Received: 5 July 2023 | Accepted: 6 October 2023

DOI: 10.1002/aaai.12165

SPECIAL TOPIC ARTICLE

The TILOS AI Institute: Integrating optimization and AI for chip design, networks, and robotics

Andrew B. Kahng[1] | **Arya Mazumdar**[2] | **Jodi Reeves**[3] | **Yusu Wang**[2]

[1]Departments of CSE and ECE, UC San Diego, La Jolla, California, USA

[2]Halıcıoğlu Data Science Institute, UC San Diego, La Jolla, California, USA

[3]School of Technology and Engineering, National University, San Diego, California, USA

Correspondence
Andrew B. Kahng, Departments of CSE and ECE, UC San Diego La Jolla, California, USA.
Email: abk@ucsd.edu

Yusu Wang, Halıcıoğlu Data Science Institute, 3234 Matthews Ln, La Jolla, CA 92093, USA.
Email: yusuwang@ucsd.edu

Funding information
National AI Research Institutes program; NSF and Intel: CCF., Grant/Award Number: 2112665; Directorate for Computer and Information Science and Engineering

Abstract

Optimization is a universal quest, reflecting the basic human need to *do better*. Improved optimizations of energy-efficiency, safety, robustness, and other criteria in engineered systems would bring incalculable societal benefits. But, fundamental challenges of scale and complexity keep many such real-world optimization needs beyond reach. This article describes The Institute for Learning-enabled Optimization at Scale (TILOS), an NSF AI Research Institute for Advances in Optimization that aims to overcome these challenges in three high-stakes use domains: chip design, communication networks, and contextual robotics. TILOS integrates foundational research, translation, education, and broader impacts toward a new nexus of optimization, AI, and data-driven learning. We summarize central challenges, early progress, and futures for the institute.

INTRODUCTION

TILOS, The Institute for Learning-enabled Optimization at Scale, is an NSF AI Institute partially supported by Intel Corporation. The institute is a partnership of six universities: UC San Diego (lead), the Massachusetts Institute of Technology, National University, the University of Pennsylvania, the University of Texas at Austin, and Yale University. It began operations in November 2021, with a mission to “make impossible optimizations possible, at scale and in practice”. TILOS aims to discover a new integration of AI, optimization, and the leading edge of practice for three high-stakes use domains: chip design, communication networks, and contextual robotics. These domains collectively underpin future innovations of information and communication technology, along with cyberphysical systems. Advancements in these domains critically hinge on better optimization: all these areas involve complex engineered systems with many pieces that need to be optimized individually and also holistically. For example, how can more circuits fit in a smaller region on the chip while using less energy? With the scale and complexity of these systems skyrocketing in the modern era, the real-world optimization needs have surpassed the reach of traditional methods. TILOS aims to use the power of AI to significantly accelerate optimization in these three use domains. At the same time, optimization is a key component of many AI frameworks: the

