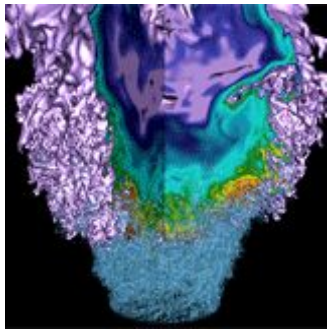
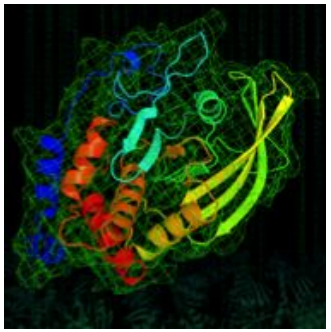
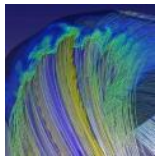
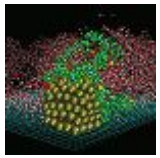
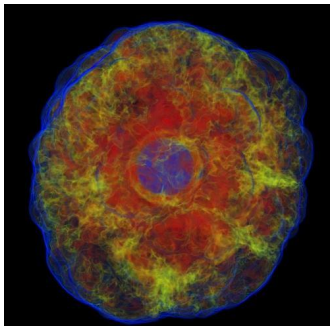
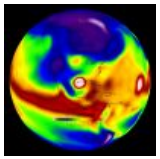
NUG Monthly Telecon

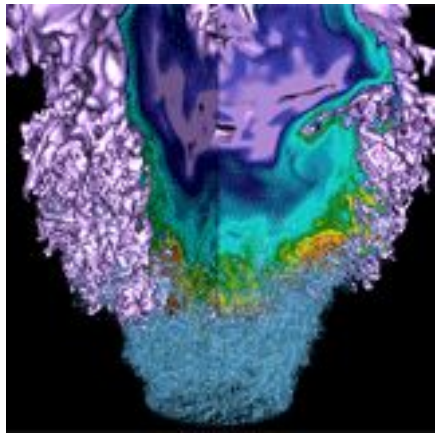

NERSC 40 YEARS at the FOREFRONT 1974-2014

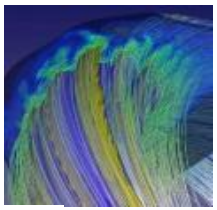
April 21 2016

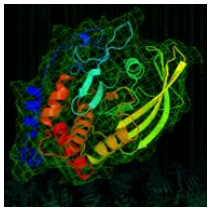
U.S. DEPARTMENT OF ENERGY | Office of Science

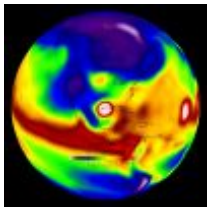
BERKELEY LAB
Lawrence Berkeley National Laboratory

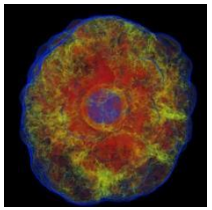
Recent Queue Changes on Cori and Edison

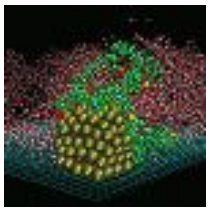












Helen He

NUG Meeting, 04/21/2016

Changes on Mar 23: Cori and Edison

- New SLURM Scheduling algorithm went into effect
- It enabled us to evaluate as many jobs as possible for backfill on the system instead of limited numbers with the previous algorithm.
- Significantly decreased debug and small regular wait time
- Increased overall system utilization

Changes on Mar 31 and Apr 4: Cori

- **Mar 31: Run limit for shared increased from 500 to 1000**
- **Apr 4: Largest debug job size reduced from 112 nodes to 64 nodes**
- **Apr 4: Disabled "qos=premium" for debug**
  - to prevent interruption with draining nodes for large debug jobs

Changes on Apr 11: Cori and Edison

- **Corrections made for scavenger eligibility**
  - A job will be put into scavenger if requested hours will run repo out of time
  - However, if an individual's allowed hours are insufficient to cover the job, but the repo is not out of time, the job will simply be rejected

Changes on Apr 15: Cori and Edison

- **Edison: Increased debug reservation pool**
  - 512 nodes (M-F, 6am - 6pm Pacific)
  - 256 nodes (M-F, 6pm - 6am Pacific and all day weekends)
- **Cori: Increased debug reservation pool**
  - 160 nodes (M-F, 6am - 6pm Pacific)
  - 128 nodes (M-F, 6pm - 6am Pacific and all day weekends)
- **Edison only: Max run limit increased from 1 to 2 for debug. Submit limit is 10 (5 on Cori). Sufficient?**

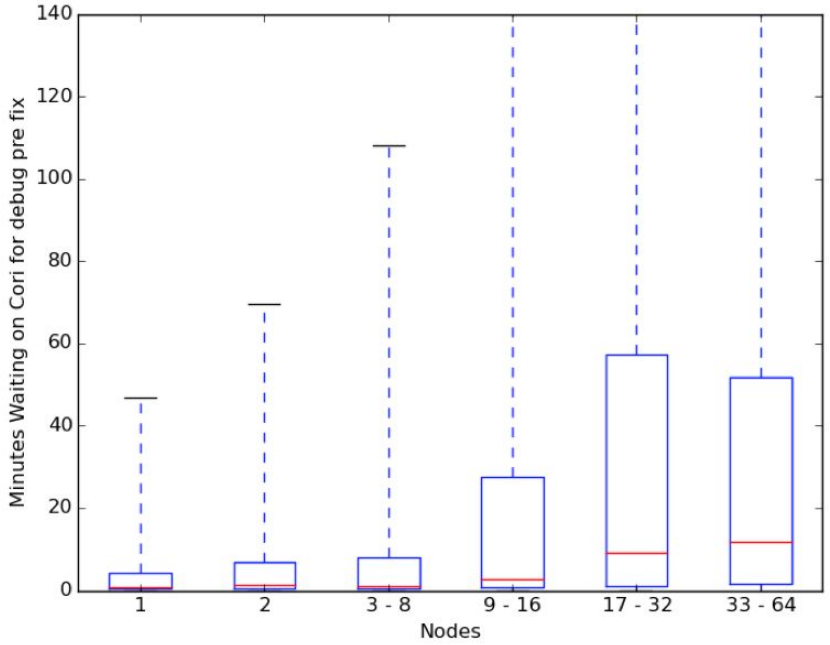
Coming Soon on Cori and Edison

- **Cori: Jobs > 512 nodes will be held until Friday at 9pm Pacific time, when they will be released to run over the weekend**
  - Minimize effects of draining for large jobs
  - Exploring possible implementation for Edison too
- **Cori only: Enable 4-day long, 1-node jobs**

Reminder: Edison is configured to support large jobs through priority boosts and discounts for jobs over 683 nodes

Cori Debug Wait Before and After Changes

Before 03/23/16



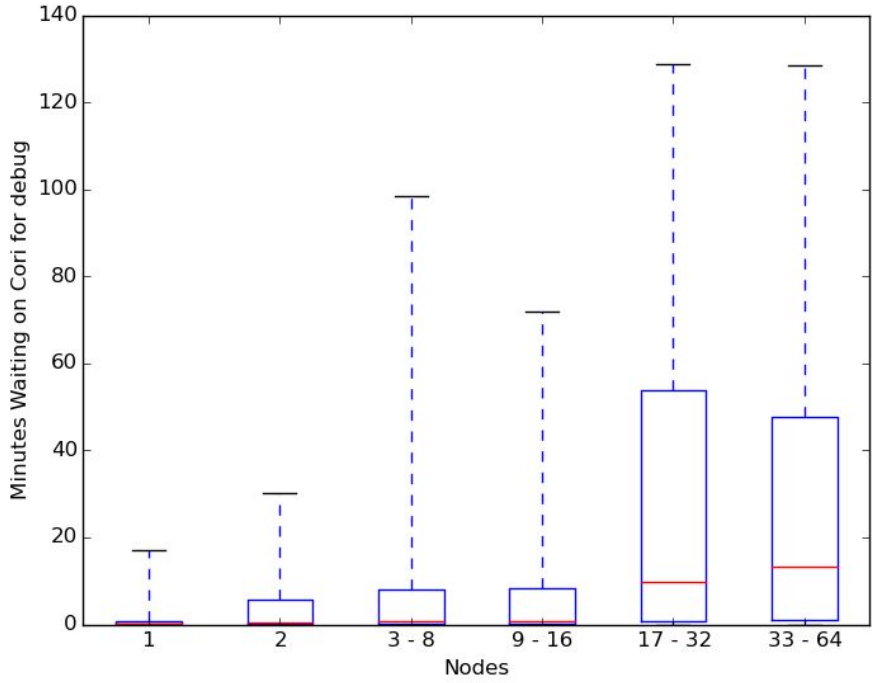
Whisker = 95% of jobs

Box = 25% to 75% of jobs

Red line = median

03/23/16 – 04/20/16

(exclude 3/31-4/7)

