Department: Head
Editor: Name, xxxx@email

Interconnects for DNA, Quantum, In-Memory and Optical Computing: Insights from a Panel Discussion

Amlan Ganguly
Rochester Institute of Technology, Rochester, NY, USA

Sergi Abadal
Universitat Politècnica de Catalunya, Barcelona, Spain

Ishan Thakkar
University of Kentucky, Lexington, KY, USA

Natalie Enright Jerger
University of Toronto, Toronto, ON, Canada

Marc Riedel
University of Minnesota, Minneapolis, MN, USA

Masoud Babaie
Delft University of Technology, Delft, Netherlands

Rajeev Balasubramonian
University of Utah, Salt Lake City, UT, USA

Abu Sebastian
IBM Research, Zurich, Switzerland

Sudeep Pasricha
Colorado State University, Fort Collins, CO, USA

Baris Taskin
Drexel University, Philadelphia, PA, USA

ABSTRACT: ***The computing world is witnessing a proverbial Cambrian explosion of emerging paradigms propelled by applications such as Artificial Intelligence, Big Data and Cy-***

 Published by the IEEE Computer Society

bersecurity. The recent advances in technology to store digital data inside a DNA strand, manipulate quantum bits (qubits), perform logical operations with photons and perform computations inside memory systems are ushering in the era of emerging paradigms of DNA computing, quantum computing, optical computing and in-memory computing. In an orthogonal direction, research on interconnect design using advanced electro-optic, wireless and microfluidic technologies has shown promising solutions to the architectural limitations of traditional von-Neumann computers. In this article, experts present their comments on the role of interconnects in the emerging computing paradigms, and discuss the potential use of chiplet-based architectures for the heterogeneous integration of such technologies.

Keywords:

DNA Computing; Optical Computing; Quantum Computing; In-Memory Computing; Photonic Interconnects; Wireless Interconnects

INTRODUCTION

Moore's law has conventionally enabled increasing integration; however, fundamental physical limitations have slowed the rate of transition from one technology node to the next, and the costs of new fabrication facilities are exponentially increasing. On the other hand, modern workloads, such as machine learning, have a seemingly insatiable appetite for more compute and memory bandwidth. These workloads also demand extreme-scale computational energy efficiency, which cannot be fulfilled by using traditional, CMOS-implemented, von Neumann computing systems. Therefore, to meet these computational demands, workload-specific accelerator chips that are implemented using emerging, beyond-Moore computing paradigms have garnered an increased attention. In particular, workload-specific chips based on computing paradigms such as Dioxyribonucleic Acid (DNA) computing and storage, quantum computing, optical computing and in-memory computing have been shown to provide disruptive benefits.

We envision that as the fabrication techniques for implementing accelerator chips based on novel computing paradigms mature, it will be possible to integrate discrete chips (or chiplets, as they are often called) from disparate technology domains and computing paradigms to create chiplet-based heterogeneous systems. For instance, Figure 1 illustrates our visualization of such a heterogeneous system using an interposer as the integration platform. The interposer is essentially a large die that has minimal logic but abundant wiring resources [1], which can be utilized to provide very high bandwidth connections between chiplets (that can be integrated on the interposer as a socket). This type of interposer-integrated, chiplet-based heterogeneous computing systems provide benefits such as design flexibility to integrate multiple computing paradigms with different emerging interconnect technologies such as wireless, silicon photonics and microfluidics, leading to improved computational capacity and energy efficiency, as delivered by the accelerators based on the aforementioned new computing paradigms.

Despite these benefits, however, such heterogeneous computing systems open up new research problems in interconnect design. In fact, because accelerator chiplets do not have components such as branch predictors and complex instruction decoders, data movement is an even larger bottleneck in them [2]. Therefore, interconnect innovations can have a much larger impact in such systems. Moreover, from the perspective of interposer-based integration of such heterogeneous systems, several research questions regarding the design and implementation of inter-chiplet interconnection networks remain open: What should be the target bandwidth, energy, and end-to-end latency for the networks? Which interconnect technologies can be utilized to achieve the performance targets of the networks? Should the networks be heterogeneous, requiring inter-domain and inter-technology conversion of data signals? What novel, modular topologies can tailor to different inter-chiplet communication patterns? What inter-technology compatibility requirements and critical design challenges should be addressed?

