

Phase Change Nanoelectromechanical Relay for Nonvolatile Low Leakage Switching

James T. Best, Mohammad Ayaz Masud,* Maarten P. de Boer,* and Gianluca Piazza*

The design, modeling, and experimental validation of a highly scalable phase change electromechanical relay are presented. The Phase Change NEMS Relay (PCNR) is a nonvolatile mechanical relay actuated by the volumetric expansion of phase change material. GeTe is used as the active phase change material, and nonvolatile relay states are changed by converting it between amorphous and crystalline phases, which differ in volume by 10%. Phase conversion is induced by Joule heating an adjacent metal layer. Finite element analysis (FEA) models are developed to predict actuator temperature distributions and quench times for varied actuation pulses. Additional models are developed to simulate actuator deflection including reflow of molten phase change material during an actuation pulse. The smallest fabricated device has a heater footprint of $3\ \mu\text{m} \times 1\ \mu\text{m}$, with a 20 nm air-gap. The FEA model, which calculates the actuation voltage, energy, and ON–OFF current ratio for this device to be 1.1 V, 42 nJ, and 10^8 , respectively, are experimentally verified. A PCNR scaled down to a $15\ \text{nm} \times 5\ \text{nm}$ heater footprint with a 2 nm air-gap is predicted to actuate at 400 mV and 1.7 pJ.

number of devices per area, but does not improve the power efficiency of an individual device. The advent of the internet of things has generated increased demand for ultralow power systems. A number of emerging technologies are being developed as ways to augment or replace CMOS for various computing applications. Some of these technologies include: resistive memory,^[3,4] phase change memory,^[5,6] magnetoresistive memory,^[7,8] and ferroelectric field effect transistors.^[9,10] A key feature common to these technologies is nonvolatility,^[11] which is the ability of a device to hold its state without a power supply. These emerging memories suffer from their limited difference in resistance between the ON and OFF states (ON–OFF ratio). A low ON–OFF ratio poses a great challenge toward cascading long arrays of devices in memory intensive applications because the sensing margin is poor during

the read operation. Furthermore, the nonzero OFF current leads to high leakage, resulting in greater static power consumption.

MEMS-based mechanical relays, another emerging technology, have a large ON–OFF ratio due to the near zero leakage off state associated with an open mechanical contact.^[12] MEMS relays have been demonstrated as nonvolatile devices as well,^[13] but a major bottleneck to the adoption of mechanical relays is their scalability to cell sizes that are comparable to state-of-the-art CMOS devices. Common MEMS relays use a flexure,^[14] which limits scalability due to material limits and anchor sizes. We earlier reported a $11\ \mu\text{m} \times 1.5\ \mu\text{m}$ nonvolatile mechanical actuator based on the volumetric expansion of GeTe phase change material.^[15] This reversible volume expansion can be used to directly close an air-gap, eliminating the need for flexures and making for an inherently highly scalable device. A prototype nonvolatile device was then presented.^[16] Here we report the detailed design and demonstration of a smaller device ($3\ \mu\text{m} \times 1\ \mu\text{m}$) and a family of nonvolatile relays. These results demonstrate a path to scaling a MEMS relay with smaller area and lower voltage than a future foreseeable CMOS transistor.

1. Introduction

Over the last five decades, complementary metal oxide semiconductors (CMOS) have dominated all aspects of computing from logic to memory by aggressively reducing transistor size. The associated advances have led to the complex, high speed, and efficient computers of today.^[1] With each process node, such scaling becomes more demanding and is expected to reach its limit around the year 2028.^[2] Further increase in CMOS functionality will be achieved through 3D integration, where devices are fabricated on more than one layer. This increases the

J. T. Best, M. A. Masud, G. Piazza
Electrical and Computer Engineering
Carnegie Mellon University
Pittsburgh, PA 15213, USA
E-mail: mmasud@andrew.cmu.edu; piazza@ece.cmu.edu

M. P. de Boer
Mechanical Engineering
Carnegie Mellon University
Pittsburgh, PA 15213, USA
E-mail: mpdebo@andrew.cmu.edu

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/aelm.202200085>.

© 2022 The Authors. Advanced Electronic Materials published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

DOI: 10.1002/aelm.202200085

2. Device Operation and Results

2.1. Operating Principle

The Phase Change NEMS relay (PCNR) is designed to address the limited scalability of existing MEMS relay designs while continuing to use mechanical contacts for a high ON–OFF ratio and while maintaining nonvolatile states. The PCNR takes

advantage of the mechanical expansion of phase change material (PCM), which occurs when converting between the amorphous and crystalline states (see **Figure 1**). **Figure 1a** shows a schematic and **Figure 1b** shows a scanning electron microscope (SEM) image of a fabricated device. The PCNR consists of two main components, the contact pair and the PCM actuator. The contact pair is formed by a pair of suspended drain and source contacts and a metal bonded to the top surface of the actuator. The metallic contact lies on the top surface of the actuator (**Figure 1a,b**) and acts as a channel to bridge the drain and source when in the ON state. The actuator is comprised of a resistive heater, a layer of PCM, and an encapsulation for the PCM (**Figure 1d**). The actuator expands and contracts as the PCM is converted between the crystalline (smaller) and amorphous (larger) states. Although other phase change materials are candidates for the PCNR actuator, GeTe is chosen due to its documented 10% volume change,^[17] and its ability to retain its

amorphous state at higher temperature than most commonly used PCM, such as Ge₂Sb₂Te₅ (GST).^[18,19] All metal layers consist of tungsten. With GeTe's high melting temperature (T_m^{GeTe}), tungsten endures the actuator's high temperature operation. The encapsulation layers of the PCM are formed by Al₂O₃, which is deposited conformally by atomic layer deposition.

Figure 1d,e demonstrates the switching mechanism of a PCNR in a schematic cross-section. The PCNR is fabricated in the OFF state with the PCM in the crystalline (smaller volume) state, which allows for the air-gap to be set by sacrificial material (air-gap region in **Figure 1d**). The gap is sized to ensure that the expected thickness change in the PCM layer is sufficient to make contact between the source and drain. A current pulse is applied to the heater to switch to the ON state, which in turn melts a portion of the PCM. After the pulse, the PCM cools and quenches into the amorphous state. A quench is considered sufficiently fast when the time for the PCM to cool from T_m^{GeTe}

