LA&S Senior Seminar

Argumentative Research Paper

3 October, 2025

Beyond-CMOS Computing Efficiency as Common Ground in Data Center Expansion

The exponential growth of artificial intelligence has sparked an unprecedented race among nations and corporations to expand data center infrastructure, creating a complex tension between technological advancement and environmental sustainability. On one side, industry leaders argue compellingly that rapid deployment is essential for maintaining competitive advantage in the global AI marketplace, creating jobs, and attracting investment capital that fuels innovation. Their position is understandable: the stakes of falling behind in AI development carry genuine strategic and economic consequences, and delays in the acquisition of cutting-edge hardware make delays potentially catastrophic. On the other side, environmental advocates raise valid concerns about the carbon footprint of this explosive growth, pointing to the troubling reality that despite decades of promises, renewable energy expansion has not kept pace with the industry's appetite for power. Both perspectives stem from legitimate priorities, economic vitality and national competitiveness on one hand, environmental responsibility and long-term sustainability on the other.

What unites these seemingly opposed viewpoints is a shared interest in the viability and success of the data center industry itself. Neither camp benefits from an unsustainable trajectory; industry cannot thrive without reliable, affordable energy, and environmental goals cannot be achieved by halting technological progress that billions now depend upon. This common ground suggests that the solution lies not in choosing between rapid deployment and environmental responsibility, but in fundamentally reimagining the efficiency of computing infrastructure itself. By examining the current state of data center energy consumption alongside emerging technologies in materials engineering and microelectronics—specifically Beyond-CMOS alternatives like magnetoresistive random access memory—we can identify a path forward that honors both the urgency of technological advancement and the imperative of environmental stewardship.

While positions that datacenters provide leverage for green planning are valid viewpoints within the disciplines of business and environmental planning, they are not novel. Data center energy concerns have existed in the public eye since the early 2000s, and it has been since that each of these arguments have been shared. It is then valuable to examine how these problems have been addressed within the last decade. I argue that such emissions and air quality, coupled with stagnation in infrastructure roll out within United States, fuel stark outlooks on the sustainability of datacenters from a grid perspective.

Since 2010, "efficiency measures avoided energy-related carbon dioxide (CO_2) emissions equivalent to nearly 20% of the global total in 2023", stated to be much more than the entire European Union and India combined (IEA, 2025). Improvements in energy efficiency reduce overall fuel consumption, which lowers emissions of pollutants such as sulfur dioxide and fine particulates (IEA, 2025). Simultaneously, the heightened use of natural gas and its by-products, natural gas plant liquids (NPGLs), has further improved air quality by displacing coal in power generation and industrial applications. NGPLs and natural gas emit far less sulfur, mercury, and particulate matter than coal, leading to measurable gains in regional air quality, though careful management of methane leakage remains important.

Despite these improvements, which come as a result of greater efficiency in the energy sector, non-renewable energy sources still dominate in the United States. Coal is the sole non-renewable energy source to have reduced since 2010, falling from >5 PWh to almost 3 in 2024. Natural gas, crude oil, and NPGL have all risen to at least double, representing a combined 22 PWh annually. Renewables, on the other hand, have not been expanded much, with nuclear, hydro-electric, geo-thermal, solar, wind, and biomass sources representing <5 PWh of production in 2024 (EIA 2025). This serves to contrast the argument that data centers offer a strongarm to barter for greener solutions alone. This is a rather troubling conclusion, as breakthroughs in AI have encouraged trillions in industry investment, largely being put towards data center construction. It is predicted that by 2030 data center power consumption will double globally, up to 3% of worldwide energy consumption, with problems becoming much more drastic if not addressed immediately. Given insufficient growth in green energy capacity in the last decade, and a move towards non-renewable sources incentivized via tax breaks by recent U.S. tax law, it appears that there is a much more promising to encourage efficient data center design (One Big Beautiful Bill Act, Pub. L. No. 119-21).

