CS528

Power and Energy Aware Design and Scheduling

A Sahu

Dept of CSE, IIT Guwahati

Outline

- Power Awareness
- Taxonomy of Power/Energy Consumption Model
- Power Aware Computing, Thermal Aware Computing
- Power Aware Scheduling in Cloud
- Migration and Management
  – Work Consolidation or VM Consolidation
  – Reduce number of active machine and run at critical frequency

Power Aware (PA) Computing

- Objective of PA computing/communications is
  - To improve power management and consumption
  - Using the awareness of power consumption of devices.
- Power consumption is most important considerations
  - In mobile devices due to limitation battery life.

Power Aware Computing

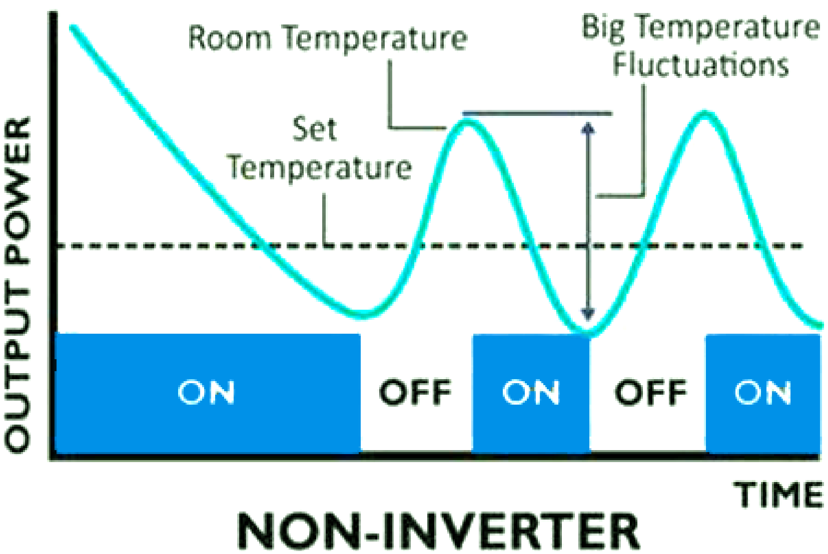
- System level power management
- Recent devices support multiple power modes.
  - CPU, disk, communication links, etc.
- **Resource Management and Scheduling Systems**
  - Can use these multiple power modes
  - To reduce the power consumption.

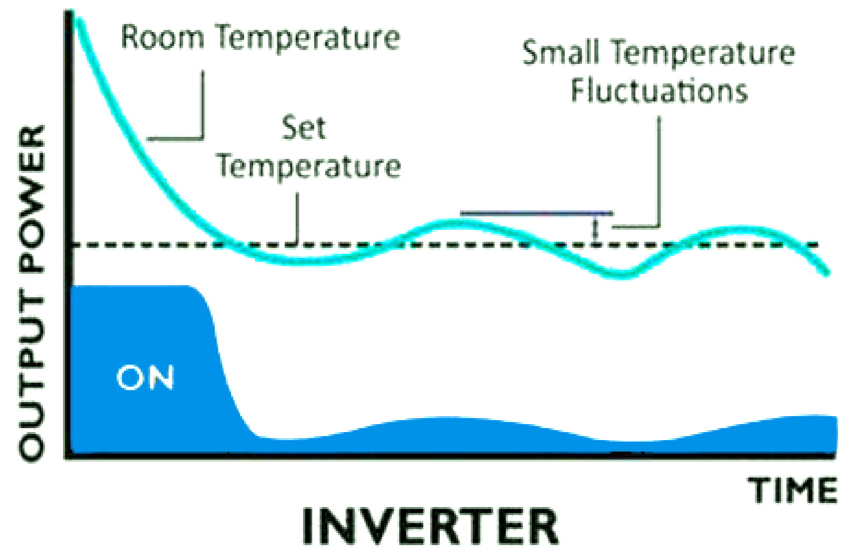
Real Life Issue: Inverter AC

- Inverter AC vs Non-Inverter AC
- Non-Inverter AC : Run fast and rest
- Non-Inverter AC: switch-of and switch-on mode
  - Sound, Fan on-off
- Inverter AC : Quit and required
  - Run at required speed : **Fun to compare with EMI**
  - Quieter than a mosquito

Real Life Issue: Inverter AC

- Eco Friendly, less power consumption
- Makes little sound, Efficient Cooling/Heating
- No Voltage Fluctuation caused by compressor
- Can be run on solar panels





DPM vs DVFS

- Inverter AC vs Non-Inverter AC
- Non-Inverter AC : Run fast and rest
- DPM : switch-of and switch-on mode
  – Sound, Fan on-off
- DVFS : Quit and required mode
  – Quieter than a mosquito
  – Run at required speed

DVFS

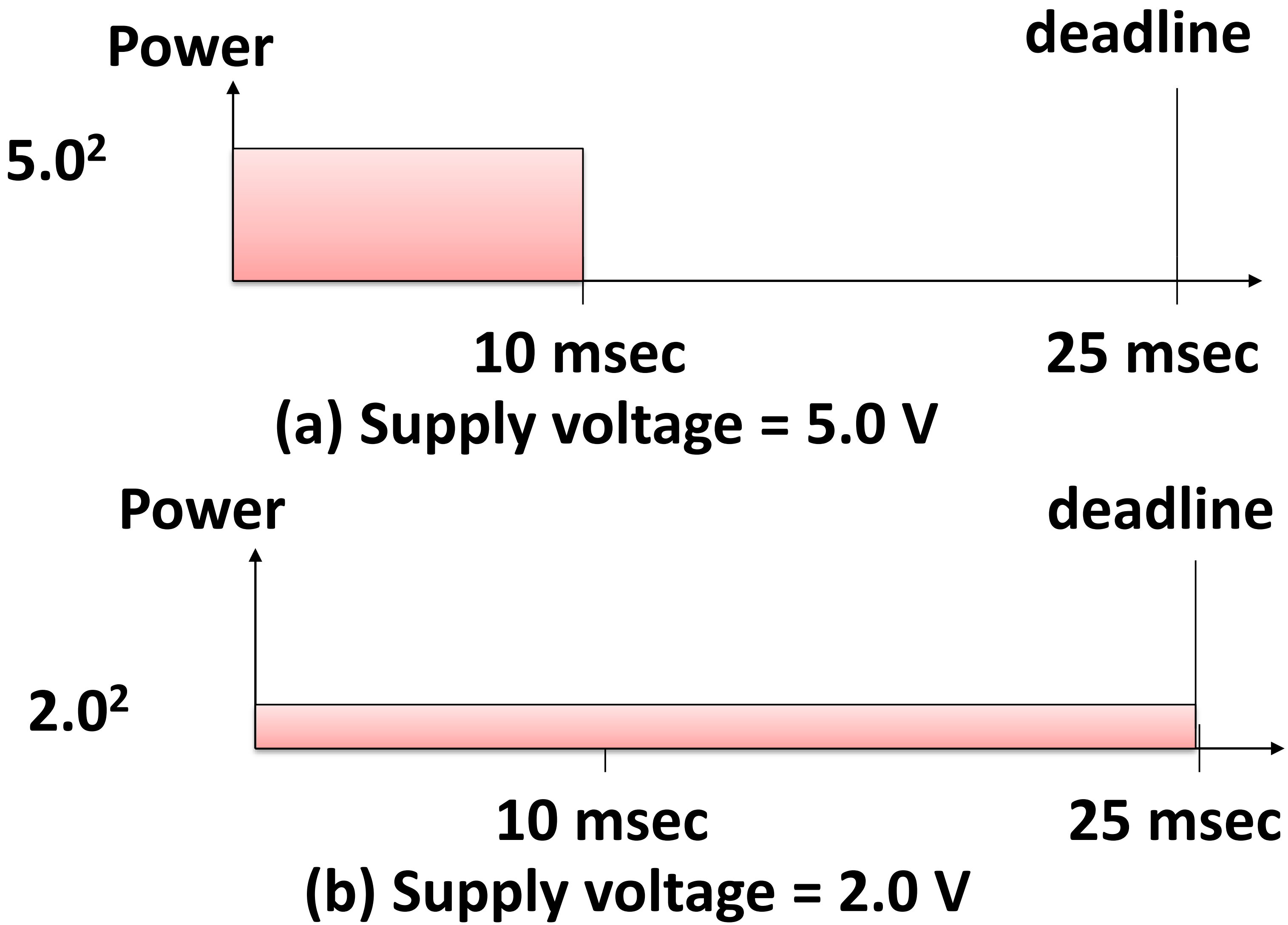
- Dynamic Voltage and Frequency Scaling
  - Intel SpeedStep
  - AMD PowerNow
- Started in laptops and mobile devices
- Now used in servers

