Applied Physics Letters ARTICLE scitation.org/journal/apl

Engineering thermal and electrical interface properties of phase change memory with monolayer MoS_2

Cite as: Appl. Phys. Lett. **114**, 082103 (2019); doi: 10.1063/1.5080959
Submitted: 12 November 2018 · Accepted: 7 February 2019 ·
Published Online: 27 February 2019

Christopher M. Neumann,[1] Kye L. Okabe,[1] Eilam Yalon,[1,2] Ryan W. Grady,[1] H.-S. Philip Wong,[1] and Eric Pop[1,3,a)]

AFFILIATIONS

[1]Department of Electrical Engineering, Stanford University, Stanford, California 94305, USA
[2]Department of Electrical Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel
[3]Department of Materials Science and Engineering, Stanford University, Stanford, California 94305, USA

[a)]Author to whom correspondence should be addressed: epop@stanford.edu

ABSTRACT

Phase change memory (PCM) is an emerging data storage technology; however, its programming is thermal in nature and typically not energy-efficient. Here, we reduce the switching power of PCM through the combined approaches of filamentary contacts and thermal confinement. The filamentary contact is formed through an oxidized TiN layer on the bottom electrode, and thermal confinement is achieved using a monolayer semiconductor interface, three-atom thick MoS_2. The former reduces the switching volume of the phase change material and yields a 70% reduction in reset current versus typical 150 nm diameter mushroom cells. The enhanced thermal confinement achieved with the ultra-thin (~6 Å) MoS_2 yields an additional 30% reduction in switching current and power. We also use detailed simulations to show that further tailoring the electrical and thermal interfaces of such PCM cells toward their fundamental limits could lead up to a sixfold benefit in power efficiency.

Phase change memory (PCM) is an emerging storage class memory technology, wherein the typical cell consists of a chalcogenide-based phase change material (commonly $Ge_2Sb_2Te_5$) contacted by top and bottom electrodes (TE and BE, respectively). The metal electrodes are used to apply voltage or current pulses to the phase change material, inducing its (reversible) transformation between amorphous and crystalline phases.[1] The phases possess electrical resistivity differing by up to four orders of magnitude, which can be read out to store binary or analog logic states (i.e., by gradual or partial programming) for neuromorphic applications.[2,3]

While PCM is already being used in computing systems, concerns remain over its relatively high reset current and power. To mitigate this problem, two approaches have often been taken: (1) reducing the volume of the phase change material or (2) improving the thermal confinement of the cell.[1] Reducing the switching volume entails scaling the contact area, either lithographically or by using nanoscale electrodes such as oxide filaments,[4–6] carbon nanotubes,[7–9] or graphene edge contacts.[10] Improving the thermal confinement requires trapping Joule heat by confining current flow through the phase change material,[11] using more thermally resistive materials,[12,13] or adding interfacial layers to the electrode contacts.[14–17] In earlier experiments, two-dimensional (2D) materials, such as graphene, have already been used as interfacial layers.[14,17] However, the high (lateral) thermal and electrical conductivities of graphene present some trade-offs in the attempt to confine heating.[14]

In this work, we combine enhancements from both categories above to demonstrate power-efficient PCM cells. To improve thermal confinement, we use monolayer molybdenum disulfide (MoS_2) grown by large-scale chemical vapor deposition (CVD).[18] Inserting this three-atom-thick (~6.15 Å) layer at the interface between the BE and the phase change material limits heat loss through this interface. To reduce the contact area, we form a narrow metal-oxide filament in the thin oxide on top of the TiN BE.[4–6] We also use finite element modeling to gain additional insights into the

benefits of further optimizing such interface modifications, toward fundamental physical limits, for power-efficient PCM.

We fabricate filamentary BE PCM with a MoS_2 interfacial layer [Fig. 1(a)], as well as three types of control devices: (1) filamentary BE PCM with graphene as an interfacial layer, (2) filamentary BE PCM without a 2D material, and (3) conventional PCM of the same dimensions with neither a filament nor a 2D layer. For all device types, we start with planarized TiN BEs which are ~150 nm in diameter. For filamentary devices, the BEs were first cleaned through an Ar sputtering process before being exposed to water for at least 1 h to ensure that a thin oxide layer forms. For devices with 2D materials, we use polymer-assisted wet transfer techniques[19,20] to place the 2D material layer on the BE substrate and pattern it using e-beam lithography. The patterned area covers the entire BE and ~25 nm past the BE edges to account for the overlay placement margin. Next, we DC sputter and lift-off 30 nm $Ge_2Sb_2Te_5$ (GST) capped *in situ* with 20 nm TiN. We previously showed that such GST depositions can be done with minimal damage to graphene.[14] Figure 1(b) shows Raman spectroscopy data with a strong MoS_2 signal both before and after GST deposition, confirming that monolayer MoS_2 can also withstand this process, and Figs. 1(c) and 1(d) display transmission electron microscopy (TEM) cross-sections showing the interfacial MoS_2 layer. For devices without filaments, the GST/TiN layer is deposited *in situ* after the Ar sputter cleaning to prevent any native oxide formation. After GST/TiN lift-off, we pattern and lift-off an additional 20 nm sputtered TiN and 40 nm Pt to form the probe pads and top electrode. Finally, devices are annealed in air at 180 °C for at least one hour to crystallize the GST layer into the fcc phase.

