



# **IBM EnergyScale for POWER8 Processor-Based Systems**

*November 2015*

*Martha Broyles  
Christopher J. Cain  
Todd Rosedahl  
Guillermo J. Silva*

# Table of Contents

Executive Overview.....	4
EnergyScale Features.....	5
Power Trending.....	5
Thermal Reporting.....	5
Fixed Maximum Frequency Mode.....	6
Static Power Saver Mode.....	6
Fixed Frequency Override.....	6
Dynamic Power Saver Mode.....	7
Power Management's Effect on System Performance.....	7
Dynamic Power Saver Tunable Parameters.....	8
Idle Power Saver.....	8
Power Capping.....	9
“Soft” Power Capping.....	9
Error Handling for Power Capping.....	9
Power Capping Example.....	10
Energy-Optimized Fan Control.....	10
Power Supply Idle Detection.....	10
Processor Core Nap.....	10
Processor Folding.....	11
EnergyScale for I/O.....	11
User Interfaces.....	12
Overview.....	12
ASMI.....	13
Setting System Power Management Mode.....	14
Dynamic Power Saver Tunable Parameters.....	15
Example Tunable Parameters for a Fixed Maximum Frequency.....	16
Example Tunable Parameters for More Aggressive Power Savings.....	17
Idle Power Saver.....	18
CIM Client.....	19
Script to Read Average Power.....	20
Average Power Output.....	21
Script to Read Average Frequency.....	22
Average Frequency Output.....	23
Script to Read Thermal Data.....	24
Thermal Data Output.....	25
Script to Query and Set a Power Management Mode.....	26
Example Setting Fixed Frequency Override Mode.....	29
Script to Query and Set a Power Cap.....	30
Example Setting a Power Cap.....	34
HMC.....	35
Setting System Power Management Mode.....	35
DCMI.....	36
EnergyScale Operating System Support.....	37
Processor Folding in AIX.....	38
Processor Folding in IBM i.....	39
Processor Folding in Linux.....	40
Query Idle Power Saver and Dynamic Power Saver Tunable Parameters in AIX.....	41
Appendix I: System Requirements.....	42
Release Level FW810.....	43
Feature Support.....	43

Frequency.....	44
Release Level FW810.3.....	46
Feature Support.....	46
Frequency.....	47
Release Level FW820.....	49
Feature Support.....	49
Frequency.....	50
Release Level FW830.....	51
Feature Support.....	51
Frequency.....	52
Release Level FW840.....	53
Feature Support.....	53
Frequency.....	53
Appendix II: Estimating Total System Power Saved.....	54
Appendix III: Processor Usage and Accounting.....	55
Appendix IV: Resources.....	57

# Executive Overview

The energy required to power and cool computers can be a significant cost to a business – reducing profit margins and consuming resources. In addition, the cost of creating power and cooling infrastructure can be prohibitive to business growth. In response to these challenges, IBM developed EnergyScale™ Technology for IBM Power systems. EnergyScale provides functions that help the user to understand and control IBM server power and cooling usage. This enables better facility planning, provides energy and cost savings, enables peak energy usage control, and increases system availability. Administrators may leverage EnergyScale capabilities to control the power consumption and performance of POWER processor-based systems to meet their particular data center needs.

In this paper, the functions provided by EnergyScale are described along with usage examples, and hardware and software requirements. Support and awareness of EnergyScale extends throughout the system software stack, and is included in the AIX, IBM i, and Linux operating systems. This paper focuses on features and functions found in systems based on POWER8 processors. For previous-generation IBM Power systems, please refer to a companion paper, “[IBM EnergyScale for POWER6 Processor-based Systems](#)” or “[IBM EnergyScale for POWER7 Processor-based Systems](#)”.

# EnergyScale Features

EnergyScale provides many features to measure, monitor, and control both the power consumption and energy efficiency of POWER8 Systems. Not all features may be supported, see [Appendix I: System Requirements](#) for features supported for a specific release and system.

Function	Summary
Power Trending	Collects and reports power consumption information for a server available via CIM client. Limited power consumption information is available via DCMI.
Thermal Reporting	Collects and reports inlet and exhaust temperatures available via CIM client. Additional thermal information is available via DCMI.
Fixed Maximum Frequency Mode	New power management mode to run at maximum frequency allowed. Available starting with FW810.3 and FW830. Can be set via ASMI.
Static Power Saver	Provides predictable performance with power savings by reducing CPU frequency a fixed amount. Can be set via ASMI or CIM client.
Fixed Frequency Override	Allows a user to specify a specific frequency between the minimum and maximum allowed for the system to run at. Can be set via CIM client.
Dynamic Power Saver	Allows a system to implement algorithms to adjust the processor core frequency to favor system performance (saving power where applicable) or to balance power and performance. Core frequency may exceed 100% at times. Can be set via ASMI or CIM client.
Dynamic Power Saver Tunable Parameters	Allows the behavior in Dynamic Power Saver mode to be modified for specific system or workload needs. Only settable via ASMI.
Idle Power Saver	Reduced energy consumption when system is considered idle. Can be controlled via ASMI.
Power Capping	Enforces a user-specified power budget on a system may be set via CIM client or DCMI.
Energy-Optimized Fans	System fans respond dynamically to temperatures of system components
Power Supply Idle Detection	Enables a power supply to go into an idle mode when system power usage is low
Processor Core Nap	Enables low-power modes in POWER8 when cores are unused
Processor Folding	Dynamically re-allocates which processor cores execute a task to optimize energy efficiency of the entire system
EnergyScale for I/O	Powers on I/O slots only when needed

## Power Trending

EnergyScale provides continuous collection of real-time server power consumption. This power usage data may be collected using a CIM client such as wbemcli. Administrators may use this information to predict data center power consumption at various times of the day, week, or month. In addition, data may be aggregated to identify anomalies, manage electrical loads, and enforce system-level power budgets. Please refer to the Power Capping Example section below to learn how power trending data may be used to set a power budget, also referred to as a power cap.

## Thermal Reporting

A measured ambient temperature and a calculated exhaust heat index temperature can be read from a CIM client. Please note that the ambient temperature reported may vary from system model to system

model due to variations in the placement of the ambient temperature sensor in the system. This information can help identify data center “hot-spots” that need attention. Some additional thermal information is available via DCMI.

### **Fixed Maximum Frequency Mode**

Fixed Maximum Frequency is a new mode introduced in FW810.3 and FW830 releases. When enabled the system will run at its maximum frequency regardless of workload as long as power and thermal limits allow. In addition while in Fixed Maximum Frequency mode, Idle Power Save is disabled and cannot be enabled. See [Appendix I](#) for details on the maximum frequency by system and firmware release.

### **Static Power Saver Mode**

Static Power Saver lowers the processor frequency and voltage a fixed amount, reducing the power consumption of the system while still delivering predictable performance. This percentage is predetermined to be within a safe operating limit and is not user-configurable. In addition, some EnergyScale-ready operating systems automatically enable processor folding in dedicated processor partitions when Static Power Saver mode is enabled. See [Appendix I](#) for details on the actual percentage the frequency is lowered and estimated processor power saved by system and firmware release.

ASMI is the recommended user interface to enable/disable Power Saver mode. Power Saver could be enabled based on regular variations in workloads, such as predictable dips in utilization over night, or over weekends. Power Saver can be used to reduce peak energy consumption, which can lower the cost of all power used. Please note that when Power Saver is enabled for certain workloads with low CPU utilization, workload performance will not be impacted, though CPU utilization may increase due to the reduced processor frequency.

The only time that a system does not support operating at the Power Saver voltage and frequency is during a system boot or re-boot. Power Saver may be enabled at any time; however, if Power Saver was enabled prior to a system boot the voltage and frequency will remain at the default boot values until the platform firmware reaches a standby or running state. Immediately before the platform firmware starts executing on the system's processors, the voltage and frequency will drop to the power saver mode values. If a system re-boot occurs while in Power Saver mode, the voltage and frequency will return back to boot values, and following a successful system re-boot the voltage and frequency will be dropped back to power saver mode. The power saver mode setting persists across system boots, service processor resets and loss of ac power (unless the power outage is long enough to drain the NVRAM battery).

### **Fixed Frequency Override**

Fixed Frequency Override allows a user to specify a fixed frequency for all processors in the system to run at. The specified frequency must be within the system's supported frequency range. A query of the supported frequency range can be done from a CIM client. The system will stay running at the fixed frequency regardless of workload. In cases when the fixed frequency is set above nominal the frequency may be lowered due to thermal or power limits.

Fixed Frequency Override is mutually exclusive with the other power modes (i.e. Static Power Saver, Dynamic Power Saver). Only one of these modes may be enabled at a time. Note, Idle Power Saver is an independent function from the power modes and when enabled and idleness has been detected the system will still enter Idle Power Saver when in Fixed Frequency Override.

## Dynamic Power Saver Mode

Dynamic Power Saver varies processor frequency and voltage based on the utilization of the system's POWER8 processors. Processor frequency and utilization are inversely proportional for most workloads, which says that as the frequency of a processor increases, its utilization decreases, given a constant workload. Dynamic Power Saver takes advantage of this relationship to detect opportunities to save power, based on measured real-time system utilization.

Dynamic Power Saver may favor *system performance* or *system power savings*, the user selects which one to favor when setting this mode. When Dynamic Power Saver is enabled, system firmware continuously monitors the performance and utilization of each of the computer's POWER8 processor cores. Based on this utilization and performance data, the firmware dynamically adjusts the processor frequency and voltage, reacting in milliseconds to adjust workload performance and also deliver power savings when the system is under-utilized. In “Dynamic Power Saver, Favor Power” mode, system firmware will balance performance and power consumption. See [Tunable Settings for More Aggressive Power Savings](#) section to enable more aggressive power savings than the default favor power savings mode allows. In “Dynamic Power Saver, Favor Performance” mode, system firmware will default to the maximum processor core frequency allowed for a given system's environment and configuration, and reduce frequency only when a system is very lightly utilized or idle.

Please note that while Dynamic Power Saver Mode is available on all POWER8 processor-based systems, specific frequency limits may vary based on system type and configuration. Refer to [Appendix I: System Requirements](#) for the frequency limits for each system type. Dynamic Power Saver mode is mutually exclusive with Static Power Saver and Fixed Frequency Override mode. Only one of these modes may be enabled at a time.

### **Power Management's Effect on System Performance**

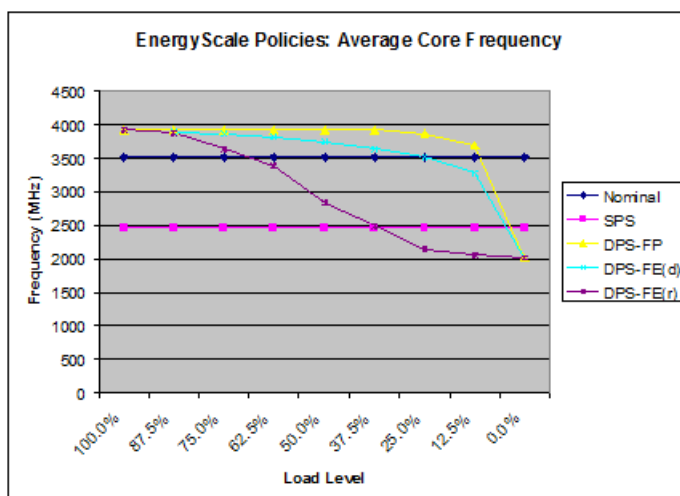
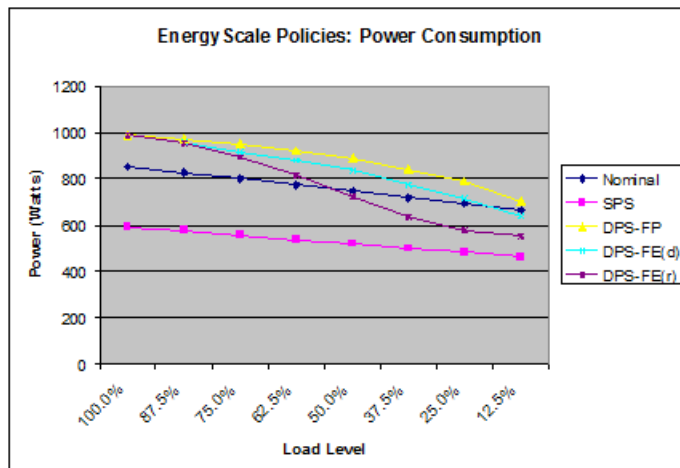
All power management modes can affect some aspects of performance. This is dependent on system configuration and how performance is measured.

