

Shape Memory Alloy Fasteners

June 29, 2022

Intrinsic Devices, Inc.
2353 3rd St. , San Francisco, CA 94107

Introduction:

Shape memory alloys have been used to produce fasteners and seals for over 40 years. Most of the applications utilize cylindrical parts that shrink in diameter to deliver a uniform radial clamping force that is extremely difficult to replicate by other methods. The early adopters of the technology were engineers in the military and aerospace industries driven by the need to maximize performance of high reliability systems. Since then use has spread to sensors, downhole tools , manufacturing equipment and medical devices. While the technology is mature, the broader engineering community remains largely unaware of this capability.

This paper describes the properties of shape memory alloys as they relate to fasteners. The discussions focus on heat shrinkable rings, although most of the concepts are valid for fasteners which operate in other modes such as radial expansion, axial contraction, or axial expansion. There are a variety of alloys that exhibit shape memory. This paper will be restricted to binary NiTi alloys and some ternary alloys such as NiTiFe and NiTiNb.

Table of Contents:

Section	Subject
1	Shape Memory Alloys
2	Crystallographic Origin of the Shape Memory Effect
3	Macroscopic Deformation Behavior
4	Free and Constrained Recovery
5	Transformation Temperatures
6	Wide Hysteresis Alloy
7	Stress-Temperature Relationship
8	Elastic Modulus
9	Operating Stresses
10	Repeatability
11	Operating Limitations

1 Shape Memory Alloys

Shape memory alloys are a unique family of metals which exhibit a temperature dependent shape change. This shape change is much more dramatic than that caused thermal expansion and is produced by a completely different mechanism. A variety of alloys such as copper-aluminum-nickel, copper-zinc-aluminum, iron-manganese-silicon and nickel-titanium exhibit the shape memory effect. The most useful of these are alloys based on a nominal composition of 50/50 atomic percent nickel titanium. Small amounts of a third element may be added to achieve specific properties. Nickel titanium alloys are frequently referred to as Nitinol, for Nickel Titanium Naval Ordinance Laboratory where the shape memory properties of the alloy were discovered.

NiTi SMAs can be deformed in tension, compression, or shear. The deformed part is stable until heated above its transformation temperature, at which point it "recovers" toward its original shape. Depending on the material processing and the deformation mode, the recovery strain ranges from 2.5% to 6%. The lower strains are for SMA's operating in compression and the higher strains for tension. If the recovery is resisted, the SMA can develop stresses up to 480 MPa (70 kpsi).

2 Crystallographic Origin of the Shape Memory Effect

This unique effect is produced by a crystalline phase change known as a martensitic transformation. This transformation occurs over a range of temperature. Above this temperature range, the material is in the austenitic phase and below it, the material is in the martensitic phase. The geometric shape of the material when austenitic is termed the memory shape. Martensite forms on cooling from austenite by a shear type of process. If stress is not applied to the material during cooling, no macroscopic shape change takes place. The martensite forms in several different orientations, each with alternating senses of shear, called twins. It is the alternating nature of these twins which allows the transformation to occur without an external shape change. This is illustrated in two dimensions in Figure 1.

By applying a low stress, these twins can be deformed by shear into a single orientation, as illustrated in the figure. This deformation is reversible, that is, it does not involve permanent plastic deformation. This deformed martensite will hold its shape indefinitely, as long as the temperature of the material is held below the transformation temperature of the alloy.

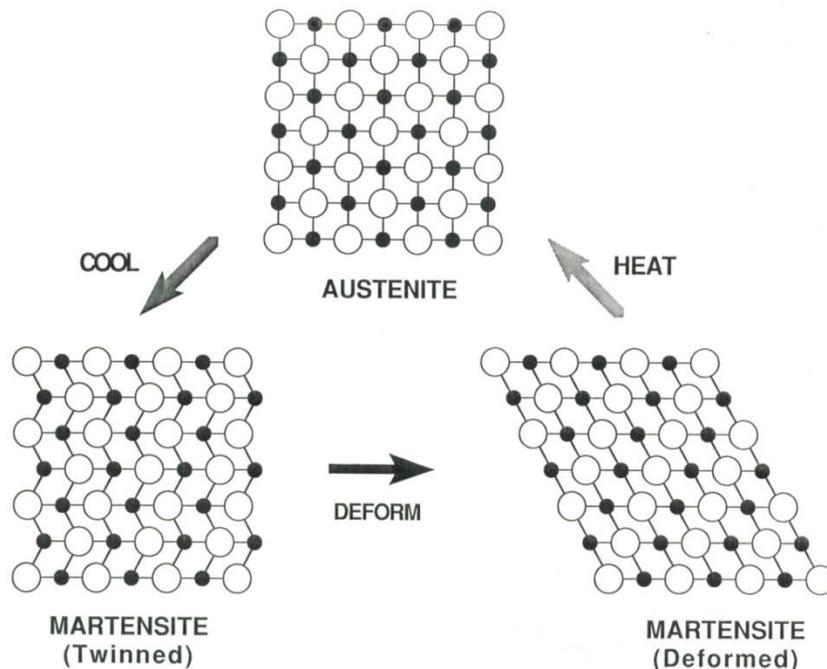


Figure 1 Schematic of Crystal Structures

If the deformed martensite is now heated through its transformation temperature range, it reverts to austenite. However, the austenitic crystalline structure cannot accommodate the deformation which

was applied to the martensite. As a result, the material must return to its memory shape as it reverts to austenite. This cycle of cooling, deformation, and shape recovery is the fundamental shape memory process.

