Western Digital

White Paper

# Storage Power Efficiency Improvement With HDD Idle Modes

Hard disk drive power-saving modes offer the ability to significantly reduce power consumption for applications where periods of inactivity are common.

*October 2023*

## Introduction

Data center power usage—and the carbon emissions associated with generating that power—take on tremendous importance in our data-fueled world. The International Energy Agency estimates that data centers represented 1-1.5% of global electricity use in 2021. With rapid growth seen between 2015 and 2021, that number may increase markedly in the future.

There are many very high-powered compute devices in a data center, such as CPUs and GPUs. In comparison, storage devices individually consume much less power than a CPU or GPU, but in most data centers they are far more numerous. Power consumed is costly, and power consumed generates heat, which further requires more power to cool the data center. Thus, improving power efficiency is of paramount importance to improve both cost and environmental impact, on a global scale.

Much thought is given to the active power consumption of devices in the data center. In order to ensure a return on investment on high-cost and high-power components like CPUs and GPUs, those components must be used as much and as intensely as possible. But there's another side to the data center—the data. Data storage devices serve those high-cost components with the data they need to act upon, but much of the storage in a data center may spend most of its time idle, with accesses only intermittently made to read "cold" data.

For data storage devices, the question becomes "How do I save storage power when I'm not actively using my storage?" Data storage devices—hard disk drives (HDDs) in particular—are already designed to minimize power when idle. But they are also designed with additional power-saving modes that may require customer action to use. These more advanced "idle" modes can offer significant power savings for applications where the application demands allow.

## HDD Idle Modes

Software designers can reduce HDD power consumption at the drive level by employing the use of power saving modes, either by issuing commands to the drive to proactively enter idle states, or by setting drive timers which inform the drive to enter these states after certain durations of inactivity.

There are several power saving modes available for HDDs:

- Idle_A: Drive Ready, but not performing IO, drive may power down selected electronics to reduce power without increasing response time
- Idle_B: Spindle rotation at 7200 RPM with heads unloaded
- Idle_C/Standby_Y: Spindle rotation at Low RPM with heads unloaded
- Standby_Z: Actuator is unloaded and spindle motor is stopped. Commands can be received immediately

Each power saving mode progressively saves more power, but the exit latency or the amount of time the drive takes to be able to service new commands, grows larger. Table 1 shows the power consumption in each mode, the default timer setting for entry, and the exit latency, for the 22TB[1] Ultrastar® DC HC570 HDD:

| Power Saving Mode | Power (W, SATA) | Power (W, SAS) | Default Timer (SATA) | Default Timer (SAS) | Exit Latency (typical, sec) |
|---|---|---|---|---|---|
| Idle_A | 5.7 | 6.0 | 2 sec | 2 sec | 0 |
| Idle_B | 3.8 | 4.1 | 10 min | 60 min | 1.5 |
| Idle_C/Standby_Y | 3.3 | 3.6 | Never | Never | 4 |
| Standby_Z | 1.3 | 1.6 | Never | Never | 15 |

Table 1: Idle Modes for the Ultrastar 22TB HC570 HDD

These power-saving modes offer the ability for a data center operator to get the most power efficient operation of the drive in cold and archival workloads. But it is critical to fully understand the workload to ensure that Quality of Service (QoS) and drive specifications can simultaneously be met.

## Balancing QoS with Power Savings and Drive Specifications

To utilize these idle modes, many competing factors must be balanced. In all cases it requires very detailed understanding of the drive workload and access patterns from the applications that utilize the HDDs.

The first consideration is QoS, which is the application requirements on how quickly data must be accessed. If data needs to be returned in milliseconds, it is not possible to use Idle_B, Idle_C, or Standby_Z idle modes. In that case, the point is moot and none of the advanced idle modes are compatible with the application's QoS. On the contrary, if an application has relaxed QoS requirements, for example an ad-supported video streaming site, the time necessary to exit a low-power mode can be easily met by playing an ad while the user waits for data. The wide panoply of possible applications will have a wide range of QoS requirements.

The second consideration involves frequency of access. This impacts both selection of idle modes and how much power can be saved using them. We often think of individual pieces of data in terms of warm or cold. However, this thought is not applicable to the scale of modern storage. Modern HDD capacities necessitate a deeper look. A 22TB drive will contain many pieces of data. If each piece of data is 1MB, for example, the drive will have **22 million** individual pieces of data residing on it! If each piece of data is deemed cold, only accessed once per year, that means the drive will have 22 million reads during the course of a year, *or more than 41 times per minute*.

1 One terabyte is equal to one trillion bytes. Actual user capacity may be less due to operating environment.

Frequency of access determines both the power savings and whether you can safely invoke specific power modes. Enterprise-class HDDs have two specifications which particularly govern idle modes: load/unload cycles and motor start/stop cycles. The drives are typically rated for 600,000 load/unload cycles during their life, and 50,000 start/stop cycles. Because Idle_B (and beyond) invokes a load/unload cycle, and because Standby_Z invokes a start/stop cycle, it is important for the health of the drive to ensure that idle modes are not invoked too often such that they exceed the specifications. But if they are not invoked often enough, the power savings will be minimal.

As an example, consider a company offering offsite backup as a service. In this case, data stored in offsite backup is not a primary storage or usage location. Data is typically written once, and the only purpose to read the data is if there is a failure of the primary storage, such as disaster recovery, ransomware, or human error corrupting/deleting the primary storage. For such an application, utilizing idle modes could save significant power.

