The Architectural Imperative: Why Brain-Inspired Hardware is the Key to Sustainable AI

itihaasa Research and Digital
Centre for Nano Science and Engineering, IISc

April 2026

itihaasa
Research and Digital

ceNSE
CENTRE FOR NANO SCIENCE AND ENGINEERING

Contents

Foreword

The AI revolution is reshaping our world, but it is colliding with a hard physical reality: an unsustainable energy demand that threatens our global power grids. To navigate this AI paradox, a fundamental architectural disruption is required, shifting our focus from raw speed to extreme energy efficiency. For India, leading this transition toward brain-inspired, memristor-based hardware is not just a technological aspiration; it is a strategic necessity.

Neuromorphic architectures offer a transformative pathway to democratize AI by enabling high-performance inference and training directly at the edge. This allows intelligence to be embedded locally in everything from wearable health monitors to agricultural sensors predicting extreme climate events in vulnerable rural regions. Crucially, by processing data on-device, this hardware paradigm ensures privacy and fortifies national data sovereignty by removing the need to excessively rely on centralized cloud infrastructure. Furthermore, energy-efficient edge computing is vital for India's strategic and defense sectors, empowering power-constrained platforms like satellites and surveillance drones with crucial onboard intelligence.

However, translating these groundbreaking molecular memristor technologies from academic laboratories into commercial realities requires robust, targeted funding. To build an indigenous hardware capability, India must channel funding to R&D projects and deeptech translational startups. Philanthropic and institutional support is already demonstrating what is possible. Notably, the Pratiksha Trust has funded a project at CeNSE, IISc, to develop a molecular neuromorphic accelerator for real-time neurofeedback, forms an integral part of an ambitious Brain Co-Processor moonshot program.

By aggressively nurturing and funding this interdisciplinary ecosystem in the domains of material science, electronics, nanoscience, molecular memristors, neuroscience, brain inspired hardware, etc. India can position itself at the global frontier of sustainable AI.

Kris Gopalakrishnan

Trustee, Pratiksha Trust

Founder and Chairperson, itihaasa Research and Digital

Executive Summary

The Problem: The AI Energy Crisis and the Memory Wall

The rapid advancement of AI is creating an unsustainable energy demand, with some projections indicating computing could consume 100% of global energy generation by 2040 if unchecked. This crisis is driven by an inherent design flaw in modern computers called the "von Neumann bottleneck". Since the processor and memory are physically separate, moving data back and forth consumes up to a thousand times more power than the actual computation. To sustainably advance AI, the industry must transition from prioritizing raw speed to maximizing energy efficiency.

The Solution: Memristors and In-Memory Computing

The Memristor Element: Memristors offer an architectural escape route by acting as a non-linear electrical component that functions simultaneously as a resistor and a memory element. This mimics the synaptic connections of a biological brain.

In-Memory Computing and Vector Matrix Multiplication: By storing and processing data in the same location, memristors eliminate the energy-draining communication overhead between the processor and memory. When arranged in dense crossbar arrays, memristors become massively parallel analog calculating engines. They perform vector-matrix multiplication (the core math of deep learning) in a single time step, decoupling computation time from AI model complexity.

Material Innovation: Traditional oxide-based memristors rely on unpredictable stochastic switching, hindering commercial scalability. Conversely, molecular memristors—such as the pioneering platform developed at the Centre for Nano Science and Engineering (CeNSE) in the Indian Institute of Science (IISc)—utilize precise thermodynamic transitions, successfully achieving 16,500 distinct conductance states.

Neuromorphic Capabilities and Performance

Brain-Inspired Emulation: Memristors naturally support the implementation of Artificial Neural Networks by mimicking synaptic plasticity and relying on high-precision analog strengths.

Privacy by Architecture: Memristor technology enables AI training and inferencing directly on the device, eliminating the need to centralize data in the cloud and inherently safeguarding data privacy.

Efficiency Benchmarks: Memristor-based neuromorphic platforms provide extreme energy efficiency, outperforming powerful conventional CPUs and GPUs by up to 460 times and 220 times respectively on key AI tasks.

Edge Intelligence: These energy gains make memristors ideal for power-constrained systems, allowing surveillance drones to analyze images locally and potentially extending their flight times tenfold. They also enable sophisticated Wearable AI and distributed IoT sensor networks.

Global Landscape and Commercialization

International Research: The race to commercialize memristors involves hubs across China, the USA, and Europe, alongside prominent developments in India via IISc. Leading US startups are actively commercializing memristor-powered neuromorphic chips.

Hybrid CMOS Integration: Memristors are designed to act as highly efficient accelerators rather than complete replacements for general-purpose CMOS. Commercialization requires integrating memristor layers onto existing silicon chips through a Back-End-of-Line (BEOL) process.

The Software Imperative: A major hurdle to adoption is the software ecosystem. Success requires Electronic Design Automation (EDA) simulators, data management protocols, and specialized application mapping stacks to route AI workloads efficiently on hybrid architectures.

India's Action Plan

IISc's Commercial Readiness: Sreetosh Goswami and Navakanta Bhat from the Centre for Nanoscience and Engineering detail that their memristor crossbar technology currently sits at TRL 7/8. They plan to incubate a startup by 2026 to develop an extremely energy-efficient System-on-Chip (SoC) for Edge AI applications like the YOLO (You Only Look Once) model.

Pragmatic Manufacturing Strategy: To secure a first-mover advantage, experts recommend fabricating the underlying silicon chips at advanced global foundries, while completing the specialized molecular memristor integration in-house. Domestic large-scale manufacturing can follow once 300mm wafer facilities are established in India.

Strategic Ecosystem Requirements and Funding: Advancing this capability requires establishing a dedicated Class 100 cleanroom pilot fabrication line to eliminate process variability. Additionally, the ecosystem requires deeptech translational grants—modeled on the US SBIR program—to support innovative hardware startups and overcome early-stage funding barriers.

1. The AI Paradox and the Global Energy Crisis

The AI revolution is real, and it is spectacular. AI now composes music, designs drugs, and drives complex strategic decisions. Yet, this achievement comes with a critical, often hidden cost: an unsustainable hunger for energy. This is the AI Paradox: our smartest systems are rapidly becoming our most resource-intensive.

1.1. The Looming Energy Cliff

If global computing trends continue unchecked, energy consumption projections are alarming, threatening to absorb 100% of global energy generation by 2040.[1] This isn't a distant environmental concern; it is an imminent global crisis, forcing tech giants to consider extreme measures, such as commissioning dedicated nuclear power plants simply to power their data centers.[2] Our progress in AI is crashing into a hard physical wall.

This crisis is compounded by the deceleration of traditional silicon scaling—the famed "Moore's Law" is slowing. The solution cannot be simply cramming more digital switches onto a chip. Instead, we must fundamentally shift our strategic goal from maximizing raw speed to maximizing **energy efficiency** (operations per Joule).[3] This moment demands a fundamental material and architectural disruption, not just incremental upgrades.

1.2. The Design Flaw: The Memory Wall

The bottleneck isn't the processing itself, but the moving of data. Every computer today is shackled by an ancient design flaw called the "von Neumann bottleneck," where the processor (the decision-maker) is physically separate from the memory (the data storage).[4]

Imagine a chef who must constantly run across town—microsecond by microsecond—just to fetch ingredients from the pantry. That continuous, energy-intensive trip between the processor and the memory consumes up to a thousand times more power than the actual calculation itself.[5] This communication overhead is the biggest drain on modern computing and the single greatest inhibitor to scalable AI.