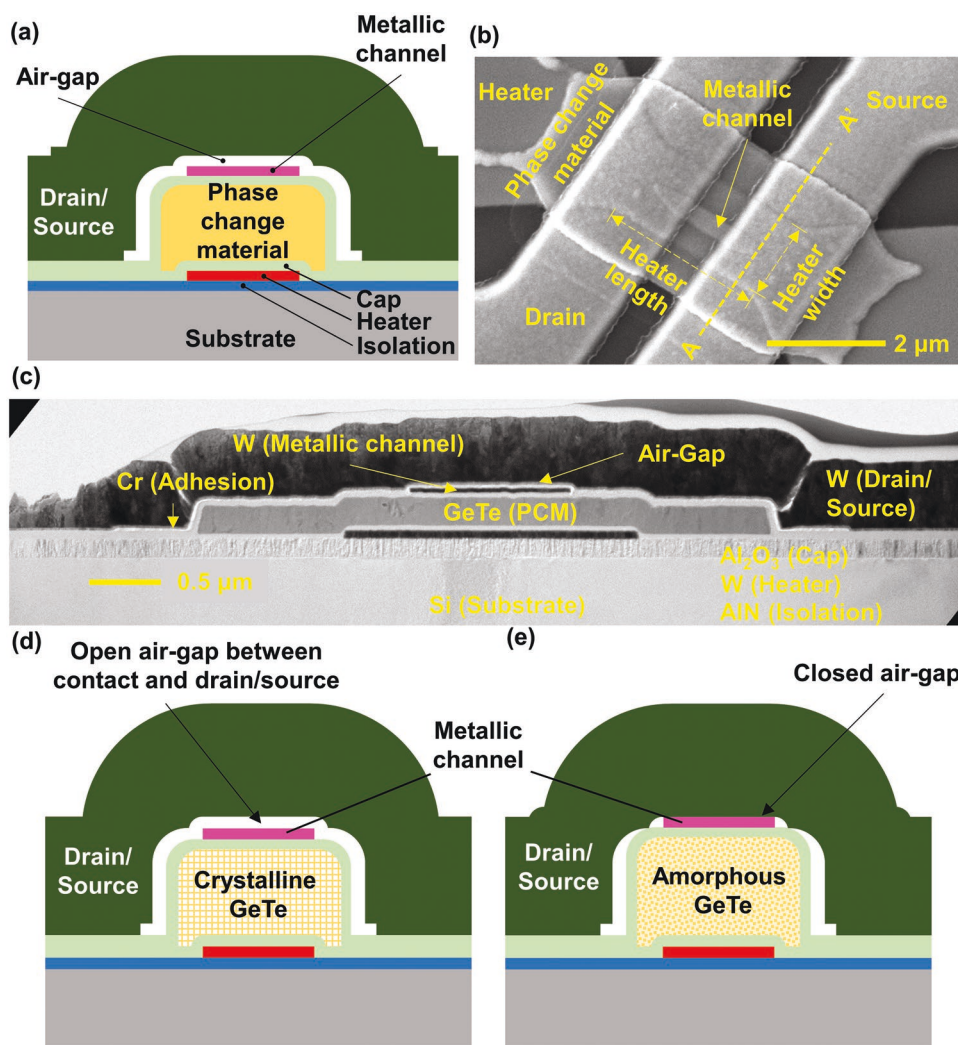


Figure 1. Phase Change NEMS Relay concept showing: a) a schematic cartoon of the device. Two major components of the relay are the actuator and the metal contact pair. The metallic channel connects the drain and source in the actuated state. b) A labeled SEM top-view of a fabricated PCNR. c) A TEM image of the A-A' cross section shows the architecture and the materials. d,e) Schematic cross-section before and after actuation, respectively. GeTe is in the crystalline, high-density state in the as-fabricated device. (d) An actuation pulse turns GeTe into amorphous state, which closes the air-gap as depicted in (e). Phase transformation occurs when appropriate current pulse is applied on the heater.

(1000 K for GeTe) to the glass transition temperature (500 K) is less than 100 ns.^[20] If cooled at a slower rate, the PCM will crystallize and the relay will not switch. The expanded PCM in the actuator presses the metallic contact into the drain and source, bridging the two contacts (Figure 1e).

A similar process returns the PCNR to the OFF state. Another pulse heats the PCM and the rise in PCM temperature facilitates rapid crystallization of any amorphous portion. Unlike the switching ON process, the PCM need not be melted to turn the relay OFF. Hence, we apply a lower power pulse for recrystallization. Also, a longer pulse is preferred as it ensures complete recrystallization of the PCM. As a result, the crystalline PCM contracts and separates the metallic contact from the drain and the source, once again creating an air-gap, turning the relay OFF. The switch ON and OFF pulses are also called actuation and recrystallization pulse, respectively.

2.2. Finite Element Modeling

The primary goal for modeling the PCNR actuator is to predict the temperature distribution and expansion in the actuator for a given actuation pulse. The thermal model also estimates quench time, which is a critical parameter for successful actuation. We develop two separate models of the PCNR in COMSOL Multiphysics finite element analysis (FEA) software—one combines the actuator electrical and thermal physics in a 3D electro-thermal model, while the other simulates mechanical actuation in a 2D cross-section. The two separate models greatly reduce computational complexity compared to a unified setup. Details of the simulation setup and material properties are discussed in the Experimental Section and in the Supporting Information.

Figure 2 demonstrates the simulation setup and results extracted from COMSOL. Figure 2a shows PCNR geometry and