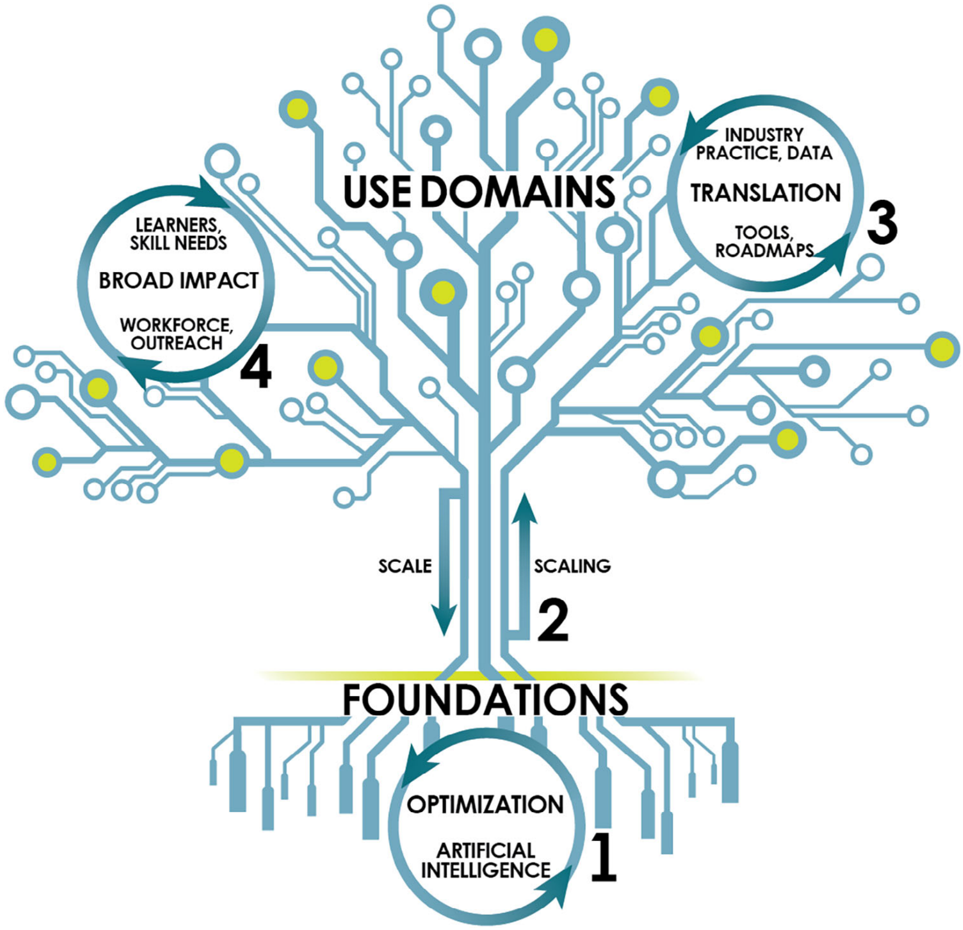

FIGURE 1 Toward integration of AI/ML and optimization: four virtuous cycles in The Institute for Learning-enabled Optimization at Scale (TILOS).

training of a deep learning neural network fundamentally involves optimizing a highly non-convex system. Hence, TILOS will develop innovations that can fuse AI and optimization, propelling their symbiotic advancement in both foundations and applications.

The approach proposed by TILOS is reflected in Figure 1, which shows four *virtuous cycles* in the institute. First, *mutual* advances of AI and optimization provide our foundations. Second, the challenges of scale in practical optimization contexts, alongside the scaling breakthroughs achieved by new AI/ML and optimization methods, bind together foundations and the use domains of chip design, networks, and robotics. Third, the cycle of translation and impact must steadily bring research and the leading edge of practice closer together. Fourth, a cycle of research, education, and broadening participation must be established in order to grow a new AI-optimization-use nexus and its workforce. This is an ambitious agenda for just under 30 faculty, with a similar number of Ph.D. students and postdoctoral scholars. Thus, as TI-LOS members create these virtuous cycles, an added mindset is required: *How can we amplify our efforts, and optimize our impacts?* This mindset shapes many basic processes, from collaboration mechanisms to curriculum creation for workforce development, to culture change (reproducibility, benchmarking, datasharing, etc.) across entire research communities.

FUNDAMENTAL RESEARCH

Fundamental research in TILOS is balanced across the foundations of AI/ML and optimization, and the three use domains.

