montblanc-project.eu | @MontBlanc_EU

Using SVE to Accelerate HPC Applicaotins
A Mont-Blanc 2020 Perspective

Adrià Armejach, Ruben Langarita and Miquel Moreto

Barcelona Supercomputing Center (BSC)

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement n° 671697 & 779877

Mont-Blanc roadmap

Today

Vision: leverage the fast growing market of mobile technology for scientific computation

2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020

EPI European Processor Initiative

- **Design SoC**
- **Develop IPs**

Prepare an industrial solution
Test market acceptance

Mont-Blanc 2020

Extend the concept and explore new possibilities

Mont-Blanc 3

Proof of concept : HPC computing based on mobile embedded technology

Mont-Blanc 2

Mont-Blanc

MONT-BLANC

MB2020 Partners

Short Name	Country
Bull	FR
SemiDynamics	ES
CEA	FR
BSC	ES
ARM	UK
Juelich	DE
Kalray	FR

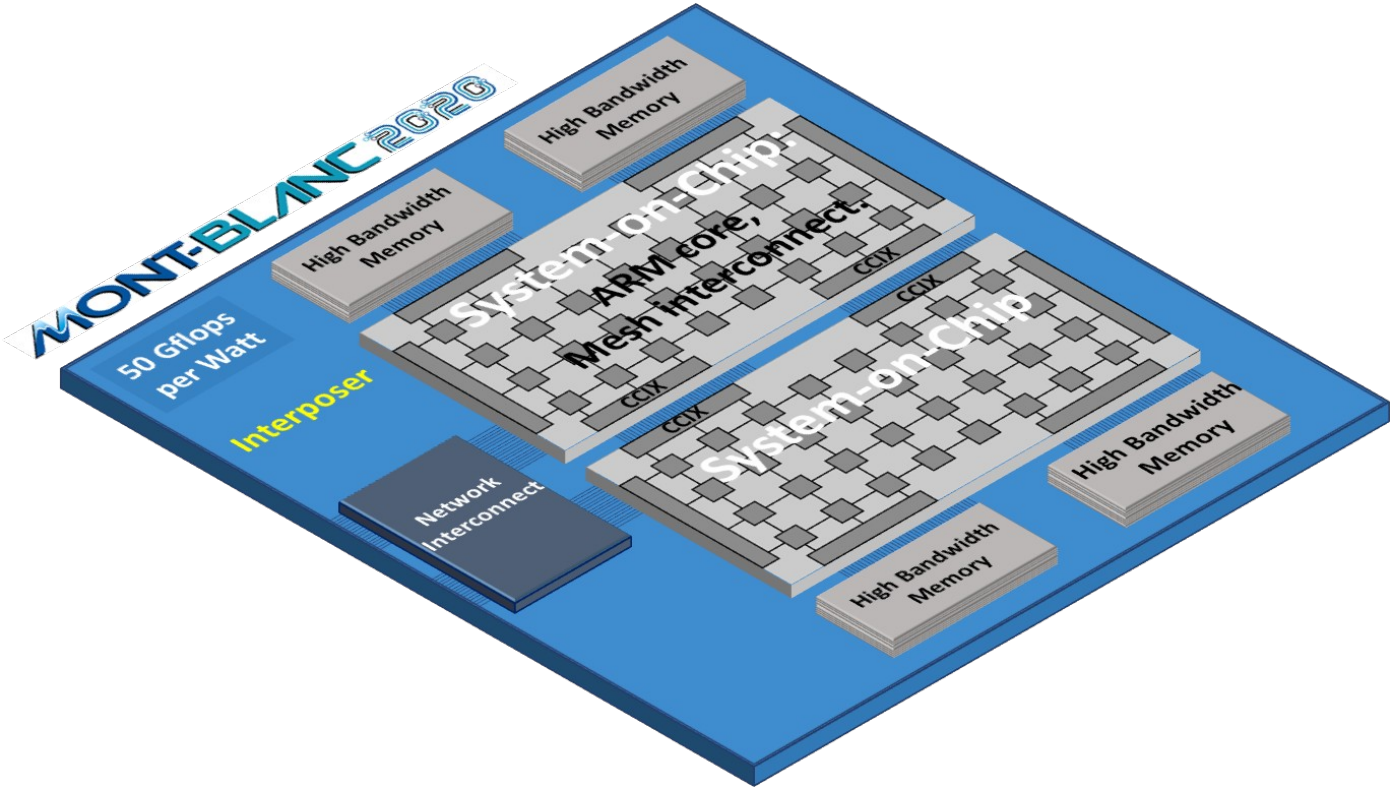
MB2020 Objectives

- **Define a low-power System-on-Chip (SoC) architecture targeting Exascale using Arm ISA**
  - Leverage Arm's strong technological relevance covering embedded to high end solutions
  - dynamic ecosystem, needed to deliver the system software and applications mandatory for successful market acceptance.

