Analysis of the deformation paths and thermomechanical parameter identification of a shape memory alloy using digital image correlation over heterogeneous tests

Yves Chemisky a,*, Fodil Meraghni a, Nadine Bourgeois b, Stephen Cornell c, Rachid Echchorfa, Etienne Patoora a

a Arts et Métiers ParisTech Metz, LEM3 UMR CNRS 7239, Metz, France
b Université de Lorraine, LEM3 UMR CNRS 7239, Metz, France
c Department of Aerospace Engineering, Texas A&M University, 3409 TAMU College Station, TX 77843-3409, USA

A B S T R A C T

With the design of new devices with complex geometry and to take advantage of their large recoverable strains, shape memory alloys (SMA) components are increasingly subjected to multiaxial loadings. The development process of SMA devices requires the prediction of their thermomechanical response for which the calibration of the material parameters for the numerical model is an important step. In this work, the parameters of a phenomenological model are extracted from tests performed on specimens with non-uniform geometry, which induce heterogeneous strain fields carried out on specimens with the same thermomechanical loading history. The digital image correlation technique is employed to measure the strain fields on the surface of the specimen and to analyze the strain paths of chosen points. Finite element analysis enables the computation of numerical strain fields using a thermodynamical constitutive model for shape memory alloys previously implemented in a finite element code. The strain fields computed numerically are compared with experimental ones obtained by DIC to find the model parameters which best match experimental measurements using a newly developed parallelized mixed genetic/gradient-based optimization algorithm. These numerical simulations are carried out in parallel using a supercomputer to reduce the time necessary to identify the set of model parameters. The major features of this new algorithm are its ability to identify the material parameters which describe the thermomechanical behavior of shape memory alloys from full-field measurements for various loading conditions (different temperatures, multiaxial behavior, heterogeneous test configurations). It is demonstrated that model parameters for the simulation of SMA structures are thus obtained based on a reduced number of heterogeneous tests at different temperatures.

1. Introduction

Shape memory alloys (SMA) are utilized in a wide range of applications and have been incorporated in system with increased complexity. Their unique ability to recover substantial deformation when subjected to particular thermomechanical load made them very attractive and suitable for actuator devices and to replace complex assemblies. The development of new design solution, combining three-dimensional models along with an integration in finite element analysis (FEA) packages, has contributed to the appearance of shape memory alloys parts with complex geometrical shape.

It has been recently noticed that the development of modern computer design and analysis tools based on three-dimensional constitutive models is essential for the development process of SMA applications [36]. These models should be however calibrated from carefully obtained experimental data and advanced material characterization.

Indeed, new SMA structures are subjected to heterogeneous loading conditions, where the material locally undergoes non-proportional loadings. Once provided with the calibrated material parameters, three-dimensional constitutive models based on thermodynamics of irreversible processes can predict accurately and efficiently the behavior of such multiaxial, non-proportional loadings. A review of such models for SMAs can be found in Patoor et al. [38] for the constitutive modeling of a single crystal and Lagoudas et al. [28] for the constitutive modeling of polycrystals. Pioneer models have been improved over the years and have been
implemented in FEA packages [40,37,27,11,43,1]. Recent constitutive models have also focused on specific features associated with the unique behavior of shape memory alloys.

The evolution of transformation induced plasticity has been further considered for conventional SMAs by Bo and Lagoudas [7] and Lagoudas and Entchev [26]. The coupling between phase transformation and plasticity has also been considered in the literature [19,50,24] and at high temperature to account for viscoplastic effects [21,10]. Recent models have also focused on specific loading-path dependent behavior such as tension–compression asymmetry and anisotropy [11,43]. The design of SMA structures and the optimization of their characteristics rely now on finite element analysis. This constitutes powerful three-dimensional design solutions for SMA components subjected to complex loading. The accuracy of numerical simulations relies on the model ability to accurately account for uniaxial and multiaxial non-proportional loadings. However, these solutions require more advanced characterization techniques than the uniaxial tension or thermal analyses recommended by the ASTM standards [14,15]. Indeed, SMAs structural computations require necessarily the correct estimation of the model parameters including those related to the loading-path dependent multiaxial behavior and the characteristics of a specific alloy. The methodology specified in the ASTM standards was specifically designed for Nickel–Titanium in the biomedical industry, and may not be suitable for the identification of a model that aims at performing the numerical simulation of SMA structures. Moreover, since the methodology described in the ASTM standards does not give information about the characterization of the complete set of parameters for these models, the development of a reliable method to identify the parameters for SMA constitutive models is an important step to perform reliable numerical simulations for the design of SMA structures, regardless of the choice of a constitutive model. A traditional method is to systematically perform uniaxial tests in several direction, shear, combined tension/shear on a series of experiments to extract the features that are loading-path dependent. Also, a uniaxial stress–temperature diagram is constructed along with the uniaxial stress/strain/temperature response, and the parameters related to phase transformation are extracted from these graphical representation [25,22,20,11]. The most common procedure is to manually obtain the phase transformation data using a tangent intersection method [46]. However, several issues are associated with this method, as no objective criteria are utilized to assess the efficiency and the robustness of the method. This can lead to significant error when estimating the model parameters of a SMA material. For homogeneous tests, a recent method has been developed to identify the material parameters of a SMA from a set of experiments, utilizing a gradient-based inverse approach [34]. A similar approach has been used to determine the parameters of an SMA model from experimental data [49]. However, experimental characterization requires testing of several different types of specimens (tensile dogbone specimen, double shear specimens, thin tubes) to obtain the whole set of material parameters required for finite element (FE) simulations if tension–compression asymmetry and/or anisotropy induced by processing conditions is considered. In the case of Ni–Ti shape Memory Alloys, the processing of these specimens strongly influences the behavior [41] and thus the extraction of a unique set of parameters that represents the material behavior is not possible. Moreover, localization effects may be observed in specimens of simple geometry subjected to homogeneous loading, but the SMA device will not necessarily exhibit such effects, depe nding on its geometry and loading conditions. Thus identification of material parameters on specimens as similar as possible to the final product shall be preferred. Moreover, the design process is accelerated if the characterization requires only a few specimens. The identification procedure which utilizes full-field kinematical fields is an appropriate method for the identification of the set of material parameters for the purpose of SMA devices design. Such set of material parameters may include parameters characteristic of loading-path dependent behavior.

