1 Vision

My research agenda focuses on developing machine learning approaches that efficiently utilize the abundance of energy data, accurately interpret physical energy system behaviors, provide cost-efficient and resilient solutions for emerging energy system challenges, and support the analysis of the societal impacts from the energy transition and climate change. I aim to identify key energy system applications that can be enabled or improved by new machine learning techniques, formulate energy system problems that highlight critical technical challenges, design efficient algorithms grounded in rigorous mathematical principles, and validate these methods using real-world testbeds or datasets.

The energy sector is undergoing an unprecedented surge in renewable integration, electrification of everything from vehicles to buildings, and a transition towards decentralized power generation to combat climate change and achieve a net zero goal by 2050. In response to the rising need for intelligent solutions, harnessing the power of data is essential. The widespread deployment of smart meters in energy systems exemplifies this potential. However, current energy analytics tools face bottlenecks in high-fidelity data processing, particularly because energy data is often partially observable. For example, Behind-the-Meter (BTM) solar systems result in grid operators observing only aggregated measurements of load and solar generation. Synchrophasor data in power transmission systems suffers from missing/corrupted values, leaving system dynamics unobservable. These challenges raise the following research questions in energy systems:

Q1: *How can we effectively extract information for decision-making, when facing large-scale, partially observable energy data?*

Meeting the critical demands of information extraction for energy systems often requires more than simply scaling up existing machine learning methods. It is essential to leverage the inherent structure of energy data to extract information effectively. Moreover, common challenges with large-scale energy data lie not only in the complexity of the processes that generate it but also in the downstream decision-making processes it aims to inform and optimize. For example, electricity price formation relies on power flow models that incorporate congestion and account for the uncertainty introduced by renewable generation variability. Strategic energy storage must consider future price uncertainty while modeling physical constraints. Understanding and modeling energy system decision processes under uncertainty is critical for accurately learning from observed behaviors. This highlights another research question in designing energy systems:

Q2: *How can we build an autonomous energy system that learns from complex, structured data, while improving decision quality under uncertainty?*

My overarching research seeks to address the above questions by developing intelligent, cost-efficient, and resilient solutions that bridge machine learning with the evolving needs of energy systems (as illustrated in Fig. 1). As a result, my research spans from developing innovative machine learning methodologies to conducting in-depth energy system analyses. In the following, I describe two major areas of my research in detail. My work on high-dimensional energy data analytics addresses the question of how to efficiently analyze and extract information from complex energy data (Q1). My research on autonomous energy storage systems

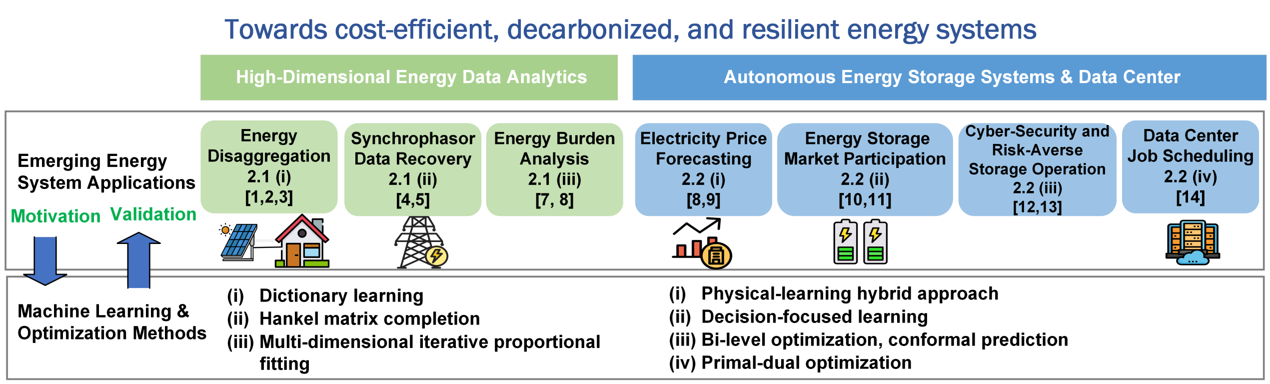

Figure 1: Overview of my research on learning-enabled sustainable energy systems. Machine learning problems are at the bottom and energy system applications are at the top (corresponding to the subsection below).

and data center job scheduling explores how to enhance decision-making in complex environments (Q2).

