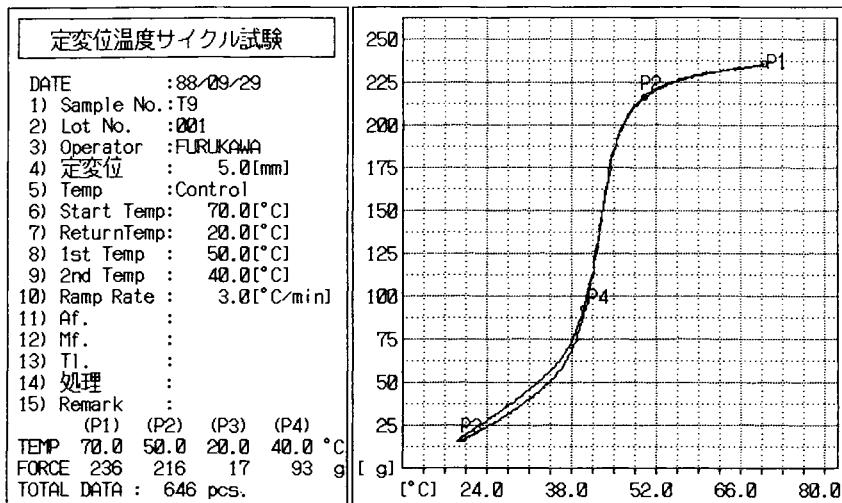
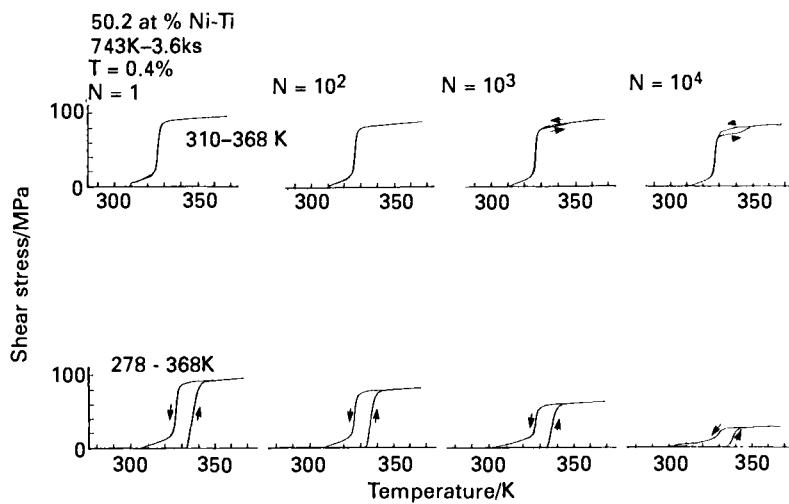


Figure 7: Example of filed data (T-P curve).

Figure 8: Comparison of the change in the T- τ curve.

Upper: The R-phase transformation occurred in a cycling.
Lower: Both the R-phase and the martensitic transformations occurred in a cycling.

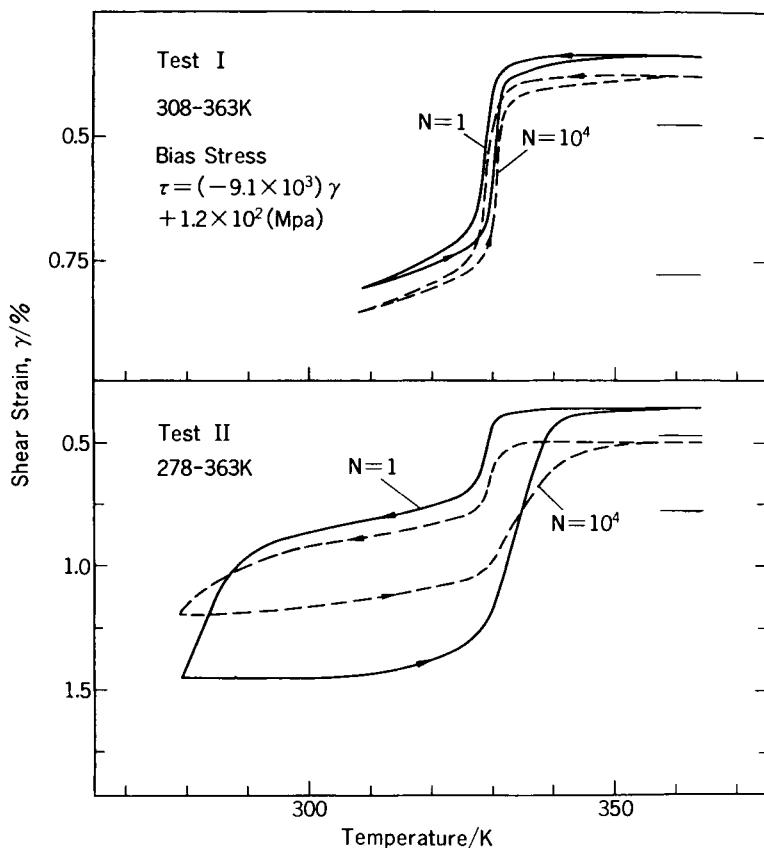


Figure 9: Comparison of the change in the T- γ curve of the two-way unit.
 Upper: The R-phase transformation occurred in a cycling.
 Lower: Both the R-phase and the Martensitic transformations occurred in a cycling.

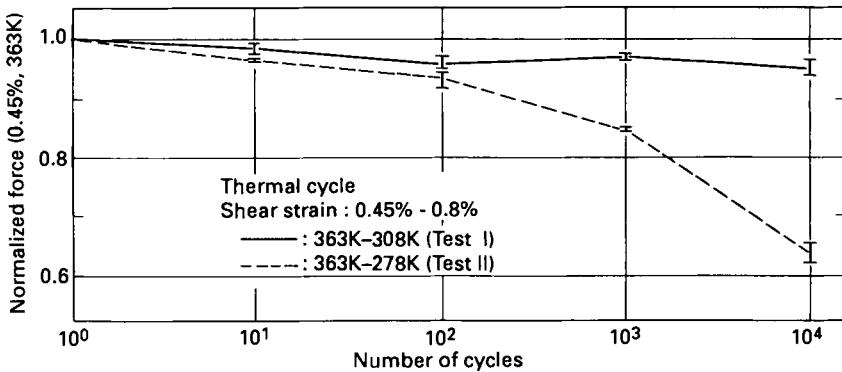


Figure 10: Change in the recovery force as a function of the number of cycles.
 Test I: The R-phase transformation occurred in a cycling.
 Test II: Both the R-phase and the martensitic transformations occurred in a cycling.

of an applied stress or stain must also be taken into account. If it is too large, the martensitic transformation is induced. Figure 11 shows the change in recovery force as a function of shear strain. The minimum temperature during cycling was set to be above M_s under no stress. The recovery force decreases stepwise at 0.8% strain. The effect of the strain becomes more pronounced as the number of cycles increases.

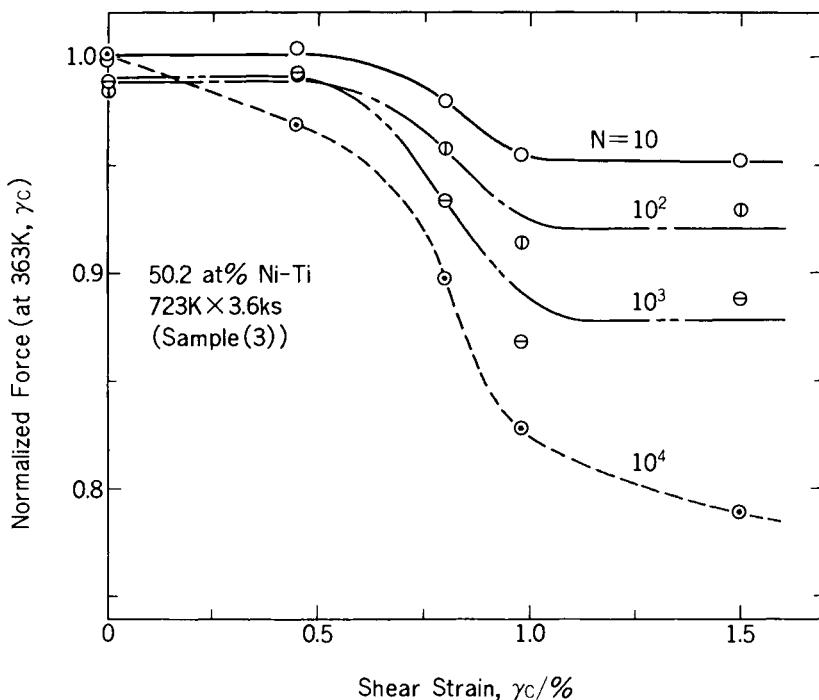


Figure 11: change in the recovery force as a function of the shear strain: (N:number of cycles).

Figures 12-(1) and (2) show similar curves with the initial T- τ curves under each strain. It is important to note that the fatigue property of an SMA component can be evaluated from the initial T- τ curve. We can then easily find the strain conditions necessary to achieve excellent fatigue lifetimes.

The occurrence of the hysteresis depends on the amount of the strain, alloy composition and heat treatment condition. Optimal combinations of these factors should be studied and accumulated as a design database. A long life SMA coil spring was designed on the basis of these concepts. Figure 13 shows the fatigue properties of the component. The temperature range of operation was also taken as a design parameter in addition to the above three. The recovery force after 10^4 cycles does not decrease, and the properties are excellent. Thermal cycling was continued up to one million cycles, and no degradation was found.

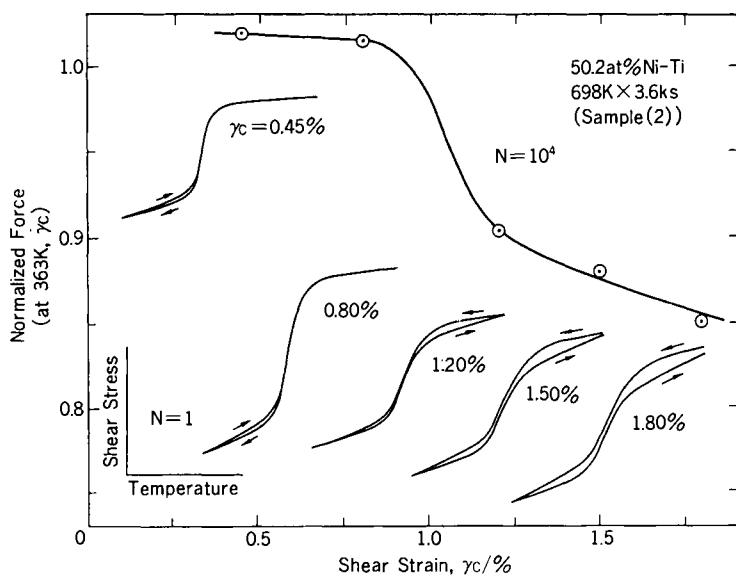
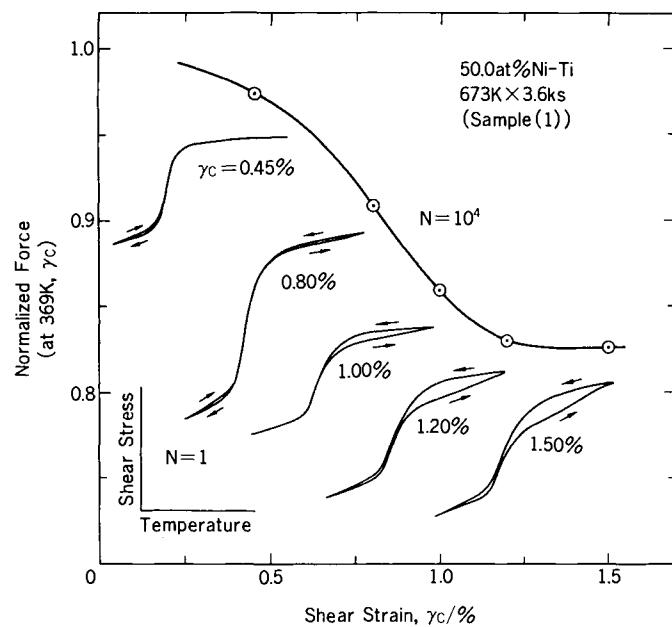


Figure 12:-(1), (2) Relation between the initial T- τ curves and the change in the recovery force. The former curves are drawn in an arbitrary unit.

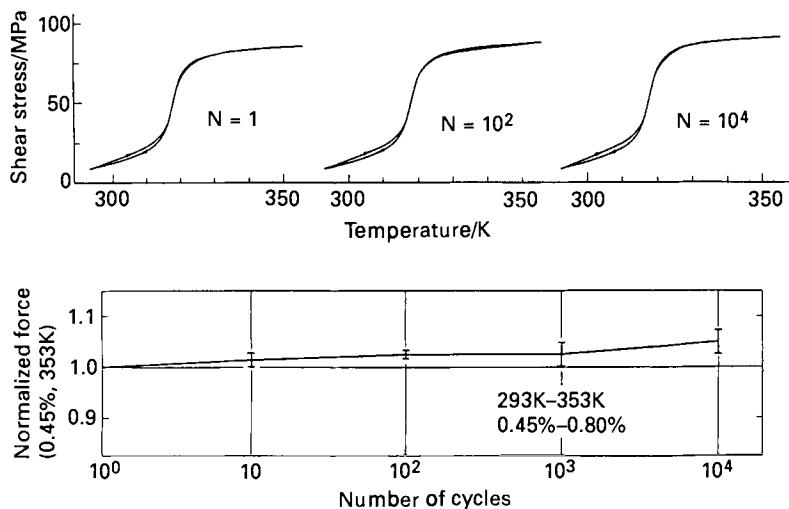


Figure 13: Fatigue properties of an SMA coil spring which was designed based on the present design concept. Note that no degradation takes place after 10⁴ cycles.

3 Conclusion

The fatigue life of Ni-Ti SMA components depends on the type of transformation. The best properties are obtained with the R-phase transformation. The most important point in achieving these excellent properties is to suppress the martensitic transformation completely. With regard to coil springs, the appearance of a narrow hysteresis in the initial T- τ curve becomes a measure of the fatigue lifetime. This fact provides a strain condition in a coil design. The suppression of the hysteresis brings about excellent fatigue lifetimes.

Fatigue of Copper-Based Shape Memory Alloys

E. Hornbogen

Institute of Materials,
Ruhr-University
Bochum, D-4630 Bochum

Fatigue implies the response of materials to cyclic loading. The sequence of events starts with accumulation of defects such as dislocation groupings, persistent bands, and surface phenomena like slip steps, extrusions, and intrusions. This stage is followed by formation of cracks and their slow and rapid growth.¹ Shape memory alloys show additional fatigue phenomena, which are similar in Ni-Ti and Cu-Zn based alloys with their bcc ordered high temperature phases^{2,3,4}, but somewhat different in Fe-Mn-Si alloys with the reversible fcc-hcp transformation.⁵ The discussion in this paper is concerned with recent results on copper based alloys.⁶⁻⁹ (some results on iron based alloys can be found in refs. 5 and 10.)

Fatigue loading in these shape memory alloys may be:

- mechanical (PE),
- thermal (2W),
- thermo-mechanical, alternating (1W),

or modes intermediate between these three cases, depending on whether the alloys are investigated in the pseudoelastic (PE), two-way (2W) or one-way (1W) condition (Figure 1).

As has been already covered in some detail, several temperature ranges exist in such alloys with different response to an external load:

$T > M_d$:	stable high temperature phase β
$M_d > T > M_s$:	range of premartensitic phenomena and phases, and stress-assisted martensite
$M_s > T > M_f$:	range of thermal transformation
$M_f > T$:	completely transformed martensite (α_m), capable of reorientation of martensite variants (α^+ , α^-) and reversible twinning.

At low temperatures the α_m martensite must be more stable than β , but it is usually not the most stable phase. It may have inherited, for example, the order from β , which in turn can be reduced by plastic deformation of α_m during mechanical fatigue and aging.

For the sake of brevity β will be used to designate the austenite and α_m the martensite, ignoring for the various crystal structures of martensite such as 2H, 9R, 18R, etc. discussed in earlier chapters.

If an alloy is exposed to cyclic mechanical loading in these four temperature ranges a qualitatively different response can be expected in each case. The effects should

include mechanical fatigue, implying accumulation of defects, formation of cracks and their stable and unstable growth (Figure 2). The structural changes induced by mechanical fatigue will modify the transformation behavior, i.e. the transformation temperatures M_s , M_f , A_s , A_f , and the hysteresis loops of the transformation (Figure 1). Thermal fatigue can be caused by cycling through the complete transformation range between $T > A_s$ and $T < M_s$.⁶ Consequently, three variables have to be considered for fatigue experiments with SMA: temperature, external stress, and strain.

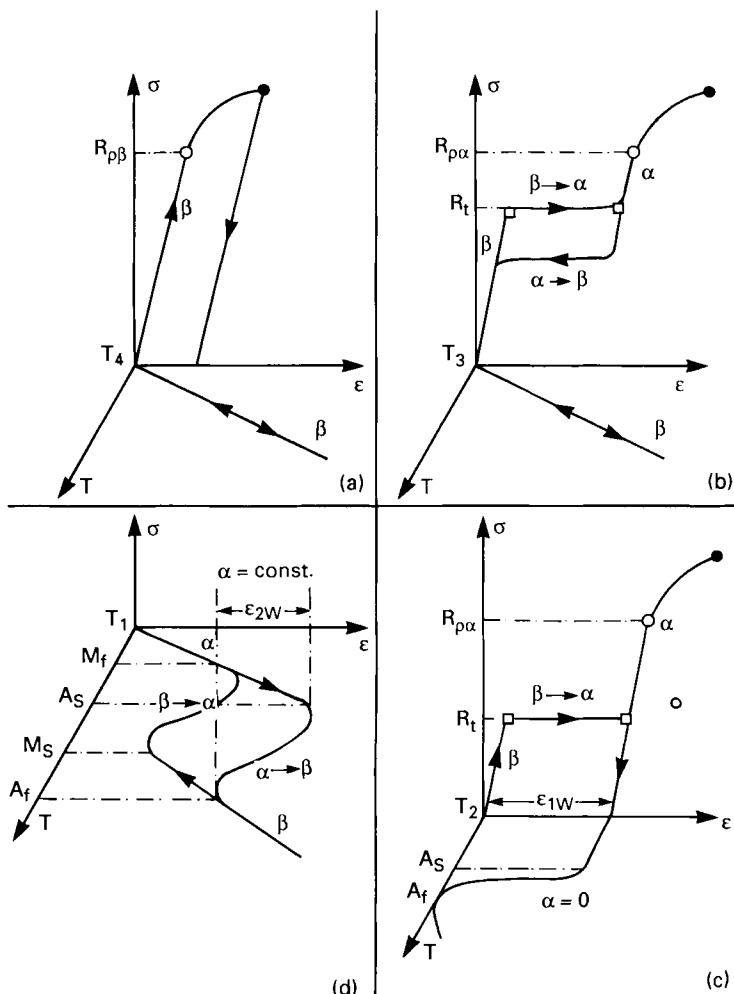


Figure 1: Different types of thermo-mechanical behavior of shape memory alloys shown in stress-strain-temperature space - generalized fatigue implies repeated cycling of any of these variables. Specifically shown are (a) normal behavior, (b) pseudoelasticity ($\sigma \neq \text{const.}$), (c) two way-effect ($T = \text{const.}$), and (d) one way-effect (alternating variables σ and T)

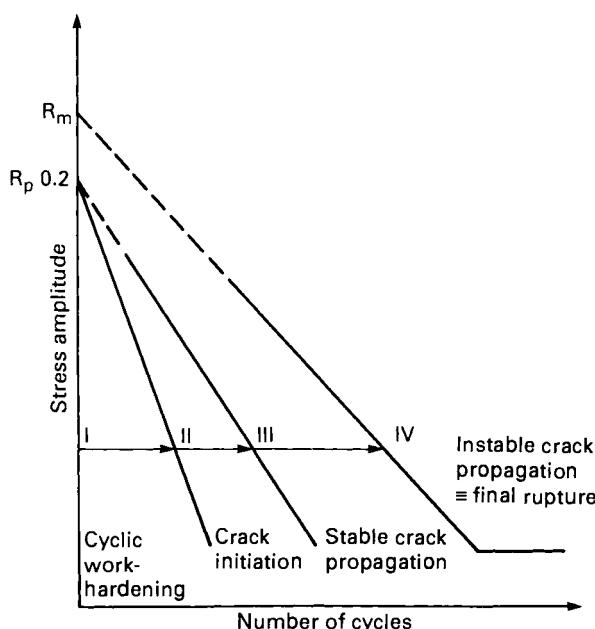


Figure 2: Stages of development of mechanical failure in a schematic σ -N-curve (see Figures 4-7). The new features are defects caused by moving transformation interfaces (β - α) and failure favored by defects in martensite such as microcracks at variant boundaries (α^+ - α^-).

Grain boundaries in bcc ordered shape memory alloys provide sites of severe incompatibility because of the high elastic anisotropy. They are often extremely brittle, due, for example, to the segregation of Zn in the intermetallic compounds based on Cu-Zn.⁷ In the martensitic state, a large number of additional crystal boundaries of different types originate (including variant and twin boundaries). With respect to cracking, they can act in a similar way as the original grain boundaries in the austenitic high temperature phase, but they form a much larger number of very small cracks. Former austenite boundaries are not preferred sites for embrittlement in martensitic microstructures.

1. Experimental Results

Chemical compositions and transformation temperatures of the investigated alloys are listed in Table 1. Details on microscopic observations can be found in several recent papers.⁶⁻⁹ These studies have been used for the interpretation of the results. Figure 3 summarizes the effects of thermal cycling at low temperatures through the whole transformation range on hardness and the stress-strain behavior of the stable β -phase. The alloy shows considerable work hardening and a tendency towards pseudoelasticity with an increasing number of cycles. The σ -N-fatigue life curve shows an increased number of cycles to failure if martensitic transformation is induced during mechanical cycling (Figure 4). A similar situation exists for alloy 2 if the load is applied at ambient temperature (Figure 5). Alloy 3 is in the completely transformed

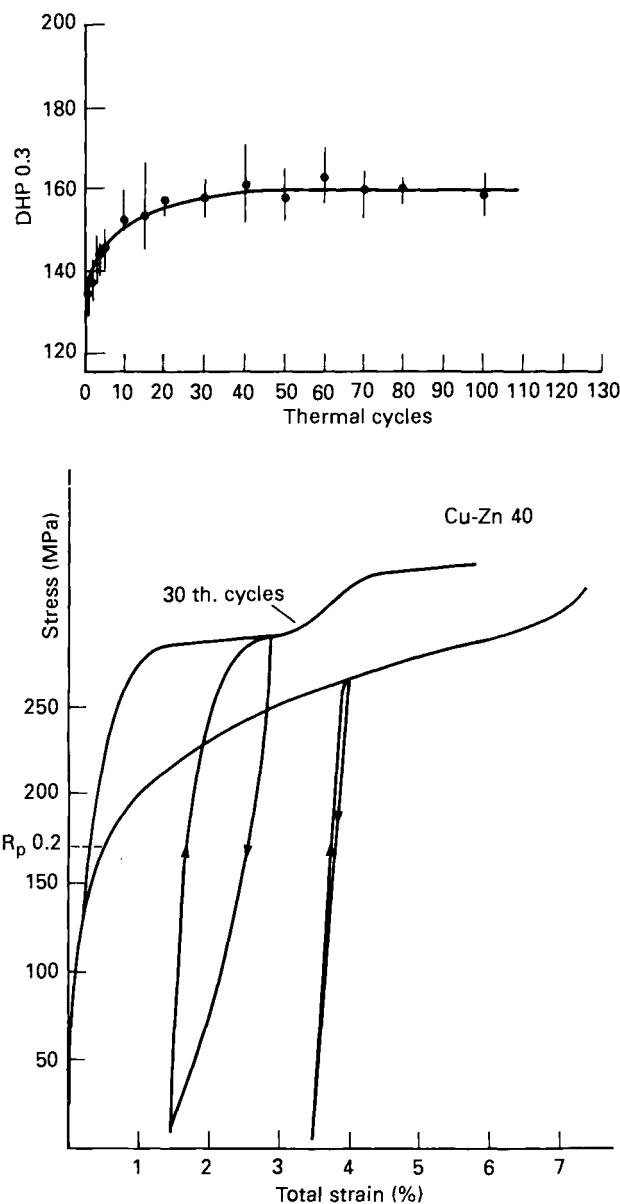


Figure 3: Effects of thermal cycling on properties of β -Cu-Zn, alloy 1. (a) hardening and (b) increase in yield stress and pseudoelastic behavior.

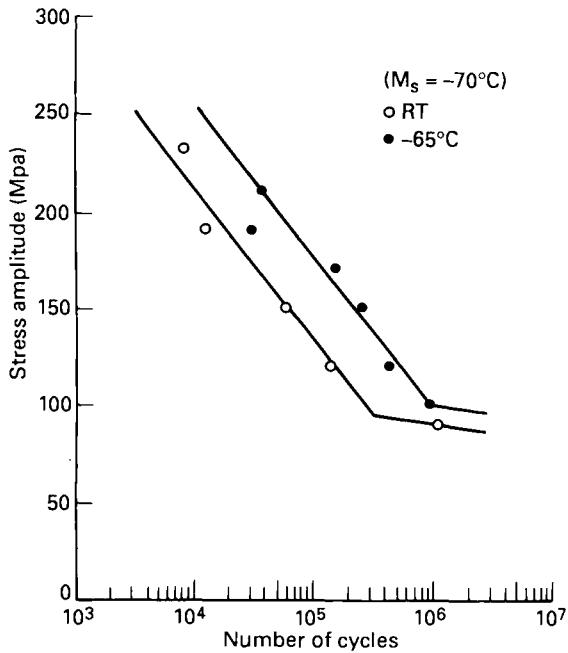


Figure 4: Fatigue life of alloy 1 far above and at the -70°C M_s .

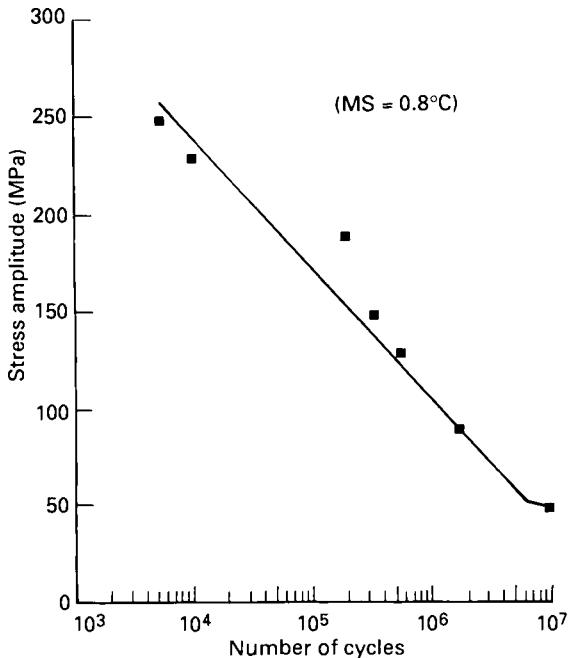


Figure 5: Fatigue of alloy 2 at 20°C corresponds to alloy 1 at -70°C (stress-induced transformation).

state if fatigued at this temperature (Figure 6). The life of alloy 1 is improved, if it had been previously subjected to a thermal cycling treatment (Figure 7).

Table I: Chemical Compositions in wt.% and M_s Values of Selected Cu-Zn-Al Alloys

Alloy	Cu	Zn	Al	M_s
1	60.48	39.51	-	-70
2	69.75	26.32	3.92	+8
3	70.50	25.72	3.54	+120
4	72.42	24.26	3.32	+10

An analysis of the fatigue crack growth functions should contribute to a better understanding of the results from the σ -N-fatigue curves (Figure 8). Fatigue crack growth is rapid for the stable β -phase in the as-betazitized state (alloy 1), much slower in the martensitic alloy 3 and in alloy 1, which had been subjected to a hot-rolling treatment during betatizing to produce rugged grain boundaries (Figure 8).⁶

In addition to work hardening and fracture, there is an effect of fatigue on the transformation behavior of shape memory alloys. Figure 9 shows how the transformation temperatures were defined from measurements of electrical conductivity. A comparison of figures 10 and 11 indicates that thermal cycling has a much stronger effect on the transformation behavior than mechanical cycling (alloy 2). There is a tendency of mechanical cycling for lowering the transformation

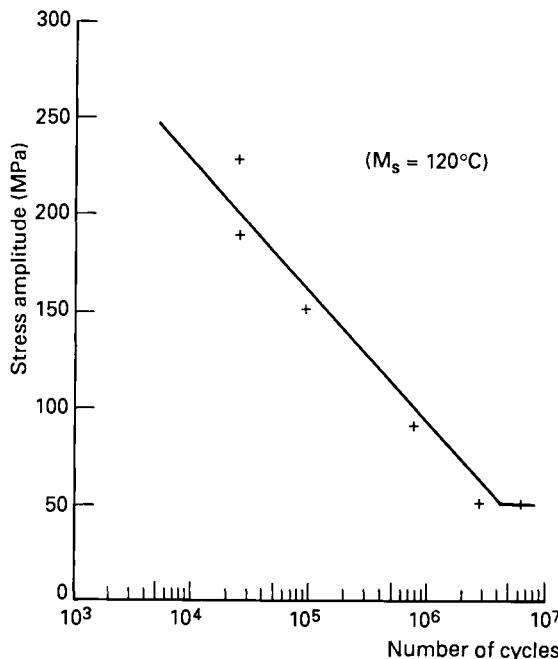


Figure 6: Fatigue of the completely transformed Alloy 3 at 20°C.

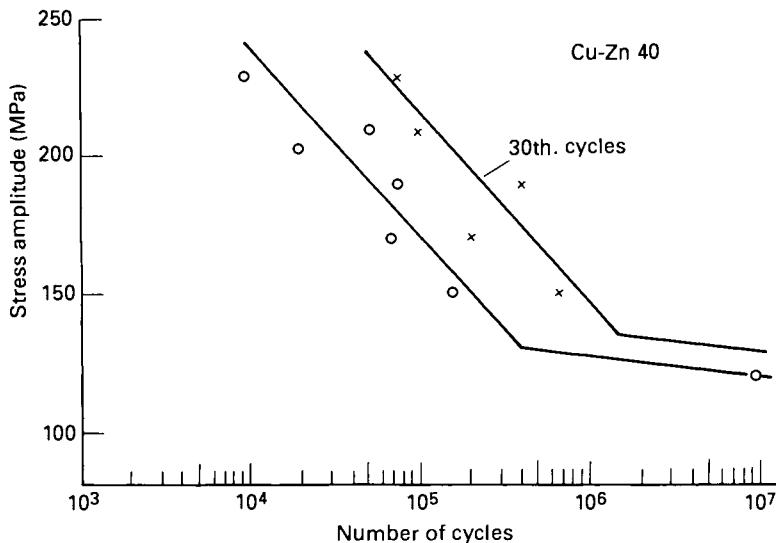


Figure 7: Effect of previous thermal cycling on the fatigue life of alloy 1, at 20°C.

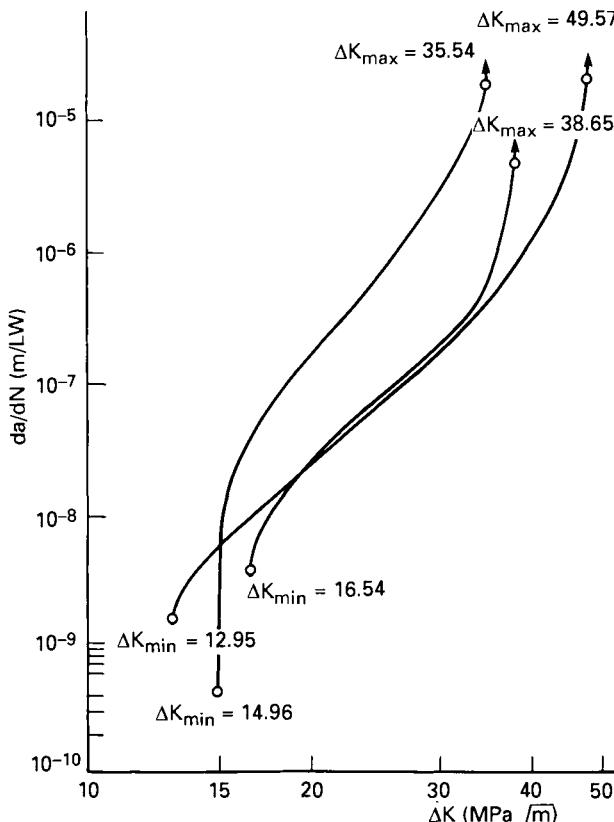


Figure 8: Fatigue crack growth in the stable β phase as betatized, hot-rolled and betatized, and martensitic.

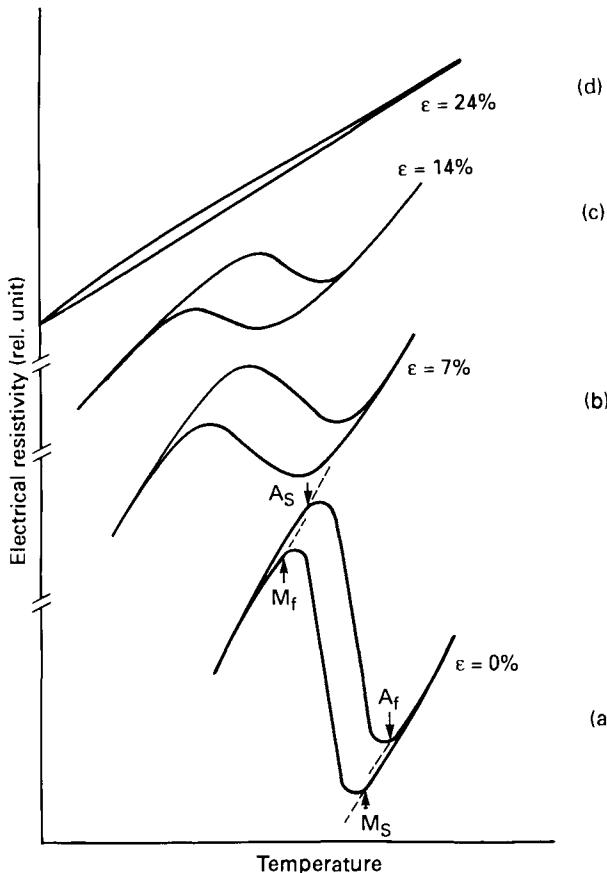


Figure 9: Definition of transformation temperatures from a typical resistivity vs. temperature curve, and the effect of plastic strain ϵ in the β -phase on the transformation.

temperatures. Thermal cycling has a similar effect on A_s and A_f . The effect is of different sign for M_s and M_f . Consequently the hysteresis is reduced by thermal cycling while it is slightly increased by mechanical loading.

Finally the effect of cycling on the pseudoelastic loop is of interest. Most remarkable is an increase in slope of the pseudoelastic part of the σ - ϵ -curve. A tendency to decrease the area of the loop can also be clearly seen (Figure 12).

Some microscopic observations in context with fatigue of shape memory alloys have been summarized in (Figure 13). Cracks nucleate and grow at grain boundaries in the homogeneous β -phase, unless a special thermo-mechanical betatizing treatment has removed intercrystalline brittleness. Intra-austenitic crack nucleation starts at inclusions which have earlier formed in martensite and during reversion to austenite. Rapid multiple crack formations take place at variant boundaries as well as inside

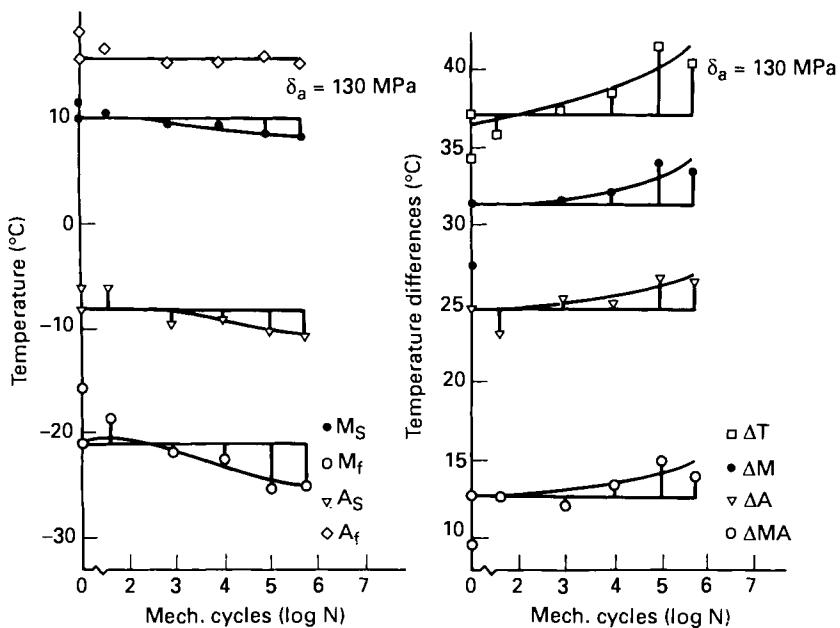


Figure 10: Mechanical cycling lowers transformation temperatures and increases hysteresis (alloy 2, $\Delta T = A_f - M_f$, $\Delta M = M_s - M_s$, $\Delta A = A_f - A_s$).

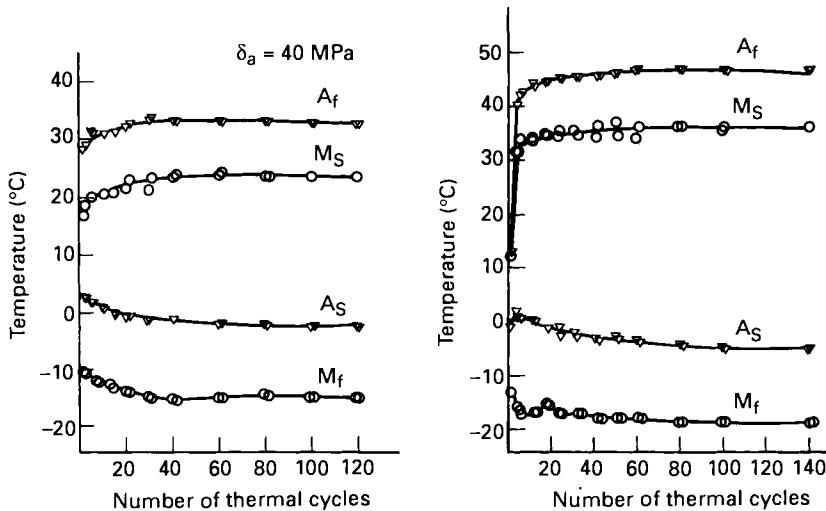


Figure 11: Thermal cycling raises temperatures of martensite formation and lowers that for the start of reversion (alloy 4).

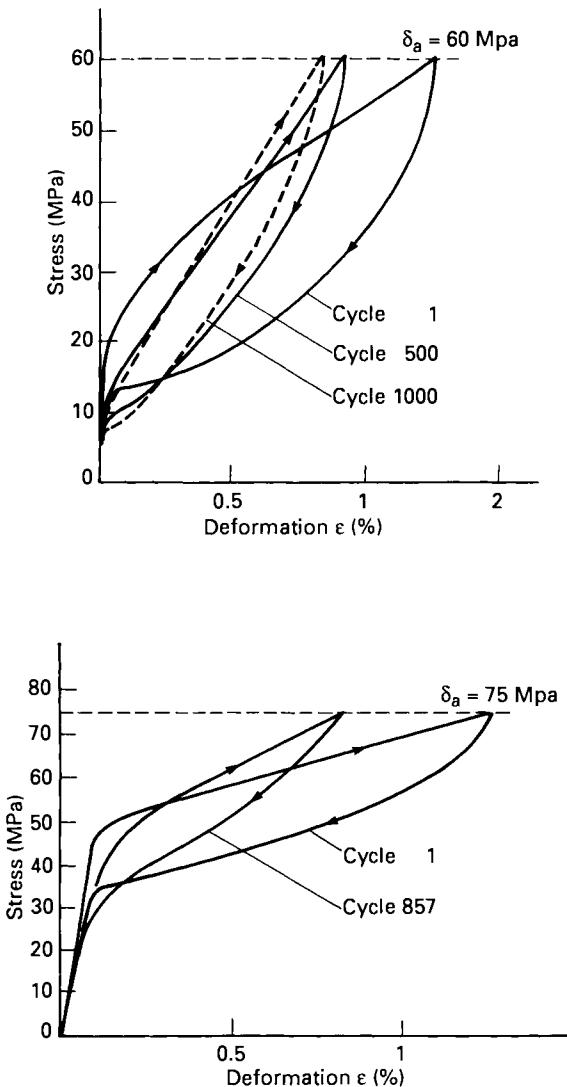


Figure 12: Effect of mechanical cycles on the shape of the pseudoelastic hysteresis loops (Alloy 4).

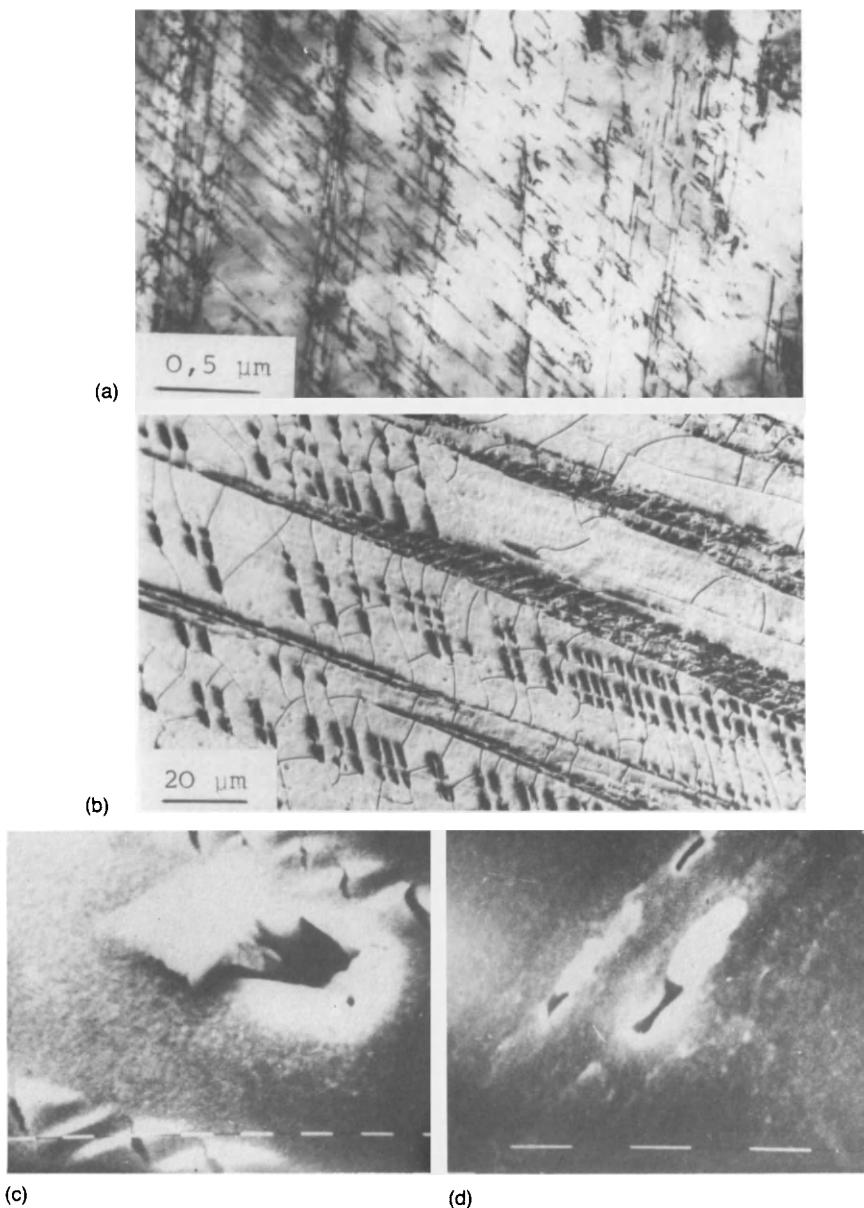


Figure 13: Examples for microstructural changes during fatigue of Cu-base SMA.

- Dislocations in deformed β -phase
- Residual martensite (dark) and subboundary-like structure left after fatigue deformation of martensite and reversion to austenite
- Square-shaped intrusions plus tongue shaped extrusions formed during mechanical fatigue in the pseudoelastic state.
- Crack initiation at $\beta\beta$ -grain boundaries

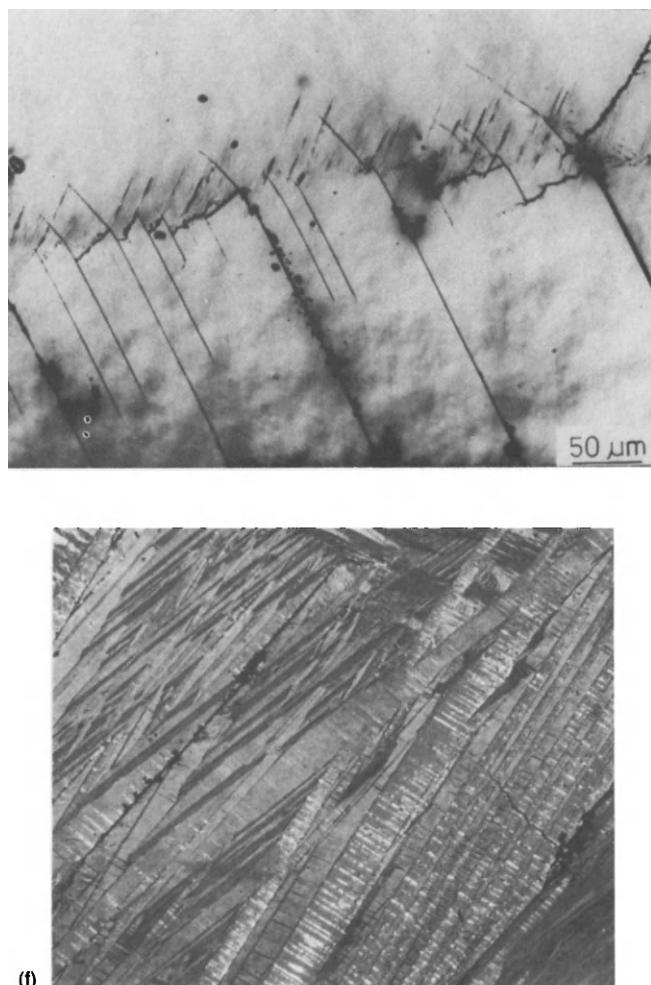


Figure 13 (cont'd) (e) slip localized on β - α grain boundaries, and (f) defects within martensite plates.

fatigued martensite. Consequent growth is slow due to energy dissipations in a highly branched crack system.

Pseudoelastic cycling through the transformation ranges produces surface steps and a particular type of tongue shaped extrusion/intrusion phenomenon. The square shaped holes which form in this connection seem to be not very effective in inducing the nuclei for the growing cracks and final fracture.

2. Summary and Conclusions

The fatigue life of copper-base shape memory alloys does not differ more than one order of magnitude regardless of the particular loading conditions. This is due to micromechanisms: crack nucleation at $\beta-\beta$ grain boundaries or $\alpha-\alpha$ variant boundaries and growth preceded by localized strain in the β -phase, and multiple crack growth in martensite. If the $\beta-\beta$ -boundaries are embrittled, the performance of the alloys is also deteriorated for all fatigue conditions. A combined hot-rolling and betatizing treatment is suitable to remove brittleness, retard intercrystalline crack nucleation and enforce a transcrystalline fatigue crack path (Figure 13).

The course of the transformation is changed by mechanical and thermal fatigue. Transformation temperatures vary up to 30°C. These changes can be explained by the following microstructures features.

- (1) Dislocation groupings as sources of internal stresses will raise M_s and induce two-way behavior.
- (2) Dislocation forests (more evenly spaced interfacial dislocations) will lower M_s and cause work hardening.
- (3) A decrease in degree of order by motion partial dislocation will stabilize martensite and raise A_s .
- (4) Residual martensite in β , will raise M_s and induce two-way effect by oriented nucleation.
- (5) Slip localized in $\beta-\alpha$ -boundaries, $\alpha^--\alpha^+$ -martensite variant boundaries as well as intra-martensitic slip in stacking form faults and twin boundaries cause microcracks.
- (6) Defects which form inside martensitic crystals and in the variant boundaries may transform during the $\alpha_m \rightarrow \beta$ reverse transformation and retard the reverse transformation and retard the reverse transformation (increase A_s).

These structural changes also cause work hardening, initiation and growth of cracks, modify the temperatures and course of the transformation and they are also responsible for the two-way effect.

Acknowledgement

The support for this work from the German Ministry of Research and Technology (BMFT Project 03M 5006) is gratefully acknowledged.

References

1. E. Hornbogen, Microstructure and Mechanisms of Fracture, Proc. ICSMA 6, Melbourne, 1059 (1982).

2. K. Shimizu, J. Electron Microscopy, **34**, 277 (1985).
3. K.N. Melton, Scripta Met., **13**, 73 (1979).
4. J. Perkins, W. E. Muesing, Met. Trans. A, **14A**, 33 (1983); Met. Trans., **15A**, 313 (1984).
5. M. Sade, K. Halter, E. Hornbogen, J. Mat. Sci. Lett., **23**, (1988).
6. M. Thumann, E. Hornbogen, Z. Metallkde., **79**, 119 (1988).
7. M. Sade, E. Hornbogen, Z. Metallkde., **79**, (1988).
8. M. Sade, J. Kumpfert, E. Hornbogen, Z. Metallkde., **79**, (1988).
9. M. Sade, E. Hornbogen, J. Mat. Sci. Lett., **23**, (1988).
10. M. Sade, K. Halter, E. Hornbogen, Z. Metallkde., **79**, (1988).

Shape Memory Actuators for Automotive Applications

Dieter Stoeckel
Raychem Corporation
300 Constitution Drive
Menlo Park, CA 94025

Actuators are devices which perform a task, like moving an object, either on demand or in response to certain changes in their environment (temperature, pressure etc). In a modern car more than 100 actuators are used to control engine, transmission and suspension performance, to improve safety and reliability and enhance driver comfort. Most of these actuators today are electric motors, solenoids, thermobimetals, wax motors, vacuum or pressure actuators.

Shape memory actuators have not yet penetrated this growing market significantly. This is due in part to low market awareness about shape memory technology, as well as little engineering data for the alloys. However, as the scientific community comes to understand the metallurgy and discuss it more openly, shape memory actuators are becoming increasingly popular for automotive applications. Japan, which leads in this area, has an impressive number of patent applications. About 100 patents for automotive applications are filed annually in Japan, outnumbering American and European applications by an order of magnitude. Japan not only leads in the development of new applications, but also in their commercialization. Shape memory actuators are also produced commercially in Europe, but no base businesses are known in the United States. This is even more astonishing, as Ni-Ti alloys have been developed in the United States and used for over 20 years (mainly in the aerospace industry).

1. Design Principles:

There are only two commercially available groups of shape memory alloys for actuator applications today, the Cu-Zn-Al alloys and the Ni-Ti alloys. For automotive applications, Ni-Ti is preferred because of a number of advantages like high strength, high electrical resistivity, large recovery strains, easy workability, and excellent corrosion resistance. Therefore, in the following, we will focus on Ni-Ti alloys only. The design of shape memory actuators is generally based on the different stress-strain curves of the material in its austenitic and martensitic condition. Although two-way shape memory actuators are available, most applications use the one-way effect with an external reset force. Figure 1 shows schematic load/deflection diagrams for shape memory springs working against a constant force (load) and against a bias spring. Work against a biasing or reset spring is the most common case in automotive actuator applications.

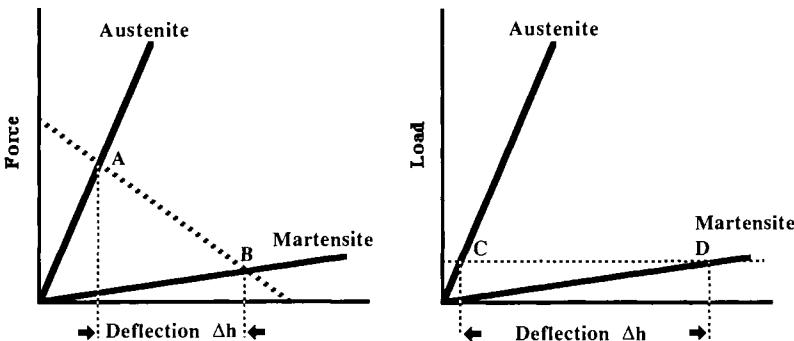


Figure 1: Design principle for shape memory actuators (left: work against bias spring; right: work against constant force).

Shape memory actuators can be used in two basically different ways: as thermal or as electrical actuators. *Thermal actuators* combine the sensing and the actuating functions, responding to a temperature change by changing shape and/or generating a force. The function of *electrical actuators*, on the other hand, is simply to move an object or perform a task on demand. Usually, a current is passed through the shape memory actuator, internally heating it above A_s to recover its shape.

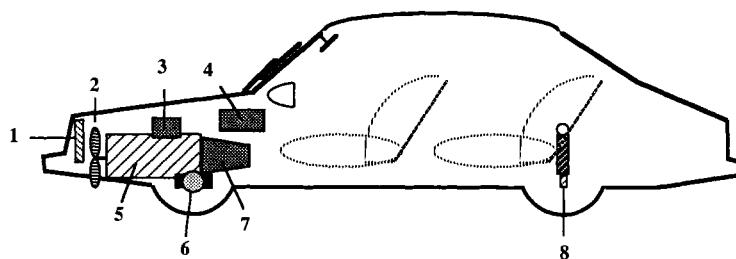
2. Thermal Actuators:

As mentioned earlier, thermal actuators respond to changes in temperature by changing their shape and/or generating a force. Shape memory actuators, in this area, generally compete with thermobimetals and wax motors, and sometimes with electric/electronic devices or vacuum/pneumatic systems. Thermal actuators have a variety of applications in cars. Figure 2 schematically shows areas of both potential and realized applications for shape memory thermal actuators.

There are three different temperature ranges, in which thermal actuators either are already used or could provide significant benefits:

- | | |
|--------------------------|---|
| $< 0^\circ\text{C}$: | <ul style="list-style-type: none"> - low temperature performance (engine, transmission, suspension, brakes) - cold start performance |
| $0 - 60^\circ\text{C}$: | <ul style="list-style-type: none"> - cold start performance - climate control |
| $60 - 120^\circ\text{C}$ | <ul style="list-style-type: none"> - cooling systems (engine, brakes) - restart ability - climate control - rattling (engine, transmission) |

Most of these temperature ranges can be covered by Ni-Ti shape memory alloys available today. However, certain limitations, like transformation temperature ranges vs. required number of cycles, hysteresis width, and stability, have to be considered. Figure 3 shows the transformation temperature ranges of the most common Ni-Ti



- 1 Radiator Shutter**
- 2 Fan Clutch**
- 3 Fuel Management**
- 4 Climate Control**
- 5 Engine Control**
- 6 Brake Ventilation**
- 7 Transmission Control**
- 8 Rattling Noise Reduction**
- 8 Suspension Adjustment**

Figure 2: Potential applications of shape memory thermal actuators in automobiles.

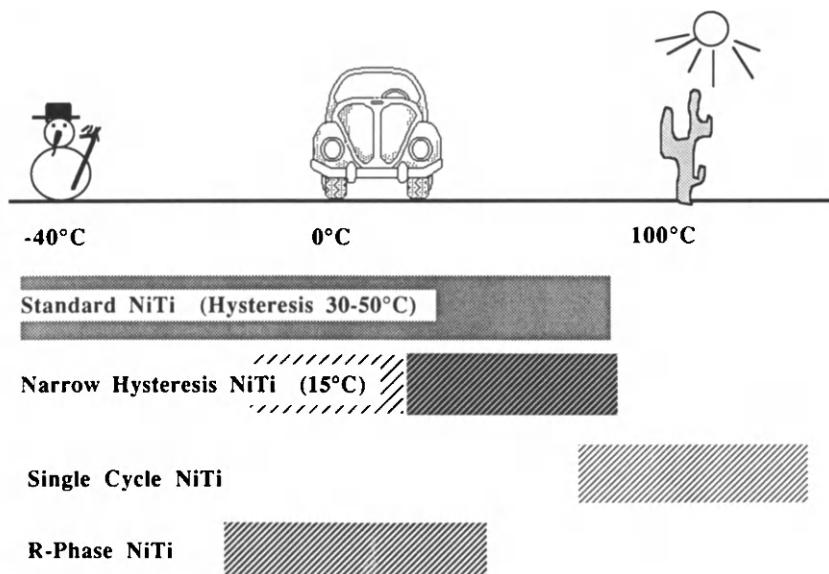


Figure 3: Range of transformation temperatures of commercially available Ni-Ti alloys.

alloys. For multiple cycle applications, standard binary Ni-Ti alloys with transformation temperatures from -50°C to approximately +70°C perform well. These alloys exhibit a hysteresis width of about 30 to 50 degrees and are reasonably stable during cycling. Binary alloys with higher transformation temperatures (up to +120°C) tend to *walk* and, therefore, can only be used for single or low cycle applications. *Walking* is usually defined as accumulated amnesia. It causes drifting of the zero-point during repeated actuation.

Most actuator applications require Ni-Ti alloys with both a narrow hysteresis and high stability of the shape memory effect. The hysteresis width of ternary and quaternary Ni-Ti-Cu alloys, with transformation temperatures from -30°C to +70°C, is only about 15°C. They show excellent stability even after 100,000 thermal cycles. Another important advantage of Ni-Ti-Cu alloys is the low martensitic strength, which allows low reset forces and thus improved work output.

Unfortunately, presently available Ni-Ti alloys having transformation temperatures above 80°C are not sufficiently stable for multiple cycle applications. However for single cycle applications (e.g. over-temperature protectors) binary and ternary Ni-Ti alloys with transformation temperatures up to 150°C are available. Besides having the highest transformation temperatures, ternary alloys can provide high austenitic and martensitic strength.

Patent literature proposes many applications for shape memory thermal actuators. One successful application is a temperature-sensitive governor valve, which controls the shifting pressure in automatic transmissions. This valve's function is shown schematically in Figure 4. At low temperatures, the spring force of a steel bias spring

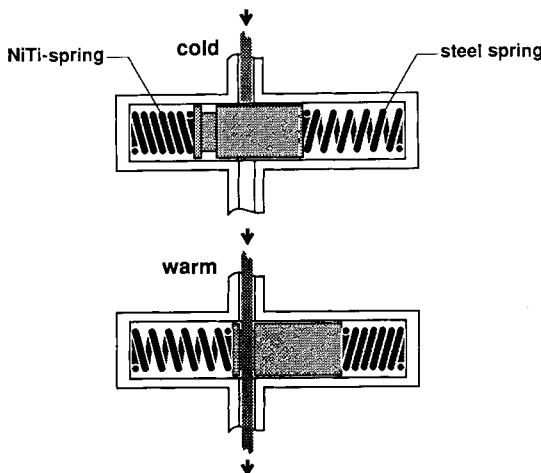


Figure 4: Function of a thermal valve.

is higher than that of the Ni-Ti shape memory spring in the martensitic state. Consequently, the steel spring can compress the Ni-Ti spring, pushing the moveable piston of the valve into the "closed" position for this particular application. When the temperature of the transmission and the transmission fluid increases to operating

temperature, the Ni-Ti spring transforms into austenite. It then expands, overcoming the steel spring force, and eventually, pushing the piston into the "open" position. This pressure regulating valve improves the cold start performance of the transmission, allowing smoother shifting at low temperatures (Figure 5). Other shape memory governor valves control the warm-up phase of automatic transmissions, reducing smog emission and fuel consumption.

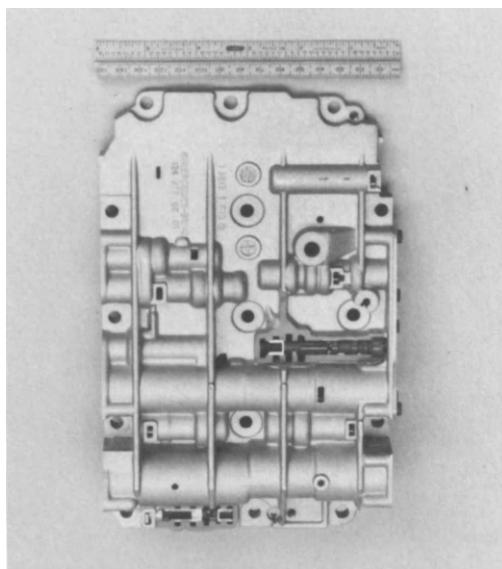


Figure 5: Valve body with two thermal valves (cut-away sections).

Figure 6 and 7 show two different versions of temperature-sensitive boost compensators, and Figure 8 a temperature-compensated valve lifter. Pressure actuators can acquire thermal response by incorporating shape memory springs into diaphragm-type pressure or vacuum actuators. The low temperature characteristics would therefore be different from the high temperature characteristics.

Another automotive application of shape memory thermal actuators, which has been described frequently in the patent literature, is using shape memory washers to reduce gearbox noise. The thermally reactive washers, either Belleville or wave-washers, are intended to compensate for the slackness in sprocket assemblies caused by the differential in thermal expansion of dissimilar materials, like aluminum and steel, when temperature increases. Washers are useful because they provide very high forces for little motion during recovery. Figure 9 and 10 show examples.

Since conventional shock absorbers tend to be too hard at very low temperatures, they don't provide comfortable driving. This is caused by the high viscosity of the oil in the shock absorber, which usually is balanced for the temperature range of 0°C to 100°C. A shape memory washer in the shock absorber's valve (Figure 11), which changes the pressure at low temperatures, compensates for the oil viscosity.

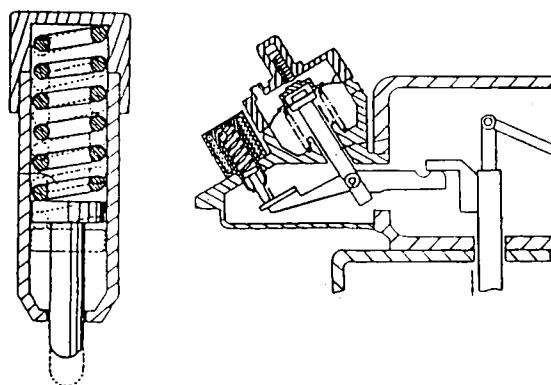


Figure 6: Temperature dependent boost compensator.¹

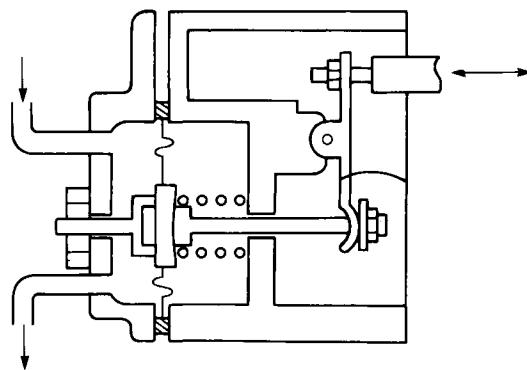


Figure 7: Temperature dependent boost compensator.²

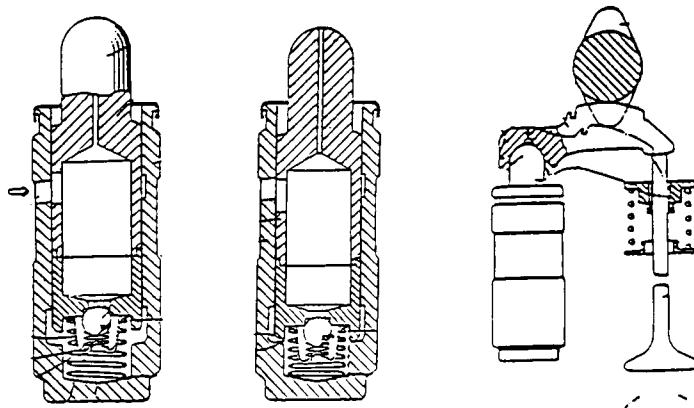


Figure 8: Temperature compensated valve lifter.³

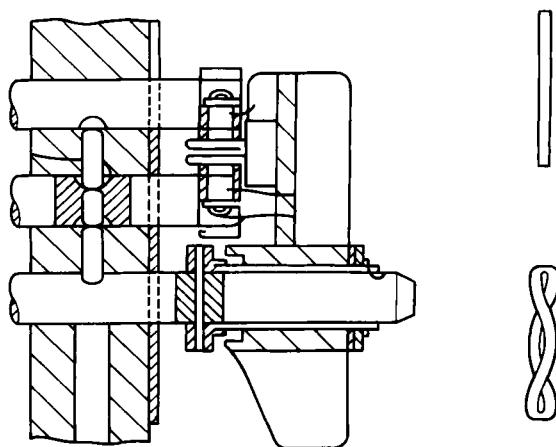


Figure 9: Rattling noise reduction with shape memory wave washer.⁴

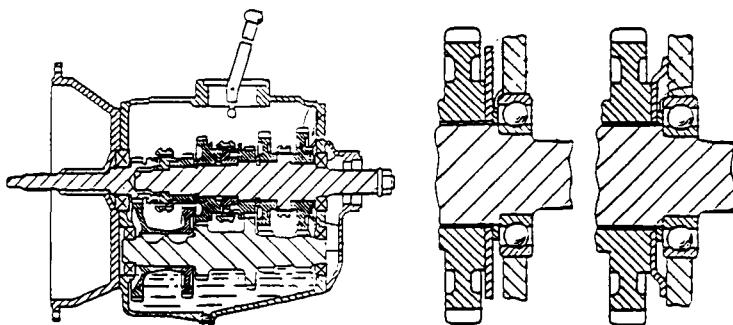


Figure 10: Rattling noise reduction with "Belleville-type" washer.⁵

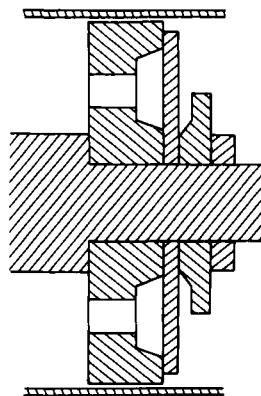


Figure 11: Temperature compensating valve for shock absorbers.⁶

There are many other areas in a car, where shape memory thermal actuators can and will provide significant advantages over competing technologies. Some of the benefits are listed below:

- size
- high force
- large motion
- high work output
- few mechanical parts
- non-linear characteristic

On the other hand, for certain applications the following limitations have to be taken into account:

- limited range of transformation temperatures
- hysteresis
- non-linear characteristic

3. Electrical Actuators:

As mentioned earlier, electrical actuators are devices which perform a task on demand. The stimulus is any voltage applied to the device, which is usually an electric motor or a solenoid. If electrically heated above A_s , such as by passing current through a wire or a spring, Ni-Ti shape memory elements become electrical actuators. They can provide interesting advantages over motors and solenoids like:

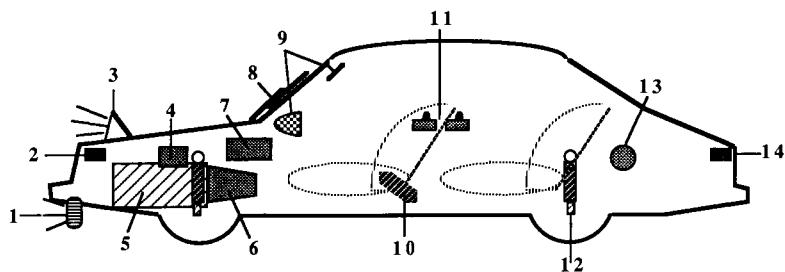
- small size
- noise-less operation
- few mechanical parts

Therefore an almost unlimited number of potential applications can be found in the patent literature. Figure 12 schematically shows the areas in a car where electrical shape memory actuators have been suggested. However, only very few have actually been implemented or seem technically and economically feasible because of the limited range of transformation temperatures of existing shape memory alloys. As shown in Figure 13, the operating temperature range of a car ranges from -40°C to approximately +100°C, with even higher temperatures in under-hood locations. In order to work properly at all temperatures, the shape memory alloy has to have an M_f temperature well above the maximum operating temperature. Today no Ni-Ti alloys with transformation temperatures above approximately 80°C are available for cyclic applications. Cu-Al-Ni shape memory alloys have transformation temperatures this high, but they are brittle and unstable. Ti-Pd-Ni alloys remain too expensive for routine applications.

Among those applications that are in production is the remote fog-lamp louver opening device. A shape memory spring is wired in series with the lamp. Turning on the fog-lamp passes the lamp current through the actuator, which heats up, contracts and opens the louver (Figure 14). A reset spring closes the louver when the lamp is turned off. Figure 15 shows the design of the actuator with an integrated over-load spring. The area where fog-lamps are located on a car is usually very well ventilated, so that the low transformation temperature of the actuator (~ 65°C) is not an issue.

The same applies for windshield wipers. Shape memory actuators could provide an elegant solution for increasing the pressure at high speeds. Figure 16 shows different designs incorporating a Ni-Ti spring or tensile wire into the wiper arm.

Because of their small size and noiseless operation, shape memory actuators have also been suggested for central locking systems (Figure 17), trunk locks and fuel tank cap locks. However, there are serious problems when ambient temperature



- 1 Foglamp Louver
- 2 Engine Hood Lock
- 3 Retractable Head-Light
- 4 Fuel Management
- 5 Engine Control
- 6 Transmission Control
- 7 Climate Control
- 8 Wiper Pressure Control
- 9 Rear-View Mirror Adjustment
- 10 Seat-Belt Adjustment
- 11 Central Locking System
- 12 Shock Absorber Adjustment
- 13 Filler Inlet Lock
- 14 Trunk Lock

Figure 12: Potential applications for electrical shape memory actuators in automobiles.

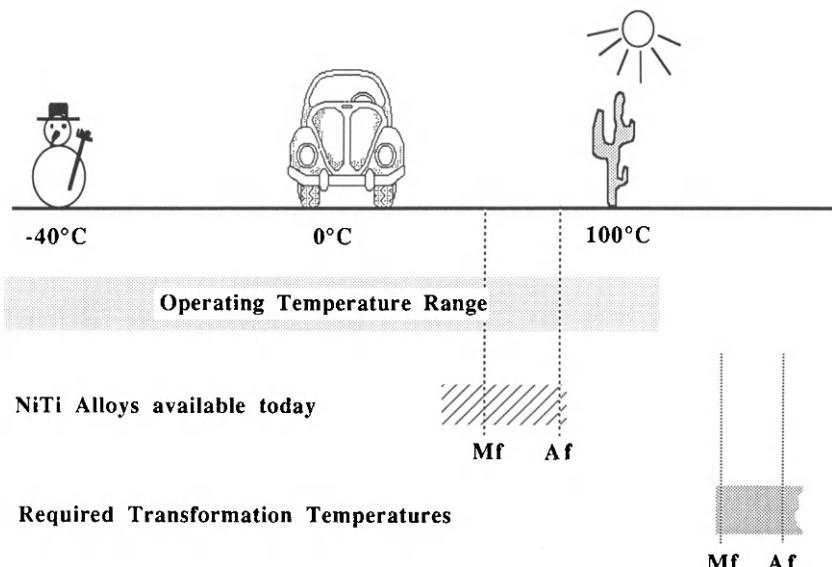


Figure 13: Operating temperature range for automobiles and transformation temperatures of Ni-Ti alloys.



Figure 14: Fog lamp with shape memory electrical actuator.

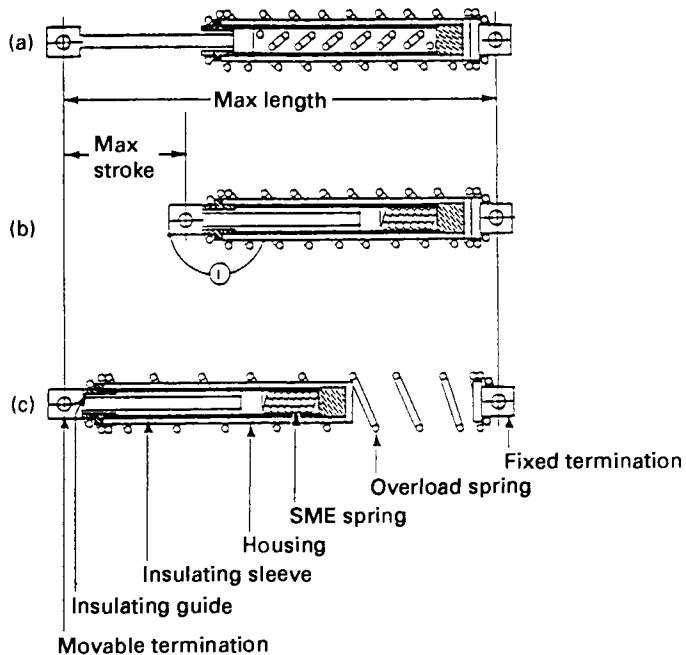


Figure 15: Design example for a linear shape memory actuator.

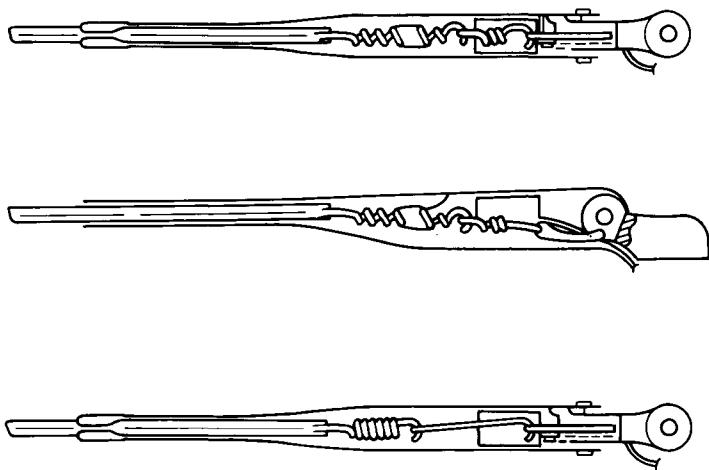


Figure 16: Windshield wiper with shape memory pressure actuator.⁷

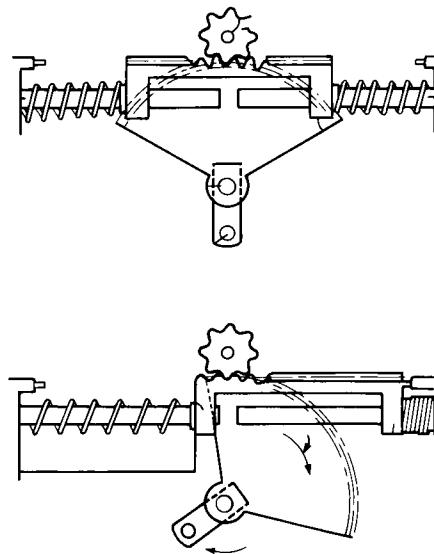


Figure 17: Door locking mechanism with shape memory springs.⁸

approaches the transformation temperature of the shape memory alloy, for instance when the car is parked in the sun. In this case, the shape memory actuator could self-actuate or fail to reset.