These considerations have motivated the development of an identification method to obtain the model parameter of a SMA constitutive law based on complex, heterogeneous loading conditions. Note that such a methodology is also necessary to properly analyze complex tests (i.e. multiaxial tests on samples with complex geometry), to be able to detect additional features, e.g. anisotropy and tension–compression asymmetry.

The requirements for such an identification procedure are summarized below:

(i) The identification procedure should be adapted to complex (multiaxial, non-proportional) thermomechanical tests utilized to characterize the constitutive behavior of materials.
(ii) The procedure should be fast enough to be of a practical interest for the identification of multiple parameters based on a reduced number of tests.
(iii) This procedure should be compatible with experimental observations, such as reaction forces and full kinematic field, measured at the surface of the sample using digital image correlation (DIC).

According to the last requirement, it is clear that the identification procedure relies on two-dimensional measurements. The identification of three-dimensional models requires additional assumptions about the anisotropic nature of their behavior, to be able to compute properly the out-of-plane response of SMA components (according to the plane where the kinematical field is measured). Identification methods coupled with kinematical fields measurement have been extensively developed in the last decades. The most used methods are described in the useful review work of Avril et al. [2]. These identification methods are classified into five categories, that are (i) the constitutive equation gap method (CEGM); (ii) the virtual field method (VFM); (iii) the equilibrium gap method (EGM); (iv) the reciprocity gap method (RGM) and (v) the finite element model updating (FEMU) method. The advantages and drawbacks of each method are described extensively in the review of Avril et al. [2].

The FEMU approach, proposed by Kavanagh and Clough [23], is most intuitive and has a lot of flexibility with regard to the definition of the cost function. This method can therefore be applied in a general framework, even if full-field displacement data with sufficiently sharp spatial resolution is not directly provided, which is necessary for the EGM and VFM methods [2]. Furthermore, the FEMU approach has the advantage to perform at the same time a parametric study of the material parameters, utilizing the numerical simulations performed for the identification, assuming that the design space has been sufficiently explored. Following these considerations, in this work a FEMU method is developed and coupled with an innovative combined genetic/gradient-based optimization algorithm. Since finite element analyses are necessary to compare the experimental data with the numerical simulations, a specific attention is devoted to the parallelization of such approach in a supercomputer.

Various optimization algorithms have been used in identification methods, i.e. deterministic algorithms such as gradient-based Levenberg–Marquardt algorithm [31,33], real space evolutionary-inspired, genetic algorithms or Bayesian statistical approaches [5]. The Leven berg–Marquardt optimization algorithm has been often adopted for the determination of material parameters for metals [45,32,16,13]. A hybrid algorithm that used an evolutionary algorithm combined with a gradient-based algorithm has been utilized by Chaparro et al. [9] to determine the parameters of an elastic–plastic constitutive model including an anisotropic criterion [3] coupled with a non-linear kinematical hardening [30]. The evolutionary algorithm was used to determine a good initial guess values for the gradient-based approach and therefore
ensure that a local minimum has been found. In the present work, NiTi SMA parameters are identified on a heterogeneous configuration test using an evolutionary algorithm combined with a Levenberg–Marquardt algorithm. The latter is utilized to further optimize the best elements of each generation provided by the genetic algorithm.

The constitutive model utilized for the description of the SMA thermomechanical behavior is the model developed by Lagoudas et al. [27]. This model has been chosen for the following reasons: (i) the model includes several features, such as a dependence of the maximum transformation strain with the stress magnitude different forward and reverse slopes in the uniaxial stress–temperature diagram and transformation hardening. This will allow us to extend the present work, that is focused on the superelastic behavior of SMAs, to other behavior such as thermal actuation; (ii) the model is formulated and derived from a thermodynamic potential and (iii) under the approximation of multiaxial proportional cases, it is shown that the material parameters identified can be directly utilized to determine some of material parameters of other models such as [37,11], as shown by Stebner et al. [46].

Based on all the previously discussed points, the proposed work focuses on the development of a parallelized hybrid identification procedure for the parameters of the model of Lagoudas et al. [27] that has the following features:

1. It is adapted to the identification of material parameters directly on an SMA structure under very different solicitation (i.e. local thermomechanical paths).
2. It is based on the definition of a cost function by means of a square difference of the experimentally and numerically computed variables (strains and forces).
3. A parallelized optimization algorithm is utilized, which combine an evolutionary algorithm with the Levenberg–Marquardt algorithm.

The organization of this work is as follows: Section 2 presents the experimental procedure for the mechanical tests utilized to validate the identification procedure proposed. Several mechanical tests on samples with notches (inspired from the geometry utilized by [35]) have been performed to test the identification methodology. This geometry has been selected since it has been shown that uniform displacement conditions at its boundary lead to strongly heterogeneous strain fields. Such geometry is considered suitable to analyze the capabilities of an identification algorithm based on a FEMU approach using strain fields obtained by DIC technique. Section 3 presents the experimental results and the analysis of local strain paths. Based on the measured spatio-temporal strain distributions, several points of the tested heterogeneous specimen have been analyzed and discussed in terms of local strain-paths. Section 4 briefly reviews the model of Lagoudas et al. [27] and presents the material parameters to be identified in the case of superelasticity. Section 5 presents the parallelized identification procedure with the definition of the hybrid optimization algorithm. The validation of such identification based on heterogeneous tests is provided in Section 6, comparing the strain fields and the local strain evolution obtained numerically with the ones obtained using the DIC technique. A comparison between the simulated uniaxial stress–strain response with the experimental characterization of uniaxial dogbone specimens is also presented and analyzed. Concluding remarks are provided in Section 6.