The above questions were discussed in a moderated panel in the thirteenth Workshop on Network-on-Chip Architectures (NoCArc 2020). In this paper, the speakers and moderators of the panel discuss various emerging computing

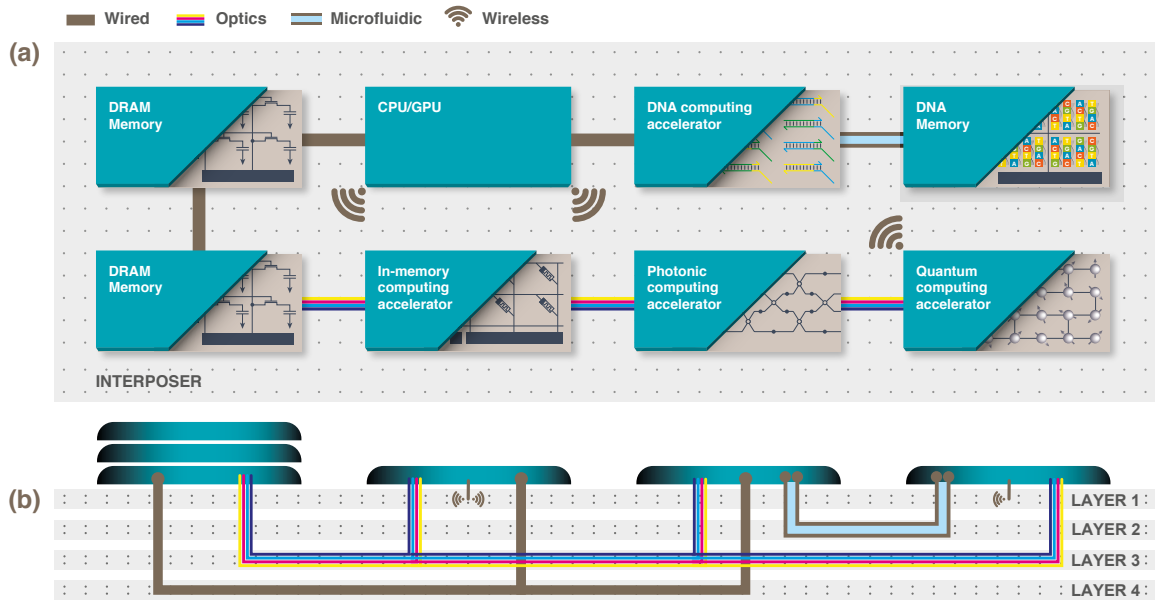


Figure 1. Interposer-based heterogeneous integration of multiple emerging computing paradigms through wired, optical, microfluidic and wireless interconnect technologies. (a) Top-view and (b) Side-view.

paradigms and their vision of a possible interconnection system that can enable the realization of heterogeneous systems with the different paradigms (Figure 1). We hope that these questions and our radically aggressive vision of a future heterogeneous computer will engender new research interest for the efficient design of inter-chiplet interconnection networks.

EMERGING COMPUTING PARADIGMS: PROMISES AND INTERCONNECTION NEEDS

Next, we discuss the performance promises and interconnection requirements of various emerging computing paradigms and architectures.

DNA STORAGE & COMPUTING

Ever since Watson and Crick first described the molecular structure of DNA, its information-bearing potential has been apparent to computer scientists. For decades, the idea of using DNA to store information for man-made computing systems was speculative and futuristic. Given its maturity, displacing conventional computer storage systems still seems far-fetched. And yet, spurred by the healthcare industry, the technology for both sequencing (reading) and synthesizing (writing) DNA has followed a Moore's law-like trajectory for the past 20 years. Sequencing 3 billion nucleotides in a human genome can be done for less than $1,000 dollars. Synthesizing a Megabyte of DNA data can be done in less than a day.

In a highly influential Science paper in 2012, the renowned Harvard genomicist George Church made the case for DNA storage purely based upon physical limits. He delineated the storage capacity of DNA: around 200 Petabytes per gram; the read-write speed: less than 100 microseconds per bit; and, most importantly, the power usage: as astonishingly little as 10^{-10} Watts per Gigabyte, hence orders of magnitude below the fJ/bit barrier targeted by other emerging technologies described below. Moreover, DNA is stable for decades, perhaps even millennia, at room temperature as DNA extraction from mammoths can attest. Therefore, DNA storage systems could outperform not only magnetic and electronic systems, but any realistic physical system that has been studied, and its chip integration could provide spectacular benefits in perpetual or hard-to-access systems.

With respect to computation, whereas electronic systems perform computation in terms of voltage, molecular systems perform computation in terms of molecular concentrations. Follow-

ing this principle, researchers have built DNA systems that perform many forms of computation: combinatorial search, signal processing, arithmetic, and more [3]. The goal is not the computation *per se*, but rather to construct the biological equivalent of embedded controllers: molecular systems engineered to perform useful computation where it is needed, for instance for drug delivery and for monitoring the effectiveness of drug therapy.

