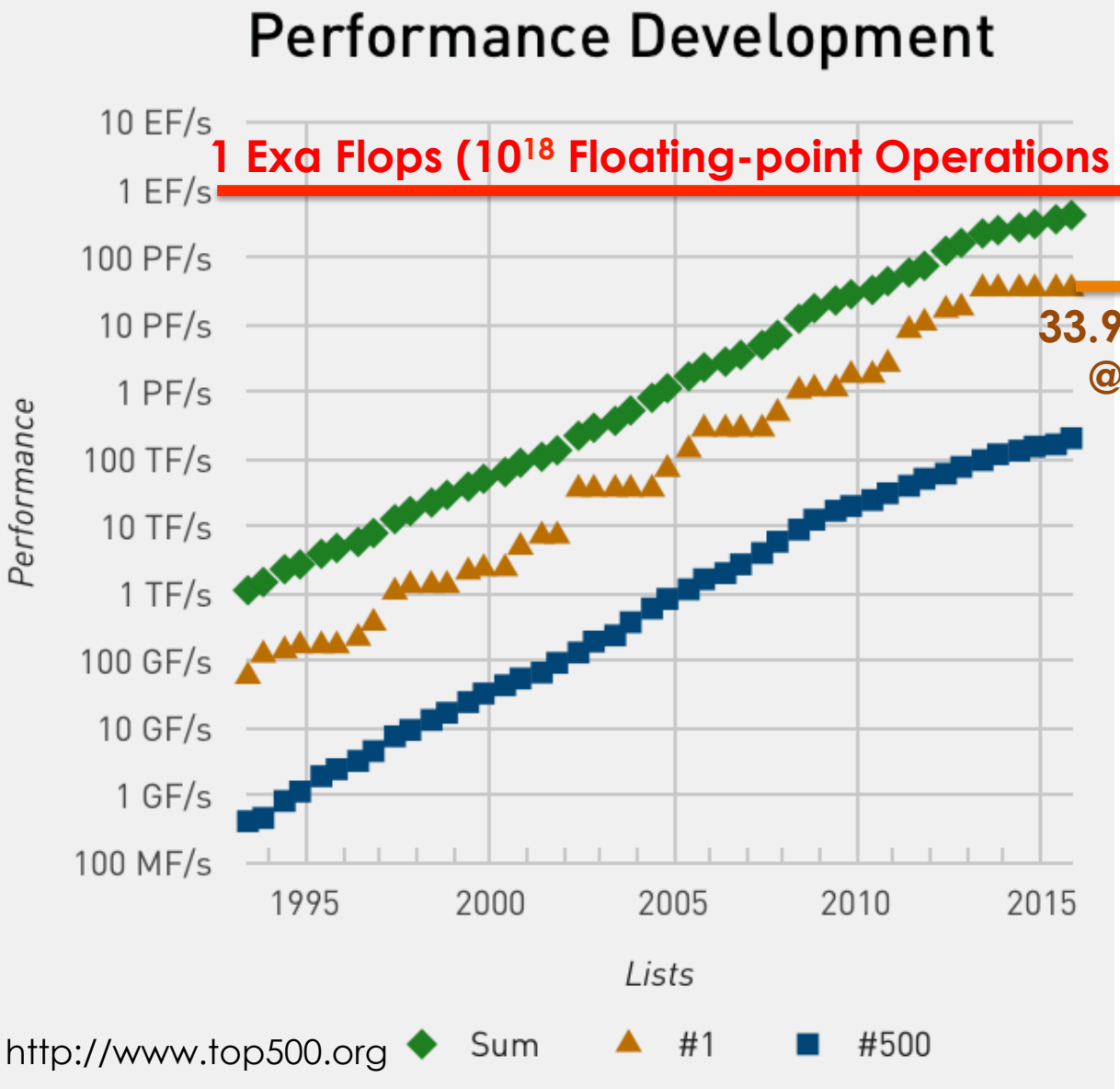
Impact of Manufacturing Variability in Power Constrained Supercomputing

Koji Inoue

Kyushu University

Trends of Supercomputing



World-Wide Next Target
1 Exa Flopas @ 20 – 30 MW

1 Exa Flops (10^{18} Floating-point Operations Per Second)

30X

33.9 Peta Flops
@17.8 MW

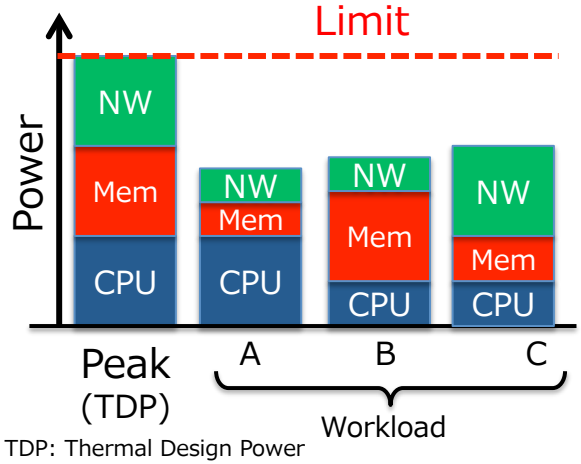
1.7X

Need to improve power efficiency!

