Designing Energy Efficient SATA Devices

Overview and Implementation Recommendations
2011 − 2013 Platforms

April 2011

Revision 1.63

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL® PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER, AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

UNLESS OTHERWISE AGREED IN WRITING BY INTEL, THE INTEL PRODUCTS ARE NOT DESIGNED NOR INTENDED FOR ANY APPLICATION IN WHICH THE FAILURE OF THE INTEL PRODUCT COULD CREATE A SITUATION WHERE PERSONAL INJURY OR DEATH MAY OCCUR.

The products described in this document may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order.

Intel and the Intel logo are trademarks of Intel Corporation or its subsidiaries in the U.S. and other countries.

*Other names and brands may be claimed as the property of others.

Copyright © 2007-2011, Intel Corporation. All rights reserved.

Implementation Recommendations

Contents

Figures

Tables

Revision History

Revision Number	Description	Revision Date
1.0	Initial release.	September 2007
1.1	Added sections on Staggered Spin Up	March 2008
1.2	Removed sections on Asynchronous Notification	March 2008
1.3	Miscellaneous Updates	April 2008
1.4	Added interlock switch platform recommendations	September 2009
1.5	Added battery life analyzer and platform 2013 recommendations	May 2010
1.6	Review Updates	July 2010
1.62	Miscellaneous Updates	November 2010
1.63		April 2011

§

1 Introduction

This document contains a series of recommendations designed to assist platform vendors, software developers and SATA device vendors to build power-friendly SATA-based platforms.

Serial ATA (SATA) is the next generation of storage interface technology that has replaced Parallel ATA (PATA) in mobile systems. SATA provides many new features including: increased data transfer rates with increased protocol efficiency; low pin count interconnect with thinner flexible cables and connectors; decreased signaling voltage; enhanced error detection; and a point to-point topology with the elimination of master/slave. It also includes support for Native Command Queuing, Native Hot Plug, Staggered Spin Up, and Port Multipliers.

These new features are designed to provide both immediate value to the SATA subsystem today, and allow for future growth in storage technology over the next several years.

SATA Link Power Management can be used to reduce the power consumption of the SATA interface, providing SATA capabilities at a minimum power cost.

§

2 SATA Link Power Management Overview

SATA Link Power Management (LPM) puts the physical layer (PHY) of the link into a low-power state. This PHY layer Link Power Management is independent of the ATA protocol power state of the disk, and as such complements the existing power management capabilities provided by the ATA command set. For example, the ATA command set reduces the power consumption of the attached device by issuing ATA protocol-level power state change requests to the disk. These requests typically instruct the device to spin down the media to save power. The rotational state of the media is completely independent of the state of the link.

Independent intelligent PHY power management has shown a significant reduction in the overall power consumption of the SATA subsystem, both in the platform and in the SATA device itself.

2.1 Link Power Management States

SATA provides two link power management states, in addition to the "Active" state. These states are "Partial" and "Slumber," that, by specification, differ only by the command sent on the bus to enter the low power state, and the return latency. Partial has a maximum return latency of 10 microseconds, while Slumber has a maximum return latency of 10 milliseconds. Typical host and device hardware implementations should be able to realize greater power savings in Slumber with its longer specified return latency. Partial, therefore, is designed to allow link power state transitions continually with minimal impact on performance. Slumber is designed to be used only when the link is expected to be idle for an extended period of time. It is not possible to transition the link directly from Partial to Slumber without passing through the Active state.

2.2 Host- and Device-Initiated Power Management

SATA Link Power Management can be broken up into two basic types: Host-Initiated Link Power Management (HIPM) and Device-Initiated Link Power Management (DIPM). SATA Link Power Management requires cooperation between the host and the device. Either the host or the device can request the link to enter a low-power state, but the corresponding host or device must accept or reject the link state change request. Each of these provides power savings by themselves; maximum power savings, however, are achieved when both are implemented together.

Host-initiated power management can be implemented either in the host hardware or the host software. In the first case, the host controller requests a link power management transition immediately after all outstanding commands to the SATA device have been completed. This allows the link to enter a low-power state immediately upon completion of the commands to the disk. Since the host has the best knowledge of what commands have been posted, or will be posted to the device, the host is able to make an immediate link power state change without invoking a time-out period.

