From: Gerry Houlder, Seagate Technology <gerry.houlder@seagate.com>
Subj: SAS-2.1 SPL: Add low power transceiver options
Date: Dec. 16, 2008

Overview

This proposal adds lower power transceiver conditions to SAS. The intent is to add low power conditions that are similar to or compatible with the low power conditions available in SATA. This should allow an expander design that can work with either SATA or SAS devices in low power condition.

This proposal allows targets, initiators, or expanders to initiate phy power management requests.

SATA uses 4 primitives for phy power management. SAS will also use 4 primitives but will use encodings chosen from the SAS encoding space. The primitives are:

(a) PS_REQ (PARTIAL) to request partial phy power condition (transceiver should recover within 10 us);
(b) PS_REQ (SLUMBER) to request slumber phy power condition (transceiver should recover within 10 msec);
(c) PS_ACK to accept change into the requested phy power condition;
(d) PS_NAK to reject the phy power condition request.

When a phy power condition request is accepted, both ends of the link stop transmitting. Either end may revive the link by transmitting COMWAKE. Both ends are required to remember whether the link is SSP or STP protocol and previously negotiated settings (speed, DFE settings, SSC, multiplexing). The intent is that both ends can restart and obtain synchronization without needing to change receiver settings or redo training. Repeating training would increase the recovery time so much that partial phy power condition may not be feasible.

When an initiator requests an expander to open a link that is in partial phy power condition, recovery time should be fast enough so any delay can be covered by returning AIP (WAITING ON CONNECTION) until the link is ready. When an initiator requests to open a link that is in slumber phy power condition, the expander should use OPEN_REJECT (RETRY) (the recovery time will be several milliseconds and we don’t want to tie up expander pathways for that long by using AIP). I think it is important to NOT pick new primitives for this to maintain compatibility with older initiators that wouldn’t recognize the new primitive.

The partial condition has fast enough recovery so that it may be used by a target-expander link without explicit approval by an initiator. The slumber condition causes enough delay so that an initiator should explicitly allow/ disallow this option based on its performance requirements. There are several possibilities for this:

(a) Use a SAS specific mode page to enable/ disable phy power management in targets.
(b) use an SMP function to enable/ disable phy power management in expanders.

There is a desire to add information to SMP functions to indicate links that are in a low phy power condition. The DISCOVER function and the DISCOVER LIST function are extended to report this.

If a drive is removed while in partial or slumber and another drive is inserted, the drive will transmit a COMINIT at power up and alert the initiator/expander that something has changed since it did not receive a COMWAKE. Similarly, if a drive is pulled and not replaced then a COMWAKE transmitted by an initiator will timeout, so these situations are covered.

SAS-2+ Changes (based on SAS-2 rev.14f):

3.1.x Active phy power management condition: The normal power condition for a phy. The phy is not in a partial or slumber power management condition (see 6.6.5).

3.1.x Partial phy power management condition: A phy low power consumption condition that meets the recovery time requirement for the partial power management condition (see 6.6.5).

3.1.x Slumber phy power management condition: A phy low power consumption condition that meets the recovery time requirement for the slumber power management condition (see 6.6.5).

3.1.x Phy ready state: The condition of a SAS phy or expander phy when its SP state machine is in the SP15:SAS_PHY_Ready state (see 6.8.4.9).

6.6 Out of band (OOB) signals

[Clauses 6.6.1 through 6.6.4 are unchanged.]

6.6.5 Phy power management conditions

During phy power management conditions (partial or slumber), the phy shall transmit D.C. idle. The phy requirements specified in Table 58 and Table 61 apply.

Table new5 defines the maximum recovery times from phy power management conditions.

Table new5 — Phy power management maximum recovery times

Description	Partial phy power management condition	Slumber phy power management condition
Phy wakeup time [a]	10 us	10 ms
[a] See figure new2.		

6.7 Phy reset sequences

[Clauses 6.7.1 through 6.7.5 are unchanged.]

6.7.6 SAS transceiver low power sequences

6.7.6.1 Transition from active to low phy power management condition

Figure new1 shows the sequence to transition from active to low phy power management condition.

