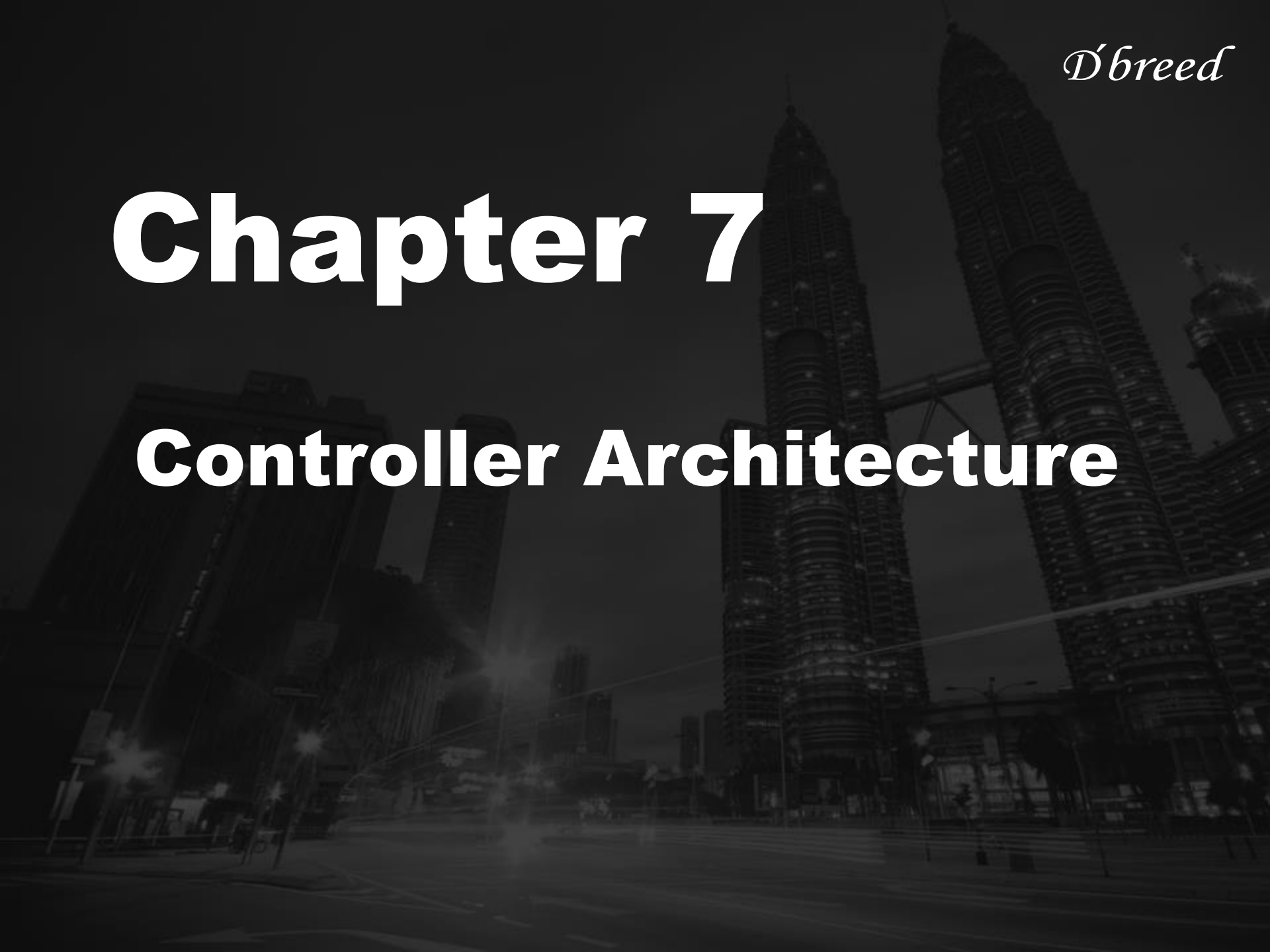


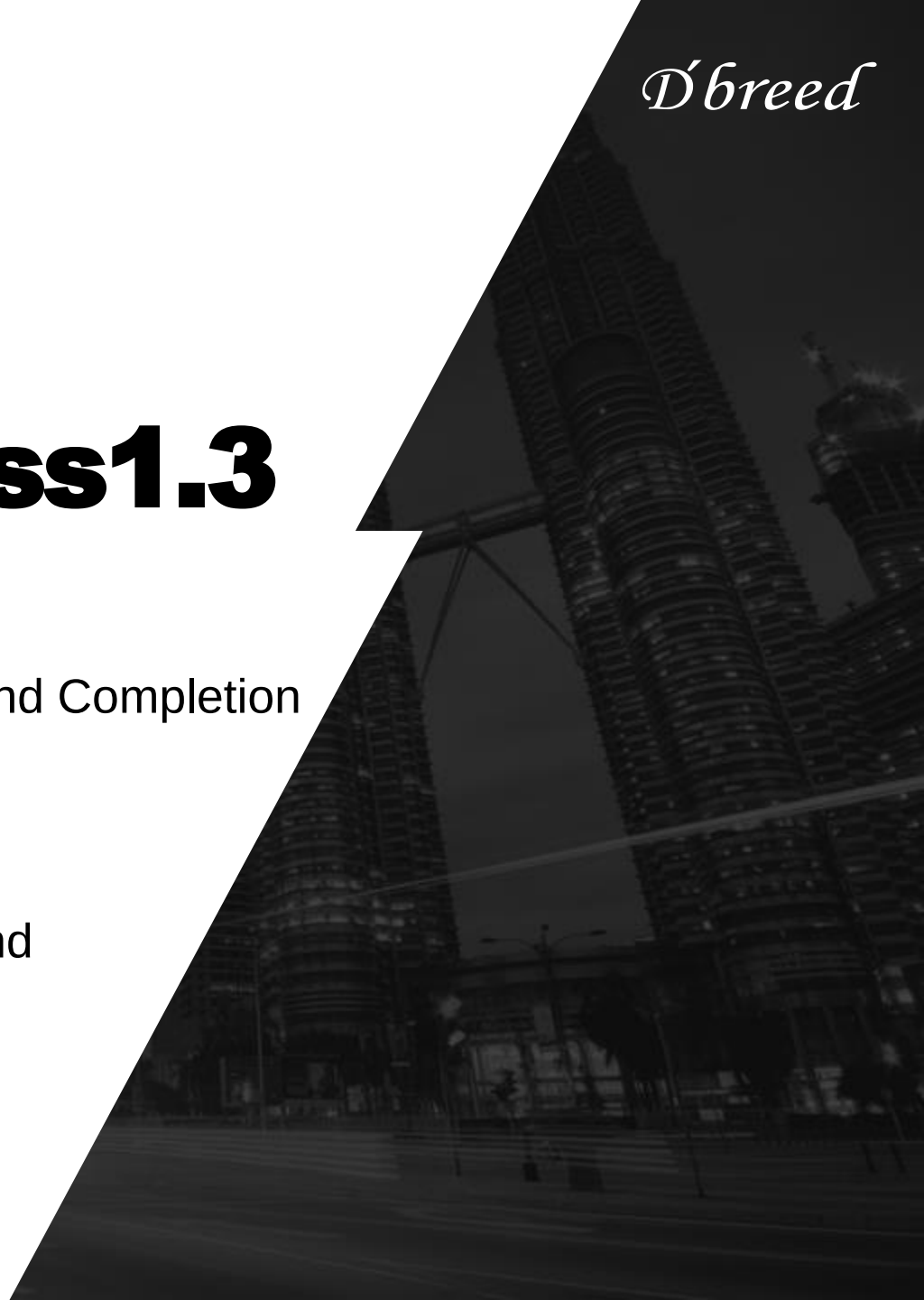
Chapter 7

Controller Architecture

Dbreed

NVM Express1.3

01. Introduction
02. Command Submission and Completion
03. Resets
04. Queue Management
05. Interrupts
06. Controller Initialization and Shutdown Processing



