Modeling and Control of Shape Memory Alloy Actuators

A Tutorial

Mahbubeh Abuie Nia Mehrizi

The material in this tutorial is based in part on Actuators: modeling and control of shape memory alloy actuators (1st Ed., 2008) by J. Jayender and my own research. For more information, please write to baharabuie@email.kntu.ac.ir.

© 2020 K. N. Toosi University of Technology

1. INTRODUCTION

What is Neuro Shape memory alloy?

A Shape Memory Alloy (SMA) is able to memorise and recover its original shape, after it has been deformed by heating over its transformation temperature. This unique effect of returning to an original geometry after a large inelastic deformation (near 10%) is known as the Shape Memory Effect (SME). SMAs have been used in a wide variety of applications because of their unique thermomechanical characteristics. Two quite successful applications are SMA-made eyeglass frame and the antenna of mobile phone.

Basic Type of SMA-Based Actuators

Both one-way SMA and two-way SMA may be used for deployable structures. Although two-way SMA can perform in two directions due to its two-way shape memory mechanism, transformation strain
associated with it is normally only half of that in one-way SMA. An alternative solution is to put two one-way SMAed based actuators one against another to generate mechanical two-way performance, i.e. heating SMA in one actuator to get forward motion, and heating SMA in another actuator to reverse. The advantage of mechanical two-way actuator is higher motion and higher force than that in material two-way actuator, while the advantage of material two-way actuator is simpler, compact and much less element involved.

![Figure 1. SMA actuator](image)

**Why use shape memory alloys?**

The shape memory behavior of interest must be identified, i.e., the shape memory effect or superelasticity (Ma et al. 2010; Sun et al. 2012; Duerig et al. 1990; Otsuka and Wayman 1998). This behavior should have potential advantages relative to other solutions and technologies.

**What are the application requirements?**

These requirements will identify what properties and performance characteristics are needed from the SMA element. Applications can require the SMA element to provide a single operational cycle or upto millions of cycles. Life of an SMA component is driven by the desired response time and thermomechanical loading conditions. These requirements in turn drive the choice of material, form, size, and control methods.

**What are the cost/expenditure limits?**

The costs include the raw SMA material, processing, and fabrication. The resulting cost per device is critical to the business case for many applications to be commercially viable.

**What is the availability and size of the SMA element?**

The available input (power) and required output (work) of the SMA element often drive material and form selection. SMA elements can be of a variety of forms (e.g., strips, rods, sheets, wires, springs, tubes, etc...) in various sizes. Another key consideration is the availability and required volume of the material from a commercial supplier or other source.

**What efficiency and response time is needed?**

Both energy and mechanical efficiencies of SMA components
may be drivers in evaluating SMA technology. However, many applications, particularly in the aerospace industry, are weight sensitive; thus weight savings (mass efficiency) may be of higher priority. The maximum available cyclic frequency can also drive the evaluation.

**What is the proposed environment?**

The environment, especially thermal conditions, will greatly influence properties of SMAs. Moreover, SMA selection can be limited by the commercial availability of alloys, depending on manufacturing volume of the application. Currently, transformation temperatures are limited to \(115^\circ C\), but this is expected to increase in the near future with the development of high temperature SMAs (Ma et al. 2010; Benafan et al. 2012a). Other concerns may include vibration, humidity, corrosive elements, and bio-compatibility.

**What relevant standard and documents are available?**

This is especially critical in applications that require certification. Examples of such documents include ASTM standards: F2004-05, F2005-05, F2063-05, F2082-06, F2516-07, F2633-07, application specific documents, certification documentation, and supplier data.

**What other components/system will be required?**

Finally, the SMA element is often one component in a system that can include other mechanical and electrical elements. One must consider if required system integration technologies are available or if development is needed, potentially resulting in increased application cost.

