Holistic Cost Analysis of Running a Computing Center

Zhangqier Wang

Throughput Computing 2025

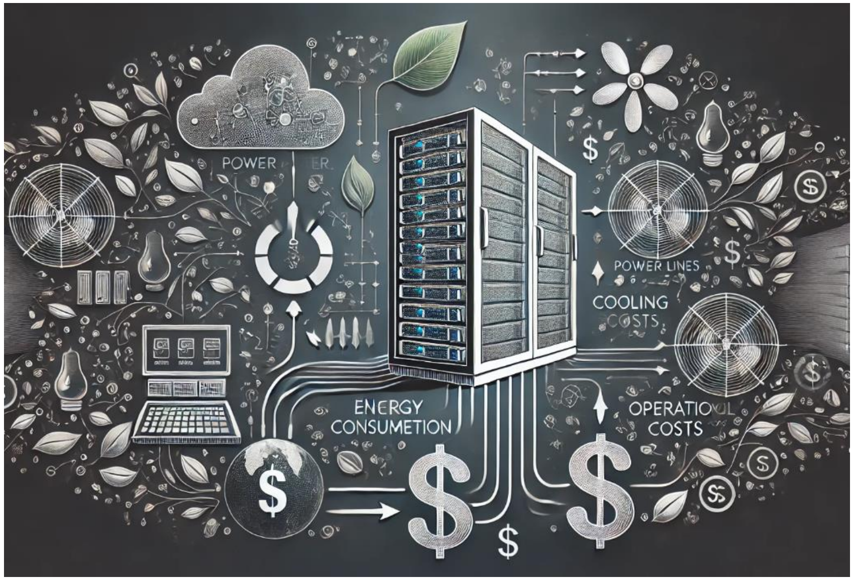
June. 4th 2025

Outline

- Introduction
- Holistic cost analysis on CMS T2
  - Estimate the cost of each type of hardware
  - Policy for procurement plan
- Cost analysis for a general computing center
  - A general way to form the policy

Introduction

- The increasing volume of data (ex. HL-LHC), new physics exploration, and AI applications are driving a surge in computing resource requirements
- Beyond performance metrics, more factors to consider
  - Computing demand is increasing, leading to high power consumption
  - **Power consumption** including cooling, a key contributor to operational expenses
  - Low-power hardware advancements exist, but the cost-benefit of switching is unclear.
- Computing in general faces a major re-design to align with energy efficiency and sustainability goals.



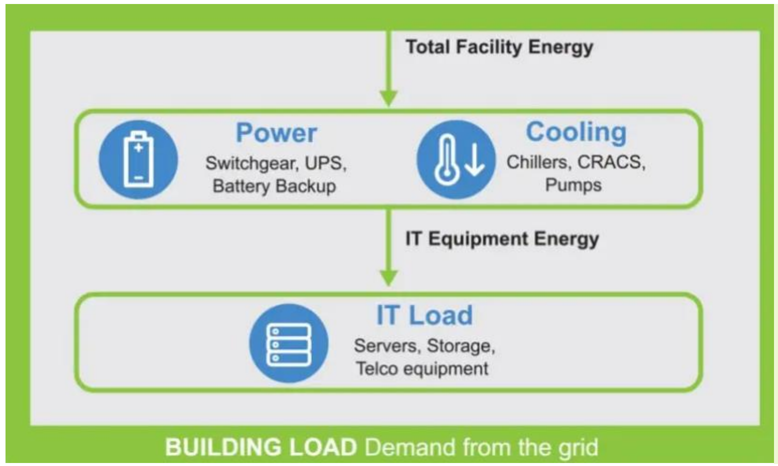
HTC25

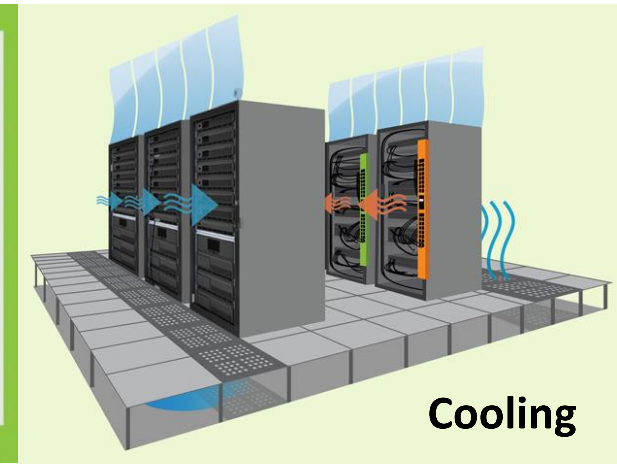
Cost compositions

- **Cost of computing is complicated**, hardware purchase is just a fraction of it
  - Including power usage, cooling, space usage.

Hardware Purchase

Racks: on average ~$5000/year

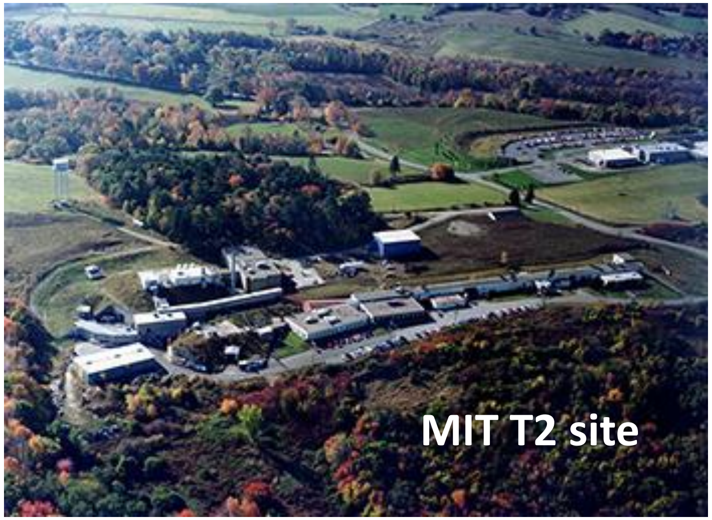




PUE = Total Facility Energy / IT Equipment Energy

MIT Tier 2 Center

- The MIT Tier-2 Center is a high-performance computing facility dedicated to processing, storing, and analyzing data for the CMS, LHCb, and other experiments
  - Comprises approximately **700 machines**, providing **25k CPU cores** and **16.5 PB of storage**.
  - Compute/Storage mix model
- Worker nodes are also used as storage devices
  - Re-design to have dedicated compute and storage servers
- Cost evaluation determine hardware retirement policy
  - Prepare MIT T2 for HL-LHC (data x10)
  - Improve energy efficiency
  - Spend less to provide same amount of computation



