Elastocaloric Cooling: State-of-the-art and Future Challenges in Designing Regenerative Elastocaloric Devices

Parham Kabirifar – Andrej Žerovnik – Žiga Ahčin – Luka Porenta – Miha Brojan – Jaka Tušek*
University of Ljubljana, Faculty of Mechanical Engineering, Slovenia

The elastocaloric cooling, utilizing latent heat associated with martensitic transformation in shape-memory alloys, is being considered in the recent years as one of the most promising alternatives to vapour compression cooling technology. It can be more efficient and completely harmless to the environment and people. In the first part of this work, the basics of the elastocaloric effect (eCE) and the state-of-the-art in the field of elastocaloric materials and devices are presented. In the second part, we are addressing crucial challenges in designing active elastocaloric regenerators, which are currently showing the largest potential for utilization of eCE in practical devices. Another key component of elastocaloric technology is a driver mechanism that needs to provide loading for active elastocaloric regenerators in an efficient way and recover the released energy during their unloading. Different driver mechanisms are reviewed and the work recovery potential is discussed in the third part of this work.

Keywords: elastocaloric effect, cooling, active elastocaloric regenerator, fatigue life, driver mechanism

0 INTRODUCTION

Modern society is totally dependent on cooling, refrigeration and air-conditioning. In Europe, 70 % of food is chilled or frozen; 90 % of all vaccines must be chilled; and air-conditioning is essentially everywhere, in our homes, cars, offices, etc. In the USA, 87 % of households use air-conditioning, the energy consumption of which is more than Africa’s total electricity output. Moreover, the energy consumption required for cooling and air-conditioning is growing exponentially. Nowadays, in the developed countries it accounts for around 20 % of the total energy consumption; by 2060, the energy used for cooling will overtake that of heating and by the end of the century, cooling will outstrip heating by 60 % [1]. Nevertheless, that is only half of the problem. Nearly all cooling and air-conditioning devices out there, as well as heat-pumps, are based on vapour compression technology, which is over a century old. In fact, this is one of the oldest electricity-based technologies still in use, without a viable alternative [2]. Over the decades of development, vapour compression technology has increased its efficiency and specific power (and therefore its compactness) along with reducing its environmental impact. However, from the thermodynamic perspective, the efficiency of the vapour compression cooling cycle has reached its limit and no significant improvements should be expected in the future. Despite decades of development, vapour compression technology has still three important disadvantages. The first problem is its efficiency. The exergy efficiency, which is the ratio between the actual and Carnot’s coefficient of performance (COP), of the vast majority of the cooling systems in use is well below 30 %. The second problem is the refrigerants that could leak into the atmosphere and harm the environment/people when the refrigerator is discarded or damaged. Nowadays, these refrigerants are no longer ozone depleting (such as chlorofluorocarbons), as they have been banned since 1989 by the Montreal Protocol on Substances that Deplete the Ozone Layer [3]. We have replaced them mostly with hydrofluorocarbons, which have another problem; they cause a significant greenhouse effect (up to a few thousand times more than carbon dioxide). As a consequence, these types of refrigerants are already subject to a ban (according to the Kigali Agreement to the Montreal Protocol [3] signed in 2016) and in the near future need to be replaced preferably with natural refrigerants such as hydrocarbons (e.g. isobutene), carbon dioxide and ammonia. Nonetheless, such alternatives can be explosive in large quantities (hydrocarbons), require...
very high pressures (carbon dioxide) and be toxic (ammonia) [4] and [5]. Therefore, the problem of refrigerants is not yet sufficiently solved and further research and development is required to fulfil all the environmental and safety requirements. The third weakness of vapour compression technology, which can be very important for some applications, is the presence of vibrations and noises that are inevitable during the operation of a compressor.

In recent decades, several alternatives to vapour compression were discovered and developed but owing to various reasons such as either low efficiency or low specific power or impractical applications, none of them has proven to be a serious alternative for the everyday cooling needs. However, more recently the so-called caloric cooling technologies have shown significant potentials as serious alternatives to vapour compression technology [6] and [7]. By caloric cooling technologies, we mean magnetocaloric (magnetic), electrocaloric, elastocaloric and barocaloric cooling. The basis of caloric cooling is to exploit the latent heat of a solid-state phase transformation of a ferroic material by applying an external magnetic, electric or mechanical field. These technologies can be significantly more efficient (with theoretical exergy efficiency above 50 %), completely harmless to the environment (by applying solid-state refrigerants) and potentially vibration and noise free [8]. Among the caloric cooling technologies, the most matured and developed is the magnetic (magnetocaloric) refrigeration with up to 100 prototypes developed in universities and laboratories around the world [8] and [9]. However, more than 20 years of research on magnetocaloric materials and magnetic refrigeration systems has still failed to provide the necessary breakthrough that could lead to a commercial realisation of this technology and satisfy the urgent global need for a more efficient and environmentally harmless refrigeration. The possibility of inducing a solid-state phase transformation via martensitic phase transformation and generating a caloric effect by means of mechanical stress i.e. the elastocaloric effect (eCE) in superelastic shape-memory alloys (SMAs), opens up new avenues for solid-state refrigeration. According to the great potential of eCE, the US Department of Energy in 2014 [2] and more recently also the European Commission in 2016 [10] selected the elastocaloric cooling technology as the most promising alternative to the vapour compression refrigeration in the future.

