بسم الله الرحمن الرحیم

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 $\int \sigma d\varepsilon$, 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

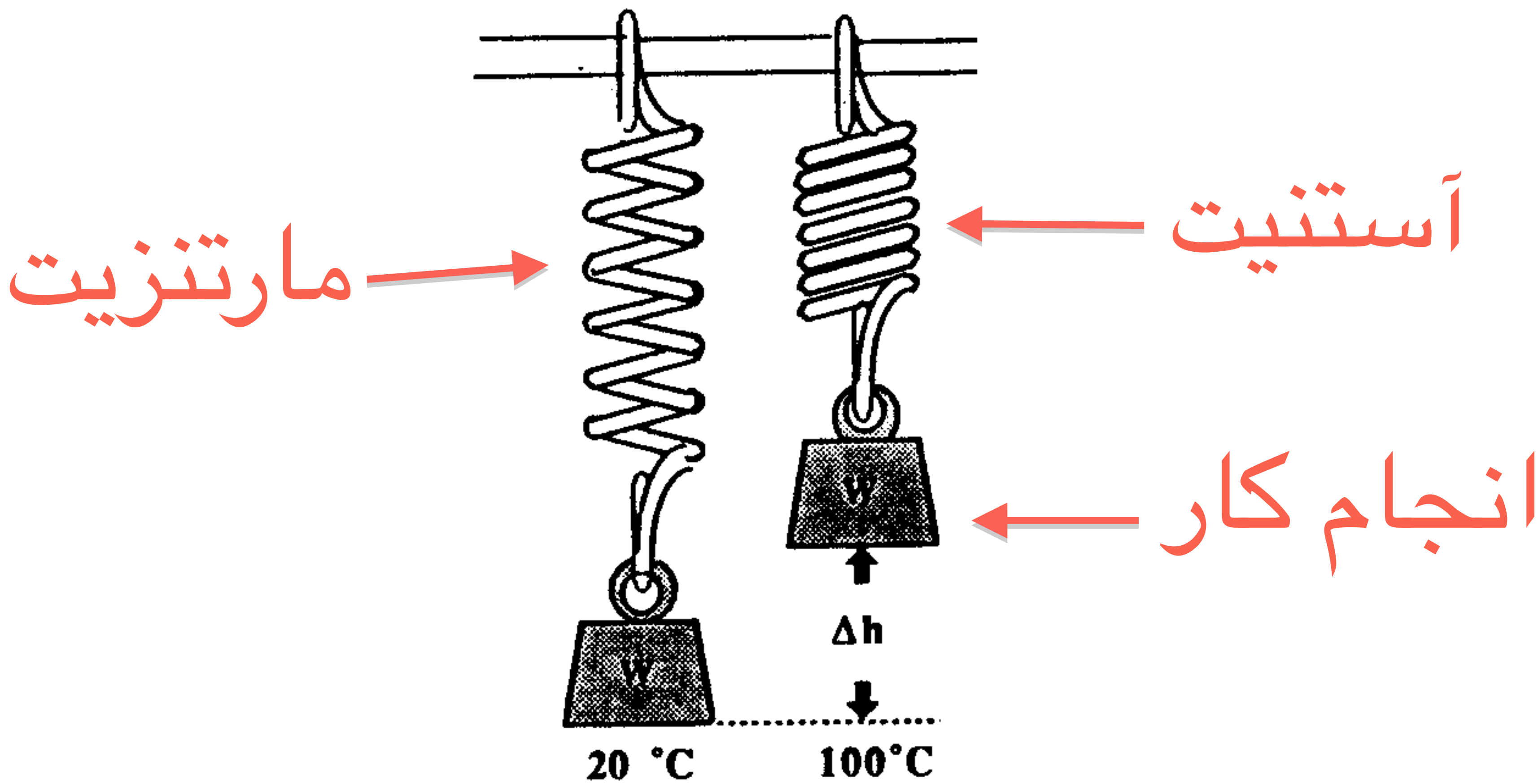

Figure 1: The classic and most simple SMA actuator consists simply of a shape memory spring with a weight on the end: as the spring is heated, the weight is lifted and when the spring is cooled, the weight stretches again.

applied by the weight (σ_o) is less than the recovery stress of the alloy, the weight will be lifted and the amount of work done will be $\sigma_o \Delta\varepsilon$. This of course describes only the work production during heating. If the applied stress is greater than the martensitic yield strength of the material $\sigma_y{}^m$, 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

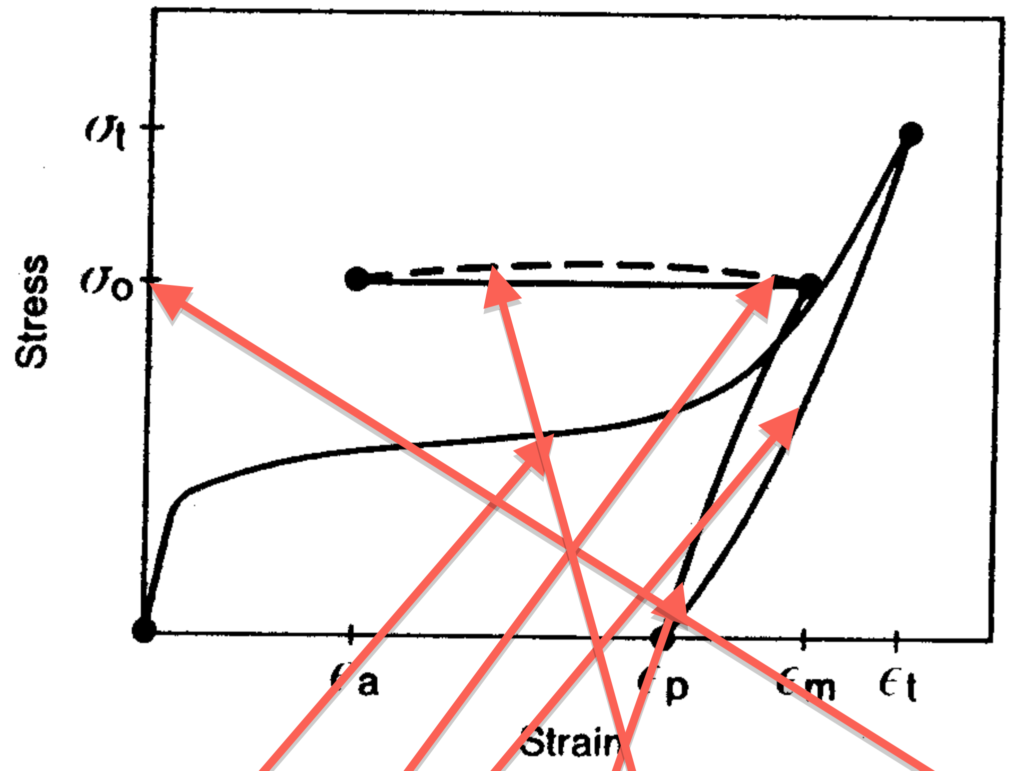

Figure 2: A stress-strain perspective of the actuator or work production event showing deformation, unloading, loading to the applied stress level σ_o, and heating to recover to ε_A. Subsequent cooling (dashed line) returns the deformed shape, ε_M.

