THE SHAPE MEMORY EFFECT • Phenomenon, Alloys, Applications

Dieter Stöckel NDC, Nitinol Devices & Components, Inc., Fremont, CA

Introduction

Certain metallic materials will, after an apparent plastic deformation, return to their original shape when heated. The same materials, in a certain temperature range, can be strained up to approx. 10% and still will return to their original shape when unloaded. These unusual effects are called thermal shape memory and superelasticity (elastic shape memory) respectively [1]. Both effects depend on the occurrence of a specific type of phase change known as thermoelastic martensitic transformation. Shape memory and superelastic alloys respond to temperature changes and mechanical stresses in nonconventional and highly amazing ways. They are, therefore, something called "smart materials". The shape memory effect can be used to generate motion and/or force, while superelasticity allows energy storage. Both effects have fascinated scientists and engineers for almost three decades, drawing them to conferences and seminars in great numbers. However, very few developments made it to the market, and can be considered economic successes. Recent successes some mainly from medical applications utilizing the superelasticity and biocompatibility of Ni-Ti alloys.

Shape Memory Effect

"Shape Memory" describes the effect of restoring the original shape of a plastically deformed sample by heating it. This phenomenon results from a crystalline phase change known as "thermoelastic martensitic transformation". At temperatures below the transformation temperature, shape memory alloys are martensitic. In this condition, their microstructure is characterized by "self-accommodating twins". The martensitic is soft and can be deformed quite by de-twinning. Heating above the transformation temperature recovers the original ahape and converts the material to its high strength, austenitic, condition (Fig.1).



Fig. 1: Schematic representation of the shape memory effect and superelasticity

The transformation from austenite to martensite and the reverse transformation from martensite to austenite do not take place at the same temperature. A plot of volume fraction of martensite, or more practically, the length of a wire loaded with a constant weight, as a function of temperature provides a curve of the following temperatures: austenit start temperature (Ass), austenit finish temperature (Af), martensite start temperature (Ms) and martensite finish temperature (Mf).

If a stress is applied to a shape memory alloy in the temperature range between Af and a maximum temperature Md, martensite can be stress-induced. Less energy is needed to stress-induce and deform martensite than deform the austenite by conventional

mechanisms. Up to 10% strain can be accommodated by this process (single crystals of specific alloys can show as much as 25% pseudoelastic strain in certain directions). Ass austenite is the thermodynamically stable phase at this temperature under no-load conditions, the material springs back into its original shape when the stress is no longer applied. This extraordinary elasticity is also called pseudoelasticity or transformational superelasticity.

It becomes increasingly difficult to stress-induce martensite at increasing temperatures above Af. Eventually, it is easier to deform the material by conventional mechanisms than by inducing and defomring martensite. The temperature at which martensite is no longer stress-induced is called Md. Above Md, the alloys are deformed like ordinary materials. Thus, superelasticity is only observed over a narrow temperature range.



Fig. 2 (left): Schematic representation of the hysteresis loop Fig. 3 (right): Stress/strain curves at different temperatures

The design of shape memory components, e.g. fasteners or actuators, is based on the distinctly different stress/strain curves of the martensite and austenite, and their temperature dependence. Figure 3 shows tensile curves of a Ni-Ti alloy at various temperatures. While the austenitic curve (T>Md) looks like that of a "normal" material, the martensitic one (T<Mf) is quite unusual. On exceeding a first yield point, several percent strain can be accumulated with only little stress increase. After that, stress increases rapidly with further deformation. The deformation in the "plateau region" can be recovered thermally. Deformation exceeding a second yield point cannot be recovered. The material is then plastically deformed in a conventional way. At temperatures T>Af, again, a plateau is observed upon loading. In this case, it is caused by stress induced martensite. Upon unloading, the material transforms back into austenite at a lower stress (unloading plateau). With increasing temperature, both loading and unloading plateau stress increase linearly [2].