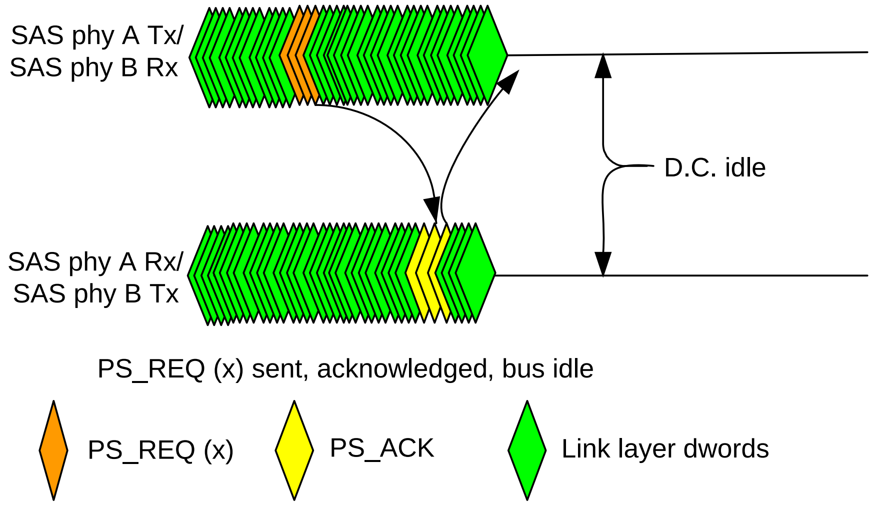

Figure new1 – SAS transition to phy power management

Figure new1 shows a transition from both ends transmitting link layer dwords to D.C. idle. The sequence proceeds as follows:

1) phy A transmits PS_REQ (x) primitive to phy B;
2) phy B transmits PS_ACK primitive to phy A;
3) both phys remember link rate, training settings, SSC setting, and multiplexing setting; and
4) both phys transition to the requested D.C. idle.

6.7.6.2 COMWAKE sequence to recover from low phy power management conditions

The sequence to recover from either partial or slumber phy power management condition is shown in figure new2.

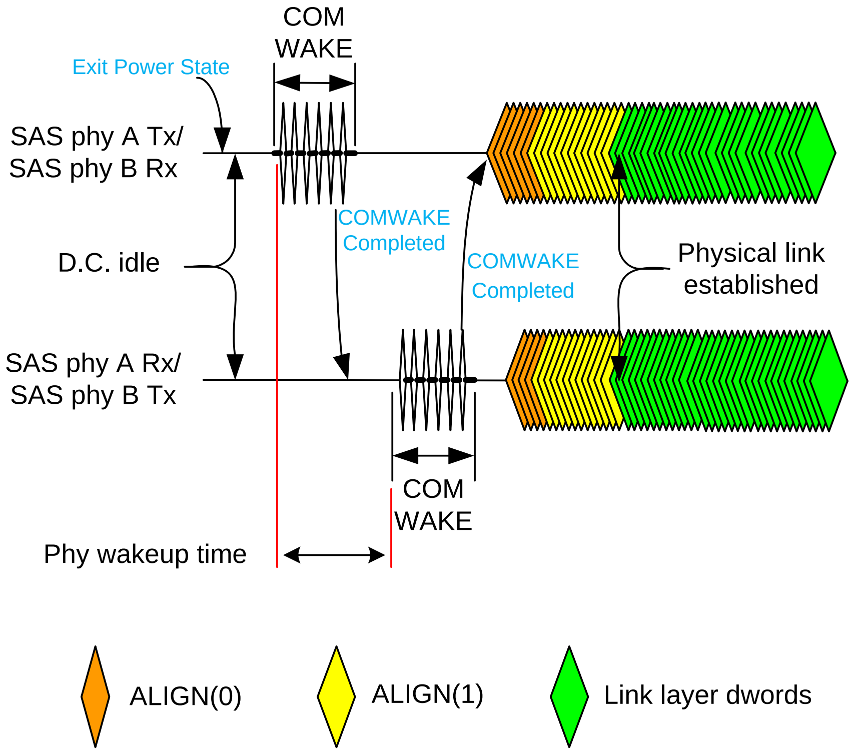

Figure new2 – SAS COMWAKE wakeup sequence

The COMWAKE recovery sequence is as follows:

