JOURNAL OF MATERIALS SCIENCE 19 (1984) 1827–1835

Shape memory heat engines

R. SALZBRENNER
Sandia National Laboratories, Albuquerque, NM 87185, USA

The mechanical shape memory effect associated with a thermoelastic martensitic transformation can be used to convert heat directly into mechanical work. Laboratory simulation of two types of heat engine cycles (Stirling and Ericsson) has been performed to measure the amount of work available per cycle in a Ni–45 at% Ti alloy. Tensile deformations at ambient temperature induced martensite, while a subsequent increase in temperature, caused a reversion to the parent phase during which a load was carried through the strain recovery, i.e. work was accomplished. The amount of heat necessary to carry the engines through a cycle was estimated from calorimeter measurements and the work performed per cycle. The measured efficiency of the system tested reached a maximum of 1.4% which was well below the theoretical (Carnot) maximum efficiency of 35.6%.

1. Introduction

A number of articles [1–9] have reported on the possibility of using the shape memory effect in energy conversion schemes. A prominent theme in these articles is the practicality of constructing a device which employs a shape memory material as the working substance to convert the heat extracted from a thermal reservoir into useful work. These studies have necessarily been concerned with the prediction of the maximum net work output, as well as the maximum efficiency that these heat engines might be expected to exhibit. The net work and efficiency have been predicted from a variety of theoretical arguments which rely on data from thermodynamic and mechanical property measurements. The predicted efficiencies range from 2 to 9%.

The present paper reports on the direct measurement of the net work available from two different types of work cycles performed with a nickel–45 at% titanium (Nitinol) alloy. Each of the work cycles employed four steps, two of which occurred under isothermal conditions. The net work available was measured as a function of the maximum strain the alloy was allowed to undergo during a particular cycle. The maximum efficiency possible from each work cycle was bracketed in a straightforward manner from the net work values and calorimetric measurements made on the Nitinol alloy.

2. Heat engines

An engine can be described by plotting the volume–pressure or the displacement–force changes which occur per engine cycle [10]. A generalized diagram of this type is shown in Fig. 1. The working medium is expanded along curve abc, and the area enclosed by abcnm represents the work done by the engine during the expansion process. The system is returned to its original state along curve cda, and during this portion of the cycle, the working medium is compressed. The area cdamn represents the work done on the engine during the compression cycle. The net work done by the engine per cycle is thus the area abcd (i.e. the crosshatched area in Fig. 1). The net work of an engine per cycle can be determined by measuring the pressure–volume, or force–displacement changes during a complete expansion–compression work cycle.

Heat is one of the least mechanically recoverable forms of energy [11]. Heat energy can be converted into work by a suitable type of "heat" engine, in which a working medium responds to heating (or cooling) by undergoing a length or volume change. The working medium is usually a fluid which is successively compressed, heated to a high temperature and expanded (doing work in the process), and finally cooled to the original temperature. Most practical heat engines employ a

0022–2461/84 $03.00 + .12

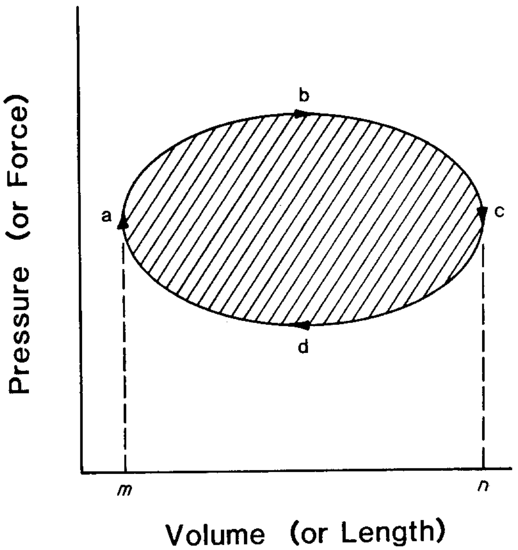

Figure 1 A (reversible) work cycle generated by an arbitrary working medium.

gas as the working fluid, since large displacements are necessary to reach reasonable efficiencies. The efficiency, e, of an engine is defined as the ratio of the net work performed per cycle, to the heat input per cycle. The maximum percentage of the total heat energy that can be converted into work (i.e. the maximum theoretical efficiency) is a function of the temperature change of the working medium before and after expansion. The maximum efficiency, e_{max}, is [10]:

$$e_{max} = \frac{T_H - T_L}{T_H} \quad (1)$$

where T_H is the highest temperature (in degrees absolute) reached during the cycle, and T_L is the lowest temperature in the cycle (generally ambient). This is the well known Carnot efficiency, which indicates the maximum possible efficiency a heat engine can attain, and is independent of the working substance or method of expansion/compression. Shape memory alloys have been considered as candidates for the working medium in heat engines because of the martensitic phase changes which occur during expansion and compression operations. The phase transformation may increase the actual efficiency towards the Carnot maximum;however, the measured efficiency cannot exceed the Carnot value.

