Materials Science Forum Vol. 583 (2008) pp 257-275
online at http://www.scientific.net
© (2008) Trans Tech Publications, Switzerland

Shape Memory Alloys: Material Modeling and Device Finite Element Simulations

Ferdinando Auricchio[1,2,3,a], Alessandro Reali[1,2,3,b]

[1]Dipartimento di Meccanica Strutturale, Università degli Studi di Pavia, Italy
[2]European Centre for Training and Research in Earthquake Engineering, Pavia, Italy
[3]Istituto di Matematica Applicata e Tecnologie Informatiche del CNR, Pavia, Italy

[a]auricchio@unipv.it, [b]alessandro.reali@unipv.it

Keywords: shape memory alloys, phenomenological model, device simulation, finite element method.

Abstract. The use of shape memory alloys (SMA) in an increasing number of applications in many fields of engineering, such as biomedical engineering, is leading to a growing interest toward an exhaustive modeling of their macroscopic behavior in order to construct reliable simulation tools for SMA devices.
In this paper we review a robust three-dimensional model able to reproduce both pseudo-elastic and shape-memory behaviors and we report numerical studies where it is used for the simulation of SMA-based biomedical devices.

Introduction

The great and always growing interest in shape memory alloys (SMA) (cf. [1, 2]) and in their industrial applications in many branches of engineering is deeply stimulating the research on constitutive laws. As a consequence, many models able to reproduce one or both of the well-known SMA macroscopic behaviors, referred to as *pseudo-elasticity* and *shape-memory effect*, have been proposed in the literature in the last years (refer for instance to [3–12]). In particular, the constitutive law proposed in [13] and improved in [14] seems to be attractive. Developed within the theory of irreversible thermodynamics, this model is in fact able to describe both pseudo-elasticity and shape-memory effect and the corresponding solution algorithm is simple and robust as it is based on a plasticity-like return map procedure. The robustness of such a model makes it particularly suitable for implementation within finite element codes, allowing in this way the simulation of the behavior of complex SMA devices.

In this paper we review the properties of this model and we report some results of its application in order to perform simulations of SMA-based biomedical devices such as stents, spinal spacers, and micro-grippers.

3D SMA Phenomenological Model

In this first part of the paper we present and discuss in detail the 3D phenomenological model for SMA introduced in [13–15], within the context of thermo-electro-mechanical coupling.

Time-continuous Frame. The model assumes the total strain $\boldsymbol{\varepsilon}$ and the absolute temperature T as control variables, the transformation strain $\mathbf{e}^{tr}$ as internal one. The second-order tensor $\mathbf{e}^{tr}$ describes the strain associated to the transformation between the two solid phases

All rights reserved. No part of contents of this paper may be reproduced or transmitted in any form or by any means without the written permission of the publisher: Trans Tech Publications Ltd, Switzerland, www.ttp.net. (ID: 193.206.66.82-20/05/08,15:00:59)

referred to as martensite and austenite. Here, this quantity has a fully reversible evolution and can be completely recovered when unloading to a zero stress state. Moreover, we require that

$$\|\mathbf{e}^{tr}\| \leq \varepsilon_L, \tag{1}$$

where $\|\cdot\|$ is the usual Euclidean norm and ε_L is a material parameter corresponding to the maximum transformation strain reached at the end of the transformation during an uniaxial test.

Assuming a small strain regime, justified by the fact that the approximation of large displacements and small strains is valid for several applications, the following standard additive decomposition can be considered

$$\boldsymbol{\varepsilon} = \frac{\theta}{3}\mathbf{1} + \mathbf{e},$$

where $\theta = \mathrm{tr}(\boldsymbol{\varepsilon})$ and $\mathbf{e}$ are respectively the volumetric and the deviatoric part of the total strain $\boldsymbol{\varepsilon}$, while $\mathbf{1}$ is the second-order identity tensor. The free energy density function Ψ for a polycrystalline SMA material is then expressed as the convex potential

$$\begin{aligned}\Psi(\theta, \mathbf{e}, T, \mathbf{e}^{tr}) \; &= \frac{1}{2}K\theta^2 + G\|\mathbf{e} - \mathbf{e}^{tr}\|^2 - 3\alpha K\theta(T - T_0) + \beta\langle T - M_f\rangle\|\mathbf{e}^{tr}\| + \\ &+ \frac{1}{2}h\|\mathbf{e}^{tr}\|^2 + (u_0 - T\eta_0) + c\left[T - T_0 - T\log\left(\frac{T}{T_0}\right)\right] + \mathcal{I}_{\varepsilon_L}(\mathbf{e}^{tr}),\end{aligned}$$

where K and G are respectively the bulk and the shear modulus, α is the thermal expansion coefficient, β is a material parameter related to the dependence of the critical stress on the temperature, M_f is the temperature below which only martensite phase is stable, h defines the hardening of the phase transformation, c is the heat capacity, and u_0, η_0 and T_0 are, respectively, the internal energy, the entropy and the temperature at the reference state. Moreover, we make use of the indicator function

$$\mathcal{I}_{\varepsilon_L}(\mathbf{e}^{tr}) = \begin{cases} 0 & \text{if } \|\mathbf{e}^{tr}\| < \varepsilon_L \\ +\infty & \text{otherwise,} \end{cases}$$

in order to satisfy the transformation strain constraint of Eq. 1; we also introduce the positive part function $\langle\cdot\rangle$, defined as

$$\langle a\rangle = \begin{cases} a & \text{if } a > 0 \\ 0 & \text{otherwise.} \end{cases}$$

We remark that, since we use only a single internal variable second-order tensor to describe phase transformations, at most it is possible to distinguish between a generic parent phase (not associated to any macroscopic strain) and a generic product phase (associated to a macroscopic strain). Accordingly, the model does not distinguish between the austenite and the twinned martensite, as both these phases do not produce macroscopic strain.

We furthermore highlight that, for the sake of simplicity, the present model does not reflect the differences existing between the austenite and the martensite elastic properties.

Starting from the free energy function Ψ and following standard arguments, we can derive the constitutive equations

$$\left\{\begin{array}{rcccl} p & = & \dfrac{\partial\Psi}{\partial\theta} & = & K\left[\theta - 3\alpha(T-T_0)\right], \\ \mathbf{s} & = & \dfrac{\partial\Psi}{\partial\mathbf{e}} & = & 2G(\mathbf{e}-\mathbf{e}^{tr}), \\ \eta & = & -\dfrac{\partial\Psi}{\partial T} & = & \eta_0 + 3\alpha K\theta - \beta\|\mathbf{e}^{tr}\|\dfrac{\langle T-M_f\rangle}{|T-M_f|} + c\log\left(\dfrac{T}{T_0}\right), \\ \mathbf{X} & = & -\dfrac{\partial\Psi}{\partial\mathbf{e}^{tr}} & = & \mathbf{s} - \beta\langle T-M_f\rangle\dfrac{\mathbf{e}^{tr}}{\|\mathbf{e}^{tr}\|} - h\mathbf{e}^{tr} - \gamma\dfrac{\mathbf{e}^{tr}}{\|\mathbf{e}^{tr}\|}, \end{array}\right. \tag{2}$$

where $p = \mathrm{tr}(\boldsymbol{\sigma})/3$ and $\mathbf{s}$ are respectively the volumetric and the deviatoric part of the stress $\boldsymbol{\sigma}$, $\mathbf{X}$ is a thermodynamic stress-like quantity associated to the transformation strain $\mathbf{e}^{tr}$, and η is the entropy. The variable γ results from the indicator function subdifferential $\partial\mathcal{I}_{\varepsilon_L}(\mathbf{e}^{tr})$ and it is defined as

$$\left\{\begin{array}{ll} \gamma = 0 & if\ \|\mathbf{e}^{tr}\| < \varepsilon_L, \\ \gamma \geq 0 & if\ \|\mathbf{e}^{tr}\| = \varepsilon_L, \end{array}\right.$$

so that $\partial\mathcal{I}_{\varepsilon_L}(\mathbf{e}^{tr}) = \gamma\mathbf{e}^{tr}/|\mathbf{e}^{tr}\|$.