Shape Memory Alloys

The shape memory effect as the result of a martensitic transformation has been known since the mid 1950's, when the effect was discovered in copper base alloys. In the early sixties, researchers at the Naval Ordnance Laboratory found the shape memory effect in Ni-Ti alloys (Nitinol - Ni-Ti Naval Ordnance Lab). Today, these alloys are the most widely used shape memory and superelastic alloys, combining the most pronounced shape memory effect and superelasticity, corrosion resistance and biocompatibility, and superior engineering properties. Copper based alloys like Cu-Zn-Al and Cu-Al-Ni are commercially available, too. These alloys are less stable and more brittle than Ni-Ti, and therefore, although less expensive, have found only limited acceptance. In recent years, iron based shape memory alloys have been widely advertised. However, with their limited shape memory strain, lack of ductility and other essential properties, these alloys will have to prove themselves as viable engineering materials.

The transformation temperatures of shape memory alloys can be adjusted through changes in composition. Ni-Ti as well as Cu-Zn-Al alloys show transformation temperatures between - 100°C and +100°C, Cu-Al-Ni alloys up to 200°C. Unfortunately, Cu-Al-Ni alloys are not stable in cyclic applications. some ternary Ni-Ti-Pd [3], Ni-Ti-Hf and Ni-Ti-Zr [4] alloys also are reported to exhibit transformation temperatures over 200°C. Although not commercially available today, these alloy could eventually expand the applicability of the shape memory effect to much higher temperatures. In the following, only Ni-Ti alloys will be reviewed.

The hysteresis is an important characteristic of the heating and cooling behavior of shape memory alloys and products made from these alloys. Depending on the alloy used and/or its processing, the transformation temperature as well as the shape of the hysteresis loop can be altered in a wide range. Binary Ni-Ti alloys typically have transformation temperatures (Af) between 0°C and 100°C with a width of the hysteresis loop of 25°C to 40°C. Copper containing Ni-Ti alloys exhibiting a premartensitic transformation (commonly called R-phase). On the other hand, a very wide hysteresis of over 150°C can be realized in Niobium containing Ni-Ti alloys after a particular thermomechanical treatment. Although low transformation temperatures (Af << 0°C) can be reached with binary Ni-Ti alloys, those alloys tend to be brittle and difficult to process. For cryogenic uses, therefore, Fe-containing Ni-Ti alloys are commonly used.



Fig. 4 (left): Influence of processing on the shape of the hysteresis loop (schematic) Fig. 5 (right): Influence of applied stress on the transformation temperatures

The standard thermomechanical processing of Ni-Ti alloys generates a steep hysteresis loop (a greater shape change with a lesser change in temperature), which generally is desirable in applications where a certain function has to be performed upon reaching or exceeding a certain temperature. Special processing can yield a hysteresis loop with a more gradual slope, i.e. a small shape change with temperature. This behavior is preferred in applications where proportional control is required [5].

The shape of the hysteresis loop is not only alloy and processing dependent, but is also influenced by the application itself. If a wire (standard processing) works against a constant load, e.g. by lifting a certain weight, the transition from martensite to austenite or vice versa occurs in a very narrow temperature range (typically 5°C). However, if the wire works against a biasing spring, the transition is more gradual and depends on the rate of the spring.

Engineering Aspects

The shape memory effect can be used to generate motion and/or force, while superelasticity can store deformation energy. The function of the different events as shown the stress/strain perspective in Fig. 6 [6] can be explained in simple terms using the example of a straight tensile wire. The wire is fixed at one end. Stretching it at room temperature generates an

elongation after unloading. The wire remains in the stretched condition until it is heated above the transformation temperature of this particular alloy. It will then shrink to its original length As no load is applied, this called *free recovery*. Subsequent cooling below the transformation temperature does not cause a macroscopic shape change.

If, after stretching at room temperature, the wire is prevented from returning to its original length, i.e. if constrained to the extended length upon heating above the transformation temperature, it can generate a considerable force. This so-called *constrained recovery* is the basis of many successful applications [7].



Fig. 6: Shape memory events in the stress/strain perspective [6]

If the opposing force can be overcome by the shape memory wire, it will generate motion against a force, and thus do work. Upon heating, the wire will contract and lift a load, for instance. Upon cooling, the same load will stretch the now martensitic wire and reset the mechanism. This effect is called *two-way-effect with external reset force* [8].

Depending on the kind of biasing mechanism, different force/displacement characteristics can be obtained [9]. Figure 7, five commonly used scenarios are compared with regard to the force/displacement response. The level of the force in Fig. 7a obviously is given by the weight of the "dead load", while the slope of the force/displacement line in Fig. 7b represents the spring rate of the biasing steel spring. In Fig. 7c, two shape memory wires are working in opposing directions. When wire 1 is heated (e.g. by electrically heating), it contracts, moves an object, and simultaneously stretches wire 2.

