POWER9 EnergyScale Introduction

Note: This document was originally copied from https://www.ibm.com/developerworks/community/wikis/home?lang=en-us#!/wiki/Power%20Systems/page/POWER9%20EnergyScale%20Introduction

Author: C Mehta

This is the first of a two part blog discussing how POWER9 hardware and firmware can dynamically alter the server characteristics to improve performance or reduce the amount of energy usage. This blog will provide a basic overview of EnergyScale and the second blog will cover how to configure and manage energy management on a server.

Overview

The announcement of POWER9 scale-out models (IBM S9xx models) introduced new features for EnergyScale[tm], including new variable processor frequency modes that provide a significant performance boost beyond the static nominal frequency. The ability of Power Systems to change frequency isn't new. In fact, variable frequency modes for energy management and performance enhancement have existed for over 10 years (since POWER6). However, there are some changes with the POWER9 system capabilities, and some new explanation is required.

EnergyScale modes

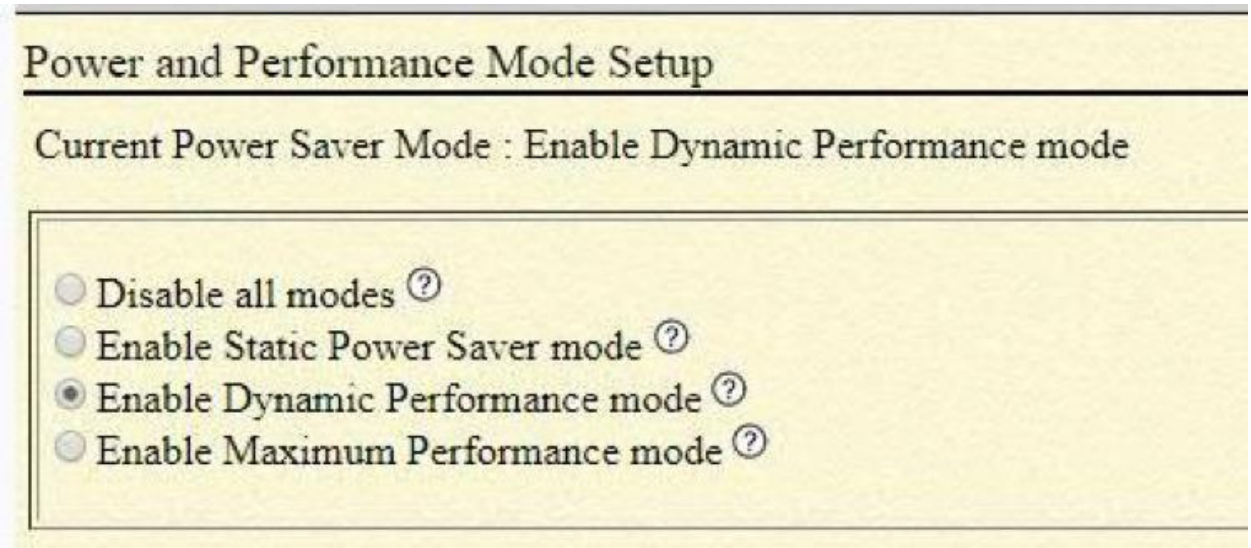
In POWER8, EnergyScale included the modes of Dynamic Power Saver – Favor Power, and Dynamic Power Saver – Favor Performance. These modes have been replaced in POWER9 with Dynamic Performance mode and Maximum Performance mode. It isn't just that the names have changed, the modes act a little differently in POWER9.

As we learn about the evolution of Energy Scale, there are three frequencies to keep in mind:

- Power saver or minimum frequency
- Nominal frequency
- Maximum frequency

The maximum frequency is approximately 20% better than nominal.

There are three modes that the system can be put in, as well as an all modes disabled option (which could be considered a mode). The menu shown below is a partial screen shot showing the options available on ASMI for POWER9.



The first option of *Disable all modes* is a mode in which the system will run at the fixed nominal frequency. The nominal frequency is the guaranteed frequency that the system will achieve when running within the specified environmental parameters for this model (meaning under the max ambient temperature and elevation). The option of *Disable all modes* was the default for all systems prior to POWER9. While providing a constant frequency, it does not maximize performance of the system.

The second option in the list is *Static Power Saver mode*. Static Power Saver mode is what you might use when you know that there is not a critical workload that will be running, and a slower clock speed will be OK. When in Static Power Saver mode, the system will run at the minimal frequency all the time, regardless of workload. Static Power Saver mode is intended to reduce electricity costs.

When in Dynamic Performance mode, the system will generally run above the nominal frequency and may even get to the maximum frequency if the workloads are light enough, or many cores are not being used. The determining factor for what frequency the CPU runs at in Dynamic Performance mode, is power. The system limits the socket power draw to a base wattage (this varies somewhat by chip and system). When in Dynamic Performance mode, the frequency will vary above the static nominal value depending on available power headroom, i.e. the power draw at the socket. With a heavy workload and all the cores being used, the system will run at least at the nominal frequency. If some cores are not being used, the system can run at much higher frequencies before the power limit is reached. If there are enough cores not being used, the maximum frequency may be reached in Dynamic Performance mode.

Note that the frequency is managed at the socket level, so different sockets may run at different frequencies. However, the mode setting is a system wide setting.

Also, Dynamic Performance mode is the only mode that lowers the processor frequency if the entire processor socket is idle for 100s of milliseconds, thus, providing both a performance boost and a power savings when possible – a great balance. Dynamic Performance mode is great for customers that want the best performance possible across the full range of environmental conditions, but have acoustic concerns.

Dynamic Performance mode is the default mode for the S914 system. All other POWER9 scale-out systems default to the Maximum Performance mode.

Maximum Performance mode takes advantage of lower active core counts and normal utilization workloads (just like Dynamic Performance mode). However, Maximum Performance mode will allow the system to reach the maximum frequency under more conditions, thus providing maximum performance (hence the name). The same constraint of power exists, but the system takes extra steps to extend the limit. When in Maximum Performance mode, the voltage regulators will allow the socket to draw more power than in the other modes. The increased power draw may be up to 100W higher.

In order to provide adequate cooling, Maximum Performance mode will increase fan speeds, which can increase the associated acoustics by up to 15 decibels, and increase the power that fans are using. If the datacenter ambient environment is less than 25C (77F), the frequency in Maximum Performance mode will consistently be in the upper range of the maximum frequency (roughly 10% to 20% better than nominal).

Note that the increased noise and wattage varies by system model, configuration, core count, and other factors.

Additionally, there is no power reduction in Maximum Performance mode due to idleness – the system is always keeping the frequency at the maximum value possible given the running workloads.

Maximum Performance mode is best for customers that have no acoustic concerns, are in favorable ambient conditions, and want top performance.

The figure below shows an example of the frequency range seen in each mode.

As mentioned, we will soon be publishing a second energy management blog that will describe how you can adjust and monitor the energy management settings on your server.

Contacting the PowerVM Team

Have questions for the PowerVM team or want to learn more? Follow our discussion group on LinkedIn IBM PowerVM