

# A Case Study on the Application of Real Phase-Change RAM to Main Memory Subsystem

Suknam Kwon, Dongki Kim, Youngsik Kim, Sungjoo Yoo, Sunggu Lee

Department of Electrical Engineering

Pohang University of Science and Technology (POSTECH)

{redstorm113, dongki.kim, kaengsik, sungjoo.yoo, slee}@postech.ac.kr

## ABSTRACT

*Phase-change RAM (PCM) has the advantages of better scaling and non-volatility compared with the DRAM which is expected to face its scaling limit in the near future. There have been many studies on applying the PCM to main memory in order to complement or replace the DRAM. One common limitation of these studies is that they are based on synthetic PCM models. In our study, we investigate the feasibility and issues of applying a real PCM to main memory. In this paper, we report our case study of characterizing the PCM and evaluating its usefulness in the main memory. Our results show that the PCM/DRAM hybrid main memory with a modest DRAM size can give comparable performance to that of the DRAM only main memory. However, the hybrid memory with small DRAMs or large footprint programs can suffer from performance degradation due to the long latency of both PCM writes and write preemption penalty, which requires architectural innovations for exploiting the full potential of PCM write performance.*

## 1. Introduction

DRAM is expected to face scaling limit in sub-20nm technology [1]. There have been presented new emerging memories, phase-change RAM (PCM), spin-torque transfer (STT) RAM, memristor, etc. Among them, the PCM is a promising candidate to compensate for the limitation of DRAM in the near future in the form of hybrid PCM/DRAM and, possibly, to replace DRAM (in some applications). Compared with the other candidates of emerging memory, the PCM technology is much more mature and already has the capability of large capacity and mass production [2][3][4].

The advantages of PCM are scalability (a 5nm implementation is already demonstrated [5]) and non-volatility. However, it has limitations: long read ( $\sim 4X$  DRAM latency) and write ( $>10X$ ) latency, high power consumption in the read ( $>2X$ ) and write ( $>10X$ ) operations, and poor write endurance ( $10^8$  vs.  $10^{15}$ ) compared with the DRAM. Recently, there have been active studies to overcome the limitations in applying the PCM to main memory subsystem in [6][7][8][9][10][11][12]. Most of the previous works utilize synthetic PCM models which assume simplistic architectures, e.g., one shot write back from the row buffer to the data array [6], write preemption without preemption penalty [7], etc.

In this paper, we present a study based on a real PCM. We use Samsung 1Gb PCM [3]. Our primary goal is to investigate the feasibility of state-of-the-art PCM for main memory applications covering the PCM/DRAM hybrid memory as well as the PCM only main memory. We demonstrate that the PCM/DRAM hybrid main memory gives comparable performance to that of the DRAM only counterpart. In addition, we quantify the impact of long write

latency of PCM on the performance of PCM-based main memory subsystem since the write latency of real PCM is much longer than that of the synthetic PCM models used in previous works. We also demonstrate that the write preemption for read latency reduction [8] needs to be applied considering non-negligible preemption penalty.

This paper is organized as follows. Section 2 introduces PCM and reviews related work. Section 3 explains our characterization system and key chip-level characterization results. Section 4 reports our system-level simulation results. Section 5 gives lessons learned from the characterization and system-level simulation. Section 6 concludes the paper.

## 2. PCM Operations and Previous Works

### 2.1 PCM Operations

The PCM stores information by modifying the phase of phase-change material between amorphous (with high resistance) and crystalline (with low resistance) states. The PCM cell consists of phase change material (typically, GST), heater and access transistor (or diode). The resistance (i.e., stored information) of phase change material can be read by applying voltage between the two terminals of the material and measuring the current level. The write operation requires heating the phase change material at different temperature levels:  $\sim 600^\circ\text{C}$  for amorphous (i.e., reset) state and  $\sim 300^\circ\text{C}$  for crystalline (i.e., set) states. Compared with the conventional DRAM, the PCM has asymmetry in read and write operations. The write operation takes longer latency and higher power consumption than the read operation. In the write operations, set and reset have different levels of latency (longer latency for set) and power consumption (higher peak power for reset).