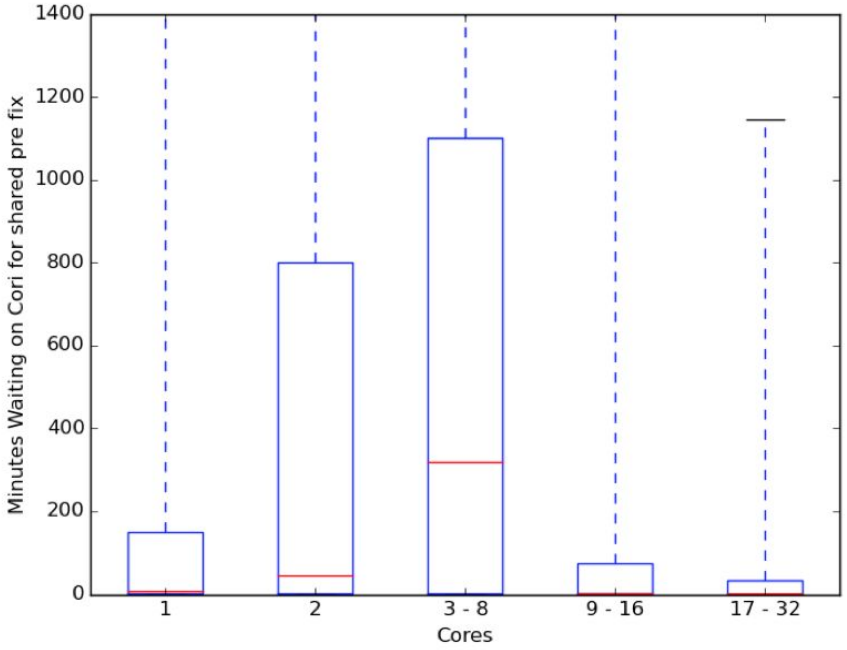


Keep in mind with max_run_limit=1, your 2nd, 3rd, … jobs will wait longer

Plots courtesy of Lisa Gerhardt

Cori Shared Wait Before and After Changes

Before 03/23/16



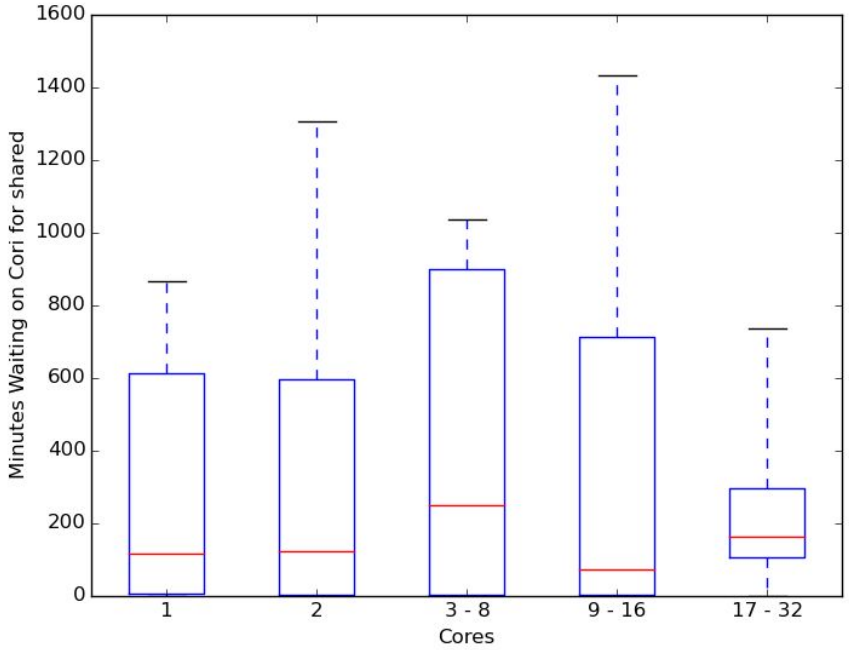
Whisker = 95% of job

Box = 25% to 75% of jobs

Red line = median

03/23/16 – 0/20/16

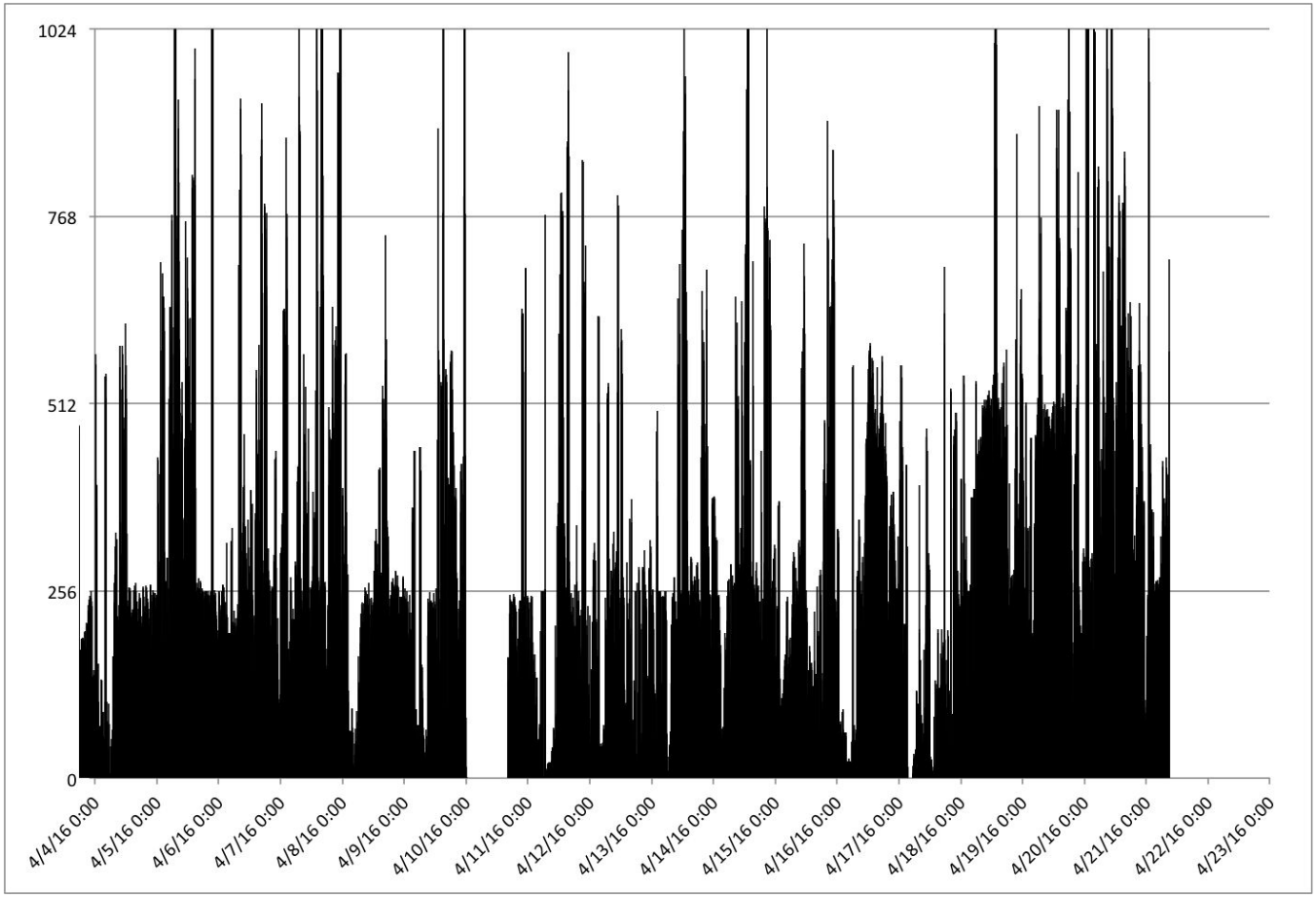
(exclude 3/31-4/7)



Keep in mind with max_run_limit, your extra submitted jobs will wait longer

Plots courtesy of Lisa Gerhardt

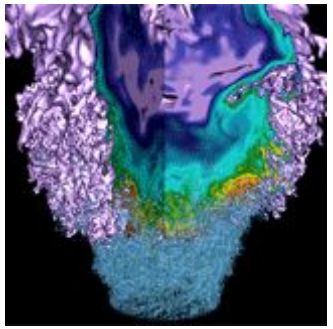
Cori Shared Wait Before and After Changes

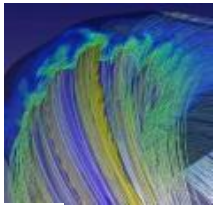


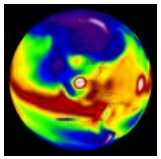
Places and Tools to Check Job Status

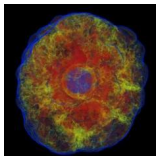
- **Completed jobs web page:**
  - https://my.nersc.gov/completedjobs.php
  - https://www.nersc.gov/users/job-logs-statistics/completed-jobs/
- **MyNERSC Queues display**
  - https://my.nersc.gov/queues.php
- **Queue Wait Times**
  - http://www.nersc.gov/users/queues/queue-wait-times/
- **Scripts described on Queue Monitoring Page**
  - https://www.nersc.gov/users/computational-systems/cori/running-jobs/monitoring-jobs/

