



SIPEARL

OneMath Arm Performance Libraries backends

SiPearl in a nutshell

Building the European high-performance low-power microprocessor



Incorporated

In June 2019



Funded

By the European Union



Arm architecture

Energy-efficiency quick time to market, proven ecosystem



Financing

Series-A to date: €113m
(€105m equity + €8m bank loans)

arm

European Innovation Council Fund

Banque européenne d'investissement

EVIDEN



Key partnerships

Joint-offering with

AMD

intel

nvidia

EVIDEN

Hewlett Packard Enterprise



Identified customers

Server manufacturers based on user specifications: First, EuroHPC ecosystem before going global.

+200

Employees

from

Atos

MEDIATEK

ST

MARVELL

NXP

Infineon

intel

NOKIA



7 locations

Maisons-Laffitte (HQ), Barcelona, Bologna, Duisburg, Grenoble, Massy, Sophia Antipolis

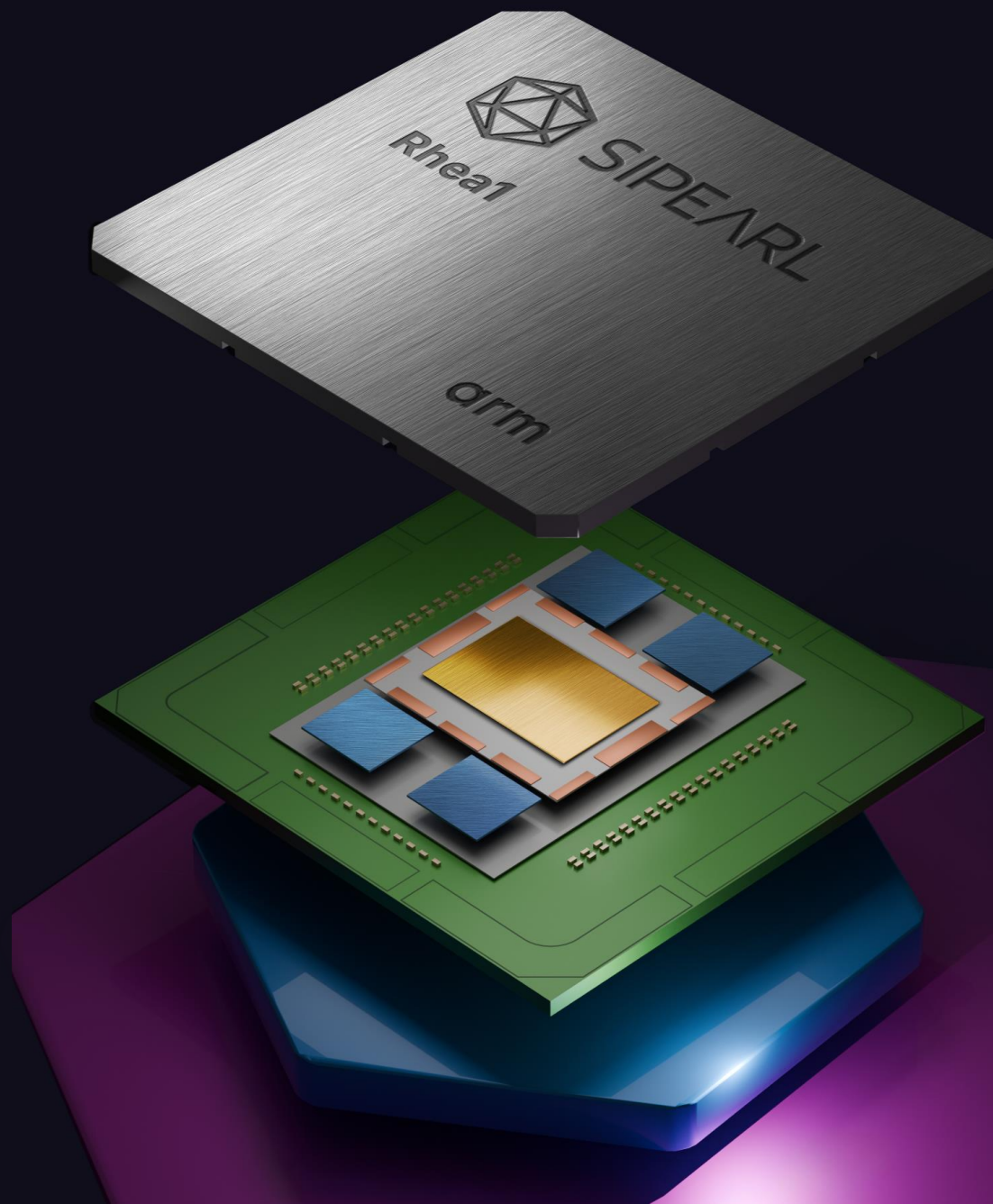
RHEA1

HPC and AI microprocessor

80 Arm® Neoverse V1 cores
with 2 × 256 SVE each

4 x HBM2e

4 x DDR5 interfaces



Arm Performance Libraries (ArmPL)

<https://developer.arm.com/Tools and Software/Arm Performance Libraries>

Set of mathematical libraries optimized for 64-bit Arm based processors, with vectorization (SVE/Neon) support and OpenMP.

