Managing DDR4 Memory Correctable Errors Ensuring Memory Sub-System Health

Technical Note by:

Vadhi SN
Amit Shah
Mark Farley

SUMMARY

This technical white paper details the naturally increasing rates of memory errors as memory speeds increase and DRAM storage cells shrink, and explains Dell EMC's use of proprietary predictive algorithms to proactively identify and notify the user when a DIMM (Dual inline Memory module) exhibits a correctable error signature that indicates an elevated probability of future failure. Although it is impossible to predict every failure, this memory warning is intended to allow the user to plan the DIMM replacement during the next scheduled maintenance event, before the potential of a cascading DRAM component failure can lead to an uncorrectable error. The algorithms guiding a decision to notify the user are a carefully determined balance of variables based on process technology, memory operating speed, and nature of channel errors detected.

Background

With each generation of server, demand for a larger memory footprint and higher memory performance grows. To meet this customer requirement, memory channel speeds and module densities have increased with every Dell EMC server generation.

DDR4 has replaced DDR3 as the de facto memory technology in today's servers. DDR4 technology brings an increase in memory channel speeds with an accompanying reduction in power consumption compared to the earlier DDR3 standard. Dell EMC first introduced DDR4 technology on the 13G servers in 2014—these servers support memory channel speeds at 2133 MT/s, an increase from the prior generation's DDR3 speed of 1866 MT/s. Dell EMC has since introduced higher DDR4 memory channel speeds with an update to the original 13G (2400 MT/s) and in the 14G (2666 MT/s) server product offerings from 2017.

Memory module densities have also kept pace with the increasing speeds. Module density increases are generally made possible due to increases in the density of the individual DRAM components that are assembled on the DIMM modules. Individual DRAM component densities have increased 16x since 2008, from 512 Mb (Megabits) chips to 8 Gb (Gigabits) chips today. DIMMs based on 16 Gb DRAM components are expected to arrive in the market in 2018. A key driver of the increase in DRAM component densities is the introduction by the semiconductor suppliers of newer technology nodes (aka "geometries"), which allow more data cells to be packed into the same physical space.

Transitioning to a smaller geometry node has other benefits too, such as reduced power consumption, smaller die-area, etc. Along with the geometries, the operating voltages have come down as the transition of memory standards has taken place. For instance, DDR3 memory technology operated at 1.5 V, while DDR4 operates at 1.2 V. Although a reduction in the operating voltage on the memory modules brings down the overall DIMM and system power, it also results in increased difficulty assuring an error-free transmission and receipt of data between the CPU and DIMM.

While the increase in the overall DIMM densities and memory channel speeds have ushered in an increase in memory performance generation over generation, the higher speeds, smaller geometries, and reduced operating voltage margins also have increased the likelihood of data transmission errors in the memory controller, memory channel (the path between the CPU and DIMM), and the DRAM modules. Memory RAS (Reliability, Availability, and Serviceability) features have evolved with every server generation to maintain the robustness of the memory sub-system and address the increased likelihood of memory errors resulting from the reduced immunity to transient electrical noise.

Dell EMC has made significant investment in System RAS capabilities to maintain the highest quality of the memory sub-system on our server products.

This overview of PowerEdge system RAS capabilities developed specifically for DDR4 memory technology and their effectiveness on our server quality.

Memory Errors and Memory Error Types

The memory sub-system comprises a number of components:

- BIOS (including the reference code)
- CPU Memory Controller
- Channel (Motherboard, CPU connectors, and DIMM connectors)
- DIMM modules (DRAM components and raw card)

> **Note:** As long as the errors detected by the memory controller are correctable, the requesting target gets corrected data, and the system functions normally

Design limitations, system noise, environmental changes (temperature, humidity etc.), contamination, and failure of any DIMM sub-component (DRAM component or register, for example) can result in data errors during transmission across the memory channel between CPU and DIMM. As one of the memory RAS features, servers typically use DIMMs with "Error Checking and Correcting" (ECC) capability, so that the memory controller can detect errors on the data received from the DIMMs.

PowerEdge servers use sophisticated ECC codes as supported on the CPU to help detect and prevent data errors that can lead to silent data corruption.

However, uncorrectable memory errors detected by the ECC logic and other RAS mechanisms can result in the entire system or an application crash, based on the error type:

- Fatal uncorrectable errors—usually lead to a system crash
- Non-fatal uncorrectable errors—usually result in the Operating system (OS) exiting the applications after detecting an uncorrectable error.