Given the rate of energy growth, I argue it is therefore infeasible to attack the issue from the side of energy production; instead, it will instead be necessary to explore beyond-CMOS technologies as a key step in achieving the necessary efficiency to secure a path to a sustainable future.

Within data centers, not all components contribute equally to efficiency (IEA, 2025). Rack servers are known to be the main contributor, and are the primary load, physically and energetically, in a typical data center. Efficiency of servers is defined as transactions per second per watt. Consequently, an increase in performance by 50% means a server is 50% more efficient, and vice versa. This creates a dynamic where it is actually the largest ‘hyperscale’ data centers which are most efficient. Moreover, in all cases heating effects contribute up to 30% inefficiency. This is a simultaneous factor of physical layout of the data center and server components. The latter is of significance, as topics in current research, such as development of devices incorporating novel two-dimensional materials, offer unique contributions to offset heating and efficiency problems.

The coupled performance-thermal wall we see in component efficiency coincides with the slowing of Moore's Law scaling, creating what researchers term the 'Beyond-CMOS imperative'—the urgent need for alternative device technologies that can deliver performance improvements without proportional increases in power consumption (Beyond CMOS). This has driven current research in the direction of two-dimensional magnetic semiconductors, which are positioned as a viable alternative to charge-based devices in ultrafast memory and logic applications due to inherently reduced heating effects.

There are multiple application spaces which magnetoresistive random access memory (MRAM) has the potential to shine in. Indeed, spin-transfer torque (STT)-MRAM has embedded itself as a replacement for flash memory (Nyugen et al.). In terms of performance, STT-MRAM is three times smaller than conventional flash memory and is at best able to switch in the few-nanosecond regime (Yang et al.) (Hellenbrand et al.). Recent improvements to switching efficiency and signal-to-noise-ratio have enabled fabs to replace conventional embedded flash memory and static random-access memory with STT-MRAM in high-end electronics. At the cache level, spin-orbit torque (SOT)-MRAM is positioned as a viable technology due to improvements in durability and efficiency compared to STT-MRAM (Nyugen et al.). Both MRAM architectures are highly valuable for integration in data center computing environments, particularly enterprise-class data centers which waste, on average, 3 times as much energy cooling than hyperscale data centers. Improvements to heating effects enable hardware to run at higher load, facilitating multiplicative gains in the two most inefficient areas of data center architecture.

While memory serves as a critical component to modern day computing, there is an inherent disadvantage in specialized use-cases, such as the training of AI models, in that the physical separation of memory and compute units reduces performance and power-efficiency. Moreover, features such as "non-volatility, stochasticity, and oscillations in MRAM, provide new feasibility for novel computing to solve combinatorial optimization problems, which are notoriously difficult for conventional computers" (Nguyen et al.). Given the emphasis on AI-focused datacenters in the last decade, it is highly desirable, then, to approach the issue of power usage and cooling using these technologies (Shehabi et al.). This unique architectural suitability towards AI workloads positions MRAM technology as a cornerstone to reducing power inefficiency and furthering technological progress.

The challenge of reconciling data center expansion with environmental sustainability reveals not a fundamental conflict of values, but rather a shared recognition that our current trajectory is untenable for all stakeholders involved. Industry leaders are correct that competitive pressures and strategic imperatives demand swift action, just as environmental advocates are correct that existing infrastructure approaches threaten both climate goals and air quality standards. The historical record demonstrates that relying solely on green energy procurement agreements and market mechanisms has not produced the renewable energy buildout necessary to offset the industry's exponential growth in power consumption. The emergence of Beyond-CMOS technologies, particularly magnetoresistive random access memory architectures, offers a tangible solution that addresses the core concerns of both positions. By dramatically improving the power efficiency and thermal performance of computing hardware, these technologies enable the rapid infrastructure deployment that industry requires while simultaneously reducing the environmental burden that concerns energy planners and environmental advocates. The multiplicative efficiency gains possible through MRAM implementation—especially in the energy-intensive domains of memory operations and cooling infrastructure, represent a pragmatic bridge between competing priorities.

