I am a *computer systems researcher focused on improving the sustainability of computing*. My work rethinks the design and operation of computer systems to address emerging challenges, including rapidly rising computing demand, increasing grid power constraints, and growing complexity in datacenter energy systems. These are challenges our current infrastructure cannot solve. To tackle these issues, I take a multidisciplinary approach that uniquely integrates domain-specific knowledge with advanced computing methods to develop impactful solutions. In manifesting real-world impact, my work has enhanced the resource efficiency of hyperscale datacenters [1] and powered testbeds for carbon-efficient applications [2]. In my work, I utilize the system stack of software-defined infrastructure, distributed systems, resource management, and performance evaluation.

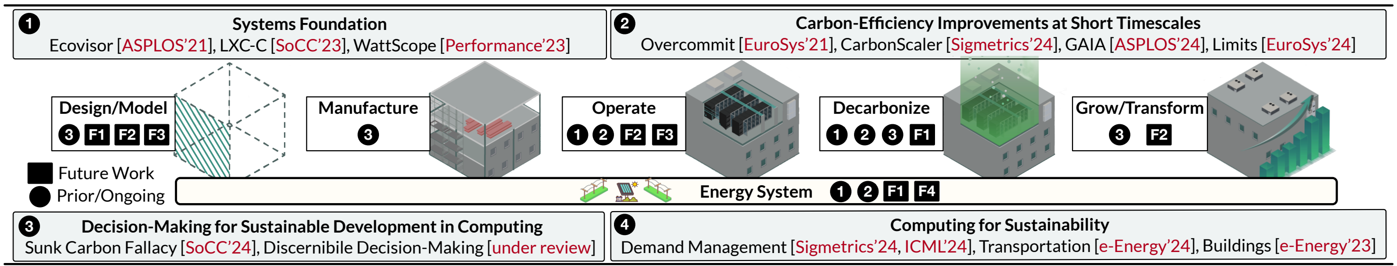

Fig.1: **Intersection of my prior (❶ – ❹) and future work (F1 – F4) with computer systems' lifecycle stages.**

Research Overview Over the past two decades, energy efficiency optimizations have increased computing's economic productivity but have not reduced its aggregate energy demand or environmental impact. Additionally, we are approaching several physical limits of energy efficiency that have tempered growth in energy demand, with the end of Dennard Scaling and the slowdown of Moore's Law [3, 4]. Moving forward, *mitigating computing's lifecycle environmental impact will require prioritizing carbon efficiency – measured by the work done per unit of carbon (and other greenhouse gases) emitted,* in addition to exploiting the dwindling opportunities to improve energy efficiency [5]. To enhance carbon efficiency, computing's operational emissions (produced by energy use) can be reduced by doing more work when and where low-carbon energy is available. Reducing embodied carbon emissions (from the production and disposal of computing hardware and infrastructure) requires reevaluating hardware design, procurement, and capacity provisioning strategies. However, optimizing operational carbon, embodied carbon, and their tradeoffs to improve computing's carbon efficiency presents fundamental computer systems challenges. My work has made key contributions to addressing those challenges; I discuss a few representative contributions below.

❶ **Systems Foundations.** Computing applications lack visibility and control over their energy supply, preventing them from adjusting power usage based on energy's carbon intensity or renewable energy availability. To address this issue, I have *done foundational work on virtualizing datacenter energy systems and exposing software-defined control to applications* [6]. Inspired by Exokernel, these abstractions enable applications to manage clean energy's variability within their software stack directly, aligning performance needs with sustainability goals by leveraging one or more dimensions of software flexibility and fault tolerance. `Ecovisor`'s software ecosystem is open-source and deployed on a community testbed [2, 7–9].

❷ **Carbon Efficiency Improvements at Short Timescales.** At timescales ranging from seconds to days, reducing the carbon footprint requires continuously optimizing workload execution to enhance applications' carbon efficiency. In developing carbon-aware scheduling policies, my work highlighted that simultaneously optimizing for carbon, energy, and performance is impossible. Using this insight, I have *designed systems for various applications that strategically trade energy or performance to achieve carbon-efficiency improvements.* For instance, I developed `CarbonScaler` [10], a carbon-aware autoscaler that scales up (energy inefficiently) during low-carbon periods (carbon efficient) without increasing job completion time (maintaining performance). The artifacts from my work on carbon-aware policies are available as open-source [7, 11].