01 Introduction

SATA

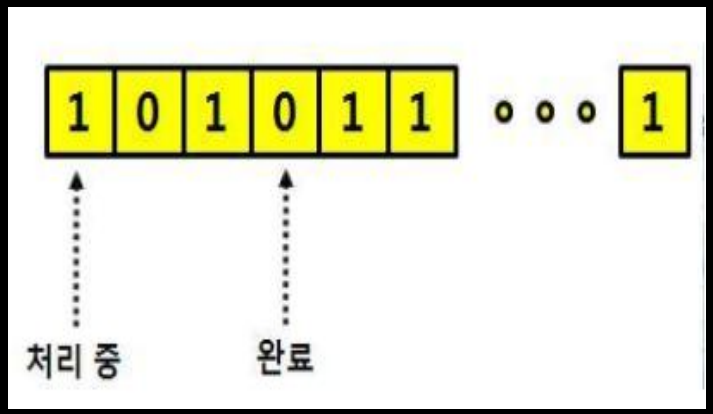
Un-cacheable register access

-9 times per command on AHCI **vs** 2 times per command on NVMe

	AHCI	NVME
Maximum Queue Depth	1 command queue 32 commands per Q	64K queues 64K Commands per Q
Un-cacheable register accesses	9 per queued command	2 per command
MSI-X	Single interrupt	2K MSI-X interrupts
Parallelism & Multiple Threads	Requires Synchronization lock to issue command	No locking

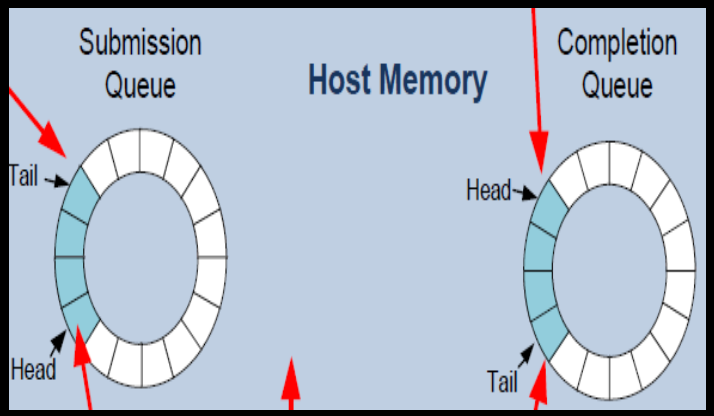
Dbreed

AHCI



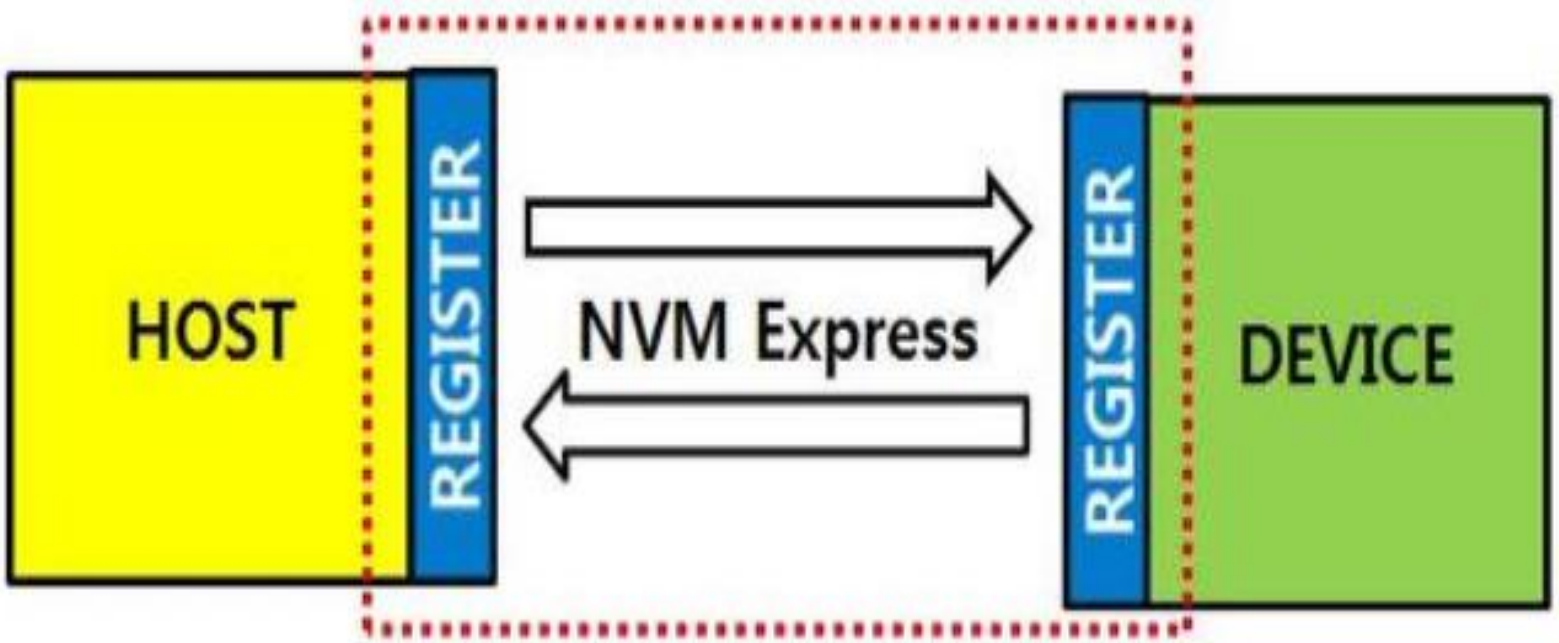
- 하나의 command List로 나누어서 command를 이슈하고 완료.
- HBA의 DMA가 command를 전달하고 처리.

