Actual FLOPs/watt to evaluate total operation efficiency of computing centers

June 22, 2017

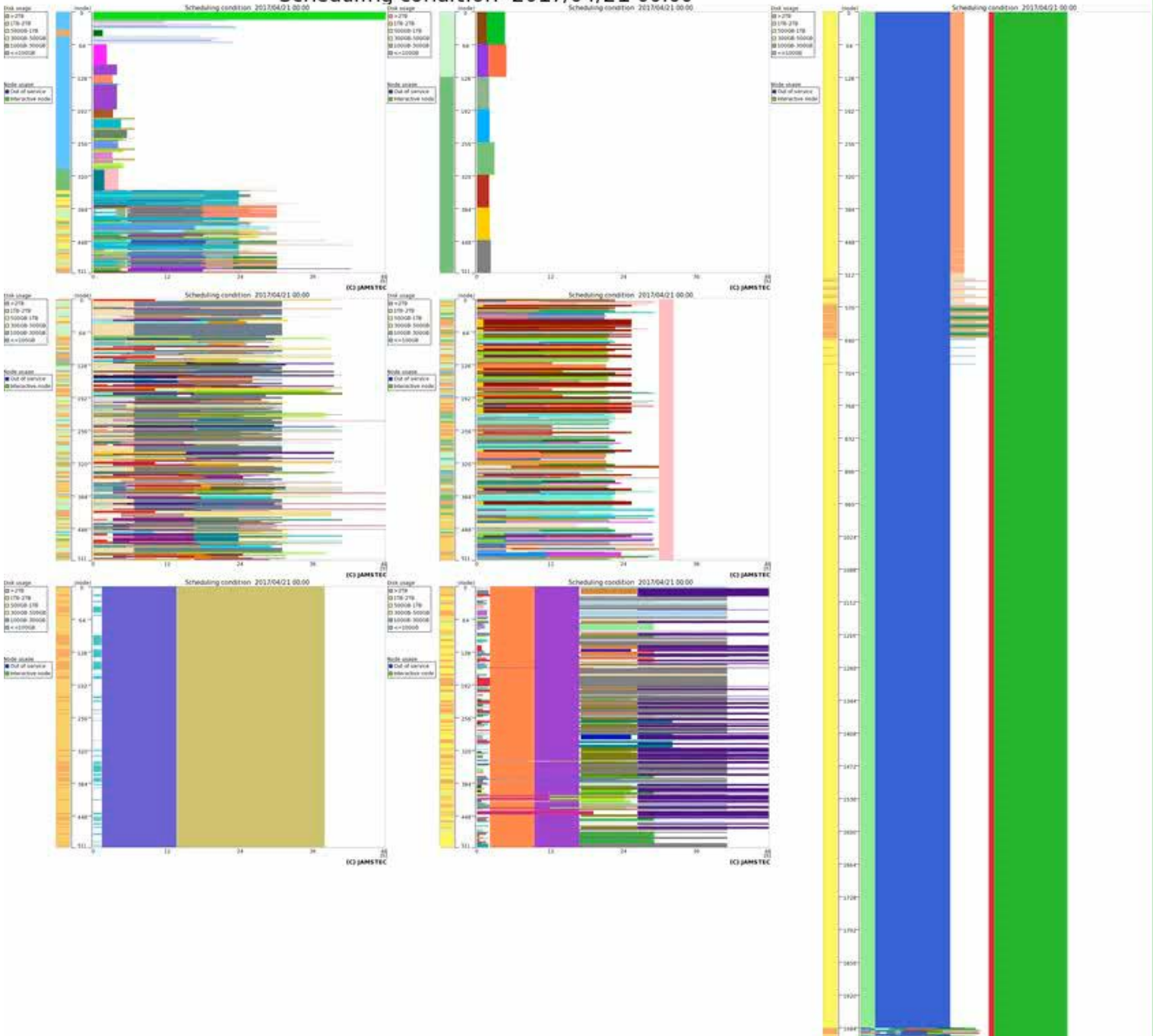
Makoto Tsukakoshi

Information Systems Department

Center for Earth Information Science and Technology

JAMSTEC

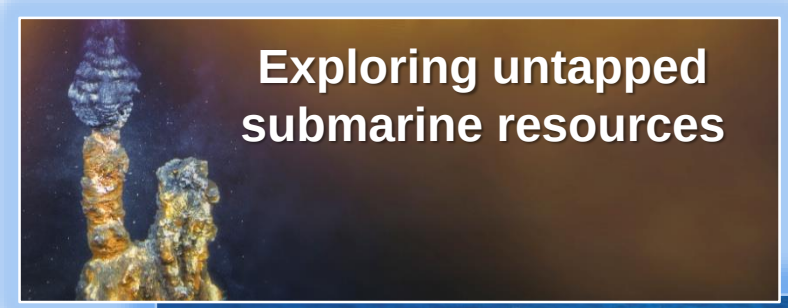
Scheduling condition 2017/04/21 00:00

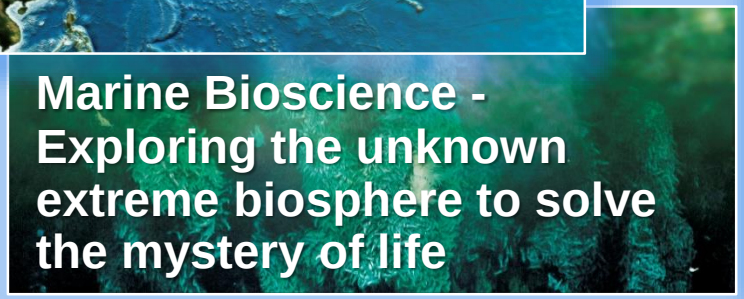


Japan Agency for Marine-Earth Science and Technology

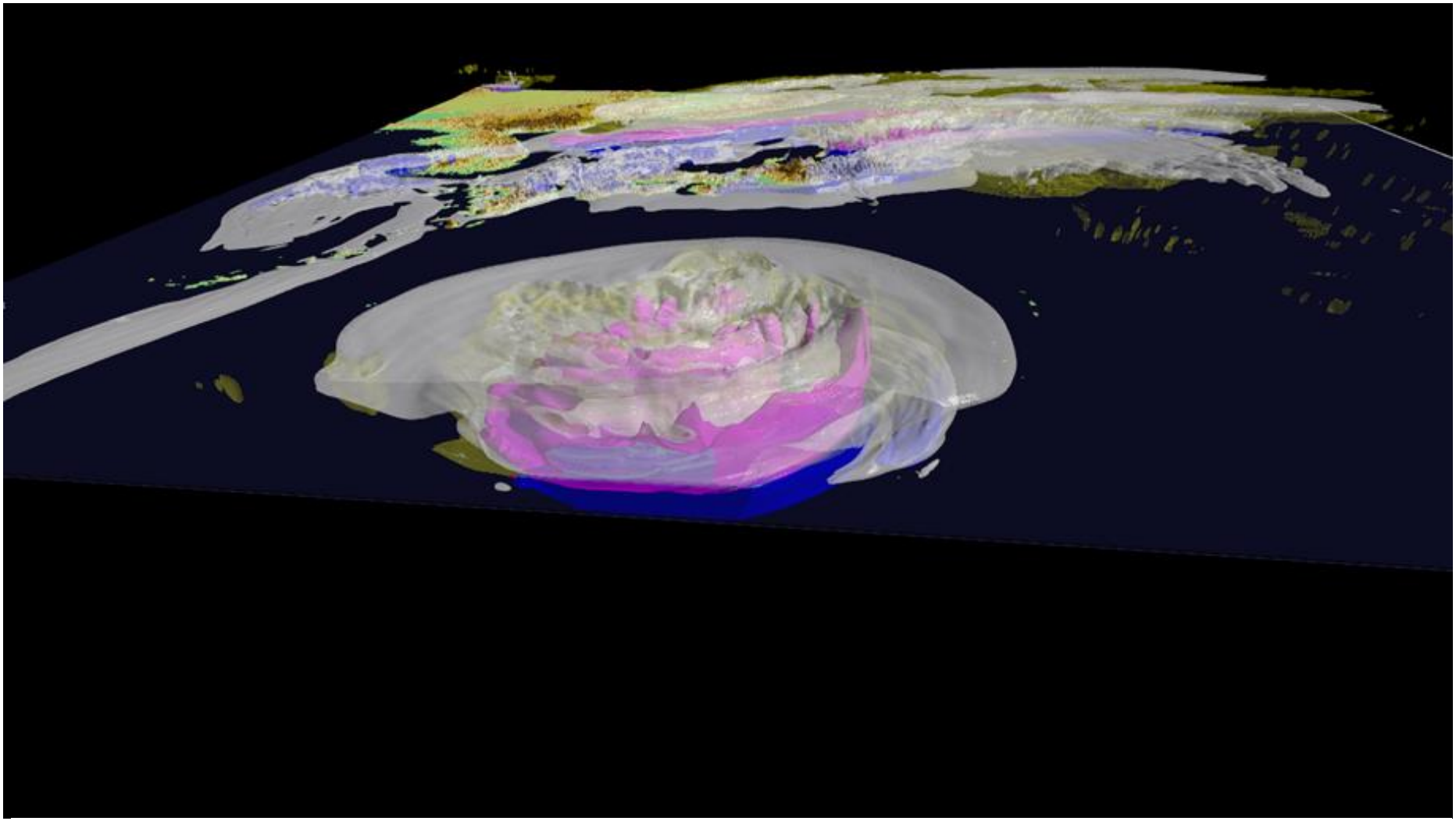
The main seven research and development issues during the third mid-term plan

During the third mid-term plan, we set and address the seven research and development issues with all our strength due to promote strategic and focused research and development based on the national and social needs.





