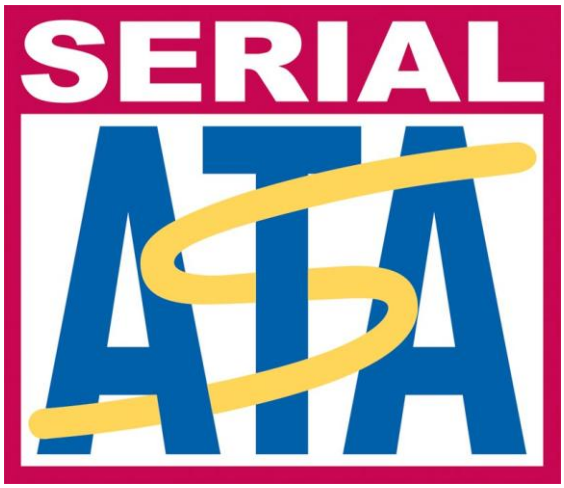
Serial ATA
Device Sleep (DevSleep) and Runtime D3 (RTD3)



December 2011

A WHITEPAPER BY:
Intel Corporation
SanDisk Corporation

Table of Contents

Overview

In the past, storage devices attached to a platform were (typically) placed into a power off state as a result of platform power state transitions associated with:

- Platform hibernation
- Platform standby
- Platform off

With the introduction of a new generation of mobile devices that focus on power efficiency and responsiveness, the ability of new platforms and operating systems to closely manage the power of one or more storage devices attached to the platform independently of other platform components (e.g. USB, networking, graphics, etc.) is becoming increasingly important.

This paper provides a basis for understanding the SW/HW ramifications regarding the power management of Serial ATA devices on a new generation of platforms through the use of the new SATA Device Sleep (DevSleep) feature. Additionally this paper will provide a basis for understanding the Runtime D3 feature and how it differs from, and can be used in conjunction with, DevSleep.

Keywords

The following terms will be used throughout this paper:

- **ATA/ATAPI Command Set** – A standard that specifies the AT Attachment (ATA) command set that is used to communicate between host systems and storage devices. Available from www.t13.org.
- **D0**[1] – Device power state in which a device is on and running. It is receiving full power from the system and is delivering full functionality to the user. All devices must support this power state.
- **D0active** – Device power state where the device has been configured and enabled by software and is functional.
- **D3hot** – Device power state that occurs when a device transitions to D3, yet still has Vcc applied.
- **D3cold** – Device power state that occurs when a device transitions to D3, but Vcc is not applied
- **HBA** – Host Bus Adapter. In this paper the HBA refers to the host hardware that is used to communicate with a SATA storage device.

[1] A detailed description of system states S0, S3, S4, S5 and device states D0, D3hot and D3cold can be found in the Advanced Configuration and Power Interface (ACPI) Specification (www.acpica.org) and the PCI Bus Power Management Interface Specification (www.pcisig.com)

- **IDENTIFY DEVICE** – A command that is defined via the ATA/ATAPI Command set. Used by a host system to retrieve a storage device's IDENTIFY DEVICE data which contains information regarding optional feature or command support provided by a storage device.
- **PBA** - Pre-Boot Authentication: A program used for authenticating the system's user and unlocking the storage device.
- **Runtime D3 (RTD3)** – Refers to the placement of a device into D3hot/cold while the rest of the platform remains in a S0 state.
- **S0** – System power state. While the system is in the S0 state, it is in the system working state. Device states are individually managed by the operating system software and can be in any device state (D0 or D3).
- **S3** - System power state (also referred to as system sleeping state). While the system is in S3, the processors are not executing instructions and power is usually removed from the devices; system DRAM context is maintained. Some system BIOS is usually required to initialize the system on transition to S0.
- **S4** – System power state (also referred to as hibernation). While the system is in S4, the processors are not executing instructions and power is usually removed from the devices; system DRAM context is not maintained. System BIOS is required to initialize the system on transition to S0 (i.e. Power-On Self-Test (POST)). System context is maintained via non-volatile memory (e.g. storage device).
- **S5** – System power state (also referred to as soft-off). Similar to S4 except that system context is not maintained via non-volatile memory.
- **SET FEATURES** - A command that is defined via the ATA/ATAPI Command set. Used by a host system to enable certain device features.
- **Zero Power ODD (ZPODD)** – Refers to a class of optical disk drives that can be placed into D3cold and then placed back into D0 as a result of a tray open/close event.

Device Sleep (DevSleep)

The need to consume less power and provide extended battery life is a critical part of today's mobile devices. To meet the ever more aggressive power/battery life requirements in this new environment, the SATA interface is evolving. DevSleep is a new addition to the SATA specification, which enables SATA-based storage solutions to reach a new level of low power operation.