[1] Jones, N. (2018). "How to stop data centres from gobbling up the world's electricity", *Nature*.
[2] Waldrop, M. M. (2016). "The chips are down for Moore's law", *Nature*.
[3] Khan, S. M. and Alexander, W. B. (2021). "The AI-Energy Paradox", *IEEE Spectrum*.
[4] Backus, J. (1977). "Programming Be Liberated from the von Neumann", *Communications of the ACM*.
[5] Wulf, W. A., & McKee, S. A. (1995). "Hitting the memory wall: implications of the obvious", *ACM SIGARCH Computer Architecture News*.

2. Memristors: The Synapse of the Future

The technological escape route from the memory wall and the power crisis is the memristor. This device is a radical departure from traditional electronics, enabling computation that mirrors the efficiency of the biological brain.

2.1. Defining the Core Element: Non-Linear Resistance

A memristor is a two-terminal, non-linear electrical component whose electrical resistance, or conductance, is not constant; it possesses a history-dependent quality, meaning its resistance depends on the electric charge that has previously flowed through it. Crucially, this unique characteristic allows the device to function as both a resistor and a memory element simultaneously, enabling stateful logic.[6] This analog, modifiable conductance state serves as the direct electrical equivalent of the biological **synaptic weight**—the strength of connection between neurons—making it the ideal building block for biologically inspired computation.[7]

2.2. The In-Memory Computing Revolution

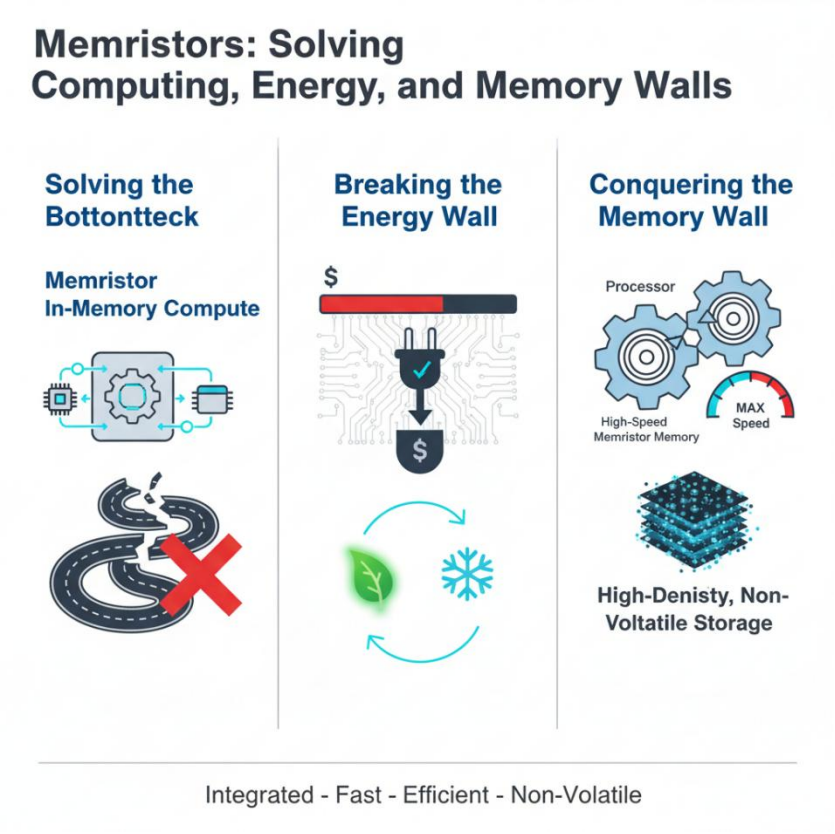
By merging memory and processing, the memristor enables **In-Memory Computing (IMC)**.[8]

1. **Eliminating the Commute:** Since data is stored and processed in the same location, the costly "commute" between the processor and memory is eliminated, dissolving the von Neumann bottleneck and saving massive amounts of energy and latency.
2. **The Crossbar Accelerator:** To deploy this power, memristors are arranged into dense, nanoscale **crossbar arrays**.[1] This array is not just storage; it's a massive, parallel analog calculating engine. It is perfectly designed to perform the core mathematical operation of all deep learning: **vector-matrix multiplication (VMM)**, often called the "dot product".
3. **Constant-Time Computation:** The breakthrough is simple physics: the time required for VMM computation is a **single time step**, regardless of the size of the matrix. This ability to physically decouple computation time from the complexity of the AI model is a game-changer for scalability.

The result is a leap in efficiency that defies traditional scale. For key AI tasks, memristor platforms have been proven to crush conventional hardware, delivering up to **460 times the energy efficiency compared to powerful CPUs and 220 times that of high-end GPUs.**[9] This is an architectural escape route from the energy crisis.

[6] Chua, L. (1971). "Memristor-The Missing Circuit Element", *IEEE Transactions on Circuit Theory*.
[7] Strukov, D. B., et al. (2008). "The missing memristor found", *Nature*.
[8] Ielmini, D., and Wong, H. S. P. (2018). "In-memory computing with resistive switching devices", *Nature Electronics*.
[9] Zidan, M. A., et al. (2018). "The future of electronics based on memristive systems", *Nature Electronics*.

Figure 1



2.3. Material Platforms: From Bulk Oxides to Molecular Precision

The practical viability of memristors hinges on the specific material platform chosen, which determines reliability, precision, and scalability. Researchers globally are investigating three primary material classes:

- **Oxide-Based Memristors:** This category, including transition metal oxides like hafnium oxide or tantalum oxide, is the most widely explored due to its accessibility and relative simplicity of fabrication.[10] These devices typically operate via the formation of conductive filaments—defects like oxygen vacancies—in response to voltage, a process that inherently resembles a "soft breakdown" mechanism. **This reliance on an inherently probabilistic, stochastic switching mechanism is a critical challenge.** It leads to significant variability and poor reproducibility, which are major hurdles for industrial-scale neuromorphic deployment.[1] For neuromorphic computing to achieve reliable, brain-like precision, this challenge must be solved by developing materials with **more controllable and predictable switching mechanisms**, replacing stochastic defect movement with consistent thermodynamic transitions (e.g., electronic or phase changes).
- **Two-Dimensional (2D) and Novel Materials:** Research is advancing into alternative materials such as silicon dioxide, molybdenum oxide, molybdenum disulfide, and other transition metal

[10] Goswami, S., et al. (2017). “Robust resistive memory devices using metal–organic ligands”, *Nature Materials*.

dichalcogenide monolayers. These materials promise unique properties and better control, often exploring electronic or phase transitions rather than the unreliable defect filament mechanism.

- **Molecular Memristors:** This represents the pinnacle of material control, utilizing transition metal complexes designed through bespoke chemical synthesis.[1] This approach aims to replace the unreliable soft breakdown mechanism with a consistent, controlled, and uniform thermodynamic process.[1] These advanced systems have demonstrated far superior performance and control compared to oxide-based memristors. For example, a pioneering molecular platform developed at the CeNSE in IISc, has achieved an astonishing **16,500 distinct, high-precision conductance states** within a single molecular film.[11] This level of analog control is transformative, moving the memristor from a variable lab curiosity to a reliable, high-precision component essential for scalable neuromorphic systems.[12]

Crucially, achieving the next generation of reliable materials requires dismantling research "silos."[1] **This endeavor is fundamentally interdisciplinary**, demanding deep collaboration among chemists (for material synthesis), material scientists (for fabrication), physicists (for understanding switching mechanisms), engineers (for integration), and computer scientists (for architecture).[1] This interdisciplinary collaboration, where living in silos would not help, represents a significant strength and opportunity for research ecosystems like India's.