DVS (Dynamic Voltage Scaling)

- Reducing the dynamic energy consumption
  - By lowering the supply voltage at the cost of performance degradation
- Recent processors support such ability
  - To adjust the supply voltage dynamically.
- The dynamic energy consumption
  - $\alpha * Vdd^2 * Ncycle$

    Vdd : the supply voltage, Ncycle : the number of clock cycle

DVS (Dynamic Voltage Scaling)



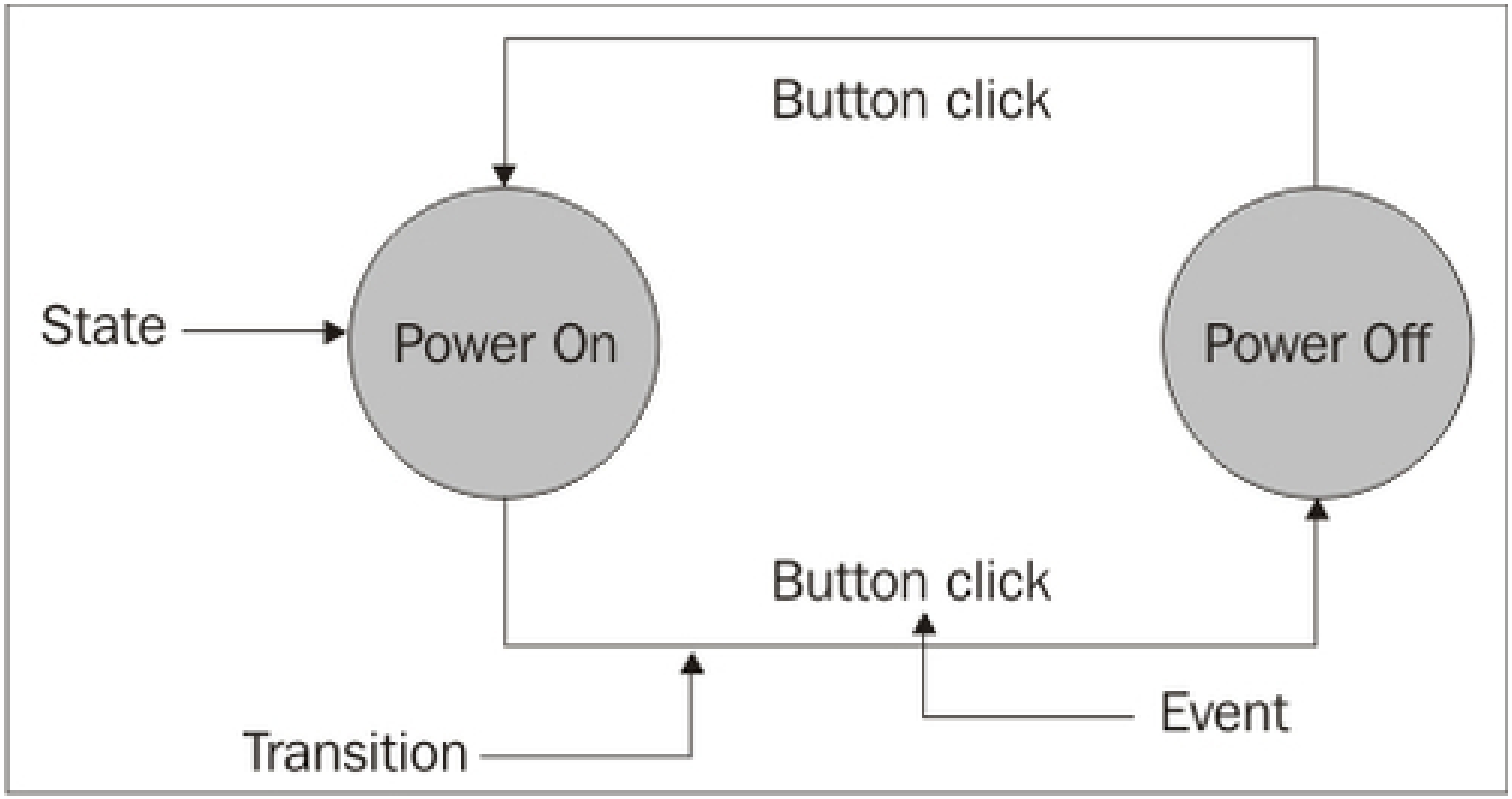
DVFS-based Power Aware Scheduling : Motivation

- Develop Resource Management and Scheduling Algorithms
  - That aim at minimizing the energy consumption
  - At the same meet the job deadline.
- Exploit industrial move towards
  - Utility Model/SLA-based Resource Allocation for Cloud Computing

