

Hauptseminar Hochleistungsrechner: Aktuelle Trends und Entwicklungen Wintersemester 2017/2018 **Speichertechnologien und ihre Nutzung**

Maurice Rang
Ludwig-Maximilians-Universität München

30.01.2018

Abstract

Die Bedeutung von nicht-flüchtigen Speichertechnologien sowie deren Verwendung steigt nach wie vor an. Die vorliegende Seminararbeit liefert einen kurzen Überblick über die in nicht-flüchtigen Speichertechnologien verwendete Hardware und kann als Einstieg in das Thema genutzt werden. Dazu wird zunächst deren grundlegender Aufbau vorgestellt. Darüberhinaus werden Unterschiede zwischen den einzelnen Technologien erläutert und analysiert. Abschließend werden in diesem Kapitel verschiedene Programmiermodelle und Einsatzmöglichkeiten für nicht-flüchtige Speichertechnologien aufgezeigt. Zum Ende dieser Arbeit gibt es eine kurze Darstellung der Speichertechnologie 3D Dram und dazu zwei passende Beispiele.

1 Einleitung

In den letzten Jahren ist die Datenmenge, die durch die Menschheit produziert wurde, auf ein sehr hohes Niveau angewachsen. Schaut man sich das Jahr 2007 an stellt man fest, dass sich hier die Kapazität des weltweiten Datenbestandes noch auf etwa 295 Exabyte (EB) beziffern ließ [14]. Circa 10 Jahre später im Jahre 2016 ist dieser Bestand schon auf 16 Zettabyte (ZB) angewachsen und damit um über 5000% angestiegen. Schenkt man Prognosen

einiger Experten Glauben soll sich der weltweite Datenbestand im Jahre 2025 auf über 163 Zettabyte (ZB) belaufen und damit nochmals um mehr als 1000% im Vergleich zum Jahr 2016 anwachsen [13]. Eine Begleiterscheinung, die durch diese rasante Entwicklung entsteht, ist die immer größer werdende Nachfrage nach neuen und verbesserten Speichertechnologien, um diese Menge an Daten effizient und schnell verarbeiten zu können. Parallel dazu ist die Weiterentwicklung und konstante Verbesserung der benötigten Mikroprozessoren genauso unabdingbar, da die Verarbeitung dieser Größenordnung an Daten ein hohes Maß an Rechenleistung benötigt. In Bezug auf die Entwicklung von Mikroprozessoren stellte Gordon Moore einer der Mitgründer von Intel im Jahr 1965 eine Theorie auf, die von da an als Mooresches Gesetz bekannt wurde. Diese Theorie besagt, dass "sich die Anzahl der Komponenten, die sich auf einem Chip befindet, jedes Jahr ungefähr um das Doppelte erhöht" [10].

Auf den nachfolgenden Seiten werden zunächst vier Speichertechnologien vorgestellt, die zur Familie der nicht flüchtigen Speicher gehören. Dabei werden von jeder Technologie die technischen Grundlagen und ihre Funktion dargelegt. Im Anschluss daran wird auf die potenziellen Einsatzgebiete und Programmiermodelle von nicht-flüchtigen Speichern eingegangen. Den Schlussteil der vorliegenden

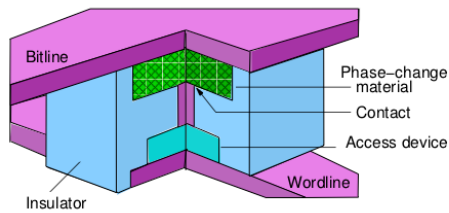


Abb. 1: Aufbau eines Phase-Change Memory Device [12]

Arbeit bildet das Kapitel 3D Stacked DRAM. Hier wird wiederum zunächst die technische Funktionalität dieser Technologie und dann deren Funktionalität beschrieben.

2 NVRAM

Im folgenden Kapitel werden vier unterschiedliche Speichertechnologien mit ihrem technischen Aufbau und ihrer Funktion vorgestellt, die zur Familie der nicht-flüchtigen Speicher gehören.