1) phy A initiates exit power management action by transitioning to active condition;
2) phy A transmits COMWAKE;
3) phy B detects COMWAKE, initiates exit power management action and transmits COMWAKE back to phy A;
4) both phys transmit ALIGN (0) primitives at previously negotiated settings;
5) when each phy receiver synchronizes on ALIGN (0)s, that phy transmitter changes to transmitting ALIGN (1)s;
6) when each phy receiver synchronizes on ALIGN (1)s, that phy transmitter changes to link layer dwords; and
7) if phys are multiplexed, the link layer transmits the multiplexing sequence (see 6.7.4.3) to align the logical links.

The link is now re-established with the same transfer rate and SNW3 settings (e.g., SSC mode and multiplexing mode) that were negotiated before the link was placed in phy power management condition.

6.7.6.3 Recovery from hot plug events during phy power management

Figure new3 shows examples of hot-plug scenarios that may occur during phy power management recovery.

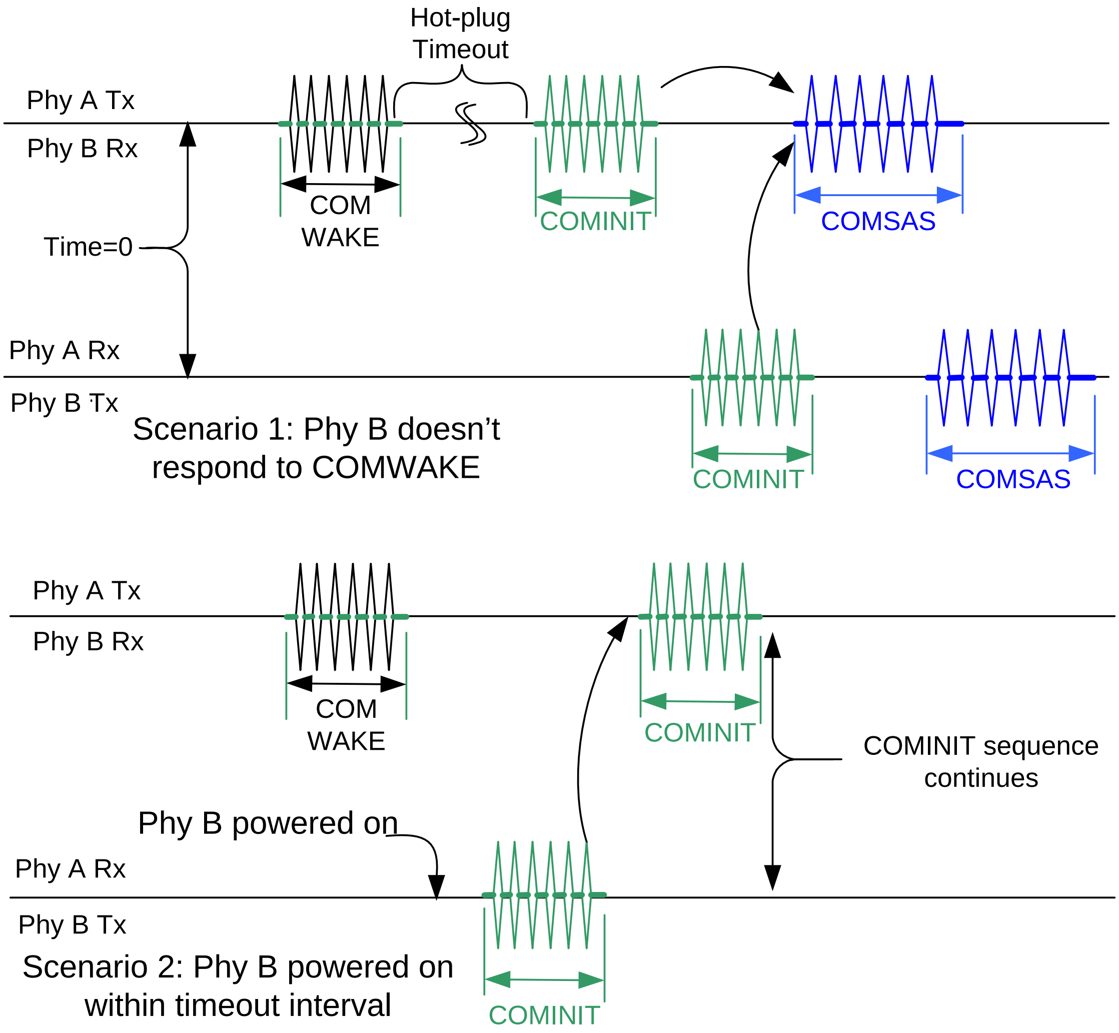

Figure new3 – COMWAKE hot-plug examples

Scenario 1 depicts:

1) phy A starting a phy power management exit request;
2) phy B target doesn't respond within timeout limit; and
3) phy A recovers by changing to COMINIT sequence.

Scenario 2 depicts:

1) phy A starting a phy power management exit request;
2) phy B is powered on and begins COMINIT sequence within the timeout limit; and
3) phy A proceeds by continuing the COMINIT sequence.

6.8 SP (phy layer) state machine

[Changes to this clause are detailed in latest revision of 08-206.]

7 Link layer

7.1 Link layer overview

The link layer defines primitives, address frames, and connections. Link layer state machines interface to the port layer and the phy layer and perform the identification and hard reset sequences, connection management, and SSP, STP, and SMP specific frame transmission and reception.

7.2 Primitives

