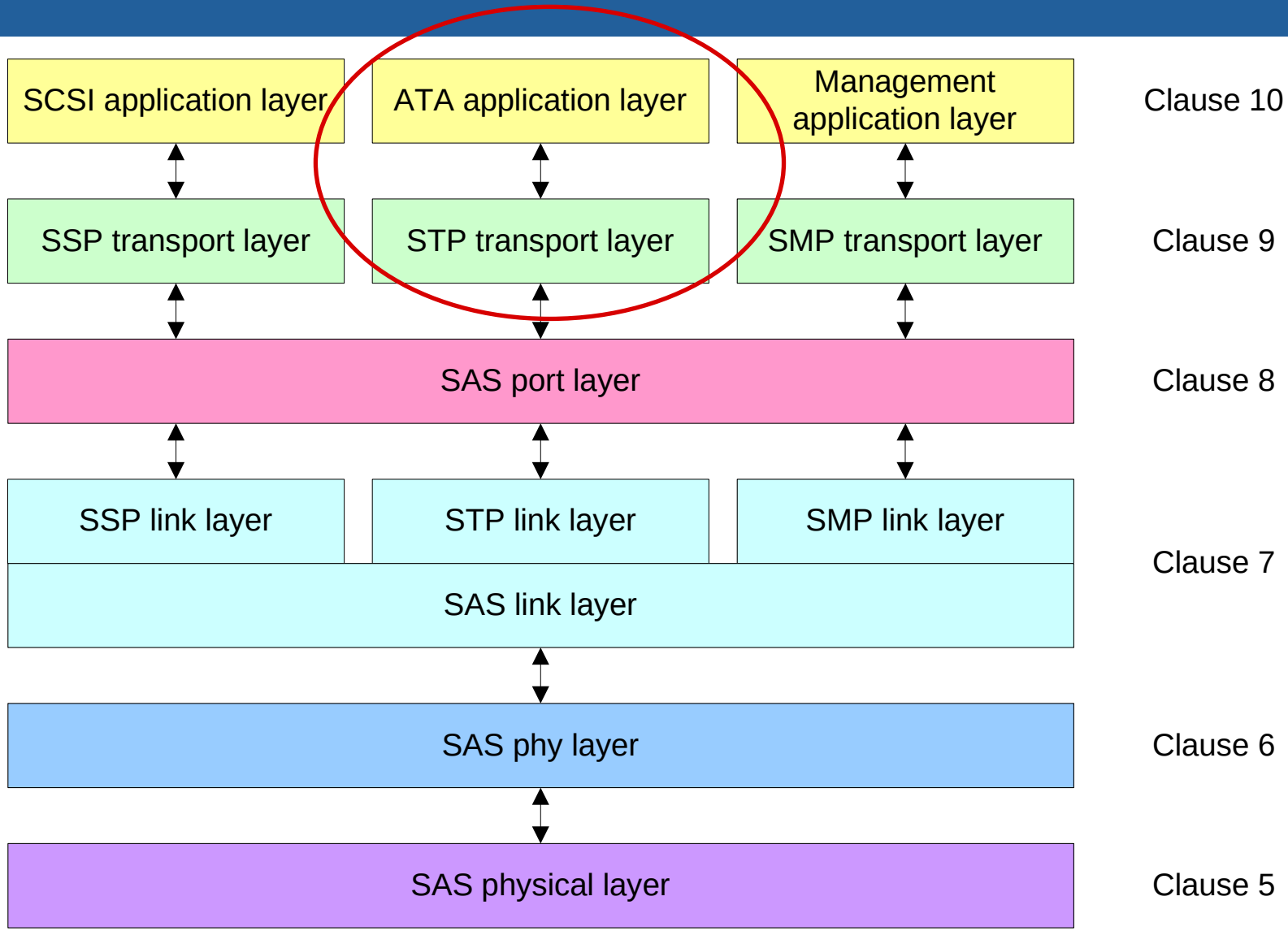
Serial Attached SCSI
ATA upper layers

by Rob Elliott
HP Industry Standard Servers
Server Storage Advanced Technology
elliott@hp.com http://www.hp.com
30 September 2003

Notice

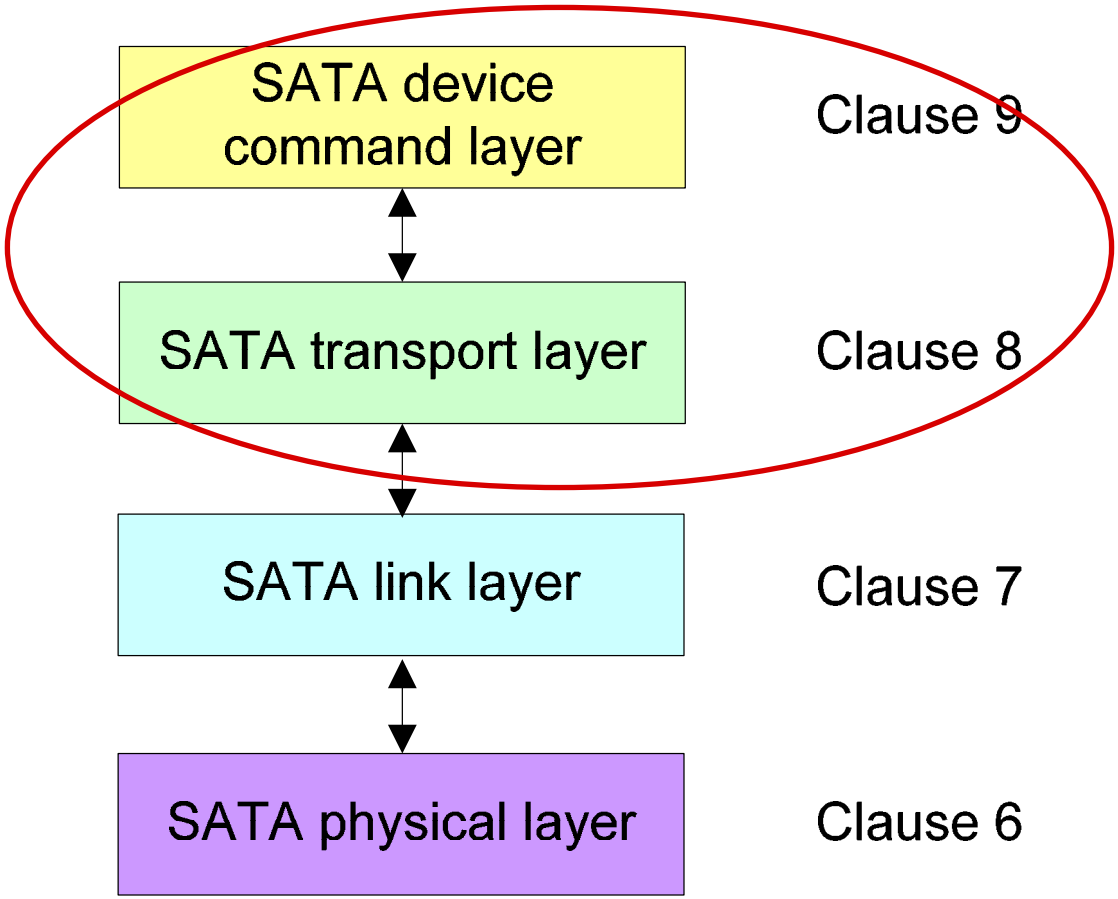
- These slides are freely distributed by HP through the SCSI Trade Association (http://www.scsita.org)
- STA members are welcome to borrow any number of the slides (in whole or in part) for other presentations, provided credit is given to the SCSI Trade Association and HP
- This compilation is © 2003 Hewlett-Packard Corporation

SAS standard layering



Serial ATA 1.0a standard layering

- For SATA 1.0a from the private Serial ATA working group



ATA/ATAPI-7 standard layering

- For the public standard ATA/ATAPI-7
- Subject to change by T13 standards committee

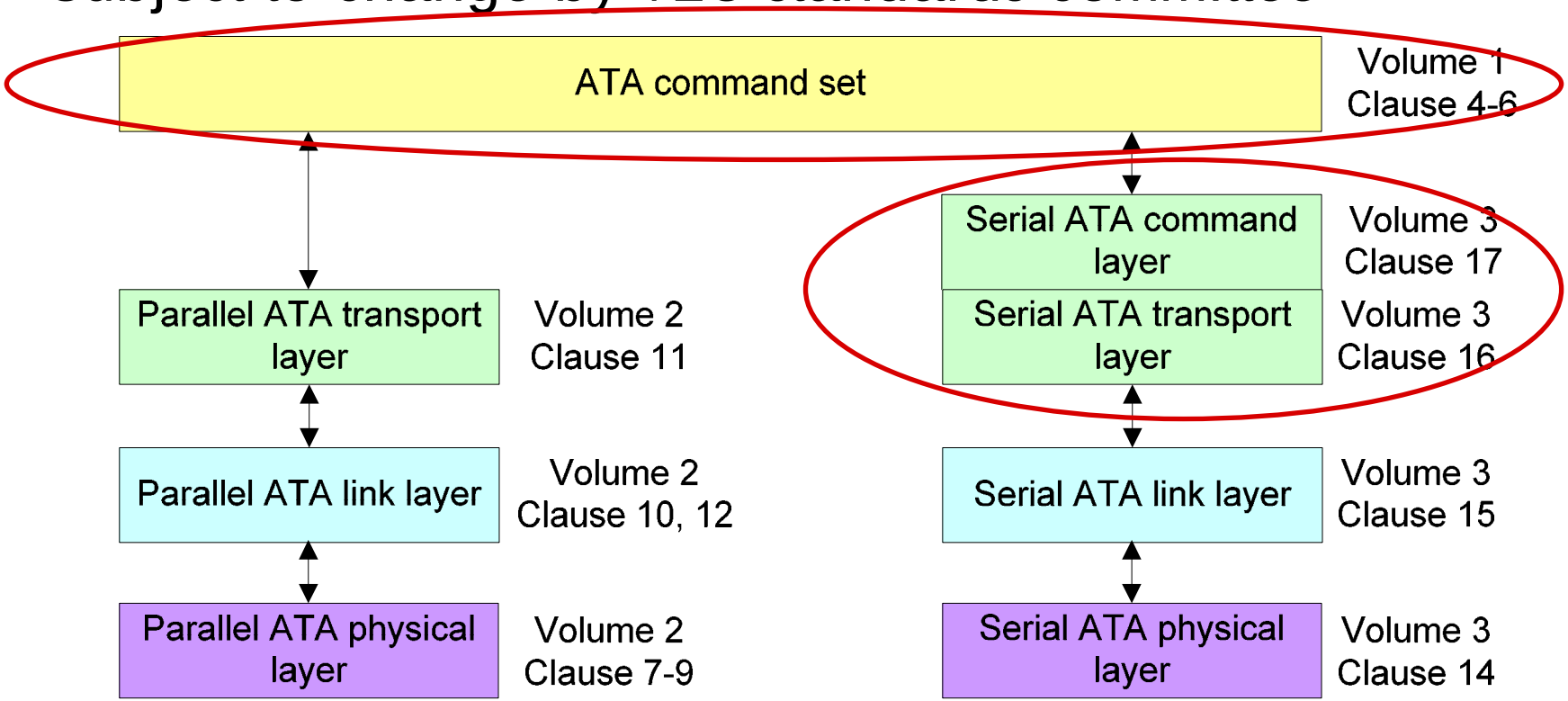


Table of contents

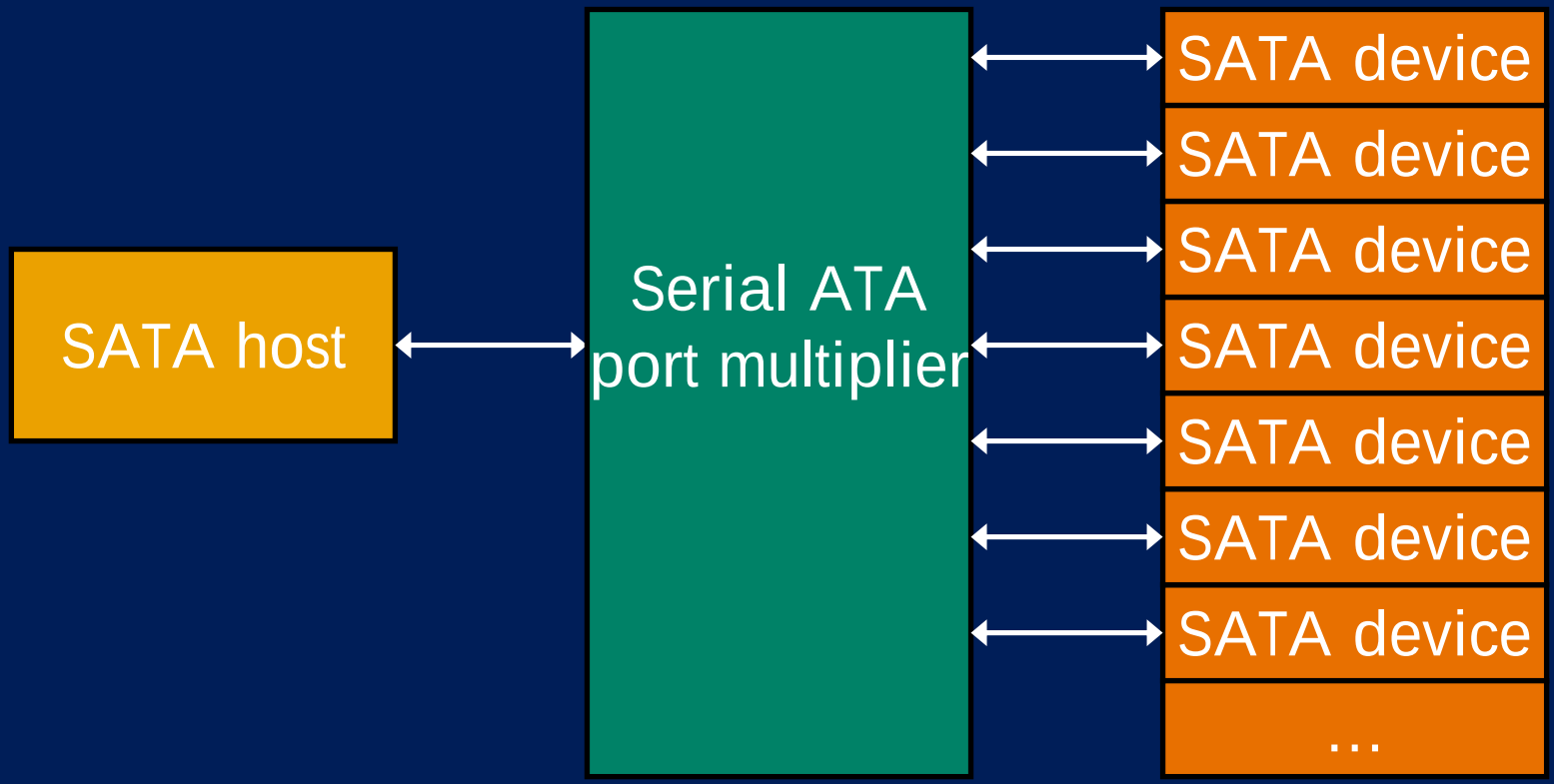
ATA architecture

ATA overview

- ATA hosts send ATA commands to ATA devices
- Two fundamentally different ATA device types
  - ATA
    - Disk drives and flash cards are pure ATA devices
  - ATAPI (ATA Packet Interface)
    - Accept SCSI commands via the ATA PACKET command
    - CDs, DVDs, and tape drives are all ATAPI devices
    - SCSI commands as defined in SPC-3, MMC-5, SSC-2, etc.
    - However, lots of limitations

