Optimizing Robotic Systems at All Scales

Brian Plancher (brian.plancher@gmail.com)

Intelligent field robots are a promising solution to many societal challenges from combating epidemics, to scaling global supply chains, to providing home health care to the elderly [1, 2, 3]. However, today, robots are mostly limited to laboratory settings as they cannot react to dynamic, real-world environments in real-time. As such, my core research question is: *how can we construct computational systems that enable robots to intelligently, flexibility, and reliably operate in the field?*

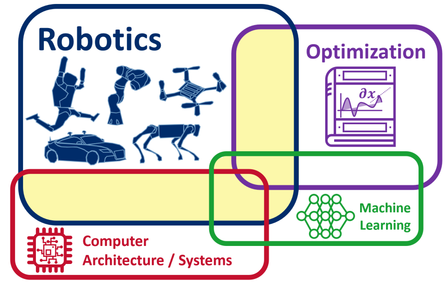

Figure 1: My interdisciplinary research is focused on computational co-design for real-world robotics.

My research seeks to address this problem by developing, optimizing, implementing, and evaluating next-generation algorithms and edge computational systems, at all scales, through algorithm-hardware-software co-design. This approach requires designing theoretically sound optimization- and learning-based algorithms (e.g., model predictive control) that run at order-of-magnitude faster rates on edge computational hardware ranging from small-scale MCUs, to large-scale GPUs and FPGAs, and even to custom ASICs and non von Neumann architectures (e.g., neuromorphic processors). As such, I work across the computational stack, designing algorithms, software systems, and computational hardware at the intersection of robotics, optimization, computer architecture / systems, and machine learning.

In my two years at Barnard College, with the support of undergraduate and masters researchers, and through collaborations across academia and industry, I have multiple tier 1 publications in robotics (e.g., 6 ICRA), computer architecture / systems (e.g., DAC, ISCA), and computer science education (e.g., SIGCSE-TS), and have raised over $800K from the National Science Foundation to support my research agenda. Looking forward, I will build on these successes and develop new algorithms, custom computation hardware, and open-source software to power dynamic and globally useful robots. And, I am very excited by the opportunity to collaboratively build this interdisciplinary agenda at your college or university.

Prior Work – Accelerating Motion Planning and Control Across Scales

Model Predictive Control (MPC) transforms robot motion planning and control problems into (often nonlinear) optimization problems that are repeatedly solved online at high rates. This approach has been shown to generate highly dynamic and environment-aware motions for complex robots [4, 5, 6]. However, to enable its widespread use, a fundamental question remains: *how can we overcome MPC's high computational complexity, while still capturing complex dynamics and providing reliable convergence, for field robots of all scales?*

Much of my prior research has focused on overcoming this challenge by designing theoretically sound algorithms that leverage a combination of offline and online compute to be compressed onto MCUs on tiny robots, or parallelized and accelerated on GPUs, FPGAs, or custom ASICs for large-scale robots (e.g., manipulators, quadrupeds).

This work has been complemented by efforts to develop accelerated middleware and benchmarking solutions, as well as memory efficient edge reinforcement learning. Combined, these results enable all robots, regardless of their size or connectivity, to execute intelligent planning and control at the edge.

Real-time Nonlinear MPC through GPU Accelerated Co-Design

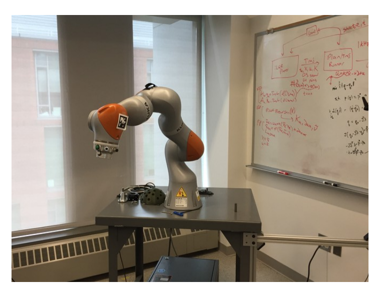

Figure 2: Kuka manipulator using PDDP on the GPU [7].