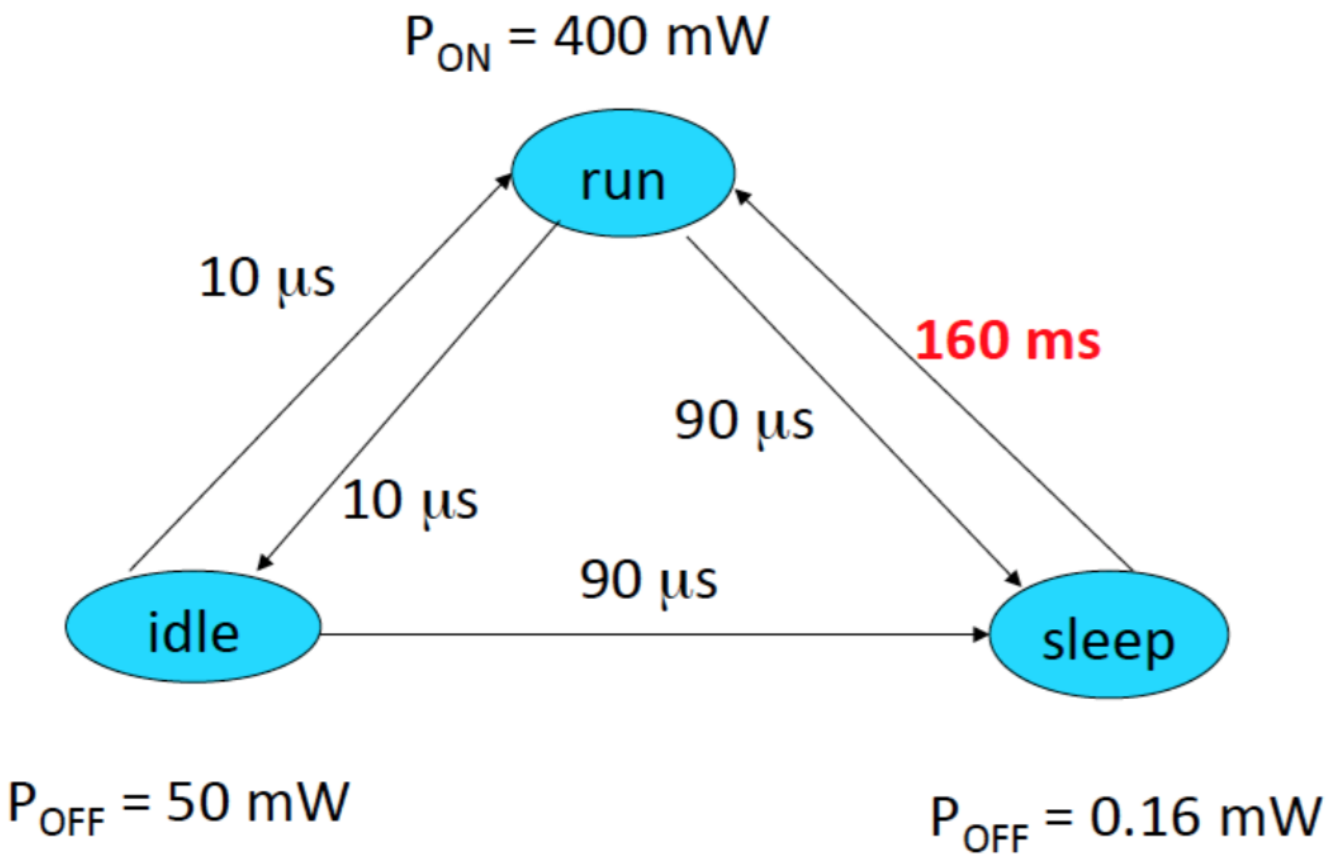
Static PM vs Dynamic PM

- Static PM
  - Invokes by user does not depends on user activities
  - Power down mode: c0, c1, …cm
    - Off, dose, nap, sleep, run
  - Mode exit upon receiving an interrupt
  - Power State machine
- Dynamic PM
  - Control power based on dynamic activity in CPU
  - Dynamically change freq, shut some parts
  - **Do when in Run State**