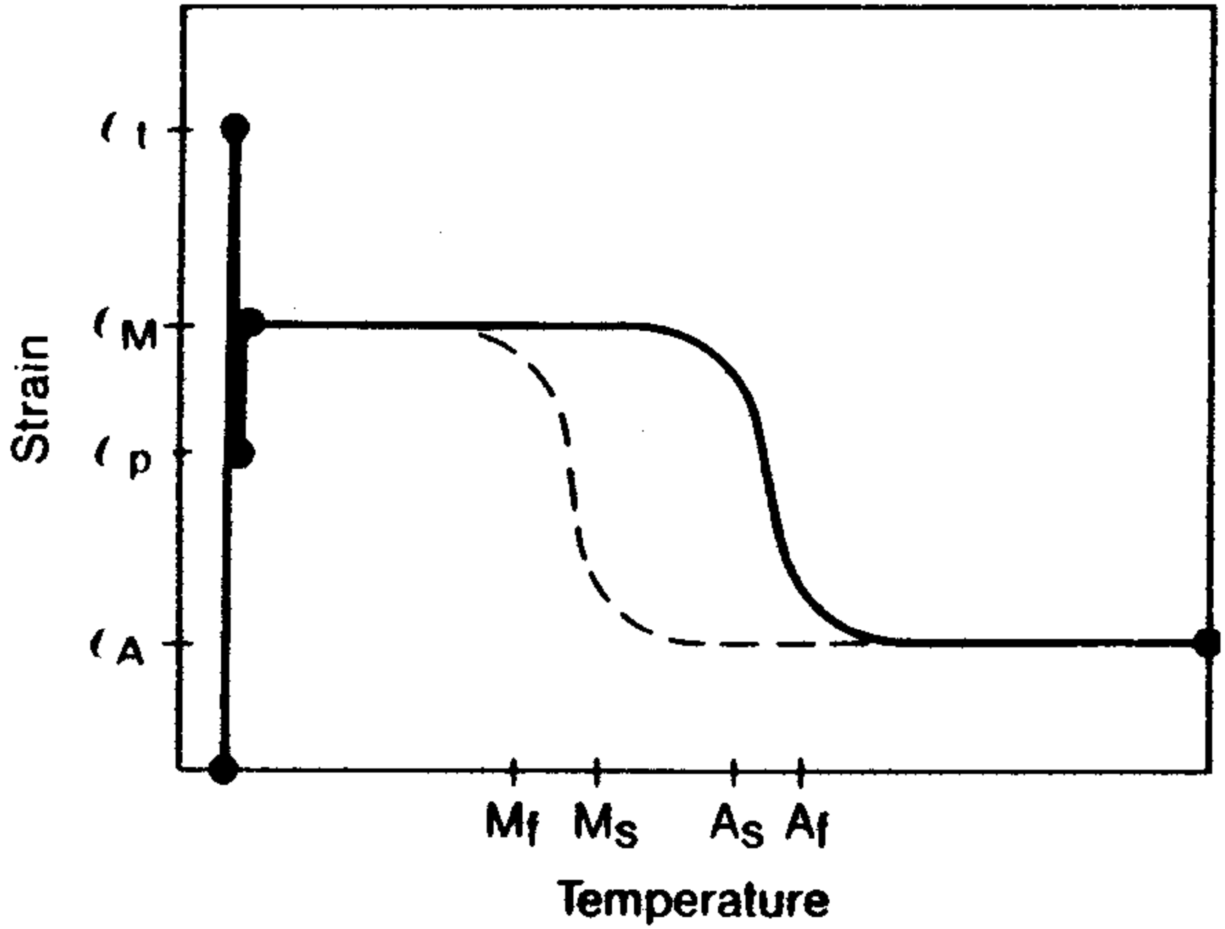

Figure 3: The strain-temperature perspective of the event described in Figure 1, showing the four transformation temperatures.

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_o). 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 $\sigma_y{}^m$. 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_o.)

