MODELLING THE CONSTITUTIVE BEHAVIOUR OF MARTENSITE AND AUSTENITE IN SHAPE MEMORY ALLOYS USING CLOSED - FORM ANALYTICAL CONTINUOUS EQUATIONS

Arathi Pai[*], **Thomas Niendorf**[**], **Phillip Kroos**[**], **Isabel Koke**[†], **Ansgar Traechtler**[*,†], **Mirko Schaper**[**]

[*]Fraunhofer Institute for Production Technology, Controls Engineering
Zukunftsmeile 1, 33102 Paderborn, Germany
arathi.pai@ipt.fraunhofer.de, ansgar.traechtler@ipt.fraunhofer.de

[**]Department of Materials Science, University of Paderborn
Warburger Strasse 100, 33098 Paderborn, Germany
Thomas.Niendorf@iwt.tu-freiberg.de, krooss@lwk.upb.de, schaper@lwk.upb.de

[†]Heinz Nixdorf Institute, Controls Engineering and Mechatronics
Fürstenallee 11, 33102 Paderborn, Germany
isabel.koke@uni-paderborn.de, ansgar.traechtler@ipt.fraunhofer.de

Summary: *This paper presents a phenomenological constitutive model for martensite and austenite in SMAs with a focus on computational efficiency, implementational simplicity and the model's eventual use in a control strategy. The model is built around the mathematical description of the typical form present in stress-strain curves in SMAs using closed form, continuous, differentiable equations and continuity conditions imposed when the strain changes direction. The parameters are few, physical and easy to identify from a monotonic tensile experiment and a simple parameter identification process. The model is able to predict the behavior of martensite and austenite when exposed to both monotonic as well as cyclic loading and unloading. Tensile iso-thermal experiments are then performed for validation and the model predictions are shown to be in good agreement with experimental data. Since the model is based entirely on continuous analytical equations, it is extremely computationally efficient and, thus, can be used as a basis towards the development of model-based control algorithms.*

1. INTRODUCTION

Shape Memory Alloys (SMAs) display two kinds of shape recovery effects: the one-way-shape-memory-effect and pseudoelasticity [1]-[5], both of which can be exploited for various applications such as switches, latches, stents, orthodontic braces, vibration damping etc. [6],

[7]. These effects arise due to phase transformations between two microscopic SMA phases: a parent austenitic phase and a martensitic phase. Characterization of both phases is done on the basis of the phase transformation temperatures: M_s, M_f (martensite start and finish) and A_s, A_f (austenite start and finish), where, generally, $M_f < M_s < A_s < A_f$. At temperatures below M_f, where the alloy is initially fully martensitic, mechanical loading causes a macroscopic shape change imposed by variant selection and detwinning, respectively. This shape is stable until the martensite is heated above A_f, triggering a transformation to austenite and a macroscopic shape recovery. This thermally induced shape recovery is the one-way-shape-memory-effect. Pseudoelasticity, also coined superelasticity occurs for temperatures above A_f, where the alloy is initially fully austenitic. Mechanical loading above certain critical stresses causes a transformation to martensite and a concurrent macroscopic shape change. Removal of the driving stress, triggers an (almost) complete reverse transformation from martensite to austenite, and a respective shape recovery.

Due to these effects, SMAs are attractive materials for a variety of applications and are particularly promising in the creation of compact and powerful actuators due to their high energy density and high specific actuation stresses. Additionally, they offer various other advantages such as high reversible strains, smooth and silent actuation, scalability (down to micrometers) etc. [7], [6]. SMAs, however, are characterized by extreme non-linear and hysteretic behaviour and therefore designing actuators for sophisticated applications, e.g. position control, is non-trivial. One of the approaches used to facilitate the development of control algorithms for SMA actuators is model-based design, where mathematical models that predict the non-linearities are directly included online in control strategies. In the past 20 years, models to describe SMA behaviour have been developed from various perspectives: Thermodynamics [8]-[14], phenomenological and thermomechanical [15]-[23], micromechanical [24]-[26], Finite element [27],[28], constitutive [29], [30] etc.

While the aforementioned models are successful to various degrees in describing SMA behaviour from various domains, there is a divergence between model accuracy and computational efficiency i.e. the models that are accurate are computationally expensive and although some numerical solutions exist (e.g. [29]), they are still not conducive to be used in control loops. On the other hand, the models that are computationally inexpensive are often too elementary and their implementation, while straightforward is error-prone because of the use of series of conditional statements (e.g. [23]).

This paper presents a phenomenological constitutive model in section 2 for martensite and austenite in SMAs with a focus on computational efficiency, implementational simplicity and the model's eventual use in a control strategy. The model is built around the mathematical description of the typical form present in stress-strain curves in SMAs using closed form, continuous, differentiable equations and continuity conditions imposed when the strain changes direction. The parameters are few, physical and easy to identify from a monotonic tensile experiment and a simple parameter identification process. The model is able to predict the behavior of martensite and austenite when exposed to both monotonic as well as cyclic loading and unloading. Tensile iso-thermal experiments are then performed in section 3 for validation and the
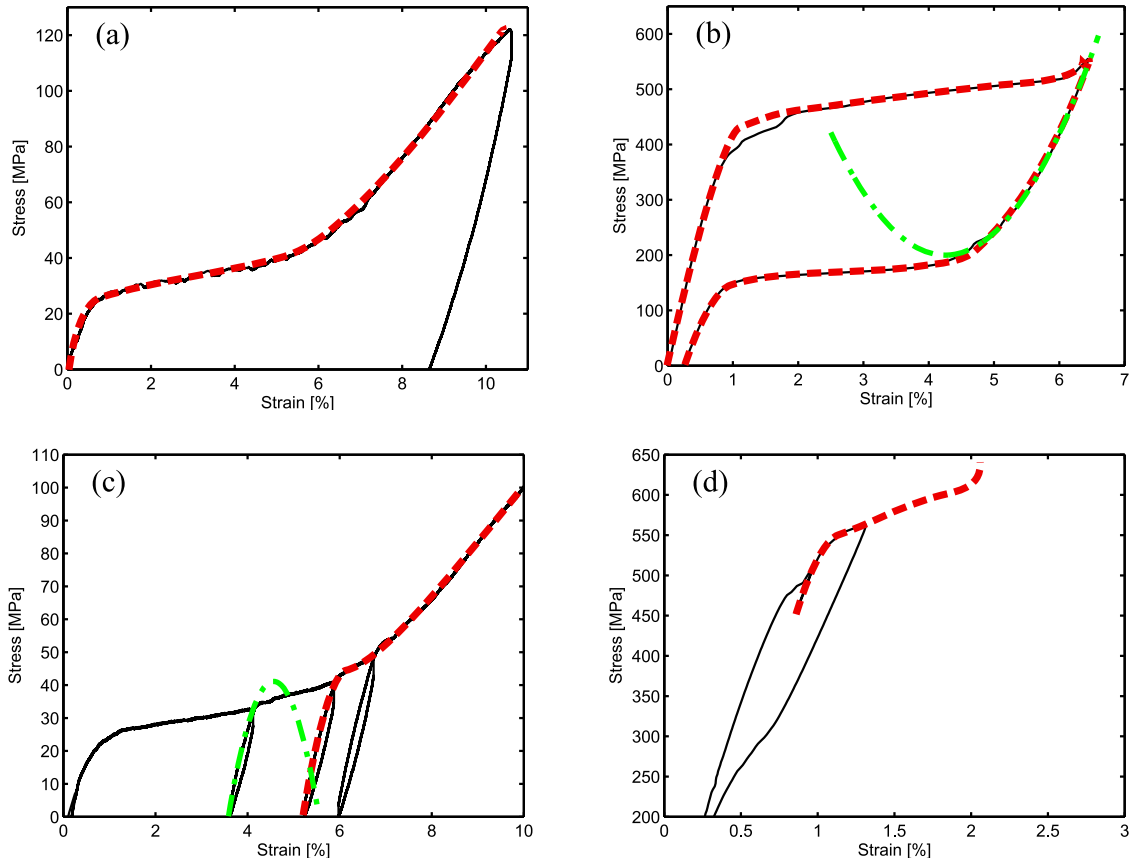
Figure 1. Typical stress-strain curves for martensite and austenite: monotonic loading (plots (a) and (b)), cyclical loading (plots (c) and (d)) [31]