2.1 Definition

NVRAM ist eine Form des neuen und revolutionären Speichers, der auf der einen Seite Zugriffscharakteristiken von herkömmlichen RAM Speichern wie DRAM oder SRAM bereitstellt und zum anderen die nicht-flüchtigen Speichereigenschaften von Sekundärspeichern wie Flash miteinander vereint. Da die Halbleitertechnologie sehr große Fortschritte macht ist davon auszugehen, dass NVRAM zukünftig ein fester Bestandteil von eingebetteten Systemen und Computern wird [6].

NVRAM weist im Vergleich zu anderen Speichertechnologien zahlreiche Vorteile auf, vor allem gegenüber Flash-Speichern. Ein großer Vorteil, den nicht-flüchtige Speicher im Gegensatz zu Flash-Speichern haben ist ihre schnelle Zugriffszeit, die hauptsächlich auf Schreibzugriffe bezogen ist. Während die Zugriffszeit bei Flash-Speichern in

Parameter	DRAM	NAND Flash	NOR Flash	PCM
Density	1X	4X	0.25X	2X-4X
Read Latency	60ns	25 us	300 ns	200-300 ns
Write Speed	≈1 Gbps	2.4 MB/s	0.5 MB/s	≈100 MB/s
Endurance	N/A	10 ⁴	10 ⁴	10 ⁶ to 10 ⁸
Retention	Refresh	10yrs	10yrs	10 yrs

Abb. 2: PCM im Vergleich [12]

einem Bereich von mehreren hundert Millisekunden liegt, beträgt die Schreibzeit bei nicht-flüchtigen Speichern wenige Nanosekunden. Dadurch eignet sich nicht-flüchtiger Speicher besonders für Situationen, in denen häufige Schreibvorgänge auftreten [6].

2.2 Phase-Change Memory (PCM)

Phase-Change Memory (PCM) ist eine Speichertechnologie, die zur Familie der nicht-flüchtigen Speicher gehört und ein sogenanntes "phase-change material" zur Sicherung von Daten verwendet. Dieses Material kann zwei physische Zustände annehmen: kristallin oder amorph. Der Vorteil dieses Materials liegt darin, dass es seinen Zustand sehr zuverlässig, schnell und oft wechseln kann. Der amorphe Zustand hat eine geringe optische Reflexivität und einen hohen elektrischen Widerstand. Der kristalline Zustand dagegen hat eine hohe Reflexivität und einen geringen Widerstand. Der Aufbau eines Phase-Change Memory Device ist in Abbildung 1 dargestellt. Das Material befindet sich zwischen einer oberen und unteren Elektrode mit einem Heizelement, das die Erweiterung zur unteren Elektrode darstellt. Das Kristallisieren des Materials wird als sogenannte SET-Methode bezeichnet. Diese Operation wird durch moderate Leistung sowie eine längere Dauer von Elektroimpulsen kontrolliert. Dabei befindet sich die Speicherzelle in einem Zustand mit geringem Widerstand. Die RESET-Methode dagegen zeichnet sich durch hohe Stromimpulse aus, die die Speicherzelle in einen Zustand mit hohem Widerstand versetzen. Die gespeicherten Daten der Zelle können mit

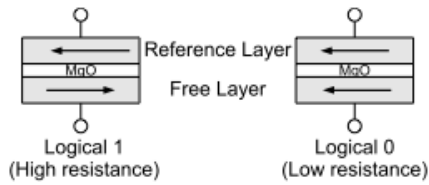


Abb. 3: Magnetic Tunnel Junction (MTJ) [16]

Hilfe von einer sehr geringen Stromzuführung aus-
gelesen werden. In der oberen Tabelle wird die
Speichertechnologie PCM mit anderen Technologien
in Bezug auf die Parameter Dichte, Lese-Latenz,
Schreibgeschwindigkeit, Standhaftigkeit und Spe-
icherung verglichen. Schaut man sich nun den Ver-
gleich an sieht man, dass PCM die vielversprechen-
sten Charakteristiken besitzt. In Bezug auf die
Dichte weist PCM (2X - 4X) ähnliche Werte auf
wie sie bei NAND Flash (4X) zu finden sind. Sieht
man sich nun den Parameter Lese-Latenz an, stellt
man fest, dass es hier Parallelen zu NOR Flash (300
ns) gibt, welche circa vier mal langsamer ist als die
Lese-Latenz bei DRAM (60 ns). Vergleicht man
abschließend den Parameter Schreibgeschwindigkeit
ist zu erkennen, dass diese bei PCM mit circa 100
MB/s höher ist als die bei den zwei Flash-Speichern
[12].

