European Processor Initiative

EUROPEAN PROCESSOR INITIATIVE

FRAMEWORK PARTNERSHIP AGREEMENT IN EUROPEAN LOW-POWER MICROPROCESSOR TECHNOLOGIES

Description

- Domain specific accelerator for scientific computing, designed for the resolution of large ill-conditioned systems of equations
- Designed for the accurate computation (up to 256 bit fractional parts) of large systems
- Targets 10x to 100x acceleration of variable precision computation (compared to software solutions)
- A single VRP processor features:
    - Fully functional branch control for efficient convergence
    - 32 internal registers for up to 256 bit arithmetic operations
    - Transparent cache access for arbitrary size of variables

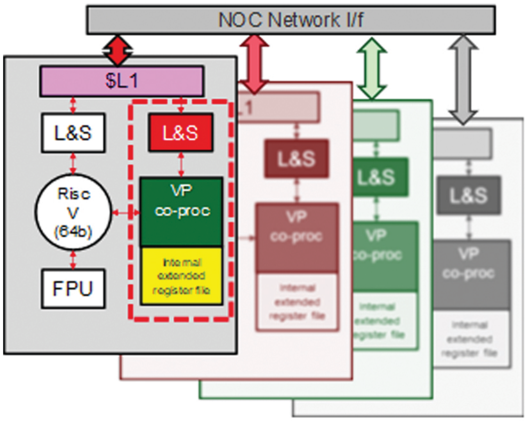
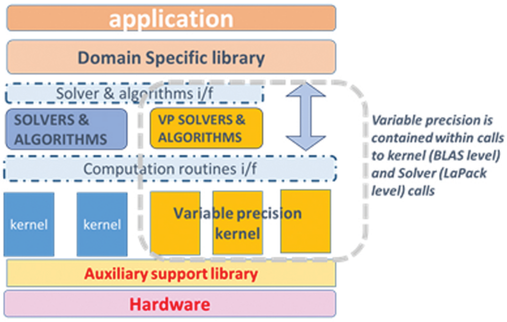

NOC Network I/f
$L1
L&S
L&S
Risc V (64b)
VP co-proc
Internal extended register file
FPU
L&S
VP co-proc
Internal extended register file
L&S
VP co-proc
Internal extended register file
L&S
VP co-proc
Internal extended register file

Figure 1 Variable precision tile structure



Variable precision is contained within calls to kernel (BLAS level) and Solver (LaPack level) calls

Figure 2 Layered programming model for variable precision

Domain Specific Accelerator in EPI – VRP (VaRiable Precision Processor)

The VaRiable-Precision Unit (VRP) enables efficient computation in scientific domains with extensive use of iterative linear algebra kernels, such as physics and chemistry. Augmenting accuracy inside the kernel reduces rounding errors and therefore improves computation's stability. Usual solutions for this problem have a very high impact in memory and computation time (e.g. use double precision in the intermediate calculations).

The hardware support of variable precision, byte-aligned data format for intermediate data optimizes both memory usage and computing efficiency. When the standard precision unit cannot reach the expected accuracy with standard precision (aka double), the variable precision unit takes the relay and continues with gradually augmenting precision until the tolerance error constraint is met. The offloading from the host processor (GPP) to the VRP unit is ensured with zero-copy handover thanks to IO-coherency between EPAC and GPP.

The VRP is embedded as a functional unit in a 64-bits RISC-V processor pipeline. The unit extends the standard RISC-V Instruction with hardwired arithmetic basic operations in variable precision for scalars: add, subtract, multiply and type conversions. It implements other additional specific instructions for comparisons, type conversion and transfers to cache.

The unit features a dedicated register file for storing up to 32 scalars with up to 256 bits of mantissa precision. Its architecture is pipelined for performance, and it has an internal parallelism of 64-bits. Thus, internal operations with higher precisions multiple of 64 bits are executed by iterating on the existing hardware.

The VRP's programming model is meant for a smooth integration with legacy scientific libraries such as BLAS, MAGMA and linear solver libraries. The integration in the host memory hierarchy is transparent for avoiding the need of data copy, and the accelerator offers a standard support for C programs. The libraries are organized in order to expose the variable precision kernels as compatible replacements of their usual counterparts in the BLAS and solver libraries. The complexity of arithmetic operations is confined as much as possible within the lower level library routines (BLAS), as represented below. Consistently, the explicit control of precision is exclusively handled at solver level.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826647