CORNAMI Intelligent Computing

Next Generation Green Computing

*Cornami is creating a new "green" semiconductor category by increasing the efficiency (performance per watt) **of computation by orders of magnitude above any other commercially available processor***. Data center power consumption and cooling costs are growing dramatically with server power consumption increasing overall by 266% since 2017[1]. While great efforts have been made to address the efficiency and cost of providing data center power and cooling[2], Cornami's reconfigurable and scalable computational fabric architecture increases processing efficiency by reducing the overall power and cooling requirement needed for today's computational loads. Cornami's computational fabric:

- Allows application developers to take full advantage of the concurrency and pipelining intrinsic within an algorithm, directly in hardware, increasing overall computational efficiency by assembling data values in the same production line approach as a factory produces goods.

- Eliminates the requirement for caches & memory to speed software execution reducing overall power consumption for a given computational load, and,

- Enables reconfigurable systolic arrays to provide maximum efficiency in application execution.

Cornami's computational fabric operates on a stream of continuous data and its native control structure is defined as a topology (e.g., Gantt chart). By assembling data values in the same way that factory production lines assemble goods, maximum algorithmic concurrency and pipelining are easily defined. Performance and throughput are increased by expanding the size of the computational fabric. Current processor architectures restrict a developer's ability to take full advantage of the inherent concurrency and pipelining within an algorithm resulting in the inefficient use of silicon resources. Cornami liberates developers from those restrictions and allows the full potential of an algorithm to be expressed to the hardware for execution. This delivers several orders of magnitude higher utilization of the silicon resource for application processing, several orders of magnitude higher application performance, and the ability to further scale performance proportional to the size of the scalable computational fabric. As an example, a secure variant of blockchain called Zero Knowledge Proof (ZKP) executes 100x faster on a Cornami server than a 96-core x86 server drawing comparable power. In almost every case, Cornami's topology-based architecture dramatically boosts performance per watt relative to any other available approach.

Cache and memory operations are a critical part of application execution on conventional processors and represent a significant portion of the overall power expended and latency in its response. Cornami's computational fabric approach does not require these structures, significantly reducing both power and latency. The linkages between processor cores in the computational fabric are dynamically configured to match the application topology and data flows from one processor core to another without the need for intermediate storage. In this way, cache and/or memory operations are not required as the data remains in flight during application execution. The power and latency savings that result from this approach are substantial and can represent approximately four orders of magnitude reduction for power and at least two orders of magnitude reduction in latency for a comparable computational load.

Cornami's scalable computational fabric architecture introduces the concept of a reconfigurable systolic array that dramatically increases performance per watt by easily matching application requirements with a custom systolic array geometry. Systolic arrays are a network of coupled processors that compute partial results from upstream data and then pass it downstream. The benefits of systolic arrays are well known to processor designers and have become a key feature in almost every current CPU and GPU offering.

However, in today's conventional processor designs, systolic arrays are a one-size-fits-all approach, often a non-reconfigurable rectangle, statically defined at the time the chip is designed. In most cases, an application is not a perfect fit. Either a systolic array is too small, requiring processing to be segmented and worked on iteratively reducing overall performance, or too large leaving much of the array idle and wasting valuable chip resources, or- it is fundamentally different from a rectangle. The use of one-size-fits-all systolic arrays waste valuable resources as each application has different requirements. Because Cornami's computational fabric can be reconfigured to match an application's needs, application-specific systolic arrays are easily implemented and are not bound by a fixed chip design which delivers higher performance per watt efficiencies than any other existing processor architecture.

Cornami's reconfigurable and scalable computation fabric allows for substantially higher application performance and throughput per watt than conventional processor offerings; reducing the power and latency needed to perform today's data-intensive computational workloads. Cornami's architecture addresses the challenge of today's high economic problems in a highly efficient way that eases the growing data center power crisis by 1) enabling application developers unrestricted use of the available concurrency and pipelining within an algorithm, 2) maintaining data in flight to eliminate the need for caches and memory, and 3) allowing for application-specific systolic arrays to be dynamcially defined. Cornami's ability to scale performance with the size of a computational fabric and effect a lower overall footprint in power usage represents the next generation in green computing.

1. https://datacentremagazine.com/articles/efficiency-to-loom-large-for-data-centre-industry-in-2023
2. https://www.datacenterfrontier.com/cloud/article/21439020/the-eight-trends-that-will-shape-the-data-center-industry-in-2023