Sub-10 nm Scaling of Phase-Change Memory: Thermoelectric Physics, Carbon Nanotube and Graphene Electrodes

A. Behnam[1], F. Xiong[1,2], K. L. Grosse[1], A. Cappelli[3], S. Hong[1], N. Wang[1,2], M.-H. Bae[1], Y. Dai[1], A.D. Liao[1], E.A. Carrion[1], D. Ielmini[4], E. Piccinini[3], C. Jacoboni[3], W. P. King[1], and E. Pop[1,2*]
[1]Univ. Illinois Urbana-Champaign, IL 61801, USA
[2]Now at Stanford Univ., Stanford, CA 94305, USA. [*]Contact: epop@stanford.edu
[3]Univ. Modena & Reggio Emilia, Modena, Italy; [4]Politecnico di Milano, Milano, Italy

ABSTRACT

We probe the scalability of phase change memory (PCM) to sub-10 nm dimensions by taking advantage of the atomically thin diameter and thinness of carbon nanotube (CNT) and graphene electrodes. Using self-aligned PCM nanowires with individual CNT electrodes we achieve ultralow programming currents (~0.1 μA set, ~1.6 μA reset) and high ON/OFF ratios (~1000x) at few-nanometer bit dimensions. We also examine patterned graphene electrodes, better suited for large-scale integration. Such devices also switch at few-Volts and few-μA, enabled by the sharp contact area with the atomically thin graphene, although endurance remains to be improved. We further measure power dissipation in PCM devices by scanning Joule expansion microscopy (SJEM) and observe that the temperature rise is dominated by Joule heating within the PCM material, whereas it is dominated by a combination of Peltier and current crowding effects at the contacts. Knowledge of such effects is essential for energy-efficient design of future PCM technology. Our results probe the fundamental scaling limits of PCM technology, also yielding devices ready for integration into flexible or transparent electronics with strict low-power requirements.

Key words: phase-change, low-power, carbon nanotube, graphene, Joule, Peltier, contacts.

1. INTRODUCTION

Electrically-programmable Phase Change Memories (PCMs) have captivated wide interest for applications in non-volatile memory[1] and reprogrammable circuits[2] due to low power operation, fast access times, and high endurance[3,4]. Data in PCMs are stored by the large ratio (>10^3) in electrical resistance between amorphous and crystalline states of the material. A drawback of PCMs is their high programming current (>0.1 mA), as Joule heat must be coupled to a finite bit volume. A central issue of PCMs concerns their fundamental scaling limits, i.e. the smallest and most energy-efficient devices that can function reliably, and what their key fundamental physics of operation are.

In this paper we review our work using carbon nanotubes (CNTs) and graphene ribbons as electrodes to induce reversible phase change in an extremely small volume of the chalcogenide-based PCM bit (in this case $Ge_2Sb_2Te_5$, GST). This approach significantly lowers the programming current and power of PCM devices and provides an excellent platform to study their scalability. Graphene electrodes also allow for large scale fabrication of PCM devices even under limited thermal budget that is suitable for flexible electronics. In addition to cell dimensions, electrical contacts and thermal interfaces to PCMs are also important for heat generation and thermal confinement. Here we also describe our work measuring the nanometer-scale temperature distributions in lateral PCM devices using scanning Joule expansion microscopy[7-8] (SJEM), which is an atomic force microscopy (AFM) based thermometry technique. The spatial and temperature resolutions are sub-50 nm and ~0.2 K. We find that Joule heating dominates the temperature rise of the GST, while the temperature rise at the metal-GST contacts consists of Joule, Peltier, and current crowding effects. Comparing SJEM measurements to a finite element analysis (FEA) model uncovers the thermopower (350 ± 150 μV K^{-1}) of a 25 nm thick film of fcc-GST.