New foundations of learning and optimization

New, "modern" vistas for foundational research in optimization have opened up in the past decade, with the confluence of areas, emerging computational resources, and consequences of automation. The TILOS *Foundations* team, with 12 faculty from four institutions, focuses on five core research thrusts at the interface of AI and optimization: (1) bridging continuous and discrete optimization; (2) parallel, distributed and federated optimization; (3) optimization on manifolds; (4) dynamic decision-making in uncertain environments; and (5) nonconvex optimization in deep learning. These thrusts are respectively motivated by five aspects of the modern context for optimization, as exemplified by our three core use-domains: (1) continuous models and methods are being deployed for inherently discrete problems; (2) distributed and federated models of data storage and algorithms are replacing centralized ones; (3) spaces with richer geometric structures than Euclidean (e.g., the configuration space of an articulated robot can be better modeled by a manifold) are increasingly used; (4) optimization must increasingly be performed in unknown and dynamic environments; and (5) nonconvex models and methods are required to explain and approach modern machine learning.

Interplay at the interface of foundations and use

The use domains of chip design, networks, and robotics bring diverse optimization challenges but inspire shared solutions with commonalities such as physical embeddedness, hierarchical-system context, underlying graphical models, safety, and robustness as first-class concerns, and the bridging of human-guided and autonomous systems. Moreover, practical optimizations in each of these domains bring further, common challenges: (1) instances have enormous scale; (2) representations and abstractions are crucial to success; (3) objectives are hazy, particularly with multi-stage optimizations and dynamic settings; (4) optimization tools must provide reliability and generalization; and (5) scaling of productivity increasingly demands new ways to learn and optimize using modern compute fabrics.

Addressing the above challenges requires us to identify the right representations, develop the machine learning methods suitable for those representations, and leverage the interplay between learning and optimization. Indeed, representations should incorporate domain knowledge, structure in data (e.g., low-dimensional, non-Euclidean), as well as the mathematical structures behind the problems at hand. The right models and representations are fundamental to optimization performance and generalization. As optimization is a fundamental component of modern machine learning, while conversely learning can help solve difficult optimization problems, we need to better understand and leverage this close interplay, form new connections and interactions, and co-evolve both.

For example, TILOS team members' investigations have advanced fundamental understanding of the capacity of graph neural networks (GNNs) in terms of representation learning and optimization (Jegelka, 2022). Examples of these advances include articulating the precise classes of functions that can be represented by GNNs, studying robustness via suitable graph distances, understanding how to better capture long-range interactions in large, sparse (hyper)graphs (critical for netlists in chip design), and developing new sign- and basis-invariant GNNs (which are crucial given the importance of meaningful position encoding in graph learning). TILOS team members are further developing more effective graph learning and optimization models for, for example, chip design applications (Kahng et al., 2024; Luo et al., 2024).

TILOS researchers have made fundamental progress on non-Euclidean optimization and sampling from Riemannian manifolds, providing a key step toward building the theory of computational complexity of solving stochastic differential equations on manifolds. Other highlights include a new framework for continuous neural set extensions to facilitate learning/optimization with discrete functions, progress on first-order methods for min–max optimization, and understanding of optimization dynamics of neural networks and the convergence of stochastic gradient descent. These are just a very small sample of examples: references of the aforementioned results, and many other research outcomes from TILOS can be found at (TILOS).

Use-domain challenges

Fourteen faculty are engaged in TILOS use-domain research: six in robotics and five each in networks and chips, with two faculty jointly pursuing both networks and chips research. There is no shortage of unsolved problems and challenges; indeed, the institute's 5-year strategic plan includes nearly 60 distinct projects across 35 research topics.

Robotics exemplifies how AI and optimization connect to the physical world. TILOS robotics researchers pursue optimization and AI-based methodologies for manufacturing autonomous, adaptive, heterogeneous teams of sensors, robots, and intelligent machines that can work with humans. Key challenges include concurrency, online processing, and minimizing the need for labeled data. This gives rise to three major research thrusts: (1) metric, semantic, and dynamic artificial perception in a physical world; (2) hierarchical semantic mapping at the edge; and (3) multi-robot perception, planning, and communication. Current research outcomes of the institute (see e.g., publications at [TILOS]) show ties to optimization foundations (e.g., solving continuous and discrete optimization problems at scale to facilitate perception, mapping, planning, communication, and reconfiguration), networks (autonomous networking provides the backbone for perception, control, and learning in teams), and chips (hardware design to support computation, communication, and sensing on SWaP-constrained autonomous robots).

