

## Non-Volatile Computer Memory: Other Dimensions, Other Domains

By Tiffany Fox

*This article was provided by the Jacob's School of Engineering.*

San Diego, Calif., March 31, 2016 — Some sources estimate that by 2025, the total amount of information generated and stored by the world's computing devices will be upwards of 3.7 zettabytes.

To conceptualize zettabytes, think of them this way: If the 11-ounce cup of coffee on your desk is the equivalent of one gigabyte, one zettabyte would have the same volume as the Great Wall of China. And in this case, we're talking 3.7 Great Walls of China.

Another way to look at it: The amount of data we're about to dump into cyberspace is the equivalent of Netflix streaming its entire catalog of movies simultaneously to almost 10 million users. That's a lot of bits. A global tsunami of bits. And the "Internet of Things" has barely gotten started.

Against this backdrop of intense digital desire and demand, computer engineers in both industry and academia find themselves on a seemingly never-ending quest to develop novel ways to store and retrieve all that data quickly, reliably, and on the cheap.

It's no surprise, then, that the 7th annual Non-Volatile Memories workshop at the University of California, San Diego elicited the interest of more than 185 researchers from around the world, representing both academia and industry. They were there to hear where things might be headed for non-volatile computer memory, or NVM, in a world where bigger, better, faster and stronger are not just the goal, but the expectation.

NVMs are crucial components of modern computing systems — components that make it possible to store increasingly large amounts of information in smaller spaces, at faster data transfer speeds and at lower cost to the consumer. NVM, in its most basic sense, is what makes it possible for you to turn off your computer (or unplug your flash drive, a form of NVM) and have it still 'remember' the last draft of the novel you've been working on. It's also relatively low-power, boasts fast random access speed and is known for being rugged since it doesn't require a spinning disk or other moving mechanical parts.

But even if there were no disks spinning at NVMW this year, there were certainly heads spinning. The workshop featured a full three days of talks on advances that have not only made it possible to store data in three dimensions, but also inside an entirely new material: strands of DNA.



More than 185 researchers from around the world, representing both academia and industry, attended the 7th annual Non-Volatile Memories Workshop to hear where things might be headed for non-volatile computer memory. Photos by Alex Matthews and Joel Polizzi for the Qualcomm Institute. For more photos, [click here](#).

## Another day, another dimension



Frank Hady, an Intel Fellow and Chief Architect of 3D XPoint Storage.

Although NVMs have been a field of active research for decades, engineers always make room for improvement, which explains the conference-goers' excitement about rumors that a representative from Intel would provide an inside look at a ground-breaking method for storing data.

The rumor, it turns out, was true: In a talk called "Wicked Fast Storage and Beyond," Frank Hady, an Intel Fellow in Intel Corporation's Non-Volatile Memory Solutions Group, made a case for why Intel's 3D XPoint Technology -- which stores data in three dimensions in an entirely new way -- could transform the industry.

3D XPoint (pronounced "3D crosspoint") derives its name from its structure, which incorporates layers of memory stacked in a crosshatch pattern in three dimensions and makes it possible to pack an array of memory cells at a density 10 times that of conventional NAND memory, an industry standard in solid-state memory.

Notably, the memory element in 3D XPoint doesn't include a transistor for switching electrical signals: instead, the memory cells are comprised of a material that changes the cell's physical properties, giving it either a high or low electrical resistance, representing the 1s and 0s of binary code. No transistor means the memory can be made very dense -- specifically, 8 to 10 times greater than DRAM, a type of random-access memory that stores each bit of data in a separate cell.

Contrast 3D XPoint with NAND, which works by shifting electrons to either side of a "floating gate" to change from 1 to 0 and back. The problem with NAND memory is that to rewrite one cell, a computer has to rewrite an entire block of cells, which slows down the process. 3D Xpoint makes it possible to both read and write, on each individual cell, at the byte level.

This capability greatly impacts processing speed: 3D Xpoint has a data rate of 77K Input/Output Operations Per Second (IOPS) compared to NAND's data rate of 11K IOPS. NAND is also a 2D technology, and one less dimension means that much less density, which can adversely affect overall device size.

Paul Siegel is a professor of Electrical Computing and Engineering at UC San Diego and past director of the Center for Memory and Recording Research (CMRR, formerly the Center for Magnetic Recording Research), a co-sponsor of the workshop. "3D storage eases the pressure to make the memory component smaller," said Siegel. "You can pack the bits more efficiently, because now you have a cube to work with instead of just a flat plane."



