

Pushing the limits of Online Auto-tuning:  
Core Configurations

Fernando Endo

Damien Couroussé

Henri-Pierre Charles

March 31, 2016

Table 1: Main parameters of the simulated cores.

| Parameter                                  | Single-issue                                     | Dual-issue          | Triple-issue          |
|--|--|---------------------|-----------------------|
| Pipeline type                              | IO only  | IO or OOO           | IO or OOO             |
| Core clock                                 | 1.4 GHz  | 1.6 GHz             | 2.0 GHz               |
| DRAM                                       | 256 MB/933 MHz/81                                | 256 MB/933 MHz/81   | 256 MB/933 MHz/81     |
| L2   | 512 kB/8/3/8/16                                  | 1024 kB/8/5/8/16    | 2048 kB/16/8/11/16    |
| L1-I                                       | 32 kB/2/1/2                                      | 32 kB/2/1/2         | 32 kB/2/1/2           |
| L1-D                                       | 32 kB/4/1/4/4                                    | 32 kB/4/1/5/8       | 32 kB/2/1/6/16        |
| Stride prefetch.                           | 1/1/8  | 1/1/12              | 2/1/16                |
| Branch pred.                               | Global/local history entries (bits)<br>256 (2)/- | 4096 (2)/-          | 4096 (2)/1024 (3)     |
| Front-end/back-end width                   | 256/8  | 4096/16             | 4096/48               |
| INT/FP pipeline depth (+ extra OOO stages) | 1/1  | 2/4                 | 3/7                   |
| Physical INT/FP registers                  | 8/10   | 8/12 (+3)           | 9/18 (+6)             |
| ITLB/DTLB/IQ/LSQ/ROB entries               | -  | 82/256              | 90/256                |
| INT units                                  | 32/32/16/8 each/-                                | 64/64/32/12 each/40 | 128/128/48/16 each/60 |
| FP/SIMD                                    | ALU/MUL execution ports<br>1/1                   | 2/1                 | 2/1                   |
| Load/store                                 | ADD/MUL cycles<br>1/4                            | 1/4                 | 1/4                   |
|  | Execution ports<br>1                             | 1 or 2              | 1, 2 or 3             |
|  | VADD/VMUL/VMLA cycles<br>3/4/6                   | 4/5/8               | 10/12/20              |
|  | Execution ports<br>1 shared                      | 1 shared            | 1 for each            |
|  | Load/store cycles<br>1/1                         | 2/1                 | 3/2                   |

Over-dimensioned to compensate the lack of L2-TLB.  
For OOO only.

Table 2: Abbreviation of the simulated core designs and CPU areas

| Abbrev. | Width | Type | VPUs | Area (mm <sup>2</sup> ) |      |       |
|---------|-------|------|------|-------------------------|------|-------|
|         |       |      |      | Core                    | L2   | Total |
| SI-I1   | 1     | IO   | 1    | 0.45                    | 1.52 | 1.97  |
| TI-I1   | 3     | IO   | 1    | 1.81                    | 5.88 | 7.70  |
| TI-I2   | 3     | IO   | 2    | 2.89                    | 5.88 | 8.78  |
| DI-I1   | 2     | IO   | 1    | 1.00                    | 3.19 | 4.19  |
| TI-I3   | 3     | IO   | 3    | 3.98                    | 5.88 | 9.86  |
| DI-I2   | 2     | IO   | 2    | 1.48                    | 3.19 | 4.67  |
| TI-O1   | 3     | OOO  | 1    | 2.08                    | 5.88 | 7.97  |
| DI-O1   | 2     | OOO  | 1    | 1.15                    | 3.19 | 4.34  |
| TI-O2   | 3     | OOO  | 2    | 3.21                    | 5.88 | 9.10  |
| DI-O2   | 2     | OOO  | 2    | 1.67                    | 3.19 | 4.86  |
| TI-O3   | 3     | OOO  | 3    | 4.35                    | 5.88 | 10.2  |

Table 3: Execution time (seconds) of the benchmarks with the original and specialized reference kernels, and with the online auto-tuned and the best statically auto-tuned kernels, in the real platforms (all run-time overheads included).

