GATSched: Multi-Objective Graph Attention Networks for Energy-Efficient HPC Job Scheduling

Kyrian Adimora
adimora.kyrian@ku.edu
University of Kansas
Lawrence, Kansas, USA

Hongyang Sun (Advisor)
University of Kansas
Lawrence, USA
hongyang.sun@ku.edu

Abstract

High-Performance Computing (HPC) systems face an urgent sustainability crisis, with leading facilities consuming 10–60 MW and incurring multimillion dollar annual energy costs. Traditional schedulers such as SLURM and PBS treat energy as secondary, leading to energy waste of 30 to 50% above theoretical optimal levels. We present GATSched, a multi-objective Graph Attention Network scheduler that models HPC workloads as dynamic graphs with specialized attention heads. Our approach jointly optimizes energy efficiency, performance, and resource utilization using four attention mechanisms: energy, performance, balance, and temporal. Through trace-driven simulation validation on 389,604 production jobs in three HPC architectures, GATSched achieves a 27–35% energy reduction while maintaining substantial resource utilization.

CCS Concepts

• **Hardware** → **Power and energy**; • **Computing methodologies** → **Machine learning**; **Neural networks**; • **Mathematics of computing** → *Graph algorithms*; • **Networks** → *Network resources allocation.*

Keywords

Energy-aware scheduling, Graph attention networks, Multi-objective optimization, Reinforcement learning, High-performance computing, Power management

ACM Reference Format:
Kyrian Adimora and Hongyang Sun (Advisor). 2018. GATSched: Multi-Objective Graph Attention Networks for Energy-Efficient HPC Job Scheduling. In *Proceedings of The International Conference for High Performance Computing, Networking, Storage, and Analysis (SC '25).* ACM, ST., LOUIS, USA, 3 pages. https://doi.org/XXXXXXX.XXXXXXX

1 The Critical Energy Challenge

Modern HPC systems consume extraordinary amounts of energy: Frontier operates at 29 MW, Fugaku at 30 MW, and Summit at 13 MW [2, 14, 18]. These power requirements create three critical problems: (1) unsustainable operational costs exceeding about $30M annually per facility, (2) significant environmental impact with carbon footprints comparable to small cities, and (3) infrastructure limitations that restrict scientific progress as power becomes the limiting factor for future large-scale systems [15].

Traditional schedulers such as SLURM, PBS Pro, and LSF treat energy as a secondary concern, focusing primarily on computational throughput [9]. This approach creates a fundamental mismatch: Although these systems optimize for job completion rates, they ignore the complex relationships between energy consumption, job characteristics, and resource allocation that determine overall system efficiency [3]. The result is substantial energy waste, often 30-50% above theoretical optimal consumption, which directly impacts both environmental sustainability and the allocation of research funding.

2 GATSched: A Paradigm Shift

GATSched introduces the graph attention network scheduler that simultaneously optimizes three critical objectives: energy efficiency, computational performance, and resource utilization. Our key insight is that HPC scheduling can be modeled as a dynamic graph problem where jobs and their complex interdependencies are captured through specialized attention mechanisms [19].

2.1 Technical Innovation

Our approach focuses on four breakthrough innovations. First, we model scheduling as dynamic graphs $G(t) = (V(t), E(t))$ where job nodes contain comprehensive 8-dimensional feature vectors encoding CPU, memory, execution time, priority, deadlines, wait time, power profiles and GPU requirements. This representation captures the full complexity of production workloads while maintaining computational efficiency, addressing limitations identified in previous energy-aware DAG scheduling approaches [4, 11].

Second, we employ multi-head Graph Attention Networks with four specialized heads: Energy Head for power optimization, Performance Head for throughput maximization, Balance Head for resource distribution, and Temporal Head for deadline management [23]. Each head learns distinct patterns through:

$$\mathbf{h}_i^{(l+1)} = \Vert_{k=1}^{4} \sigma \left(\sum_{j \in \mathcal{N}_i} \alpha_{ij}^{(l,k)} \mathbf{W}^{(l,k)} \mathbf{h}_j^{(l)} \right) \tag{1}$$

where $\mathbf{h}_i^{(l)}$ is the feature vector of job i at layer l, $\alpha_{ij}^{(l,k)}$ are learned attention weights for head k, and $\mathbf{W}^{(l,k)}$ are trainable weight matrices.