As an example, a new collaboration has been formed between the Robotics and Foundations teams on the topic of learning dynamics from trajectory data that respect the underlying structure and physical laws of dynamical systems. In particular, mobile and articulated robots often have Lie groups (viewed as a manifold) as configuration spaces. It is therefore highly desirable to formulate learning and control of robot dynamics that evolve on Lie groups. In a joint effort across the Robotics and Foundations teams, TILOS members develop a new structure-preserving deep learning architecture that can learn controlled Lagrangian or Hamiltonian dynamics on Lie groups, either from position-velocity or position-only data (Duruisseaux et al., 2023). They are further developing safe control synthesis methods with provable safety guarantees for problems with uncertain dynamics and constraints.

Communication networks also depend on AI and optimization, for example, for distributed optimization of a large number of discrete and continuous variables over nonconvex geometric structures induced by interference. A core challenge is to manage at scale radio resources across many devices distributed in space, with comprehension of the physics of signal propagation and information-theoretic limits. This gives rise to three major research thrusts: (1) multi-scale network optimization; (2) automated network fine-tuning; and (3) integration of human experts and physics. Current research outcomes of the institute (see e.g., publications at [TILOS]) highlight the ties to modern optimization foundations (e.g., sequential sampling, federated, and deep learning methods), as well as commonalities with other use domains (e.g., physical embeddedness, underlying graphical models) that enable cross-disciplinary bridges.

One particular work exemplifying the collaboration between the Networks and Foundations teams is the deployment of federated learning in a common three-tier IoT network architecture (Yu et al., 2023). Federated learning of multi-agent systems has been used here in an asynchronous setting with varying network delays, and shown to be effective with highly heterogeneous datasets. TILOS researchers also show their method to be highly resilient to system heterogeneity and dropouts (Vardhan, Ghosh, & Mazumdar, 2023; Yu et al., 2023).

Chip design brings challenges that include hierarchical-system context, extreme cost, and sensitivity of training data, "multi-everything" (physics, objectives) constrained optimization, and pervasive security aspects. Given its decades-long history as a driver of applied optimization and automation, chip design also highlights augmenting rather than rediscovering domain expertise, by encoding expert knowledge and intuition to serve optimization and decision-making agents. This gives rise to four major research thrusts: (1) direct generation of layout from circuit descriptions; (2) breakthrough scaling of verification methods; (3) quantifying the intrinsic cost of robustness in optimization and learning, with respect to aspects such as data anonymity, data integrity, and privacy in federated and distributed settings; and (4) data, benchmarking, and road mapping to improve reproducibility and relevance of research, along with translation into real-world contexts (Kahng, 2022). Research outcomes of the institute include several multi-organizational collaborative efforts seen in the TILOS organizational GitHub (TILOS Organization GitHub), ranging from new machine learning contests to open research enablements to breakthrough results for the classic hypergraph partitioning optimization problem.

EDUCATION AND WORKFORCE DEVELOPMENT

New pathways to lifelong learning for diverse students, along with development of shareable and scalable curricula, are overarching objectives of TILOS efforts in Education and Workforce Development. The six TILOS institutions have a shared goal of (1) making education in optimization, AI, computing, robotics, networking, and chip design more accessible to a diverse group of students; and (2) providing opportunities for those already in the workforce to keep current with the latest developments. These aims are well-aligned to national needs, such as the revival of U.S. semiconductor technology leadership and a diverse domestic workforce, per the 2022 CHIPS and Science Act (CHIPS for America; CHIPS & Science Act, H.R.4346).