4. Conclusion:

Although shape memory actuators can provide significant advantages over conventional devices in certain areas, they have only slowly been penetrating the automotive market. This is mainly caused by poor information and the lack of engineering data for shape memory alloys. However, with the better understanding of the metallurgy and a more open discussion in the scientific community, shape memory actuators are becoming increasingly popular for automotive applications.

References

1. Hino Motor Industry Kokai No. 63-40543 (1988).
2. Nissan Diesel Industry Kokai No. 62-169230 (1987).
3. Mitsubishi Motor Kokai No. 62-160706 (1987).
4. Toyota Motor Kokai No. 62-50361 (1987).
5. Hino Motor Industry.
6. Aishin Seiki.
7. Mazda Kokai No. 60-173455 (1985).
8. Kokusan Metal Industry Kokai No. 62-173472 (1987).

The Use of Shape Memory Alloys for Fire Protection

W. Van Moorleghem and D. Otte
Proteus N.V.
Stasegemsteenweg 110E
B-8500 Kortrijk, Belgium

The role of shape memory alloys in the field of fire protection is clear: a shape memory element can detect an increase in temperature and at the same time react to the increase by doing mechanical work. The different applications of shape memory alloys in fire protection can be divided into three groups according to the function of the element: detection, detection and actuation, and simply actuation. In the last group, the shape memory element is only used to directly transform electrical energy to mechanical energy. In the case of simple actuation, all of the advantages of electronics can be used.

1. Detection

The detection of a temperature change can be accomplished through the detection of resistivity changes associated with the martensite-austenite transformation, or through the shape memory effect itself. Both aspects will be discussed in turn.

1.1 Resistivity

Like conventional materials, memory alloys show a slow increase in resistivity with increasing temperature, but when heating is continued into the transformational temperature range, there will be a sudden and marked reduction in resistivity of some 10 to 20% (Figure 1). This change is quite complex in Ni-Ti alloys, but very distinct in Cu-based alloys. When a measuring circuit registers a certain rate of resistance loss, it indicates that the temperature has increased and that the transformation temperature has been reached. On the other hand, a sudden increase in resistance indicates a drop in temperature through the transformation regime. Thus a system can be secured against overheating, e.g. a fire-alarm, or an electrical installation can be secured against overheating and undercooling.

These types of detectors become most interesting when large installations or great lengths, e.g. electrical cables, must be protected. This can be easily accomplished by the use of a shape memory alloy wire. Figure 2 shows an overheating signal generator: the input of the circuit is a Wheatstone bridge (9) which contains a shape memory wire (15). When the bridge is out of balance, the difference will be amplified by the op-amps (4 and 12) which activate a relay (13). The feedback (20) on the first op-amp (consisting of a clock-pulse generator, an up-down counter and a digital-to-

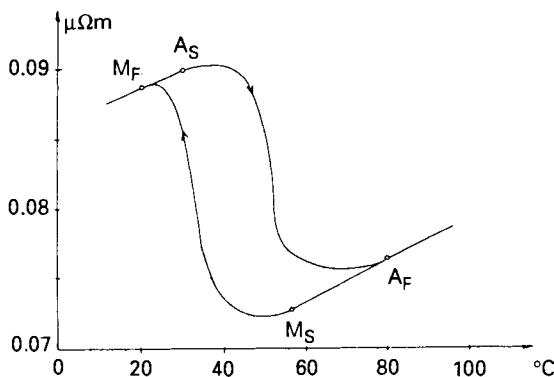


Figure 1: Resistivity of a Cu-based SMA as a function of temperature.

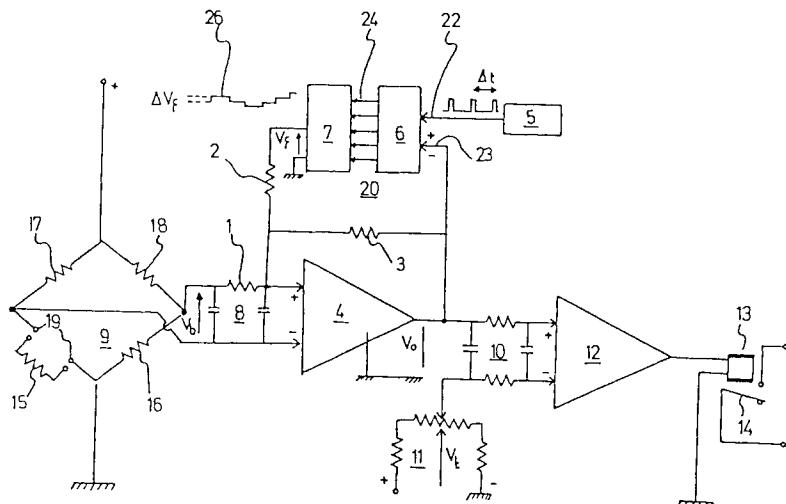


Figure 2: Overheating signal generator.

analogue converter) assures that the circuit only reacts to a sudden drop in resistance as a result of crossing the transformation temperature regime, and not to the gradual decrease in resistance outside of the transformation area.

1.2 The Shape Memory Effect

Another possibility for making a linear detector is to use the memory effect itself. Figure 3 shows the shrinkage of a Cu-Zn-Al wire in function of temperature. This shrinking is sufficient to activate an alarm system by means of a microswitch or equivalent device. An example of such an application is pictured in Figure 4. It represents a cotton spinning mill, which is difficult to protect in any other way.

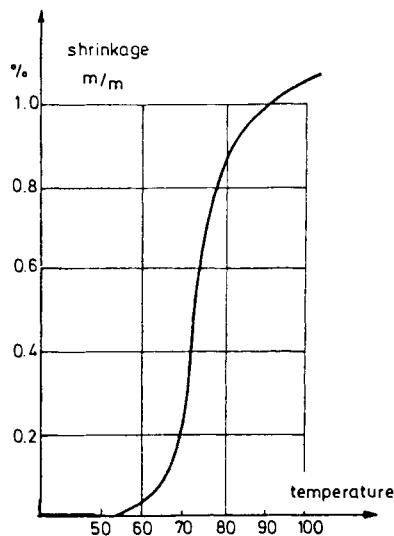


Figure 3: Shape memory effect of a Cu-based wire.

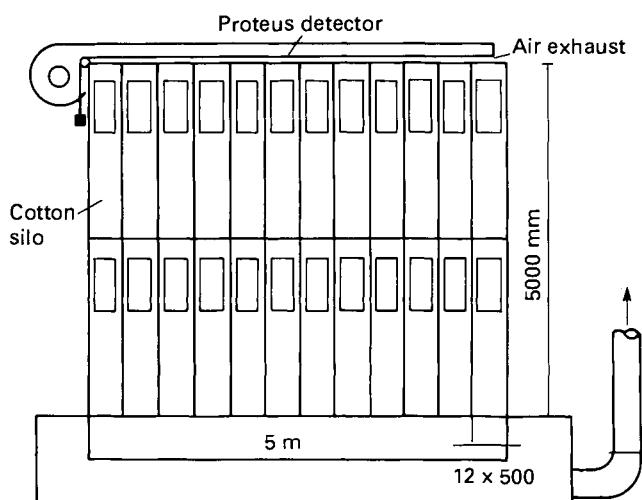


Figure 4: Application of a linear detector.

If two elements with different transformation temperatures are placed in series, a two step indicator can be realized (Figure 5). Such a detector can activate an alarm before activating a fire-extinguishing system.

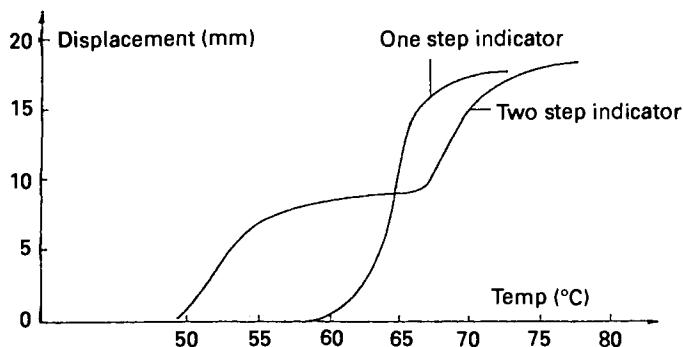


Figure 5: One and two step indicator.

2. Detection and Actuation

The application presented here is the Proteus Gas Valve (Figure 6). This gas valve has been specifically designed to cut off gas flow in case of fire. The valve is provided with a Cu-Zn-Al helical spring for the detection and actuation. If the ambient temperature exceeds the transformation temperature of the shape memory alloy, the spring expands pushing a steel ball through a stainless steel retaining ring (Figure 7). The ball closes the valve which can be reset only manually (Figure 8). In order to guarantee a quick reaction time, an aluminum heat exchanger can be mounted on the valve. Typical reaction times are shown in Figure 9.

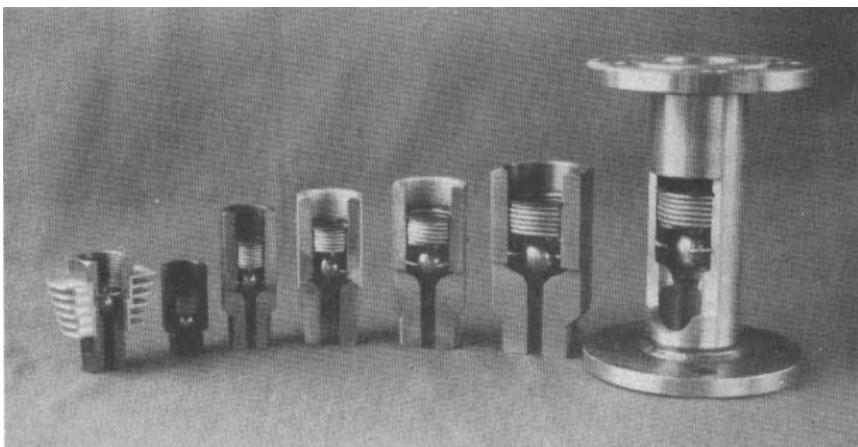


Figure 6: Proteus thermic valves.

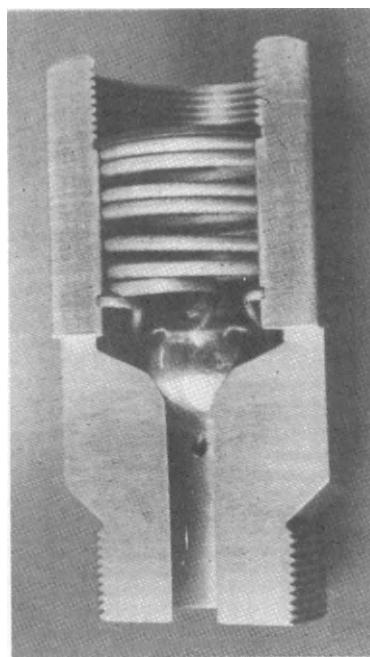


Figure 7: Proteus thermic valve.

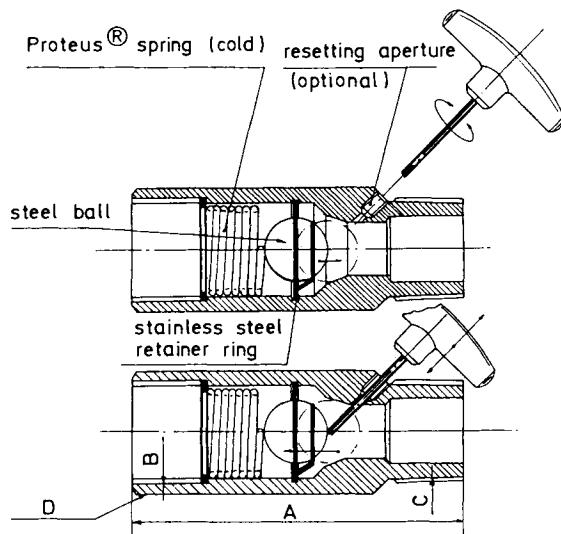


Figure 8: Resetting mechanism.

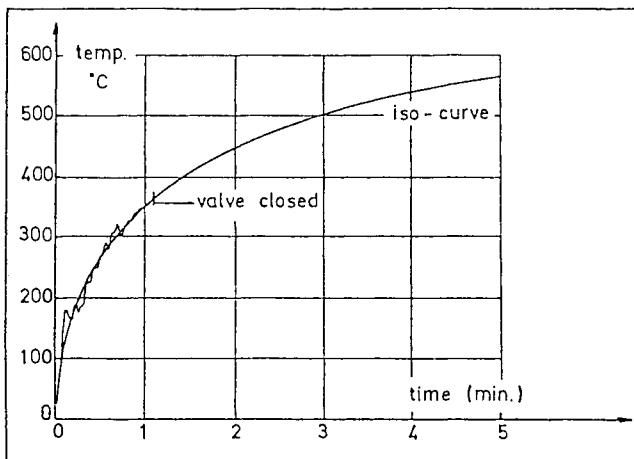


Figure 9: Reaction time of a Proteus thermic valve.

A second example is the Proteus Cu-Zn-Al link, (Figure 10) which is a temperature responsive fastener. Figures 11 and 12 demonstrate how the two parts of the link move perpendicularly with respect to the stress with which it is attached. This method, with the shape memory effect and external forces acting in different directions, is a very good way of working with shape memory alloys. Figure 13 shows an example of how the Proteus link operates a sprinkler, and Figure 14 shows a control unit of a fire valve.

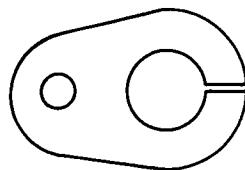


Figure 10: Proteus link.

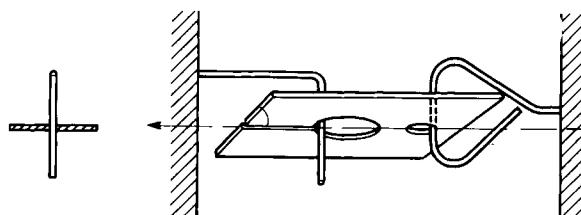


Figure 11: Illustration how Proteus link functions.

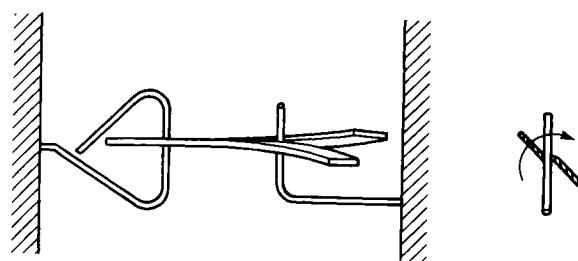


Figure 12: Illustration how Proteus link functions.

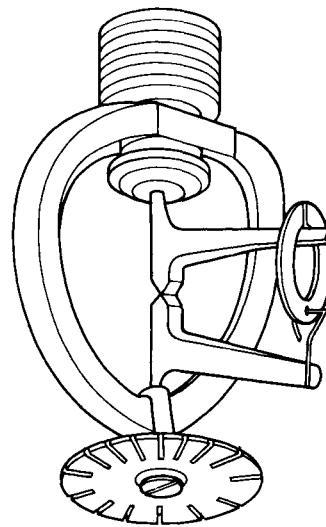


Figure 13: Sprinkler with Proteus link.

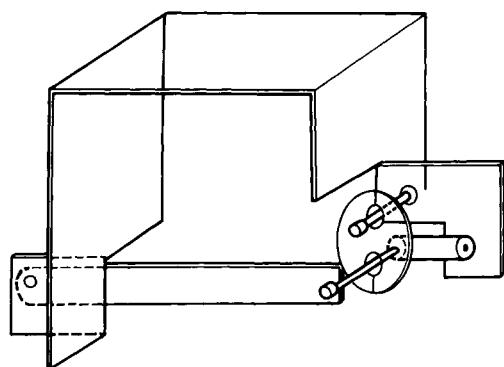


Figure 14: Control unit of a fire valve (curtain type).

3. Actuation

At the moment, there is a tendency to use precision electronics for detection and shape memory alloys only as compact and inexpensive energy transformers (actuators). An example is shown in Figure 15: on this link, one or more PTC's (Positive Temperature Coefficient resistors) are mounted. A PTC has the advantage that it cannot cause overheating. The principle of the associated electronic circuit is represented in Figure 16: a temperature responsive element activates an op-amp, which in turn activates the PTC, which heats the shape memory element. It is also possible to replace the PTC temperature responsive element by a photocell, smoke detector, etc.

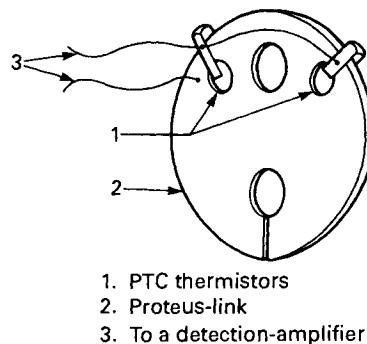


Figure 15: Proteus link activated by a PTC.

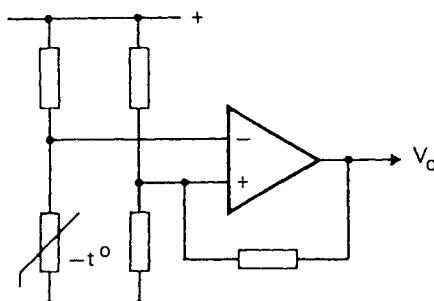


Figure 16: Electronic detector and amplifier for driving element in Figure 15.

In Line Anti-Scald Safety Valve

Stephen M. Fisher
Memry Corporation
83 Keeler Ave.
Norwalk, CT 06854

Hospital discharge summaries and emergency room logs indicate hot liquids as the major cause of nonfatal burns. In fact, many more children are burned by hot liquids, then are killed or injured by house fires. Children under the age of five are most frequently victims of scalds burns with 46 scalding deaths and 1,700 hospitalized injuries during 1985. The vehicles most often associated with these burns are hot beverages and foods encountered in the kitchen and dining area and hot tap water, particularly that encountered in the bath or shower. Tap scald burns are worse than hot liquid spills. The burns which result from hot tap water are usually more severe, cover a larger portion of the body and result in hospitalization more often than burns from other hot liquids. Many adults are not aware that tap water causes such severe injury.

Clearly, there are many interventions. Unlike scald burns from spilled hot liquids, which for the most part can only be prevented by behavior changes such as remembering to turn pot handles towards the back of the stove, there are many points of intervention for tap water scald burns. One of these is "point-of-use" control of water temperature.

The present chapter describes one approach to point-of-use thermal protection, using a Cu-Zn-Al shape memory thermal actuator. In contrast to many of the previous papers, the description that follows is not keyed on shape memory, but the product itself. This is a reminder, in a sense, that products are not designed around shape memory, and the success of a product is, more often than not, controlled by the design of the surroundings of an SMA element.

1. Traditional Control Systems

Early hot water delivery systems were basic in design and construction. The heat source was the same boiler which provided heat to the building. For the bather, there were separate faucets and spouts for hot and cold water. The user simply regulated each faucet until a comfortable bathing temperature was achieved. Showers were relatively unknown. As hot water systems become more commonplace, controls progressed very rapidly. Mixing valves which combined hot and cold water using two handles and a single tube spout became the standard. A shower arm was added, and the American bathroom began to resemble its modern counterpart in design.

1.1 Manual Mixing Valve

The most basic control valve is a simple mixing valve. This valve combines hot and cold water sources by manual manipulation of two handles. The user turns the two valves until a desired temperature is achieved. Many versions of this basic mixing valve are manufactured to very high quality standards and perform well. However, there are low cost valves on the market which are difficult to adjust. These valves are common in lower cost residential housing and budget hotels. The mixing valve's low cost and simplicity explain its popularity and its widespread use. This control valve offers no anti-scald safety benefit.

1.2 Pressure Balancing Valve

An improvement over the simple mixing valve is the pressure balancing valve. This valve is usually controlled by a single lever. It contains a valving mechanism designed to compensate for pressure fluctuations or loss on either hot or cold water lines. The valve is easier for the user to regulate, and it provides a measure of safety greater than the simple mixing valve.

The pressure balancing valve's basic limitation is it cannot sense temperature changes. Also, if the hot water source is increased in temperature, the valve requires an adjustment of its limit stops. The limit stops are built into the valve in an effort to provide a high temperature limit and prevent scald injuries. In practice, the stops are ineffective. They are often improperly adjusted, or fail to function as intended with variations in hot water demand. In certain designs, hard water and lime scale can render the pressure balance valve ineffective with no evidence to the user.

This valve is used in many institutions such as hotels, and is also widely used in medium range, residential housing. In some areas it is now required in all new construction.

1.3 Thermostatic Mixing Valve

The thermostatic control valve is far more effective in compensating for temperature variations than the pressure balance valve. As its name implies, the valve is designed to compensate for changes in water temperature. When the unit is functioning properly, it is very effective in providing safety to the user. Most hospitals and health care facilities use this valve, as do luxury hotels and upscale residential housing. Because of its high relative cost, the thermostatic valve is not as widely used as simple mixing or pressure balancing valves. Its safety shortcoming is its susceptibility to malfunction in water where corrosion and lime scale deposits are a problem. It can require maintenance on a periodic basis to ensure its temperature limits stops are properly adjusted and functioning.

In large commercial or institutional buildings where hot water is distributed at very high temperature (65°C to 90°C), branch lines to each floor or wing utilize master thermostatic mixing valves to reduce the water temperature to a safer range (43°C to 48°C). Point of use safety mixing valves are still required to assure bather safety.

1.4 Safety valves

There have been attempts in the past to develop a safety valve that acts as a safety "switch" to reduce the flow of water if an "over temperature" condition develops. There

were a number of designs developed which were based around bi-metallic or wax cartridge technology. One wax cartridge product was marketed for a short period and discontinued due to problems associated with use. Now, shape memory effect technology (SME), possesses the physical characteristics needed for the design of such a product.

2. Thermal Responsive Valve Design

Shape memory technology, due to its unique physical properties, lent itself to the design of a thermal responsive safety valve. The first step in the design process was to evaluate the environment. This included reviewing all the physical parameters that may be encountered in service such as temperature, pressure, flow rate and mechanical limitations. The next step was to collect information on the performance requirements by evaluating conditions precedent to causing a burn and reviewing available information on the various plumbing and building codes. Once this was completed, the desired functional parameters were formulated for product performance. These included items such as fast response, precise temperature control, and ease of installation. Then came the actual mechanical design. One major item of consideration was to make the design independent of as many external forces as possible, such as pressure fluctuations and system friction. After taking and assimilating all the foregoing information a decision was made to design the device as a bi-stable thermal switch. The concept was to have the device remain fully open up to the requisite "shut-off" temperature at which time the valve should close in minimal time. In order to operate as a bi-stable switch the valve was designed to close with the water flow. This required the minimum force to close the valve and thus provides lower force requirements on the SME elements, greater temperature sensitivity and faster response. A general design concept was established, see Figure 1, and then optimized to cover a full range of pressure (flow) and temperature conditions.

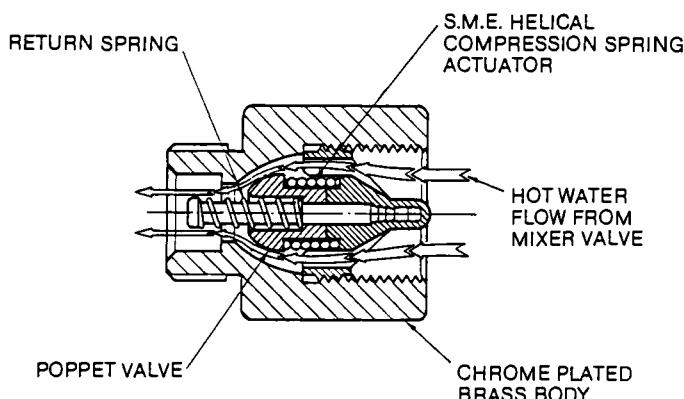


Figure 1: The Shower Gard® Anti-Scald Valve protects against hot water scalding by stopping water flow in less than one second when water temperatures reach dangerous levels.

2.1 Dynamic Analysis

The analysis was done in three parts: a quasi-steady state analysis including the entire test setup and neglecting any coupling; an analysis for the pressure related forces on the poppet including the effect of the poppet motion, but neglecting the acceleration (dV/dt) of the fluid; and a dynamic force balance on the poppet and solution for poppet velocity and location versus time.

2.1.1 Quasi steady state analysis

The analysis model is shown in Figure 2. The valve housing and poppet are represented as cylinders. This simplification is more amenable to analysis and still

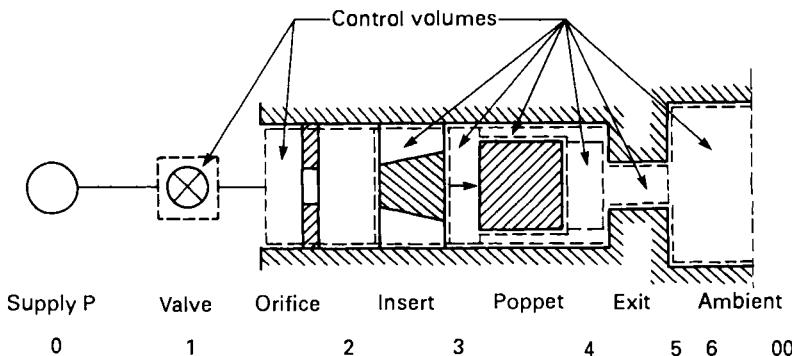


Figure 2: The analysis model used shows idealized key locations in the valve, from the supply source (o) to the outlet (oo).

retains the most important characteristic of the valve namely the reduction in flow area as the valve closes.

Upon applying conservation of mass and energy for the control volumes indicated the entrance velocity, V_1 , is given by:

$$V_1^2 = \frac{2/\rho(P_o - P_{oo})}{C_v + C_o + C_i(A_1/A_3)^2 + C_{sp}(A_1/A_3)^2 + C_p(A_1/A_5)^2 + C_e(A_1/A_5)^2 + C_{eo}(A_1/A_6)^2} \quad (1)$$

where the areas A_1 through A_6 are the cross sectional areas at the points shown in Figure 2, ρ is fluid density, P_o is the static inlet pressure, P_{oo} is the incremental pressure due to flow, and subscripted "C"s are various loss coefficients: specifically, C_v is the valve loss coefficient (determined from test data to be 61.5), C_o is the orifice loss coefficient (determined from test data to be 436), C_i is the inlet loss coefficient, determined from an assumed contraction to be 0.158, C_{sp} is the loss coefficient from the sides of the poppet (assumed to be negligible), C_e is the loss coefficient from the exit (assumed to be 0.24), C_{eo} is the loss from the expansion to ambient (assumed to

be 1.0). C_p , the loss due to the change in flow area as the poppet goes from open ($x=0$) to closed ($x=L=3.1\text{ mm}$) is approximated by:

$$A_4 / A_5 = 0.462 / (1-X) \quad \text{where } X=x/L \quad (2)$$

leading to two conditions:

$$C_p = (1 / \{0.62 + 0.38(A_5/A_4)^3\} - 1)^2 \quad \text{for } A_4 / A_5 < 1 \quad (3)$$

and

$$C_p = (A_5/A_4 - 1)^2 \quad \text{for } A_4 / A_5 > 1 \quad (4)$$

A plot of C_p versus X is shown in Figure 3. The areas used in the evaluation are as follows: $A_1 = 248. \text{ mm}^2$, $A_2 = 248. \text{ mm}^2$, $A_3 = 157. \text{ mm}^2$, $A_4 = 107.(1-X) \text{ mm}^2$, $A_5 = 49.5 \text{ mm}^2$ and $A_6 = 96.8 \text{ mm}^2$.

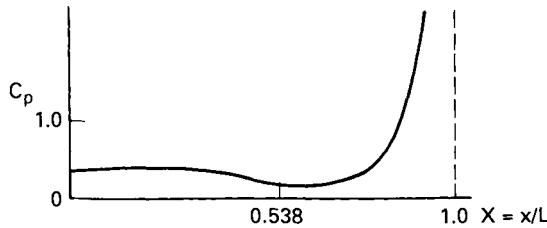


Figure 3: C_p , the loss coefficient due to the change in flow area as the poppet goes from open to closed is plotted as a function of the normalized location of the poppet.

The flow rate is given by $Q = A_1 V_1$. The pressure drop from the inlet-2 to the outlet-5, corresponding to the data taken in the tests, is given by:

$$(P_2 - P_5) = (0.5\rho V_1^2)[(A_1/A_5)^2 - (A_1/A_2)^2 + C_p(A_1/A_5)^2] = C_i(A_1/A_3)^2 \quad (5)$$

The pressure differential across the poppet (3-4 in Figure 2) is given by:

$$(P_3 - P_4) = (0.5\rho V_1^2)[(A_1/A_4)^2 - (A_1/A_3)^2] = (0.5\rho(Q/A_3)^2)[(A_3/A_4)^2 - 1] \quad (6)$$

This last equation shows that the pressure differential across the poppet is due to the increase in velocity as the flow area at 4 is reduced. Plots of V_1 , Q (the volume flow rate), $(P_2 - P_5)$ and $(P_3 - P_4)$ are shown in Figure 4.

2.1.2 Pressure Differential

The motion of the poppet causes an expansion (increase) in the fluid volume above the poppet and a compression (decrease) in the fluid volume at the closure side of the poppet (Figure 5). Neglecting any unsteadiness in the fluid ($dV/dt=0$), and assuming that the overall mass rate of flow is unaffected by the poppet motion, eg $Q_1=Q_5$ in a quasi-static manner, conservation of mass for the upper and lower control volumes

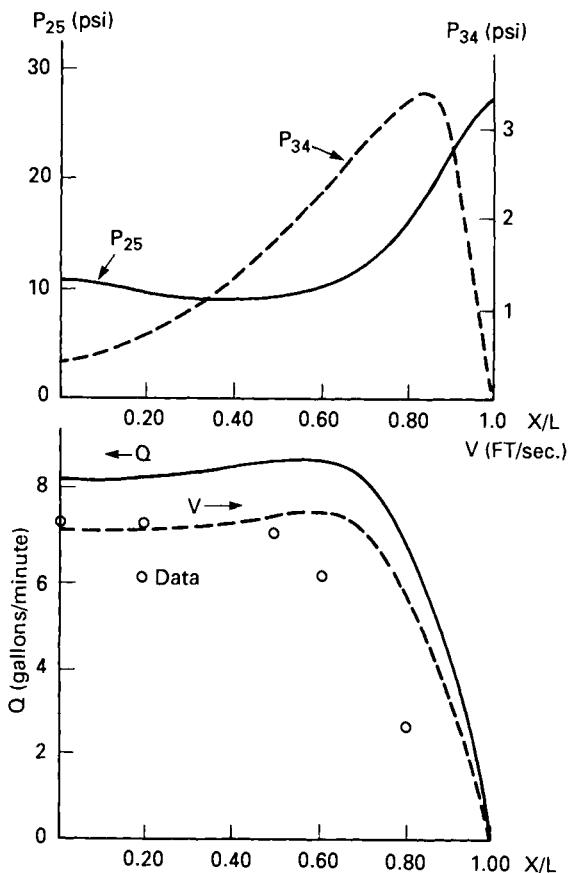


Figure 4: The results of a quasi-steady state analysis of the valve.

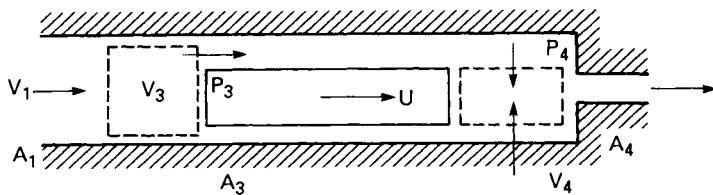


Figure 5: Model control volumes for the analysis of pressure differential with respect to poppet motion.

when combined with conservation of energy for the same volume results in the following equation for $(P_3 - P_4)$, as shown in Figure 5.

$$(P_3 - P_4) = [0.5\rho(Q/A_3)^2] [(A_3/A_4)^2 - 1] [1 - (A_2 U/A_1 V_1)]^2 \quad (7)$$

where $U = dx/dt$, the velocity of the poppet. Note that the first two terms in brackets are the expression for the pressure differential neglecting the poppet motion. The last bracketed term is the effect of the coupling between the motion of the poppet and the pressure difference across the poppet. The coupling is represented by the ratio of poppet velocity, U and fluid velocity, V_1 , times the area ratio A_2 [the area below the inlet flow divider] divided by inlet area A_1 . The current analysis assumes these areas are equal. Equation (7) predicts a reduction in the pressure differential across the poppet due to the coupling, with the value becoming zero at U equal to V_1 . For values of U greater than V_1 , the analysis predicts the possibility of reverse flow with an increase in $(P_3 - P_4)$.

2.1.3 Dynamic Force Balance

Based on the following assumptions; neglect damping, assume both the memory spring and bias spring are linear, a force balance on the poppet; see Figure 6.

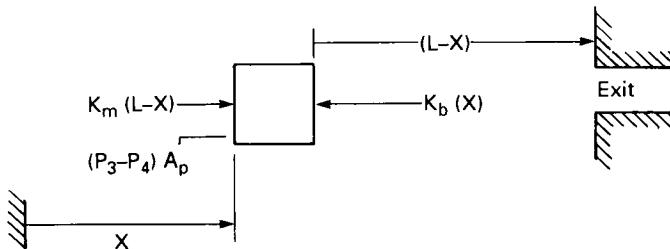


Figure 6: A schematic of the force balance on the poppet.

$$M \frac{d^2x}{dt^2} - K_m(L-x) - K_b(x) - (P_3 - P_4)A_p = 0 \quad (8)$$

where A_p is the cross sectional area of the poppet, and K_m and K_b are the spring rate constants for the memory and biasing springs respectively. Upon substitution of $U = dx/dt$ and $X = x/L$, and equation (7) for $(P_3 - P_4)$, the above equation becomes;

$$M \frac{d^2x}{dt^2} - k_m(1-x) - k_b(x) = \rho/2(V_f/M) (Q/A_3)^2 [(A_3/A_4)^2 - 1] [1 - (A_2 U/Q)^2] \quad (9)$$

where M is the mass of the poppet, $k_m = K_m L^2/M$, $k_b = K_b L^2/M$, and $V_f = A_p L$. Note that (K/M) is the square of the uncoupled frequency of the memory spring and the bias spring. The above equation is evaluated in the next section for typical values of the parameters.

2.2 Numerical Results

Equation (9) was evaluated numerically for the case where $K_b = K_m = 0.3 \text{ N/mm}$ with $L=3.2 \text{ mm}$, $M = 3.5 \text{ grams}$, giving the general equation:

$$U^2/2 = K_m(X - X^2/2) - K_b(X^2/2) = \text{Sum}\{Q, V, U=f(x)\} \Delta X \quad (10)$$

where the Sum term is last term of equation 9 and ΔX is the incremental value of X (taken as $L/100$). The following incremental cases were then evaluated (with $P_o - P_{\infty}$ equal to 1.9 bar):

Table I:

Case	C_o	K_m	K_b	Coupling
1.	0	1350	1350	No
2.	0	1350	1350	Yes
3.	436	1350	1350	No
4.	436	1350	1350	Yes
5.	0	1350	2700	Yes
6.	0	1350	675	Yes
7.	436	1350	2700	Yes
8.	436	1350	675	Yes

The results are shown in Figure 7. Figure 7a are the results of cases 1 and 2; no flow orifice without and with coupling. The no-coupling case shows an steadily increasing value of velocity. The case with coupling shows a gradual decrease in velocity until a valve of (U/V_1) equal to one is approached. After this the velocity increases rapidly with X . Figure 7b, cases 3 and 4, show basically the same result, though due to the orifice the flow is reduced resulting in a reduction of the pressure force across the poppet, thus a lower poppet velocity. Again, the uncoupled cases shows a gradual increase in U as X approaches unity, whereas the coupled case against shows an abrupt increase.

Though the analysis may not be valid as X approaches 1 it does show the tendency of the poppet motion to become less stable near this limit. The analysis indicates that as the poppet moves down the fluid velocity is decreased at both the top and bottom portions (Figure 8a). At the point at which the poppet and fluid velocity are equal ($V_1=U$) no flow is transferred from the top to the bottom since the poppet is traveling at the same rate at which fluid is entering (Figure 8b). As the poppet velocity increases to a value greater than the fluid velocity the analysis predicts that flow can reverse, causing fluid from the bottom of the poppet to flow to the top. This results in an increase in pressure differential across the poppet which further increases the poppet velocity (Figure 8c).

Figure 7c shows the results of cases 5 and 6; the effect of changing the poppet velocity by varying the spring constant of the bias spring. The results show only a minor effect of the bias spring, though the increase in spring constant does show the effect of moving the instability point to be closer to closure ($X=1$).

Figure 7d shows the results of case 7 and 8; insertion of an upstream orifice with changes in the value of bias spring constant. The combination of the orifice and a stronger bias spring (case 7) indicates a stable closure of the poppet.

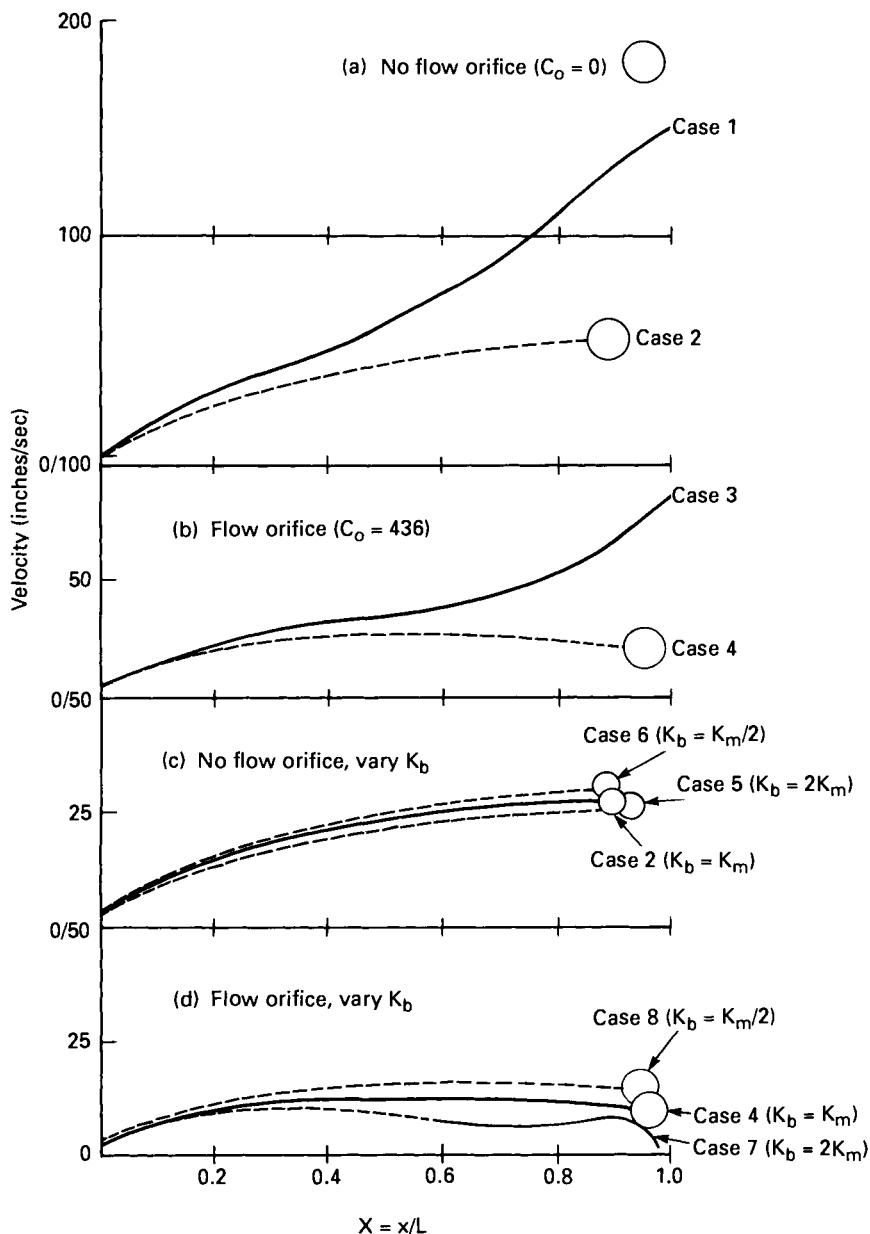


Figure 7: The results of the force balance analysis on the poppet.

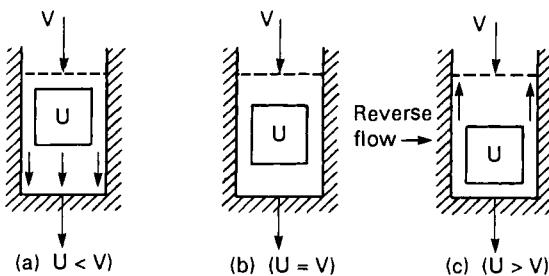
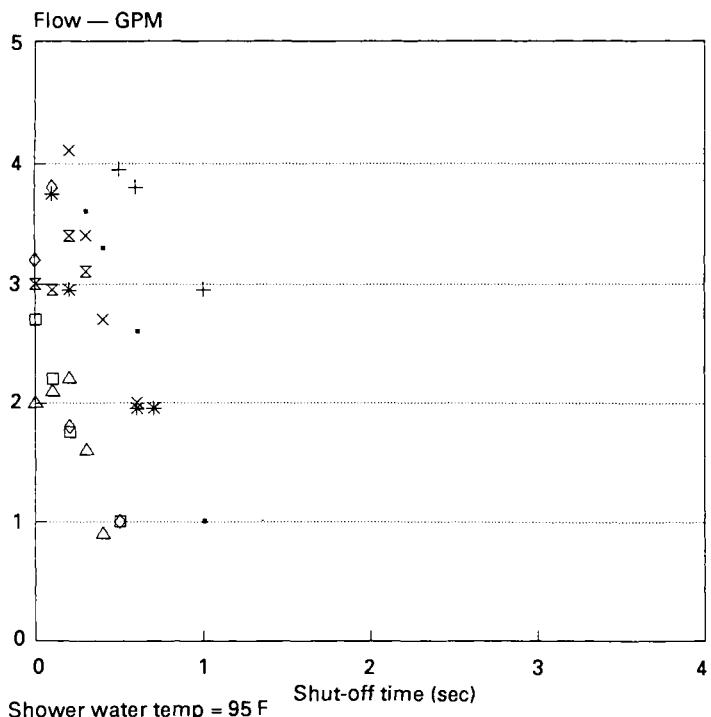


Figure 8: The position of the poppet with various flow configurations.

3. Valve Operation

The water flow passes through the valve with minimal flow restriction (40 liters per minute maximum flow). When the temperature exceeds the maximum as calibrated during manufacturing, the SME actuator expands and pushes the poppet valve further



into the fluid flow stream forcing it to close against the housing. With the poppet closed, the valve will have a residual flow of not greater than 20 liters per minute preventing cross flow from the hot to the cold. The valve is then reset by cooling the shape memory actuator and relieving the water pressure. The return spring moves the poppet valve back to its original position.

3.1 Valve Performance

Performance of the temperature responsive valve is shown in Figures 9 and 10. Figure 9 displays the response time from full open to closed (restricted flow) for the valve at flow rates from 4 liters per minute to 20 liters per minute. The response time is less than one second for a full range of pressures from 1.7 to 5.5 bar and hot water temperatures from 49°C to 71°C. Figure 10 displays the shut off (restricted flow) temperature at flow rates from 4 liters per minute to 20 liters per minute covering a full range of pressures from 1.7 to 5.5 bar and hot water temperatures from 49°C to 71°C. The results show valve operating temperature going from 49°C to 52°C over this full range of conditions.

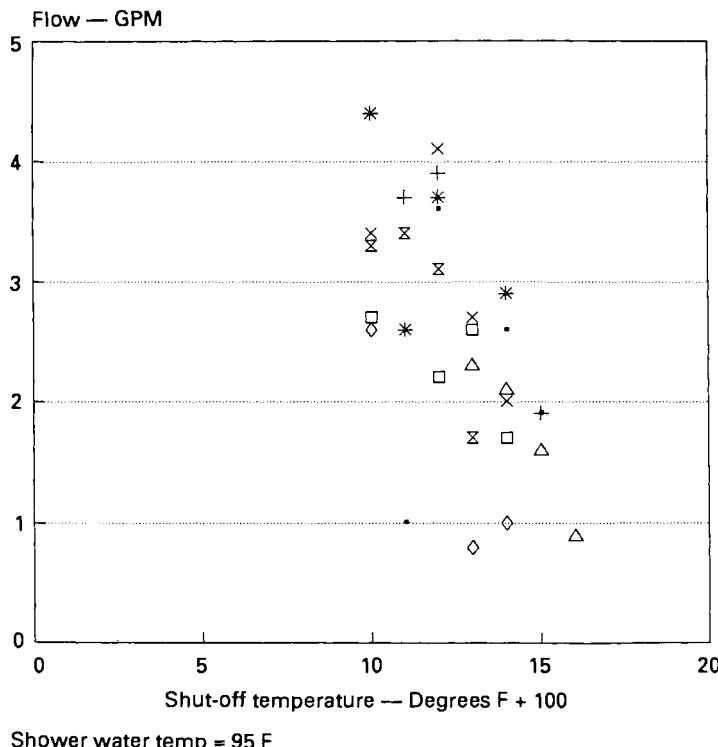


Figure 10: The Temperature at which shut-off occurs. Each symbol represents the mean performance for a group of 10 valves operating at 36°C and suddenly subjected to inlet water between 49°C and 71°C at a pressure between 1.7 and 5.5 bar.

From these extensive tests a general performance specification was developed for the thermal responsive valve, conforming with ASTM standards F444-78 and F445-78. The finished product will shut down water flow within 5 seconds after reaching 43°C and before reaching 49°C to a residual trickle of less than 2 liters per minute. Although the unit's performance will vary with installation and operating conditions, it will never exceed the limits state above assuming it is properly maintained.

4. Other Design Options

Shape memory technology lends itself to numerous other methods of controlling the flow of a fluid. The "bi-stable switch" design is extremely fast, however, has the disadvantage of having to relieve the water pressure to restore use. Other designs which act more on a modulating principle can provide similar performance while allowing for automatically resetting and continual non interrupted use. These would lend themselves well to being incorporated in other valves where no thermal control now exists or as a secondary backup. In addition, SME technology could also be used as a control element to mix hot and cold water in a thermostatic mixing valve. The SME alloys have a very high degree of corrosion resistance, are economical, and have excellent performance characteristics.

5. The Future Is Now

As new technology moves into the plumbing industry, more and more innovative product ideas will surface affording the industry an opportunity to provide greater comfort and safety to its customers. Shape memory technology will play a role in shaping the future of the plumbing industry.

Shape Memory Sensor and Actuator for Air Conditioners

Tsunehiko Todoroki
Air Conditioning Research Laboratory
Matsushita Electric Industrial Company Ltd.

The world's total demand for room air conditioners was 11 million units in 1987: four million in Japan, 3.5 million in North America, 1.1 million in the Middle East, 1 million in Southeast Asia and 1.4 million in other countries. The Japanese market has grown from 3.35 million units in 1985 to 3.5 million in 1986, 4 million in 1987 and now 4.45 million units are expected in 1988.

The leading type of air conditioner in Japan is the split type, consisting of outdoor and indoor units, as shown in Figure 1. The indoor unit is generally installed in the upper part of the room wall in order to make the best use of indoor space. In the case of a system used to control room temperature by turning on and off a compressor, the temperatures of inlet and exit air change during operation are as shown in Figure 2. When the air conditioner is installed in the upper part of the wall, it is necessary to change the direction of the air according to the temperature of the exit or diffused air; that is, hot air rises in the room so it is necessary to blow the diffused air downwards in order to warm the air near the floor, and when the diffused air temperature is lower than ambient temperature, it is necessary to direct the diffused air horizontally so that the person in the room will not be exposed to cold drafts. The conventional automatic air direction changing mechanism consists of a flap installed at the air outlet port, which is shifted by a motor, and uses a thermistor sensor to detect changes in the diffused air temperature and an IC to control the entire system. Shape memory has been used as a low-cost air direction changing mechanism replacing the conventional system.

In applying the shape memory effect, it was necessary to follow the specification of the conventional system. The intended characteristics are shown in Figure 3. To satisfy these requirements with a single shape memory element, the element must provide both the temperature sensor and the actuator functions. It has been determined that two-way actuation utilizing the R-phase transformation satisfies the intended requirements.

1. Two-Way Motion and Associated Phase Transformations

A two-way actuator is required in this application. There are two types of two-way actuation: *intrinsic* two-way motion brought about by training, ageing or severe deformation, and *extrinsic* motion introduced by applying a biasing load opposite to the recovery direction. For Ni-Ti alloys, it is more common to use an extrinsic two-way motion. Schuerch¹, Cross et al², and Hsu and Wechsler³ have found that the transformation temperature and strain associated with the martensitic transformation increase with increasing biasing load. Khachin et al⁴ confirmed the possibility of

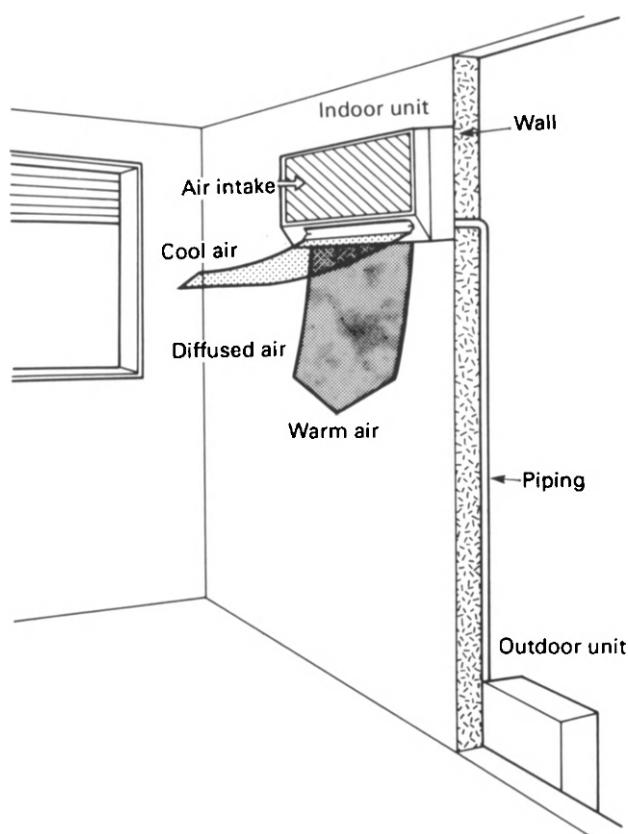


Figure 1: A typical room air conditioner in Japan.

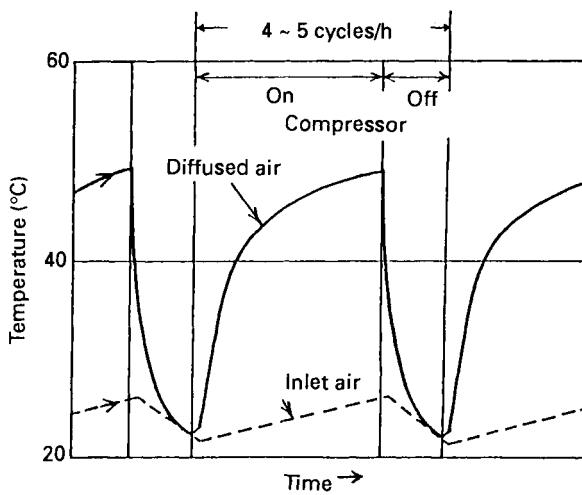


Figure 2: The variation of exit (diffused) and inlet air temperatures during operation.

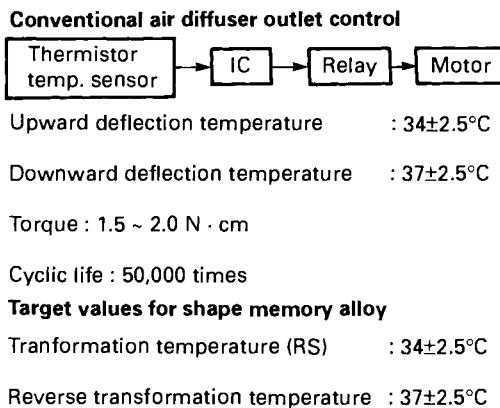


Figure 3: Specifications for the air deflector mechanism.

obtaining two-way actuation from the R-phase in a Ni-Ti-Fe alloy but did not investigate the influences of stress. Since a shape memory alloy is exposed to external stresses when in use, it is important to study these effects.

When extrinsic two-way actuation is employed, a suitable method for evaluating performance is to monitor motion, or deflection, during cooling and heating with a biasing load applied to the sample - a so called *fixed load test*. Figure 4 shows the effect of biasing load on the two-way motion of a helical Ni-Ti spring as determined by such a test^{5,6}. The hysteresis is quite small when the biasing load is small, and increases with the biasing load. The reasons for this behavior will be discussed in this chapter.

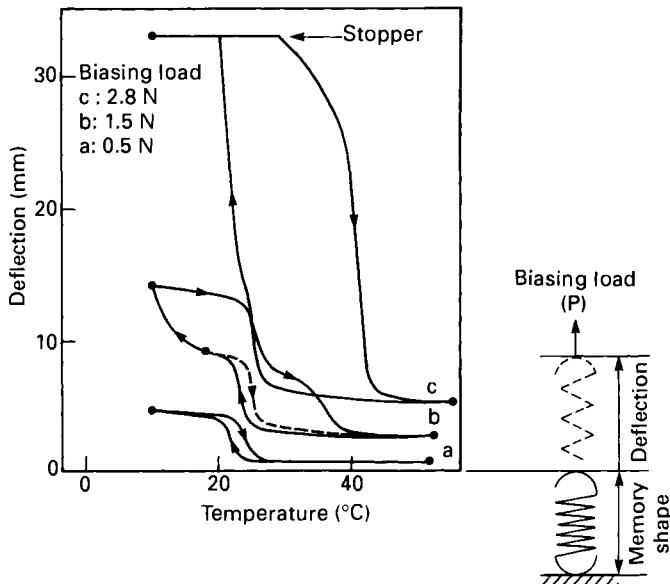


Figure 4: The effect of biasing load on the two-way motion of a Ti-50.2%Ni spring heat treated at 475°C . Dimensions for the spring were: $d=0.75\text{mm}$, $D=5.6\text{mm}$, and $n=10$.

Figure 5 shows the effect of biasing loads on motion during cooling of a 49.7 at.% Ni alloy coil heat treated at 500°C, and Figure 6 shows the associated resistivity-temperature curve (typical of a sample exhibiting the R-phase transformation⁷). The phase transformations of this composition under zero stress conditions are found by DSC to be austenite→R-phase→martensite ($B_2 \rightarrow R \rightarrow M$) in cooling and $M \rightarrow B_2$ in heating. The stresses and strains shown in Figure 5 are the maximum shear stresses

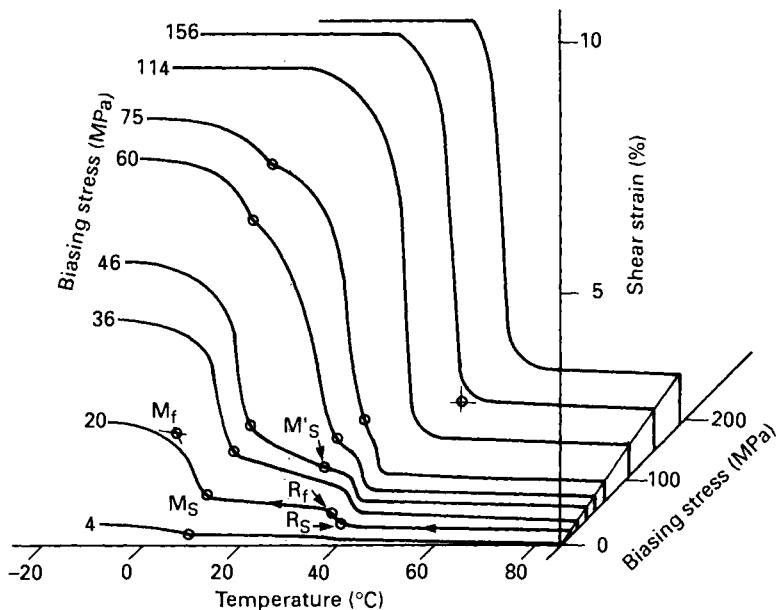


Figure 5: The effect of biasing stress on the motion of a Ti-49.7%Ni spring heat treated at 500°C⁶. For simplicity, only the motion on cooling is shown.

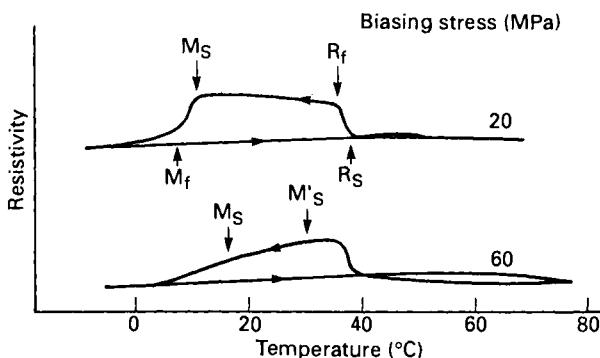


Figure 6: Electrical resistance shown as a function of temperature for the conditions outlined in Figure 5⁶.

and strains (τ and γ), calculated from the deflections and loads (δ and P) according to the usual spring design formulae:

$$\tau = 8kPD/\pi d^3 \quad (1)$$

$$\gamma = kd\delta/\pi nD^2 \quad (2)$$

$$k = C/(C-1)+1/2C \quad \text{where } C=D/d \quad (3)$$

where d is wire diameter, D is mean coil diameter, n is the number of coil windings and k is stress correction coefficient (Wood's correction). At low stresses (20 MPa) a distinct two-step behavior is shown: an initial deflection between R_s and R_f , followed by a large deformation between M_s and M_f . When the stress reaches 46 MPa, a new point of inflection ($M_{s'}$) appears. Resistivity measurements show a steady decrease in resistivity, confirming that both movements are in fact due to martensite. The strain associated with M_s -to- M_s' increases with stress, while the strain occurring between M_s' and M_f decreases. This inflection point at $M_{s'}$ is observed with coil springs but not in straight wires⁸. At stresses of 114 MPa, the deformation occurs in a single step; this is the direct transformation from austenite to martensite without the intermediate appearance of the R-phase.

The information gained by experiments such as the above give the temperature and stress boundaries for the martensite and R-phase stabilities, which in turn can be used to construct stress-temperature phase diagrams such as those shown in Figure 7. In this figure six final heat treatments of three different alloys are considered. Of course these heat treatments only have meaning after a specific cold working schedule - if one changes the amount of cold work prior to heat treatment, the effect of the given treatments would be different. The temperature range between R_s and M_s is the region in which the R-phase is found. Generally speaking, lower heat treatment temperatures and Ni-enrichment increase the range of R-phase stability. Industrially, 400-500°C heat treatments are generally optimum for producing R-phase memory. It should also be mentioned that at 400°C, the second inflection in motion at $M_{s'}$ is not observed.

A great deal can be learned by examining the schematic example of a stress-temperature phase diagram shown in Figure 8. The lines for M_s and M_f have their usual meaning, demarking the temperatures at which martensite forms upon cooling. The meaning of the A_s line, however, must be interpreted slightly differently: it corresponds to the temperature at which martensite becomes unstable - it could, upon reaching that temperature, transform to either austenite or the R-phase, depending upon one's location with respect to the R_s line. Thus it is possible to have two A_s temperatures: A_s , where the martensite becomes unstable, and A_s' where the R-phase becomes unstable.

It is perhaps useful at this point to carefully review the various transformation temperatures that have been introduced. Clearly the intervention of the R-phase complicates the simple definitions used earlier in this book.

- M_s and M_f can be understood in their simplistic forms: the temperatures at which martensite transformation starts and finishes, whether from austenite or the R-phase.
- R_s and R_f are the start and finish temperatures for the formation of the R-phase from austenite.
- A_s and A_f must be redefined as the temperatures at which reversion from martensite begins and ends - even though reversion can be to either the R-phase or austenite.
- A_s and A_f are the temperatures at which reversion from the R-phase to austenite begins and finishes (in cases where the R-phase is included in the reversion sequence). In general A_s' and A_f' are within a couple of degrees of R_f and R_s , and one need not make the distinction.

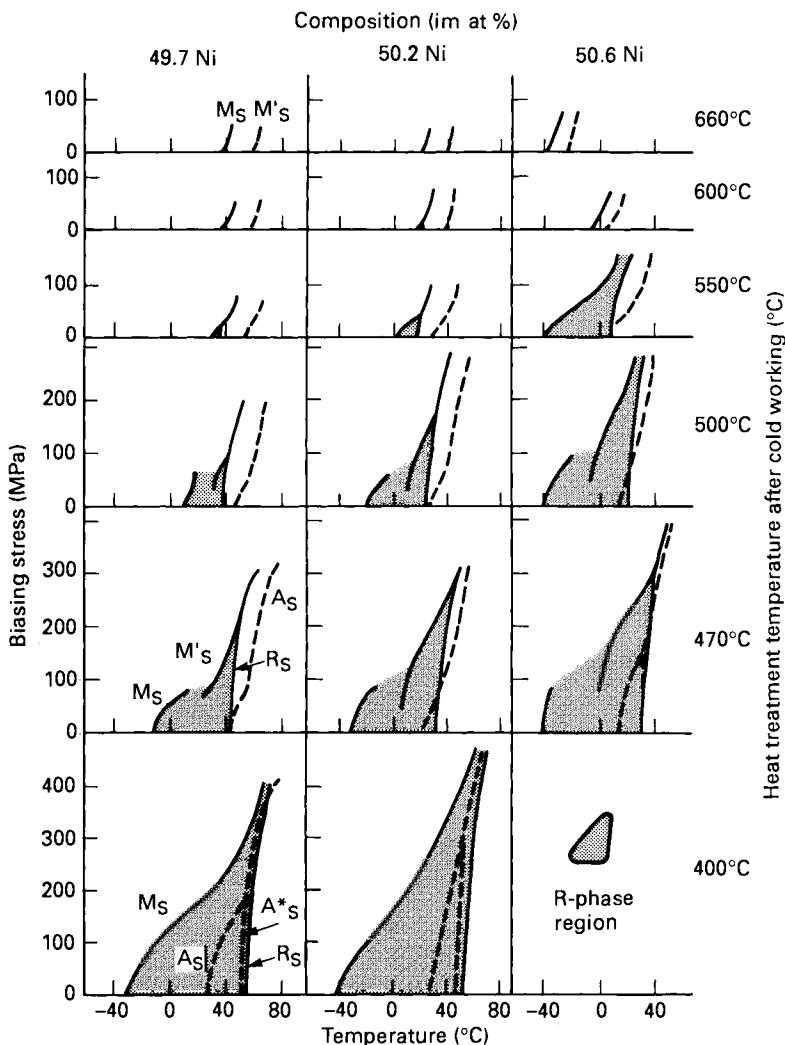


Figure 7: Stress-temperature phase diagrams for three Ni-Ti alloys after various heat treatments - a detailed description of the diagrams can be found in Figure 8.

The R-phase is formed upon heating when A_s is below A_s^* . Clearly larger stresses tend to narrow the R-phase stability region, and due to the larger hysteresis of the martensite, the R-phase field is larger in cooling than in heating.

In order to clarify the importance of diagrams such as Figure 8, it is interesting to look at a few specific cases as they would apply to actuators. At high stresses (illustrated by the trace 1-2), heating and cooling would result in a large hysteresis actuator, never entering the R-phase regime. At intermediate stresses (the trace 3-6), the R-phase transformation appears during cooling but not during heating; upon heating, one

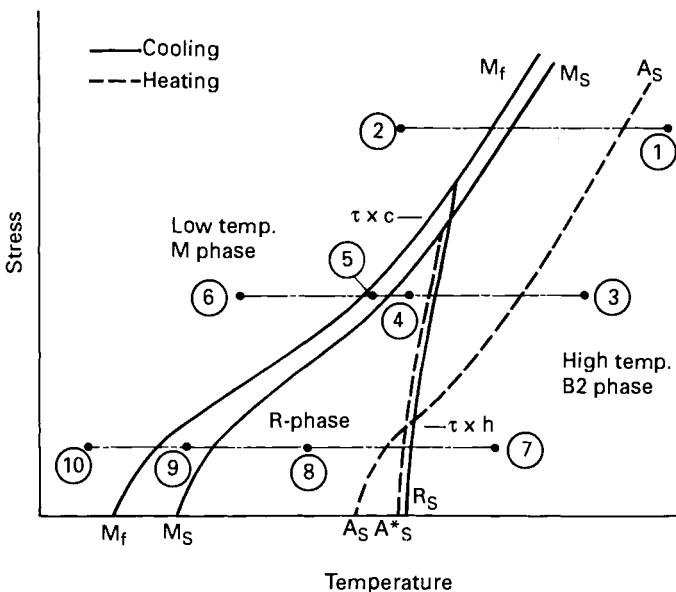


Figure 8: A typical stress-temperature phase diagram; a complete description of the three iso-stress traces is shown contained the text and in figure 9. It is important to note that the A_s line is not the stability line for austenite, but the instability line for martensite: martensite is unstable above A_s but either austenite or the R-phase can form depending upon the which side of the R_s line one is.

passes through the R-phase stability field before the martensite becomes unstable (the A_s temperature). More specifically, one finds the R-phase upon cooling to temperature 4, but upon heating to the same point, martensite is still stable and the R-phase never has a chance to form. At low stresses, again the transformation passes through the R-phase field during cooling, but this time also during heating; here A_s is below R_s , and thus the martensite loses stability while the R-phase is still the most stable alternative. The temperature range of the R-phase region is narrower at higher stresses and is larger in its extent in cooling (τ_{xc}) than in heating (τ_{xh}). As shown in Figure 7, the existence and size of R-phase region are also influenced by composition and heat treatment.

The phase diagram shown in Figure 8 is determined by measurements of the shape memory strain, and it can thus be understood in those terms. Figure 9 shows a schematic curves relating to the biased two-way memory motion associated with each of the stress levels.

- Heating and cooling between temperatures 1 and 2 corresponds to the $B2 \leftrightarrow M$ transformation and thus produces a large hysteresis. This type of deformation is always seen in samples heat treated at 550°C or above.
- Cooling from temperature 3 to 6 brings about the sequence $B2 \rightarrow R \rightarrow M$ and thus shows a two-stage movement. When heating, the martensite reverts directly to austenite since martensite stability persists to above R_s . If instead cooling is terminated at point 4, martensite is not formed and the reversion is from the R-phase to austenite and a low hysteresis is found. If cooling is continued to point 5, some martensite is formed and some R-phase is retained; in this case the

martensite will not revert to the R-phase but directly to austenite and a two step recovery is observed.

- Cooling from 7 to 10 passes follows the sequence B2 \rightarrow R \rightarrow M on cooling, and reverses the sequence during heating. In this case the low hysteresis R-phase motion is observed even though martensitic transformation is completed. This remains true even in the case of interrupted cooling to point 9.

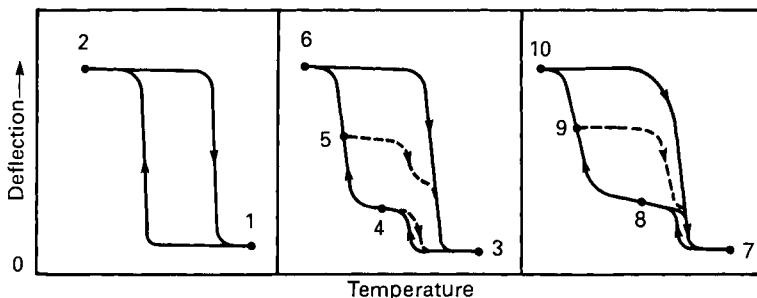


Figure 9: Two-way actuator motions are shown which correspond to the three traces shown in Figure 8.

Cooling between 3 and 4, or between 7 and 8 allows only the R \rightarrow B2 transformation, with a hysteresis of only 1.5 to 3°C, and a strain of 0.4 to 0.6%. Returning to Figure 4, curves a, b, and c are respectively the two-way transformations corresponding to 7 \leftrightarrow 8, 3 \leftrightarrow 5, 3 \leftrightarrow 6 respectively. In order to assure two-way motion due only to the low hysteresis R-phase, one must assure that the lowest service temperature remains in the R-phase region, or that A_s be below A_s^* . In order to increase the stability range of the R-phase, one must lower the heat treatment temperature. Figure 10 shows a more detailed perspective of the R-phase motion, in this case in a 50.2 at.% Ni alloy, heat treated for 60 minutes at 475°C.

Figure 11 illustrates the effects of heat treatment temperature and time on the characteristic values of Figure 10; the relationship between the transformation temperatures and heat treatment temperature is nearly linear. Hysteresis is essentially unaffected by heat treatment temperature. As the heat treatment temperature decreases, the motion associated with the R-phase becomes less sharp and the transformation strain decreases. Since the transformation temperature changes with heat treatment, the heat treatment temperature as well as the composition can be adjusted to achieve the desired transformation temperature.

2. Design of Air Direction Changing Mechanism

There are basically three ways to practically bias a Ni-Ti spring and create an extrinsic two-way motion, as shown in Figure 12. In all examples the Ni-Ti is a coil spring. For the air conditioner device, the moment method (Figure 12c) is best suited. The conventional biasing configuration shown in Figure 12b applies a continuously increasing force to the Ni-Ti spring during recovery, but the arrangement in Figure 12c essentially has a negative spring rate, applying its maximum force while the Ni-Ti spring is in its cold, or closed, configuration and its minimum force when recovery is nearing completion. Since the Ni-Ti spring exerts its greatest force when it is fully extended, this makes far better use of the memory spring output, and increases the available stroke.

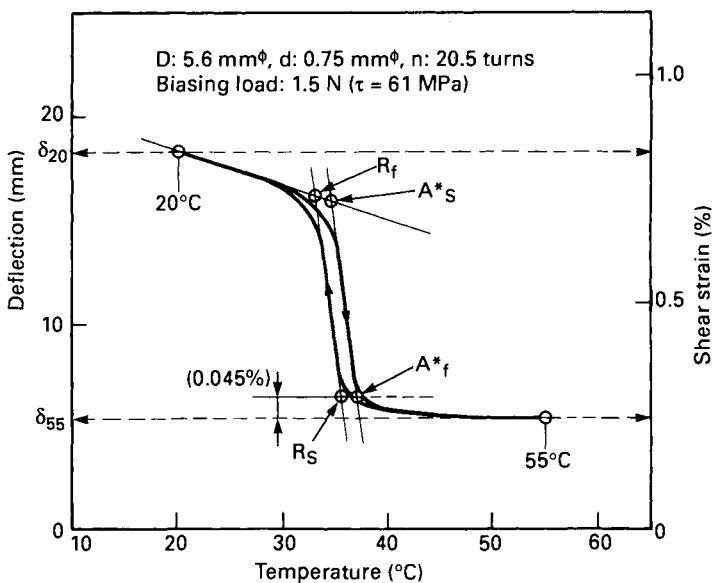


Figure 10: An expanded illustration of the performance of an actuator to use only the R-phase.

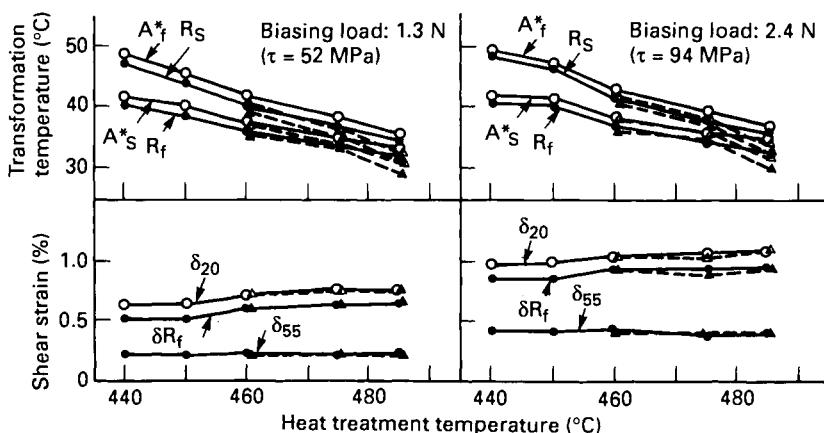


Figure 11: The effect of heat treatment temperature on actuator characteristics. The solid lines are 30 minute heat treatment times and the dashed are 60 minutes.

Figure 13 gives the specific characteristics for coil springs using the R-phase. Here, the R_f curve and the A'_f curve are the stress dependencies of strain at those temperatures as determined by the fixed load test. The difference between the two curves is the stroke of a two-way actuator. Thus as one superimposes the characteristics of the biasing spring, one obtains the anticipated end points of motion. The 0°C curve shows the stress-strain dependency at 0°: above 100 MPa, the spring yields at 0°C, indicating that martensitic transformation is occurring, and as discussed before, this must be avoided if one is to maintain the desired low hysteresis behavior. This particular data sheet can be applied for springs heat treated at 465-485°C and used at about 0°C. If the final heat treatment temperature is lower, the R_f and 0°C curves shift to smaller strains and the yield stress of the 0°C curve increases.

