Intel Ivybridge vs. AMD Opteron: performance and power implications

HEPiX Spring 2014

Tony Wong (BNL)

Acknowledgements

Thanks to Chris Hollowell and Shuwei Ye (BNL) for doing most of the work. We also thank Dell, HP and Penguin Computing for making the hardware available for us to evaluate.

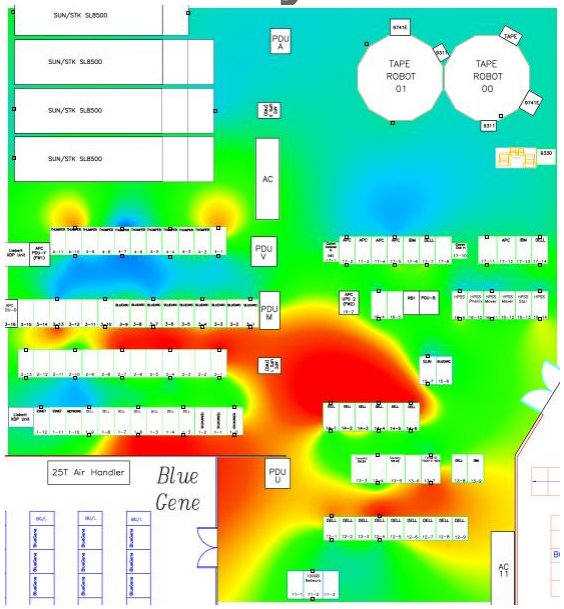
Background

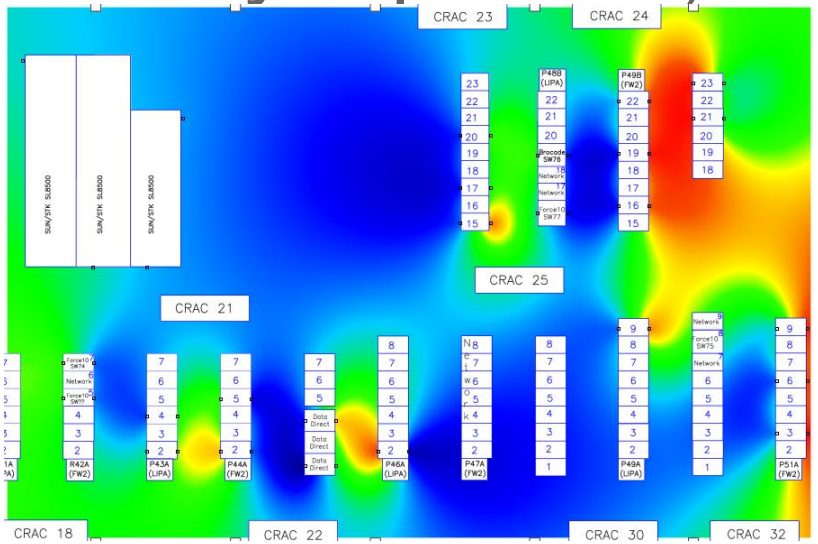
- Annual purchase cycle for RHIC-ATLAS Computing Facility (RACF) at BNL
- Data Center constraints (space, power and cooling)
- Vendor constraints (few AMD options)
- Experimental requirements (computing and storage)
- Other constraints
  - Migration to 10 Gb connectivity for RHIC
  - Compatible with Infiniband solutions?

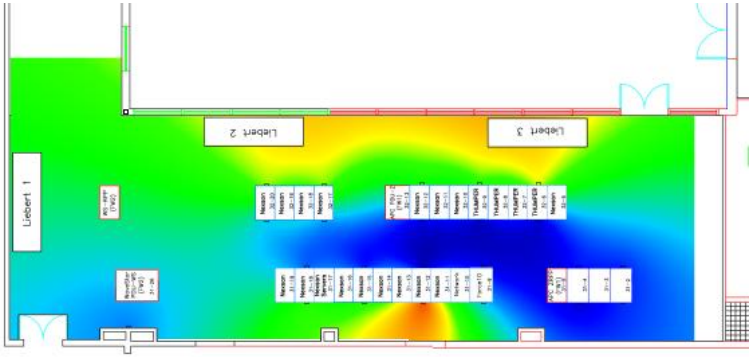
Data Center constraints

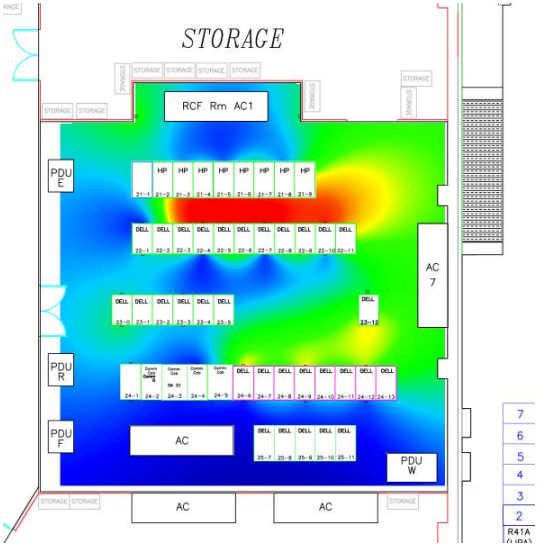
- Space
  - Approximately ~70% of 15,000 ft^2 (~1,400 m^2) data center taken
  - Remaining floor space requires power/cooling upgrades
- Power
  - 2 MW of usable UPS-backed power
  - Current usage ~1.1 MW (55% of maximum)
  - Cannot go much above ~80% due to configuration inefficiencies (ie, pdu-level redundancy for critical components)
- Cooling
  - 2 MW capacity
  - Few CRAC units on UPS power

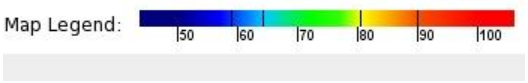
Facility Heatmap (from Synapsense)



