A thermomechanical finite strain shape memory alloy model and its application to bistable actuators

Received: 8 January 2022 / Revised: 15 April 2022 / Accepted: 19 April 2022 / Published online: 6 July 2022
© The Author(s) 2022

Abstract This work presents a thermomechanical finite strain shape memory alloy model that utilizes a projection method to deal with the incompressibility constraint on inelastic strains. Due to its finite strain formulation, it is able to accurately predict the behavior of shape memory alloys with high transformation strains. The key feature of this model is the thermomechanical modeling of the shape memory effect and superelastic behavior by optimizing a global, incremental mixed thermomechanical potential, the variation of which yields the linear momentum balance, the energy balance, the evolution equations of the internal variables as well as boundary conditions of Neumann- and Robin-type. The proposed thermal strain model allows to properly capture transformation induced volume changes, which occur in some shape memory alloys. A finite strain dissipation potential is formulated, which incorporates the disappearance of inelastic strains upon austenite transformation. This important property is consistently transferred to the time-discrete potential using a logarithmic strain formulation. Yield and transformation criteria are derived from the dual dissipation potential. The implementation based on an active set search and the algorithmically consistent linearization are discussed in detail. The model is applied in three-dimensional simulations of a bistable actuator design to explore its capabilities.

1 Introduction

Since the discovery of their unique properties, shape memory alloys are used in many medical and engineering fields in various applications. Their frequent appearance stems from their unique features, like the shape memory effect and superelasticity. Both effects emerge from the characteristic first-order phase transition from the austenite to the martensite state and vice versa. While the occurring crystallographic effects involving the detwinning of the martensite phase, which enables the shape memory effect, are well understood, the thermomechanical modeling of the occurring effects is not straightforward. In recent times, there has been a high effort to improve shape memory alloy models and to obtain fitting models for specific use cases (see, for reviews, Section 4 of Lester et al. [27] or Cisse et al. [8]). They can be roughly categorized into three classes: models based on statistical thermodynamics, models founded in micromechanics and phenomenological models.

The models based on statistical thermodynamics rely on finding the phase equilibrium through a minimization of a three well potential energy (e.g., Seelecke and Müller [43] or Govindjee et al. [16]). Since these models yield results that also consider the microstructure of the materials, they come mostly with a computational cost.
which is too high for large, structural simulations. Additionally, gathering required micromechanical material parameters is sometimes an elaborate task.

On the other hand, models based on micromechanics usually consider the mechanics of shape memory alloy (SMA) single-crystals. Many models are then extended into the regime of polycrystal modeling by usage of homogenization techniques (e.g., see Patoor et al. [35] and Lagoudas et al. [25] or the more recent models by Mirzaeifar et al. [29] and Yu et al. [55]). While these models consider the deep, underlying phenomena of shape memory alloys, this advantage again comes with a high computational cost. This makes it really challenging to use these models in large structural simulations of shape memory alloy actuators.

The third group of models is the class of phenomenological models.

Usually, they come with the advantage of only having macroscopic material parameters, which mostly are obtained through tensile tests at different temperatures. In recent years, due to the plethora of shape memory alloys effects and applications, many new shape memory alloy models were published, which try to include more and more physical phenomena. They can be divided into two subgroups: models which include a geometrically linear theory (see, e.g., Auricchio et al. [3], Auricchio et al. [4] and Sedlak et al. [42]), and models which include a geometrically nonlinear theory.

The two main ways to include a geometric nonlinearity in the shape memory alloy model is to either employ a multiplicative split of the deformation gradient going back to [10,24,26] or to make use of an additive split (see Nemat-Nasser [32]) of the rate of deformation, which are both well known from plasticity. While models based on the additive split are computationally enticing, they only allow for small strains, while still capturing large rotations well (see, e.g., Qidwai and Lagoudas [37], Müller and Bruhns [31] or Zhang and Baxevanis [57]). On the other hand, models employing a multiplicative split are computationally more elaborate, but can represent finite stretches well (see, e.g., Reese and Christ [38], Arghavani et al. [1] or Wang et al. [49]). Additionally, there exist many new models considering geometric nonlinearities, which are limited to superelasticity (see, e.g., Bellini et al. [5], Wang et al. [48] and Rezaee-Hajidehi et al. [39]). Furthermore, some models also consider transformation induced plasticity (see, e.g., Hartl et al. [19], Xu et al. [53] and it’s extension to partial phase transformations by Scalet et al. [41]).

Many of these models capture thermomechanics, martensite reorientation and detwinning as well as superelasticity at different temperatures or allow for different elastic properties of the materials while being numerically efficient and robust. However, when trying to model bistable shape memory actuators (see the actuator design in Arivanandhan [2]), one needs a numerically robust, fully thermomechanically coupled finite strain model that also models thermal expansion and volumetric effects during phase transition (see, e.g., Potapov et al. [36]). To our knowledge, there is no model in the literature fulfilling all of the aforementioned requirements, which led to the model described in this paper. For example, the model of Wang et al. [49] considers a thermomechanically coupled finite strain theory which is capable of modeling the shape memory effect as well as superelasticity, neglecting volumetric effects due to transformation as well as different expansion coefficients of the SMA phases. The model at hand falls into the aforementioned category of phenomenological models. It is embedded into the generalized standard materials framework developed by Halphen and Nguyen [17], which was extended to thermomechanics by Yang et al. [54] and which allows to ensure thermodynamic consistency. The energies as well as the dissipation potential can be seen as an extension of Sedlak et al. [42] to the finite strain case. Because the satisfaction of the incompressibility of inelastic strains for finite strains is not as straightforward as for the small strain case, a projection method developed for plasticity is incorporated into the model (see Hurtado et al. [21] and Sielenkämper et al. [44]). Further, due to the character of the energies used in this model, special numerical treatment is necessary to solve the model equations using a Newton scheme. Since our aim is to model microactuators in which the R-Phase is not present, it is not incorporated into the model. Additionally, tension-compression anisotropy, which is an important effect in many shape memory alloys (for experimental publications see, e.g., Gall et al. [15] or Wang and Zhu [47], and for modeling approaches see, e.g., Zaki et al. [56] or Sedlak et al. [42]), is not included in the model.

In the past, numerous advancements to current microelectromechanical systems (MEMS) based on electrostatics, magnetism and electrothermal principals have been made. For example, Hoffmann et al. [20] proposed a microactuator based on electrothermal activation making use of bimetal effects. This, however, comes with the downside of low actuation frequencies and a high power consumption. Han et al. [18] proposed an electrostatic actuator-based micro-switch for photonics. Devices based on electrostatics usually can be actuated with high frequencies and are adaptable to many applications. Devices using optomechanics were developed by, e.g., Eichenfield et al. [11], which come with a rather small tuning range, but allow for a very high operation speed [9]. Despite their unique advantages, current MEMS devices are challenging to use in downsized applications, where a high work output combined with a high power efficiency and bistability is crucial. The
microactuator modeled in this paper is based on a concept published by Winzek et al. [50], which utilizes a high temperature shape memory alloy with a large thermal hysteresis. This concept may overcome the aforementioned weaknesses of other actuation principles, as shape memory alloys usually have a large work output density and favorable downscaling capabilities (see, e.g., Kohl [22]). Additionally, downscaling this design is expected to drastically increase the possible actuation frequency in comparison to other, electrothermally activated actuators due to the decreasing masses and increasing thermal gradients. One further key aspect is, that the proposed actuator design requires no power in the stable states.

The paper is structured as follows: First, the energies as well as the dissipation potential is derived. Then, in Sect. 3, numerical strategies necessary to solve the model equations are discussed. The numerical results are subsequently shown in Sect. 4 before a summary and outlook concludes the paper in Sect. 5.

Notation
Throughout this paper, a direct tensor notation is preferred. Scalars and scalar valued functions are typeset using light-face italic characters, e.g., $a$ or $A$. First and second-order tensors and tensor-valued functions are represented by bold-face italic letters, e.g., $\mathbf{a}$ or $\mathbf{A}$. Further, blackboard bold-faced letters are used to denote fourth-order tensors, e.g., $\mathbb{C}$ or $\mathbb{C}$.

Additionally, the transpose of a second-order tensor is designated by $\mathbf{A}^T$, while the major transpose of a fourth-order tensor is given by $\mathbb{C}^T$. The symmetric and deviatoric part of a second-order tensor $\mathbf{A}$ are denoted by $\text{sym}(\mathbf{A}) = \frac{1}{2}(\mathbf{A} + \mathbf{A}^T)$ and $\mathbf{A}' = \mathbf{A} - \frac{1}{3}\text{tr}(\mathbf{A})\mathbf{I}$, respectively. Here, $\mathbf{I}$ is the second-order identity tensor and $\text{tr}(\mathbf{A})$ denotes the trace of $\mathbf{A}$. A double contraction of two tensors $\mathbf{A}$ and $\mathbf{B}$ is denoted by $\mathbf{A}:\mathbf{B}$, while the dyadic product is denoted by $\mathbf{a} \otimes \mathbf{b}$. A determinant of a tensor $\mathbf{A}$ is either designated by $\det(\mathbf{A})$ or by $\mathbf{A}$’s third invariant $\text{III}_A$.

2 Modeling of shape memory alloys

2.1 Kinematics

The deformation gradient $\mathbf{F}$ maps a line element from the reference configuration of a body with volume $V_0$ into the current configuration of a body with volume $V$ and is defined as

$$\mathbf{F} = \text{Grad}(\mathbf{x}(\mathbf{X}, t)), \quad (1)$$

where $\text{Grad}(\bullet)$ refers to the gradient with respect to the reference configuration while $\mathbf{X}$ and $\mathbf{x}$ are the position vectors of a material point in the reference and current configuration, respectively. We consider a multiplicative split of the deformation gradient in the form

$$\mathbf{F} = \mathbf{F}^e \mathbf{F}^i \mathbf{F}^\theta \quad (2)$$

(see Wang et al. [49]), where $\mathbf{F}^e$ is the elastic, $\mathbf{F}^i$ the isochoric part of the deformation due to transformation\(^1\) and $\mathbf{F}^\theta$ the part which describes the volume change due to thermal expansion and transformation\(^2\) of the deformation gradient. This is fairly similar to the multiplicative split in plasticity going back to the works of Eckart [10], Kröner [24] and Lee [26]. Further, we define the determinant $J^\theta = J^\theta(\theta, \xi) = \det(\mathbf{F}^\theta)$ to be a function of the absolute temperature $\theta$ and the martensite volume fraction $\xi \in [0, 1]$. Since the thermal deformation is assumed to be volumetric, we can express $\mathbf{F}^\theta$ in terms of $J^\theta$ as

$$\mathbf{F}^\theta = (J^\theta)^{\frac{1}{3}} \mathbf{I}. \quad (3)$$

As is commonly done and will be useful later, we define the elastic and inelastic left Cauchy-Green tensors as

$$\mathbf{b}^e = \mathbf{F}^e \mathbf{F}^{eT}, \quad \mathbf{b}^i = \mathbf{F}^i \mathbf{F}^{iT}. \quad (4)$$

Likewise, we define the inelastic right Cauchy-Green tensor and the inelastic Green-Lagrange strain

$$\mathbf{C}^i = \mathbf{F}^{iT} \mathbf{F}^i, \quad \mathbf{E}^i = \frac{1}{2}(\mathbf{C}^i - \mathbf{I}). \quad (5)$$

\(^1\) For brevity, $\mathbf{F}^i$ is called the inelastic part of the deformation gradient throughout this paper.

\(^2\) For brevity, $\mathbf{F}^\theta$ is called the thermal part of the deformation gradient throughout this paper.
Motivated by the observation that the shape-memory effect is caused by an almost¹ volume preserving transformation of the crystal lattice, we assume \( C^i \) to be volume preserving. Therefore, because \( \det(F^i) = 1 \) has to hold, the determinant of the deformation gradient is given by

\[
J = \det(F) = J^0 J^e,
\]

with \( J^e = \det(F^e) \). Additionally, we define the velocity gradient \( l \) and its symmetric part \( d \) as

\[
l = \dot{F} F^{-1}, \quad d = \text{sym}(l).
\]

Finally, we define the inelastic 'velocity gradient' \( L^i \) and its symmetric part \( D^i \) in analogy to \( l \) and \( d \) by

\[
L^i = \dot{F}^i F^{-1}, \quad D^i = \text{sym}(L^i).
\]

2.2 (Im-)Balance equations

2.2.1 Momentum balances

The quasistatic linear momentum balance for a body with volume \( V \) in the current configuration is given by

\[
\text{div}(\sigma) + \rho \mathbf{b} = 0 \quad \text{in} \quad V,
\]

where \( \mathbf{b} \) is the body force and \( \rho \) is the mass density. Further, Cauchy’s lemma \( t = \sigma n \) as well as the angular momentum balance \( \sigma = \sigma^T \) are assumed to hold.

2.2.2 Energy balance

Next, the balance of the energies is given by

\[
\dot{u} = \tau : d - J \text{div}(q) + w,
\]

where \( u \) is the internal energy density per unit reference volume, \( \tau \) denotes the Kirchhoff stress tensor, \( q \) is the heat flux vector in the current configuration and \( w \) represents the energy source term. Further, the divergence of the heat flux vector \( Q = J F^{-1} q \) with respect to the reference configuration is then given by \( J \text{div}(q) = \text{Div}(Q) \).

2.2.3 Clausius–Planck inequality

To be thermodynamically consistent, the Clausius–Planck inequality

\[
\mathcal{D} = \tau : d + \theta \dot{s} - \dot{u} \geq 0,
\]

where \( \mathcal{D} \) is the mechanical dissipation density per unit reference volume and \( s \) is the entropy density per unit reference volume, has to be fulfilled at any time. Further, the Helmholtz free energy, which is defined by \( \psi = u - \theta s \), is introduced. We assume \( \psi \) to be a function of the deformation gradient \( F \), the absolute temperature \( \theta \) and the internal variables \( \xi \) and \( C^i \), i.e., \( \psi = \psi(F, C^i, \xi, \theta) \). Furthermore, we make the common assumption that

\[
\mathcal{D} \bigg|_{C^i, \xi} = 0,
\]

i.e., there is no energy dissipation when the internal variables are virtually fixed. As Eq. (12) must hold for arbitrary processes, one can show the following standard results:

\[
\tau = \frac{\partial \psi}{\partial F} F^T, \quad s = -\frac{\partial \psi}{\partial \theta}.
\]

Now, it is easy to show that

\[
\mathcal{D} = \mathbf{\Sigma}^i : D^i + q \dot{\xi} \geq 0
\]

with shorthand notations for the effective Mandel stress with respect to the intermediate configuration \( \mathbf{\Sigma}^i = -2F^i (\partial \psi/\partial C^i) F^{iT} \) and the thermodynamic force \( q = -\partial \psi/\partial \xi \) associated with the martensite volume fraction \( \xi \).

³ Actually, for some shape memory alloys, there exists a difference in density between the martensite and austenite phase. For NiTiHf, see the article by Potapov et al. [36] and for modeling approaches for this effect in NiTi the work of Qidwai and Lagoudas [37]. We also model this phenomenon in Sect. 2.3.4.
2.3 Helmholtz free energy

The Helmholtz free energy density $\psi$ for this model is assumed to be a sum of elastic, chemical and hardening-like contributions in the form

$$\psi = \psi_e(b^e, \xi) + \psi_c(\xi, \theta) + \psi_h(C^i, \xi).$$  \hspace{1cm} (15)

2.3.1 Elastic energy

The elastic energy $\psi_e$ is assumed to be isotropic and to follow a modified Neo-Hookean formulation:

$$\psi_e(b^e, \xi) = \frac{\lambda(\xi)}{4} (J^e - 1 - 2 \ln J^e) + \frac{\mu(\xi)}{2} (\text{tr}(b^e) - 3 - 2 \ln J^e).$$  \hspace{1cm} (16)

Here, $\lambda(\xi)$ and $\mu(\xi)$ are the Lamé parameters in dependence of the martensite volume fraction $\xi$. Further, we use a Reuss-like mixture rule to estimate the elastic constants, i.e.,

$$\mu(\xi) = \left( \frac{\lambda_M}{\mu_M} + \frac{1 - \xi}{\mu_A} \right)^{-1}, \quad \lambda(\xi) = \left( \frac{\xi}{\lambda_M} + \frac{1 - \xi}{\lambda_A} \right)^{-1}.$$  \hspace{1cm} (17)

Here, and subsequently, the indices $\bullet_A$ and $\bullet_M$ refer to the austenite and martensite phase, respectively. Manipulating Eq. (13), one can show that the ordinary form

$$\tau = 2b^e \frac{\partial \psi_e}{\partial b^e}$$  \hspace{1cm} (18)

holds for elastic isotropy. Using Eqs. (14) and (18), one can show that

$$\Sigma^i = -2F^i \frac{\partial \psi}{\partial C^i} F^{iT} = \Sigma^e - 2F^i \frac{\partial \psi_h}{\partial C^i} F^{iT},$$  \hspace{1cm} (19)

where $\Sigma^e = C^e F^{-1} \tau F^{-T}$ is the Mandel stress with respect to the intermediate, elastically unloaded configuration, which is symmetric due to the assumption of elastic isotropy.