After a write operation is completed at a PCM cell, the resistance still changes mainly due to two reasons. Just after the write operation is finished (which takes about 100~200ns), the resistance value drifts (typically toward high resistance value) and is stabilized after a latency called *R drift latency* (at least, order of 10s of  $\mu\text{s}$ ) [13]. Thus, the total write latency is typically the sum of both cell write latency and R drift latency since the recently written PCM cell can be accessed only after R drift latency elapses since the termination of write operation. Typically, R drift latency (e.g., 10 $\mu\text{s}$ ) is much larger than cell write latency (e.g., 150ns). Thus, the R drift latency has a crucial impact on write performance as will be shown in our experiments. The other factor affecting dynamic resistance change is temperature. At high operating temperature, the phase change material undergoes re-crystallization thereby giving smaller resistance value.

The high power consumption of write operation in the PCM limits the data width of internal write operation to the data array, i.e., internal write bandwidth, and makes the write latency a function of write data size. For instance, in [14], the internal write operation allows only 16b data to be written to the data array at a time. Thus,

for instance, in order to write back a 32b data from the row buffer to the data array, two internal write operations (i.e., two times longer cell write latency) are required, which finally increases total write latency.

## 2.2 Previous Works on PCM in Main Memory

Most of previous works are focused on overcoming the poor characteristics of PCM write in terms of performance, power and endurance. In [15], a differential write method is proposed in order to write only updated bits by comparing the existing and new write data in a bit-by-bit manner. In [16], an invert coding called *flip-n-write* is proposed to reduce the number of bit updates by inverting data in case that more than half the bits need to be updated. In [11], Zhang and Li show a value dependency of PCM reliability that the (current level of) reset operation determines the lifetime of PCM cells. In [17], Lastras-Montano et al. present data encoding methods to minimize the number of reset operations. The latency of set operation is larger than that of reset operation. Thus, in [18], sets (resets) are first grouped and then a group of sets (resets) is written to the corresponding bit locations at the same time.

The lifetime of PCM device is determined by the first PCM cell which passes its write endurance limit. Thus, wear leveling is required to evenly distribute writes across PCM cells. Recently, several rotation-based wear leveling methods are presented at the granularity of cache line [7], page [10] and super-page [9]. In [19], Qureshi et al. present a start-gap and randomization method. In [20], Seong et al. present a security refresh method which dynamically applies a two-level randomization in order to cope with malicious processes, e.g., repeat address attacks.

Error correction is an important method to increase PCM lifetime even after the write endurance limit is reached and errors occur for some cells. In [21], Ipek et al. show a method which sacrifices capacity for memory lifetime by replicating data in multiple places. In [22], Schechter et al. propose utilizing an error correction pointer (ECP) based on the fact that PCM bit errors are persistent. In [23], Seong et al. present a method called SAFER which gives low area overhead for error correction exploiting the fact that the PCM errors are stuck-at faults.

In [6], Lee et al. show that the PCM alone, if its interface architecture is properly modified to mitigate long read/write latency, can give performance comparable to that of DRAM. In order to exploit the benefits of both PCM (non-volatility and large capacity) and DRAM (high performance), a PCM/DRAM hybrid memory is proposed in [7][9][10][12]. In [7], Qureshi et al. show that the PCM/DRAM hybrid memory consisting of a small DRAM and a large PCM gives performance comparable to that of large DRAM. In [12], Park et al. applies a decay concept to DRAM data in order to reduce DRAM refresh power in the hybrid memory.

The long write latency can degrade read performance when read requests are blocked by long-running write operations. In [8], Qureshi et al. present a write preemption method which preempts the current write operation in order to serve newly arrived read requests to the same PCM bank where the write operation is being performed. The write preemption has a potential to improve system performance which is typically sensitive to read latency. However, in reality, the write preemption has non-negligible latency penalty. In our system-level simulation in Section 4, we quantify the effect of non-negligible preemption penalty.

Recently, Akel et al. present a solid-state disk system based on real PCM chips having the NOR interface [24]. Our difference is that we focus on main memory applications with the PCM chip having the LPDDR2-N interface.

## 3. PCM Characterization

In this section, we explain the PCM chip used in our study and our characterization of data-dependent write latency.