Third, our multi-objective policy dynamically balances competing objectives through adaptive weighting, extending beyond traditional single-objective approaches [13, 10]:

SC '25, ST., LOUIS

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

$$r_t = \alpha_t \cdot r_t^{\text{energy}} + \beta_t \cdot r_t^{\text{perf}} + \gamma_t \cdot r_t^{\text{balance}} \quad (2)$$

where weights $[\alpha_t, \beta_t, \gamma_t]$ sum to 1 and adjust in real time based on the state of the system and the characteristics of the workload, allowing optimal trade-offs as the conditions evolve.

Fourth, our constraint-aware assignment ensures that power caps, SLA requirements, and resource limits are never violated while maintaining sub-50ms decision latency for production deployment, addressing critical power management concerns in modern HPC systems [7, 1]. The GAT architecture maintains computational efficiency through sparse attention mechanisms, enabling sub-50ms decision latency for production deployment.

3 Comprehensive Validation

We conducted a rigorous evaluation using production traces from three diverse HPC architectures, totaling 389,604 real jobs spanning 5+ years. The dataset includes: Polaris (241,772 jobs, NVIDIA A100, 2023-2024), Mira (52,154 jobs, IBM Blue Gene/Q, 2018-2019) and Cooley (95,678 jobs, Intel Haswell, 2018-2019). This represents authentic production workloads with realistic dependencies and resource patterns, providing more comprehensive validation than typical synthetic workload studies [21, 16].

Our GAT architecture uses 3 layers (64-32-16 dimensions) with 4 attention heads per layer, trained on 70/15/15 splits using Adam optimization. We compare against four production schedulers: SLURM with backfilling, PBS Pro with power awareness, LSF with resource optimization, and Volcano for cloud workloads.

4 Breakthrough Results

GATSched delivers consistent, statistically significant improvements in all architectures tested ($p < 0.001$). In Polaris, we achieve 34.87% energy reduction with 95.85% resource utilization, a combination that does not exist in current scheduler approaches. Mira shows 27.52% energy savings with 87.62% utilization, while Cooley demonstrates 27.33% reduction with 66.81% utilization.

The comparative analysis reveals substantial advantages: 34-57% better energy performance than SLURM, 39-53% improvement over PBS Pro, 44-57% better than LSF and 52-54% superior to Volcano. Critically, we achieve these energy savings while reducing job wait times by up to 99% (from 3.93h to 0.04h) and maintaining computational efficiency of 15.5-17.7 FLOPS/J. These results significantly exceed the 12-23% energy savings reported in previous HPC energy-aware scheduling studies [22, 6, 5].

Ablation studies confirm our architectural choices: GAT provides 12.3% better energy savings than standard GCN, multi-head attention yields 8.7% improvement over single-head, and dynamic weighting delivers 15.2% better adaptability than static approaches.

5 Transformative Impact

These results translate into transformative real-world impact. A single major HPC facility using GATSched can achieve annual savings of $10M+ through reduced energy consumption and gains in infrastructure efficiency. Environmental benefits include CO_2 reductions exceeding 15,000 tons annually per facility, equivalent to removing 3,000 cars from the roads. For the broader HPC community, this enables sustainable scaling to future exascale systems without proportional increases in power infrastructure [17].

Our open source framework provides immediate value to the community, with portable framework integration for SLURM and PBS systems. The 389K+ job validation demonstrates the scalability and reliability essential for deployment in critical research infrastructure.

6 Future Directions and Conclusion

GATSched establishes a new paradigm for sustainable HPC scheduling by proving that energy efficiency and computational performance are not competing objectives but can be simultaneously optimized through intelligent graph-based learning [20]. Our comprehensive validation across diverse architectures and massive production workloads demonstrates both the technical feasibility and transformative potential of this approach.