Nvme



command를 처리하기 위한
SUBMISSION QUEUE와
처리가 완료된 커맨드 정보를 저장하기 위한
COMPLEITION QUEUE가지고
MULTIPLE OUTSTANDING 방식으로
커맨드를 보내고 완료.

*NVM Express



- SSD와 같은 저장 장치와 호스트 소프트웨어간 통신하는 레지스터 레벨의 인터페이스

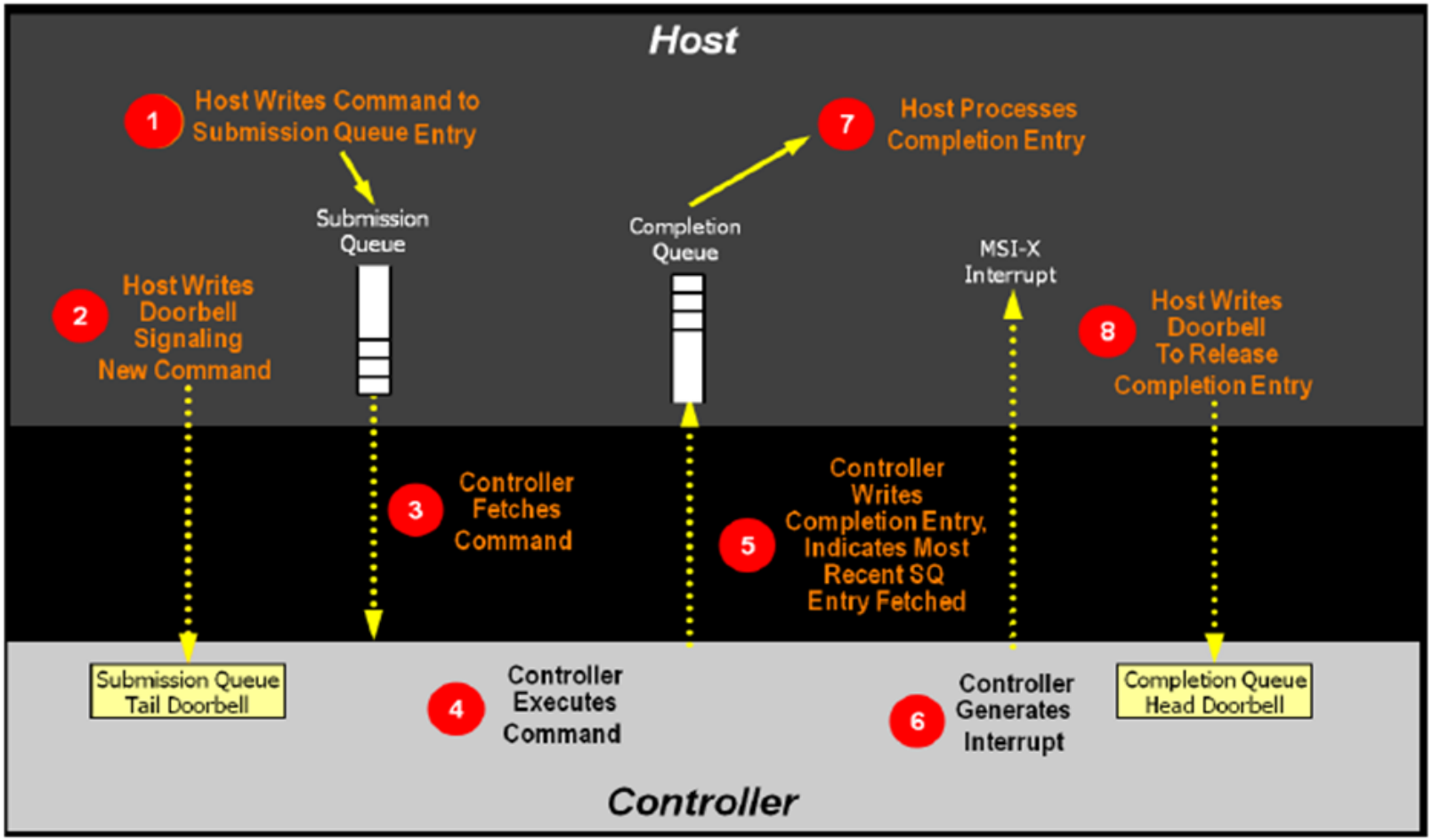
Host와 device간의 통신하는 방법에 대해 기술한 spec

02

Command Submission and Completion

1) Command Processing

Figure 178: Command Processing



2) Basic Steps when Building a Command

Figure 10: Command Dword 0

Bit	Description
31:16	**Command Identifier (CID):** This field specifies a unique identifier for the command when combined with the Submission Queue identifier.
15	**PRP or SGL for Data Transfer (PSDT):** This field specifies whether PRPs or SGLs are used for any data transfer associated with the command. If cleared to '0', the command uses PRPs for any associated data or metadata transfer. If set to '1', the command uses SGLs for any associated data or metadata transfer. PRPs shall be used for all Admin commands.
14:10	Reserved
