



## DISSERTATION DEFENSE



# SUBHANKAR PAL

## Towards Closing the Programmability-Efficiency Gap using Software-Defined Hardware

Wednesday, July 21, 2021, 11:30 am  
3725 Beyster and  
[Virtual \(Passcode 523876\)](#)

**ABSTRACT:** The past decade has seen the breakdown of two important trends in the computing industry: Moore's law, an observation that the number of transistors in a chip roughly doubles every eighteen months, and Dennard scaling, that enabled the use of these transistors within a constant power budget. This has caused a surge in domain-specific accelerators that deliver superior energy-efficiency compared to CPUs. However, the fast pace of algorithmic innovation and non-recurring engineering costs have deterred the widespread use of such accelerators, since they are catered to single applications, thus creating a programmability-efficiency gap.

This dissertation proposes to close this gap with a reconfigurable system that morphs parts of the hardware on-the-fly to tailor to the requirements of each phase of the application. This system is designed to deliver near-accelerator-level efficiency across a broad set of applications, while retaining CPU-like programmability.

The dissertation first presents a solution that uses hardware-software co-design and memory reconfiguration to accelerate sparse matrix multiplication, which forms the basis of many applications in graph analytics and scientific computing. A prototyped 40nm chip demonstrates energy-efficiency and performance-per-die-area improvements of 12.6x and 17.1x over a high-end CPU.

The next piece of the dissertation enhances the hardware architecture with reconfigurability of the dataflow and resource sharing modes for general-purpose acceleration. Moreover, this architecture uses commercial cores and a software stack for CPU-level programmability. The system is evaluated on a diverse set of compute-bound and memory-bound applications in graph analytics, machine learning, image processing, etc. and shows average performance and energy-efficiency gains of 5.0x and 18.4x over the CPU.

The final part of the dissertation proposes a runtime control framework that monitors hardware counters to predict the next best configuration upon detecting a change in phase. The dynamically reconfigurable system achieves the performance of the best-average non-reconfiguring system with 23% lower energy across real-world datasets.

**CHAIR:** Prof. Ron Dreslinski