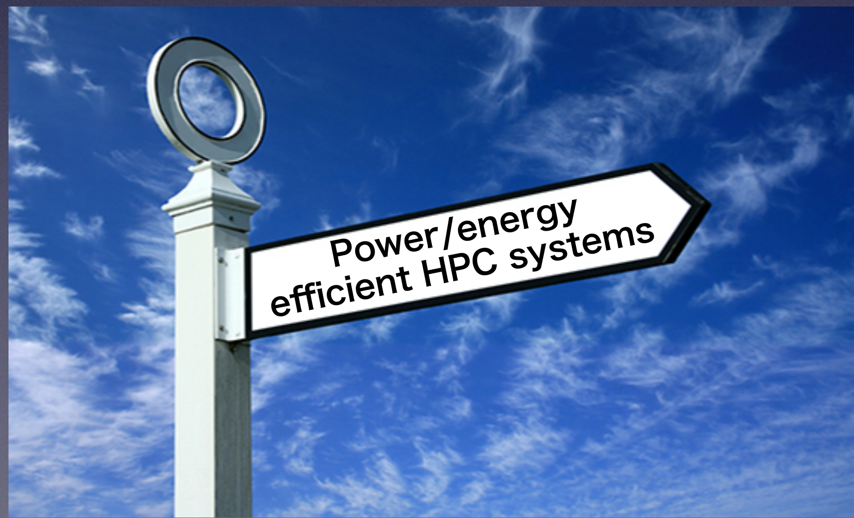
Performance and Power Analysis of SX-ACE

Ryusuke Egawa (Tohoku University)
Yoko Isobe (NEC, Tohoku University)
Soya Fujimoto (NEC, Tohoku University)

WSSP @ Stuttgart

Motivation

- Demands for power/energy efficient operation of HPC systems
  - Users might pay per power/energy instead of time in a future!?
    - Detailed power/energy consumption should can be easily obtained in the future systems
- Performance and power analysis of existing system becomes a “finger post” for future HPC system design.



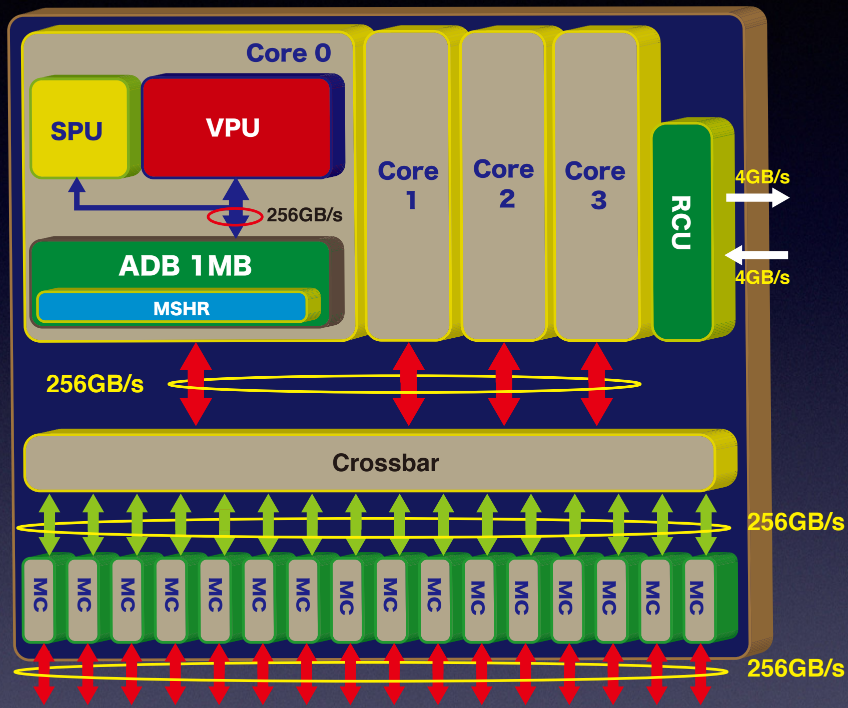
- Since the SX-ACE does not has a function/interface to easily obtain the power profile, this study measures power/energy behavior of SX-ACE using common benchmark programs

Today’s Topic

- Highlights of performance of SX-ACE using HPL, HPCG, and HPGMG
- Power Consumption of SX-ACE
  - Environments
  - Results
  - Discussions
  - How code characteristics/optimizations affect the power efficiency
- Summary

SX-ACE

- 2,560 nodes
  - 1,024 parallel jobs are available @ cyberscience center
- A 4-Core Vector Processor
  - 1GHz clock frequency
  - 272Gflop/s of VPU + 4Gflop/s of SPU per socket
    - 4-core configuration, 68Gflop/s + 1Gflop/s per core
  - 256 GB/s memory bandwidth
    - 1B/F in 4-core Multiply-Add operations
      ～ 4B/F in 1-core Multiply-Add operations
    - 128 memory banks per socket
  - 1MB private ADB per core (4MB per socket)
    - 4x compared with SX-9
    - 4-way set-associative
    - MSHR with 512 entries (address+data)
    - 256GB/s to/from Vec. Reg.
      - 4B/F for Multiply-Add operations
- Other improvement and new mechanisms to enhance vector processing capability, especially for efficient handling of short vectors operations and indirect memory accesses
  - Out of Order execution for vector load/store operations
  - Advanced data forwarding in vector pipes chaining
  - Shorter memory latency than SX-9