In this work, the basics of eCE and the state-of-the-art in elastocaloric materials and devices are presented. Since the most promising results (in terms of cooling and heat-pumping characteristics) were achieved using a regenerative elastocaloric cooling cycle, the second part of this work focuses on the challenges of designing an elastocaloric porous structure to be utilized in an active elastocaloric regenerator, which is the first key component of a regenerative elastocaloric cooling device. The second key component of an elastocaloric cooling device is a driver mechanism that exerts mechanical stress on the elastocaloric material. Different elastocaloric driver mechanisms are reviewed and discussed in the third part of the paper.

1 STATE-OF-THE-ART IN ELASTOCALORIC TECHNOLOGY

1.1 Shape-memory Materials and the Basics of the eCE

The eCE is strongly related to the superelasticity of SMAs and the latent heat that is released/absorbed during the forward/reverse martensitic phase transformation. Martensitic transformation is a reversible solid-state displacive (non-diffusional) crystalline phase transformation dominated by a shear between a high-symmetry high-temperature parent phase and a low-symmetry low-temperature product phase (ordered B2 austenite parent phase and B19’ monoclinic phase in the case of the near-equiatomic Ni-Ti alloy) [11]. The martensitic phase transformation is responsible for two unique properties of SMAs i.e. shape-memory effect and superelasticity. In shape-memory effect, a mechanically deformed SMA can “remember” its initial shape (before being deformed) and regain it upon being heated. In superelasticity, the SMA can be loaded to very large strains (up to 8.5 %) and release the transformation heat. Upon being unloaded, it can regain its initial shape (without any permanent deformation) and absorb the heat. Occurrence of shape-memory effect or superelasticity in a particular SMA sample depends on its temperature. Above austenite stability temperature ($A_f$), a SMA sample is superplastic and below martensite stability temperature ($M_f$), the sample exhibits shape-memory behaviour. These two properties make SMAs very attractive to various applications in different areas such as the automotive and aircraft industries (actuators, dampers, variable geometry chevrons, etc.), robotics, civil engineering (concrete beam reinforced with SMA) and in particular in medical and biomedical applications (implants, orthodontic wires, self-expandable stents, minimally invasive surgical instruments, etc.) [12]. The most well-known SMA is the binary near-equiatomic Ni-Ti alloy, which was first reported in 1963 and remains by far the most widely
applied SMA in various applications owing to its superior fatigue behaviour (among other SMAs) and biocompatibility. Other shape-memory (superelastic) alloys which are developed and characterized are Ni-Ti based (alloyed with Cu, Co, Pd, Fe, etc.), Cu-based (alloyed with Al, Ni, Zn, etc.), Fe-based (alloyed with Pd, Mn, Si, Ni, etc.) and magnetic SMAs (e.g. Ni-Fe-Ga and Ni-Mn-Co). The shape-memory effect and superelasticity are also exhibited by polymers (e.g. natural rubber) although such behaviours in polymers are not caused by the martensitic phase transformation [12].

Even though the latent heat release and absorption during the martensitic phase transformation in superelastic alloys and the temperature changes associated with them have been known for almost 40 years [13], it was not until recently that the eCE was recognized as a potential cooling or heat-pumping mechanism [14]. In Ni-Ti alloys, the released transformation (latent) heat can be as high as 35 J/g [15]. The basics of an elastocaloric cycle are the following. When an SMA in the austenite phase (at temperatures above $A_f$) is strained/stressed, the exothermic forward martensitic transformation (austenite to martensite) occurs. The generated heat could be slowly dissipated (under isothermal conditions) or under adiabatic conditions, heat the sample up and enable it to subsequently transfer the generated heat to the surroundings (or the heat sink). When the stress is removed, the endothermic reverse transformation (martensite to austenite) occurs, which under adiabatic conditions, cools down the sample and enables it to absorb the heat from the surrounding (or the heat source). An elastocaloric cooling cycle is shown in Fig. 1.

The eCE can be quantitatively described by isothermal entropy change (Eq. (1)) and/or adiabatic temperature change (Eq. (2)), which can be calculated using the following equations (derived from Maxwell relation):

$$\Delta s_{iso} = \frac{1}{\rho} \int_0^\sigma \left( \frac{\partial c}{\partial T} \right)_T d\sigma,$$

$$\Delta T_{ad} \approx -\frac{1}{\rho} \frac{T_0}{c} \Delta s_{iso}.$$

In the above equations, $\rho$, $\epsilon$, $\sigma$, $c$ and $T_0$ are material’s density, applied strain, applied stress, specific heat of the material and material’s temperature, respectively. It should be noted that in the presented form, Eq. (2) can be used only for estimating the adiabatic temperature changes since the specific heat ($c$) is taken as a constant value and the temperature at which adiabatic temperature change is calculated does not include the sample’s self-heating and self-cooling under the adiabatic conditions. Details on a more accurate way of calculating the adiabatic temperature changes could be found elsewhere [16].

