What Exactly is the Status of Moore's Law?

Is Moore's Law Dead? Is Moore's Law Alive and Well?

This white paper seeks to expound on what Moore's Law is, the current hurdles facing processor manufacturers related to Moore's Law, possible innovations, and how Coherent Logix's HyperX® technology has the potential to extend the benefits of Moore's Law into the next ten years.

Gordon Moore: A Fairly Accurate Visionary

In 1965, Gordon Moore predicted that the number of transistors that could fit on a microchip would double every two years; and as a result, computers would become twice as powerful in the same period. This is known as "Moore's Law".

> *"Making those chips smaller creates bigger, better yields, which allows them to actually keep the price the same or lower it with more performance over time."*
>
> **- Avi Greengart, President and Lead Analyst at Techsponential.**

Over the last 17 years, innovation surrounding the miniaturization of the transistor has fallen behind the exponential growth that Moore predicted. Processors coming off assembly lines today have billions of transistors. Based on Moore's Law, the typical computing power available to the average person should be much more powerful and much more affordable than what is currently available. Miniaturization has been hindered by specific obstacles that chip designers and manufacturers now confront.

Transistors: Size Matters...But Why

In 1971, the transistor was 10 microns wide and the gap between transistors was visible to the naked eye. Today, some high-end consumer devices, such as Apple's M1 processors, are making the move to 5 nanometers (nm), which is only 10 atoms wide. IBM has gone even further and built a prototype chip with 2nm transistors, allowing the company to place 50 billion transistors onto a chip the size of a fingernail, but it will take years before it can be an affordable, accessible product.

As chip size decreases, the risk of defects and failure mechanisms increases. Although producers have seen power and performance gains, further miniaturization has exponentially expanded the time, cost, and effort required to deliver a viable solution.

Given the challenges that miniaturization introduces, we have hit what is colloquially known as "Moore's Wall".

The Wall of Physics

As transistors get smaller than 7nm, designers must watch for a phenomenon known as "Quantum Tunnelling". Quantum tunnelling is when electrons are no longer contained by their barriers, resulting in unexpected and incorrect computations. As transistors get smaller, the need to mitigate the risks associated with quantum tunnelling grows.

Additionally, manufacturers struggle to offer consumers cost-effective processors above 4GHz. This is due to another stumbling block known as "Dennard Scaling", whereas power usage stays proportional to the area of that transistor. As a result, power (that would otherwise be utilized by the processor for crunching computations) escapes as heat.

Chips with smaller transistors are designed to process more while utilizing the same amount of power. Performance can be increased further with additional power, but there's potential for power loss due to thermal leakage and excessive heat. Today's processors require sophisticated cooling mechanisms to avoid overheating because of the millions of transistors in operation in conjunction with their power requirements. By 2006, the power lost to heat on the average processor became so significant that attempting to design a faster processor became impractical because more energy would be lost to heat dissipation than utilized for processing data.

With conventional processing gains becoming harder to produce, some manufacturers opted to shift towards multi-processor designs and allocate more and more functions to Graphics Processing Units (GPUs). This was a practical way to increase overall performance without having to directly overcome design challenges associated with miniaturization.

With the laws of physics seeming to create a hard stop, many manufacturers are wondering if it makes fiscal sense to try to overcome them or work around them.

The Wall of Cost

In additional to the physics-related challenges manufacturers have with creating a smaller transistor, there is often an added, significant expenditure: "respins". A respin is necessary after an application-specific integrated circuit (ASIC) is finalized within a chip but fails with real hardware. Any issues must be analyzed, properly identified, and re-engineered. And the associated ASIC production must be overhauled to incorporate the fixes. This is a very timely and expensive process.

The Wilson Research Group / Siemens found that only 32% of 2020 chip designs claimed first-silicon success and roughly one in four ASIC respins are due to power issues, according to Harry Foster, Chief Verification Scientist at Siemens EDA.

A chip can only utilize as much power as it is able to dissipate, while maintaining a temperature that will not cause failures. Robust thermal analysis must be performed, but that is difficult for vendors because of the timescales involved. When considering that devices may be operating at multiple GHz, the vector sets subject to analysis are large. And runtimes for parsing associated vector sets are lengthy.

The Wall of Practicality

Moore's Law, via yet to be realized innovations, may ultimately remain prophetic. IBM's 2nm transistor will undoubtedly be a huge leap forward in performance and heat consumption, with boasts of a potential 45% performance increase. But production of these 2nm chips isn't slated to begin until 2025.

And with transistors that are roughly five atoms wide, initial production costs are sure to be fantastic and rigid standards will need to be implemented at foundries to accommodate the intricacies required for successful, consistent production.

Intel placed a new $20 billion semiconductor manufacturing site in New Albany, Ohio – a suburb of Columbus. In 2022, the groundbreaking was recognized as the largest single private-sector investment in state history. But production is not expected to come online until sometime in 2025.

The future looks bright. But design costs and production costs will certainly be baked into upcoming chips.

According to International Business Strategies (IBS), integrated circuit (IC) design costs have jumped from $51.3 million for a 28nm planar device to $297.8 million for a 7nm chip. Costs have increased to $542.2 million for 5nm transistors, which is over 10x the cost of previous designs. Anything 3nm and below will most likely reach a staggering $1.5 billion in design costs.

> *"3nm will cost $4 billion to $5 billion in process development; and the fab cost for 40,000 wafers per month will be $15 billion to $20 billion."*
>
> **- Handel Jones, Chief Executive of IBS**

For these reasons, manufacturers may stay at certain larger chips longer - such as 16nm, 14nm, and 7nm. And given the potential costs and reliability issues that may manifest in a 2nm transistor, it may be over a decade before it becomes feasible to make an investment.