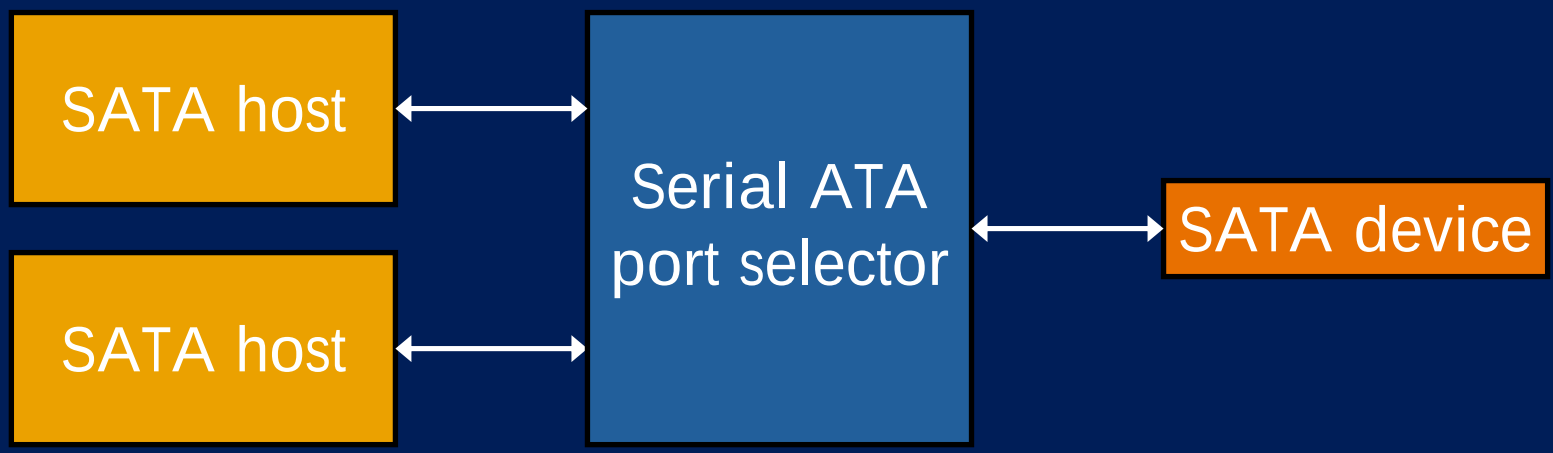
Serial ATA II port multiplier

- Serial ATA II port multipliers attach up to 15 devices to a host
  - Stores and forwards frames from the host based on a field in the frame header
  - Fills in that field for frames from devices



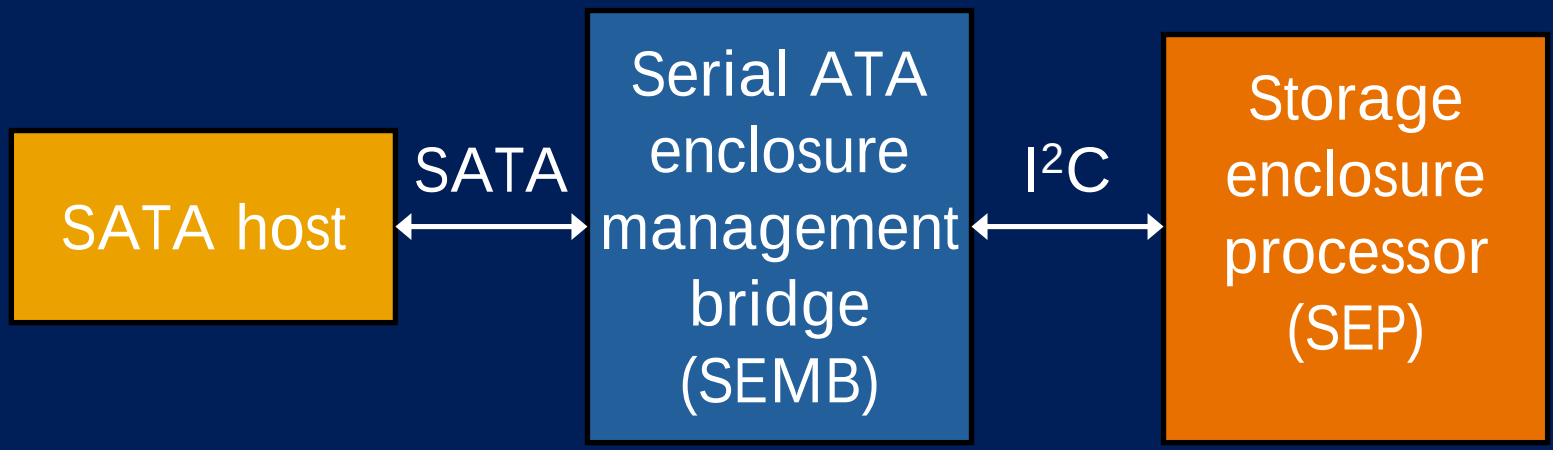
Serial ATA II port selector

- Serial ATA II port selector let a device talk to two hosts
  - First host to run OOB is active; the latecomer is inactive
  - Switch over
    - Inactive host can take control by sending a particular sequence of COMINIT OOB signals
    - Sideband signals also allowed



Serial ATA II enclosure management

- Serial ATA II enclosure management bridge (SEMB)
  - Standalone SATA device
  - Part of a SATA II port multiplier (virtual device)
- Accepts new ATA commands that carry SCSI Enclosure Services (SES) or SAF-TE payloads
- Forwards payloads to an SEP over I^2C



ATA task file

ATA task file

- ATA is defined with a task file register interface
- ATA/ATAPI-7 Volume 1 defines the task file registers

Offset	Reads	Writes	Notes
Cmd+0	Data		
Cmd+1	Error	Feature	
Cmd+2	Sector Count		
Cmd+3	LBA Low		Formerly Sector Number
Cmd+4	LBA Mid		Formerly Cylinder Low
Cmd+5	LBA High		Formerly Cylinder High
Cmd+6	Device		
Cmd+7	Status	Command	
Ctrl+2	Alternate Status	Device Control	

- Sector Count, LBA Low, LBA Mid, and LBA High actually point to two 8 bit registers
  - Writes: accesses take turns
  - Reads: bit in Device Control register selects which one

ATA task file addressing

- Registers historically mapped to x86 PC I/O space
- PCI standard BAR assignments

Channel	Command block	Control block
Primary	PCI BAR 1 01F0h	PCI BAR 1 03F6h
Secondary	PCI BAR 3 0170h	PCI BAR 3 0376h

- One PCI function supports 2 channels
- PCI BAR 4 points to bus master control registers
  - DMA command support
- Alternative interfaces possible
  - Better queuing offload, SATA II native queuing, more channels
- ATA Host Adapters – 1 defines the PCI configuration space implementation

ATAPI task file

- When talking to an ATAPI device, the same register banks are used with some different register names

Offset	Reads	Writes	ATA name
Cmd+0	Data		
Cmd+1	Error	Feature	
Cmd+2	Interrupt Reason	Reserved	Sector Count
Cmd+3	Reserved		LBA Low
Cmd+4	Byte Count Low		LBA Mid
Cmd+5	Byte Count High		LBA High
Cmd+6	Device		
Cmd+7	Status	Command	
Ctrl+2	Alternate Status	Device Control	

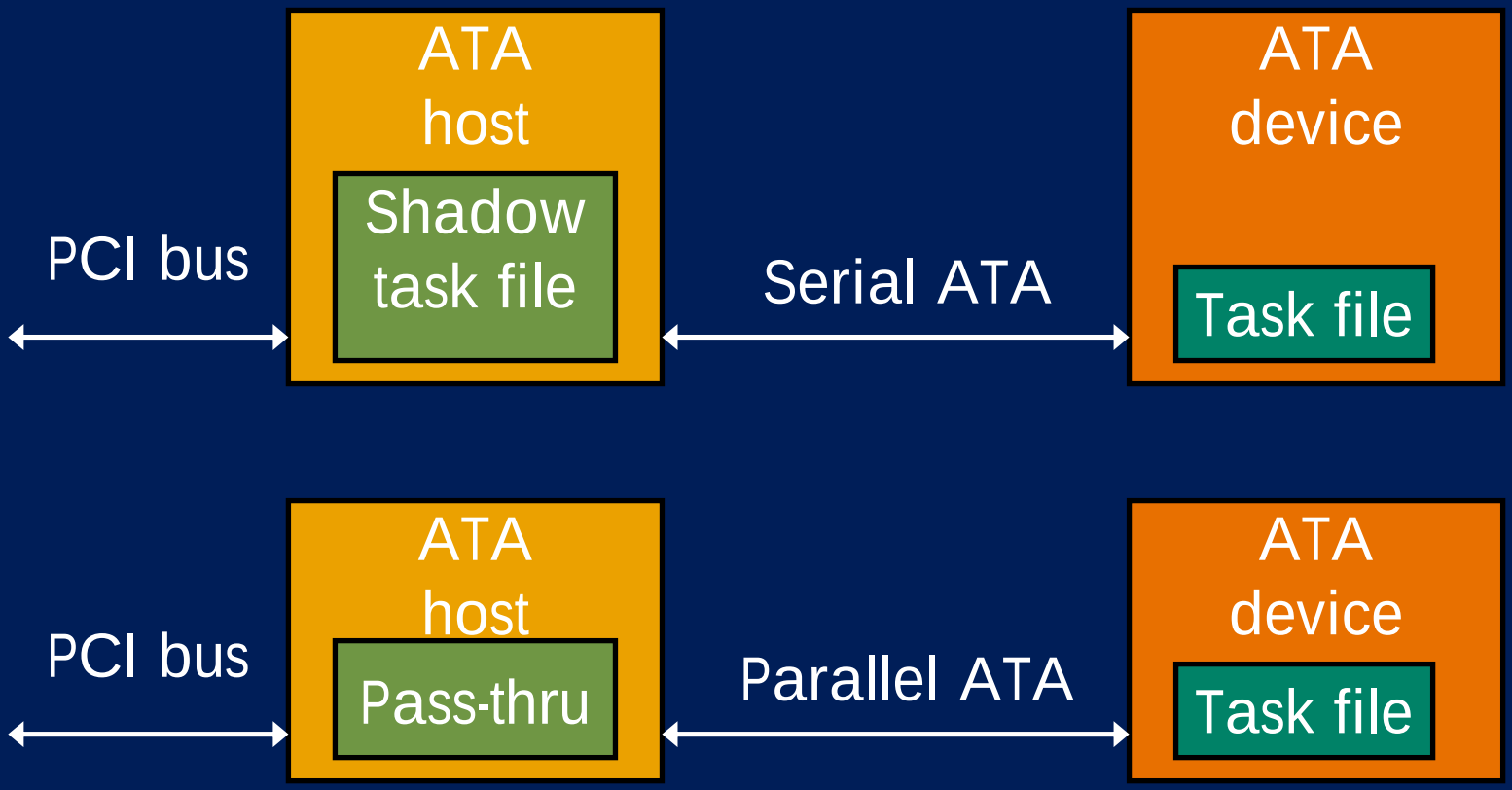
Serial ATA registers

- SATA defines an additional register bank
- Location is vendor-specific
- Since there is no standard location, all that matters is that equivalent functionality is provided

Offset	Register	Description
0	SStatus	Information about current state
1	SError	Reports SATA-specific errors
2	SControl	Control SATA-specific features
3	SActive	SATA II native queuing commands outstanding
4 to 15	Reserved	

Task file location

- The task file is in the ATA device
- In parallel ATA, accesses to these registers result in parallel ATA traffic
- In serial ATA, a Shadow Task file register bank is also managed by the host to mirror the ATA device's task file



Task file power up signature

- After reset or running the EXECUTE DEVICE DIAGNOSTIC command, devices place certain patterns in their task file registers

Register	ATA	ATAPI	SATA port multiplier	SEMB	SEMB without SEP
Sector Count	01h	01h	01h	01h	FFh
LBA Low	01h	01h	01h	01h	FFh
LBA Mid	00h	14h	69h	3Ch	FFh
LBA High	00h	EBh	96h	C3h	FFh
Device	00h	00h or 10h[1]	00h	00h	FFh

[1] May be 10h if a DEVICE RESET, IDENTIFY DEVICE, or READ SECTORS command has completed

ATA Command register

- Write-only, offset Cmd+7
- Host sends a Command Code to the device
  - Like the SCSI Operation Code field
- Starts execution of the command
  - Must write all the other registers first
- Annex in ATA/ATAPI-7 Volume 1 lists all command codes
- Each command is specified in ATA/ATAPI-7
  - list of the task file contents used to invoke the command, similar to the SCSI CDB
  - list of the task file contents after running the command, similar to SCSI Status and Sense Data

Offset	7	6	5	4	3	2	1	0
Cmd+7	Command Code							

ATA Data register

- Read-Write, offset Cmd+0
- Although it looks like an 8-bit register, 16-bit accesses can be done to offset 0h
  - Does not affect the neighboring Error/Features registers at offset 1h
- For PIO commands, this is the portal for software to read or write data
- For DMA commands, the host adapter accesses memory directly
  - BAR 4 register bank points to a PRD (physical region descriptor table) scatter-gather list which directs the data transfers

Offset	7	6	5	4	3	2	1	0
Cmd+0	Data[7:0]							
Cmd+1	Data[15:8] (only accessible via 16-bit accesses to Cmd+0)							

ATA Device register

- Read-Write, offset Cmd+6
- Device Select bit
  - Used to select device 0 vs. device 1 in parallel ATA
  - In Serial ATA, always 0
    - Host adapter could pretend to support two devices per channel
    - Microsoft discouraged that
- All other bits are command-specific

Offset	7	6	5	4	3	2	1	0
Cmd+6			Device Select					

ATA Device Control register

- Write-only, offset Ctrl+2
- HOB (high-order) bit
  - Used to read extra address bits for 48-bit commands
- SRST (Software Reset) bit
  - Used to force a soft device
  - Decodes of writes to this bit could be done by low-power hardware in parallel ATA
    - Not as simple for serial ATA
- nIEN (Interrupt Enable) bit
  - allows device to interrupt the host
  - 0 = interrupt enabled
  - 1 = interrupt disabled