To describe phase transformation and inelasticity evolution, we choose a limit function F defined as

$$F(\mathbf{X}) = \|\mathbf{X}\| - R \tag{3}$$

where R is the radius of the elastic domain in the deviatoric space.

Considering an associative framework, the flow rule for the internal variable takes the form

$$\dot{\mathbf{e}}^{tr} = \dot{\zeta}\frac{\partial F}{\partial\mathbf{X}} = \dot{\zeta}\frac{\mathbf{X}}{\|\mathbf{X}\|}, \tag{4}$$

where $\dot{\zeta}$ is the consistency parameter.

The model is then completed by the classical Kuhn-Tucker conditions

$$\left\{\begin{array}{c} \dot{\zeta} \geq 0, \\ F \leq 0, \\ \dot{\zeta}F = 0. \end{array}\right. \tag{5}$$

Following classical arguments [16], we can also compute the evolution of the current internal energy u, using the first principle of thermodynamics, as

$$\dot{u} = \dot{\Psi} + T\dot{\eta} + \dot{T}\eta = \boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}} + r - \nabla\cdot\mathbf{q}, \tag{6}$$

being r and $\mathbf{q}$, respectively, the heat source and flux vector.

Moreover, since SMA actuators are often electrically activated by Joule effect, we extend the model to take into consideration also thermo-electro-mechanical coupling. Limiting the discussion to the steady-state, we recall that the electrical problem in an isotropic conductive material with an applied stationary potential difference is described by the following equations

$$\begin{array}{ll} \nabla\cdot\mathbf{j} & = 0, \\ \mathbf{E} & = -\nabla V, \\ \mathbf{j} & = \sigma^{el}\mathbf{E}, \end{array} \tag{7}$$

with $\mathbf{j}$ the current density, $\mathbf{E}$ the electric field, V the electric potential, σ^{el} the electric conductivity constant. Accordingly, between electrical and thermal fields, a coupling takes place through the Joule effect, i.e., when a current density $\mathbf{j}$ passes through a conductive body unit volume, the electric field $\mathbf{E}$ generates a power density in form of a heat production

$$H_{elc} = \mathbf{E} \cdot \mathbf{j}. \tag{8}$$

Hence, under our hypotheses of small deformations, no internal heat source ($r = 0$), and isotropic Fourier law ($\mathbf{q} = -K^{th}\nabla T$, with K^{th} thermal conductivity constant), we can write the thermo-electro-mechanical coupled problem using the free energy definition of Eq. 2 into Eq. 6, thus obtaining

$$c\dot{T} - K^{th}\nabla^2 T = w'_{ch}. \tag{9}$$

We highlight that in the above equation w'_{ch} is the volume heat source, which is defined as

$$w'_{ch} = H_{elc} + D_{mech} + H_{tmc}, \tag{10}$$

where H_{elc} is the heat production due to thermo-electrical coupling defined by Eq. 8, D_{mech} is the heat production due to mechanical dissipation

$$D_{mech} = \boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}} - \left(\frac{\partial \Psi}{\partial \theta}\,\dot{\theta} + \frac{\partial \Psi}{\partial \mathbf{e}}\,\dot{\mathbf{e}} + \frac{\partial \Psi}{\partial \mathbf{e}^{tr}}\,\dot{\mathbf{e}^{tr}} \right), \tag{11}$$

and H_{tmc} is the heat production due to thermo-mechanical coupling (thermo-elastic effect and phase transformation latent heat)

$$H_{tmc} = T \left(\frac{\partial^2 \Psi}{\partial T \partial \theta}\,\dot{\theta} + \frac{\partial^2 \Psi}{\partial T \partial \mathbf{e}}\,\dot{\mathbf{e}} + \frac{\partial^2 \Psi}{\partial T \partial \mathbf{e}^{tr}}\,\dot{\mathbf{e}^{tr}} \right). \tag{12}$$

We finally remark that for this model we have

$$\begin{aligned} D_{mech} &= \mathbf{X} : \dot{\mathbf{e}^{tr}}, \\ H_{tmc} &= T \left(\beta \frac{\mathbf{e}^{tr}}{\|\mathbf{e}^{tr}\|} : \dot{\mathbf{e}^{tr}} - 3K\alpha\dot{\theta} \right). \end{aligned} \tag{13}$$

Observation 1. The proposed model is thermodynamically consistent since it satifies the second principle of thermodynamics in the form of the Clausis-Duhem inequality (for more details see [14]).

Observation 2. By exploiting basic Convex Analysis tools (see, e.g., [17]) we can rewrite our constitutive model defined by Eqs. 2–5 in the equivalent form

$$\begin{pmatrix} -p \\ -\mathbf{s} \\ \eta \\ \partial D \dot{\mathbf{e}}^{tr} \end{pmatrix} + \partial \Psi \begin{pmatrix} \theta \\ \mathbf{e} \\ T \\ \mathbf{e}^{tr} \end{pmatrix} \ni \mathbf{0}. \tag{14}$$

Here ∂D stands for the subdifferential of the function D defined as

$$D(\mathbf{e}^{tr}) = \sup_{F(\mathbf{A}) \leq 0} \left\{ \mathbf{A} : \mathbf{e}^{tr} \right\}, \tag{15}$$

which is the dissipation function associated to the phase transformation mechanism ($\mathbf{A}$ is a generic thermodynamic stress-like tensor). It can be shown that $D(\mathbf{e}^{tr}) = R||\mathbf{e}^{tr}||$, as well as that it is the Fenchel-Legendre conjugate of the indicator function of the non-empty, convex, and closed domain

$$\mathcal{E} = \{\mathbf{A} \,:\, F(\mathbf{A}) \leq 0\}.$$

Hence, it is easy to check that D is positively 1-homogeneous, that is

$$D(\lambda \mathbf{e}^{tr}) = \lambda D(\mathbf{e}^{tr}) \quad \forall \lambda > 0.$$

Namely, the time-evolution of $\mathbf{e}^{tr}$ is of rate-independent type since we readily have that

$$\partial D(\lambda \mathbf{e}^{tr}) = \partial D(\mathbf{e}^{tr}) \quad \forall \lambda > 0.$$

The formulation of rate-independent evolution problems in terms of a doubly-nonlinear differential inclusion as in Eq. 14 has recently attracted a good deal of attention. In particular, the mathematical treatment of relations as Eq. 14 is nowadays fairly settled and existence, uniqueness, and time-discretization results are available. The interested reader is referred to the recent survey [18] where a comprehensive collection of mathematical results on doubly-nonlinear rate-independent problems is provided.

Observation 3. We explain here (as done in [19]) the model capability to undergo fully reversible phase transformations.