The following issues must be considered before turning on any power management mode or feature:

1. Systems running at low utilization (and consequently, low frequency) may maintain processor throughput. However, response time to a particular task may be affected. Also, reaction time to an incoming workload may be affected.
2. Any system set-up that limits the amount of processing allowed, such as running with capped partitions, can cause the frequency to be reduced. Even though a partition may be running at 100% of its entitled capacity, the system as a whole may not be heavily utilized.
3. Using virtual shared processor pools also may limit the overall system utilization and cause lower processor frequencies to be set.
4. Some external workload managers also have the effect of limiting system processing by adjusting workloads to a point where frequency will be lowered.
5. Since the processor frequency will be variable, performance monitoring tools may be affected.

As shown in the graphs below from a representative POWER8 system, enabling the various power savings modes may directly affect power consumption as a workload varies. For example, “Dynamic Power Saver, Favor Performance” mode may deliver higher workload throughput than either of the other modes at the expense of system power consumption. At less than peak utilization, “Dynamic Power Saver, Favor Power” mode with the default parameters [DPS-FE(d) in graphs] will deliver

power savings, and may still be delivering adequate workload throughput. With the [more aggressive power saving settings](#) [DPS-FE(r) in graphs] the power savings is even greater. It is important to note that trade-offs must be made between energy consumption, workload response times, and throughput.



### Dynamic Power Saver Tunable Parameters

The tunable parameters can be used to modify the system behavior while dynamic power saver is enabled. This may be useful to properly balance the performance required with the energy savings desired. These parameters should not be changed unless the user is working directly with an IBM representative or has the proper level of expertise in the effects of these parameter changes. While the parameters will be shown in this paper, no attempt will be made to explain in detail the system energy saving or performance that will result from the parameter modification.

### Idle Power Saver

This mode, which is enabled/disabled independently from all other modes and functions, reduces the energy usage to a low level *when the system is determined to be idle*. When idle, the voltage/frequency of the processor is reduced to the minimum and static power saver mode is reported to the OS to enable processor folding. When not idle the system is managed in accordance to the configured power mode



(i.e. Dynamic Power Saver, Fixed Frequency Override). Idle Power Saver can be enabled/disabled via ASMI. Additionally, the utilization levels that determine idleness and the time delays for entry/exit can also be modified from ASMI. See the ASMI section for more details. See [Appendix I](#) for system support details such as which systems support this mode and which systems have it enabled by default.

## **Power Capping**

Power Capping enforces a user-specified limit on power consumption. The user may set and enable a power cap using a CIM client or DCMI. In most data centers and other installations, when a server is installed, a certain amount of power is allocated to it. Generally, the amount is what is considered to be a “safe” value, and it typically has a large margin of reserved, extra power that is never used. This is called the *margin power*. The main purpose of the power cap is not to save power but rather to give a data center operator the capability to reallocate power from existing systems to new systems by reducing the margin assigned to the existing servers. That is, power capping gives an operator the capability to add extra servers to a data center which previously had all available power allocated to its existing systems. It does this by guaranteeing that a system will not use more power than assigned to it by the operator.

Previously, the data center administrator had to plan for the power consumption of the data center based on the Underwriters' Laboratories (UL) rating on the back of the servers being installed. The UL rating (commonly referred to as “label power”) on today's servers indicates the most power that a system could ever draw and is based on the capacity of the power supplies. It has to take into account a fully-configured system with the highest power-usage parts installed at the highest possible utilization.

### **“Soft” Power Capping**

There are two power ranges into which the power cap may be set. When a power cap is set in the guaranteed range (described above), the system is guaranteed to use less power than the cap setting. In order to meet this guarantee, extreme system configuration and environmental conditions must be accounted for. Setting a power cap in this region allows for the recovery of the margin power, but in many cases cannot be used to save power. Soft power capping extends the allowed power capping range further, beyond a region that can be guaranteed in all configurations and conditions. By setting a power cap in this soft region, the system can be set to save power by running at a lower power/performance point. If the power management goal is to meet a particular power consumption limit, then soft power capping is the mechanism to use. The performance impacts of a particular power cap setting can be determined by reading the power and CPU speed information from a CIM client.

### **Error Handling for Power Capping**

Certain failures can cause the system firmware to enter a “safe mode.” All power and thermal trending data will no longer be reported. There will always be an error log entry generated when safe mode is entered. When this mode is entered, the firmware will drop the system to a safe low power mode to guarantee that any hard power cap set by the client is still held.

Note that the failure to enforce a “soft” power cap below the minimum guaranteed range is not an error, and will not result in the system firmware entering “safe mode.” The system will continue to operate normally with all EnergyScale features at the minimum-supported frequency until the soft power cap is disabled or raised.

To re-enable full EnergyScale functionality, a reboot, firmware update, FRU replacement, or complete A/C power cycle of the system must be completed. If the problem which originally caused the system to enter safe mode is still present, the system will re-enter safe mode and generate an additional error log entry.

### **Power Capping Example**

An installation<sup>1</sup> has twelve JS22 servers. The UL rating on a JS22 server is ~402w (dc). Based on that value, the data center manager has budgeted ~4.8KW (dc) of power in the data center. The system operators connect Active Energy Manager to the twelve JS22s, and notice that the systems are actually only consuming a maximum of 308W (dc) of power while running their normal workloads. The firmware, based on configuration, allows a minimal power cap of 368W (dc). Based on the trending the operators have seen, they are confident in setting a power cap at the minimum of 368W (dc). The operator sets the power cap of each of the twelve systems to 368W (dc), and the data center manager is now guaranteed that the average power over time for each system will never exceed 368W (dc). This frees up ~408W(dc) power within the data center as indicated below.

$$(402W(dc) - 368W(dc)) * 12 = 408W(dc)$$

The data center manager can now add a thirteenth JS22 if needed.

### **Energy-Optimized Fan Control**

Cooling fans contribute significantly to the overall power consumption of a given computer. In order to minimize energy expended on cooling and to minimize the energy wasted “over-cooling” a system, firmware on all POWER8 systems will adjust fan speed in response to real-time temperatures of the system components. Note that in comparison to previous-generation systems, exhaust temperatures may increase, however all components will still be within allowed RAS temperature envelopes. This is a natural product of the optimization of fan speed, component temperature, and fan power consumption.

### **Power Supply Idle Detection**

When the system power usage is low, one supply will go into an idle mode to reduce power consumption. The other will stay active. In the case of an increased load, the power supply will automatically come out of the idle mode and begin to source current. There is no risk of system shutdown due to this function. Additionally, it is normal, under low load conditions, to measure reduced current draw from the power supply that is in idle mode and increased current draw from the power supply that is active.

### **Processor Core Nap**

IBM POWER processors uses a low-power mode called Nap that stops processor execution when there is no work to do on a particular processor core. The latency of exiting Nap falls within a partition dispatch (context switch) such that the hypervisor firmware can use it as a general purpose idle state. When the Operating System detects that a processor thread is idle, it yields control of a hardware thread to the hypervisor. The hypervisor immediately puts the thread into Nap. If the processor core is in a shared processor pool (the set of cores being used for micro-partition dispatching) and there is no micro-partition to dispatch, the hypervisor puts the thread into Nap mode. When all hardware threads running on a given processor core enter Nap mode, the whole core enters Nap mode, which allows the hardware to clock off most of the circuits inside the processor core. Reducing active power consumption by turning off the clocks allows the temperature to fall, which further reduces leakage (static) power of the circuits resulting in a cumulative effect. On some systems, unlicensed cores are kept in nap mode until they are licensed, and return to nap mode when they are unlicensed again.

---

<sup>1</sup> It is expected that all electrical installation guides are followed.

## **Processor Folding**

While Processor Core Nap provides substantial energy savings when processors become idle, additional savings can be realized if processors remain idle by intent rather than by happenstance. Processor Folding is a consolidation technique that dynamically adjusts, over the short-term, the number of processors available for dispatch to match the number of processors demanded by the workload. As the workload increases, the number of processors made available increases; as the workload decreases, the number of processors made available decreases. Processor Folding increases energy savings during periods of low to moderate workload because unavailable processors remain in low-power idle states longer than they otherwise would. Since the idle condition is intentional, the hypervisor is also advised to exploit special purpose idle states available on some POWER8 systems that can reduce power consumption even further than with Nap mode alone, but without the stringent latency requirement. Processor Folding achieves power savings similar to those that could be achieved by intelligent, utilization-based logical partition (LPAR) configuration changes, but it does so with much greater efficiency and fidelity, and without impacting the configuration or processor utilization of the LPAR.

Some EnergyScale-ready Operating Systems automatically enable Processor Folding in dedicated processor LPARs when Static Power Saver mode is enabled. Additionally, when a system enters Idle Power Saver mode this enables processor folding and allows deeper processor idle power states to be entered.

## **EnergyScale for I/O**

IBM Power Systems automatically power off pluggable PCI adapter slots that are not being used to save approximately 14 watts per slot. A PCI adapter slot is considered not being used when the slot is empty, when the slot is not assigned to a partition, or when the partition to which the slot is assigned is not powered on. A PCI slot is powered off immediately by system firmware when it is dynamically removed from the partition to which it was assigned, and when the partition to which it is assigned is powered off. Furthermore, system firmware automatically scans all pluggable PCI slots at regular intervals looking for those that meet the criteria for being not in use and powers them off. This ensures among other things that slots left on after platform power-on are subsequently powered off if they are not in use. This is supported on all POWER8 processor-based systems, and the expansion units that they support. Note that it applies to hot-pluggable PCI slots only. Power controls for other types of I/O features and built-in, or embedded, PCI adapters are not available and so they cannot be powered off independently from their enclosure power.

# User Interfaces

## Overview

POWER8 systems do not support Active Energy Manager (AEM). A CIM client (i.e. wbemcli) can be used to access EnergyScale features that were previously supported by AEM.

The table below summarizes the ASMI, CIM interface, and HMC support. Refer first to [Appendix I: System Requirements](#) to know if the specific feature is supported for a particular release and system.