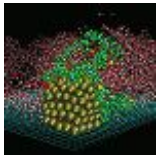
MyNERSC Update











NERSC 40 YEARS at the FOREFRONT

MyNERSC Upate

Most Parts of Site Returned to Normalcy After SLURM Update:

- Completed Jobs (Including SRUN Info)
- Queues
- Usage Stats
- Now Computing

Nodes	100
CPUs	2,400
Wall Hours Used	0.026
Wall Hours Requested	4.000
Raw Core Hours Used	63.33

Lustre I/O | STD-ERR | STD-OUT | Batch Script

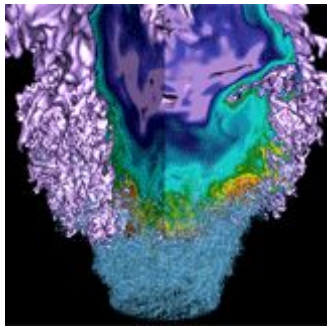
Aprun Information:

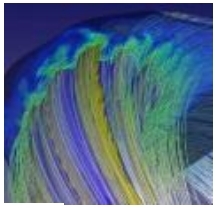
#	Command	Nodes	Time	Memory
0	epsilon.cplx.x	100	92s	61MB
Nodes:	nid02299			

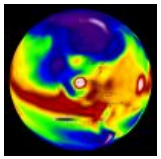
MyNERSC Upate

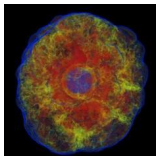
Demos

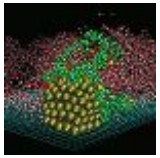
NUG 2016 Update











NERSC 40 YEARS at the FOREFRONT

NUG 2016 Wrap-Up

Thanks everyone for attending/organizing!

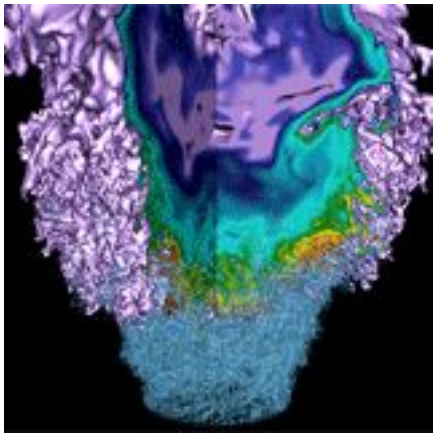
4 Days Packed With Science/Training/Tutorials

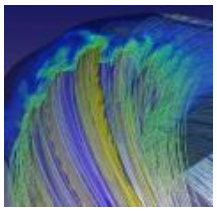
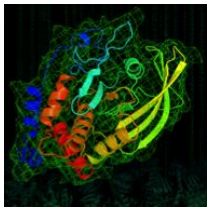
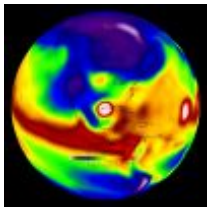
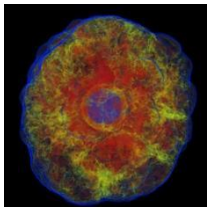
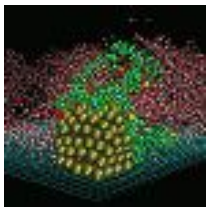
- **Cray on site for KNL Hack-a-thon**
- **New User Training**
- **KNL Application Readiness Day**
- **Science and Technology Day**
- **NERSC Business Day**

NERSC 2016 Award Winners

- **High Impact Science -** Charles Koven, William Riley (Berkeley Lab) David Lawrence (NCAR) Simulations on the effect of CO2 emmisions from melting permafrost
- **High Impact Early Career -** Nathan Howard (MI) High resolution multi-scale simulations in plasma turbulent simulations in Fusion reactor physis.
- **Innovative use of HPC** - Scott French (Google) helping seismologists create a unique 3D scan of the Earth's interior that resolved some long-standing questions about mantle plumes and volcanic hotspots.
- **Innovative use of HPC Early Career** - Min Si, (University of Tokyo, ANL), pioneering work in developing novel system software (CASPER) in the context of MPI-3 one-sided communication async progress.

Optimization of the particle-in-cell code WARP/PICSAR: preparation to Cori phase 2



Mathieu Lobet, Henri Vincenti, Remi Lehe, Ankit Bhagatwala, Jack Deslippe, Jean-Luc Vay

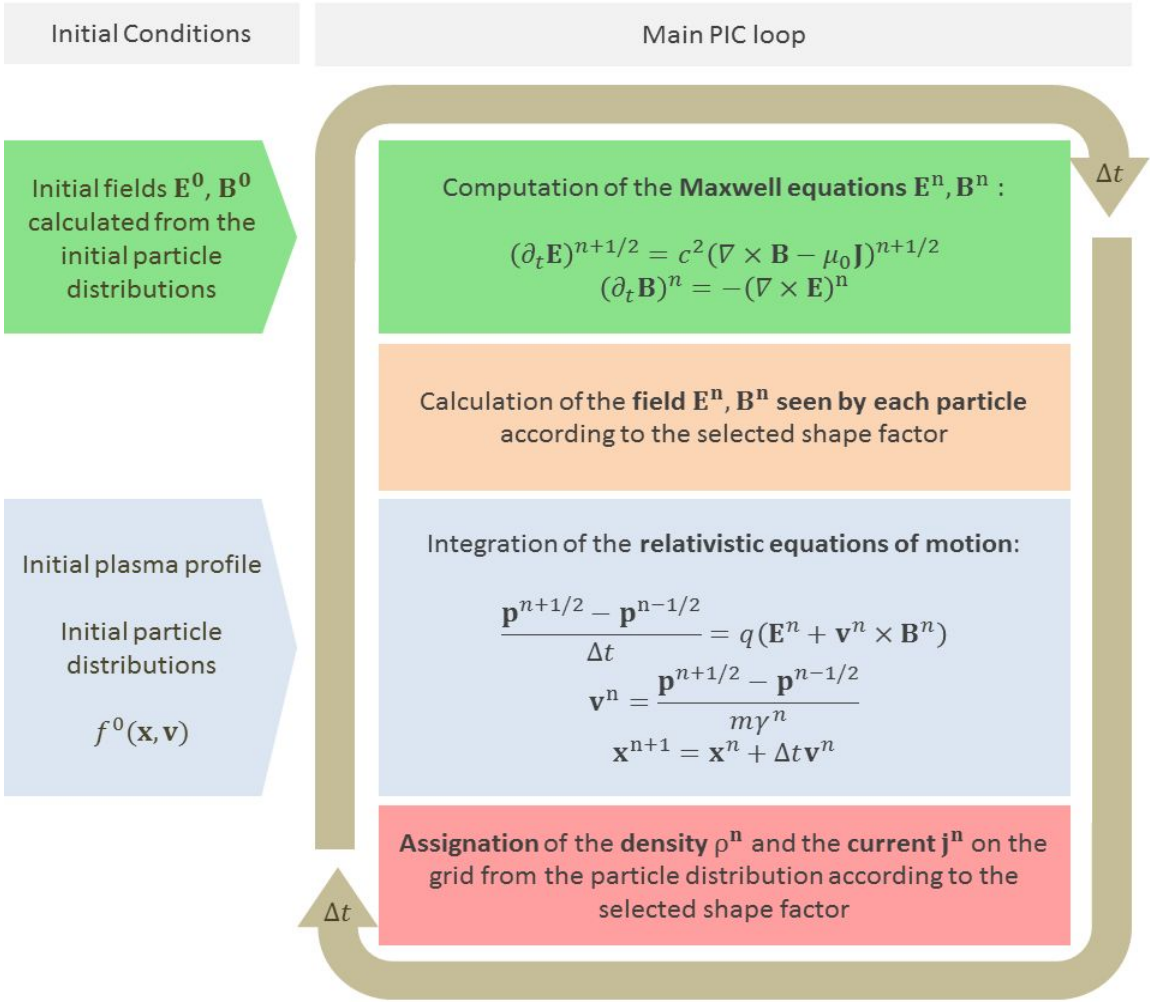
NERSC
April 20 2016

U.S. DEPARTMENT OF ENERGY | Office of Science

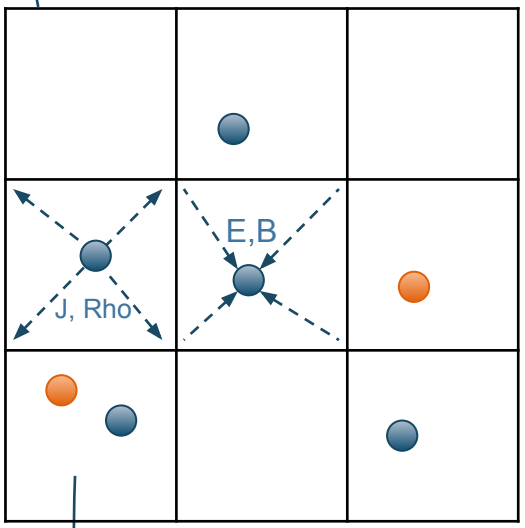
BERKELEY LAB
Lawrence Berkeley National Laboratory

What is a particle-in-cell (PIC) code ?