In samples with oxidized BEs, we must first form a filament before we see memory operation. This is done by applying 1/50/1 ns rise/width/fall pulses of increasing bias until the filament is formed, and we observe a current of >200 μA during the pulse. In the more than 50 measured devices, this forming occurs between 1.5 and 2 V. To demonstrate the filamentary nature of these oxidized BEs, we measured devices of various BE sizes with and without oxidation. Figure 2(a) shows that the current required to reset devices without oxide increases with the increasing BE area, as is expected for a PCM device. However, in Fig. 2(b), devices with the oxide layer show no dependence on the BE area and exhibit significantly lower reset currents. Given this trend, we conclude that in devices with an oxide layer, the effective BE size is related to the filament's size rather than the physical BE size.

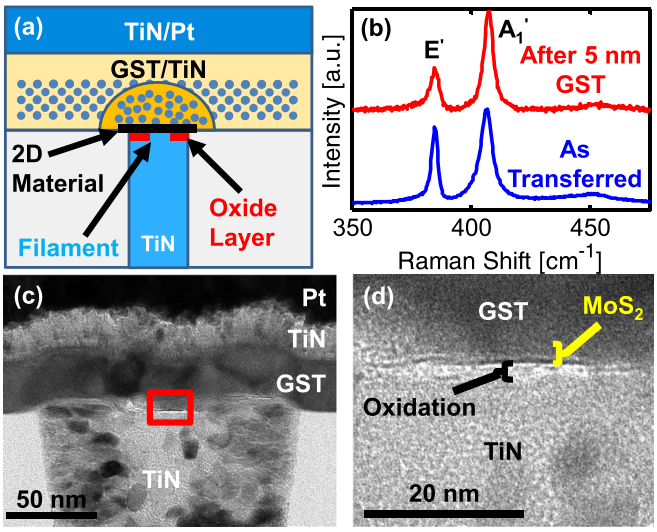

FIG. 1. (a) Cross-sectional schematic showing the position of the oxide filament and 2D material layers in a PCM device. (b) Monolayer MoS_2 Raman spectra before (bottom, blue) and after (top, red) 5 nm deposition of GST, probed with a 532 nm laser. The strong signal after GST sputtering indicates that MoS_2 remains present. The A_1' peak shows little change in position (405.9 cm^{-1} before, 406.3 cm^{-1} after) and full width half maximum (FWHM) (6.0 cm^{-1} before, 5.8 cm^{-1} after). The E′ peak position is also unchanged (384.9 cm^{-1} before, 385.0 cm^{-1} after), but the FWHM is broadened (3.5 cm^{-1} before, 5.2 cm^{-1} after). The peak center and FWHM are extracted from peak fits. (c) TEM cross-sections of a PCM device fabricated for this work, with the MoS_2 interfacial layer. The BE diameter is ~150 nm. (d) Zoomed in image of the red boxed area in (c) showing a ~2 nm oxidation layer on TiN, just below the monolayer of MoS_2.

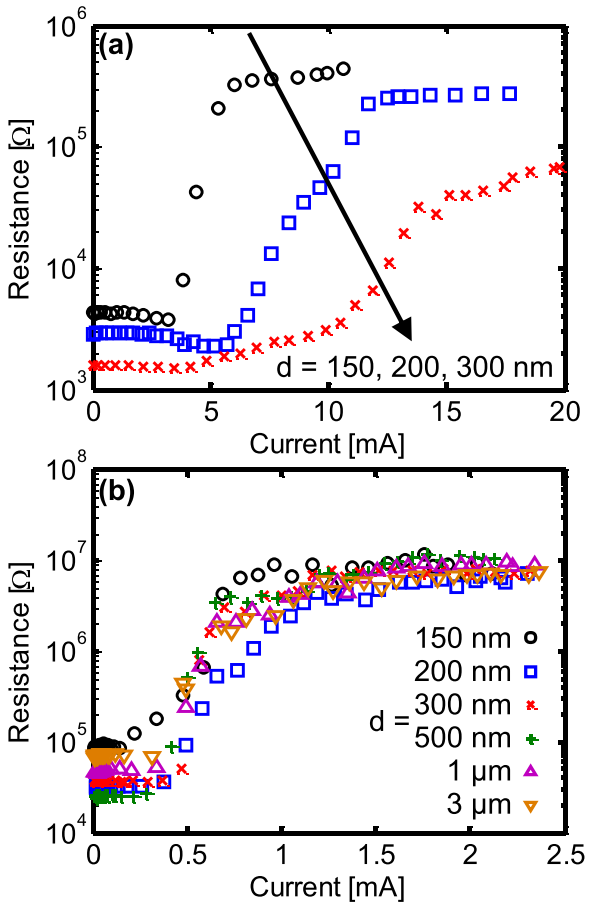

FIG. 2. DC read resistance vs. pulse current magnitude for various nominal BE diameters, *d*, without (a) and with (b) thin oxide layers on the BE. Cells without the oxide layer show expected scaling behavior, whereas those with oxidized BE have no dependence on BE diameter, indicating filamentary conduction through the oxide.