It is worth noting that the more familiar martensitic materials cannot be deformed in this manner. In most steels, for example, the formation of martensite is associated with a large hardness increase.

3 Macroscopic Deformation Behavior

The deformation behavior of the martensite is manifested macroscopically by the stress-strain curve of a shape memory alloy, Figure 2. The martensitic curve shows an initial low yield, on the order of 200 MPa (30 kpsi), which is the stress required to reorient the structure. Beyond this, several percent strain is accumulated with little stress increase. The reorientation process is essentially complete at the end of this plateau. Subsequent deformation is conventional in nature, that is, elastic loading followed by permanent plastic deformation at higher stresses. In the austenitic phase, this twin reorientation mechanism is not possible and so the yield strength is much higher, from 410 MPa (60 kpsi) to 550 MPa (80 kpsi). Martensite has a higher ultimate tensile strength and lower elongation to failure since it exists, and is tested, at a lower temperature than austenite.

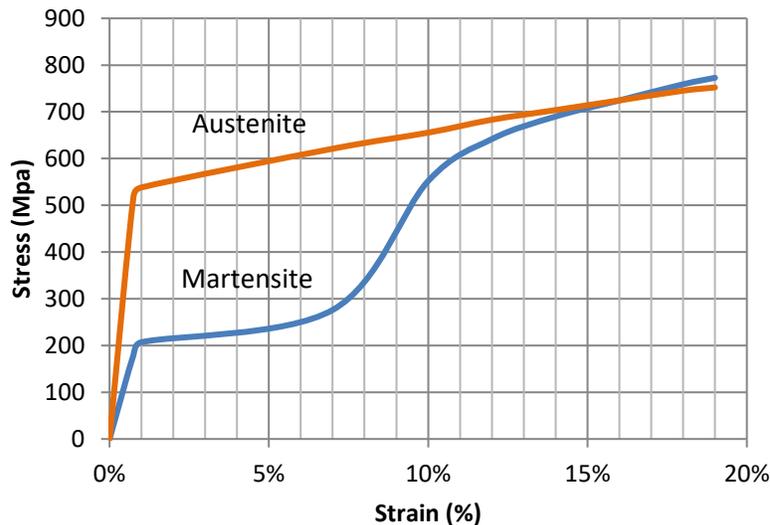


Figure 2 Comparison of the Stress-Strain Behaviors of Austenite and Martensite

4 Free and Constrained Recovery

Once the martensite has been deformed, it is stable in this deformed state until heated above its transition temperature to austenite. Figure 3 illustrates the stretching, unloading, and subsequent recovery to austenite by heating. Two possible recovery curves are shown. In one, the material is allowed to freely return to its memory shape at zero stress.

The second path shows a case where the material is heated to recover freely to some strain, at which point the strain is fixed. As the material is heated further it builds up a stress in attempting to return to

its memory shape. The peak stress generated is a function of the unresolved recovery strain, or strain between the constrained shape and the freely recovered shape. If a series of these tests were run, fixing the SMA at various unresolved recovery strains and measuring the peak stress generated, a plot of those strain-stress points would look like Figure 4. The stress rises steeply up to 1.5% unresolved recovery, at which point a plateau is reached where stress increases only slightly with increasing strain. This plateau stress is on the order of 410 MPa (60 ksi) for most alloys.