- Laser-matter interaction, plasma physics (laboratory and space media), charged beam acceleration/propagation



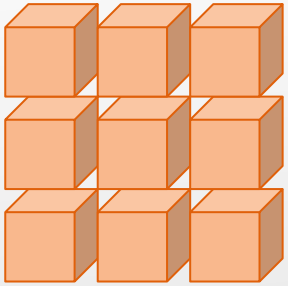
Grids for electromagnetic fields (E,B), current (J) and charge (Rho)



Super-particles of different species: group of real particles

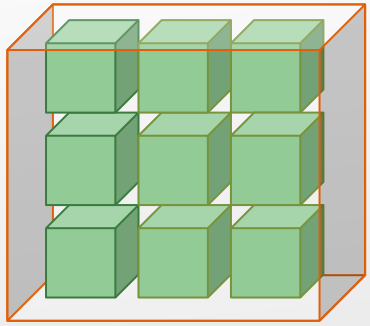
First optimization: a two-stage hybrid parallelization for memory locality

Domain decomposition (using MPI)



- Parallelization available in WARP
- Each subdomain contains their own particle arrays and field arrays

Subdomain decomposition into tiles (using OpenMP)



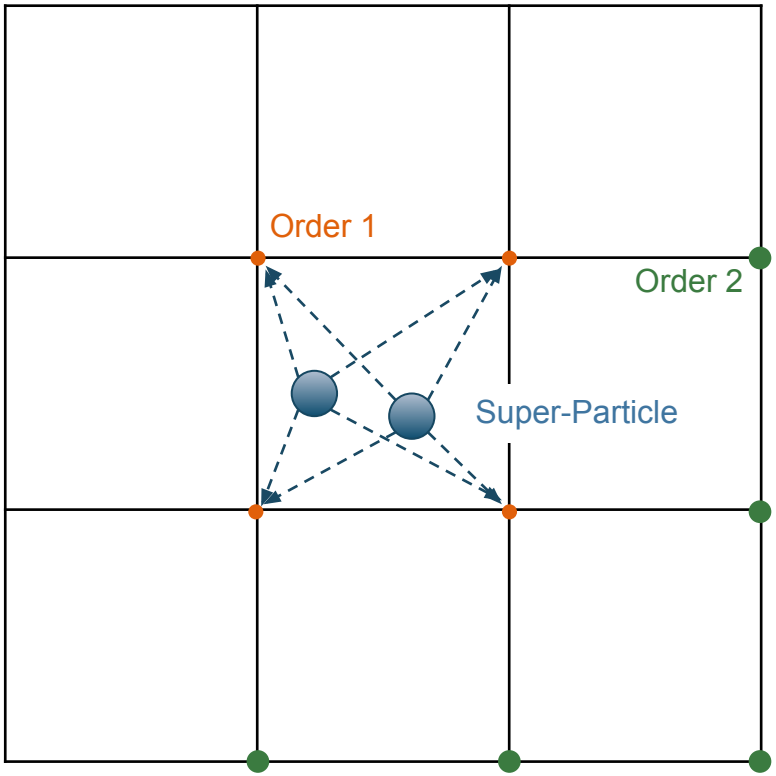
- Parallelization available via PICSAR
- Each tile contains their own particle arrays but fields are still shared

- On Xeon Phi KNL
  - 1 MPI task per NUMA domain

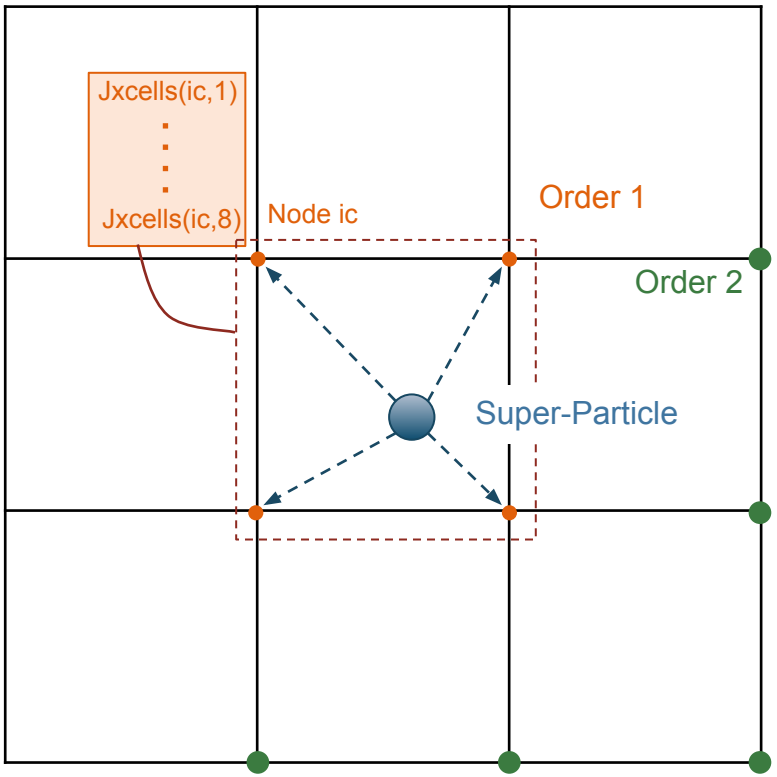
- On Xeon Phi KNL
  - 1 OpenMP thread per core
  - More tiles that OpenMP threads: load balancing
  - Tiles must fit in L2 cache for the fields/L3 or HBM for the particles

Second optimization: the current deposition

- Current grids Jx(NCELLS), Jy(NCELLS), Jz (NCELLS)
- Charge(NCELLS)

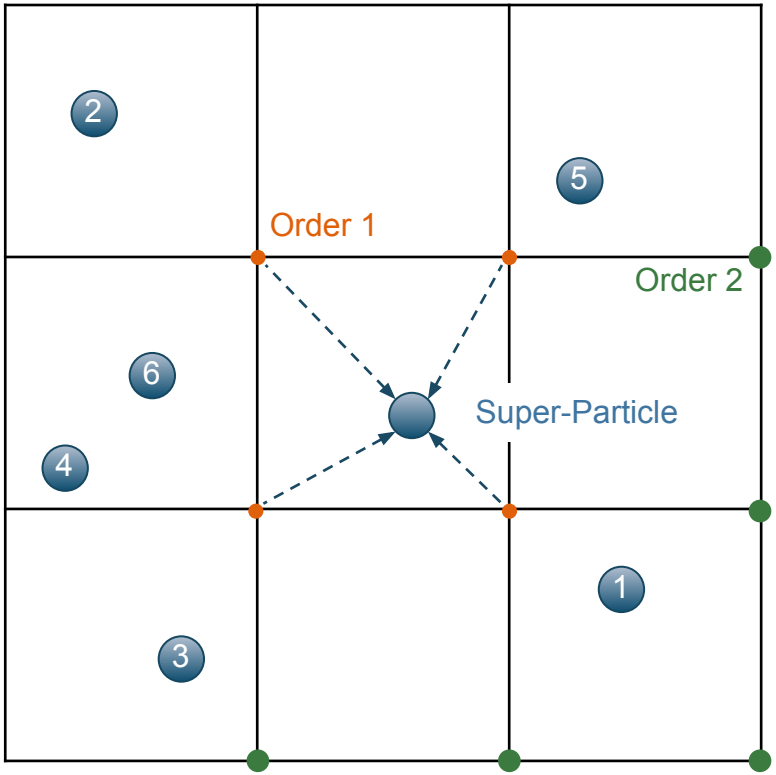


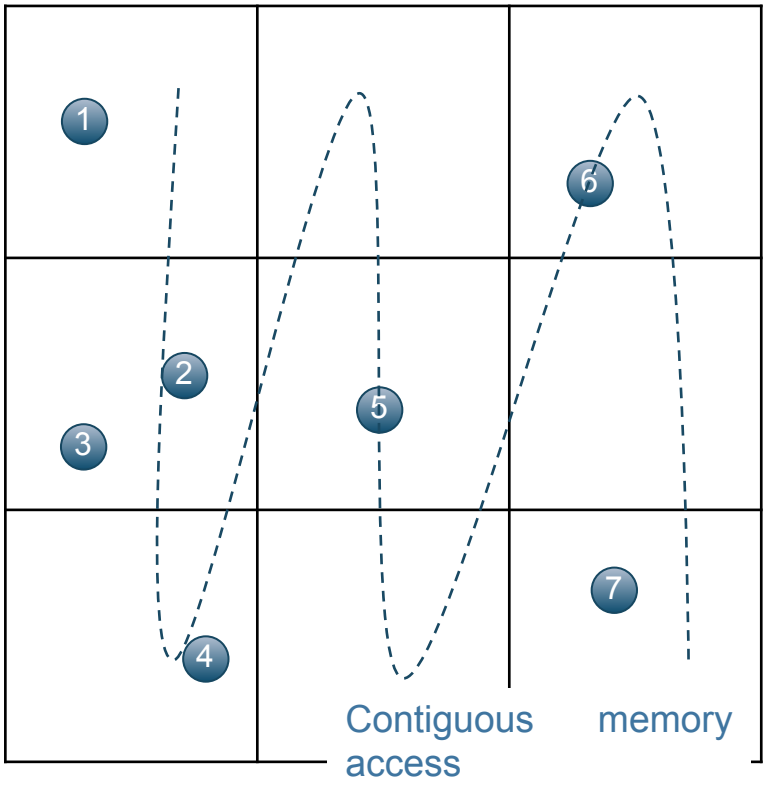
- Current grids: Jx(NCELLS), Jy(NCELLS), Jz (NCELLS)
- Charge(NCELLS)
- Temporary current arrays: Jxcells(8,NCELLS), Jycells(8,NCELLS), Jzcells(8,NCELLS)