2. PCM NANOWIRES SELF-ALIGNED WITH CARBON NANOTUBE ELECTRODES

We first discuss our approach for fabrication and characterization of PCM nanowire devices self-aligned with CNT electrodes.[3] Figure 1 presents a schematic of our approach and AFM images of CNT devices at different fabrication stages. We begin with a CNT spanning two Pd electrodes on a SiO_2/Si substrate (Figures 1a-b). We spin a thin layer of polymethyl methacrylate (PMMA), typically ~50 nm onto the device (Figure 1c). We then apply a voltage across the Pd pads in vacuum, flowing current through the CNT such that localized Joule heating along its length causes the PMMA covering it to evaporate, leaving behind a narrow trench self-aligned with the CNT (Figure 1d-e). We then create a nanogap (~20 to 150 nm) in the exposed CNT by electrical breakdown at high bias in air under Ar flow (Figure 1f). We then sputter ~10 nm of GST over the device, filling the nanogap and nanotrench as shown in Figure 1g. We lift-off the remaining PMMA leaving behind a GST nanowire (NW) that spans the nanogap and is perfectly aligned with the CNT electrodes (Figures 1h-i).

The AFM image and electrical characteristics of final devices are shown in Figure 2. Figure 2b shows current-voltage characteristics under DC current sweep, demonstrating memory SET switching from the high resistance amorphous (OFF) phase ($R_{OFF} \approx 2.5$ GΩ) to the low-resistance crystalline (ON) phase ($R_{ON} \approx 1.3$ MΩ). The SET switching is initiated at a threshold voltage (V_T) through a field-induced transition of the amorphous phase;[9] Joule heating then crystallizes the bit at ~150 ºC into the conductive state. The device V_T decreases by 20-30% after the first few switching cycles (Figure 2b), which is consistent with previous reports.[1,10] This "burn-in" is beneficial as it stabilizes the memory bit and allows lower power operation in the long run. Reversible memory switching is achieved with pulsed operation and shown in Figure 2c. The bit is re-amorphized (RESET) with a current pulse which heats up the crystalline GST (c-GST) to its melting point (~620 ºC) then quenches it back to a disordered amorphous GST (a-GST) state during the short falling edge of the pulse. This device has $R_{OFF}/R_{ON} \approx 2000$, effectively approaching the intrinsic resistivity ratio between a-GST and c-GST (10^3-10^4).[1,10]

The self-aligned structure presents several benefits allowing us to approach the fundamental limits of switching in such small PCM bits. The narrow constriction of the self-aligned NW (Figure 2a inset) enables the ultra-high ON/OFF ratio by eliminating parasitic leakage paths around the small PCM bit. The NW constriction also improves device endurance (Figure 2d) compared to previous results with CNT electrodes,[1,10] by limiting the size of the so-called crystalline "halo" that can form around the bit region[9] after several switching events. The programming currents are reduced by ~100× compared to industrial state of the art[11,12] by the use of CNT electrodes which have much smaller diameter (~2 nm) than typical metal electrodes (~20-80 nm); in addition, the combination of CNT electrodes and nar-