model predictions are shown to be in good agreement with experimental data. Since the model is based entirely on continuous analytical equations, it is extremely computationally efficient and, thus, can be used as a basis towards the development of model-based control algorithms. It is important to note that the developed model focussed on modelling macroscopic phenomena observed in SMAs, as these are most relevant for actuator development.

2. MODELLING ALGORITHM

In Figure 1, the expected curves for tensile loading and unloading experiments given a starting point of either 100% twinned martensite or 100% austenite are shown (monotonic loading - unloading in martensite (a) and austenite (b) and cyclical loading-unloading in martensite (c) and austenite (d)). The data in the figure shows a repeatable 's-shaped' curve (shown as red dotted curves in Figure 1. Note that in plot (b), two s-shaped curves, one for loading and the other for unloading behaviour, are present). This curve, plotted by the blue solid line on the stress-strain ($\sigma - \varepsilon$) plane in Figure 2(a), is characterised by three slopes connected by two 'knees' and

describes, physically, the evolution of phases in the SMA during an iso-thermal tensile test. For a test done below M_f (100% martensite), twinned martensite is elastically deformed in the first segment $a - b$. Detwinning starts at ε_1 and proceeds until complete conversion to detwinned martensite at ε_2, after which the detwinned martensite is elatically deformed until plasticity (not shown) sets in. For a test conducted above A_f (100% austenite), in segment $a - b$, the austenite is elastically deformed. In segment $c - d$, austenite to stress-induced-martensite (SIM) transformation proceeds up to ε_2. Then, for the rest of $e - f$, 100% SIM is elastically deformed until plastic deformation starts.
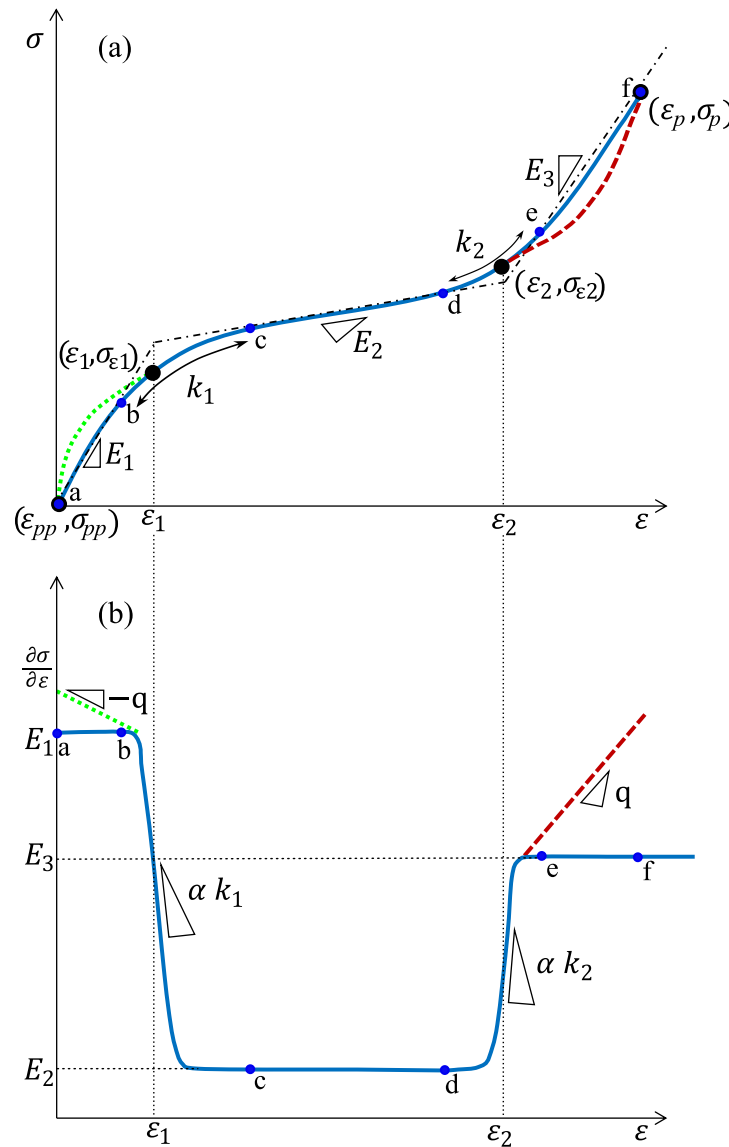

Figure 2. General characteristic of SMA stress-strain curve (a), derivative of stress vs strain (b)

Since this curve is ubiquitous in the SMA stress-strain diagrams, the modelling algorithm is as follows:

1. Mathematical description of the s-shaped curve using a closed-form, differentiable, analytical equation, the 'base equation' (see section 2.1).

2. Identification of required parameters in the base equation (see section 2.2).

3. Imposition of continuity conditions at the points where the strain input changes direction (see section 2.3).

4. Update of parameters to model loading, unloading and reloading behaviour with the base equations (see section 2.4).

The next sub-sections describe each of these steps in detail.

2.1 Base equation

In order to mathematically describe the s-shaped curve, the derivative of the stress with respect to strain ($\frac{\partial \sigma}{\partial \varepsilon}$) is plotted against the strain as shown by the solid blue curve in Fig. 2(b), where E_1, E_2 and E_3 are the slopes of the σ-ε curve in segments $a-c$, $c-d$ and $d-f$, respectively, ε_1 and ε_2 are the strains at the 'knees' and k_1 and k_2 describe the curvatures of the function at the knees. Their values are proportional to $\frac{\partial^2 \sigma}{\partial \varepsilon^2}$, evaluated at ε_1 and ε_2. This derivative can be mathematically described using a combination of scaled and shifted sigmoid functions as shown in (1).

$$\frac{\partial \sigma}{\partial \varepsilon} = \frac{(E_1 - E_2)}{1 + e^{k_1(\varepsilon - \varepsilon_1)}} + \frac{(E_3 - E_2)}{1 + e^{-k_2(\varepsilon - \varepsilon_2)}} + E_2 \tag{1}$$

Taking the integral of (1) yields the 'base equation', σ_{bq}, in (5) where σ_1, σ_2 and σ_3 in (2), (3) and (4) describe the function, approximately, in segments $a-c$, $c-d$ and $d-f$, respectively. (ε_p, σ_p) are the coordinates of any point on the curve and arise from the integration constants. In the model, they are specifically set to the coordinates of the end points of the base curve, although this is mathematically not required. Additionally the starting point of the base curve is also preserved in the coordinates (ε_{pp}, σ_{pp}) as shown in Figure 2(a). Equation (5) succeeds in describing the base curve in a closed form, analytical and differentiable equation that is extremely simple to implement. The vast majority of existing models use the dash-dotted line in the top plot of Fig. 2(a) as an approximation. In this case, a plot of $\frac{\partial \sigma}{\partial \varepsilon}$ is piecewise continuous, the curvatures at the knees in the σ-ε curve are replaced by sharp corners and the base curve, while linear, is not differentiable, and is described with a series of conditional statements that are tedious to program and error-prone when considering several loading-unloading cycles.

