Today’s computing challenges: opportunities for computer hardware design

Woorham Bae Corresp. 1, 2

1 Circuits Department, Ayar Labs, Santa Clara, CA, United States

2 Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, CA, United States

Corresponding Author: Woorham Bae
Email address: wrbae@eecs.berkeley.edu

Due to 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 limit, and therefore they 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 big challenge for computer hardware industry. However, at the same time it also provides great opportunities to 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.

Today's computing challenges: Opportunities for Computer hardware design

Woorham Bae[1,2]

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

Corresponding Author:
Woorham Bae
Olcott Street, Santa Clara, CA, 95054, USA
Email address: wrbae@eecs.berkeley.edu

Abstract

Due to 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 limit, and therefore they 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 big challenge for computer hardware industry. However, at the same time it also provides great opportunities to 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.

I. 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 [1], [2]. That is mainly because the world needs more electronics to handle the data. As mentioned above, the world is pushing to realize whole bunch of things (such as smart city, security, autonomous vehicle…), 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, Figure 1 shows Cisco's two reports on the amount of data creation in the world, which were released in 2017 and 2019 [3], [4]. Both reports predict 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 the electronics is responsible to handle 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 five years. Nvidia gives a bit more aggressive prediction, such that the number of total connected devices will increase by 16 times in seven years [5]. 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 Figure 1, where the Cisco's projection is shown, the amount of digital data is going to explode. If so, what happens 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. Known that we are already consuming the largest portion of energy in the Earth for handling the data with electronics [6]–[8], 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 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 [9], [10], enables a transistor to be faster and consume even less power [11]–[14]. 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 [15]. For the memory side, the scaling of device footprint enabled a higher memory density [16]. 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 [17]–[19]. 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 [20], [21]. 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 [21]–[25]. They also introduced time-interleaving technique, which is something like the parallelism, to achieve very high speed even above the transistor limit [26]–[29].

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” [14], [30].

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. And it will be more tough 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 ten years.

The remainder of this paper is organized as follows. Section II presents potential solutions for addressing the challenges of the logic and opportunities for the computer hardware design industry and engineers. Section III 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. Section IV 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 Section V.

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.

II. Logic (System Semiconductor)

A. 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. So, let us take a quick look at where the scaling limit comes from. In his talk at ISSCC2015 [2], the president of Samsung Electronics told that the physical limit of transistor dimension is around 1.5nm, which is given from Heisenberg's uncertainty principle. However, he also told that he expected that the practical limit will be 3nm. After five years, now, the 7nm technology is already widely available in the industry. And the leading foundries such as TSMC and Samsung Electronics are already working on 5nm and 3nm technology development, which means that we are almost there.

As a result, recalling the energy discussion in the section I, 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 45nm CLO process, which is specialized to silicon photonics [32].

159
160 **B. Productivity problem of specialization**
161 We found that the specialization would be a potential solution to resolve the energy problem and
162 to retain the continuous advance of computing. However, there are also some downsides of the
163 specialization, so we need to investigate how profit is made in the new age with the
164 specialization. In a simplified model in Figure 4(a), a fabless company shipped 1 million units of
165 a generic chip before, but they are planning to design 10 specialized chips in 10 different
166 processes to meet the better efficiency requirement. At the same time, they are expecting they
167 can ship 2 million chips in total as there will be more demand of electronics. In the model, the
168 company is currently making $3 million profit. On the right side of Figure 4(a), a linear
169 extrapolation is made to when the company designs 10 specialized chips and total shipping is
170 doubled. Note that all the cost is extrapolated in linearly proportional to the amount of
171 production.
172 However, it is too optimistic projection. Figure 4(b) shows a bit more realistic model. The
173 revenue and production cost are indeed proportional to the amount of shipping. However, does it
174 make sense to extrapolate other expenses? Of course, the answer is no. For example, the amount
175 of manpower cannot be scaled linearly. To design a single complete chip, they need analog
176 engineers, digital engineers, manufacturing engineers and more. Therefore, it makes no sense
177 that only 4 engineers can make a chip which used to be made by 20 engineers, even though there
178 must be some amount of efforts that can be shared among the chip designs. So, the model in
179 Figure 4(b) assumes 10 engineers can design a specialized chip A0. If so, the profit becomes
180 minus. The calculation here is very rough, but at least we can observe a large fraction of design
181 cost is not scaled with the amount of production. The company would raise the price, but
182 customers will not be happy with that. Then, is the specialization a false dream?
183 The most reasonable solution here is to reduce the design time, since such design costs are
184 proportional to the design time, as shown in Figure 4(c). For example, if they can reduce the
185 design time by half, they can reduce the expenses by half, then they can make more profit. As
186 mentioned earlier, they are designing 10 different but similar chips, and there is some amount of
187 sharable efforts. That means, if they maximize the amount, they should be able to reduce the
188 design time considerably.
189
190 **C. Reducing design time by reusing design**
191 Then what should we try to maximize shareability? Generally speaking, we can say the analog
192 and mixed signal (AMS) circuit design is usually the bottleneck of reducing design time. That is
193 mainly because AMS circuit designs highly rely on human's heuristic knowledge and skills,
194 compared to the digital design. Moreover, the design complexity has been increased as
195 technology scales down, due to the complex design rules and digital-friendly scaling of CMOS
196 technology, which is represented by the number of design rules shown in Figure 5(a), where we
197 can find the design complexity has been increased exponentially as the technology scales down
198 [33]–[35]. Figure 5(b) 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 Figure 5(a). 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 [36], [37], so they spend most of the design time constraining the tools, which is not very efficient [38]. 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-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 [39]–[44]. 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 Figure 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 [45]–[47]. They are quite different each other as summarized in Figure 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 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 Figure 7) or they should spend more time to make it as good as custom design (right side of Figure 7). For either way, a good script has to have flexible parameterizations [38]. 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 Figure 5(b). 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 [48]. 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.

III. Memory and Storage

A. 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 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 section I. 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) [50]–[60]. 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 [51]–[53], [61]. 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 advancements on processing units, burden more on the interconnect side for higher bandwidth and lower latency [53], [62], [63]. 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 [64]–[68]. 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 [61], [69]. In addition, the stacked DRAM dies are connected by TSV links whose characteristic is also very different [70]–[72]. 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 [73], [74], which should be overcome by hardware design.

B. 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 Figure 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 [75]–[82]. 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 [83]; 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 [84]–[88]. On the other hand, some circuit design techniques can be introduced to mitigate the snapback current [89]–[91]. Also, circuit designers can propose variation-tolerant or variation-compensated techniques to address the variation issue [92]–[95], or sneak-current cancellation scheme for the sneak current issue [96]–[98]. 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 [99]–[105]. 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.

IV. Interconnect

A. Trend survey and challenges

In this section, the challenges and potential solutions of computer communication interconnect are presented. Recalling the section I, 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 channel (wire), and then de-serializes the serial input to parallel at the receive side [106]–[109]. 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 11(a) shows a survey from the state-of-the-art published works[110]–[142], 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 11(b) 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 [133]–[142]. 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 Figure 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? Figure 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. 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 Figure 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 [143].

B. 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 [117], [144]– [149]. The bit-error-rate (BER) of an interconnect is a function of the amplitude noise (SNR) and the timing noise (jitter) [150]. 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 [137]–[142], [151]–[154], 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. 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 [155]–[158]. 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 Figure 13(a) [159]. 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 Figure 13(b). 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 [159]. 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 1000x lower than that of electrical interconnects [160]. 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 [160]–[163]. 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 [160], 6-pJ/bit 112-Gb/s PAM-4 [164], so the time will really come soon.

V. 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 Figure 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.

References

[1] Horowitz, M., 2014, February. 1.1 computing's energy problem (and what we can do about it). In *IEEE ISSCC Dig. Tech. Papers* (pp. 10-14). IEEE.

[2] Kim, K., 2015, February. 1.1 silicon technologies and solutions for the data-driven world. In *IEEE ISSCC Dig. Tech. Papers* (pp. 1-7). IEEE.

[3] Index, C.V.N., 2017. Forecast and Trends. *Cisco Systems Inc*, pp.2017-2022.

[4] Index, C.V.N., 2019. Forecast and Trends, 2017–2022 White Paper. *Cisco Systems Inc.: San Jose, CA, USA*.

[5] Shao, S. 2019, October. Towards an Intelligent Edge: Wireless Meets AI. BWRC Fall Retreat. Berkeley Wireless Research Center.

[6] Bae, W., Jeong, G.S. and Jeong, D.K., 2016. A 1-pJ/bit, 10-Gb/s/ch forwarded-clock transmitter using a resistive feedback inverter-based driver in 65-nm CMOS. *IEEE Trans. Circuits Syst. II, Exp. Briefs*, *63*(12), pp.1106-1110.

[7] J. Whitney and P. Delforge, “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.

[8] Pierce, F., 2018. Energy Hogs: Can World’s Huge Data Centers Be Made More Efficient?. *Yale Environment*, *360*.

[9] Moore, G.E., 1965. Cramming more components onto integrated circuits.

[10] Moore, G.E., 1975, December. Progress in digital integrated electronics. In *Electron devices meeting* (Vol. 21, pp. 11-13).

[11] Holt, W.M., 2016, January. 1.1 Moore's law: A path going forward. In *IEEE ISSCC Dig. Tech. Papers* (pp. 8-13). IEEE.

[12] Bohr, M.T. and Young, I.A., 2017. CMOS scaling trends and beyond. *IEEE Micro*, *37*(6), pp.20-29.

[13] Mak, P.I. and Martins, R.P., 2010. High-/mixed-voltage RF and analog CMOS circuits come of age. *IEEE Circuits and Systems Magazine*, *10*(4), pp.27-39.

[14] Yeric, G., 2019, April. IC Design After Moore’s Law. In *2019 IEEE CICC* (pp. 1-150). IEEE.

[15] Danowitz, A., Kelley, K., Mao, J., Stevenson, J.P. and Horowitz, M., 2012. CPU DB: recording microprocessor history. *Queue*, *10*(4), pp.10-27.

[16] Hwang, C.G., 2002, February. Semiconductor memories for IT era. In *ISSCC Dig. Tech. Papers* (Vol. 1, pp. 24-27). IEEE.

[17] Mueller, W., Aichmayr, G., Bergner, W., Erben, E., Hecht, T., Kapteyn, C., Kersch, A., Kudelka, S., Lau, F., Luetzen, J. and Orth, A., 2005, December. Challenges for the DRAM cell scaling to 40nm. In *IEEE IEDM Tech. Dig.* (pp. 4-pp). IEEE.

[18] Sung, M., Jang, S.A., Lee, H., Ji, Y.H., Kang, J.I., Jung, T.O., Ahn, T.H., Son, Y.I., Kim, H.C., Lee, S.W. and Lee, S.M., 2015, December. Gate-first high-k/metal gate DRAM technology for low power and high performance products. In *IEEE IEDM Tech. Dig.,* (pp. 26-6). IEEE.

[19] Jang, S.H., Lim, J., Han, J., Jang, J., Yeo, J., Lee, C., Baek, S., Lee, J., Lee, J.H., Yamada, S. and Lee, K., 2019, December. A Fully Integrated Low Voltage DRAM with Thermally Stable Gate-first High-k Metal Gate Process. In *2019 IEEE IEDM Tech. Dig.,* (pp. 28-4). IEEE.