Offset	7	6	5	4	3	2	1	0
Ctrl+2	HOB					SRST	nIEN	

ATA Status registers

- Status register
  - Read-only, offset Cmd+7
  - Reading this clears a pending interrupt

Offset	7	6	5	4	3	2	1	0
Cmd+6	BSY	DRDY	DF		DRQ			ERR

- Alternate Status register
  - Read-only, offset Ctrl+2
  - Same contents as Status register
  - Reading this register does not clear a pending interrupt

Offset	7	6	5	4	3	2	1	0
Ctrl+2	BSY	DRDY	DF		DRQ			ERR

ATA Status registers – BSY and DRDY bits

- BSY (Busy) bit
  - 1 indicates the device is busy
    - Device controls the task file
    - Host writes to task file are prohibited (except sending a DEVICE RESET command)
    - If host reads from the task file, only the BSY bit is valid
  - 0 indicates the device is not busy
    - Host controls the task file
- DRDY (Device Ready) bit
  - 1 indicates the device is ready to accept all commands
  - 0 indicates the device only accepts a few commands
    - DEVICE RESET, EXEUTE DEVICE DIAGNOSTIC, IDENTIFY PACKET DEVICE, PACKET

ATA Status registers – DF, DRQ, and ERR bits

- DF (Device Fault) bit
  - 1 indicates the device had a major error processing a command
- DRQ (Data request) bit
  - 1 indicates the device is ready for data transfer
- ERR (Error) bit
  - 1 indicates an error occurred during execution of the previous command
  - For ATAPI devices, this is called the CHK bit and indicates a SCSI Status of CHECK CONDITION

Other ATA registers

- Error register
  - Read-only, offset Cmd+1
  - ABRT (Abort) bit
    - 1 if command code or a parameter was invalid or some other problem occurred
  - All other bits are command-specific
- Features, LBA Low, LBA Mid, LBA High, and Sector Count registers
  - All bits are command-specific

ATA Sector Count and LBA registers

- Originally, ATA supported CHS (cylinder head sector) addressing
- Later, it added 28-bit LBA (logical block address) addressing
  - 137 GB maximum
- Recently, ATA/ATAPI-6 added 48-bit addressing
  - Sector Count and LBA registers treated as 2-deep FIFOs
    - Write the extended bits first (e.g. LBA [31:24], then the normal bits (e.g. LBA [7:0])
  - New commands, all with EXT in their name, are 48-bit commands
    - e.g. READ DMA EXT is 48-bit, READ DMA is 28-bit
  - 28-bit commands unchanged

Offset	Register	Description
Cmd+2	Sector Count	Sector Count [7:0] and [15:8]
Cmd+3	LBA Low	LBA [7:0] and [31:24]
Cmd+4	LBA Mid	LBA [15:8] and [39:32]
Cmd+5	LBA High	LBA [23:16] and [47:40]
Cmd+6	Device	Includes LBA [27:24] in 28-bit addressing mode

Serial ATA SStatus register

- DET (device detection) field reports basic phy status

Value	Description
0	No device attached
1	Device attached
3	Device attached and running
4	Phy is disabled or in loopback mode
Others	Reserved

- SPD (speed) field reports the speed being used

Value	Description
0	No device attached
1	1.5 Gbps
2	3.0 Gbps
3 to 15	Reserved

- IPM (interface power management) field reports the power management state of the phy

Value	Description
0	No device attached
1	Active state
2	Partial state
3	Slumber state
4 to 15	Reserved

Serial ATA SError register

- DIAG field reports specific errors

Bit	Description
B	10b8b decode error
C	CRC error
D	Disparity error
F	Unrecognized FIS Type
I	Phy internal error
N	PhyRdy has changed
H	Handshake error – R_ERR received
S	Link sequence error – primitive sequence problems
T	Transport state transition error – frame sequence problems
W	COMWAKE received
X	Exchanged – device presence changed

- ERR field indicates which errors occurred

Value	Description
C	Non-recovered error
E	Internal error
I	Recovered error
M	Recovered error
P	Protocol error
T	Non-recovered error

Serial ATA SControl register

- DET (device detection) field reports basic phy status

Value	Description
0	No device attached
1	Force a phy reset sequence
4	Disable the phy
Others	Reserved

- SPD (speed) field limits the speeds that can be used

Value	Description
0	No device attached
1	Limit to 1.5 Gbps
2	Limit to 1.5 and 3.0 Gbps
4 to 15	Reserved

- IPM (interface power management) field controls whether the phy honors power management requests (e.g. via PMREQ primitives)

Value	Description
0	Honor requests
1	Ignore requests for Partial state
2	Ignore requests for Slumber state
3	Ignore requests for both Partial and Slumber states
4 to 15	Reserved

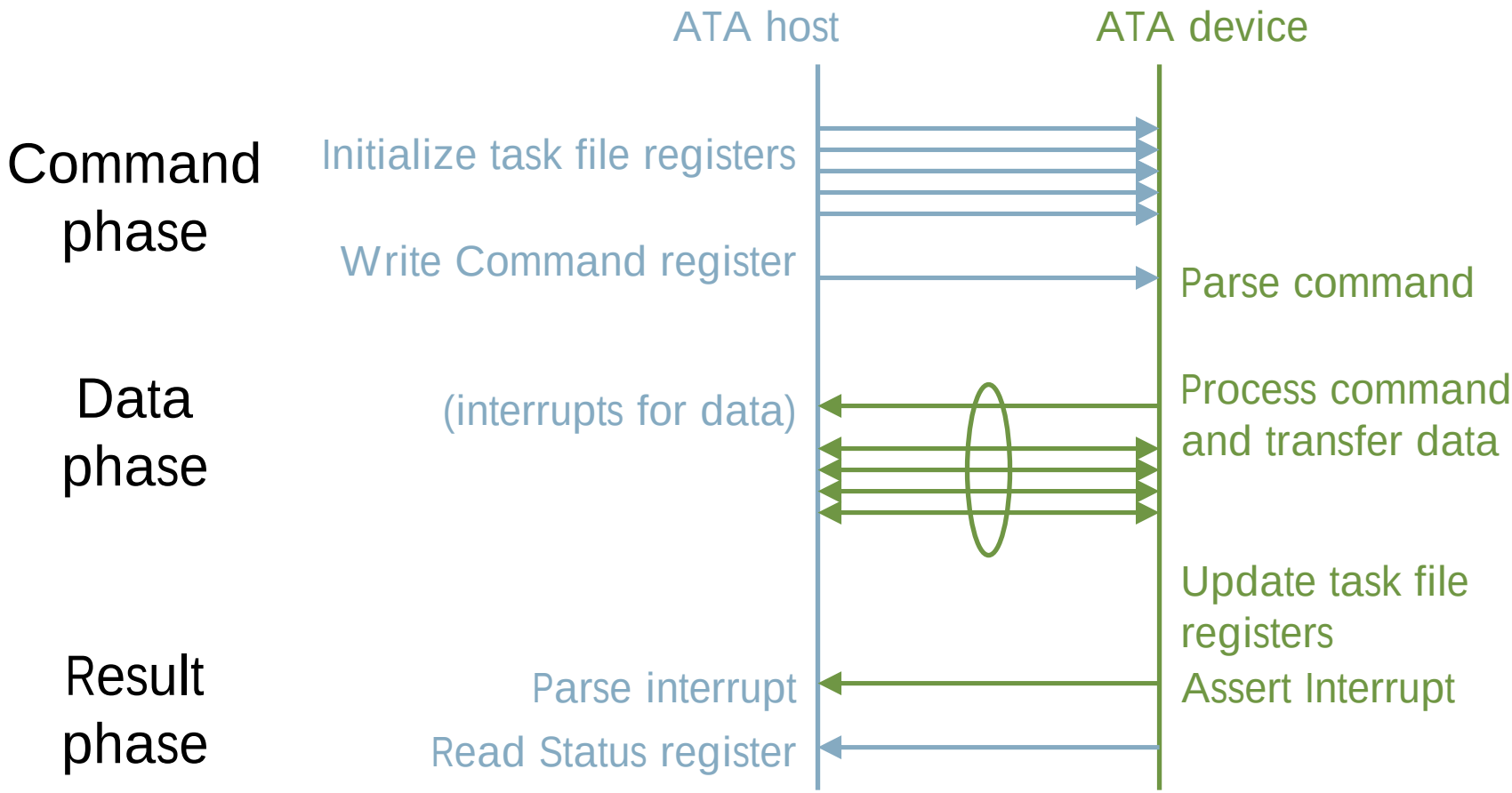
Serial ATA SActive register

- 32-bit register defined by Serial ATA II
- One bit per queue tag
  - Bit 0 corresponds to tag 00h
  - Bit 31 corresponds to tag 01Fh
  - 1 = command with that tag is pending
  - 0 = command with that tag is not pending
- The host sets bits to 1 by writing the register
  - Host can only clear bits by issuing a reset
- The device clears bits with the Set Device Bits FIS

ATA commands

ATA basic command sequence

- In parallel ATA, all register accesses go over the ATA bus to the device
- In Serial ATA, register updates are batched together



ATA command types

- Different sequences of transfers on the bus for different commands

Type of command	Description
Device Reset	Reset a device
Execute Device Diagnostic	Run diagnostics
Non-data	No data transferred
PIO read	Software-based read
PIO write	Software-based write
DMA read	DMA controller manages read
DMA write	DMA controller manages write
DMA read queued	DMA controller read with legacy queuing (rarely used)
DMA write queued	DMA controller write with legacy queuing (rarely used)
FPDMA read	Serial ATA native DMA read
FPDMA write	Serial ATA native DMA write
Packet	Encapsulated SCSI commands (PIO or DMA based)

ATA feature sets

- ATA commands are grouped into feature sets
- IDENTIFY DEVICE command often indicates if an entire feature set is supported

Feature set	Feature set
General	48-bit address
PACKET command	Device configuration overlay
Power management	Media card pass through
Advanced power management	Streaming
Security mode	Time-limited read / write
SMART (Self-monitoring, analysis, and reporting)	General purpose logging
Host protected area	Overlapped
CompactFlash Association	Queued
Removable media	Serial ATA II Enclosure management
Power-up in standby mode	Serial ATA II First-party DMA
Automatic acoustic management	

ATA commands – mandatory for ATA devices - part 1

Command	Type	Description
CHECK POWER MODE	N	Report if device is in Active, Idle, or Standby mode
EXECUTE DEVICE DIAGNOSTIC	EDD	Perform internal diagnostics
FLUSH CACHE	N	Flush the write cache
IDENTIFY DEVICE	PIO R	Return lots of information about the device, including capabilities, Serial number, Model number, Firmware revision, features, version numbers, etc.
IDLE, IDLE IMMEDIATE	N	Send device into Idle power management mode
READ DMA	DMA R	Read data (using DMA)
READ MULTIPLE	PIO R	Read multiple sectors of data (using PIO)
READ SECTORS	PIO R	Read data (using PIO)
READ VERIFY SECTORS	N	Read data from media but don't return it to the host

ATA commands – mandatory for ATA devices - part 2

- ATAPI devices have different requirements (shown later)

Command	Type	Description
SET FEATURES	N	Control optional features
SET MULTIPLE MODE	N	Establishes block count for READ / WRITE MULTIPLE commands
SLEEP	N	Send device into Sleep power management mode
STANDBY, STANDBY IMMEDIATE	N	Send device into Standby power management mode
WRITE DMA	DMA W	Write data (using DMA)
WRITE MULTIPLE	PIO W	Write multiple sectors of data (using PIO)
WRITE SECTORS	PIO W	Write data (using PIO)

ATA commands – prohibited for ATA devices

- Several commands are not allowed for non-ATAPI devices

Command	Restriction
DEVICE RESET	Packet only
IDENTIFY PACKET DEVICE	Packet only
PACKET	Packet only

ATA commands – mandatory for ATAPI devices

Command	Type	Description
CHECK POWER MODE	N	Report if device is in Active, Idle, or Standby mode
DEVICE RESET	DR	Reset the device
EXECUTE DEVICE DIAGNOSTIC	EDD	Perform internal diagnostics
IDENTIFY DEVICE	PIO R	Return the PACKET feature set signature
IDENTIFY PACKET DEVICE	PIO R	Return lots of information about the device, including capabilities, Serial number, Model number, Firmware revision, features, version numbers, etc.
IDLE IMMEDIATE	N	Send device into Idle power mode
NOP	N	Always fails with an error
PACKET	P	Transfer a SCSI command
READ SECTORS	PIO R	Read data (using PIO)
SET FEATURES	N	Control optional features
SLEEP	N	Send device into Sleep power mode
STANDBY IMMEDIATE	N	Send device into Standby power mode

ATA commands – prohibited for ATAPI devices

- Several commands are not allowed for ATAPI devices

Command	Restriction
DOWNLOAD MICROCODE	Non-packet only
FLUSH CACHE EXT	Non-packet only
MEDIA EJECT, MEDIA LOCK, MEDIA UNLOCK	Non-packet only
READ BUFFER, WRITE BUFFER	Non-packet only
READ DMA, READ DMA EXT, READ DMA QUEUED, READ DMA QUEUED EXT, READ MULTIPLE, READ MULTIPLE EXT, READ SECTORS EXT, READ VERIFY SECTORS, READ VERIFY SECTORS EXT	Non-packet only
SET MULTIPLE MODE	
SMART	Non-packet only
WRITE DMA, WRITE DMA EXT, WRITE DMA FUA EXT, WRITE DMA QUEUED, WRITE DMA QUEUED EXT, WRITE DMA QUEUED FUA EXT, WRITE MULTIPLE, WRITE MULTIPLE EXT, WRITE MULTIPLE FUA EXT, WRITE SECTORS, WRITE SECTORS EXT	

Serial ATA frames

Serial ATA frame format

- Frame Information Sequence (FIS) is the SATA name for a frame
- FIS Type field indicates the type of frame
- PM Port field specifies which device behind a SATA II port multiplier the frame is to or from
- CRC covers the entire frame
- Remember that the frame and CRC are sent between SOF and EOF and are scrambled

Byte	Field(s)	
0	FIS Type	
1	FIS-specific	PM Port
1 to (n-4)	FIS-specific	
(n-3) to n	CRC	

Serial ATA frame types

- Different FISes are used for different types of commands

FIS Type	Size	FIS name	Retry?	Used for
27h	24	Register – Host to Device	Yes	N, PIO, DMA, FPDMA
34h	24	Register – Device to Host	Yes	N, PIO, DMA, FPDMA
A1h	12	Set Device Bits – Device to Host	Yes	DMAQ
5Fh	20	PIO Setup – Device to Host	Yes	PIO, ATAPI
39h	8	DMA Activate – Device to Host	Yes	DMA
41h	32	DMA Setup – Device to Host	Yes	FPDMA
46h	8 to 8196	Data – Device to Host Data – Host to Device	No	PIO, DMA, FPDMA
58h	16	BIST Activate – Device to Host BIST Activate – Host to Device	Yes	BIST