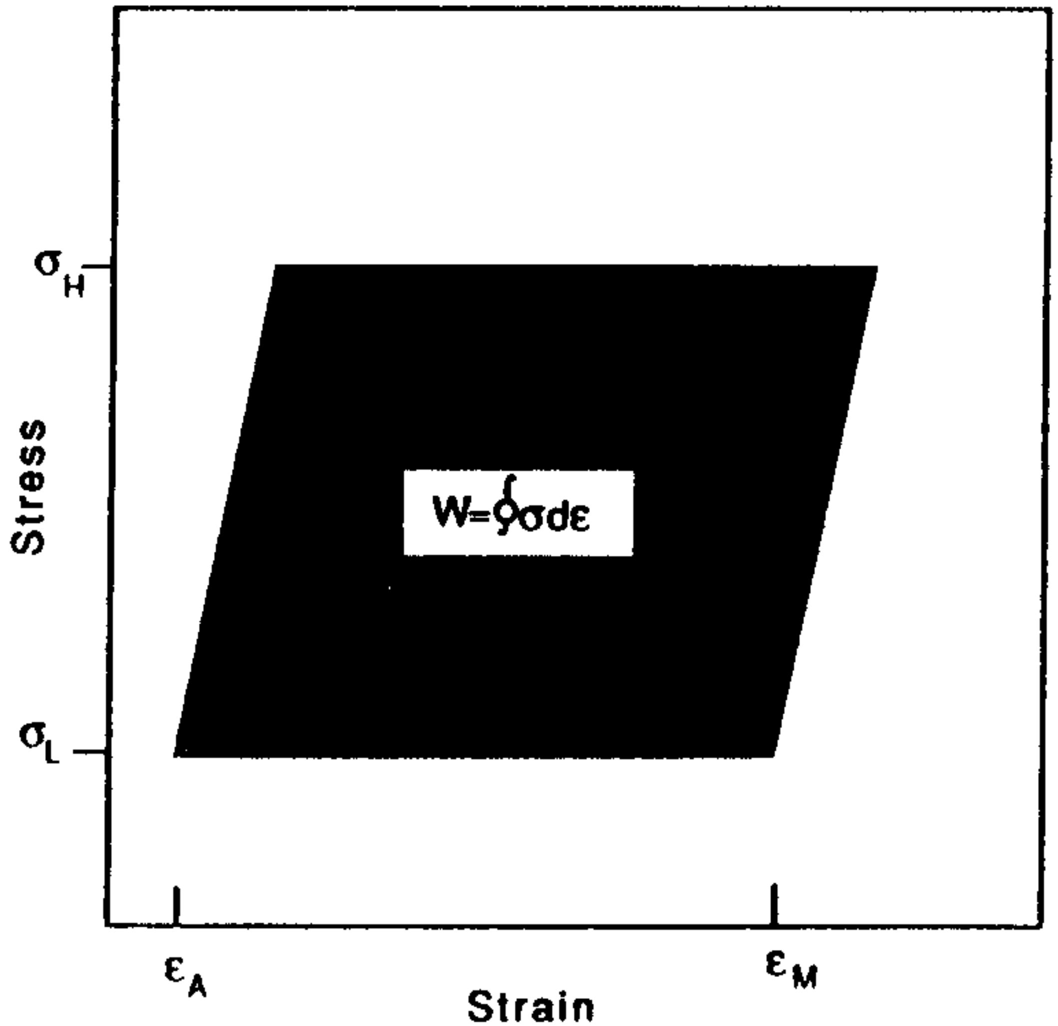

Figure 4: A typical actuator recovers against a higher load (σ_H) than is applied during cooling for resetting (σ_L), resulting in a <u>net</u> work output (shaded area).

One should be aware that σ_L can in fact be below $\sigma_y{}^m$ 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

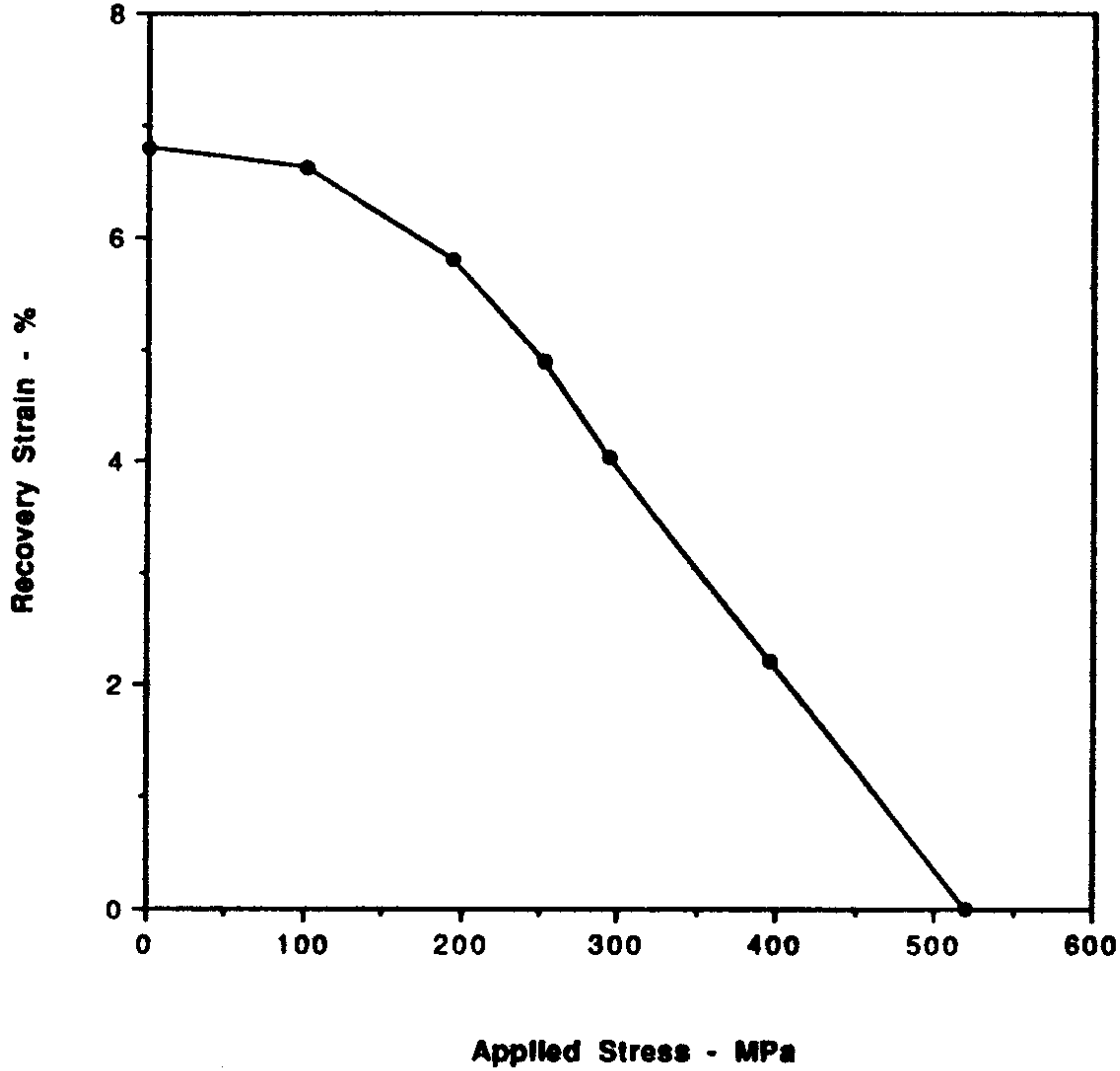

Figure 5: The recovery strain of a Ni-Ti-Fe alloy after a total deformation of 8% shows that the recovery strain (ε_M-ε_A) is reduced by an applied load. This particular alloy is fully annealed; stronger alloys would show full recovery against far greater stresses.

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.