The host controller on the host can automatically put the link into either Slumber or Partial after the command completes. Typically, for performance reasons, this will be Partial. However, after some period of idleness, it is generally assumed that the link will be inactive for an extended period of time and it is desirable to transition the link from Partial to Slumber. This can be done either by the host software or the device.

Device-initiated power management is implemented by the SATA device. The SATA device knows best how long a specific command might take to complete, and is best equipped to request a link power management state change while processing the command.

Since the host is best equipped to manage the PHY between commands and the device best within a command, the best power management is obtained when the host and device cooperate.

2.3 Link Power Management and Device State

The operating system can put devices into a D-state as defined by the Advanced Configuration and Power Interface (ACPI) specification. The link, however, can be put into its power management state independent of the D-state of the host controller or the device.

2.4 Power Management Protocol

The protocols used for entering and exiting Partial and Slumber states are different. To enter one of the low-power states, the standard communication protocol between the host and device can be used. Once the PHY has been placed into a low-power state, the standard communication protocol between the host and SATA device cannot be used; and another Out of Band (OOB) mechanism is required.

2.4.1 Entry Signaling Protocol

Either the host or the device can initiate a request to enter into the Partial or Slumber power state. The request is signaled by transmitting either the PMREQ_P or PMREQ_S primitive. PMREQ_P is used for a request to transition to the Partial power state, and PMREQ_S is used to transition to the Slumber power state. The corresponding host or

device must then respond with a PMACK acknowledge or PMNAK negative acknowledge primitive. If the request is acknowledged with a PMACK, both the host and device transition into the corresponding power state. To ensure the PMACK received is reliable without having to handshake the handshake, the target sends several PMACK's before entering a low-power state. If a PMNAK is received, no power state change occurs.

2.4.2 Exit Signaling Protocol

Once the PHY is in a low-power state, SATA specifies a low-level signaling mechanism to bring the interface back to active state. An OOB signal "COMWAKE" is sent to the device that acts as a wake-up call and causes communication to be reestablished.

2.4.3 Hardware/Software Control

Link Power Management is only enabled when the host controller and the device report that they are capable of issuing or receiving Link Power Management requests, and the OS software driver is capable of enabling the host hardware and SATA device. The host controller reports this capability to software in the Capabilities Register of the AHCI host controller. The SATA device reports the ability to support these commands in the IDENTIFY_DEVICE data structure returned by the device during device enumeration. The host must specifically enable DIPM on the SATA device via the ATA SET_FEATURES command upon initialization to enable DIPM on the SATA device. See the ATA and SATA specifications for details.

2.4.4 Listen Mode

The AHCI specification allows host software to put a port into listen mode when no device is connected to the port. A port in Listen mode has the equivalent power consumption on the host as a port in slumber, but allows the port to detect device insertions on that particular port.

2.4.5 Automatic Partial to Slumber

Automatic partial to slumber allows the host and the SATA device to do an automatic or implicit transition to slumber from partial without requiring the link to transition to active. An implicit transition means that the host or SATA device can take up to the 10 ms specified for slumber to return from the partial state. No explicit PMREQ_S commands are sent on the bus to transition it to slumber and no transition to the active state is required. As with DIPM, the SATA device sets a bit in the IDENTIFY DEVICE data structure to indicate that the device is capable of transitioning the link on the device. The device sets another bit in the IDENTIFY DEVICE data structure indicates that the SATA device allows the host to auto transition the link.

As with DIPM the host must specifically enable APS on the device via the ATA SET_FEATURES command.

2.5 Host vs. Device Link Control

There are two fundamental times the link can be put into slumber. These are between commands, or between command bursts, and within a command or command burst on the bus.

Between commands, the host is generally best equipped to put the link into a low power state. The host and host software has a wide variety of information available to it that it can use to determine the timing frequency of the transition into the two low power states.

Within a command or command burst, the SATA device is best equipped to transition the link into a low power state. The SATA device has advanced knowledge of the time it will take to respond to any specific command or command burst.

2.6 Host/Device Design Recommendations and Interaction

Between commands, the host hardware should transition the link into Partial after every command or command burst. Partial provides the best tradeoff between power savings and performance.

In the absence of a host-initiated power management transition, the device should attempt to transition the link into Partial after some appropriate short timeout.

After a longer period of timeout, it can be assumed that the link will remain idle for a longer period of time. Either the host, through the host software, or the device should at this point transition the link to Slumber.