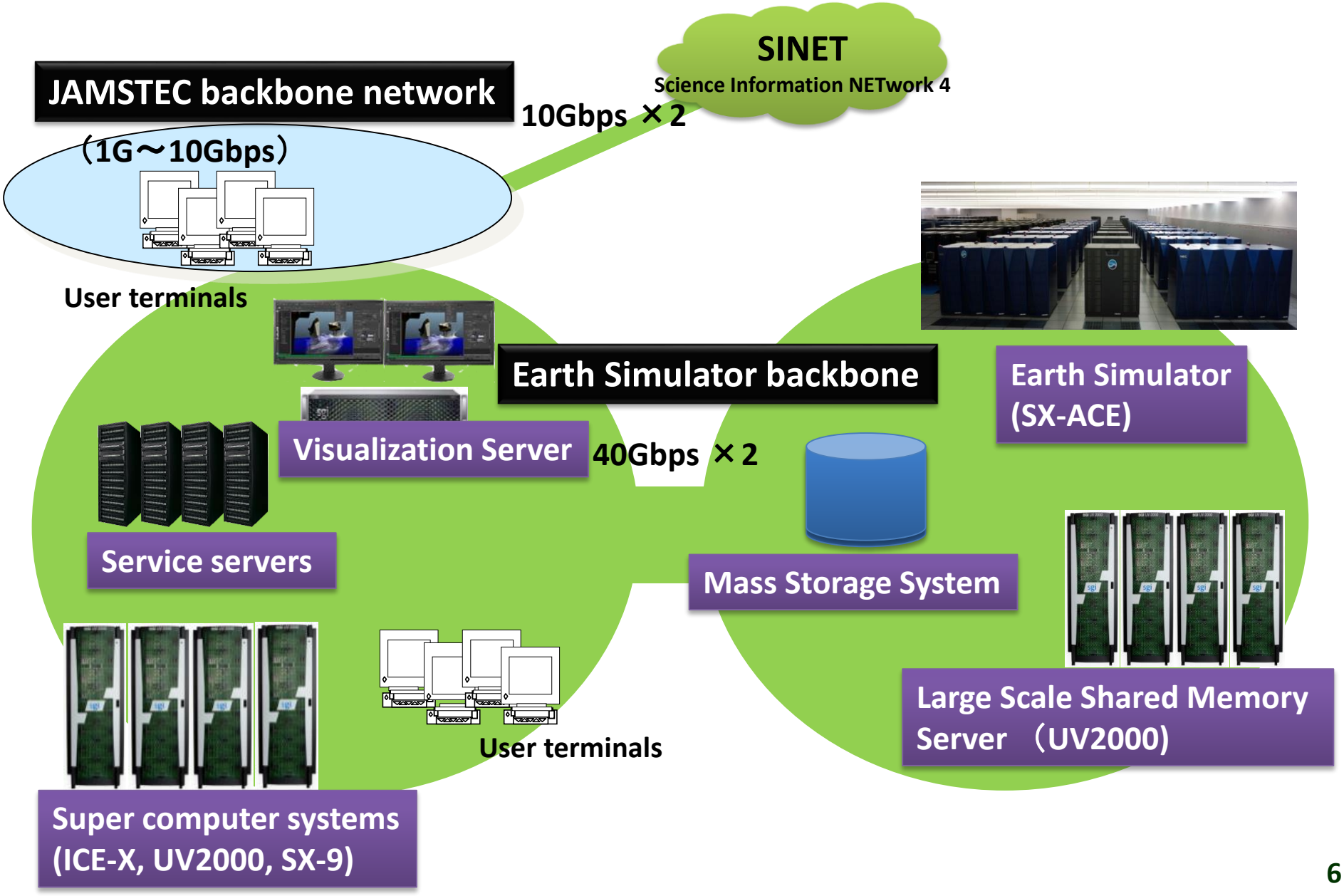
Typhoon-marine interaction using nonstatic atmospheric wave ocean coupling model



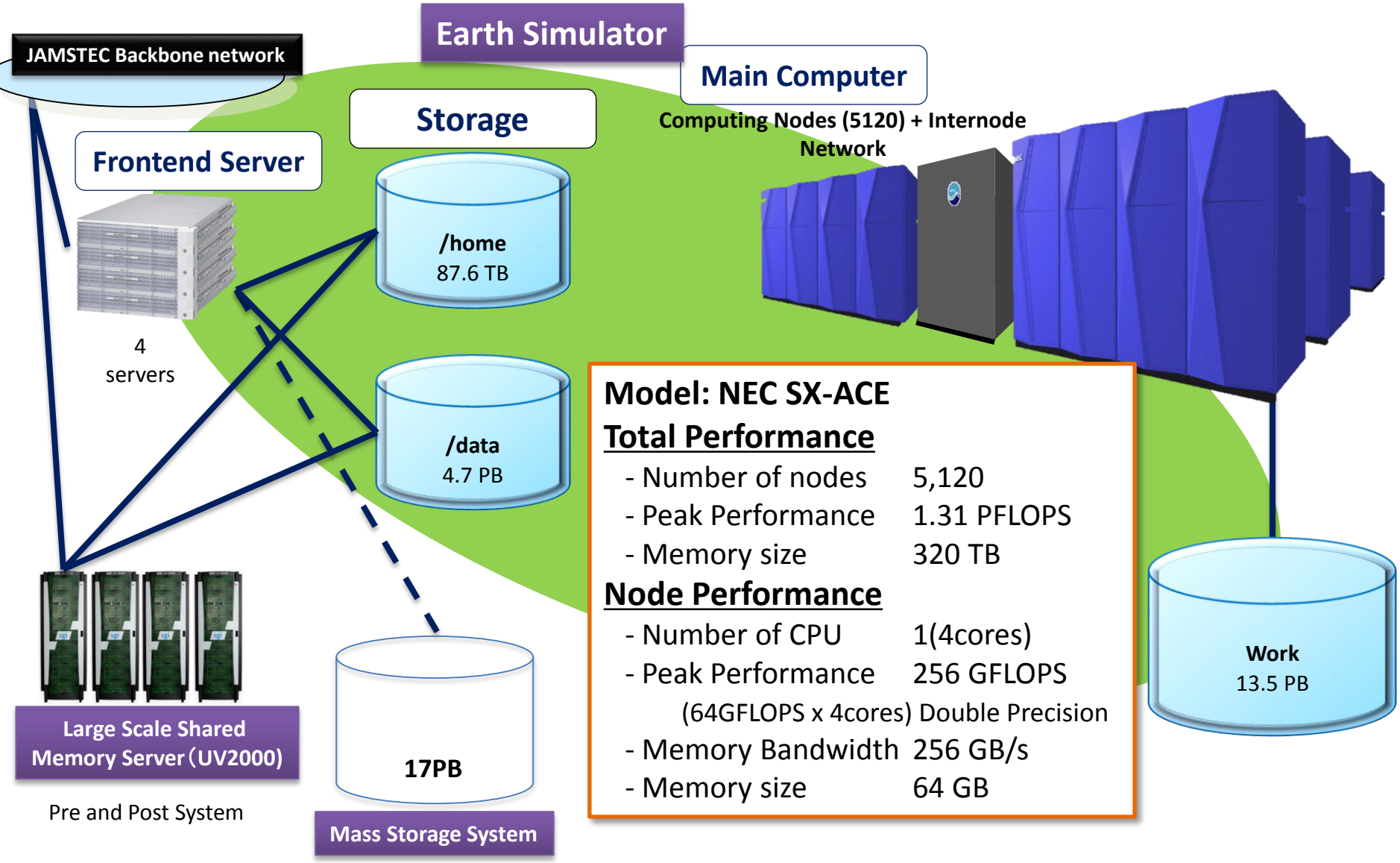
Typhoon Vera in 1959 by Kazuhisa Tsuboki (2015)