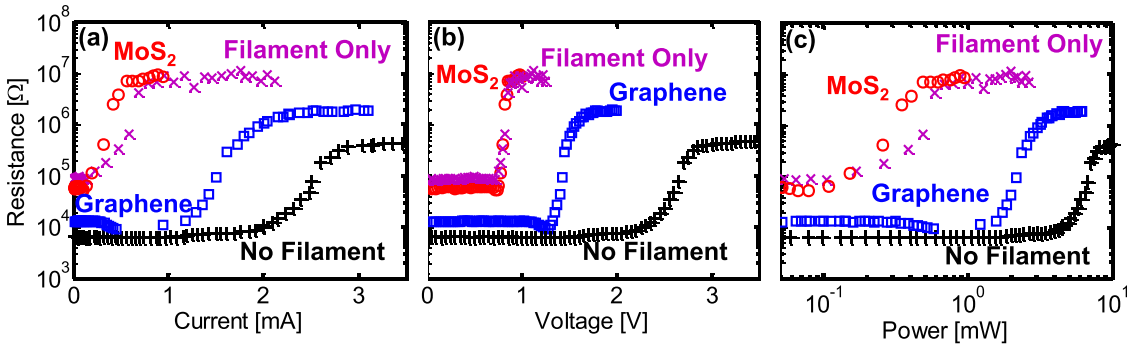

FIG. 3. DC read resistance vs. (a) current, (b) voltage, and (c) power during the reset for the PCM with a MoS_2 interfacial layer and an oxide filament (red circle) and control devices: oxide filament only (purple x), graphene interfacial layer and oxide filament (blue square), and without the oxide filament (black cross). Note that devices with the MoS_2 interfacial layer show more than 30% reduction in switching current and power compared to devices with a filament only.

We measure the resistance of all devices with a 50 mV DC bias. For set and reset programming, we use 1/50/2000 ns and 1/50/1 ns rise/width/fall pulses, respectively. The general layout of the measurement system and endurance information can be found in the supplementary material. Next, we compare the reset current of the filament-contacted devices that have a MoS_2 interfacial layer with the control devices. To do this, we first cycle all measured devices at least 1000 times to ensure consistent and reliable operation. Subsequently, cells are set into the low resistance state. We apply a series of increasing amplitude reset pulses to the device and capture the transient current through an oscilloscope. To obtain the current, we measure the voltage across the 50 Ω input of the oscilloscope. We note that, where a pulse current is given, we refer to the peak current rather than the average. For peak power calculations, we assume nominal applied voltage and peak current.

In Fig. 3, we plot the peak current, nominal voltage, and peak power required to reset devices of each type. Comparing filament-contacted devices with and without the MoS_2 layer, we see about a 30% reduction in the current, but the reset voltage remains constant. We attribute this reduction to the additional electrical and thermal resistances at the BE-GST interface. Conversely, for the graphene-interfaced PCM, we see a significant increase in the switching current. This is similar to observations in previous work[14] where a carefully patterned graphene layer gave a current reduction, but a graphene area larger than the BE led to larger programming current. In our devices, the graphene is patterned to be slightly larger than the BE, but the filament's effective area is much smaller. Because graphene has relatively high in-plane thermal and electrical conductivities, it acts as a "spreader" for the heat and current. However, as shown in Fig. 4, MoS_2 has significantly lower (~20×) in-plane thermal conductivity than graphene. In addition, the electrical conductivity of undoped MoS_2 is orders of magnitude lower than graphene, and this 2D layer can effectively be considered an insulator.[21,22] Because of this, the cell power-efficiency can be improved even when MoS_2 is not precisely patterned.

We performed finite-element simulations to examine how much of an impact we can expect by modifying the thermal and electrical interfaces of a typical mushroom cell device. The key is to determine the expected power benefits of adding a 2D material layer and also to understand the "ultimate limits" of a perfect interface material with ideal thermal and electrical properties. We assume a device structure similar to our fabricated design with 100 nm diameter TiN BE and 30 nm thick GST. The top metal stack is identical with 40 nm TiN and Pt. To simplify the simulation, we chose not to include the filamentary structure, and the 2D material layer is modeled as a simple increase in the electrical and thermal resistances of the interface. We take into account the temperature-dependent electrical contact resistance, $\rho_c(T) = \rho(T)t_{eff}$, by tying it to an effective GST thickness, t_{eff}, where the electrical resistivity of GST, $\rho(T)$, is obtained from Ref. 23. In other words, an effective GST thickness of 10 nm would give an interface that has the same electrical resistance as 10 nm GST at that temperature. The temperature dependence