Maximum work

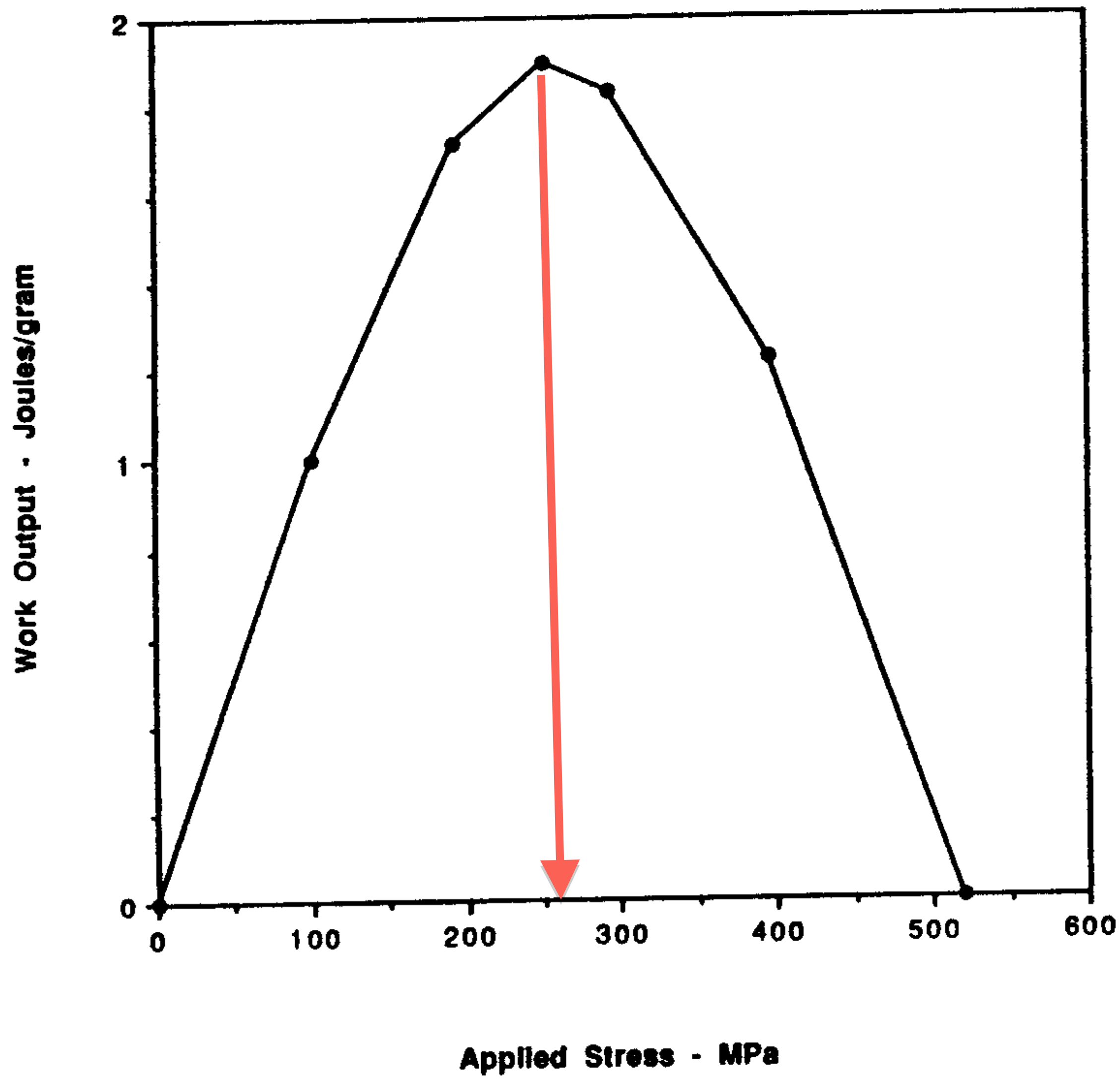

Figure 6: The work output of the same Ni-Ti-Fe alloy shown in Figure 4 is plotted against the opposing stress level, showing that the highest work output is obtained when working against a stress of 250 MPa.

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.

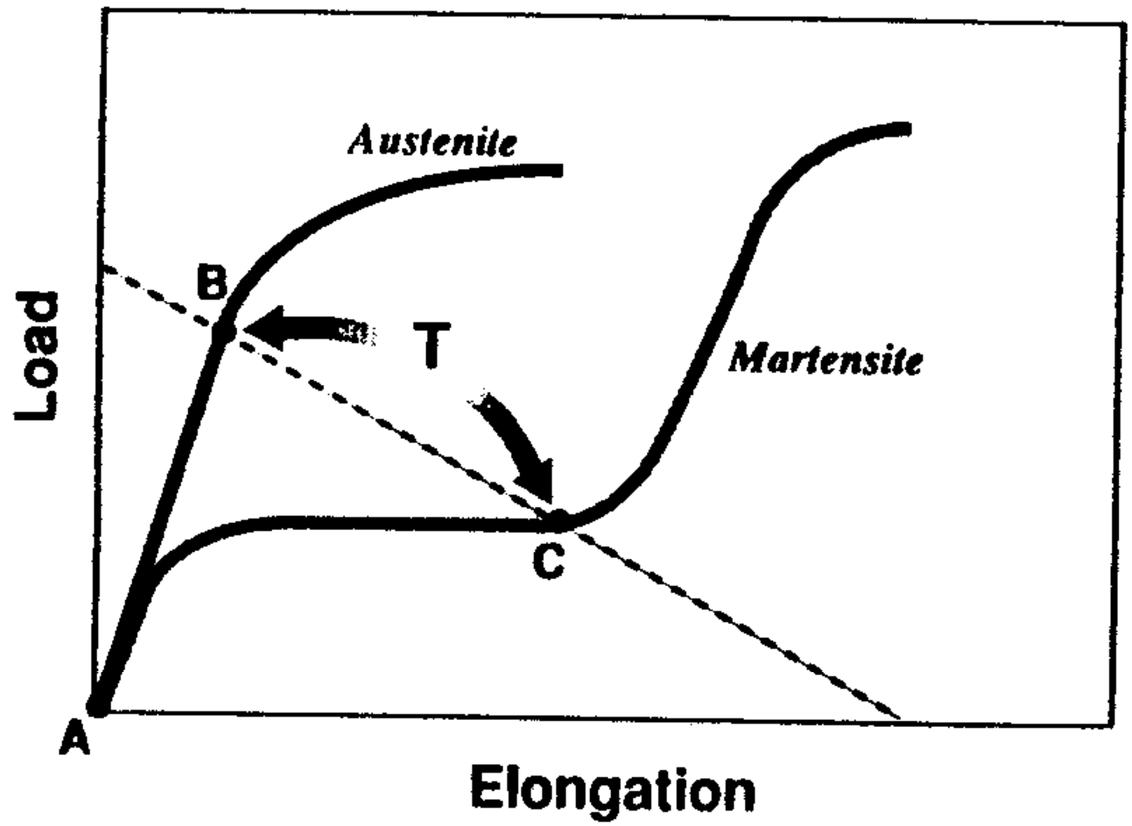

Figure 7: Figure 2 showed the actuation of a constant load, where as Figure 7 demonstrates recovery against a spring. In this case, motion ideally occurs between points B and C, the intersection points with the austenitic and martensitic stress-strain curves.

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 $\mathbf{A}_{\sigma_H}$ (A_s or A_f with an applied stress of σ_H) to slightly below $\mathbf{M}_{\sigma_L}$ (M_s or M_f with an applied stress of σ_L).

