Optimization of Compute-Bound Applications

Stanford University, ME344 — Summer 2018

Andrey Vladimirov, Colfax International

Compute-Bound Performance

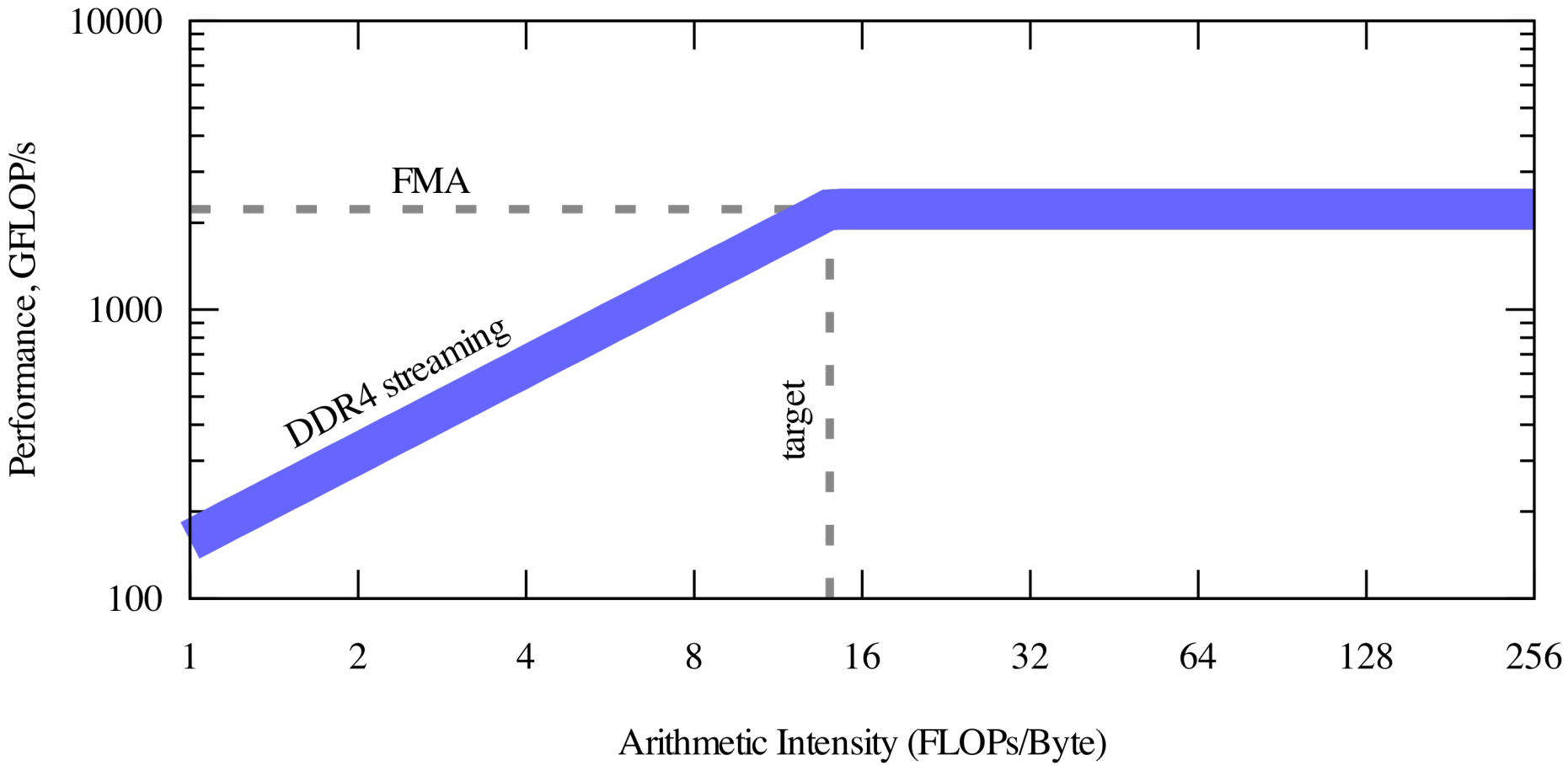
Theoretical Peak Performance of an Intel Xeon Gold 6128 CPU =

6 cores/socket
× 2 sockets
× $2.9 \cdot 10^9$ cycles/second
× 16 floating-point values/vector
× 2 instructions/cycle
× 2 FLOPs/instruction =
= $2227.2 \cdot 10^9$ FLOP/s = **2227.2 GFLOP/s**