There is no practical example of a Carnot-cycle engine, which requires two adiabatic and two isothermal steps per cycle. The Stirling and Ericsson cycles, however, are two reversible heat engine cycles which can be approximately incorporated in experimental designs [11]. Since they are theoretically reversible, they are equivalent to the Carnot cycle in efficiency. Each of these cycles employs four steps, two of which involve isothermal expansion or contraction. No adiabatic steps are included in either heat engine cycle. Fig. 2a illustrates the Stirling cycle. The medium is first compressed at T_L (Step I), the pressure is then increased at constant volume by raising the temperature to T_H, (Step II), then the medium is allowed to expand at T_H (Step III), and finally the temperature is lowered to T_L with a decrease in pressure. A distinguishing feature of the Stirling cycle is the two constant volume (or constant length) steps during which the pressure (or force) is changed as a result of an external change in temperature. The Ericsson cycle is illustrated in Fig. 2b. The first step is an isothermal compression (Step I), followed by an expansion at constant pressure caused by increasing the temperature (Step II), an isothermal expansion (Step III), and finally a reduction in temperature at constant pressure with a concomitant volume contraction. The Ericsson cycle is characterized by two constant pressure (or force) steps during which the volume (or length) is varied by changes in temperature.

3. Laboratory simulation of NiTi heat engines

Stirling and Ericsson cycles can be performed on shape memory materials by employing a mechanical testing system which has the capability of controlling strain as well as load. Engaging strain control (specifically, holding strain constant during appropriate portions of the test) enables a Stirling cycle to be coursed as shown in Fig. 3a. Alternately, load control can be employed to hold force constant as the temperature is changed, resulting in an Ericsson cycle as depicted in Fig. 3b. Fig. 3 shows the direction of each cycle, including the isothermal length/load change steps, and the temperature change steps during which length or load is held constant. The primary difference between the pressure–volume diagrams of Fig. 2 (schematically shown for gas), and the force–length diagrams of Fig. 3 (for a shape memory material) is the direction of response of the working substance. Instead of a hydrostatically applied pressure on the gas, an axial (tensile)

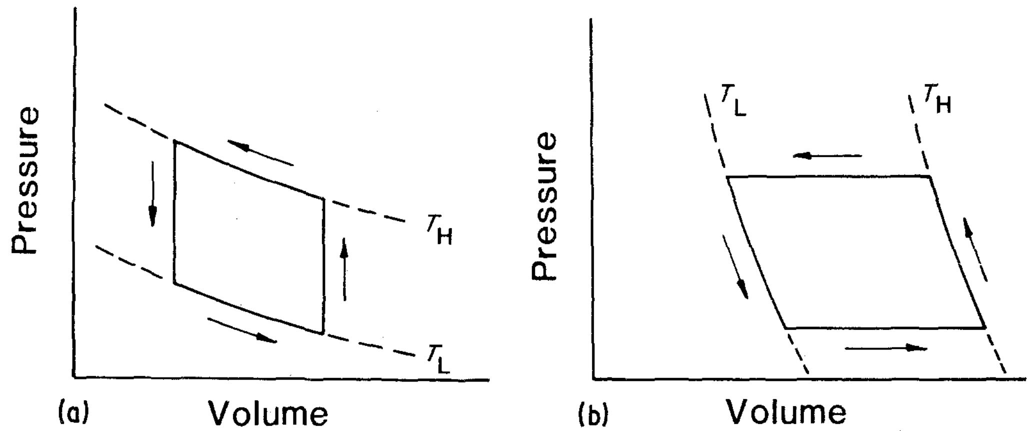

Figure 2 Schematic diagrams for work cycles which include isothermal expansions (and contractions); (a) Stirling cycle and (b) Ericsson cycle.

load applied to the shape memory alloy causes an increase in length due primarily to the martensitic transformation. An increase in temperature causes the martensite to begin to revert to the parent phase which in turn results in either an increase in force or a decrease in length. The martensitic transformation and reversion are capable of producing relatively large changes in either length or force as the temperature is changed, thus allowing the shape memory alloy to do a substantial amount of work. The area enclosed by the force–displacement curve provides a measure of the net work done per cycle.

