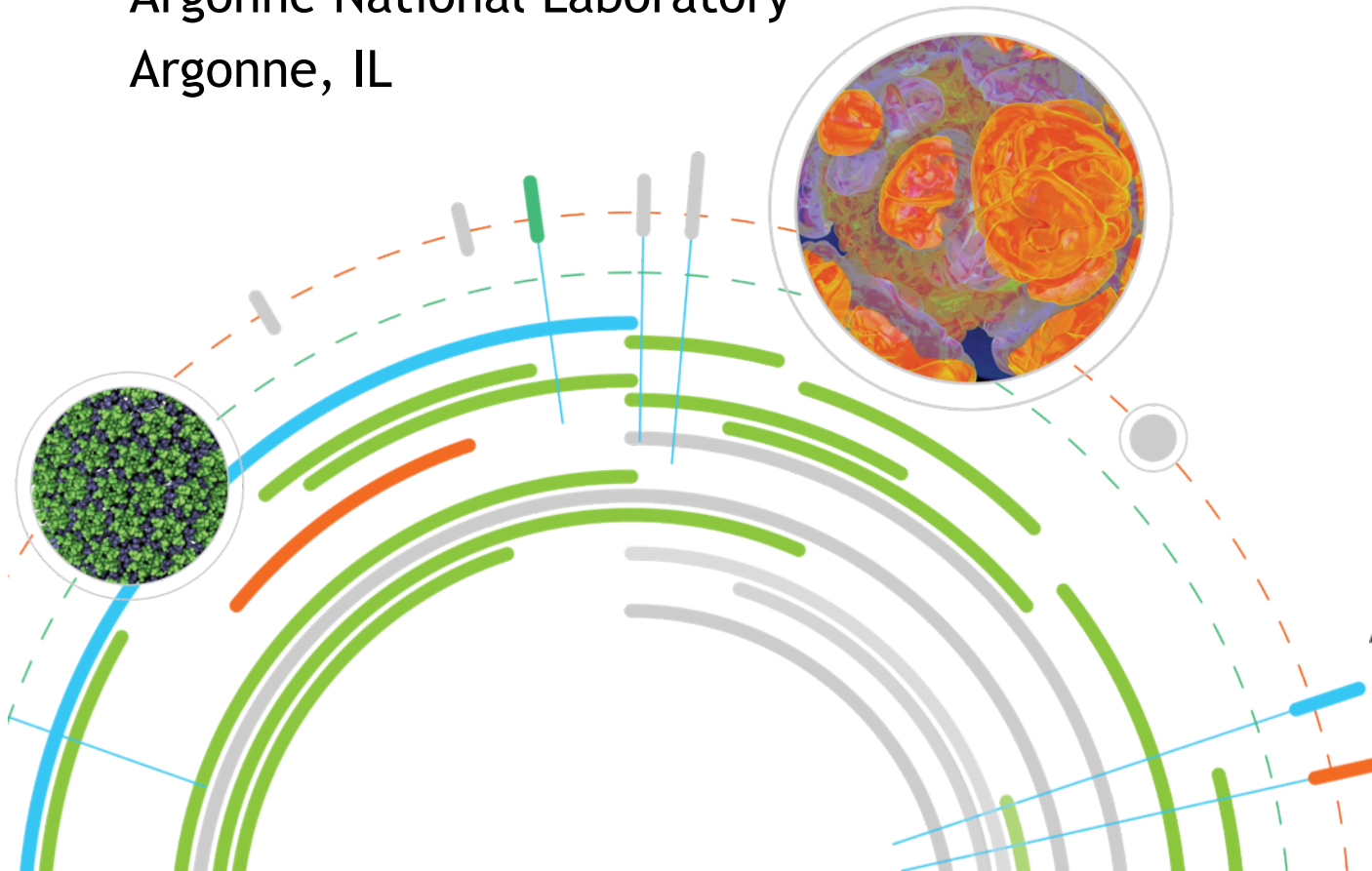


KNL B0 bin3 Progress Update

Vitali Morozov

Argonne National Laboratory
Argonne, IL



Argonne Leadership
Computing Facility



Instruction Throughput: 2 IPC, 90-92% of Flop peak

```
..B1.8: # (8*4+3)/4=8.75 instructions, 16 flops, 18 cycles
vfmadd231ps %zmm0, %zmm0, %zmm0
vfmadd231ps %zmm1, %zmm1, %zmm1
vfmadd231ps %zmm2, %zmm2, %zmm2
vfmadd231ps %zmm3, %zmm3, %zmm3
vfmadd231ps %zmm20, %zmm20, %zmm20
vfmadd231ps %zmm21, %zmm21, %zmm21
vfmadd231ps %zmm22, %zmm22, %zmm22
vfmadd231ps %zmm23, %zmm23, %zmm23
vfmadd231ps %zmm24, %zmm24, %zmm24
vfmadd231ps %zmm25, %zmm25, %zmm25
vfmadd231ps %zmm26, %zmm26, %zmm26
vfmadd231ps %zmm27, %zmm27, %zmm27
vfmadd231ps %zmm28, %zmm28, %zmm28
vfmadd231ps %zmm29, %zmm29, %zmm29
vfmadd231ps %zmm30, %zmm30, %zmm30
vfmadd231ps %zmm31, %zmm31, %zmm31
vfmadd231ps %zmm0, %zmm0, %zmm0
vfmadd231ps %zmm1, %zmm1, %zmm1
vfmadd231ps %zmm2, %zmm2, %zmm2
vfmadd231ps %zmm3, %zmm3, %zmm3
vfmadd231ps %zmm20, %zmm20, %zmm20