The object can be moved in the opposite direction by heating wire 2 after cooling of wire 1. So called reverse biasing is shown in Figure 7d and e. The magnet causes the shape memory wire to generate a high static force, that drops sharply when the magnet is separated from its holding plate. A slower drop in force can be achieved by using a cam arrangement with a decreasing lever during actuation of the shape memory wire. Reverse biasing is beneficial when high cyclic stability is important.



Fig. 7: Biasing Mechanisms and their effect on force/displacement characteristics [9]

Under optimum conditions and no load the shape memory strain can be as high as 8%. However, for cyclic applications the usable strain is much less. The same applies for the stress; for a one-time actuation the austenitic yield strength may be used as maximum stress. Much lower values have to be expected for cyclic applications.

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 temperatures shape, even without external force [10]. However, it is smaller and its cyclic behavior is not as well understood as that of the one-way-effect. Because there is no special treatment necessary, the cyclic use of the one-way-effect with external reset force in many cases is the more economic solution.

The forth event is *superelasticity*. A wire is loaded at temperatures above Af, but below Md. After reaching the first yield pont, it can be elongated to approx. 8% strain with no significant stress increase. Upon unloading, the wire recovers its original length elasticity, although with a stress hysteresis.



Fig. 8 (left): Tensile behavior of a superelastic wire at different temperatures Fig. 9 (right): Comparison of the flexibility of a stainless steel and a superelastic wire

Applications of Shape Memory and Superelastic Alloys

In the following, applications will be categorized according to the function of the shape memory alloy itself, as suggested by Duerig and Melton [6]. The early product development history of Ni-Ti has been full of failures and disappointments [11]. This can be attributed to the lack of understanding of the effects and the unavailability of engineering data, unreliable melting techniques and plain over-expectation. One major disadvantage of shape memory is its spectacular showing. It shows off as if it could solve all the problems in the world (browsing through the patent literature February1990 reveals: vacuum cleaner, sleeping device, method of manufacturing shoes, racket gut, shape recoverable fabric, diapers, toy boat, necktie, oilcooler bypass valve, throttle mechanism, concrete processing method ...). Obviously, it doesn't. In the meantime, after many million \$ lost on attempts to build the perpetuum mobile and to compete with thermostatic bimetals and other alternatives, the technology finally has come of age. Engineers understand the benefits, but also the limitations of the material, fabrication methods are reliable, and prices are at an acceptable level. Most new volume applications are based on the superelastic effect, which doesn't require as tight a transformation temperature control as the shape memory effect, as used for actuators, for instance.

The first technical successes clearly were uses of the constrained recovery event for joining and fastening purposes [7]. In the late sixties and early seventies, Raychem Corp. pioneered the development of tube and pipe couplings for aircraft, marine and other applications. The concept is straightforward: a sleeve is machined with an I.D. that is approx. 3% smaller than the diameter of the tubing it is designed to join. It is then cooled to its martensitic state and

radially expanded eight percent, making it large enough to slip over two tube ends. When heated, the sleeve shrinks onto the tube ends and, while generating a high force, joins the tubes. Most couplings are made from cryogenic Ni-Ti-Fe alloys and have to be stored in liquid nitrogen after expansion. While this does not seem to pose a problem for aircraft manufacturers, it is a logistics issue for most commercial users. Therefore, wide-hysteresis Ni-Ti-Nb alloys have been developed, which can be stored and shipped at room temperature after expansion at low temperatures, and have to be heated to 150°C for installation [13]. These alloys remain in their high strength, austenitic state even after cooling to below -20°C.



Fig. 10: Coupling, machined and expanded (top), after free recovery (middle) and installed on a tube (bottom) [12]



Fig. 11: Cut-away view of a shape memory coupling installed on a stainless steel tube [12]

To join large diameter pipes, or to create high compressive stresses near weld joints of such pipes for fatigue improvement, prestrained Ni-Ti-Nb wire or ribbon can be wound around the pipe and then thermally recovered. This wire wrap technology was recently development by ABB [14] for nuclear applications. It has to be mentioned, however, that NiT-Ti cannot be used in the high temperature, high pressure lines of a PDR, because of severe hydrogen embrittlement.