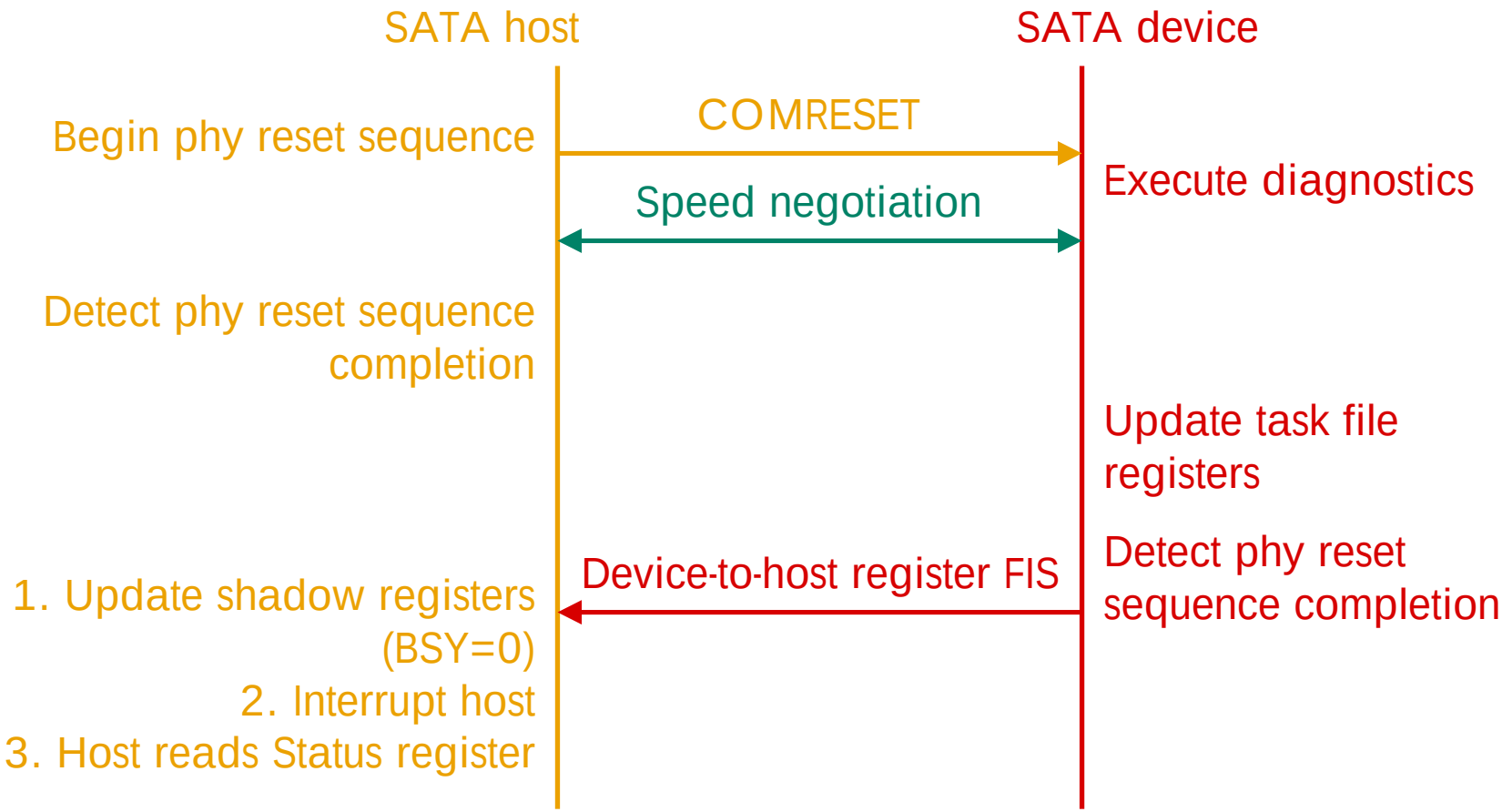
Serial ATA frame sequences

Serial ATA device command layer state machines

- One state machine per type of command controls frame sequences
  - DHR - Power-on and COMRESET
  - DI – Device Idle
  - DSR – Software reset
  - DEDD – Execute Device Diagnostic command
  - DDR – Device Reset command
  - DND – Non-data command
  - DPIOI – PIO data-in command (read)
  - DPIOO – PIO data-out command (write)
  - DDMAI – DMA data in command (read)
  - DDMAO – DMA data out command (write)
  - DDMAQI – DMA queued data-in command (read)
  - DDMAQO – DMA queued data-out command (write)
  - DFPDMAQ – FPDMA queued command (Serial ATA II, read or write)
  - DP – PACKET command (PIO or DMA, read or write)

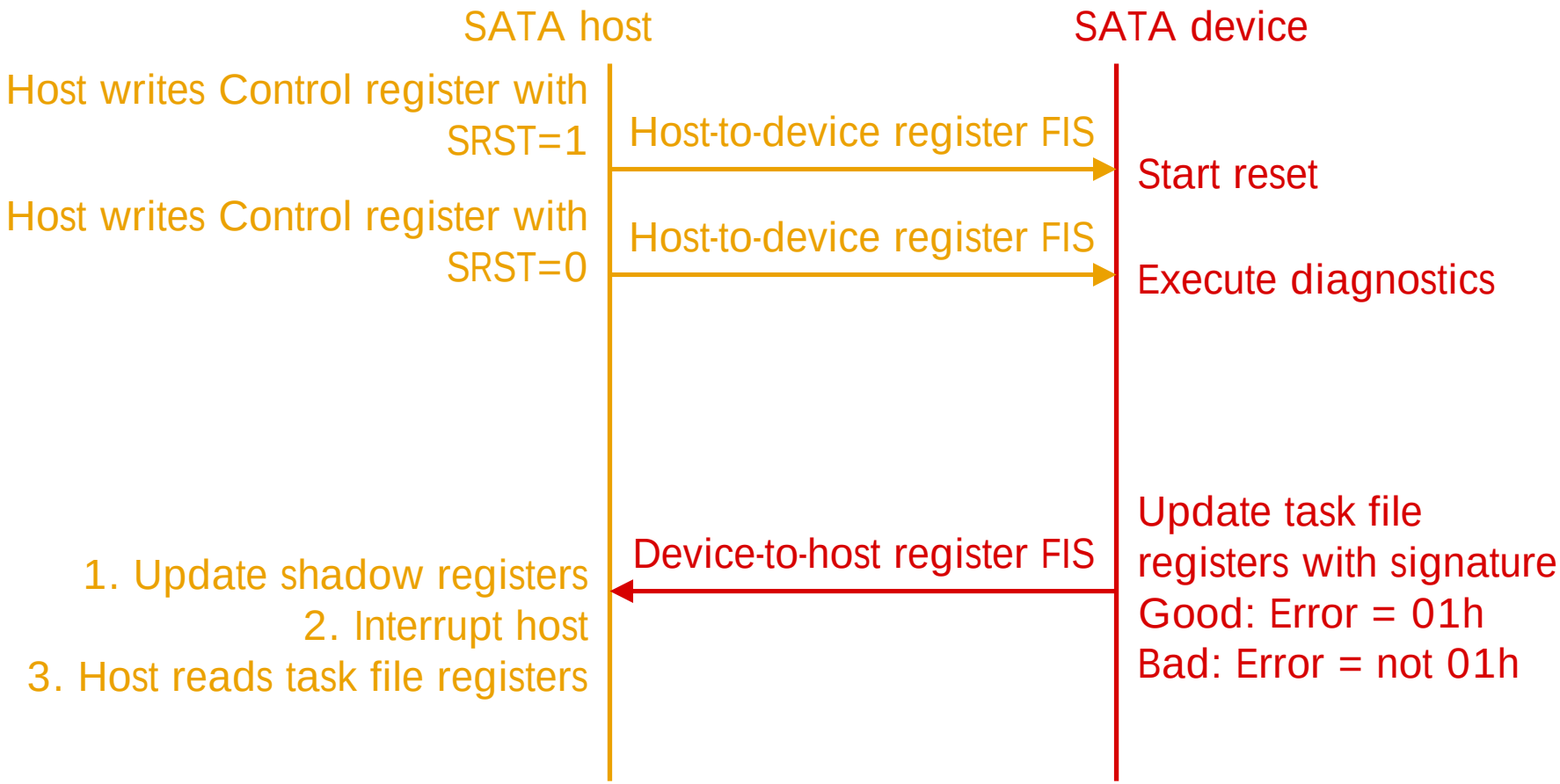
Serial ATA after COMRESET

- After completing a phy reset sequence, device runs diagnostics and transmits a FIS with the results
- SAS expanders store this FIS; read via SMP