2 Past and Ongoing Research

2.1 High-Dimensional Energy Data Analytics

My research in energy data analytics centers on inferring missing information from partial observations. Energy disaggregation estimates load and solar generation from aggregated measurements. Synchrophasor data recovery infers unobservable system dynamics from partially observed information. My work leverages the low-rank structure of energy data and introduces frameworks based on low-rank models, dictionary learning and matrix completion, to address challenges in these applications. In energy burden analysis, I tackle the challenge of inferring joint distributions of residential characteristics given marginal distributions.

(i) **Energy Disaggregation with Behind-the-Meter Solar Generation.** As most meters in power systems only measure the net point of aggregate loads, solar generation is usually Behind-the-Meter (BTM) and is *invisible* to the power system operator. Data-driven methods for energy disaggregation face challenges in collecting individual load data at the substation level due to the invisibility of BTM solar. I formulated the energy disaggregation problem as a deterministic dictionary learning problem with *partial labels* [1]. After analyzing the intrinsic column sparsity of our problem, I developed an algorithm by adding a ℓ_{21} norm constraint and proposed an incoherence term, improving accuracy by 20-30% over traditional methods. In [2], my collaborators and I developed disaggregation and forecasting combination frameworks for BTM solar generation to enhance the accuracy of net load forecasting. Additionally, because renewable generations are volatile, it is inevitable that the disaggregation of power consumption contains errors. I extended Bayesian dictionary learning algorithms to provide *uncertainty modeling* for load disaggregation [3]. The proposed method learns the probabilistic distributions of the dictionaries and the coefficients by Gibbs sampling. The key insight is to employ Monte-Carlo integration to *approximately* compute the predictive mean and variance of the distribution of estimated loads. My work is the first to model uncertainty in both Bayesian dictionary learning and energy disaggregation.

(ii) **Synchrophasor Data Recovery for Resilient Power Grids.** Phasor Measurement Units (PMUs) provide high-resolution, real-time synchronized measurements and thus offer better visibility into power grid dynamics. However, communication congestions or PMU malfunctions often lead to synchrophasor data quality issues like missing data and cyber-attacks. The existing data recovery method fails when the missing data is *simultaneous and consecutive across all channels*, and no existing work addresses *quantifying the uncertainty* of the data recovery results. I incorporated the Hankel structure into Bayesian matrix completion to leverage the *spatial-temporal relationships* in time series data. Mean-field variational inference was employed to approximate posterior distributions and provided uncertainty measures to assess the confidence levels of each estimation [4]. Moreover, existing data recovery methods assume power systems dynamics can be approximated by a linear dynamical system, but their performance degrades significantly when handling *nonlinear dynamics*. I formulated the nonlinear data recovery as a Bayesian high-rank matrix completion problem [5]. The central idea is to exploit the *kernel trick* and lift the Hankel matrix of the nonlinear data into a higher dimensional space such that *the lifted Hankel matrix is low-rank*. The proposed approach recovers simultaneous and consecutive missing data/corruptions, even in highly nonlinear power systems, improving recovery accuracy by 50% compared to low-rank Hankel methods. My two algorithms have been integrated into the Streaming Synchrophasor Data Quality tool, a PMU data processing product developed in collaboration with the Electric Power Research Institute (EPRI).