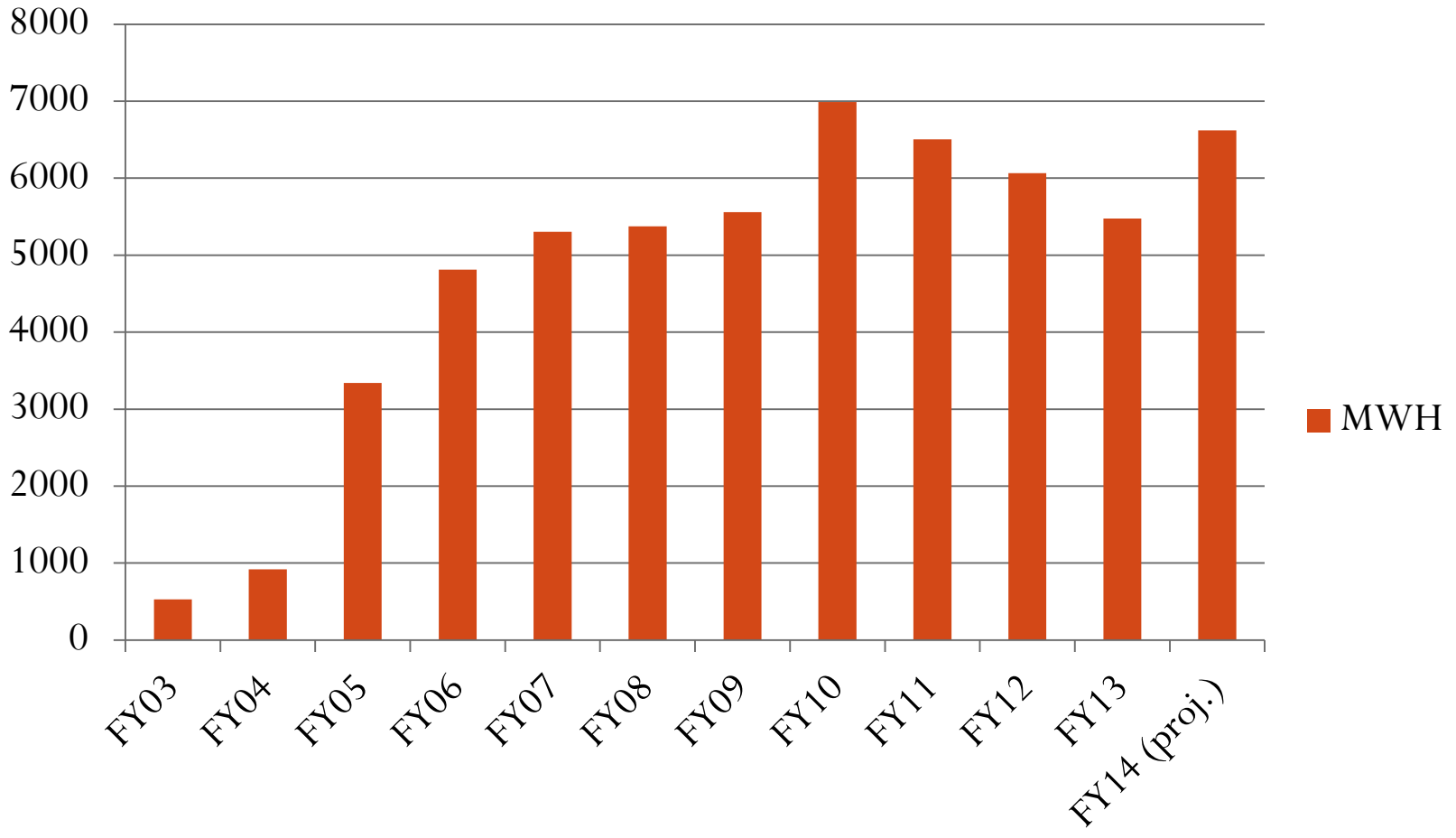




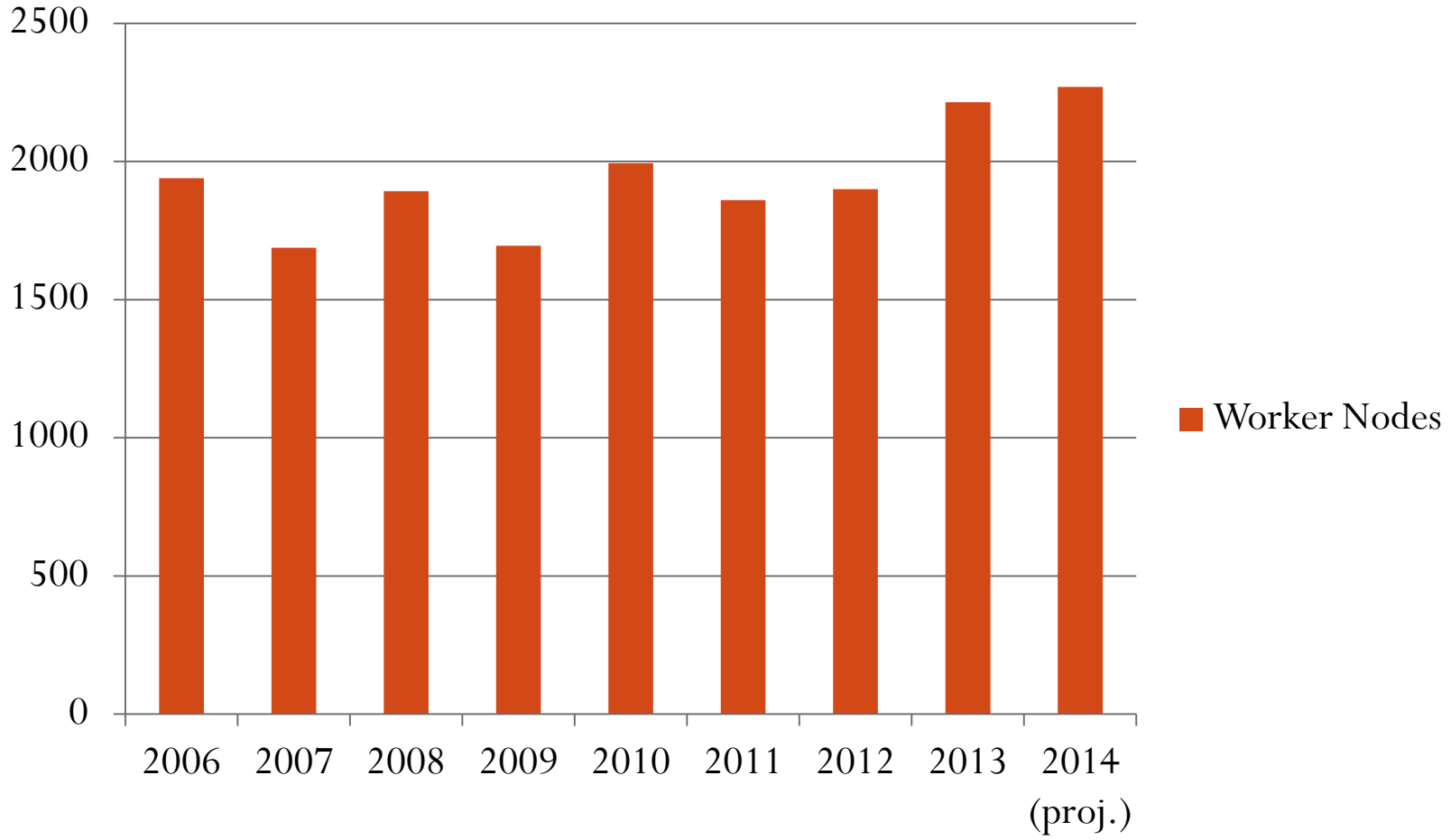




RACF Historical Power Usage



Historical Worker Node Count



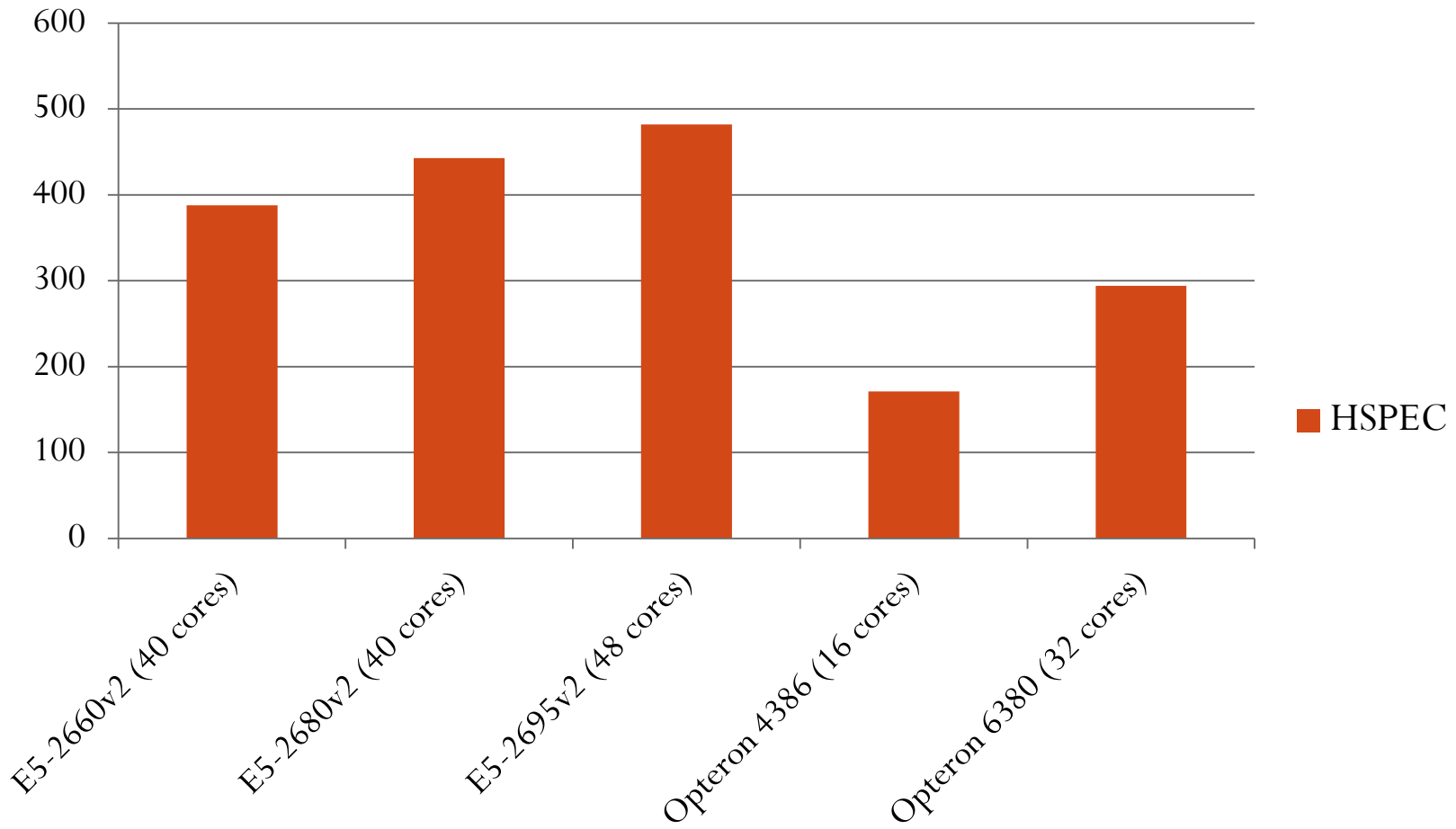
Experimental Requirements

- RHIC
  - Disk-heavy worker nodes (2-U, dual-socket, multiple large SATA drives)
  - 2+ GB of RAM per physical (Opteron) or logical (Ivybridge) core
  - 10 Gb connectivity (new for 2014)
- USATLAS
  - Disk-light worker nodes (1-U, dual-socket, multiple small SATA drives)
  - 2+ GB of RAM per physical (Opteron) or logical (Ivybridge) core