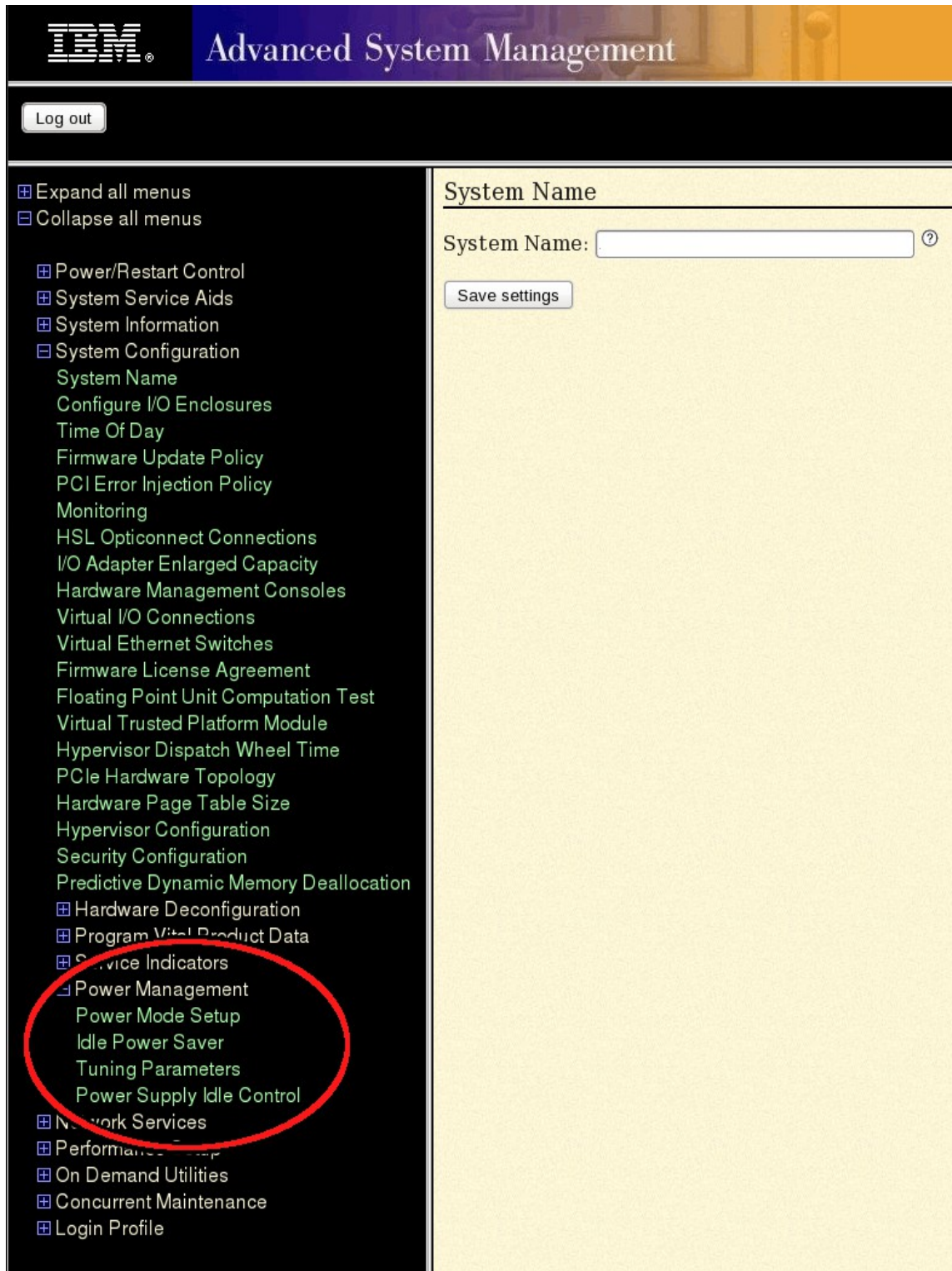
\* HMC support was made available starting with HMC release 8.2.0.

\*\* DCMI support was made available on some systems starting with FW840 release.

	<i>ASMI</i>	<i>CIM Client</i>	<i>HMC*</i>	<i>DCMI**</i>
<b>Power Trending</b>	N	Y	N	Y
<b>Thermal Reporting</b>	N	Y	N	Y
<b>Fixed Maximum Frequency Mode</b>	Y	N	Y	N
<b>Static Power Saver</b>	Y	Y	Y	N
<b>Fixed Frequency Override</b>	N	Y	N	N
<b>Dynamic Power Saver</b>	Y	Y	Y	N
<b>Dynamic Power Saver Tunable Parameters</b>	Y	N	N	N
<b>Idle Power Saver</b>	Y	N	N	N
<b>Power Capping</b>	N	Y	N	Y
<b>Power Supply Idle</b>	Y	N	N	N

# ASMI

Supported EnergyScale features can be found on the Advanced System Management Interface (ASMI) under the “Power Management” menu.



# Setting System Power Management Mode



Log out

- Expand all menus
- Collapse all menus

- Power/Restart Control
- System Service Aids
- System Information
- System Configuration
  - System Name
  - Configure I/O Enclosures
  - Time Of Day
  - Firmware Update Policy
  - PCI Error Injection Policy
  - Monitoring
  - HSL Opticonnect Connections
  - I/O Adapter Enlarged Capacity
  - Hardware Management Consoles
  - Virtual I/O Connections
  - Virtual Ethernet Switches
  - Firmware License Agreement
  - Floating Point Unit Computation Test
  - Virtual Trusted Platform Module
  - Hypervisor Dispatch Wheel Time
  - PCIe Hardware Topology
  - Hardware Page Table Size
  - Hypervisor Configuration
  - Security Configuration
  - Predictive Dynamic Memory Deallocation
- Hardware Deconfiguration
- Program Vital Product Data
- Service Indicators
- Power Management
  - Power Mode Setup**
  - Idle Power Saver
  - Tuning Parameters
  - Power Supply Idle Control
- Network Services
- Performance Setup
- On Demand Utilities
- Concurrent Maintenance
- Login Profile

### Power Mode Setup

Current Power Saver Mode : Disable Power Saver mode

- Disable Power Saver mode ?
- Enable Static Power Saver mode ?
- Enable Dynamic Power Saver (favor power) mode ?
- Enable Dynamic Power Saver (favor performance) mode ?

Note: Enabling any of the Power Saver modes will cause changes in the processor frequencies, changes in processor utilization, changes in power consumption, and performance to vary. Other effects are possible as well. Please see the EnergyScale™ white paper for more information on power saving modes.

Continue ?

## Dynamic Power Saver Tunable Parameters

Note: These parameters are for advanced users only. Do not adjust them unless directed to do so by IBM. The “Reset Parameters” button on the bottom will restore all parameters to their default value.

**IBM** Advanced System Management Copyright © 2002, 2014 IBM Corporation. All rights reserved.

Log out

- Expand all menus
- Collapse all menus
- Power/Restart Control
- System Service Aids
- System Information
- System Configuration
  - System Name
  - Configure I/O Enclosures
  - Time Of Day
  - Firmware Update Policy
  - PCI Error Injection Policy
  - Monitoring
  - HSL Opticonnect Connections
  - I/O Adapter Enlarged Capacity
  - Hardware Management Consoles
  - Virtual I/O Connections
  - Virtual Ethernet Switches
  - Firmware License Agreement
  - Floating Point Unit Computation Test
  - Virtual Trusted Platform Module
  - Hypervisor Dispatch Wheel Time
  - PCIe Hardware Topology
  - Hardware Page Table Size
  - Hypervisor Configuration
  - Security Configuration
  - Predictive Dynamic Memory Deallocation
  - Hardware Deconfiguration
  - Program Vital Product Data
  - Service Indicators
  - Power Management
    - Power Mode Setup
    - Idle Power Saver
      - Tuning Parameters**
      - Power Supply Idle Control
  - Network Services
  - Performance Setup
  - On Demand Utilities
  - Concurrent Maintenance
  - Login Profile

### Tuning Parameters

Utilization threshold for increasing frequency  
Current value: 99.9%  
New value:  Range: MinVal- 0.0% MaxVal- 100.0%

Utilization threshold for decreasing frequency  
Current value: 99.9%  
New value:  Range: MinVal- 0.0% MaxVal- 100.0%

Number of samples for computing utilization statistics  
Current value: 4  
New value:  Range: MinVal-1 MaxVal-1024

Step size for going up in frequency  
Current value: 0.8%  
New value:  Range: MinVal- 0.1% MaxVal- 100.0%

Step size for going down in frequency  
Current value: 0.8%  
New value:  Range: MinVal- 0.1% MaxVal- 100.0%

Delta percentage for determining active cores  
Current value: 18%  
New value:  Range: MinVal-0% MaxVal-100%

Utilization threshold to determine active cores with slack  
Current value: 98.0%  
New value:  Range: MinVal- 0.0% MaxVal- 100.0%

Enable/disable frequency delta between cores  
Current value: Disabled  
New value:

Maximum frequency delta between cores  
Current value: 10%  
New value:  Range: MinVal-10% MaxVal-100%

Note: It is recommended that only advanced users modify these values since it can result in unexpected behavior and performance impacts. Please see the EnergyScale™ white paper for more information on tuning parameters.

Save settings ?

**Reset Parameters ?**

### ***Example Tunable Parameters for a Fixed Maximum Frequency***

Shown below are changed parameter settings to have the system to attempt to run at its maximum processor frequency regardless of utilization.

NOTE: Starting in FW810.3 and FW830 releases this may also be achieved by setting the power management mode to “Fixed Maximum Frequency”.

<b>Utilization threshold for increasing frequency</b>	Default
<b>Utilization threshold for decreasing frequency</b>	0.0%
<b>Number of samples for computing utilization statistics</b>	Default
<b>Step size for going up in frequency</b>	100.0%
<b>Step size for going down in frequency</b>	Default
<b>Delta percentage for determining active cores</b>	0.0%
<b>Utilization threshold to determine active cores with slack</b>	0.0%
<b>Enable/disable frequency delta between cores</b>	Default
<b>Maximum frequency delta between cores</b>	Default



### **Example Tunable Parameters for More Aggressive Power Savings**

Shown below are changed parameter settings to have the system more aggressively save power than the default Dynamic Power Saver – Favor Power mode allows.

NOTE: Starting in FW810.3 and FW830 releases this is no longer needed as the defaults for Dynamic Power Saver – Favor Power mode were changed to the more aggressive power saving setting.

<b>Utilization threshold for increasing frequency</b>	Default
<b>Utilization threshold for decreasing frequency</b>	Default
<b>Number of samples for computing utilization statistics</b>	16
<b>Step size for going up in frequency</b>	Default
<b>Step size for going down in frequency</b>	Default
<b>Delta percentage for determining active cores</b>	Default
<b>Utilization threshold to determine active cores with slack</b>	Default
<b>Enable/disable frequency delta between cores</b>	Default
<b>Maximum frequency delta between cores</b>	Default

# Idle Power Saver



Log out

- Expand all menus
- Collapse all menus
  
- Power/Restart Control
- System Service Aids
- System Information
- System Configuration
  - System Name
  - Configure I/O Enclosures
  - Time Of Day
  - Firmware Update Policy
  - PCI Error Injection Policy
  - Monitoring
  - HSL Opticonnect Connections
  - I/O Adapter Enlarged Capacity
  - Hardware Management Consoles
  - Virtual I/O Connections
  - Virtual Ethernet Switches
  - Firmware License Agreement
  - Floating Point Unit Computation Test
  - Virtual Trusted Platform Module
  - Hypervisor Dispatch Wheel Time
  - PCIe Hardware Topology
  - Hardware Page Table Size
  - Hypervisor Configuration
  - Security Configuration
  - Predictive Dynamic Memory Deallocation
- Hardware Deconfiguration
- Program Vital Product Data
- Service Indicators
- Power Management
  - Power Mode Setup
    - Idle Power Saver**
    - Tuning Parameters
    - Power Supply Idle Control
- Network Services
- Performance Setup
- On Demand Utilities
- Concurrent Maintenance
- Login Profile

### Idle Power Saver

#### Idle Power Saver Enable

Current value: Enabled

New value:

#### Delay Time to Enter Idle Power

Current value: 240seconds

New value:  Range: MinVal-10seconds MaxVal-600seconds

#### Utilization Threshold to Enter Idle Power

Current value: 8%

New value:  Range: MinVal-1% MaxVal-95%

#### Delay Time to Exit Idle Power

Current value: 10seconds

New value:  Range: MinVal-10seconds MaxVal-600seconds

#### Utilization Threshold to Exit Idle Power

Current value: 12%

New value:  Range: MinVal-5% MaxVal-95%

Note: Selecting a utilization threshold to enter idle power that is higher than the utilization threshold to exit idle power will result in unexpected behavior. Please see the EnergyScale™ white paper for more information on Idle Power Saver.

?

?

## ***CIM Client***

This section will provide a few example scripts for some features, this is not meant to show everything that can be done via CIM. For more information search for “EnergyScale” or “EnergyStar” on [www.ibm.com](http://www.ibm.com) to find the “Manual For Using WBEM CLI to Fetch Flexible Service Processor CIM Data”.