In summary, the intersection points of the A'_f and R_f curves with the stress-strain characteristics of the biasing spring provide the high and low temperature stresses respectively. D and d, and n must then be calculated through equations 1 and 2, so that the stress-strain coordinates correspond to the required actuator motions and loads. In case of an air conditioner flap drive mechanism the low temperature stress is 80 MPa, the high temperature stress is 20 MPa and the external force to be applied by the system is 1.5 Newtons. The Ni-Ti spring dimensions are thus calculated as: $d=.75$ mm, $D=6.5$ mm, and $n=21.5$.

The assembled actuator unit of Ni-Ti spring, biasing spring and flap is shown in Figure 14.9 When the Ni-Ti coil temperature exceeds 37°C, the Ni-Ti alloy coil contracts to shift the flaps down so that the air is diffused downward. When the diffused air temperature decreases and the Ni-Ti coil temperature becomes less than 34°C, the Ni-Ti alloy coil is expanded by the biasing spring and the flaps shift upwards to diffuse the air horizontally.

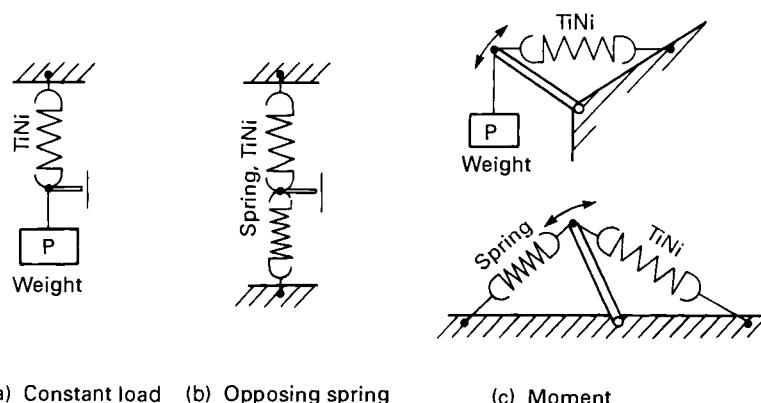


Figure 12: The three basic methods of applying a biasing load: constant load, opposing spring, and a cantilevered spring. The advantage of the last method is that the maximum biasing stress is applied as the memory springs begins to recover, opposite to what is experienced in a simple opposing spring design.

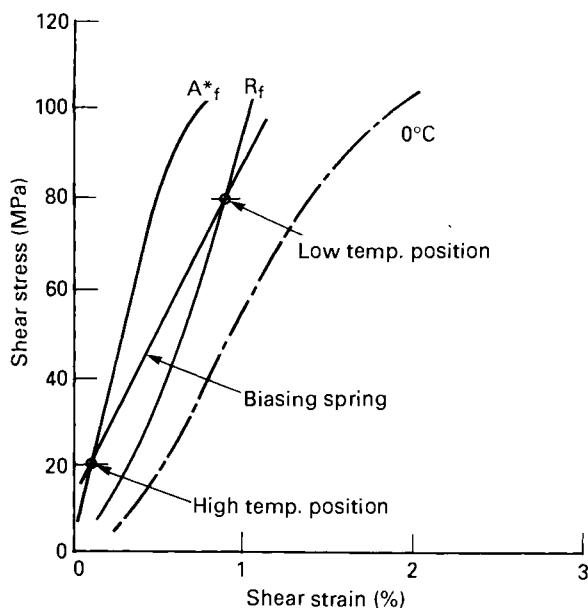


Figure 13: Stress-strain curves and for both the austenite and the R-phase. These can be used in designing a spring by superimposing the behavior of the biasing spring and extracting the low and high temperature spring positions. The 0°C curve shows that the spring yields at stresses above 100 MPa - yielding in this case, means the formation of martensite and an expansion of the hysteresis.

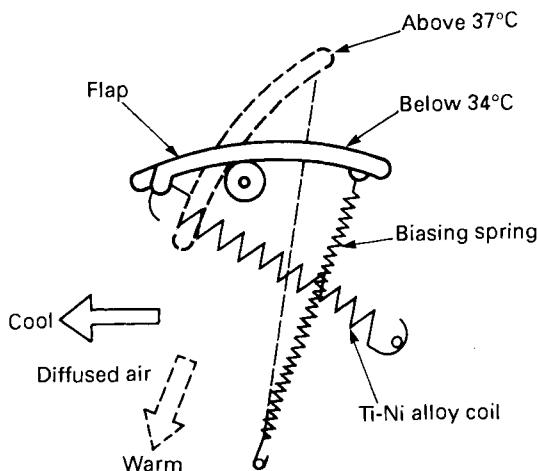


Figure 14: A schematic of the assembled R-phase actuator.

3. Reliability of Two-Way Actuation by R-phase Transformation

It is known that there is almost no change in the R-phase transformation temperature due to thermal cycling without stress¹⁰, but little is known about changes due to thermal cycling with an external stress applied. As shown in Figure 2, this particular application requires 70,000 cycles between room temperature and about 55°C (corresponding to a 10 year life). During such service, the shift in temperature should be less than 2.5°C. In order to evaluate the actuator, tests were conducted using a jig emulating the operation shown in Figure 14. The temperature extremes were controlled by 55°C water and 20°C air. The results are summarized in Figure 15: it was determined that the actuator characteristics remain nearly unchanged even after 500,000 cycles.

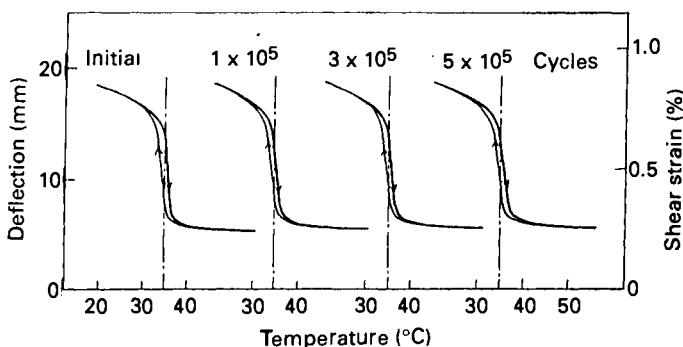


Figure 15: The characteristics of an R-phase spring are monitored during thermal cycling between 20°C and 55°C with loads of 1 N and 3 N ($\tau = 40$ MPa and 120 MPa). The sample was heat treated at 475°C.

A second mode of actuator deterioration is that of self deformation: if the memory spring is brought to a high temperature while being constrained, it can be unintentionally shape set into that configuration. In order to test these constrained ageing effects, coils were left at 70°C and 150°C (both above A_s) while being constrained in a position away from their closed configuration. This deflected strain was called $\Delta\gamma$. In addition to measuring geometrical shifts as a function of time, the springs were cycled in a fixed load test to measure their transformation temperatures. In Figure 16a, it is shown that constrained ageing at 70°C has no effect on the transformation temperatures. In Figure 16b, both the transformation temperatures and the equilibrium geometry are monitored during ageing at 150°C: one must conclude that the R-phase actuator is not suitable for prolonged use at 150°C when accurate performance is required.

4. Features of Air direction changing Mechanism

Figure 17 shows an air conditioning unit with a Ni-Ti coil employed in the air direction changing mechanism. The Ni-Ti coil is located on the right-hand side of the air outlet, and the bias spring is installed on the left side. The relationship between the transformation temperature, the diffused air temperature and flap operation are shown in Figure 18: the linear relationship makes it quite easy to adjust the flap operation

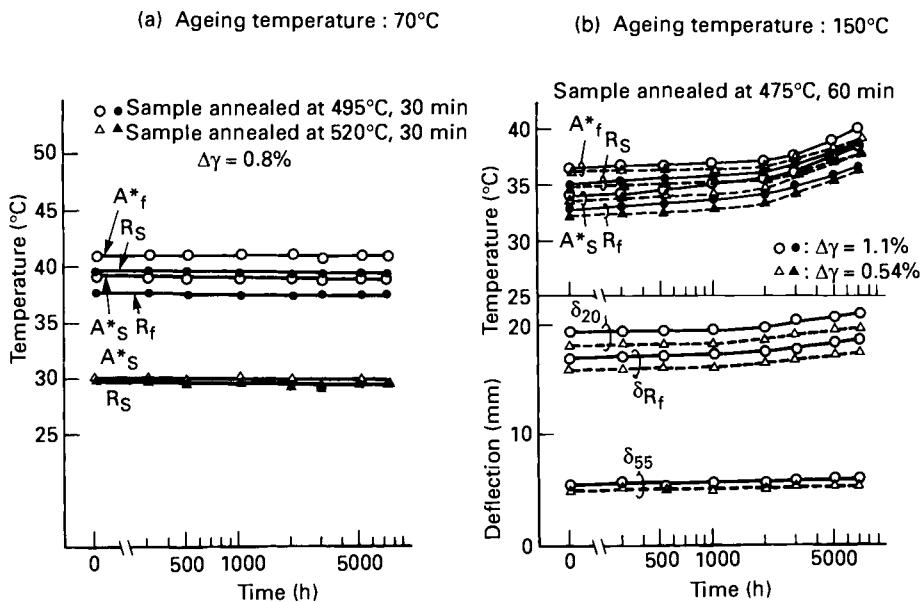


Figure 16: The effects of constrained ageing on the characteristics of an R-phase actuator: The curves on the left correspond to ageing above A_s but below M_d , while those on the right are above M_d .

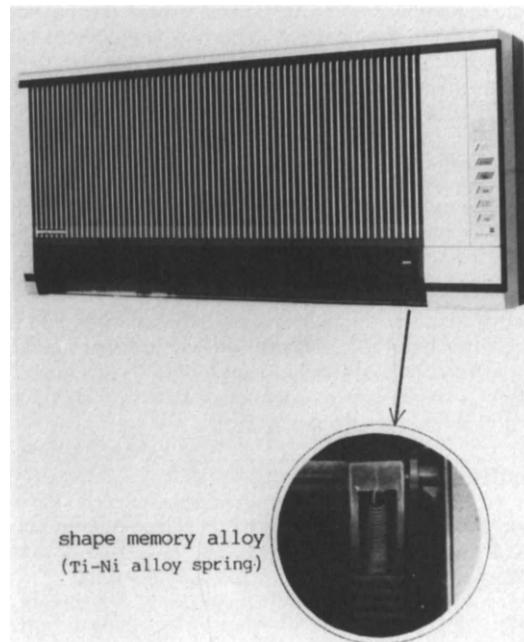


Figure 17: A photograph of the complete air conditioner unit: the inset shows the shape memory spring itself.

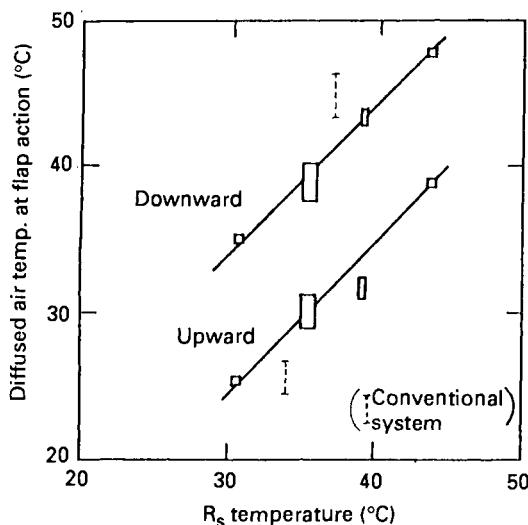


Figure 18: The relation between transformation temperature, the diffused air temperature, and flap motion.⁹

temperature according to the transformation temperature of Ni-Ti alloy used. The hysteresis of two-way motion due to R-phase transformation is about 1.5°C, while the hysteresis of the flap operation is about 10°C. This is because it is not the temperature of diffused air that controls movement but rather the temperature of Ni-Ti alloy itself, which changes in a more sluggish fashion.

As compared with the conventional system, the advantages of the shape memory sensor/actuator are as follows:

1. Since the shape memory alloy serves both as sensor and actuator, fewer parts are required, and their size and weight are reduced.
2. The Ti-Ni alloy exhibits excellent corrosion resistance and does not require a protective coating; consequently it is able to respond more quickly to changes in the diffused air temperature.
3. The operation is quieter since no motors are used.

Air conditioners using shape memory in the air direction changing mechanism were first put on sale in September 1983, and over one million sets have been sold so far. This is only the beginning of a widespread usage of Ni-Ti actuators in consumer and household appliances. A variety of products using the highly reliable R-phase transformation have now been developed in Japan.

Acknowledgement:

The author is grateful to Mr. K. Fukuda and Mr.T. Hayakumo of Air-conditioner Division, Matsushita Electric Industrial Co. Ltd., and to Dr. Y. Suzuki and Mr. H. Tamura of Furukawa Electric Company, for their cooperation with this work.

References:

1. Schuerch, H.U.: 'Certain Physical Properties and Application of Nitinol', NASA CR-1232. (1968).

2. Cross, W.B., Kariotis, A.H., and Stimler, F.J. 'Nitinol Characterization Study', NASA CR-1433, (1969).
3. Hsu, C.H. and Wechsler, M.S.: 'The Effect of Applied Stress on the Transformation Behavior of Ti-Ni', Proc. Int. Conf. Solid-Solid Phase Transformation, 1293 (1981).
4. Khachin, et. al.: Phys. Met. Metall. 46(3), 49 (1979).
5. Todoroki, T.: J. Japan Inst. Metals. 49(6), 439 (1985).
6. Todoroki, T. and Tamura, H.: J. Japan Inst. Metals. 50(6), 538 (1986).
7. Dautovich, D.P. and Purdy, G.R.: Canadian Metall. Quart. 4(2), 129 (1965).
8. Todoroki, T. and Tamura H.: J. Japan Inst. Metals. 50(6), 546 (1986).
9. Todoroki, T., Fukuda, K. and Hayakumo, T.: Kogyo-zairyo. 32(7), 85 (1984).
10. Miyazaki, S., Igo, Y. and Otsuka, K.: Acta Metall. 32(10), 2045 (1986).

Shape Memory Actuators in Circuit Breakers

Erhard Runtzsch
Asea Brown Boveri AG
Heidelberg, West Germany

Electrical appliances must be protected against overheating. Overheating can be caused by short-circuits or by overload. It is the purpose of the circuit breaker to switch off the current instantaneously in the case of short-circuits or after a specific period of time in the case of overload. Thus the current/time characteristic of the circuit breaker must be adjusted according to the protective switching application.

One of the most important applications of circuit breakers is the protection of electric lines and cables against overheating according to IEC 364-4-43. The insulation of wires and cables is rated for a specific constant operating temperature and a short-time over-temperature. This can be converted into a load-limit-curve for a particular wire or cable, as shown in Figure 1, which cannot be exceeded, either during short

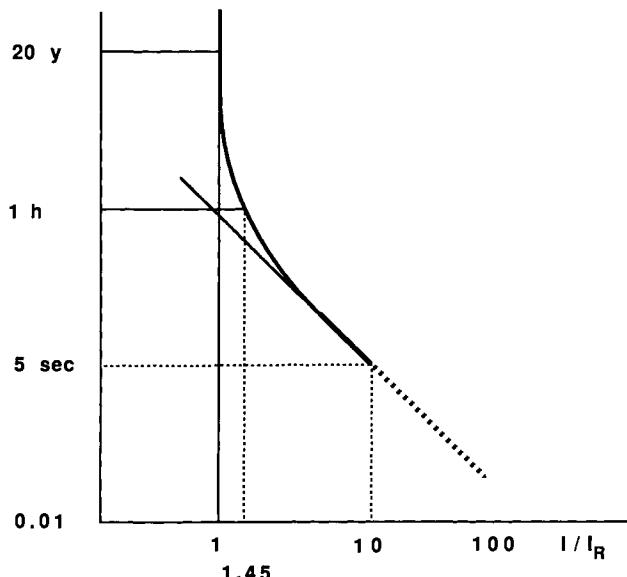


Figure 1: Load limit curve for insulated wires.

circuit or during overload. This means that the current has to be switched off before the wire reaches the critical temperature. This can be done either by fuses or reversibly by automatic circuit breakers.

Circuit breakers are thus subject to two types of tripping events: short circuits and steady-state overloads. Circuit breakers usually switch off short-circuits instantaneously by means of magnetic tripping. Overload tripping on the other hand is achieved by time-delaying, temperature-dependent thermobimetal actuators. Thermobimetal actuators are wired into the main circuit of the circuit breaker and are passed directly by the operating current. (Some designs heat the thermobimetal indirectly by passing the current through a nearby stainless steel plate, but the principles are the same). The operating current causes the temperature of the actuator to increase and the actuator to bend. Upon reaching a predetermined temperature, and thus deflection, a latch is released. The tripped switch mechanism opens the contacts and interrupts the overcurrent.

Ever since the discovery of shape memory, there has been a great deal of interest in shape memory circuit breakers, with projects at nearly all the major circuit breaker manufacturers. Prototypes have been made and tested using Ni-Ti, Cu-Zn-Al, and Cu-Al-Ni switching elements. This paper is directed at one circuit breaker development program, in which Cu-Al-Ni was considered to be the most appropriate SMA element material.

1. Requirements:

The allowable constant operating current is determined by the allowable operating temperature. The current-carrying capacity of wires and cables is given by the ambient temperature, the cross-section, the material and the lay-out (bundling etc.). Insulated wires can carry an overload for a limited time. An international standard calls out an overload of 45% being carried for one hour maximum. Temperatures of about 140°C can typically be reached before it is necessary to interrupt the circuit. The tripping characteristic of the circuit breaker corresponds to the load limit curve of the insulated wire. As shown in Figure 2, the circuit breaker should carry an overcurrent of $1.13 \times I_r$ (I_r being the rated current) without tripping. It should trip, however, after one hour at a current of $1.45 \times I_r$ and after one minute at $2.55 \times I_r$. At $3-5 \times I_r$ magnetic tripping occurs instantaneously.

Actuators in circuit breakers have requirements for the following characteristics:

- tripping temperature
- overload stability
- short circuit stability
- total deflection

As already mentioned cable temperatures can be as high as 140°C. Therefore, the tripping temperature of the actuator cannot be lower than 150°C. This all but rules out Ni-Ti.

Actuators in circuit breakers have to carry currents up to $5 \times I_r$ (before magnetic tripping) for short periods of time (10 sec). Temperatures up to 200°C can thus be reached. The overload stability of actuators is tested by cyclically loading the actuator 1000 times with $2 \times I_r$ for 70 seconds per cycle and a mechanical force of 0.3N. No permanent shape change of the actuator is allowed after this procedure.

Circuit breakers also have to handle short-circuit currents of more than 10000 Amps, which are limited to 5% of I^2xt . However, the actuator still has to be able to carry a

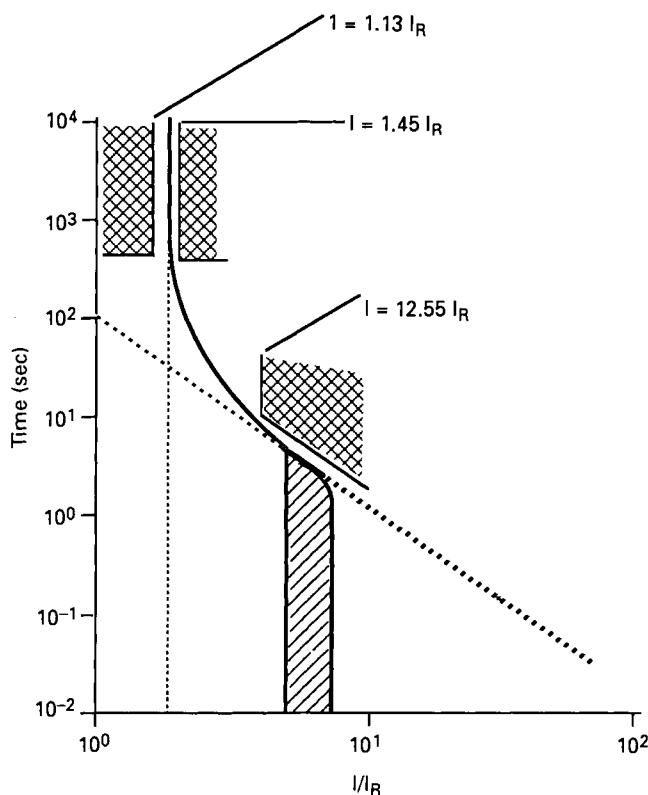


Figure 2: Tripping characteristic of a circuit breaker.

tripping load I^2xt of about 50000 Amp²-sec without any significant change in performance. An "operating constant C_o "

$$C_o = \int I^2 dt / F^2 \quad (1)$$

with I = effective current

F = cross-sectional area of the wire

is considered to be necessary for sufficient stability during service life. Test requirements are 10 short circuit cycles, after which the tripping characteristics of the circuit breaker are permitted to change by no more than +/- 10%. Note that in this case tripping is magnetic, and thus the SMA is exposed to very high temperatures but without a superimposed load. Even so, the heat that is generated during these cycles severely overheats Cu-Zn-Al alloys, making them unsuitable for this application.

2. Actuators for Circuit Breakers:

Thermobimetal actuators basically meet all the requirements for use in a circuit breaker. However, because of the proportional deflection/ temperature characteristics

of thermobimetals the actuator reaches the tripping point in a creeping mode and any change in ambient temperature causes a shape change of the actuator (Figure 3). Tolerances in force and deflection can thereby cause large variations in tripping times. Thus, a rather sophisticated mechanical latching system is required and every system has to be adjusted individually.

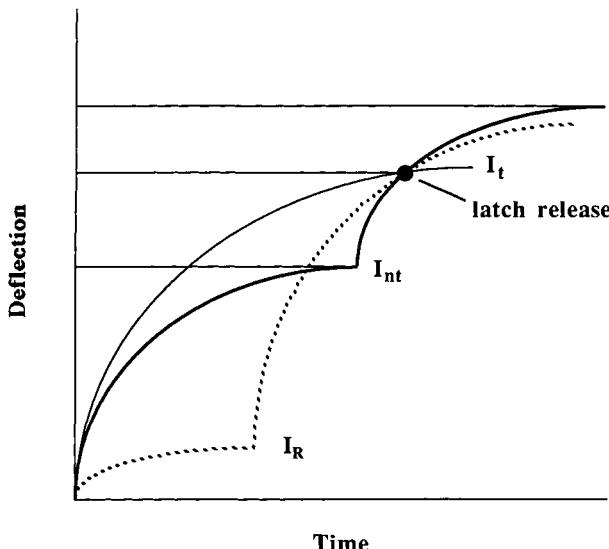


Figure 3: Tripping characteristic of a Thermo-Bimetal Actuator.

Shape memory actuators on the other hand offer large motions (deflections) in a very narrow temperature range and therefore don't have to be temperature compensated. Tolerances in force and deflection have practically no effect on the tripping time (Figure 4). The only practical shape memory alloy which meets the requirements with regard to tripping temperature is Cu-Al-Ni. Therefore a Cu-Al-Ni alloy with a *switching temperature* of 170°C was chosen as an actuator alloy for a 20 Amp circuit breaker. By switching temperature (T_s) we refer to the temperature at which a certain deflection is achieved (obviously between A_s and A_f). Elements with 1.7mm x 1.6mm x 35mm were cut out of sheet material and a two-way effect with a total deflection of 3mm at the free end (nearly 2%) was induced into the elements via a constrained ageing treatment¹.

3. Performance of the Elements:

One of the first problems that was encountered was that Cu-Al-Ni is inherently quite brittle, and that ultra-fine grain sizes are needed in order to obtain a useful material. This problem was solved by going to a powder processing route, which reduced the average grain size from 150 microns to 15 microns, and increased ductility from 1% to over 6%.

The performance (deflection vs. temperature and load) of these elements was tested using a set-up as shown in Figure 5. The transformation temperatures and deflections were quickly determined to be adequate for this application. It became apparent that

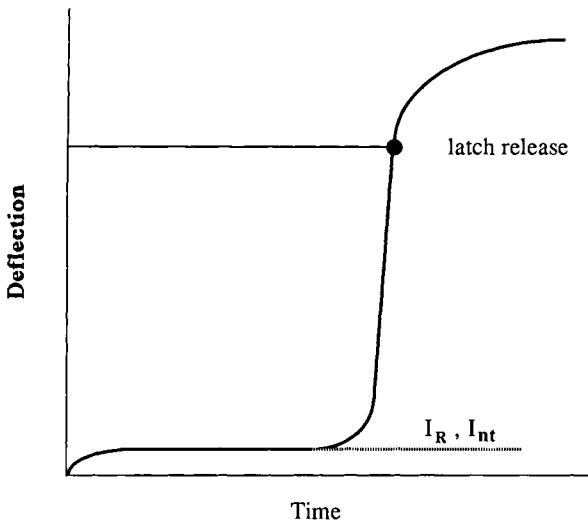


Figure 4: Tripping characteristic of a Shape Memory Actuator.

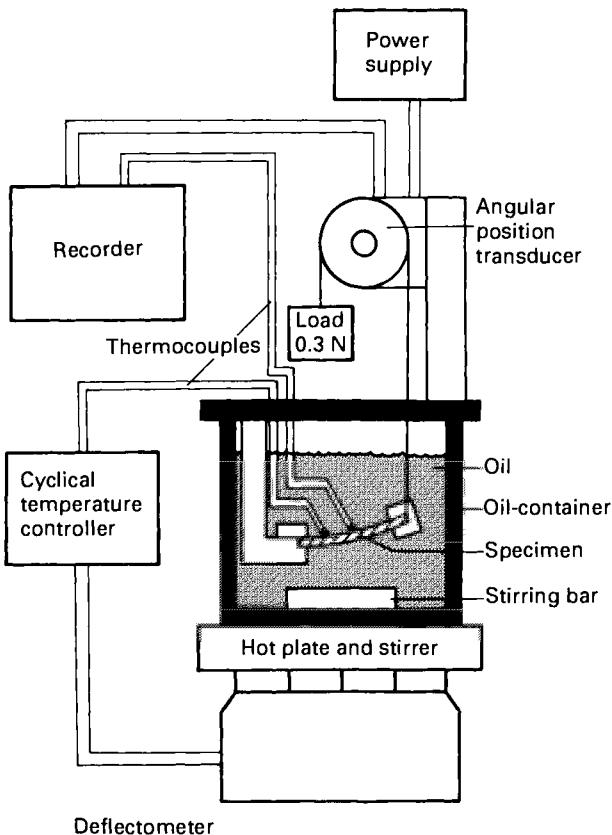


Figure 5: Equipment used to measure shape memory element performance.

the critical issue was, in fact, thermal stability. Thermal stability of the material was measured by immersing the elements in molten salt for various times and temperatures and then testing in the set-up shown in Figure 5. Three types of degradation were observed: reduction in the two-way motion, change in T_s , and walking (or shift in the high temperature shape). The requirement was that a one hour exposure at 300°C would result in no more than a 25% reduction in the two-way strain, a 10% shift in A_t , and that the zero point strain would shift by no more than 7.5%. The origin of these requirements is complex and will not be treated here, except to say that the 300°C and 1 hour figures were anticipated to be comparable to the complex overheating events described earlier.

Figure 6 shows the results of the 300°C exposure test. All three parameters are plotted as a function of time. In fact one hour represented the actual performance limit for this material. Note that at 300°C, it is T_s that first becomes out of spec. Tests were conducted at other temperatures, allowing the construction of a heat ageing survival curve (Figure 7). What is interesting is that different attributes are first to become critical depending upon the exposure temperature.

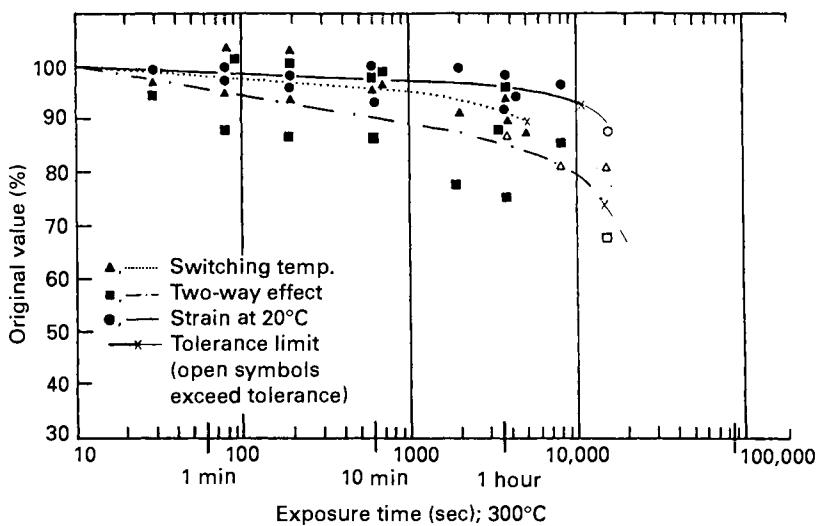


Figure 6: Stability during 300°C exposure without load.

4. Performance of the Elements in Circuit Breakers:

Circuit breakers with these elements as actuators were built and tested (Figure 8). The circuit breakers did, in fact, show the required tripping characteristics. However, although the thermal stability tests showed satisfying results, short-circuit stability in the actual breakers did not quite meet the requirements. It appears that the dynamic effect of short-circuit electrical heating stresses the material more than the 300°C static test, which is supposed to simulate the short-circuit condition. Further work is necessary to understand the dynamic heating effects.

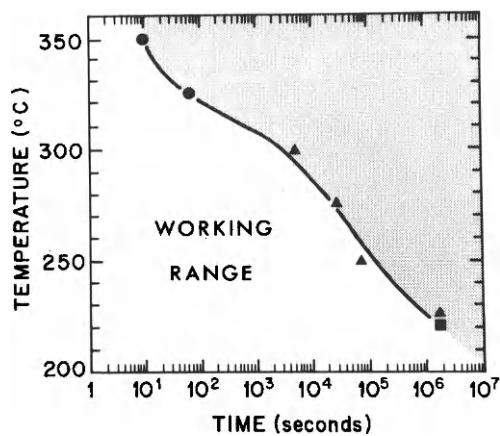


Figure 7: Ageing stability envelope, with the switching temperature (Δ), two-way strain (■) and transformation temperature (●) marked according to which "failed" first.

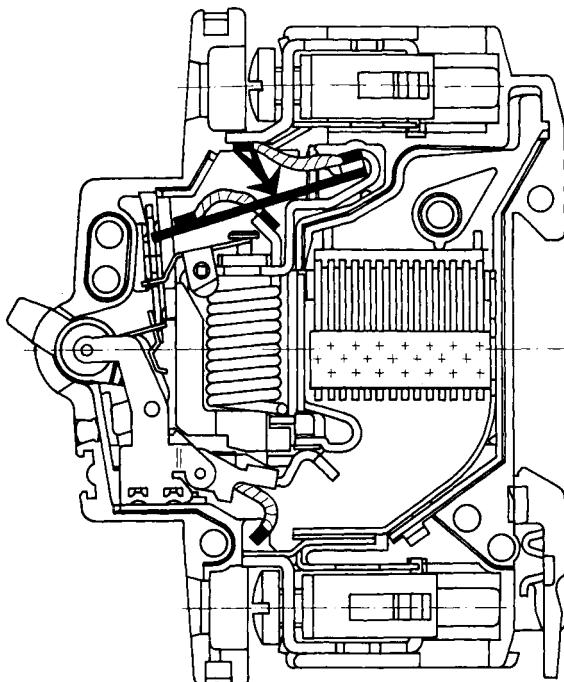


Figure 8: Typical circuit breaker: the arrow points to the location of the SMA element.

If the stability of the alloy can be improved and some of the difficulties associated with powder processing eliminated, Cu-Al-Ni shape memory actuators could be successfully used in circuit breakers. The benefit would be simplification of the design and assembly process. A 100,000,000 parts per year market (in Europe alone) is waiting.

Reference:

1. T.W. Duerig, J. Albrecht and G.H. Gessinger: J. Metals page 12 (December 1982).

Shape Memory Actuators for Robotic Applications

Y. Furuya and H. Shimada
Materials Processing Engineering
Tohoku University
Sendai 980, JAPAN

Shape memory alloys represent a new class of material, capable of transforming thermal energy into mechanical work. If the heating and cooling of these alloys is controlled by pulsed direct electrical current, repeated cyclic motions can be achieved and used in robotic applications as thermal-mechanical actuators.

1. Principles of Actuators:

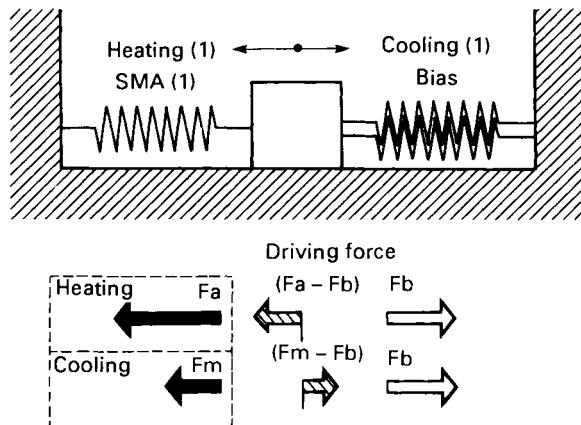
Electrically driven SMA actuators can be generally divided into two groups: biased and differential. Figure 1a shows the principle of a *biased actuator*:^{1,2} the SMA spring applies a force- F_a upon heating and a force F_m upon cooling, while the biasing spring applies a temperature independent force, F_b . The driving force of the actuator device is the difference between these two forces: $F_a - F_b$ upon heating and $F_b - F_m$ upon cooling. In this case the net force is asymmetric (different in the two movement directions). In the case of the *differential actuator* (Figure 1b) two SMA springs are used and alternately activated. Here the performance is symmetrical, with a net driving force of $F_a - F_m$ in both directions. Generally speaking the net force of the differential type is two times that of the biased.

Figure 2 shows the principles of biased actuator design in more detail. Radiating from the origin, the force-displacement characteristics of the SMA spring in both the cold and warm state are shown. Superimposed upon this are the biasing spring characteristics (in this case both a linear and non-linear spring are shown). The distance the actuator can move is thus controlled by the points at which the biasing spring curve is intersected by the martensitic and austenitic SMA spring curves. For obvious reasons, non-linear biasing springs are generally preferred for large displacement actuators. Figure 3 shows how this type of actuator can be used to affect rotation in a robotic joint.

1.1 Control Systems:

Figure 4 shows a comparison of control logic between actuators using conventional devices such as motors or hydraulics, to those employing an SMA spring.³ In the sequential control systems, the SMA spring is able to combine the functions of gears and linkages of drive motors and those of the control system. In feedback control systems, the change in resistivity observed as the SMA goes through its transformation is used to provide positional feedback, thus acting as an actuator and position detector. The resulting simplifications in design and the fact that the SMA device has no rotating or sliding parts, make super-miniturization and integration possible.

(a) Bias type



(b) Differential type

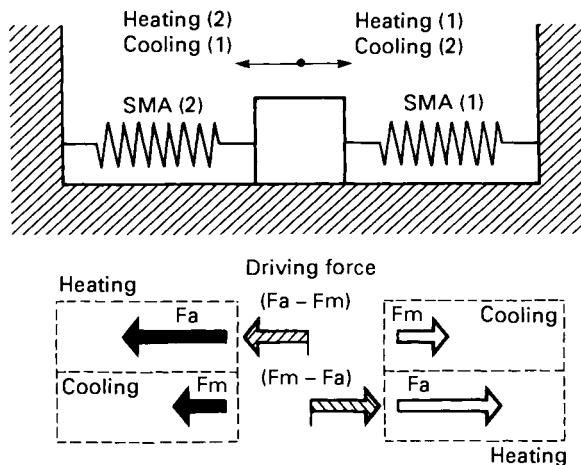


Figure 1: The fundamental ways of applying shape memory springs in the construction of electrical actuators: biasing and differential.

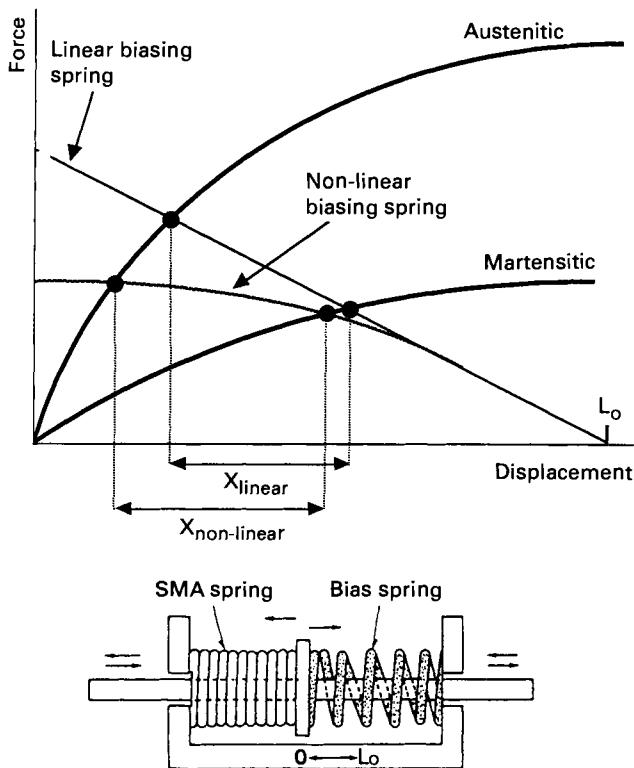


Figure 2: Force-displacement curves such as those shown above can be used to determine the characteristics of biased actuators: the intersection points of the austenitic and martensitic force-displacement curves with the characteristics of the biasing spring (whether linear or non-linear) give the total displacement expected. Changing the prestrain (L_0) of the biasing stiffness can have large influences upon the stroke of the actuator.

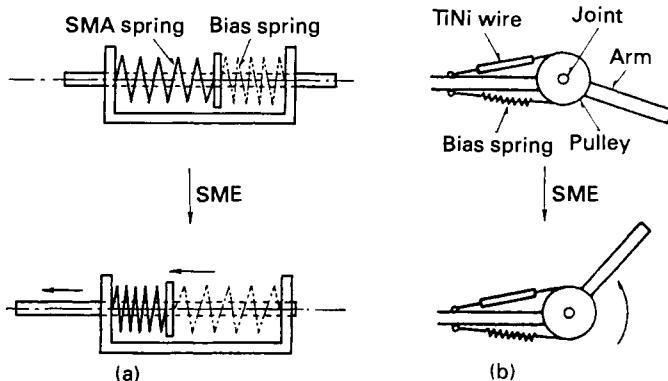


Figure 3: Linear (a) and rotational (b) manifestations of a biased SMA actuator shown in both hot and cold configurations.

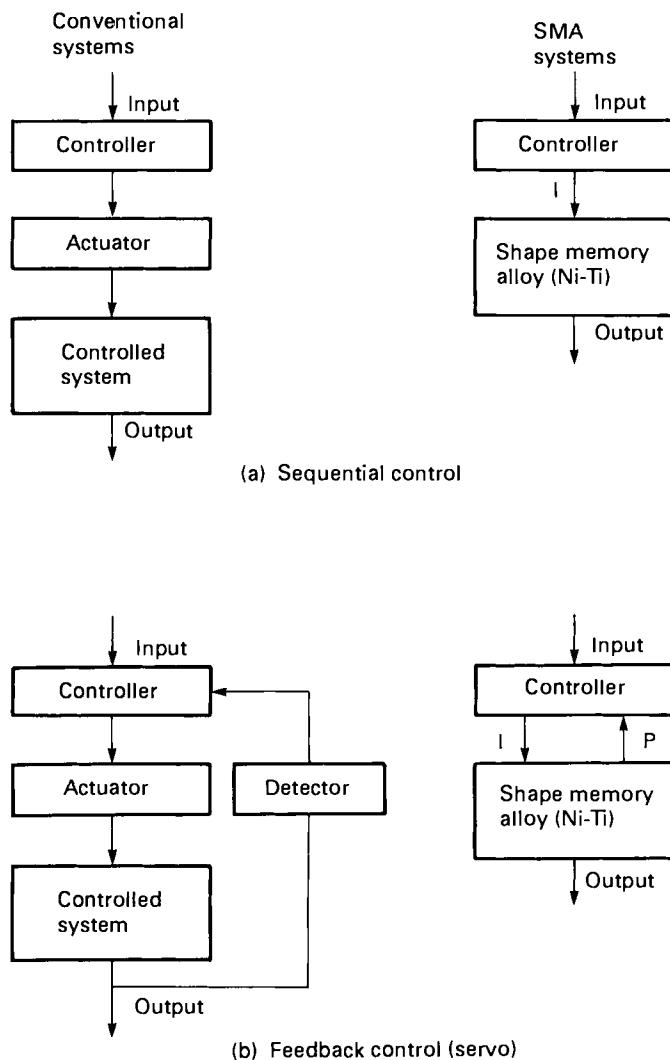


Figure 4: Shape memory systems show significant advantages for both sequential and feedback controllers, in both cases providing simplification through combining the actuation and control functions.

Figure 5 shows an example of how sequential control can be used to maintain a specific position. This is accomplished via pulse width modulation (PWM), a name coined by Miwa¹. Direct current pulses of specific frequencies, amplitude and duration are used, leading to the definition of a "duty ratio" (D), defined as the ratio of "on" time to the time between pulses (P_1). Figure 6 shows the dependence of motion on the duty ratio. It should also be pointed out that the recovery stress increases with the heating rate due to adiabatic heating. It is for this reason that pulsed direct current is generally preferred for actuators.

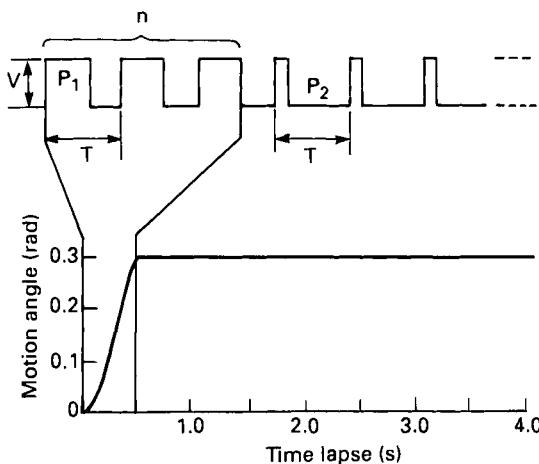


Figure 5: Sequential control through shape memory can be accomplished by controlling the pulse characteristics to control the angle of motion. Here two duty ratios are used, a P_1 of 0.5 to bring the angle to the desired value, and a P_2 of 0.15 to maintain that position.

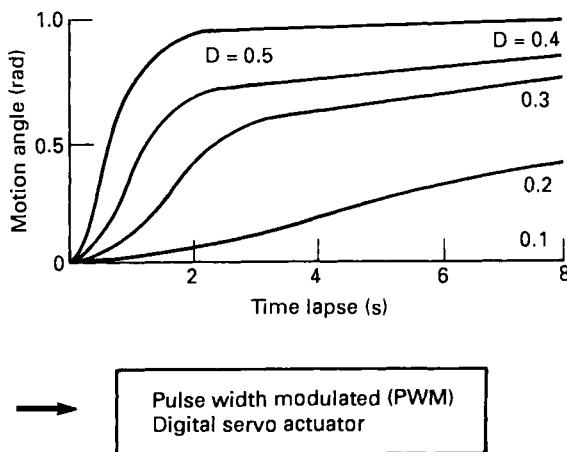


Figure 6: A systematic relationship exists between motion and duty ratio.

1.2 Response Time:

The response speed of an actuator is controlled by the cooling; by increasing the current density one can easily obtain rapid motion on the heating cycle, but cooling can only be affected by thermal conduction from the surface. Thus from an engineering point of view, measures generally have to be taken to increase the rate of cooling.²

The most commonly used methods of increasing the rate of heat loss are to increase the surface area to volume ratio (via films, ribbons, and the like) or by bonding a heat-sinking material to the wire (such as a silicone rubber or copper). Considerable improvements in response time can be achieved by combining these two methods, but it can often be difficult to maintain a high integrity bond. The fastest times are achieved by using water as an operating environment.¹ One should be aware, however, that the energy consumption of an actuator device inevitably increases as these methods to reduce response time are employed.

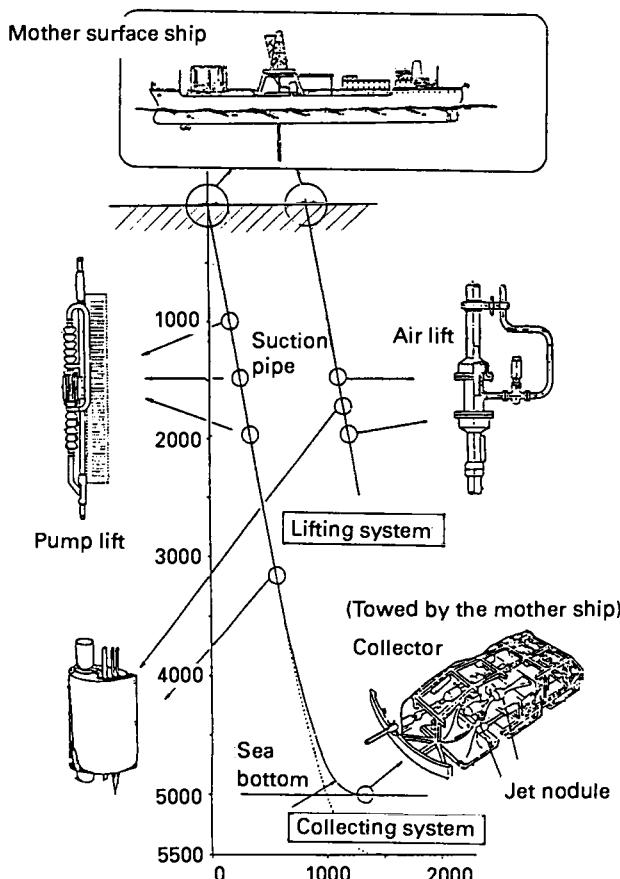


Figure 7: The basic scheme for undersea nodule mining involves a surface ship, a collection system, and a lifting system.

2. SMA Robotics for the Development of Submarine Resources:

As a specific example of a robotic actuator, the development of a submarine robotic device⁴ will be described in some detail. Robotic devices using Ni-Ti are much simpler in operation than those using motors or hydraulics; moreover, they are characterized by high power/weight ratios, corrosion resistance and strength. Thus they are particularly well suited for operation in extreme environments such as might be found in space, radioactive environments and in the deep sea; in these cases the control and maintenance of motors and hydraulic units is very difficult (largely due to corrosion effects). The specific case described here is designed to operate undersea, to investigate and collect resources such as manganese nodules by remote control from a mother ship.

2.1 The Basic Development Concept:

Figure 7 shows a schematic view of the manganese nodule mining system now used by the Japanese industrial agency (MITI) since 1981. The system consists of three main components: a collection system on the ocean bottom which uses a jet module towed by the mother ship, a lifting system which employs cable and a suction pipe, and the control system based on the mother ship. The overall arrangement, however, has several weaknesses:

- (a) The central control system is complex, and any failure causes an entire system shutdown.
- (b) The collector is towed by cables and can be easily overturned on rugged ocean bed terrain.
- (c) Jet mining causes environmental pollution of the ocean bed.

To address these weaknesses an independent "walking" SMA robot capable of examining the seabed has been proposed.

The original design for the SMA robot⁶ is shown in Figure 8. Externally, it is designed to imitate a crab, which appears to be the ideal configuration for movement on land and the sea floor. The use of SMA actuators in the robotic crab enables each of the joints to move as a real crab would. The "muscles" of the robotic crab thus consist of SMA actuators heated electrically and cooled by the surrounding water itself. By using a microcomputer to control electrical pulses, the robot can move smoothly, almost biologically, on the sea floor.

2.2 Joint Characterization:

A model of a single leg was made to investigate its basic dynamic characteristics (Figure 9). Ni-Ti wire of 0.4 mm diameter was stretched over a pulley 20mm diameter and fastened to the end of the leg. Thus small elongations of the Ni-Ti wire are able to effect large angular displacements. The "knee joint" of the leg is opened by the recovery (shrinkage) of the SMA wire. The return of the knee joint to the bent configuration is brought about by the biasing springs once the current is removed and the SMA wire is cooled back to the martensitic state. The torque generated by the actuator mechanism (T) was measured at different joint angles by the arrangement shown in Figure 9. The motion and response times were measured via high speed photography.

2.3 Shape Memory Characteristics:

The relationship between the torque applied by the SMA arm and the joint angle is shown in Figure 10. Torque decreases in inverse proportion to the joint angle. Both torque and the angular motion of the arm increase as the prestrain in the wire is

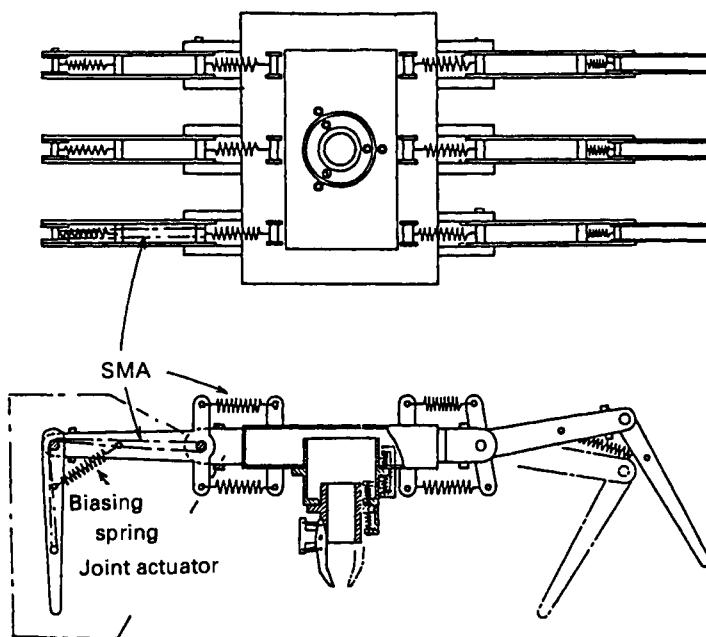


Figure 8: The basic scheme for the SMA undersea robot resembles a crab, with six legs, each with two joints.

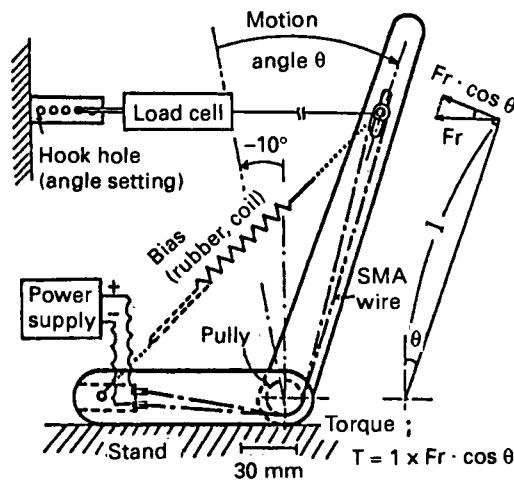


Figure 9: The basic SMA elbow joint consists of a SMA wire circling a pulley, and a biasing spring.

increased. Figure 10 also shows that a rubber-like biasing spring with nonlinear elastic behavior is the most effective for a biasing spring. An additional factor affecting recovery force is the heating rate. Figure 11 shows that larger forces can be achieved by more rapidly heating the wire. This may be caused by the exothermic nature of the transformation.

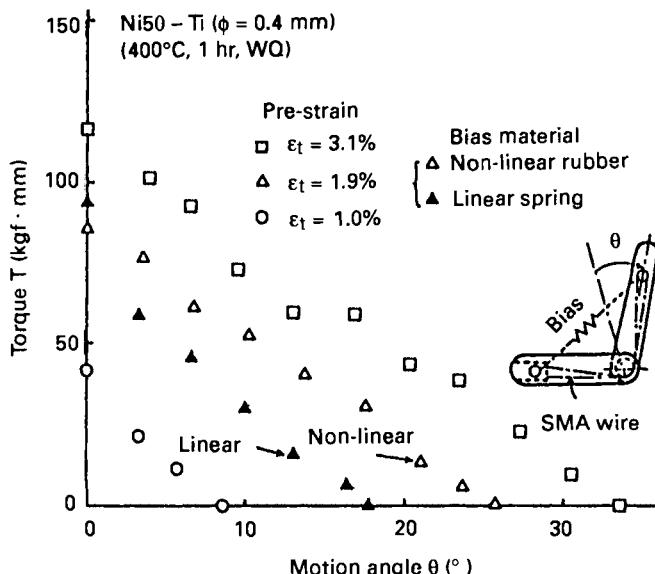


Figure 10: The Torque generated by the SMA joint is strongly dependent upon the angle of the arm, the nature of the biasing spring, and the prestrain.

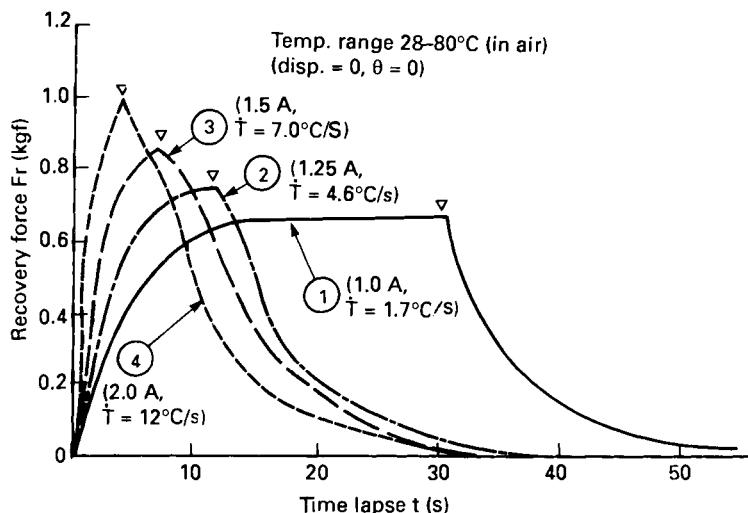


Figure 11: Recovery stress increases with the speed of recovery. The inverted triangle marks where current was shut-off.

As discussed earlier the cooling response of an actuator immersed in water should be dramatically better than in air. Measurement of the "crab" joint confirm this (Figure 12).

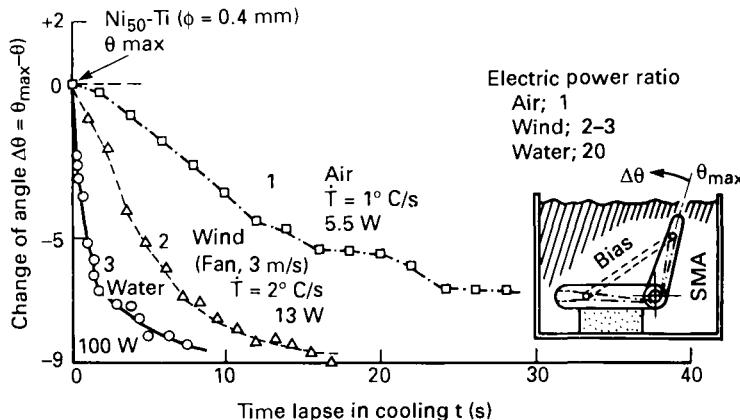


Figure 12: The response speed of the joint actuator in three environments, verifying the large improvements one obtains from a water heatsink.

But as was also pointed out earlier, a significant penalty must be paid in terms of the electrical power required for heating. In this case a tenfold increase in reset speed is penalized by a 20 fold increase in power requirements.

2.4 Degradation Due to Fatigue:

SMA robotic actuators are generally expected to function consistently over a great number of cycles, and thus the degradation of the shape memory during cycling must be examined. In the case of the crab joint, it is found that the range of motion, ΔD , decreases rapidly during cycling, with most the degradation occurring during the first 10 to 20 percent of the total life (Figure 13). Actual failures were found to occur most frequently at the pulleys. Improvements in cyclic performance can be brought about by a short training cycle and by decreasing the prestrain in the wire.¹⁰

2.5 Movement of the Crab Robot:

Three models of the robotic crab have been developed since 1985. The first, Figure 14a, was constructed using SMA coil springs, the second and third versions could be actuated remotely using a microcomputer. The third model (Figure 14c) is equipped with a micro-video camera and can move and search under water. Figure 15 shows the general form of the robot, though as yet the equipment for collecting and sucking nodules has not been included in the actual robotic device. The joints at the root of each leg are designed to minimize cyclic reduction in driving force and rotate in a semi-circular motion. Pulses of current are applied via a microcomputer in order to cause a walking motion with three legs contacting ground at any given time (Figure 16). All SMA springs are enveloped by fiberglass pipes in order to avoid excessive heat loss from the memory spring and to reduce power consumption. The crab moves at a speed of 7.5 cm/sec and the power used is approximately 10 kW in water and 1 kW in air.

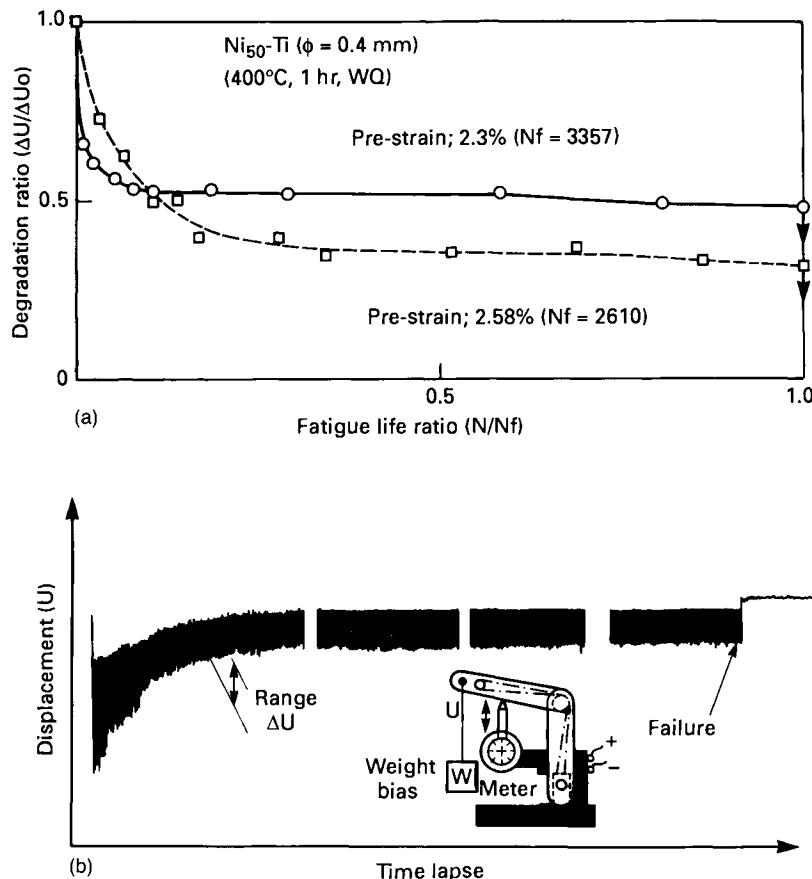


Figure 13: Testing shows that both the actuator stroke and positional range are affected by cycling (bottom). Failure in this case occurs in thousands of cycles with strains of 2.5%, but lifetimes can be extended beyond a million by reducing the strain to under 1%. By normalizing the positional shifts with cycling to the total fatigue life (above) one sees that most drift occurs during the first 15-20% of life. (ΔU = the cyclic range of knee joint motion, and ΔU_0 = the range of motion during the first cycle.)

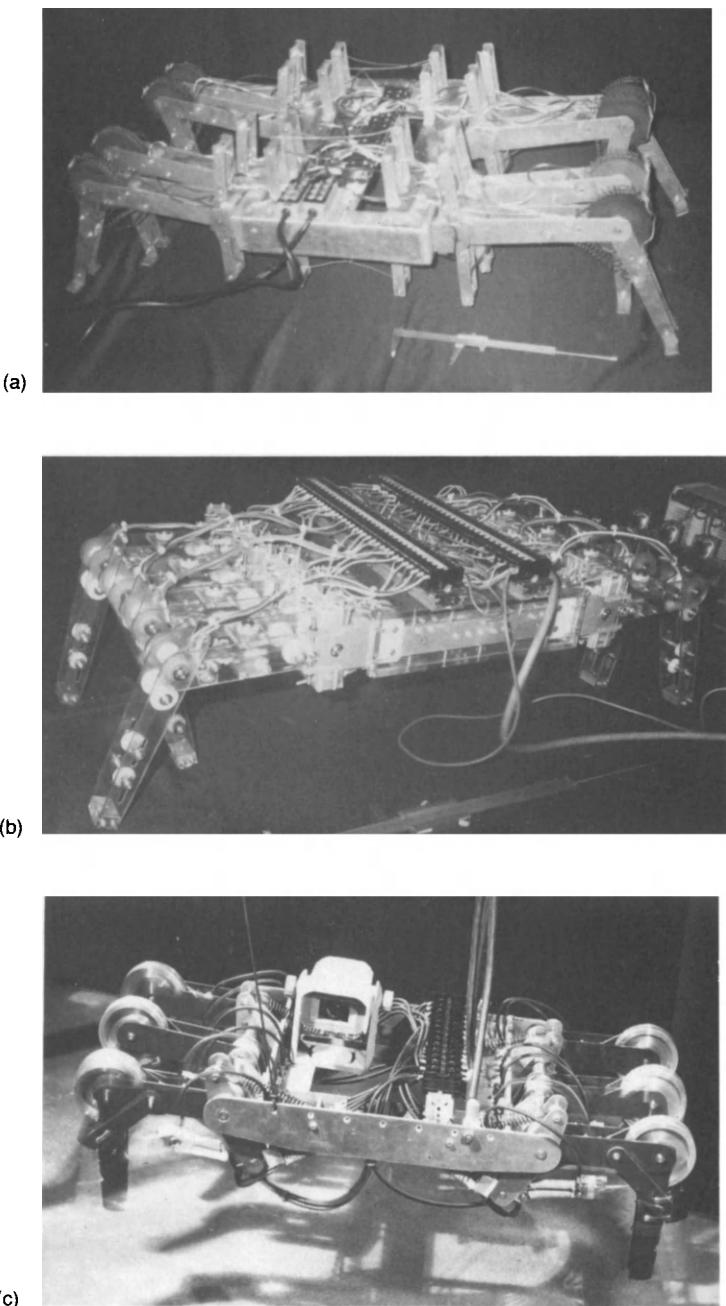


Figure 14: Three generations of robotic crab are shown. The second (b) was controlled by microprocessor, the third could be controlled remotely.

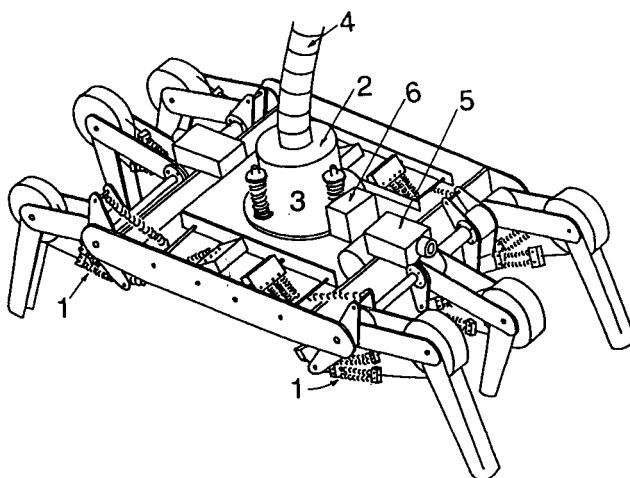


Figure 15: A schematic drawing of the robotic crab shows the 6 basic components: (1) the SMA wire or spring, (2) the nodule collector, (3) the collecting claws, (4) the suction pipe, (5) the ultrasonic sensor, and (6) a TV camera.

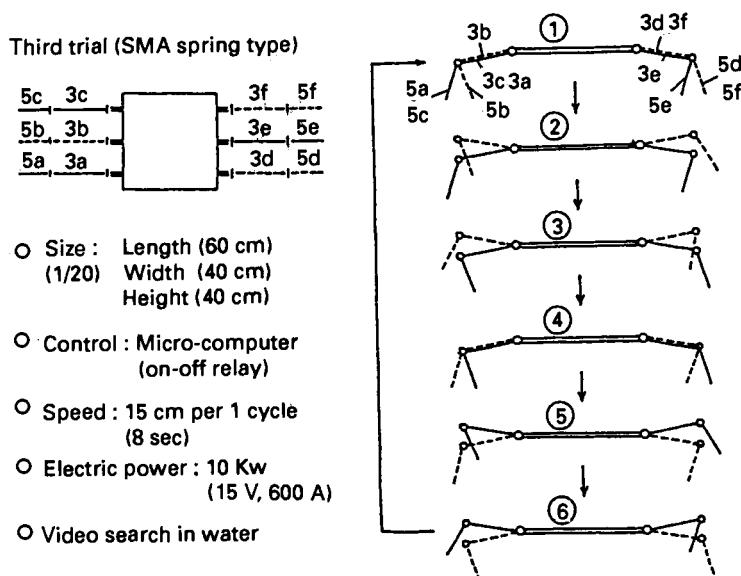


Figure 16: Motion of the crab is accomplished by the coordinated movement of the 6 legs as shown above. The size of the unit was 60 cm (length) by 40 cm by 40 cm. The speed was 15 cm per cycle (8 seconds), with a power consumption of 10 kW (15V at 600 amps).

3. A Summary of the Use of Shape Memory in Robotics:

Several prototypes of three dimensional manipulators and walking robots have been manufactured by both universities and industries. Some of these are shown in Figures 17 through 22 and are described in the captions of those figures. These robots have for the most part been developed in Japan during the last 5 years. SMA's are used to provide compact designs, light weights, soft and smooth motions, and three dimensional movement. These characteristics should provide particular advantages in harsh environments, such as ocean water, space, nuclear and medical. Many technical issues still have to be resolved however, such as response speed, precise control, fatigue and efficiency. Nevertheless, SMA robots provide the opportunity to construct robots that better resemble, and are in fuller harmony with, nature and human life.

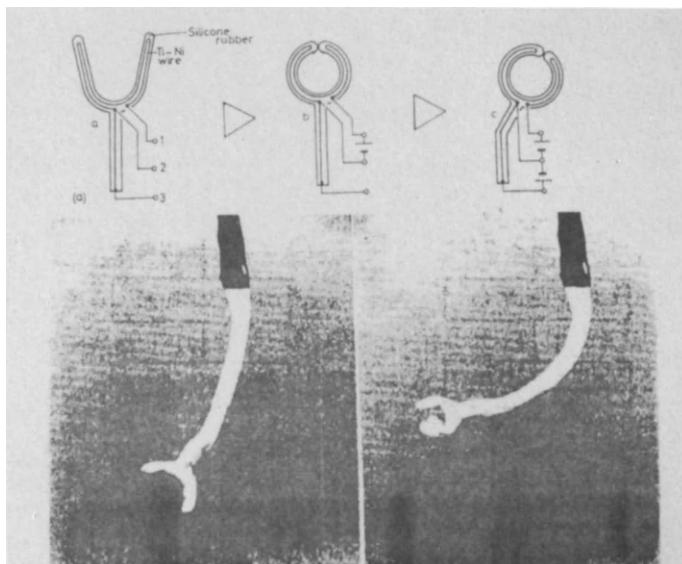


Figure 17: An example of a soft SMA micro-manipulator covered by silicon rubber. The rubber acts as a biasing "spring" for the underlining 0.2 mm diameter Ni-Ti wire. The actuator cycles 2 times per second when immersed in water.

In present projects for developing manganese nodules in advanced countries, engineering systems towed by a mother ship have been proposed⁹. Such systems inevitably necessitate a very large-scaled control method from ship to collector; therefore it is very difficult to achieve total control, leaving a constant danger of overturn of the collector on the rugged sea bottom. In contrast, Ni-Ti SMA actuators are characterized by high strength, large power/weight ratios, three-dimensional free motions, exceptional corrosion resistance, etc. Therefore, an SMA robot allows simpler construction and the possibility to move smoothly without environmental pollution of the ocean floor. Based on these characteristics, we will probably be able to reduce the complexity of the machine components and weight.¹¹ Consequently, the manufacturing maintenance costs would be considerably reduced.

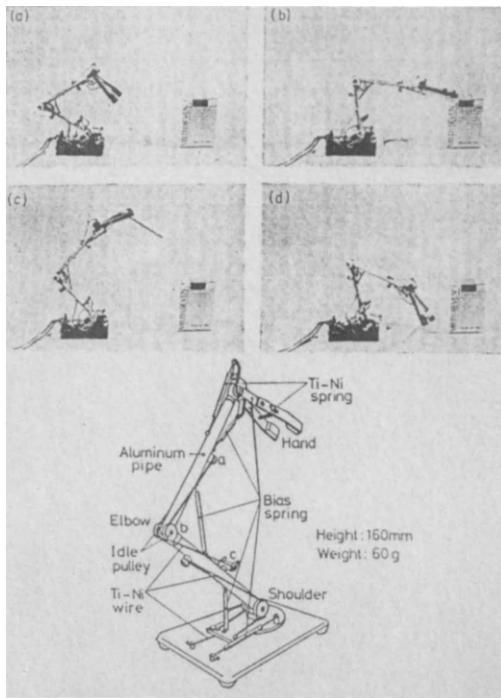


Figure 18: A skeleton-muscle type of SMA manipulator. The shoulder, elbow and hand are constructed from Ni-Ti wire and biasing spring. The very fine wire diameter (0.2 mm) allows rapid motion.

As mentioned earlier, there are two main technical problems. One is the low energy efficiency in transforming electrical power into motions of the SMA actuator (i.e. torque of the arm). In general, the efficiency of a conventional robot system is about 40-50% while that of the SMA robot is only 5-6%. One way to improve the energy efficiency is to adopt a thermally closed system by alternately circulating hot and cold water through the tubes connected to each SMA actuator. A second problem is the fatigue and shape memory degradation following the continued use of an SMA actuator. Improvements in mechanical design may be necessary, but this may not be a major technical problem considering that SMA Ni-Ti materials have relatively long fatigue lives.⁸

Other technical problems related to the precise control of the SMA robot and the collection system remain, but it is expected that these would be more than compensated by the advantages of an SMA robot in such extreme environments.¹² Still other sub-sea opportunities exist and are being researched, such as an SMA heat engine that generates electrical power from the small differences in temperature between the sea surface (25°C) and the sea bottom (4°C), and an artificial fishery using SMA's to provide large-scale circulation of nourishing sub-surface sea water.

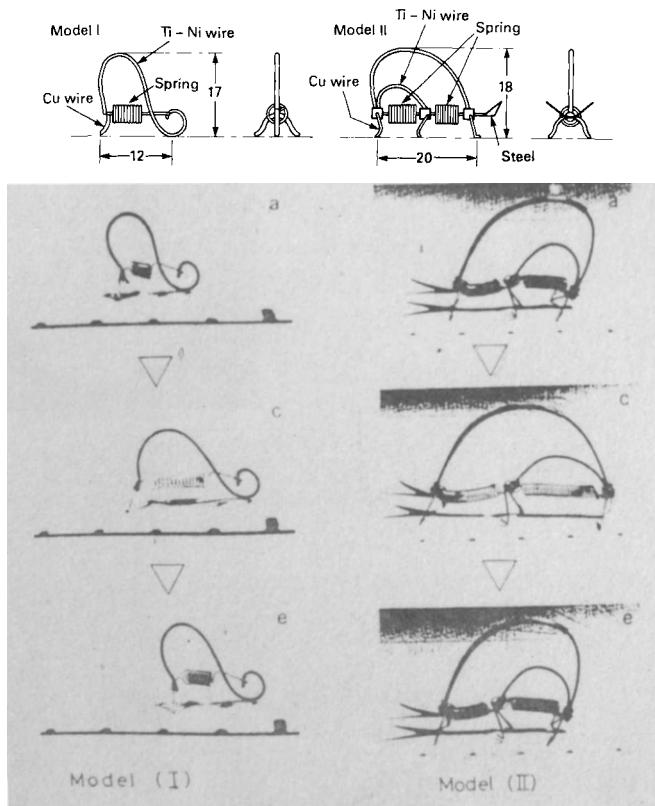


Figure 19: Two examples of a mechanical "animal", which can move smoothly without noise, very much as an insect would move. Motions are controlled by a microcomputer.

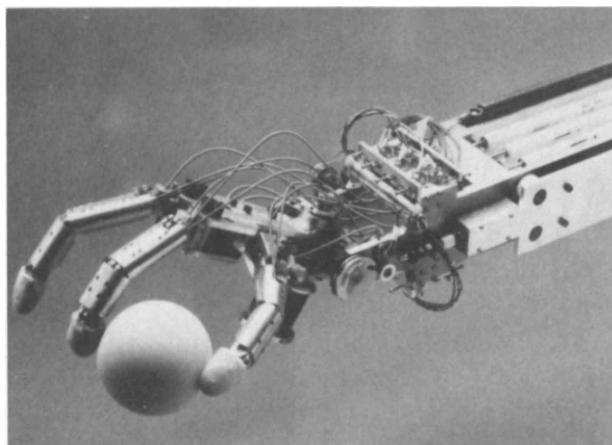


Figure 20: An SMA robotic arm built by Hitachi resembles a human hand and can grasp a very delicate ball without damage.

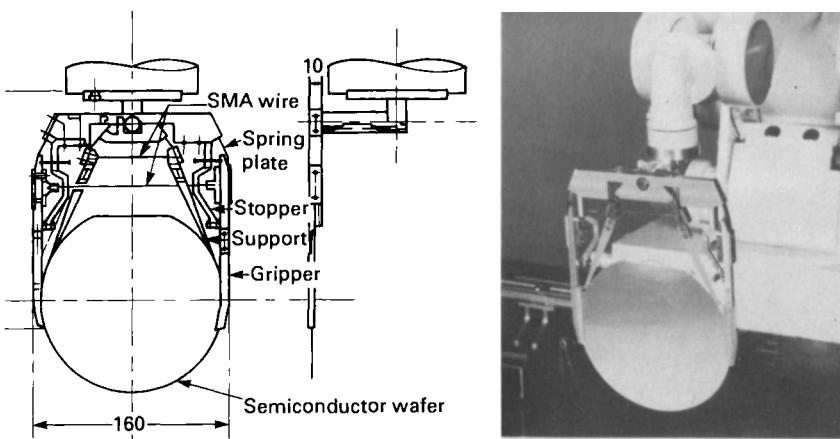


Figure 21: Similar to the arm shown in Figure 20, this SMA gripper is used to grip and transport semi-conductor wafers without damage. The actuator is simply constructed and helps in maintaining very clean, dustless environments, especially in vacuum chambers.