2.3.2 Chemical energy

For the chemical energy, we assume a standard relationship (see, e.g., Lexcellent et al. [28] or Panico and Brinson [34]):

$$\psi_c = (u_0^A - \theta s_0^A) - \xi (\Delta u^{AM} - \theta \Delta s^{AM}) + c \left( \theta - \theta_0 - \theta \ln \frac{\theta}{\theta_0} \right)$$

$$= u_0^A - \theta s_0^A + \xi (\theta - \theta_0) \Delta s^{AM} + c \left( \theta - \theta_0 - \theta \ln \frac{\theta}{\theta_0} \right),$$  \hspace{1cm} (20)

where $c$ is the specific heat capacity and $\Delta s^{AM}$ is the difference in specific entropy of the austenite and martensite phase: $\Delta s^{AM} = s_0^A - s_0^M$. Here, $c$ is assumed constant (compare [42]). In Eq. (20), we made use of the definition of the equilibrium temperature of austenite and martensite, which is $\theta_0 = \Delta u^{AM}/\Delta s^{AM}$. Further, we assume for the equilibrium temperature $A_s > \theta_0 > M_s$, where $A_s$ is the temperature where the reverse transformation starts and $M_s$ is the starting temperature of the forward transformation.
2.3.3 Hardening energy

Since the inelastic strains vanish as $\xi \to 0$, the inelastic strains $E^i$ are assumed to satisfy the relation $E^i = \xi E^t$ (see Otsuka and Ren [33]), where $E^t$ is a measure for the effective transformation strain. Now, the hardening-like energy adapted from Sedlak et al. [42] for finite strains is assumed to be given by

$$\psi_h = k E^\text{int} \xi^2 \frac{\langle E^t \rangle^2}{1 - \langle E^t \rangle}, \quad \langle E^t \rangle = \sqrt{\frac{2}{3} \left\| E^t \right\|^2 k}. \quad (21)$$

Here, $k$ is the maximum transformation strain, $E^\text{int}$ is a hardening related parameter and $\langle E^t \rangle$ is a modified von Mises equivalent strain. Obviously, $\psi_h \to \infty$ for $\langle E^t \rangle \to 1$. This is desired, since it captures the martensite becoming fully detwinned, which is assumed to cost additional energy.

2.3.4 Thermal strains

Modeling thermal strains of multi-phase materials requires attention and special treatment. In this work, the determinant of the thermal part of the deformation gradient is connected to the coefficients of thermal expansion (CTEs) by

$$J^\theta = \left(1 + \epsilon^\theta(\xi, \theta)\right)^3, \quad \epsilon^\theta = \xi \alpha_M (\theta - \theta_{\text{ref}M}) + (1 - \xi) \alpha_A (\theta - \theta_{\text{ref}A}). \quad (22)$$

Here, $\epsilon^\theta(\xi, \theta)$ is the thermal strain. Additionally, $\alpha_M$ and $\alpha_A$ are the CTEs of the martensite and austenite phase, respectively. Further, $\theta_{\text{ref}M}$ and $\theta_{\text{ref}A}$ are the reference temperatures for austenite and martensite. Here, we want to note that it is necessary to distinguish the two in order to properly represent the transformation-induced volume change which is present in some shape memory alloys (see below).

**Example**

We assume a one-dimensional SMA rod with only one reference temperature. Further, we simplify the SMA model by assuming that martensite and austenite transformation occur at the distinct temperatures $\theta_M$ and $\theta_A$, and not at temperature ranges (see Fig. 1). The thermal strains are then given by $\epsilon^\theta = \alpha_{AM}(\theta - \theta_{\text{ref}})$, with $\alpha_{AM} = \xi \alpha_M + (1 - \xi) \alpha_A$. Furthermore, we assume that we know the shape of our rod at the thermal annealing temperature $\theta_{\text{ref}}$ (see Fig. 1).

During forward transformation, the jump of $\epsilon^\theta$ at $\theta_M$ can be analytically calculated and is given by

$$\Delta \epsilon^\theta = (\theta_{\text{ref}} - \theta_M)(\alpha_A - \alpha_M). \quad (23)$$

Here, we can not only see that this jump is dependent on the reference temperature, but that it can be positive or negative, depending on the reference temperature $\theta_{\text{ref}}$ being larger or smaller than the forward transformation temperature $\theta_M$.

Using two reference temperatures $\theta_{\text{ref}M}$ and $\theta_{\text{ref}A}$, we can not only circumvent this problem, but also model the magnitude of this jump in compliance with, e.g., the experimental results of Potapov et al. [36], where they calculated the jump $\Delta V$ in volume from the lattice parameters for Ni$_{49.8}$Ti$_{35.2}$Hf$_{15}$ to be 0.47%. However, we want to emphasize that this effect does not play a significant role in many other shape memory alloys.

![Fig. 1](image-url) Thermal strains when only considering one reference temperature $\theta_{\text{ref}}$
2.4 Dissipation potential

The dissipative behavior of the shape memory alloy is assumed to be governed by the dissipation potential

$$\phi(\dot{\xi}, D^i, \xi) = \begin{cases} 
\phi_M, & \text{if } \dot{\xi} \geq 0 \land \text{tr}(D^i) = 0, \\
\phi_A, & \text{if } \dot{\xi} < 0 \land \text{tr}(D^i) = 0, \\
\infty, & \text{else}.
\end{cases}$$

(24)

Here, depending on the direction of transformation, either $\phi_M$ or $\phi_A$ are used as the active dissipation potential. As is commonly done for plasticity, the dissipation is assumed to be infinite if tr$(D^i) \neq 0$, which guarantees volume-preserving inelastic deformations (compare Rockafellar [40]). Additionally, we define the dissipation potential for an increasing martensite volume fraction as

$$\phi_M = \frac{\dot{\xi}}{\xi} \Delta s^{AM}((\theta_0 - M_s) + \xi(M_s - M_f)) + \sqrt{\frac{2}{3}} \sigma^{reo} \| D^i \|,$$

(25)

where $M_f$ is the finish temperature of forward transformation and $\sigma^{reo}$ is the reorientation stress. Likewise, the dissipation potential when going to a higher austenite volume fraction is defined as

$$\phi_A = \frac{\dot{\xi}}{\xi} \Delta s^{AM}((\theta_0 - A_f) + \xi(A_f - A_s)) + \sqrt{\frac{2}{3}} \sigma^{reo} \left( \left\| \frac{\dot{\xi}}{\xi} \right\| + \left\| \frac{D^i - \frac{\dot{\xi}}{\xi} \epsilon^i}{\xi} \right\| \right).$$

(26)

Here, $A_f$ is the finish temperature of reverse transformation and $\epsilon^i = \frac{1}{2} \ln b^i$. The significance of using the logarithmic inelastic strain $\epsilon^i$ as a strain measure for the dissipation is explained below. The dissipation potential represents a geometrically nonlinear generalization of the small strain potential proposed in Sedlak et al. [42], which is based on the works of Bernardini and Pence [6], Panico and Brinson [34] and Moumni et al. [30].

2.5 Transformation/yield criteria and inelastic evolution equations

It is assumed that the evolution of $C^i$ and $\xi$ follows from the minimization problem

$$\inf_{D^i, \dot{\xi}} \Psi + \phi,$$

(27)

which is equivalent to the Legendre–Fenchel transformed problem (compare Eqs. (14), (19))

$$\sup_{D, \dot{\xi}} (\Sigma^e - \Sigma^b) : D^i + q \dot{\xi} - \phi(\dot{\xi}, D^i, \xi) =: \phi^*(\Sigma^i, q, \xi),$$

(28)

where $\phi^*$ is the dual dissipation potential of $\phi$:

$$\phi^*(\Sigma^i, q, \xi) = \sup_{D^i, \dot{\xi}} \Sigma^i : D^i + q \dot{\xi} - \begin{cases} 
\phi_M, & \text{if } \dot{\xi} \geq 0, \\
\phi_A, & \text{if } \dot{\xi} < 0.
\end{cases}$$

(29)

Due to the minimization in Eq. (27), there is a tendency that for $\dot{\xi} < 0$ we get $D^i - (\dot{\xi}/\xi) \epsilon^i = 0$ as a result of the last term in Eq. (26). In that case $b^i \rightarrow I$ as $\xi \rightarrow 0$ in time (for a proof, see Appendix A), i.e., for vanishing martensite content, the inelastic strain vanishes, which is a physical necessity. With the reparametrization $D^i = D^i - (\dot{\xi}/\xi) \epsilon^i$ (compare Eq. (26)) we find

$$\phi^*(\Sigma^i, q, \xi) = \begin{cases} 
\sup_{D^i, \dot{\xi} \geq 0} \Sigma^i : D^i - \sqrt{\frac{2}{3}} \sigma^{reo} \| D^i \| + (q - \sigma^{reo} \| \frac{\dot{\xi}}{\xi} \|), & \text{if } \dot{\xi} \geq 0, \\
\sup_{D^i, \dot{\xi} \leq 0} \Sigma^i : D^i - \sqrt{\frac{2}{3}} \sigma^{reo} \| D^i \| + \left( \frac{1}{\xi} \Sigma^i : \epsilon^i + q - \sigma^{reo} \| \frac{\dot{\xi}}{\xi} \| \right) \dot{\xi}, & \text{if } \dot{\xi} < 0.
\end{cases}$$

(30)
where $Q_M$ and $Q_A$ are defined in Eqs. (25) and (26) and capture the transformation hysteresis due to temperature change. Further reparametrizing $D^{ij} = \lambda N$ with $\lambda \geq 0$ and the obvious solution

$$N = \frac{\Sigma^i}{\|\Sigma^i\|},$$

(31)

it follows that

$$\phi^*(\Sigma^i, q, \xi) = \sup_{\lambda \geq 0, \xi \geq 0} \lambda f(\Sigma^i) + \hat{\xi} g_M(q)$$

$$= \begin{cases} 0, & \text{if } f \leq 0 \land g_M \leq 0 \land g_A \geq 0, \\ \infty, & \text{else}. \end{cases}$$

(32)

Thus, we find the following transformation and yield criteria in the classical Karush-Kuhn-Tucker form as well as evolution equations:

$$f(\Sigma^i) = \|\Sigma^i\| - \sqrt{3} \sigma^{\text{reo}} \leq 0; \quad \lambda f = 0; \quad \lambda \geq 0; \quad D^i = \lambda \frac{\partial f}{\partial \Sigma^i};$$

$$g_M(q) = q - Q_M(\xi) \leq 0; \quad (\hat{\xi}) g_M = 0; \quad \hat{\xi} > 0 \text{ possible if } g_M = 0;$$

$$g_A(q, \Sigma^i) = \frac{1}{\hat{\xi}} \Sigma^i : \epsilon^i + q - Q_A(\xi) + \sqrt{3} \hat{\xi} \|\epsilon^i\| \geq 0; \quad (\hat{\xi}) g_A = 0; \quad \hat{\xi} < 0 \text{ possible if } g_A = 0;$$

(33)

with the Macaulay bracket $(\bullet) = (\bullet + |\bullet|)/2$ and its modified form $(\bullet)_- = (\bullet - |\bullet|)/2$. It is then straightforward to prove thermodynamic consistency (see Eq. (14)). The '0' branch in Eq. (32) implies that (see Eq. (28))

$$\mathcal{D} = \Sigma^i : D^i + q \hat{\xi} = \phi,$$

(34)

Thus, we find with Eqs. (4) and (10) the following alternative form of the energy balance:

$$\theta \hat{s} = \phi - \text{Div}(\mathbf{Q}) + w.$$

(35)

2.6 Rate potential

To simplify the discussion, we start by the isothermal case, i.e., in a first step, the temperature $\theta$ is considered to be a given parameter. The rate potential and its time discretized form read

$$\pi = \psi + \phi \simeq \frac{1}{\Delta t} \left( \psi(F, C^i, \xi, \theta) - \psi_n + \phi_\Delta \right).$$

(36)

Here, $\psi_n$ refers to the Helmholtz free energy at the previous time, $\Delta t$ is the time step from $t_n$ to $t_{n+1}$ and $\phi_\Delta$ is the time discretized version of the dissipation potential multiplied by $\Delta t$, defined by

$$\phi_\Delta = \begin{cases} \sigma^{\text{reo}} \Delta \alpha + Q(C^i, \xi_n + \frac{1}{2}, \text{sg} (\Delta \xi)) \Delta \xi, & \text{if } \mathbf{III}^i = 1, \\ \infty, & \text{else}, \end{cases}$$

(37)

with $\text{sg}(\bullet)$ referring to the sign function and $\Delta \xi = \xi - \xi_n$. Additionally, $\mathbf{III}^i$ is the third invariant of $C^i$, i.e.,

$$\mathbf{III}^i = \text{det}(C^i).$$

(38)

The constraint $\mathbf{III}^i = 1$ in Eq. (37) is consistent with the requirement $\text{tr}(D^i) = 0$ (see Eq. (24)). Additionally, we compute the effective inelastic strain increment $\Delta \alpha$ as

$$\Delta \alpha = \sqrt{\frac{3}{2}} \left\| \frac{1}{2} U_n^{i-1} \Delta t C^i U_n^{i-1} \right\| =: \sqrt{\frac{3}{2}} \left\| \frac{1}{2} \Delta \epsilon C^i \right\| U_n^{i-1},$$

(39)

\footnote{For brevity and simplicity, we dropped the index $n + 1$ whenever we deemed it to be not helpful for the presentation.}
where the inelastic right stretch is \( U^i = \sqrt{C^i}, \Delta t C^i = C^i - C^i_0 \) and the shorthand notation \( \|\cdot\|_{U_{n-1}} = \|U_{n-1}^i \cdot U_{n-1}^{i-1}\| \) is used. Further, we define \( C^i_t \) as

\[
C^i_t = (C^i_n)^{1 + \frac{\Delta t}{\xi_n}}.
\]

Now, for \( \xi \) going to zero in any given step, i.e., \( \xi = \xi_n + \Delta \xi = 0 \Leftrightarrow \Delta \xi = -\xi_n \rightarrow C^i_t = (C^i_n)^0 = I, C^i_t \) goes back to unity again. Thus, the time discrete potential (37) based on definition (40) of \( C^i_t \) is the key to ensure that the inelastic strain consistently disappears during the transformation from martensite to austenite. Further, the function \( Q(C^i_n, \xi_n + \frac{1}{2}, \text{sgn}(\Delta \xi)) \) is given by

\[
Q(C^i_n, \xi_n + \frac{1}{2}, \text{sgn}(\Delta \xi)) = \left\{ \begin{array}{ll}
\Delta \sigma^{AM}((\theta_0 - M_s) + \xi_n(M_s - M_t)), & \text{if } \Delta \xi \geq 0, \\
\Delta \sigma^{AM}((\theta_0 - A_t) + \xi_n(A_t - A_s)) - \sqrt{2} \frac{\sigma_{\text{reo}}}{\xi_n} \| \varepsilon^i_n \|, & \text{if } \Delta \xi < 0,
\end{array} \right.
\]

where \( \xi_n + \frac{1}{2} = (\xi + \xi_n)/2 \) is the midpoint evaluation of \( \xi \), which is employed to obtain a reasonable transformation when the material is not stressed (see Frost et al. [14] for a similar concept). Moreover, we used \( C^i_t \) and \( \xi_n \) for the reverse transformation in Eq. (41) to circumvent the eigenvalue problem as well as its linearization in every local Newton iteration. In this way, it suffices to solve the eigenvalue problem once per time step.