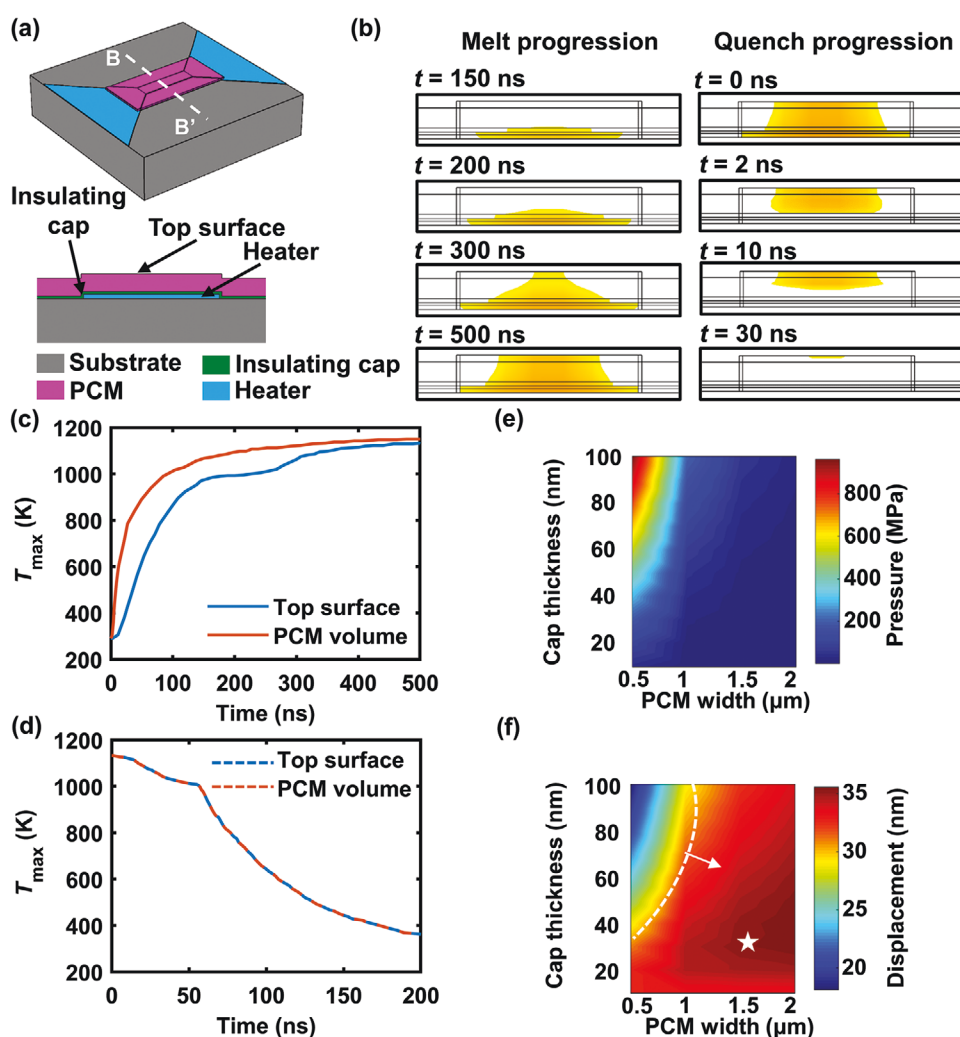


Figure 2. a) A representative PCNR device for electrical and thermal COMSOL simulation. B-B' cross-section at the center of the heater includes the most contributing layers for thermal transport. b) Simulation results show the progression of molten GeTe during an actuation pulse and subsequent quench progression. t is set to 0 at the beginning of the quench. c,d) Maximum surface and volume temperatures in the GeTe during melt and quench, respectively. The fast quenching step is necessary for successful phase transformation. e,f) Mechanical COMSOL simulation results for 200 nm thick molten GeTe showing (e) the pressure and (f) the maximum vertical displacement generated by the molten GeTe as a function of cap thickness and molten GeTe width. The high displacement region (right side of the white dashed line) in (f) provides the acceptable combinations of PCM width and cap thickness for optimum displacement. The star represents the combination with maximum displacement.

corresponding cross-section used for modeling the melting and quenching progressions, respectively, in the actuator. Cross-sections in Figure 2b show molten sections of the PCM at various times during an actuation pulse from the OFF to the ON state. The plots highlight regions at temperatures above T_m^{GeTe} . Time for melt progression starts at the initiation of an input electrical pulse. By $t = 150$ ns, the PCM closest to the heater begins to melt as it has the lowest thermal resistance to the heater. The molten section of PCM progresses up and outward from the center of the heater. A sufficient actuation pulse will melt the PCM through the entire thickness as depicted at $t = 300$ ns and $t = 500$ ns. We turn off the input pulse at $t = 500$ ns and reset the clock to measure quench time. PCM solidifies from the bottom up as lower sections of PCM have lower thermal resistance to the substrate. Most of the PCM solidifies within first 30 ns. We plot the maximum temperature of the PCM and its top surface as a function of time in Figure 2c,d. The temperature versus time trend can be seen to flatten around T_m^{GeTe} due to latent heat of transformation. As expected, the top surface melts last but retains heat longer than any other part of the PCM. However, the entire PCM block, including the top surface, solidifies within the first 100 ns of quenching, resulting in amorphous GeTe.

Mechanical modeling is separate from the electrical and thermal modeling as the main goal is to determine the possible vertical deflection from a given geometry of the PCNR actuator. The actuator is designed to confine molten PCM within a cap layer to ensure that the volumetric expansion of GeTe leads to a vertical expansion. Proper estimation of this expansion is required to determine the air-gap thickness of the relay.

As evident from the electrothermal model, only the region of PCM directly over the heater reaches T_m^{GeTe} and is surrounded by solid PCM (Figure 2b, 500 ns). Hence, the molten PCM width is assumed to be the same as the heater width. We sweep this parameter from 500 nm (lower bound for optical lithography) to 2 μm to see its effect on the expansion of the actuator. By contrast, a thick encapsulation layer may be better suited for confining the liquid PCM, but will limit the expansion. We sweep the cap layer thickness from 10 to 100 nm. Figure 2e,f shows the resulting pressure and displacement, respectively; as a function of the cap thickness and molten PCM width for a 200 nm thick GeTe PCM layer. We find that the vertical displacement is slightly larger than 10% of the film thickness because reflow of the PCM in the liquid phase focuses the expansion to the center of the actuator. Pressure in the liquid GeTe and the displacement are relatively insensitive to the geometry until the cap layer becomes very stiff. As a result, most combinations of the two input parameters yield a displacement above 30 nm. Thick cap layers with narrow PCM widths create a significantly stiffer cap, as depicted in the top left corner of Figure 2f (left side of the white dashed line).

The encapsulated liquid PCM is comparable to a rectangular pressure vessel with two sides constrained by solid PCM and the bottom is constrained by the substrate. The encapsulation layer is deformed by the vertical expansion of the liquid PCM. The expansion in a typical pressure vessel is inversely proportional to the thickness of its wall. However, when the wall is substantially thin compared to the radius of the vessel, a geometric nonlinear effect reduces the overall expansion.^[21] This effect is

evident in Figure 2f as a peak for maximum displacement is obtained at a cap thickness around 30 nm for the device dimensions used in this simulation. For walls thicker than 30 nm, we observe lower displacement, as expected. When this layer is thinner than 30 nm, it is subjected to large deformation due to the 10% volume expansion of the PCM. Under the geometric nonlinear assumption, this large deformation significantly increases the stiffness of the deformed part of the cap layer. As a result, very thin cap layers produce less displacement than the 30 nm layer. It should be noted that this optimum cap thickness is strongly related to cap material properties, PCM thickness, and the width of the molten PCM region. For our current fabrication process, we use 30 nm Al_2O_3 as the cap layer for 200 nm thick PCM.

2.3. Scaling

The relays designed for initial proof of concept fabrication are large compared to the envisioned final form of the PCNR. Two scaling methods are developed to bring the PCNR to an equivalent CMOS scale device. The first is isotropic in which all dimensions (length, width, and thickness) are scaled equally. The isotropic scaling parameter (K_V) indicates how many times smaller all dimensions are compared to the nominal geometry. For example, an isotropic scaling parameter of 10 indicates all dimensions are ten times smaller. The second method is lateral where layer thicknesses are held constant and lateral dimensions (length and width) are scaled equally. The lateral scaling parameter (K_L) indicates how many times smaller all lateral dimensions are compared to the nominal geometry. Volumetric scaling factor stops at 10 as this reduces the air-gap to 2 nm. Smaller air-gaps are likely impractical due to field emission or tunneling leakage. Further lateral scaling can shrink the heater down to dimensions of 15 nm (length) $\times$ 5 nm (width) $\times$ 5 nm (thickness). Compared to the nominal geometry, area is reduced by a factor of 40 000, heater voltage is reduced by a factor of 3, quench time is reduced by a factor of 40, and actuation energy is reduced by four orders. At the smallest heater size, the PCNR will occupy an area and operate at an actuation voltage that are each lower than current CMOS technology nodes.^[1,2] Table S1 of the Supporting Information summarizes a nominal geometry, and its volumetric and lateral scaling trends.

2.4. Experimental Demonstration

To validate the device concept and the modeling approach and support the theory associated with device scaling, microscale prototypes are fabricated and tested (see the Experimental Section for detailed fabrication process). The FEA results provide a valuable guideline for selecting the amplitude and duration of the input actuation pulse, but are not expected to precisely represent fabricated relays. These parameters are then adjusted to actuate fabricated relays. The switching and contact characteristics of these devices are measured from experimental results, and are reported in Sections 2.4.1 and 2.4.2, respectively.