4. Experimental measurement of net work for NiTi heat engines

The sample employed as the working medium was a Ni–45 at% Ti tensile specimen with a 2.5 cm (1.0 inch) gauge section and a 0.13 cm^2 (0.02 inch^2) rectangular cross-section. This particular batch of NiTi underwent the martensitic transformation on cooling (with no applied stress) starting at about 19° C (M_s). The reverse martensite-to-parent transformation on heating was completed at about 38° C (A_f). The transformation temperatures were determined from differential scanning calorimeter curves. The sample

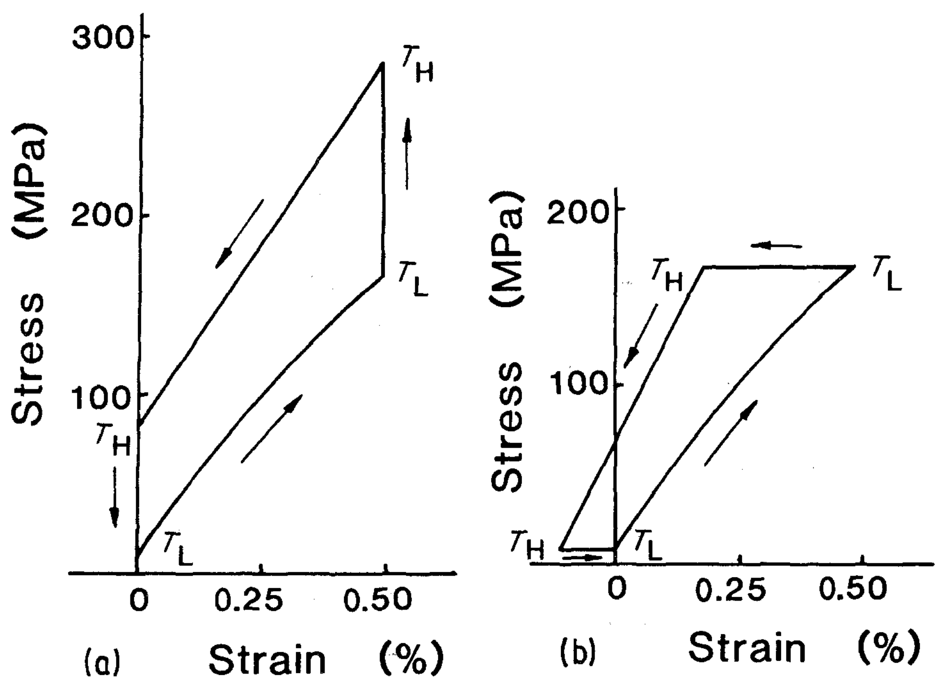

Figure 3 Work cycle conducted on NiTi using (a) strain control (Stirling cycle), and (b) load control (Ericsson cycle).

was mounted in an MTS* electrohydraulic tensile frame, and tests were conducted using load control and strain control. The sample was loaded (strained) at a constant rate and ambient temperature while the load and strain were simultaneously measured. Strain rates of approximately 10^{-3} sec^{-1} were used throughout these tests.

The loading of the sample at a temperature a few degrees above M_s caused the NiTi alloy to undergo a stress-assisted martensitic transformation [6]. The shape change associated with the transformation provided the plastic-like deformation behaviour. At a desired strain, the machine applied loading ramp was stopped. The sample gauge section was then heated to the desired temperature with a hot air gun. The temperature of the gauge section was monitored by a thermocouple welded to the sample. Care was taken to heat only the sample, and not the extensiometer (a 5 cm long arm extensiometer was used for these tests). Raising the temperature of the deformed NiTi sample imparted the driving force for reversion to the crystal structure of the parent phase. For tests in which load control was employed, the strain was allowed to vary (decrease) as the temperature was increased, that is strain was recovered by the martensite-to-parent reversion. Under strain control, the load increased in response to the thermally induced reverse transformation. After the sample was stabilized at a particular temperature (ranging from 60 to 190° C), it was unloaded either to zero load (for load control) or to zero strain (for strain control). If the sample had not undergone complete reversion to the parent phase on heating during the load or strain control step, then the reverse transformation of martensite-to-parent continued as the sample was unloaded while still at the elevated temperature. The cycle was completed by returning to ambient temperature under either load or strain control. Actual examples of the strain control, or Stirling, cycle and the load control, or Ericsson, cycle are shown in Figs. 4 and 5, along with a summary of the changes which occurred during each step. Included in the summary are the direction and cause of the heat flow with respect to the NiTi working volume. The heat flow effects are discussed in more detail in the following section.