HTC25

Holistic Cost Analysis on MIT T2

Machine Categories

- CPU models categorized into 8 types
  - power consumption, and cpu, memory usage are checked
  - average year represents the age of the machine

CPU model	Avg year	Production Process	HS06	Cores	HS06/core
Intel(R)_Xeon(R)_E5310-5410	2008-2013	65 nm	69	8	8.6
Intel(R)_Xeon(R)_X5647	2017	32 nm	155	16	9.7
Intel(R)_Xeon(R)_E5520-5620	2018	45 nm	120-140	16	8.1
Intel(R)_Xeon(R)_E5-series	2018	14/22 nm	169-449	8-40	11.1
Intel(R)_Xeon(R)_Silver	2019	14 nm	530-706	48-64	11.0
Intel(R)_Xeon(R)_Gold	2021	10 nm	904	64	14.1
AMD_EPYC_9754_128-Core_Processor	2023	5 nm	7450	512	14.6

Power Cross-check

- The power consumption is monitored using “ipmitool” and “omreport”.
- The current is measured on two servers and compared to the current from the monitoring
  - Measured using clamp meter and AC splitter
  - Load CPUs using linux stress command “stress --cpu N --timeout 100”
  - Load CPU with CMS actually process using 16/32 cores.

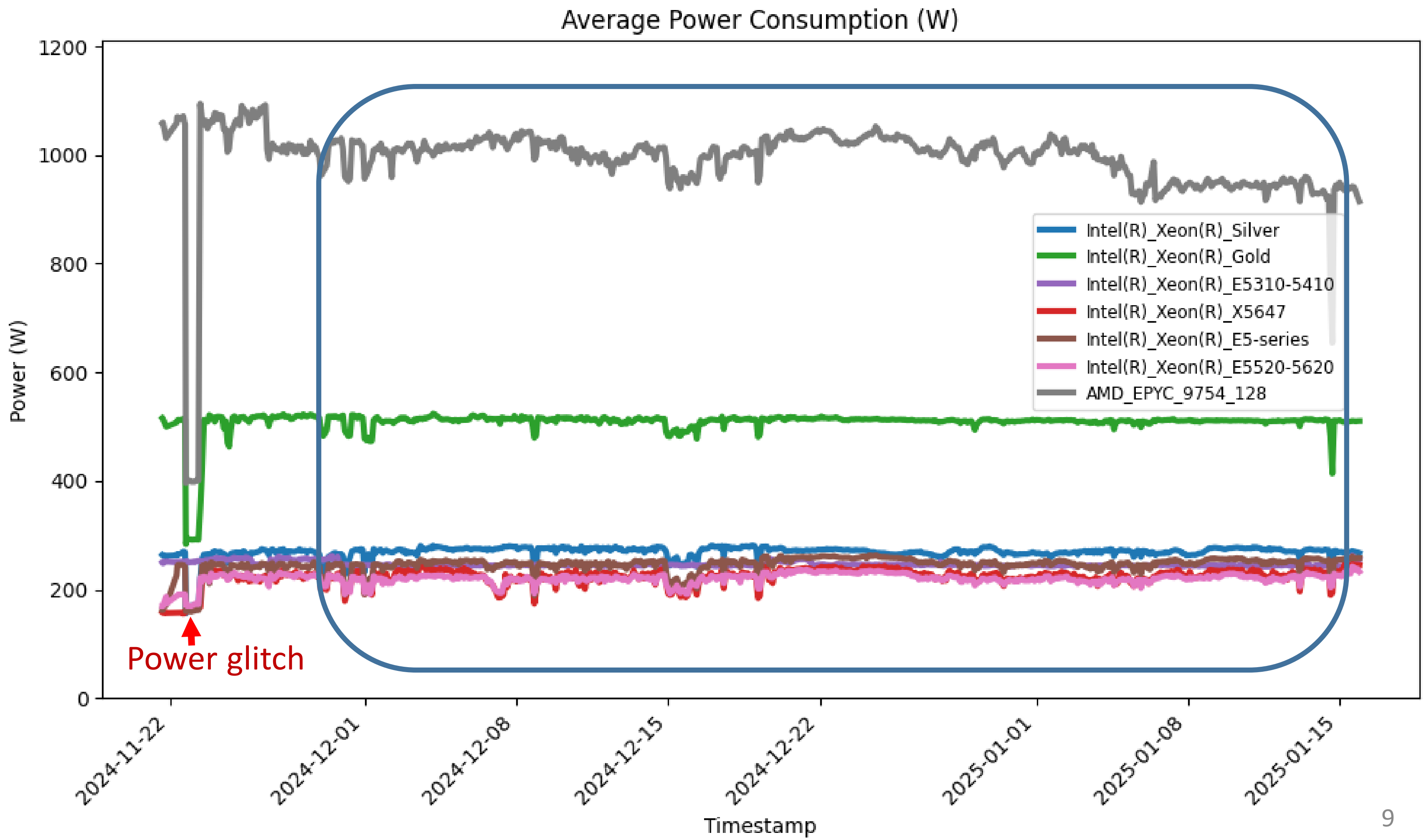
Current (A)		Base	16 cpu	32 cpu	48 cpu	64 cpu	CMS 16 cores	CMS 32 cores
Server1	Meter	2.80	3.72	4.44	4.63	4.81	4.07	5.00
	Monitor	2.8	3.8	4.4	4.6	4.8	4.0	5.0
Server2	Meter	2.49	3.44	4.13	4.37	4.49	3.76	4.64
	monitor	2.4	3.4	4.0	4.2	4.4	3.7	4.6