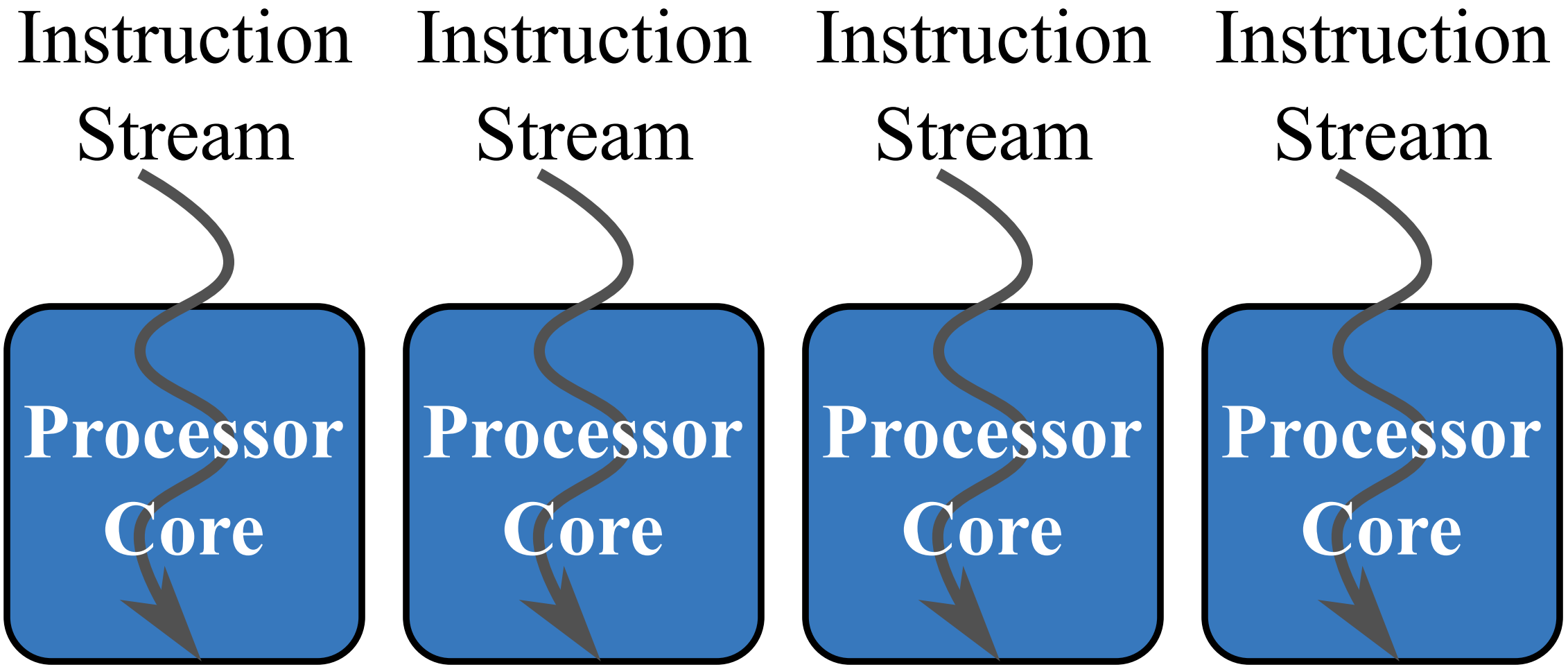
DRAM Bandwidth = **160 GB/s**

Target arithmetic Intensity = **2227.2 GFLOP/s** / **160 GB/s** ≅ **14 FLOPs/Byte**