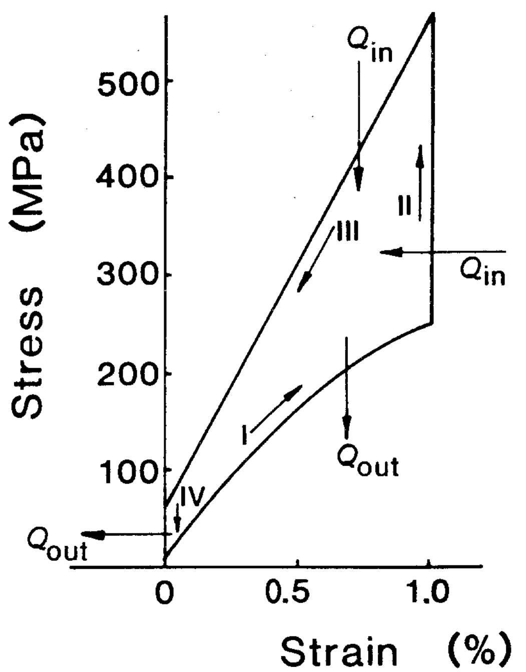

Figure 4 Schematic descriptions of the work cycle for the Stirling cycle. Step I: change σ, ϵ; hold T constant. Q_{out} because of σ-assisted P → M transformation. Step II: change T, σ; hold ϵ constant. Q_{in} to raise T and drive M → P reverse transformation. Step III: change σ, ϵ; hold T constant. Q_{in} to hold T constant as M → P transformation occurs. Step IV: change T, σ; hold ϵ constant. Q_{out} to lower T and from P → M transformation.

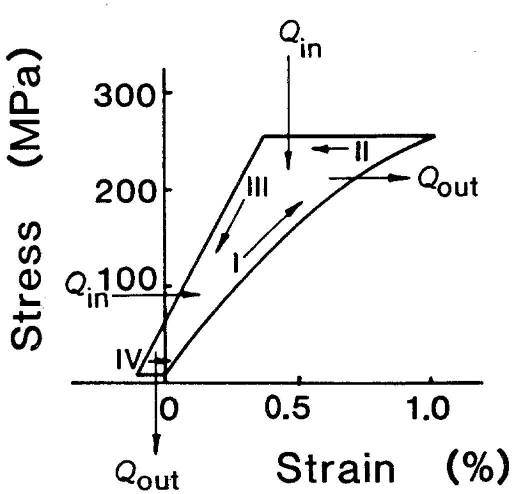

Figure 5 Schematic description of the work cycle for the Ericsson cycle. Step I: change σ, ϵ; hold T constant. Q_{out} due to P → M transformation. Step II: change T, ϵ; hold σ constant. Q_{in} to raise T and drive M → P reverse transformation. Step III: change σ, ϵ; hold T constant. Q_{in} to hold T constant as M → P transformation occurs. Step IV: change T, ϵ; hold σ constant. Q_{out} to lower T and from P → M transformation.

*MTS Systems, Inc.

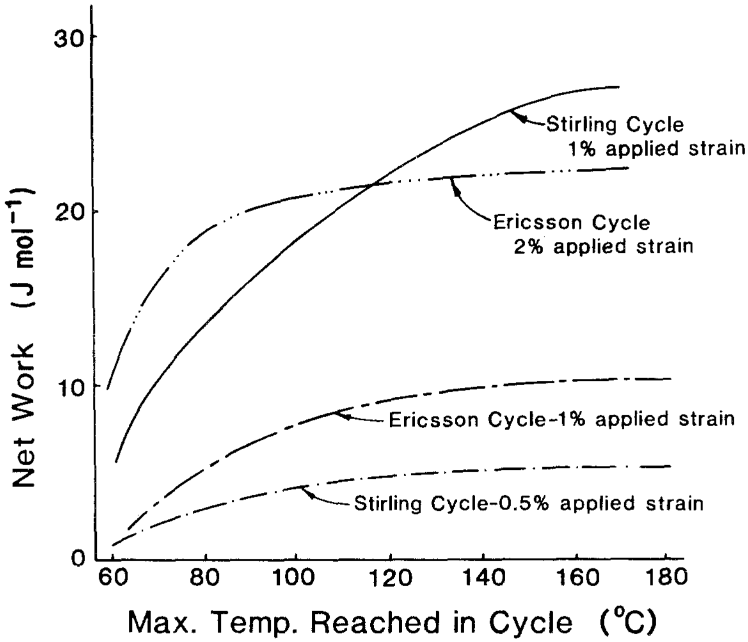

Figure 6 Measured net work increases as a function of maximum temperature up to a maximum value which is governed by the applied strain.