1.2 Elastocaloric Materials

Generally speaking, all SMAs (and also shape-memory polymers) can be considered as potential elastocaloric materials if their transformation temperatures ($A_f$) are below the working temperature of the device, which is a precondition for reversible superelasticity. Table S1 of the Supplementary Data shows the basic elastocaloric properties for the most interesting elastocaloric materials published to date. The elastocaloric materials can be divided into three main groups: shape-memory alloy-based (Ni-Ti-based, Cu-based and Fe-based), shape-memory polymer-based (SMP) and magnetic shape-memory alloy-based (MSMA). In addition to the adiabatic temperature changes (and their corresponding applied stress/strain), Table S1 also shows other parameters such as sample preparation, sample’s microstructure and geometry, sample’s history (stabilization and training), loading mode (tension or compression) and measuring technique (thermocouples or IR camera) that can all significantly affect the eCE. For example, it has been demonstrated [17] that eCE is decreasing during the initial cycling before the sample is stabilized. Moreover, it is well known that
SMAs exhibit tension-compression asymmetry [18] and that their behaviour strongly depends on their microstructure (e.g. grain size, texture, presence and type of precipitates, etc.) [19] and [20]. It is therefore important to specify such parameters when comparing different elastocaloric materials.

The most studied eCE is that of the binary Ni-Ti alloy under tensile loading, where adiabatic temperature changes of up to around 25 K were measured at room temperature [21] and [22]. Much less explored is the eCE in Ni-Ti alloys under compressive loading, nevertheless, the initial results show that adiabatic temperature changes similar to those of tension can be obtained under compression as well. It has been demonstrated that alloying binary Ni-Ti with Cu, V and/or Co improves its fatigue life and/or decreases its hysteresis but at the expense of decreasing its adiabatic temperature changes [17] and [23] to [26]. In Cu-based SMAs, such as Cu-Zn-Al and Cu-Al-Ni, adiabatic temperature changes of up to 15 K were measured. The main advantage of these alloys is that such adiabatic temperature changes are generated under smaller transformation stresses compared to Ni-Ti [13], [14] and [27]. In Fe-based SMAs, such as Fe-Rh and Fe-Pd, adiabatic temperature changes of up to 5 K were measured along with a minor hysteresis, which is the main advantage of these alloys [28] and [29]. Nonetheless, high prices of Rh and Pd prevents practical applications of these alloys.

The adiabatic temperature changes in MSMA are up to around 10 K at moderate stresses but since most of these alloys are brittle, their application in practical cooling devices would be impractical [30] and [31]. Very recently, a new elastocaloric alloy (Ni-Mn-Ti-B) was discovered in which a reversible adiabatic temperature change of 31.5 K was measured, which is the largest directly measured reproducible caloric effect so far [32]. With that, the authors showed that there is still plenty of room for improvements in the eCE by designing new elastocaloric materials. In the near future, we can, therefore, expect elastocaloric materials with even higher eCE.

Among shape-memory polymers, the most studied for elastocaloric proposes is natural rubber [33], where adiabatic temperature changes of up to 12 K were measured at very small stresses (up to 2 MPa), however, its main disadvantages for practical elastocaloric applications are poor thermal (heat transfer) properties and in particular, very high required strains (up to 600 %), which makes application of elastocaloric polymers very unpractical.

A recent comprehensive review on elastocaloric alloys can be found in [34].

### 1.3 Elastocaloric Devices

In recent years, concurrent with the research and development in elastocaloric materials, a significant progress was made in the design and development of the elastocaloric cooling/heat-pumping devices [35] to [48] and the corresponding numerical models [46] to [49] for simulation and optimization of their performance. In general, the proposed concepts of elastocaloric devices can be divided into two main groups:

- concepts based on single elastocaloric element(s) with solid-to-solid contact heat transfer;
- concepts based on porous elastocaloric structures with convective heat transfer.

The most widely applied thermodynamic cycle for exploitation of eCE (for both of the above-mentioned concepts) is Brayton thermodynamic cycle, which is based on four basic operational steps (Fig. 2):

- adiabatic loading (where austenite (A) to martensite (M) transformation occurs), which owing to the eCE, heats up the elastocaloric material;
- heat transfer from the elastocaloric material to the heat sink (by contact or convective heat transfer);
- adiabatic unloading (where martensite (M) to austenite (A) transformation occurs), which owing to the eCE, cools down the elastocaloric material;
- heat transfer from the heat source to the elastocaloric material (by contact or convective heat transfer).

In general, elastocaloric materials in cooling/heat-pumping devices can also operate based on other thermodynamic cycles such as Ericsson, Carnot or some combination of them [50]. Moreover, cooling/heat-pumping devices can operate in compliance with single-stage or multi-stage thermodynamic cycles. The latter can be further divided into heat-recovery, active regenerative and cascade cycles.

The main difference between a single-stage and a multi-stage thermodynamic cycle is in the generated temperature span. In single-stage devices, the maximum temperature span is limited by the adiabatic temperature changes ($\Delta T_{ad}$) of the material. Therefore, such devices can only operate around a surrounding temperature ($T_{sur}$) as shown in Fig. 2c. On the other hand, multi-stage thermodynamic cycles enable an increase in the temperature span ($T_{span}$) between the heat sink and the heat source. Among them, currently, the largest potential for practical applications is demonstrated by an active regenerative thermodynamic cycle (owing to its compactness...
Elastocaloric Cooling: State-of-the-art and Future Challenges in Designing Regenerative Elastocaloric Devices