2.3 Spin-Transfer Torque RAM (STT-RAM)

Spin-Transfer Torque RAM (STT-RAM) ist eine
entstehende nicht-flüchtige Speichertechnologie, die
als Universalspeicher eingesetzt werden kann [15].
Sie gehört zu den sogenannten magnetoresistiven
Speicherarten (MRAM). Wie bei herkömmlichen
magnetoresistiven Speicherarten nutzt eine STT-
RAM Zelle zum Speichern von binären Daten
eine Magnetic Tunnel Junction (MTJ). Eine MTJ
besteht aus zwei ferromagnetischen Schichten, dem
Reference Layer und Free Layer und zusätzlich aus
einem Tunnel Barrier Layer (MgO). Die magnetische
Ausrichtung des Reference Layer ist beständig
und ändert sich nicht, wohingegen die magnetische
Ausrichtung des Free Layers geändert werden

	(A) SRAM	(B) STT-RAM	eDRAM	
			(C) 1T1C	(D) Gain cell
Cell schematic				
Process	CMOS	CMOS + MTJ	CMOS + Cap	CMOS
Cell size (F^2)	120 - 200	6 - 50	20 - 50	60 - 100
Data storage	Latch	Magnetization	Capacitor	MOS gate
Read time	Short	Short	Short	Short
Write time	Short	Long	Short	Short
Read energy	Low	Low	Low	Low
Write energy	Low	High	Low	Low
Leakage	High	Low	Low	Low
Endurance	10^{16}	$> 10^{15}$	10^{16}	10^{16}
Retention time	-	-	< 100 ns *	< 100 ns *
Features	(+) Fast (-) Large area (-) Leakage	(+) Non-volatile (+) Potential to scale (-) Extra process (-) Long write time (-) High write energy (-) Poor stability	(+) Low leakage (+) Small area (-) Extra process (-) Destructive read (-) Refresh	(+) Low leakage (+) Decoupled read/write (-) Refresh

* 32 nm technology node

Abb. 4: STT-RAM im Vergleich [16]

kann. Die magnetische Ausrichtung zwischen den
beiden Schichten erzeugt unterschiedliche Wider-
stände beim MTJ, die zum repräsentieren der in
den Zellen gespeicherten Daten genutzt werden.
Wenn das magnetische Feld zwischen den beiden
Schichten parallel zueinander ist und der Wider-
stand des MTJ niedrig ist, repräsentiert dieser Zu-
stand eine logische 0. Sobald das Feld antiparallel
ist und die MTJ einen hohen Widerstand hat, wird
eine logische 1 repräsentiert (siehe Abbildung 3).
Während des Lesevorgangs liegt eine kleine nega-
tive Spannung zwischen der Bit und Word line
an. Um eine logische 0 schreiben zu können, wird
eine positive Spannung zwischen der Bit und Word
line angelegt. Eine logische 1 wird durch eine nega-
tive Spannung geschrieben, die einen Strom in
die entgegengesetzte Richtung erzeugt [16]. Abbil-
dung 4 zeigt STT-RAM im Vergleich zu den Spe-
ichertechnologien SRAM und eDRAM hinsichtlich
verschiedener Kriterien. Sieht man sich nun die
einzelnen Punkte an stellt man fest, dass STT-
RAM nicht viel schlechter abschneidet als die an-
deren beiden Technologien. Ein wichtiger Vorteil,
den STT-RAM gegenüber den anderen Technolo-
gien besitzt ist, dass die Eigenschaft des nicht
flüchtigen Speichers. Nachteile, die diese Tech-
nologie im Vergleich zu SRAM und eDRAM den-
noch aufweist sind eine lange Schreibzeit, ein hoher
Energieverbrauch beim Schreiben und eine geringe
Beständigkeit.