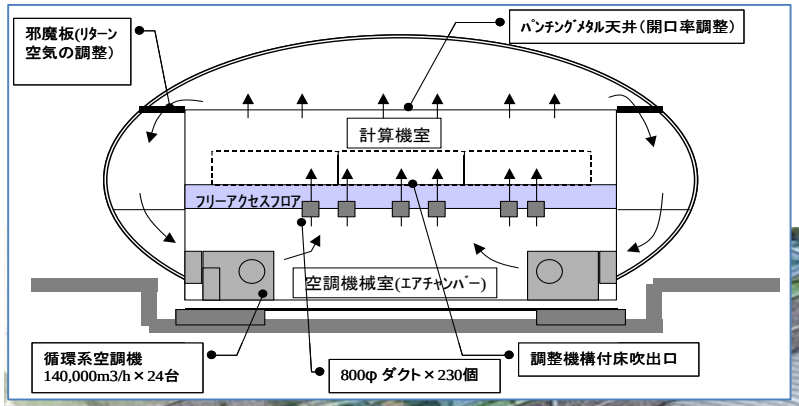
Typhoon-Ocean Interaction Study Using the Coupled Atmosphere-Ocean Non-hydrostatic Model: With Careful Consideration of Upper Outflow Layer Clouds of Typhoon

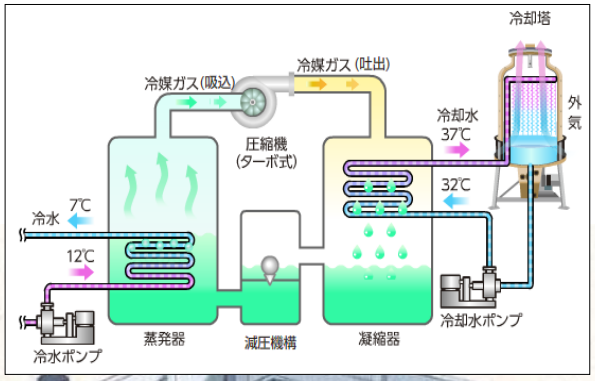
JAMSTEC Information System Outline

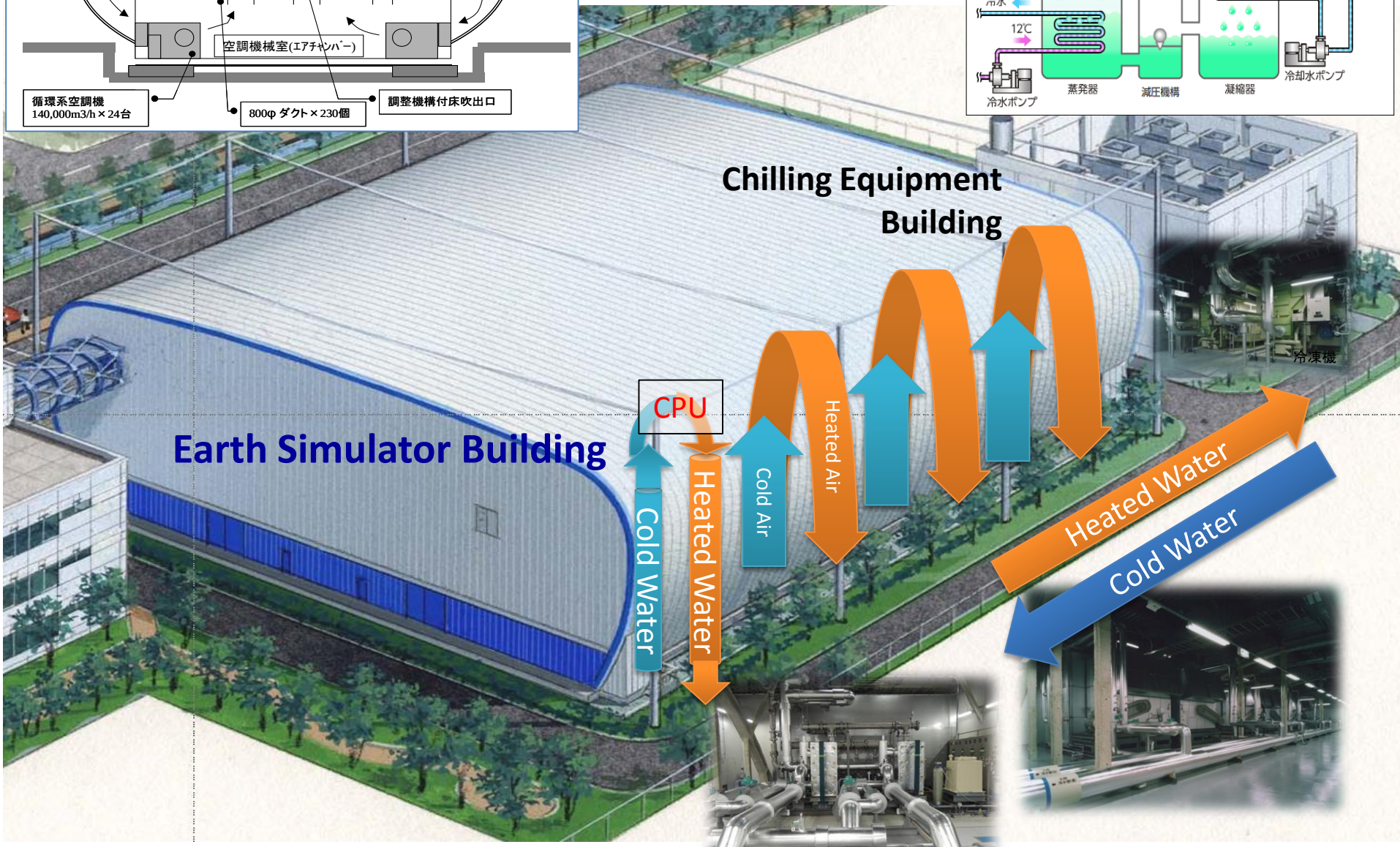


Earth Simulator Outline




邪魔板(リターン空気の調整)
パンチングメタル天井(開口率調整)
計算機室
フリーアクセスフロア
空調機械室(エアチャンバー)
循環系空調機
140,000m3/h×24台
800φ ダクト×230個
調整機構付床吹出口


冷却塔
冷媒ガス(吸込)
冷媒ガス(吐出)
圧縮機
(ターボ式)
冷却水
37℃
32℃
外気
冷水
7℃
12℃
冷水ポンプ
蒸発器
減圧機構
凝縮器
冷却水ポンプ


Chilling Equipment Building
Earth Simulator Building
CPU
Cold Water
Heated Water
Cold Air
Heated Air
Heated Water
Cold Water
冷凍機

Outline of Power Monitoring System

windows server

IP Network

SNMP Manager

Monitor
CEC
SNMP Adopter ND-FA02

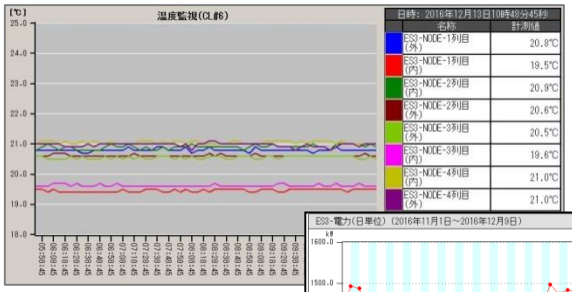
Modbus

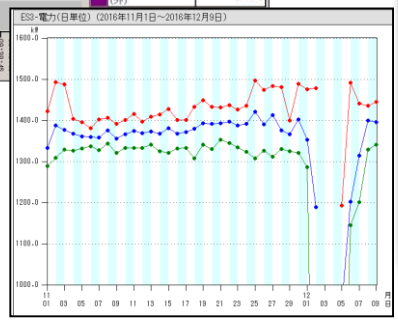
Monitoring Software: CEC SND-UM01

Panasonic
EcoPowerMeter
KW2G-H

SD Card
32GB

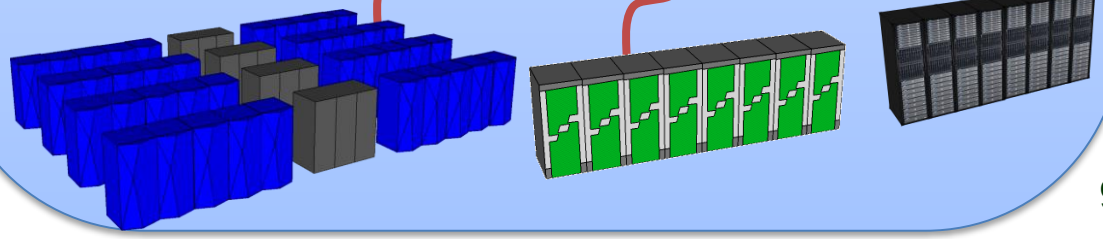
Distribution board





EcoPowerMeter
Accumulate :
1GB/Month

Current sensor (split type; 100 A or 50 A)