This concept was first presented for utilizing the magnetocaloric effect in magnetic refrigeration [8] and enables generating temperature spans along the length of the regenerator that can be several times larger than the adiabatic temperature changes of the magnetocaloric or elastocaloric material itself. This is crucial for application in practical cooling or heat-pumping devices where temperature spans of 30 K or more are usually required. An active elastocaloric regenerator is a porous structure made of an elastocaloric material through which a heat-transfer fluid is pumped in the counter-flow direction. If the operating conditions (frequency, mass flow rate, etc.) are appropriate, a temperature span between the heat source temperature \( T_h \) and the heat sink temperature \( T_c \) is established under the steady-state conditions. Accordingly, each element of the elastocaloric material along its length performs its own thermodynamic cycle at a slightly different temperature (between \( T_h \) and \( T_c \)) as shown in Fig. 2d. As also shown in Fig. 2d, the entire elastocaloric material in the regenerator does not generate useful cooling \( (q_c) \) or heating energy \( (q_h) \) since a portion of the energy generated by the eCE is spent for establishing the temperature span. In general, the active regenerative cycle is similar to the cascade system, where several thermodynamic cycles (connected in series) are used to increase the temperature span. As explained in [8], the major difference between a cascade cycle and an active regenerative one is in the fact that in the latter, the heat is not pumped directly between the neighbouring particles and instead, all the particles, which are indirectly coupled through the heat-transfer fluid, simultaneously accept or reject the heat to the heat-transfer fluid.

**Fig. 2.** A schematic representation of four basic operational steps of: a) a single-stage elastocaloric device with a solid-to-solid contact heat transfer mechanism, and b) a regenerative elastocaloric device with a convective (through porous material) heat transfer mechanism [51] and temperature (T)-entropy (s) diagrams: of c) a single-stage, and d) a regenerative elastocaloric Brayton thermodynamic cycle
In the recent years, researchers from Saarland University [36] and Karlsruhe Institute of Technology [39] and [40] have presented several different single-stage cooling/heat-pumping systems that operate based on a single (or double) elastocaloric foil that is loaded in tension and moves between the heat sink and the heat source and transfers the heat by contacting them directly. The maximum temperature span measured on this type of device was up to 14 K [39]. Such a concept was suggested to be applied in novel microelectromechanical systems (MEMS) such as temperature control of lab-on-chip systems or micro cooling in biomedical applications. Very recently, a 3-stage cascade elastocaloric device with a solid-to-solid contact heat transfer mechanism was developed and tested [45]. The authors demonstrated that, compared with a single-stage system, a 3-stage cascade system can increase the temperature span by a factor of two and reach a maximum temperature span of 15 K. On the other hand, the specific cooling power and the COP was reduced by almost a factor of 3 compared to a single-stage operation. In 2015, Qian et al. [35] presented an elastocaloric cooling system with Ni-Ti tubes loaded in compression, through which a heat-transfer fluid was pumped. The system utilizes the so-called heat recovery principle by using a bypass between the elastocaloric structure (tubes) and the external heat exchangers that stores a part of the energy produced by the eCE and then preheats or precools the heat transfer fluid with an aim of increasing the generated temperature span. The maximum measured temperature span of this device was 4.7 K and the maximum specific cooling power was 600 W per kg of the elastocaloric material. Even better results were obtained using active regenerative principle as shown by Tušek et al. [37] and Engelbrecht et al. [38]. This device, which was based on a set of thin Ni-Ti plates under tension, operated as a heat-pump and generated a maximum measured temperature span of 20 K, a corresponding specific heating power of up to 800 W per kg of the elastocaloric material, a maximum COP value of up to 7 and an exergy efficiency of 19 % (at a temperature span of 13 K). In 2018, Kirsch et al. [41] and [42] presented a new rotary elastocaloric device based on a set of Ni-Ti wires loaded in tension and air as the heat transfer medium. This device can be considered as the first pre-industrial prototype of an elastocaloric device, but full experimental results are not yet available. Recently, also a cascade elastocaloric cooling system with Ni-Ti tubes loaded in compression was proposed [43].

In order to increase the convective heat transfer (similar to a heat-pipe concept). Unfortunately, no cooling characteristics other than the temperature span of the device are published to date. In 2019, another three-stage cascade elastocaloric device with a single Ni-Ti wire loaded in tension was designed, built and tested [44]. A maximum temperature span of 28.3 K was measured, which is currently the largest temperature span measured for elastocaloric devices. Table 1 shows the overview and comparison of the developed proof-of-the-concept elastocaloric devices to date. Maximum temperature span, maximum specific cooling/heating power per mass of the elastocaloric material and the maximum COP and exergy efficiency are presented in Table 1 (if available). The COP is defined as the ratio between the cooling/heating power and the input work, while exergy efficiency is defined as the ratio between the actual COP and the maximum theoretical COP (also called the Carnot COP) that would be possible over a certain temperature span. It should be noted that all the COP values published for elastocaloric devices are based on the assumptions that a perfect recovery of the released mechanical energy is achieved during the unloading and that the input work is only related to the area enclosed in the stress-strain diagram (hysteresis) – see Section 3.2. for more details on mechanical work recovery.

2 DEVELOPMENT OF SUPERELASTIC POROUS STRUCTURES (ACTIVE ELASTOCALORIC REGENERATORS)

The focus of this section is on the development of elastocaloric (superelastic) porous structures that will serve as active elastocaloric regenerators. The crucial challenges in designing efficient and fatigue-resistant active elastocaloric regenerators are addressed and discussed.