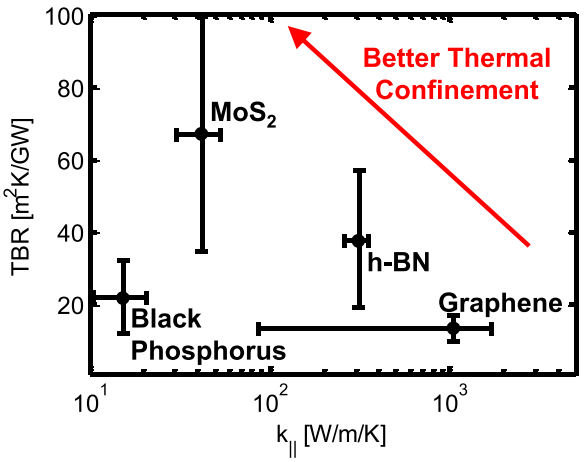

FIG. 4. Comparing the thermal properties of some common 2D materials, where experimental data are available: range of measured thermal boundary resistance (TBR) vs. in-plane thermal conductivity ($k_{||}$),[28–37] near room temperature. The lower end of the $k_{||}$ range corresponds to confined samples (i.e., graphene nanoribbons)[37] or to samples with higher defect density. Overall, better thermal confinement for PCM could be achieved using 2D material interfaces with lower $k_{||}$ and higher TBR, e.g., 2D materials with heavier atomic masses (like WTe_2)[38] or with 2D material heterostructures.[26,27] At the higher PCM operating temperatures, the TBR and $k_{||}$ will likely be lower due to increased phonon population and scattering, respectively.

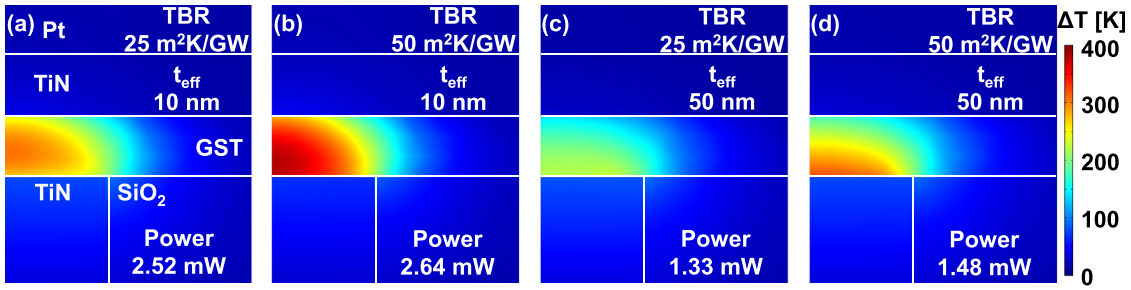

FIG. 5. Simulated temperature rise (ΔT) for a 100 nm BE after 50 ns of 1.2 V bias for devices with (a) nominal boundary resistances as listed, (b) increased thermal boundary resistance (TBR), (c) increased electrical boundary resistance (as t_{eff} of GST), and (d) increased both electrical and thermal boundary resistances. The values of TBR and the effective thickness of electrical resistance (t_{eff}) are listed in each panel. Increasing TBR improves thermal confinement in the phase change layer, while increasing the electrical contact resistance raises the power density at the interface and generates more heat per input power.

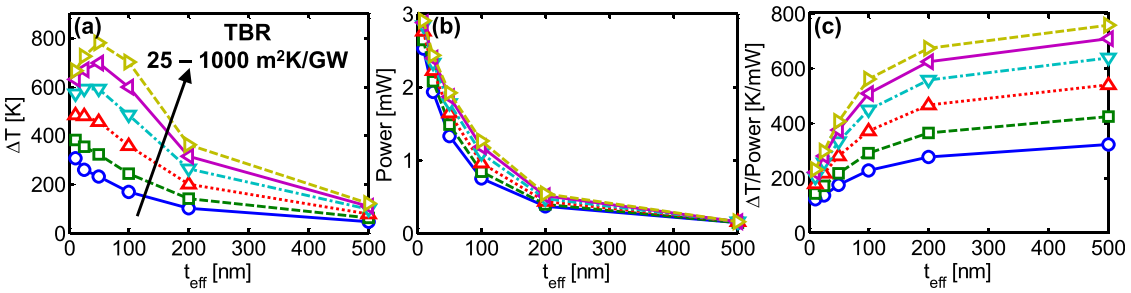

FIG. 6. (a) Maximum temperature rise and (b) power vs. effective thickness of electrical resistance (t_{eff}) for a 100 nm wide BE with varying thermal boundary resistances (TBR) under identical bias conditions (1.2 V for 50 ns). (c) Temperature rise per input power, corresponding to the thermal resistance of the device. Increasing both the electrical and thermal boundary resistances improves the heating efficiency (as $\Delta T/P$) of such devices.

of the thermal boundary resistance (TBR) is not taken into account as it is expected to be relatively negligible for MoS_2[24] above room temperature.

In Fig. 5, we show the simulated cross-sectional temperature profiles of the device at the end of a 1.2 V, 50 ns pulse. For a given voltage, we note that while increasing the thermal resistance of the interface results in a higher temperature, increasing the electrical resistance reduces the temperature due to the lower current. It is important to note that, for simplicity, we are assuming that the PCM is the dominant electrical resistance in our "circuit." However, if we increase both boundary resistances, we see that the temperature is similar. In Figs. 6(a) and 6(b), we plot the maximum temperature change and current against a wide range of GST-BE interface parameters. We then calculate the relative power efficiency of the cell by normalizing the temperature increase with the power expended, as shown in Fig. 6(c). From these results, we see that increasing the thermal or electrical resistance can provide a 2× improvement in efficiency, and maximizing both shows a 6× benefit.