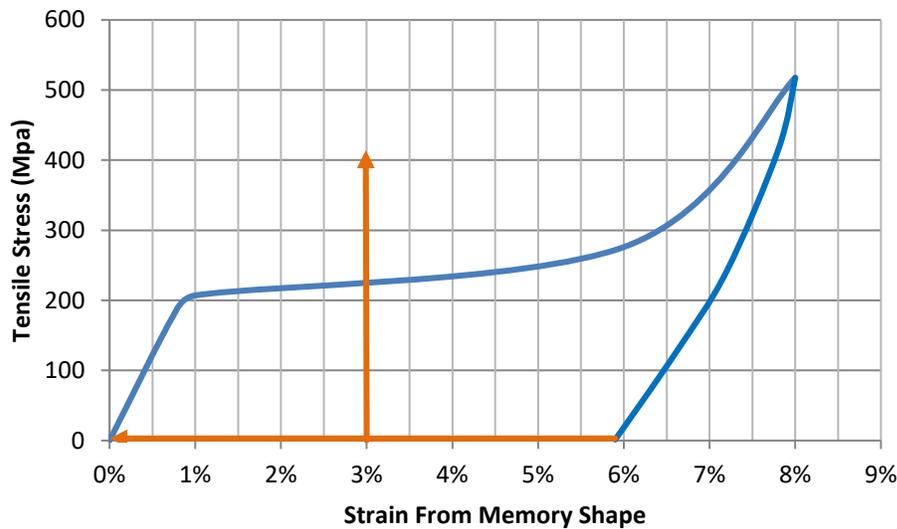


Figure 3 Deformation of Martensite, Unloading, and Recovery to Austenite

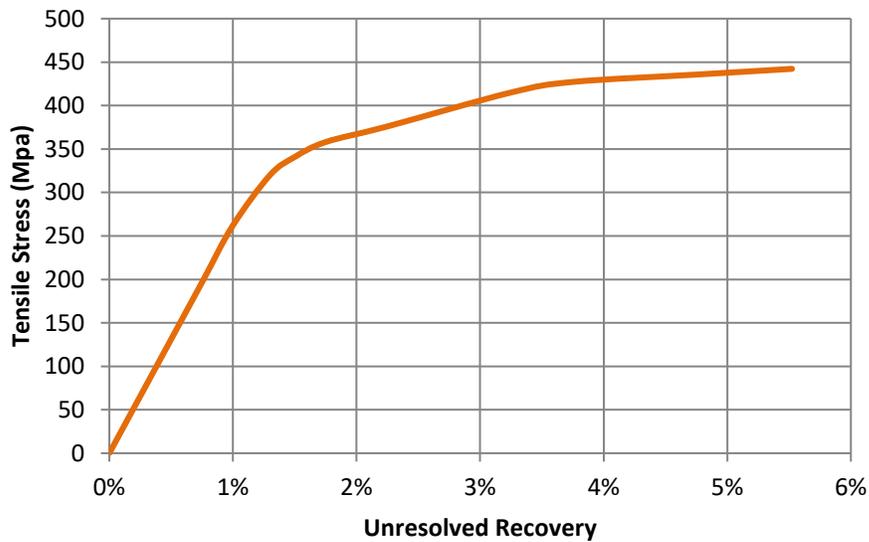


Figure 4 Recovery Stress vs Unresolved Recovery

The stress versus unresolved recovery curve looks similar to the austenitic stress-strain curve but the stresses are lower and the knee in the curve that relates to the yield point on the stress –strain curve is more rounded and occurs at a higher strain. Common design practice is to size SMA rings and the parts they will be installed on such that the unresolved recovery will be above 1.5%. Because of the shallow slope of the curve above 1.5%, the hoop stress generated by the rings will be consistent within a range of +/-13%. Even higher and more consistent stress generation can be achieved by designing with greater than 1.5% minimum unresolved recovery. Practical designs can be achieved at less than 1.5% unresolved recovery, but care must be taken in qualification testing to insure that worst case dimensions are used for testing.

Our fastener ring drawings tabulate a recommended Minimum Substrate Diameter for each ring that corresponds to 1.5% unresolved recovery.

During constrained recovery, the SMA yields itself in trying to return to its memory shape. Because plastic deformation has occurred, the SMA part will have a new larger memory shape. This new shape will be less than the constraint size by the amount of stored elastic strain in the part. For example, a part installed with 3% unresolved recovery will generate a recovery stress of 410 MPa (60 kpsi). Since the austenitic elastic modulus is about 83 GPa (12 x 10⁶ psi), this elastic strain in the SMA part will be 0.5%. The new memory size is 2.5% larger than the original memory size. See Figure 5.

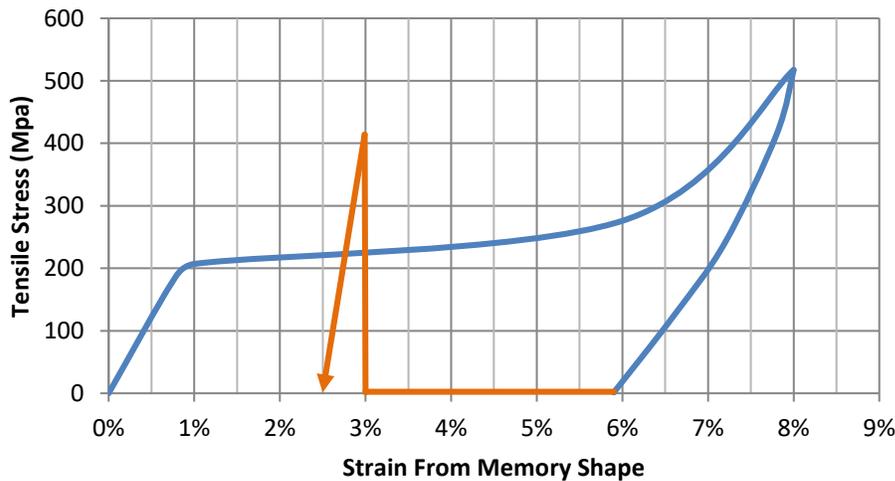


Figure 5 Elastic Unloading After Recovery on a Constraint

The curves above are for parts deformed in tension. The curves are similar for parts deformed in compression, except the strains are about half as large.