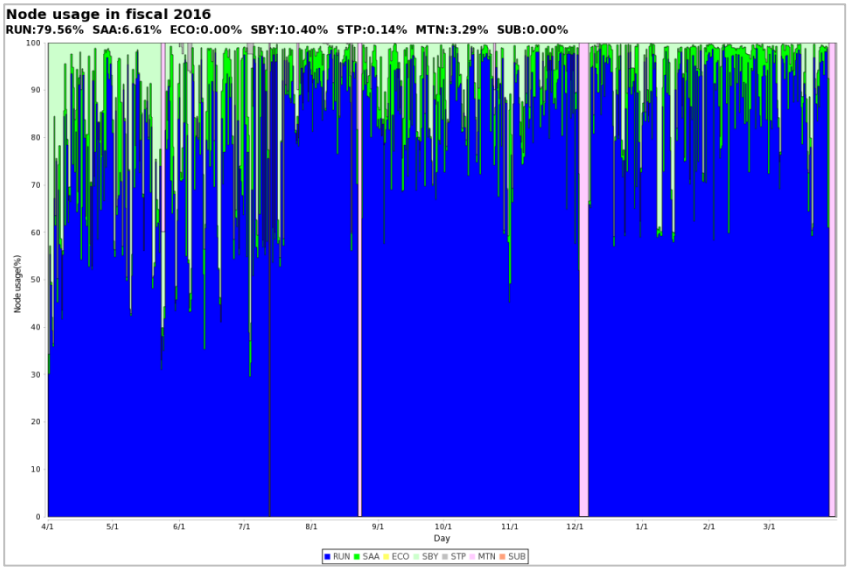
Computation Statistics of Earth Simulator (ES2 vs New ES)

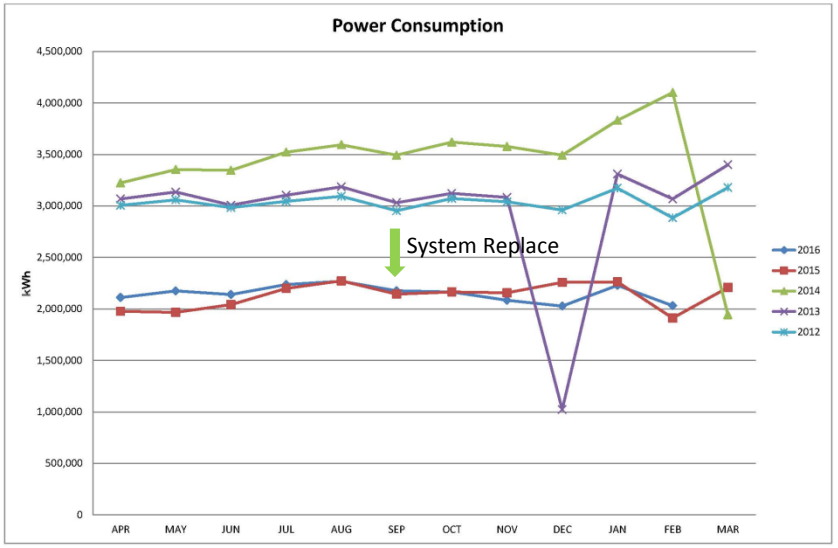
SYSTEM	ES2(2012)	New ES	Ratio
Total #JOB/Year	74,244	227,024	
Total #JOB/Year(Normalized*)		320,512	4.3
Ave. #nodes of JOB (#CPU)	5(40)	32(32)	
Ave. Memory size/JOB (GB)	640	2,048	3.2
Ave. I/O per JOB (byte)	175,427,617,865	374,632,801,438	2.1
Ave. Performance of JOB(GFLOPS)	240.28	663.56	2.8
Ave. Efficiency of JOB (per peak)	5.65%	8.08%	
Total floating operations/Year (GFLOPs)	260,638,065,109	2,055,356,205,065	
Total floating operations/Year (Normalized*) (GFLOPs)		2,901,679,348,327	11.1

* New ES was half system in April and May in 2015. And the statistics is as of January 15 2016. Normalization was done proportionally.

Our fundamental efforts in supercomputer operation
(my key index and issues)

- **Maintain system availability**
  Prevent HW/SW failure. Minimize down time. ***99.86% in FY16***
- **Keep high utilization**
  Optimize operation policy, scheduling and user environment.
  89.07% in FY16
- **Keep high performance**
  Optimize user programs by consultation and tuning.
- **Lower power consumption**
  Although facilities are somehow *fixed*, we have been trying to optimize air flow and temperature.





(proposal) Use actual FLOPs per watt to evaluate total energy efficiency of computing centers

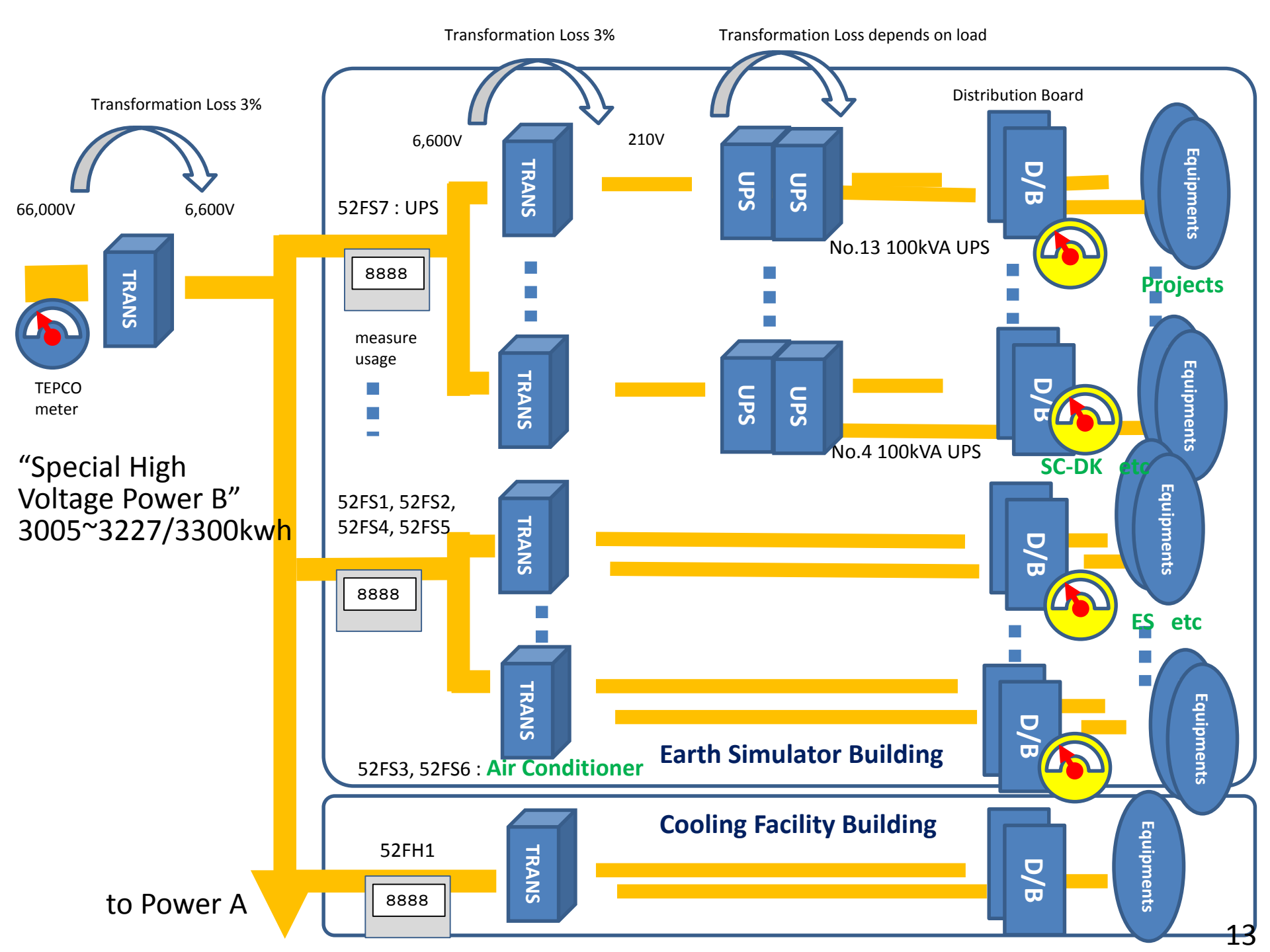
- **Efficiency should be "actually achieved *Performance/Total energy consumed*"**
- *Peak* or *Linpack* Performance : Don't represent real work loads ("Green500")
- *HPCG* or other *Benchmark* suites : Better but do not fit each center's AP spectrum (each centers should have been optimized to its own application work loads.)
- Benchmarks show only capability. Do not represent the center's operational efforts.
- Power consumption must be actual and total include cooling.

