Power Management on OpenVMS

Clarete Riana
OpenVMS Engineering

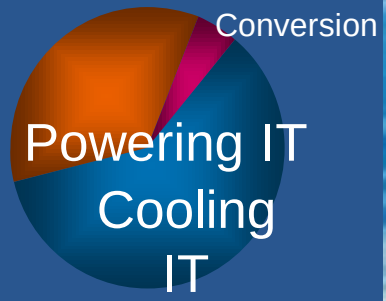
Agenda

- Business Challenges of Today
- HP strategy for power management
- Processor power saving
- Power Monitoring and control
- Power saving toolbox
- OpenVMS and power management
- Q & A

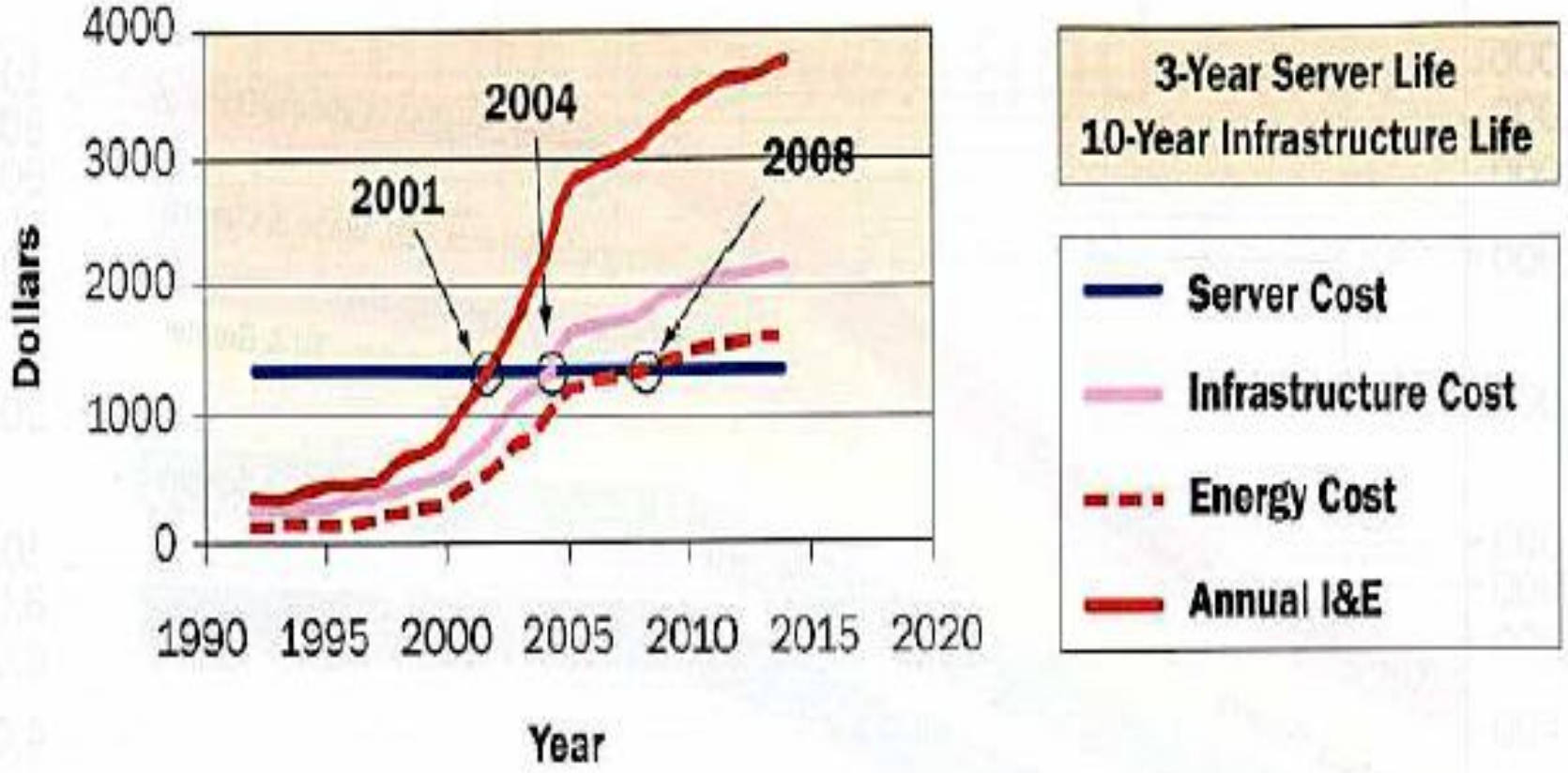
Business challenges of Today

Rising consumption of energy

- Cost of energy is rising
- Compute density is increasing
- Systems require more energy to power and cool
- Costs more to cool a server than to power it~



Business challenges of Today



*

Figure 3. Annual Amortized Costs in the Data Center for a 1U server.