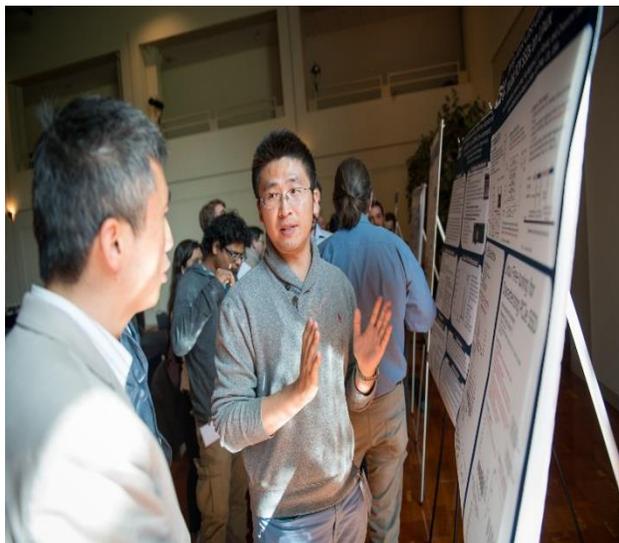
Prof. Paul Siegel, professor of Electrical Computing and Engineering at UC San Diego.

Hady, the chief architect of the technology, noted that the unique design of 3D XPoint also makes it tough to beat in terms of latency, or the time between IOPS. Acknowledging that latency has long been a thorn in the side of systems engineers, Hady cited an anonymous saying from the world of networking: "Bandwidth problems can be cured with money. Latency problems are harder because the speed of light is fixed. You can't bribe God."

But that doesn't mean there aren't ways to work around the limitations of the universe.

"3D XPoint has a 12x latency advantage and that's why I'm calling it wicked fast," he added. "Because of this low latency, performance shoots up very quickly, and you will see that kind of performance increase as you run your apps."

"Every three decades or so, we see a leap in latency that makes a huge impact on the system," he continued. "The really neat thing is we're about to make that three orders of magnitude jump again. The memory guys are now kind of where the networking guys have been -- for the first time in decades, there's a reason to go tune the system."



NVMW included a poster session and tutorial, as well as keytone talks and panel discussions with research in both industry and academia.

Referring to the technology as a "fertile systems research area," Hady -- who was visibly excited throughout his presentation - - claimed it will pave the way for "rethinking applications for big improvements"

"If you are a systems grad student, you picked a good time," he said. "You don't always have these technological shifts to base your research on and you have one now."

As for how "high-performance" 3D XPoint is, only time will tell. No 3D XPoint devices have been independently tested for bit-error rates, but a paper presented at NVMW by Jian Xu and Steven Swanson of the UC San Diego Department of Computer Science and Engineering concludes that there are a number of integration issues with "managing, accessing, and maintaining consistency for data stored in NVM."

For one, existing file systems won't be able to take advantage of 3D NVM technology until new software can be written and the system re-engineered. Xu and Swanson, who is co-director of the Non-Volatile Systems Lab (NVSL), a co-sponsor of the workshop, propose in their paper that 3D NVMs like 3D XPoint be used in conjunction with DRAM solid-state drives rather than as replacements for flash SSDs.

Hady seemed to agree that DRAM isn't going anywhere. When an audience member asked if it would replace NAND flash, he replied "I think there's a long life for both." And as for cost, because 3D XPoint is proprietary, Hady couldn't reveal much other than to say it will be "somewhere between DRAM and NAND."

## Taking Data's Temperature

Intel isn't the only company working at the forefront of 3D memory -- Toshiba is another major player in the global effort to cope with the world's data tsunami.

In his NVMW keynote address entitled "Advances in 3D Memory: High Performance, High Density Capability for Hyperscale, Cloud Storage Applications, and Beyond," Jeff Ohshima, a member of the Semiconductor and Storage Products executive team at Toshiba Corporation, discussed three new technologies in the works at Toshiba (which, along with Western Digital, Huawei, Samsung, Intel and several other corporations, is a corporate sponsor of the workshop).

The trio of technologies described in his talk are designed to "fit into a hierarchy of bits," Ohshima explained -- a hierarchy made ever more complex by the sheer volume of data being generated by server farms, datacenters and smartphones.

Fortunately, not all of the data is being used all the time. Engineers refer to bits as being either "hot," "warm" or "cold," meaning hot bits (data in active use) are typically stored in a computer's DRAM or MRAM (types of random-access memory) to allow for fast retrieval, while warm bits (less frequently accessed bits) are stored in a high-performance solid-state drive (SSD) and cool bits are stored in an archival SSD.



Jeff Ohshima, member of the Semiconductor and Storage Products executive team at Toshiba Corporation.