$$\sigma_1 = (E_1 - E_2)\left[\varepsilon - \varepsilon_p - \frac{1}{k_1}\ln\left(\frac{1 + e^{k_1(\varepsilon - \varepsilon_1)}}{1 + e^{k_1(\varepsilon_p - \varepsilon_1)}}\right)\right] \tag{2}$$

$$\sigma_2 = E_2(\varepsilon - \varepsilon_p) + \sigma_p \tag{3}$$

$$\sigma_3 = (E_3 - E_2)\left[\varepsilon - \varepsilon_p + \frac{1}{k_2}\ln\left(\frac{1 + e^{-k_2(\varepsilon - \varepsilon_2)}}{1 + e^{-k_2(\varepsilon_p - \varepsilon_2)}}\right)\right] \tag{4}$$

$$\sigma_{bq}(E_{1,2,3}, \varepsilon_{1,2}, k_{1,2}, (\varepsilon_p, \sigma_p), \varepsilon) = \sigma_1 + \sigma_2 + \sigma_3 \tag{5}$$

Further inspection of the experimental data has shown that parabolic loci can occur in either the first segment (see Figure 1(c), where a parabola is shown exemplary by the green dash-dotted curve) or in the third segment (see Figure 1(b)). To this end, (5) is augmented with σ_q in (6) for a parabola in segment $\varepsilon_2 - f$ (see the red dashed line in Figure 2) or with σ_q in (7) for a parabola in segment $a - \varepsilon_1$ to model the green dotted line in Figure 2. The parameter q is the parabolic constant. Consequently, the final base equation is given by (8). Note that should a parabolic locus be absent, then $q = 0$, rendering $\sigma_q = 0$ and (8) is identical to (5).

For unloading:

$$\sigma_q = q(\varepsilon - \varepsilon_2)\left[\varepsilon - \varepsilon_2 - \frac{1}{k_2}\ln\left(1 + e^{k_2(\varepsilon - \varepsilon_2)}\right)\right]$$
$$- q(\varepsilon_p - \varepsilon_2)\left[\varepsilon_p - \varepsilon_2 - \frac{1}{k_2}\ln\left(1 + e^{k_2(\varepsilon_p - \varepsilon_2)}\right)\right] \tag{6}$$

For loading:

$$\sigma_q = q(\varepsilon - \varepsilon_1)\left[\varepsilon - \varepsilon_1 + \frac{1}{k_1}\ln\left(1 + e^{k_1(\varepsilon_1 - \varepsilon)}\right)\right]$$
$$- q(\varepsilon_p - \varepsilon_1)\left[\varepsilon_p - \varepsilon_1 + \frac{1}{k_1}\ln\left(1 + e^{k_1(\varepsilon_1 - \varepsilon_p)}\right)\right] \tag{7}$$

$$\sigma_{bq}(E_{1,2,3}, \varepsilon_{1,2}, k_{1,2}, (\varepsilon_p, \sigma_p), q, \varepsilon) = \sigma_1 + \sigma_2 + \sigma_3 + \sigma_q \tag{8}$$

The parameter identification process is presented in the following section.

2.2 Parameter identification of base equation

Parameter identification is based on experimental data of monotonic loading-unloading experiments on martensite and austenite such as in Figure 1(a) and (b). To describe these curves, the base equation in (8) is used, with $q = 0$, since the initial loading curve does not contain parabolic components. The required parameters here are: $E_1, E_2, E_3, \varepsilon_1, \varepsilon_2, k_1, k_2, (\varepsilon_p, \sigma_p)$.

Since $(\varepsilon_p, \sigma_p)$ arise from integration constants, they can be set to any value on the curve and for parameter identification, they are set to $(0,0)$. This is justified because at the beginning of the loading experiment, the stress and the strain are commonly both 0. The remaining 7 parameters are identified as follows:

1. E_1, E_2, E_3 are calculated as the slopes of the lines passing through any two points in the segments $a-b$, $c-d$, $e-f$ (see Figure 2), respectively. The lines passing through these point will be called ab, cd and ef, respectively, and are shown by the dash-dotted lines in Figure 2.

2. ε_1 is calculated as the intersection of line ab with cd and similarly, ε_2 is the intersection of cd with ef.

3. σ_{ε_1} and σ_{ε_2}, the stresses at ε_1 and ε_2, as calculated in the step above, are extracted from the experimental data.

4. k_1, k_2 are then calculated with equations (9) and (10) where $(\varepsilon_{ab}, \sigma_{ab})$ and $(\varepsilon_{ef}, \sigma_{ef})$ are the coordinates of any point in segment $a-b$ and $e-f$, respectively.

$$k_1 = -\ln(2)\left(-\varepsilon_1 + \frac{\sigma_{\varepsilon_1} - \sigma_{ab} - E_2\varepsilon_1 + E_1\varepsilon_{ab}}{E_1 - E_2}\right)^{-1} \tag{9}$$

$$k_2 = \ln(2)\left(-\varepsilon_2 + \frac{\sigma_{\varepsilon_2} - \sigma_{ef} - E_2(\varepsilon_1 - \varepsilon_{ef})}{E_3 - E_2}\right)^{-1} \tag{10}$$

Additionally, for austenite, the 'width' of the hysteresis loop is required and this value is indirectly calculated from ε_1^{uld}, the strain at the first unloading knee and E_2^{uld}, the slope of the unloading plateau. Both of these parameters are easy to identify from the data. For martensite, no further parameters are required. In the next section, the continuity conditions are presented.

2.3 Imposition of continuity conditions

Consider that the SMA being tested in the experiment is first in its 100% austenitic state at no load. Consider a simple tensile experiment with a strain input as shown by the red dashed curve in Figure 3 i.e. monotonic loading upto 4% strain, unloading to 0% strain and reloading to 8% strain. Consider a second SMA sample in its 100% twinned martensitic state at no load and a strain as shown by the solid blue curve in Figure 3 i.e. monotonic loading upto 4% strain, unloading to 2% strain and reloading to 8% strain. Assume further that identification experiments as described in section 2.2 above have been carried out and the parameters for both austenite and martensite are known.