2. Material selection and experimental procedure

2.1. NiTi SMA description and sample preparation

The material utilized for the experiments was a NiTi (Ti 50.4 Ni) shape memory alloy. It has been received in the form of plates, cold-worked (2.5 mm thickness), provided by Nimesis Technology. Samples were cut using waterjet hyperbaric machining and annealed (400 °C, 30 min) followed by a water-quenching. To proceed with Digital Image Correlation (DIC) technique each specimen was prepared accordingly for the purpose of full-field strain measurements by generating gray levels using a speckle pattern. DIC technique is full-field optical technique based on the recognition of geometrical changes in the grayscale distribution of surface patterns before and after straining [47,42]. It is particularly suitable for measuring kinematic fields for non-standard and heterogeneous thermomechanical tests. To this end, a white background was painted on the specimen using a flat white spray paint, and on the top a black speckle pattern was applied. This final layer was a diluted acrylic paint, applied using an airbrush. This step is very important since the spatial resolution of the DIC technique depends on the quality of the applied speckle. For the experiments, the illumination of the sample surface has been adjusted to obtain a large and rather uniform distribution of gray levels. Indeed, accurate adjustment of the speckle to the correlation pattern size is a crucial factor since each pattern must reach an adequate grayscale distribution, mandatory for the strain fields extraction [29].

2.2. Heterogeneous tests and DIC full field measurements

The mechanical test is a non-standard tensile test inspired from the work by Meuwissen et al. [35]. This non-standard test configuration gives rise to heterogeneous and non-uniform in-plane strain fields, namely longitudinal (in the direction of loading), transverse and shear strain components. By using such a heterogeneous specimen, the behavior of shear strain and stress may be demonstrated in conjunction with the principal strain and stress on the same specimen. In the present study, the original Meuwissen sample geometry has been modified and optimized to maximize the area of interest, i.e. with heterogeneous local loading paths. The dimensions of the tested samples are given in Fig. 1. In this figure, the Zone of Interest (ZI) has been defined, which is the zone where the DIC measurements are compared with numerical simulations. While the sample geometry is complex, it is loaded using a uniaxial tensile test machine. The specimen is loaded under applied displacement into a thermal chamber with a cross-head velocity of 0.5 mm/min. This value has been set from previous experiments to obtain a maximal quasi-static strain rate of the range of $10^{-3}$ s$^{-1}$ even in the area where the strains are localized, i.e. between the notches. By incrementing the applied traction, one obtains hence a spatio-temporal distribution of the in-plane strain components, which are measured using optical measurements and computed using digital image correlation (DIC). Three experiments were performed at different temperatures: 50 °C, 60 °C and 70 °C.

![Fig. 1. Geometry of the specimen for the heterogeneous tests. The dimensions are expressed in mm.](image-url)
The DIC data was analyzed using the VIC-2D software by Correlated Solutions, Inc. In this study, the pattern size was chosen as $16 \times 16$ pixels. The subset size is defined hence as the spatial resolution and equals $16$ pixels. It must be noticed that the larger the pattern size, the smaller the uncertainty. Whereas, by increasing the pattern size, the spatial resolution would be degraded. During the experiments, pictures were taken by a $2$ megapixel cooled camera ($4096$ gray level) placed in front of the specimen surface outside the thermal chamber. In order to avoid the well-known problems of non-convergence of the correlation algorithm due to the large motions of the specimen, images of the sample surface were recorded at the beginning and after each traction step corresponding to a crosshead displacement of $0.01$ mm. The displacement resolution is $2 \cdot 10^{-2}$ pixels (smallest displacement that can be detected). The strain resolution is estimated to be of the range of $2 \cdot 10^{-4}$. The resulting experimental data was extracted and stored to a pixel based grid; with one data point every $5 \times 5$ pixels. The scale from pixels to millimeters was similar for each test: one pixel every $0.065$ mm ($\pm 0.02$ mm). Thus, the DIC strain measurements were extracted and reported on a grid with one data point every $0.33 \times 0.33$ mm ($\pm 0.05 \times 0.05$ mm). The zone of interest has been defined in Fig. 1. The elongation of this zone is defined as the difference of the average of the longitudinal displacement (along $X$) between limit line 1 and limit line 2. The displacement along limit line 1 and limit line 2 obtained on the surface of the specimen from DIC. This ensures that the elongation of this zone is due to the deformation of the zone, and not to other effects (e.g. sliding in the grips).

### 2.3. Test matrix

The objective is to analyze the super-elastic behavior of SMA structures at three different temperatures, loading up to the maximum strain before unloading. Since the strain field is heterogeneous due to the structure geometry, preliminary tests were conducted where samples were loaded up to failure (tests $\#1$, $\#3$, $\#5$ in Table 1). By performing DIC analysis of these tests to up-failure, the load level at which the point of maximum longitudinal strain reaches $6.0\%$ has been determined for the three tested temperatures. This load is further utilized to perform the subsequent tests that are further utilized for the parameter identification (tests $\#2$, $\#4$, $\#6$ in Table 1). The specimens were indeed loaded until the point on the heterogeneous specimen with the maximum longitudinal strain (in the direction of loading) reaches about $6.0\%$ strain. By incrementing the applied load, a spatio-temporal distribution of the strain is obtained and is utilized to fully identify the constitutive model parameters.