vs

Consistent current reading from monitoring and measurements

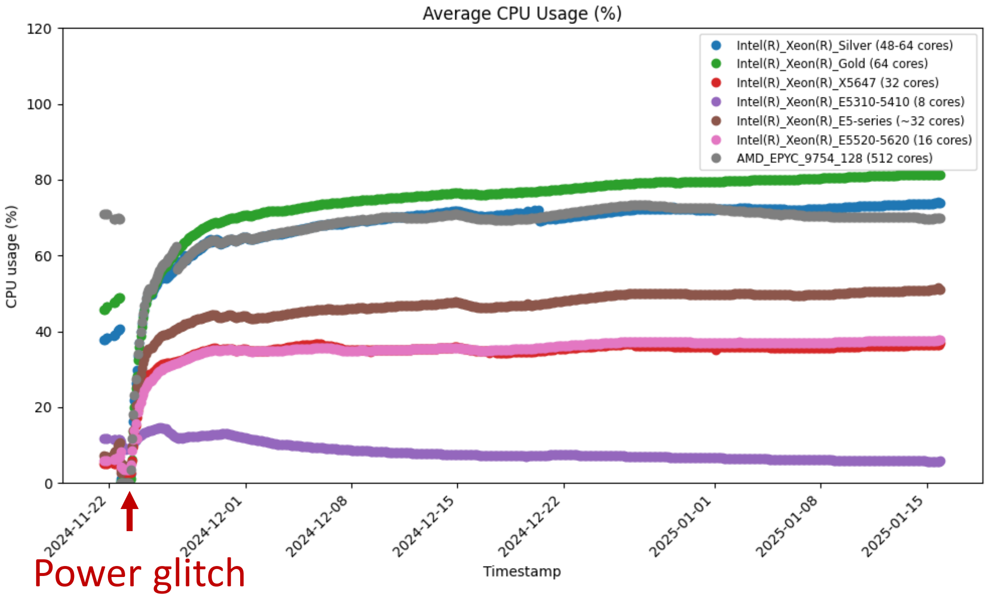
MIT Tier 2 Power Usage

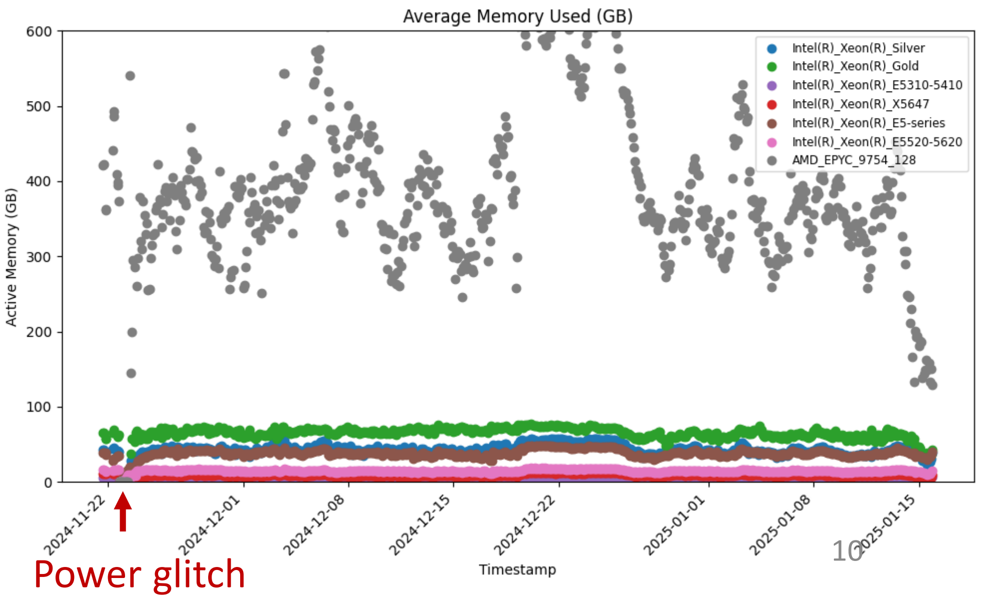
- Power consumption is relatively stable for Tier 2 operation.
  - Cost analysis based on data from the “plateau” region Nov.30th -Jan 15th



MIT Tier 2 CPU usage

- Memory usage extracted from active memory via ‘vmstats –s’
- CPU usage and memory usage is highly correlated
  - average 1.2 GB/core
- No correlation found between power usage and disk activity
- Power consumption and CPU usage is highly correlated.
- Estimate the computing resource and its connection to the power consumption.
  - **Evaluated by Power/HS06**