### 3.1 PCM Chip

In our study, we utilize a commercial chip consisting of 3D-stacked 1Gb PCM and 512Mb DRAM [3]. Table 1 shows the characteristics of the PCM. The chip has the LPDDR2-N interface which has asymmetric read and write paths [25]. The read path is like DRAM. The PCM has four row buffers (each 32B) for read operations. The write path is like NOR and NAND Flash memory. The interface has a program buffer. For a write operation, we give the start address, write data size, and write data to the interface. On receiving the write initiation command, the PCM initiates the internal write operation to move data from the program buffer to the data array. The status of internal write operation is monitored by polling the status register. The write preemption penalty is 25 $\mu$ s as shown in Table 1, which means that in order to serve a new read request by preempting the current write operation, the read request has at least an additional latency of 25 $\mu$ s for the PCM chip to become ready to serve the read request.

Table 1 PCM characteristics

Number of banks, bit width	16 banks, 16b
$t_{RCD}$	80ns
Read/write latency	6/3 clocks @ 667MHz
Total write latency	25 $\mu$ s (for 2B~32B)
Write preemption latency	25 $\mu$ s
I/O bits and bandwidth	16b, up to 800Mbps (400MHz DDR)

### 3.2 Chip-level Characterization

We performed a PCM chip-level performance characterization. We utilized a commercial Verilog model of the PCM. In this subsection, we focus on the relationship between total write latency (=cell write latency + R drift latency) and write data size since the long write latency can have significant impacts on system performance when the main memory consists of only PCM chips or the PCM/DRAM hybrid memory has a small DRAM.

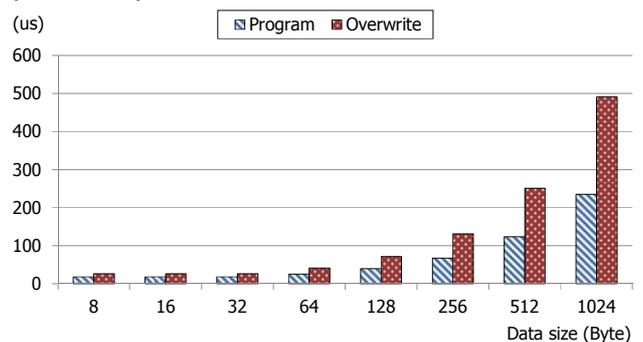


Figure 1 Write latency (program and overwrite) vs. data size

We developed in Verilog a PCM controller supporting the LPDDR2-N specification and performed RTL simulation in order to quantify the relationship between total write latency and write data size utilizing a testbench where we perform write operations at the same location while varying the data size with a fixed bit difference rate. Figure 1 shows the relationship between total write latency and write data size. The figure shows that there are two regions of write latency. When the write data size is up to 32B, the write (overwrite) latency is constant, about 25 $\mu$ s. The constant write latency for small data sizes is considered to be due to the R drift latency. As

mentioned previously, the total write latency is the sum of cell write latency and R drift latency. Note that the write latency (e.g., 25 $\mu$ s for 2B) is much larger than that used in synthetic PCM models (e.g., 150ns [6]) in previous works. The long write latency for short write data can significantly impact system performance. Thus, we investigate the effects of the long write latency of real PCM in our study as will be given in Section 4. In the second region for data sizes larger than 32B, the write latency has a linear relationship with the write data size (the slope of about 11 $\mu$ s/32B). The linear relationship is the expected behavior due to the limitation of internal write bandwidth as mentioned in Section 2.1.

## 4. System-level Evaluation

### 4.1 Evaluation Methodology

For the system-level simulation, we utilize an event-driven multi-core simulation environment called McSim [26]. It performs functional simulation of x86 programs based on the Pin environment [27]. The timing model covers x86 in-order cores with branch prediction and TLB, L1/L2 caches, on-chip network, and main memory subsystem. Table 2 lists the architectural parameters.

**Table 2 Architectural parameters**

CPU and L1 cache	x86 in-order, 32KB I/D, 4-way, 1 cycle hit latency, 2GHz
L2 cache	1MB unified, 16-way, 6 pipe stages, 2GHz
DRAM	64MB~1GB (cache), 2GB (DRAM only), LPDDR2-800, 4x16b, $t_{CL}=t_{RP}=t_{RCD}=15$ ns
PCM	LPDDR2-800, 4x16b (details in Table 1)