Memory errors can be broadly classified into two categories—Hard errors and Soft errors.

Hard Errors

Hard errors are persistent in nature and do not go away with time, system power cycle, or reset. A hard error may result because of physical damage due to improper handling of part, electrostatic discharge (ESD) or electrical overstress (EOS), a DRAM fabrication process or module assembly defect, or a flaw in the DRAM module itself. In general, most of the hard errors that PowerEdge servers have seen on the DDR4 modules are because of reliability failure of cells in the DRAM component or physical damage to the DIMM (e.g. scratches to the board, fractured sub-components, etc.).

Some hard errors result in correctable data errors (for example, single bad cell in a DRAM component while others result in uncorrectable data failures (multiple bad bits beyond the correction capability of ECC). Dell EMC has recently observed an accelerated aging phenomenon (that is, a decreased reliability lifetime on DRAM components, when subjected to peak operation and elevated temperatures over an extended period of time) that can also lead to hard errors

While most correctable hard errors will not degrade to uncorrectable errors, there are causes of hard errors related to silicon defects that may be an early warning of cascading failure over time, ultimately leading to uncorrectable errors if not removed.

DIMM suppliers and Dell EMC have multiple safeguards in place to help prevent use of DIMMs with hard errors at various stages in our server development and validation processes, and extensive test during system build processes. Dell EMC also continually monitors the field quality and reliability characteristics of our memories and collaborates with our CPU and memory suppliers for process and reliability improvements to further protect the customer's experience with Dell EMC memory products.
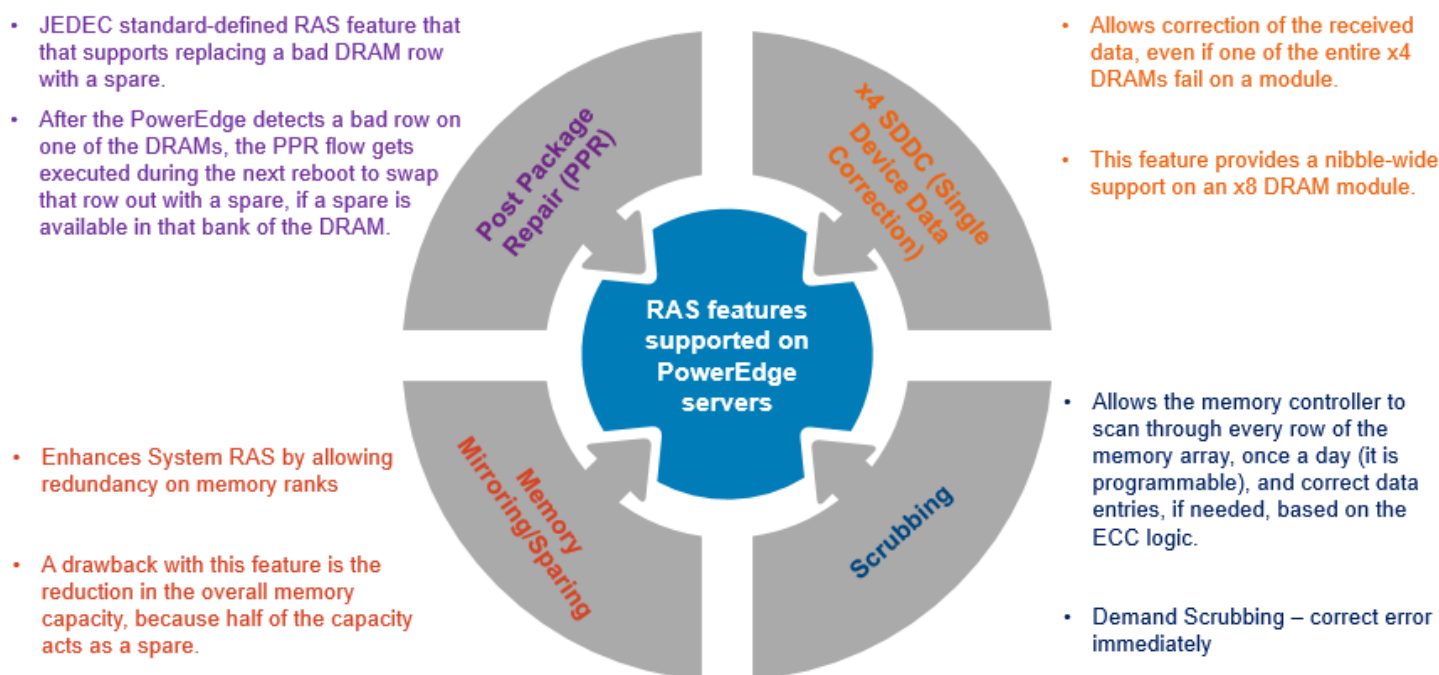
DELLEMC