Within a command, the SATA device should attempt to put the link into partial or slumber based on the time the SATA device will take to respond to the command.

2.7 Device Removal during Power Management

SATA Link Power Management disables the PHY on the SATA TX/RX transmit and receive lines between the host and device. Because the PHY is disabled, a device removal cannot be immediately detected without waiting for the next command to be sent to the SATA device. The AHCI specification provides the capability for a second electrical signal to be provided to the host controller to notify the host that the SATA device has been removed. Typically this will be implemented as a Mechanical interlock switch on the mobile platform.

3 Recommended Host and Device Behavior

The 2013 platform will contain additional power savings features in the core subsystem that require the SATA links to be in the low power state. The following recommendations will help OS and device vendors to take advantage of these capabilities.

3.1 Recommended Host Behavior

Host Software should enable Host Initiated power management on the host with the following capabilities listed below.

If the device supports Device Initiated Link Power Management, the host should enable device Initiated link power management on the device.

3.1.1 Host Behavior between Commands

The host software should enable Aggressive Link Power Management on the host controller, and set the host controller to auto transition to partial. Or the host SW should have the link enter partial as quickly as possible within 1-2 micro seconds of the command completion.

Host software should attempt to put the link into slumber after 10ms of inactivity delay.

If the device supports APS the host should enable APS on the device, and do an implicit (no PM_REQ) transition of the link into slumber after a 10ms timeout.

APS support in the host is optional.

3.1.2 Host Behavior within a Command

The device is best equipped to manage the link within a command. The host software should not put the link into partial or slumber within a command.

3.2 Recommended Device Behavior

Because not all currently shipping OS's and OS policies support HIPM slumber transitions, the device should support Device Initiated Power Management slumber transitions in the device. Device support of APS is optional.

3.2.1 Between Commands

The device firmware should put the link into partial very quickly (ideally in a few micro seconds after the command completion). This delay allows the host to put the link into partial if HIPM is enabled in the system.

The device should also transition the link into slumber after 10ms of inactivity.

In devices that support APS, and in systems where the host has enabled APS on the device, the device should do an implicit (no PM_REQ) transition to slumber after 10 ms.

Devices should not NAK host link requests at any time

3.2.2 Device Behavior within a Command

Within a command the device should put the link into partial or slumber immediately after the command is received, potentially taking into account the expected response time of the device and the expected response time of the system.

Within a command, the policy device uses to put the link into partial or slumber is determined by the device implementation. A simple implementation that puts the link into partial immediately after receiving a command is probably adequate for most device implementations. The suggestions below are designed to assist a device vendor in implementing a best of class power and performance implementation.

Max resume latency of the core system and the link combined is expected to be 1 millisecond for slumber, and 100 μ second for partial.

Partial is recommended, slumber is optional and for many devices the power savings may not be worth the implementation cost of slumber within a command.

Devices that support APS and slumber should support an explicit slumber transition using PM_REQ on the bus, rather than an implicit (no PMREQ) transition on the bus.

The following diagram shows a typical device response to a command. The diagram shows 3 major events: 1) when the command is initially received by the device, 2) when the device should begin to wake up the link and 3) when the device attempts respond to the host with data.

Figure 3-1. Typical Command Response by Device
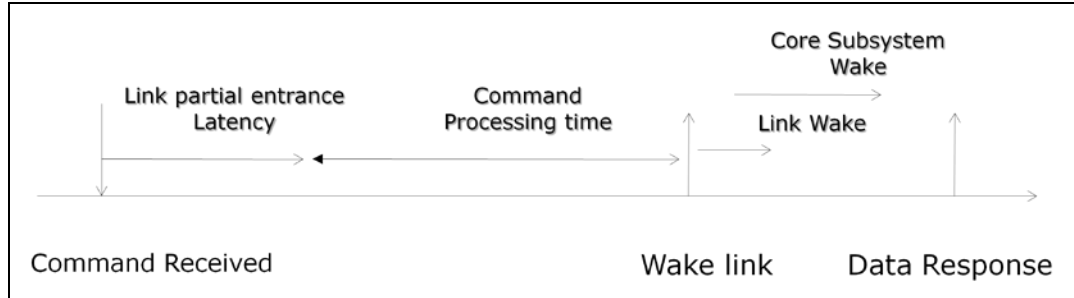
After the command is received, the link takes some time (typically 1 microsecond) to enter into a low power state. Then there is some time the device requires to process the command. There is some time to wake up the link, and then there is some time that the core subsystem may take above and beyond the link wake up time before data can be transferred from the device to the host.