❸ **Decision-Making for Sustainable Development in Computing.** The decision-making for sustainable development in computing , such as greener chip design, server procurement, and datacenter siting, requires accurate data on operational and embodied carbon footprint. However, real-world estimates are fraught with uncertainty, which prior work tends to overlook. My work has *built tools to quantify the uncertainty in embodied and operational carbon estimates and highlight its impact on decision discernibility.* In this effort, I extended the Product Attributes and Impact Algorithm (PAIA), a lifecycle analysis (LCA) tool used by ICT companies [12], to conduct uncertainty-driven quantitative assessments of key decisions in sustainable computing [13]. I have also extensively evaluated the incentives provided by various carbon-based metrics and their impact on holistic carbon footprint reduction [14].

❹ **Computing for Sustainability.** During my Ph.D., I *extensively used computational tools, including analytical modeling, applied machine learning, algorithm design, and system prototyping, to enable decarbonization of energy systems in other societal sectors*, such as the electric grid [15], residential buildings [16], and transportation [17].

Across these threads, I have enabled computing stakeholders to reliably quantify and significantly reduce the lifecycle environmental impact of AI demand while demonstrating how AI capabilities can accelerate societal decarbonization.

Prior & Ongoing Work Figure 1 describes the main threads in my work that span technical contributions in systems, continuous scheduling optimizations, uncertainty-aware decision-making, and computing for sustainability.

❶ – Systems Foundations | Rep. Work: `Ecovisor` [6]. A few fundamental challenges hinder carbon efficiency optimizations in existing systems: i) limited visibility into both their operational and embodied carbon footprint; ii) insufficient flexibility to adapt to variations in carbon intensity and energy availability; iii) inadequate programmability to expose software-defined interfaces for automated monitoring and control of energy and computing resources. These gaps in visibility, flexibility, and programmability obstruct the development of carbon-efficient applications.

To address these challenges, I *designed abstractions that virtualize the energy system to provide applications with visibility and software-defined control of physical energy systems* (Figure 2). Specifically, I developed the `Ecovisor` prototype [6], which extends a Container Orchestration Platform (COP) to enable fine-grained monitoring and control of physical energy system components by exposing privileged API access to applications. The `Ecovisor` prototype also supports stateful applications and vertical scaling through LXD. It operates on a central server, exposing a REST API that allows applications to register callback functions for energy and carbon monitoring and control. Virtualizing the energy system required solving unique challenges related to software-defined power management in solar- and storage-based grid-connected energy systems. For example, I developed the `Ecovisor` prototype as a custom hardware-software solutions using off-the-shelf battery charge controllers, direct current (DC) power supplies, and inverters. I also used Helios, a programmable software-defined solar module that I previously designed [18], which provides a high-level interface to a DC-DC power optimizer, enabling remote, real-time control of a solar module's output via a REST API. Finally, I evaluated the `Ecovisor` prototype to demonstrate its benefits for a wide range of applications. As one example, `Ecovisor` reduced the carbon footprint of training ResNet34 in PyTorch by 20%, comparable to a system-level suspend-resume policy, while reducing training time by 2.58×.

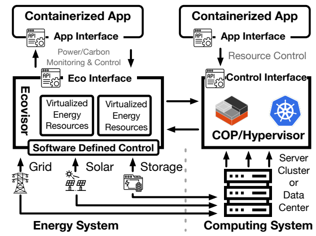

Fig.2: ***Ecovisor's architecture.***

`Ecovisor` enables the broadest range of policies by default to accommodate users and applications with varying characteristics, goals, strategies, and tolerances for reducing carbon and energy. However, most applications do not need to modify their logic and interact directly with the virtual energy system – third-party libraries and services can manage this complexity. To illustrate this, I *developed Carbon Containers (`LXC-C`), a system-level abstraction that dynamically regulates application-level carbon emissions* through performance-aware vertical scaling and migrations [9]. I also developed supporting tools to enable non-intrusive energy monitoring, thermal energy management, and a fair distributed rate control [19–21].

❷ – Carbon-Efficiency Improvements at Short Timescales | Rep. Work: Overcommit [1] . Over the span of seconds-to-days, the carbon footprint of computing and AI workloads can be reduced by improving efficiency in two dimensions: first, doing more work with the same hardware to improve embodied carbon efficiency; second, doing more work with the same amount of energy or carbon to enhance operational carbon efficiency.