Wide hysteresis alloys are also used in a variety of fastening applications. For example, rings may be used to [15]:

- terminate electromagnetic shielding braid to connectors
- terminate heat shielding braid to oxygen sensors
- fix the location of bearings or gears at any point on a shaft, if desired, locking in a controlled axial preload force
- assemble clusters of radially disposed elements by compressing them with controlled uniform radial pressure
- provide very high retention forces and low contact resistance in high amperage connectors.



Fig. 12: Electromagnetic shielding braid termination with fastener rings [12]



Fig. 13: Installing braid termination rings with conductive heating [12]



Fig. 14: Heat shielding braid termination on oxygen sensor with fastener ring [15]



Fig. 15: High amperage pin/socket connector with fastener ring installed [16]

A similar concept is used for ZIF (zero insertion force) connectors. In a technically highly successful pin/socket version of a such a connector, a Ni-Ti ring surrounds the outwardbending tangs of a fork contact. When cooled (with liquid nitrogen, for instance), the ring weakens as it transforms to its martensitic phase, enabling the springy tangs to force it open. The mating pin then can be inserted or removed freely. Nearly one million contacts have been produced for the Trident program [15]. Other connectors incorporate U-shaped actuators that force open a spring clamp when heated with a foil heater attached to the actuator [17].



Fig. 16: Cryofit ® pin/socket connector [12]



Fig. 17: Printed circuitboard connector [17]

Shape memory actuators respond to a temperature change with a shape change [18]. The change in temperature can be caused by a change of ambient temperature or by electrically heating the shape memory element. In the first case, the shape memory alloy acts as a sensor and an actuator (thermal actuator). In the second case, it is an electrical actuator that performs a specific task on demand. Thermal as well as electrical shape memory actuators combine large motion, rather high forces and small size, thus they provide high work output. They usually consist of only a single piece of metal, e.g. a straight wire or a helical spring, and do not require sophisticated mechanical systems. Although originally considered most important, actuators are the technically and economically least successful applications of the shape memory effect, when measured as outcome vs. development effort. The reason for the limited success of shape memory actuators are technical insufficiencies as well as cost. Design requirements usually include transformation temperature on heating, reset temperature (hysteresis), force (stress), displacement (strain), cyclic stability (fatigue), response time on heating and cooling, dimensions, over-temperature and over-stress tolerance, etc..



Fig. 18: Thermostatic control valve (cut-away)



Fig. 19: as 18, function schematic [19]

An example of a technically as well as (at least for the user) economically successful application of a thermal shape memory is the thermally responsive pressure control valve in the Mercedes-Benz automatic transmission. To improve the shifting comfort, the shifting pressure of the transmission is reduced during cold start situations and increased again when the transmission reaches operating temperature [19]. Introduced in model year 1989 Mercedes cars, this system has operated extremely reliably. Why is this application so successful? The required Af temperature is 60°C with a comfortable ± 5°C tolerance, the spring is completely immersed in the transmission fluid, thus heating and cooling is slow and very uniform, the required force is low (approx. 5 N), very small displacement, maximum ambient temperature is 130°C, only 20,000 cycles expected. This fortunate combination of design parameters is seldom found. There has been a wealth of suggested shape memory applications for automotive use, like the "smart idle screw", carburetor ventilation valve, oilcooler bypass valve to name a few [20]. Other applications of thermal shape memory actuators marketed today include viscosity compensating devices, ventilation valves, antiscald valves, fire detection and prevention devices, air conditioning and ventilation devices, etc..



Fig. 20: "Smart idle screw" (prototype [20]



Fig. 22: Oilcooler bypass valve (prototype)



Fig. 21: Carburetor ventilation valve (prototype)



Fig. 23: as 22, schematic function



Fig. 24: Clogging indicators for oil coolers [21]



Fig. 25: Automatic gas line shut-off valves [22]



Fig. 26: anti-Scald valve



Fig. 27: Motion mechanism in toys

Electrical shape memory actuators have been suggested to replace solenoids, electric motors etc. By controlling the power during electrical actuation, specific levels of force and/or specific positions can be maintained. A variety of valves, triggering devices, animated objects, toys etc. are presently being marketed. The integration of Ni-Ti wires in composite structures has been suggested, to allow the structure to change shape on demand. These "smart composites" can also actively attenuate acoustic noise in structures by having fundamental control over structural stiffness. Strain-compliant shape memory composites can be used as integrated members in truss structures, performing passive and active roles in vibration and shape control. Recently, a system to dampen the low frequency swing of large antennas or reflectors during space shuttle maneuvers has been proposed, using a shape memory controlled hinge system [23].