We focus on the case $\beta\langle T - M_f\rangle > R$, which is indeed the most interesting, and for simplicity we start considering scalar quantities instead of second-order tensors. This is equivalent to study a proportional loading process (i.e. developed along a fixed direction) so that the scalars X, s, and e^{tr} assume the physical meaning of norms for the corresponding tensor-valued quantities. Then, we suppose to be in the condition $F = 0$, that is

$$|X| = \left| s - \beta\langle T - M_f\rangle \frac{e^{tr}}{|e^{tr}|} - h e^{tr} \right| = R, \tag{16}$$

which implies

$$s = \beta\langle T - M_f\rangle \frac{e^{tr}}{|e^{tr}|} + h e^{tr} \pm R. \tag{17}$$

Fig. 1 reports the graphical representations of the relation $s = s(e^{tr})$ obtained from Eq. 17 for the two cases of unloading from compression and unloading from tension, respectively. The figure shows that for each case there exists an interval for s inside which $e^{tr} = 0$. Considering the intersection of the intervals for the two cases, it is possible to conclude that

$$s \in [-\beta\langle T - M_f\rangle + R, \beta\langle T - M_f\rangle - R] \Rightarrow e^{tr} = 0, \tag{18}$$

i.e., if s belongs to the interval $[-\beta\langle T - M_f\rangle + R, \beta\langle T - M_f\rangle - R]$, e^{tr} is necessarily equal to zero. Accordingly, this implies that, due to the continuity of the considered functions, if we are coming from a state with s outside the indicated interval and we are unloading, e^{tr} necessarily approaches zero as s approaches the extreme of such an interval.

Indeed, the very same conclusion holds also in a 3D framework. Still referring to the unloading situation, one can prove that, whenever $\|\mathbf{s}\|$ approaches $\beta\langle T - M_f\rangle - R$ and $\mathbf{e}^{tr}$ is such that

$$\|\mathbf{X}\| = \left\| \mathbf{s} - \beta\langle T - M_f\rangle \frac{\mathbf{e}^{tr}}{\|\mathbf{e}^{tr}\|} - h\mathbf{e}^{tr} \right\| = R,$$

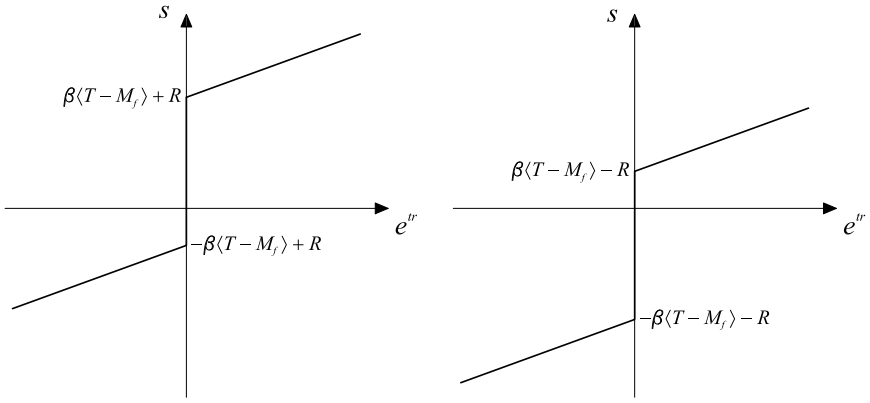

Fig. 1: Plots of $s = s(e^{tr})$ in the cases of unloading from compression (left) and from tension (right).

then $\|\mathbf{e}^{tr}\|$ tends to zero. A proof follows.

Assume that this is not the case. Hence, there exists $\varepsilon > 0$ such that, for all $\alpha > 0$, there exist $\mathbf{s}_\alpha, \mathbf{e}^{tr}_\alpha$ such that

$$\|\mathbf{s}_\alpha\| - \beta\langle T - M_f\rangle + R < \alpha,$$

$$\left\|\mathbf{s}_\alpha - \beta\langle T - M_f\rangle \frac{\mathbf{e}^{tr}_\alpha}{\|\mathbf{e}^{tr}_\alpha\|} - h\mathbf{e}^{tr}_\alpha\right\| = R,$$

but

$$\|\mathbf{e}^{tr}_\alpha\| > \varepsilon.$$

Then, it suffices to choose $\alpha \leq h\varepsilon$ and check that

$$-\|\mathbf{s}_\alpha\| > -\beta\langle T - M_f\rangle + R - \alpha,$$

$$\left\|\beta\langle T - M_f\rangle \frac{\mathbf{e}^{tr}_\alpha}{\|\mathbf{e}^{tr}_{\alpha_*}\|} + h\mathbf{e}^{tr}_\alpha\right\| > \beta\langle T - M_f\rangle + h\varepsilon.$$

Now, we take the sum of the latter relations and exploit the Lipschitz continuity of the norm in order to obtain that

$$R < \left\|\beta\langle T - M_f\rangle \frac{\mathbf{e}^{tr}_\alpha}{\|\mathbf{e}^{tr}_\alpha\|} + h\mathbf{e}^{tr}_\alpha\right\| - \|\mathbf{s}_\alpha\| \leq \left\|\mathbf{s}_\alpha - \beta\langle T - M_f\rangle \frac{\mathbf{e}^{tr}_\alpha}{\|\mathbf{e}^{tr}_\alpha\|} - h\mathbf{e}^{tr}_\alpha\right\| = R,$$

which is clearly a contradiction and proves the correctness of our thesis.

Time-discrete Frame. Let us now focus on the crucial issue of computing the stress and internal variable evolution of a SMA sample in a strain-driven situation. We shall directly concentrate ourselves on the solution of the time-incremental problem. Namely, we discretize the time-interval of interest $[0, t_f]$ by means of the partition $I = \{0 = t_0 < t_1 < ... < t_{N-1} < t_n = t_f\}$, assume to be given the state of the system $(p_n, \mathbf{s}_n, \eta_n, \mathbf{e}^{tr}_n)$ at time t_n, the actual total

strain $(\theta, \mathbf{e})$ and temperature T at time t_{n+1} (note that for notation simplicity here and in the following we drop the subindex $n+1$ for all the variables computed at time t_{n+1}), and solve for $(p, \mathbf{s}, \eta, \mathbf{e}^{tr})$. For the sake of numerical convenience, instead of solving Eq. 2 we prefer to perform some regularization. Indeed, we let $\overline{\|\cdot\|}$ be defined as

$$\overline{\|\mathbf{a}\|} = \sqrt{\|\mathbf{a}\|^2 + \delta} - \sqrt{\delta},$$

(δ is a user-defined parameter controlling the smoothness of the norm regularization) and introduce the regularized free energy density $\overline{\Psi}$ as

$$\begin{aligned}\overline{\Psi}(\theta, \mathbf{e}, T, \mathbf{e}^{tr}) &= \frac{1}{2}K\theta^2 + G\|\mathbf{e} - \mathbf{e}^{tr}\|^2 - 3\alpha K\theta(T - T_0) + \beta\langle T - M_f\rangle\overline{\|\mathbf{e}^{tr}\|} + \\ &+ \frac{1}{2}h\|\mathbf{e}^{tr}\|^2 + (u_0 - T\eta_0) + c\left[T - T_0 - T\log\left(\frac{T}{T_0}\right)\right] + \mathcal{I}_{\varepsilon_L}(\mathbf{e}^{tr}).\end{aligned}$$

Finally, the updated values $(p, \mathbf{s}, \eta, \mathbf{e}^{tr})$ for regularized constitutive model can be computed from the following relations

$$\begin{cases} p &= K[\theta - 3\alpha(T - T_0)], \\ \mathbf{s} &= 2G(\mathbf{e} - \mathbf{e}^{tr}), \\ \eta &= \eta_0 + 3\alpha K\theta - \beta\overline{\|\mathbf{e}^{tr}\|}\dfrac{\langle T - M_f\rangle}{|T - M_f|} + c\log\left(\dfrac{T}{T_0}\right), \\ \mathbf{X} &= \mathbf{s} - \beta\langle T - M_f\rangle\dfrac{\mathbf{e}^{tr}}{\sqrt{\|\mathbf{e}^{tr}\|^2 + \delta}} - h\mathbf{e}^{tr} - \gamma\dfrac{\mathbf{e}^{tr}}{\overline{\|\mathbf{e}^{tr}\|}}, \\ \mathbf{e}^{tr} &= \mathbf{e}^{tr}_n + \Delta\zeta\dfrac{\mathbf{X}}{\|\mathbf{X}\|}, \\ F &= \|\mathbf{X}\| - R, \end{cases} \tag{19}$$

along with the requirements

$$\begin{cases} \gamma &\geq 0, \\ \|\mathbf{e}^{tr}\| &\leq \varepsilon_L, \\ \Delta\zeta &\geq 0, \quad F \leq 0, \quad \Delta\zeta F = 0, \end{cases} \tag{20}$$

where $\Delta\zeta = \zeta - \zeta_n = \int_{t_n}^{t_{n+1}} \dot{\zeta}\mathrm{d}t$ is the time-integrated consistency parameter.