While it may be expected that thermal and electrical resistances should be maximized to increase the power efficiency, there are both consequences and limits to doing so. Large increases in electrical resistivity will increase the switching voltage (requiring higher voltage transistors and introducing additional array-level power consumption) and decrease the memory window (limiting multi-level cell designs). In the case of increased thermal resistivity, we need to consider two possible trade-offs: the effect that thermal confinement has on the quenching process and the finite TBR of a single 2D material interface. For our mushroom cell, it is impractical to reach the quenching limit[25] as heat will dissipate through the TE and bulk GST even if the BE interface is very thermally resistive (details can be found in section S3 of the supplementary material). While a thicker layer could be engineered to have very high thermal resistance, some heat generated in this layer would be trapped further from the phase change region. Therefore, we expect that using layered 2D heterostructures (e.g., a MoS_2/WTe_2 bilayer) will further increase the TBR while preserving the atomic thinness of the interface. These van der Waals heterostructures have been suggested to have highly tunable properties[26] in addition to high TBR which exceeds that of bulk material interfaces.[27] However, future work must consider the effect of the relatively high (when compared to bulk insulators) in-plane thermal conductivity of these materials as well.

In conclusion, we demonstrated PCM devices which combine the area reduction of a metal-oxide filament with the thermal confinement of monolayer MoS_2 at the BE-GST interface. These reduce the switching current by 70% and an additional 30%, respectively, compared to control devices. Finite element simulations illustrate that increasing electrical or thermal boundary resistances individually can improve the power efficiency of typical PCM by more than 2× and by 6× if both are

increased up to near the fundamental limits. However, reaching these enhanced thermal boundary resistances will require novel approaches, such as the use of 2D heterostructures at the interface.

See supplementary material for additional information on the device measurement setup and transient response.

This work was performed at the Stanford Nanofabrication Facility (SNF) and Stanford Nano Shared Facilities (SNSF), supported by the National Science Foundation (NSF) as part of the NNCI under Award No. 1542152. This work was supported by member companies of the Stanford Non-volatile Memory Technology Research Initiative (NMTRI) and by the NSF EFRI 2-DARE under Grant No. 1542883. K.L.O. acknowledges funding from the Semiconductor Research Corporation (SRC) task 2826. R.W.G. acknowledges the support from the NSF Graduate Research Fellowship under Grant No. DGE-1656518.