3. Architecting Efficiency: The Physics and Implementation of In-Memory Computing

The true paradigm shift enabled by memristors is In-Memory Computing (IMC). IMC fundamentally alters the computational structure by allowing processing to occur where the data resides, achieving orders-of-magnitude improvements in energy efficiency.

3.1. Eliminating Communication Overhead: Co-locating Storage and Logic

Memristors' ability to both store and process data at the same physical location directly addresses the memory wall—the high latency and energy consumption incurred by continuous data movement. By co-locating storage and processing, memristive architecture eliminates the communication overhead that currently constitutes the largest contributor to overall computing energy expenditure. This transition substantially reduces both the time delays and energy consumption for computations, a crucial step for resource-heavy workloads like AI.

The implementation of IMC can take several forms. For digital applications, the transition involves using

[11] Goswami, S., et al. (2020). "Decision-making in a molecular memristor", *Nature*. (Referencing the IISc breakthrough regarding high-precision conductance states).
[12] Goswami, S., Goswami, S., Venkatesan, T. (2020). "An organic approach to low energy memory and brain inspired electronics." *Applied Physics Reviews*, 7(2).

stateful in-memory computing, where fundamental Boolean operations still form the basis of computation, but the logic inputs and outputs are stored directly as non-volatile conductance states in the memristors, streamlining the entire process. While the ultimate goal is true in-memory computing, an intermediate stage of near-memory computing, where memristors are placed proximate to the processor to minimize communication distance, also provides significant gains.

3.2. Analog Compute Acceleration: Memristors as High-Speed Dot Product Engines

The most profound utility of IMC in the context of the AI boom is the acceleration of vector-matrix multiplication (VMM). VMM, or the dot product, is the fundamental mathematical operation underlying all deep learning algorithms, defining the computational cost of both training and inference.

The memristor crossbar array functions as a dedicated analog VMM accelerator, often referred to as a "dot product engine".[13] The operating principle harnesses the inherent physics of the crossbar:

1. The elements of the input matrix (which represent the synaptic weights of the neural network) are programmed into the corresponding cell conductances (G) within the memristor array.
2. An input vector is applied as a set of voltages (V) across the rows of the crossbar.
3. Based on Ohm's Law (I=GV) and Kirchhoff's Current Law (summing currents at the column output), the output currents generated on the columns naturally and instantaneously represent the dot product sum of the vector with the matrix.

This architectural breakthrough yields an unmatched level of parallel efficiency. The time required to perform the entire VMM computation is a single time step, regardless of the size of the matrix. This property represents a critical mathematical scalability breakthrough. Traditional digital processing time scales linearly or polynomially with the size of the matrix, meaning latency increases alongside model complexity. In contrast, memristor IMC offers a constant-time operation for the most resource-intensive step of deep learning, physically decoupling computation time from the complexity of the AI model. This provides orders-of-magnitude improvements in speed and energy expenditure compared to conventional CMOS ASICs.[14]

This capability confirms that memristors will not replace general-purpose CMOS entirely, but are **optimally suited to function as highly efficient, application-specific accelerators (ASICs) integrated into existing silicon** platforms. This hybrid approach, combining the best of scaled silicon with new, specialized devices, is necessary to extend useful computing utility beyond the limitations of current silicon technology.

[13] Hu, M., et al. (2016). "Dot-product engine for neuromorphic computing: programming 1T1R crossbar arrays", *Advanced Materials*.
[14] Sun, X., et al. (2019). "Exploiting memristor-based in-memory computing for AI acceleration", *IEEE Design & Test*.

4. Neuromorphic Computing: Mirroring the Brain's Synaptic Dynamics

Memristors are more than just fast accelerators; their inherent properties allow them to serve as foundational elements for neuromorphic computing, hardware designed to mirror the massively parallel and low-energy functionality of the biological brain.

4.1. Beyond Binary: Leveraging Multi-Level Conductance States

The ability to store information as a high-precision, continuous analog value—the conductance state—is key to efficient neuromorphic emulation.[15] The biological brain relies on the analog strength of synaptic connections. Molecular memristors are ideally suited to represent this complexity, having demonstrated the capability to store up to 16,500 distinct conductance states.

This capability is vital because digital computers, while excelling at precision, are not inherently optimized for the "perception tasks" that are critical to modern AI (e.g., image recognition, natural language processing). Memristor-based analog platforms, which inherently support the implementation of Artificial Neural Networks (ANNs), offer a natural and significant advantage in processing these AI workloads.[16]

4.2. Emulating Synaptic Plasticity and the Neuroscience Gap

In the brain-inspired computing paradigm, processing and memory are governed by the functions of neurons and synapses.[17] The tunable, non-volatile conductance of the memristor directly mimics **synaptic plasticity**, the biological mechanism by which the brain learns through the strengthening or weakening of connections based on neural activity.

This bio-mimicry offers a path to revolutionize AI training. The high analog precision achieved by advanced molecular materials allows the acceleration of **both training and inferencing directly on the chip**. This capability is transformative because it eliminates the current requirement for training data to be centralized in remote data centers. By keeping data processing and training local to the device, this approach ensures that data never leaves the platform, safeguarding privacy and significantly enhancing security—a crucial feature given the rising concerns over data breaches associated with centralized AI training.[18] This feature elevates memristors from merely a performance enhancement tool to a fundamental governance and security enabler, providing **privacy by architecture**.

---

[15] Mead, C. (1990). “Neuromorphic electronic systems”, *Proceedings of the IEEE*.
[16] Indiveri, G., et al. (2011). “Neuromorphic silicon neuron circuits.”, *Frontiers in Neuroscience*.
[17] Bi, G. Q., & Poo, M. M. (1998). “Synaptic modifications in cultured hippocampal neurons: dependence on spike timing”, *Journal of Neuroscience*.
[18] Sebastian, A., et al. (2020). “Memory devices based on unconventional computing”, *Nature Nanotechnology*.

The challenge, however, extends beyond physics. True neuromorphic innovation is currently hampered by a **significant gap in fundamental knowledge from neuroscience**. While hardware can mimic simple synaptic functions, the complex, non-linear dynamics required for advanced applications like Spiking Neural Networks (SNNs) and **chaotic computing** often lack detailed biological blueprints to guide architectural design. Unlocking the full potential of these brain-inspired systems, which necessitates a complete reimagining of computing architectures, requires deeper, ongoing collaboration with biologists and neuroscientists to inform the next generation of hardware.

5. Performance Benchmarks and Strategic Application Domains

The transition to memristor-based computing is driven by quantifiable, transformative performance metrics, particularly in the domain of energy efficiency. These gains open up high-value strategic application domains previously constrained by power budgets and data security requirements.

5.1. Comparative Energy Efficiency Analysis

Empirical studies confirm that for critical AI and signal processing tasks, memristor-based neuromorphic platforms deliver orders-of-magnitude improvements in energy efficiency over state-of-the-art conventional hardware.[19] Specific benchmarks indicate that for certain workloads, memristor platforms can outperform todays CPUs and GPUs by a factor of over 100x. These figures demonstrate that memristors are the highly promising technological components for mitigating the escalating global computing energy crisis. By providing the power efficiency required for AI growth, they offer a viable path to relieve the severe stress that current power-hungry data centers impose on national power grids.