Third optimization: the field gathering + particle bin sorting

- Maxwell grids Ex, Ey, Ez, Bx, By, Bz
- Particle cell sorting for cache reuse

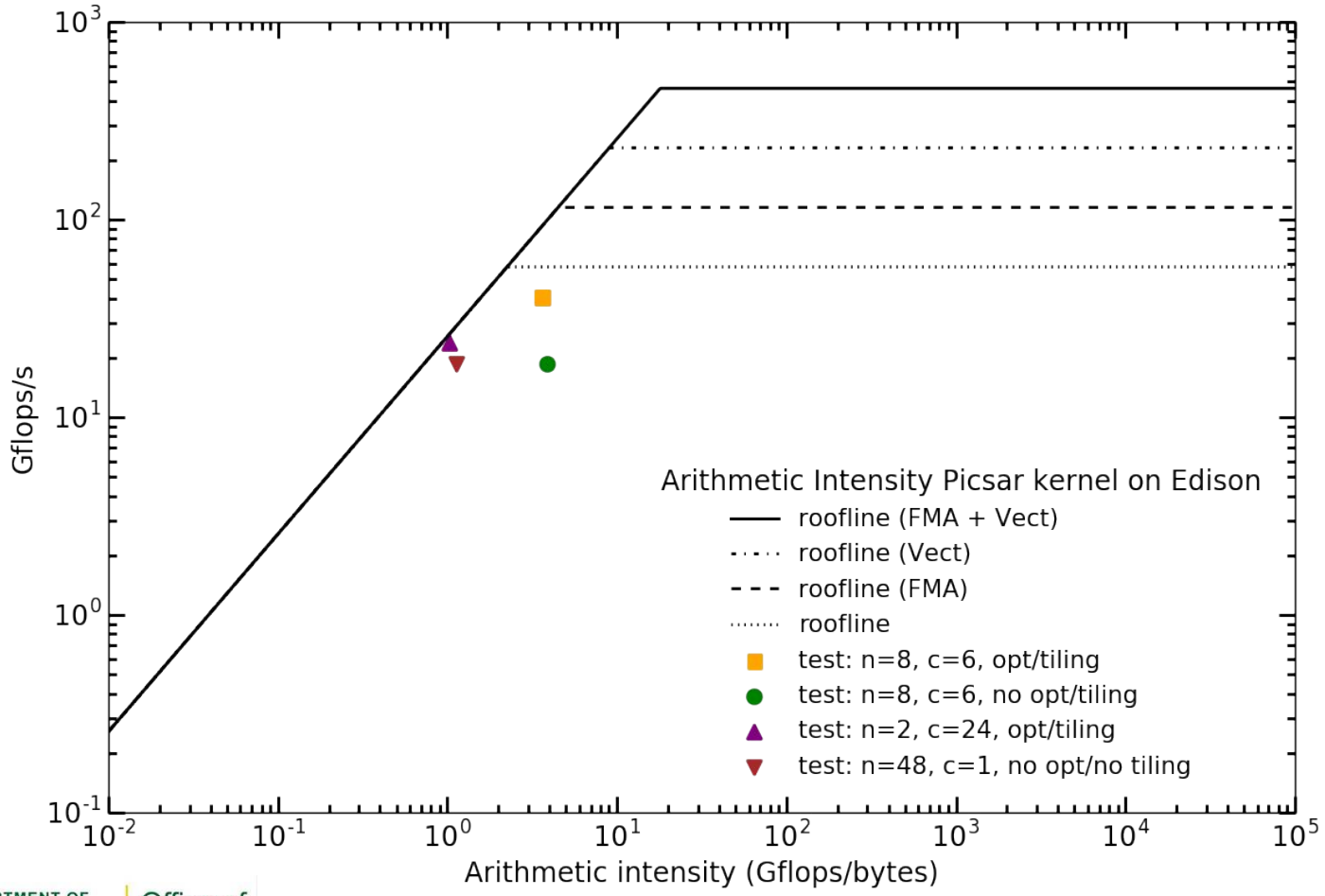




Roofline performance model for the PICSAR kernel on Edison

Opt = optimized subroutines + sorting, **tiling** = with tiling decomposition

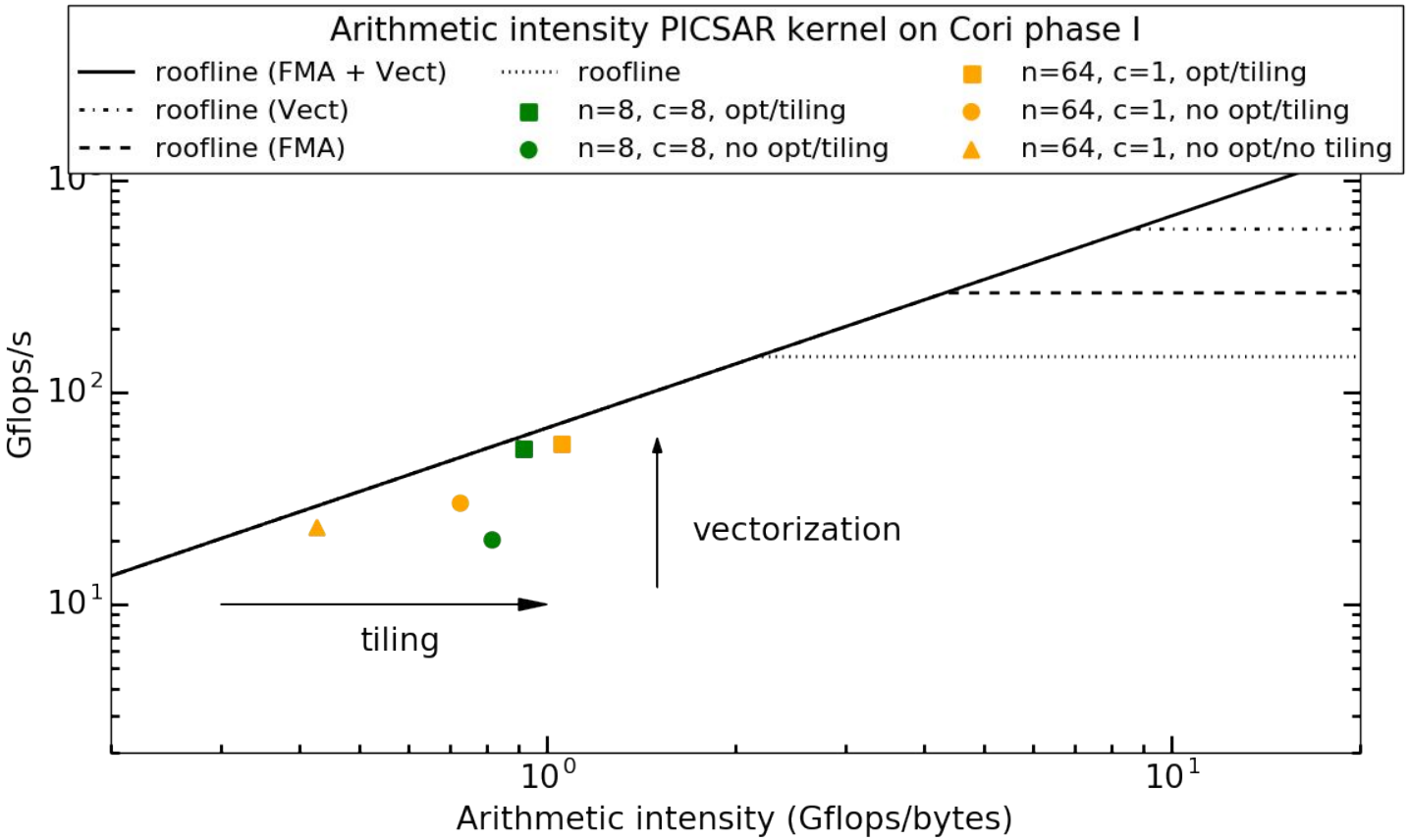
Maxwellian homogeneous plasma, 120x120x120 cells with 20 super-particles per cells, **6x6x6 tiles**
Tile sizes: 2 Mo for particles, 63 Ko for fields



