PUBLIC SUBMISSION

Received: May 24, 2025
Tracking No. mb2-gibn-3kzh
Comments Due: May 28, 2025 **Submission Type:** API

Docket: NSF-2025-OGC-0001
NITRD_FRDOC_0001

Comment On: NSF-2025-OGC-0001-0001
Request for Information: Development of a 2025 National Artificial Intelligence Research and Development Strategic Plan

Document: NSF-2025-OGC-0001-DRAFT-0106
Comment on FR Doc # 2025-07332

Submitter Information

Name: Joerg Appenzeller

General Comment

See attached file(s)

Attachments

Appenzeller RFI AI R_D input

Joerg Appenzeller (Purdue University) Response to AI R&D RFI

This document is approved for public dissemination. The document contains no business-proprietary or confidential information. Document contents may be reused by the government in developing the 2025 National AI R&D Strategic Plan and associated documents without attribution.

Overview: Today's nearly ubiquitous –and rising– demands for smarter generative Artificial Intelligence (AI), faster search engines, and more accurate predictive modeling vastly outweigh any foreseeable incremental improvement to conventional von Neumann computer architectures. The core of this problem is the *mismatch* between evolving probabilistic computational needs and the entirely deterministic nature of current computers that operate sequentially with a focus on precision. Contrary to the capabilities of current computer systems, many real-world problems, such as drug discovery, data encryption and decryption, supply chain logistics, and large data handling, require computers to embrace "uncertainty" in complex data to provide a weighted range of possible answers. The critical absence of such probabilistic computer ***hardware*** capabilities remains a barrier to accelerating next-generation computational power for AI and Machine Learning (ML) applications. This grand challenge defines the urgent need to build next-generation AI hardware beyond deep learning with new architectures that do not solely depend upon the use of silicon complementary metal oxide semiconductor (CMOS) technology.

To design and fabricate probabilistic computer hardware, naturally stochastic probabilistic materials and devices are required for true random number generation (TRNG) with suitable methods to demonstrate tunability. Fluctuation speeds, device dimensions, and energy efficiency are key areas of fundamental research required to concatenate individual stochastic devices efficiently. Industry is unlikely to fund research on these fundamental questions due to key gaps remaining in the design and fabrication of probabilistic hardware, but several companies have expressed strong interest in hardware accelerators if government funding can first expand upon the basic science of this field.

While the global focus is on using existing CMOS technology and design layouts to improve on the performance of AI, the US can maintain its competitive leadership in AI with disruptive and energy efficient probabilistic hardware. These new computing architectures can be developed in the US if fundamental research on the suitable novel materials, devices and circuits for probabilistic computing (p-computing) is conducted.

Advancements in National Security: For our nation's security, maintaining global AI research and development leadership is imperative; however, current development of CMOS architectures is insufficient to meet future AI computational demands. P-computing capabilities will provide US companies with the computational hardware to output solutions imperative to innovative product development for use in commercial, industrial, and public security applications. Our military will also benefit from these advanced technologies used to protect confidential data from cyberattacks through enhanced encryption beyond what is possible with current CMOS-based computing. A key aspect of p-computing decryption and encryption of data is the source of stochastic electrical behavior at room temperature, which may involve non-semiconducting materials. Advancements also include modified stochastic magnetic tunnel junctions (sMTJs) that are naturally radiation hard, making them ideal for defense applications.

Advancement in Public Infrastructure Resilience: The semiconductor industry has developed specialized computer chips, i.e. graphics processor units (GPUs) and tensor processing units (TPUs) to enable building more powerful computers, but at the price of enormous energy consumption. Morgan Stanley estimates that "by 2027, generative AI could use as much energy as Spain needed to power itself in 2022." To meet the computational demands for ever increasing AI usage from industry and the public, p-computing can help alleviate these energy concerns due to its reduced power consumption e.g. from use of stochastic magnetic tunnel junctions. The energy savings from use of p-computing for AI systems would allow for continued growth of AI development and application without an energy crisis.

Advancements in Scientific Discovery: Fundamental advances in key scientific fields are critical to make probabilistic computing a reality:

Variability: Understanding how various parameters that characterize stochastic devices ultimately enter into the performance of probabilistic circuits (p-circuits) is critical to delivering the desired system-level and technology-level testbeds. While stochastic devices, by definition, are different from each other, it remains unclear what "similarities" the devices must show, i.e. what variability between devices is acceptable.

Magnetic dynamics: Building on existing accomplishments using stochastic magnetic tunnel junctions (sMTJs), gaining a better understanding of the dynamics of coupled probabilistic bits (p-bits) are imperative. These insights will enable the engineering of fast, functional technology, improve circuit architectures, and optimize individual p-bit hyper-parameters.

Geometry impact: The p-bit geometry is known to be a key player for the actual energy barrier between the two states that characterizes a magnetic tunnel junction (MTJ). A high resistive state is characterized by an antiparallel orientation between the magnetization of the two magnets constituting a conventional MTJ, while a low resistive state results from their parallel arrangement. Fluctuating randomly between these two states at room temperature constitutes the p-bit core. While the industry has a good idea of the geometrical impact on stable MTJs, much less is known about the design rules for p-bits. Advancing the fundamental knowledge in this context will be central to the research effort.

Materials impact: The MTJ material greatly influences the quality of the *tunable* true random number generator (*tunable* TRNG), including the choice of materials for both magnetic layers, the insulating spin filter between them, as well as their respective thicknesses and geometrical layout. Moreover, choosing magnets with in-plane anisotropy (IMA) versus perpendicular anisotropy magnets (PMA) influences the switching dynamics and, therefore, the switching speed of p-bits. The limited fundamental knowledge in this critical area restricts substantial advances in the development of a probabilistic computer.