## Script to Read Average Power

```
#!/usr/bin/perl
my $system="";
my $uid="HMC";
my $pwd="xxxxxx";
my $verbose=0;

if ($#ARGV >= 0)
{
    $system="$ARGV[$ii]";
}
else
{
    print "ERROR: Must supply system name or IP address\n";
    exit 1;
}
if ($verbose)
{
    print "==> wbemcli -nl -noverify ei \"https://${uid}:${pwd}\@$
{system}:5989/root/ibmsd:fips_powermetricvalue\" | grep -e \"-InstanceID\" -e \"-
MeasuredElementName\" -e \"-TimeStamp\" -e \"-MetricValue\"\n";
}
@result=`wbemcli -nl -noverify ei "https://${uid}:${pwd}\@$
{system}:5989/root/ibmsd:fips_powermetricvalue" | grep -e "-InstanceID" -e "-
MeasuredElementName" -e "-TimeStamp" -e "-MetricValue"`;
if ( $#result < 0 )
{
    print "ERROR: Failed to find Power Trending Data\n";
}
else
{
    print "Average Power Trending Data for $system: (Watts)\n";
    print "  Timestamp                Watts Instance Identifier\n";
    print "  -----                -----
-----\n";
    foreach my $line (@result)
    {
        if ($line =~ /-MetricValue=(\d*)/)
        {
            $value=$1;
        }
        elsif ($line =~ /-TimeStamp=(\S*)/)
        {
            $timestamp=$1;
        }
        elsif ($line =~ /-MeasuredElementName=("[^"]*)/)
        {
            $element=$1;
        }
        elsif ($line =~ /-InstanceID=("[^"]*)/)
        {
            $instance=$1;
            if ($instance =~ /Avg/)
            {
                printf("  $timestamp    %5d $element\n", $value);
            }
        }
    }
}
exit 0;
```

### Average Power Output

Timestamp	Watts	Instance Identifier
20140606140700.000000+000	343	CECDrawer 00001E00
20140606140730.000000+000	343	CECDrawer 00001E00
20140606140800.000000+000	343	CECDrawer 00001E00
20140606140830.000000+000	343	CECDrawer 00001E00
20140606140900.000000+000	343	CECDrawer 00001E00
:	:	:
:	:	:
20140606150400.000000+000	342	CECDrawer 00001E00
20140606150430.000000+000	342	CECDrawer 00001E00
20140606150500.000000+000	342	CECDrawer 00001E00
20140606150530.000000+000	342	CECDrawer 00001E00
20140606150600.000000+000	342	CECDrawer 00001E00
20140606140700.000000+000	226	Power Supply U78C9.001.D123456-E1 00001000
20140606140730.000000+000	226	Power Supply U78C9.001.D123456-E1 00001000
20140606140800.000000+000	226	Power Supply U78C9.001.D123456-E1 00001000
20140606140830.000000+000	226	Power Supply U78C9.001.D123456-E1 00001000
20140606140900.000000+000	226	Power Supply U78C9.001.D123456-E1 00001000
:	:	:
:	:	:
20140606150400.000000+000	226	Power Supply U78C9.001.D123456-E1 00001000
20140606150430.000000+000	226	Power Supply U78C9.001.D123456-E1 00001000
20140606150500.000000+000	226	Power Supply U78C9.001.D123456-E1 00001000
20140606150530.000000+000	226	Power Supply U78C9.001.D123456-E1 00001000
20140606150600.000000+000	226	Power Supply U78C9.001.D123456-E1 00001000
20140606140700.000000+000	194	Power Supply U78C9.001.D123456-E2 00001001
20140606140730.000000+000	194	Power Supply U78C9.001.D123456-E2 00001001
20140606140800.000000+000	194	Power Supply U78C9.001.D123456-E2 00001001
20140606140830.000000+000	194	Power Supply U78C9.001.D123456-E2 00001001
20140606140900.000000+000	194	Power Supply U78C9.001.D123456-E2 00001001
:	:	:
:	:	:
20140606150400.000000+000	194	Power Supply U78C9.001.D123456-E2 00001001
20140606150430.000000+000	194	Power Supply U78C9.001.D123456-E2 00001001
20140606150500.000000+000	194	Power Supply U78C9.001.D123456-E2 00001001
20140606150530.000000+000	194	Power Supply U78C9.001.D123456-E2 00001001
20140606150600.000000+000	194	Power Supply U78C9.001.D123456-E2 00001001

## Script to Read Average Frequency

```
#!/usr/bin/perl
my $system="";
my $uid="HMC";
my $pwd="xxxxxx";
my $verbose=0;

if ($#ARGV >= 0)
{
    $system="$ARGV[$i]";
}
else
{
    print "ERROR: Must supply system name or IP address\n";
    exit 1;
}

if ($verbose)
{
    print "==> wbemcli -nl -noverify ei \"https://${uid}:${pwd}@${system}:5989/root/ibmsd:fips_cpuousagemetricvalue\" | grep -e \"-InstanceID\" -e \"-MeasuredElementName\" -e \"-TimeStamp\" -e \"-MetricValue\" \n";
}
@result=`wbemcli -nl -noverify ei "https://${uid}:${pwd}@${system}:5989/root/ibmsd:fips_cpuousagemetricvalue" | grep -e "-InstanceID" -e "-MeasuredElementName" -e "-TimeStamp" -e "-MetricValue"`;
if ( $#result < 0 )
{
    print "ERROR: Failed to find Processor Frequency Trending Data\n";
}
else
{
    print "Processor Frequency Trending Data for $system: (MHz)\n";
    print "  Timestamp                MHz    Instance Identifier\n";
    print "  -----                ----"
    -----\n";
    foreach my $line (@result)
    {
        if ($line =~ /-MeasuredElementName=.(["]*)/)
        {
            $instance=$1;
        }
        elsif ($line =~ /-TimeStamp=(\S*)/)
        {
            $timestamp=$1;
        }
        elsif ($line =~ /-MetricValue=(\d*)/)
        {
            $value=$1;
            printf("  $timestamp    %4d  $instance\n", $value);
        }
    }
}
exit 0;
```

### **Average Frequency Output**

Timestamp	MHz	Instance Identifier
20140606140700.000000+000	2060	CECDrawer 00001E00
20140606140730.000000+000	2060	CECDrawer 00001E00
20140606140800.000000+000	2060	CECDrawer 00001E00
20140606140830.000000+000	2060	CECDrawer 00001E00
:	:	:
:	:	:
20140606145530.000000+000	2060	CECDrawer 00001E00
20140606145600.000000+000	2060	CECDrawer 00001E00

## Script to Read Thermal Data

```
#!/usr/bin/perl
my $system="";
my $uid="HMC";
my $pwd="xxxxxx";
my $verbose=0;

if ($#ARGV >= 0)
{
    $system="$ARGV[$ii]";
}
else
{
    print "ERROR: Must supply system name or IP address\n";
    exit 1;
}

if ($verbose)
{
    print "==> wbemcli -nl -noverify ei \"https://${uid}:${pwd}\@$
{system}:5989/root/ibmsd:fips_thermalmetricvalue\"\n";
}
@result=`wbemcli -nl -noverify ei "https://${uid}:${pwd}\@$
{system}:5989/root/ibmsd:fips_thermalmetricvalue"`;
if ( $#result < 0 )
{
    print "ERROR: Failed to find Thermal Trending Data\n";
}
else
{
    print "Thermal Trending Data for $system: (1/100th degree C)\n";
    print "   Timestamp                Temp   Instance Identifier\n";
    print "   -----\n";
    -----\n";
#   print "   20140606094200.000000+000   3300   IBM:ExhaustAirTemp_U78C9.001.D123456
00001E00_10711";

    foreach my $line (@result)
    {
        if ($line =~ /-InstanceID=([\^]*)./)
        {
            $instance=$1;
        }
        elsif ($line =~ /-TimeStamp=(\S*)/)
        {
            $timestamp=$1;
        }
        elsif ($line =~ /-MetricValue=(\d*)./)
        {
            $value=$1;
            printf("   $timestamp   %4d   $instance\n", $value);
        }
    }
}

exit 0;
```



### Thermal Data Output

Thermal Trending Data for tul73fp: (1/100th degree C)

Timestamp	Temp	Instance Identifier
20140606135600.000000+000	3400	IBM:ExhaustAirTemp_U78C9.001.D123456 00001E00_11726
20140606135630.000000+000	3400	IBM:ExhaustAirTemp_U78C9.001.D123456 00001E00_11728
20140606135700.000000+000	3400	IBM:ExhaustAirTemp_U78C9.001.D123456 00001E00_11731
20140606135730.000000+000	3400	IBM:ExhaustAirTemp_U78C9.001.D123456 00001E00_11733
20140606135800.000000+000	3400	IBM:ExhaustAirTemp_U78C9.001.D123456 00001E00_11735
:	:	:
:	:	:
20140606145200.000000+000	3300	IBM:ExhaustAirTemp_U78C9.001.D123456 00001E00_11950
20140606145230.000000+000	3300	IBM:ExhaustAirTemp_U78C9.001.D123456 00001E00_11952
20140606145300.000000+000	3300	IBM:ExhaustAirTemp_U78C9.001.D123456 00001E00_11955
20140606145330.000000+000	3300	IBM:ExhaustAirTemp_U78C9.001.D123456 00001E00_11957
20140606145400.000000+000	3300	IBM:ExhaustAirTemp_U78C9.001.D123456 00001E00_11959
20140606135600.000000+000	2400	IBM:InletAirTemp_U78C9.001.D123456-D1 00006000_11727
20140606135630.000000+000	2400	IBM:InletAirTemp_U78C9.001.D123456-D1 00006000_11729
20140606135700.000000+000	2400	IBM:InletAirTemp_U78C9.001.D123456-D1 00006000_11730
20140606135730.000000+000	2400	IBM:InletAirTemp_U78C9.001.D123456-D1 00006000_11732
20140606135800.000000+000	2400	IBM:InletAirTemp_U78C9.001.D123456-D1 00006000_11734
:	:	:
:	:	:
20140606145200.000000+000	2300	IBM:InletAirTemp_U78C9.001.D123456-D1 00006000_11951
20140606145230.000000+000	2300	IBM:InletAirTemp_U78C9.001.D123456-D1 00006000_11953
20140606145300.000000+000	2300	IBM:InletAirTemp_U78C9.001.D123456-D1 00006000_11954
20140606145330.000000+000	2300	IBM:InletAirTemp_U78C9.001.D123456-D1 00006000_11956
20140606145400.000000+000	2300	IBM:InletAirTemp_U78C9.001.D123456-D1 00006000_11958