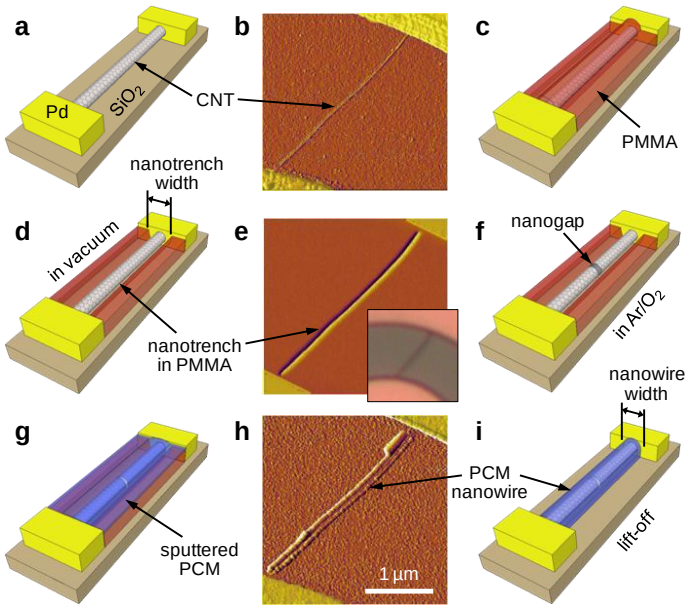

Figure 1. Schematics and AFM imaging of self-aligned devices, after Ref. 3. **a**, CNT between two Pd electrodes. **b**, AFM of a CNT with length $L \sim 3.1$ µm and diameter $d \sim 2.2$ nm. **c**, The CNT device is coated with a thin layer (~50 nm) of PMMA. **d**, Current flow in the CNT leads to Joule heating and nanotrench formation along it as the PMMA evaporates (in vacuum). **e**, AFM imaging of nanotrench (~90 nm wide) in PMMA. Inset shows nanotrench is visible under the optical microscope, enabling quick detection. **f**, CNT nanogap is formed by electrical cutting under Ar/O_2 flow. **g**, PCM deposition covers the device and fills the nanogap and nanotrench. **h-i**, AFM imaging and schematic of self-aligned nanowire (NW) with CNT electrodes obtained after PMMA lift-off. Some devices were further encapsulated with a ~10 nm layer of evaporated SiO_2. See Ref. 3.

row NW constriction further reduces programming current by 3-5× with respect to devices which used CNT electrodes alone.[1]

3. PCM WITH GRAPHENE ELECTRODES

In addition to CNT electrodes that are useful for exploring the limits of scaling in PCM devices, we have fabricated and characterized lateral PCM devices with graphitic electrodes. Graphene is a strong candidate as electrode/interconnect for future wearable/flexible electronics due to its simultaneous (good) conductance, current capacity, transparency and flexibility.[13] Recently, chemical vapor deposition (CVD) has facilitated wafer-scale and low cost growth of high quality graphene that can be transferred easily to various substrates for device fabrication.[14] As a result, CVD graphene electrodes can play a role in future development of high-density flexible non-volatile memories.

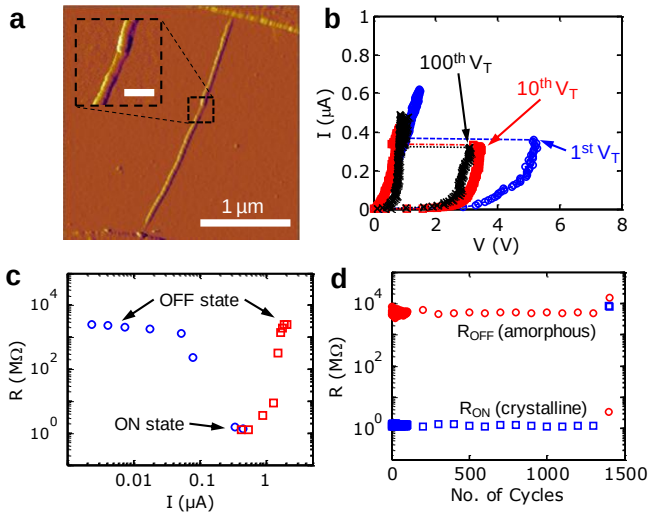

Figure 2. PCM-CNT electrical characteristics. **a**, AFM image of a self-aligned NW with CNT electrodes of d ~ 2.5 nm. The GST nanowire is ~40 nm wide, ~10 nm tall, and capped by ~10 nm SiO_2. Inset zoom-in shows nanogap region (scale bar 150 nm). **b**, Electrical characteristics of the 1st, 10th and 100th SET switch, showing the threshold voltage stabilizes at V_T ~ 3.2 V. **c**, Resistance switching after a series of current pulses with increasing amplitude. SET (RESET) pulses have 300 ns (100 ns) width and rising/falling edges of 50 ns (2 ns). The SET (RESET) current is ~0.4 µA (~1.9 µA). The ratio R_{OFF}/R_{ON} = 2.5 GΩ/1.3 MΩ, nearly ~2000×. **d**, Endurance test for ~ 1500 cycles of operation.

We have developed two different lateral structures for characterizing different aspects of graphene-based PCM devices (Figure 3). For both structures, we use substrates of 90 nm SiO_2 on highly doped Si, and transfer multi-layer or single-layer CVD graphene onto their surface. We then pattern large probing contacts (Figure 3a) using photolithography and e-beam evaporation. For the first structure, we continue fabrication and use e-beam lithography to define smaller metal contacts and then graphene nanoribbon (GNR) electrodes with a small gap in the middle (Figure 3b). We remove the rest of graphene by O_2 plasma etching. We then fill the gap with sputtered GST (10 nm) and then evaporate a protective SiO_2 layer (10 nm) on top to complete the fabrication of GNR structures. The lengths (L) and widths (W) of the GNR electrodes are 0.5 - 2 µm and 30 - 400 nm, respectively. The gap length (L_G) is 30 - 100 nm.

