PeerJ Computer Science

Today's computing challenges: opportunities for computer hardware design

Woorham Bae[1,2]

[1] Circuits Department, Ayar Labs, Santa Clara, CA, USA
[2] Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, CA, USA

ABSTRACT

Due to the explosive increase of digital data creation, demand on advancement of computing capability is ever increasing. However, the legacy approaches that we have used for continuous improvement of three elements of computer (process, memory, and interconnect) have started facing their limits, and therefore are not as effective as they used to be and are also expected to reach the end in the near future. Evidently, it is a large challenge for computer hardware industry. However, at the same time it also provides great opportunities for the hardware design industry to develop novel technologies and to take leadership away from incumbents. This paper reviews the technical challenges that today's computing systems are facing and introduces potential directions for continuous advancement of computing capability, and discusses where computer hardware designers find good opportunities to contribute.

Subjects Computer Aided Design, Computer Architecture, Computer Networks and Communications
Keywords Computer, Hardware, Silicon, Processor, Memory, Interconnect, Moore's law, Circuit design

Submitted 4 November 2020
Accepted 9 February 2021
Published 30 March 2021

Corresponding author
Woorham Bae,
wrbae@eecs.berkeley.edu

Academic editor
Arun Somani

Additional Information and Declarations can be found on page 18

DOI 10.7717/peerj-cs.420

Copyright
2021 Bae

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

INTRODUCTION

These days, the world has been evolving very fast in various areas. The focus of technology development has been also switching to realize better human experience, convenience, and happiness, rather than old focuses such as mass production, automation, or cost reduction. Such rapid changes severely impact on the silicon industry, which has been responsible for the computing capability of the planet for several decades. The impact can be either positive or negative; it can provide more opportunities but also introduce many challenges at the same time. In fact, the opportunities are all about data (*Horowitz, 2014*; *Kim, 2015*). That is mainly because the world needs more electronics to handle the data. As mentioned above, the world is pushing to realize a whole bunch of things (such as smart cities, security, autonomous vehicles…) for better human experience, convenience, and happiness. In order to do that, we need to create, replicate, and process all the data. Accordingly, all the surveys predict that the amount of digital data will increase exponentially in the next 10 years. For example, Fig. 1 shows Cisco's two reports on the amount of data creation in the world, which were released in 2017 and 2019 (*Cisco Visual Networking Index, 2017*; *Cisco Visual Networking Index, 2019*). Both reports predict

How to cite this article Bae W. 2021. Today's computing challenges: opportunities for computer hardware design. **PeerJ Comput. Sci. 7: e420 DOI 10.7717/peerj-cs.420**

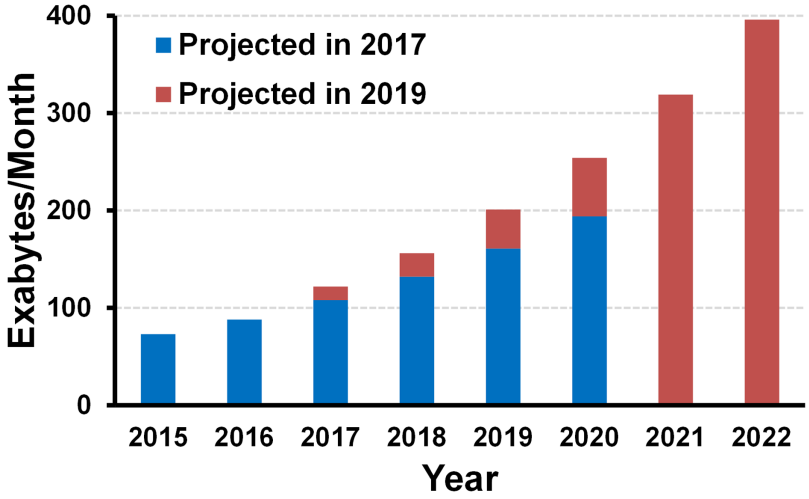

Figure 1 **Amount of digital data creation trend and projection (source: Cisco (*Cisco Visual Networking Index, 2017*, *2019*)).** Full-size DOI: 10.7717/peerj-cs.420/fig-1

that the amount of data will grow exponentially, but the 2019 report tells that the data has been created more than that expected 2 years before, and it will increase more rapidly.

The world is driven by data, and electronics are responsible for handling those data, which means that we need to create more and more electronics devices. Cisco predicts that the number of electronic devices will increase by almost twice in 5 years. Nvidia gives a bit more aggressive prediction, such that the number of total connected devices will increase by 16 times in 7 years (*Shao, 2019*). No matter how much it is, everybody expects that there will be more needs for electronic devices.

On the other hand, the challenges are also all about data. Here are some critical concerns one can raise in the age of such exploding data. How do we process such amount of data? Where should we store the data? How do we communicate with the data? And, what happens if we keep the same energy efficiency while the amount of data is exploding? Going back to the Fig. 1, where the Cisco's projection is shown, the amount of digital data is going to explode. If so, what happens if we keep the same energy efficiency for processing, storing, and communicating? Then the energy consumption will increase at the same rate as the data explosion, which is definitely not affordable. It is known that we are already consuming the largest portion of energy in the Earth for handling the data with electronics (*Bae, Jeong & Jeong, 2016*; *Whitney & Delforge, 2014*; *Pierce, 2018*); definitely such amount of data should not be affordable. From this observation, we would say that the energy efficiency must be improved proportional to the data explosion, at least to keep the same amount of energy consumption.

In fact, such explosion of data is not something that started yesterday, even though there might be some difference in degree. Hence, it is worthwhile trying to learn something from the history, how we did handle such exploding data before. Figure 2 shows a simplified computing system, where we can see a logic (processor) IC and a memory IC, and an interconnect link between them. Basically, in order to handle more data, we need

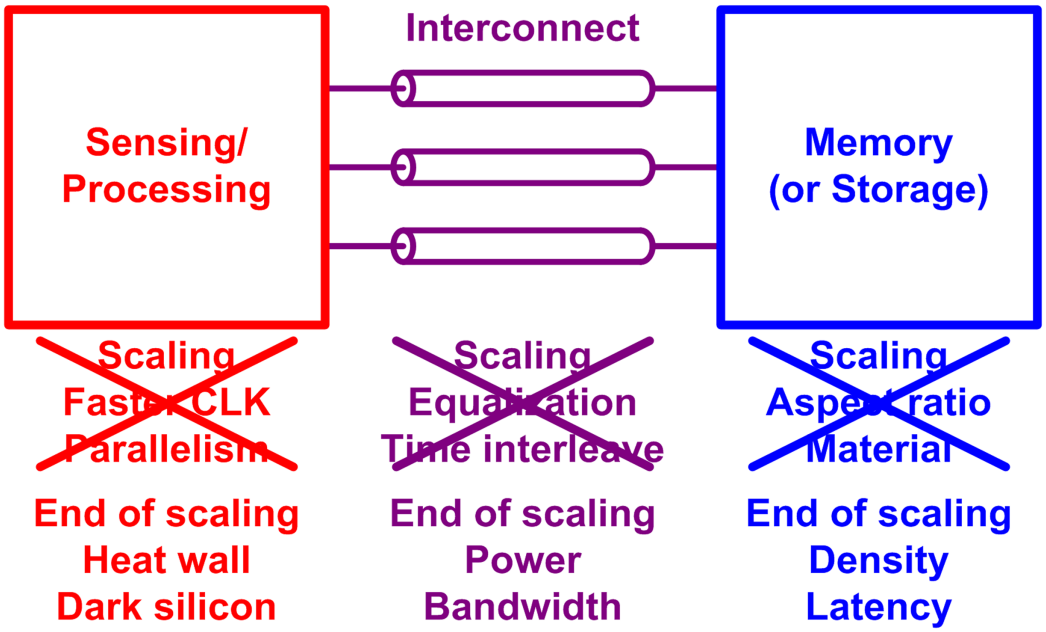

Figure 2 **Summary of how we have made a better computer and why that will not work for future computers.** Full-size DOI: 10.7717/peerj-cs.420/fig-2

higher processing speed, interconnect bandwidth, and memory density. Figure 2 also shows a simple summary of how we managed to enable it. For the processor side, the CMOS (complementary metal-oxide silicon) technology scaling, which is generally represented by Moore's law (*Moore, 1965*, *1975*), enables a transistor to be faster and consume even less power (*Holt, 2016*; *Bohr & Young, 2017*; *Mak & Martins, 2010*; *Yeric, 2019*). Once we have a faster transistor, we can raise the clock rate for faster processing. Once after the power scaling of transistor has been retarded due to some physical reasons (i.e., leakage current), people introduced parallelism such as multi-core processing to increase the processing speed without increasing the clock rate (*Danowitz et al., 2012*). For the memory side, the scaling of device footprint enabled a higher memory density (*Hwang, 2002*). However, extensive scaling led to many challenges, which were overcome by the memory industry with process innovations such as higher aspect ratio of DRAM, and material innovations like high-k materials (*Mueller et al., 2005*; *Sung et al., 2015*; *Jang et al., 2019*). For the interconnect side, the transistor scaling has been also a key enabler for a higher bandwidth, because a faster transistor makes a circuit faster (*Daly, Fujino & Smith, 2018*; *Horowitz, Yang & Sidiropoulos, 1998*). However, the electrical channels (wires) which bridge separate ICs cannot be scaled with the silicon technology, as they present in the physical world, not in the digital IC world. That is, an electrical channel has a finite bandwidth so that high-frequency components of transmitted signal attenuate over the channel. As a result, interconnect engineers had to make many innovations in equalization circuits which compensate the channel loss at high frequency, that is to equalize the channel response at low and high frequency (*Horowitz, Yang & Sidiropoulos, 1998*; *Dally & Poulton, 1997*; *Zerbe et al., 2003*; *Stojanovic et al., 2005*; *Choi, Hwang & Jeong, 2004*).

They also introduced time-interleaving technique, which is something like the parallelism, to achieve very high speed even above the transistor limit (*Kim & Horowitz, 2002*; *Lee, Dally & Chiang, 2000*; *Musah et al., 2014*; *Bae et al., 2017*).

However, these legacy approaches cannot be good solutions for these days and the future. First of all, we are about to lose the almighty scaling. The scaling has not been fully finished yet; however, it has been a while since the power scaling started being retarded as discussed earlier. As a result, increasing the clock frequency is no longer available because we do not want to burn the chip out. The parallelism was introduced to overcome such challenge, but it has also hit the limit because of the same heat dissipation issue. Only a fraction of the multi-cores can be turned on at the same time, which is called "dark silicon" (*Yeric, 2019*; *Esmaeilzadeh et al., 2011*).

The similar issue happened to the memory, that is, the scaling has been retarded which limits the increase in the memory density. The scaling also introduced many non-idealities so that there are many higher-level assistances which burden memory module and increase the latency of the memory. For the interconnect side, the channel loss becomes very significant as the required interconnect bandwidth increases, so the equalization circuitry consumes too much power. It will be tougher as the scaling is ending because we can no longer take advantage of faster transistors. To summarize this section, the legacy solutions for handling data explosion will not be as effective as they used to be for the today's and future computer. From the following section, we will discuss on the possible solutions that enable the continuous advance in computing capability for the next 10 years.

The remainder of this article is organized as follows. "Logic (System Semiconductor)" presents potential solutions for addressing the challenges of the logic and opportunities for the computer hardware design industry and engineers. "Memory and Storage" describes the recent innovations from the memory industry and discusses the future direction. In addition, the opportunities for design engineers from the revolution of the memory devices will be discussed. "Interconnect" discusses the recent trend of the interconnect technology and potential solutions to resolve the challenges that the interconnect technology is facing. Finally, conclusions are provided in "Conclusions".

Survey methodology

This review was conducted in Sept.–Oct. 2020. Three different approaches were used to collect research articles:

1. Searching Google scholar and IEEE Xplore with various keywords such as Moore's law, CMOS scaling, high-bandwidth memory, V-NAND, crosspoint memory, transceiver, PAM-4, and silicon photonics.
2. Starting from an initial pool of articles and then move back and forth between their citations and references.
3. Selecting articles based on their impact and credibility; Prioritizing articles with high citations or from top conferences and journals of the fields, such as JSSC, TCAS, TCAD, TED, AELM, ISSCC, VLSI, IEDM.

LOGIC (SYSTEM SEMICONDUCTOR)

Efficient computing with specialized IC

In this section, the technology directions for silicon logic to maximize the opportunity for the hardware design is discussed. While dealing with technology development of computation logic, it is inevitable to discuss the scaling limit of semiconductor process technology, the end of Moore's law. In fact, since 2014, there have been at least one of plenary talks at *International Solid-State Circuits Conference* (ISSCC) that discuss on preparing the end of the Moore's law, for example *Kim (2015)* and *Vandersypen & van Leeuwenhoek (2017)*. So, let us take a quick look at where the scaling limit comes from. In his talk at ISSCC2015 (*Kim, 2015*), the president of Samsung Electronics told that the physical limit of transistor dimension is around 1.5 nm, which is given from Heisenberg's uncertainty principle. However, he also told that he expected that the practical limit will be 3 nm. After 5 years, now, the 7 nm technology is already widely available in the industry. And the leading foundries such as TSMC and Samsung Electronics are already working on 5 nm and 3 nm technology development, which means that we are almost there.

As a result, recalling the energy discussion in the "Introduction", the appropriate question for this point should be how we can improve the energy efficiency without scaling. We can find some hints from today's mining industry, the cryptocurrency mining, where the computing energy efficiency is directly translated to the money. Recalling 2017, when the cryptocurrency value hit the first peak, the readers may remember that the graphics processing unit (GPU) price became very expensive. It is simply because the GPU is much more efficient than central processing units (CPU), so mining with GPU gave more profit margin. Then, why the GPU is much efficient compared to CPU?