### 3. Experimental results and analysis of local strain paths

The DIC strain fields at the maximum load are shown in Fig. 2 for the three temperatures, i.e. $50\ ^\circ C$, $60\ ^\circ C$ and $70\ ^\circ C$. In the same figure, the overall response of the specimen is shown as a function of time, i.e. the elongation of the zone of interest and the force. The evolution of strains ($\varepsilon_{xx}$, $\varepsilon_{yy}$, $\varepsilon_{xy}$) in six points named A–F and located in Fig. 2a were extracted from DIC results. Point B is at the center of the specimen and point D is situated almost at the lateral surface in the center of one notch. The analysis of these kinematical fields (see Fig. 2) and local loading paths of the six points shown in Fig. 3 indicates that a rather large zone between notches is undergoing transformation. The two points B and D are in the middle of the strain and transformation localization band. The solicitation at point F, far from the notches, is close to uniaxial tension. The zone located around A near the corner and below A near the border is almost free of loading. Points C and D are situated in the zones where shear strains are important.

The analysis of local strain fields obtained from DIC (at a temperature of $50\ ^\circ C$) shows that maximal longitudinal, transverse and shear strains are observed in the notched region. But strain gradients are spread over a large zone surrounding the notches. Although the imposed loading is uniaxial, due to the asymmetrical position of notches, local shear strains are locally rather important. The non-linear evolution of von Mises equivalent strain as a function of time shown in Fig. 3(a) is clear characteristic of a non-linear behavior of the material. Indeed, strain localization appears when the

Table 1  
<table>
<thead>
<tr>
<th>Number</th>
<th>Temperature (°C)</th>
<th>Load (kN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
<td>Up to failure</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>22.6</td>
</tr>
<tr>
<td>3</td>
<td>60</td>
<td>Up to failure</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
<td>24.7</td>
</tr>
<tr>
<td>5</td>
<td>70</td>
<td>Up to failure</td>
</tr>
<tr>
<td>6</td>
<td>70</td>
<td>26.7</td>
</tr>
</tbody>
</table>

Fig. 2. Longitudinal strain fields measured by DIC for the three temperatures and the location of the six particular material points (a) $\varepsilon_{yy}$ for $T = 50\ ^\circ C$, (b) $\varepsilon_{yy}$ for $T = 60\ ^\circ C$, (c) $\varepsilon_{yy}$ for $T = 70\ ^\circ C$. 

\[ T = 50\ ^\circ C \quad T = 60\ ^\circ C \quad T = 70\ ^\circ C \]
SMA undergoes a martensitic transformation close to the notched area. The changes in the slope of the curves of equivalent strains as a function of time at points C and E (as at points B and D) are thus a signature of the martensitic transformation, since it corresponds to a strain localization (increase/decrease of strain rate). The maximum equivalent strain is located near the center of the notch on the surface of the specimen (point D). The equivalent total strain reaches 6% at the maximum displacement imposed. The equivalent strain near point A is almost zero; no transformation exists near this point. Similarly, no change in rate of the strain is observed for point F, it is concluded that the material at this point does not undergo martensitic transformation. It is thus shown that the transformation starts near point D, then propagates up to the center of the sample to point B. Afterwards, the localization and transformation band enlarges towards the points C and D. Transformation at points D and B is completed before the load is reversed as suggested by the second change in the slope. It is thus observed that at the same load, a wide range of local strain paths are present in the specimen (see Fig. 3). Whereas the strain paths seem to be on the same line in terms of longitudinal strain versus transversal strain, they are scattered in terms of shear strain as a function of longitudinal or transversal strains. Strains at points A and F are really small compared to the other points. As expected, the highest values of the longitudinal and transversal strains (6% and −2.6%, respectively) are located at point D, near the far end of the notch. Longitudinal and transversal strains at the point B (4.6% and −2.4%, respectively) in the center of the specimen are smaller than those at the point D, but this point is also in the transformation band. Shear strain at point B (0.1%) is less than that at point D (0.3%) and both are small compared to their respective longitudinal and transversal strains \( \epsilon_{yy} \) and \( \epsilon_{xx} \). On the other hand, shear strains at points E and C reach 1% and −0.5%, respectively, while the longitudinal and transversal strains at this point are, respectively, 2.3% and −0.7% (respectively 3% and −1.4% for the point C). As mentioned, the variety in strain paths is mainly due to the local shear strains induced by the specimen geometry. It is noted that this variety of local strain paths is very important to be able to identify the parameters of three-dimensional constitutive models. This is especially true not only for models that include tension–compression asymmetry or anisotropy, but also for models that account for non-proportional loadings.

Fig. 3. Local strain paths for six particular locations and overall elongation and force for the test at 50 °C: (a) von Mises equivalent strains versus time, (b) transverse strains \( \epsilon_{xx} \) versus shear strains \( \epsilon_{xy} \), (c) longitudinal strains \( \epsilon_{yy} \) versus transverse strains \( \epsilon_{xx} \), (d) longitudinal strains \( \epsilon_{yy} \) versus shear strains \( \epsilon_{xy} \), (e) elongation of the zone of interest versus time and (f) force version time.
4. Thermomechanical model of martensitic transformation for SMAs and model parameters

The model proposed by Lagoudas et al. [27] describes the behavior of a Shape Memory Alloy representative volume element (RVE), considering thermoelastic and martensitic strain mechanisms that contribute to the macroscopic strain when a thermomechanical load is applied:

\[ \varepsilon = \varepsilon^c + \varepsilon^{th} + \varepsilon^{fwd}, \]

where \( \varepsilon \) is the total strain, \( \varepsilon^c \) is the elastic strain, \( \varepsilon^{th} \) is the thermal strain and \( \varepsilon^{fwd} \) is the transformation strain. The evolution of inelastic transformation strains is set to be only possible due to a change in the martensitic volume fraction, following Qidwai and Lagoudas [40], the link between the rates of \( \varepsilon^{th} \) and \( \dot{\varepsilon} \) is thus expressed:

\[ \dot{\varepsilon} = \dot{\Lambda}^t \dot{\varepsilon}, \quad \dot{\Lambda}^t = \begin{cases} \Lambda^t_{\text{fwd}}, & \dot{\varepsilon} > 0, \\ \Lambda^t_{\text{rev}}, & \dot{\varepsilon} < 0. \end{cases} \]