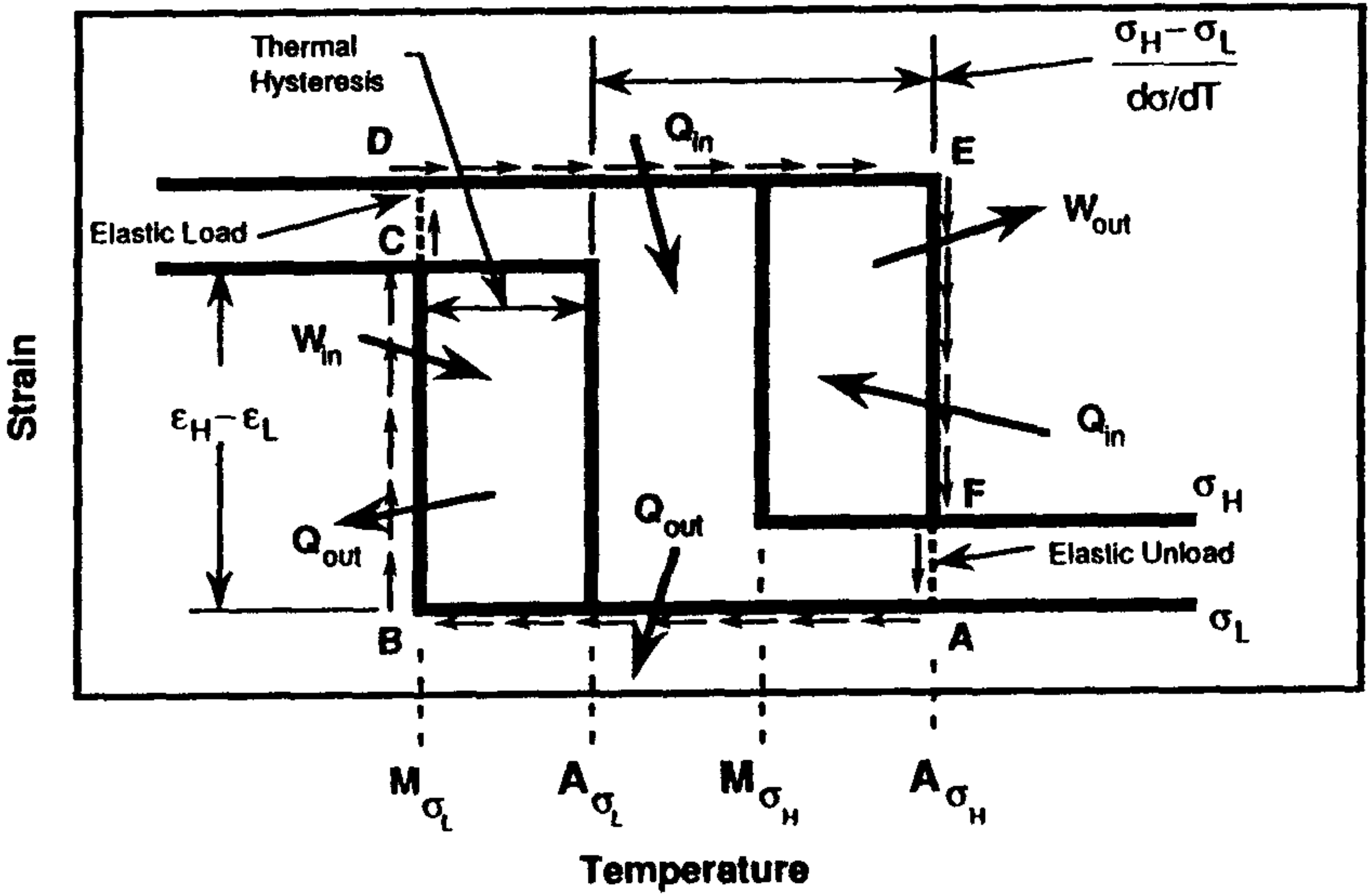

Figure 8: Two idealized strain-temperature recovery profiles, one at σ_L and the other at σ_H. A typical actuator would follow the path marked with arrows, from 'A' to 'F'. See text for details.

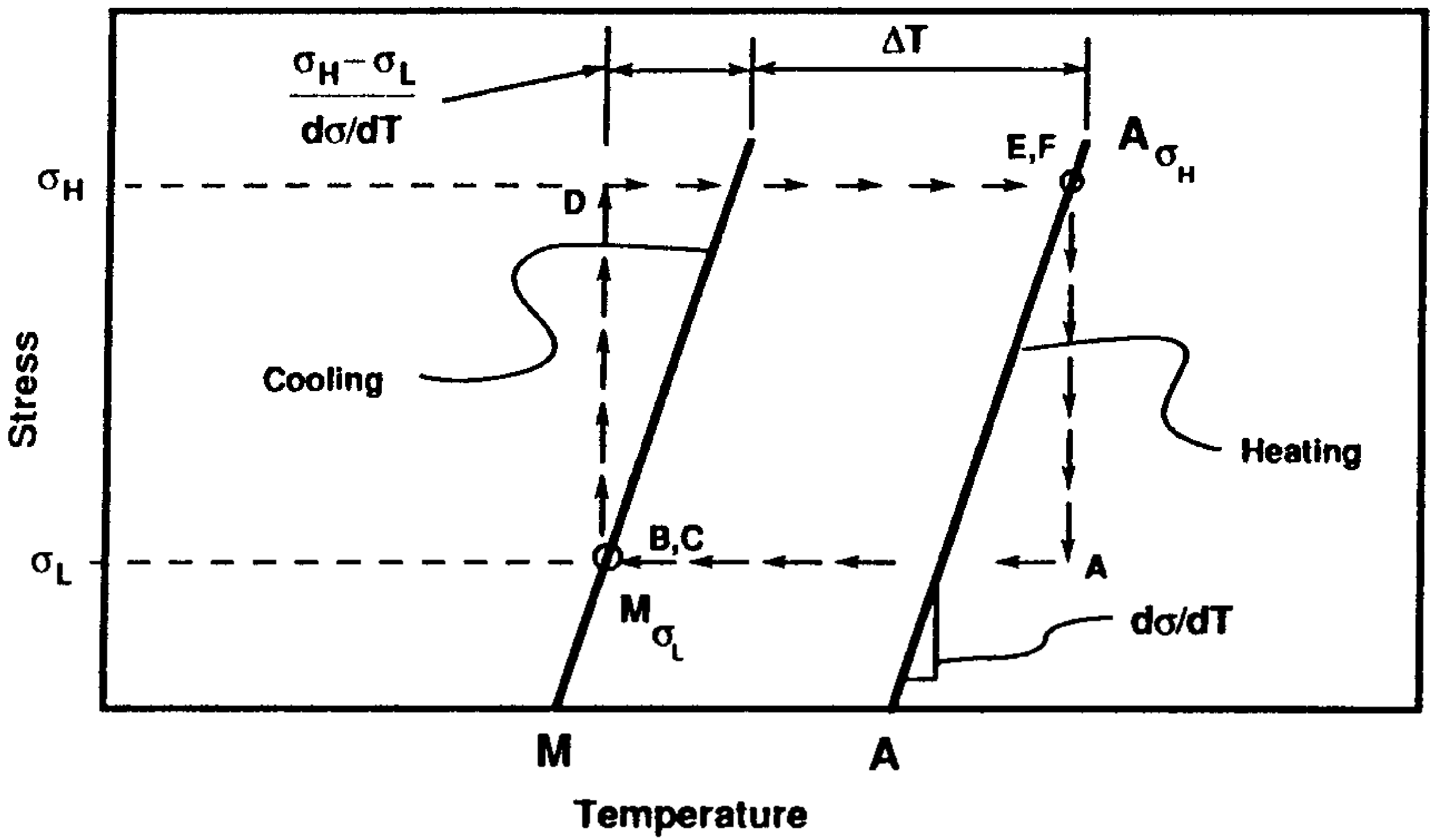

Figure 9: By "crossplotting" curves such as that shown in Figure 8, one arrives at the stress-temperature curve shown in Figure 9. Again the path of an actuator is shown by the arrows from 'A' to 'F'. See text for details.