(iii) **Residential Electrification and Energy Burden Analysis with Climate Change.** Residential electrification is a key driver of decarbonization, replacing carbon-heavy appliances with electric alternatives to create more efficient homes. However, this transition raises concerns about *energy affordability*, particularly for low-income households. In [6], I presented a novel approach to evaluate energy burden distribution during this electrification shift. We developed a data-driven Reduced Complexity Building Energy Model (RC-BEM) [7] to simulate energy demand across diverse climate conditions. With the help of this model, I introduced multi-dimensional proportional fitting method to estimate energy burden distributions based on housing characteristics. Our analysis reveals that annual energy burdens are generally manageable but surge during extreme summer and winter months. By disentangling the effects of electrification from climate change, we provide policymakers with new insights to balance climate goals with energy equity for a sustainable future.

2.2 Learning-Enabled Autonomous Energy Storage Systems and Data Center.

The central theme of my postdoctoral research is enhancing decision-making for energy systems under uncertainty. Energy storage systems must *strategically* plan their operations based on future price forecasts to quantify their opportunity value, considering their limited energy capacity. These systems integrate various modules—electricity price prediction, market participation, cybersecurity, and risk-averse operation—into a cohesive framework for optimal decision-making. My research spans these critical components to advance the capabilities of autonomous energy storage operations. Data center job scheduling follows a similar rationale, where limited resources require efficient job allocation while accounting for uncertainties in future demand.

(i) **Learning-Physical Hybrid Approach for ERCOT Price Forecast**. Most learning-based electricity price prediction models ignore the *price formation principles* behind locational marginal prices (LMP). I designed a data-driven forecasting model grounded in the electricity price formation of the ERCOT market, decomposing LMP into system lambda, congestion price, and price adders, and building separate models for each. I led a team of four students in processing over 100 GB of data and built a learning-physical hybrid prediction model. Our system enables energy storage arbitrage, achieving 80% profit attainable with perfect forecasting. We also studied uncertainty quantification using conformal prediction [8], a model-agnostic, distribution-free method that enables *dynamic adjustment* and provides *coverage guarantees* for uncertainty intervals. In [9], my collaborators and I explored designing storage bids for the day-ahead market based on real-time price predictions and arbitrage strategies in two-settlement markets to maximize profits.

(ii) **Strategic Energy Storage: Bridging Prediction and Optimization with Decision-Focused Learning.** Energy storage systems typically employ Model Predictive Control (MPC) policy, which decouples price prediction from storage optimization. However, this separation introduces challenges, as machine learning models often fail to account for *how* predictions interact with energy storage optimization. I developed an end-to-end model that integrates price prediction and optimization [10], ensuring the training objective is directly aligned with final storage decisions. To ensure differentiability in linear storage models, the *perturbation idea* was introduced into the decision-focused loss function and provided a rigorous theoretical analysis of the perturbed loss. Our surrogate loss function *bypasses* expensive Jacobian computations, providing a more efficient model training. This method effectively addresses challenges in self-scheduling storage arbitrage, and storage behavior forecasting. In [11], I extended the decision-focused framework to tackle storage bidding, a more complex challenge where storage systems must place bids in advance rather than making real-time charge/discharge decisions. Along with degradation costs, the bidding model must account for opportunity costs that result from limited storage capacity. Given future price forecasting, the key idea for bidding design is to link the opportunity cost to the dual variable of the state of charge (SoC). Training an end-to-end bidding framework with a decision-focused loss is challenging due to the dual optimization layers for bid generation and market clearing. Leveraging the implicit function theorem, the decision loss is backpropagated through dual differentiable optimization layers to update the predictor's weights. This approach yields higher profits and generalizes well to *price-maker* scenarios, providing a versatile solution for real-world applications.