**Heating Methods**

Basically, SMAs can be heated by the following three different methods,

- By passing an electrical current through them. This method is only applicable where a small diameter SMA wire or spring is used, otherwise the electrical resistance is too small to produce enough heating. The main advantage is simplicity, while the big disadvantage is that the SMA element needs to be electrically insulated.
- By passing an electrical current through a high resistance wire or tape wrapped around the SMA element. This method is available for SMA bars or tubes. The electrical wire needs to be electrically insulated, but the insulator should have good thermal conductivity to let the heat flow to the SMA element.
- By exposing the SMA component to thermal radiation. This may be the simplest way in space, since a component exposed to the light of the sun will heat to...
temperature of 150°C or more. No additional heating system is required. The main disadvantage of this method is that it is inflexible, and it could be very difficult to react the structure.

An advantage of using heating by electrical current is that by controlling the electrical current, deployment and retraction can be made controllable.

II. THEORY AND MODELING

The model consists of three equations.
- Phase Transformation
- Temperature Dynamics
- Constitutive Equation

**Modeling of Phase Transformation**

Since an SMA exists only in Martensite and Austenite phases, it can be modeled as a two-state system, like an electron. The Fermi–Dirac statistics, which describe the distribution of electrons in two states depending on their energy levels, has been found to provide a good model for the state of an SMA in Martensite and Austenite forms. We use two modeling equations based on whether the alloy is being heated or cooled due to hysteresis with two different transition temperatures. Since the SMA is in the Martensite form at lower temperatures, the phase transformation equation during heating is described by analogy with the Fermi–Dirac statistics in the form

\[
\xi = \frac{\xi_m}{1+\exp\left(\frac{T_f - T}{\sigma_a} + K_a \cdot \sigma\right)}.
\]  

(1)

Where \(\xi\) is the fraction of the Austenite phase, \(\xi_m\) is the fraction of the Martensite phase prior to the present transformation from Martensite to Austenite, \(T\) is the temperature, \(T_f\) is the transition temperature from Martensite to Austenite, \(\sigma_a\) is an indication of the range of temperature around the transition temperature \(T_f\) during which the phase change occurs, \(\sigma\) is the stress, and \(K_a\) is the stress curve-fitting parameter which is obtained from the loading plateau of the stress-strain characteristic with no change in temperature.

On cooling, the Austenite phase gets converted to the Martensite phase and the modeling equation during cooling is described by analogy with the Fermi–Dirac statistics in the form

\[
\xi = \frac{\xi_a}{1+\exp\left(\frac{T_f - T}{\sigma_m} + K_m \cdot \sigma\right)}.
\]  

(2)

Where \(\xi_a\) is the fraction of the Austenite phase prior to the present transformation from Austenite to Martensite, \(T\) is the temperature, \(T_f\) is the transition temperature from Austenite to Martensite, \(\sigma_m\) is an indication of the range of temperature around
the transition temperature $T_{fm}$ during which the phase change occurs, $\sigma$ is the stress, and $K_m$ is the stress curve-fitting parameter which is obtained from the unloading part of the stress-strain characteristic. The parameters $K_a$ and $K_m$ indicate the response of the SMA to application of external stress. This model includes both twinned and detwinned Martensite phases. Since the SMA is modeled as a two-component system, at any given time, the sum of the mole fractions of the Austenite and Martensite phase is 1, i.e.,

$$\xi_a + \xi_m = 1$$  \hspace{1cm} (3)

The time derivatives of (1) and (2) are as follows.

