To: T10 Technical Committee
From: Rob Elliott, HP (elliott@hp.com)
Date: 13 October 2002
Subject: 02-360r1 SAS spinup

Revision History
Revision 0 (10 September 2002) First revision
Revision 1 (13 October 2002) Includes standardese and converted to FrameMaker from Word. Ties SPINUP to any power condition change, not just START STOP UNIT which is block device specific.

Related Documents
sas-r02a - Serial Attached SCSI revision 2a
spi5r02 - SCSI Parallel Interface 5 revision 2 (SCA-2 connector features)
sbc2r07 - SCSI Block Commands 2 revision 7 (START STOP UNIT and power conditions)
spc3r09 - SCSI Primary Commands 3 revision 9 (power conditions)
rbc-r10a - Reduced Block Commands revision 10a (START STOP UNIT and power conditions)

Overview
Parallel SCSI disk drives offer a variety of mechanisms to control spin-up, all controlled by the RMT_START and DLYD_START pins on the SCA-2 connector. Defined behaviors are:

a) pin-up automatically after power on;
b) pin-up automatically after power on after delaying for (up to 12 seconds) * (the SCSI ID assigned by the SEL_ID pins). This means that a drive with SCSI ID 0 powers on immediately, while a drive with SCSI ID 7 waits up to 84 seconds;
c) spin-up under software control with START STOP UNIT.

The Serial Attached SCSI connector lacks the pins to replicate these features, and SAS does not have enclosure-assigned addresses on which to base any form of delayed spin-up. Therefore, only START STOP UNIT is currently supported.

Serial ATA disk drives, in contrast, will spin-up automatically after power on (specifically, after the phy reset sequence completes).

Enclosures are often designed with power supplies that cannot tolerate all their disk drives spinning up at the same time.

Problem
Serial Attached SCSI offers the ability to connect more initiators and more drives than parallel SCSI. The initiators are not guaranteed to have software coordinating access to the drives; e.g. they could be separate servers each using a dedicated disk drive for booting. With software control of spin-up, they could all choose to spin-up drives at the same time (e.g. if they all boot simultaneously).

To avoid more drives spinning up than an enclosure supports, the enclosure requires some control of spin-up sequencing.

Even after initial power on, independent initiators are not prevented from causing simultaneous spin-up - they could all issue START STOP UNIT requesting a START at the same time. It would be very difficult for the enclosure to intercept these SCSI commands and delay them somehow.

Suggested Changes
Create a new SPINUP primitive. Immediately after power on, it is used to trigger automatic spin-up; afterwards, it interacts with the START STOP UNIT command to delay when software-requested spin-up actually occurs.

3.x Normative references
[Add SBC-2 and RBC since START STOP UNIT is referenced]

7 Link layer

7.1 Primitives
7.1.2 Primitives summary
[Add SPINUP; Use = All, From = I, E; To = T; Type = Single]

7.1.4 SAS primitives
7.1.4.n SPINUP

SPINUP is transmitted by an initiator port or expander port and is used to indicate to a target device that it may temporarily consume excess power (e.g. while spinning-up rotating media). The length of time the target device consumes excess power and the amount of excess power is vendor-specific. SPINUP shall interact with the device's power condition state, controlled by the Power Conditions mode page (see SPC-3) and/or the START STOP UNIT command (see SBC-2 and RBC), as described in 10.3.1.

Initiator devices and expander devices shall transmit SPINUPs while attached to target devices (i.e., devices that report any target protocol support in their IDENTIFY address frames). They shall transmit one SPINUP after power on when the enclosure is ready for initial target device spin-up. After the initial SPINUP, they shall transmit SPINUP periodically.

SPINUP shall be sent in place of an ALIGN. It may or may not affect the ALIGN sequencing (i.e.., rotation through ALIGN(0), ALIGN(1), ALIGN(2), or ALIGN(3)). SPINUP shall not be transmitted until at least three ALIGNs have been transmitted since the previous SPINUP. Otherwise, the selection of when and how often to transmit SPINUP is vendor-specific. If an initiator device or expander device has no vendor-specific controls, it shall default to transmitting SPINUP on every phy at least once every 1 ms.