We shall clearly state that our numerical experiments are not performed on the discrete version of the non-regularized model described in the time-continuous frame, but rather on the above-introduced δ-regularized discrete version. This choice turns out to be quite convenient from the numerical viewpoint and preserves most of the characteristic features of the model. Moreover, it can be proved that the δ-regularized model converges to the original one as the regularization parameter δ goes to 0. This fact along with additional mathematical analysis of the model is addressed in [20].

Solution Algorithm. The solution of the discrete model is performed by means of an elastic-predictor inelastic-corrector return map procedure as in classical plasticity problems (cf. [21]). An elastic trial state is evaluated keeping frozen the internal variables, then a trial value of the limit function is computed to verify the admissibility of the trial state. If this is not verified, the step is inelastic and the evolution equations have to be integrated.

We remark that we distinguish two inelastic phases in our model: a non-saturated phase ($\|\mathbf{e}^{tr}\| < \varepsilon_L$, $\gamma = 0$) and a saturated one ($\|\mathbf{e}^{tr}\| = \varepsilon_L$, $\gamma \geq 0$). In our solution procedure we start assuming to be in a non-saturated phase, and when convergence is attained we check if our assumption is violated. If the non-saturated solution is not admissible, we search for a new solution considering saturated conditions.

For each inelastic step, we have to solve the nonlinear system of Eq. 19. Then, also the equations related to electrical and thermal fields have to be solved taking into account thermo-electro-mechanical coupling. However, as the aim of this part of the paper is just to introduce the model that we will use for our subsequent developments, without focusing on algorithmic problems, we do not describe here the algorithm that we implemented in FEAP [22] to find a solution to such nonlinear system. We just say that we used a staggered formulation where we consider a mechanical, an electrical, and a thermal partition, and we refer interested readers to [15] for more details.

Observation 4. In the same spirit of Observation 2, we shall now recast the aforementioned algorithm defined by Eqs. 19 and 20 in terms of dissipation. Exactly as above, we assume to be given the current state of the system $(p_n, \mathbf{s}_n, \eta_n, \mathbf{e}_n^{tr})$ at time t_n and the actual total strain $(\theta, \mathbf{e})$ and temperature T at time t_{n+1}. Then, Eqs. 19 and 20 are nothing but the Euler-Lagrange relations for the following minimum problem

$$\min_{\mathbf{e}_*^{tr}} \left\{ D(\mathbf{e}_*^{tr} - \mathbf{e}_n^{tr}) + \overline{\Psi}(\theta, \mathbf{e}, \eta, \mathbf{e}_*^{tr}) \right\} \tag{21}$$

along with Eqs. 19_1–19_4.

The minimum problem of Eq. 21 corresponds in this setting to the Euler method where, nevertheless, the usual incremental quotients are replaced by the weaker distance $D(\mathbf{e}_*^{tr} - \mathbf{e}_n^{tr})$.

Following [18] and [20], it can be shown that the problem of Eq. 21 is uniquely solvable and that the incremental solutions arising from the step-by-step solution of the minimization problem converge to a time-continuous solution of the constitutive relation as the diameter of the time partition I goes to zero. Moreover, the model is stable with respect to the regularization parameter $\delta > 0$. In particular, solutions to the incremental problem of Eq. 21 converge to the unique minimizer of problem

$$\min_{\mathbf{e}_*^{tr}} \left\{ D(\mathbf{e}_*^{tr} - \mathbf{e}_n^{tr}) + \Psi(\theta, \mathbf{e}, \eta, \mathbf{e}_*^{tr}) \right\}, \tag{22}$$

as δ goes to zero.

Numerical Tests on the Model

To show the model capability of reproducing the macroscopic behavior of SMA materials, we perform a couple of proportional tension-compression stress-driven (temperature-parameterized) numerical experiments. For a complete testing of the model we refer to [14, 15]. We consider the material properties specified in Table 1, describing an idealized Cu-based SMA; E and ν are respectively the Young's modulus and the Poisson's ratio, while all the other material constants have already been introduced previously.

In the first experiment, we test the pseudo-elastic effect, i.e. we perform a tension-compression uniaxial test at a constant temperature $T = 313$ K $> M_f$. As it is possible to see in Fig. 2, the pseudo-elastic effect is perfectly reproduced and a complete strain recovery is obtained as the stress is driven to zero.

parameter	value	unit
E	$53 \cdot 10^3$	MPa
ν	0.3	-
β	2.1	MPa K^{-1}
M_f	293	K
h	1000	MPa
R	20	MPa
ε_L	4	%
δ	10^{-8}	-

Table 1: Material parameters for the numerical tests.

The second experiment aims at reproducing the shape-memory effect and consists of a tension-compression uniaxial test performed at a temperature $T = M_f$. The result is reported in Fig. 3, where it is shown that at the end of the loading cycle a residual strain is obtained. Such residual strain is however completely recovered heating the material (here we heat it up to $T = 313$ K).

We finally highlight that we performed both our tests choosing different stress-increments in order to prove the robustness of the algorithm. In particular, we used 0.1 MPa and 10 MPa increments, which are represented on the figures using solid lines and dots, respectively. Since dots always lie on solid lines, the robustness of the algorithm is evident.

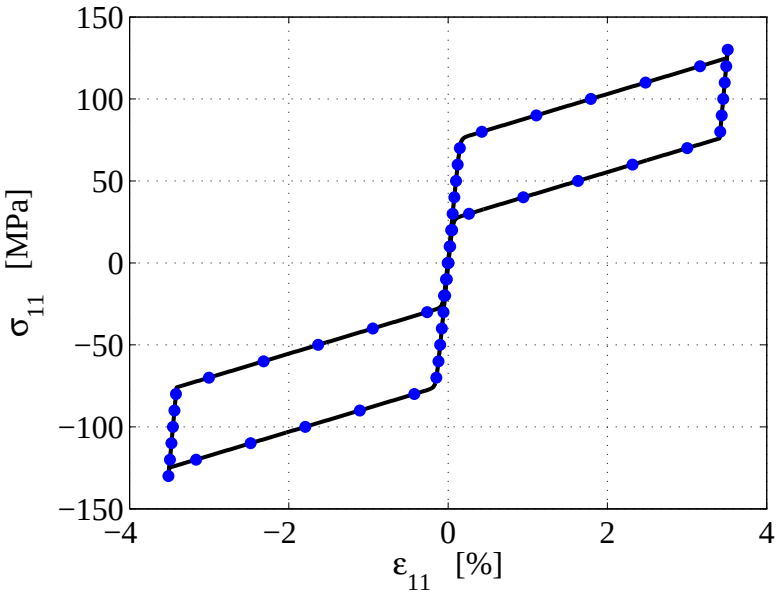

Fig. 2: Pseudo-elastic effect. $\sigma_{11} - \epsilon_{11}$ response to a tension-compression cycle at $T = 313$ K $> M_f$. Solid line: $\Delta\sigma = 0.1$ MPa; dots: $\Delta\sigma = 10$ MPa.