- **To design, implement and leverage new technologies to improve the performance, power efficiency, reliability and security of the processor:**
  - Exploit SVE compute units
  - specific Network-on-Chip (NoC) with high bandwidth and low power characteristics
  - specific SoC Power Management controller
  - innovative packaging technologies
  - state of the art silicon process

- **To improve on the economic sustainability of processor development through a modular design that allows to retarget the SoC for different markets**

Mont-Blanc 2020 SoC



Requirements for SVE: Design-space exploration

- **Explore hardware trade-offs**
  - Different vector lengths (128 to 2048 bits)
    - Off-chip memory bandwidth demands
      - Pin-based with DDR interfaces
      - Silicon interposer with HBM interfaces
  - Sizing of key core structures – e.g., load/store queue
  - Number of vector units – i.e., 2 x 256bit vs. 1 x 512bit
  - Recommended memory bandwidth per core

- **Characterize performance of applications**
  - Demanded bytes per floating point operation – operational intensity (OI)
    - End goal: Understand how applications perform under different vector lengths and memory bandwidth constrains

Requirements for SVE – Simulation parameters

- **Resemble architecture envisioned in MB2020**

- **Gem5 simulation parameters**

	Description
Processor size	8 cores - 2 clusters of 4 cores each
Cores	3-wide issue/retire, 64-entry instruction queue, 192-entry ROB, 48LDQ + 48STQ, 2 Vector processing units (VPU), 2GHz
L1I	64KB, 4-way, 2 cycle, 8MSHR
L1D	64KB, 4-way, 2 cycle, 24MSHR
L2 (private)	256KB, 2-way, 7 cycle, 24MSHR, stride prefetcher
Last-level Cache	16MB in 8 banks, 16-way, 20 cycles, 64MSHR
NoC	Coherent crossbar, 128-bit wide, 2 cycles
Main memory	4 DDR4-2400, 2 ranks/channel, 16 banks/rank, 8KB row-buffer 76.8 GB/s peak bandwidth
	1 HBM stack, 8 channels/stack 128-bit wide, 8 banks/channel 128 GB/s peak bandwidth