Static PM with Power States

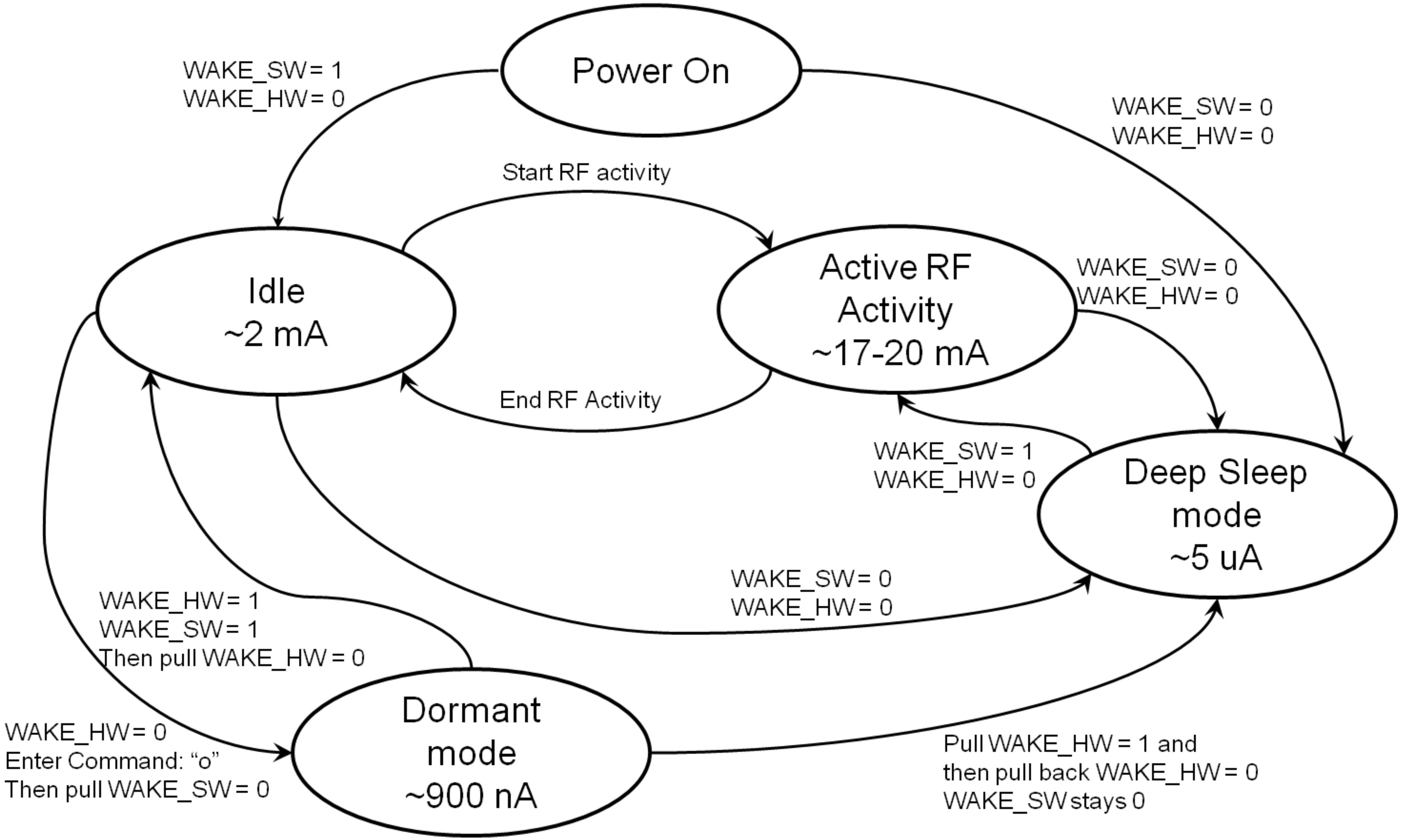


Static PM with Power States

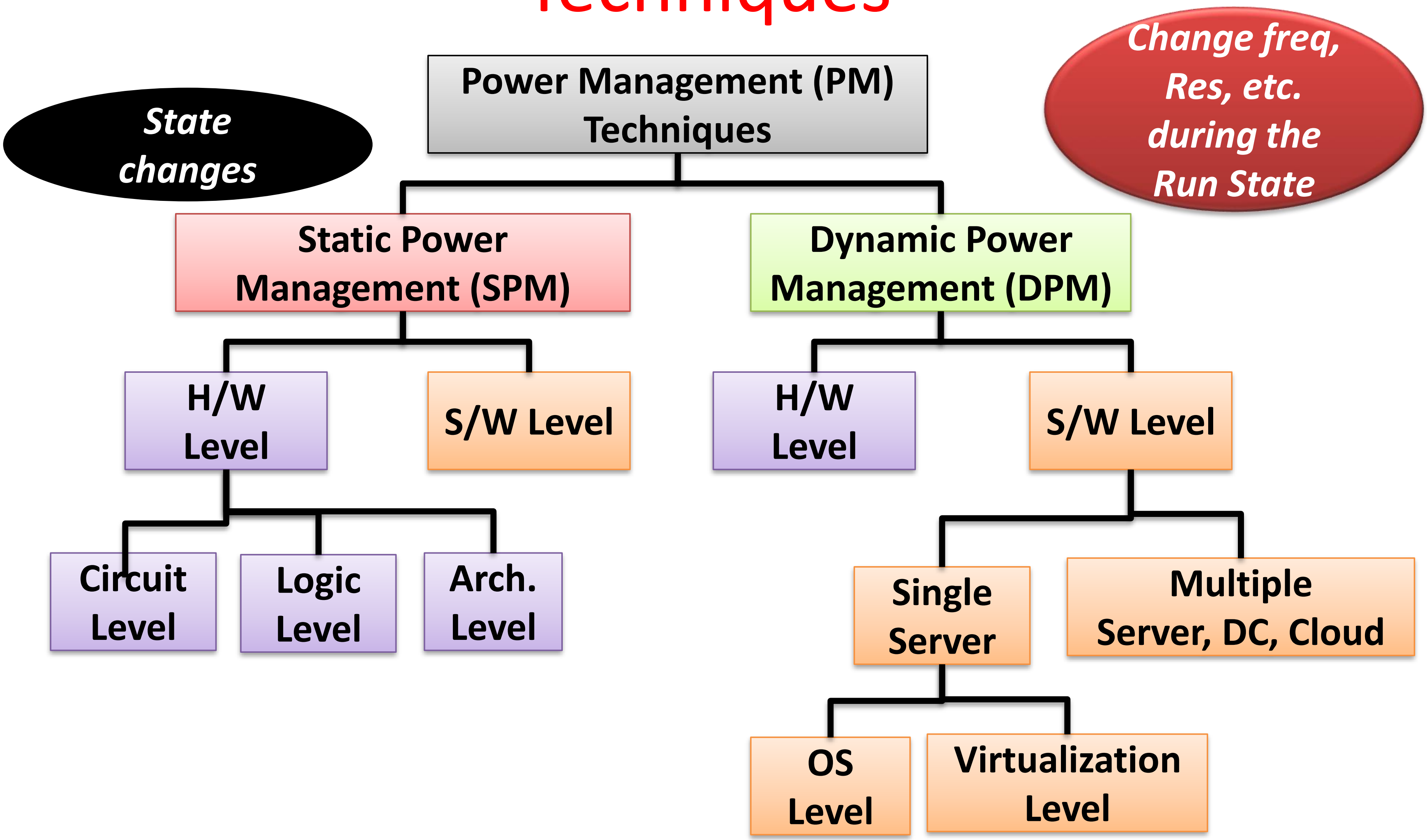


$P_{TR} = P_{ON}$

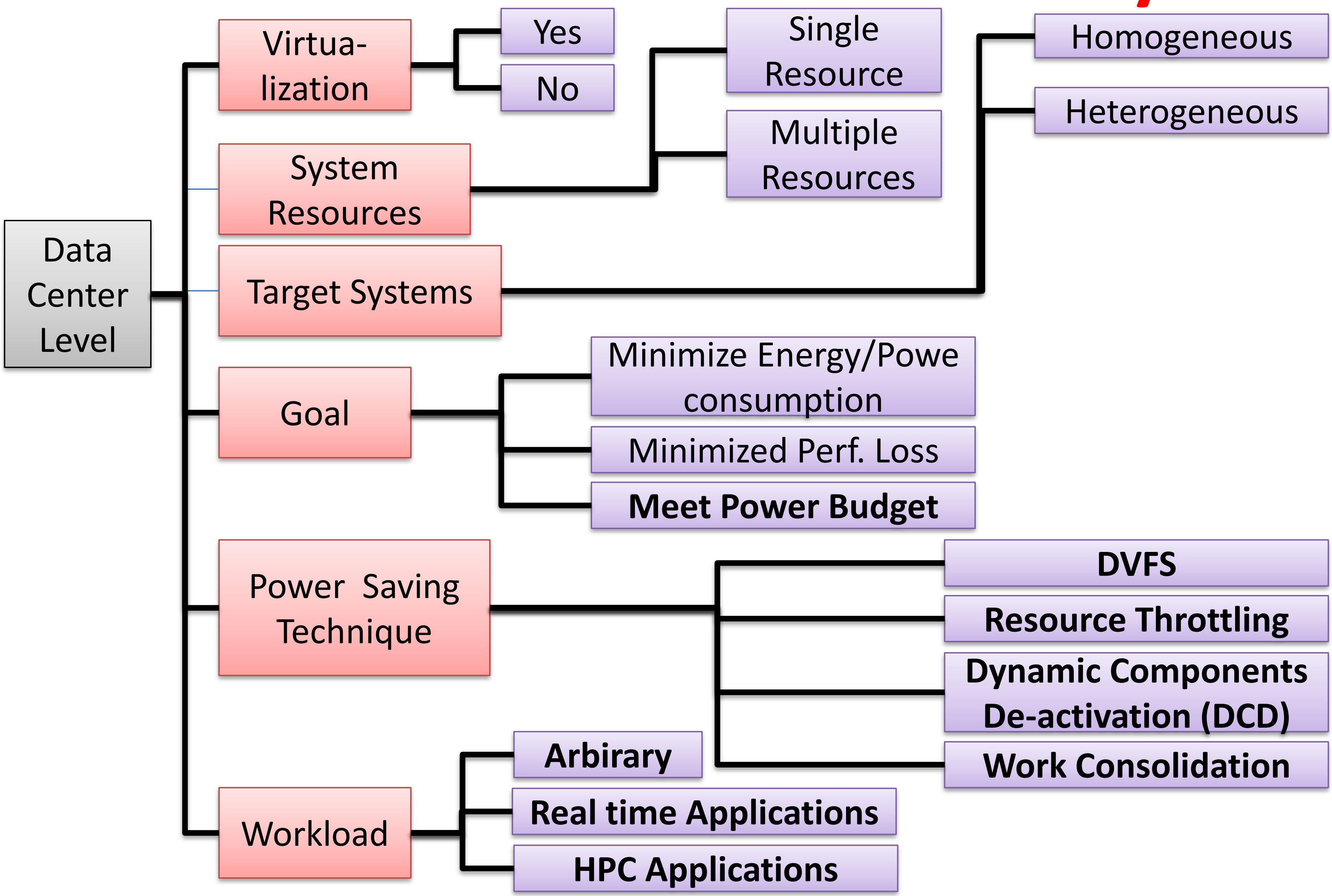
Static PM with Power States



Taxonomy of Power Management Techniques



Data Center Level : Taxonomy



Throttling Vs Over clocking

- Throttling
  - to hold somebody tightly by the throat and stop him/her breathing
  - Put a cut-off mark: Example car governor
  - Some thing going wrong: reduce activity
  - **Thermal/Power Throttling**
- Overclocking (If necessary): Turbo Boost
  - Put maximum doable afford
  - Run at maximum speed
  - Urgency to do more work

Cloud Providers EC Measures

- Cloud service: profit margin is reduced due to high energy costs
- Amazon.com’s estimate EC data centres
  - Amount to 42% of the total budget
  - Direct power consumption and cooling infrastructure
  - Amortized over a 15-year period.

Cloud Providers EC Measures

- Google, Microsoft, and Yahoo are building large DCs
  - in barren desert land surrounding the Columbia River, USA
  - to exploit cheap hydroelectric power.
- Increasing pressure from Governments worldwide
  - to reduce carbon footprints, which significant impact on climate change
  - **Carbon Tax (July 2012 in Australia) on industries**

Green Computing

- Performance/Watt is not following Moore's law.
- Advanced scheduling schemas to reduce energy consumption.
  - **Power aware, Thermal aware**
- Data center designs to reduce Power Usage Effectiveness.
  - **Cooling systems,**
  - **Rack design**

Research Opportunities

- There are a number of areas to explore in order to conserve energy within a Cloud environment.
  - Schedule VMs to conserve energy.
  - Management of both VMs and underlying infrastructure.
  - Minimize operating inefficiencies for non-essential tasks.
  - Optimize data center design.