The device should enter a low power state within a command when the link entrance latency plus the combined link wake latency and core subsystem wake latency are less than the time the device will take to respond. The device may attempt to wake the link some time before the data is ready to transfer to ensure the host system is ready to receive the data when the device is ready to send it.

3.3 The Following Table Summarizes the Host and Device Recommendations

Table 3-1. Link Recommendations between Commands (No APS)

No APS	Host Initiated	Device Initiated	Comments
Slumber Timeout	10 ms	10 ms	Explicit Slumber transition using PMREQ_S
Partial Timeout	<1 µsec (ALPM) 1-2 µsec (SW initiated)	As soon as possible (within microseconds)	5 µsec allows host to transition to partial first

Table 3-2. Link Recommendations between Commands (APS enabled)

APS	Host Initiated	Device Initiated	Comments
Slumber Timeout	Implicit – 10 ms	Implicit – 10 ms recommended device time out	Implicit Slumber transition
Partial Timeout	<1 µsec (ALPM)	As soon as possible (within microseconds)	

Table 3-3. Link Recommendations within Commands (w/wo APS)

	Host Initiated	Device Initiated	Comments
Slumber Timeout	None	Optional. Use and duration of slumber calculated by device assuming 1 millisecond max system resume latency. Devices that enable APS should use explicit slumber transition.	Power savings probably not worth the implementation effort
Partial Timeout	None	Calculated by device assuming a maximum 100 µsec system resume latency	Partial timeout is recommended

4   *Implementation Recommendations*

SATA LPM is an optional feature of the SATA specification. As such, implementing this technology correctly in the platform requires cooperation between the BIOS, Platform Hardware, OS storage driver, and the SATA device.

In particular:

- The host controller in the Intel chipset must inform OS driver software that it is capable of implementing Link Power Management
- The SATA device must inform the OS driver software that it is capable of implementing Link Power management
- The OS driver software must support LPM

4.1   System BIOS Support

Full support of the SATA Link Power Management requires that the SATA host controller be put in AHCI mode. The BIOS should support setting the Host Controller in this mode.

Also, because LPM is an optional feature in the SATA specification, not all AHCI host controllers are required to support it. The host controller must declare whether it supports the feature to SW. In Intel chipsets, this is done through BIOS at configuration time.

These specific AHCI Host Controller capability registers must be set to 1:

- CAP.SALP – informs the SW that the host Controller is capable of supporting Host initiated, hardware initiated (called Aggressive Link Power Management in the AHCI specification) in the Host controller Hardware.
- CAP.SMPS – reports whether the AHCI host controller is capable of supporting an interlock switch
- CAP.SSS – Informs the SW that the host controller is capable of supporting Staggered Spin Up, and enables the Intel® RST driver to put the ports in listen mode. Staggered Spin Up also requires that the BIOS initially spin up the SATA devices (PxCMD.SUD) after host controller reset.
- CAP.SSC – Informs the SW that the host controller is capable of supporting slumber state transitions on the SATA link.
- CAP.PSC – informs the SW that the host controller is capable of supporting partial state transitions on the SATA link

In addition to the Host Controller capability registers the following registers must be set on a per port basis to inform the host software of the physical configuration of the SATA device attached to the port

- PxCMD.MPSP – reports whether there is an interlock switch implemented in the specific SATA port
- PxCMD.HPCP – reports whether the port is hot plug capable.
- PxCMD.ESP – reports whether the port is an ESATA port

As described in the AHCI specification, PxCMD.HPCP indicates that the device on the SATA port is removable from the system, and that the signal and power is on the same physical connector.

PxCMD. ESP indicates that the device on the SATA port is connected on an ESATA connector. An ESATA connector is signal only, and the device is independently powered through another source.

For further details please consult the *Intel BIOS Writers Guide*, available from your Intel Field Service Representative.

4.2   Platform Hardware (Interlock Switch)

Link Power Management puts the SATA link in a neutral logic state. With the signal lines in a neutral logic state, it is impossible for an AHCI Host Controller to detect when the device has been removed from the bus. As such currently shipping OS system software can't detect when the SATA device has been removed until it tries to access it again.