Roofline performance model for the PICSAR kernel on Cori

Opt = optimized subroutines + sorting, **tiling** = with tiling decomposition

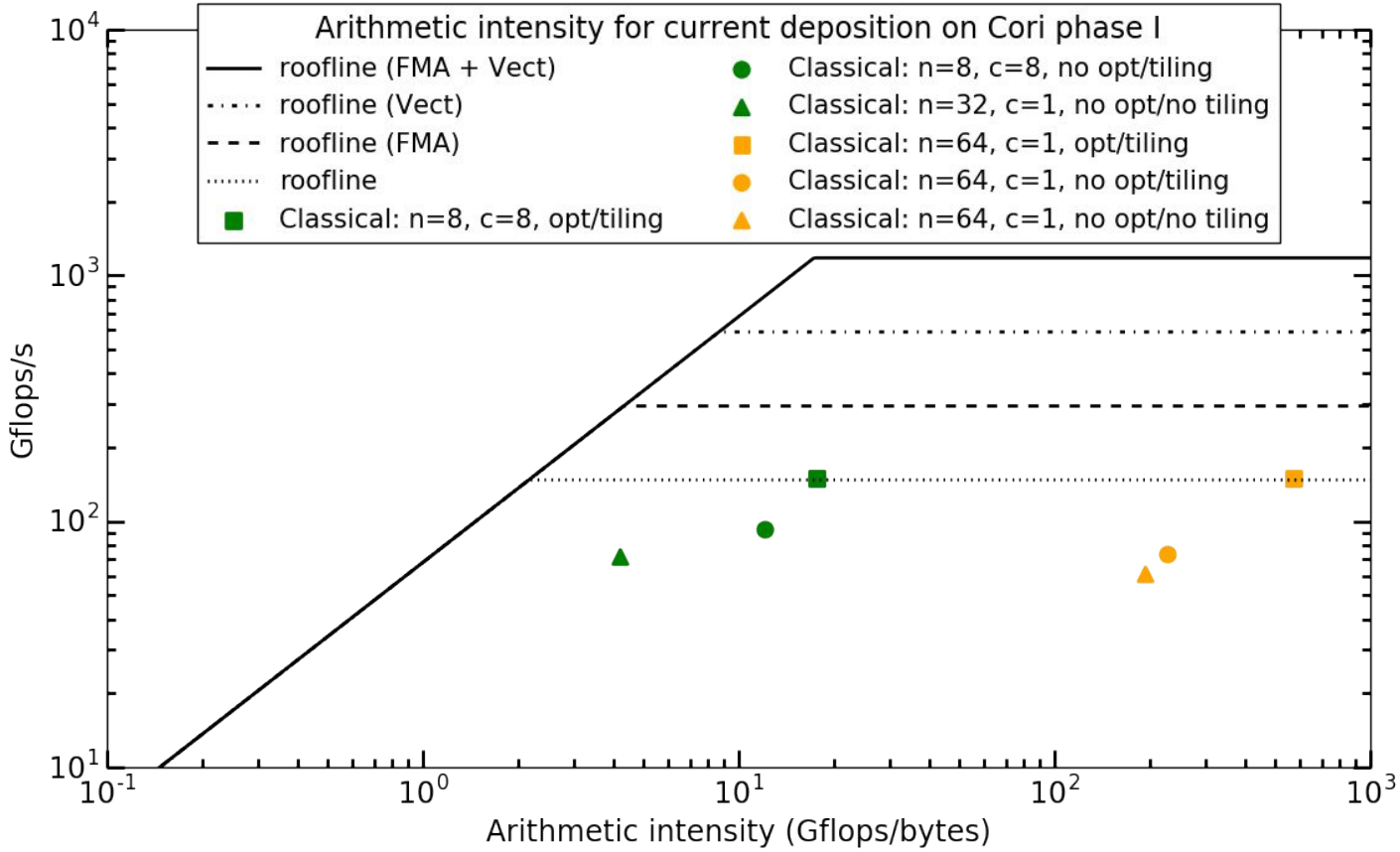
Maxwellian homogeneous plasma, 120x120x120 cells with 20 super-particles per cells, **6x6x6 tiles**
Tile sizes: 2 Mo for particles, 63 Ko for fields



Roofline performance model for current deposition on Cori

Opt = optimized subroutines + sorting, **tiling** = with tiling decomposition

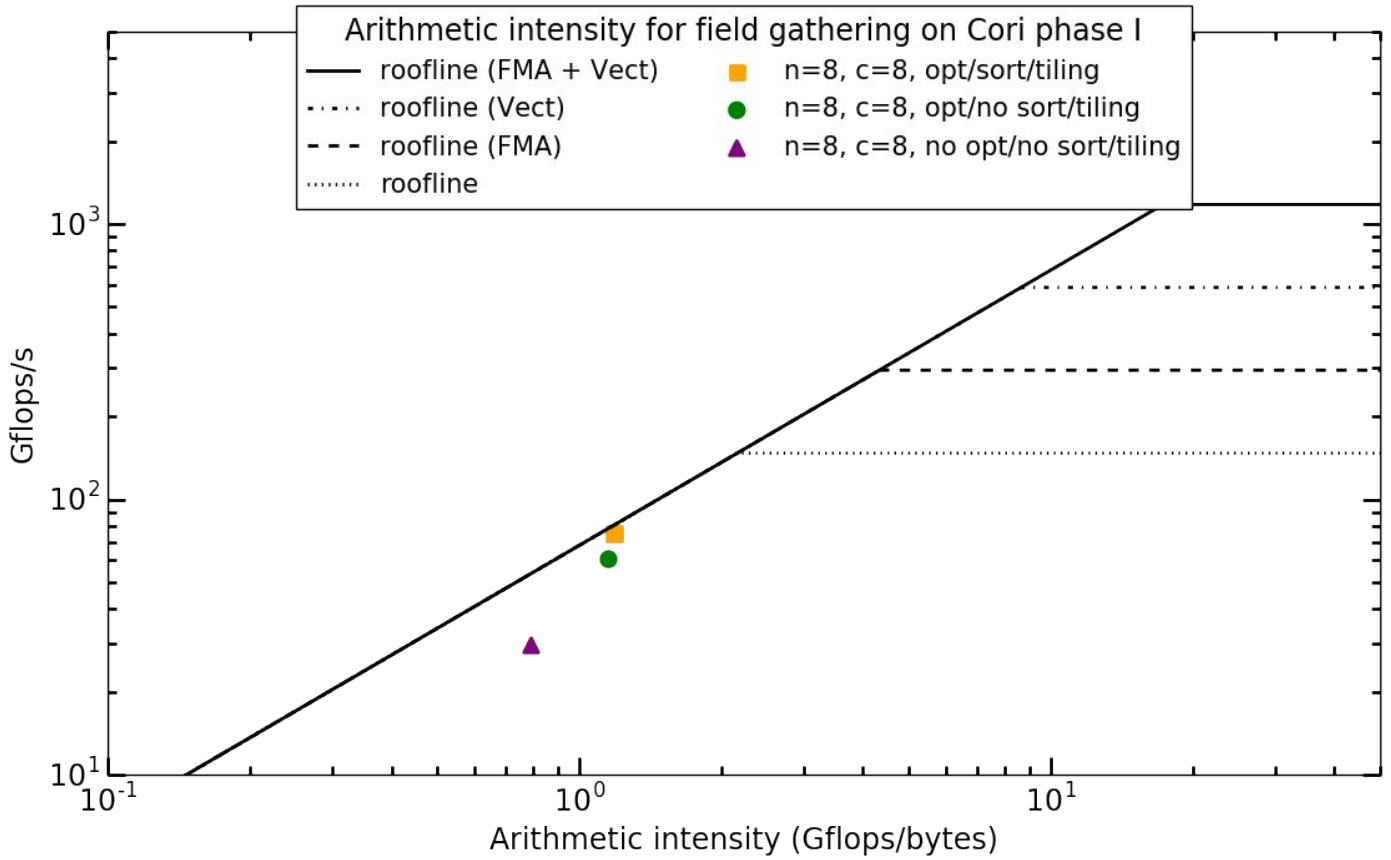
Maxwellian homogeneous plasma, 120x120x120 cells with 20 super-particles per cells, **6x6x6 tiles**
Tile sizes: 2 Mo for particles, 63 Ko for fields



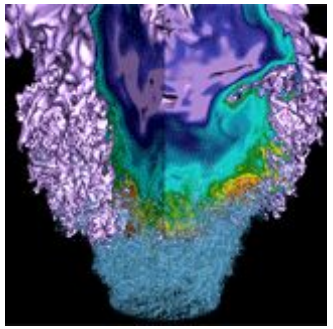
Roofline performance model for field gathering on Cori

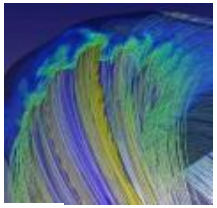
Opt = optimized subroutines, **sort** = with sorting, **tiling** = with tiling decomposition

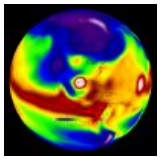
Maxwellian homogeneous plasma, 120x120x120 cells with 20 super-particles per cells, **6x6x6 tiles**
Tile sizes: 2 Mo for particles, 63 Ko for fields