Soft Errors

Soft errors are transient in nature and are caused by a brief electrical disturbance in any of the sub- components. They can further be classified into two categories by the duration of their existence—isolated and burst. Multiple factors can cause soft errors, including clock jitter, electrical noise in components, the memory channel or voltage regulator, impedance changes due to a change in system operating temperature, capacitive loading, and naturally-occurring radiation (alpha particles).

A key observation Dell EMC made at the introduction of DDR4-based servers is an increased incidence of soft errors, mostly occurring in bursts. These errors could arise because of any of the system transients, such as power noise or channel noise, and typically last from hundreds to thousands of cycles when the system is stressed at peak levels.

RAS features supported for DDR4 on PowerEdge 13G & 14G Servers

The objective of Memory RAS is to prolong the uptime of servers by helping detect and correct the transient correctable errors and proactively prevent the occurrence of uncorrectable errors. Dell EMC, in collaboration with our CPU and DIMM partners, has been strengthening the memory sub-system in every generation of PowerEdge servers to meet the challenges posed by higher capacities and speeds, smaller geometries, and reduced operational voltages.

- • JEDEC standard-defined RAS feature that that supports replacing a bad DRAM row with a spare.

- • After the PowerEdge detects a bad row on one of the DRAMs, the PPR flow gets executed during the next reboot to swap that row out with a spare, if a spare is available in that bank of the DRAM.

- • Enhances System RAS by allowing redundancy on memory ranks

- • A drawback with this feature is the reduction in the overall memory capacity, because half of the capacity acts as a spare.

Post Package Repair (PPR)

x4 SDDC (Single Device Data Correction)

RAS features supported on PowerEdge servers

Memory Mirroring/Sparing

Scrubbing

- • Allows correction of the received data, even if one of the entire x4 DRAMs fail on a module.

- • This feature provides a nibble-wide support on an x8 DRAM module.

- • Allows the memory controller to scan through every row of the memory array, once a day (it is programmable), and correct data entries, if needed, based on the ECC logic.

- • Demand Scrubbing – correct error immediately

Along with these system RAS features, Dell EMC continues to improve the proprietary predictive failure mechanism (algorithm) used to help detect the frequency and nature of correctable errors on every DIMM and warn the user when the errors exceed a technology-specific threshold. Our new algorithm discriminates transient errors from hard errors and warns the customer beforehand to help prevent hard correctable errors from potentially becoming uncorrectable errors.

Effectiveness of DDR4 Predictive Failure Analysis Algorithm

Through the use of the predictive failure analysis algorithm, Dell EMC has seen a significant improvement in our customers' experience by increasing our ability to differentiate between normal transient errors and potentially degrading components. A recent update of the BIOS' algorithm has resulted in a 27% reduction in memory warnings due to transient errors that would otherwise have advised replacement of the DIMM during the next maintenance opportunity.

DELLEMC

The continuous improvement to our error management processes benefit the customer through a reduction of unnecessary replacements driven by natural transient errors, while at the same time increasing DIMM quality improvement focus on hard failures

Conclusion

An effective server memory error management strategy is crucial, because as the memory speeds and capacities increase, voltage margins decrease, and newer semiconductor fabrication technologies continue to drive smaller DRAM cells. Dell EMC has invested in continually monitoring and analyzing the reliability performance of memories in the field to fine-tune the proprietary BIOS algorithm used to proactively identify and notify the user of those correctable memory errors which may lead to future uncorrectable errors. The result of this warning allows the system to continue to operate uninterrupted until the DIMM can be replaced at a scheduled maintenance event.

Customers can experience an improvement in memory reliability, availability, and serviceability by maintaining a regular schedule of BIOS updates to assure they have in place the most effective protection.

DELLEMC