Improving Embodied Carbon Efficiency. Datacenters often overcommit resources – that is, they schedule more tasks on a server than its physical capacity – to increase typically low utilization rates of 30% to 60% [22]. However, determining the right level of overcommitment is challenging: undercommitting wastes resources, while overcommitting risks performance degradation. Despite research on advanced overcommit policies, production policies remain simple due to the need for lightweight solutions and the significant effort required to test arbitrary policies. To address these challenges in Google's Borg scheduler, I *developed dynamic overcommit policies that predict future resource usage and adjust the per-machine overcommit factor using lightweight predictors* [1]. To test an arbitrary predictor before deployment in production, I built an Apache Beam-based open-source simulator that mimics Borg's machine-level component, Borglet. To quantify potential performance degradation, I introduced violation rate, a metric that correlates with 99th percentile CPU scheduling latency – an important Borg Service Level Objective (SLO) related to task wait times in the OS CPU scheduler. *Initial deployment on 25,000 machines increased usable capacity by 10-16%, leading to a full rollout across all Google datacenters, reducing overall infrastructure needs.* I am currently collaborating with researchers at Google to further enhance overcommitment by leveraging workload periodicity [23].

Improving Operational Carbon Efficiency. Improving operational carbon efficiency requires a nuanced approach: when systems are energy inefficient, increasing energy efficiency directly enhances carbon efficiency; otherwise, improving carbon efficiency may necessitate sacrificing energy efficiency [24]. My analysis revealed that while there are limits, the carbon footprint of computing can be significantly reduced by strategically exploiting an application's flexibility based on its carbon intensity variability, performance flexibility, and workload characteristics [25]. To illustrate this for federated learning (FL), I *designed `EcoLearn`, which minimizes the carbon footprint of FL while maintaining model accuracy and training time* [26]. `EcoLearn` achieved this by integrating carbon awareness into multiple aspects of FL training, including (i) selecting clients with high data utility and low carbon, (ii) provisioning more clients during initial training rounds, and (iii) mitigating stragglers using dynamic carbon-aware client over-provisioning. I implemented `EcoLearn` in the Flower framework and showed that it reduced the carbon footprint of training by up to $10.8\times$, while maintaining model accuracy and training time within $\sim 1\%$ of state-of-the-art carbon-agnostic approaches. Finally, I demonstrated that these carbon footprint reduction opportunities are available for a broad set of workloads – such as data processing, scientific computing, and AI training – and in various computing environments, such as public clouds, on-premise datacenters, edge computing, and hybrid clouds [10, 27–33].

❸ – Decision-Making for Sustainable Development in Computing | Rep. Work: Sunk Carbon Fallacy [14] Sustainable development in computing requires rethinking decisions around hardware design, server procurement, datacenter placement, and carbon accounting [34–36]. Unlike continuous short-term optimizations, these decisions are made infrequently over months or years but have long-lasting effects. In this thread, I have worked on enabling discernable decision-making under uncertainty and analyzed the choice of metrics for various decisions in computing's lifecycle.

I have shown that embodied and operational carbon estimates, which inform such decisions, are riddled with uncertainty due to the complexity of supply chains and variability in computer systems' real-world performance [13]. In quantifying uncertainty for the embodied carbon, I extended the Product Attributes and Impact Algorithm (PAIA), a lifecycle analysis tool used by ICT companies, to generate distributions of estimates based on chip size, fabrication location, technology node, and other key attributes. For operational carbon, I demonstrated that performance variability can significantly impact operational carbon estimates, even in highly controlled environments. To reduce this uncertainty, *I tailored the embodied and operational carbon distributions to be decision-specific, exposing uncertainty only in attributes that decision-makers can affect*. For example, emissions from server transportation can be ignored when designing chips, while application performance variations can be set aside when procuring servers. Finally, for an example of choosing a processor, I showed that there is only 1% chance of discernibly picking a low-carbon processor, even if the chosen processor has almost 10% less expected embodied carbon value.