TILOS institute members work closely with corporate partners to understand the latter's workforce training needs. This guides creation of new curriculum modules, courses, and programs for university education (both in-person and online, undergraduate and graduate) as well as professional certificates, tutorials, short courses, and summer camps. San Diego-based National University (NU) is a key driver and motivating force for these efforts in TILOS. NU is a 53-year-old, non-profit university that educates a diverse group of students from across the U.S. with over 230,000 alumni and approximately 29,000 active students. NU student demographics include 60% female, 38% male, 14% active-duty military, 14% veteran, and an average age of 33 years. NU is also a Hispanic Serving Institution and a member of the Hispanic Association of Colleges and Universities.

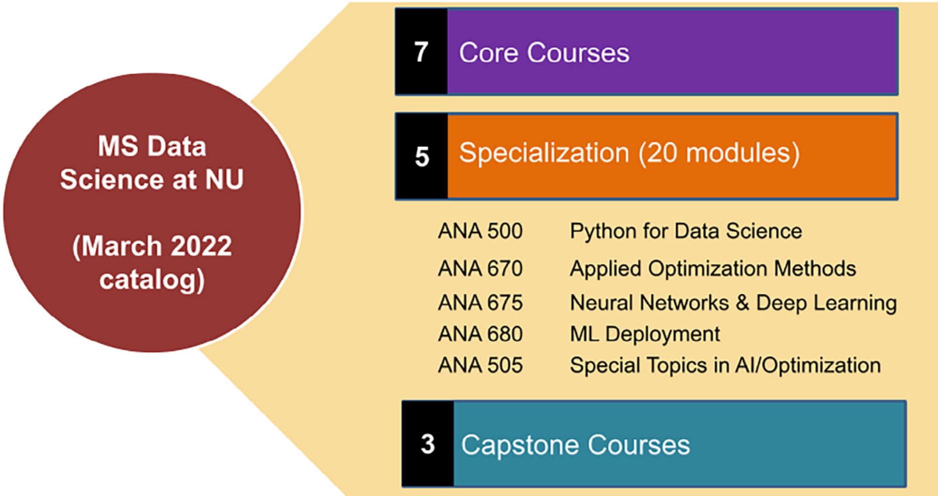

FIGURE 2 Five courses, comprising 20 modules, in the new AI/Optimization specialization of the M.S. Data Science degree program at National University.

A highlight of the past 2 years has been the development, teaching, and assessment of 20 new curriculum modules for a new five-course AI/Optimization specialization in NU's M.S. Data Science degree program (see Figure 2). Enrollment in the specialization has increased from 10 students in the calendar year 2022 to 21 students in Q1 2023 to 32 students in Q2 2023, with 56% of enrolled students being military-affiliated (veteran, retired military, active duty, and active reserve). More recently, a deep collaboration with Intel has enabled National University to integrate several of Intel's "AI for Workforce Development" curriculum modules into newly developed courses in the brand new B.S. Data Science degree at NU, with concentrations in AI/ML, cybersecurity analytics, and bioinformatics.

KNOWLEDGE TRANSFER

Recall from Figure 1 that *translation* at the interface between industry and academia is the third virtuous cycle of TILOS. TILOS has built strong ties with various industries, facilitated by events such as TILOS Industry Day. In an idealized life cycle of translation, real-world practitioners supply problems and data, researchers bridge foundations and use domains to discover new methods, and these results go back into the real world. Unfortunately, today this picture is complicated by various technical and cultural obstacles. In addition to the many fruitful interactions TILOS has with our industry partners, TILOS also aims to help mitigate, if not remove, such obstacles. Four examples of obstacles to knowledge transfer and translation, along with potential mitigations from TILOS, are as follows.