The 0.5% installed elastic strain in a SMA ring is much larger than what can typically be achieved by other fastening techniques such as crimps, thermal shrink fits, or bolted joints. This means the SMA part has more “follow” to accommodate changes due to differential thermal expansion, cold flow of polymers, vibration, etc.

It should be noted that the transformation to Martensite, deformation of Martensite, and recovery to Austenite are all essentially constant volume processes. So, as a previously stretched part is recovered, its cross-sectional area will increase. For a shrinking ring, the radial thickness and axial length will each increase 0.5% for every 1% of diameter contraction.

5 Transformation Temperatures

The temperature at which the material transforms from martensite to austenite is controlled by alloy composition and processing. Figure 6 shows the transformation curves for two alloys. There is a hysteresis between the heating curve, martensite to austenite, and the cooling curve, austenite to martensite. The temperature at which Martensite begins to form is called the Martensite Start Temperature, M_s . Similarly, the temperature at which austenite begins to form is called A_s . These temperatures are determined graphically by plotting a measured characteristic, drawing in tangent lines, and then reading the temperature at the intersection of the two lines. For example, the transformation temperatures for the “warm” alloy on the right are:

Austenite Start, $A_s = 75^\circ\text{C}$
Austenite Finish, $A_f = 79^\circ\text{C}$
Martensite Start, $M_s = 45^\circ\text{C}$
Martensite Finish, $M_f = 41^\circ\text{C}$

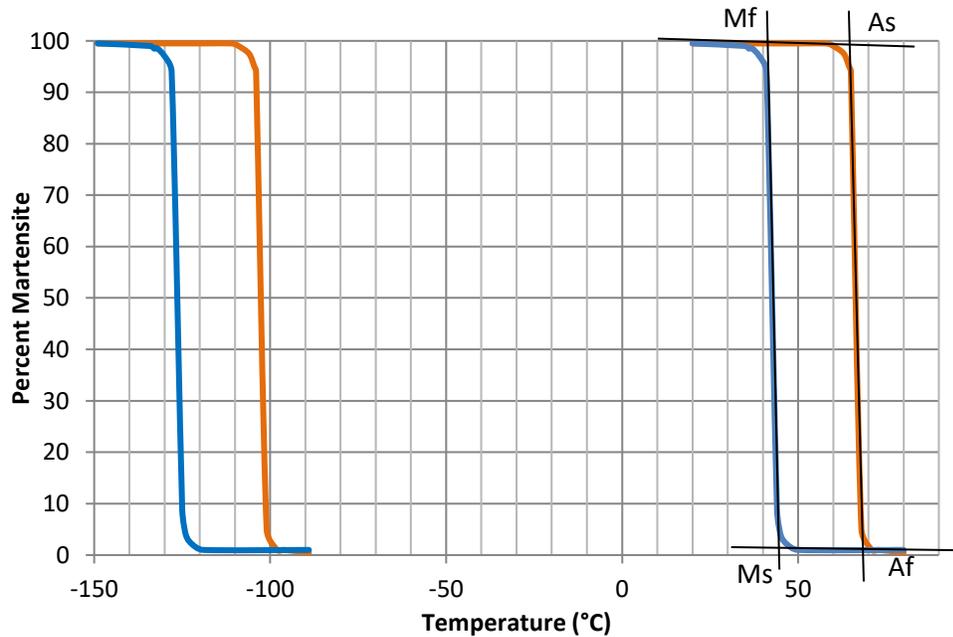


Figure 6 Transformation Cycles For Two Alloys

Consider a ring made of the material with the higher transition temperatures. It would be martensitic at room temperature. It could be expanded to a larger diameter and held indefinitely at room temperature. If the ring were placed over a substrate with a diameter larger than the memory diameter

of the ring, and heated to above 79°C (174°F), the ring would recover and be locked in place. However, when the ring cooled back down to room temperature it would transform back to martensite, and its gripping force would be reduced to an unusable level.

In order to make a practical fastener, a lower transition temperature alloy must be used. The reversion to martensite must take place at a temperature that is lower than the minimum required operating temperature of the installed ring. A ring made of the "cold" alloy shown would be expanded at a temperature less than -125°C (-193°F). Such a ring would be shipped and stored in liquid nitrogen until ready for installation, at which point it would be pulled out and put into place. On warming to room temperature it would recover, shrinking onto the substrate. As long as the temperature of the ring stayed above -125°C (-193°F) it would maintain its grip. A material like this could be used to produce functional fasteners.

Alloys with very low transformation temperatures like this one are termed "cryogenic". Intrinsic Devices NiTiFe Alloy C is an example. While binary alloys can be formulated with excess nickel to have very low transformation temperatures, the characteristics are difficult to control and the transformation temperatures can shift over time due to precipitation of nickel from the matrix at potential operating temperatures. Adding iron depresses transformation temperatures more controllably without the excess nickel precipitation problem. NiTiFe was the first commercial shape memory alloy, originally used to make hydraulic couplings for the F-14A military aircraft beginning in 1971.