HP Delivering a holistic Energy Efficient Solution

Optimizing from chip to chiller

chip

Energy Saving Solutions from the Server **Chip** to the Data Center Air **Chillers** and everything in-between

chiller

Itanium 9100 9300 Series, Xeon, AMD
Low power processors, Demand based switching,

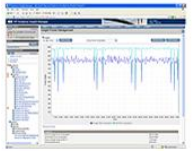
Efficient Systems
efficient power supplies

climate savers
smart computing

HP BladeSystem
HP Thermal Logic

Power Distribution Rack & MCS
3 phase UPS and liquid cooling

Insight Power Manager & iLO
Monitor and regulate energy./power



Virtualization and Consolidation
HP Virtual Server Environment

VSE

Storage Thin Provisioning / Dynamic Capacity Mgt

HP Services: Thermal Mapping
Thermal Assessments

HP Dynamic Smart Cooling: up to 45% cooling cost savings w/Mapping

HP Delivering a holistic Energy Efficient Solution

Optimizing from chip to chiller

chip

Energy Saving Solutions from the Server **Chip** to the Data Center Air **Chillers** and everything in-between

chiller

Itanium 9100 9300 Series, Xeon, AMD
Low power processors, Demand based switching,

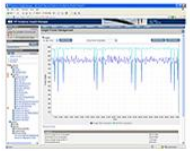
Efficient Systems
efficient power supplies

climate savers
smart computing

HP BladeSystem
HP Thermal Logic

Power Distribution Rack & MCS
3 phase UPS and liquid cooling

Insight Power Manager & iLO
Monitor and regulate energy./power



Virtualization and Consolidation
HP Virtual Server Environment

VSE

HP-UX 11i

Storage Thin Provisioning / Dynamic Capacity Mgt

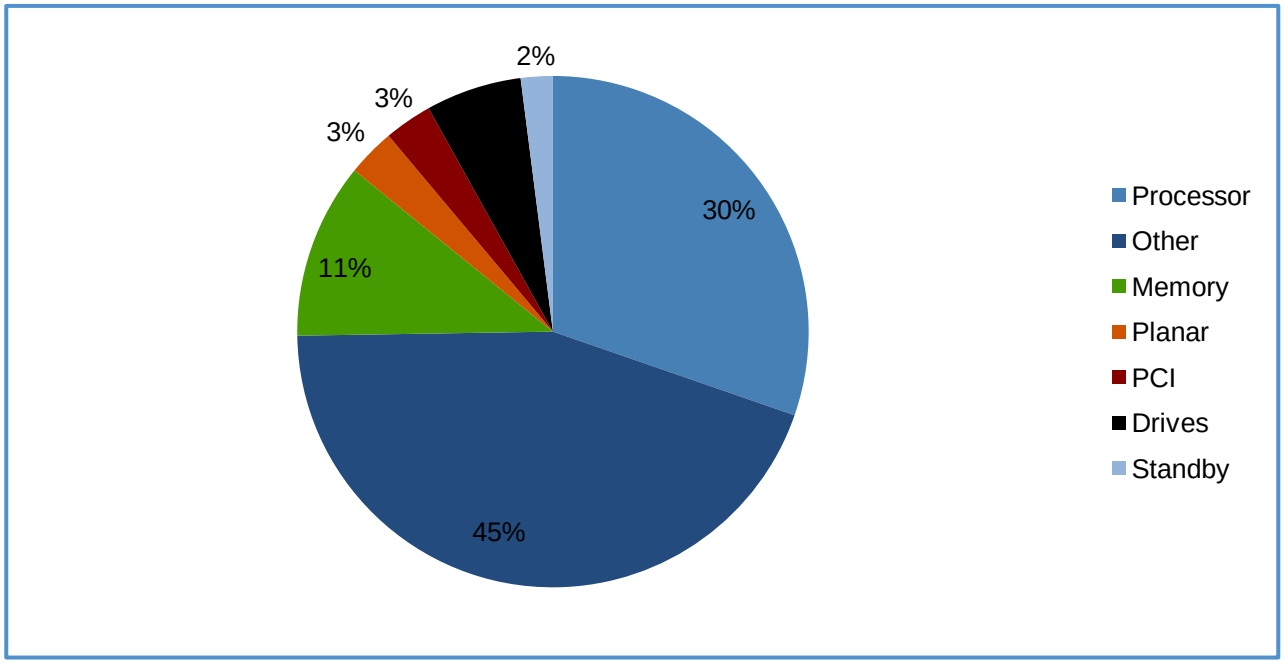
HP Services: Thermal Mapping
Thermal Assessments