Using the base equation (8) and the parameters identified, the first loading sequence from 0% to 4% can be described mathematically and is shown in plot (a) of Figure 4, where the left column is for austenite and the right column for martensite. (The red dashed curve is included
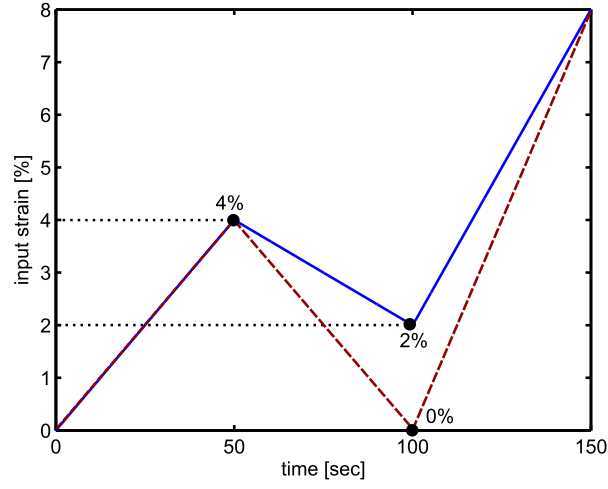
Figure 3. Strain input for simple tensile experiment

only as a reference for the locus of the loading curve). Recall that $(\varepsilon_{pp}, \sigma_{pp})$ and $(\varepsilon_p, \sigma_p)$ track the start and end coordinates of the function. At the beginning of loading, both are initialised at (0,0). At the point where the strain input changes direction i.e. at the end of loading at 4% strain, the value of $(\varepsilon_p, \sigma_p)$ is updated to the end points of the loading curve i.e. (4, 475) for austenite and (4, 170) for martensite. For the next unloading, sequence, (8) is used again to describe the next s-shaped curve shown with the red dashed curve Figure 4(b). The parameters required to describe this curve need to be calculated through a parameter update process as presented in the next section. The selection of $(\varepsilon_p, \sigma_p)$ as the end points of the previous sequence ensures continuity of the curve. Note that in martensite, although the strain input sinks to 2%, values of stress below 0 are cut off. When the strain input changes direction again i.e. at the end of the unloading section, the values of both $(\varepsilon_{pp}, \sigma_{pp})$ and $(\varepsilon_p, \sigma_p)$ are updated to (4, 475) and (0.1, 0) for austenite and (4, 170) and (3.4, 0) for martensite in preparation for the reloading sequence. The reloading curve is described similarly using (8) and once again updated parameters. Note that only the tracking of the end points in each sequence through $(\varepsilon_p, \sigma_p)$ is sufficient to ensure continuity. $(\varepsilon_{pp}, \sigma_{pp})$ is preserved in order to have a record of the loading - unloading history of the SMA. In the next section, the parameter update process is presented.

2.4 Parameter update

The parameter update algorithm is used to calculate new parameters when the strain input changes direction. It is based on the parameters identified from the identification process in section 2.2, the parameters $(\varepsilon_{pp}, \sigma_{pp})$ and $(\varepsilon_p, \sigma_p)$, which track the history of the SMA during the course of the experiment as well as phenomena and hypothesis from materials science. The updated parameters will be appended with the superscript * to differentiate them from the identified parameters. The update algorithm will be presented separately for loading and unloading in the following sub-sections.
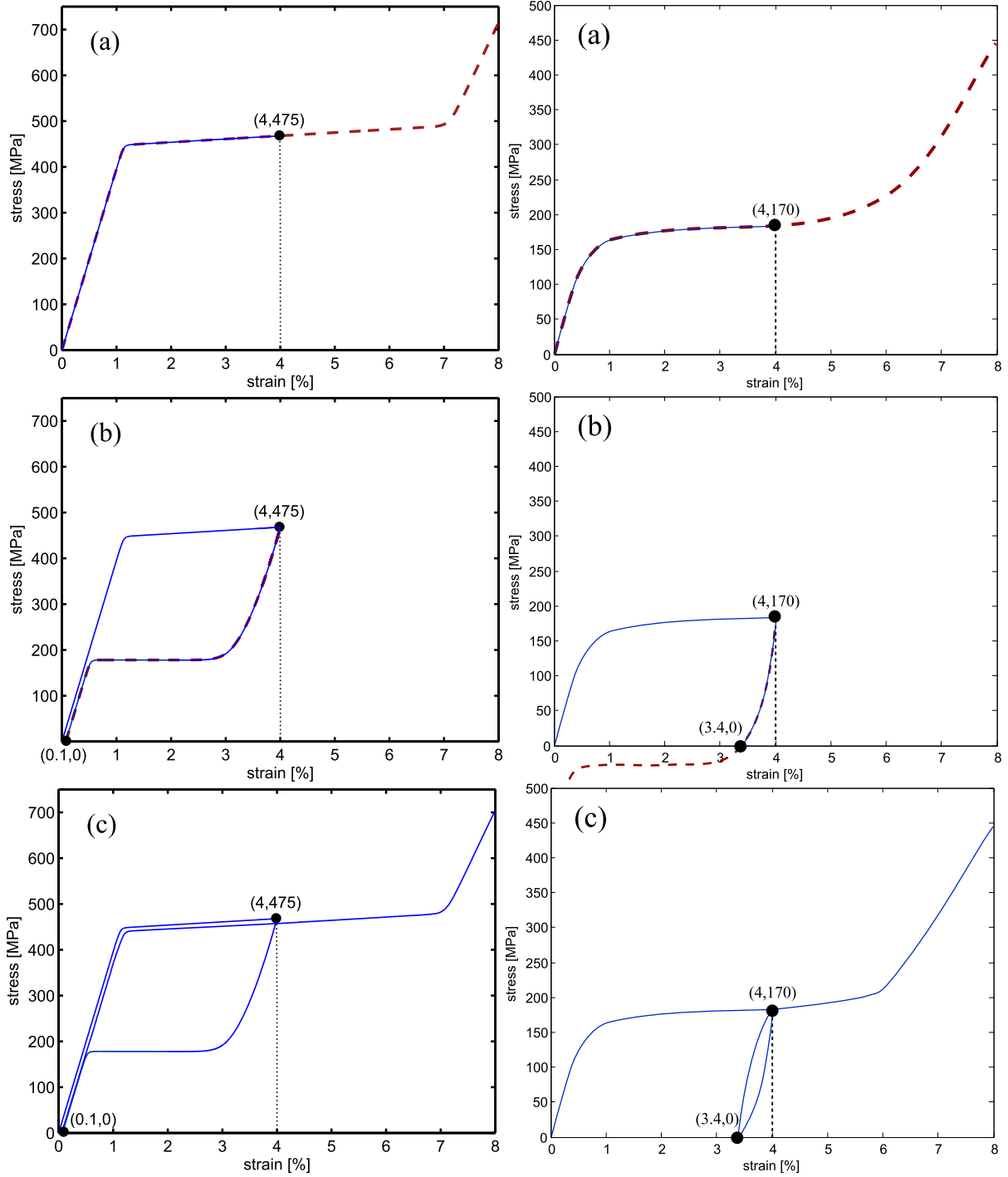
8

Figure 4. Continuity conditions for loading (a), unloading (b), reloading (c) in austenite (right column) and martensite (left column)

2.4.1 Unloading

In order to model unloading behaviour, the following parameters need to be calculated: $E_3^*, \varepsilon_1^*, \varepsilon_2^*, k_2^*, q^*$ (see Figure 5). The rest of the parameters remain the same i.e. $E_1^* = E_1$, $E_2^* = E_2$ and $k_1^* = k_1$. For austenite, the value of $E_2^* = E_2^{uld}$. Additionally, since unloading behaviour in both austenite and martensite is characterised by residual strain, an additional parameter, ε_r, is introduced. In austenite, the build up of residual strain is attributed, among others, to plastic deformation of SIM, micro-yielding effects and degradation [3, 4]. The amount of residual strain saturates after about 50 cycles according to a negative exponent (c.f. equation (11), where n is the number of cycles) due to strain hardening (training effect). This is consistent with findings available in literature, e.g. in [30].

$$\varepsilon_r = (0.37)(1 - e^{-0.035(n-2)}) + 0.05 \tag{11}$$

In martensite, residual strain occurs because no driving force for a reverse transformation is present, thus the unloading path is characterized by an almost elastic behaviour. To this end, ε_r is calculated with (12) where E_3^* is in (16).