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'.

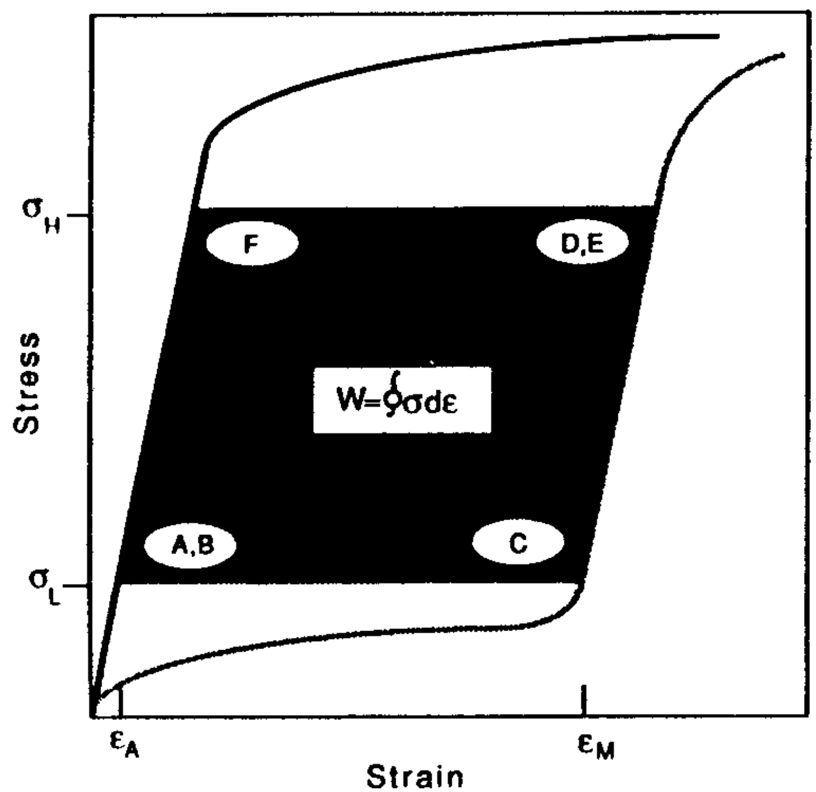

Figure 10: The actuator path described in Figures 8 and 9 are shown in terms of martensitic and austenitic stress-strain curves.

L to S

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:

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

Net work output is written as:

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

The heat input, Q_{in}, is the sum of the sensible heat needed to warm the metal from $\mathbf{M}_{\sigma_L}$ to $\mathbf{A}_{\sigma_H}$ 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 $\mathbf{M}_{\sigma_L}$ to $\mathbf{A}_{\sigma_H}$ is then:

$$Q_s = \rho C(\mathbf{A}_{\sigma_H} - \mathbf{M}_{\sigma_L}) = \rho C\left\{\Delta T + \frac{\sigma_H - \sigma_L}{d\sigma/dT}\right\} \tag{3}$$

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:

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

and then finally the thermal efficiency of an actuator becomes:

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

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 (ε_A-ε_M) improves efficiency (Figure 11), though clearly fatigue damage increases with increases in ε_A-ε_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