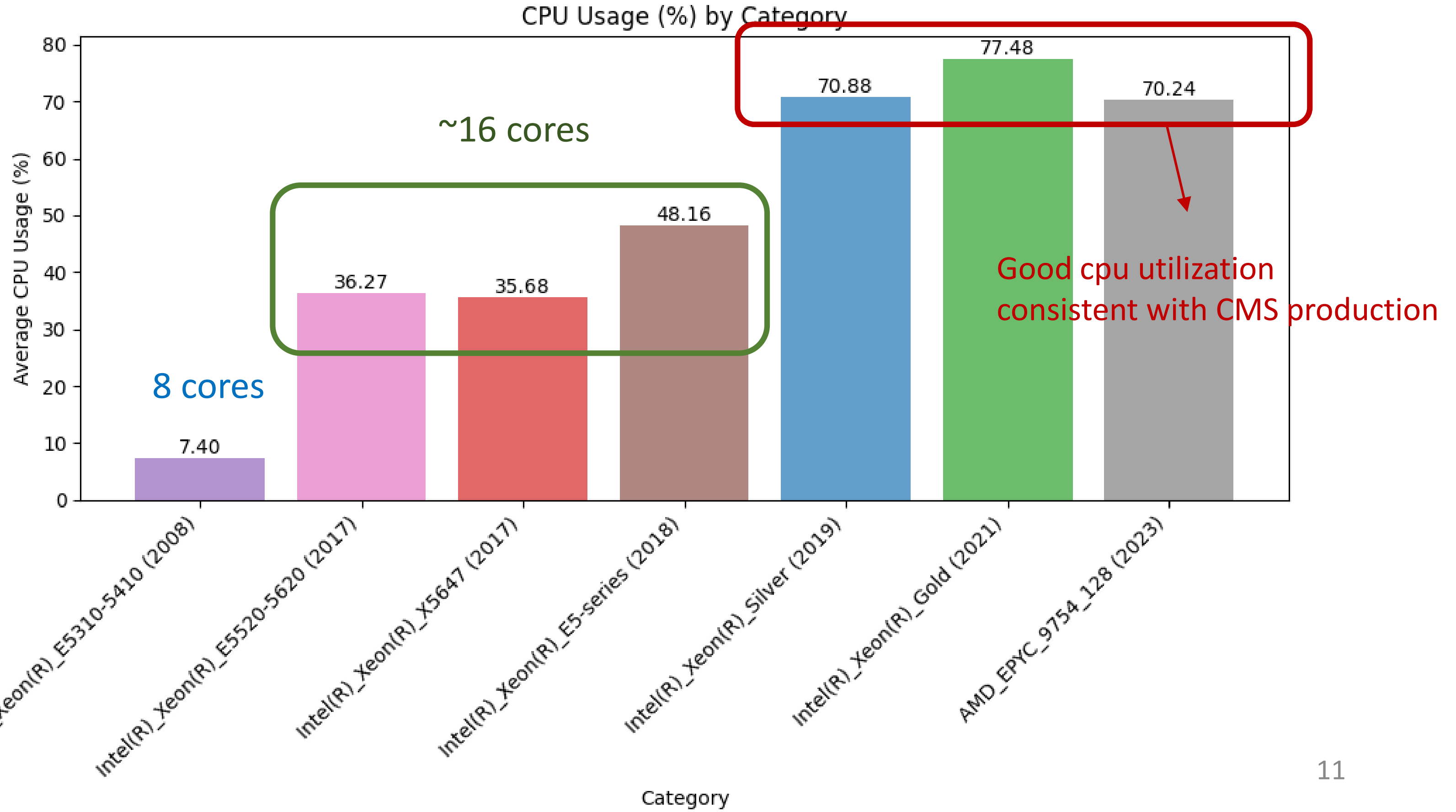




HTC25

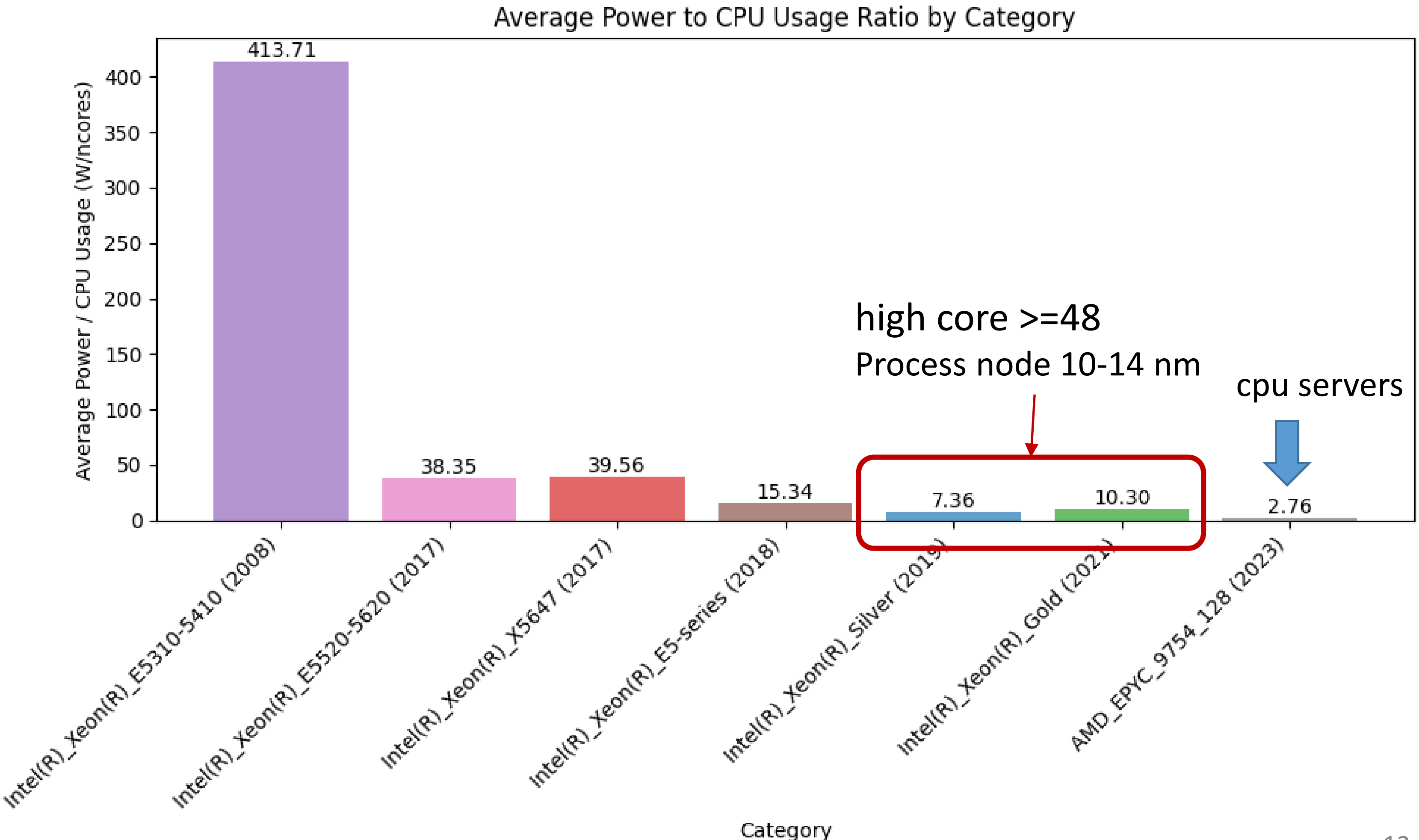
CPU Usage Comparison

- Old machines tend to be less used, due to job mismatches
  - Intel(R)_Xeon(R)_E5310-5410 have only 8 cores per machine, no longer suitable for modern computation needs
  - Production CMS pilots using 8 cores, CPU usage is very low for low core machines (fragmentation issue).