[19] Shafiee, A., et al. (2016). “A Convolutional Neural Network Accelerator with In-Situ Analog Arithmetic in Crossbars”, *ISCA*. (Benchmark data on the 100x-400x gains).

Table 1: Comparative Efficiency of Memristor Architectures

Feature/Metric	Conventional CMOS (Von Neumann)	Memristor-based Accelerator (IMC)	Strategic Significance
Data Flow Bottleneck	Severe (Memory Wall/Microsecond latency)	Eliminated (Co-location)	Shifts the computing cost model entirely.
Key AI Operation	Sequential Digital/CPU or Pipelined Parallel/GPU	Analog/Massive Parallelism (VMM)	Optimized for neural network perception tasks.
Matrix Multiplication Time Cost	Scaled by Matrix Size	Single Time Step (Size Independent)	Breaks the scalability barrier of complex models.
Energy Efficiency (AI Workloads)	High Power Consumption	Orders-of-Magnitude Improvement (220x - 460x)	Essential mitigation for the 2040 energy crisis.
Data Privacy	Centralized Training Required	On-Chip Training Possible	Creates architectural data security.

5.2. Edge Intelligence and Power-Constrained Systems

The massive energy efficiency gains translate directly into superior performance for distributed and power-constrained computing systems, particularly at the edge of the Internet of Things (IoT).

- **IoT and Sensor Networks:** Memristors enable local intelligence, allowing data processing to occur directly at the sensor or edge device. This drastically reduces the data traffic between edge devices and the cloud, which accounts for a substantial portion of overall computing energy consumption.[20] Through precise chemical design, memristors can even serve as dual-purpose sensors capable of both storage and computation, minimizing energy use in IoT systems.

[20] Chen, Y., et al. (2019). "Deep Learning on the Edge: A Neuromorphic Computing Perspective", *Proceedings of the IEEE*.

- **Strategic Power Budgets:** For platforms operating under tight power constraints, such as surveillance drones and satellites, memristor technology allows for the creation of low-power, on-board inferencing systems. For instance, a drone equipped with a memristor-based system could analyze images locally and only relay those determined to be perceived threats, potentially extending its flight time by a factor of 10 times. Similarly, satellites, which also operate in power-constrained environments, benefit significantly from this technology.
- **Wearable AI:** Memristor technology enables wearable devices, such as smart watches, to host complex AI inferencing engines locally. This capability allows for real-time applications like language translators or health diagnostics, processed securely without reliance on cloud resources.

6. Global Research and Commercialization Landscape

Researchers across the world are intensely working on memristor technology given its promise to converge computing and memory within a single device, which can lead to high-speed, low-energy computing critical for the AI era.

6.1. International Research Highlights

Research hubs across the globe are driving advancements in memristor materials and application-specific accelerators:

- **China:** Tsinghua University is a leader in memristor research, having showcased a fabricated full-system-integrated chip that supports efficient on-chip learning, claiming it consumes about 3% of the energy of an ASIC system for the same task. Potential uses include AI, autonomous driving, and wearable devices.[21]
- **USA:** Researchers from the University of Rochester have developed high-performance, bipolar phase-change memristors, indicating potential for robust, fast, low-power, and scalable memory and neuromorphic computing systems. Leading startups like Tetramem and Crossbar are at the forefront of commercializing neuromorphic chips powered by memristors.[22]
- **Europe (EU):** European researchers are focusing on both materials and applications:
    - **France:** Cea-Leti, a research institute, developed memristors to execute computationally intensive Bayesian neural network tasks, successfully differentiating nine classes of heart arrhythmia from ECG recordings, nearly matching software performance.[23]
    - **Germany:** Researchers at Forschungszentrum Jülich developed a new type of memristor that prevents low-power edge AI chips from "forgetting" when switching between different AI models (a key challenge for agentic AI).
    - **UK/EU Projects:** The University of Southampton, through the EU-funded MENESIS project, is

[21] Wan, W., et al. (2022). “A compute-in-memory chip based on resistive random-access memory”, *Nature*.
[22] Prezioso, M., et al. (2015). “Training a machine learning classifier with memristive networks”, *Nature*.
[23] Dalgaty, T., et al. (2021). “In situ learning using intrinsic memristor variability via Markov chain Monte Carlo sampling.”, *Nature Electronics*.

developing memristor-enabled neuromorphic systems for autonomous, efficient, and radiation-hardened computing directly onboard satellites. Other EU-funded initiatives focus on developing memristors from novel materials like tellurene for synaptic behavior emulation and advancing foundational neuromorphic circuits for AI applications. Cranfield University (UK) also focuses on large-area flexible memristors for sustainable wearable electronics.

- **India:** The Indian Institute of Science (IISc), has developed a memristor-based brain-inspired analog computing platform capable of storing and processing data in an astonishing **16,500 conductance states** within a molecular film. This breakthrough in high-precision molecular memristors is critical for bringing complex AI tasks, like training Large Language Models, to personal devices. [24]

7. Commercialization Roadmap: Challenges and Strategic Imperatives

While the theoretical promise and empirical performance of memristors are immense, commercial readiness hinges on overcoming fundamental material and infrastructural challenges.

7.1. The Critical Materials Challenge: From Stochasticity to Control

A significant obstacle preventing the widespread commercialization of memristors is the material science bottleneck. Most existing devices rely on defective oxides driven by an inherently probabilistic, stochastic filamentary mechanism. This mechanism is difficult to control and is not scalable, leading to variability that prevents industrial adoption.

For memristors to become a stable, scalable technology, this soft breakdown mechanism must be replaced by a consistent, controlled, and uniform thermodynamic process, such as an electronic or phase transition. Molecular memristors, such as the high-precision platform developed at CeNSE in IISc, represent the most optimal material solution currently realized. Advancing this technology requires specialized synthesis, fabrication, and integration expertise that spans multiple scientific disciplines.

7.2. System Integration: CMOS Compatibility and Infrastructure Imperatives

Memristor technology is not a full replacement for silicon CMOS but rather a necessary accelerator. Integration must therefore follow a hybrid system model, with the memristor memory being integrated with silicon chips fabricated by existing foundries.

This integration is performed as a **Back-End-of-Line (BEOL) process**, meaning the memristor layers are fabricated on top of the already-processed silicon chips, similar to precedents set by integrating unconventional technologies like OLEDs. While the molecular materials are not traditionally handled by

[24] S, H., Bhat, N. & Goswami, S. “Neuromorphic pathways for transforming AI hardware”, *Nature Electronics* 8, 752–756 (2025). https://doi.org/10.1038/s41928-025-01432-z

industrial fabs, the necessary processes developed for memristor integration are CMOS-compatible.

7.3. The Software Imperative for Hybrid Architectures

The hardware innovation of memristors must be paired with comprehensive software development. Commercialization requires robust software stacks to manage data routing and application mapping between the silicon processor and the memristor memory.[25]

Specifically, three layers of software are necessary:

1. **Electronic Design Automation (EDA) Simulators:** These simulators are needed to integrate the characteristics of high-precision analog memristors with other EDA platforms.
2. **Data Management Protocols:** A specialized software stack is required for the appropriate and efficient management of data flow between the host silicon processor chip and the integrated memristor memory.
3. **Application Mapping Stacks:** These are critical for mapping existing AI inference algorithms, currently fine-tuned for conventional GPU chips, onto the novel hybrid silicon-memristor architecture to ensure efficient utilization of the new hardware.