- **Evaluate the energy efficiency by "*Executed FLOPs per watt*" !**
- **Where, the energy is measured total inclusive amount.**

(Example)

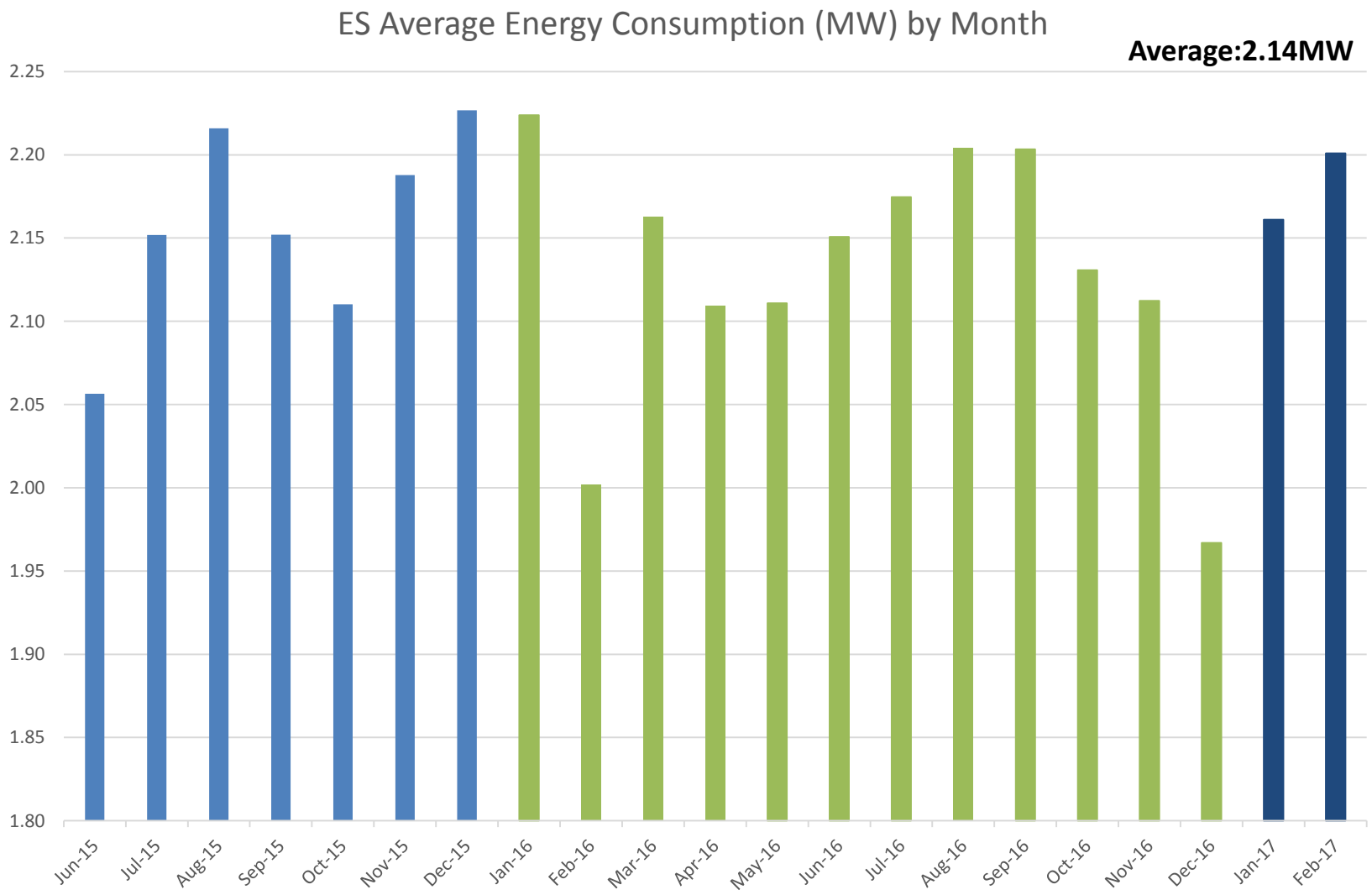
In 2016 (JAN-DEC), Earth Simulator (Entire System including cooling facilities) has actually achieved 178,956 GFLOPS/Kwh (7.6% of theoretical peak x 100% non-stop operation)

Using HPCG/peak performance ratio (5.2%) and annual GWh from the annual report 2014, efficiency of "K-Computer" is estimated 109,221 GFLOPS/Kwh


Transformation Loss 3%
Transformation Loss depends on load
Transformation Loss 3%
66,000V
6,600V
TEPCO meter
TRANS
"Special High Voltage Power B" 3005~3227/3300kwh
6,600V
210V
Distribution Board
52FS7 : UPS
8888
measure usage
UPS
D/B
Equipments
No.13 100kVA UPS
Projects
No.4 100kVA UPS
SC-DK etc
52FS1, 52FS2, 52FS4, 52FS5
ES etc
52FS3, 52FS6 : Air Conditioner
Earth Simulator Building
Cooling Facility Building
52FH1
to Power A

