Nature computes better

Opportunity space

v1.1

Suraj Bramhavar, Programme Director

CONTEXT

This document describes an opportunity space - an area that we believe is likely to yield breakthroughs, from which one or more funding programmes will emerge.

You can find out more about opportunity seeds within this space and meet the Creators **here**. (www.aria.org.uk/nature-computes-better-seeds)

In tandem, our programme hypothesis related to this opportunity space has now been published. You can read this document **here.** [PDF] (https://www.aria.org.uk/wp-content/uploads/2023/11/ARIA-Unlocking-AI-compute-hardware-v1.0-1.pdf)

We have also launched a programme, Scaling Compute, in this opportunity space. Find out more **here**. (aria.org.uk/scaling-compute)

This opportunity space is not currently soliciting feedback — you can stay up to date with this opportunity space, plus others across ARIA, **here**.

An ARIA opportunity space should be

+ important if true (i.e. could lead to a significant new capability for society),
+ under-explored relative to its potential impact, and
+ ripe for new talent, perspectives, or resources to change what's possible.

SUMMARY

We can redefine the way computers process information by exploiting principles found ubiquitously in nature. We can better understand how the natural world around us performs computation and build dramatically more efficient computers.

BELIEFS

The core beliefs that underpin/bound this area of opportunity.

1. The growth of AI exacerbates an already unsustainable demand for compute → we need alternative scaling pathways.

2. Natural systems are orders of magnitude more efficient than silicon microprocessors at a wide range of computational tasks → a stronger understanding of how living systems compute is needed to advance both engineering biology and the creation of new hardware.

3. Investigating the role of statistical physics and nonlinear dynamics in novel hardware represents a significantly underexplored opportunity → exploiting these approaches is likely to yield new modalities for AI processing.

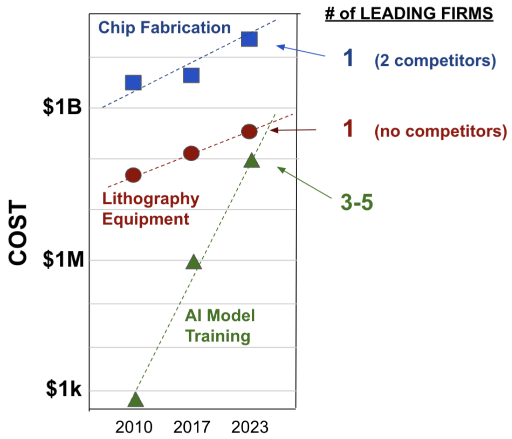
4. Modern AI has massive and broad applicability but is underpinned by a narrow set of mathematical kernels → this presents a unique opportunity for the development of next-generation computing paradigms.

OBSERVATIONS

Some signposts as to why we see this area as important, underserved, and ripe

1. For the first time in computing history, increasing performance requires exponentially increasing costs. The economics of Moore's Law are fundamentally broken.

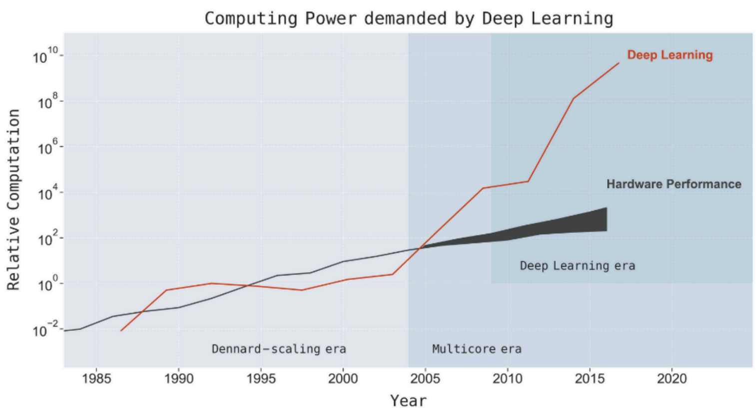
2. Entry costs to innovate at the leading edge have skyrocketed, leaving few firms with the ability to participate.



3. Figure 1: This graph shows the cost increases of three different major parts (chip fabrication, lithography equipment, AI model training) of the ecosystem over the function of time. It shows a small number of firms remaining with the ability to operate at the cutting edge.

4. AI is now the primary economic driver of computational power, and its capabilities are primarily derived from well-known mathematical primitives.
   Note from Suraj - it provides a unique combination of specificity and impact.

5. Access and governance of this critical technology has become a major geopolitical issue



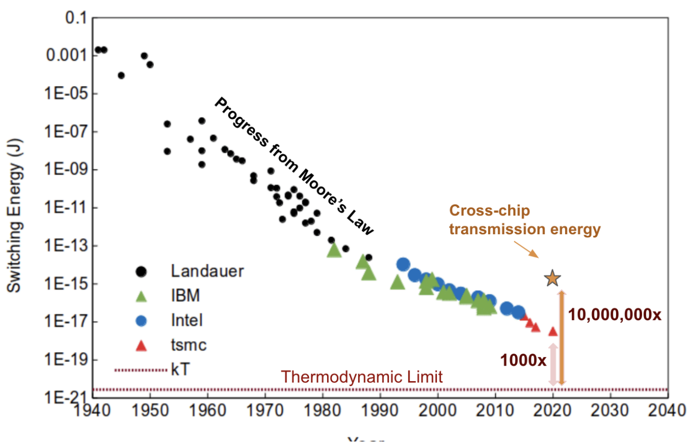
6. Figure 2: This graph shows exponentially increasing demand of compute power as a result of deep learning

7. Algorithmic advances are plentiful (and being investigated exhaustively), but rarely change the fundamental dynamics of the industry. Alternative hardware

paradigms present a riskier alternative but can ultimately lead to much larger performance improvements.

8. Recent popular demonstrations (e.g. Stable Diffusion, ChatGPT) are increasingly showing impressive new capabilities by incorporating phenomena found in nature (stochasticity/probability) as core features.

    Note from Suraj - these features are wastefully recreated within a paradigm designed specifically to avoid them.



9. Figure 3: This graph shows energy required to flip a bit over the function of time and the gap between this, and fundamental physical limits.

10. The energy required to flip a bit in the digital domain has plateaued at about 1000x the fundamental limit. Sending that bit across a chip requires yet another 10,000x more energy. A single user's session with chatGPT requires ~150x more power than a human brain consumes performing ALL functions.

    Note from Suraj — there is ample room for improvement.

11. Vibrant communities have developed around quantum computing and neuromorphic computing, but far less attention has been paid to energy-minimization in physical systems as an extremely efficient computational mechanism.

12. Scalability, manufacturability, and true economic advantage have historically limited the adoption of new computing technologies. These commercial

constraints can serve as an opportunity to galvanise creative approaches and improve the chances for new technologies to break through.

SOURCES

A compiled, but not exhaustive list of works helping to shape our view and frame the opportunity space (for those who want to dig deeper)

- The End of Moore's Law
- Compute Demand in the AI Era (Fig 2)
- The Geopolitics of AI and Semiconductors (1,2,3,4,5)
- The Role of Probabilities in ChatGPT
- Energy-Based AI Algorithms
- Energy-Minimization in Hardware
- Thermodynamic Linear Algebra
- Computing with Biology
- A Burgeoning Community
- Computing with Light
- AI & Novel Electronic Memories
- The Evolution of Transistor Switching (Fig 3)

ENGAGE

This opportunity space is not currently soliciting feedback – you can stay up to date with this opportunity space, plus others across ARIA, **here** (www.aria.org.uk/opportunity-space-updates/).