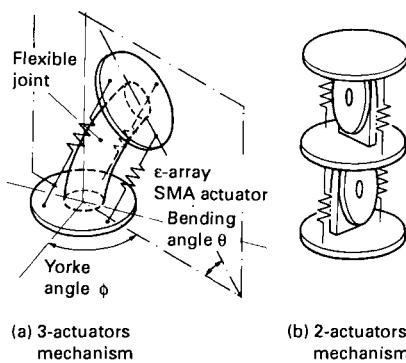
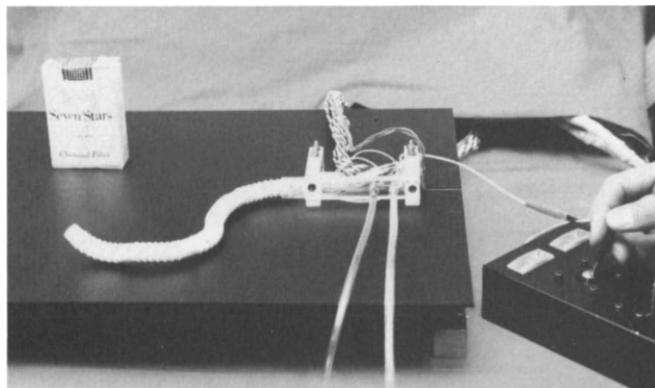


Figure 22: A robotic endoscope uses several articulated segments (top) to provide an arm capable of making complex, articulated motions in three-dimensions.

Acknowledgement

I have received many useful comments from Prof. K. Miwa (Waseda University), Prof. S. Hirose (Tokyo Institute of Technology) and Dr. Y. Hosoda (Hitachi). The SMA submarine robot has been developed jointly by the engineering faculty of Tohoku University and Honda Seiki Co., Ltd.

References

1. Miwa, Y., Honma, D.: Bull. Japan Inst. Metals **24**, 61.(1985).
2. Hosoda, Y., "Application of Shape Memory Alloy for Robots", JSME Seminar, 648, 53 (1987).
3. Honma, D., Miwa, Y. and Iguchi, N., "Application of shape memory effect to digital control actuator", Bull. JSME, 27-230, 1737.(1984).
4. Furuya, Y., Goto, Y., Shimada, H. and Honda, R., "An extreme operation submarine robot actuated by shape memory alloy", MRS'88 Meeting (Tokyo, 1988, June).
5. Kurabayashi, K., "Control of shape memory actuator", System and Control (Japan), 29, 288.(1985).
6. Furuya, Y., Shimada, H., Goto, Y., Honda, R., "Mechanical animal shape memory robot", 4th Japan robot symp., P. 461. (1986)
7. Hirose, S., Oita, K., Tsukamoto M., et al, "Three-dimensional active endoscope actuated by shape memory effect", 2nd Japan Robot Symp., .123.(1984).
8. McNichols, J.L. Jr., Brookes, P.C., Corny, J.S.: J. Appl. Phys. **52**, 7442.(1981).
9. Manganese nodules mining system projected by MiTi, Rep. Tech. Rese. Assoc. of MNMS (1982, Japan).
10. Furuya, Y. et al.: J. Japan Inst. Metals **52**, 139 (1988).
11. Ishibashi, T., "Mechanical animal robot crab", Marine Museum published by Tokai University, Japan, 13-2 (1983).
12. Furuya, Y., Shimada, H., Goto, Y., Honda, R., "An extreme operation robot designed to develop the deep submarine manganese nodules resources", Proc. of Symp. International Co-operation on Industrial Robots '87, P.261, (SICIR-'87, Tokyo, 1987).

A Shape Memory Arming Device

David Goldstein and Ens. Alex Weiner, USN
Naval Surface Warfare Center
10901 New Hampshire Avenue
Silver Spring, MD 20903-5000

Customarily, the fuzes used in military explosives protect against premature detonation. Following some specified time or event, the fuze enables detonation to occur. These functions together are called *safing and arming*, and commonly are performed by a clockwork mechanism. The manual removal of an external pin unlocks the mechanism, initiating the timing out cycle leading to arming.

The concept of using shape memory alloys for safing and arming has intrigued designers for many years. For example, Eckelmeyer¹ investigated alloying of Ni-Ti in 1975 with the intent of increasing the transformation temperature to meet requirements for the maximum permissible storage temperature for weapons. Scarzello, Lenko, et al² has received a Notice of Allowability for a US Patent which presents a scheme for coupling time-controlled discharge of a capacitor through a Ni-Ti wire as a means of obtaining both safing and arming in a compact device.

Figure 1 shows an artist's rendition of a simple prototype device used to demonstrate feasibility of the arming concept. A standard 9 volt alkaline battery pulses the 24 cm

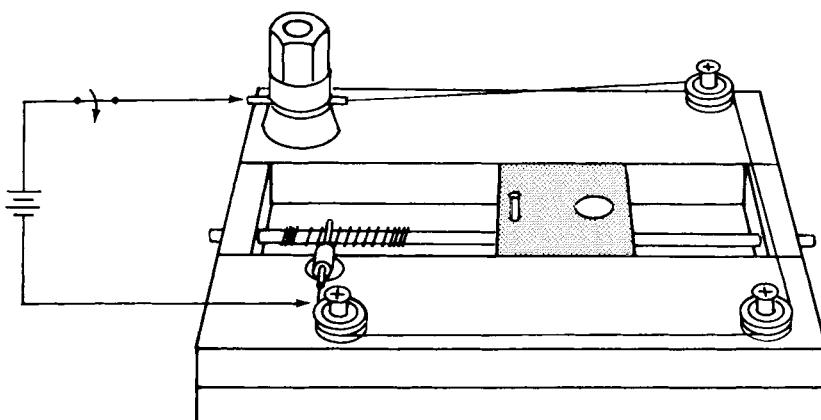


Figure 1: Prototype test bed for safing and arming after activation of Ni-Ti wire.

(9-1/2 inch) long Ni-Ti wire, which then contracts. This unlatches the pin mounted on the slide plate. A spring force drives the plate forward along its track to a position that aligns the elements in the explosive lead. The spring biasing the contraction of the Ni-Ti opposes accidental release of the slide plate from vibration or shock, an important consideration.

A second feature of this prototype is straightforward and will be included in a subsequent prototype to be constructed. Its purpose is fire safing, i.e., preventing arming due to heating during a conflagration. This too will be done with a Ni-Ti wire, configured similarly to the one shown in Figure 1, but it will pull a latch into a position blocking the slide plate from moving.

Table I: The Temperature Requirements for Two NiTi Wires.

Temperature (°C)	Conflagration		Arming
	Safing		
-55 to 100	Inert		Inert
100 to 120	Activation		Inert
120 to 130	--		Inert
130 to 160	--		Activation

Temperature requirements for activation of each of the two Ni-Ti wires are shown in Table 1. Temperature requirements for the assembled component of the fuze are that it be stable (inactive) from -55 to +80°C, that the conflagration safing wire activates between 100 and 120°C; and that arming occurs by electrically heating the Ni-Ti wire from 130 to 160°C. In addition to meeting these temperature requirements, the Ni-Ti must contract 6 mm (1/4 inch) under 172 MPa (25 ksi) load upon discharge of a 10 Joule capacitor into it. The ability of a single alloy of Ni-Ti to transform at different temperatures as a function of stress, particularly for one-time uses, was presented by Goldstein et al.³

1. Experimental Results and Discussion

Alloy 83825 is a slightly Ti rich binary, induction melted and reduced to wire by hot swaging and conventional wire drawing. Its transformation temperatures, as annealed and with several tensile strain levels, are shown by the differential scanning calorimetry (DSC) curves in Figure 2. Calorimetry was done in Perkin Elmer DSC 2. The 83825 wire in the annealed condition (425°C for 1/2 hr) was not strained following annealing. Tensile strains above 6% cause the transition temperatures to shift to higher values as shown by the shift of the peaks away from the vertical index line at

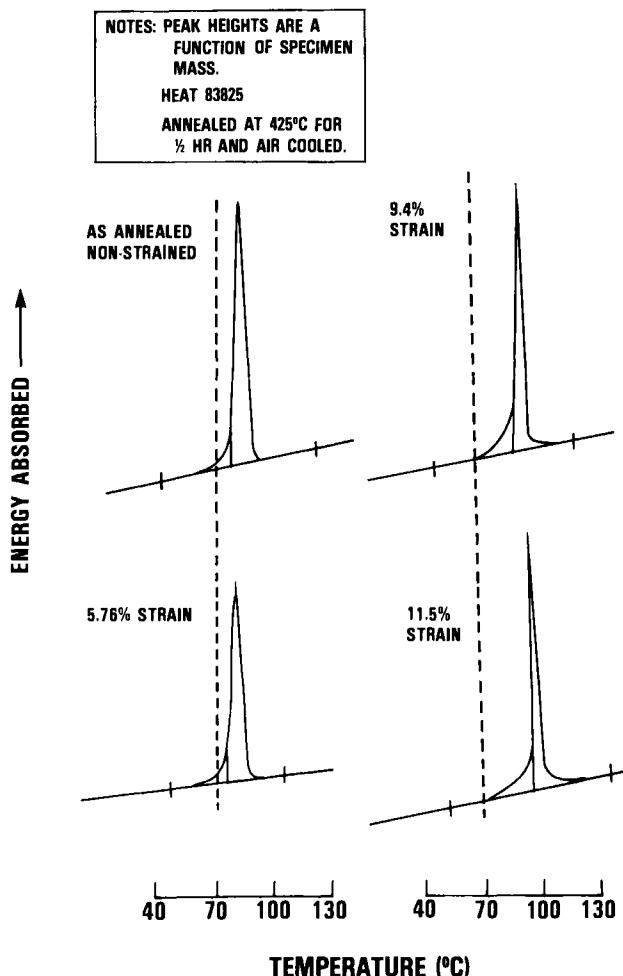


Figure 2: Transition temperatures of Ni-Ti during heating.

70°C. It appears that an 11.5% strain will suffice for the conflagration safing wire, with its major transformation (under no-load conditions) occurring between 90 and 105°C. Figure 3 shows the contraction of 1/4mm (10 mil) wires which were strained between 5 and 6%, and heated under different external dead loads in an air circulating furnace. Temperature control was within $\pm 2^\circ\text{C}$. Contraction is estimated to be accurate within $\pm 0.2\%$. Figure 3 also shows quantatively how the transition temperature finish value (A_f) and the transition temperature range, vary as function of the load. Note that the amount of contraction varies inversely with load. For example, at 130°C, if the load is reduced from 1.3 to 0.82 Kg (2.8 to 1.8 pounds), the contraction increases from 0.8 to 3.7%.

The energy requirements for contraction under the various loads are shown in Figure 4. The wires tested were 25 cm (10 inches) in length and were pre-strained between 5

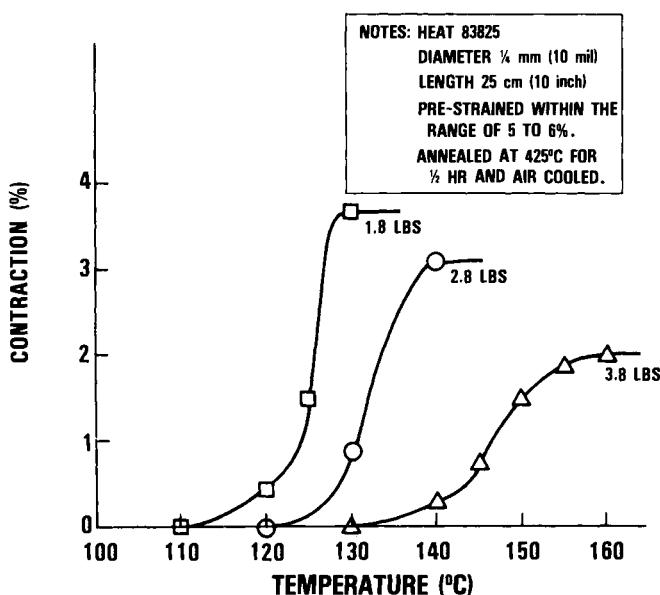


Figure 3: Effect of applied stress during shape recovery of Ni-Ti.

and 6%. Nine joules will lift a 8.2 Newton load, equivalent to 159 MPa (1.8 pounds, 23 ksi) stress. This increases to 12 joules for a 13 N load (2.8 pound, 35.7 ksi), for the same 4% contraction.

The energy values in joules were calculated from the voltage drop across a shunt device place in series with the Ni-Ti wire which is pulsed for known times. The voltage trace was displayed on an oscilloscope.

Since the pulse durations were less than 0.2 seconds, it was assumed that heating was essentially adiabatic, i.e., insignificant heat could be lost to the atmosphere in so brief a time. Peak currents from the power supply were less than 5 amperes at a constant potential of 18 volts. The energy values on the ordinate of Figure 4 are shown specifically for a 25 cm (10 inch) long wire, and as normalized for weight. The energy was used to (1) raise the wire temperature to the transformation value, (2) provide 25 joules per gram minimum required for the latent heat of transformation, and (3) lift the dead weight.

Capacitor discharge, with an effective energy of about 6 joules in the first time constant, contracted a 5-inch length of wire about 4.5% under no-load conditions. As shown in Figure 5, the contraction amount decreased smoothly to 1% under about 117 MPa load. This is for 1/4 mm (10 mil) wire, stressed initially to have residual elongation between 5 and 6%. Note that pre-strains of 6.25 and 6.87% respectively have contractions of only 3.2% under no-load, suggesting that a maximum of 6% strain was the limit for good contraction of the wire.

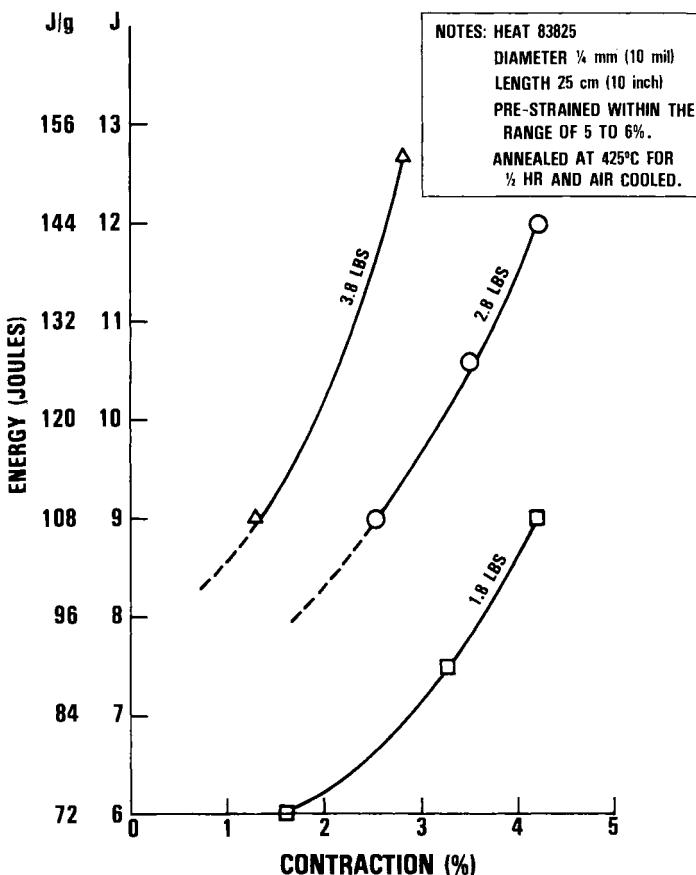


Figure 4: Energy of transformation of Ni-Ti under load.

2. Conclusions

Fine Ni-Ti wire prestressed between 5 and 6% and placed under load during electrical heating (pulsing) showed 4% contraction. The load on the $\frac{1}{4}$ mm (10 mil) wire was 13 N (2.8 pounds), equivalent to a stress of 248 MPa (36 ksi). The energy required was 12 joules (144 joules per gram). In furnace tests, the completed contraction with the same load occurred at temperatures of 140°C and higher. Capacitor discharge produced 4.5% contraction in a $\frac{1}{4}$ mm (10 mil) diameter, 12.7 mm (5 inch) length of Ni-Ti under no-load conditions. The contraction under a 6 N (152 MPa) load (1.35 pound, 22 ksi) was 0.9%. Increased electrical energy is required for increased load lifting if the percentage contraction of the Ni-Ti is fixed. Increased contraction can be "traded" for decreased load at a constant energy level.

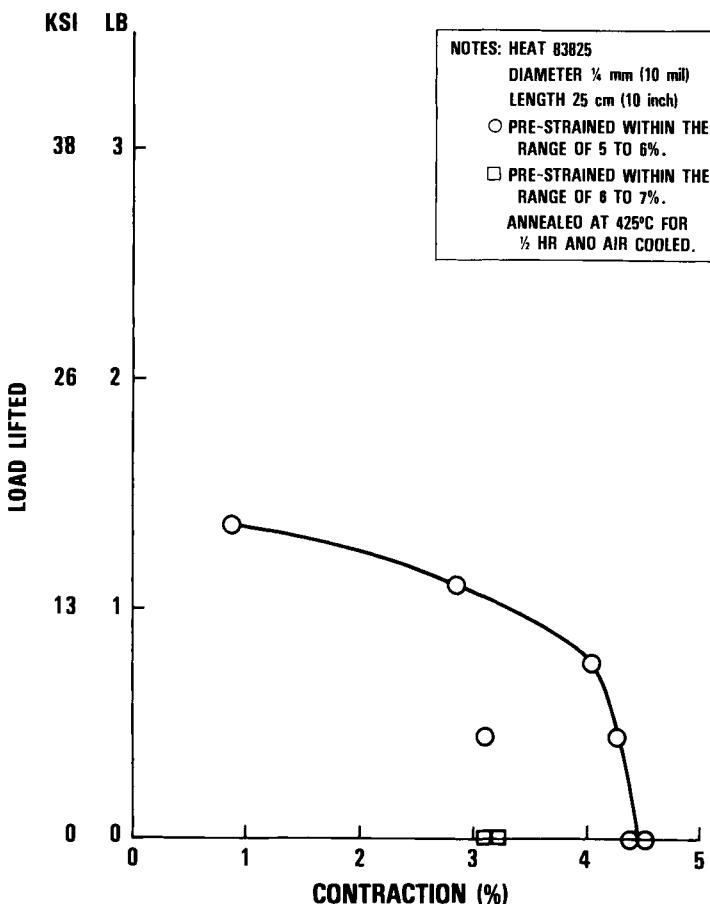


Figure 5: Shape recovery of Ni-Ti wire by capacitor pulsing.

During heating, both the temperature for completion of transformation and the range of the transformation temperature increase with increasing load.

3. References

1. Eckelmeyer, K.H., "The Effects of Alloying on the Shape Memory Phenomenon in Nitinol", SAND 74-0418, Sandia Laboratories (Mar. 1975).
2. Lenko, D., Scarzello, J., et al., NSWC Memorandum from office of Patent Counsel, Patent Application, Silver Spring, MD (4 May 1988).
3. Goldstein, D. et al., "Stress Effects on NITINOL Phase Transformations", J. of Metals, pp. 19-26 (Mar. 1987).

Using Shape Memory for Proportional Control

Darel E. Hodgson,
Shape Memory Applications, Inc.
10201 Torre Ave., Suite 310
Cupertino, CA 95014

Shape memory alloys have been applied to a wide range of devices, but almost all are two position products. The device is expected to have a temporary martensitic shape or position which is changed to the "remembered" austenitic shape when the alloy is heated. The product is only asked to exhibit these two ends of the position spectrum of which it is capable. It is rare that an application demands that the shape memory element undergo only partial transformation and interrupt its shape change at any one of the infinite number of positions between fully martensitic and fully austenitic. When this does occur, and the position is specifically chosen and then achieved by the device, we can say that it is being used for proportional control.

Using a shape memory alloy (SMA) for proportional control seems an obvious course since the transformation does not occur in a "step" fashion but rather as a continuous change over a temperature range. Using the effect to achieve a desired position or level of stress is therefore merely a matter of interrupting the progress of the transformation at the desired point.

The advantages of SMA for proportional control are that one can build a device which has:

- a) smooth and continuous motion - i.e. no motion steps
- b) total silence
- c) extremely small to extremely large forces
- d) any needed range of motions
- e) extremely fine resolution
- f) very small space requirements
- g) different ranges of temperature actuation
- h) actuation with any chosen heat source

1. Modifying the Rate of Shape Recovery

To accomplish proportional control with SMA, though, it is desirable that one modify the response of the shape change from the usual large amount of recovery with a small temperature rise to a much more gradual change with temperature. There are a number of ways one can achieve this recovery behavior with a much more gradual slope. The first technique is to design so the SMA element works against a spring with a high spring rate (K) which strongly increases the opposition stress as the SMA elements tries to accomplish its shape recovery. Because of the stress sensitivity of the recovery temperature, this will force the shape recovery to occur over a wider

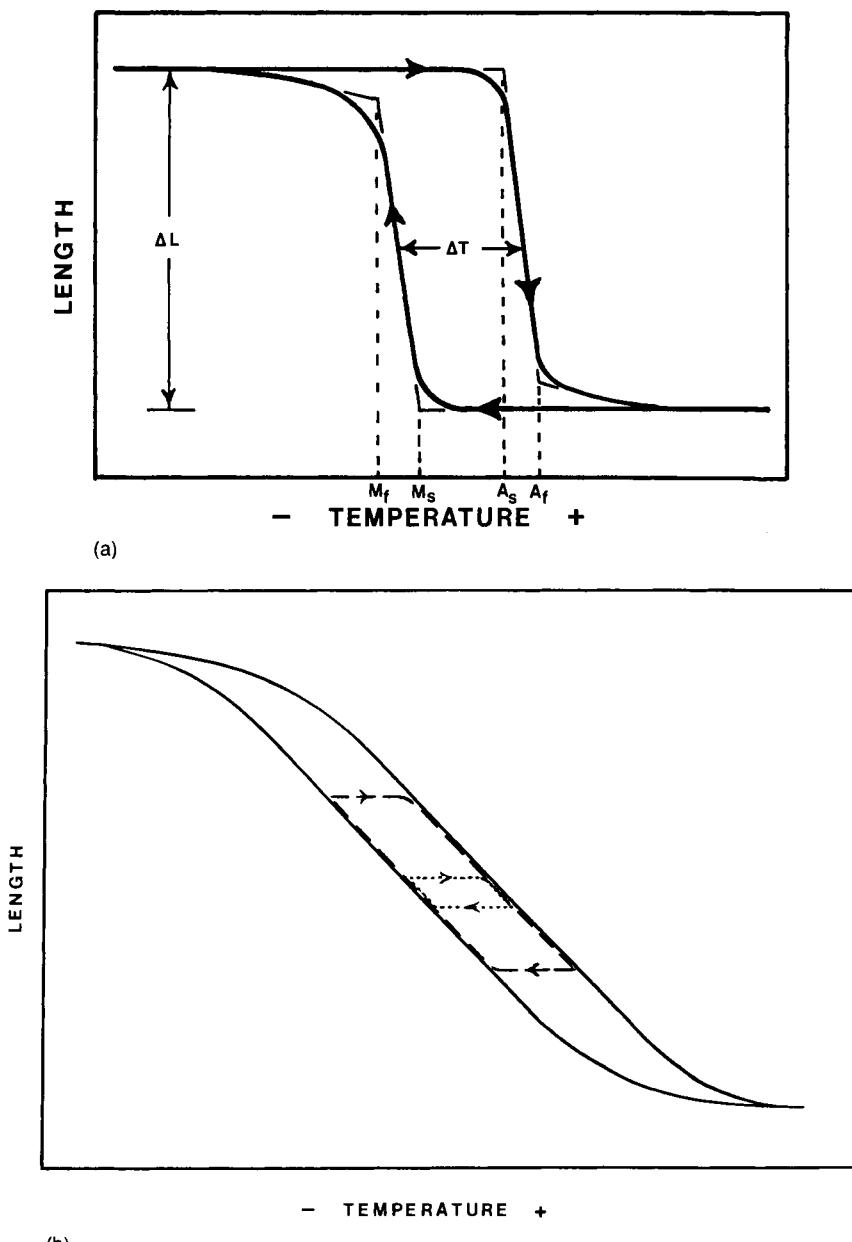


Figure 1(a) Motion vs temperature for a Shape Memory Alloy in uniaxial tension after standard thermomechanical processing (note the very steep curve giving large changes in length with small change in temperature), and (b) motion vs temperature for an SMA element processed to show gradual shape change with temperature and with improved control function response.

temperature range and thus give a smaller amount of motion for a given temperature change. The next way that one can increase the temperature range of the shape change is to design the SMA element so that it has a range of stress and strain in it as it acts rather than a single value of stress or strain as in a straight wire in tension. This will lead to motion over a wider range of temperature also. Finally, it is possible to heat treat and mechanically process the SMA element in such a way that the normally sharp shape change with temperature occurs over a much wider span of temperature. The "normal" curve of motion versus temperature and one which has been optimized for proportional control are shown in Figure 1. Note the much more slopey appearance of the modified curve and the fact that a small temperature change gives a much less drastic amount of motion from the SMA element.

2. Allowing for Hysteresis

Next, one must accommodate the hysteresis of the transformation. As can be seen by following the dashed line in the motion curve of Figure 1, even small changes in desired position mean that one must move across the temperature hysteresis gap if the device was previously moving in the opposite direction. That is to say, if one was moving position by slowly cooling the SMA element and then wanted to reverse the direction a small amount by heating the element one would need to heat the element all the way through the hysteresis before further heating would cause motion in the element. It is certainly possible with proper control circuits to handle this aspect of the control function, but one must be fully aware of it and allow for a more complex controlling function than simply proportionally increasing or decreasing the drive power to move the element either direction.

3. Position Feedback

Depending on the accuracy of the control one wishes to achieve, it is very likely that one will need to measure the actual control position or function continuously and feed that information back to the controlling circuitry. Without this feedback function, one can achieve a modest degree of proportional control as in the Matsushita air conditioner¹ using an SMA spring to direct the flow of air depending on its temperature. It is not, however, possible at this time to achieve better than approximate control.

By using position or function measurement and feedback to the control circuit, the limit of control is determined by the accuracy of measurement and the ability of the control and power circuitry to adjust the power to the SMA element in extremely small increments. The theoretical positioning capability of an SMA element is as small as atomic dimensions, and positioning to within a few millionth of an inch has been achieved². A block function diagram of the elements necessary to achieve proportional control using an SMA element is shown in Figure 2.

4. Mechanical Design

The actual design of the control device can take many forms and the SMA element can be nearly any form used in other SMA applications - springs, wire, bent beams, rings or more complex shapes. A fairly simple design is shown in Figure 3, which is a valve controller used to adjust the rate of liquid flow through a small plastic tube. The shaft with its sharp tip is pushed against the tube by the bias spring with sufficient force to

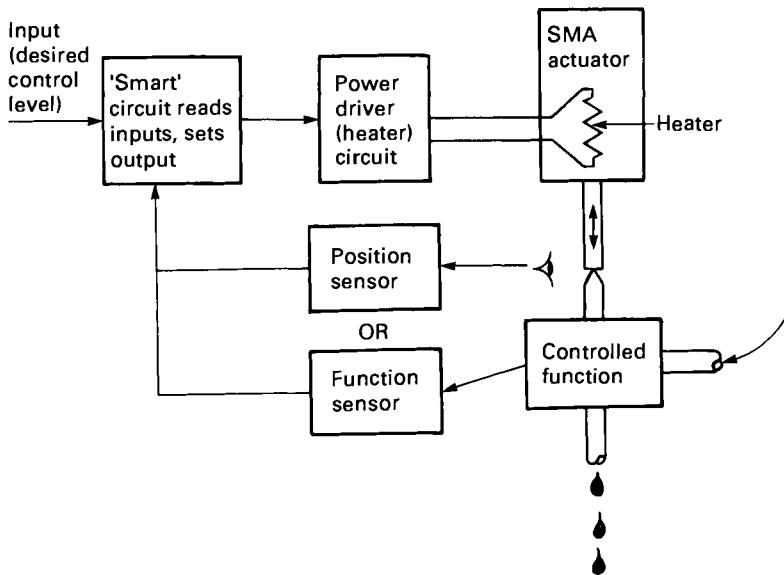


Figure 2: Block function diagram of the elements needed in a typical shape memory alloy proportional control device.

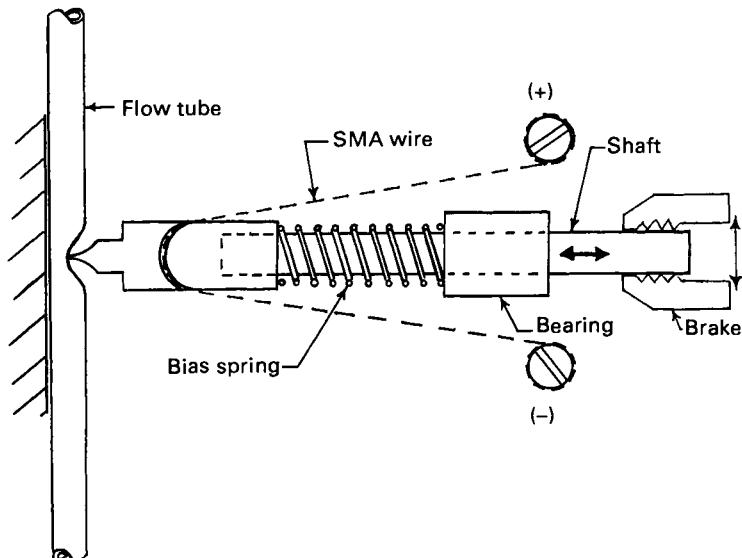


Figure 3: Simple shape memory alloy proportional control device, showing the SMA wire to adjust the valve, the bias spring to close the valve and sketch the wire upon cooling, and a brake mechanism to hold the valve in place without constant heating of the wire.

shut off the flow. The shape memory wire is then in the stretched, martensitic condition. By applying electrical current through the SMA wire, the level of heating can be adjusted until the shape recovery which tries to shorten the wire is just sufficient to pull the valve open against the force of the bias spring. Precise control of the current in the wire is able to adjust and hold the valve position within about 0.25 microns.

Another aspect of the design in Figure 3 which should be mentioned is the brake on the shaft. This allows one to achieve a desired level of control and then hold the shaft in that position so the power to the SMA wire can be turned off. In some cases, such as a battery powered device, such power saving is an important issue to the service life of the device. By careful attention to design items such as this, shape memory can be used in control devices where other solutions may have been impossible to implement due to power or space requirements.

5. Conclusions

Shape memory alloys are usually used in a bimodal fashion where the component has a particular shape or position in its cold, martensitic condition and another shape after heating to restore it to the austenitic condition. A wide range of applications are possible, though, in which the SMA component is deliberately interrupted part way through the transformation. Upon either heating or cooling, this allows one to achieve a specific level of force or desired position from the continuous spectrum exhibited by the SMA as it transmits the entire transformation.

Special processing of the SMA element to achieve uniform rates of transformation over an extended temperature range is useful to accomplish proportional control, and a control feedback system to confirm acquisition of the desired level is usually needed. If these are provided, SMA can yield control precision which is extraordinary.

As the many advantages of shape memory alloys in proportional control are understood by more design engineers, it is expected that a much wider range of applications will emerge. Those who provide the alloys can accelerate this process by supplying the best mechanical characterization of the material which they can, and by determining processing treatments which yield properties most conducive to control devices.

6. References

1. Todoroki, T., "Shape Memory Actuator for A/C Systems", Proc. of Engineering Aspects of Shape Memory Alloys, Michigan, (to be published), August, (1988).
2. Clappier, J., Beta Phase, Inc., private communication, March, (1988).

An Engineer's Perspective of Pseudoelasticity

T.W. Duerig and R. Zadno
Raychem Corporation
Menlo Park, CA

The fourth type of shape memory event or application is different from the previous three in the sense that it is completely isothermal. In the most general sense, any non-linearity in a stress-strain curve during unloading can be referred to as pseudoelasticity. Typical examples are shown in Figures 1 and 2. Figure 1 shows a small non-linearity which can

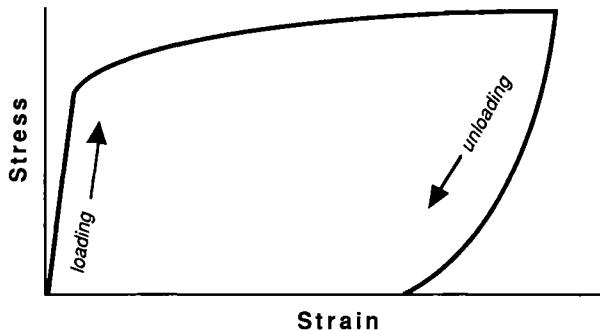


Figure 1: An idealized stress-strain curve showing generic pseudoelasticity, defined by a non-linearity during unloading.

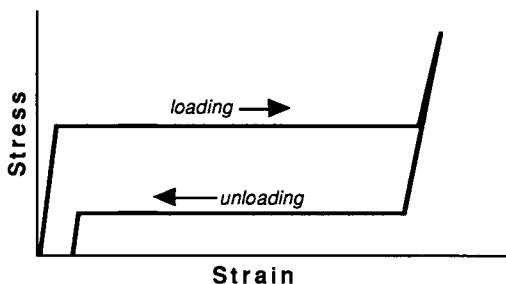


Figure 2: Idealized superelasticity, showing a clear unloading plateau and an extremely large elastic range.

present very real problems to engineers trying to base designs on a constant elastic modulus, but as will be shown, also provides one of the most important benefits of shape memory coupling devices. Figure 2 shows a more useful aspect of pseudoelasticity: under certain circumstances, a pseudoelastic metal can be deformed more than 10 times as far as a conventional metal and still completely springback to its original undeformed shape. Because of the enormous amount of elastic deformation available in these metals, this second type of pseudoelasticity is often referred to as "*superelasticity*", and represents one of the most useful manifestations of the shape memory effect.

The terms superelasticity and pseudoelasticity have often been used interchangeably. For the purposes of this paper, pseudoelasticity will be used as the more general term, referring to any non-linearity during unloading. Superelasticity will be a more specific term, used to describe pseudoelastic materials that show a plateau during unloading. Since tensile curves are often not as clearly defined as in Figure 2, we use the presence of an inflection point during unloading as evidence of a plateau. In this paper we will overview the various causes of pseudoelasticity, discuss some of its mechanical aspects, and then look at the various ways to develop the effect. In general the emphasis will be on Ni-Ti since that remains the preferred superelastic alloy. Although the use of superelastic materials looks at first to be very appealing in many designs, there are two factors that have significantly reduced the scope of application: an extreme temperature dependence of some of the key mechanical characteristics and some comparatively complicated fatigue effects. The first of these will be discussed in some detail in this paper, while the second will be treated in detail in the following paper.

1. The Mechanisms of Pseudoelasticity

Pseudoelasticity can be caused by either twinning or by a stress induced phase transformation. Though both mechanisms will be considered in this section, the emphasis in subsequent parts will be with transformational pseudoelasticity since practical *superelasticity* has only been achieved through this mechanism.

1.1 Twinning Pseudoelasticity

Twinning pseudoelasticity is caused by the reversible motion of twin boundaries. In materials that are deformed by either the formation of twins or by the motion of twin boundaries, it is often the case that the positions of the twins in the deformed state are not stable, and that there is a driving force causing them to return to their original positions during unloading¹. We will look at a few examples of such phenomena.

Figure 3 shows a typical tensile curve of Ni-Ti in the austenitic phase, well above M_d so one can be sure that no transformational effects are involved. In this case, the material is known to deform by mechanical twinning². This process is different from conventional twinning in that it leaves the lattice improperly ordered and thus in a higher energy state. This is called *pseudotwinning*, and might, rigorously speaking, be called a phase transformation. When the stress is removed, some of the twins will shrink or even disappear in order to return the preferred lattice ordering.

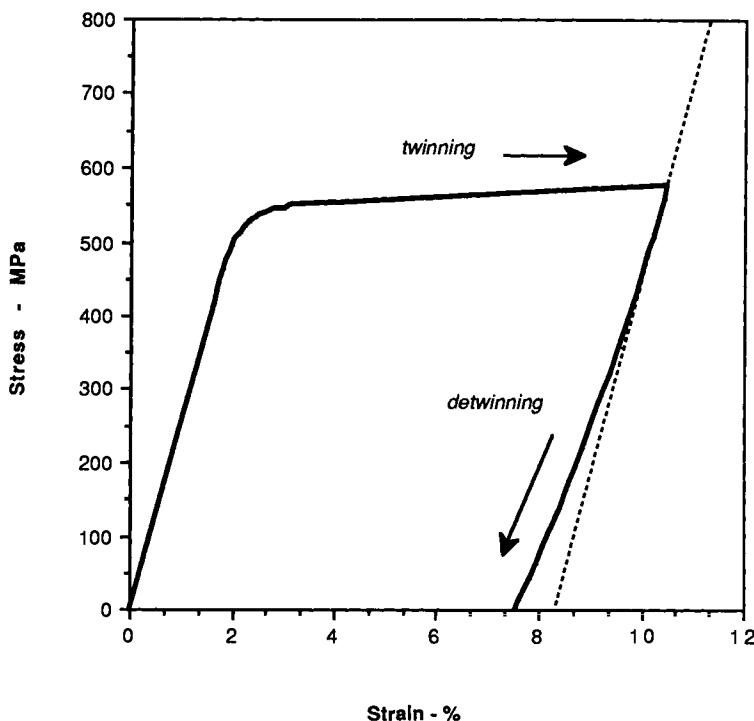


Figure 3 A room temperature stress strain curve of a NiTiFe shape memory alloy showing pseudoelastic behavior due to pseudotwinning. M_d in this alloy is -30°C, so no martensitic transformation is involved.

This provides a springback strain above and beyond the normal elastic effects - represented by the light line tangent to the unloading line in Figure 3. Although the pseudoelastic effect is small, it is of great practical importance in couplings and fasteners since it reduces the apparent modulus of the installed Ni-Ti device. Reducing modulus creates what is often called a "live" joint, meaning that it is better able to compensate for differential thermal expansion and other small motions tending to separate the recovered part from the substrates being joined. Note also that we can define three moduli for the unloading portion of Figure 3: a tangent modulus at the onset of the unloading, a tangent modulus at zero load (always lower than the first), and a secant modulus (simply connecting the starting and finishing points) - which of these is the most relevant depends upon its intended use. Figure 4 shows some other interesting aspects of pseudoelasticity that concern cyclic loadings. The most obvious is that reloading does not follow the same path as unloading, so that a strain hysteresis develops. Secondly, though less obvious without careful data analysis, the unloading secant modulus monotonically decreases as the deformation is increased³.

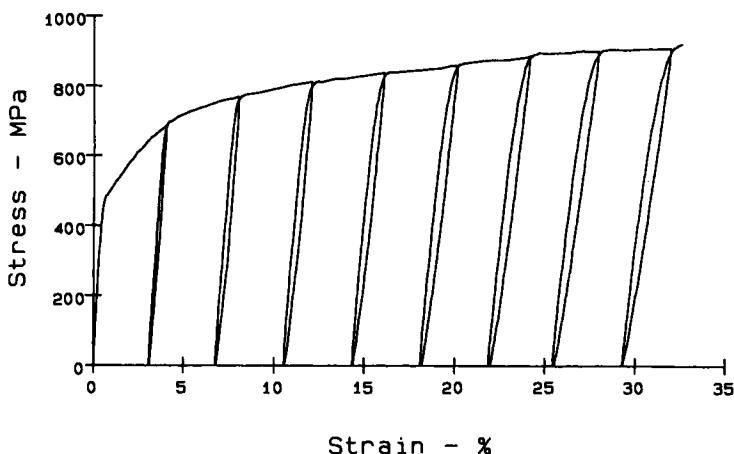


Figure 4 Repeated loading and unloading cycles of the same alloy shown in Figure 3 illustrate that pseudoelasticity can result in a hysteresis, which increases in width with total strain.

The martensitic phase of Ni-Ti also shows twinning pseudoelasticity, as shown in Figure 5. In this case deformation proceeds by the motion of martensite twin boundaries. Although the motion does not cause disorder, there nevertheless appears to be a driving force for the twins to return to their original positions when the stress is removed. It is not clear whether this is due to residual stresses, a desire for accommodation dislocations in the twin boundaries to maintain an even spacing, or is

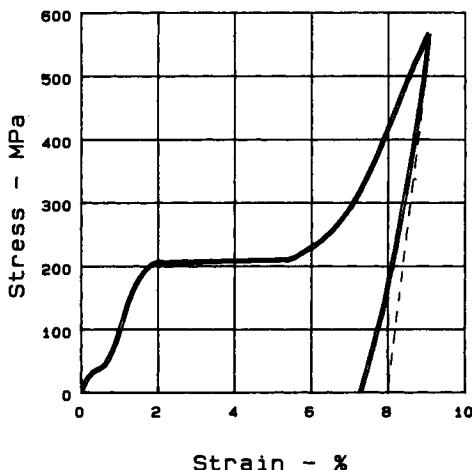


Figure 5 Twinning pseudoelasticity in NiTiFe martensite. In this case deformation was at -196°C while A_s is -100°C . (The first small inflection at 0.5% strain is due to a reorientation of the R-Phase.)

related to reducing the tapered portion of martensite plates⁴. Martensitic pseudoelasticity appears to have no constructive value, but does play an important role in design in that it complicates calculations of springback and the compliance of a device.

The previous two examples of twinning pseudoelasticity were small effects with no direct practical application. Larger twinning superelastic effects are observed in many systems such as AuCu^{5,6} and InTi⁷. In the case of AuCd, the martensite structure is orthorhombic, and undergoes mechanical pseudotwinning. Unlike the austenitic phase of Ni-Ti, proper variant selection of the pseudotwins in AuCd can result in springback strains of several percent. This enhanced elasticity has also been called *ferroelasticity* due to certain similarities with magnetism⁵. One interesting aspect of this, however, is that aging at room temperature while in the deformed state will lock-in that shape and prevent superelastic springback⁸. This aging effect is due to a reordering of the pseudotwins, bringing them back to their more stable configuration of atomic order. Due to the chemical make-up of these special alloys there has not been a substantial effort to commercialize them.

1.2 Transformational Superelasticity:

For the purposes of this paper, we will only discuss transformational pseudoelasticity in its fully evolved form: superelasticity. Transformational superelasticity requires that a stress induced transformation occurs, usually stress inducing a martensitic phase from an austenitic. This is illustrated in Figure 6. The austenitic phase is

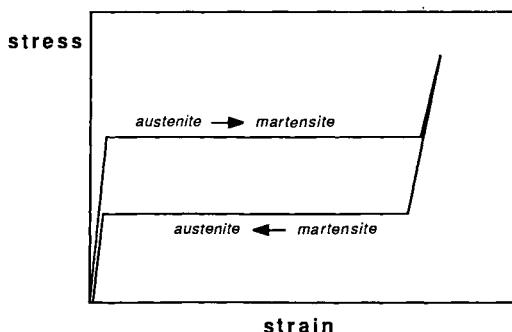


Figure 6 Superelasticity results when martensite must be stress induced, and when austenite again becomes the stable phase when the stress is removed. The plateaus themselves are caused by the ability of martensite to accommodate the applied stress by selecting those variants most favorably inclined.

stable before the application of a stress, but at some critical stress the martensite becomes the more stable, causing yielding and a stress plateau. The plateau is caused by the ability of martensite to form in several different variants just as in the case of shape memory, the

difference being that in this case the variants are selected during the transformation itself instead of through the progressive consumption of some variants by others. The details of deformation along the stress induced plateau will be treated later in more detail; now, it is enough to observe that at some strain the structure becomes completely martensitic and the stress again begins to increase in a linearly elastic way, but this time according to the martensitic modulus. Since the martensite is only stable because of the applied stress, the austenite structure again becomes stable during unloading, and as discussed in previous sections, the original undeformed shape must be returned. Thus the reverse transformation causes an unloading plateau, but at a lower stress level due to the transformational hysteresis.

Transformational superelasticity can only be realized if the temperature of the material is below M_d and above A_s . If the temperature exceeds M_d the martensite cannot be stress induced; if the temperature is below A_s the stress induced martensite will remain stable during unloading and again no unloading plateau will be observed. In fact for full superelasticity we require the application temperature to be above A_f . The details of this temperature dependency are well summarized in Figure 7, and will be examined in more detail in the section 2.2.

Figure 6 showed an example of superelastic behavior in Ni-Ti, the most commonly used alloy. Significant superelasticity can also be found in Cu-based memory alloys. In the case of Cu-Al-Ni, large strains have only been

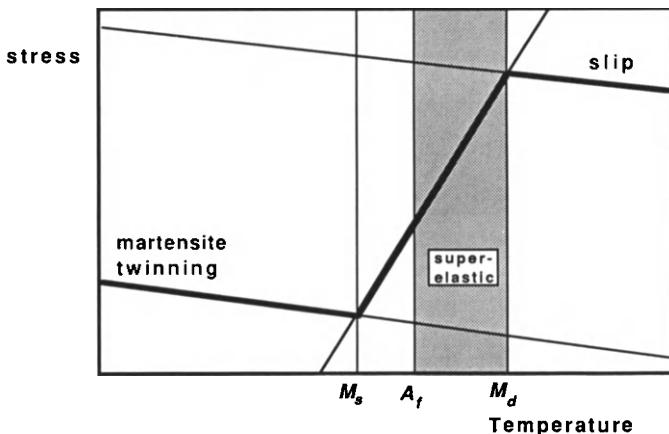


Figure 7 A stress-temperature phase diagram illustrating the temperature range in which superelasticity can be found. The heavy line represents the variation in yield stress usual to shape memory alloys. Below M_s , deformation occurs by martensite twinning; between M_s and A_f , the martensite is stress induced, but once induced, is stable; between A_f and M_d the stress induced martensite becomes unstable during unloading and superelasticity is observed; above M_d the deformation is due to slip since martensite can no longer be stress induced.

reported in single crystals⁹, where superelasticity is enhanced by successively stress inducing different martensite structures (Figure 8).

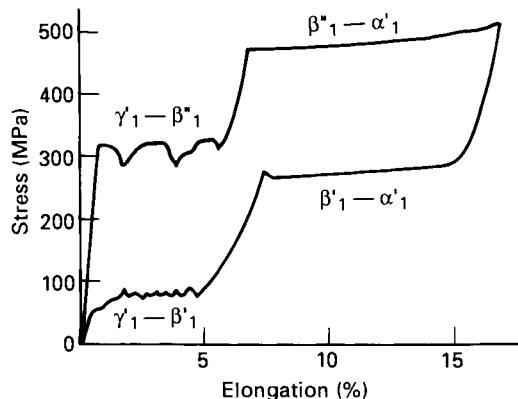


Figure 8 Single crystals of CuAlNi are known to exhibit extremely large superelastic strains (nearly 20% in certain crystallographic directions). These strains are the result of successive martensitic transformations from a DO_3 austenite. (Figure from Otsuka and Shimizu⁹).

Perhaps the most interesting aspect of this sequence is that one of the stress induced martensites (β'_1) leads to a hysteresis-free superelasticity¹⁰. Still, polycrystalline Cu-Al-Ni alloys tend to be inherently brittle, subject to fatigue and to have much smaller superelastic strains (on the order of 1%)¹¹. For these reasons, they are not particularly interesting from an engineering point of view. Larger transformational superelasticity (on the order of 5%) can be found in polycrystalline Cu-Zn^{12,13} and various ternaries based on that system¹⁴, best known of which is Cu-Zn-Al^{15,16}. There are currently no superelastic applications of Cu-Zn-Al, so the engineering principles are less well known than in the Ni-Ti system.

2. A Mechanical description of superelasticity:

Since the stress-strain curve shown in Figure 2 is unusual, it is necessary that several new physical descriptors be defined - these are shown in Figure 9: the loading plateau stress (σ_l), the unloading plateau stress (σ_u), the total deformation strain (ϵ_t), the permanent set or unrecovered strain (ϵ_p), and the stored energy (E) (defined as the area under the unloading curve). Figure 9 is very clearly defined, but this is not always the case; the plateaus can often be difficult to clearly distinguish, particularly in

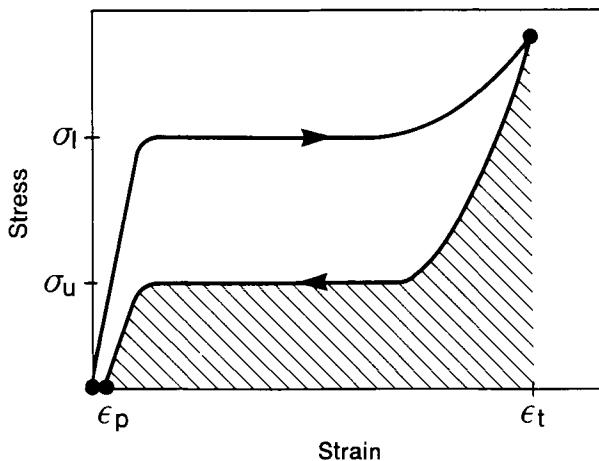


Figure 9 A schematic superelastic stress-strain curve defining various key mechanical descriptors: the loading stress - σ_l , and unloading stress - σ_u , the total strain - ϵ_t , and the permanent set - ϵ_p . The two plateau stresses are defined by the inflection points, the springback strain (not shown) is defined by the difference between the total and plastic strains, and the stored elastic energy is defined by the area under the unloading curve (shaded).

polycrystalline Cu-based alloys, so it is necessary to define σ_l and σ_u in such a way as to be easily measured in ill-defined curves. The best definition appears to be to use the stress at which inflection occurs. The difference between σ_l and σ_u is the superelastic hysteresis, which is a direct consequence of the thermal hysteresis in the shape memory event, and can be quantitatively connected to the thermal effect through the stress rate from the Clausius-Clapeyron equation^{1,15}.

2.1 Strain dependencies

The first mechanical aspect of superelasticity to be discussed is the dependence of the above superelastic properties upon the total deformation strain - ϵ_t . Clearly the permanent set (ϵ_p) should increase as the total strain exceeds some critical value and slip begins to contribute to the overall deformation. Figure 10 shows a typical relationship, with no measurable permanent set at strains below about 8%, and with permanent set increasing from that point on. The largest recoverable strain observed in polycrystalline Ni-Ti is 11% - this includes both the elastic and superelastic contributions. Figure 11 shows a second important effect of the total strain. Although the loading plateau of course remains constant, the unloading plateau is decreased as the total deformation is increased. This means, in effect, that the stress hysteresis increases as the material is deformed. One can easily imagine that some materials are

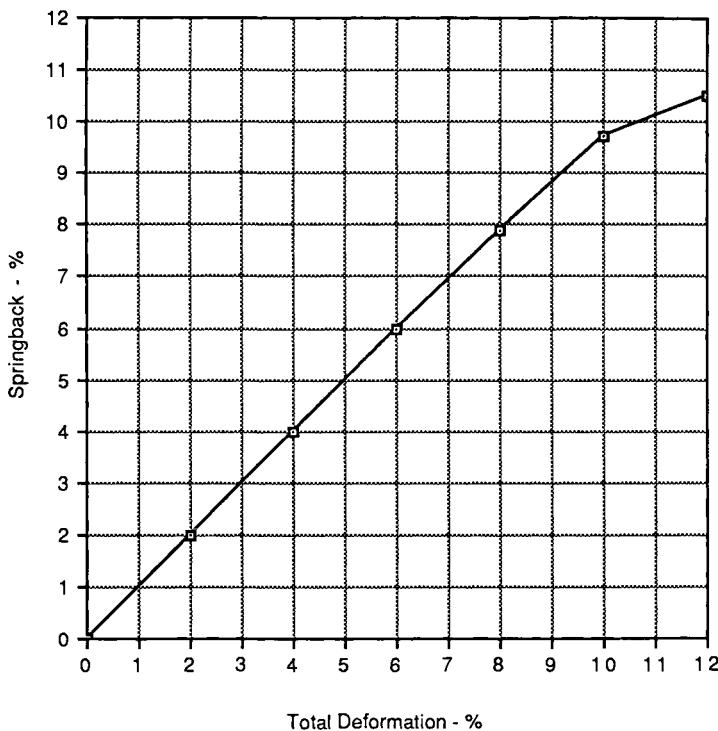


Figure 10 The dependence of springback strain ($\epsilon_t - \epsilon_p$) upon the total tensile deformation. The results are for a Ni-Ti wire with 50.8% Ni after cold working 40% and annealing at 375°C for 30 minutes.

superelastic only after small deformations. One curious but important implication of Figure 11 is that one does not necessarily store more energy by deforming further - increasing deformation beyond 6% increases springback, but the ensuing reduction in stiffness is substantial and can actually reduce the springback energy. This same effect has been observed in Cu-Zn-Al¹⁷.

2.2 Temperature dependencies

As pointed out earlier, the extreme temperature dependency of superelasticity is one of the two factors most limiting to its use. This is shown quantitatively in Figures 12 and 13. Figure 12 shows that very low ϵ_p values can only be obtained over a 60°C temperature range. Below this ideal window, recovery during unloading is incomplete but heating after unloading will complete recovery. Above the ideal window the temperature approaches M_d and the applied deformation can no longer be accommodated without some dislocation movement; in this case, heating

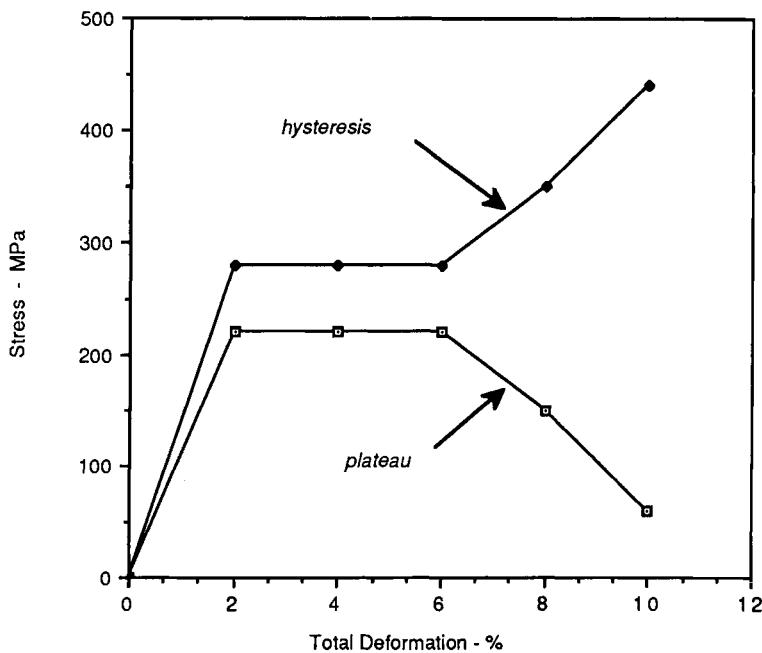


Figure 11 The dependence of the unloading plateau stress upon total deformation in the same wire described in Figure 10.

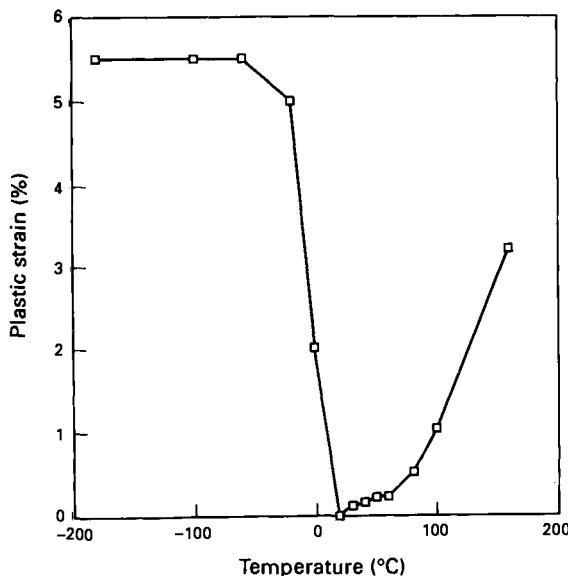


Figure 12 The temperature dependence of plastic strain, or permanent set, after an 8% tensile deformation. The wire was Ni-Ti with 50.8 at. % Ni and was cold worked 40% and annealed at 375 for 30 minutes before testing.

after unloading should cause no additional recovery. As we will see in section 3, there are a variety of ways to move the ideal window to lower or higher temperatures, but windows wider than 80°C have not as yet been observed in Ni-Ti.

Figure 13 shows a second aspect of this temperature dependency: the effect of temperature upon σ_l and σ_u . As the temperature increases, so does the difference between ambient temperature and M_s , and so do the two plateau stresses. The rate of increase is linear and follows the Clausius-Clapeyron equation. Typical values for the rate of increase, or stress rate, range from 3 to 20 MPa/°C in superelastic Ni-Ti alloys¹⁸. This increase is extremely significant in design even when the expected temperature changes are small. In the example shown in Figure 13, one would expect the unloading stiffness to more than double between room temperature and body temperature. It is therefore essential that medical product testing be done at body temperature.

One should be aware that the phase transformation has a rather high latent heat of transformation - about 5 cal/g in Ni-Ti, but strongly dependent upon M_s ¹⁹. Thus although we frequently say that the superelastic event is isothermal, this is often not the case due to self-heating in the specimen. As strain rates increase, and conditions become

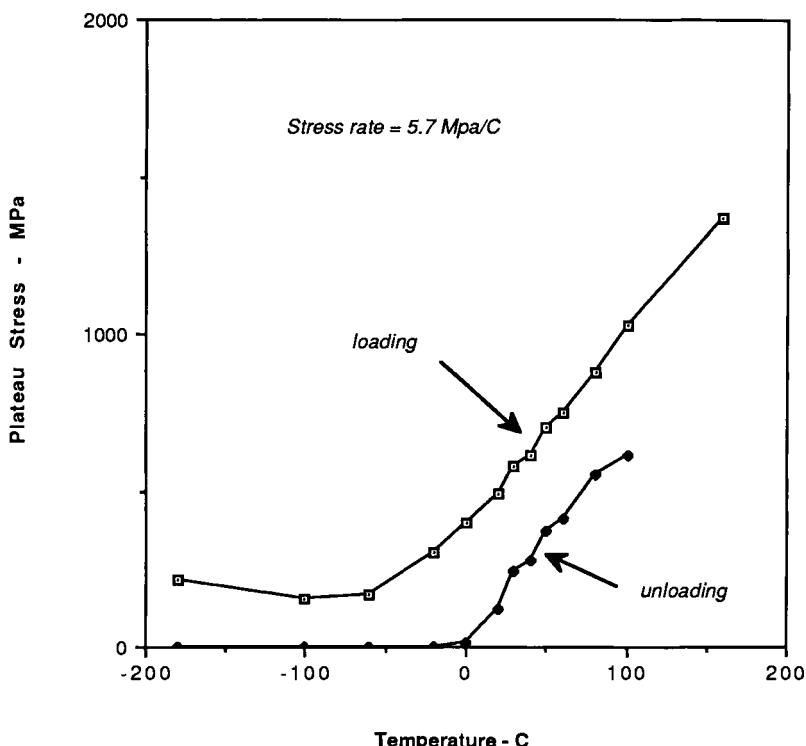


Figure 13 The temperature dependence of the plateau stresses in the same wire described in Figure 12. The stress increase follows the Clausius-Clapeyron equation.

more adiabatic, this effect can become quite substantial, often causing a temperature rise upon deformation of well over 5°C (and of course a similar temperature drop upon unloading). This thermal effect has been observed in both Ni-Ti²⁰ and in Cu-Zn-Al²¹. One result of this is to introduce some apparent strain rate effects.

2.3 Energy storage

The storage of large amounts of elastic energy is, more often than not, the primary driving force for using a superelastic material. Elastic energy is defined as $\int \sigma d\varepsilon$, or the area under the unloading portion of the stress-strain curve (as shown in Figure 9). Table I compares typical elastic energy storing capacities for some of the better known spring materials with that of superelastic Ni-Ti and Cu-Zn-Al. The advantage of superelasticity is clear. Still, one must be aware that these values must be derated for fatigue effects, and as will be discussed in the next paper, the derated values for Ni-Ti and Cu-Zn-Al are far less favorable than those shown in Table I. The values for Ni-Ti and Cu-Zn-Al shown in the tables are approximate (but very conservative) numbers for room temperature pseudoelasticity - at 50°C, it would be possible to significant better these figures. It should also be noted that the numbers for Cu-Zn-Al may not represent a maximum since a systematic and complete analysis has never been done. These are numbers calculated from typical literature data^{13,15}.

Table I: Comparison of the Elastic Energy Storage Capacity of Various Spring Materials.

Material	Maximum Springback Strain	Stored Energy
Steel	0.8%	8. Joules/cc
Titanium	1.7%	14. Joules/cc
Ni-Ti	10.0%	42. Joules/cc
Cu-Zn-Al	5.0%	14. Joules/cc

2.4 Detailed stress and strain analysis

The stress applied by a pseudoelastic spring is not dependent upon deflection, but on temperature; exactly opposite to a conventional spring. For many applications, this is one of the most appealing features of superelasticity: one can apply a constant stress over a very wide range in strain. But again one must be careful. The curves shown in Figures 2 and 9 are tensile stress-strain curve; torsional or bending curves would look substantially different due to the unusual non-linear strain distribution across the cross-section (common bending and torsional models assume elasticity to derive the strain distribution). Absolutely flat tensile stress-strain curves such as shown in Figure 2 are quite easily produced in Ni-Ti, but the same material will show substantial work hardening in bending and torsion²²⁻²⁴. This is illustrated in Figure 14, where actual tensile properties of a superelastic Ni-Ti alloy are used in a finite

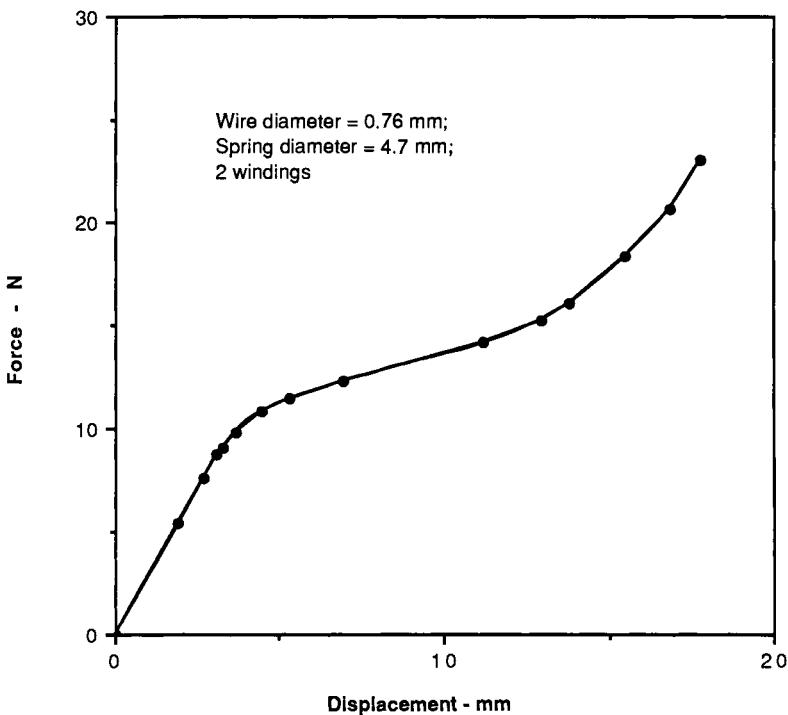


Figure 14: One complication of superelastic product design is that conventional bending and torsional formulae are no longer valid, and finite element analyses, though complicated, must be used. Shown is a numerical calculation of load-deflection in a helical spring using the program MARC, and actual data from an ideally superelastic (flat loading plateau) Ni-Ti alloy. It is impossible to achieve a flat loading plateau in torsion or bending.

element analysis of torsion.²⁴ Compressive properties too greatly differ from tensile even though strain distributions are uniform.²⁵

Finally, it is worthwhile taking a closer look at what is occurring as deformation proceeds along the loading and unloading plateaus. As the engineering stress-strain curves can be perfectly flat in this region, one might wonder what the true stress-strain curves would look like - for many design approaches this information is essential. Traditionally one would assume a uniform change in the cross-sectional area and conservation of volume, then calculate the true stress-strain values from areas based on elongation. As is so often the case with these alloys, this approach is not successful. Deformation along the plateau occurs by the growth of Lüder's bands²⁵ - thus we do not have a uniform cross section, but locally necked regions which grow in length until they consume the entire gauge length. The area is either the deformed area or the undeformed - there is no in-between. As deformation proceeds, the integrated length of the necked regions grow linearly with strain until the

end of the plateau. So, in fact, the entire concept of true stress-strain and engineering stress-strain must be realigned to fit superelastic deformations.

3. Processing to produce and maximize superelasticity

Nearly all shape memory alloys are superelastic at some temperature. To be useful, the superelastic temperature range must be made as wide as possible, and must be centered around the application temperature. The process of shifting the window can be accomplished rather easily by the methods described in earlier chapters. Here we will predominately deal with methods used to maximize the window width, which directly results in reducing the permanent set. Because there is a great deal more data available for the Ni-Ti system than for the Cu-based, our examples will key upon Ni-Ti, but the principles involved are valid for all alloys.

There are four property modifications that will directly increase the width of the superelastic window:

1. Increasing austenitic strength: This has the effect of delaying the onset of dislocation movement, thereby increasing M_d and increasing the upper temperature bound for the superelastic window.
2. Decreasing the stress rate: Given a critical stress for the onset of dislocation movement, decreasing the rate at which stress will be increased will also raise M_d .
3. Decreasing hysteresis and/or increasing the steepness of the transformation: Both will tend to keep A_f closer to M_s and therefore further separated from M_d . One can also think of this in terms of decreasing the stress hysteresis, thereby ensuring an unloading plateau at lower temperatures.
4. Eliminating or suppressing the R-Phase: If the R-Phase transformation occurs above A_f , it will then be controlling of the point at which the window for complete superelasticity begins. This is actually of significant importance since the high strength conditions that normally show superelasticity also show an R_s above room temperature.

For the most part, the logic of each of the above four can be seen in Figure 7. Although all are important, the most important is the first. There are two practical methods of accomplishing this in Ni-Ti: cold working and age hardening. Since Ni-Ti alloys with less than 50.6 at.% Ni are stable²⁷, this clearly limits the use of the second approach to Ni-rich compositions. It should also be noted that it may not be possible to have the first without the second in Ni-rich alloys. We will now briefly discuss both approaches.

3.1 Cold worked superelastic Ni-Ti:

As has been discussed in earlier²⁷, Ni-Ti can be strengthened by cold working, but in order to preserve the ability to stress induce martensite the cold working operation must be followed by a recovery anneal. To illustrate these effects upon superelasticity, we will follow the behavior

of a Ni-Ti V alloy, though the same trends exist in the binary alloys. The reason for choosing this particular alloy is that it is stable, and allows us to ignore aging effects which could complicate the issue in the binary alloys.

As the annealing temperature is decreased, more cold work is retained and the austenitic phase is strengthened while ductility is reduced (illustrated in Figure 15). On this basis, one would then expect better

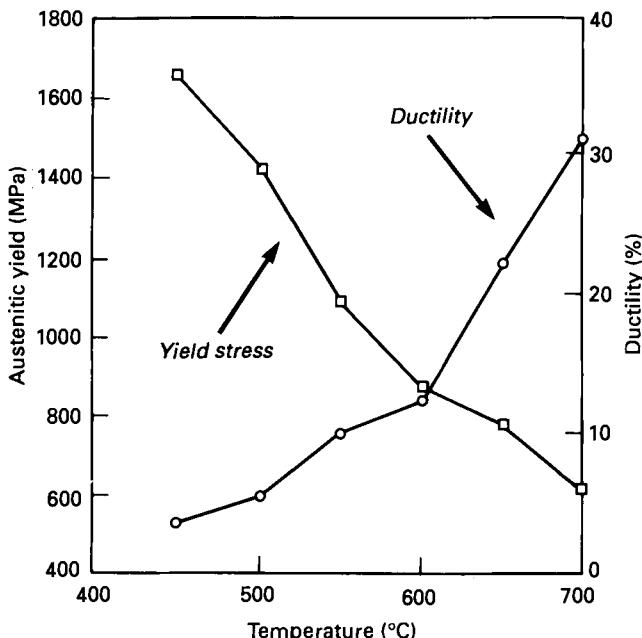


Figure 15 Strengthening response of a superelastic Ni-Ti-V wire after cold working 40% and annealing for 1 hour at the temperature shown on the ordinate.

superelasticity after lower temperature anneals. Figure 16 shows that the stress rate is rather dramatically reduced by lowering the annealing temperature, again indicating superior superelasticity. Figure 17 shows that M_s decreases as more cold work is retained; since A_f is below room temperature in all cases (Figure 17 shows the transformation temperatures under load - the unloaded values are roughly 15 degrees lower), we expect superelasticity at all cases. Also shown in Figure 17 is that the thermal hysteresis is decreased by increasing the annealing temperature, a factor favoring superior superelasticity after more complete anneals. Finally, we might expect higher superelastic stress plateaus after lower annealing temperatures since the difference between the application temperature and A_f is greater. The net effect of all these factors can be seen in figures 18 and 19. Lower annealing temperatures

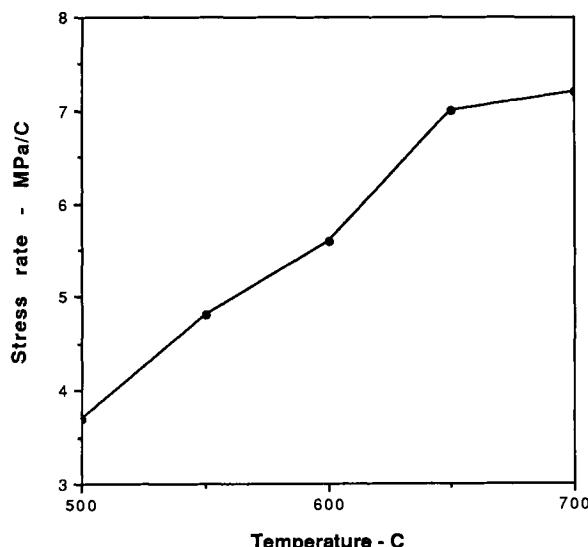


Figure 16 Stress rate is dramatically influenced by the amount of residual cold work. This trend of decreasing stress rate with increasing resistance to slip is common and quite powerful, but not fully understood. The alloy is as described in Figure 15.

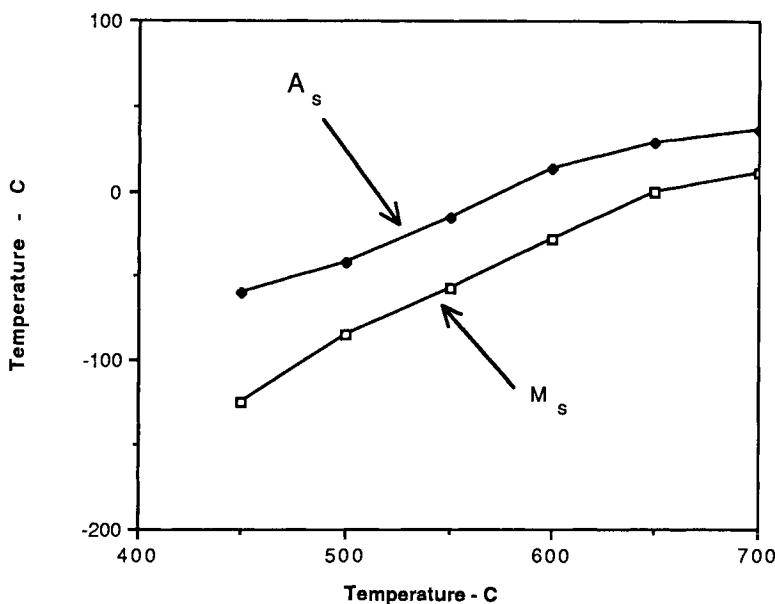


Figure 17 The transformation temperatures A_s and M_s measured mechanically at 150 MPa (the unloaded values would be less by an amount determined by the stress rate). The general trend of decreasing M_s with increasing residual cold work is usual in Ni-Ti alloys. The alloy is as described in Figure 15.

decrease permanent set after room temperature deformations (Figure 18) and increases the superelastic stiffness (Figure 19). We can see that the smaller thermal hysteresis found after the hotter annealing treatments manifests itself as a smaller stress hysteresis in Figure 19.

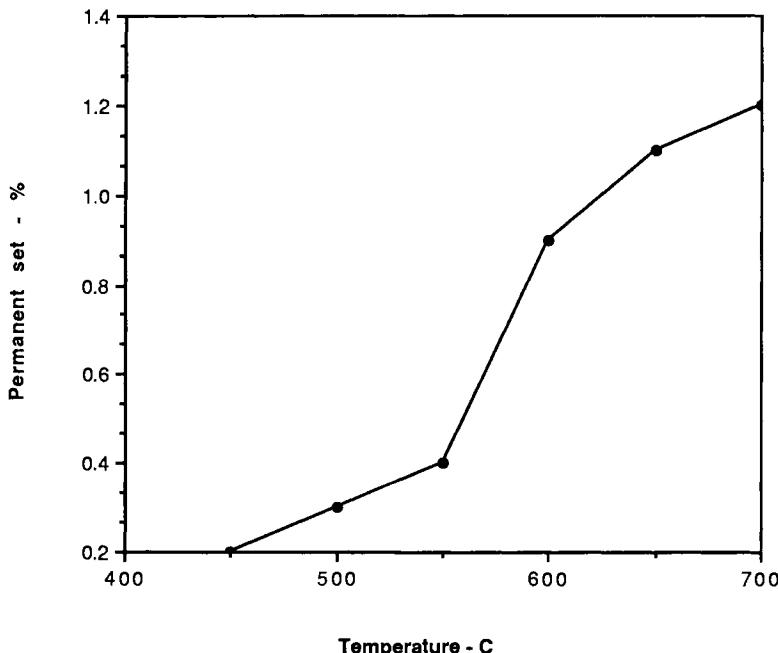


Figure 18 Permanent set (ϵ_p) after deforming 8.0% at 20°C is shown as a function of annealing temperature. The alloy is as described in Figure 15.