As DNA storage is coming online, DNA computing has reoriented itself in a way relevant to the scenario at hand. Instead of just storing data, there is the impetus to perform *in-memory* computation in order to avoid having to sequence (read) the DNA, compute in the digital domain, and then synthesize (write) the result back in the DNA memory. A particularly promising approach is to encode data by "nicking" DNA with editing enzymes such as PfAgo and CRISPR-Cas9 [4]. Data can be stored on potentially long DNA strands, divided into "registers", each storing a single bit. Nicks and denaturing create open toeholds in each register and, then, toehold-mediated strand displacement [3] is used to implement computation on the stored values. Bio-compatibility of DNA storage and compute devices is yet another benefit of this technology.

Interconnection needs: Perhaps the biggest challenge facing DNA computing and storage, in the long term, is interconnecting it with other forms of computation and storage. None of these systems will be plug-and-play. From handling liquid, to supplying chemical reagents, to disposing of waste, DNA systems will not only have different footprints, they will operate on very different signals (i.e., molecular/bio-chemical signals compared to electrical signals of traditional computers), and at different time scales (e.g., at hours scale compared to nanoseconds scale for traditional computers). However, the bio-compatibility of the DNA technologies and ultra high density, small footprint and low power mean that when successfully integrated with a heterogeneous system, as envisioned here, it can augment human experiences and memory in unprecedented ways. In the future world of augmented realities, DNA storage and computing can be predicted to play a key role.

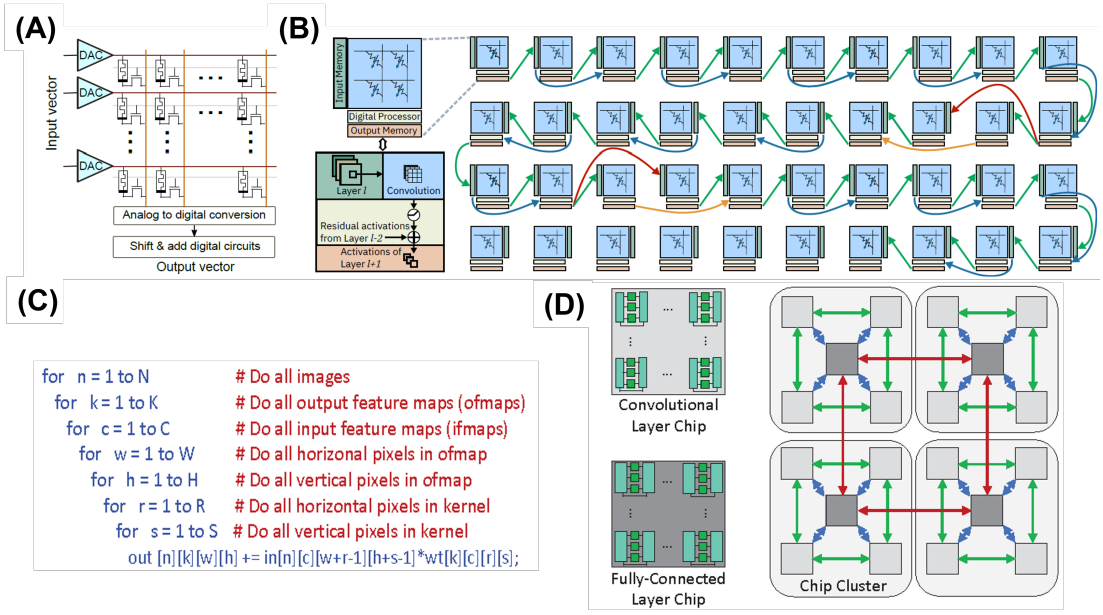
QUANTUM COMPUTING

Quantum computers (QCs) exploit the superposition and entanglement phenomena of individual particles to tackle classically intractable computational problems. Such a fundamental change of rules of computing is expected to lead to exponential speedups in critical tasks such as deep learning or combinatorial optimization, not achievable with any other technology.

A QC consists of a set of quantum bits (qubits) at extremely low temperatures to store/process the information, and a classical electronic system to control and read out the qubits' state. Currently, QCs contain tens of qubits that are connected to electronic modules at room temperature through hundreds of coaxial cables [5]. However, this approach is not scalable due to the sheer interconnect complexity and poor system reliability. Given that the simplest non-trivial algorithms require more than 100 logical qubits, scalability is thus a very important problem, which would possibly require electronics to be placed in close proximity with the qubits at a similar temperature. Another alternative to enhance scalability is to interconnect multiple QC cores in an analogy to classical multicore computers.

The reliability of existing qubit technologies is typically not sufficient to be used as computational qubits directly. It is thus indispensable to use quantum error correction algorithms wherein a *logical* qubit is encoded into several *physical* qubits. This correction loop demands a readout time much shorter than the qubit decoherence time. With current technologies, this implies reading out each qubit with 1 Mb/s rate [6].