Interface Power Management in SATA today

In the SATA spec today, the host or device can place the interface (PHY) into reduced power states as follows:

1. Partial - PHY is in a reduced power mode; exit time < 10 microseconds

2. Slumber - PHY in a reduced power mode (lower power than Partial); exit time < 10 milliseconds

The tradeoff for going into a reduced interface power state is that the SATA device cannot respond as quickly to commands as when the interface is fully powered; that is, there is a tradeoff of exit latency vs. power savings when using interface power states. Both the Partial and Slumber interface states use so-called “in band” signaling; that is, the commands used by the host and device to change the interface power state are transmitted over the SATA bus itself.

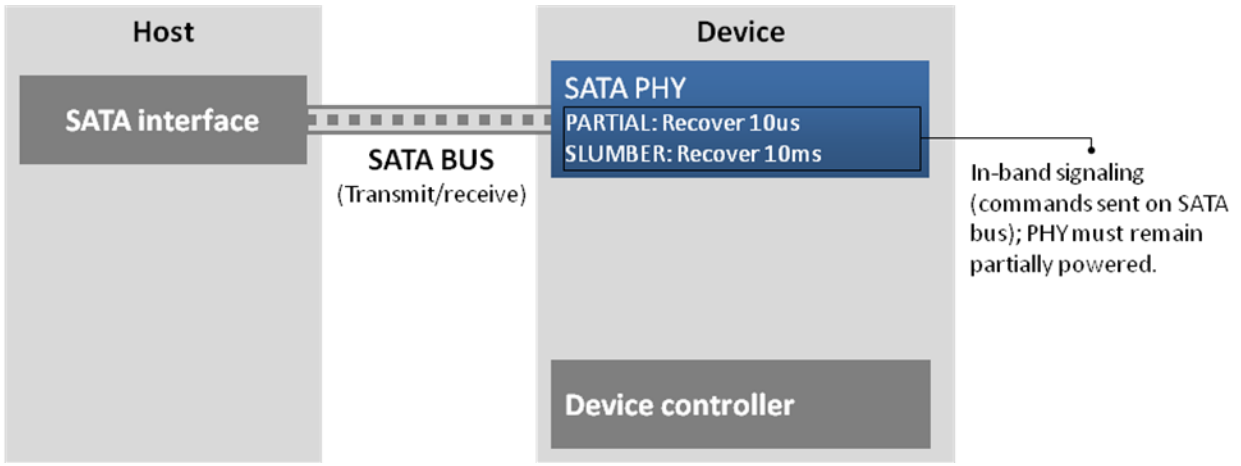


Figure Error! No text of specified style in document.**-1. Host-Device In-Band Signaling**

The existing “in-band” SATA power management scheme means that the SATA PHY cannot be fully powered down; it must remain powered to process the state change commands. Thus, if a host wishes to save additional power on the SATA interface, then only two options exist:

1. From the host controller side, the associated host port can be placed into the offline state (if supported by the SATA HBA architecture). Unfortunately, the placement of a host SATA port into offline mode does not also place the associated storage device’s PHY into a similar state.

2. Completely power off the SATA device (as shown in the figure below); however, this can cause much longer exit latency, and may even use more net energy (depending on the frequency of power removal) than remaining in Partial or Slumber, due to powering the entire device off and back on. Removal of power from a device is sometimes referred to as Runtime D3 (RTD3). This is discussed in the Runtime D3 (RTD3) section of this white paper.