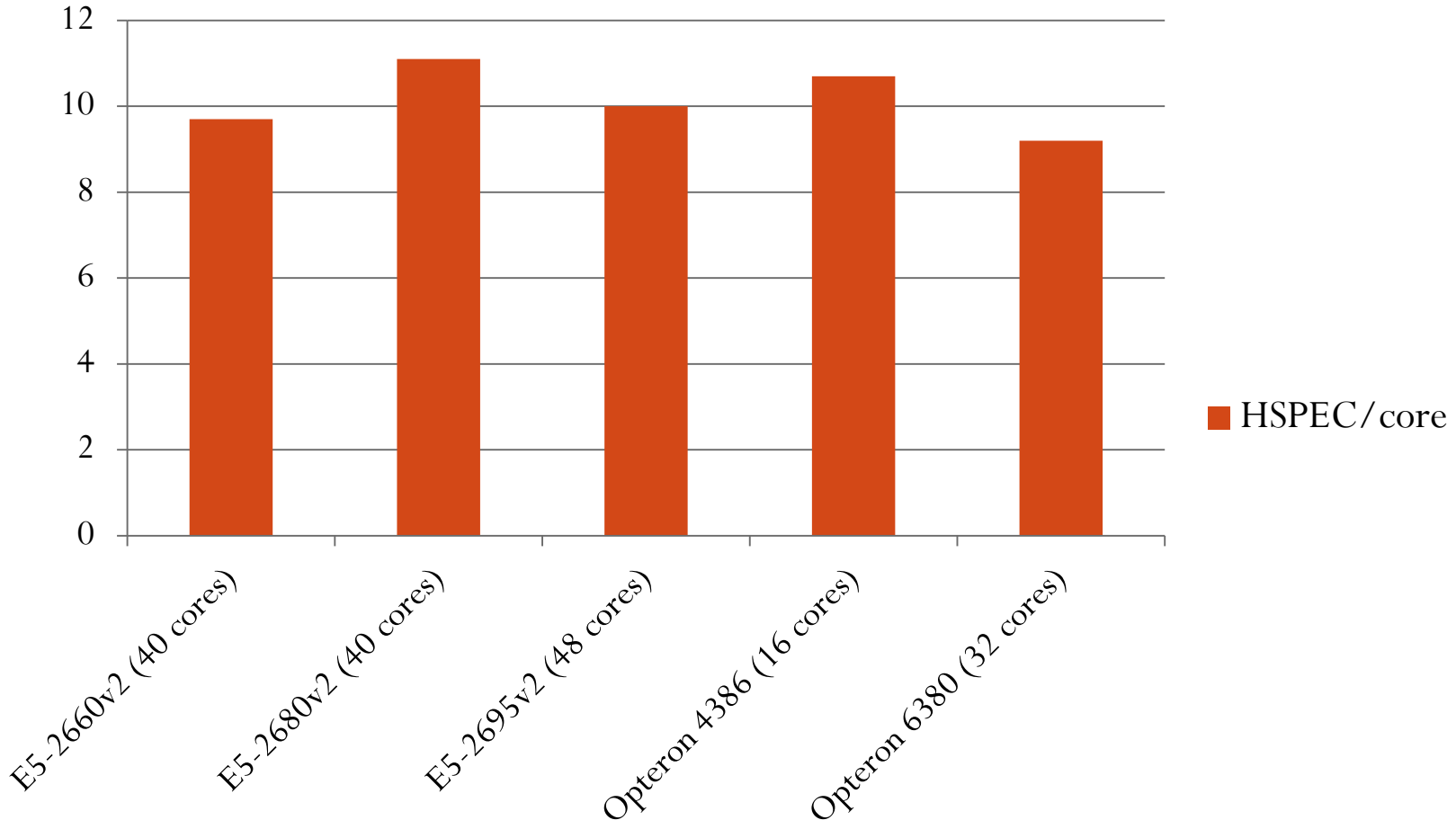
Hardware Evaluation

- CPU
  - E5-2695v2 (12 physical or 24 logical cores – Ivybridge 115 W TDP)
  - E5-2680v2 (10 physical or 20 logical cores – Ivybridge 115 W TDP)
  - E5-2660v2 (10 physical or 20 logical cores – Ivybridge 95 W TDP)
  - Opteron 4386 (8 physical cores)
  - Opteron 6380 (16 physical cores)
- Storage
  - 4 x 2 TB or 8 x 1 TB SATA drives (1-U)
  - 12 x 4 TB SATA drives (2-U)
- Vendors
  - HP (Ivybridge)
  - Dell (Ivybridge)
  - Penguin Computing (Opteron)
- Want to validate HSPEC results with real-life applications

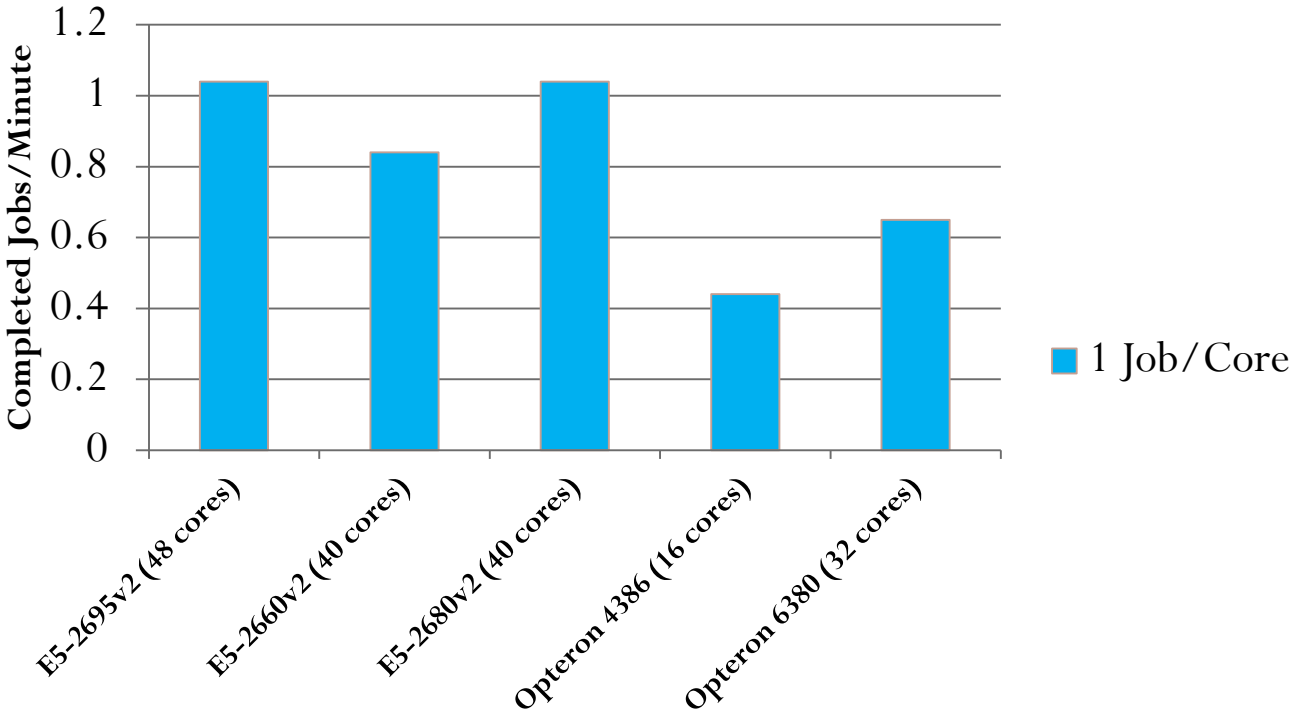
HSPEC


600
500
400
300
200
100
0
E5-2660v2 (40 cores)
E5-2680v2 (40 cores)
E5-2695v2 (48 cores)
Opteron 4386 (16 cores)
Opteron 6380 (32 cores)
HSPEC