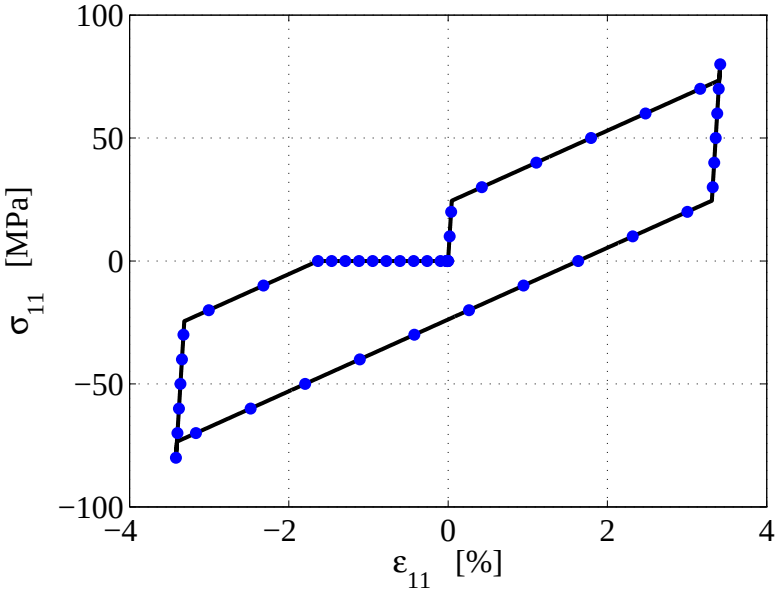

Fig. 3: Shape-memory effect. $\sigma_{11} - \epsilon_{11}$ response to a tension-compression cycle at $T = M_f$, followed by heating strain recovery up to $T = 313$ K ($\Delta T = 1$ K). Solid line: $\Delta\sigma = 0.1$ MPa; dots: $\Delta\sigma = 10$ MPa.

Finite Element Simulation of SMA-based Devices

In this second part of the paper, we focus on the simulation of devices, based on the peculiar properties of SMA, using within a finite element framework the constitutive model just presented. In particular, we show some numerical results obtained in the last years on the modeling of some biomedical SMA-based devices such as stents, spinal spacers and micro-grippers.

Simulation of the Behavior of a Self-expandable SMA Stent. In this section we report some numerical studies on the behavior of a self-expandable SMA stent, taken from [23]. We start with a brief introduction on this kind of biomedical device. More details and numerical results on this topic, as well as many references to the related literature, can be found in [23].

A vascular stent is a small metal tube, which is inserted into an artery at the site of a narrowing to act as internal scaffolding or as a support to the blood vessel. Two types of coronary stents are available on the market based on two different expansion principles: balloon-expandable and self-expandable stents. The former is mounted on a catheter supporting a balloon and is positioned by inflating the balloon in the site of blockage. Under the pressure of the balloon, the stent deploys itself to mechanically support the vessel wall. When the balloon is deflated and retracted, the stent remains in place and keeps the artery open. On the other hand, the self-expandable stent is mounted on a catheter and then compressed by a protective sheath. When the sheath is retracted the stent expands by itself into the artery. In both cases, in the weeks following the insertion of the stent, a thin film of cells grows over the metallic frame lining the walls of the artery like a skin. Until the skin has grown over the metal there is a risk for blood elements to stick to the stent. This is a phase of the in-stent restenosis process, which includes an inflammatory phase, a granulation or cellular proliferation phase, and a phase of remodeling involving extracellular matrix protein synthesis. It is worthwhile noting that the restenosis is also present after an intervention of angioplasty, but there are profound differences

between vascular biological responses to balloon- and stent-induced injury. If compared to angioplasty, stenting improves the short-term success rate and the safety of the percutaneous coronary intervention. In the longer term, it decreases the restenosis and the need for repeated revascularization. However, stenting does not eliminate restenosis.

Moreover, many observations present in the literature support the notion that wall stress is a potent stimulus for pathophysiological adaptations of the vessel wall. As a consequence the knowledge of the stress state generated during the intervention may help in understanding some aspects of the restenotic process, which could be as one of the outcome of a modified mechanical stress state. It is well known, indeed, that atherosclerotic lesions occur where there are local variations in the arterial mechanical and fluid dynamics environments. The outward force of the stent against the vessel wall creates non-physiologic stresses and strains. The initial induced process is an acute elastodamage with abnormal stresses, followed by a stress-mediate injury response. In this light, the stent design as well as the modality of stent expansion could produce a different arterial response to the mechanical action induced by the stent itself. The mechanical stresses produced during the expansion should be reduced as much as possible not to hinder the beneficial effects of the treatment.

Indeed, stress-strain analysis, force and tension calculations are important parameters in evaluating the response to these implants. In the scientific literature there has been an increasing effort to study angioplasty and stenting procedures by means of computational structural analysis with the aim of predicting and calculating the stress state generated after a percutaneous coronary intervention.

The following numerical tests (originally reported in [23]) aim at investigating the stress state induced on the vascular wall by the expansion of a SMA stent using the finite element method.

A geometry very similar to the SciMED Radius stent (SciMED Live System, Maple Grove MN, USA) is considered. Its peculiar structure, with tubular-like rings, having the main function to sustain the vessel after the stent expansion, and bridging members (links), having the main function to assure the stent flexibility by allowing mutual rotation between adjacent rings, allows to study only a stent unit composed by two rings and the links between them. Artery and plaque are modeled with 6338 and 6177 3-D 10-nodes hybrid modified tetrahedral elements, respectively: a stenosis of 36% is considered. The constitutive reference model of the plaque is used. Both the extremities of the artery are constrained along the longitudinal direction to simulate the remaining part of the vessel. The stent is discretized by means of 7144 3D 10-nodes hybrid modified tetrahedral elements with a corresponding number of nodes of 17585 (see Fig. 4). The constitutive model described in the first part of the paper is used with NiTi typical material parameters. The stent rigid body motion is avoided constraining three points, positioned in the middle of the three links respectively, in the longitudinal and tangential directions. The numerical results of a simulation of a stent self-expansion showing the Von Mises stresses induced on the vessel are finally reported in Fig. 5. This data, as highlighted above, are fundamental in order to evaluate the response, and the effectiveness, of these implants.

The commercial code ABAQUS (Hibbit Karlsson & Sorenses, Inc., Pawtucket, RI, USA) with the large deformation option is used to provide the non-linear force/displacement response of the device.

Simulation of the Behavior of a SMA Spinal Spacer. In this section we report some numerical studies on the behavior of a SMA spinal spacer, taken from [24]. More details on

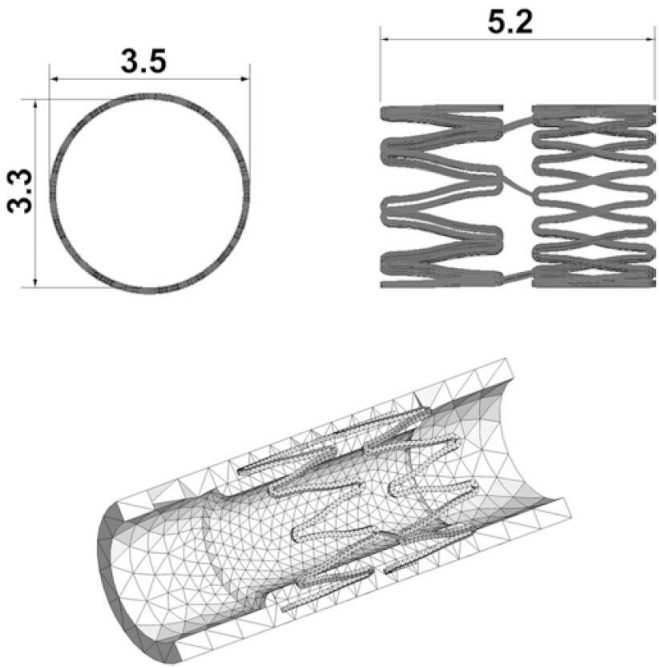

Fig. 4: CAD model of a unit of the "SciMED Radius" stent (top) and entire model with artery and plaque (bottom). The figure is taken from [23].