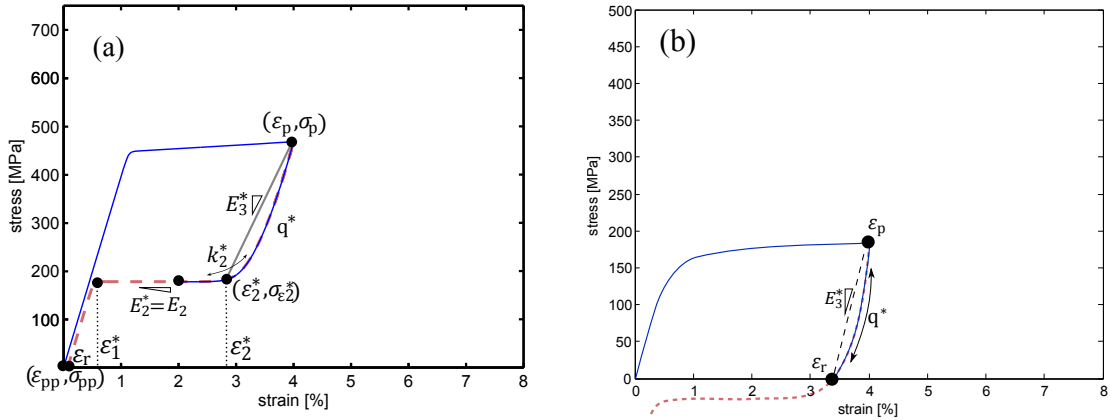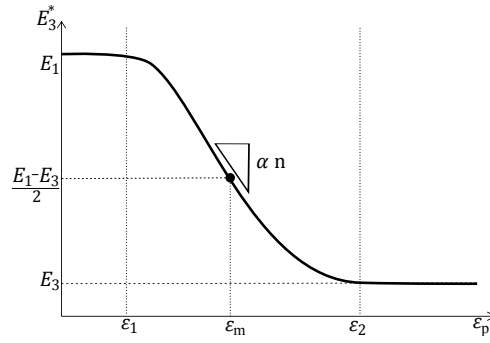$$\varepsilon_r = \varepsilon_p - \frac{\sigma_p}{E_3^*} \tag{12}$$

Figure 5. Parameter update in unloading for austenite (a), martensite (b)

The rest of the parameters are identical for austenite and martensite and are calculated as follows:

- ε_2^*: This parameter is updated using (13), where ε_r is the residual strain, calculated with (11) or (12) and σ_1 and σ_2 are calculated with equations (2) and (3) respectively, evaluated at $\varepsilon = \varepsilon_r$ and with $\varepsilon_1 = \varepsilon_1^{uld}$.

$$\varepsilon_2^* = \frac{1}{k_2} \ln \left(\frac{e^{k_2 \left(\frac{\sigma_1 + \sigma_2}{E_2 - E_3} + \varepsilon_p - \varepsilon_r \right)} - 1}{e^{-k_2 \varepsilon_r} - e^{k_2 \left(\frac{\sigma_1 + \sigma_2}{E_2 - E_3} - \varepsilon_r \right)}} \right) \tag{13}$$

10

- E_3^*: The value of E_3^* follows the curve in Figure 6, where the x-axis is ε_p and the y-axis is E_3^*. What this curve implies is that should unloading occur before the first knee i.e. $\varepsilon_p \leq \varepsilon_1$, then $E_3^* = E_1$. This is reasonable considering that when $\varepsilon_p \leq \varepsilon_1$, then the SMA is microscopically either fully austenite or twinned martensite and unloading from ε_p is elastic with modulus E_1. Consider, however, that the SMA is loaded to a value ε_p such that $\varepsilon_p \geq \varepsilon_2$. In this case the material is microscopically either fully SIM or detwinned martenstite and unloading from ε_p is also elastic, but with modulus E_3. Therefore, $E_3^* = E_3$. If however, unloading occurs between the knees i.e. $\varepsilon_1 \leq \varepsilon_p \leq \varepsilon_2$, then the microscopic composition of the SMA is either a mixture of austenite and SIM or twinned and detwinned martensite. The unloading slope, E_3^*, in this case is hypothesised to be a weighted linear combination of the slopes E_1 and E_3, where the weight is given by the volume fraction of SIM or detwinned martensite, v, present in the material. This volume fraction is given by the relative position of ε_p to the knees i.e. when ε_p is at ε_1, $v = 0$ and when ε_p is at ε_2, $v = 1$. In between the knees v varies linearly.

Figure 6. Update of E_3 in unloading

The curve in Figure 6 is mathematically described by (16), where n in (14) is proportional to the slope of the curve at ε_m in (15), the midpoint between the knees.

$$n = \frac{4}{\varepsilon_2 - \varepsilon_1} \tag{14}$$

$$\varepsilon_m = \frac{\varepsilon_2 + \varepsilon_1}{2} \tag{15}$$

$$E_3^* = \frac{E_3 - E_1}{1 + e^{-n(\varepsilon_p - \varepsilon_m)}} + E_1 \tag{16}$$

- q^* and k_2^*: In addition to the unloading slope calculated above, the experimental data shows a parabolic locus. In the SMA, grains of all possible orientations can be expected to be found. As grain orientation corresponds to critical stress levels for phase transformation of single grains [32, 4, 1], small volume fractions of the wire might show a premature and others sluggish phase transformation. As a consequence, unloading is characterized

11

by the smooth curvatures depicted. Should this sluggish behaviour be absent, then the material would unload elastically (and linearly) until ε_2^* in (13). This effect is exploited to calculate $\sigma_{\varepsilon_2^*}$, the expected stress at ε_2^* (see Figure 5(a)) using (17). The calculation of the parabolic coefficient q^* then proceeds with the (simplified) quadratic formula in (18). k_2^* is subsequently calculated with (19) to ensure a smooth transition between the parabola and the rest of the curve.

$$\sigma_{\varepsilon_2^*} = E_3^*(\varepsilon_2^* - \varepsilon_p) + \sigma_p \tag{17}$$

$$a = (\varepsilon_p - \varepsilon_2^*)^2; \quad b = \sigma_{\varepsilon_2^*} - \sigma_p - E_3^*(\varepsilon_2^* - \varepsilon_p); \quad c = ln(2)(E_3^* - E_2^*)^2$$

$$q^* = \frac{b - \sqrt{b^2 - ac}}{2a} \tag{18}$$

$$k_2^* = \frac{4q^*}{E_3^* - E_2^*} \tag{19}$$

- ε_1^*: With all the other values, ε_1^* is then calculated with equation (20) where σ_3, σ_2 and σ_q are calculated with equations (4), (3) and (6), evaluated at $\varepsilon = \varepsilon_r$ and $\sigma_r = 0$.

$$\varepsilon_1^* = \frac{1}{k_1} \ln \left(\frac{e^{k_1 \varepsilon_p} - e^{k_1 \left(\frac{\sigma_r - \sigma_3 - \sigma_2 - \sigma_q}{E_1 - E_2} + \varepsilon_p \right)}}{e^{k_1 \left(\frac{\sigma_r - \sigma_3 - \sigma_2 - \sigma_q}{E_1 - E_2} + \varepsilon_p - \varepsilon_r \right)} - 1} \right) \tag{20}$$