vfmadd231ps %zmm21, %zmm21, %zmm21
vfmadd231ps %zmm22, %zmm22, %zmm22
vfmadd231ps %zmm23, %zmm23, %zmm23
vfmadd231ps %zmm24, %zmm24, %zmm24
vfmadd231ps %zmm25, %zmm25, %zmm25
vfmadd231ps %zmm26, %zmm26, %zmm26
vfmadd231ps %zmm27, %zmm27, %zmm27
vfmadd231ps %zmm28, %zmm28, %zmm28
vfmadd231ps %zmm29, %zmm29, %zmm29
vfmadd231ps %zmm30, %zmm30, %zmm30
vfmadd231ps %zmm31, %zmm31, %zmm31
addq      $64, %r11
cmpq      %rax, %r11
jb        ..B1.8
```

1 thread pre core

```
4496: IPC=2.0003 SF/cycle=58.5240 Peak rate=91.4437
4512: IPC=1.9911 SF/cycle=58.2539 Peak rate=91.0217
4528: IPC=1.9825 SF/cycle=58.0029 Peak rate=90.6296
4544: IPC=2.0268 SF/cycle=59.2980 Peak rate=92.6532
4560: IPC=2.0024 SF/cycle=58.5847 Peak rate=91.5386
4576: IPC=1.9954 SF/cycle=58.3810 Peak rate=91.2204
4592: IPC=1.9849 SF/cycle=58.0714 Peak rate=90.7366
4608: IPC=2.0293 SF/cycle=59.3702 Peak rate=92.7659
4624: IPC=2.0047 SF/cycle=58.6503 Peak rate=91.6412
4640: IPC=1.9859 SF/cycle=58.1030 Peak rate=90.7859
```