In the above, \( \Lambda^t_{\text{fwd}} \) is the transformation direction tensor during forward transformation and \( \Lambda^t_{\text{rev}} \) is the form during reverse. The two are defined as (cf. [67]):

\[ \Lambda^t_{\text{fwd}} = \frac{3}{2} \frac{H^{\text{fwd}} \sigma}{\varepsilon}, \quad \Lambda^t_{\text{rev}} = \frac{\varepsilon - \varepsilon^{fwd}}{\varepsilon} \]

where \( H^{\text{fwd}} \) is the uniaxial transformation strain magnitude for full transformation and is considered equal to the maximum transformation strain magnitude throughout this paper. Also, the effective stress \( \bar{\sigma} \) is defined in the Mises sense. This model has been implemented in the FEA package Abaqus, which is used in this work to perform the finite element simulations. The model parameters can be deducted from a uniffed set of material parameters proposed by Stebner et al. [46] (see Table 2).

Considering that the SMA is subjected only to superelastic loadings, it is unnecessary to identify the two parameters associated with martensitic reorientation \( \sigma^r \) and \( \sigma^l \). Moreover, since it is assumed that the transformation hardening is proportional to the volume fraction of martensite, a relationship is assumed between \( M_f - M_s \), \( C^M \) and \( A_f - A_s \). An additional parameter is introduced to represent such relation, already utilized by Peulitier et al. [39] and Chemisky et al. [11] has been defined:

\[ H_f = \varepsilon_{\text{actmax}}^l a C^M \phi(M_f - M_s) - \sigma_{\text{actmax}}^l a C^\phi(0 - A_s), \]

such that the knowledge of \( \varepsilon_{\text{actmax}}^l a C^M \phi \), \( C^\phi \), \( M_s \), \( A_f \) and \( H_f \) is sufficient for the knowledge of the martensitic transformation characteristics. Furthermore, the effect of thermal expansion is neglected for superelastic loadings and the elastic parameters are assumed to be the same for the austenitic and martensitic phases. These assumptions are physically motivated and reduce the number of material parameters to be identified, keeping a good balance between the model accuracy and the complexity of the identification procedure. With such conditions, the remaining eight independent parameters are \( E \) (MPa), \( \nu \), \( H_f \) (MPa), \( \varepsilon_{\text{actmax}}^l a C^M \phi \), \( C^\phi \), \( M_s \) (MPa C\(^{-1}\)), \( C^M \) (MPa C\(^{-1}\)), \( M_s \) (C) and \( A_f \) (C).

Note that these assumptions reduce the identification problem, but the identification of the full model is also possible. However, depending on the experimental database, it might be hard to identify some of the parameters if their sensitivity is weak or null for the corresponding numerical simulation (i.e. parameters related to reorientation from superelastic loading paths).

5. Parameter identification using finite element updating method coupled with an hybrid optimization algorithm

5.1. Definition of the objective function

Having a constitutive model with a specific set of material parameters, the identification problem consists in the determination of material parameters that minimize the difference between computed data and a set of experimental data. The numerical model is typically a finite element model, with the same geometry and boundary conditions than the experimental setup, and considering the aforementioned constitutive model. A part of the measured displacement at the boundary of the specimens is then utilized to define (or to complete the definition) of the boundary value problem to solve using the finite element method and another part is utilized to define a cost function to be minimized.

The cost function is considered to be:

\[ C(p) = \frac{1}{2} \sum_{i=1}^{n} \frac{1}{N_i} (\varepsilon_{\text{num}}^l i(p) - \varepsilon_{\text{exp}}^l i)^2 \]

where \( C(p) \) is the cost function, \( \varepsilon_{\text{num}}^l i(p) \) is the \( i \)-th information obtained with the numerical simulation, \( \varepsilon_{\text{exp}}^l i \) is the \( i \)-th information obtained from the set of experiments conducted and \( N_i \) is a weight factor. Note that all these information are potentially obtained from a number of experiments at different times and at different spatial positions.

In the methodology proposed here, the boundary conditions are determined using the displacement experimentally measured using DIC at the lower and the upper boundary (as \( y \) is considered as the vertical axis) of the zone of interest. Considering a thin specimen, on the left and right sides and on the front and back faces it is assumed that no tractions are applied to complete the identification area of interest is required to be compatible with the load cell in the experimental setup. Note that with this definition the area of interest is required to be compatible with a measure of force considering an appropriate section. The grid for the experimental DIC data has thus to be compared with numerical results obtained from finite element analyses. It is necessary to interpolate the results from DIC grid over the FEA grid. A bilinear interpolation algorithm has been developed for such interpolation, where interpolated values of the new grid are obtained based on the values of the four nearest neighbors forming a quad surrounding the interpolated point. For the specific case on non-convex geometries such as Mewissen samples, a control of the quad geometry is necessary to remove badly shaped (flat) quad and therefore interpolate uniquely from values at geometric points inside the geometric shape of the specimen.