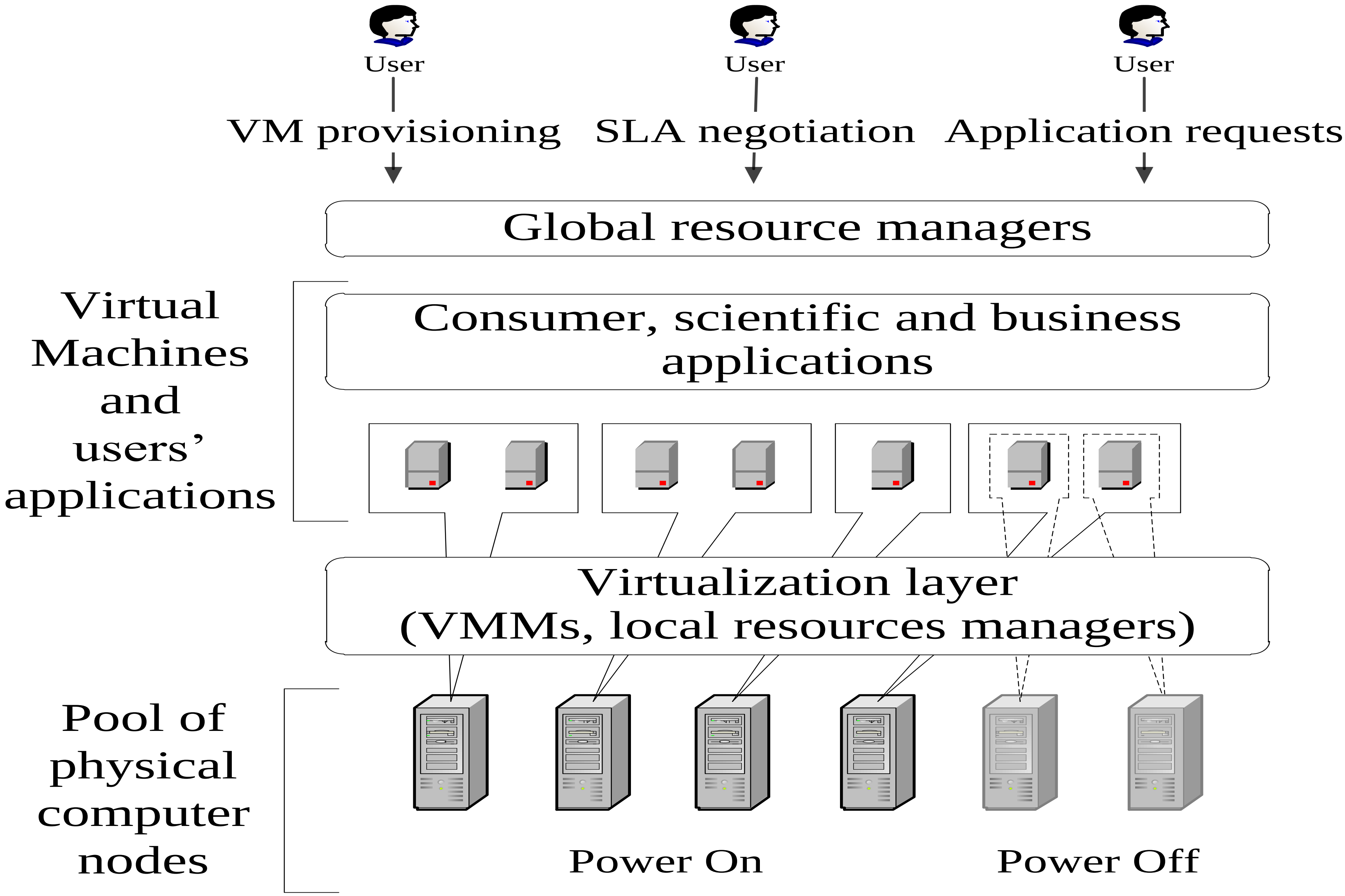
“Power-Aware” Algorithms

- Host overload detection
  - Adaptive utilization threshold based algorithms
  - Regression based algorithms
- Host underload detection algorithms
  - Migrating the VMs from the least utilized host

“Power-Aware” Algorithms

- VM selection algorithms
  - Minimum Migration Time policy (MMT) : **Select VM with small size, so that MT will be minimum**
  - Random Selection policy (RS)
  - Maximum Correlation Policy (MC) : effective VM
- VM placement algorithms
  - **Where to put the selected VMs on the Machines**
  - Heuristic for the bin-packing problem
  - Power-Aware Best Fit Decreasing algorithm (PABFD) : for the current selected VM, choose the best PM for power saving

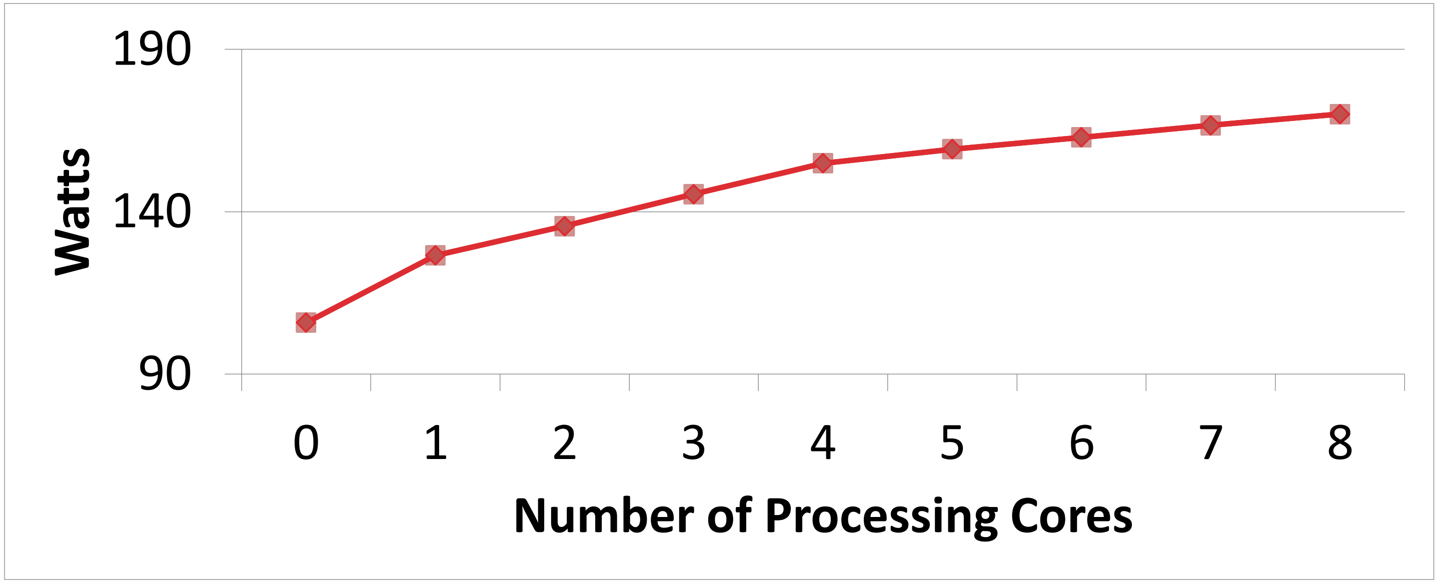
Dynamic VM Consolidation



Three Sub-Problems

- When to migrate VMs?
  - Host overload detection algorithms
  - Host underload detection algorithms
- Which VMs to migrate?
  - VM selection algorithms
- Where to migrate VMs?
  - VM placement algorithms