Using some further shorthand notations for constant terms in \( Q \), we obtain

\[
\phi_\Delta = \sigma^{\text{reo}} \Delta \alpha
\]

\[
= \frac{Q^{M_0}}{Q^{M_0}} - \frac{\Delta \sigma^{AM}((\theta_0 - M_s) + \xi_n(M_s - M_t))}{\Delta \sigma^{AM}((\theta_0 - A_t) + \xi_n(A_t - A_s)) - \sqrt{2} \frac{\sigma_{\text{reo}}}{\xi_n} \| \varepsilon^i_n \|} \Delta \xi + \frac{1}{2} \Delta \sigma^{AM}(A_t - A_s) \Delta \xi^2,
\]

with the time-continuous theory is trivially proven except for \( \Delta \alpha \). Therefore, we have a look at the approximation of \( a^{1+x} \) at \( x \ll 1 \):

\[
a^{1+x} \approx a + x \ln a
\]

as \( \ln a^x \approx a^x - 1 \). Hence, we can approximate \( C^i_t \) as

\[
C^i_t \approx C^i_n + C^i_n \ln(C^i_n) \left( \frac{\Delta \xi}{\xi_n} \right).
\]

We can use this result for computation of the time-discrete derivative of \( \alpha \):

\[
\frac{\Delta \alpha}{\Delta t} \approx \sqrt{\frac{1}{2}} \| U_{n-1}^i \| \frac{1}{\Delta t} C^i U_{n-1}^i \Delta \xi U_{n-1}^{i-1} - \frac{1}{2} \ln(C^i_n) \left( \frac{\Delta \xi}{\Delta t \xi_n} \right) \|.
\]
Therefore, with the time step width going to zero, we obtain
\[
\Delta t \to 0 \Rightarrow \sqrt{\frac{2}{3}} \left\| D^i - \frac{\dot{\xi}}{2\xi} \ln (b^i) \right\| = \dot{\alpha},
\]
which is consistent with the time-continuous Eq. (26). Furthermore, since \( \mathbf{III}^i = \det(C^i) = 1 \), we get for the time derivative of \( \mathbf{III}^i \):
\[
\dot{\mathbf{III}^i} = \dot{\mathbf{III}^i} C^i : \dot{C}^i = 0.
\]
Hence, since \( \phi_\Delta \) is time-continuous, we can use the discretized potential to solve for \( C^i \) and \( \xi \):
\[
\left( C^i, \xi \right) = \arg \inf_{C^i,\xi} \pi_\Delta(F, C^i, \xi, \theta).
\]

3 Numerical strategies

The potential \( \pi_\Delta \), as we formulated it in Eq. (48) carries two major numerical difficulties. First, \( C^i \) is constrained to be volume preserving, i.e., \( \det(C^i) = 1 \), which is very important to be exactly satisfied to comply with physics and not accumulate errors. Second, \( \pi_\Delta \) is not differentiable at \( \Delta t C^i = 0 \). The strategies employed to overcome these and other numerical difficulties as well as general numerical approaches are presented in this section.

3.1 Inelastic volume preservation

To deal with the constraint in \( C^i \), we employ a strategy using a projection of \( C^i \) into the space of unimodular tensors, which was introduced by Hurtado et al. [21] for crystal plasticity. Our approach here closely follows the approach presented in Sielenkämper et al. [44].

First, we express \( C^i \) in terms of the unconstrained inelastic auxiliary right Cauchy-Green tensor \( \hat{C}^i \):
\[
C^i = \mathbf{III}^{i-\frac{1}{2}} \hat{C}^i, \quad \mathbf{III}^i = \det(\hat{C}^i).
\]
Thus, \( \det(C^i) = 1 \) is automatically satisfied. This idea is borrowed from various formulations in hyperelasticity, where similar approaches are used to decouple volumetric and deviatoric deformations (see, e.g., Flory and Volkenstein [13] or Simo et al. [45]). Now, we replace \( C^i \) by \( C^i(\hat{C}^i) \) in the minimization problem in Eq. (48):
\[
\left( C^i, \xi \right) = \arg \inf_{\hat{C}^i,\xi} \pi_\Delta(F, C^i(\hat{C}^i), \xi, \theta).
\]
However, while this removes the constraint from the minimization problem, \( \pi_\Delta \) is now invariant with respect to changes of \( \mathbf{III}^i \), and is therefore not uniquely solvable. For this reason, we add the regularization energy
\[
\psi_r = \frac{A}{2} (\hat{\mathbf{III}^i} - B)^2
\]
which has no effect on the solution. This is due to the fact that \( \hat{\mathbf{III}^i} = B \) is exactly satisfied after converging to a solution, since \( \psi_r \) is the only term in \( \pi_\Delta \) that is dependent on \( \mathbf{III}^i \). Therefore, \( \psi_r \) is zero at any solution of the minimization problem. Further, the constants \( A \) and \( B \) can be chosen arbitrarily. They do not have to be chosen particularly large. In this paper, we chose \( B = 1 \), but any other positive value could be chosen and would lead to the exact same results. In this special case, \( C^i = \hat{C}^i \) is exactly satisfied once the solution algorithm is converged. However, we want to note that this equality does not hold in the not yet converged state.

The following tensor, which is used in the residuals and stiffness terms later, is also known from hyperelastic models:
\[
P_1(\hat{C}^i) = \mathbf{III}^{i-\frac{1}{2}} \left( \mathbf{I}^s - \frac{1}{3} \hat{C}^i \otimes \hat{C}^{i-1} \right) = \frac{\partial C^i}{\partial \hat{C}^i},
\]
where \( \mathbf{I}^s \) is the fourth-order identity on symmetric second-order tensors.
3.2 Differentiability at \( \Delta_t C^i = 0 \)

The ansatz \( C^i = \begin{pmatrix} \mathbf{I} \, \frac{1}{2} \end{pmatrix} C^i \) renders the solution \( C^i \) a priori volume preserving, but it does not solve the lack of differentiability of \( \phi_{\Delta} \) for \( C^i = C^i_\iota \). However, this is achieved by the following reparametrization:

\[
\hat{C}^i(\Delta \gamma, \tilde{N}^s) = C^i_\iota + 2 \Delta \gamma \frac{\tilde{N}^s}{\| \tilde{N}^s \|_{U_{n}^{i-1}}}, \quad \Delta \gamma \geq 0,
\]

where \( \tilde{N}^s \) is an unconstrained, symmetric 2nd-order tensor. The final unconstrained minimization problem now reads

\[
(\Delta \gamma, \tilde{N}^s, \xi) = \arg \min_{\Delta \gamma > 0, \tilde{N}^s, \xi} \pi_{\Delta}(F, C^i, (\hat{C}^i(\Delta \gamma, \tilde{N}^s)), \xi, \theta).
\]

It is easy to show that \( \pi_{\Delta} \) is invariant with respect to \( \| \tilde{N}^s \|_{U_{n}^{i-1}} \). In order to render the solution unique, we further modify \( \psi_r \):

\[
\psi_r = \frac{A}{2} \left( (\mathbf{I}^i - 1)^2 + \left( \| \tilde{N}^s \|_{U_{n}^{i-1}} - 1 \right)^2 \right) + \frac{\epsilon}{2} (\tilde{N}^s - \tilde{N}^s_{ir})^2,
\]

where the choice of \( A \), with the same argument as in Sect. 3.1, still has no effect on the solution. Additionally, the last term with a very small and constant \( \epsilon \) is explained in Sect. 3.5. Now, the minimization problem from Eq. (54) is computed by solving the stationary conditions:

\[
\frac{\partial \pi_{\Delta}}{\partial \Delta \gamma} = 0, \quad \frac{\partial \pi_{\Delta}}{\partial N^s_{ir}} = 0, \quad \frac{\partial \pi_{\Delta}}{\partial \xi} = 0,
\]

where in general only a subset of these three equations is involved, depending on the 'active' variables. The active set of variables is determined by the activation or yield criteria presented in the sequel (see Sects. 3.3 and 3.10).

3.3 Algorithmic yield criterion

In a given time step, \( C^i \) will only change (i.e., \( C^i \neq C^i_\iota \Leftrightarrow \Delta \gamma > 0 \)), if this decreases (minimizes) the potential. In order to decide whether the couple \( (\Delta \gamma, \tilde{N}^s) \) is activated for a given state \( (F, C^i_\iota, \xi, \theta) \), we evaluate the algorithmic yield condition

\[
f := -\frac{\partial}{\partial \Delta \gamma} \inf_{N^s} \pi_{\Delta}(F, C^i(\hat{C}^i(\Delta \gamma, \tilde{N}^s)), \xi, \theta) \bigg|_{\Delta \gamma \to 0}.
\]

If \( f > 0 \), \( \Delta \gamma \) is activated, i.e., \( (\Delta \gamma, \tilde{N}^s) \) are put into the active set. Otherwise, \( \Delta \gamma = 0 \) minimizes \( \pi_{\Delta} \), which corresponds to the case where \( f \leq 0 \). If \( (\Delta \gamma, \tilde{N}^s) \) are part of the active set, the related equations and concerning residuals are: \( \partial \pi_{\Delta} / \partial \Delta \gamma \bigg|_{\Delta \gamma = 0} = 0 \) and \( \partial \pi_{\Delta} / \partial \tilde{N}^s = 0 \). For a given \( F, \xi \) and \( \theta \), we obtain the explicit form of \( f \) by variation of \( \pi_{\Delta} : \Delta \gamma = 0 \) if \( \pi_{\Delta}(F, C^i_\iota, \xi, \theta) \) is minimized by \( C^i = C^i_\iota \). In other words, \( \Delta \gamma = 0 \) is minimizer of \( \pi_{\Delta} \) if

\[
\delta \pi_{\Delta} \geq 0 \quad \forall \delta C^i = \frac{\partial C^i}{\partial \tilde{C}^i} : (2 \Delta \gamma \tilde{N}^s),
\]

with the shorthand notation

\[
N^s = \tilde{N}^s / \| \tilde{N}^s \|_{U_{n}^{i-1}}
\]

for an arbitrary \( \tilde{N}^s \neq 0 \) and \( \delta \gamma \geq 0 \). Further, taking a closer look at the variation of \( \pi_{\Delta} \), we obtain (compare Appendix B)

\[
\delta \pi_{\Delta} = \left. \frac{\partial \psi_r}{\partial C^i} \right|_{C^i = C^i_\iota} : F_{\iota}(C^i_\iota) : (2 \Delta \gamma N^s) + \sqrt{2} \sigma^{re0} \delta \gamma \| U_{n}^{i-1} N^s U_{n}^{i-1} \|_{1}.
\]
Using the definition of the unimodular projector in Eq. (52), we obtain
\[ \delta \pi_\Delta \overset{(52)}{=} \left( \frac{2}{3} \frac{\partial \psi}{\partial C_i^1} \bigg| C^i_1 = C^i_1 \right) \left( C^i_1 : 2 \frac{\partial \psi}{\partial C_i^1} \bigg| C^i_1 = C^i_1 \right) C_i^{i-1} : N^s + \sqrt{3} \sigma^\text{reo} \delta \gamma. \]  
(61)

With the definitions of \( N^s \) in Eq. (59) and \( \text{DEV} C_i \) in Eq. (61) as well as making use of the fact that \( U_n^{i-1} \) and \( C_i^{i-1} \) are coaxial, we can simplify Eq. (60) to
\[ \delta \pi_\Delta = \left( U_n^i \text{DEV} C_i \left( 2 \frac{\partial \psi}{\partial C_i^1} \bigg| C^i_1 = C^i_1 \right) U_n^i \right) \left( U_n^{i-1} \tilde{N}^s U_n^{i-1} \right)^{-1} \left( U_n^{i-1} \tilde{N}^s U_n^{i-1} \right) \delta \gamma. \]  
(62)

Now, since inequality (58) must hold for arbitrary symmetric \( \tilde{N}^s \), we get
\[ \delta \pi_\Delta \geq - \left\| \text{DEV} C_i \left( 2 \frac{\partial \psi}{\partial C_i^1} \bigg| C^i_1 = C^i_1 \right) \right\| U_n^i \geq \sqrt{3} \sigma^\text{reo} \delta \gamma \geq 0 \quad \forall \delta \gamma \geq 0. \]  
(63)

Thus, the algorithmic yield criterion reads
\[ f = \left\| \text{DEV} C_i \left( 2 \frac{\partial \psi}{\partial C_i^1} \bigg| C^i_1 = C^i_1 \right) \right\| U_n^i - \sqrt{3} \sigma^\text{reo} \leq 0. \]  
(64)

However, note that for \( \xi = 0 \), and thus \( C_i^1 = C^i_1 \), we have a simplified yield criterion
\[ f = \left\| R_n^i \left( 2 U_n^i \frac{\partial \psi}{\partial C_i^1} \bigg| C^i_1 = C^i_1 \right) U_n^i \right\| - \sqrt{3} \sigma^\text{reo} \]  
(65)

where the superscript 'tr' denotes the trial state and \( R^i = F^i U^{i-1} \). This result again shows the consistency with the time-continuous theory (see Eq. (33)).

3.4 Algorithmic transformation criteria

In analogy to Sect. 3.3, the activation criteria for forward ('M', \( \Delta \xi \geq 0 \)) and reverse ('A', \( \Delta \xi < 0 \)) transformations read
\[ g_M = - \frac{\partial}{\partial \xi} \pi_\Delta \bigg| \Delta \xi \rightarrow 0^+, \quad g_A = - \frac{\partial}{\partial \xi} \pi_\Delta \bigg| \Delta \xi \rightarrow 0^-, \]  
(66)

where \( 0^+ \) and \( 0^- \) denote the right- and left-hand limit (recall that \( \text{sg} (\xi) \) is an argument of \( Q \) in Eq. (41)).

3.5 Initial guess for \( \tilde{N}^s \)

In general, the direction of inelastic flow \( \tilde{N}^s \) is determined by the minimization of \( \pi_\Delta \) through a Newton scheme, which requires a reasonable initial guess when \( \Delta \gamma \) is activated. Coincidentally, an analytical solution for \( \Delta \gamma \rightarrow 0 \) exists for \( \tilde{N}^s \):
\[ \tilde{N}^{s,\text{tr}} = \frac{C_n^i \text{DEV} C_i \left( -2 \frac{\partial \psi}{\partial C_i^1} \bigg| C^i_1 = C^i_1 \right) C_n^i}{\text{DEV} C_i \left( -2 \frac{\partial \psi}{\partial C_i^1} \bigg| C^i_1 = C^i_1 \right) U_n^i}. \]  
(67)

The proof is given in Appendix B.
It is noted that for $\Delta \gamma = 0$, it follows that $C^i = C^i_r$, which is independent of $\tilde{N}^s$. That means that any choice of $\tilde{N}^s$ minimizes $\pi_\Delta$, i.e., there is no unique solution. For very small $\Delta \gamma$, the sensitivity of $\pi_\Delta$ with respect to $\tilde{N}^s$ is also small, which can lead to a bad condition of the nonlinear equation system, which needs to be solved to minimize $\pi_\Delta$. To stabilize the solution process, the last term in Eq. (55) is added to $\pi$, since for $\Delta \gamma \to 0$ it is known that $\tilde{N}^s = \tilde{N}^{s,lf}$ minimizes $\pi_\Delta$. In theory, $\epsilon$ can be chosen arbitrarily small, in practice a finite value is necessary due to the limited numerical accuracy. Thus, the last term in Eq. (55) is used to ‘guide’ the algorithm toward the right solution when it is no longer able to find it by itself.

3.6 Computing derivatives of $\Delta \alpha$ for $\Delta \gamma \to 0$

The derivatives of $\Delta \alpha$ are numerically tough to obtain for $\Delta \gamma \to 0$, due to an almost zero denominator of $\tilde{M}$ (see Eq. (112) and preceding equations). To overcome this issue, we derived the derivatives of $\Delta \gamma$ with respect to $\xi$, $\tilde{N}^s$ and $\Delta \gamma$ separately for $\Delta \gamma \to 0$, which are then numerically feasible to obtain. However, this seems to be a rather theoretical as we never observed the case that $\Delta \gamma$ was too close to zero to obtain the derivatives using the usual way shown in Appendix C. Therefore, and for brevity, we do not discuss them here.

3.7 Regularization of $\psi_h$

Numerically, $\psi_h$ is challenging, since with $\langle E^i \rangle \to 1$, $\psi$ and $\partial \psi / \partial \langle E^i \rangle$ go to infinity. This is very demanding, because values of $\langle E^i \rangle \geq 1$ can occur during the iterative solution process of the Newton scheme, leading to a loss of convergence. The term (compare to Eq. (21))

$$\varphi(\langle E^i \rangle) = \frac{(\langle E^i \rangle)^2}{1 - (\langle E^i \rangle)^4}$$

is illustrated in Fig. 2 in black.

To stabilize the solution process, we partially replace the function $\varphi$ by a linear approximation in the region where $\langle E^i \rangle > c$, where $c$ is close to 1 (see Fig. 2 in green). This regularization approach is adopted from crystal plasticity, where it successfully improved the numerical treatment of the power law in Wulfinghoff and Böhlke [52]. In order to prevent the solution $\langle E^i \rangle$ from taking nonphysical solutions beyond 1, we add a penalty-type energy as illustrated in Fig. 3. We penalize all states $(\langle E^i \rangle, \xi)$ (see Fig. 3) below the dashed half-line starting at $\xi_0$ by the energy

$$\psi_p = \begin{cases} 
\frac{1}{2} H_p l_F^2 & \xi < \frac{\langle E^i \rangle}{c} + \xi_0 \\
0 & \text{else}
\end{cases}$$

where $l_F$ is the minimum distance of the point $(\langle E^i \rangle, \xi)$ from the half-line and $H_p$ is a large penalty parameter.