The key aspect is to evaluate how often drive accesses occur and determine if using short-duration timers for modes like Idle_B can be invoked without exceeding the drive specifications. If data is accessed more often than every 4.4 minutes on average, unloading the heads will exceed 600,000 load/unload cycles within the drive's 5 year warranty period. If data is accessed only once every 10 minutes, however, unloading the heads can be done safely while remaining within the spec.

In an application averaging one access every 10 minutes, relying on the drive's default Idle_B timer will save very little power. The drive will only enter Idle_B when 10 minutes have elapsed, and statistically will be very near to the next drive access which will return the drive to Idle_A. The 1.9W of power saving will not be used for the majority of the time the drive is idle. But if the offsite backup company sets the timer more aggressively, such as 30 seconds, the drive will be in Idle_B, on average, during roughly 95% of its usage time. That 1.9W of power savings 95% of the time will result in 15.8 kWh of power saving over a year. Multiplied across thousands of HDDs, the savings on power and cooling can add up to significant levels.

Such a calculation can additionally change the way the application is designed. For offsite backup, the demands of the application that lead to excessive drive access is unlikely to be reads; the data is only read as an exception. However, the ingest of data from customers all across a country—or even globally—might preclude the use of idle modes. However, if the application is designed with an ingest tier and a storage tier, data can be staged in the ingest tier until it reaches a certain utilization and then moved to the storage tier only when needed. In this case, it may be possible to even achieve drive access frequencies that enable the use of Standby_Z mode, saving 4.4W per drive and—since idle time will likely be well in excess of 95% of drive usage—power savings approach 38.5 kWh per year per drive. Across a large data center, the ingest tier might literally "pay for itself" in power cost savings.

For an example[2] hyperscaler datacenter, the Idle_B power savings in the above drive access scenario could translate to an annual savings of $4.7M, while the Standby_Z power savings would translate to an annual savings of $11.4M respectively, as described in Table 2. Furthermore, if we consider the power savings over the typical 5 year lifetime of a HDD, the savings scale up by a factor of 5.

| Power savings when compared to Idle_A | Idle_B | Standby_Z |
|---|---|---|
| Annual power savings per drive (kWh) | 15.8 | 38.5 |
| Annual HDD power savings for an example hyperscale datacenter (MWh) | 46,944 | 114,391 |
| Annual savings for an example hyperscaler datacenter (USD) | $4.7M | $11.4M |

Table 2: Power Savings Potentially Realized by Invoking Idle Modes[2]

Thus, it is incumbent on the application designer to understand access frequency in great detail, to determine whether idle mode power savings are practicable for their own application. As shown by the example, the sensitivity to configuration of how an application is set up may be significant. Care must be taken to ensure that power savings do not push drives beyond their load/unload or start/stop cycle specifications. Of course, idle modes must also meet the QoS demands of the application.

2 Data center with 62 MW total power consumption, containing ~2.97M HDDs, with a power cost of $0.10/kWh.

## Implementation Options

The power saving modes described above are standardized within ATA (SATA) and SCSI (SAS) command sets, and can be set by the user via standardized means. For ATA, this is handled using the Extended Power Conditions (EPC) portion of the Set Features command, and for SCSI, this is handled using SCSI mode page 1Ah, Power Control. A user has two options to cause the drive to enter power saving mode.

- Utilizing power condition timers: The power condition timers can be set based upon a specific time since the last host access to automatically enter a power saving mode. They can be set individually for each idle and standby mode named above.
- Commanded entry: The host can send a command instructing the drive to immediately enter a power saving mode. This can be commanded into any of the idle or standby modes named above.

Because of the short default timer (2s) to enter Idle_A, it is unlikely that a user will need to either change the timer or command the drive to enter this state.


There are free open-source industry standard tools available which provide the ability to set or enter these modes.

For SATA drives, the tool *hdparm* offers options for both commanded entry into the Standby_Z power mode as well as setting the Standby_Z timer. However, it does not provide the ability to command entry into or set the timers for Idle_A, Idle_B, or Idle_C. Changing these will require a user to issue the Set Features command directly. Western Digital customers can contact their Field Applications Engineer for assistance if needed.

For SAS drives, the tool *sg_modes* (part of the *sg3_utils* package) allows the user to easily read mode page 1Ah. The tool *sg_wr_mode* offers the ability to directly write the bits in mode page 1Ah to set the timers. The tool *sg_start* proves access to the SCSI START STOP UNIT command, which will allow commanded entry into desired power modes.

The description of the commands are covered in detail in the Product Manual for SATA and SAS Ultrastar HDDs available at the Western Digital web site. For SATA, the Set Features command described in the Ultrastar manual is applicable to other Western Digital SATA HDDs brands such as WD Gold®, WD Red®, or WD Purple®.

## Conclusion

Saving power is a goal that affects the world's economy and the planet's future. While much attention is devoted to increasing the efficiency of data centers during activity, it might be true that the low-hanging fruit is to reduce power consumption of devices when they are inactive. Hard disk drive power saving modes offer the ability to significantly reduce power consumption for applications where periods of inactivity are common.

## Learn More

"IEA: Data Centers and Data Transmission Networks"
https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks

"How To Choose Storage Solutions For Your Datacenter"
https://www.westerndigital.com/solutions/data-center

Western Digital.

5601 Great Oaks Parkway
San Jose, CA 95119, USA
www.westerndigital.com