Overprovisioned Systems

Under-provisioned (Conventional)

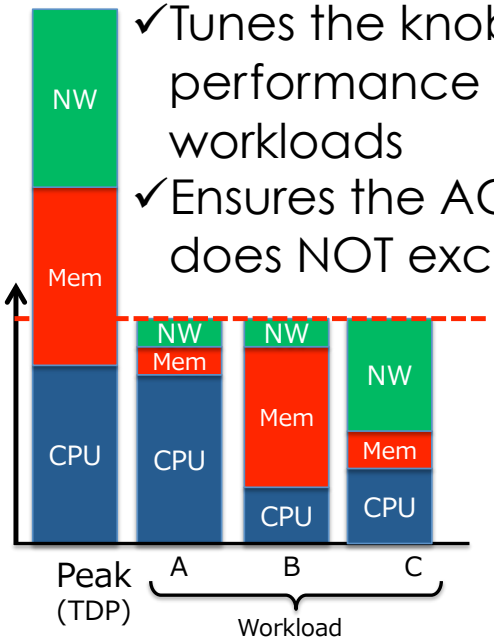
- HW Design
  - ✓Ensures the PEAK system power does NOT exceed the limit
- SW Design
  - ✓Tries to maximize the activity of HW components



TDP: Thermal Design Power

Over-provisioned

- HW Design:
  - ✓Allows to install HWs w/o considering the power limit
  - ✓Provides power-performance knobs
- SW Design:
  - ✓Tunes the knobs to maximize the performance based on SW workloads
  - ✓Ensures the ACTUAL system power does NOT exceed the limit



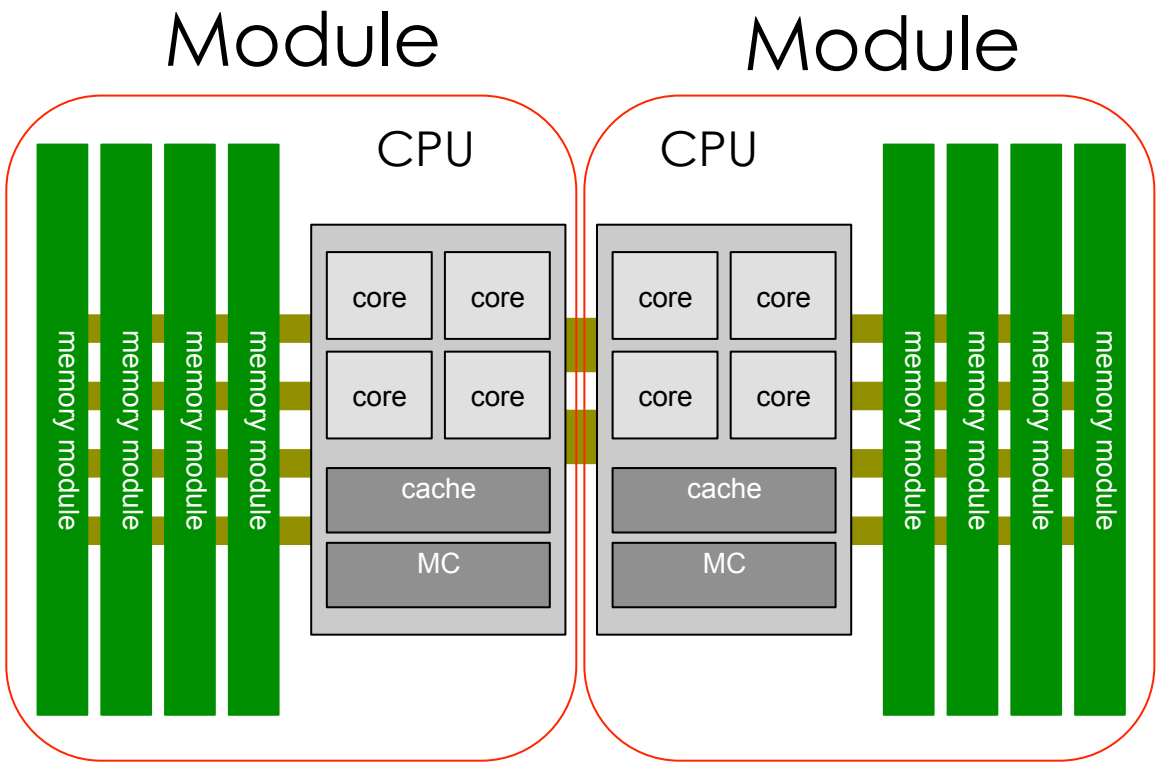
Experimental Setup

- HPC Challenge: star DGEMM, star STREAM(Triad)
- NPB: BT, SP, EP
- Magneto Hydro-Dynamics(MHD) simulation
  - Typical stencil app. to simulate space plasma
  - Calculations and communications appear in turn
- Fiber benchmark suite: mVMC-mini (mVMC)
  - Variational Monte-Carlo simulation for strongly correlated electron system

Blue=EP type
Red=With Comm. & Sync.