![Fig. 2 Linear approximation of $\psi_h$](image-url)
3.8 Viscosity of the martensite volume fraction $\xi$

To further improve the numerical robustness, we added an artificial viscosity-like term for the martensite volume fraction. Since it is a dissipative term, we add this term to the discretized dissipation potential:

$$\phi_{\Delta \eta} = \frac{1}{2 \Delta t} \eta_\xi \Delta \xi^2,$$

where $\eta_\xi$ is a small, positive constant.

3.9 Thermomechanical coupling

Like the displacements, the temperature is an unknown for our model. Both fields are coupled through the thermomechanical energy and through the dissipative terms, which lead to a heating of the material. Additionally, the temperature changes the material behavior and leads to thermal strains. Both fields are coupled using a potential-based monolithic approach similar to the one presented in Yang et al. [54]. Further, we assume Fourier’s heat conduction law with heat conductivity $\kappa$. The thermomechanic quasistatic time-discrete potential then reads

$$\pi_{\Delta} = \psi - (u_n - \theta s_n) + \theta_n \phi_{\Delta} - \frac{\Delta t}{\theta_n} \kappa \| \text{Grad}(\theta) \|^2 + \Delta t \frac{\theta}{\theta_n} \kappa \| \text{Grad}(\theta_n) \|^2,$$

where we neglect the influence of body forces. The integral form $\Pi_{\Delta}$ of the local potential $\pi_{\Delta}$ is then given by

$$\Pi_{\Delta} = \int_{V_0} \pi_{\Delta} dV - \int_{\partial V_0^t} \tilde{t} \cdot udS - \int_{V_{0Q}} \frac{\Delta t}{\theta_n} \bar{Q} \theta dS + \int_{V_0} \frac{\Delta t}{\theta_n} \theta w dV,$$

where $w$ denotes the heat source and $\tilde{t}$ as well as $\bar{Q}$ denote prescribed tractions and normal heat flux at the Neumann-type boundaries $\partial V_0^t$ and $\partial V_{0Q}$, respectively. Now, the classical weak form of the linear momentum balance is obtained by variation of Eq. (71) with respect to $u$:

$$\delta_u \Pi_{\Delta} = \int_{V_0} \tau : d_\delta dV - \int_{\partial V_{0u}} \tilde{t} \cdot \delta u dS \overset{1}{=} 0,$$

with $d_\delta = \text{sym}(\text{Grad}(\delta u) F^{-1})$. Likewise, we obtain the weak form of the energy balance by variation of Eq. (71) with respect to $\theta$:
Subsequently, we find \( u, \theta, \xi, C^i \) by finding the solution to the saddle point problem

\[
(u, \theta, \xi, C^i) = \arg \inf_{u \in \mathcal{K}_u} \sup_{\theta \in \varnothing} \inf_{z} \Pi_{\Delta},
\]

where \( \mathcal{K}_u = \{ u : u = \bar{u} \text{ on } \partial V_0 \} \) is the set of admissible displacements satisfying the Dirichlet boundary conditions imposed on the boundary \( \partial V_0 \) and \( \varnothing = \{ \theta : \theta = \bar{\theta} \text{ on } \partial V_0 \} \) is the set of admissible temperatures satisfying the Dirichlet boundary conditions imposed on the boundary \( \partial V_0 \). Additionally, \( z \) is the vector of internal variables, i.e., \( z = (\xi, C^i) \).

Manipulating Eq. (74) using the definition \( s = -\partial \psi / \partial \theta \) as well as applying Gauss theorem, we obtain

\[
\delta_{\theta} \Pi_{\Delta} = \int_{V_0} \left( -\Delta s + \frac{1}{\theta_n} \phi_n + \frac{\kappa \Delta t}{\theta_n} \text{Div} (\text{Grad}(\theta)) - \frac{\kappa \Delta t}{\theta_n^2} \text{Grad}(\theta) \cdot \text{Grad}(\delta \theta) + \frac{\Delta t}{\theta_n} w \right) \delta \theta dV + \int_{\partial V_0 Q} -\Delta t \kappa \cdot \text{Grad}(\theta) \cdot \frac{N}{\theta_n} \delta \theta dS - \int_{\partial V_0 Q} \frac{\Delta t}{\theta_n} \bar{Q} \delta \theta dS = 0,
\]

where \( N \) is the external normal on the boundary \( \partial V_0 \) in the reference configuration. To show consistency with the time-continuous theory, we take a look at the integrand over the volume integral in Eq. (76). By multiplying with \( \theta_n / \Delta t \), we get

\[
\frac{\theta_n}{\Delta t} \frac{\Delta s}{\Delta t} = \frac{1}{\Delta t} \phi_n - \text{Div} (-\kappa \text{Grad}(\theta)) + w + \frac{\kappa}{\theta_n} \text{Grad}(\theta_n) \cdot (\text{Grad}(\theta_n) - \text{Grad}(\theta)).
\]

Here, we want to note that this requires \( \kappa \) to be independent of the temperature \( \theta \). However, one possibility is to consider \( \kappa = \kappa(\theta_n) \), which could circumvent this limitation.

Now, for \( \Delta t \to 0 \), we get (also see Eq. (35))

\[
\theta \dot{s} = \phi - \text{Div} (-\kappa \text{Grad}(\theta)) + w.
\]

Hence, we arrived at the energy balance (Eq. (35)) with Fourier’s law \( Q = -\kappa \text{Grad}(\theta) \), which proves the consistency with the time-continuous theory. Further, the surface integrals in Eq. (76) imply the Neumann boundary condition \( \bar{Q} = -\kappa \text{Grad}(\theta) \cdot N \) on \( \partial V_0 Q \). Alternatively, one can include Robin-type boundary conditions into the model by replacing the integrand \( \Delta t \bar{Q} \phi / \theta_n \) in Eq. (72) by \( 1/2 \Delta t h (\theta - \theta_s)^2 / \theta_n \), where \( h \) is the heat convection coefficient and \( \theta_s \) is the temperature of the surrounding medium. In that case, the variation of \( \Pi_{\Delta} \) yields the boundary condition \(-\kappa \text{Grad}(\theta) \cdot N = h (\theta - \theta_s) \) on \( \partial V_0 Q \).

### 3.10 Active set search

When solving the set of Eq. (56), one has to decide which variables will evolve using the activation criteria given in Eqs. (64) and (66). They will then be put into the active array of variables \( A \). If then, at a later state, an activation criterion is inactive, the variable is taken from \( A \) again. This is done using the active set search algorithm outlined in Algorithm 1.

The algorithm is structured as follows: First, if \( A = \emptyset \) or no variable from \( A \) was rescaled (see the end of this subsection for an explanation of rescaling) during the last iteration, we need to evaluate the activation criteria. With them at hand, we determine our new set \( A^{j+1} \). The details for determining which variable to activate are given in Algorithm 2. Subsequently, we check if the active set changed from the last iteration. If this is not the case, the solution from the last iteration is confirmed as solution of the equation system. In that case, we compute the algorithmic tangent (see Appendix E for details), save the history variables and exit the material routine. Otherwise, we solve the minimization problem for the current active set \( A^{j+1} \). This is done
Algorithm 1 Active set algorithm solving the minimization problem in Eq. (56).

\[
\text{loop} \\
\quad j \leftarrow j + 1 \\
\quad \text{if no rescaling done in last step or } A^j = \emptyset \text{ then} \\
\quad \quad \text{determine } A^{j+1} \text{ (see Algorithm 2)} \\
\quad \text{end if} \\
\quad \text{if } A^{j+1} = A^j \text{ then} \\
\quad \quad \text{compute } c^j, \frac{\partial \pi}{\partial \gamma^j}, \frac{\partial \psi}{\partial \gamma^j}, \frac{\partial \tau}{\partial \gamma^j} \\
\quad \quad \text{save history variables} \\
\quad \quad \text{exit} \\
\quad \text{end if} \\
\quad \text{end loop} \\
\]

\[
\xi_0^* = \xi^j, \Delta \gamma_0^* = \Delta \gamma^j, \hat{N}^{\xi \gamma}_0 = \hat{N}_{0}^{\xi \gamma} \\
\text{loop} \\
\quad k \leftarrow k + 1 \\
\quad \text{get } \frac{\Delta \xi_k^*}{\Delta \gamma_k^*} \text{ based on } \xi_k^*, \Delta \gamma_k^* \text{ and } \hat{N}^{\xi \gamma}_k \\
\quad \text{if } k > \text{maxIter or } \left\| \partial \pi / \partial A^{j+1} \right\| < \text{tolNwtn} \text{ then} \\
\quad \quad \text{exit} \\
\quad \text{else} \\
\quad \quad \text{calculate } \Delta \xi_k^*, \Delta \gamma_k^* \text{ if in } A^{j+1} \\
\quad \quad \text{if } \xi_k^* + \Delta \xi_k^* < 0 \text{ then} \\
\quad \quad \quad \text{rescale } \Delta \xi_k^*, \Delta \gamma_k^* \text{ and } \hat{N}^{\xi \gamma}_k \text{ s.t. } \xi_k^* = \xi \text{ hard} \\
\quad \quad \text{end if} \\
\quad \quad \text{update } \xi_{k+1}^* = \xi_k^* + \Delta \xi_k^*, \Delta \gamma_{k+1}^* = \Delta \gamma_k^* + \Delta \gamma_{k+1}^*, \hat{N}_{k+1}^{\xi \gamma} = \hat{N}^{\xi \gamma}_k + \Delta \hat{N}^{\xi \gamma}_k \text{ according to } A^{j+1} \\
\quad \quad \text{end if} \\
\quad \text{end loop} \\
\quad \text{if } \Delta \xi_{k+1}^* \text{ goes in dir. not matching } g_A/g_M \text{ then} \\
\quad \quad \text{rescale } \Delta \xi_{k+1}^*, \Delta \gamma_{k+1}^* \text{ and } \hat{N}^{\xi \gamma}_{k+1} \\
\quad \quad \text{xiJustDeactivated } \leftarrow \text{true} \\
\quad \text{else} \\
\quad \quad \text{xiJustDeactivated } \leftarrow \text{false} \\
\quad \text{end if} \\
\quad \Delta \gamma_{j+1}^* = \Delta \gamma_j^* + \Delta \xi_{k+1}^* \text{ and } \hat{N}^{\xi \gamma}_{j+1} = \hat{N}^{\xi \gamma}_j + \Delta \hat{N}^{\xi \gamma}_j \\
\text{end loop} \\
\]

Algorithm 2 Algorithm deciding the priority when updating $A$.

\[
\text{compute } f \\
\text{if } \xi \notin A^j \text{ then} \\
\quad \text{compute } g_A \text{ and } g_M \\
\text{end if} \\
\text{if } A^j = \emptyset \text{ then} \\
\quad \text{if } g_A < 0 \text{ and not } \text{xiJustDeactivated} \text{ then} \\
\quad \quad \text{add } \xi \text{ to } A^{j+1} \\
\quad \text{else if } g_M > 0 \text{ and not } \text{xiJustDeactivated} \text{ then} \\
\quad \quad \text{add } \xi \text{ to } A^{j+1} \\
\quad \text{else if } f > 0 \text{ then} \\
\quad \quad \text{add } \Delta \gamma \text{ to } A^{j+1} \\
\quad \text{end if} \\
\quad \text{else if } \xi \in A^j \text{ and } \Delta \gamma \notin A^j \text{ then} \\
\quad \quad \text{if } f > 0 \text{ then} \\
\quad \quad \quad \text{add } \Delta \gamma \text{ to } A^{j+1} \\
\quad \quad \quad \text{compute new } \hat{N}^{\xi \gamma} \\
\quad \quad \text{end if} \\
\quad \text{else if } \Delta \gamma \in A^j \text{ and } \xi \notin A^j \text{ and not } \text{xiJustDeactivated} \text{ then} \\
\quad \quad \text{if } g_A < 0 \text{ then} \\
\quad \quad \quad \text{add } \xi \text{ to } A^{j+1} \\
\quad \quad \text{else if } g_M > 0 \text{ then} \\
\quad \quad \quad \text{add } \xi \text{ to } A^{j+1} \\
\quad \quad \text{end if} \\
\quad \text{else if } f > 0 \text{ then} \\
\quad \quad \text{compute new } \hat{N}^{\xi \gamma} \\
\quad \text{end if} \\
\]

\[\text{Eq. (65), (165), (171), (172), (87), (119)-(124)}\]
using a Newton scheme. First, we compute our residual and stiffness concerning active set \( A^{j+1} \), i.e., only derivatives with respect to the active set are computed. If the maximum iterations are exceeded or the norm of the residual is lower than the tolerance, we exit the loop. Otherwise, we check if updating \( \xi^{j+1} \) would result in a negative \( \xi \) when applying the Newton step. If this is the case, we rescale, i.e., multiply the increments by a scalar such that \( \xi_\text{hard}^{j+1} \) is set to \( \xi_\text{min} \), which is \( 5 \times 10^{-5} \) in this work. Subsequently, we compute \( \Delta \xi^{k+1} \), \( \Delta Y^{j+1} \), and \( \tilde{N}^{s,j+1} \). In fact, we do not update \( \xi^{j+1} \), \( \Delta Y^{j+1} \), or \( \tilde{N}^{s,j+1} \) yet. We update them only after exiting the Newton scheme and being sure that \( \xi \) incremented in a direction matching \( g_A \) and \( g_M \), which were calculated before the Newton loop. If, due to a bad starting solution, this is not the case, we rescale the Newton step such that \( \xi^{j+1} = \xi_n \). If this happens, we make sure that in the next iteration of the active set algorithm, we won’t activate \( \xi \) again.

4 Numerical results

In this section, the previously presented model is tested for different examples. First, to test the model’s time convergence behavior and to show the superelastic as well as martensite reorientation behavior at high and low temperatures, respectively, thermomechanical Gauss point evaluations are conducted. Finally, a full actuator model is simulated. To cope with the thin structures occurring in the actuator, the SMA model at hand is embedded into a hexahedral element formulation with reduced integration and hourglass stabilization for the displacement, while the heat conduction terms are fully integrated.

4.1 Gauss point evaluations

For the Gauss point evaluations, thermal expansion as well as the transformation induced volume change, which is discussed in Sect. 2.3.4, are neglected for simplicity. The simulations are conducted using a reduced integration hexahedral element with hourglass stabilization, which is embedded into the finite element program FEAP [46]. The material constants used in the Gauss-point are given in Table 1.

Here, it is noted that \( \theta_{\text{ref}A} \) and \( \theta_{\text{ref}A} \) are chosen such that \( \Delta V = 0.47\% \) (see Eq. (23)), matching the findings reported in Potapov et al. [36]. The numerical parameters are summarized in Table 2.

Figure 4 shows a tensile test at a temperature of 270°C for 49 and 40,000 time steps. Clearly, convergence with regard to time step width is not an issue, as both time step widths yield accurate results.

To demonstrate that the model results indeed do not depend on the numerical parameters, we compare results for two different sets of numerical parameters in a tensile test. The first set of numerical parameters is the one given in Table 2, which is used in the remainder of this paper. The second set is defined in Table 3 just

| Table 1 Material parameters |
|-----------------------------|
| **Elastic constants**       |
| \( \lambda_A \) [GPa]       | \( \lambda_M \) [GPa] | \( \mu_A \) [GPa] | \( \mu_M \) [GPa] |
| 23.076                      | 13.461                | 34.615            | 20.192            |
| **Dissipation potential parameters** |
| \( \theta_0 \) [°C]         | \( M_s \) [°C]       | \( M_f \) [°C]    | \( A_s \) [°C]    | \( A_f \) [°C]    | \( \sigma^{\text{reo}} \) [MPa] |
| 80                          | 70                    | 50                 | 120               | 140               | 160               |
| **Chemical and hardening parameters** |
| \( \Delta \lambda^{AM} \) [MPa K\(^{-1}\)] | \( E^{\text{int}} \) [MPa] | k [-] |
| 0.383                       | 30                    | 0.12               |
| **Thermal constants**       |
| \( \alpha_M \) [10\(^{-6}\) K\(^{-1}\)] | \( \alpha_A \) [10\(^{-6}\) K\(^{-1}\)] | c [mJ mm\(^{-3}\) K\(^{-1}\)] | \( \kappa \) [W m\(^{-1}\)] | \( \theta_{\text{ref}M} \) [°C] | \( \theta_{\text{ref}A} \) [°C] |
| 6.6                         | 11.0                  | 2.9                | 8.6               | 588.28            | 500.0             |
Table 2 Numerical parameter set 1 for simulation

| Numerical parameters | c [-] | $H_p$ [MPa] | A [MPa] | $B$ [-] | $\xi_0$ [-] | $\epsilon$ [MPa] | $\eta_\xi$ [MPas$^{-1}$] |
|----------------------|-------|-------------|--------|--------|------------|-------------|------------------|
| 0.99                 | $10^8$ | $10^5$      | 1      | $10^{-4}$ | $10^{-4}$  | 0.01        |                  |

Fig. 4 Tensile test for 49 and 40,000 time steps. No error due to large load steps is visible for this comparison. The results are shown in Fig. 5. Clearly, the numerical parameters do not have a noticeable effect on the results.

