Research and Teaching Statement

Yuyang Wang

Research Vision and Agenda

The evolving field of distributed computing is seeing an exponential growth in traffic demands, fueled by a deluge of data-intensive workloads. This trend is prominently exemplified by the rapid expansion of data-intensive artificial intelligence (AI) applications. The recent advent of large language models is propelling the broad adoption of ever-larger deep learning models and datasets, marking a significant milestone toward the era of data ubiquity. The continued scaling of these applications has pushed the limits of computational hardware, yet outpacing the evolution of the underlying communication infrastructure, rendering chip-to-chip data movement a formidable barrier impeding performance and energy efficiency. My research endeavors to find **transformative connectivity solutions**, maximally harnessing the potential of integrated silicon photonics (SiPh). In this pursuit, I have devised a dual-thrust research agenda for my independent career. The first thrust focuses on **reconfigurable system connectivity**. It aims to develop optical interconnects that not only provide unprecedented bandwidth but also adapt in real-time to the ever-evolving demands of emerging applications. The second thrust looks into **innovative system architectures**. It targets redefining chip-to-chip communication with groundbreaking optical I/O technologies, thereby pioneering new computational paradigms and interconnect functionalities. My research agenda is situated at the system level, leveraging the latest breakthroughs in device designs and link architectures, while simultaneously informing their future advancements from a system application perspective. This cross-layer approach introduces unique design challenges, which I am equipped to tackle with my interdisciplinary research experience (Fig. 1), ensuring the readiness of essential design tools and methodologies for these advanced connectivity solutions.

1 Thrust 1: Reconfigurable System Connectivity

With the advent of augmented reality (AR), virtual reality (VR), and Metaverse applications, distributed machine learning frameworks are seeing an increase in data privacy concerns that were previously confined to sectors with sensitive information, such as banking and healthcare. These sectors typically handle smaller volumes of data with more flexible latency requirements. In response, decentralized learning frameworks like federated learning have received growing popularity, as they allow the exchange of model parameters over raw data. Yet, certain applications still prioritize data parallelism to meet stringent requirements on model accuracy. Consequently, the data landscape in distributed computing is evolving toward both larger volumes and greater heterogeneity. This evolution, coinciding with the expansion of large models like GPTs, necessitates the next generation of optical interconnects to further excel in traffic adaptability, in addition to bandwidth and energy efficiency.

In this research thrust, my objective is to significantly enhance traffic adaptability by co-designing reconfigurable link architectures along with dynamic reconfiguration strategies (Fig. 1-T1). Building upon the SiPh transceiver developed during my postdoctoral research [1, 2]—which stands out for its leading bandwidth capacity and energy efficiency among state-of-the-art solutions—I aim to incorporate greater reconfigurability into its design. My prior work, namely on runtime laser power scaling and link bandwidth reconfiguration [3–5], serves as a proof-of-concept for the effectiveness of traffic-adaptable tuning knobs in improving both the performance and the energy efficiency of optically connected computing systems. Moving forward, I anticipate the success of this research thrust to be contingent on the following critical tasks:

1. Profiling and characterizing the traffic patterns of a diverse range of distributed computing applications, expected to exhibit greater heterogeneity and temporal dynamics compared to the collective communications typically observed in current computing clusters, as referenced in [5].