6 Wide Hysteresis Alloy H

For some applications cryogenic installation is an advantage. The substrate is not heated. The installation is very operator insensitive. No heating tool is required. However, in many instances assembly procedures do not allow the use of a part that will recover shortly after it is removed from a dewar of liquid nitrogen. Because of this limitation, an alloy was developed that has an extremely wide one-time hysteresis. This is our Alloy H which is formulated with 14 weight % niobium. The niobium increases the hysteresis and stabilizes the deformed martensite. Figure 7 is a plot of strain versus temperature for an Alloy H ring that has been previously expanded in diameter. The deformed martensitic ring can be stored/handled at temperatures up to 50°C (122°F). The ring is installed by heating above this. At just above 50°C (122°F) the ring begins to transform to Austenite. If nothing is inside the ring resisting its shrinkage, it will be fully recovered by 100°C (212°F). In order to generate full recovery stress, the ring must be heated further. This is explained in the next section. Once installed, the part maintains its grip down to -100°C (-148°F).

A number of important behaviors are illustrated in Figure 7. When the ring is heated for the first time, it begins to shrink sharply at about 50°C (122°F). We refer to this temperature as As' . On cooling under no load, the ring spontaneously expands in diameter. This is termed two-way shape memory. The two-way strain for Alloy H is on the order of 0.75%, much less than the initial recovery strain of about 5.5%. Since a ring installed on a rigid substrate has about 0.5% elastic interference, the two-way strain is just enough to loosen the ring. Some SMAs with higher transformation temperatures exhibit larger two-way strains, up to about 3%.

While the As' is 50°C (122°F) on the first heating cycle, on all subsequent cycles it exhibits the lower value of $As = -50C$. This is because the deformation during ring expansion stabilizes the martensite, leading to the higher transformation temperature to austenite on the first cycle. The martensite is actually in a metastable condition at room temperature. An expanded ring can be transformed to austenite at room temperature by repeated bending or by radial compression.

Another thing worth noting on the curve is the initial downward slope when the ring is first heated above room temperature. This shrinkage is referred to as premature recovery. The magnitude of this strain is about 0.25%.

If you measure the as-supplied diameter of Alloy H rings, you will notice they are substantially larger than the tabulated Minimum Supplied Inside diameter. The rings are produced oversized so that they will still meet the Minimum Supplied Inside Diameter after being heated to 45°C (113°F), the specified stability temperature for shipping and handling.

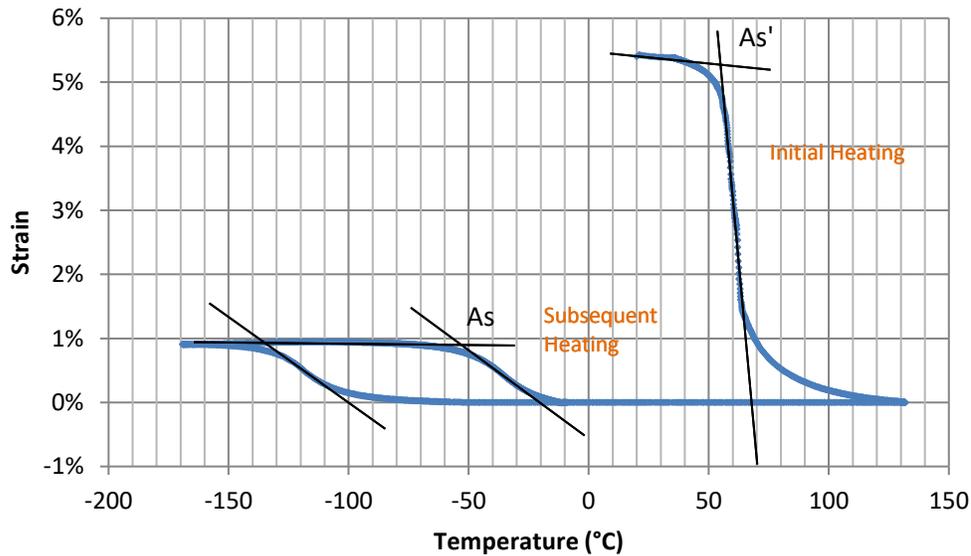


Figure 7 Transformation Cycle For Wide Hysteresis Alloy

7 Stress-Temperature Relationship

The transformation temperatures of a given alloy are actually a function of the stress that is applied to the material. An applied stress increases the transformation temperatures, just as applying pressure to water raises its boiling temperature. The rates vary from alloy to alloy going from 0.25°C/MPa (3.1°F/kpsi) for cryogenic alloys to 0.11°C/MPa (1.38°F/kpsi) for the warmest binary NiTi alloys. For Alloy H, Ms temperatures increase with applied stress at a rate of about 0.21°C/MPa (2.57°F/kpsi).