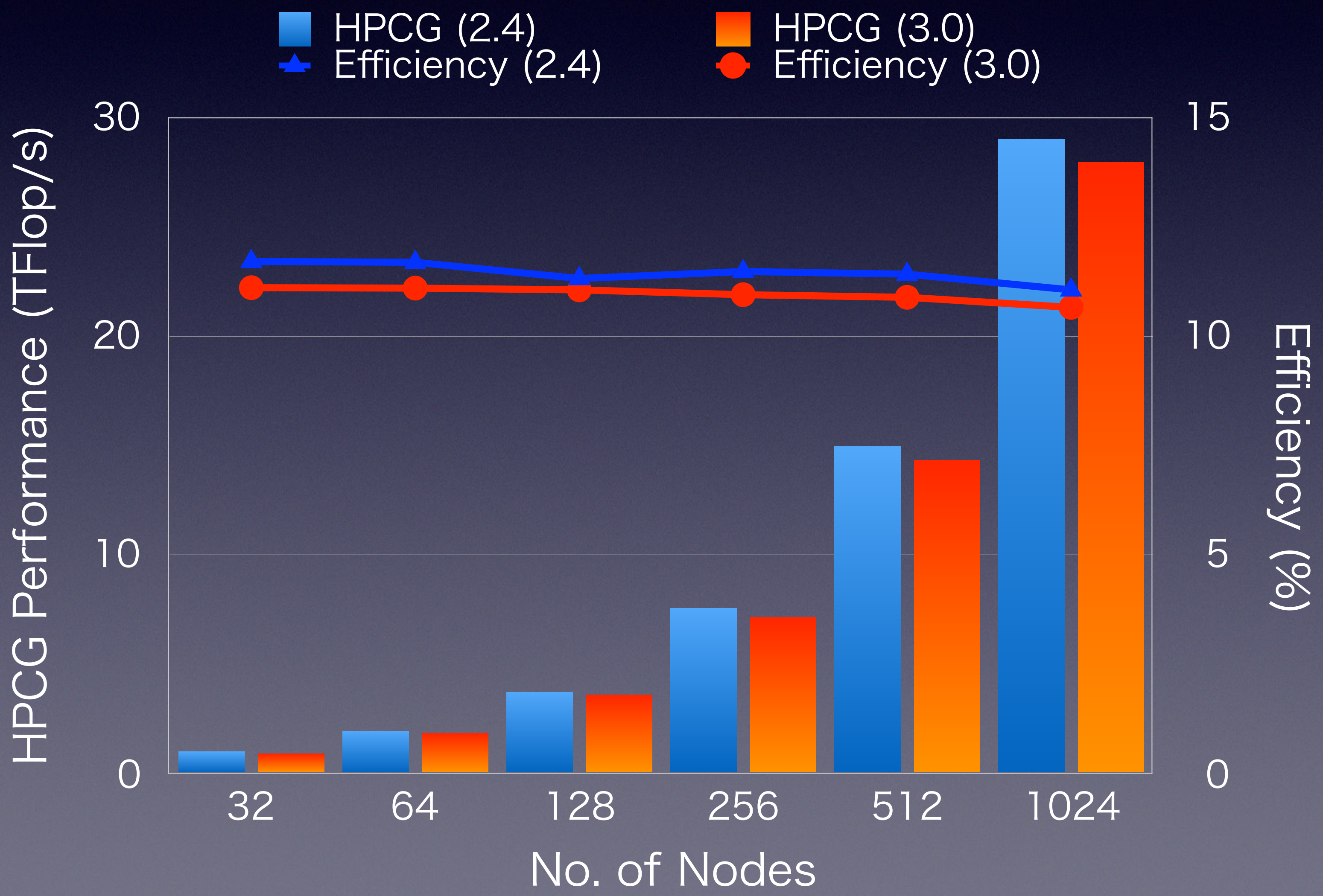


HPL & HPCG Results

HPL

512node: 123.1TF(93.9%)　1,024node: 246.2(93.9%)

HPCG



HPGMG

- HPGMG-FV (High Performance Geometric Multigids)
  - solves variable-coefficient elliptic problems on isotropic cartesian grids
  - Using the **finite volume method (FV)** and **Full Multigrid (FMG)**.
- **Filling the gap between HPL and HPCG**
  - Tracking real application's behavior
    - memory bound, cache friendly
      - 120 points stencil
    - MPI, OpenMP, OpenACC implementations are available
      - Enabling fair comparison with GPUs, Accelerators

Benchmark	Kernel	Required B/F
HPL	DGEMM	< 0.1
HPGMG	GSRB	> 1
HPCG	SpMV, SYMGS	> 4

Benchmarking HPGMG on SX-ACE

- We didn't apply special (code level) optimizations
  - Problem size selection
    - **128^3, 256^3, 512^3 $\log_2$box_dim = 9**
  - Number of nodes to use
    - for better load balancing, number of processes should be $[\text{integer}]^3$

# of nodes	# of Processes	DOF/s	GFLOPS	Eff. [%]
2	8	1.469E+08	176.28	34.43
16	64	1.136E+09	1363.20	33.28
54	216	3.778E+09	4553.60	32.80

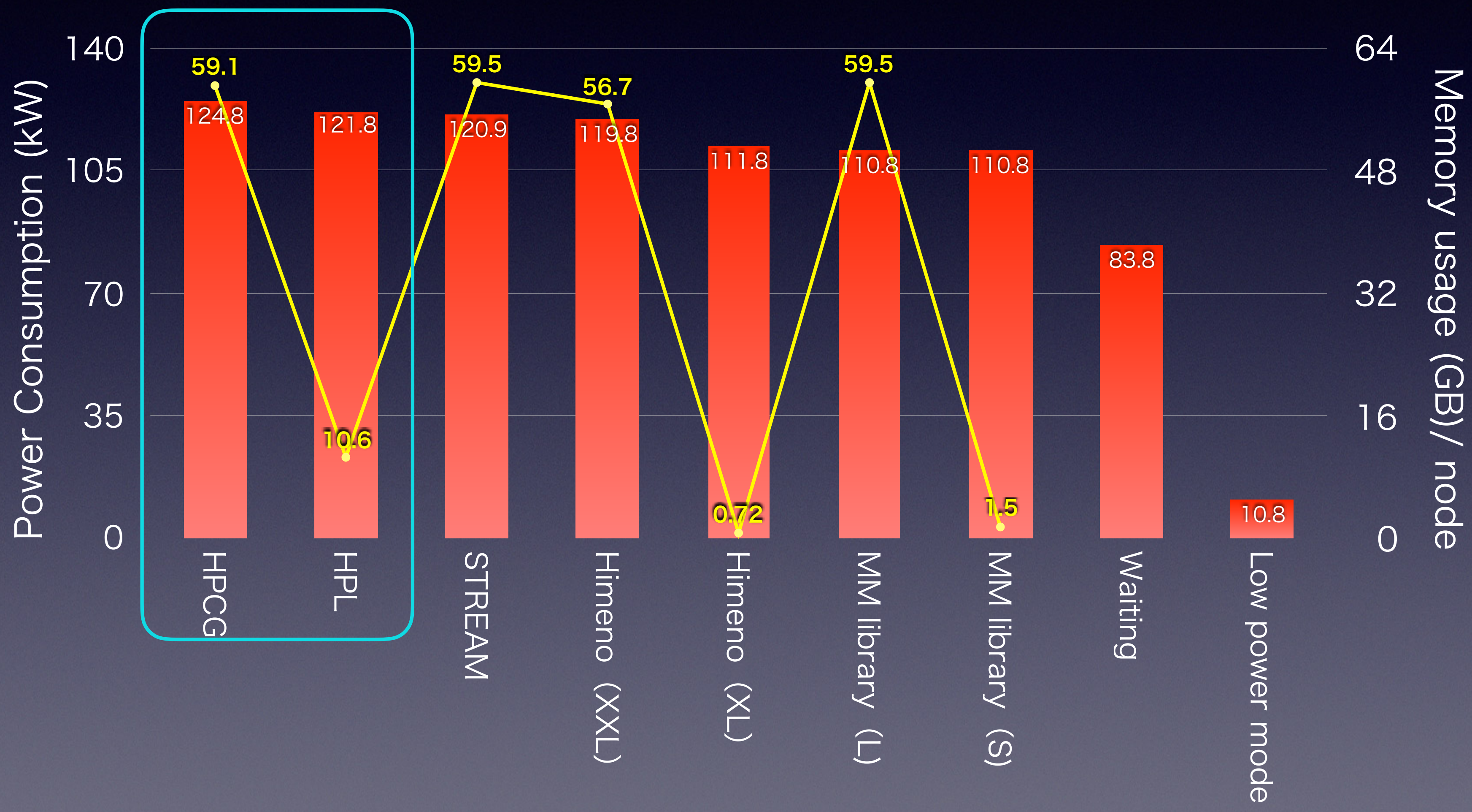
routines	log2_box_dim	Box Size	MFLOP	Eff. [%]	Ave. VL	Code B/F	Act. B/F	ADB Hit
Smooth	7	128	8842.4	13.82	32.9	5.03	1.26	84.77
	8	256	15638.4	24.4	60.5	5.03	1.24	84.03
	9	512	22364.5	34.94	126.6	5.03	1.36	82.21
residual	7	128	19106.7	29.85	73.5	4.83	1.58	51.42
	8	256	25740.7	40.22	135.4	4.83	1.52	51.27
	9	512	24471.6	38.24	212.9	4.83	1.68	51.74

HPGMG-Rank 2016 Nov.