Interconnection needs: Even considering an optimistic 0.1% error per qubit operation, a logical qubit will need 10,000 physical qubits to achieve an error rate of 10^{-15} [5]. Hence, error correction implies moving 10 Gb/s of classical data per logical qubit from the CPU to the QC. To host the minimum target of 100 logical qubits, an interconnect sustaining 1 Tb/s with end-to-end latency of less than 10 ns will be needed. Moreover, although room temperature qubits are being investigated, currently such high-speed interconnects must operate inside dilution fridges where the cooling power is extremely limited, enforcing a stringent target energy requirement



An example dataflow for a convolutional layer in a deep neural network. The loops can be re-ordered and each loop can be tiled.

A ScaleDeep System with 16 convolutional layer chips and 4 fully connected layer chips, organized as 4 chip clusters. The topology includes a wheel-and-spoke within a chip cluster and a ring of chip clusters.

Figure 2. **(A)** An example circuit performing in-memory dot-products. **(B)** Mapping of a neural network on an array of IMC cores. **(C)** An example dataflow. **(D)** Interconnect topologies in a ScaleDeep system.

of a few fJ/bit. Further, the support of quantum-coherent transfers of the qubits' states between multiple QC cores would be desirable, but it is not achievable with standard interconnects.

IN-MEMORY COMPUTING

In-memory computing (IMC) is an emerging paradigm where certain computational tasks are performed in the memory itself by exploiting the physical attributes of the memory devices, their array-level organization, the peripheral circuitry and the control logic. In recent years, both charge-based and resistance-based memory (memristive) devices have been employed for IMC (e.g., memristive crossbar based dot-product engine shown in Figure 2A). Although IMC has found applications in a wide range of areas (including high-precision scientific computing, low-precision stochastic computing, signal processing, and optimization), one of the most prominent application domains for IMC is deep learning. For example, to perform deep learning inference, the pre-trained synaptic weights are mapped to an array of IMC cores (Figure 2B) where each core performs the dot products corresponding to each layer. Over the past few years, not only memristor crossbar arrays, but also 1T1R/1T1C type of volatile and non-volatile RAMs (SRAM, DRAM, RRAM, PCRAM) have been utilized to demonstrate IMC.

Interconnection needs: Let us consider an example application domain of deep learning. The communication fabric needed to feed the IMC cores with data or to facilitate the efficient movement of activations from one IMC core to another plays an oversized role in both memristor-based and RAM-based implementations of IMC. Both are more amenable to highly pipelined dataflows, and this opens up new research directions. For example, an accelerator based on a communication fabric with a 5-parallel prism topology can achieve is a remarkable throughput of 40,000 Images/s for ResNet32 on CIFAR-10 (Figure 2B) [8]. Moreover, the dataflows of deep learning applications (e.g., Figure 2C) are also amenable to chiplet-based accelerator platforms (Figure 2D), as discussed in detail below. One challenge, however, is the need to overprovision the communication fabric. Wireless interconnects with their potential for a plastic topology and

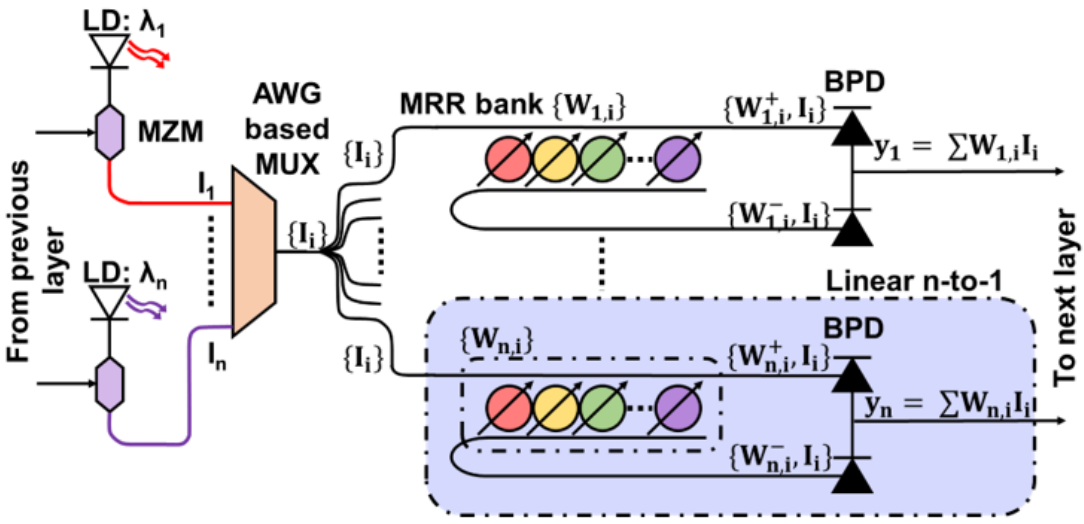


Figure 3. Example of silicon photonic computing. The figure shows a photonic neuron implementation based on the noncoherent Broadcast-and-weight (B&W) protocol [7].

multicasting capability could play a key role in this context. IMC implementations are capable of approaching 1 Tb/s internal bandwidth (as in the Micron HMC) or 10 TOPS/mm^2 (with memristive dot-product engines). Moreover, to address the requirements of latency-critical AI workloads, the interconnects for IMC systems should provide low end-to-end latency of 10-100 ns.