For so-called "warm data" Toshiba last year developed a form of NAND flash memory using a technology called Through Silicon Via (TSV), which achieves an input-output data rate of over 1 Gbps. That's higher than any other NAND flash memory and also reduces power consumption by approximately 50 percent with low voltage supply.

Toshiba has also developed its own version of 3D NVM called BiCS Flash, so-called for the technology's "bit column stackable" design. BiCS Flash was first introduced in 2007, but the newest generation device features the triple-level cell, or TLC, technology, which makes it possible to store three bits per cell.

As if solid-state drive development wasn't fast and furious enough, Ohshima also pointed out Toshiba's solution for storing cool data: Quadruple Level Cell or QLC, which is capable of storing four bits per cell. QLC and other technologies are expected to drastically increase SSD drive size in the near future, resulting in capacities of up to 128 terabytes by 2018.

But even with bigger, denser drives, Ohshima notes that all of the existing storage media represent only a third of what's needed in the coming decade to store the aforementioned 3.7 zettabytes of data expected to saturate cyberspace. Imagine a warehouse stacked to the rafters with 250 billion DVDs (that's billion with a 'b'), add 2.7 identical warehouses and maybe you get some idea for why engineers are so excited about storing more bits per cell, and in an extra dimension.

"The emergence of these 3D technologies across the storage industry represents a significant inflection point in the trajectory of computer memory capabilities," said Siegel, who works on modulation and error-correction coding techniques, or algorithms that modify the information before storing it in order to help prevent, detect and fix errors that might occur when the information is retrieved from the memory. "They are driving new paradigms in computing system architecture and sparking the invention of entirely new data encoding and management techniques."

But what if bits and cells and disks and drives weren't the only way to go about storing data? What if data could be stored in an entirely new way, a way that might change the world of computing as we know it?

## The DNA Domain



Ryan Gabrys, electrical engineer at SpaWar Systems Center in San Diego.

When it comes down to it, DNA -- the same stuff that serves as a blueprint for all organic life -- is really not much different from binary code. Instead of using 1s and 0s to encode information as magnetic regions on a hard drive, DNA uses four nucleobases (known by the abbreviations T, G, A and C) to encode information onto strands comprising a double helix. To "sequence" DNA essentially means to decode it from those TGAC bases into binary code.

But that also means, of course, that binary code can be converted into nucleobases -- a new area of research that has major implications for data storage, and for human society.

DNA, first of all, is extremely dense: you can store one bit per base, and a base is only a few atoms large. It's also extremely durable. While most archival computing systems need to be kept in subzero temperatures, DNA (as we know from paleontological digs) can still be read hundreds of thousands of years later, even after being dug up out of the Gobi Desert.

Imagine once again that coffee cup on your desk. DNA is so dense and so robust that singer Justin Bieber's catalog could be stored more than 66 million times in about .03 ounces of that coffee, or the equivalent of a single gram of DNA. What's more, Bieber's music, when stored this way, wouldn't be destroyed for hundreds of thousands of years (for better or worse).

"This is what makes DNA a form of super-archival data storage," explained Ryan Gabrys, an electrical engineer at Spawar Systems Center in San Diego who gave a talk at NVMW on "Coding for DNA-Based Storage (the Bieber analogy came from his talk)." "Using (artificial) DNA, you could store 6.4 gigabytes of information within a human cell," or the equivalent of about 100 hours of music. In one human cell.

"All the world's data," added Gabrys, "could be stored in an area the size of a car."

Eitan Yaakobi, an assistant professor of computer science at the Technion – Israel Institute of Technology and an affiliate of CMRR, acknowledged that “DNA storage may sound like science fiction at this point.” But, he added, “I believe that in the near future we will see tremendous progress in this direction and literally the sky will be the limit. It will be a great opportunity to combine efforts from researchers from all different fields starting from biology to systems and coding.”

And yet, as Gabrys pointed out, “everything’s complicated with DNA.” The technologies available to sequence DNA are prone to substitution- and deletion-based errors (hence Gabrys’ work to develop error-correction codes) and the process to convert binary code into nucleobases requires proprietary and expensive technology.

“For this to even be a consideration we’re talking six to seven years down the road,” he said.

Which means just about the only thing that’s guaranteed in the wild world of data storage is that there will be a lot to talk about at the Non-Volatile Memories Workshop come six or seven years from now.

#### **Media Contacts**

**Tiffany Fox**  
(858) 246-0353  
[tfox@ucsd.edu](mailto:tfox@ucsd.edu)

#### **Related Links**

[Non-Volatile  
Memories Workshop  
2016](#)