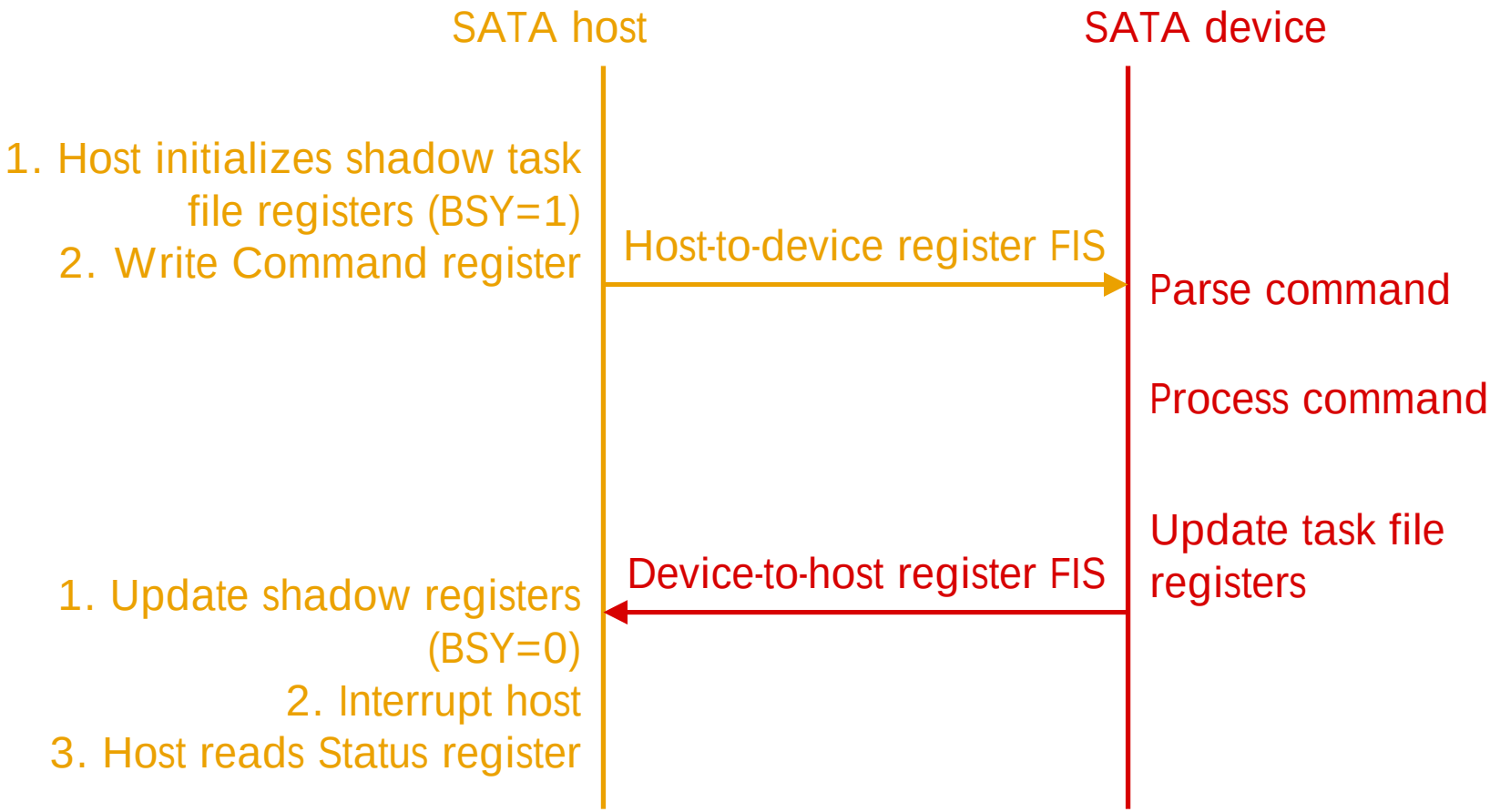
Serial ATA software reset sequence

- Causes diagnostics to be run, returning the appropriate signature

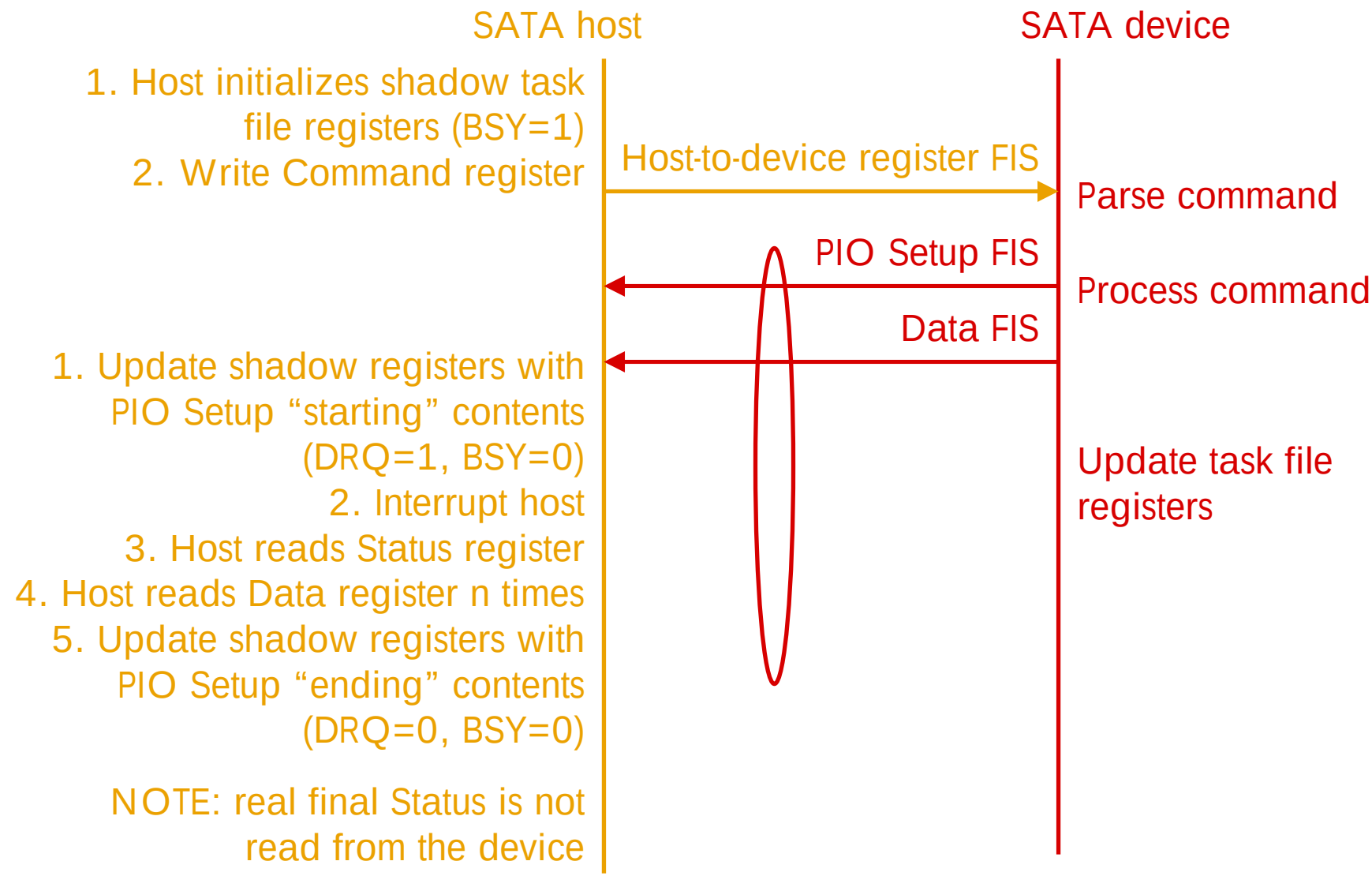


Serial ATA non-data command sequence

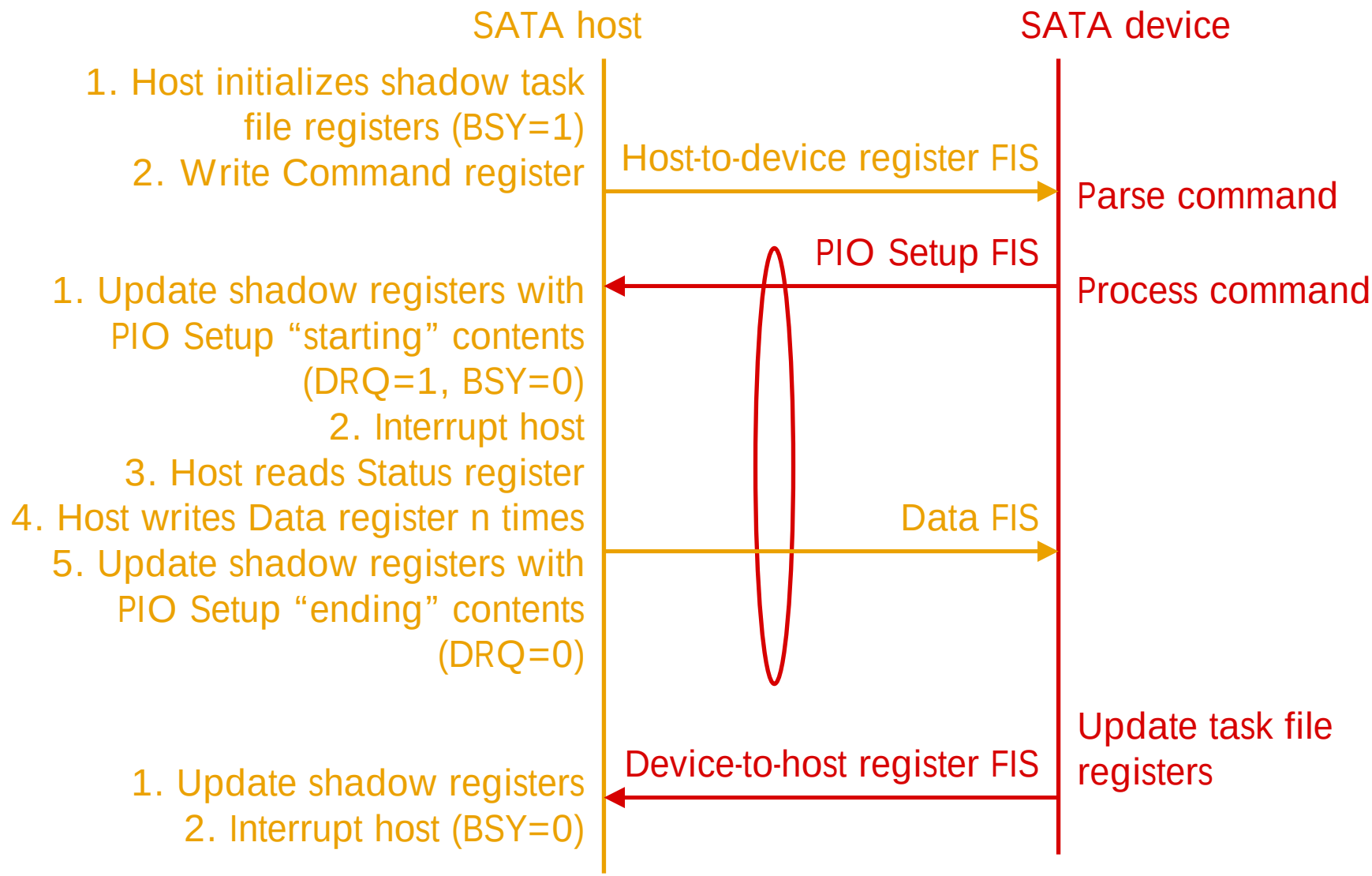
- Non-data commands just exchange Register FISes
- DEVICE RESET and EXECUTE DEVICE DIAGNOSTIC commands use same sequence



