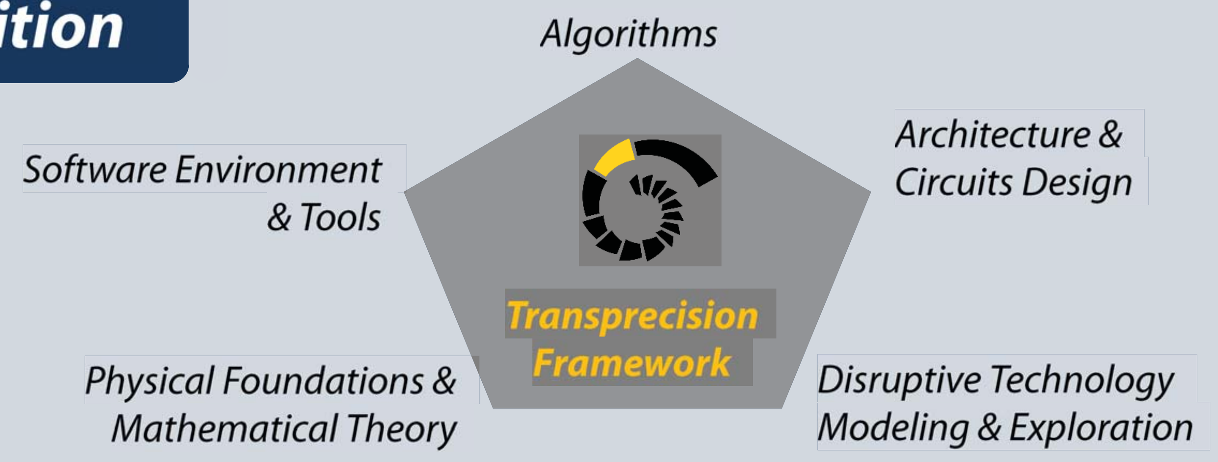
OPRECOMP

Open Transprecision Computing

European Commission

This project is funded by the European Union's H2020 - FET Proactive research and innovation program under grant agreement #732631.

Ambition



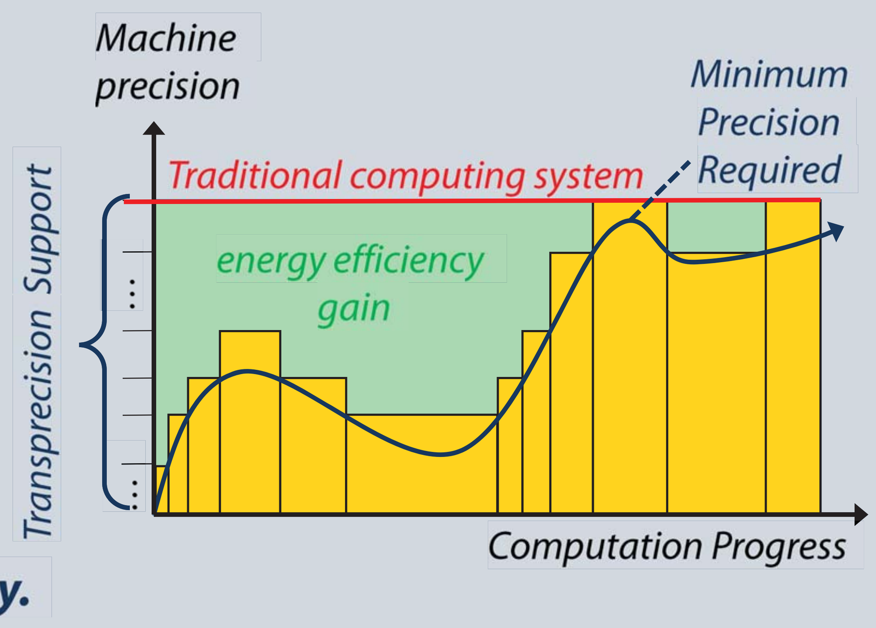
OPRECOMP *aims to build an innovative, reliable foundation for computing based on transprecision analytics demolishing the ultra-conservative "precise" computing abstraction and replacing it with a more flexible and efficient one, namely* ***transprecision computing****. The "disruptive" mission is to demonstrate this idea in the domains of* ***Big Data Analytics****,* ***Deep Learning*** *and* ***High Performance Computing****.*

Physical Foundations & Transprecision

Computations are performed by ***physical systems****, and as such the changes in the physical system determine the* ***energetic cost*** *associated with the acto of computing.*

OPRECOMP *aims at:*

- *Identifying fundamental bounds (laws of physics) and technological constraints*
- *Determine trade-offs between accuracy and energetic costs*
- *Providing a* ***holistic transprecision framework*** *enabling to deliver* ***accurate*** *results with* ***10x less energy.***