It is because it is specialized. CPU is more generic, but the GPU is more specific. That is, there is a computing trade-off between the flexibility and the efficiency. After finding that, people went to field-programmable gate array (FPGA) for cryptocurrency mining for better efficiency, and eventually they end up with designing application-specific integrated circuit (ASIC) just for the mining. Figure 3 shows the survey of various cryptocurrency miners, where we can find an ASIC miner provides 10^4 times better efficiency than a CPU miner. From the observation, we can conclude that such a huge gain comes from the design of specialized ICs. To summarize, making specialized ICs is one of the top promising solution for the efficient computing. In accordance with that, the foundry companies would diversify their process technology instead of scaling it down, for example the Global Foundries 45 nm CLO process, which is specialized to silicon photonics (*Rakowski et al., 2020*).

Productivity problem of specialization

We found that the specialization would be a potential solution to resolve the energy problem and to retain the continuous advance of computing. However, there are also some downsides of the specialization, so we need to investigate how profit is made in the new age with the specialization. In a simplified model in Fig. 4A, a fabless company shipped

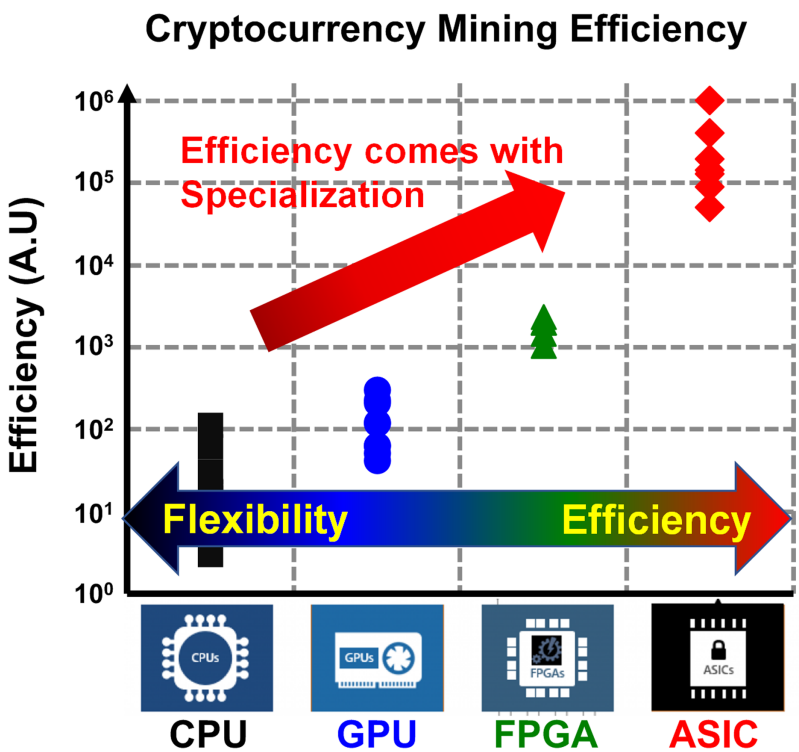

Figure 3 **Survey of energy efficiency of various cryptocurrency miners.**
Full-size DOI: 10.7717/peerj-cs.420/fig-3

1 million units of a generic chip before, but they are planning to design 10 specialized chips in 10 different processes to meet the better efficiency requirement. At the same time, they are expecting they can ship 2 million chips in total as there will be more demand of electronics. In the model, the company is currently making $3 million profit. On the right side of Fig. 4A, a linear extrapolation is made to when the company designs 10 specialized chips and total shipping is doubled. Note that all the cost is extrapolated in linearly proportional to the amount of production.

However, it is too optimistic projection. Figure 4B shows a bit more realistic model. The revenue and production cost are indeed proportional to the amount of shipping. However, does it make sense to extrapolate other expenses? Of course, the answer is no. For example, the amount of manpower cannot be scaled linearly. To design a single complete chip, they need analog engineers, digital engineers, manufacturing engineers and more. Therefore, it makes no sense that only 4 engineers can make a chip which used to be made by 20 engineers, even though there must be some amount of efforts that can be shared among the chip designs. So, the model in Fig. 4B assumes 10 engineers can design a specialized chip A0. If so, the profit becomes minus. The calculation here is very rough, but at least we can observe a large fraction of design cost is not scaled with the amount of production. The company would raise the price, but customers will not be happy with that. Then, is the specialization a false dream?

The most reasonable solution here is to reduce the design time, since such design costs are proportional to the design time, as shown in Fig. 4C. For example, if they can

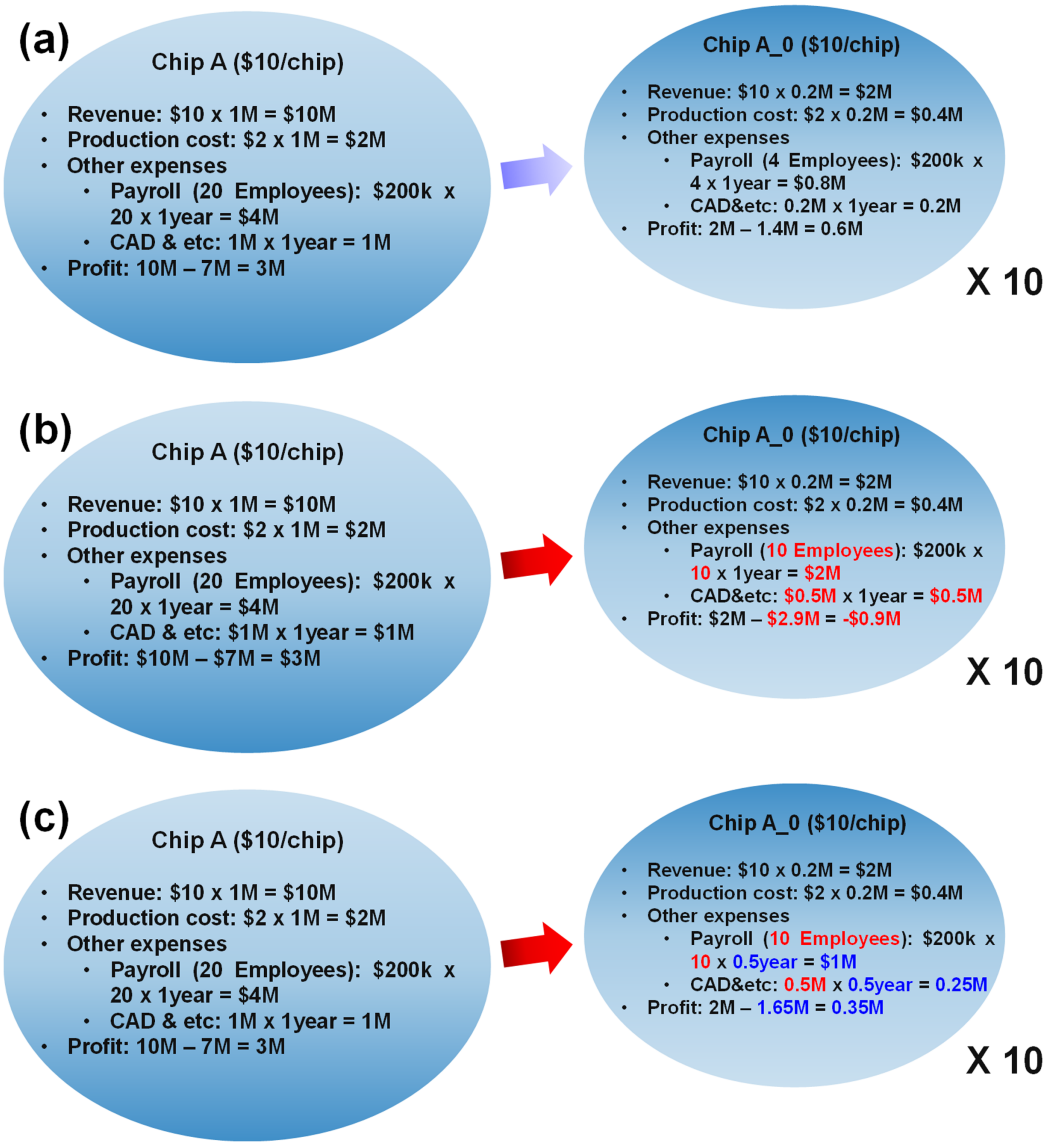

Figure 4 Productivity issue of specialization. Case study of (A) ideal case, (B) practical case, (C) practical case with reduced design time. Full-size DOI: 10.7717/peerj-cs.420/fig-4

reduce the design time by half, they can reduce the expenses by half, then they can make more profit. As mentioned earlier, they are designing 10 different but similar chips, and there is some amount of sharable efforts. That means, if they maximize the amount, they should be able to reduce the design time considerably.

Reducing design time by reusing design

Then what should we try to maximize shareability? Generally speaking, we can say the analog and mixed signal (AMS) circuit design is usually the bottleneck of reducing design time. That is mainly because AMS circuit designs highly rely on human's heuristic knowledge and skills, compared to the digital design. Moreover, the design complexity has been increased as technology scales down, due to the complex design rules and digital-

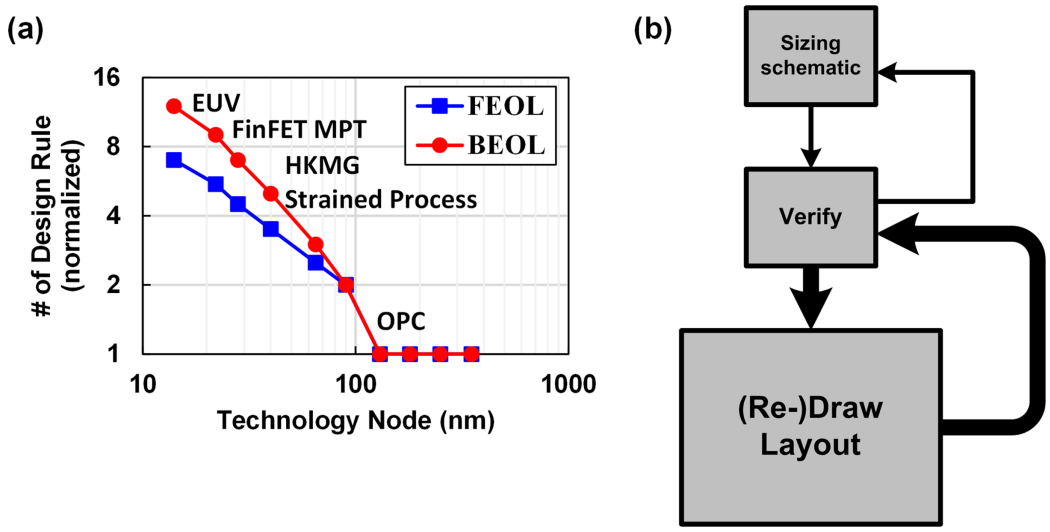

Figure 5 **(A) Silicon design complexity across technology node, (B) general design flow of AMS circuit.** Full-size DOI: 10.7717/peerj-cs.420/fig-5

friendly scaling of CMOS technology, which is represented by the number of design rules shown in Fig. 5A, where we can find the design complexity has been increased exponentially as the technology scales down (*White*; *Whitcombe & Nikolic, 2019*; *Han et al., 2021*). Figure 5B shows a general design flow of an AMS circuit. Once we decide a circuit topology, we carefully size the transistor dimensions based on some calculations, and run simulation using CAD tools. If the simulation result is not positive, we go back and tweak the sizing. Once we meet the spec with the schematic simulation, we proceed to draw the layout mask, after that we run parasitic extracted (PEX) simulation and check the result again. Based on the result, we have to go back and forward many times until the performance of the circuit is fully optimized. The main issue here is that most of time is spent for drawing layout, and its complexity has been increasing as shown in Fig. 5A. One may ask why we do not try to automate the analog design as we did for the digital design. However, in fact, it is hard to say we can do it for the layout design in the near future because there are only a few ways to do it right, however there are billions of ways to do it wrong. That means, to make the automation tool work correctly, a designer should constrain the tools very precisely (*Habal & Graeb, 2011*; *Lin, Chang & Lin, 2009*), so they spend most of the design time constraining the tools, which is not very efficient (*Chang et al., 2018*). That is the main reason why the engineers in this field rarely use such automation tools.

In fact, a better way is reusing, because reuse is a bit easier than automation. For example, we can just grab a good designer who knows how to do it correctly and let him/her do the design. At the same time, we enforce him/her to write down every single step he would do to create a correct output into an executable script (often called as a generator). Then the script has the way of doing right of the good designer, so the output should be correct no matter who run the script. However, because transistor shapes are different between process technologies, it is hard to automatically capture a design-rule-

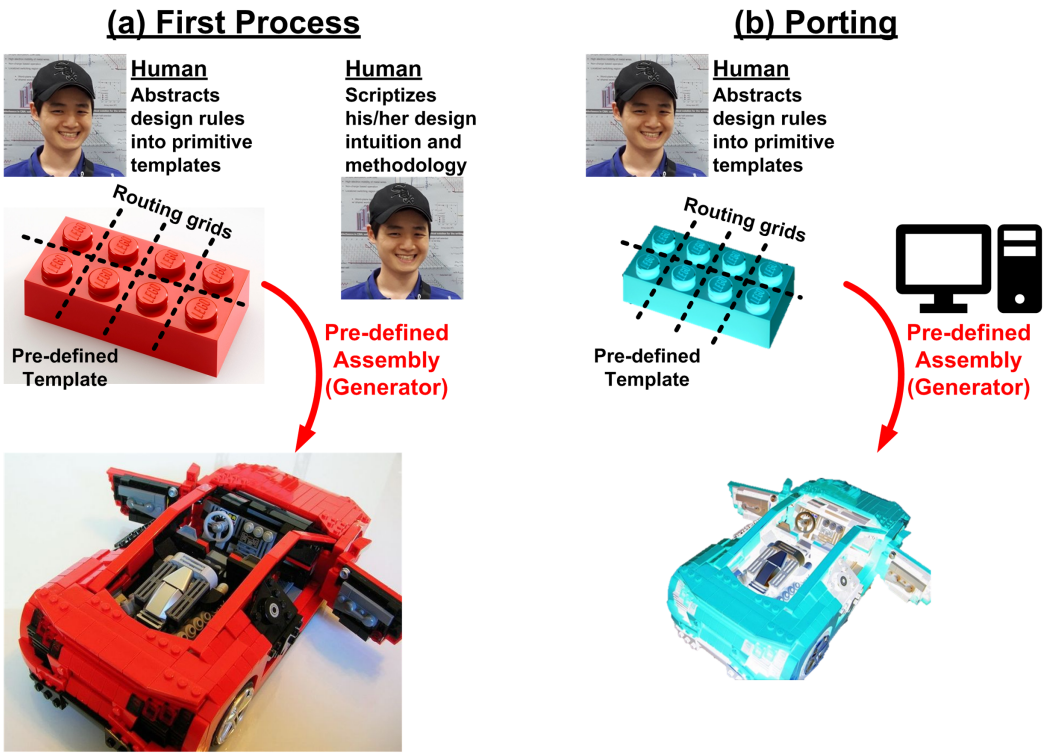

Figure 6 **Design flow based on reuse of design process using executable script (generator): (A) for the first process, (B) for porting designs to another process.**

Full-size DOI: 10.7717/peerj-cs.420/fig-6

compatible shape only with the script, without intervention of designer's heuristic knowledge. Therefore, such script-based approach works well in a single process technology, however it would face many challenges when ported to another technology. To address such portability issue, template-based approaches have been proposed in many works (*Crossley et al., 2013*; *Yilmaz & Dundar, 2008*; *Castro-Lopez et al., 2008*; *Martins, Lourenco & Horta, 2013*; *Kunal et al., 2019*; *Wang et al., 2019*). Instead of letting a layout script draw a layout from scratch, designers prepare design-rule-aware templates of primitive components. The script assembles the templates by following the way an expert designer pre-defined. It is like a Lego block, when we buy a Lego package, there are many unit blocks (templates) and an assembly manual (script), as shown in Fig. 6.