My research has shown that developing theoretically sound algorithms that are optimized to take advantage of the large-scale parallelism available on GPUs can significantly improve the performance of nonlinear MPC. This has included the development of novel preconditioners for provably better numerical conditioning [8], as well as GPU-friendly variants of trajectory optimization solvers ranging from Parallel Differential Dynamic Programming solvers [9, 7], to Conjugate Gradient-based Direct Methods [10], as well as key underlying kernels such as Rigid Body Dynamics [11, 12]. Implementations of these co-designed algorithms which both leverage these theoretical improvements, as well as expose and exploit GPU-friendly sparsity and parallelism patterns, run at faster-than-state-of-the-art rates for larger problems. This enables real-time, long-horizon optimal control, and has been validated on real robot hardware (Figure 2) [7]. This work has been supported by the National Science Foundation (NSF) through my CRII and new CSSI grants. My leadership in this area has also spurred a number of current academic collaborations to overcome bottleneck optimization-based computations across the robotics pipeline with GPUs, as well as an active industry collaboration (and pending grant) with the Toyota Research Institute (TRI) to develop novel GPU-accelerated risk-sensitive algorithms and solvers for autonomous driving at the edge of handling.

Toward Custom Hardware Accelerators for MPC – Rigid Body Dynamics Building Blocks

I have also shown that key bottleneck computations in the MPC pipeline can be further accelerated through the use of FPGAs and custom ASICs, directly encoding the embodied algorithms' structured sparsity and parallelism patterns into computational hardware. For example, my collaborators and I have shown that custom hardware accelerators for rigid body dynamics can provide as much as a 58x and 519x speedup over the CPU and GPU respectively [11, 13, 14]. This line of work has also set the stage for the development of automated design flows, interdisciplinary collaborations, and heterogeneous architectures that will enable a future of performant and useful custom robotics chips [14, 15]. My core collaborators and I have also been at the forefront of this growing field of custom robotics accelerators, building a vibrant and inclusive community around this effort, with special sessions and workshop at major conferences (e.g., RSS 2022, MICRO 2022-24, DAC 2024, and ICRA 2025 under review).

Dynamic Tiny Robots through Real-Time MPC on MCUs

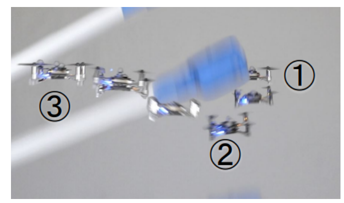

Figure 3: TinyMPC on an MCU enables dynamic obstacle avoidance [16].

Small, low-cost, globally-accessible robots face stringent power, memory, and compute limitations, historically preventing the use of sophisticated algorithms like MPC. As a step toward enabling dynamic planning and control for tiny robots, in work that *won the IEEE ICRA 2024 Best Paper Award in Automation*, my collaborators and I showed that a combination of theoretical and computational advances can compress MPC and enable it to run on the embedded computing devices (e.g., MCUs) found on such tiny robots. Using our TinyMPC algorithm and implementation, we demonstrated high-speed obstacle avoidance and trajectory tracking [16, 17] (Figure 3). I've also taken steps toward overcoming the power limitations on such systems, and have work under review using a combination of different edge systems to enable laser light to power and give high-level commands to tiny robots in the field, enabling more sophisticated and longer-range deployments.

Accelerated Middleware and Benchmarking for Robotic System Architectures

End-to-end robotics applications can only take advantage of accelerated algorithms, like those presented in the previous three lines of research, if they can be run on, and tested through, accelerated middleware and benchmarking frameworks. In this research thrust, my collaborators and I have developed such benchmarking frameworks, middleware interconnects, and runtime systems with adaptive parameter tuning. Importantly, all of these solutions are natively integrated with ROS and ROS2. This enables, for example, end-to-end evaluations that were previously impossible, as much as 4.5x improvements in UAV mission time and energy, and nearly 25% improvements in kernel runtimes when including data-transfer overheads [18, 19, 20, 21].