09:08	**Fused Operation (FUSE):** In a fused operation, a complex command is created by "fusing" together two simpler commands. Refer to section 6.1. This field specifies whether this command is part of a fused operation and if so, which command it is in the sequence.
07:00	**Opcode (OPC):** This field specifies the opcode of the command to be executed.

Value	Definition
00b	Normal operation
01b	Fused operation, first command
10b	Fused operation, second command
11b	Reserved

-> 호스트는 이렇게 만들어진 command를 제출하기 위해
submission queue tail doorbell을 write한다.

3) Processing Completed Commands

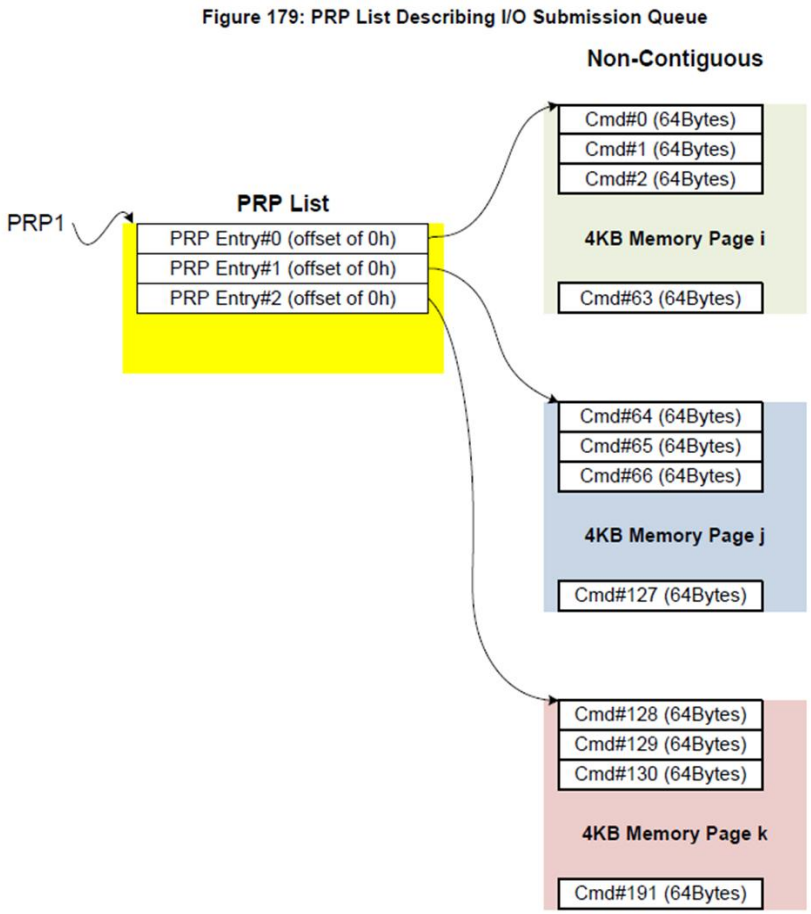
Figure 26: Completion Queue Entry Layout – Admin and NVM Command Set

	31	23	15	7	0
DW0	Command Specific				
DW1	Reserved				
DW2	SQ Identifier		SQ Head Pointer		
DW3	Status Field		P	Command Identifier	

①. 호스트는 해당 CQ 로부터 CQE 를 읽는다.
②. 호스트는 해당 completion 을 생성한 SQE 를 알아내기 위해 CQE 를 처리한다. DW2.SQID에 SQ ID 가 나타나고, DW3.CID 에 completion 을 생성한 커맨드가 나타난다.
③. DW3.SF는 completion 의 상태를 나타낸다.
④. 호스트 소프트웨어는 CQyHDBL을 업데이트함으로써 사용 가능한 CQ slot 을 알리고 관련된 interrupt는 clear 된다.