HP Dynamic Smart Cooling: up to 45% cooling cost savings w/Mapping

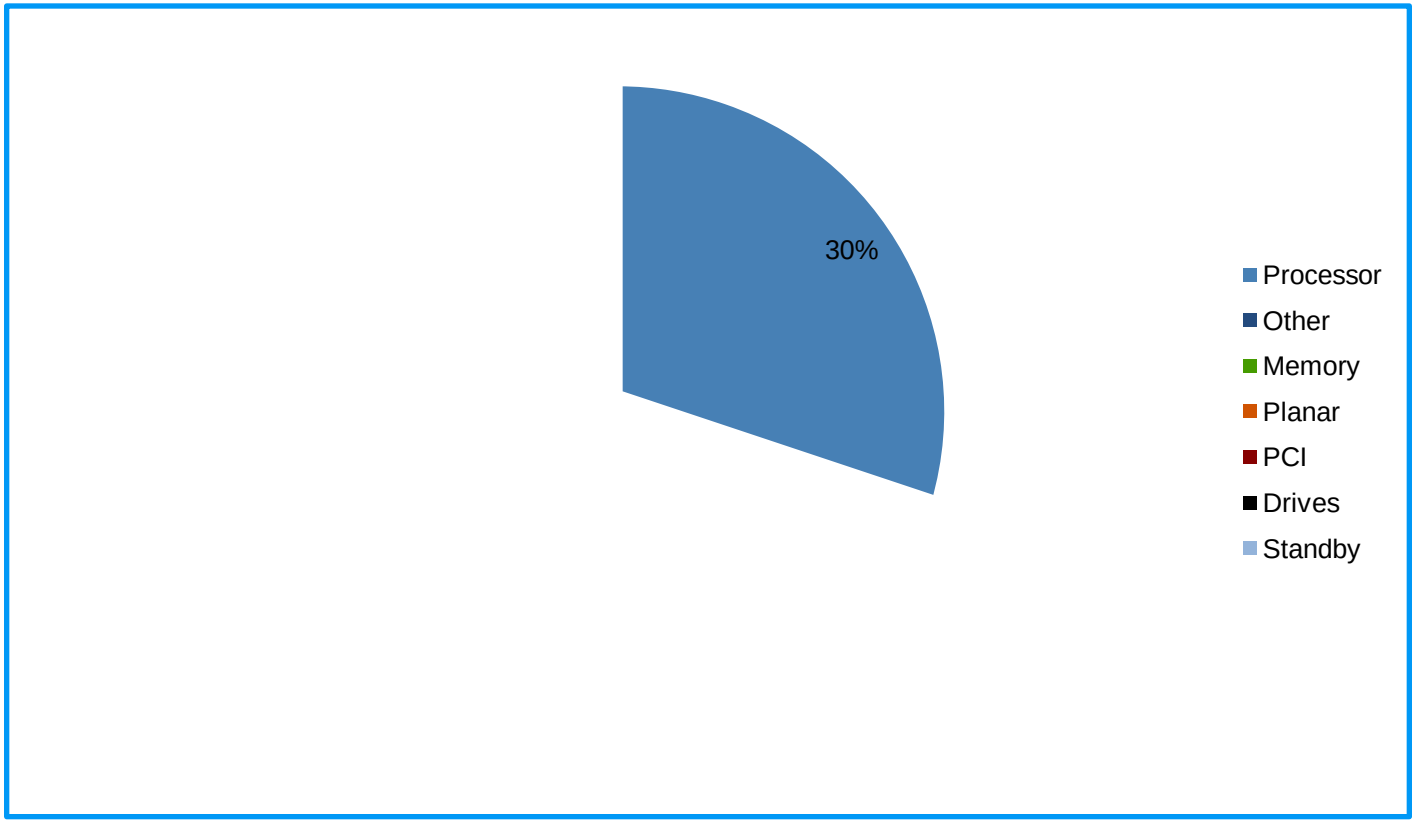
Power Utilization

– Data Center Power Utilization

– 50% for air-conditioning and power supply



