Performance of Devito on HPC-Optimised ARM Processors

H. Senger [1,2] J. Freire [1] E. Gomi [2] F. Luporini [3] G. J. Gorman [3]

[1]Universidade Federal de São Carlos (UFSCar), Brazil [2]Universidade de São Paulo, Brazil [3]Imperial College London, UK

In this work we evaluate the performance of Devito, a domain specific language (DSL) for finite differences, on Arm ThunderX2 processors. Experiments with two common seismic computational kernels demonstrate that Arm processors can achieve competitive performance compared to Intel Xeon processors.

HPC-Optimised ARM Processors

HPC-Optimised Arm processors are emerging: Huawei (Kunpeng 920), Ampere (eMAG), Fujitsu (A64FX), and Marvell (ThunderX2). In recent studies, Arm based supercomputers have demonstrated capable of providing levels of performance competitive with state-of-the-art HPC-optimized processors (e.g. Intel Skylake and Broadwell) for a wide range of applications with very attractive performance per Dollar ratio.

The Isambard [a] system - the largest Arm based HPC production system in Europe:

- Cray XC50 (Scout) with Aries interconnect
- Each node has 2 Arm based processors - 32-core Marvell ThunderX2 CPUs, 256 GB of DDR4 DRAM
- 42 blades with 4 nodes
- The whole system: 10,752 Armv8 cores

[a]http://gw4.ac.uk/isambard/

Devito is a finite difference DSL for creating highly optimised finite-difference operators from concise symbolic expressions built upon SymPy. Creating operators symbolically allows application developers to:

- Build complex solvers from a few lines of code
- Utilise automated (symbolic, and loop) performance optimisation
- Adjust numerical discretization at run time
- To develop and integrate other mathematical operators that fall outside the stencil programming model
- Develop high-performance solvers in hours, not months

Currently, parallelism is supported by OpenMP and MPI which are integrated to the Devito compiler stack. Code generation for GPUs is under development.

Setup of Experiments

- Experiments on three processor types (see Table 1)
- Two benchmarks:
  - **acoustic** isotropic wave equation.
  - Tilted Transverse Isotropy (**TTI**) model: a high-order stencil representative of the state-of-the-art for seismic imaging.
- Equations are discretized with second order in time and varying space orders of 4, 8, 12 and 16
- Grid sizes: 512^3, 768^3, and 1024^3 points, spacing of 20m
- The time step is modelled for one second (i.e., 327 time steps for acoustic, and 415 time steps for TTI)
- Devito 3.4, Cray's CCE 8.7.9 (for Arm), and GCC 8.2
- Thread pinning (one thread per physical core)

Table 1: Processor types used (specifications per socket). Intel processors can slow down clock when all cores execute AVX512.

	ThunderX2	Intel 5120	Intel 6126
Cores/socket	32	14	12
Threads/core	1,2 or **4**	2	2
Clock frequency (GHz)	2.2	2.2/**1.6**	2.6/**2.3**
Vector unit width	128-bit	512-bit	512-bit
Max. # FP64 ops./cycle	8	16	32
Max. perf.(FP64) Gflop/s	563.2	358.4	883.2
Linpack (FP64) Gflop/s	410.5	345.4	695.0
L1 cache (core)	32 KB	32 KB	32 KB
L2 cache (core)	256 KB	1 MB	1 MB
L3 cache (shared)	32 MB	19 MB	19 MB
Memory	128 GB	192 GB	96 GB
Memory channels	8	6	6
Maximum bandwidth	160 GB/s	107.3 GB/s	119.2 GB/s
Measured bandwidth	119 GB/s	65 GB/s	71 GB/s

Experimental Results

Devito generates stencil code on-the-fly, allowing various types of performance optimisation across Arm and Intel platforms:

- Parallelism, and SIMD vectorization (OpenMP >= 4)
- Loop tiling (best block shape obtained via auto-tuning)
- Domain-specific symbolic optimisation:
  - Common sub-expression elimination and factorisation for Flop reduction (especially relevant for high-order stencils like TTI)
  - Polynomial approximations for trigonometry terms
  - Heuristic hoisting of time-invariant expressions

Table 2: The execution times, speedup (S), and the amount of GFLOPS (G) for the scaling experiment on a single socket, 512^3 grid, and SO=4.