For the second (GSTNR) structure, we again start from the large contacts and then pattern the graphene into microribbons (width of 2 - 10 µm) using photolithography (Figure 3c). We use a gold-based shadow evaporation technique[15] and follow it by O_2 plasma etching to controllably define a gap of 20 - 100 nm in the graphene ribbons. The last two steps are definition and deposition of a GST nanoribbon (width

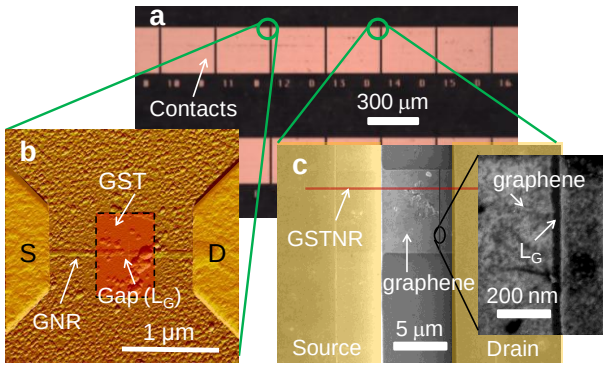

Figure 3. Design of graphene-based memory structures. **a**, Optical image of the large Ti/Au or ti/Ni contacts fabricated on Si/SiO_2 substrates. **b**, AFM image of the structure of a lateral GNR PCM cell. L = 1.5 µm, W ~ 50 nm and gap length (L_G) ~ 50 nm. GST thickness is ~ 10 nm and GST window is 1 µm by 700 nm. GNRs (including the gap) and GST window are both defined by e-beam lithography. **c**, Scanning electron microscope (SEM) image of the structure of a lateral GSTNR PCM cell. Graphene microribbons are defined by photolithography and the gaps are created by a shadow evaporation technique. Microribbon dimensions are L ~ 7 µm, W = 8 µm. Average L_G is 50 nm. The inset is a zoomed-in image of the created gap in graphene. GST nanoribbon is defined by e-beam lithography with 50 nm width.

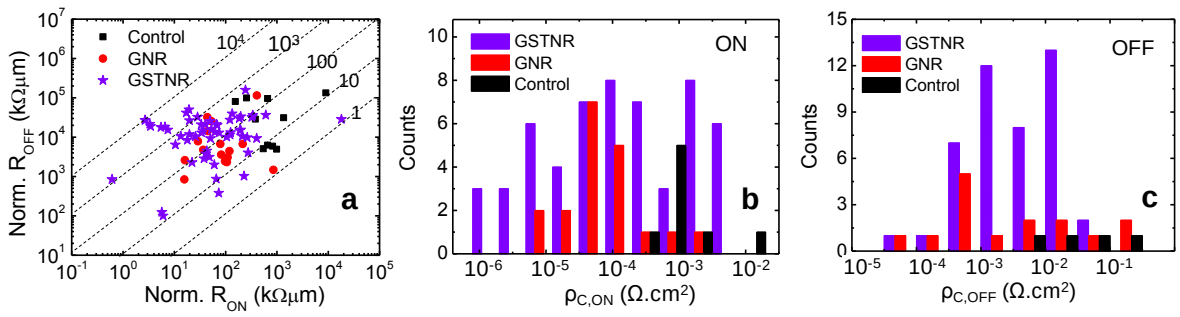

Figure 4. Results for GSTNR, GNR, and control PCM devices. **a**, Normalized ON vs. OFF resistances. Dashed lines show the ON/OFF ratio limits. **b,c** Extracted contact resistivities between graphene and GST in ON and OFF states, respectively.

~50 nm) across the gap in the graphene electrodes (using e-beam lithography) and evaporation of the protective SiO_2 layer (10 nm) on top. While GNR structures allow us to evaluate the scalability of the PCM devices with graphitic electrodes, GSTNR structures facilitate the extraction of device parameters such as contact resistance between GST and graphene due to the well-defined structure of the GST nanoribbon.