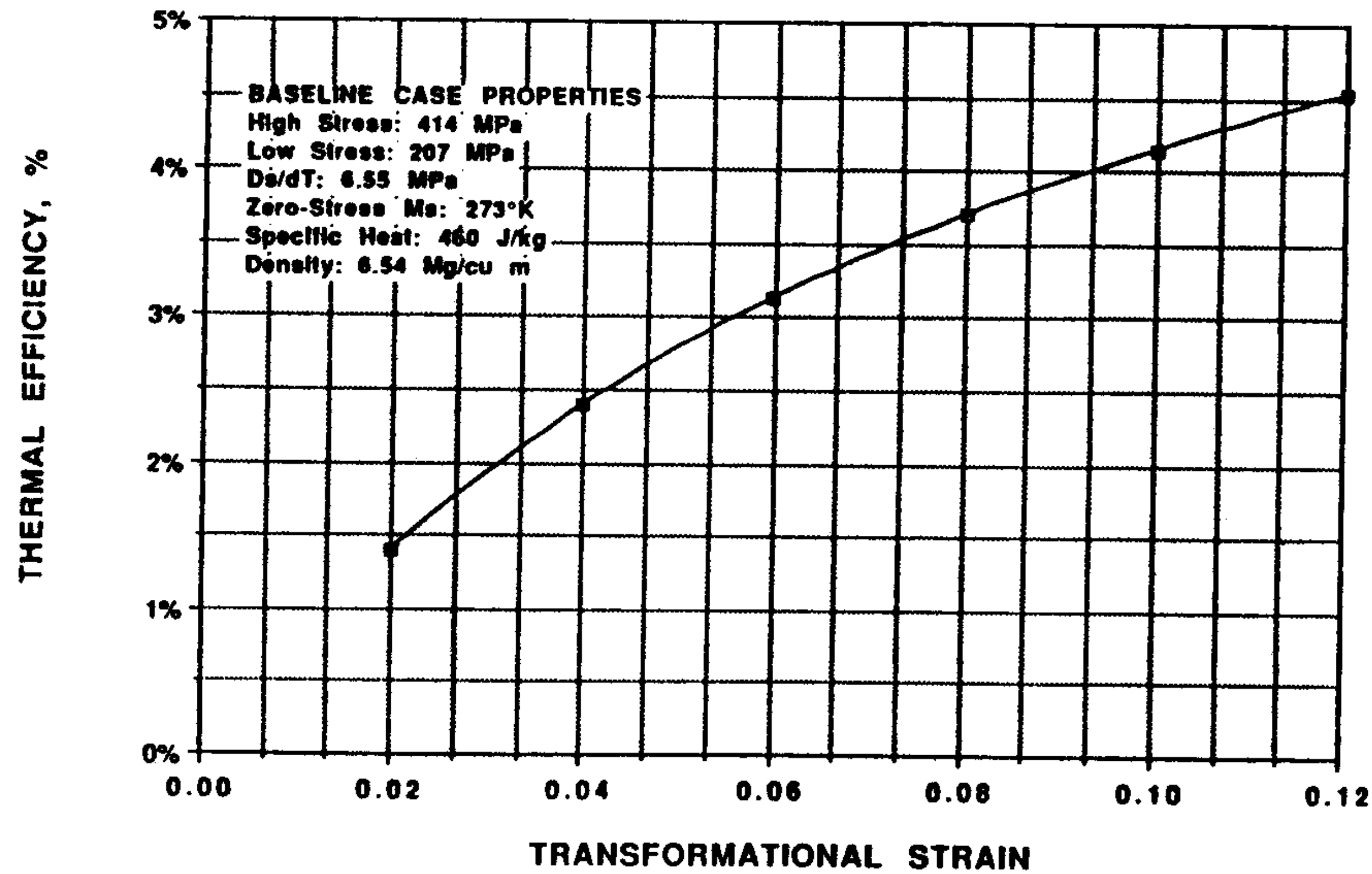

Figure 11: The thermal efficiency of an actuator increases dramatically with transformational strain. Unfortunately, fatigue considerations limit the practical strain values to under 4%.

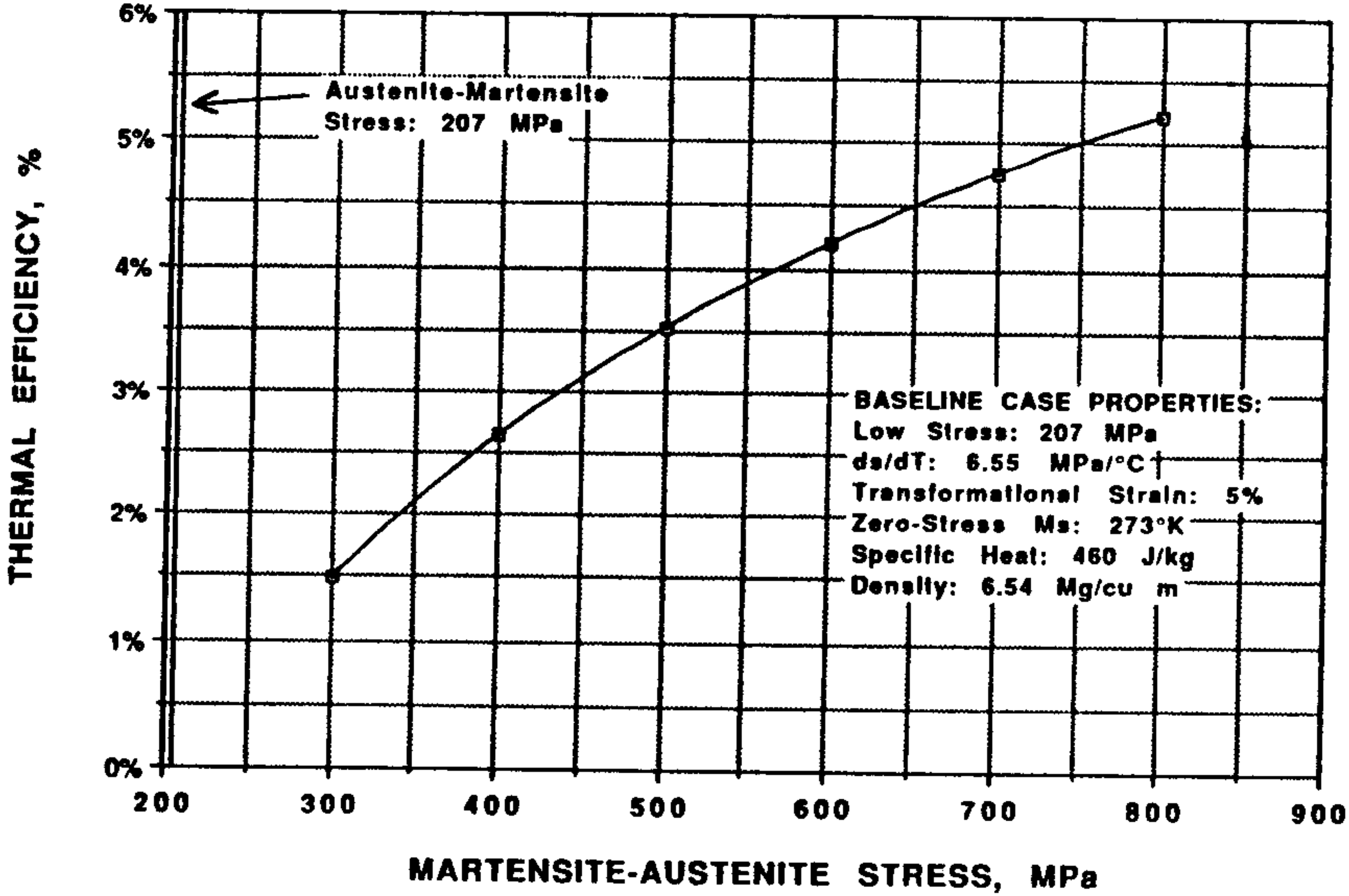

Figure 12: The importance of using alloys with high austenitic yield strengths is shown. Unfortunately there are limits that have to be placed on σ_H due to the effects shown in Figure 5, and fatigue limitations. 400 MPa appears now to be a very optimistic value when coupled with strains of 4%.

