Appears in the Proceedings of the 37th International Symposium on Computer Architecture (ISCA-37), June 2010

Re-Architecting DRAM Memory Systems with Monolithically Integrated Silicon Photonics

Scott Beamer[†] Chen Sun[‡] Yong-Jin Kwon[†]
Ajay Joshi[§] Christopher Batten[¶] Vladimir Stojanović[‡] Krste Asanović[†]

[†] Dept. of EECS
University of California
Berkeley, CA
{sbeamer,kwon,krste}@berkeley.edu

[‡] Dept. of EECS
Massachusetts Institute of Technology
Cambridge, MA
{sunchen,vlada}@mit.edu

[§] Dept. of ECE
Boston University
Boston, MA
joshi@bu.edu

[¶] School of ECE
Cornell University
Ithaca, NY
cbatten@cornell.edu

ABSTRACT

The performance of future manycore processors will only scale with the number of integrated cores if there is a corresponding increase in memory bandwidth. Projected scaling of electrical DRAM architectures appears unlikely to suffice, being constrained by processor and DRAM pin-bandwidth density and by total DRAM chip power, including off-chip signaling, cross-chip interconnect, and bank access energy. In this work, we redesign the DRAM main-memory system using a proposed monolithically integrated silicon-photonic technology and show that our photonically interconnected DRAM (PIDRAM) provides a promising solution to all of these issues. Photonics can provide high aggregate pin-bandwidth density through dense wavelength-division multiplexing. Photonic signaling provides energy-efficient communication, which we exploit to not only reduce chip-to-chip interconnect power but to also reduce cross-chip interconnect power by extending the photonic links deep into the actual PIDRAM chips. To complement these large improvements in interconnect bandwidth and power, we decrease the number of bits activated per bank to improve the energy efficiency of the PIDRAM banks themselves. Our most promising design point yields approximately a 10× power reduction for a single-chip PIDRAM channel with similar throughput and area as a projected future electrical-only DRAM. Finally, we propose optical power guiding as a new technique that allows a single PIDRAM chip design to be used efficiently in several multi-chip configurations that provide either increased aggregate capacity or bandwidth.

Categories and Subject Descriptors

B.3.1 [**Memory Structures**]: Semiconductor Memories—*DRAM*; B.4.3 [**Input/Output and Data Communications**]: Interconnections—*fiber optics, physical structures, topology*

General Terms

Design

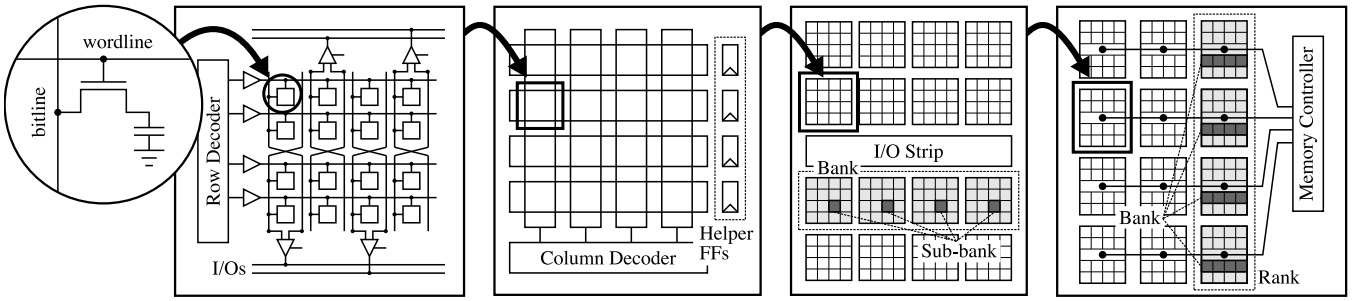
ISCA'10, June 19–23, 2010, Saint-Malo, France.

1. INTRODUCTION

The move to parallel microprocessors would appear to continue to allow Moore's Law gains in transistor density to be converted to gains in processing performance. Unfortunately, off-chip memory bandwidths are unlikely to scale in the same way and could ultimately constrain achievable system performance. Area and power overheads of high-speed transceivers and package interconnect limit the number of pins. Improved per-pin signaling rates are possible but only at a significant cost in energy efficiency and therefore will not necessarily improve aggregate off-chip bandwidth in a power-constrained system. It seems unlikely that pin bandwidth will increase dramatically without a disruptive technology. Even if we remove pin-bandwidth limitations, memory system performance could be constrained by the energy consumption of other components in current DRAM architectures. Apart from the I/O energy required to send a bit from the processor to the DRAM chip, considerable energy is expended traversing the intra-chip interconnect from the DRAM chip I/O to the desired bank and then accessing the actual storage cell.

In this paper, we propose photonically interconnected DRAM (PIDRAM) which uses a monolithically integrated silicon-photonic technology to tackle all of these challenges. Dense wavelength-division multiplexing (DWDM) allows multiple links (wavelengths) to share the same media (fiber or waveguide) yielding two orders of magnitude greater bandwidth density than electrical technology. Silicon photonics also demonstrates significantly better energy efficiency, supporting far larger off-chip bandwidths at a reasonable power budget. Monolithic integration allows energy-efficient photonic links to not only replace electrical I/O in DRAM chips, but to also extend across the DRAM chip to greatly reduce intra-chip interconnect energy. By redesigning DRAM banks to provide greater bandwidth from an individual array core, we can supply the bandwidth demands with much smaller pages thereby reducing bank activation energy. Our results show a promising design for a single-chip PIDRAM memory channel that provides a 10× improvement in throughput at similar power. Surprisingly, this does not incur an area penalty as higher bandwidth from each array core means fewer larger array blocks are required.

DRAMs are commodity parts, and ideally a single mass-produced part should be usable in a wide variety of system configurations. We propose optical power guiding as a technique to enable greater scalability of PIDRAM configurations, where we use guided photonic buses to direct optical power only to the PIDRAM chips that are actually accessed on a channel. Optical power guiding



(a) Array Core (b) Array Block (c) Bank and Chip (d) Channel

Figure 1: DRAM Memory System – Each inset shows detail for a different level of current electrical DRAM memory systems.

gracefully scales the requirements for optical loss and power delivery, and allows a flexible tradeoff between optical power, capacity, and bandwidth, while using the same PIDRAM part and only customizing the memory controller. Our results show optical power guiding scales significantly better than shared and split photonic bus implementations.

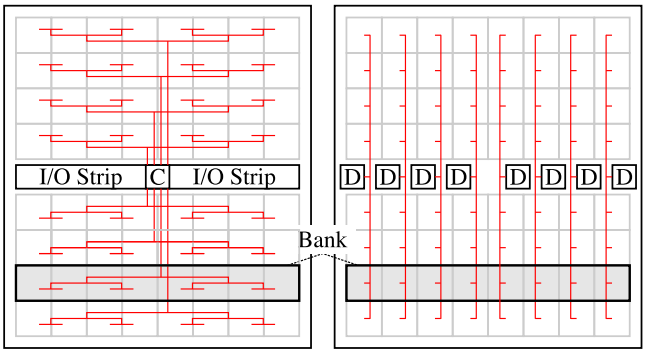
2. DRAM TECHNOLOGY

Figure 1 shows the structure of modern DRAMs, which employ multiple levels of hierarchy to provide fast, energy-efficient access to billions of storage cells. At the lowest level, each *cell* contains a transistor and a capacitor and holds one bit of storage.

Cells are packed into 2D arrays and combined with the periphery circuitry to form an *array core* (Figure 1(a)). Each row shares a wordline with peripheral wordline drivers, and each column shares a bitline with peripheral sense-amplifiers. Differential sense-amplifiers are used to amplify and latch low-swing signals when reading from the bitlines and to regenerate full-rail voltages to refresh the cell or write new values into the cell. The array core is sized for maximum cell density for a reasonable delay and energy per activation or refresh. In this paper, we model a folded bitline DRAM, which provides better common-mode noise rejection for the sense amp [12]. However, our general assumptions are also valid for the open bitline architecture which is making a comeback due to better scalability and area efficiency. Array cores are limited to a modest size that grows very slowly with respect to technology scaling due to intrinsic capacitances, so we assume a typical array core size of 512 wordlines by 1024 bitlines. Even though on any activation of an array core every cell in the activated row is read, only a few bits will be transferred over the array core I/O lines during a column access. A few row hits are possible for some workloads, but most of the other bits read from a row are never accessed before a different row is activated.

An *array block* is a group of array cores that share circuitry such that only one of the array cores is active at a time (Figure 1(b)). Each array core shares its sense-amplifiers and I/O lines with the array cores physically located above and below it, and the array block provides its cores with a global predecoder and shared helper flip-flops for latching data signals entering or leaving the array block. As a result, the access width of an array block is equivalent to the number of I/O lines from a single array core.