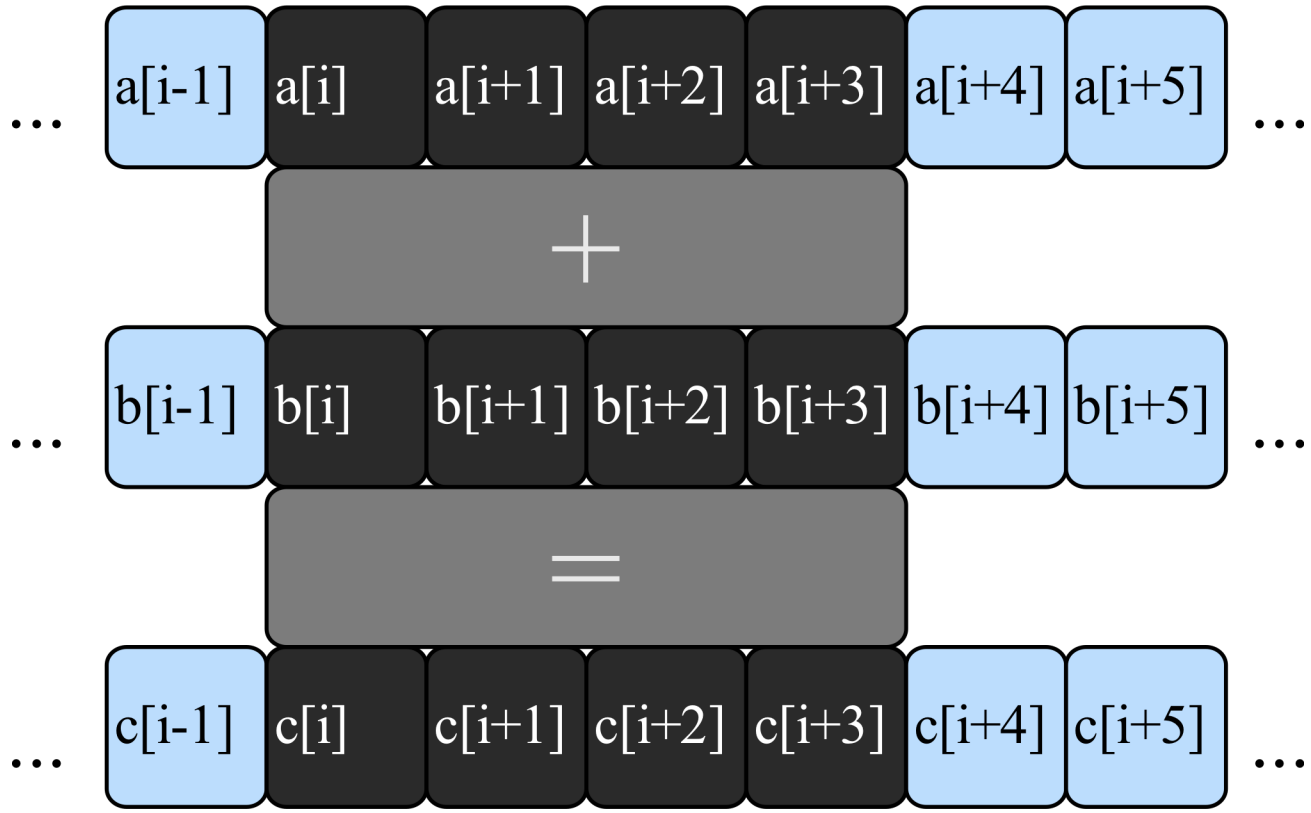
Roofline Diagram



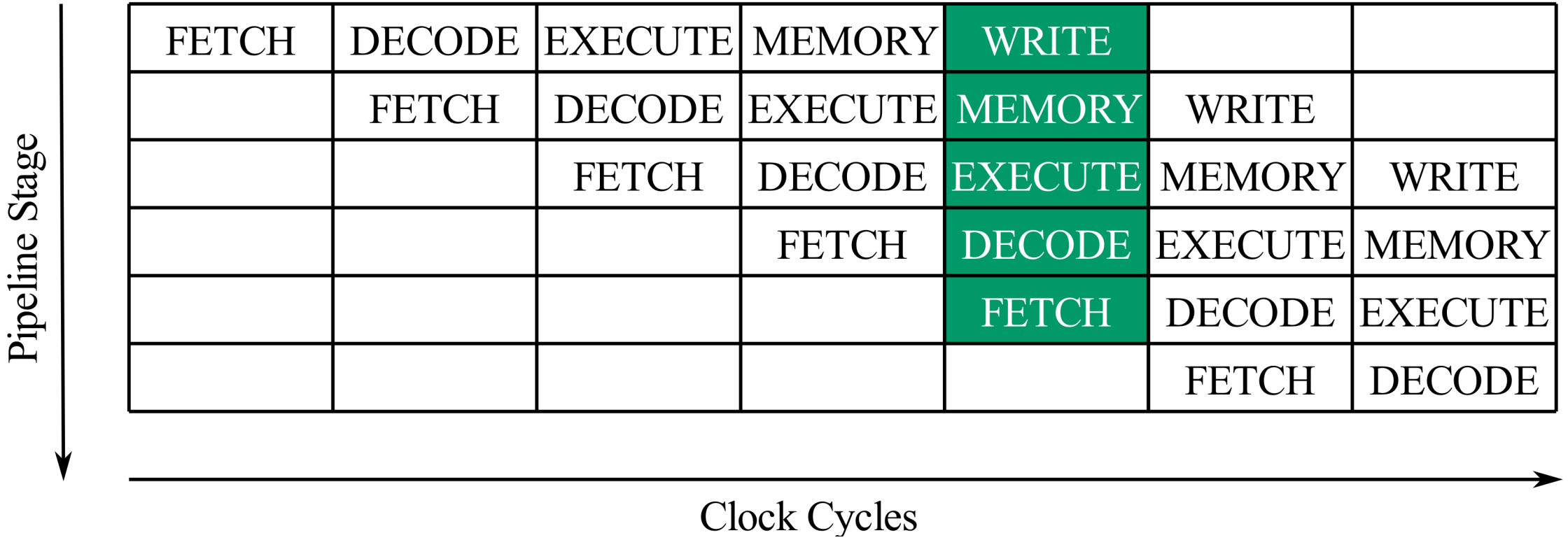
Memory Controllers and Threads



Vectorization

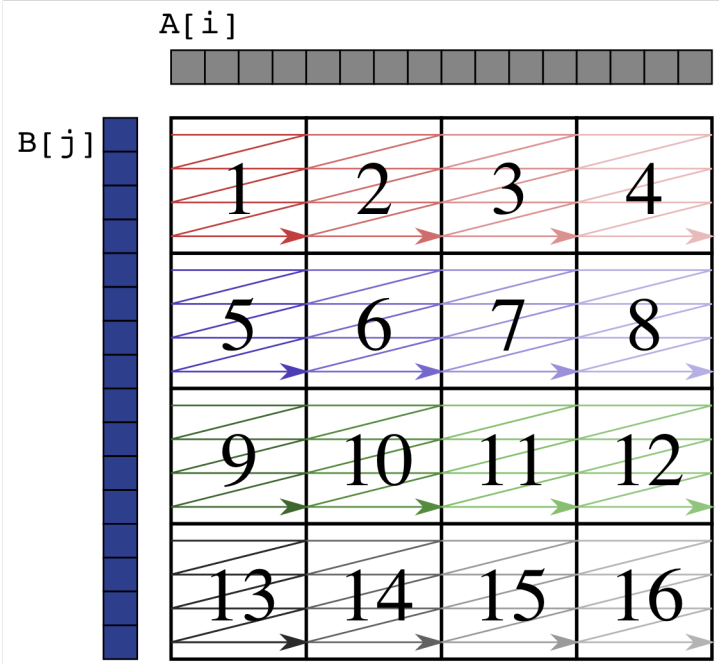


Instruction Pipelining

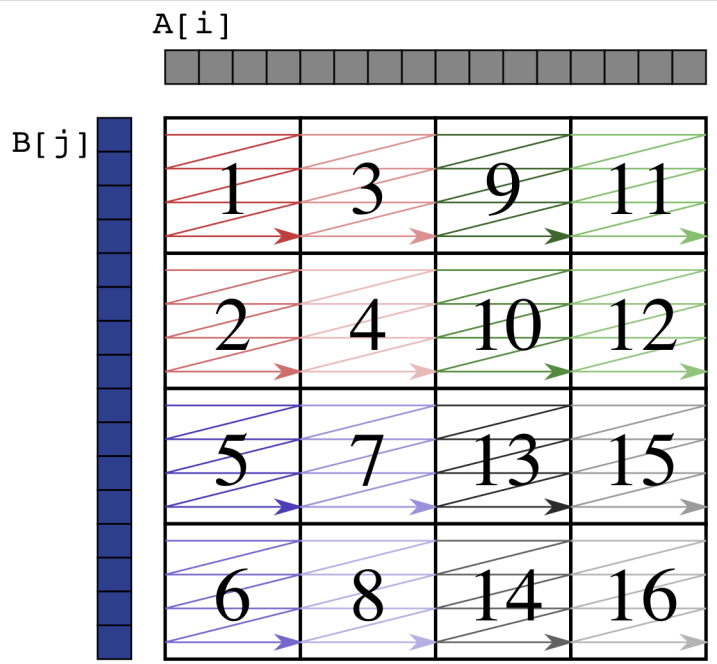


FETCH	DECODE	EXECUTE	MEMORY	WRITE		
	FETCH	DECODE	EXECUTE	MEMORY	WRITE	
		FETCH	DECODE	EXECUTE	MEMORY	WRITE
			FETCH	DECODE	EXECUTE	MEMORY
				FETCH	DECODE	EXECUTE
					FETCH	DECODE