The metrics quantifying computing's carbon footprint and making decisions are still evolving. I have evaluated these metrics to assess the incentives they create for carbon-aware optimizations and their impact on holistic carbon footprint. For example, I showed that while lifecycle emissions metrics help with procurement, they can increase emissions by 24% when applied to job placement [14]. I also examined how confidential power purchase agreements (PPAs) influence public estimates of carbon intensity, leading to overestimating carbon reductions by up to 55.1% and inadvertently increasing emissions by over 3% for carbon-aware optimizations [37]. Collectively, my research enhances the understanding of computing's environmental implications and provides tools for responsible development in computing and AI [34].

❹ – Computing for Sustainability The net sustainability implications of computing and AI tools can be improved by using them to quantify and reduce emissions in other societal sectors. I have done extensive work on using computational methods to accelerate the decarbonization of the electric grid, buildings, and transportation sectors. First, I have *designed spatiotemporal scheduling algorithms for learning-augmented online optimization that use AI predictions to achieve better average-case performance without sacrificing worst-case competitive guarantees* [38–41]. These algorithms apply to a broad set of sustainability problems, such as carbon-aware electric vehicle charging and computing workload execution. Second, to improve the programmability of networked energy systems, I *used physical models and data-driven ML methods* for solar PV performance modeling and forecasting [15, 42, 43], Bayesian methods for anomaly detection in solar panels [44, 45], and distributed rate control approaches from computer networks for controlling distributed solar capacity [21, 46]. Third, to support energy transition in the buildings sector, I have developed tools for *tactical energy transition* from gas-based heating to electric heat pumps [16, 47], incentive design for solar energy adoption [48–50], and devising smart load-shedding solutions [51, 52]. Finally, I have devised *carbon- and equity-aware ride assignment policies* for ridesharing platforms [17, 53, 54].

Future Research Directions My foundational work on building systems, carbon-efficient applications, and frameworks for sustainable computing and AI is evergreen. However, its capabilities must expand, and its application to improving computing's carbon efficiency will need to evolve, presenting a rich set of challenges as the insatiable demand for AI workloads grows, new application frameworks emerge, tangible incentives to reduce carbon footprints are introduced, or planetary limits on materials and emissions are reached. Below, I envision my research addressing existing and future challenges in enabling sustainable AI.

F1 – Designing and Operating Sustainable Datacenters. Meeting the growing demand for AI workloads responsibly requires a sustainability-aware, multidisciplinary approach that both looks inward to address user, application, and infrastructure challenges in datacenters and outward to understand electric grid constraints and the challenges it faces for a reliable operation.

My work will lay the groundwork for sustainable datacenters by benchmarking their architecture, hardware, and workloads to assess tradeoffs between performance, energy, and sustainability metrics, such as carbon and water footprint. In this effort, I will develop system support that enables datacenters to automatically adjust their operations at minimum performance impact based on grid conditions at short time scales while optimizing design for mutually beneficial coordination with the electric grid over the long term. I will develop higher-level frameworks for the sustainable operation of datacenters, respecting the constraints and objectives of users, datacenter operators, and grid utilities. This will involve creating carbon-centered service level agreements (cSLAs) that allow cloud platforms to offer sustainable solutions while enabling users to optimize sustainability and financial goals. Finally, I will develop frameworks for datacenter–grid coordination, leveraging game-theoretic approaches to design datacenter demand response solutions that offer meaningful incentives for participation.

F2 – Co-Adapting Emerging Applications and Heterogeneous Hardware. Modern cloud-native and AI inference-driven applications are being deployed on heterogeneous (specialized and aging) hardware and are driving much of the increase in computing demand. Prior work does not tackle optimizing the carbon efficiency of this evolving software-hardware ecosystem.

Interestingly, the defining aspects of emerging applications – scalability, resiliency, and redundancy – are also desired characteristics for using power from highly variable and unreliable renewable energy sources. I will develop new abstractions for developers and cloud operators to deploy modern applications on specialized (for performance) and aging (to reduce embodied carbon) hardware run on intermittent renewable power (to reduce operational carbon) while balancing ease of use against deep optimizations. However, these applications' sheer scale and distributed nature make deploying any optimizations challenging, and simple heuristics do not work well. While modern black-box ML/AI tools for systems are being used, they are generally reserved for non-critical and particular use cases due to their poor generalization and lack of worst-case guarantees. I will continue my work on combining AI advice with robust algorithms to provide good average-case performance and worst-case guarantees when using AI tools for systems, ensuring they remain adaptive and robust in dynamic and uncertain environments.