Rank	Site	System	10^9DOF/s	MPI	OMP	Acc	DOF/ Process	Top500 Rank	HPCG Efficiency	Efficiency	System Arch.
1	ALNL	Mira	500	49152	64	0	36M	9	1.66	5.96	BlueGene
2	HLRS	Hazel Hen	495	15408	12	0	192M	14	1.86	8.03	CrayXC40 (Xeon)
3	OLNL	Titan	440	16384	4	1	32M	3	1.83	N/A	CRAY (Xeon+K20)
4	KAUST	ShaheenII	326	12288	16	0	144M	15	1.57	5.41	CrayXC40 (Xeon)
5	NERSC	Edison	296	10648	12	0	128M	60	3.06	14.48	CrayXC30 (Xeon)
6	CSCS	Liz Daint	153	4096	8	1	32M	8	1.60	14.97	CrayXC30 (Xeon+K20)
7	Tohoku Univ.	SX-ACE	73.8	4096	1	0	128M	-	10.64	33.78	NEC SX
8	LRZ	Super MUC	72.5	4096	8	0	54M	36	2.62	12.29	Idataplex (Xeon)
9	NREL	Peregrine	10.0	1024	12	0	16M	-	N/A	2.65	Apollo 8000 (Xeon)
10	NREL	Peregrine	5.29	512	12	0	16M	-	N/A	5.38	Apple 8000 (Xeon)
11	HLRS	KABUKI	3.24	256	1	0	32M	-	11.45	23.71	NEC SX
12	NERSEC	Babagge	0.726	256	45	0	8M	-	N/A	N/A	Intel Xeon Phi

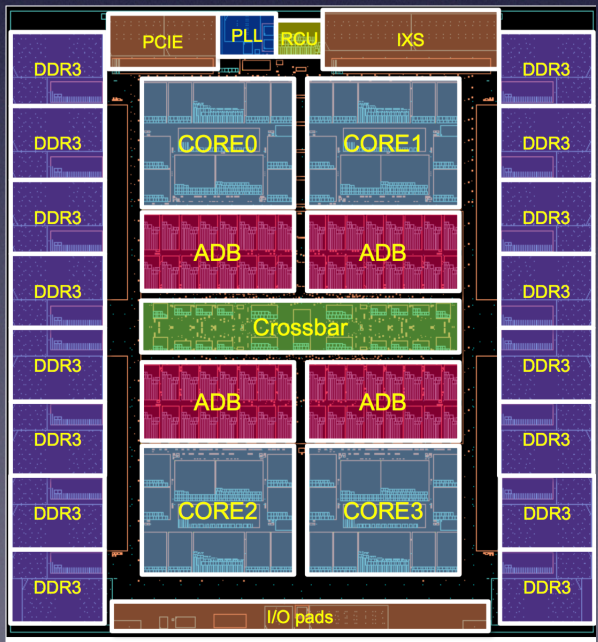
Power Measurements

Power Consumption of SX-ACE(1 Cluster)

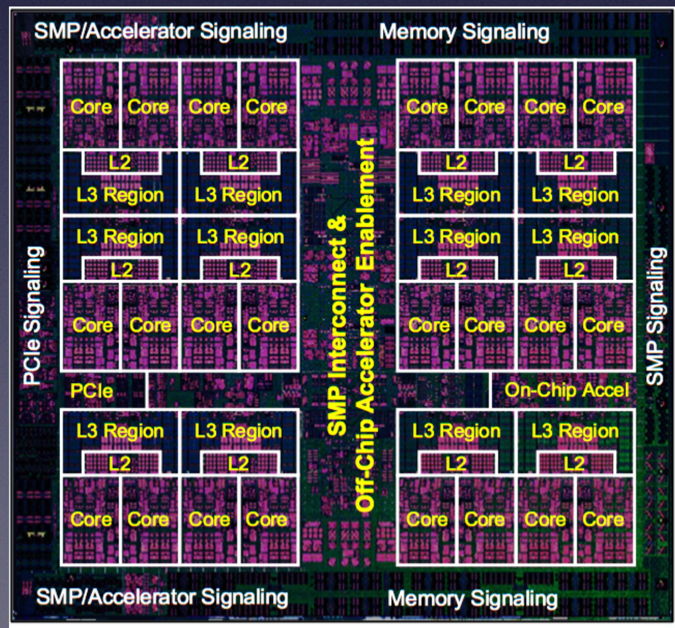


SX-ACE

- A penalty of silicon budgets are used to obtain a certain memory bandwidth compared to scalar type processors
  - memory power seems dominant on SX-ACE