7.2.1 Primitives overview

[No changes to this clause.]

7.2.2 Primitive summary

Table 112 defines the deletable primitives.
[No changes to table 112.]

Table 113 defines the primitives not specific to the type of connection.
[Add these items to table 113.]

Primitive	Use	from			to			Primitive Sequence Type
		I	E	T	I	E	T	
PS_ACK	NoConn	I	E	T	I	E	T	Extended
PS_NAK	NoConn	I	E	T	I	E	T	Extended
PS_REQ (PARTIAL)	NoConn	I	E	T	I	E	T	Extended
PS_REQ (SLUMBER)	NoConn	I	E	T	I	E	T	Extended

Table 114 defines the primitives used only inside SSP and SMP connections.
[No changes to table 114.]

Table 115 lists the primitives used only inside STP connections and on SATA physical links.
[No changes to table 115.]

7.2.3 Primitive encodings

Table 116 defines the primitive encoding for deletable primitives.
[No changes to table 116.]

Table 117 defines the primitive encoding for primitives not specific to type of connection.
[Note: Add these items to table 117. Need to assign primitive encodings.]

Primitive	character				hexadecimal
	1st	2nd	3rd	4th	
PS_ACK	K28.5	D16.7	D27.4	D30.0	BCF09B1Eh
PS_NAK	K28.5	D24.0	D27.4	D02.0	BC189B02h
PS_REQ (PARTIAL)	K28.5	D07.3	D02.0	D04.7	BC6702E4h
PS_REQ (SLUMBER)	K28.5	D30.0	D24.0	D02.0	BC1E1802h

Table 118 defines the primitive encodings for primitives used only inside SSP and SMP connections.
[No changes to table 118.]

Table 119 lists the primitive encodings for primitives used only inside STP connections and on SATA physical links.
[No changes to table 119.]

7.2.4 Primitive sequences
7.2.5 Deletable primitives
[No changes to the above clauses.]

7.2.6 Primitives not specific to type of connections
7.2.6.1 AIP (Arbitration in progress)
AIP is ~~sent~~ transmitted by an expander device after a connection request to specify that the connection request is being processed and specify the status of the connection request.

The versions of AIP representing different statuses are defined in table 124.