Site	Node Micro-Architecture	Total nodes	Procs. Per Node	Cores Per Procs.	Power Msrmt.
Cab(LLNL)	Intel E5-2670 Sandy Bridge	1,296	2	8	RAPL
BG/Q Vulcan (LLNL)	IBM PowerPC A2	24,576	1	16(compute)	EMON
Teller (SNL)	AMD A10-5800K Piledriver	104	1	4	PI
HA8K(Kyushu Univ.)	**Intel E5-2697v2 Ivy Bridge**	**965**	**2**	**12**	**RAPL**

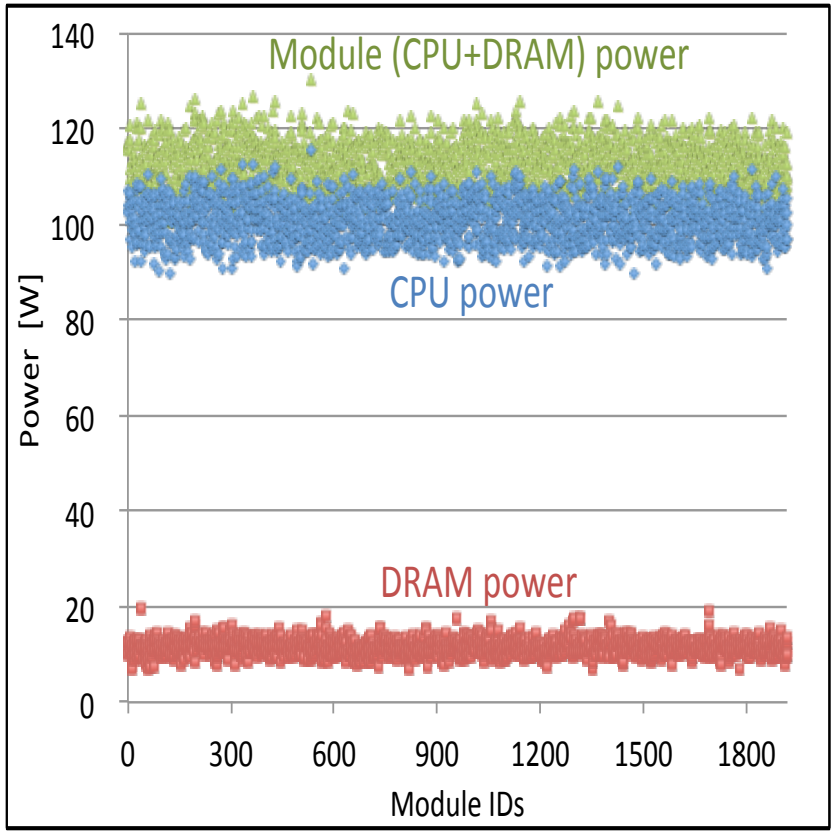
Terminology



CPU = Processor chip (including cores, cache, MC, etc.)
Module = A pair of a CPU and DRAMs directly connected to it

