Research Statement

Yuyang Wang

December 1, 2023

In the dynamic landscape of distributed computing, the exponential growth in traffic demands within data centers and high-performance computing systems has been distinctive, fueled by a deluge of data-intensive workloads. This trend is prominently exemplified by the rapid expansion of machine learning, big data analytics, and most notably, deep learning (DL)–driven artificial intelligence (AI) applications. The recent advent of large language models, which has revolutionized natural language processing and creative content generation, is propelling the broad adoption of ever-larger DL 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, notably via increased parallelism and specialization. Yet, this rapid progress has outpaced the evolution of the underlying communication infrastructure, rendering chip-to-chip data movement a formidable barrier impeding performance and energy efficiency. This communication bottleneck has become the grand challenge to the quest of upscaling the computing systems toward exascale.

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. The synergy of these thrusts introduces unique design challenges, which I am equipped to tackle with my interdisciplinary research experience, ensuring the readiness of essential design tools and methodologies for these advanced connectivity solutions.

My research agenda is situated at the system level, squarely fitting into the interdisciplinary nexus of the fields related to Device Science and Nanotechnology that bridges silicon photonics, networked systems, computer architecture, design automation, and AI/machine learning. I look forward to the possibility of collaborating with the esteemed colleagues in the Department of Electrical and Computer Engineering at the Texas A&M University to confront the grand challenge of data movement in future computing infrastructures across the full system stack.

1 Research Accomplishments and Skills

My doctoral and postdoctoral work have established a solid foundation for my anticipated research (Fig. 1) and equipped me with the skills necessary to address the upcoming research challenges. I was among the first to integrate accurate **compact models and simulation methodologies** of silicon photonic devices into widely used electronic design automation (EDA) platforms like Cadence Virtuoso [1, 2]. This integration is crucial for the accurate co-simulation of electronic and photonic components, enabling the efficient development of complex photonic integrated circuits (PICs) with reliable performance estimation. My expertise in **process variation characterization, mitigation, and tolerance** ensures the robustness and energy efficiency of fabricated designs [3–5]. This is particularly important for advanced technologies that often rely on emerging fabrication processes and require post-fabrication tuning. My work in these areas has been recognized in leading design automation conferences like DAC and ICCAD, respected photonics venues like OFC and JLT, and a forthcoming book chapter with Springer, effectively connecting the electronics and photonics research communities.

These design enabling techniques have been practically utilized in creating two generations of SiPh transceiver chips, featuring a **scalable link architecture** that facilitates unprecedented channel parallelism and delivers a chip I/O bandwidth of over 16 Tbps with energy consumption below 1 pJ/b [6, 7]. Fabricated in partnership with AIM Photonics through two full-wafer runs, each chip, measuring ~70 mm², densely integrates over 2,000 microresonators. The chip layout process was fully scripted and automated, showcasing not only significant technological advancements but also remarkable design efficiency. The highlighted link architecture was instrumental in securing a $35M SRC JUMP 2.0 grant with 23 principal investigators, a program to which I have contributed through proposal writing and ongoing research efforts. This work has also resulted in invited papers and presentations at both photonics and electronics design conferences (Photonics West and CICC), and an invited journal submission to *Nature Communications Physics*.

These accomplishments have advanced my research into exploring **traffic adaptability** for optical interconnects in distributed computing systems, grounding them in credible performance models and hardware validation. Notably, I have delved into runtime adjustments of parameters such as laser power and link bandwidth, aiming at accelerating distributed machine learning applications with reduced energy consumption [8–10]. A promising off-chip prototype, designed to redistribute wavelength channels across multiple ports at application runtime, is currently under review for publication. Additionally, an on-chip implementation is being fabricated at AIM Photonics and slated for testing in May 2024. These investigations underscore the significance of integrating architectural innovations and optimization strategies at the system level, a process which—without meticulous execution—could inadvertently counteract the advancements achieved at both device and link levels. This realization is a key driver behind my future research directions.