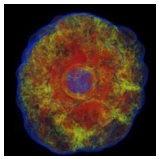


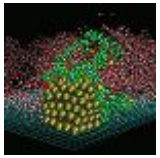
Move Update







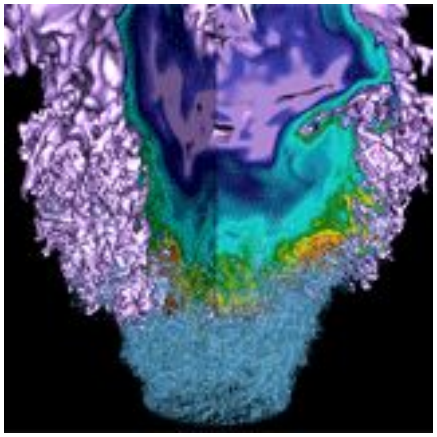


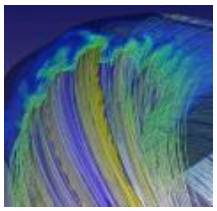


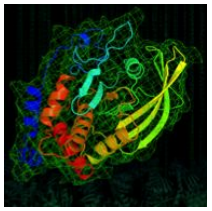
Systems Update

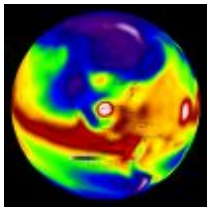
- **Babbage (KNC Testbed) Returns**
- **KNL Whiteboxes being Prepped**
- **Cori Outage (~1 Month Expected) This Summer For Cori-Phase 2 Installation and Integration**

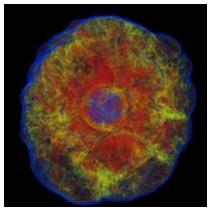
PostDoc Program

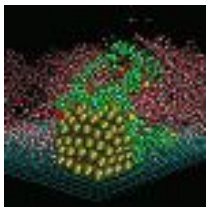












NERSC

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB
Lawrence Berkeley National Laboratory

PostDocs

Taylor Barnes
Quantum ESPRESSO

Brian Friesen
Boxlib

Andrey Ovsyannikov
Chombo-Crunch

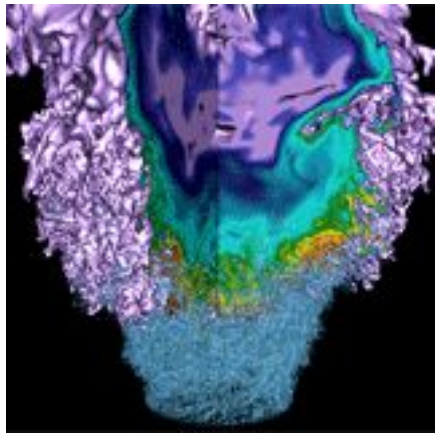
Mathieu Lobet
WARP

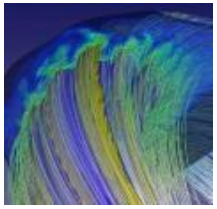
Tuomas Koskela
XGC1

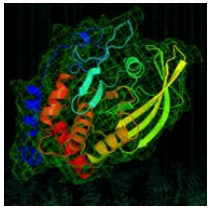
Tareq Malas
EMGeo

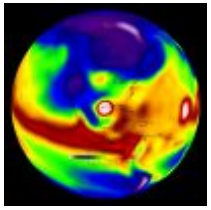
2 Positions Still Open: https://lbl.taleo.net/careersection/jobdetail.ftl?job=81356&lang=en

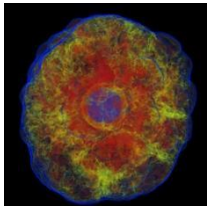
Extra

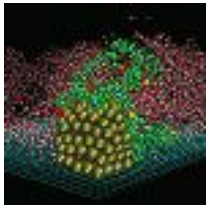












NERSC

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB
Lawrence Berkeley National Laboratory

Cori Regular Wait Before Changes

03/11-03/22/16

	Hours Requested																																																	
Nodes	<1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	48+
1	17	36	53	23	131	47	59	61	34	42	48	48	66	63	91	0.1	56		88		50	23	53	87	78			71			59			58			84				78		76				68	71	86	
2	33	41	37	16	37	13	43	49	37	37	37		52		68		21	65			47				47	70					46			62	62	64	118				46							113	72	
3	28	49	36	0.6	48	58	27			50	63		47								0.0			9.6	52				62								133													
4	8.1	30	34	41	39	30	49		49		33	58	52	66		422	81							21	66											61	113													
5	45		13	8.4	35	242	14				71	72				69																					138													
6	11	50	0.3	37	0.3	0.2	59		0.0				68			429			71						70				91		69						82													
7			0.7	33	3.9	45	59		28	62	68		91						87						65																									
8	4.8	21	26	8.7	44	12	0.0		25	0.6	44	55	0.0		60	3.5	52				62		93	72	66												101													
9	18	27	61	60									75																																					
10	10	14	30		1.1	17	25	3.6	6.0	11		56	37		3.7				5.2		102		93		70												101													
11	0.0	0.0																							58	67																								
12	18	58					42					60	69																																					
13		56																																																
14							66																																											
15	0.0		56				75						66																								120													
16	6.4	29	51	26	24	2.8	33	1.0		40	68		45								43				17							270	441			87	203												442	
17-19	17	59		0.0			68						69														60																							
20-23	37	58	56		61		54				0.3		53																								0.2													
24-31	53	54	22	68	0.1	48	55				57		64				70	63	77		66			73													0.9													
32-47	28	79	71	47	33	12	51		0.0		63	72	73	71		433									69						0.0						91													
48-63	34	9.6	3.2	0.3					66		35																																							
64-127	15	28	67	45	44		75	64	9.4		47		73			0.6									124								2.2			4.1	63													
128-255	22	16	68		5.1	50	60		66		62		62																							10	7.6													
256-511	24	6.4	37	27		241	60		4.3	71	72		25								71				262																									
512-1023	43	3.1								73																																								
1024-1535	75	19		85	123																																													

