Decoupled Execution of Workload on Co-Processors

TECHNOLOGY NUMBER: 2020-478

OVERVIEW

A method to distribute computing workloads on multiple co-processors

- Allows independent control units to independently make decisions
- Can be toggled between automatic and manual control

BACKGROUND

A processing workload consists of reading data, computing, and then writing data. Recent trends in data science have placed importance on big data analytics, requiring high performance computing (HPC) strategies that can streamline data flow and task management. Common computer solutions for streamlining data flow and task management include parallel computing, restructuring software, developing more powerful processors, or combining different methods to make a highly efficient computer architecture. One method by which to increase efficiency processing involves offloading the tasks to computing processing units, or tiles, that can control the data flow and execution flow without interacting with the host core. The tiles utilize their own execution controllers and perform in burst mode to increase the speed of data movement without burdening the host core. The benefits of this approach suggest a need to propagate methods by which to improve its efficiency and implementation.

INNOVATION

Researchers have devised a method to distribute computing workloads on multiple co-processors with separate control units that can independently process and make decisions. This layout improves performance by efficiently and continually offloading the workload to each of the processors. The controllers can manage data flow and execution flow without any back-and-forth communication with the host core, speeding up processes and allowing multiple tasks to be processed simultaneously. Depending upon user preference, these autonomous data handling responsibilities can be toggled to manual control. This technology therefore has the

Technology ID
2020-478

Category
Hardware
Engineering & Physical Sciences
Semiconductors, MEMS, and
Electronics

Inventor
Junkang Zhu
Zhengya Zhang

Further information
Joohee Kim
jooheek@umich.edu

View online

potential to speed up computing while maintaining performance and low latency. The structure will benefit industries that rely on or are branching into big data analytics or machine learning. For instance, the concurrent use of multiple processors should aid efforts in defense, healthcare, and autonomous vehicle operation.

Explore other available products test at University of Michigan