An interlock switch is dedicated hardware, resident in the platform that asserts a signal to the chipset that notifies the driver that the device may have changed state. This hardware is platform dependent, and asserting this signal does not specify that a hot plug event necessarily occurred, only that a hot plug event may have occurred.

The interlock switch asserts the AHCI Mechanical Presence bits in the AHCI register space. See the AHCI specification for details.

Intel® RST will use this signal to determine whether or not the SATA device is present in the system when the link is in a low power state. If the platform reports that an interlock switch is built into the system, Intel® RST will enable LPM on ports that are hot swap capable.

If the SATA device is connected to the host through a SATA slimline connector, the existing Device Present Pin on the connector can be coupled with a pull up resistor on the motherboard to provide the necessary interlock switch signaling to the chipset.

This provides an inexpensive way to provide the necessary interlock switch capability without the cost of an independent switch.

4.3   Operating System Storage Driver

The Operating System Storage driver level software must also support AHCI mode, and must enable LPM in the host controller hardware, if the host controller hardware reports it's capable of supporting LPM. We recommend that host software implement both HIPM partial and HIPM slumber transitions as described in this document.

Figure 4-1. Interlock Switch on Slimline Connected Drives
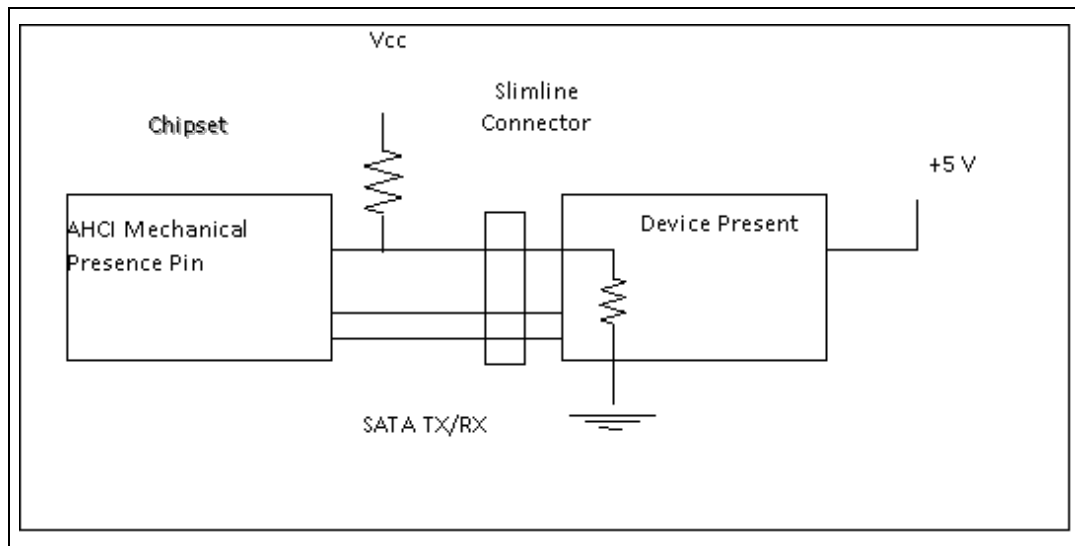
Intel® RST software supports the native SATA AHCI interface with all of the native SATA AHCI capabilities, including Link Power Management.

4.4   Intel® Rapid Storage Technology (Intel® RST)

4.4.1   Link Power Management

Intel® RST implements both host-initiated and device-initiated power management. For host-initiated power management, the driver enables hardware-initiated (called Aggressive Link Power Management in the AHCI specification) via the AHCI PxCMD.ALPE bit. It configures the hardware to enter into Partial after every command or when command queue is empty (NCQ). (PxCMD.ASP in the AHCI specification.)

Current versions of Intel® RST shipping in the 2011 time frame do not transition the link to slumber after transitioning the link into partial but future versions of Intel® RST may do so.

Intel® RST will enable Device Initiated Link Power Management in SATA devices that report they are capable.

4.4.2 Registry Settings

This default behavior is controlled and alterable by registry settings. These registry settings can be created to alter the behavior of the Intel® RST driver for Mobile ICH6M, 7M, 8M, and 9M, and PCH I/O controller hub SKUs of the chipset. They can also be created in 2011 platform desktop chipsets.