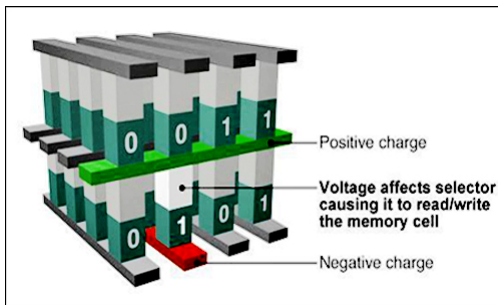


Abb. 5: 3D-Crosspoint-Architektur [1]

2.4 Resistive Random Access Memory (R-RAM)

Resistive Random Access Memory (R-RAM) ist ein weiteres Beispiel für eine nicht-flüchtige Speichertechnologie. Diese Technologie besteht im Grunde aus zwei Operationen: Unipolares Switching und bipolares Switching. Innerhalb dieses Kontextes ist das unipolare Switching für das Ausführen von Programmen und das Löschen zuständig. Dies geschieht mit Hilfe von kurzen und langen Impulsen oder mit Hilfe von hoher und niedriger Stromspannung mit der selben Spannungspolarität. Im Gegensatz dazu werden bei bipolarem Switching kurze Impulse mit einer umgekehrten Spannungspolarität verwendet. Ein Beispiel für unipolares Switching ist ein filament-basiertes R-RAM Gerät. Ein Filament wird in Form eines isolierten Dielektrikums erzeugt nachdem eine hohe Spannung angesetzt wurde. Nachdem das Filament geformt wurde, kann es durch angemessene Spannung gesetzt oder zurückgesetzt werden. Ein typisches Beispiel für bipolares Switching ist ein PCM Gerät, welches aus zwei metallischen Elektroden besteht. Ein dünner Film aus Elektrolyten befindet sich zwischen den Elektroden. Wenn eine negative Spannung an der inaktiven Elektrode angelegt wird fließen metallische Ionen in die Elektrolyte und werden durch die inaktive Elektrode vermindert. Abschließend formen die Ionen einen kleinen "Nanodraht" zwischen den Elektroden. Daraus resultiert, dass der Widerstand zwischen den Elektroden stark gesenkt wird. Soll die Zelle gelöscht werden

wird eine positive Spannung angelegt was zur Folge hat, dass die Ionen zurück in die Elektrolyte fließen. Der erzeugte "Nanodraht" wird zerstört und der Widerstand wieder erhöht [9].

2.5 3D XPoint

3D XPoint ist eine von Intel und Micron entwickelte nicht-flüchtige Speichertechnologie, die auf der sogenannten CrossPoint-Architektur basiert. In Abbildung 5 ist ein Querschnitt dieser Architektur zu sehen. Die Besonderheit hierbei liegt in der Adressierung und in der Kontroll-Architektur der Speicherzellen. Der Einsatz von Silizium für die Herstellung der elektronischen Komponenten ist nach wie vor gegeben. Der Unterschied zu den anderen Technologien liegt jedoch darin, dass nun keine Notwendigkeit mehr darin besteht für jeden Bereich einer einzelnen Zelle oder mehrerer Zellen einen Transistor einzusetzen. Dadurch wird es ermöglicht die Speicherdichte auf der einzelnen Fläche eines Chips um das Zehnfache zu erhöhen. Jede einzelne Speicherzelle besteht aus einem Speicherwert-Selektor (in Abbildung 5 hellgrau markiert) und einem Speicher-Bit (in Abbildung 5 grün markiert). Diese werden in mehreren Schichten gestapelt und von sogenannten Adressierungsleitungen miteinander verbunden (in Abbildung 5 als graue Balken dargestellt). Durch die Adressierungsleitungen liegt an jeder Speicherzelle eine Spannung an, die es ermöglicht zu bestimmen, ob es sich bei der auszuführenden Aktion um einen Lese- oder Schreibvorgang handelt. Das Weglassen der Transistoren für jede einzelne Zelle und das Zusammenschließen mehrerer Zellen hat zur Folge, dass die Architektur des Speichers stabiler und langlebiger wird [2]. Setzt man 3D XPoint nun in Vergleich mit NAND und DRAM Speichern ist festzustellen, dass 3D XPoint ähnliche Werte in Sachen Lese- und Schreibgeschwindigkeit aufweist wie sie bei DRAM vorkommen. NAND Speicher sind bezüglich dieses Parameters langsamer als die beiden anderen Technologien. Einen Vorteil, den NAND sowie 3D XPoint Speicher aufweisen sind die niedrigen Kosten. Diese sind viel geringer als die bei DRAM