A *bank* is an independently controllable unit that is made up of several array blocks working together in lockstep (Figure 1(c)). The number of array blocks per bank sets the bank's access width. Array blocks from the same bank do not need to be placed near each other,



(a) Command Bus (b) Read- & Write-Data Buses

Figure 2: DRAM Chip Organization – Example DRAM chip with eight banks and eight array blocks per bank: (a) command bus is often implemented with an H-tree to broadcast control bits from the command I/O pins to all array blocks on the chip, (b) the read- and write-data buses and array blocks are bit-sliced across the chip to match the data I/O pins. (C = off-chip command I/O pins, D = off-chip data I/O pins, on-chip electrical buses shown in red)

and they are often striped across the chip to ease interfacing with the chip I/O pins. When a bank is accessed, all of its array blocks are activated, each of which activates one array core, each of which activates one row. The set of activated array cores within a bank is the *sub-bank* and the set of all activated rows is the *page*.

A *chip* includes multiple banks that share the chip's I/O pins to reduce overheads and help hide bank busy times (Figure 1(c)). Figure 2 shows how the I/O strip for the off-chip pads and drivers connects to the array blocks in each bank. The DRAM command bus must be available to every array block in the chip, so a gated hierarchical H-tree bus is used to distribute control and address information from the centralized command pins in the middle of the I/O strip (Figure 2(a)). The read- and write-data buses are striped across the chip such that all array blocks in a column are connected to the same data bus pin in the I/O strip (Figure 2(b)).

A *channel* uses a memory controller to manage a collection of banks distributed across one or more DRAM chips (Figure 1(d)). The channel includes three logical buses: the command bus, the read-data bus, and the write-data bus. To increase bandwidth, multiple DRAM chips are often ganged in parallel as a *rank*, with a slice of each bank present on each chip. To further scale bandwidth, the system can have multiple memory channels. To increase capacity, multiple ranks can be placed on the same channel, but with only one accessed at a time.

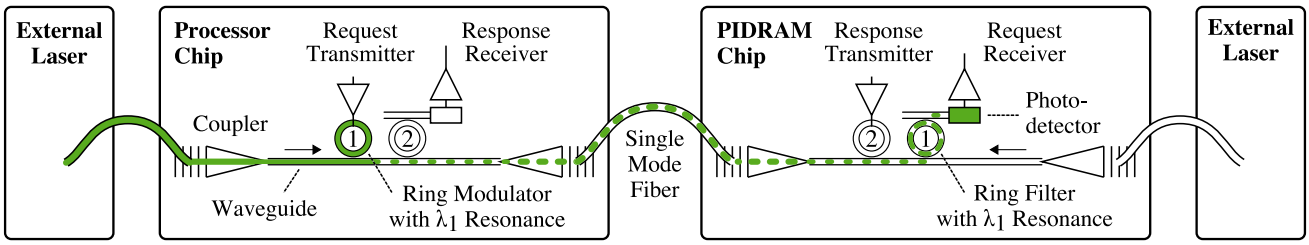


Figure 3: PIDRAM Link – Two DWDM links in opposite directions between a memory controller in a processor chip and a bank in a PIDRAM chip. λ_1 is used for the request and λ_2 is used for the response in the opposite direction on the same waveguides and fiber.

	Transmitter Energy (fJ/bt)			Receiver Energy (fJ/bt)		
I/O Technology	Data Dependent	Fixed	Thermal Tuning	Data Dependent	Fixed	Thermal Tuning
Electrical	1050	1450	n/a	1050	1450	n/a
Photonic (aggressive)	40	5	16	20	10	16
Photonic (conservative)	100	20	32	50	30	32

Table 1: Projected Electrical and Photonic I/O Energy – fJ/bt = average energy per bit-time assuming 50% bit transition probability, fixed energy includes clock and leakage, thermal tuning energy assumes 20 K temperature range. Electrical I/O projected from an 8 pJ/bt at 16 Gb/s design in a 40 nm DRAM process [14], to a 5 pJ/bt at 20 Gb/s design in a 32 nm DRAM process. Photonic I/O runs at 10 Gb/s/wavelength. Photonic projections based on our own preliminary test chips and ongoing circuit designs.

3. SILICON PHOTONIC TECHNOLOGY

Monolithically integrated silicon photonics is a promising new technology for chip-level interconnects. Photonics offers improved energy efficiency and bandwidth density compared to electrical interconnect for intra-chip and especially inter-chip links (e.g. memory channels). In this section, we first describe our assumed photonic technology and then discuss the various costs involved in implementing a unified on-chip/off-chip PIDRAM link.

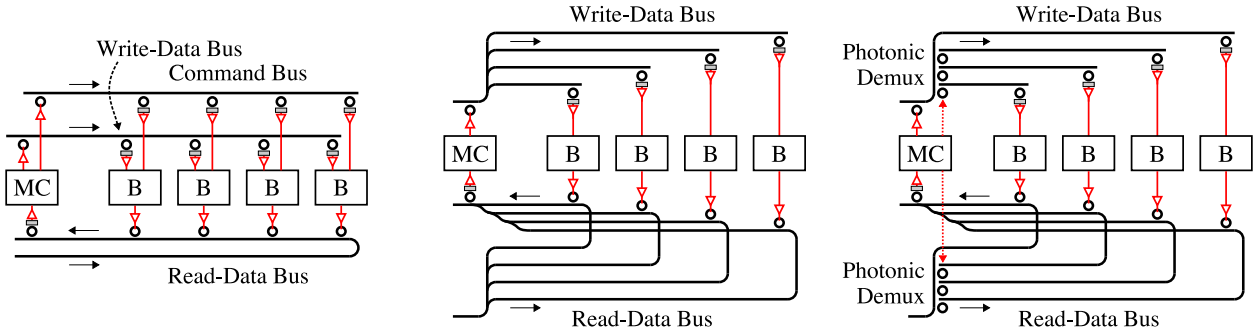
The various components in a PIDRAM link are shown in Figure 3. For a command or write data, light from an external broadband laser source is coupled into a photonic waveguide on the processor chip. The light passes along a series of ring resonators in the memory controller that each modulate a unique wavelength. The modulated light is then transmitted to the PIDRAM chip on a single-mode fiber. At the receiver side, the light is filtered by the tuned ring filters and dropped onto the photodetector. The photodetector converts light into electrical current which is sensed by the electrical receiver. For the read data, light is sent in the reverse direction on the same waveguides and fiber from the PIDRAM chip back to the processor chip. In this example, two wavelengths are multiplexed onto the same waveguide, but the real potential of silicon photonics lies in its ability to support dense wavelength-division multiplexing (DWDM) with dozens of wavelengths per waveguide. There are times when it is advantageous to filter a set of wavelengths at a time, and this can be accomplished using a bank of small single-wavelength rings or multi-wavelength comb filters. These ring filter banks can be either passively tuned to a fixed frequency or actively tuned to enable optical switching.

Both 3D and monolithic integration of photonic devices have been proposed in the past few years to implement processor-to-memory photonic networks. With 3D integration, the processor chips, memory chips, and a separate photonic chip are stacked in a variety of configurations. The photonic devices can be implemented in monocrystalline silicon-on-insulator (SoI) dies with thick layer of buried oxide (BOX) [6], or in a separate layer of silicon nitride (SiN) deposited on top of the metal stack [2]. Since the photonic devices are on a separate layer, engineers can employ customized processing steps to improve photonic device performance (e.g. like introducing ridge waveguides or epitaxial Ge for photodetectors).

Photonic Device Parameter	**Value**
Optical fiber loss	0.5e-5 dB/cm
Coupler loss	0.5–1 dB
Splitter loss	0.2 dB
Non-linearity loss at 30 mW	1 dB
Modulator insertion loss	1 dB
Waveguide loss	2–4 dB/cm
Waveguide crossing loss	0.05 dB
Filter through loss	1e-4–1e-3 dB
Filter drop loss	1 dB
Photodetector loss	1 dB
Laser efficiency	30–50%
Receiver sensitivity	-20 dBm

Table 2: Projected Photonic Device Parameters – Based on coupler designs in [21], waveguide losses from [6], filter designs in [19] as well as our preliminary photonic transceiver test chips and ongoing device work.

In this work, we assume monolithic integration, where photonic devices have to be designed using the existing process layers of a standard logic and DRAM process. The photonic devices can be implemented in polysilicon on top of the shallow-trench isolation (STI) in a standard bulk CMOS process [8,19] or in monocrystalline silicon with advanced thin BOX SoI. Photodetectors can be implemented using the silicon-germanium that is already present in the majority of sub-65 nm processes (and proposed for future DRAM processes [10]). Although monolithic integration may require some post-processing, its manufacturing cost should be much lower than 3D integration. Monolithic integration decreases the area and energy required to interface electrical and photonic devices, but it requires active area for waveguides and other photonic devices. It also requires an additional step in a DRAM process to deposit un-



(a) Shared Photonic Buses (b) Split Photonic Buses (c) Guided Photonic Buses

Figure 4: Photonic Implementations of Command, Write-Data, and Read-Data Buses – (a) *shared photonic buses* where optical power is broadcast to all banks along a shared physical medium, (b) *split photonic buses* where optical power is split between multiple direct connections to each bank, and (c) *guided photonic buses* where optical power is actively guided to a single bank. For clarity, command bus is not shown in (b,c), but it can be implemented in a similar fashion as the corresponding write-data bus. (MC = memory controller, B = bank)

doped polysilicon, since, unlike in a logic process, all polysilicon layers in a DRAM process are deposited heavily doped to minimize fabrication cost and resistivity of polysilicon interconnect.