Accelerated and Memory Efficient Edge Reinforcement Learning

Machine learning-based robotics applications are increasing in prevalence and importance. Whether structured as foundation-models, pixels-to-actions policies, or as learned hyperparameters or models for MPC, these algorithms will need to run on the edge to support the next generation of field robots. This will be particularly challenging for tiny robots which have limited memory and compute [22]. As a step toward alleviating these issues, I have worked to reduce the challenging memory requirements and long runtimes of deep reinforcement learning training. Through differential encodings, quantization, and vectorization, we have enabled as much as a 16.7× reduction in memory and a 32% reduction in latency without impacting training performance [23, 24].

Promoting Responsible and Accessible Robotics

Throughout all of this research, I have worked to promote a responsible and accessible future for global robotics. Beyond open-sourcing all of our software, and running a research lab that is majority undergraduate female, I have: explored research to understand global diversity, equity, inclusion, and belonging in robotics and computing more broadly [25, 26]; designed and documented new interdisciplinary, project-based, open-access courses that lower the barrier to entry of cutting edge topics like robotics and machine learning [27, 28, 29, 30]; and collaboratively promoted a sustainable and privacy preserving future for autonomous systems and edge computing [31, 32, 33, 34].

Future Work – Optimizing Robotic Systems at All Scales

At your college or university, I will pursue interdisciplinary and collaborative robotics research through co-design, building off of my past efforts and extending them to enable increased edge autonomy for global field robots (Figure 4).

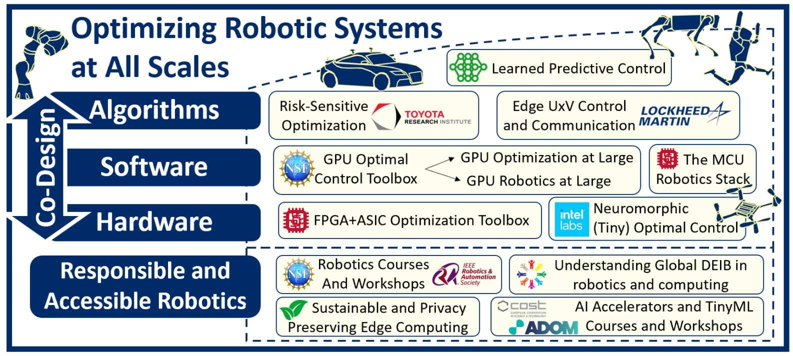

Figure 4: My future research expands the breadth and impact of computational co-design for real-world robotics.

Learned Predictive Control

Maximizing the benefits and minimizing the weaknesses of algorithmic classes (e.g., learning, sampling, and optimization) will be needed to develop high-performance, generalizable, and explainable edge systems, especially where access to cloud computing may be limited or unavailable. Towards this end, I am currently working on ML-guided, mixed-integer solvers to enable real-time contact-implicit MPC, and GPU-batched trajectory optimization for sample-efficient actor critic learning. I also have a collaborative proposal in preparation to develop generalizable drone swarm controllers through provably convergent meta-learning. At your college or university, I will build on these efforts and develop combined algorithmic approaches for efficient edge deployments on field robots.

The Hardware-Accelerated Edge Optimization Toolbox

With support from NSF through my new CSSI grant, and through collaboration with TRI, I will generalize our prior GPU-accelerated work into a broadly applicable toolbox, explicitly consider uncertainty, adding APIs in high level languages, and integrating our solvers with popular machine learning and core robotics frameworks. At your college or university, I aim to also go beyond the GPU and develop the "one-stop-shop" edge optimization toolbox that is broadly accessible, useful, and highly performant across a variety of hardware backends. In fact, my collaborators and I are already developing FPGA-based workflows with the eventual aim of fabricating our own custom robotics ASICs, and have an NSF proposal in preparation to support this work. I will also move beyond our initial focus on MPC, to broader classes of both optimization and robotics problems, ideally though many opportunities for collaboration at your college or university. Finally, I have also begun working with Intel Neuromorphic Computing Labs (NCL) to develop next-generation optimization solvers using custom neuromorphic processors for energy efficient operation.