For heating

$$\dot{\xi} = \frac{\xi^2}{\xi_m} \cdot \left[ \exp \left( \frac{T_{fa} - T}{\sigma_a} + K_a \cdot \sigma \right) \right] \cdot \left[ \frac{\dot{T}}{\sigma_a} - K_a \cdot \dot{\sigma} \right]$$  \hspace{1cm} (4)

For cooling

$$\dot{\xi} = \frac{\xi^2}{\xi_a} \cdot \left[ \exp \left( \frac{T_{fm} - T}{\sigma_m} + K_m \cdot \sigma \right) \right] \cdot \left[ \frac{T}{\sigma_m} - K_m \cdot \dot{\sigma} \right]$$  \hspace{1cm} (5)

\[\text{B. Modeling of Temperature Dynamics}\]

The SMA actuator is heated by the process of Joules heating by applying a voltage across the SMA. The loss of heat from the SMA is through natural convection. Mathematically the dynamics of the temperature are given by the following equation.

$$\dot{T} = \frac{1}{mc_p} \left[ \frac{V^2}{R} - hA(T - T_a) \right]$$  \hspace{1cm} (6)

where $m$ is the mass per unit length, $c_p$ is the specific heat capacity, $V$ is the voltage applied across the SMA, $R$ is the resistance per unit length, $h$ is the coefficient of convectional cooling, $A = \pi d$ is the circumferential area of cooling, $d$ is the diameter of the wire, $T$ is the temperature, and $T_a$ is the ambient temperature. The coefficient $h$ is assumed to have the characteristics of a second-order polynomial to enhance the rate of convection at higher temperatures as observed in open-loop results

$$h = h_0 + h_2 T^2$$  \hspace{1cm} (7)

\[\text{C. Constitutive Equation}\]

The constitutive equation relating changes in stress, strain, temperature, and mole fraction is given by the following equation.

$$\dot{\sigma} = D \dot{\varepsilon} + \theta_t \dot{T} + \Omega \dot{\xi}$$  \hspace{1cm} (8)

where $\sigma$ is the stress in the SMA, $D$ is the Young’s modulus of the alloy, $\varepsilon$ is the strain, $\theta_t$ is the thermal expansion factor, $\Omega = -D \varepsilon_i$ is the phase transformation contribution factor, and $\varepsilon_i$ is the initial strain in the SMA. The model is capable of explaining the shape memory and super-elastic properties of Ni-Ti alloys. It should be noted here that in (8), $D$ can be assumed to be the average of the Young’s moduli for the Martensite and Austenite
phases to model the shape memory effect while a more precise Young’s modulus based on the composition of the alloy would be required to model the super-elastic property. This brief, however, deals only with the shape memory effect of Ni-Ti alloys.

Table 1: PARAMETERS OF THE SMA

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass per unit (m in kg.m⁻¹)</td>
<td>4.54e⁻⁴</td>
</tr>
<tr>
<td>Specific heat capacity (c_p in J.kg⁻¹.K⁻¹)</td>
<td>320</td>
</tr>
<tr>
<td>Resistance per unit length (R in Ω.m⁻¹)</td>
<td>13.0677</td>
</tr>
<tr>
<td>Young’s Modulus(Austenite)(D_a in N.m⁻²)</td>
<td>75e⁹</td>
</tr>
<tr>
<td>Young’s Modulus(Martensite)(D_m in N.m⁻²)</td>
<td>28e⁹</td>
</tr>
<tr>
<td>Thermal Expansion (θ in N.m⁻², K⁻¹)</td>
<td>−11e⁻⁶</td>
</tr>
<tr>
<td>SMA initial strain (ε_i)</td>
<td>0.03090</td>
</tr>
<tr>
<td>Heat convection coefficient h₁ in J.m⁻².s⁻¹.K⁻¹</td>
<td>28.552</td>
</tr>
<tr>
<td>Heat convection coefficient h₂ in J.m⁻².s⁻¹.K⁻¹</td>
<td>4.060e⁻⁴</td>
</tr>
<tr>
<td>Diameter of wire (in m)</td>
<td>304e⁻⁶</td>
</tr>
<tr>
<td>Length of wire (in m)</td>
<td>0.24</td>
</tr>
<tr>
<td>Ambient temperature (T_a in °C)</td>
<td>20</td>
</tr>
<tr>
<td>Martensite to Austenite transformation temperature (T_f) in °C</td>
<td>85</td>
</tr>
<tr>
<td>Austenite to Martensite transformation temperature (T_f) in °C</td>
<td>42</td>
</tr>
<tr>
<td>Spread of temperature around T_f (σ_a in °C)</td>
<td>6</td>
</tr>
<tr>
<td>Spread of temperature around T_f (σ_m in °C)</td>
<td>4.5</td>
</tr>
</tbody>
</table>

In addition, the experimental results obtained for the control of the SMA actuator further justify the model since an accurate model is crucial for obtaining good results. The parameters of the model are listed in Table I.

