CS528
Energy/Power Aware Scheduling of Tasks

A Sahu

Dept of CSE, IIT Guwahati

Outline

- Power Aware
- Task with Hard Deadlines
- Energy Efficiency
- Energy Efficient Scheduling
- Real Time Tasks

Announcement

Mid-Sem Paper Showing

Friday 5PM to 6PM, CSE Seminar Room

Class AVG: **26.09 out of 50**

Power Aware Scheduling Vs Energy Aware Scheduling

- Power Budget should not exceed
  - Minimized
  - Monthly Expenses: CAP ===> Solution is EMI
  - Power CAP: If your system have 100W design, at any instance of time you should not run things above 100W
    - Suppose you have 3KW wiring in your home, you have 3 AC with each of 1.5KW rating, At a given time, you can run maximum of 2 AC.
- Total energy budget should not exceed
  - Battery capacity, mah (mobile), AH (UPS)
  - Minimized: EC
  - Power and Time

Top 500 HPC System

https://www.top500.org/lists/top500/2021/11/

1. Fugaku remains the No. 1 system. Japan
   - 7,630,848 cores : **442 Pflop/s** on HPL Benchmarks.
   - This puts it 3x ahead of the No. 2 system in the list.
2. Summit, an IBM-built system at the Oak Ridge National Laboratory (ORNL), USA,
   - **148.8 Pflop/s** on the HPL benchmark
   - 4,356 nodes, each housing
     - two Power9 CPUs with 22 cores each
     - and six NVIDIA Tesla V100 GPUs, each with 80 streaming multiprocessors (S.M.).
   - Nodes are linked together with a Mellanox dual-rail EDR InfiniBand network.
3. Sierra, at Lawrence Livermore National Lab, USA,
   - Architecture is very similar to Summit.
   - 4,320 nodes with two Power9 CPUs and four NVIDIA Tesla V100 GPUs. Sierra achieved **94.6 Pflop/s.**

Top 500 HPC System

https://www.top500.org/lists/top500/2022/11/

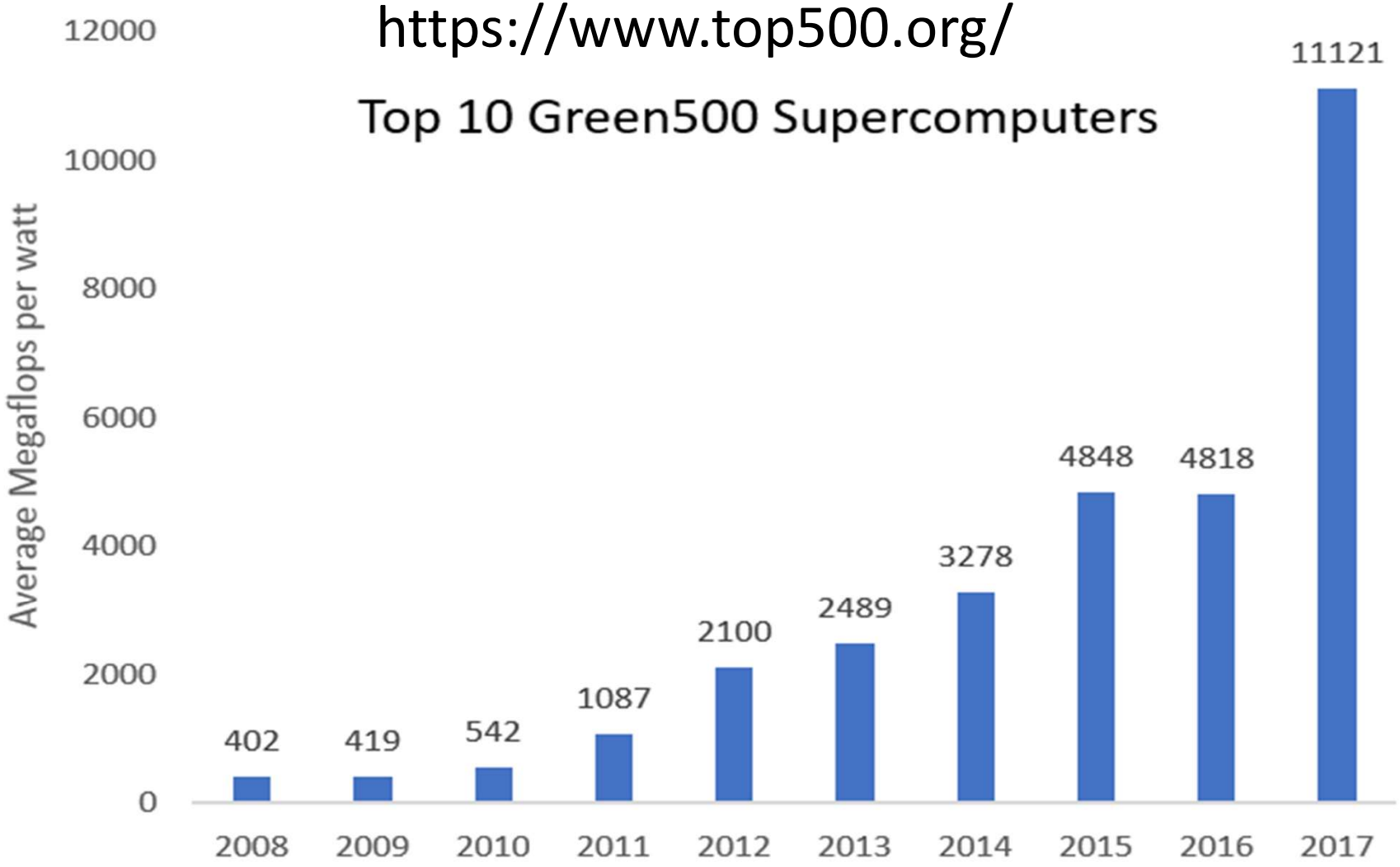
		Cores	Pmax-Pflop/s	Ppeak-Pflop/s	Power KW
1	**Frontier** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE DOE/SC/Oak Ridge National Laboratory United States	8,730,112	1,102.00	1,685.65	21,100
2	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442.01	537.21	29,899
3	**LUMI** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE EuroHPC/CSC Finland	2,220,288	309.10	428.70	6,016
4	**Leonardo** - BullSequana XH2000, Xeon Platinum 8358 32C 2.6GHz, NVIDIA A100 SXM4 64 GB, Quad-rail NVIDIA HDR100 Infiniband, Atos EuroHPC/CINECA Italy	1,463,616	174.70	255.75	5,610