Table 124 — AIP primitives

Primitive	Description
AIP (NORMAL)	Expander device has accepted the connection request. This may be ~~sent~~ transmitted multiple times (see 7.12.4.3).
AIP (RESERVED 0)	Reserved. Processed the same as AIP (NORMAL).
AIP (RESERVED 1)	
AIP (RESERVED 2)	
AIP (WAITING ON CONNECTION)	Expander device has determined the routing for the connection request, but either the destination phys are all being used for connections or there are insufficient routing resources to complete the connection request. This may be ~~sent~~ transmitted multiple times (see 7.12.4.3).
AIP (WAITING ON DEVICE)	Expander device has determined the routing for the connection request and forwarded it to the output physical link. This is ~~sent~~ transmitted one time (see 7.12.4.3).
AIP (WAITING ON PARTIAL)	Expander device has determined the routing for the connection request, but the destination phys are all busy with other partial pathways. This may be ~~sent~~ transmitted multiple times (see 7.12.4.3).
AIP (RESERVED WAITING ON PARTIAL)	Reserved. Processed the same as AIP (WAITING ON PARTIAL).

See 7.12 for details on connections.

7.2.6.2 BREAK
7.2.6.3 BREAK_REPLY
7.2.6.4 BROADCAST
7.2.6.5 CLOSE
7.2.6.6 EOAF (End of address frame)
7.2.6.7 ERROR
7.2.6.8 HARD_RESET
7.2.6.9 OPEN_ACCEPT
[Clauses 7.2.6.2 through 7.2.6.9 are unchanged.]

7.2.6.10 OPEN_REJECT
OPEN_REJECT specifies that a connection request has been rejected and specifies the reason for the rejection. The result of some OPEN_REJECTs is to abandon (i.e., not retry) the connection request and the result of other OPEN_REJECTs is to retry the connection request.

All of the OPEN_REJECT versions defined in table 127 shall result in the originating port abandoning the connection request.

Table 127 — Abandon-class OPEN_REJECT primitives

[Table 122 is unchanged.]

All of the OPEN_REJECT versions defined in table 128 shall result in the originating port retrying the connection request.

Table 128 — Retry-class OPEN_REJECT primitives

Primitive	originator	Description
OPEN_REJECT (NO DESTINATION) [a]	Expander phy	An expander device in the pathway is not configuring and determines that: a) there is no such destination phy; b) the connection request routes to a destination expander phy in the same expander port as the source expander phy and the expander port is using the subtractive routing method; or c) the SAS address is valid for an STP target port in an STP/SATA bridge, but the initial Register - Device to Host FIS has not been successfully received (see 10.4.3.12).
OPEN_REJECT (PATHWAY BLOCKED) [b]	Expander phy	An expander device determined the pathway was blocked by higher priority connection requests.
OPEN_REJECT (RESERVED CONTINUE 0) [c]	Unknown	Reserved. Process the same as OPEN_REJECT (RETRY).
OPEN_REJECT (RESERVED CONTINUE 1) [c]		
OPEN_REJECT (RESERVED INITIALIZE 0) [a]	Unknown	Reserved. Process the same as OPEN_REJECT (NO DESTINATION).
OPEN_REJECT (RESERVED INITIALIZE 1) [a]		
OPEN_REJECT (RESERVED STOP 0) [b]	Unknown	Reserved. Process the same as OPEN_REJECT (PATHWAY BLOCKED).
OPEN_REJECT (RESERVED STOP 1) [b]		
OPEN_REJECT (RETRY)	Destination phy or zoning expander phy	Either: a) a phy with destination SAS address exists but is temporarily not able to accept connections (see 7.16.1, 7.17.5, and 7.18.3); b) an expander device in the pathway is configuring and would otherwise have returned OPEN_REJECT (NO DESTINATION)(see 4.7.2 and 7.12.4.2.5); c) an expander device in the pathway is locked and would otherwise have returned OPEN_REJECT (ZONE VIOLATION)(see 4.9.3.5 and 7.12.4.2.5); ~~or~~ d) an expander device in the pathway has reduced functionality (see 4.6.8 and 7.12.4.2.5); or e) a phy with destination SAS address exists but is

		in the slumber phy power management condition (see 7.x).

a If the I_T Nexus Loss timer is already running, it continues running; if it is not already running, it is initialized and started. Stop retrying the connection request if the I_T Nexus Loss timer expires.
b If the I_T Nexus Loss timer is already running, it continues running. Stop retrying the connection request if the I_T Nexus Loss timer expires.
c If the I_T Nexus Loss timer (see 8.2.2) is already running, it is stopped.