[20] Daly, D.C., Fujino, L.C. and Smith, K.C., 2018. Through the looking glass-the 2018 edition: trends in solid-state circuits from the 65th ISSCC. *IEEE Solid-State Circuits Magazine*, *10*(1), pp.30-46.

[21] Horowitz, M., Yang, C.K.K. and Sidiropoulos, S., 1998. High-speed electrical signaling: Overview and limitations. *IEEE Micro*, *18*(1), pp.12-24.

[22] Dally, W.J. and Poulton, J., 1997. Transmitter equalization for 4-Gbps signaling. *IEEE Micro*, *17*(1), pp.48-56.

[23] Zerbe, J.L., Werner, C.W., Stojanovic, V., Chen, F., Wei, J., Tsang, G., Kim, D., Stonecypher, W.F., Ho, A., Thrush, T.P. and Kollipara, R.T., 2003. Equalization and clock recovery for a 2.5-10-Gb/s 2-PAM/4-PAM backplane transceiver cell. *IEEE J. Solid-State Circuits*, *38*(12), pp.2121-2130.

[24] Stojanovic, V., Ho, A., Garlepp, B.W., Chen, F., Wei, J., Tsang, G., Alon, E., Kollipara, R.T., Werner, C.W., Zerbe, J.L. and Horowitz, M.A., 2005. Autonomous dual-mode (PAM2/4) serial link transceiver with adaptive equalization and data recovery. *IEEE J. Solid-State Circuits*, *40*(4), pp.1012-1026.

[25] Choi, J.S., Hwang, M.S. and Jeong, D.K., 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 J. Solid-State Circuits*, *39*(3), pp.419-425.

[26] Kim, J. and Horowitz, M.A., 2002. Adaptive supply serial links with sub-1-V operation and per-pin clock recovery. *IEEE J. Solid-State Circuits*, *37*(11), pp.1403-1413.

[27] Lee, M.J., Dally, W.J. and Chiang, P., 2000. Low-power area-efficient high-speed I/O circuit techniques. *IEEE J. Solid-State Circuits*, *35*(11), pp.1591-1599.

[28] Musah, T., Jaussi, J.E., Balamurugan, G., Hyvonen, S., Hsueh, T.C., Keskin, G., Shekhar, S., Kennedy, J., Sen, S., Inti, R. and 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 J. Solid-State Circuits*, *49*(12), pp.3079-3090.

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

[30] Esmaeilzadeh, H., Blem, E., Amant, R.S., Sankaralingam, K. and Burger, D., 2011, June. Dark silicon and the end of multicore scaling. In *2011 38th Annual international symposium on computer architecture (ISCA)* (pp. 365-376). IEEE.

[31] Vandersypen, L. and van Leeuwenhoek, A., 2017, February. 1.4 quantum computing-the next challenge in circuit and system design. In *IEEE ISSCC Dig. Tech. Papers* (pp. 24-29). IEEE.

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

[33] M. White, “Are you really ready for your next node?" https://blogs.mentor.com/calibre/blog/2017/01/11/are-you-really-ready-for-your-next-node/.

[34] Whitcombe, A. and Nikolic, B., 2019. Configurable Data Converters for Digitally Adaptive Radio.

[35] Han, J., Bae, W., Chang, E., Wang, Z., Nikolic, B., and Alon, E., LAYGO: A Layout Generation Framework to Enhance Design Productivity in Advanced CMOS Technologies. *IEEE Trans. Circuits Syst. I, Reg. Papers*, in press.

[36] Habal, H. and Graeb, H., 2011. Constraint-based layout-driven sizing of analog circuits. *IEEE Trans. Computer-Aided Design Integr. Circuits Syst.*, *30*(8), pp.1089-1102.

[37] Lin, P.H., Chang, Y.W. and Lin, S.C., 2009. Analog placement based on symmetry-island formulation. *IEEE Trans. Computer-Aided Design Integr. Circuits Syst.*, *28*(6), pp.791-804.

[38] Chang, E., Han, J., Bae, W., Wang, Z., Narevsky, N., NikoliC, B. and Alon, E., 2018, April. BAG2: A process-portable framework for generator-based AMS circuit design. In *2018 IEEE CICC* (pp. 1-8). IEEE.

[39] Crossley, J., Puggelli, A., Le, H.P., Yang, B., Nancollas, R., Jung, K., Kong, L., Narevsky, N., Lu, Y., Sutardja, N. and An, E.J., 2013, November. BAG: A designer-oriented integrated framework for the development of AMS circuit generators. In *2013 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)* (pp. 74-81). IEEE.

[40] Yilmaz, E. and Dundar, G., 2008. Analog layout generator for CMOS circuits. *IEEE Trans. Computer-Aided Design Integr. Circuits Syst.*, *28*(1), pp.32-45.

[41] Castro-Lopez, R., Guerra, O., Roca, E. and Fernández, F.V., 2008. An integrated layout-synthesis approach for analog ICs. *IEEE Trans. Computer-Aided Design Integr. Circuits Syst.*, *27*(7), pp.1179-1189.

[42] Martins, R., Lourenco, N. and Horta, N., 2013. LAYGEN II—Automatic layout generation of analog integrated circuits. *IEEE Trans. Computer-Aided Design Integr. Circuits Syst.*, *32*(11), pp.1641-1654.

[43] Kunal, K., Madhusudan, M., Sharma, A.K., Xu, W., Burns, S.M., Harjani, R., Hu, J., Kirkpatrick, D.A. and Sapatnekar, S.S., 2019, June. ALIGN: Open-source analog layout automation from the ground up. In *Proceedings of the 56th Annual Design Automation Conference 2019* (pp. 1-4).

[44] Wang, A., Bae, W., Han, J., Bailey, S., Ocal, O., Rigge, P., Wang, Z., Ramchandran, K., Alon, E. and Nikolić, B., 2019. A real-time, 1.89-GHz bandwidth, 175-kHz resolution sparse spectral analysis RISC-V SoC in 16-nm FinFET. *IEEE J. Solid-State Circuits*, *54*(7), pp.1993-2008.

[45] Berkeley analog generator, main framework. [Online]. Available: https://github.com/ucb-art/BAG_framework

[46] Berkeley analog generator, layout with gridded objects (Laygo). [Online]. Available: https://github.com/ucb-art/laygo

[47] Ayar Custom Generator (ACG). [Online]. Available: https://github.com/AyarLabs/ACG

[48] Hakhamaneshi, K., Werblun, N., Abbeel, P. and Stojanović, V., 2019, November. BagNet: Berkeley Analog Generator with Layout Optimizer Boosted with Deep Neural Networks. In *2019 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)* (pp. 1-8). IEEE.

[49] Settaluri, K., Haj-Ali, A., Huang, Q., Hakhamaneshi, K. and Nikolic, B., 2020, January. AutoCkt: Deep Reinforcement Learning of Analog Circuit Designs. In *Design, Automation & Test in Europe Conference & Exhibition (DATE)* (pp. 1-6).

[50] Lee, D.U., Kim, K.W., Kim, K.W., Lee, K.S., Byeon, S.J., Kim, J.H., Cho, J.H., Lee, J. and Chun, J.H., 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 J. Solid-State Circuits*, *50*(1), pp.191-203.

[51] O'Connor, M., 2014, June. Highlights of the high-bandwidth memory (hbm) standard. In *Memory Forum Workshop*.

[52] Tran, K., 2016. Start Your HBM/2.5 D Design Today. *High-Bandwidth Memory White Paper*, *6*.

[53] Jun, H., Cho, J., Lee, K., Son, H.Y., Kim, K., Jin, H. and Kim, K., 2017, May. Hbm (high bandwidth memory) dram technology and architecture. In *2017 IEEE International Memory Workshop (IMW)* (pp. 1-4). IEEE.

[54] Xu, Q., Siyamwala, H., Ghosh, M., Suri, T., Awasthi, M., Guz, Z., Shayesteh, A. and Balakrishnan, V., 2015, May. Performance analysis of NVMe SSDs and their implication on real world databases. In *Proceedings of the 8th ACM International Systems and Storage Conference* (pp. 1-11).

[55] Kim, H.J., Lee, Y.S. and Kim, J.S., 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)*.

[56] Kim, W., Choi, S., Sung, J., Lee, T., Park, C., Ko, H., Jung, J., Yoo, I. and Park, Y., 2009, June. Multi-layered vertical gate NAND flash overcoming stacking limit for terabit density storage. In *Proc. Symp. VLSI Technology* (pp. 188-189). IEEE.

[57] Kim, C., Kim, D.H., Jeong, W., Kim, H.J., Park, I.H., Park, H.W., Lee, J., Park, J., Ahn, Y.L., Lee, J.Y. and Kim, S.B., 2017. A 512-gb 3-b/cell 64-stacked wl 3-d-nand flash memory. *IEEE J. Solid-State Circuits*, *53*(1), pp.124-133.

[58] Tanaka, T., Helm, M., Vali, T., Ghodsi, R., Kawai, K., Park, J.K., Yamada, S., Pan, F., Einaga, Y., Ghalam, A. and Tanzawa, T., 2016, January. 7.7 A 768Gb 3b/cell 3D-floating-gate NAND flash memory. In *IEEE ISSCC Dig. Tech. Papers* (pp. 142-144). IEEE.

[59] Im, J.W., Jeong, W.P., Kim, D.H., Nam, S.W., Shim, D.K., Choi, M.H., Yoon, H.J., Kim, D.H., Kim, Y.S., Park, H.W. and Kwak, D.H., 2015, February. 7.2 A 128Gb 3b/cell V-NAND flash memory with 1Gb/s I/O rate. In *IEEE ISSCC Dig. Tech. Papers* (pp. 1-3). IEEE.

[60] Park, K.T., Nam, S., Kim, D., Kwak, P., Lee, D., Choi, Y.H., Choi, M.H., Kwak, D.H., Kim, D.H., Kim, M.S. and Park, H.W., 2014. Three-dimensional 128 Gb MLC vertical NAND flash memory with 24-WL stacked layers and 50 MB/s high-speed programming. *IEEE J. Solid-State Circuits*, *50*(1), pp.204-213.

[61] Ko, H.G., Shin, S., Oh, J., Park, K. and Jeong, D.K., 2020, February. 6.7 An 8Gb/s/μm FFE-Combined Crosstalk-Cancellation Scheme for HBM on Silicon Interposer with 3D-Staggered Channels. In *IEEE ISSCC Dig. Tech. Papers* (pp. 128-130). IEEE.

[62] Patterson, D.A., 2004. Latency lags bandwith. *Communications of the ACM*, *47*(10), pp.71-75.

[63] Hsieh, K., Ebrahimi, E., Kim, G., Chatterjee, N., O'Connor, M., Vijaykumar, N., Mutlu, O. and Keckler, S.W., 2016. Transparent offloading and mapping (TOM) enabling programmer-transparent near-data processing in GPU systems. *ACM SIGARCH Computer Architecture News*, *44*(3), pp.204-216.

[64] Vučinić, D., Wang, Q., Guyot, C., Mateescu, R., Blagojević, F., Franca-Neto, L., Le Moal, D., Bunker, T., Xu, J., Swanson, S. and 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)* (pp. 309-315).