Moving forward requires recognizing that innovation in computing efficiency is not merely a technical consideration but a strategic necessity for achieving both economic and environmental objectives. Rather than framing the conversation as industry versus environment, we must unite around the principle that sustainable technological advancement depends on transformative improvements in how we compute. By prioritizing the development and deployment of Beyond-CMOS technologies alongside continued efforts in renewable energy expansion, we can honor the legitimate urgencies on both sides of this debate and chart a course toward a data center industry that is simultaneously competitive, economically robust, and environmentally responsible. The path to a green future in computing lies not in constraining innovation, but in fundamentally reinventing the efficiency of the technologies that power it.

"24/7 by 2030: Realizing a Carbon-Free Future" *Google Sustainability*, https://sustainability.google/reports/247-carbon-free-energy/. Accessed 3 Oct. 2025.

"Beyond CMOS: The Future of Semiconductors" *IEEE IRDS™*. https://irds.ieee.org/home/what-is-beyond-cmos. Accessed 7 Sep. 2025.

Bi, Xiuyuan, et al. "STT-RAM Cell Design Considering CMOS and MTJ Temperature Dependence." *IEEE Transactions on Magnetics*, vol. 48, no. 11, Nov. 2012, pp. 3821–24. *IEEE Xplore*, https://doi.org/10.1109/TMAG.2012.2200469.

Cai, Kaiming, et al. "Spin-Based Magnetic Random-Access Memory for High-Performance Computing." *National Science Review*, vol. 11, no. 3, Oct. 2023, p. nwad272. *PubMed Central*, https://doi.org/10.1093/nsr/nwad272.

Dieny, B., et al. "Opportunities and Challenges for Spintronics in the Microelectronics Industry." *Nature Electronics*, vol. 3, no. 8, Aug. 2020, pp. 446–59. *www.nature.com*, https://doi.org/10.1038/s41928-020-0461-5.

Global Optimism and Amazon. "The Climate Pledge." *The Climate Pledge*, https://www.theclimatepledge.com. Accessed 3 Oct. 2025.

Hellenbrand, Markus, et al. "Progress of Emerging Non-Volatile Memory Technologies in Industry." *MRS Communications*, vol. 14, no. 6, Dec. 2024, pp. 1099–112. *Springer Link*, https://doi.org/10.1557/s43579-024-00660-2.

Hu, G., et al. "Key Parameters Affecting STT-MRAM Switching Efficiency and Improved Device Performance of 400°C-Compatible p-MTJs." *2017 IEEE International Electron Devices Meeting (IEDM)*, 2017, p. 38.3.1-38.3.4. *IEEE Xplore*, https://doi.org/10.1109/IEDM.2017.8268515.

Joshi, Vinod Kumar, et al. "From MTJ Device to Hybrid CMOS/MTJ Circuits: A Review." *IEEE Access*, vol. 8, 2020, pp. 194105–46. *IEEE Xplore*, https://doi.org/10.1109/ACCESS.2020.3033023.

Khalili Amiri, Pedram, et al. "Electric-Field-Controlled Magnetoelectric RAM: Progress, Challenges, and Scaling." *IEEE Transactions on Magnetics*, vol. 51, no. 11, Nov. 2015, pp. 1–7. *IEEE Xplore*, https://doi.org/10.1109/TMAG.2015.2443124.

Kim, Woojin, et al. "VCMA-MTJ: Towards Ghz Operation Low Power MRAM." *2023 IEEE International Magnetic Conference - Short Papers (INTERMAG Short Papers)*, 2023, pp. 1–2. *IEEE Xplore*, https://doi.org/10.1109/INTERMAGShortPapers58606.2023.10228494.

Lam, Chester. *Alder Lake's Caching and Power Efficiency*. 24 Aug. 2025, https://chipsandcheese.com/p/alder-lakes-caching-and-power-efficiency.