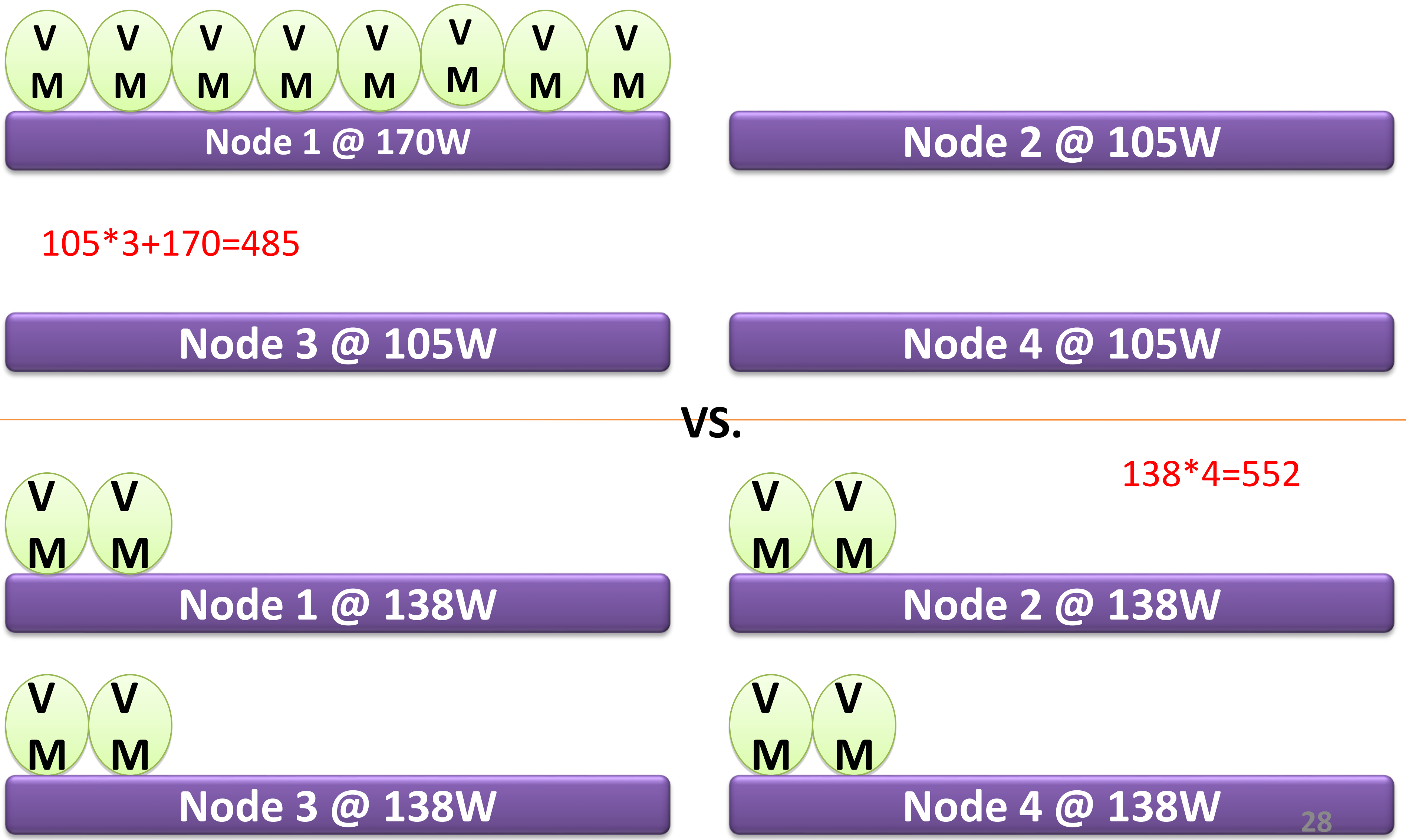
VM scheduling on Multi-core Systems



Power consumption curve on an Intel Core i7 920 Server (4 cores, 8 virtual cores with Hyperthreading)

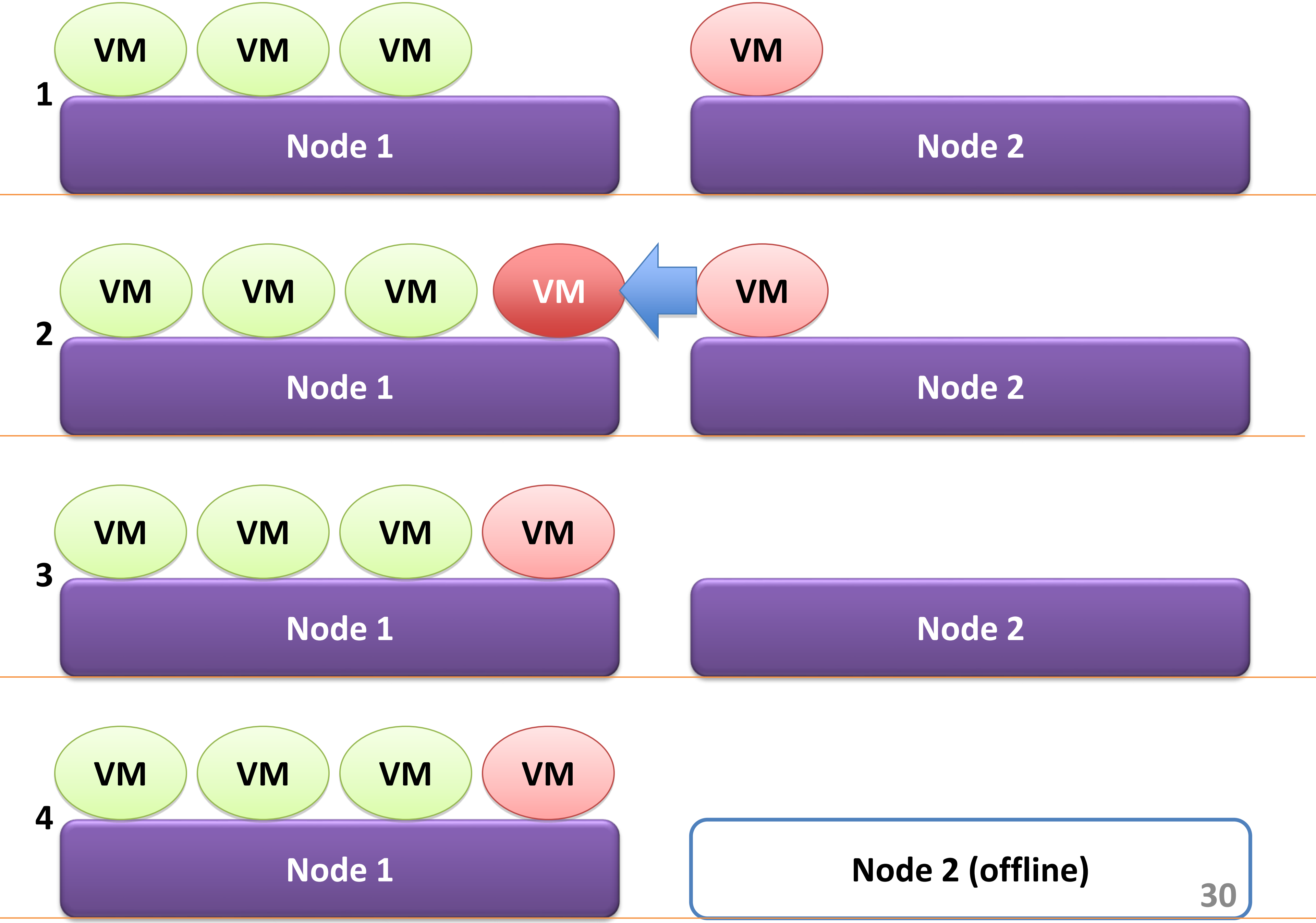
- Nonlinear relationship between the number of processes used and power consumption
- We can schedule VMs to take advantage of this relationship in order to conserve power

485 Watts vs. 552 Watts



VM Management

- Monitor Cloud usage and load.
- When load decreases:
  - Live migrate VMs to more utilized nodes.
    - Live Migration: Service of the VM is undisrupted during migration (in case of web/db/file server)
    - At the time of migration: service continues at both Source and Target
    - Grace full Migration: Suppose you want to shift your shop from IITG CORE I to CORE V, First get Space at CORE V, Start Service from CORE V, after some time Stop Service at CORE I
  - Shutdown unused nodes.
- When load increases:
  - Start up waiting nodes.
  - Schedule new VMs to new nodes.


1
VM
VM
VM
VM
Node 1
Node 2
2
VM
VM
VM
VM
VM
Node 1
Node 2
3
VM
VM
VM
VM
Node 1
Node 2
4
VM
VM
VM
VM
Node 1
Node 2 (offline)

Energy Minimization Problem

- Given N tasks with (e_i, d_i), $a_i=0$
- Given M machines with Power model of Machines

$$P(t) = P_{min} + \alpha . u(t)^3$$

Execute all the tasks without missing deadlines and goal is to minimize energy.