The net work performed per cycle increased with increases in the maximum temperature reached during a cycle as shown in Fig. 6. This was due to the increased driving force for reversion of the martensite to the parent phase as the temperature was increased. The temperature necessary to completely revert the sample to the parent phase was dependent on the maximum strain (or load) applied to the sample. High strains (and/or loads) caused a greater amount of stress-induced martensite to be formed, which in turn required a larger thermal driving force to fully revert the sample back to the parent phase. The measured net work per cycle increased until the sample was transformed completely during the constant strain step or constant load step.

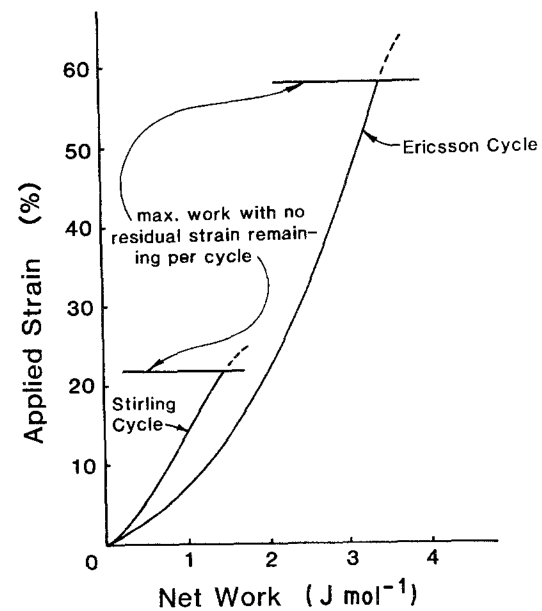

Figure 7 Net work measured as a function of applied strain for Stirling and Ericsson heat engine cycles with NiTi as the working medium.

Fig. 7 shows the increase in measured net work per cycle with increasing strain applied during the isothermal loading step (Step I). The data in this figure apply to cycles in which a maximum amount of reversion (while under load) on heating (during Step II) was obtained. The low temperature portion of the cycle was room temperature (25° C), while the upper temperature reached was 190° C. The maximum amount of work that could be performed per cycle without non-recoverable plastic deformation was 58.2 J mol^{-1} for the Ericsson cycle and 21.8 J mol^{-1} for the Stirling cycle. The working volume for the tensile specimen which produced these values was 0.33 cm^3 (0.020 $inch^3$). The maximum net work for both types of heat engines was limited by the point where recurring, non-recoverable strain was accumulated after each cycle. If a stress of approximately 685 MPa (100 ksi) was exceeded during either the Ericsson of the Stirling cycles, not all of the applied strain was recoverable. The accumulation of non-recoverable plastic strain (by applying an excessive load) caused the sample to fail after a few cycles. The strain control (Stirling) cycle was capable of

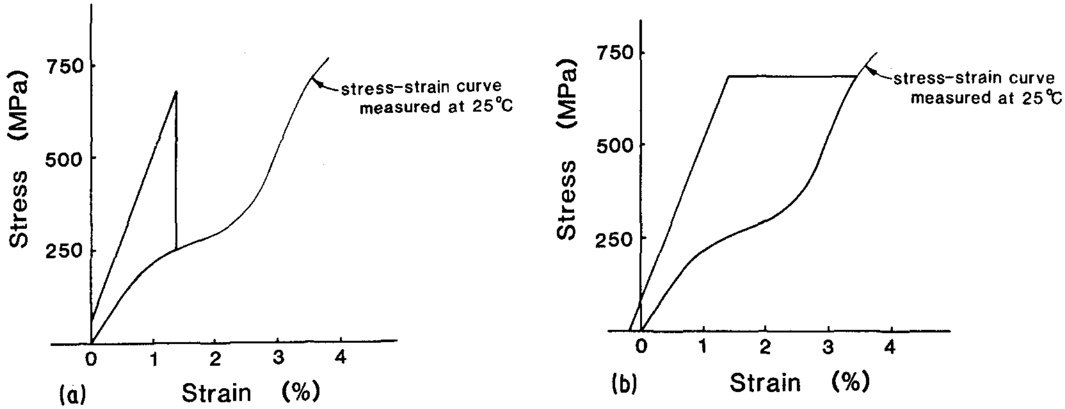

Figure 8 Maximum work which can be performed by the NiTi is limited by the stress above which accumulated plastic damage occurs; (a) the Stirling cycle, and (b) the Ericsson cycle.

producing more work than the load control cycle up to about 1.5% strain, since a greater load can be carried through the same distance. Fig. 8 schematically depicts the maximum work available from each type of cycle. The load control (Ericsson) cycle was capable of performing a greater amount of work than that possible with strain control, when the applied strain was in the 2.2 to 3.5% range.