HSPEC per core

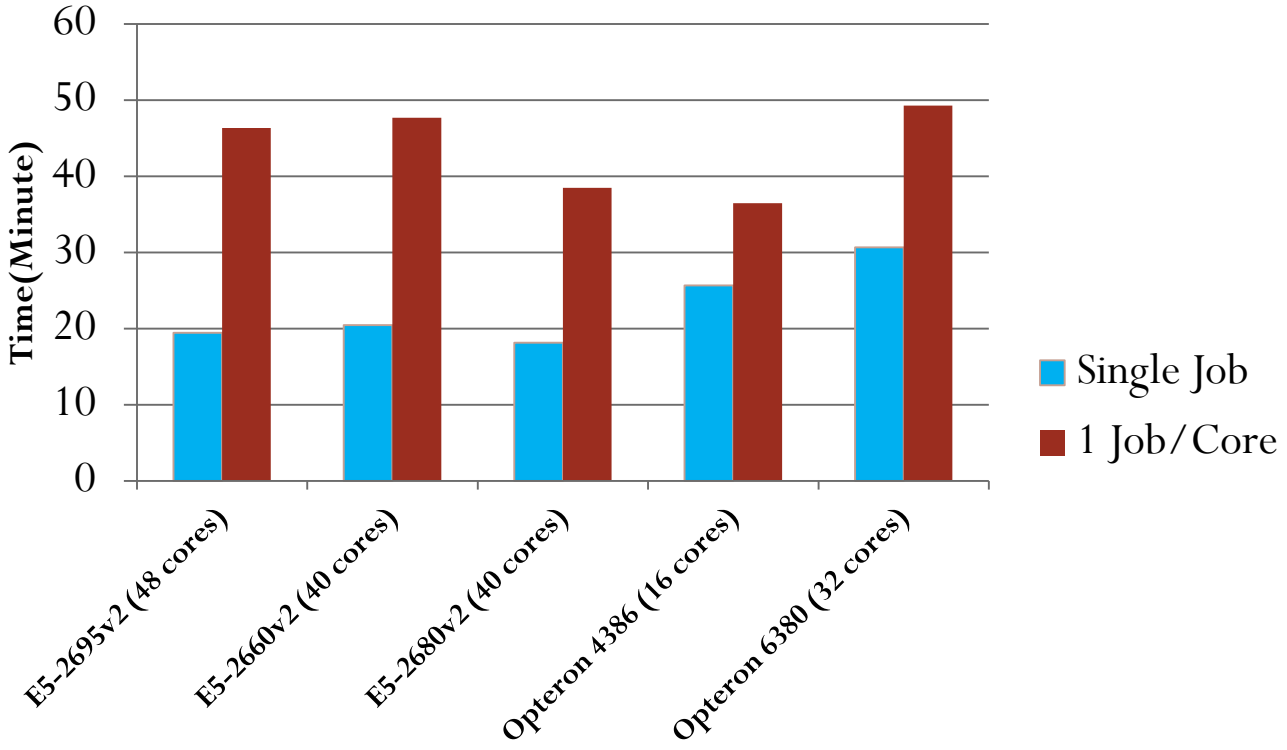


ATLAS Full Simulation



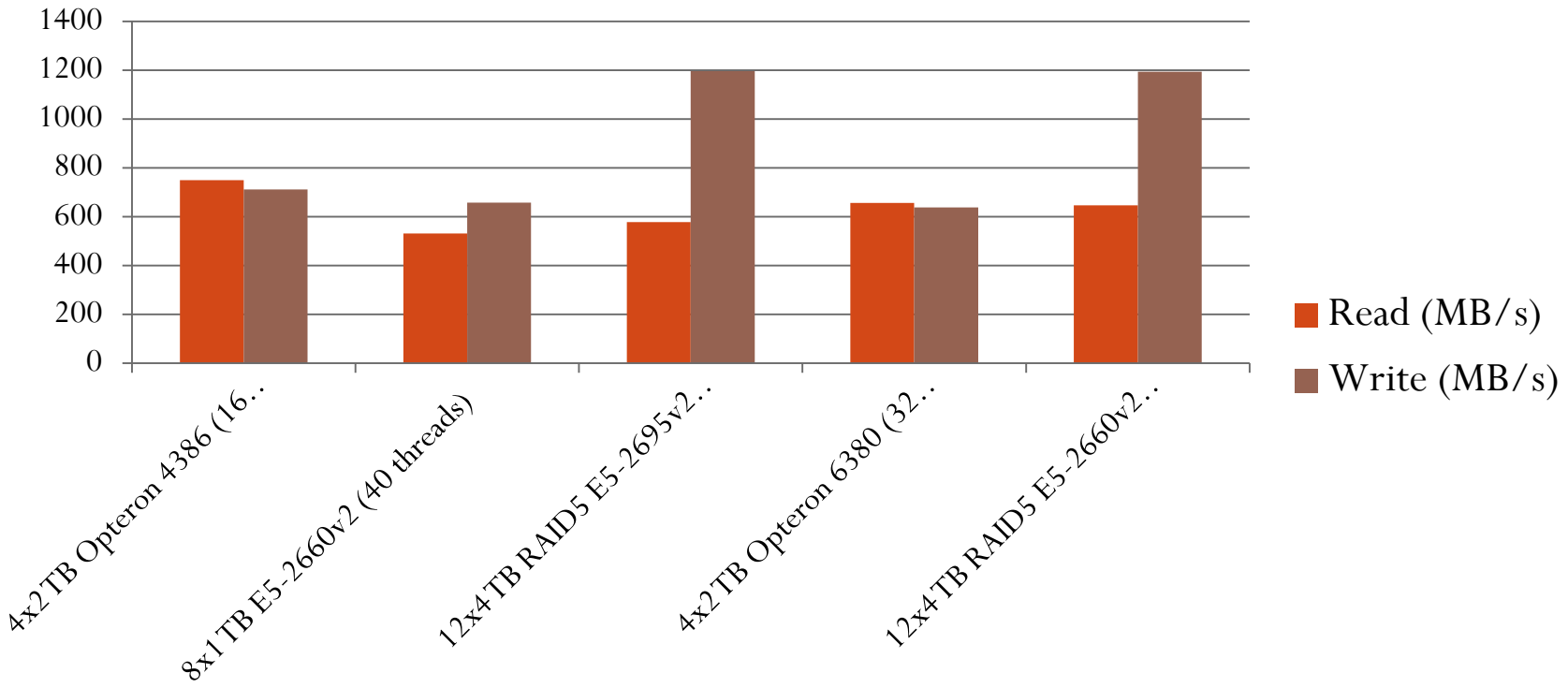
- Ivybridge at higher Thermal Design Power (TDP) perform better than those at lower TDP (ie, E5-2695v2 vs. E5-2660v2)
- Opteron 6380's throughput is ~50% higher than 4386's
- Opteron 6380's throughput is ~63% of E5-2680v2's and ~77% of E5-2660v2's