Green500: evolution

- 2008: best result = 536 MFlops/Watt
- 2009: best result = 723 MFlops/Watt
  - Cell cluster, ranking 110 in top500
- 2010: best result = 1684 MFlops/Watt
  - IBM BlueGene/Q prototype 1, ranking 101 in top500,
  - Peakperf: 65 TFlops;
- 2011: best result = **2097 MFlops/Watt**
  - IBM BlueGene/Q prototype 2
  - power consumption: 41 kW / Peak 85 TFlop/s

Green500: evolution

https://www.top500.org/



https://www.top500.org/lists/green500/2022/11/

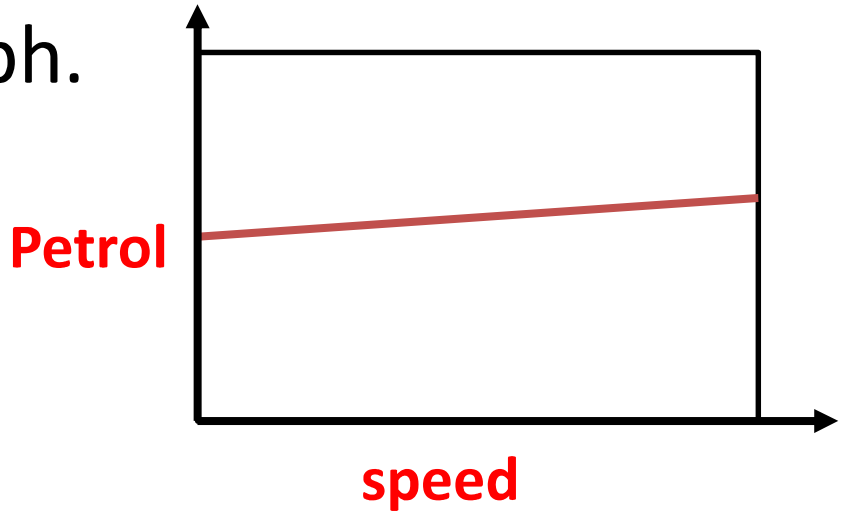
1	**Top 500 rank** :405	**Henri** - Lenovo ThinkSystem SR670 V2, Intel Xeon Platinum 8362 2800Mhz (32C), NVIDIA H100 80GB PCIe, Infiniband HDR, Lenovo Flatiron Institute United States	Cores: 5,920	PF/S 2.04	Power KW 31	GF/W 65.091
2	32	**Frontier TDS** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE DOE/SC/Oak Ridge National Laboratory United States	120,832	19.20	309	62.684
3	11	**Adastra** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE Grand Equipement National de Calcul Intensif - Centre Informatique National de l'Enseignement Suprieur (GENCI-CINES) France	319,072	46.10	921	58.021