Data Reuse with Tiling and Recursion

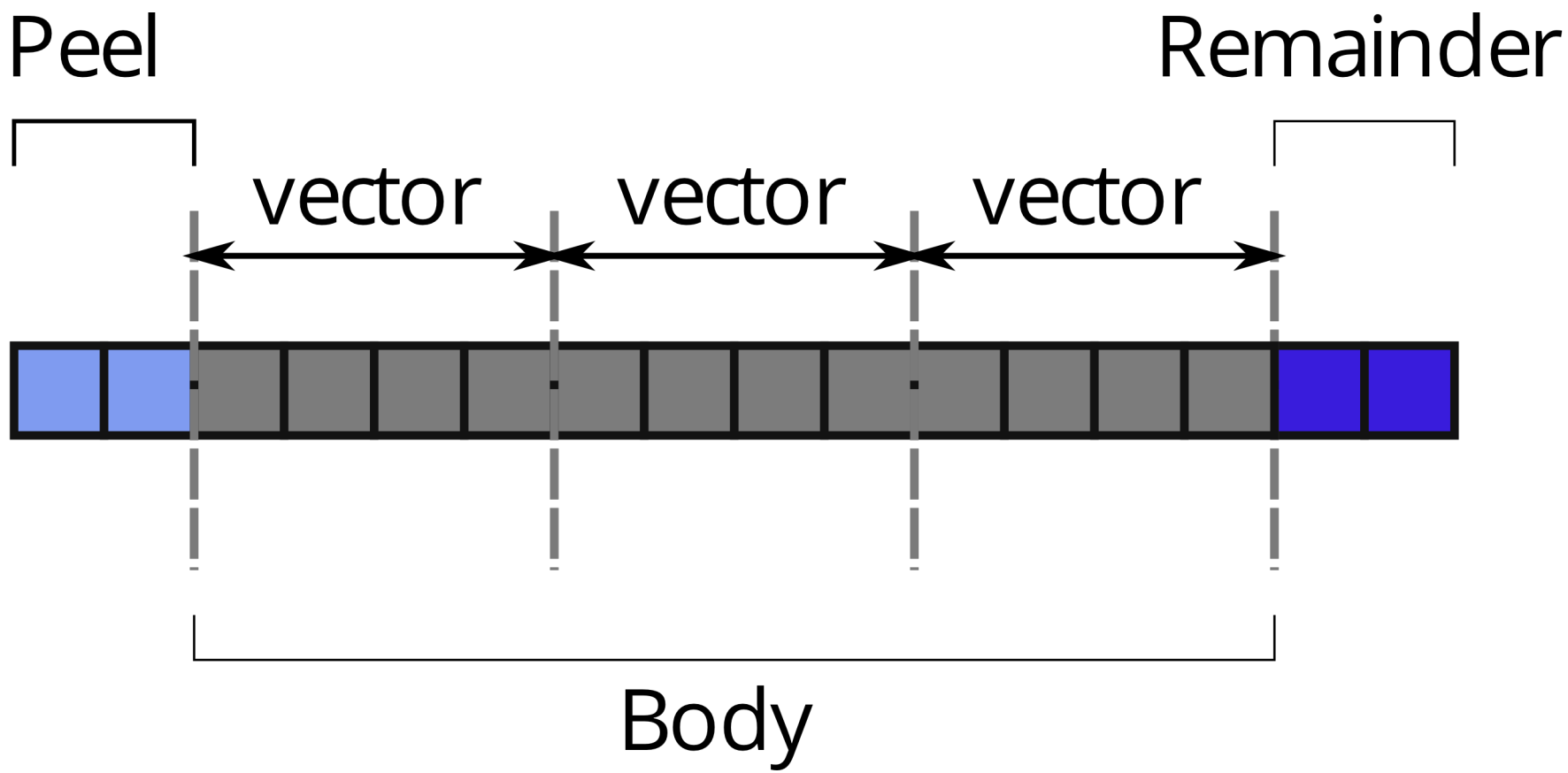



```
for (jj=0...n)
for (ii=0...n)
for (j=jj...jj+TILE)
for (i=ii...ii+TILE)
W(A[i], B[j])
```



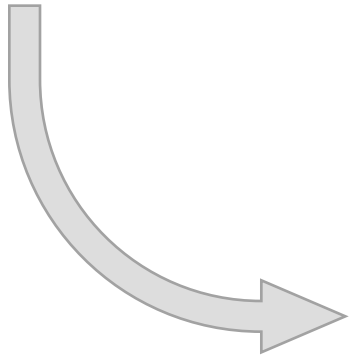

```
if (Threshold(nn,mm)) Loop();
else { if (nn > mm)
Recurse(nn/2, mm);
else
Recurse(nn, mm/2); }
```

Loop Overhead



FMACrunch: Synthetic Benchmark

```
for (v = 0; v < N_CHAINED_FMAS; v++)
  for (l = 0; l < VECTOR_WIDTH; l++)
    a[v*VECTOR_WIDTH + l] = a[v*VECTOR_WIDTH + l]*b[l] + c[l];
```



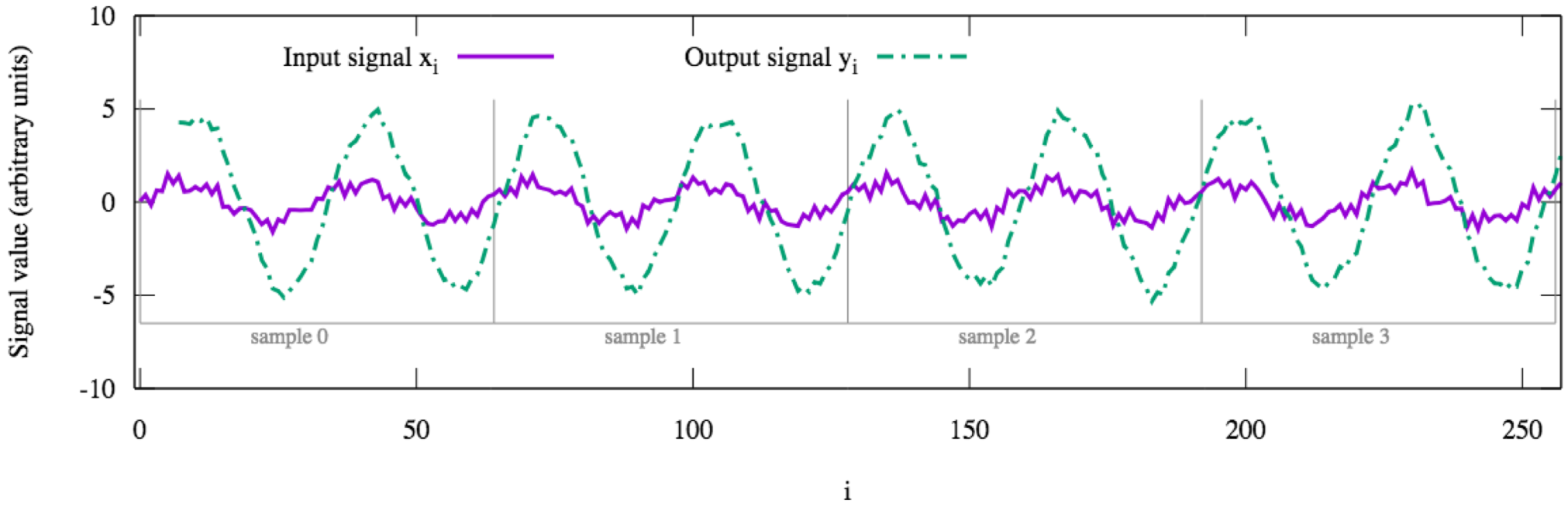
```
..B1.28:
        incq      %rdx
        vfmadd213ps %zmm8, %zmm9, %zmm7
        vfmadd213ps %zmm8, %zmm9, %zmm6
        vfmadd213ps %zmm8, %zmm9, %zmm5
        vfmadd213ps %zmm8, %zmm9, %zmm4
        vfmadd213ps %zmm8, %zmm9, %zmm3
        vfmadd213ps %zmm8, %zmm9, %zmm2
        vfmadd213ps %zmm8, %zmm9, %zmm1
        vfmadd213ps %zmm8, %zmm9, %zmm0
        cmpq      %rax, %rdx
        jb        ..B1.28       # Prob 82%
```