Dbreed

4) Creating an I/O Submission Queue



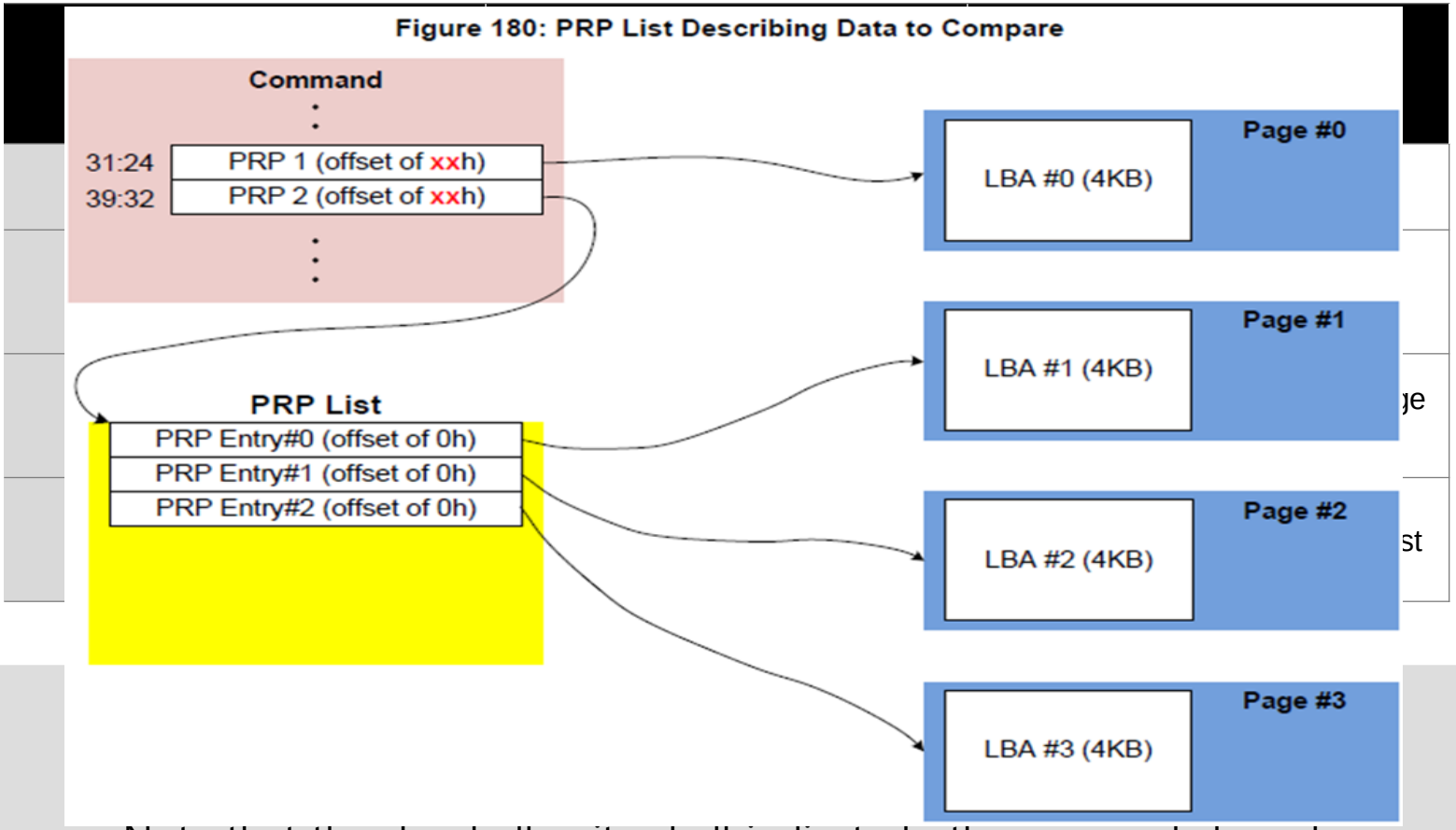
①. Create the I/O Completion Queue that the SQ uses with the Create I/O Completion Queue command

②. Builds a Create I/O Submission Queue command for the Admin Submission Queue.