[65] Ajanovic, J., 2009, August. PCI express 3.0 overview. In *Proceedings of Hot Chip: A Symposium on High Performance Chips* (Vol. 69, p. 143).

[66] Budruk, R., 2007. Pci express basics. In *PCI-SIG Developers Conference*.

[67] Cheng, K.H., Tsai, Y.C., Wu, Y.H. and Lin, Y.F., 2010. A 5-Gb/s inductorless CMOS adaptive equalizer for PCI express generation II applications. *IEEE Trans. Circuits Syst. II, Exp. Briefs*, *57*(5), pp.324-328.

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

[69] Liu, H., Ding, Q. and Jiang, J., 2018, October. 112G PAM4/56G NRZ Interconnect Design for High channel count packages. In *2018 IEEE 27th*

673 *Conference on Electrical Performance of Electronic Packaging and Systems*
674 *(EPEPS)* (pp. 237-239). IEEE.
675 [70] Lee, H., Cho, K., Kim, H., Choi, S., Lim, J. and Kim, J., 2015, August. Electrical
676 performance of high bandwidth memory (HBM) interposer channel in terabyte/s
677 bandwidth graphics module. In *2015 International 3D Systems Integration*
678 *Conference (3DIC)* (pp. TS2-2). IEEE.
679 [71] Kim, H., Cho, J., Kim, M., Kim, K., Lee, J., Lee, H., Park, K., Choi, K., Bae,
680 H.C., Kim, J. and Kim, J., 2012. Measurement and analysis of a high-speed
681 TSV channel. *IEEE Transactions on Components, Packaging and*
682 *Manufacturing Technology*, *2*(10), pp.1672-1685.
683 [72] Lee, J.C., Kim, J., Kim, K.W., Ku, Y.J., Kim, D.S., Jeong, C., Yun, T.S., Kim, H.,
684 Cho, H.S., Kim, Y.O. and Kim, J.H., 2016. 18.3 A 1.2 V 64Gb 8-channel
685 256GB/s HBM DRAM with peripheral-base-die architecture and small-swing
686 technique on heavy load interface. In *IEEE ISSCC Dig. Tech. Papers* (pp. 318-
687 319). IEEE.
688 [73] Sohn, K., Yun, W.J., Oh, R., Oh, C.S., Seo, S.Y., Park, M.S., Shin, D.H., Jung,
689 W.C., Shin, S.H., Ryu, J.M. and Yu, H.S., 2016. A 1.2 V 20 nm 307 GB/s HBM
690 DRAM with at-speed wafer-level IO test scheme and adaptive refresh
691 considering temperature distribution. *IEEE J. Solid-State Circuits*, *52*(1),
692 pp.250-260.
693 [74] Ko, H.G., Shin, S., Kye, C.H., Lee, S.Y., Yun, J., Jung, H.K., Lee, D., Kim, S.
694 and Jeong, D.K., 2019, June. A 370-fJ/b, 0.0056 mm 2/DQ, 4.8-Gb/s DQ
695 Receiver for HBM3 with a Baud-Rate Self-Tracking Loop. In *Proc. Symp. VLSI*
696 *Circuits* (pp. C94-C94). IEEE.
697 [75] Wong, H.S.P., Raoux, S., Kim, S., Liang, J., Reifenberg, J.P., Rajendran, B.,
698 Asheghi, M. and Goodson, K.E., 2010. Phase change memory. *Proceedings of*
699 *the IEEE*, *98*(12), pp.2201-2227.
700 [76] Wong, H.S.P., Lee, H.Y., Yu, S., Chen, Y.S., Wu, Y., Chen, P.S., Lee, B.,
701 Chen, F.T. and Tsai, M.J., 2012. Metal–oxide RRAM. *Proceedings of the*
702 *IEEE*, *100*(6), pp.1951-1970.
703 [77] Bae, W., Yoon, K.J., Hwang, C.S. and Jeong, D.K., 2017. Extension of two-port
704 sneak current cancellation scheme to 3-D vertical RRAM crossbar array. *IEEE*
705 *Transactions on Electron Devices*, *64*(4), pp.1591-1596.
706 [78] Yoon, K.J., Kim, Y. and Hwang, C.S., 2019. What Will Come After V-NAND—
707 Vertical Resistive Switching Memory?. *Advanced Electronic Materials*, *5*(9),
708 p.1800914.
709 [79] Foong, A. and Hady, F., 2016, May. Storage as fast as rest of the system.
710 In *2016 IEEE 8th International Memory Workshop (IMW)* (pp. 1-4). IEEE.

[80] Kau, D., Tang, S., Karpov, I.V., Dodge, R., Klehn, B., Kalb, J.A., Strand, J., Diaz, A., Leung, N., Wu, J. and Lee, S., 2009, December. A stackable cross point phase change memory. In *IEEE IEDM Tech. Dig.,* (pp. 1-4). IEEE.

[81] Yoon, K.J., Kim, G.H., Yoo, S., Bae, W., Yoon, J.H., Park, T.H., Kwon, D.E., Kwon, Y.J., Kim, H.J., Kim, Y.M. and Hwang, C.S., 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), p.1700152.

[82] Liu, T.Y., Yan, T.H., Scheuerlein, R., Chen, Y., Lee, J.K., Balakrishnan, G., Yee, G., Zhang, H., Yap, A., Ouyang, J. and Sasaki, T., 2013. A 130.7-$\hbox {mm}^{2} $2-Layer 32-Gb ReRAM Memory Device in 24-nm Technology. *IEEE J. Solid-State Circuits*, *49*(1), pp.140-153.

[83] Yoon, K.J., Bae, W., Jeong, D.K. and Hwang, C.S., 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), p.1600326.

[84] Bae, W. and Yoon, K.J., 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*, pp.553-564.

[85] Wang, T., 2017, April. Modelling multistability and hysteresis in ESD clamps, memristors and other devices. In *2017 IEEE CICC* (pp. 1-10). IEEE.

[86] Kvatinsky, S., Friedman, E.G., Kolodny, A. and Weiser, U.C., 2012. TEAM: Threshold adaptive memristor model. *IEEE Trans. Circuits Syst. I, Reg. Papers*, *60*(1), pp.211-221.

[87] Linn, E., Siemon, A., Waser, R. and Menzel, S., 2014. Applicability of well-established memristive models for simulations of resistive switching devices. *IEEE Trans. Circuits Syst. I, Reg. Papers*, *61*(8), pp.2402-2410.

[88] Chen, P.Y. and Yu, S., 2015. Compact modeling of RRAM devices and its applications in 1T1R and 1S1R array design. *IEEE Transactions on Electron Devices*, *62*(12), pp.4022-4028.

[89] Kim, K. and Ahn, S.J., 2005, April. Reliability investigations for manufacturable high density PRAM. In *2005 IEEE International Reliability Physics Symposium, 2005. Proceedings. 43rd Annual.* (pp. 157-162). IEEE.

[90] Redaelli, A.N.D.R.E.A., Pirovano, A., Pellizzer, F., Lacaita, A.L., Ielmini, D.A.N.I.E.L.E. and Bez, R., 2004. Electronic switching effect and phase-change transition in chalcogenide materials. *IEEE Electron Device Letters*, *25*(10), pp.684-686.

[91] Parkinson, W.D., Ovonyx Inc, 2011. *Reading a phase change memory*. U.S. Patent 8,077,498.

[92] Athmanathan, A., Stanisavljevic, M., Papandreou, N., Pozidis, H. and Eleftheriou, E., 2016. Multilevel-cell phase-change memory: A viable technology. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, *6*(1), pp.87-100.

[93] Park, H.K., Lee, K.W., Song, S.H., Lee, K.G., Shin, J.H., Gangasani, V., Shin, Y.S., Kang, D.H., Park, J.H., Song, K.W. and Koh, G.H., 2017, June. A novel write method for improving RESET distribution of PRAM. In *Proc. Symp. VLSI Technology* (pp. T96-T97). IEEE.

[94] Hwang, Y.N., Um, C.Y., Lee, J.H., Wei, C.G., Oh, H.R., Jeong, G.T., Jeong, H.S., Kim, C.H. and Chung, C.H., 2010, June. MLC PRAM with SLC write-speed and robust read scheme. In *Proc. Symp. VLSI Technology* (pp. 201-202). IEEE.

[95] Bae, W., Yoon, K.J., Song, T. and Nikolić, B., 2018. A Variation-Tolerant, Sneak-Current-Compensated Readout Scheme for Cross-Point Memory Based on Two-Po0rt Sensing Technique. *IEEE Trans. Circuits Syst. II, Exp. Briefs*, *65*(12), pp.1839-1843.

[96] Vontobel, P.O., Robinett, W., Kuekes, P.J., Stewart, D.R., Straznicky, J. and Williams, R.S., 2009. Writing to and reading from a nano-scale crossbar memory based on memristors. *Nanotechnology*, *20*(42), p.425204.

[97] Shevgoor, M., Muralimanohar, N., Balasubramonian, R. and Jeon, Y., 2015, October. Improving memristor memory with sneak current sharing. In *2015 33rd IEEE International Conference on Computer Design (ICCD)* (pp. 549-556). IEEE.

[98] Bae, W., Yoon, K.J., Hwang, C.S. and Jeong, D.K., 2016. A crossbar resistance switching memory readout scheme with sneak current cancellation based on a two-port current-mode sensing. *Nanotechnology*, *27*(48), p.485201.

[99] Alibart, F., Zamanidoost, E. and Strukov, D.B., 2013. Pattern classification by memristive crossbar circuits using ex situ and in situ training. *Nature communications*, *4*(1), pp.1-7.

[100] Prezioso, M., Merrikh-Bayat, F., Hoskins, B.D., Adam, G.C., Likharev, K.K. and Strukov, D.B., 2015. Training and operation of an integrated neuromorphic network based on metal-oxide memristors. *Nature*, *521*(7550), pp.61-64.

[101] Yoo, H.J., 2019, February. 1.2 intelligence on silicon: From deep-neural-network accelerators to brain mimicking AI-SoCs. In *IEEE ISSCC Dig. Tech. Papers* (pp. 20-26). IEEE.

[102] Xue, C.X., Chen, W.H., Liu, J.S., Li, J.F., Lin, W.Y., Lin, W.E., Wang, J.H., Wei, W.C., Chang, T.W., Chang, T.C. and Huang, T.Y., 2019, February. 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 *IEEE ISSCC Dig. Tech. Papers* (pp. 388-390). IEEE.

[103] Kim, K.M. and Williams, R.S., 2019. A family of stateful memristor gates for complete cascading logic. *IEEE Trans. Circuits Syst. I, Reg. Papers*, *66*(11), pp.4348-4355.

[104] Yoon, K.J., Han, J.W. and Bae, W., 2020. A novel stateful logic device and circuit for in-memory parity programming in crossbar memory. *Advanced Electronic Materials*, p. 202000672

[105] Wang, Z., Li, C., Lin, P., Rao, M., Nie, Y., Song, W., Qiu, Q., Li, Y., Yan, P., Strachan, J.P. and Ge, N., 2019. In situ training of feed-forward and recurrent convolutional memristor networks. *Nature Machine Intelligence*, *1*(9), pp.434-442.

[106] Bae, W., 2020. Supply-Scalable High-Speed I/O Interfaces. *Electronics*, *9*(8), p.1315.