Access Online Materials

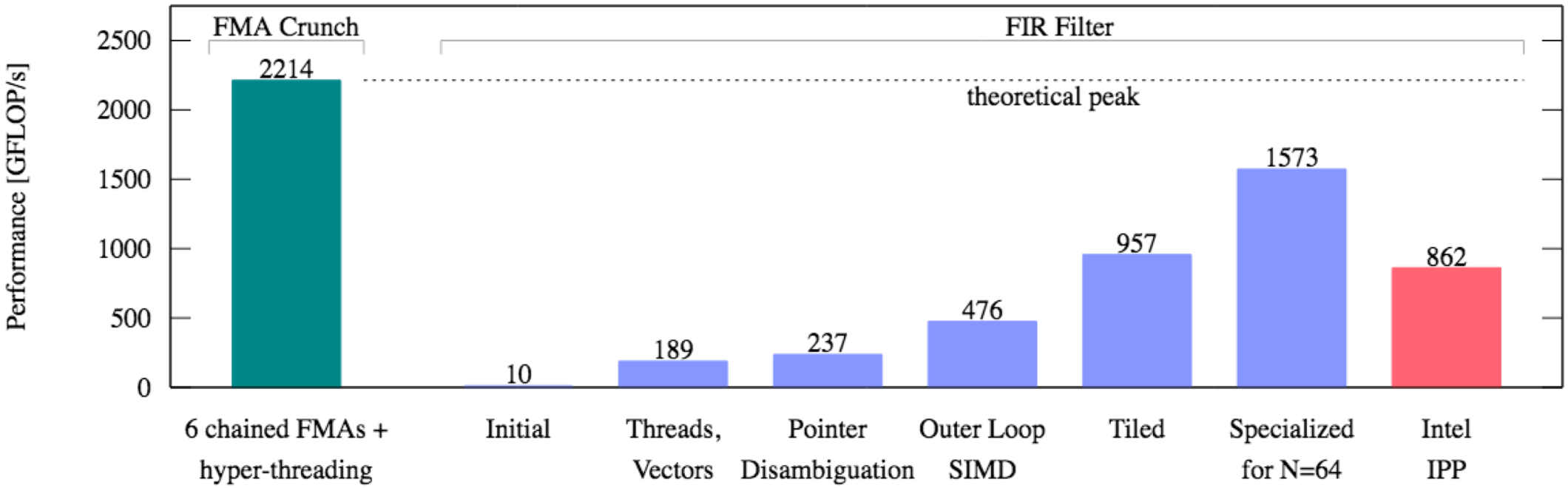
mc2learning.com/course/optimization-of-compute-bound-applications-c-lin/