Next Generation Intelligence for Tiny Robots

I will also build new systems and solutions to overcome the challenges of imbuing tiny robots with advanced intelligence. I am currently extending our TinyMPC planning and control library into a full MCU robotics stack, aiming for a perceptive tiny drone racing demonstration by the end of the academic year. I will also integrate the aforementioned low-power neuromorphic chips onto such tiny drones though novel ASIC fabrication, and build on our laser power and communication project to develop heterogeneous edge robotic systems that enable long duration operation for swarms of tiny robots. Finally, I have a collaborative AFOSR proposal with Lockheed Martin (LM) in preparation to support work on more advanced UxV edge control and communication.

Responsible and Accessible Robotics

At your college or university, I will research and develop new STEM learning models, courses, and outreach programs to improve student access and outcomes in robotics, embedded machine learning (TinyML), and other cutting-edge computing topics. With support from my new NSF CSSI and RAS-TEP grants, I will be developing a series of open-source robotics courses and workshops. I have also been in talks with Adom Inc., to develop the first remote-hands-on electronics course for AI accelerators and have a pending collaborative proposal for an EU COST Action on TinyML. I will also continue to work to understand global DEIB in robotics and computing, as well as promote a sustainable and privacy preserving future for edge computing. Finally, my work on tiny robots will unlock order-of-magnitude cheaper edge intelligence, broadening access to autonomous robotics research globally.

Current and Future Funding Sources

My research has been and will be supported by the NSF (GRFP, CRII, CSSI). I have also been in conversation with, have submitted in the past, and am actively working on whitepapers and collaborative grants for multiple DOD agencies (e.g., AFOSR, ONR, DIU, DARPA), as well as the DOE. I am also planning to continue to pursue awards for early career research (e.g., NSF CAREER, ONR YIP, DARPA YFA, DOE ECRP). I have also assisted multiple students in securing fellowships including the NSF GRFP, NDSEG, and NVIDIA Research, and anticipate that my students will also win such awards. Finally, as demonstrated by my active and pending projects with TRI, Intel NCL, LM, and Adom Inc., I work closely with industrial partners, and anticipate many future collaborative and direct funding opportunities.

References

[1] G.-Z. Yang, B. J. Nelson, R. R. Murphy, H. Choset, H. Christensen, S. H. Collins, P. Dario, K. Goldberg, K. Ikuta, N. Jacobstein, D. Kragic, R. H. Taylor, and M. McNutt, "Combating covid-19—the role of robotics in managing public health and infectious diseases," 2020.

[2] J. Lavelle and B. Ruane, "Gartner hype cycle shows supply chain adoption of mobile robots will far outpace drones over next three years," August 2023.

[3] J. Shaw, "The coming eldercare tsunami," *Harvard Magazine*, 2020.

[4] Y. Tassa, T. Erez, and E. Todorov, "Synthesis and stabilization of complex behaviors through online trajectory optimization," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012.

[5] M. Posa, C. Cantu, and R. Tedrake, "A direct method for trajectory optimization of rigid bodies through contact," *The International Journal of Robotics Research*, 2014.

[6] S. Kuindersma, R. Deits, M. Fallon, A. Valenzuela, H. Dai, F. Permenter, T. Koolen, P. Marion, and R. Tedrake, "Optimization-based locomotion planning, estimation, and control design for the atlas humanoid robot," *Autonomous Robots*, 2016.

[7] **B. Plancher** and S. Kuindersma, "Realtime model predictive control using parallel ddp on a gpu," in *Toward Online Optimal Control of Dynamic Robots Workshop at IEEE International Conference on Robotics and Automation*, 2019.

[8] X. Bu and **B. Plancher**, "Symmetric stair preconditioning of linear systems for parallel trajectory optimization," in *IEEE International Conference on Robotics and Automation*, 2024.

[9] **B. Plancher** and S. Kuindersma, "A performance analysis of parallel differential dynamic programming on a gpu," in *Workshop on the Algorithmic Foundations of Robotics*, 2018.