Next, we modeled tensile tests at 50 °C, 100 °C, 160 °C and 200 °C. The resulting stress–strain curves are shown in Fig. 6.

For low temperatures, i.e. 50 °C and 100 °C, the model predicts a remaining martensite reorientation. On the other hand, for rather high temperatures, the model captures superelastic material behavior when unloading. However, we want to emphasize that these results do not consider any damage or plastic deformations which might already occur.

Finally, the model captures the shape memory effect well. This is shown in Fig. 7, where we applied different stresses (160 MPa, 300 MPa, 400 MPa and 500 MPa) and then started a thermal cycle.

Table 3 Numerical parameter set 2, only for comparison in Fig. 5

| Numerical parameters | c [-] | $H_p$ [MPa] | A [MPa] | $B$ [-] | $\xi_0$ [-] | $\epsilon$ [MPa] | $\eta_\xi$ [MPas$^{-1}$] |
|----------------------|-------|-------------|--------|--------|------------|-------------|------------------|
| 0.98                 | $10^7$ | $10^6$      | 1      | $10^{-3}$ | $10^{-3}$  | $10^{-3}$   |                  |

Fig. 5 Investigation of the numerical parameters influence (set 1 is given in Table 2, set 2 in Table 3)
A thermomechanical finite strain shape memory alloy model

Fig. 6 Tensile test at various temperatures

\[
\tau_{11} = 160 \text{ MPa} \quad \tau_{11} = 300 \text{ MPa} \quad \tau_{11} = 400 \text{ MPa} \quad \tau_{11} = 500 \text{ MPa}
\]

Fig. 7 Thermal cycling tests with different stresses (160 MPa, 300 MPa, 400 MPa and 500 MPa)

For 160 MPa, one obtains only twinned martensite, i.e., \(E^t = 0\) in the context of this work. Therefore, one only gets a small hysteresis due to the different elastic constants of austenite and martensite. For the larger prestresses, the austenite is transformed into detwinned martensite, i.e., \(\langle E^t \rangle \neq 0\). This leads to the typical shape memory effect.

4.2 Plate with a hole

In this example, we simulate a plate with a cylindrical hole (see Fig. 8), which is at first loaded by a traction \(\tilde{t}\), then unloaded and subsequently heated, which lets it recover the initial shape.

The material and numerical parameters are unchanged from Sect. 4.1, except the reference temperatures are now \(\theta_{\text{refA}} = 80^\circ \text{C}\) and \(\theta_{\text{refM}} = 87.66^\circ \text{C}\). Due to the symmetry, using appropriate symmetry conditions, only one-eighth of the entire plate is simulated.

During the loading by the traction, the temperature at the upper and lower end is held constant at the initial temperature of \(\bar{\theta} = 80^\circ \text{C}\) \((M_s < \bar{\theta} < A_s)\). Additionally, in the initial state, the plate is fully austenite. First, the traction \(\tilde{t}\) is increased linearly to a maximum of 330 MPa in longitudinal direction over the duration of 50s. Subsequently, the traction \(\tilde{t}\) is decreased to zero over the duration of 50s. Finally, the temperature at both ends of the plate is increased to \(200^\circ \text{C}\) over the course of 20s.

The mesh used in the eighth of the plate consists of 2907 uniformly distributed elements and is shown in Fig. 9.

For the entire simulation, 293 time steps were required, taking a total CPU-time of 2524s.
At first, during the loading, almost the entire plate is transformed to martensite. The plate at maximum traction is shown in Fig. 10a. Here, the $E_{xx}$ strain at the edge of the hole reaches roughly up to 18% (see Fig. 11).

Additionally, the plate heats up slightly due to the latent heat and mechanical dissipation (Fig. 10a). However, most of the heat is conducted out of the plate at the temperature Dirichlet boundaries due to the small size of the plate and the long simulated time. Subsequently, after unloading, the plate does not transform back, which is shown in Fig 10b. Finally, when increasing the temperature at both ends, the martensite is transformed back to austenite, leading to a recovery of the initial shape (see Fig. 10c). Further, Fig. 12 shows the temperature and displacement of point A (see Fig. 8) over time. Here, the increase in slope of the displacement at roughly 15s is the point where we have a forward transformation in the vicinity of the hole. Then, the second change in slope of the displacement at roughly 36s is the rest of the plate starting to undergo forward transformation. At 50s, when releasing the traction, the displacements decrease. Then again, due to thermal expansion, the displacement increases slightly after increasing the temperature at the ends at 100s. However, when the temperature reaches the backward transformation temperatures, the material almost

---

5 A separate simulation with “zero heat flux” boundary conditions led to a temperature increase in approximately 50°C.
recovers the initial shape. The nonzero displacement at $t = 120\text{s}$ stems from thermal expansion and the $\Delta V$ effect.

Overall, the examples have been chosen such that the stresses in the vicinity of the hole reach unphysical values, at which a real material is expected to show irrecoverable strains or even fracture. Thus, this simulation rather serves as an example showing that the implementation is able to find a solution, even under severe loads.

4.3 Finite element actuator model

In this section, we model a bistable shape memory microactuator using in the finite element program FEAP [46]. The actuator concept was published by Winzek et al. [50] for large structures and is built with three layers: a bottom layer of molybdenum, a middle layer of NiTiHf and a top layer of polymethyl methacrylate (PMMA) (Winzek et al. [50] used PMMA layers on both sides). For simplicity, a hexahedral element formulation with reduced integration and hourglass stabilization for the displacements is used for all materials. The material and numerical parameters for the shape memory alloy remain unchanged (see Tables 1 and 2).

4.3.1 Actuating principle

The actuator works through interlacing of the large hysteresis of NiTiHf with the polymers hysteresis. This is shown in Fig. 13.

It is actuated by joule heating to specific temperatures and cooling back to room temperature. Depending on the heat cycle, the martensite state or austenite state is held in place by the polymer. This is understood best when looking at the following example, which is depicted in Fig. 14, where we neglected the CTE of PMMA.
A bimorph with SMA and molybdenum (left) in comparison to the bistable actuator with an additional layer of PMMA (right). The temperatures are indicated by different thermometer colorings.

Fig. 14 A bimorph with SMA and molybdenum (left) in comparison to the bistable actuator with an additional layer of PMMA (right). The temperatures are indicated by different thermometer colorings.

For simplicity. Here, we compare a bimorph of SMA and molybdenum on the left to the proposed actuator on the right. When heating from ambient temperature (blue, Fig. 14a) to a temperature above $A_f$ (red), the glass transition temperature is reached, which drastically decreases the stiffness of the PMMA. Subsequently, the SMA in the bimorph as well as the bistable actuator reach reverse transformation temperatures, which makes them bend up (Fig. 14b). Now, upon cooling, the actuator reaches $\theta_g$, which ‘freezes’ its current shape. Thus, when subsequently reaching forward transformation temperatures, unlike the bimorph, it will not revert into the original shape at room temperature (Fig. 14c). Now, we can heat the actuator to a temperature above $\theta_g$ but below $A_s$ (orange in Fig. 14d) to soften the polymer, which makes the actuator adopt the deformation of the bimorph. When now cooling back to room temperature again, the actuator is in its initial stable state again and the actuation cycle could be started again. The two stable states at room temperature are given by Figs. 14c, e.

For the actuation behavior, it is important to have a homogeneous temperature profile in the actuator. Therefore, a wing-shape actuator developed in Arivanandhan et al. [2] is used to obtain rather homogeneous temperatures in the double beam cantilever.

To rapidly optimize the actuators properties, it turned out that building and modeling only the beam section of the actuator is advantageous. Therefore, we simulated just the beam part with suitable boundary conditions instead of simulating the entire wing structure including the attached wafer material. This thin film actuator with a zoom into the layers is shown in Fig. 15, where the Dirichlet boundary condition for the displacements and the Robin boundary condition for the temperature are indicated as well.

Fig. 15 Zoom into the layers of the actuator (molybdenum in red, NiTiHf in blue and PMMA in orange)
4.3.2 Polymer and molybdenum model

Since the interest in modeling the actuator lies rather in the states at $\theta_1$, $\theta_2$ and $\theta_3$, and less in the states between them, we chose a very simple polymer model. It is governed by a thermally coupled viscoelastic Maxwell model for finite strains (Young’s modulus and Poisson’s ratio are 500 MPa and 0.4, respectively), where the viscosity is high ($10^7$ MPa) at low temperatures and low (1 MPa) at high temperatures. Therefore, it has almost no stiffness at high temperatures while it behaves almost elastically at low ones. Additionally, the glass transition temperature is given by $\theta_g = 77^\circ$C. The molybdenum is modeled with a thermally coupled Neo-Hookean elastic model. For the molybdenum, Young’s modulus, Poisson’s ratio, the thermal expansion coefficient as well as the reference temperature are $E = 65 \times 10^3$ MPa, $\nu = 0.31$, $\alpha = 5 \times 10^{-6} \text{K}^{-1}$ and $\theta_{ref} = 500^\circ$C, respectively.

4.3.3 Mesh convergence

The actuator considered is shown in Fig. 15. It is clamped on the left side. It has a length of 10 mm, a width of 5 mm. The layer thicknesses of the molybdenum and TiNiHf are for this work 20 um and 10 um, respectively. The polymer layer thickness is 160 um for now, before different layer thicknesses are compared in Sect. 4.3.4. For the initial conditions, we assume zero displacements as well as a temperature of 500 $^\circ$C. The polymer layer is 160 um thick for now, before different layer thicknesses are compared in Sect. 4.3.4.

The actuator is thermally annealed in a flat state. Furthermore, we assume the material to be in its austenite state at $t = 0$, which directly implies that initially $C^t = I$. Subsequently, it is cooled down to room temperature at 20 $^\circ$C, which bends the actuator due to the mismatch in the coefficients of thermal expansions and difference in cell volume between the austenite and martensite phase.

The heating cycle is realized through applying a heat source in the Mo and NiTiHf material. The heat source magnitude is modeled by a sine function, which is cut off when below 0. It is given by

$$w_h = \max(2250(\sin(0.03t) - 0.6), 0) \text{ mW mm}^{-3}, \quad w_l = \max(0.65 \cdot 2250(\sin(0.03t) - 0.6), 0) \text{ mW mm}^{-3}$$

for the higher and lower heat cycles, respectively.

At the top and at the bottom face of the actuator a Robin boundary condition is applied, which cools the thin film to room temperature over time. Due to their small areas, the heat transfer at the lateral faces is neglected. The surrounding air’s temperature is 20 $^\circ$C while the convective heat transfer coefficient is assumed to be 70 W m$^{-2}$ K$^{-1}$ in accordance with Kohl et al. [23].

For these thin structures, a sufficiently fine mesh is crucial to obtain converged results. Therefore, we conducted convergence studies with regard to the necessary elements in each layer and the amount of elements needed over the length and width of the actuator. First, it turned out that due to the bending deformation of the thin film, it is sufficient to only use one element over the width of the actuator. Then, the convergence with respect to the amount of elements over the length is tested using four elements over the thickness for each material layer. The results are shown in Fig. 16, where the stroke and temperature of the SMA at the tip is plotted over time. Here, only the temperature for 20 elements over the length is shown, since there was virtually no difference in temperature for the different discretizations. Further, we concluded that using 20 elements over the length leads to an acceptable error in stroke while the main features of the actuator are conveyed well. Additionally, roughly 160 time steps were used for one actuation cycle, e.g., from 0 s to 180 s in Fig. 16.

Next, the convergence with regards to the elements used over the depth for each material was studied. The results are depicted in Fig. 17.

Here, for the number of elements over the SMA thickness, one element is already enough to obtain results with an acceptable error. For the molybdenum and polymer layer, one needs at least two elements to obtain converged results.

To illustrate the importance of modeling the $\Delta V$ effect, Fig. 18 shows the stroke and temperature over time for $\Delta V$ being 0 and 0.49%. Here, one can see that the volume change contributes to large parts of the achieved stroke between the two stable states at room temperature.

4.3.4 Influence of the polymer thickness

The polymer layer thickness plays a key role for this actuator design—if it is too thin, the high temperature shape cannot be held at room temperature by the polymer layer. On the other hand, it should not be too thick, since body forces and manufacturing problems come up in that case. Further, since the polymer insulates
the actuator thermally to one side, thinner polymer layers lead to the possibility of faster actuation cycles. Therefore, we tested several polymer layer thicknesses, the results are shown in Fig. 19.

First, at (1) (see Fig. 19), the actuator is heated. At first, the larger CTE of the polymer makes the actuator bend down. Then, when reaching $\theta_g$, the polymer softens up and the actuator relaxes. Subsequently, at (2), the actuator is cooled down to room temperature by the surrounding air. The polymer hardens again and the device reaches stable configuration (3) at room temperature. Now, when heating again, the actuator bends down due to the large CTE of the PMMA at (4). As soon as the temperature reaches $\theta_g$, the polymer softens up. Afterward, at (5) the shape memory alloy reaches $A_f$ and the martensite is transformed back to austenite, which makes the actuator deflect up. Finally, when we remove the source at (6), the polymer hardens before $M_f$ is reached. Thus, depending on the polymer thickness, the polymer may hold the actuators shape or release some of it’s stroke. Finally, the actuator reaches it’s second stable state (6) at room temperature, which is also it’s initial state.
A thermomechanical finite strain shape memory alloy model

3083

Fig. 19 Stroke (left y-axis) and temperature (right y-axis, dashed in black) over time for different PMMA layer thicknesses

For many actuator designs, one wants to maximize the achievable stroke while keeping the device as small and power efficient as possible. Therefore, we try to maximize the difference in stroke between states $A$ and $B$. In turn, one must choose a sufficient polymer thickness, such that the actuator does not release too much stroke at $5^\circ$. For example, 0.1mm of PMMA is not thick enough to hold the shape, while 0.3 mm holds the shape perfectly (see Fig. 19 at $5^\circ$). Depending on the actuation frequency in mind and necessary achievable stroke, either value in between might be a suitable choice.

5 Summary and outlook

In this paper, a new thermomechanical shape memory alloy model for finite strains is presented. It uses a projection method to fulfill the incompressibility constraint for the inelastic stretches. Further, the model is realized in the generalized standard material formulation being extended to thermomechanics. The optimization of the global, incremental mixed thermomechanical potential by variation yields the mechanical balance principles as well as the evolution equations of the internal variables and boundary condition integrals. The presented model employs a thermal strain formulation for the shape memory alloy which allows to describe the transformation induced volume changes found in some shape memory alloys. Using a logarithmic strain formulation, a finite strain dissipation potential incorporates vanishing inelastic strains upon austenite transformation in a manner consistent with the time-continuous case. Additionally, yield and transformation criteria as well as the algorithmically consistent tangent for the coupled problem are given and discussed. Due to numerical difficulties, a regularization of the hardening energy term is implemented.

The numerical results show, that the model is capable of producing reasonable results in tensile tests. The dependency on the temperature at which the tensile tests are carried out is captured accurately. Furthermore, it enables the solution of thin film problems.

Using the model, we are now able to estimate viable layer thicknesses and device sizes as well as fitting joule heating parameters. Additionally, it turned out that only applying polymer to one side increases the maximum actuation frequency while also introducing more deflection between the stable states and requiring more power. Moreover, we found that the $\Delta V$ effect as well as the thermal expansion of the SMA, due to the design being very sensitive to volume changes, is fairly important for the actuators stroke and therefore needs to be modeled accurately as well. Furthermore, the model results show that the bistability at room temperature, which is enabled by the polymer locking in the deformation, is achievable for the shown actuator concept.

In the future, it remains interesting to find ways of solving the model problem without any penalty terms, which are not reasoned physically and are numerically challenging to deal with. A possible solution to this problem could be to make use of the Fischer-Burmeister [12] complementary function. In fact, this has been done in Auricchio et al. [4] for shape memory alloys and in Brepols et al. [7] for damage-plasticity with great success. Furthermore, there is still space for improvements to the algorithms convergence behavior which could increase the speed of the proposed material model, especially under severe loading conditions. Additionally, an inclusion of functional fatigue properties, e.g., a shifting of the TiNiHf transformation temperatures over several transformation cycles into the model would enable to predict the long-term behavior of the actuators.
Acknowledgements  Financial support of subproject A3 Cooperative Actuator Systems for Nanomechanics and Nanophotonics: Coupled Simulation of the Priority Programme SPP 2206 by the German Research Foundation (DFG) (Grant WU847/3-1) is gratefully acknowledged by the authors.

Funding  Open Access funding enabled and organized by Projekt DEAL.