Registry settings may be freely modified from their default settings for test purpose. For product based implementations please consult the Intel® RST team for details.

4.4.2.1 Host-Initiated Power Management

KEY_LOCAL_MACHINE\System\CurrentControlSet\Services\iaStor\Parameters\Portx, where 'x' can be from 0 to 31.

Table 4-1. Host Initiated Power Management Settings

Key Name	Values	Default	Definition
LPM	1 = Enabled 0 = Not Enabled	1	Enables Host Initiated Link Power Management
LPMSTATE	0 = Partial 1 = Slumber	Partial	Specifies whether the hardware transitions to partial or slumber on command completion

4.4.2.2 Device-Initiated Power Management

Table 4-2. Device Initiated Power Management Settings

Key Name	Values	Default	Definition
DIPM	1 = Enabled 0 = Not Enabled	1	Specifies whether the driver enables Device-Initiated Link Power Management on the device: **1** means the driver enables Device-Initiated Link Power Management on the device **0** means the driver does not enable Device-Initiated Link Power Management the device

4.5 SATA Device

For the best implementation of LPM, both Host Initiated and Device Initiated Link Power Management should be supported in the SATA device. A SATA device must declare to the driver the features it will support in its IDENTIFY Device data structure.

4.5.1 Host-Initiated Link Power Management

The SATA device must report it is capable of supporting Link Power Management in Word 76 bit 9 of the IDENTIFY_DEVICE data Structure. Once the capability has been reported to the host, the host may send link power management requests to the SATA device. The SATA device must acknowledge the requests to achieve the power savings.

4.5.2 Device-Initiated Power Management

The SATA device must declare support and current status in the IDENTIFY_DEVICE data structure. Word 78 bit 3 indicates whether the feature is supported, Word 79 bit 3 indicates whether the feature is currently enabled. The driver enables the features by setting sector Count 03 in SET FEATURES command

5 Battery Life Analyzer Tool

The Intel Battery Life Analyzer tool provides insight into the configuration and current operation of the power savings features of the SATA subsystem. This document is designed to provide supplementary information to the BLA tool to assist system manufacturers in tuning and configuring their systems for maximum power savings in the SATA subsystem.

The battery life analyzer shows the following information for the SATA subsystem

Figure 5-1. Battery Life Analyzer Example Output
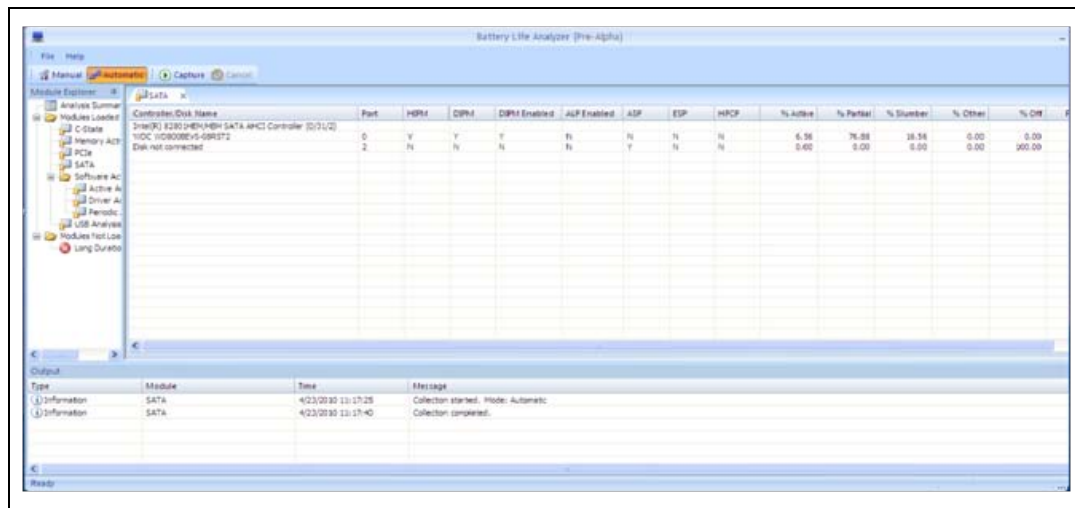
Below is a list of the various values in the table, with an interpretation of the value, and a recommendation for system integrators