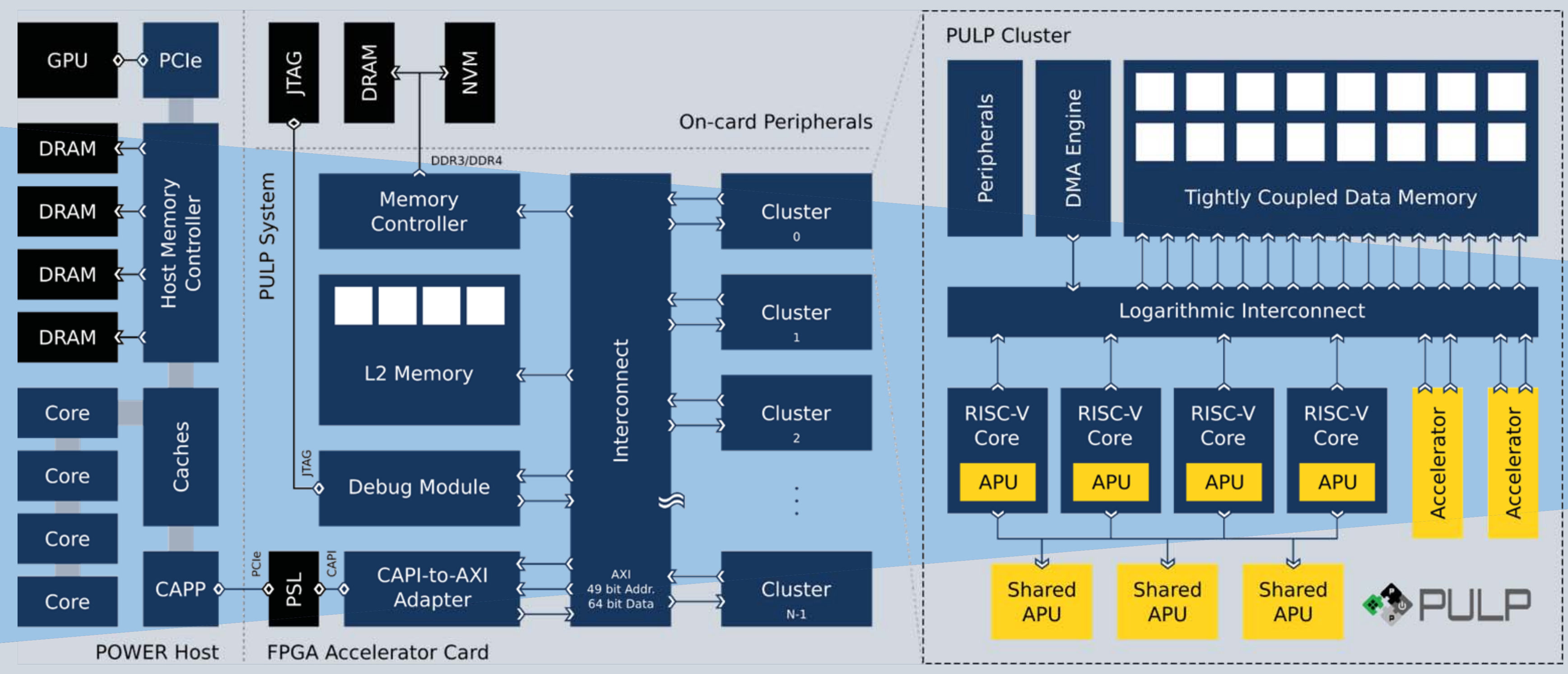


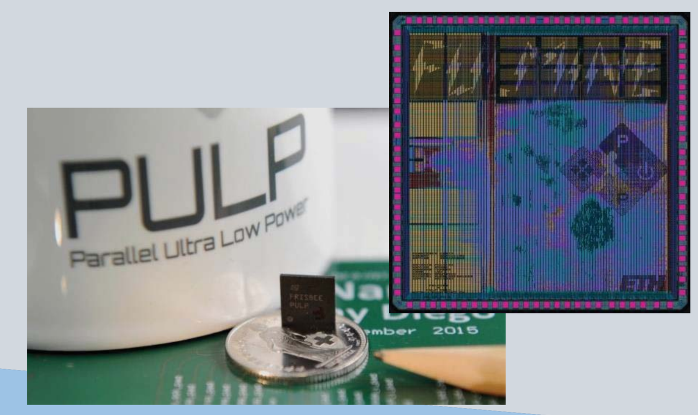
Platforms

kW

kW Anchor:
IBM POWER8/9
CAPI+FPGA / GPU

FPGA-based Parallel Ultra Low Power (PULP) platform emulator with transprecision support