Such reuse-based approach is very attractive for the future of specialization, however there are some hurdles that the designers must overcome. In fact, the hurdle is not a matter of development of elegant CAD tools. Here is an example based on the author's engineering experiences. The author has used three different frameworks for helping such reusing process, the Laygo, XBase, and ACG (*Berkeley Analog Generator, 2021a*, *2021b*; *Ayar Custom Generator, 2021*). They are quite different each other as summarized in Fig. 7, for example the Laygo defines the templates more strictly so it more limits the degree of freedom, whereas the ACG has loose template definitions. There are pros and cons; the Laygo reduces the number of ways to do in wrong way for easier portability at the cost of sacrificing the degree of freedom. The ACG allows freedom however it

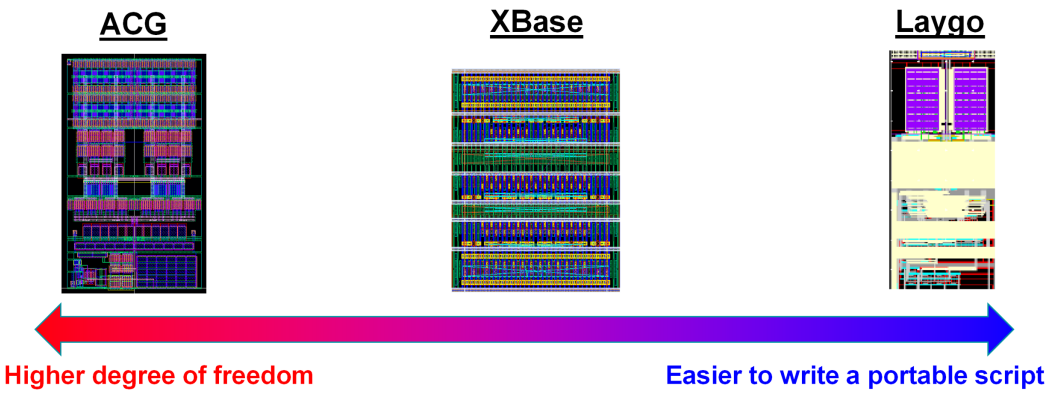

Figure 7 **Comparison of 3 different frameworks to support reusing process.**
Full-size DOI: 10.7717/peerj-cs.420/fig-7

burdens a designer spend more time on writing a portable script. That is, to summarize, there is just a trade-off. Designers should spend more time to make it portable (left side of Fig. 7) or they should spend more time to make it as good as custom design (right side of Fig. 7). For either way, a good script has to have flexible parameterizations (*Chang et al., 2018*). So, it is not a matter of which tool we would use. Instead, what is more important is whether a designer is willing to use this methodology or not, because analog designers are not generally familiar with such parameterization. In addition, writing a design script requires more skillsets and insights compared to custom designs. As a result, to take full benefit of the reusing, the designers must be patient and be willing to learn something, which is the main hurdle.

Once we overcome the hurdle, there will be more opportunities to further improve the productivity. For example, it allows a machine to accomplish the entire design iteration shown in Fig. 5B (*Settaluri et al., 2020*). Conventionally, it was believed that the design space is too huge to fully automate the optimization, even with schematic-only simulation. However, recent progress in deep learning technology enable handling such huge space, so a machine can handle the schematic optimization (*Hakhamaneshi et al., 2019*). However, as mentioned earlier, the layout automation is almost impossible so the machine must struggle with the layout loop. The script-based layout reuse can bridge the gap:
(1) The machine sizes the schematic parameters. (2) The layout script generates a layout from the parameters. (3) The machine runs PEX simulations and checks the results.
(4) Based on the results, the machine resizes the parameters and repeats (1)–(3) until the circuit is fully optimized. Many efforts should be preceded to fully realize such AI-based design, but it is evident that there will be tons of opportunities along the way.

MEMORY AND STORAGE

Memory scaling limit and 3-D integration

In the previous section, we discussed that the specialization and reuse of the design process will be one of the solutions for the challenges that the logic side is facing. In this section, recent progress and future technology for memory will be presented, and then the

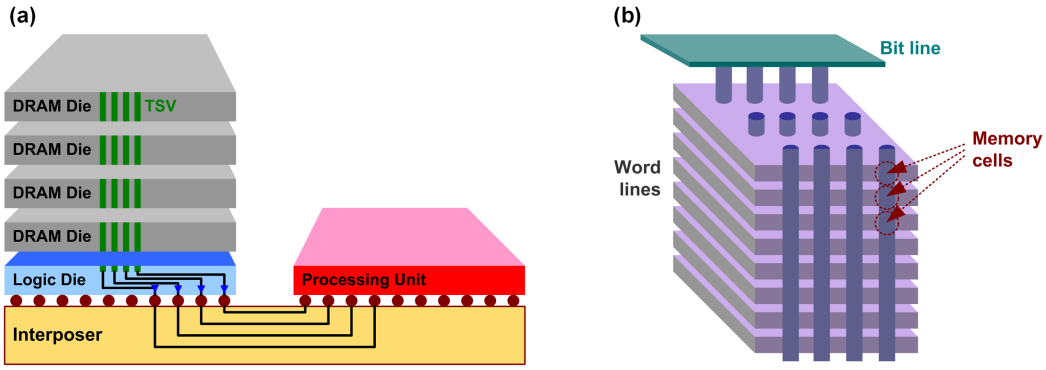

Figure 8 **Conceptual diagram of (A) HBM and (B) V-NAND.**
Full-size DOI: 10.7717/peerj-cs.420/fig-8

opportunities for hardware designers to contribute to the technology innovation will be discussed. In fact, in the memory industry, physics and device engineering have played more critical role compared to design engineering. For example, circuit topology of bit-line sense amplifier in memory module has not been changed for decades while the memory devices have been evolving. This trend is likely to continue in the future, however it is expected that the memory industry will need more innovations from design.

Let us briefly review the challenges that current memory is facing, which is mainly because of the scaling limit as discussed in the "Introduction". Basically, higher memory density is the top priority which has been enabled by the process scaling. For DRAM, however, lower capacitance due to extensive scaling results in many challenges such as short data retention, poor sensing margin, and interference. As a result, the scaling is no longer as effective as it used to be. Similarly, NAND flash also experiences many non-idealities introduced by the extensive scaling, such as short channel effect, leakage, and interference. Again, the scaling is not useful as it used to be. Recently, however, memory industry has found a very good way rather than pushing the device scaling too hard, they found the solutions from 3D stacking. Figure 8 shows the recent innovations with 3D stacking that have been developed for DRAM and NAND flash, high-bandwidth memory (HBM) and vertical NAND (V-NAND) (*Lee et al., 2014*; *O'Connor, 2014*; *Tran, 2016*; *Jun et al., 2017*; *Xu et al., 2015*; *Kim, Lee & Kim, 2016*; *Kim et al., 2009*, *2017*; *Tanaka et al., 2016*; *Im et al., 2015*; *Park et al., 2014*). In HBM, multiple DRAM dies are stacked, and they are connected by through silicon vias (TSV). A base logic die can be used to buffer between the DRAM stack and the processing unit (host SoC). The logic die and the processing unit are connected through micro-bumps and silicon interposer. Because the memory stack and the processing unit are not integrated in 3D manner, the HBM is often considered as 2.5D integration. Unique features of the HBM such as low capacitance of TSV, 2.5D integration, and high interconnect density of silicon interposer enable high capacity (not always), low power, and high bandwidth compared to legacy DRAM (*O'Connor, 2014*; *Tran, 2016*; *Jun et al., 2017*; *Ko et al., 2020*). On the other hand, in NAND flash, the memory cells themselves are stacked. Interestingly, nowadays it is higher than 100 layers. In fact, these much of innovations on the capacity, as well as

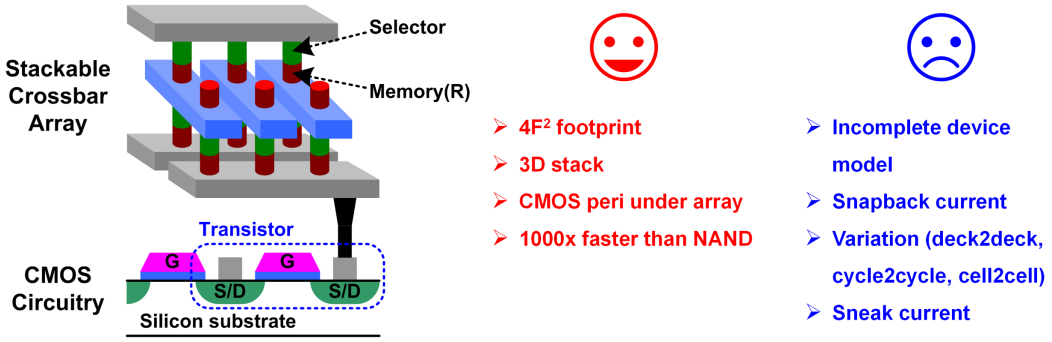

Figure 9 **New memory device with crossbar array structure.**
Full-size DOI: 10.7717/peerj-cs.420/fig-9

advancements on processing units, burden more on the interconnect side for higher bandwidth and lower latency (*Jun et al., 2017*; *Patterson, 2004*; *Hsieh et al., 2016*). In other word, it requires more contributions from interconnect design so that it is an opportunity for design engineers. For example, as solid-state drive (SSD) capacity has dramatically increased with the V-NAND, the legacy serial-ATA (SATA) interface is not fast enough to provide enough bandwidth. As a result, recent SSD products use NVM Express (NVMe) protocol which is based on peripheral component interconnect express (PCIe) interface. In fact, the PCIe is one of the standards that is evolving very quickly; the industry was working on 16-Gb/s PCIe gen4 in 2016, but started working on 32-Gb/s gen5 since 2018, and 64-Gb/s gen6 specification is going to be released soon (*Vučinić et al., 2014*; *Ajanovic, 2009*; *Budruk, 2007*; *Cheng et al., 2010*; *Li et al., 2018*).

Since multiple dies are stacked in the HBM, there are more interconnects that are required, and there are unique challenges which can be distinguished from a conventional interconnects, which means there are plenty of works that the interconnect design has to do. For example, the stacked DRAM is communicating with processing unit through the silicon interposer channel, which is quite different to the conventional channels such as channel response and crosstalk (*Ko et al., 2020*; *Liu, Ding & Jiang, 2018*). In addition, the stacked DRAM dies are connected by TSV links whose characteristic is also very different (*Lee et al., 2015*, *2016*; *Kim et al., 2012*). And there is also a logic die where a HBM PHY is used to bridge the DRAM stack and the host SoC. There are also unique issues, for example thermal stability issue due to the stacking (*Sohn et al., 2016*; *Ko et al., 2019*), which should be overcome by hardware design.

Introducing new memory devices

In addition to those efforts discuss above, the memory industry is trying to introduce new non-volatile memory (NVM) devices, for example phase-change RAM (PRAM) or resistive RAM (RRAM, also referred to as memristor), whose conceptual diagram is shown in Fig. 9. These devices have only two ports so that it has a smaller footprint of $4F^2$, and they are able to be integrated in crossbar array and easy to stack (*Wong et al., 2010*, *2012*; *Bae et al., 2017*; *Yoon, Kim & Hwang, 2019*; *Foong & Hady, 2016*; *Kau et al., 2009*; *Yoon et al., 2017*; *Liu et al., 2013*). In addition, they can be formed in back-end process

so that they can be integrated on top of the CMOS peripheral circuits, which makes their effective density even higher and realizes a true sense of 3D integration. Moreover, the devices themselves are much faster than NAND device. Note that a faster device means that we need a faster interconnect not to degrade the memory performance due to the interconnect. That is, there will be more demand on high performance interconnect design, similar to what happens on the HBM and V-NAND cases.