2.4.1. Characterization of the Switching Pulses

One switching cycle of PCNR involves an actuation pulse and a recrystallization pulse. With the FEA model as a starting point, the pulses are calibrated using PCNR actuators. These are devices fabricated without the drain and source contacts, which leave the actuation area visible for optical imaging. For actuation, we adjust the amplitude of the pulse until a dark spot appears in the actuator, a characteristic associated with converting the PCM from the crystalline to the amorphous state.^[15] A recrystallization pulse removes the dark spot as the material reverts to the crystalline phase. These calibrated pulses are then used as a starting point in the full relay where amplitude is increased if necessary to connect the drain and source contacts.

This process is then repeated for a range of heater dimensions. A simple analytical model for the switching voltage demonstrates how switching voltage varies with heater length and width. We compare this model with our FEA model and experimental results to assess the accuracy of our modeling scheme. The switching voltage, V_s on the heater is modeled as

$$V_s = \sqrt{\frac{R_H \Delta T}{R_{th}}} \quad (1)$$

where R_H is the heater resistance, ΔT is the rise in heater temperature required for phase change, and R_{th} is the thermal resistance of the PCM. For a 200 nm layer of PCM, R_{th} is dominated by the thermal boundary resistance,^[22] which is inversely proportional to the interface area between the heater and PCM.^[23] If the length of the heater is reduced by a factor of S ($S > 1$), R_{th} increases proportionally, which decreases switching voltage. Furthermore, R_H is inversely proportional to S . Since the required ΔT is the same for devices independent of scale, the switching voltage decreases by the same factor of S .

The heater acts as a voltage divider connected in series with the traces and the internal impedance of the pulse generator, and the voltage across the heater increases with temperature. Therefore, we report the voltage at the onset of the actuation pulse, and call it the starting actuation voltage. As shown in **Figure 3a**, experimental actuation voltage and simulation results are in good agreement and follow the trend calculated from Equation (1).

Switching measurements in the time domain are also important as they define the maximum operating speed of the relay as well as the propagation delay. Measurements are taken for a single relay cycling on and off five times. **Figure 3b** shows the actuation voltage on the heater, heater current density, and drain/source current for the smallest demonstrated device. The relay has a 1 μm wide and 3 μm long heater and is actuated

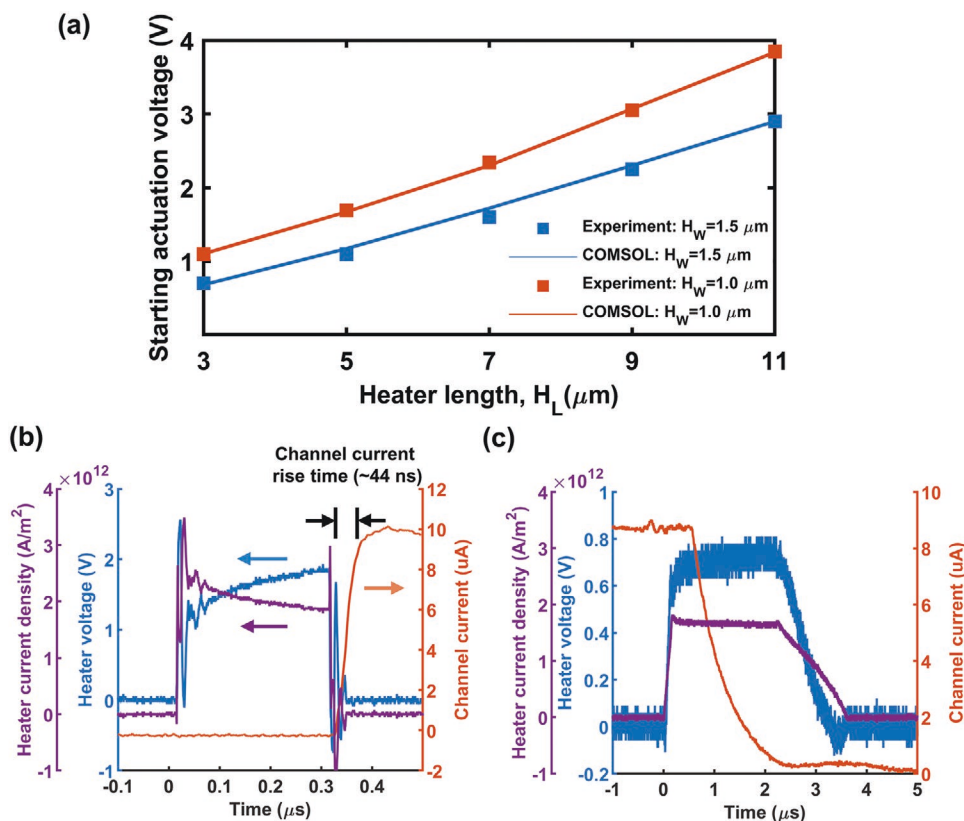


Figure 3. Experimental switching results. a) Comparison of experimental starting actuation voltage with FEA results. This plot shows good agreement between the experimental and FEA results for various heater dimensions. Corresponding heater dimensions (heater length, H_L and heater width, H_W) are marked in the SEM. b,c) Measurements of extracted heater voltage, heater current density, and channel current of a PCNR switching (b) ON and (c) OFF. The reported switching operation is performed on a relay with a 3 $\mu\text{m} \times 1 \mu\text{m}$ heater. Extracted heater voltage in (b) and (c) is the voltage drop only across the heater. Heater current density is the heater current normalized to the 1.0 μm wide and 50 nm thick heater cross-section.

by 300 ns wide pulses with 2 ns rise and fall times. Heater voltage (blue) increases throughout the actuation pulse duration. The temperature coefficient of resistivity of the tungsten heater, combined with the high operating temperatures, cause the resistance to rise from about 15 to 35 Ω . The fast rise and fall times of the actuation pulse excite ripples in the heater voltage and current due to reflections from imperfect impedance matching in the test setup between the pulse generator, oscilloscope, and relay. The heater current density (purple) is simultaneously plotted to check if the current density is too high to operate without causing heater failure. Current through the channel connecting the drain and source (orange) begins to rise toward the end of the actuation pulse. These switching pulses are chosen for “perfect” actuation, where the pulse is just the right combination of amplitude and width to actuate the device. “Perfect” actuation, by definition, turns the relay switch on as the pulse turns off. The channel current rise time observed does not necessarily represent the rate of contact resistance change. Rise time is limited by the rate at which the cables used in the test setup can be discharged through the relay contact resistance. The rise time of the channel current (44 ns) is significantly longer than the rise time of the input pulse (2 ns) and is attributed to a higher contact resistance, which can vary widely for a “perfect” actuation pulse.

Figure 3c shows heater voltage, heater current density, and drain/source current for the same device turning off. Recrystallization pulses are 2 μ s wide with 100 ns and 1 μ s rise and fall times, respectively. We first note the lack of large oscillations in the heater voltage and current. The significantly slower rise and fall times of the recrystallization pulse (blue) do not excite the reflections as in Figure 3b. The fall time in particular is chosen to be slow to minimize the cooling rate such that the amorphous state of the phase change material cannot exist in the event that any phase change material is inadvertently melted. Channel current (orange) begins to fall at 0.6 μ s.