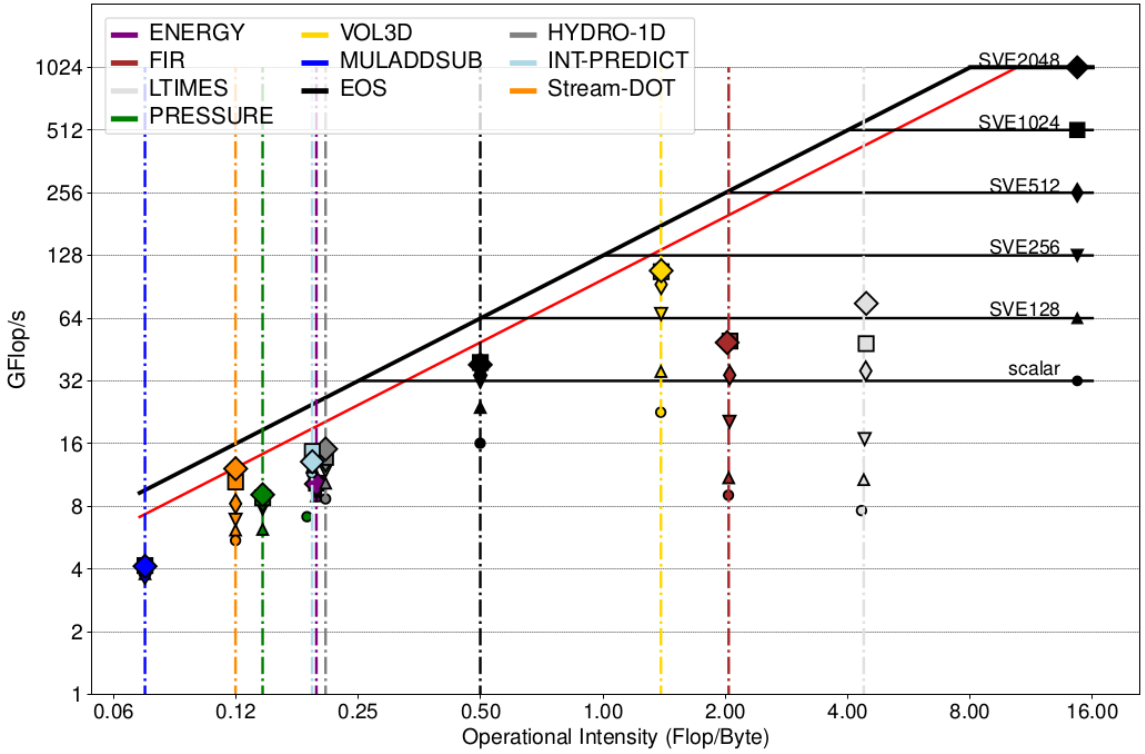
Requirements for SVE – Benchmarks

- **Simulate a representative subset of the RAJAPerf loops**
  - Good auto-vectorization coverage

Benchmark	Description
MULADDSUB	Simple synthetic loop for quick testing
HYDRO_1D	Main computational loop of the HYDRO application
EOS	Calculates the equation of state
INT_PREDICTOR	Integral predictor
ENERGY	Kernel to calculate energy states, extracted from an LLNL application
PRESSURE	Kernel to calculate pressure in a system, extracted from an LLNL application
FIR	Finite impulse filter, extracted from an LLNL application
LTIMES	Kernel extracted from an LLNL application
Stream-DOT	Stream benchmark, dot-product version
VOL3D	Kernel to calculate a 3D volume, extracted from an LLNL application

Requirements for SVE – Vector length results (HBM)

- **Peak performance (compute ceilings)**
  - Peak SVE128 = 2GHz (cycles/s) × 8 (#cores) × 2 (flop/instr) × 2 (instr/cycle) = **64GFlops/s**
- **Memory ceilings:**
  - Peak 128 GB/s
  - 16GB/s per core
- **Operational intensity**
  - Most benchmarks <1
  - Common values
- **HBM improves SVE scalability**
  - Memory bound benchmarks benefit greatly
  - EOS shows significantly better performance with SVE256
- **SVE 256 likely to be the best candidate**
  - Need at least OI == 2 to reach peak SVE 512 bit performance



Requirements for SVE – Summary of Requirements

Summary

- Use of HBM technology mandatory to scale to more than 8 cores per socket
- Recommend the use of SVE 256 or SVE 512 bits
  - SVE 512bits starts to be beneficial with OI > 1.5
  - SVE 512bits offers marginal benefits for OI < 1
- Contemporary aggressive OoO HW structures sufficient for SVE 256 & 512 bits
  - 48 LQ, 48SQ, 92 IQ, 192 ROB
  - 33% larger HW structures offer modest improvements – 17% max. 5% avg.
- We recommend a minimum of 16GB/s of off-chip memory bw per core

Next Steps (to be presented in the next JLESC…)

Next steps

- Look at a wider range of benchmarks
  - HACCKernels, HPCG, MiniAMR, SWFFT, XSBench, …
- Validate simulations results vs Fujitsu's A64FX processor
  - Issues with cluster installation at BSC (affected by covid19)
  - Reproduce RajaPerf performance figures with 256- and 512-bit SVE
  - Extend to other benchmarks of interest
- Evaluate large-scale applications on Fujitsu's A64FX processor
  - HPC
  - Bioinformatics applications: genomics mappers and variant calling
- **Interested in collaborating with other JLESC partners in that direction!**

montblanc-project.eu | @MontBlanc_EU

Using SVE to Accelerate HPC Applicaotins
A Mont-Blanc 2020 Perspective

Adrià Armejach, Ruben Langarita and Miquel Moreto

Barcelona Supercomputing Center (BSC)

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement n° 671697 & 779877