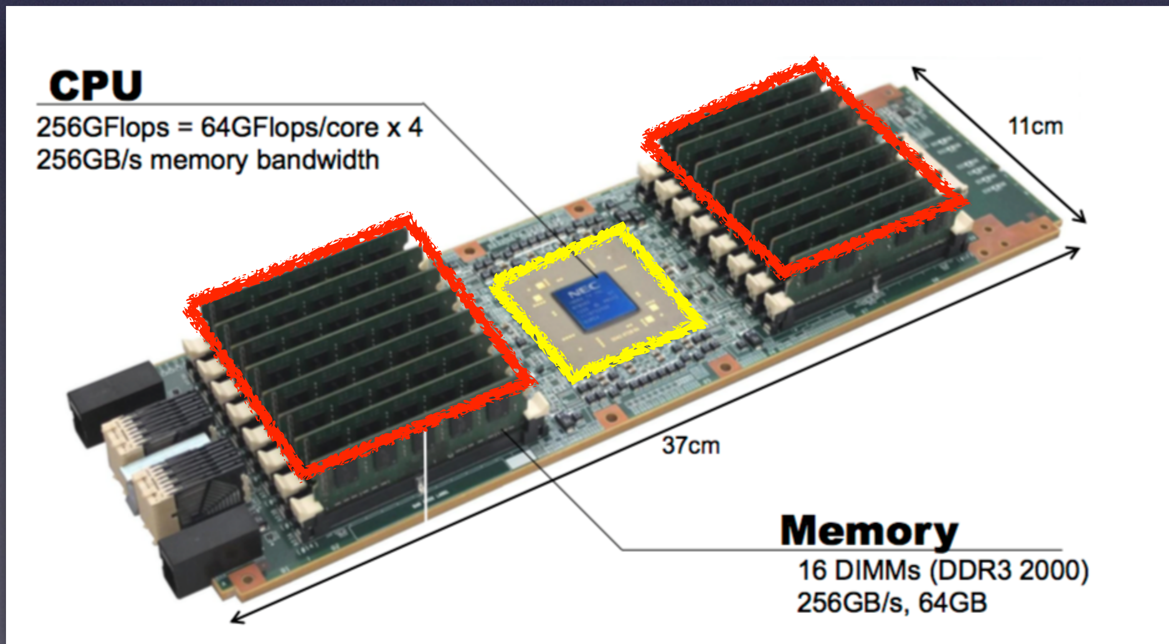
SX-ACE



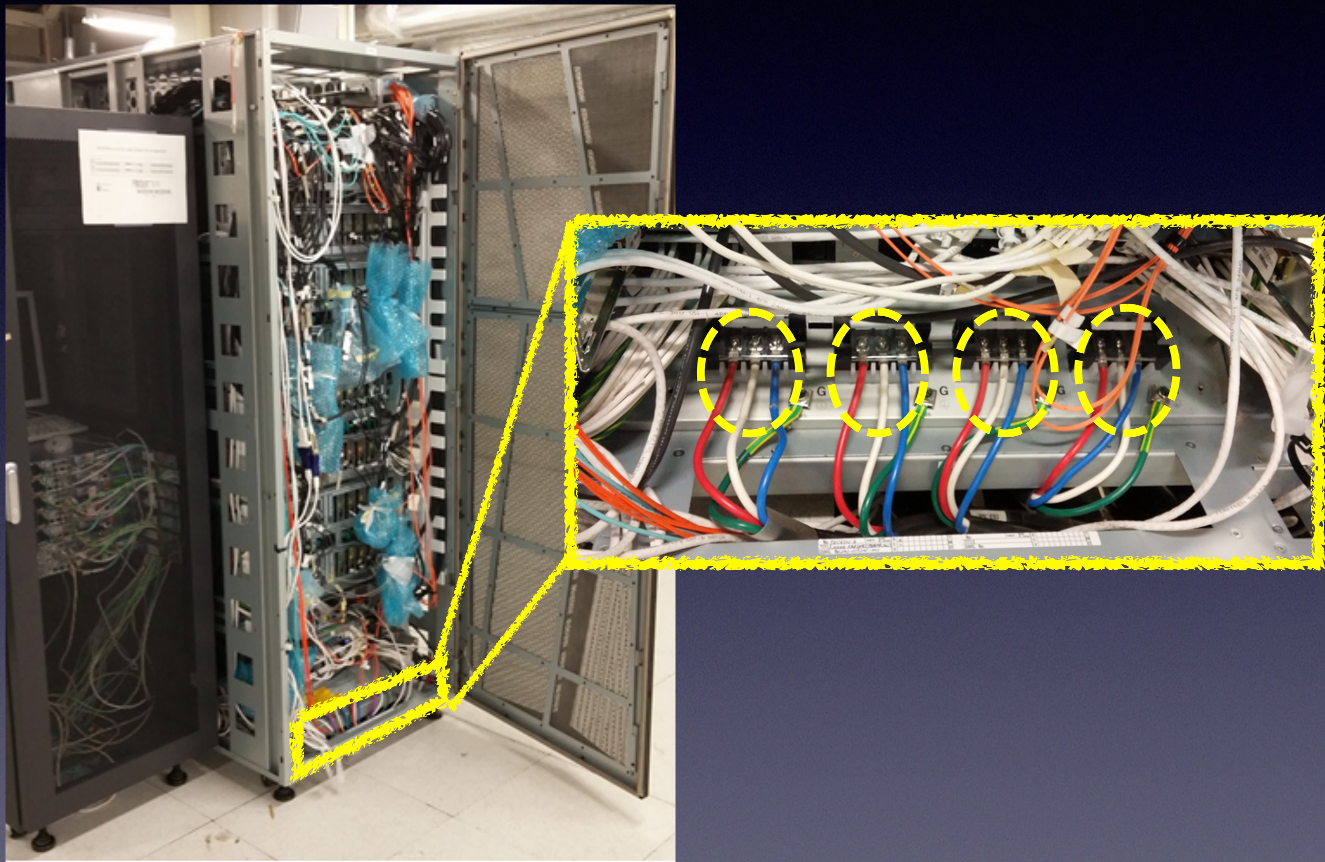
Power 9

Power Measurement Environments

- Focusing on Power of CPU, memory, and others
  - There is no API/UIs for measuring power of CPU, Memory, system on SX-ACE
    - Intel has RAPL ・・・ :-<
  - Only measured CPU power can be obtained by the maintenance console



Power Measuring Environments



Clump Meter

Fujimoto-san

Power Consumption of SX-ACE

Baseline OS Only

P_{Others}

approximation based on measurements of PD

P_{CPU}

47.1W | 63.4W | 59.0W

measured power(CPU)

measured power (Total)

P_{total}

$$P_{Memory} = P_{total} - (P_{CPU} + P_{Others})$$

With apps

approximation based on measurements of PD

Δ 47.1W | 63.4W | 59.0W σ

measured power(CPU)

measured power (Total)

$$P_{Memory} = P_{total} - (P_{CPU} + P_{Others})$$

Δ σ

are included in P_{Totol}

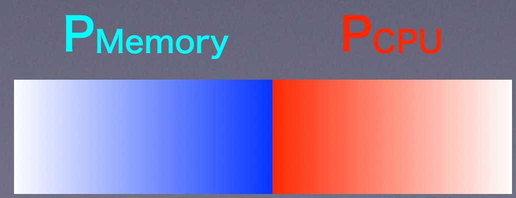
(also in PCPUmeasured data)

Let's imagine how the power is distributed among CPU and Memory

- HPL
  - N_{mem}=1,314,575,046,530 N_{comp}= 40,921,856,965,740
    - **N_{mem} : N_{comp} ≒ 1 : 30**

- HPCG
  - N_{mem}=108,545,870,653,100 N_{comp}=73,577,230,636,248
    - **N_{mem} : N_{comp} ≒ 1 : 0.6**