2 threads per core

```
4496: IPC=1.9220 SF/cycle=56.2317 Peak rate=87.8620
4512: IPC=1.9902 SF/cycle=58.2286 Peak rate=90.9822
4528: IPC=1.9816 SF/cycle=57.9762 Peak rate=90.5878
4544: IPC=2.0262 SF/cycle=59.2794 Peak rate=92.6241
4560: IPC=2.0009 SF/cycle=58.5397 Peak rate=91.4683
4576: IPC=1.9847 SF/cycle=58.0675 Peak rate=90.7305
4592: IPC=1.9758 SF/cycle=57.8048 Peak rate=90.3201
4608: IPC=2.0288 SF/cycle=59.3582 Peak rate=92.7472
4624: IPC=2.0055 SF/cycle=58.6763 Peak rate=91.6817
4640: IPC=1.9974 SF/cycle=58.4377 Peak rate=91.3089
```

4 threads per core

```
4496: IPC=1.8628 SF/cycle=54.5007 Peak rate=85.1573
4512: IPC=1.9783 SF/cycle=57.8808 Peak rate=90.4388
4528: IPC=1.9810 SF/cycle=57.9594 Peak rate=90.5616
4544: IPC=2.0257 SF/cycle=59.2649 Peak rate=92.6015
4560: IPC=2.0005 SF/cycle=58.5287 Peak rate=91.4511
4576: IPC=1.9937 SF/cycle=58.3292 Peak rate=91.1393
4592: IPC=1.9835 SF/cycle=58.0329 Peak rate=90.6763
4608: IPC=2.0276 SF/cycle=59.3221 Peak rate=92.6908
4624: IPC=2.0047 SF/cycle=58.6511 Peak rate=91.6424
4640: IPC=1.9960 SF/cycle=58.3978 Peak rate=91.2466
```

- Cleaning polynomial evaluation (bug reported)
- Unroll 4
 - / and sqrtf to rsqrt(f*f*f)
- Removing rsqrt refinement step
- Inverting subtraction and combine with loads (bug reported)
- Removing $r2 > 0.0f$ comparison
- Re-computing updates

HACCmk: 1.8 IPC, 67% of Flop Peak

```
400: time=178618 result 36825.56095735 IPC = 1.1785 SF/cycle = 27.8269 Peak rate = 43.4795
416: time=177483 result 39489.90405742 IPC = 1.2335 SF/cycle = 29.1250 Peak rate = 45.5078
432: time=181366 result 42233.55036113 IPC = 1.2535 SF/cycle = 29.5977 Peak rate = 46.2464
448: time=200031 result 45059.59772295 IPC = 1.1786 SF/cycle = 27.8299 Peak rate = 43.4842
464: time=202267 result 47980.09570992 IPC = 1.2072 SF/cycle = 28.5051 Peak rate = 44.5392
480: time=201535 result 50980.66289856 IPC = 1.2534 SF/cycle = 29.5952 Peak rate = 46.2426
496: time=205518 result 54074.49616614 IPC = 1.2701 SF/cycle = 29.9891 Peak rate = 46.8579
512: time=226246 result 57244.23585822 IPC = 1.1909 SF/cycle = 28.1202 Peak rate = 43.9378
```

1 thread/core
1.2 IPC, 46% of peak

```
400: time=147051 result 36825.56095735 IPC = 1.4315 SF/cycle = 33.8005 Peak rate = 52.8132
416: time=150424 result 39489.90405742 IPC = 1.4554 SF/cycle = 34.3642 Peak rate = 53.6940
432: time=153251 result 42233.55036113 IPC = 1.4835 SF/cycle = 35.0277 Peak rate = 54.7307
448: time=163363 result 45059.59772295 IPC = 1.4432 SF/cycle = 34.0764 Peak rate = 53.2443
464: time=166483 result 47980.09570992 IPC = 1.4667 SF/cycle = 34.6320 Peak rate = 54.1125
480: time=169353 result 50980.66289856 IPC = 1.4916 SF/cycle = 35.2191 Peak rate = 55.0298
496: time=172129 result 54074.49616614 IPC = 1.5165 SF/cycle = 35.8061 Peak rate = 55.9470
512: time=185523 result 57244.23585822 IPC = 1.4524 SF/cycle = 34.2928 Peak rate = 53.5825
```