5. Heat flow and efficiency

The primary purpose of a heat engine is to provide a means of converting a quantity of heat directly into work. Measurements of the heat available to be converted, as well as the amount of work actually done are fundamental to describing the effectiveness of such an engine. The previous section described the measurement of the net work performed per cycle for Ericsson and Stirling-type engines in which the work was performed by a uniaxial (tensile) deformation and the subsequent recovery of the NiTi shape memory alloy. The heat available for conversion by the working volume of the tensile specimen was not measured directly during the Stirling or Ericsson cycles, but it nonetheless must be evaluated in order to calculate engine efficiency. The heat available for conversion into work was determined from calorimetry measurements, which are detailed later in this section. The efficiency, e, of a heat engine was defined earlier as the ratio of net work done by the engine per cycle, W_{net}, to the heat taken in per cycle, Q_{in}, and thus can be expressed as:

$$e = \frac{W_{\text{net}}}{Q_{\text{in}}} \qquad (2)$$

A qualitative description of the heat flow with respect to the NiTi specimen for Steps I through IV in each heat engine cycle is shown in Figs. 4 and 5. Heat-flow-in occurred during Steps II and III in each cycle, and was necessary to raise the sample temperature, and to compensate for the endothermic martensite-to-parent transformation. The quantity of heat necessary for changing the temperature of a NiTi sample (under no applied load) was measured with a differential scanning calorimeter (Perkin-Elmer Model DSC II). The plots of the heat flow rate against temperature are shown in Fig. 9 for both heating and cooling of an unconstrained sample. The calorimetry curve produced by heating the sample provided values that were necessary to estimate the heat-in during an engine cycle. The calorimeter curve on heating generated values for the specific heat and the transformation enthalpy, from which the heat-in necessary to raise the temperature from 25 to 190° C of an unstressed sample was calculated. However, specific heat and transformation enthalpy can be affected by an applied stress. The effect of stress on the specific heat is small and will be neglected [9]. Stress has a larger influence on the transformation enthalpy. The measured magnitude of the enthalpy from a thermoelastic martensitic transformation has been shown to decrease in response to external constraints. The decrease (in magnitude) can be thought of in terms of transferring transformation enthalpy (heat) directly into stored strain energy [12] some of which can be used to produce work. Alternately, the difference between the enthalpy measured for an unrestrained sample and the enthalpy of a stressed transformation can be estimated as the