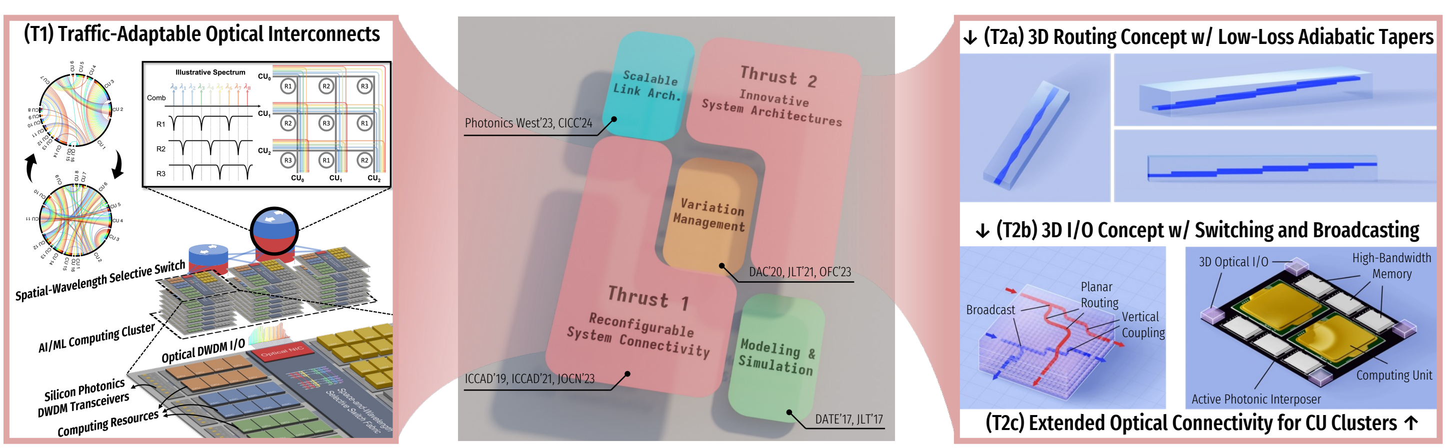

Figure 1: Overview of my research accomplishments and proposed research directions.

2. Introducing additional reconfigurable parameters beyond laser power and link bandwidth, such as wavelength allocation and switching/routing, and developing runtime reconfiguration strategies tailored to these characterized traffic patterns.
3. Conducting system-level simulations to assess the energy and performance impacts of the proposed reconfigurability, supported by credible performance models that accurately reflect real link designs.
4. Designing and integrating reconfiguration modules with state-of-the-art SiPh transceiver implementations, and validating the enhanced interconnect architecture with reconfiguration strategies on a hardware testbed driven by realistic/production network traces.

Throughout this endeavor, I also anticipate deriving valuable insights from a system application perspective. These insights will be instrumental in informing the design of SiPh devices and circuits, focusing on essential aspects such as tuning range and reconfiguration speed, to meet key performance metrics at the system level. This collaborative synergy across multiple design hierarchies is essential to maintain cutting-edge system connectivity in an ever-changing data landscape.

2 Thrust 2: Innovative System Architectures

Complementary to the first research thrust aimed at advancing chip-to-chip connectivity, the second thrust strives to address the notable gap between on-chip and off-chip communication bandwidths. This gap is particularly pronounced in accelerator systems comprising clusters of computing units (CUs) that frequently access data from both on-chip memory banks and off-chip memory pools. Expanding the number of on-chip high-bandwidth memory (HBM) stacks is increasingly impractical as the bandwidth capacity of electronic interposers approaches its limits. Conventional approaches using optical fibers to connect CU clusters and memory pools are also constrained by the size and pitch of fiber arrays. Nonetheless, the emerging concept of 3D optical I/Os, benefiting from dense waveguide routing across multiple layers, presents a promising avenue to scale up CU clusters with optical connectivity that stays on-board with extended reach (Fig. 1-T2a–c). My contribution to assisting the formulation of this concept, which was successfully showcased at the 2023 DARPA ERI Summit, has inspired me to further explore this cutting-edge area. The key challenges I plan to address in this research thrust include:

1. Formulating the 3D routing problem with objectives such as maximized density and minimized loss, and developing efficient routing algorithms that draw from traditional EDA expertise and the latest in machine learning techniques.
2. Informing the design of 3D routing elements with performance and area constraints, and optimizing their physical design employing recent advances in areas such as photonic inverse design and topology optimization.
3. Conducting system-level design space explorations for computing architectures with transformed memory connectivity to delineate optimal system configurations, such as the ideal size of CU clusters that benefit from the expanded reach of on-board optical connectivity, and the optimal balance between on-chip and off-chip memory capacities.