Domains:

- BLAS – CBLAS and XBLAS interfaces, with most of MKL extensions, batched and interleaved batched calls.
- LAPACK – LAPACKE interface
- FFT – FFTW interface
- Sparse Linear Algebra (SpMV, SpMM, SpSV)
- RNG – OpenRNG
- Optimized/Vectorized math/string libraries

Standalone versions available for GCC/NVHPC/Arm compilers, and MacOS/MSVC compilers

OneMath ArmPl backend

Rationale:

- Provide good support for oneAPI in SiPearl processors with CPU/GPU targets
- GPU targets tested by other parties (Durham University, as said in last call minutes)
- No existing CPU target for aarch64 CPUs
 - Intel sycl compiler (Open DPC++) was working, with host (now with native_cpu) or pocl OpenCL backend
 - AdaptiveCpp working as well (omp backend)
- Started as an internship couple years ago

BLAS backend integration (PR [#629](#))

Porting:

- Added aarch64 CPU target support (For ArmPI and Netlib)
- Porting of most calls straightforward from Netlib backend
- Other extensions from MKL added later
- Batched calls added
 - Loops when no direct support
- Tests made with interleaved batched calls
 - Not kept, as changing the data layout everytime was too costly
- Some bugs found in the process (fixed)
 - Iterations with Arm to improve ArmPI (zaxpby, gemmt, omatcopy...) due to issues found in oneMath tests
 - Issues in OneMath were reported and fixed: csrot/zdrot tests ([#195](#)), BLAS tests not testing the correct backend due to cblas symbol collision ([#204](#))
- Supports most of oneMath API
 - omatadd2/copy2, batch or int8/bfloat16 gemm flavours unsupported in ArmPI as of 24.10
- Around 2PM of work

Lapack backend

- No Netlib backend available
- ArmPI provides LAPACKE interface
- Split calls using LAPACKE_*_work functions to query scratchpad size
- Small issues with scratchpad sizes:
 - OneMath interface only allows one work array/scratchpad size per kernel
 - Some LAPACKE calls (hegv, heevd) have up to three separate work arrays/sizes
 - For now default to LAPACKE_ direct call (which allocates arrays each time) for these.
- No support for batched calls for now:
 - Not yet supported in ArmPI, would be loops over single calls
- 100% tests passed, 0 tests failed out of 796
 - That's not right (skipped batch)
- PR [#633](#) submitted, and merged yesterday

RNG backend

- ArmPI released OpenRNG library (<https://gitlab.arm.com/libraries/openrng>)
- OpenRNG:
 - Drop in replacement for Intel VSL library
 - Included in ArmPL
 - Several generators and distributions implemented (mcg31 mcg59 mrg32k3a mt19937 nondeterm philox4×32×10 r250 sfmt19937 sobol)
- Implementation mostly straightforward in OneMath, just small differences from OneMKL
- Mrg32k3a and philox4x32x10 only for now, as for other backends
- Test status : 97% tests passed, 4 tests failed out of 154
 - OpenRNG issue with int32 uniform generator detected during OneMath testing and reported to ARM
- PR [#634](#) submitted

Found issues

- **BLAS tests with cblas symbol conflicts when linking with reference BLAS**
 - Issue fixed by oneMath team [#204](#)
- **BLAS tests not working properly (csrot/zdrot)**
 - Issue fixed by oneMath team [#195](#)
- **LAPACK tests with LAPACKE symbol conflicts when linking with reference Lapack**
 - PR ready, pending internal review to upstream
- **LAPACK tests not reported skipped tests (success instead)**
 - PR ready, pending internal review to upstream