F3 – Digital Twins-in-the-Loop (DTIL) Datacenters. There is no production-scale deployment of sustainable computing solutions, such as spatiotemporal workload migrations and datacenter demand response. Beyond the lack of incentives, the key obstacle is an obscured view of the financial, technical, and infrastructure implications of sustainability-driven optimizations.

My research will lead an effort to build holistic end-to-end models for distributed datacenter infrastructure that accounts for cost, energy, carbon, and performance impacts of carbon-aware optimizations. For instance, I will develop realistic models for workload migrations' cost and carbon impacts in the network. Similarly, I will also focus on the challenging and pressing issue of modeling the tradeoffs between redundancy, availability, and embodied carbon. The modeling effort will drive a reassessment of the relentless pursuit of marginal gains in performance – without demonstrated system-scale benefits – and identify the realistic carbon-aware optimizations at scale. In the long run, guided by the modeling efforts, I aim to leverage AI tools to design digital twin-in-the-loop (DTiL) datacenters that have a symbiotic relationship with the physical infrastructure. The DTiL datacenters will use digital twins to optimize operations at short timescales and inform design at long timescales.

F4 – Computing-Energy-Society Nexus. A push towards electrification and embedded intelligence across various societal sectors creates interdependencies that did not exist before. For instance, smart electric cars are changing the landscape of personal transportation: they require computing resources across the stack (device, edge, cloud) and create electricity couplings between residential and commercial buildings. Ultimately, datacenters, the electric grid, and other societal sectors are increasingly coupled due to reliance on computing and electric grids. This means that a siloed focus on improving resource efficiency in each sector is unlikely to yield practical solutions that lead to societal-scale decarbonization [55]. My research will design computing solutions (infrastructure and software) that power computational approaches to holistic cross-domain decarbonization. The geographical distribution of demand (e.g., roadside, buildings, cloud-based) and its varying temporal characteristics (e.g., ephemeral on the roadside, periodic in buildings, continuous analytics in the cloud) will require hardware-software solutions that sustainably serve cross-sectoral approaches instead of domain-specific over-provisioned solutions.

Concluding Remarks The defining technological trend of our times is the AI-driven exponential growth in computing demand and the challenges the current infrastructure faces in sustainably satisfying that demand. Consequently, my research topics are the top priorities for technology companies, academic institutions, the National Science Foundation, and the Department of Energy. However, computing's lifecycle emissions span and intersect multiple societal sectors, meaning that *no single field of science, government entity, or large tech company can mitigate the existential threat posed by climate change and computing's growing contributions to it*. As a result, I have fully embraced the power of a multidisciplinary approach to problem-solving, uniting computer science, engineering, and traditional sciences to accelerate the development of sustainable solutions. My role as the Computing & Climate Impact Fellow at the highly interdisciplinary MIT Climate & Sustainability Consortium (MCSC) exemplifies it. I am uniquely positioned to make impactful contributions to the field of sustainable computing due to my interdisciplinary academic training (electrical power engineering, energy systems engineering, and computer science), a track record of impactful publications across domains (ACM SIGs on Energy, Metrics, Architecture, Operating Systems, and Computers & Society), and industry collaborations (Google, Amazon, Meta, IBM, VMWare, Telefonica, and many startups).

References

[1] **Noman Bashir**, Nan Deng, Krzysztof Rzadca, David Irwin, Sree Kodak, and Rohit Jnagal. "Take it to the Limit: Peak Prediction-driven Resource Overcommitment in Datacenters". In: ACM EuroSys. 2021.

[2] David Irwin, Prashant Shenoy, and Michael Zink. *CCRI: New: A Community Testbed for Designing Carbon-Efficient Cloud Applications*. `https://www.nsf.gov/awardsearch/showAward?AWD_ID=2213636`. (Accessed on October 8, 2024). 2022.

[3] Charles E. Leiserson, Neil C. Thompson, Joel S. Emer, Bradley C. Kuszmaul, Butler W. Lampson, Daniel Sanchez, and Tao B. Schardl. "There's plenty of room at the Top: What will drive computer performance after Moore's law?" In: *Science* (2020).

[4] Hadi Esmaeilzadeh, Emily Blem, Renee St. Amant, Karthikeyan Sankaralingam, and Doug Burger. "Dark Silicon and the End of Multicore Scaling". In: *SIGARCH Comput. Archit. News* (2011).