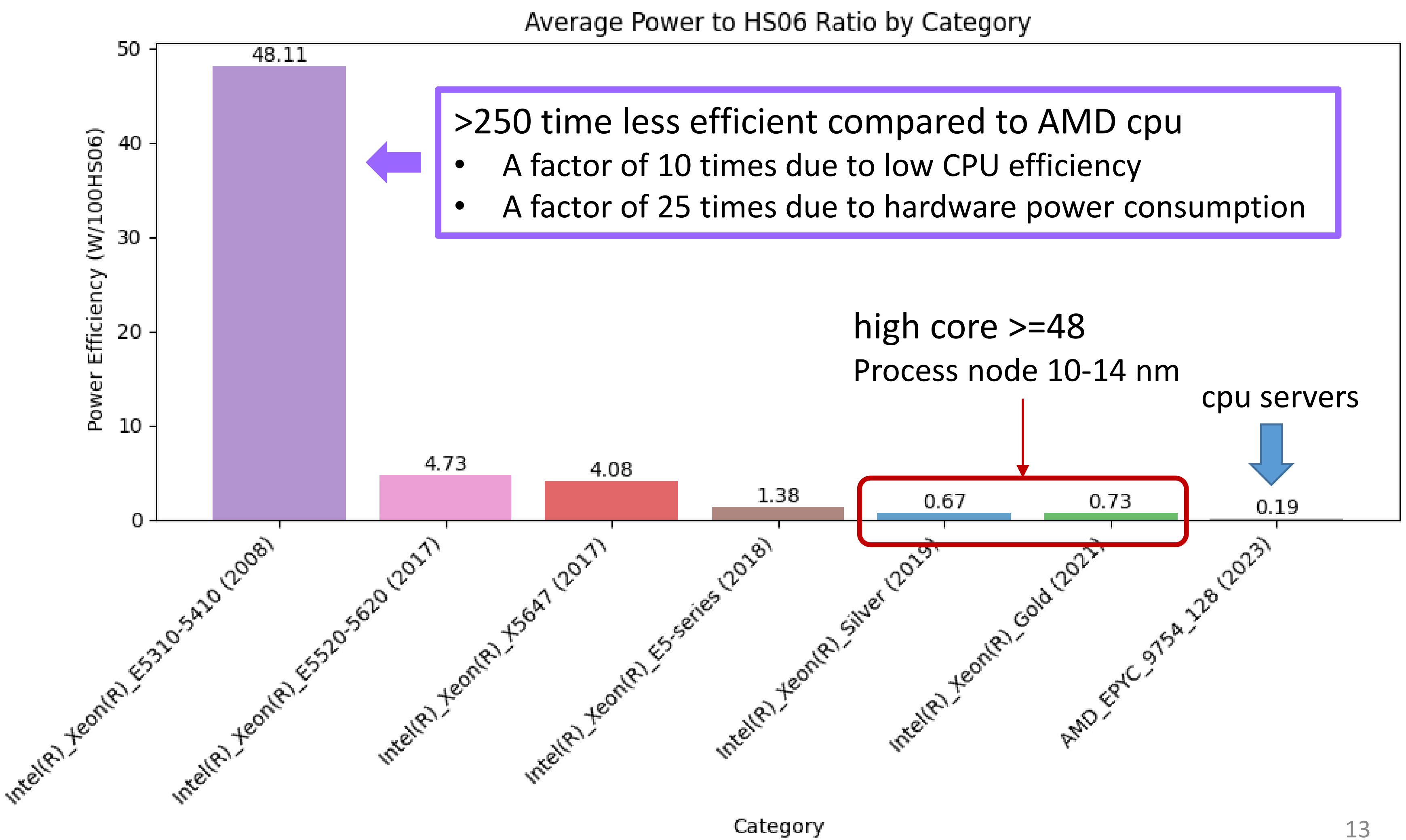
Power/Core Comparison

- The average power consumption for delivering 1 computing core



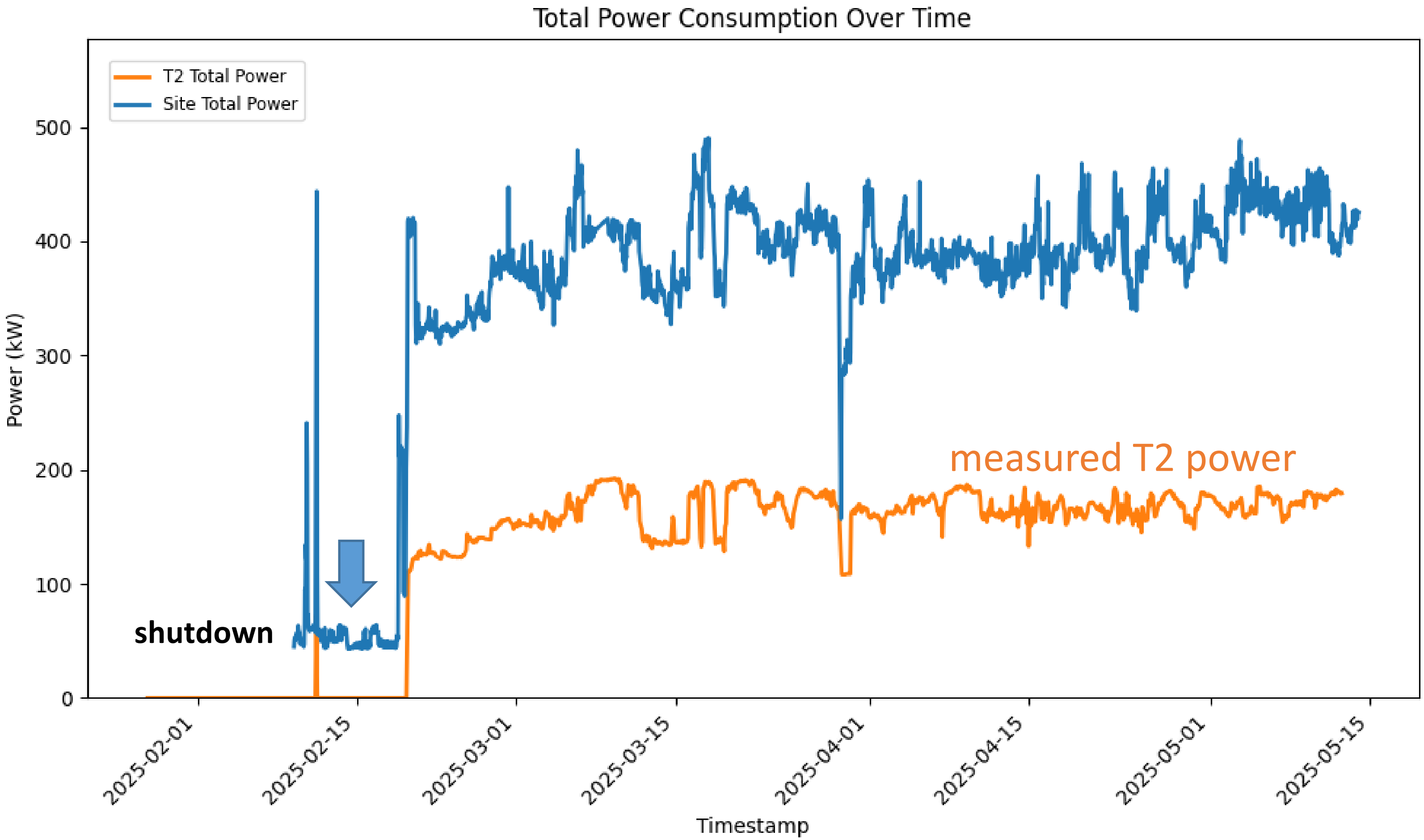
Power/HS06 Comparison

- The average power consumption for delivering 100 HS06 of compute



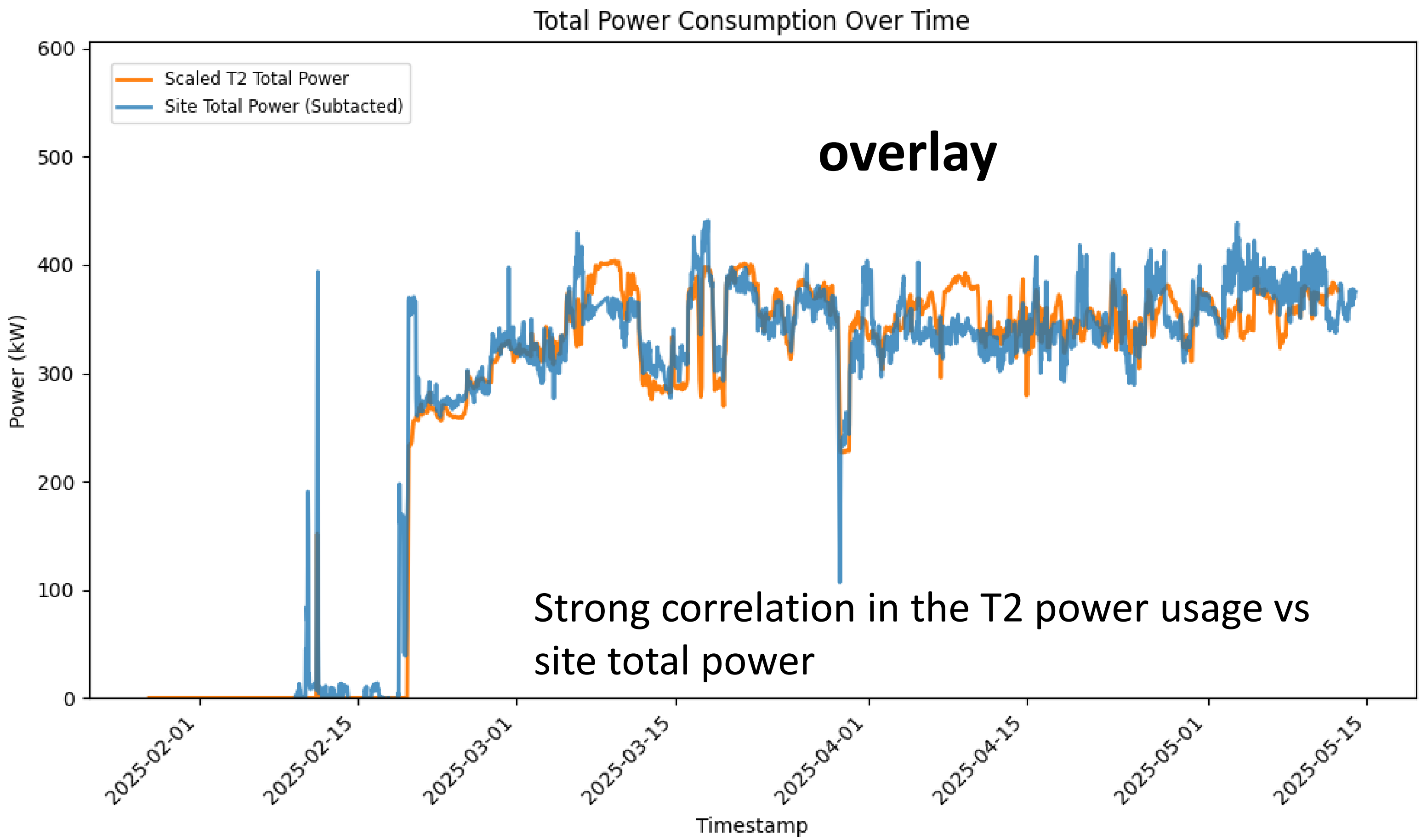