In addition to eliminating the bandwidth taper at chip boundaries and allowing for continued upscaling of CU clusters, this research thrust also promises to expand the role of optical interconnects beyond traditional data communication. For instance, certain computational tasks, such as matrix multiplication, can be offloaded to the optical domain, for which existing explorations have been limited by the vast difference in physical dimensions of electronic and photonic implementations. This limitation can be significantly alleviated by the manifolded density of optical components enabled by 3D routing. This thrust, therefore, not only addresses current technological limitations but also fosters the development of new computing paradigms, where optical interconnects assume a more dynamic and integral role in future computing system architectures.

3 Research Collaborations and Initiatives

My research experience has been deeply rooted in multidisciplinary collaboration, a skill I mastered during my postdoctoral training at the Columbia University. There, I led research initiatives within our group, guided by my supervisor's mentorship and backed by funding from agencies like DARPA, SRC, and ARPA-E. These initiatives required seamless teamwork with colleagues from academia, industry, and governmental bodies. In addition, I have a proven track record in assisting both my doctoral advisor and postdoctoral supervisor with fundraising activities. My responsibilities also encompassed preparing and compiling reports and materials, as well as participating in presentations at quarterly reviews to fulfill the requirements of our funded projects. Given the interdisciplinary essence of my research agenda, I am enthusiastic about the opportunity to collaborate with the diverse faculty in the Department of Electrical Engineering and contribute my experience and enthusiasm to your esteemed institution. I keenly anticipate the chance to work with a community that resonates my commitment to innovation and making a meaningful impact on the future of technology.

References

[1] Y. Wang *et al.*, "Scalable architecture for sub-pJ/b multi-Tbps comb-driven DWDM silicon photonic transceiver," in *SPIE Photonics West*, Mar. 2023.

[2] Y. Wang *et al.*, "Silicon photonics chip I/O for ultra high-bandwidth and energy-efficient die-to-die connectivity," in *IEEE Custom Integrated Circuits Conference (CICC)*, 2024, under review.

[3] Y. Wang *et al.*, "Task Mapping-Assisted Laser Power Scaling for Optical Network-on-Chips," in *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, Nov. 2019.

[4] Y. Wang *et al.*, "Traffic-Adaptive Power Reconfiguration for Energy-Efficient and Energy-Proportional Optical Interconnects," in *IEEE/ACM International Conference On Computer Aided Design (ICCAD)*, Nov. 2021.

[5] Z. Wu, L. Y. Dai, Y. Wang *et al.*, "Flexible silicon photonic architecture for accelerating distributed deep learning," *J. Opt. Commun. Netw.*, 2023, to appear.

Teaching Philosophy and Interest

My educational journey, influenced by remarkable and dedicated mentors, has led me to value the transformative power of effective teaching. I strive for a pedagogy that balances engagement, conveyance, and inspiration, guided by five introspective questions when creating course materials: 1) What insights will students gain from the synchronization of spoken words and slides? 2) What about when they independently examine the slide content at lecture pace? 3) What are their takeaways when revisiting the slides after class? 4) How can I ensure consistency in key takeaways across various learning contexts? 5) How can I structure slides to effectively convey key messages without direct reading? These practices have shaped my teaching philosophy, which emphasizes the effective use of visualization.

Teaching Philosophy Visualization is central to my instructional approach, leveraging the human brain's rapid image processing ability, reportedly surpassing text interpretation by 6x–600x[1]. This enables more efficient extraction of information from slide content, allowing for profound engagement. In STEM education, where theories and equations can become overwhelming, visualization serves to translate intricate ideas into comprehensible and memorable images. It also addresses the challenge of conveying essential concepts without merely reading from slides—a practice that hinders critical thinking. Beyond classroom benefits, effectively visualizing data and concepts is an indispensable skill for students' academic pursuits and research careers. Building on this, I address the challenge of ensuring consistent takeaways in different learning contexts by integrating concise bullet points alongside visuals.