2 threads/core
1.5 IPC, 55% of peak

```
400: time=133354 result 36825.56095735 IPC = 1.5785 SF/cycle = 37.2720 Peak rate = 58.2375
416: time=135981 result 39489.90405742 IPC = 1.6100 SF/cycle = 38.0140 Peak rate = 59.3969
432: time=138189 result 42233.55036113 IPC = 1.6452 SF/cycle = 38.8454 Peak rate = 60.6960
448: time=147200 result 45059.59772295 IPC = 1.6017 SF/cycle = 37.8182 Peak rate = 59.0910
464: time=150969 result 47980.09570992 IPC = 1.6175 SF/cycle = 38.1910 Peak rate = 59.6734
480: time=153791 result 50980.66289856 IPC = 1.6425 SF/cycle = 38.7830 Peak rate = 60.5984
496: time=156220 result 54074.49616614 IPC = 1.6709 SF/cycle = 39.4526 Peak rate = 61.6447
512: time=164534 result 57244.23585822 IPC = 1.6376 SF/cycle = 38.6673 Peak rate = 60.4176
```

4 thread/core
1.8 IPC, 67% of peak

```
2848: time=828398 result 1278312.89843813 IPC = 1.8093 SF/cycle = 42.7201 Peak rate = 66.7501
2864: time=830200 result 1290935.00380351 IPC = 1.8155 SF/cycle = 42.8668 Peak rate = 66.9794
2880: time=838919 result 1303614.36513717 IPC = 1.8067 SF/cycle = 42.6583 Peak rate = 66.6536
2896: time=842234 result 1316360.45010925 IPC = 1.8096 SF/cycle = 42.7265 Peak rate = 66.7601
2912: time=846028 result 1329148.42210893 IPC = 1.8114 SF/cycle = 42.7699 Peak rate = 66.8279
2928: time=850088 result 1341949.83329287 IPC = 1.8126 SF/cycle = 42.7995 Peak rate = 66.8742
2944: time=858274 result 1354864.52650320 IPC = 1.8052 SF/cycle = 42.6229 Peak rate = 66.5983
2960: time=860064 result 1367795.59051550 IPC = 1.8112 SF/cycle = 42.7654 Peak rate = 66.8209
2976: time=863812 result 1380780.63478190 IPC = 1.8131 SF/cycle = 42.8099 Peak rate = 66.8905
2992: time=865702 result 1393825.84251841 IPC = 1.8189 SF/cycle = 42.9462 Peak rate = 67.1034
3008: time=874331 result 1406923.12567775 IPC = 1.8105 SF/cycle = 42.7497 Peak rate = 66.7964
```

HACC Full Application: 5x core-to-core ratio!

320³: 16 cores, 1 node of BG/Q, ¼ of the node of KNL

Cores	RPN	OMP	TH	BG/Q Time, s	KNL B0, Time, s	Ratio
16	4	4	16	4297	616.3308	6.98
16	4	8	32	2677	543.7294	4.92
16	4	16	64	2504	530.2267	4.72
16	8	2	16	4362	544.7519	8.00
16	8	4	32	2571	459.5265	5.59
16	8	8	64	2278	437.2058	5.18
16	16	4	64	2581	468.5037	5.50

512³: 64 cores, 4 nodes of BG/Q, 1 node of KNL

Cores	RPN	OMP	TH	BG/Q Time, s	KNL B0, cache mode Time, s	KNL B0, flat mode Time, s	Ratio
64	16	4	64	4542	678.7571	678.2269	6.69
64	16	8	128	2823	606.1815	609.2007	4.66
64	16	16	256	2556	587.2716	587.4443	4.35
64	32	2	64	4747	620.7261	621.2356	7.65
64	32	4	128	2824	536.1650	534.9907	5.27
64	32	8	256	2503	503.0927	501.8637	4.98
64	64	4	256	2539	510.3745	506.7107	4.98