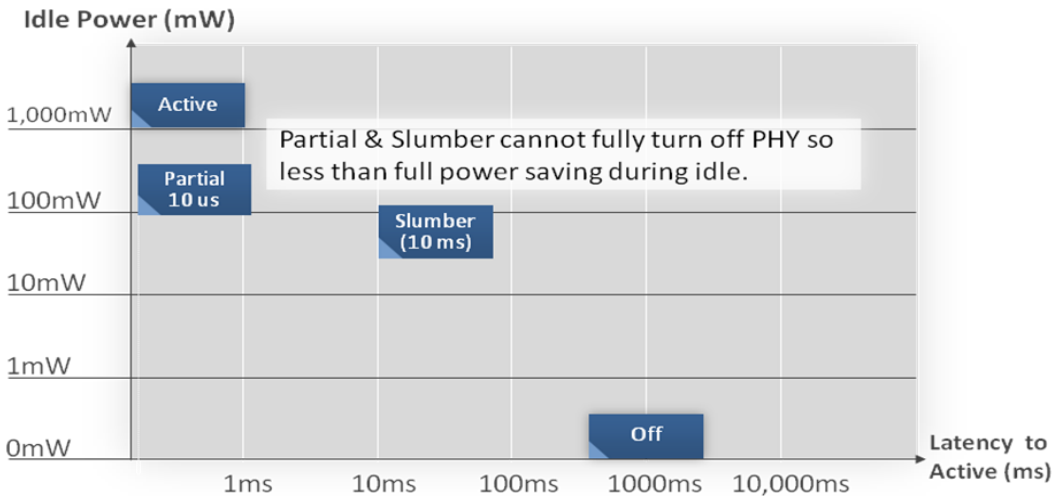


Figure Error! No text of specified style in document.**-2. Power vs. Latency With RTD3**

DevSleep – Less Power, Fast Exit Latency, Better overall energy profile
With the addition of DevSleep, hosts/devices have a new power management option in which the host and device can completely power down their respective PHY's and other link-related circuitry. Devices might also choose to power down additional sub-systems while in DevSleep, enabling even further power savings. However, the exit latency, and overall transition energy to/from DevSleep, is much lower compared to RTD3.

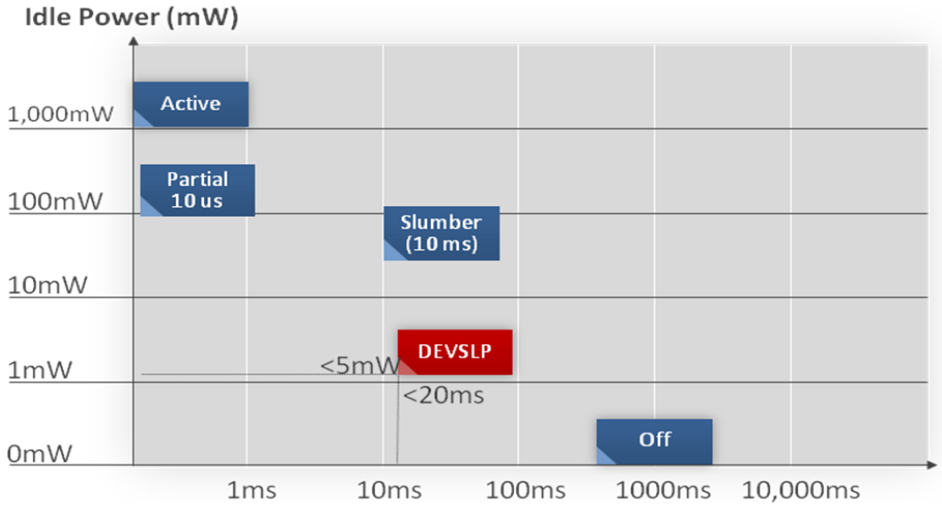


Figure Error! No text of specified style in document.**-3. Power vs. Latency With RTD3**

and DevSleep
With DevSleep enabled, a host has a middle ground between today's interface power management states (Partial and Slumber) and "off" (RTD3). It can now go into a low latency power mode where both the host and device PHY can be completely powered off, as well as possibly other sub-systems, but still maintain an exit latency much closer to Slumber than to a full shutdown (RTD3).

The DevSleep specification does not state what power levels a device will reach while in the DevSleep state, but SSDs are targeting 5mW or less.

DevSleep Theory of Operation
DevSleep works by defining a new signal (DEVSLP) which is connected between the host and storage device. When the host asserts the DEVSLP signal, the device enters the DevSleep interface power state for as long as the host asserts the DEVSLP signal. When the host negates the DEVSLP signal, the device returns from the DevSleep state. The DevSleep specification allows implementers flexibility regarding what the device actually does when the DEVSLP signal is asserted. The device may completely power down its PHY, and it may also choose to power down other subsystems, as long as it can meet the exit latency requirements.