2.4.2 Reloading

When the SMA is in its austenitic state at no load and the first loading-unloading sequence is complete i.e. unloading to 0 MPa Stress, as shown in Figure 7(a), then the parameter update requires no additional calculations and the identified parameters are directly used as the reloading parameters (the * parameters). In the martensite case, the values of ε_1^*, E_2^* and q^* are the only parameters that need updating as shown in Figure 7(b). ε_1^* is simply set to the value of ε_{pp}, the point at which the previous unloading took place. E_2^* is only updated when the value of $\varepsilon_p \geq \varepsilon_2$. In this case, $E_2^* = E_3$ as reloading joins the original loading curve after the plateau as shown exemplarily in Figure 8 for martensite. The value of q^* is calculated using equation (21) with $\varepsilon_q = \varepsilon_1^* = \varepsilon_{pp}$ and σ_1, σ_2 and σ_3 are calculated with equations (2), (3) and (4) evaluated at $\varepsilon = \varepsilon_q = \varepsilon_{pp}$.

$$q^* = \frac{\sigma_1 + \sigma_2 + \sigma_3}{\varepsilon_q - \varepsilon_p} \tag{21}$$

With the algorithm described in the above three sections namely, parameter identification, imposition of continuity conditions and parameter update, the model is then validated using experiments. The results are presented in the next section.
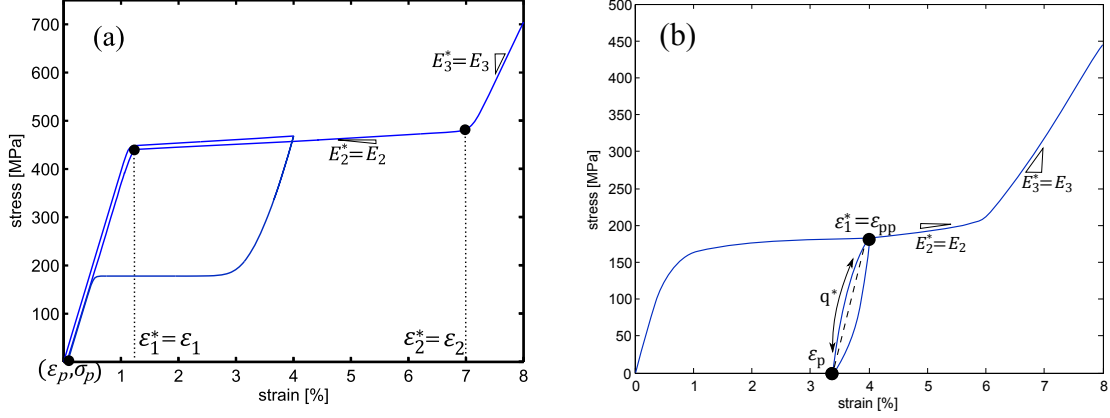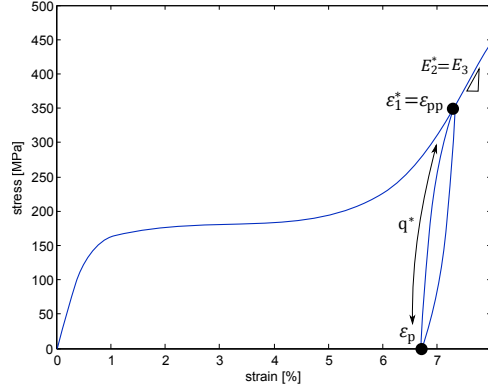
Figure 7. Parameter update in reloading for austenite (a), martensite (b)

Figure 8. E_2 update for reloading beyond ε_2

3. EXPERIMENTAL VERIFICATION

3.1 Experimental Set-up

In order to validate the model, tensile experiments were performed on poly-crystalline NiTi wire samples with a servo-hydaulic testing machine. The wires with 0.5 mm diameter were cut to a nominal length of about 60 mm. They were then attached to custom built fixing grips with grooves of appropriate diameter to facilitate alignment and to ease installation within the testing machine. The experiments were carried out under displacement control and with a constant cross head displacement of 2 mm/min. The experiments were all conducted at room temperature. In order to characterise both martensite and austenite, two sets of wires were used: one set with an A_f of $-25\,^\circ$C, meaning that the wire samples from this set were fully austenitic at room temperature and no load. The other set had an A_f of $95\,^\circ$C, meaning that the wire samples from this set were fully (twinned) martensitic at room temperature and no load. For both wire samples, initially, a monotonic experiment was first carried out. The data was then used to

13

identify the required parameters according to section 2.2. Subsequently, the same wire sample was then exposed to a cyclical loading and unloading experiment and the model was used to predict expected behaviour using the identified parameters, (8), the continuity conditions and the parameter update process. The following subsections present the experimental data and the model for austenite and martensite separately.

3.2 Austenite: monotonic experiments and parameter identification

In this experiment, the specimen was a NiTi wire with A_f of $-25\,°C$, therefore the wire was austenitic at room temperature and no load. The wire sample was loaded to with a constant velocity of 2 mm/min to a maximum displacement of 5 mm (approximately 8.5% strain) followed by an unloading ramp to a force of 5 N. The results are plotted as solid black curves in Figure 9. Using this data, the model parameters are extracted using the steps outlined in section 2.2 and are shown in Table 1 . The model results are plotted as the red dash-dotted curve in Figure 9. The results show good correlation with experimental data with a root-mean-square-error (RMS-error) of 6.6 MPa. The model shows good prediction especially at the knees, where other models tend to have large errors. Note that the austenite unloading curve is characterised by a 'peak' at 5.5% strain. These peaks are due to nucleation in the SMA [33] and for simplicity, they are not accounted for in the model.

	E_1	E_2	E_3	ε_1	ε_2	k_1	k_2	q	E_2^{uld}	ε_1^{uld}
	GPa	GPa	GPa	%	%	–	–	GPa	GPa	%
Austenite	40	0.7	25	1.13	7.45	3500	1400	0	0	0.34

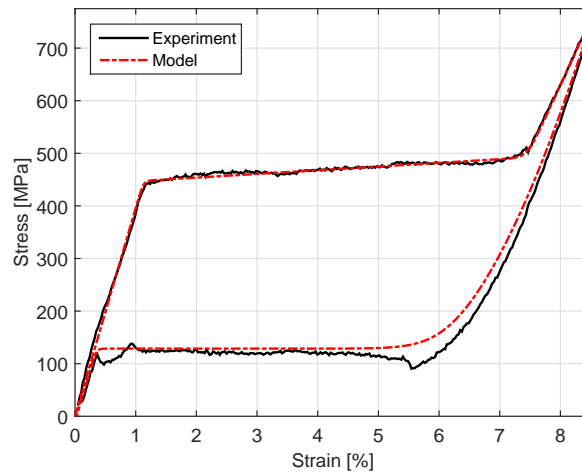Table 1. Model parameters for austenite

Figure 9. Austenite: Monotonic loading,unloading - experiment (solid) and model (dashed)

14

3.3 Austenite: cyclical experiments

In this experiment, the same wire sample as in the monotonic experiment above, is exposed to a displacement input as in Figure 10 i.e. 25 cycles of loading and unloading, where the maximum displacement of each cycle is increased for 0.2 mm and the minimum displacement for each cycle is 0 mm. Therefore, the material is loaded up to different strains in each of the three areas of the s-shaped curve and completely unloaded before the start of the next cycle.
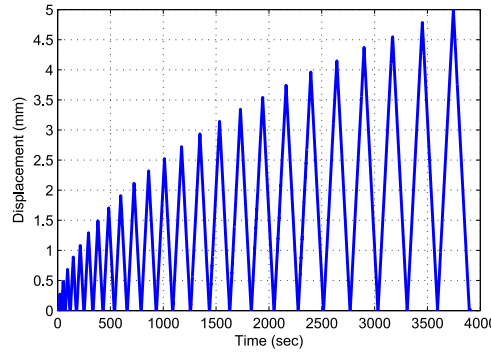

Figure 10. Input displacement versus time

The results are plotted by the solid black curve in Figure 11. Using the parameters in Table 1, imposition of the continuity conditions each time the strain direction changes and the parameter update process, the model prediction is shown as the red dash-dotted line in Figure 11. The model shows good overall correlation with experimental data with an RMS error of 7.4 MPa . Additionally, the model is able to predict following observed phenomena:

- 'Shifts' of the pseudoelastic loop each cycle. This is due to the fact that residual strain builds up in the material.