Open Access  This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Declarations  

Conflict of interest  The authors have no competing interests to declare that are relevant to the content of this article.

Appendix  

A Discussion of the dissipation potential $\phi_\Delta$

In this section, we have a close look at the term

$$\| D^i - \xi \| = \| D^i - \frac{\xi}{2\xi} \ln b^i \| \quad \text{for } \xi \leq 0$$

from Eq. (26). If this term is zero, then $b^i \to I$ for $\xi \to 0$.

Proof  Consider $\xi(t) = \xi_0 (1 - \frac{t}{T})$ and $D^i = \frac{\xi}{\xi_0} e^i_0 = -\frac{1}{2T} \ln (b^i_0)$. Then

$$b^i(t) = \exp(2t D^i) b^i_0 = \exp \left( -\frac{t}{T} \ln(b^i_0) b^i_0 \right) = (b^i_0)^{1 - \frac{t}{T}} \to b^i(T) = I.$$  

While this only shows that $b^i$ goes to $I$ for one special $\xi(t)$, one could easily rescale the time to obtain arbitrary $\xi(t)$, for which $b^i \to I$ with $\xi \to 0$.

B Proof of $\tilde{N}^{\text{str}}$ minimizing $\pi_\Delta$ for $\Delta \gamma \to 0$

Proof  We are looking for $\tilde{N}^s$ minimizing $\pi_\Delta$ for $\Delta \gamma \to 0$:

$$\delta \Pi = \frac{\partial \psi}{\partial C^t} C^i = \frac{\partial \psi}{\partial C^t} C^i : \Pi^r_t : (2 \Delta \gamma \delta N^s) \to 0,$$

where the variation is done exclusively with respect to $\tilde{N}^s$. Here, the dissipation potential $\phi_\Delta$ is neglected because

$$\delta \phi_\Delta = \delta \left( \sqrt{\frac{2}{3}} \sigma^{\text{reco} \| 1 \| \Delta \gamma, C^i \| U_{i-1}^{-1} \right) = \frac{1}{2} \delta \left( \sqrt{\frac{2}{3}} \sigma^{\text{reco} \| 1 \| \Delta \gamma, C^i \| U_{i-1}^{-1} \right).$$

Since $\pi_\Delta$ is constructed such that it is minimized by $\Pi^r_t = 1$, we constrain $\tilde{N}^s$ to be in the subspace of symmetric 2nd-order tensors satisfying $\det(C^i) = \det(C^i + 2 \Delta \gamma N^s) = 1$ (i.e., we only search this subspace for the minimizer). Hence, $\Pi^r_t = 1$ and

$$\delta \phi_\Delta = \delta \left( \sqrt{\frac{2}{3}} \sigma^{\text{reco} \| 1 \| \Delta \gamma, C^i \| U_{i-1}^{-1} \right),$$

(83)
which is independent of $\tilde{N}^s$.

With the definition of the tensor

$$\tilde{F}^s_{N^i} = \frac{\partial N^s}{\partial N^i} = \frac{1}{\|\tilde{N}^s\|U^i_{n-1}} (\tilde{\delta}^s - N^s \otimes C_n^{i-1} N^s C_n^{i-1}) ,$$

Eq. (81) becomes

$$\delta \pi = \Delta \gamma U^i_n \left( [\tilde{F}^s_{N^i}]^T (C^i) : \left( 2 \frac{\partial \psi}{\partial C^i} \right) C^i = C^i \right) U^i_n : \delta \frac{U^i_{n-1} \tilde{N}^s U^i_{n-1}}{\|U^i_{n-1} \tilde{N}^s U^i_{n-1}\|} = 0. \quad (85)$$

Since $N^i$ is normalized (i.e., its norm is one), the solution reads

$$N^i = \frac{U^i_n \text{DEV} C^i}{\left\| \text{DEV} C^i \right\|} \left( \left[ \tilde{F}^s_{N^i} \right]^T \left( \frac{-2 \frac{\partial \psi}{\partial C^i} \big|_{C^i = C^i} \right) U^i_n \left( 2 \frac{\partial \psi}{\partial C^i} \big|_{C^i = C^i} \right) \right] \right), \quad (86)$$

which leads to $\tilde{N}^{s,\text{tr}}$ in Eq. (67).

**C First derivatives of $\pi_\Delta$**

To minimize the potential using the active set algorithm including the Newton scheme, we need to obtain first and second-order derivatives of the potential with respect to the set of internal variables and temperature. The differential of the discretized potential with respect to the internal variables ($u$ and $\theta$ are fixed) is given by

$$d\pi_\Delta = \left( \frac{\partial (\psi_c + \psi_h + \psi_r + \psi_p)}{\partial C^i} \right) : \left( \frac{\partial^2 C^i}{\partial \Delta \alpha} + \delta \frac{\partial \psi_c + \psi_h + \psi_r + \psi_p}{\partial \Delta \xi} + \delta \frac{\partial \psi_c + \psi_h + \psi_r + \psi_p}{\partial \Delta \xi} \right) + Q(\xi, \Delta \xi)$$

$$+ H(\text{sgn} (\Delta \xi)) \Delta \xi \right) d\Delta \xi$$

$$+ \left( \frac{\partial (\psi_c + \psi_h + \psi_r + \psi_p)}{\partial C^i} \right) : \left( 2N^s \right) \left( \delta \frac{\partial \psi_c + \psi_h + \psi_r + \psi_p}{\partial \Delta \gamma} \right) d\Delta \gamma$$

$$+ \left( \frac{\tilde{F}^s_{N^i}}{2 \Delta \gamma} \right) \left( \delta \frac{\partial (\psi_c + \psi_h + \psi_r + \psi_p)}{\partial C^i} \right) + \delta \frac{\partial \psi_c + \psi_h + \psi_r + \psi_p}{\partial \Delta N^s} d\Delta N^s, \quad (87)$$

where, for simplicity, the factor $\theta / \theta_t$ is neglected in front of $\delta \frac{\sigma^{\text{reo}}}{\partial \Delta \xi}$. Here, the tensor $\tilde{F}^s_{N^i}$ is given by (see Eq. (59))

$$\tilde{F}^s_{N^i} = \frac{\partial N^s}{\partial N^i} = \frac{1}{U^i_{n-1} \tilde{N}^s U^i_{n-1}} (\tilde{\delta}^s - N^s \otimes C_n^{i-1} N^s C_n^{i-1}) . \quad (88)$$

Further, the derivative of $C^i$ with respect to $\Delta \xi$ is (see Eq. (40))

$$\frac{\partial C^i}{\partial \Delta \xi} = \frac{H_-(\Delta \xi)}{\xi_n} \ln(C^i_n) C^i_n, \quad (89)$$

where $H_-(\Delta \xi)$ is the double negative Heaviside function: $H_-(\Delta \xi) = -H(-\Delta \xi)$. The remaining occurring derivatives are given by

$$\frac{\partial (\psi_c + \psi_h + \psi_p)}{\partial C^i} \left( \frac{\tilde{F}^s_{N^i}}{2 \Delta \gamma} \right) = \frac{\partial (\psi_c + \psi_h + \psi_p)}{\partial C^i} , \quad (90)$$
\[ \frac{\partial \psi_e}{\partial C^i} = -\frac{1}{2} F^T \frac{\partial \psi_e}{\partial b^e} F C^{i-1}. \]  

(91)

\[ \frac{\partial \psi_e}{\partial b^e} = \left( \frac{\lambda(\xi)}{4} \left( \text{det}(b^e) - 1 \right) - \frac{\mu(\xi)}{2} \right) B^{e-1} + \frac{\mu(\xi)}{2} I. \]  

(92)

\[ \nabla \lambda \left( \frac{1}{\lambda_A} + \frac{\xi}{\lambda_M} \right)^{-2} \left( \frac{1}{\lambda_A} + \frac{1}{\lambda_M} \right), \]  

(94)

\[ \frac{\partial \lambda}{\partial \xi} = -\left( \frac{1 - \xi}{\lambda_A} + \frac{\xi}{\lambda_M} \right)^{-2} \left( \frac{1}{\lambda_A} + \frac{1}{\lambda_M} \right), \]  

(95)

\[ \frac{\partial \psi_e}{\partial \xi} = \left( \theta - \theta_0 \right) \Delta s^{AM}. \]  

(96)

\[ \frac{\partial \psi_h}{\partial C^i} = \frac{1}{2} \tilde{\gamma} \frac{\partial (E^i)}{\partial E^i} \frac{\partial (E^i)}{\partial C^i}. \]  

(97)

\[ \frac{\partial (E^i)}{\partial E^i} = \frac{1}{\tilde{\gamma}^{\frac{1}{2}}} \frac{E^i}{\|E^i\|}. \]  

(100)

\[ \frac{\partial \psi_h}{\partial \xi} = \frac{\partial \psi_h}{\partial (E^i)} \frac{\partial (E^i)}{\partial \xi} + \frac{\partial (E^i)}{\partial \xi} \frac{\partial (E^i)}{\partial E^i} \frac{\partial (E^i)}{\partial C^i}, \]  

(99)

\[ \frac{\partial (E^i)}{\partial \xi} = -\frac{(E^i)}{\xi^2}. \]  

(101)

\[ \frac{\partial \psi_e}{\partial \xi} = A \left( \tilde{\lambda}^i - 1 \right) \tilde{\lambda}^{i-1} \tilde{C}^{i-1}. \]  

(103)

\[ \frac{\partial \psi_e}{\partial N^s} = A \left( \tilde{N}^s \right)_{U_{n-1}} - 1 \left( \tilde{C}^{i-1} \tilde{N}^s \right)_{U_{n-1}} \right), \]  

(104)

\[ \frac{\partial \psi_p}{\partial C^i} = \frac{\partial \psi_p}{\partial (E^i)} \frac{\partial (E^i)}{\partial C^i} \]  

(105)

\[ \frac{\partial \psi_p}{\partial (E^i)} = H_p \frac{\xi}{\sqrt{c + 1}} \]  

(107)

\[ \frac{\partial \psi_p}{\partial \xi} = \begin{cases} H_p \frac{\xi}{\sqrt{c + 1}} & \xi \geq \xi_0 - c(E^i), \\
H_p \xi \left( \xi - \xi_0 + \xi \left( E^i \right)^2 \right) & \text{else} \end{cases} \]  

(108)
\[
\frac{\partial \Delta \alpha}{\partial \Delta \xi} = \sqrt{\frac{2}{3} \frac{\partial}{\partial \Delta \xi} \hat{M}_C : (\hat{P}_1 - t^s) : \frac{\partial}{\partial \Delta \xi} C_i^j},
\]
(109)

\[
\frac{\partial \Delta \alpha}{\partial \Delta \gamma} = \sqrt{\frac{2}{3} \frac{\partial}{\partial \Delta \gamma} \hat{M}_C : \hat{P}_1 : N^s,
\]
(110)

\[
\frac{\partial \Delta \alpha}{\partial N^s} = \sqrt{\frac{2}{3} \frac{\partial}{\partial N^s} \hat{M}_C : \hat{P}_1^T : \hat{M}_C^i},
\]
(111)

\[
\hat{M}_C = \frac{\partial}{\partial C_i^j} \frac{\partial \Delta C^i}{\partial U^{i,j}_{s-1}} = \frac{C_n^{i,j} - 1 \Delta C_i^{i,j} C_n^{i,j-1}}{\|\Delta C_i^j\|^2 U^{i,j}_{s-1}}.
\]
(112)

Here, \(a_1\) is the slope of the linear approximation in Fig. 2 at \(\langle E^0 \rangle = c k\) and \(a_0\) is the value of the function \(\phi\) at that same point. Further, the terms related to \(\psi_p\) are set to zero if \(\xi > \langle E^1 \rangle/c + \xi_0\).

**D Second derivatives of \(\pi_\Delta\)**

The second derivatives of the potential \(\pi_\Delta\) are necessary to compute the algorithmic tangent. They are briefly given in this appendix. First, we define the symmetrizing box product as

\[
A \square B : C = \text{Asym}(C) B,
\]
(113)

where \(A\) and \(B\) are arbitrary second-order tensors. Now, we compute the differential of the operator \(\hat{P}_{N^s}:\)

\[
d(\hat{P}_{N^s} : B) = \Box_{N^s}(B) : d\hat{N}^s + \hat{P}_{N^s}^T : d\hat{B},
\]
(114)

where \(B\) is an arbitrary symmetric second-order tensor and

\[
\Box_{N^s}(B) = -\frac{1}{\|\hat{N}^s\|^{2}} \left( f(B) \otimes C_n^{i,j-1} \hat{N}^s C_n^{i,j-1} + C_n^{i,j-1} \hat{N}^s C_n^{i,j-1} \otimes f(B) \right) - \frac{1}{\|\hat{N}^s\|^{2}} \left( \hat{N}^s : B C_n^{i,j-1} \hat{N}^s C_n^{i,j-1} \otimes C_n^{i,j-1} \hat{N}^s C_n^{i,j-1} + (\hat{N}^s : B) C_n^{i,j-1} \square C_n^{i,j-1} \right).
\]
(115)

Here, \(B\) is defined as

\[
f(B) = B - \frac{1}{\|\hat{N}^s\|^{2}} (\hat{N}^s : B) C_n^{i,j-1} \hat{N}^s C_n^{i,j-1}.
\]
(116)

Similarly, we define the derivative of \(\hat{P}_1\) via

\[
d(\hat{P}_1^T : B) = \Box_1(B) : d\hat{C}_i + \hat{P}_1^T : d\hat{B},
\]
(117)

with

\[
\Box_1(B) = -\frac{1}{3} \hat{C}_i^{i,j-1} \left( \left( B - \frac{1}{3} \hat{C}_i : B \right) \hat{C}_i^{i,j-1} \right) \otimes \hat{C}_i^{i,j-1} + \hat{C}_i^{i,j-1} \otimes \left( B - \frac{1}{3} \hat{C}_i : B \right) \hat{C}_i^{i,j-1} \right) \left( \hat{C}_i^{i,j-1} \square \hat{C}_i^{i,j-1} - \frac{1}{3} \hat{C}_i^{i,j-1} \otimes \hat{C}_i^{i,j-1} \right).
\]
(118)

Subsequently, we find for the derivatives of \(\pi_\Delta:\)

\[
\frac{\partial^2 \pi_\Delta}{\partial \Delta \xi^2} = \frac{\partial}{\partial \Delta \xi} \frac{\partial}{\partial \Delta \xi} \left( \Box_1 \left( \frac{\partial (\psi_e + \psi_h + \psi_p)}{\partial C_i^j} \right) + \hat{P}_1^T : \frac{\partial^2 (\psi_e + \psi_h + \psi_p)}{\partial C_i^j^2} : \hat{P}_1 + \frac{\partial^2 \psi_f}{\partial C_i^j^2} : \frac{\partial}{\partial \Delta \xi} \right).
\]
\[ + 2 \frac{\partial \mathbf{C}_i}{\partial \Delta \xi} : \mathbf{P}_i^\top : \frac{\partial^2 (\psi_e + \psi_h + \psi_p) + \partial (\psi_e + \psi_h + \psi_p)}{\partial \mathbf{C}_i} + \frac{\partial^2 \mathbf{C}_i}{\partial \Delta \xi^2} + \sigma^{\text{reo}} \frac{\partial^2 \Delta \alpha}{\partial \Delta \xi^2} \]
\[ + \frac{\partial^2 (\psi_e + \psi_h + \psi_p)}{\partial \Delta \xi^2} + H(\text{sgn} (\Delta \xi)), \]
\[ \frac{\partial^2 \pi_\Delta}{\partial \Delta \gamma^2} = 2 \mathbf{N}^s : \left( D_i \left( \frac{\partial (\psi_e + \psi_h + \psi_p)}{\partial \mathbf{C}_i} \right) + \mathbf{P}_i^\top : \frac{\partial^2 (\psi_e + \psi_h + \psi_p)}{\partial \mathbf{C}_i^2} : \mathbf{P}_i + \frac{\partial^2 \psi_i}{\partial \mathbf{C}_i^2} \right) : 2 \mathbf{N}^s \]
\[ + \sigma^{\text{reo}} \frac{\partial^2 \Delta \alpha}{\partial \Delta \gamma^2}, \]
\[ \frac{\partial^2 \pi_\Delta}{\partial \Delta \xi \partial \Delta \gamma} = 2 \Delta \gamma \left( \mathbf{P}_i^\top : \frac{\partial^2 (\psi_e + \psi_h + \psi_p)}{\partial \mathbf{C}_i^2} : \mathbf{P}_i + \frac{\partial^2 \psi_i}{\partial \mathbf{C}_i^2} \right) : \frac{\partial \mathbf{C}_i}{\partial \Delta \xi} + \sigma^{\text{reo}} \frac{\partial^2 \Delta \alpha}{\partial \Delta \xi \partial \Delta \gamma} \]
\[ + 2 \mathbf{N}^s : \mathbf{P}_i^\top : \frac{\partial^2 (\psi_e + \psi_h + \psi_p)}{\partial \mathbf{C}_i^2} \]
\[ \frac{\partial^2 \pi_\Delta}{\partial \Delta \gamma \partial \mathbf{N}^s} = \sigma^{\text{reo}} \frac{\partial^2 \Delta \alpha}{\partial \Delta \gamma \partial \mathbf{N}^s} + \mathbf{P}_i^\top : \left( 2 \Delta \gamma \left( \mathbf{P}_i^\top : \frac{\partial^2 (\psi_e + \psi_h + \psi_p)}{\partial \mathbf{C}_i^2} : \mathbf{P}_i + \frac{\partial^2 \psi_i}{\partial \mathbf{C}_i^2} \right) : \frac{\partial \mathbf{C}_i}{\partial \Delta \gamma} \right) : 2 \mathbf{N}^s \]
\[ + 2 \left( \mathbf{P}_i^\top : \frac{\partial (\psi_e + \psi_h + \psi_p)}{\partial \mathbf{C}_i} + \frac{\partial \psi_i}{\partial \mathbf{C}_i} \right). \]