III. STATE-SPACE REPRESENTATION

The dynamic characteristics of the SMA are completely defined by (4) or (5) (heating or cooling), together with (6) and (8). We can also define \( \sigma_e \) as the integral of the error, i.e.,

\[
\dot{\sigma}_e = \epsilon - \epsilon_{ref}
\]

Where \( \epsilon \) is the strain of the SMA actuator and \( \epsilon_{ref} \) is the reference trajectory. The dynamic equations of the SMA along with (9) can be represented in the state-space form

\[
\dot{\bar{z}} = f(\bar{z}, u, t)
\]

Where

\[
\bar{z} = [\epsilon \ T \ \xi \ \sigma_e]^T
\]

and \( u \) is the input voltage to the SMA wire. The nonlinear equations are linearized about a set of operating points \((\epsilon_0 \ T_0 \ \xi_0 \ u_0)\) on the reference trajectory. To obtain the operating points, \( \epsilon_0 \) is chosen as the value of the reference strain at that instant of time. \( T_0 \) and \( \xi_0 \) are obtained by integrating (4) or (5) and (8), depending on whether the SMA is being heated or cooled, for a given value of \( \epsilon_0 \). The value of \( u_0 \) is obtained from (6) for a given value of \( T_0 \), assuming steady-state conditions.

Equation (10) is linearized about the calculated operating points, assuming the no-load case, to obtain linear models in the form
\[
\dot{z} = A\dot{z} + Bu, \quad y = C\dot{z} \tag{11}
\]

Where

\[
A = \begin{bmatrix}
\frac{\partial f}{\partial \dot{z}} & 0 & 0 \\
0 & H_{func} & 0 \\
0 & H_{func} & 0 \\
1 & 0 & 0
\end{bmatrix}_{e_0,T_0,\xi_0,u_0}
\]

\[
B = \begin{bmatrix}
\frac{2u}{m \cdot c_p \cdot R} \left(-\dot{\theta}_t - \Omega \cdot G_{func}\right) \\
\frac{2u}{m \cdot c_p \cdot R} \\
\frac{2u}{m \cdot c_p \cdot G_{func}} \\
0
\end{bmatrix}_{e_0,T_0,\xi_0,u_0}
\]

The closed form expressions of A and B are given as

\[
A = \begin{bmatrix}
0 & 1 & \frac{H_{func}(\theta - \Omega \cdot G_{func})}{\epsilon_0, T_0, \xi_0, u_0}
\end{bmatrix}
\]

\[
B = \begin{bmatrix}
\frac{2u}{m \cdot c_p \cdot R} \left(-\dot{\theta}_t - \Omega \cdot G_{func}\right) \\
\frac{2u}{m \cdot c_p \cdot R} \\
\frac{2u}{m \cdot c_p \cdot G_{func}} \\
0
\end{bmatrix}_{e_0,T_0,\xi_0,u_0}
\]

Where

\[
H_{func} = \frac{-hA - 2h_2AT(T - T_a)}{mc_p} \tag{12}
\]

\[
G_{func} = \frac{K \exp\left(\frac{T_f - T}{\sigma_i}\right)}{\sigma_i \left(1 + \exp\left(\frac{T_f - T}{\sigma_i}\right)\right)^2} \tag{13}
\]

Where \(T_f\) is chosen either as \(T_{fa}\) or \(T_{fm}\) according to whether the SMA actuator is being heated or cooled. Correspondingly, \(\sigma_i\) is chosen as either \(\sigma_m\) or \(\sigma_a\) and \(K\) is chosen as \(\xi_m\) or \(\xi_a\).