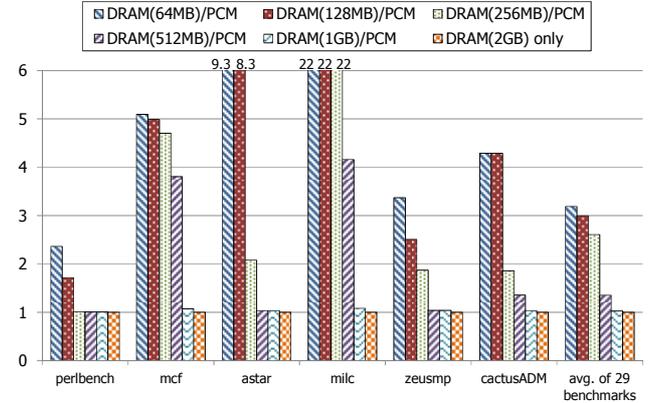
The main memory subsystem is configured in three cases: (1) DRAM only, (2) PCM only, and (3) PCM/DRAM hybrid memory. The PCM and DRAM models in the McSim are equipped with the performance and power consumption data from both the PCM datasheet and our chip-level characterization in Section 3.2 (for PCM write latency). We calculated the power consumption of each of PCM and DRAM utilizing the Micron power calculator [28] with the corresponding information in the datasheet.

In the PCM/DRAM hybrid memory, the DRAM plays the role of last level cache where the tags are managed at the granularity of DRAM row (4KB). Thus, in case of DRAM cache miss, a 4KB victim line is evicted (if dirty, after being written back to the PCM) and a new line is fetched from the PCM to the DRAM. In order to reduce the overhead of PCM writes, the differential writes and *flip-n-write* are applied in the PCM. Thus, only dirty bits in the 4KB cache line are written in the PCM and the size of bit updates is at most 2KB. The data-dependent long write latency of PCM first affects the latency of such write-back operations. We applied write preemption during the write-back operations in order to serve newly arrived read requests to the same bank where the write-back operation is being performed. We used SPEC2006 benchmarks.

### 4.2 Evaluation Results

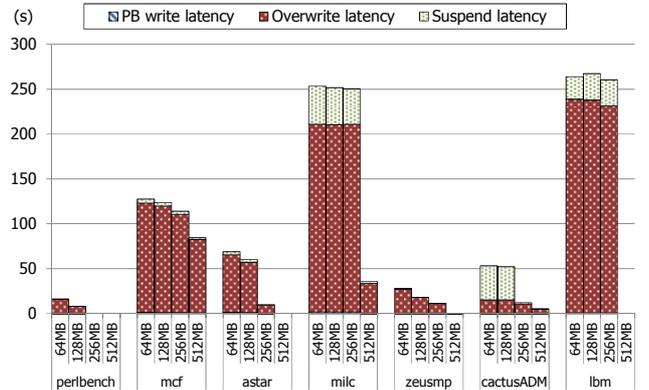
In our system-level evaluation, our goal is to evaluate (1) the performance and energy consumption of PCM-based main memory and (2) the impact of write (preemption) latency on system performance. Figure 2 shows the performance comparison between the DRAM only (2GB) and the PCM/DRAM hybrid main memories. In the figure, we do not show the results of the PCM only main memory (2GB) since its performance is significantly low (10~100 times worse than the DRAM only cases) due to the long write latency. The figure shows that the hybrid memory gives comparable performance to that of the DRAM only memory when the DRAM cache size is larger than 256MB. The figure also shows that too

small DRAM caches significantly degrade system performance mainly due to the long write latency in the PCM.



**Figure 2 System performance comparison (normalized CPI)**

Figure 3 shows the decomposition of total write latency. The figure shows that the accumulated write preemption penalty occupies 0.4~71.5% of total write latency. Thus, for further performance improvements, the write preemption penalty needs to be reduced.



**Figure 3 Total write latency**

Figure 4 shows the energy consumption. With small caches, the PCM program power is dominant because of significant evictions from the cache. As the DRAM cache size gets larger, the PCM portions become smaller because the larger DRAM cache reduces read/write accesses to the PCM. The DRAM only memory gives 14.1% (67.33%) less energy consumption than the PCM/DRAM hybrid with 1GB (512MB) DRAM cache. The program power and latency can be reduced by device scaling. For example, assuming that the program latency and power scale linearly with device scaling, the hybrid memory having a 28nm PCM and 1GB (512MB) DRAM consumes 9.3% (48.8%) less energy than the one with the current 58nm PCM.