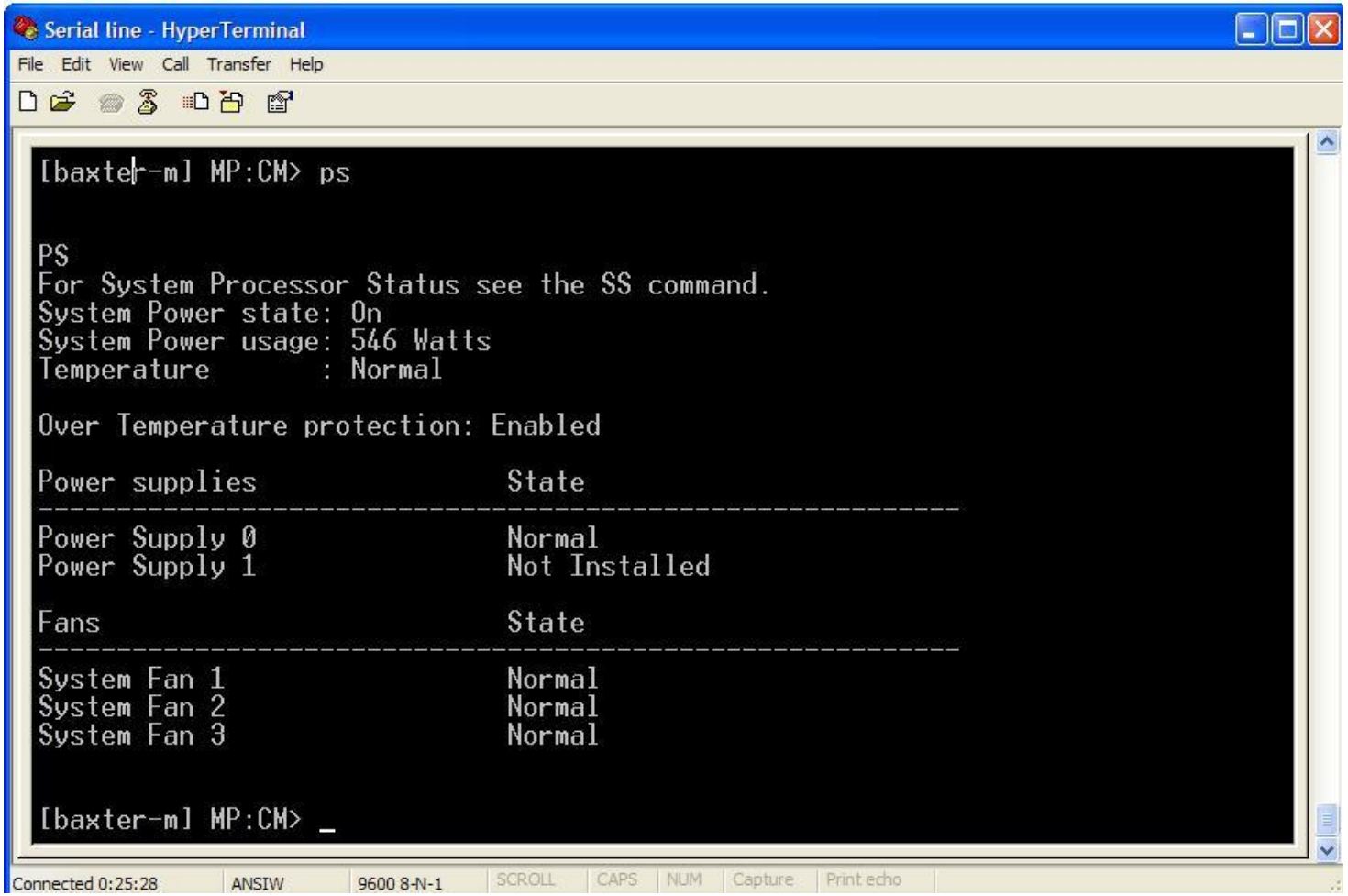
This talk covers…



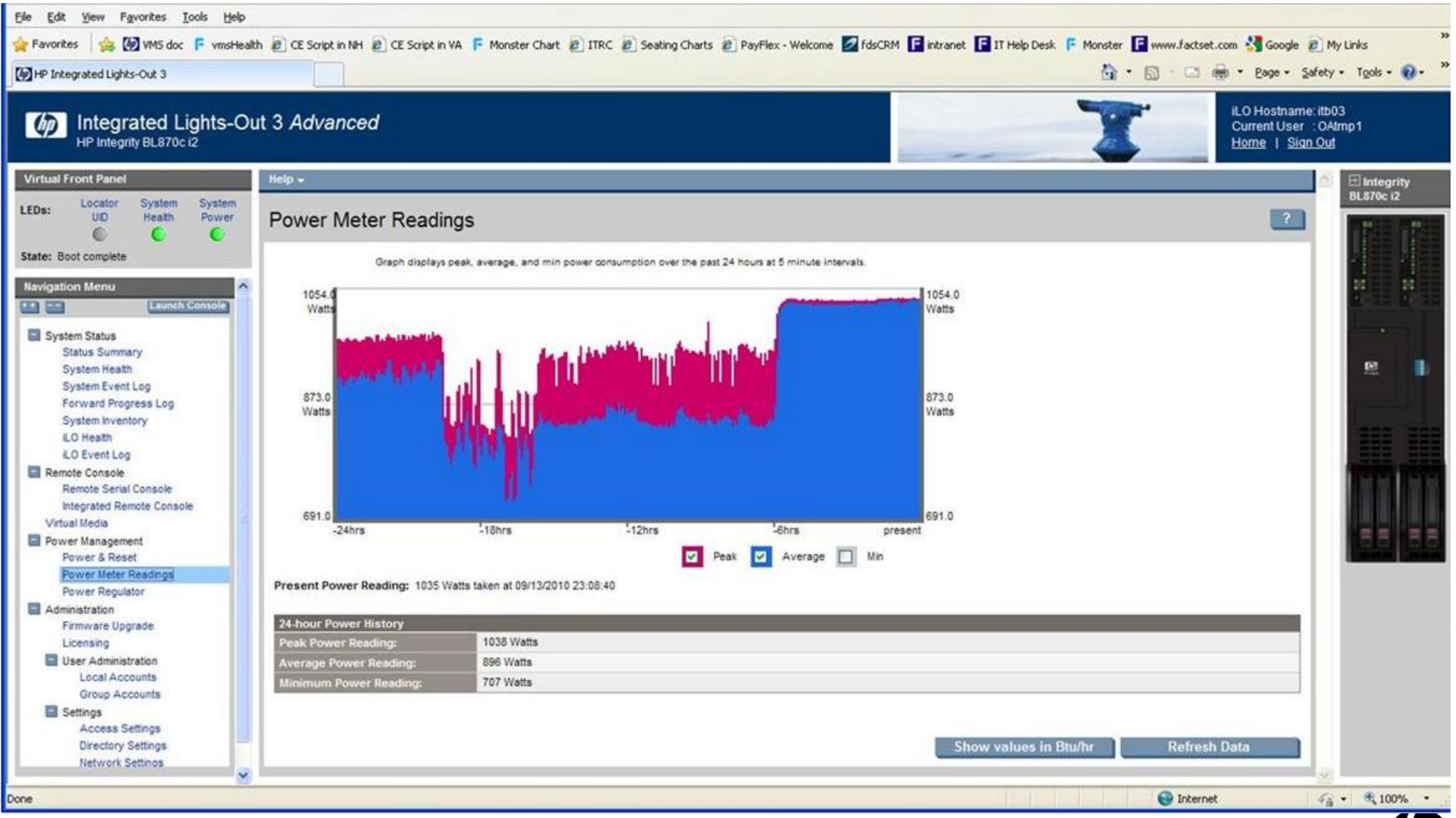
Power Monitoring

- Done autonomously by platform
- No support from OS required
- Available on many recent platforms
- Available via text, web

Power Monitoring via iLO Text Interface



Power Monitoring via iLO Web Interface



Power saving toolbox

–What does the OS have available to save power on Integrity Platforms (right now)?