Current falls slowly due to the large RC time constant associated with charging the test setup cable capacitance. Current flow is limited by a 100 k Ω resistor connected in series with the source, which increases the observed time constant. Unlike the actuation pulses, the recrystallization pulses continue after the channel current begins to change. The extended pulse time ensures a complete recrystallization of the PCM. The micro-scale PCNR has longer ON–OFF cycling time compared to conventional mushroom cell PCM architecture, since complete transformation of the PCM layer is necessary for PCNR operation. However, as discussed in the scaling analysis, a thinner PCM layer enables faster switching in scaled PCNR.

2.4.2. Contact Characterization

Key contact parameters measured for the PCNR are shown in Figure 4. ON-state resistance varies depending on the applied actuation pulse. Two types of actuation are defined for the device: “perfect” and “over” actuation. As stated above, a “perfect” actuation is the one with the exact actuation pulse amplitude and width required to turn on the relay. “Perfect” actuation is used for the switching time measurements. In an “over” actuation, a higher amplitude pulse is used to turn the relay on. Devices switched with “perfect” actuation pulses exhibit more cycles before failure, but higher contact resistance. Devices switched with “over” actuation pulses show significantly lower contact resistance, but fewer cycles before failure. Unless otherwise specified, all the results are obtained for perfectly actuated device.

ON-state resistance is measured for an “over” actuated device to get a lower bound on contact resistance. Figure 4a shows the current and resistance of a relay in the ON-state as a function of voltage over the drain and source contacts. A total contact resistance of 2.3 k Ω is measured for the relay, which

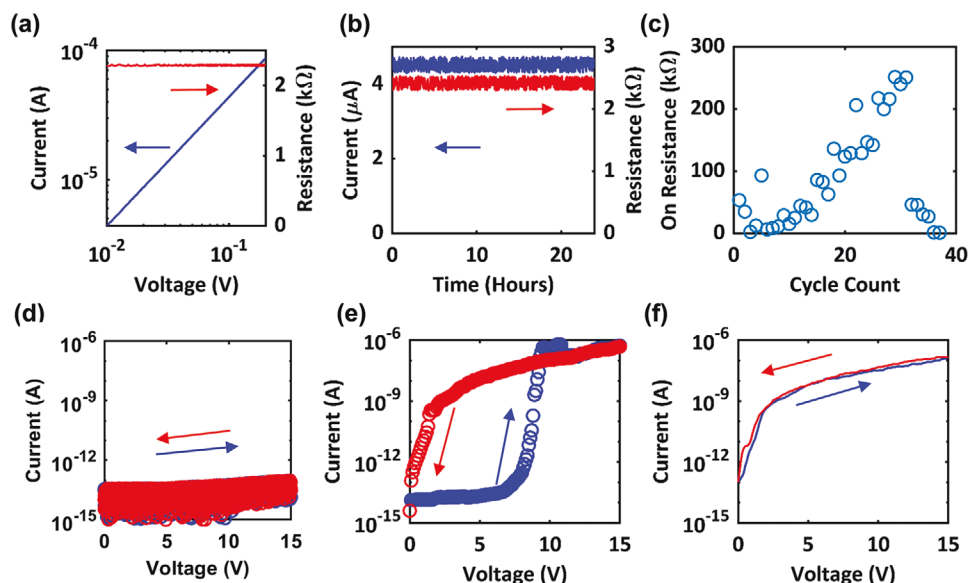


Figure 4. Key contact parameter measurements for the PCNR including a) ON-state contact resistance versus voltage, b) ON-state contact resistance over 24 h, c) ON-state contact resistance as a function of actuation cycle, d) OFF-state leakage of an unreleased air-gap, e) OFF-state leakage breakdown of the released air-gap, and f) OFF-state leakage of the released air-gap after breakdown.

corresponds to an average of 1.15 k Ω per contact since current must flow through both drain and source contacts. The contacts are ohmic as resistance is constant over the applied voltage range. Nonvolatility is assessed through limited testing of contact resistance over time for a “perfectly” actuated device. A “perfectly” actuated device is conservatively chosen instead of an “over” actuated device as less contraction of the actuator is required to break contact. However, this “perfectly” actuated device too has a relatively low ON-state resistance of 2.5 k Ω . ON-state resistance is measured for 24 h by applying a constant 10 mV across the drain and source contacts. Figure 4b shows the resistance as a function of time. Resistance is constant throughout the 24 h test, which demonstrates nonvolatility but does not prove long term stability. Accelerated testing at elevated temperatures is required to extrapolate the state retention time at room temperature. Contact resistance does change with actuation cycles. Figure 4c shows ON-state resistance as a function of cycle count for a relay “perfectly” actuated until failure. The 36 cycles shown represent the most cycles before failure of a tested PCNR device. A single amplitude actuation pulse is used until the 32nd cycle, at which point the amplitude is increased to counter the increasing ON-state resistance. The additional amplitude melts and converts more phase change material to the amorphous state, thus improving the contact resistance. The relay eventually fails in the ON state, which is typical of other high cycle count relays. Investigations on the failure mechanisms are ongoing. The two primary sources of failure are the PCM layer and the contact area. Degradation of the phase change properties of PCM has been extensively explored by the research community and solutions employed for conventional phase change memory is applicable for PCNR as well.^[24,25] On the other hand, contact degradation is one of the primary challenges for NEM relays. We are exploring the suitability of different durable contact materials, including conducting oxides, to increase the endurance of the relay. Our speculation on maximum achievable cycle count is preliminary, but with proper understanding of the failure mechanisms, our target is to achieve the same level of reliability as conventional phase change memory.

OFF state leakage is important to characterize as low leakage is one of the defining characteristics of the air-gap associated with an NEMS relay and it sets the ON–OFF ratio. OFF-state leakage measurements for the PCNR are taken under various conditions. Leakage current between the drain and source contacts is measured before and after the release etch. The release etch process (details in the Experimental Section) is used to isotropically remove the 20 nm SiO₂ layer deposited to define the air-gap between the contact pair. Figure 4d shows the leakage current for an unreleased device as voltage is swept up to 15 V and back down. The unreleased relay exhibits around 10 fA of leakage current and the sacrificial SiO₂ does not break down with up to 15 V across the drain and source. Figure 4e shows the leakage current for the same relay after release. At a nominal D/S operating voltage (<1 V), off-state leakage current is 14 fA. After the release etch the relay undergoes a breakdown event starting at 8 V which is substantially higher than the nominal drain voltage of 1 V. The leakage current follows a different relationship with voltage after the breakdown event. Figure 4f shows a subsequent leakage current measurement

after the breakdown event. At the limit of nominal D/S operating voltage (1 V), leakage current reaches a high value of 30 pA. The relay has undergone a permanent change in leakage current as the hysteresis observed in the initial breakdown no longer occurs, but instead the breakdown path is followed. Also, a linear *I*–*V* relation is observed which indicates the growth of a postbreakdown resistive path. The single breakdown event is likely caused by field emission current and subsequent electrode vaporization.^[26]

3. Discussion

The microscale prototype is the first experimental demonstration of the proposed device topology. It validates the device principle of operation and verifies the FEA model. We use imaging techniques to analyze the condition of the contacts and gain more insights on possible routes of device failure. With this knowledge, we can compare PCNR with existing relay devices and predict its performance in CMOS-comparable technology nodes.