The final version of the cost function is written as:

\[ C(p) = \frac{1}{2} (\varepsilon_{\text{exp}}^l \varepsilon_{\text{num}}^l p - \varepsilon_{\text{exp}}^l)^2 + \sum_{i=1}^{n} \sum_{j=1}^{4} \frac{\varepsilon_{\text{exp}}^l \varepsilon_{\text{num}}^l p - \varepsilon_{\text{exp}}^l)^2}{\sum_{i=1}^{n} \sum_{j=1}^{4} \varepsilon_{\text{exp}}^l \varepsilon_{\text{num}}^l p - \varepsilon_{\text{exp}}^l)^2} \]

Table 2

<table>
<thead>
<tr>
<th>Material parameters for typical NiTi materials.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E ) (Elastic Modulus, Austenite (Martensite)) (MPa)</td>
</tr>
<tr>
<td>( \nu ) (Poisson ratio of Austenite (Martensite))</td>
</tr>
<tr>
<td>( \sigma^r ) (Coeff. Thermal Exp., Austenite (Martensite))</td>
</tr>
<tr>
<td>( M_s ) (Mart. Start Temperature, ( \sigma = 0 ))</td>
</tr>
<tr>
<td>( M_f ) (Mart. Finish Temperature, ( \sigma = 0 ))</td>
</tr>
<tr>
<td>( A_s ) (Mart. Start Temperature, ( \sigma = 0 ))</td>
</tr>
<tr>
<td>( A_f ) (Mart. Finish Temperature, ( \sigma = 0 ))</td>
</tr>
<tr>
<td>( C^M ) (Uniaxial stress—temperature slope, forward transformation)</td>
</tr>
<tr>
<td>( C^\phi ) (Uniaxial stress—temperature slope, reverse transformation)</td>
</tr>
<tr>
<td>( \sigma^r ) (Martensitic reorientation start)</td>
</tr>
<tr>
<td>( \sigma^l ) (Martensitic reorientation finish)</td>
</tr>
<tr>
<td>( \varepsilon_{\text{actmax}}^l ) (Maximum recoverable transformation strain)</td>
</tr>
</tbody>
</table>
5.2. Optimization algorithm

The definition of an optimization algorithm has to take into account the computational time necessary to perform a numerical simulation for a given set of parameters. Indeed, it often depends on the model utilized to obtain values entering in the definition of the cost function according to a given set of design variables. Especially, the efficiency of the algorithms depends on the time necessary to compute a numerical solution and then evaluate the corresponding cost to be able to evaluate the prescribed set of parameters within the set of admissible solutions. Also, the form of the cost function might impact the decision of an optimization algorithm, e.g. if local minima are expected. Indeed, the main issue associated with gradient-based techniques is that the method ensures the convergence to local minima. Heuristic such as genetic algorithm should be therefore utilized in such cases to determine preferential sets of parameters from a large population that are close to the global minima [9]. In the approach proposed, the genetic algorithm and the gradient-based method are used simultaneously, with the following procedure:

1. An initial population of \( C_0 \) individuals (each individual is a set of parameters) is generated. The selection of individuals can be aleatory, given limitations of the material parameters, or can be generated using design of experiments (DOE).
2. The numerical simulation for all the individuals is computed in parallel.
3. All the individuals are scored by computing their cost value.
4. The individuals of the initial generation are classified, and a number \( n \) of them are selected (the best ones). These constitute the ‘current’ generation \( n \).
5. A set of children of the current generation (with \( n \) members) is determined using the crossover technique. A mutation probability has been added to increase the diversity of the children generation. For the same purpose a small deviation operator from the parent’s parameters is also applied.
6. The best members of the current generation are selected. The set of numerical simulation to apply the finite differences derivation technique is determined.
7. The numerical simulation for all the children and the simulations required for the finite differences derivation are computed in parallel.
8. The cost function of all the individual is determined.
9. The parameters of a number \( N_{opt} \) of the best members are updated using a Levenberg–Marquardt algorithm.
10. The current generation \( n+1 \) is determined from the best individuals among the current generation \( n \) and the children.
11. Stationary condition test for the best individual of the current generation compared to the previous best. If needed, reloop from item 5.

An overview of the identification procedure in general is presented in Fig. 4.

5.3. Finite element model and experimental boundary conditions

A finite element model has been developed using Abaqus. Only an area close to the notches is considered for the model that corresponds to the window recorded by the camera. As previously mentioned, the prescribed displacement at the boundary are set from the DIC evaluation of the displacements. A total of 377 linear elements with full integration (C3D8H denomination in Abaqus) is utilized to discretize the simulated domain of the tested specimen. Note that the model utilized in this work [27] assumes an isotropic behavior for the elastic strain mechanisms and the phase transformation strain mechanism, and the numerical simulation of the three-dimensional response of SMA structures are performed under this assumption. The information of in-plane strain on the surface corresponding to the one analyzed using DIC is extracted from the elements centroids close to the notches, to maximize the heterogeneity of the strain field entering in the identification procedure. It results in a total of 247 material points where the three in-plane components of strain \( \varepsilon_{xx}, \varepsilon_{yy}, \) and \( \varepsilon_{xy} \) are compared. Over a full loading–unloading cycle, the values of the strain at the material points are compared on 20 increments, uniformly spaced according to the applied displacement.

6. Parameter identification results and validation of the identification procedure

6.1. Identification of the material parameters based on the heterogeneous Meuwissen experiments

The identification method presented in Section 5 is applied to the case of a set of experimental data corresponding to three isothermal tests performed at 50 °C, 60 °C and 70 °C. Several tests have been conducted with different parameters of the hybrid genetic–gradient based algorithm, which are resumed in Table 3.

The different control parameters of the algorithm are the following:

- The initial number of individuals of the first generation \( N_{ini} \).
- The number of sons generated and the number of retained individual for the current generation \( N_{cur} \).
- The number of best individuals selected to be optimized via the Levenberg–Marquardt gradient based algorithm \( N_{cb} \).
- The probability of mutation of the parameters set to 5%. This means that for each parameter of all the new individual, there is a 5% chance that the concerned parameter is set randomly inside the design space rather than being selected from the crossover technique. This low value is generally utilized when a low population is defined to keep a certain stability of the generations, still slowly introducing new parameters to authorize the exploration of new points in the design space.
- The first (best) half of each generation is considered for the crossover technique.

For each case the identification was stopped after a maximum of 30 generations, or after a stabilization of the cost function. The algorithm has been run on a cluster to be able to parallelize the numerical simulations. This is achieved using a developed software coupled with TORQUE Resource Manager for the management of the simulation on the cluster.