A photonic link consumes several types of power: laser power, data-dependent and fixed power in the electrical transmitter and receiver circuits (where fixed power includes clock and static power), and thermal tuning power which is required to stabilize the frequency response of the thermally sensitive ring resonators. Table 1 shows our predictions based on current ongoing designs for the data-dependent and fixed power spent in the electrical circuits and in the in-plane heaters for thermal tuning. The laser power depends on the amount of optical loss that any given wavelength experiences as it travels from the laser, through various optical components, and eventually to the receiver (see Table 2). Some optical losses, such as non-linearity, photodetector loss, and filter drop loss, are independent of the main-memory bandwidth and layout. We will primarily focus on losses (waveguide loss, through-ring loss, and coupler loss) affected by the required main-memory bandwidth and layout of the photonic interconnect. These components contribute significantly to the total optical path loss and set the required optical laser power and correspondingly the electrical laser power.

For our photonic links, we assume that with double-ring filters and a 4 THz free-spectral range, up to 128 wavelengths modulated at 10 Gb/s can be placed on each waveguide (64 λ in each direction, interleaved to alleviate filter roll-off requirements and crosstalk). A non-linearity limit of 30 mW at 1 dB loss is assumed for the waveguides. The waveguides are single mode and a pitch of 4 µm minimizes the crosstalk between neighboring waveguides. The diameters of a regular modulator/filter ring is ≈10 µm and that of a comb filter ring is ≈40 µm. In the following photonic layouts, we conservatively assume that these photonic components can fit in a 50 µm STI trench around each waveguide, when monolithically integrated. We also project from our ongoing circuit designs that the area of the photonic E/O transceiver circuit is around 0.01 mm^2 for modulator driver, data, and clock receivers and associated SerDes datapaths. From [14] we assume that area for an electrical I/O transceiver will be mostly bump-pitch limited at around 0.25 mm^2. While chip-to-chip links with energy as low as 1 pJ/bt at rates of 10-20 Gb/s have been demonstrated to date in advanced logic process nodes [5, 18], they operate over very mild channels with roughly -10 dB loss using weak forms of equalization. At 20 Gb/s, memory channels typically have roughly -20 dB of loss, which in combination with slower transistors on the DRAM chip makes it difficult to preserve such low energy while increasing the amount of equalization and transmit-swing or receiver gain necessary to operate at this higher loss. Hence, in our analysis we use the results from an emulated DRAM transceiver chip [14], scaled optimistically to both lower energy per bit and higher data rate per pin.

4. PIDRAM CHANNEL ORGANIZATION

As described in Section 2, a DRAM memory channel uses a memory controller to manage a set of DRAM banks that are distributed across one or more DRAM chips. The memory system includes three logical buses: a command bus, a write-data bus, and a read-data bus. Figure 4 illustrates three ways to implement these buses using the photonic components discussed in the previous section. For now we assume that a PIDRAM bank never needs to be distributed across multiple chips, and later in this section we will revisit this assumption.

Figure 4(a) shows a *shared photonic bus*, which is similar to several photonic DRAM proposals in the literature [7, 24] and logically works like a standard electrical bus. In this implementation, the memory controller first broadcasts a command to all of the banks and each bank determines if it is the target bank for the command. For a PIDRAM write command, just the target bank will then tune-in its photonic receiver on the write-data bus. The memory controller places the write data on this bus; the target bank will receive the data and then perform the corresponding write operation. The interaction of the command and write-data bus resembles the single-writer multiple-reader buses described in other nanophotonic networks [13, 26]. For a PIDRAM read command, just the target bank will perform the read operation and then use its modulator on the read-data bus to send the data back to the memory controller. The read-data bus resembles the multiple-writer single-reader buses described in other nanophotonic networks [24], except that the memory controller schedules the read-data bus to avoid any need for global arbitration.

At first glance, the shared photonic bus seems attractive since, when the bus is active, all of the optical laser power is fully utilized. Unfortunately, each bus is shared and the losses multiply together making the optical laser power an *exponential* function of the number of banks. If all of the banks are on the same PIDRAM chip, then the losses can be manageable. However, to scale to larger capacities, we will need to "daisy-chain" the shared photonic bus

through multiple PIDRAM chips. Each PIDRAM chip adds two coupler losses, waveguide losses, and extra ring losses. The coupler losses are particularly difficult to mitigate, since even with optimized designs and process steps, the coupler loss is usually more than 0.5 dB [21]. Large coupler losses and the exponential scaling of laser power combine to make the shared photonic bus feasible only for connecting banks within a PIDRAM chip as opposed to connecting banks across PIDRAM chips.

Figure 4(b) shows an alternative implementation that we call a *split photonic bus*, which divides the long shared bus into multiple branches. In the command and write-data bus, modulated laser power is still sent to all receivers, and in the read-data bus, laser power is still sent to all modulators. The split nature of the bus, however, means that the total laser power is roughly a *linear* function of the number of banks. If each bank was on its own PIDRAM chip, then we would use a couple of fibers per chip (one for modulated data and one for laser power) to connect the memory controller to each of the PIDRAM chips. Each optical path in the write-data bus would only traverse one optical coupler to leave the processor chip and one optical coupler to enter the PIDRAM chip regardless of the total number of banks. This implementation reduces the extra optical laser power as compared to a shared photonic bus at the cost of additional splitter and combiner losses in the memory controller. It also reduces the effective bandwidth density of the photonic bus, by increasing the number of fibers for the same effective bandwidth. In this example with four banks per bus, we will need to use four waveguides/fibers where we only needed one in the shared photonic bus. The significant bandwidth density of our photonic technology can potentially make this a reasonable tradeoff.

To further reduce the required optical power, we introduce the the concept of a *guided photonic bus*, shown in Figure 4(c), which uses optical power guiding in the form of photonic demultiplexers to actively direct power to just the target bank. Each photonic demultiplexer uses an array of either ring or comb filters, and these filters are actively tuned by the memory controller to guide the light down the desired branch, leaving the unused branches dark. For the command and write-data bus, the photonic demultiplexer is placed after the modulator to direct the modulated light to the target bank. For the read-data bus, the photonic demultiplexer is placed before the modulators to allow the memory controller to manage when to guide the light to the target bank for modulation. Since the optical power is always guided down a single branch, the total laser power is roughly *constant* and independent of the number of banks. The optical loss overhead due to the photonic demultiplexers and the reduced bandwidth density due to the branching make a guided photonic bus most attractive when working with relatively large per-bank optical losses.

Figure 5 illustrates in more detail our proposed PIDRAM memory system. The figure shows a processor chip with multiple independent PIDRAM memory channels; each memory channel includes a memory controller and a PIDRAM DIMM, which in turn includes a set of PIDRAM chips. To prevent the on-chip processor network from becoming a bottleneck, the multiple PIDRAM channels could be distributed around the processor chip and connected to the cores with one of the recently proposed on-chip photonic networks [9, 13, 24, 26]. In the rest of this paper, we will focus on the implementation of a single PIDRAM memory channel.

Each PIDRAM chip contains a set of banks, and each bank is completely contained within a single PIDRAM chip. We use a hybrid approach to implement each of the three logical buses. The memory scheduler within the memory controller orchestrates access to each bus to avoid conflicts. The command bus is implemented with a single wavelength on a guided photonic bus. The command wavelength is actively guided to the PIDRAM chip containing the target bank. Once on the PIDRAM chip, a single receiver converts the command into the electrical domain and then electrically broadcasts the command to all banks in the chip. Preliminary analysis suggested that an electrical broadcast will require significantly less energy than optically broadcasting the command wavelength to all banks. Very high bandwidth channels supporting many banks may require a second command wavelength to ensure sufficient command bandwidth. Both the write-data and read-data buses are implemented with a guided photonic bus to actively guide optical power to a single PIDRAM chip within a PIDRAM DIMM, and then they are implemented with a shared photonic bus to distribute the data within the PIDRAM chip. Section 5 will describe the exact floorplan for the PIDRAM chip, and Section 6 will provide more detail on the PIDRAM bank design.