T2 Total Power Usage

- MIT CMS T2 is the major computing center running at the Site
- There are other computing servers not accounted for



T2 Total Power Usage

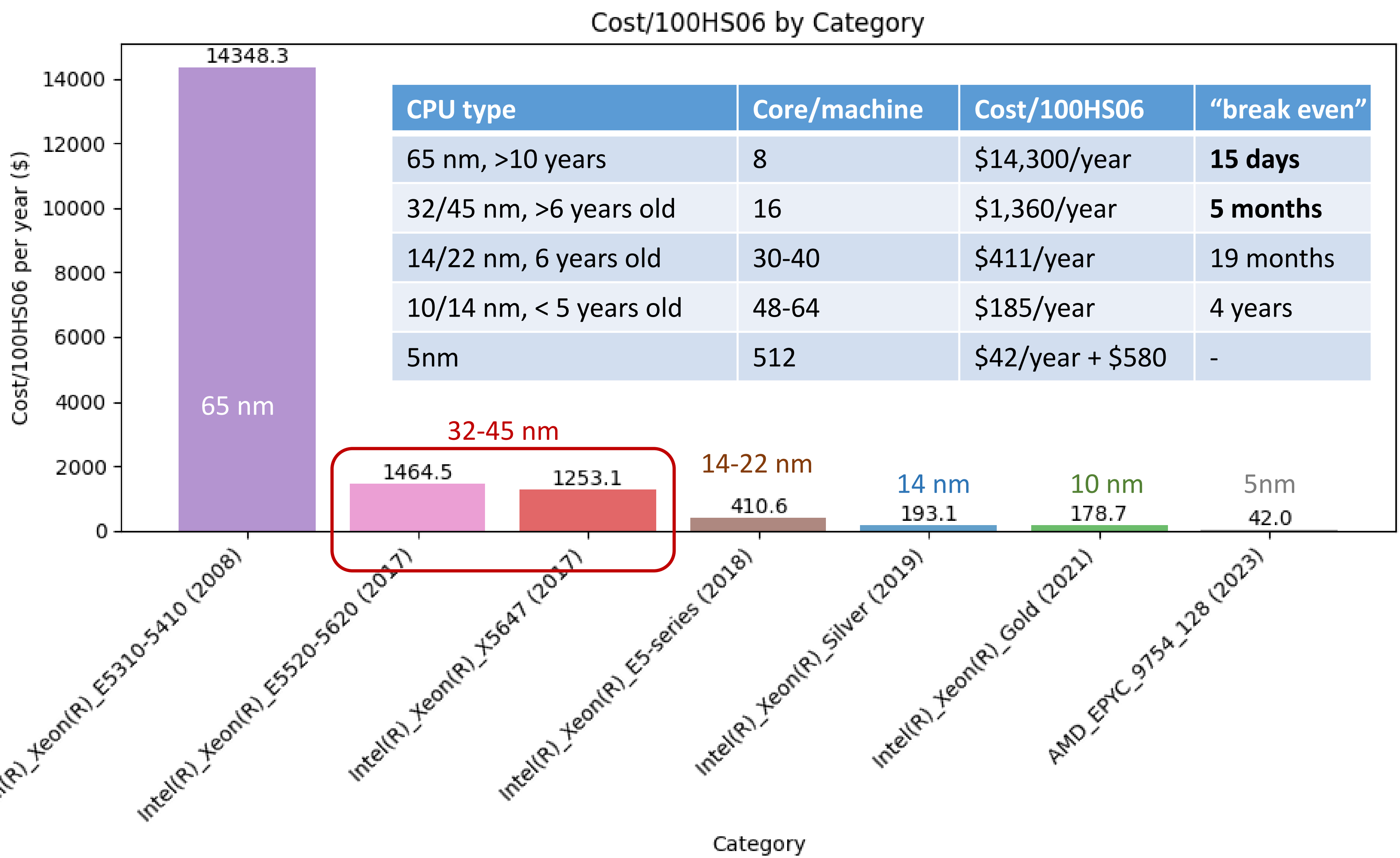
- MIT CMS T2 is the major computing center running at the Site
  - Scale T2 power to check overlay