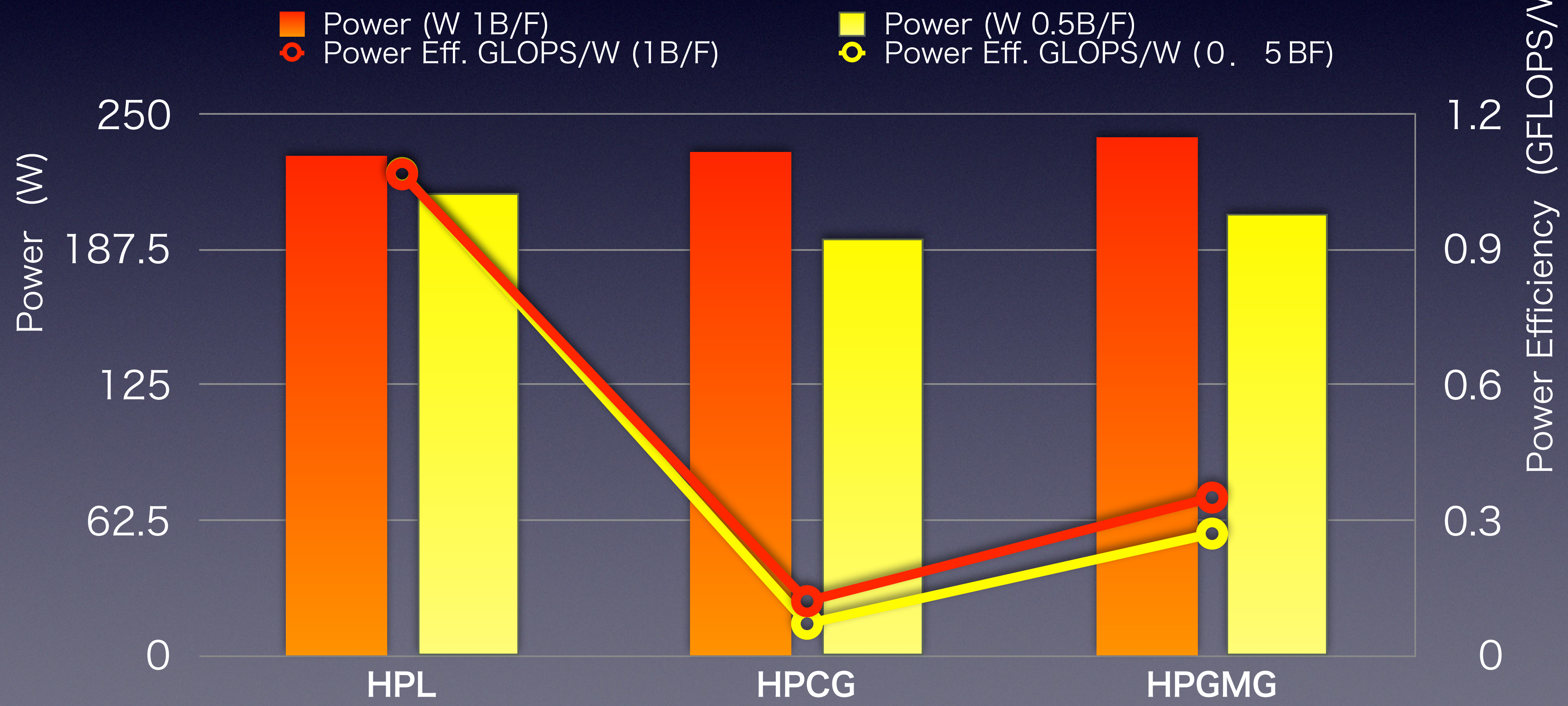
- HPGMG
  - N_{mem}=2,948,629,297,204 N_{comp}=7,603,493,066,782
    - **N_{mem} : N_{comp} ≒ 1 : 2.5**



Power Consumption

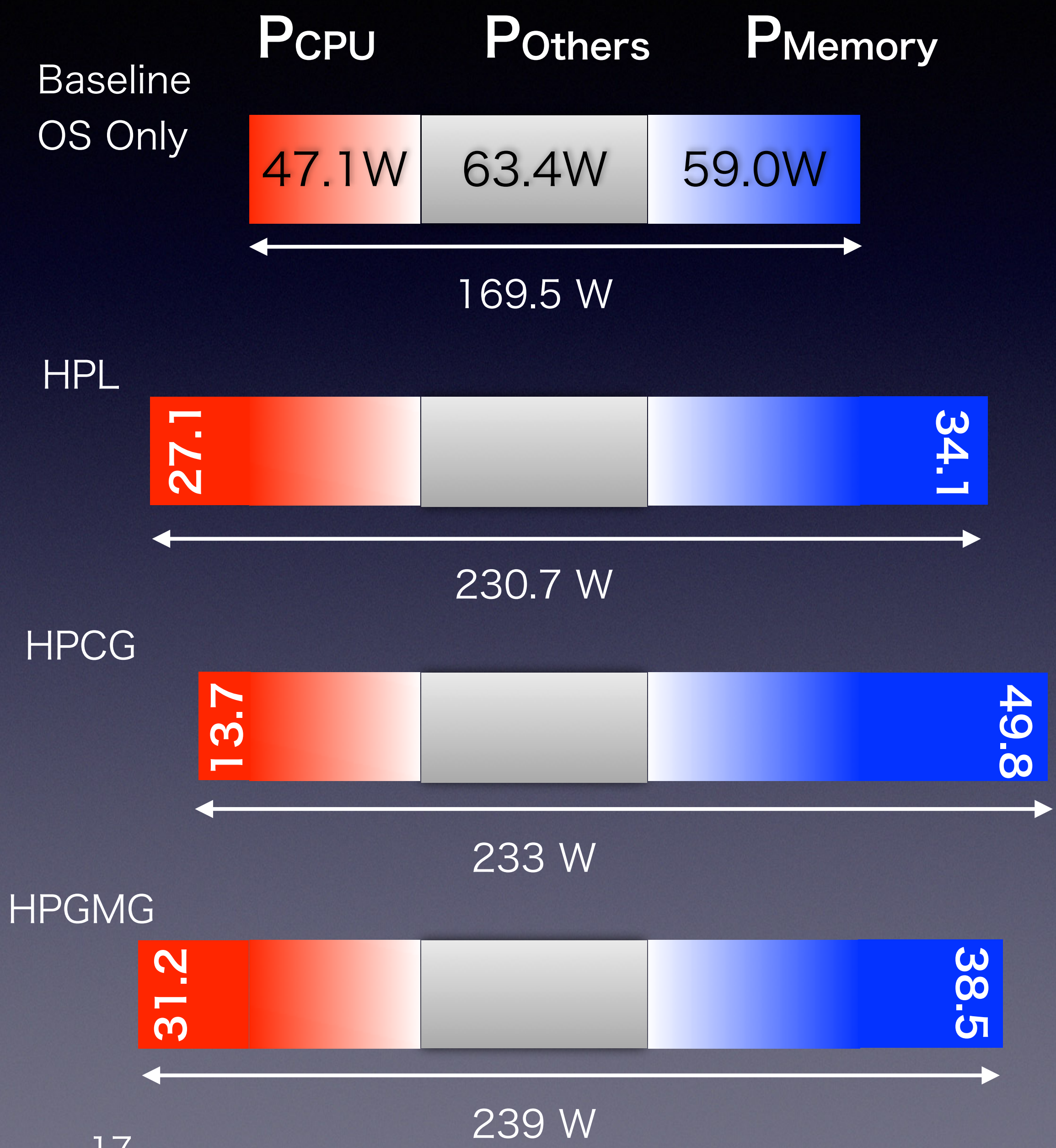
- Performance and power are evaluated using single node of SX-ACE