The set of initial parameters is bounded to define the design space (see Table 4):

The final material parameters have been identified for each of the four configurations. In Fig. 5, the cost function associated with the best element of each generation is shown for these cases. The non-smooth gaps from one generation to another (e.g., first generation of test case 3) is characteristic of the appearance of a new individual from the evolutionary algorithm while the gradual decrease of the cost function is associated with the Levenberg–Marquardt optimization. It is shown here that cases #1, #3 and #4 are converging to similar minima in terms of the value of the cost function. The case #2 is stuck to local minima. This yields the importance of the number of individuals in the first generation to ensure both genetic and gradient-based algorithms to be efficient. The first algorithm requires a sufficient population to be able to efficiently find a better configuration, while the second requires a good starting point close enough to the global minima. When several individuals are optimized through a Levenberg–Marquardt algorithm, it is shown that again local minima are often reached, even with initial points close to the final value of the cost function. Only one individual has reached the minimal value comparable to those observed for the cases #1 and #3. According to the complexity of the behavior, the number of experimental points and the noise introduced by the experimental measurement of DIC, local minima close to the global minimum solution are expected. To obtain a response as close as possible to experimental values, a large initial number of individual for the first generation are recommended. The number of individuals in each generation and the number of individuals to be optimized via gradient-based algorithm depend on the accuracy versus computational time. It is found here that the configuration #3 gives the best results and will be further utilized for the comparison of strain fields and force versus displacement response. The final identified parameters from this configuration are shown in Table 5.

The comparison between the strain field obtained using DIC and the numerical simulation with the identified parameters is shown in Fig. 6. Globally, the correlation between the experimental and simulated fields is satisfying, especially in the area where the in-plane strain components are rather important due to the localization effect induced by the presence of notches. It is further noticed that differences appear for the longitudinal strain $\varepsilon_{yy}$ and the in-plane shear strain $\varepsilon_{xy}$ around the notches.

To investigate further the accuracy of the identification procedure, the evolution of local strain fields computed in two points of interest, namely points B and D where position are recalled in Fig. 6(a), is compared with experimental data to see if the numerical model with identified parameters is able to track the local evolution of strain. This temporal evolution is shown in Fig. 7 for the longitudinal strain $\varepsilon_{yy}$. It is seen that the evolution of local strain is correctly described for both loading and unloading for the longitudinal and transverse strains of the center point B. However, differences are observed over the surface in the center of one notch (point D). The temporal evolution is indeed different, since the transformation strains appear sooner in the simulations. This indicates that the model might not be able to accurately track local multiaxial non-proportional loadings. It is worth to notice that in this work the evolution of transformation strains in different directions is normal to an isotropic transformation surface.

Note that the prediction of local strain and stress fields is very important to determine, for instance, the fatigue lifetime of a SMA component. Indeed, it has been shown that the fatigue life of SMA actuators is linked to the actuation work produced over an actuation

---

**Table 3**

Parameters of the hybrid genetic–gradient based algorithm for the four identification tests.

<table>
<thead>
<tr>
<th>Number</th>
<th>$N_{test}$</th>
<th>$N_{var}$</th>
<th>$N_{c}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>150</td>
<td>50</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>150</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

**Table 4**

Min and max bounds for the material parameters.

<table>
<thead>
<tr>
<th>Material parameter</th>
<th>$E$ (MPa)</th>
<th>$\nu$</th>
<th>$H_f$ (MPa)</th>
<th>$\varepsilon_{uc}^f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E$ (MPa)</td>
<td>50 000</td>
<td>0.2</td>
<td>0.5</td>
<td>0.03</td>
</tr>
<tr>
<td>$\nu$</td>
<td>80 000</td>
<td>0.45</td>
<td>6</td>
<td>0.05</td>
</tr>
<tr>
<td>$c^H$ (MPa C$^{-1}$)</td>
<td>67 500</td>
<td>0.394</td>
<td>1.79</td>
<td>0.0416</td>
</tr>
<tr>
<td>$c^L$ (MPa C$^{-1}$)</td>
<td>8.74</td>
<td>9.66</td>
<td>−14.5</td>
<td>5.5</td>
</tr>
</tbody>
</table>

**Table 5**

Identified material parameters NiTi materials.

<table>
<thead>
<tr>
<th>$E$ (MPa)</th>
<th>$\nu$</th>
<th>$H_f$ (MPa)</th>
<th>$\varepsilon_{uc}^f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>67 500</td>
<td>0.394</td>
<td>1.79</td>
<td>0.0416</td>
</tr>
<tr>
<td>$c^H$ (MPa C$^{-1}$)</td>
<td>67 500</td>
<td>0.394</td>
<td>1.79</td>
</tr>
<tr>
<td>$c^L$ (MPa C$^{-1}$)</td>
<td>8.74</td>
<td>9.66</td>
<td>−14.5</td>
</tr>
</tbody>
</table>
cycle [17]. The overall mechanical response of the sample, in terms of the evolution of force with respect to the average displacement at the boundary of the DIC window, is compared in Fig. 8 at three different temperatures, i.e. 50°C, 60°C and 70°C. It is shown that globally the overall behavior during loading is accurately represented by the numerical model using the identified parameters. The behavior during unloading is correctly simulated for the test performed at 50°C. More pronounced differences appear during reverse transformation, especially at 60°C and 70°C. It is noted here that the transformation hardening characteristics have

Fig. 6. Contour plots of the experimental strain fields (a, d, g), the numerical strain fields (b, e, h) and their absolute difference (c, f, i) for the components $\varepsilon_{xx}$, $\varepsilon_{yy}$ and $\varepsilon_{xy}$, respectively, at 50°C with a tensile load of 22.7 kN.
been selected to be similar between forward and reverse transformation. The selection of different hardening parameters could have improved the results, with the cost of increasing the number of parameters to identify.