[107] Chang, K.Y., Wei, J., Huang, C., Li, S., Donnelly, K., Horowitz, M., Li, Y. and Sidiropoulos, S., 2003. A 0.4-4-Gb/s CMOS quad transceiver cell using on-chip regulated dual-loop PLLs. *IEEE J. Solid-State Circuits*, *38*(5), pp.747-754.

[108] Mooney, R., Yeung, E., Kennedy, J., Canagasaby, K., Mansuri, M., O'Mahony, F., Jaussi, J. and Casper, B., 2006, February. A 20Gb/s embedded clock transceiver in 90nm CMOS. In *IEEE ISSCC Dig. Tech. Papers* (pp. 1334-1343). IEEE.

[109] Bulzacchelli, J.F., Meghelli, M., Rylov, S.V., Rhee, W., Rylyakov, A.V., Ainspan, H.A., Parker, B.D., Beakes, M.P., Chung, A., Beukema, T.J. and Pepeljugoski, P.K., 2006. A 10-Gb/s 5-tap DFE/4-tap FFE transceiver in 90-nm CMOS technology. *IEEE J. Solid-State Circuits*, *41*(12), pp.2885-2900.

[110] Tamura, H., Kibune, M., Takahashi, Y., Doi, Y., Chiba, T., Higashi, H., Takauchi, H., Ishida, H. and Gotoh, K., 2001, February. 5 Gb/s bidirectional balanced-line link compliant with plesiochronous clocking. In *IEEE ISSCC Dig. Tech. Papers* (pp. 64-65). IEEE.

[111] Haycock, M. and Mooney, R., 2001, February. 3.2 GHz 6.4 Gb/s per wire signaling in 0.18/spl mu/m CMOS. In *IEEE ISSCC Dig. Tech. Papers* (pp. 62-63). IEEE.

[112] Tanaka, K., Fukaishi, M., Takeuchi, M., Yoshida, N., Minami, K., Yamaguchi, K., Uchida, H., Morishita, Y., Sakamoto, T., Kaneko, T. and Soda, M., 2002, February. A 100 Gb/s transceiver with GND-VDD common-mode receiver and flexible multi-channel aligner. In *IEEE ISSCC Dig. Tech. Papers* (pp. 264-465). IEEE.

[113] Lee, B.J., Hwang, M.S., Lee, S.H. and Jeong, D.K., 2003. A 2.5-10-Gb/s CMOS transceiver with alternating edge-sampling phase detection for loop characteristic stabilization. *IEEE J. Solid-State Circuits*, *38*(11), pp.1821-1829.

[114] Lee, H.R., Hwang, M.S., Lee, B.J., Kim, Y.D., Oh, D., Kim, J., Lee, S.H., Jeong, D.K. and Kim, W., 2004, February. A fully integrated 0.13/spl mu/m CMOS 10

Gb Ethernet transceiver with XAUI interface. In *IEEE ISSCC Dig. Tech. Papers* (pp. 170-520). IEEE.

[115] Krishna, K., Yokoyama-Martin, D.A., Wolfer, S., Jones, C., Loikkanen, M., Parker, J., Segelken, R., Sonntag, J.L., Stonick, J., Titus, S. and Weinlader, D., 2005, February. A 0.6 to 9.6 Gb/s binary backplane transceiver core in 0.13/spl mu/m CMOS. In *IEEE ISSCC Dig. Tech. Papers* (pp. 64-585). IEEE.

[116] Landman, P., Brouse, K., Gupta, V., Wu, S., Payne, R., Erdogan, U., Gu, R., Yee, A.L., Parthasarathy, B., Ramaswamy, S. and Bhakta, B., 2005, February. A transmit architecture with 4-tap feedforward equalization for 6.25/12.5 Gb/s serial backplane communications. In *IEEE ISSCC Dig. Tech. Papers* (pp. 66-585). IEEE.

[117] Casper, B., Jaussi, J., O'Mahony, F., Mansuri, M., Canagasaby, K., Kennedy, J., Yeung, E. and Mooney, R., 2006, February. A 20Gb/s Forwarded Clock Transceiver in 90nm CMOS B. In *IEEE ISSCC Dig. Tech. Papers* (pp. 263-272). IEEE.

[118] Palermo, S., Emami-Neyestanak, A. and Horowitz, M., 2008. A 90 nm CMOS 16 Gb/s transceiver for optical interconnects. *IEEE J. Solid-State Circuits*, *43*(5), pp.1235-1246.

[119] Kim, J.K., Kim, J., Kim, G., Chi, H. and Jeong, D.K., 2008, June. A 40-Gb/s transceiver in 0.13-μm CMOS technology. In *Proc. Symp. VLSI Circuits* (pp. 196-197). IEEE.

[120] Lee, J., Chen, M.S. and Wang, H.D., 2008, February. A 20Gb/s duobinary transceiver in 90nm CMOS. In *IEEE ISSCC Dig. Tech. Papers* (pp. 102-599). IEEE.

[121] Amamiya, Y., Kaeriyama, S., Noguchi, H., Yamazaki, Z., Yamase, T., Hosoya, K., Okamoto, M., Tomari, S., Yamaguchi, H., Shoda, H. and Ikeda, H., 2009, February. A 40Gb/s multi-data-rate CMOS transceiver chipset with SFI-5 interface for optical transmission systems. In *IEEE ISSCC Dig. Tech. Papers* (pp. 358-359). IEEE.

[122] Chen, M.S., Shih, Y.N., Lin, C.L., Hung, H.W. and Lee, J., 2011, February. A 40Gb/s TX and RX chip set in 65nm CMOS. In *IEEE ISSCC Dig. Tech. Papers* (pp. 146-148). IEEE.

[123] Takemoto, T., Yamashita, H., Kamimura, T., Yuki, F., Masuda, N., Toyoda, H., Chujo, N., Kogo, K., Lee, Y., Tsuji, S. and Nishimura, S., 2012, June. 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 *Proc. Symp. VLSI Circuits* (pp. 106-107). IEEE.

[124] Raghavan, B., Cui, D., Singh, U., Maarefi, H., Pi, D., Vasani, A., Huang, Z.C., Çatlı, B., Momtaz, A. and 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 J. Solid-State Circuits*, *48*(12), pp.3219-3228.

[125] Navid, R., Chen, E.H., Hossain, M., Leibowitz, B., Ren, J., Chou, C.H.A., Daly, B., Aleksić, M., Su, B., Li, S. and Shirasgaonkar, M., 2014. A 40 Gb/s serial link transceiver in 28 nm CMOS technology. *IEEE J. Solid-State Circuits*, *50*(4), pp.814-827.

[126] Zhang, B., Khanoyan, K., Hatamkhani, H., Tong, H., Hu, K., Fallahi, S., Abdul-Latif, M., Vakilian, K., Fujimori, I. and Brewster, A., 2015. A 28 Gb/s multistandard serial link transceiver for backplane applications in 28 nm CMOS. *IEEE J. Solid-State Circuits*, *50*(12), pp.3089-3100.

[127] Upadhyaya, P., Savoj, J., An, F.T., Bekele, A., Jose, A., Xu, B., Wu, D., Turker, D., Aslanzadeh, H., Hedayati, H. and Im, J., 2015, February. 3.3 A 0.5-to-32.75 Gb/s flexible-reach wireline transceiver in 20nm CMOS. In *IEEE ISSCC Dig. Tech. Papers* (pp. 1-3). IEEE.

[128] Norimatsu, T., Kawamoto, T., Kogo, K., Kohmu, N., Yuki, F., Nakajima, N., Muto, T., Nasu, J., Komori, T., Koba, H. and Usugi, T., 2016, February. 3.3 A 25Gb/s multistandard serial link transceiver for 50dB-loss copper cable in 28nm CMOS. In *IEEE ISSCC Dig. Tech. Papers* (pp. 60-61). IEEE.

[129] Gopalakrishnan, K., Ren, A., Tan, A., Farhood, A., Tiruvur, A., Helal, B., Loi, C.F., Jiang, C., Cirit, H., Quek, I. and Riani, J., 2016, January. 3.4 A 40/50/100Gb/s PAM-4 Ethernet transceiver in 28nm CMOS. In *IEEE ISSCC Dig. Tech. Papers* (pp. 62-63). IEEE.

[130] Shibasaki, T., Danjo, T., Ogata, Y., Sakai, Y., Miyaoka, H., Terasawa, F., Kudo, M., Kano, H., Matsuda, A., Kawai, S. and Arai, T., 2016, February. 3.5 A 56Gb/s NRZ-electrical 247mW/lane serial-link transceiver in 28nm CMOS. In *IEEE ISSCC Dig. Tech. Papers* (pp. 64-65). IEEE.

[131] Peng, P.J., Li, J.F., Chen, L.Y. and Lee, J., 2017, February. 6.1 a 56Gb/s PAM-4/NRZ transceiver in 40nm CMOS. In *IEEE ISSCC Dig. Tech. Papers* (pp. 110-111). IEEE.

[132] Han, J., Lu, Y., Sutardja, N. and Alon, E., 2017, February. 6.2 A 60Gb/s 288mW NRZ transceiver with adaptive equalization and baud-rate clock and data recovery in 65nm CMOS technology. In *IEEE ISSCC Dig. Tech. Papers* (pp. 112-113). IEEE.

[133] Upadhyaya, P., Poon, C.F., Lim, S.W., Cho, J., Roldan, A., Zhang, W., Namkoong, J., Pham, T., Xu, B., Lin, W. and Zhang, H., 2018, February. A fully adaptive 19-to-56Gb/s PAM-4 wireline transceiver with a configurable ADC in 16nm FinFET. In *IEEE ISSCC Dig. Tech. Papers* (pp. 108-110). IEEE.

[134] Wang, L., Fu, Y., LaCroix, M.A., Chong, E. and Carusone, A.C., 2018. A 64-Gb/s 4-PAM transceiver utilizing an adaptive threshold ADC in 16-nm FinFET. *IEEE J. Solid-State Circuits*, *54*(2), pp.452-462.

[135] Depaoli, E., Monaco, E., Steffan, G., Mazzini, M., Zhang, H., Audoglio, W., Belotti, O., Rossi, A.A., Albasini, G., Pozzoni, M. and Erba, S., 2018, February. A 4.9 pJ/b 16-to-64Gb/s PAM-4 VSR transceiver in 28nm FDSOI CMOS. In *IEEE ISSCC Dig. Tech. Papers* (pp. 112-114). IEEE.

[136] Tang, L., Gai, W., Shi, L., Xiang, X., Sheng, K. and He, A., 2018, February. A 32Gb/s 133mW PAM-4 transceiver with DFE based on adaptive clock phase and threshold voltage in 65nm CMOS. In *IEEE ISSCC Dig. Tech. Papers* (pp. 114-116). IEEE.

[137] LaCroix, M.A., Wong, H., Liu, Y.H., Ho, H., Lebedev, S., Krotnev, P., Nicolescu, D.A., Petrov, D., Carvalho, C., Alie, S. and Chong, E., 2019, February. 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 *IEEE ISSCC Dig. Tech. Papers* (pp. 114-116). IEEE.