Electrical measurements, AFM and Raman spectroscopy confirm good quality and uniformity of our graphitic electrodes[13]. Our best devices switch at threshold voltages as low as ~3 V with low programming currents (<1 μA SET, <10 μA RESET) and good ON/OFF ratios (>100), enabled by the sharp contact area with their atomically thin graphene electrodes (Figure 4a). There is, however, a wide scatter in device performance and the operation is only stable within the first 10-100 SET/RESET cycles. As a result, a better understanding of the interface between GST and graphene electrodes and improving device quality are necessary for performance optimization.

Estimated minimum contact resistivities between graphene and GST in the crystalline and amorphous states (Figures 4b and 4c, respectively) are 0.2 and 50 $k\Omega\cdot\mu m^2$ respectively, comparable to values reported for other GST contacts (e.g. with TiW[16,17]). For extracting contact resistivity values we have considered the design of devices and subtracted all the additional contributors to total device resistance (e.g. GST resistance, contact resistance between graphene and metal electrodes or graphene electrode access resistance). We have also fabricated control devices with structures similar to GNR devices, but with metal electrodes contacting lateral GST bits instead. Overall performance of graphene-based devices, particularly those with small dimensions, is comparable to or better than the control devices due to a more efficient contact and smaller effective area between GST and graphene.

4. THERMOELECTRIC PHYSICS AT GST-METAL CONTACTS

In highly scaled PCM devices electrical and thermal interfaces are also important for obtaining energy-efficient devices. Electrical and thermal interfaces can cause heat generation and thermal confinement in devices, which in turn affect PCM programming power. We therefore present the results of our SJEM analysis of Joule, Peltier, and current crowding (interface) effects for GST-based lateral PCM devices.[20]

Figure 5a shows the lateral GST test devices. GST (25 nm) was deposited on SiO_2 on highly p-doped Si substrate. For electrical contact, 100 nm TiW (10/90 % weight) was patterned by photolithography and deposited by sputtering. Devices were encapsulated by 10 nm of evaporated SiO_2. Fabrication was completed by spin coating 60 nm of PMMA on the samples. The PMMA protects the devices from oxidation, and amplifies thermo-mechanical expansions of the PCM device during operation, as needed for the SJEM technique.[8] Before starting the SJEM measurements we crystallized the GST into predominately face-centered cubic (fcc) phase by baking the entire device on a hot plate at 200 ^oC for 5 minutes.[19] X-ray diffraction (XRD) analysis confirms the fcc films and the measured electrical resistivity is also in the generally accepted range for thin fcc films (10^{-2} –10^{-1} Ω·cm).

Figure 5a also shows a schematic of the SJEM experiment. A sinusoidal waveform at 28 kHz with peak voltage *V* was applied to resistively heat the device. The associated thermo-mechanical expansions

of the sample were measured by the AFM cantilever, laser, and photodiode. A lock-in amplifier at the heating frequency f_H with a low-pass filter bandwidth of 3 - 27 Hz recorded the peak surface expansion Δh. The spatial resolution was ~50 nm and temperature resolution was ~0.2 K based on our previous reports for similar geometries.[8]

Figure 5b shows the surface expansion Δh overlaid on the topography of a device. The GST peak temperature rise ΔT is proportional to Δh and is calculated by Finite Element Analysis (FEA) modeling.[8] The device was biased with $V = \pm 8.9$ V at 28 kHz, and Δh was recorded at $f_H = 56$ kHz, as Joule heating occurs at twice the applied frequency for a bipolar sine wave. Further increasing f_H decreases Δh as the thermal diffusion length decreases, which decreases the amount of material which thermo-mechanically expands.[19] At the micrometer-scale Δh is uniform across the device indicating uniform lateral heating, electric field, and resistivity distribution.