Future work includes adaptation of online learning for evolving workloads, federated learning across multisite HPC federations, quantum-classical hybrid scheduling integration, and real-time workload prediction using transformer architectures for proactive energy management [12, 8]. This research provides the foundation for sustainable high-performance computing essential for next-generation scientific discovery while maintaining the computational excellence that drives innovation across all scientific domains.

The combination of 27-35% energy savings, superior resource utilization, and novel GAT-based deployment capabilities makes GATSched not just a research contribution but a practical solution to one of HPC's most pressing challenges: achieving sustainability without sacrificing scientific progress.

Acknowledgments

This research is supported in part by the US National Science Foundation grant #2441633. We thank the Argonne Leadership Computing Facility (ALCF) for providing access to the Polaris, Mira, and Cooley system workload traces that made this comprehensive evaluation possible.

References

[1] Anders Aaen Springborg, Michele Albano, and Samuel Xavier-de-Souza. 2023. Automatic energy-efficient job scheduling in hpc: a novel slurm plugin approach. In *Proceedings of the SC '23 Workshops of The International Conference on High Performance Computing, Network, Storage, and Analysis* (SC-W '23). Association for Computing Machinery, Denver, CO, USA, 1831–1838. ISBN: 9798400707858. doi:10.1145/3624062.3624265.

[2] Scott Atchley et al. 2023. Frontier: exploring exascale. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (SC '23) Article 52. Association for Computing Machinery, Denver, CO, USA, 16 pages. ISBN: 9798400701092. doi:10.1145/3581784.3607089.

[3] Mario Bambagini, Mauro Marinoni, Hakan Aydin, and Giorgio Buttazzo. 2016. Energy-aware scheduling for real-time systems: a survey. *ACM Trans. Embed. Comput. Syst.*, 15, 1, Article 7, (Jan. 2016), 34 pages. doi:10.1145/2808231.

[4] Sanjeev Baskiyar and Rabab Abdel-Kader. 2010. Energy aware dag scheduling on heterogeneous systems. *Cluster Computing*, 13, 4, (Dec. 2010), 373–383. doi:10.1007/s10586-009-0119-6.

[5] Danilo Carastan-Santos, Georges Da Costa, Millian Poquet, Patricia Stolf, and Denis Trystram. 2024. Light-weight prediction for improving energy consumption in hpc platforms. In *Euro-Par 2024: Parallel Processing*. Jesus Carretero, Sameer Shende, Javier Garcia-Blas, Ivona Brandic, Katzalin Olcoz, and Martin Schreiber, (Eds.) Springer Nature Switzerland, Cham, 152–165. ISBN: 978-3-031-69577-3. doi:10.1007/978-3-031-69577-3_11.

[6] Briag Dupont, Nesryne Mejri, and Georges Da Costa. 2020. Energy-aware scheduling of malleable hpc applications using a particle swarm optimised greedy algorithm. *Sustainable Computing: Informatics and Systems*, 28, 100447. doi:https://doi.org/10.1016/j.suscom.2020.100447.

[7] Vadim Elisseev, John Baker, Neil Morgan, Luigi Brochard, and Terry Hewitt. 2016. Energy aware scheduling study on bluewonder. In *Proceedings of the 4th International Workshop on Energy Efficient Supercomputing* (E2SC '16). IEEE Press, Salt Lake City, Utah, 61–68. ISBN: 9781509038565.

[8] E. A. Kiselev, A. V. Yarovoy, P. N. Telegin, and A. V. Baranov. 2024. Scheduling supercomputer jobs under existing power consumption constraints. *Lobachevskii Journal of Mathematics*, 45, 10, 5082–5091. doi:10.1134/S199508022460612X.

[9] Bartłomiej Kocot, Paweł Czarnul, and Jerzy Proficz. 2023. Energy-aware scheduling for high-performance computing systems: a survey. *Energies*, 16, 2. doi:10.3390/en16020890.

[10] Jingbo Li, Xingjun Zhang, Zheng Wei, Jia Wei, and Zeyu Ji. 2021. Energy-aware task scheduling optimization with deep reinforcement learning for large-scale heterogeneous systems. *CCF Transactions on High Performance Computing*, 3, 4, 383–392. doi:10.1007/s42514-021-00083-8.