For example, Alloy H with a -100°C (-148°F) Ms at zero stress would have an Ms of -43°C (-45°F) when under a 275 MPa (40 kpsi) tensile stress. As another way of looking at this, consider a stress-strain test of this material at -43°C (-45°F). At the start of the test the material is fully austenitic. As the stress is gradually increased, the material will remain austenitic until the stress reaches 275 MPa (40 kpsi), at which point martensite will begin to form and the material will begin to deform in response to the stress. When martensite is produced in this manner, it is called stress induced martensite.

This interplay between stress and temperature manifests itself during the initial heating to install a ring as well as subsequent cooling to room temperature and below.

Consider an expanded Alloy H ring positioned over a shaft that is 2% larger than the memory diameter of the ring. The ring shrinks as it is heated, contacting the shaft at just above 50°C (122°F). As the ring is heated further, the stress rises from zero at a rate of 4.76MPa/°C (.389°kpsi/°F). As the stress gets higher, the curve rolls over, eventually reaching a maximum at about 200°C (392°F). See Figure 8.

The recommended installation temperature for Alloy H rings is 165°C (330°F). This allows the material to deliver over 90% of the maximum clamping force and allows thermal compatibility with a wider range of materials. Alloy H parts can be supplied with temperature indicating paint that changes color at 165°C. For reference, this temperature is below temperatures required for soldering. If the full clamping stress is not required, Alloy H parts can be installed at temperatures as low as 100°C (212°F).

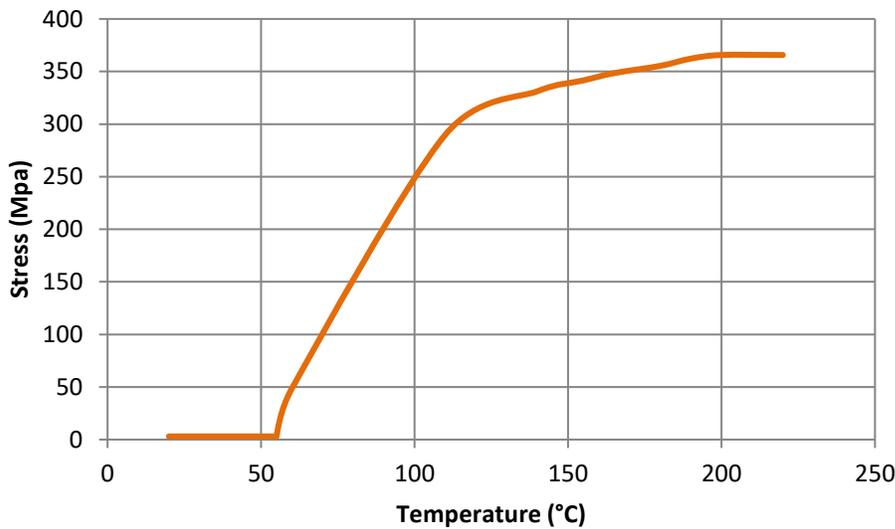


Figure 8 Stress Buildup During Constrained Recovery of Alloy H ring

Consider the ring described above with a zero stress M_s of -100°C (-148°F). At 165°C it has an installed hoop stress of 345 MPa (50 kpsi). At this stress level, the M_s will be -25°C (-13°F), as calculated using the stress-temperature dependence described earlier. So, as the assembly is cooled from room temperature the stress will be constant until -25°C (-13°F) is reached. At this point, the stress in the ring will begin to decay at a rate of 4.6 MPa/°C (0.37 kpsi/°F), until -100°C (-148°F) is reached and the stress has gone essentially to zero. See Figure 9.

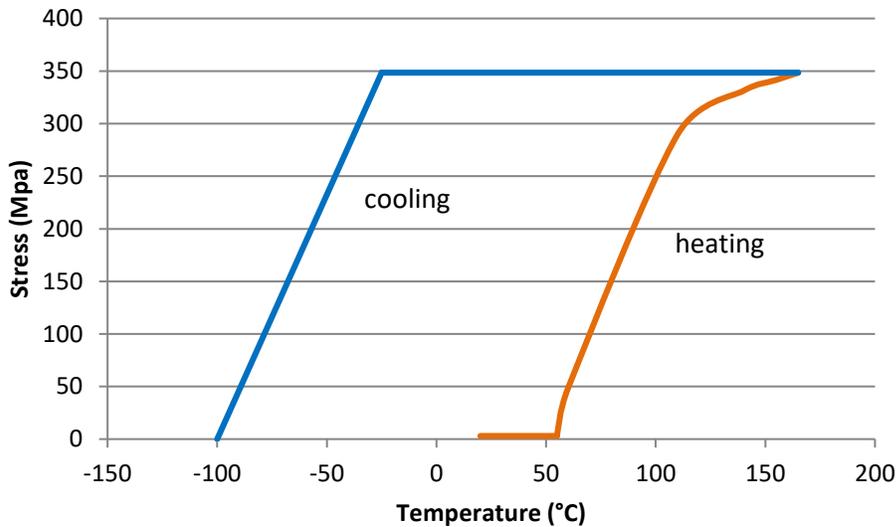


Figure 9 Stress Decline With Cooling Due to Martensite Formation (Modulus Change Excluded)