	1BF	0.5BF
Peak Perf.	256	256
Mem. BW	256	128



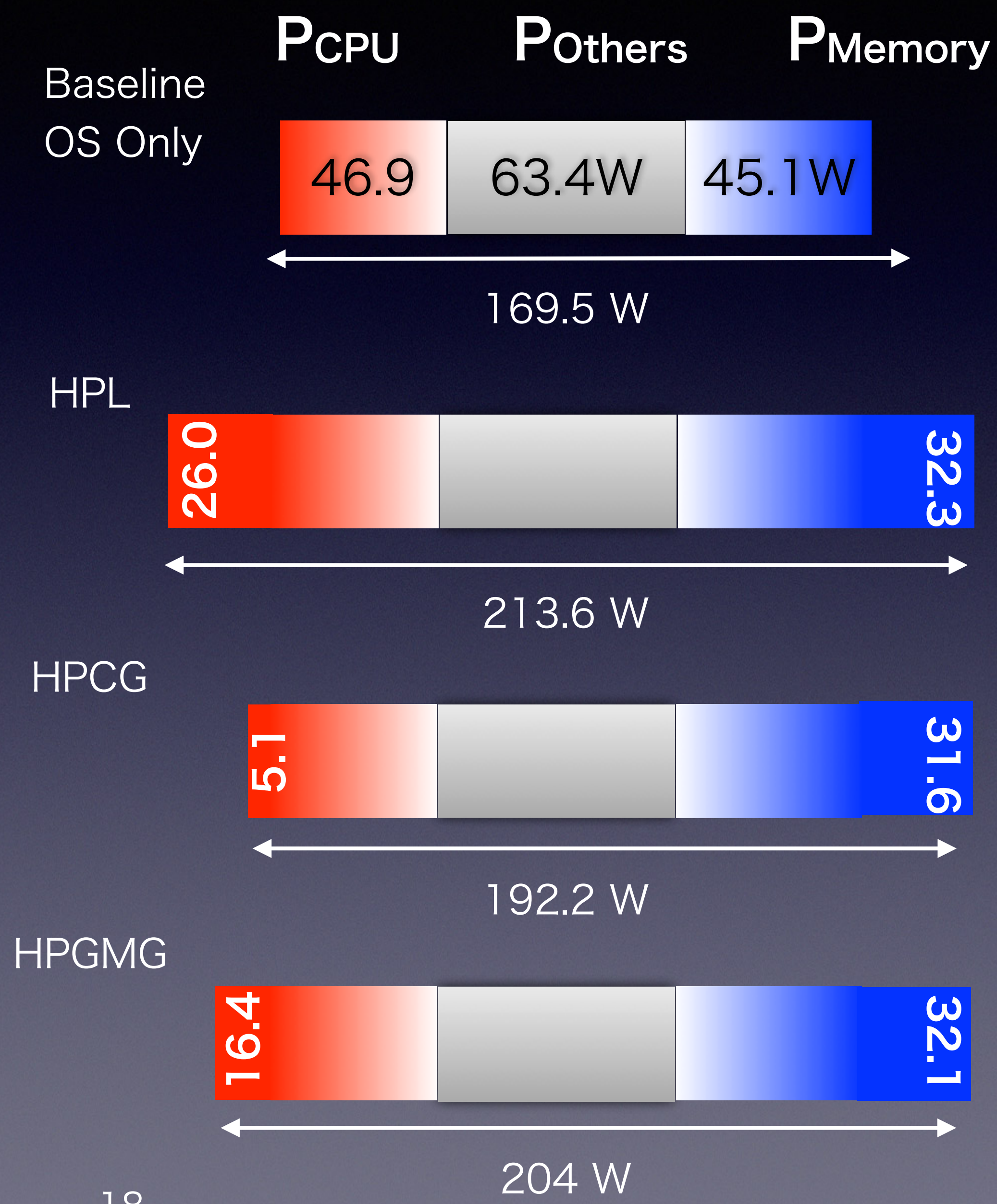
Power Comparison (Base Line)

	Base Line (1 B/F)		
Peak BF (GFlop/s)	256		
Memory BW (B/sec)	256		
System B/F	1		
	HPL	HPCG	HPGMG
Performance (GFlop/s)	246.3	28.8	84.3
Max Power (W)	230.7	233.0	239.2
Efficiency	96.2%	11.2%	32.9%
Performance /Watt	1.067	0.12	0.35



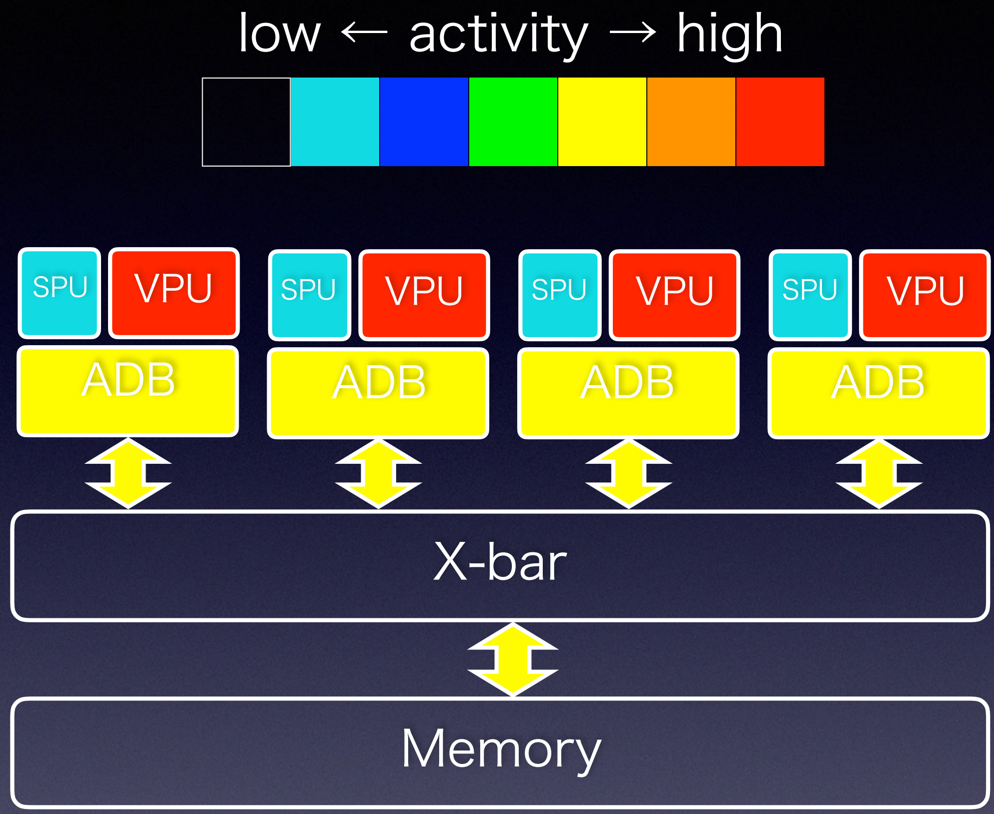
Power Comparison (0.5)

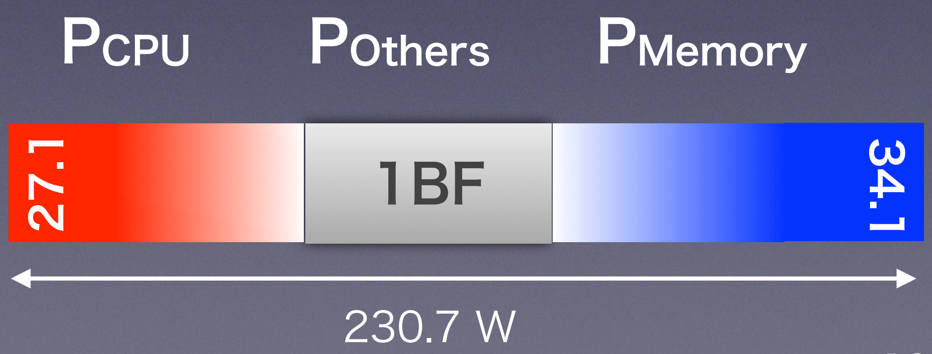
	Base Line (0.5 B/F)		
Peak BF (GFlop/s)	256		
Memory BW (B/sec)	128		
System B/F	0.5		
	HPL	HPCG	HPGMG
Performance (GFlop/s)	230.2	13.6	56
Max Power (W)	213.6	192.2	204
Efficiency	89.9%	5.3%	21.8%
Performance /Watt	1.07	0.07	0.27