2.1 Potential Geometries of an Active elastocaloric regenerator

Active elastocaloric regenerators can be made from different basic shapes of SMAs such as wires, rods, plates, tubes and blocks. Using these shapes, it is possible to make parallel-plate (Fig. 3a) [37] and [38] or wire-based (Fig. 3b) [41] and [42] regenerator for tensile loading. Wires are particularly challenging as they are not in a dog-bone shape and the stress concentration at the grip section usually results in their premature failure. On the other hand, potential geometries for compressive loading can be made from tubes with supporting elements placed along their length (Fig. 3c) or as a set of tubes in a holder (Fig. 3d).
that prevents the tubes from buckling. The major
difference between these two geometries is the flow
of the heat-transfer fluid. In the case of the geometry
shown in Fig. 3c, the heat-transfer fluid flows in the
cross-flow regime, thus, it does not flow inside the
tubes as is the case of the geometry shown in Fig. 3d.
This is an important feature for practical applications.
As explained below in this section, it is desirable that
the fluid flow is in the loading direction of the active
elastocaloric regenerator. Therefore, in contrast to the
case of the regenerator shown in Fig. 3d, in which the
heat-transfer fluid enters from the top and through the
bulky loading grips, the heat-transfer fluid can enter
from the side in the case of the regenerator shown in
Fig. 3c. In this way, the dead volume (the volume
between the regenerator and the heat sink/source) and
the heat losses to the grips as well as the complexity
of the grips can be reduced. Another potential porous
geometry for compressive loading can be made of
blocks with holes (Fig. 3e), but since Ni-Ti-based
materials (as the most promising elastocaloric
materials) are hard-to-machine materials, this can be

<table>
<thead>
<tr>
<th>Institution</th>
<th>Concept/ principle</th>
<th>eCM</th>
<th>Max. temp. span</th>
<th>Max. specific cooling/heating power</th>
<th>Max. COP / Max. exergy eff.</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Univ. of Maryland (2015)</td>
<td>Heat recovery</td>
<td>Ni-Ti tubes in compression</td>
<td>4.7 K</td>
<td>600 W/kg (at $\Delta T_{\text{span}}=0$ K)</td>
<td>n/a</td>
<td>[35]</td>
</tr>
<tr>
<td>Saarland Univ. (2015)</td>
<td>Single-stage</td>
<td>Ni-Ti sheet in tension</td>
<td>7 K</td>
<td>n/a</td>
<td>n/a</td>
<td>[36]</td>
</tr>
<tr>
<td>Technical Univ. of Denmark and Univ. of Ljubljana (2016, 2017)</td>
<td>Active elastocaloric regeneration</td>
<td>Ni-Ti sheets in tension</td>
<td>20 K</td>
<td>800 W/kg (at $\Delta T_{\text{span}}=15$ K)</td>
<td>7 (at $\Delta T_{\text{span}}=5$ K) / 19 % (at $\Delta T_{\text{span}}=13$ K)</td>
<td>[37] and [38]</td>
</tr>
<tr>
<td>Karlsruhe Inst. of Technology (2017)</td>
<td>Single-stage</td>
<td>Ni-Ti-Fe and Ti-Ni-Cu-Co sheets in tension</td>
<td>14 K</td>
<td>n/a</td>
<td>6.7 (at $\Delta T_{\text{span}}=6.3$ K) / 15 % (at $\Delta T_{\text{span}}=10.5$ K)</td>
<td>[39] and [40]</td>
</tr>
<tr>
<td>Saarland Univ. (2018)</td>
<td>Active elastocaloric regeneration</td>
<td>Ni-Ti wires in tension</td>
<td>9 K</td>
<td>n/a</td>
<td>n/a</td>
<td>[41] and [42]</td>
</tr>
<tr>
<td>Fraunhofer IPM (2018)</td>
<td>Cascade (heat pipe – like)</td>
<td>Ni-Ti tubes in compression</td>
<td>10 K</td>
<td>n/a</td>
<td>n/a</td>
<td>[43]</td>
</tr>
<tr>
<td>Cornell Univ. (2019)</td>
<td>3-stage cascade</td>
<td>Ni-Ti wires in tension</td>
<td>28.3 K</td>
<td>n/a</td>
<td>7.7 (at $\Delta T_{\text{span}}=5.2$ K) / n/a</td>
<td>[44]</td>
</tr>
<tr>
<td>Karlsruhe Inst. of Technology (2019)</td>
<td>3-stage cascade</td>
<td>Ti-Ni-Cu-Co sheets in tension</td>
<td>15 K</td>
<td>n/a</td>
<td>1.9 (at $\Delta T_{\text{span}}=5$ K) / n/a</td>
<td>[45]</td>
</tr>
</tbody>
</table>

**Fig. 3. Potential geometries of an active elastocaloric regenerator**
a very time consuming and expensive solution. Fig. 3f shows a honeycomb geometry made from corrugated plates with an enhanced buckling stability and good heat transfer properties [52]. However, the geometry made of block with holes and the honeycomb geometry also require fluid flow inlet/outlet through the grips, which is their disadvantage (as already mentioned above).

Using conventional methods to fabricate complex geometries and porous (or hollow) structures, such as the ones shown in Fig. 3, could be challenging since multiple fine elements should be assembled and joined together. Such elements (structures) might have been manufactured by different methods and consequently have different properties. Furthermore, it is difficult to machine and weld Ni-Ti [53] and [54] and the large number of required joints (and interfaces) along with the probable variations in their quality and properties can weaken the structure and reduce its performance and fatigue life. Therefore, methods such as additive manufacturing could be adopted to fabricate a uniform jointless active elastocaloric regenerator. Recently, additive manufacturing, in particular selective laser sintering (SLS) and selective laser melting (SLM), have become potential alternative methods for making porous superelastic implants for biomedical applications [55]. Fabrication of Ni-Ti using different additive manufacturing techniques, especially SLM [56], has been the subject of multiple studies [57]. Achieving superelasticity in Ni-Ti samples that are fabricated using SLM, usually requires a post-process heat treatment, nevertheless, a recent study has demonstrated the possibility of achieving superelasticity in as-fabricated SLM samples by adjusting the parameters of the process [58]. Thermomechanical response of additively manufactured Ni-Ti parts, including their fatigue life as well as their elastocaloric response, has been investigated and some promising results have been achieved. Nonetheless, there is still room for further improvements by investigating new techniques, chemical compositions, process parameters, heat treatments etc. [57] and [59] to [62].