NOTE 49 - Some SAS logical phys compliant with earlier versions of this standard also transmit OPEN_REJECT (RETRY) if they receive an OPEN address frame while their SL_CC state machines are in the SL_CC5:BreakWait state (see 7.14.4.7).

7.2.6.11 SOAF (Start of address frame)

7.2.6.12 TRAIN

7.2.6.13 TRAIN_DONE

[Clauses 7.2.6.11 through 7.2.6.13 are unchanged.]

7.2.6.14 PS_ACK

PS_ACK specifies the positive acknowledgement of a phy power management request primitive.

See 7.10 for details of phy power management.

7.2.6.15 PS_NAK

PS_NAK specifies the negative acknowledgement of a phy power management request primitive.

See 7.10 for details of phy power management.

7.2.6.16 PS_REQ

PS_REQ specifies a request to transition to a phy power management condition. All the versions defined in table new6 request a specified level of phy power savings.

Table new6 — PS_REQ primitives

Primitive	originator	Description
PS_REQ (PARTIAL)	Expander phy or SAS phy	Request to enter partial phy power condition (see x.x).
PS_REQ (SLUMBER)	Expander phy or SAS phy	Request to enter slumber phy power condition (see x.x).

7.2.7 Primitives used only inside SSP and SMP connections

7.2.8 Primitives used only inside STP connections and on SATA physical links

[Clauses 7.2.7 and 7.2.8 are unchanged.]

7.3 Physical link rate tolerance management

7.4 Idle physical links

7.5 CRC

7.6 Scrambling

7.7 Bit order of CRC and scrambler

[Clauses 7.3 through 7.7 are unchanged.]

7.8 Address frames

[Changes to clauses 7.8 and 7.9 are described in latest revision of 08-249.]

7.10 Power management

SAS phy power management may be supported on SAS phys and expander phys. Phy power management is only allowed outside of connections.

Phy power management is enabled in SAS target devices using the SAS Protocol Specific mode page (see 10.2.7.4). Phy power management is enabled is SAS expander devices using the SMP PHY CONTROL function (see 10.4.3.28).

If phy power management is enabled and the received IDENTIFY address frame has the PARTIAL CAPABLE bit set to one (see 7.8.2), then SAS phys and expander phys may generate PS_REQ (PARTIAL). If phy power management is enabled and the received IDENTIFY address frame has the SLUMBER CAPABLE bit set to one (see 7.8.2), then SAS phys and expander phys may generate PS_REQ (SLUMBER). If phy power management is enabled, then SAS phys and expander phys may reply with PS_ACK to accept a phy power management request. If phy power management is disabled, SAS phys and expander phys shall reject a phy power management request by replying with PS_NAK.

If an expander phy is in partial phy power management condition and the expander device receives a connection request routed to that expander phy, then the expander device initiates the exit power management procedure (see 7.x) on that expander phy and responds with AIP (NORMAL) until the OPEN address frame is transmitted to the expander phy.

If an expander phy is in slumber phy power management condition and the expander device receives a connection request routed to that expander phy, then the expander device initiates the exit power management procedure (see 7.x) on that expander phy and responds with OPEN REJECT (RETRY) until a phy ready state (see 6.8.4.9) is established with that expander phy.

If an expander phy is in partial or slumber phy power management condition and the BPP is requested to forward a Broadcast to that expander phy, then the BPP initiates the exit power management procedure (see 7.x) on that expander phy, forwards the Broadcast, and initiates the procedure to return that phy to its previous phy power management condition.

If a SAS initiator phy or expander phy is in partial or slumber phy power management condition and that phy is requested to transmit a NOTIFY, then the NOTIFY is not transmitted and that phy remains in the same phy power management condition.