Thus decreasing the annealing temperature has every advantage in terms of superelasticity (lower permanent set, greater stiffness and greater energy storage capacity) but ductility is decreased as an exchange; the need to preserve a useful ductility is, in fact, one of the factors most limiting to the quality, or extent of, superelasticity. Conversely, we could say that an inherently more ductile Ni-Ti alloy would allow us to further enhance its superelastic character.

Although we have concentrated on the width of the superelastic window, a few words on some other aspects are in order. As we reduce the annealing temperature (increase strength) the superelastic plateau becomes flatter. Also, since M_s is decreased, the window is shifted to lower temperatures. We should also note that all of the above figures depicted binary Ni-Ti that was cold worked 40%. Decreasing the amount of cold work is exactly equivalent to increasing the annealing temperature, thus the detailed effects of cold working will not be presented. The effect of annealing time, however, is more complex. While it is true that shorter times are generally equivalent to lower temperatures, the effect is not straightforward. There appears to be a step in the strength/time curve.

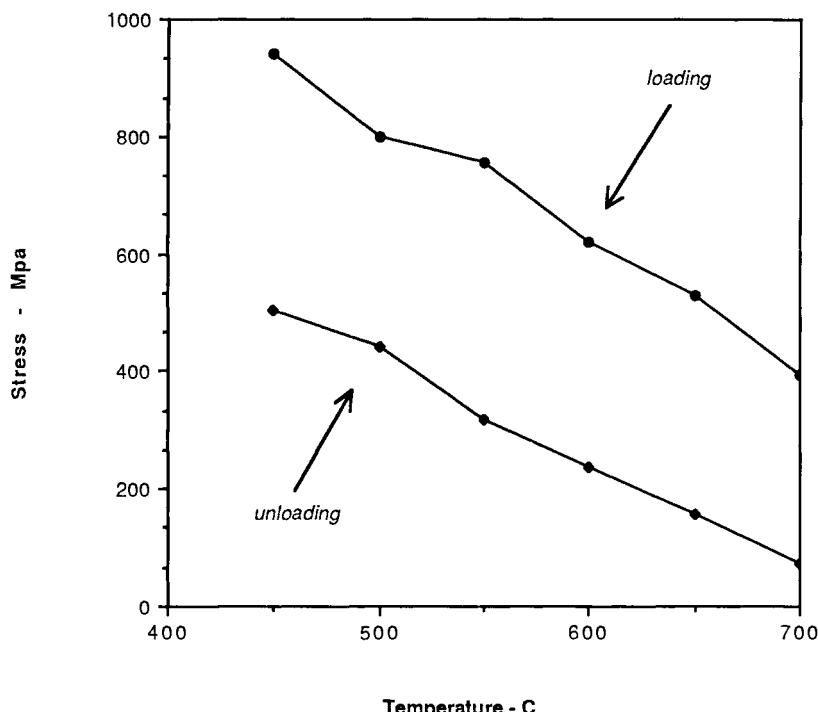


Figure 19 Superelastic plateau stresses shown as a function of the annealing temperature; note that the stress hysteresis decreases with increasing annealing temperature, in good agreement with figure 16. The alloy is as described in Figure 15.

Very short times at high temperatures appear to maintain high strength without increasing M_s and without bringing out the R phase. While these short (under 5 minute) heat treatments are usually impractical in the production of anything except straight wire, they do tend to lead to somewhat superior superelastic behaviors.

3.2 Aging

As pointed out earlier in these proceedings²⁷, Ni-rich Ni-Ti alloys decompose during aging through a series of more stable compounds: $Ni_{14}Ti_{11}$, Ni_4Ti_3 , Ni_3Ti_2 and finally Ni_3Ti . These precipitates strengthen the Ni-Ti matrix, which enhances superelasticity but also cause a shift in the transformation temperatures²⁸. Many of these effects are complex, but to illustrate general trends we will follow the aging of a Ni-Ti wire with 50.8 atomic percent Ni after solution treating at 850°C and water quenching. To simplify matters, we have chosen an aging temperature of 325°C and will not consider that to be a variable.

The shifts in the transformation temperatures during aging are shown in Figure 20. M_s increases steadily while A_f initially drops then again rises. The R-phase becomes evident after the first minute of aging, then R_s climbs rather quickly to nearly 60°C . On this basis alone one would expect room temperature superelasticity after aging one minute, and for superelasticity to disappear again after 100 minutes - at this point R_s

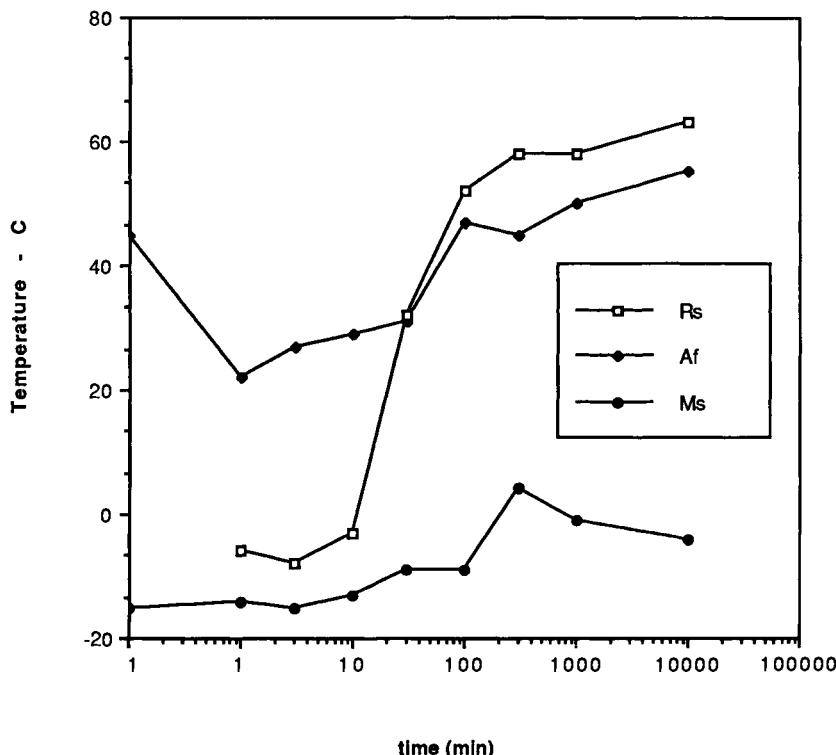


Figure 20: Aging effects on transformation temperatures are shown in a Ni-Ti wire with 50.8% Ni. The wire was annealed for 30 minutes in vacuum at 850°C and quenched in water prior to aging at 350°C for the time shown on the ordinate. Note that the reported transformation temperatures are under a 150 MPa stress.

would be above ambient. Figure 21 shows that the austenite is strengthened by aging, but the maximum strength levels are only moderate, and start to increase after 10 minutes. Also shown in Figure 21 is the dependence of stress rate upon aging time, showing a sharp reduction until about the 10 minute mark. If one compares the requirements of superelasticity discussed in section 3.0 with Figures 20 and 21, one would conclude that optimum superelasticity should be found after aging times of 10 to 100 minutes. Figures 22 and 23 verify this.

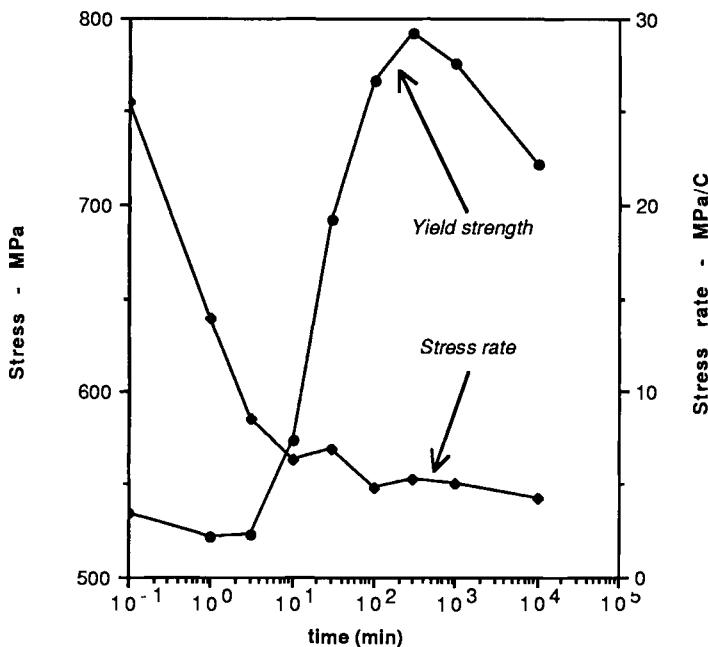


Figure 21 Age hardening of the wire described in the caption of Figure 20 is shown. Note that the stress rate change is marked and generally opposite to the resistance to slip.

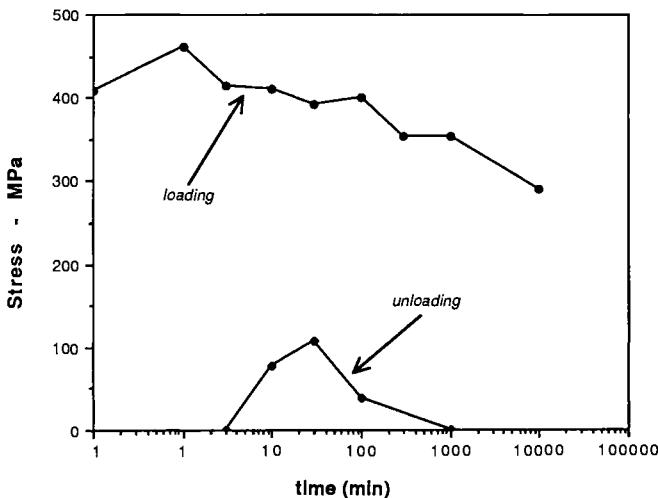


Figure 22 The loading and unloading plateau stresses are plotted as function of aging time, showing a general decrease in loading stiffness (commensurate with Figure 20), and a superelastic plateau that is only observable after 10 to 100 minutes of aging. The wire is again that described in the caption of Figure 20.

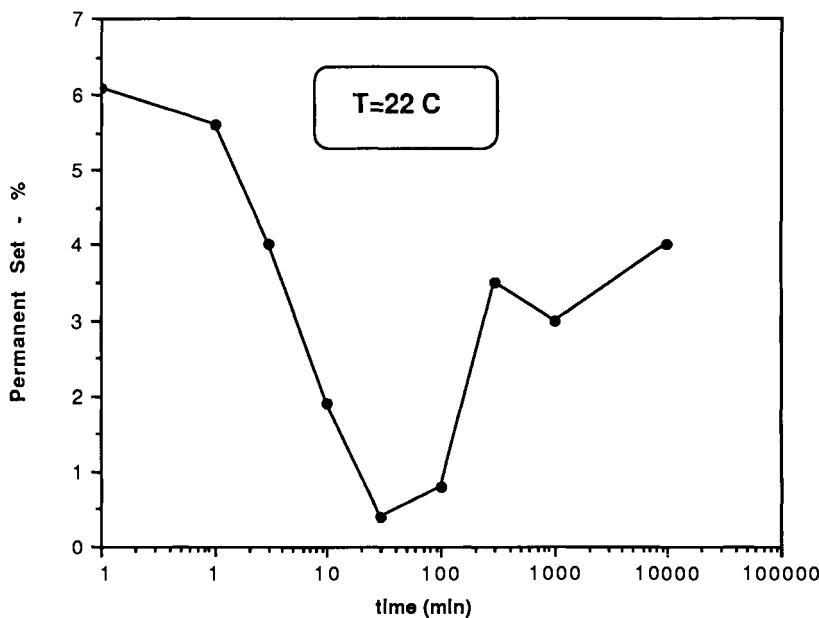


Figure 23 Aging effects on permanent set in the same wire described in Figure 20, again indicating superelastic effects only between 10 and 100 minutes of aging.

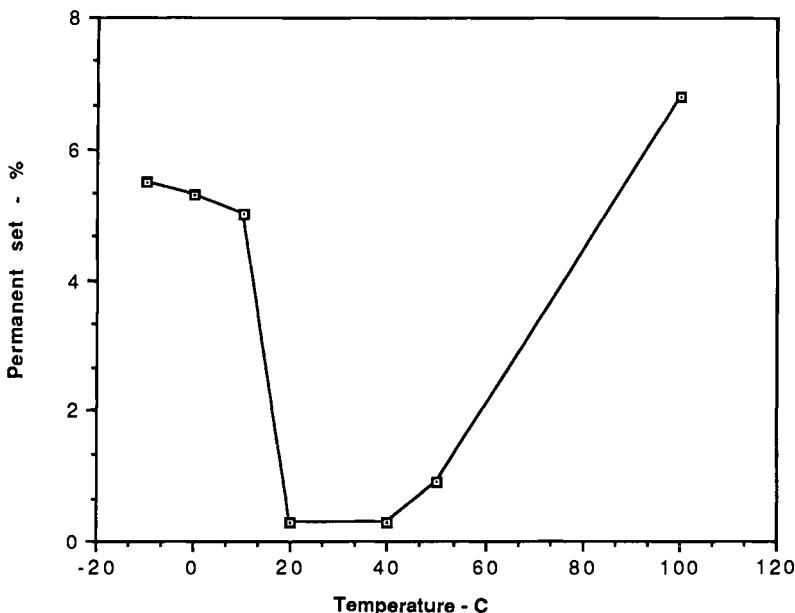


Figure 24 A more detailed look at the superelastic temperature window of a Ni-Ti wire with 50.8% Ni after aging 30 minutes at 350°C. The window is only 30°C wide.

Figure 22 shows the loading plateau is relatively constant (the slight decrease being commensurate with the increasing M_s in Figure 20), and that the unloading plateau stress reaches a maximum after 30 minutes of aging. With times less than 3 minutes, no unloading plateau is found. Figure 23 shows that the room temperature permanent set reaches a marked minimum between 30 and 100 minutes. Finally we can look at permanent set as a function of temperature (Figure 24). The window is substantially narrower than in the cold worked binary shown in Figure 12, but that is expected since the austenitic strength levels in the aged material are substantially lower.

Whether the aging or the cold working approach is better depends strongly upon application specifics. Generally speaking it is possible to achieve superelasticity with lower strength levels and greater ductilities through aging, but the windows are narrower. The material shown in Figure 24 exhibited over 60% elongation to failure; the material of Figure 12 only 10%. Also, the constant stress plateaus are longer and flatter in the aged material as a rule. Finally, greater stiffness can be obtained at room temperature in the cold worked material.

The above example is only that: an example. In alloys of greater Ni content, aging effects can be substantially faster and strengths higher. Also at higher temperatures aging can be more rapid. Still the trends are valid and can be generally applied. The choice of an optimum aging temperature may also depend upon the exact alloy and condition - Miyazaki et al²⁹ aged a 50.6% Ni alloy at 300, 400 and 500°C and concluded that 400°C was optimum. More recent work³⁰ suggests that better properties can be obtained by aging at 325°C to 350°C.

3.3 Compositional effects

Binary Ni-Ti alloys are limited to superelastic applications at room temperature and above - and at room temperature, their usefulness is quite limited by the rather low unloading stress plateau level. The reason is that it is difficult to lower A_f below room temperature while maintaining a high austenitic strength. To understand this, recall that to achieve a low M_s one goes to the Ni-rich side of the phase diagram. Here M_s is indeed decreased after quenching; but the alloys also become susceptible to aging effects which increase the transformation temperatures. In order to obtain the austenitic strength levels required for superelasticity, one must age the Ni-rich alloys, even if inadvertently during the recovery anneal following cold working. In addition to increasing M_s , aging introduces the R-phase and increases hysteresis. Thus room temperature represents a practical lower limit for superelasticity in binary Ni-Ti.

To obtain superelastic windows at low temperatures, or to increase the stiffness at room temperature, two alloying strategies have been used. The first is to add an M_s suppressing addition³¹ such as vanadium, cobalt, iron, aluminum or chromium. In proper proportions these additions suppress M_s without causing an instability to aging. Thus the cold working approach to increasing strength can be invoked while keeping both M_s and R_s below room temperature. An example of the benefits of such a modified alloy are shown in Figure 25: here the vanadium alloy is clearly stiffer

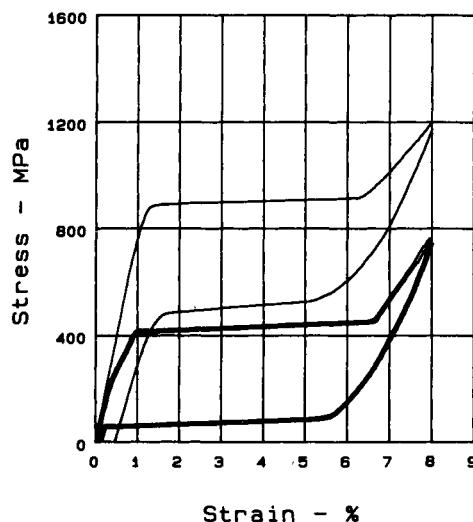


Figure 25 Third element additions such as vanadium suppress M_s and can lead to far greater superelastic stiffnesses and energy storage abilities. The case shown here is a Ni-Ti wire with 6 at.% V after cold working 40% and annealing at 500°C for one hour. The energy storage efficiency is more than 5 times that of the binary.

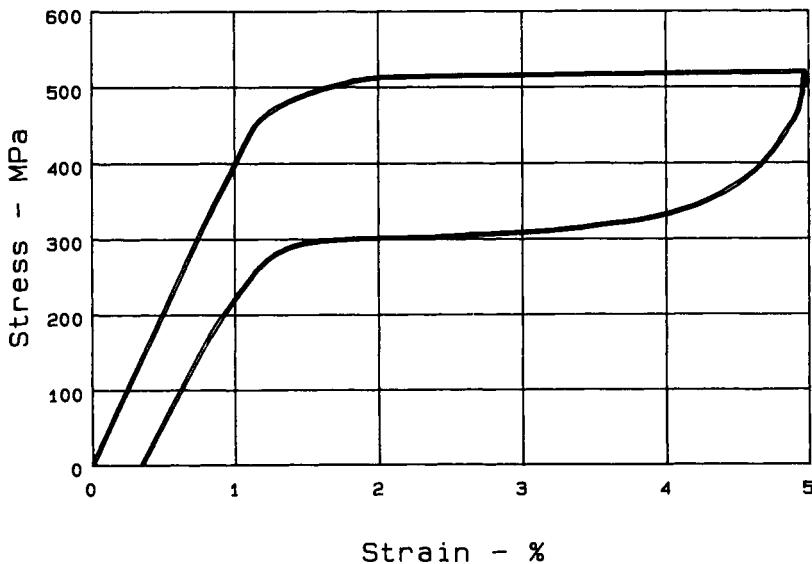


Figure 26 An extreme example of low temperature superelastic windows is illustrated by this Ni-Ti-Fe rod deformed at -196°C and still showing excellent superelasticity.

(higher unloading plateau stress) than the binary. In this case the stored energy efficiency is more than 5 times that of the binary. A still more exaggerated case is shown in Figure 26, where a Ni-Ti-Fe alloy is shown to be fully superelastic even at -196°C. Some care must be taken, however, since these doped alloys are often less ductile than the binaries. The second alloying approach is to add copper³². Copper has the advantage of reducing hysteresis and eliminating the R-phase. There is, however, little in the way of quantitative design data for these alloys.

4. Summary:

We have presented here an overview of pseudoelasticity, and more specifically of superelasticity in Ni-Ti. The key engineering parameters (stiffness, hysteresis, permanent set and stored energy) have been defined, showing that Ni-Ti has the theoretical capability of reducing the size and weight of springs by nearly an order of magnitude. The value of Ni-Ti in any particular application must be judged, however, with an understanding of what temperature variations are expected. Two approaches to producing superelasticity were presented: cold working and aging. Which may be preferred in a particular case would depend upon the alloy composition, the shape of the device to be made, and the properties desired.

References:

1. Otsuka, K., Shimizu, K., Int. Met. **31**, 3, 93 (1986).
2. Goo, E., Duerig, T.W., Melton, K., Sinclair, R., Acta Metall. **33**, 9, 1725 (1985).
3. Miyazaki, S., Otsuka, K., Suzuki, Y., Scripta Met. **15**, 287 (1981).
4. Otsuka, K., Sakamoto, H., Shimizu, K., Scripta Met. **11**, 41 (1977).
5. Lieberman, D.S., Shmerling, M.A., Katz, R.W., Shape Memory Effects in Alloys, J. Perkins editor, Plenum Press, 203 (1975).
6. Birnbaur, H.K. and Reed, T.A., Trans. AIME, **218**, 662 (1960).
7. Burkart, M.W. and Reed, T.A., Trans. AIME, **1516** (1953).
8. Zangwell, A., Bruinsma, R., Materials Research Society, 529 (1984).
9. Otsuka, K., Sakamoto, H., Shimizu, K., Scripta Met. **10**, 983 (1976).
10. Otsuka, K., Wayman, C.M., Nakai, K., Sakamoto, H., Shimizu, K., Acta Met. **24**, 207 (1976).
11. Yang, N.Y.C., Laird, C., Pope, D.P., Met. Trans. **8A**, 955 (1977).
12. Suoninen, E.J., M.I.T., PhD Thesis (1954).
13. Schroeder, T.A., Cornelis, I., Wayman, C.M., Met. Trans. **7A**, 535 (1976).
14. Groel, J.F., Wire Journal 53 (1971).
15. Adnyana, D.N., Springs 31 (1988).
16. Delaey, L., Janssen, J., Van De Mosselaer, D., Dullenkopf, G., Deruyttere, A., Scripta Met. **12**, 373 (1978).
17. Sugimoto, K., Nakamura, Y., Delaey, L., J. Physique **42** 10, CS-1019 (1983).
18. Melton, K.N., Mercier, O., Acta Metall. **29**, 393 (1981).
19. Aioldi, G., Rivolta, B., Turco, C., ICOMAT-86, Jap. Inst. Metals., 691 (1987).
20. Mukherjee, K., Sircar, S., Dahotre, N., Materials Sci. and Eng. **74**, 75-84 (1985).

21. Mai, C., Guenin, G., Morin, M., Livet, F., Gobin, P.F., Mat. Sci. and Eng. **45**, 217 (1980).
22. Miura, F., Mogi, M., Ohura, Y., Hamanaka, H., Ortho. & Dento. Orthopedics **90**, 1 (1986).
23. Hua-Cheng, T., Feng-Zhi, Y., Burstone, C.J., Qin, B., Proc. of SMA '86, 421-426 (1986).
24. Zadno, Z., Yu, W., Duerig, T.W., to be published.
25. Wosilewski, R.J., Met. Trans. **2A**, 2973 (1971).
26. Miyazaki, S., Imai, T., Otsuka, K., Suzuki, Y., Scripta Met. **15**, 853 (1981).
27. Melton, K.N., this volume
28. Nishida, M., Wayman, C.M., Mater. Sci. Eng. **93**, 191 (1987).
29. Miyazaki, S., Ohmi, Y., Otsuka, K., Suzuki, Y., Journal De Physique (1982).
30. Zadno, R., Duerig, T.W., Unpublished research, Raychem Corp. (1986).
31. Otsuka, K., Miyazaki, S., Spec. Pro. Res. on Energy in aid of Sci. Reserch of Ministry of Educ. Sci. Culture, Japan, SPEY 14, 51-58 (1985).
32. Zakharova, N.N., Kuz'min, S.L., Likhachev, V.A., Phys. Metals **3** 5, 840 (1981).

Thermal and Stress Cycling Effects and Fatigue Properties of Ni-Ti Alloys

S. Miyazaki

Institute of Materials Science

University of Tsukuba

Tsukuba, Ibaraki 305, Japan

In many applications of SMA's, the shape memory function is used many times by heating and cooling and/or loading and unloading. Therefore, the stability of material characteristics, i.e. transformation temperatures and mechanical behaviour, during thermal or stress cycling is an important considerations in developing suitable alloys. Although shape memory alloys can be subjected to both thermal and stress cycling, it is necessary to investigate these effects separately in order to development a basic understanding of each.

Superelasticity (or pseudoelasticity) is also one of the most important properties of the shape memory alloys. Since superelasticity appears only by loading and unloading above A_f , thermal cycling is not involved and stress-cycling effects are the main consideration. By cyclically applying stress, the alloys finally show fatigue failures, which place a limit upon practical uses. The present paper reviews the above cycling effects and fatigue life of Ni-Ti alloys, and also indicates how one can suppress cycling effects and extend superelastic fatigue life.

1 Thermal Cycling Effects

Since solution treated conditions of Ni-Ti have no special internal structures such as dislocations and/or precipitates, they are highly susceptible to thermal cycling effects¹. Figure 1 shows the electrical resistance vs. temperature curves as a function of the number of thermal cycles for three solution treated Ni-Ti alloys, with 49.8, 50.6 and 51.6 at.% Ni². The arrows indicate the transformation temperatures: T_i , (the "pre-martensitic" incommensurate transformation start temperature) M_s and M_f . All solution treated specimens show M_s and M_f decreasing during thermal cycling, while T_i is constant. The height of the peak in the electrical resistance vs. temperature curve also increases during thermal cycling. A more quantitative representation of the thermal cycling effect on the transformation temperatures is shown in Figure 2, where T_i , M_s and M_f are plotted against the number of thermal cycles for a Ti-49.8 at% Ni alloy². Since the decrease in M_f is greater than that in M_s , the temperature difference ($M_s - M_f$) increases during cycling. It is also noticed that M_s and M_f decrease rapidly during the initial stage of thermal cycling, while they change more gradually in the later stage.

The phenomena resulting from thermal cycling are summarized as follows:

- (a) decrease in M_s and M_f ²⁻⁸

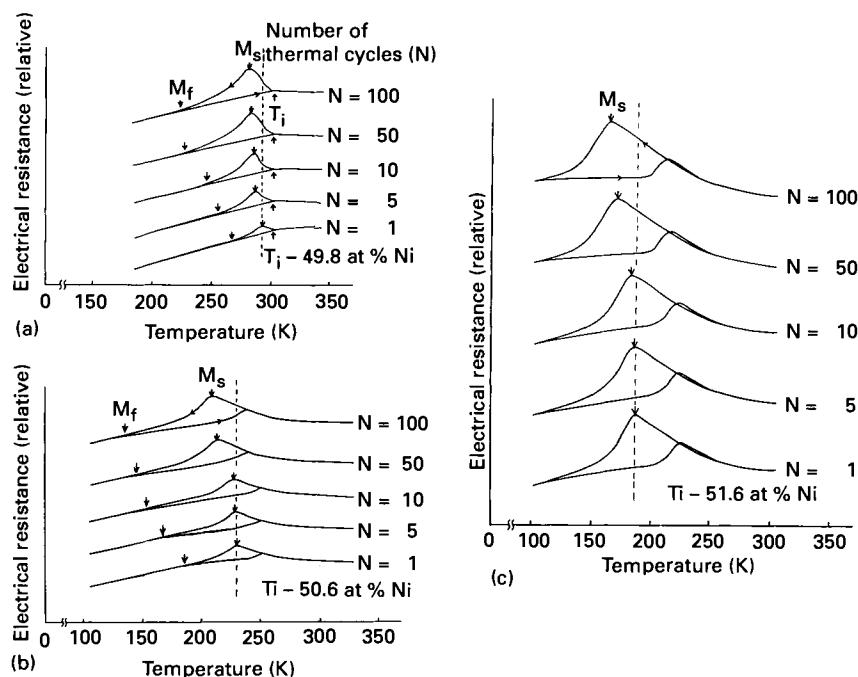


Figure 1: Effect of thermal cycling on the electrical resistance vs. temperature curve for (a) Ti-49.8at%Ni, (b) Ti-50.6at%Ni and (c)Ti-51.6at%Ni alloys which were solution-treated at 1000°C for one hour followed by quenching into ice water.

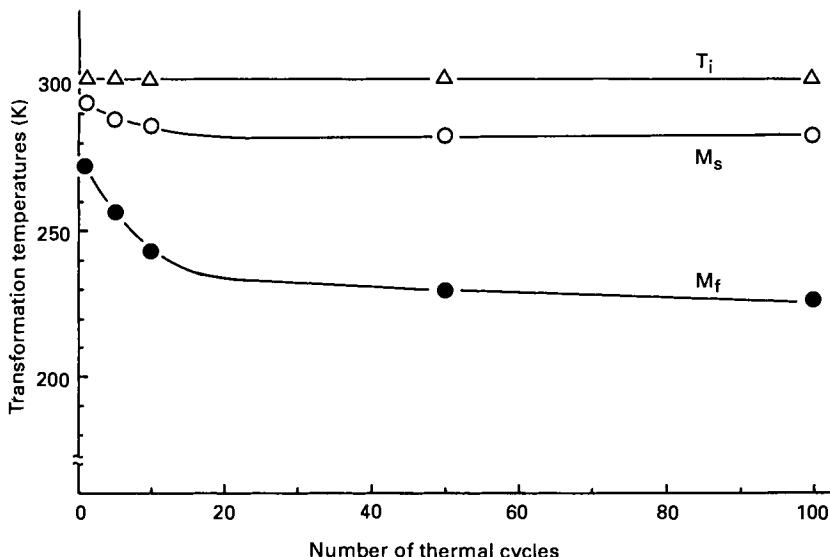


Figure 2: Transformation temperatures (T_i , M_s and M_f) as a function of the number of thermal cycles for the Ti-49.8at%Ni alloy which was solution-treated at 1000°C for one hour followed by quenching into ice water.

- (b) increase in the temperature difference $(M_s - M_f)^2$
- (c) rapid change in M_s and M_f during initial cycling stage followed by a slow change in the later cycling stage²
- (d) increase in the peak height of the electrical resistance vs. temperature curve²⁻⁷
- (e) constancy of T_i ²

Among these, (d) can be attributed to (a) and (e): since M_s decreases while T_i remains constant, the temperature range where the incommensurate phase exist widens while, the electrical resistance keeps increasing upon cooling between T_i and M_s .

1.1 Causes for Thermal Cycling Effects

Transmission electron microscopy reveals that dislocations are introduced during thermal cycling, as shown in Figure 3(a)². With further thermal cycling, dislocations tangle and the density increases as shown in Figures 3(b) and (c). The stress field

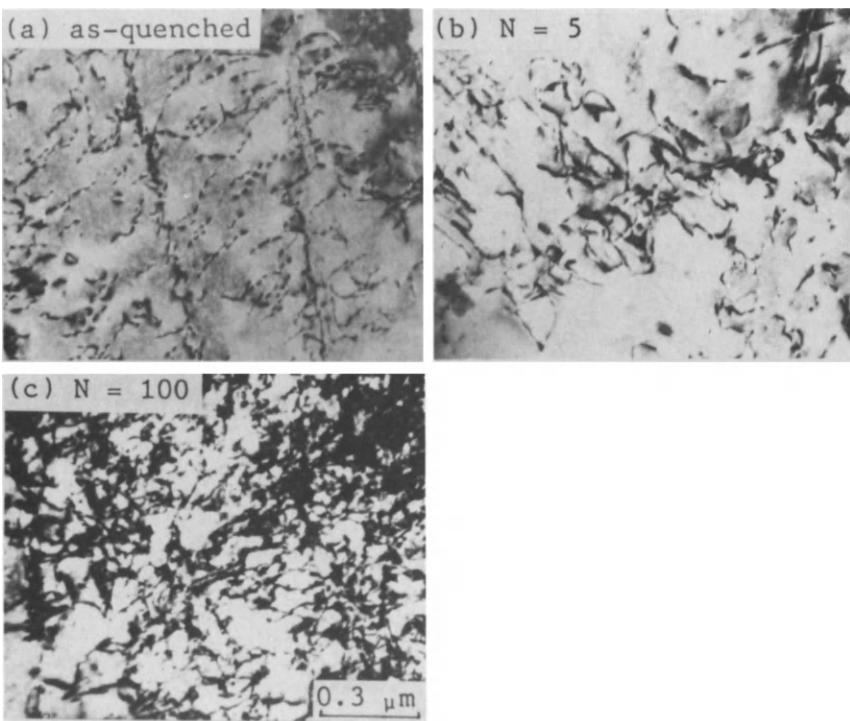


Figure 3: Transmission electron micrographs of changes of internal structure due to thermal cycling between temperatures below M_f and above A_f in a Ti-49.8at%Ni alloy which was solution-treat at 1000°C for one hour followed by quenching into ice water: (a) as-quenched ($N=1$ cycle), (b) $N=5$ cycles and (c) $N=100$ cycles.

formed by these dislocations evidently has the effect of suppressing the martensitic transformation, since M_s and M_f decrease. The increase in the temperature difference ($M_s - M_f$) can also be explained in terms of the internal stresses due to these dislocations: the intensity of the internal stresses decreases with the distance from the dislocations, and thus the effective transformation temperature differs from place to place in the specimen. The increase in dislocation density with thermal cycling causes the introduction of dislocations to become more difficult due to work hardening. This is consistent with the fact that M_s and M_f decrease rapidly during initial cycling but change more gradually later. If the stress field formed by dislocations is assumed to be the cause for the thermal cycling effects, it also consistently explains the fact that T_i remains constant during thermal cycling: the incommensurate phase transformation is not accompanied by any lattice distortion and thus T_i is insensitive to the stress field. Other conceivable causes for thermal cycling effect, e.g. change in the degree of order and aging effects which occur in Cu-based shape memory alloys, cannot consistently explain all the thermal cycling effects mentioned above for Ni-Ti alloys. Moreover, such causes will not occur during thermal cycling in Ni-Ti alloys².

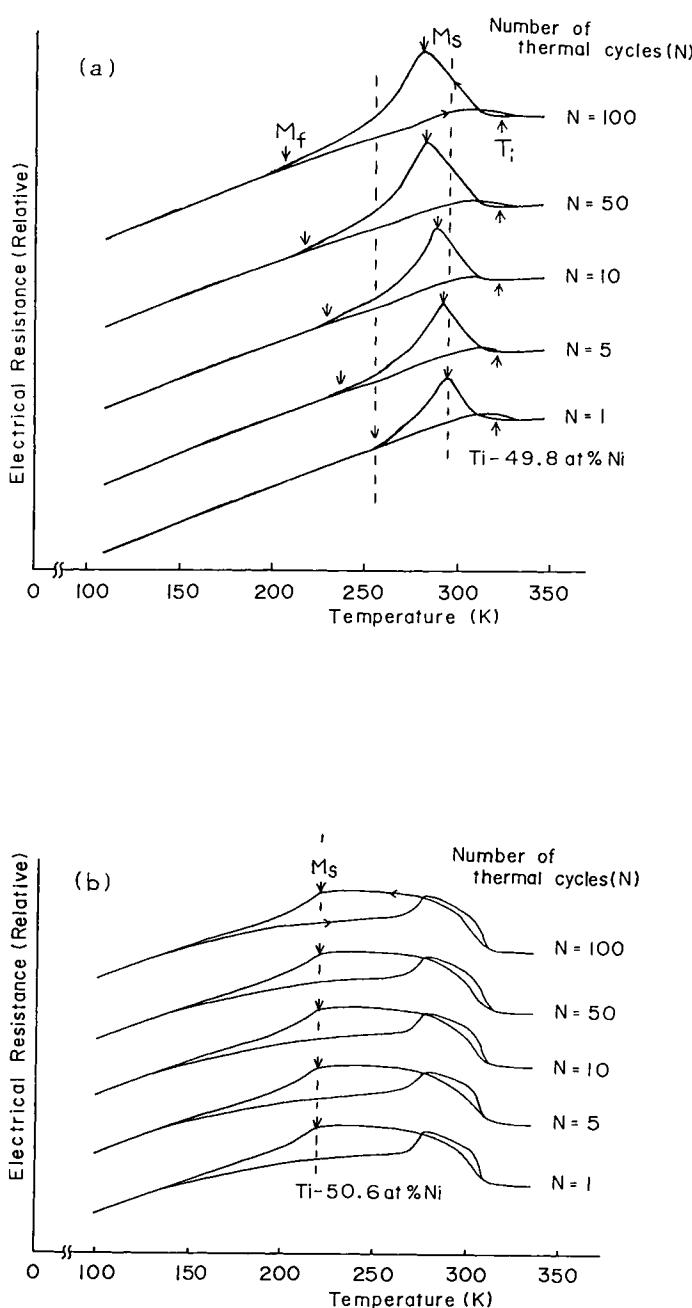
1.2 Stabilization of Transformation Characteristics against Thermal Cycling

The above conclusion is that the cause for the thermal cycling effects in Ni-Ti alloys is the introduction of dislocations. It is expected then, that thermal cycling effects can be suppressed by introducing fine internal structures which can raise the critical stress for slip deformation.

Figure 4 shows the effect of thermal cycling on three aged Ni-Ti alloys. The 49.8 at% Ni alloy shows thermal cycling effects similar to those of solution treated specimens, while the other alloys show no change in transformation characteristics. Transmission electron microscopy revealed fine Ti_3Ni_4 precipitates in the latter alloys, with no special internal structure in the former. Ti_3Ni_4 can be formed in Ni-Ti alloys with more than 50.5 at% Ni, therefore it is clear that fine precipitates are effective in stabilizing the martensitic transformation temperature against thermal cycling. By aging at higher temperatures, these precipitates grow and their density becomes lower, resulting in a decrease of the critical stress for slip. In such cases, M_s decreases with increasing thermal cycling, while T_R , the rhombohedral transition temperature, increases⁸.

Annealing at temperatures which are lower than the recrystallization temperature thermally rearranges dislocations which were introduced by the preceding cold work. Such thermally rearranged dislocations can also raise the critical stress for slip, resulting in stable transformation characteristics against thermal cycling. This is the case even in Ti-49.8 at% Ni (Figure 5), where the transformation temperatures do not change with thermal cycling.

It is also interesting to examine whether dislocations introduced by tensile deformation also stabilize the transformation characteristics against thermal cycling. Figure 6 shows the change in M_s due to the thermal cycling of solution treated Ti-49.8 at% Ni after pre-straining². Pre-strained specimens show an increase in M_s while unstrained specimens showed a decrease in M_s . The change in M_s decreases with increasing pre-strain; this indicates that the effect of thermal cycling is also suppressed by work hardening. In case of a small pre-strain, new dislocations may be induced and the existing dislocations may be rearranged by thermal cycling. This will lead to a change in the internal stress field, shifting M_s . In a heavily work hardened state, dislocations



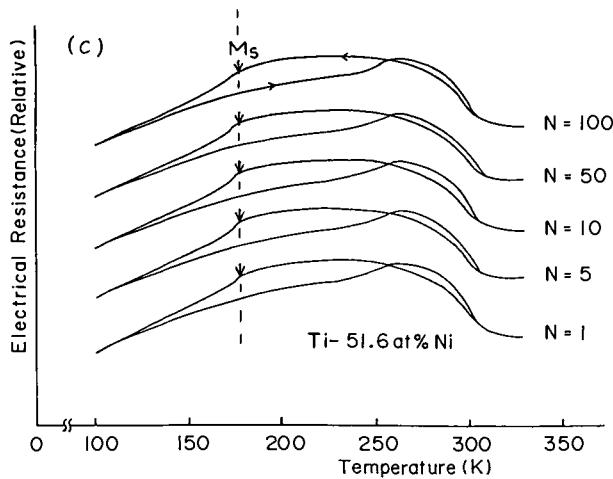


Figure 4: Effect of thermal cycling on the electrical resistance vs. temperature curve for (a) Ti-49.8at%Ni, (b) Ti-50.6at%Ni and (c) Ti-51.6at%Ni alloys which were age-treated at 400°C for one hour after solution-treatment at 1000°C.

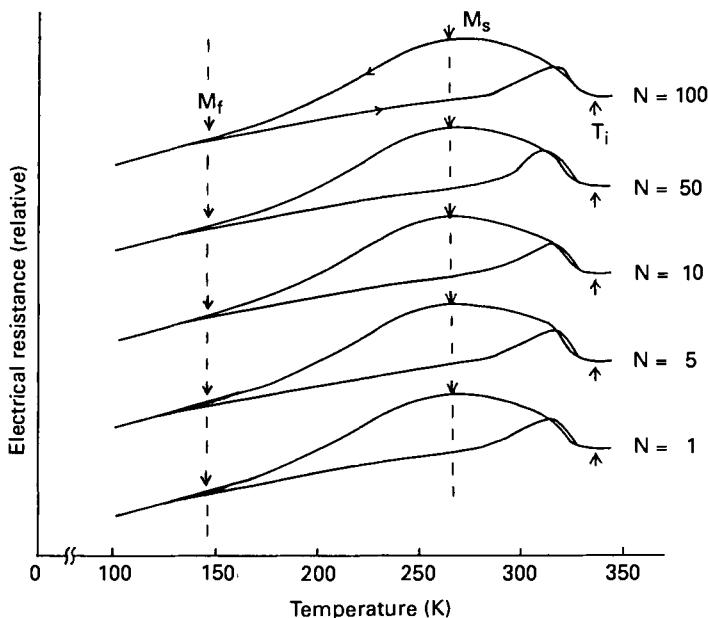


Figure 5: Effect of thermal cycling on the electrical resistance vs. temperature curve for Ti-49.8at%Ni alloy which was annealed at 1000°C for one hour without preceding solution-treatment after cold work.

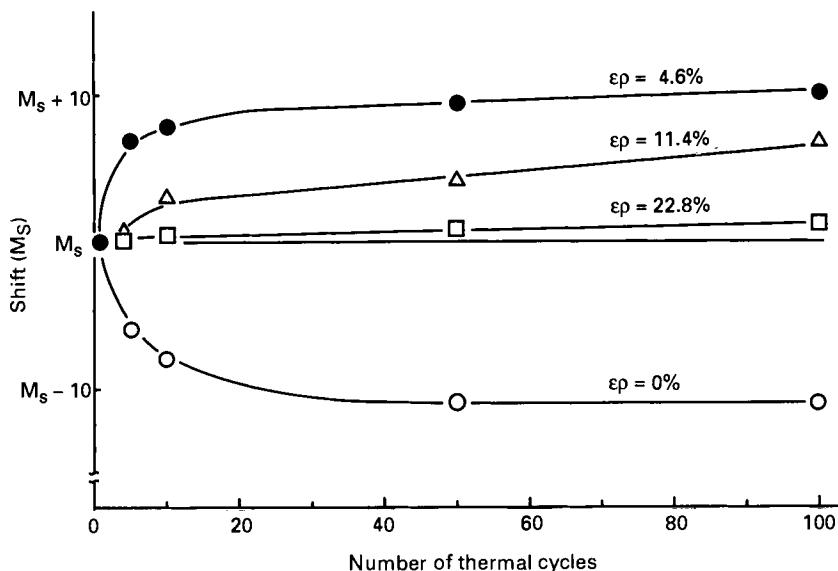


Figure 6: Temperature shift in M_s due to thermal cycling for specimens with various pre-strains.

may hardly move during thermal cycling, and thus M_s becomes insensitive to thermal cycling.

The above results on thermal cycling effects are summarized in Table 1, which shows that fine Ti_3Ni_4 precipitates and a high density of dislocations are effective in suppressing thermal cycling effects.

Table I: Effects of Thermal Cycling on the Ni-Ti Alloys

	Ni-Content (at%)	Heat-Treatment	Changes by Thermal Cycling	Internal Structure
(1)	49.8		YES	NONE
	50.6	1000°C+IQ	YES	NONE
	51.6		YES	NONE
(2)	49.8		YES	NONE
	50.6		NO	PRECIPITATES
(3)	51.6	1000°C+IQ → 400°C+IQ	NO	PRECIPITATES
	49.8	400°C + IQ	NO	TRD
(4)	49.8	1000°C+IQ, followed by prestraining beyond 20%	NO	SWHS
(5)	49.8	1000°C+IQ, followed by prestraining to 4.6%	YES	WWHS

TRD: Thermally Rearranged Dislocations

SWHS: Strongly Work Hardened Structure

WWHS: Weakly Work Hardened Structure

2 Stress Cycling Effects

In order to evaluate the effects of cyclic stresses on superelasticity, it is convenient to first characterize superelasticity itself. Figures 7b, c and d show schematic stress-strain curves representing superelasticity, while Figure 7a shows elasticity in a normal

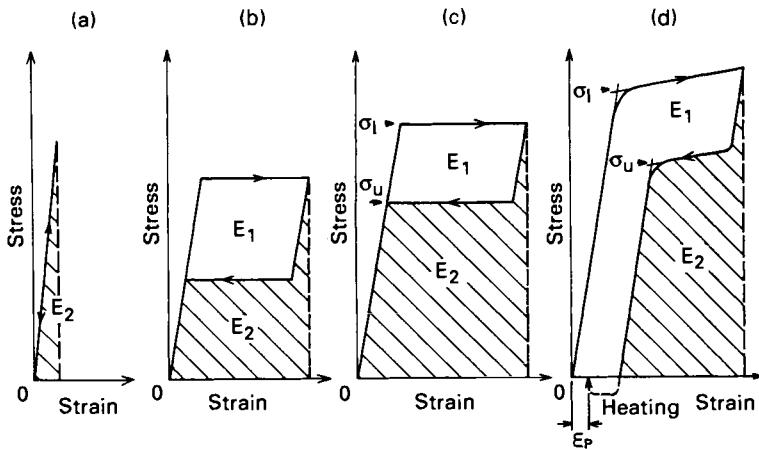


Figure 7: Schematic stress-strain curves representing (a) the elastic deformation and (b)-(d) the superelasticity characteristics.

material. In the figures, σ_1 is the critical stress for inducing martensites. E_1 is the energy density which is dissipated during one cycle, while E_2 the energy density per unit volume which is stored and available upon unloading. Since the total work done on a specimen is (E_1+E_2) , the efficiency of energy storage, η , is defined $E_2/(E_1+E_2)$. By comparing Figures 7b and 7c, it is understood that both E_2 and η increase with increasing σ_1 . However, in case σ_1 exceeds the critical stress for slip (σ_s), superelasticity becomes incomplete since a permanent residual strain (ϵ_p) is introduced (as shown by Figure 7d). Thus it is also important to raise σ_s in order to improve the superelasticity characteristics.

2.1 Effects of Cyclic Stress

Figure 8 shows the effect of cyclic stresses on the stress-strain curve of a Ti-50.5 at.% Ni alloy; the stress is controlled by choosing a suitable test temperature. General features of stress cycling effects are that residual strain increases (as shown by the deviation of the starting points of the stress-strain curves), the critical stress for inducing martensites decreases (as shown by the arrows), and the strain or stress hysteresis becomes smaller.

The cause for the residual strain is probably the occurrence of slip during the preceding deformation (as pointed out by Melton and Mercier⁹). Internal stresses formed by slip may assist the formation of the stress-induced martensite; thus the critical applied stress σ_1 decreases. As the stress-induced transformation occurs like a

Lüders deformation in stage I¹⁰, the stress for inducing martensite is constant for the first cycle. However, since the internal stress field after cyclic deformation has a gradient in strength, σ_i increases with increasing strain even in stage I. As a consequence, the hysteresis becomes smaller by cyclic deformation. Figure 8 also

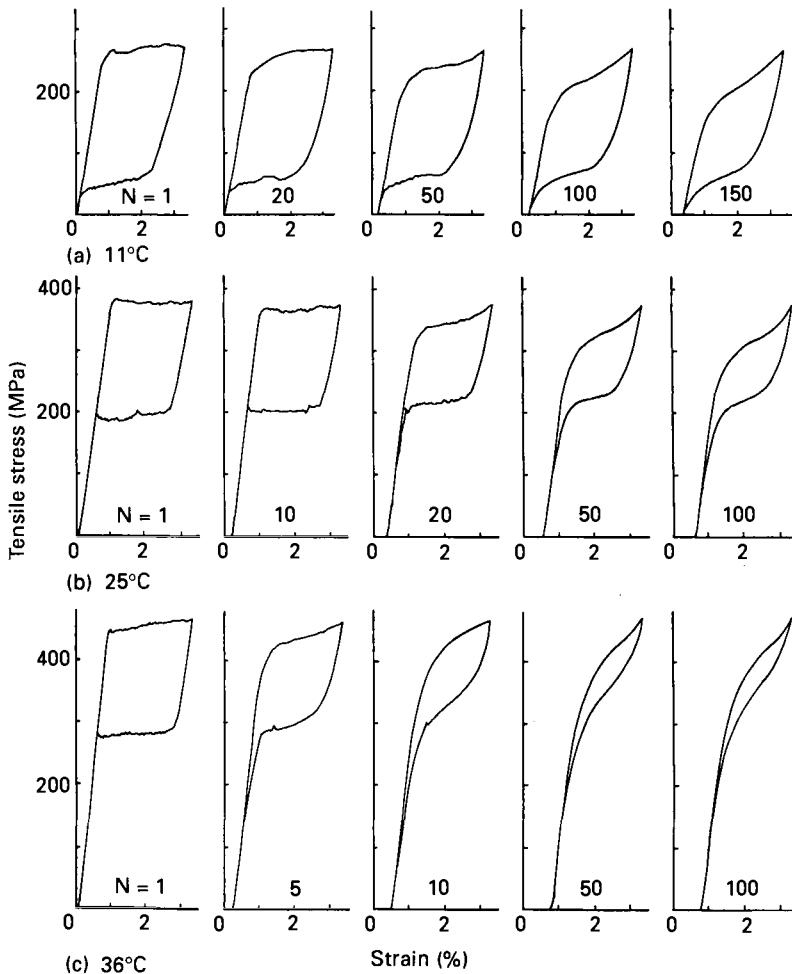


Figure 8: Effect of cyclic deformation on stress-strain curves at various temperature in a Ti-50.5at%Ni alloy.

shows that the degree of change in the residual strain, the critical stress for inducing martensites and the strain hysteresis increase with increasing cyclic stress.

The above cycling effects are shown quantitatively in Figure 9, where 9a, 9b and 9c correspond to the residual strain, the critical stress for inducing martensites and the strain hysteresis, respectively. However, all these values become insensitive to cycling with increasing number of cycles, i.e. a steady state of the superelasticity characteristics exists corresponding to the cyclic stress level. This fact can be explained by the work hardening due to introduction of dislocations during cycling.

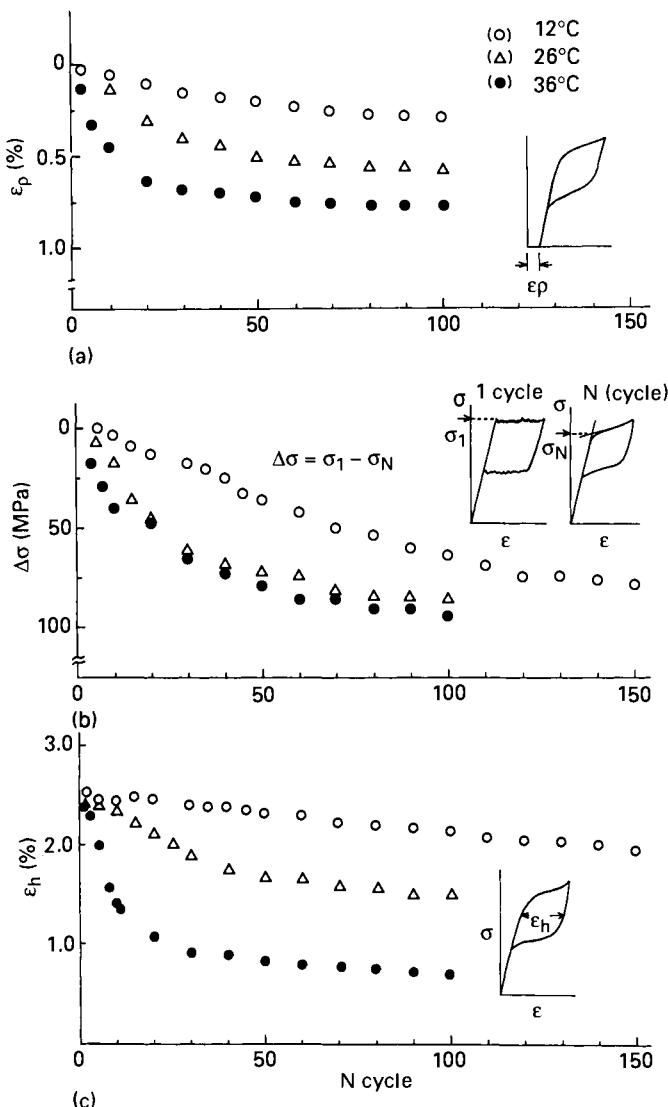


Figure 9: Effect of cyclic deformation on (a) residual strain, (b) critical stress for inducing martensites, and (c) strain hysteresis.

Therefore, it is important to raise the critical stress for slip in order to stabilize the superelasticity characteristics against cyclic deformation. The critical stress for slip is sensitive to the internal structure, which depends on thermo-mechanical treatment and Ni-content^{11,12}; this point will be described in more detail in the following section.

2.2 Stabilization of Superelasticity

As mentioned above, two mechanisms are available to raise the critical stress for slip in Ni-Ti: precipitation hardening and hardening due to a high dislocation density.

Transmission electron microscopy reveals the internal structures developed in Ni-Ti alloys. Figure 10 shows the effect of aging temperature on the internal structure of a Ti-50.6 at.% Ni alloy which was solution treated at 1000°C and aged for one hour at 400°C, 500°C and 600°C. Figures 10a and 10b show fine precipitates. Diffraction patterns also support the above observation, i.e. the small diffraction spots in 10a and 10b come from the precipitates. Comparison of micrographs in 10a and 10b shows

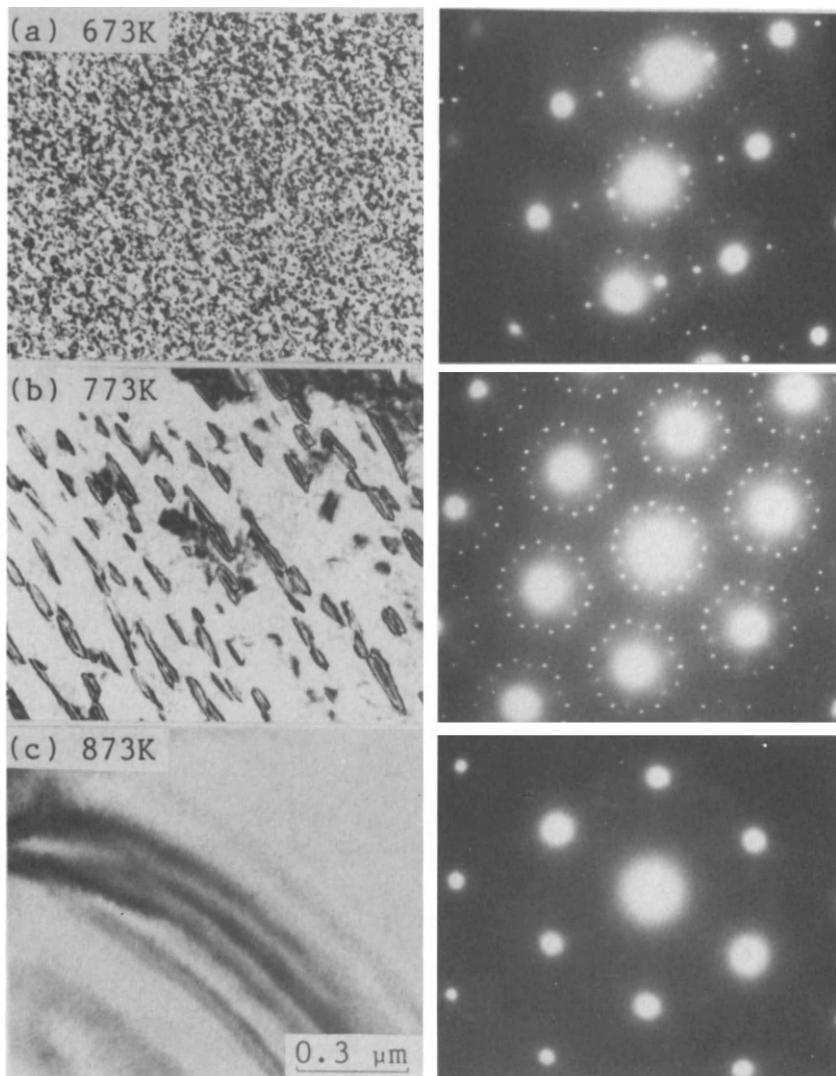


Figure 10: Effect of aging temperature on the internal structure of a Ti-50.6at%Ni alloy which was aged for one hour after solution-treatment, the aging temperature being (a) 400°C, (b) 500°C or (c) 600°C.

that aging at 400°C is more effective than aging at 500°C in order to derive a higher critical stress for slip, because the density of the precipitates is higher in the former. If a specimen is aged at a temperature below 400°C, precipitation is insufficient for suppressing slip. Therefore, the optimum aging temperature is around 400°C.

In order to retain and thermally rearrange dislocations which were introduced during preceding cold work, annealing at intermediate temperatures is necessary. Figure 11 shows that a high density of dislocations remains in a specimen annealed at 400°C (as shown in 11b) similarly to those in an as-rolled specimen (as shown in 11a). However, such dislocations annihilate and recrystallization occurs by annealing at 500°C (as shown in 11c), where many small recrystallized grains can be seen. Annealing at 600°C makes such recrystallized grains grow (as shown in 11d). Therefore, the optimum annealing temperature for thermally rearranging dislocations which were introduced during preceding cold work is around 400°C.

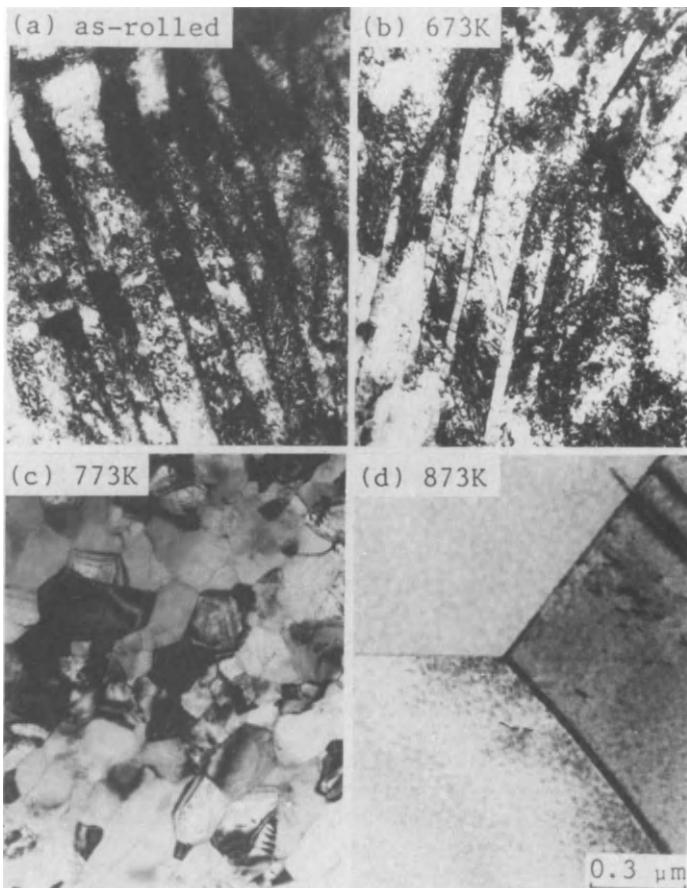


Figure 11: Effect of annealing temperature on the internal structure of a Ti-49.8at%Ni alloy which was annealed for 3.6ks without preceding solution-treatment after cold work, the annealing temperature being (b) 400°C, (c) 500°C or (d) 600°C.

Since the optimum aging temperature and the optimum annealing temperature are almost the same, a combined effect of the fine precipitates and a high density of thermally rearranged dislocations can be expected in a Ni-rich specimens annealed at 400°C without a preceding solution-treatment. The critical stress for slip for such conditions is as high as 620 MPa. The superelastic characteristics of these Ni-Ti alloys are listed in Table II.

Table II. Superelasticity in Ni-Ti after various Heat Treatments

Ni-Content (at.%)	Heat Treatment	σ_s (MPa)	$E_2(\text{max})$ (MJ/m ³)	η (max)
(1) 50.6	1000°C + IQ	110	8.9	0.18
	1000°C + IQ → 500°C + IQ	320	15.1	0.40
	1000°C + IQ → 400°C + IQ	470	33.3	0.68
(2) 49.8	1000°C + IQ → 300°C + IQ	320	20.5	0.50
	1000°C + IQ → 400°C + IQ	<100	3.8	0.21
	" "	230	6.7	0.27
(3) 50.1	" "	470	33.3	0.68
	" "	650	37.7	0.78
	1000°C + IQ	<100	6.8	0.18
(3) 50.6	500°C + IQ	280	15.7	0.48
	400°C + IQ	500	31.5	0.70
	1000°C + IQ → 400°C + IQ	<100	3.8	0.21
(4) 50.6	400°C + IQ	620	42.2	0.81

Figure 12 shows the effect of thermo-mechanical treatment on the change of stress-strain behavior during stress cycling; (a) and (b) represent aged specimens including precipitates, and (c) represents an annealed specimen including dislocations. The plateau in the first stress-strain curve changes to a gradually increasing slope during cycling in (a) and (b), while the plateau is retained even after 100 cycles in (c). It is not clear at present what causes this difference.

As mentioned above, the combined effect of precipitates and dislocations is available in Ni-rich specimens annealed at around 400°C without a preceding solution-treatment. Figure 12 (d) shows such a case. The change in stress-strain behavior is small compared with the other cases (a, b and c), showing that the combined effect is effective in stabilizing the superelastic characteristics against stress cycling.

Although the change in the superelasticity is rapid during initial cycling, the change after 100 cycles is hardly noticeable. Thus, we can anticipate to deriving a more stable superelasticity from cyclic training prior to service. The result of this is shown in Figure 12(e), where N' indicates the number of cycles after the preceding 100 cyclic deformation. The stress-strain curve hardly shows any change, even after 100 cycles.

The data of Figure 12 are normalized by the value of the first cycle, and shown in Figure 13 as a function of the number of cycles. Figure 13(a) shows the critical stress for inducing transformation, (b) the available stored energy and (c) the efficiency of energy storage. The first two decrease with increasing cycles, while the third one increases. On the basis of these results, it is concluded that the combined effect accompanied by the prior training is effective in stabilizing the superelasticity characteristics.

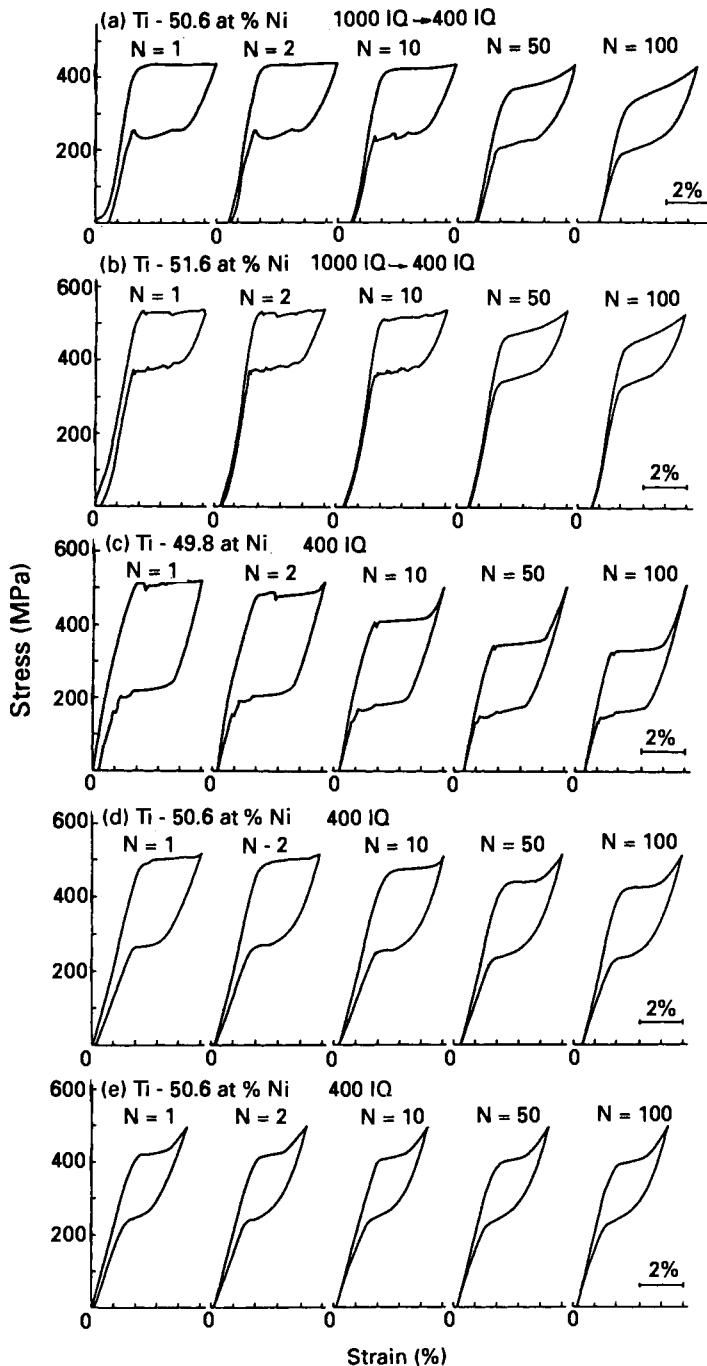


Figure 12: Effect of stress cycling on the stress-strain curves of Ni-Ti alloys after various heat-treatments.

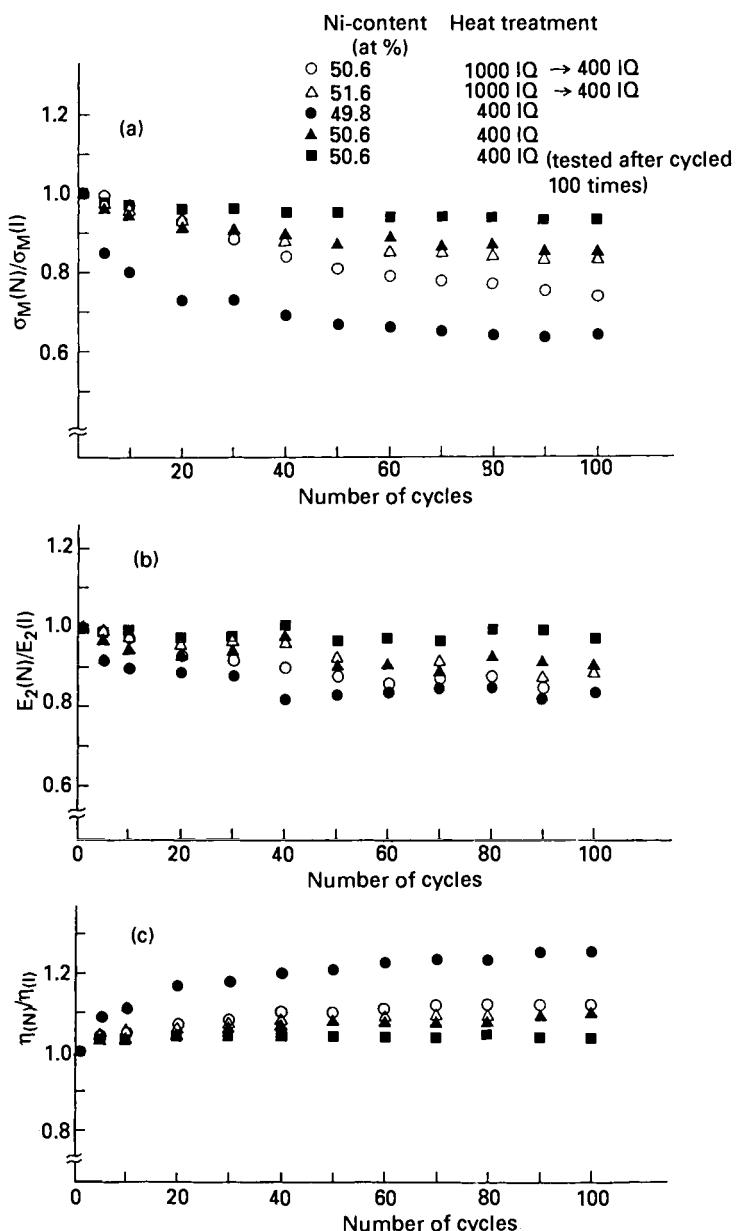


Figure 13: Effect of stress cycling on the superelasticity characteristics of Ni-Ti alloys after various heat-treatments.

3 Fatigue Properties

As mentioned in the above sections, both thermal and stress cycling effects can be effectively suppressed by applying special thermo-mechanical treatments which raise the critical stress for slip. However, by applying a cyclic stress exceeding a certain level, the Ni-Ti alloy shows fatigue failure. Therefore, it is important to understand the fatigue characteristics of Ni-Ti and then to improve the fatigue life in order to use this alloy in many practical applications.

3.1 Fatigue Life

The fatigue test mode for normal materials commonly consists of a tension-compression cycling, because they are subjected to such a cycling modes when used as structural materials. Melton and Mercier have reported fatigue data obtained in such a cycling mode⁹. However, Ni-Ti is a functional material characterized by the superelastic and shape memory effects; the fatigue mode appropriate to such a material in use is commonly a loading-unloading cycling with or without heating after each unloading.

In shape memory alloys, it is necessary to know how the martensitic transformation affects the fatigue life. Since in general stress is one of the important factors which determine the fatigue life, it is important to consider that the critical stress for inducing the martensitic transformation depends on test temperature under a condition of fixed M_s .

Figure 14 shows the effect of test temperature on the fatigue life of Ti-50.8at%Ni which was annealed at 400°C for 1 hour after cold work without a preceding solution treatment. In order to understand the basic characteristics of the cyclic tensile stress vs. fatigue life relationship (S-N curve), the data obtained at 50°C will be discussed first. The S-N curve is represented by two straight lines, the slope of the line in the

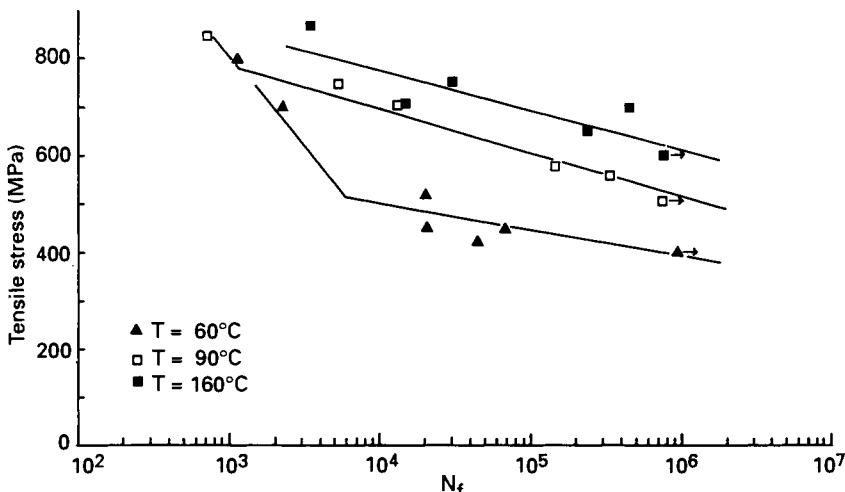


Figure 14: Effect of test temperature on the fatigue life of a Ti-50.8at%Ni alloy, which was annealed at 400°C for one hour and then cooled in a furnace gradually.

shorter life region being steeper than that in the longer life region. The intersection of the two lines nearly corresponds to the critical stress for inducing martensites at the test temperature.

The intersection of the two fatigue life lines becomes higher as shown by the S-N curve obtained at 90°C. However, the fatigue life becomes almost insensitive to test temperature above M_d , above which no martensite can be induced. The fatigue life at 160°C is presumably such a case, since the test temperature is well above A_f . The fatigue life at 160°C is excellent: 10⁶ cycles at 600 MPa.

Figure 15 shows the S-N curve at room temperature for the Ti-50.8at%Ni alloy subjected to various types of thermo-mechanical treatments. If we are concerned with

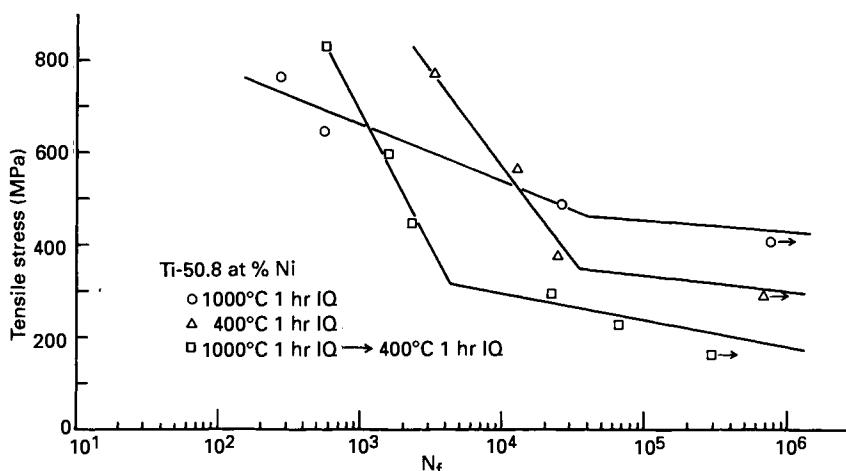


Figure 15: Effect of heat-treatment on the fatigue life a Ti-50.8at%Ni alloy which was tested at room temperature.

superelasticity, we need to concentrate our attention on the shorter fatigue life regions of an age-treated specimen (1000°C IQ → 400°C IQ) and an annealed specimen (400°C IQ). The fatigue life for the latter is about ten times that of the former. It is concluded therefore that the combined effect of precipitates and dislocations is also effective in improving the fatigue life to a certain extent.