## Script to Query and Set a Power Management Mode

```
#!/usr/bin/perl
my $system="";
my $uid="HMC";
my $pwd="xxxxxx";
my $verbose=0;

if ($#ARGV >= 0)
{
    $system="$ARGV[$i]";
}
else
{
    print "ERROR: Must supply system name or IP address\n";
    exit 1;
}

my $freq="";
my $currentMode = 0;
my $ffoFreq=0;
# Query Frequency Points and Power Mode
if ($verbose)
{
    print "==> wbemcli -nl -noverify ei \"https://${uid}:${pwd}@${system}:5989/root/ibmsd:CIM_PowerUtilizationManagementService\" | grep -e \"-FixedFrequency\" -e \"-PowerUtilizationMode\"\n";
}
$result=`wbemcli -nl -noverify ei "https://${uid}:${pwd}@${system}:5989/root/ibmsd:CIM_PowerUtilizationManagementService" | grep -e "FixedFrequency" -e "PowerUtilizationMode"`;
if ( $result eq "" )
{
    print "ERROR: Failed to find Frequency Points\n";
}
else
{
    print "Frequency Points for $system:\n";
    $result =~ /-FixedFrequencyPoints=(\d*),(\d*),(\d*),(\d*)/;
    my @freqType = ($1, $2, $3, $4);
    $result =~ /-FixedFrequencyPointValues=(\d*),(\d*),(\d*),(\d*)/;
    my @freqValue = ($1, $2, $3, $4);
    $result =~ /-FixedFrequencyOverrideFreq=(\d*)/;
    $ffoFreq = $1;
    for ($i = 0; $i <= $#freqValue; $i++)
    {
        my $freqName="Unknown";
        if ($freqType[$i] == 2) { $freqName = "Nominal"; }
        elsif ($freqType[$i] == 3) { $freqName = "Maximum"; }
        elsif ($freqType[$i] == 5) { $freqName = "Minimum"; }
        if ($freqType[$i] != 4)
        {
            printf(" %8s: %5d MHz\n", $freqName, $freqValue[$i]);
        }
    }
    printf(" Override: %5d MHz\n\n", $ffoFreq);

    $result =~ /-PowerUtilizationMode=(\d*)/;
    $currentMode = $1;
    printf(" Power Mode: %d (%s)\n", $currentMode, modeString($currentMode));
}
print "\n";

# Set Power Mode
```

```

print " Modes: 2 = Nominal, 4 = Static Power Save, 3 = Dynamic Power Save (Favor
Power),\n";
print "          32768 = Dynamic Power Save (Favor Performance), 32769 = Fixed
Frequency Override\n\n";
print "Enter new Power Mode for $system [$currentMode]? ";
$answer = <STDIN>;
chomp($answer);
if (" $answer" ne "")
{
    my $newMode = $answer;

    if ($newMode == 32769)
    {
        print "Enter new Fixed Frequency in MHz for $system [$ffoFreq]? ";
        $answer = <STDIN>;
        chomp($answer);
        if (" $answer" ne "")
        {
            $ffoFreq = $answer;
        }
        print "Set Power Mode $newMode @ $ffoFreq MHz (".modeString($newMode).")\n";
    }
    else
    {
        print "Set Power Mode $newMode (".modeString($newMode).")\n";
    }

    # Service parm that is needed for setting the Power Mode
    my $pwrUtilMgtService="";
    if ($verbose)
    {
        print "=> wbemcli -nl -noverify ein https://${uid}:${pwd}\@$
{system}:5989/root/ibmsd:CIM_PowerUtilizationManagementService\n";
    }
    @result=`wbemcli -nl -noverify ein https://${uid}:${pwd}\@$
{system}:5989/root/ibmsd:CIM_PowerUtilizationManagementService`;
    if ( $#result < 0 )
    {
        print "ERROR: Failed to find pwrUtilMgtService\n";
    }
    else
    {
        $pwrUtilMgtService=$result[0];
        chomp($pwrUtilMgtService);
        if ($verbose) { print "pwrUtilMgtService: $pwrUtilMgtService\n"; }
    }

    if ($newMode == 32769)
    {
        # Set Fixed Frequency Override Frequency
        if ($verbose)
        {
            print "=> wbemcli -nl -noverify mi 'https://${uid}:${pwd}\@$
{pwrUtilMgtService}' FixedFrequencyOverrideFreq=${ffoFreq}\n";
        }
        $result=`wbemcli -nl -noverify mi 'https://${uid}:${pwd}\@$
{pwrUtilMgtService}' FixedFrequencyOverrideFreq=${ffoFreq}`;
        if ( $#result < 0 )
        {
            print "ERROR: Failed to set Fixed Frequency Override to ${ffoFreq}\n";
        }
        else
        {

```

```

        print "Set Fixed Frequency Override to ${ffoFreq} completed\n";
    }
}

# Set Power Mode
if ($verbose)
{
    print "==> wbemcli -nl -noverify mi 'https://${uid}:${pwd}\@$
{pwrUtilMgtService}' PowerUtilizationMode=${newMode}\n";
}
$result=`wbemcli -nl -noverify mi 'https://${uid}:${pwd}\@$
PowerUtilizationMode=${newMode}`;
if ( $#result < 0 )
{
    print "ERROR: Failed to Set Power Mode to $newMode\n";
}
else
{
    print "Set Power Mode to ${newMode} completed\n";
}

print "Re-querer Power Mode for $system:\n";
if ($verbose)
{
    print "==> wbemcli -nl -noverify ei \"https://${uid}:${pwd}\@$
{system}:5989/root/ibmsd:CIM_PowerUtilizationManagementService\" | grep \"-
PowerUtilizationMode\" \n";
}
$result=`wbemcli -nl -noverify ei "https://${uid}:${pwd}\@$
{system}:5989/root/ibmsd:CIM_PowerUtilizationManagementService" | grep
"PowerUtilizationMode"`;
if ($result =~ /-PowerUtilizationMode=(\d*)/)
{
    $currentMode = $1;
    printf("Current Power Mode for $system: %d (%s)\n", $currentMode,
modeString($currentMode));
}
else
{
    print "ERROR: Failed to find Frequency Points\n";
}
}

sub modeString
{
    my $string = "Unknown";
    ($mode) = @_ ;
    if ($mode == 2) { $string = "Nominal"; }
    elsif ($mode == 3) { $string = "Dynamic Power Saver (Favor Power)"; }
    elsif ($mode == 4) { $string = "Static Power Saver"; }
    elsif ($mode == 32768) { $string = "Dynamic Power Saver (Favor Performance)"; }
    elsif ($mode == 32769) { $string = "Fixed Frequency Override"; }
    return $string;
}

exit 0;

```

## ***Example Setting Fixed Frequency Override Mode***

Frequency Points for tul73fp:

Nominal: 3425 MHz  
Maximum: 3691 MHz  
Minimum: 2029 MHz  
Override: 3425 MHz

Power Mode: 2 (Nominal)

Modes: 2 = Nominal, 4 = Static Power Save, 3 = Dynamic Power Save (Favor Power),  
32768 = Dynamic Power Save (Favor Performance), 32769 = Fixed Frequency  
Override

Enter new Power Mode for tul73fp [2]? 32769  
Enter new Fixed Frequency in MHz for tul73fp [3425]? 3691  
Set Power Mode 32769 @ 3691 MHz (Fixed Frequency Override)  
Set Fixed Frequency Override to 3691 completed  
Set Power Mode to 32769 completed  
Re-querer Power Mode for tul73fp:  
Current Power Mode for tul73fp: 32769 (Fixed Frequency Override)

## Script to Query and Set a Power Cap

```
#!/usr/bin/perl
my $system="";
my $uid="HMC";
my $pwd="xxxxxxx";
my $verbose=0;

if ($#ARGV >= 0)
{
    $system="$ARGV[$i]";
}
else
{
    print "ERROR: Must supply system name or IP address\n";
    exit 1;
}

# Object parms that are needed for some CIM queries
my $sysObject="";
if ($verbose)
{
    print "==> wbemcli -nl -noverify ein https://${uid}:${pwd}\@$
{system}:5989/root/ibmsd:FipS_CEC\n";
}
@result=`wbemcli -nl -noverify ein https://${uid}:${pwd}\@$
{system}:5989/root/ibmsd:FipS_CEC`;
if ( $#result < 0 )
{
    print "ERROR: Failed to find sysObject\n";
    exit 1;
}
else
{
    $sysObject=$result[0];
    chomp($sysObject);
    if ($verbose) { print "sysObject: $sysObject\n"; }
}

my $pwrAllocObject="";
if ($verbose)
{
    print "==> wbemcli -nl -noverify ai \"https://${uid}:${pwd}\@$
{sysObject}\" -ac
CIM_SettingsDefineState -arc CIM_PowerAllocationSettingData\n";
}
@result=`wbemcli -nl -noverify ai 'https://${uid}:${pwd}\@$
{sysObject}' -ac
CIM_SettingsDefineState -arc CIM_PowerAllocationSettingData`;
if ( $#result < 0 )
{
    print "ERROR: Failed to find pwrAllocObject\n";
    exit 1;
}
else
{
    $pwrAllocObject=$result[0];
    chomp($pwrAllocObject);
    if ($verbose) { print "pwrAllocObject: $pwrAllocObject\n"; }
}

my $pwrAllocCapabObject="";
if ($verbose)
{
```

```

    print "==> wbemcli -nl -noverify ai 'https://${uid}:${pwd}\@${pwrAllocObject}'
-ac CIM_ElementCapabilities -arc CIM_AllocationCapabilities\n";
}
@result=`wbemcli -nl -noverify ai 'https://${uid}:${pwd}\@${pwrAllocObject}' -ac
CIM_ElementCapabilities -arc CIM_AllocationCapabilities`;
if ( $#result < 0 )
{
    print "ERROR: Failed to find pwrAllocCapabObject\n";
}
else
{
    $pwrAllocCapabObject=$result[0];
    chomp($pwrAllocCapabObject);
    if ($verbose) { print "pwrAllocCapabObject: $pwrAllocCapabObject\n"; }
}

# Power Cap Limits
my $pcapLimits="";
if ($verbose)
{
    print "==> wbemcli -nl -noverify ai \"https://${uid}:${pwd}\@${
{pwrAllocCapabObject}\" -ac CIM_SettingsDefineCapabilities -arc
CIM_PowerAllocationSettingData | grep -e \"Limit=\"
-e \"PowerAllocationSettingPurpose=\"\n";
}
@result=`wbemcli -nl -noverify ai "https://${uid}:${pwd}\@${pwrAllocCapabObject}" -ac
CIM_SettingsDefineCapabilities -arc CIM_PowerAllocationSettingData | grep -e "Limit="
-e "PowerAllocationSettingPurpose="`;
if ( $#result < 0 )
{
    print "ERROR: Failed to find Power Cap Limits\n";
}
else
{
    print "Power Cap Limits for $system:\n";
    foreach my $line (@result)
    {
        if ($line =~ /-PowerAllocationSettingPurpose=(\d*)/)
        {
            $purpose=$1;
            if ($purpose == 3)
            {
                printf("    Hard Minimum Power Cap: %5d W\n", $limit);
            }
            elsif ($purpose == 4)
            {
                printf("        Maximum Power Cap: %5d W\n", $limit);
            }
            elsif ($purpose == 1)
            {
                printf("    Soft Minimum Power Cap: %5d W\n", $limit);
            }
        }
        elsif ($line =~ /-Limit=(\d*)/)
        {
            $limit=$1;
        }
    }
    print "$result";
}

# Current Power Cap

```

```

my $pcapCurrent="";
if ($verbose)
{
    print "==> wbemcli -nl -noverify gi \"https://${uid}:${pwd}\@${pwrAllocObject}\"
    | grep -e \"Limit=\" -e \"PowerAllocationSettingPurpose=\"\n";
}
$result=`wbemcli -nl -noverify gi "https://${uid}:${pwd}\@${pwrAllocObject}" | grep
-e "Limit=" -e "PowerAllocationSettingPurpose="`;
if ( $#result < 0 )
{
    print "ERROR: Failed to find Current Power Cap\n";
}
else
{
    $result =~ /-Limit=(\d*)/;
    $limit=$1;
    printf("          Current Power Cap: %5d W\n", $limit);
}
print "\n";

# Set Power Cap
print "Enter new Power Cap in Watts (0 to disable, or just press enter for no
change)? ";
$answer = <STDIN>;
chomp($answer);
if (" $answer" ne "")
{
    my $newcap = $answer;

    my $pwrUtilMgtService="";
    if ($verbose)
    {
        print "==> wbemcli -nl -noverify ein https://${uid}:${pwd}\@$
(system):5989/root/ibmsd:CIM_PowerUtilizationManagementService\n";
    }
    @result=`wbemcli -nl -noverify ein https://${uid}:${pwd}\@$
(system):5989/root/ibmsd:CIM_PowerUtilizationManagementService`;
    if ( $#result < 0 )
    {
        print "ERROR: Failed to find pwrUtilMgtService\n";
    }
    else
    {
        $pwrUtilMgtService=$result[0];
        chomp($pwrUtilMgtService);
        if ($verbose) { print "pwrUtilMgtService: $pwrUtilMgtService\n"; }
    }

    if ($verbose)
    {
        print "==> wbemcli -nl -noverify cm 'https://${uid}:${pwd}\@$
{pwrUtilMgtService}' ApplyPowerAllocationSettings.PowerAllocationSettings='<INSTANCE
CLASSNAME=\"FipS_PowerAllocationSettingData\"><PROPERTY NAME=\"Limit\"
TYPE=\"uint64\"><VALUE>${newcap}</VALUE></PROPERTY><PROPERTY NAME=\"InstanceID\"
TYPE=\"string\"><VALUE>IBM:PowerAllocationSettingData:6</VALUE></PROPERTY></INSTANCE>
'\n';
    }
    $result=`wbemcli -nl -noverify cm 'https://${uid}:${pwd}\@${pwrUtilMgtService}'
ApplyPowerAllocationSettings.PowerAllocationSettings='<INSTANCE
CLASSNAME="FipS_PowerAllocationSettingData"><PROPERTY NAME="Limit"
TYPE="uint64"><VALUE>'${newcap}'</VALUE></PROPERTY><PROPERTY NAME="InstanceID"
TYPE="string"><VALUE>IBM:PowerAllocationSettingData:6</VALUE></PROPERTY></INSTANCE>'`
;