Power Aware Scheduling Vs Energy Aware Scheduling

- Power Budget should not exceed
  - Minimized
  - Monthly Expenses: CAP ===> Solution is EMI
  - Power CAP: If your system have 100 design, at any instance of time you should not run things above 100W
    - Suppose you have 3KW wiring in your home, you have 3 AC with each of 1.5KW rating, At a given time, you can run maximum of 2 AC.
- Total energy budget should not exceed
  - Battery capcity, mah (mobile), AH (UPS)
  - Minimized: EC
  - Power and Time

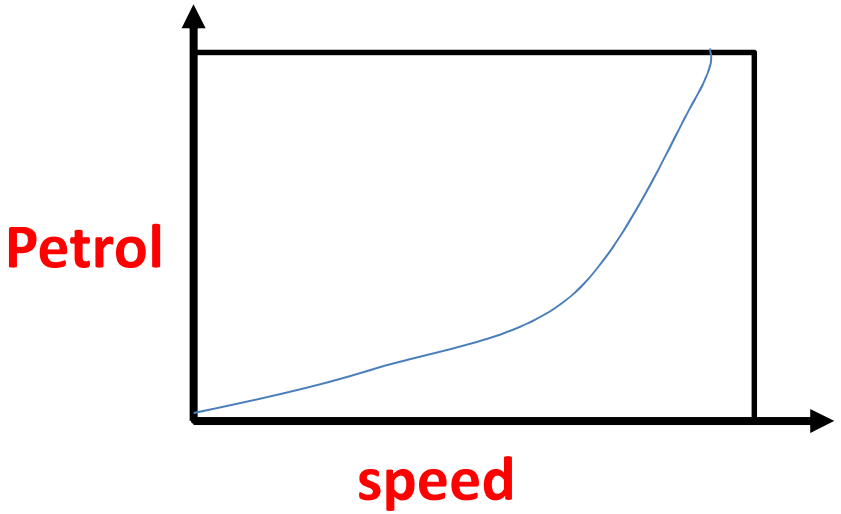
Speed Matters or Not : I

- Assume you have an Royal Enfield Bullet, you need to go from IITG to GS road, 30KM in 1 hours
- Petrol consumption is almost same at any speed. Example it 2ml/minute at 10kmph and 2.1ml/minute at 100kmph.
- How to save petrol ?
  - Sol: Go at higher controllable speed



Speed Matters or Not : I

- Assume you have an Bike, you need to go from IITG to GS road, **30KM in 1 hours**
- Petrol consumption is exponentially/quadratic increasing with speed. Example it 2ml/minute at 10kmph and 20ml/minute at 100kmph.
- How to save petrol ?
  - Sol: Go at slower speed to meet the deadline
  - Above example 30kmph
  - **Critical Speed**



Power and Energy Consumptions

- CPU: dynamic power $P_d = C_{ef} * V_{dd}^2 * f$
  - C_{ef} : switch capacitance, V_{dd} : supply voltage
  - f : processor freq → linear related to V_{dd}