3.2 Directions Towards Improved Fatigue Life

As mentioned in the preceding section, the fatigue life for cyclic elastic deformation is high enough in Ni-Ti alloys. Moreover, the fatigue life for cyclic loading induced by thermal cycling was found to be more than 10⁷ cycles¹³. However, the fatigue life for cyclic superelastic deformation is still insufficient for many applications^{14,15}. The special thermo-mechanical treatments mentioned above do not result in a sufficient superelastic fatigue life. In order to obtain a clue for developing materials with a long fatigue life, it is important to know the nucleation sites of fatigue cracks and the crack nucleation mechanism. Scanning electron microscope observation of the fractured surface revealed that fatigue cracks can nucleate at TiC inclusions in Ni-rich specimens, as shown in Figure 16 showing a TiC inclusion at the crack nucleation site.

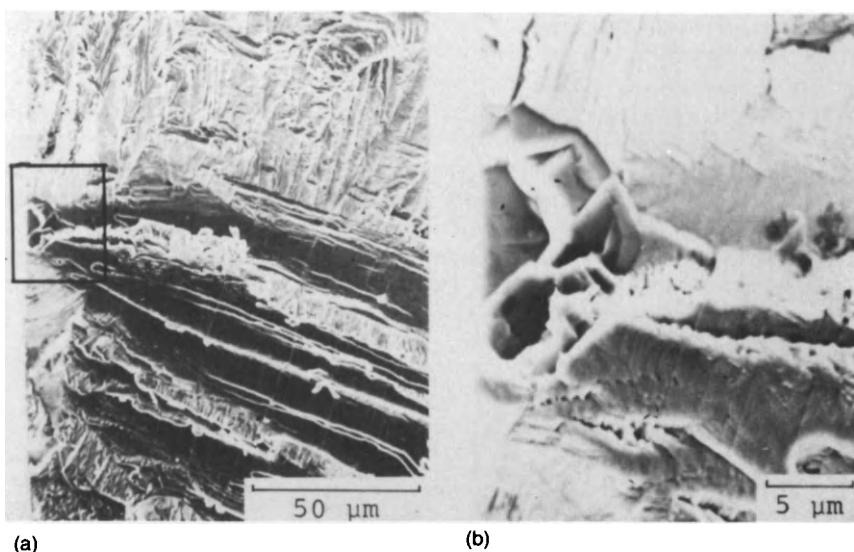


Figure 16: Scanning electron micrographs of a fatigue fracture surface of a Ti-50.8at%Ni alloy, where a TiC inclusion can be seen.

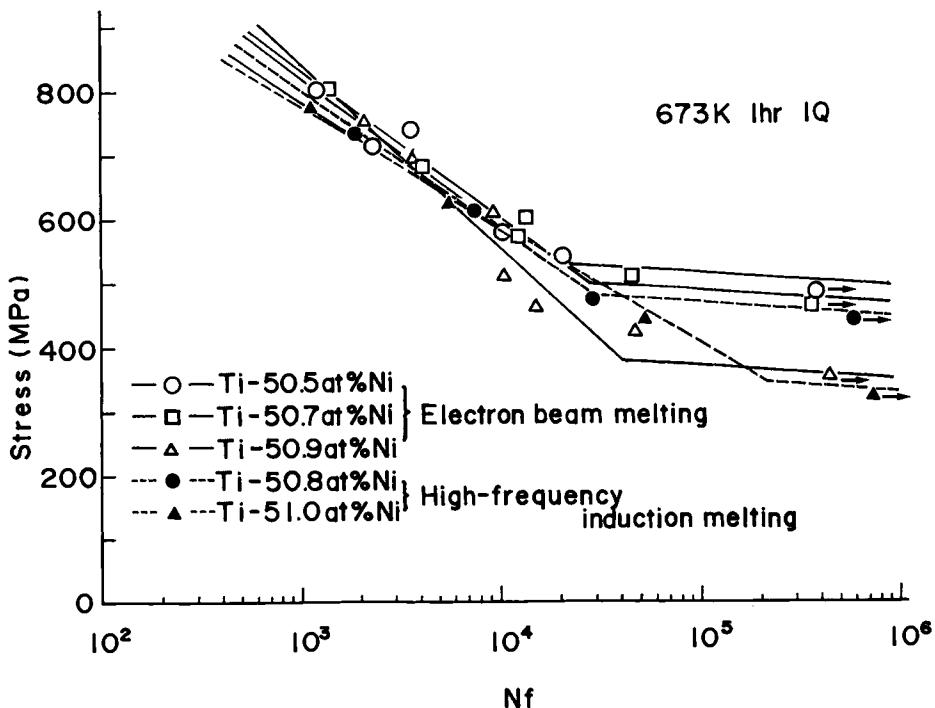


Figure 17: Fatigue lives of Ni-Ti alloys which were made by electron beam melting and high-frequency vacuum induction melting.

If a specimen does not contain TiC inclusions, it is natural to expect that the fatigue crack nucleation will be delayed even in Ni-rich specimens. In fact, no TiC inclusions were observed at the crack nucleation site in such a case. The fatigue lives of Ni-Ti alloys produced by high frequency induction melting in a carbon crucible and electron beam melting method are compared in Figure 17. Contrary to expectations, fatigue lives for superelastic cyclic deformation are almost the same irrespective of the melting method. This indicates that there is another place where the fatigue crack nucleates as easily as at a TiC inclusion. Optical microscopy revealed that cracks nucleated at grain boundaries in specimens produced by an electron beam melting method, a typical example being shown in Figure 18. Large stress concentrations are expected to occur at both TiC particles and at grain boundaries.

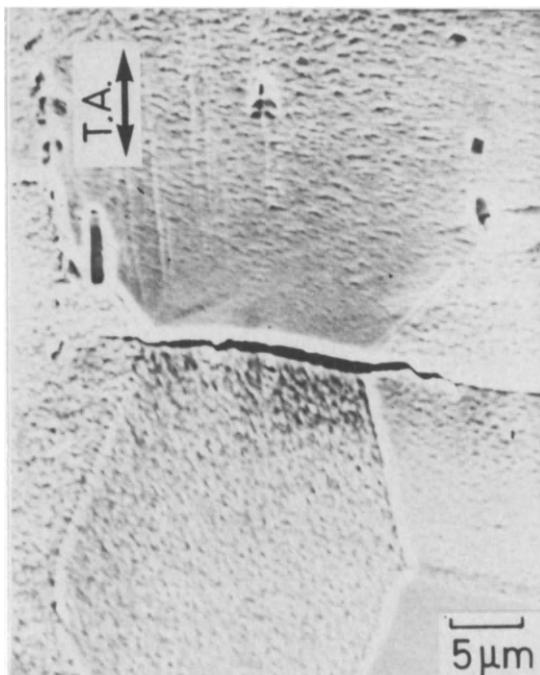


Figure 18: Fatigue crack nucleation along a grain boundary in a Ni-Ti alloy which was made by an electron beam melting method.

On the basis of the above discussion, it is possible to improve the fatigue life of Ni-Ti alloys by introducing the combination of purification of alloys and the introduction of a favorable texture; such treatment removes nucleation sites such as TiC inclusions and grain boundaries where a large stress concentration occurs.

References

1. Miyazaki, S. and Otsuka, K., *Met. Trans. A.* 17, 53, 1986
2. Miyazaki, S., Igo, Y. and Otsuka, K., *Acta Met.*, 34, 2045, 1986

3. Wang, F.E., DeSavage, B.F. and Buehler, W.J.: *J. Appl. Phys.*, **39**, 2166 (1968).
4. Wayman, C.M., Cornelis, I. and Shimizu, K.: *Scripta Met.*, **6**, 115 (1972).
5. Sandrock, G.D., Perkins, A.J. and Hehemann, R.F.: *Met. Trans.*, **12**, 2769. (1971).
6. Perkins, J., *Met. Trans.*, **4**, 2709. (1973).
7. Airoldi, G., Bellini, G. and Francesco, C.D., *J.Phys.F.Metall.Phys.* **14**, 1983 (1984).
8. Tadaki, T., Nakata, Y. and Shimizu, K., *Trans. JIM*, **28**, 883 (1987).
9. Melton, K.N., Mercier, O., *Acta Met.*, **27**, 137 (1979).
10. Miyazaki, S., Imai, T., Otsuka, K. and Suzuki, Y., *Scripta Met.*, **15**, 853 (1981).
11. Miyazaki, S., Ohmi, Y., Otsuka, K. and Suzuki, Y., *J. dePhysique*, **43**, Suppl. 12, C4-255 (1982).
12. Saburi, T., Tatsumi, T. and Nenno, S., ibid., C4-261 (1982).
13. McNichols, J.L.Jr., Brokes, P.C., Cory, J.S., *J. Appl. Phys.*, **52**, 7442 (1981).
14. Melton, K.M. and Mercier, O., *Mater. Sci. Eng.*, **40**, 81. (1979).
15. Miyazaki, S., Sugaya, Y. and Otsuka, K., *Proc. of Inter. Meeting on Advanced Materials*, **251**, 257 (1988).

Linear Superelasticity in Cold-Worked Ni-Ti

G.R. Zadno and T.W. Duerig

Raychem Corporation

300 Constitution Drive

Menlo Park, CA 94025

In their as-cold worked state, Ni-Ti alloys exhibit nearly hysteresis-free linear pseudoelasticity with elastically recoverable strains (springback) as high as 4%. Although the deformation mechanism of this alloy in this condition has not been thoroughly investigated, it is clear that the stress induced martensitic transformation observed in a conventional superelastic material is not the controlling factor. This fundamental difference provides significant potential advantages to the cold worked material such as a temperature and compositional independence of the mechanical properties.

Figure 1 compares the loading and unloading curves of the as-cold worked Ni-Ti with stainless steel. The former exhibits a low modulus "elastic" zone with 4% recoverable strain compared to less than 1% in stainless steel. Although there is a small hysteresis

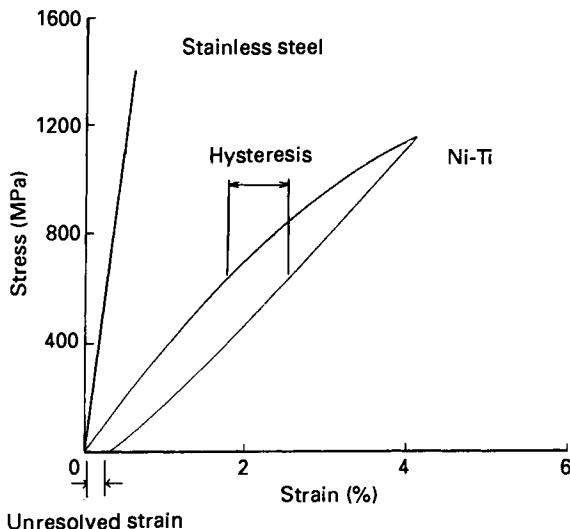


Figure 1: Stress-strain curves at room temperature of the cold worked martensitic Ni-Ti and a typical stainless steel.

between the loading and the unloading curve, it is a reasonable approximation for design purposes to assign a Young's modulus to Ni-Ti and use conventional elastic stress-strain equations. The unresolved strain in Figure 1 increases with the total strain and is independent of the amount of cold work (Figure 2). This presents

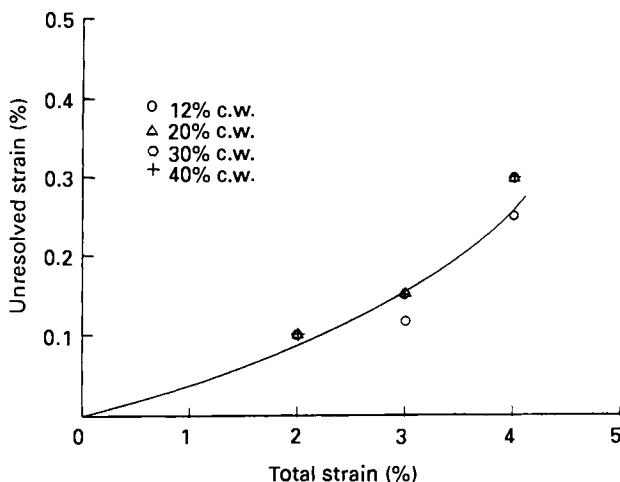


Figure 2: The unresolved strain in cold worked Ni-Ti depends on the total strain but not on the amount of cold work. Samples pulled to 4% at room temperature.

manufacturing advantages since small variations in die size can result in large variations in the amount of the retained cold work and therefore the properties of conventional superelastic Ni-Ti wire. The hysteresis in the loading and the unloading curve also increases with increasing total strain and it is independent of the amount of cold work. The Young's modulus increases slightly with the amount of cold work and decreases with total strain (Figure 3). This decrease is primarily due to the deviation from linearity in the loading and the unloading curves.

It has been shown by many authors^{1,2,3} that superelastically recoverable strains as high as 10% can be obtained in Ni-Ti alloys either by cold working and annealing at a low temperature, or by solution treating and ageing. In these conditions the stress-strain curve is non-linear; not only does Hook's law fail, but the restoration stress in a spring is nearly independent of displacement. The deformation mechanism in this case is the stress induced austenite-to-martensite transformation and therefore depends directly on the ambient temperature. If the temperature is too high (above M_d) the austenite is deformed plastically and at lower temperatures, (below M_s), the deformation occurs through the motion of the martensite twin boundaries; in both cases low recoverable strains are obtained upon unloading. The temperature range over which non-linear superelasticity is observed is around 80°C which creates limitations to possible applications. As we will discuss below the linear pseudoelasticity in the cold worked Ni-Ti is not controlled by the austenite-martensite transformation and is therefore observed over a much wider temperature range. It is shown in Figures 4 and 5 that Young's modulus, hysteresis and the unresolved strain are nearly constant from -200°C to 150°C.

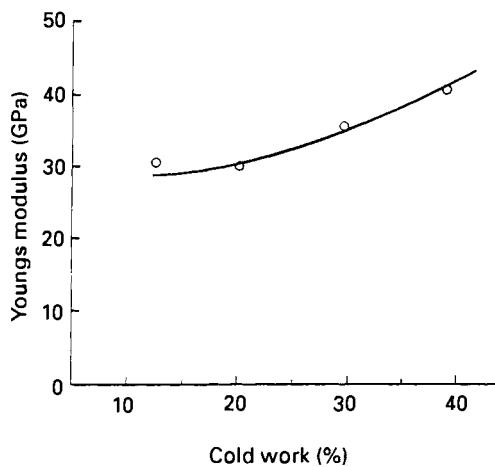


Figure 3: Young's modulus increases with the amount of cold work in Ni-Ti.

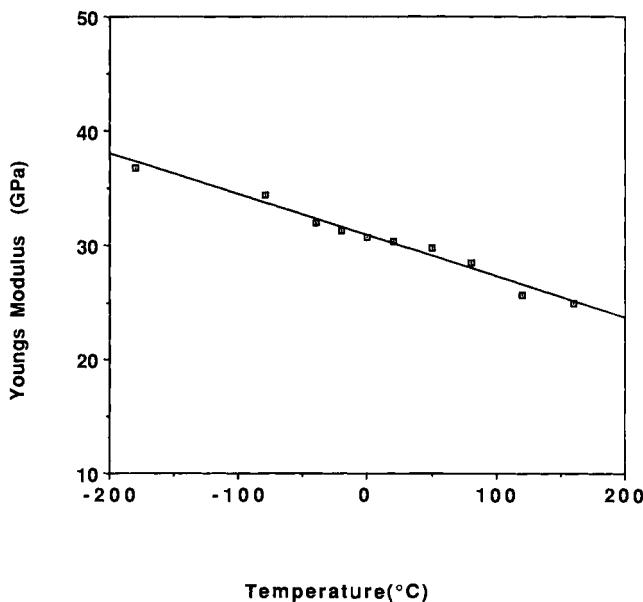


Figure 4: Young's modulus of the as-drawn Ni-Ti wire at various temperatures.
Young's modulus = stress at σ (4%) / ϵ .

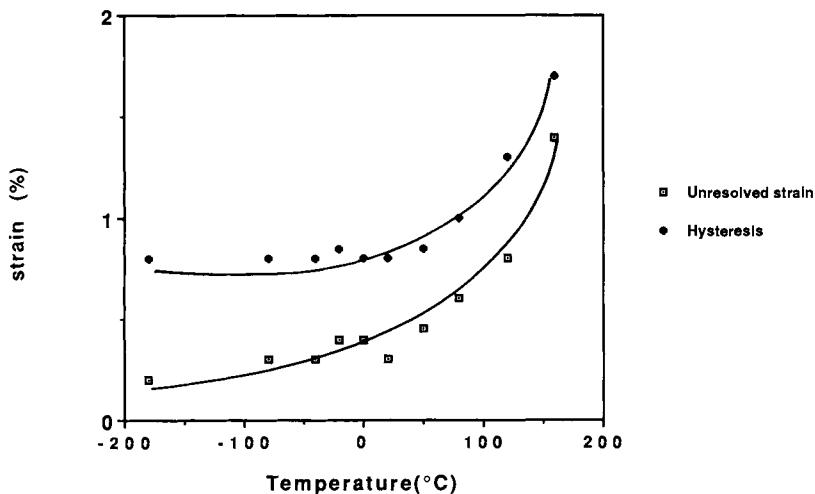


Figure 5: Hysteresis and the unresolved strain in cold worked Ni-Ti wire pulled to 4% and unloaded are almost constant over a wide temperature range.

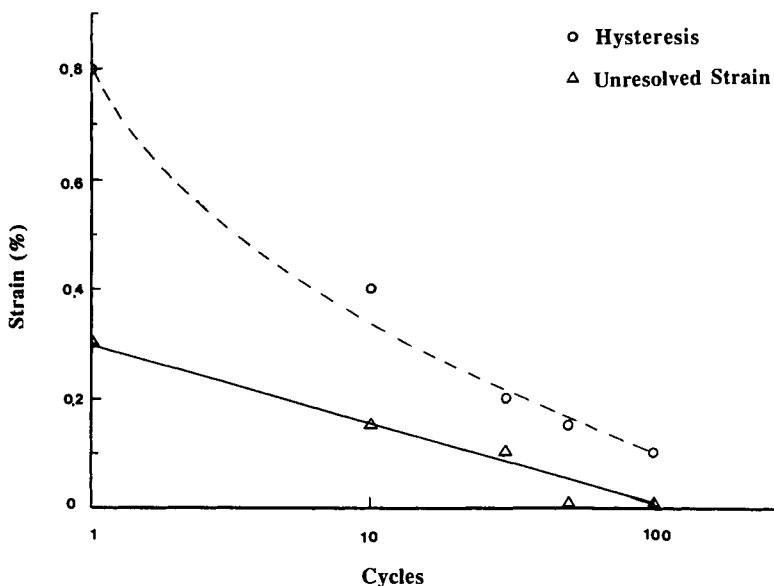


Figure 6: The variations of the hysteresis and the unresolved strain with cyclic deformation in cold worked Ni-Ti.

Figure 6 shows the influence of mechanical cycling at room temperature on hysteresis and the unresolved strain after unloading from 4% total strain in every cycle. It is seen that both parameters decrease in value with the number of cycles and after 80 cycles the unresolved strain becomes zero. In this condition the material behaves like a conventional alloy but with the advantage of having a much larger elastic range. The nonlinear pseudoelastic Ni-Ti behaves quite similarly during cycling. However in this condition the austenite-martensite transformation during loading and unloading is accompanied by the creation of dislocations which in turn result in plasticity and reduction in elasticity. If there are any dislocations created during cycling the cold worked material it is not because of a phase transformation and they do not produce plasticity; in fact the unresolved strain reduces with the number of cycles. It should be noted that before cycling the unresolved strain in the cold worked Ni-Ti is large for total strains about 4%. Although in the present study we have not followed the influence of the total deformation on the unresolved strain during and after cycling one might expect that larger recoverable strains than 4% are obtained after 100 cycles.

Cold worked Ni-Ti also has a high energy storage capacity due to both low modulus and high yield stress. This energy (defined as the area under the unloading stress-strain curve) is about 5 J/cc in steel, and about 20 J/cc in cold worked Ni-Ti. Conventional superelastic Ni-Ti alloys would give values similar to those of cold work Ni-Ti.

It is important to note that the above properties were also observed in Ni-Ti wires with 49.4, 50.0 and 50.8% Ni. Such a composition insensitivity is impossible with superelastic Ni-Ti since the M_s is extremely sensitive to the composition.

As shown above, linear superelasticity is introduced in Ni-Ti through cold working.^{3,4} Similar results are also obtained by neutron irradiation.⁵ The detailed mechanism for linear superelasticity has not been thoroughly investigated. It is, nevertheless believed that the cold worked matrix is composed of twinned and untwinned highly dislocated martensite.^{3,5} In this condition the modulus is low and transformation from martensite to austenite becomes difficult even at temperatures as high as 150°C. The growth of austenite plates is probably pinned by the high defect concentration from cold working. Some detwinning could occur during loading but apparently does not strongly influence the tensile behavior except to introduce a small mechanical hysteresis.

In summary, the tensile properties of the as-drawn Ni-Ti wire are unique in that exceptionally elasticity is observed over a wide temperature range. Despite lower recoverable strains compared to conventional superelastic Ni-Ti, this material could be used in a wide range of industrial applications. Some specific characteristics of this material are:

- (a) Cold worked Ni-Ti exhibits low modulus linear pseudoelasticity with around 4% recoverable strain.
- (b) Minor variations in Ni content and the amount of cold work do not have a first order influence on this behavior.
- (c) Linear pseudoelasticity is observed over a very wide temperature range.
- (d) Cold worked Ni-Ti has a high energy storage capacity; more than 4 times that of steels springs and comparable to superelastic Ni-Ti.
- (e) The primary limitation of cold worked Ni-Ti is shape setting which must be done with over deformation, since normal shape setting with heat would destroy the as-cold worked microstructure.

References

1. Miyazaki, S. et al., J de Physique, C4(43), 255 (1982).
2. Miyazaki, S. et al., Met. Trans. 17A, 115 (1986).
3. Zadno, G.R. et al., MRS conference (to be Published) (1988).
4. Mercier, O. et al., J de Physique, C4(43), 267 (1982).
5. Hoshiya, T. et al., ICOMAT-86, Jap. Inst. Met., 685 (1986).
6. Tadaki, T. et al., Script. Met., 14, 911 (1980).

Eyeglass Frames and SMA - The Challenge and the Product

J. David Chute, The Beta Group
and Darel E. Hodgson, Shape Memory Applications, Inc.

The concept of using shape memory alloys (SMA) in eyeglass frames is not new. The first patent application was in 1975¹ and to date at least 20 additional applications have been filed. The reasons for using SMA, and especially Ni-Ti, in eyeglass frames are numerous. Some of them are:

- The shape memory can allow a frame which is accidentally bent to be restored to like-new shape merely by heating.
- By using the superelastic property of an SMA, frame components which can withstand extreme deformation and spring back completely are possible.
- Ni-Ti alloy is only about 70% of the density of the most common frame materials such as nickel-silver.
- Ni-Ti is extremely corrosion resistant and needs no coating or electroplating to protect it from perspiration or skin oils. Also, there has been no problem with any allergic response to skin contact with Ni-Ti.

In spite of the above advantages, there had been only one minor use of SMA in frames (as a small superelastic lens retaining wire) until the development effort described in this report:

1. Problems Using Ni-Ti

The main reasons which one suspects have hindered the use of Ni-Ti in frames are the following:

- If annealed martensite is used so the frame will have significant shape memory recovery, the martensite will likely exhibit far too low a yield strength to function as an acceptable frame. One can see this in the stress-strain curve shown for annealed martensite in Figure 1(a).
- If the material is used in the superelastic range with material properties as shown in Figure 1(b), then as soon as the ambient temperature drops several degrees the material transforms to martensite and is too weak to serve as a frame as shown in Figure 1(a).
- If the material is used as heavily work hardened martensite (superelastic), it tends to be too stiff, with close to linear elastic properties which are not the most desirable, and also suffers small permanent deformations if bent

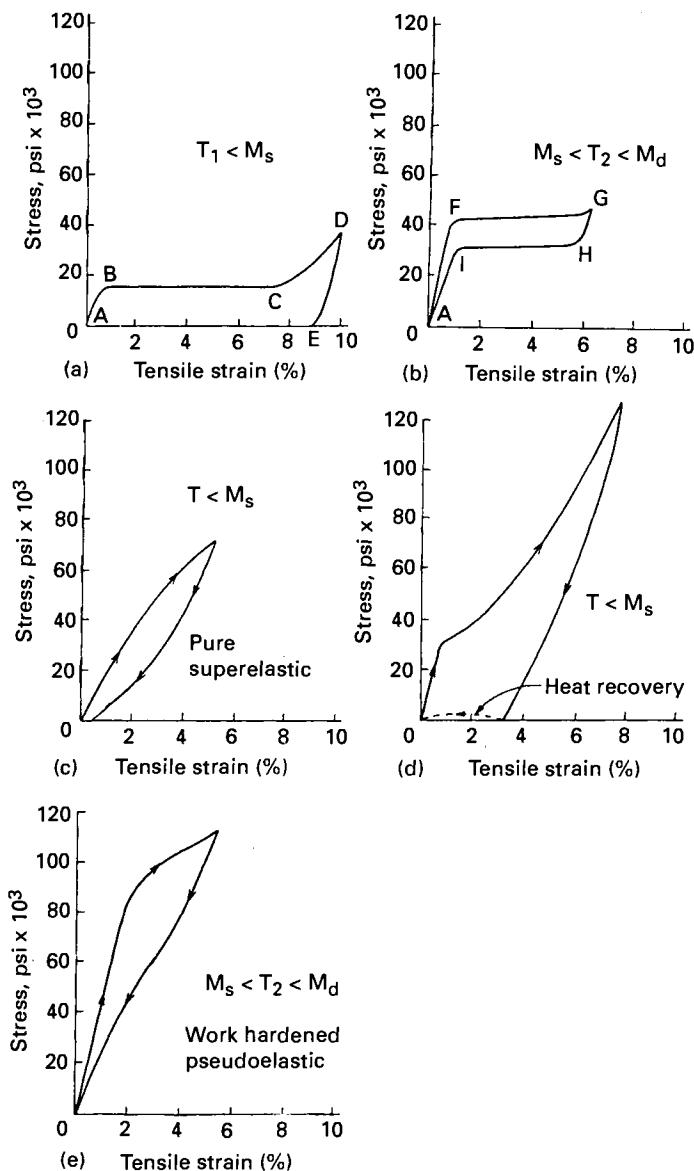


Figure 1 (a) Stress-strain property of annealed martensitic Ni-Ti, (b) stress-strain property of pseudoelastic annealed Ni-Ti, strained above the austenitic recovery temperature and below the maximum stress induced martensite temperature, (c) stress-strain property of heavily cold worked martensitic Ni-Ti, (d) stress-strain property of Ni-Ti properly processed to give adequate stiffness of eyeglass frames and complete shape recovery if deformed and (e) stress-strain property of Ni-Ti processed to give "Optimized" elasticity which is very elastic at room temperature and retains this elasticity to low service temperatures.

significantly and will not give full shape recovery. The stress-strain curve showing this condition is seen in Figure 1(c).

- The Ni-Ti is extremely hard to fabricate, especially with the type of operations and tools normally used in the eyeglass frame industry. The frame industry uses large reductions per pass in its metalworking operations and rarely has to anneal the materials it uses. Ni-Ti, on the other hand, will tolerate only a small reduction per working pass and must be reannealed frequently. Machining of it is very difficult and the type of cutting tools normally used in the frame industry will not suffice. The Ni-Ti is difficult to solder, braze or weld and presents problems if one attempts to electroplate or polish it as one would normal frame materials.
- Finally, one cannot deny that the Ni-Ti is expensive. At first, it seems prohibitively so, and any frame which uses any significant amount of it must sell for a premium over standard frames. Careful analysis of the business opportunity is needed to justify the risk taken in funding the developing effort.

In spite of these problems, Beta Phase undertook the task of determining if the problems could be overcome and the advantages realized. The first part of the frame to be considered was the temple, which is the piece that sets over the ear and along the side of the head and connect to the front of the frame. An attempt was made to develop a part which had a combination of good elastic properties with complete shape memory so that any deformation put into the temples would either spring back immediately or be removed by merely heating in hot water. Also, in the course of the initial experiments and market studies, it was decided that a temple having superelastic-like properties without the tendency to become weak in low ambient temperatures would be desirable.

2. Technical Achievements

By pursuing an extensive study of different processing steps when making the temples, we were able to make parts with elastic properties as shown in Figures 1(d) and 1(e). The curve shown in Figures 1(d) describes the characteristics of temples which display an initial adequate stiffness, significant resilience when bent a large amount, and an ability to recover their original shape when heated if any retained deformation occurs. The elasticity achieved in Figure 1(e), termed "optimized" elasticity², is neither the superelasticity obtained from cold worked martensite nor the superelasticity seen in annealed material deformed above A_f and below M_d . The properties achieved in each of these temples were obtained by careful choice of alloys with specific transformation properties, selected types and amounts of work hardening during fabrication, and heat treatments during and after fabrication chosen.

The elasticity and shape memory properties of the temples resulting from the development were quite acceptable. Temples designed to have shape memory, which were 1mm thick, could be wrapped around a 25mm diameter mandrel and would immediately spring back to a retained bend with a radius of about 50mm when released. The entire amount of this deformed was removed upon heating the temple and the as-manufactured shape was restored. Temples developed for "optimized" elasticity properties could be wrapped around a 25mm diameter mandrel and would totally spring back to the initial shape upon release. This elastic behavior was retained to a temperature of 0°C or below when the temple was cooled.

In addition to the elastic and memory properties in the temples, a number of other features were also developed to create market ready products. Acceptable fabrication schedules and tooling were defined which allowed manufacture of the temples at costs which meant a viable product. Jewelry quality finishes were achieved on the Ni-Ti temples at acceptable cost. Electroplating processes were developed which yielded gold plated parts that could be bent back and forth through severe deformations with no cracking or flaking of the coating nor embrittlement of the Ni-Ti. Brazing and welding techniques were found which allowed attachment of hinge pieces of other metals to the Ni-Ti temples. Finally, a process and piece of equipment were invented which allowed custom setting of the temples' memory shape to each unique customer's facial shape in the optometrists office. The combination of these features plus the elastic and memory properties in the temples resulted in one of the most attractive and sophisticated shape memory alloy products of which these authors are aware. A marketing photograph used to promote the frames is shown in Figure 2.

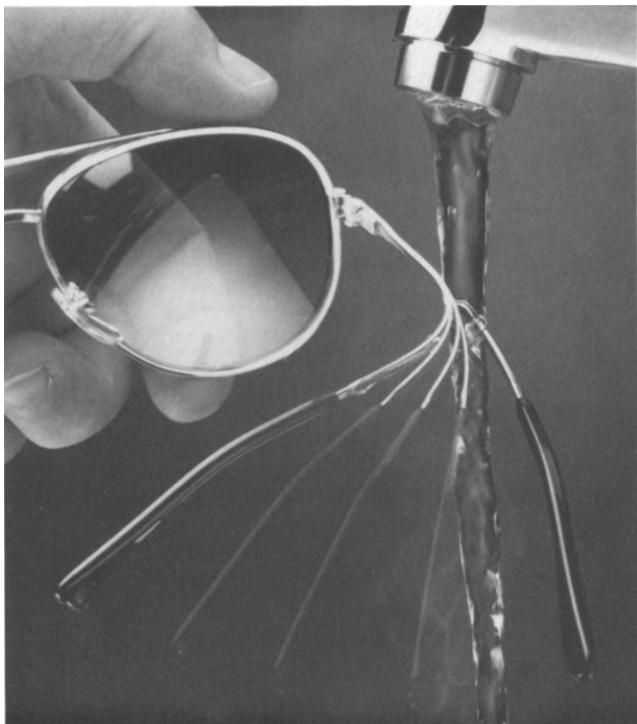


Figure 2: Marketing photograph of an eyeglass frame showing the shape recovery of a bent frame by heating in hot water.

3. Business History and Lessons

Product development viewed from the technologist's standpoint is a sequence of solved problems that result in a final product which performs more or less as it is supposed to. Product development viewed from the marketer's standpoint is a black

void of modern day wizardry that results (hopefully) in something that may or may not be salable in a time frame short enough to affect one's career. Overall, of course, product development is some of each of these views. In general, it can be characterized as producing a product never done before by too few workers in an underfunded environment with unrealistic time constraints. It is as much art as science. The development of Ni-Ti components for eyeglass frames follows this norm, and may serve as an interesting example of what to do and what to avoid when launching a product. The business history of the development can be summarized as follows:

- (a) The product and memory setting properties were first conceived in a small, very under-capitalized company. This forced the introduction of a well capitalized partner to fund the development.
- (b) The funding partner was excited about the product and gave strong capital support, but had its own set of flaws. They did not know the eyeglass frame industry well, there was poor agreement on the ideal product, they were impatient; and, finally, they allowed their short term profit orientation, combined with a quarterly loss, to cause an end to funding at precisely the wrong time.
- (c) The first technical partner in the frame industry was interested in the product, had excellent fabrication skills and was making strong progress, but the correct business deal was not together to allow them to participate in the product. Their support was eventually lost.
- (d) The second technical partner was capable but a small operator, a bit of a rogue and too separated from the developed by distance and culture to allow any control. The result was that product delivery was late, if at all, and most of the technology was transferred to the partner with very little *quid pro quo*.
- (e) The second funding partner was very strong in the frame industry, but had poor cash flow due to being a leveraged buy-out. The company depended upon delivery of the Ni-Ti temples to accomplish their business plan and to remain solvent. Therefore, any delay or extra expense put both partners in jeopardy.
- (f) Finally, the need for sales revenue and pressure created by media attention to this new "magic" product led to premature market introduction. Temple supply was unreliable. The product was not as sophisticated as the marketplace required, and the shape setting protocol for the optometrists was not well enough developed. Thus, initial excellent sales response quickly turned to disappointment and a "black eye" for the product.

The business lessons in most of the above are obvious. Choose your partners extremely carefully - both financial and technical ones. Be certain that your objectives are clearly in accord, and then work very hard to keep communications open and be sure all parties remain in accord. In addition, the following guides are worth heeding:

- (a) Design for user benefit - not product features. Because the product incorporated elasticity and memory, it was much flashier to demonstrate than the purely elastic temple. This caused the development team to produce it first. Unfortunately, it did not incorporate as much actual user benefit as the elastic version of the material and did not sell as well. As it turns out, the elastic version of the product ultimately demonstrated market success.

- (b) If your product is sold into a market which is new to you, you must either get immersed in that market or get a good partner who is already fully acquainted with it.
- (c) Stay focused on what must be done. Don't attempt too much. Make sure the technical progress and required product development proceeds in a timely fashion. Add extra bells and whistles after the original timetable has been met.
- (d) Be certain everyone involved in the project agrees on the product definition, technical objectives and milestones, who is responsible for what, and the marketing thrust and timetable. Then, keep them agreeing. Things can change, but all must agree they're changing.
- (e) If possible, find a committed product champion. Then, give that individual authority to do the job. Support him with the needed resources.
- (f) As a last thought - over optimism is a killer! If a team member says something might be possible, it will be heard as something that can be done, then as something that should and will be done, and finally as a missed target or deadline when it doesn't happen. Any qualifiers or reservations are almost immediately forgotten. Optimism is good, but almost all jobs take much longer and cost much more than first thought. Don't let people's expectations run away with you.

4. Conclusions

The technical development of Ni-Ti temples for eyeglass frames should be counted as a success by most measures. It took longer than hoped and cost more, but the product performed exceptionally well. Processes to plate, coat, fabricate and shape the temples were developed and the economics were close to target. The problems in getting the product into the marketplace were ones of not coordinating the development steps with managements expectations, of having partners with different objectives as the development team, and in understanding what the market really needed instead of what could be developed. As an individual asked to do product development, one must be sensitive to these issues and help control them, if possible, to do one's job completely. It is not enough to simply provide technical and product development contributions without considering and acting on the organizational and business issues affected by them.

References

1. Nihon Kinzoku Co., Ltd., Japanese Patent Application No. U50-133425 (1975).
2. Zider et al., CVI/Beta Ventures, U.S. Patent Application S.N. 876,077 (1986).

The Use of Ni-Ti as an Implant Material in Orthopedics

Dr. J. Haasters,
Ostseeklinik Damp, Ostseeklinik Damp GmbH,
2335 Damp 2, W. Germany

Prof. G. v.Salis-Solio
Medizinische Universität Lübeck
Klinik für Orthopädie, Ratzeburger Allee 160
2400 Lübeck, W. Germany

Dr. G. Bensmann
Krupp Medizintechnik GmbH
Harkortstr. 65, 4300 Essen 1, W. Germany

The specific properties of Ni-Ti have already been reported in detail during the course of this book. The present authors have been concerned with the use of Ni-Ti as an implant material in orthopedics for more than 12 years. The results of this work are outlined in this chapter.

1. Possible Medical Applications of the Memory Effect¹

The properties of memory alloys predestine them for a range of applications both in engineering and in medicine, especially in surgical orthopedics. The potential applications outlined below are in pertinent areas of medicine, but the list is by no means complete and is meant only as food for thought.

1.1 Osteosynthesis plates²

Osteosynthesis is the surgical treatment of bone fractures. During surgery to re-set fractures, especially in the extremities, osteosynthesis plates are attached to the bone on both sides of the fracture with bone screws.

Healing proceeds more rapidly if the fracture faces are under uniform compressive stress. F. Pauwels has developed a mechanically and biologically well-founded theory for this, which states that the deviator portion of the stress tensor favors the formation of connective tissue and the spherical tensor portion that of cartilage. Special plates (dynamic compression plates) have been developed to produce uniform pressure conditions. The aim is to produce the necessary compressive stress in the fracture gap by exploiting the memory effect.

Plates such as shown in Figure 1 were produced for model tests. Before the holes were drilled, these plates were stretched in a tensile testing machine by 8%. In addition to stretching, the plates were ground to finished size and drilled. During machining, care had to be taken to ensure that the plates did not heat up above the transformation temperature. Otherwise, contraction would have occurred at this stage. A plate produced in this way was attached by screws through its outer holes to two araldite plates, simulating a plane model of a bone in such a way that an initial gap of 5mm was produced between them. This gap was only intended to demonstrate the contraction attainable with a memory plate. Spacings as great as this are not possible with conventional metal implants, since every effort is made in fixing the plate to keep

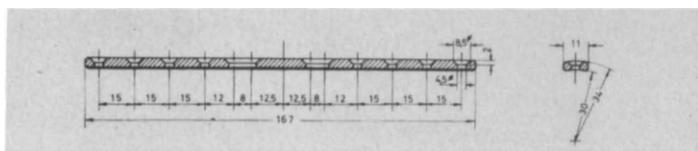


Figure 1 Osteosynthesis plate of Ni-Ti.

the fracture gap as small as possible. The screws adjacent to the gap were only loosely screwed into the longitudinal holes for guidance. The plate was then heated, the temperature being controlled by a thermocouple. Figure 2 shows the complete course of the experiment.

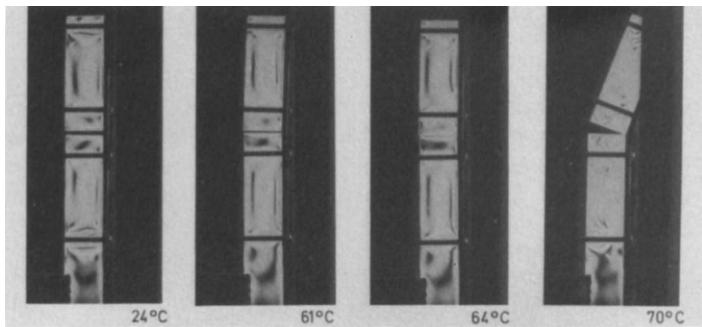


Figure 2 Osteosynthesis experiment with Ni-Ti plate.

At 24°C the araldite plates are 5mm apart. After A_s is exceeded, the gap rapidly becomes smaller and at 61°C is only 2.5mm. At 64°C the gap is completely closed. At this point the operator will cut off the heat supply and the bone-plate assembly remains in the set configuration. If the heat supply is not interrupted, the eccentric loading causes gaping of the fracture on the plate side and to a stress concentration on the side close to the plate. This non-homogeneous state of stress in the gap, which can normally be corrected by initial deformation of the plate, can with a memory plate also be turned into a state of uniform stress by giving the plate an appropriate initial configuration. In the experiment, the total pull which the plate can exert was deliberately demonstrated.

1.2 Jaw plates³

Jaw plates, which are used on fractures of the lower jaw, have functions similar to osteosynthesis plates. In addition to the function of fixation, any gaping of the fracture gap when the lower jaw undergoes stress by mastication must also be prevented. This is particularly important since, for anatomical reasons, the plate can only be fixed to the lower area of the jaw bone.

To demonstrate the operating principle of memory osteosynthesis plates a casting of jaw was sawn in two and prepared for the attachment of a memory jaw plate. Figure 3 shows the jaw casting with the memory plate screwed in place but not yet activated.

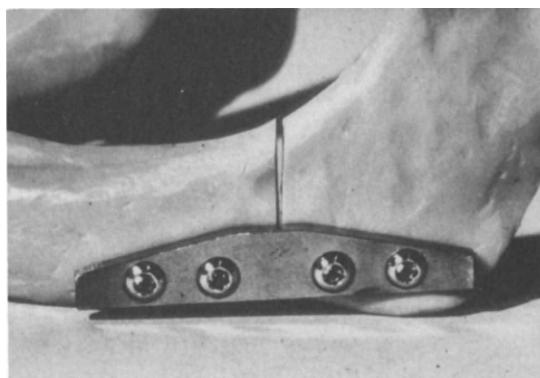


Figure 3 Jaw Casting with Ni-Ti plate and fracture gap before heating.

The "fracture gap" is clearly visible. The memory plate was subsequently heated with hot water. As can be seen from Figure 4 the fracture gap then closed. The deflection of the memory plate can clearly be seen. The fracture gap was then permanently opened by manually simulating mastication. Subsequent heating of the memory plate

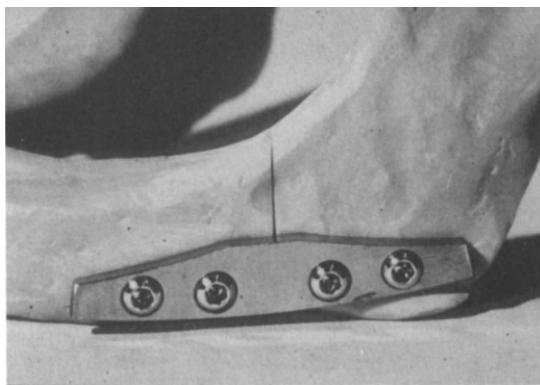


Figure 4 Jaw Casting with an Ni-Ti plate and fracture gap after heating.

again brought about complete closure of the fracture gap. This should prove to be very advantageous in any necessary post-treatment.

1.3 Staples

To immobilize fractures and osteosyntheses, especially in the area of the lower extremities, Blount staples are frequently used. Technically speaking, these are nothing but simple clamps. The problem of attaching them arises from the fact that it is difficult to achieve compression of the osteosynthesis. A gap of greater or lesser size frequently remains. A memory staple can well be used to remedy this problem. Figure 5 shows the function of the memory staple.

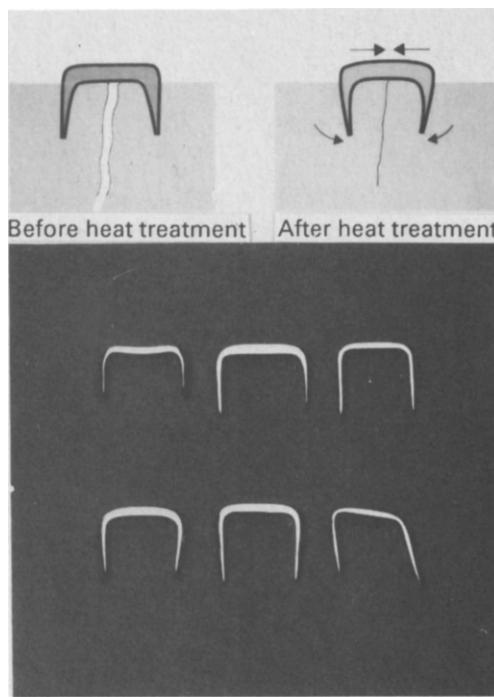


Figure 5 Function of the memory staple.

1.4 Medullary nails⁵

In certain circumstances fractures are not immobilized with osteosynthesis plates but with medullary nails. This applies particularly to clean, closed fractures of the major long bones. For this purpose the medullary cavity of the bone is drilled out, and a medullary peg, as shown at the lower left of Figure 6, is driven into the blind hole.

As can be seen at the top Figure 6, the memory medullary nail, which basically consists of a suitably pre-formed slit or closed tube, is introduced undersize into the pre-drilled medullary cavity. After heating, for example, by rinsing with appropriately heated sterile solution, the slit tube expands to come into contact with the wall of the medullary cavity. Once the fracture is healed, the medullary peg is cooled down with cold water to below M_s . It then contracts again and can be removed without any application of force.

1.5 Spacers⁶

Operations to reinforce the lumbar spine still represent an unsolved problem in orthopedic surgery. The onset of ankylosis is dependent on two factors: that bone chips must be inserted between the vertebrae to be reinforced, and that adequate absence of movement in this area must be ensured during the healing phase, since the relative motion of the two vertebrae to be reinforced impedes osseous intergrowth.

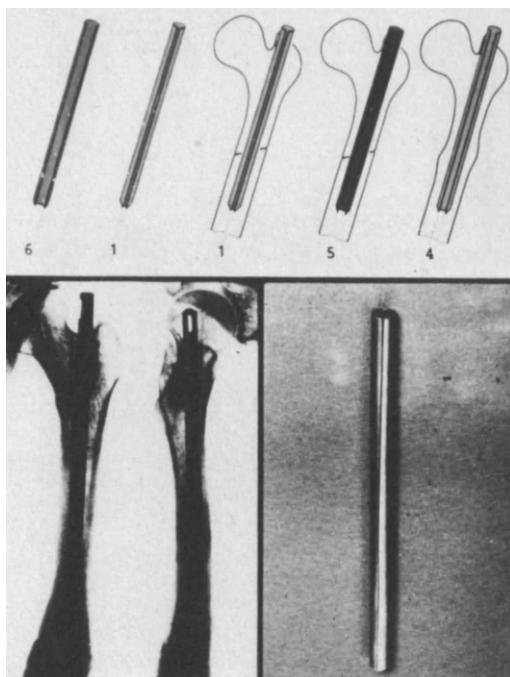


Figure 6 Concept of a medullary peg of memory alloy (before deformation - after deformation, insertion and heating, in the functional state, and cooling and withdrawal).

This last problem - relative motion of the interactive motion of the two vertebrae - is the very one which can be held responsible for many failures. The use here of Ni-Ti components which brace the two vertebrae against one another, seems to promise success. Figure 7 shows a ready-to-insert spacer in the pre-deformed (right) and the activated configuration (the spacer is placed into the disk space).



Figure 7 Spacer for bone-chip arthrodesis of a spinal column before and after heating.

2. Animal Experiments

Several types of animal experiments have been conducted and will be reviewed in turn.

2.1 Tissue compatibility tests on rats

For initial tissue reaction tests white Han-Vistar rats were obtained and Ni-Ti components implanted in them. The 40 rats were provided with specimens on both sides. Tissue compatibility tests are frequently carried out with Davos-type specimen cylinders. Figure 8 shows these specimen cylinders. In order to permit comparison with the studies described in the literature, this type of specimen cylinder was also used for the compatibility tests with Ni-Ti. In addition, rings as shown in Figure 9, were manufactured and implanted. The metal rings were selected because, according to other workers, the tissue inside the metal ring is largely static and at rest. The implants were left in situ for 6, 12, 18 and 24 weeks.

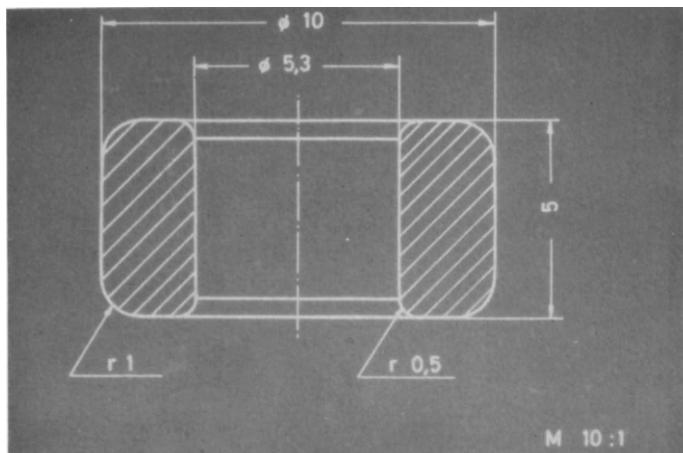


Figure 8 Specimen for testing the tissue compatibility of implant materials.

On removal of the metal specimens no corrosion was found. Histologically, typical inflammatory reactions were observed around the foreign body; fibrocytes, lymphocytes and a distinct increase in the formation of capillaries in the tissue were apparent, but no giant cells induced by the presence of the foreign body were found. For comparison, specimens with the same shape and of conventional materials as used for making prostheses, were implanted in the same animal. Here, the same connective tissue reaction was visible around the foreign body. With these specimens, too, the typical inflammatory reactions with fibrocytes and macrophages were found initially. Furthermore, measurement of the capsules showed a distinct decrease in thickness as a function of implantation period.

2.2 Implantation of jaw plates³

After the tissue compatibility tests, Ni-Ti lower jaw plates for the osteosynthesis of a simulated lower jaw fracture were implanted in minipigs. Operations were performed

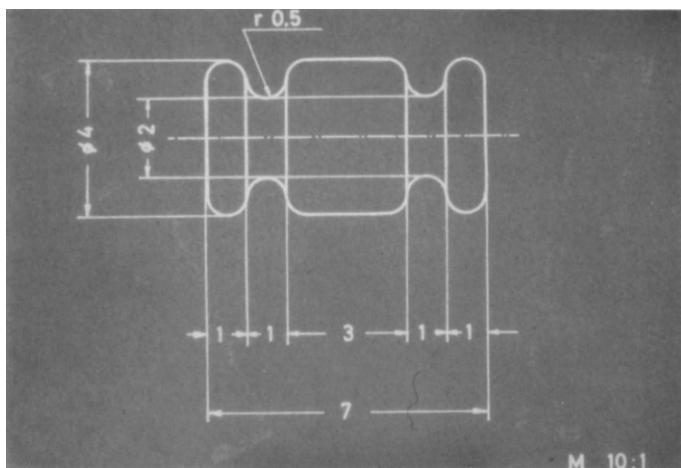


Figure 9 Specimen ring for testing the tissue compatibility of implant materials.

on a total of 9 pigs. Because one pig having died under anesthesia, it was possible to evaluate only 8 pigs histologically.

The operations took place after insufflation anesthesia with the pigs in the dorsal position. After preparation and exposure, the lower jaw was first cut with an oscillating saw, followed by careful haemostasis, repositioning of the bone and matching of a previous gas-sterilized memory jaw plate. Subsequently, after drilling of the holes, the plate was screwed to the lower jaw with osteosynthesis screws (Figure 10a). When the plate was firmly screwed in position, the lower jaw was heated with an appropriate probe connected to a temperature sensor. After the plate had been heated to approximately 55°C, it recovered in the desire manner and complete closure of the osteotomy gap was achieved (Figure 10b).

2.3 Implantation of staples

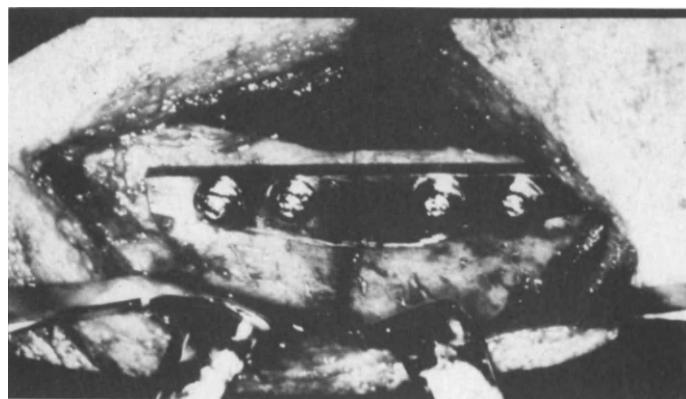
At the end of the lower jaw operation and under the same anesthetic the iliac crest was exposed in each case and a simple iliac osteotomy performed with a chisel. The osteotomy was subsequently re-bridged with a memory staple. The memory staple was heated to approximately 55°C using two electrodes and the osteotomy gap closed completely. Figure 11 shows an Ni-Ti staple inserted in the iliac crest before heating. The results, determined after the animals were sacrificed, showed complete osseous intergrowth of the osteotomy gap. No complications arose in activating the memory staples.

2.4 Implantations of spacers⁶

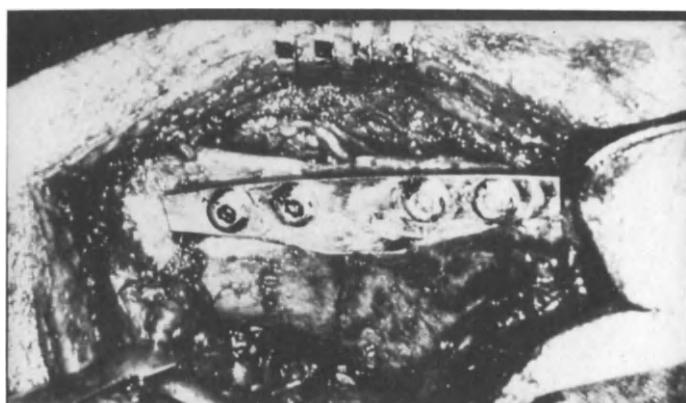
In spinal surgery two possible applications appear particularly interesting:

- (a) The correction of lateral curvature (scoliosis).
- (b) The arthrodesis of two vertebrae under traction.

The latter possibility of traction spondylodesis using memory implants has been studied in animal tests.



(a)



(b)

Figure 10 Memory jaw plate animal test (a) before heating and (b) after heating).



Figure 11 Ni-Ti-staple in the iliac wing (Ni-Ti staple after heating).

An implant was therefore developed which fits accurately between two vertebrae and has cavities and openings permitting the intercalation of bone chips. The implant is made high enough to exceed the distance between the vertebrae by several millimeters, but can be accurately fitted in deformed state. Once placed, it is heated, leading as expected to thermal expansion, i.e. to an increase in height, thus firmly wedging it between the vertebrae and possibly even causing slight spreading of the vertebrae (traction). Figure 12 shows this in basic outline.

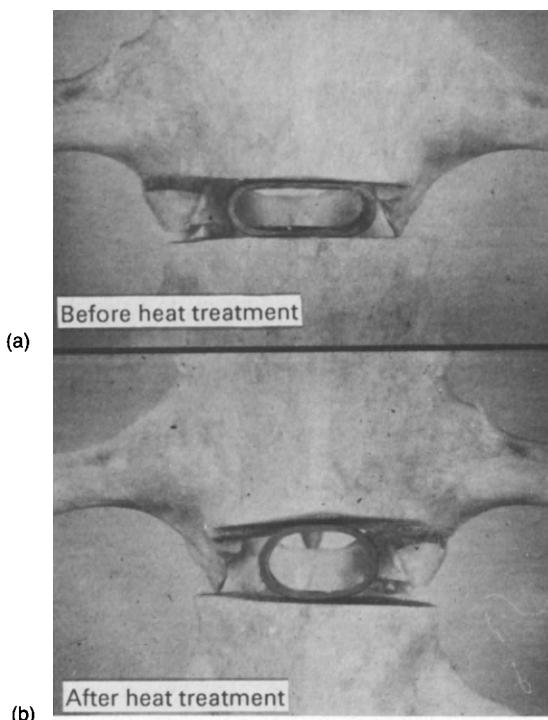


Figure 12 Ni-Ti spacer for bone chip arthrodesis of spinal column segment (before and after heating).

The implant increases in height by heating and exerts traction on the two vertebrae to be reinforced. In every case, wedging or traction ensures the necessary absence of motion in the reinforced area so that the bone chips to be finally inserted in the gaps of the implant can lead to osseous intergrowth.

The operative technique was performed on 25 sheep and was, in most cases, not attended by complications. Figure 13 shows the implant inserted between the vertebrae after activation of the memory effect before and after the space created by the removal of the intervertebral disc had been filled with bone chips. The interior of the implant was also filled with bone chips, as can clearly be seen.

Expansion of the implant and its retention between the vertebrae gave an absence of motion such as has hitherto been unattainable with conventional operative techniques. X-ray checks carried out since then have shown the implants to be properly and firmly positioned and in some cases the desired result - ankylosis - had occurred a short time after the operation. Figure 14 shows the spinal column removed

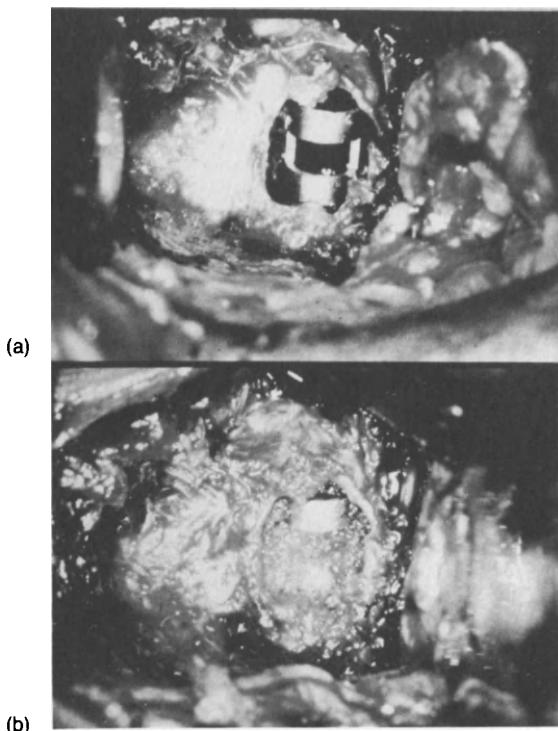


Figure 13 Ni-Ti spacer in a sheep's spinal column before and after filling with bone chips.

from a sheep. The complete ankylosis of the two vertebrae can clearly be seen. In no case was the implant rejected by the animal, so that satisfactory compatibility can be assumed.

3. Clinical applications

Memory osteosynthesis staples are used for the same applications where Blount's bone staples⁷ made of conventional implant materials are employed. Compared to Blount's staples, the memory staples have the advantage of allowing the osteotomy gap to be closed tighter after the device has been implanted. The aim of osteosynthesis should always be to achieve the firmest possible apposition between the two ends of the bone in order to accomplish rapid and reliable healing. This can best be done in operative ankylosis on the foot or in corrective osteotomy at the knee-joint in genu valgum or genu varum (Figure 15).

3.1 Handling of the Staples

3.1.1 Foot

Subtalar arthrodesis for fixation of the talocalcaneonavicular joint is an important operation to stabilize the foot and ensure a plantigrade gait. It leaves the talocrural

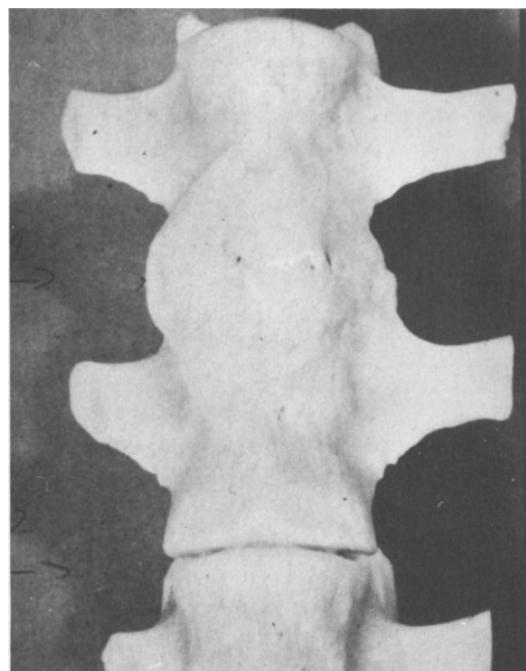


Figure 14 Sheep's spinal column fused with Ni-Ti spacer.

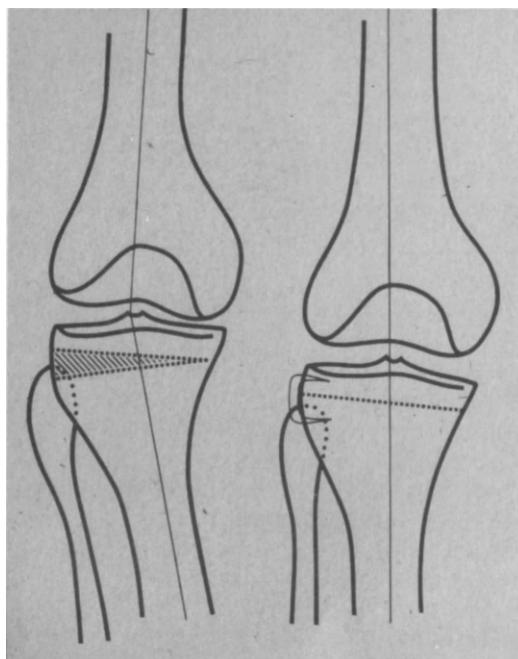


Figure 15 Correction of lower leg by wedge osteotomy and subsequent fixation with memory staple.

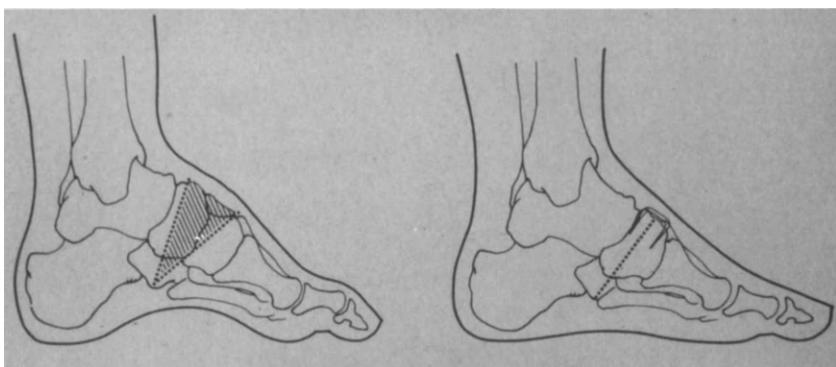


Figure 16 Correction of pes cavus with memory staple: (left) front wedge resection, (right) condition after the procedure.

joint fully functional and is performed in reconstructive orthopedic surgery for deformities or paralysis of the foot and comminuted fractures (Figure 16)⁸.

Memory staples have been shown to be successful for fixation after osteotomy. They provide good compression of the surfaces at the resection site and thus make a reliable bony union possible.

After resection of the articular surfaces and correction of the foot position, the staple is clamped in a driving guide and driven firmly into the bone with a mallet. Subsequently, two electrodes are brought into contact with the corners of the staples and resistance heating occurs (Figure 17). Current is applied under vision until the osteotomy surfaces are firmly approximated. Usually two or three staples are needed to stabilize the apposition. A padded plaster split is applied for the duration of wound

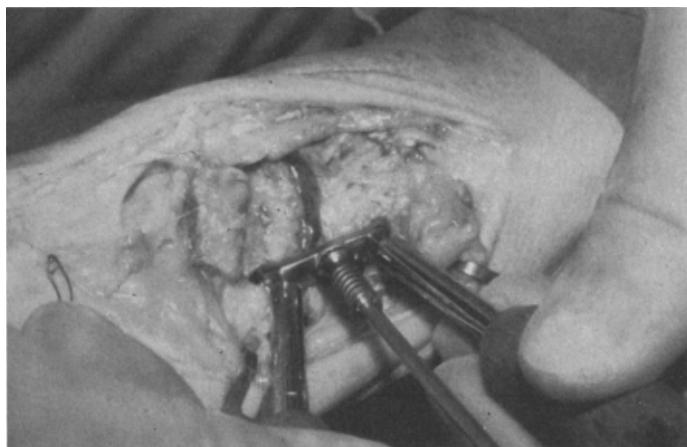


Figure 17 Resistance heating of memory staple after resection osteotomy on foot. Two electrodes and temperature sensor are connected to bone staple.

healing. Figure 18 presents an X-ray image of a resection osteosynthesis fixed with nickel-titanium staples.

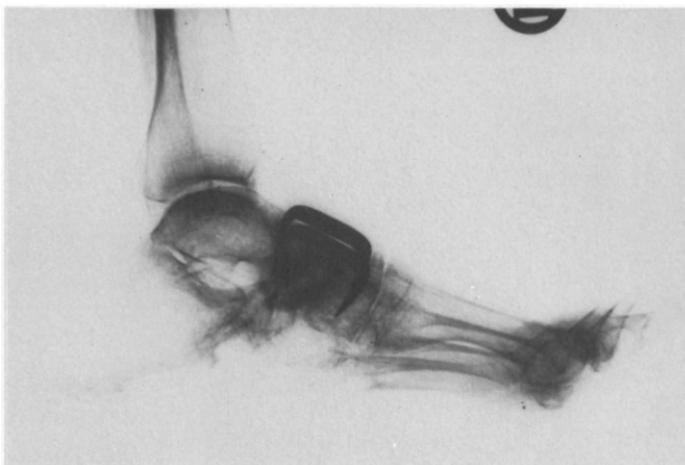


Figure 18 X-ray photograph of resection osteosynthesis eight weeks after implantation.

3.1.2 Knee-joint

In osteotomy at the knee joint (osteotomy of the proximal tibia to treat genu valgum or genu varum), fixation with two or three memory staples provides sufficient support to permit exercise³. Handling of the staples is as simple as in the foot. Here again the staple is implanted in the customary manner and subsequently made to recover by applying hot water or resistance heating. Figure 19 shows the condition of a high osteotomy of the tibia fixed with two memory staples.

3.2 Postoperative examination

3.2.1 Histological examination

Early results of the implantation of memory staples in humans are already available. All osteosyntheses of ankle and knee joints healed by first intention. Bone formation through the union proceeded without complication. Tissue covering the implant which was obtained on removal of the staples was subjected to histological examination. Inspection by the naked eye showed the memory implant to be covered, like other implants made of conventional materials, by a fine layer of connective tissue. Inflammatory rejections, increased blood supply or other rejection mechanisms were not observed. Figure 20 shows the condition of the tissue after removing a memory staple from the foot. Exposure of the examined tissue to memory metal ranged between 6 and 16 months.

Histological examination revealed partly parallelly oriented collagenous connective tissue containing numerous small vessels in foci and fragments of elastic fibers, this tissue must thus be regarded as well vascularized scar tissue. The vascularized



Figure 19 Osteotomy of proximal tibia fixed with memory staples, eight weeks after operation.



Figure 20 Site of operation following removal of memory staple 16 month after implantation .

portions are on the outer side of the mesothelium-covered capsule and exhibited no evidence of an inflammatory reaction or an immune system reaction (Figure 21).

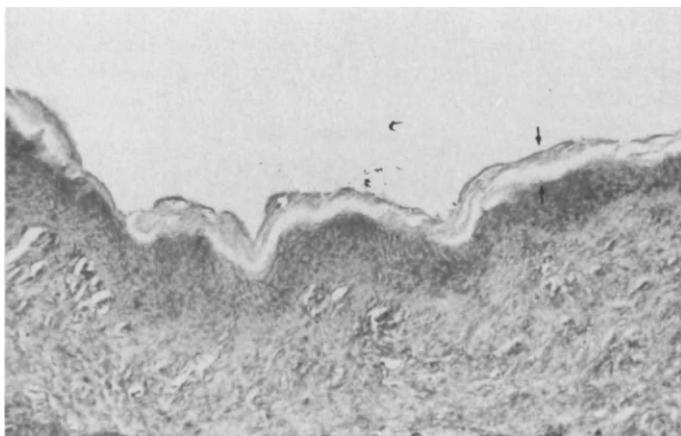


Figure 21 Histological section through tissue surrounding reoperated memory staple.

For comparison with connective tissue membranes covering memory implants, identical specimens were taken from membranes covering previously used metal implants after ensuring that the implants had been in place for comparable periods of time. Histological examination revealed essentially the same appearance as for the memory implants (Figure 22). Compared with the preceding histological investigations on animals, this permits the conclusion that the memory implants did not produce any pathological response in the patients observed and examined.

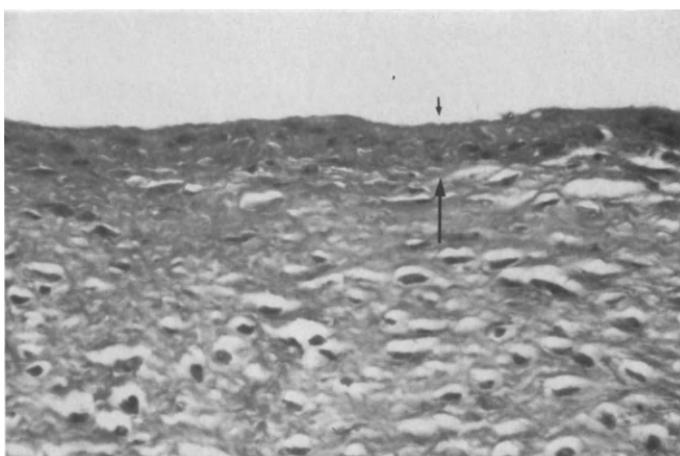


Figure 22 Histological section through tissue surrounding conventional Blount's bone staple.

3.2.2 Examination of reoperated staples

Several memory staples which had been implanted for up to 1.6 years were examined following removal by surgery. Macroscopic study showed all the staples had a uniform appearance with bright metallic surfaces. Slight scratches are attributable to mechanical damage during implantation and extraction. No corrosion was observed on visual inspection.

Supplementary to macroscopic examination, the staples were scrutinized for corrosion by scanning electron microscope. The image reproduced in Figure 23 shows the

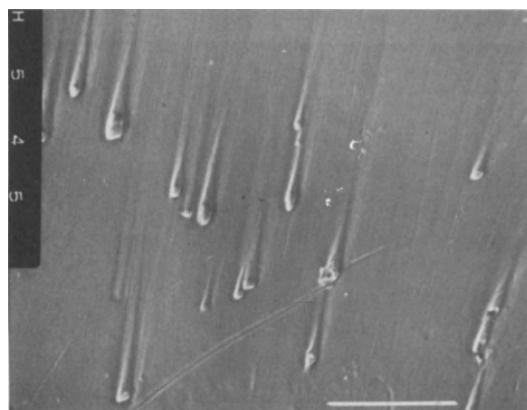


Figure 23 Scanning electron micrograph of surface of nickel-titanium implant after being implanted for 1.6 years.

typical appearance of the staple surface in 1000X magnification. It exhibits the features of a ground and polished metal surface with fine grinding marks. Numerous titanium carbide particles protrude from the metal matrix, not having been abraded because of their high hardness. Slight elevations of the metal matrix are visible behind the titanium carbide particles in the grinding direction, having been in the "leeward" side of the titanium carbide particles during the grinding operation. No areas of corrosive attack were detected on the surface of the implants.

Moreover, results of microprobe spectrometry established that no change had occurred in the chemical composition on the surfaces of the implants compared with the core material. Accordingly the material used for the staples may be regarded as adequately resistant to corrosion.

3.3 Clinical implantation of a spacer

Following the thoroughly convincing results obtained in tests on animals, the orthopedic clinic of the Lubeck School of Medicine has been using this technique on human beings since the autumn of 1982.

The operative technique initially corresponds to the conventional ventral arthrodesis technique:

- Transperitoneal or retroperitoneal access.
- Frontal removal of the invertebral disk, paying special attention to the ventral neuro-vascular structures, avoiding opening the vertebral canal.

- Revivification (with spongiosa) of the neighboring vertebra end plates.
- Ventral insertion of the memory implant; after warming with hot water the implant expands until it is firmly clamped between the neighboring vertebrae.
- The inside of the implant and the remaining disk region are filled with cancellous bone graft.
- After the operation the usual radiological check is carried out utilizing two projections.
- After four weeks' rigid confinement to bed, a further radiological check is carried out and, if the finding is favorable, the patient can get on his feet but has to wear a corset.

By the end of 1987, 95 patients had undergone this operation and no case of dislocation of the implant has been recorded to date. Radiological examinations have clearly revealed signs of incipient bone fusion in patients whose implants have been in place for lengthy periods. Figures 24 to 25 are X-ray photographs illustrating the position of the implant and the increasing bone fusion.

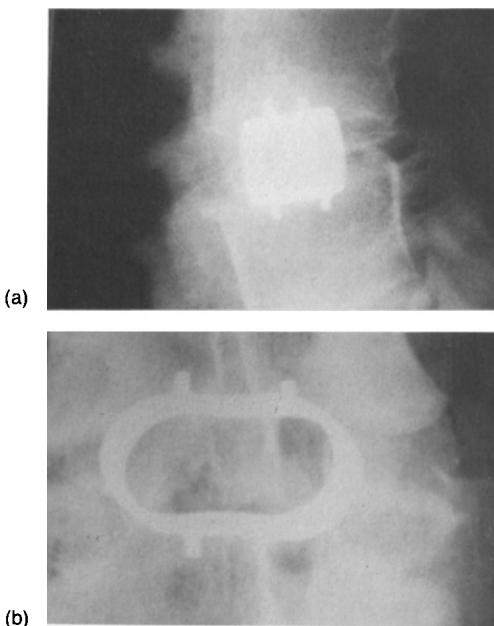


Figure 24 Radiological finding, postoperative (a.p. and lateral).