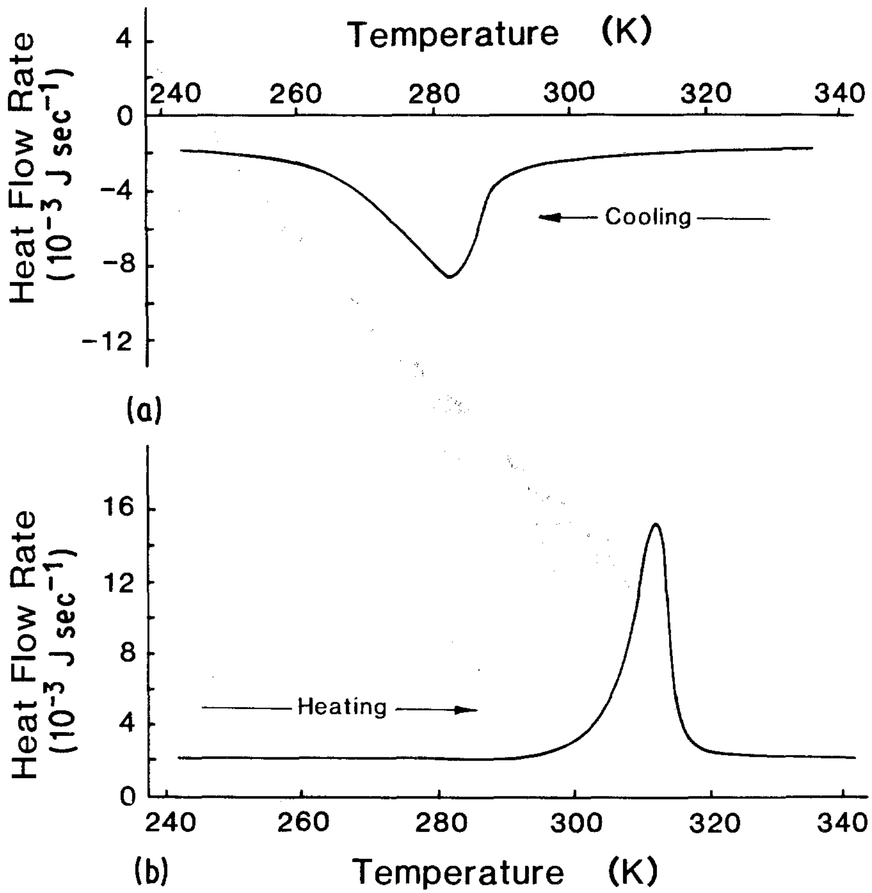

Figure 9 Differential scanning calorimeter curves for a NiTi sample (48.3 mg) on (a) cooling, and (b) heating.

useful work performed during the transformation [3, 9, 13]. Thus,

$$\Delta H^{\sigma=0} - \Delta H^{\sigma} \cong W_{net} \quad (3)$$

where ΔH is the transformation enthalpy from the martensite-to-parent transformation, and the superscript σ refers to whether or not an external stress was applied during the transformation. Note that this is only an approximation for calculation purposes. It will be shown that for the particular heat engine described here that the transformation enthalpy does not substantively change the efficiency. The total heat-in per engine cycle can be approximated as:

$$\begin{aligned} Q_{in} &\cong \Delta H^{\sigma} + C_p(T_H - T_L) \\ &= \Delta H^{\sigma=0} - W_{net} + C_p(T_H - T_L) \quad (4) \end{aligned}$$

if ΔH^{σ} is independent of temperature, and C_p (parent) is equal to C_p (martensite). The calorimeter studies of NiTi have verified these assumptions and yielded values of $\Delta H^{\sigma=0}$ and C_p (heat capacity) as $1190\,J\,mol^{-1}$ and $25.1\,J\,mol^{-1}\,K^{-1}$, respectively. The maximum value of W_{net} reported in Fig. 6 was $58.2\,J\,mol^{-1}$ for the 2.17 g NiTi working mass (i.e. $0.33\,cm^3$ working volume and a density of $6.6\,gm\,cm^{-3}$). Equation 4 was used to calculate the heat-in per cycle as $5270\,J\,mol^{-1}$. The measured efficiency was determined as 1.10% by using Equation 2.

The calculated efficiency is dominated by the value of C_p (i.e. $25.1\,J\,mol^{-1}\,K^{-1}$) times the change in temperature per cycle if ΔT is greater than a few tens of degrees. The quantity $C_p\Delta T$ (for cycles between 25 and 190° C) is equal to $4140\,J\,mol^{-1}$, which compared to the magnitude of $\Delta H^{\sigma=0}$ (which is $1190\,J\,mol^{-1}$) is significantly larger. Since the transformation enthalpy under stress is decreased compared to the stress-free condition, the maximum efficiency can be overestimated by entirely disregarding the contribution of the transformation enthalpy to the heat-in. Thus,

$$e < 1.4\% = \frac{W_{net}}{C_p\Delta T} \quad (5)$$

where values for W_{net}, C_p, and ΔT were previously stated. By parallel reasoning, the efficiency will be underestimated if the full value of the stress-free transformation enthalpy is used in the calculation. Using the values from above,

$$e > 1.1\% = \frac{W_{\text{net}}}{C_p \Delta T + \Delta H^{\sigma=0}} \qquad (6)$$

Combining the results from equations 4 and 5,

$$1.1\% < e < 1.4\% \qquad (7)$$

for the case of the tensile Ericsson cycle between 25 and 190° C, which provided the maximum net work for this study.

Assume for the sake of argument, that the same net work could be obtained with a change in temperature of 75 K. Using the same arguments as above to overestimate the actual efficiency, the efficiency would be less than 3.1% if ΔT is 75 K and the values of W_{net} and C_p are the same as those used previously. The efficiency is still low even though a value of W_{net} which is much too high is used (extensive measurements of W_{net} using cycle temperatures of 25 and 100° C have not been made, but the maximum value of W_{net} was less than 33 J mol^{-1}). The important conclusion is that the net work is small in comparison to the heat-in required per cycle whether or not the transformation enthalpy is considered in the calculations; the efficiency is thus constrained to be a small number.

The upper and lower temperatures which were used to determine W_{net} (25 and 190° C), were used to calculate the Carnot efficiency (from Equation 1) as 35.6%. The actual efficiency for the tensile shape memory engines was indeed small in comparison with the maximum possible efficiency found from the Carnot cycle.