$$P \ \alpha \ f^3$$

  - Battery Powered System Reduce Energy usage

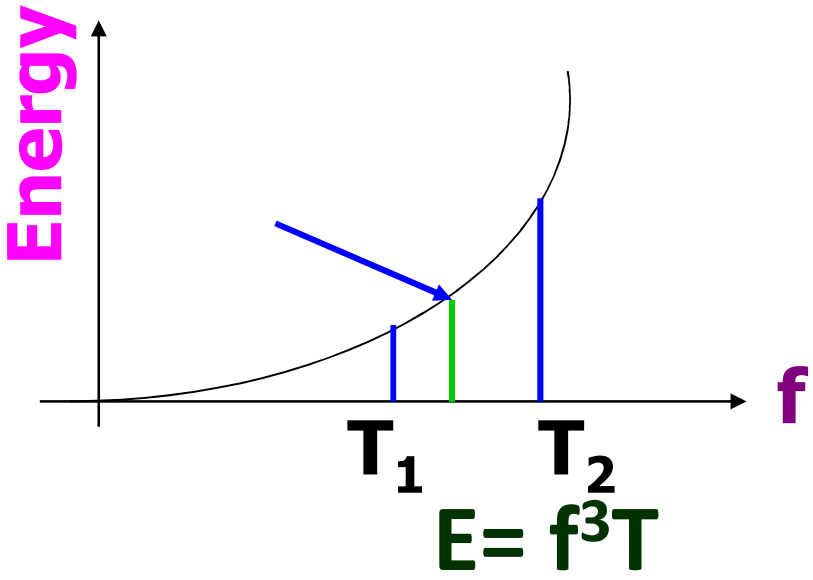
$$E = P.\ t \quad \alpha \quad f^3 t$$

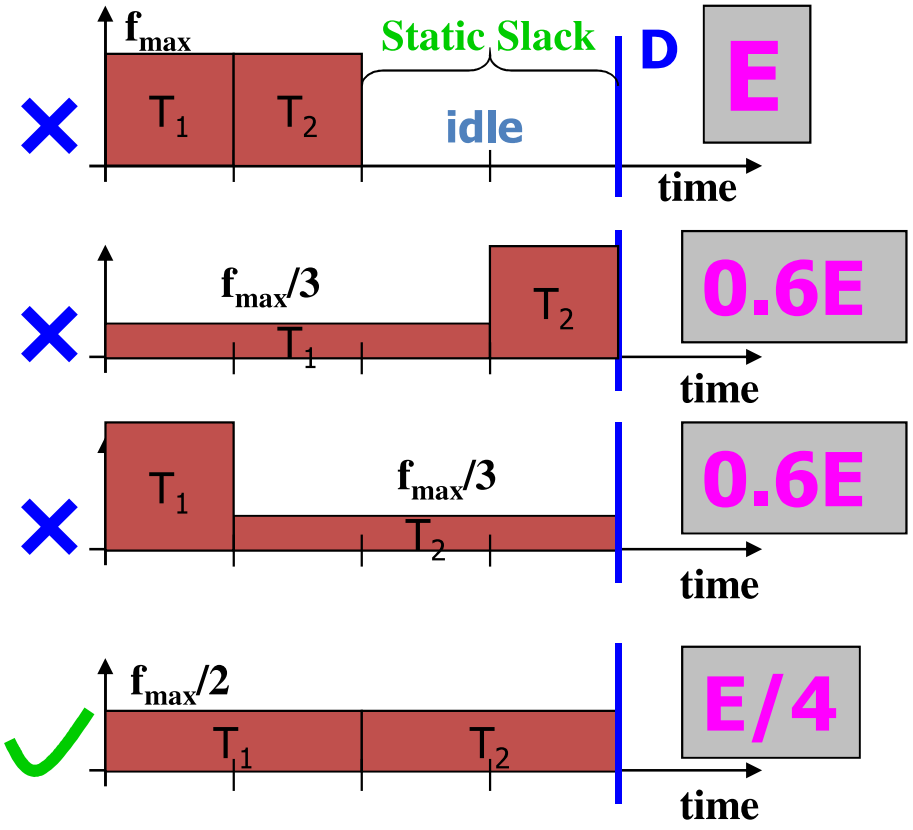
  - Execution time t is inverse to f, $t \ \alpha \ 1/f$