The arthrodesis of two lumbar vertebrae represents a major orthopedic operation. Provided that posterior spinal fusion is not necessary, preference should be given to ventral arthrodesis of the vertebrae because of the availability of the large surface areas of the neighboring vertebrae, which is favorable in terms of fusion. Unfortunately, ventral spondylodesis has a very high failure rate of over 30%. The key problem here seems to be inadequate mechanical stabilization. Similar considerations in relation to posterior spinal fusion have already led to the use of

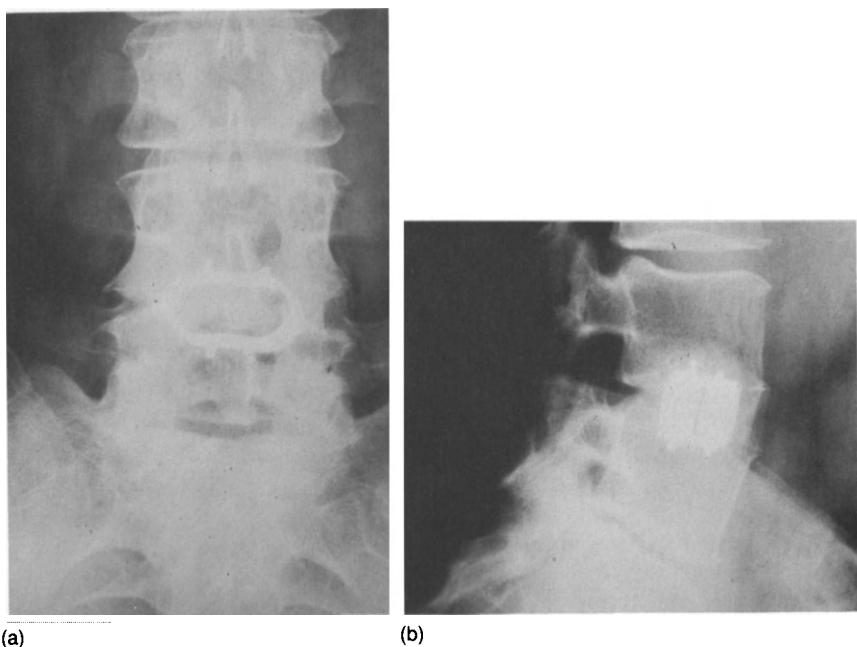


Figure 25 Radiological finding, postoperative after 9 months (a.p. and lateral).

posterior metal implants for additional stabilization. Tests carried out on animals have shown that a memory implant inserted between the vertebrae permits both sufficient mechanical stabilization and bone fusion in the affected segment. Before the operation, allergy tests are always carried out, especially to determine if there is a sensitivity to nickel and titanium.

The progress of patients who have undergone the operation has so far been highly satisfactory. No case of a dislocation of the implant has been recorded and in all patients bone fusion in the affected segment has been observed after just a few months. The operative technique has so far proved free of complications. The procedure employed hitherto depended on warming the memory implant with hot water. However this method of warming the implant will soon be replaced by a more reliable method permitting more accurate implant recovery. Should the stabilizing effect of the memory implant continue to prove its worth, the next step will be to consider gradually reducing the post-operative period of rigid bed confinement (currently 4 weeks) in order to reduce the risks inherent in long periods of recumbancy (thrombosis, embolism).

4. Conclusions

Based on the clinical and histological results obtained with a small number of patients, it is possible to say that implants made from memory alloys show every sign of completely fulfilling the demands made with respect to biofunction and biocompatibility. An essential advantage offered by these new implants is the ability to induce a change of shape by simple application of heat, thus ensuring reliable compression of the osteotomy gap and thereby promoting successful healing as

intended. The early results of implantation of this novel bone surgery material in humans as described herein are such as to encourage continuation of efforts in the direction taken and encourage further work on standardizing and simplifying the method, especially with regards to sterilization and heating of the implants, with a view to expanding a broad and unproblematical field of application.

References:

1. Bensmann, G. et al.: Tech. Mitt. Krupp. Forsch.-Ber. **37**, 21 (1979).
2. Baumgart,F., Bensmann,G. and Hartwig,H.: Tech. Mitt. Krupp. Forsch.-Ber. **35**, 157 (1977).
3. Schettler,D., Baumgart,F., Bensmann.G., and Haasters,J.: Dr. Z. f. Mund-, Kiefer- und Gesichtschir. **2**, 448 (1978).
4. Bensmann,G., Baumgart,F., and Haasters,J.: Tech. Mitt. Krupp. Forsch.-Ber. **40** 123 (1982).
5. Haasters, J., Becker, J. and Bensmann, G.: Orthop. Praxis **20**, 409 (1984).
6. Bensmann,G. and v.Salis-Soglio,G.: Tech. Mitt. Krupp. Forsch.-Ber. **42**, 25 (1984).
7. Blount, W.P.: J. of Bone and Joint Surg. **36**, 688 (1954).
8. Hackenbroch, M. and Witt, A.N. Rutt: A.: Orthop.-chir. Oper.. Stuttgart. **5**, (1973).

Medical Applications of Ni-Ti Alloys in China

Shibi Lu, M.D.
Chairman and Professor of Orthopedic Department,
General Hospital of PLA, Beijing, China

Ni-Ti combines the unique characteristics of shape memory and superelasticity with excellent fatigue life, wear characteristics and corrosion resistance. Not only does it have many industrial applications; its good biocompatibility makes it an ideal biological engineering material, especially in orthopedic surgery and orthodontics. Its function cannot be compared to any of the conventional materials.

1. Fundamental Medical Studies

Ni-Ti alloys have been applied to the medical field in China for ten years. Before clinical applications, a series of fundamental medical studies were performed, such as corrosion resistance, cell growth inhibition, trace nickel analysis of hair and the tissue response to sample coupons implanted in animals.

1.1 Corrosion Tests

Corrosion testing was carried out by quantitative *in vitro* testing¹. Seven kinds of media were used, simulating the conditions in the mouth and human body. Specimens were put into the media and sealed at 37°C for 72 hours. The weight loss was measured on equiaxial Ni-Ti coupons with the dimensions of 19.38 x 0.63 x 14.87 mm using comparative standards of 316L stainless steel (Table 1).

Table 1: Corrosion Rates of Ni-Ti

Media	Mass corrosion rate (mm/yr)	Corrosion Resistance Grade
Synthetic Saliva	2.9×10^{-5}	A
Synthetic Sweat	2.8×10^{-5}	A
Hank's Solution	0	A
1% sodium chloride sol.	5.5×10^{-5}	A
1% Lactic Acid	5.7×10^{-5}	A
0.05% hydrochloric Acid	0	A
0.1% sodium sulphate acid	6.9×10^{-5}	A

An experiment to determine the corrosion of implanted Ni-Ti coupons in rabbits was also carried out² using coupons of 10x10x2 mm in size. The results are shown in Table 2.

Table 2: Corrosion of Ni-Ti Implanted in Rabbits

No.	Implanted time (Days)	Weight increase (g)	Corrosion rate (mm/yr.)
1	3	-9.1x10 ⁻⁵	5.6x10 ⁻³
2	7	-3.8x10 ⁻⁵	1.1x10 ⁻³
3	15	-2.4x10 ⁻⁵	3.1x10 ⁻⁴
4	30	-7.2x10 ⁻⁵	4.5x10 ⁻⁴
5	90	-3.4x10 ⁻⁵	6.7x10 ⁻⁵
6	180	-4.7x10 ⁻⁵	4.8x10 ⁻⁵
7	360	-4.2x10 ⁻⁵	2.4x10 ⁻⁵

1.2 *In Vitro* Toxicity³

In vitro toxicity was evaluated by cell growth inhibition tests, adhesion tests or cell adhesion, which had advantages of being short term tests and at the same time being highly repeatable and reliable. Mouse fibroblast L-cells were grown in Eagle's medium with the addition of 10% calf's serum. The Ni-Ti specimens were put into the culture chamber. A cell suspension was prepared and pipetted into the chamber containing a humidified carbon dioxide atmosphere at 37°C for 24 hours. Finally, they were fixed and compared with a negative control. The results showed that the cells grew well on the surface of the alloy.

1.3 Histological Observation⁴

In order to observe tissue tolerance to implants, 60 Ni-Ti specimens were implanted in the femur and subcutaneous area of rats for periods of 3 to 10 months. X-ray photographs taken after one week showed newly formed bone tissue in close contact with the Ni-Ti. Radiographic examination of the implants showed normal bone healing with no evidence of reaction and/or resorption of bone and no rejection phenomenon over a 10 month period.

The histological observations indicated that newly formed bone tissue was already growing between the compact bone and the small separated bone. After one week, part of the sponge tissue covered the Ni-Ti surface. After two weeks, the gap between the compact bone and Ni-Ti was nearly filled with newly formed bone. From 4 to 8 weeks, the amount of compact bone in contact with the Ni-Ti increased with time. After 8 weeks normal bone tissue was observed. Subsequent stages showed only an increase in the compactness of the bone. The tissue response consisted of a thin pseudo-membrane surrounding the test pieces; microscopic examination shows no inflammatory reaction, and the biocompatibility of Ni-Ti was good. As a result, it is considered that Ni-Ti is a valuable implant material for dental and medical applications.

1.4 The Analysis of Ni-Content in Dog Implants

Though a variety of alloys containing nickel are widely used in medicine, pure nickel has been demonstrated to be harmful to the human body. In order to investigate whether the nickel in the Ni-Ti implants may migrate to other parts of the body, a trace nickel analysis of dogs' hair was performed by means of X-ray spectrograph, both preimplantation and postimplantation.

Four young dogs were used: two for preliminary tests and two for formal tests. Three Ni-Ti disks were used per dog, each 7.5mm in diameter and 2mm in thickness. The specimens were embedded in the soft tissue of the dogs' right hind legs. The anterior

dorsal region of the thorax was shaved to obtain the hair used for analysis. As a control, the hair was cut 1-3 days prior to implantation. When the hair grew postoperatively, it was shaved again for testing. The hair was cut from the same region after 6, 9, 12, 18 and 24 months. The results, given in Table 3, show that the nickel content in the dogs' hair increased only slightly after implantation. The surfaces and weight of the specimens changed very little.

Table 3: Nickel Contents in Dogs' Hair Pre- and Postimplantation

	Nickel content (ppm)
preimplantation	0.50-0.60
6 months	0.20-0.30
9 months	0.10-0.63
12 months	0.96-0.92
18 months	0.87-1.15
2 years	0.86-0.54

2. Clinical Applications

Since 1981 Ni-Ti devices have been used clinically in Beijing and Shanghai, in fields such as orthodontics, oral-maxillofacial surgery, orthopedic surgery, plastic surgery and gynecology.

2.1 Shape Memory Ni-Ti Clip for Tubal Sterilization

In 1981, a new tubal clip for female sterilization was designed: it could be opened in ice water and closed at 40°C. Four rabbits were initially tested. A month after the application, the clips were found to be covered by a smooth serosal layer without any adhesion. From December 1981 to March 1985, these types of clips were applied to 325 women for sterilization with satisfactory results⁴.

The clips were 2 mm in width, either 15 or 17 mm in length, and had a space of 0.1 mm between the two arms to avoid cutting the tube and prevent fistula formation. The inner parts of the two arms were toothed. Their weight varied from 0.4 to 0.47 gm.

Minilaparotomy was done under local anesthesia and the uterine tube was traced out to the fimbriae routinely. With the isthmic and ampullary portions of the tube held up by Allis forceps, the opened clip was removed from ice water with mosquito forceps and placed directly over the isthmus 30 mm from the uterine cornu. The two arms closed spontaneously after 40°C warm saline was poured over the clip. Then the tube was replaced into the abdominal cavity. The total operative time was recorded from anesthesia to the end of the operation. Any feeling of discomfort during the application of clips was noted and recorded. Patient follow-up was done 6 months after the operation and yearly thereafter.

The average operation time was 14 minutes, with times ranging from 6 to 45 minutes. Patients did not complain of pain either during or after the operation. There was neither traumatic injury nor hematoma. In only one case was a mild infection of the abdominal incision noted.

One hundred and forty-four cases were followed for more than 2 years and 124 cases for more than 3 years. X-rays of the pelvis in 10 cases revealed a normal position of the clips without distortion. Hysterosalpingograms in 3 cases revealed blockages of the tubes at the position of the clips. Follow-up showed that 29 cases had very mild backache. One had irregular periods which diagnostic curettage proved to be functional uterine bleeding with cystic hyperplasia of the endometrium. Six had oligomenorrhea, and none had menorrhagia. Four women became pregnant 5-10 months after sterilization, all of whom were treated in the first 6 months of our project. On second operations performed on three women, the clips were found to be in correct positions, with fistula formation in one while in the other two, the tubes were very thick and large. The tubes were not excised or tied: the procedure brought about very little

trauma. There were 4 failures with preliminary unrefined clips made during the first six months of the project. No ectopic pregnancy was noted. Among these 4 failures, two cases were performed in the postmenstrual period giving a failure rate of 2/127 (1.63%), and two cases were postabortion (a failure rate of 6.4%). Postpartem and postabortion sterilizations yielded failure rates 2 to 5 times higher than that of postmenstrual cases, probably due to edema and congestion of the tissue after conception. In second operations, the clips were found to be 15 and 17 mm in length. Clamps with increased forces at the tips of two arms were designed (over 75 grams). During the last three years, 250 women were operated upon with no failures. In total, the failure rate was 1.23%.

In conclusion, the clips with a narrow width of 2 mm are conveniently clamped on the isthmic portion of the fallopian tube for sterilization. A more reliable reversal is possible by microsurgical reanastomosis.

2.2 Intertrial Fixation in Orthopedic Surgery Via Compressive Staples⁵

Since 1980, the shape memory characteristic of Ni-Ti has been used to manufacture an internal fixation staple. From 1981 onward these staples have been used in clinical cases with promising results. The alloy is Ni-44 wt %Ti. The most commonly used staple is 1.5 mm in diameter and consists of an undulated body and 2 pointed arms attached to each end of the body at an angle of 60 degrees. The undulated body can be straightened between 0 and 5°C. Once the pointed arms are inserted into the bone, body temperature and hot saline pads induce a recovery of the straightened staple to its original undulated shape, thereby exerting a compressive force on the bone ends. The recovery force (bending moment) has been measured to be 3 kg-mm after straightening the staple 5 mm in ice water and heating it to 37°C. If it is then heated to a higher temperature and then cooled again to 37°C, the recovery force increases to 5.5 kg-mm. In order to enhance the recovery force, the temperature of the saline pads should be above 37°C but well within the range that can be tolerated by the patient. The Ni-Ti shape memory alloy possesses a better wear resistance than the commonly used Ti-alloys and medical-grade stainless steel. It also possesses good tissue compatibility and can resist corrosion by body fluids.

From December 1981 to July 1985 Ni-Ti staples were used in cases of bone and joint operations, including fractures of the ankle, triple arthrodesis of the foot, arthrodesis of the wrist joint, arthrodesis of the hip joint, fracture of the patella, fracture of the olecranon, and metacarpal and phalangeal fractures. In addition to internal fixation with the staples, patient immobilization was used. All cases showed either bony union or sound ligament healing. With the exception of arthrodesis cases, functional recovery was also satisfactory. Follow-up examination of those patients retaining the staples for more than 2 years showed no signs of inflammation and no tenderness over the operated area. X-ray examination revealed solid bony union without loosening of the staples and without absorption of the fixed bones.

The compressive staple can be easily applied. It is only necessary to make a small incision on the periosteum and to drill a small hole through the incised periosteum in each bone fragment. The periosteum of the bone ends can be retained. Thus damage to the endosteum, as occurs during intramedullary nailing, and damage to the periosteum, as is produced during extensive stripping for plate fixation, can be obviated. The surgical trauma is minimized and favorable conditions for bone healing are provided. However, as the area thus fixed is rather limited as is the fixation force, external immobilization with a cast is necessary in the early post-operative period.

Histologically, in 8 cases no obvious foreign body reaction was detected in the soft tissues abutting the staples which were removed after bone union.

2.3 Shape Memory Pins in Oral-Maxillofacial Surgery⁶

Since November 1981, Ni-Ti pins have been used in oral-maxillofacial surgery, such as in the fixation of fractures of the mandible, in orthognathic surgery, in reconstructive

surgery of maxillofacial bone defects, in iliac or rib bone transplantation post-mandibulectomy, etc. From clinical analysis and observations of 18 cases, a complete 100% success rate has been achieved. The results show that Ni-Ti is an excellent biomaterial for use in oral and maxillofacial surgery, and has bright prospects for medical applications.

2.4 Shape Memory Double-Cup Hip Prosthesis⁷

Double-cup total hip surface replacement has certain advantages, but its clinical application is limited due to postoperative complications such as loosening or displacement of the prosthesis, avascular necrosis of the femoral head or delayed fracture of the femoral neck. In the double-cup prosthesis now available, the diameter of the rim of the femoral component should be larger than that of the bottom. Otherwise, it cannot be positioned over the femoral head which is the main cause of loosening and displacement. Luck modified the femoral cup to a cylindrical form, so that it could prevent the cup from rotating into varus, but it required a wider trimming of the femoral head and neck. Perhaps this is the cause of delayed fracture of the femoral neck after replacement.

The lip of the femoral component (made of NT-2 shape memory alloy) is able to shrink by means of raising the local temperature after the component is applied to the femoral head. Over trimming of the femoral head can be avoided and the stability is thus significantly increased.

The design of the shape memory cup of the femoral head and its clinical trial in 16 cases (17 hips) is presented in the present work. The clinical indications were osteoarthritis, central dislocation, joint stiffness and pathological dislocation. The follow-up period was 4-48 months. All of these cases were found to have their femoral components in the proper position and no loosening or fracture of femoral neck was found. The femoral cup was cracked in one case after a bad fall two years after the operation. Another case had constant postoperative hip pain even though the cup still remained in place. The cups were removed one year after operation and a total hip replacement was then performed.

2.5 Intervertebral Artificial Joints⁸

The artificial joint is shaped from a Ni-Ti shape memory sheet 0.8 mm thick and 8 mm wide. The maximum distances between 2 arched arms are 13, 14, 15 and 16 mm. A two mm long inverted spur projects anteriorly and laterally on both the upper and lower arched arms. There is a vertical "resisting" plate over each end to prevent posterior slippage. The length of the plate is 6 mm with a central notch which is clamped by pliers. The depths of the concave arches are 16, 17, 18 mm, smaller than the average the sagittal diameter of the cervical vertebra in order to prevent compression of the spinal cord by protrusion into the vertebral canal.

There are five characteristics of the joint: its shape can be freely changed in 4 to 10°C water and recovered at normal body temperature (36-37°C); the loading capacity of the anterior opening of the joint prosthesis when compressed to half distance is 164 MPa - larger than the net weight of the skull; the fatigue lifetime of the alloy at 490 MPa is over 2.5×10^7 cycles, and the alloy exhibits good wear resistance; the alloy is non-toxic to the human body, not harmful to tissue, and corrodes slower than 0.001 mm/year: the magnetic conductivity of the alloy is less than 1.002.

In the operation, the artificial joint is pressed by a hemostat clamp in 5 to 10°C aseptic physiological saline solution to make the distance between upper and lower arched arms about half of the original distance, and to completely close the opening so that the artificial joint can be easily introduced by passing through the narrow outer opening of the intervertebral space into the deep part of the intervertebral space until the resisting plates are situated just lateral to the anterior margin of the vertebra. The implant then recovers rapidly to its original shape by warming to normal body

temperature. The implant, acting as a support, is prevented from slipping by means of the arched upper and lower arms and the inverted spurs. The patient is allowed to sit up with a plaster collar and to be ambulatory the day after the operation. Neck collar fixation is maintained for about six weeks. The movement of the cervical vertebral with the artificial joint is schematical.

Our patient mainly suffer from adult osteogenic cervical spondylosis and cervical fracture-dislocation complicated with compression symptoms of the spinal cord. The corresponding size of the joint is selected for the patient, and the patient should be examined again to determine spinal column mobility and to identify any hidden fractures that may be present. In all, 32 cases have been reported (24 male and 8 female) with ages ranging from 24 to 65 years (46.5 year average). Of these, 16 cases suffered from cervical fracture dislocation complicated with paralysis and 16 suffered from cervical spondylosis. In total there were 37 intervertebral space operations, in which 4 involved 2 vertebrae and one underwent two operations. As to the distribution of the vertebral levels, 6 cases were C4-5, 23 cases C5-6 and 8 cases C6-7. The operations were all quite smooth with surgery times ranging from 60 to 120 minutes. The advantages of using the artificial joint are that bone grafting is not necessary, and that suffering and complications related to surgery can be prevented because the operation time is shortened. There were no complications due to inward or outward slippage, and no absorption of bone graft or pain over the bone graft donor area were observed. Mobility of the diseased vertebra is preserved (or increased) which follows with the human body's anatomical and physiological principles. The incidence of degenerative changes occurring in the neighboring vertebrae due to transmitted local movement after fusion-fixation of the diseased vertebra is reduced. The device also possesses a certain spreading action (the force is 8-16 kg when compressed to half its height, which exceeds the weight of skull). Thus it not only can prevent kyphotic deformities which often appear after resection of diseased bone, but it can also improve the original angular deformity (mostly seen in patients with vertebral trauma). Surgical procedures and manipulating techniques are simplified and fusion-fixation is not necessary after decompression.

Through clinical application, it is suggested that this kind of intervertebral joint prosthesis possesses articulating and supporting action; long term follow-up is necessary to observe the duration of its function and the reaction to the implantation.

2.6 Scoliosis Correction Rods²

The diameters of the Ni-Ti rods vary from 6.8 to 7.8 mm, and the lengths vary from 200 to 400 mm depending upon the individual. The shape recovery temperatures range from 35°C to 39°C. Three kinds of rod are available: straight, "L"-shaped and curved. The diameter of the fixation wire is 1 mm. The sterilized instruments are put into refrigerator preoperatively.

After exposure of the vertebral laminae and the articular processus, the facet joint surface is destroyed bilaterally and parts of the articular processes are removed from the apex of the convex side to facilitate corrective manipulation. Two memory rods are fixed by wires using the sublaminar method. When the trunk is manipulated, hot saline solution moistened gauze (34 to 50°C) is spread over the rods. When the rods warm to A_s, they soon recover their original shape and produce a corrective force. Throughout the manipulation period, the patient is monitored with spinal stimulating potentials. Facet joint fusion is then carried out by placing cancellous bone in the spaces of laminae and articular processus. The incision is then washed and sutured - postoperative external immobilization is not required. The patient is kept in bed for 6 weeks.

Between 1982 and 1985, operations were carried out on 70 patients with idiopathic scoliosis. Of these, the preoperative Cobb's angles ranged from 35° to 105° with a mean of 53.8°. The postoperative Cobb's angles ranged from 9° to 67° with a mean of 25.8°, resulting in corrections ranging from 13° to 62° with a mean of 25.9°. In percentage, the angles were corrected 23.4% to 75% (a mean of 51.8%). The

postoperative incremental increases in height ranged from 20 to 60 mm (mean of 39 mm). The lengths of the rods were from 22 to 40 cm, and the numbers of vertebrae which were fixed ranged from 5 to 13. The upper-ends of the rods protruded into the subcutaneous tissues in 2 cases, both because of insufficient wire fixation.

Shape memory scoliosis rods can offer an efficient corrective effect; the percentage of angle correction averaged 5.18% in patients treated in the above described way; Luque reported 65 cases of scoliosis (25 idiopathic, 40 paralytic), of which the percentages of angle correction were 53-72%. He employed traction, such as Holo traction pre-operatively, and observed only 15°-19° correction during operation. If preoperative correction was combined with the above procedure, the effect would be better. Fracture does not occur in the fulcrum: the fixation of memory alloy instrumentation is distributed over multiple laminae and spinous processes, so the force is dispersed. The shape memory rods are fixed with wire to the laminae, according to the curvature of the spinal deformity. Thus the forces generated during heating and the corrective forces on each segment of the spine are even. The above procedure need not employ a plaster body cast postoperatively. There has also been no corrective angle loss detected in follow-up observation. Kahn reported 42 cases: the corrective angles in Harrington's group were 40% six weeks after operation and 30% in Luque's group; after 19 months of follow-up the corrective angles were 29% in Harrington's group and 33% in Luque's. There has been no appearance of loss of correction in our cases; on the contrary a few cases have shown improvement.

References

1. Zhang, X.F, et al., "A Study of Shape Memory Alloy for Medicine": Shape Memory Alloy 86' Proceeding of the International Symposium of Shape Memory Alloys, China Academic Publishers (1986).
2. Lu, S.B, et al., "Scoliosis Correction Rods of TiNi Alloy and Clinical Application", ibid, 379.
3. Xue, M. and Jia, W.T.: "Application of TiNi Shape Memory Alloy to Medicine and Dentistry". ibid, 411.
4. Xue, P, et al.: "TiNi Shape Memory Alloy Clip for Tubal Sterilization", ibid, 444.
5. Yang, P.J, et al.: "Internal fixation with NiTi Shape Memory Alloy Compressive Staples in Orthopedic Surgery". ibid, 438.
6. Pan, K.F, et al.: "Clinical Application of NiTi Memory Alloy Pin in Oral-Maxillo-Facial Surgery", ibid, 427.
7. Dai, K.R, et al.: "Orthopedic application of NiTi Shape Memory Alloy", ibid, 430.
8. Zhao, D.L, et al.: "Study and Clinical applications of the Cervical Intervertebral Artificial Joint", ibid, 433.

Superelastic Ni-Ti Alloys in Orthodontics

Rohit C.L. Sachdeva^{*} and Shuichi Miyazaki[†]

* Department Of Orthodontics
Baylor College Of Dentistry
Dallas, Texas USA

† Suichi Miyazaki
Institute Of Material Science
University Of Tsukuba
JAPAN

The development of new alloys in dentistry has contributed significantly to the evolution of orthodontic appliance treatment. Traditionally gold, stainless-steel and chrome cobalt nickel¹ have been amongst the most popular alloys used in orthodontic therapy. In the last decade or so the cold worked martensitic Ni-Ti²⁻⁵ and beta titanium alloys⁵ have been introduced to the field of orthodontics. These alloys demonstrate both excellent springback and low stiffness characteristics; properties that are considered suitable for orthodontic appliance design.²⁰ Superelastic Ni-Ti alloys (SE NiTi) were recently added to the list of materials available to the orthodontic practitioner.⁶⁻¹⁴ This alloy exhibits a unique blend of mechanical properties. The most significant of all is the generation of a constant stress when the material undergoes transformational deformation (i.e. superelasticity). This behavior has great ramifications in orthodontic mechanotherapy.

1. Biologic Basis for the Use of SE NiTi

Tooth movement during orthodontic therapy is achieved as a result of the bone remodelling processes propagated by the application of forces to the dentition. Such forces are created by elastically deforming an orthodontic wire and allowing its stored energy to be released to the dentition over a period of time.

At the subclinical level these forces are translated into stress fields in the periodontium. This mechanical stimulus triggers the chemical mediators responsible for bone remodelling through a complex chain of events.²²⁻²⁴ Stresses that are tensile in nature tend to favor bone apposition while compressive stresses encourage bone resorption.²⁵⁻²⁹ It is generally believed that optimal tooth movement is achieved by applying forces that are low in magnitude and continuous in nature. Such forces minimize tissue destruction and also sustain a relatively constant stress in the periodontal ligament during tooth movement. Both these factors are important in achieving efficient tooth movement. In contrast, forces that are high in magnitude encourage hyalinization of the periodontal ligament and may cause irreversible tissue damage such as root resorption (Figures 1a and 1b).³⁰⁻⁵⁷ Also, if these forces dissipate readily, fluctuating stress fields are created within the periodontal ligament which often cause tissue damage and lead to erratic tooth motion. As a consequence of these side effects, tooth movement is retarded.

It has been noted by various clinicians that appliances fabricated from SE NiTi appear

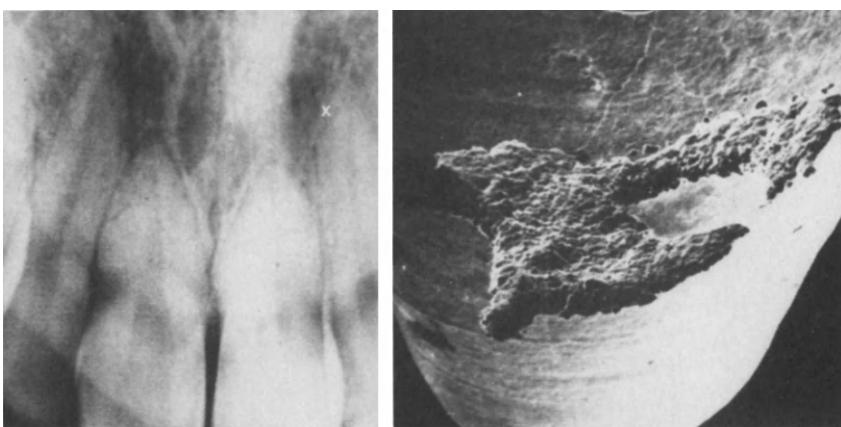


Figure 1: (a) Root resorption caused as a result of heavy forces being placed on the upper central incisors during orthodontic therapy, and (b) photomicrograph of a root surface of a tooth showing root resorption as a result of excessive forces applied on the dentition.

to move teeth with greater efficiency and in a shorter time when compared to other orthodontic alloys.¹²⁻¹⁴, and it is especially true in situations requiring large deflections of an orthodontic arch wire such as the preliminary bracket alignment stage in orthodontic therapy. This phenomenon may readily be explained by comparing the load deflection characteristics of an SE NiTi alloy with other orthodontic materials. (Figure 2a). Stainless steel orthodontic arch wire produces very high forces for a small deflection: for instance a 0.41 mm stainless steel orthodontic arch wire when displaced 3mm in a three point bending test produces close to 750g of force. Clinically, such a force level is considered to be high in magnitude and is liable to cause irreversible tissue damage.^{56,57} In addition, these forces dissipate linearly and rapidly, thus causing fluctuating stress fields in the periodontal ligament. Application of such a force system to the dentition is likely to cause unfavorable biological sequelae such as hyalinization and root resorption and at the clinical level orthodontic delay tooth movement. In contrast to other alloys, SE NiTi generates lower force magnitudes under similar loading conditions. SE NiTi also demonstrates an unloading curve with three distinct phases each reflective of the predominant crystalline structure occurring in the material at a given loading circumstance. The initial and final slopes are defined by the modulus of elasticity of the austenitic and martensitic phases respectively (Figure 2b). In the transition between these two slopes the forces produced by this alloy remains fairly constant over a considerable range of wire activation. This characteristic is not demonstrated by any of the other alloys used in orthodontics. It is this mechanical property of SE NiTi that allows for the application of constant forces to the dentition over long activation spans which results in a desirable biological response.

It has been hypothesized that the transient changes in the load produced by SE NiTi alloys as a result of changes in mouth temperature may accelerate tooth motion (Figures 3a and 3b).⁸ Such temperature changes may be induced during the ingestion of cold or hot food. A change in the loading pattern experienced by the

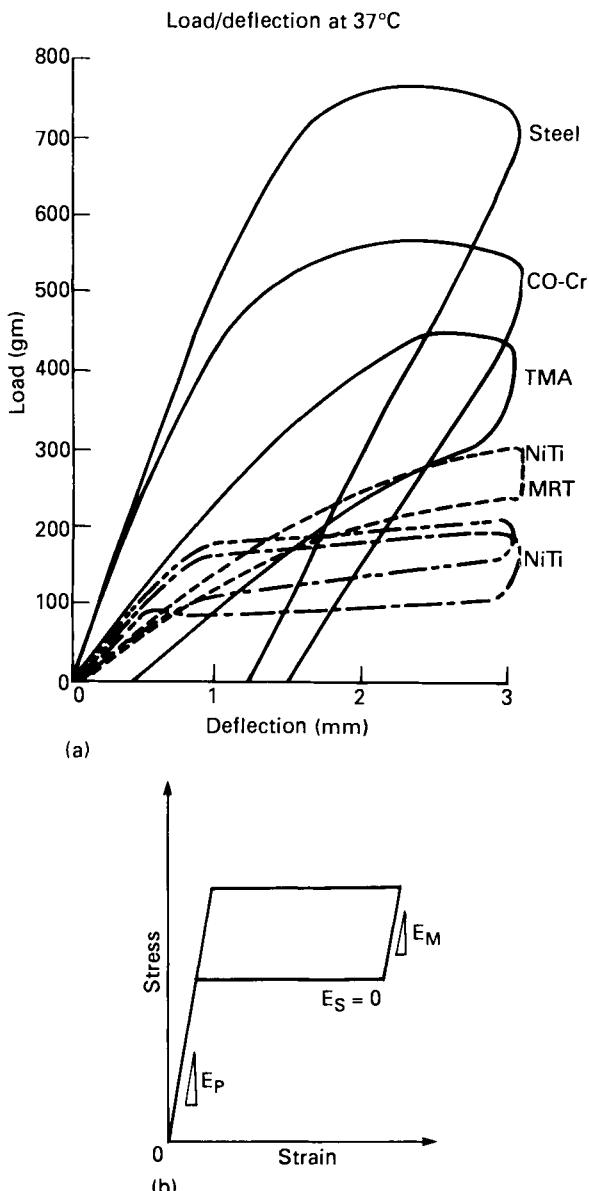


Figure 2: (a) The load-deflection characteristics of various orthodontic alloys available in the market today (wire diameter of 0.16" tested in 3-point bending): Co-Cr (Elgiloy), TMA (Titanium Molybdenum Alloy), NiTi-MRT (cold worked martensitic Ni-Ti), NiTi (superelastic Ni-Ti). Note all alloys except for Ni-Ti have relatively high load deflection characteristics. (b) Schematic figure showing the stress-strain curve of SE NiTi, where E_p is the austenitic modulus, E_m if the martensitic modulus and E_s is the modulus of the reverse transformation.

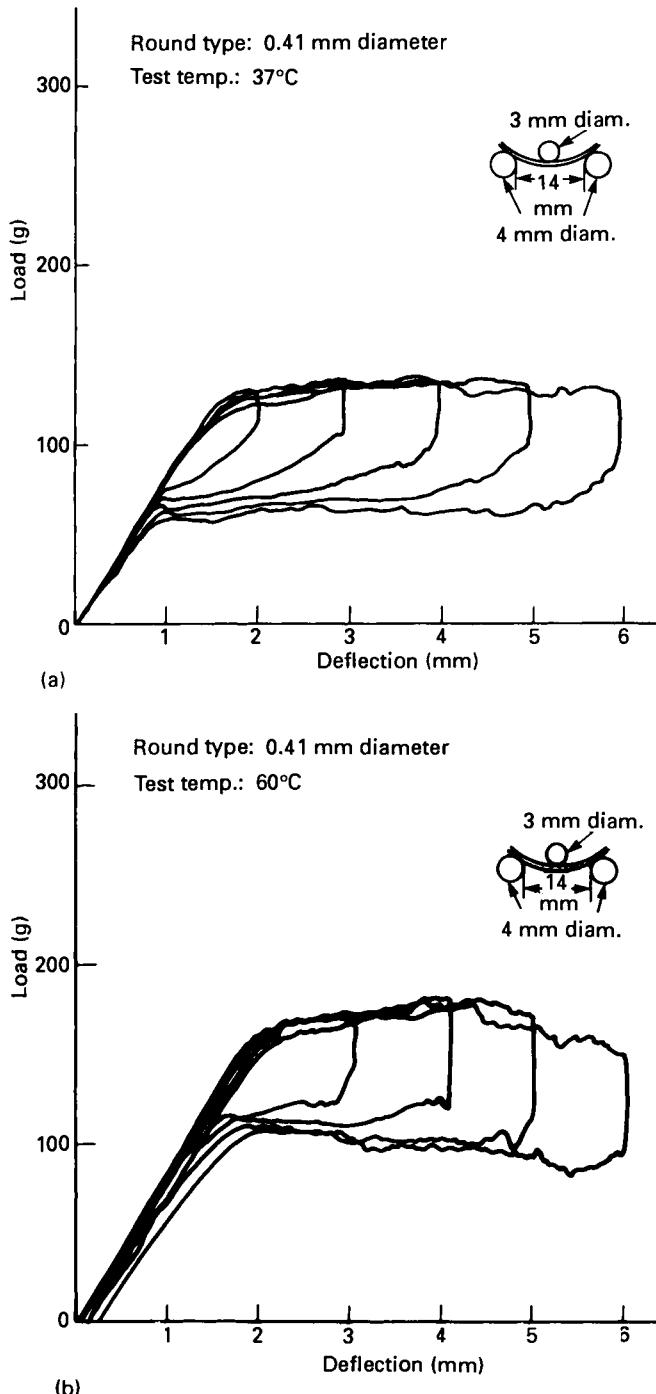


Figure 3: Deformation temperature effects on the stresses generated by SE NiTi. The highest stresses are produced at elevated temperatures.

dentition as a result of variation in mouth temperature is considered to encourage blood flow in areas where blood stasis has occurred due to capillary strangulation. Thus the cellular elements involved in the repair processes during tooth movement are replenished with essential nutrients which maintain their vitality and normal function. These events encourage the optimal biological response.

A disadvantage of the thermally induced stress change is that it cannot be easily measured or controlled by the orthodontist in the oral environment. However, in the mouth this phenomenon is transient in nature and is not likely to adversely affect tooth movement.

2. Clinical Applications of SE NiTi Archwire

SE NiTi alloy has found its greatest use in the fabrication of orthodontic archwires. Two primary mechanical characteristics of this alloy are taken advantage of in the design of this orthodontic appliance. Firstly, superelasticity, a unique property which provides for the application of constant stresses to the dentition. (The biological advantages of which have been discussed earlier) and secondly, the materials ability to show complete recovery even when deformed to strains in the order of 6 to 8 percent.^{58-60, 67} The latter property provides the clinician with the distinct advantage of being able to activate a SE NiTi orthodontic archwire over a long span without risk of permanently deforming the appliance, thus increasing its working range. Also, it is important to note that the range of optimal force delivered to the dentition is increased as a result of the constant stresses being applied over a greater working range of the superelastic appliance (Figures 4a and 4b). This may explain why it is possible for us

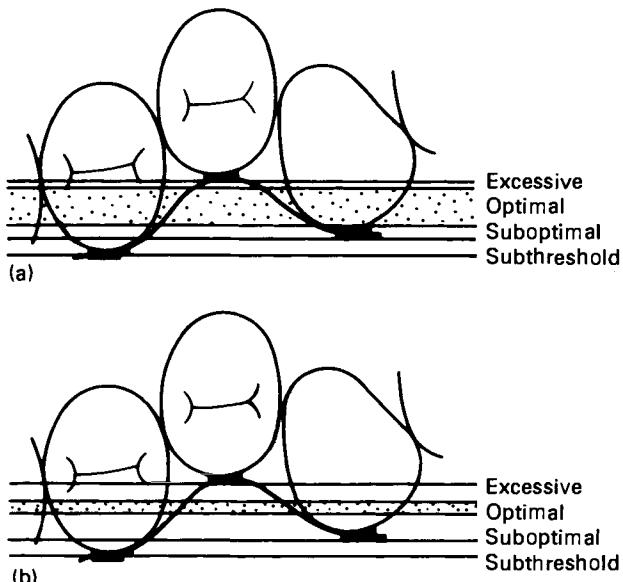


Figure 4: Schematic diagram showing that the range of optimal forces delivered by SE NiTi is greater than that of conventional alloys. Also note that conventional alloys tend to impart higher forces to the dentition over a greater range of tooth movement: (a) Ni-Ti, (b) Stainless steel.

to achieve "levelling" of the dentition during the preliminary bracket alignment stage of orthodontic therapy by using a single SE NiTi orthodontic archwire (Figures 5a, b,c and d). In comparison, when using conventional alloys during this phase of treatment,

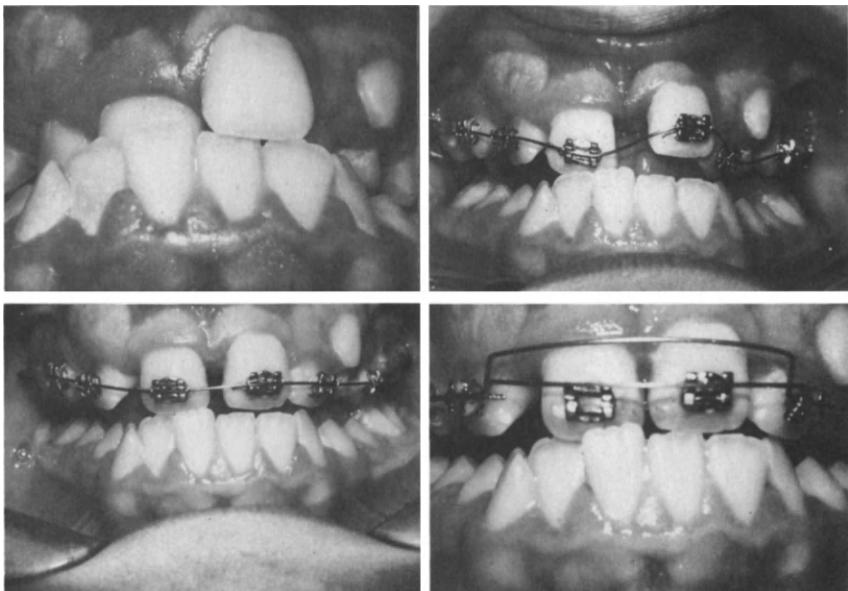


Figure 5: A case demonstrating the use of SE NiTi in the preliminary bracket alignment stage of orthodontic therapy: (a) malocclusion at the start of treatment, (b) 1 week after insertion of wire, (c) 2 weeks into treatment, and (d) 3 weeks into treatment - malocclusion corrected.

a clinician is obliged to use a series of wires to achieve the desired results. The non superelastic orthodontic arch wires are prone to permanent deformation for a relatively small amount of strain.¹⁵⁻¹⁷ This factor imposes a severe limitation on the working range of such appliances. A clinician may, therefore, activate these orthodontic wires only over a short distance. This limits the amount of tooth movement achieved between appointments.

An orthodontist usually attempts to increase the working range of the traditional orthodontic archwires by overbending it (Figure 6). This step has a number of disadvantages; firstly, a tooth may be displaced beyond its desired position in the arch form, secondly, the wire may be inadvertently deformed over the larger activation distance and finally, the configuration of the appliance changes, which influences the bracket wire geometry resulting in the generation of spurious force systems which may lead to unwanted tooth motion. Placement of such bends is not required in SE NiTi archwires because of their increased working range. However, an orthodontist may further increase the range of optimal forces generated by SE NiTi orthodontic archwires by simply overactivating the wire prior to or after engaging it into the brackets (Figure.7).

A characteristic of SE NiTi orthodontic archwire which is of clinical importance is that its load deflection characteristics may be controlled by the operator to suit the needs of

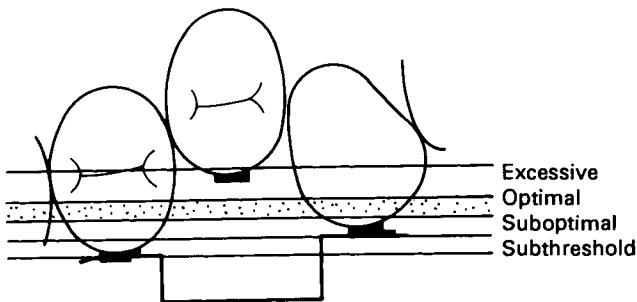


Figure 6: Overbending the wire may increase the zone of optimal force slightly but at the risk of applying the excessive forces to the dentition. This step is not recommended [see text for further details]

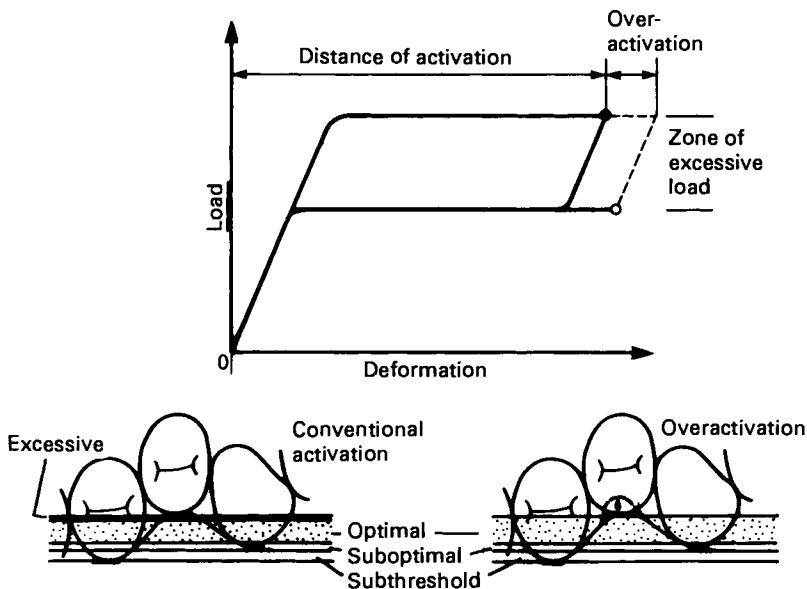


Figure 7: Overactivation of SE NiTi wire eliminates the zone of excessive force and increases the optimal force range.

his treatment. For instance, during the finishing stage of orthodontic treatment, it is appropriate to take advantage of linear elasticity exhibited by the austenitic phase. On the other hand in clinical situations that require large amounts of tooth movement it is best to utilize the wire in its superelastic mode. Either of these loading circumstances may be obtained by influencing the amount the archwire is displaced or stressed. For stresses below the critical stress for inducing martensite, linear elasticity is exhibited by the wire.

Another disadvantage of traditional orthodontic archwires is that slight errors in their activation may considerably influence the applied stresses. This is because of their high load deflection characteristics.²⁰ In contrast, SE NiTi orthodontic wires may be activated over a wide activation range without changing the magnitude of the stresses applied to the dentition. In other words, the appliance is forgiving to errors in activation.

Complex loop designs are usually incorporated in traditional orthodontic archwires to lower their load deflection characteristics. In SE NiTi orthodontic archwires these are not required, because of their favorable load deflection properties.

In our clinical experience the use of SE NiTi archwires in orthodontic treatment reduces chairside time spent on a patient because frequent adjustments of the appliance are not required. Besides this factor, the biological and mechanical advantages of using SE NiTi outweigh the use of any of the conventional materials presently used in orthodontics.

3. Variable Load, Constant Force Uniform Cross-section Orthodontics, A New Clinical Strategy

Traditionally the magnitude of forces applied to the dentition has been controlled by varying the cross-section of the orthodontic arch wire used in a patient. A small change in the cross-section of the wire may significantly affect its stiffness.¹⁶⁻¹⁸ In fact, the load deflection rate of a round wire increases as the fourth power of its diameter, and thus smaller cross-section wires are used to generate lighter forces and larger for heavier forces.

With the advent of newer alloys for appliance fabrication the variable modulus approach in orthodontic treatment has evolved.¹⁵ This treatment philosophy advocates the use of a uniform cross-section of wire, preferably rectangular, to maintain torque control during all phases of orthodontic therapy. The desirable orthodontic force levels are achieved by choosing a range of materials possessing the appropriate moduli of elasticity. A low modulus alloy is used in situations requiring low and continuous forces, whereas a high modulus material is used when high forces are required.

Stresses generated by SE NiTi can be easily influenced in a controlled manner by either compositional change and/or heat treatment.⁶¹ This characteristic provides the clinician with the unique opportunity to control the magnitude of the forces applied to the dentition through the use of a single alloy system. In essence, the practice of this philosophy of treatment requires that a series of similarly sized orthodontic arch wires with different nickel compositions and/or heat treatments be used to generate the desired force levels through all phases of orthodontic therapy (Figures 8a,b,c, and d). This treatment strategy allows the operator to take full advantage of the superelastic properties of Ni-Ti during all phases of orthodontic treatment. An added benefit of using this approach in orthodontic therapy is that a clinician need familiarize himself

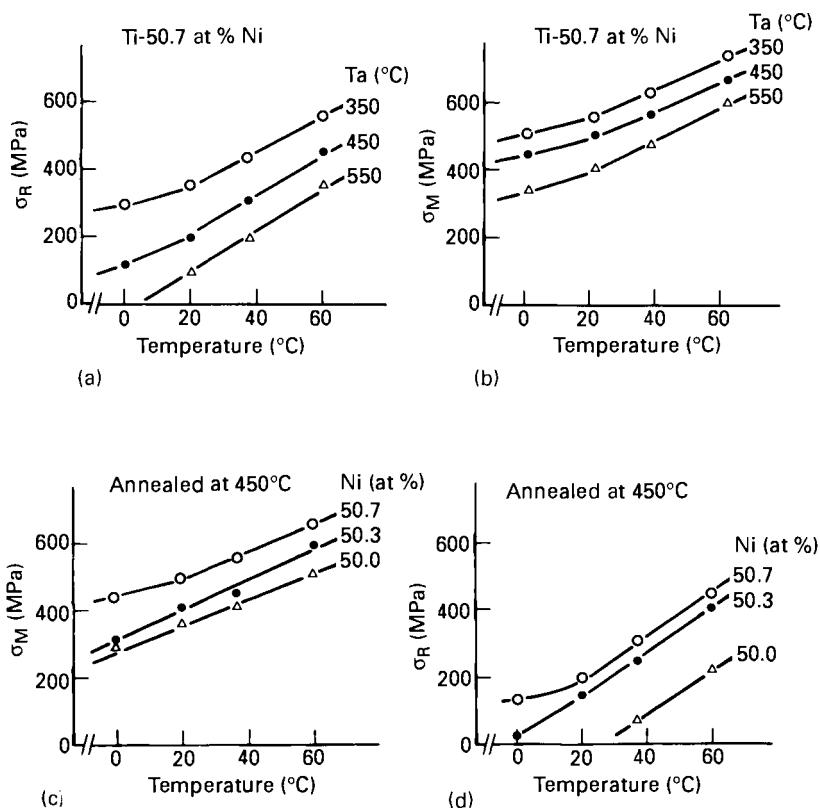


Figure 8: Stresses generated by SE NiTi can be affected by heat treatment (σ_m = critical stress for inducing martensite and σ_r = critical stress for inducing reverse transformation. Figures (c) and (d) show that another way of influencing SE NiTi alloys is by varying nickel composition.

with only a single alloy system. This treatment strategy is termed "variable load, constant force, uniform cross-section orthodontics."

4. SE NiTi Coil Springs

Coil springs have been used in orthodontic treatment for either opening or closing of extraction space for close to half a century. In the past they have been fabricated from stainless steel and chrome cobalt-nickel alloys. Recently coil springs fabricated from SE NiTi were introduced to the orthodontic profession. This appliance takes advantage of the superelastic properties of Ni-Ti.¹⁹ The mechanical characteristics of the SE NiTi coil spring provide for a better orthodontic appliance than those fabricated from either stainless steel or chrome-cobalt-nickel (Figure 9). Again in our clinical experience, the SE NiTi coil springs appear to provide greater efficiency in tooth movement than coil springs fabricated from other materials (Figures 10a and b).

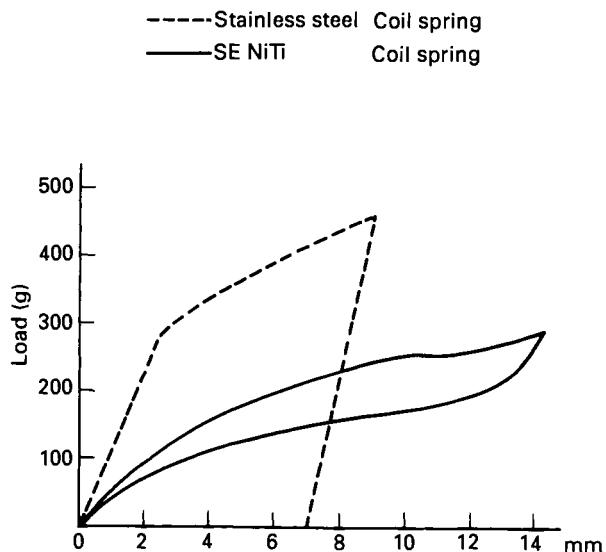


Figure 9 Load deflection characteristics of stainless steel and SE NiTi closed coil springs. Note that the stainless steel coil generates very high forces and shows a high residual strain. In comparison SE NiTi coil spring produces relatively constant forces that are low in magnitude and shows no permanent deformation for similar or higher activations.

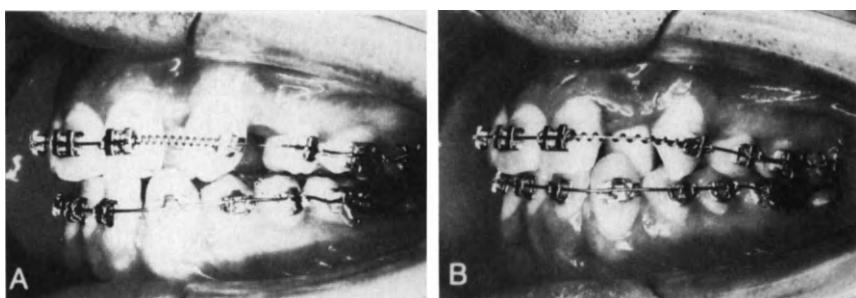


Figure 10: Distal movement of canine being accomplished by the use of SE NiTi coil spring [courtesy of F Muira American Journal of Orthodontics, Vol. 94, No.2, page 95, 1988].

5. Biocompatibility of Ni-Ti Alloys

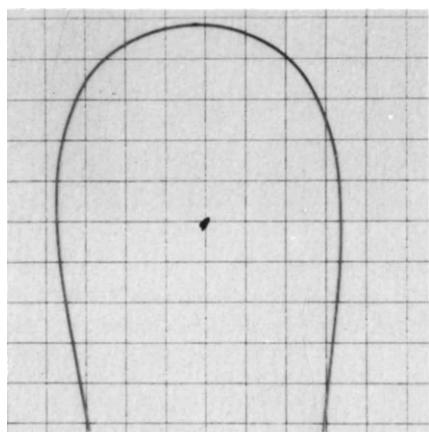
Stoichiometric Ni-Ti in its cold worked martensitic state was introduced to the field of orthodontics in the mid seventies. The binary SE NiTi alloy does not significantly differ in composition from this alloy. An exhaustive review of the medical literature by the authors has not revealed any case report regarding a hypersensitivity reaction in a patient as a direct result of using Ni-Ti alloys in the mouth.

Various *in vivo* and *in vitro* studies have determined that Ni-Ti is as biocompatible as the presently accepted materials.⁸³⁻⁸⁵ Cutright et al⁸⁶ examined Ni-Ti implanted subcutaneously in rats over a period of 9 weeks. No differences in inflammatory response compared with stainless steel was observed during the same period of implantation. No evidence exists in the literature to contraindicate the use of SE NiTi orthodontic appliances in nickel sensitive patients. However, this does not preclude the necessity of taking a full medical history of a patient by the orthodontist and alerting the patient to the fact that a nickel containing alloy is being used in his treatment.

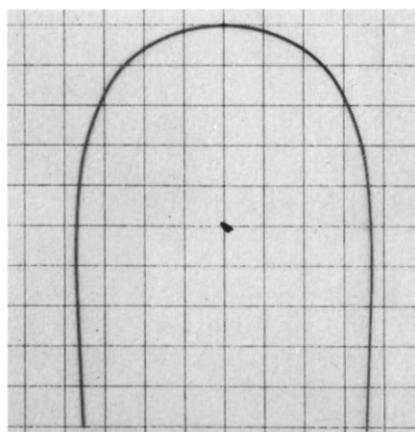
6. Two-way Shape Memory Effect

SE NiTi orthodontics arch wires exhibit the two way shape memory effect (Figures 11a,b, and c). This is a direct result of the manufacturing processes involved in the fabrication of the arch wire; it is repeatedly thermal cycled through the transition temperature range in a constrained condition to achieve the desired form. Therefore training of the material occurs and the arch wire demonstrates the two way shape memory⁶⁹ effect.⁷⁸

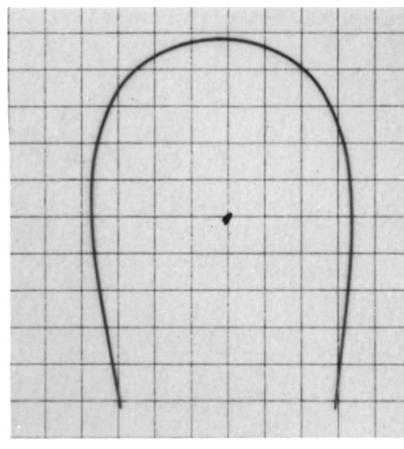
Recovery stresses generated by change in the shape of the arch wire are difficult to elucidate at the present. This is primarily due to the fact that the wire is constrained at multiple points of attachment, and render the system statically indeterminate. However, in our experience, this effect is of little consequence and aberrant tooth movement at the clinical level has not been observed. Any changes in the arch form in the orthodontic arch wire are temporary since temperatures in the oral cavity do stabilize towards the norm rapidly after ingestion of either hot or cold food.



(a)



(b)



(c)

Figure 11: Two way shape memory effect demonstrated by SE NiTi orthodontic arch wire: (a) orthodontic arch wire at room temperature (20°C), (b) the same arch wire at 37°C (mouth temperature), and (c) arch wires returns to original shape when cooled to room temperature (20°C). Note the changes in arch form are minimal.

7. Characterization and Standardization of SE NiTi Orthodontic Products

Any effort to optimize orthodontic mechanotherapy requires that an orthodontic appliance function in a reliable and reproducible manner when in clinical use. As a first step towards achieving this goal the manufacturing and the clinical communities need to develop both product specifications and their manufacturing tolerances.⁶⁶ As a second step the manufacturer needs to develop quality control measures that closely adhere to the developed product specifications. Finally the clinician must use these appliances in accordance with the suggested operational guidelines. These considerations become even more important when developing SE NiTi products, since their mechanical characteristics are very sensitive to slight variations in the manufacturing processes. Nickel composition, oxygen content, heat treatment, the amount of cold working and deformation temperature are some of the factors that can readily influence the mechanical properties of SE NiTi alloy.⁵⁸⁻⁶¹

The lack of standardization in the industry with regard to SE NiTi orthodontic products has led to great variation in the mechanical performance of these appliances. Two examples are discussed here to emphasize the necessity for standardization of SE NiTi orthodontic products. In the first example, tensile tests were performed at 37°C on two SE NiTi orthodontic archwire 0.46mm in diameters. These were selected from different batches of orthodontic archwires sold by the same manufacturer. The stresses generated by these products are different (Figure 12a), and as expected a significant difference exists between the martensitic start temperatures (M_s) between the two test samples (Figure 12b).

In the second case, the differences in the mechanical characteristics of the SE NiTi orthodontic products is further appreciated. The load deflection characteristics of

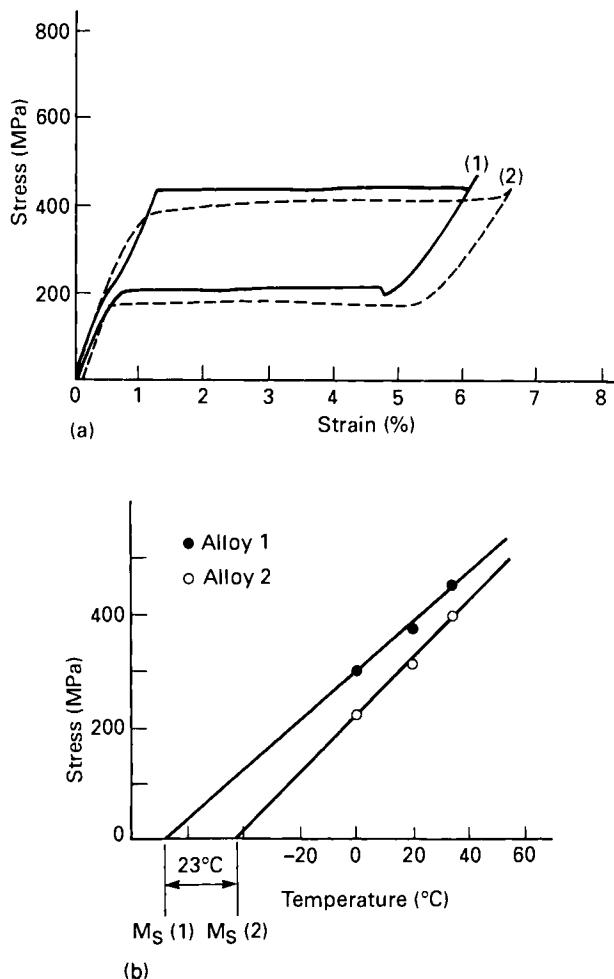


Figure 12: (a) Stress/strain curves of SE NiTi alloys 0.46mm in diameter sold by the same manufacturer. The difference in stresses produced by the wires is attributed to the lack of manufacturing tolerances for SE NiTi orthodontic products. The plot in (b) demonstrates the differences in the M_s temperatures between the two wires shown in (a).

0.41mm diameters SE NiTi orthodontic archwires sold by two different companies were tested at 37°C using a 3-point bending test (Figures 13a and b). Note only

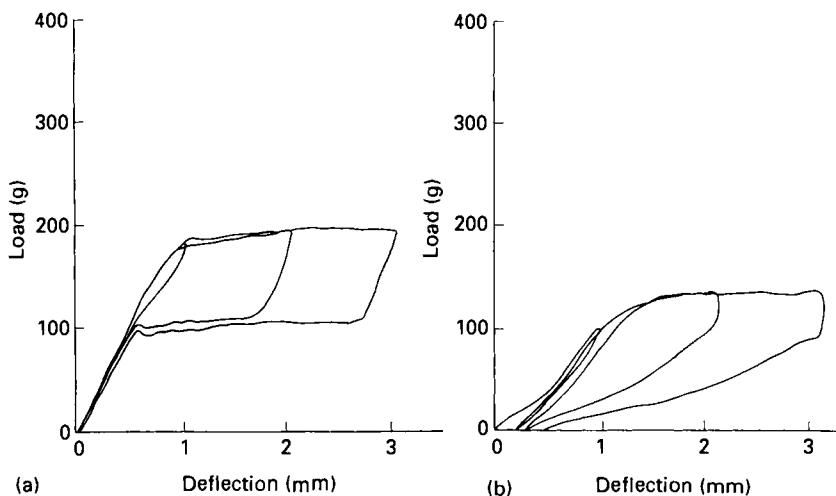


Figure 13: Load deflection characteristics of SE NiTi orthodontic archwires sold by two different companies. (Test temperature 37°C wire diameter 0.41mm, Test mode: 3-point bending.) Wire A exhibits ideal superelasticity. Wire B does not demonstrate ideal superelasticity and also shows considerable residual strain.

orthodontic archwire 'A' exhibits ideal superelasticity at mouth temperature. This appliance would certainly deliver the optimal force system to the dentition and elicit the desired biological response. Orthodontic archwire 'B' does not exhibit ideal superelasticity, which may be because of slip deformation occurring as result of a lower critical stress for slip. Also note that this wire shows considerable residual strain at mouth temperature. The M_s temperature of SE NiTi alloys is extremely sensitive to the slightest departures in the manufacturing processes of the alloy.^{60,61,80} It may therefore be used as a guide to establish product specifications and a check for manufacturing tolerances.

It is important to define the mechanical characteristics of SE NiTi orthodontic product at 37°C; the operational temperature of the appliance in the mouth and not room temperature as is frequently practiced. This is required since the deformation temperature affects the stresses generated by SE NiTi alloys above M_s .

The slope of the stress versus temperature plot of SE NiTi orthodontic products is of special interest to the orthodontist. A high value for this slope signifies the generation of very high in response to slight changes in deformation temperature. These stresses may well exceed the pain threshold and cause a great discomfort to a patient. In addition, such stresses even though transient in nature may cause cellular necrosis and retard tooth movement.

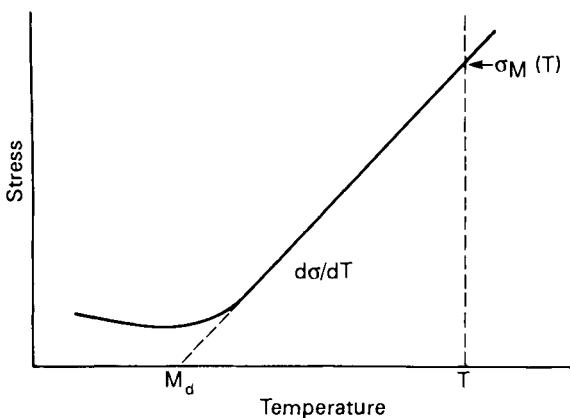


Figure 14: Schematic diagram demonstrating the parameters that should be characterized in SE NiTi orthodontic products: M_s , T_d (the deformation temperature), σ_{MS} (the critical stress for inducing martensitic, and $d\sigma/dT$ (the stress rate).

8. Summary

The superelastic characteristics of new equiatomic Ni-Ti have recently been taken advantage of in designing orthodontic appliances. The efficiency of orthodontic tooth movement has been greatly enhanced by employing orthodontic appliances fabricated from this alloy. This is especially true in situations requiring large deflections, i.e. in correcting very irregular teeth in patients. The future holds great promise for the use of SE NiTi alloy in all phases of orthodontic appliance treatment.

Acknowledgements

The authors express their sincerest thanks to Professor Marvin Wayman, University of Illinois, Dept. of Metallurgy and Mining Engineering, and Professor Otsuka, University of Tsukuba, Institute of Material Sciences, for all their encouragement and the support they have given us in our research. We thank Susan C. Brown for her secretarial assistance.

References:

1. "Elgiloy": Rocky Mountain Orthodontic Brochure (1977).
2. Andreasen, G.F., And Morrow, R.E.: Am. J. Orthod., **73**, 142 (1978).
3. Andreasen, G.F. And Hillerman, T.B.: J. Am. Dent. Assoc., **82**, 1373 (1971).
4. Andreasen, G.F.: Am. J. Orthod., **78**, 528 (1980).
5. Goldberg, A.J. And Burstone, C.J.: J. Dent. Res., **58**, 593 (1979).
6. Watanabe K.: J. Dent. Mat., **1**, 47 (1982).
7. Watanabe, K., et. al.: J. Jpn. Dent. Mat. (Special Issue), **2**, 9 (1983).
8. Ohura,Y., et al.: J. Jpn. Orthod. Soc., **42**, 494 (1983).
9. Ohura,Y., et al.: J. Jpn. Orthod. Soc., **43**, 590 (1984).

10. Ohura, Y.: J. Jpn. Orthod. Soc., 43, 71 (1984).
11. Ohura, Y., et al.: J. Jpn. Orthod. Soc., 41, 784 (1982).
12. Miura, F., Mogi, M., Ohura, Y., And Mamanaka, M.: Amer. J. Orthod., 90, 1 (1986).
13. Burstone, C.J., Qin, B., Morton, J.Y.: Am. J. Orthod., 87, 445 (1985).
14. Aiba, I., et al.: J. Dent. Res., Vol. 67 (1988).
15. Burstone, C.J.: Am. J. Orthod., 80, 1 (1981).
16. Kusy, R.P. And Stush, A.M.: Dent. Mat., 3, 207 (1987).
17. Drake, S.R., et al.: Am. J. Orthod., 82, 206 (1982).
18. Kusy, R.P.: Am. J. Orthod., 83, 374 (1983).
19. Muira, I., Mogi, M., Ohura, Y. And Karibe, M.: 94, 89 (1988).
20. Burstone, C.J.: Current Orthodontic Concepts and Techniques, Led. 2, Graber, T.M. And Swain, C.V. Mosby, eds., pp 23 (1985). .
21. Thurow, R.C.: Edgewire Orthodontics 4th Ed., C.V. Mosby, ed. (1982).
22. Rodan, G.A., Bourret, L.A., Harvey, A. And Mensi, T.: Science, 189, 467 (1975).
23. Kraus, M., et al.: Society For Biomaterials, 80 (1977).
24. Rodan, G.A.: Orthodontics - The State Of The Art, H.G. Barrer, ed.: University Of Pennsylvania Press Philadelphia, 315-322 (1981).
25. Deangelis, V.: Am. J. Orthod., 58, 284 (1970).
26. Epker, B.N. And Frost, H.M.: J. Dent. Res., 44, 33 (1965).
27. Justus, R. And Luft, H.J.: Calcif. Tissue Res., 5, 222 (1970).
28. Storey, E.: Aust. Dent. J., 57, 57 (1953).
29. Utley, R.K.: Am. J. Orthod., 54, 167 (1968).
30. Oppenheim, A.: Angle Orthodont., 34, 244 (1964).
31. Oppenheim, A.: Amer. J. Orthod. Oral Surg., 28, 263 (1942).
32. Roberts, W.E., et al.: Dent. Clin. North Am., 25, 3 (1981).
33. Zander, H.A. And Muhlemann, H.R.: Oral Surg., 9, 380 (1956).
34. Reitan, K.: Acta Odontol. Scand. (Suppl. 6) (1951).
35. Reitan, K.: Am. J. Orthod., 43, 32 (1957).
36. Reitan, K.: Am. J. Orthod., 46, 881(1960).
37. Reitan, K.: Angle Orthod., 34, 244 (1964).
38. Rygh, P.: Acta Odontol. Scand., 31, 109 (1973).
39. Rygh, P.: Scand. J. Dent. Res., 82, 57 (1974).
40. Oppenheim, A.: Amer. J. Orthod. Oral Surg., 30, 277 (1944).
41. Oppenheim, A.: Amer. J. Orthod. Oral Surg., 3, 57 (1911).
42. Oppenheim, A.: Acta Odont. Scandinav., 7, 115 (1947).
43. Oppenheim, A.: Amer. J. Orthodont., 43, 32 (1957).
44. Oppenheim, A.: Dent. Rec., 74, 271 (1954).
45. Oppenheim, A.: Amer. J. Orthod., 46, 881 (1960).
46. Oppenheim, A.: Dent. Rec., 73, 539 (1953).
47. Schwarz, A.M.: Internat. J. Ortho- Dont., 18, 331 (1932).
48. Henry, J., And Weinmann, J.P.: J. Am. Dent. Assoc., 42, 280 (1951).
49. Hotz, R.: Fortschr. Kieferorthop., 27, 220 (1966).
50. Nakamura, M.: J. Osaka Dent. Univ., 2, 1 (1968).
51. Tayer, B.H., Gianelly, A.A., And Ruben, M.P.: Am. J. Orthod., 54, 515 (1968).
52. Stuteville, O.H.: J. Am. Dent. Assoc., 24, 1494 (1937).
53. Storey, E. And Smith, R.: Aust. J. Dent., 56, 11 (1952).
54. Oppenheim, A.: Am. J. Orthod., 28, 263 (1942).
55. Wainwright, W.M.: Am. J. Orthod., 64, 278 (1973).
56. Reitan, K.: Current Orthodontic Concepts and Techniques, vol. 1, Graber,T.M., Swain, B.F. and C.V. Mosby, eds., 101-159 (1981).

57. Gianelly, A.A. And Goldman, H.M: Biological Basis of Orthodontics, Lea & Febiger, 116-204 (1971).

References on the general behavior of Ni-Ti:

58. Shimizu, K.: Bull. J. Inst. Met., 24, 13 (1985).
59. Miyazaki, S., Ohmi, Y., Otsuka, K. And Suzuki, Y.: Journal De Physique, Colloque C4, (Suppl. 12), 43 (1982).
60. Delaey, L., Krishnan, R.V., Tas, H. And Warlimont, H.: J. Mater. Sci., 9, 1521 (1974).
61. Miyazaki, S., And Sachdeva, R.: J. Dent. Res., 66 (Special Issue):349 (1987).
62. Sachdeva, R. And Miyazaki, S.: J. Dent. Res., 66, 1939 (1987).
63. Melton, K.N., And Mercier, O.: Acta Metall., 27, 137 (1979).
64. Edwards, G.R., Perkins, J., And Johnson, J.M.: Scr. Metall., 9, 1167 (1975).
65. Mercier, O. And Melton, K.N.: Scr. Metall., 10, 1075 (1976).
66. Duerig, T.W. And Melton, K.N., Shape Memory Alloy '86, China Academic Publishers, 397. (1986).
67. Bush, R., Leudeman, R.T., And Gross, P.M.: Amra Cr, 65, 02-1, Ad 629726, U.S. Army Materials Research Agency (1966).
68. Saburi T. And Wayman, C.M.: Acta Metall., 27, 970 (1979).
69. Honma, T.: Titanium And Zirconium, 27, 60 (1979).
70. Miyazaki, S., Otsuka, K., And Suzuki, Y.: Scr. Met., 15, 287 (1981).
71. Wang, F. And Buehler, W.: Appl. Phys. Lett., 21, 105 (1972).
72. Jackson, C.M., Wagner, H.J. And Wasilewski, R.J.: Nasa Sp - 5110, (1972).
73. Tas, H., Delaey, L., And Deruyttere, A.: J. Less-Common Met., 28, 141 (1972).
74. Wasilewski, R.J.: Metall. Trans., 2, 2973 (1971).
75. Wasilewski, R.J.: Scr. Metall., 5, 207 (1971).
76. Olsen, G.B. And Cohen, M.: Scr. Metall., 9, 1247 (1975).
77. Mohamed, H.A. And Cohen, M.: Met. Trans., 7A, 1041 (1976).
78. Wayman, C.M. And Shimizu, K.: Met. Sci. J., 6, 175 (1972).
79. Otsuka, K. And Shimizu, K.: Proc. Icomat, 79, 607 (1979).
80. Ostuka, K., And Wayman, C.M.: Reviews on the Deformation Behavior of Materials, Freud Publishing House, Tel-Aviv, Israel, 81-172 (1977).
81. Ostuka, K. And Shimizu, K.: Solid State Phase Transformations, H. Aaronson, ed., AIME, Pittsburgh Conference, 1267-1286 (1982).
82. Golestaneh, A.A.: Proc. Icomat 79, Mit Press, Cambridge, 679 (1979).

References on the biocompatibility of Ni-Ti:

83. Castleman, L.S., Motzkin, S.M., Alicandri, F.P., Bonawit, V.L. And Johnson, A.A.: J. Biomed. Mater. Res., 10, 695 (1976).
84. Castleman, L. S.: Proc. 5th Annl. Int. Biomaterials Symp., Clemson University, Clemson, S.C. (1973).
85. Motzkin, S.M., et al.: Proc. 4th New England Bioengineering Conf., Yale University, New Haven, Conn., 301 (1976).
86. Cutright, D.E., et al.: Oral Surg., 35, 578 (1973).
87. Spiechowicz, et al.: Contact Dermatitis, 10, 206 (1984).
88. Heisterkamp, C.A., Buehler, W.J. And Wang, F.E.: " 55-Nitinol -- a New Biomaterial", 8th annual conference on Medical And Biomedical Engineering, Chicago (1979).