- C-states
- P-states

Power Saving Toolbox

- C-states
  - Idle states; processor can't do work while saving power
  - C0 is normal "run" state
  - C1 is "stop processor but keep everything coherent"
  - On Itanium, C1 is entered with PAL_HALT_LIGHT
  - Exit C1 via an interrupt
  - Available on all VMS-supported I64 processors
    - Power reduction varies with processor
  - Increases interrupt latency since CPU must "turn back on"

Power Saving Toolbox

- P-states
  - Power/performance states: used when processor is active
  - P0 is the highest performance (and probably highest power)
  - Pn uses less power (and probably has lower performance) the higher ‘n’ goes. Nmax varies with processor
  - Only available on some variants of recent processors (starting with 9100)
  - Power/performance tradeoff varies with processor

iLO Commands

Static High Performance

- Don’t try to save any power.

Static Low Power

- Save power at the expense of performance

Dynamic (or efficiency)

- Use OS-defined scheme to make a reasonable compromise between power savings and performance

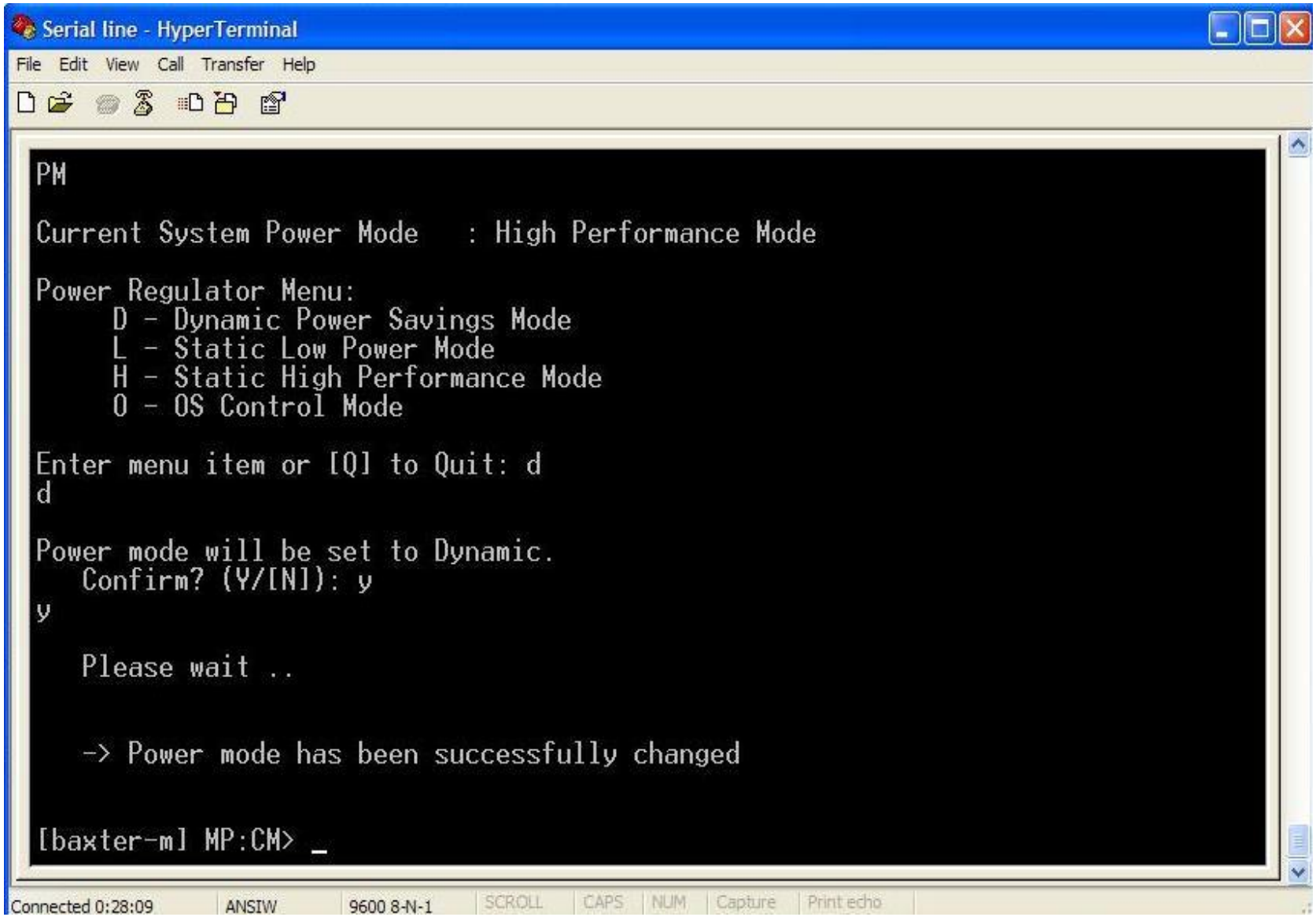
OS Control

- Use OS-defined interface to control power/performance decisions

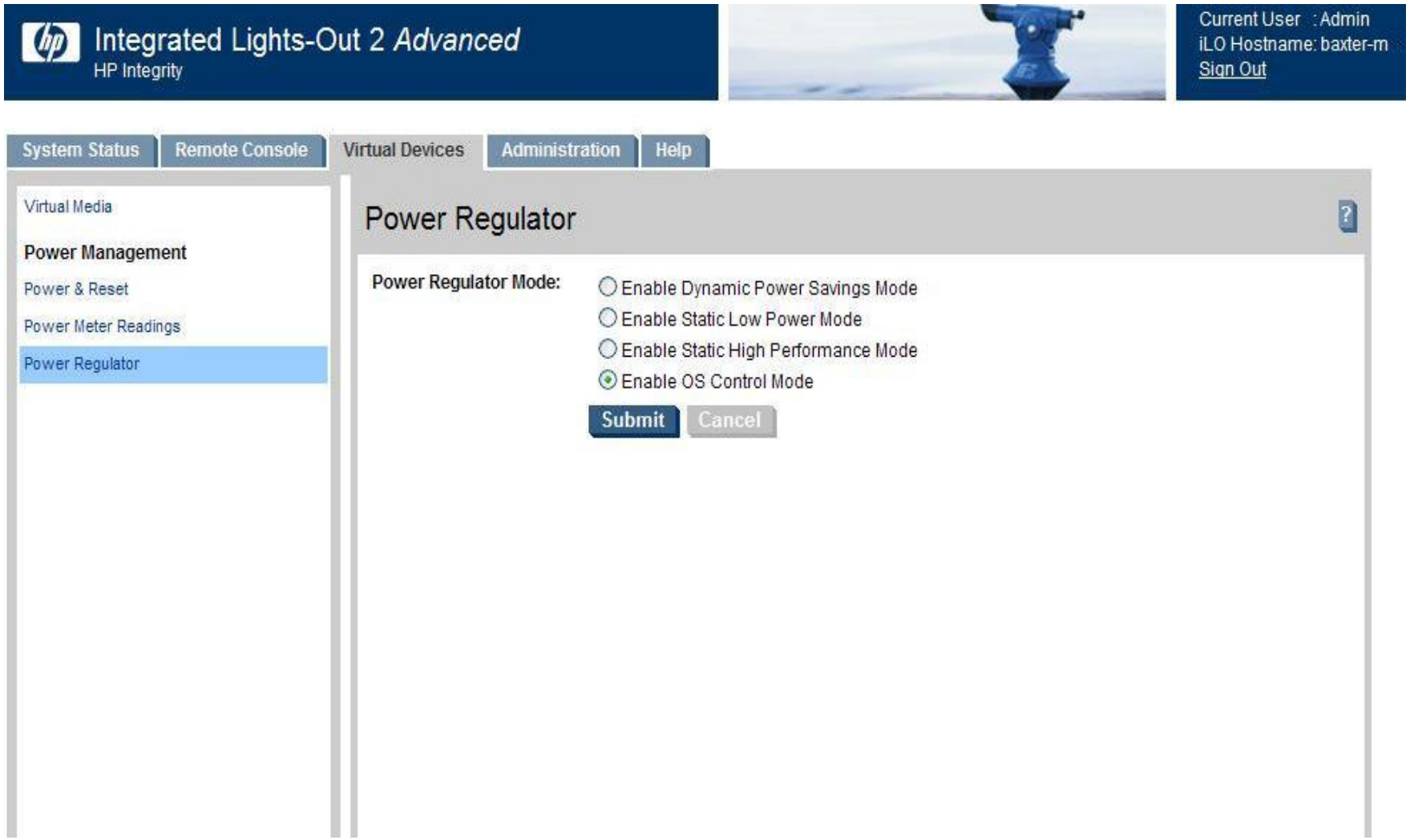
Power Control or Management

- Different controls and interfaces available in different platforms
- Interface available via iLO text, iLO web, or OS-specific.