ATLAS Full Simulation



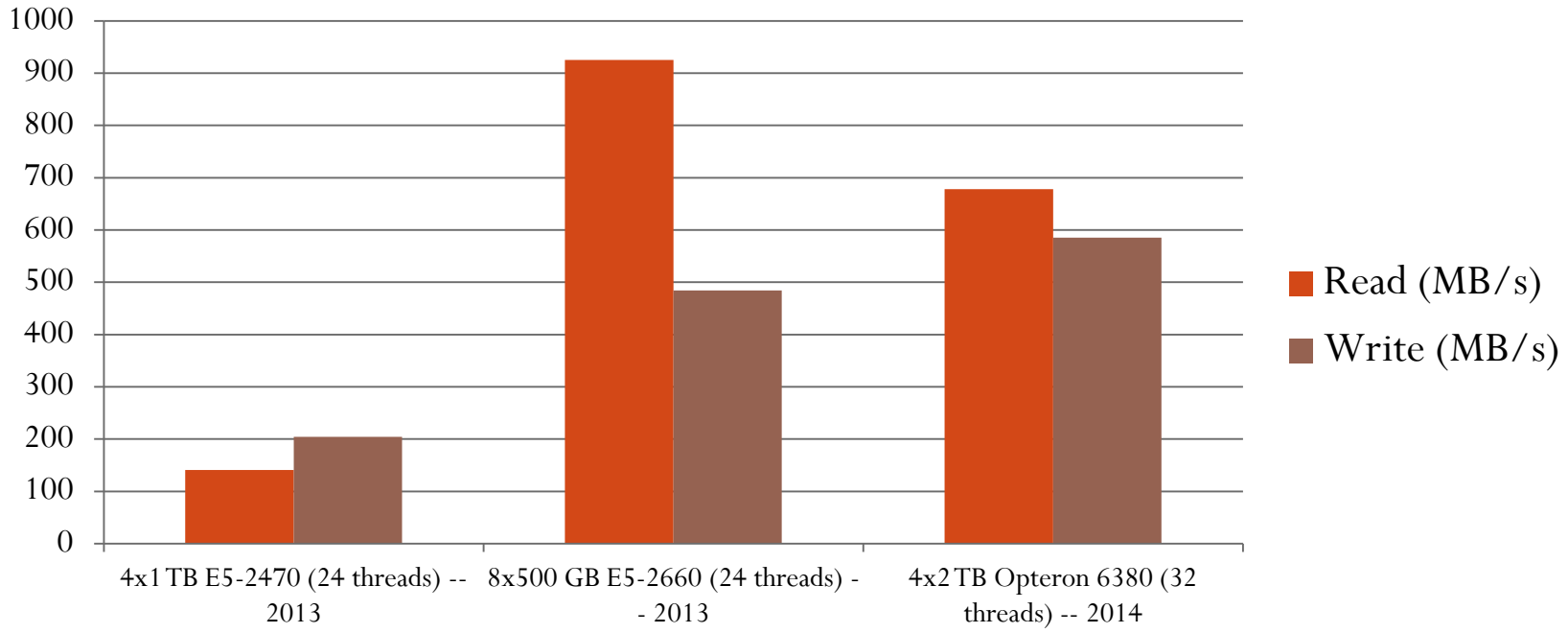
- Multi-job throughput is significantly worse for Ivybridge's when compared to single-job performance due to hyperthreading

Local Disk (random) I/O with Bonnie++ (aggregate)



- Mostly dependent on # of drives, quality of drives and controller
- Controller impacted 8x1 TB E5-2660v2 results negatively
- More (high-quality) drives improves I/O but increases cost/server

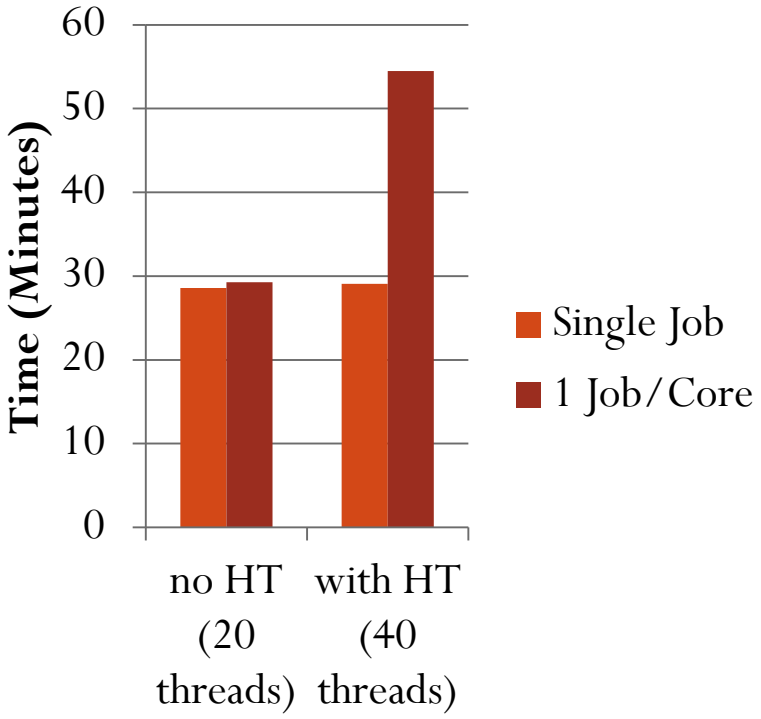
Local Disk (Random) I/O with Bonnie++ (aggregate)