89. Schmerling, M.A., et al.: J. Biomed. Mater. Res., **10**, 879 (1976).
90. Hughes, J.L.: "Evaluation of Nitinol for Use as a material in the Construction of Orthopaedic Implants", Damd., 17-74-C-4041. U.S. Army Medical Research And Development Command, Fort Detrick, Frederick, MD (1977).
91. Vreeburg, K.J.J., et al: J. Dent. Res., **63**, 124 (1984).
92. Haasters, J., Bensmann, G. And Baumgart, T.: "Memory Alloys - New Material for Implantation in ORthopedic Surgery, Part 2", in Current Concepts of Internal Fixation of Fractures, Uhthoff. H.K., ed., Springer-Verlag, New York, p128 (1980)

Shape Memory Implants

Sekio Fukuyo^{1,4}, Yuichi Suzuki^{2,4}, Kazuo Suzuki^{3,4} and Eikou Sairenji^{1,4}

1)Nihon University School of Dentistry

2)Furukawa Denko

3)Matsumoto Dental College

4)The Academy of Shape Memory Material for Medical Use

The Shape memory implant (SMI) is a dental endosseous implant made of Ni-Ti shape memory alloy. The device was given official approval by the Ministry of Health and Social Welfare of Japan for use as a medical implant on June 1, 1985. In this paper, information regarding the biocompatibility of the device will be presented, as well as how the SMI is used in clinical operation, and the indications and advantages of the implant compared with the other dental implants.

1. Biocompatibility

Cell cultures were studied using L181 cells, which is anchoring-dependent cell from connective tissue from the rear leg of a rat. Figures 1 and 2 show low and high magnification enlargements of the Ni-Ti specimen; the L181 cell has clearly grown onto the base material. No toxicity of Ni-Ti is observed.

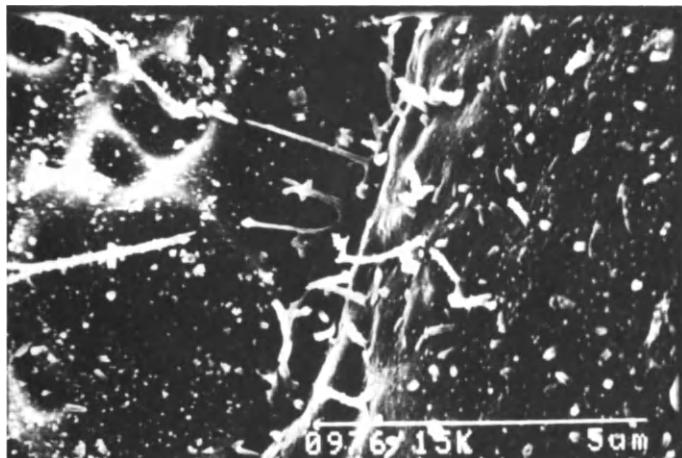


Figure 1: Low magnification SEM micrograph of Ni-Ti after implantation in a rat.

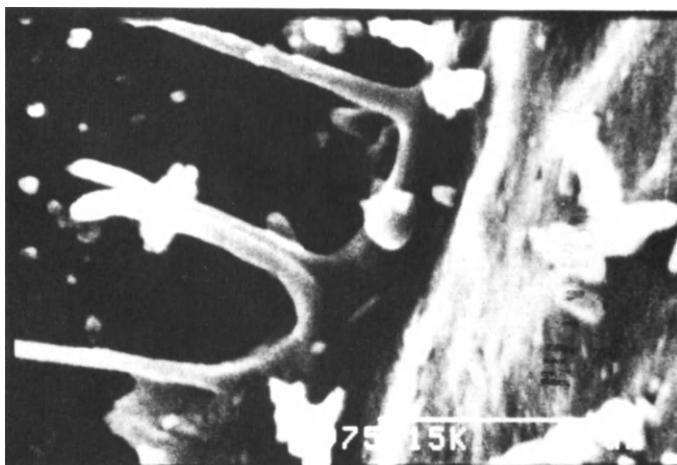


Figure 2: High magnification SEM micrograph of Ni-Ti after implantation in a rat.

We have studied histology by implanted SMI's into the jawbones of Japanese monkeys and dogs. Figure 3 shows an x-ray photograph 6 months after such an operation and Figure 4 shows the apex of the SMI. Figure 5 shows the same part 6 months after operation. Newly generated bone tissue is observed surrounding the Ni-Ti shape memory implant. New bone contacted or attached tightly to shape memory implant and connective tissue is not observed. From these results, we conclude that the shape memory implant has good biocompatibility.

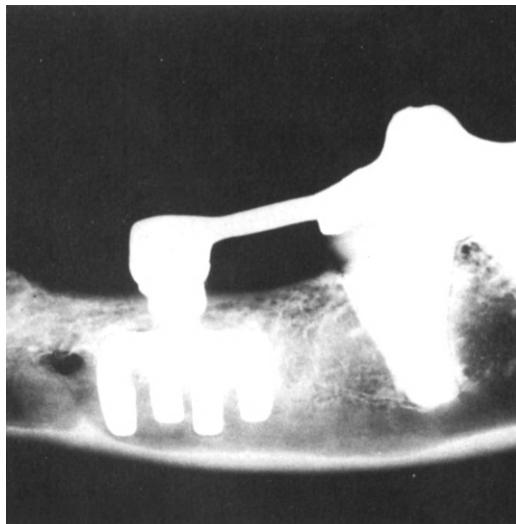


Figure 3: An X-ray of an SMI implant 6 months after installation.

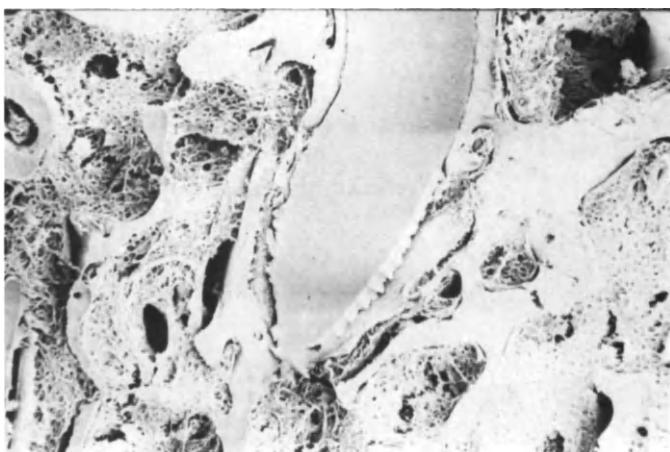


Figure 4: An SEM photograph showing the apex of an implant.

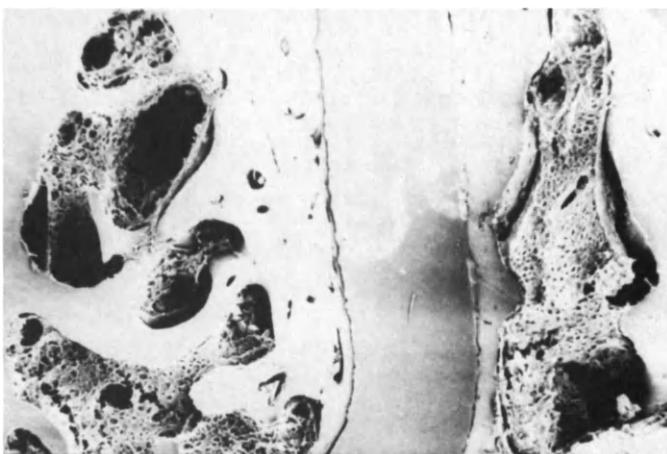


Figure 5: An SEM photograph showing the body of an implant.

2. Design of the Shape Memory Implant

There are now two types of shape memory implants: single wing and double wing. Figure 6 shows the single wing type shape memory implant; part of the apex opens bucco-lingual after insertion into jaw bone. Figure 7 shows double wing type; parts of the apex and vent opened as in the case of the single wing type. This effect has a very strong ability of mastication compared with the ordinary non-opening blade type implant.

The stresses in the implant were analyzed by the Finite Element Method. Figure 8 shows a cross-sectional cut of the ordinary blade implants. Bite forces of 60 kg from a vertical direction were supported before the implant was reduced to half its height. Figure 9 shows a similar analysis of the single wing SMI: here the height reduction at the same 60 kg load is half that of the ordinary implant. Figure 10 shows that the double wing SMI compresses only 1/4 as much as the ordinary implant.

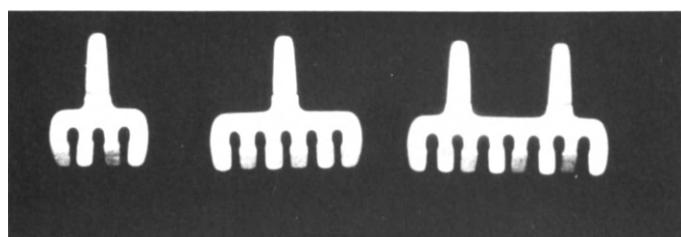


Figure 6: Photograph showing the single wing type of implant.

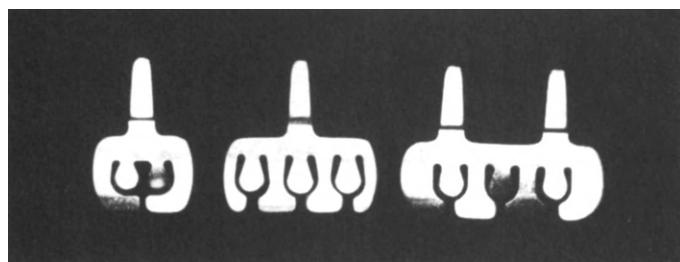


Figure 7: Photograph showing the double wing type of implant.

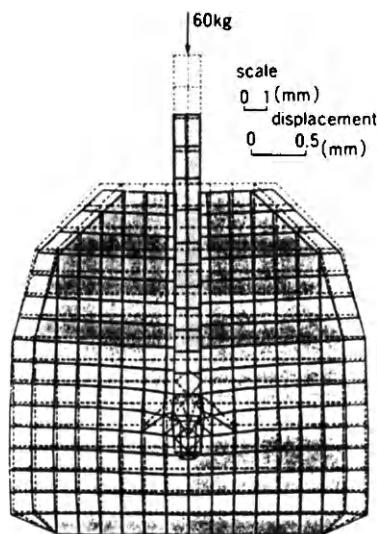


Figure 8: A finite element analysis of an installed conventional dental implant showing significant deflection under a 60 kg load.

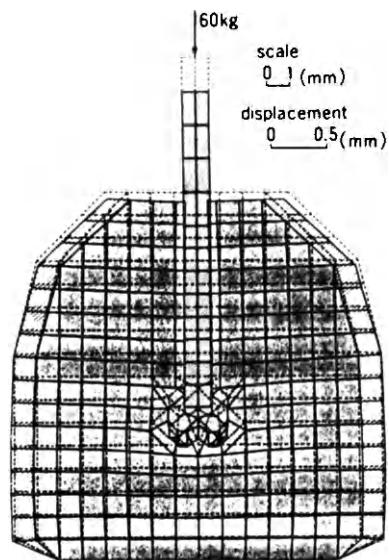


Figure 9: A finite element analysis of a single wing SMA implant showing a much smaller deflection than the conventional implant.

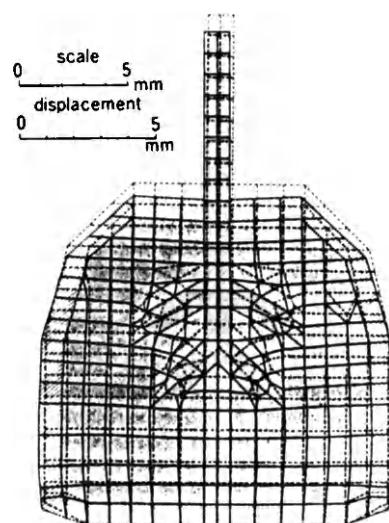


Figure 10: A finite element analysis of a double wing SMA implant showing still further reductions in deflection.

3. Clinical Studies

Since 1985, there have been approximately 15,000 successful clinical applications of the SMI in Japan. Figure 11 and Figure 12 shows upper edentulous and lower bilateral cases 6 years after operation. Figure 13 and Figure 14 shows upper and lower bilateral cases 4 years after operation. Clinical results have shown good mastication.

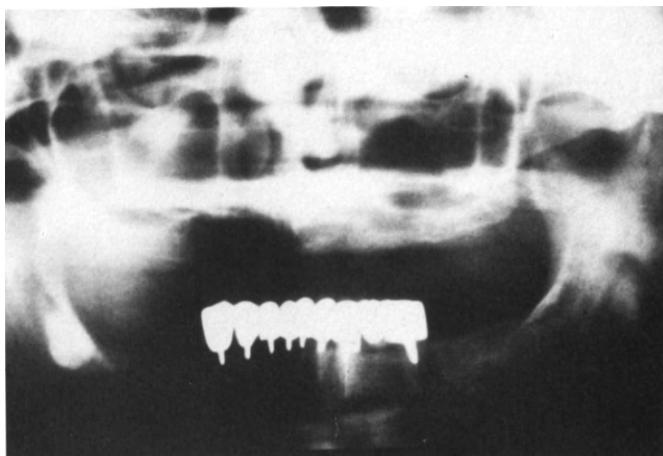


Figure 11: X-ray showing an upper edentulous case 6 years after operation.

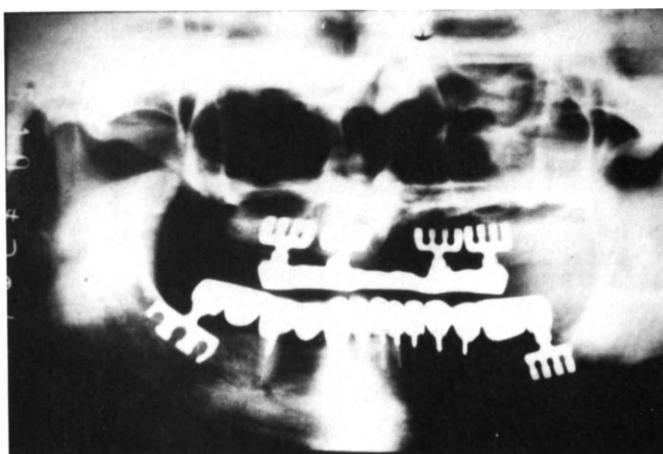


Figure 12: X-ray showing a lower edentulous case 6 years after operation.

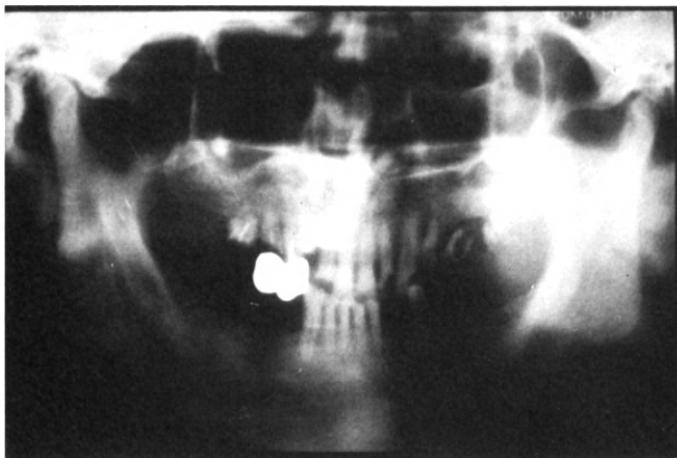


Figure 13: An upper bilateral case 4 years after operation.

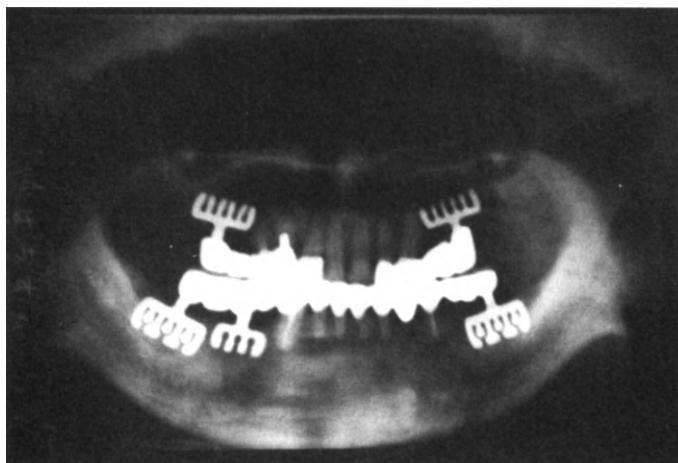


Figure 14: A lower bilateral case 4 years after operation.

The advantages of the shape memory implants are that they exert strong and continuous forces of mastication, have good initial fixation in jaw bone, are easily installed and require a simple operation, and that they have a good stress dispersion from the shape memory effect.

References

1. Fukuyo, S., et al., Development of Shape Memory Implant, 71st International Dental Federation Tokyo, (1983).
2. Fukuyo, S., Shape Memory Implant, Jap. Medical Culture Center.Tokyo, (1985).
3. Fukuyo, S., et al., Dental Implants, 1, (1986).
4. Fukuyo, S., et al., Iijon University Dental J. 59, 614-625 (1985).
5. Fukuyo, S., International Journal of Oral & Maxillofacial Implants.2(3), (1988).

The Use Of Ni-Ti In The Homer Mammalok®

James P. O'Leary*, James E. Nicholson† and Roland F. Gatturnat

* Engineering Design
Tufts University
Medford, Massachusetts

† Mitek Surgical Products, Inc.,
Dedham, Massachusetts

The Homer Mammalok is a needle wire localizer designed to utilize the superelastic characteristic of nickel-titanium to meet an important and demanding need in the medical field. We will explain the nature of the problem, describe the history of the design, and discuss the ramifications of the material choice in the evolution of the product. It is hoped that our experience with this material will be helpful to others who design products incorporating Ni-Ti alloys.

1. Problem Setting

The high incidence of breast cancer has led to very aggressive efforts at early diagnosis and therapy. It is now recommended that healthy women over the age of forty have mammograms every other year¹. This procedure frequently identifies small lesions which must be removed and examined microscopically. The procedure for removing the lesion is made difficult because there are no landmarks to guide the surgeon. The lesion can not be easily distinguished in any obvious way from the surrounding tissue. There is also a requirement that there be as little as possible excess tissue removed. This is one of a class of problems in "localization". Soft tissue, although having definite structures, can change shape so radically that locating these structures by conventional means such as measurements from fixed points is impossible.

The present method of dealing with the problem is for the radiologist to place a needle in the breast at the site of the lesion as a guide to the surgeon during the removal². Briefly, the steps followed to excise the lesion are:

- (1) Take two mammograms showing the lesion in a top and a sideview.
- (2) Insert the needle into the breast using the mammograms as a guide to the estimated location of the lesion.
- (3) Take two additional mammograms to confirm the correct positioning of the needle.
- (4) Surgically remove the lesion with sufficient margin of healthy tissue, leaving the removed segment and the needle intact.
- (5) X-ray the removed specimen to verify that the entire lesion has been excised.
- (6) Pathologically examine frozen sections from the specimen.

The mammograms are taken in radiology. The patient is then taken to the operating room for surgery. The needle must be secured in transit. If there is motion of the needle when the patient is transported, the localizer will move, therefore marking the wrong tissue to be removed surgically.

2. Problem Definition:

The need for a better localizing product was defined; constrain the needle preventing accidental advancement or withdrawal during transit. A number of solutions had been worked out by various groups and some products were on the market but the problems of needle motion remained. It was necessary to secure the localizer in place. The first attempt was to attach the device at the skin. Some have done this with tape and others have used sutures. Although the device may be fixed at the skin there is no means to prevent internal motion. An internal tag is needed to mark the lesion within the breast. One obvious way to attach a needle to the tissue near the lesion is to place a barb, as on a fish hook, at the end of a needle³. This would keep the needle from moving out of the breast, but if the needle is reasonably sharp, it could advance and even puncture the lung. A second disadvantage is that the needle can not be repositioned if the mammogram after placement shows that the needle was improperly positioned. The widely used localizer product presently in use incorporates a barb like design fashioned by a reverse bend in the end of the wire (Figure 1). The wire with the hook

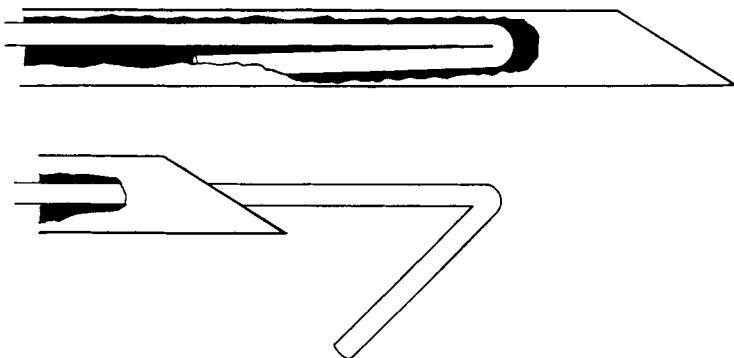


Figure 1. Concept drawing of barb formed from bent wire.

is placed within a cannula (hypodermic tube), and inserted into the breast by the radiologist using the procedure previously outlined. When position is verified by a mammogram the wire is pushed out of the cannula deploying the barb. This design requires that the wire be very thin since it is doubled in the tube, and a reasonable amount of recovery of position is needed to define the barb. The thin wire is regarded as a problem because it can be inadvertently cut during surgery.

The idea which was the starting point for the new product was a hook which would be self deploying (Figure 2). Originally this was to be thermally activated using a bimetallic structure which would deploy when the wire reached body temperature. A search for the right material combination was fruitless but it turned up the suggestion that Ni-Ti be looked into for its shape memory behavior. The search turned toward

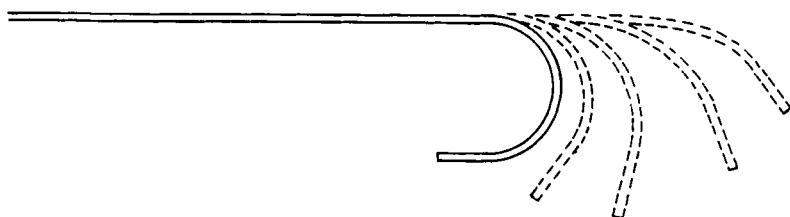


Figure 2. Thermally activated hook concept.

finding an alloy which would transform at about 33°C. The design would be a wire inserted as a straight element and form a hook as it approached body temperature (37°C). It was thought that there would be sufficient time to reposition, if it were necessary, before the hook formed.

In exploring Ni-Ti the superelastic characteristic came to our attention which opened the way for the current design⁴. This consists of a preformed curved hook which is held straight in a cannula and returns to its curved shape as it is pushed out (Figure 3).

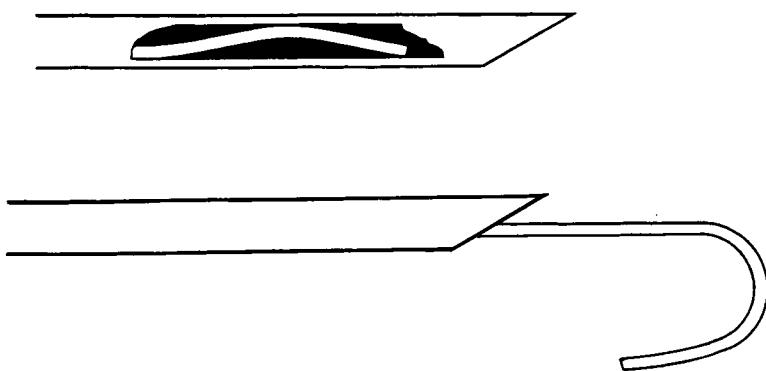


Figure 3. Preformed hook concept: with wire in withdrawn position and deployed position.

Because the wire is following a curved path as it is deployed, it only meets resistance at the tip. The wire is never trying to move laterally through the tissue. With this design it seemed that a heavier wire was practical. There seemed to be potential for a very good solution.

3. Design Requirements:

The following requirements were identified as imperative:

- (a) Full recovery of the preformed hook.
- (b) High strength against transection in surgery
- (c) Reasonable hook size
- (d) Function with #20 gauge hypodermic needle (0.6mm or 0.024 in. I.D.).
- (e) Biocompatibility of the materials which invade the breast.

4. Geometry:

The design details developed around the radius of the hook. A size was needed which would satisfy the surgeon and radiologist that an appropriate specimen had been tagged. The radius chosen was 9mm (0.36 in.). It was also decided that a very conservative factor of safety be used because of the nature of the application. The major loading which occurs is due to the straightening of the wire in the cannula. Since this kind of load is not subject to any severe fluctuations, an operating strain of four percent was considered safe. This is a single use disposable product and does not get cycled more than four or five times in normal use. The wire diameter is 0.4mm (0.15 in.) which is sufficient to prevent accidental cutting with a scalpel. This utilizes the internal diameter of the hypodermic tube much more efficiently than could a folded design.

Comparing the design obtained in Ni-Ti with an implementation in stainless steel with the same base point is startling. If the same diameter, stainless steel wire is used the hook will have a radius of 50mm (2.0 in.). If the hook radius is maintained, the wire diameter becomes 0.05mm (0.002 in.). Using titanium is only slightly better with wire diameter of 0.1mm or a hook radius of 25mm.

5. Biocompatibility:

Ni-Ti has been used in a variety of medical applications but the issue of material compatibility does not go away with a product that invades the body. There have been no cases of adverse tissue reaction resulting from Ni-Ti⁵ so from a design viewpoint there was no difficulty. Each melt of material used is tested for toxicity by an independent laboratory as part of the qualification process.

6. Difficulties:

The question of fatigue failures and design requirements was mentioned earlier. Due to the nature of the design, it is unlikely that the product will exhibit mechanical failure from the low stress and strain imposed during use. The normal "life" of the product during fabrication is that the wire is formed and heat treated, cleaned and assembled in the cannula with the hook extended. The assembly is then packaged and sterilized. At use, the sterile package is opened and the wire is retracted into the cannula until the hook just disappears within the cannula. The device is then inserted by the radiologist into the breast and the hook is deployed. If the mammograms show correct positioning of the hook relative to the lesion, the patient is transferred to surgery where the lesion is removed. One or two repositionings may be required until the wire is positioned satisfactorily with respect to the lesion. Several more are not beyond reason. No fatigue failures have occurred in this setting. If on the other hand, the product is pulled out to demonstrate its use in a non-sterile setting, an entirely new range of activities may begin. The hook may be cycled hundreds of times, the wire may get bent to a very sharp radius to demonstrate its flexibility, and other forms of experiment may ensue leading to either permanent offset of the hook or fatigue and fracture of the wire.

This latter sequence of activities is in one sense not at all related to the performance of the product but it can not be ignored. The perceived durability of the product will have much to do with its acceptance. This is not a problem unique to this product but here the line is very clear. In the opinion of the designers, the developed design is the

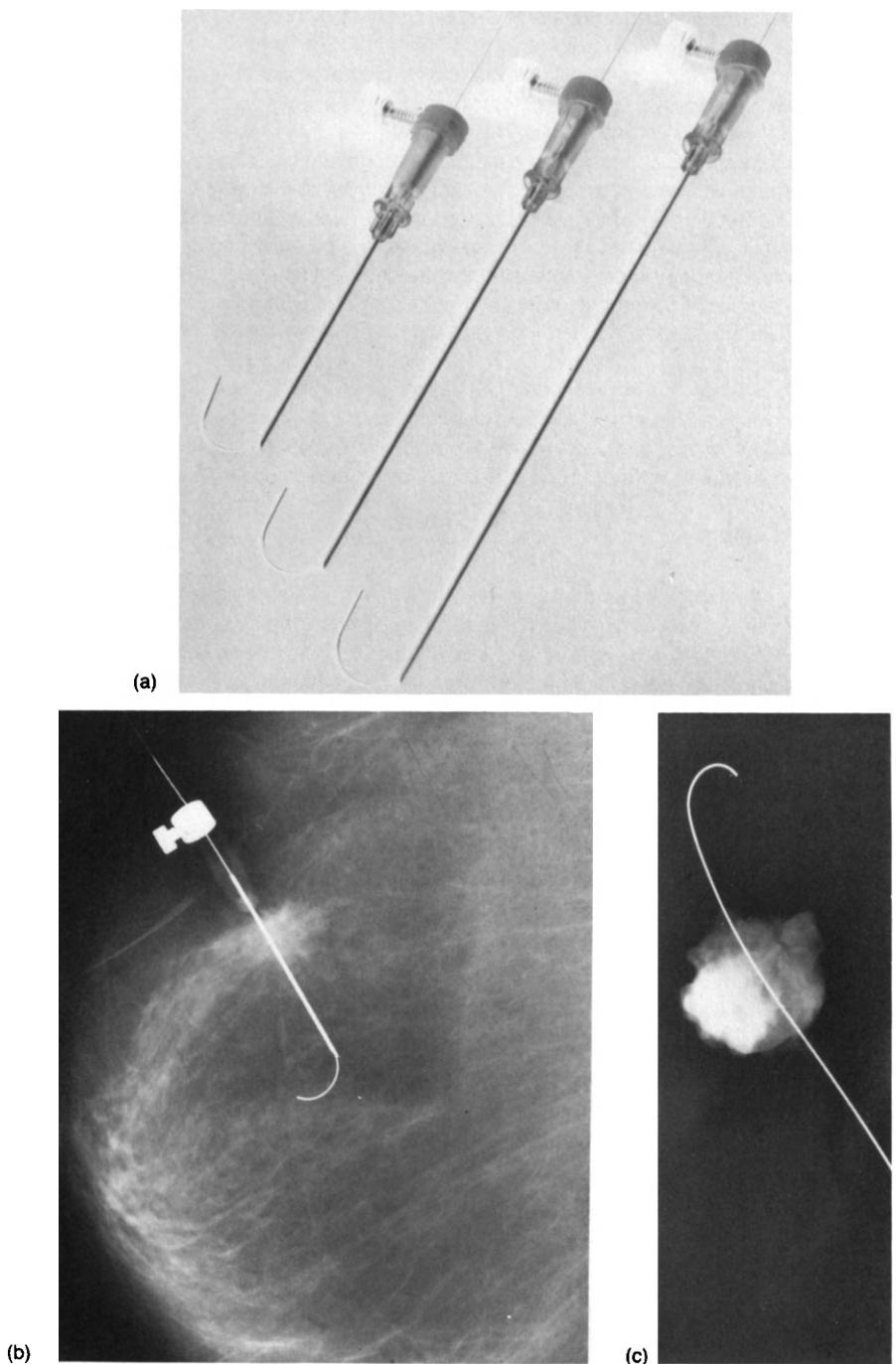


Figure 4. (a) photograph of Homer Mammalok, (b) radiograph of device in use, and (c) X-ray of lesion after excision.

optimum for product performance. Should one alter the design to obtain a configuration which may be more durable in incidental handling yet inferior in some small way to the original design? The preferred course of action is that the design remain fundamentally unchanged and that the package or presentation be altered in some way to minimize extracurricular manipulation. That avenue is currently under study but it is not an easy task. This aspect of the problem is included here because we feel that this will be a recurring issue when products are largely Ni-Ti based. In other applications where product is composed of several materials in addition to the Ni-Ti element, the issue may not be important. The fact that the user is sophisticated (i.e., a radiologist or a surgeon) does not help the fundamental design problem. Other minor problems have come up along the way as in any design. Surface quality of the wire became especially important to smooth operation of the wire within the cannula. Low production runs and having few people experienced with the material seems to turn every "secondary operation" into a primary development project. As more and more parts are manufactured at high volume with these alloys this problem will diminish, but for now it should not be taken for granted that a few prototypes can be easily turned into many thousands of products which perform consistently.

7. Conclusion:

The Homer Mammalok® was designed to meet a set of design criteria as previously outlined. From the engineering point of view, these criteria have all been met. The product has also been well received in the medical community, being accepted by the radiologists who place the units, the surgeons who excise the lesions with the unit still attached and the pathologists who are responsible for evaluating the results of this procedure. In the commercial sector, the product has gained a significant fraction of the market. There are more than 500,000 procedures in this country annually, and that number is expected to grow. The Homer Mammalok® has been used in over 300,000 of these cases and there have been no product problems. The authors are pleased that it performs as designed and has generated no surprises.

References

1. American Cancer Society, Cancer Facts and Figures, page 10, (1987).
2. Homer, M.J., Smith, T.J., Marchant, D.J.: JAMA, 252, 2452 (1984).
3. Kopans, D.B., Meyer, J.E.: AJR; 138, 586 (1982).
4. Homer, M.J.: Radiology, 157, 259 (1985).
5. Castleman, L.S., et al.: J Biomed Res; 10 695 (1976).

The Use of Superelasticity in Guidewires and Arthroscopic Instrumentation

Jim Stice,
Flexmedics Corporation
420 North Fifth Street, #620
Minneapolis, MN 55401

Two different medical devices which use Ni-Ti superelasticity have been developed for clinical evaluation. Each provides benefits that should advance the use of less invasive surgical procedures. One is a guidewire for use in a number of different medical specialities. The second is an arthroscopic needle for suturing meniscal tears. A short description of the clinical application for both devices will be followed by a discussion of device design.

There is a general trend in the medical services sector towards cost containment. Each source of financing for medical services has adopted various forms of limits on payments for particular kinds of procedures. The response of hospitals and other service providers has been to attempt to reduce costs so as to be profitable with constrained revenues. One means to reduce costs and still maintain quality service is to shorten the patient's hospital stay by using less invasive and more effective procedures. Less invasive procedures often require the ability to enter the body through natural openings of very small incisions and to maneuver diagnostic or therapeutic devices to the desired position in the body.

The limited access available and sought of less invasive procedures makes for limited design options, thus devices must have very small profiles. They are generally as long as needed with the smallest possible circular cross section. In addition, it is critical not to deviate too much from more standard devices when creating new devices, since "feel" during use is important and dramatic changes will often be rejected despite other promising performance factors. While these devices might be relatively simple from a design perspective, the processing required to achieve these "simple" devices can become quite complex and challenging. One of the biggest processing problems is creating safe and low cost joints and assemblies for multiple component devices using Ni-Ti alloys.

Both of devices to be described use the superelastic properties of Ni-Ti alloys, rather than the thermal memory. The relatively small temperature changes tolerated by tissue and the limited access available when doing surgical procedures make it difficult to engineer devices using Ni-Ti thermal memory properties (shape memory effect). In, addition, both of the following devices use Ni-Ti wire since it is readily available and has the required circular cross section for low profile. Ni-Ti is the preferred shape memory alloy because of its superior mechanical and biocompatibility properties.

1. Guidewires

A guidewire is a long, thin metallic wire which is passed into the body by a physician through natural openings of small incisions. It serves as a guide for safe introduction of various therapeutic and diagnostic devices.

A guidewire is used because its special mechanical properties provide the necessary balance between controllability and flexibility needed to safely and effectively negotiate the circuitous paths of the body's inner core. Once the guidewire is in place, a hollow tube or catheter is inserted over the guidewire and into the desired location. The wire may then be removed and the appropriate diagnostic or therapeutic procedure performed. Guidewires are predominantly used in cardiology and radiology and to a lesser extent in gastroenterology and urology.

There are two typical methods of construction for guidewires. One type is a single strand of wire with a flexible coiled tip construction (Figure 1a). The second is a helical

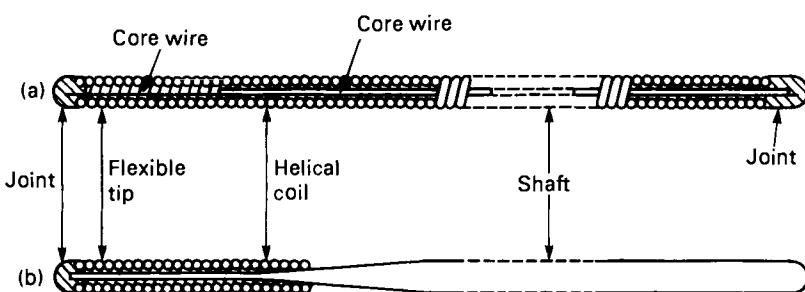


Figure 1: Standard guidewire configurations.

coil of wire with one or several straight single strand wires different mechanical properties are available. One specialty design uses a plastic tube as a partial replacement for the spring coil, but retains a single strand core wire.

Guidewires come in a variety of sizes. Lengths range from about 25 cm to 400 cm, while outside diameters vary from 0.35mm to 1.0mm. The tips which are advanced into the body are more flexible than the shaft to reduce the possibility of vessel perforations. This increased tip flexibility is achieved by using a smaller diameter core wire at the tip.

Other design considerations for guidewires include the use of PTFE (Teflon) coatings to reduce frictional drag while moving the guidewire in the vessel. Many times the guidewire movement is monitored by the physician with X-ray equipment. In such instances, it is critical that the wire be visible by absorbing X-rays. Guidewires will come with straight or curved tips to increase maneuverability in the vessel. Finally, the strength of the coil to core wire joint is important so it does not fail during use.

Stainless steel is now used almost exclusively for guidewires. The anticipated benefits of superelastic Ni-Ti to replace stainless steel include (a) diminishing the complication of the guidewire taking a permanent kink (which may be difficult to remove from the patient without injury), and (b) increased steerability to pass the wire to the desired location due to an anticipated improvement in torquability (an ability to translate a twist at one end of the guidewire into a turn of nearly identical degree at the other end).

The design of Figure 1a is preferred for Ni-Ti wire since it has greater torquability. Stainless steel wires using design 1a cannot be matched in shaft stiffness when using Ni-Ti alloys and design 1b. It is uncertain at this time how lower stiffness will affect the clinical performance of a Ni-Ti wire.

Several PTFE coatings have been successfully applied with standard processes to Ni-Ti wire. Some care is required since the cure temperatures of the coatings are high enough in some instances to change Ni-Ti mechanical properties. No design modifications are needed when replacing stainless steel with Ni-Ti for purposes of radiopaqueness since it is as good as stainless steel in this regard. Ni-Ti guidewires with both curved and straight tips have been fabricated.

The biggest technical challenge has been getting tip strengths of 0.5 kg or greater. Welding or soldering of the core wire to the coil is used predominantly in stainless wires. This has proven to be difficult with Ni-Ti using the same techniques. However, sufficient strength has been achieved by a change in joining materials and configuration of the core wire tip.

These findings indicate that a safe wire has been developed. Clinical studies will determine its effectiveness.

2. Arthroscopic Suture Needles

The use of fiberoptic scopes in orthopedics began to grow significantly in the late 1970's. Initially the scopes were used for diagnostic purposes, but surgeons have increasingly made it a practice to use a scope therapeutically, avoiding large surgical incisions. Arthroscopy is being used increasingly to achieve less-invasive orthopedic surgeries.

The meniscus (cartilage which prevents upper and lower leg bones from rubbing) has been traditionally removed (either totally or partially by surgeons) to repair an injury or tear. However, the removal of the meniscus has been shown to contribute to arthritis. This finding has led to the development of methods to suture the meniscus as a means of treating an injury.

Guiding cannulas have been developed to assist in the delivery of needles to tissue otherwise unreachable to the arthroscopist without significant tissue trauma. Such cannulas are typically rigid, hollow, tubular devices, usually constructed of stainless steel and may be straight or curved.

Metallic needles are used to suture a tear in the meniscus. These needles are 20 to 25 cm long with a diameter of 0.8mm to 1.0mm. The needle includes a sharp end and means to pull a suture from its other end. The needle is able to bend along the path of the cannula. A pair of such needles are attached at their back ends to opposite ends of a suture. Figure 2a shows one style with the suture permanently attached by means of swaging. The other needle type, Figure 2b, has a hole at one end which allows for suture attachment by passing the suture through the hole.

In use, a curved cannula is inserted into the knee with the forward end of the cannula adjacent to the operative location. One of the needles is then inserted into the cannula from the back end. The sharp end of the needle is forced through the forward end of the cannula and emerges in a substantially straight orientation as it advances toward the operative location. The needle is then advanced through the appropriate tissue, exiting out the other side of the knee. This is repeated with the second needle on the opposite side of the meniscus tear. The needles are then separated from the suture and the suture is snugged up and tied off to hold the tissue together.

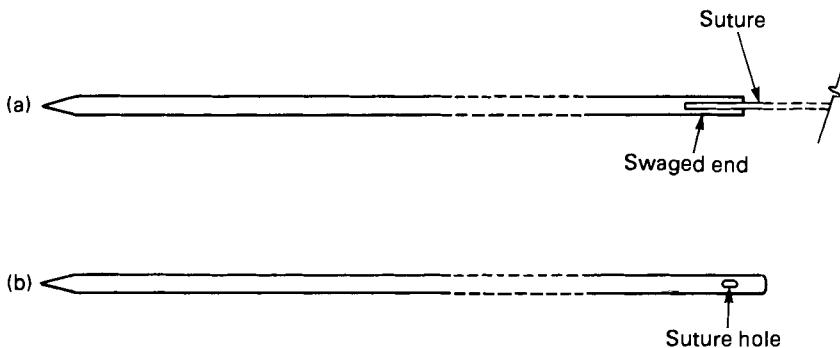


Figure 2: Standard arthroscopic needles.

Meniscal suture needles have been fabricated of stainless steel. The advantages of a superelastic Ni-Ti wire comes when using a curved cannula. In this instance, the Ni-Ti needle takes much less set or shape change when passed through a curved cannula. The anticipated advantage is greater controllability with a "straighter" needle. Critical design attributes for a meniscal suture needle include (a) small outside diameter, (b) sharp point, (c) resistance to buckling and breaking, and (d) secure suture attachment.

The lower stiffness of Ni-Ti superelastic alloys creates a lower buckling resistance when compared to a stainless steel needle of the same outside diameter. Otherwise, the Ni-Ti needles are expected to perform similarly to stainless needles without taking a significant permanent set when used with curved cannulas. Secure suture attachment has been accomplished without sacrificing breakage resistance with several slight configuration changes.

Side by side testing of stainless steel and Ni-Ti needles of comparable size has been undertaken. The test was comprised of pushing both needle types, initially straight, through a 35mm radius curved tip cannula. The Ni-Ti had a final shape was substantially straight, while the stainless steel had a final shape radius 30 cm.

Bench test results suggest Ni-Ti needles will be a safe device. Clinical tests will be undertaken to establish whether it creates a more effective procedure.

3. Summary

Two different medical applications for superelastic Ni-Ti properties have been presented. These device suggest the high shape recovery resulting from the Ni-Ti alloy could be valuable as new less invasive procedures are developed. In some instances Ni-Ti will replace stainless steel in current procedures for more effective and safer procedures.

The most significant barriers to greater use of Ni-Ti alloys in medical devices include (a) lower stiffness than stainless steel, (b) more difficult joining and assembly of multiple component devices, and (c) higher alloy costs.

The benefits of Ni-Ti superelasticity for safer and more effective procedures will have to be balanced against its disadvantages. In some instances, these disadvantages can be overcome with slight design and processing changes.

4. References

1. DeHaven, K.: Peripheral meniscus repair - three to seven year results. Presented at the Third Congress of the International Society of the Knee, Gleneagles, Scotland, April, (1983).
2. Dotter, C.T., Judkins, M.P., Frische, L.H.: American Journal of Roentgenology, 9, 957 (1966).
3. Judkins, M.P., Kidd, H.J., Frische, L.H., Dotter, C.T.: Radiology, 88, 1127 (1967).

Some Notes on Mechanical Damping of Shape Memory Alloys

Manfred Wuttig
Department of Chemical Engineering
University of Maryland
College Park, MD 20742

In this note the damping properties of shape memory alloys, i.e. the characteristics of the imaginary parts of the mechanical response coefficients will be highlighted. The generally frequency dependent linear complex compliance of a material is given by:

$$S(f) = S_r(f) + iS_i(f). \quad (1)$$

In such a material the strain lags behind the stress by an angle $\tan^{-1}(S_i/S_r)$ and a relative amount of mechanical energy:

$$\delta W/W = 2\pi S_i/S_r \quad (2)$$

is absorbed per cycle as the material vibrates. The fraction given by Eq. (2) is called the damping or internal friction of the material. A discussion of the damping characteristics of shape memory alloys must thus consider the various mechanisms by which energy can be absorbed internally.

Micromechanically, the time delay of the strain causing the absorption of energy is caused by the motion of structural imperfections in the material. Prominent examples of imperfections are point defects such as interstitials, dislocations as well as structural boundaries between grains and twins. Fully developed or incipient parent-product and twin boundaries are primarily responsible for the exceptionally large ($\delta W/W$ as large as 10%) damping of shape memory alloys. An appreciation of the controlling features of internal friction caused by the motion of boundaries will thus help in selecting shape memory alloys for damping applications.

Structural boundaries are positioned at local minima determined by the superposition of external and internal stresses. They move whenever these positions change upon the applications of an external stress. The rate of motion is usually controlled by a thermally activated process. These facts have two important implications.

First, if the externally applied stresses are small in comparison to internal stresses, the boundaries will move by insignificant distances causing technologically insignificant damping. Structurally superior alloys are hardened by a variety of means, all of which result in an increase of the level of the internal stresses. Therefore, it is unrealistic to expect that a shape memory alloy with improved yield characteristics features simultaneously high damping. Compromises must be made if both the static and dynamic properties of the alloy are important in a design.

Second, the damping is temperature and frequency dependent and peaks in the vicinity of the martensitic transformation temperature. This characteristic requires that for any application the transformation and use temperatures as well as frequencies are matched.

The special damping mechanism by which the amount of product phase is changed by the application of stress is unique to shape memory alloys. This type of damping occurs in a wider temperature range than the damping mentioned above. However, at temperatures above the martensite start temperature this damping mechanism is only accessible at large ($>10^{-3}$) strains. Due to the irreversible nature of the first order martensitic transformation this type of damping fatigues significantly. In addition, the load in the structural member for which a high damping is desired will alter the damping characteristics and must be considered as an important design parameter. In summary, shape memory alloys are also high damping alloys reaching damping capacities of the order of ten percent. Their usefulness is limited to the temperature region around and below the transformation which, however, is the temperature range in which the shape memory operates. It is also important to keep in mind the load and frequency limitations when designing for vibration free structures with shape memory alloys. The compromises which have to be made are complex.

Index

In subheadings metals are referred to by their chemical symbol. Other abbreviations used are as follows: SM — shape memory; SMA — shape memory alloy

Accelerometers, SMA ring clamping, 166–168
Accommodation in transformation, 5–6
in SM effect, 11–12
Acoustic properties, and transformation, 106, 110
Actuators, 181–194, 245–255
see also Electrical actuators; Nickel-titanium actuators; Thermal actuators; Wax actuators
see also by particular application
calibration techniques, 248–250
compression springs, 247–248
cost, 245
design, 184–187
 biasing springs, 186–187
 chart, 251
 examples, 251–254
 and stress rate, 185–186
 and work output, 184–185
efficiency, 187–192
 prediction for ideal alloy, 189–190,
 191, 192
 thermal cycle, 187–189
 thermal relationships in, 189
 variations from ideal, 190, 192
hysteresis, 224, 225, 226, 227, 254–255
interdependent variables, 255
two-way SM properties, 245–247
types, 192–193
work production, 181, 182–184
 design and, 184–185
 resetting stress, 183–184
 stress/strain/temperature
 relationships, 182–183, 221–224

Aging
of Cu-based alloys, 78–83
 direct quenching, 79, 81
 isothermal, 81
 and Ni content, 81–82
 post-quench aging, 78–79
 and pseudoelasticity, 82–83
and martensite stabilization, 99
of Ni-Ti, for superelasticity, 386–390
for stabilization of superelasticity,
 404–405
for suppression of thermal cycling
 effects, 397–399

Air conditioner control, conventional, 315
versus SMA system, 328
Air conditioner control, SMA for, 315–329
advantages, versus conventional,
 328
design of mechanism, 322, 328–325
operation of unit, 326–328
reliability using R-phase transition,
 326, 327
two-way motion phase transitions,
 315, 317–322
fixed load tests, 317–319
and heat treatment temperature,
 322, 323
SM motions, 321–322
stress–temperature phase
 diagrams, 319–321
Aircraft, SMA hydraulic couplings, 137,
 137–141
see also Cars, SM actuators in
'All-around' SM, 205
 for thermal actuators, 209, 210
Aluminium, *see* Copper-based alloys
Amnesia in Ni-Ti alloy
 in actuators, and work output, 184–185
 and annealing, 30, 33
 strain, 117, 118
Aneurysm clips, two-way SM for, 200
Ankylosis, SM spacers for, 429–430
Annealing of Ni-Ti alloys
 and amnesia, 30, 33
 and cold working, 28–30, 31, 32
 of Ni-Ti-Cu alloys, and
 microstructure, 50
 and R-Phase transition, 41
 and stability, 23–24
for superelasticity, 382, 385–386,
 405–408
for thermal cycling effect
 suppression, 397, 399
Anti-scald safety value, 305–314
concept, 305
dynamic analysis, 306
dynamic force balance, 309
environment evaluation, 305
numerical results, 310–312
operation, 312–313
performance, 312, 313–314
pressure differential, 307–309
quasi-steady state analysis, 306–307, 308

- Archwires in orthodontics, 456–459
 mechanical characterization, 463–466
 two-way SM effect, 462–463
 variable load/constant force strategy, 459–460
- Arming devices, SM in, 356–361
 fire safety, 357
 prototype, 356–357
 wire testing, 357–361
- Arthrodesis
 foot, 435, 437, 438
 spinal, SM spacers for, 429–430, 411–443
 animal experiments, 432–435, 436
- Arthroscopic suture needles, 483, 485–486
- Athermal transformations, 3
- Atomic ordering in alloys, 6–7
- Automobiles, *see* Cars, SM actuators in
- Backshell braid fasteners, *see* Braid termination, SMA fasteners for Bain strain, 4–5
- Bearings, preloading, SMA rings for, 165–166
- Belleville washers, 213
- Beta-phase overheating, 98–99
- Betaflex connector, 151, 157
- Betalloy, 131–132
- Betatizing
 of Cu-based alloys
 and grain-refining elements, 73
 and transformation temperatures, 73–74
 and fatigue, 279
- Bias springs in actuators, 241–243
 design, 186–187
- Bimetals, thermostatic, 211–214
 in circuit breakers, 331, 332–333
 examples, 213, 214
 materials, 211, 212
 properties, 211–212
 versus SM actuators, 216–217
 thermal-deflection characteristics, 212–213
- Biocompatibility of Ni-Ti
 breast lesion hook, 480
 of dental implants, 470–472
 in orthodontics, 462
 of orthopedic implants, 431, 438–440
- Biomedical applications, *see* Medical applications of Ni-Ti
- Blount orthopedic staples, 428, 435
- Bones, *see* Orthopedic implants
- Boron, in grain refining, 73
- Braid termination, SMA fasteners for, 133, 134, 135, 164–165
- Brazing of Cu alloys, 86
- Breast lesion localization
see also Guidewires, medical conventional procedure, 477–478
- needles for
 conventional, 478
 SM 478–482
- Cables, insulated, 330, 331
- Cadmium-gold transformation measurement, 107
- CANTIM -125, *see* Copper-aluminium-nickel-magnese alloy
- Cars, SM actuators in, 283–294
see also Aircraft, SMA hydraulic couplings
 design principles, 283–284
 electrical actuators, 290–294
 applications, 290–294
 temperature ranges, 290, 291
- market penetration, 283
- thermal actuators, 284–290
 applications, 284–285, 286–290
 stability, 286
 temperature ranges, 284–286
- Circuit breakers, 330–337
see also Electrical actuators
 requirements, 331–332
- SM actuators, 333
 performance as assembled, 335
 performance testing, 333, 335, 336
 potential, 337
- thermobimetal actuators, 331, 332–333
 tripping events, 330–331
- Clausius–Clapeyron equation, 15
- Coatings, of Cu-based alloys, and SM effect, 86
- Cold working
 of Cu-Zn-Al alloys, 72
 of Ni-Ti alloys
 and annealing, 28–30, 31, 32
 and R-phase transition, 41
 and superelasticity, 382–386, 414–419
- Color change in transformation monitoring, 107
- Connectors, *see* Couplings; Electrical connectors; Fastener rings
- Constrained recovery, 115, 119–129
see also Couplings; Fastener rings, SMA
 applications, 128–129
 joints, 124–128
 compliance, 124–126
 self-inspecting, 126, 127
 stress decay, 126–128
- measurement methods, 115–116
 quantification of, 119–120
 recovery stress
 and austenite stress–strain curve, 124

- Constrained recovery (*cont.*)
 recovery stress (*cont.*)
 and contact strain, 120–121
 and deformation strain, 122
 and substrate compliance, 122–124
- Contact strain, 120
- Contact temperature, 120
- Control systems, SM actuators in, 338, 341, 342
 see also Proportional control
- Copper addition to Ni-Ti, and superelasticity, 329
- Copper-aluminium-nickel alloys
 see also Nickel-titanium-copper alloys
 for circuit breaker actuators, 333
 performance, 333, 335
- Copper-aluminium-nickel-manganese alloy, 89–95
 development, 89–90
 recovery force, 91–93, 94
 SM effect, 90–91
 thermal stability, 91
- Copper-based alloys, 69–88
 see also Fatigue of Cu-based alloys;
 Fire protection
 alloy elements
 and equilibrium phases, 71–73
 grain-refining, 73
 and lattice structure, 70
 SM properties and, 69–73
 and transformation, 70–71
 classification, 68, 69
 cost-effectiveness, 86
 environmental effects, 85–86
 grain refining
 alloy elements, 73
 powder metallurgy for, 85
 rapid solidification for, 85
 joining, 86
 mechanical properties, 83–85
 elasticity/pseudoelasticity, 85
 stress-strain relationships, 83–85
 versus Ni-Ti alloys, 69
 pseudoelasticity in, 85
 and aging effects, 82–83
 source binary alloys, 69
 space system application, 170
 and transformation
 color change in, 107
 thermal effects on 73–83
 thermoelasticity, 97
 wide-hysteresis, for heat-to-shrink applications, 131–133
 comparison with Ni-Ti alloys, 136
 composition, 131
 preconditioning, 132–133
- Copper-zinc alloy, superelastic behavior, 16, 17
- Copper-zinc-aluminium alloys, *see* Copper-based alloys
- Corrosion of Ni-Ti alloys, 34
 see also Stress-corrosion cracking, of Cu-based alloys of medical implants, 441, 445–446
- Couplings, 137–148
 see also Electrical connectors, SMA; Fastener rings; Joints, SMA
 constrained recovery for, 128–129
 corrosion resistance of Ni-Ti in, 34
 for industrial pipes, 143–146
- Permacouple, 133
 predeformed, phase stabilization in, 99
- in space systems, 171–172
- success, 148
- testing, 141–143
- tolerances, 147–148
- for tubes, 137–141
 access advantages, 140
 installations, 137, 138–139
 manufacture, 137
 modelling, 140–141, 142
 tools, 137–138
- two-way SM devices, 197, 198
- wire, *in situ* wound, 173, 176
- Creep in constrained recovery joints, 126
- Cryocon contact, 151
- Cryofit coupling, 140, 141
- Cryogenic alloys for joints, 130
- Cryotact
 contact, 154
 DIP connectors, 149, 150, 155
- Crystallographic measurements, and SM effect, 110
- Crystallography of martensite
 see also Transformations,
 microscopic perspective symmetry, 6
 transformation, 4–7, 8
- Cycling. *see* Stress cycling of Ni-Ti alloys; Thermal cycling of Ni-Ti alloys
- Damping of SMAs, 488–489
 in space structures
 out-of-phase vibration, 174–175
 stiffness control, 175, 177
- Definitions
 Constant strain, 120
 fatigue of Ni-Ti alloys, 256
 martensite/austenite, 4
 proportional control, 362
 pseudoelasticity, 370
 superelasticity, 370
 transformations, 3
- Dental SM implants, 470–476
 see also Orthodontics, Ni-Ti alloys in;
 Orthopedic implants
 advantages, 476
 biocompatibility, 470–472
 clinical results, 475–476
 design, 472–474

- Detwinning, 6
- Dezincification, of Cu-based alloys, 85
- Differential Scanning Calorimetry (DSC), 107, 109
- Diffusional transformations, 3
- Elastic energy in hysteresis, 97
see also Ferroelasticity;
 Pseudoelasticity; Superelasticity;
 Thermoelasticity of Cu alloys
- Electrical actuators, 192–193, 219–233
see also Circuit breakers,
 Thermoelectric power
- alloys
 processing, 221
 selection, 219–221
- ambient temperature effects, 224–226, 227–230
- in cars, 290–294
 applications, 290–294
 temperature ranges, 290, 291
- design tradeoffs, 219
- energy evaluation, 230, 232
- hysteresis evaluation, 224, 225, 226, 227
- stress–strain evaluation 221–224
- temporal evaluation, 226, 230, 231, 232
- Electrical connectors, SMA, 149–157
see also Resistance/resistivity
- constrained recovery for, 129
- design, 150–155
 alloy choice, 150–151
 compliance, 153–155
 constraints, 157
- Cryocon contact, 151
- operating path, 152–153
- preliminary questions, 150
- and SMA stress–strain diagram, 151–152
- stroke, 153–154
- performance, 149–150
 corrosion resistance of Ni-Ti, 34
 shock vibration, 149, 150
 zero insertion/high retention forces, 149
- systems in use, 155–157
- Expansion, thermal, 107, 110
- Explosives, *see* Arming devices, SM in
- Eyeglass frames, SMAs for, 420–425
 advantages, 420
- business lessons from development, 424–425
- development for manufacture, 422–423
- problems, 420–422
- Fastener rings, SM, 158–169
see also Couplings; Joints, SMA
- applications, 158
 bearing preload, 165–166
- Fastener rings, SM (*cont.*)
 applications (*cont.*)
 braid termination, 133, 134, 135, 164–165
 engineering, 197, 198
 fire protection, 300–301
 hermetic seal, 168–169
 medical, 199
 missile retaining ring, 166, 167
 radial assembly, 166–168
 constrained recovery for, 128–129
 corrosion resistance of Ni-Ti in, 34
 design, 162–164
 stress–strain analysis, 162–163
 testing, 163–164
- elastic interference, 160, 161
- installation temperature, 160, 161
- operating temperature range, 162
- operator insensitive installation, 160, 162
- radial pressure, 160
- stress generation, 159–160
- tolerances, 159
- two-way SM devices, 197, 198, 199
- Fatigue of Cu-based alloys, 267–280
 fatigue crack growth analysis, 272, 273
 fatigue life, 269, 271, 272, 273
 and grain boundaries, 269
 loading modes, 267
 microstructure changes, 274, 277, 278, 279
 and pseudoelasticity, 274, 276
 and thermal cycling, low temperature, 269, 270
- thermo-mechanical behavior, 267–269
 and transformation behavior, 272, 274, 275, 279
- Fatigue of Ni-Ti alloys, 256–266, 409–412
see also Stress cycling of Ni-Ti alloys;
 Thermal cycling of Ni-Ti alloys
- definitions, 256
- deterioration, 257–259
- fatigue life, 259, 409–410
- improving, 410–412
- and low-temperature phase structure, 259
- in use
 breast lesion hook, 480, 482
 coil springs, 259–266
 robotic actuators, 347, 348
- Feedback in control, 338, 341, 364, 365
 robotic arms, 107
- Ferroelasticity, 373
- Fire protection, 295–302
see also Anti-scald safety valve; Fire
 safing of arming device; Hot water
 control systems
- actuation, 302
 Proteus link, 300–301

- Fire protection (*cont.*)
 detection, 295–298
 resistivity changes, 295–296
 SM effect, 296–298
 detection with actuation, 298–301
 Proteus Gas Valve, 298–300
 Proteus link, 300–301
 SM action, and stabilization, 99
- Fire safing or arming device, 357
 alloy wire testing for, 358
- Flow proportional control valve, 364, 365, 366
 see also Anti-scald safety valve; Hot water control systems
- Fog-lamps for cars, louver actuator, 290, 292
- Foot arthrodesis, SM staples for, 435, 437–438
- Free recovery, 116–118
- Fuzes, *see* Arming devices, SM in
- Galvanic series, Ni-Ti alloys in, 34
- Grain-refining of Cu-based alloys, 73
- Guidewires, medical, 483–485
 see also Breast lesion localization
- Habit planes, 7
- Hall coefficient, 110
- Hardness measurement during transformation, 107
- Heart muscles, artificial, two-way SM for, 199
- Heat engines
 using one-way SM, 193
 using two-way SM, 199
- Heat-to-shrink alloys, *see* Wide hysteresis alloys
- Hermetic sealing with SM rings, 168–169
- Hip prosthesis, SMA, 449
- Homer Mammalok, 478–482
 biocompatibility, 480
 concept, 478–479
 design requirements, 479
 geometry, 480
 problems
 fatigue, 480, 482
 of full-scale production, 482
- Hooke's law, 234
- Hot water control systems
 see also Anti-scald safety valve
 design options, 314
 traditional, 303–305
- Hysteresis in SMAs, 97–98
 see also Wide hysteresis alloys
 for actuators, 224, 225, 226, 227, 254–255
- Ni-Ti alloys, 133, 135
 and Cu addition, 51–53
 versus Cu-based alloys, 136
 and fatigue, 261–266
- Hysteresis in SMAs (*cont.*)
 proportional control and, 364
 for springs, 235, 236
 temperature, 9, 10
 in R-phase transition, 36, 37
- Implants, *see* Orthopedic implants
- Iron addition to Ni-Ti, and superelasticity, 391, 392
- Iron-manganese-silicon alloys, space system application, 170
- Jaw SM implants
 see also Dental implants; Orthodontics
 pins, 448
 plates, 427–428
 animal experiments, 431–432, 433
- Joints (bone), *see* arthrodesis, Arthroscopic suture needles
- Joints, SMA, 124–128
 see also Couplings; Fastener rings; Welding, of Cu-based alloys
- compliance, 124–126
- cryogenic alloys, 130
- self-inspecting, 126, 127
- stress decay, 126–128, 130
- Knee osteotomy, SM staples for, 438
- Latent heat of transformation, 107
 superelastic, 379
- Lattice structures
 see also Crystallography of martensite; Microstructure; transformations, microscopic perspective
 of Cu-based alloy elements, 70
 deformation in transformations, 4
- Lattice invariant shear, 5
- Lüder's bands, in superelasticity, 381
- Magnetic changes, 110
- Mammography, lesion localization
 conventional procedure, 477–478
 needles for
 conventional, 478
 SM, 478–482
- Manganese, in Cu-based alloys, 73
- Manganese nodules, *see* Submarine resources, robotics in development
- Medical applications of Ni-Ti, 445–451
 see also Dental implants; Orthodontics, Ni-Ti alloys in; Orthopedic implants
- arthroscopic suture needles, 483, 485–486
- breast lesion localization, 478–482
 guidewires, 483–485
 heart muscles, artificial, 199

- Medical applications of Ni-Ti (*cont.*)**
- pre-clinical trials
 - in vitro, 445–446
 - in vivo, 446–447
 - sterilization, tubal, 447–448
 - two-way SM, 199–200
- Medullary nails, 429**
- Melt extraction/spinning, for Cu-based alloy production, 85**
- Microstructure of alloys**
- see also* Crystallography of
 - martensite; Transformations, microscopic perspective
 - and Cu-based alloy fatigue, 274, 277, 278, 279
 - of martensite plates, 7–9
 - of Ni-Ti alloys, and thermal cycling, 396–397
 - of Ni-Ti-Cu alloys
 - annealing and, 50
 - and Cu addition, 46–51
 - of two-way SM effect, 200–202
- Missile SMA retaining ring, 166, 167**
- Moduli in pseudoelastic behavior, 371**
- see also* Young's modulus
- Nickel, in Cu-based alloys, and aging, 81–82**
- Nickel-titanium actuators, 192–193, 234–244**
- see also* Cars, SM actuators in; Electrical actuators; Thermal actuators
 - bias spring designs, 241–243
 - corrosion resistance, 34
 - cyclic stability, 34
 - design concepts, 237–239
 - load-deflection behavior, 238–239
 - spring deflections, 237
 - design example, 240–241
 - equations, 244
 - hysteresis behavior, 235, 236
 - spring behavior, 234–235
 - terms, 243–244
 - transformation temperatures, 28
- Nickel-titanium alloys, 21–35**
- see also* Fatigue of Ni-Ti alloys; Pseudoelasticity; R-phase transitions; Stress cycling of Ni-Ti alloys; Thermal cycling of Ni-Ti alloys
 - see also by particular application*
 - and alloying elements, 34–35
 - for constrained recovery joints, 125–126
 - versus Cu-based alloys, 69, 86
 - corrosion behavior, 34
 - discovery of properties, 21
 - hysteresis expansion, 133–135
 - versus Cu-based alloys, 136
 - mechanical properties, 24–28
 - modulus, 28
- Nickel-titanium alloys (*cont.*)**
- mechanical properties (*cont.*)
 - stress-strain curves, 24–26, 151–152
 - transformation temperatures, 27–28
 - yield stress, 26–27
 - metallurgy of, 21–24
 - stability, and heat treatment, 23–24
 - stoichiometry, 21
 - transformation temperature and composition, 22–23
 - superelasticity in, linear, 414–419
 - and thermo-mechanical processing, 28–34
 - cold working/annealing, 28–30, 31, 32
 - cyclic behavior, 30, 33–34
 - transformation
 - measurement, 106–107, 108
 - temperatures, 27–28
 - thermoelasticity, 97
- Nickel-titanium-copper alloys, 46–57**
- microstructure, and Cu addition, 46–51
 - lattice parameters, 46, 47
 - oxide phase, 51
 - and processing conditions, 50
 - transformation and temperature, 50
 - twinning, 47, 48, 49
 - SM properties, and Cu addition, 51–56
 - cyclic behavior, 54–55
 - hysteresis, 51–53
 - and M_s temperature, 51
 - two-step property changes, 55–56
 - and yield strength, 53–54
- Niobium, in Ni-Ti for hysteresis expansion, 133–135**
- Orthodontics, Ni-Ti alloys in, 452–469**
- see also* Dental implants
 - archwires, 456–459
 - biocompatibility, 462
 - characterization/standardization of products, 463–466
 - coil springs, 460–461
 - forces, wire materials compared, 452–456
 - two-way SM effect, 462–463
 - variable load/constant force strategy, 459–460
- Orthopedic implants, 426–444**
- see also* Arthroscopic suture needles; Dental implants
 - animal experiments, 431–435
 - with jaw plates, 431–432, 433
 - with spacers, 432–435, 436
 - with staples, 432, 433
 - tissue compatibility, 431
 - hip prosthesis, 449
 - intervertebral joints, 449–450
 - jaw plates, 427–428

- Orthopedic implants (cont.)**
- medullary nails, 429
 - osteosynthesis plates, 427–428
 - pins in oral-maxillofacial surgery, 448
 - scoliosis correction rods, 450–451
 - spacers, 429–430
 - clinical use, 441–443
 - staples, 428–429, 448
 - condition after use, 441
 - for foot arthrodesis, 435, 437–438
 - histological examination after implantation, 438–440
 - for knee osteotomy, 438
 - Osteosynthesis plates, 427–428
 - Oxides, of Ni-Ti-Cu alloys, 51
- Palladium, *see* Titanium-palladium-nickel/Ti-Pt-Ni alloys**
- Passive film, on Ni-Ti alloys, 34**
- Permacouple, 133**
- Pin Grid Array Package (PGAP) connectors, 149, 150, 155, 156**
- Pipe SM couplings, 143–146**
- Plate structure of martensite, 7–9**
- Platinum, *see* Titanium-palladium-nickel/Ti-Pt-Ni alloys**
- Positive Temperature Coefficient (PTC) resistor, 302**
- Powder processing of Cu based alloys, 85**
 - for circuit breakers, 333
- Preconditioning in hysteresis expansion, 132–133**
- Proportional control, 362–366**
 - advantages, 362
 - definition, 362
 - design, 364, 365, 366
 - and hysteresis, 364
 - position feedback, 364, 365
 - recovery rate modification, 362–364
- Proteus Gas Valve, 298–299**
 - reaction time, 300
- Proteus link, 300–301**
 - PTC resistor activated, 302
- Pseudoelasticity, 369–393**
 - see also* Superelasticity
 - in Cu-based alloys, 85
 - and aging effects, 82–83
 - and fatigue, 274, 276
 - definitions, 370
 - mechanisms, 370–375
 - transformational superelasticity, 373–375
 - twinning, 370–373
 - non-linearity examples, 369–370
 - Pseudotwinning, 370–373
- Quenching**
 - and Cu-based alloy transformation, 79, 81
 - rate, 74–78
 - step-quenching, 97
 - vacancies, and stabilization, 99
- R-phase transitions, 36–57**
 - in air conditioner actuation, 319–322, 323
 - in mechanism design, 324, 325
 - reliability, 326, 327
 - conditions for, 39–41
 - and constant load, 41–43, 44
 - lattice changes in, 37–38
 - resistance/temperature curves, 36–37
 - SM effect in, 38, 39
 - under cyclic conditions, 43, 44
 - studies on, 43–44
 - superelasticity in, 38, 39
 - in Ti-Pd-Ni/Ti-Pt-Ni alloys, 59, 60
 - for thermal actuators, 209
- Recovery, free, 116–118**
 - see also* Constrained recovery
- Resistance/resistivity, electrical**
 - in fire detection, 295–296
 - and thermal cycling, 394, 395
 - Ti-Pd-Ni/Ti-Pt-Ni transformation behavior and, 60, 61
 - in transformation monitoring, 107, 108
- Resistors, Positive Temperature Coefficient (PTC), 302**
- Ring (acoustic), *see* Acoustic properties, and transformation**
- Ring fasteners, *see* Fastener rings, SM**
- Robotic arms, feedback control, 107**
- Robotics, SM actuators in, 338–355**
 - development projects, 351–354
 - technical problems, 352
 - principles, 338–343
 - biased actuators, 338, 339, 340
 - control systems, 338, 341, 342
 - differential actuators, 338, 339
 - response time, 343
 - for submarine resource development, 344–350, 351
 - design concept, 344
 - 'knee joint', 344, 345, 346–347
 - movement of robot, 347, 349, 350
- Rockwell hardness measurement in transformation, 107**
- Safety valves, *see* Anti-scald safety valve; Valves**
- Scalding, 303**
 - see also* Anti-scald safety valve; Hot water control systems
- Scoliosis correction, SMA in**
 - rods, 450–451
 - spacers, 432
- Shear, lattice-invariant, 4**
- Shower Gard Anti Scald Valve, *see* Anti-scald safety valve**
- Slip in accommodation, 5, 6**
- Soldering of Cu-based alloys, 86**
- Space systems, SMA in, 170–177**
 - commercial alloys, 170
 - protective shrouds, 173, 176
 - release mechanisms, 177

- Space systems, SMA in (*cont.*)
 requirement for ease of assembly, 170–171
 seals for rockets, 177
 space stations, 170, 173
 vibration damping, 174–175, 177
 temperature limits, 172
 tube couplings, 171, 172
 wire, for light couplings, 173, 176
- Spacers, orthopedic, 429–430, 432, 435, 436, 441–443
- Spectacles, *see* Eyeglass frames, SMAs for
- Spinal surgery, SMAs in
 intervertebral joints, 449–450
 scoliosis correction rods, 450–451
 spacers, 429–430
 animal experiments, 432–435, 436
 arthrodesis, 441–443
- Springs, SM, 234–235
see also Actuators; Ni-Ti actuators
 for air conditioner direction control, 322, 324–325
 bias, in actuators, 241–243
 design, 186, 187
 Ni-Ti, fatigue measurement, 259–266
 in orthodontics, 460–461
- Stability of SMAs, 96–105
 behavior differences, design/actual, 102–104
 external parameters, 96
 and hysteresis, 97–98
 martensitic stabilization, 99–101
 in wide-hysteresis alloys, 132
 overheating in beta-phase, 98–99
 reference treatment and, 97
 stabilization tendency, 96
 two-way SM effect degradation, 102
- Staples, orthopedic, 428–429, 448
 animal experiments, 432, 433
 for foot arthrodesis, 435, 437–438
 histological examination, 438–440
 for knee osteotomy, 438
- Sterilization (medical procedure), 447–448
- Strain-temperature of actuator work, 182
- Stress cycling of Ni-Ti alloys, 401–408
see also thermal cycling of Ni-Ti alloys
 stabilization of superelasticity, 403–408
 and stress/strain relationships, 401–403
 and superelasticity, 394
- Stress rate, 16
 in actuator design, 185–186
 constant strain definition, 120
- Stress-corrosion cracking, of Cu-based alloys, 86
- Stress-induced martensite, 14–16
see also Superelasticity
- Stress–strain relationships
 of actuators, 182–183, 221–224
 in constrained recovery, 119–120, 124
 of Cu-based alloys, 83–85
 of fastener rings, 162–163
- Stress–strain relationships (*cont.*)
 in free recovery, 116–118
 for Ni-Ti alloys, 24–26
 couplings, 151–152
 SM effect/superelasticity in, 38, 39, 40
 in stress cycling, 401–403
 of superelasticity, 380–382
- Structure of alloys, *see* Crystallography of martensite; Microstructure of alloys; Transformations, microscopic perspective
- Submarine resources, robotics in development, 344–350, 351
 design concept, 344, 345
 'knee joint', 344, 345, 346–347
 robot movement, 347, 349, 350
- Superelasticity, 16–18, 20, 369, 370, 375–382, 401
see also Stress-induced martensite
 definition, 370
 energy storage, 380
 and fatigue life, 411–412
 linear, and cold-working, 414–419
 in medical applications, guidewires/needles, 483–486
see also Orthodontics, Ni-Ti alloys in
 in Ni-Ti-Cu alloys, 51–53
 physical descriptors, 375–376
 processing for, 382–392
 aging, 386–390
 cold work and recovery anneal, 382–386
 and composition, 390–392
 property modifications, in 382
 and R-phase transition, 38, 39
 strain dependencies, 376–377, 378
 and stress cycling, 394, 403–408
 stress–strain analysis, 380–382
 temperature dependencies, 377–380
 transformational, 373–375
- Surface coatings, of Cu-based alloys, 86
- Surgery, *see* Medical applications of Ni-Ti
- Swaged fittings, versus SM, 138
- Symmetry, crystallographic, 6, 7
- Thermal actuators, 193
see also Electrical actuators; Ni-Ti actuators
 in cars, 284–290
 applications, 284–285
 stability, 286
 temperature ranges, 284–286
 hysteresis, 36
 versus conventional types, 216–217
- SM effect, 207–211
 alloys, 207, 208
 one-way, 208
 properties of actuators, 209, 211
 R-phase, 209
 two-way, 208–209, 210