The HyperX® Difference

Smashing Through Moore's Wall

Even when using the smallest transistors, chips with Field Programmable Gate Arrays (FPGAs) have a big dilemma.

By design, all FPGAs encounter a bottleneck when the flow of data is routed off-chip and assigned to registers in memory for computation. Whether using 2nm transistors or 5nm, the flow and processing of data are throttled by access to tangential memory.

Coherent Logix's HyperX architecture demonstrates that it's exponentially more efficient to incorporate the memory with the data flow, rather than detour the data flow off-chip.

With HyperX, memory is directly integrated with processing elements. Because processing data in memory is local, increases in transistor size correlates to a performance increase, whereas common chip design would see a decrease in performance. This makes HyperX ideal for organizations that need to process large amounts of data.

With data flow and computing optimized, HyperX can process multiple, parallel functions, making it a reliable solution for wide data streams (such as video, 5G / 6G, conferencing, and real-time analytics). And because HyperX architecture uses larger, easy-to-manufacture transistors, powerful processing potential maintains a reasonable cost.

The HyperX next-generation processors overcome the challenges invoked by the potential thresholds of Moore's Law. And HyperX accomplishes this using a revolutionary approach that are exclusive to Coherent Logix's processor design and functionality.

Designed To Work Smarter...Not Harder

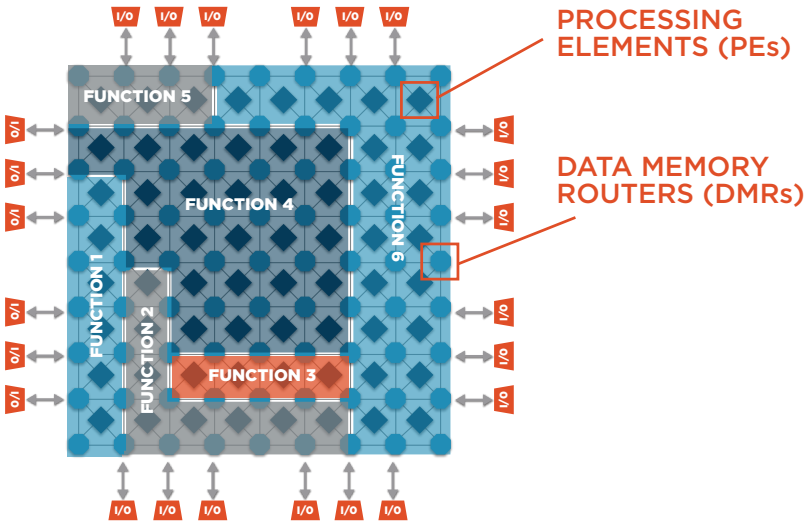
The vast majority of processors currently being manufactured are based on Von Neumann architecture, which consists of shared memory for programs and data. The Von Neumann processor operates fetching and execution cycles by constantly moving data to and from memory in a consecutive fashion. Bottlenecks are unavoidable when data flows rely on this common architecture.

HyperX, however, employs HyperX Fabric, which is comprised of an array of Processing Elements (PE) and Data Memory and Routing (DMRs) units located together with PEs having direct memory access, eliminating the need to fetch and load data. The layout of these arrays is such that each PE is surrounded by four Data Memory and Routing (DMRs) units and each DMR is surrounded by four PEs. This fabric enables end-to-end data pipelining and massive parallel processing.

PE's within the HyperX® Fabric only utilize power and resources when they are actively working on something. Unlike Von Neumann-based chips, when HyperX PE's are idle, they consume virtually no power. And when they are actively working, they are 4x to 10x more powerful than competitors due to their instructions and memory being local.

Smashing Through The Wall of Physics

HyperX processors are designed with transistors that are optimized at 12nm. This mitigates the risk of quantum tunnelling. Additionally, HyperX's ultra-low power consumption, remarkably low heat generation, and dynamic power utilization greatly minimize the need for thermal management. Overheating due to Dennard Scaling is a challenge that HyperX processors easily avoid.



HyperX Fabric

Smashing Through The Wall of Exponentially Increasing Costs

HyperX has improved the cost model for high-performance chips through the use of 12nm transistors, which reduces the need for costly respins. With the incorporation of software-defined hardware, HyperX expedites time-to-market by reducing the time and cost of development by approximately 75% when compared to the leading FPGAs. And HyperX streamlines the development cycle via its cycle-accurate simulator, intuitive debugging options, and by enabling code portability across use cases. With streamlined development process combined with software-defined hardware, the cost-to-value potential is unmatched.

Smashing Through The Wall of Impractical Inefficiency

HyperX Technology incorporates a disruptive approach that combines powerful concepts to create a comprehensive solution for any application. These innovations include:

- A software-definable processor architecture that allows application code to be stored and executed within the node – and is not subjugated to traditional external boundary interactions, such as accessing external memory.
- The optimization of data movement via Software Defined Functions (SDFs) – whereas data can be applied in a pipelined fashion. SDFs working in conjunction with other SDFs simultaneously can take full advantage of powerful parallel processing.

Given all these benefits, combined with the fact that HyperX can be developed or upgraded directly in the field, Coherent Logix has managed to avoid Moore's Wall by circumventing Moore's Law. HyperX Technology costs remain competitive, and value remains high because of a common architecture, C-programmability with the most advanced embedded development environment, and unparalleled computing efficiency.

Moore's Wall may become a factor for Coherent Logix's HyperX Technology, someday. But for those looking to accelerate a successful go-to-market strategy, there are no limits to the progress and potential that can be experienced using HyperX.