8 Elastic Modulus

Actually, stress induced martensite is not the only factor that controls stress in a fastener design. SMAs exhibit a significant variation in elastic modulus with temperature. At temperatures more than about 250°C (450°F) above the zero stress M_s , the elastic modulus of NiTi alloys is 83 GPa (12×10^6 psi). On cooling toward M_s , the modulus decreases. At M_s , the modulus is on the order of 21 GPa (3×10^6 psi).

Consider the Alloy H ring above, installed on a shaft with the same coefficient of thermal expansion as the SMA. As the assembly is cooled from 165°C, the elastic interference remains the same and the SMA modulus decreases. Since the stress is equal to the strain times the modulus, the stress in the SMA part will decrease in direct proportion to the modulus. The modulus of Alloy H is about 41 GPa (6×10^6 psi) at -55°C (-67°F). Thus, a part with an initial installed stress of 345 MPa (50 kpsi) would decline to 172 MPa (25 kpsi) at -55°C (-67°F). As you can see, this stress is below the 207 MPa (30 kpsi) which would be required to stress induce martensite. This is shown schematically in Figure 10. As the temperature is reduced further, martensite formation becomes the controlling factor and the stress falls at that sharper rate.

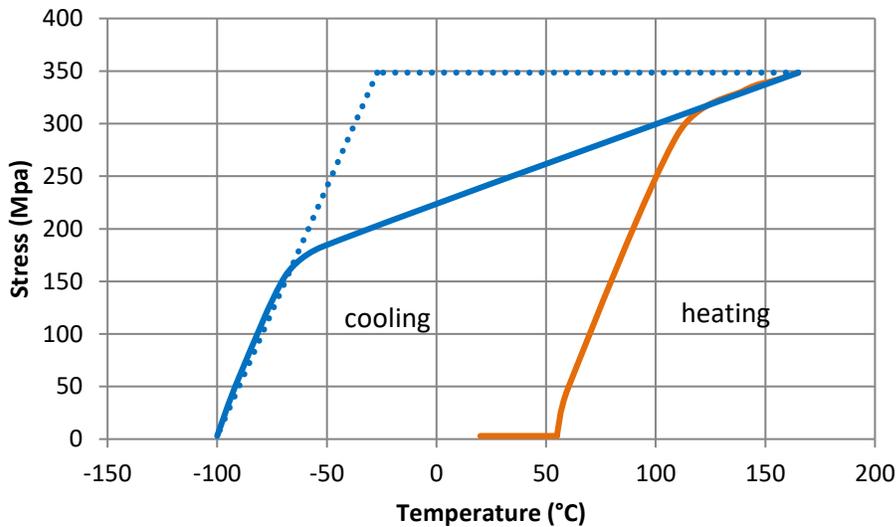


Figure 10 Stress Decline With Cooling Due to Martensite Formation and Modulus Decrease for Alloy H

9 Operating Stresses

When designing assemblies and testing them, keep in mind the maximum ring hoop stress will occur at the maximum operating temperature and the minimum hoop stress at the minimum operating temperature. Substrate stiffness and coefficient of thermal expansion also affect the ring hoop stress at any temperature. A rigid substrate with a high CTE, like a solid aluminum shaft, will produce higher stresses at high temperatures and lower stresses at low temperatures than a compliant low CTE substrate. The installation method also influences the operating stresses. If the ring is heated for installation while the substrate remains at room temperature, the ring hoop stresses will be higher at all temperatures. This is because the thermal contraction of the ring on cooling after installation results in a higher elastic strain in the ring.

The various effects discussed above can be combined into an analysis that is similar to that of a preloaded bolt. The initial preload is calculated from the stress versus unresolved recovery curve. A force balance is set up between the SMA and the parts it is installed on. Elastic deformations are determined at the starting point of 165C. Then the temperature is changed and the effects of differential thermal expansion, SMA modulus change and elastic interaction are used to calculate the new stress in the SMA. This is compared to the martensite formation stress at the new temperature. Whichever stress is lower will be the actual value. Intrinsic Devices has developed a code for these calculations which we routinely run for users of Alloy H rings.

On the Alloy H ring drawings a Nominal Clamping Force is tabulated for each part. This is intended to give a rough starting point for design. The clamping force is based on the hoop stress that would exist in a ring installed on a aluminum cylinder at 2% unresolved recovery after heating the ring and the cylinder together to 165C and then cooling to room temperature. This stress is 193 MPa (28 kpsi).

The document “Clamping Force Calculation” available on our website, presents equations to calculate radial clamping force and clamping pressure based on the ring dimensions and hoop stress.