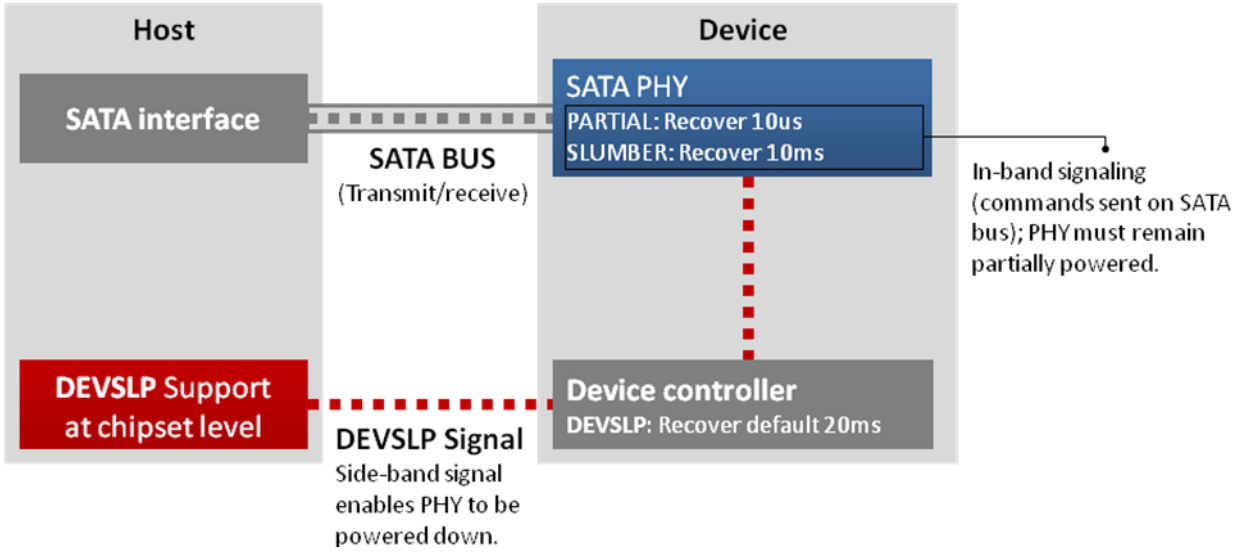


Figure Error! No text of specified style in document.**-4. DEVSLP Signal**

DevSleep operates as follows:

The host may assert the DEVSLP signal from any state, provided that:

- Device supports the Device Sleep feature (per ATA IDENTIFY DEVICE command)
- The Device Sleep feature is enabled by host (per ATA SET FEATURES command)
- There are no commands outstanding

On DEVSLP Assertion

- Host must assert DEVSLP for >= 10ms, or as specified in Identify Device Data Log;
- Host and device may power down PHY and other systems (e.g., PLL's, clocks, media);
- Neither host nor device shall initiate PHY communications while DEVSLP asserted
- All PHY communications ignored by host and device while DEVSLP asserted

On DEVSLP Negation

- Device must detect OOB in <= 20ms, or as specified in Identify Device Data log
- Host and device can use COMWAKE or COMRESET/COMINIT for renegotiation

Runtime D3 (RTD3)

While the new DevSleep feature provides excellent low power and exit latency characteristics it is still a fact that energy will continue to be consumed as long as Vcc is applied. Complete removal of power from devices that have been idle for long periods of time (e.g. hours, days) enables the maximum possible power conservation for platforms with stringent power requirements.

It is possible to completely remove power (D3cold) from devices while the system remains in a powered state (S0); this is sometimes referred to as runtime D3 (RTD3) support. This is typically done to conserve power on the system during long periods of device idleness. The ability to support RTD3 is dependent on both system (e.g. the system chipset, the motherboard, ACPI etc.) and OS support. While this paper specifically addresses storage devices, RTD3 can be supported by other, non-storage devices present in a system.

A system with RTD3 storage support implements the following:

1. A storage controller that can be put into D3cold; depending on implementation, the controller may support placement into D3cold or it may support D3hot.
2. One or more storage devices whose power can be programmatically applied and removed (usually via a power FET implemented on the system board).

Any non-optical storage device attached to a platform with storage subsystem RTD3 support can be used without specific hardware modifications. However, for best end-

user experience, storage device manufacturers should optimize their device's power-on to ready latency so that the transition from D3 to D0active is not humanly discernible. Optical devices that support RTD3 should be compliant with the Mt. Fuji Commands for Multimedia Devices Version 7 INF-8090i v7 (refer to Appendix K - SATA ODD Zero Power Effort Notes). These devices (referred to as ZPODD) implement specific mechanisms that allow the ZPODD to be placed into RTD3 when the device is idle/unused and then subsequently placed into D0Active when the end user presses the media eject button.

DevSleep and RTD3 are orthogonal technologies and can function independently of one another. It is anticipated that a storage subsystem with these technologies will use them in a complimentary manner.

For example, a platform designer may omit power FETs on non-ZPODD SATA port(s) if a storage device with DevSleep support provides extremely low power usage (e.g. 0.5mW) when DevSleep is active. This type of strategy is beneficial for several reasons:

- The omission of a FET reduces system BOM cost
- DevSleep exit latency is generally less than the latency associated with the powering up (D0) of a device from a fully powered down state (D3cold).
- DevSleep exit energy consumption is generally less than the energy required to bring a device out of D3cold.