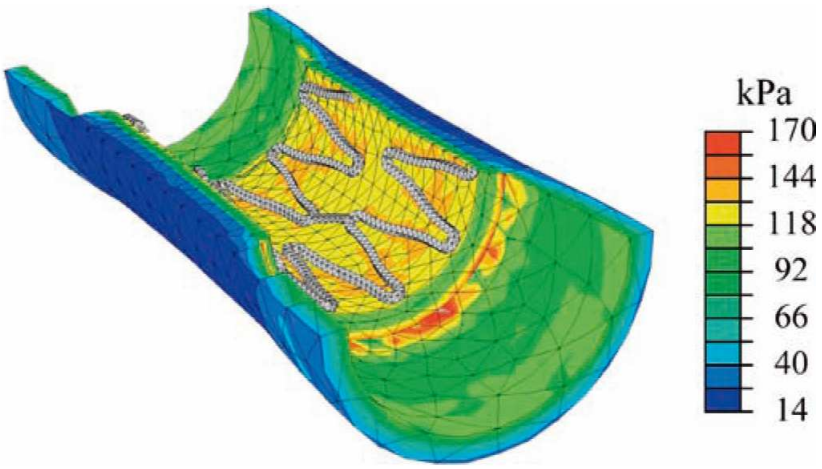

Fig. 5: Von Mises stress distribution induced on the vessel by the self-expansion of the "SciMED Radius" SMA stent. The figure is taken from [23].

the simulations we show in the following, as well as references to the related literature, can be found in [24].

The intervertebral disk has several important functions, including functioning as a spacer, as a shock absorber and as a motion unit: it maintains the separation distance between adjacent vertebral bodies allowing spine large motion in several directions as well as avoiding nerve compression; it allows the spine to compress and rebound during activities as jumping and running and it resists the downward pull of gravity on the head and trunk during prolonged sitting and standing; last, it allows the spinal segment to flex, rotate, and bend on the side. A spinal vertebrae spacer is a device that should be able to substitute for a damaged intervertebral disk. Accordingly, metals such as stainless steel and titanium, characterized by high stiffness, if compared with biological materials, and low damping are not suitable for realizing spacers. The higher compliance and the ability of damping of NiTi alloys, coupled with the possibility of reducing the device dimension, crimping it during implantation and then recovering the original shape through shape memory effect, indicates SMA as the best choice for manufacturing effective spinal vertebrae spacers.

We now show a model which is used to compare the response of different NiTi spinal spacers during implant and physiological loading history (see [24]). The device is compressed in martensitic phase, assuming a reduced shape which helps the insertion between the vertebrae. Once positioned into the body, it recovers its original expanded shape by residual deformation thermal recovery (shape-memory effect) and starts to work opposing force to spinal compressive load. Different spacers, in material and geometry, are considered to compare different possible design solutions. In particular, two NiTi alloys (referred to as SMA1 and SMA2, cf. [24]) are investigated. Using SMA1, which has M_f at body temperature (M_f =311 K), the spacer is compressed and implanted when in martensitic phase and hence it is heated to recover the original shape into the body. Using SMA2, which has M_f lower than room temperature (M_f =289 K), the spacer is initially cooled up to M_f to perform compression and implantation in martensitic phase and hence it returns to the body temperature recovering the original shape. The geometry of the spacer (see Fig. 6) is chosen referring to pictures available in the literature and three different models (referred to as A, B, and C) are compared with some different dimensions in thickness and holes (see again [24]). The models are discretized by means of ten-node tetrahedral elements: the number of elements is 3336, 3163, and 3313 and the number of nodes is 6233, 5653, and 5917, for model A, B, and C, respectively. These values are suggested by the sensitivity analysis performed for each device varying the number of elements between about 1500 and 6000. Boundary conditions are chosen in order to replace symmetry conditions on the planes 23 and 13 (cf. Fig. 6): only one fourth of the device in the circumferential direction is studied. Moreover, one of the nodes belonging to the symmetry plane is constrained in direction 3 to avoid rigid body motion, allowing structure transversal deformation. An analytical rigid surface moving against the spacer is considered to simulate both the device compression performed before the implantation and the following compressive action of the adjacent vertebral bodies.

Fig. 7 summarizes the device (material type SMA1, model A) response in terms of displacement history of a central node of the model during crimping, thermal shape recovery and physiological loading steps. For the most significant steps, deformed versus undeformed shapes are also reported (insets of Fig. 7). Fig. 8 compares the responses obtained for the spacer changing geometry data (models A, B, and C) in the case of material type SMA2. In particular, the performed simulations highlight the excessive flexibility of spacer A, unable to support the physiological force, versus the good and quite similar performance of spacers B and C. More-

over, a lack of differences between SMA1 and SMA2 spacers is detected from the mechanical point of view. The preferred choice depends on the working condition and on the physiological temperature range. Finally, the implemented routine allows to catch the temperature dependence of the spacer response and to identify the zones where the transformation takes place and eventually to optimize the device shape. Indeed, the shape-memory effect is maximized as much as the transformation is diffused without overcoming the limit transformation strain ε_L.

Also in this tests, the commercial code ABAQUS (Hibbit Karlsson & Sorenses, Inc., Pawtucket, RI, USA) with the large deformation option is used to provide the non-linear force/displacement response of the device.

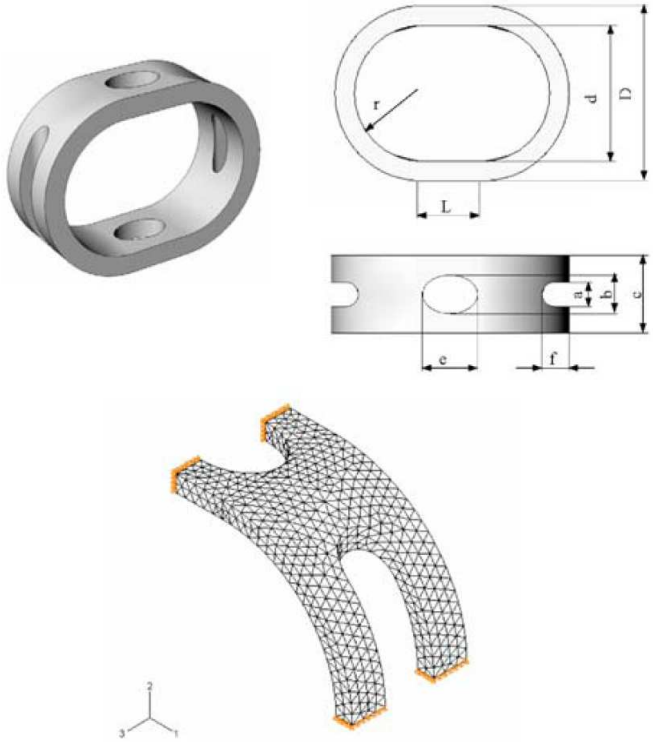

Fig. 6: Three-dimensional CAD model of the spinal vertebrae spacer. Due to the symmetry only one fourth of the device is simulated. The mesh of the spacer and the boundary conditions are reported below. The figure is taken from [24].