We observe heat generation at the GST-TiW interface due to current crowding and Peltier effects. Current crowding is independent of carrier flow direction and occurs due to a finite interface resistivity, ρ_C, between the GST channel and TiW contacts. On the other hand, the Peltier effect is dependent on the direction of current flow through junction and heats (cools) the junction as carriers flow into (out of) the contact due to the difference in GST and TiW thermopower. The TiW-GST interface properties were found by fitting the predicted and measured Δh. A GST resistivity $\rho_{GST} = 1.7\times10^{-2}$ Ω·cm and a contact resistivity $\rho_C = 3\times10^{-5}$ Ω·cm^2 fit the measurement results. In addition, ρ_C and ρ_{GST} values are also close to values obtained from Transmission Line Method (TLM) measurements.[20] Joule heating dominates power dissipation in the GST as expected. The majority of heat generation at the contacts is, however, due to the finite ρ_C and associated current crowding effect. Peltier heating and cooling is observed as the change in ΔT with hole flow direction. At $|V| = 1.6$ and 3.2 V the difference in ΔT with carrier flow at the channel edge is ~1.5 and 3 K (~63 and ~32 % of the channel ΔT).

Figure 5c shows the fitting of predictions to measurements to determine the GST thermopower, S_{GST}. The difference in Δh for hole flow left and right (due to Peltier heating and cooling of the contacts) is denoted $\Delta h_{Peltier} = \Delta h(j+) - \Delta h(j-)$, where $j+$ and $j-$ denote hole flow left and right. Figure 5d shows the coefficient of determination R^2 for predictions at both contacts for each bias condition. The average R^2 curve has a maximum $R^2 = 0.65$ which predicts $S_{GST} \approx 350 \pm 150$ μV K^{-1} for fcc phase GST, similar to previous studies.[18,21]

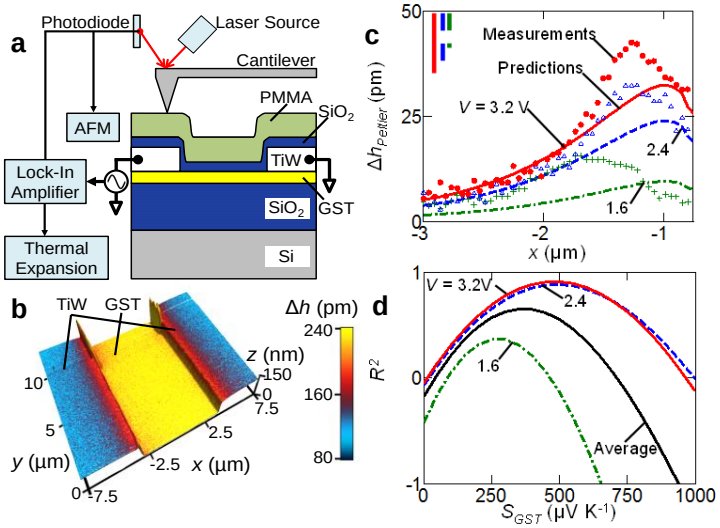

Figure 5. a, Schematic of phase PCM device and SJEM experiment. Lateral test devices consisted of 60 nm PMMA, 10 nm SiO_2, 100 nm TiW, 25 nm GST, and 300 nm SiO_2 on a Si substrate, from top to bottom. SJEM operates by supplying a periodic voltage waveform to resistively heat the GST device, while AFM measures the resulting peak surface thermo-mechanical expansion Δh. **b**, Measured Δh overlaid on topography for a device. The peak GST temperature rise (ΔT) is proportional to Δh. **c**, Measured and predicted $\Delta h_{Peltier}$ at the GST-TiW interface for the $L = 1.5$ μm device. Three bias conditions are shown $|V| = 1.6$, 2.4, and 3.2 V in green dash-dot line and crosses; blue dashed line and triangles; and red solid line and dots. The GST thermopower $S_{GST} = 250$, 500, and 500 μV K^{-1} for $|V| = 1.6$, 2.4, and 3.2 V. Bars located in the top-left show the standard deviation of the measurements over 32 scans. Measurements are shown by markers and predictions are shown as lines. **d**, The coefficient of determination R^2 for predictions from both contacts. The three bias conditions are shown similar to **c**. The average R^2 is shown as a black solid line. Negative R^2 indicates the measurement average is a better fit than predictions. The S_{GST} uncertainty was estimated from 0.1 decrease below the maximum R^2.

5. CONCLUSION

In summary, we have developed a novel technique to fabricate sub-50 nm

PCM nanowires that are self-aligned to CNT electrodes without the need for complex lithography. This enabled us to study PCM devices with bits of a few hundred cubic nanometers, confined by the CNT nanogap and NW width. Such self-aligned PCM devices show ultra-small power consumption, improved endurance, and extremely high ON/OFF ratios approaching the intrinsic limits of the PCM. The powerful yet simple nanofabrication method could also serve as an excellent platform to study other nanoelectronic and molecular devices and materials self-aligned with CNT electrodes.