Example Cost Analysis

- **Translate power efficiency to cost efficiency**
- For each type of machines, the cost includes power, space, and cooling.
  - Power price at T2: 14 – 18 cents / kWh
  - Power usage effectiveness (PUE) is 1.4 as a typical example
  - Space usage: >$5000 / (40 unit rack) every year
  - **Yearly cost = PUE * power * $0.16/kWh * year + $5000 * (rack space)**
- **Cost of providing 100 HS06 computation**
  - **Replace with new CPU server** AMD_EPYC_9754 (**5nm**)
    - **Cost = $580 (purchase) + $42/year**
    - 100 HS06 is provided by 1.9% of a single server
    - Among $42 per year, $2.5 comes from rack usage, $40 comes from power bill.
  - Intel(R)_Xeon(R)_E5310-5410 (**65nm, >10 years old**)
    - **Cost = $14,300/year**
    - If replaced, after 2 weeks it will break even
  - Similar estimations for other CPU models

Cost Summary



CPU type	Core/machine	Cost/100HS06	"break even"
65 nm, >10 years	8	$14,300/year	**15 days**
32/45 nm, >6 years old	16	$1,360/year	**5 months**
14/22 nm, 6 years old	30-40	$411/year	19 months
10/14 nm, < 5 years old	48-64	$185/year	4 years
5nm	512	$42/year + $580	-

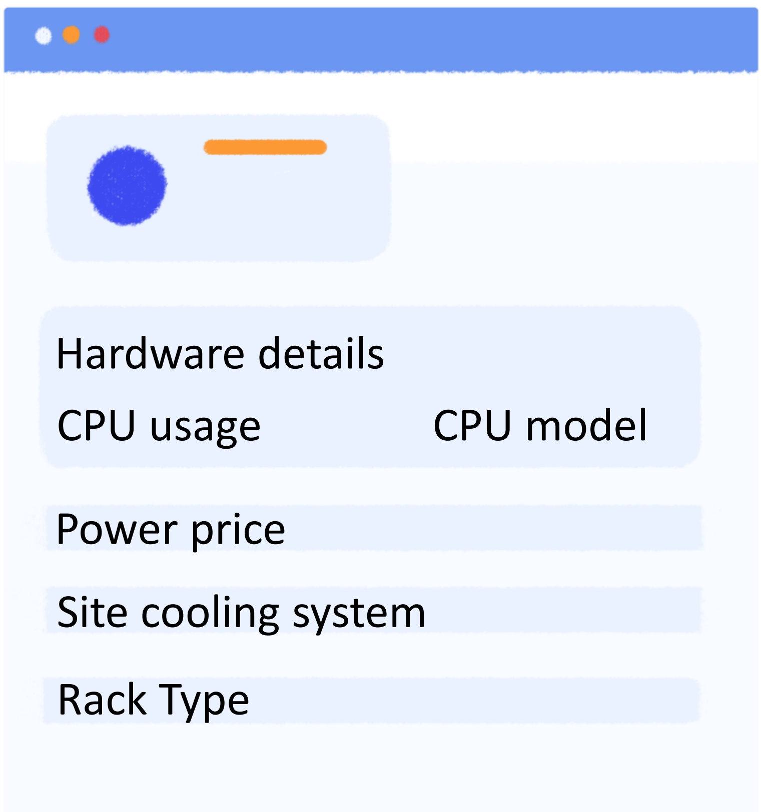
break even *: make a positive reform of investment

Universal Cost Analysis Tool

Development of the Tool

- To Develop a tool to assess computing hardware and suggest cost-effective upgrades
  - A Python package to analyze full cost of running **existing computing hardware**

Web interface to enter the parameters



Fetch existing data base

Holistic cost analysis

- Analyze hardware type
- Evaluates performance and power consumption
- Cost breakdowns in power, cooling, and racks

- Purchase of the new hardware to provide equivalent computing resources
  - How much operation time to save money with new hardware
- **Upgrade recommendations** to maintain performance at reduced cost.

Calculation

- CPU usage pattern matters
  - Pattern 1: constantly active with a CPU usage at certain level (like T2)
  - Pattern 2: Active at high CPU usage, inactive at ~0 CPU usage

$$Cost(100HS) = \frac{100HS}{F_{HS} \times \varepsilon}\left[\underbrace{P \cdot PUE \cdot Price}_{\text{Power bill}} + \underbrace{\frac{\$5000}{\min\left(\frac{40}{n}, \frac{P_{rack}}{P}\right)}}_{\text{Rack price}}\right] \times T$$

New hardware purchase

$$\frac{100HS}{F_{HS} \times \varepsilon} M(AMD)$$

$M(AMD)$ is the price of AMD hardware: $30,000

- Parameters
  - F_{HS}: HS score of a CPU model
  - P: Power consumption
  - n: Units of a machine

  → Determined from data bases

  - ε: CPU usage efficiency
  - PUE: power usage effectiveness (cooling effectiveness)
  - $5000: yearly spend of a rack
  - P_{rack}: Maximum power supply of a rack
  - Price: electric price in the unit of $/(W*year)