| Benchmark     | Input  | Version | Cortex-A8 |            |       | Cortex-A9 |       |            |       |       |
|---------------|--------|---------|-----------|------------|-------|-----------|-------|------------|-------|-------|
|               |        |         | Ref.      | Spec. Ref. | O-AT  | BS-AT     | Ref.  | Spec. Ref. | O-AT  | BS-AT |
| Streamcluster | Small  | SISD    | 9.75      | 10.2       | 9.26  | 9.06      | 3.26  | 4.00       | 2.66  | 2.47  |
|               |        | SIMD    | 3.84      | 3.79       | 3.74  | 3.51      | 3.33  | 2.90       | 2.51  | 2.24  |
|               | Medium | SISD    | 19.9      | 21.8       | 17.9  | 17.8      | 7.54  | 11.1       | 5.87  | 5.68  |
|               |        | SIMD    | 7.13      | 7.05       | 6.59  | 5.93      | 9.09  | 8.86       | 5.09  | 4.84  |
|               | Large  | SISD    | 46.8      | 46.1       | 41.0  | 40.8      | 14.8  | 14.7       | 12.0  | 11.3  |
|               |        | SIMD    | 15.1      | 15.0       | 11.1  | 10.2      | 17.2  | 15.1       | 10.1  | 9.84  |
| VIPS lintra   | Small  | SISD    | 0.841     | 0.842      | 0.676 | 0.640     | 0.502 | 0.504      | 0.456 | 0.443 |
|               |        | SIMD    | 0.556     | 0.563      | 0.584 | 0.510     | 0.455 | 0.454      | 0.471 | 0.442 |
|               | Medium | SISD    | 2.30      | 2.29       | 1.76  | 1.73      | 1.47  | 1.41       | 1.37  | 1.24  |
|               |        | SIMD    | 1.47      | 1.48       | 1.40  | 1.36      | 1.31  | 1.41       | 1.31  | 1.26  |
|               | Large  | SISD    | 26.6      | 24.3       | 25.1  | 22.9      | 10.1  | 9.63       | 9.88  | 9.49  |
|               |        | SIMD    | 24.7      | 24.9       | 24.0  | 22.2      | 10.4  | 10.0       | 9.94  | 9.54  |

Spec. ref.: Reference kernel specialized as in the auto-tuned versions.

O-AT: Online auto-tuned kernel.

BS-AT: Best statically auto-tuned kernel.

Table 4: Statistics of online auto-tuning in the Cortex-A8 and A9 (SISD / SIMD separated, or average if minor variations).

| Bench.             | Input set | Explo-<br>rable<br>versions | Exploration<br>limit in one<br>run | Kernel<br>calls | Run-time regeneration and space exploration |               |                    |                |             |       |
|--------------------|-----------|-----------------------------|------------------------------------|-----------------|---|---------------|--------------------|----------------|-------------|-------|
|                    |           |                             |                                    |                 | Explored                                    |               | Overhead to bench. |                | run-time    |       |
|                    |           |                             |                                    |                 | A8  | A9            | A8                 | A9             | A8          | A9    |
| Stream-<br>cluster | Small     | 390                         | 43-49                              | 49              | 49  | 0.4 % (11 ms) | 0.2 % (11 ms)      | 0.4 % (9.2 ms) | 13 / 4.4 %  | 32 %  |
|                    | Medium    | 510                         | 55-61                              | 58              | 61  | 0.2 % (17 ms) | 0.2 % (17 ms)      | 0.3 % (15 ms)  | 6.3 / 2.7 % | 22 %  |
|                    | Large     | 630                         | 67-73                              | 67              | 73  | 0.2 % (30 ms) | 0.2 % (30 ms)      | 0.2 % (26 ms)  | 5.6 / 1.8 % | 15 %  |
| VIPS               | Small     | 858                         | 106-112                            | 44              | 28  | 4.2 % (26 ms) | 4.2 % (26 ms)      | 2.5 % (12 ms)  | 100 %       | 100 % |
|                    | Medium    | 330                         | 39-45                              | 40              | 42  | 0.9 % (14 ms) | 0.9 % (14 ms)      | 1.0 % (14 ms)  | 18 %        | 66 %  |
|                    | Large     | 596                         | 73-79                              | 75              | 71  | 0.3 % (71 ms) | 0.3 % (71 ms)      | 0.8 % (78 ms)  | 28 %        | 86 %  |