We have also presented the first study of PCM devices with patterned graphene electrodes for wafer-scale integration. The thin structure of these devices (with thin PCM and graphene layers) makes them ideal for flexible and transparent electronics which have strict low-power requirements. In addition, graphene interconnects can also be integrated with conventional CMOS substrates. Our devices switch at threshold voltages as low as ~3 V with low programming currents (<1 μA set, <10 μA reset) and good ON/OFF ratios (>100), enabled by the sharp contact area with their atomically thin graphene electrodes.

We also analyzed Joule, Peltier, and current crowding effects in a fcc phase GST device with TiW contacts using SJEM with ~50 nm spatial and ~0.2 K temperature resolution. Joule heating dominated power dissipation in the GST channel, while power dissipation at the GST-TiW contacts was a combination of Peltier and current crowding effects. Comparing measurements and modeling predictions, we obtained $S_{GST} \approx 350 \pm 150\ \mu VK^{-1}$ for a 25 nm thick film of fcc phase GST. The large measured thermopower of GST could reduce the energy consumption by >50 % in highly scaled PCM devices due to Peltier heating, compared to scenarios which only utilize Joule heating. PCM energy consumption can be further reduced by optimizing the GST-electrode contacts and their thermoelectric properties.

REFERENCES: [1]F. Xiong *et al.*, Science **332** (2011) 568. [2]S. M. Yoon *et al.*, IEEE EDL **30** (2009) 371. [3]F. Xiong *et al.*, Nano Lett. **13** (2013) 464. [4]D. Loke *et al.*, Science **336** (2012) 1566. [5]D. L. Kencke, *et al.*, IEEE IEDM (2007) 323. [6]J. P. Reifenberg, *et al.*, IEEE EDL **29** (2008) 1112. [7]J. Varesi and A. Majumdar, Appl. Phys. Lett. **72** (1998) 37. [8]K. L. Grosse, *et al.*, Nat. Nanotechnol. **6** (2011) 287. [9]D. Ielmini, Phys. Rev. B **78** (2008) 035308. [10]J. Liang, *et al.*, IEEE TED **59** (2012) 1155. [11]M. J. Kang, *et al.*, IEEE IEDM (2011) 39. [12]G. Servalli, IEEE IEDM. 2009, 113. [13]A. Behnam, *et al.*, Nano Lett. **12** (2012) 4424. [14]S. Bae, *et al.*, Nat. Nanotechnol. **5** (2010) 574. [15]Y. Naitoh, *et al.*, Nanotechnol. **17** (2006) 5669. [16]D. Roy, *et al.*, IEEE EDL **31** (2010) 1293. [17]E.K. Chua, *et al.*, Appl. Phys. Lett. **101** (2012) 012107. [18]J. Lee, *et al.*, J. Appl. Phys. **112** (2012) 014902. [19]X. Xie, *et al.*, ACS Nano **6** (2012) 10267. [20]K.L. Grosse, *et al.*, Appl. Phys. Lett. **102** (2013) 193503. [21]T. Kato, *et al.*, Jpn. J. Appl. Phys. **44** (2005) 7340.

BIOGRAPHY: Eric Pop is an Associate Professor of Electrical Engineering (EE) at Stanford University. Previously he was with the University of Illinois Urbana-Champaign, first as an Assistant Professor (2007-12) and then as an Associate Professor (2012-13) of Electrical & Computer Engineering. His research interests are in energy efficient electronics, novel 2-D and 1-D devices and materials, nanoscale energy conversion and harvesting. He received his Ph.D. in EE from Stanford (2005), the M.Eng./B.S. in EE and B.S. in Physics from MIT. Between 2005-2007 he did post-doctoral work at Stanford and worked at Intel on non-volatile memory. He received the Presidential Early Career (PECASE) Award from the White House (2010) and Young Investigator Awards from the ONR (2010), NSF (2010), AFOSR (2010) and DARPA (2008). He is an IEEE Senior member and currently serves as the Technical Program Chair of the IEEE Device Research Conference (DRC).