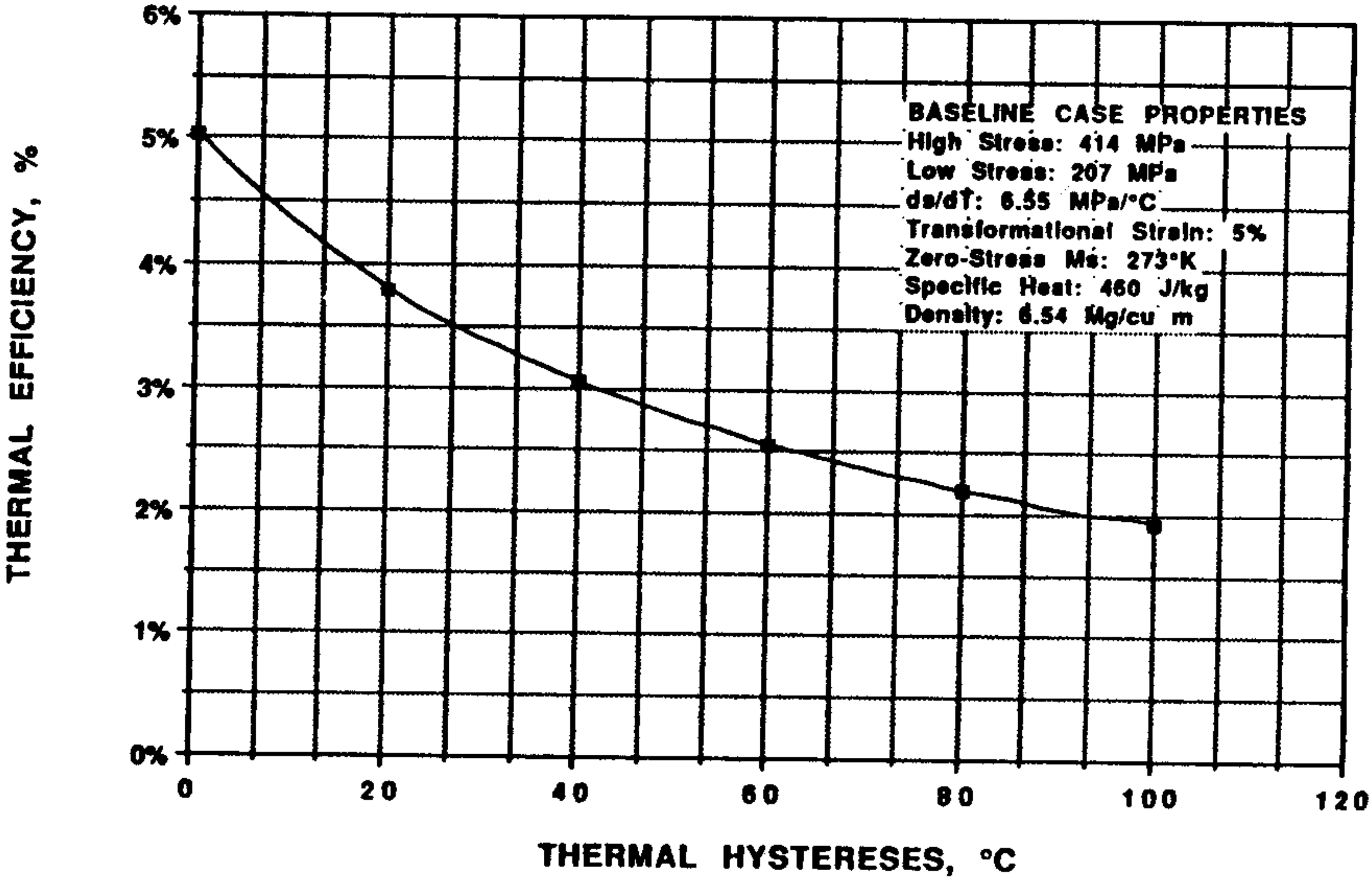

Figure 13: Thermal hysteresis decreases efficiency, providing hope for more efficient heat engines and actuators using Ni-Ti-Cu alloys, or by using only the R-phase,

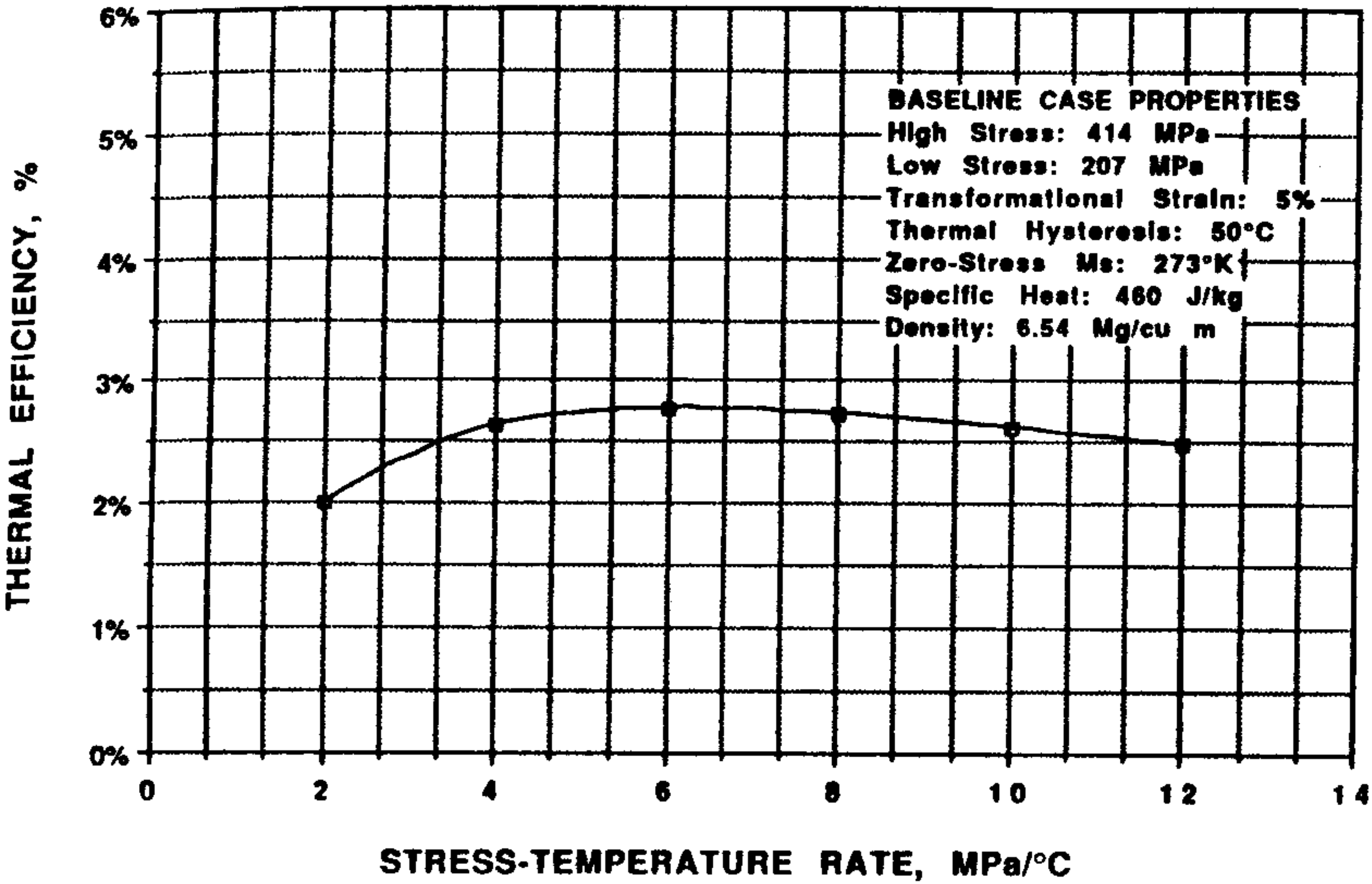

Figure 14: The effects of stress rate, dσ/dT, are unclear, depending upon what other properties are.

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, servo-motors, 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.