Speichern der Fall sind. Zu erwähnen ist jedoch, dass 3D XPoint Speicher immer noch ungefähr zehn mal teurer sind als NAND Speicher. Einen weiteren Vorteil, den 3D XPoint und NAND gegenüber DRAM haben ist, dass sie nicht-flüchtig sind [3].

3 Programmiermodelle für NVRAM

Die nicht-flüchtige Eigenschaft von Speichern erlaubt eine Vielzahl an unterschiedlichen Programmiermodellen für nicht-flüchtige Speicher, die diese gegen Programm- und auch gegen Systemabstürze sicher machen. Einige dieser Programmiermodelle werden nun im Folgenden erläutert.

Ein erster Ansatz ist das sogenannte "Mnemosyne" ein Interface, das von Volos et al. zum Programmieren von SCM's entwickelt wurde. Dieses Interface erlaubt es Applikationen statische Variablen zu deklarieren deren Werte selbst über einen Systemneustart hinaus bestehen bleiben. Darüber hinaus lässt es zu, dass Applikationen Speicher reservieren, der durch persistenten Speicher unterstützt wird. Die Schreibreihenfolge für persistente Speicher wird hierbei durch eine Software geregelt, die non-cached Schreibmodi, cache-line flush Anweisungen und Speicherbarrieren verwendet.

Ein weiterer Ansatz, der nach Coburn et al. benannt wurde ist der NV-heap. Dies ist ein persistentes Objektsystem, das eine transaktionale Semantik bereitstellt. Um sicherstellen zu können, dass die Anwendungen konsistent ausgeführt werden, werden bei NV-heap die Daten in flüchtige und nicht-flüchtige Daten separiert.

Ein Ansatz, der sich von den beiden oben Dargestellten differenziert wurde von Narayanan et al. entwickelt. Dieser präsentierte eine Möglichkeit, bei der der gesamte Systemspeicher nicht-flüchtig ist. Dazu verwenden ein Prinzip, das als "flush-on-fail" bezeichnet wird. Bei diesem Prinzip wird der vorübergehende Status, der sich in den Registern der CPU und den Cache-lines befindet erst dann

gespeichert wird, wenn es zu einem Fehlerzustand kommt. Bei einem Neustart des Systems werden die aktuellen Applikations- und OS-Stati durch einen transparenten Checkpoint wieder hergestellt.

Zhao et al. haben eine Herangehensweise präsentiert, die einen nicht-flüchtigen last-level cache und nicht-flüchtigen Systemspeicher verwendet, um eine persistente Speicherhierarchie schaffen zu können. Hierbei ist es so, dass wenn eine Cache-line aktualisiert wurde beinhaltet sie die aktuelle Version während die Daten, die sich im nicht-flüchtigen Speicher befinden noch die alte Version beinhalten. Wenn die "dreckige" Cache-line bereinigt wurde, wird automatisch der alte Stand im nicht-flüchtigen Speicher durch den neuen Stand ersetzt.