③. Submits the command for execution by writing the Admin Submission Queue doorbell (SQ0TDBL)

④. Maintain the PRP List unmodified in host memory until the Submission Queue is deleted.

5) Executing a Fused Operation



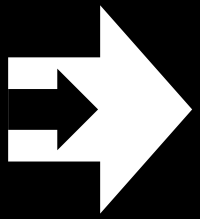
- Note that the doorbell write shall indicate both commands have been submitted at one time

03

Resets

1) Five primary Controller Level Reset mechanisms

- NVM Subsystem Reset
- Conventional Reset
- PCI Express transaction layer Data Link Down status
- Function Level Reset
- Controller Reset



- The controller stops processing any outstanding Admin or I/O commands.
- All I/O Submission Queues are deleted.
- All I/O Completion Queues are deleted.
- The controller is brought to an Idle state.

2) Queue Level

A queue level reset is performed by deleting and then recreating the queue.

①. 호스트는 삭제할 queue의 ID를 명시하는 Delete I/O SQ 또는 CQ command를 Admin Queue에 제출한다.

②. 삭제 후, 호스트는 Create I/O SQ 또는 CQ command로 queue를 재 생성한다.

NOTE

- 호스트는 SQ 또는 CQ를 삭제하기 전에 해당하는 queue가 idle 상태임을 확인해야 한다.
- queue level reset이 실행되면 동일한 명령 내에서 CQ를 사용하는 SQ도 reset됨을 주의해야 한다.

Dbreed

04 Queue Management

1) Queue Setup and Initialization

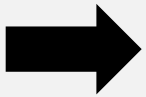
①. ASQ, ACQ 레지스터들을 알맞게 초기화 함으로서 ASQ와 ACQ를 설정한다.

②. 원하는 I/O SQ, CQ의 수를 Set Features command를 통해 제출한다.

③. Queue 당 최대 지원하는 엔트리 수와 queue가 물리적으로 연속되어야 하는지 결정한다.

④. 컨트롤러에 의해 할당된 수의 범위와 Create I/O CQ command의 설정 값 내에서 원하는 I/O CQ를 생성한다.

⑤. 컨트롤러에 의해 할당된 수의 범위와 Create I/O SQ command의 설정 값 내에서 원하는 I/O SQ를 생성한다.



The desired I/O Submission Queues and I/O Completion Queues have been setup and initialized

2) Queue Coordination & Abort

Coordination

- 여러 개의 I/O queue 조합에 대응하는 하나의 Admin queue 조합이 있다.
- Admin command는 여러 I/O queue의 조합에 영향을 줄 수 있다.
- 호스트는 에러를 피하기 위해 Admin 작업들과 I/O queue 조합에 관여하는 thread들 사이가 잘 조정되어야 한다.

Abort

- 많은 수의 command를 취소하기 위해서는 I/O SQ를 삭제하고 재 생성해야 한다.
- Queue가 성공적으로 삭제되면 호스트는 Create I/O SQ command를 제출함 으로서 queue를 재 생성한다.

Dbreed

05 Interrupts

1) MSI Behavior

controller

Completion Queue에 미 확인 큐 존재.

&

Create I/O Completion Queue command에서 interrupt 사용 가능.