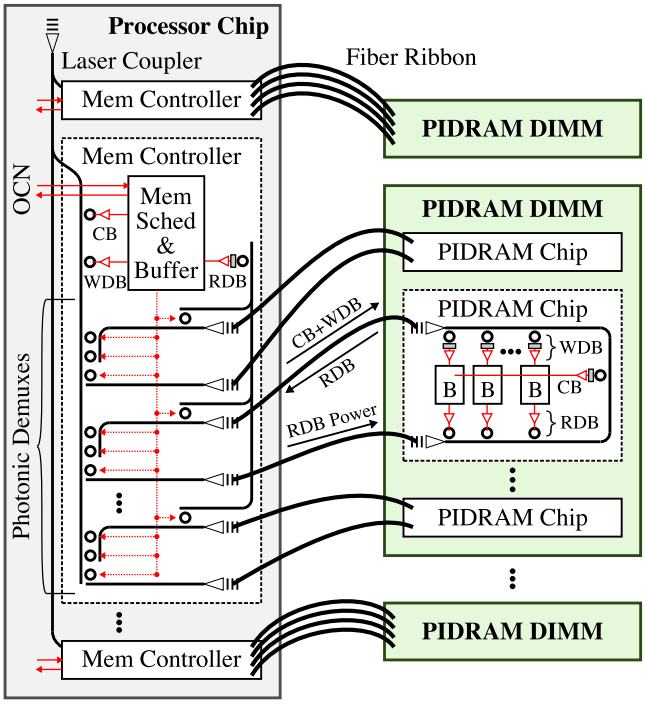
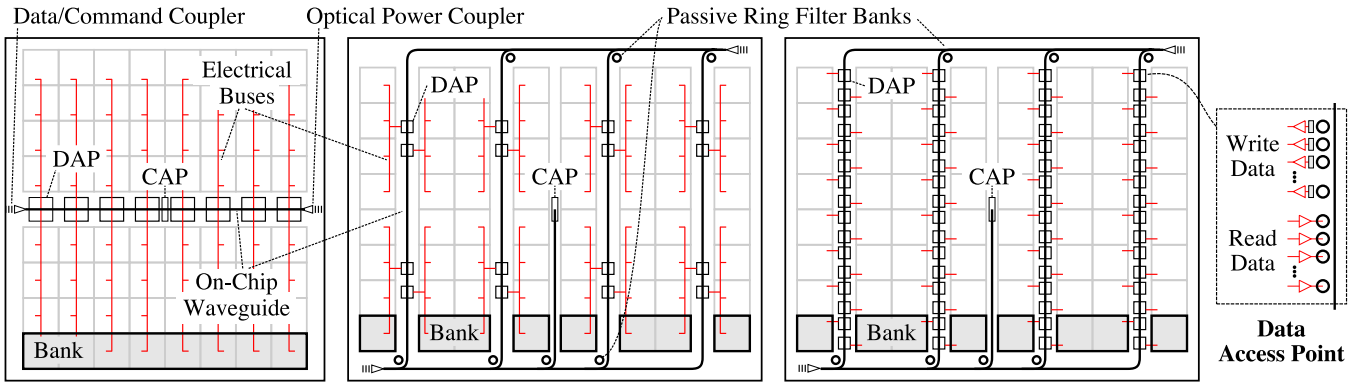


Figure 5: PIDRAM Memory System – Each PIDRAM memory channel connects to a PIDRAM DIMM via a fiber ribbon. The memory controller manages the command bus (CB), write-data bus (WDB), and read-data bus (RDB), which are wavelength division multiplexed onto the same fiber. Photonic demuxes guide power to only the active PIDRAM chip. (OCN = on-chip network, B = PIDRAM bank, each ring represents multiple rings for multi-wavelength buses, on-chip electrical wiring shown in red)

Providing one laser per PIDRAM chip is not economical, so a handful of large but efficient lasers are used with their optical power brought on to the processor chip and subsequently divided among the memory channels and guided to the PIDRAM chips. This approach places fibers destined for the same PIDRAM chip in physical proximity, allowing them to be grouped into a fiber ribbon to ease assembly. For economic and packaging reasons, we assume each PIDRAM chip will only have two fibers: one for the three buses and one for the unmodulated read-data optical power. Figure 5 illustrates how the command, write-data, and read-data buses



(a) P1 Floorplan (b) P2 Floorplan (c) P8 Floorplan

Figure 6: PIDRAM Chip Floorplans – Three floorplans are shown for an example PIDRAM chip with eight banks and eight array blocks per bank. For all floorplans, the photonic command bus ends at the command access point (CAP), and an electrical H-tree implementation efficiently broadcasts control bits from the command access point to all array blocks. For clarity, the on-chip electrical command bus is not shown, but it is similar to that shown in Figure 2(a). The data buses shown in the floorplans gradually extend the photonics deeper into the PIDRAM chip: (a) *P1* uses photonic chip I/O for the data buses but fully electrical on-chip data bus implementations, (b) *P2* uses seamless on-chip/off-chip photonics to distribute the data bus to a group of four banks, and (c) *P8* uses photonics all the way to each bank. (CAP = command access point, DAP = data access point, on-chip electrical buses shown in red)

are wavelength division multiplexed onto the same fiber. Routing the read-data optical power through the processor chip also enables a guided photonic bus implementation, since the corresponding photonic demultiplexer can be positioned within the appropriate memory controller.

So far in this section we have assumed that a bank is completely contained within a single PIDRAM chip, but this is in contrast to traditional electrical DRAM memory systems where banks are almost always distributed across multiple chips. Electrical DRAM chips are usually pin-bandwidth limited to a few bits per bus clock cycle. For example, to obtain all 512 bits needed for a 64-byte cache line, a bank might be striped across eight different chips and each access activates all chips in parallel. Unfortunately, this leads to large page sizes and wasted energy as most of the activated page is unused. Alternatively, electrical DRAM memory systems could activate just a single chip and wait for the entire cache line to stream out. Unfortunately, the limited pin bandwidth per chip will significantly increase serialization latency. A PIDRAM memory channel supports much higher bandwidth to each PIDRAM chip. A single data fiber can provide 80 GB/s in each direction, enabling an entire cache line to be stored in a single chip without incurring additional serialization latency. By locating a bank (and a cache-line access) on a single PIDRAM chip, the page size can be reduced by a factor of eight or more resulting in significant activation energy savings. Of course there are additional chip-level and bank-level considerations with providing this amount of memory bandwidth that will be discussed in the next two sections.

5. PIDRAM CHIP ORGANIZATION

In the previous section we motivated using guided photonic buses to implement the inter-chip portion of the command, write-data, and read-data buses while using shared photonic buses for the intra-chip portion of these buses. There is an important design trade-off in terms of how much of the on-chip portion of these buses should be implemented photonically versus electrically. This design choice is primarily driven by trade-offs in area and power.

Figure 6(a) illustrates the approach labeled *P1*, where the electrical I/O strip in Figure 2 is replaced with a horizontal waveguide and multiple *photonic access points*. Each photonic access point converts the corresponding bus between the optical and electrical domains. The on-chip electrical H-tree command bus and vertical electrical data buses remain as in traditional electrical DRAM shown in Figure 2.

Figures 6(b) and 6(c) illustrate our approach for implementing more of the on-chip portion of the data buses with photonics to improve cross-chip energy-efficiency. We use a *waterfall floorplan*, where the waveguides are distributed across the chip. The horizontal waveguides contain all of the wavelengths, and the optically passive ring filter banks at the bottom and top of the waterfall ensure that each of these vertical waveguides only contains a subset of the channel's wavelengths. Each of these vertical waveguides is analogous to the electrical vertical buses in *P1*, so a bank can still be striped across the chip horizontally to allow easy access to the on-chip photonic interconnect. Various waterfall floorplans are possible that correspond to more or less photonic access points. For a *Pn* floorplan, *n* indicates the number of partitions along each vertical electrical data bus. All of the photonic circuits have to be replicated at each data access point for each bus partition. This increases the fixed link power due to link transceiver circuits and ring heaters. It can also potentially lead to higher optical losses, due to the increased number of rings on the optical path. In Section 7 we further evaluate these trade-offs. Our photonic floorplans all use the same on-chip command bus implementation as traditional electrical DRAM: a command access point is positioned in the middle of the chip and an electrical H-tree command bus broadcasts the control and address information to all array blocks.

6. PIDRAM BANK ORGANIZATION

After redesigning the DRAM memory channel and DRAM chip to use an energy-efficient photonic interconnect, the next limiting factor is the power consumed within the banks themselves. During a bank access, every constituent array block activates an array core, which activates an entire array core row, of which only

a handful of bits are used. The energy spent activating the other bits is wasted, and this waste dominates bank energy consumption. Reducing wasted accesses while keeping the bank access size constant requires either decreasing the array core row size or increasing the number of I/Os per array core and using fewer array cores in parallel. Reducing the array core row size results in a greater area penalty due to less amortization of fixed area overheads, so we propose increasing the number of array core I/Os to improve access efficiency. Increasing the number of I/Os per array core, while keeping the bank size and access width constant, will have the effect of decreasing the number of array blocks per bank. Currently there is little incentive to make this change because the energy savings within the bank are small compared to the electrical inter-chip and intra-chip interconnect energy. Even if there were energy savings, current designs are also pin-bandwidth limited, so there would be no benefit to supporting such a wide bank access width, since that would increase serialization latency. The improvements in bandwidth density and energy efficiency realized through photonic bus implementations are key enablers for increasing the number of array core I/Os and thus reducing the bank activation energy.