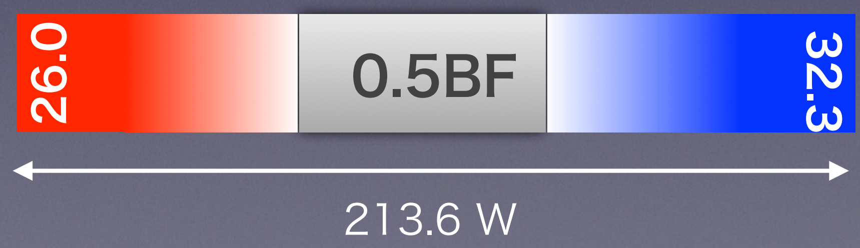


HPL

- VFA and VFM has almost 100% workload
- ADB Hit rates 77.8%
  - Code B/F = 0.36
  - Actual B/F = 0.32

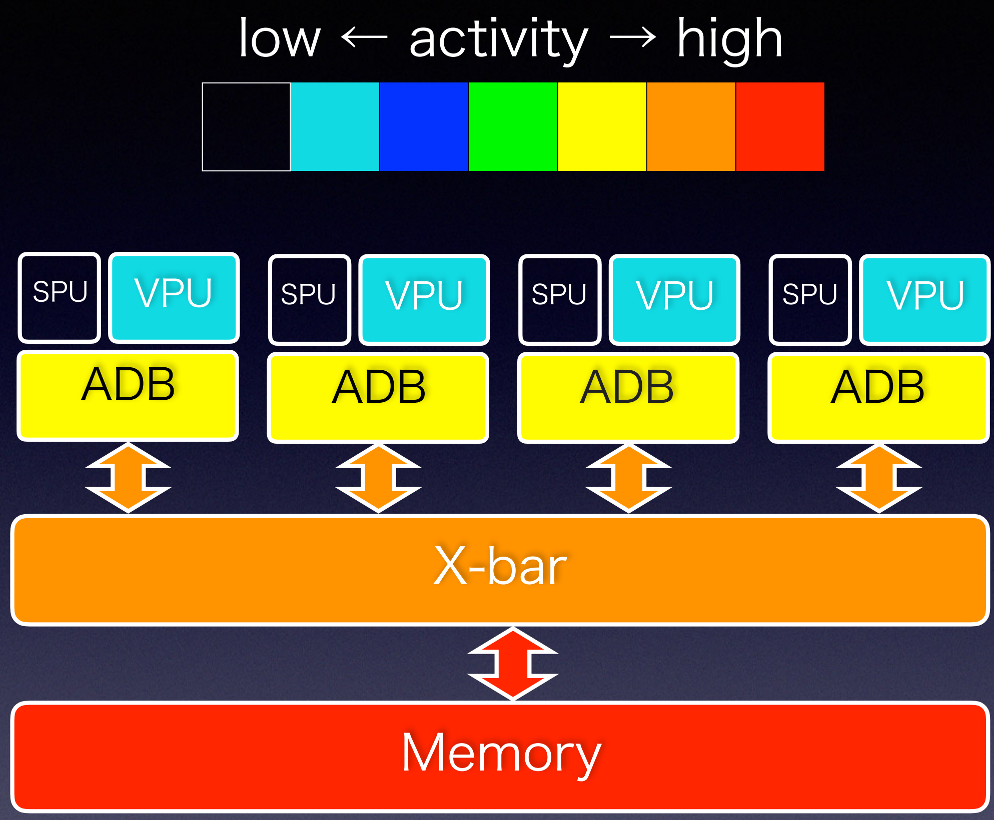




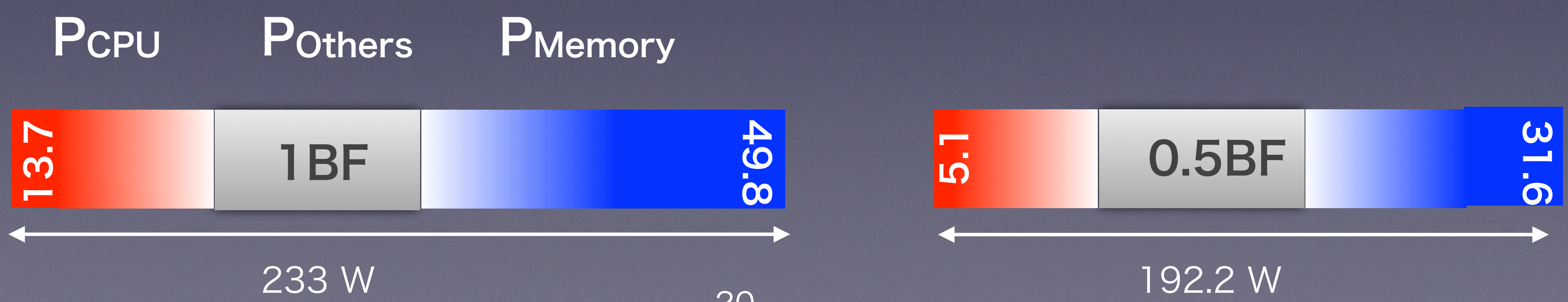


HPCG

- High memory pressure
- ADB also shows a high activity
  - Actual BF = 6.5
  - ADB Hit rate = 61.54%
- Since the memory cannot feed enough data to the VPU, the workloads of VPU are relatively low.