mW

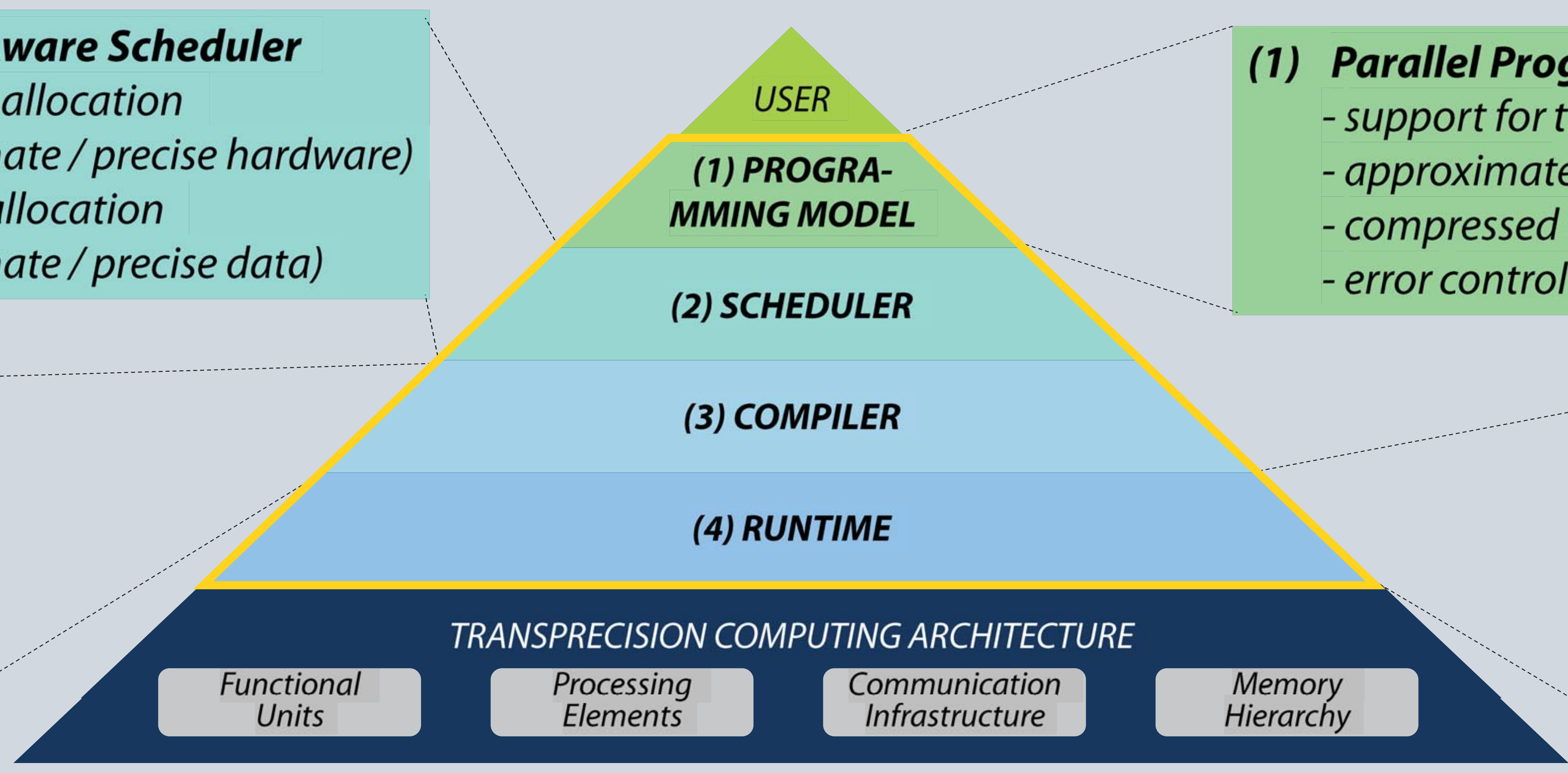
mW Anchor:
ASIC implementations of PULP with transprecision support

- *Enable silicon measurements*
- *Derive model for large-scale implementation (kW anchor)*

Software Stack

(2) Accuracy-Aware Scheduler
- *workload allocation (approximate / precise hardware)*
- *memory allocation (approximate / precise data)*

(1) Parallel Programming Model (e.g. OpenMP)
- *support for transprecision arithmetic*
- *approximate communication*
- *compressed data transfer*
- *error control for heterogeneous architectures*



(3) Dynamic Code Generation
- *variable vector length*
- *variable datatypes*
- *exploit inherent algorithmic error resilience*
- *accuracy / energy tradeoff*

(4) Energy-Efficient Runtime
- *monitor/control variable precisions at runtime*
- *autotuning and scheduler task management*

Micro-Benchmarks

3 Relevant Application Domains

Deep Learning	***Big-Data and Data Analytics***	***HPC and Scientific Computing***
CNN PCA SGD LSTM	*PageRank K-Means SVM KNN*	*GLQ FFT Stencil SparseSolve*

Quality Metrics / Baseline Implementations / Transprecision Implementations / mW and kW Implementations

These micro-benchmarks are composed of several distinct computational kernels. For each kernel, different algorithms using transprecision will be implemented covering single-threaded and multi-threaded code for both mW and kW platforms using OpenMP and CUDA for parallelization. Quality metrics are defined for each problem to assess the correctness and to measure the balance between quality loss against energy gains.

Partners

Technische Universität Kaiserslautern
Universitat Jaume·I
GreenWaves Technologies
ETH zürich
cea
CINECA
Queen's University Belfast
Alma Mater Studiorum A.D. 1088
Studium Generale Civitatis Perusii A.D. MCCCVIII
IBM

www.oprecomp.org
@oprecompProject