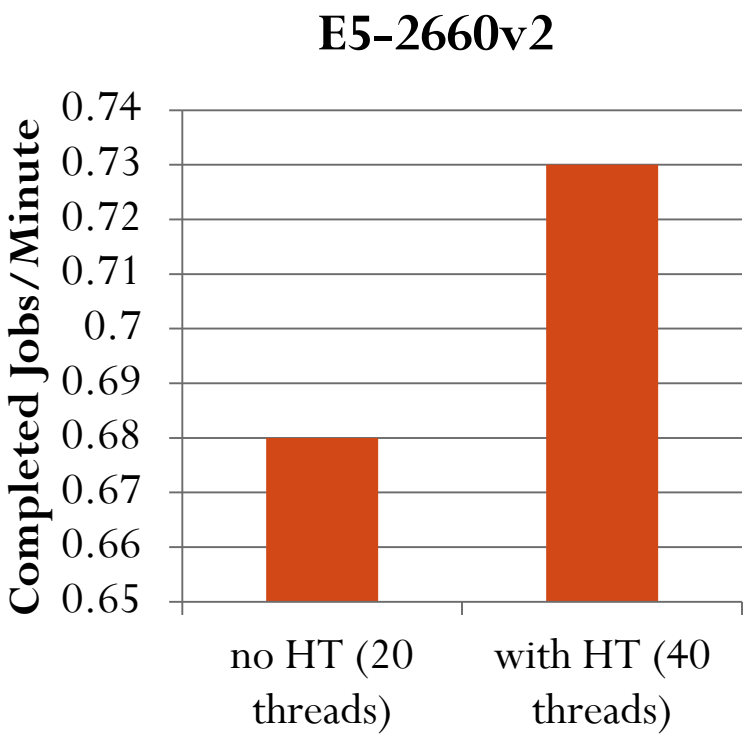


- Test of Sandybridge cpu (2013) shows the superior performance of 8-disk configuration vs. 4-disk configuration
- In 2014, results show 12-drive I/O is better than 8-drive I/O —> expect 8-drive to have better I/O performance than 4-drive configuration
- Note 4-drive write performance in 2014 is already **HIGHER** than 8-drive write performance in 2013 and much higher than 4-drive configuration in 2013
- Variations in SATA link rate (3 Gbps and 6 Gbps) accentuate results but does not alter general trends

The Effect of HT

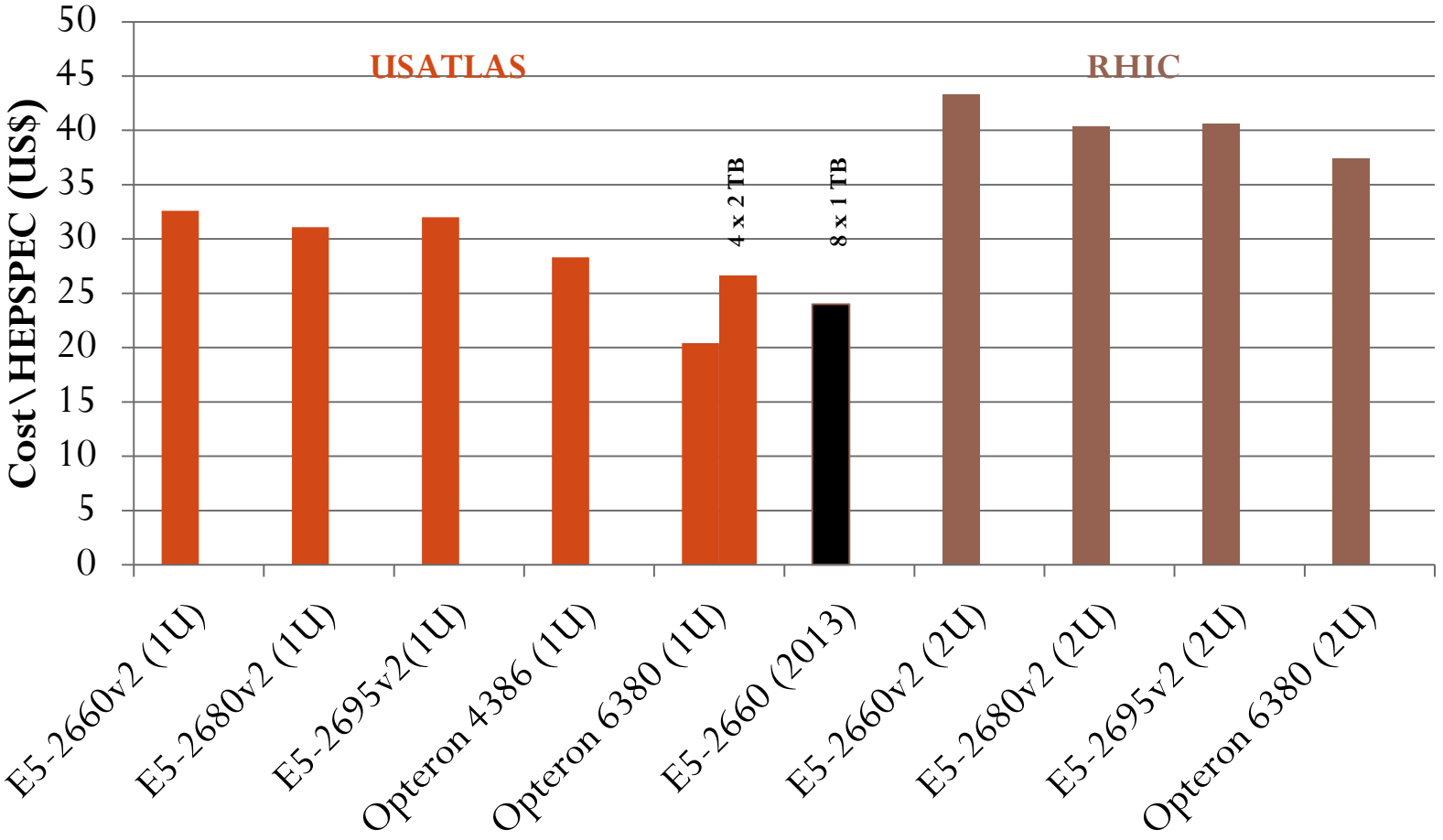
- HSPEC boost of 18-27% with HT enabled
- I/O scales linearly with HT disabled