Storage Device RTD3 and Device Context

Unlike Partial, Slumber and DevSleep, a device placed into RTD3 completely loses its context; any settings made by the host prior to placing a storage device into RTD3 must be restored by the host during re-application of power: These include (but are not necessarily limited to) the following ATA commands (refer to the ATA/ATAPI Command specification available from www.t13.org):

- SET FEATURES
- DEVICE CONFIGURATION FREEZE LOCK
- SET MAX FREEZE LOCK
- SET MAX ADDRESS (If V_V attribute is not used)
- SECURITY FREEZE LOCK
- Etc.

Additionally, when ATA security or Opal security is enabled on a storage device, upon entry into RTD3 the storage device reverts to the locked state. Upon application of power to the storage device, the host is required to unlock the storage device before attempting to access it. Unlike when the platform transitions from S3/S4/S5->S0, there

is (typically) no system BIOS participation (other than via ACPI) involved in taking a storage device out of RTD3. This implies that BIOS cannot be used to collect the user credentials (when a storage device has ATA security enabled) nor can BIOS be used to load and execute a PBA associated with an Opal Security enabled storage device – these activities, if applicable, are required to be handled by the host OS.

SATA HBA RTD3

The concept of RTD3 is not limited to the children (e.g. storage device) of a parent device (e.g. SATA HBA); it can also be applied to the parent device itself. While the mechanism and policy for placing a SATA HBA is out of scope for this paper, there are some device, PHY and HBA power states that are valid. The following figure shows the valid combinations:

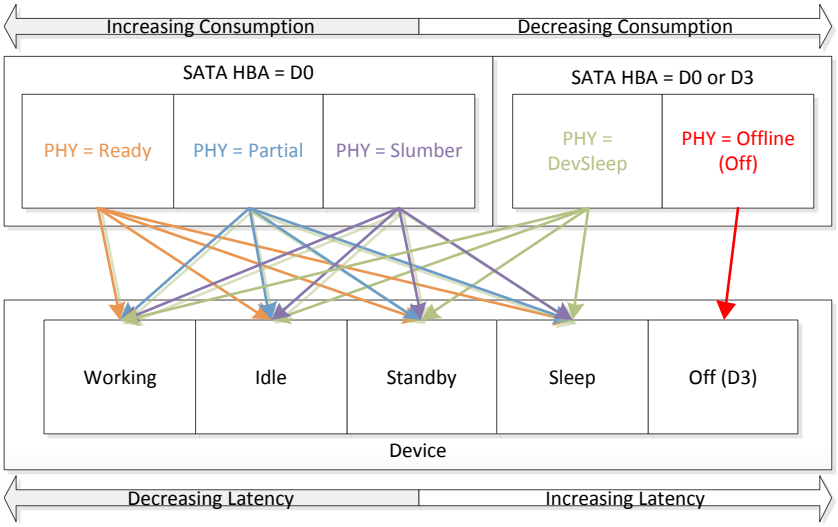


Figure Error! No text of specified style in document.**-5. Device-HBA Power Combinations**

Notes:

- A device is considered in the Working state when it is not in Idle, Standby, Sleep or Off. While in the Working state the device may or may not be actively processing host initiated commands. It is a requirement that the host shall not initiate DevSleep until the device has completed all commands issued by the host.
- Sleep invoked via ATA SLEEP command sent to device.
- Standby invoked via ATA STANDBY (IMMEDIATE) command sent to device.
- Idle invoked via ATA IDLE (IMMEDIATE) command sent to device.
- Device working state occurs when device not in IDLE, STANDBY, SLEEP or Off (RTD3).
- It is not typical to place the SATA HBA in D3hot/cold while the attached device(s)

remains in D0 due to lack of perceived value (DevSleep may be an exception depending on device's power consumption when in DevSleep). Typically when the storage HBA is in D3hot/cold, then the associated storage devices are also in D3cold.

Summary

To meet the ever more aggressive power/battery life requirements of today's mobile devices, the SATA interface is evolving with the addition of the DevSleep interface power state. DevSleep enables hosts and devices to completely shut down the SATA interface, saving more power vs. the existing Partial and Slumber interface power states, which require that the PHY be left powered. DevSleep will help to enable a new generation of power friendly SATA-based mobile devices.

Additionally, new power stringent systems will place an increased emphasis on the ability of the system and OS to completely power down storage devices and the storage controller during long periods of idleness. Implementing both DevSleep and RTD3 support on the platform provides for a hierarchical power management solution that allows the system to choose between power and latency in a dynamic and efficient manner.