[10] E. Adabag, M. Atal, W. Gerard, and **B. Plancher**, "Mpcgpu: Real-time nonlinear model predictive control through preconditioned conjugate gradient on the gpu," in *IEEE International Conference on Robotics and Automation*, 2024.

[11] **B. Plancher**, S. M. Neuman, T. Bourgeat, S. Kuindersma, S. Devadas, and V. Janapa Reddi, "Accelerating robot dynamics gradients on a cpu, gpu, and fpga," *IEEE Robotics and Automation Letters*, 2021.

[12] **B. Plancher**, S. M. Neuman, R. Ghosal, S. Kuindersma, and V. Janapa Reddi, "Grid: Gpu-accelerated rigid body dynamics with analytical gradients," in *International Conference on Robotics and Automation*, 2022.

[13] S. M. Neuman, **B. Plancher**, T. Bourgeat, T. Tambe, S. Devadas, and V. Janapa Reddi, "Robomorphic computing: a design methodology for domain-specific accelerators parameterized by robot morphology," in *ACM International Conference on Architectural Support for Programming Languages and Operating Systems*, 2021.

[14] S. M. Neuman, R. Ghosal, T. Bourgeat, **B. Plancher**, and V. Janapa Reddi, "Roboshape: Using topology patterns to scalably and flexibly deploy accelerators across robots," in *International Symposium on Computer Architecture*, 2023.

[15] S. M. Neuman, **B. Plancher**, and V. Janapa Reddi, "Invited: The magnificent seven challenges and opportunities in domain-specific accelerator design for autonomous systems," in *ACM/IEEE Design Automation Conference*, 2024.

[16] K. Nguyen, S. Schoedel, A. Alavilli, **B. Plancher**, and Z. Manchester, "Tinympc: Model-predictive control on resource-constrained microcontrollers," in *IEEE International Conference on Robotics and Automation*, 2024.

[17] S. Schoedel, K. Nguyen, E. Nedumaran, **B. Plancher**, and Z. Manchester, "Code generation for conic model-predictive control on microcontrollers with tinympc," 2024.

[18] B. Boroujerdian, R. Ghosal, J. Cruz, **B. Plancher**, and V. Janapa Reddi, "Roborun: A robot runtime to exploit spatial heterogeneity," in *ACM/IEEE Design Automation Conference (DAC)*, 2021.

[19] B. Boroujerdian, H. Genc, S. Krishnan, B. P. Duisterhof, **B. Plancher**, K. Mansoorshahi, M. Almeida, W. Cui, A. Faust, and V. Janapa Reddi, "The role of compute in autonomous micro aerial vehicles: Optimizing for mission time and energy efficiency," *ACM Transactions on Computing Systems*, jul 2022.

[20] V. Mayoral-Vilches, S. M. Neuman, **B. Plancher**, and V. Janapa Reddi, "Robotcore: An open architecture for hardware acceleration in ros 2," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2022.

[21] V. Mayoral-Vilches, J. Jabbour, Y.-S. Hsiao, Z. Wan, M. Crespo-Álvarez, M. Stewart, J. M. Reina-Muñoz, P. Nagras, G. Vikhe, M. Bakhshalipour, M. Pinzger, S. Rass, S. Panigrahi, G. Corradi, N. Roy, P. B. Gibbons, S. M. Neuman, **B. Plancher**, and V. Janapa Reddi, "Robotperf: An open-source, vendor-agnostic, benchmarking suite for evaluating robotics computing system performance," in *IEEE International Conference on Robotics and Automation*, Yokohama, Japan, 2024.

[22] S. M. Neuman, **B. Plancher**, B. P. Duisterhof, S. Krishnan, C. Banbury, M. Mazumder, S. Prakash, J. Jabbour, A. Faust, G. C. de Croon, *et al.*, “Tiny robot learning: Challenges and directions for machine learning in resource-constrained robots,” in *2022 IEEE 4th International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, 2022.