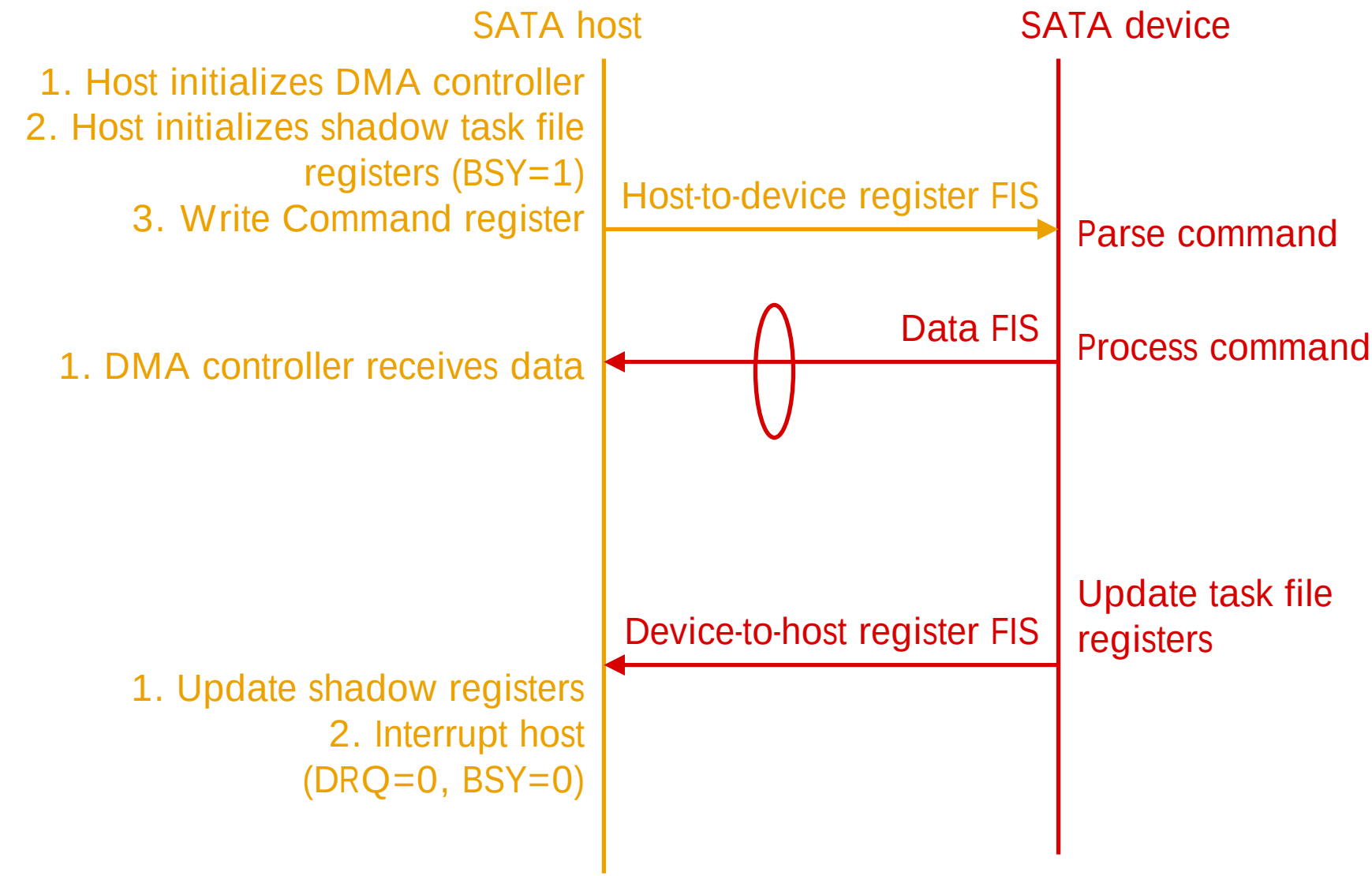
Serial ATA PIO read command sequence



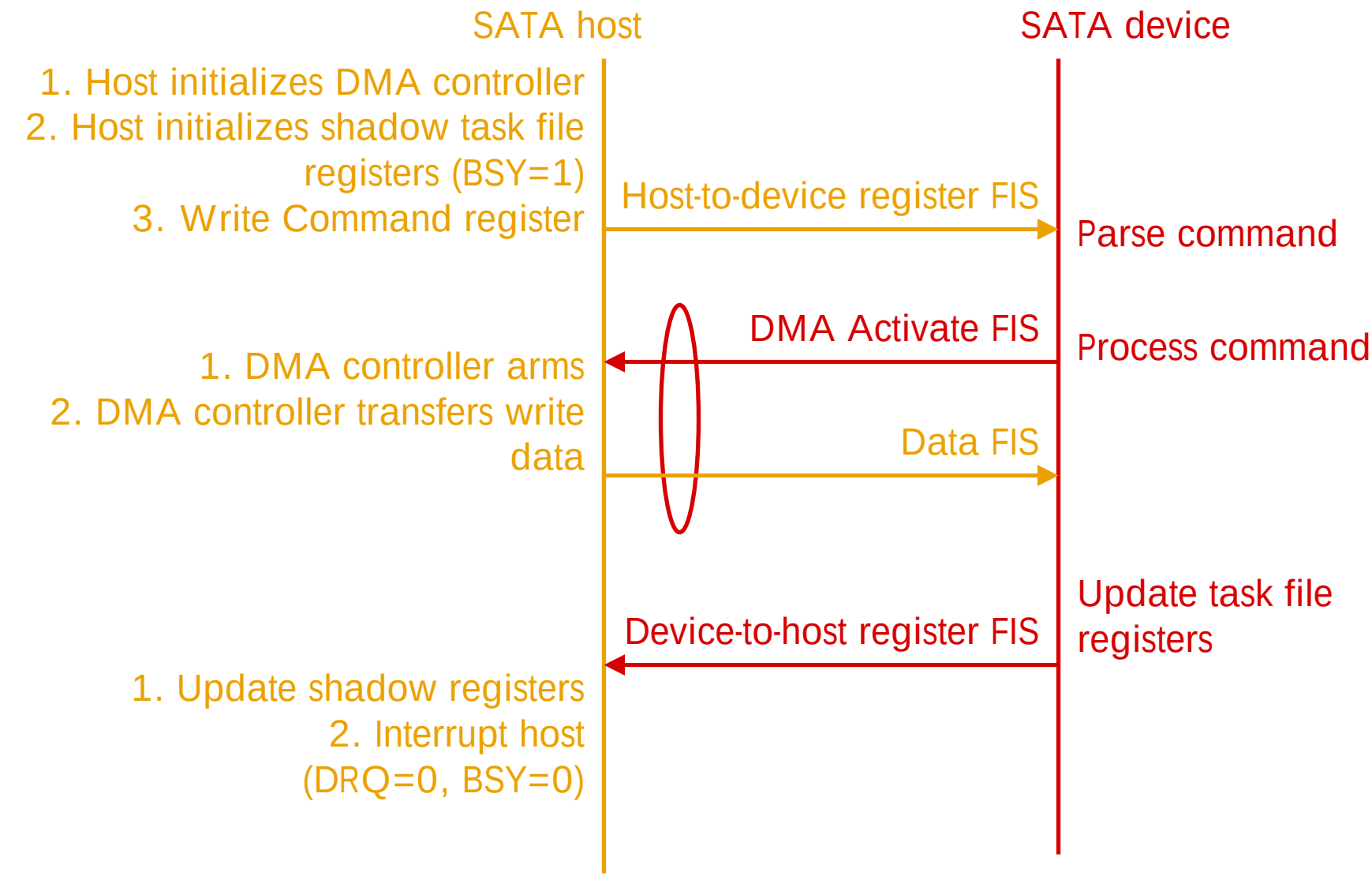
Serial ATA PIO write command sequence



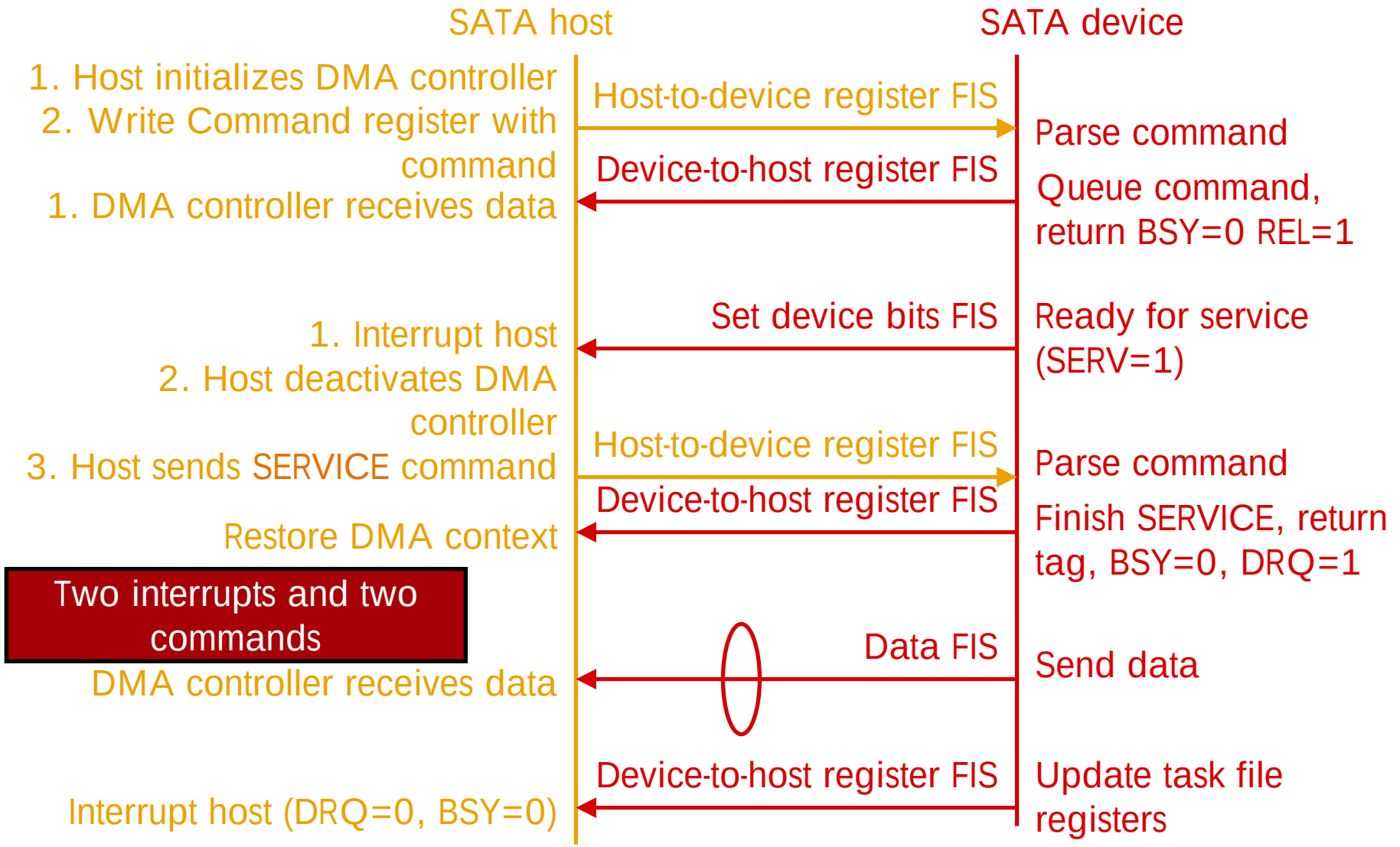
Serial ATA DMA read command sequence



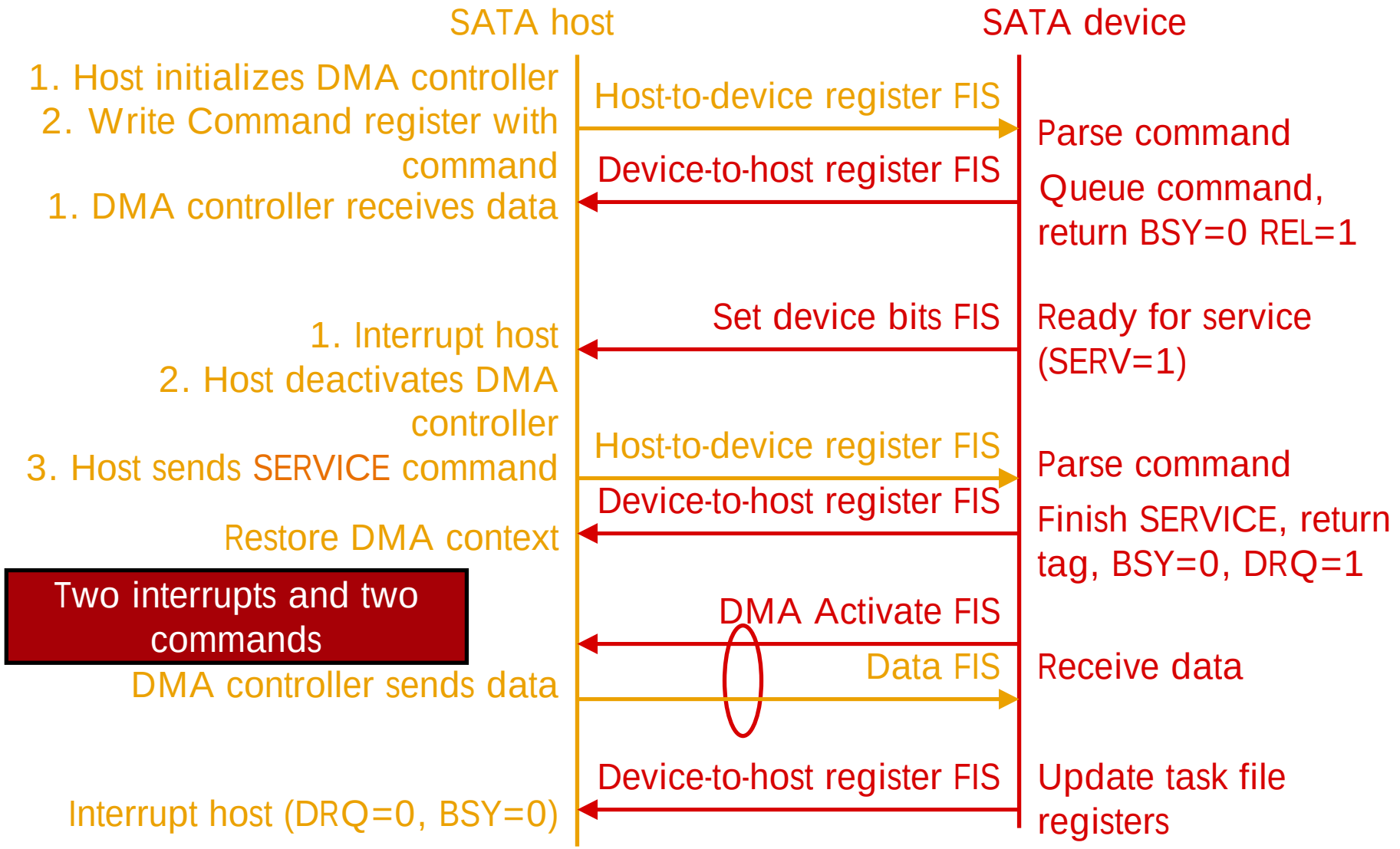
Serial ATA DMA write command sequence



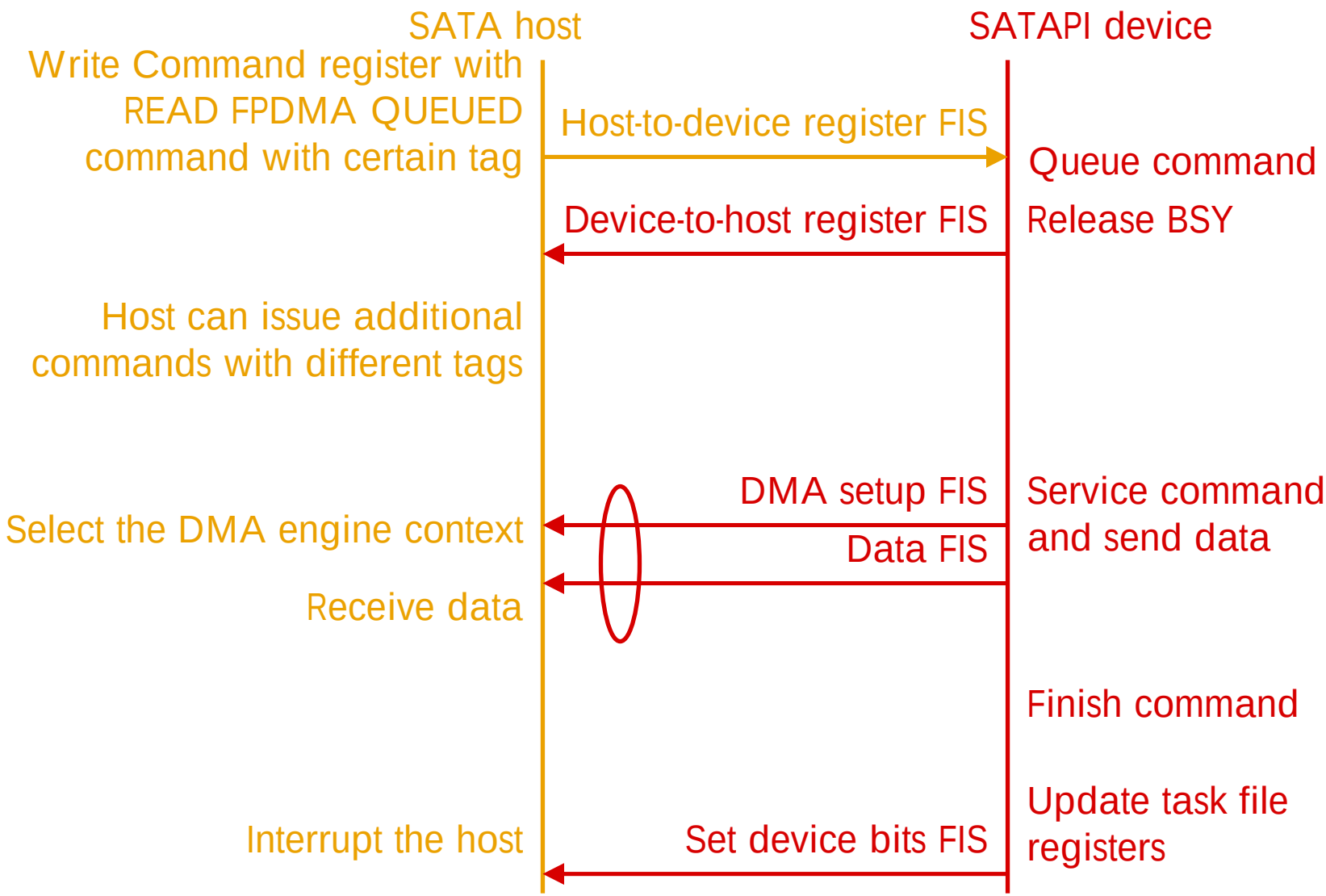