NOTE 1 The initiator device or expander device uses SPINUP to avoid exceeding enclosure power supply capabilities during spin-up of multiple target devices. It may choose to rotate transmitting SPINUP across all of its ports, distributing it to N ports at a time if the enclosure power supply is capable of powering N target devices spinning up at a time. An expander device may allow this timing to be configured by a NVROM programming with enclosure-specific sequencing patterns, or may employ more complex, dynamic interaction with the enclosure power supply.

NOTE 2 SPINUP should be transmitted as frequently as possible to avoid incurring application layer timeouts.

SPINUP shall not be forwarded through expander devices.

I_T nexus loss, logical unit reset, and hard reset shall not cause a target device to spin-up automatically on receipt of SPINUP.

Target devices with multiple target ports shall accept SPINUPs from all target ports (e.g., SPINUP received on target port A serves as a wakeup for a START STOP UNIT command received through target port B).

10 SCSI layer
10.3 Power conditions
The logical unit power condition states from the Power Condition mode page (see SPC-3) and START STOP UNIT command (see SBC-2 and RBC), if implemented, shall interact with the SPINUP primitive (see 7.x) to control temporary consumption of extra power (e.g., spin-up of rotating media) as described in this subclause.
:
a) after power on, if the target device has not received a START STOP UNIT command with the START bit set to zero, spin-up automatically after receiving SPINUP. This automatically spins-up the media after power on without waiting for the application client; and
b) after power on, if the target device has previously received a START STOP UNIT command with the START bit set to zero when it receives a START STOP UNIT command with the START bit set to one, spin-up after receiving the next SPINUP. The application client's spin-up request is effectively delayed until SPINUP arrives.

The SCSI application layer Power Condition (SA_PC) state machine describes how the target device processes logical unit power condition state change requests and SPINUP. The SA_PC state machine is an enhanced version of the logical unit power condition state machines described in SPC-3, SBC-2, and RBC. The SA_PC states are as follows:

a) SA_PC_0:Powered_on;
b) SA_PC_1:Active;

c) SA_PC_2:Idle;
d) SA_PC_3:Standby;
e) SA_PC_4:Stopped (specific to SBC-2 and RBC devices);
f) SA_PC_5:Active_Wait (specific to SAS devices); and
g) SA_PC_6:Idle_Wait (specific to SAS devices);

The SA_PC state machine stall start in the SA_PC_0:Powered_on state after power on.

Figure xx describes the SA_PC state machine.