6. Discussion

The actual efficiency for the simulated heat engine was low in comparison with the maximum theoretical efficiency. The net work that can be accomplished is limited, since the distance the force can be moved through is restricted to the recoverable strain, which is a few percent at most (~ 1 to 8). The net work for the shape memory material is, however, several times greater than that available from an alloy which does not exhibit a shape memory effect. Without the shape memory effect, the distance through which the load can be carried is a small fraction of that which can be employed by a shape memory alloy such as NiTi.

The transformation required to accomplish the shape memory effect increases the amount of heat-in which is necessary to raise the temperature because of the enthalpy of transformation. The enthalpy associated with the (stress free) transformation is about thirty percent of the heat-in required from the heat capacity term. The heat-in/cycle required for a heat engine using a non-shape memory alloy may therefore be smaller than that required for an engine using the shape memory effect. However, the net work would also drop by at least two orders of magnitude. The net effect on efficiency would be to decrease it by an order of magnitude or more. Although the efficiency of the shape memory material (used as the working medium) is quite low when compared with the Carnot efficiency, the net effect of employing the transformation to perform work is to increase the efficiency drastically above that available from a non-shape memory alloy.

7. Conclusions

The laboratory simulation of two types of heat engines was described in which a NiTi specimen was used as the working medium. The net work available per cycle was measured between room temperature and 190° C as a function of applied strain. The maximum work available for the two types of cycles was restricted by the load at which non-recoverable plastic deformation occurred. The maximum amount of work available from this NiTi sample, between 25 and 190° C, was measured as 58.2 J mol^{-1} per cycle.

Calorimetry measurements were made on the NiTi alloy, and provided values for the heat capacity and the transformation enthalpy (for the martensite-to-parent transition) of an unstressed sample. These quantities were found to be 25.1 J mol^{-1} K^{-1} and 1190 J mol^{-1}, respectively. The magnitude of the transformation enthalpy was approximated as being decreased by the amount of work done, that is, the transformation can be thought of as providing the mechanism for transforming heat into mechanical work. The heat-in per heat-engine cycle was thus estimated from the heat-in for unstressed material (known from calorimetry results) minus the work done (i.e. the heat converted into work). Through measurement of work performed and calorimetric parameters the efficiency of the heat engines which used a NiTi tensile specimen as the working substance was between 1.1 and 1.4%. The measured efficiency was well below the theoretical maximum, or Carnot, efficiency of 35.6% for heat engines operating between 25 and 190° C. Hence, the practicality of using an engine based on the tensile behaviour of NiTi is at best limited to specialized applications.

Other sample orientations, or other types of deformations besides tensile, may produce larger recoverable deflections and change the working loads. This will in turn change the net work output. Similarly, changes in alloy composition (to change the temperature, and stress assisted transformation range), and/or changes in the operating temperatures of the cycle will affect the net work per cycle, as well as the efficiency. The methods described in this article to measure the net work available per cycle can be applied to engines which employ different types of deformation, temperature ranges, or alloy compositions.

Acknowledgement
This work was supported by the US Department of Energy under contract No. DE-AC04-76 DP00789.

References

1. M. AHLERS, *Scripta Metall.* 9 (1975) 71.
2. A. A. GOLESTANEH, *J. Appl. Phys.* **49** (1978) 1241.
3. B. M. MOGUTNOV, *Fiz. Metal. Metalloved.* **45** (1978) 428.
4. H. A. MOHAMED, *J. Mater. Sci.* **13** (1978) 1364.
5. *Idem, ibid.* **13** (1978) 2729.
6. *Idem, ibid.* **14** (1979) 1339.
7. H. A. MOHAMED and J. WASHBURN, *ibid.* **12** (1977) 469.
8. K. MUKHERJEE, *Scripta Metall.* **14** (1980) 405.
9. P. WOLLANTS, J. R. ROOS and L. DELAEY, *ibid.* **14** (1980) 1217.
10. D. HALLIDAY and R. RESNICK, "Physics" Parts I and II, (John Wiley & Sons, Inc., New York, 1966).
11. R. U. AYRES and R. P. McKENNA, "Alternatives to the Internal Combustion Engine", (Johns Hopkins University Press, Baltimore and London, 1973).
12. R. J. SALZBRENNER and M. COHEN, *Acta Metall.* 27 (1979) 739.
13. H. B. CALLEN, "Thermodynamics" (John Wiley & Sons, Inc., New York, 1966) p. 111.

Received 19 October 1982
and accepted 21 September 1983