Eine letzte Technik, die noch zu erwähnen bleibt ist die von Condit et al. Sie beschäftigten sich mit einer Methodik, bei dem ein transaktionales Filesystem zum Einsatz kommt, das die byte adressierbarkeit von Speichern so beeinflusst, sodass die Menge an Metadaten, die während einer Aktualisierung geschrieben wird, drastisch reduziert wird. Dieses Filesystem soll garantieren, dass das Schreiben im Filesystem dauerhaft wird und jede Operation des Filesystems automatisch in einer bestimmten Reihenfolge ausgeführt wird [11].

4 Einsatzmöglichkeiten von NVRAM

Da jede einzelne der vorgestellten Speichertechnologien ihre Vor- und Nachteile besitzt haben sich einige Wissenschaftler gedacht, dass man die Vorteile besser nutzen könnte, wenn man die einzelnen Technologien miteinander kombiniert. Im folgenden Abschnitt werden nun drei dieser Kombinationen dargestellt [11].

4.1 Flash + PCM

Sun et al. hat die Kombination zwischen Flash-Speicher und PCM vorgeschlagen. Hierbei soll PCM als Protokollumgebung für den NAND Flash-Speicher genutzt werden. Dieses Verfahren reduziert die Lese- sowie Löschoptionen für Flash,

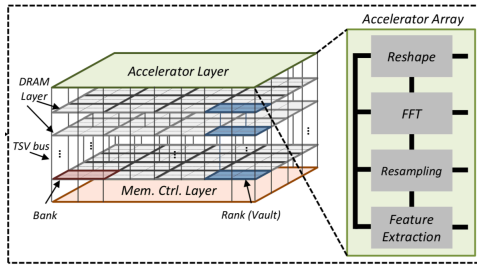


Abb. 6: Architektur eines 3D Stacked DRAM Systems [7]

was die Lebenszeit, die Performance und Energieeffizienz des Flash-Speichers verbessert. Durch die byte-adressierbarkeit von PCM wird zusätzlich die Performance für Leseoperationen verbessert.

4.2 DRAM + PCM

Qureshi et al. haben einen Ansatz hervorgebracht, der aus einem hybriden Hauptspeicher besteht, bei dem DRAM als "Page Cache" genutzt wird, um die Vorteile in Bezug auf Latenz des DRAMs mit den Vorteilen der Kapazität von PCM kombinieren zu können. Bei einem Seitenfehler wird die Seite, die von der Hard Disk geholt wurde, nur auf den DRAM Speicher geschrieben. Die Seite wird nur dann auf den PCM Speicher geschrieben, wenn sie aus dem DRAM Speicher entfernt wurde als "dreckig" gekennzeichnet wurde.

4.3 PCM + Flash + HDD

Kim et al. haben das Potential von PCM in Speicherhierarchien bezüglich Kosten und Performance bewertet. Sie beobachteten, dass das Schreiben auf Flash-Speicher basierend auf dem materiellen Level wesentlich langsamer ist als das Schreiben auf PCM. Basierend auf dem Systemlevel ist das Schreiben auf flash-basierte SSD's schneller als das Schreiben auf PCM-basierte SSD's was durch den Stromverbrauch bedingt ist. Weiterhin beobachteten sie, dass das Hinzunehmen von PCM die Performance in einem abgestuften Speichersystem verbessert und

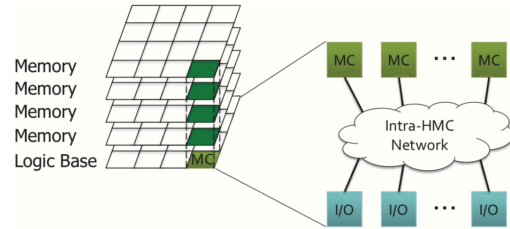


Abb. 7: Architektur eines HMC [8]

auch eine höhere Performance pro Dollar hervorgerufen kann.