With increased bandwidth, it also becomes advantageous to have more banks per chip, to hide bank busy times by interleaving between multiple parallel banks. Currently, however, rapid bank interleaving puts strain on the power delivery network of a DRAM chip because activates draw significant instantaneous current. To reduce the cost of the power delivery network, modern DRAM standards include two new timing constraints, t_{RRD} and t_{FAW}, which mandate minimum intervals between activate commands [16]. These constraints are the result of aggressive cost-cutting, and could potentially handicap any benefits from additional banks by limiting the number of activates per unit time. Recent industry focus on improving DRAM core efficiency have yielded designs that reduce t_{RRD} and t_{FAW} significantly [17]. Furthermore, implementing inter-chip communication with photonics frees up a number of pins, which can be used as power pins to improve power delivery. We can also reduce instantaneous power draw by decreasing the number of bits activated. Increasing the number of I/Os per array core reduces the number of array blocks activated, which reduces the total number of bits activated. But increasing chip access width forces each chip to activate more bits. Since we scaled access width up to an entire 512-bit cache line per chip, the number of I/Os per array core must also scale proportionally if we want to keep the number of activated bits constant.

7. EVALUATION OF SINGLE-CHIP PIDRAM MEMORY CHANNEL

In this section, we compare various PIDRAM configurations and floorplans to a baseline electrical DRAM implementation with the same capacity. This baseline design is labeled *E1* and is similar to that described in Section 2. In this section, we limit our study to a single-chip PIDRAM memory channel, and we ignore bandwidth-density constraints for the electrical baseline. In the next section, we will explore scaling to multi-chip PIDRAM memory channels.

7.1 Methodology

To evaluate the energy efficiency and area tradeoffs of the proposed DRAM architectures, we use a heavily modified version of the Cacti-D DRAM modeling tool [22]. Though we were able to use some of Cacti-D's original models for details such as decoder sizing, gate area calculations and technology parameter scaling, the design space we explored required a complete overhaul of Cacti-D's assumed DRAM organization and hierarchy. To this end, we built our own architectural models for the DRAM core, from circuit-level changes at the array core level, to the array block level and higher bank organization levels as shown in Figure 1, while relying on Cacti-D's original circuit models to handle most of the low-level circuit and process technology details. To validate our electrical models, we tested them against known points for a range of processes and configurations. In addition to covering the baseline electrical DRAM design, we accounted for the overhead of each relevant photonic design in our models and developed a comprehensive methodology for calculating the power and area overheads of off-chip I/O for both the electrical and photonic cases of interest. Since silicon photonics is an emerging technology, we also explore the space of possible results with both aggressive and conservative projections for photonic devices and photonic link circuits. All energy and area calculations presented are for a 32 nm DRAM process. For all of our designs the photonic command-bus power was a small enough fraction of the data-bus power that we ignore it for the purposes of our evaluation.

To quantify the performance of each DRAM design, we use a detailed cycle-level microarchitectural C++ simulator. We use synthetic traffic patterns to issue loads and stores at a rate capped by the number of in–flight messages. The memory controller converts requests into DRAM commands which are issued based on a round-robin arbitration scheme and various timing constraints based on contemporary timing parameters found in the Micron DDR3 SDRAM data sheet [16]. These timing parameters are also in agreement with our modified CACTI-D models. We simulate a range of different designs by varying: floorplan, number of I/Os per array core, number of banks, and the channel bandwidth. We use the events and statistics from the simulator to animate our DRAM and photonic device models to compute the energy per bit.

We find that for random traffic, a bank with a 512-bit access width has a bi-directional data bandwidth of approximately 10 Gb/s independent of system size, which matches our analytical model. Since each wavelength (λ) has a uni-directional bandwidth of 10 Gb/s, this translates to an equivalent bandwidth of 1/2 λ in each direction under balanced read and write workloads. Accordingly, we find the knee in the curve of sustained random bandwidth versus number of banks occurs when the number of λ per direction is half the number of banks.

For streaming traffic the effective bank bandwidth is higher, however, we believe random traffic is more representative of expected system traffic in future systems. In the manycore era, even if every core has locality in its access stream, there will be so many of them, that from the point of view of any memory controller, accesses will appear random. An intelligent memory controller could reorder the accesses to re-extract some of the locality, but this is unlikely to scale to many cores. Consequently, we perform most of our design and analysis assuming random traffic.

Latency is not an important figure of merit for this work because we do not expect PIDRAM to affect it significantly. We do not change the array core internals, which sets many of the inherent latencies for accessing DRAM. Moreover, our bank bandwidths are sufficiently sized such that the serialization latency is not significant, especially for random traffic, when compared to the inherent DRAM latencies. As to be expected, as the channel approaches peak utilization, the latency does rise dramatically due to queuing delays.

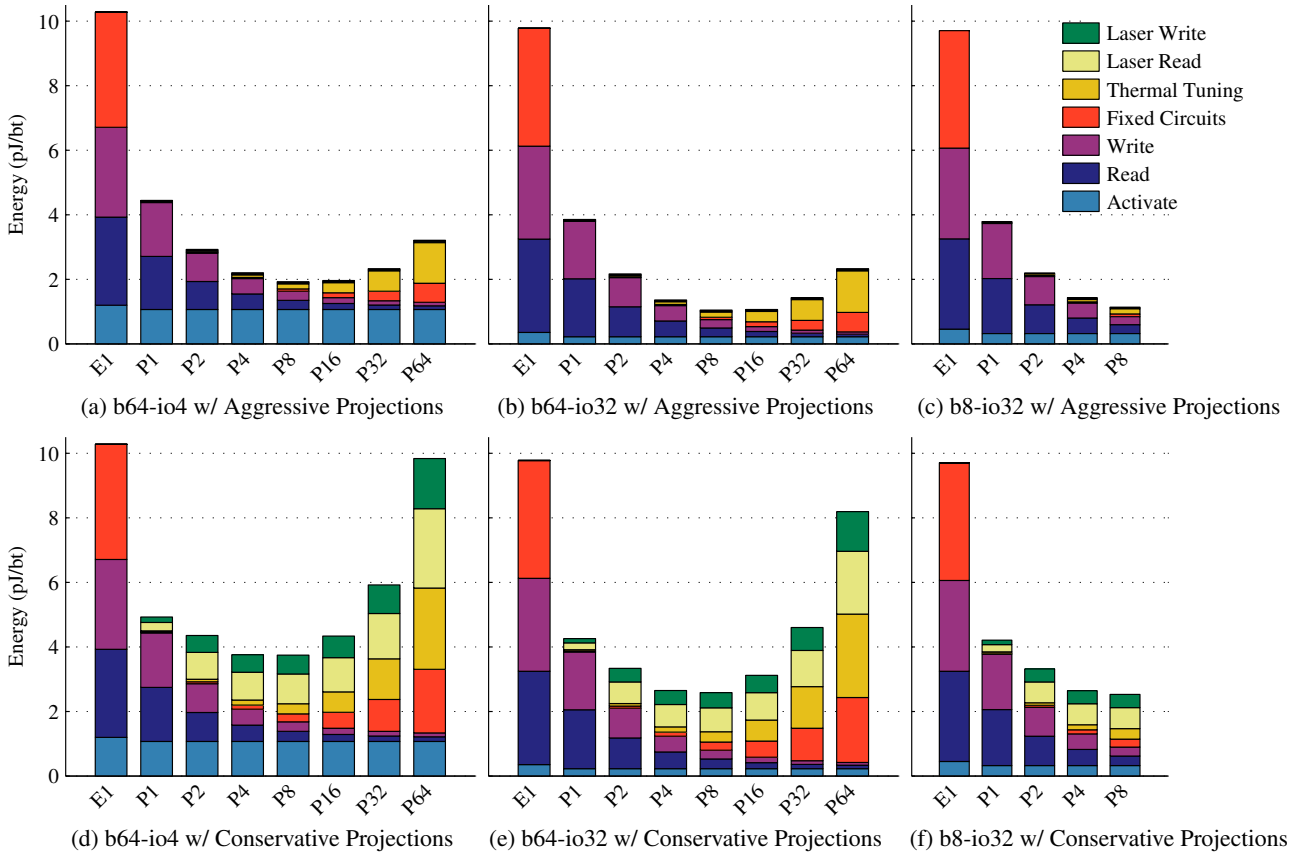


Figure 7: Energy Breakdown for Various DRAM Designs – (a–c) assume aggressive projections for photonic devices, while (d–f) assume more conservative projections as discussed in Section 3. Results for (a,b,d,e) are at a peak bandwidth of ≈500 Gb/s and (c,f) are at a peak bandwidth of ≈60 Gb/s with random traffic. Thermal tuning energy assumes 20 K temperature range, and fixed circuits energy includes clock and leakage. Read energy includes chip I/O read, cross-chip read, and bank read energy. Write energy includes chip I/O write, cross-chip write, and bank write energy. Activate energy includes chip I/O command, cross-chip row address energy, and bank activate energy.