If a SAS target device is in a SA_PC state that requires receipt of a NOTIFY (ENABLE SPINUP) to transition from that state (see 10.2.10), then the SAS target device shall not generate PS_REQ (PARTIAL) or PS_REQ (SLUMBER) requests and shall not accept PS_REQ (PARTIAL) or PS_REQ (SLUMBER) requests until a NOTIFY (ENABLE SPINUP) is received.

[Note: Do we need more guidance on when partial or slumber conditions should be invoked and what responses should result?]

SATA interface power management is not supported in STP.

STP initiator ports shall not generate SATA_PMREQ_P, SATA_PMREQ_S, or SATA_PMACK. If an STP initiator port receives SATA_PMREQ_P or SATA_PMREQ_S, it shall reply with SATA_PMNAK.

If an expander device receives SATA_PMREQ_P or SATA_PMREQ_S from a SATA device while an STP connection is not open, it shall not forward it to any STP initiator port and shall reply with SATA_PMNAK. If one of these primitives arrives while an STP connection is open, it may forward the primitive to the STP initiator port.

SCSI idle and standby power conditions, implemented with the START STOP UNIT command (see SBC-3) and the Power Condition mode page (see SPC-4), may be supported by SSP initiator ports and SSP target ports as described in 10.2.10.

ATA idle and standby power modes, implemented with the IDLE, IDLE IMMEDIATE, STANDBY, STANDBY IMMEDIATE, and CHECK POWER MODE commands (see ATA8-ACS), may be supported by STP initiator ports. The ATA sleep power mode, implemented with the SLEEP command, shall not be used.

7.11 SAS domain changes (Broadcast (Change) usage)

[Changes to clauses 7.11 through 7.18 are in latest revision of 08-249.]

10.2.7.7 Enhanced Phy Control mode page

The Enhanced Phy Control mode page contains parameters that affect SSP target phy operation. If the mode page is implemented by one logical unit in a SCSI target device, then it shall be implemented by all logical units in the SCSI target device that support the MODE SELECT or MODE SENSE commands.

The mode page policy (see SPC-4) for this mode page shall be shared.

Table 227 defines the format of this mode page.

[Note: Table 227 and following text is unchanged and not shown here.]

Table 228 defines the enhanced phy control mode descriptor.

Table 228 — Enhanced phy control mode descriptor

Byte \ Bit	7	6	5	4	3	2	1	0
0	Reserved							
1	PHY IDENTIFIER							
2	(MSB)	DESCRIPTOR LENGTH (0010h)						
3								(LSB)
4	PROGRAMMED PHY CAPABILITIES							
7								
8	CURRENT PHY CAPABILITIES							
11								
12	ATTACHED PHY CAPABILITIES							
15								
16	Reserved							
17								
18	Reserved			NEGOTIATED SSC	NEGOTIATED PHYSICAL LINK RATE			
19	Reserved					ENABLE SLUMBER	ENABLE PARTIAL	HARDW MUXING SUPPORTD

The DESCRIPTOR LENGTH field contains the length in bytes that follow in the descriptor and shall be set to the value defined in table 228.

An ENABLE SLUMBER bit set to one specifies that the device server shall enable support for slumber phy power management (see 7.10). An ENABLE SLUMBER bit set to zero specifies that the device server shall disable support for slumber phy power management.

An ENABLE PARTIAL bit set to one specifies that the device server shall enable support for partial phy power management (see 7.10). An ENABLE PARTIAL bit set to zero specifies that the device server shall disable support for partial phy power management.

The fields in the enhanced phy control mode descriptor not defined in this subclause are defined in the SMP DISCOVER response (see 10.4.3.10). These fields shall not be changeable with the MODE SELECT command.

[Changes to clause 10.4.3 SMP functions are defined in 08-420.]

Annex K (informative)
Primitive encoding

[Note: Update table K.1 to include coding of the PS_REQ (PARTIAL), PS_REQ (SLUMBER), PS_ACK, and PS_NAK primitives.]