Another important factor in designing active elastocaloric regenerators is the aspect ratio (cross section to length). Since relatively high stresses (up to 1000 MPa) are required to induce the martensitic transformation, the cross sectional area needs to be as small as possible to reduce the required force. This further means that in order to keep a constant mass of the elastocaloric material (which is directly related to the generated cooling/heating power), the regenerator’s length needs to be increased. In addition, as shown previously for active magnetic regenerators made of ordered structures with straight fluid flow channels [63] (e.g. parallel-plate, set of tubes, etc.), it is desirable that the active elastocaloric regenerator is relatively long (above 50 mm) in order to increase the length of the path along which a temperature span can be established. Therefore, the heat-transfer fluid should flow along the length and the loading direction of the active elastocaloric regenerator.

Furthermore, under the steady-state conditions a temperature profile is established along the regenerator in the fluid flow direction. As each part of the elastocaloric material along the regenerator works at a different temperature (see Fig. 2b) and in accordance with the dependence of transformation strain on the temperature of the SMA [21], that further means that the coldest part of the regenerator would be strained much more than its hottest part and the eCE in the coldest part of the regenerator would be larger compared to its hottest part. In order to overcome this issue and ensure an equal eCE along the regenerator under the steady-state conditions, a layered active elastocaloric regenerator [64] should be applied (analogous to a layered active magnetic regenerator [65] and [66]). That means that different elastocaloric materials with different transformation temperatures should be stacked along the length of the regenerator (in the fluid flow direction).

### 2.2 Tension vs. Compression

In general, the active elastocaloric regenerator can be loaded in tension or compression, each of which has some pros and cons concerning the elastocaloric cooling. Although very interesting, other loading modes such as bending [67] or torsion [68] will not be considered here since they cannot provide a homogenous strain distribution over the porous structure. Homogenous strain distribution that results in homogenous distribution of the eCE is the first precondition of an efficient active elastocaloric regenerator. The main advantage of tensile loading is the possibility of utilizing thin elements with high specific heat transfer area per unit volume and small hydraulic diameters that result in good heat transfer properties, while its main disadvantage is a reduced fatigue life compared to compressive loading. It is known [69] that any cracks on the surface or impurities in the material, which are practically inevitable, will grow and expand during tensile loading and eventually lead to the failure of the material (see Fig. 4). Nevertheless, under compressive loading, those cracks and impurities do not have
much space to grow. This can significantly prolong the fatigue life of the material, which is an important advantage of compressive loading. On the other hand, thin elements, which would enhance the heat transfer characteristics of an active elastocaloric regenerator, cannot be compressed without solving the buckling stability problem associated with them. It is, therefore, challenging to design a porous geometry that can be compressed in a stable manner and at the same time assure an efficient and rapid heat transfer.

Fig. 4. Schematic representation of crack growth during tension and compression [69]

Fig. 5a shows the stress-strain behaviour of a superelastic Ni-Ti alloy under isothermal (strain rate of 0.0002 s⁻¹) and adiabatic (strain rate of 0.06 s⁻¹) conditions where tension-compression asymmetry [18] is clearly observable. Dog-bone shaped thin sheets (with a thickness of 250 μm) were tested under tension [70], while thin-walled tubes (with an outer diameter of 3 mm and a wall thickness of 250 μm) were tested under compression. In order to stabilize the stress-strain behaviour of the samples before the experiments, both samples were subjected to a mechanical stabilization (training), which consisted of 100 superelastic cycles at a strain-rate of 0.002 s⁻¹ up to the stresses/strains that corresponded to the end of the transformation plateau. From Fig. 5a, it is evident that in compressive loading, the slope of the transformation plateau is significantly steeper compared to the tensile loading, which means that a higher stress is required to complete the transformation, while the resulting transformation strain is smaller. Fig. 5b shows positive and negative adiabatic temperature changes for tensile and compressive loadings as functions of the applied strain and stress.

Fig. 5. a) Stress-strain response of Ni-Ti tubes and dog-bone shaped sheets under compression and tension, and b) the corresponding adiabatic temperature changes as functions of the applied stress and strain

From the adiabatic temperature change-strain projection of Fig. 5b, one can see that larger adiabatic temperature changes can be obtained in compression, where (positive) adiabatic temperature changes of up to 28 K were measured. The reason for this could be related to the smaller transformation strain in compression compared to tension, which means that a more complete transformation (with a larger
transformation latent heat) is obtained at smaller strains under compression. This could be clearly observed in the stress-strain projection of Fig. 5b. It is also seen that the typical (almost flat) transformation plateau region that is observable in tension, turns into a strain-hardening region in compression. As a consequence, the transformation start stress and the stress that is needed to complete the transformation under tension are much smaller than under compression. Therefore, as seen from the adiabatic temperature change-stress projections of Fig. 5b, in tension, larger adiabatic temperature changes can be obtained at small and moderate stresses compared to compression. This is where tension is superior to compression. It is further evident from Fig. 5b that positive adiabatic temperature changes (during loading) are in all cases larger than negative ones (during unloading), which is due to the intrinsic mechanical dissipative heat of the internal friction that manifests itself as the hysteresis.