[138] Pisati, M., De Bernardinis, F., Pascale, P., Nani, C., Sosio, M., Pozzati, E., Ghittori, N., Magni, F., Garampazzi, M., Bollati, G. and Milani, A., 2019, February. 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 *IEEE ISSCC Dig. Tech. Papers* (pp. 116-118). IEEE.

[139] Ali, T., Yousry, R., Park, H., Chen, E., Weng, P.S., Huang, Y.C., Liu, C.C., Wu, C.H., Huang, S.H., Lin, C. and Wu, K.C., 2019, February. 6.4 A 180mW 56Gb/s DSP-Based Transceiver for High Density IOs in Data Center Switches in 7nm FinFET Technology. In *IEEE ISSCC Dig. Tech. Papers* (pp. 118-120). IEEE.

[140] Im, J., Zheng, K., Chou, A., Zhou, L., Kim, J.W., Chen, S., Wang, Y., Hung, H.W., Tan, K., Lin, W. and Roldan, A., 2020, February. 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 *IEEE ISSCC Dig. Tech. Papers* (pp. 116-118). IEEE.

[141] Ali, T., Chen, E., Park, H., Yousry, R., Ying, Y.M., Abdullatif, M., Gandara, M., Liu, C.C., Weng, P.S., Chen, H.S. and Elbadry, M., 2020, February. 6.2 A 460mW 112Gb/s DSP-Based Transceiver with 38dB Loss Compensation for Next-Generation Data Centers in 7nm FinFET Technology. In *IEEE ISSCC Dig. Tech. Papers* (pp. 118-120). IEEE.

[142] Yoo, B.J., Lim, D.H., Pang, H., Lee, J.H., Baek, S.Y., Kim, N., Choi, D.H., Choi, Y.H., Yang, H., Yoon, T. and Chu, S.H., 2020, February. 6.4 A 56Gb/s 7.7 mW/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 *IEEE ISSCC Dig. Tech. Papers* (pp. 122-124). IEEE.

[143] Shannon, C.E., 1948. A mathematical theory of communication. *The Bell system technical journal*, *27*(3), pp.379-423.

[144] Li, H., Chen, S., Yang, L., Bai, R., Hu, W., Zhong, F.Y., Palermo, S. and Chiang, P.Y., 2014, June. 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 *Proc. Symp. VLSI Circuits* (pp. 1-2). IEEE.

[145] Ragab, A., Liu, Y., Hu, K., Chiang, P. and Palermo, S., 2011. Receiver jitter tracking characteristics in high-speed source synchronous links. *Journal of Electrical and Computer Engineering*, *2011*.

[146] Casper, B. and O'Mahony, F., 2009. Clocking analysis, implementation and measurement techniques for high-speed data links—A tutorial. *IEEE Trans. Circuits Syst. I, Reg. Papers*, *56*(1), pp.17-39.

[147] Hossain, M. and Carusone, A.C., 2011. 7.4 Gb/s 6.8 mW source synchronous receiver in 65 nm CMOS. *IEEE J. Solid-State Circuits*, *46*(6), pp.1337-1348.

[148] Chung, S.H. and Kim, L.S., 2012, June. 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 *Proc. Symp. VLSI Circuits* (pp. 144-145). IEEE.

[149] Bae, W., Ju, H., Park, K., Cho, S.Y. and Jeong, D.K., 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 J. Solid-State Circuits*, *51*(10), pp.2357-2367.

[150] Bae, W., Jeong, G.S., Park, K., Cho, S.Y., Kim, Y. and Jeong, D.K., 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 Trans. Circuits Syst. I, Reg. Papers*, *63*(9), pp.1393-1403.

[151] Harwood, M., Warke, N., Simpson, R., Leslie, T., Amerasekera, A., Batty, S., Colman, D., Carr, E., Gopinathan, V., Hubbins, S. and Hunt, P., 2007, February. A 12.5 Gb/s SerDes in 65nm CMOS using a baud-rate ADC with digital receiver equalization and clock recovery. In *IEEE ISSCC Dig. Tech. Papers* (pp. 436-591). IEEE.

[152] Chen, E.H. and Yang, C.K.K., 2010. ADC-based serial I/O receivers. *IEEE Trans. Circuits Syst. I, Reg. Papers*, *57*(9), pp.2248-2258.

[153] Wang, L., Fu, Y., LaCroix, M.A., Chong, E. and Carusone, A.C., 2018. A 64-Gb/s 4-PAM transceiver utilizing an adaptive threshold ADC in 16-nm FinFET. *IEEE J. Solid-State Circuits*, *54*(2), pp.452-462.

[154] Palermo, S., Hoyos, S., Cai, S., Kiran, S. and Zhu, Y., 2018. Analog-to-digital converter-based serial links: an overview. *IEEE Solid-State Circuits Magazine*, *10*(3), pp.35-47.

[155] Miller, D.A., 2000. Rationale and challenges for optical interconnects to electronic chips. *Proceedings of the IEEE*, *88*(6), pp.728-749.

[156] Young, I.A., Mohammed, E., Liao, J.T., Kern, A.M., Palermo, S., Block, B.A., Reshotko, M.R. and Chang, P.L., 2009. Optical I/O technology for tera-scale computing. *IEEE J. Solid-State Circuits*, *45*(1), pp.235-248.

[157] Jeong, G.S., Bae, W. and Jeong, D.K., 2017. Review of CMOS integrated circuit technologies for high-speed photo-detection. *Sensors*, *17*(9), p.1962.

[158] Thraskias, C.A., Lallas, E.N., Neumann, N., Schares, L., Offrein, B.J., Henker, R., Plettemeier, D., Ellinger, F., Leuthold, J. and Tomkos, I., 2018. Survey of photonic and plasmonic interconnect technologies for intra-datacenter and high-performance computing communications. *IEEE Communications Surveys & Tutorials*, *20*(4), pp.2758-2783.

[159] Cho, H., Kapur, P. and Saraswat, K.C., 2004. Power comparison between high-speed electrical and optical interconnects for interchip communication. *Journal of lightwave technology*, *22*(9), p.2021.

[160] Sun, C, et al., 2020, June. TeraPHY™: An O-Band WDM Electro-Optic Platform for Low Power, Terabit/s Optical I/O. In *Proc. Symp. VLSI Technology* (pp. 1-2). IEEE

[161] Sun, C., Wade, M.T., Lee, Y., Orcutt, J.S., Alloatti, L., Georgas, M.S., Waterman, A.S., Shainline, J.M., Avizienis, R.R., Lin, S. and Moss, B.R., 2015. Single-chip microprocessor that communicates directly using light. *Nature*, *528*(7583), pp.534-538.

[162] Narasimha, A., Analui, B., Liang, Y., Sleboda, T.J., Abdalla, S., Balmater, E., Gloeckner, S., Guckenberger, D., Harrison, M., Koumans, R.G. and Kucharski, D., 2007. A Fully Integrated 4$\times $10-Gb/s DWDM Optoelectronic Transceiver Implemented in a Standard 0.13um CMOS SOI Technology. *IEEE J. Solid-State Circuits*, *42*(12), pp.2736-2744.

[163] Sun, C., Georgas, M., Orcutt, J., Moss, B., Chen, Y.H., Shainline, J., Wade, M., Mehta, K., Nammari, K., Timurdogan, E. and Miller, D., 2015. A monolithically-integrated chip-to-chip optical link in bulk CMOS. *IEEE J. Solid-State Circuits*, *50*(4), pp.828-844.

[164] Li, H., Balamurugan, G., Kim, T., Sakib, M.N., Kumar, R., Rong, H., Jaussi, J. and Casper, B., 2020. A 3-D-Integrated Silicon Photonic Microring-Based 112-Gb/s PAM-4 Transmitter With Nonlinear Equalization and Thermal Control. *IEEE J. Solid-State Circuits*.

Figure 1

Amount of digital data creation trend and projection (source: Cisco [3], [4])

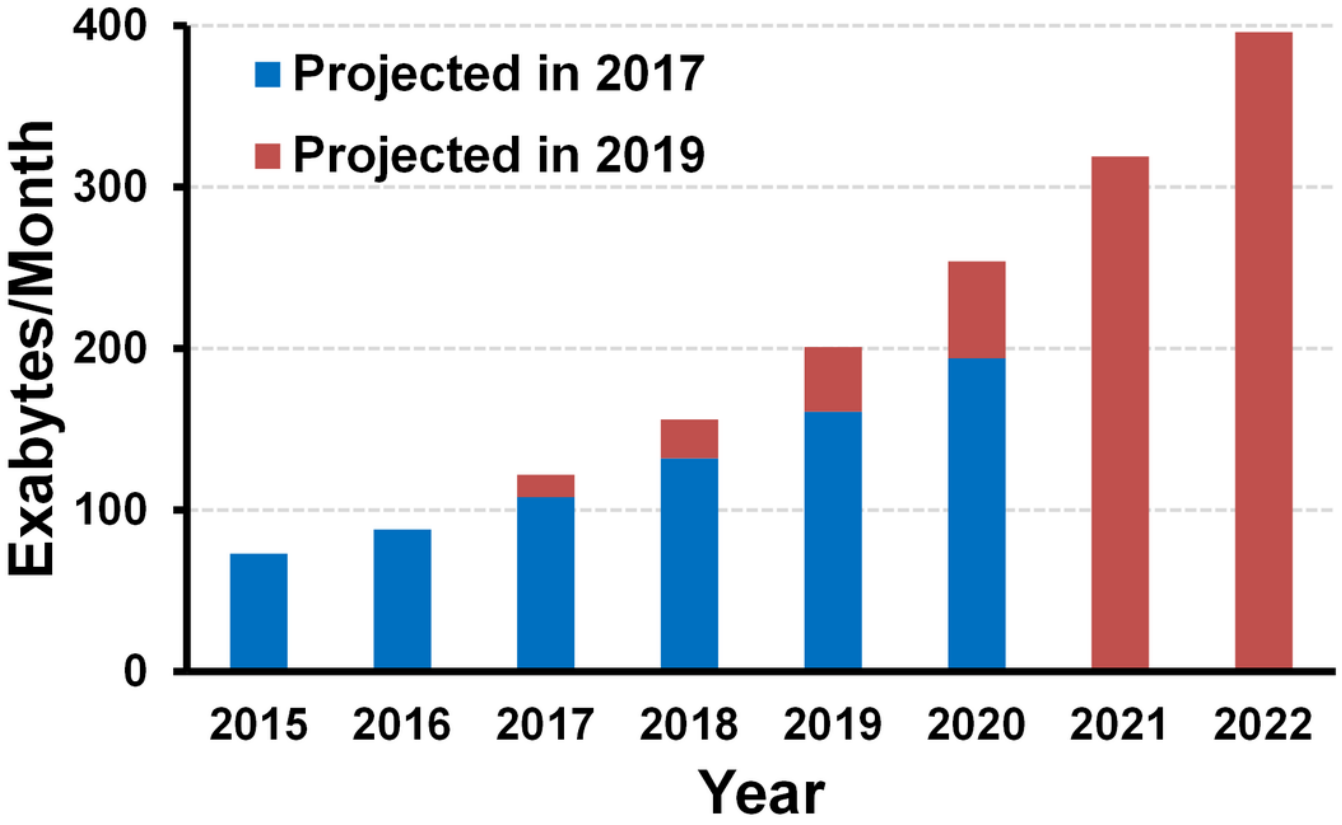

Figure 2

Summary of how we have made a better computer and why that will not work for future computers