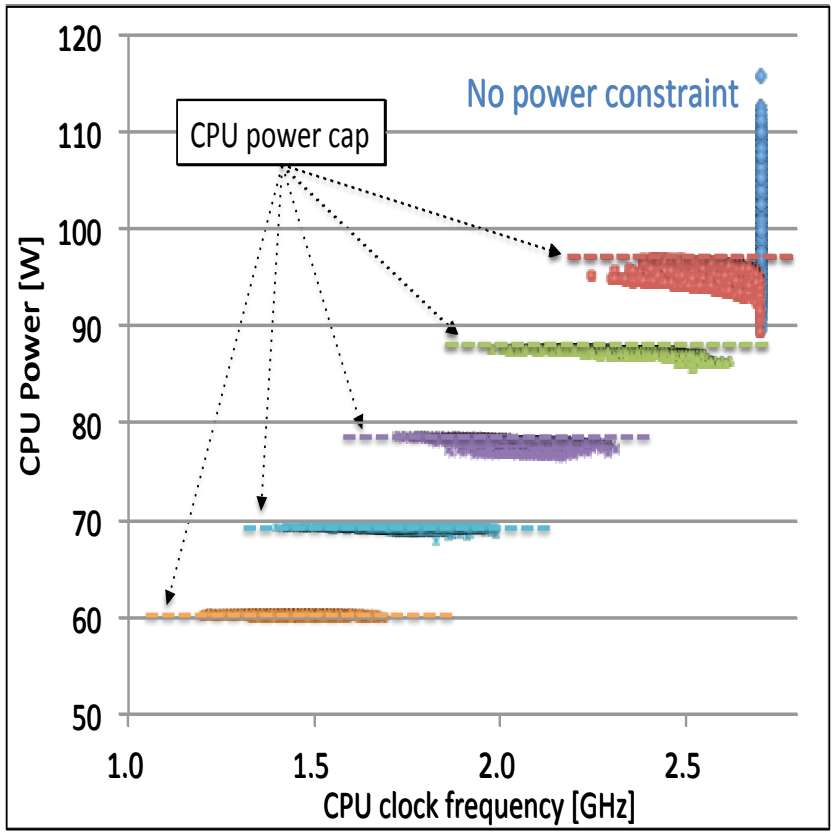
Impact on CPU Frequency

star DGEMM



30%

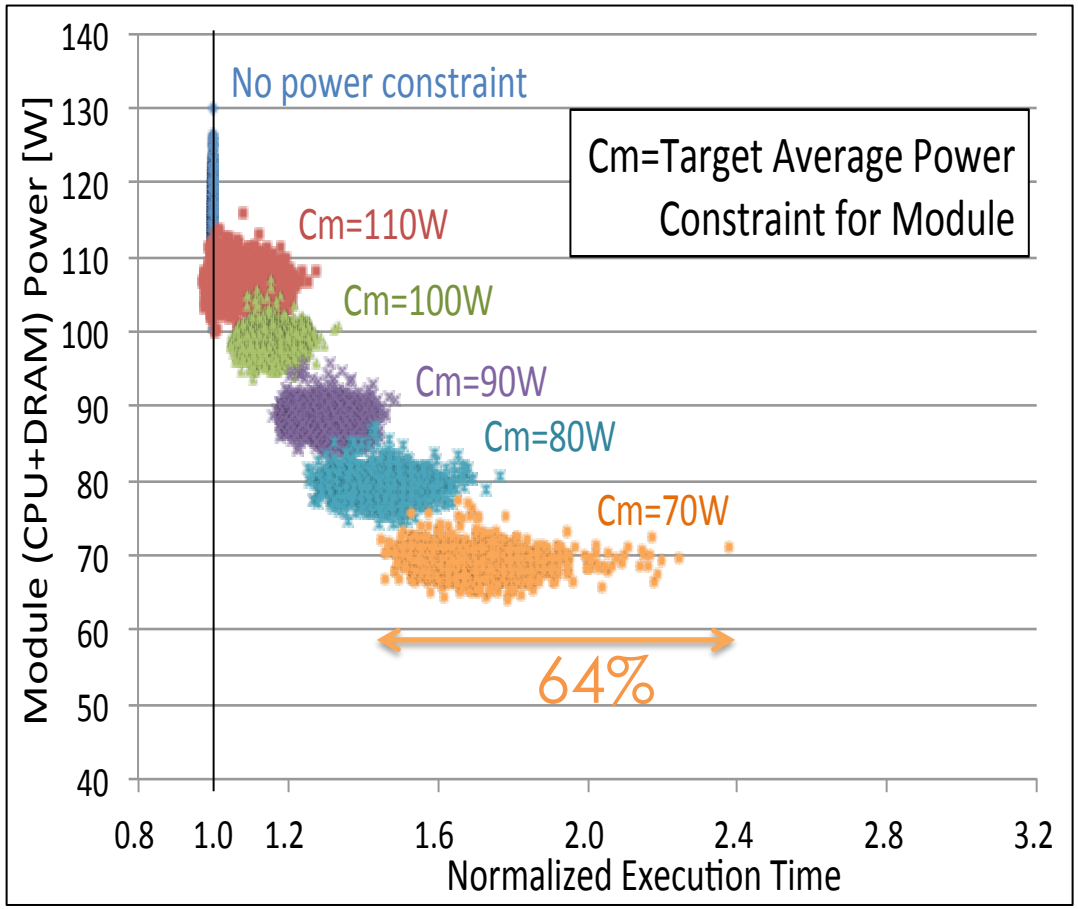
w/ a uniform power constraint



Power variation is translated into CPU frequency variation applying uniform power constraint!