Table 5-1. Battery Life Analyzer Output Recommendations

Column Title	Interpretation	Recommendations
Controller/Disk Name	The SATA Controller or SATA disk name on that specific port	
Port	The specific Port number that the SATA device connected to on the AHCI host controller	
HIPM	Does the Device Report that the device supports HIPM in the ATA IDENTIFY DEVICE data structure	Devices that support both HIPM and DIPM are recommended. Consider purchasing a device that supports both

Column Title	Interpretation	Recommendations
DIPM	Does the Device Report that the device supports DIPM in the ATA IDENTIFY DEVICE data structure	Devices that support both HIPM and DIPM are recommended. Consider purchasing a device that supports both
DIPM Enabled	Has DIPM been enabled on the device by the host?	If this is N, either the device does not support DIPM, or the SW driver has not enabled it. Verify your driver and OS system configuration
ALP enabled	Has Aggressive Link Power Management been enabled on the host controller?	Should be yes. If not please consult your BIOS writers guide for more details
ASP	When Aggressive Link Power Management is enabled, determines which link state (partial or slumber) the link should transition to after each command. Y = Slumber, N = Partial	Should be set to N (Partial)
ESP	Is this an ESATA port?	If the port is routed to an ESATA connector, this bit should be set to Yes. If not, the bit should be set to No.
HPCP	Is the port Hot Plug capable?	If the port is hot plug capable, then the bit should be yes. If not the bit should be set to No.
% Active	Shows how much time the link is in active during the capture run	Idle systems should have the link in slumber > 90% of the time
% Partial	Shows how much time the link is in partial during the capture run	Idle systems should have the link in slumber > 90% of the time
% Slumber	Shows how much time the link is in slumber during the capture run	Idle systems should have the link in slumber > 90% of the time

5.1 Debugging SATA LPM issues with BLA

5.1.1 SATA Link States

On an idle system (one with very little activity and very little SATA traffic) run a capture and look at the link states. You should run this test on a machine with the OS and drivers installed, but very few applications loaded on the system, and very little traffic to the disk. In this configuration all ports should be in slumber most of the time.

Hard drives used as the boot drive should show a > 90% slumber residency when the machine is idle and the disk not accessed. Hard drives not used as a boot drive should show 100% slumber.

Optical drives on OS's that poll the optical drive should show ~90% slumber on that port when the drive is unused. Optical drives on systems that support AN should show ~100% slumber when the drive is unused.

If the slumber state residency for the port is less than these numbers, then the following recommendations will help you achieve the maximum power savings.

5.1.2   Port Recommendations

ESATA ports cannot have SATA LPM enabled on them. A port that is not physically connected to an ESATA port cannot be marked ESATA.

Hot Plug Capable Ports cannot have SATA LPM enabled on them unless there is an interlock switch configured to the port as described in this document. Only Ports connected to drives that are physically removable should have the hot plug capable bit set. Ports connected to drives that are not physically removable should NOT have the bit set.

If the port is physically removable, and interlock switch should be configured in the system as described in this paper.

5.1.3   Device Recommendations

Make sure the device supports both HIPM and DIPM.

5.1.4   DIPM not Enabled on the Device

Make sure the device has DIPM enabled on it.

Intel® RST in its default configuration will enable DIPM when the drive supports it and when the bits in the Port Recommendations above are configured to allow LPM.

Other OS's or drivers may have different policies. For instance at the time of the development of this paper Microsoft Windows 7* will support Link Power Management differently based on the OS Power Policy setting. In particular Microsoft Windows 7 will support only HIPM in the balanced and high performance power policies. It will support HIPM and DIPM in the power saver power policy.

5.1.5   Device Behavior

If the platform complies with the above recommendations, and the link is still not showing significant slumber state residencies, then it is possible that the device is not

effective in transitioning the link into a low power state, or is not allowing the host to put the link into a low power state. There are two things to check for in driver design

1) The device claims to support DIPM and does not initiate a slumber transaction, or does not do so quickly enough. If the device does not initiate a slumber transaction, or does so after a very long timeout, the link may not transition as efficiently into the slumber state as otherwise. We recommend that the drive transition the link into slumber after about 100 ms in the 2011-2012 timeframe and 10 ms in the 2013 timeframe.

2) The device NAKs requests from the host to go into either the partial or slumber state. We recommend that the drive not NAK any of the partial or slumber requests from the host.

Contact the device vendor for more information on the behavior of the device. If necessary, debug of these problems will require a SATA line analyzer.