Power Management – Serial Line



Power Management – iLO Web Interface



What has VMS done?

V8.2-1and V8.3

- C1 state when we predict CPU will be idling frequently
- Controlled by SYSGEN parameter CPU_POWER_MGMT and CPU_POWER_THRSH
- Default: Turned on
- Processors: All supported (different results)

V8.3-1H1

- Same algorithm available, but default is off

V8.4

- Use of p-states, Improved algorithm, default is on

What has VMS done in V8.4?

- Stopped (and iCAP) CPUs will go into C1 state
- When supported, iLO/IPM controls take precedence
- If not supported, VMS falls back to “OS Control”

Reminder: iLO Commands

Static High Performance

- Don’t try to save any power.

Static Low Power

- Save power at the expense of performance

Dynamic (or efficiency)

- Use OS-defined scheme to make a reasonable compromise between power savings and performance

OS Control

- Use OS-defined interface to control power/performance decisions

High Performance

- No C1 state in idle; just loop in P0/C0
- Never use any p-state other than P0
- In other words, just like V8.3-1H1's default

Low Power

- All CPUs run in Pn state (where n is the lowest power available)
- Always use C1 in idle

Dynamic/Efficiency

- Non-idle CPUs will be in P0 state
- Idle CPUs will choose C1 state using a new algorithm (by default)

OS Control

- Sysgen Parameter CPU_POWER_MGMT sets the default behavior when you switch the mode to OS_CONTROL
  - 0 – Just like high performance mode
  - 1 – Just like low power mode
  - 2 – Just like dynamic mode
- Parameter is dynamic. If you change it while in OS_CONTROL mode, power use changes

OS Control

- Can use new system service $POWER_CONTROL
- $POWER_CONTROL can choose
  - POWER$C_HIGH_PERF
  - POWER$C_LOW_POWER
  - POWER$C_EFFICIENCY
- Advantage: You can write a program to use in a batch job run on a schedule, or any other scheme you wish
- $POWER_CONTROL is flexible enough for us to add additional features later (no specific plans)
- Returns SS$_WRONGSTATE if called when not in OS_CONTROL state

OS Control

- Calling system service also changes the sysgen parameter
  - In other words, via the sysgen parameter, VMS remembers the last value set in OS_CONTROL mode
  - Sounds complex but I think it follows “principle of least surprise”

Idle Power Algorithms

- General idea
  - Go into C1 state in idle if interrupts are “not expected”
  - Algorithm decides when to use C1 state in idle
  - Trigger point determines when we stop using C-states

Idle Power Algorithms

- Trigger measurement
  - Old algorithm:
    - Percentage of time in idle for previous second
    - Measured by sampling idleness every 1 ms.
    - Threshold % chosen by CPU_POWER_THRSH
    - C-state decision made each second based on past second
  - New algorithm:
    - Number of exits from idle
    - Count each interrupt and each scheduler exit
    - If threshold exceded immediately stop C1 use

Idle Power Algorithms

- Resume power savings after threshold exceeded
  - Old algorithm
    - If percent of idle samples in previous second exceeds required threshold
  - New algorithm
    - If no 10-millisecond interval during the previous second exceeded the number-of-idle-exits threshold

New Idle Power Algorithm

–Summary

- New algorithm uses criteria more relevant to the behavior you want
- New algorithm switches off C-state idle to avoid interrupt latency more quickly
- New algorithm switches back to power savings and higher latency fairly slowly like previous algorithm

Summary

OpenVMS has had power saving mechanisms on Integrity since V8.2-1!

With V8.4—

OpenVMS takes part in HP's common power saving program

Power saving has better interrupt latency than before

Simplest interface via Management Processor web or serial interface-just three choices: High, low, compromise

System service available for more flexibility

Q & A