Impact on Application Performance

star DGEMM w/ a uniform power constraint



KYUSHU UNIVERSITY

Problem and Goal

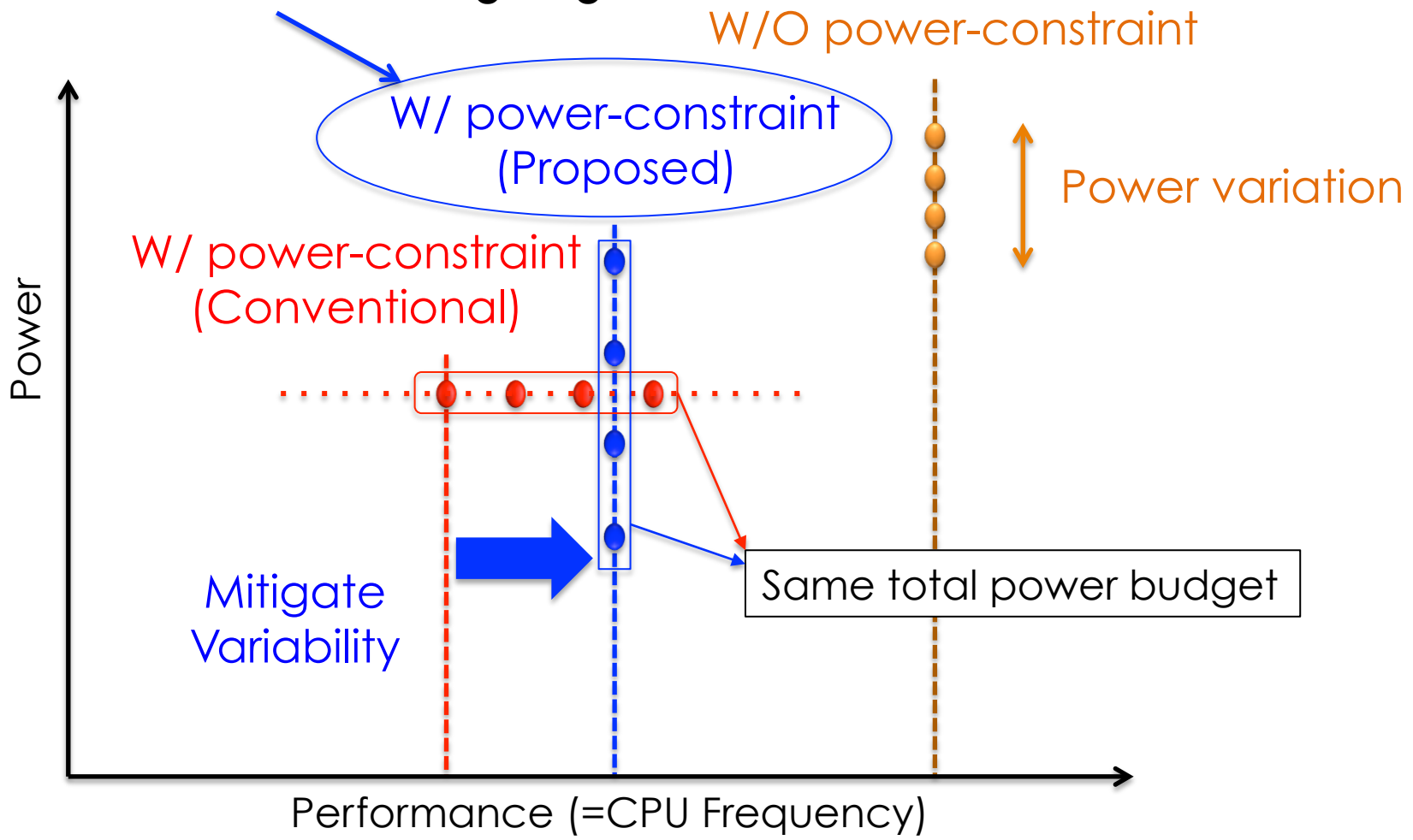
- Power-Constraint Supercomputing
  - will be applied to future HPC systems
- Manufacturing Variability
  - leads to performance variation under power constraint

Our Goal

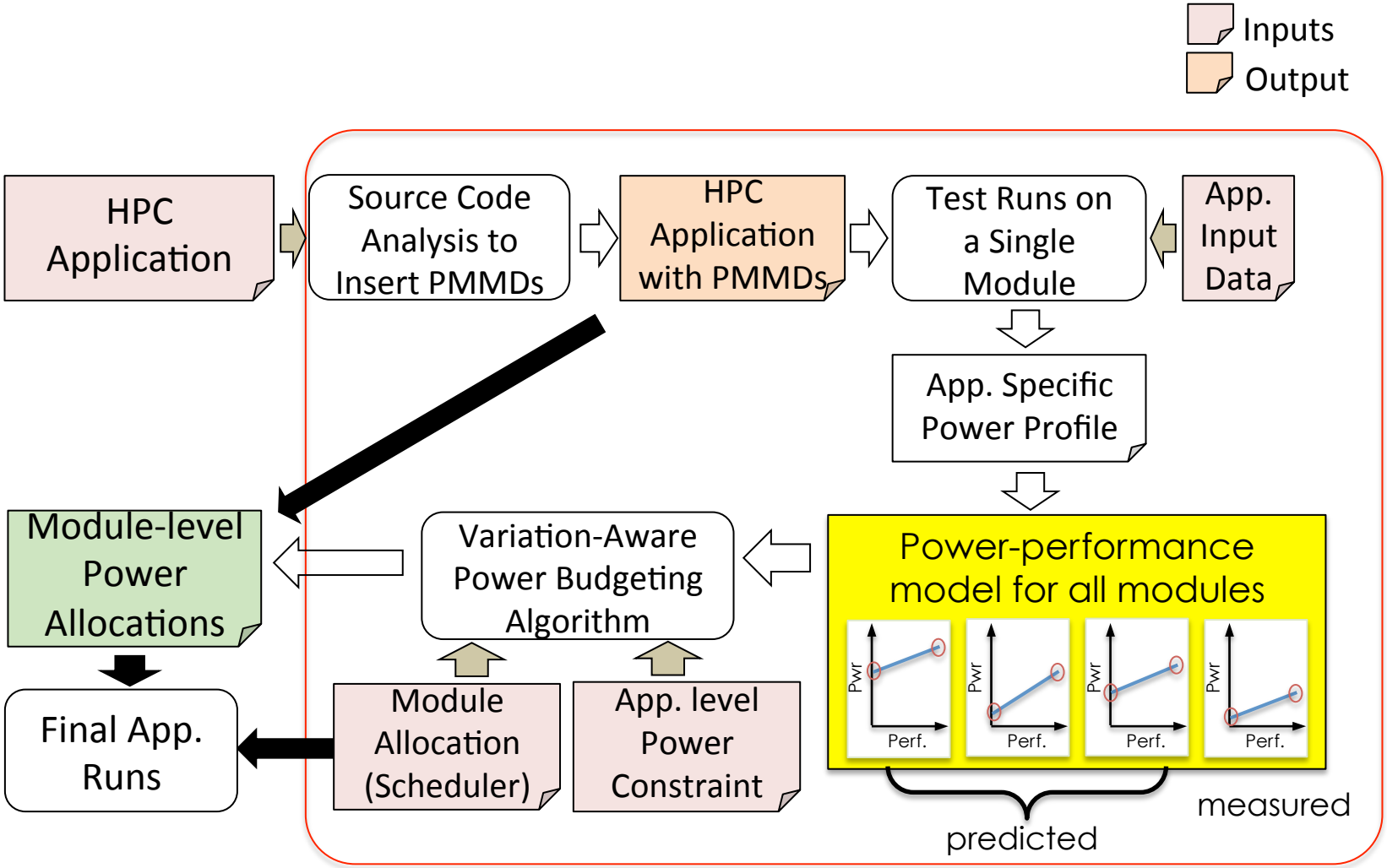
Mitigate the impact of manufacturing variability on performance of HPC apps. under power constraint !