- The consecutive decrease of the onset of forward transformation (the critical stress for SIM formation), which is induced by micro-plasticity in the SMA [3, 4].

Further, the hypothesis that the unloading slope E_3^* is based on the volume fraction of SIM present in the material at the unloading strain, ε_p gives good results. It was found that the lowest RMS error tended to occur when the slope of the curve in Figure 6 is double the amount that was originally hypothesised i.e. (16) is amended to (22) for austenite.

$$E_3^* = \frac{E_3 - E_1}{1 + e^{-2n(\varepsilon_p - \varepsilon_m)}} + E_1 \tag{22}$$

3.4 Martensite: monotonic experiments and parameter identification

In this experiment, the wire specimen was a NiTi wire with A_f of $95\,°C$, therefore the wire was martensitic at room temperature. Prior to the experiment the wire sample was placed in a
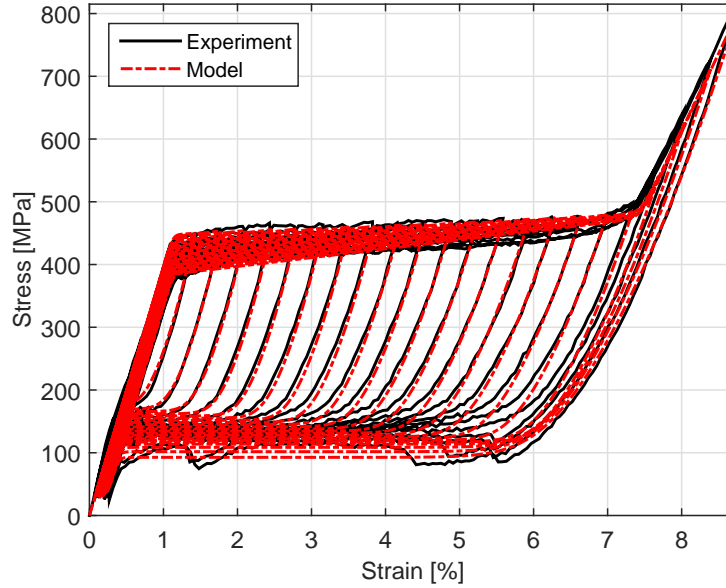
15

Figure 11. Austenite complete cyclical loading - experiment (solid) and model (dashed)

furnace at $400\,^\circ$C for 10 seconds and allowed cool to room temperature to remove all residual strains and to ensure that the martensite was 100% twinned martensite at no load. The experiment was carried out exactly as in the monotonic austenite case in section 3.2 above. The results are plotted as solid black curves in Figure 12. Using this data, the model parameters are extracted using the steps outlined in section 2.2 and are shown in Table 2 . The model results are plotted as the red dash-dotted curve in Figure 12. The results show excellent correlation with experimental data with an RMS-error of 1.9 MPa.

	E_1	E_2	E_3	ε_1	ε_2	k_1	k_2	q
	GPa	GPa	GPa	%	%	—	—	GPa
Martensite	29	0.39	15.8	0.58	6.43	580	170	0

Table 2. Model parameters for martensite

3.5 Martensite: cyclical experiments

In this experiment, the same wire sample as in the monotonic experiment above was used. Before the start of the experiment, the wire was placed in a furnace at $400\,^\circ$C for 10 seconds and allowed cool to room temperature to remove the residual strain from the monotonic experiment. The rest of the testing procedure and input displacement is identical to that of cyclical loading in austenite as in section 3.3. The experiment results are plotted as the solid black curves in Figure 13 and show, as expected, an increasing residual strain at every cycle. Further, the unloading and reloading cycles are characterised by minor hysteresis loops (see inset in Figure 13). Using
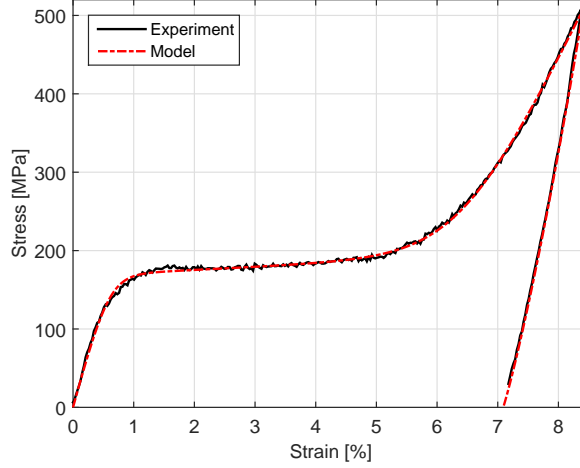
16

Figure 12. Martensite monotonic loading - unloading: experiment (solid) and model (dashed)

the parameters in Table 2, imposition of the continuity conditions each time the strain direction changes and the parameter update process, the model prediction is shown as the red dashed line in Figure 13. The model shows good overall correlation with experimental data with an RMS error of 5.5 MPa . Note that the inclusion of parabolic loci for the unloading and loading curves allows the model to automatically predict the minor hysteresis loops. Additionally, similar to the austenite case, the hypothesis that the unloading slope E_3^* is based on the volume fraction of detwinned martensite present in the material at the unloading strain, ε_p gives good results. It was found that the lowest RMS error tended to occur when (23) is used instead of (16).

$$E_3^* = \frac{2E_3 - E_1}{1 + e^{-n(\varepsilon_p - \varepsilon_m)}} + E_1 \tag{23}$$

4. CONCLUSIONS

This paper presents a novel phenomenological constitutive model that can be used to predict the behaviour of either martensite or austenite subjected to monotonic or cyclical loading and unloading cycles. The model is based on the mathematical description of a ubiquitous s-shaped curve present in typical stress-strain uniaxial tensile experiments on martensite and austenite with an analytical, continuous and differentiable equation without the use of conditional statements. The model parameters are few, physical and easy to identify and the parameter identification process has to be carried out only once using a simple monotonic loading and unloading experiments for both austenite and martensite. The modelling of loading and unloading behaviour is achieved through the imposition of continuity conditions and a parameter update algorithm that uses the parameters identified in the identification algorithm and the tensile history of the SMA in the experiment. In addition to accurately modelling the curvature at the knees, the model includes experimentally observed phenomena such as quadratic loci for major and minor loops in both phases, the variation of the unloading slopes based on the volume
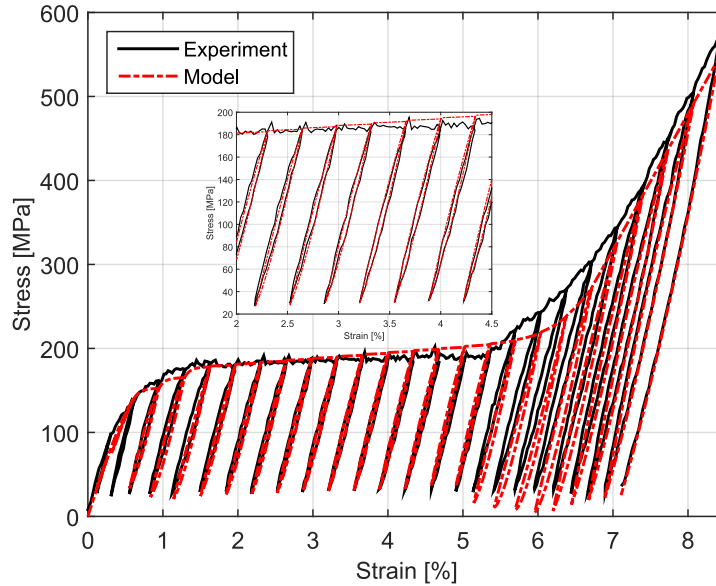
Figure 13. Martensite cyclical loading - experiment (solid) and model (dashed)