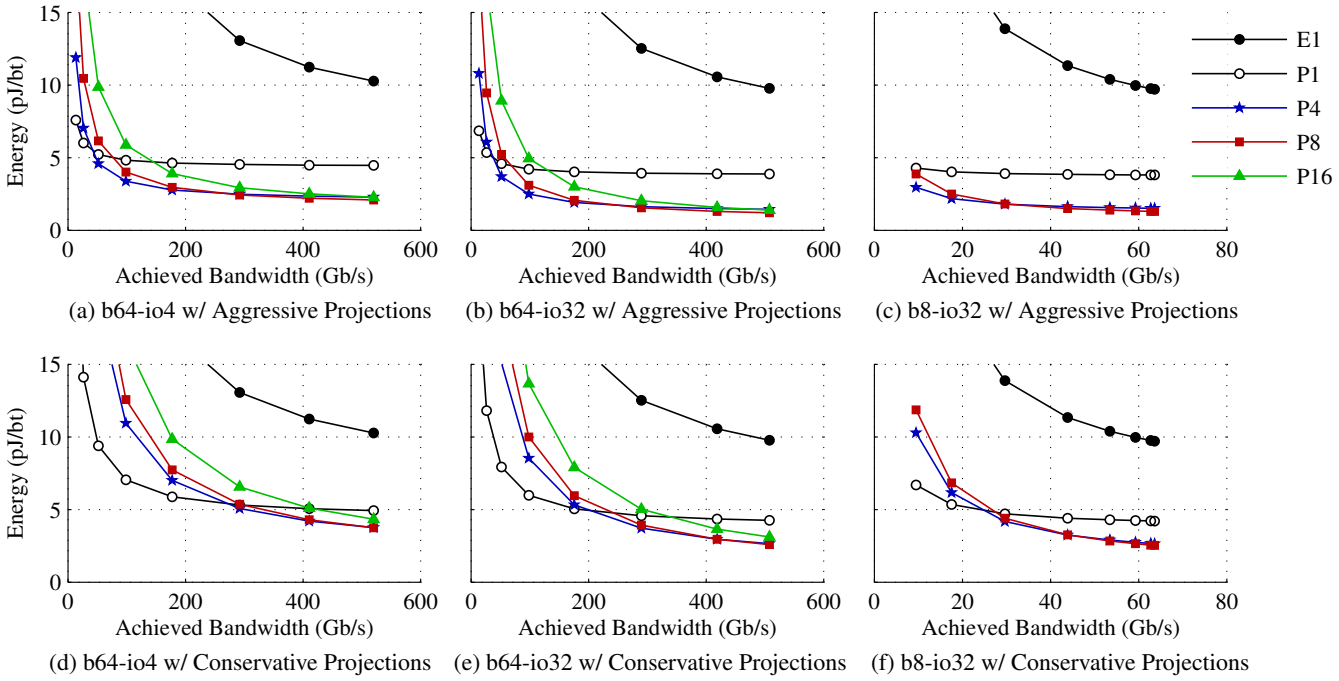
Although we evaluate hundreds of design points with our methodology, we will limit the rest of this section to the three configurations shown in Table 3. These configurations are either optimal or representative for their given parameters. The *b64-io4* configuration and the *b64-io32* configuration represent high-bandwidth PIDRAM chips (we include both to demonstrate the tradeoffs for changing the number of array core I/Os) and the *b8-io32* configuration represents low-bandwidth PIDRAM chips. All of our configurations are for a capacity of 8 Gb, which yields a reasonably sized chip given the 32 nm DRAM process technology. The DRAM-chip access width (bits per request) is 512 bits, which is scaled up from the 64 bits in contemporary DRAM. This is to enable the transfer of a 64-byte cache line from a single chip with a single request.

Parameter	b64-io4	b64-io32	b8-io32
Banks	64	64	8
Bandwidth (λ / direction)	32	32	4
I/Os per Array Core	4	32	32

Table 3: Representative Configurations – We explored designs consisting of 8, 16, 32, and 64 banks, each with 4, 8, 16, 32 and 64 λ/dir, and 4, 8, 16, and 32 I/Os per array core. All configurations were evaluated for all floorplans possible with that configuration.

7.2 Energy Breakdown

Figure 7 shows the energy-efficiency breakdown for various floorplans implementing our three representative PIDRAM configurations. Each design is subjected to a random traffic pattern at peak utilization and the results are shown for the aggressive and conservative photonic technology projections. Across all designs it is clear that replacing the off-chip links with photonics is advantageous, as *E1* towers above the rest of the designs. How far photonics is taken on chip, however, is a much richer design space. To achieve the optimal energy efficiency requires balancing both the data-dependent and data-independent components of the overall energy. For Figure 7, the data-independent energy includes: electrical laser power for the write bus, electrical laser power for the read bus, fixed circuit energy including clock and leakage, and thermal tuning energy. As shown in Figure 7(a), *P1* spends the majority of the energy on intra-chip communication (write and read energy) because the data must traverse long global wires to get to each bank. Taking photonics all the way to each array block with *P64* minimizes the cross-chip energy, but results in a large number of photonic access points (since the photonic access points in *P1* are replicated 64 times in the case of *P64*), contributing to the large data-independent component of the total energy. This is due to the fixed energy cost of photonic transceiver circuits and the energy spent on ring thermal tuning. By sharing the photonic access points across eight banks, the opti-



(a) b64-io4 w/ Aggressive Projections (b) b64-io32 w/ Aggressive Projections (c) b8-io32 w/ Aggressive Projections

(d) b64-io4 w/ Conservative Projections (e) b64-io32 w/ Conservative Projections (f) b8-io32 w/ Conservative Projections

Figure 8: Energy vs. Utilization – (a–c) assume aggressive projections for photonic devices, while (d–f) assume more conservative projections as discussed in Section 3. To reduce clutter, we only plot the three most energy efficient waterfall floorplans (*P4*, *P8*, *P16*). *P16* is not shown for (c,f) since *b8-io32* only has eight banks.

mal design is *P8*. This design balances the data-dependent savings of using intra-chip photonics with the data-independent overheads due to electrical laser power, fixed circuit power, and thermal tuning power.

Once the off-chip and cross-chip energies have been reduced (as in the *P8* floorplan for the *b64-io4* configuration), the activation energy becomes dominant. Figure 7(b) shows the results for the *b64-io32* configuration which increases the number of bits we read or write from each array core to 32. This further reduces the activate energy cost, and overall this optimized design is 10× more energy efficient than the baseline electrical design. Figure 7(c) shows similar tradeoffs for the low-bandwidth *b8-io32* configuration.

Figure 7(d–f) shows the same designs as Figure 7(a–c), but for conservative silicon-photonic technology assumptions. Replacing the off-chip links with silicon photonics still helps significantly, but bringing photonics across the chip closer to the array blocks is less of an improvement. This is a consequence of the lossier components which require more laser power. The optimal floorplan still appears to be *P8*, but it has a smaller margin over *P1*. Changing the number of I/Os per array core still proves to be beneficial, but this improvement is diluted.

7.3 Energy vs. Utilization

Although low power at peak throughput is important, a system designer is often just as concerned about energy efficiency at low utilization. For a given design, we scale back the utilization by reducing the number of messages that can be in flight, and the results are shown in Figure 8(a–c). As expected, the energy per bit increases as utilization goes down due to the data-independent power components. Although there are laser power, fixed circuit, and thermal tuning overheads for our PIDRAM designs, the fixed circuit overheads for the electrical baseline are significant enough to result in poor energy-efficiency regardless of utilization.

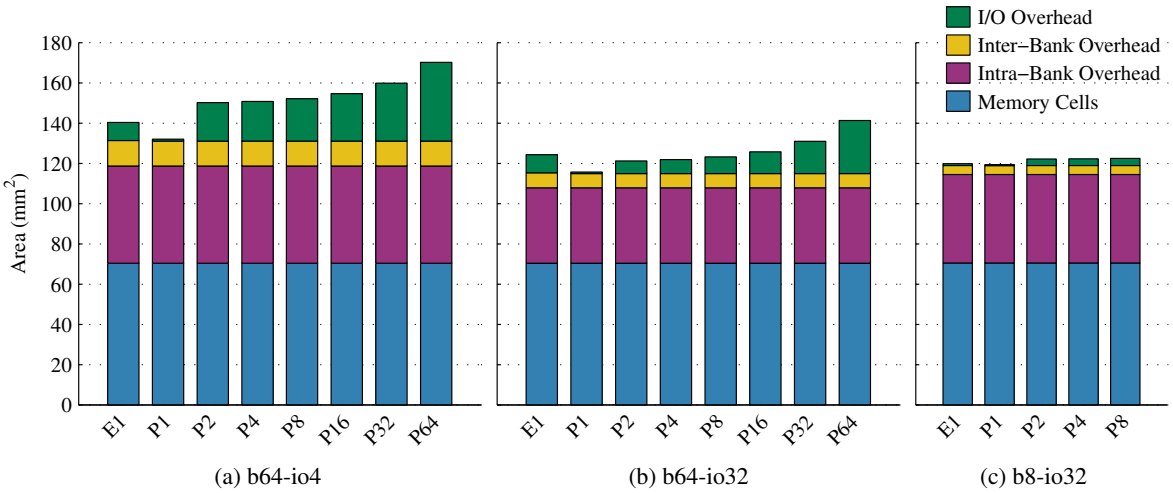
Systems with higher data-independent power will have a steeper slope, and this tradeoff can clearly be shown when comparing *P8* and *P16* to *P1*. The higher numbered *Pn* floorplans do better for the high-utilization cases because the global electrical wires connecting the array blocks to the photonic access points are shorter. However, they do worse than the *P1* floorplan for low utilization because the data-independent power of their rings and idle photonic circuits adds up. Essentially, this is a trade-off between the data-dependent and data-independent power components, and the target system utilization will determine the appropriate design.