Here, for the occurring derivatives we find:
\[ \frac{\partial^2 \psi_e}{\partial \Delta \xi^2} = \frac{1}{4} \frac{\partial^2 \lambda}{\partial \Delta \xi^2} \left( \text{det}(b^2) - 1 - 2 \ln J^2 \right) + \frac{1}{2} \frac{\partial^2 \mu}{\partial \Delta \xi^2} \left( \text{tr}(b^2) - 3 - 2 \ln J^2 \right), \]
\[ \frac{\partial^2 \lambda}{\partial \Delta \xi^2} = 2 \left( \frac{1 - \xi}{\lambda_A} + \frac{\xi}{\lambda_M} \right)^3 \left( \frac{-1}{\lambda_A} + \frac{1}{\lambda_M} \right)^2, \]
\[ \frac{\partial^2 \mu}{\partial \Delta \xi^2} = 2 \left( \frac{1 - \xi}{\mu_A} + \frac{\xi}{\mu_M} \right)^3 \left( \frac{-1}{\mu_A} + \frac{1}{\mu_M} \right)^2, \]
\[ \frac{\partial^2 \psi_e}{\partial \mathbf{C}_i^2} = \mathbf{E}^e_\xi : \frac{\partial^2 \psi_e}{\partial \mathbf{b}^2} : \mathbf{E}^e_\xi^\top - C^{i-1} s \frac{\partial \psi_e}{\partial \mathbf{C}_i} - \frac{\partial \psi_e}{\partial \mathbf{C}_i} C^{i-1}, \]
\[ \mathbf{E}^e_\xi = (C^{i-1} F^T) \frac{\lambda}{\Delta} \left( FC^{i-1} \right), \]
\[ \frac{\partial^2 \psi_e}{\partial \mathbf{b}^2} = \frac{\lambda(\xi)}{4} e^2 b^{i-1} \otimes b^{i-1} - \left( \frac{\lambda(\xi)}{4} (J^{i-2} - 1) - \frac{\mu}{2} \right) b^{i-1} \otimes b^{i-1}, \]
A thermomechanical finite strain shape memory alloy model

\[ \frac{\partial^2 \psi_e}{\partial \xi \partial C^i} = -\bar{\varepsilon}_i^e : \frac{\partial^2 \psi_e}{\partial \xi \partial b^e}, \]  

(131)

\[ \frac{\partial^2 \psi_e}{\partial \xi \partial b^e} = \frac{1}{4} \frac{\partial \mu}{\partial \xi} \left( \left( J^e - \frac{1}{J^e} \right) J^e b^{e-1} \right) + \frac{1}{2} \frac{\partial \mu}{\partial \xi} (I - b^{e-1}), \]  

(132)

\[ \frac{\partial^2 C_i^j}{\partial \Delta \xi^2} = \frac{H_\perp(\xi)}{\Delta \xi^2} C_i^j \ln^2(C_i^j), \]  

(133)

\[ \frac{\partial^2 \psi_r}{\partial C^i} = \left( \frac{\partial^2 \psi_r}{\partial \tilde{\imath}^i} + \frac{\partial \psi_r}{\partial \tilde{\imath}^i} \right) \tilde{\imath}^i \hat{\imath}^i \otimes \hat{\imath}^i - \frac{\partial \psi_r}{\partial \tilde{\imath}^i} \tilde{\imath}^i \hat{\imath}^i \otimes \hat{\imath}^i, \]  

(134)

\[ \frac{\partial^2 \psi_r}{\partial N^i} = A \left( C_i^{n-1} \otimes C_i^{n-1} \left( 1 - \frac{1}{N^i} \right) + \frac{1}{N^i} C_i^{n-1} N^i C_i^{n-1} \otimes C_i^{n-1} N^i C_i^{n-1} \right), \]  

(135)

\[ \frac{\partial^2 \psi_h}{\partial \xi^2} = 2 \frac{\partial^2 \psi_h}{\partial \tilde{\imath}^i} \frac{\partial \tilde{\imath}^i}{\partial \tilde{\imath}^j} \frac{\partial \tilde{\imath}^j}{\partial \tilde{\imath}^i} + \frac{\partial^2 \psi_h}{\partial \tilde{\imath}^i} \frac{\partial \tilde{\imath}^i}{\partial \tilde{\imath}^j} \frac{\partial \tilde{\imath}^j}{\partial \tilde{\imath}^i}, \]  

(136)

\[ \frac{\partial^2 \psi_h}{\partial \tilde{\imath}^i} \frac{\partial \tilde{\imath}^i}{\partial \tilde{\imath}^j} \frac{\partial \tilde{\imath}^j}{\partial \tilde{\imath}^i} \left( \begin{array}{c} \{ \tilde{E}^i \} \leq c, \\ \{ \tilde{E}^i \} > c, \end{array} \right) \]  

(137)

\[ \frac{\partial^2 \psi_h}{\partial \tilde{\imath}^i} \frac{\partial \tilde{\imath}^i}{\partial \tilde{\imath}^j} \frac{\partial \tilde{\imath}^j}{\partial \tilde{\imath}^i} = \left\{ \begin{array}{c} 2 \{ \tilde{E}^i \}, \\ 0, \end{array} \right. \]  

(138)

\[ \frac{\partial^2 \psi_h}{\partial \tilde{\imath}^i} \frac{\partial \tilde{\imath}^i}{\partial \tilde{\imath}^j} \frac{\partial \tilde{\imath}^j}{\partial \tilde{\imath}^i} = \frac{\partial^2 \tilde{E}^i}{\partial \tilde{\imath}^i} \frac{k E^i}{\tilde{\imath}^i}, \]  

(139)

\[ \frac{\partial^2 \psi_h}{\partial \tilde{\imath}^i} \frac{\partial \tilde{\imath}^i}{\partial \tilde{\imath}^j} \frac{\partial \tilde{\imath}^j}{\partial \tilde{\imath}^i} = \frac{1}{2} \left( \frac{\partial^2 \psi_h}{\partial \tilde{\imath}^i} \frac{\partial \tilde{\imath}^i}{\partial \tilde{\imath}^j} \frac{\partial \tilde{\imath}^j}{\partial \tilde{\imath}^i} - \frac{\partial^2 \psi_h}{\partial \tilde{\imath}^i} \frac{\partial \tilde{\imath}^i}{\partial \tilde{\imath}^j} \frac{\partial \tilde{\imath}^j}{\partial \tilde{\imath}^i} \right), \]  

(140)

\[ \frac{\partial^2 \psi_h}{\partial \tilde{\imath}^i} \frac{\partial \tilde{\imath}^i}{\partial \tilde{\imath}^j} \frac{\partial \tilde{\imath}^j}{\partial \tilde{\imath}^i} = \frac{1}{4} \left( \frac{\partial^2 \psi_h}{\partial \tilde{\imath}^i} \frac{\partial \tilde{\imath}^i}{\partial \tilde{\imath}^j} \frac{\partial \tilde{\imath}^j}{\partial \tilde{\imath}^i} + \frac{\partial^2 \psi_h}{\partial \tilde{\imath}^i} \frac{\partial \tilde{\imath}^i}{\partial \tilde{\imath}^j} \frac{\partial \tilde{\imath}^j}{\partial \tilde{\imath}^i} \right), \]  

(141)

\[ \frac{\partial^2 \tilde{E}^i}{\partial \tilde{\imath}^i} = \sqrt{\frac{3}{k}} \frac{1}{\| \tilde{E}^i \|} \left( \| \tilde{E}^i \| \otimes \tilde{E}^i \right), \]  

(142)

\[ \frac{\partial^2 \psi_p}{\partial \xi^2} = \left\{ \begin{array}{c} \frac{1}{\sqrt{c + (\{ \tilde{E}^i \})^2}}, \\ \frac{1}{\sqrt{c + (\{ \tilde{E}^i \})^2}}, \end{array} \right. \]  

(143)

\[ \frac{\partial^2 \psi_p}{\partial \xi^2} = \frac{1}{4} \left( \frac{\partial^2 \psi_p}{\partial \tilde{\imath}^i} \frac{\partial \tilde{\imath}^i}{\partial \tilde{\imath}^j} \frac{\partial \tilde{\imath}^j}{\partial \tilde{\imath}^i} + \frac{\partial^2 \psi_p}{\partial \tilde{\imath}^i} \frac{\partial \tilde{\imath}^i}{\partial \tilde{\imath}^j} \frac{\partial \tilde{\imath}^j}{\partial \tilde{\imath}^i} \right), \]  

(144)
\[
\frac{\partial^2 \psi_p}{\partial E^i \partial \xi} = \frac{1}{2 \xi} \left( \frac{\partial^2 \psi_p}{\partial E^i} - \frac{\partial^2 \psi_p (E^i)}{\xi^2} - \frac{1}{\xi} \frac{\partial \psi_p}{\partial E^i} \right) \frac{\partial (E^i)}{\partial E^i},
\]

(145)

\[
\frac{\partial^2 \psi_p}{\partial \xi \partial (E^i)} = \begin{cases} 
H_p \left( \frac{\xi}{\sqrt{1+c}} \right)^2 & \xi \geq \xi_0 - c(E^i), \\
H_p \xi^2 & \text{else},
\end{cases}
\]

(146)

\[
\frac{\partial^2 \psi_p}{\partial \xi \partial (E^i)} = \begin{cases} 
H_p \left( \frac{(E^i) - c}{\sqrt{1+c}} + l_p \frac{1}{\sqrt{1+c}} \right) & \xi \geq \xi_0 - c(E^i), \\
2H_p \xi (E^i) & \text{else},
\end{cases}
\]

(147)

\[
\frac{\partial^2 \Delta \alpha}{\partial \Delta \xi^2} = \sqrt{\frac{2}{3}} \left( \frac{\partial C^i}{\partial \xi} : \left( P_i^T - 0^\circ \right) \right) \left( \frac{1}{2} \frac{\partial \tilde{M}_C}{\partial C^i} : \left( P_i - 0^\circ \right) + \frac{1}{2} \frac{\partial \tilde{M}_C}{\partial C^i} : \left( P_i - 0^\circ \right) \right) \frac{\partial C^i}{\partial \xi}
\]

(148)

\[
\frac{\partial^2 \Delta \alpha}{\partial \Delta \gamma^2} = \sqrt{\frac{2}{3}} 2N^s : \left( \frac{1}{2} \frac{\partial \tilde{M}_C}{\partial C^i} \right),
\]

(149)

\[
\frac{\partial^2 \Delta \alpha}{\partial \Delta \xi \partial \Delta \gamma} = \sqrt{\frac{2}{3}} \left( \frac{1}{2} \frac{\partial \tilde{M}_C}{\partial C^i} : \left( P_i^T - 0^\circ \right) \right) + \frac{1}{2} \frac{\partial \tilde{M}_C}{\partial C^i} : \left( P_i - 0^\circ \right) \frac{\partial C^i}{\partial \xi}
\]

(150)

\[
\frac{\partial^2 \Delta \alpha}{\partial \Delta \xi \partial \Delta \gamma} = \sqrt{\frac{2}{3}} \left( \frac{1}{2} \frac{\partial \tilde{M}_C}{\partial C^i} : \left( P_i^T - 0^\circ \right) \right) + \frac{1}{2} \frac{\partial \tilde{M}_C}{\partial C^i} : \left( P_i - 0^\circ \right) \frac{\partial C^i}{\partial \xi}
\]

(151)

\[
\frac{\partial^2 \Delta \alpha}{\partial \Delta \xi \partial \Delta \gamma} = \sqrt{\frac{2}{3}} 2\Delta \xi \left( \frac{1}{2} \frac{\partial \tilde{M}_C}{\partial C^i} : \left( P_i^T - 0^\circ \right) \right) \frac{\partial C^i}{\partial \xi}
\]

(152)

\[
\frac{\partial^2 \Delta \alpha}{\partial \Delta \gamma \partial \Delta N} = \sqrt{\frac{2}{3}} \left( \frac{1}{2} \frac{\partial \tilde{M}_C}{\partial C^i} : \left( P_i^T - 0^\circ \right) + \frac{1}{2} \frac{\partial \tilde{M}_C}{\partial C^i} : \left( P_i - 0^\circ \right) \right) \frac{\partial C^i}{\partial \xi}
\]

(153)

\[
\frac{\partial \tilde{M}_C}{\partial C^i} = \frac{1}{\sqrt{N_i}} \left( C_n^{-1} \otimes C_n^{-1} \right) \left( \tilde{M}_C \otimes \tilde{M}_C \right)
\]

(154)
**E Consistent tangent**

When using the finite element method to solve boundary value problems, we make use of the consistent tangent operator. To derive it with respect to \( \frac{\partial}{\partial d} \) from the potential, we start with the virtual work of the internal forces

\[
\int_{V_0} \tau : d\delta \, dV. \tag{155}
\]

Here, \( d\delta \) is defined as

\[
d\delta = \text{sym} \left( \text{Grad} (\delta u) F^{-1} \right). \tag{156}
\]

Now, we need to calculate the differential of virtual work of the internal forces:

\[
\int_{V_0} d(\tau : d\delta) \, dV, \tag{157}
\]

where through using the symmetry of \( \tau \) we can simplify to

\[
d(\tau : d\delta) = d(\tau : I_{\delta}) = d(FS) : \delta F. \tag{158}
\]

Now, the differential reads

\[
d(FS) : \delta F = (dFF^{-1} FSF^T (\delta F F^{-1})^T) : I + (F \delta SF^T) : I_{\delta} = (I_{\delta} I_{\delta}^T) : I + (e^a : d_d) : d\delta, \tag{159}
\]

where the first term, i.e., \( I_d \tau I_{\delta}^T \) is related to the so-called geometric tangent and the second term arises from the incremental material stiffness. Further, \( I_d \) and \( d_d \) are defined in analogy to \( I_{\delta} \) and \( d\delta \). Here, we neglected, in a first step, the influence of perturbations \( d\theta \) of the temperature (for the full linearization, see below). Now, we can split \( e^a \) into the elastic and inelastic parts:

\[
e^a : d_d = \left( F \Box F^T : \frac{\partial S}{\partial E} : F^T \Box F \right) : d_d + \frac{\partial (FSF^T)}{\partial A} : \frac{\partial A}{\partial E} : dE
\]

\[
= e^e : d_d + \frac{\partial \tau}{\partial E} : \frac{\partial A}{\partial E} : F^T d_d F, \tag{160}
\]

where (with a small abuse of notation) \( A \) denotes the vector of active variables and may contain \( \xi \), \( \Delta \gamma \) and \( \tilde{N}^s \). Additionally, the elastic tangent \( e^e \) is given by

\[
e^e = (\lambda(\xi) \text{det}(b^e)) I \otimes I + (2\mu(\xi) - \lambda(\xi(\text{det}(b^e) - 1))))b^e. \tag{161}
\]