fraction of the phases present, load history effects such as the build up of residual strain and the presence of minor hysteresis loops in cyclical martensite experiments. The model's simplicity guarantees computational efficiency without compromising accuracy in predicting observed behaviour, as was verified with monotonic and cyclic loading-unloading experiments. The model can therefore form the basis for the development of real-time control algorithms for SMA actuators. Further, although the experimental verification was carried out using NiTi wires, the phenomena that are modelled are universal for all other SMAs such as iron based, copper based alloys etc. To model the behaviour of these alternative SMAs, only the parameter identification step is required. Future work will concentrate, amongst others, on improving current accuracy, extension of the model to partial cyclical loading cases and extension of the model to intermediate temperatures, where the SMA material in the unloaded condition (zero stress), is composed of martensite and austenite simultaneously.

5. ACKNOWLEDGEMENTS

References

[1] K. Otsuka and X. Ren. Physical metallurgy of Ti-Ni-based SMAs. *Progress in Materials Science*, 50(5):511–678, 2005.

18

[2] Y. Liu, Z. Xie, J. Van Humbeeck, and L. Delaey. Some results on the detwinning process in NiTi shape memory alloys. *Scripta materialia*, 41(12):1273–1281, 1999.

[3] K. Gall et al. On the mech. behv. of single crystal NiTi SMAs and related polycrystalline phenomenon. *Materials Sci. and Eng.*, :85-92, 2001.

[4] P. Krooss, T. Niendorf, I. Karaman, Y. Chumlyakov, and H.J. Maier. Cyclic deformation behaviour of aged FeNiCoAlTa single crystals. In *Functional Material Letters*, 2012.

[5] J. Frenzel et al. Influence of Ni on martensitic phase transformations in NiTi SMAs. *Acta Materialia*, 58(9):3444–3458, 2010.

[6] J. Van Humbeeck. SMAs: A material and a technology. *Adv. Eng. Mat.*, 3:837–850, 2001.

[7] N. Huber, N. Fleck, and M. Ashby. The select. of mech. act. based on perf. idices. *Int. J. of Plasticity*, 2(11):59–72, 1986.

[8] M. Achenbach. A model for an alloy with shape memory. *International Journal of Plasticity*, 5(4):371–395, 1989.

[9] J.G. Boyd and D.C. Lagoudas. A thermodynamical constitutive model for SMAs. Part 1. The monolithic SMA. *International Journal of Plasticity*, 12:805–842, 1996.

[10] L. C. Brinson and M. S. Huang. Simplifications and comparisons of SMA constitutive models. *Intelligent Material Systems and Structures.*, 7:108–114, 1996.

[11] Y. Ivshin and T. J. Pence. Thermomechanical model for a 1-variant SMA. *Journal of Intellelligent Material Systems and Structures*, 5:455–473, 1994.

[12] S. Seelecke and I. Muller. SMA actuators in smart structures: Modeling and simulation. *Applied Mechanics Review*, 57, 2004.

[13] John A. Shaw. A thermomechanical model for a 1-D SMA wire with propargating instabilities. *International Journal of Solids and Structures*, 39:1275–1305, 2002.

[14] R. Smith. *Smart Material Systems: Model Development*. SIAM, 2005.

[15] F. Preisach. Über die magnetische Nachwirkung. *Zeit. für Physik*, 94, 1935.

[16] S.M. Dutta and F.H. Ghorbel. Differential hysteresis modeling of an SMA wire actuator. *IEEE/ASME Transactions on Mechatronics*, 10:189–197, 2005.

[17] K. Ikuta, M. Tsukamoto, and S. Hirose. Mathematical model and experimental verification of SMA for designing micro actuator. In *Proc. of IEEE MEMS*, pages 103–108, 1991.

[18] K. Tanaka, S. Kobayashi, and Y. Sato. Thermomechanica of transfromation, pseudoelasticity and SME. *International Journal of Plasticity*, 2(11):59–72, 1986.

[19] C. Rogers and C. Liang. 1-D thermomechanical constitutive relations for SMAs. *Journal of International Material Systems and Structures*, 8(4):285–302, 1997.

[20] L.C. Brinson. 1-D constitutive behaviour of SMAs: thermomech. deriv. with non-constant material functions and redefined martensite internal variable. *Intelligent Material Systems and Structures*, 4(2):229–242, 1993.

[21] M. Elaninia and M. Ahmadian. An enhanced SMA phenomenological model: I and II. *Smart Materials and Structures*, 14:1297–1308, 2005.

[22] D.Madill and D.Wang. Modelling and L2-stability of a SMA position control system. *IEEE Transactions on Control Systems Technology*, 6(4):473–481, 1998.

[23] A. Pai and R.B. Gorbet. Extension of madill's SMA model to include time-varying stress. In *CANSMART 2006 Smart Materials and Structures*, pages 60–70, Canada, 2006.

[24] B.Goo and C.Lexcellent. Micromechanics-based modeling of 2-way memory effect of a single crystal SMA. *Acta Materialia*, 45, 1997.

[25] Gao X. et al. A multivariant micromechanical model for smas: Part 1: Crystallographic issues for single crystal model. *International Journal of Plasticity*, 16:1345–1369, 2000.

[26] A. Vivet and C. Lexcellent. Micromechanical modelling for tension-compression pseudoelastic behavior of AuCd single crystals. *EPJ Applied Physics*, 2:125–132, 1998.

[27] L.C. Brinson and R. Lammerung. FE analysis of the behavior of SMAs and their applications. *International Journal of Solids and Structures*, 30:3261–3280, 1993.

[28] J. J. Amalraj, A. Bhattacharyya, and M. G. Faulkner. Finite-element modeling of phase transformation in SMA wires with variable material properties. *Smart Materials and Structures*, 9(5):622–631, 2000.

[29] Auricchio et al. Theoretical and numerical modeling of shape memory alloys accounting for multiple phase transformations and martensite reorientation. *International Journal of Plasticity*, 59:30–54, 2014.

[30] W. Ren, H. Li, and G. Song. Phenomenological modeling of the cyclic behavior of superelastic shape memory alloys. *Smart Materials and Structures*, 16:1083–1089, 2007.

[31] A. Pai. *A phenomenological model of shape memory alloys including time varying stress.* Masters thesis, University of Waterloo, 2007.

[32] E. Hornbogen. *Advanced Structural and Functional Materials.* Springer, 1991.

[33] John A. Shaw and Stellos Kyriakides. On the nucleation and propagation of phase transformation fronts in a NiTi alloy. *Acta Materialia*, 45(2):683–700, 1997.