Figure 8(d–f) shows the effects of less capable photonic devices, which result in a relatively large penalty for low utilization of high-bandwidth systems. This most notably affects the *P4*, *P8*, and *P16* floorplans.

7.4 Area

Figure 9 shows the total area breakdown of each design. Increasing the number of I/Os per array core results in significant area savings for all floorplans (less intra-bank and inter-bank overhead for *b64-io32* vs. *b64-io4*) because each array block has fixed area overheads. Recall from Section 6, increasing the number of I/Os per array block reduces the number of array blocks when keeping the access width constant.

Replacing the off-chip links with photonics results in significant area savings (*E1* vs. *P1*) due to the large size of bump-pitch limited electrical off-chip I/Os. Taking photonics deeper on-chip results in the jump in I/O overhead area between the *P1* and *P2* floorplans which can be explained by the move from the single photonic strip in *P1* to the waterfall in *P2* and above. Since we assume a very conservative 50 µm width for each photonic trench in the area cal-



(a) b64-io4 (b) b64-io32 (c) b8-io32

Figure 9: Area Breakdown for Various DRAM Designs with Aggressive Photonic Projections– The I/O overhead includes the area costs of circuits that form the access points as well as the area overhead of the waveguides. The inter-bank overhead is the area cost of wires and buffers used to bring bits from the banks to the photonic access points. The intra-bank overhead refers to the area taken up by the peripheral circuitry of each bank (e.g., decoders and sense-amplifiers). The memory cell area is the area taken up by the actual DRAM cells.

culations, the vertical trenches needed by the waterfall present a noticeable area increase. Interestingly, the losses in the conservative case require the laser power per wavelength to be so high that only 8 wavelengths can be supported per waveguide to stay within the 30 mW nonlinearity limit. This requires 16 waveguides instead of one in *P1* and two per column instead of one for the waterfall of *P2* and above. This however results in less than 1 mm^2 additional area overhead due to the compact waveguide pitch in each trench.

Much of the area savings come from increasing the number of I/Os per array core, but the most energy efficient design (*b64-io32-P8*), has slightly smaller die area than the electrical baseline (*b64-io32-E1*). Additionally, the electrical baseline I/O area is mostly bump-pitch limited and unlikely to scale much in more advanced process nodes, setting the upper bound on memory area efficiency for a required I/O bandwidth. Photonic access points, on the other hand, are relatively small both in size and number of vertical couplers. They are allowed to scale due to dense wavelength division multiplexing and continue to shrink with the scaling of electrical back-end circuits.

8. SCALING TO MULTI-CHIP PIDRAM MEMORY CHANNELS

When the number of PIDRAM chips per channel is scaled to increase capacity, the primary concern is the amount of laser power needed to overcome the extra losses that result from the overhead of adding more chips. In this section, we first quantitatively examine the laser power trade-offs between the shared, split, and guided photonic bus approaches, before qualitatively discussing 3D integration as a complementary technique for further capacity scaling.

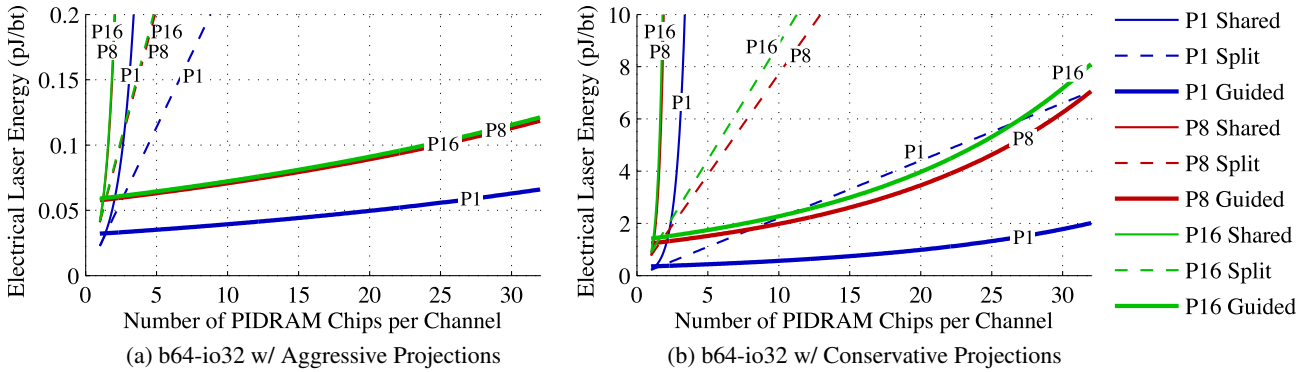
8.1 Optical Power Guiding

For our -20 dBm receiver sensitivity and 30% laser efficiency, an optical path loss in the range of roughly 15–25 dB is needed to keep the background laser power below the link energy cost. With a daisy-chained shared bus approach, the optical loss grows exponentially by the loss through a PIDRAM chip (3.5–7 dB aggressive or 7–13 dB conservative, depending on the floorplan) for each additional chip on the channel. With 5.5 dB (aggressive) to 10 dB (conservative) already lost in the memory controller waveguides, couplers, and rings, this approach becomes impractical beyond one or two chips. With 32 *b64-io32-P8* chips sitting on a channel implemented as a shared bus, the optical loss grows to 213 dB and 407 dB for the aggressive and conservative projections, respectively.

The split bus approach fares significantly better than shared bus as the required laser power grows roughly linearly with the number of chips per channel. For a single *b64-io32-P8* chip channel, the optical loss is 12 dB aggressive and 22 dB conservative, and grows to 27 dB and 37 dB when 32 *b64-io32-P8* chips are attached the channel for the aggressive and conservative projections, respectively.

With a guided bus, the laser power is sent only to the necessary chip. The fixed loss in the memory controller increases by 2–3 dB due to the extra power guiding ring and the need to also couple the read path laser in and out of the memory controller. A second increase in the memory controller loss results from the power guiding rings added to the memory controller with each additional chip. More rings along the path means more ring through loss and longer waveguides within the memory controller, amounting to an extra 0.1 dB to 0.3 dB loss for each additional chip. A guided bus channel with 32 *b64-io32-P8* chips has an optical loss of 17 dB and 33 dB for the aggressive and conservative projections, respectively.

Figure 10 shows how much the laser power contributes to the overall energy/bit for several floorplans of the *b64-io32* configuration. We can see that the guided bus designs have much more room to scale, as the shared and split bus approaches quickly become infeasible after only a few chips. As expected, designs that do not go as far into the PIDRAM chip consume less power, which makes sense since the PIDRAM chips themselves contribute less loss to the optical critical path. Interestingly, with conservative components, the split bus in Figure 10(b) can outperform the guided bus for smaller number of chips per channel, because the loss-overhead of guiding on the memory controller side is bigger than the linear increase in power required for the split bus.



(a) b64-io32 w/ Aggressive Projections

(b) b64-io32 w/ Conservative Projections

Figure 10: Electrical Laser Power Scaling for Multi-Chip PIDRAM Memory Channels – Laser power increases more slowly with the proposed guided photonic bus implementation versus either the shared or the split photonic bus implementations.

8.2 3D Integration

Three dimensional stacking is a complementary technology to photonics. This technology can be used to increase the capacity of PIDRAM memory channels without additional fiber wiring and packaging overhead. We introduce the concept of a PIDRAM cube, which is a collection of stacked PIDRAM dies (e.g., as in [11]), connected electrically by through-silicon vias and optically by vertical coupling in through-silicon via holes.

Stacking can be especially useful for high-capacity systems, where a significant fraction of the fibers would be unused with optical power guiding. By stacking these chips in a PIDRAM cube and adding a second stage of power guiding within the stack, we can reduce the number of packages and fibers in the system while maintaining the same capacity and bandwidth. The first stage of power guiding determines which PIDRAM cube gets the channel, while the second stage determines which die in the cube gets the channel.

With our photonic design, all of the dies in the stack after the base die will be the same, which greatly reduces the manufacturing costs. For example, for a stack of eight die, the generic die needs a total of 16 couplers. Only one in each direction will be active in any given die, and the others will be drilled-out by the TSV holes. Although stacking DRAM chips on top of the CPU die may increase the DRAM chip temperature (and hence refresh power), stacking PIDRAM chips in a cube away from the CPU should have minimal temperature effects on PIDRAM leakage and photonic components as overall PIDRAM chip power dissipation is relatively small.