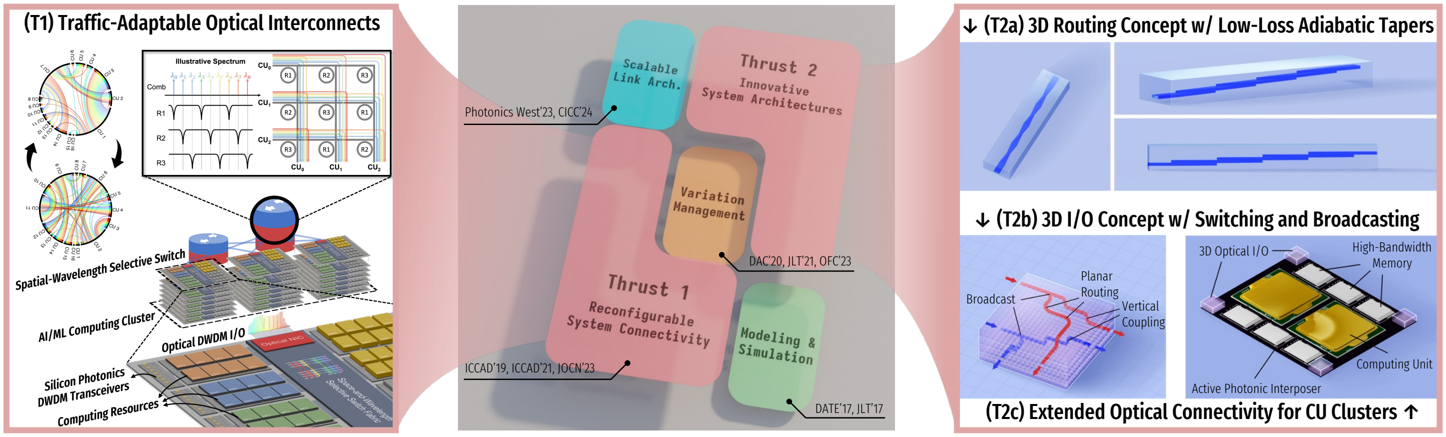


Figure 1: Overview of my research accomplishments and proposed research directions.

2 Research Vision and Agenda

My research agenda is set to continue 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. In light of the evolving data landscape, I plan to focus on two complementary and synergistic research thrusts in pursuit of groundbreaking connectivity solutions.

2.1 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 [6, 7]—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 [8–10], 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 [10].
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.

2.2 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 at the Texas A&M University. The expertise present in areas such as *CAD of VLSI interconnects* (Linda Katehi), integrated optics (Christi Madsen), nanophotonics (Rusty Harris), and other exceptional faculty members working in device physics (Philip Hemmer, Laszlo Kish, Pao-Tai Lin, Peter Rentzepis, Mark Weichold, Jun Zou) is particularly compelling. I am excited to contribute my experience and enthusiasm to your esteemed institution, keenly anticipating 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] R. Wu, Y. Wang, Z. Zhang, C. Zhang, C. L. Schow, J. E. Bowers, and K.-T. Cheng, "Compact modeling and circuit-level simulation of silicon nanophotonic interconnects," in *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2017*. Lausanne, Switzerland: IEEE, Mar. 2017, pp. 602–605.

[2] Z. Zhang, R. Wu, Y. Wang, C. Zhang, E. J. Stanton, C. L. Schow, K.-T. Cheng, and J. E. Bowers, "Compact Modeling for Silicon Photonic Heterogeneously Integrated Circuits," *Journal of Lightwave Technology*, vol. 35, no. 14, pp. 2973–2980, Jul. 2017.

[3] Y. Wang, J. Hulme, P. Sun, M. Jain, M. A. Seyedi, M. Fiorentino, R. G. Beausoleil, and K.-T. Cheng, "Characterization and Applications of Spatial Variation Models for Silicon Microring-Based Optical Transceivers," in *2020 57th ACM/IEEE Design Automation Conference (DAC)*. IEEE, Jul. 2020, pp. 1–6.

[4] Y. Wang, P. Sun, J. Hulme, M. A. Seyedi, M. Fiorentino, R. G. Beausoleil, and K.-T. Cheng, "Energy Efficiency and Yield Optimization for Optical Interconnects via Transceiver Grouping," *Journal of Lightwave Technology*, vol. 39, no. 6, pp. 1567–1578, Mar. 2021.