These devices have many attractive features, however, there are plenty of challenges that need to be overcome to make them succeed in the industry. For example, their operation and side effects are not yet fully modeled; and the PRAM has a reliability issue which is called snapback current during write operation; and the RRAM has a sneak current issue which distorts readout operation as well as write (*Yoon et al., 2016*); and the variation effect is much larger than the legacy devices because of their intrinsic non-linearity. In fact, these kinds of challenges fall into categories where design engineers can do better than device engineers. For example, they can build a good physics-aware model to bring these devices into an accurate and complex hardware simulation, to enable collaborative optimization between circuits and devices. Because of their non-linearity and hysteresis, some special techniques need to be developed to ensure that they converge in a huge array-level simulation, while capturing the realistic behavior (*Bae & Yoon, 2020*; *Wang, 2017*; *Kvatinsky et al., 2012*; *Linn et al., 2014*; *Chen & Yu, 2015*). On the other hand, some circuit design techniques can be introduced to mitigate the snapback current (*Kim & Ahn, 2005*; *Redaelli et al., 2004*; *Parkinson, 2011*). Also, circuit designers can propose variation-tolerant or variation-compensated techniques to address the variation issue (*Athmanathan et al., 2016*; *Park et al., 2017*; *Hwang et al., 2010*; *Bae et al., 2018*), or sneak-current cancellation scheme for the sneak current issue (*Vontobel et al., 2009*; *Shevgoor et al., 2015*; *Bae et al., 2016*). In addition, looking further forward, RRAM is regarded to be a promising candidate for in-memory computing or neuromorphic computing, because of its capability to store analog weights (*Alibart, Zamanidoost & Strukov, 2013*; *Prezioso et al., 2015*; *Yoo, 2019*; *Xue et al., 2019*; *Kim & Williams, 2019*; *Yoon, Han & Bae, 2020*; *Wang et al., 2019*). These approaches are believed to overcome the limitation of the current computer architecture, where we need tons of inter-disciplinary research opportunity to realize them.

To summarize this section, the introduction of the 3D integration and the new memory devices is believed to overcome the scaling limit of memory devices, and it needs a lot of supports from hardware designers and gives many opportunities to contribute.

INTERCONNECT

Trend survey and challenges

In this section, the challenges and potential solutions of computer communication interconnect are presented. Recalling the "Introduction", increase in data and advancement in computing require higher speed interconnect, however the electrical channel becomes more and more inefficient as the data rate increases. Figure 10 shows an architectural diagram of a general interconnect, which serializes parallel input to high-speed non-return-to-zero (NRZ) bitstream and transmits it through electrical

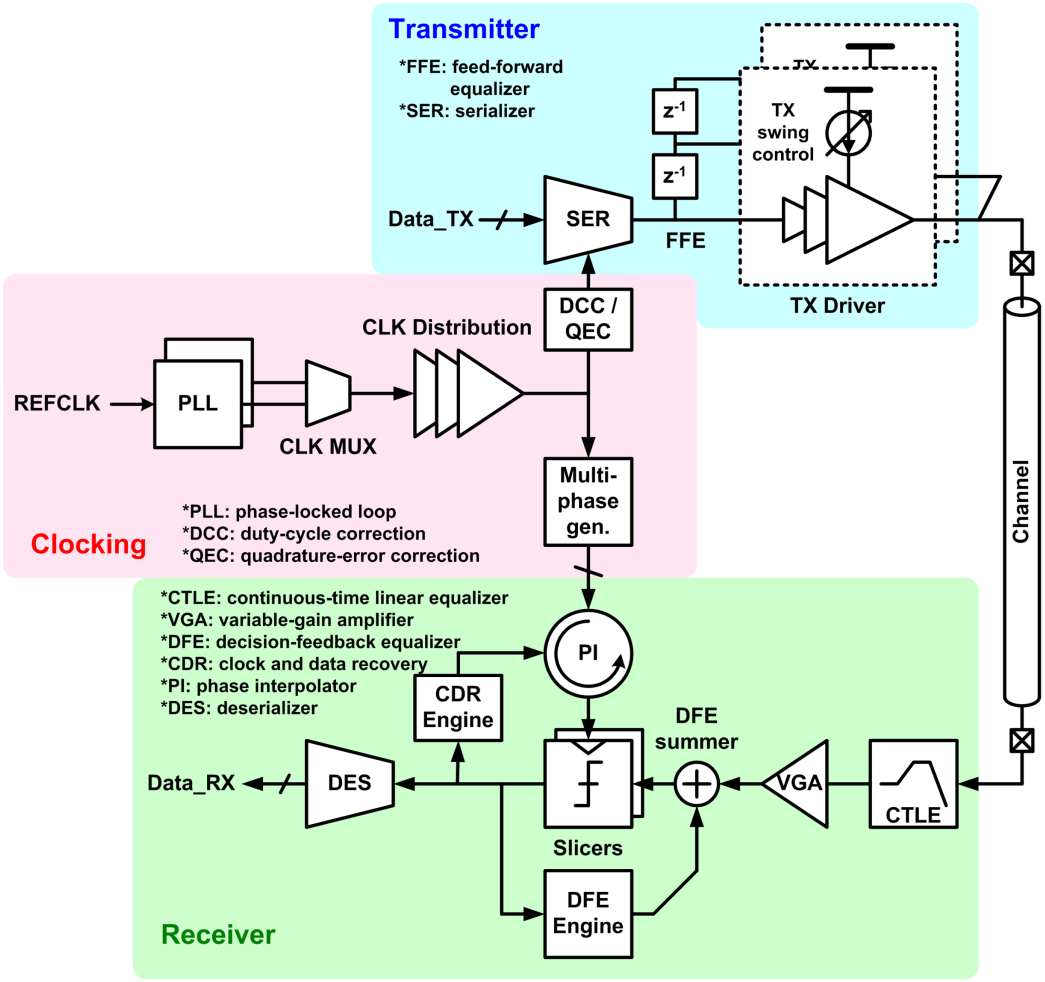

Figure 10 **Block diagram of general interconnect architecture.**
Full-size DOI: 10.7717/peerj-cs.420/fig-10

channel (wire), and then de-serializes the serial input to parallel at the receive side (*Bae, 2020*; *Chang et al., 2003*; *Mooney et al., 2006*; *Bulzacchelli et al., 2006*). It is notable that this architecture has not been changed over last 15 years. Since then, the advancements mainly focus on improving building blocks of the given golden architecture, such as designing a better equalizer to provide a better compensation for the channel loss.

Let us have a deeper look at what causes the challenges on the computer interconnect. As mentioned earlier, the electrical channels do not scale with the silicon technology. However, the interconnect partially takes advantage of the technology scaling, because faster transistors enable a better circuit to overcome the increased channel loss. Figure 11A shows a survey from the state-of-the-art published works (*Tamura et al., 2001*; *Haycock & Mooney, 2001*; *Tanaka et al., 2002*; *Lee et al., 2003*, *2004*; *Krishna et al., 2005*; *Landman et al., 2005*; *Casper et al., 2006*; *Palermo, Emami-Neyestanak & Horowitz, 2008*; *Kim et al., 2008*; *Lee, Chen & Wang, 2008*; *Amamiya et al., 2009*; *Chen et al., 2011*; *Takemoto et al., 2012*; *Raghavan et al., 2013*; *Navid et al., 2014*; *Zhang et al., 2015*;

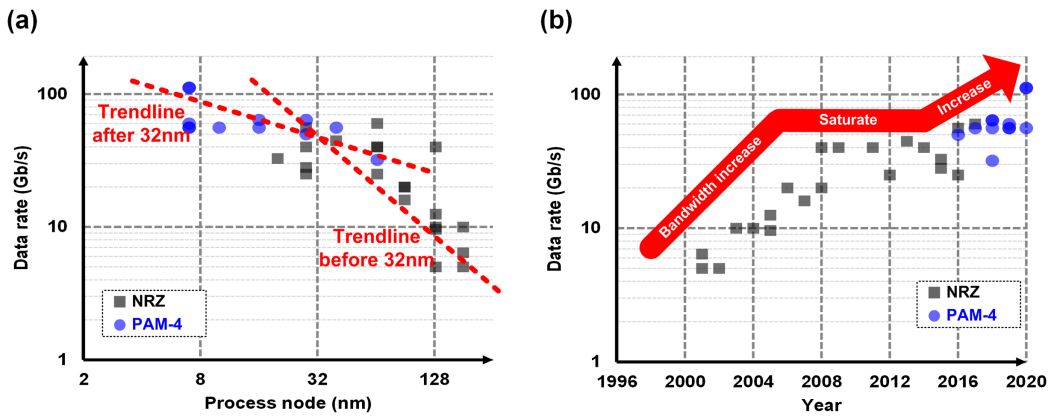

Figure 11 **Survey and trend of interconnects with respect to (A) technology nodes (B) published years.** Full-size DOI: 10.7717/peerj-cs.420/fig-11

Upadhyaya et al., 2015; *Norimatsu et al., 2016*; *Gopalakrishnan et al., 2016*; *Shibasaki et al., 2016*; *Peng et al., 2017*; *Han et al., 2017*; *Upadhyaya et al., 2018*; *Wang et al., 2018*; *Depaoli et al., 2018*; *Tang et al., 2018*; *LaCroix et al., 2019*; *Pisati et al., 2019*; *Ali et al., 2019*, *2020*; *Im et al., 2020*; *Yoo et al., 2020*), where we can confirm the correlation between the technology node and the data rate. On the other hand, however, overcoming the increased channel loss has become more and more expensive as the loss is going worse as the bandwidth increases; the equalization circuits consume too much power to compensate the loss, which makes people hesitant to increase the bandwidth. As a result, the tendency has been weakened after 32-nm node. Figure 11B shows the bandwidth trend over years, which evidently shows the bandwidth increase has saturated at around 28–40 Gb/s for years.

Recently, a dramatic change has been made to break the ice. An amplitude modulation technique, which is called 4-level pulse-amplitude modulation (PAM-4), has been adopted in the industry (*Upadhyaya et al., 2018*; *Wang et al., 2018*; *Depaoli et al., 2018*; *Tang et al., 2018*; *LaCroix et al., 2019*; *Pisati et al., 2019*; *Ali et al., 2019*, *2020*; *Im et al., 2020*; *Yoo et al., 2020*). With PAM-4, the interconnect can transmit two bits in one-bit period, which doubles the effective bandwidth over NRZ. This dramatic change enables the interconnect bandwidth higher than 50 Gb/s as observed in Fig. 11, and most of latest specifications whose speed are higher than 50 Gb/s employ the PAM-4. In addition, all the golden architecture except for very front-end circuits do not have to be changed with PAM-4, which makes it more attractive.

However, we have to ask if this approach is sustainable or not. We doubled the data rate by adopting PAM-4, then can we do the same with PAM-8 or PAM-16? Fig. 12 shows the comparison between those modulations. The basic concept of PAM-4 is to transmit two bits at the same time, so it achieves 2x higher data rate at the same Nyquist frequency. However, there are 4 signal levels (3 stacked eyes) instead of 2 levels (1 eye), the signal-to-noise ratio (SNR) degrades by 3x, or 9.5 dB. It also introduces some other non-idealities such as non-linearity and CDR complexity, so it can be worse. These days, PAM-4 is justified because the benefit from the higher bandwidth exceeds the SNR loss.

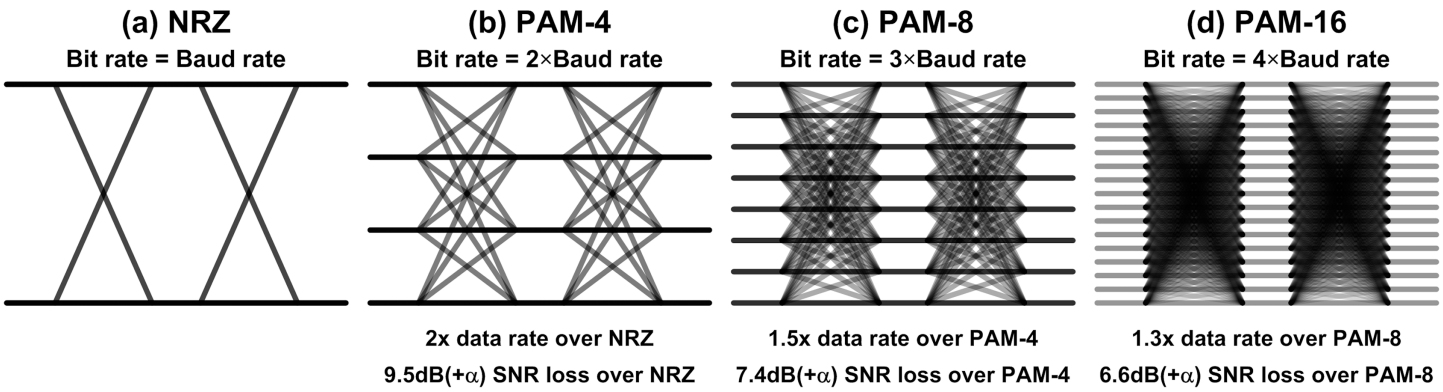

Figure 12 **Comparison of (A) NRZ, (B) PAM-4, (C) PAM-8, and (D) PAM-16.** Full-size DOI: 10.7717/peerj-cs.420/fig-12

We can do the same calculation for PAM-8. It transmits 3 bits while PAM-4 transmits 2 bits, so we get 1.5x higher bandwidth, whereas there are 7 eyes over PAM-4's 3 eyes, which is equivalent to 7.4-dB SNR degradation. That is, the benefit of PAM-8 is lower than what we can get from PAM-4. The same calculation for PAM-16 is also given in the Fig. 12, where we can find the benefit gets even smaller than PAM-8. From the observation, we can conclude that the amplitude modulation will not be a sustainable solution while the channel capacity and the noise keep the same (*Shannon, 1948*).

Future directions

As an alternative, we would rather start modifying the golden architecture. One of the potential candidates is a forwarded-clock architecture, which has been explored in several literatures (*Casper et al., 2006*; *Li et al., 2014*; *Ragab et al., 2011*; *Casper & O'Mahony, 2009*; *Hossain & Carusone, 2011*; *Chung & Kim, 2012*; *Bae et al., 2016*). The bit-error-rate (BER) of an interconnect is a function of the amplitude noise (SNR) and the timing noise (jitter) (*Bae et al., 2016*). If the SNR becomes worse as the channel loss increases or PAM is used, we can try cancel it out by improving the timing noise. However, in the conventional architecture, way of reducing the timing noise is very limited other than burning more power. Instead, we can forward the transmitter clock to the receiver along with data. Because the timing noise of the forwarded clock and the data are correlated, sampling the data with the forwarded clock cancels the correlated component out hence the effective timing noise at the receive side is minimized. With that, the signaling power and the CDR complexity can be significantly reduced at the same BER, at the cost of just one extra clock channel.