Figure 5 shows a transmission electron microscope (TEM) image of an actuated device sectioned along the length of the heater. In regions where the channel contact overlaps the drain and the source, the air-gap should be reduced to zero in an ideally actuated device. However, the TEM images in Figure 5c–e show that the contact is closed only at the inside edges of the source or drain, and that the air-gap gradually opens as we go further away from the center. This is caused by a nonuniform spatial temperature distribution along the length of the heater. Maximum temperature is reached at the center, resulting in a bowing effect as the PCM expands. This problem will be mitigated in a scaled device as temperature is more uniform in nanoscale heaters. An electron energy loss spectroscopy (EELS) analysis verifies that the device is indeed released prior to actuation, as confirmed by the absence of any material in the air-gap (Figure 5f).

The low cycle count can be attributed to failure in the PCM layer as repeated thermal cycling is reported to have caused void formation and phase segregation.^[25,27] Other sources of PCM failure, such as high through current and high electric field, are not relevant to our device. While the alumina layers electrically isolate GeTe from both the heater and the contact pair, its careful placement on the heater mitigates the chances of field-induced separation of elements. Out-of-plane mechanical distortion of the drain and source electrodes due to plastic deformation can also cause failure. However, since most devices failed in the ON state, we conclude that it may not have played a significant role in limiting the cycle count.

Figure 6 shows a scaling comparison of the area and actuation voltage of the PCNR, an existing MEMS electrostatic cantilever design and its theoretical limit, and CMOS devices (see Section S1, Supporting Information, for the PCNR scaling model).^[1,2,28,29] The PCNR shows a significant reduction in area and voltage compared to the MEMS cantilever relay and offers comparable area and actuation voltage to current CMOS devices. Although voltage is not the only metric used to evaluate a technology, this comparison shows that an aggressively scaled PCNR can be integrated with CMOS to provide area-efficient

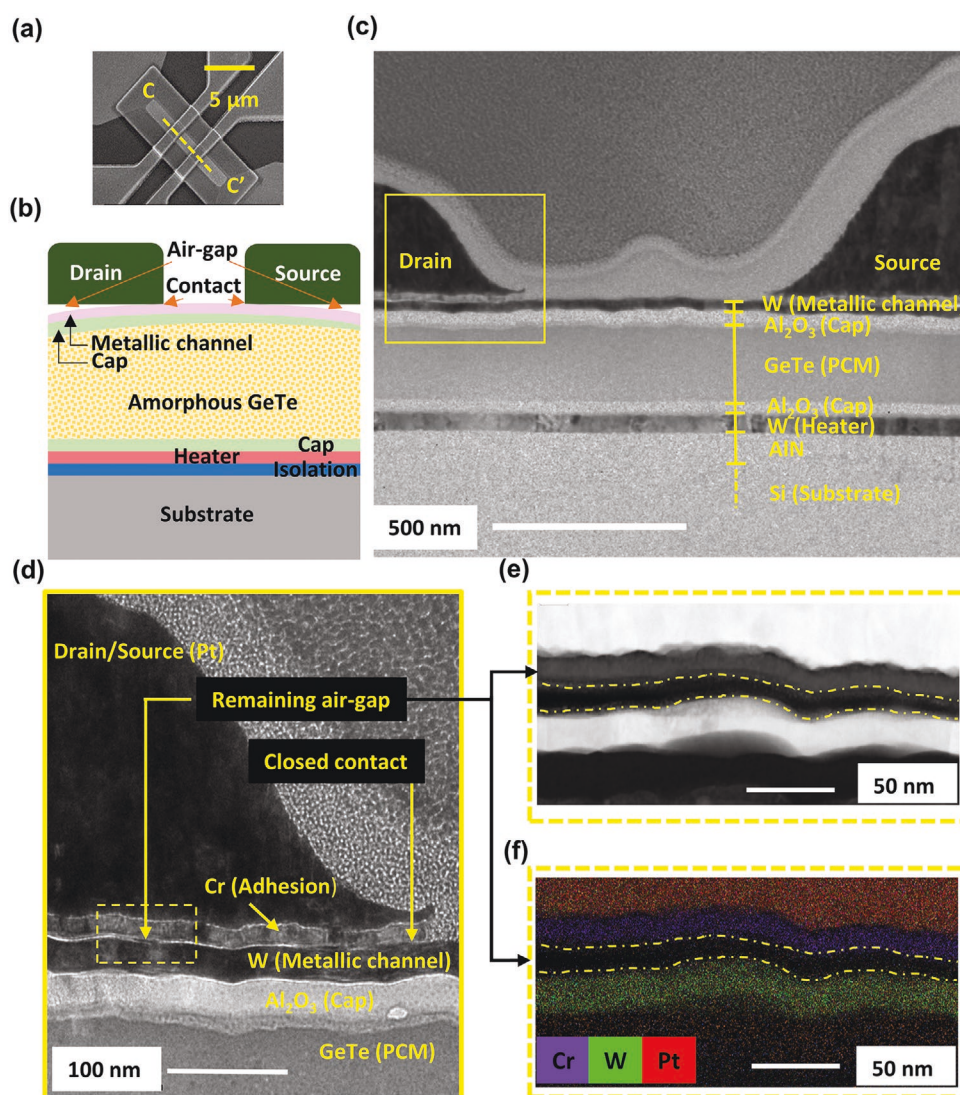


Figure 5. TEM of a cross-section along the length of the heater and electron energy loss spectroscopy (EELS) elemental map of the materials in and around the PCNR air-gap after release and actuation. a) SEM of the actuated device marking the line for cross-section. b) A schematic cartoon of a “nonideal” actuated device. Nonuniform heating causes the PCM to bow at the middle, resulting in contacts formed only near the center of the heater. c) TEM image on an actuated device. d) Magnified view of the yellow rectangle in (c) showing the contact between the drain and top metal contact. The contact is made near the edge of the drain and an air-gap slowly appears as we move to the left. e) Further magnification of the dashed rectangle in (d) shows the gap (between two dashed boundary lines) between the two metal layers and f) EELS element analysis demonstrates the absence of any material in the gap.

nonvolatile functionalities. Further improvements in actuation energy and voltage are possible by adjusting the geometry to better thermally couple the heater to the PCM (see Section S2, Supporting Information). Reducing thermal resistance between PCM and heater or increasing thermal resistance between the PCM and substrate can reduce temperature differential and improve energy efficiency. One way to reduce the temperature differential between the PCM and the heater is by adjusting the relative surface area between the PCM and the heater and the PCM and the substrate—thus increasing the thermal coupling between the PCM and the heater while reducing the thermal coupling between the PCM and the substrate.

Switching energy can also be improved for highly scaled devices by increasing the thermal resistance between the heater and the substrate. Higher thermal resistance reduces the input

power required to maintain a temperature, but there are limits to increasing thermal resistance between the heater and substrate as high thermal resistance slows the quench time. The thick layers of PCM used in this initial demonstration, due to their high thermal resistance, are close to the quenching speed requirement. The thin layers of PCM used in a highly scaled design can benefit from increased thermal resistance between the heater and substrate to decrease switching energy while maintaining support for the amorphous state.