5 3D Stacked DRAM

3D Stacked DRAM ist eine Speichertechnologie, bei der mehrere DRAM Matrizen und logische Schichten übereinander gestapelt werden und durch sogenannte TSV's (Through Silicon Vias) verbunden sind. Durch viele TSV Verbindungen ist eine Kommunikation mit hoher Bandbreite und geringer Latenz zwischen den einzelnen Schichten möglich. Abbildung 6 zeigt den Aufbau einer 3D DRAM Architektur. Hierbei ist zu erkennen, dass diese Struktur aus mehreren DRAM Schichten besteht, die an sich auch nochmal aus mehreren Teilstücken (in der Abbildung als *banks* gekennzeichnet) bestehen. Das vertikale Aufeinanderstapeln von banks wird als *vault* bezeichnet. Jedes vault hat seine eigene TSV-Verbindung und seinen eigenen Vault-Controller. Die logische Schicht (in Abbildung 6 als Mem. Ctrl. Layer bezeichnet) besteht aus dem Memory Controller, einem Crossbar Switch und aus Vault und Link Controllern. Diese Kontrolleinheiten belegen nicht den gesamten Platz der logischen Schicht, sondern lassen einen Raum für Erweiterungen [4].

5.1 Hybrid Memory Cube (HMC)

Als Beispiel für die oben erläuterte Speichertechnologie wird die Technologie Hybrid Memory Cube (HMC) gewählt. Die Architektur von HMC ist äh-

lich der in Abbildung 6 dargestellten Struktur. Der Cube besteht aus mehreren übereinander gestapelten DRAM Schichten und einer logischen Schicht am unteren Ende. Die einzelnen Schichten sind durch TSV's miteinander verbunden. Jede Speicherschicht des Cubes ist in einzelne Segmente aufgeteilt und das vertikale Zusammenschließen mehrerer Segmente wird als Vault bezeichnet. Der Prozessor kommuniziert mit dem HMC-seitigen Memory Controller indem er Nachrichten schickt, die Speicheraktionen, wie Schreiben oder Lesen, Speicheradressen und/oder Speicherdaten enthalten sendet. Innerhalb der logischen Schicht wird zusätzlich ein Switch benötigt, um die verschiedenen Vault Memory Controller und die I/O-Ports zu verbinden [8].

5.2 High Bandwidth Memory (HBM)

Ein weiteres Beispiel für 3D DRAM Speicher ist High Bandwidth Memory (HBM). Diese Technologie wird vorwiegend von Graphikprozessoren und als Nachfolger für GDDR5 verwendet. HBM wird von AMD, Nvidia und Hynix entwickelt. Dieser Standard zielt auf leistungsstarke Grafikanwendungen und Netzwerkanwendungen ab, die eine hohe Performance benötigen. HBM hat einen niedrigen Stromverbrauch, ultra-weite Kommunikationssleitungen und eine neuartige Stapelstruktur. Es ist vorgesehen, dass bis zu 8 Speicherchips übereinander gestapelt werden und mit Hilfe von TSV's verbunden werden. Die Prozessoranordnung bei HBM besteht aus einem Kernprozessor und vier Speicherstapeln. Die Speicherstapel an sich befinden sich auf einer logischen Schicht. Der Silizium Interposer ist ein Chip mit metallischen Schichten, die keine Logik besitzen. Der Aufbau von HBM ist nochmals detailliert in Abbildung 8 dargestellt [1].

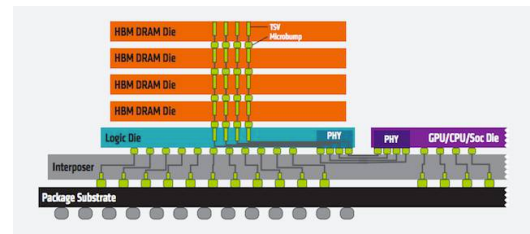


Abb. 8: Architektur eines HBM [1]