On the other hand, we can also make a bigger change on the architecture. In an analog-to-digital converter (ADC)-based interconnect or digital signal processing (DSP)-based interconnect (*LaCroix et al., 2019*; *Pisati et al., 2019*; *Ali et al., 2019*; *2020*; *Im et al., 2020*; *Yoo et al., 2020*; *Harwood et al., 2007*; *Chen & Yang, 2010*; *Wang et al., 2018*; *Palermo et al., 2018*), the analog front-end circuits of the receiver are replaced by a high-speed ADC, and a large fraction of the equalization and CDR stuffs are done in the digital domain. With that, an extensive equalization with dense digital logic is enabled.

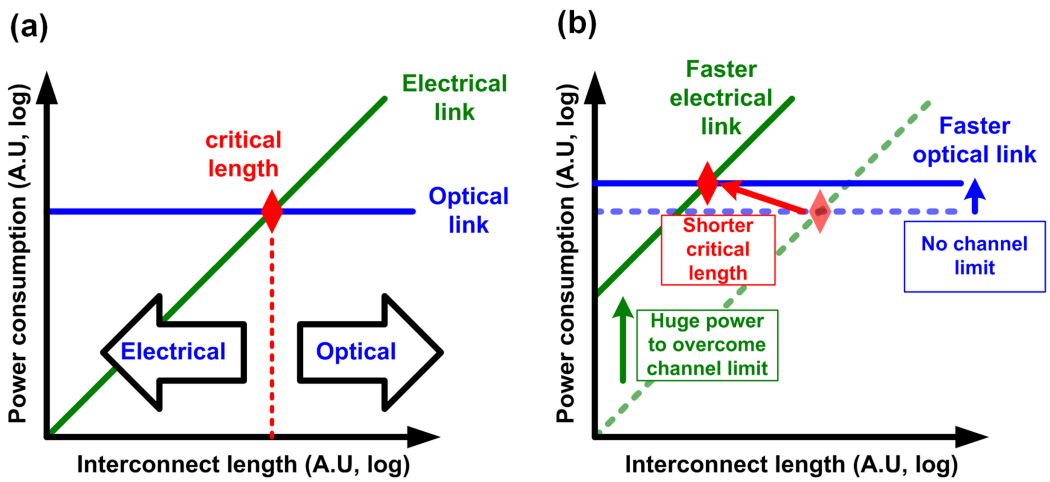

Figure 13 **(A) Power comparison between electrical and optical interconnects and definition of critical length. (B) Reduction of critical length at higher speed.**

Full-size DOI: 10.7717/peerj-cs.420/fig-13

In addition, PAM-4 justifies the use of ADC because it already requires simple ADC-like front-end as it transmits and receives multiple data levels. The DSP-based interconnect is maturing rapidly these days, however there are still lots of works to come, for example design techniques for building high-speed ADC or resolving high latency of DSP-based receiver.

For a long-term solution, more dramatic change would be required, because the fundamental limit comes from the limited bandwidth of electrical channels. As a result, replacing the electrical channel with optical channel whose bandwidth is almost infinite is believed to be a very promising and eventual solution (*Miller, 2000*; *Young et al., 2009*; *Jeong, Bae & Jeong, 2017*; *Thraskias et al., 2018*). Conventionally, the optical interconnect has been used for long-distance telecommunication whereas the electrical interconnect is responsible for short-distance computer communication. It is mainly because the optical interconnect consumes much higher power because of power-hungry optical devices and electrical-optical interfaces. On the other hand, because of its lossless nature, the communication distance has little impact on the optical communication performance. However, the electrical interconnect exhibits lower power consumption at short-reach communication, however its power consumption dramatically increases as the communication distance increases because the electrical channel loss increases exponentially with the distance. As a result, there is a critical length where the optical interconnect becomes more efficient than the electrical interconnect, as shown in Fig. 13A (*Cho, Kapur & Saraswat, 2004*). In similar manner, when the required data rate increases, the power consumption of the electrical interconnects increases exponentially even at the same distance, however it has little impact on the optical interconnect as shown in Fig. 13B. Therefore, the critical length is expected to become shorter as the data rate keep increasing, which make us believe the optical interconnect will be eventually used for computer communication (*Cho, Kapur & Saraswat, 2004*). However, to realize it, the energy efficiency of optical interconnects must be improved a lot. Currently, the

bandwidth-efficiency product of commercial optical interconnects (long-reach) is almost 1,000x lower than that of electrical interconnects (*Sun et al., 2020*). Then why does a present optical interconnect consume that much power? There can be many reasons, but one of the main reasons is that it is not monolithically integrated. When we look into an optical communication module, there are multiple ICs such as photonics transmitter, receiver, electronic driver IC, retimer IC, and microcontroller. As a result, there are so many interfaces even in a single communication module, where electrical signals come out to the real analog world and experience bulky parasitics, which leads to such poor energy efficiency. That is, monolithic integration where optical devices and VLSI circuits are integrated in a single chip can be a solution for reducing the power consumption (*Sun et al., 2015a*, *2015b*, *2020*; *Narasimha et al., 2007*). In addition to the monolithic integration, dense wavelength division multiplexing (DWDM) enables transmitting multiple data streams through a single optical fiber, which significantly improves the bandwidth density of optical interconnect. DWDM can be regarded as another modulation, but it does not degrade SNR as much as PAM. In fact, it has been more than 30 years ago that the optical interconnect began to gain attention, but there have been no succeed until recently. However, recently, the accumulated efforts are coming out with promising engineering samples such as 5-pJ/bit monolithic DWDM (*Sun et al., 2020*), 6-pJ/bit 112-Gb/s PAM-4 (*Li et al., 2020*), so the time will really come soon.

CONCLUSIONS

In this paper, the challenges that the current computing system (logic, memory, interconnect) is facing are reviewed. For the logic, the cryptocurrency miners are surveyed which leads to the future direction of specialization, but the downside of specialization is also discussed with an example of a fabless company. For the memory, the challenges and opportunities for design engineering in conjunction with device engineering are reviewed, whereas other reviews tend to focus on devices. For the interconnect, the state-of-the-art works are surveyed, and the recent trends and challenges are discussed. From the reviews and surveys for each part, the solutions and opportunities for those challenges are discussed, which are summarized in Fig. S2. For the logic side, the specialization is proposed for achieving higher efficiency after Moore's law, and the reusing is also proposed for addressing the productivity issue of the specialization. On the memory side, 3D integration of memory dies or cells and introduction of new NVM devices are expected to overcome the memory density issue. At the same time, they request substantial assistances from design engineers, for example high-performance interconnects, robust physics-aware device modeling, and tons of design techniques to overcome the device limits. Finally, the interconnect side needs to innovate its conventional architecture which has not been changed for a while, and eventually it must drive the optical interconnect.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The authors received no funding for this work.

Competing Interests

Woorham Bae is employed by Ayar Labs. Woorham Bae is also an Academic Editor for PeerJ.

Author Contributions

- Woorham Bae conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

This work does not include any code as it is a literature review.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj-cs.420#supplemental-information.

REFERENCES

Ajanovic J. 2009. PCI express 3.0 overview. In: *Proceedings of Hot Chip: A Symposium on High Performance Chips.* Vol. 69. 143.

Ali T, Chen E, Park H, Yousry R, Ying YM, Abdullatif M, Gandara M, Liu CC, Weng PS, Chen HS, Elbadry M. 2020. 6.2 A 460mW 112Gb/s DSP-based transceiver with 38dB loss compensation for next-generation data centers in 7nm FinFET Technology. In: *2020 IEEE International Solid-State Circuits Conference—(ISSCC).* Piscataway: IEEE, 118–120.

Ali T, Yousry R, Park H, Chen E, Weng PS, Huang YC, Liu CC, Wu CH, Huang SH, Lin C, Wu KC. 2019. 6.4 A 180mW 56Gb/s DSP-based transceiver for high density IOs in data center switches in 7nm FinFET Technology. In: *2019 IEEE International Solid-State Circuits Conference—(ISSCC).* Piscataway: IEEE, 118–120.

Alibart F, Zamanidoost E, Strukov DB. 2013. Pattern classification by memristive crossbar circuits using ex situ and in situ training. *Nature Communications* **4(1)**:1–7 DOI 10.1038/ncomms3072.

Amamiya Y, Kaeriyama S, Noguchi H, Yamazaki Z, Yamase T, Hosoya K, Okamoto M, Tomari S, Yamaguchi H, Shoda H, Ikeda H. 2009. A 40Gb/s multi-data-rate CMOS transceiver chipset with SFI-5 interface for optical transmission systems. In: *2009 IEEE International Solid-State Circuits Conference—Digest of Technical Papers.* Piscataway: IEEE, 358–359.

Athmanathan A, Stanisavljevic M, Papandreou N, Pozidis H, Eleftheriou E. 2016. Multilevel-cell phase-change memory: a viable technology. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* **6(1)**:87–100 DOI 10.1109/JETCAS.2016.2528598.

Ayar Custom Generator. 2021. AyarLabs/ACG. *Available at https://github.com/AyarLabs/ACG* (accessed 19 March 2021).

Bae W. 2020. Supply-scalable high-speed I/O interfaces. *Electronics* **9(8)**:1315 DOI 10.3390/electronics9081315.

Bae W, Jeong GS, Jeong DK. 2016. A 1-pJ/bit, 10-Gb/s/ch forwarded-clock transmitter using a resistive feedback inverter-based driver in 65-nm CMOS. *IEEE Transactions on Circuits and Systems II: Express Briefs* **63(12)**:1106–1110 DOI 10.1109/TCSII.2016.2618896.

Bae W, Jeong GS, Park K, Cho SY, Kim Y, Jeong DK. 2016. A 0.36 pJ/bit, 0.025 mm^2, 12.5 Gb/s forwarded-clock receiver with a stuck-free delay-locked loop and a half-bit delay line in 65-nm CMOS Technology. *IEEE Transactions on Circuits and Systems I: Regular Papers* **63(9)**:1393–1403 DOI 10.1109/TCSI.2016.2578960.

Bae W, Ju H, Park K, Cho SY, Jeong DK. 2016. A 7.6 mW, 414 fs RMS-jitter 10 GHz phase-locked loop for a 40 Gb/s serial link transmitter based on a two-stage ring oscillator in 65 nm CMOS. *IEEE Journal of Solid-State Circuits* **51(10)**:2357–2367 DOI 10.1109/JSSC.2016.2579159.

Bae W, Ju H, Park K, Han J, Jeong DK. 2017. A supply-scalable-serializing transmitter with controllable output swing and equalization for next-generation standards. *IEEE Transactions on Industrial Electronics* **65(7)**:5979–5989.

Bae W, Yoon KJ. 2020. Comprehensive read margin and BER analysis of one selector-one memristor crossbar array considering thermal noise of memristor with noise-aware device model. *IEEE Transactions on Nanotechnology* **19**:553–564 DOI 10.1109/TNANO.2020.3006114.

Bae W, Yoon KJ, Hwang CS, Jeong DK. 2016. A crossbar resistance switching memory readout scheme with sneak current cancellation based on a two-port current-mode sensing. *Nanotechnology* **27(48)**:485201 DOI 10.1088/0957-4484/27/48/485201.

Bae W, Yoon KJ, Hwang CS, Jeong DK. 2017. Extension of two-port sneak current cancellation scheme to 3-D vertical RRAM crossbar array. *IEEE Transactions on Electron Devices* **64(4)**:1591–1596 DOI 10.1109/TED.2017.2664863.

Bae W, Yoon KJ, Song T, Nikolić B. 2018. A variation-tolerant, sneak-current-compensated readout scheme for cross-point memory based on two-po0rt sensing technique. *IEEE Transactions on Circuits and Systems II: Express Briefs* **65(12)**:1839–1843 DOI 10.1109/TCSII.2018.2868460.

Berkeley Analog Generator. 2021a. Layout With Gridded Objects (Laygo). *Available at https://github.com/ucb-art/laygo* (accessed 19 March 2021).

Berkeley Analog Generator. 2021b. Main Framework: ucb-art/BAG_framework. *Available at https://github.com/ucb-art/BAG_framework* (accessed 19 March 2021).

Bohr MT, Young IA. 2017. CMOS scaling trends and beyond. *IEEE Micro* **37(6)**:20–29 DOI 10.1109/MM.2017.4241347.

Budruk R. 2007. PCI express basics. In: *PCI-SIG Developers Conference*.

Bulzacchelli JF, Meghelli M, Rylov SV, Rhee W, Rylyakov AV, Ainspan HA, Parker BD, Beakes MP, Chung A, Beukema TJ, Pepeljugoski PK. 2006. A 10-Gb/s 5-tap DFE/4-tap FFE transceiver in 90-nm CMOS technology. *IEEE Journal of Solid-State Circuits* **41(12)**:2885–2900 DOI 10.1109/JSSC.2006.884342.

Casper B, Jaussi J, O'Mahony F, Mansuri M, Canagasaby K, Kennedy J, Yeung E, Mooney R. 2006. A 20Gb/s forwarded clock transceiver in 90nm CMOS B. In: *2006 IEEE International Solid State Circuits Conference—Digest of Technical Papers.* Piscataway: IEEE, 263–272.

Casper B, O'Mahony F. 2009. Clocking analysis, implementation and measurement techniques for high-speed data links—a tutorial. *IEEE Transactions on Circuits and Systems I: Regular Papers* **56(1)**:17–39 DOI 10.1109/TCSI.2008.931647.

Castro-Lopez R, Guerra O, Roca E, Fernández FV. 2008. An integrated layout-synthesis approach for analog ICs. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* **27(7)**:1179–1189 DOI 10.1109/TCAD.2008.923417.