Fig. 28: Active damping system [23]

Fig. 29: Smart composites for shape control [24]

Limiting factors for the use of shape memory alloys in electrical actuators are the transformation temperatures available today and the lack of control over cooling times. In order to work properly, the Mf temperature of the shape memory alloy must be well above the maximum operating temperature of the actuator. Commercially available alloys that are sufficiently stable in cyclic applications, have maximum transformation temperatures (Mf) of around 70°C. Thus, an electrical actuator made from this alloy would fail to reset when ambient temperature reaches 70°C. Correspondingly, the actuator would self-trigger when ambient reaches its Ass temperature. For applications with high operating temperatures (e.g. automotive), alloys with transformation temperatures well above 150°C are required. As mentioned above, Ni-Ti-Pd alloy with transformation temperatures up to 200°C might eventually become available.

The use of shape memory for actuator for robots has often been proposed, and several prototypes have been presented. However, as the shape memory effect is a thermal phenomenon, response time is dictated by the heating and cooling of the material. While heating can be controlled through the power supplied to the actuator, cooling is less controllable. Depending on the size of actuator (wire diameter, mass), cooling times can be seconds to minutes.

As mentioned earlier, applications using superelastic Ni-Ti have seen explosive growth during the last two years, with antennae, brassieres and eyeglass frames being the volume leaders. followed by dental archwires and guidewires. The first application of superelastic Nitinol was as orthodontic archwire during the 1970s. The advantages that Nitinol provides over conventional materials, obviously are the increased elastic range and a nearly constant stress during unloading [25].

Superelastic Nitinol guidewires are increasingly used because of their extreme flexibility and kink resistance. They also show enhanced torquability (the ability to translate a twist at one end of the guidewire into a turn of nearly identical degree at the other end) [26], thus significantly improving steerability. The low force required for bending the wire is considered to cause less trauma than stainless steel guidewires. Kink resistance and steerability are also the main reasons for using Nitinol in stone retrieval and fragmentation baskets. The shaft as well as the basketwires can be made from superelastic Nitinol.

More recently, shape memory and superelastic Nitinol alloys have been used very effectively for self-expanding stents. The small profile of the compresses stent facilitates safe, atraumatic placement of the stent. After being released from the delivery system, the stent self-expands either elastically or thermally and exerts a constant, gentle radial force on the vessel wall.



Fig. 30: Self-expanding Nitinol stent [27]



Fig. 31: Non-kinking microsurgical instrument

Medical device manufacturers are increasingly using Nitinol in instruments and devices for minimally invasive procedures [28]. The concept is to enter the body with a minimum profile through small incisions with or without a portal, and then changing shape inside the body cavity. One of the first instruments to use superelastic Nitinol was the Mitek Mammalok ® needle wire localizer, used to locate and mark breast tumors so that subsequent surgery can be more exact and less invasive [29]. The concept of constraining a curved superelastic component inside a cannula during insertion into the body is used in a variety of instruments

for minimally invasive surgery. Figure 32 shows a dissecting spatula, the curvature of which is increased by progressive extrusion of the superelastic blade. Different blade configurations are used for variable curvature suture and sling passers [30]. Instruments with deflectable distal ends use curved superelastic components which are constrained in a cannula during insertion into the body and deployed once inside the body. Graspers, needle holders and scissors can be inserted through straight trocar cannulae. Once inside the peritoneal cavity, they can change into their curved configuration, thus increasing the degrees of freedom for manipulation [31].





Fig. 32: Retractable spatula [30]

Fig. 33: Hingeless instruments [32]

In a new electrosurgical device for transurethral ablation of prostatic tissue, radiofrequency energy is delivered into the prostate via two side-deploying needles. These needles, made from superelastic Nitinol, are deflected from the axis of the catheter around a sharp bend to be deployed radially through the urethral wall into the prostate tissue. After passing the guiding channel, they protrude straight out of the catheter tip [33].