```



```
if ( $#result < 0 )
{
    print "ERROR: Failed to Set Power Cap to $newcap\n";
}
else
{
    print "Set Power Cap ${newcap} W completed\n";
}
}
exit 0;
```

### ***Example Setting a Power Cap***

```
Power Cap Limits for tul73fp:  
  Hard Minimum Power Cap: 1260 W  
    Maximum Power Cap: 1507 W  
  Soft Minimum Power Cap:  322 W  
    Current Power Cap:    0 W
```

```
Enter new Power Cap in Watts (0 to disable, or just press enter for no change)?
```

```
1500
```

```
Set Power Cap 1500 W completed
```

# HMC

The HMC (Hardware Management Console) is a management console that controls managed systems, logical partitions, managed frames, other features provided through the managed objects, and the HMC itself. The HMC provides both graphical user interface (GUI) and command line interfaces. Users can use the user interfaces to configure or manage various features offered by the managed objects.

## Setting System Power Management Mode

The user can list which power management modes are supported using the `lspwrmgmt` command on the command line as illustrated below:

```
lspwrmgmt -m <managed system name> -r sys -F supported_power_saver_mode_types
```

Example output:

```
"static,dynamic_favor_power,dynamic_favor_perf,fixed_max_frequency"
```

A user can then enable one of the supported power management modes using the `chpwrmgmt` command on the command line as illustrated below enabling Fixed Maximum Frequency mode:

```
chpwrmgmt -m <managed system name> -r sys -o enable -t fixed_max_frequency
```

To query the current power management mode use `lspwrmgmt` command on the command line as illustrated below:

```
lspwrmgmt -m <managed system name> -r sys
```

Example output:

```
curr_power_saver_mode=Enabled,curr_power_saver_mode_type=fixed_max_frequency,desired_power_saver_mode=Enabled,desired_power_saver_mode_type=fixed_max_frequency,"supported_power_saver_mode_types=static,dynamic_favor_power,dynamic_favor_perf,fixed_max_frequency"
```

NOTE: The new mode may not take effect immediately. Normally, if the operation is performed before the system is powered on, the desired mode won't take effect until the system is up and running. If the mode is in transition, any changes will be blocked.

To Disable power management mode:

```
chpwrmgmt -m <managed system name> -r sys -o disable
```

With the HMC GUI, users can reach this task by selecting the managed system -> Operations -> Power Management.

## ***DCMI***

On some systems and FW level the Data Center Manageability Interface (DCMI) may be used to access additional data such as processor temperatures, baseboard temperatures, power and the ability to set a power limit. Refer to the DCMI specification for full command list. See [Appendix I: System Requirements](#) for system and required FW that supports DCMI.

### NOTES:

- Power readings and power limits are input (AC) power.
- Set Power Limit exception actions cannot be changed on POWER8 systems. The Set Power Limit command will be rejected with 0x8A OEM completion code if the exception action is anything besides 0x11 (log event only)
- Set Power Limit correction time limit cannot be changed from 1000ms. The Set Power Limit command will be rejected with 0x85 completion code for correction time out of range if any other value is sent
- Set Power Limit sampling period cannot be changed from 1s. The Set Power Limit command will be rejected with 0x89 completion code for sampling period out of range if any other value is sent
- If a Set Power Limit command is sent while a power limit is already active the new power limit will take effect immediately without another activate power limit command being sent.

# EnergyScale Operating System Support

Operating system support for EnergyScale has been included since the introduction of the POWER6 processor. Certain capabilities, however, require more recent editions of each operating system. Please refer to the table below for a summary of available features by operating system level. As a best practice, the Fix Level Recommendation Tool (<http://www-304.ibm.com/support/customer/flare/>) and/or System Software Maps (<http://www-304.ibm.com/support/docview.wss?uid=ssmlmaps>) should be used to determine the latest recommended code level when planning system installs and upgrades.

<b>Feature</b>	<b>AIX</b>	<b>IBM i</b>	<b>Linux</b>
<b>Scaled Processor Time API(s)</b> Provides programmatic access to POWER8 SPURR <sup>2</sup> registers	5.3 TL9 6.1 TL2 7.1	6.1.1	SLES11 RHEL 5.5
<b>Performance Tool Support for Scaled Processor Time</b> PURR / SPURR support in OS Performance Tools	5.3 TL11 6.1 TL4 7.1	6.1.1	n/a
<b>Normalized Process and Job Accounting</b> Job accounting is SPURR-based	5.3 TL 9 6.1 TL2 7.1	7.1	n/a
<b>Processor Folding</b> Processor folding using processor idle states	5.3 TL 9 6.1 TL2 7.1	6.1.1	SLES11 SP1 RHEL6
<b>Processor Folding Control / Status</b> Processor folding may be enabled, disabled, or tuned from the OS	5.3 TL9 6.1 TL2 7.1	7.1	SLES11 SP1 RHEL6
<b>EnergyScale Configuration Status</b> EnergyScale Power Mgmt mode exposed through the OS	6.1 TL6 7.1	7.1	SLES11 SP1 RHEL6
<b>Query Idle Power Saver and Dynamic Power Saver Tunable Parameters</b> Query only, parameters are not settable from the OS.	7.1 TL4 SP1 7.2	n/a	n/a

<sup>2</sup> Refer to Appendix III: Processor Usage and Accounting for details about the SPURR register.

## Processor Folding in AIX

In AIX the processor folding policy can be configured from the command line via the `schedo` command. The `vpm_fold_policy` tunable is a 4-bit value where each bit indicates the configuration of a different setting. The following table shows the various settings that are controlled.

<i>Bit</i>	<i>Setting</i>
<b>0</b>	=1 processor folding is enabled when the partition is using shared processors
<b>1</b>	=1 processor folding is enabled when the partition is using dedicated processors
<b>2</b>	=1 disables the automatic setting of processor folding when the partition is in static power saver mode
<b>3</b>	=1 processor affinity will be ignored when making folding decisions

Table 1: `vpm_fold_policy` is a 4-bit value, in which each bit controls an aspect of folding.

The following command displays the current setting of `vpm_fold_policy`:

```
# schedo -L vpm_fold_policy
NAME                                CUR    DEF    BOOT    MIN    MAX    UNIT  TYPE
DEPENDENCIES
-----
vpm_fold_policy                    1      1      1       0     15     D
-----
```

To enable processor folding on a partition using dedicated partitions when the current value of `vpm_fold_policy` is set to 1, the following command would be issued to set the value to 3:

```
# schedo -o vpm_fold_policy=3
```

To disable processor folding, the value of `vpm_fold_policy` can be set to the value 4 using the following command:

```
# schedo -o vpm_fold_policy=4
```

By default, AIX will attempt to consider processor affinity (or topology) information when making processor folding decisions. This allows for the workload to remain spread across the processor nodes (e.g., chips depending on the system) and benefit from improved performance. Bit 3 of the `vpm_fold_policy` tunable allows this default behavior to be disabled. For example, if `vpm_fold_policy` is currently set to 6, indicating that processor folding is enabled when the partition is using dedicated partitions and that the partition will automatically enable processor folding when in static power saver mode, the following command would change the setting to indicate that the operating system should no longer consider processor affinity when making folding decisions:

```
# schedo -o vpm_fold_policy=14
```

For more information see the help information via the `-h` option of the `schedo` command:

```
# schedo -h vpm_fold_policy
```

## Processor Folding in IBM i

In an IBM i partition, processor folding is configured and controlled by the operating system by default. On POWER6, POWER7 and POWER8 servers, the operating system enables processor folding by default when Static Power Saver mode is enabled. On POWER8 servers, the operating system also enables processor folding by default in shared processor LPARs. Beginning in IBM i 7.1, operating system control of processor folding may be overridden via the **QWCCTLSW** limited availability API, which provides a key-based control language programming interface to a limited set of IBM i tunable parameters. Processor folding control is accessed via QWCCTLSW key 1060. The following sequence of calls cycles through the various options. Changes to key 1060 take effect immediately but do not persist across IPLs.

Get current status:

```
> CALL QWCCTLSW PARM('1060' '1')
KEY 1060 IS *SYSCTL.
KEY 1060 IS SUPPORTED ON THE CURRENT IPL.
KEY 1060 IS CURRENTLY ENABLED.
```

Explicitly disable processor folding:

```
> CALL QWCCTLSW PARM('1060' '3' )
KEY 1060 SET TO *OFF.
```

Explicitly enable processor folding:

```
> CALL QWCCTLSW PARM('1060' '2' 1)
KEY 1060 SET TO *ON.
```

Re-establish system control of processor folding:

```
> CALL QWCCTLSW PARM('1060' '2' 2)
KEY 1060 SET TO *SYSCTL.
```

## Processor Folding in Linux

It is essential to install a daemon package based on the host OS to enable utilization-based processor folding for Static Power Saver and Idle Power Saver modes:

```
pseries-energy-1.4.0-1.el7.ppc64.rpm  
pseries-energy-1.4.0-1.el6.ppc64.rpm  
pseries-energy-1.4.0-1.sles11.ppc64.rpm
```

Version 5.4 has the necessary user space tools required to enable CPU Folding.<sup>3</sup>

Once this package is installed, the `energyd` daemon will monitor the system power mode and activate processor folding when system power mode is set to "Static Power Saver" and deactivate processor folding in all other modes. The utilization-based CPU folding daemon will deactivate unused cores and transition them to low power idle states until the CPU utilization increases and those cores are activated to run a workload.

Utilization-based processor folding can be manually disabled using the following commands:

```
/etc/init.d/energyd stop #Stop daemon now, activate all cores  
chkconfig energyd off    #Do not restart daemon on startup
```

-or-

```
rpm -e pseries-energy    #un-install the package completely
```

Alternatively, CPU cores can be folded or set to low power idle state in any power mode manually using the following command line:

```
echo 0 > /sys/devices/system/cpu/cpuN/online #Where N is the  
logical CPU number
```

Please note that all active hardware threads of a core needs to be taken off-line using the above command in order to move the core to a low power idle state.

The cores can be activated again with the following command:

```
echo 1 > /sys/devices/system/cpu/cpuN/online #Where N is the  
logical CPU number
```

---

<sup>3</sup> Refer to <http://www14.software.ibm.com/webapp/set2/sas/f/lopdiags/installtools/home.html> for more details.