Chang E, Han J, Bae W, Wang Z, Narevsky N, NikoliC B, Alon E. 2018. BAG2: a process-portable framework for generator-based AMS circuit design. In: *2018 IEEE Custom Integrated Circuits Conference (CICC).* Piscataway: IEEE, 1–8.

Chang KY, Wei J, Huang C, Li S, Donnelly K, Horowitz M, Li Y, Sidiropoulos S. 2003. A 0.4-4-Gb/s CMOS quad transceiver cell using on-chip regulated dual-loop PLLs. *IEEE Journal of Solid-State Circuits* **38(5)**:747–754 DOI 10.1109/JSSC.2003.810045.

Chen MS, Shih YN, Lin CL, Hung HW, Lee J. 2011. A 40Gb/s TX and RX chip set in 65nm CMOS. In: *IEEE Journal of Solid-State Circuits.* Piscataway: IEEE, 146–148.

Chen EH, Yang CKK. 2010. ADC-based serial I/O receivers. *IEEE Transactions on Circuits and Systems I: Regular Papers* **57(9)**:2248–2258 DOI 10.1109/TCSI.2010.2071431.

Chen PY, Yu S. 2015. Compact modeling of RRAM devices and its applications in 1T1R and 1S1R array design. *IEEE Transactions on Electron Devices* **62(12)**:4022–4028 DOI 10.1109/TED.2015.2492421.

Cheng KH, Tsai YC, Wu YH, Lin YF. 2010. A 5-Gb/s inductorless CMOS adaptive equalizer for PCI express generation II applications. *IEEE Transactions on Circuits and Systems II: Express Briefs* **57(5)**:324–328 DOI 10.1109/TCSII.2010.2047311.

Cho H, Kapur P, Saraswat KC. 2004. Power comparison between high-speed electrical and optical interconnects for interchip communication. *Journal of Lightwave Technology* **22(9)**:2021–2033 DOI 10.1109/JLT.2004.833531.

Choi JS, Hwang MS, Jeong DK. 2004. A 0.18-/spl mu/m CMOS 3.5-gb/s continuous-time adaptive cable equalizer using enhanced low-frequency gain control method. *IEEE Journal of Solid-State Circuits* **39(3)**:419–425 DOI 10.1109/JSSC.2003.822774.

Chung SH, Kim LS. 2012. 1.22 mW/Gb/s 9.6 Gb/s data jitter mixing forwarded-clock receiver robust against power noise with 1.92 ns latency mismatch between data and clock in 65nm CMOS. In: *2012 Symposium on VLSI Circuits (VLSIC).* Piscataway: IEEE, 144–145.

Cisco Visual Networking Index. 2017. *Forecast and Trends*. San Jose: Cisco Systems Inc, 2017–2022.

Cisco Visual Networking Index. 2019. *Forecast and Trends, 2017–2022 White Paper*. San Jose: Cisco Systems Inc.

Crossley J, Puggelli A, Le HP, Yang B, Nancollas R, Jung K, Kong L, Narevsky N, Lu Y, Sutardja N, An EJ. 2013. BAG: a designer-oriented integrated framework for the development of AMS circuit generators. In: *2013 IEEE/ACM International Conference on Computer-Aided Design (ICCAD).* Piscataway: IEEE, 74–81.

Dally WJ, Poulton J. 1997. Transmitter equalization for 4-Gbps signaling. *IEEE Micro* **17(1)**:48–56 DOI 10.1109/40.566199.

Daly DC, Fujino LC, Smith KC. 2018. Through the looking glass-the 2018 edition: trends in solid-state circuits from the 65th ISSCC. *IEEE Solid-State Circuits Magazine* **10(1)**:30–46 DOI 10.1109/MSSC.2017.2771103.

Danowitz A, Kelley K, Mao J, Stevenson JP, Horowitz M. 2012. CPU DB: recording microprocessor history. *Queue* **10(4)**:10–27 DOI 10.1145/2181796.2181798.

Depaoli E, Monaco E, Steffan G, Mazzini M, Zhang H, Audoglio W, Belotti O, Rossi AA, Albasini G, Pozzoni M, Erba S. 2018. A 4.9pJ/b 16-to-64Gb/s PAM-4 VSR transceiver in 28nm FDSOI CMOS. In: *2018 IEEE International Solid—State Circuits Conference—(ISSCC).* Piscataway: IEEE, 112–114.

Esmaeilzadeh H, Blem E, Amant RS, Sankaralingam K, Burger D. 2011. Dark silicon and the end of multicore scaling. In: *2011 38th Annual international symposium on computer architecture (ISCA).* Piscataway: IEEE, 365–376.

Foong A, Hady F. 2016. Storage as fast as rest of the system. In: *2016 IEEE 8th International Memory Workshop (IMW).* Piscataway: IEEE, 1–4.

Gopalakrishnan K, Ren A, Tan A, Farhood A, Tiruvur A, Helal B, Loi CF, Jiang C, Cirit H, Quek I, Riani J. 2016. 3.4 A 40/50/100Gb/s PAM-4 Ethernet transceiver in 28nm CMOS. In: *2016 IEEE International Solid-State Circuits Conference (ISSCC).* Piscataway: IEEE, 62–63.

Habal H, Graeb H. 2011. Constraint-based layout-driven sizing of analog circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* **30(8)**:1089–1102 DOI 10.1109/TCAD.2011.2158732.

Hakhamaneshi K, Werblun N, Abbeel P, Stojanović V. 2019. BagNet: Berkeley analog generator with layout optimizer boosted with deep neural networks. In: *2019 IEEE/ACM International Conference on Computer-Aided Design (ICCAD).* Piscataway: IEEE, 1–8.

Han J, Bae W, Chang E, Wang Z, Nikolic B, Alon E. 2021. LAYGO: a layout generation framework to enhance design productivity in advanced CMOS technologies. *Transactions on Circuits and Systems I: Regular Papers* **68(3)**:1012–1022 DOI 10.1109/TCSI.2020.3046524.

Han J, Lu Y, Sutardja N, Alon E. 2017. 6.2 A 60Gb/s 288mW NRZ transceiver with adaptive equalization and baud-rate clock and data recovery in 65nm CMOS technology. In: *2017 IEEE International Solid-State Circuits Conference (ISSCC).* Piscataway: IEEE, 112–113.

Harwood M, Warke N, Simpson R, Leslie T, Amerasekera A, Batty S, Colman D, Carr E, Gopinathan V, Hubbins S, Hunt P. 2007. A 12.5 Gb/s SerDes in 65nm CMOS using a baud-rate ADC with digital receiver equalization and clock recovery. In: *2007 IEEE International Solid-State Circuits Conference. Digest of Technical Papers.* Piscataway: IEEE, 436–591.

Haycock M, Mooney R. 2001. 3.2 GHz 6.4 Gb/s per wire signaling in 0.18 /spl mu/m CMOS. In: *IEEE International Solid-State Circuits Conference. Digest of Technical Papers. ISSCC.* Piscataway: IEEE, 62–63.

Holt WM. 2016. 1.1 Moore's law: a path going forward. In: *2016 IEEE International Solid-State Circuits Conference (ISSCC).* Piscataway: IEEE, 8–13.

Horowitz M. 2014. 1.1 computing's energy problem (and what we can do about it). In: *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers.* IEEE, 10–14.

Horowitz M, Yang CKK, Sidiropoulos S. 1998. High-speed electrical signaling: overview and limitations. *IEEE Micro* **18(1)**:12–24 DOI 10.1109/40.653013.

Hossain M, Carusone AC. 2011. 7.4 Gb/s 6.8 mW source synchronous receiver in 65 nm CMOS. *IEEE Journal of Solid-State Circuits* **46(6)**:1337–1348 DOI 10.1109/JSSC.2011.2131730.

Hsieh K, Ebrahimi E, Kim G, Chatterjee N, O'Connor M, Vijaykumar N, Mutlu O, Keckler SW. 2016. Transparent offloading and mapping (TOM) enabling programmer-transparent near-data processing in GPU systems. *ACM SIGARCH Computer Architecture News* **44(3)**:204–216 DOI 10.1145/3007787.3001159.

Hwang CG. 2002. Semiconductor memories for IT era. In: *2002 IEEE International Solid-State Circuits Conference. Digest of Technical Papers.* Vol. 1. Piscataway: IEEE, 24–27.

Hwang YN, Um CY, Lee JH, Wei CG, Oh HR, Jeong GT, Jeong HS, Kim CH, Chung CH. 2010. MLC PRAM with SLC write-speed and robust read scheme. In: *2010 Symposium on VLSI Technology.* Piscataway: IEEE, 201–202.

Im JW, Jeong WP, Kim DH, Nam SW, Shim DK, Choi MH, Yoon HJ, Kim DH, Kim YS, Park HW, Kwak DH. 2015. 7.2 A 128Gb 3b/cell V-NAND flash memory with 1Gb/s I/O rate. In: *2015 IEEE International Solid-State Circuits Conference—(ISSCC) Digest of Technical Papers.* Piscataway: IEEE, 1–3.

Im J, Zheng K, Chou A, Zhou L, Kim JW, Chen S, Wang Y, Hung HW, Tan K, Lin W, Roldan A. 2020. 6.1 A 112Gb/s PAM-4 long-reach wireline transceiver using a 36-way time-interleaved SAR-ADC and inverter-based RX analog front-end in 7nm FinFET. In: *2020 IEEE International Solid-State Circuits Conference—(ISSCC).* Piscataway: IEEE, 116–118.

Jang SH, Lim J, Han J, Jang J, Yeo J, Lee C, Baek S, Lee J, Lee JH, Yamada S, Lee K. 2019. A fully integrated low voltage DRAM with thermally stable gate-first high-k metal gate process. In: *2019 IEEE International Electron Devices Meeting (IEDM).* Piscataway: IEEE, 28.4.1–28.4.3.

Jeong GS, Bae W, Jeong DK. 2017. Review of CMOS integrated circuit technologies for high-speed photo-detection. *Sensors* **17(9)**:1962 DOI 10.3390/s17091962.

Jun H, Cho J, Lee K, Son HY, Kim K, Jin H, Kim K. 2017. Hbm (high bandwidth memory) dram technology and architecture. In: *2017 IEEE International Memory Workshop (IMW).* IEEE, 1–4.

Kau D, Tang S, Karpov IV, Dodge R, Klehn B, Kalb JA, Strand J, Diaz A, Leung N, Wu J, Lee S. 2009. A stackable cross point phase change memory. In: *2009 IEEE International Electron Devices Meeting (IEDM).* Piscataway: IEEE, 1–4.

Kim K. 2015. 1.1 silicon technologies and solutions for the data-driven world. In: *Digest of Technical Papers—IEEE International Solid-State Circuits Conference.* Piscataway: IEEE, 1–7.

Kim K, Ahn SJ. 2005. Reliability investigations for manufacturable high density PRAM. In: *2005. Proceedings. 43rd Annual 2005 IEEE International Reliability Physics Symposium.* Piscataway: IEEE, 157–162.

Kim H, Cho J, Kim M, Kim K, Lee J, Lee H, Park K, Choi K, Bae HC, Kim J, Kim J. 2012. Measurement and analysis of a high-speed TSV channel. *IEEE Transactions on Components, Packaging and Manufacturing Technology* **2(10)**:1672–1685 DOI 10.1109/TCPMT.2012.2207900.

Kim W, Choi S, Sung J, Lee T, Park C, Ko H, Jung J, Yoo I, Park Y. 2009. Multi-layered vertical gate NAND flash overcoming stacking limit for terabit density storage. In: *2009 Symposium on VLSI Technology.* Piscataway: IEEE, 188–189.

Kim J, Horowitz MA. 2002. Adaptive supply serial links with sub-1-V operation and per-pin clock recovery. *IEEE Journal of Solid-State Circuits* **37(11)**:1403–1413 DOI 10.1109/JSSC.2002.803937.

Kim C, Kim DH, Jeong W, Kim HJ, Park IH, Park HW, Lee J, Park J, Ahn YL, Lee JY, Kim SB. 2017. A 512-gb 3-b/cell 64-stacked wl 3-d-nand flash memory. *IEEE Journal of Solid-State Circuits* **53(1)**:124–133 DOI 10.1109/JSSC.2017.2731813.

Kim JK, Kim J, Kim G, Chi H, Jeong DK. 2008. A 40-Gb/s transceiver in 0.13-μm CMOS technology. In: *2008 IEEE Symposium on VLSI Circuits.* Piscataway: IEEE, 196–197.

Kim HJ, Lee YS, Kim JS. 2016. Nvmedirect: a user-space i/o framework for application-specific optimization on nvme ssds. In: *8th {USENIX} Workshop on Hot Topics in Storage and File Systems (HotStorage 16).*

Kim KM, Williams RS. 2019. A family of stateful memristor gates for complete cascading logic. *IEEE Transactions on Circuits and Systems I: Regular Papers* **66(11)**:4348–4355 DOI 10.1109/TCSI.2019.2926811.

Ko HG, Shin S, Kye CH, Lee SY, Yun J, Jung HK, Lee D, Kim S, Jeong DK. 2019. A 370-fJ/b, 0.0056 mm 2/DQ, 4.8-Gb/s DQ receiver for HBM3 with a baud-rate self-tracking loop. In: *2019 Symposium on VLSI Circuits (pp. C94-C94).* Piscataway: IEEE.

Ko HG, Shin S, Oh J, Park K, Jeong DK. 2020. 6.7 An 8Gb/s/μm FFE-combined crosstalk-cancellation scheme for HBM on silicon interposer with 3D-staggered channels. In: *2020 IEEE International Solid-State Circuits Conference—(ISSCC).* Piscataway: IEEE, 128–130.

Krishna K, Yokoyama-Martin DA, Wolfer S, Jones C, Loikkanen M, Parker J, Segelken R, Sonntag JL, Stonick J, Titus S, Weinlader D. 2005. A 0.6 to 9.6 Gb/s binary backplane transceiver core in 0.13/spl mu/m CMOS. In: *2005 IEEE International Digest of Technical Papers. Solid-State Circuits Conference.* Piscataway: IEEE, 64–585.