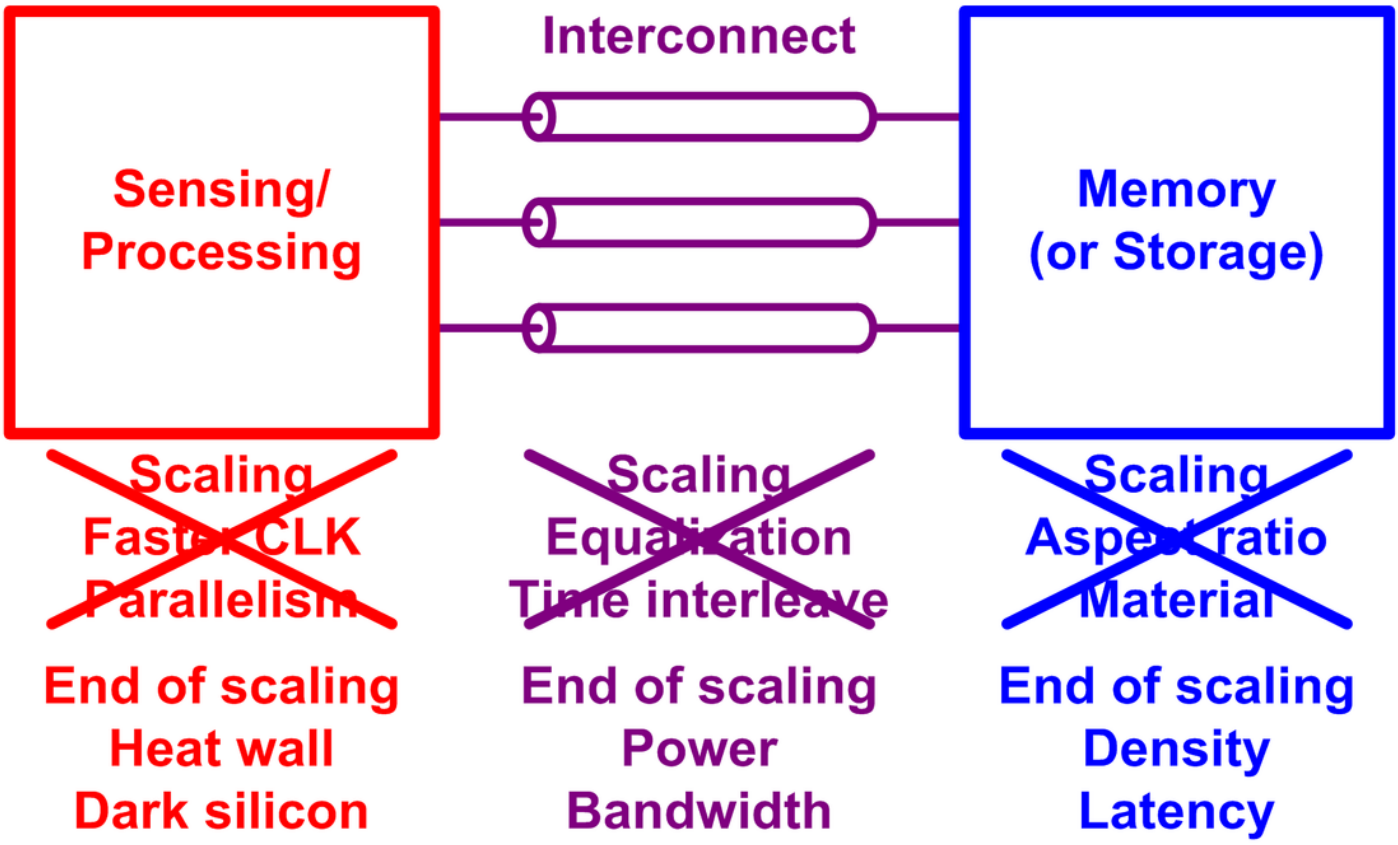

Figure 3

Survey of energy efficiency of various cryptocurrency miners

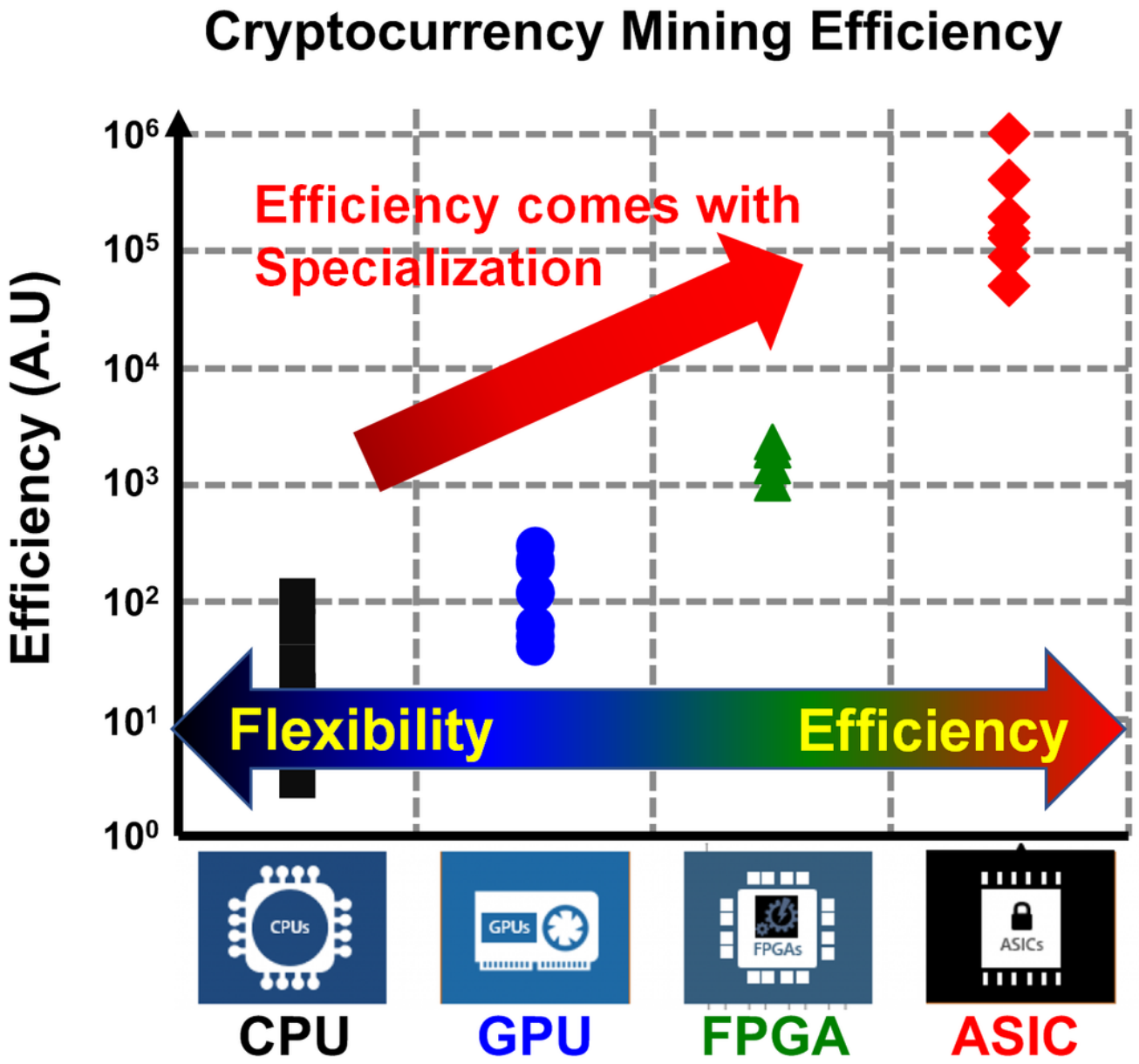

Figure 4

Productivity issue of specialization

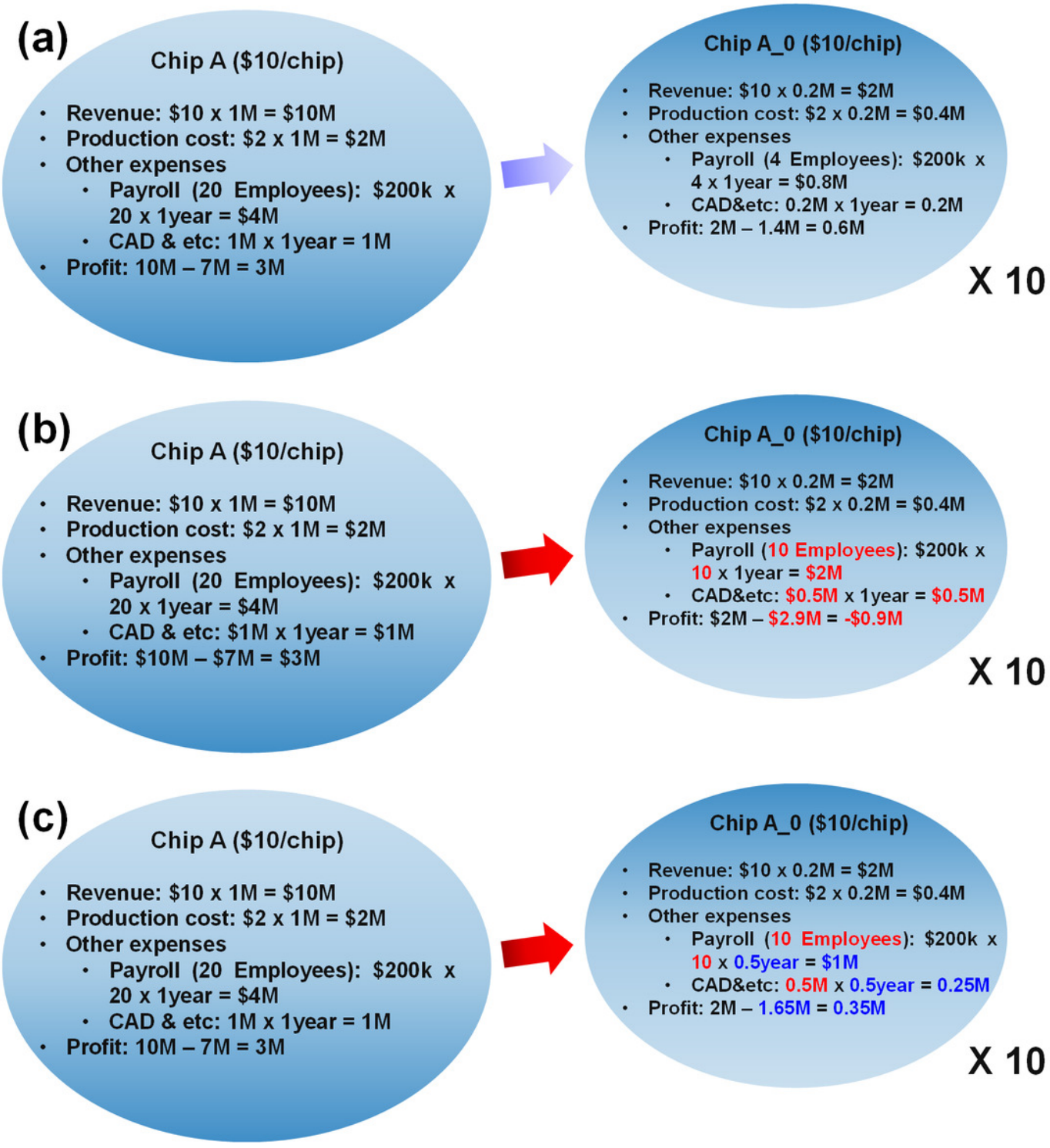

Figure 5

(a) Silicon design complexity across technology node, (b) general design flow of AMS circuit