[5] **Noman Bashir**, Tian Guo, Mohammad Hajiesmaili, David Irwin, Prashant Shenoy, Ramesh Sitaraman, Abel Souza, and Adam Wierman. "Enabling Sustainable Clouds: The Case for Virtualizing the Energy System". In: ACM SoCC. 2021.

[6] Abel Souza, **Noman Bashir**, Jorge Murillo, Walid Hanafy, Qianlin Liang, David Irwin, and Prashant Shenoy. "Ecovisor: A Virtual Energy System for Carbon-Efficient Applications". In: ACM ASPLOS. 2023.

[7] *Ecovisor Software Prototype*. `https://github.com/carbonfirst/Ecovisor`. 2023.

[8] *Carbon Containers Software Prototype*. `https://github.com/carbonfirst/CarbonContainers`. 2023.

[9] John Thiede, **Noman Bashir**, David Irwin, and Prashant Shenoy. "Carbon Containers: A System-level Facility for Managing Application-level Carbon Emissions". In: ACM SoCC. 2023.

[10] Walid Hanafy, Qianlin Liang, **Noman Bashir**, David Irwin, and Prashant Shenoy. "CarbonScaler: Leveraging Cloud Workload Elasticity for Optimizing Carbon-Efficiency". In: ACM SIGMETRICS. 2023. **Best Student Paper Award**.

[11] *CarbonScaler Software Prototype*. `https://github.com/umassos/CarbonScaler`. 2024.

[12] *Product Attributes to Impact Algorithm*. `https://paia.mit.edu/`. 2024.

[13] **Noman Bashir**, Anagha B. Subramanya, Julia Xia, Varun Gohil, Ajay Gupta, Melissa Zgola, Greg Norris, Elsa Olivetti, and Christina Delimitrou. "Discernible Decision Making under Uncertainty in Sustainable Computing". In: preparation. 2024.

[14] **Noman Bashir**, Varun Gohil, Mohammad Shahrad, David Irwin, Anagha B. Subramanya, Elsa Olivetti, and Christina Delimitrou. "The Sunk Carbon Fallacy: Rethinking Carbon Footprint Metrics for Effective Carbon-Aware Scheduling". In: ACM SoCC. 2024.

[15] **Noman Bashir**, Dong Chen, David Irwin, and Prashant Shenoy. "Solar-TK: A Data-Driven Toolkit for Solar PV Performance Modeling and Forecasting". In: IEEE MASS. 2019.

[16] Adam Lechowicz, **Noman Bashir**, John Wamburu, Mohammad Hajiesmaili, and Prashant Shenoy. "Equitable Network-Aware Decarbonization of Residential Heating at City Scale". In: ACM e-Energy. 2023.

[17] Mahsa Sahebdel, Ali Zeynali, **Noman Bashir**, Prashant Shenoy, and Mohammad Hajiesmaili. "LEAD: Towards Learning-Based Equity-Aware Decarbonization in Ridesharing Platforms". In: submission. 2024.

[18] **Noman Bashir**, David Irwin, and Prashant Shenoy. "Helios: A Programmable Software-defined Solar Module". In: ACM BuildSys. 2018.

[19] Xiaoding Guan, **Noman Bashir**, David Irwin, and Prashant Shenoy. "WattScope: Non-intrusive Application-level Power Disaggregation in Datacenters". In: IFIP Performance. 2023.

[20] **Noman Bashir**, Yasra Chandio, David Irwin, Fatima M. Anwar, Jeremy Gummeson, and Prashant Shenoy. "Jointly Managing Electrical and Thermal Energy in Solar- and Battery-powered Computer Systems". In: ACM e-Energy. 2023.

[21] **Noman Bashir**, David Irwin, Prashant Shenoy, and Jay Taneja. "Enforcing Fair Grid Energy Access for Controllable Distributed Solar Capacity". In: ACM BuildSys. 2017.

[22] Muhammad Tirmazi, Adam Barker, Nan Deng, Md E. Haque, Zhijing Gene Qin, Steven Hand, Mor Harchol-Balter, and John Wilkes. "Borg: the Next Generation". In: ACM EuroSys. 2020.