## 5. Lessons Learned and Future Work

Our evaluation results show that the PCM/DRAM hybrid main memory is a promising candidate to complement the DRAM when the DRAM faces its scaling limit. We also found that the long write latency and write preemption penalty play a crucial role in the performance of PCM-based main memory. In order to apply the PCM to general cases, e.g., where small DRAM caches need to be adopted for cost reasons or applications with large footprint run as in the server applications, the PCM write latency needs to be reduced significantly. Compared with the cell write latency (100s of nanoseconds), the write latency in Table 1, e.g., about 25 $\mu$ s for 2B~32B data write, is significant and needs to be reduced.

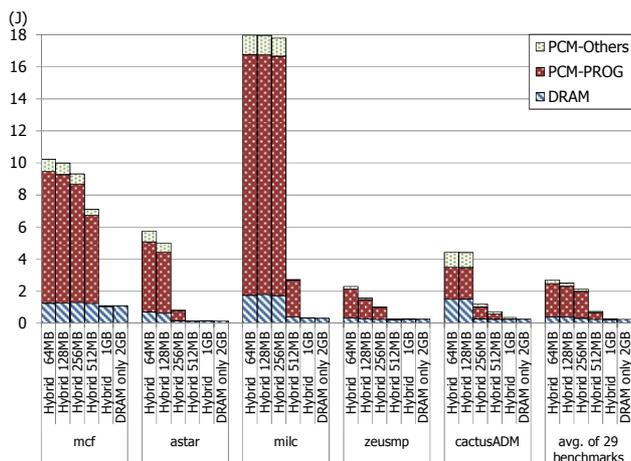


Figure 4 Memory energy consumption

We analyze that there are two factors which result in the gap between cell-level write latency and chip-level write performance. They will be peak write current limit and R drift latency. The constraint of peak write current is given as the chip specification considering the application areas, e.g., low peak current for mobile applications. Thus, in case that the peak current constraints become less stringent or more advanced technology, e.g., 28nm is adopted, the internal write bandwidth can be increased thereby reducing latency for large data. There will also be possibilities of further improvement regarding the factor of R drift latency. We analyze that the PCM chip imposes 14 $\mu$ s as the R drift latency.<sup>2</sup> We envision that such a long latency can be hidden by architectural improvements. For instance, the R drift latency of a preceding data can be hidden by the write operation of a subsequent data.

We analyze that write preemption penalty is also related with R drift latency since subsequent reads can be performed only after the R drift latency of previous internal write operation is completely finished. We expect that the write preemption penalty can also be hidden by overlapping a new write operation to a different bank during the period of R drift latency.

Note that we did not evaluate the cases where the PCM can give advantages over the conventional DRAM, e.g., reduction of page faults by large PCM capacity [7] and fast boot-up by fetching the OS image from the fast PCM instead of the Flash memory storage.

## 6. Conclusion

In this study, we evaluated the state-of-the-art PCM for main memory applications. We performed a chip-level characterization to obtain the relationship between PCM write latency and write data size. In our system-level simulations, we demonstrated that the PCM/DRAM hybrid main memory gives a comparable performance to the DRAM only main memory. We also reported that the long write latency and write preemption penalty can incur significant performance degradations in the hybrid main memory when a small DRAM is utilized and/or large footprint applications run on the hybrid memory. In summary, in order for the PCM to be widely utilized in the main memory, the write performance (including write preemption penalty) and power need to be improved via architectural innovations as well as device scaling.

<sup>2</sup> The write latency of 32B data is 25 $\mu$ s. The slope of write latency is 11 $\mu$ s/32B as shown in Figure 1. Thus, we analyze that the additional 14 $\mu$ s is due to R drift.