Li, Yueting, et al. "Work-in-Progress: Toward Energy-Efficient Near STT-MRAM Processing Architecture for Neural Networks." *2022 International Conference on Hardware/Software Codesign and System Synthesis (CODES+ISSS)*, 2022, pp. 13–14. *IEEE Xplore*, https://doi.org/10.1109/CODES-ISSS55005.2022.00013.

Manchon, A., et al. "Current-Induced Spin-Orbit Torques in Ferromagnetic and Antiferromagnetic Systems." *Reviews of Modern Physics*, vol. 91, no. 3, Sep. 2019, p. 035004. *DOI.org (Crossref)*, https://doi.org/10.1103/RevModPhys.91.035004.

United States, Energy Information Administration. *Monthly Energy Review. Aug. 2025*, www.eia.gov/totalenergy/data/monthly/pdf/mer.pdf. Accessed 3 Oct. 2025.

International Energy Agency. "Emission Reductions." *Multiple Benefits of Energy Efficiency*, IEA, 2019, iea.blob.core.windows.net/assets/7d547730-44ff-404b-af52-0812a340aee5/MultipleBenefitsofEnergyEfficiency-EmissionReductions.pdf. Accessed 3 Oct. 2025.

Nguyen, V. D., et al. "Recent Progress in Spin-Orbit Torque Magnetic Random-Access Memory." *Npj Spintronics*, vol. 2, no. 1, Oct. 2024, p. 48. *www.nature.com*, https://doi.org/10.1038/s44306-024-00044-1.

Paullin, Charlie. "Legislative Report: Development of Data Centers Could Triple Energy Demands If Unconstrained • Virginia Mercury." *Virginia Mercury*, 9 Dec. 2024, https://virginiamercury.com/2024/12/09/legislative-report-development-of-data-centers-could-triple-energy-demands-if-unconstrained/.

United States, Congress, House. *One Big Beautiful Bill Act.* Congress.gov, www.congress.gov/bill/119th-congress/house-bill/1. 119th Congress, H.R. 1, enacted 4 July 2025.

Shehabi, Arman, et al. *United States Data Center Energy Usage Report*. LBNL--1005775, 1372902, 1 Jun. 2016, p. LBNL--1005775, 1372902. *DOI.org (Crossref)*, https://doi.org/10.2172/1372902.

Song, Cheng, et al. "Recent Progress in Voltage Control of Magnetism: Materials, Mechanisms, and Performance." *Progress in Materials Science*, vol. 87, Jun. 2017, pp. 33–82. *ScienceDirect*, https://doi.org/10.1016/j.pmatsci.2017.02.002.

Total Energy Monthly Data - U.S. Energy Information Administration (EIA). https://www.eia.gov/totalenergy/data/monthly/. Accessed 9 Sep. 2025.

"Tracking Clean Energy Progress 2023 – Analysis." *IEA*, 12 Jul. 2023, https://www.iea.org/reports/tracking-clean-energy-progress-2023.

Van Geet, Otto, and David Sickinger. *Best Practices Guide for Energy-Efficient Data Center Design*. NREL/TP--5R00-89843, DOE/GO--102024-6283, 2417618, 25 Jul. 2024, p. NREL/TP--5R00-89843, DOE/GO--102024-6283, 2417618. *DOI.org (Crossref)*, https://doi.org/10.2172/2417618.

Yang, Hyunsoo, et al. "Two-Dimensional Materials Prospects for Non-Volatile Spintronic Memories." *Nature*, vol. 606, no. 7915, Jun. 2022, pp. 663–73. *www.nature.com*, https://doi.org/10.1038/s41586-022-04768-0.

Yusuf, Alaba, et al. "Domain-Specific STT-MRAM-Based In-Memory Computing: A Survey." *IEEE Access*, vol. 12, 2024, pp. 28036–56. *IEEE Xplore*, https://doi.org/10.1109/ACCESS.2024.3365632.