Concept

Variation-Aware Power Budgeting



Variation-Aware Power Budgeting Strategy



Power Model Calibration

Application-independent Power Variation Table (PVT)

Module ID	Normalized Power
1	1.0
:	:
k	1.2
:	:
N	0.8

Obtained once at system installation

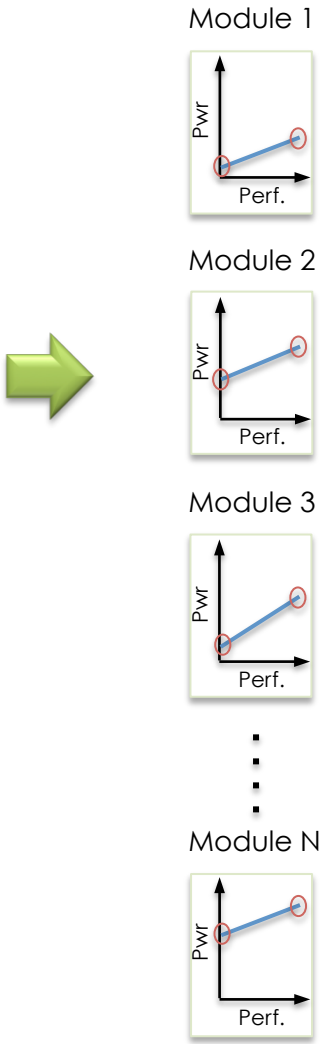
Estimated Application Specific Power Consumption

Module ID	Power Consumption
1	
:	
k	120W
:	
N	

Measured power on single module k

Module ID	Power Consumption
k	120W

Test run on a module!



Power Model Calibration

Application-independent Power Variation Table (PVT)

Module ID	Normalized Power
1	1.0
:	:
k	**1.2**
:	:
N	0.8

Estimated Application Specific Power Consumption

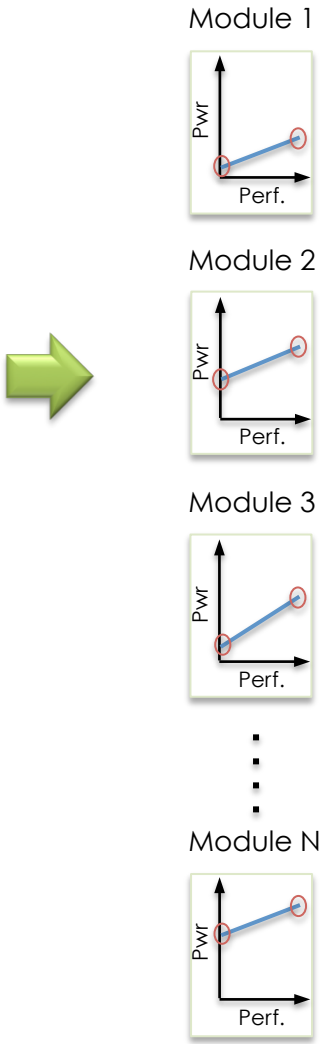
Module ID	Power Consumption
1	
:	
k	**120W**
:	
N	

120W/1.2

application dependent average power

Measured power on single module k

Module ID	Power Consumption
k	**120W**



Power Model Calibration

Application-independent Power Variation Table (PVT)