(1) Relevant (real) datasets may be proprietary and shared (if at all) only with very few researchers. TILOS research seeks new democratizations: Can we develop a science of "data virtual reality", enabling the generation of shareable, proxy research data that is artificial but indistinguishable from real *from the perspective of optimization methods and real-world practitioners*? A complementary need is to develop trusted tests for, for example, identity leakage.
(2) Real data may be extremely scarce and expensive, for example, a single execution of the chip design flow may take several weeks, using tool licenses that cost millions of dollars. The research need is to learn to optimize with less real data, and to improve the reliability of methods for data augmentation and transfer learning.
(3) Research may be inherently irreproducible. (i) Some fields have not yet gone through the stages of introspection (Hutson, 2018) and subsequent adoption of "papers with code" as a cultural norm. Here, TILOS directly aims to change long-standing cultures. A recent TILOS effort (Cheng et al., 2023; TILOS MacroPlacement GitHub) provides new open-source (code, data) research enablement, toward transparent assessment of a deep learning method for chip placement. (ii) Irreproducibility may also be due to proprietary tool scripting languages, or report and logfile formats, which cannot be published. Here, TILOS efforts elicited policy changes from major electronic design automation (EDA) tool vendors during the second half of 2022 (Junkin, 2022).
(4) Benchmarking can be forbidden by suppliers of optimization software tools. At the same time, progress of optimization methods requires a clearly illuminated

leading edge, as even very well-studied optimizations may be far from well-solved. (For example, the state of the art for the classic hypergraph min-cut partitioning optimization remained static for a quarter-century, until results last year in Bustany et al. (2022)). The resulting challenge for both research and culture change is to develop new principles and mechanisms that provide a foundation for fair benchmarking.

Through the mitigation of the above obstacles, ongoing research, and efforts to change community culture, TILOS aims to scale *people* in addition to optimization in practice, by pioneering new democratizations, new cultural and scientific or technical norms, and principled bases for looking forward and investing resources.

CONCLUSION

Our world is at the brink of an era where AI becomes an integral part of our daily lives, enhancing our capabilities and shaping a brighter future for humanity. Modern challenges for optimization include helping existing systems to learn from data and adapt to changing circumstances, so as to achieve improved accuracy, speed, and efficiency. In their collective pursuit of optimization in and via AI, TILOS researchers aim to advance foundational mathematics and domain sciences, fueling innovation and enabling us to unlock the full potential of AI technologies. We invite readers to learn more at the institute website, *tilos.ai*.

ACKNOWLEDGMENTS

This material is based upon work supported by the National AI Research Institutes program supported by NSF and Intel under Award No. 2112665. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict.

ORCID

Andrew B. Kahng https://orcid.org/0000-0002-4490-5018

REFERENCES

Bustany, I., A. B. Kahng, Y. Koutis, B. Pramanik, and Z. Wang. 2022. "SpecPart: A Supervised Spectral Framework for Hypergraph Partitioning Solution Improvement." In *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*.

Cheng, C.-K., A. B. Kahng, S. Kundu, Y. Wang, and Z. Wang. 2023. "Assessment of Reinforcement Learning for Macro Placement." In *Proceedings of the* ACM/IEEE *International* Symposium on Physical Design, 158–66.

CHIPS and Science Act, H.R.4346, 117th Congress (2021–2022), Public Law No: 117-167 (08/09/2022). https://www.congress.gov/bill/117th-congress/house-bill/4346/text

CHIPS for America, National Institute of Standards and Technology (NIST), U.S. Department of Commerce. https://www.nist.gov/chips

Duruisseaux, V., T. Duong, M. Leok, and N. Atanasov. 2023. "Lie Group Forced Variational Integrator Networks for Learning and Control of Robot Systems." *PMLR Learning for Dynamics and Control Conference (L4DC)*.

Hutson, M. 2018. "Artificial Intelligence Faces Reproducibility Crisis." *Science* 359(6377): 725–26.

Jegelka, S. 2022. "Theory of Graph Neural Networks: Representation and Learning." In *Proceedings of the International Congress of Mathematicians (ICM)*.

Junkin, D. 2022. "Supporting the Scientific Method for the Next Generation of Innovators." https://open-source-eda-birds-of-a-feather.github.io/doc/slides/BOAF-Junkin-DAC-Presentation.pdf.