Fig. 6 shows the distribution of the temperature changes over the sample during loading and unloading Ni-Ti thin-walled tubes under compression and Ni-Ti dog-bone shaped thin sheets under tension. One can see that in compression, the temperature change in the sample is fully homogenous (neglecting heat losses through the grips), while in tension, temperature bands related to the Lüders-type transformation [71] are clearly visible, which is in agreement with the presence of strain-hardening region and flat plateau under compressive and tensile loading respectively. With regard to the eCE, the homogenous temperature change is an advantage of compressive loading.

Fig. 6. Distribution of temperature changes [°C] over the elastocaloric sample during tensile; a) loading, and b) unloading under a strain of 4.6 %, and during compressive c) loading, and d) unloading under a strain of 3.8 %

2.3 Fatigue Life of Elastocaloric Materials

As already mentioned above, the major difference between tensile and compressive loading is the fatigue life. It is estimated that a cooling or heat-pumping device should withstand more than 10 million loading/unloading cycles with no functional and structural fatigue in its lifetime (10 years). Fig. 7 shows the adiabatic temperature changes of Ni-Ti that correspond to different strains/stresses as functions of fatigue life (number of cycles until failure or runout) obtained from dog-bone shaped thin sheets under tension [70] and thin-walled tubes under compression. As shown in Fig. 7, in tensile loading, high-cycle fatigue regime (>10⁶ cycles) can be reached only by using small strains (stresses) that do not correspond to a completed phase transformation and thus can only result in limited adiabatic temperature changes of below 8 K. It should be noted that, in most cases (for tensile loading), Ni-Ti samples withstand 10⁸ fatigue cycles have also reached 10⁹ or even 10¹⁰ cycles as shown in [72]. Furthermore, Chluba et al. [25] and [73] developed an ultra-low fatigue Ni-Ti-Cu-Co thin film by sputtering deposition that could withstand 10 million loading cycles under tension with strains of up to 2.50 % and adiabatic temperature changes of up to 10 K. The results are highly encouraging, but the disadvantage is that the applied sputtering deposition technique is still too time-consuming for large-scale production of this material that would be required in practical cooling or heating applications. On the other hand, our preliminary results on thin-walled tubes as well as other authors’ results on bulk blocks or cylinders [74] and [75] show that it is possible to reach well above 1 million loading and unloading cycles under compression with strain (stress) amplitudes

Fig. 7. Adiabatic temperature changes of Ni-Ti alloy as functions of fatigue-life under tension and compression at different applied strains/stresses
high enough to cover the entire transformation plateau (note that in compression, transformation strain of a fully transformed sample is smaller than that of the tension). This can result in stable adiabatic temperature changes of up to around 25 K even after 1 million cycles as shown in Fig. 7. Accordingly, compressive loading shows a very high potential for designing fatigue-resistant elastocaloric cooling systems capable of reaching up to 10 million operation cycles. However, a precondition for reaching a long fatigue life in compression is the buckling stability of the specimen.

3 DEVELOPMENT OF THE DRIVER MECHANISM

In addition to the active elastocaloric regenerator, another key component of the elastocaloric technology is the driver mechanism that provides the elastocaloric material with the mechanical loading. There are two crucial requirements that present a challenge in developing a driver mechanism. The first requirement is providing the elastocaloric material with a sufficiently high mechanical load since phase transformation occurs at relatively high stresses. The second requirement is to ensure that the work that is released during unloading the material is effectively exploited (work recovery). The efficient use of the recoverable work helps to increase the efficiency of the driver mechanism and the entire elastocaloric device.

3.1 Review of Different Driver Concepts Applied in Elastocaloric Technology

The prototypes and proof-of-the-concept elastocaloric devices developed to date are based on a hydraulic actuator [35], linear actuators with a spindle motor [36], [37] and [44] and a rotary driver using a cam-disk [41]. Different potential driver mechanisms for elastocaloric technology are presented and reviewed also in [76]. In general, one can distinguish between linear driver mechanisms (such as pneumatic or hydraulic actuators, linear motors, piezo actuators, SMA actuators and magnetostriction actuators) and rotary driver mechanisms (where using an electromotor seems to be the most reasonable option). The linear drivers are mostly suitable for loading an individual regenerator or a pair of regenerators operating simultaneously. On the other hand, by utilizing rotary drivers, several regenerators distributed in a ring can be loaded and the whole system can be more compact.

Pneumatic actuators usually operate with relatively small working pressures of up to 10 bar and therefore the high forces, which are required to induce and complete the phase transformation in the elastocaloric material, can only be generated using small contact surfaces. The hydraulic actuators on the other hand, can provide higher pressures (up to 300 bar), but are only useful for laboratory testing owing to their poor overall efficiency. Linear motors allow of a direct tensile or compressive loading with relatively high efficiencies owing to the absence of transmissions, but have difficulty in providing the required high forces [76]. Solid-state actuators, utilizing ferroic properties of functional materials, could be interesting alternatives. Among them, the most widely applied are piezo actuators [77], but their small obtainable displacements (below 0.1 mm) are not enough to generate large eCE in an active elastocaloric regenerator. In 2018, Huo et al. [78] applied a magnetostriction actuator (Terfenol-D), which could generate strains of up to 0.2 % to load/unload a Cu–Al–Mn elastocaloric alloy. In 2019, Qian et al. [79] proposed a linear heat-driven SMA actuator to load and unload the elastocaloric materials. It can generate the required stresses and strains, but its drawback is its relatively low operating frequency.