OPTICAL COMPUTING

Optical computing refers to the paradigm where computational operations, such as matrix-vector multiplications, can be performed entirely in the optical domain. By communicating, detecting, and processing information directly in the optical domain, silicon photonics-based processors have the potential to provide very high footprint efficiencies in the hundreds of TMAC/s/mm^2 with energy efficiencies of sub-fJ/MAC [9], [10].

Figure 3 shows an example of a photonic computation unit to illustrate the principles of optical computing. The computation unit is a noncoherent photonic neuron configuration, relying on multiple wavelengths of light to perform parallel processing. The figure shows n neurons in a layer, with the colored-dotted box representing a single neuron. With appropriate orchestration of the involved photonic components (Figure 3), a weighted summation of input optical signals can be achieved in each neuron. After summation, nonlinearity in the neuron can be implemented with optoelectronic devices such as excitable lasers or electro-absorption modulators. Although optical computing is still in its infancy, there have been some exciting recent developments, such as the unveiling of an optical computing accelerator for machine learning at HotChips 2020 by a startup called LightMatter.

Interconnection needs: For optical computing to reach its true potential, the networks interconnecting the optical computing units should support $>$10 Tb/s bandwidth, 1-10 ns latency, and $<$10 fJ/bit energy. This can be achieved if the optical computing units would naturally use photonic interconnects to support intra- and inter-unit communication, without requiring frequent electrical-to-optical and optical-to-electrical conversions that limit the computational throughput, latency, and energy-efficiency of the optical computing core.

INTERCONNECT DESIGN FOR NEXT-GENERATION HETEROGENEOUS SYSTEMS

Based on the particular promises of various emerging computing paradigms and their specific interconnection needs, we have identified several key features related to the interconnection network design for our envisioned heterogeneous computing system (Figure 1). Table 1 lists the

identified design features, which we discuss next.

INTERCONNECT ARCHITECTURES AND TOPOLOGIES FOR INTERPOSER-INTEGRATED CHIPLET PLATFORMS

Interconnect innovations can have a transformative impact on our envisioned heterogeneous computing system (Figure 1). The best practice for designing interconnects for such systems has been to tailor the interconnect architecture to the specific data movement need of the target workload. For example, consider deep learning training and inference workloads. The representative kernel in deep learning workloads is the nested for loops shown in Figure 2C that perform a convolution. The for loops can be tiled, their ordering can be permuted, they can be partitioned across compute units in different ways; each of these many *dataflows* exhibits a different data movement pattern and would benefit from a tailored network. It is therefore possible to arrive at highly efficient design points by adapting the algorithm and network to be amenable to each other. A similar design approach was utilized in the ScaleDeep system [11] to derive the custom interconnect topologies (Figure 2D) for a deep learning training workload in multi-chip architectures; a methodology that could be generalized to other workloads and architectures.

For our envisioned heterogeneous system (Figure 1), unconventional inter-chiplet interconnect topologies are worth exploring given the different data reuse and movement patterns across the target workloads listed in Table 1. The design space is expected to be even more interesting as we move to larger scale systems in the future that incorporate deeper hierarchies and more exotic technologies, e.g., interposer-based inter-chiplet communication as in SIMBA, or wafer-scale integration as in Cerebras.

To realize our vision of custom chiplet solutions, the cost of communication between chiplets on an interposer needs to be similar to the cost of communication within a monolithic system. In reality, bandwidth and latency through the interposer or package substrate may be negatively impacted. Clock crossings between chiplets and interposers manufactured in different processes must be managed. It is clear that this future heterogeneous interconnection platform has to be modular and networking-based, similar to a Network-on-Chip rather than a solely shared-medium based approach, to support the increasing number of chiplets in the systems of the future [1]. Inheriting trusted and true techniques (e.g., Globally Asynchronous and Locally Synchronous (GALS) architectures) and morphing them into modern incarnations (such as locally technology-domain compatible incarnations), while minimizing the need for inter-domain conversion coupled with novel dataflow-aware topologies (e.g., topologies with short network diameters and high local connectivity such as small-world graphs), will be the key to addressing these challenges.

THE PROMISE OF EMERGING INTERCONNECT TECHNOLOGIES

In this section, we examine the emerging photonic, wireless, and microfluidic interconnect technologies, which when coupled with novel architectures as discussed above and complementing or replacing conventional wires within an interposer as shown in Figure 1, provide critical features and unique benefits to potentially aid the successful realization of our envisioned heterogeneous system.