Serial ATA DMAQ read command sequence



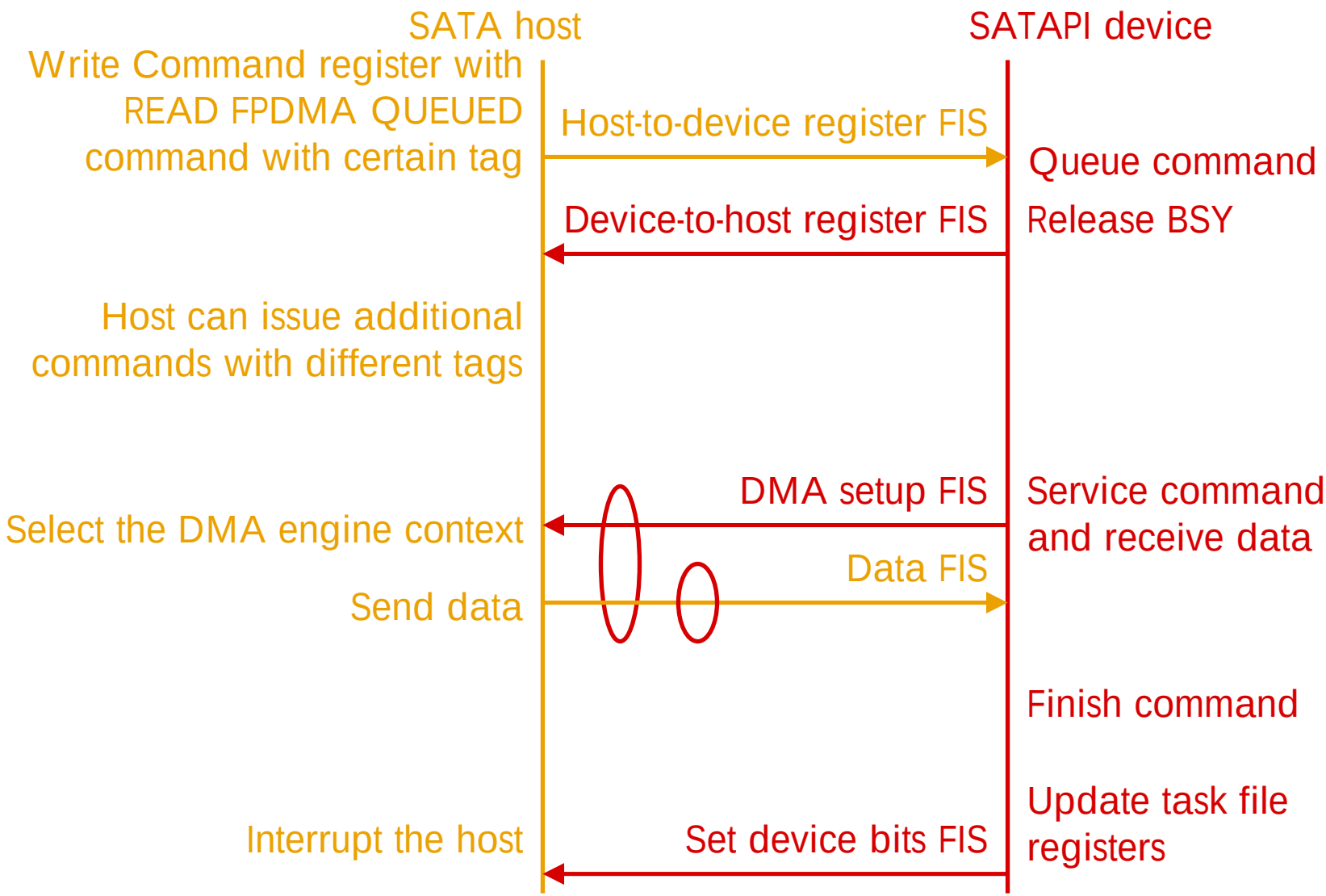
Serial ATA DMAQ write command sequence



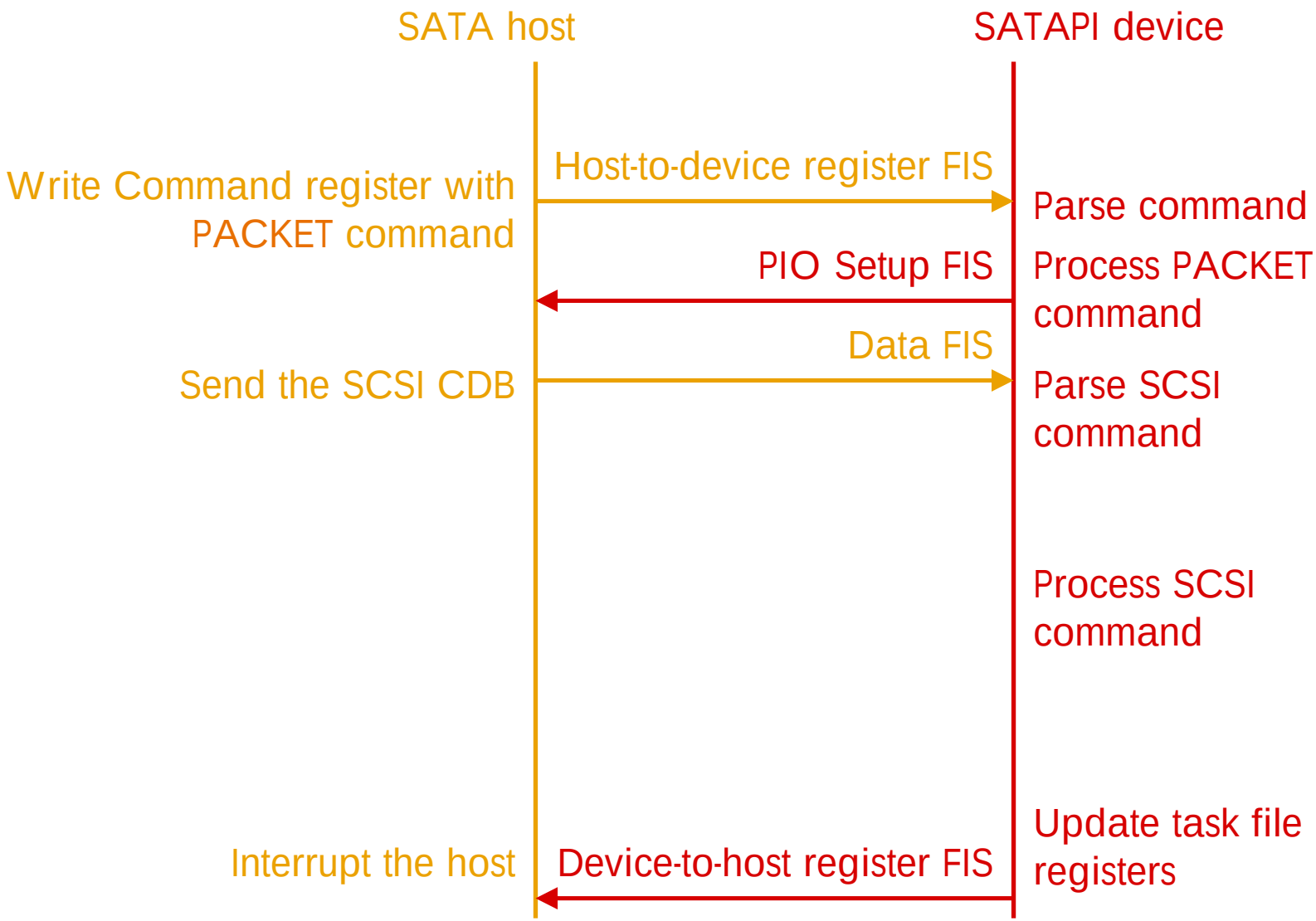
Serial ATA FPDMA read command sequence



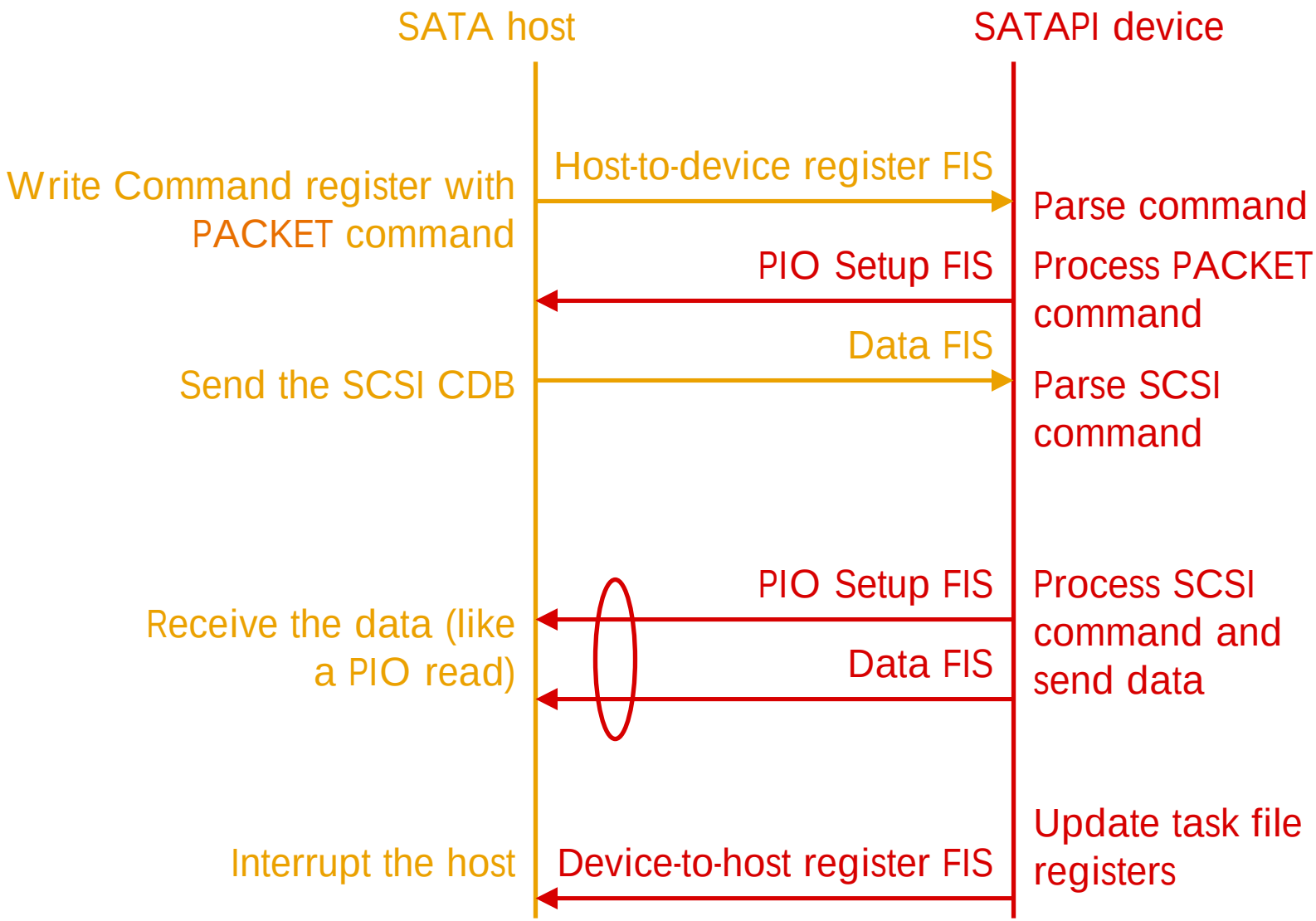
Serial ATA FPDMA write command sequence



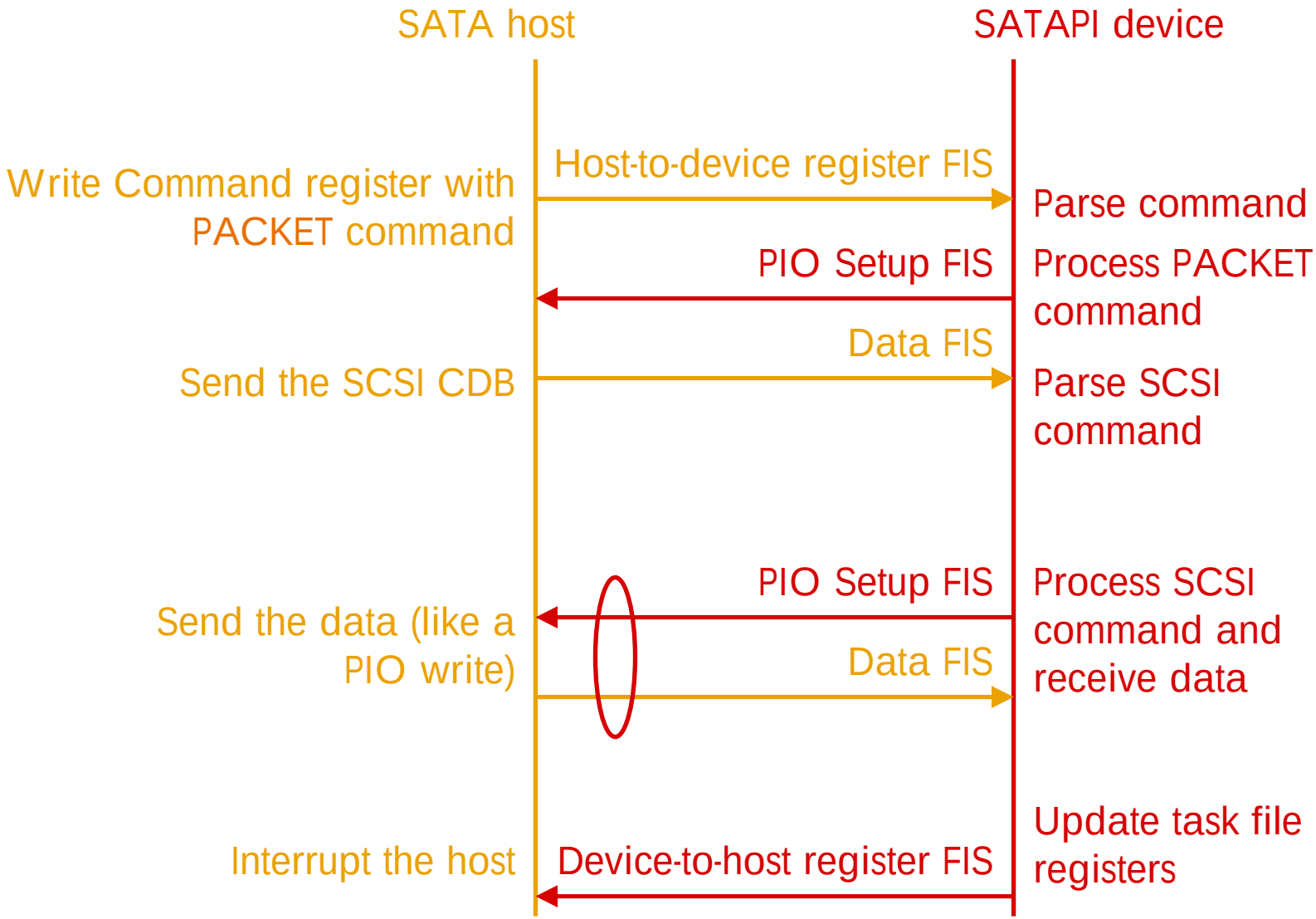
Serial ATAPI PIO non-data command sequence