der Fokus auf den technischen Aufbau dieser Technologien und deren Verwendung gelegt. Jede der dargelegten Technologien hat eine andere Funktionsweise und ist in sich anders aufgebaut. Durch diese Unterschiede weisen sie bezüglich der wichtigsten Parameter für Speichertechnologien Dichte, Schreib- und Lesegeschwindigkeit sowie Dauerhaftigkeit unterschiedliche Werte auf. Vergleicht man nun abschließend die vier Speichertechnologien untereinander in Bezug auf die zuvor genannten Parameter stellt man fest, dass jede einzelne Speichertechnologie ihre Vor- und auch Nachteile besitzt. Beispielsweise weist STT-RAM in Bezug auf die Lesegeschwindigkeit exzellente Werte auf wohingegen die Lesegeschwindigkeit bei PCM eine Dauer von 200 bis 300 ns hat. Abschließend werden in diesem Kapitel unterschiedliche Programmiermodelle und Einsatzmöglichkeiten für NVRAM erläutert. In Abschnitt 5 wurde zunächst der grundlegende Aufbau von 3D DRAM Speichern dargelegt. Hierbei lag der Fokus wiederum auf dem hardwareseitigen Aufbau und dessen Funktionalität. Abschließend wurden zwei Beispiele aufgeführt, die die Verwendung von 3D DRAM Speichern verdeutlichen sollten. Die Bedeutung von nicht-flüchtigen Speichertechnologien wird in Zukunft noch mehr an Bedeutung gewinnen, da sie universell und vielseitig einsetzbar sind.

6 Zusammenfassung

Im vorliegenden Paper wurden zu Beginn vier Speichertechnologien vorgestellt, die zur Familie der nicht-flüchtigen Speicher gehören. Hierbei wurde

Literatur

- [1] Amd high bandwidth memory (hbm). http://www.semiconductorscentral.com/memory_page.html.
- [2] 3d xpoint technologie: 1000mal schneller als ssd. <https://www.krollontrack.de/blog/3d-xpoint-technologie-1000mal-schneller-als-ssd/> 4928/, December 2015.
- [3] Intel® optane™: Die revolution mit ultraschneller speichertechnik beginnt 2016. <https://www.intel.de/content/www/de/de/it-managers/non-volatile-memory-idf.html>, 2015.
- [4] B. Arkin, J. Hoe, and F. Franchetti. Hardware accelerated memory layout transform within 3d-stacked dram. Technical report, Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh.
- [5] M. Chang, P. Rosenfeld, S. Lu, and B. Jacob. Technology comparison for large last-level caches (l3cs): Low-leakage sram, low write-energy stt-ram, and refresh-optimized edram. Technical report, University of Maryland.
- [6] I. Doh, D. Lee, J. Choi, and S. Noh. Exploiting non-volatile ram to enhance flash file system performance. Technical report, Hongik University and Dankook University, 2007.
- [7] Q. Guo, N. Alachiotis, B. Akin, F. Sadi, G. Xu, T. Low, L. Pileggi, J. Hoe, and F. Franchetti. 3d-stacked memory-side acceleration: Accelerator and system design. Technical report, Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh.
- [8] G. Kim, J. Kim, J. Ahn, and J. Kim. Memory-centric system interconnect design with hybrid memory cubes. Technical report, Seoul National University, KAIST, 2013.
- [9] H. Li and Y. Chen. An overview of non-volatile memory technology and the implication for tools and architectures. Technical report, Alternative Technology Group Seagate Technology LLC, 2009.
- [10] C. Mack. Fifty years of moore’s law, 2011.
- [11] S. Mittal and J. Vetter. A survey of software techniques for using non-volatile memories for storage and main memory systems. Technical report, IEEE, 2015.
- [12] M. Qureshi, V. Srinivasan, and J. Rivers. Scalable high performance main memory system using phase-change memory technology. Technical report, IBM Research, 2009.
- [13] D. Reinsel, J. Gantz, and J. Rydning. Data age 2025: The evolution of data to life-critical. Technical report, IDC Headquarter, 2017.
- [14] Christoph Schrader. Explosion des cyberspace. *Süddeutsche Zeitung*, 2011.
- [15] C. Smullen, V. Mohan, A. Nigam, S. Gurumurthi, and M. Stand. Relaxing non-volatility for fast and energy-efficient stt-ram caches. Technical report, Department of Computer Science and Department of Electrical and Computer Engineering University of Virginia.
- [16] P. Zhou, B. Zhao, J. Yang, and Y. Zhang. Energy reduction for stt-ram using early write termination. Technical report, Department of Electrical and Computer Engineering, Department of Computer Science University of Pittsburgh.