9. RELATED WORK

Techniques such as microthreading [25] and multicore DIMMs [1] reduce the effective page size in current electrical memory modules by partitioning a wide multichip DRAM interface into multiple narrow DRAM interfaces each with fewer chips. Since both approaches use commodity electrical DRAM chips, they result in either smaller access sizes or longer serialization latencies. Architectural modifications to commodity DRAM, such as single subarray access and selective bitline access [23], also trade-off number of activated bits with serialization latency and area. Although our approach also reduces the ratio of activated to accessed bits, the energy-efficiency and bandwidth-density advantages of silicon photonics allows us to maintain the original access size and access latency while in addition reducing on-chip and off-chip interconnect energy.

Several researchers have proposed leveraging alternative technologies such as 3D stacking [11, 15, 20] and proximity communication [4] to address the manycore memory bandwidth challenge. Both technologies would offer either local DRAM physically packaged with the processor chip, or a tightly integrated multichip memory module connected to the processors via standard electrical interconnect. 3D stacking relies on through-silicon vias (TSVs) to communicate between layers, but monolithically integrated silicon photonics can also use the TSV holes for free-space optical communication. An optical scheme would have significantly higher bandwidth density than recently demonstrated stacked DRAM [11], requiring fewer TSVs while improving yield, as metal is not required to fill the TSV. Even a more advanced TSV technology with a 10 µm pitch at 20 Gb/s per via, would offer 5–10× lower bandwidth-density compared to an integrated optical vertical coupler. Furthermore, 3D stacking does not improve the horizontal communication required to connect any processing core to any memory bank. Although tight integration of a few DRAM chips and compute logic in a single package can serve some markets well, other systems demand much larger ratios of DRAM to compute logic. Stacked memory modules improve DRAM capacity but are connected to the processor chip through energy-inefficient and comparatively low-bandwidth electrical interconnect. In contrast, PIDRAM provides a flexible and scalable way to support different system configurations with high bandwidth density and energy efficiency.

Previous studies have illustrated the advantages of using an optical channel between on-chip memory controllers and a *buffer chip* positioned near a rank of DRAM chips. These schemes used either shared buses with arbitration at the memory controller [7] or point-to-point links with arbitration at the buffer chip [3]. Our work examines the channel-level, chip-level, and bank-level implications of fully integrating photonics into the actual DRAM chip, and our analysis shows the importance of considering all of these aspects to realize significant energy-efficiency gains. The Corona system briefly mentions a photonic memory-controller to buffer-chip channel, but then proposes using 3D stacked DRAM to mitigate the buffer chip to DRAM energy [24]. Although this mitigates some of the disadvantages of 3D stacking mentioned earlier, the Corona scheme relies on daisy-chained memory modules to increase capacity. We have found that this channel-level organization places stringent constraints on the device optical loss parameters, especially waveguide and coupler loss. In this work, we have proposed optical power guiding as a new way to increase capacity with less aggressive devices. The Corona work assumed a single DRAM or-

ganization, but our studies have shown that the advantages of photonics vary widely depending on the channel-level, chip-level, and bank-level configurations.

10. CONCLUSION

Photonic technology continues to improve rapidly, and could well form an essential ingredient in future multiprocessor performance scaling by improving interconnect performance. We have developed new photonically integrated DRAM architectures, and while photonics is a clear win for chip-to-chip communication, leveraging this new technology on the PIDRAM chip itself requires a careful balance between fixed and dynamic power. Background power is as important as active power, and should be addressed in future photonic device development, e.g., by providing fine-grain control of laser power to give better energy-efficiency at low to zero utilization. Surprisingly, despite the need to increase the number of array core I/O lines to obtain greater energy savings, PIDRAM area does not necessarily grow thanks to the high bandwidth density. High performance, low cost, and energy efficiency at the chip level are not sufficient for PIDRAMs to become successful high-volume commodity parts, they must also support a wide range of multi-chip configurations with different capacity and bandwidth tradeoffs. Our new laser power guiding technique can be used to construct a scalable high-capacity memory system with low background power by only illuminating active paths in the memory system.

ACKNOWLEDGMENTS

This work was supported in part by Intel Corporation and DARPA awards W911NF-08-1-0134, W911NF-08-1-0139, and W911NF-09-1-0342. Research also supported in part by Microsoft (Award #024263) and Intel (Award #024894) funding and by matching funding by U.C. Discovery (Award #DIG07-10227).

We would like to acknowledge the MIT photonic team, including J. Orcutt, A. Khilo, M. Popovic, C. Holzwarth, B. Moss, H. Li, M. Georgas, J. Leu, J. Sun, C. Sorace, F. Kaertner, J. Hoyt, R. Ram, and H. Smith. We would also like to thank David Patterson and members of the Berkeley ParLab for feedback.

REFERENCES

[1] J. Ahn et al. Multicore DIMM: An energy-efficient memory module with independently controlled DRAMs. *Computer Architecture Letters*, 8(1):5–8, Jan/Jun 2009.

[2] T. Barwicz et al. Silicon photonics for compact, energy-efficient interconnects. *Journal of Optical Networking*, 6(1):63–73, Jan 2007.

[3] C. Batten et al. Building manycore processor-to-DRAM networks with monolithic CMOS silicon photonics. *IEEE Micro*, 29(4):8–21, Jul/Aug 2009.

[4] R. Drost et al. Challenges in building a flat-bandwidth memory hierarchy for a large scale computer with proximity communication. *Int'l Symp. on High-Performance Interconnects*, Aug 2005.

[5] K. Fukuda et al. A 12.3 mW 12.5 Gb/s complete transceiver in 65 nm CMOS. *Int'l Solid-State Circuits Conf.*, Feb 2010.

[6] C. Gunn. CMOS photonics for high-speed interconnects. *IEEE Micro*, 26(2):58–66, Mar/Apr 2006.

[7] A. Hadke et al. OCDIMM: Scaling the DRAM memory wall using WDM based optical interconnects. *Int'l Symp. on High-Performance Interconnects*, Aug 2008.

[8] C. Holzwarth et al. Localized substrate removal technique enabling strong-confinement microphotonics in bulk Si CMOS processes. *Conf. on Lasers and Electro-Optics*, May 2008.

[9] A. Joshi et al. Silicon-photonic Clos networks for global on-chip communication. *Int'l Symp. on Networks-on-Chip*, May 2009.

[10] I. Jung et al. Performance boosting of peripheral transistor for high density 4 Gb DRAM technologies by SiGe selective epitaxial growth technique. *Int'l SiGe Technology and Device Mtg.*, 2006.

[11] U. Kang et al. 8 Gb 3D DDR3 DRAM using through-silicon-via technology. *Int'l Solid-State Circuits Conf.*, Feb 2009.

[12] B. Keeth et al. *DRAM Circuit Design: Fundamental and High-Speed Topics*. Wiley-IEEE Press, 2008.

[13] N. Kırman et al. Leveraging optical technology in future bus-based chip multiprocessors. *MICRO*, Dec 2006.

[14] H. Lee et al. A 16 Gb/s/link, 64 GB/s bidirectional asymmetric memory interface. *IEEE Journal of Solid-State Circuits*, 44(4):1235–1247, Apr 2009.

[15] G. Loh. 3D-stacked memory architectures for multi-core processors. *ISCA*, June 2008.

[16] Micron DDR SDRAM products. Online Datasheet, `http://www.micron.com/products/dram/ddr3`.

[17] T.-Y. Oh et al. A 7 Gb/s/pin GDDR5 SDRAM with 2.5 ns bank-to-bank active time and no bank-group restriction. *Int'l Solid-State Circuits Conf.*, Feb 2010.

[18] F. O'Mahony et al. A 47x10 Gb/s 1.4 mW/(Gb/s) parallel interface in 45 nm CMOS. *Int'l Solid-State Circuits Conf.*, Feb 2010.

[19] J. Orcutt et al. Demonstration of an electronic photonic integrated circuit in a commercial scaled bulk CMOS process. *Conf. on Lasers and Electro-Optics*, May 2008.

[20] H. Sun et al. 3D DRAM design and application to 3D multi-core systems. *IEEE Design and Test of Computers*, 26(5):36–47, Sep/Oct 2009.

[21] D. Taillaert, P. Bienstman, and R. Baets. Compact efficient broadband grating coupler for silicon-on-insulator waveguides. *Optics Letters*, 29(23):2749–2751, Dec 2004.

[22] S. Thoziyoor et al. A comprehensive memory modeling tool and its application to the design and analysis of future memory hierarchies. *ISCA*, June 2008.

[23] A. Udipi et al. Rethinking DRAM design and organization for energy-constrained multi-cores. *ISCA*, June 2010.

[24] D. Vantrease et al. Corona: System implications of emerging nanophotonic technology. *ISCA*, June 2008.

[25] F. Ware and C. Hampel. Improving power and data efficiency with threaded memory modules. *Int'l Conf. on Computer Design*, Oct 2007.

[26] P. Yan et al. Firefly: Illuminating on-chip networks with nanophotonics. *ISCA*, June 2009.