Module ID	Normalized Power
1	1.0
:	:
k	1.2
:	:
N	**0.8**

Estimated Application Specific Power Consumption

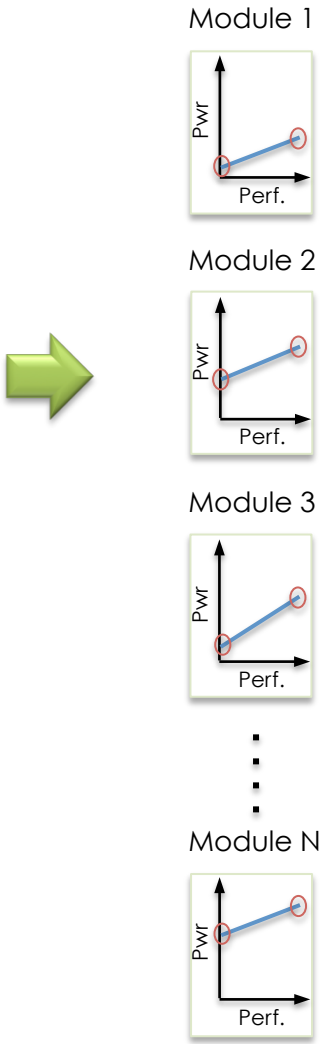
Module ID	Power Consumption
1	
:	:
k	120W
:	:
N	**80W**

(120W/1.2) x 0.8

application dependent power on module-N

Measured power on single module k

Module ID	Power Consumption
k	120



KYUSHU UNIVERSITY

Options for Power Setting

Two options for power settings

- ***Power Capping*** (Pc) using RAPL
- ***Frequency Selection*** (Fs) using CPUFreqlibs

	Power Capping (Pc)	**Frequency Selection (Fs)**
Power Constraint	◎ Guaranteed	△ Not guaranteed
Performance Equivalence	△ Not guaranteed	◎ Guaranteed