[5] Y. Wang, S. Wang, A. Novick, A. James, R. Parsons, A. Rizzo, and K. Bergman, "Dispersion-Engineered and Fabrication-Robust SOI Waveguides for Ultra-Broadband DWDM," in *Optical Fiber Communication Conference (OFC) 2023*. Optica Publishing Group, 2023, p. Th3A.4.

[6] Y. Wang, A. Novick, R. Parsons, S. Wang, K. Jang, A. James, M. Hattink, V. Gopal, A. Rizzo, C.-P. Chiu, K. Hosseini, T. T. Hoang, and K. Bergman, "Scalable architecture for sub-pJ/b multi-Tbps comb-driven DWDM silicon photonic transceiver," in *Next-Generation Optical Communication: Components, Sub-Systems, and Systems XII*, G. Li, K. Nakajima, and A. K. Srivastava, Eds. SPIE, Mar. 2023, p. 55.

[7] Y. Wang, S. Wang, R. Parsons, A. Novick, V. Gopal, K. Jang, A. Rizzo, C.-P. Chiu, K. Hosseini, T. T. Hoang, S. Shumarayev, and K. Bergman, "Silicon photonics chip I/O for ultra high-bandwidth and energy-efficient die-to-die connectivity," in *2024 IEEE Custom Integrated Circuits Conference (CICC)*. IEEE, 2024, under review.

[8] Y. Wang and K.-T. Cheng, "Task Mapping-Assisted Laser Power Scaling for Optical Network-on-Chips," in *2019 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*. Westminster, CO, USA: IEEE, Nov. 2019, pp. 1–6.

[9] Y. Wang and K.-T. Cheng, "Traffic-Adaptive Power Reconfiguration for Energy-Efficient and Energy-Proportional Optical Interconnects," in *2021 IEEE/ACM International Conference On Computer Aided Design (ICCAD)*. Munich, Germany: IEEE, Nov. 2021, pp. 1–9.

[10] Z. Wu, L. Y. Dai, Y. Wang, S. Wang, and K. Bergman, "Flexible silicon photonic architecture for accelerating distributed deep learning," *Journal of Optical Communications and Networking*, 2023, to appear.

Teaching Statement

Yuyang Wang

November 30, 2023

My educational journey, profoundly shaped by the dedication and expertise of several remarkable teachers and mentors, have led me to a firm belief in the transformative power of good teaching on a student's life path. Over the years, I have contemplated and reflected on the quintessential pedagogy that balances the act of engaging, conveying, and inspiring. Specifically, in developing my course materials, I am guided by a series of introspective questions that aims at ensuring that the content not only is informative but also stimulates critical thinking.

I continually ask myself five key questions:

1. What insights will students gather while actively engaging with both my spoken words and the slide content?
2. What insights will students gather while independently examining the slide content at my lecture pace?
3. What are the takeaways for students when they revisit the slides on their own time?
4. How can I guarantee consistency in the key takeaways gathered across these varied learning contexts?
5. How do I structure the slide content to convey key messages to the students without them directly reading from it?

These practices have shaped my teaching philosophy, which emphasizes the effective use of visualization.

1 Teaching Philosophy

Visualization is central to my instructional approach, leveraging the human brain's rapid image processing ability, which reportedly surpasses text interpretation by 6x–600x[1] [1]. This capability enables students to extract information from visual aids alongside verbal explanations far more efficiently than text alone, allowing for profound engagement in class. In the realm of STEM education, where abstract theories and complex equations can be overwhelming, visualization serves as a vital bridge, translating intricate ideas into comprehensible and memorable images. It also elegantly addresses the pedagogical challenge of conveying essential concepts without resorting to simply reading from the slides—a practice that hinders critical thinking. Beyond the immediate classroom benefits, the ability to visualize data and concepts is an indispensable skill for students, one that is increasingly critical in both their academic pursuits and future research careers. Building on this philosophy, I address the challenge of ensuring consistent takeaways from the course materials in different learning contexts—whether inside or outside the classroom—by integrating concise bullet points alongside visuals, ensuring key messages being conveyed.