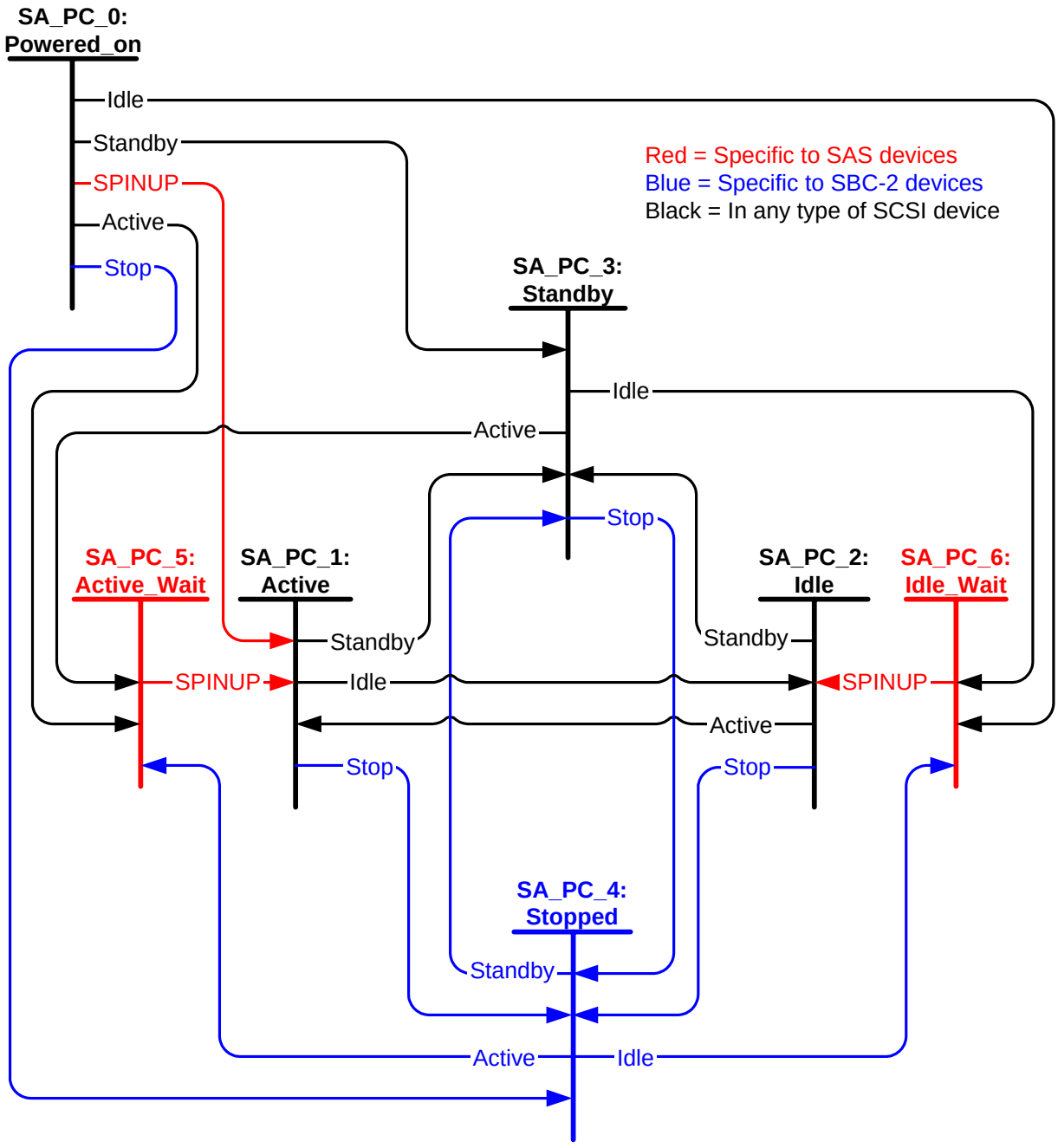

Figure 1 — SCSI application layer power condition (SA_PC) state machine

0.0.1 SA_PC_0:Powered_on state

0.0.1.1 State description

This state shall be entered upon power on.

0.0.1.2 Transition SA_PC_0:Powered_on to SA_PC_1:Active

This transition shall occur if a SPINUP is received.

0.0.1.3 Transition SA_PC_0:Powered_on to SA_PC_3:Standby

This transition shall occur if:

a) a START STOP UNIT command with the POWER CONDITION field set to STANDBY is received;
b) a START STOP UNIT command with the POWER CONDITION field set to FORCE STANDBY is received; or
c) the Power Condition mode page indicates the device shall start in the standby power condition.

0.0.1.4 Transition SA_PC_0:Powered_on to SA_PC_4:Stopped

This transition shall occur if a START STOP UNIT command with the START bit set to zero is received.

0.0.1.5 Transition SA_PC_0:Powered_on to SA_PC_5:Active_Wait

This transition shall occur if:

a) a START STOP UNIT command with the START bit set to one is received; or
b) a START STOP UNIT command with the POWER CONDITION field set to ACTIVE is received.

0.0.1.6 Transition SA_PC_0:Powered_on to SA_PC_6:Idle_Wait

This transition shall occur if:

a) a START STOP UNIT command with the POWER CONDITION field set to IDLE is received;
b) a START STOP UNIT command with the POWER CONDITION field set to FORCE IDLE is received; or
c) the Power Condition mode page indicates the device shall start in the idle power condition.

0.0.2 SA_PC_1:Active state

0.0.2.1 State description

While in this state, rotating media in block devices shall be active.

0.0.2.2 Transition SA_PC_1:Active to SA_PC_2:Idle

This transition shall occur if:

a) a START STOP UNIT command with the POWER CONDITION field set to IDLE is received;
b) a START STOP UNIT command with the POWER CONDITION field set to FORCE IDLE is received; or
c) the Power Condition mode page idle timer expires.