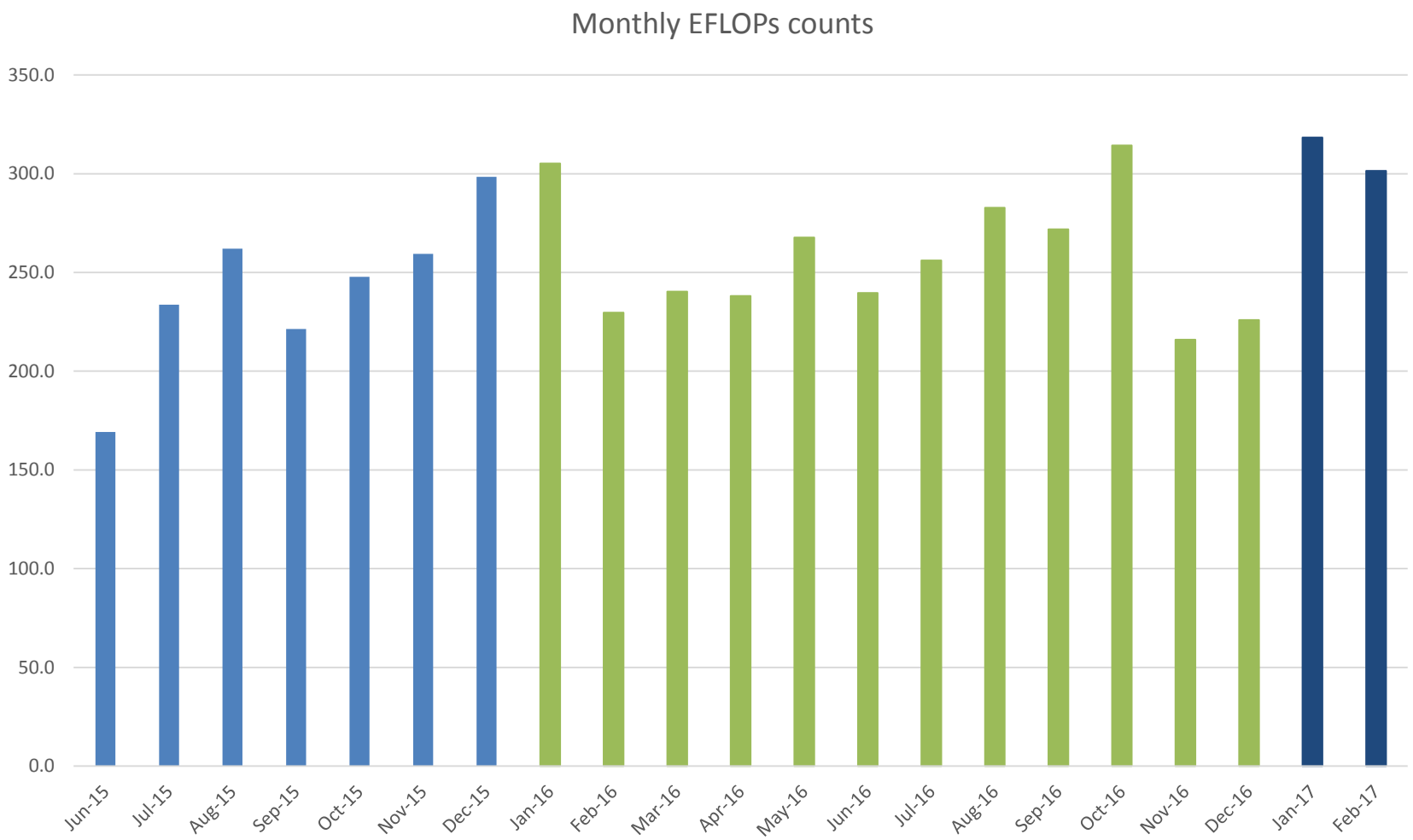
Power Consumption of Entire ES System

including Storage, Air Conditioner and Chiller

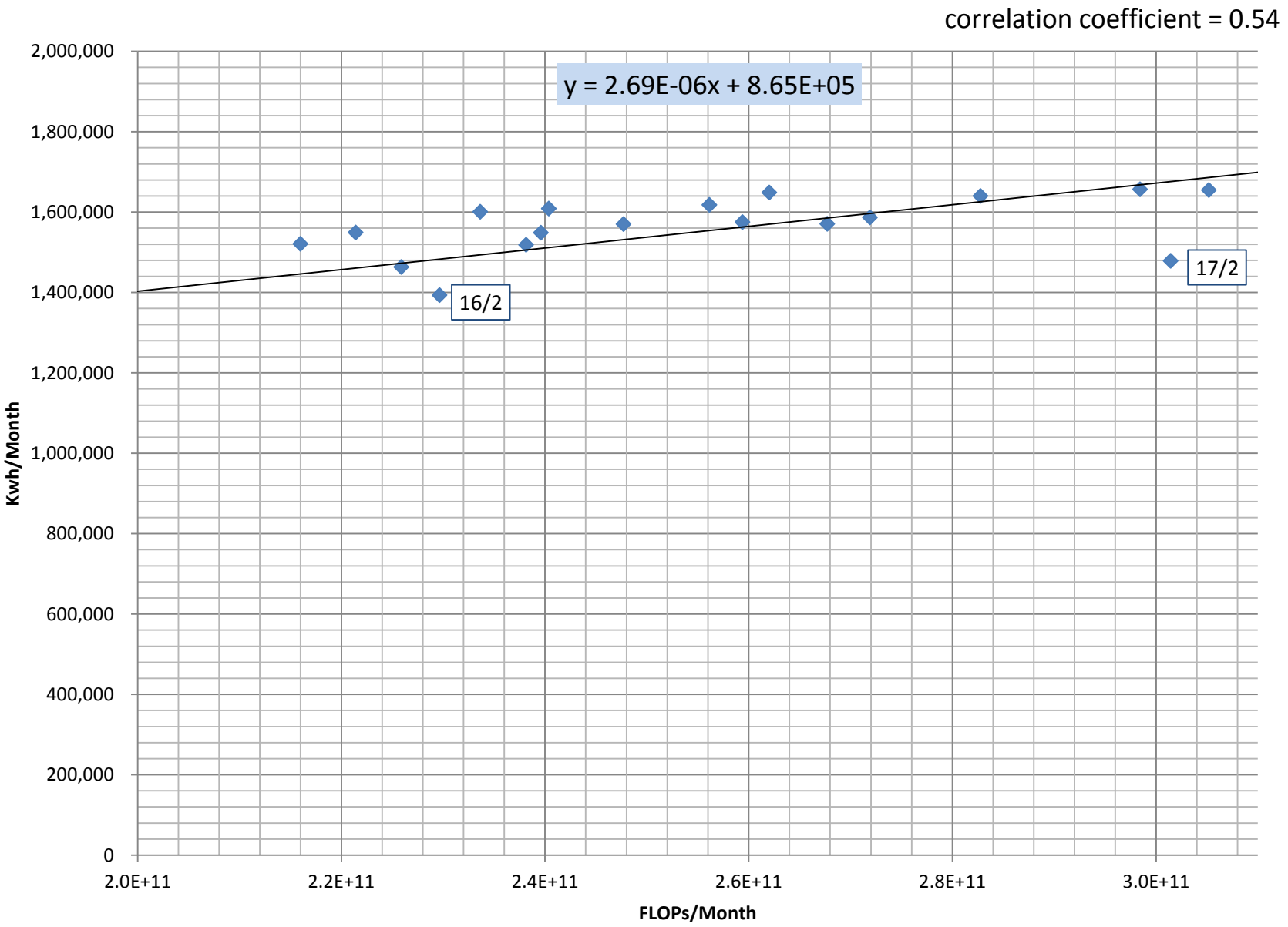


Monthly Exa FLOPs executed on ES System

Average:254.9EFLOPs

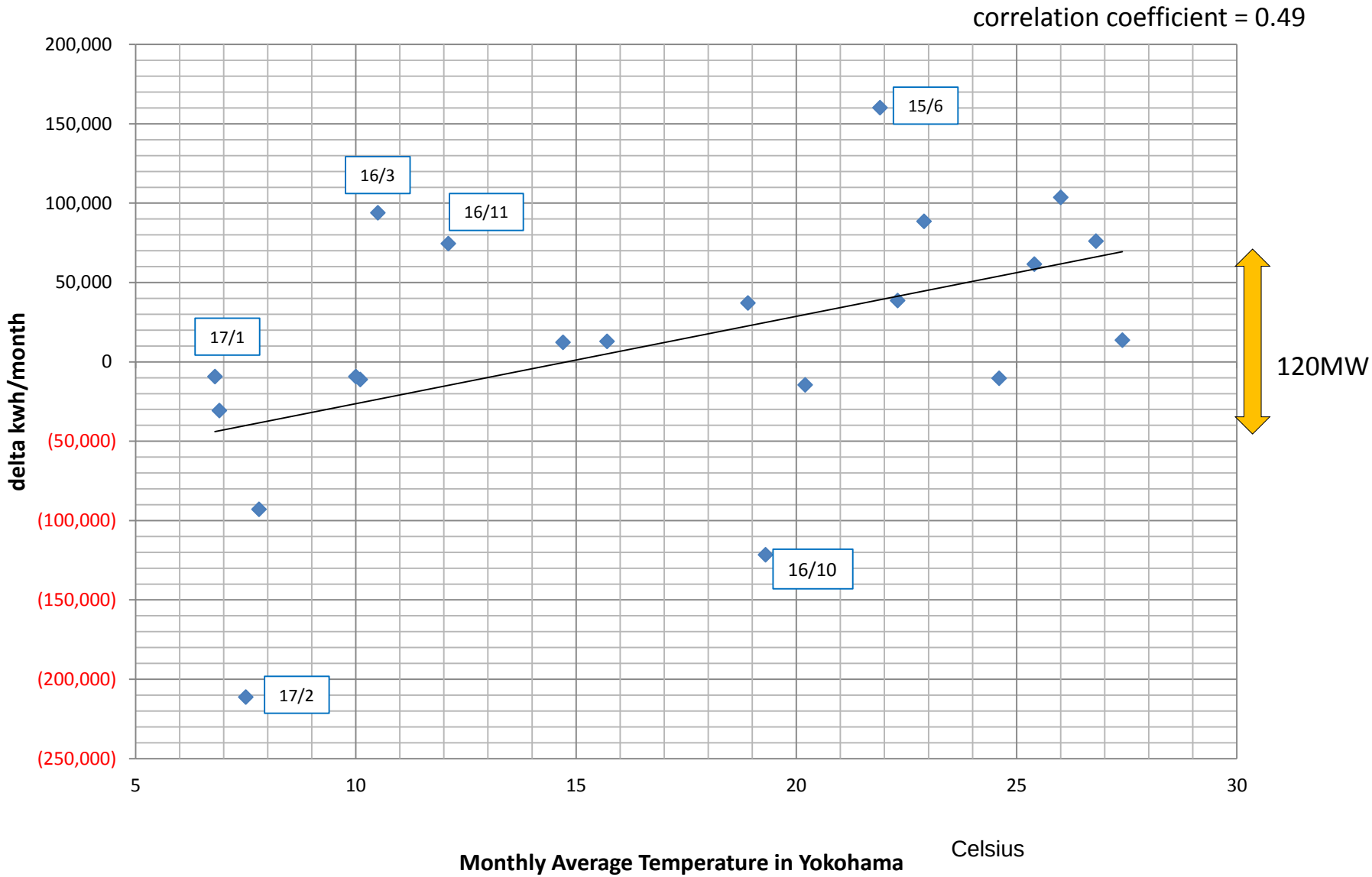