4. Conclusions

The PCNR is an alternative class of NEMS relay that addresses the scalability concerns associated with other MEMS relays. The

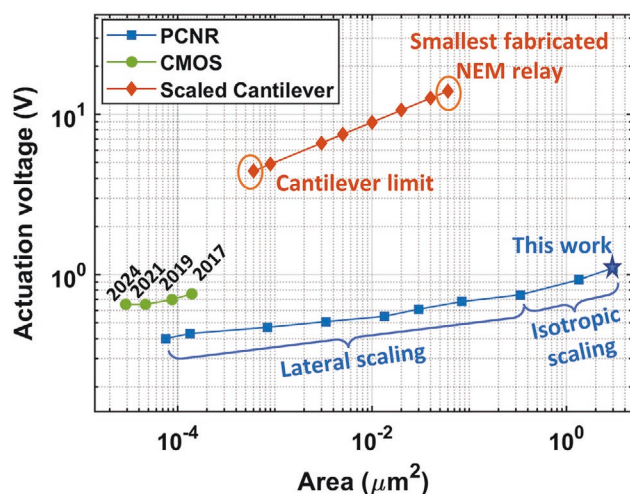


Figure 6. Comparison between different switching technologies by area and actuation voltage. Even without considering the anchor area, the shortest flexure-based NEM relay has four order larger footprint than a CMOS device.^[28] Further scaling is difficult since the minimum cantilever size is constrained by the material properties.^[29] PCNR, on the other hand, is already operated at a lower voltage and promises aggressive scalability.

key component is a layer of PCM that expands and contracts when converted between amorphous and crystalline states. In addition to shrinking actuator size by directly opening and closing the relay contacts with material contraction and expansion, the PCNR exhibits nonvolatile states because the PCM supports room-temperature stable amorphous and crystalline states. FEA models predict the switching characteristics of a given PCNR device and guide scaling laws and performance optimization. Models are verified experimentally via microscale prototypes and validate device switching and contact characteristics. Highly scaled geometries that operate at low voltages (400 mV), low energy (1.7 pJ), and high switching speeds (<2 ns), while preserving low leakage and nonvolatility, are possible to fabricate. Lateral heat flow in the actuator is identified as a source of inefficiency in highly scaled devices. Further improvements in actuation energy can be achieved by adjusting the geometry to better couple heater energy into the PCM instead of the substrate. Future work will focus on developing a design which increases the surface area of the PCM in contact with the heater while decreasing the surface area of the PCM in contact with the substrate—thus maintaining the fast quench time required to support the amorphous state while reducing temperature difference between the heater and PCM.

5. Experimental Section

Electrical and Thermal FEA Model: A model of the PCNR actuator is built in COMSOL Multiphysics finite element analysis software that combines the electrical and thermal physics of the actuator (see Section S3, Supporting Information). The COMSOL model only includes the heater layer's electrical properties as all other portions are insulators or are electrically insulated from the heater. The necessary high heater temperature operation requires the inclusion of its temperature coefficient of resistivity (TCR), which causes a ≈ 2.5 times increase in

resistivity at peak heater temperature. The heater resistivity and TCR used in the simulations were taken from measurements on a test structure at temperatures ranging from 170 to 370 K in a Lakeshore Cryogenic probing station. Resistivity at 293 K for the W heater was $1.67 \times 10^{-7} \Omega \text{m}$ with a TCR of $0.17\% \text{K}^{-1}$.

The thermal model includes the substrate, heater, PCM, and insulating layers. The same PCM thermal conductivity ($1 \text{ W m}^{-1} \text{K}^{-1}$) and latent heat of phase change (68 kJ kg^{-1}) were used for both phases of the material.^[30] ALD Al_2O_3 and sputtered W have thermal conductivity of $2 \text{ W m}^{-1} \text{K}^{-1}$ and $40\text{--}70 \text{ W m}^{-1} \text{K}^{-1}$, respectively.^[31] The heater acted as a source of thermal energy by coupling the electrical model to the thermal model through joule heating. Thermal boundary resistances (TBRs) were included at the interfaces between each dissimilar material.^[32] The TBR was modeled as a fixed thermal resistance per area ($1.42 \times 10^{-8} \text{ K m}^2 \text{W}^{-1}$ between the heater and PCM) and was a dominant factor in highly scaled devices. The cap and metallic contact were not modeled as the latent heat of the PCM was dominant for thermal timing. As the PCM melted, it absorbed large amounts of heat due to the latent heat of fusion while maintaining an approximately constant temperature. Melting was modeled as a sharp increase in heat capacity at a narrow temperature range around the melting temperature such that the added heat equaled the latent heat of fusion.

Actuator Expansion FEA Model: The actuator model is built in 2D as the long aspect ratio of the actuator allows for a plane strain condition at the center of the cross-section (see Section S4, Supporting Information). Rather than specifying a strain within a solid model of the PCM, a hydrostatic pressure was applied to the boundary of the molten PCM and was proportional to the volume change of the molten area. The hydrostatic pressure better modeled the forces generated by the molten PCM as it did not support shear forces at the boundary between PCM and structure. The PCM was only included in the mesh to measure the volume change used when calculating the hydrostatic pressure. The hydrostatic pressure on the boundary best modeled the pressure generated from the molten PCM as this allowed for lateral motion that would otherwise be constrained.

Fabrication: Fabrication of the PCNR was completed following a 4 mask process that is illustrated in 1–8 subfigures of Figure S5 (Supporting Information). Figure 1 shows a cross-section (Figure 1c) and top view (Figure 1b) image of a fabricated PCNR device taken with an SEM. The chosen insulation layer material (AlN) mitigated electrical conduction between heater and substrate while allowing low thermal resistance, which facilitated rapid quenching. The cap layer choice was also crucial since it insulated the PCM from metallic heater and contact layers, and also encapsulated molten PCM during phase changes. SiO_2 was used as the sacrificial layer as it was isotropically etched in vapor HF while keeping the alumina cap layer intact. Finally the drain and source layer was fabricated by lifting off a thick W layer so that unwanted tensile stress-induced bending of the contact pair during off state could be limited.

Switching Test Setup: The test setup (shown in Figure S6, Supporting Information) was used to evaluate the switching time for the fabricated PCNR devices. Actuation pulses were generated by an Agilent B1110A pulse generator with a 50Ω internal resistance.

Each heater size required a different actuation pulse for optimal switching. Actuation pulse variation was limited to amplitude, so all actuation pulses shared the same rise time, fall time, and duration. Each relay was actuated ON and OFF until failure. Testing was conducted in both air and vacuum with minimal change in actuation voltage, although devices tested in vacuum withstood more cycles before failure. Failure of the relay was defined as when the device stopped responding to actuation pulses.