Simulation and Design of SMA Micro-grippers. In this section we show some numerical studies on the behavior of SMA micro-grippers. Details on design procedures and more numerical results can be found in [25].

Micro-grippers belong to the family of micro-actuators and their main applications in the biomedical field can be of two types, both in mini-invasive surgery. The first one is as a micro-device directly used to obtain tissue samples during an operation; for this use it is important the capability of the device to be very small but to show high gripping forces (and to this end the properties of NiTi SMA seem to be particularly suitable). A second application is to use the gripper not directly as surgical instrument, but as a device to be utilized in the assemblage of other devices (for instance, to assemble laser devices for laparoscopy operations).

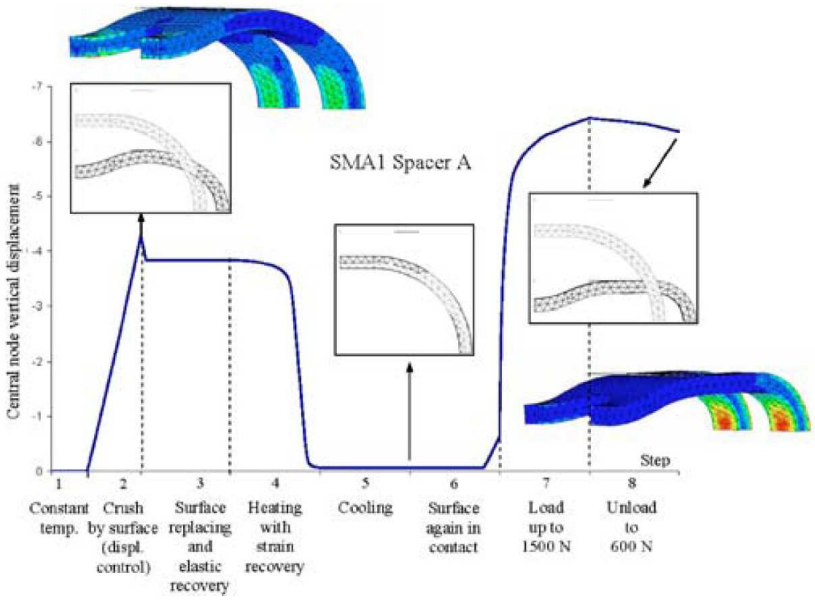

Fig. 7: SMA1 Spacer A: central node vertical displacement (in mm) at different steps and some deformed configurations (black color in the inset). The corresponding Von Mises contour maps of the deformed configurations are reported as well. The figure is taken from [24].

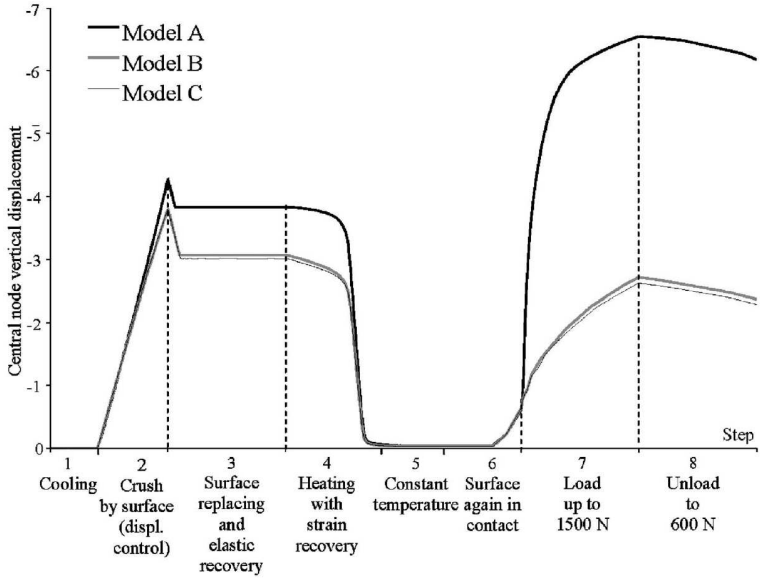

Fig. 8: SMA2 Spacer A, B, and C: central node vertical displacement (in mm) at different steps. The figure is taken from [24].

Independently on the final goal of the device, a well-designed micro-gripper has to show the following desired properties:

- very small dimensions to make possible mini-invasive surgical applications and the handling of tiny objects;
- simple activation system, allowing also a possible position control;
- practical operability frequencies, i.e. "reasonable" device opening and closing times;
- high "enough" gripping force.

In a recent work [25], an interesting design strategy for SMA micro-grippers has been proposed. The first studied device is taken from the literature [26] and consists of a monolithic gripper. Then, a non-monolithic version of the same device has been designed and analyzed as well. In the following we present the simulations performed for the two versions of micro-gripper under consideration. The shape of the overall device (cf. [26]) is shown in Fig. 9, where we report its geometrical properties on the coarsest mesh we created for the model in the finite element code FEAP [22], which is the program used to perform all the numerical simulations. We highlight that including the constitutive model in FEAP (and similarly in ABAQUS) is quite simple, due to the presence of a user material model interface which manages all the needed history variables (as explained in [22]); so, at each Gauss point, it is only required to implement the constitutive law which computes the stress tensor given the strain and to compute the consistent material tangent.

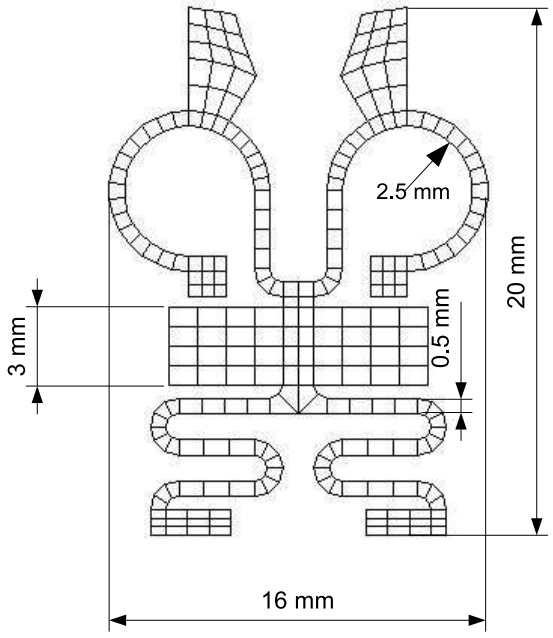

Fig. 9: Shape and size of the micro-gripper.

We first started with a study of the monolithic micro-gripper proposed by [26]. It is composed of two main components, i.e. a *Gear Actuator*, which can transform a linear force into a gripping force between its jaws, and a *Linear Actuator*, which can provide a linear force to be supplied

to the Gear Actuator. After designing the micro-gripper (as detailed in [25]), we simulated the behavior of the whole device, obtaining the results reported in Fig. 10. It is possible to observe an actual movement of the gripper to the equilibrium points we have found during the design of the device. Moreover, we remark that the two actuators are always activated independently and they are never in the austenitic condition at the same time.

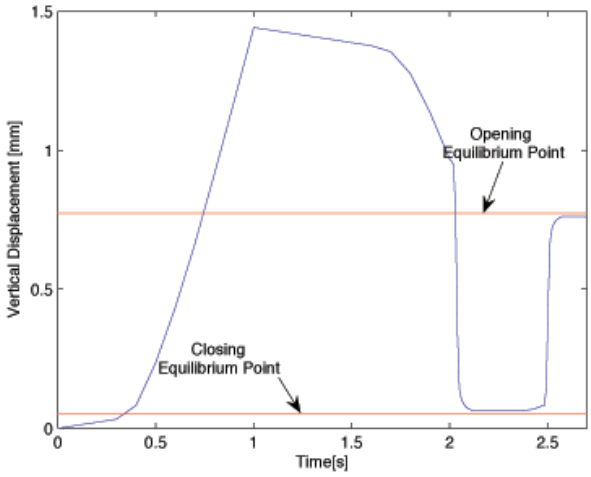

Fig. 10: Results of the simulation of a monolithic micro-gripper.