&

interrupt가 마스크 되어 있지 않을 때

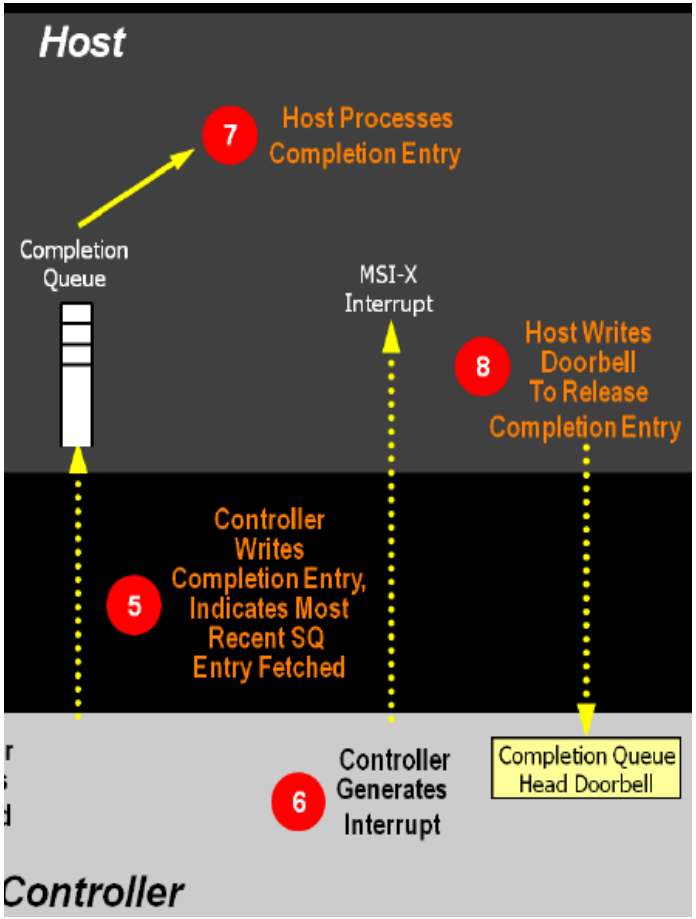
➡ IS (interrupt status register) bit is set to '1'

Host

- 인터럽트 서비스 내에서 INTMS 레지스터를 통해 레지스터 비트를 '1'로 설정
- 모든 CQE가 처리되면 INTMC 레지스터를 통해 레지스터 비트를 '0'으로 clear

Dbreed

2) Interrupt Example



①. Controller는 IS 레지스터의 비트를 '1'로 설정한다.
->controller는 호스트로 인터럽트를 발생시킨다.

②. 호스트는 새로운 CQE의 위치를 알아내기 위해 모든 I/O CQ를 검색한다.

③. 호스트는 I/O CQ와 연결된 interrupt에 mask 하기위해 INTMS 레지스터에 08h 를 쓴다.

④. Controller는 INTMS 레지스터에 쓴 것을 근거로 interrupt에 mask한다.

⑤. 호스트는 I/O CQ의 새로운 CQE를 처리한다.

⑥. CQE 처리를 완료했다면 CQyHDBL을 업데이트한다.

⑦. 호스트는 INTMC 레지스터에 08h를 write하여 인터럽트를 unmask한다.

Dbreed

06

Controller Initialization and Shutdown Processing

1) Initialization

①. 시스템 설정에 기초하여 PCI 및 PCIe 레지스터를 설정한다.

②. 호스트는 컨트롤러가 이전 리셋 작업이 완료될 때까지 대기한다.

③. Admin queue를 설정한다.

④. 컨트롤러 셋팅이 이루어져야 한다.

⑤. CC.EN을 '1'로 설정함 으로서 컨트롤러가 활성 된다.

⑥. 컨트롤러는 CSTS.RDY(Ready)가 '1'로 설정되면서 커맨드 처리 준비가 완료된다.

⑦. 호스트는 Set features 커맨드에 지원되는 I/O SQ, CQ 의 개수를 결정한다.
결정되면 MSI and/or MSI-X 레지스터가 설정되어야 한다.

⑧. 호스트는 컨트롤러가 지원하는 수만큼 I/O CQ/SQ 를 할당한다.

➡ The controller may be used for I/O commands.

2) Shutdown

①. 새로운 I/O command 제출을 중단하고 미 처리된 command를 처리한다.

②. 호스트는 모든 I/O SQ를 삭제하고 미 처리 되어있던 command는 취소된다.

③. 호스트는 모든 I/O CQ를 삭제한다.

④-ⓐ. 일반적인 shutdown에서는 호스트는 Shutdown Notification필드를 01b로 설정하여 normal shutdown operation 을 알린다.
컨트롤러는 Shutdown Status필드를 10b 로 셋팅하여 shutdown 처리가 완료되었음을 알린다

④-ⓑ. 갑작스러운 shutdown에서는 호스트는 Shutdown Notification필드를 10b로 설정하여 abrupt shutdown operation 을 알린다.
컨트롤러는 Shutdown Status필드를 10b 로 셋팅하여 shutdown 처리가 완료되었음을 알린다.

➡ Shut down operation 후에 컨트롤러에 command 실행을 시작하기 위해 reset이 필요하다.


Dbreed
THANK
YOU