Contact Characterization Test Setup: Contacts were characterized by sweeping voltage and measuring current to obtain I – V curves. Contacts were probed with an Agilent B1500A Semiconductor Device Parameter Analyzer using only two probes. Two-point probe measurements included contact resistance as well as probe resistance. Two-point measurements were chosen over more accurate four-point measurements due to limited probing pad size as well as the negligible resistance of the probe compared to the device resistance.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

The authors acknowledge the Kavcic-Moura Endowment Fund and the U.S. National Science Foundation (NSF) (Award No. 1854702) for supporting this work. The authors are grateful to the staff members of the cleanroom facility at Claire & John Bertucci Nanotechnology Laboratory in Carnegie Mellon University. The authors also acknowledge the use of the CMU Materials Characterization Facility at Carnegie Mellon University supported by Grant No. MCF-677785.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

J.T.B. and M.A.M. contributed equally to this work. J.T.B. designed, fabricated, and characterized the PCNR devices. M.A.M. worked in scaling analysis, fabrication, and characterization. M.P.d.B. and G.P. supervised the work as principal investigators.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

memory, nanoelectromechanical systems, nonvolatile, phase change, relay

Received: January 22, 2022
Revised: March 11, 2022
Published online:

- [1] *International Roadmap for Devices and Systems (IRDS)*, 2018 ed., IEEE, Piscataway, NJ **2018**.
- [2] *International Roadmap for Devices and Systems (IRDS)*, 2021 ed., IEEE, Piscataway, NJ **2021**.
- [3] P. Chi, S. Li, C. Xu, T. Zhang, J. Zhao, Y. Liu, Y. Wang, Y. Xie, *SIGARCH Comput. Archit. News* **2016**, 44, 27.
- [4] L. Zhu, J. Zhou, Z. Guo, Z. Sun, *J. Mater.* **2015**, 1, 285.
- [5] B. C. Lee, P. Zhou, J. Yang, Y. Zhang, B. Zhao, E. Ipek, O. Mutlu, D. Burger, *IEEE Micro* **2010**, 30, 143.
- [6] G. W. Burr, M. J. Breitwisch, M. Franceschini, D. Garetto, K. Gopalakrishnan, B. Jackson, B. Kurdi, C. Lam, L. A. Lastras, A. Padilla, B. Rajendran, *J. Vac. Sci. Technol., B: Nanotechnol. Microelectron.: Mater., Process., Meas., Phenom.* **2010**, 28, 223.
- [7] D. Apalkov, B. Dieny, J. M. Slaughter, *Proc. IEEE* **2016**, 104, 1796.
- [8] S. Tehrani, *IEEE Int. Electron Devices Meet*, IEEE, San Francisco, CA **2006**, p. 330.
- [9] H. Mulaosmanovic, J. Ocker, S. Müller, U. Schroeder, J. Müller, P. Polakowski, S. Flachowsky, R. van Bentum, T. Mikolajick, S. Slesazek, *ACS Appl. Mater. Interfaces* **2017**, 9, 3792.
- [10] D. Reis, K. Ni, W. Chakraborty, X. Yin, M. Trentzsch, S. D. Dunkel, T. Melde, J. Müller, J. Beyer, S. Datta, M. T. Niemier, *IEEE J. Explor. Solid-State Comput. Devices Circuits* **2019**, 5, 103.
- [11] S. Lai, *IEEE Int. Electron Devices Meet*, IEEE, San Francisco, CA **2008**, p. 121.
- [12] U. Sikder, G. Usai, L. Hutin, T. J. K. Liu, *IEEE Electron Devices Technology and Manufacturing Conf. EDTM 2018 – Proc.*, IEEE, Japan **2018**, p. 128.
- [13] R. Gaddi, C. Schepens, C. Smith, C. Zambelli, A. Chimenton, P. Olivo, *IEEE Int. Reliability Physics Symp. Proc.*, IEEE, Monterey, CA **2011**, pp. 2G.2.1–2G.2.6.
- [14] R. Nathanael, V. Pott, H. Kam, J. Jeon, T.-J. K. Liu, *IEEE Int. Electron Devices Meet.*, IEEE, San Francisco, CA **2009**, p. 223.
- [15] J. Best, G. Piazza, *32nd IEEE Int. Conf. Micro Electro Mechanical Systems*, IEEE, Korea **2019**, p. 962.
- [16] J. T. Best, M. A. Masud, M. P. de Boer, G. Piazza, *IEEE Int. Electron Devices Meet*, IEEE, San Francisco, CA **2019**, pp. 34.1.1–34.1.4.
- [17] R. Cheng, Y. Ding, S. M. Koh, A. Gyanathan, F. Bai, B. Liu, Y. C. Yeo, *IEEE Symp. on VLSI Technology*, IEEE, Honolulu, HI **2012**, p. 93.
- [18] A. Fantini, L. Perniola, M. Armand, J. F. Nodin, V. Sousa, A. Persico, J. Cluzel, C. Jahan, S. Maitrejean, S. Lhostis, A. Roule, *IEEE Int. Memory Workshop*, IEEE, Monterey, CA **2009**, <https://doi.org/10.1109/IMW.2009.5090585>.
- [19] L. Perniola, V. Sousa, A. Fantini, E. Arbaoui, A. Bastard, M. Armand, A. Fargeix, C. Jahan, J. F. Nodin, A. Persico, D. Blachier, *IEEE Electron Device Lett.* **2010**, 31, 488.
- [20] J. Pries, Y. Yu, P. Kerres, M. Häser, S. Steinberg, F. Gladisch, S. Wei, P. Lucas, M. Wuttig, *Phys. Status Solidi RRL* **2021**, 15, 2000478.
- [21] W. M. Cho, B. E. Lee, S. H. Koo, Y. S. Lee, *Comput. Struct.* **1995**, 55, 1063.
- [22] M. Jeong, J. P. Freedman, H. J. Liang, C. M. Chow, V. M. Sokalski, J. A. Bain, J. A. Malen, *Phys. Rev. Appl.* **2016**, 5, 014009.
- [23] E. T. Swartz, R. O. Pohl, *Rev. Mod. Phys.* **1989**, 61, 605.
- [24] A. Padilla, G. W. Burr, C. T. Rettner, T. Topuria, P. M. Rice, B. Jackson, K. Virwani, A. J. Kellock, D. Dupouy, A. Deunne, R. M. Shelby, *J. Appl. Phys.* **2011**, 110, 054501.
- [25] Y. Wang, Y. Yuan, Y. Wang, S. Yuan, X. Chen, *IEEE Trans. Device Mater. Reliab.* **2021**, 21, 236.
- [26] F. W. Strong, J. L. Skinner, P. M. Dentinger, N. C. Tien, *Proc. SPIE* **2006**, 6111, 611103.
- [27] Y. J. Park, T. Y. Yang, J. Y. Cho, S. Y. Lee, Y. C. Joo, *Appl. Phys. Lett.* **2013**, 103, 073503.
- [28] W. W. Jang, J. B. Yoon, M. S. Kim, J. M. Lee, S. M. Kim, E. J. Yoon, K. H. Cho, S. Y. Lee, I. H. Choi, D. W. Kim, D. Park, *Solid-State Electron.* **2008**, 52, 1578.
- [29] L. Hutin, W. Kwon, C. Qian, T. J. K. Liu, *IEEE Trans. Electron Devices* **2014**, 61, 1382.
- [30] S. R. Choi, D. Kim, S. H. Choa, S. H. Lee, J. K. Kim, *Int. J. Thermophys.* **2006**, 27, 896.
- [31] E. A. Scott, J. T. Gaskins, S. W. King, P. E. Hopkins, *APL Mater.* **2018**, 6, 058302.
- [32] J. Martan, N. Semmar, C. Boulmer-Leborgne, P. Plantin, E. Le Menn, *Nanoscale Microscale Thermophys. Eng.* **2006**, 10, 333.