We then decided to work on a non-monolithic variant of the device, obtained replacing the gear actuator with a pseudo-elastic component with the same shape and size. We propose to heat the linear actuator using a thermoelectric component known as Peltier device, which can be used both for heating and cooling. Thanks to such a device, able to heat and cool off the linear actuator, we can control the temperature the actuator itself would reach. Thus we can think of a gripper which has more than one closing condition (that is, the condition where the jaws get closer; we remark that no contact with an object is considered for this condition), in fact the temperature of the linear actuator determines how much the gripper jaws would close: the higher the temperature of the actuator, the more the jaws will close. So, we were able to design a gripper with two closing positions (cf. again [25]), the extension to the case of multi-closing conditions being straightforward. We finally simulated the overall behavior of the device and the results are those provided in Fig. 11. As it is shown, we observe the behavior for which we designed the device: i.e., we can actually close partially the device and then close it completely; however, we cannot achieve a partial reopening of the device once we have closed it completely.

Summary

In the present paper, an effective 3D constitutive model for describing the macroscopic behavior of SMA within a thermo-electro-mechanical framework has been described (see [13–15]). Such a model has been successfully employed in order to simulate and study the behavior of SMA biomedical devices. In particular, simulations of SMA-based stents, spinal spacers, and micro-grippers are reported and commented.

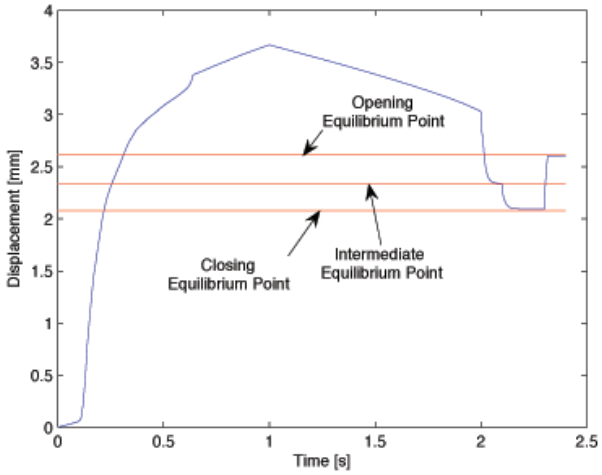

Fig. 11: Results of the simulation of a non-monolithic micro-gripper with two closing conditions.

Acknowledgements

For the fruitful discussions on the subject of the present paper as well as for providing the figures and data of their previous works reported in this paper, the authors would like to thank Prof. Gabriele Dubini, Mr. Paolo Massarotti, Prof. Francesco Migliavacca, Dr. Lorenza Petrini, and Ms. Silva Schievano (Politecnico di Milano); Dr. Ulisse Stefanelli (IMATI-CNR, Pavia); and Mr. Angelo Tardugno (Imperial College, London).
This work has been partially supported by Regione Lombardia through the INGENIO research program "Laser processing of SMA micro-devices"; by the Ministero dell'Istruzione, dell'Università e della Ricerca (MiUR) through the PRIN research program "Shape memory alloy active microactuators and devices for biomedical applications: constitutive modeling, structural analysis, design, use of laser techniques for prototype implementation and experimental validation"; and by the European Science Foundation through the EUROCORES S3T project "SMARTeR: Shape Memory Alloys to Regulate Transient Responses in civil engineering".

References

[1] T.W. Duerig, K.N. Melton, D. Stoekel and C.M. Wayman: *Engineering aspects of shape memory alloys*, Butterworth-Heinemann (1990).

[2] T.W. Duerig and A. Pelton: *SMST-2003 Proceedings of the International Conference on Shape Memory and Superelastic Technology Conference*, ASM International (2003).

[3] C. Bouvet, S. Calloch and C. Lexcellent,: Eur. J. Mech. A/Sol. Vol. 23 (2004), pp. 37–61.

[4] S. Govindjee and C. Miehe: Comput. Methods Appl. Mech. Engrg. Vol. 191 (2001), pp. 215–238.

[5] D. Helm and P. Haupt: Int. J. Sol. Struct. Vol. 40 (2003), pp. 827–849.

[6] D.C. Lagoudas and P. Entchev: Mech. Mater. Vol. 36 (2004), pp. 865–892.

[7] S. Leclercq and C. Lexcellent: J. Mech. Phys. Sol. Vol. 44 (1996), pp. 953–980.

[8] V.I. Levitas: Int. J. Sol. Struct. Vol. 35 (1998), pp. 889–940.

[9] V.I. Levitas and D.L. Preston: Phys. Rev. B Vol. 66 (2002), pp. 134206:1–9.

[10] V.I. Levitas and D.L. Preston: Phys. Rev. B Vol. 66 (2002), pp. 134207:1–15.

[11] B. Peultier, T. Benzineb and E. Patoor: J. Phys. IV Fr. Vol. 115 (2004), pp. 351–359.

[12] B. Raniecki and C. Lexcellent: Eur. J. Mech. A/Sol. Vol. 13 (1994), pp. 21–50.

[13] A.C. Souza, E.N. Mamiya and N. Zouain: Eur. J. Mech. A/Sol. Vol. 17 (1998), pp. 789–806.

[14] F. Auricchio and L. Petrini: Int. J. Num. Meth. Engrg. Vol. 61 (2004), pp. 807–836.

[15] F. Auricchio and L. Petrini: Int. J. Num. Meth. Engrg. Vol. 61 (2004), pp. 716–737.

[16] J. Lemaitre and J.L. Chaboche: *Mechanics of Solid Materials*, Cambridge University Press (1990).

[17] F.H. Clarke: *Optimization and nonsmooth analysis. Second edition*, Classics in Applied Mathematics, 5. Society for Industrial and Applied Mathematics, SIAM (1990).

[18] A. Mielke, in: Handbook of Differential Equations, edited by C. Dafermos and E. Feireisl, Elsevier (2005).

[19] F. Auricchio, A. Reali and U. Stefanelli: Int. J. Plast. Vol. 23 (2007), pp. 207–226.

[20] F. Auricchio, A. Mielke and U. Stefanelli: Math. Mod. Meth. Appl. Sci. Vol. 18 (2008), pp. 125–164.

[21] J.C. Simo and T.J.R. Hughes: *Computational Inelasticity*, Springer (1998).

[22] R.L. Taylor: *FEAP: A Finite Element Analysis Program, programmer manual*, http://www.ce.berkeley.edu/∼rlt/feap/, (2001).

[23] F. Migliavacca, L. Petrini, P. Massarotti, S. Schievano, F. Auricchio and G. Dubini: Biomech. Mod. Mechanobiol. Vol. 2 (2004), pp. 205–217.

[24] L. Petrini, F. Migliavacca, P. Massarotti, S. Schievano, G. Dubini and F. Auricchio: J. Biomech. Engrg. (Trans. ASME) Vol. 127 (2005), pp. 716–725.

[25] F. Auricchio, A. Reali and A. Tardugno, to appear in: Recent research development in computational materials science, engineering methods and industrial applications, edited by G. Costanza, Research Signpost (2008).

[26] M. Kohl: *Shape Memory Microactuators*, Springer (2004).