## Query Idle Power Saver and Dynamic Power Saver Tunable Parameters in AIX

With a minimum level of AIX and FW840 system firmware the Idle Power Saver and Dynamic Power Saver Tunable parameters may be queried in AIX. The parameters are not settable from the OS. See [EnergyScale Operating System Support](#) for required AIX level. Example query from AIX:

```
# Iparstat -P
Dynamic Power Saver Tunables
-----
Util threshold for increasing frequency : 99.0%
Util threshold for decreasing frequency : 97.0%
Number of samples for computing util stats : 16 samples
Step size for going up in frequency    : 0.8%
Step size for going down in frequency   : 0.8%
Delta %% for determining active cores   : 18%
Util threshold to determine active cores...: 98.0%
Enable/Disable freq delta between cores : Disabled
Maximum frequency delta between cores   : 10%
Idle Power Saver
-----
Idle Power Saver Enable                 : Enabled
Delay time to enter Idle Power Saver    : 240 seconds
Util threshold to enter Idle Power Saver : 8%
Delay time to exit Idle Power Saver     : 10 seconds
Util threshold to Exit Idle Power Saver : 12%
AIX Power Saving Action                  : -
```

# Appendix I: System Requirements

Due to differences in each release, this appendix details the systems and EnergyScale features supported by release including the actual frequency limits for various power management modes, as well as the estimated processor power saved in Power Saver mode.

# Release Level FW810

## Feature Support

Refer to the EnergyScale Features chapter earlier in this document for a definition of each feature.

	Any system in OPAL Hypervisor Mode	Power S814* (8286-41A)	Power S822* (8284-22A)	Power S824* (8286-42A)	Power S812L** (8247-21L)	Power S822L** (8247-22L)
Power Trending	Y	Y	Y	Y	Y	Y
Thermal Reporting	Y	Y	Y	Y	Y	Y
Static Power Saver	N	Y	Y	Y	Y	Y
Fixed Frequency Override	N	Y	Y	Y	Y	Y
Dynamic Power Saver	N	Y	Y	Y	Y	Y
Dynamic Power Saver Tunable Parameters	N	Y	Y	Y	Y	Y
Idle Power Saver	N	Y – Enabled by Default	Y – Enabled by Default	Y – Enabled by Default	Y	Y
Power Capping	N	Y	Y	Y	Y	Y
Power Supply Idle Detection	N	N	N	N	N	N

### Support Notes

\*Capabilities shown are for the system default PowerVM Hypervisor Mode

\*\*Capabilities shown are for PowerVM. The system default is OPAL Hypervisor Mode. Refer to “Any system in OPAL Hypervisor Mode” column for default capabilities

## Frequency

Depending upon the Power savings setting selected, the maximum and minimum frequency may change. For a definition of each setting, please refer to the EnergyScale Features chapter earlier in this document. Note that processor power savings values below are statistical representations, and may vary from system-to-system. All frequencies are relative to nominal frequency for a particular system.

### Support Notes

<sup>1</sup> Note that CPU frequencies in excess of 100% are *not* guaranteed. The actual maximum frequency may vary based on environmental conditions, system configuration, firmware version, component tolerances, and workload.

	<i>Static Power Saver Frequency (fixed)</i>	<i>Static Power Saver % Processor Power Saved</i>	<i>Maximum Frequency<sup>1</sup></i>	<i>Minimum Frequency</i>
<b>IBM Power S814 (8286-41A) @3.026GHz</b>	70%	30-40%	110%	67%
<b>IBM Power S814 (8286-41A) @3.724GHz</b>	70%	30-40%	108%	54%
<b>IBM Power S822 (8284-22A) @3.425GHz</b>	70%	30-40%	108%	59%
<b>IBM Power S822 (8284-22A) @3.891GHz</b>	70%	30-40%	109%	52%
<b>IBM Power S822 (8284-22A) @4.157GHz</b>	70%	30-40%	104%	50%
<b>IBM Power S824 (8286-42A) @3.525GHz</b>	70%	30-40%	111%	58%
<b>IBM Power S824 (8286-42A) @3.891GHz</b>	70%	30-40%	109%	52%
<b>IBM Power S824 (8286-42A) @4.157GHz</b>	70%	30-40%	104%	50%
<b>IBM Power S812L (8247-21L) @3.026GHz</b>	70%	30-40%	110%	67%
<b>IBM Power S812L (8247-21L) @3.425GHz</b>	70%	30-40%	108%	59%
<b>IBM Power S822L (8247-22L) @3.026GHz</b>	70%	30-40%	110%	67%

	<i>Static Power Saver Frequency (fixed)</i>	<i>Static Power Saver % Processor Power Saved</i>	<i>Maximum Frequency<sup>1</sup></i>	<i>Minimum Frequency</i>
<b>IBM Power S822L (8247-22L) @3.425GHz</b>	70%	30-40%	108%	59%
<b>IBM Power S822L (8247-22L) @4.157GHz</b>	70%	30-40%	104%	50%

# Release Level FW810.3

## Feature Support

FW810.3 supports new Fixed Maximum Frequency Mode and new system 8247-42L.

Refer to the EnergyScale Features chapter earlier in this document for a definition of each feature.

	Any system in OPAL Hypervisor Mode	Power S814* (8286-41A)	Power S822* (8284-22A)	Power S824* (8286-42A)	Power S812L** (8247-21L)	Power S822L** (8247-22L)	Power S824L** (8247-42L)
Power Trending	Y	Y	Y	Y	Y	Y	Y
Thermal Reporting	Y	Y	Y	Y	Y	Y	Y
Fixed Maximum Frequency Mode	N	Y	Y	Y	Y	Y	Y
Fixed Frequency Override	N	Y	Y	Y	Y	Y	Y
Dynamic Power Saver	N	Y	Y	Y	Y	Y	Y
Dynamic Power Saver Tunable Parameters	N	Y	Y	Y	Y	Y	Y
Idle Power Saver	N	Y – Enabled by Default	Y – Enabled by Default	Y – Enabled by Default	Y	Y	Y
Power Capping	N	Y	Y	Y	Y	Y	Y
Power Supply Idle Detection	N	N	N	N	N	N	N

### Support Notes

\*Capabilities shown are for the system default PowerVM Hypervisor Mode

\*\*Capabilities shown are for PowerVM. The system default is OPAL Hypervisor Mode. Refer to “Any system in OPAL Hypervisor Mode” column for default capabilities

## Frequency

Depending upon the Power savings setting selected, the maximum and minimum frequency may change. For a definition of each setting, please refer to the EnergyScale Features chapter earlier in this document. Note that processor power savings values below are statistical representations, and may vary from system-to-system. All frequencies are relative to nominal frequency for a particular system.

	<i>Static Power Saver Frequency (fixed)</i>	<i>Static Power Saver % Processor Power Saved</i>	<i>Maximum Frequency<sup>1</sup></i>	<i>Minimum Frequency</i>
<b>IBM Power S814 (8286-41A) @3.026GHz</b>	70%	30-40%	110%	67%
<b>IBM Power S814 (8286-41A) @3.724GHz</b>	70%	30-40%	108%	54%
<b>IBM Power S822 (8284-22A) @3.425GHz</b>	70%	30-40%	108%	59%
<b>IBM Power S822 (8284-22A) @3.891GHz</b>	70%	30-40%	109%	52%
<b>IBM Power S822 (8284-22A) @4.157GHz</b>	70%	30-40%	104%	50%
<b>IBM Power S824 (8286-42A) @3.525GHz</b>	70%	30-40%	111%	58%
<b>IBM Power S824 (8286-42A) @3.891GHz</b>	70%	30-40%	109%	52%
<b>IBM Power S824 (8286-42A) @4.157GHz</b>	70%	30-40%	104%	50%
<b>IBM Power S812L (8247-21L) @3.026GHz</b>	70%	30-40%	110%	67%
<b>IBM Power S812L (8247-21L) @3.425GHz</b>	70%	30-40%	108%	59%
<b>IBM Power S822L (8247-22L) @3.026GHz</b>	70%	30-40%	110%	67%
<b>IBM Power S822L (8247-22L) @3.425GHz</b>	70%	30-40%	108%	59%
<b>IBM Power S822L (8247-22L) @4.157GHz</b>	70%	30-40%	104%	50%

	<i>Static Power Saver Frequency (fixed)</i>	<i>Static Power Saver % Processor Power Saved</i>	<i>Maximum Frequency<sup>1</sup></i>	<i>Minimum Frequency</i>
<b>IBM Power S824L (8247-42L) @3.026GHz</b>	70%	30-40%	110%	67%
<b>IBM Power S824L (8247-42L) @3.425GHz</b>	70%	30-40%	108%	59%
<b>IBM Power S824L (8247-42L) @3.525GHz</b>	70%	30-40%	111%	58%
<b>IBM Power S824L (8247-42L) @4.157GHz</b>	70%	30-40%	104%	50%

**Support Notes**

<sup>1</sup> Note that CPU frequencies in excess of 100% are *not* guaranteed. The actual maximum frequency may vary based on environmental conditions, system configuration, firmware version, component tolerances, and workload.



# Release Level FW820

## Feature Support

Refer to the EnergyScale Features chapter earlier in this document for a definition of each feature.

	Power E870 (9119-MME)
	Power E880 (9119-MHE)
Power Trending	N
Thermal Reporting	N
Static Power Saver	Y
Fixed Frequency Override	Y
Dynamic Power Saver – Favor Performance	Y*
Dynamic Power Saver – Favor Power	N
Dynamic Power Saver Tunable Parameters	N*
Idle Power Saver	N
Power Capping	N
Power Supply Idle Detection	N

### Support Notes

\* No support for tunable parameters. Dynamic Power Saver Favor Performance will use the settings for a fixed maximum frequency by default and cannot be changed. See [Example Tunable Parameters for a Fixed Maximum Frequency](#).

## Frequency

Depending upon the Power savings setting selected, the maximum and minimum frequency may change. For a definition of each setting, please refer to the EnergyScale Features chapter earlier in this document. Note that processor power savings values below are statistical representations, and may vary from system-to-system. All frequencies are relative to nominal frequency for a particular system.

	<i>Static Power Saver Frequency (fixed)</i>	<i>Static Power Saver % Processor Power Saved</i>	<i>Maximum Frequency<sup>1</sup></i>	<i>Minimum Frequency</i>
<b>IBM Power E870 (9119-MME) @4.02GHz</b>	76%	25-30%	104%	76%
<b>IBM Power E870 (9119-MME) @4.19GHz</b>	76%	25-30%	106%	76%
<b>IBM Power E880 (9119-MHE) @4.35GHz</b>	76%	25-30%	104%	76%

### Support Notes

<sup>1</sup> Note that CPU frequencies in excess of 100% are *not* guaranteed. The actual maximum frequency may vary based on environmental conditions, system configuration, firmware version, component tolerances, and workload.

# Release Level FW830

## Feature Support

Refer to the EnergyScale Features chapter earlier in this document for a definition of each feature.

	Power S814 (8286-41A)	Power S822 (8284-22A)	Power S824 (8286-42A)	Power E850 (8408-E8E)	Power E870 (9119-MME)  Power E880 (9119-MHE)
Power Trending	Y	Y	Y	Y	N
Thermal Reporting	Y	Y	Y	Y	N
Fixed Maximum Frequency Mode	Y	Y	Y	Y	Y
Static Power Saver	Y	Y	Y	Y	Y
Fixed Frequency Override	Y	Y	Y	Y	Y
Dynamic Power Saver – Favor Performance	Y	Y	Y	Y	Y*
Dynamic Power Saver – Favor Power	Y	Y	Y	Y	N
Dynamic Power Saver Tunable Parameters	Y	Y	Y	Y	N*
Idle Power Saver	Y – Enabled by Default	Y – Enabled by Default	Y – Enabled by Default	Y – Enabled by Default	N
Power Capping	Y	Y	Y	Y	N
Power Supply Idle Detection	N	N	N	N	N

### Support Notes

\*No support for tunable parameters. Dynamic Power Saver Favor Performance acts the same as Fixed Maximum Frequency mode.

## Frequency

Depending upon the Power savings setting selected, the maximum and minimum frequency may change. For a definition of each setting, please refer to the EnergyScale Features chapter earlier in this document. Note that processor power savings values below are statistical representations, and may vary from system-to-system. All frequencies are relative to nominal frequency for a particular system.