8. Action plan for nurturing a capability in memristor technologies

8.1 Fabrication capability

At present, memristors cross bars are being fabricated in an academic fab in IISc. The next immediate stage of translating the integration of a silicon chip and a memristor to a higher TRL will require a pilot line where there will not be any variabilities. This is a cleanroom, where the memristors using molecular technology can be implemented as a back-end-of-line process on top of foundry-fabricated silicon chips. This IISc team is working with Ministry of Electronics and IT (MeitY) to set-up a R&D facility for 200mm wafer processing. Once the technology development is completed, the team can then move to large scale manufacturing through fabrication capabilities available in commercial foundries.

Given the exponential growth in AI, India needs to be extremely agile and pragmatic on fabrication capabilities. Currently, we do not have 300mm commercial wafer processing capabilities in India for both FEOL and BEOL. While we protect and secure all the IPs related to Molecular Neuromorphic AI technology, design, and products, we should fast track the large-scale manufacturing and production through leading foundries abroad to capture the market and establish industry leadership. This is even more crucial for Semiconductor chips for Edge AI inference and training, where the demand is humongous and the required volumes are astronomical. As and when the 300mm wafer fabrication

[25] Gokmen, T., & Vlasov, Y. (2016). "Acceleration of Deep Neural Network Training with Resistive Cross-Point Devices", *Frontiers in Neuroscience*.

facilities are established in India, we can then focus on domestic manufacturing, which is also important from strategic perspectives.

8.2 Associated software stacks

Developing a comprehensive EDA and software stack is also essential. Even lower-precision oxide memristor platforms typically provide device simulators that integrate with established EDA environments, enabling circuit- and system-level co-design. A comparable ecosystem must be built here in India. At the hardware level, accurate device and array models are required for integration with standard EDA flows. At the system level, a dedicated software layer must manage data movement and synchronization between the silicon processor and the memristor crossbar arrays.
Crucially, a compiler framework is needed to translate high-level AI models into hardware-aware instructions. Models currently optimized for GPUs must be re-partitioned, quantized, and mapped onto a heterogeneous architecture in which control logic resides in silicon while compute-intensive matrix operations execute within analog crossbars. Such a compiler would schedule operations, allocate precision, manage memory placement, and minimize analog–digital transitions to fully exploit the hardware's efficiency.
The capability to develop these integrated EDA tools, runtime systems, and compiler stacks already exists within India's semiconductor and software ecosystem.

8.3 Funding

Promising technologies that also develop fundamental strategic capabilities need to be nurtured with adequate funding. India's AI mission[26] is creating a scalable AI computing ecosystem with over 10,000 GPUs through public-private partnerships. Along with this, the AI mission also needs to explore allocating grant funds to promising translational startups that are designing innovative hardware devices that create an indigenous capability. Translational startups that are working on innovative SOC plus memristor based memory technologies should be encouraged. India's Semiconductor Mission[27] has an ambitious target of setting up fabs and design centers in India and provides attractive incentives for companies doing so. We believe that the Semiconductor Mission should also encourage the companies setting up fabs and design centers to invest in fab equipment, software, and training in our nationally funded higher educational institutions. Pratiksha Trust has funded a project for developing a high-precision molecular neuromorphic accelerator for real-time neurofeedback and signal conditioning at CeNSE in IISc as part of a larger program to develop a Brain Co-Processor moonshot program. The key objective is to design and implement a high-precision analog computation platform using molecular memristor crossbars.

Grant funding for deeptech translational startups like those developing memristor technologies is the

[26] https://indiaai.gov.in/
[27] https://www.ism.gov.in/

need of the hour. This can be modelled on the Small Business Innovation Research (SBIR)[28] program in the USA. Qualcomm, a world leader in SOCs for mobile phones has received multiple grants from SBIR in the late 1980s and early 1990s. We are hopeful that the initiatives like the RDI (Research Development and Innovation) Fund[29] will provide a leadership for funding in the deeptech domain.

9. Expert Q&A: Perspectives on Indigenous Capability

The work on memristors at the Indian Institute of Science (IISc) is spearheaded by Sreetosh Goswami and Navakanta Bhat at the Centre for Nano Science and Engineering (CeNSE)[30] which is a state-of-the-art nanofabrication facility funded by Office of the Principal Scientific Advisor (PSA) to the Govt. of India, Ministry of Electronics and Information Technology (MeitY), and the Department of Science and Technology (DST). This section is a Q&A with Sreetosh and Navakanta where they explain different aspects of memristors and the potential impact of memristors on India.

9.1 About CeNSE, IISc

Navakanta: CeNSE is a comprehensive 14,000 sq. ft. cleanroom facility benchmarked against best university facilities in the world. It supports advanced technology development in a variety of areas including Nanoelectronics, Nano photonics, 2D materials, GaN technology, MEMS and NEMS, nanobiotechnology, neuromorphic devices, semiconductor equipment and processes.

The Micro Nano Characterization Facility is perhaps only such facility in the world offering a wide diversity of characterization equipment under one umbrella, for electrical, mechanical, material and optical characterization of nanoscale phenomena. The Packaging and Systems Facility (PASF) offers a unique platform to build a complete package and electronic system around the materials and devices innovations coming out of CeNSE research labs. In addition to above major facilities, there are many thematic labs, such as photonics lab. photovoltaics lab, gas sensors lab, neuromorphic science and engineering lab etc. to serve the needs of different application domains of nanotechnology. The Indian Nanoelectronics User's Programme provides access to CeNSE research facilities to other academic institutes, strategic labs, startups and industries. In addition to research and technology development, a major focus at CeNSE, since its inception, has been to translate the outcome of research into products through incubation of start-ups. The Nanotechnology Incubation Centre at CeNSE, has enabled the creation of about 20 start-ups over the last decade, across a wide range of applications.

9.2 Role played by CeNSE in nurturing memristor technology

Sreetosh: It all began when I relocated to CeNSE after having spent nearly a decade working on molecular memristors (at the National University of Singapore), I was grateful for IISc's exceptional

[28] https://www.sbir.gov/
[29] https://rdifund.anrf.gov.in/
[30] https://www.cense.iisc.ac.in/

support, including a very generous startup grant that facilitated my return to India from Singapore. From the outset, CeNSE provided extensive support, ensuring we could hit the ground running from day one.

Subsequently, a team naturally formed around me, including my father, Sreebrata Gowami, an inorganic chemist, who is the brain behind the materials we now use as memristors. The development of materials and devices also sparked the interest of Navakanta Bhat, who became deeply involved in building CMOS compatible processes, circuits and systems for the memristor crossbars for the technological demonstrations.

The National Nanofab, Nano Characterization facility, Packaging and Systems labs at CeNSE were the key enablers to convert the materials and device knowhow on memristors into a complete chip and system to demonstrate practical applications of neuromorphic technology.

Navakanta: During the early days of my interactions with Sreetosh, after his joining CeNSE, I could quickly recognize the amazing potential of his molecular memristor platform for the future of computing and storage, in the context of neuromorphic technology. To take his insights from materials and devices into computing engines (dot product engine), the obvious next step was building crossbars and interfacing the crossbar with mixed signal electronic circuits. One of my Ph.D. students was working on crossbar architecture for gas sensor array, along with the associated excitation and readout circuits. We could immediately realize that we can pivot this platform to build the dot product engine for memristors. This prior knowhow at CeNSE provided the right platform for Sreetosh to achieve an accelerated system level translation.