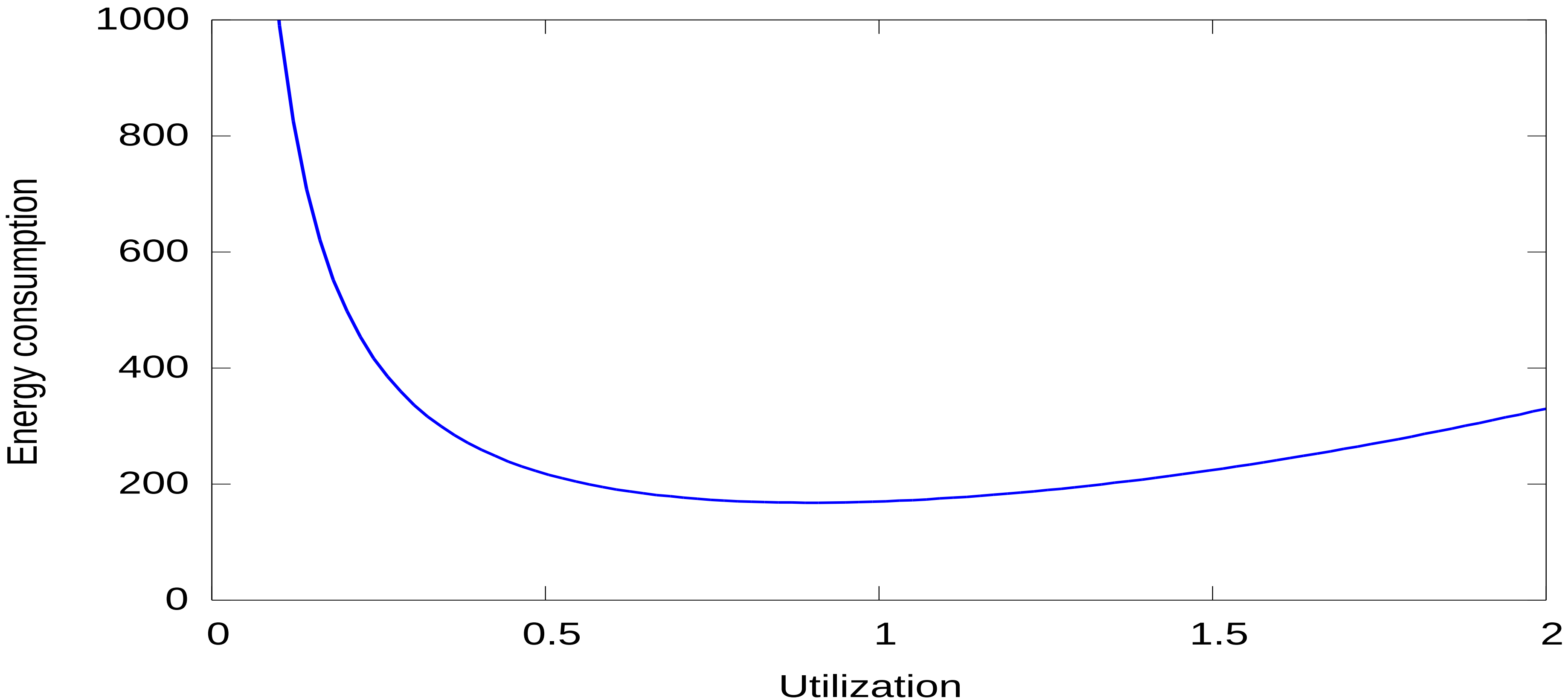
Energy Model of Server/Host

- Processor is the major contributor
- Considered both static and dynamic EC

$$E = \int_{t_1}^{t_2} P(t)dt$$

$$P(t) = P_{min} + \alpha.u(t)^3$$

Energy Model of Server/Host



- When a task runs with utilization u,

$$E = (P_{min} + \alpha . u^3).e/u = (P_{min}/u + \alpha . u^2).e$$

Critical Utilization

Minimum utilization

- Utilization required by a task to finish at deadline

$$u_{i\_min} = e_i / d_i$$

Least feasible VM

- Closest VM type which can support u_{i_min}

Critical utilization

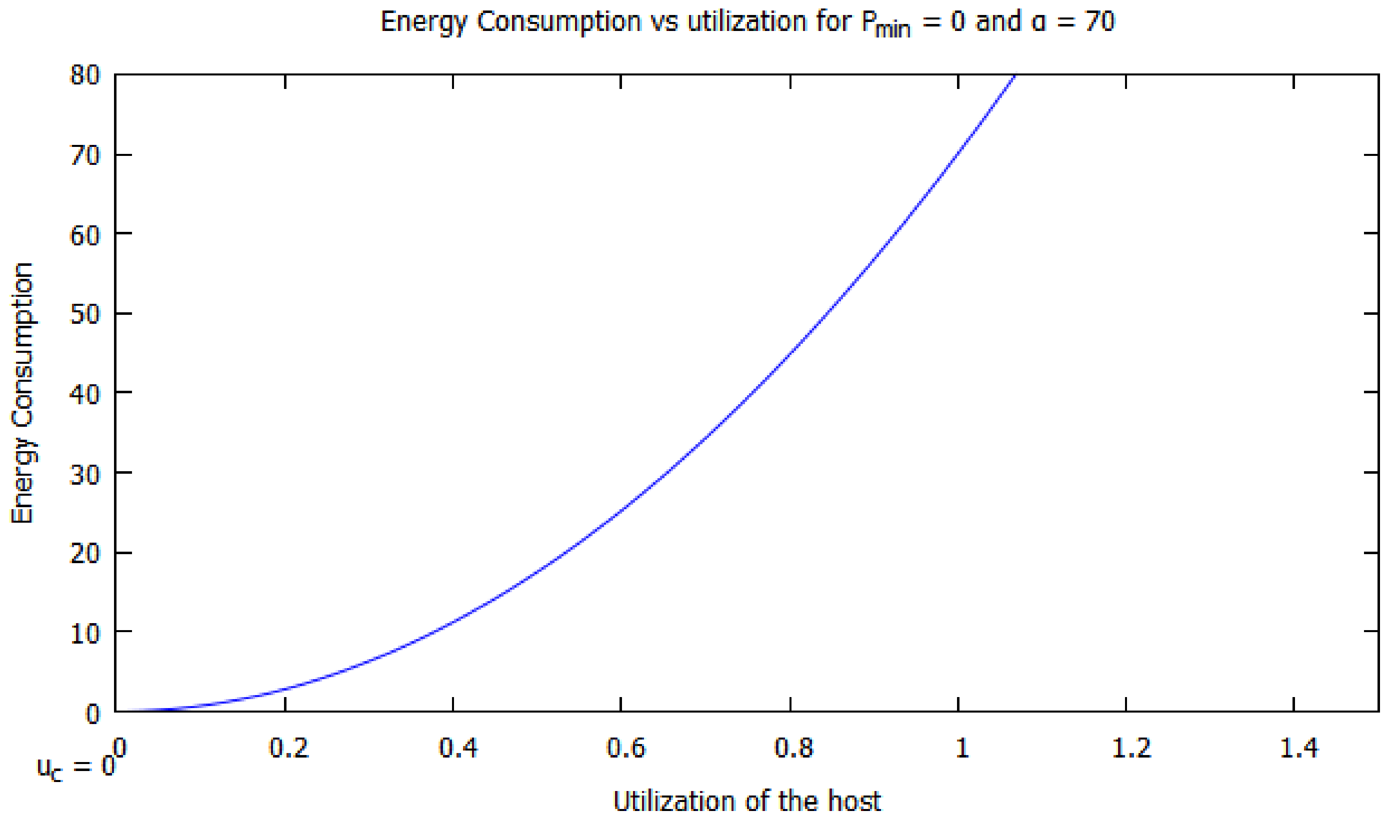
- Utilization of a host when energy consumption is minimum

$$u_c = \sqrt[3]{P_{min}/2\alpha}$$

Classify of host based on U_c

- Case 1 : Negligible static power (i.e. $P_{min} = 0$)
  => u_c=0
- Case 2 : Very high static power
  => u_c>1
- Case 3 : General case
  => 0<u_c<=1