Kunal K, Madhusudan M, Sharma AK, Xu W, Burns SM, Harjani R, Hu J, Kirkpatrick DA, Sapatnekar SS. 2019. ALIGN: open-source analog layout automation from the ground up. In: *Proceedings of the 56th Annual Design Automation Conference 2019.* 1–4.

Kvatinsky S, Friedman EG, Kolodny A, Weiser UC. 2012. TEAM: threshold adaptive memristor model. *IEEE Transactions on Circuits and Systems I: Regular Papers* **60(1)**:211–221 DOI 10.1109/TCSI.2012.2215714.

LaCroix MA, Wong H, Liu YH, Ho H, Lebedev S, Krotnev P, Nicolescu DA, Petrov D, Carvalho C, Alie S, Chong E. 2019. 6.2 A 60Gb/s PAM-4 ADC-DSP transceiver in 7nm CMOS with SNR-based adaptive power scaling achieving 6.9 pJ/b at 32dB loss. In: *2019 IEEE International Solid-State Circuits Conference—(ISSCC).* bxfd: Piscataway: IEEE, 114–116.

Landman P, Brouse K, Gupta V, Wu S, Payne R, Erdogan U, Gu R, Yee AL, Parthasarathy B, Ramaswamy S, Bhakta B. 2005. A transmit architecture with 4-tap feedforward equalization for 6.25/12.5 Gb/s serial backplane communications. In: *ISSCC. 2005 IEEE International Digest of Technical Papers. Solid-State Circuits Conference.* Piscataway: IEEE, 66–585.

Lee J, Chen MS, Wang HD. 2008. A 20Gb/s duobinary transceiver in 90nm CMOS. In: *2008 IEEE International Solid-State Circuits Conference—Digest of Technical Papers.* Piscataway: IEEE, 102–599.

Lee H, Cho K, Kim H, Choi S, Lim J, Kim J. 2015. Electrical performance of high bandwidth memory (HBM) interposer channel in terabyte/s bandwidth graphics module. In: *2015 International 3D Systems Integration Conference (3DIC) (pp. TS2-2).* Piscataway: IEEE.

Lee MJ, Dally WJ, Chiang P. 2000. Low-power area-efficient high-speed I/O circuit techniques. *IEEE Journal of Solid-State Circuits* **35(11)**:1591–1599 DOI 10.1109/4.881204.

Lee BJ, Hwang MS, Lee SH, Jeong DK. 2003. A 2.5-10-Gb/s CMOS transceiver with alternating edge-sampling phase detection for loop characteristic stabilization. *IEEE Journal of Solid-State Circuits* **38(11)**:1821–1829 DOI 10.1109/JSSC.2003.809519.

Lee HR, Hwang MS, Lee BJ, Kim YD, Oh D, Kim J, Lee SH, Jeong DK, Kim W. 2004. A fully integrated 0.13 /spl mu/m CMOS 10 Gb Ethernet transceiver with XAUI interface. In: *2004 IEEE International Solid-State Circuits Conference.* Piscataway: IEEE, 170–520.

Lee JC, Kim J, Kim KW, Ku YJ, Kim DS, Jeong C, Yun TS, Kim H, Cho HS, Kim YO, Kim JH. 2016. 18.3 A 1.2 V 64Gb 8-channel 256GB/s HBM DRAM with peripheral-base-die architecture and small-swing technique on heavy load interface. In: *2016 IEEE International Solid-State Circuits Conference (ISSCC).* Piscataway: IEEE, 318–319.

Lee DU, Kim KW, Kim KW, Lee KS, Byeon SJ, Kim JH, Cho JH, Lee J, Chun JH. 2014. A 1.2 V 8 Gb 8-channel 128 GB/s high-bandwidth memory (HBM) stacked DRAM with effective I/O test circuits. *IEEE Journal of Solid-State Circuits* **50(1)**:191–203 DOI 10.1109/JSSC.2014.2360379.

Li H, Balamurugan G, Kim T, Sakib MN, Kumar R, Rong H, Jaussi J, Casper B. 2020. A 3-D-integrated silicon photonic microring-based 112-Gb/s PAM-4 transmitter with nonlinear equalization and thermal control. *IEEE Journal of Solid-State Circuits* **56(1)**:19–29 DOI 10.1109/JSSC.4.

Li H, Chen S, Yang L, Bai R, Hu W, Zhong FY, Palermo S, Chiang PY. 2014. A 0.8 V, 560fJ/bit, 14Gb/s injection-locked receiver with input duty-cycle distortion tolerable edge-rotating 5/4X sub-rate CDR in 65nm CMOS. In: *2014 Symposium on VLSI Circuits Digest of Technical Papers.* Piscataway: IEEE, 1–2.

Li S, Spagna F, Chen J, Wang X, Tong L, Gowder S, Jia W, Nicholson R, Iyer S, Song R, Li L. 2018. A power and area efficient 2.5-16 Gbps gen4 PCIe PHY in 10nm FinFET CMOS. In: *2018 IEEE Asian Solid-State Circuits Conference.* Piscataway: IEEE, 5–8.

Lin PH, Chang YW, Lin SC. 2009. Analog placement based on symmetry-island formulation. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* **28(6)**:791–804 DOI 10.1109/TCAD.2009.2017433.

Linn E, Siemon A, Waser R, Menzel S. 2014. Applicability of well-established memristive models for simulations of resistive switching devices. *IEEE Transactions on Circuits and Systems I: Regular Papers* **61(8)**:2402–2410 DOI 10.1109/TCSI.2014.2332261.

Liu H, Ding Q, Jiang J. 2018. 112G PAM4/56G NRZ interconnect design for high channel count packages. In: *2018 IEEE 27th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS).* Piscataway: IEEE, 237–239.

Liu TY, Yan TH, Scheuerlein R, Chen Y, Lee JK, Balakrishnan G, Yee G, Zhang H, Yap A, Ouyang J, Sasaki T. 2013. A 130.7mm2 2-layer 32-Gb ReRAM memory device in 24-nm technology. *IEEE Journal of Solid-State Circuits* **49(1)**:140–153 DOI 10.1109/JSSC.2013.2280296.

Mak PI, Martins RP. 2010. High-/mixed-voltage RF and analog CMOS circuits come of age. *IEEE Circuits and Systems Magazine* **10(4)**:27–39 DOI 10.1109/MCAS.2010.937880.

Martins R, Lourenco N, Horta N. 2013. LAYGEN II—automatic layout generation of analog integrated circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* **32(11)**:1641–1654 DOI 10.1109/TCAD.2013.2269050.

Miller DA. 2000. Rationale and challenges for optical interconnects to electronic chips. *Proceedings of the IEEE* **88(6)**:728–749 DOI 10.1109/5.867687.

Mooney R, Yeung E, Kennedy J, Canagasaby K, Mansuri M, O'Mahony F, Jaussi J, Casper B. 2006. A 20Gb/s embedded clock transceiver in 90nm CMOS. In: *2006 IEEE International Solid State Circuits Conference—Digest of Technical Papers.* Piscataway: IEEE, 1334–1343.

Moore GE. 1965. Cramming more components onto integrated circuits. *Available at https://download.intel.com/sites/channel/museum/Moores_Law/Articles-Press_Releases/Gordon_Moore_1965_Article.pdf*.

Moore GE. 1975. Progress in digital integrated electronics. In: *Electron Devices Meeting.* Vol. 21. 11–13.

Mueller W, Aichmayr G, Bergner W, Erben E, Hecht T, Kapteyn C, Kersch A, Kudelka S, Lau F, Luetzen J, Orth A. 2005. Challenges for the DRAM cell scaling to 40nm. In: *IEEE InternationalElectron Devices Meeting, 2005. IEDM Technical Digest.* Piscataway: IEEE, 4.

Musah T, Jaussi JE, Balamurugan G, Hyvonen S, Hsueh TC, Keskin G, Shekhar S, Kennedy J, Sen S, Inti R, Mansuri M. 2014. A 4-32 Gb/s bidirectional link with 3-tap FFE/6-tap DFE and collaborative CDR in 22 nm CMOS. *IEEE Journal of Solid-State Circuits* **49(12)**:3079–3090 DOI 10.1109/JSSC.2014.2348556.

Narasimha A, Analui B, Liang Y, Sleboda TJ, Abdalla S, Balmater E, Gloeckner S, Guckenberger D, Harrison M, Koumans RG, Kucharski D. 2007. A fully integrated 4$ times $10-Gb/s DWDM optoelectronic transceiver implemented in a standard 0.13um CMOS SOI technology. *IEEE Journal of Solid-State Circuits* **42(12)**:2736–2744 DOI 10.1109/JSSC.2007.908713.

Navid R, Chen EH, Hossain M, Leibowitz B, Ren J, Chou CHA, Daly B, Aleksić M, Su B, Li S, Shirasgaonkar M. 2014. A 40 Gb/s serial link transceiver in 28 nm CMOS technology. *IEEE Journal of Solid-State Circuits* **50(4)**:814–827 DOI 10.1109/JSSC.2014.2374176.

Norimatsu T, Kawamoto T, Kogo K, Kohmu N, Yuki F, Nakajima N, Muto T, Nasu J, Komori T, Koba H, Usugi T. 2016. 3.3 A 25Gb/s multistandard serial link transceiver for 50dB-loss copper cable in 28nm CMOS. In: *2016 IEEE International Solid-State Circuits Conference (ISSCC).* Piscataway: IEEE, 60–61.

O'Connor M. 2014. Highlights of the high-bandwidth memory (hbm) standard. In: *Memory Forum Workshop*.

Palermo S, Emami-Neyestanak A, Horowitz M. 2008. A 90 nm CMOS 16 Gb/s transceiver for optical interconnects. *IEEE Journal of Solid-State Circuits* **43(5)**:1235–1246 DOI 10.1109/JSSC.2008.920330.

Palermo S, Hoyos S, Cai S, Kiran S, Zhu Y. 2018. Analog-to-digital converter-based serial links: an overview. *IEEE Solid-State Circuits Magazine* **10(3)**:35–47 DOI 10.1109/MSSC.2018.2844603.

Park HK, Lee KW, Song SH, Lee KG, Shin JH, Gangasani V, Shin YS, Kang DH, Park JH, Song KW, Koh GH. 2017. A novel write method for improving RESET distribution of PRAM. In: *2017 Symposium on VLSI Technology (pp. T96-T97).* Piscataway: IEEE.

Park KT, Nam S, Kim D, Kwak P, Lee D, Choi YH, Choi MH, Kwak DH, Kim DH, Kim MS, Park HW. 2014. Three-dimensional 128 Gb MLC vertical NAND flash memory with 24-WL stacked layers and 50 MB/s high-speed programming. *IEEE Journal of Solid-State Circuits* **50(1)**:204–213 DOI 10.1109/JSSC.2014.2352293.

Parkinson WD, Ovonyx Inc. 2011. U.S. Patent 8,077,498. *Available at https://patents.google.com/patent/US8427862*.

Patterson DA. 2004. Latency lags bandwith. *Communications of the ACM* **47(10)**:71–75 DOI 10.1145/1022594.1022596.

Peng PJ, Li JF, Chen LY, Lee J. 2017. 6.1 a 56Gb/s PAM-4/NRZ transceiver in 40nm CMOS. In: *2017 IEEE International Solid-State Circuits Conference (ISSCC).* Piscataway: IEEE, 110–111.

Pierce F. 2018. Energy hogs: can world's huge data centers be made more efficient? *Yale Environment* **360**:3–4.

Pisati M, De Bernardinis F, Pascale P, Nani C, Sosio M, Pozzati E, Ghittori N, Magni F, Garampazzi M, Bollati G, Milani A. 2019. 6.3 A Sub-250mW 1-to-56Gb/s continuous-range PAM-4 42.5 dB IL ADC/DAC-based transceiver in 7nm FinFET. In: *2019 IEEE International Solid-State Circuits Conference—(ISSCC).* Piscataway: IEEE, 116–118.

Prezioso M, Merrikh-Bayat F, Hoskins BD, Adam GC, Likharev KK, Strukov DB. 2015. Training and operation of an integrated neuromorphic network based on metal-oxide memristors. *Nature* **521(7550)**:61–64 DOI 10.1038/nature14441.

Ragab A, Liu Y, Hu K, Chiang P, Palermo S. 2011. Receiver jitter tracking characteristics in high-speed source synchronous links. *Journal of Electrical and Computer Engineering* **2011(10)**:1–15 DOI 10.1155/2011/982314.

Raghavan B, Cui D, Singh U, Maarefi H, Pi D, Vasani A, Huang ZC, Çatlı B, Momtaz A, Cao J. 2013. A sub-2 W 39.8-44.6 Gb/s transmitter and receiver chipset with SFI-5.2 interface in 40 nm CMOS. *IEEE Journal of Solid-State Circuits* **48(12)**:3219–3228 DOI 10.1109/JSSC.2013.2279054.

Rakowski M, Meagher C, Nummy K, Aboketaf A, Ayala J, Bian Y, Harris B, Mclean K, McStay K, Sahin A, Medina L. 2020. 45nm CMOS-Silicon Photonics Monolithic Technology (45CLO) for next-generation, low power and high speed optical interconnects. In: *Optical Fiber Communication Conference.* Optical Society of America, T3H.3.

Redaelli A, Pirovano A, Pellizzer F, Lacaita AL, Ielmini D, Bez R. 2004. Electronic switching effect and phase-change transition in chalcogenide materials. *IEEE Electron Device Letters* **25(10)**:684–686 DOI 10.1109/LED.2004.836032.

Settaluri K, Haj-Ali A, Huang Q, Hakhamaneshi K, Nikolic B. 2020. AutoCkt: deep reinforcement learning of analog circuit designs. In: *Design, Automation & Test in Europe Conference & Exhibition (DATE).* 1–6.

Shannon CE. 1948. A mathematical theory of communication. *Bell System Technical Journal* **27(3)**:379–423 DOI 10.1002/j.1538-7305.1948.tb01338.x.

