**Parallel Computing of Graph-based Functions in Re-RAM**

CMOS is reaching its physical boundaries because to the ongoing reduction in feature size, prompting the search for viable successor technologies beyond the scaling limit. ReRAM is a nonvolatile memory technology with low power consumption, built-in computing capabilities, and excellent logic synthesis efficiency.

In the Binary Decision Diagram (BDD) technique, the Multiply-Accumulate (MAC) operation is employed instead of logic primitives. The BDD nodes are mapped to parallel MAC operations right away.

The And-Inverter Graph (AIG) is an automated compilation approach based on in-memory computer architecture. Any Boolean function can be converted.

Wires connecting two-input AND gates that correspond to nodes are shown by the edges of graph-based representations, which can be augmented to depict inverters between nodes.

ReRAM MAC Computation is used to execute several MAC processes at the same time.

Because they need the fewest operations and devices, MIGs are the state-of-the-art graph structure for ReRAM-based synthesis in Graph Based Computation.

If a calculation uses one wordline and multiple bitlines, it is wordline parallel; if it uses one bitline and multiple wordlines, it is bitline parallel; and if it uses both wordlines and bitlines in parallel, it is mixed parallel.

BDD-based Parallel Computation requires that each node be implemented as a 2x1 multiplexer.

In AIG-based Parallel Computation, all progeny of both nodes must be computed. There must be no data dependencies between the nodes, and they must share a wordline operand with host devices on the same wordline, with the content not being used for any other calculations.

In M-AIG-based Parallel Computation, each node represents an m-Input And Gate or PI; each input edge can be linked to the constant 1 or to a child node, and if the input edge is attached to a node, it can be complemented to indicate inversion.

The suggested solution greatly reduces the number of necessary devices and procedures. It reduces the number of procedures by 66 percent on average. BDD and AIG thrive in both area and operation. For lesser values of m, m-AIG has beaten AIG.

**Power Aware Computing**

Processor designs are influenced by power, energy, and temperature. CPU clock frequency stagnation and reliance on parallelism will increase future energy efficiency. In addition to hardware, software design may have a significant impact. A must-have feature is the ability to measure power and energy consumption. The PAPI library, which offers a general and portable access to hardware counters coupled to the CPU and other components, was used in this investigation. The Xeon Phi Knights Landing (KNL) processor architecture was used in the experiments. Kernels of Dense Linear Algebra (DLA) were employed in the study (BLAS kernels).

Kernels prevalent in high-performance computing applications were chosen to analyse and assess the effect of application of power consumption and energy demands. Level 1 addresses scalar and vector operations, level 2 addresses matrix-vector operations, and level 3 addresses matrix-matrix operations. Compute Intensive Level 3 is in the Compute Intensive Class, whereas Memory Bond Levels 1 and 2 are in the Memory Bond Class.

The memory bound class dgemv and the compute costly procedure dgemm were investigated and analysed.

PAPI is a library that collects performance counter data from diverse hardware and software components in a uniform manner. PAPI is made up of a variety of components that enable for the monitoring of power consumption and utilisation through various interfaces.

According to dgemm Kernel behaviour, the FLAT mode uses the MCDRAM as physical addressable memory space rather than a cache.

With the exception of DDR4, which performs four times slower than MCDRAM, the performance of the dgemv Kernel decreases between the two storages, and the results are the same as in Hybrid Mode.

This study found that using high bandwidth MCDRAM on KNL is crucial for high efficiency and low power consumption, and that Hybrid mode is the best option if the application requires a lot of computing.

**Temperature-Aware Computer Systems Opportunities and Challenges**

It hasn't been enough to design with power in mind to stop the flow of problems like heat density. The rate of localised heating is much faster than the chip-wide heating rate. Even in the worst-case situation, most high-power applications are still 20% or more below the worst-case scenario.

The requirement for Architectural-Level Thermal Management stems from the fact that each computing system's architectural domain is unique, and workload development is required to control instruction level parallelism. The design manual for the computer system includes hot spots and temperature gradients. In this situation, the role of system design and operating system is important.

Thermal modelling at the design stage is necessary to prevent thermally induced temporal and spatial nonuniformities in computing.

The compact model of a parametric microarchitecture in a computing system must track temperatures at the microarchitectural unit level, be modelled so that a new compact model for various microarchitectures can be created, solve the RC circuit's differential equations quickly, and be boundary and initial-condition independent.

Temperature-tracking Dynamic Frequency Scaling plots temperature versus average power density for gcc with a power averaging period of 0.033 seconds.

Because carrier mobility in CMOS is temperature sensitive, frequency in Temperature-tracking Dynamic Frequency Scaling is similarly linearly dependent on the operating temperature. When an application reaches its temperature limit, it may simply adjust the frequency to compensate for the higher temperature.

Dynamic Voltage Scaling is a method of thermal control. Because circuits switch more slowly as the operating voltage approaches the threshold voltage, lowering the processor voltage must be accompanied by a decrease in frequency.

Migrating Computation is the best DTM technique at 0.8 K/W, according to the findings, because the floorplan alone is sufficient to lower the operating temperature of the primary integer register file, MC can use ILP to hide the extra latency of the spare register file, and complete elimination of activity in the primary register file allows it to cool quickly, reducing the use of the slower secondary register file.