So $E \ \alpha \ f^2$

Power Aware Scheduling

Static slack: uniformly slow down all tasks





I. $f^3T+f^3T=2.f^3T=E$

II. $3(f/3)^3T+f^3T=0.57.E$ ✓

III. $f^3T+3(f/3)^3T=0.57.E$

IV. $2(f/2)^3T+ 2(f/2)^3T=E/4$

Energy Aware Scheduling

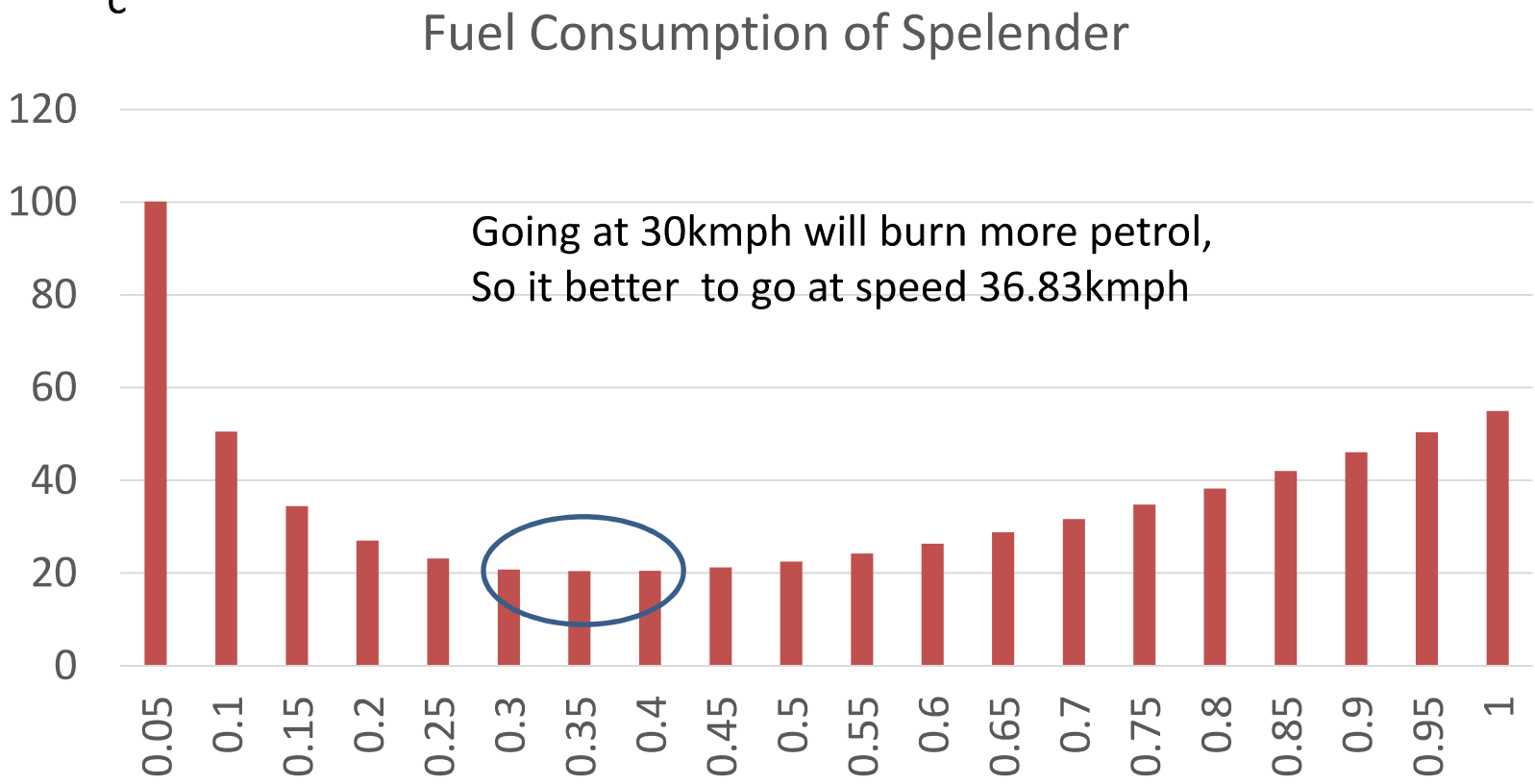
- E= P * T, More refined model $P=P_s+\alpha*f^3$
- Suppose f=[0:1].
- $E= (P_s+ \alpha *f^3)*T/f = T*[P_s/f + \alpha.f^2]$
- Min at $dE/df=0$, $-P_s*1/(f^2)+2.\ \alpha.f =0$

➔ $2.\alpha.f=P_s/(f^2)$ ➔ $f^3 =P_s/2.\ \alpha$

➔ $\mathbf{f_c = \sqrt[3]{P_s/2.\alpha}}$

Full consumption of Splendor

- Instantaneous Petrol Consumption: P= 5+50f^3
- Distance to travel 30km in 60 minutes deadline
  - F_c=0.368399



Full consumption of Bullet

- P= 200+20f^3
- Distance 30km in 60 minutes
  - Fc =1.709