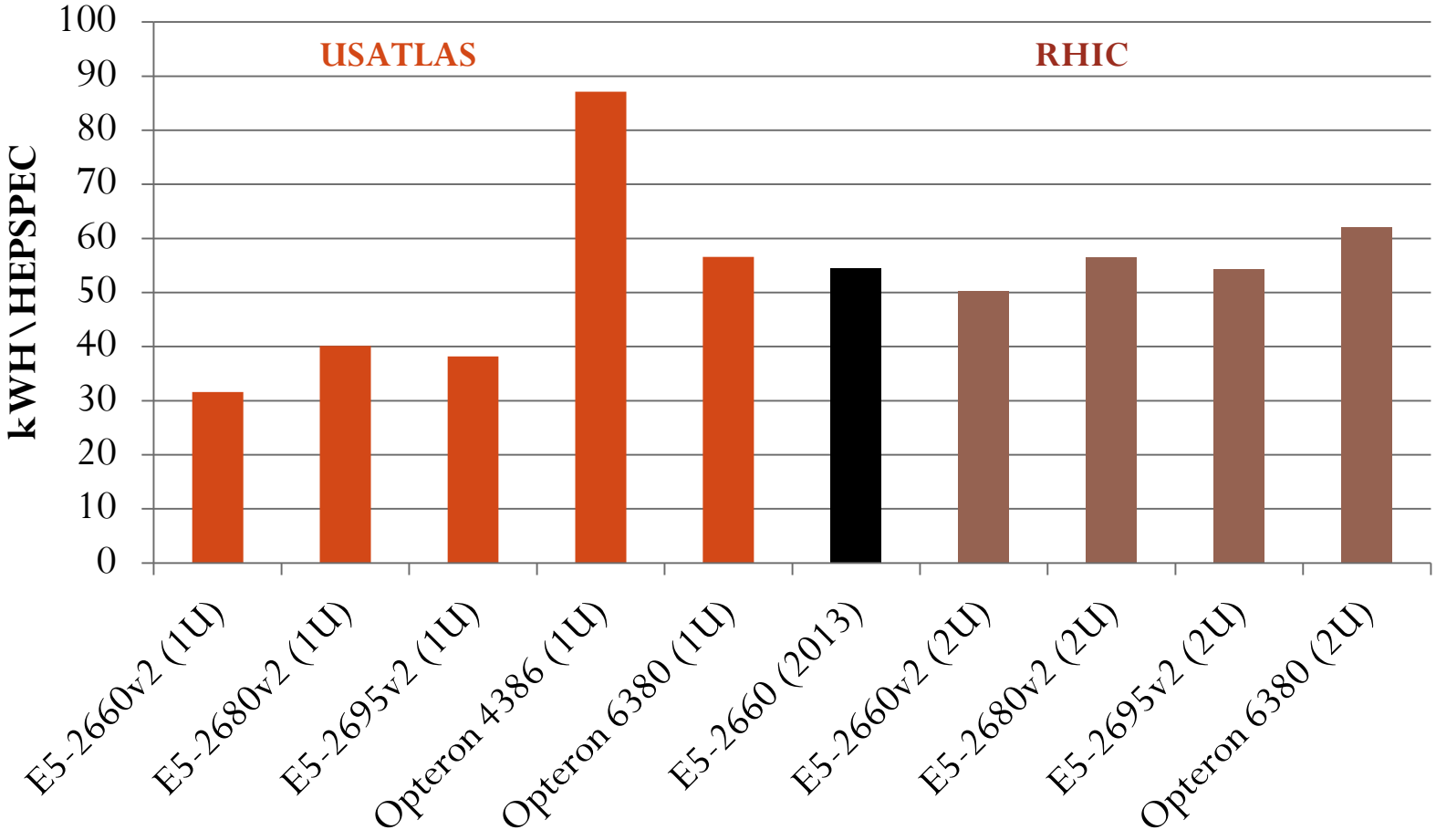


- HT boosts ATLAS job throughput by ~15%
- Turning off HT (and cutting back on RAM) increases the price competitiveness of Ivybridge by ~5% --not enough to overcome the price-performance advantages of the Opteron platforms

Cost per HSPEC for each configuration



Power Usage over 5 years



Cost Breakdown and Power Considerations

CPU	Cores	Server Cost	10 GbE	Composite List Price	Power Usage	5-yr Power Cost
E5-2660v2 (1U)	40	$12,639	$473	$13,289 (128 GB RAM, no 10 GbE)	280 W	$736
E5-2680v2 (1U)	40	$13,759	$473	$14,409 (128 GB RAM, no 10 GbE)	406 W	$1,067
E5-2695v2 (1U)	48	$15,439	$473	$16,089 (128 GB RAM, no 10 GbE)	420 W	$1,104
Opteron 6380 (1U)	32	$5,985(4x2) $7,820(8x1)	Incl.	$6,635 (96 GB RAM, no 10 GbE) $8,470 (96 GB RAM, no 10 GbE)	380 W	$999
E5-2660v2 (2U)	40	$16,790	$473	$17,263	445 W	$1,169
E5-2680v2 (2U)	40	$17,910	$473	$18,383	570 W (est.)	$1,498
E5-2695v2 (2U)	48	$19,590	$473	$20,063	598 W	$1,572
Opteron 6380 (2U)	32	$10,980	Incl.	$10,980	417 W (est.)	$1,096

- List prices (in US dollars) for servers (current as of May 7, 2014)
- Cores assume dual-socket servers (no hyperthreading on Opteron)
- Power cost based on BNL historical average ~6 cents/kwh
- RAM upgrade costs $325 for each incremental 16 GB

Procurement Guidance Summary

CPU	Composite List Price	kHSPEC	Computing cores	Storage (TB)	Power (kW)	Space in ft^2 (Racks)
E5-2660v2 (1U)	$13,289	21.0	2,160	432	15	34 (2)
E5-2680v2 (1U)	$14,409	21.7	1,960	392	20	34 (2)
E5-2695v2 (1U)	$16,089	21.2	2,112	352	18	34 (2)
Opteron 6380 (1U)	$6,635(4x2) $8,470(8x1)	31.8 25.0	3,456 2,720	864 680	41 32	68 (4) 51 (3)
E5-2660v2 (2U)	$17,263	30.3	3,120	3,744	35	85 (5)
E5-2680v2 (2U)	$18,383	32.3	2,920	3,504	42	85 (5)
E5-2695v2 (2U)	$20,063	32.3	3,216	3,216	40	85 (5)
Opteron 6380 (2U)	$10,980	35.9	3,904	5,856	51	136 (8)

- Composite price (in US dollars) includes server, RAM upgrade (USATLAS) and 10 GbE (RHIC)
- Rightmost five columns normalized to a fixed, hypothetical budgetary constraint
- Historical 25% discount **NOT** applied to composite cost
- Final FY2013 final prices were $5.7k/server (USATLAS) and $7.3k/server (RHIC)

Effect of 2014 acquisitions

- Facility infrastructure
  - Net power usage increase under ~100 kW
  - Net footprint increases up to ~200 ft^2 (~19 m^2)
  - Additional infrastructure (CRAC units and PDU's) installed
- Cost/worker node
  - Minimal ~3% increase for RHIC due to10 Gb connectivity
  - Ivybridge is pricy compared to Sandybridge and Opteron, even after dropping core count/socket and taking a historical ~25% discount
  - Optional memory upgrade increases cost 5-10%

Implications for the future

- With limited space, power and cooling until ~2020, several trends developing:
  - De-emphasize core count to optimize local disk I/O
  - Throughput more important than raw cpu performance
  - Reduce power footprint
- Haswell to be released late in 2014
  - To be marketed as E5-26xxv3 series (up to 14 cores?)
  - Cannot time a FY14 procurement with Haswell release in hopes of a price drop for Ivybridge
  - Reported TDP goes up to 160 W – is that a bad omen?
- Clouds and decreasing sales volume turning servers into a niche (expensive) market for hardware makers—consolidation among hardware brands a concern. Expand pool of acceptable brands?

Back-up slides

Local Disk (Random) I/O with Bonnie ++ (per core)