REFERENCES

[1]S. W. Fong, C. M. Neumann, and H.-S. P. Wong, IEEE Trans. Electron Dev. **64**, 4374 (2017).
[2]A. Sebastian, M. Le Gallo, G. W. Burr, S. Kim, M. BrightSky, and E. Eleftheriou, J. Appl. Phys. **124**, 111101 (2018).
[3]S. R. Nandakumar, M. Le Gallo, I. Boybat, B. Rajendran, A. Sebastian, and E. Eleftheriou, J. Appl. Phys. **124**, 152135 (2018).
[4]B. J. Choi, S. H. Oh, S. Choi, T. Eom, Y. C. Shin, K. M. Kim, K.-W. Yi, C. S. Hwang, Y. J. Kim, H. C. Park, T. S. Baek, and S. K. Hong, J. Electrochem. Soc. **156**, H59 (2009).
[5]Q. Hubert, C. Jahan, A. Toffoli, L. Perniola, V. Sousa, A. Persico, J.-F. Nodin, H. Grampeix, F. Aussenac, and B. de Salvo, in *Proceedings of the European Solid-State Device Research Conference*, Helsinki, Finland (2011), pp. 95–98.
[6]C. Xu, Z. Song, B. Liu, S. Feng, and B. Chen, Appl. Phys. Lett. **92**, 062103 (2008).
[7]J. Liang, R. G. D. Jeyasingh, H.-Y. Chen, and H.-S. P. Wong, IEEE Trans. Electron Dev. **59**, 1155 (2012).
[8]F. Xiong, A. D. Liao, D. Estrada, and E. Pop, Science **332**, 568 (2011).
[9]F. Xiong, M.-H. Bae, Y. Dai, A. D. Liao, A. Behnam, E. A. Carrion, S. Hong, D. Ielmini, and E. Pop, Nano Lett. **13**, 464 (2013).
[10]A. Behnam, F. Xiong, A. Cappelli, N. C. Wang, E. A. Carrion, S. Hong, Y. Dai, A. S. Lyons, E. K. Chow, E. Piccinini, C. Jacoboni, and E. Pop, Appl. Phys. Lett. **107**, 123508 (2015).
[11]Y. N. Hwang, S. H. Lee, S. J. Ahn, S. Y. Lee, K. C. Ryoo, H. S. Hong, H. C. Koo, F. Yeung, J. H. Oh, H. J. Kim, W. C. Kim, W. C. Jeong, J. H. Park, H. Horii, Y. H. Ha, J. H. Yi, G. H. Koh, G. T. Jeong, H. S. Jeong, and K. Kim, in *IEDM Technical Digest*, Washington, DC, USA (2003), pp. 37.1.1–37.1.4.
[12]S. W. Fong, C. M. Neumann, E. Yalon, M. M. Rojo, E. Pop, and H. S. P. Wong, IEEE Trans. Electron Dev. **64**, 4496 (2017).
[13]J. Y. Wu, M. Breitwisch, S. Kim, T. H. Hsu, R. Cheek, P. Y. Du, J. Li, E. K. Lai, Y. Zhu, T. Y. Wang, H. Y. Cheng, A. Schrott, E. A. Joseph, R. Dasaka, S. Raoux, M. H. Lee, H. L. Lung, and C. Lam, in *IEDM Technical Digest*, Washington, DC, USA (2011), pp. 3.2.1-3.2.4.
[14]C. Ahn, S. W. Fong, Y. Kim, S. Lee, A. Sood, C. M. Neumann, M. Asheghi, K. E. Goodson, E. Pop, and H.-S. P. Wong, Nano Lett. **15**, 6809 (2015).
[15]R. Feng, S. Zhitang, G. Yuefeng, W. Liangcai, F. Songlin, and C. Bomy, Nanotechnology **19**, 445706 (2008).
[16]C. Kim, D.-S. Suh, K. H. P. Kim, Y.-S. Kang, T.-Y. Lee, Y. Khang, and D. G. Cahill, Appl. Phys. Lett. **92**, 013109 (2008).
[17]C. Zhu, J. Ma, X. Ge, F. Rao, K. Ding, S. Lv, L. Wu, and Z. Song, Appl. Phys. Lett. **108**, 252102 (2016).
[18]K. K. H. Smithe, C. D. English, S. V. Suryavanshi, and E. Pop, 2D Mater. **4**, 011009 (2017).
[19]Y.-C. Lin, C. Jin, J.-C. Lee, S.-F. Jen, K. Suenaga, and P.-W. Chiu, ACS Nano **5**, 2362 (2011).
[20]K. K. H. Smithe, A. V. Krayev, C. S. Bailey, H. R. Lee, E. Yalon, Ö. B. Aslan, M. Muñoz Rojo, S. Krylyuk, P. Taheri, A. V. Davydov, T. F. Heinz, and E. Pop, ACS Appl. Nano Mater. **1**, 572 (2018).
[21]N. C. Wang, E. A. Carrion, M. C. Tung, and E. Pop, Appl. Phys. Lett. **110**, 223106 (2017).
[22]K. K. H. Smithe, S. V. Suryavanshi, M. Muñoz Rojo, A. D. Tedjarati, and E. Pop, ACS Nano **11**, 8456 (2017).
[23]K. Cil, F. Dirisaglik, L. Adnane, M. Wennberg, A. King, A. Faraclas, M. B. Akbulut, Y. Zhu, C. Lam, A. Gokirmak, and H. Silva, IEEE Trans. Electron Dev. **60**, 433 (2013).
[24]E. Yalon, Ö. B. Aslan, K. K. H. Smithe, C. J. McClellan, S. V. Suryavanshi, F. Xiong, A. Sood, C. M. Neumann, X. Xu, K. E. Goodson, T. F. Heinz, and E. Pop, ACS Appl. Mater. Interfaces **9**, 43013 (2017).
[25]R. Jeyasingh, S. W. Fong, J. Lee, Z. Li, K.-W. Chang, D. Mantegazza, M. Asheghi, K. E. Goodson, and H.-S. P. Wong, Nano Lett. **14**, 3419 (2014).
[26]K. Kim, J. He, B. Ganeshan, and J. Liu, J. Appl. Phys. **124**, 055104 (2018).
[27]S. Vaziri, E. Yalon, M. Muñoz Rojo, S. V. Suryavanshi, C. McClellan, C. S. Bailey, A. J. Gabourie, V. Chen, S. Deshmukh, K. K. H. Smithe, and E. Pop, in *MRS Fall Meeting*, Boston, MA, USA, EP03.17.02 (2018).
[28]J. P. Feser and D. G. Cahill, Rev. Sci. Instrum. **83**, 104901 (2012).
[29]H. Jang, J. D. Wood, C. R. Ryder, M. C. Hersam, and D. G. Cahill, Adv. Mater. **27**, 8017 (2015).
[30]I. Jo, M. T. Pettes, J. Kim, K. Watanabe, T. Taniguchi, Z. Yao, and L. Shi, Nano Lett. **13**, 550 (2013).
[31]J. Liu, G.-M. Choi, and D. G. Cahill, J. Appl. Phys. **116**, 233107 (2014).
[32]Z. Luo, J. Maassen, Y. Deng, Y. Du, R. P. Garrelts, M. S. Lundstrom, P. D. Ye, and X. Xu, Nat. Commun. **6**, 8572 (2015).
[33]E. Pop, V. Varshney, and A. K. Roy, MRS Bull. **37**, 1273 (2012).
[34]X. Li, Y. Yan, L. Dong, J. Guo, A. Aiyiti, X. Xu, and B. Li, J. Phys. D: Appl. Phys. **50**, 104002 (2017).
[35]E. Yalon, C. J. McClellan, K. K. H. Smithe, M. Muñoz Rojo, R. L. Xu, S. V. Suryavanshi, A. J. Gabourie, C. M. Neumann, F. Xiong, A. B. Farimani, and E. Pop, Nano Lett. **17**, 3429 (2017).
[36]R. Yan, J. R. Simpson, S. Bertolazzi, J. Brivio, M. Watson, X. Wu, A. Kis, T. Luo, A. R. Hight Walker, and H. G. Xing, ACS Nano **8**, 986 (2014).
[37]M.-H. Bae, Z. Li, Z. Aksamija, P. N. Martin, F. Xiong, Z.-Y. Ong, I. Knezevic, and E. Pop, Nat. Commun. **4**, 1734 (2013).
[38]M. J. Mleczko, R. L. Xu, K. Okabe, H.-H. Kuo, I. R. Fisher, H.-S. P. Wong, Y. Nishi, and E. Pop, ACS Nano **10**, 7507 (2016).