0.0.2.3 Transition SA_PC_1:Active to SA_PC_3:Standby

This transition shall occur if:

a) a START STOP UNIT command with the POWER CONDITION field set to STANDBY is received;
b) a START STOP UNIT command with the POWER CONDITION field set to FORCE STANDBY is received; or
c) the Power Condition mode page standby timer expires.

0.0.2.4 Transition SA_PC_1:Active to SA_PC_4:Stopped

This transition shall occur if:

a) a START STOP UNIT command with the START bit set to zero is received.

0.0.3 SA_PC_2:Idle state

0.0.3.1 State description

While in this state, rotating media in block devices shall be active.

0.0.3.2 Transition SA_PC_2:Idle to SA_PC_1:Active

This transition shall occur if:

a) a START STOP UNIT command with the START bit set to one is received;
b) a START STOP UNIT command with the POWER CONDITION field set to ACTIVE is received; or
c) a command is received which requires the active power condition.

0.0.3.3 Transition SA_PC_2:Idle to SA_PC_3:Standby

This transition shall occur if:

a) a START STOP UNIT command with the POWER CONDITION field set to STANDBY is received;
b) a START STOP UNIT command with the POWER CONDITION field set to FORCE STANDBY is received; or
c) the Power Condition mode page standby timer expires.

0.0.3.4 Transition SA_PC_2:Idle to SA_PC_4:Stopped

This transition shall occur if:

a) a START STOP UNIT command with the START bit set to zero is received.

0.0.4 SA_PC_3:Standby state

0.0.4.1 State description

While in this state, rotating media in block devices shall be stopped.

0.0.4.2 Transition SA_PC_3:Standby to SA_PC_4:Stopped

This transition shall occur if:

a) a START STOP UNIT command with the START bit set to zero is received.

0.0.4.3 Transition SA_PC_3:Standby to SA_PC_5:Active_Wait

This transition shall occur if:

a) a START STOP UNIT command with the POWER CONDITION field set to ACTIVE is received; or
b) a command is received which requires the active power condition.

0.0.4.4 Transition SA_PC_3:Standby to SA_PC_6:Idle_Wait

This transition shall occur if:

a) a START STOP UNIT command with the POWER CONDITION field set to IDLE is received;
b) a START STOP UNIT command with the POWER CONDITION field set to FORCE IDLE is received; or
c) a command is received which requires the idle power condition.

0.0.5 SA_PC_4:Stopped state

0.0.5.1 State description

This state is only implemented in block devices.

While in this state, rotating media shall be stopped.

0.0.5.2 Transition SA_PC_4:Stopped to SA_PC_3:Standby

This transition shall occur if:

a) a START STOP UNIT command with the POWER CONDITION field set to STANDBY is received.

Editor's Note 1: FORCE_STANDBY and standby timer expiration purposely not included.

0.0.5.3 Transition SA_PC_4:Stopped to SA_PC_5:Active_Wait

This transition shall occur if:

a) a START STOP UNIT command with the START bit set to one is received; or
b) a START STOP UNIT command with the POWER CONDITION field set to ACTIVE is received (see SBC-2).

0.0.5.4 Transition SA_PC_4:Stopped to SA_PC_6:Idle_Wait

This transition shall occur if:

a) a START STOP UNIT command with the POWER CONDITION field set to IDLE is received (see SBC-2).

Editor's Note 2: FORCE_IDLE and idle timer expiration purposely not included.

0.0.6 SA_PC_5:Active_Wait state

0.0.6.1 State description

This state is only implemented in SAS devices.

While in this state, rotating media in block devices shall be stopped.

0.0.6.2 Transition SA_PC_5:Active_Wait to SA_PC_1:Active

This transition shall occur if:

a) a SPINUP is received.

0.0.7 SA_PC_6:Idle_Wait state

0.0.7.1 State description

This state is only implemented in SAS devices.

While in this state, rotating media in block devices shall be stopped.

0.0.7.2 Transition SA_PC_6:Idle_Wait to SA_PC_2:Idle

This transition shall occur if:

a) a SPINUP is received.