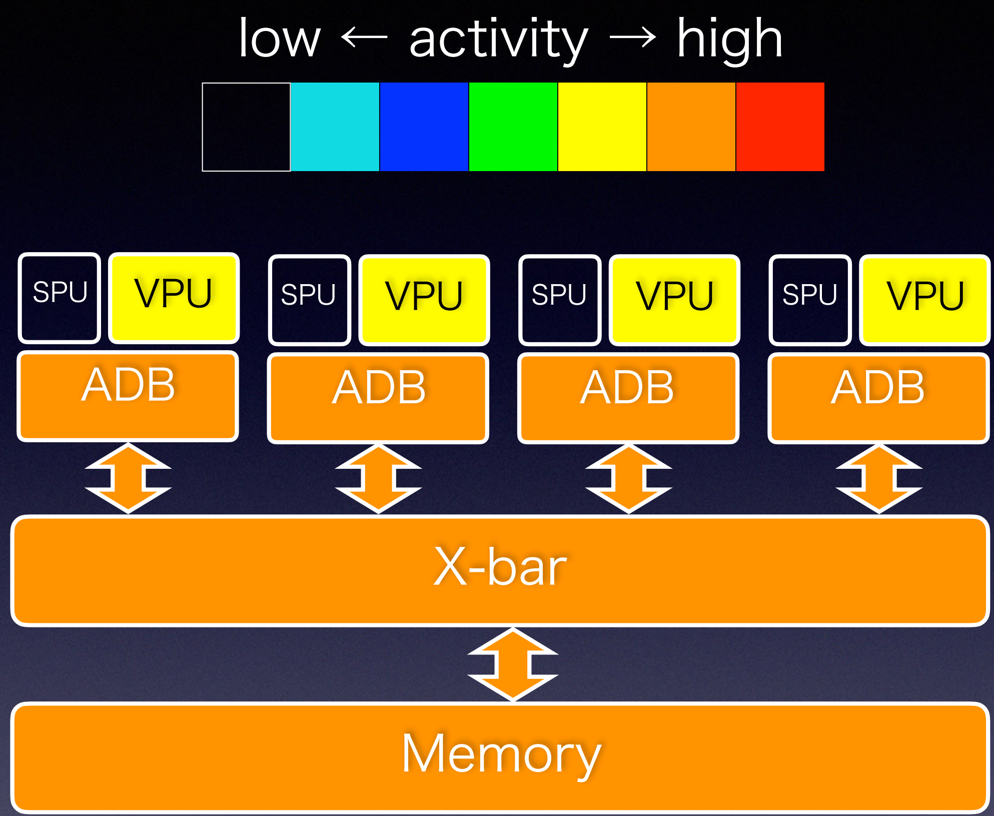


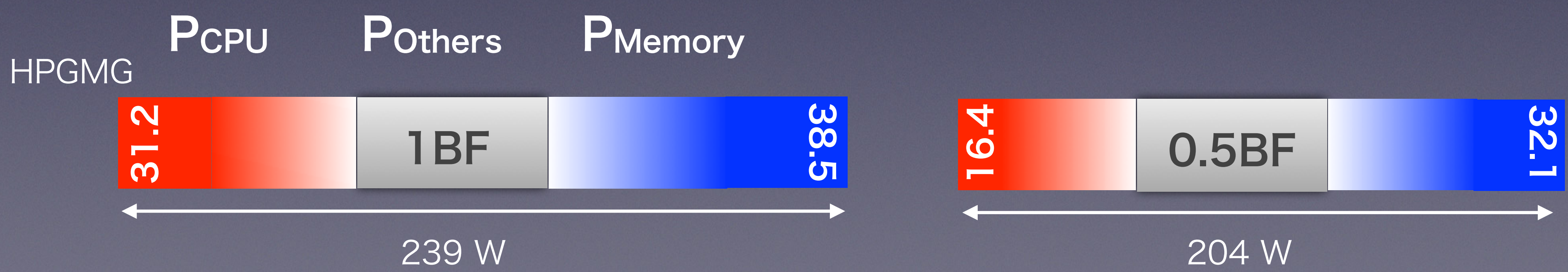
HPCG



HPGMG

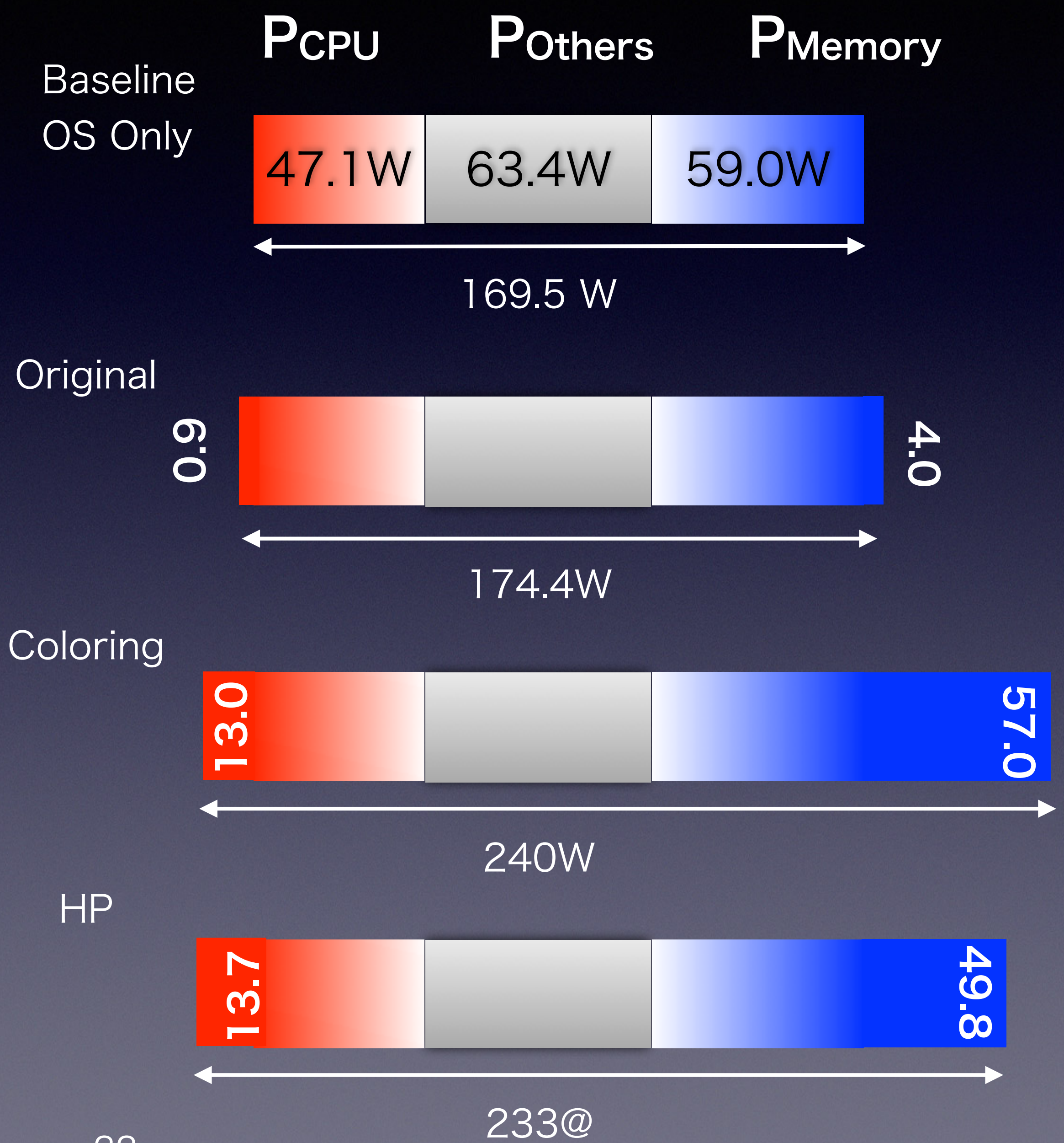
- High-pressure on X-bar and Memory
  - Act.B/F=1.5
  - High ADB Activity (Hit rate)
    - Stensil calculation
- Since the efficiency of HPGMG is 33%, activities of VPU become
  - HPL>HPGMG>HPCG
- ADB activity is higher than HPL
  - Thus , CPU power is larger than HPL





Effects of Code optimizations on Performance and Power

	ORG	Color	HP
Performance [GF]	0.58	27.5	28.8
Power [W]	174	240	233
Prog B/F		11.5	11.9
Actual B/F		7.0	6.4



Memory optimization is also effective for power reduction on SX-ACE

The relationship between code optimization and power/performance should be examined

Conclusions

- Performance and Power Analysis of SX-ACE
  - Sharing the measurement data :-)
- Memory affects the performance and power of SX-ACE significantly
- Memory access optimization seems effective for the both of performance and power consumption
- Needs to have API/UI for measuring/tracking the power/performance behavior for the future system
- More detailed analysis of changes in power with profiling data is our future work.