9.3 Compelling features of memristors that will make it a technology of the future

Sreetosh: The key advantage of memristors lies in their ability to address several critical bottlenecks in modern computing. Two of them are significant.

- **Overcoming the Memory Wall**: In conventional computers, data is stored in memory but processed in the arithmetic logic unit (ALU). This necessitates continuous data transfer between memory and the processing unit, resulting in significant time delays (latency) and energy consumption—a challenge widely known as the "memory wall." Memristors, with their ability to both store and process data, enable in-memory computing where data can be stored and processed at the same physical location. This would substantially reduce both energy and time consumption for various computations.
- **Enhancing AI and Machine Learning Performance:** Memristors support the development of accelerators, such as dot product engines or spiking neural networks, which can be integrated with existing silicon-based technologies. These accelerators dramatically enhance performance, particularly for machine learning and artificial intelligence applications. While digital computers excel at precision computing, they are not inherently optimized for perception tasks critical to AI.

Memristor-based accelerators are uniquely suited for implementing artificial neural networks, offering a natural advantage in AI workloads.

Navakanta: While memristors offer a unique value proposition to enable analog in-memory computing, it is essential for them to co-exist with transistor based digital electronics. One of the biggest advantages of the memristor technology, in particular molecular memristor, is its compatibility with back end of the line (BEOL) CMOS, with processing temperatures well below 300^{o}C. This would enable monolithically integrated systems on chip (SoC) solutions, with memristor arrays interfaced with digital and mixed signal circuits realized on advanced CMOS nodes. This will extend the performance envelope of current digital systems, beyond what is possible in foreseeable future.

9.4 A brief history of R&D in memristors

Sreetosh: Memristors, or resistive memory devices, have been a topic of research since the 1980s. Early observations of conductance switching with hysteresis often occurred during measurements of thin films due to soft breakdown phenomena. However, the real challenge has always been to translate these effects into practical applications, such as storage or computing, rather than merely treating them as intriguing phenomena of condensed matter physics. A pivotal moment for the field came with a 2008 *Nature* paper demonstrating that such devices could perform computational tasks, sparking global interest in neuromorphic computing.

In India, like elsewhere, numerous studies reported resistive memory effects, making it difficult to highlight just a few. Notably, during my undergraduate years (2010-2012), I collaborated with my father, who was then a professor at the Indian Association for the Cultivation of Science, and we also reported molecular resistive memories. However, harnessing memristors as computing elements introduces significant complexities, and this remains a relatively rare pursuit in India.

Today, besides us at IISc, researchers in IIT Bombay like Udayan Ganguly and in IIT Delhi are advancing neuromorphic technologies with transition metal oxides, though their approaches often involve alternative, non-memristive methodologies.

9.5 Promising memristor materials and architectures

Sreetosh: The most widely explored materials for memristor technology are transition metal oxides such as hafnium oxide, vanadium oxide, and tantalum oxide. However, whether these are the most optimal materials is still up for debate. These materials operate by forming conductive filaments of defects, like oxygen vacancies, in response to applied voltage, altering their conductance. Essentially, this process resembles a soft breakdown mechanism. Interestingly, what might be considered “bad films” in condensed matter physics often turn out to be useful candidates for memristors.

One of the reasons for the popularity of these oxides is their accessibility. They can be fabricated using widely available physical vapor deposition techniques like pulsed laser deposition, atomic layer deposition, chemical vapor deposition, or sputtering. However, the major challenge lies in their reproducibility—variability in performance remains a significant hurdle.

Ideally, this soft breakdown mechanism should be replaced by a consistent, thermodynamic process, such as an electronic or phase transition, which is both uniform and controllable. Molecular memristors based on transition metal complexes have demonstrated far superior performance compared to oxide-based memristors. Despite this, their popularity remains limited due to the specialized nature of their synthesis, fabrication, and integration with silicon technology. This requires multi-disciplinary expertise spanning chemistry, physics, material science, fabrication and electrical engineering.

However, given the substantial advantages offered by molecular memristors, it is high time the scientific and technological community invested more effort into overcoming these challenges and exploring their full potential.

Navakanta: Bringing in any new material platform into CMOS (Complementary metal–oxide–semiconductor) technology platform requires comprehensive assessment of contamination control in the silicon fabs. Hence it is relatively easy to repurpose an existing material from the CMOS process flow. As a result, one very specific memristive material from the class of transition metal oxide, Hafnium oxide, has been adopted by some foundries, since it was already being used as gate oxide of CMOS transistors. While this has been adequate for Resistive RAM applications (RRAM) as single bit storage cell, its utility in truly analog memristor application has been limited due to inherent noise in storing the exact value of resistance, with fine resolution, given the statistical nature of soft beak down process in this class of materials. In contrast the deterministic nature of molecular material coupled with the ease of integration in CMOS process flow make it an ideal choice for truly analog in-memory computing applications.

9.6 Memristors compared to traditional CMOS technology in terms of energy efficiency, performance, and scalability

Sreetosh: It depends on the application. Unlike the general-purpose computers that dominated the last two decades, today's computing hardware is increasingly being tailored for specific uses, known as application-specific integrated circuits (ASICs). Memristor-based neuromorphic elements are unlikely to replace CMOS technology entirely, but they can significantly enhance existing CMOS systems. For instance, in our recent *Nature* paper[31], we have shown that for certain AI and signal processing tasks, memristor-based neuromorphic platforms can deliver upto 460 times the energy efficiency compared to powerful CPUs and 220 times that of high-end GPUs.

Navakanta: The debate should not be whether memristors will replace traditional digital CMOS technology. Instead, the narrative should be on utilizing the memristors to enable very substantial value addition to the existing digital CMOS technology by building AI accelerators for all tasks involving vector matrix multiplication such as perception, inference, training etc., which form the basis of AI workflow.

[31] Sharma, D., Rath, S.P., Kundu, B. *et al.* Linear symmetric self-selecting 14-bit kinetic molecular memristors. *Nature* 633, 560–566 (2024). https://doi.org/10.1038/s41586-024-07902-2

9.7 Memristors and in-memory computing

Sreetosh: Modern memory devices can write data in nanoseconds, and processors operate just as quickly. The real challenge lies in the communication between these two components. Because memory and processing units are physically separate, data transfer takes microseconds and consumes up to a thousand times more energy than the actual switching of memory elements. By combining computing and storage in the same physical location, we can eliminate this communication overhead, which is currently the biggest contributor to overall computing energy consumption.

For digital applications, the transition to such in-memory computing is relatively straightforward. It requires new data management protocols and a modified architecture, but the fundamental computing model remains intact. Boolean operations will still form the basis of computation—the difference is that in stateful in-memory computing, the logic inputs and outputs are stored directly as conductance states in memristors, streamlining the entire process.

Memristors are also ideally suited to replace flash memories. These can be used as non-volatile memories. It has been shown that the endurance of these devices are a million times better than flash memory devices.