6 Compliance and Testing Recommendations

These recommendations are designed to be a minimum set of recommendations that will assist all SATA device vendors in implementing a SATA device that will work in all systems with a variety of BIOS and Software configurations. They are not intended to be a comprehensive validation plan for the SATA device, or even LPM within the SATA device. Platform vendors may choose to implement some or all of these recommendations in a peripheral qualification plan.

6.1 Devices

6.1.1 Key Challenges for Device Vendors

The key challenge a device vendor faces is that SATA LPM is an optional feature in the SATA specification. The Intel® RST will implement all of the features of LPM, but not all platform's BIOS or other drivers will. It is critical therefore that the SATA device be validated in all possible combinations of LPM and other optional features of SATA such as Native Command Queuing, Asynchronous Notification, etc.

Two critical levels of validation should take place – functional validation and power tuning.

Functional validation means that the SATA device continues to operate under all possible combinations of optional features in the SATA device. For instance, the SATA spec defines several of these optional features including: Native Command Queueing, Host Initiated LPM, Device Initiated LPM, and Asynchronous notification. Each of these may or may not be enabled by the host SW, or potentially the host BIOS. A validation matrix should include all of these features.

Experience has shown that if a comprehensive test of all possible combinations of these optional features is not performed there is a significant risk of a SATA device firmware bug that will prevent reliable operation of the SATA device in some scenarios. This will result in either a low adoption rate of the SATA device in the marketplace, or more importantly, reluctance on behalf of the Platform Vendor or OS/Driver vendor to implement the LPM on a widespread basis.

In addition to functional validation a SATA device must be competitive in the marketplace, that is, the SATA device must show good value when compared against its peers under typical configurations.

6.1.2   Hard Drive Functional Testing

A Hard drive must, of course, operate cleanly in each of the following configurations of HIPM, DIPM and NCQ. Typical system configurations that are known to exist in the industry today are called out. This table is illustrative and does not cover all system configurations that might exist.

Table 6-1. Hard Drive System Configurations

HIPM	DIPM	NCQ	Typical System Config
0	0	0	IDE Compatible Driver, Intel® RST with limited system configuration, or equivalent configurations with other OS drivers
0	0	1	Intel® RST with limited system configuration, or equivalent configurations with other OS drivers
0	1	0	IDE Compatible Driver with DIPM enabled, Intel® RST with HIPM disabled, or equivalent configurations with other OS drivers
0	1	1	IDE Compatible DIPM and NCQ enabled, or Intel® RST with HIPM disabled, or equivalent configurations with other OS drivers
1	0	0	Intel® RST with limited system configuration, or equivalent configurations with other OS drivers
1	0	1	Intel® RST with limited system configuration, or equivalent configurations with other OS drivers
1	1	0	Intel® RST with limited system configuration, or equivalent configurations with other OS drivers
1	1	1	Intel® RST or equivalent

The drive should be particularly power optimized for the last configuration.

APS even though not part of this table should be tested as well. The drive should remain functional whether or not the host enables APS on the drive, or implements it on the host.

Functional testing is as simple as running a validation suite under the various configurations.

Power optimization testing should include using a SATA line analyzer to view the trace and make sure that:

1. The drive acknowledges the LPM requests from the host
2. Implements an appropriate slumber requests to keep the link in a low power state.

6.1.3   Optical Drives

Optical drives must also operate in all possible combinations of the following optional features.

Table 6-2. Optical Drive System Configurations

HIPM	DIPM	Typical System Config
0	0	IDE Compatible Driver, Intel® RST with limited system configuration, or equivalent configurations with other OS drivers
0	1	IDE Compatible Driver, Intel® RST with limited system configuration, or equivalent configurations with other OS drivers
1	0	IDE Compatible Driver, Intel® RST with limited system configuration, or equivalent configurations with other OS drivers
1	1	Intel® RST or equivalent

Optical drives should be particularly power optimized for the last configuration as it is the most common / preferred implementation.

Asynchronous Notification and APS, even though not part of this table, should be tested as well. The drive should remain functional whether or not the host enables APS on the drive, or implements it on the host.

Power optimization testing should include using a SATA line analyzer to view the trace and make sure that:

3. The drive acknowledges the LPM requests from the host
4. Implements an appropriate slumber requests to keep the link in a low power state during long periods of idleness

§