In the proposed vision, conventional signaling through a silicon interposer is a faster alternative to traditional Printed Circuit Board (PCB), and a less expensive option than monolithic 3D integration. This is because, by being mainly used for inter-chiplet wiring, the interposer can be manufactured in a different technology node (with higher yield) than chiplets, which in turn may employ technologies amenable to quantum, optical, in-memory, or DNA computing and DNA storage. Hence, interposers seem like the perfect integration platform for systems that incorporate a host, multiple accelerator chiplets, along with stacked DRAMs or DNA storage.

Despite these advantages, interposers have remained underutilized. This can be attributed to importing interconnection architectures and topologies from the traditional on-chip networks domain which do not fully exploit the abundant wiring resources of the interposer. Novel topologies and routing protocols are needed to deliver the overall throughputs necessary to support integration of computing and memory chiplets

Table 1. Summary of characteristics of emerging computing technologies and their interconnect requirements.

	Emerging Computing Technology			
	Quantum	**Optical**	**In-Memory**	**DNA**
Applications	Cryptography, unstructured search, combinatorial optim., generative chemistry	Deep neural networks, scientific computing	Deep neural networks, genomics, signal processing, recomm. systems, data analytics	Ultra-massive storage of data, bio-compatible processing, theranostics
Required Bandwidth	1 Tb/s	10 Tb/s	1 Tb/s	1 Gb/s
Required Efficiency	1–10 fJ/bit	1–10 fJ/bit	0.1-1 pJ/bit	$<$1 aJ/bit
Required Latency	1–10 ns	1–10 ns	10–100 ns	$\sim$1 hour
Interconnect Alternatives	Electrical, Photonic, Wireless	Photonic	Electrical, Photonic, Wireless	Microfluidic, Photonic, Electrical, Wireless
Analog–Digital Conversion?	Depends on the qubit control/readout scheme	Yes (if computing is implemented using analog data signals)	Yes (if computing is implemented using analog data signals)	Yes
Inter-Technology Data Conversions?	Depends on the qubit readout and quantum state transfer schemes, which could be optical	Depends on whether the architecture is all-optical or not	None, if electrical or wireless interconnects are used	From/to the biochemical domain via photo-chemistry, microfluidics, EM transduction
Challenges for Interconnect Design	Cryogenic operation, thermalization, limited cooling power of dilution refrigerator, latency and bandwidth	Thermal and fabrication variations, crosstalk, aging, tuning power, side-channel attacks	Thermal, analog noise, ADC area/energy/ bandwidth, fabrication challenges, DRAM cost, programming interface	High error rates, slow operation, domain conversion, waste management.
Compatibility requirements from memory/storage subsystems	Compatible with most memory/storage technologies, though at cryogenic operation	High-speed optical transceivers at the storage/memory interface	Implementations vary, but compatible with most memory/storage technologies	Suitable for long-term, massive storage and bio-compatible operation

providing the performance of monolithic 3D integration while costing orders of magnitude less. Moreover, conventional signaling may fall short in providing the speed, efficiency, or versatility demanded by heterogeneous architectures, leaving space for other interposer compatible interconnect technologies to fill this gap as discussed next.

Silicon photonic interconnects: To achieve the target performance values listed in Table 1, silicon photonic interconnects appear as an extremely efficient communication substrate, both on-chip or through an interposer. Moreover, by both communicating and computing in the optical domain, an entirely new class of intuitive, highly energy-efficient (a few fJ/bit) optical computing architectures will become viable. Silicon photonic interconnects will also be essential to support the high bandwidths required for future 3D stacked memory and IMC architectures [12]. The stability of most silicon photonic devices at low temperatures and the ability to transport single photons may also make photonic interconnects a very effective high-bandwidth and even quantum-coherent communication medium for quantum computing. Moreover, as recently demonstrated DNA technologies such as photonic DNA memory and photonic translation of DNA strands mature, photonic interconnects will enable efficient photo-chemical interfaces with DNA computing and storage units.

However, to realize the potential of silicon photonics, many challenges remain to be overcome. To address them, it will be imperative to take a cross-layer optimization approach, wherein foundry-true behavioural and energy models of silicon-photonic devices and circuits inform critical choices for the design of silicon-photonic interconnect architectures. Such a cross-layer approach where microring resonator device widths and layouts are co-designed with device tuning circuits and router architectures (that aggregate banks of these devices) can achieve much better communication performance and energy-efficiency than conventional approaches that optimize the photonic device, circuit, and architecture layers separately [7].