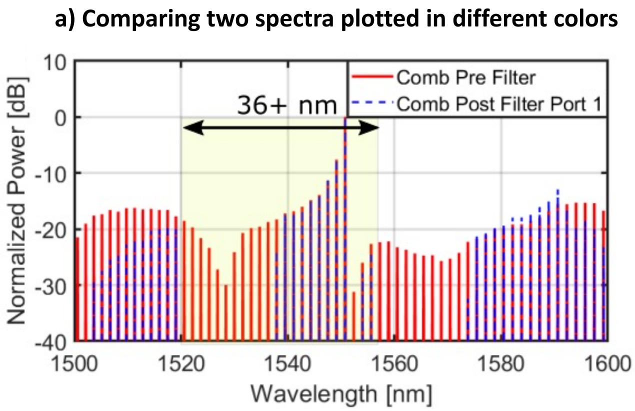
2 Pitfalls and Lessons Learned

Designing visuals that effectively encode information demands thoughtful consideration and attention to detail, ensuring accessibility for all learners. Informed by my personal experience with minor color vision deficiency, I am acutely aware of key pitfalls in visual information delivery, such as solely relying on color contrast to embed information. For instance, Fig. 1 demonstrates how using color as the sole differentiator between two spectra can make the data difficult to interpret for those with color vision impairments. The prevalence of color blindness, affecting approximately 8% of males and 0.5% of females [2], might surprise many. However, this statistic virtually guarantees that in any moderately sized classroom, at least one individual is likely to have a form of color vision deficiency. Acknowledging this, I am committed to employing multiple modes of differentiation in my teaching materials, such as patterns, textures, and annotations, to ensure that all students, regardless of visual ability, have equal access to the information presented.

3 Inclusive Teaching

My commitment to inclusivity extends beyond color vision awareness to embrace all facets of diversity and accessibility in education. I recognize that students come with a broad spectrum of cultural backgrounds, personal histories, and educational experiences, all of which influence their learning needs. In light of this, I will strive to create a classroom environment that is not only physically accessible but also cognitively and culturally welcoming. This entails the use of language that is inclusive and bias-free, as well as the incorporation of diverse examples in my teaching materials, practices that I will regularly reflect on to ensure adherence.

[1]Despite the discrepancy with the unfounded internet meme claiming 60,000x, as called out in *The 60,000 Fallacy* (https://policyviz.com/2015/09/17/the-60000-fallacy/), this is still substantial enough to warrant extra attention to the use of visualization in teaching.



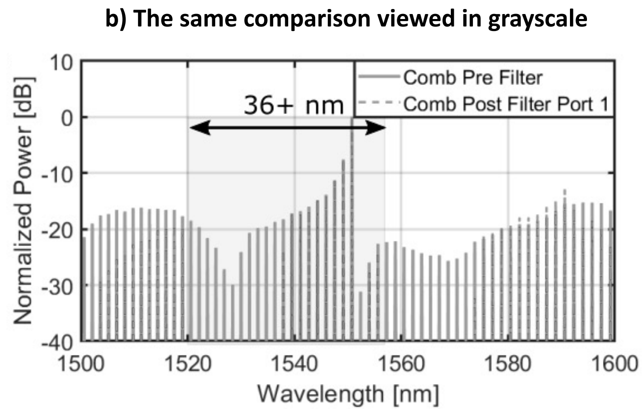


Figure 1: Example of non-robust information encoding solely with color contrast. **a)** Two spectra distinguishable by people with normal color vision. **b)** Illustrative rendering of the same two spectra possibly perceived by people with color vision deficiency.

4 Teaching Plans

In harmony with the esteemed curriculum of the Department of Electrical and Computer Engineering at the Texas A&M University, I am eager to contribute by teaching a range of courses that intersect with my expertise, such as

- Optical Communication Systems (ECEN 462),
- Integrated Optoelectronics (ECEN 675), and
- Optical Interconnects Circuits and Systems (ECEN 721).

Believing that the preparation for teaching materials deepens my own understanding of the subject matter, I am also open to teaching courses beyond my primary expertise, such as

- Computer Architecture and Design (ECEN 350).

Moreover, I am enthusiastic about the prospect of designing new courses, currently not offered at the Texas A&M University, such as

- Electronic Photonic Design Automation,

which could expand and enrich the department's already distinguished curriculum.