[23] L. Grossman and **B. Plancher**, “Just round: Quantized observation spaces enable memory efficient learning of dynamic locomotion,” in *IEEE International Conference on Robotics and Automation*, London, UK, 2023.

[24] L. Grossman and **B. Plancher**, “Differentially encoded observation spaces for perceptive reinforcement learning,” in *IEEE International Conference on Robotics and Automation*, Yokohama, Japan, 2024.

[25] E. Fields, C. Ho, M. J. Kim, Z. Wu, and **B. Plancher**, “Underrepresentation of women in robotics research [women in engineering],” *IEEE Robotics & Automation Magazine*, 2024.

[26] E. Selman-Housein and **B. Plancher**, “Improving the representation of undergraduate women in cybersecurity: A literature review,” in *ACM Technical Symposium on Computer Science Education*, 2025.

[27] S. Karaman, A. Anders, M. Boulet, J. Connor, K. Gregson, W. Guerra, O. Guldner, M. Mohamoud, **B. Plancher**, R. Shin, and J. Vivilecchia, “Project-based, collaborative, algorithmic robotics for high school students: Programming self-driving race cars at mit,” in *IEEE integrated STEM education conference*, 2017.

[28] V. Janapa Reddi, **B. Plancher**, S. Kennedy, L. Moroney, P. Warden, L. Suzuki, A. Agarwal, C. Banbury, M. Banzi, M. Bennett, B. Brown, S. Chitlangia, R. Ghosal, S. Grafman, R. Jaeger, S. Krishnan, M. Lam, D. Leiker, C. Mann, M. Mazumder, D. Pajak, D. Ramaprasad, J. E. Smith, M. Stewart, and D. Tingley, “Widening access to applied machine learning with tinyml,” *Harvard Data Science Review*, 1 2022, https://hdsr.mitpress.mit.edu/pub/0gbwdele.

[29] **B. Plancher**, S. Buttrich, J. Ellis, N. Goveas, L. Kazimierski, J. Lopez Sotelo, M. Lukic, D. Mendez, R. Nordin, A. Oliva Trevisan, M. Pavan, M. Roveri, M. Rüb, J. Tum, M. Verhelst, S. Abdeljabar, S. Adebayo, T. Amberg, H. Aworinde, J. Bagur, G. Barrett, N. Benamar, B. Chaudhari, R. Criollo, D. Cuartielles, J. A. Ferreira Filho, S. Gizaw, E. Gousev, A. Grande, S. Hymel, P. Ing, P. Manandhar, P. Manzoni, B. Murmann, E. Pan, R. Paskauskas, E. Pietrosemoli, T. Pimenta, M. Rovai, M. Zennaro, and V. Janapa Reddi, “Tinyml4d: Scaling embedded machine learning education in the developing world,” in *AAAI Spring Symposium Series*, 2024.

[30] **B. Plancher**, “Parallel optimization for robotics: An undergraduate introduction to gpu parallel programming and numerical optimization research,” in *IEEE International Parallel and Distributed Processing Symposium Workshops*, 2024.

[31] S. Prakash, M. Stewart, C. Banbury, M. Mazumder, P. Warden, **B. Plancher**, and V. Janapa Reddi, “Is tinyml sustainable? assessing the environmental impacts of machine learning on microcontrollers,” *Communications of the ACM*, Nov 2023.

[32] P. Warden, M. Stewart, **B. Plancher**, S. Katti, and V. Janapa Reddi, “Machine learning sensors: A design paradigm for the future of intelligent sensors,” *Communications of the ACM*, Nov 2023.

[33] M. Stewart, P. Warden, Y. Omri, S. Prakash, J. Santos, S. Hymel, B. Brown, J. MacArthur, N. Jeffries, **B. Plancher**, and V. Janapa Reddi, “Datasheets for machine learning sensors,” 2023.

[34] M. Stewart, E. Moss, P. Warden, **B. Plancher**, S. Kennedy, M. Sloane, and V. Janapa Reddi, “Materiality and risk in the age of pervasive ai sensors,” 2024.