	<i>Static Power Saver Frequency (fixed)</i>	<i>Static Power Saver % Processor Power Saved</i>	<i>Maximum Frequency<sup>1</sup></i>	<i>Minimum Frequency</i>
<b>IBM Power E870 (9119-MME) @4.02GHz</b>	76%	25-30%	104%	76%
<b>IBM Power E870 (9119-MME) @4.19GHz</b>	76%	25-30%	106%	76%
<b>IBM Power E850 (8408-E8E) @3.026GHz</b>	70%	30-40%	100%	67%
<b>IBM Power E850 (8408-E8E) @3.359GHz</b>	70%	30-40%	100%	60%
<b>IBM Power E850 (8408-E8E) @3.724GHz</b>	70%	30-40%	100%	54%
<b>IBM Power E880 (9119-MHE) @4.02GHz</b>	76%	25-30%	106%	76%
<b>IBM Power E880 (9119-MHE) @4.35GHz</b>	76%	25-30%	104%	76%

### Support Notes

<sup>1</sup> Note that CPU frequencies in excess of 100% are *not* guaranteed. The actual maximum frequency may vary based on environmental conditions, system configuration, firmware version, component tolerances, and workload.

# Release Level FW840

## Feature Support

New for FW840 is DCMI support and the ability to [query Idle Power Saver and Dynamic Power Saver Tunable Parameters in AIX](#). Refer to the EnergyScale Features chapter earlier in this document for a definition of each feature.

	Any system in OPAL Hypervisor Mode	Power S814* (8286-41A)	Power S812L** (8247-21L)	Power E870 (9119-MME) Power E880 (9119-MHE)
		Power S822* (8284-22A)	Power S822L** (8247-22L)	
		Power S824* (8286-42A)	Power S824L** (8247-42L)	
Power Trending	Y	Y	Y	N
Thermal Reporting	Y	Y	Y	N
Fixed Maximum Frequency Mode	N	Y	Y	Y
Static Power Saver	N	Y	Y	Y
Fixed Frequency Override	N	Y	Y	Y
Dynamic Power Saver – Favor Performance	N	Y	Y	Y <sup>+</sup>
Dynamic Power Saver – Favor Power	N	Y	Y	N
Dynamic Power Saver Tunable Parameters	N	Y	Y	N <sup>+</sup>
Idle Power Saver	N	Y – Enabled by Default	Y	N
Power Capping	N	Y	Y	N
DCMI	Y <sup>#</sup>	Y	Y	N

### Support Notes

\*Capabilities shown are for the system default PowerVM Hypervisor Mode

\*\*Capabilities shown are for PowerVM. The system default is OPAL Hypervisor Mode. Refer to “Any system in OPAL Hypervisor Mode” column for default capabilities

<sup>+</sup>No support for tunable parameters. Dynamic Power Saver Favor Performance acts the same as Fixed Maximum Frequency mode.

<sup>#</sup>Set/Get Power Limit commands are not supported in OPAL Hypervisor Mode due to power capping not supported

## Frequency

Refer to the frequency table for the system in previous FW release levels.

# Appendix II: Estimating Total System Power Saved

The amount of power that a system saves with the EnergyScale features varies based on configuration and workload. The following is a guideline to estimating the amount of power that may be saved:

## **Step 1:**

Determine the power usage for a specific system configuration by using the system power calculator located at this url:

<http://www.ibm.com/systems/support/tools/estimator/energy>

## **Step 2:**

Determine the power used by the processors.

For example, by adding/removing processors using the tool above, it may be shown that a 4.2 GHz dual-core module consumes 268W.

## **Step 3:**

Determine the power saved from power saver mode.

Assuming this system supports a 30% frequency drop for Power Saver Mode, this will save 20-30% of processor power. This yields a savings between 50-80W per dual core module (DCM).

## **Step 4:**

Determine the power saved from processor nap.

Processor nap saves roughly 15% of the processor power or an additional 10% if the system is in power saver mode. Assuming that the system is in power saver mode and that the system is idle (all cores napping), the per-DCM power savings is between 30-40% (80 – 107W).

## **Step 5:**

Determine the power saved as a percentage of the total system power.

Assume we have a system with two DCMs that consumes 1032W while running a high stress workload. If that workload is run with power saver mode enabled, the power usage will drop to between 932W and 872W which is a 10-16% savings. If that system is idle and power saver mode is enabled, the power usage will drop to between 872W and 818W. This is a 16-21% savings of the total system power.

# Appendix III: Processor Usage and Accounting

For historical reasons, process accounting charges and processor utilization are usually formulated in terms of time. The processors used in early time-sharing computer systems were fixed-frequency and single-threaded; processing capacity per unit of time was relatively constant. Consequently, it was convenient to express process accounting charges in terms of processor time and processor utilization as the percentage of time that the processor was not idle over an interval of interest.

With the introduction of multi-threaded processors (i.e. processors capable of executing multiple programs simultaneously), it was no longer desirable to report processor utilization based on the percentage of time that the processor was not idle. Consider a processor that supports 2-way Simultaneous Multi-Threading (SMT). When both threads are idle, utilization should be 0%. When both threads are not idle, utilization should be 100%. When one thread is idle and the other is not idle, there are several options:

1. Treat the processor as idle (0% utilization). This is obviously wrong, as the process is consuming more than 50% of the processing capacity.
2. Continue to treat the processor as not idle and charge the process with 100% of the time (100% utilization). This is a little better than the first option, but it causes utilization to be over-reported since it does not recognize the capacity available in the idle thread. The process is also significantly over-charged compared to when it shares the processor with another process.
3. Treat processor threads as individual processors, charging the process with 100% of the time and reporting processor-thread based utilization (50% utilization). This causes utilization to be under-reported because SMT efficiencies are much less than 100%, that is, the not idle thread typically represents only 15%-30% additional capacity. The process is also significantly over-charged compared to when it shares the processor with another process.

To address this problem, the Processor Utilization Resource Register (PURR) was introduced on POWER5™ for the purpose of apportioning processor time among the processor's threads. The PURR is defined such that  $\sum (\Delta \text{PURR}) = \Delta \text{TIME}$  over any interval. PURR ticks are apportioned among the processor threads based on relative instruction rate, so it is an indication of the relative processor capacity used by each thread. By expressing processor utilization as the ratio of not idle PURR ticks to available time and by expressing process accounting in terms of PURR ticks, the historical definition and relationship between processor utilization and process accounting was maintained.

IBM POWER8™ processor-based systems with EnergyScale employ variable processor speed technology to dynamically optimize the speed and energy usage of the processor to the demand of the workload. Because the PURR ticks at a constant rate independent of processor speed, PURR-based processor utilization remains a useful and accurate metric, but PURR-based process accounting charges can vary depending on processor speed. To address this problem, the IBM POWER6 processor included a new per-thread processor timekeeping facility to normalize the relationship between processor time and processor speed. The new facility was named the Scaled Processor Utilization

Resource Register (SPURR), and it represented processor time at nominal (i.e. 100%) speed.<sup>4</sup> The SPURR was primarily intended to improve process accounting consistency, but it can also be used in conjunction with the PURR in processor speed and capacity calculations. For example, a SPURR to PURR percentage of 85% indicates that the processor operated at 85% nominal speed over a sample interval.

While the PURR and SPURR provide the information necessary to measure accurate processor utilization and to perform consistent process accounting in the variable processor speed environment, they do not address the ambiguity of CPU time in the historical context. Simply stated, whether a CPU time value in a legacy software interface should represent PURR-based CPU time or SPURR-based CPU time is open to some interpretation. There is no single best choice to handle all cases, and in fact, the issue has not been uniformly addressed by all operating systems or even among versions of the same system. There is general agreement that SPURR-based process accounting is preferable to PURR-based accounting, since the results are more consistent across EnergyScale modes and within modes that vary the processor speed dynamically. This is particularly true for POWER8, which can operate across a wider range of processor speeds than POWER6.

---

<sup>4</sup> When the POWER8 processor is operating at full speed, the PURR and SPURR tick in lockstep; when the POWER8 processor is operating at reduced speed, the SPURR ticks slower than the PURR; when the POWER8 processor is operating in excess of full speed, the SPURR ticks faster than the PURR.



# Appendix IV: Resources

## **AIX:**

<http://publib.boulder.ibm.com/infocenter/pseries/v6r1/index.jsp>

(Search for "spurr" to discover references regarding enablement in APIs and/or tools)

## **DCMI:**

<http://www.intel.com/content/www/us/en/data-center/dcmi/data-center-manageability-interface.html>

## **IBM i:**

[http://www-01.ibm.com/support/knowledgecenter/ssw\\_ibm\\_i/welcome](http://www-01.ibm.com/support/knowledgecenter/ssw_ibm_i/welcome)

(Search for "Energy management", "Processor folding", "Scaled processor time attribute", "Processor time", "Scaled processor time", "Processor utilized time", "Processor scaled utilized time", "Processor interrupt time", "Processor scaled interrupt time", "Processor stolen time", "Processor scaled stolen time", "Processor donated time", "Processor scaled donated time", "Processor idle time", "Processor scaled idle time")

## **EnergyScale:**

<http://www-03.ibm.com/systems/power/software/energy/about.html>

## **IBM EnergyScale for POWER7 Processor-Based Systems:**

<http://www-03.ibm.com/systems/power/hardware/whitepapers/energyscale7.html>

## **IBM EnergyScale for POWER6 Processor-Based Systems:**

<http://www-03.ibm.com/systems/power/hardware/whitepapers/energyscale.html>

## **Power Systems:**

<http://www.ibm.com/systems/power>

## **IBM Systems Workload Estimator:**

<http://www.ibm.com/systems/support/tools/estimator/>

## **IBM Systems Energy Estimator:**

<http://www.ibm.com/systems/support/tools/estimator/energy>

## **System Power Calculators:**

<http://publib.boulder.ibm.com/infocenter/systems/scope/hw/topic/iphdl/systemcalculators.htm>

The Power Architecture and Power.org wordmarks and the Power and Power.org logos and related marks are trademarks and service marks licensed by Power.org.

UNIX is a registered trademark of The Open Group in the United States, other countries or both.

Linux is a trademark of Linus Torvalds in the United States, other countries or both.

Java and all Java-based trademarks and logos are trademarks of Sun Microsystems, Inc. in the United States and/or other countries.

All performance information was determined in a controlled environment. Actual results may vary. Performance information is provided "AS IS" and no warranties or guarantees are expressed or implied by IBM. Buyers should consult other sources of information, including system benchmarks, to evaluate the performance of a system they are considering buying.

Photographs show engineering and design models. Changes may be incorporated in production models.

Copying or downloading the images contained in this document is expressly prohibited without the written consent of IBM.



© IBM Corporation 2014  
IBM Corporation  
Systems and Technology Group  
Route 100  
Somers, New York 10589

Produced in the United States of America  
June 2014  
All Rights Reserved

This document was developed for products and/or services offered in the United States. IBM may not offer the products, features, or services discussed in this document in other countries.

The information may be subject to change without notice. Consult your local IBM business contact for information on the products, features and services available in your area.

All statements regarding IBM future directions and intent are subject to change or withdrawal without notice and represent goals and objectives only.

IBM, the IBM logo, ibm.com, AIX, BladeCenter, EnergyScale, i5/OS, Power, POWER, POWER6, POWER7, POWER8, System i and System p are trademarks or registered trademarks of International Business Machines Corporation in the United States or other countries or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml)

Other company, product, and service names may be trademarks or service marks of others.

IBM hardware products are manufactured from new parts, or new and used parts. In some cases, the hardware product may not be new and may have been previously installed. Regardless, our warranty terms apply.

This equipment is subject to FCC rules. It will comply with the appropriate FCC rules before final delivery to the buyer.

Information concerning non-IBM products was obtained from the suppliers of these products or other public sources. Questions on the capabilities of the non-IBM products should be addressed with those suppliers.

When referring to storage capacity, 1 TB equals total GB divided by 1000; accessible capacity may be less.

The IBM home page on the Internet can be found at: <http://www.ibm.com>.

The IBM Power Systems home page on the Internet can be found at: <http://www.ibm.com/systems/power/>

POW03125-USEN-04