(iii) **Cybersecurity and Risk-Averse Operation for Strategic Energy Storage.** Strategic energy storage decisions are heavily influenced by *future price expectations*. Storage systems may charge even during high-price periods in anticipation of future spikes. However, high prices signal supply-demand imbalances, and manipulated price predictions can make the grid more vulnerable. Attackers can generate false price signals, causing multiple storage systems to charge simultaneously, leading to demand surges and threatening grid stability. This is the first work to reveal these vulnerabilities in coordinated energy storage responses to falsified predictions in a real distribution network [12]. Since price predictions can be manipulated, energy storage systems must operate with an *awareness of the risks* associated with *imperfect predictions*. We introduced a conformal controller that extends conformal prediction to calibrate energy storage decisions based on realized opportunity values, ensuring low-risk operational decisions [13].

(iv) **Data Center Job Scheduling in the Era of AI.** The rise of large language models (LLMs) like ChatGPT has dramatically increased the power demands of data centers, pushing the limits of the current power system. The stochastic nature of AI workloads has made electricity consumption more *volatile* than ever, posing a critical challenge to maintaining a stable power supply. I am developing a strategic approach to dynamically pace computational resource capacity, ensuring a smooth and efficient response to these fluctuating workloads. I bridge control theory with *primal-dual* methods, enabling data centers to *dynamically adjust* resource allocation through dual variable tuning for precise management of uncertain demands [14]. We are currently validating this algorithm on synthetic data and will test it on real LLM simulators, such as VLLM and SGLang, to further showcase its potential.

3 Future Plan: Energy System Cost Efficiency, Decarbonization, and Resilience

I will continue conducting interdisciplinary research spanning energy systems, machine learning, optimization, networking, and environmental science. I will seek collaborators from these fields to work together to tackle emerging energy system challenges. My future research aims to advance new data-driven algorithms to achieve more reliable integration of DERs and data centers, and to build a foundation for my long-term vision of enhancing the electric energy system's decarbonization, cost-efficiency, and resilience. In the following, I detail three crucial steps towards this goal.

Learning-Enabled Decision-Making for Sustainable Energy Systems.

Humans solve complex challenges through intelligence and adaptability. Achieving similar capabilities in energy systems requires: (1) developing effective *representation learning models*, such as renewable and data center workload forecasting, and (2) designing *intelligent and efficient optimization strategies*. I believe that decision-focused learning bridges the gap between predictive models and decision-making, positioning it as a key driver in future energy systems. In [10, 11], the uncertainty is present in the objective function. One extension in the decision-focused setting is to handle constraint-based uncertainty, which poses greater challenges because the predicted solutions may *not be feasible* under the true constraints. In sustainable data center job scheduling, data centers aim to fully use renewable energy. Jobs must be scheduled in advance, but renewable energy availability is only revealed closer to the execution time, creating a mismatch between predictions and reality, and forcing costly last-minute adjustments. I will extend the decision-focused approach by introducing a penalty function for mismatched decisions and adapting end-to-end learning for this setting.

Collaborative Learning in Multi-Agent Systems for Grid Autonomy.

As energy resources become increasingly decentralized, with distributed energy storage systems and renewable generations, the *coordination* between multiple agents becomes essential. These agents must operate collaboratively in real time to balance supply and demand, respond to grid fluctuations, and optimize resource usage. I will start by focusing on multi-agent energy storage dispatch, building upon my previous work [10] and incorporating graph neural networks (GNNs) to capture the spatial and temporal correlations between distributed storage units. This will enable more efficient real-time coordination of multi-unit energy storage. I will also explore the coordination of solar and wind generation with energy storage, addressing the challenges of managing intermittent renewable resources by optimizing storage dispatch to maintain supply-demand balance.

Adaptive Strategies for Resilient Energy Systems.