  → Input parameters

  - T: time in unit year

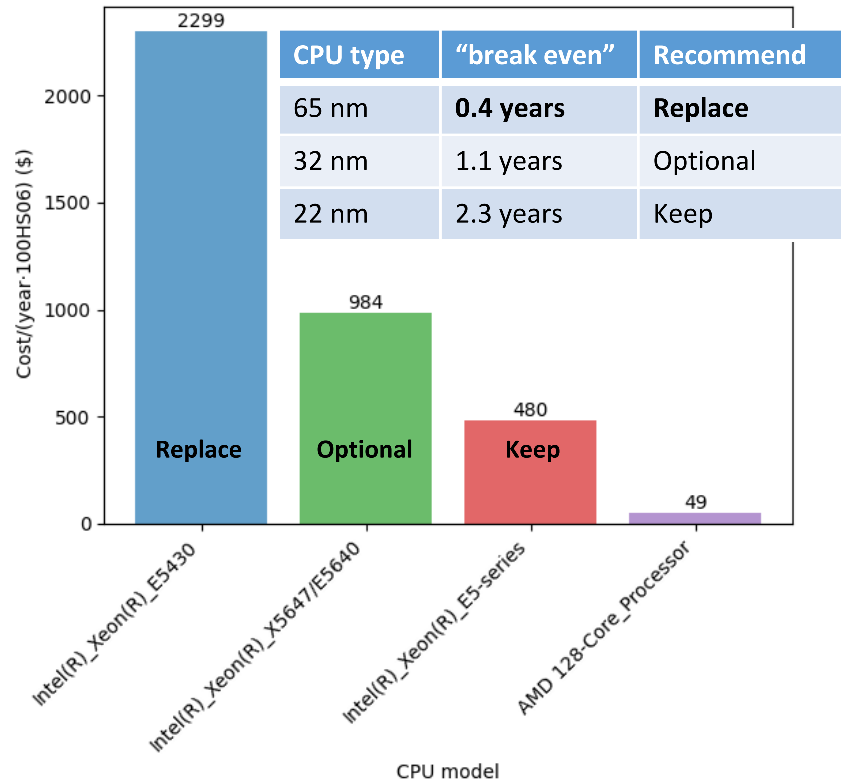
Real-World Application

- Example of usage using MIT Tier 3 - 1000 Cores site
  - Three types of hardware

CPU model	Process node	HS06	Cores	Power @100% [W]
Intel(R)_Xeon(R)_E5430	65 nm	69	8	260
Intel(R)_Xeon(R)_X5647/E5640	32 nm	155	16	300
Intel(R)_Xeon(R)_E5-series	22 nm	355	32	360
Replace with new hardware ($30,000 each machine)				
AMD 128-Core_Processor	5 nm	7450	512	1100

Input variable:
- CPU average usage: 40%
- Site power usage efficiency: 1.4
- Power price 14 cents/kWh
- Rack Type 12500W maximum

Cost for each type & replacement recommendation



CPU type	"break even"	Recommend
65 nm	**0.4 years**	**Replace**
32 nm	1.1 years	Optional
22 nm	2.3 years	Keep

Summary

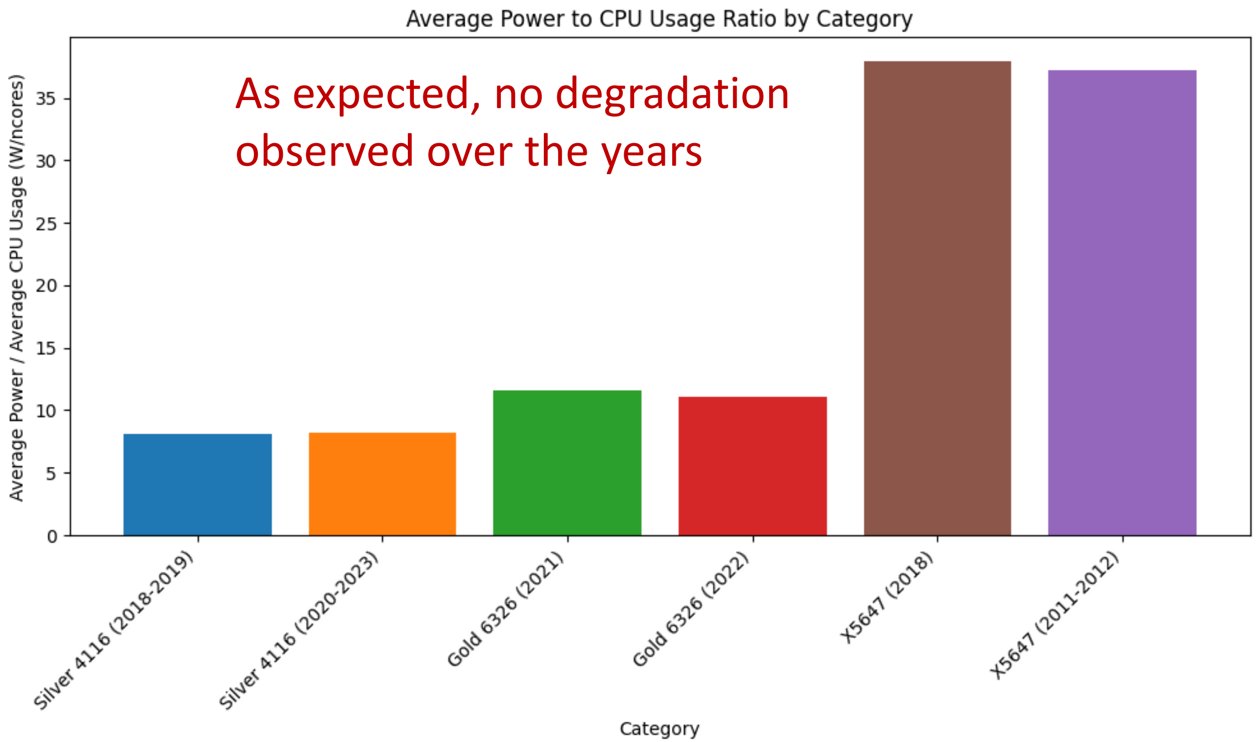
- Holistic Cost Analysis is presented for running a computing center
  - Cost from the power consumption can be enormous, strongly depends on the hardware and age
    - Smaller computing setups can draw substantial amount of power
- **Old hardware** should be replaced to reduce long-term operational costs
- A rough guideline
  - Hardware older than 10 years should be replaced immediately
  - Hardware older than 7 years should be replaced
    - A node with >30 nm process
    - Cost savings typically realized within 1–2 years.
- Provide a general tool/way to conduct the cost analysis

The End

Back up

Machine Age

- Machines with the same CPU model may be produced in different years and have been running longer.
  - Aging effect on power consumption is checked
- Check the power usage on 3 CPU models
  - Intel(R)_Xeon(R)_Silver_4116_2.10GHz: 104 machines (2018-2019), 2 machines (2021-2023)
  - Intel(R)_Xeon(R)_Gold_6326_2.90GHz: 36 machines (2021), 22 machines (2022)
  - Intel(R)_Xeon(R)_X5647_2.93GHz: 2 machines (2011-2012), 63 machines (2017)



HS06 Comparison

- The average HS06 per machine