[23] Xiaoding Guan, **Noman Bashir**, Prashant Shenoy, and David Irwin. "Ahead of the Curve: Leveraging Periodicity to Improve Job Scheduling in Data Centers". In: preparation. 2024.

[24] Walid Hanafy, Roozbeh Bostandoost, **Noman Bashir**, David Irwin, Mohammad Hajiesmaili, and Prashant Shenoy. "The War of the Efficiencies: Understanding the Tension between Carbon and Energy Optimization". In: HotCarbon. 2023.

[25] Thanathorn Sukprasert, Abel Souza, **Noman Bashir**, David Irwin, and Prashant Shenoy. "On the Limitations of Carbon-Aware Temporal and Spatial Workload Shifting in the Cloud". In: ACM EuroSys. 2024.

[26] Talha Mehboob, **Noman Bashir**, Jesus Omana Iglesias, Michael Zink, and David Irwin. "EcoLearn: Optimizing the Carbon Footprint of Federated Learning". In: submission. 2024.

[27] Pradeep Ambati, **Noman Bashir**, David Irwin, and Prashant Shenoy. "Good Things Come to Those Who Wait: Optimizing Job Waiting in the Cloud". In: ACM SoCC. 2021.

[28] Pradeep Ambati, **Noman Bashir**, David Irwin, and Prashant Shenoy. "Waiting Game: Optimally Provisioning Fixed Resources for Cloud-Enabled Schedulers". In: SC. 2020.

[29] Walid Hanafy, Qianlin Liang, **Noman Bashir**, Abel Souza, David Irwin, and Prashant Shenoy. "Going Green for Less Green: Optimizing the Cost of Reducing Cloud Carbon Emissions". In: ACM ASPLOS. 2024.

[30] Qianlin Liang, Walid Hanafy, **Noman Bashir**, Ahmed Ali-Eldin, David Irwin, and Prashant Shenoy. "Dělen: Enabling Flexible and Adaptive Model-serving for Multi-tenant Edge AI". In: ACM/IEEE IoTDI. 2023.

[31] Qianlin Liang, Walid Hanafy, **Noman Bashir**, David Irwin, and Prashant Shenoy. "Energy Time Fairness: Balancing Fair Allocation of Energy and Time for GPU Workloads". In: IEEE/ACM SEC. 2023.

[32] **Noman Bashir**, Adam Lechowicz, Rohan Shenoy, Mohammad Hajiesmaili, Adam Wierman, and Christina Delimitrou. "Learning Carbon-Aware Scheduling Algorithms for Data Processing Clusters". In: preparation. 2024.

[33] Varun Gohil, **Noman Bashir**, and Christina Delimitrou. "URJA: Request-Level Power Capping for Microservice". In: preparation. 2024.

[34] **Noman Bashir**, Priya Donti, James Cuff, Sydney Sroka, Marija Ilic, Vivienne Sze, Christina Delimitrou, and Elsa Olivetti. "The Climate and Sustainability Implications of Generative AI". In: An MIT Exploration of Generative AI (2024).

[35] **Noman Bashir**, David Irwin, Prashant Shenoy, and Abel Souza. "Sustainable Computing – Without the Hot Air". In: HotCarbon. 2022.

[36] Yichen Gao, **Noman Bashir**, Christopher Hill, and Jeremy Gregory. "Enabling Proactive Sustainability Interventions in Datacenters". In: submission. 2024.

[37] Diptyaroop Maji, **Noman Bashir**, David Irwin, Prashant Shenoy, and Ramesh K Sitaraman. "The Green Mirage: Impact of Location- and Market-based Carbon Intensity Estimation on Carbon Optimization Efficacy". In: ACM e-Energy. 2024.

[38] Adam Lechowicz, Nicolas Christianson, Bo Sun, **Noman Bashir**, Mohammad Hajiesmaili, Adam Wierman, and Prashant Shenoy. "Online Conversion with Switching Costs: Robust and Learning-Augmented Algorithms". In: ACM SIGMETRICS. 2024.

[39] Adam Lechowicz, Nicolas Christianson, Jinhang Zuo, **Noman Bashir**, Mohammad Hajiesmaili, Adam Wierman, and Prashant Shenoy. "The Online Pause and Resume Problem: Optimal Algorithms and An Application to Carbon-Aware Load Shifting". In: ACM SIGMETRICS. 2023.