Shao S. 2019. *Towards an intelligent edge: wireless meets AI. BWRC Fall Retreat*. Berkeley: Berkeley Wireless Research Center.

Shevgoor M, Muralimanohar N, Balasubramonian R, Jeon Y. 2015. Improving memristor memory with sneak current sharing. In: *2015 33rd IEEE International Conference on Computer Design (ICCD).* Piscataway: IEEE, 549–556.

Shibasaki T, Danjo T, Ogata Y, Sakai Y, Miyaoka H, Terasawa F, Kudo M, Kano H, Matsuda A, Kawai S, Arai T. 2016. 3.5 A 56Gb/s NRZ-electrical 247mW/lane serial-link transceiver in 28nm CMOS. In: *2016 IEEE International Solid-State Circuits Conference (ISSCC).* IEEE, 64–65.

Sohn K, Yun WJ, Oh R, Oh CS, Seo SY, Park MS, Shin DH, Jung WC, Shin SH, Ryu JM, Yu HS. 2016. A 1.2 V 20 nm 307 GB/s HBM DRAM with at-speed wafer-level IO test scheme and adaptive refresh considering temperature distribution. *IEEE Journal of Solid-State Circuits* **52(1)**:250–260 DOI 10.1109/JSSC.2016.2602221.

Stojanovic V, Ho A, Garlepp BW, Chen F, Wei J, Tsang G, Alon E, Kollipara RT, Werner CW, Zerbe JL, Horowitz MA. 2005. Autonomous dual-mode (PAM2/4) serial link transceiver with adaptive equalization and data recovery. *IEEE Journal of Solid-State Circuits* **40(4)**:1012–1026 DOI 10.1109/JSSC.2004.842863.

Sun C, Georgas M, Orcutt J, Moss B, Chen YH, Shainline J, Wade M, Mehta K, Nammari K, Timurdogan E, Miller D. 2015a. A monolithically-integrated chip-to-chip optical link in bulk CMOS. *IEEE Journal of Solid-State Circuits* **50(4)**:828–844 DOI 10.1109/JSSC.2014.2382101.

Sun C, Jeong D, Zhang M, Bae W, Zhang C, Bhargava P, Van Orden D, Ardalan S, Ramamurthy C, Anderson E, Katzin A, Lu H, Buchbinder S, Beheshtian B, Khilo A, Rust M, Li C, Sedgwick F, Fini J, Meade R, Stojanović V, Wade M. 2020. TeraPHYTM: an O-band WDM electro-optic platform for low power, Terabit/s Optical I/O. In: *Proceedings of Symposium VLSI Technology.* Piscataway: IEEE, 1–2.

Sun C, Wade MT, Lee Y, Orcutt JS, Alloatti L, Georgas MS, Waterman AS, Shainline JM, Avizienis RR, Lin S, Moss BR. 2015b. Single-chip microprocessor that communicates directly using light. *Nature* **528(7583)**:534–538 DOI 10.1038/nature16454.

Sung M, Jang SA, Lee H, Ji YH, Kang JI, Jung TO, Ahn TH, Son YI, Kim HC, Lee SW, Lee SM. 2015. Gate-first high-k/metal gate DRAM technology for low power and high performance products. In: *2015 IEEE International Electron Devices Meeting (IEDM).* Piscataway: IEEE, 26.6.1–26.6.4.

Takemoto T, Yamashita H, Kamimura T, Yuki F, Masuda N, Toyoda H, Chujo N, Kogo K, Lee Y, Tsuji S, Nishimura S. 2012. A 25-Gb/s 2.2-W optical transceiver using an analog FE tolerant to power supply noise and redundant data format conversion in 65-nm CMOS. In: *2012 Symposium on VLSI Circuits (VLSIC).* Piscataway: IEEE, 106–107.

Tamura H, Kibune M, Takahashi Y, Doi Y, Chiba T, Higashi H, Takauchi H, Ishida H, Gotoh K. 2001. 5 Gb/s bidirectional balanced-line link compliant with plesiochronous clocking. In: *2001 IEEE International Solid-State Circuits Conference. Digest of Technical Papers. ISSCC.* Piscataway: IEEE, 64–65.

Tanaka K, Fukaishi M, Takeuchi M, Yoshida N, Minami K, Yamaguchi K, Uchida H, Morishita Y, Sakamoto T, Kaneko T, Soda M. 2002. A 100 Gb/s transceiver with GND-VDD common-mode receiver and flexible multi-channel aligner. In: *2002 IEEE International Solid-State Circuits Conference. Digest of Technical Papers.* Piscataway: IEEE, 264–465.

Tanaka T, Helm M, Vali T, Ghodsi R, Kawai K, Park JK, Yamada S, Pan F, Einaga Y, Ghalam A, Tanzawa T. 2016. 7.7 A 768Gb 3b/cell 3D-floating-gate NAND flash memory. In: *2016 IEEE International Solid-State Circuits Conference (ISSCC).* Piscataway: IEEE, 142–144.

Tang L, Gai W, Shi L, Xiang X, Sheng K, He A. 2018. A 32Gb/s 133mW PAM-4 transceiver with DFE based on adaptive clock phase and threshold voltage in 65nm CMOS. In: *2018 IEEE International Solid—State Circuits Conference—(ISSCC).* Piscataway: IEEE, 114–116.

Thraskias CA, Lallas EN, Neumann N, Schares L, Offrein BJ, Henker R, Plettemeier D, Ellinger F, Leuthold J, Tomkos I. 2018. Survey of photonic and plasmonic interconnect technologies for intra-datacenter and high-performance computing communications. *IEEE Communications Surveys & Tutorials* **20(4)**:2758–2783 DOI 10.1109/COMST.2018.2839672.

Tran K. 2016. Start your HBM/2.5 D design today. *High-Bandwidth Memory White Paper* **6**:1–2.

Upadhyaya P, Poon CF, Lim SW, Cho J, Roldan A, Zhang W, Namkoong J, Pham T, Xu B, Lin W, Zhang H. 2018. A fully adaptive 19-to-56Gb/s PAM-4 wireline transceiver with a configurable ADC in 16nm FinFET. In: *2018 IEEE International Solid—State Circuits Conference—(ISSCC).* Piscataway: IEEE, 108–110.

Upadhyaya P, Savoj J, An FT, Bekele A, Jose A, Xu B, Wu D, Turker D, Aslanzadeh H, Hedayati H, Im J. 2015. 3.3 A 0.5-to-32.75 Gb/s flexible-reach wireline transceiver in 20nm CMOS. In: *2015 IEEE International Solid-State Circuits Conference—(ISSCC) Digest of Technical Papers.* Piscataway: IEEE, 1–3.

Vandersypen L, van Leeuwenhoek A. 2017. 1.4 quantum computing-the next challenge in circuit and system design. In: *2017 IEEE International Solid-State Circuits Conference (ISSCC).* Piscataway: IEEE, 24–29.

Vontobel PO, Robinett W, Kuekes PJ, Stewart DR, Straznicky J, Williams RS. 2009. Writing to and reading from a nano-scale crossbar memory based on memristors. *Nanotechnology* **20(42)**:425204 DOI 10.1088/0957-4484/20/42/425204.

Vučinić D, Wang Q, Guyot C, Mateescu R, Blagojević F, Franca-Neto L, Le Moal D, Bunker T, Xu J, Swanson S, Bandić Z. 2014. {DC} Express: shortest latency protocol for reading phase change memory over {PCI} express. In: *12th {USENIX} Conference on File and Storage Technologies ({FAST} 14).* 309–315.

Wang T. 2017. Modelling multistability and hysteresis in ESD clamps, memristors and other devices. In: *2017 IEEE CICC.* Piscataway: IEEE, 1–10.

Wang A, Bae W, Han J, Bailey S, Ocal O, Rigge P, Wang Z, Ramchandran K, Alon E, Nikolić B. 2019. A real-time, 1.89-GHz bandwidth, 175-kHz resolution sparse spectral analysis RISC-V SoC in 16-nm FinFET. *IEEE Journal of Solid-State Circuits* **54(7)**:1993–2008 DOI 10.1109/JSSC.2019.2913099.

Wang L, Fu Y, LaCroix MA, Chong E, Carusone AC. 2018. A 64-Gb/s 4-PAM transceiver utilizing an adaptive threshold ADC in 16-nm FinFET. *IEEE Journal of Solid-State Circuits* **54(2)**:452–462 DOI 10.1109/JSSC.2018.2877172.

Wang Z, Li C, Lin P, Rao M, Nie Y, Song W, Qiu Q, Li Y, Yan P, Strachan JP, Ge N. 2019. In situ training of feed-forward and recurrent convolutional memristor networks. *Nature Machine Intelligence* **1(9)**:434–442 DOI 10.1038/s42256-019-0089-1.

Whitcombe A, Nikolic B. 2019. Configurable data converters for digitally adaptive radio. *Available at https://digitalassets.lib.berkeley.edu/techreports/ucb/text/EECS-2019-156.pdf.*

White M. 2021. Are you really ready for your next node? *Blog. Available at https://blogs.sw.siemens.com/calibre/2017/01/11/are-you-really-ready-for-your-next-node/* (accessed 19 March 2021).

Whitney J, Delforge P. 2014. "Data center efficiency assessment—scaling up energy efficiency across the data center industry: evaluating key drivers and barriers," NRDC and Anthesis, Rep. IP:14-08-A, Aug. 2014. *Available at https://www.nrdc.org/sites/default/files/data-center-efficiency-assessment-IP.pdf*.

Wong HSP, Lee HY, Yu S, Chen YS, Wu Y, Chen PS, Lee B, Chen FT, Tsai MJ. 2012. Metal-oxide RRAM. *Proceedings of the IEEE* **100(6)**:1951–1970 DOI 10.1109/JPROC.2012.2190369.

Wong HSP, Raoux S, Kim S, Liang J, Reifenberg JP, Rajendran B, Asheghi M, Goodson KE. 2010. Phase change memory. *Proceedings of the IEEE* **98(12)**:2201–2227 DOI 10.1109/JPROC.2010.2070050.

Xu Q, Siyamwala H, Ghosh M, Suri T, Awasthi M, Guz Z, Shayesteh A, Balakrishnan V. 2015. Performance analysis of NVMe SSDs and their implication on real world databases. In: *Proceedings of the 8th ACM International Systems and Storage Conference.* New York: ACM, 1–11.

Xue CX, Chen WH, Liu JS, Li JF, Lin WY, Lin WE, Wang JH, Wei WC, Chang TW, Chang TC, Huang TY. 2019. 24.1 A 1Mb multibit ReRAM computing-in-memory macro with 14.6 ns parallel MAC computing time for CNN based AI edge processors. In: *2019 IEEE International Solid-State Circuits Conference—(ISSCC).* Piscataway: IEEE, 388–390.

Yeric G. 2019. IC design after Moore's Law. In: *2019 IEEE CICC.* Piscataway: IEEE, 1–150.

Yilmaz E, Dundar G. 2008. Analog layout generator for CMOS circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* **28(1)**:32–45 DOI 10.1109/TCAD.2008.2009137.

Yoo HJ. 2019. 1.2 intelligence on silicon: from deep-neural-network accelerators to brain mimicking AI-SoCs. In: *2019 IEEE International Solid-State Circuits Conference—(ISSCC).* Piscataway: IEEE, 20–26.

Yoo BJ, Lim DH, Pang H, Lee JH, Baek SY, Kim N, Choi DH, Choi YH, Yang H, Yoon T, Chu SH. 2020. 6.4 A 56Gb/s 7.7mW/Gb/s PAM-4 wireline transceiver in 10nm FinFET using MM-CDR-Based ADC timing skew control and low-power DSP with approximate multiplier. In: *2020 IEEE International Solid-State Circuits Conference—(ISSCC).* Vol. 6.. Piscataway: IEEE, 122–124.

Yoon KJ, Bae W, Jeong DK, Hwang CS. 2016. Comprehensive writing margin analysis and its application to stacked one diode-one memory device for high-density crossbar resistance switching random access memory. *Advanced Electronic Materials* **2(10)**:1600326 DOI 10.1002/aelm.201600326.

Yoon KJ, Han JW, Bae W. 2020. A novel stateful logic device and circuit for in-memory parity programming in crossbar memory. *Advanced Electronic Materials* **6(12)**:202000672.

Yoon KJ, Kim Y, Hwang CS. 2019. What will come after V-NAND—vertical resistive switching memory? *Advanced Electronic Materials* **5(9)**:1800914 DOI 10.1002/aelm.201800914.

Yoon KJ, Kim GH, Yoo S, Bae W, Yoon JH, Park TH, Kwon DE, Kwon YJ, Kim HJ, Kim YM, Hwang CS. 2017. Double-layer-stacked one diode-one resistive switching memory crossbar array with an extremely high rectification ratio of 109. *Advanced Electronic Materials* **3(7)**:1700152 DOI 10.1002/aelm.201700152.

Young IA, Mohammed E, Liao JT, Kern AM, Palermo S, Block BA, Reshotko MR, Chang PL. 2009. Optical I/O technology for tera-scale computing. *IEEE Journal of Solid-State Circuits* **45(1)**:235–248 DOI 10.1109/JSSC.2009.2034444.

Zerbe JL, Werner CW, Stojanovic V, Chen F, Wei J, Tsang G, Kim D, Stonecypher WF, Ho A, Thrush TP, Kollipara RT. 2003. Equalization and clock recovery for a 2.5-10-Gb/s 2-PAM/4-PAM backplane transceiver cell. *IEEE Journal of Solid-State Circuits* **38(12)**:2121–2130 DOI 10.1109/JSSC.2003.818572.

Zhang B, Khanoyan K, Hatamkhani H, Tong H, Hu K, Fallahi S, Abdul-Latif M, Vakilian K, Fujimori I, Brewster A. 2015. A 28 Gb/s multistandard serial link transceiver for backplane applications in 28 nm CMOS. *IEEE Journal of Solid-State Circuits* **50(12)**:3089–3100 DOI 10.1109/JSSC.2015.2475180.