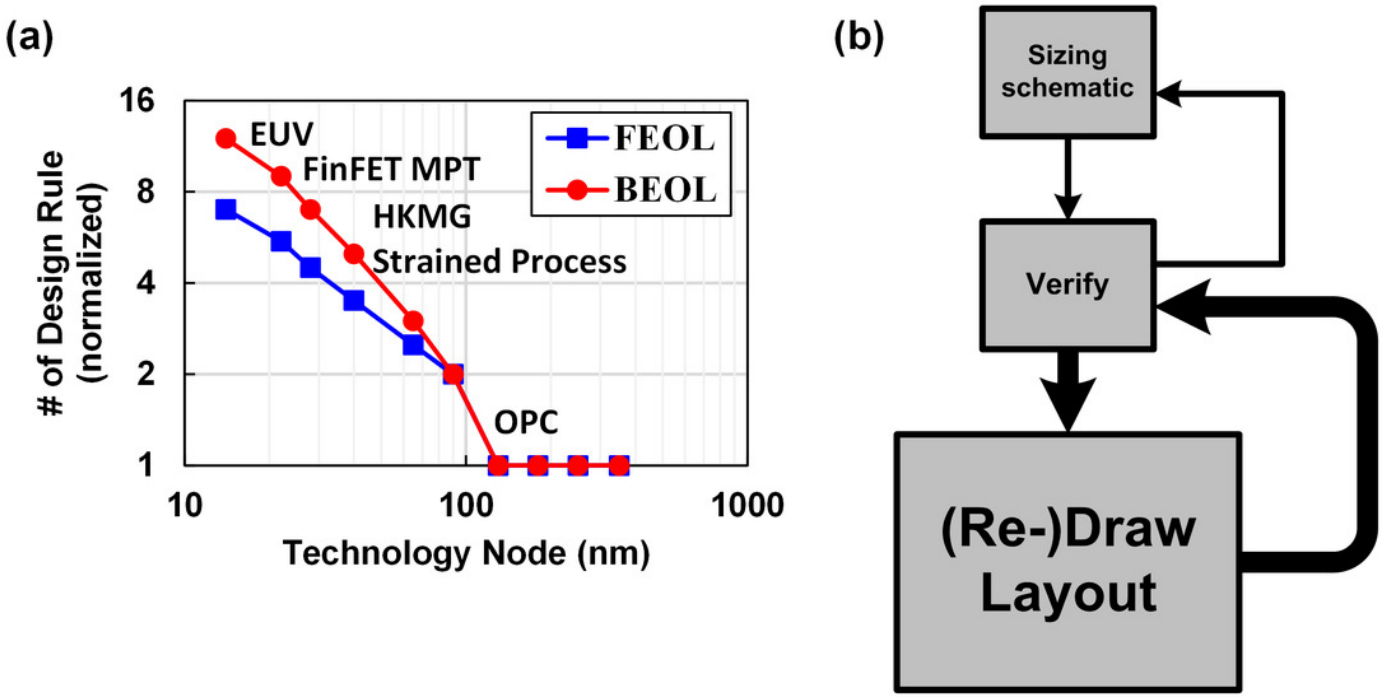

Figure 6

Design flow based on reuse of design process using executable script (generator)