Monthly Power Consumption and FLOPs



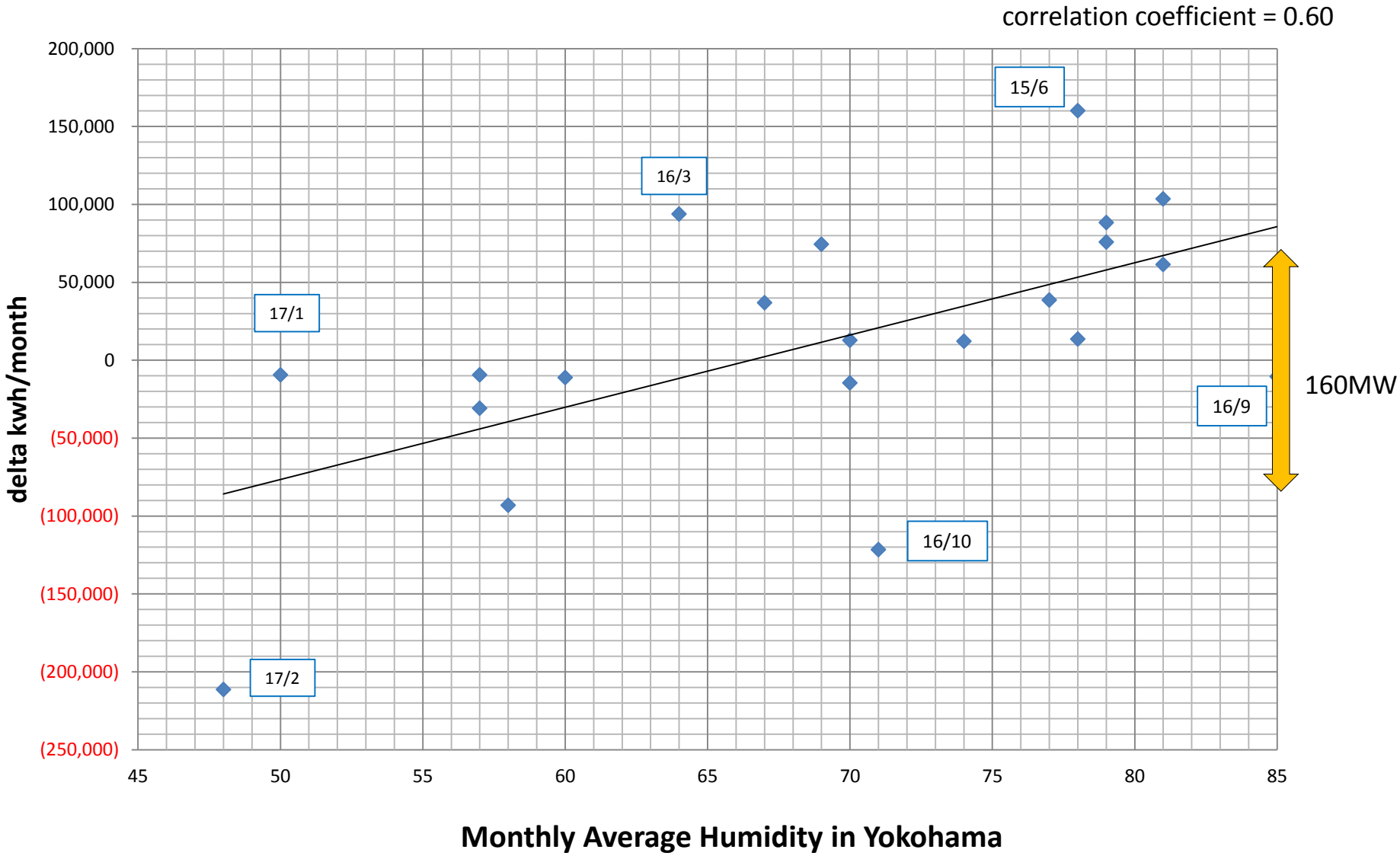
Analysis of *delta*(t,h)

assuming power consumption = a * FLOPs + b + *delta* (temperature, humidity)

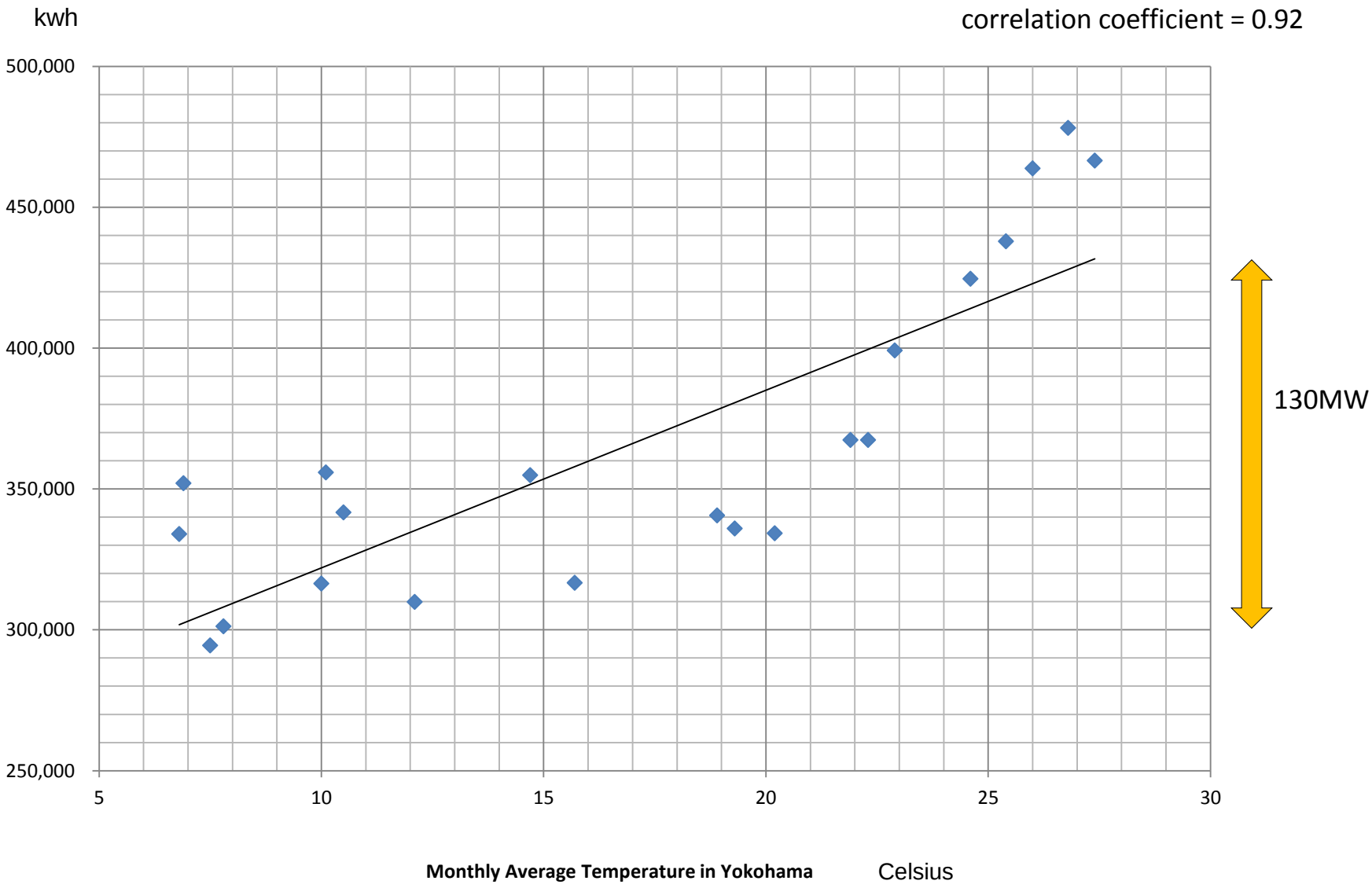


Analysis of *delta*(t,h)

