

High-Performance Host System for Heterogenous Mesh-of-Tiles Architectures

Recent advances in open and modular processor architectures are enabling a new generation of heterogeneous many-core systems that tightly integrate general-purpose cores with domain-specific hardware accelerators. While much attention has been devoted to the accelerator fabric itself, the overall efficiency and programmability of these systems is increasingly bottlenecked by the host: the general-purpose subsystem responsible for orchestrating computation, managing data movement, running the operating system and software stack, and dispatching workloads to the accelerator tiles. As mesh-of-tiles accelerators such as MAGIA scale to hundreds of tightly coupled tiles, the host must provide enough single-thread performance, memory-level parallelism, and I/O bandwidth to keep the fabric fed, without compromising the energy efficiency that motivates these architectures in the first place. Open and high-performance RISC-V cores - such as the application-class PULP-C910 and the out-of-order XiangShan core - now offer a credible, fully open foundation on which to build such a host, enabling co-design of the control plane and the acceleration plane within a single, coherent, open hardware ecosystem.

This Incarico di Ricerca position, which is aligned with the activity of the **ARCHYTAS** project, focuses on the design and integration of a high-performance, application-class host system for accelerators based on the 'mesh-of-tiles' paradigm, with MAGIA as the reference accelerator platform. The incaricato di ricerca will focus on: 1) evaluation, integration and co-design of open high-performance RISC-V cores (e.g. PULP-C910, XiangShan) as the host of a mesh-of-tiles accelerator; 2) design of the coherent interconnect, memory hierarchy and host-to-fabric interfaces required to sustain accelerator throughput; 3) bring-up of the operating system, software stack and runtime for workload dispatch, data movement and synchronization between host and accelerator tiles.

High-Performance Host System for Heterogenous Mesh-of-Tiles Architectures

I recenti progressi nelle architetture di processori aperti e modulari stanno abilitando una nuova generazione di sistemi many-core eterogenei che integrano strettamente core general-purpose con acceleratori hardware domain-specific. Sebbene gran parte dell'attenzione sia stata dedicata al tessuto di accelerazione in sé, l'efficienza complessiva e la programmabilità di questi sistemi sono sempre più limitate dall'host: il sottosistema general-purpose responsabile dell'orchestrazione del calcolo, della gestione del movimento dei dati, dell'esecuzione del sistema operativo e dello stack software, e del dispatch dei carichi di lavoro verso le tile dell'acceleratore. Man mano che gli acceleratori a mesh-of-tiles come MAGIA scalano fino a centinaia di tile strettamente accoppiate, l'host deve fornire prestazioni single-thread, parallelismo a livello di memoria e larghezza di banda di I/O sufficienti a mantenere il tessuto alimentato, senza compromettere l'efficienza energetica che motiva queste architetture. I core RISC-V aperti e ad alte prestazioni — come il core application-class PULP-C910 e il core out-of-order XiangShan — offrono oggi una base credibile e completamente aperta su cui costruire tale host, abilitando il co-design del piano di controllo e del piano di accelerazione all'interno di un unico ecosistema hardware aperto e coerente.

Questa posizione di Incarico di Ricerca, allineata con le attività del progetto **ARCHYTAS**, si focalizza sulla progettazione e integrazione di un sistema host ad alte prestazioni, di classe applicativa, per acceleratori basati sul paradigma "mesh-of-tiles", con MAGIA come piattaforma di acceleratore di riferimento. L'incaricato di ricerca si concentrerà su: 1) valutazione, integrazione e co-design di core RISC-V aperti ad alte prestazioni (ad es. PULP-C910, XiangShan) come host di un acceleratore mesh-of-tiles; 2) progettazione dell'interconnessione coerente, della gerarchia di memoria e delle interfacce host-tessuto necessarie a sostenere il throughput dell'acceleratore; 3) bring-up del sistema operativo, dello stack software e del runtime per il dispatch dei carichi di lavoro, il movimento dei dati e la sincronizzazione tra host e tile dell'acceleratore.