Navakanta: Memristors will enable both in-memory computing and near memory computing. Many AI applications require billions of parameters and they are necessarily stored in off-chip DRAMs through High Bandwidth Memory (HBM) architecture. Since HBM relies on 2 (or more) different chips (dice) with interposer like interconnect technology, the data movement tends to be very inefficient. It is quite possible that 3D integration of memristor based ultra-high density non-volatile memory along with in-memory mixed signal computing chip can enable on-chip storage of entire set of parameters for AI models (such as LLM), completely eliminating the off-chip DRAM access. This fusion of In-memory compute along with near-memory compute can be a game changer in edge AI inference and training of sophisticated models.

9.8 Potential applications of memristors beyond memory – brain inspired and edge computing

Sreetosh: Memristors have the potential to enable a wide range of applications and computing architectures, but their impact depends heavily on the specifics of their use. For instance, if memristors are employed as dot product engines, accelerators that enable one-stepped matrix multiplication, there is no need for entirely new algorithms or architectures. While integrating these platforms with existing technology does require process development and a robust software stack for efficient data routing, it is more of a straightforward enhancement than a complete overhaul. This "add-on" approach can significantly boost performance.

However, if we envision using memristors for bio-inspired chaotic computing, the challenge becomes much more complex. This would necessitate a complete reimagining of computing architecture, starting with a deeper understanding of how the brain itself operates. Memristor-based computing platforms

offer a range of applications beyond memory storage. As discussed earlier, they can enable in-memory computing, where logic calculations are performed directly within the memory, eliminating the need to physically move data. Additionally, memristors can act as dot product engines, accelerating matrix multiplications, a fundamental operation in modern computing.

Looking ahead, these platforms hold potential for numerous futuristic applications, including oscillator-based computing, spiking neural networks, and chaotic computing. These approaches, inspired by the biological brain's working principles, offer significant promise but would require a comprehensive reimagining of computing architectures.

Memristor-based accelerators can also enhance sensor networks at the edge of the Internet of Things (IoT). By enabling local intelligence at the edge, they can significantly reduce data traffic between edge devices and the cloud, which accounts for a substantial portion of computing energy consumption. Through precise chemical design, memristors can even serve as dual-purpose sensors capable of both storage and computation, drastically minimizing energy use in IoT systems.

On the hardware front, the primary challenge lies in scaling up and integrating these platforms with existing silicon chips to create a system-on-chip. Our current efforts are focused on addressing this challenge. To achieve optimal computing performance, robust data management and routing software stacks are equally essential, and we are actively working on these aspects as well.

Navakanta: **In the near term, the application of brain inspired analog in-memory computing would be for edge inference and model fine tuning using memristors**. This requires judicious partitioning of computing between a tiny digital engine which does most pf the house keeping on the chip, and multiple tiles of memristor based cross-bars for analog MAC (Multiply-Accumulate) operations. There are interesting opportunities in developing new system level architectures for such mixed signal brain inspired computing. In the long-term edge training would transform the compute capabilities of edge devices.

9.9 Memristors can play in addressing global challenges like computing energy crisis, data privacy and climate prediction

Sreetosh: Neuromorphic computing based on memristive architectures has the potential to confront several pressing global challenges while advancing the broader vision of sustainable and democratized AI.

- **The computing energy crisis:** Digital infrastructure is becoming one of the largest energy consumers on the planet. Data centers already account for a substantial and rapidly growing fraction of global electricity demand, driven largely by AI workloads. As model sizes and training cycles scale exponentially, energy consumption risks outpacing grid capacity in many regions. Neuromorphic architectures, by performing computation in-memory and in analog, dramatically reduce data movement — the dominant source of energy loss in modern systems. This shift from

data shuttling to physics-based computation offers a pathway toward sustainable AI, where performance scales without proportional increases in power consumption.

- **Democratization of AI and data sovereignty:** Today, AI training is overwhelmingly centralized in large data centers, concentrating computational power and raising concerns about access, privacy and equity. High-precision molecular neuromorphic hardware opens the possibility of performing both inference and training directly at the edge. When models can be trained and adapted locally — on-device and in analog — data need not leave the point of generation. This preserves privacy, enhances security, and enables communities, institutions and even individuals to participate in AI development without reliance on centralized cloud infrastructure.
- **Climate resilience and distributed intelligence:** Neuromorphic systems are well suited to real-time, distributed sensing and adaptive modeling. By embedding low-power intelligence in local weather stations, agricultural sensors and environmental monitoring nodes, edge-based neuromorphic platforms can analyze real-time data alongside historical climate patterns to forecast extreme events. Such distributed intelligence could provide early warnings in vulnerable rural regions, supporting climate adaptation strategies while operating within strict energy constraints.

Together, these capabilities position neuromorphic computing not merely as a new hardware paradigm, but as an enabling foundation for energy-efficient, privacy-preserving and widely accessible artificial intelligence.

Navakanta: **This would become a foundational technology with capabilities that would influence almost all application domains.** This would be similar to multiple inflexion points we have had in the evolution of Silicon technology over last several decades, such as transition from BJTs to MOSFETs; NMOS to CMOS; planar transistors to FinFETs to name a few.

9.10 Memory, AI ML, and IoT will be the big markets for memristors

Sreetosh: The potential applications of memristor-based technologies span a wide range, making this an incredibly exciting opportunity. Due to their analog nature, enabling both storage and computing, these devices can significantly impact several sectors.

Flash memory is likely to be one of the first practical applications, followed by transformative advancements in AI and ML. Since our platform emphasizes data privacy, it is particularly suited for sectors where confidentiality is paramount. **Two areas that immediately come to mind are the strategic sector, including defense and military, and healthcare.** We are already in extensive discussions with India's defense research laboratories regarding potential collaborations.

Another key area of impact is the energy-efficient implementation of IoT systems. By enabling distributed computing at the edge, memristor-based platforms can dramatically reduce the energy consumption associated with data transmission over the internet. AI accelerators, high density memories, edge computing are likely to be the killer applications.

Navakant: As I said, the application domains will be so vast that we would be limited only by our imagination.

9.11 Trends in commercialization of memristors

Sreetosh: Memristor technology began its journey in 2008, and while it has been 17 years, we are yet to see a product that comprehensively outperforms a CPU or GPU. The potential of this technology has always been immense, but a critical technical issue has not been adequately addressed.

The fundamental problem lies in the materials, more specifically the inherently probabilistic stochastic filamentary mechanism that drives existing memristors. While this mechanism works, it is inherently difficult to control and is not scalable. For memristors to become a viable and scalable technology, a more controlled and predictable mechanism is essential. As a result, most research publications in this field continue to focus on addressing material challenges.

Defective oxides remain the most widely used materials for memristors because they are readily available and relatively simple to fabricate into devices. In many cases, laboratories repurpose devices that were initially deemed "failed" or "imperfect" for condensed matter studies as memristors. This approach highlights the need for a paradigm shift in materials research for memristor technology. **Once the materials challenge is overcome, it can unlock rapid progress across circuit design and the associated software stack. We believe the memristor platform we have developed represents one of the most optimized material systems realized to date, with the potential to fundamentally reshape the trajectory of neuromorphic computing.**

Navakant: For any new technology to come to mainstream we need a killer application. **For memristors, the ideal application is AI, which is all pervasive.** Given the exponential growth in AI, cutting across all domains, we believe that we are at the right place at the right time to make this a big commercial success.