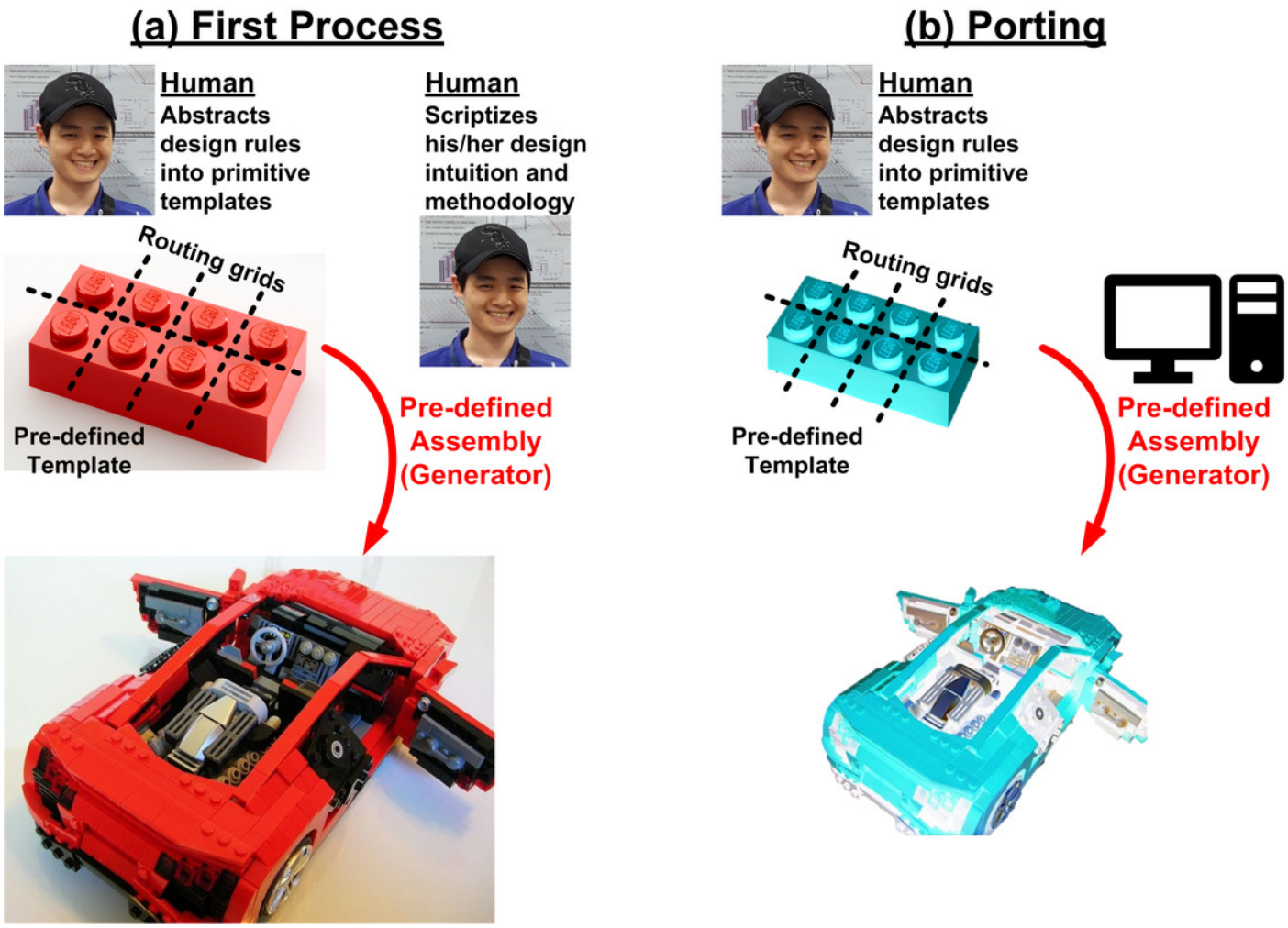

Figure 7

Comparison of 3 different frameworks to support reusing process

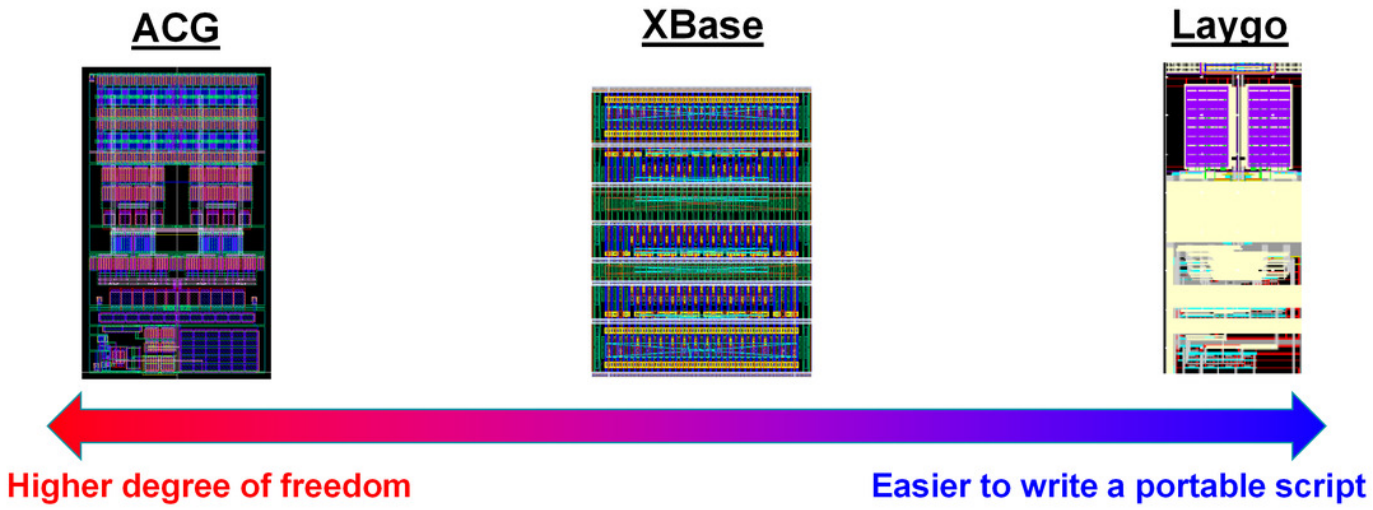

Figure 8

Conceptual diagram of (a) HBM and (b) V-NAND

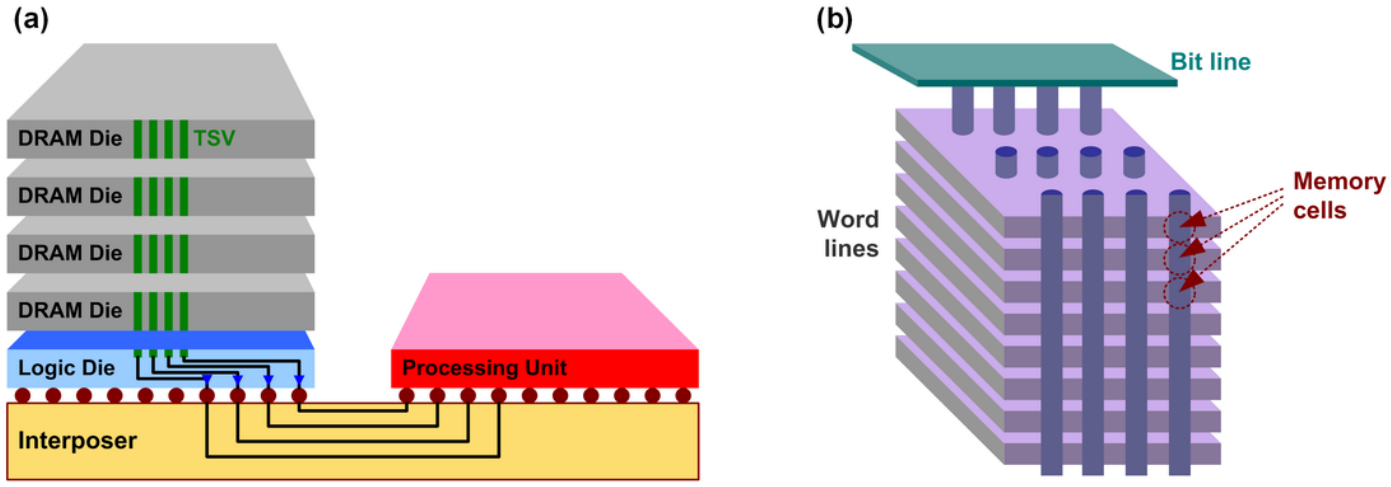

Figure 9

New memory device with crossbar array structure

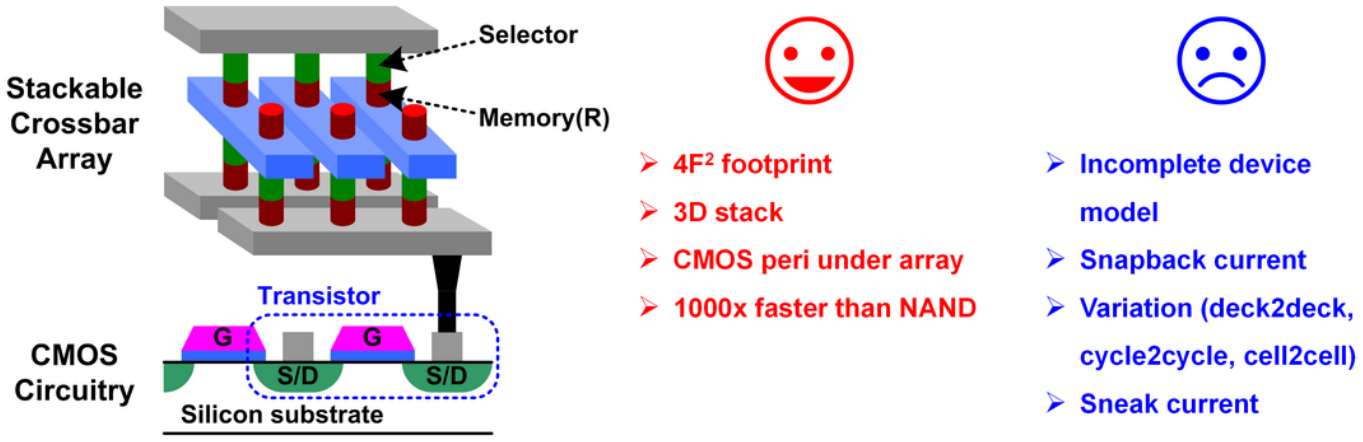

Figure 10

Block diagram of general interconnect architecture

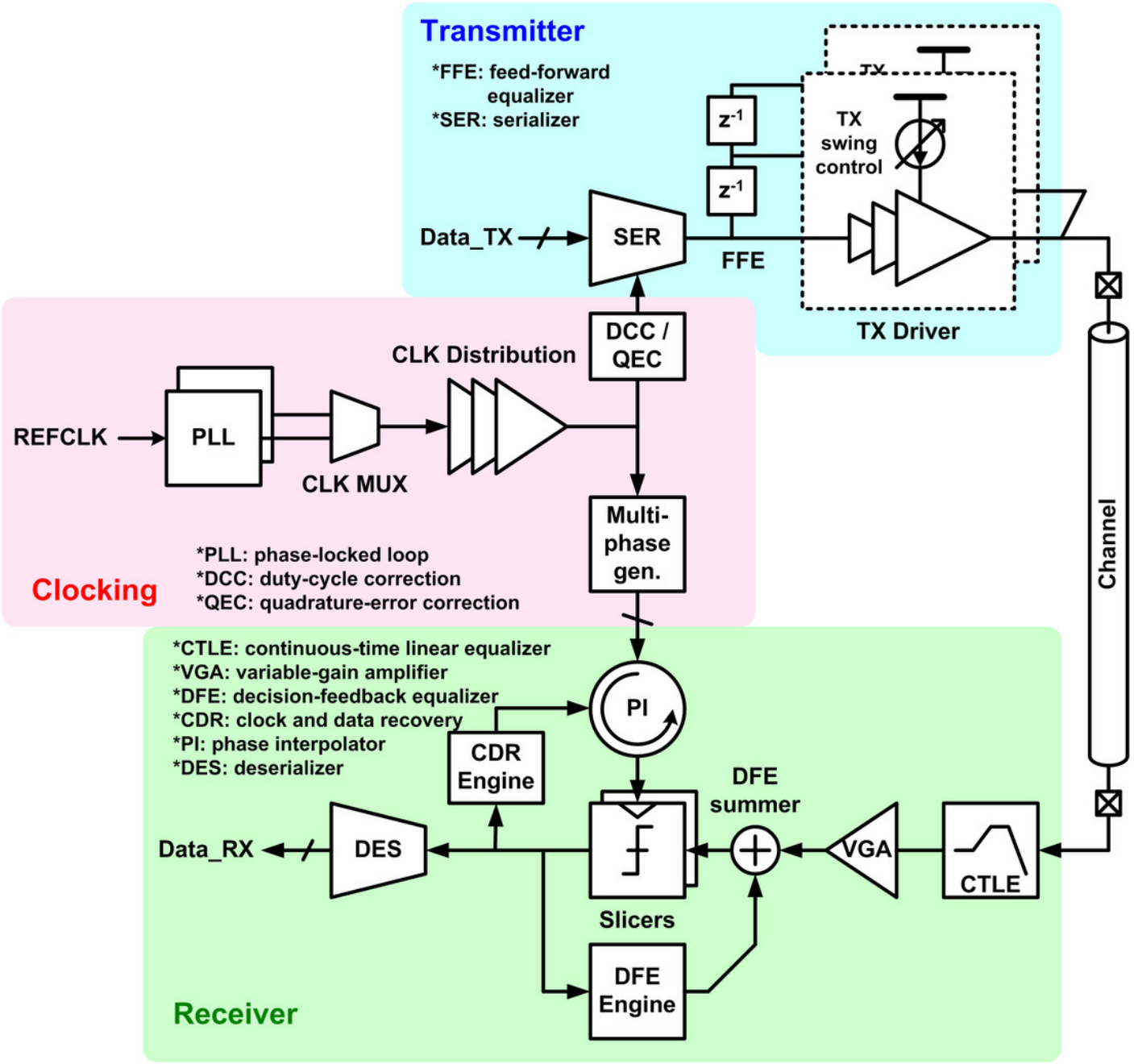

Figure 11

Survey and trend of interconnects with respect to (a) technology nodes (b) published years

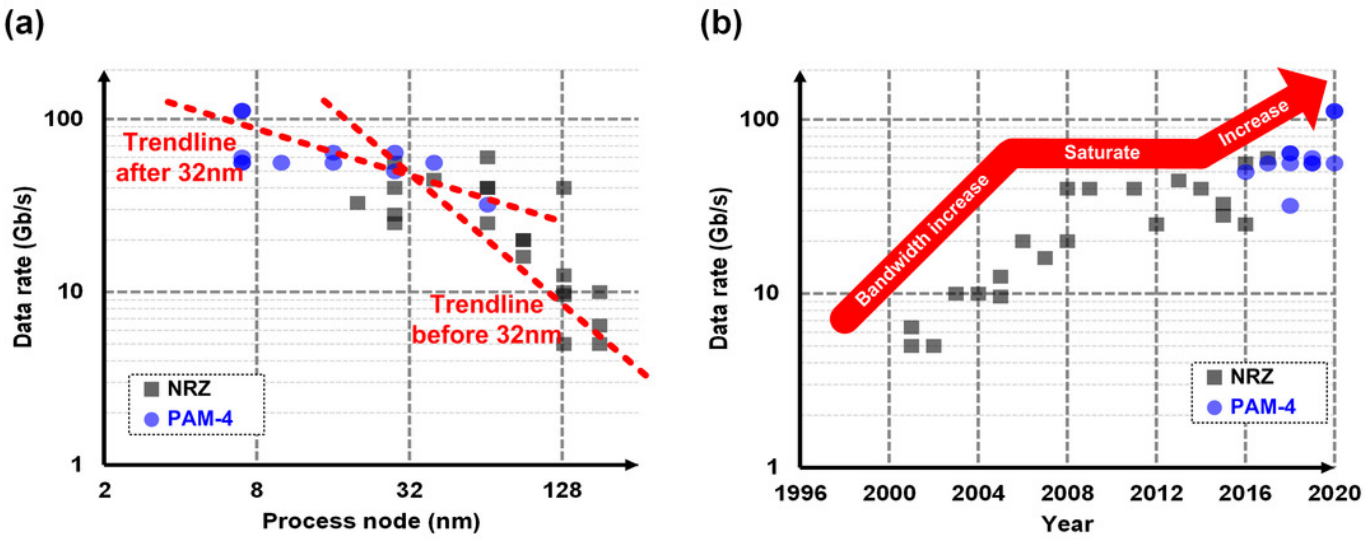

Figure 12

Comparison of NRZ, PAM-4, PAM-8, and PAM-16

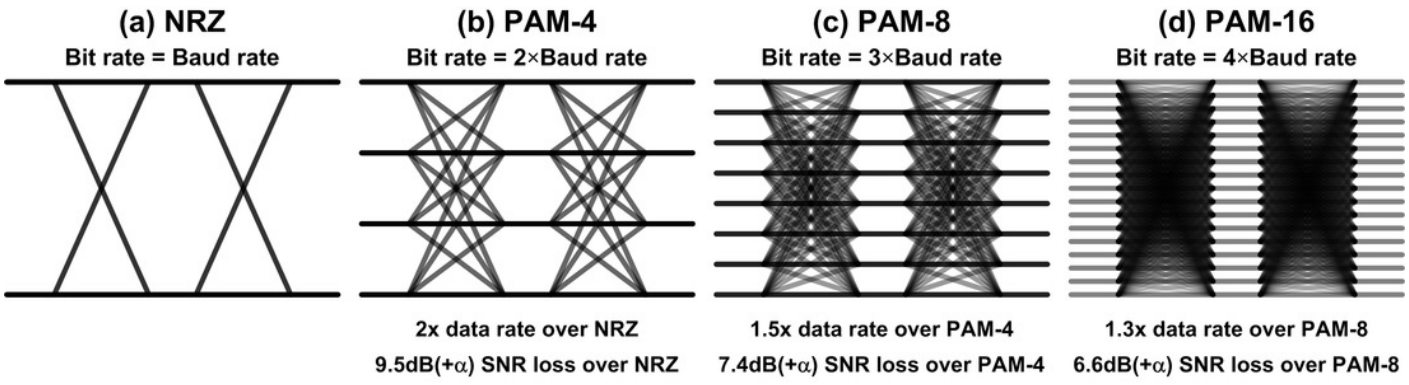

Figure 13

Power comparison between electrical and optical interconnects and definition of critical length