6.2. Comparison with identification procedure based on homogeneous uniaxial tests

To further validate the identification procedure, numerical simulations using the identified values are utilized to simulate the stress–strain response of homogeneous tensile tests. Uniaxial experiments have been performed on dogbone specimens of the same material, with the same thermomechanical treatment than the Meuwissen specimens. These samples were also cut using waterjet hyperbaric machining and annealed (400 °C, 30 min) followed by a water-quenching. Three uniaxial tests were performed up to a maximum strain of 6%, at three temperatures, 30 °C, 43 °C and 50 °C, at a strain rate of the order of $10^{-4}$ s$^{-1}$. The comparison between the experimental response of the uniaxial sample and the numerical simulation of the uniaxial response, based on the model parameters identified from the DIC fields on the Meuwissen samples, is shown in Fig. 9.

The comparison between the numerical simulations and the experimental results leads to the following remarks:

- The elastic regime is well predicted, especially at 43 °C and 50 °C, while it seems that the elastic response is softer at 30 °C. It is noted that while the expected value of the Young modulus for NiTi systems is about 80 GPa, the identification of such low Young modulus from uniaxial tests is often reported, e.g. in Hartl and Lagoudas [18]. Actually, from the definition of the phenomenological model, the target parameter is rather a linear apparent modulus than the elastic Young's modulus. In this linear stage martensitic transformation may be present, which explain the high value of the Poisson ratio, since martensitic transformation appears almost without volume change for NiTi systems. Superelastic uniaxial tests at different temperatures have shown that the apparent moduli decrease when the test temperature gets closer to transformation temperatures [8,12]. This phenomenon can be attributed to different factors, i.e. R-phase transformation [44], elastic softening and diffuse or multiple stage transformation [41]. However such effects are not described in the model utilized in this study [27] and are hence not simulated here. Since the uniaxial tests were performed at a temperature lower than the test on notched specimens, the differences observed for the elastic properties identified are consistent with previous observations of the characteristic behavior of NiTi.
- The simulation of the change of strain mechanism from martensitic transformation to a fully elasticity of martensite arise at a level of strain that corresponds to the change of regime observed experimentally. Note that the identification procedure integrates in the cost function the difference between the simulated strain measured on the Meuwissen samples up to a maximum of 6% von Mises equivalent strain. Only a few material points are loaded up to this level, so the identification of this stage is not expected to be very accurate. Also, the model considers a sharp change of the two regimes, which is not observed experimentally. At this point, a strong non-linear transformation has to be considered to increase the simulation of SMA devices up to high local stress–strain states.
- The critical stress for the onset of transformation is not very well captured for the three temperatures. A constant shift of 20–50 MPa is observed. According to the identified slopes in the uniaxial stress–temperature diagram from the full-field measurements, this corresponds to a shift of the transformation temperatures of 2.3–5.8 °C. Nevertheless, the critical stress for the onset of reverse transformation is very well predicted. Such differences have to be accounted for in the design procedure that utilizes the proposed methodology. The numerical simulation of SMAs should integrate the variability of the material response. This is a general statement about the design of SMAs components since it is recalled here that a very small change of composition in NiTi system leads to a drastic change of the transformation temperatures [48]. The precipitation of No-rich Ni during heat treatments also modify the local composition of alloys and lead to a drastic change of the overall behavior [4].
- The stress–temperature dependence (slopes in the uniaxial stress–temperature diagram $C^0$ and $C^1$) is very well predicted. This yields the capability of the developed procedure to simulate SMA components response at different temperatures.

7. Conclusions

In this work, an inverse identification procedure has been developed to determine the material parameters of SMAs directly on structures using full field measurements. The proposed identification method, using a parallelized hybrid genetic–gradient based optimization algorithm, is able to properly identify the material...
parameters of a Shape Memory Alloy. It has been demonstrated that a large initial population is important to be able to rapidly select a good starting point for a gradient-based Levenberg–Marquardt algorithm, avoiding hence local minima. The simulated in-plane strain field and the global force–displacement of the Meuwissen sample are in good agreement with the experimental observation. This validates the capability of the proposed identification procedure to find a suitable set of material parameters that describe accurately the material behavior. The comparison between the simulation of the uniaxial stress–strain response has led to several important recommendations on the use of the proposed identification procedure in the design strategy of SMA components:

- While the uniaxial simulations are in good agreement with the experiments, it is necessary to integrate the variability of the thermomechanical response.
- Since the identification procedure relies on two-dimensional measurements, the three-dimensional simulation of SMA thermomechanical response requires additional assumptions about the anisotropic nature of their behavior.
- In the case of heterogeneous kinematical fields, since only a few material points are loaded up to the maximum strain level, the identification of the thermomechanical response corresponding to this stage is not expected to be very accurate and the results should be taken with caution.

Such an identification method is nevertheless of great importance for material that are extremely sensitive to thermomechanical processing conditions, e.g. NiTi materials where it is very difficult to test the material properties since processing conditions of the
device could not be easily reproduced on the simple test samples.
Thus, it is very difficult to proceed in such cases with the legacy calibration methods and identification procedures have to utilize the thermomechanical characterization of the structures, often implying heterogeneous conditions. Furthermore, exploiting the heterogeneity of the local strain paths, this method can also be utilized to extract material parameters that are difficult to obtain based on uniaxial experiments, such as material parameters characteristic of anisotropy and tension-compression asymmetry, since this method applies to heterogeneous tests with local multiaxial non-proportional conditions. Further work will emphasize on the development of anisotropic criteria for phase transformation to be able to calibrate and simulate the response of three-dimensional shape memory alloys with an increasing accuracy.

Acknowledgment

The support of the NSF International Institute of Multifunctional Materials for Energy Conversion (IIMEC), award #0841082, is acknowledged.

References

[34] Schreier H, Orteu J-J, Sutton M. Image correlation for shape, motion and deformation measurements: basic concepts, theory and applications; 2009.