ThunderX2

	acoustic			TTI		
T	time	S	G	time	S	G
1	288.1	1.0	6.0	1879.4	1.0	7.6
2	145.2	2.0	11.9	977.1	1.9	14.6
4	73.5	3.9	23.5	489.1	3.8	29.2
8	37.6	7.7	46.0	245.5	7.7	58.2
16	19.3	14.9	89.7	126.3	14.9	113.1
32	11.0	26.3	157.7	67.3	27.9	212.3

Xeon Gold 5120

	acoustic			TTI		
T	time	S	G	time	S	G
1	93.5	1.0	18.5	573.0	1.0	24.9
2	46.9	2.0	36.9	277.2	2.1	51.5
4	25.2	3.7	68.7	162.2	3.5	88.1
8	16.8	5.5	102.6	102.1	5.6	139.9
14	14.7	6.4	117.7	69.2	8.3	206.5

Xeon Gold 6128

	acoustic			TTI		
T	time	S	G	time	S	G
1	85.9	1.0	20.1	505.3	1.0	28.3
2	42.6	2.0	40.5	246.3	2.1	58.0
4	23.0	3.7	75.2	131.4	3.8	108.7
8	14.9	5.8	116.2	81.3	6.2	175.7
12	13.6	6.3	126.9	61.4	8.2	232.6

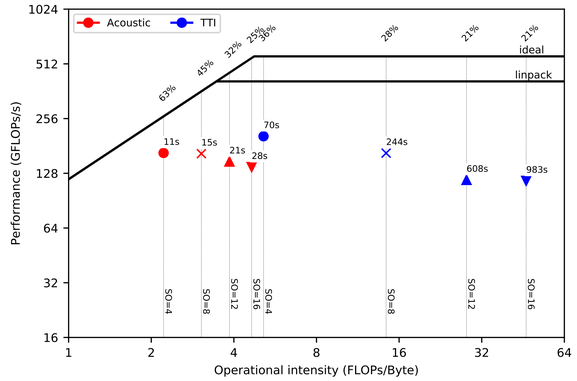

Figure 1: Roofline plot showing performance on Arm ThunderX2.

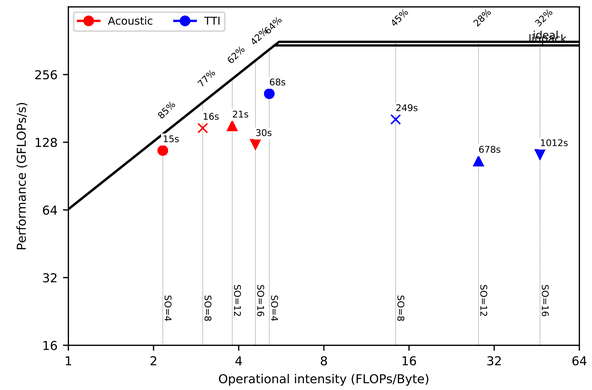

Figure 2: Roofline plot showing performance on Xeon 5120.

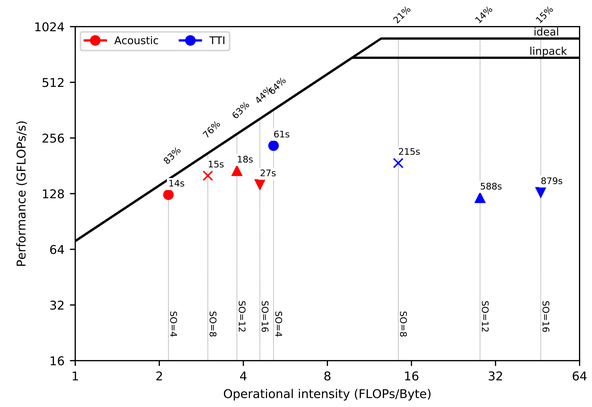

Figure 3: Roofline plot showing performance on Xeon 6126.

Main Findings

- With Devito, Arm based processors are capable of delivering performance similar to state-of-the-art Intel Xeon processors for the execution of seismic inverse problems
- Devito is shown to be capable of generating efficient high performance code for Arm processor
- All models compiled and ran successfully, and no architecture specific code tuning was necessary to obtain high performance other than sepecifying the compiler and its options

https://bitbucket.org/ws-5-stmi/performance-of-devito-on-arm

Imperial College London

POLI USP

ufscar

Research Centre for Gas Innovation