10 Repeatability

The science described in the previous sections has been used for over 40 years to produce couplings, seals, fasteners and electrical connectors. The processes used to melt these alloys and turn them into products with consistent properties is proven. As an example, we have been producing the same BHE type welded wire ring for over 18 years. As part of the lot qualification testing, we install a sample ring on an instrumented puck to measure the radial clamping force exerted by the ring. See Figure 11. Over this time period, a succession of ingots and wire drawing and fabrication lots, the clamping force has only varied by +/- 10%.

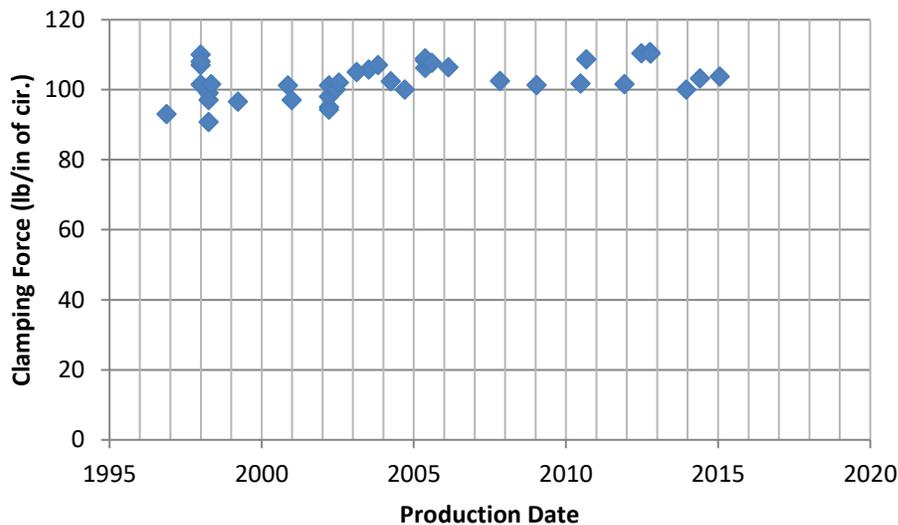


Figure 11 Ring Clamp Force From Multiple Production Lots

11 Operating Limitations

Low Temperature – As discussed earlier, ring clamping stress falls off as temperature is reduced, going to zero once below M_s . Many successful designs operate down to -55°C or -65°C . A combination of analysis and testing can be used to determine the ring cross section needed to deliver adequate clamping force at the minimum temperature.

High Temperature – Thermally activated creep, stress relaxation, becomes a concern when operating temperatures approach 300°C . The ring hoop stress will fall off gradually with time as it is held at elevated temperature. Figure 12 shows how long it takes a part to lose 20% of its initial hoop stress as a function of operating temperature.

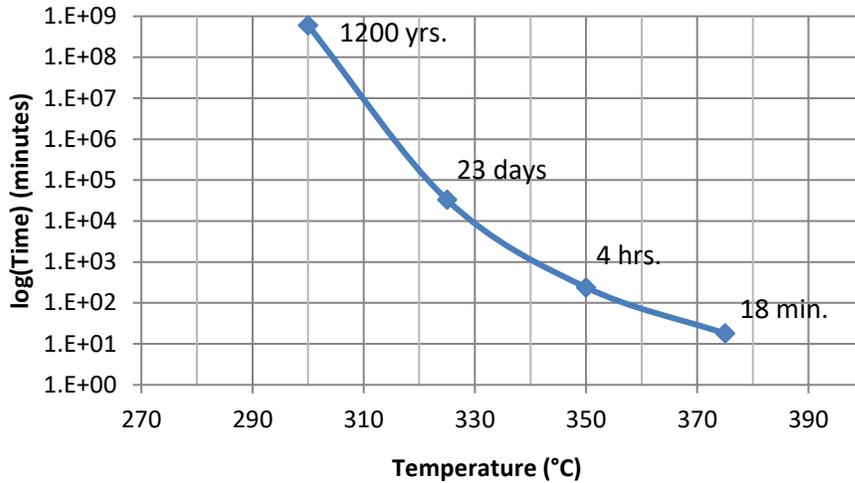


Figure 12 20% stress loss due to stress relaxation

Corrosion – The corrosion resistance of NiTi alloys is generally better than 316 stainless steel but not as resistant as pure titanium. Corrosion failures are uncommon. There have been tunneling corrosion failures in parts subjected to aqueous environments combining low pH with a high chloride concentration. Hydrogenated water at temperatures on the order of 300C are of concern for hydrogen embrittlement. Liquid metal embrittlement has also caused failures. Researchers have found that NiTiNb alloys offer superior corrosion resistance to other NiTi alloys. Since we started producing Alloy H products in 1993, there have been no reported corrosion failures. Please ask us about your specific environment and we will share the relevant information we have collected.

Fatigue – Because of the way SMA rings are used, stress fluctuations are generally due to thermal cycling and not mechanically induced. The magnitude of thermally induced fluctuations are small. There have been no reported fatigue failures of Alloy H parts.