Maximum Payload Size (MPS) vs. Maximum Read Request Size (MRS)

Sinan Kaya

okaya@codeaurora.org

Maximum Payload Size

7:5	**Max_Payload_Size** – This field sets maximum TLP payload size for the Function. As a Receiver, the Function must handle TLPs as large as the set value. As a Transmitter, the Function must not generate TLPs exceeding the set value. Permissible values that can be programmed are indicated by the Max_Payload_Size Supported field in the Device Capabilities register (see Section 7.8.3).	RW

Defined encodings for this field are:

000b	128 bytes max payload size
001b	256 bytes max payload size
010b	512 bytes max payload size
011b	1024 bytes max payload size
100b	2048 bytes max payload size
101b	4096 bytes max payload size
110b	Reserved
111b	Reserved

Maximum Read Request Size

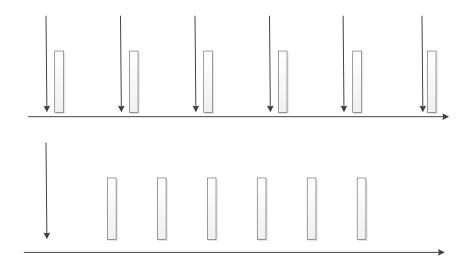| Bit Location	Register Description
14:12	**Max_Read_Request_Size** – This field sets the maximum Read Request size for the Function as a Requester. The Function must not generate Read Requests with a size exceeding the set value. Defined encodings for this field are:
	000b 128 bytes maximum Read Request size
	001b 256 bytes maximum Read Request size
	010b 512 bytes maximum Read Request size
	011b 1024 bytes maximum Read Request size
	100b 2048 bytes maximum Read Request size
	101b 4096 bytes maximum Read Request size
	110b Reserved
	111b Reserved
	Functions that do not generate Read Requests larger than 128 bytes and Functions that do not generate Read Requests on their own behalf are permitted to implement this field as Read Only (RO) with a value of 000b.
	Default value of this field is 010b.

- Note: Max_Payload_Size applies only to TLPs with data payloads; Memory Read Requests are not restricted in length by Max_Payload_Size. The size of the Memory Read Request is controlled by the Length field

- Completions must not include more data than permitted by Max_Payload_Size.
  - Receivers must check for violations of this rule. Refer to Section 2.2.

- Memory Read Requests may be completed with one, or in some cases, multiple Completions

- Software must set Max_Read_Request_Size of an isochronous-configured device with a value that does not exceed the Max_Payload_Size set for the device.

Kernel Command Line Overrides

pcie_bus_tune_off   Disable PCIe MPS (Max Payload Size) tuning and use the BIOS-configured MPS defaults.

pcie_bus_safe   Set every device's MPS to the largest value supported by all devices below the root complex.

pcie_bus_perf   Set device MPS to the largest allowable MPS based on its parent bus. Also set MRRS (Max Read Request Size) to the largest supported value (no larger than the MPS that the device or bus can support) for best performance.

pcie_bus_peer2peer   Set every device's MPS to 128B, which every device is guaranteed to support. This configuration allows peer-to-peer DMA between any pair of devices, possibly at the cost of reduced performance. This also guarantees that hot-added devices will work.

Tuning Possibility

- If the PCIe endpoint is doing a lot of reads from the system, increasing Max_Read_Request_Size saves round-trip time
- 10% performance bump was observed while running FIO workload with LSI SAS card.

Summary

- We don't trust FW.

- Maximum read request size and maximum payload size are not the same thing.

- There is an opportunity to improve performance.