Most actuators that operate continuously and with a relatively high efficiency are based on a rotary motion (e.g. car engine, electric motor, existing vapour compression cooling systems, etc.). Rotary drivers in combination with a camshaft, a cam-disc or a cam-ring, are therefore promising solutions for designing compact elastocaloric systems. The active elastocaloric regenerators can be positioned radially or axially with respect to the direction of rotation. The ability to adjust the shape of the cam-disk or cam-ring and the fluid flow regime of the heat transfer medium can ensure adiabatic, isothermal or combined loading of the elastocaloric material [41].

Another interesting and promising solution is a hybrid system [80], which is based on exploiting the pressure difference between the evaporation and condensation phases in a classic vapour compression refrigeration system. Such a pressure difference could be used to run an actuator that would further load and unload the elastocaloric material. Since elastocaloric systems require high loads, combining them with a CO₂ vapour compression system (owing to the high pressure difference between their evaporator and condenser) could result in the most appropriate hybrid system. A hybrid system could be utilized to design a vapour compression system with a higher efficiency.
3.2 Recovery of the Released Mechanical Work

The first step in assessing the viability of work recovery is a theoretical analysis of the amount of recoverable work available in an active elastocaloric regenerator (material). Such an analysis was performed on the two extremes of loading rate i.e. adiabatic and isothermal loading conditions (data from Fig. 5). Fig. 8 shows an example of force variations during phase-shifted tensile loading and unloading of two regenerators under adiabatic and isothermal conditions, where the normalized force ($F_{\text{norm}}$) is taken into account. It is assumed that during the first half of the cycle, one regenerator is loaded (blue curves) and another one is simultaneously unloaded (red curves), while in the second half of the cycle, the opposite happens. Such an operation in an elastocaloric device was demonstrated in [35] and [40]. In Fig. 8, the area between the loading curve and the x-axis represents the invested or required work ($F_{\text{norm}}$ above 0), while the area between the unloading curve and the x-axis represents the available or recoverable work ($F_{\text{norm}}$ below 0). The difference between the loading and unloading forces (the net force) is shown with the green curves in Fig. 8. One can see that at the very beginning of the cycle, the force released during the unloading is higher than the required force for loading (net force is negative). If both active elastocaloric regenerators are directly connected to each other (without utilizing the natural frequency, a flywheel or a similar system to store the kinetic energy) the negative net force cannot be recovered since the regenerator that is being loaded at that time, requires a much smaller force than can be provided by the regenerator being unloaded. Under such circumstances, only 78 % and 77 % of the available work can be recovered under isothermal and adiabatic loading conditions respectively. That means that, in order to overcome the hysteresis losses and the lost negative net force, the driver needs to provide 73 % and 47 % of the required invested work in the case of each adiabatic and isothermal loading cycle respectively. However, by introducing a flywheel, which is a mechanical device specifically designed to efficiently store rotational energy (kinetic energy), the negative net force can also be recovered. By using a flywheel, the input work, which is provided by the driver in each loading cycle, is only related to the hysteresis losses that are 65 % and 32 % of the required invested work in the cases of adiabatic and isothermal loadings respectively and the rest of the required invested work can be recovered. This is the main reason that isothermal loading can result in a more efficient operation of an elastocaloric device as already shown in the literature [50].

Another way to recover the negative net force is utilization of natural frequency work recovery systems. Such systems make it possible to recover the mechanical unloading work by storing it in an oscillating mass that oscillates within the natural frequency range of the system. Nevertheless, the natural frequency of a typical active elastocaloric regenerator [37] is in the kHz range, which is significantly higher than the operating frequency required for an efficient elastocaloric cooling (up to 10 Hz). In general, the natural frequency could be reduced by methods such as using an additional spring system in order to decrease the stiffness of the system.

5 CONCLUSIONS

In conclusion, the state-of-the-art and future challenges of elastocaloric cooling technology, as one of the most promising alternatives to vapour compression cooling technology, are reviewed. Active elastocaloric regenerators show a promising performance regarding the generated temperature span, specific cooling/heating power and COP. Adding a work recovery mechanism to such systems can decrease the required input work and improve the overall efficiency of the system.

Fatigue life of the elastocaloric materials is a challenging issue concerning the commercialization of elastocaloric cooling devices. This issue can be tackled by designing new material systems, utilizing new fabrication methods and developing new devices that operate under compressive loading. In order to define the optimal geometry of an active elastocaloric regenerator under compressive loading, buckling
stability and heat transfer properties should be considered. To assure a fast and efficient heat transfer between an elastocaloric material and a heat-transfer fluid, thin-walled structures with small hydraulic diameters and therefore high specific heat transfer areas are required. Nevertheless, such geometries are, in general, not favourable from the buckling stability point of view. Consequently, a compromise between these two opposing requirements needs to be reached.

6 ACKNOWLEDGEMENTS

The work was supported by European research Council (ERC) under Horizon 2020 research and innovation program (ERC Starting Grant No. 803669).

7 REFERENCES


Elastocaloric Cooling: State-of-the-art and Future Challenges in Designing Regenerative Elastocaloric Devices