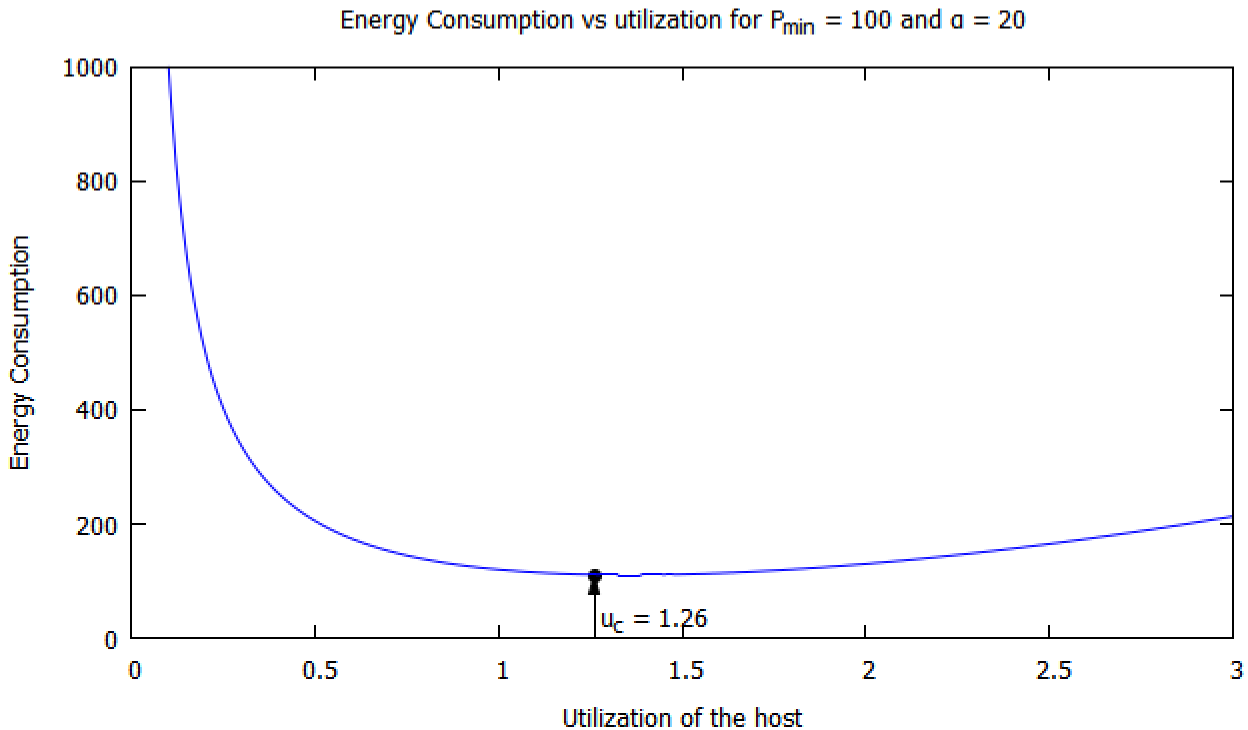
Scheduling for Case I : P_{min}=0



- For every task, choose the least feasible VM type
- Allocate the VM to a new host

$E = \alpha . u_t^2 . e$ $\quad\quad$ $u_1^2 + u_2^2 < (u_1 + u_2)^2$

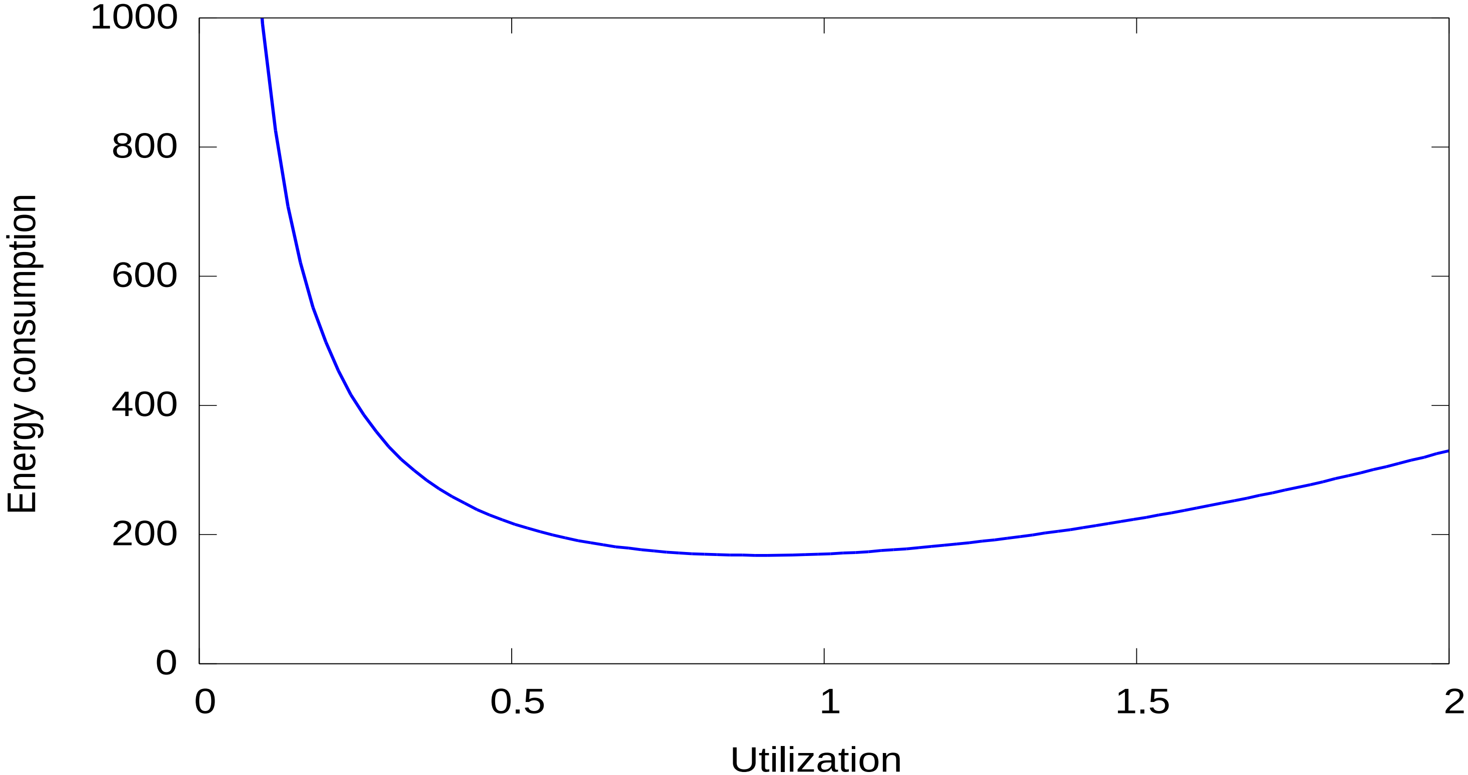
Scheduling for Case II : $U_c > 1$



- Maximum utilization of a running host is 1
- Closest feasible utilization value is 1
- Solved using general case with utilization 1

Energy Model of Server/Host

- Power consu. of a host, $P = P_{min} + \alpha u^3$



$P_{min} = 100$
$\alpha = 70$

- When a task runs with utilization u,

$$E = (P_{min} + \alpha u^3).l/u = (P_{min}/u + \alpha u^2).l$$

Energy Model of Server/Host

Least feasible VM type

- A VM with utilization $\lceil u_{i\_min} \rceil$
- $u_{i\_min} = l_i / d_i$ (uti. to finish a task just at deadline)

Critical utilization

- utilization of a host when energy consumption is minimum

$$u_c = \sqrt[3]{P_{min}/2\alpha}$$

Target of the scheduler is to keep the host utilization at u_c