9.12 Commercialization readiness of CeNSE's memristor technology

Sreetosh: **Today, our memristor crossbar technology is at TRL 7/8 and can be integrated onto a printed circuit board to create an electronic card capable of accelerating matrix multiplication for various applications including image, signal processing and few AI workloads. The next step involves developing a complete System-on-Chip (SoC) by integrating the crossbar with silicon chips fabricated in foundries, along with the creation of software stacks to demonstrate its utility in specific computing applications.**

Navakant: We are already working on multiple fronts for commercialization. **We are developing a Minimum Viable Product (MVP), a mixed signal SoC on 22nm CMOS capable of achieving compute performance of multiple TOPS (Terra Operations Per Second) with extreme energy efficiency. In parallel we are also developing capabilities to perform post-processing on foundry made chips so that the memristor based cross bars for in-memory computing can be integrated monolithically onto the**

single chip. We expect to demonstrate edge AI applications such as implementation of YOLO model on single chip by end of 2026. We will be incubating a start-up at IISc sometime during 2026 for commercialization of this technology.

9.13 What capabilities are required to commercialize CeNSE's memristor technology

Sreetosh: Ideally, both the silicon SoC and the memristor stack would be fabricated entirely within India. However, in the absence of advanced foundries comparable to TSMC, the most pragmatic and scalable pathway is to have the silicon base fabricated at established global foundries and subsequently integrate our molecular memristor technology using custom back-end protocols developed in-house.

To move the platform beyond TRL 7/8, the immediate priority is the establishment of a dedicated pilot fabrication line within a Class 100 cleanroom, where the molecular layer can be implemented as a back-end-of-line process on foundry-fabricated chips. While we currently utilize the shared CeNSE cleanroom facility, this arrangement introduces variability and limits process control. **A dedicated pilot line is therefore the first critical step toward reproducibility and scale. Once the process is fully stabilized and optimized, it can be transferred to a commercial fabrication facility for volume manufacturing.** On the chip design side, the next phase should be driven through a start-up structure, enabling the recruitment of top-tier chip designers and the focused development of a fully integrated SoC combining a processor core with memristor crossbar arrays on a single platform.

Navakant: The key is to decouple the front end of the line (FEOL) process on advanced CMOS nodes (sub-22 nm, going all the way down to 3nm) and back end of the line (BEOL) process which requires relatively less demanding lithography capabilities. Thus, the wafer processing can potentially be done either in single foundry or different foundries. The molecular materials will have to be integrated with CMOS compatible processes on 300mm wafers in BEOL process, leveraging CMOS compatible spin coating process. We also need to have facility to be able to produce the starting material (molecular material) at scale so that large scale commercial wafer production is possible.

Having worked in semiconductor sector for more than three decades, I am of the strong view that the first mover advantage is very crucial to capture the market and establish industry leadership. This requires us to be extremely agile and pragmatic. **While we protect and secure all the IPs related to Molecular Neuromorphic AI technology, design, and products, we should fast track the large-scale manufacturing and production through leading foundries abroad, since we do not have 300mm wafer processing capabilities in India for both FEOL and BEOL. This is even more crucial for Semiconductor chips for Edge AI inference and training, where the demand is humongous and the required volumes are astronomical. As and when the 300mm wafer facilities are established in India, we can then focus on domestic manufacturing, which is also important from strategic perspectives.**

9.14 Challenges in integrating memristors with existing CMOS fabrication processes

Sreetosh: Molecular materials used to make memristors are not traditionally handled by industrial fabs, so their integration requires some deviation from existing practices. However, the processes we use to develop this technology are CMOS-compatible, which is a significant advantage. Moreover, there are precedents for integrating unconventional technologies, such as OLEDs and solar cells, with silicon circuits. Similarly, memristors will be integrated with silicon chips not as a front-end-of-line process but as a back-end-of-line (BEOL) process.

Navakant: CMOS fabs tend to be conservative in introducing any new materials. The advantage here is that the material can be introduced at the very end of the CMOS process flow. We still need to work with some commercial fabs with large manufacturing capabilities who will adopt this material into their foundry. We are exploring multiple options to achieve this.

9.15 Examples of companies and startups that have commercialized similar memristor technologies

Sreetosh: Currently, leading startups like Tetramem and Crossbar (based in the US) are at the forefront of commercializing neuromorphic chips powered by memristors. Major companies such as IBM and Intel have given their shot at neuromorphic computing in the previous decade. They experimented with CMOS technologies to build a brain like platform, but those early attempts achieved limited success. The major learning was that CMOS is not suitable for this and increased the interest in alternative technologies including memristors. IBM has since been exploring various alternative approaches such as phase change memories, memristors, spintronics. Beyond the tech giants, strategic sectors such as DARPA are significant funders of neuromorphic research in the US and represent potential customers for this technology.

Today, China, particularly Tsinghua University, is perhaps the biggest hub for memristor research. Tsinghua has developed its own 25nm fabrication facility for neuromorphic chips and is making rapid progress in AI chip development. While commercialization details remain scarce, the pace of technological advancement and investment suggests they already have a substantial market.

9.16 Why are memristors important for India

Navakanta and Sreetosh: Memristor technologies hold significant promise across strategic sectors, edge computing, and data security.

- **Strategic sectors:** Platforms such as surveillance drones operate under extremely tight power budgets. Memristor-based, low-power onboard inference engines could process imagery locally and transmit only frames flagged as potential threats, rather than streaming all captured data. This selective communication could substantially extend operational endurance — potentially by

an order of magnitude. A similar advantage applies to satellites and other space-based systems, where energy efficiency directly translates into mission longevity and capability.

- **Edge computing:** Wearable devices, including smartwatches, can integrate memristor-enabled AI inference engines to perform real-time tasks such as language translation without relying continuously on cloud connectivity. The same capability extends to distributed IoT sensor networks in smart cities and intelligent buildings, where local inference reduces latency, bandwidth use, and energy consumption.
- **Data security:** Beyond defense, space, or atomic energy applications, edge-based inference is increasingly relevant for personal data — particularly health information. Memristor-based systems enable processing directly at the point of data generation, minimizing the need to transmit sensitive information to the cloud and thereby strengthening privacy and data sovereignty.

About itihaasa Research and Digital

itihaasa Research and Digital (www.itihaasa.com) is a non-profit Section 8 company that analyses the evolution of technologies, R&D, and innovation in India. We have published reports on the R&D landscape of brain science, quantum technologies, and AI/ML in India. We were the knowledge partner for the US - India Artificial Intelligence (USIAI) initiative of the Indo-U.S. Science and Technology Forum (IUSSTF). Kris Gopalakrishnan, co-founder of Infosys, is the founder and Chairperson of itihaasa Research and Digital. Our flagship project is itihaasa history of Indian IT, a first-of-its-kind free digital museum that recounts the history of Indian IT since the 1950s. We have also authored the award-winning book Against All Odds: The IT Story of India. You can reach out to N Dayasindhu Ph.D. (dayasindhu@itihaasa.com) and Krishnan Narayanan (krishnan@ itihaasa.com) for any queries on the study.

About CeNSE, IISc

The Centre for Nano Science and Engineering (CeNSE) was established in 2010 to pursue interdisciplinary research across several disciplines with a focus on nanoscale systems. Current research topics include, but are not limited to nanoelectronics, MEMS/NEMS, microfluidics, nanomaterials and devices, photonics, nano-biotechnology, solar cells, neuro-biology, quantum and neuromorphic computing, sensors, devices, thin-films, and optoelectronics.

You can reach out to Sreetosh Goswami (sreetosh@iisc.ac.in) and Navakanta Bhat (navakant@iisc.ac.in) for any queries on CeNSE.