Supplementary Material

Engineering Thermal and Electrical Interface Properties of Phase Change Memory with Monolayer MoS_2

C. M. Neumann,[1] K. L. Okabe,[1] E. Yalon,[1,2] R. W. Grady,[1] H.-S. P. Wong,[1] and E. Pop[1,3]

[1]*Dept. of Electrical Engineering, Stanford University, Stanford, California 94305, USA*
[2]*Dept. of Electrical Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel*
[3]*Dept. of Materials Science and Engineering, Stanford University, Stanford, California 94305, USA*

S1. Measurement Setup and Cycling

The electrical measurement setup has two key functions: pulsed writing and DC reading. For pulsed switching and current measurements, we use an Agilent 81160A Function Generator (FG) to produce the pulses and an Agilent MSO7104A Mixed Signal Oscilloscope to measure the current. To measure the read resistance for all devices, we apply a 50 mV DC bias with an Agilent B1500A Semiconductor Parameter Analyzer (SPA). Finally, a Keithley 707B Switch Matrix is used to switch between these pieces of equipment as part of the write/read process. The general layout of the measurement system is shown in Fig. S1(a). This setup was used for both endurance testing (as shown in Fig. S1(b)) as well as transient measurements (see section S2).

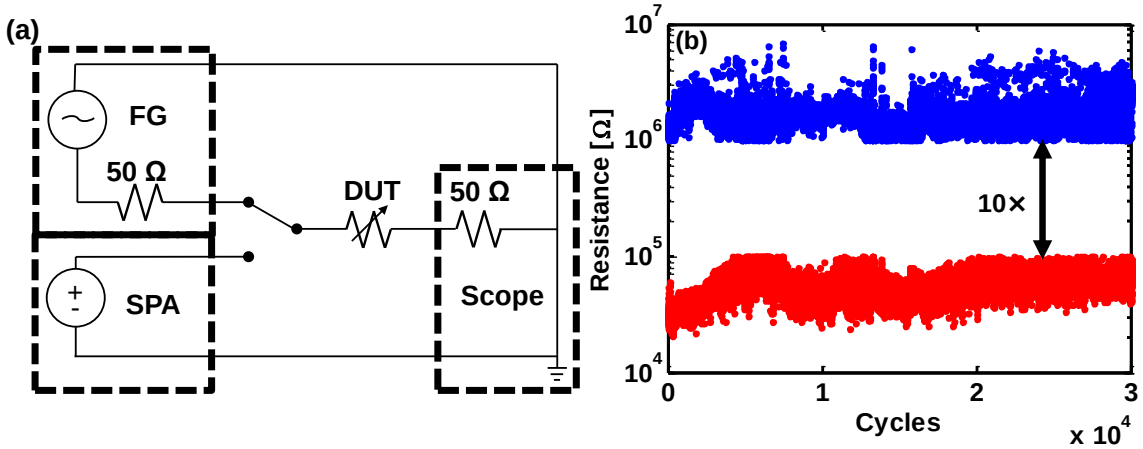

Fig. S1. (a) The measurement setup layout using a function generator (FG), semiconductor parameter analyzer (SPA), and oscilloscope (Scope). A switch matrix controls the connection to alternate between pulsed and DC measurement. We note that the FG and oscilloscope both have a 50 Ω internal impedance which is in series with the resistance of the measured device. **(b)** Cycling of a PCM device with MoS_2 and oxide filament showing that a 10× resistance window can be maintained for more than 30k cycles using a write-verify scheme.

S2. Transient Behavior

While performing pulsed switching on devices, we measure the full transient response across the 50 Ω input on the "downstream" oscilloscope. This allows us to see how the devices behave over time as well as determine the transient resistance (i.e. the resistance of the device during the pulse), which is much lower than the DC read resistance. In Fig. S2(a), we compare example pulse responses from filament-contacted PCM devices with and without MoS_2 under identical pulse conditions (1 V, 1/50/1 ns rise/width/fall time). Looking at these measurements, we note two major differences.

First, the current in the PCM device with MoS_2 interfacial layer is significantly lower. Given that the voltage is constant, we can attribute this change to the resistivity added by the MoS_2 layer. This is consistent with Fig. S2(b) which shows that the MoS_2 device has a higher transient resistance for all pulsed bias conditions. It should also be noted that this is the same set of pulse data shown in Fig. 3 of the main text, where the filament-contacted PCM device without MoS_2 is shown to have *higher* DC read resistance before/after pulsing.