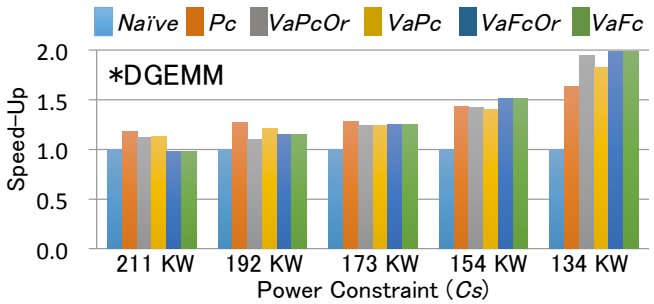
KYUSHU UNIVERSITY

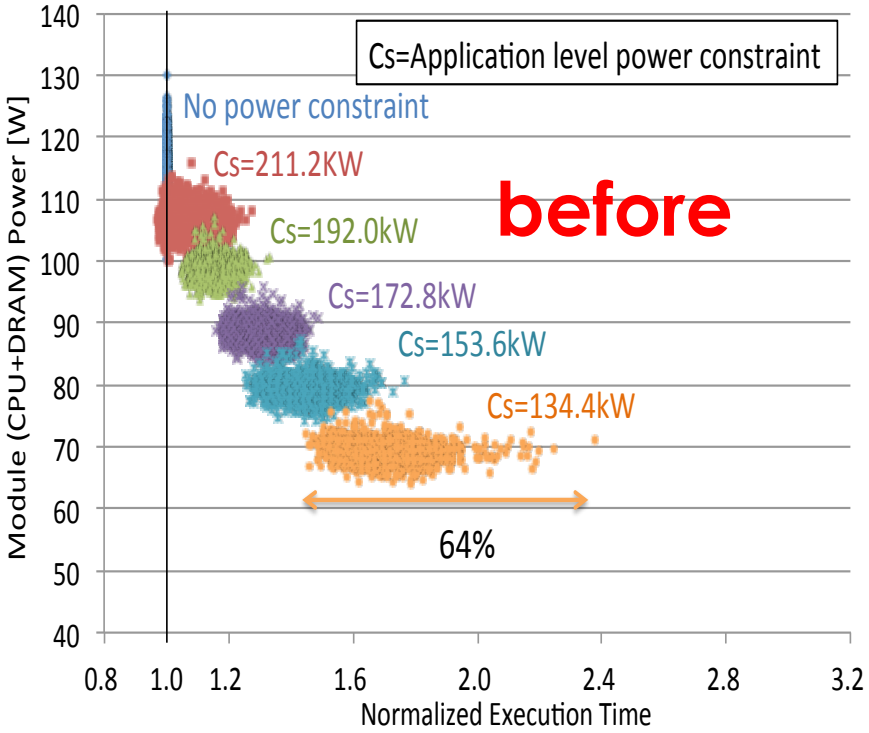
Tested Power Budgeting Methods

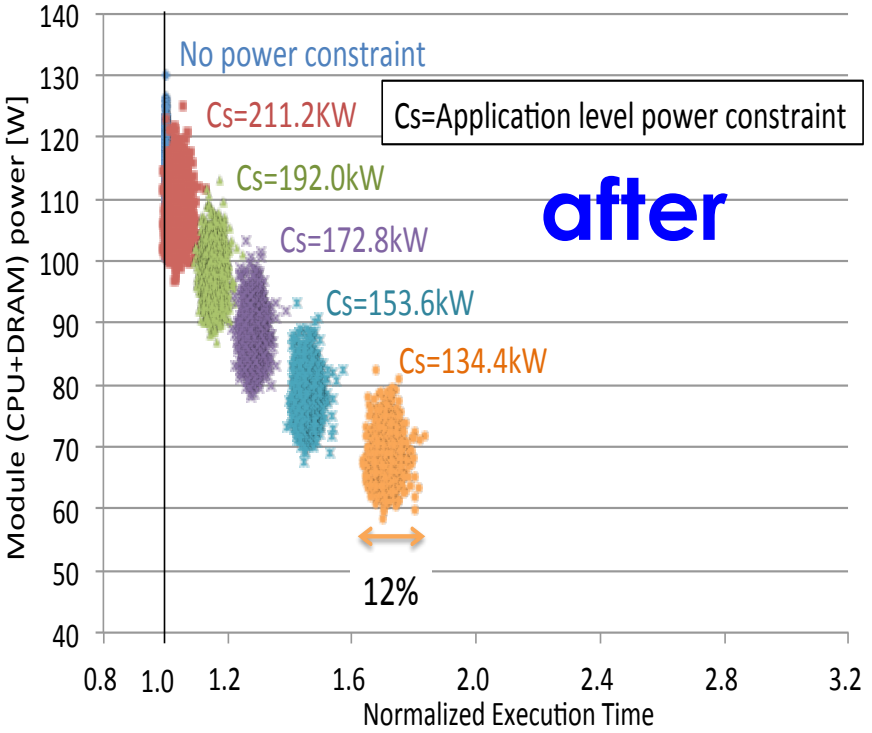
Method Name	Application Specific	Variation Aware	Power-Performance Model	Pwr. Set.
Naive	No	No	-----	Power Cap
Pc	Yes	No	Calibration	Power Cap
VaPc	Yes	Yes	Calibration	Power Cap
VaFs	Yes	Yes	Calibration	Freq. Sel.
VaPcOr	Yes	Yes	Oracle	Power Cap
VaFsOr	Yes	Yes	Oracle	Freq. Sel.

Va=Variation-Aware, Pc=Power Capping, Fs=Frequency Selection
Or=Observed power data are used

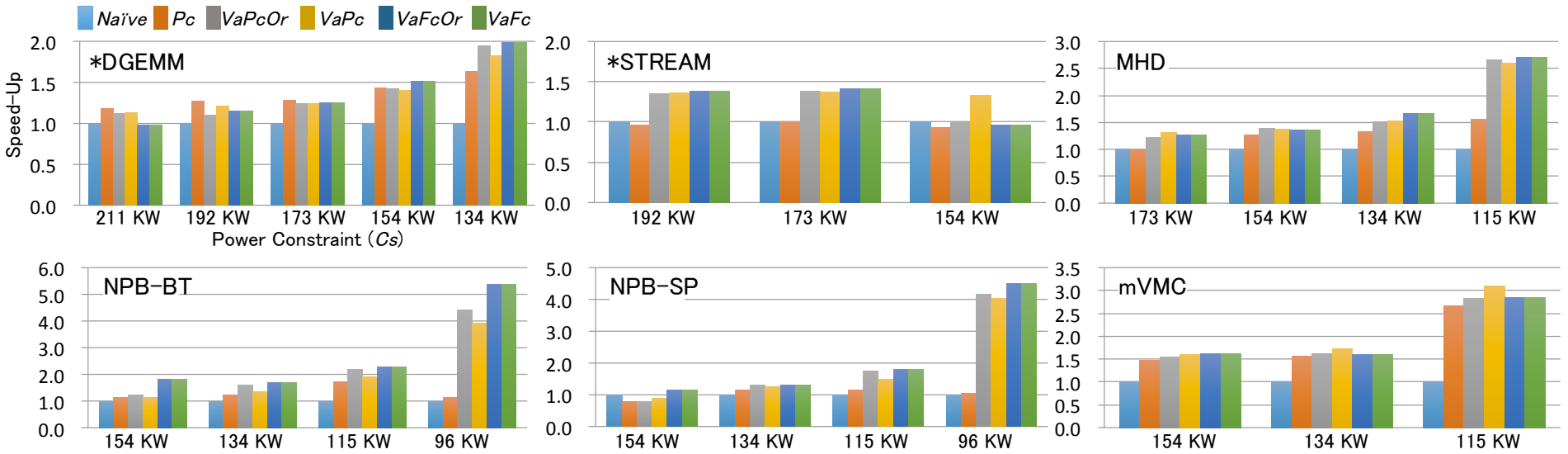
Speedup Ratios Normalized to Naïve (star DGEMM on 1,920 modules)







Speedup Ratios Normalized to Naïve (All results on 1,920 modules)



5.4X speedup at maximum (NPB-BT)

1.8X speedup in average

KYUSHU UNIVERSITY

Conclusions

Power constrained computing becomes main-stream!

Manufacturing variability causes serious performance issue!

Optimize power resource allocation!

KYUSHU UNIVERSITY

Acknowledgements

- This research was supported by JST CREST.
- Special thanks to Dr. Yuichi Inadomi, Prof. Masaaki Kondo, Dr. Tapasya Patki, Dr. Martin Schulz, and other all members of this project.