[11] Jing Mei, Kenli Li, and Keqin Li. 2014. Energy-aware task scheduling in heterogeneous computing environments. *Cluster Computing*, 17, 2, (June 2014), 537–550.

[12] Aditya Priya et al. 2024. Energy-minimizing workload splitting and frequency selection for guaranteed performance over heterogeneous cores. In *Proceedings of the 15th ACM International Conference on Future and Sustainable Energy Systems* (e-Energy '24). Association for Computing Machinery, Singapore, Singapore, 308–322. ISBN: 9798400704802. doi:10.1145/3632775.3661968.

[13] Mohammad Sajid and Zahid Raza. 2017. Energy-aware stochastic scheduler for batch of precedence-constrained jobs on heterogeneous computing system. *Energy*, 125, 258–274. doi:https://doi.org/10.1016/j.energy.2017.02.069.

[14] Mitsuhisa Sato, Yuetsu Kodama, Miwako Tsuji, and Tesuya Odajima. 2022. Co-design and system for the supercomputer "fugaku". *IEEE Micro*, 42, 2, (Mar. 2022), 26–34. doi:10.1109/MM.2021.3136882.

[15] Woong Shin, Vladyslav Oles, Ahmad Maroof Karimi, J. Austin Ellis, and Feiyi Wang. 2021. Revealing power, energy and thermal dynamics of a 200pf pre-exascale supercomputer. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (SC '21) Article 12. Association for Computing Machinery, St. Louis, Missouri, 14 pages. ISBN: 9781450384421. doi:10.1145/3458817.3476188.

[16] Georgios L. Stavrinides and Helen D. Karatza. 2017. Simulation-based performance evaluation of an energy-aware heuristic for the scheduling of hpc applications in large-scale distributed systems. In *Proceedings of the 8th ACM/SPEC on International Conference on Performance Engineering Companion* (ICPE '17 Companion). Association for Computing Machinery, L'Aquila, Italy, 49–54. ISBN: 9781450348997. doi:10.1145/3053600.3053611.

[17] C. Terboven et al. 2025. Ee-hpc a framework for energy efficient hpc system management. In *Proceedings of the SC '24 Workshops of the International Conference on High Performance Computing, Network, Storage, and Analysis* (SC-W '24). IEEE Press, Atlanta, GA, USA, 1878–1882. ISBN: 9798350355543. doi:10.1109/SCW63240.2024.00236.

[18] Sudharshan S Vazhkudai et al. 2018. Power analysis of the summit supercomputer. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, 1–12.

[19] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph attention networks. In *International Conference on Learning Representations (ICLR)*. doi:10.17863/CAM.48429.

[20] Zhibao Wang, Shuaijun Chen, Lu Bai, Juntao Gao, Jinhua Tao, Raymond R. Bond, and Maurice D. Mulvenna. 2023. Reinforcement learning-based task scheduling for environmentally sustainable federated cloud computing. *Journal of Cloud Computing*, 12, 1, 174. doi:10.1186/s13677-023-00553-0.

[21] Xu Yang, Zhou Zhou, Sean Wallace, Zhiling Lan, Wei Tang, Susan Coghlan, and Michael E. Papka. 2013. Integrating dynamic pricing of electricity into energy aware scheduling for hpc systems. In *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis* (SC '13) Article 60. Association for Computing Machinery, Denver, Colorado, 11 pages. ISBN: 9781450323789. doi:10.1145/2503210.2503264.

[22] Xu Yang, Zhou Zhou, Sean Wallace, Zhiling Lan, Wei Tang, Susan Coghlan, and Michael E. Papka. 2013. Integrating dynamic pricing of electricity into energy aware scheduling for hpc systems. In *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis* (SC '13) Article 60. Association for Computing Machinery, Denver, Colorado, 11 pages. ISBN: 9781450323789. doi:10.1145/2503210.2503264.

[23] Mingjun Zhao, Lei Mo, Ji Liu, Jie Han, and Dan Niu. 2024. Gat-based deep reinforcement learning algorithm for real-time task scheduling on multicore platform. In *2024 36th Chinese Control and Decision Conference (CCDC)*, 5674–5679. doi:10.1109/CCDC62350.2024.10587381.