FIR Filters in Computing

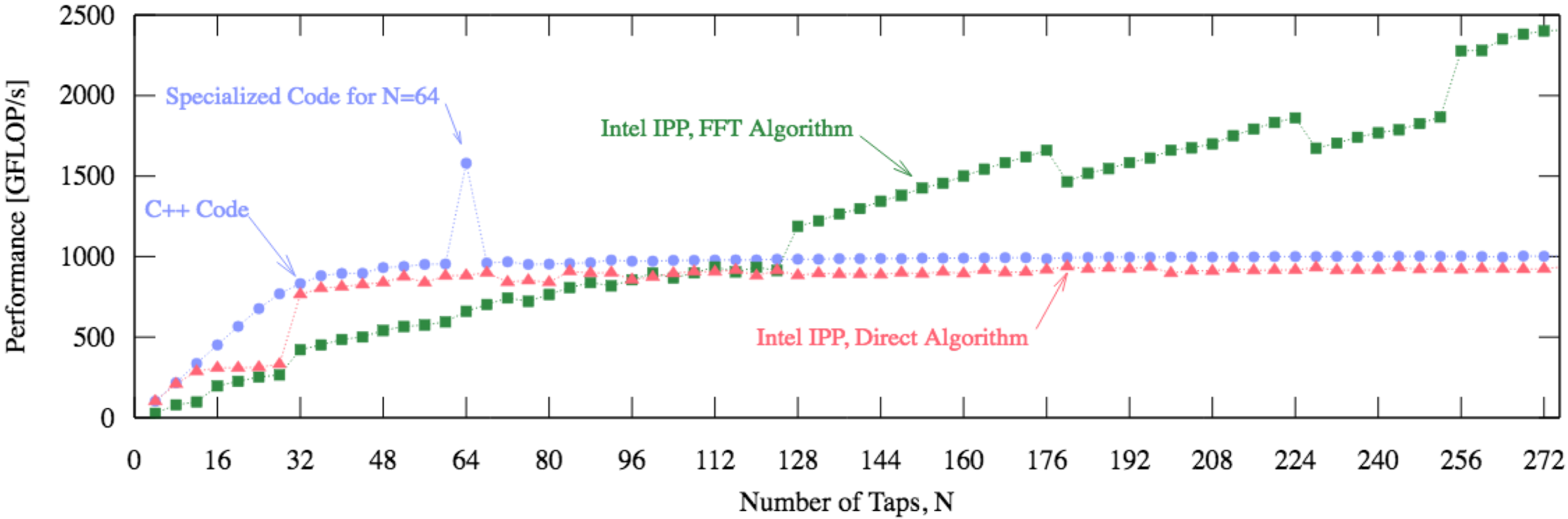
$$y_j = \sum_{k=0}^{N-1} w_k x_{j-k}$$



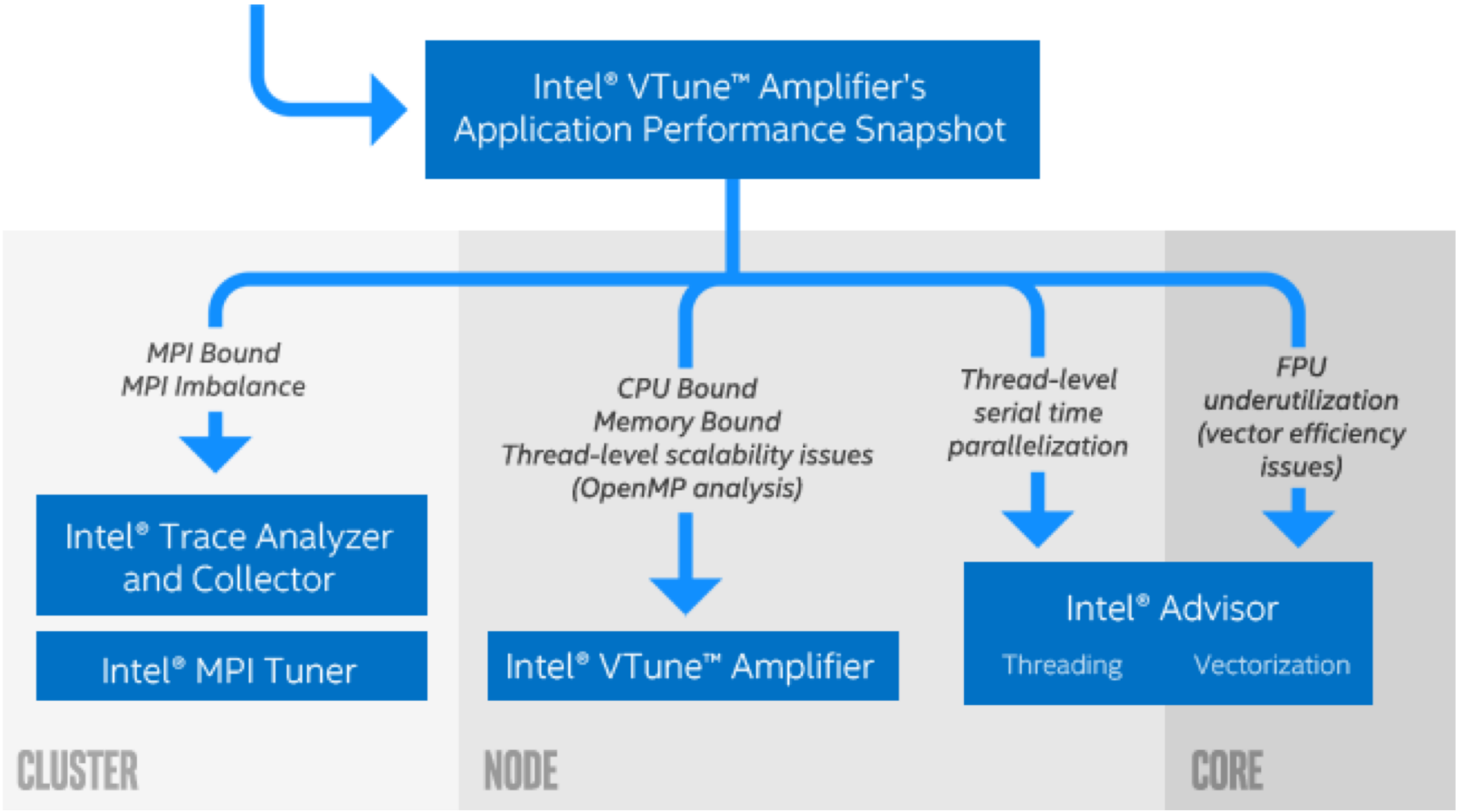
Optimization



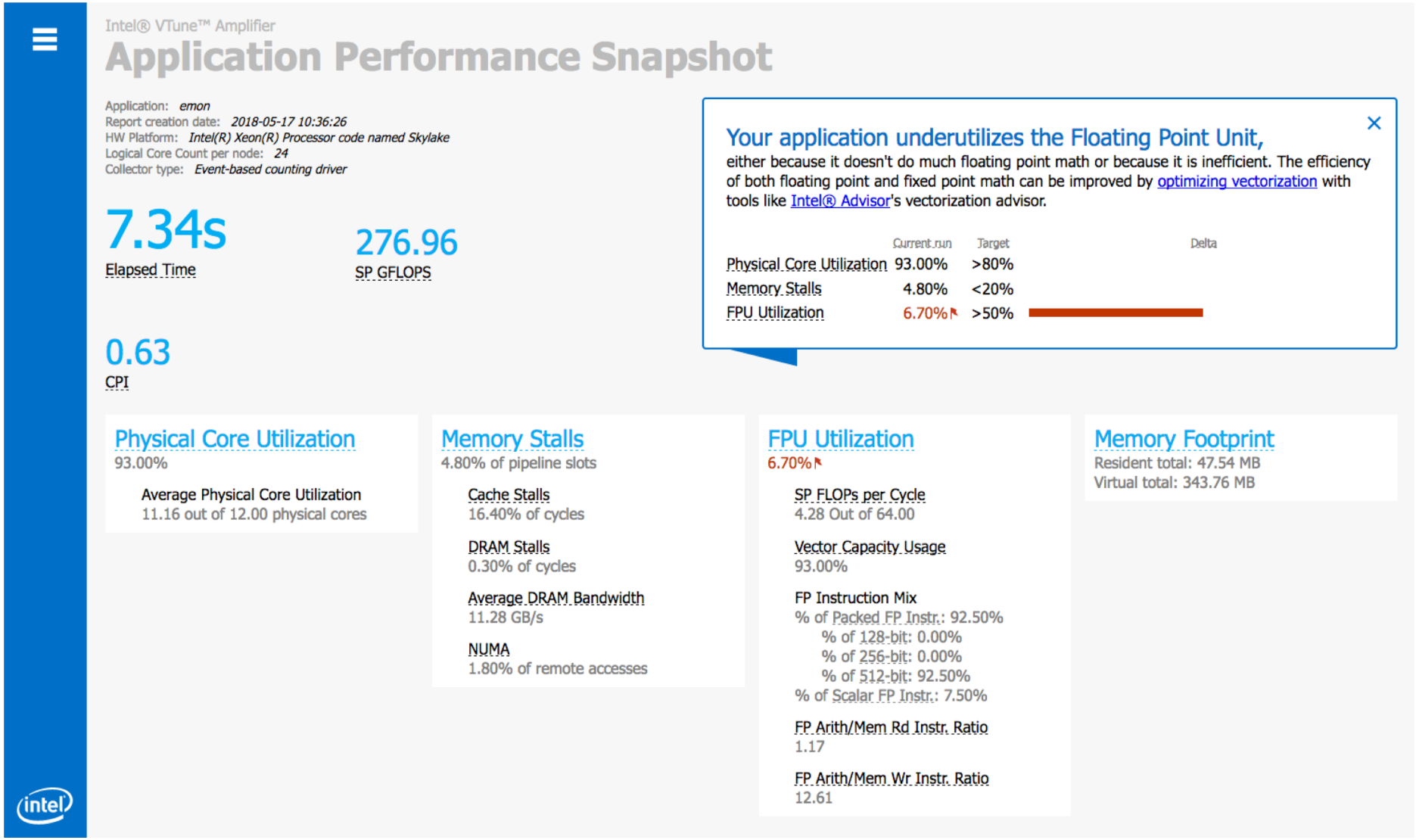
Value Proposition of Intel IPP



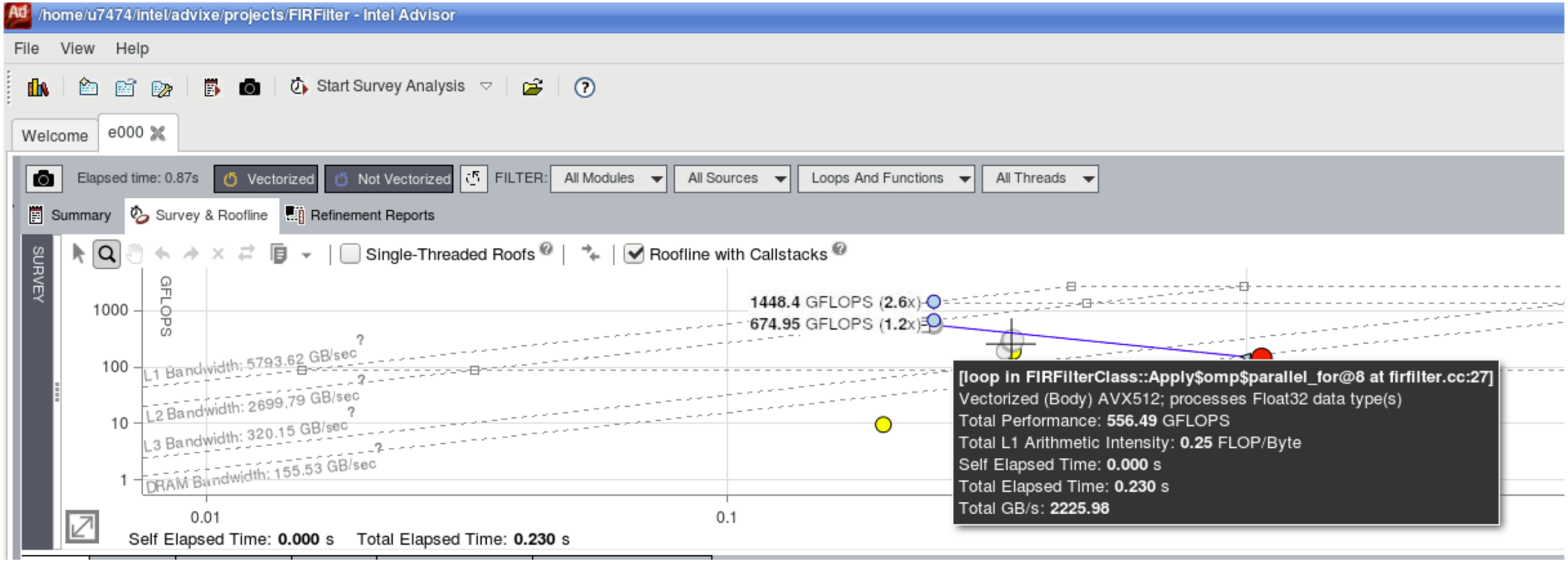
Intel Parallel Studio XE Tools



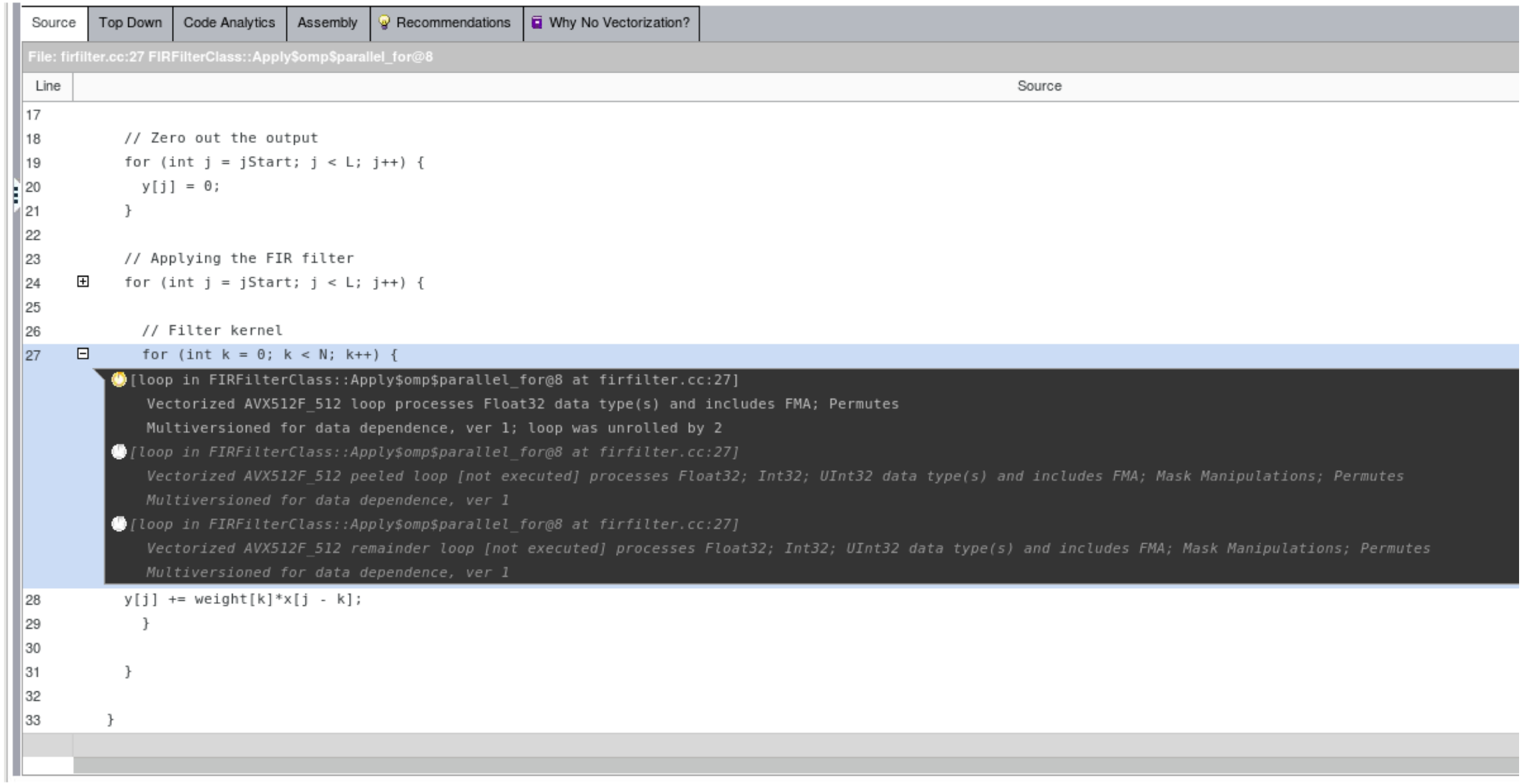