Hingeless instruments use the elasticity of spring materials instead of pivoting joints to open and close the jaw of grasping forceps or the blades of scissors. Because of their simple design without moving parts and hidden crevices, they are easier to clean and sterilize. A new generation of hingeless instruments uses superelastic Nitinol for the actuating component of these instruments, which provides elasticity higher than stainless steel by at least a factor of 10. This results in an increased opening span and/or reduced displacment of the constraining tube for ergonomic handling. In many cases the functional tip can be a monolithic superelastic component, vs. multiple intricate, precision machined components and linkages of conventional instruments. This allow the design of instruments with very small profiles [32].

Long and thin instruments, e.g. like forceps used in urology, tend to be very delicate and can kink easily, destroying an expensive tool. Using superelastic Nitinol for the outer tube and a superelastic actuation rod, makes the instrument very flexible and kink resistant. Superelastic tubes have only recently been made available by different suppliers. They are also used for biopsy needles, e.g. for interventional computer tomography or magnetic resonance imaging. In these techniques Nitinol instruments can be clearly detected without artifacts (glow) [34].

References

- [1] T.W.Duerig, K.N.Melton, D.Stoeckel, C.M.Wayman: "Engineering Aspects of Shape Memory Alloys", Butterworth-Heinemann, 1990
- [2] T.W.Duerig, R. Zadno, in [1] 369
- [3] S.M.Tuominen, R.J.Biermann, J.o.Metals 1988, 32
- [4] J.H.Mulder et al., Proc. of SMST (A.R. Pelton et al.eds.) (1994) to be published
- [5] D.Stoeckel, Proc. ADPA Conf. "Smart Structures" Washington (1990)
- [6] T.W.Duerig, K.N.Melton, Pro. SMA 86, Guilin (1986) 397
- [7] J.L.Proft, T.W.Duerig, in [1] 115
- [8] D.Stoeckel, Advanced Materials & Processes, 138 (1990) Oct., 33
- [9] R.G.Gilbertson: "Working With Shape Memory Wires", Mondotronics 1991
- [10] J.Perkins, D.Hodgson, in [1] 195
- [11] T.W.Duerig, Proc.MRS Conf. Boston (1994) to be published
- [12] "Tinel Shape Memory Alloys", RAYCHEM company literature
- [13] T.W.Duerig, K.N.Melton, J. Proft, in [1] 130
- [14] H. Kornfeld, in [4] to be published
- [15] D. Stoeckel, T. Borden, Metall 46 (1992) 668
- [16] H.P. Kehrer, H. Nußkern, VDI/VDE "Actuator 94" Bremen (1994) 324
- [17] J.F. Krumme, Connection Technology, April (1987)
- [18] D. Stoeckel, T. Waram, SPIE Vol 1543 "Active and Adaptive Optical Components" 1991,382
- [19] D. Stoeckel, J. Tinschert, SAE Technical Paper Series #910805
- [20] D. Stoeckel, Springs Vol. 30, Oct. (1991) 35
- [21] J. F. Tertre, D. Stoeckel, Proc. "Actuator 92", VDI/VDE Bremen (1992), 100
- [22] "FIRECHECK" MEMRYTECH company literature
- [23] T. Stevens, Mat. Eng., Vol. 108, Oct. 1991, 18
- [24] Proc. ADPA Conf "Smart Structures" Washington (1990)
- [25] R. Sachdeva, S. Miyazaki, in [1] 452
- [26] J. Stice, in [4] to be published
- [27] ANGIOMED AG, company lieterature
- [28] A. Melzer, D. Stoeckel, in [4] to be published
- [29] J.P.O'Leary, J.E.Nicholson, R.F.Gatturna, in [1] 477
- [30] STORZ, company literarure
- [31] P.P.Ponzet, R. Zadno, in [4] to be published
- [32] D. Stoeckel, A. Melzer, Proc. 8th CIMTEC Florence (1994) to be published
- [33] TUNA, VIDAMED Int'l, company lieterature
- [34] D.H.W. Grönemeyer, R.M.M. Seibel, "Interventionelle Computertomographie", Ueberreuter Wissenschaft, Wien Berlin (1989) 308