## 7. References

- [1] International Technology Roadmap for Semiconductors (ITRS), available at [www.itrs.net](http://www.itrs.net).
- [2] C. Villa, et al., "A 45nm 1Gb 1.8V Phase-Change Memory," Proc. ISSCC, 2010.
- [3] H. Chung, et al., "A 58nm 1.8V 1Gb PRAM with 6.4MB/s Program BW," Proc. ISSCC, 2011.
- [4] EE Times, Samsung to ship MCP with phase-change, <http://www.eetimes.com/electronics-news/4088727/Samsung-to-ship-MCP-with-phase-change>.
- [5] Numonyx, "Phase Change Memory (PCM): A new memory technology to enable new memory usage models," available at [www.numonyx.com/en-us/MemoryProducts/PCM/Pages/PCM.aspx](http://www.numonyx.com/en-us/MemoryProducts/PCM/Pages/PCM.aspx).
- [6] B. C. Lee, et al., "Architecting Phase Change Memory as a Scalable DRAM Alternative," Proc. ISCA, 2009.
- [7] M. K. Qureshi, V. Srinivasan, and J. A. Rivers, "Scalable High Performance Main Memory System Using Phase-Change Memory Technology," Proc. ISCA, 2009.
- [8] M. K. Qureshi, et al., "Improving Read Performance of Phase Change Memories via Write Cancellation and Write Pausing," Proc. HPCA, 2010.
- [9] P. Zhou, et al., "A Durable and Energy Efficient Main Memory Using Phase Change Memory Technology," Proc. ISCA, 2009.
- [10] G. Dhiman, R. Ayoub, and T. Rosing, "PDRAM: A Hybrid PRAM and DRAM Main Memory System," Proc. DAC, 2009.
- [11] W. Zhang and T. Li, "Characterizing and Mitigating the Impact of Process Variations on Phase Change based Memory Systems," Proc. MICRO, 2009.
- [12] H. Park, S. Yoo, and S. Lee, "Power Management of Hybrid DRAM/PRAM-based Main Memory," Proc. DAC, 2011.
- [13] D. Ielmini, A. L. Lacaita, and D. Mantegazza, "Recovery and Drift Dynamics of Resistance and Threshold Voltages in Phase-Change Memories," IEEE Trans. on Electron Devices, vol. 54, no. 2, Feb. 2007.
- [14] K. Lee, et al., "A 90nm 1.8V 512Mb Diode-Switch PRAM with 266MB/s Read Throughput," IEEE J. Solid-State Circuits, vol. 43, no. 1, pp. 150-162, Jan. 2008.
- [15] B. D. Yang, et al., "A Low Power Phase-Change Random Access Memory Using a Data-Comparison Write Scheme," Proc. ISCAS, 2007.
- [16] S. Cho and H. Lee, "Flit-N-Write: A Simple Deterministic Technique to Improve PRAM Write Performance, Energy and Endurance," Proc. MICRO, 2009.
- [17] L. A. Lastras-Montano, et al., "On the lifetime of multilevel memories," Proc. ISIT, 2009.
- [18] G. Sandre, et al., "A 90nm 4Mb Embedded Phase-Change Memory with 1.2V 12ns Read Access Time and 1MB/s Write Throughput," Proc. ISSCC, 2010.
- [19] M. K. Qureshi, et al., "Enhancing Lifetime and Security of PCM-Based Main Memory with Start-Gap Wear Leveling," Proc. MICRO, 2009.
- [20] N. H. Seong, et al., "Security Refresh: Prevent Malicious Wear-out and Increase Durability for Phase-Change Memory with Dynamically Randomized Address Mapping," Proc. ISCA, 2010.
- [21] E. Ipek, et al., "Dynamically Replicated Memory: Building Reliable Systems from Nanoscale Resistive Memories," Proc. ASPLOS, 2010.
- [22] S. Schechter, et al., "Use ECP, not ECC, for Hard Failures in Resistive Memories," Proc. ISCA, 2010.
- [23] N. Seong, et al., "SAFER: Stuck-At-Fault Error Recovery for Memories," Proc. MICRO, 2010.
- [24] A. Akeel, et al., "Onyx: A Prototype Phase Change Memory Storage Array," Proc. HotStorage, 2011.
- [25] JEDEC Standard, Low Power Double Data Rate 2 (LPDDR2), JESD209-2E, April 2011.
- [26] S. Li, et al., "McPAT: An Integrated Power, Area, and Timing Modeling Framework for Multicore and Manycore Architectures," Proc. MICRO, 2009.
- [27] C. Luk, et al., "Pin: Building Customized Program Analysis Tools with Dynamic Instrumentation," Proc. PLDI, June 2005.
- [28] Micron Technology Inc., "Calculating Memory System Power for DDR2," TN-47-04, [www.micron.com/support/dram/power\\_calc.html](http://www.micron.com/support/dram/power_calc.html).