Application Performance Snapshot



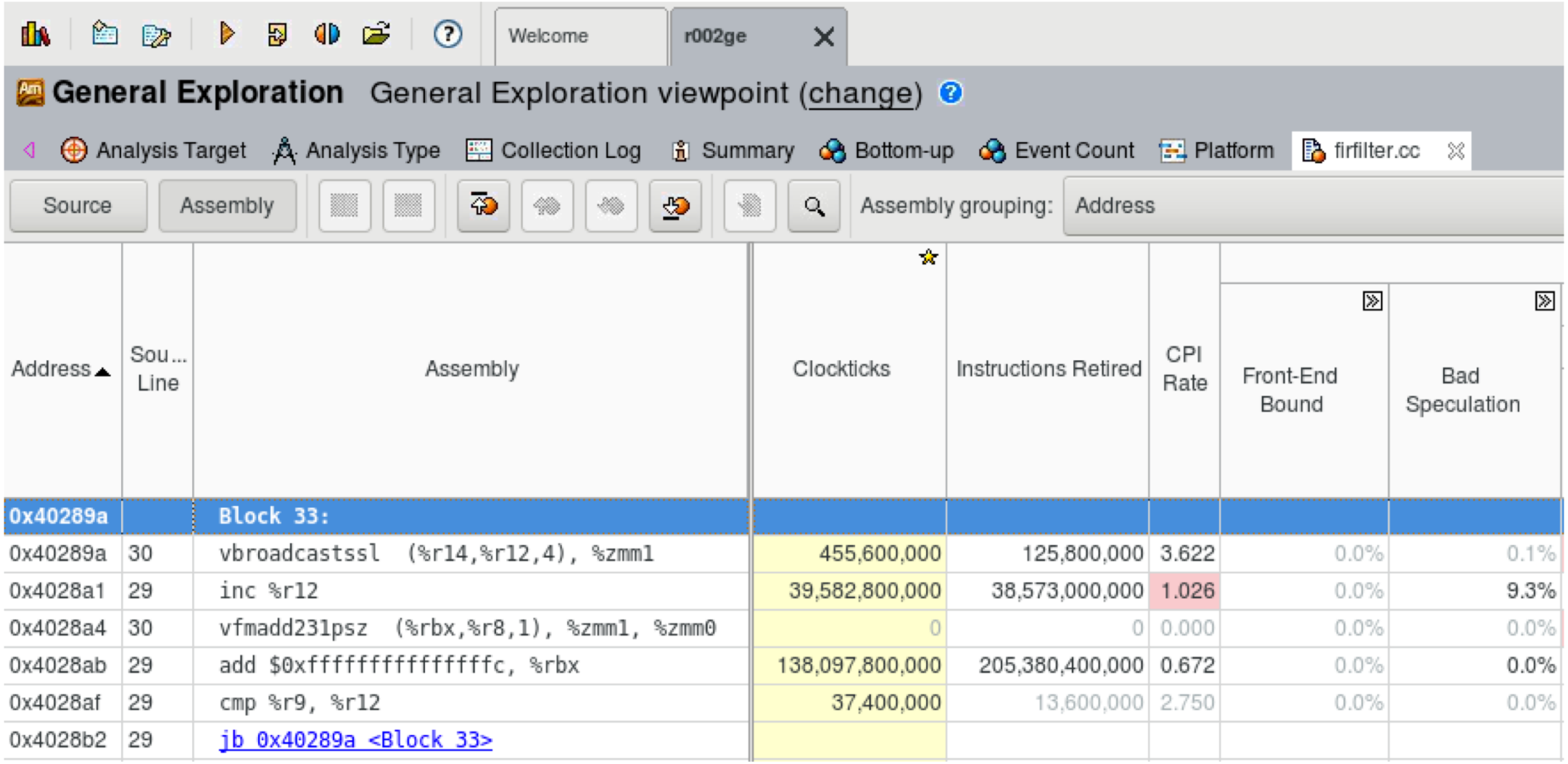
Intel Advisor XE



Intel Advisor XE



Intel VTune Amplifier XE



Address	Sou... Line	Assembly	Clockticks	Instructions Retired	CPI Rate	Front-End Bound	Bad Speculation
0x40289a		Block 33:					
0x40289a	30	vbroadcastssl (%r14,%r12,4), %zmm1	455,600,000	125,800,000	3.622	0.0%	0.1%
0x4028a1	29	inc %r12	39,582,800,000	38,573,000,000	1.026	0.0%	9.3%
0x4028a4	30	vfmadd231psz (%rbx,%r8,1), %zmm1, %zmm0	0	0	0.000	0.0%	0.0%
0x4028ab	29	add $0xfffffffffffffffc, %rbx	138,097,800,000	205,380,400,000	0.672	0.0%	0.0%
0x4028af	29	cmp %r9, %r12	37,400,000	13,600,000	2.750	0.0%	0.0%
0x4028b2	29	jb 0x40289a <Block 33>					