Resilient energy systems are not just about surviving disruptions—they must *anticipate, adapt, and recover* swiftly to ensure stability. As the energy landscape evolves with the integration of DERs and the rapid expansion of data centers, current energy systems are more vulnerable than ever. I will enhance energy system visibility and develop adaptive strategies for better planning and operation of future-proof energy systems. A key step in this effort is to address the gap in current load disaggregation methods that often overlook critical physical dynamics such as SoC transitions. By extending my previous work on load disaggregation [4] and physics-informed energy storage frameworks [10], I aim to improve the disaggregation of measurements in systems with behind-the-meter energy storage. Furthermore, volatility in demand from data centers can be exploited for cyberattacks by manipulating job scheduling, creating sudden surges or drops in demand that destabilize the grid. Building on my work [14], I plan to investigate these vulnerabilities and develop robust, general-purpose algorithms that ensure energy system resilience in the face of such cyber threats.

Potential Funding Opportunities. My past research has been funded by NSF, DOE, EPRI, AFOSR, ARO, and Concord Energy. Moving forward, I aim to secure both government and industry funding to support my work. My research plan aligns with several key funding programs, such as Energy, Power, Control, and Networks (EPCN) programs from NSF, and Advanced Research Projects Agency-Energy (ARPA-E) program from DOE. Additionally, I plan to explore industry-sponsored opportunities, including Amazon Research Award for Sustainability and EPRI University Global Research Award.

References

[1] W. Li*, **Ming Yi***, and M. Wang, "Real-time energy disaggregation at substations with behind-the-meter solar generation," *IEEE Transactions on Power Systems*, 2020, * equal contribution.

[2] A. Stratman, T. Hong, **Ming Yi**, and D. Zhao, "Net load forecasting with disaggregated behind-the-meter pv generation," *IEEE Transactions on Industry Applications*, 2022.

[3] **Ming Yi** and M. Wang, "Bayesian energy disaggregation at substations with uncertainty modeling," *IEEE Transactions on Power Systems*, 2021.

[4] **Ming Yi**, M. Wang, E. Farantatos, and T. Barik, "Bayesian robust hankel matrix completion with uncertainty modeling for synchrophasor data recovery," *ACM SIGENERGY Energy Informatics Review*, 2022.

[5] **Ming Yi**, M. Wang, T. Hong, and D. Zhao, "Bayesian high-rank hankel matrix completion for nonlinear synchrophasor data recovery," *IEEE Transactions on Power Systems*, 2023.

[6] **Ming Yi**, S. Nawawi, and P. Vaishnav, "How do electrification and climate change affect the distribution of energy burdens: an analysis of 10,000 buildings in 30 us cities," 2024, under preparation.

[7] S. Nawawi, **Ming Yi**, M. Craig, T. Detjeen, and P. Vaishnav, "Cross-sectoral trade-offs in a changing climate: Surrogate models to balance home energy bills, occupant comfort, and power system externalities," *submitted to Joule*, 2024.

[8] S. Alghumayjan, **Ming Yi**, and B. Xu, "Risk-averse uncertainty quantification in electricity price forecasting with conformal prediction," *submitted to IEEE PES General Meeting*, 2024.

[9] S. Alghumayjan, J. Han, N. Zheng, **Ming Yi**, and B. Xu, "Energy storage arbitrage in two-settlement markets: A transformer-based approach," *Electric Power Systems Research*, 2024.

[10] **Ming Yi**, S. Alghumayjan, and B. Xu, "Perturbed decision-focused learning for strategic energy storage," *submitted to IEEE Transactions on Smart Grid*, 2024.

[11] **Ming Yi**, Y. Wu, J. Anderson, and B. Xu, "Decision-focused bidding design for energy storage arbitrage," 2024, under preparation.

[12] **Ming Yi**, Y. Wu, J. Anderson, and G. Zussman, "Stealthy cyber-attack for strategic behind-the-meter energy storage," 2024, under preparation.

[13] Y. Wu, **Ming Yi**, B. Xu, and J. Anderson, "Risk-averse energy storage arbitrage," 2024, under preparation.

[14] **Ming Yi** and B. Xu, "A dual-based pid controller for job scheduling in data centers with uncertain demand," 2024, under preparation.