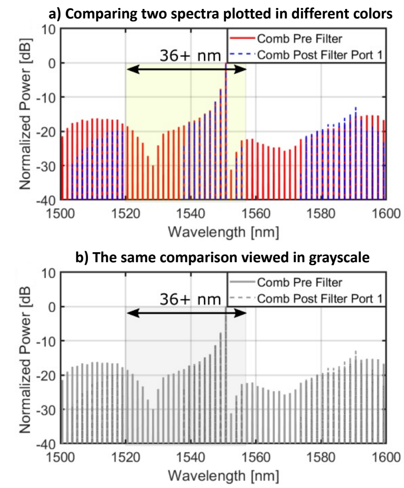

Figure 2: Example of non-inclusive color use: **a)** two spectra distinguishable with normal color vision; **b)** indistinguishable as possibly rendered by impaired color vision.

Inclusive Teaching Informed by my personal experience with minor color vision deficiency, I design visual materials with inclusivity in mind. For example, as shown in Fig. 2, using color as the sole differentiator between two spectra being compared can be challenging for those with color vision impairments. The seemingly surprising prevalence of color blindness, affecting approximately 8% of males and 0.5% of females[2], virtually guarantees that any moderately sized classroom will have students with a form of color vision deficiency. To ensure accessibility for all, I always use multiple modes of differentiation, such as patterns, textures, and annotations. My commitment to inclusivity further extends to all diversity and accessibility aspects in education. Understanding the varied cultural, personal, and educational backgrounds of students, I aim to create a classroom environment that is physically, cognitively, and culturally welcoming, employing inclusive language and diverse examples in my teaching materials, practices that will be regularly reviewed for adherence.

Teaching Interest Aligned with the esteemed curriculum of Stanford University's Department of Electrical Engineering, I am eager to contribute by teaching a range of courses that intersect with my expertise, such as *Digital System Design* (EE 108), *Advanced Optical Fiber Communications* (EE 348), and *Interconnection Networks* (EE 382C). Believing that the preparation for teaching materials deepens my own understanding of the subject, I am also open to teaching courses beyond my immediate expertise, such as *Computer Systems Architecture* (EE 282), and *Electronic Design Automation (EDA) and Machine Learning Hardware* (EE 292A). Moreover, I am enthusiastic about the prospect of designing new courses, currently not offered at the Stanford University, such as *Electronic Photonic Design Automation*, and *Silicon Photonics Optical Interconnects*, which could expand and enrich the department's already distinguished curriculum.

Mentoring My mentoring philosophy extends beyond knowledge sharing and intellectual guidance. I aim to provide holistic support for students in both academic and professional growth, through creating a collaborative research environment that encourages them to excel academically, uphold professionalism, and pursue their research interests towards becoming independent researchers. My postdoctoral experience at the Columbia University involved guiding several Ph.D. students to key achievements, such as publishing papers as lead authors at premier conferences and obtaining their initial industry positions. I look forward to further developing and applying these mentorship practices at the Stanford University.

Post–COVID-19 Considerations In response to the post–COVID-19 new norm of hybrid teaching and research, I have adapted and am ready to further refine my approaches to meet the evolving challenges. Plans include enhancing slides with necessary annotations and animations to compensate for the reduced interaction of physical whiteboard sessions. Furthermore, I intend to organize regular coffee hours or lunch meetings to facilitate in-person discussions when possible, following university guidelines. This adaptability is essential in the changing educational landscape, and I am committed to continually innovating my teaching and mentoring methods to ensure accessible and effective education for all students.

[1]Despite the discrepancy with the unfounded meme claiming 60,000x, as called out in *Research: Is a Picture Worth 1,000 Words Or 60,000 Words in Marketing?* (https://www.emailaudience.com/research-picture-worth-1000-words-marketing/), this is still substantial enough to warrant extra attention to the use of visualization in teaching.

[2]*Types of Colour Blindness* (https://www.colourblindawareness.org/colour-blindness/types-of-colour-blindness/)