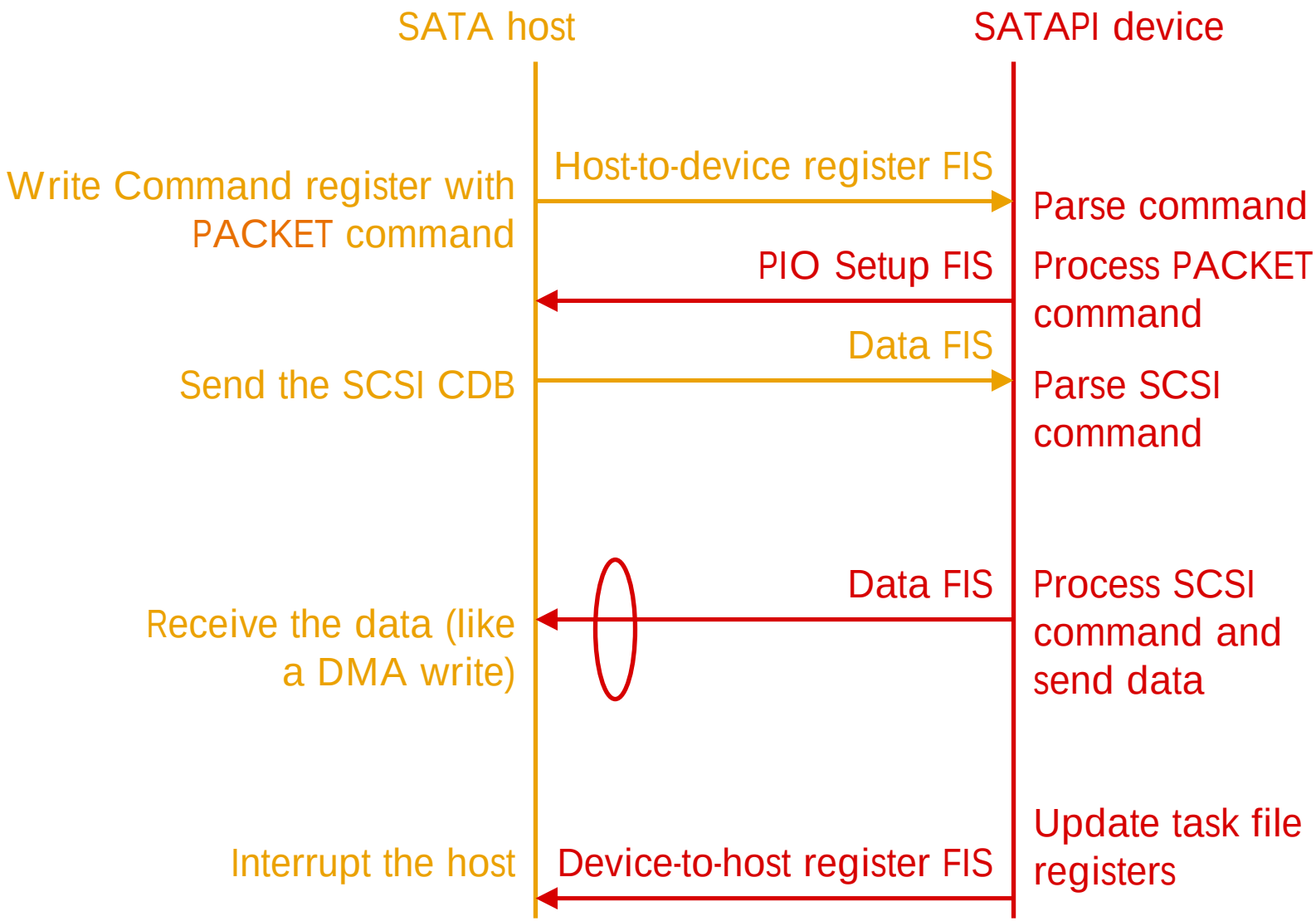
Serial ATAPI PIO read command sequence



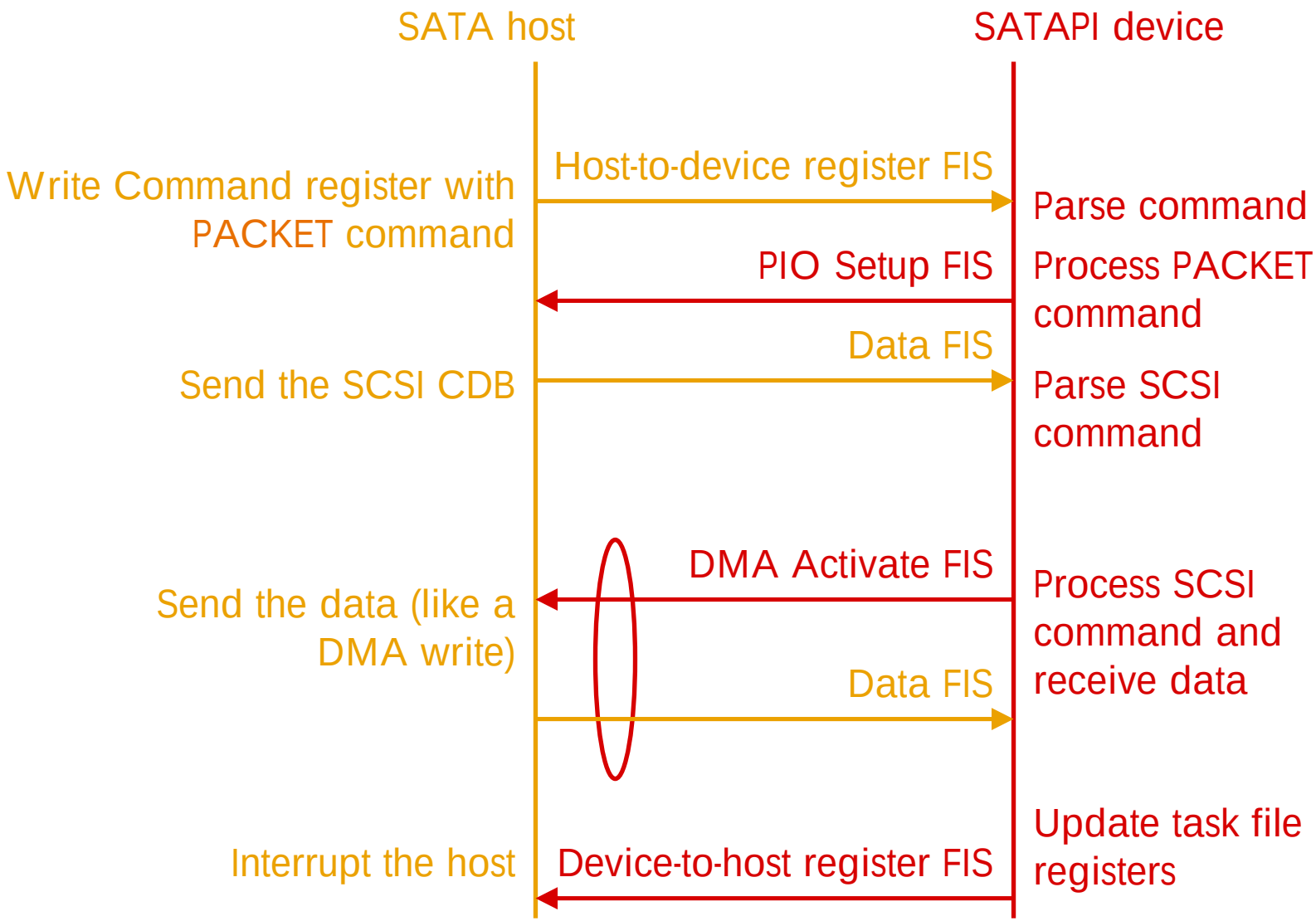
Serial ATAPI PIO write command sequence



Serial ATAPI DMA read command sequence



Serial ATAPI DMA write command sequence



Serial ATA frame contents

SATA Register – Host to Device FIS

- Sends updated task file register contents to the device
- When sent
  - Write to Command register (C=1)
  - Write to Control register, if the values change (C=0)
- Data register not included

Byte	Field(s)				
0	FIS Type (27h)				
1	C	R	R	R	PM Port
2	Command				
3	Features				
4	LBA Low (Sector Number)				
5	LBA Mid (Cylinder Low)				
6	LBA High (Cylinder High)				
7	Device				
8	LBA Low (Sector Number) exp				
9	LBA Mid (Cylinder Low) exp				
10	LBA High (Cylinder High) exp				
11	Features exp				
12	Sector Count				
13	Sector Count (exp)				
14	Reserved				
15	Control				
16 to 19	Reserved				
20 to 23	CRC				

SATA Register – Device to Host FIS

- Sends updated register block to the host
- When sent
  - Power on
  - Command completion
  - For legacy queuing, when device wants to disconnect

Byte	Field(s)				
0	FIS Type (34h)				
1	R	I	R	R	PM Port
2	Status				
3	Error				
4	LBA Low (Sector Number)				
5	LBA Mid (Cylinder Low)				
6	LBA High (Cylinder High)				
7	Device				
8	LBA Low (Sector Number) (exp)				
9	LBA Mid (Cylinder Low) (exp)				
10	LBA High (Cylinder High) (exp)				
11	Reserved				
12	Sector Count				
13	Sector Count (exp)				
14	Reserved				
15	Reserved				
16 to 19	Reserved				
20 to 23	CRC				

SATA Set Device Bits – Device to Host FIS

- Sends updated Error and Status bits to the host
- Does not alter BSY (bit 7) or DRQ (bit 3) of the Status register
- Optional

Byte	Field(s)				
0	FIS Type (A1h)				
1	R	I	R	R	PM Port
2	Status				
3	Error				
4 to 7	SActive				
8 to 11	CRC				

- Used for legacy queuing
  - To set the SERV bit in the Status register for queued commands
- Used for FPDMA native queuing
  - SActive bits indicate which commands are queued and which are completed
  - 32 tags supported

SATA PIO Setup – Device to Host FIS

- Sent to host to request a Data FIS containing PIO write data
- Sent to host before sending a Data FIS containing PIO read data
  - Ending status (E_Status) is returned before the PIO Read data transfer
  - *Serial ATA Storage Architecture and Applications* book by Knut Grimsrud & Hubbert Smith (Intel) recommends not using PIO Reads because of this

Byte	Field(s)				
0	FIS Type (5Fh)				
1	R	I	D	R	PM Port
2	Status				
3	Error				
4	LBA Low (Sector Number)				
5	LBA Mid (Cylinder Low)				
6	LBA High (Cylinder High)				
7	Device				
8	LBA Low (Sector Number) (exp)				
9	LBA Mid (Cylinder Low) (exp)				
10	LBA High (Cylinder High) (exp)				
11	Reserved				
12	Sector Count				
13	Sector Count (exp)				
14	Reserved				
15	E_Status				
16 to 17	Transfer Count				
18 to 19	Reserved				
20 to 23	CRC				

SATA DMA Activate – Device to Host FIS

- Tells the host to start running a DMA transfer of write data
- Must be received prior to each write Data FIS
- When sent
  - When device is ready to start accepting some DMA write data
- Not used on DMA reads
- SATA II Auto-activate bit in DMA Setup FIS eliminates need for this for FPDMA

Byte	Field(s)				
0	FIS Type (39h)				
1	R	R	R	R	PM Port
2 to 3	Reserved				
4 to 7	CRC				

SATA DMA Setup FIS

- Only device-to-host (so far)
- Only for SATA II first party DMA (should be called FPDMA Setup FIS)
- A (auto-activate) bit
  - Device will not send a DMA Activate FIS to throttle write data; host can send Data FIS immediately
- D (direction) bit
  - 1=write, 0=read
- DMA Buffer Identifier fields
  - Bottom 5 bits of Low field carry the Tag (selects the DMA context)
  - All other bits are zero

Byte	Field(s)				
0	FIS Type (41h)				
1	A	I	D	R	PM Port
2 to 3	Reserved				
4 to 7	DMA Buffer Identifier Low				
8 to 11	DMA Buffer Identifier High				
12 to 15	Reserved				
16 to 19	DMA Buffer Offset				
20 to 23	DMA Transfer Count				
24 to 27	Reserved				
28 to 31	CRC				

SATA FPDMA buffer offset modes

- Non-zero buffer offset mode (optional)
  - DMA Buffer Offset can be any value
  - After delivering all data for a DMA Setup FIS, can send another for a different tag
  - Can fill data for a command in any order
    - e.g. end of buffer first, wrapping to start of buffer
  - Requires new context lookup in host for each DMA Setup FIS
    - Might have to rewind a scatter-gather list
- Zero buffer offset mode
  - DMA Buffer Offset must follow previous DMA Buffer Offset + DMA Transfer Count
  - Must deliver all data for a command together
  - No interleaving of different commands' DMA Setup FISes and data FISes

SATA Data FIS

- Bidirectional
- Contains read (device-to-host) or write (host-to-device) data
- Minimum: 4 bytes
- Maximum: 8196 bytes

Byte	Field(s)				
0	FIS Type (46h)				
1	R	R	R	R	PM Port
2 to 3	Reserved				
4 to n	Data				
n to n+3	CRC				

SATA BIST Activate FIS

- Bidirectional
- Place the receiver into a test mode
  - Loopback modes
    - Far-end analog
    - Far-end retimed
    - Near-end
  - Transmit mode

Byte	Field(s)				
0	FIS Type (58h)				
1	R	R	R	R	PM Port
2	Pattern Definition				
3	Reserved				
4 to 7	Data				
8 to 11	Data				
12 to 15	CRC				

Serial ATA transport layer state machines

SATA transport layer state machines

- Separate host and device state machines
  - HT – Host transport layer
  - DT – Device transport layer
- Each is divided in numerous groups of states
- Transport layer builds the frames and has the link layer send them

SATA host transport layer state machines

- General states
  - HTI – idle
- States that transmit frames
  - HTCM – transmit command FIS
  - HTCR – transmit control FIS
  - HTDMASTUP – transmit DMA Setup FIS
  - HTXBIST – transmit BIST Activate FIS
- States that receive (decompose) frames
  - HTR – receive Register FIS
  - HTDB – receive Set Device Bits FIS
  - HTDA – receive DMA Activate FIS (and transmit or receive Data FISes)
  - HTPS – receive PIO Setup FIS (and transmit or receive Data FISes)
  - HTDS – receive DMA Setup FIS
  - HTRBIST – receive BIST Activate FIS

SATA device transport layer state machines

- General states
  - DTI – idle
- States that transmit frames
  - DTR – transmit Register FIS
  - DTDB – transmit Set Device Bits FIS
  - DTPIOSTUP – transmit PIO Setup FIS
  - DTDMAACT – receive DMA Activate FIS
  - DTDMASTUP – transmit DMA Setup FIS
  - DTDATA – transmit Data FIS
  - DTXBIST – transmit BIST Activate FIS
- States that receive (decompose) frames
  - DTCMD – receive Register FIS
  - DTDATAO – receive Data FIS
  - DTSTP – receive DMA Setup FIS
  - DTRBIST – receive BIST Activate FIS

SATA transport layer state machine notes

- Host leaves HTI0:HT_HostIdle on:
  - Writes to Command or Control register (send Register FIS)
  - Receiving a frame
  - Application layer requests
- Device leaves DTI0:DT_DeviceIdle on
  - Receiving a frame
  - Application layer requests

ATAPI (ATA Packet Interface)

ATAPI overview

- PACKET command used to send a SCSI CDB
- When command completes with an error, the Error register contains a few fields with pieces of the SCSI Sense Data
  - Sense Key field
  - EOM (end of medium) bit
  - ILI (incorrect length indication) bit
- Nearly all MMC (multimedia commands) devices are using ATAPI
  - Although PACKET theoretically works with the ATA overlap and queuing features, they are never used
  - Because they have been focused on parallel ATA, to avoid hogging the bus shared with a disk drive, most commands are “immediate” types that complete immediately

ATAPI vs. SCSI part 1

- No logical unit numbers
  - logical unit number 0 assumed
- No SCSI task management functions
  - Just SCSI commands
  - ATA DEVICE RESET command serves as a Logical Unit Reset
- ATA/ATAPI standard supports 16 byte CDBs, but 12 byte implementations are predominant
- No variable length CDBs
- Can only return GOOD or CHECK CONDITION status
  - no BUSY, TASK SET FULL, RESERVATION CONFLICT, etc.
- No autosense
  - Sense Key is returned in the Error register
  - must use REQUEST SENSE command to retrieve complete sense data after receiving a CHECK CONDITION status
- Must still support numerous ATA commands like CHECK POWER MODE

ATAPI vs. SCSI part 2

- No way to transfer odd amounts of data with DMA transfers
  - However, PIO transfers can do this
- Only 32 ATA tags possible
  - Tags and queuing usually not supported
- No ATAPI protocol identifier in SCSI
  - IDENTIFY PACKET DEVICE must be used along with SCSI INQUIRY
- ATAPI devices “shall not check” reserved fields (e.g. in CDBs)
  - SCSI devices "may or may not check"
- Certain tape commands complete immediately
  - unless ATAPI overlap feature is supported (it is usually not)
  - ERASE, LOAD, LOCATE, REWIND, SPACE, and WRITE FILEMARK
  - In normal SCSI targets, they don’t return Status until complete
  - Software must use READ POSITION to check the status of the command

Wrap up

Serial Attached SCSI tutorials

- General overview (~2 hours)
- Detailed multi-part tutorial (~3 days to present):
  - Architecture
  - Physical layer
  - Phy layer
  - Link layer
    - Part 1) Primitives, address frames, connections
    - Part 2) Arbitration fairness, deadlocks and livelocks, rate matching, SSP, STP, and SMP frame transmission
  - Upper layers
    - Part 1) SCSI application and SSP transport layers
    - Part 2) ATA application and STP/SATA transport layers
    - Part 3) Management application and SMP transport layers, plus port layer
  - SAS SSP comparison with Fibre Channel FCP

Key SCSI standards

- Working drafts of SCSI standards are available on http://www.t10.org
- Published through http://www.incits.org
  - Serial Attached SCSI
  - SCSI Architecture Model – 3 (SAM-3)
  - SCSI Primary Commands – 3 (SPC-3)
  - SCSI Block Commands – 2 (SBC-2)
  - SCSI Stream Commands – 2 (SSC-2)
  - SCSI Enclosure Services – 2 (SES-2)
- SAS connector specifications are available on http://www.sffcommittee.org
  - SFF 8482 (internal backplane/drive)
  - SFF 8470 (external 4-wide)
  - SFF 8223, 8224, 8225 (2.5”, 3.5”, 5.25” form factors)
  - SFF 8484 (internal 4-wide)

Key ATA standards

- Working drafts of ATA standards are available on http://www.t13.org
  - Serial ATA 1.0a (output of private WG)
  - ATA/ATAPI-7 Volume 1 (architecture and commands)
  - ATA/ATAPI-7 Volume 3 (Serial ATA standard)
- Serial ATA II specifications are available on http://www.t10.org and http://www.serialata.org
  - Serial ATA II: Extensions to Serial ATA 1.0
  - Serial ATA II: Port Multiplier
  - Serial ATA II: Port Selector
  - Serial ATA II: Cables and Connectors Volume 1

For more information

- International Committee for Information Technology Standards
  - http://www.incits.org
- T10 (SCSI standards)
  - http://www.t10.org
  - Latest SAS working draft
  - T10 reflector for developers
- T13 (ATA standards)
  - http://www.t13.org
  - T13 reflector for developers
- T11 (Fibre Channel standards)
  - http://www.t11.org
- SFF (connectors)
  - http://www.sffcommittee.org
- SCSI Trade Association
  - http://www.scsita.org
- Serial ATA Working Group
  - http://www.serialata.org
- SNIA (Storage Networking Industry Association)
  - http://www.snia.org
- Industry news
  - http://www.infostor.com
  - http://www.byteandswitch.com
  - http://www.wwpi.com
  - http://searchstorage.com
- Training
  - http://www.knowledgetek.com


hp
invent