Kahng, A. B. 2022. "Leveling Up: A Trajectory of Open-ROAD, TILOS and Beyond." In *Proceedings of the* ACM/IEEE *International Symposium* on Physical Design, 73–79.

Kahng, A. B., R. R. Nerem, Y. Wang., and C. Yang. 2024. "NN-Steiner: A Mixed Neural-Algorithmic Approach for the Rectilinear Steiner Minimum Tree Problem." In *Proceedings of 38th Annual AAAI Conference on Artificial Intelligence (AAAI)*.

Luo, Z., T. Hy, P. Tabaghi, M. Defferrard, E. Rezaei, R. Xarey, R. Davis, R. Jain, and Y. Wang. 2024. "DE-HNN: An effective neural model for Circuit Netlist representation." In *Proceedings of the 27th Internaltional Conference on Artificial Intelligence and Statistics (AISTATS)*.

TILOS. https://www.tilos.ai/publications/.

TILOS MacroPlacement GitHub. https://github.com/TILOS-AI-Institute/MacroPlacement.

TILOS Organization GitHub. https://github.com/TILOS-AI-Institute/.

Vardhan, H., A. Ghosh, and A. Mazumdar 2023. "A Convergent Federated Clustering Algorithm without Initial Condition." In Federated Learning and Analytics in Practice: Algorithms, Systems, Applications, and Opportunities.

Yu, X., L. Cherkasova, H. Vardhan, Q. Zhao, E. Ekaireb, X. Zhang, A. Mazumdar, and T.Š. Rosing. "Async-HFL May 2023. Efficient and Robust Asynchronous Federated Learning in Hierarchical IoT Networks." In *Proceedings of the 8th ACM/IEEE Conference on Internet of Things Design and Implementation*, 236–48.

How to cite this article: Kahng, A. B., A. Mazumdar, J. Reeves, and Y. Wang. "The TILOS AI Institute: Integrating optimization and AI for chip design, networks, and robotics." *AI Magazine* 45: 54–60. https://doi.org/10.1002/aaai.12165

AUTHOR BIOGRAPHIES

Andrew B. Kahng, University of California, San Diego, is the Founding PI and director (2021–2023), Chips team co-lead. He is Distinguished Professor of CSE and ECE in the Jacobs School of Engineering at UCSD, where he holds the endowed chair in high-performance computing. He is a Fellow of ACM and IEEE, and the 2019 Ho-Am Prize laureate in Engineering. His interests span many aspects of IC design automation, including fundamental algorithms, machine learning, and open source. ORCID: 0000-0002-4490-5018

Arya Mazumdar, University of California, San Diego, is the Deputy Director and Associate Director for Research at TILOS, Co-PI and Foundations team co-lead, Associate Professor, The Halıcıoğlu Data Science Institute. He is a Distinguished lecturer of the IEEE Information Theory Society and a recipient of the NSF CAREER Award, Jack K. Wolf Paper Award, and a European Association for Signal Processing Best Paper Award among others. Mazumdar's current research interests include information theory, statistical machine learning and estimation, and distributed/federated optimization. ORCID: 0003-4605-7996

Jodi Reeves, National University, is the Associate Director for Education, Diversity, and Outreach and chair of the Education and Workforce Development Committee, Professor and Department Chair of Data Science, Academic Program Director for M.S. Data Science and B.S. Data Science. ORCID: 0001-9240-1722

Yusu Wang, University of California, San Diego, is the PI and Director (2023-present), Foundations team and formerly Associate Director, Research (2021–2023), Professor at The Halıcıoğlu Data Science Institute, Faculty co-lead for Foundations of Data Science CoP at Translational Data Analytics Institute at the Ohio State University from 2018 to 2020. NSF CAREER Award and DOE ECPI Award. Member of the Computational Geometry Steering Committee, AATRN Advisory committee, and SIGACT CATCS Committee. ORCID: 0000-0001-7950-4348