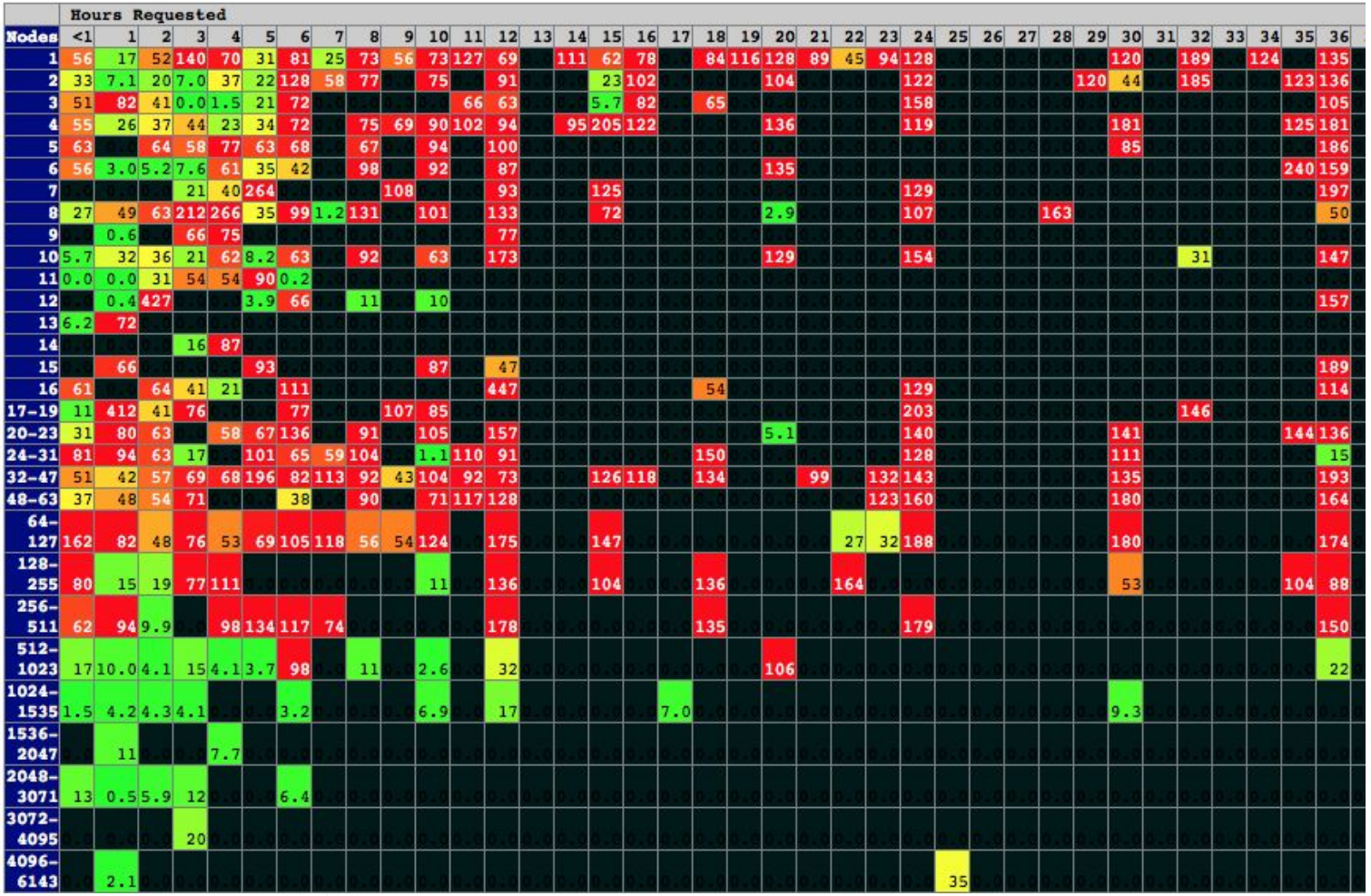
Cori Regular Wait After Changes

04/11-04/20/16

Wait time for shorter, smaller jobs have decreased

Edison Regular Wait Before Changes

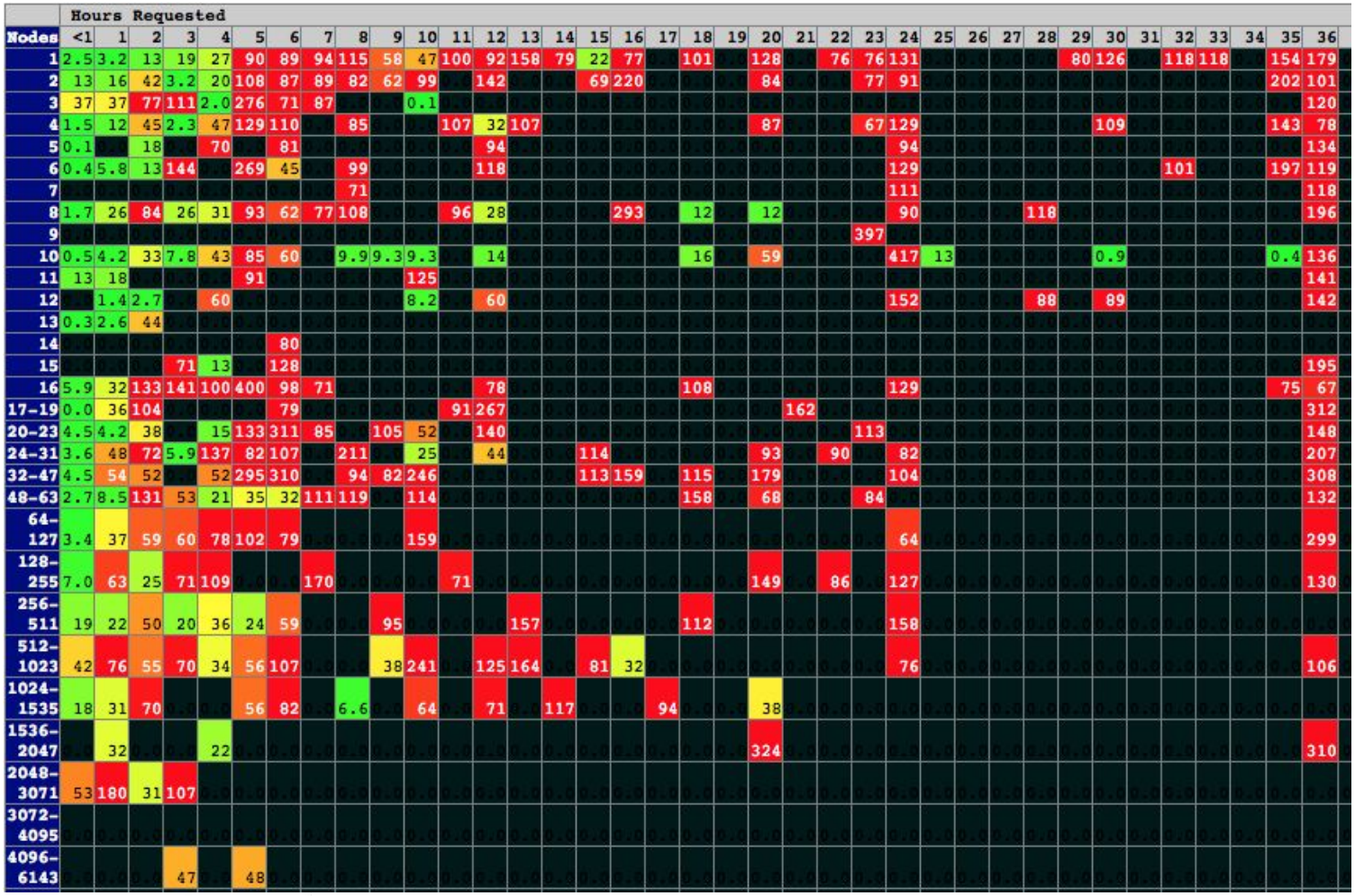
03/11-03/22/16



Nodes	Hours Requested																																		
	<1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
1	56	17	52	140	70	31	81	25	73	56	73	127	69		111	62	78		84	116	128	89	45	94	128						120		189		124		135
2	33	7.1	20	7.0	37	22	128	58	77		75		91			23	102				104				122					120	44	185			123	136
3	51	82	41	0.0	1.5	21	72					66	63			5.7	82		65						158											105
4	55	26	37	44	23	34	72		75	69	90	102	94		95	205	122				136				119						181				125	181
5	63		64	58	77	63	68		67		94		100																		85					186
6	56	3.0	5.2	7.6	61	35	42		98		92		87								135														240	159
7				21	40	264				108			93			125									129											197
8	27	49	63	212	266	35	99	1.2	131		101		133			72					2.9				107				163							50
9		0.6		66	75								77																							
10	5.7	32	36	21	62	8.2	63		92		63		173								129				154								31				147
11	0.0	0.0	31	54	54	90	0.2																														
12		0.4	427			3.9	66		11		10																										157
13	6.2	72																																			
14				16	87																																
15		66				93					87		47																								189
16	61		64	41	21		111						447						54						129											114
17-19	11	412	41	76			77			107	85														203								146				
20-23	31	80	63		58	67	136		91		105		157								5.1				140						141				144	136
24-31	81	94	63	17		101	65	59	104		1.1	110	91						150						128						111					15
32-47	51	42	57	69	68	196	82	113	92	43	104	92	73			126	118		134			99		132	143						135					193
48-63	37	48	54	71			38		90		71	117	128											123	160						180					164
64-127	162	82	48	76	53	69	105	118	56	54	124		175			147							27	32	188						180					174
128-255	80	15	19	77	111						11		136			104			136				164								53				104	88
256-511	62	94	9.9		98	134	117	74					178						135						179											150
512-1023	17	10.0	4.1	15	4.1	3.7	98		11		2.6		32								106															22
1024-1535	1.5	4.2	4.3	4.1			3.2				6.9		17					7.0													9.3						
1536-2047		11			7.7																																
2048-3071	13	0.5	5.9	12			6.4																														
3072-4095				20																																	
4096-6143		2.1																								35											

Edison Regular Wait After Changes

04/11-04/20/16



	Hours Requested																																		
Nodes	<1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
1	2.5	3.2	13	19	27	90	89	94	115	58	47	100	92	158	79	22	77		101		128		76	76	131					80	126		118	118		154	179
2	13	16	42	3.2	20	108	87	89	82	62	99		142			69	220				84			77	91										202	101
3	37	37	77	111	2.0	276	71	87			0.1																									120
4	1.5	12	45	2.3	47	129	110		85			107	32	107							87			67	129					109					143	78
5	0.1		18		70		81						94												94												134
6	0.4	5.8	13	144		269	45		99				118												129								101			197	119
7									71																111												118
8	1.7	26	84	26	31	93	62	77	108			96	28				293		12		12				90				118								196
9																								397													
10	0.5	4.2	33	7.8	43	85	60		9.9	9.3	9.3		14						16		59				417	13				0.9					0.4	136
11	13	18				91					125																										141
12		1.4	2.7		60						8.2		60												152				88		89						142
13	0.3	2.6	44																																		
14							80																														
15				71	13		128																														195
16	5.9	32	133	141	100	400	98	71					78						108						129											75	67
17–19	0.0	36	104				79					91	267									162															312
20–23	4.5	4.2	38		15	133	311	85		105	52		140											113													148
24–31	3.6	48	72	5.9	137	82	107		211		25		44			114					93		90		82												207
32–47	4.5	54	52		52	295	310		94	82	246					113	159		115		179				104												308
48–63	2.7	8.5	131	53	21	35	32	111	119		114								158		68			84													132
64–127	3.4	37	59	60	78	102	79				159														64												299
128–255	7.0	63	25	71	109			170				71									149		86		127												130
256–511	19	22	50	20	36	24	59			95			157						112						158												
512–1023	42	76	55	70	34	56	107			38	241		125	164		81	32								76												106
1024–1535	18	31	70			56	82		6.6		64		71		117			94			38																
1536–2047		32			22																324																310
2048–3071	53	180	31	107																																	
3072–4095																																					
4096–6143				47		48																															