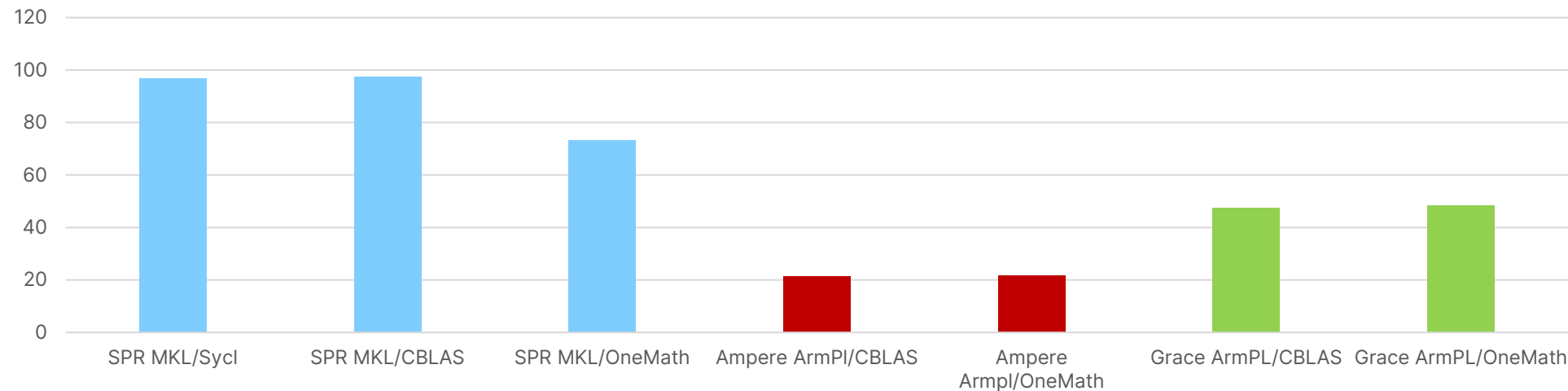
Usecases

- Drop in replacement for some OneMKL codes, assessing performance portability
- BigDFT (ab initio DFT code using wavelets) :
 - replace BLAS/MKL/GPU code path by single OneMath/Sycl codepath and ensure portability
 - need BLAS, DFT backend
- Experiments with several subdevices (one Host, one Accelerator, CPU or GPU)
- Evaluation of various strategies (interleaved batch calls)

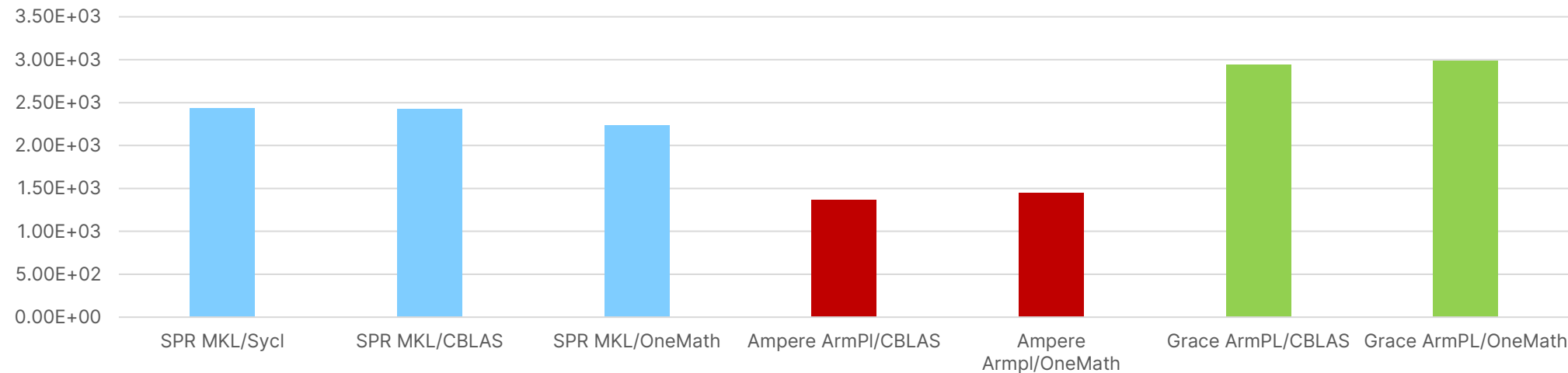
Performance – Dgemm

SPR: Intel Sapphire Rapids Xeon(R) CPU Max 9460, 40 core (no HT, no improvement with), HBM
Ampere: Ampere altra, 80 core ARM Neoverse N1 (NEON)
Grace: Nvidia Grace, 72 core ARM Neoverse V2 (SVE)

Single Core (Gflops)



Full socket (Gflops)



Conclusion

- BLAS/LAPACK/RNG domains available for aarch64 CPUs
- Performance close to native
- DFT in the works (need to map to FFTW)
- Sparse BLAS next ?

About... SiPearl

SiPearl is building the European high-performance low-power microprocessor dedicated to AI inference and supercomputing. This new generation of microprocessors will first target EuroHPC Joint Undertaking ecosystem, which is deploying world-class supercomputing infrastructures in Europe for solving major challenges in medical research, security, energy management and climate with a reduced environmental footprint.

SiPearl is working in close collaboration with its 30 partners from the European Processor Initiative (EPI) consortium - leading names from the scientific community, supercomputing centres and industry - which are its stakeholders, future clients and end-users.

SiPearl employs more than 200 people in France (Maisons-Laffitte, Grenoble, Massy, Sophia Antipolis), Germany (Duisburg), Italy (Bologna) and Spain (Barcelona).



Contact

Augustin Degomme
HPC benchmark engineer - Grenoble
augustin.degomme@sipearl.com