Wireless interconnects: A new window of opportunity for wireless interconnects has emerged by bridging the horizontal and vertical axes. The vertical through-silicon-via (TSV)-based antennas (TSV-As) can pierce through the interposer within the 2.5D/3D multi-die packages and radiate laterally, with the interposer acting as a waveguide with only a 3 dB loss over 10 mm of distance [13]. However, the wireless interface

may incur significant area and power overheads to achieve the competitive 10–100 Gb/s rates. Due to this limitation, it becomes challenging to adopt wireless interconnects to achieve the target performance needs of quantum computing, optical computing, and in-memory computing paradigms (Table 1). Nevertheless, wireless interconnects which do not have physical layouts, are perfectly suitable for supporting massive broadcast or establish fast links between distant chips [13]. Due to this capability, wireless interconnects can interface with a massively parallelized DNA memory, to realize the full potential of the DNA storage technology, or to implement an ultra-fast control channel across the entire heterogeneous system which can result in completely novel memory coherency and other control protocols.

Microfluidic interconnects: DNA computing modules best communicate among themselves using microfluidic channels [14], as such channels can easily carry reagents necessary for communication among these chiplets. Microfluidic channels within silicon substrates have been proposed and created for a variety of purposes from liquid cooling to communication in Lab-on-Chip devices. Advanced fabrication techniques like 3D printing have been shown to produce low-cost substrates in polymers with microfludic interconnect channels. Such techniques can be used to create an interposer layer with an array of microfludic channels to interconnect chiplets in its technology domain. This will eliminate the need for unnecessary inter-domain signal conversion while communicating strictly within the DNA domain for example, between a DNA storage and a DNA computer chiplet.

CROSS-TECHNOLOGY DESIGN: CHALLENGES AND OPPORTUNITIES

The interconnection sub-system of our envisioned heterogeneous system will connect the chiplets of various computing paradigms and memories together. It will consist of disparate technologies, potentially using an efficient combination of photonic, wireless, and electrical interconnect technologies. We envision implementing data communication over the interconnection sub-system in the digital domain, for high error tolerance and ease of implementation. But some computing chiplets can benefit from implementing the processing in the analog domain (e.g., in-memory, optical, and DNA computing). Therefore, the interconnection sub-system may require analog-to-digital and digital-to-analog conversion, which can incur undesired area, energy, and bandwidth overheads (Table 1). Similarly, while inter-interconnect-technology data conversion can be avoided for optical computing chiplets if connected among each other with optical interconnects, chiplets of all the other computing paradigms will need inter-interconnect-technology data conversions to be transmitted through the interconnect sub-systems. For example, memory access from the DNA archives can only be communicated to a processing element or another DNA-based unit after conversion into electronic, electromagnetic, or photonic domain through microfluidics. For instance, a conversion from DNA to photonic domain can be achieved through the use of a photo-chemical interface where, utilizing the specific tunability of various fluidic substances, wavelength division multiplexing can be achieved. Likewise, interaction of EM waves with microstructures in microfluidic channels can provide a novel optofluidic interface platform for domain conversion between optical and microfluidics.

All these insights will guide the future interconnect systems for processing platforms that consist of heterogeneous technologies. However, it is worth noting that the coexistence of inter-domain and inter-technology interconnects, may present additional challenges in some computing paradigms, such as cryogenic temperatures in quantum computers or existence in bio-chemical environments for DNA archives. While optical processors and optical interconnects can eliminate such inter-domain coexistence challenges, it has the challenge of thermal drift in tuning. Therefore, designers of the future will have the option of choosing from an array of interconnect technologies depending on the requirements of the processing chiplets and available resources.

We envision the future interconnect subsystems for heterogeneous chiplet based computers to incorporate a multi-layered interposer with electrical, wireless, optical and microfluidic layers to cater to the disparate technologies of the individual chiplets as shown in Figure 1. Intelligent inter-layer routing and floorplanning needs to be

adopted through simple passage holes to allow the interconnects of different technologies vertically traverse through the interposer layers to reach its corresponding layer to enable inter-chiplet connectivity. The most important challenge for future designers of such interconnections would be to achieve the correct balance between power-performance goals and over-provisioning of real-estate, while catering to such widely varying demands from the interconnects exemplified by the range of throughputs among disparate technologies.

CONCLUSIONS

In this article, we discussed glimpses of the current state of knowledge in the diverse emerging computing paradigms of DNA computing, quantum computing, in-memory computing, and optical computing. To live up to their early promise of disrupting the performance bounds of processing data-centric applications, the massively heterogeneous computers based on these emerging paradigms demand for highly efficient and dependable data provisioning. Initial discussions indicate that the emerging electro-optic, microfluidic, and wireless interconnection technologies can meet this demand if their full potential can be realized, for which the co-design of processor-interconnection subsystems holds the key. We also envision the need for multi-layered interposer structures to enable the massively heterogeneous computing platforms of the future.