Second, although the pulse conditions and measurement setup are identical, the oscilloscope appears to measure a shorter pulse for the PCM device with MoS_2. In this case, it is important to recall that the oscilloscope is placed downstream from the device and will only measure the current is able to conduct through that device. Also, both devices were already in the reset state prior to the pulse and are very resistive initially. Therefore, we interpret this delay in current to be the time it takes for the GST region to melt and become conducting.

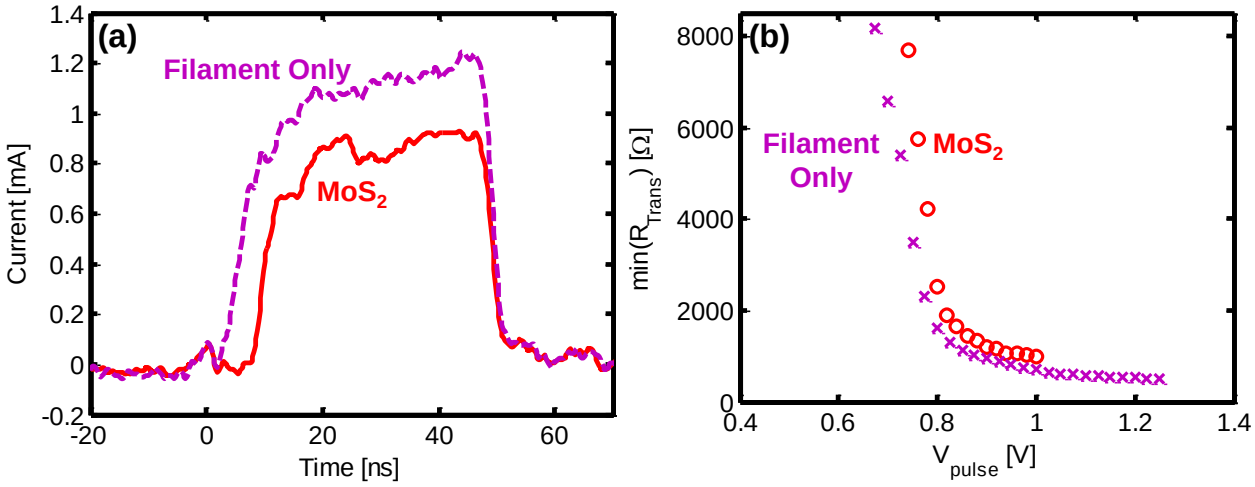

Fig. S2. (a) The current measured during a typical reset pulse (1 V, 1/50/1 ns rise/width/fall time). **(b)** By finding the maximum current point during each pulse, we are able to calculate the minimum transient resistance of each device.

S3. Quenching Simulation

When attempting to reset a PCM device, it is critical for the PC material to quickly "quench" in order to avoid recrystallization. While improved thermal confinement is often the goal for increasing power efficiency, it is important to make sure that a device is not so well confined that it is no longer able to quench. Because of this concern, we expanded on the finite element simulations in the main text to study the cooling process in a device that has high TBR (500 m^2K/GW) and t_{eff} (100 nm). In Fig. S3, we plot the maximum temperature of a cell that, after being biased at 1.2V for 50 ns, has the voltage abruptly removed. From this we see that, even at very high TBR, the device cools to below the crystallization temperature within 5 ns, sufficiently faster than what is needed to quench the device.[1] However, while quenching is not a concern for this geometry, it is important to note that an extremely well confined device may still be able to reach the quenching limit and should be considered when designing a PCM cell.

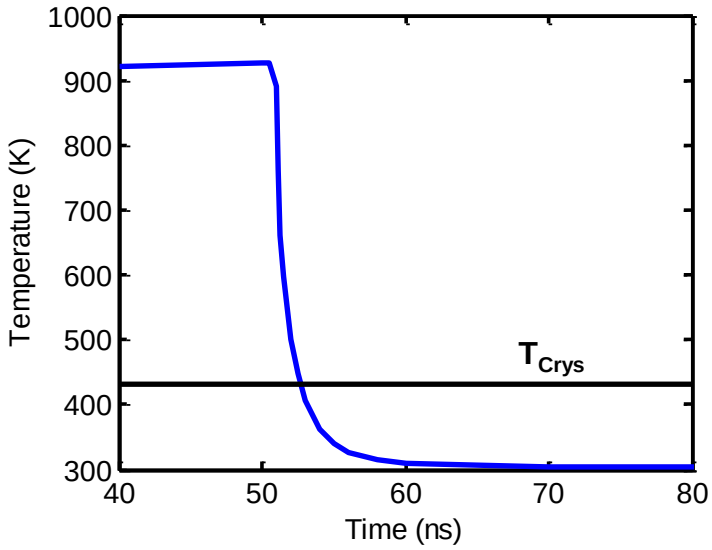

Fig. S3. Peak temperature over time for a device which undergoes a 1.2 V pulse, turned off at 50 ns. Even at the hottest point, the temperature drops below the crystallization temperature within 5 ns.

Citations

[1]R. Jeyasingh, S. W. Fong, J. Lee, Z. Li, K.-W. Chang, D. Mantegazza, M. Asheghi, K. E. Goodson, and H.-S. P. Wong, Nano Lett. **14**, 3419 (2014).