Next, we need to obtain \( \frac{\partial A}{\partial E} \). Therefore, we take a look at the local Newton scheme, where

\[
\frac{\partial \pi_{\Delta}}{\partial A} = 0 \tag{162}
\]

holds in the converged state. Now, calculating the differential of Eq. (162), we obtain

\[
0 = d \left( \frac{\partial \pi_{\Delta}}{\partial A} \right)
\]

\[
= \frac{\partial^2 \pi_{\Delta}}{\partial A^2} dA + \frac{\partial^2 \pi_{\Delta}}{\partial A \partial E} : dE = 0
\]

\[
= \frac{\partial^2 \pi_{\Delta}}{\partial A^2} dA + \left( \frac{\partial S}{\partial A} \right)^T : dE = 0. \tag{163}
\]
Rearranging the result from Eq. (163) we obtain
\[
\frac{dA}{\partial E} = -\left(\frac{\partial^2 \pi}{\partial A^2}\right)^{-1} \frac{\partial \pi}{\partial A} : dE
\]
\[
= -\left(\frac{\partial^2 \pi}{\partial A^2}\right)^{-1} \frac{\partial S}{\partial A} : F^T \mathbb{S} F : dA
\]
\[
= -\left(\frac{\partial^2 \pi}{\partial A^2}\right)^{-1} \frac{\partial \tau}{\partial A} : dA,
\]
which we can insert into Eq. (160) to finally obtain the algorithmic tangent:
\[
e^a = e^c - \frac{\partial \tau}{\partial A} \left(\frac{\partial^2 \pi}{\partial A^2}\right)^{-1} \left(\frac{\partial \tau}{\partial A}\right)^T.
\]
In here, the term \(\frac{\partial \tau}{\partial A}\) is computed as
\[
\frac{\partial \tau}{\partial A} = \begin{pmatrix}
\frac{\partial \tau}{\partial \xi} \\
\frac{\partial \tau}{\partial \Delta y} \\
\frac{\partial \tau}{\partial \Delta z}
\end{pmatrix}
\]
with the additional derivatives
\[
\frac{\partial \tau}{\partial \xi} = \frac{\partial C^i}{\partial \xi} \frac{\partial^2 \psi_e}{\partial \xi^2} + 2b^e \frac{\partial^2 \psi_e}{\partial b^e \partial \xi},
\]
\[
\frac{\partial \tau}{\partial \Delta y} = 2 \frac{\partial \tau}{\partial C^i} : N^s,
\]
\[
\frac{\partial \tau}{\partial \Delta z} = 2 \Delta y \frac{\partial \tau}{\partial C^i} : P_{N^s}.
\]
\[
\frac{\partial \tau}{\partial C^i} = -2 \left(I \mathbb{S} \frac{\partial \psi_e}{\partial b^e} + \frac{\partial \psi_e}{\partial b^e} \mathbb{S} I + \left(I \mathbb{S} b^e + b^e \mathbb{S} I\right) : \frac{\partial^2 \psi_e}{\partial b^e^2}\right),
\]
Additionally, the algorithmic consistent tangent with respect to \(\theta\) as well as the mixed contributions can be computed through linearizing the weak form in Eq. (74):
\[
\int_{V_0} d\left(\frac{\partial \pi}{\partial \theta} \delta \theta - \frac{\kappa \Delta \theta}{\theta_n} \text{Grad}(\delta \theta) \cdot \text{Grad}(\delta \theta)\right) dV
\]
\[
= \int_{V_0} \delta \theta \left(\left(\frac{\partial^2 \pi}{\partial A^2} \delta A \delta \theta - \frac{\partial \pi}{\partial \theta} \delta \theta \right) - \frac{\partial \pi}{\partial \theta} : dA \right) - \text{Grad}(\delta \theta) \left(\frac{\kappa \Delta \theta}{\theta_n} \text{Grad}(\delta \theta)\right) dV. \]
Here, the derivative \(\frac{\partial \tau}{\partial \theta}\) is given by
\[
\frac{\partial \tau}{\partial \theta} = \frac{\partial \tau}{\partial A} \left(\frac{\partial^2 \pi}{\partial A^2}\right)^{-1} \frac{\partial \pi}{\partial A \partial \theta}.
\]
with
\[
\frac{\partial^2 \pi_\Delta}{\partial A \partial \theta} = \left( \Delta \sigma^{AM} + \sigma^{reo} \frac{\partial \Delta \alpha}{\theta_n} + \frac{Q}{\theta_n} + \frac{H \Delta \xi}{\theta_n} \land \frac{\sigma^{reo} \partial \Delta \alpha}{\theta_n} \land \frac{\sigma^{reo} (\partial \Delta \alpha / \partial N_S)}{\theta_n} \right)^T.
\]  
(173)

Now, using standard formulations for the element ansatz functions \( N \) and its gradients \( B \), one can assemble the element stiffness matrix (see, e.g., Wriggers [51]).

References

1. Arghavani, J., Auricchio, F., Naghdabadi, R.: A finite strain kinematic hardening constitutive model based on Hencky strain: general framework, solution algorithm and application to shape memory alloys. Int. J. Plast. 27(6), 940–961 (2011). https://doi.org/10.1016/j.ijplas.2010.10.006
2. Arivanandhan, G., Li, Z., Curtis, S., Velvaluri, P., Quandt, E., Kohl, M.: Temperature homogenization of co-integrated shape memory-silicon bimorph actuators. Proceedings (2020). https://doi.org/10.3390/leCAT2020-08501
3. Auricchio, F., Reali, A., Stefanelli, U.: A three-dimensional model describing stress-induced solid phase transformation with permanent inelasticity. Int. J. Plast. 23(2), 207–226 (2007). https://doi.org/10.1016/j.ijplas.2006.02.012
4. Auricchio, F., Bonetti, E., Scalet, G., Ubertini, F.: Theoretical and numerical modeling of shape memory alloys accounting for multiple phase transformations and martensite reorientation. Int. J. Plast. 59, 30–54 (2014). https://doi.org/10.1016/j.ijplas.2014.03.008
5. Bellini, C., Berto, F., Di Cocco, I., Iacoviello, F.: A cyclic integrated microstructural-mechanical model for a shape memory alloy. Int. J. Fatigue 153, 10673 (2021). https://doi.org/10.1016/j.ijfatigue.2021.10673
6. Bernardini, D., Pence, T.J.: Models for one-variant shape memory materials based on dissipation function. Int. J. Non-Linear Mech. 37(8), 1299–1317 (2002). https://doi.org/10.1016/S0020-7462(02)00020-3
7. Brepolis, T., Wulffinghoff, S., Reese, S.: Gradient-extended two-surface damage-plasticity: micromorphic formulation and numerical aspects. Int. J. Plast. 97, 64–106 (2017). https://doi.org/10.1016/j.ijplas.2017.05.010
8. Cisse, C., Zaki, W., Ben Zineb, T.: A review of constitutive models and modeling techniques for shape memory alloys. Int. J. Plast. 76, 244–284 (2016). https://doi.org/10.1016/j.ijplas.2015.08.006
9. Cisse, C., Zaki, W., Ben Zineb, T.: A review of constitutive models and modeling techniques for shape memory alloys. Int. J. Plast. 76, 244–284 (2016). https://doi.org/10.1016/j.ijplas.2015.08.006
10. Eckart, C.: The thermodynamics of irreversible processes. IV. The theory of elasticity and anelasticity. Phys. Rev. 73(4), 373–373 (1948). https://doi.org/10.1103/PhysRev.73.373
11. Eichenfield, M., Camacho, R., Chan, J., Vahala, K.J., Painter, O.: A picogram-and nanometre-scale photonic-crystal optomechanical cavity. Nature 459(7246), 550–555 (2009). https://doi.org/10.1038/nature08061
12. Fischer, A.: A special newton-type optimization method. Optimization 24(3–4), 269–284 (1992). https://doi.org/10.1080/02331939208843795
13. Flory, P.J., Volkenstein, M.: Statistical mechanics of chain molecules. Biopolymers 8(5), 699–700 (1969). https://doi.org/10.1002/bip.1969.380080514
14. Frost, M., Benešová, B., Sedlák, P.: A microscopically motivated constitutive model for shape memory alloys: formulation, analysis and computations. Math. Mech. Solids 21(3), 358–382 (2016). https://doi.org/10.1177/1081286514522474
15. Gill, K., Sehitoglu, H., Chumlyakov, Y.I., Kireeva, I.V.: Tension-compression asymmetry of the stress-strain response in aged single crystal and polycrystalline niti. Acta Mater. 47(4), 3 (1999). https://doi.org/10.1016/S1359-6454(98)00432-7
16. Goddard, J., Shetty, T.K.: High-temperature shape memory microactuators. Springer, Berlin (2004)
17. Godfrey, S., Mielke, A., Hall, G.J.: The free energy of mixing for n-variant martensitic phase transformations using quasi-convex analysis. J. Mech. Phys. Solids 51(4), 1–XXVI (2003). https://doi.org/10.1016/S0022-5096(03)00015-2
18. Halphen, B., Nguyen, Q.S.: Sur les matériaux standard généralisés. (1975)
19. Han, S., Seok, T.J., Quack, N., Yoo, B.-W., Wu, M.C.: Large-scale silicon photon switches with movable directional couplers. Optica 2(4), 370–375 (2015). https://doi.org/10.1364/OPTICA.2.000370
20. Hartl, D.J., Chatzigeorgiou, G., Lagoudas, D.C.: Three-dimensional modeling and numerical analysis of rate-dependent irreversible deformation in shape memory alloys. Int. J. Plast. 26(10), 1485–1507 (2010). https://doi.org/10.1016/j.ijplas.2010.01.002
21. Hoffmann, M., Bezzaoui, H., Voges, E.: Micromechanical cantilever resonators with integrated optical interrogation. Sens. Actuators A 44(1), 71–75 (1994). https://doi.org/10.1016/0924-4247(94)00076-4
22. Hurtado, D., Stainier, L., Ortiz, M.: The special-linear update: an application of differential manifold theory to the update of isochoric plasticity flow rules. Int. J. Numer. Methods Eng. 97(4), 298–312 (2014). https://doi.org/10.1002/nme.4600
23. Iwanicki, B., Shape Memory Microactuators. Springer, Berlin (2004)
24. Kohl, M., Kreyvet, B., Just, E., Sma microgrpper system. Sens. Actuators A Phys. 97–98, 646–652. https://doi.org/10.1016/S0924-4247(01)00803-2
25. Kröner, E.: Allgemeine Kontinuumstheorie der Versetzungen und Eigenspannungen. Arch. Ration. Mech. Anal. 4(1), 273 (1959). https://doi.org/10.1007/BF02813938
26. Lagoudas, D.C., Entchev, P.B., Popov, P., Fatoor, E., Brinson, L.C., Gao, X.: Shape memory alloys, part II: modeling of polycrystals. Mech. Mater. 38(5), 430–462 (2006). https://doi.org/10.1016/j.mechmat.2005.08.003
27. Lee, E.H.: Elastic-plastic deformation at finite strains. J. Appl. Mech. 36(1), 1–6 (1969). https://doi.org/10.1115/1.365480
28. Lester, B.T., Baxevanis, T., Chemisky, Y., Lagoudas, D.C.: Review and perspectives: shape memory alloy composite systems. Acta Mech. 226(12), 3907–3960 (2015). https://doi.org/10.1007/s00707-015-1433-0
29. Lexcellent, C., Boubakar, M., Bouvet, C., Calloch, S.: About modelling the shape memory alloy behaviour based on the phase transformation surface identification under proportional loading and anisothermal conditions. Int. J. Solids Struct. 43(3), 613–626 (2006). https://doi.org/10.1016/j.jss.2005.07.004
29. Mirzaeifar, R., DesRoches, R., Yavari, A., Gall, K.: A micromechanical analysis of the coupled thermomechanical superelastic response of textured and untextured polycrystalline niti shape memory alloys. Acta Mater. 61(12), 4542–4558 (2013). https://doi.org/10.1016/j.actamat.2013.04.023

30. Moumni, Z., Zaki, W., Nguyen, Q.S.: Theoretical and numerical modeling of solid-solid phase change: application to the description of the thermomechanical behavior of shape memory alloys. Int. J. Plast. 24(4), 614–645 (2008). https://doi.org/10.1016/j.jplas.2007.07.007

31. Müller, C., Bruhns, O.: A thermodynamic finite-strain model for pseudoelastic shape memory alloys. Int. J. Plast. 22(9), 1658–1682 (2006). https://doi.org/10.1016/j.jplas.2006.02.010

32. Nemat-Nasser, S.: On finite deformation elasto-plasticity. Int. J. Solids Struct. 18(10), 857–872 (1982). https://doi.org/10.1016/0020-7683(82)90070-1

33. Otsuka, K., Ren, X.: Physical metallurgy of ti-ni-based shape memory alloys. Prog. Mater Sci. 50(5), 511–678 (2005). https://doi.org/10.1016/j.pmatsci.2004.10.001

34. Panico, M., Brinson, L.: A three-dimensional phenomenological model for martensite reorientation in shape memory alloys. J. Mech. Phys. Solids 55(11), 2491–2511 (2007). https://doi.org/10.1016/j.jmps.2007.03.010

35. Patoor, E., Lagoudas, D.C., Entchev, P.B., Brinson, L.C., Gao, X.: Shape memory alloys, part I: general properties and modeling of single crystals. Mech. Mater. 38(5), 391–429 (2006). https://doi.org/10.1016/j.mechmat.2005.05.027

36. Potapov, P., Shelyakov, A., Gulyaev, A., Svistunov, E., Matveeva, N., Hodgson, D.: Effect of hf on the structure of ni-ti martensitic alloys. Mater. Lett. 32(4), 247–250 (1997). https://doi.org/10.1016/S0167-577X(97)00037-2

37. Qiuwai, M., Lagoudas, D.: On thermomechanics and transformation surfaces of polycrystalline niti shape memory material. Int. J. Plast. 16(10), 1309–1343 (2000). https://doi.org/10.1016/S0749-6419(00)00012-7

38. Reese, S., Christ, D.: Finite deformation pseudo-elasticity of shape memory alloys—constitutive modelling and finite element implementation. Int. J. Plast. 24(3), 455–482 (2008). https://doi.org/10.1016/j.jplas.2007.05.005

39. Rezaee-Hajidehi, M., Tůma, K., Stupkiewicz, S.: Gradient-enhanced thermomechanical 3d model for simulation of transformation patterns in pseudoelastic shape memory alloys. Int. J. Plast. 128, 102589 (2020). https://doi.org/10.1016/j.jplas.2019.08.014

40. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton, 2015. ISBN 9781400873173. https://doi.org/10.1515/9781400807317

41. Scalet, G., Karakalas, A., Xu, L., Dagoudas, D.: Finite strain constitutive modelling of shape memory alloys considering partial phase transformation with transformation-induced plasticity. Shape Memory Superelast. 7(2), 206–221 (2021). https://doi.org/10.1007/s40830-021-00350-5

42. Sedlák, P., Frost, M., Benešová, B., Ben Zineb, T., Šittner, P.: Thermomechanical model for niti-based shape memory alloys including r-phase and material anisotropy under multi-axial loadings. Int. J. Plast. 39, 132–151 (2012). https://doi.org/10.1016/j.ijplas.2012.06.008

43. Seelecke, S., Müller, I.: Shape memory alloy actuators in smart structures: modeling and simulation. Appl. Mech. Rev. 57(1), 23–46 (2004). https://doi.org/10.1115/1.1584064

44. Sielenkämper, M., Dittmann, J., Wulfinghoff, S.: Numerical strategies for variational updates in large strain inelasticity with incompressibility constraint. Int. J. Numer. Methods Eng. https://doi.org/10.1002/nme.6855

45. Tamo, J., Taylor, R., Pieter, K.: Variational and projection methods for the volume constraint in finite deformation elastoplasticity. Comput. Methods Appl. Mech. Eng. 51(1), 177–208 (1985). https://doi.org/10.1016/0045-7825(85)90033-7

46. Taylor, R.L.: FEAP—finite element analysis program. http://projects.ce.berkeley.edu/feap/ (2017)

47. Wulfinghoff, S., Böhlke, T.: Equivalent plastic strain gradient crystal plasticity—enhanced power law subroutine. GAMM-Mitteilungen 36(2), 134–148 (2013). https://doi.org/10.1002/gamm.201310008

48. Xu, L., Solomou, A., Baxevanis, T., Lagoudas, D.: Finite strain constitutive modeling for shape memory alloys considering transformation-induced plasticity and two-way shape memory effect. Int. J. Solids Struct. 221, 42–59 (2021). https://doi.org/10.1016/j.ijsolstr.2020.03.009

49. Yang, Q., Stainier, L., Ortiz, M.: A variational formulation of the coupled thermo-mechanical boundary-value problem for general dissipative solids. J. Mech. Phys. Solids 54(2), 401–424 (2006). https://doi.org/10.1016/j.jmps.2005.08.010

50. Yoo, C., Kang, G., Sun, Q., Fang, D.: Modeling the martensite reorientation and resulting zero/negative thermal expansion of shape memory alloys. J. Mech. Phys. Solids 127, 295–331 (2019). https://doi.org/10.1016/j.jmps.2019.03.015

51. Zaki, W., Moumni, Z., Morin, C.: Modeling tensile-compressive asymmetry for superelastic shape memory alloys. Mech. Adv. Mater. Struct. 18(7), 559–564 (2011). https://doi.org/10.1080/15376694.2011.605016

52. Zhang, M., Baxevanis, T.: An extended three-dimensional finite strain constitutive model for shape memory alloys. J. Appl. Mech. 88(11), 08 (2021). https://doi.org/10.1115/1.4051833.111010

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.