[40] Adam Lechowicz, Nicolas Christianson, Bo Sun, **Noman Bashir**, Mohammad Hajiesmaili, Adam Wierman, and Prashant Shenoy. "Chasing Convex Functions with Long-term Constraints". In: ICML. 2024.

[41] Adam Lechowicz, Nicolas Christianson, Bo Sun, **Noman Bashir**, Mohammad Hajiesmaili, Adam Wierman, and Prashant Shenoy. "CarbonClipper: Optimal Algorithms for Carbon-aware Spatiotemporal Workload Management". In: submission. 2024.

[42] **Noman Bashir**, David Irwin, and Prashant Shenoy. "DeepSnow: Modeling the Impact of Snow on Solar Generation". In: ACM BuildSys. 2020.

[43] **Noman Bashir**, David Irwin, and Prashant Shenoy. "A Probabilistic Approach to Committing Solar Energy in Day-ahead Electricity Markets". In: IGSC/SUSCOM (2021).

[44] Menghong Feng, **Noman Bashir**, Prashant Shenoy, David Irwin, and Dragoljub Kosanovic. "SunDown: Model-driven Per-Panel Solar Anomaly Detection for Residential Arrays". In: ACM COMPASS. 2020.

[45] Menghong Feng, **Noman Bashir**, Prashant Shenoy, David Irwin, and Beka Kosanovic. "Model-driven Per-panel Solar Anomaly Detection for Residential Arrays". In: ACM TCPS (2021).

[46] **Noman Bashir**, David Irwin, Prashant Shenoy, and Jay Taneja. "Mechanisms and Policies for Controlling Distributed Solar Capacity". In: ACM TOSN. 2018.

[47] John Wamburu, **Noman Bashir**, David Irwin, and Prashant Shenoy. "Data-driven Decarbonization of Residential Heating Systems". In: ACM BuildSys. 2022.

[48] Anupama Sitaraman, **Noman Bashir**, David Irwin, and Prashant Shenoy. "No Free Lunch: Analyzing the Cost of Deep Decarbonizing Residential Heating Systems". In: IGSC. 2023. **Best Student Paper Award**.

[49] Cooper Sigrist, Adam Lechowicz, Jovan Champ, **Noman Bashir**, and Mohammad Hajiesmaili. "Lost in Siting: The Hidden Carbon Cost of Inequitable Residential Solar Installations". In: submission. 2024.

[50] Anupama Sitaraman, Adam Lechowicz, **Noman Bashir**, Xutong Liu, Mohammad Hajiesmaili, and Prashant Shenoy. "Dynamic Incentive Allocation for City-Scale Deep Decarbonization". In: submission. 2024.

[51] **Noman Bashir**, Hira Shahzad Sardar, Mashood Nasir, Naveed Ul Hassan, and Hassan A. Khan. "Lifetime Maximization of Lead-Acid Batteries in Small Scale UPS and Distributed Generation Systems". In: IEEE PowerTech. 2017.

[52] **Noman Bashir**, Zohaib Sharani, Khushboo Qayyum, and Affan A. Syed. "Delivering Smart Load-shedding for Highly-stressed Grids". In: IEEE SmartGridComm. 2015.

[53] Mahsa Sahebdel, Ali Zeynali, **Noman Bashir**, Prashant Shenoy, and Mohammad Hajiesmaili. "A Holistic Approach for Equity-aware Carbon Reduction of the Ridesharing Platforms". In: ACM e-Energy. 2024.

[54] Ali Zeynali, Mahsa Sahebdel, **Noman Bashir**, Ramesh Sitaraman, and Mohammad Hajiesmaili. "Near-Optimal Emission-Aware Online Ride Assignment Algorithm for Peak Demand Hours". In: submission. 2024.

[55] Prashant Shenoy, Andrew A Chien, David Irwin, Mohammad Hajiesmaili, Vivienne Sze, Mani Srivastava, Line Roald, Yuvraj Agarwal, Rick Adrion, Ramesh Sitaraman, Neena Thota, Priya Donti, Zico Kolter, Deepak Rajagopal, John Birge, Jimi Oke, and Ali Hortaçsu. *National Science Foundation Expeditions in Computing for Computational Decarbonization of Societal Infrastructures at Mesoscales*. `https://codecexp.us/`. (Accessed on October 8, 2024). 2024.