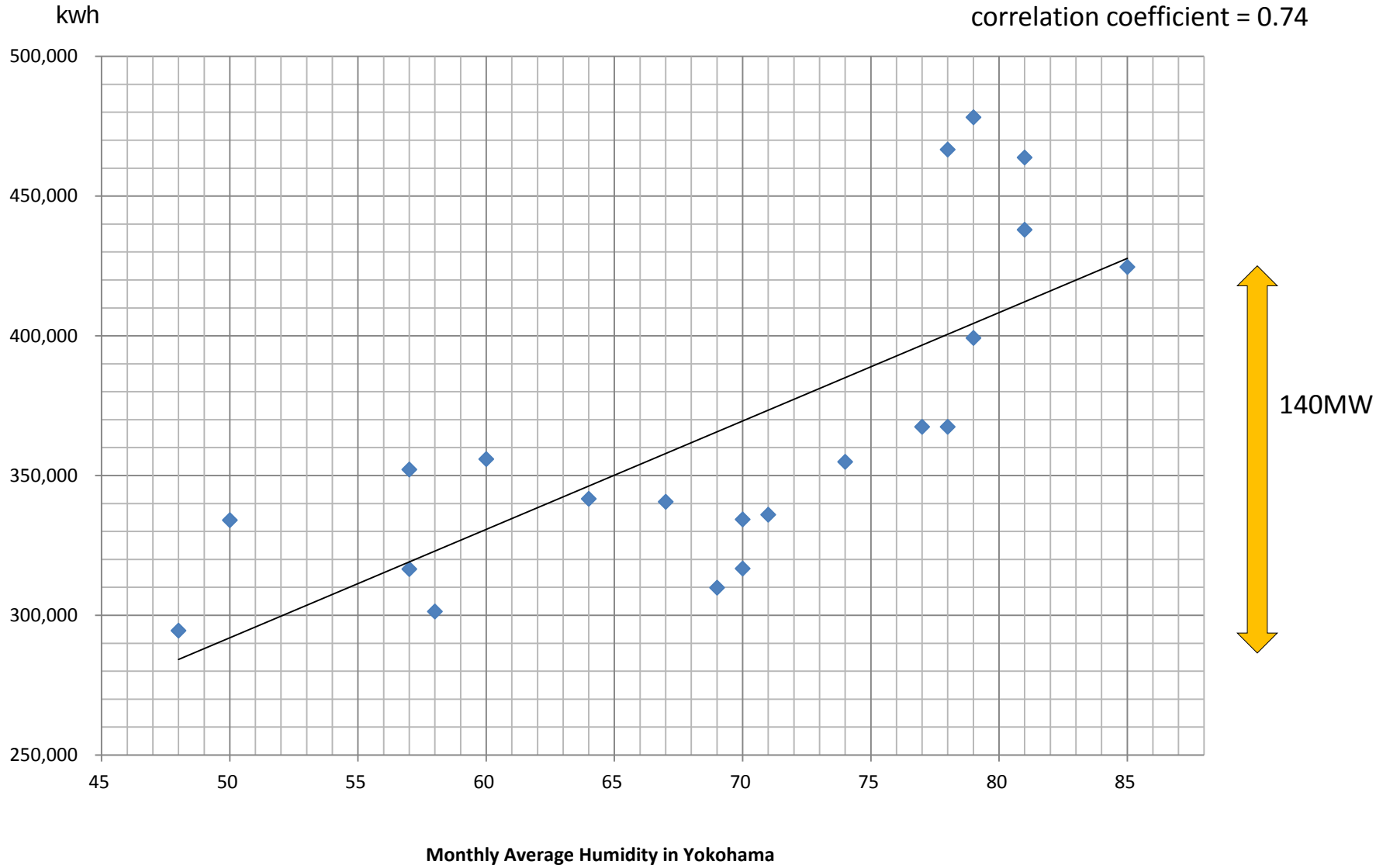
assuming power consumption = a * FLOPs + b + *delta* (temperature, humidity)



Power consumption of Chiller – temperature by Month

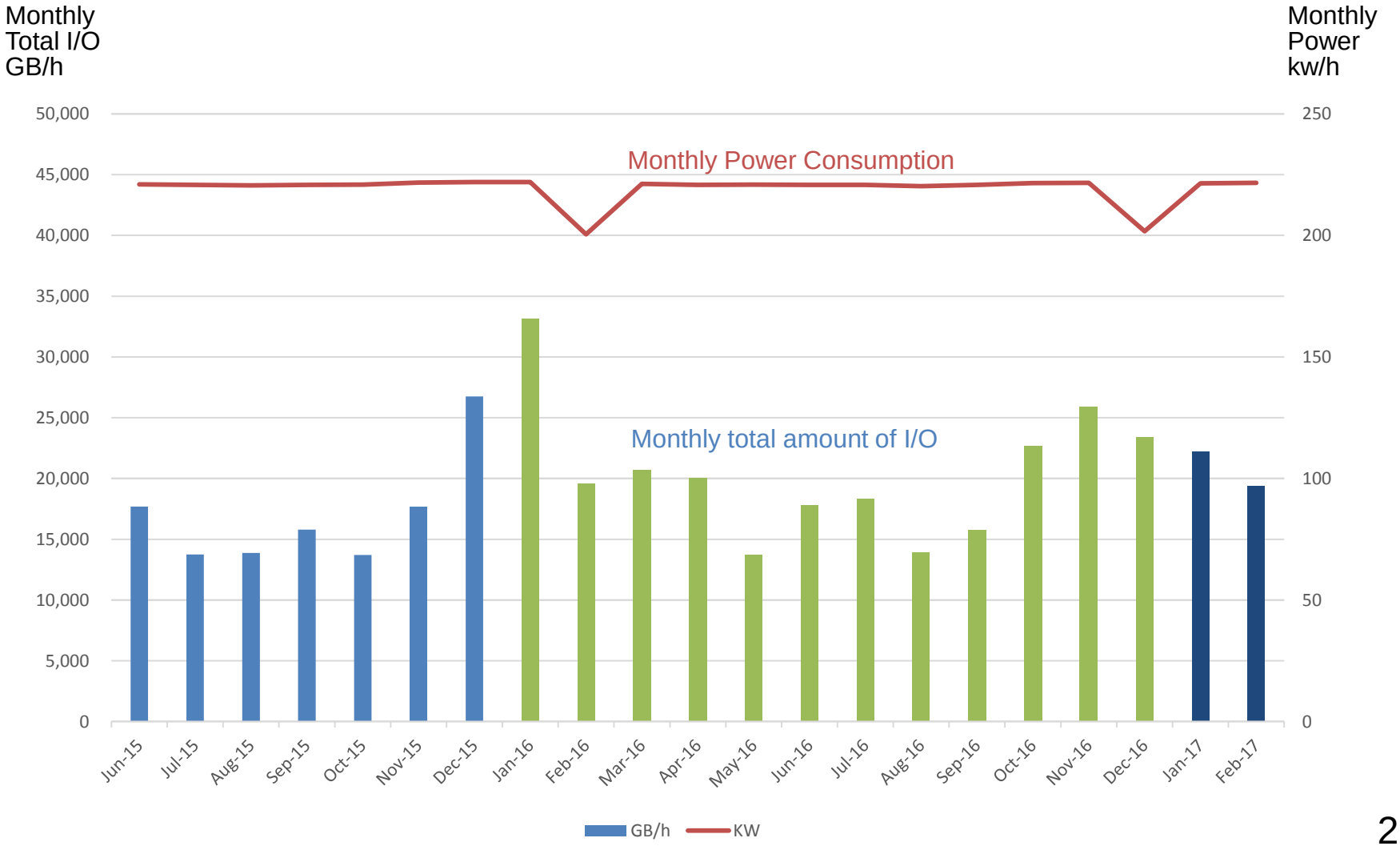


Power consumption of Chiller – humidity by Month



Power consumption of Peripherals and amount of I/O

– Power consumption is consistent: 221kwh per hour. (Not using MAID function. It drops in the month of maintenance.)



power consumption, FLOPs and temp/humidity

- Clearly, power consumption is complex sum of arithmetic units, cache, memory and so on, however, Earth Simulator data shows there is a correlation between the power consumption and actual FLOPs.
- In Earth Simulator Data, assuming:

  *Power consumption = a * FLOPs + b + delta (temperature, humidity)*

  and the data indicates *delta* comes from the chiller system.
- With average 1540Mwh/month, temperature and humidity give 8 to 10% difference.
- In Yokohama, monthly average temperature is 5.9 to 26.7 degree and humidity is 30% to 78%. Probably, An other location in northern Japan has 8 to 10% energy efficiency advantage.

Summary and issues

- Power consumption and Actual FLOPs counts are useful to analyze data center operation efficiency.
- Challenge – NEC SX-ACE has hardware function to record FLOPs. How to do it with Intel architecture and others ? The difference of HW FLOPs and AP FLOPs.
- Questions – How to deal with single, half precision or integer and logical operations ?