ACKNOWLEDGMENT

The authors thank Maurizio Palesi for providing the opportunity to organize this panel at NoCArc2020 and this article is dedicated to the memory of friend and colleague Vassos Soteriou. This work was supported in part by the US NSF CAREER Grant CNS-1553264 and EU H2020 research and innovation programme under Grant 863337.

REFERENCES

1. A. Kannan *et al.*, "Enabling interposer-based disintegration of multi-core processors," in *MICRO-48*, 2015.
2. Y.-H. Chen *et al.*, "Eyeriss: A Spatial Architecture for Energy-Efficient Dataflow for Convolutional Neural Networks," in *Proceedings of ISCA-43*, 2016.
3. D. Soloveichik *et al.*, "DNA as a universal substrate for chemical kinetics," *PNAS*, vol. 107, no. 12, pp. 5393–5398, 2010.
4. S. K. Tabatabaei *et al.*, "DNA punch cards for storing data on native DNA sequences via enzymatic nicking," *Nature Communications*, vol. 11, no. 1, p. 1742, 2020.
5. F. Arute *et al.*, "Quantum supremacy using a programmable superconducting processor," *Nature*, vol. 574, no. 7779, pp. 505–510, 2019.
6. B. Patra *et al.*, "Cryo-CMOS circuits and systems for quantum computing applications," *IEEE JSSC*, vol. 53, no. 1, pp. 309–321, 2017.
7. S. Pasricha and M. Nikdast, "A survey of silicon photonics for energy-efficient manycore computing," *IEEE MDAT*, vol. 37, no. 4, pp. 60–81, 2020.
8. M. Dazzi *et al.*, "Efficient pipelined execution of cnns based on in-memory computing and graph homomorphism verification," *IEEE Transactions on Computers*, vol. 70, no. 6, pp. 922–935, 2021.
9. F. P. Sunny *et al.*, "A survey on silicon photonics for deep learning," *ACM JETC*, vol. 17, no. 4, 2021.
10. J. Feldmann *et al.*, "Parallel convolution processing using an integrated photonic tensor core," *Nature*, vol. 589, pp. 52–58, 2021.
11. S. Venkataramani *et al.*, "SCALEDEEP: A Scalable Compute Architecture for Learning and Evaluating Deep Networks," *ISCA*, 2017.
12. I. G. Thakkar and S. Pasricha, "3D-ProWiz: An energy-efficient and optically-interfaced 3D DRAM architecture with reduced data access overhead," *IEEE TMSCS*, vol. 1, no. 3, pp. 168–184, 2015.
13. V. Pano *et al.*, "TSV antennas for multi-band wireless communication," *IEEE JETCAS*, vol. 10, no. 1, pp. 100–113, 2020.
14. A. Pfreundt *et al.*, "An easy-to-use microfluidic interconnection system to create quick and reversibly interfaced simple microfluidic devices," *Journal of Micromechanics and Microengineering*, vol. 25, no. 11, p. 115010, 2015.

Amlan Ganguly is an Associate Professor and Department Head of Computer Engineering at Rochester Institute of Technology, USA. His research interests are in wireless interconnects, non-von Neumann computers, edge computing and hardware security.

Sergi Abadal is a Distinguished Researcher at Universitat Politècnica de Catalunya (UPC), Spain. His research interests include graphene antennas, chip-scale wireless communications, and computer architecture.

Ishan Thakkar is an Assistant Professor of ECE at the University of Kentucky, USA. His research interests include In-Memory Computing, Optical Computing and Interconnects.

Natalie Enright Jerger is Professor and Canada Research Chair in Computer Architecture at the University of Toronto. Her research interests include networks-on-chip, approximate computing and internet of things.

Marc Riedel is an Associate Professor of Electrical and Computer Engineering at the University of Minnesota. His research interests include digital design, DNA storage, DNA computing, and molecular simulations.

Masoud Babaie is a tenured Assistant Professor at the Delft University of Technology, The Netherlands. His research interests include wireless communications and cryogenic electronics for quantum computation.

Rajeev Balasubramonian is a Professor in the School of Computing at the University of Utah. His research explores memory security and accelerators for machine learning and genomics.

Abu Sebastian is a Distinguished Research Staff Member and Technical Manager at IBM Research - Zurich. His research interests include neuromorphic and in-memory computing, exploratory memory devices and nanoscale dynamics and control.

Sudeep Pasricha is a Professor in the ECE Department at Colorado State University, USA. His research interests include chip-scale network and memory architectures, emerging technologies, hardware/software co-design, and cross-layer design optimizations.

Baris Taskin is a Professor in the ECE Department at Drexel University, USA. His research interests include circuits and systems design and design automation, particularly for clocking, networks-on-chip, and HW/SW co-design.