For the no-load case, which is the case considered here, \(\sigma\) and \(\dot{\sigma}\) are equal to zero. In this case, the model given by (11) is not controllable since the number of controllable states is only 2. Physically, for a given input current, the SMA reaches a desired temperature resulting in a phase transformation which causes a change in strain as given by the constitutive equation. The three states (temperature, mole fraction, strain) cannot be independently controlled to a desired value. On removing the uncontrollable modes, the following state-space model is obtained

\[
\dot{\tilde{x}} = A'\tilde{x} + B'\tilde{u} \tag{14}
\]

Where

\[
\tilde{x} = \begin{bmatrix}
\tilde{\epsilon} \\
\sigma_e
\end{bmatrix}
\]

The model is evaluated every 100 ms and the corresponding gain matrix is calculated. Since the plant dynamics are comparatively slow, the linear models closely approximate the nonlinear dynamics of the SMA around every operating point implying that the nonlinear dynamics of the system are taken into account in the models.
VI. CONTROLLER

PID control of SMA actuators

In this part, the SMA model is used in the closed loop PID control scheme to investigate the possibility of the PID algorithm in SMA control in the simulation environment before applying it to real control system. The optimal values of PID parameters are determined by using Fuzzy algorithm in order to obtain the desired output displacement. This control strategy is shown in Fig. 2.

The error between the desired displacement and the output of SMA model is the input of PID controller. The output of the PID controller is the current applied to the SMA actuator. The optimal values of PID parameters are obtained by using Fuzzy algorithm which is described in the following section. In this work, the model of SMA actuator is programmed as a Simulink block in order to be used in Simulink Toolbox of Matlab.

Application of self tuning fuzzy PID controller to SMA real time control system

The self tuning fuzzy PID controller is applied to improve the control performance. The structure of the self tuning fuzzy PID controller is shown in Fig. 2. The controller has the form of PID structure, but the PID parameters are tuned by the fuzzy inference, which provides a nonlinear mapping from the error $e(t)$ (the difference between reference and system output), and derivation of error $de(t)$ to the PID parameters $K_p$, $K_i$ and $K_d$. These parameters are tuned from the initial values which are obtained from the simulation.

The stability and robustness analyses of the systems using the fuzzy PID controller have been addressed in several . According to this paper, the Lyapunov function $V(x)$ was firstly established for a closed loop system composed of a plant and a constant parameter PID controller, then the $V(x)$ was also proved to be a Lyapunov function for the closed loop system when the fuzzy PID controller is used. This implied the asymptotic stability of this control structure. This section presents the implementation of the self tuning fuzzy PID controller for the position control of SMA actuators. The structure of the fuzzy inference block in Fig. 2 is shown in Fig. 3.
From the PID parameters obtained by fuzzy algorithm, it is convenient to determine the variable ranges for $K_p$, $K_i$ and $K_d$ as follows: 

$[K_{pmin}, K_{pmax}]$, $[K_{imin}, K_{imax}]$ and $[K_{dmin}, K_{dmax}]$, respectively. The values used in this work are: $K_p \in [0.1, 3]$, $K_i \in [0.5, 5]$ and $K_d \in [0.005, 0.2]$. In order to obtain feasible rule bases with high inference efficiency, the PID parameters must be normalized over the interval $[0, 1]$. Therefore, the real values of the PID parameters are as follows:

$K_p = 2.9K_p' + 0.1$; $K_i = 4.5K_i' + 0.5$; $K_d = 0.195K_d' + 0.005$ ,

where $K_p'$; $K_i'$ and $K_d'$ are the normalized values of $K_p, K_i$ and $K_d$, respectively.