5 Mentoring

My mentoring philosophy extends beyond knowledge sharing and intellectual guidance. I am dedicated to providing holistic support for the academic and professional development of the students. I will strive to create a supportive and collaborative research environment, inspiring students to excel in their studies, uphold professionalism, and explore their own research interests until they emerge as independent researchers. My time as a postdoctoral researcher at the Columbia University allowed me to partially implement this philosophy, guiding several Ph.D. students to significant milestones, including publishing their first papers as lead authors at premier conferences like OFC and CLEO, and securing their initial industry positions. I eagerly anticipate the opportunity to extend and refine my mentorship practices at the Texas A&M University.

6 Post–COVID-19 Considerations

In the post–COVID-19 era, I have adapted to the emerging norms of hybrid teaching and research settings. I am prepared to further refine my teaching and mentoring approaches to accommodate these evolving challenges. For example, I plan to enhance my slides with additional annotations and essential animations to offset the lack of real-time interaction that a physical whiteboard provides. Additionally, I plan to introduce regular coffee hours and/or lunch meetings within the group to facilitate in-person discussion whenever feasible and in compliance with the university guidelines. Such adaptability is critical for navigating through the evolving educational landscape, and I am committed to the continuous innovation in my teaching and mentoring practices to provide all students with accessible and effective education.

References

[1] "Research: Is A Picture Worth 1,000 Words Or 60,000 Words in Marketing? | Emailaudience." [Online]. Available: https://www.emailaudience.com/research-picture-worth-1000-words-marketing/

[2] "Types of Colour Blindness." [Online]. Available: https://www.colourblindawareness.org/colour-blindness/types-of-colour-blindness/

Diversity Statement

Yuyang Wang

December 1, 2023

Official statistics published by the U.S. government [1] indicate that individuals with doctoral degrees belong to a distinguished group characterized by the highest median weekly earnings and the lowest unemployment rates across the country. With the privilege comes responsibility. In recognizing the advantages my educational status affords me, I am committed to leveraging this privilege to foster inclusive excellence in STEM education, where diversity metrics have yet to reach an acceptable standard.

Having attended training programs centered on diversity, equity, and inclusivity (DEI) in STEM education, I have gained a solid understanding that *diversity* in academia extends beyond common identity indicators like race, class, ethnicity, citizenship, gender & gender identity, sexual orientation, religion, ability, marital status, veteran status, political affiliation, etc. It also includes educational backgrounds and life experiences such as those of transfer students, non-traditional students (e.g., aged 25 and above or those returning to education), student parents, first-generation students, academically under-prepared students, international students, and so on. Capturing the full spectrum of diversity is essential to crafting a truly inclusive educational environment.

The commitment to diversity became even more personal and tangible during a routine COVID test on campus. The young man conducting the test was unaware of what a "postdoc" was, revealing a gap in his knowledge of educational pathways, especially regarding doctoral programs. This interaction keeps on reminding me of the broader role that an educator can play. As a future faculty member, I am inspired to be a force of enlightenment and change, to open up opportunities to students who may have never known of the intellectual and life options that abound at our university. The influence of these seemingly simple acts of sharing knowledge and opening doors can be more significant than most people perceive.

Figure 1: Firsthand experience of diversity in action at the SRC JUMP 2.0 Center for Ubiquitous Connectivity (CUbiC) have profoundly shaped my understanding of inclusive practices in STEM education.

My time at the Center for Ubiquitous Connectivity (CUbiC) under the SRC JUMP 2.0 program has been instrumental in shaping my understanding of practical diversity in action. Diversity is a critical priority in the CUbiC center, as evidenced by the center featuring a female director, a female theme lead, 6 female principal investigators, alongside over 20% and growing researchers from underrepresented groups. Witnessing the center's inclusive mentoring and outreach programs has demonstrated to me the utmost importance of such initiatives in a research environment. These experiences have provided me with a solid framework for what I aspire to emulate in establishing my own research group in the Department of Electrical and Computer Engineering at the Texas A&M University.

In conclusion, it is through the combined efforts—recognizing the full spectrum of diversity, actively contributing to an inclusive academic culture, and taking every opportunity to educate and inform—that I aim to fulfill the responsibility that accompanies my privileged educational achievement. I see it as my duty to ensure that the pathways to academic and professional success are accessible to all, and to serve as a mentor for the next generation of scholars and innovators.

References

[1] "Education pays : U.S. Bureau of Labor Statistics." [Online]. Available: https://www.bls.gov/emp/tables/unemployment-earnings-education.htm