In this paper, the linguistic levels assigned to the input variables $e(t)$ and $\dot{e}(t)$ are as follows: NB: negative big; NM: negative medium; NS: negative small; ZO: zero; PS: positive small; PM: positive medium; PB: positive big. The membership functions of these fuzzy sets are shown in Fig. 3. The maximum error and the maximum derivation of error ranges are chosen from the specification of SMA actuators.
**Fuzzy rule and fuzzy inference**

Using the above fuzzy sets of the input and output variables, fuzzy rules are composed as follows:

Rule $i$: If $e(t)$ is $A_i$ and $\dot{e}(t)$ is $B_i$ then $K'_p$ is $C_i$, $K'_d$ is $D_i$ and $K'_i$ is $E_i$, $i = 1, 2, \ldots, m$, where $m$ is the number of fuzzy rules, $A_i, B_i, C_i, D_i, E_i$ are the $i^{th}$ fuzzy sets of the input and output variables of the fuzzy system. In this tutorial, triangle functions and gaussian function are used as the membership functions, denoted by $\mu$, as shown in Figs. 4 and 5.

Generally, the fuzzy rules are dependent on the plant to be controlled and the type of the controller. These rules are determined from intuition or practical experience. In this tutorial, rules are designed based on the characteristics of SMA actuators, such as slow response or nonlinear hysteresis effect, and the properties of the PID controller.

The rule sets are established and shown in surfaces in Fig. 6. In this paper, the MAX–MIN fuzzy reasoning method is used to obtain the output from the inference rule and present input. For given a specific input fuzzy set $A^l$ in $U$, the output fuzzy set $B^l$ in $O$ for $K'_p$ is computed through the inference engine as follows:

$$
\mu'_B(K'_p) = \max \left[ \sup_{x} \min \left[ \mu'_A(x), \mu'_{R_1}(e(t)), \mu'_{R_2}(de(t)), \mu'_{R_3}(K'_p) \right] \right]
$$

The output membership functions for $K'_d$ and $K'_i$ are computed similarly.

**Defuzzification**

The centroid defuzzification method used in this paper to convert the aggregated fuzzy set to a crisp output value $y^*$ from the fuzzy set $B^l$ in $V \subset R$. This work computes the weighted average of the membership function or the center of gravity (COG) of the area bounded by the membership function curves:
\[ y^* = \int_y \mu_B(y) \cdot y \, dy \]
\[ \int_y \mu_B(y) \, dy \]

We compute the crisp value of \( K_P^1, K_D^1 \) and \( K_I^1 \) by using the above expression

**V. SIMULATION RESULT**

The performance of the proposed control algorithm was investigated for different reference inputs. The sampling time was set to be 0.01 s in all experiments. Fig. 7 shows the performance of the control system with respect to step reference input. From this figure, it can be observed that the self tuning fuzzy PID controller achieves better tracking response than the conventional PID controller without fuzzy tuning.

Figs. 8 and 9 depict the performance of control system respect to sine reference input with two different frequencies. Since the SMA actuator is a slow response system, the control performance is better in the low frequency tracking. The effect of a small variation of input current was also investigated and is shown in Fig. 9. The output displacement keeps good tracking adaptively with the desired trajectory even there is a little change in the input current.

![Figure 7. step response](image7.png)

![Figure 8. System response with respect to sine reference input (f = 0.1 Hz).](image8.png)

![Figure 9. System response with respect to sine reference input (f = 0.33 Hz).](image9.png)

![Figure 10. Effect of the change of input current.](image10.png)
**VI. CONCLUSION**

In this brief, we have developed a new model for SMA actuators based on the physics of the process where we have used the Fermi–Dirac statistical model to represent the two-state process. Based on this model, we have developed and implemented experimentally a controller: self-tuning fuzzy PID controller based on fuzzy rules. The simulation and experimental results in the absence of perturbations show excellent tracking response for the SMA, thereby validating both the model and the control scheme. In particular, the results also clearly justify the use of the model for describing the transformation between Martensite and Austenite phases. The controllers have the ability to reject uncertainties in the parameters of the model.

**VII. REFERENCE**


